

東京大学農学生命科学研究科  
アグリバイオインフォマティクス人材養成プログラム  
バイオインフォマティクスリテラシーI  
平成18年6月5日(月)、7日(水) @農学部2号館化学第一講義室

# 立体構造予測 II [Web版]

二次構造予測、立体構造予測(ホモロジーモデリング)など

東京大学農学生命科学研究科  
アグリバイオインフォマティクス人材養成ユニット  
特任助手

古田 忠臣



# 講義の予定

## ■ 5月29日(月)、31日(水)

- 構造データベース:PDB

- 構造分類データベース:SCOP、CATH

構造類似性 □ 構造比較サーバー:CE、DALI/FSSP、VAST

配列類似性 □ 相同性検索: BALST、PSI-BLAST、FASTA、CLUSTALW  
1D検索

## ■ 6月5日(月)、7日(水)

2D予測

- 二次構造予測:PSIPRED、PHDsec、PREDETOR、NPS@

3D予測

- 立体構造予測 … 参考:CASP

- 比較モデリング法

- ホモロジーモデリング: **MODELLER**、**SWISS-MODEL**

- フォールド認識法: meta server (3D-Jury)

- de novo / ab initio 予測法: Robetta など

# タンパク質立体構造予測とは？

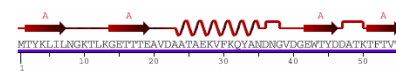
- アミノ酸配列情報(一次構造情報)【問題】を基に
- 物理化学的・情報科学的手法を用いてそのタンパク質の立体構造(三次構造情報)【答え】を予測する
- 創薬など公益性の高い分野に応用されている

【問題】 一次構造: **アミノ酸配列**

例)

MTYKLILNGKTLKGETTTEAVDAAT  
AEKVFKQYANDNGVDGEWYDDA  
TKTFTVTE

二次構造:  $\alpha$  ヘリックス  
 $\beta$  ストランド  
コイル(ターン)

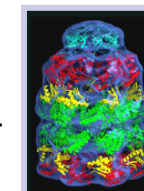


【答え】 三次構造: **タンパク質立体構造**

例) 下図



四次構造: 複合体  
超分子



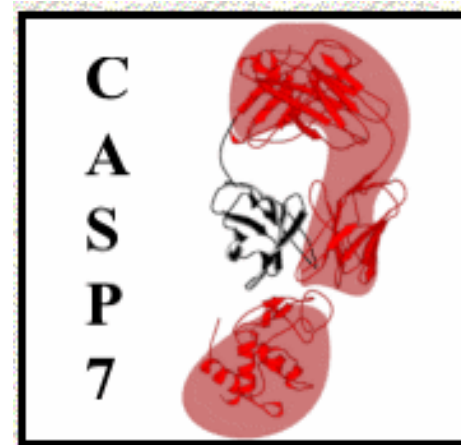


# (参考) CASP: 聖杯の探索 (The Search for the Holy Grail)

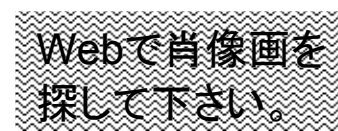
Critical Assessment of Techniques for Protein Structure Prediction

タンパク質立体構造予測の国際コンテスト [URL] <http://predictioncenter.gc.ucdavis.edu/>

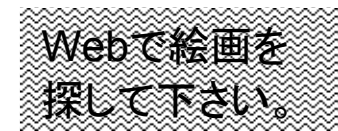
- 2年に一度開催される
- X線解析、NMR解析の実験研究者から、その年、構造が決定されるタンパク質のアミノ酸配列【問題】が提供される
- Humanは約3週間、Serverは48時間で構造を予測し、モデルを5つまで提出
- 年末の会議で、GDT\_TS等のスコアで予測構造を決定された構造【答え】と比べて評価
- 他、ドメイン予測、機能予測などもある



	Year	Targets	Predictors
CASP1	1994	33	35
CASP2	1996	42/42 (T0001-T0042)	72
CASP3	1998	43/42 (T0043-T0085)	98
CASP4	2000	43/62 (T0086-T0128)	163
CASP5	2002	67/67 (T0129-T0195)	215
CASP6	2004	64/87 (T0196-T0282)	208
CASP7	2006	(T0283- )	96



Leonardo da Vinci



The Last Supper

実践 バイオインフォマティクス ゲノム研究のためのコンピュータスキル オライリー・ジャパン 2002

10.2.1節 CASP: 聖杯の探索

# 問題の分類・難易度

易 ↑ ↓ 難	比較モデリング法 Comparative modeling (CM) <ul style="list-style-type: none"><li>・Easy(BLAST)</li><li>・Hard(PSI-BLAST)</li></ul>	Homology Based Modeling	Template-based modeling
	フォールド認識法 Fold recognition (FR) <ul style="list-style-type: none"><li>・Homologous</li><li>・Analogous</li></ul>	如何に良い鑄型(template)構造を検索し、 良いアラインメントを得るか！	
	新規フォールド New fold (NF) de novo / ab initio 予測法 de novo / ab initio prediction	Non-homology Modeling	Template-free modeling

# 構造モデル構築に用いる主な手法・ツール

## ■ CM(比較モデリング)

- ホモロジーモデリング ・BLAST, PSI-BLAST, FASTA, SSEARCHなど
- ・MODELLER, SWISS-MODELなど

- 二次構造予測 ..... PHDsec, PSIPRED, NNpredict, Jpred, NPS@など
- Transmembrane region prediction ..... SOSUI, MEMSAT, TMHMM, PHDhtmなど
- Domain search / parsing ..... Pfam, ProDom, TIGRFAM, RPS-BLAST/CDDなど
- Motif(s)/block(s) search ..... PROSITE, BLOCKSPRINTSなど

(以下、上記情報を用いる)

## ■ FR(フォールド認識)

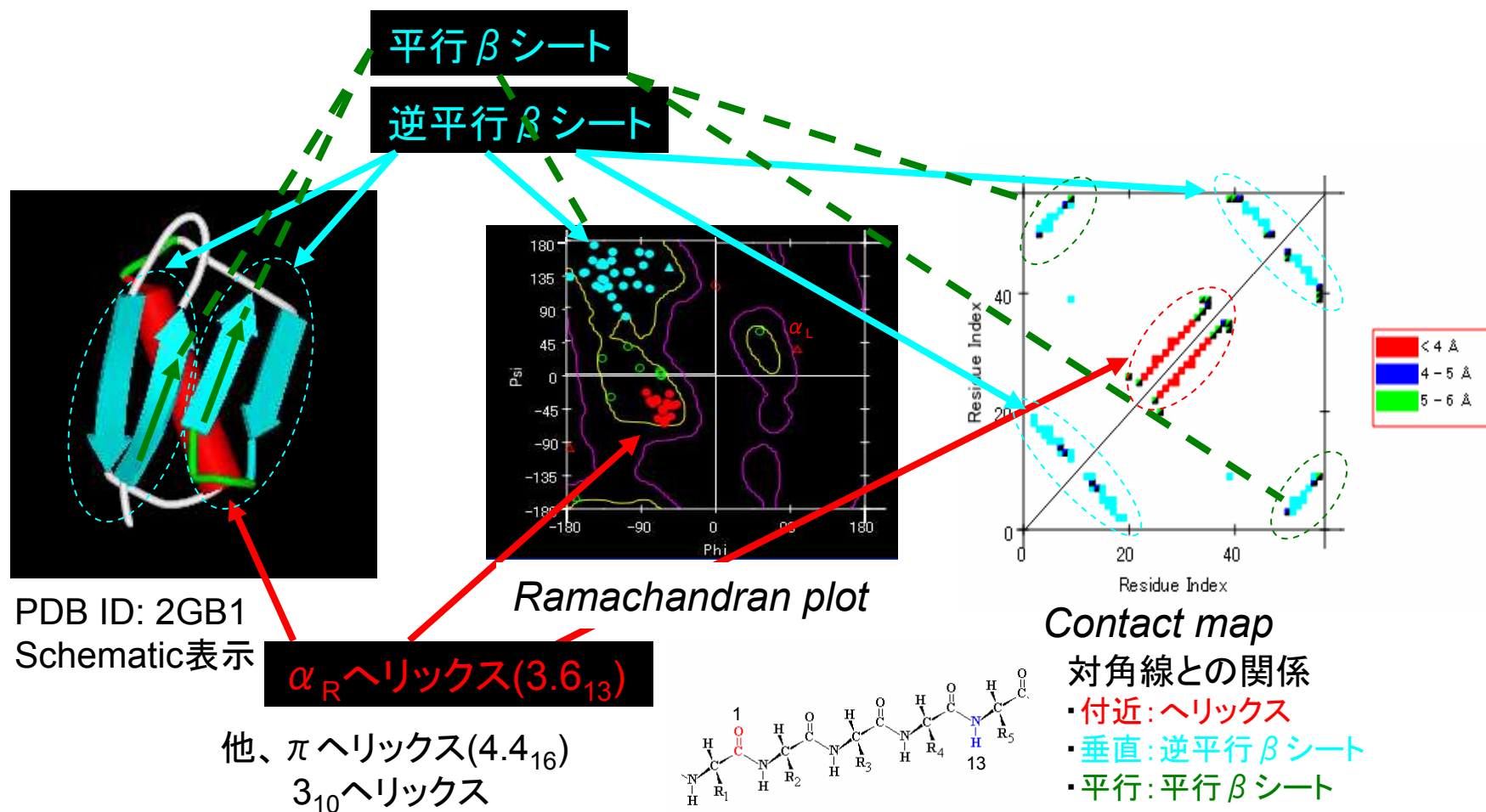
- Threading ..... 3D-PSSM, FUGUE2, mGenThreaderなど
  - Consensus ..... meta server (3D-Jury)
- Model building ..... MODELLER, SWISS-MODELなど

