

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」

HOMCOSサーバを用いた複合体立 体構造の検索とモデリング

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

創薬等支援技術基盤プラットフォーム

解析拠点

解析拠点は、タンパク質の構造解析に供する試料の調製、タンパク質の立体構造解析及び計算科学を活用したバイオインフォマティクス等に関する技術や施設及び設備等を一貫して提供し、外部研究者等のタンパク質立体構造解析研究を支援します。

生産領域: タンパク質試料の調整
解析領域: タンパク質構造解析
バイオインフォマティクス領域: 構造予測等の計算化学

制御拠点

制御拠点は、創薬シーズ等の探索のために、化合物ライブラリーとスクリーニングの技術基盤や施設及び設備等と、化合物の最適化や新規骨格の構築等を行う合成技術の基盤等を一貫して整備して外部研究者等に提供します。

ライブラリー・スクリーニング領域: 化合物ライブラリーの提供スクリーニング機器の共用
合成領域: ヒット化合物の最適化

情報拠点

情報拠点は、タンパク3000プロジェクト及びターゲットタンパク研究プログラム並びに平成23年度創薬等支援技術基盤プラットフォームの成果からなるデータベースやソフトウェアを管理・運用します。また、それらを継続的に更新し、内容の拡充や高度化を行います。

情報領域: データベース、解析ツールの提供

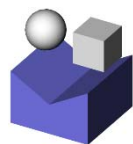
プラットフォームをご利用希望の方は各拠点情報のページから詳しい支援メニューをご覧になり、ご利用を希望する拠点の問い合わせ窓口または総合窓口(全般的なお問い合わせ)からお問い合わせください。また、お申し込みは各拠点情報のページにあるお申し込みフォームからご登録ください。

情報拠点: 生物学者のニーズを的確に捉えた 情報解析ツールの開発がミッション

- ➡ アンケートに、ご要望・ご批判をたくさん書いていただけるとありがたいです
- ➡ 立体構造を用いた情報解析(モデリング、ドッキング等)の個別相談も受け付けています。

今日の内容

1. ホモロジー・モデリングによる複合体の立体構造予測とは？



2. HOMCOSを用いたタンパク質の結合分子の予測



3. HOMCOSを用いたヘテロ複合体構造の予測

**MODE
LLER**

Modellerによる二量体のモデリング



4. HOMCOSを用いた化合物-蛋白質複合体構造の予測

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	非経験的方法 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	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,...

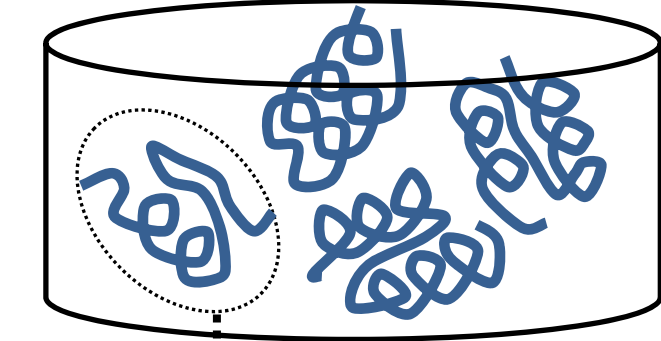
ホモロジー・モデリングによる3次構造予測

原理：立体構造はアミノ酸配列より保存しやすい。

予測対象配列（クエリ配列）

LNVANGKSVIGPALLEEVSGRD

立体構造データベース



テンプレート
構造

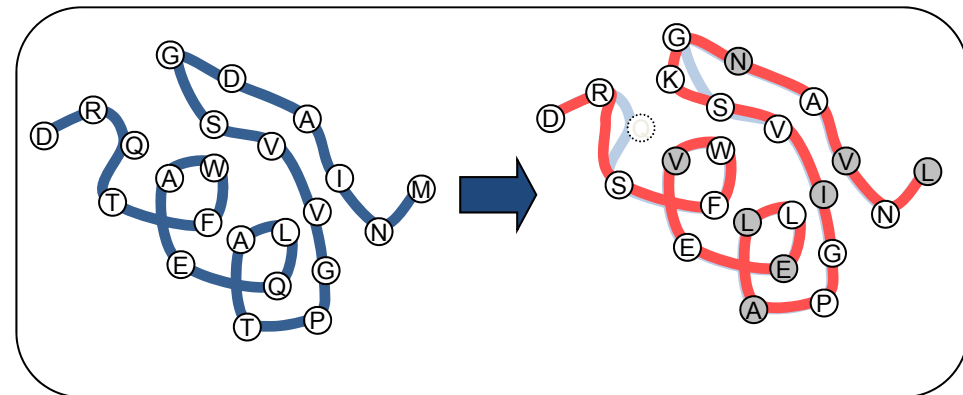
ステップ1：フォールド認識

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

BLAST, プロフィール法, スレディング法....



鋳型（テンプレート）構造とそのアラインメント

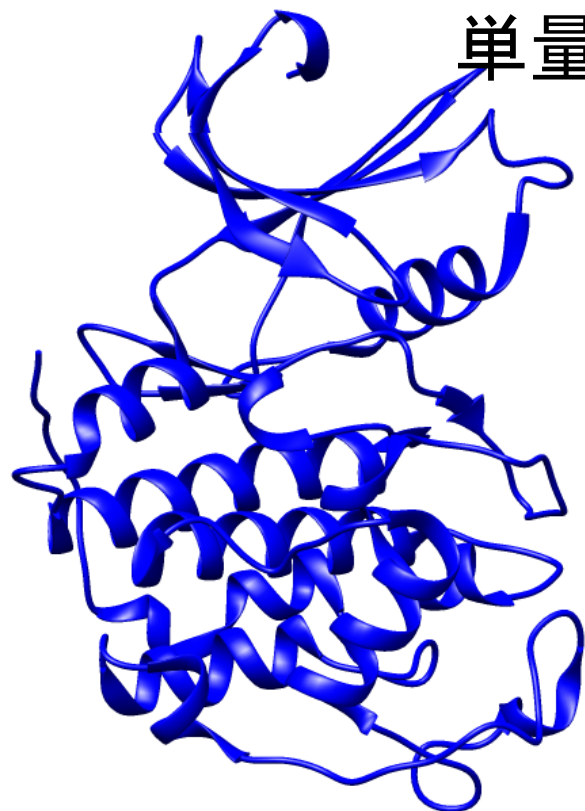


ステップ2：モデリング

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

MODELLER, FAMS,

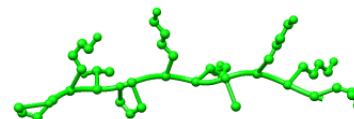
複合体立体構造は 単量体構造より機能情報が豊富



Cyclin-dependent protein kinase (CDK2)



