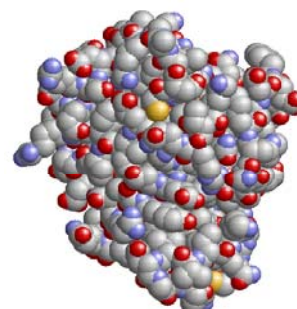


平成22年度・近畿大学・農学部・生命情報学

生体高分子の立体構造とその予測

Ras

```
MTEYKLVVVGAGGVGKSAL
TIQLIQNHVFVDEYDPTIED
SYRKQVVIDGETCLLDILD
TAGQEEYSAMRDQYMRIGE
GFLLCVFAINNTKSFEDIHQ
YREQIKRVKDSDDVPMVLV
GNKCDLAARTVESRQAQDL
ARSYGIPYIETSAKTRQGV
EDAFYTLVREIRQH
```



2010年5月25日(火)

奈良先端大・情報・蛋白質機能予測学講座

川端 猛

takawaba@is.naist.jp

<http://isw3.naist.jp/IS/Kawabata-lab/lec-ja.html>

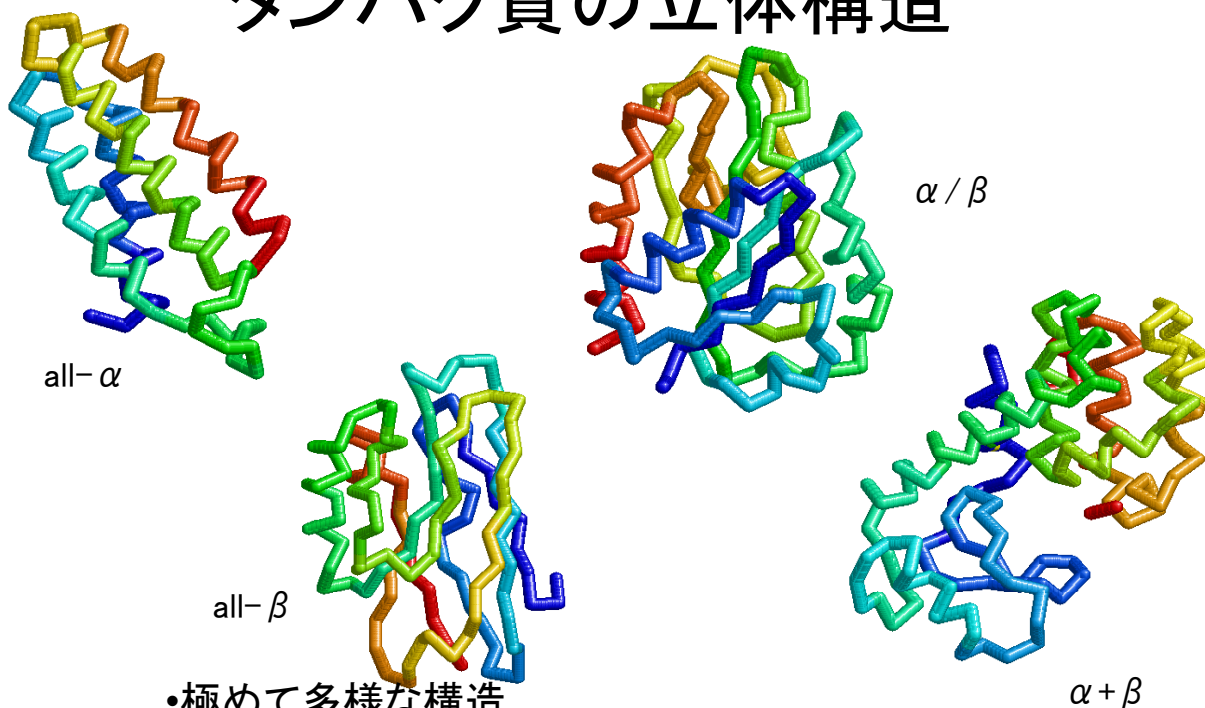
平成22年度「生命情報学 & 生命情報学実習」講義日程

	講義	生命情報学	演習	生命情報学演習	2010.3.25
4/13	川端1	分子生物学の基礎と配列データベース			
4/20	川端2	ペアワイズアライメントと配列相同性検索	川端	主要WEBデータベースと配列相同性検索	
4/27	川端3	マルチプルアライメントとその応用			
5/11	川端4	分子系統学基礎	川端	マルチプルアライメントと系統樹作成演習	
5/18	川端5	蛋白質の物理化学的性質とアミノ酸配列解析			
5/25	川端6	蛋白質立体構造データの情報解析	川端	蛋白質立体構造データの可視化	
6/1	川端	>>試験(川端 担当分)<<			
6/8	中村1	化学構造データと計算化学基礎I			
6/15	中村2	化学構造データと計算化学基礎II	中村	ChemOfficeを用いた計算化学演習	
6/22	中村	>>試験(中村担当分)<<			
6/29	金谷1	トランスクリプトーム解析			
7/6	金谷2	インタラクローム解析	金谷	発現プロフィール解析演習	
7/13	金谷3	メタボローム解析	金谷	インタラクローム・代謝物解析演習	
7/20	金谷	>>試験(金谷担当分)<<			

来週6/1(火)の川端分の試験について

- 筆記用具以外持ち込み不可
- 講義のスライドのPDFファイルは、講義 <http://isw3.naist.jp/IS/Kawabata-lab/lec-ja.html> のサイトにあるので、必要ならダウンロードして復習しておくこと。

タンパク質の立体構造

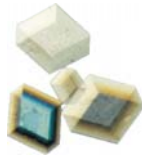


- 極めて多様な構造
- 構造はそのアミノ酸配列によって決定される
- 立体構造の多様性はその分子機能の多様性と関係

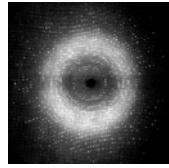
立体構造の決定法

X線結晶解析

大量発現
精製
結晶化

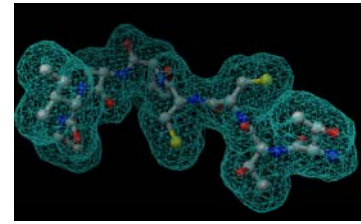


X線回折強度の測定



3次元
電子密度マップ

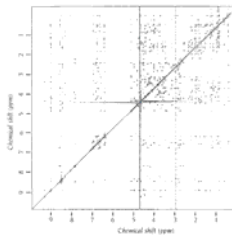
原子モデルの構築
原子モデルの精密化



NMR(核磁気共鳴法)

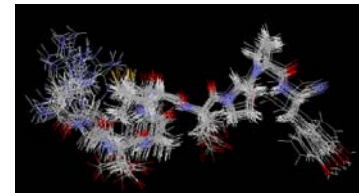
大量発現
精製

核磁気共鳴解析
(NOE解析, 帰属と距離拘束の抽出)



原子間
距離拘束

原子モデルの構築
原子モデルの精密化



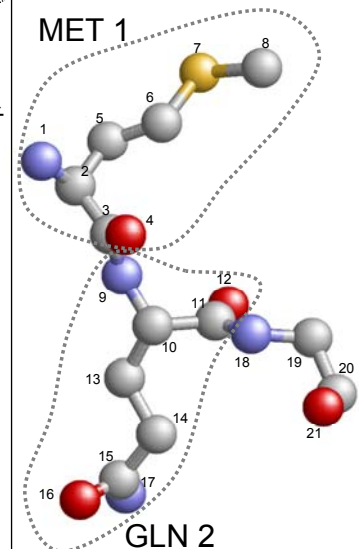
- (1)多数分子の平均構造の観察(精製が重要。大きな分子、複合体はより難しくなる)
- (2)発現・精製・結晶化のステップで、各タンパク質によって実験条件の調整が必須
- (3)実験データの収集・原子モデル構築の段階で計算機の支援が不可欠

PDBフォーマット

RCSB PDB <http://www.rcsb.org/pdb/>
MSD EBI <http://www.ebi.ac.uk/msd>
PDBj <http://www.pdbj.org>

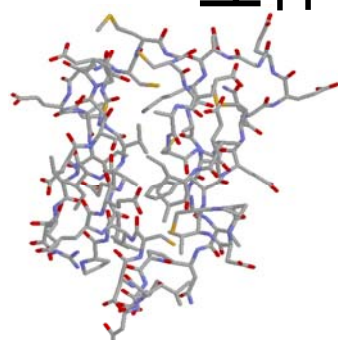
wwPDB(World Wide Protein Data Bank): 立体構造データの収集するデータベース

HEADER		UBIQUITIN		17-APR-92		1AAR			
COMPND		DI-UBIQUITIN							
SOURCE		BOVINE (BOS TAURUS) ERYTHROCYTES							
AUTHOR		W. J. COOK, L. C. JEFFREY, M. CARSON, Z. CHEN, C. M. PICKART							
原子番号	残基名	鎖識別子	残基番号	X座標	Y座標	Z座標	占有率	温度因子	
ATOM	1	N	MET A	1	15.493	30.088	14.694	1.00	8.36
ATOM	2	CA	MET A	1	14.600	29.031	15.110	1.00	8.15
ATOM	3	C	MET A	1	15.476	27.793	15.419	1.00	9.30
ATOM	4	O	MET A	1	16.571	27.561	14.871	1.00	8.96
ATOM	5	CB	MET A	1	13.500	28.837	14.105	1.00	9.89
ATOM	6	CG	MET A	1	13.823	27.997	12.931	1.00	10.21
ATOM	7	SD	MET A	1	12.312	27.711	11.891	1.00	10.33
ATOM	8	CE	MET A	1	13.174	26.595	10.726	1.00	7.30
ATOM	9	N	GLN A	2	14.968	27.014	16.326	1.00	9.75
ATOM	10	CA	GLN A	2	15.552	25.806	16.852	1.00	11.92
ATOM	11	C	GLN A	2	15.000	24.553	16.168	1.00	11.71
ATOM	12	O	GLN A	2	13.787	24.387	16.086	1.00	10.00
ATOM	13	CB	GLN A	2	15.368	25.715	18.386	1.00	12.39
ATOM	14	CG	GLN A	2	15.858	24.413	19.051	1.00	14.00
ATOM	15	CD	GLN A	2	15.676	24.587	20.561	1.00	15.59
ATOM	16	OE1	GLN A	2	16.525	25.209	21.205	1.00	18.67
ATOM	17	NE2	GLN A	2	14.564	24.203	21.152	1.00	15.39
ATOM	18	N	ILE A	3	15.960	23.745	15.687	1.00	11.79
ATOM	19	CA	ILE A	3	15.593	22.437	15.130	1.00	12.14
ATOM	20	C	ILE A	3	16.491	21.342	15.720	1.00	11.99
ATOM	21	O	ILE A	3	17.701	21.540	15.767	1.00	12.91

