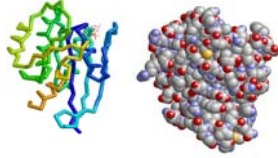


近畿大学・農学部・生命情報学  
**生体高分子の立体構造とその予測**

Ras

MTEYKLVVVGAGGVGKSAL  
 TIQLIQNHFVDEYDPTIED  
 SYRKQVVIDGETCLLDILD  
 TAGQEYSAMRDQYMRITGE  
 GFLLCFVAINNTKSFEDIHQ  
 YRQIKRVDSDVPMVFLV  
 GKNCLEAAKTVESRGAQLD  
 ARSYGIPYIEESAKTRQGV  
 EDAPYTLVREIRGH



2008年6月3日(火)

奈良先端大・情報・蛋白質機能予測学講座

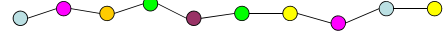
川端 猛

takawaba@is.naist.jp

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

**生体高分子の立体構造**

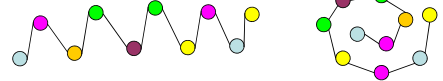
いくつかのユニットがひも状にー列に並んだ高分子(ポリマー)



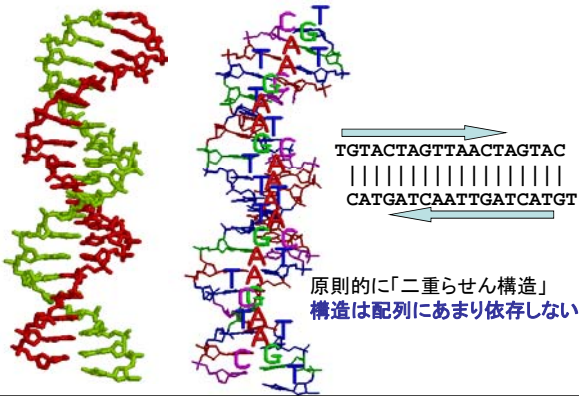
DNA、RNAは4種の塩基[AT(UG)]

タンパク質は20種類のアミノ酸[AVFPMILDEKRSYHCNQWG]

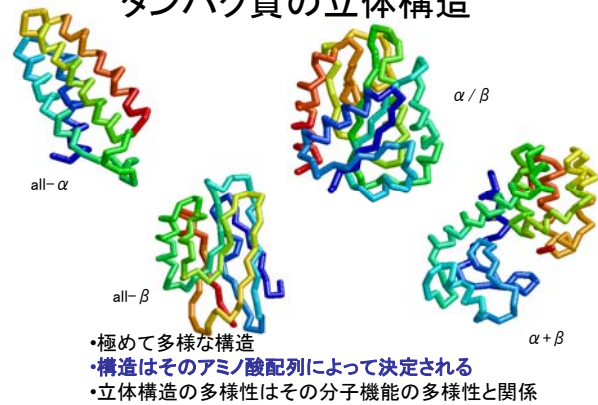
- ひもなので、とにかくフレキシブル
- 原理的にどんな形もとれる。
- ユニットの並びや環境によってどんな形を好むかが決まる



**核酸(DNA)の立体構造**



**タンパク質の立体構造**



**立体構造データからわかること**

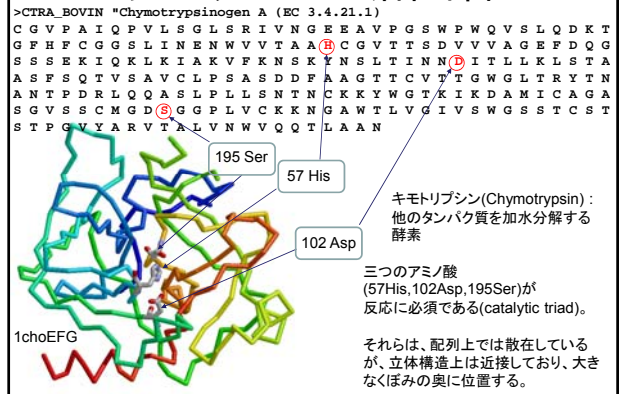
1. 機能(主に結合)に関する理解が深まる

- 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**



### 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)

2ht7A      ヒト・異常染色体      1opJA

インフルエンザH3N8 ノイラミニダーゼ      タンパク質チロシンキナーゼABL1

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？

似た構造を探す

RMSD=3.1 Å  
SeqID = 14.8 %  
Rel for Sfam = 94.2 %  
Rel for Fold = 99.2 %

機能未知立体構造      機能既知立体構造

1p9vA Hypothetical protein YddE Escherichia coli      1bwZA Diaminopimelate epimerase Hemophilus influenzae