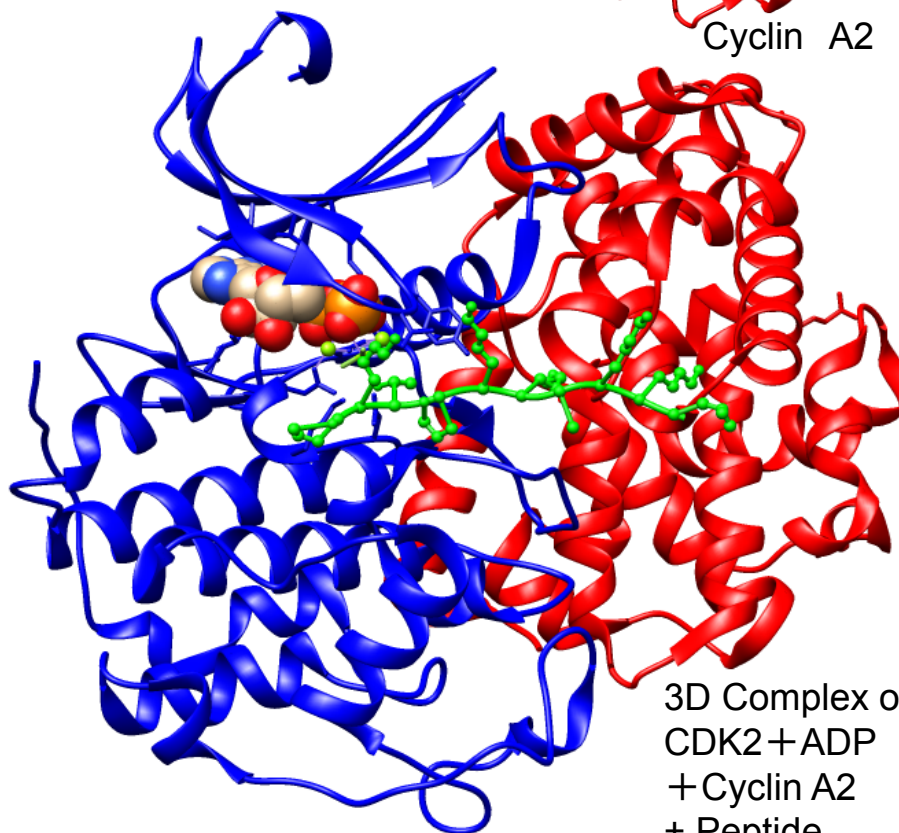
ADP



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

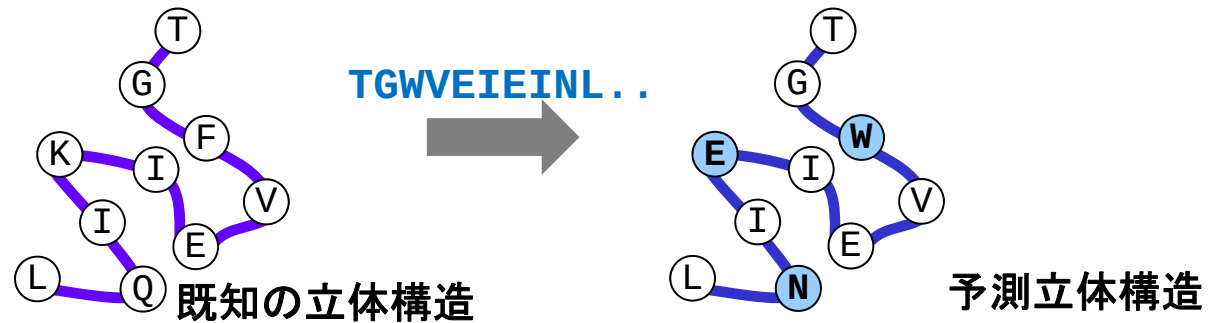
→阻害剤の設計・改変

(2)結合・反応のメカニズムの理解

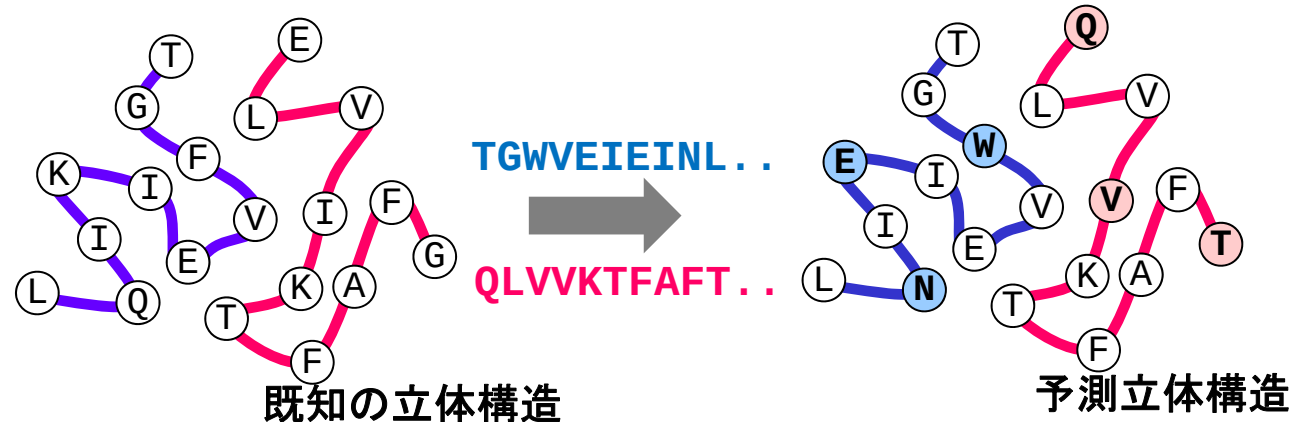
鋳型ベースのモデリング: Template-based Modeling

複合体のホモロジー・モデリング

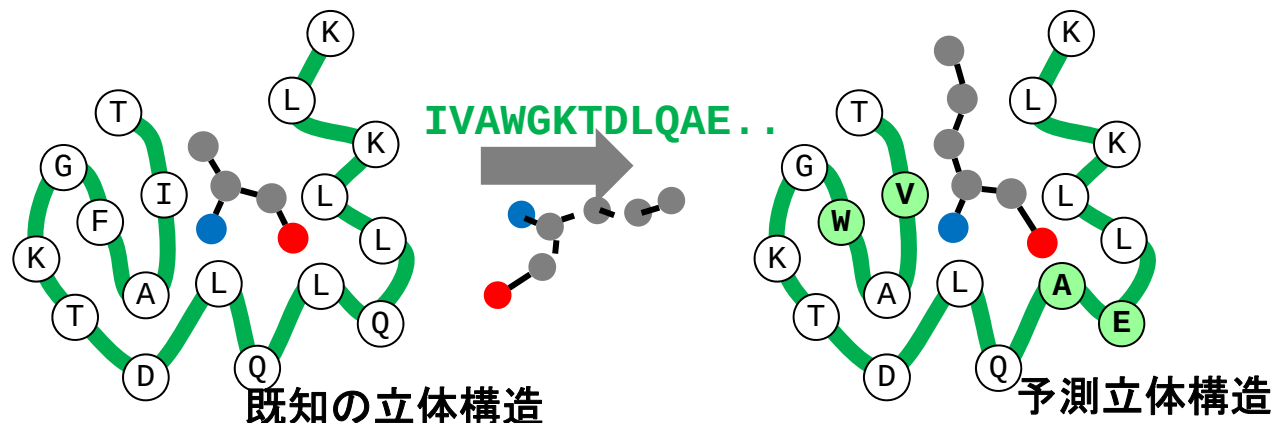
タンパク質単量体



タンパク質-タンパク質
複合体

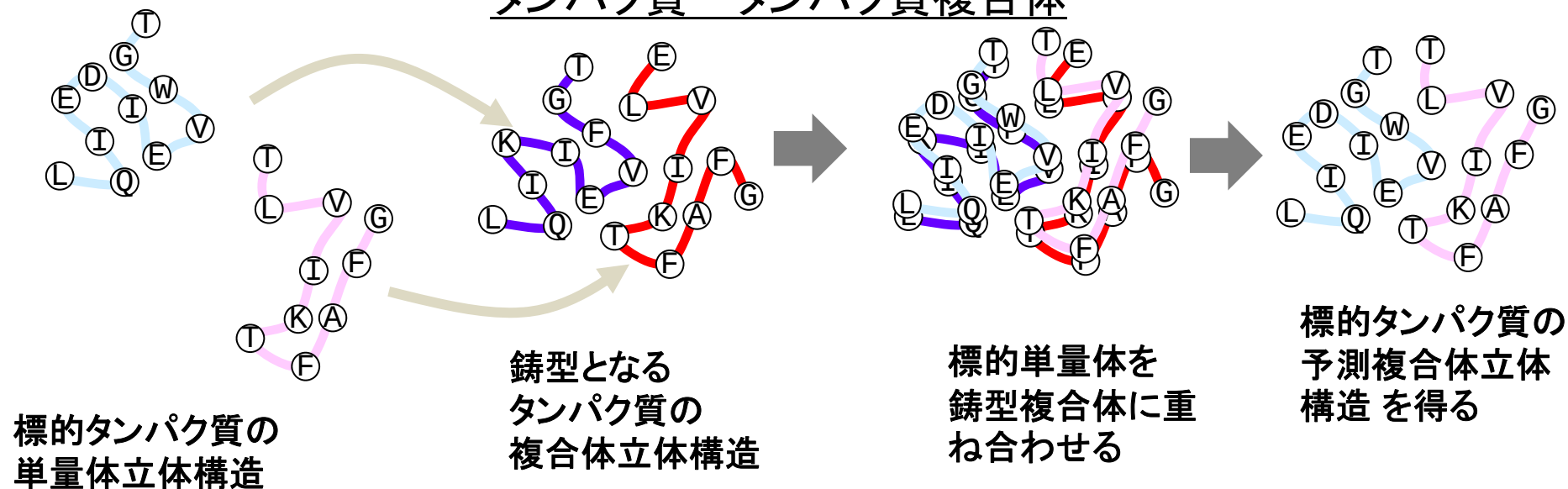


化合物-タンパク質
複合体

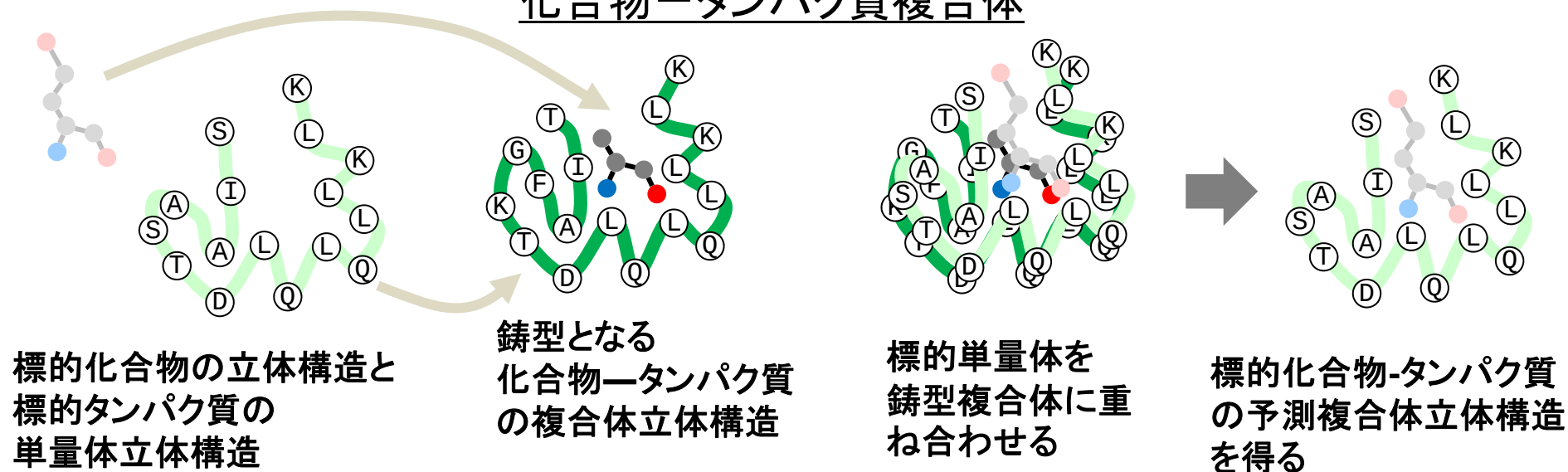


鑄型ベースのドッキング:Template-based Docking

タンパク質-タンパク質複合体



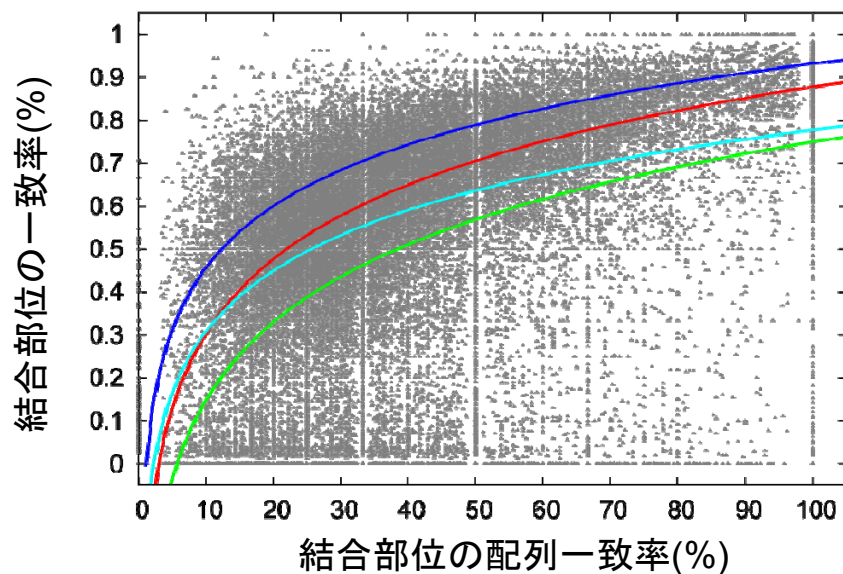
化合物-タンパク質複合体



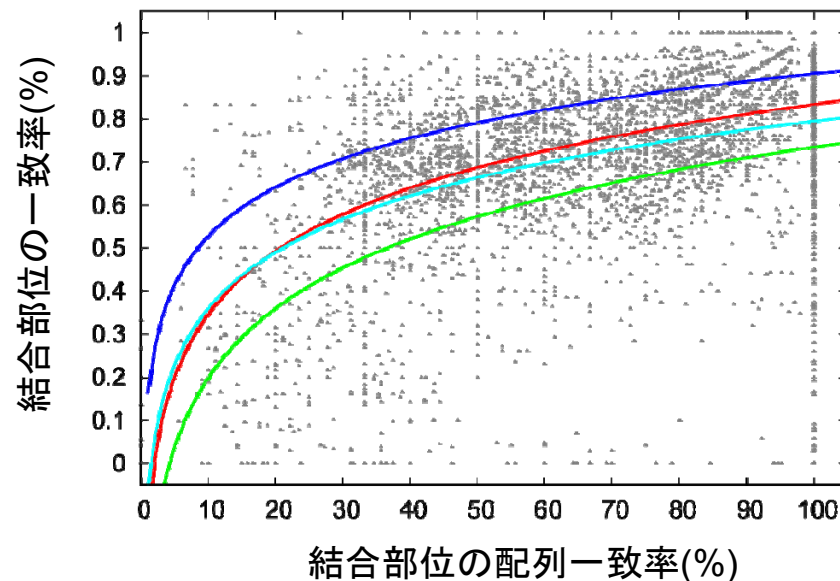
鋳型ベースの複合体予測の精度の見積もり

PDB内の相同な二量体間の相互比較

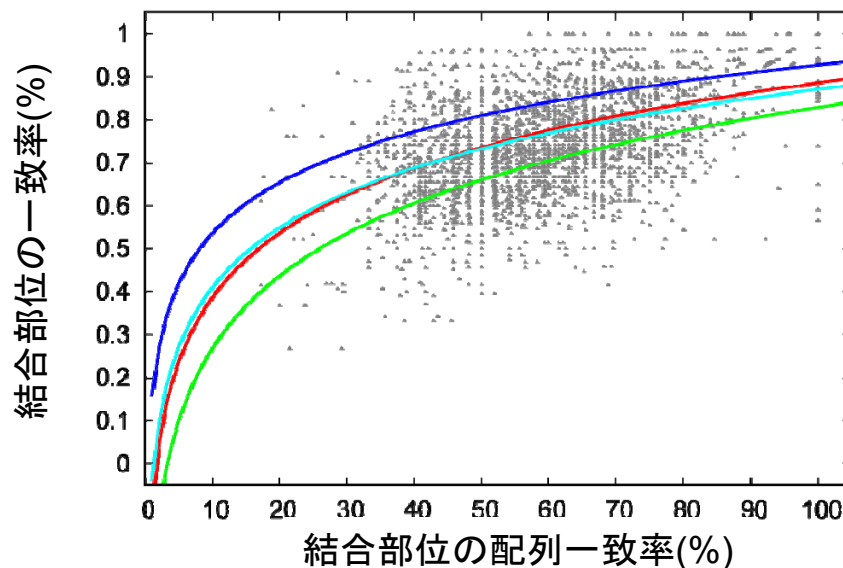
ホモ蛋白質二量体



ヘテロ蛋白質二量体



ヘム-蛋白質



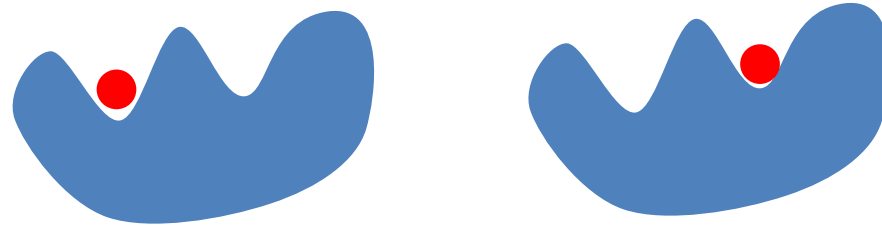
- メディアン
- 下側 25%-パーセンタイル
- 上側 25%-パーセンタイル
- 平均