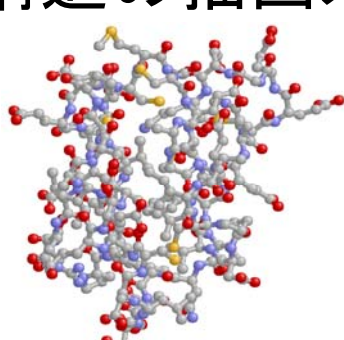


PDBは1971年に設立

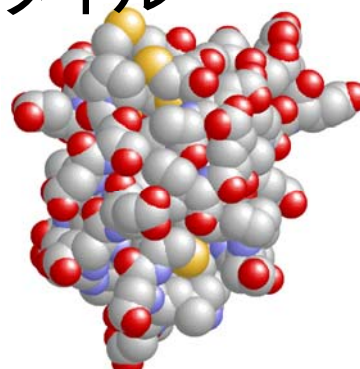
立体構造の描画スタイル



ワイアフレーム



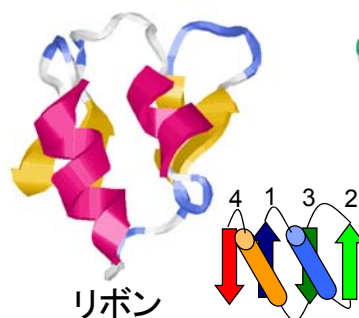
ボール&スティック



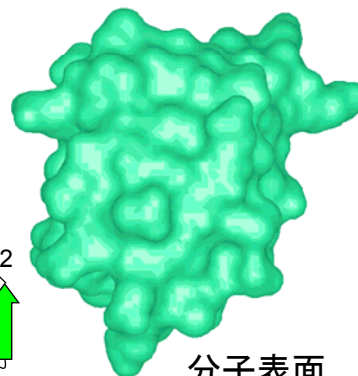
空間充填



バックボーン



リボン

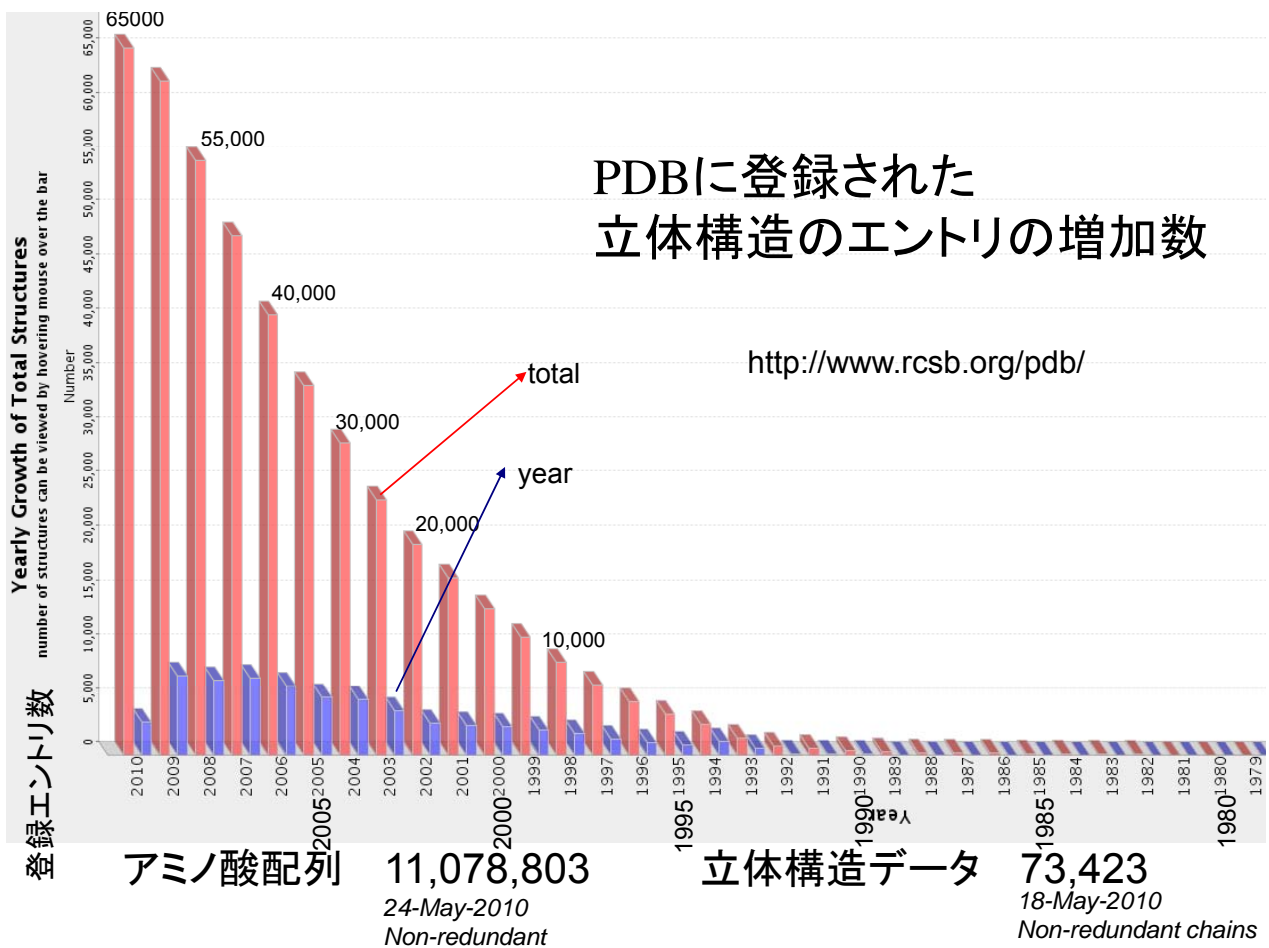


分子表面

PDBコード: 1fxd(ferredoxin II), Rasmol / Chimeプラグインで描画

立体構造描画のフリーウェア・シェアウェア

ソフト名	WebSite	対応機種	特徴
RasMol	www.openrasmol.org	Win Mac Linux (ソース公開)	独自の高速描画アルゴリズムを用い、非力なハードでも同様に高速描画が可能。コマンド言語は強力で多機能。基本的な描画法をカバーするが、分子表面の描画はできない。描画の質はやや低い。
Jmol	jmol.sourceforge.net	Win Mac Linux (ソース公開)	Java Applet によりWEBブラウザ内で動作。Chimeと同様に、WEBページに分子描画機能を加えることが可能。Javaを実行できる環境(JRE)があれば、特別なソフトウェアのインストールを必要としない。
PyMOL	pymol.sourceforge.net	Win Mac Linux (ソース公開)	OpenGLを用いた質の高い描画。Python言語で開発されており、Python言語を用いた機能拡張も可能。有料版と一部の機能が制限された無料版がある。
CHIMERA	www.cgl.ucsf.edu/chimera/	Win Mac Linux (ソース非公開)	OpenGLを用いた質の高い描画。電顕画像などの多様なデータの取り込み、ドッキング、静電計算などのインタフェース有り。Python言語を用いた機能拡張。
VMD	www.ks.uiuc.edu/Research/vmd/	Win Mac Linux (ソース非公開)	OpenGLを用いた質の高い描画。Tcl言語を用いた機能拡張が可能。分子動力学の可視化を主要目的として幅広い機能。



立体構造データからわかること

1. 機能(主に結合)に関する理解が深まる

- 原子レベルの結合・反応メカニズムの理解
- 機能に重要なアミノ酸の配置 (配列上散らばっていても、空間的には近接している)
- 立体構造データをもとにした、結合分子の推定・設計

2. 立体構造の類似から、遠い相同性の認識・大きなファミリーの構成が可能

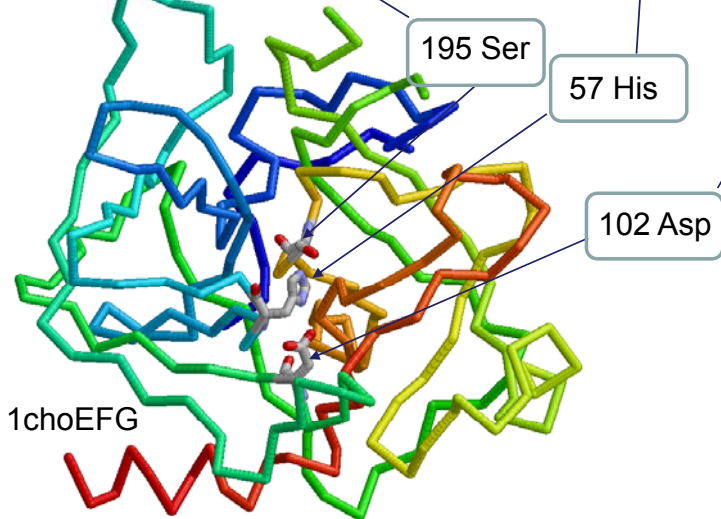
- 立体構造は配列に比べて進化的に保存しやすい

キモトリプシンの活性部位

>CTRA_BOVIN "Chymotrypsinogen A (EC 3.4.21.1)

```

C G V P A I Q P V L S G L S R I V N G E E A V P G S W P W Q V S L Q D K T
G F H F C G G S L I N E N W V V T A A H C G V T T S D V V V A G E F D Q G
S S S E K I Q K L K I A K V F K N S K Y N S L T I N N D I T L L K L S T A
A S F S Q T V S A V C L P S A S D D F A A G T T C V T T G W G L T R Y T N
A N T P D R L Q Q A S L P L L S N T N C K K Y W G T K I K D A M I C A G A
S G V S S C M G D S G G P L V C K K N G A W T L V G I V S W G S S T C S T
S T P G V Y A R V T A L V N W V Q Q T L A A N
    
```



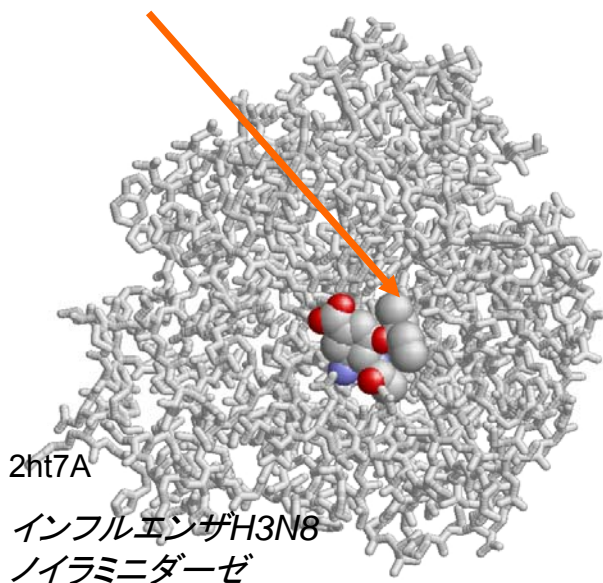
キモトリプシン(Chymotrypsin) :
他のタンパク質を加水分解する
酵素

三つのアミノ酸
(57His, 102Asp, 195Ser)が
反応に必須である(catalytic triad)。

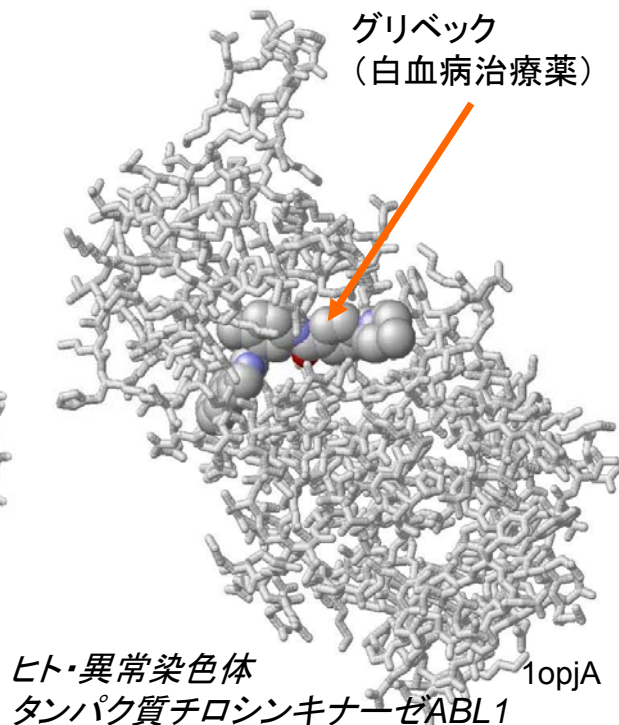
それらは、配列上では散在している
が、立体構造上は近接しており、大き
なくぼみの奥に位置する。

タンパク質に選択的に結合し機能を阻害する分子

タミフル
(インフルエンザ治療薬)