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

### 立体構造データの計測とデータフォーマット

### 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 → 原子モデルの構築 → 原子モデルの精密化

3次元電子密度マップ

**NMR(核磁気共鳴法)**

大量発現 → 精製 → 核磁気共鳴解析 (NOE解析, 偏風と距離拘束の抽出) → 原子モデルの構築 → 原子モデルの精密化

原子間距離拘束

- (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

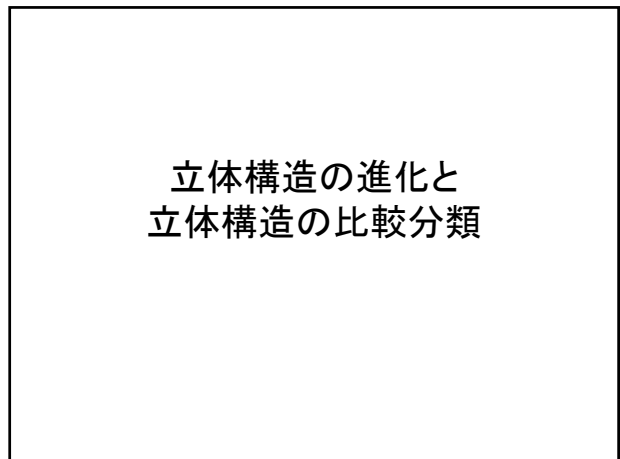
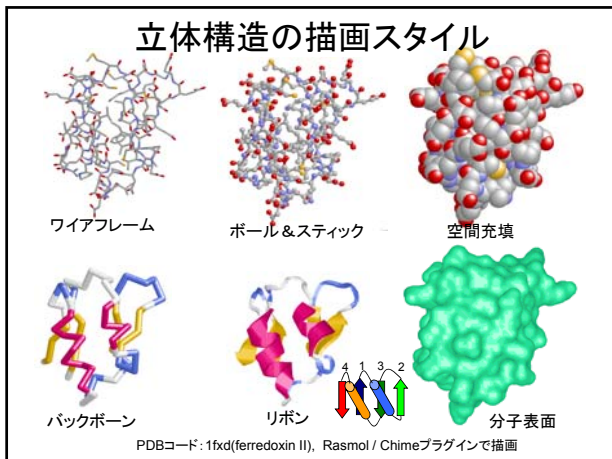
```

HEADER    UBIQUITIN                                17-APR-92    1AAR
COMPND    DI-UBIQUITIN
SOURCE    BOVINE (BOS TAURUS) ERYTHROCYTES
AUTHOR    W.-T. COOK, L. C. JEFFREY, M. CARLSON, Z. CHEN, C. M. PICKART
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

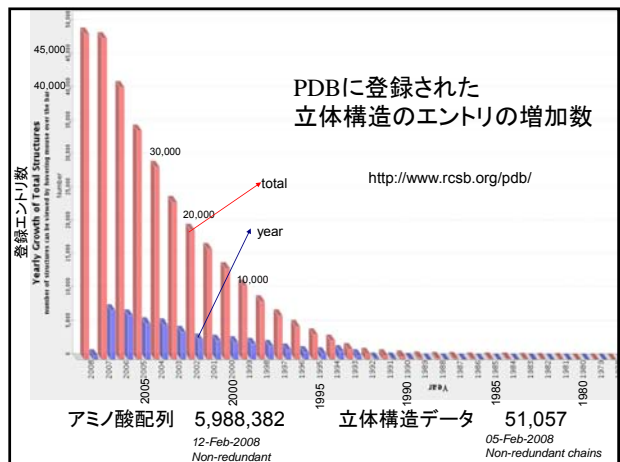
ソフト名	WebSite	対応機種	特徴
RasMol	<a href="http://www.openrasmol.org">www.openrasmol.org</a>	Win Mac Linux (ソース公開)	独自の高速描画アルゴリズムを用い、非力なハードでも同様に高速描画が可能。コマンド言語は強力な多機能。基本的な描画法をカバーするが、分子表面の描画はできない。描画の質はやや低い。
Chime	<a href="http://www.mdchime.com">www.mdchime.com</a>	Win Mac (ソース非公開)	RasMolをベースにしたWEBブラウザ内で動作するプラグイン。分子表面表示など描画機能はRasMolより拡張されている。WEBページに分子描画機能を加えることが可能。
Jmol	<a href="http://jmol.sourceforge.net">jmol.sourceforge.net</a>	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	<a href="http://pymol.sourceforge.net">pymol.sourceforge.net</a>	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、ソースも公開。Python言語を用いた機能拡張も可能。
CHIMERA	<a href="http://www.cgl.ucsf.edu/chimera/">www.cgl.ucsf.edu/chimera/</a>	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	<a href="http://www.ks.uiuc.edu/Research/vmd/">www.ks.uiuc.edu/Research/vmd/</a>	Win Mac Linux (ソース非公開)	OpenGLを用いた質の高い描画。C++言語を用いた機能拡張が可能。分子力学の可視化を主要目的として幅広い機能。
Swiss PDB Viewer	<a href="http://au.expaty.org/spdbv">au.expaty.org/spdbv</a>	Win Mac Linux (ソース非公開)	OpenGLを用いた質の高い描画。幅広い機能をサポートしているが、現在あまり活発にアップデートが進んでいない。



### タンパク質立体構造の分類と比較

#### 立体構造比較の重要性

1. 「データ整理」: 立体構造データ数は近年急増(約3万エントリー)
2. 「進化」: 立体構造は配列より進化的に保存しやすい
  - (1) 配列では見つかからない遠いホモログの発見
    - → 機能推定につながる
    - → 生物の初期進化の解明
  - (2) 正確なアライメント
3. 「物理化学」: 構造データベースの統計から、タンパク質の物理化学的な性質が明らかになる可能性
  - (1) アミノ酸の2次構造傾向
  - (2) 安定な2次構造のパッキング
  - (3) リガンド結合のパターン



### 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

配列の部分同士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...

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://www.sanger.ac.uk/Software/Pfam/>