複合体の場合も、予測精度はおおむね
配列一致度と相関する

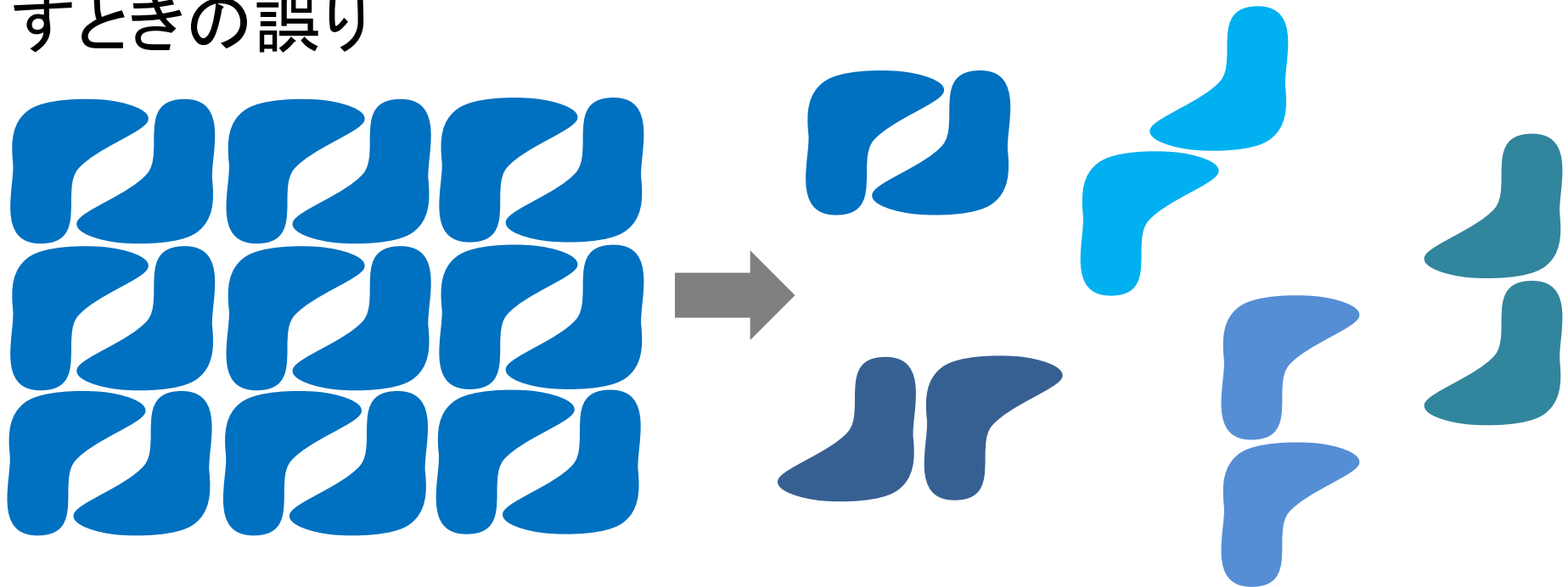
ただし、ホモ蛋白質二量体や、金属イオンで
結合部位が全く異なる予測をする場合もある

似た蛋白質でも結合部位が異なることがある理由

1. 一つのタンパク質に複数の結合部位がある



2. 結晶から 生物学的単位(biological unit)を切り出すときの誤り





HOMCOSを用いた 蛋白質の結合分子の予測

HOMCOS :複合体立体構造の検索・ホモロジーモデリングのサーバ

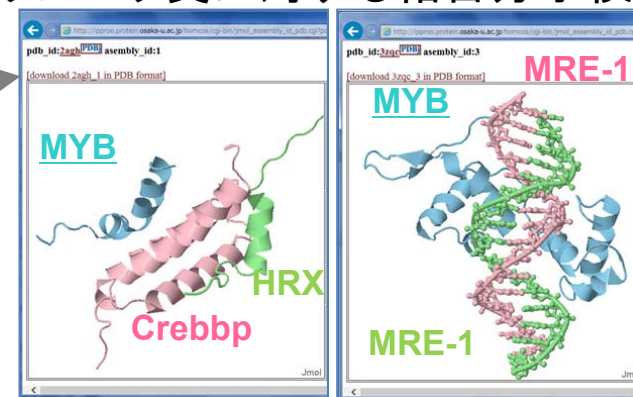
- ・PDB内の複合体の立体構造データを検索し、それを鋳型にモデリングする
- ・配列相同性検索は**BLAST**、化学構造類似性検索は**KCOMBU**を使用

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

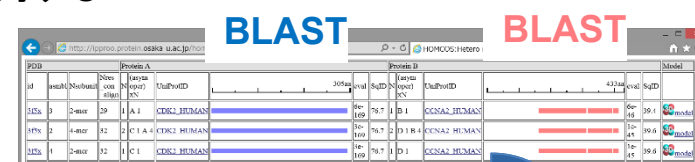
VaProSからの検索では、 **3D Interaction**に相当

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

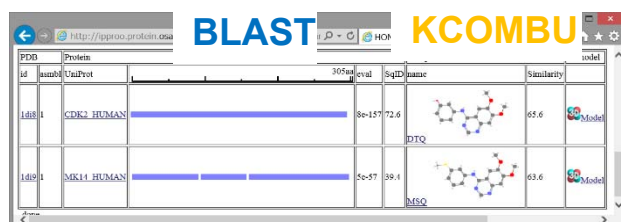
タンパク質に対する結合分子検索



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

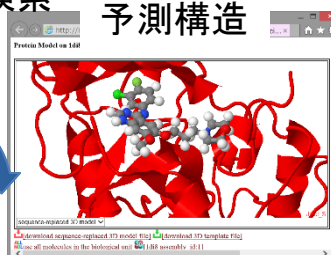


化合物-タンパク質複合体のモデル
 アミノ酸配列と化合物構造を入力、アミノ酸配列はBLASTで、化学構造はKCOMBU でPDBに対して検索

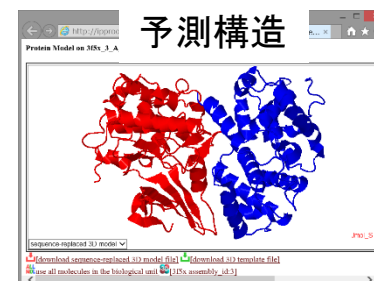


鋳型構造

予測構造



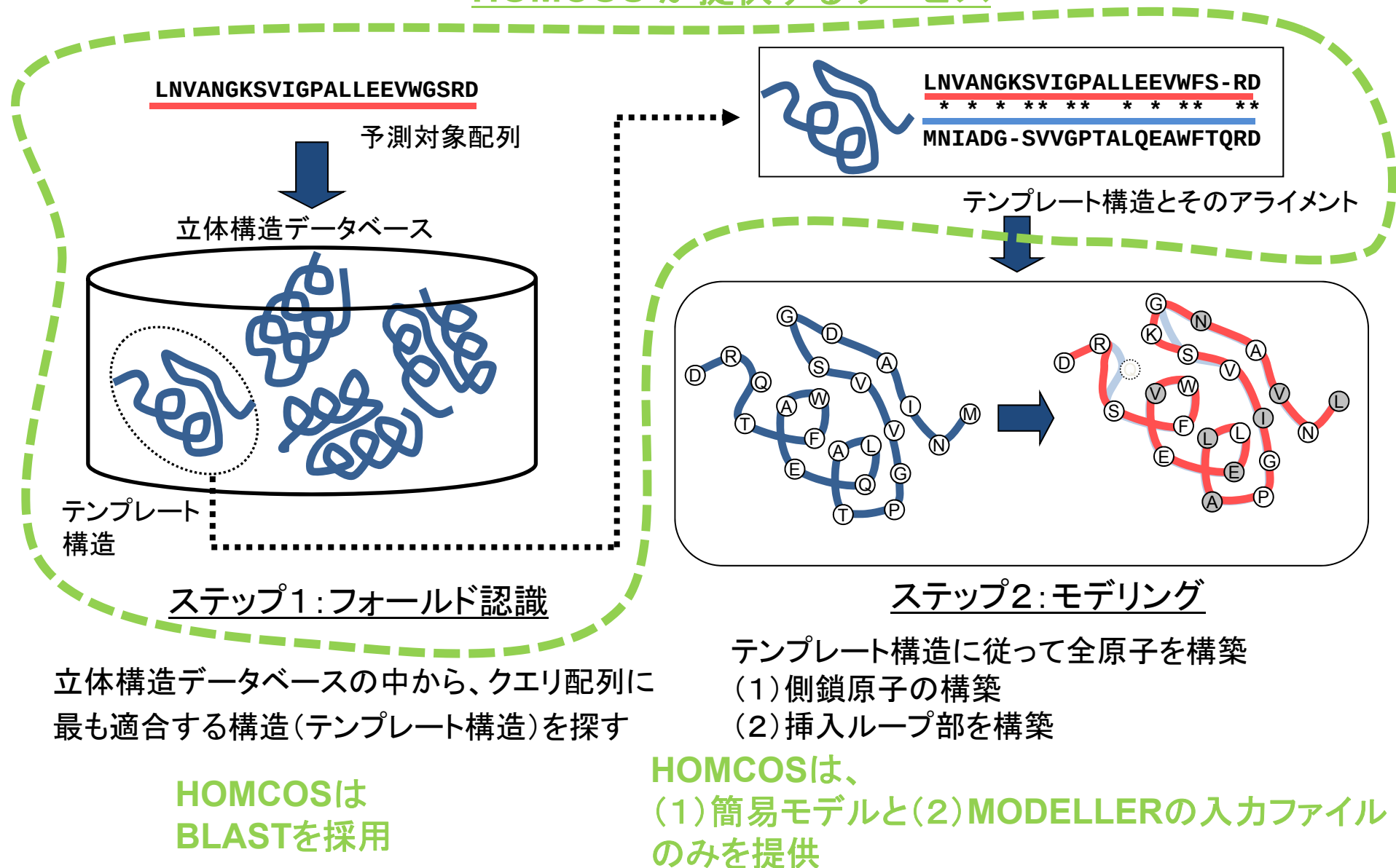
予測構造



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





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

相同なタンパク質のリスト

PDBに登録された
アミノ酸配列の
データベース

>1vwg_A

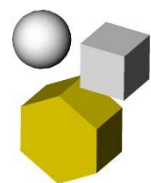
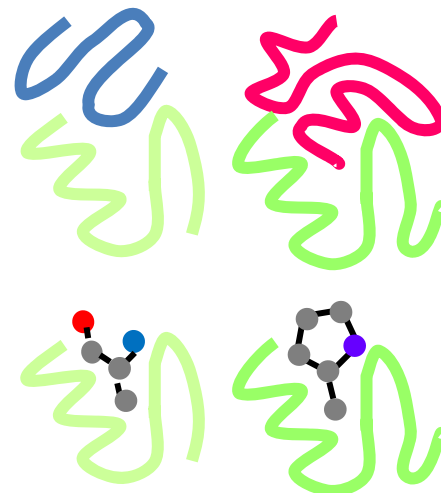
>1jsu_B

2g9xA
1w98A
1fq1B
:

1vwg_1 A1 B1
2g9x_1 A1 B1
:

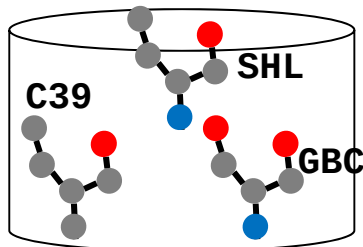
コンタクトしている分子の表

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



KCOMBU検索

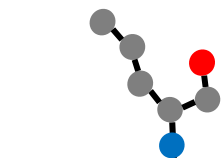
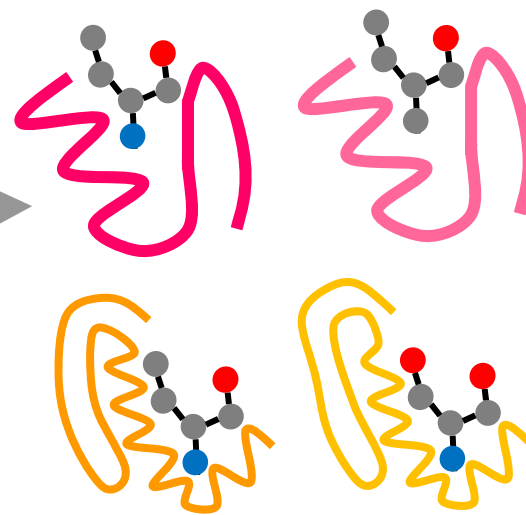
1vwg_1 A1 B1
2g9x_1 A1 B1
:

コンタクトしている分子の表

SHL
C39
GBC
:

類似した化合物のリスト

問い合わせ化合物とコンタクトする
分子の予測リスト



問い合わせ化合物



CDK3を題材とした結合分子予測

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



HOMCOS:タンパク質に対する結合分子の検索

[Go to English page] ヘルプページ

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

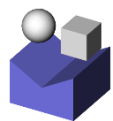
調べたいタンパク質の[PDB_ID]+[鎖識別子]、PDBファイルのアップロード、UniProtのID、あるいは1文字表記のアミノ酸配列を入力してください。:

PDB	PDB_ID: (1fin, 4hnb,...) CHAIN_ID: (A, B, ...)
PDBファイルのアップロード	参照...
UniProt_ID	CDK3_HUMAN (CDK3_HUMAN, MYB_HUMAN, ARSA_HUMAN, BRCA1_HUMAN, ADRB2_HUMAN, P15289, P38398, P24941 ...)
アミノ酸配列	

Reset values SEARCH

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、**[bars:full]**  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%に設定してある。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する。

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

HOMCOS

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

PID	QueryLength	Homolgous Sequence in PDB	UniProt Query	TITLE
43907	305	500	CDK3_HUMAN(Q00526)	RecName: Full=Cyclin-dependent kinase 3; EC=2.7.11.22;AltName: Full=Cell division protein kinase 3;
QUERYSEQ	MDMFQKVEKIGEGTYGVVYKAKNRETGQLVALKKIRLDLEMEGVPSTAIRESILLKELKHPNIVRLLDVHNKRKLYLVEFLSQDLKQYMDSTPGSELPLHLIKSYLFQLLQGVSFCHSHRVIHRDLKPQNLLINELGAIKLADFGLARAFGVPLRTYTHEVVTWYRAPEILLGSKFYTTAVDIWSIGCIFAEMVTRKALFPGDSEIDQLFRIFRMLGTPSEDTPGVTQLPDYKGSFQKTRKLEIEVNPLEPEGRDLLMQLLQYDPSQRI TAKTALAHPYFSSPEFSPAARQYVLQRFRRH			

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

UniProt Feature Tables [\[Q00526\(CDK3_HUMAN\)\]](#)

region	name	description
1-305	CHAIN	Cyclin-dependent kinase 3. /FTId
4-286	DOMAIN	Protein kinase.
10-18	NP_BIND	ATP.
127-127	ACT_SITE	Proton acceptor.
33-33	BINDING	ATP.