## ■ NF(新規フォールド)

- Fragment Assembly ..... ROBETTA, ProtInfo, ROKKYなど

# 二次構造

1. DS1.5で「File」→「Open URL...」→「2GB1」
2. 「Chart」→「Ramachandran plot」
3. 「Chart」→「Contact plot」



L. Pauling, R.B. Corey, *PNAS* **37**, 235-240 (1951),

“Atomic coordinates and structure factors for two helical configurations of polypeptide chains”

G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, *J. Mol. Biol.* **7**, 95-99 (1968),

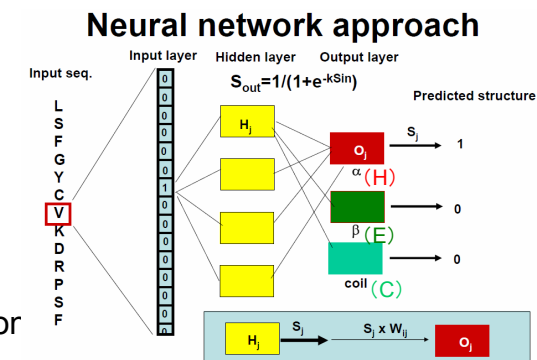
“Stereochemistry of Polypeptide Chain Configuration”

# 二次構造予測: Chou-Fasman, GOR, NN,,,

Webで顔写真を  
探して下さい。

D. Jones

- Chou-Fasman法 1974
  - 15タンパク質から二次構造頻度 $P_\alpha$ 、 $P_\beta$ を計算し、それを基にある閾値以上を $\alpha$  (4/6)、 $\beta$  (3/5)と予測する・・・精度:50-60%
- GOR(Garnier,Osguthorpe,Robson)法 1978
  - 17残基のWindowで配列をスキャンし、その情報量を基に、中心のアミノ酸の二次構造( $\alpha$  (連続4)、 $\beta$  (連続2)、 $\cdot$ )を予測・・・精度:約65%
- ニューラルネットワーク(NN)法 1988ー
  - 13-17残基のWindowでNNを学習させ、二次構造( $\alpha$ 、 $\beta$ 、コイル)を予測
    - NNpredict – Kneller et al. 1990
    - PHD – Rost, Sander 1993
    - PSIPRED – Jones 1999 75-80%
- 最近接法
  - PREDATOR – Frishman, Argos 1995



P.Y. Chou, G.D. Fasman, *Biochemistry* **13**, 222-245 (1974), "Prediction of Protein Cor

J. Garnier, D.J. Osguthorpe, B. Robson, *J. Mol. Biol.* **120**, 97-120 (1978),

"Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins"

N. Qian, T.J. Sejnowski, *J. Mol. Biol.* **202**, 865-884 (1988),

"Predicting the secondary structure of globular proteins using neural network models"

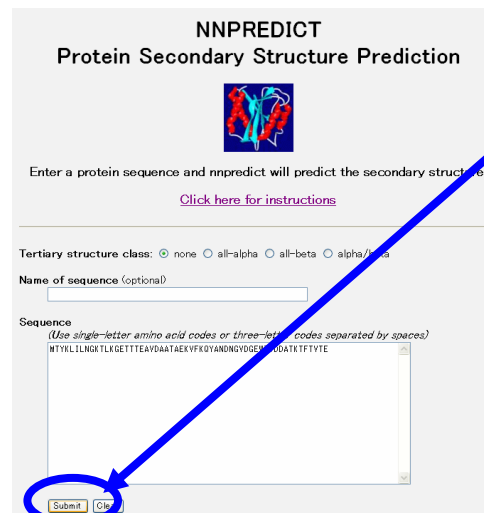
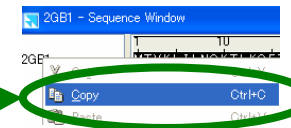
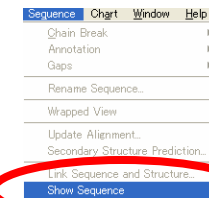
D.T. Jones, *J. Mol. Biol.* **292**, 195-202 (1999), "Protein secondary structure prediction based on position-specific scoring matrices"



# 二次構造予測: NNPREDICT

[URL] <http://www.cmpharm.ucsf.edu/%7Enomi/nnpredict.html>

- 先程の2GB1の配列を表示し
  - 「Sequence」→「Show Sequence」
- 配列を選択して、右クリックでコピー
- NNPREDICTサイトで、配列を貼り付け、「Submit」



NNPREDICT  
Protein Secondary Structure Prediction

Enter a protein sequence and nnpredict will predict the secondary structure.  
[Click here for instructions](#)

Tertiary structure class: ☒ none ☐ all-alpha ☐ all-beta ☐ alpha/beta

Name of sequence (optional):

Sequence  
(Use single-letter amino acid codes or three-letter codes separated by spaces)  
MTYLLNGKTLKGETTTEAYDAAAEKVFYKDYANDNGVGGVYDQATKTFITYE

- 予測結果と答えとを比較

- 予測結果

Results of nnpredict query

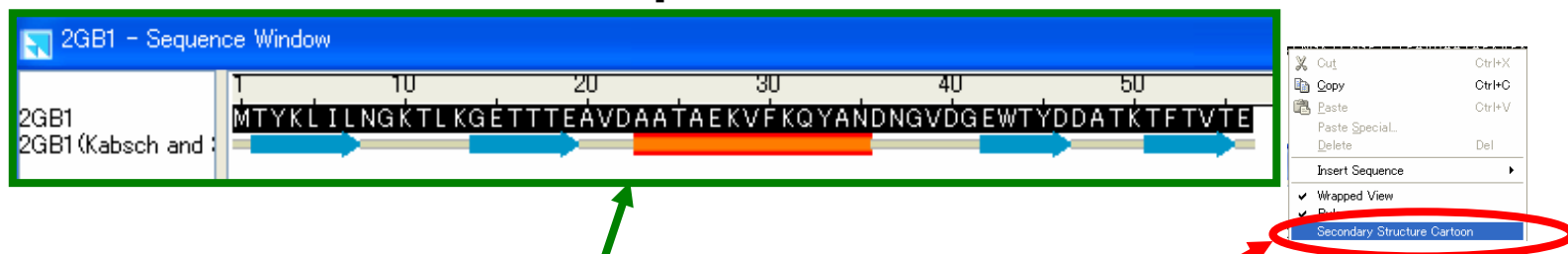
Tertiary structure class: none

Sequence:  
MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEW TYDDATKTFTVTE

Secondary structure prediction (H = helix, E = strand, - = no prediction).  
---EEEE-----HHHHHHHHHHHHHH-----E-----

予測: MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTE  
答え: MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTE

正答率: 37/56=0.66



- DS1.5の配列上で、右クリック「Secondary Structure Cartoon」を選択し、答えを表示

# 二次構造予測：（コンセンサス予測）

## NPS@

[URL] [http://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_seccons.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_seccons.html)

- NPS@サイトを開き、先程と同様に、2GB1の配列を貼り付けて、**MLRC**、**DSC**、**GORIV**、**PHD**、**PREDATOR**にチェックを入れ、「**SUBMIT**」をクリックして下さい。

Pôle BioInformatique Lyonnais  
Network Protein Sequence Analysis  
NPS@ is the IBCP contribution to PBIL in Lyon, France

[\[HOME\]](#) [\[NPS@\]](#) [\[SES\]](#) [\[HELP\]](#) [\[REFERENCES\]](#) [\[NEWS\]](#) [\[MPSA\]](#) [\[ANTHEPROT\]](#) [\[Geno2D\]](#) [\[SuMo\]](#) [\[Positions\]](#) [\[PBIL\]](#)

Wednesday, February 1st 2006: corrected chustaw options for nucleotide sequence alignments ([see news](#))  
When sending automatic requests on NPS@, please use HTTP POST method not GET.