各蛋白質のファミリーのHMMのプロフィール、マルチプルアライメントを集めたデータベース

The screenshot shows the Pfam website interface. On the left, there's a navigation menu with options like 'Home', 'About', 'FAQ', 'Help', 'Contact', 'Links', 'Downloads', 'API', 'Feedback', and 'Privacy Policy'. The main content area displays a protein family profile for 'PF00089.0' (a beta-barrel domain). It includes a sequence logo, a multiple sequence alignment, and a description of the protein family. The alignment shows conserved residues across various species, with a color scale indicating conservation levels.

## 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(Ca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)

Fold(フォールド)

Superfamily(スーパーファミリー)

Family(ファミリー)

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

The screenshot shows the SCOP database entry for 'Protein: Flavodoxin from Desulfovibrio vulgaris'. It displays the protein's lineage, PDB entry domains, and a 3D ribbon diagram of the protein structure. The lineage is as follows: 1. Root: all-α, 2. Class: Alpha and beta proteins (αβ) [51349], 3. Fold: Flavodoxin-like [52171], 4. Superfamily: Flavodoxin [52218], 5. Family: Flavodoxin [52219], 6. Protein: Flavodoxin [52220], 7. Species: Desulfovibrio vulgaris [52222]. The PDB entry domains listed are 1. Ifu (1fueA), 2. Ifu (1fueA), 3. Ifu (1fueA), and 4. Ifu (1fueA).

## クラス

The diagram illustrates four protein classes based on their structural motifs:
 

- all-α (a)**: 1mbd. α-ヘリックスが中心 (α-helix centered).
- α/β (c)**: 1n55A and 1pqaA. α-ヘリックスとβ-シートが配列上交互に現れる。平行β-シートが中心 (α-helix and β-sheet appear alternately in sequence. Parallel β-sheet centered).
- all-β (b)**: 1mqkH. β-シートが中心 逆平行が多い (β-sheet centered, many antiparallel).
- α+β (d)**: 1fxd and 1a2p. α-ヘリックスとβ-シートが混在、配列上の並びの規則はない 逆平行と平行のβ-シートが混在 (α-helix and β-sheet coexist, no sequence order rule. Antiparallel and parallel β-sheets coexist).

## 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シン (同一残基率 SeqID 44%)

ピロリ菌(1fueA)

大腸菌(1ag9A)

The image shows a comparison of the protein structures from Helicobacter pylori (1fueA) and Escherichia coli (1ag9A). The structures are shown as blue and green ribbons. The SeqID is 44% and the RMSD is 1.2 Å. Below the structures, the amino acid sequences are aligned:
 