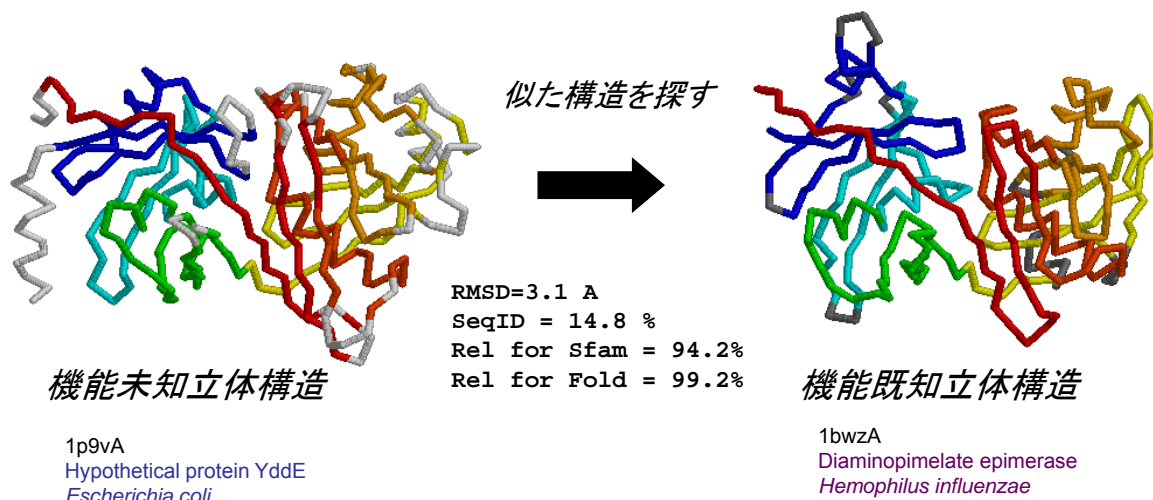
グリベック
(白血病治療薬)



標的タンパク質の結合部位の立体構造をもとに、そこに選択的に結合する分子を設計(ドラッグ
デザイン)

「形」の比較による弱い相同性認識

機能未知の立体構造データから機能を予測するには？



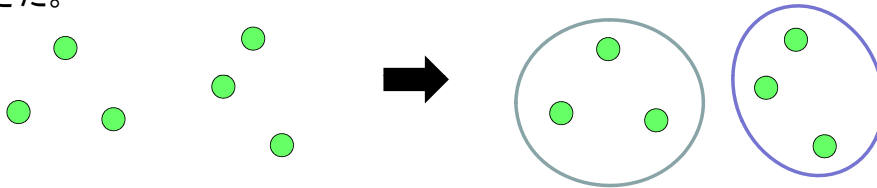
論理の流れ

1. 立体構造が似ているなら相同(ホモログス)
2. 相同(ホモログス)なら分子機能も似ているはず
3. 構造類似性は機能の類似性を意味する

配列と立体構造による タンパク質の比較分類

配列比較によるタンパク質分類

1970年代から、配列の一致度(sequence identity)に基づいて、グループを作る作業が行われてきた。



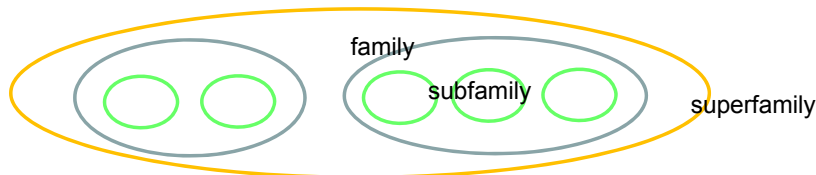
どのくらいの類似性で一つにまとめるかは様々な考えがあるが、少なくとも「**相同なタンパク質群**」を一つのグループにまとめるべき

ファミリー(family, 族):

: 進化的に関係のある(相同な)タンパク質グループのこと

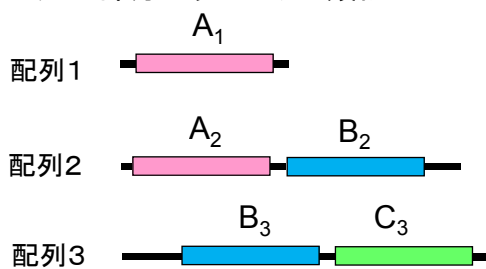
複数のファミリーがあつまった大きなグループ: **スーパーファミリー(superfamily, clan)**

一つのファミリーの中の小さなグループ: **サブファミリー(subfamily)**

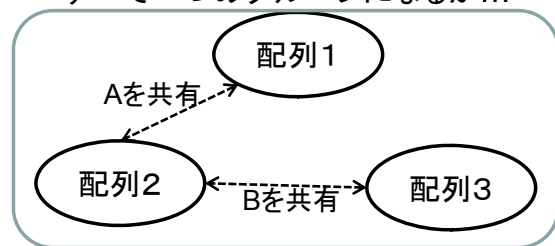


ドメイン単位による分類

配列の部分どうしだけが類似していた場合

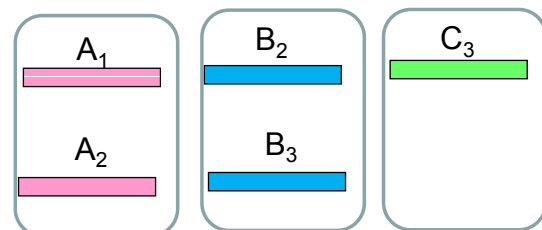
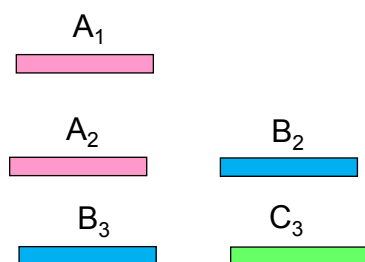


部分一致を重視すると、
すべて一つのグループになるが...



1と2が類似、2と3が類似していても
1と3は類似していない!

↓
ドメインに分割すると



きれいに分類することができる!!

Pfam : 蛋白質ファミリーのデータベース

<http://pfam.sanger.ac.uk>

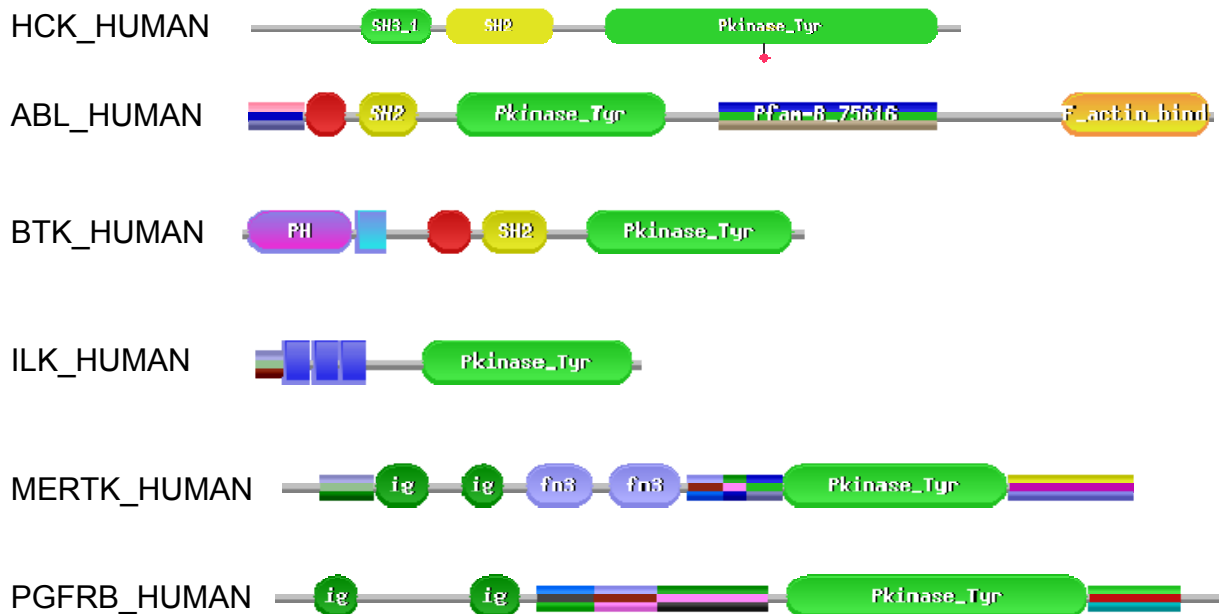
各蛋白質ファミリーのマルチプルアライメント、HMMなどを集めたデータベース

The image shows three screenshots of the Pfam website. The top-left screenshot displays the 'Browse Pfam families' page with a table of protein families. The top-right screenshot shows a sequence alignment view for a specific family. The bottom screenshot shows the detailed 'Summary' page for the 'Zinc finger, C2H2 type' (PF00096).

ID	Accession	Type	Number of Sequences	Average length	Average %id	Average coverage	Has 3D	Change status	Description
GP120	PF00516	Family	24 75195	175.5	54	87.21	✓	Changed	Envelope glycoprotein GP120
RVT_1	PF00078	Family	156 71535	165.3	67	39.98	✓	Changed	Reverse transcriptase (RNA-dependent DNA polymerase)
ABC_tran	PF00005	Domain	64 65707	184.2	26	37.89	✓	Changed	ABC transporter
Ank	PF00023	Repeat	1163 64674	30.6					
COX1	PF00115	Family	23 59394	232.7					
RVP	PF00072	Domain	50 54135	94.0					
LSR_1	PF00560	Repeat	2445 53686	22.7					
zf-C2H2	PF00096	Domain	196 52611	23.4					
Cytochrom_B_N	PF00033	Domain	8 49376	154.2					
WD40	PF00400	Repeat	1863 45685	38.7					
TPR_1	PF00515	Repeat	562 37697	32.2					
Oxidored_o1	PF00361	Family	33 36594	215.3					

Pkinase_Tyrドメインをもつタンパク質の例

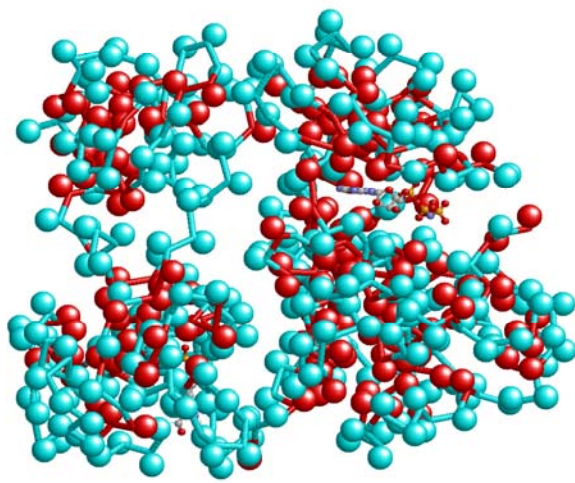
Family : Pkinase_Tyr (PF07714) : Protein tyrosine kinase



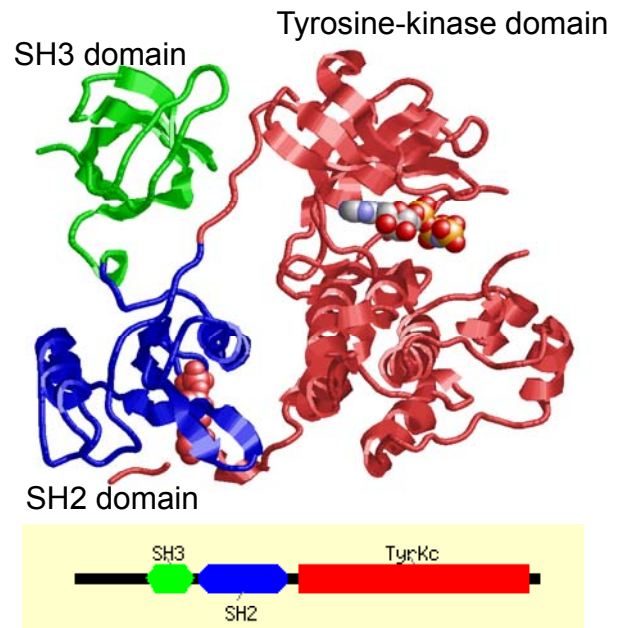
Pfamデータベース (<http://pfam.sanger.ac.uk/Software/Pfam/>) からの引用