### CONSENSUS SECONDARY STRUCTURE PREDICTION

[\[Abstract\]](#) [\[NPS@ help\]](#) [\[Original server\]](#)

Choose methods :

- ☐ SOPM (Geourjon and Deleage, 1994) [Choose parameters](#)
- ☐ SOPMA (Geourjon and Deleage, 1995) [Choose parameters](#)
- ☐ HNN (Guermeur, 1997)
- ☒ MLRC on GOR4, SIMPA96 and SOPMA (Guermeur *et al.*, 1998)  
When MLRC is checked GOR4, SOPMA and SIMPA96 predictions are not displayed.
- ☐ DDM (Deleage and Roux, 1987)
- ☒ DSC (King and Stenberg, 1996)
- ☐ GOR I (Garnier *et al.*, 1978) [Choose parameters](#)
- ☐ GOR III (Gibrat *et al.*, 1987)
- ☒ GOR IV (Garnier *et al.*, 1996)
- ☒ PHD (Rost *et al.*, 1994)
- ☒ PREDATOR (Argos *et al.*, 1996) [Choose parameters](#)
- ☐ SIMPA96 (Levin *et al.*, 1996)

Sequence name (optional) :

Paste a protein sequence below : [help](#)  
MTYKILNGKTLKGETTTEAVDAATAERVFQYANDNGVDGEWYDDATKTFVTVE

Output width :

## ■ 予測結果と答えとを比較

Consensus prediction result for : UNK\_260170

View Consensus in: [[MPSA \(Mac, UNIX\)](#) , [About...](#)] [[AnTheProt \(PC\)](#) , [Download...](#)] [[HELP](#)]

	10	20	30	40	50
UNK_260170	MTYKLILNGKTLKGETTTEAVDAATAAEKVFKQYANDNGVDGEW	TYDDATKTFTVTE			
DSC	cccccccccccccccccccc	hhhhhhhhhhhhhhhh	cccccccccccccccccccc	cccccccccccccccccccc	cccccccccccccccccccc
MLRC	cccccccccccccccccccc	hhhhhhhhhhhhhhhh	cccccccccccccccccccc	cccccccccccccccccccc	cccccccccccccccccccc
PHD	cccccccccccccccccccc	hhhhhhhhhhhhhhhh	cccccccccccccccccccc	cccccccccccccccccccc	cccccccccccccccccccc
Predator	cccccccccccccccccccc	hhhhhhhhhhhhhhhh	cccccccccccccccccccc	cccccccccccccccccccc	cccccccccccccccccccc
Sec.Cons.	cccccc?cccc?eee?hhhhhhhhhhhhhh?cccccccccccccccccccc				



予測: MTYKLILNGKTLKGETTTEAVDAATAAEKVFKQYANDNGVDGEWTYDDATKTFTVTE

答え: MTYKLILNGKTLKGETTTEAVDAATAAEKVFKQYANDNGVDGEWTYDDATKTFTVTE

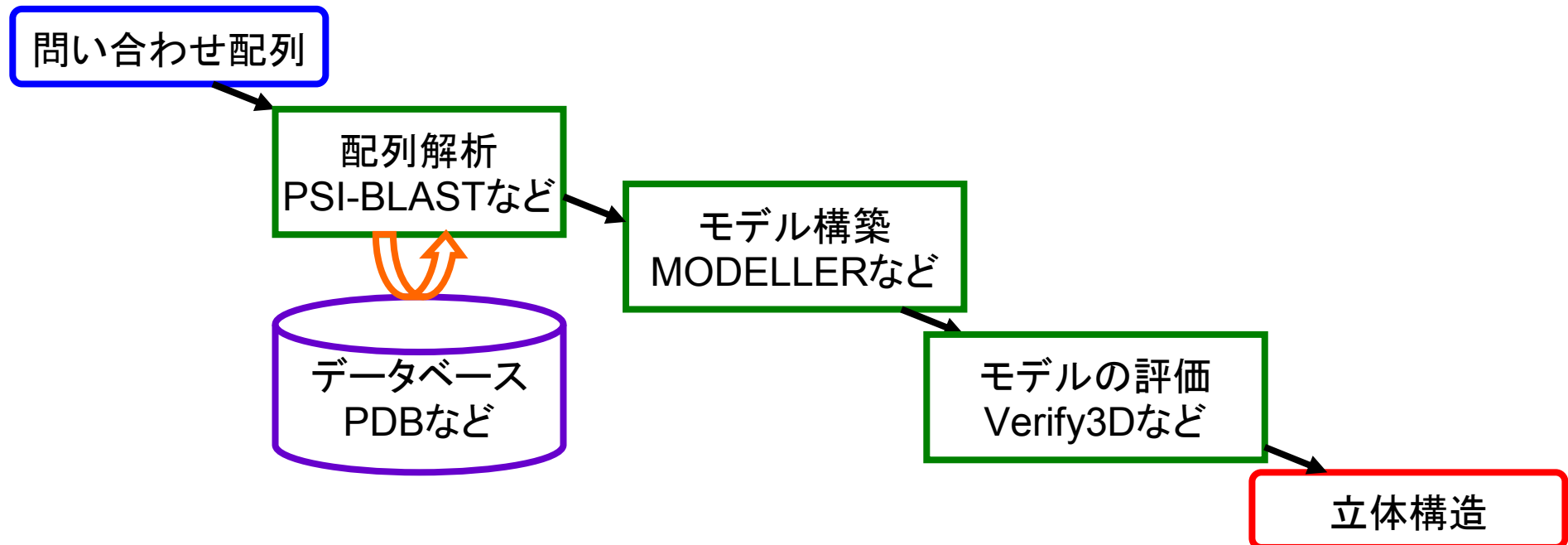
正答率: 47/56=0.83

一般に、コンセンサスを取るほうが良い予測になります。

→ 時代はコンセンサス！

# ホモロジーモデリング (比較モデリング)

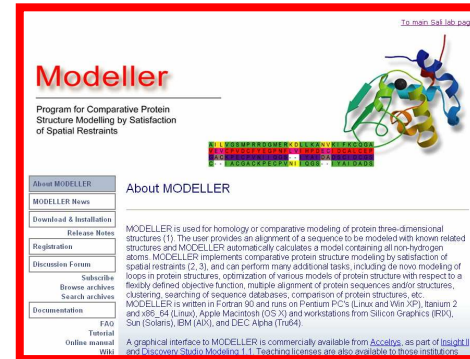
- 「問い合わせ配列」に対して、データベースを用いて配列解析を行い、検索された鋳型・アラインメントを基に「立体構造」を構築することを、ホモロジーモデリング(比較モデリング)と言います。



# ホモロジーモデリングの標準的なツール MODELLER、SWISS-MODEL

## ■ MODELLER

- [URL] <http://www.salilab.org/modeller/>
- ダウンロードして、手元で実行
  - 力場: CHARMM22

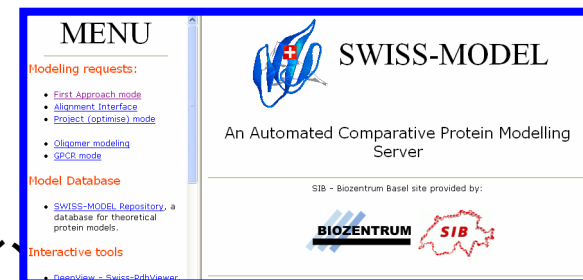


Webで顔写真を  
探して下さい。

A. Sali

## ■ SWISS-MODEL

- [URL] <http://swissmodel.expasy.org/>
  - →「First Approach mode」
  - メールアドレス、名前、タイトル、配列を入力し、「Normal mode」をチェックして、「Send Request」をクリック
  - 入力したメールアドレスに、モデル構造(PDB)が送信される
    - 力場: GROMOS96

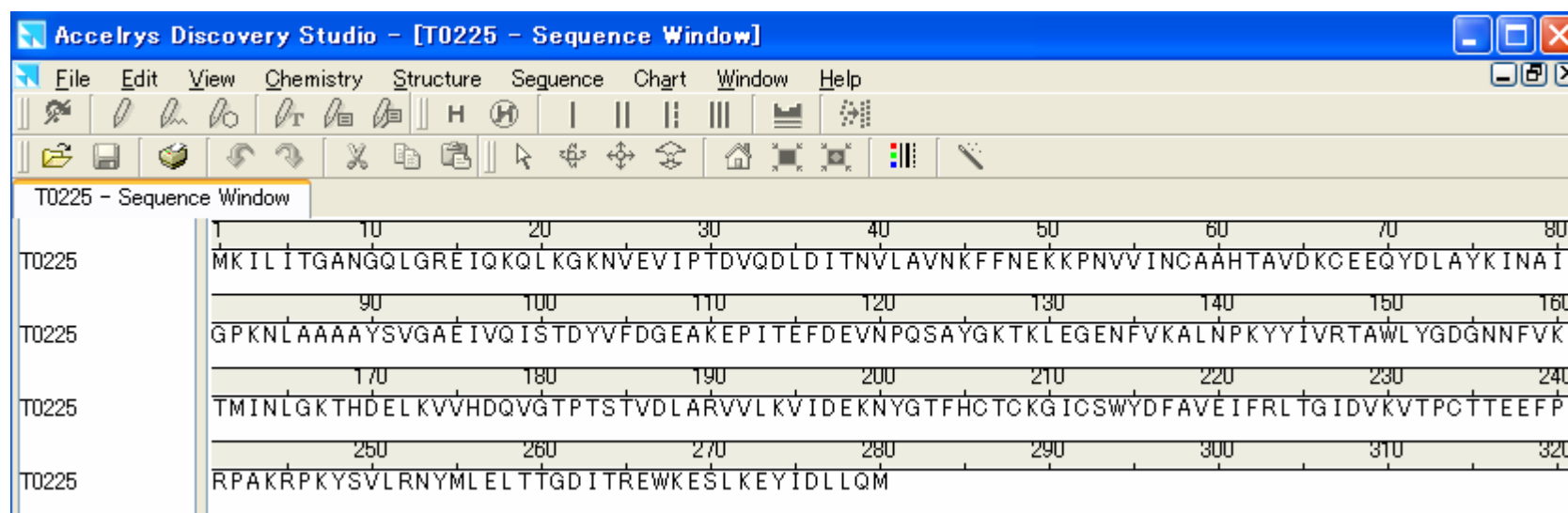


A. Sali, T.L. Blundell, *J. Mol. Biol.* **234**, 779-815 (1993), “Comparative protein modelling by satisfaction of spatial restraints”  
M.C. Peitsch, *Biochem. Soc. Trans.* **24**, 274-279 (1996),  
“ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling”