```

    1fueA 2: GKIGIFPGTDSGNAEALAEKISKAIQ--NAEVDVAKASKEQFNGFTKVLVAPTAGAGD:59
    1ag9A 2: AITGIFPGSDTGNTENIAKMIQKQLGKDVADVHDIAKSKEDLEAYDILLGIPTWYGYG:61
    1fueA 60: LQTDWEDPLGTLEASD-FANKTIGLVGLGDDQTYSETFQEGIFHYEKAK--AGKVVGGQT:116
    1ag9A 62: AQCDWDDFFPTLE-EIDFNGKLVALPGQGDQDYAEYFCDALGTIRDIIIEPRGATVGHW:120
    
```

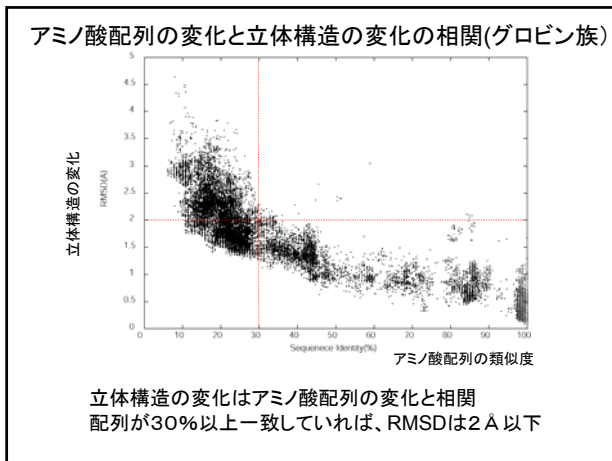
 The text below the alignment states: 配列が似ていれば立体構造も似ている (If the sequences are similar, the structures are also similar).

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

The diagram shows structural similarity between proteins with different sequences. It compares:
 

- ピロリ菌・フラボドキシシン(1fueA) and 大腸菌・フラボドキシシン(1ag9A) with a BLAST score of 44% and 1.2 Å RMSD.
- 大腸菌・変性タンパク質CheY (3chy) with a structural comparison score of 8% and 4.4 Å RMSD.
- ラット・NADPHシクロームP450還元酵素 C末ドメイン(1ja1A2) and ヒト・キノン還元酵素(104aA) with a PSI-BLAST score of 22% and 2.9 Å RMSD.
- ラット・NADPHシクロームP450還元酵素 C末ドメイン(1ja1A2) and ヒト・キノン還元酵素(104aA) with a structural comparison score of 14% and 3.2 Å RMSD.

 A secondary structure diagram at the bottom right shows a β-barrel fold with 5 strands labeled 1 to 5. The text at the bottom states: 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 (Structure is more evolutionarily conserved than sequence → possibility of discovering new homologs from structural comparison).



### 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)

ラット-NADPHシトクロームP450還元酵素 C末ドメイン (1ja1a2) ... "Flavoproteins" superfamily (c.23.5)

ヒト-キニン還元酵素(1d4aa)

### 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fhjkl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

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。「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

### スーパー・フォールド(Superfold)

多くのスーパーファミリーを含むフォールド。一般に機能も多彩。

フォールド名 (SCOP1.73による、クラスfhjkl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: 「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