ドメイン構造と疎水性の関係

Tyrosine protein kinase HCK (1ad5A, HCK_HUMAN)



3つの分断された疎水性コア
→ ドメイン構造



大きなタンパク質は、いくつかの構造上のまとまり(ドメイン)を持つことが多い

立体構造によるタンパク質の分類

- 配列による分類の延長。ドメイン単位の分類、進化的に関係のある(相同な)タンパク質を群にする(ファミリー)という考え方は同じ。
- 立体構造の類似性から、より遠縁の相同性を見つけ、より大きな群を作る(スーパーファミリー)
- スーパーファミリーをまとめたさらに大きなグループ(フォールド、クラス)
- 主にアミノ酸(Cαレベル)のラフな構造の類似性に着目
- 分類対象は主に球状タンパク質

立体構造分類データベースSCOP

<http://scop.mrc-lmb.cam.ac.uk/scop/>

・4階層で分類

Class(クラス)

<アナロジー>

Fold(フォールド)

Superfamily(スーパーファミリー)

Family(ファミリー)

<ホモロジー>

どの階層に分類するかは、
配列の類似性
立体構造の類似性
分子機能の類似性
機能部位の類似性
を専門家が総合的に判断して決める

SCOP: Protein: Flavodoxin from *Desulfovibrio vulgaris* - Microsoft Inte...

ファイル(F) 編集(E) 表示(V) お気に入り(I) ツール(T) ヘルプ(H)

アドレス(A) http://scop.mrc-lmb.cam.ac.uk/scop/data/scop_b.d.c.e.f.b.c.html

Protein: Flavodoxin from *Desulfovibrio vulgaris*

Lineage:

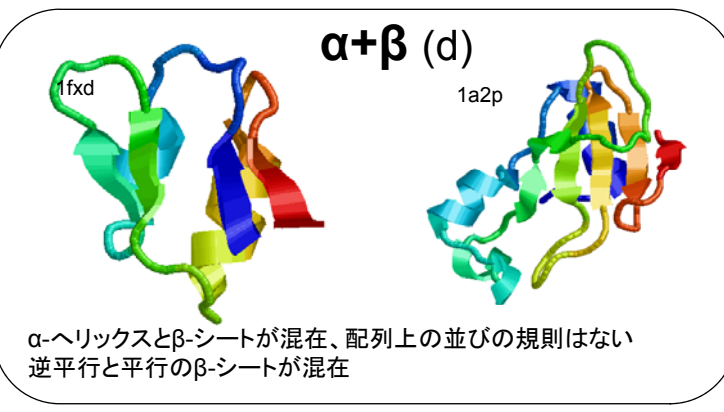
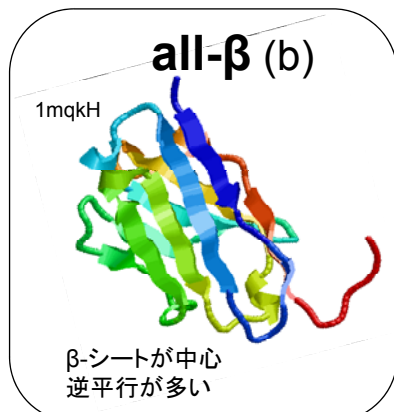
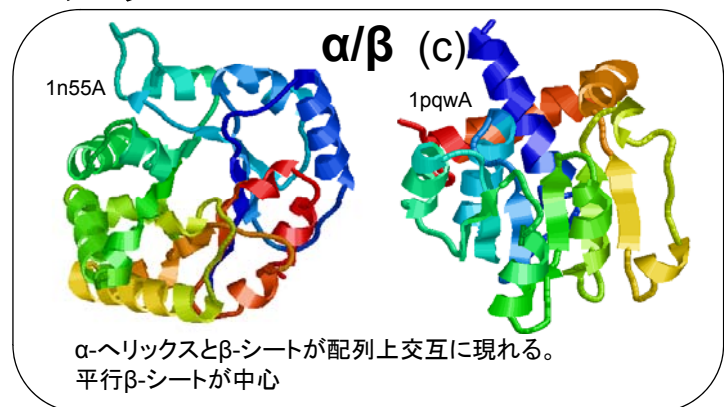
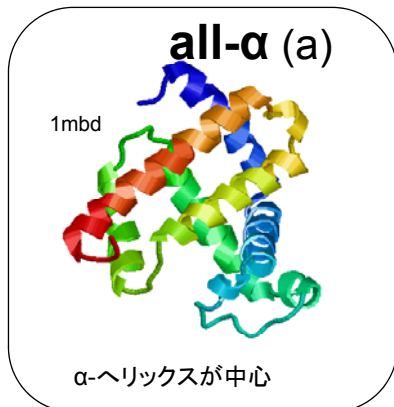
1. Root: [scop](#)
2. Class: [Alpha and beta proteins \(a/b\)](#) [51349]
Mainly parallel beta sheets (beta-alpha-beta units)
3. Fold: [Flavodoxin-like](#) [52171]
3 layers, a/b/a; parallel beta-sheet of 5 strand, order 21345
4. Superfamily: [Flavoproteins](#) [52218]
5. Family: [Flavodoxin-related](#) [52219]
binds FMN
6. Protein: Flavodoxin [52220]
7. Species: [Desulfovibrio vulgaris](#) [52222]

PDB Entry Domains:

1. [1f4p](#) [59651] [3D] [4D]
complexed with fmn; mutant
1. [cham a](#) [59651] [3D] [4D]
2. [1j8q](#) [62744] [3D] [4D]
complexed with fmn
1. [cham a](#) [62744] [3D] [4D]
3. [1j9e](#) [62756] [3D] [4D]
complexed with fmn; mutant
1. [cham a](#) [62756] [3D] [4D]
4. [1akr](#) [31152] [3D] [4D]
complexed with fmn; mutant

1akr

クラス

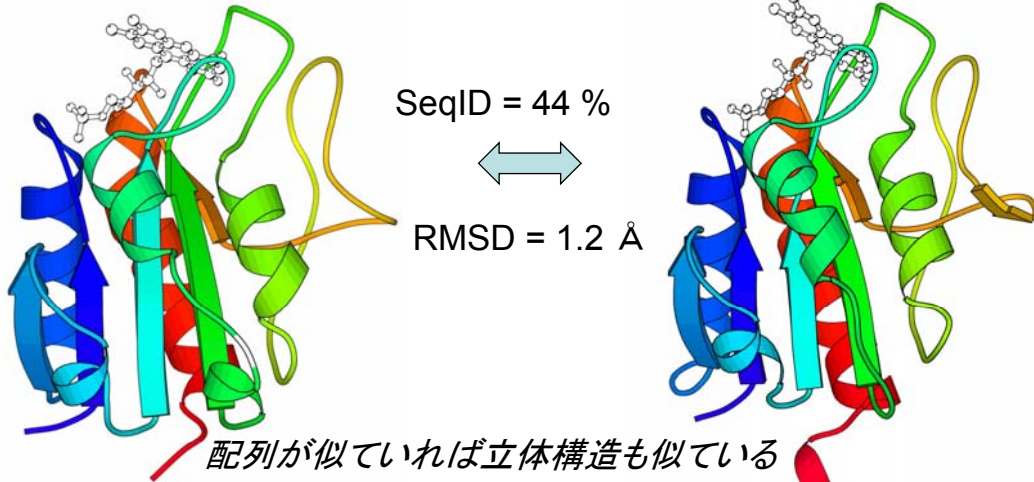


アミノ酸配列と立体構造の関係

ピロリ菌と大腸菌のフラボドキシン (同一残基率 SeqID 44%)

ピロリ菌(1fueA)

大腸菌(1ag9A)