MONOMER

pdb_id	asym_id	identity	description
3f5x	C	6.7	CDK2_HUMAN Cell division prot

 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

Protein Model on 3f5x_C_1 (TEMPLATE: PDB3f5x)

DOWNLOAD: [\[sequence-replaced 3D model\]](#) [\[3D template\]](#) [\[Modeller script\]](#)

ALL MOLECULES IN THE BIOLOGICAL UNIT: [\[assembly id:1\]](#) [\[assembly id:2\]](#) [\[assembly id:4\]](#)

CurrentView: HTML5 Change to: [JAVA (faster than HTML5)]

MODEL[1] Protein A "queryA" TEMPLATE:3f5x_C_1 SeqID=76.7%

queryA 1:MDMFQKVEKIGEGTYGVVYKAKNRETGQLVALKKIRLDLEMEGVPSTAIRESILLKELKHPNIVRLLDVHNKRKLYLVEFLSQDLKQYMDSTPGSELPLHLIKSYLFQLLQGVSFCHSHRVIHRDLKPQNLLINELGAIKLADFGLARAFGVPLRTYTHEVVTWYRAPEILLGSKFYTTAVDIWSIGCIFAEMVTRKALFPGDSEIDQLFRIFRMLGTPSEDTPGVTQLPDYKGSFQKTRKLEIEVNPLEPEGRDLLMQLLQYDPSQRI TAKTALAHPYFSSPEFSPAARQYVLQRFRRH



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。



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: [3f5x_C_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

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

Protein Model on 3f5x_C_1 (TEMPLATE: [PDBj3f5x](#))

DOWNLOAD:
[\[sequence-replaced 3D model\]](#)
[\[3D template\]](#)
[MODELLER \[Modeller script\]](#)

ALL MOLECULES IN THE BIOLOGICAL UNIT:
[ALL MOL \[assembly id:1\]](#)
[ALL MOL \[assembly id:2\]](#)
[ALL MOL \[assembly id:4\]](#)

CurrentView: HTML5 Change to: [\[JAVA \(faster than HTML5\)\]](#)

model	mark	query	asym_id oper (auth_asym_id)	type	description	contact sites
1	☒	a	A(CDK3_HUMAN)	C 1 (C)	polymer(polypeptide(L))	Cell division protein kinase 2 :CDK2_HUMAN

MODEL[1] Protein A "queryA" TEMPLATE:3f5x_C_1 SeqID=76.7%

queryA	1:MDMFQKVEKIGEGTYGVVYKAKNRETGQLVLLKKIRLDLEMEGVPS'
3f5x_C_1	1:MFNFQKVEKIGEGTYGVVYKARNLTGEVVALKKIRLDTEGVPSP'
SecStr	: TTEEEEEEE SS EEEEEETTT EEEEEEE TTT S H
ExpBur	: eeebeeeeeeeeeebbebebebeebbbbbbbeeeeeeebe

クエリ(CDK3_HUMAN)と鋳型(3f5x_C_1)とのアラインメント

3QDLKKYMDSTPGSELP: 100
***** *
EQDLKKFMDSALTGIP: 100
EEHHHHHHHTTTT :
abbeebbebeeebe:

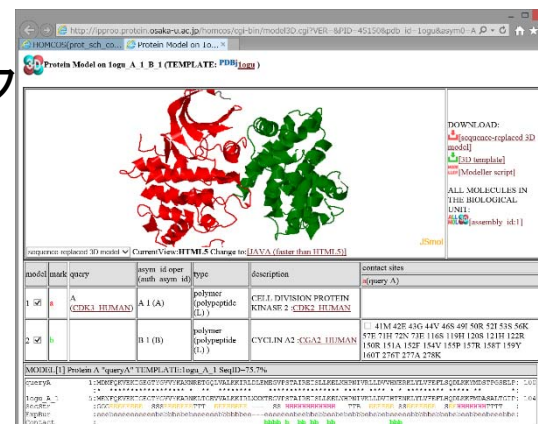
デフォルトではHTML5(Jsmol)による分子の表示。
JAVA実行環境がインストールしてあるなら、JAVAを用いたほうが表示は速い



標的分子と鑄型の同一残基率(%)
値が高いほど予測の信頼性が高い。



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





3D Protein Model on logu_A_1_B_1 (TEMPLATE: [PDBj logu](#))

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

model	mark	query	asym_id oper (auth_asym_id)	type	description	contact sites
1	☒	a	A (CDK3_HUMAN)	polymer (polypeptide (L))	CELL DIVISION PROTEIN KINASE 2 : CDK2_HUMAN	
2	☒	b	B 1 (B)	polymer (polypeptide (L))	CYCLIN A2 : CGA2_HUMAN	

MODEL[1] Protein A "queryA" TEMPLATE:logu_A_1 SeqID=75.7%

```

queryA      1:MDMFQKVEKIGEGTYGVVYKAKNRETGQLVALKKIRLDLEMEGVPSTAIREISLLKELKHPNIVRLLDVVHNERKLYLVFEFLSODLKKYMDSTPG
: * *****
logu_A_1    5:MENFQKVEKIGEGTYGVVYKARNKLTGEVVALKIRLXXXTGVPSTAIREISLLKELNHPNIVKLDVINTENKLYLVFEFLHQDLKKFMDASAL
SecStr      :GGGEEEEEE SSSEEEEETTT EEEEEEE --- SS HHHHHHHHHHHH TTB EEEEE SSSEEEE SEHHHHHHHHHTT
ExpBur      :eeebeeeeeeeeebbbbbebeeeeeebbbbbee-- eeeeeeeebbbbbebebebebebebeebbbbbebebebeebbeebbeebbeebbe:
Contact     : bbbb b bb bb bb bbb

```

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

DOWNLOAD:

[\[sequence-replaced 3D model\]](#)

[\[3D template\]](#)

[\[Modeller script\]](#)

ALL MOLECULES IN THE BIOLOGICAL UNIT:

[\[assembly id:1\]](#)

予測接触残基
標的配列(CDK3)の残基番号、残基名

☐ 41M 42E 43G 44V 46S 49I 50R 52I 53S 56K
57E 71H 72N 73E 116S 119H 120S 121H 122R
150R 151A 152F 154V 155P 157R 158T 159Y
160T 276T 277A 278K

接触残基
の文字で
れている

接触残基が'b'の文字で示されている。

CDK3

Cyclin A2



ALL
MOL

生物学的単位内の全分子を表示した場合

3D Protein Model on 1ogu assembly_id=1 (TEMPLATE: [PDBj1ogu assembly_id:1](#))

ST8という化合物 (ATP阻害剤) も表示される

CDK3 (鋳型はCDK2)

Cyclin A2

JSmol

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

DOWNLOAD:
[\[sequence-replaced 3D model\]](#)
[\[3D template\]](#)
[\[Modeller script\]](#)

model	mark	query	asym_id oper (auth_asym_id)	type	description	contact sites a(query A)
1	☒	a	A	polymer (polypeptide (L))	CELL DIVISION PROTEIN KINASE 2 : CDK2_HUMAN	
2	☒	化合物 (ST8)	3 1 (A)	non-polymer (ST8)	4- {[4-AMINO-6-(CYCLOHEXYLMETHOXY)-5-NITROSOPYRIMIDIN-2-YL]AMINO} BENZAMIDE	☐ 10I 12E 13G 18V 31A 64V 80F 81E 82F 83L 84S 85Q 86D 89K 131Q 132N 134L 145D
3	☒	c	B	polymer (polypeptide (L))	CYCLIN A2 : CGA2_HUMAN	☐ 41M 42E 43G 44V 46S 49I 50R 52I 53S 56K 57E 71H 72N 73E 116S 119H 120S 121H 122R 150R 151A 152F 154V 155P 157R 158T 159Y 160T 276T 277A 278K
4	☒	化合物 (SGM)	B)	non-polymer (SGM)	MONOTHIOGLYCEROL	



CONTACTしている
化合物の分子名

標的と鋳型の同一残基率(%)
値が高いほど予測の信頼性が高い。

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





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

CDK3
(鋳型はCDK2)

化合物名。
3文字表記は4QE

DOWNLOAD:

 [sequence-replaced 3D model]

 [3D template]

MODELLER [Modeller script]

ALL MOLECULE IN THE BIOLOGICAL UNIT:

ALL MOL 3D [4cfm assembly id:1]

予測接触残基
標的配列(CDK3)
の
残基番号、残基名

model	symbol	query	asym_id oper_expression (auth_asym_id)	type	description	contact sites
1 ☒	a	A (<u>CDK3_HUMAN</u>)	A 1 (A)	polymer (polypeptide(L))	CYCLIN-DEPENDENT KINASE 2 : <u>CDK2_HUMAN</u>	a(query A)
2 ☒	b		E 1 (A)	non-polymer (<u>4QE</u>)	6-(CYCLOHEXYLMETHOXY)-8-(2-METHYLPHENYL)-9H-PURIN-2-AMINE	☐ 10I 12E 13G 18V 31A 64V 80F 81E 82F 83L 84S 85Q 86D 89K 131Q 132N 134L 145D

MODEL[1] Protein A "queryA" TEMPLATE:4cfm A 1

```
queryA      1:MDMFQKVEKIGEGTYGVVYAKNRETGQLVALKKIRLDLEMEGVFPSTAIRESLLKLKHNPINVRLLDVVNERNKIYLVFELFSQDLKYMDSTPGSELP: 100  
            *: ***** ** **** *  
4cfm_A_1    6:MENFQKVEKIGEGTYGVVYARNLKTGEVVALKKIRLDTEGGVPSTAIREISLKLNHPNIKVLLDIHTENKIYLVFELFHQDKLFKFMDASALTGP: 105  
SecStr       : TT##### SSS##### TTT ##### SSS s h##### TTB ##### SE##### F##### TTTT :  
ExpBur       : b##### eebbbbebebebebeeeebbbsbbbbeeeeeeeebebbsbbebebebbbebebebeebbeebbebebebeebbeebbebebebe:  
Contact     : |          b bb   b           b                   b             hbbbbbb b |
```

接触残基が'b'の文字で示されている。



Site Table


[\[bars:full\]](#) [\[SiteTable\]](#)

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

Contact Bar(summary)[0.0 %]
seq_id(%): [\[0\]](#) [\[30\]](#) [\[40\]](#) [\[50\]](#) [\[60\]](#) [\[70\]](#) [\[80\]](#) [\[90\]](#) [\[95\]](#) [\[100\]](#)

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






[\[bars:summary\]](#) [\[bars:full\]](#) Site Table[0.0 %]
seq_id(%): [\[0\]](#) [\[30\]](#) [\[40\]](#) [\[50\]](#) [\[60\]](#) [\[70\]](#) [\[80\]](#) [\[90\]](#) [\[95\]](#) [\[100\]](#) [\[show\]](#) [\[download\]](#) [\[help\]](#)

PID	QueryLength	Homologous Sequence in PDB	UniProt Query	TITLE
45150	305	500	CDK3_HUMAN(Q00526)	RecName: Full=Cyclin-dependent kinase 3; EC=2.7.11.22; AltName: Full=Cell division protein kinase 3;
QUERYSEQ	MDMFQKVEKIGEGTYGVVYKAKNRETGQLVALKKIRLDLEMEGVSTAIRESILLKELKHPNIVRLDVVHNERKLYLVFEFLSQDLKKYMDSTPGSELPLHLIKSYLFQLQGVSFCHSHRVIHRDLKPQNLINELGAIKLADPGLARAFGVPLRTYTHEVVTLYWYRAPEILLGSKFYTTAVDIWSIGCIFAEMVTRKALFPGDSEIDQLFRIFRMLGTPSEDWPGVTQLPDYKGSFQWTRKGLLEIVNLEFEGRDLLMQLLYDPSQRITAKTALAHFYSSPEFSPAARQYVLQRFRH			