# CASP6のターゲット:T0225のモデル構造をホモロジーモデリングにより構築してみましょう

## ① 配列の表示

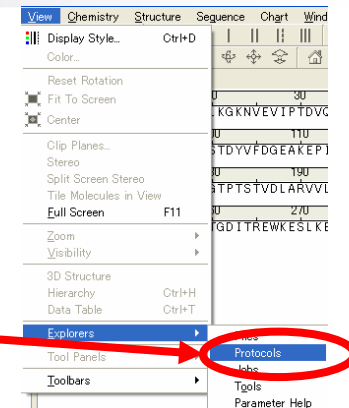
- 講義のリンクページからファイル「T0225.fasta」をクリックし、デスクトップなどに保存
- DS1.5で「File」→「Open」を選択し、今保存した「T0225.fasta」を開く
  - (ファイル名が表示されない時は直接ファイル名を入力してみてください)



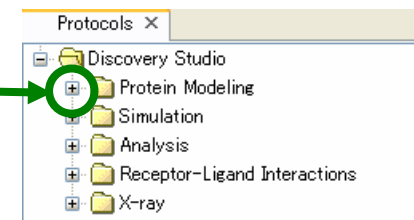
## ②BLASTプロトコルを開く

- Protocols Explorerが表示されていない方は、

- 「View」→「Explorers」→「**Protocols**」を選択

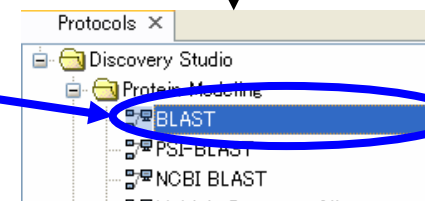


- Protocols Explorerの「Discovery Studio」→「**Protein Modeling**」フォルダを、左の「+」をクリックして開く

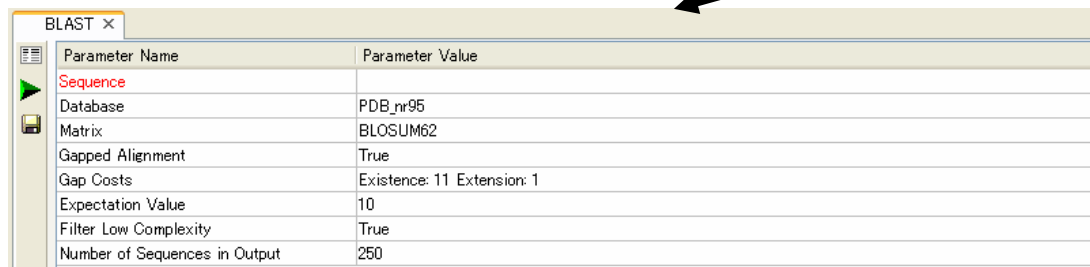


- **BLAST**をダブルクリック

- 右下にBLASTプロトコルのパラメータ設定タブが開きます



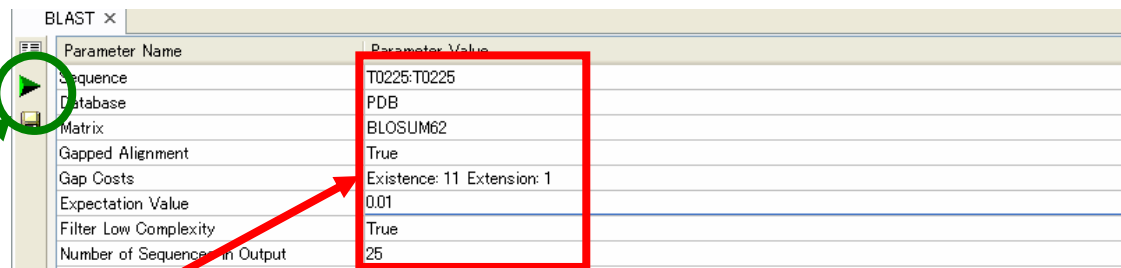
- 同時に、左下にHelp、Jobsのタブも開きます



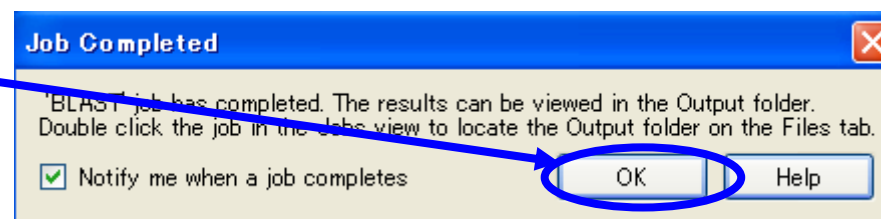


### ③ 相同性検索BLASTの実行

- 設定を  
配列を「**T0225:T0225**」  
データベースを「**PDB**」  
E-valueを「**0.01**」  
出力配列数を「**25**」と変更し  
て
- **BLASTを実行**して下さい
- Job(検索)が始まり
- 20秒ほどで終了します。
  - 「**OK**」をクリック

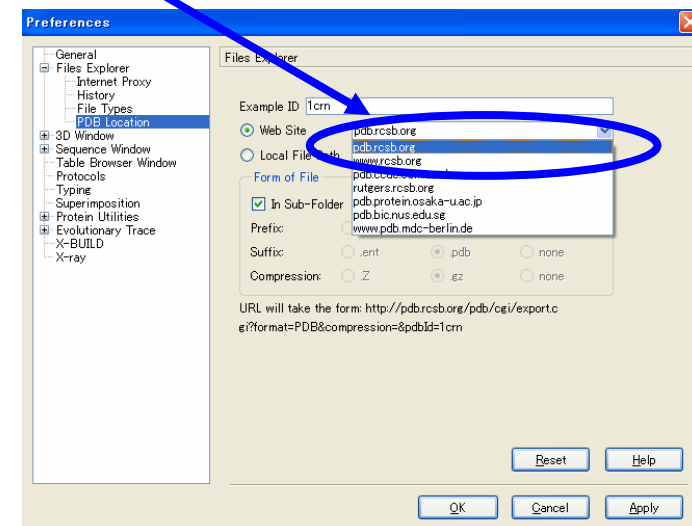


Protocol Name	Status	Elapsed Time	Details
BLAST	Running...	6 sec	
Multiple Sequenc...	Success	13 sec	Comple
Multiple Sequenc...	Success	13 sec	Comple
Build Models	Success	1 min 18 sec	Comple
Build Models	Success	60 sec	Comple
Verify Protein	Success	18 sec	Comple
Build Models	Success	1 min 15 sec	Comple



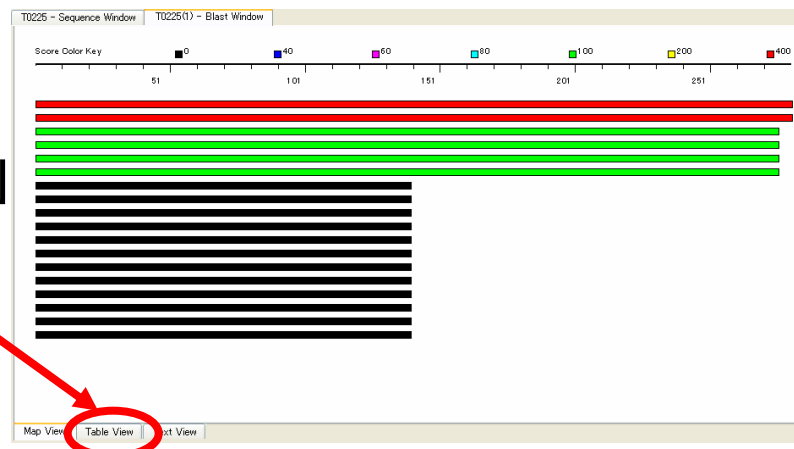
## 【注意】 他の設定は変更しないで下さい

- そのままの設定では、BLAST検索結果から構造が上手くダウンロードできませんので、以下の様に設定を変更して下さい。
  - 「Edit」→「Preferences」を選択
  - 「Files Explorer」の左の「+」をクリック
  - 「PDB Location」をクリック
  - Web Siteを「www.rcsb.org」から「[pdb.rcsb.org](http://pdb.rcsb.org)」を選択
  - 設定が変更できたので「OK」を選択



## ④アラインメントを表示し、鋳型構造をダウンロードする

- 相同性検索の結果がスコアで色分けされて表示されます。
- Blast Windowの「Table View」をクリックする



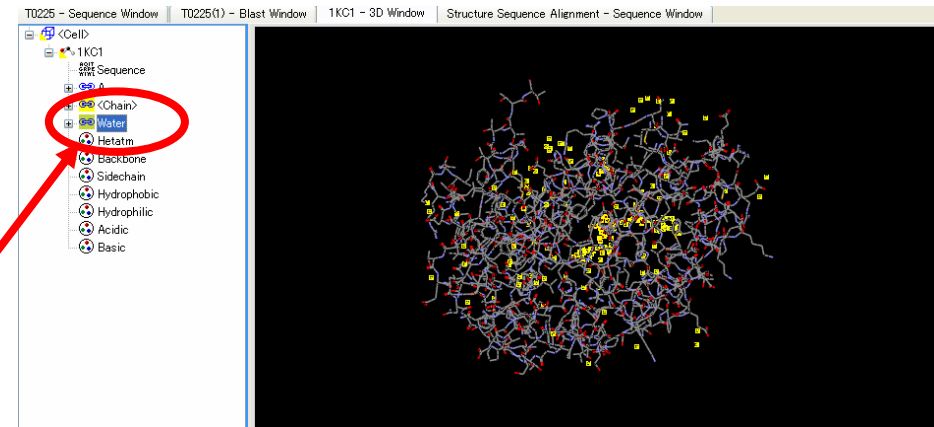
- 1VL0は答えなので、1KC1:Aを鋳型に選びましょう。
  - 複数の鋳型を選択することもできます。
- 1KC1\_Aを選択し、右クリックで「Load Structure and Alignment」を選択する