1fueA 2: GKIGIFFGTDSGNAEAI AEKISKAIG--NAEVVDVAKASKEQFNGFTKVLVAPTAGAGD: 59

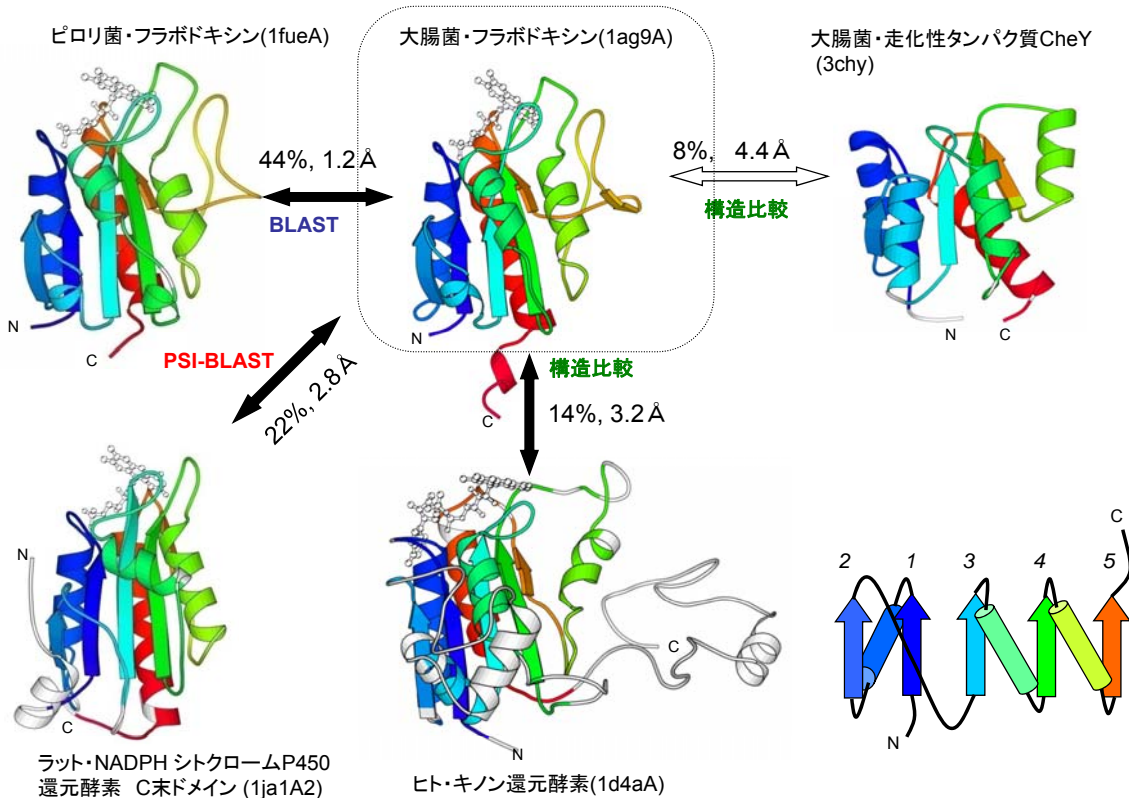
***** ** *

1ag9A 2: AITGIFFGSDTGNTENIAKMIQKQLGKDVADVHDI AKSSKEDLEAYDILLGLIPTWYYGE: 61

1fueA 60: LQTDWEDFLGTL EASD--FANKTIGLVGLGDQDTYSETFAEGIFHIYEKAK--AGKVVGQT: 116
 *

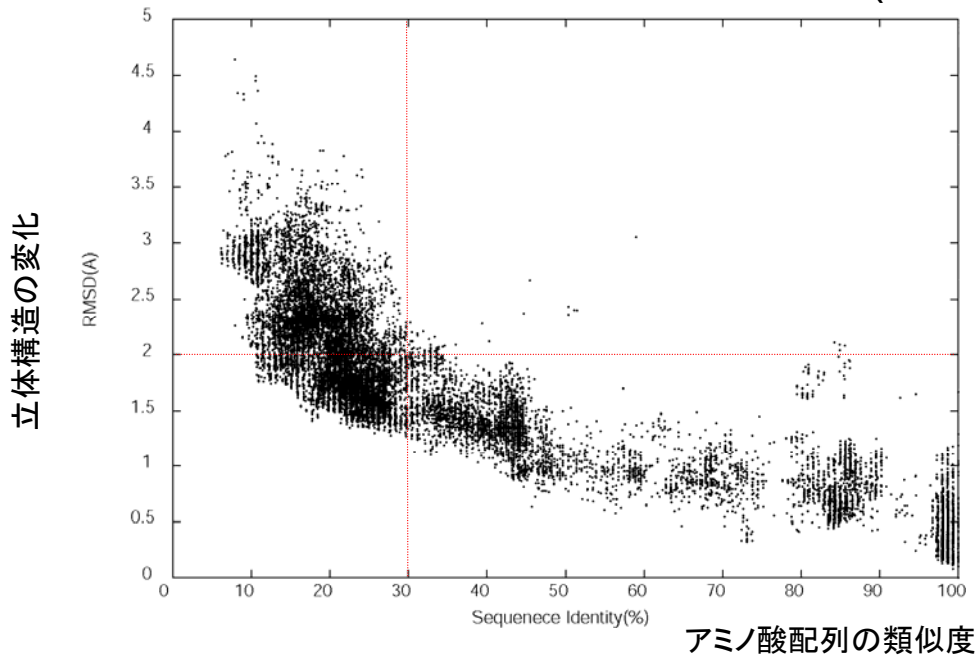
1ag9A 62: AQCDWDDFFPTLE-EIDFNGKLVALFGCGDQEDYAEYFCDALGTIRDII EPRGATIVGHW: 120

アミノ酸配列がほとんど似ていなくても立体構造は似ている場合がある



構造は配列より進化的に保存がよい → 構造比較から新たなホモログが発見できる可能性

アミノ酸配列の変化と立体構造の変化の相関(グロビン族)



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

なぜ、配列が変わっても立体構造 は変わりにくいのだろうか？

進化の過程でランダムに生じるアミノ酸変異を

- (1) 立体構造を変化させる変異
 - (2) 立体構造を変化させない変異
- の二つに分類。



タンパク質がその分子機能を発揮するためには立体構造を維持することが重要であるため、

(1)の変異が入ったタンパク質は機能を失い、
その変異が入った個体は淘汰される(子供を残せない)。

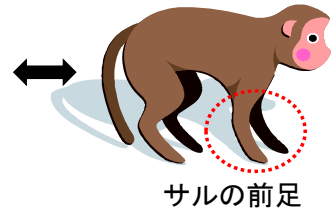
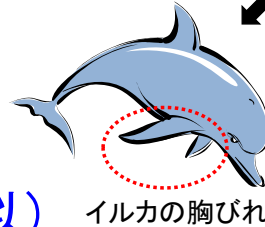
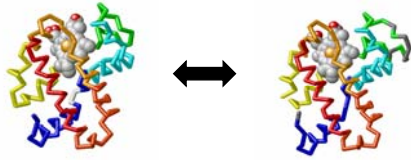


よって、(2)の変異だけが許されて生き残る。その結果、
配列が変化しても、立体構造は変わらないように見える。

2つの構造の類似性

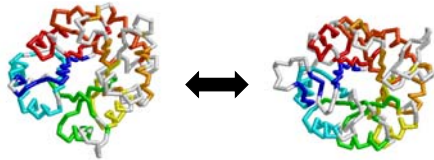
ホモロジー (homology, 相同):

進化的起源を共有することによる類似
多くの場合、分子機能なども類似している

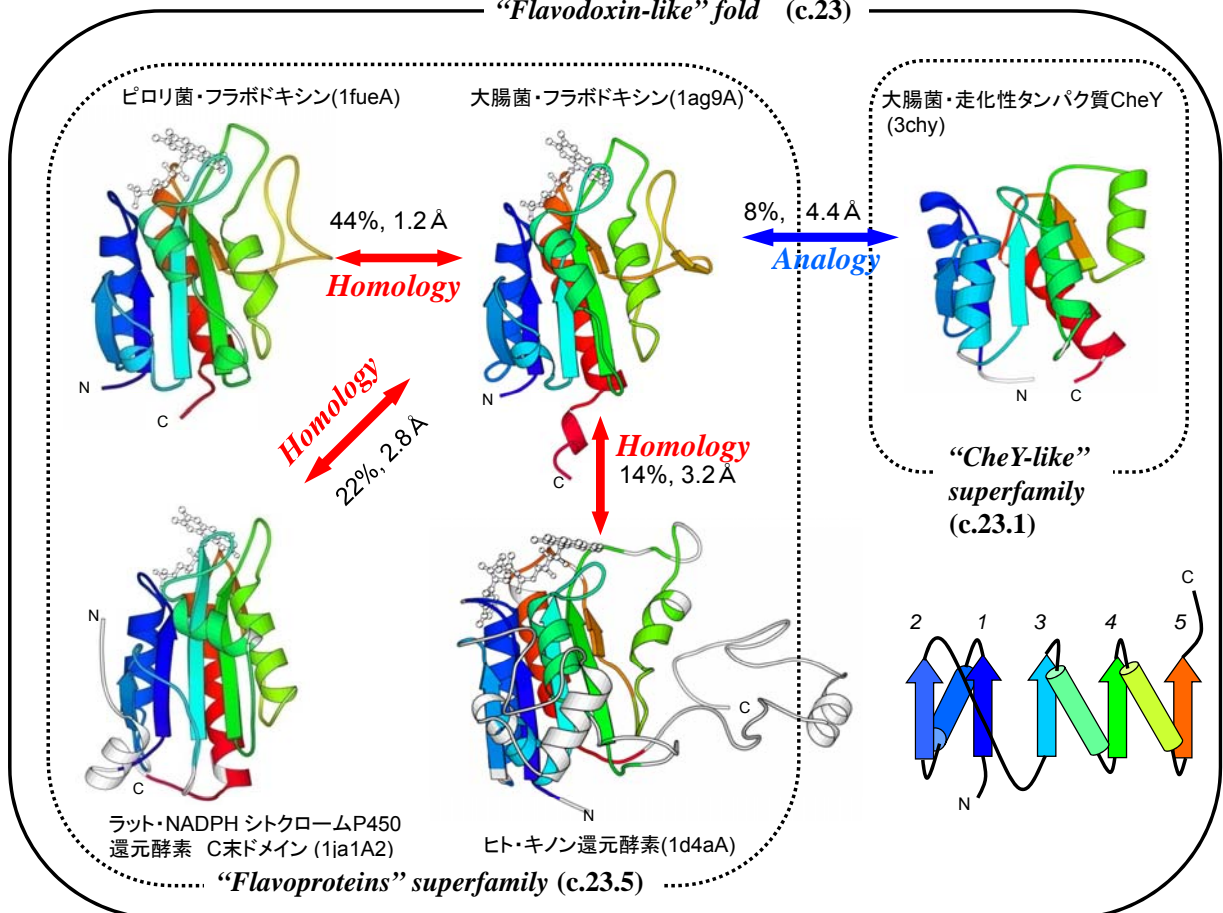


アナロジー (analogy, 相似)

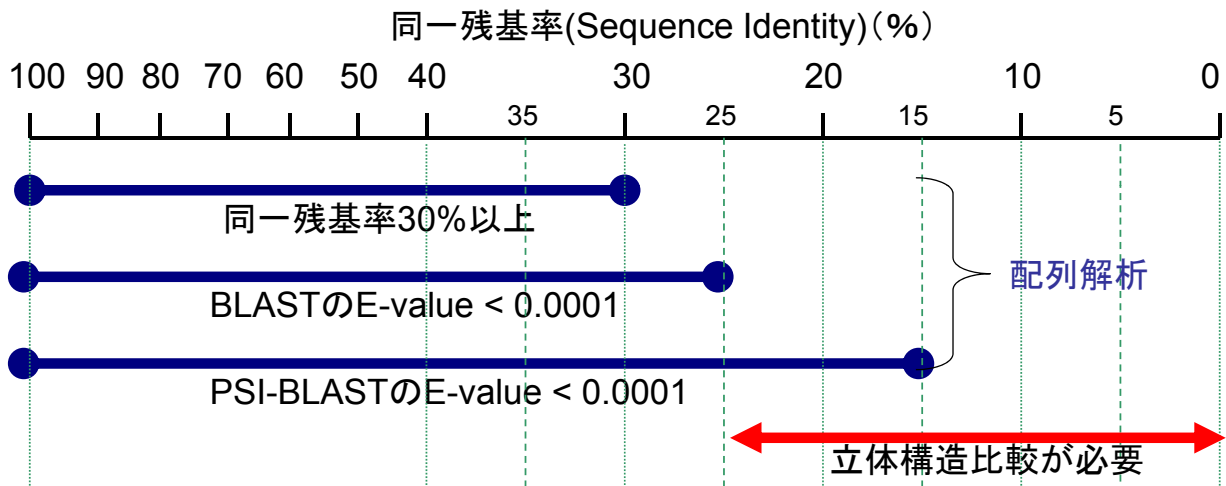
進化的起源とは無関係な類似
多くの場合、分子機能など他の属性は似ていない。
物理化学的な構造の偏好が原因とされる。



“Flavodoxin-like” fold (c.23)



タンパク質の相同性の判断基準



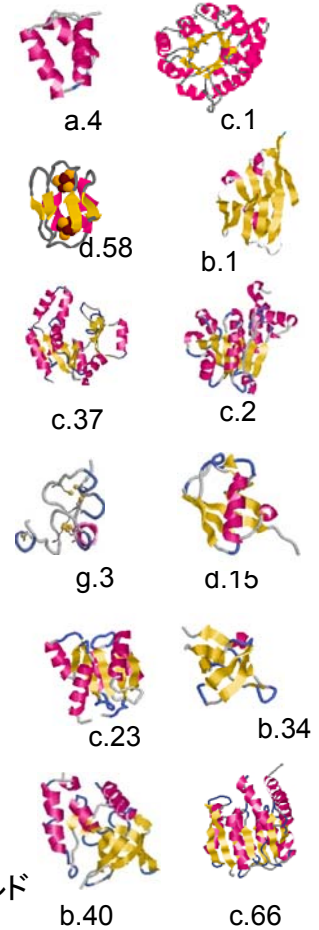
同一残基率が25%以下の場合の相同性の判断基準

- (1) 立体構造の類似性が高い(DALIのZスコア、MATRASのRdisスコア)
- (2) 同一残基率がそこそ高い(PSI-BLASTでヒットする、SeqID \geq 15%ぐらい)
- (3) 分子機能に類似性がある(補酵素、酵素反応、基質、代謝経路等の共通性等)
- (4) 基質・補酵素の結合部位が類似しており、そのアミノ酸が保存(モチーフ)
- (5) ドメイン構成の共通性
- (6) スーパーフォールドの場合は、特別な注意が必要

タンパク質の代表的な フォールドの紹介

多く現れるフォールド(出現ドメイン数順)

	フォールド名 (SCOP1.73による。クラスfhijklは除く)	ドメインの数	スーパーファミリーの数
a.4	DNA/RNA-binding 3-helical bundle	292	14
c.1	TIM beta/alpha-barrel	269	33
d.58	Ferredoxin-like	248	55
b.1	Immunoglobulin-like beta sandwich	246	27
c.37	P-loop nucleoside triphosphate hydrolase	235	1
c.2	NAD(P)-binding Rossmann-fold domains	168	1
g.3	Knottins	154	19
d.15	Beta-Grasp(ubiquitin-like)	131	13
c.23	Flavodoxin-like	118	15
b.34	SH3-like barrel	117	18
b.40	OB-fold	100	12
c.66	SAM-dependent methyltransferases	95	1