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

[n]:site number of query sequence. [a]:amino acid of query sequence. [e]:predicted exposed/buried. [acc]:predicted accessibility. [p]:contact mols]:binding other molecules [observed aa]:Observed amino acid frequency. [feature table]:UniProt Feature Table

結合分子のサマリー

n	a	s	e	acc	pdb	contact_mols	observed aa	feature table	variant
SITE 1	M		e	29.5	3f5x_C	homo precipitant	MLIVQ		
SITE 2	D	T	e	73.4	3f5x_C	hetero CCNA2_HUMAN CG2A_HUMAN homo precipitant	DENKGQST		
SITE 3	M	T	e	35.8	3f5x_C	hetero CCNA2_HUMAN precipitant	RDKNVEILQCMTSFHAG		
SITE 4	F	E	b	17.7	3f5x_C		YFWL		
SITE 5	Q	E	e	47.4	3f5x_C	MOUSE	EGKDDPVTATKQMSV		
SITE 6	K	E	e						

溶媒露出度(%)

二次構造(H:αヘリックス、E:βシート)

UniProtのアノテーション (Feature Table)

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

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

(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(CDKK3)

Browser address bar: http://ipproo.protein.osaka-u.ac.jp/homcos/cgi-bin/prot_sch_eachsite.cgi?QI

HOMCOS: HOMCOS(prot_sch_... x)

SITE Summary for the 3-rd Site(M)

PID	QueryLength	FocusSite	TITLE
45150	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

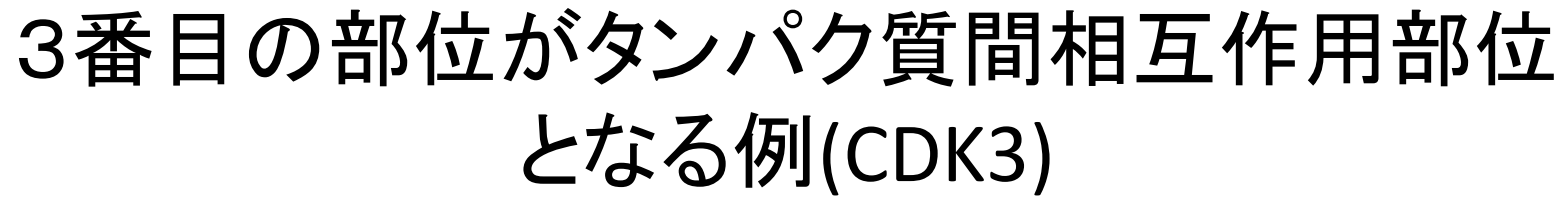
Template For Monomer	predicted SecStr	predicted ExpBur	Predicted Relative Acc(%)
3f5x	T (<i>Hbond turn</i>)	e (<i>exposed</i>)	35.8

3D Complex Information

Predicted Bind Molecules
precipitant :1 hetero :7
Templates for 3D complexes
precipitant [EDO] 3qu0 A 1 E 1 hetero [1550:CCNA2 HUMAN] 3f5x C 2 B 3 3f5x A 4 D 1 1jst A 1 D 1 2v22 C 1 B 1 3f5x C 1 B 4 2v22 A 1 D 1 3f5x A 3 D 2

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。

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



3番目のMet (鋳型ではAsn)



SPIC_HUMANの場合の 「タンパク質に対する検索」のトップ画面

SPIC_HUMAN (Transcription factor Spi-C)

PID	QueryLength	Homologous Sequence in PDB	UNIPROT_QUERY	TITLE
30875	248	54	SPIC_HUMAN (AC:Q8N5J4 ID:SPIC_HUMAN)	RecName: Full=Transcription factor Spi-C;
QUERYSEQ	MTCVEQDKLGQAFEDAFEVLQHSSTGDLQYSPDYRNLYALINHRPHVKGNSCYGLPTEEPVYNWRTVINSAADFYFEGNIHQSLQNIITENQLVQPTLLQKGGKGRKKRLRFEYLHESLYNPEMASCQWVDKTKGIFQFVSKNKEKL AELWGKRRGNRKMTYQKMARALNRYGRSEITIKRRKLTQYQSEAILQRLSPSYFLGKEIFYSQCVQPDQEYLSLNNWNNYNYTYANYHELNHDC			

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

UniProt Feature Tables [Q8N5J4(SPIC_HUMAN)]

248	region	name	description
1-248	CHAIN		Transcription factor Spi-C. /FTId=PRO_0000204140.
111-194	DNA_BIND		ETS. {ECO:0000255}PROSITE- ProRule:PRU00237.

MONOMER

248	pdb_id	asym_id	identity	description
	1pue	F	59.6	SPI1 MOUSE PROTEIN (TRANSCRIPTION FACTOR PU.1 (TF PU.1))
	4bqa	A	30.9	ETS2 HUMAN PROTEIN C-ETS-2
	1r36	A	30.4	ETS1 MOUSE C-ets-1 protein

HETERO

248	contact mol	homologue
1k78[2]	E	PAX5 HUMAN Paired Box Protein Pax5[124 aa]
1mdm[1]	C	PAX5 HUMAN PAIRED BOX PROTEIN PAX-5[124 aa]
1k78[1]	I	PAX5 HUMAN Paired Box Protein Pax5[58 aa]
4i18[1]	A	RUNX1 MOUSE Runt-related transcription factor 1[142 aa]
1awc[1]	D	GABP2 MOUSE PROTEIN (GA BINDING PROTEIN BETA 1)[153 aa]

NUCLEOTIDE

248	contact mol	homologue
1pue[2]	C	DNA (5'-D(*AP*AP*AP*AP*AP*GP*GP*GP*GP*AP*AP*GP*TP*...
1pue[2]	B	DNA (5'-D(*TP*CP*CP*CP*AP*CP*TP*CP*CP*CP*CP*TP*...
4mhg[1]	A	Specific 14 bp DNA
4mhg[1]	B	Complimentary Specific 14 bp DNA



のアイコンをクリックすると核酸-タンパク質複合体の立体構造モデルが表示される



DOWNLOAD:

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

ALL MOL 3D [1pue assembly id:2]

[illegible]



核酸タンパク質複合体(SPIC)

assembly_id=1の生物学的単位に含まれる全分子を用いたモデル

sequence-replaced 3D model CurrentView:HTML5

model	symbol	query	asym_id oper_expression (auth_asym_id)	type	description	
1 ☒	a		C 1 (C)	polymer (polydeoxyribonucleotide)	DNA (5'-D(*AP*A P*AP*AP*AP*GP*G P*GP*GP*AP*AP*G P*TP*GP*GP*G)-3 ')	167Q 171R 174R 184K 188K 189L
2 ☒	b		D 1 (D)	polymer (polydeoxyribonucleotide)	DNA (5'-D(*TP*C P*CP*CP*AP*CP*T P*TP*CP*CP*CP*C P*TP*TP*TP*T)-3 ')	☐ 110K 111L 112R 113L 154W 158K 160N 161R 162K 164M 167Q 168K 171R 176Y 188K
3 ☒	c	A (SPIC_HUMAN)	F 1 (F)	polymer(polypeptide(L))	PROTEIN (TRANSCRIPTION FACTOR PU.1 (TF PU.1)) :SPI1_MOUSE	

MODEL[3] Protein A "queryA" TEMPLATE:1pue_F_1

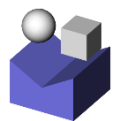
queryA 110:KLRLFEYLHESLYNPEMASCIQWVDKTKGIFQFVSKNKEKLAEIWLWGRKRGNRKMTYQKMARALRNYGRSGEITKIRKLTQFSEAIL: 198

http://ipproo.protein.osaka-u.ac.jp/homcos/cgi-bin/model3D.cgi?PID=30875&pdb_id=... ***** *
***** *
***** *
***** *
***** *
***** *

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

ALL MOLECULE IN THE BIOLOGICAL UNIT:
 [1pue assembly_id:2]

をクリックすると、このPDBの生物学的単位(Biological Unit) assembly_id=1に含まれる全分子が表示される。



ARSA_HUMANの場合のSite Tableの画面



ARSA_HUMAN(Arylsulfatase A)

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

Site Table[0.0 %]

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

PID	QueryLength	Homologous Sequence in PDB	UNIPROT_QUERY	TITLE
9107	507	36	ARSA_HUMAN (AC:P15289 ID:ARSA_HUMAN)	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(%) LAAGLAVARPNIVLIFADDLGYGDLGCGYGHPSSTTPNLDQLAAGGLRFTDFYVPSLCTPSRAALLTGRLPVRMGMPGVLPVSSRGGLPLEEVTVAEVLAAAGYLTMAGKWHLGVGPEGAFLPPHQGFHRLGIPYS
CFPPATPCDGGCDQGLVPIPLLANLSVEAQPPWT DGTADYMAFAHNTIMANORNDPFTVVA SHHTYVPQFSGQSFASRSGRPFQDGLMELDAAVGTLMTAIGDLGLLETLVIFTADNGPETM
REPALAFWPGHIAPGVTHELASSLDLLPTLAF 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 EVRGVFAVRTGKYKAHFFTQGSASDITADPACHASSSLTAHEPPLLYDLKDPGE UniProtの変異体に関する情報(VARIANT)

[BLAST file] 結合分子のサマリー [multiple] UniProtのアノテーション (Feature Table) [variant]: UniProt Human variant. [feature table]: 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 ZN NA precipitant	D	METAL Calcium.	D->N:Disease
SITE 30	D	S	b	1.2	1n2l_A	compound CSN metal CA MG NA precipitant	DQTH	METAL Calcium.	D->H:Disease
SITE 31	L		b	0.0	1n2l_A		LQMVG		
SITE 32	G		b	0.0	1n2l_A		GDRANE		G->S:Disease
SITE 33	Y	T	b	8.7	1n2l_A	homo precipitant	WVYFIETAGLSDKPR		
SITE 34	G	T	b	7.1	1n2l_A	compound NAG precipitant	GSNVIPDHMAREFKLQT		

↓ 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)

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

http://ipproo.protein.osaka-u.ac.jp/homcos/cgi-bin/prot_sch_eachsite.cgi?QPROTSEQ_ID=

SITE Summary for the 29-th Site(D)

PID	QueryLength	FocusSite	TITLE
45948	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]

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%)

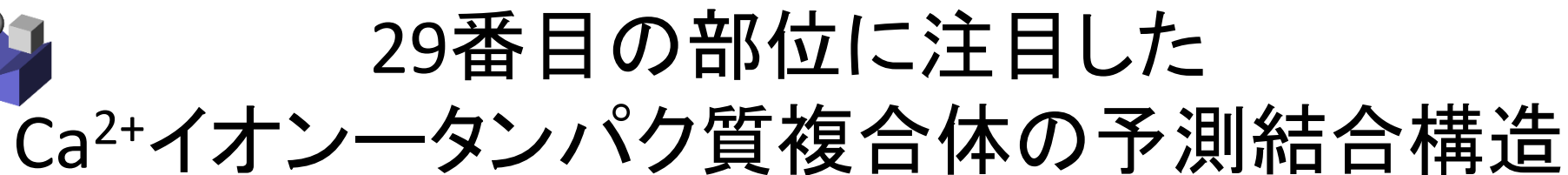
3D Structure Information

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

3D Complex Information

Predicted Bind Molecules
precipitant:2 metal:62 compound:2
Templates for 3D complexes
precipitant [SO4] 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 1 C 1 4uph D 1 K 1 1auk A 4 C 4 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 [CA] 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 4cys B 1 G 1 4cys 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

カルシウムイオン (Ca²⁺)との複合体構造が予測されている。





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

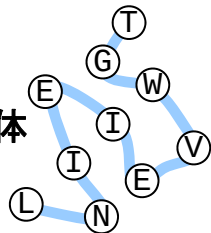
配列Aと相同な
タンパク質のリスト

問い合わせタンパク質A

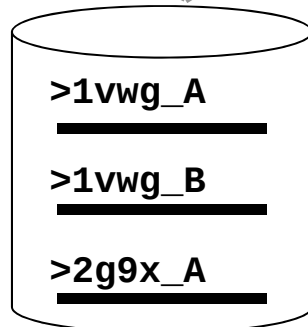
配列 TGWVEIEINL...

or

単量体
構造



BLAST検索



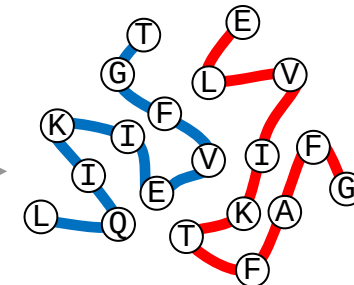
PDB内のアミノ酸配列
のデータベース

1vwgA
2g9xA
8atcA
1fq5A
:

1vwg_1 A1 B1
2g9x_1 A1 B1
1jsu_1 A1 B1
8atc_2 A1 B1
2fi5_1 E2 I1
:

結合している
分子の表

相同な複合体の一つを
鋳型として取り出す



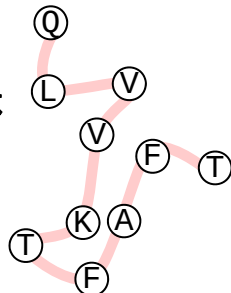
鋳型構造: 1vwg_1 A1 B1

問い合わせタンパク質B

配列 QLVVKTF AFT...

or

単量体
構造



BLAST 検索

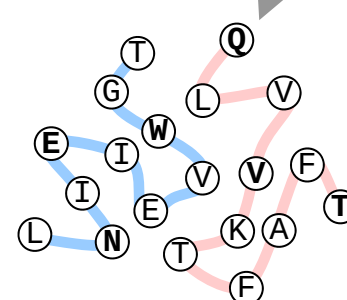
1vwgB
2g9xB
2fi5I
2eufA
:

配列Bと相同な
タンパク質のリスト

配列の
置き換え

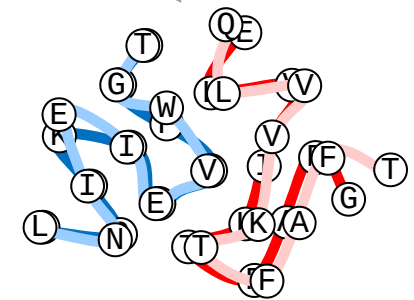
or

単量体の
重ね合わせ



予測モデル構造

Template-based Model
(Sequence-replaced model)



予測モデル構造

Template-based docking

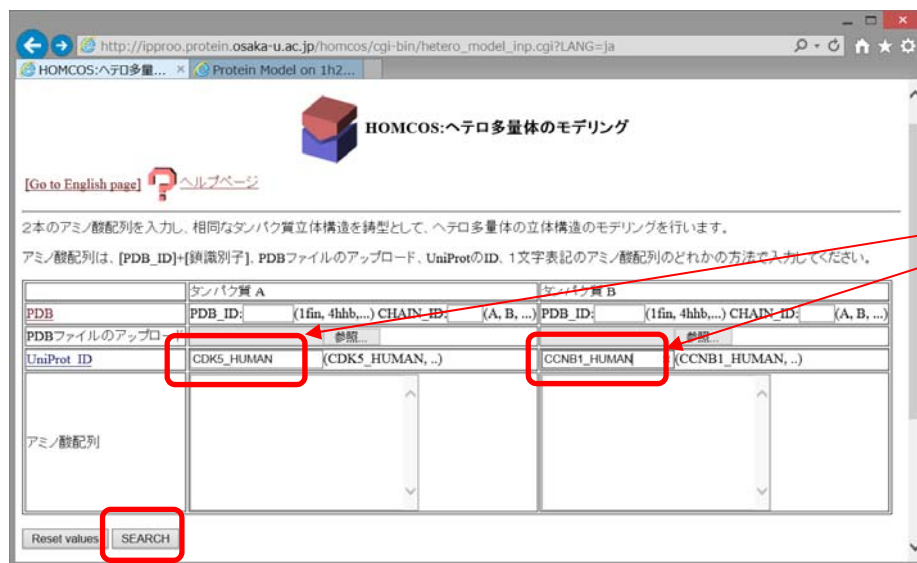


ヘテロ多量体のモデリング(2本の配列から)

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 protein 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

Search Templates to Model Hetero Multimers

proteinA	TITLE	RecName: Full=Cyclin-dependent-like kinase 5; EC=2.7.11.1; AltName: catalytic subunit; Short=TPKII catalytic subunit;
	UniProt_ID for query	CDK5_HUMAN
	QUERYSEQ (length 292)	MQYKEKLEKIGEGTYGTVFKAKNRETHEIVALKRVRLLDDDEGVFPSSALREICLLKELF FGIPVRCYSAEVVTLWYRFPDVLFGAKLYSTSIDMWSAGCIFAEANAGRPLFPNDV
	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)	MALRVTRNSKINAENKAKINMAGAKRVPTAPATSKPGLRPTALGDIGNKVSEQLQAKMPMKKEAKPSATGKVIDKKLPKLEKVPMLVPVSEPVPEPEPEPEPVKEEKLSPILVDTASPSMETSGCAPAEEDLCQAFSDVI LAVNDVDAEDGADPNLCSEYVKDIYAYLRQLEEEQMPRYLLGREVTGNMRAILIDWLQVQMKFRLQETMYMTVSIIDRFMQNNCVPKMKMLQVGVGTAMFIAASKYEEMPEIGDFAFVTDNTYTKHQIRQMEMKILRALNFGGLRP LPLHLRRASKIGEVNDEQHTLAKYLMELTMVHVFPPSQIAAGAFCLALKILDNGEWTPTLQHYLSYTEESLLFPVQHLAKNVVMVQGLTKHMTVKNKYATSKHAKISTLPQLNSALVQDLAKAVAKV
	BLAST	Nsbjseq 220 [show plain BLAST file]

PID 46271 precalc_blt T minNca 1 minRca 0 Nmodel 206

複合体のsequence-replaced 3D model

Protein Model on 1h27_A_1_B_1 (TEMPLATE: PDBj1h27)

DOWNLOAD:
[\[sequence-replaced 3D model\]](#)
[\[3D template\]](#)
[\[Modeller script\]](#)

ALL MOLECULES IN THE BIOLOGICAL UNIT:
[\[assembly id:1\]](#)

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

model	mark	query	asym_id oper (auth asym_id)	type	description	contact sites
						a(query A) b(query B)
1	☒	A (CDK5_HUMAN)	A 1 (A)	polymer (polypeptide (L))	CELL DIVISION PROTEIN KINASE 2 : CDK2_HUMAN	☐ 169E 170Y 171V 173D 174I 177Y 219L 254I 257K 258Y 259E 260E 261M 265E 266I 279K 280H 283R 286E 287M 290L 291R 294N 295F 296G 297L 298G 302P 303L
2	☒	B (CCNB1_HUMAN)	B 1 (B)	polymer (polypeptide (L))	CYCLIN A2 : CGA2_HUMAN	☐ 37L 39D 40D 41D 42E 43G 44V 46S 49L 50R 52I 53C 56K 57E 71H 72S 115G 118H 119S 120R 121N 149R 150A 151F 153I 154P 156R 157C 158Y 159S 276S 277A 278E

Homologous hetero-multimer found by BLAST

asmb1:assembly_id NproAB:number of proteins homologous to query A or B Nres_conali:number of aligned contact residues

PDB				Protein A						Protein B					
id	asmb1	NproAB	Nres_conali	N	asym oper	UniProtID		Evalue	SqID (%)	N	asym oper	UniProtID		Evalue	SqID (%)
 1h27	1	2-mer	33	1	A 1	CDK2_HUMAN		8e-122	60.7	1	B 1	CGA2_HUMAN		8e-46	39.4
 1h27	2	2-mer	33	1	C 1	CDK2_HUMAN		7e-123	60.1	1	D 1	CGA2_HUMAN		8e-46	39.4
 3f5x	1	8-mer	32	4	C 1 C 2 A 3 A 4	CDK2_HUMAN		3e-122	60.7	4	D 1 D 2 B 3 B 4	CCNA2_HUMAN		2e-45	39.6

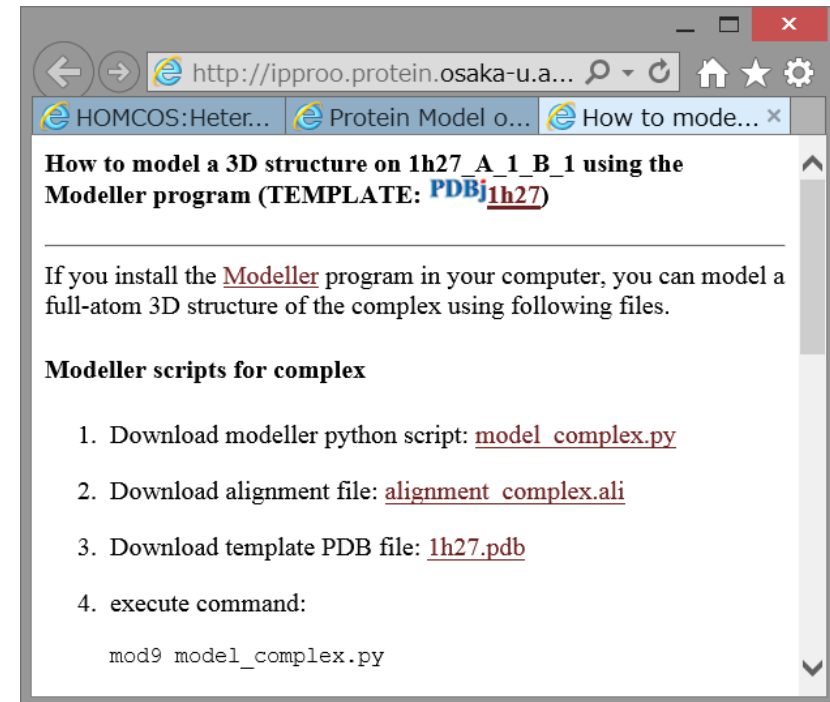
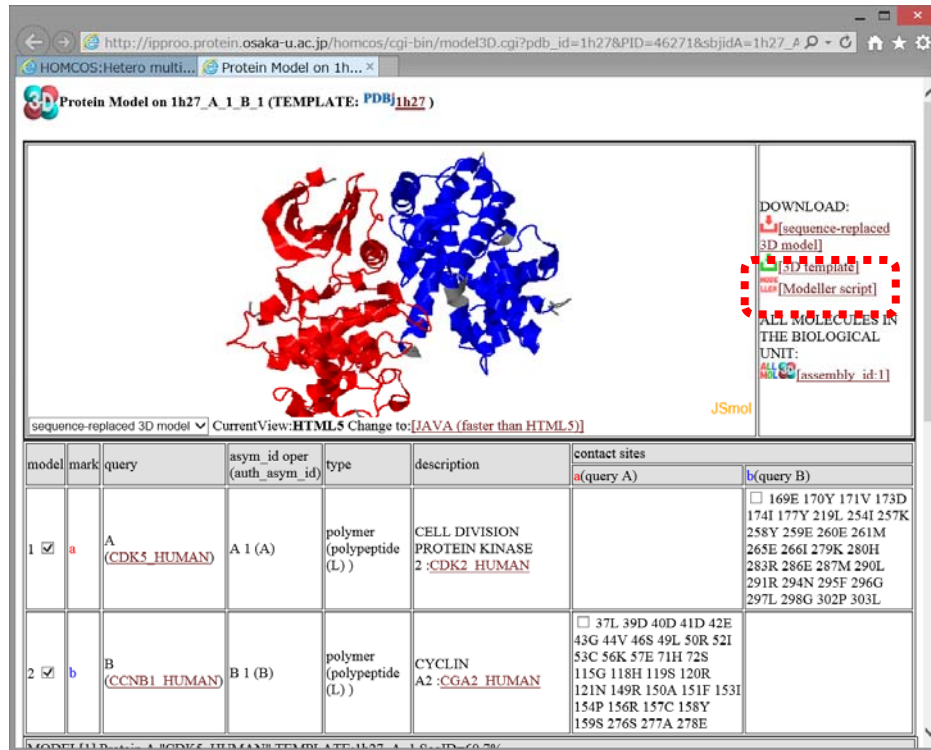
CDK5_HUMAN

CCNB1_HUMAN

MODELLER

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。

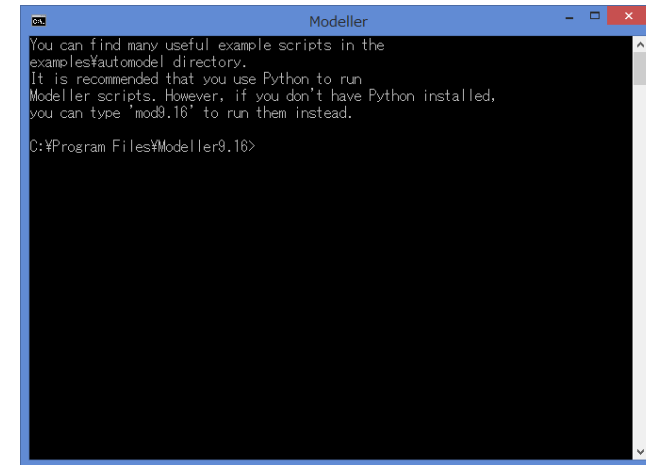
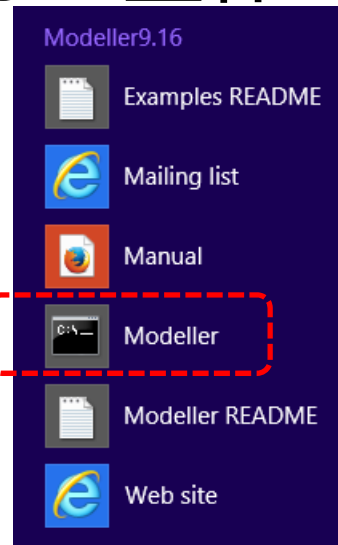
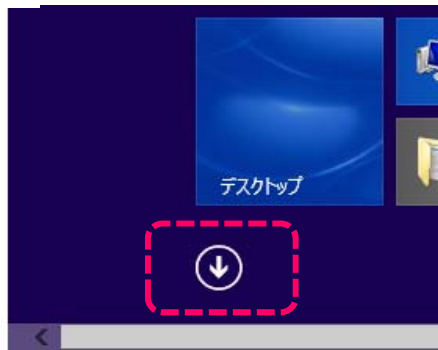