↓

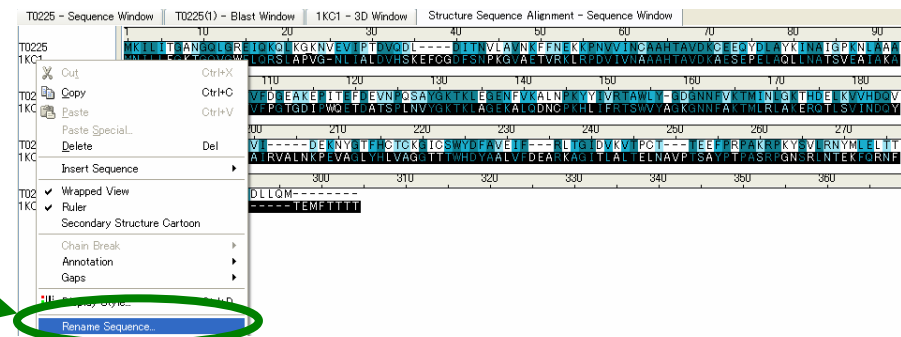
	Title/Description	Accession	Sequence Length	Alignment Length	Bit Score	E-value	Identity	Positive
1	structureX1vl0 ...	1VL0_A	281	280	571.237	1.31381e-163	100	100
2	structureX1vl0 ...	1VL0_B	280	280	571.237	1.31381e-163	100	100
3	structureX1kc1 ...	1KC1_A	298	298	7.46433e-042	36	50	50
4	structureX1kbz ...	1KBZ_A	298	298	7.46433e-042	36	50	50
5	structureX1kc3 ...	1KC3_A	298	298	7.46433e-042	36	50	50
6	structureX1n2s ...	1N2S_A	298	298	7.46433e-042	36	50	50
7	structureX1kvs ...	1KVS	338	338	0.00306592	25	40	40
8	structureX1udc ...	1UDC	338	338	0.0052295	25	40	40
9	structureX2udp ...	2UDP_A	338	338	0.0052295	25	40	40
10	structureX1kvr ...	1KVR	338	338	0.0032994	25	40	40
11	structureX1nah ...	1NAH	338	338	0.00892018	25	40	40
12	structureX1lrl ...	1LRL_A	338	164	36.965	0.00892018	25	40
13	structureX1nai ...	1NAI	338	164	36.965	0.00892018	25	40
14	structureX1xel ...	1XEL	338	164	36.965	0.00892018	25	40
15	structureX1lrk ...	1LRK_A	338	164	36.965	0.00892018	25	40
16	structureX1udb ...	1UDB	338	164	36.965	0.00892018	25	40
17	structureX1uda ...	1UDA	338	164	36.965	0.00892018	25	40
18	structureX1lrj ...	1LRJ_A	338	164	36.965	0.00892018	25	40

## ⑤構造とアラインメントの修正

- 構造がダウンロードされ
  - ☐ 1KC1 – 3D Window
- アラインメントが表示されます
  - ☐ Structure Sequence Alignment - Sequence Window



- <Chain>やWaterを削除し、chain Aのみにして下さい。
- 鋳型の配列名を右クリックし、「Rename Sequence」を選択し、「1KC1」から「1KC1\_A」に変更して下さい



## ⑥モデル構築の実行

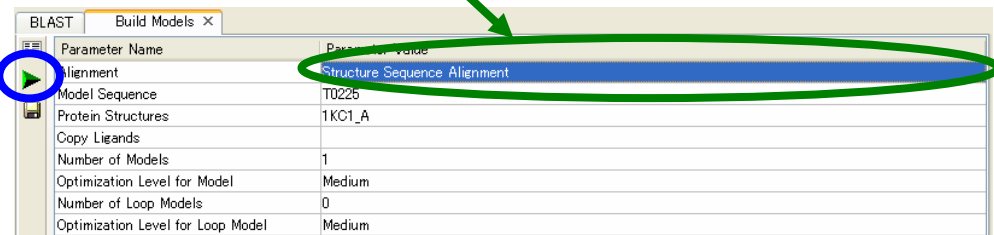
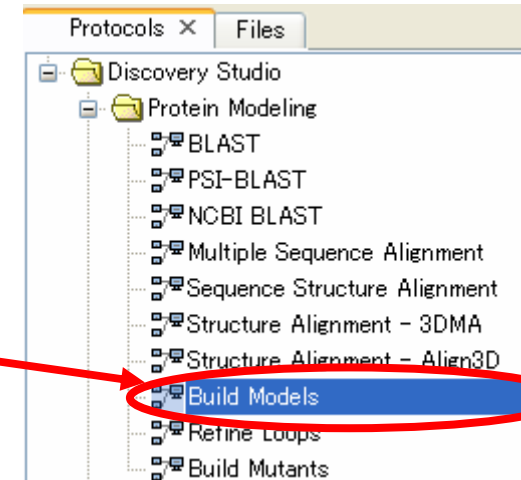
- Protocols Explorerから  
「Build Models」をダブルクリック

- 右下のBuild Modelsタブから、設定を

Alignment — Structure Sequence Alignment  
を選択する

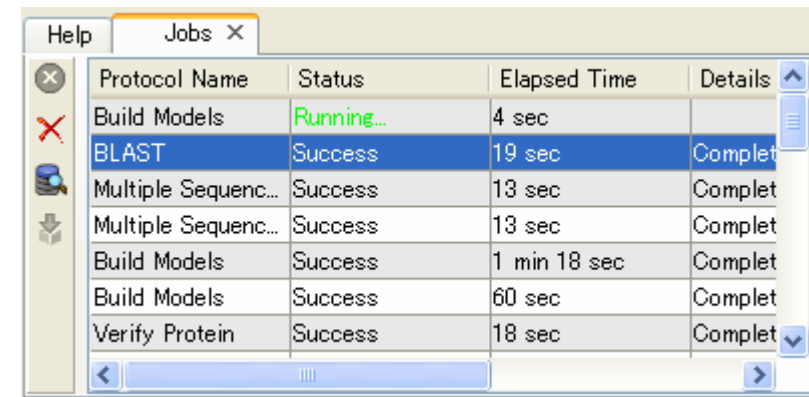
- ☐ 同時に、以下の様に変更される
  - Model Sequence – T0225
  - Protein Structures - 1KC1\_A

- 設定ができたので、  
モデル構築を実行

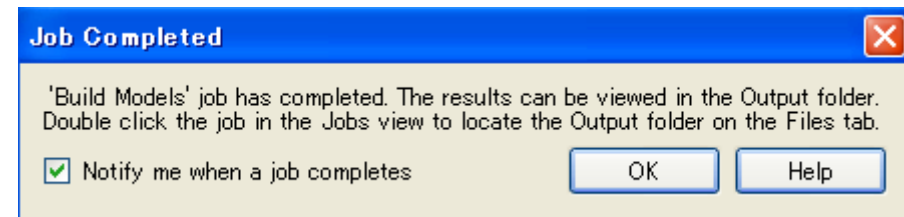


■ Jobが実行され

■ 4, 5分で終了する



Protocol Name	Status	Elapsed Time	Details
Build Models	Running...	4 sec	
BLAST	Success	19 sec	Comple
Multiple Sequenc...	Success	13 sec	Comple
Multiple Sequenc...	Success	13 sec	Comple
Build Models	Success	1 min 18 sec	Comple
Build Models	Success	60 sec	Comple
Verify Protein	Success	18 sec	Comple