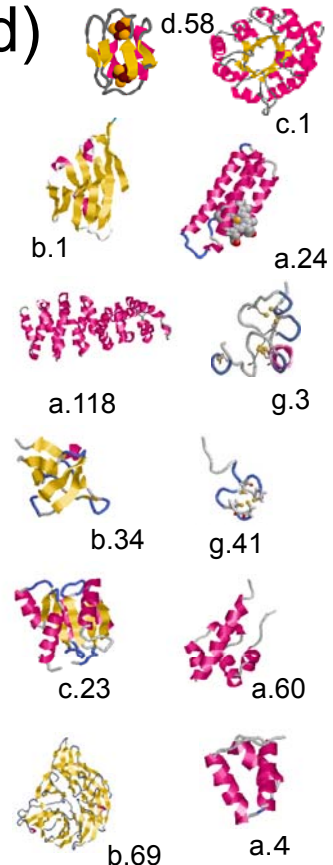


太字はスーパーフォールド: 多くのスーパーファミリーを含むフォールド

スーパー・フォールド(Superfold)

スーパーフォールド: 多くのスーパーファミリーを含むフォールド

	フォールド名 (SCOP1.73による。クラスfhijklは除く)	その下のスーパーファミリーの数
d.58	Ferredoxin-like	55
c.1	TIM beta/alpha-barrel	33
b.1	Immunoglobulin-like beta sandwich	27
a.24	Four-helical up-and-down bundle	27
a.118	Alpha-alpha superhelix	23
g.3	Knottins	19
b.34	SH3-like barrel	18
g.41	Rubredoxin-like	16
c.23	Flavodoxin-like	15
a.60	SAM(sterile alpha motif) domain-like	15
b.69	7-bladed beta-propeller	14
a.4	DNA/RNA-binding 3-helical bundle	14

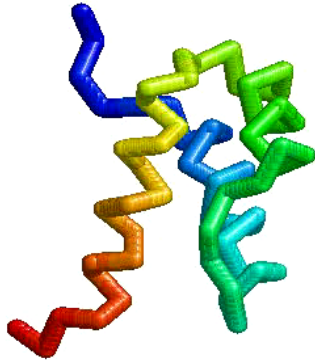


スーパーフォールドは「構造の類似と機能の類似が対応しにくいフォールド」ともいえる

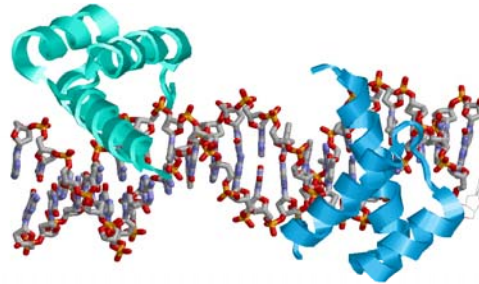
三本ヘリックスバンドル

DNA/RNA-binding 3-helical bundle(a.4)

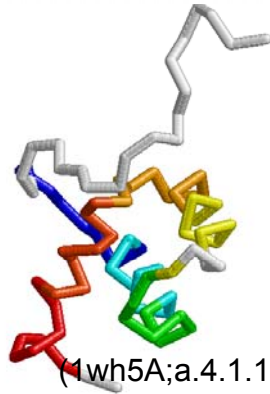
三本のヘリックスが束になった構造
DNAやRNAに結合する蛋白質が多い



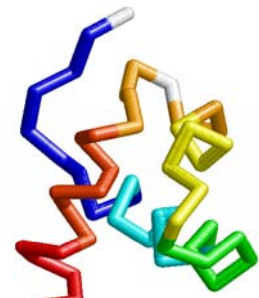
Mating type protein A2,
yeast(1k61A;a.4.1.1)



DNAと
結合した
構造(1k61)



(1wh5A;a.4.1.1)



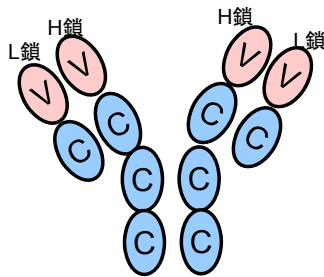
(1gvdA;a.4.1.3)

免疫グロブリン様

Immunoglobulin-like beta sandwich(b.1)



Immunoglobulin
Heavy chain variable
domain. 1mjuH1(b.1.1.1)



免疫グロブリン分子はH鎖とL鎖からなり、それぞれ、免疫グロブリン様ドメインからできている。

細胞表面あるいは分泌蛋白質が多い。免疫や分子認識に関する機能を担うことが多い。



Immunoglobulin
Heavy chain constant
domain 1mjuH2(b.1.1.2)

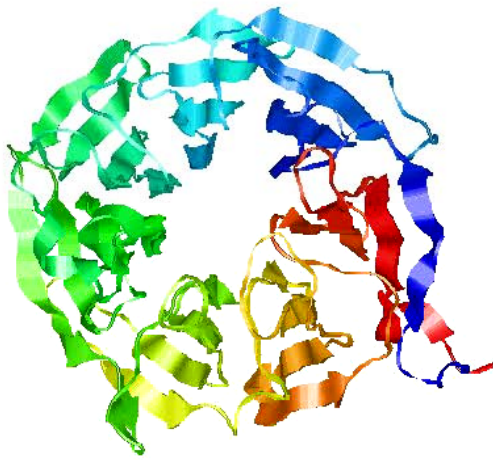


Macromomycin
1noaA (b.1.7.1)

4本のβストランドからなるβシートが二枚、サンドイッチ状に重なった構造をとる。逆平行のβシートが主。

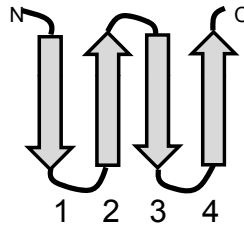
7枚羽 β プロペラ

7-bladed beta-propeller (b.69)

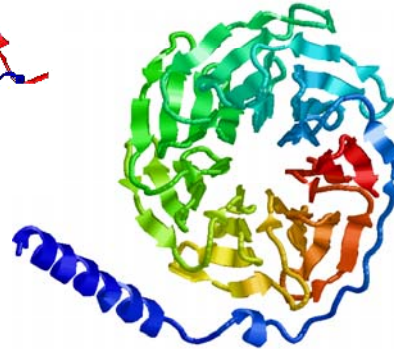


Galactose oxidase
domain3, 1k3iA3(b.69.1.1)

細胞膜付近の、シグナル伝達や糖結合に関わることが多い。



のユニットが7回、円環状に繰り返されたフォールド



Guanine nucleotide-binding
protein subunit beta-1
1gotB(b.69.4.1)

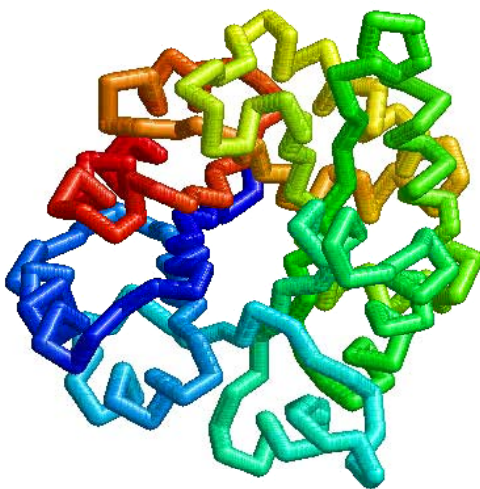


Influenza Neuraminidase
1f8dA(b.68.1.1)

6枚羽 β プロペラ

TIMバレル

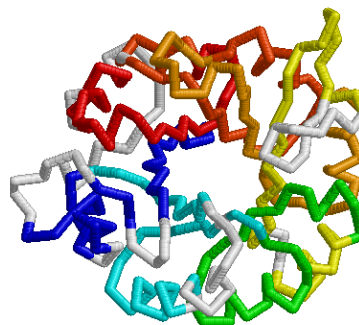
TIM beta/alpha barrel (c.1)



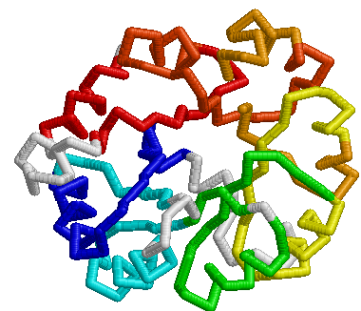
Triosephosphate isomerase
1n55A(c.1.1.1)

八本の β ストランドと α ヘリックスが交互に組み合わさり、対称的な樽(barrel)構造を形成する。中央の樽構造は平行の β シートからなる。

全般に解糖系の酵素が多い。基質、酵素反応は極めて多彩。



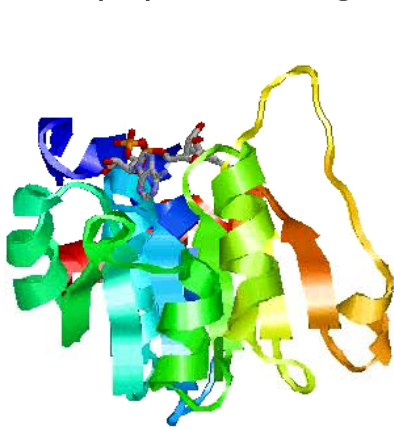
Imidazole glycerol
phosphate synthase
subunit hisF
1thfD(c.1.2.1)



KHG/KDPG aldolase
1euaA(c.1.10.1)

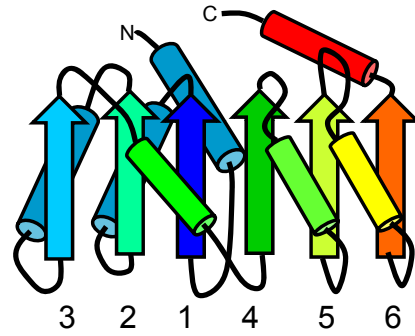
ロスマン フォールド

NAD(P)-binding Rossmann-fold domains (c.2)



Alcohol dehydrogenase gamma chain domain2, 1u3wA1(c.2.1.1)

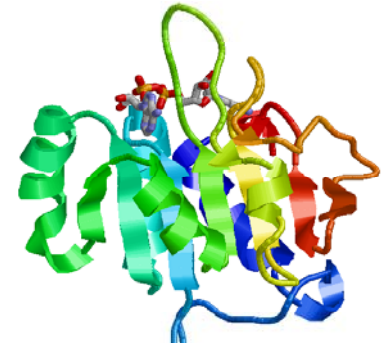
補酵素を利用したNAD(P)を脱水素酵素を持つ。基質を認識するドメインは別に存在する。



平行のβシートが主。



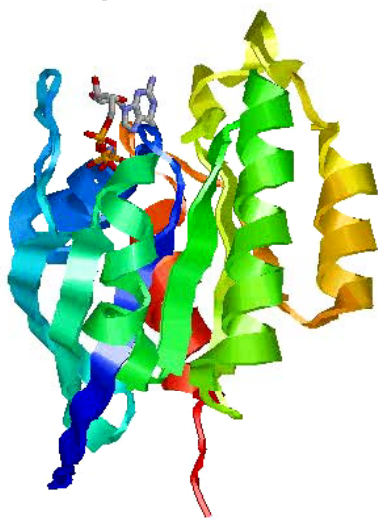
Aspartate beta-semialdehyde dehydrogenase 2gz1A1(c.2.1.3)



L-alanine dehydrogenase 1pjcA1(c.2.1.4)