(1) モデル3D構造のウィンドウで **MODELLER** をクリックする (2) Modellerのスキプのページが表示される。

(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。

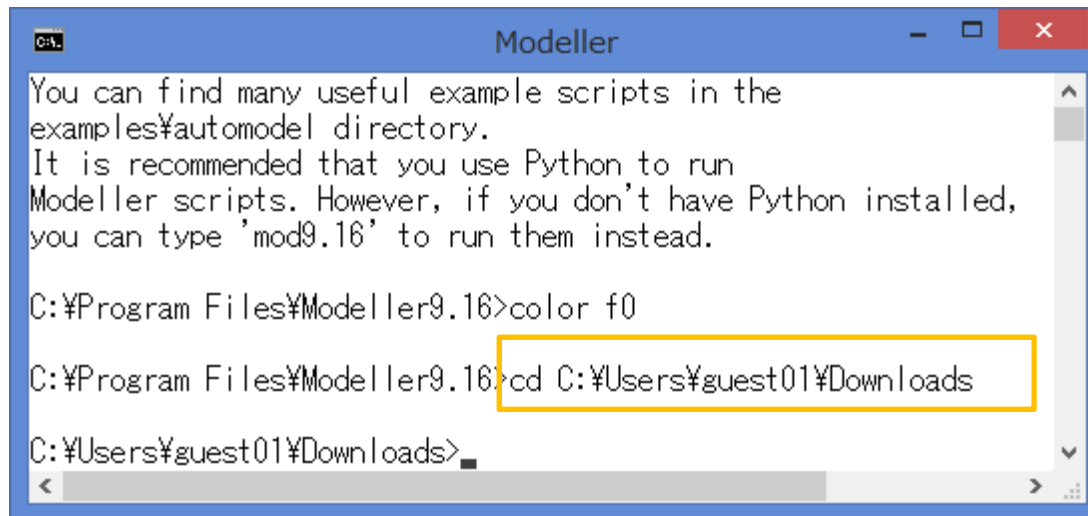
今回は、Win8では、C:\Users\guest01\Downloads というディレクトリに、
Macでは、/Users/guest/Downloads というディレクトリに保存することにする。

MODELLERによる二量体のモデリング(Win8)[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”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\guest01\Downloads**と入力する。

MODELLERによる二量体のモデリング(Win8)[3]

```
Modeller
C:\Users\guest01\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\guest01\Downloads のディレクトリ

2016/03/13  15:35    <DIR>          .
2016/03/13  15:35    <DIR>          ..
2016/03/13  15:35                363,949  1h27.pdb
2016/03/13  15:35                1,258  alignment_complex.ali
2016/03/13  15:35                1,129  model_complex.py
                3 個のファイル          366,336 バイト
                2 個のディレクトリ  460,065,701,888 バイトの空き領域

C:\Users\guest01\Downloads>
```

(8) コマンド **dir** 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。

```
Modeller
C:\Users\guest01\Downloads>mod9.16 model_complex.py
```

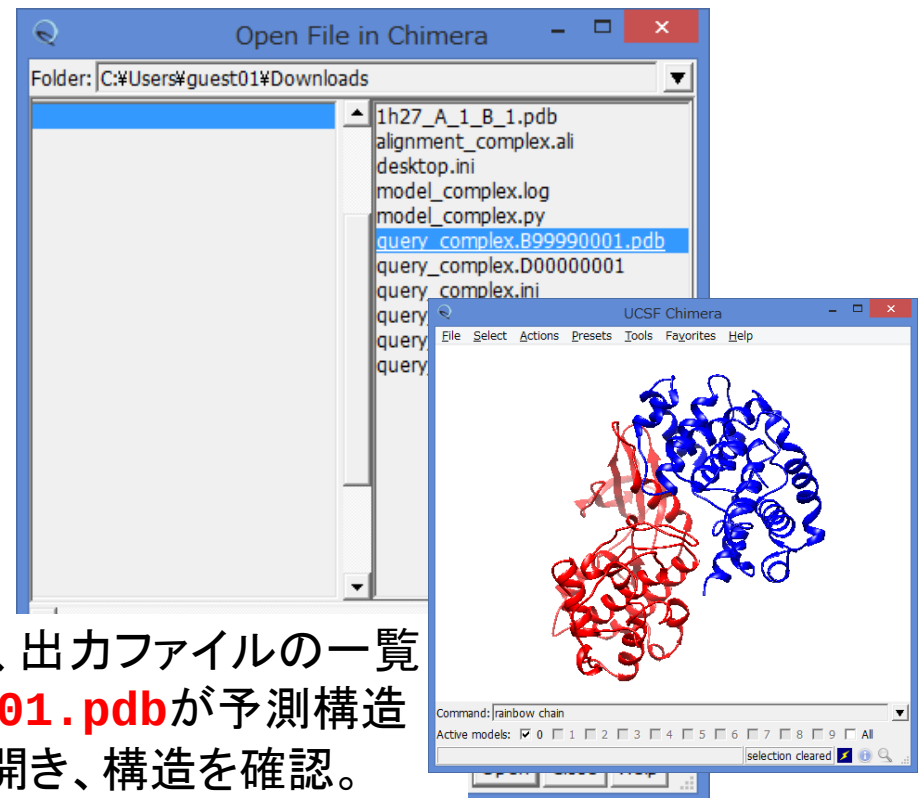
(9) コマンド **mod9.16** [スクリプトファイル] を入力し、Modellerを実行する。今回は、**mod9.16 model_complex.py** と入力する。この後、計算終了までには1分～数分程度の時間がかかる。

```
Modeller
C:\Users\guest01\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\guest01\Downloads のディレクトリ

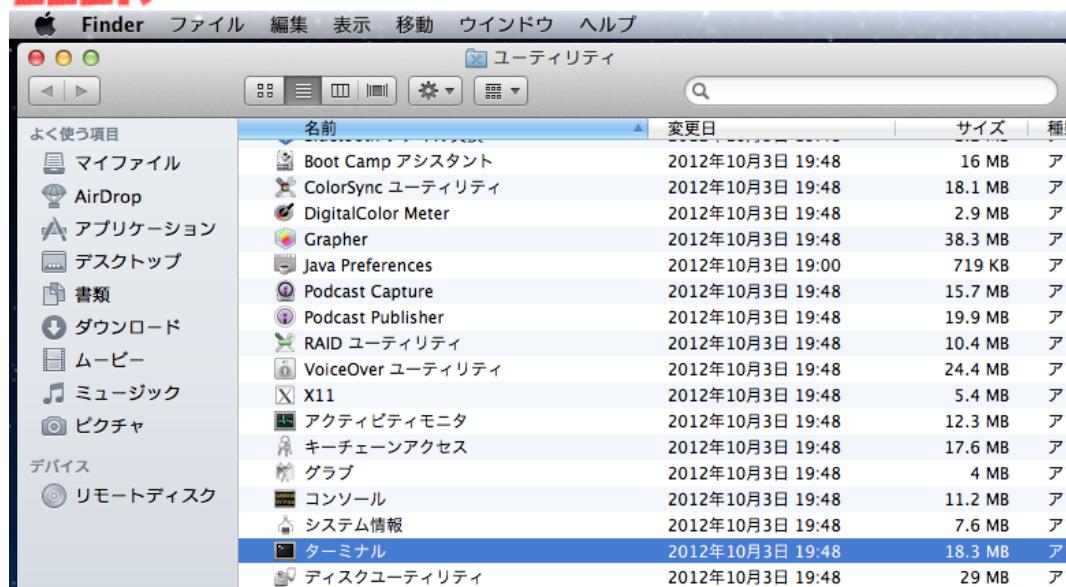
2016/03/13  15:40    <DIR>          .
2016/03/13  15:40    <DIR>          ..
2016/03/13  15:35                363,949  1h27.pdb
2016/03/13  15:35                1,258  alignment_complex.ali
2016/03/13  15:40                48,771  model_complex.log
2016/03/13  15:35                1,129  model_complex.py
2016/03/13  15:40                344,571  query_complex.B999900001.pdb
2016/03/13  15:40                20,930  query_complex.D000000001
2016/03/13  15:39                344,455  query_complex.ini
2016/03/13  15:39                3,805,818  query_complex.rsr
2016/03/13  15:39                3,704  query_complex.sch
2016/03/13  15:40                267,108  query_complex.V999900001
                10 個のファイル          5,201,693 バイト
                2 個のディレクトリ  460,060,389,376 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.B999900001.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。

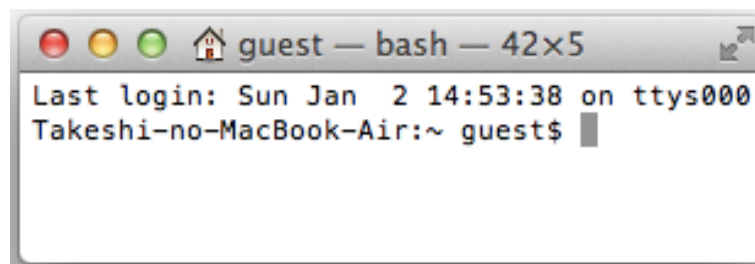


MODELLER

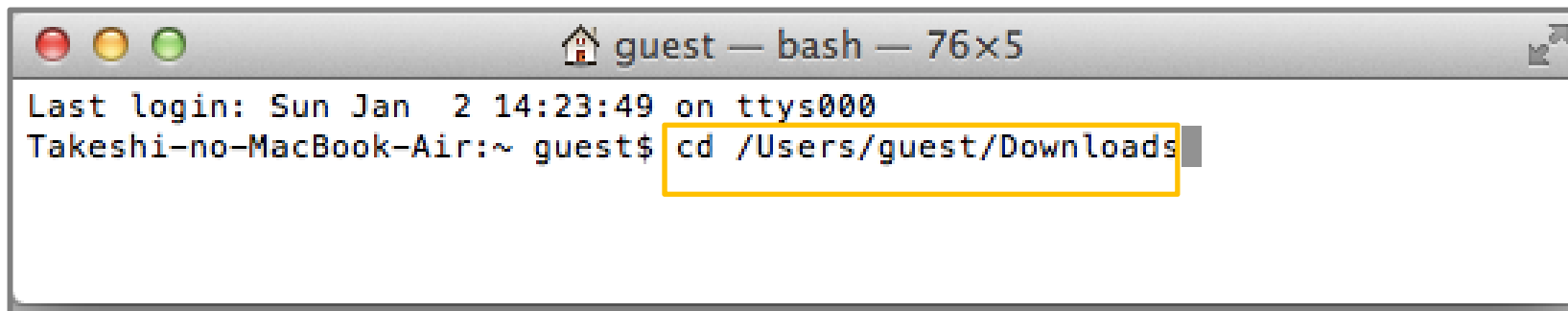
Modellerによる二量体のモデリング(Mac)[2]



(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

Modellerによる二量体のモデリング(Mac)[3]

```
Downloads — bash — 50x5
Takeshi-no-MacBook-Air:Downloads guest$ ls
1h27.pdb                alignment_complex.ali
About Downloads.lpdf    model_complex.py
Takeshi-no-MacBook-Air:Downloads guest$
```

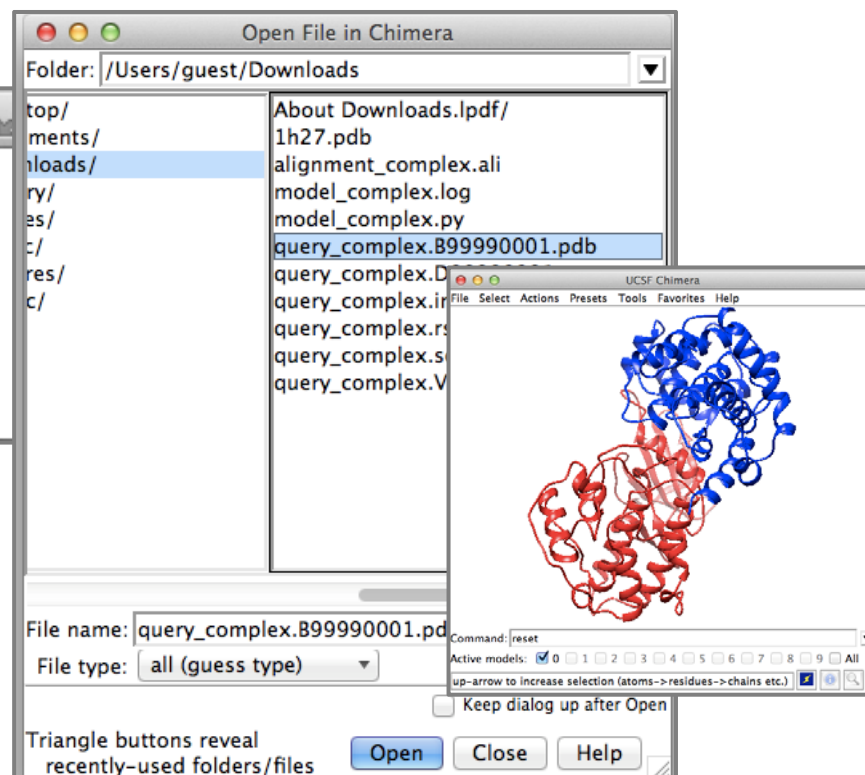
(6) コマンド **ls** を入力すると、現在のディレクトリにあるファイルの一覧を表示。ダウンロードした三つのファイルがあることを確認。