### フォールド・ファミリーの数は有限

Chothiaは、生物界には約1000のファミリーが存在すると推定 (Nature. 1992年, 357,543-544)

SCOP 1.73 (Nov 2007) (Class fhjklを除く)

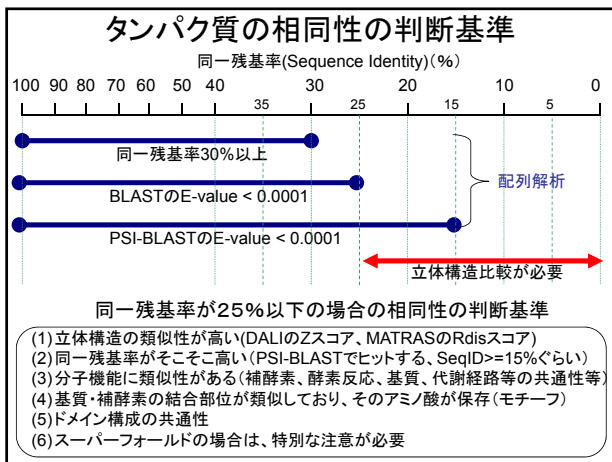
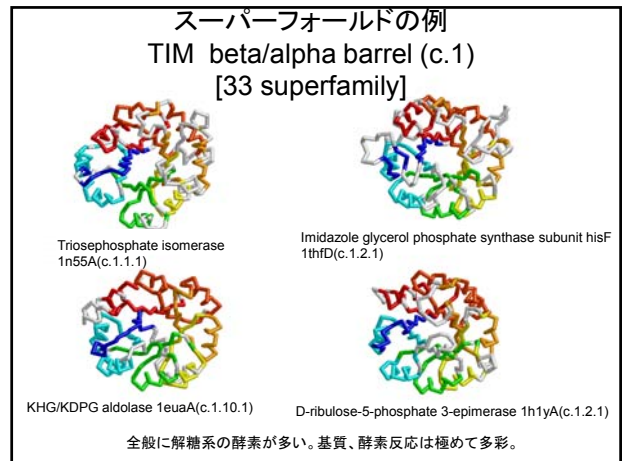
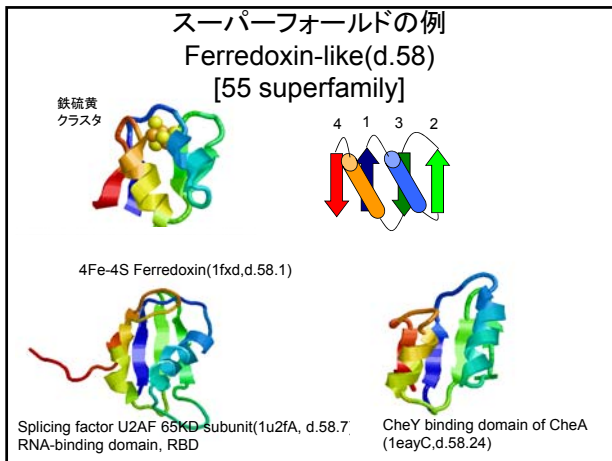
フォールドの数 : 1036  
 スーパーファミリーの数 : 1685  
 ファミリーの数 : 3360

1000個程度の立体構造を解けば、生物界のすべての基本構造がわかる? (「構造ゲノム科学」の発想の原点)

ごく少数のフォールドが多数のスーパーファミリーをかかえる

一つしかスーパーファミリーを含まないフォールドが圧倒的に多い

大多数の蛋白質はメジャーな100程度のフォールドに含まれるが、残りの蛋白質のフォールドは多様



### 立体構造比較プログラム

5万以上の立体構造から類似構造を探すには計算機の支援が不可欠

1990年ごろから多くの構造比較プログラムが開発されている

DALI (<http://www.ebi.ac.uk/dali>)

CE (<http://cl.sdsc.edu/ce.html>)

VAST (<http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml>)

MATRAS (<http://biunint.naist.jp/matras/>)

#### 構造比較プログラム開発における2つの問題

1. どうやって類似性を定義するか
2. どうやって高速に類似構造を検索するか

### 距離の差のスコアと座標の差のスコア

※原子の対応付けがわかっているとする

$$DRMS = \sqrt{\frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j=2}^N (D_{ij}^x - D_{ij}^y)^2}$$

対応する原子ペアの距離の差 (Distance-based RMSD)

※重ね合わせは不要  
※鏡像の区別はできない

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N [x_i - R(y_i - t)]^2}$$

最適に重ね合わせたときのXYZ座標の差 (Root Mean Square Deviation)

並進ベクトル*t*を引いたあと、回転行列*R*をかける  
*t* : 2つの分子の重心を合わせるように決める  
*R* : 2つの原子群の相関を最大化する行列  
→特異値分解で解析的に解ける

### 立体構造アライメントのヒューリスティックな方法

#### MATRASの階層的アライメント

STEP1  
ベクトル表現された2次構造間の対応をBuild-up法で計算

STEP2  
仮アライメントをもとに動的計画法を用いたアライメントを反復的に繰り返す

2次構造のベクトル表現

# タンパク質の立体構造予測

Amino acids sequence

## 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。

ステップ1: フォールド認識  
BLAST/FASTA, プロフィール法, ...

ステップ2: モデリング  
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 **ALIMSTKGFVS**  
Structure **LLLM---GFIT**

(2) 側鎖原子の構築

Sequence **AYVIND**  
Structure **AFVVTD**

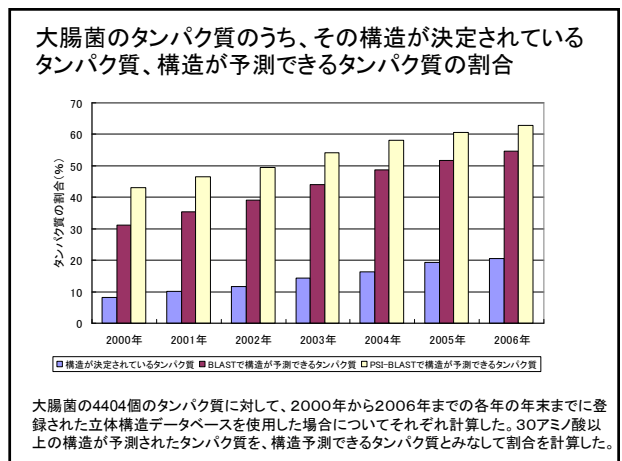
MODELLER : <http://www.sailab.org/modeller/modeller.html>  
FAMS <http://www.pharm.kitasato-u.ac.jp/biomoleculardesign/>