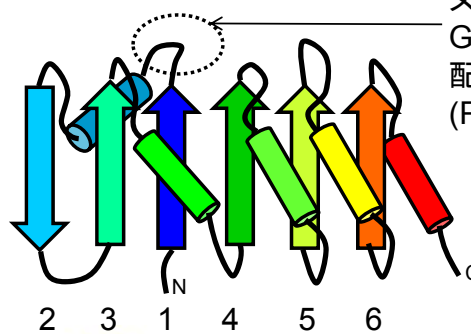
P-loop スーパーファミリー

P-loop nucleoside triphosphate hydrolase(c.37)



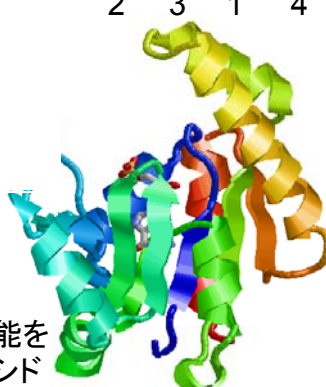
Ras-related protein RAB-5A 1r2qA(c.37.1.8)

シグナル伝達、DNA結合、モーター機能など、多彩な分子機能を担うが、ATP, GTPなどのヌクレオシドの加水分解を行う点は共通。



ヌクレオシド結合部。GxxxxGKTの配列モチーフ(P-loop)を持つ。

平行のβシートが主。トポロジーには様々な変異あり。

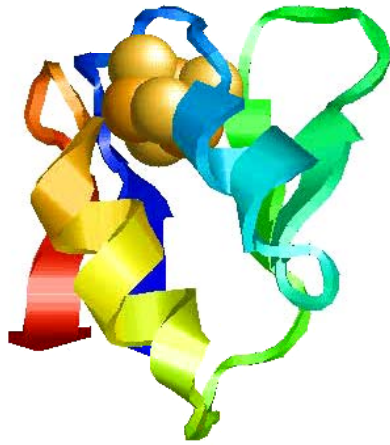


Guanylate kinase 1gkyA(c.37.1.1)

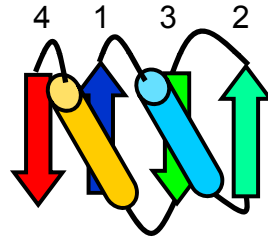


RecA protein.Domain1 1xmsA1(c.37.1.11)

フェレドキシン様 Ferredoxin-like (d.58)



4Fe-4S Ferredoxin(1fxd,d.58.1)

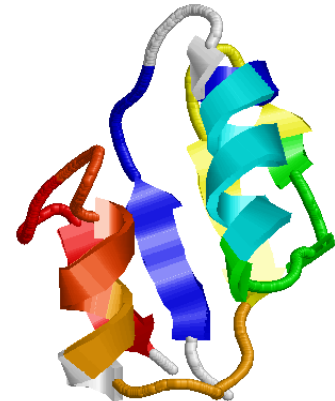


4本の逆平行のβストランドと
2本のαヘリックスからなる。

分子機能は極めて多彩。



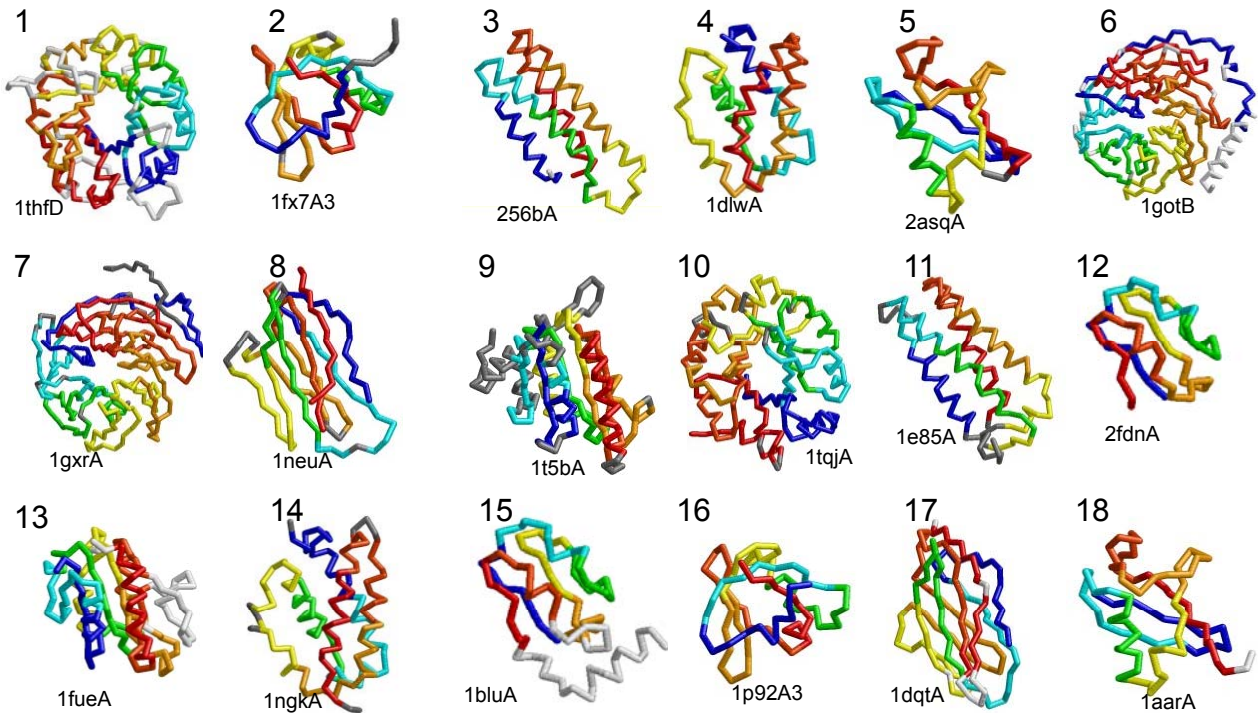
U1A RNA binding domain (1nu4A,d.58.7.1)



(1eayC,d.58.24)