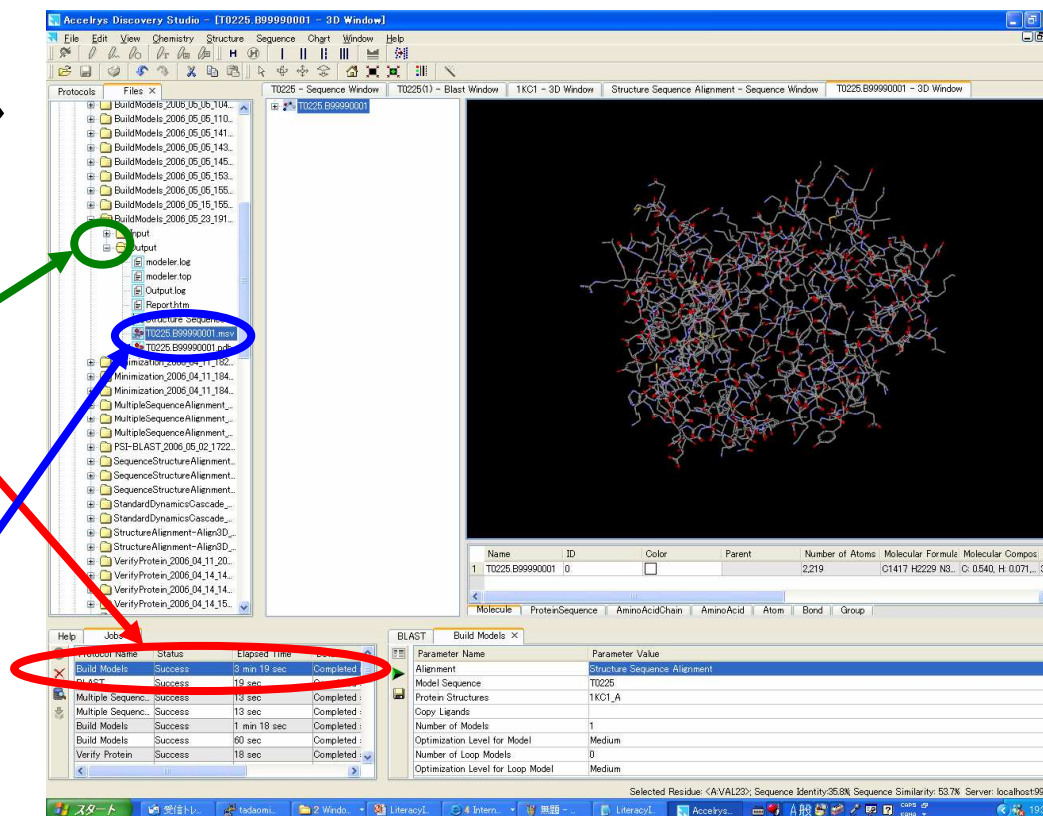


■ さて、何人上手く実行できるでしょうか？

、、、、

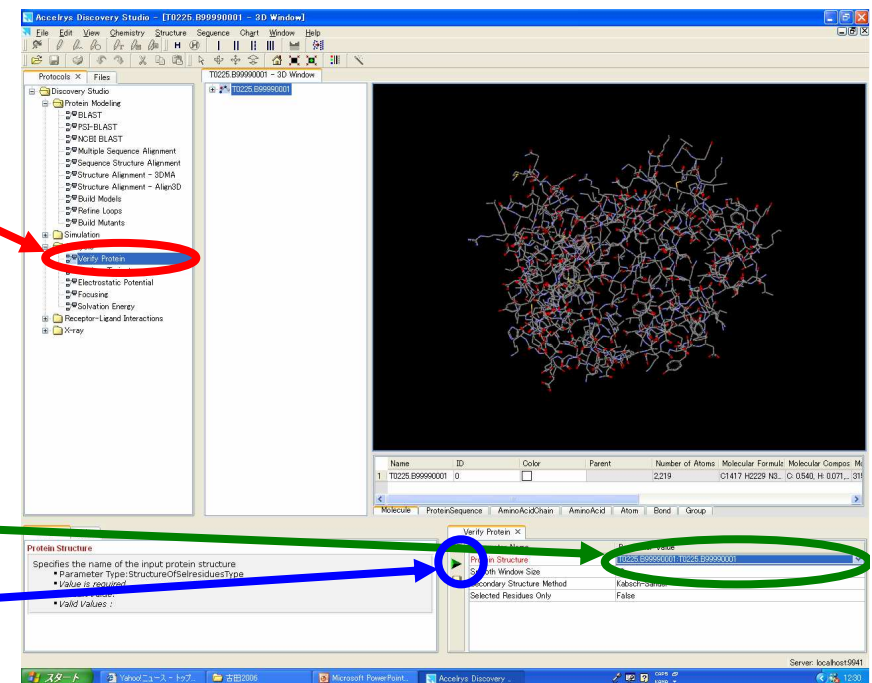
# 構築された構造の表示

- 左下のJobsタブから「Build Models」をダブルクリック
- Files ExplorerのBuildModels\_...のOutputの左の「+」をクリック
- 「T0225.B99990001.msv」をダブルクリック



# 構造の評価: Verify3D、PROCHECK、PROSA

- 構築したモデル構造を評価しましょう
- Protocols→Analysis→Verify Proteinをダブルクリック
- 右下のVerify Proteinタブで、設定で以下を選択
  - ☐ Protein Structure – T0225.B99990001: ,,,
- 実行



J.U. Bowie, R. Luthy, D. Eisenberg, *Science* **253**, 164-170 (1991),

“A method to identify protein sequences that fold into a known three-dimensional structure”

R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, *J. Appl. Cryst.* **26**, 283-291 (1993),

“PROCHECK: a program to check the stereochemical quality of protein structures”

M.J. Sippl, *Proteins* **17**, 355-362 (1993), “Recognition of Errors in Three-Dimensional Structures of Proteins”



- 20秒ほど実行されたのち終了します。

- 終了すると、**結果**が表示されます。

- 3D windowの下スクロールバーを右端までスクロールするとVerify Scoreの欄が表示されます(97.78)。

- 結果の表示を以下の様に変更してみましょう。

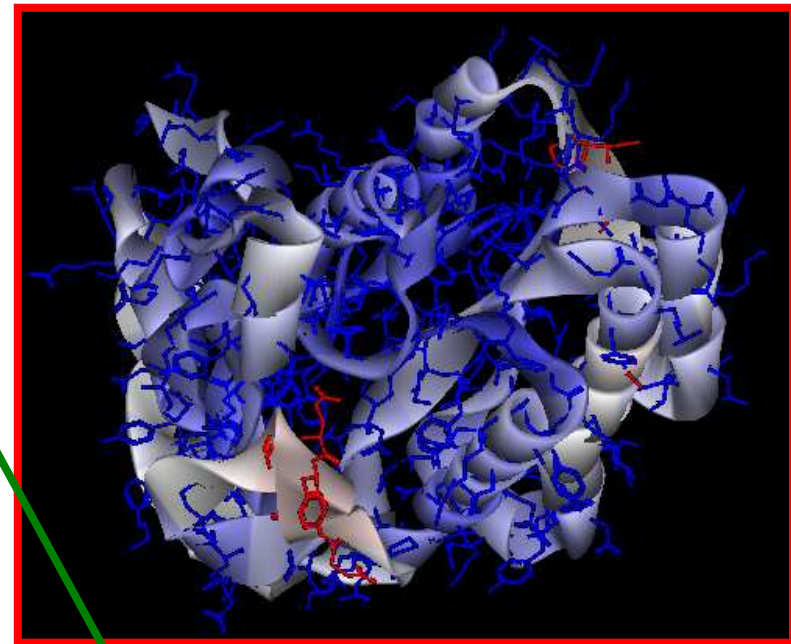
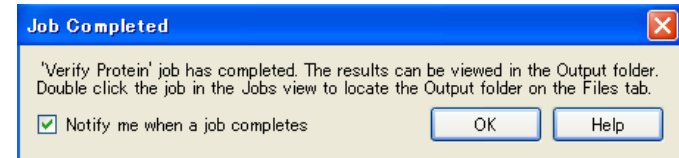
- 「Ctrl」+「D」・・・表示(Display Style)の変更

☐ Atom – None

☐ Protein – Solid Ribbon

■ Display Size 0.5

Protocol Name	Status	Elapsed Time	Details	Start Date	Output Locati
Verify Protein	Running	4 sec		2006-05-24 12:36	C:\Documents
Build Models	Success	3 min 19 sec	Completed succe	2006-05-23 19:16	C:\Documents
BLAST	Success	19 sec	Completed succe	2006-05-23 16:57	C:\Documents
Multiple Sequenc...	Success	13 sec	Completed succe	2006-05-15 15:51	C:\Documents
Multiple Sequenc...	Success	13 sec	Completed succe	2006-05-11 11:23	C:\Documents
Build Models	Success	1 min 18 sec	Completed succe	2006-05-05 15:54	C:\Documents
Build Models	Success	60 sec	Completed succe	2006-05-05 15:38	C:\Documents



Verify Score	Verify Expected H	Verify Expected L
97.78	127.341	57.3034

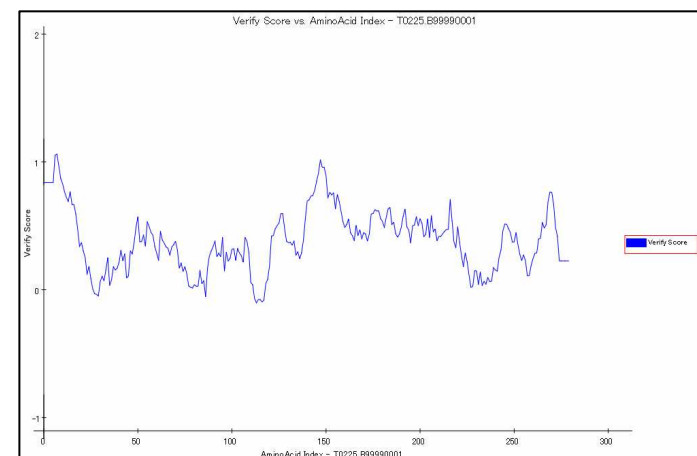
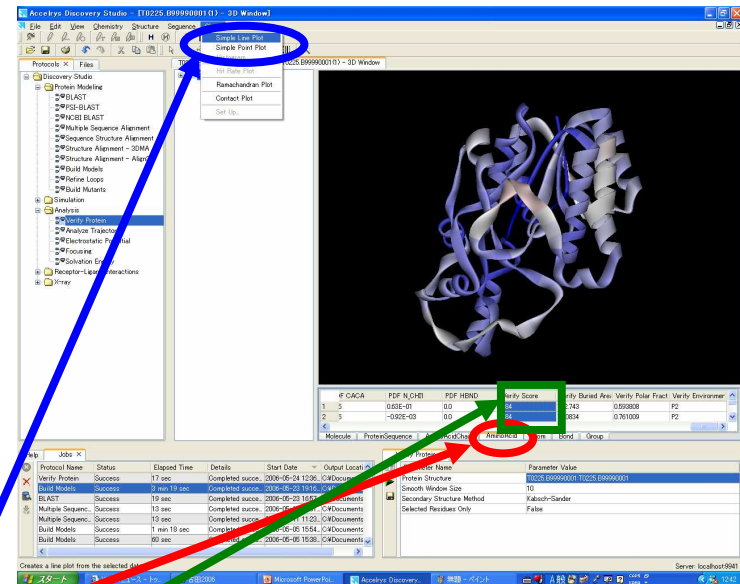
3D Windowの下の方にデータが表示されていない場合、  
「View」→「Data Table」にチェックを入れて表示させてください。