## モデリングした構造の精度と用途

精度 (MODELLER ACCURACY)	用途 (APPLICATIONS)
1.0Å 100%	反応メカニズムの理解 リガンドの設計
1.5Å 95%	ドッキング of macromolecules, prediction of protein partners 高分子のドッキング
3.0Å 80%	Virtual screening and docking of small ligands 低分子のドッキング
4.0Å 60%	Defining antibody epitopes [分子置換法による精密化]
4.0Å 60%	Molecular replacement in X-ray crystallography 部位特異的置換のサポート
4.0Å 60%	Designing oligomers, elastic, crystallizable molecules [NMRの精密化]
4.0Å 60%	Supporting site-directed mutagenesis [電顕等の粗い電子密度へのフィット]
4.0Å 60%	Refining NMR structures Ab initio
4.0Å 60%	Fitting into low-resolution electron density 保存している表面残基の発見
4.0Å 60%	Structure from sparse experimental restraints
4.0Å 60%	Functional relationships from structural similarity
4.0Å 60%	Identifying patches of conserved surface residues
4.0Å 60%	Finding functional sites by 3D motif searching

SeqID = 100 %  
SeqID = 50 %  
SeqID = 30 %

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



## 構造予測コンテスト CASP

Critical Assessment of techniques for protein Structure Prediction  
<http://predictioncenter.llnl.gov/>

- (1) ターゲットタンパク質の募集  
半年以内に立体構造が解ける予定のタンパク質を広く構造生物学者から募集
- (2) ターゲットタンパク質のアミノ酸配列だけをWEBで公示
- (3) 予測者は、期日までに、予測立体構造をサブミット
- (4) ターゲット立体構造が公表された後、予測立体構造と比較・優劣を判断

- John Moultらが主催で1994年に開始。2年おきに開催。
- 2004年にCASP6が開催。世界各国から 224チーム、65サーバが参加。
- 予測の手法・難易度によっていくつかのカテゴリに分けて審査

Comparative Modeling  
 Fold Recognition (Homologous)  
 Fold Recognition (Analogous)  
 New Fold



- CASP6ではいくつかの日本人研究者のグループが入賞  
 CBRC-3D (Kentaro Tomii) :Fold Recognition(Homologous)  
 CHIMERA (Mayuko Takeda-Shitaka) :Fold Recognition(Homologous)

## 立体構造からの分子機能の推定

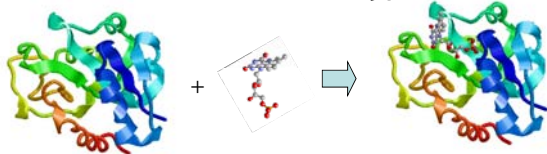
配列・構造比較による相同性からの機能推定の限界

1. 類似度が低い場合、ホモロジーとアナロジーを見分けるのは困難
2. 相同であっても、進化距離が離れている場合、分子機能の違いは大きい
3. 近縁の相同であっても、基質特異性など機能の詳細は異なる

➡ 相同性に依存しない機能推定の手法とは？

ドッキング計算  
 ポケット形状  
 静電場計算  
 分子動力学法  
 基準振動解析