```
Downloads — bash — 67x5
Takeshi-no-MacBook-Air:Downloads guest$ mod9.16 model_complex.py
```

(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。

```
Downloads — bash — 59x8
Takeshi-no-MacBook-Air:Downloads guest$ ls
1h27.pdb                query_complex.D00000001
About Downloads.lpdf    query_complex.V99990001
alignment_complex.ali  query_complex.ini
model_complex.log      query_complex.rsr
model_complex.py       query_complex.sch
query_complex.B99990001.pdb
Takeshi-no-MacBook-Air:Downloads guest$
```

(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。



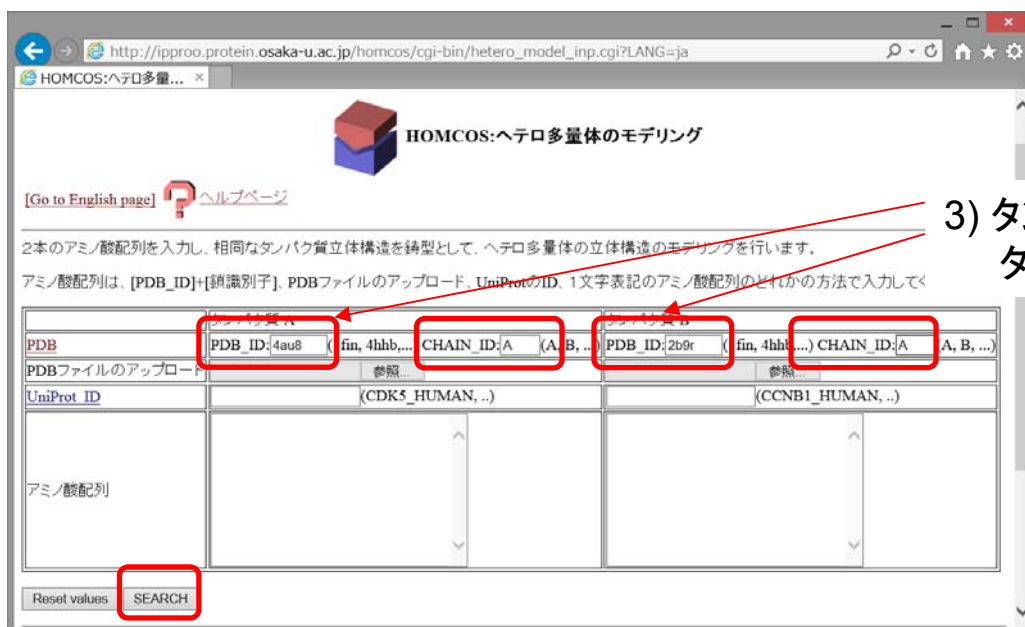


ヘテロ多量体のモデリング(2つの単量体構造から)

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PDB_IDに4au8, CHAIN_IDにA、 タンパク質BのPDB_IDに2b9r, CHAIN_IDにAを入力

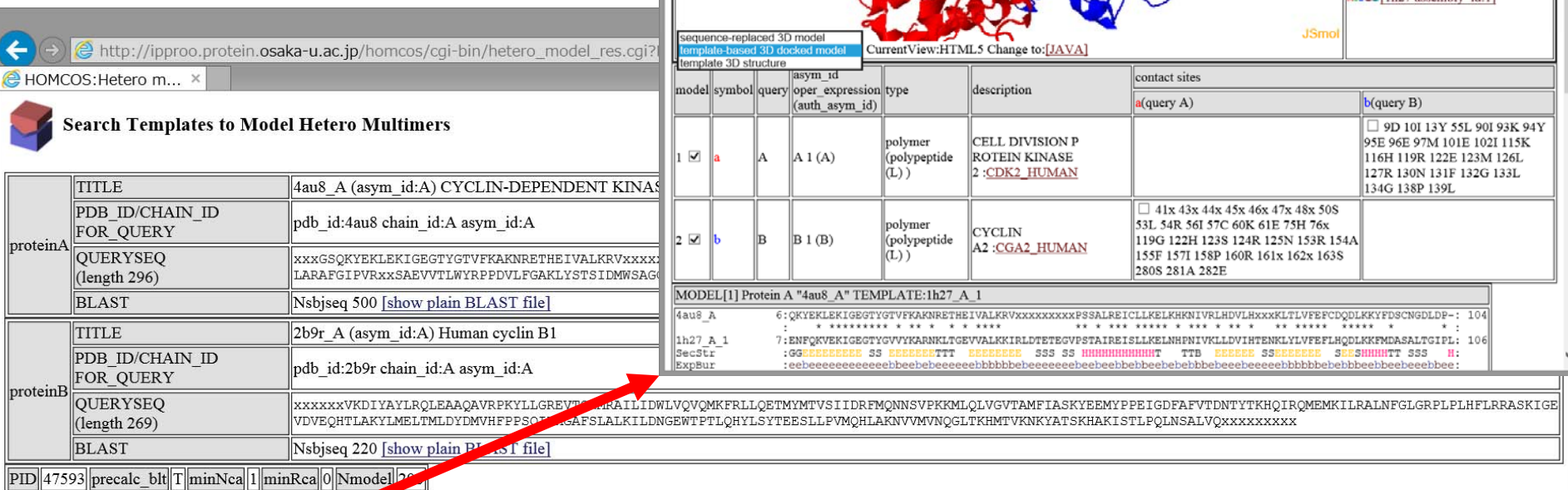
4au8A: Cyclin-dependent protein kinase 5
2b9rA: G2/mitotic-specific cyclin B1

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



-  sequence-replaced model
-  template 3D structure
-  template-based 3D docking model

の三つを、表示・ダウンロード可能



2b9rA (CCNB1)

```
asmb1:assembly id      Nres: number of proteins homologous to query A or B      Nres conali: number of aligned contact residues
```

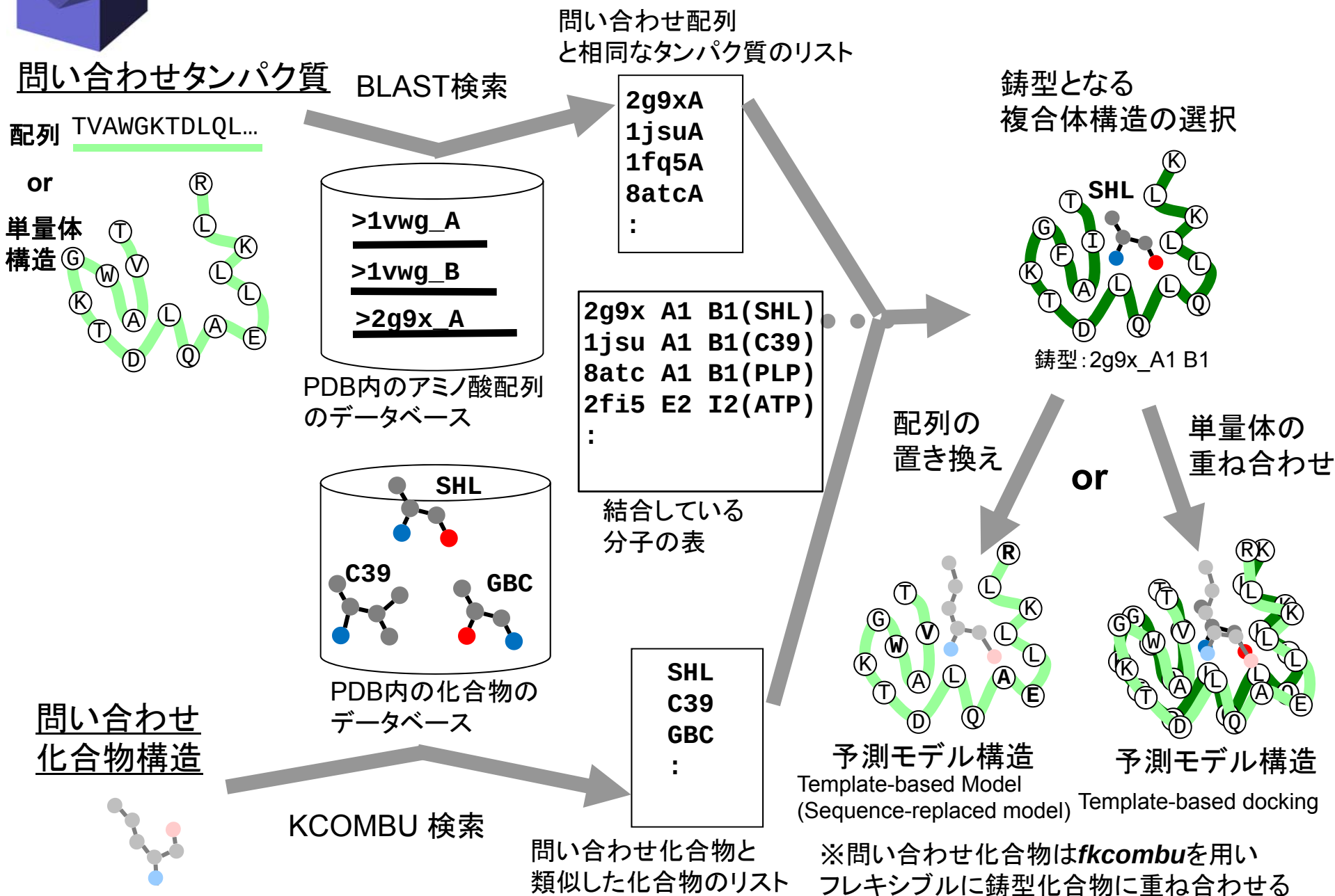
PDB				Protein A							Protein B						
id	asmb1	NproAB	Nres conali	N	asym oper	UniProtID		296aa	Evalue	SqID (%)	N	asym oper	UniProtID		269aa	Evalue	SqID (%)
 1h27	2-mer	33	1	A 1	CDK2_HUMAN		6e-112	58.6	1	B 1	CGA2_HUMAN		3e-48	39.9			
 1h27	2	2-mer	33	1	C 1	CDK2_HUMAN		4e-113	58.0	1	D 1	CGA2_HUMAN		3e-48	39.9		
 3f5x	1	8-mer	32	4	C 1 C 2 A 3 A 4	CDK2_HUMAN		2e-112	58.6	4	D 1 D 2 B 3 B 4	CCNA2_HUMAN		2e-48	39.9		



HOMCOSを用いた 化合物－蛋白質複合体の予測



化合物-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

タンパク質: PDB_ID: (2rh1, 1fin, 4hbb,...) CHAIN_ID: (A, B, ...)

化合物: PDBの3文字の化合物コード: IRE (AU, ATP, G39,...)

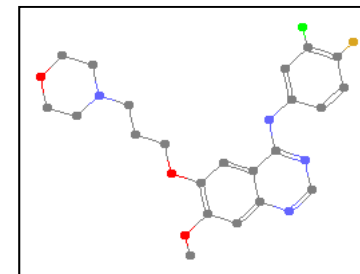
UniProt ID: CDK3_HUMAN

SMILES文字列: CC(NCC(COe1cccc3ne2cccc2e31)O)C

化合物ファイルのアップロード: SDF MOL MOL2 PDB

Reset values SEARCH

Iressa/Gefitinib (IRE)

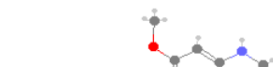


[illegible]

複合体のモデル構造が表示される

標的化合物(IRE) 鑄型化合物(DTQ)

3dmodel	PDB	Protein				Compound		
	id	asmb	UniProt	305aa	eval	SqID	name	Similarity
	1di8	1 A B	1di8					
			CDK2_HUMAN		1e-156	72.6	 DTQ	65.6

Query Compound	Template Compound(1di8_B_1:DTQ)	Similarity
		65.6(%)

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、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、その特異性を予測することは困難
- ・「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