- 構造は残基毎に以下の様に色分けされて表示されています。

- 良い — 青
- 普通 — 白
- 悪い — 赤

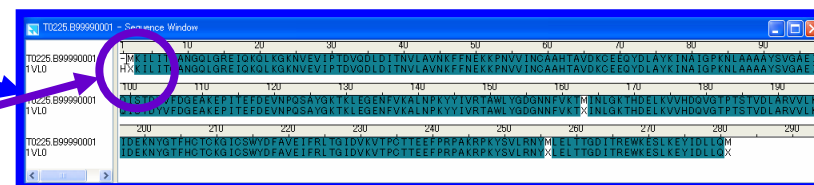
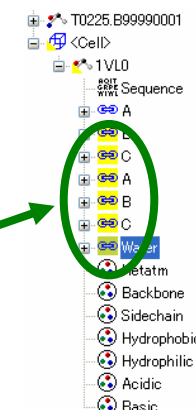
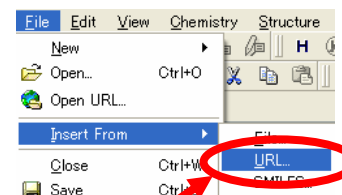
- それでは、残基毎のスコアの図を作成してみましょう。

- Amino Acidタブをクリック
- 右にスクロールし、Verify Scoreの列を選択
- 「Chart」→「Simple Line Plot」を選択
- 右図の様なプロットが表示されます。
  - $S < 0$  悪い(ミスフォールド!?)
  - $0 < S < 0.5$  普通
  - $0.5 < S$  良い

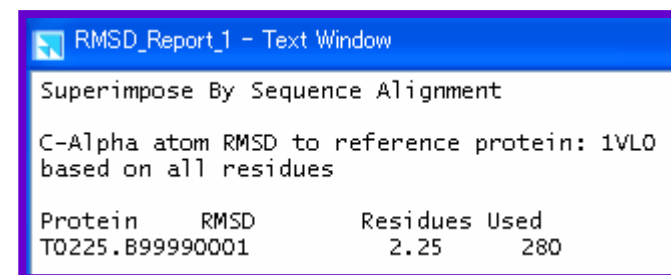
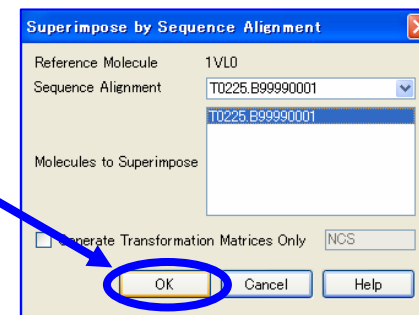
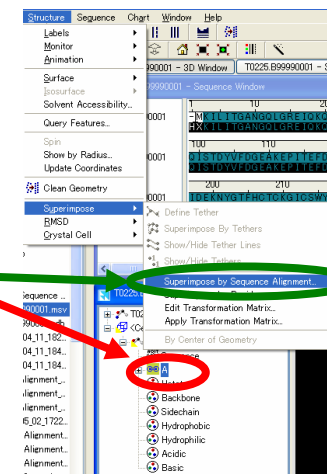
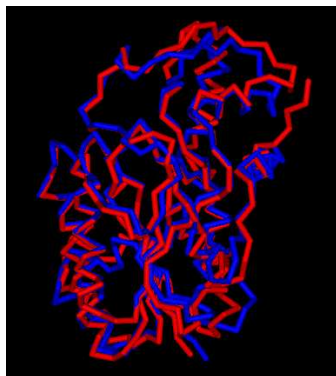


# RMSDを計算する

- それでは、答え(1VL0)とどれくらい近い構造が構築できたか、RMSDを計算してみましょう。
- (Verify3Dではなく)構築したモデルの3D Windowをアクティブにする
- 「File」→「Insert From」→「URL」を選択
- 1VL0を入力し、構造をダウンロード
- Chain Aの他を選択し削除
- 「Sequence」→「Show Sequence」を選択し、配列を表示
- T0225の先頭にSpace一つ入力し、アラインメントを合わせる。
  - XはMETの代わりにMSEの為
    - MSE = SELENOMETHIONINE

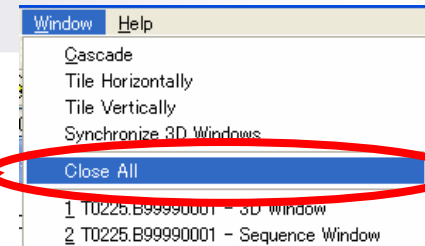


- 3D Windowでchain Aをアクティブにする
- 「Structure」→「Superimpose」→「Superimpose by Sequence Alignment」を選択
- 「T0255,,,」を選択し、「OK」をクリック
- 「Text Window」が表示され、280残基を用いて、RMSDが2.25と表示されました
- 表示を変えると構造がどれくらい似ているか見やすくなります。



参考までに、リガンドを含めてモデリングすると  
(Build Modelsの設定で、Copy Ligands – 1KC1\_A::NDP901)  
RMSD=2.20、Verify Score=100.28に改善されます。

# アラインメントの修正



- 「Window」→「Close All」で全てのWindowを閉じてください
- アラインメントを修正してより良いモデルを構築してみましょう。
- Jobsタブから、先程実行した「Build Models」をダブルクリック
- 「Input」フォルダから以下の2つのファイルを開いてください。
  - 1KC1\_A.pdb
  - Structure Sequence Alignment.bsml
- アラインメントを以下の様に修正してください。

Structure Sequence Alignment - Sequence Window

	1	10	20	30	40	50	60	70	80
1KC1_A	MNILLFGKTGQVGVWELQRSLA-PVGNLIALDVHSKEFCGDFSNPKGVAETVRKL	RPDVIVNAAHAHTAVDKAESEPELAQLLNATSVEA							
T0225	MKILITGANGQLGREIQKQLKGKNVEVIPTDVQD----	LDITNVLAVNKKFNEKKPNVVINCAHAHTAVDKCEEQYDLAYKINAIGPKN							
1KC1_A	IAKAANETGAWVHYSTDYVFPSTGDI	PWQETDATSP	LNVDYGKTKLAGEKALQDNC	PKHL	IFRTSWVYAGKGNNAFAKTMRLAKERQT				
T0225	LAAAYSVAEIVQISTDYVFDGEAKER	ITEFDEVNPQSA	YGKTKLEGENFVKALN	PKYY	IVRTAWLYG-DGNNFVKTMINLGKTHDE				
1KC1_A	LSVINDQYGAPTGAEL	LADCTAHAI	IRVALNKPEVAGL	YHLVAGGTT	TWHDYAALVF	DEARKAGITLALTELNAVPT	SAYPT	PASRP	GN
T0225	LKVVDQVGTPST	STVDLARVVL	KVIDEK----	NYGTFHCTCKG	ICSWYDFAVE	IFRLTG-----	IDVKVTPCTTEEF	PRPAKRPKY	
1KC1_A	SRINTEKFORNFDL	ILPOWELGVKRM	LITEMFTTTT						
T0225	SVLRNYMLELTTGD	ITREWKESLKEYIDLLQM---							

- 先程と同様に、モデル構築を行い、Verify Score、RMSDを計算してみてください。

#### □ モデル構築

- 「Protocols」→「Protein Modeling」→「Build Models」
- Alignment - Structure Sequence Alignment
- Model Sequence – T0225
- Protein Structures – 1KC1\_A
- 実行

#### □ Verify Score計算

- 「Protocols」→「Analysis」→「Verify Protein」
- 構造を選択し、実行

#### □ RMSD計算

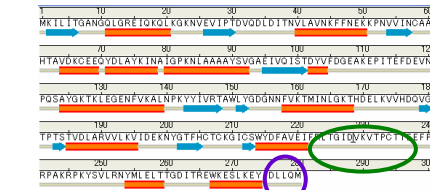
- 答え(1VL0)を挿入
- アラインメントを合わせる
- 「Structure」→「Superimpose」→「Superimpose by Sequence Alignment」

- アラインメントの修正により、以下の様にモデルが改良されました。

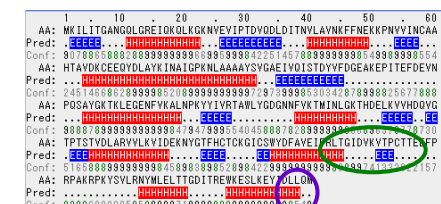
- Verify Score: 97.78 → 98.44
- RMSD: 2.25 → 1.37

Verify Score	Verify Expected T	Verify Expected L	MOLID
98.44	127.341	57.3034	3D192DEC2F1C4...