## ドッキング計算



**ドッキング計算**: 結合分子と蛋白質の相互作用エネルギーが最小となる位置を探索することで、**結合位置と結合エネルギー**を推定する計算のこと

- 多くの有償のプログラム (GOLD, ICM, Glide, FlexX...) のほかにいくつかの教育機関無料のプログラム (AutoDock, DOCK) がある。
- 結合低分子のフレキシビリティは考慮するが、蛋白質側の自由度は固定する場合が多い。
- 一般に、結合位置の予測より、結合エネルギーの予測のほうが難しい。

**バーチャル・スクリーニング**: 多数の候補結合分子に対して、ドッキング計算を繰り返すこと。  
**SBDD (Structure-based Drug Design)**: 結合部位の構造をもとに結合分子を設計すること。

## AutoDockプログラム

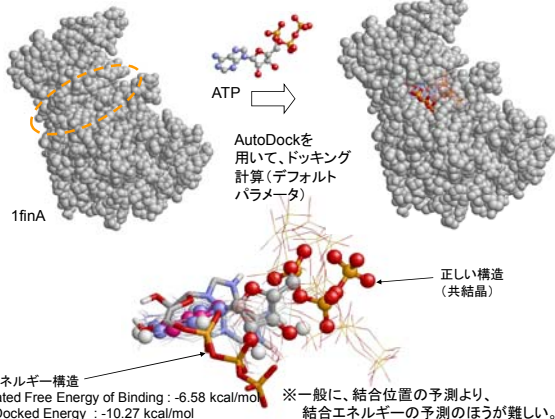
<http://autodock.scripps.edu/>

ポテンシャル関数

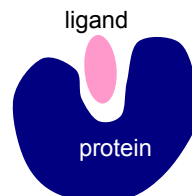
$$\Delta G = w_{vdw} \sum_{i,j} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + w_{bond} \sum_{i,j} E(t) \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^6} \right) + w_{elec} \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} + w_{tor} N_{tor} + w_{sol} \sum_{i,j} (S_i V_j + S_j V_i) \exp[-r_{ij}^2 / 2\sigma^2]$$

$i$ : リガンド原子,  $j$ : 蛋白質原子  
 $E(t)$ : 水素結合のドナー、アクセプターの角度  $t$  に依存する関数  
 $N_{tor}$ : リガンド分子の自由度 ( $sp^3$  結合の数)  
 $S_j$ : 原子  $i$  の水合パラメータ,  $V_j$ : 原子  $j$  のまわりの溶媒接触体積

## 試験計算例 (1finA: Cyclin-dependent protein kinase)



## ポケット形状の探索による結合部位候補の推定



低分子は、一般にタンパク質のポケット (溝、凹部) に結合する傾向

どうしてポケットが好まれるのか？

- (i) 大きな相互作用面積 → 大きな結合エネルギー
- (ii) 水を排除することは酵素の触媒反応に有利
- (iii) 結合による水の排除体積の上昇を抑える

幾何学的にポケット形状を探索することは、結合部位の候補発見に有効

既に様々なポケット認識プログラムがある

SURFNET: <http://www.biochem.ucl.ac.uk/~roman/surfnet/surfnet.html>

PASS: <http://www.ccl.net/ccl/software/UNIX/pass/overview.shtml>

CASTp: <http://cast.engr.uic.edu/cast/>



### 計算されたポケットの例

1qjaA :14-3-3 protein zeta/delta (protein kinase inhibitor protein1)

実際に結合するペプチド

SURFNETプログラムで認識されたポケット

PHECOMプログラムで認識されたポケット (Kawabata & Go, 2007)

⇒ドッキング計算、局所原子配置比較の候補サイトを絞りこむために有効

### 静電相互作用計算による核酸結合部位の推定

各原子に部分電荷  $q_i$  を設定し、静電ポテンシャル  $\phi(r)$  を求める

(1)クーロン相互作用の式

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon} \sum_i \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|}$$

(2)ポアソン-ボルツマン方程式

水の配向効果を連続体として考慮