蛋白質の立体構造比較: 学籍番号: _____

氏名: _____



上の蛋白質群の中から、以下のフォールドに属する蛋白質の番号を記せ。

TIMバレル: _____

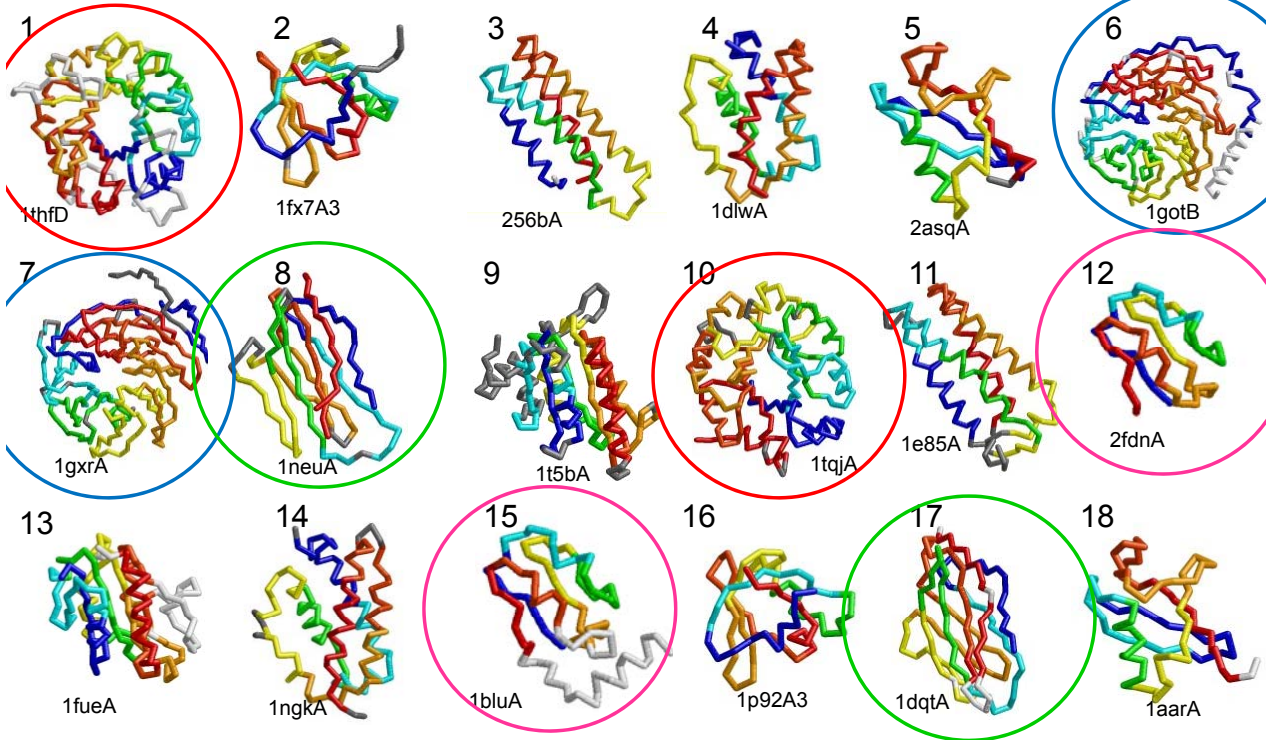
免疫グロブリン様: _____

βプロペラ: _____

フェレドキシン様: _____

蛋白質の立体構造比較: 学籍番号: _____

氏名: _____



上の蛋白質群の中から、以下のフォールドに属する蛋白質の番号を記せ。

TIMバレル: 1, 10

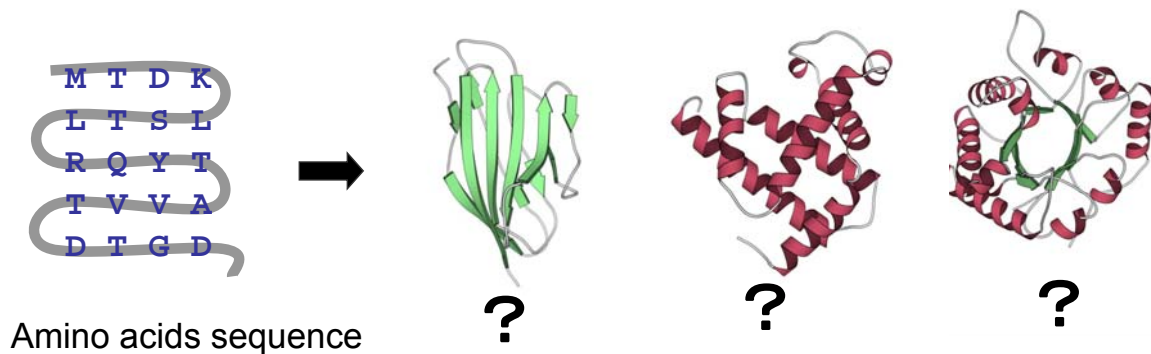
免疫グロブリン様: 8, 17

βプロペラ: 6, 7

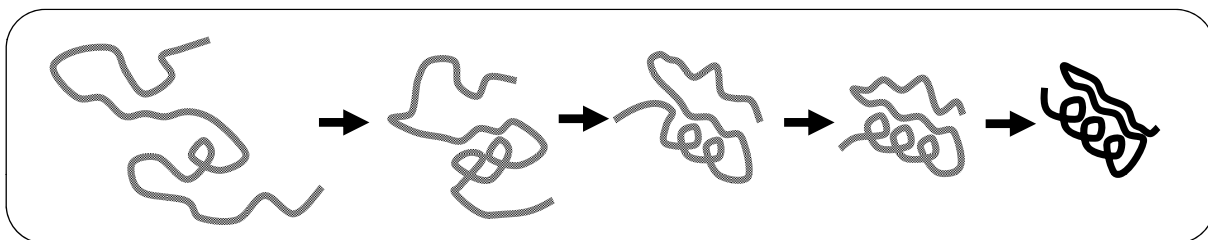
フェレドキシン様: 12, 15

H22 生命情報学
2010.5.25

タンパク質の立体構造予測



3次構造予測(1): *Ab initio* 予測



原子モデルとポテンシャルエネルギー関数を設定
より低いエネルギー値になるように構造を変形していく

分子動力学法、モンテカルロ法、
エネルギー最小化計算。

大きな計算量が必要。

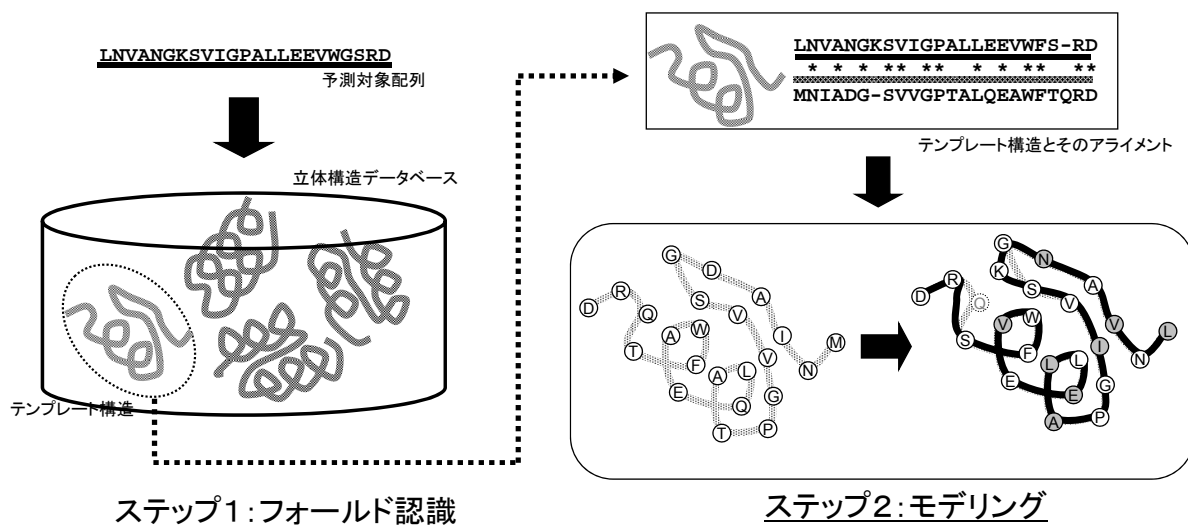
近年フラグメントアセンブリの手法が進展

Baker グループ (U. Washington) <http://depts.washington.edu/bakerpg/newindex.html>

高田グループ (神戸大学) <http://theory.chem.sci.kobe-u.ac.jp/indexj.html>

3次構造予測(2): 比較モデリング (ホモロジー・モデリング)

原理: 立体構造はアミノ酸配列より保存しやすい。



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

BLAST/FASTA, プロフィール法, ...

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

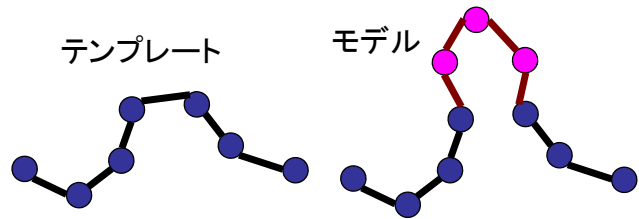
MODELLER, FAMS, ...

モデリング

テンプレート構造を元にした全原子の構築
(MODELLER, FAMS)

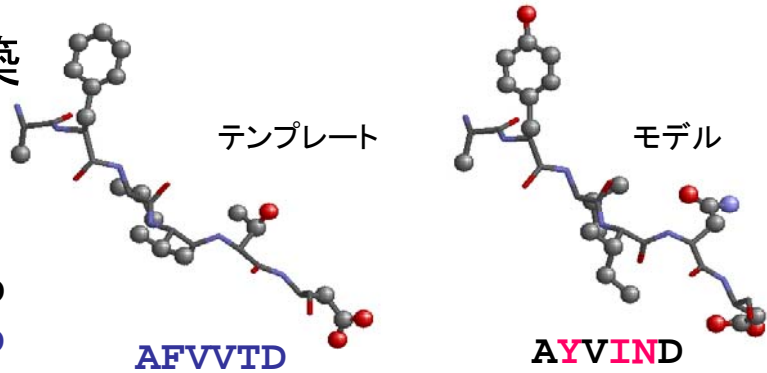
(1) ループの構築

Sequence ALIM**STK**GFVS
Structure LLLM---GFIT



(2) 側鎖原子の構築

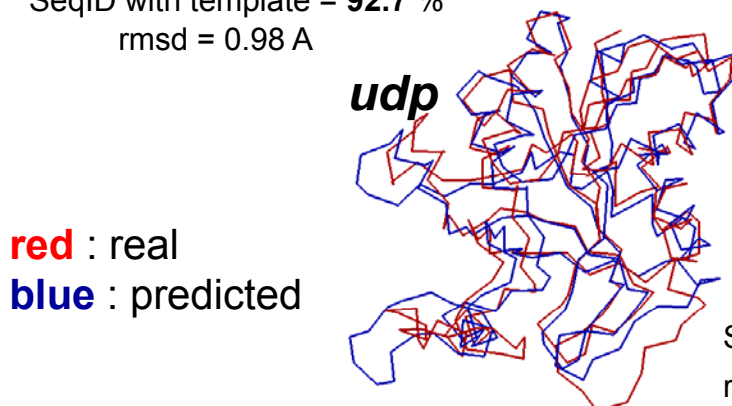
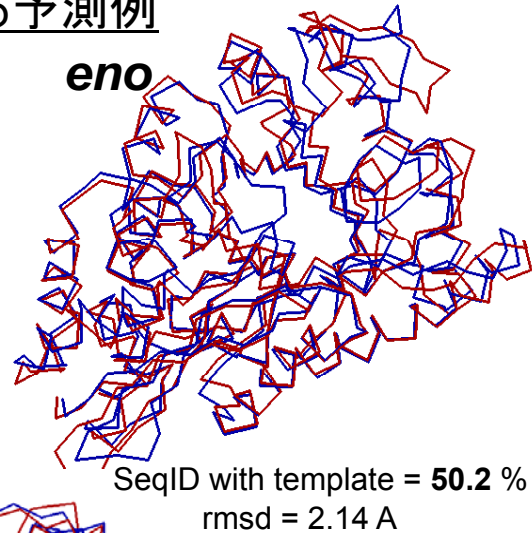
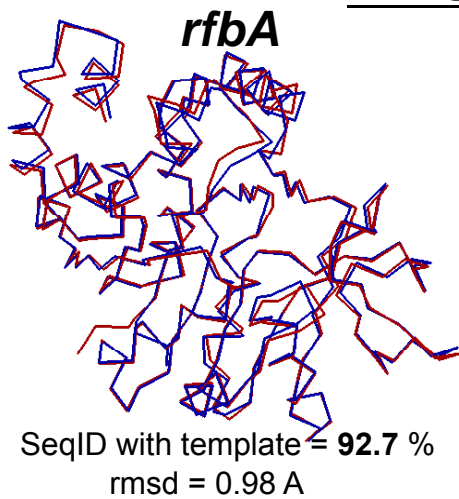
Sequence **AYVIND**
Structure AFV**VTD**



MODELLER : <http://www.salilab.org/modeller/modeller.html>

FAMS <http://www.pharm.kitasato-u.ac.jp/biomoleculardesign/>

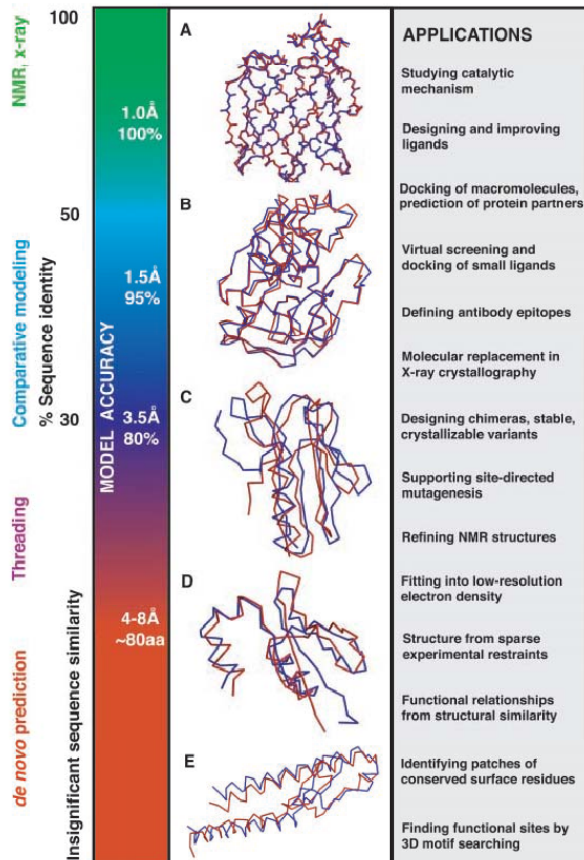
BLASTによる予測例



red : real
blue : predicted

モデリングした構造の精度と用途

SeqID = 100 %



反応メカニズムの理解

リガンドの設計

SeqID = 50 %

高分子のドッキング

低分子のドッキング

SeqID = 30 %

[分子置換法による精密化]

部位特異的置換のサポート

[NMRの精密化]

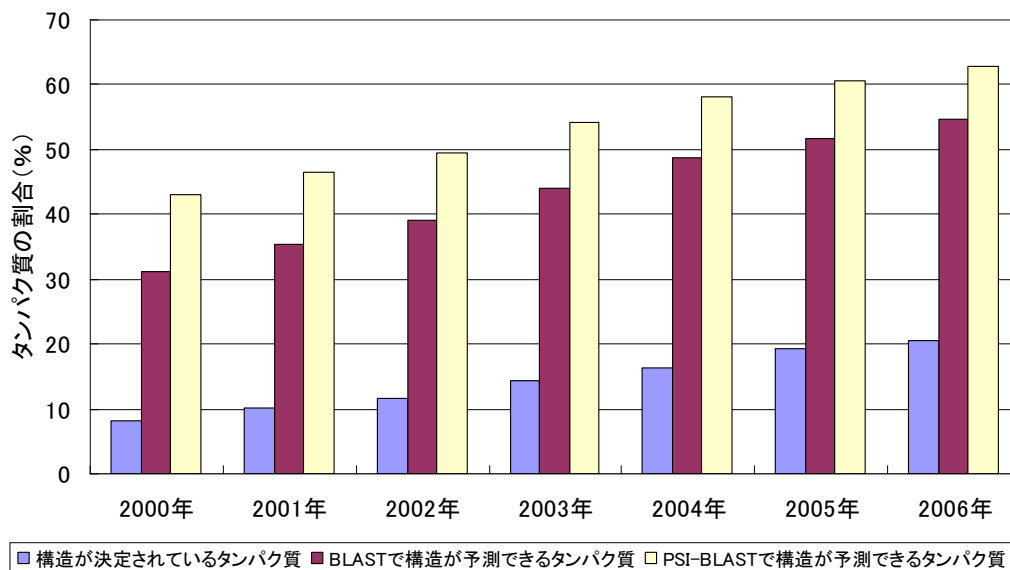
[電顕等の粗い電子密度へのフィット]

Ab initio

保存している表面残基の発見

D.Baker and A.Sali Science Vol 294 93-96

大腸菌のタンパク質のうち、その構造が決定されているタンパク質、構造が予測できるタンパク質の割合



大腸菌の4404個のタンパク質に対して、2000年から2006年までの各年の年末までに登録された立体構造データベースを使用した場合についてそれぞれ計算した。30アミノ酸以上の構造が予測されたタンパク質を、構造予測できるタンパク質とみなして割合を計算した。