BLASTアラインメントでの二次構造



PSIPREDの二次構造予測



修正したアラインメントでの二次構造



RMSD_Report_2 - Text Window		
Superimpose By Sequence Alignment		
C-Alpha atom RMSD to reference protein: 1VL0 based on all residues		
Protein	RMSD	Residues Used
T0225.B99990001	1.37	280

アラインメントの修正は、Verify3Dスコアの低い部分および二次構造予測に基づく修正が功を奏しています。



# フォールド認識法 Threadingなど

Webで顔写真を  
探して下さい。

K. Ginalski

- これまでは、BLAST、PSI-BLAST等による相同性検索を用いて主に近縁の配列を検索し、その鋳型・アラインメントを基にモデル構築をおこないましたが、
- マルチプルアラインメント、プロファイル(PSSM)を有効に用いたり、構造配列相関を用いることにより、より遠縁の鋳型を検索することができます。
- これらフォールド認識法を用いた多くのサーバーが存在します。
  - 3D-PSSM, FUGUE2, Sam-T02, mGenThreaderなど
- さらに、それらいくつかのサーバーのメタサーバー(コンセンサス予測をする)もあります。

□ 3D-Jury

□ [URL] <http://bioinfo.pl/meta/>

やはり、時代はコンセンサス！？

The BioInfoBank Meta Server offers a gateway to well-benchmarked protein structure and function prediction methods. Structural models collected from the predictions servers are assessed using the powerful 3D-jury consensus approach.

Submit Queue Servers Benchmarks Join Help

Ads by Google  
**Find Secondary Structures**  
RAPTOR: casp5 winner, gives quality structure prediction. Free Demo.  
[www.bioinformatics.org/raptor](http://www.bioinformatics.org/raptor)

**Sequence analysis tool**  
Easy-to-use free software for sequence alignment & tree building  
[www.genious.com](http://www.genious.com)

**DNA Analysis Software**  
Download a demo of Sequencer and start analyzing your own data today  
[www.genecodes.com](http://www.genecodes.com)

**Structure Prediction Meta Server Input Page**  
0 jobs from 133.11.221. in the last week

Your E-mail:

Target Name:

Amino Acid Sequence only (in one letter code):

Reset Clear Format Submit

Submit domains separately  
Remove coiled coil regions  
Check LiveBench for evaluation of the reliability of the servers  
Results are stored only for 1 month  
Jobs queued for more than 7 days for servers with queue>30 are skipped  
Use is limited to 30 jobs per week per domain  
Contact us in case of problems with interpretation of results  
Some servers return only models, no alignments (target sequence is shown)  
Results published on this server are public and can not be used for patenting

Skip: Queue:

<input type="checkbox"/> PDB-Blast	
<input type="checkbox"/> ESIYPred3D	
<input type="checkbox"/> GRDB	6
<input type="checkbox"/> FFAS03	
<input checked="" type="checkbox"/> Sam-T02	32
<input type="checkbox"/> Superfamily	
<input checked="" type="checkbox"/> INUB	268
<input type="checkbox"/> FUGUE2	3
<input type="checkbox"/> 3D-PSSM	
<input type="checkbox"/> mGenThreader	9
<input type="checkbox"/> pspired	9
<input type="checkbox"/> profsec	1
3D-Jury	102

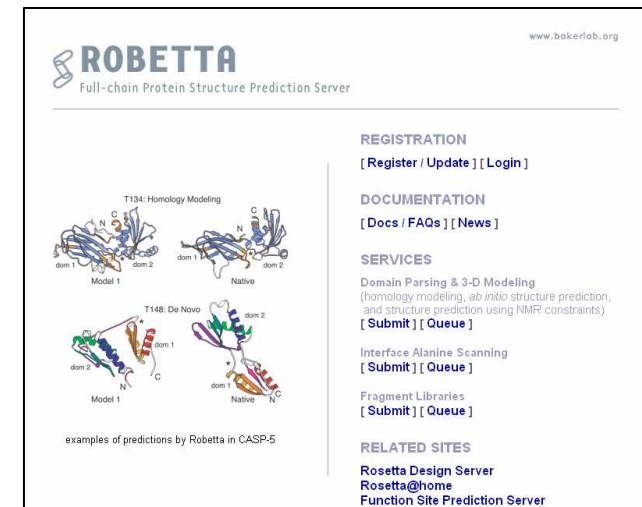
# ab initio / de novo予測法 Fragment Assembly法

Webで顔写真を  
探して下さい。

D. Baker

- 鋳型構造がない、つまり新規フォールドの予測に関して、PSSM相関等で集めた部分構造(フラグメント)を利用する、フラグメントアセンブリ法が主流です。
- D. Bakerが普及させました(Rosetta法)。
  - ROBOTTA = Robot + Rosetta
  - [URL] <http://robetta.bakerlab.org>

立体構造予測において、  
プロフィール(PSSM)は大変重要です。



K.T. Simons et al., *J. Mol. Biol.* **268**, 209-225 (1997),

“Assembly of protein tertiary structures from fragments with similar local sequences using simulate annealing and Bayesian scoring functions”

D. Chivian et al., *Proteins* **53**, 524-533 (2003), “Automated prediction of CASP-5 structures using the Robetta server”



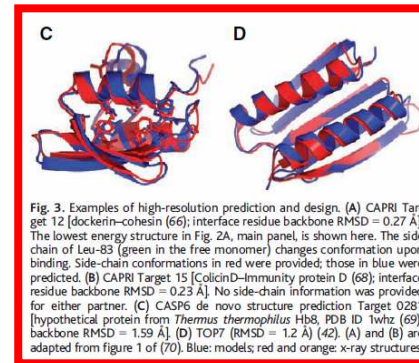
# CASPでの歴史的な予測

## BAKER group:

### T0281 in CASP6

- BAKER groupは、FR/AのターゲットT0281を鋳型を基づかないab initio / de novo予測で**RMSD=1.59 Å**という成功を収めました。
- しかし、BAKER groupでも**NF、FR/Aのターゲット**(25個)に対する予測は、数個の良い予測(RMSD<5 Å)はあるものの、半分以上がRMSD>10 Åというのが現状です。
- ちなみに、**T0281(1WHZ)の構造比較**を行うと、上記の成功は素晴らしいことがわかります。

- Bakerはデザインでも精力的な研究を行っています。
- 最近の話題: **Rosetta@home**
  - SETI, folding@homeの次! ?
  - [URL] <http://http://boinc.bakerlab.org/rosetta/>



Webで顔写真を  
探して下さい  
(Baker & ROKKY)

T0281(1WHZ)  
CE none  
DALI 1cb1 47 RMSD=3.15959  
VAST 1DQ3 52

	RMSD	GDT_TS	a.a.
NF			
T0201	6.063	48.94	94
T0209_2	<b>4.396</b>	57.46	57
T0216_1	24.472	14.11	209
T0216_2	39.184	12.91	164
T0238	22.340	26.52	181
T0241_1	15.794	25.00	117
T0241_2	16.589	21.85	119
T0242	13.249	25.87	115
T0248_2	12.234	31.89	87
FR/A			
T0198	<b>4.907</b>	51.11	225
T0199_3	13.146	25.61	82
T0209_1	12.480	20.61	108
T0212	6.017	55.84	124
T0215	8.081	43.40	53
T0230	10.277	49.27	102
T0235_2	12.152	30.23	43
T0239_1	6.786	46.43	70
T0248_1	<b>3.515</b>	68.35	79
T0248_3	10.079	44.54	87
T0262_1	15.072	30.90	72
T0272_1	<b>3.658</b>	58.53	85
T0272_2	8.493	34.59	99
T0273	36.02	12.583	186
T0280_2	11.727	39.70	51
T0281	<b>1.59</b>	81.78	70

O. Schueler-Furman, et al., *Science* **310**, 638-642 (2005), "Progress in modeling of protein structures and interactions"


B. Kuhlman, et al, *Science* **302**, 1364-1368 (2003), "Design of a Novel Globular Protein Fold with Atomic-Level Accuracy"



## 【課題3】 ホモロジーモデリング

T0229のホモロジーモデリングを行い、結果をPowerPointにまとめよ

1. 実習のリンクページからターゲットの配列 (T0229.fasta)をダウンロード
2. BLAST検索を実行する
3. 答え(1VLA)以外の一致度の高い鋳型を用いてモデル構築を行う
4. 構造の評価 (Verify3D)を行う
5. 答えとのRMSDを計算する
6. 上記を図も含めて、PowerPointに記述する



時間に余裕がある方は以下の課題を行ってください

## 【課題4】 二次構造予測

1CSPの二次構造予測を行い、結果をPowerPointにまとめよ

1. コールドショックプロテイン(PDB ID: 1CSP)をダウンロードし、配列をコピー
2. NPS@サイト([URL] [http://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_seccons.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_seccons.html))に貼り付け、コンセンサス二次構造予測をする
3. PowerPointで、予測結果を答えと比較し、正答率等を含めて記述する



## <課題の提出>

- 上記、【課題3】、（時間に余裕がある方は【課題4】）をPowerPointで2ページまでにまとめる
- PowerPointファイルを添付し、E-mailで以下のメールアドレスへ送信する
  - E-mail address: [tadaomi@iu.a.u-tokyo.ac.jp](mailto:tadaomi@iu.a.u-tokyo.ac.jp)