$$\Delta(\epsilon(\mathbf{r})\Delta f(\mathbf{r})) - k^2 f(\mathbf{r}) + r(\mathbf{r}) = 0$$

空間をグリッドに切り、差分方程式として解く

Grasp : <http://trantor.bioc.columbia.edu/grasp/>

Delphi : <http://trantor.bioc.columbia.edu/delphi/>

### 静電相互作用の計算例

Trp Repressor(1tro)とDNAの複合体

Chime-pluginによる真空のクーロン相互作用の計算・表示

- ・赤がマイナス、青がプラスのポテンシャルを示す
- ・単純化したCHARMMの部分電荷の値を使用

→ DNA結合面はプラスの電荷が多く分布している

### リガンド結合に伴う立体構造の変化

Maltodextrin-binding protein(1omp, 1anf)

異なる構造をつないだアニメーションを集めたサーバ  
Database of Macromolecular Movements (<http://www.molmovdb.org/>)

### リガンド結合に伴う立体構造の変化

Repressor of the diphtheria toxin gene(1dpr, 1f5t)

Fe<sup>2+</sup>の結合による構造変化で、DNAとの結合を制御

非結合状態。開いている。(1dprAB) 結合状態。閉じている。(1f5tABEF)

### 分子動力学法(Molecular Dynamics, MD)

ニュートンの運動方程式

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i = - \frac{\partial V}{\partial \mathbf{r}_i}$$

運動方程式に従って構造を変形

を数値的に解くことで、時間にそった構造変化の履歴を計算する

$m_i$  : 原子  $i$  の質量       $\mathbf{r}_i$  : 原子  $i$  の位置ベクトル  
 $\mathbf{F}_i$  : 原子  $i$  にかかる力       $V$  : ポテンシャルエネルギー

原則的に、小さな時間ステップでの変化を積み重ねた差分法解く  
 1ステップはピコ(10<sup>-12</sup>)秒からフェムト(10<sup>-15</sup>)秒  
 圧力や温度を制御したアンサンブルを生成するために特殊な技法が必要

- ・効率的で緻密なサンプリングが可能。
- ・平衡状態だけでなく、分子の動く過程を観察できる
- ・サンプリングの方法を工夫することで、様々な発展的な使用が可能  
 拡張アンサンブル法、熱力学的摂動法 など

ソフトウェア: AMBER(<http://amber.scripps.edu/>), CHARMM(<http://www.charmm.org/>)

### 生体高分子シミュレーション用の典型的なポテンシャル関数

- ・ポテンシャルエネルギー: 分子の位置、構造に依存するエネルギー  
力場(force field)とも呼ばれる
- ・古典力学的な経験的なエネルギー関数(量子効果は扱わない)
- ・原子中心を作用単位とする ・AMBER, CHARMMが有名

$$V = \sum_{bonds} \frac{k_l}{2} (l_i - l_i^0)^2$$

結合長



$$+ \sum_{angles} \frac{k_\theta}{2} (\theta_i - \theta_i^0)^2$$

結合角



$$+ \sum_{torsions} \frac{V_n}{2} (1 + \cos(n\omega - \gamma))$$

二面角



$$+ \sum_{i=1}^N \sum_{j=i+1}^N \epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

van der Waals相互作用



$$+ \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{4\pi\epsilon_{ij}}$$

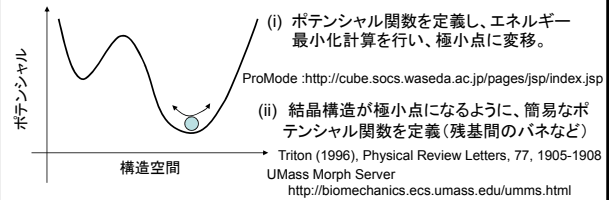
静電相互作用 (水素結合)



### 基準振動解析法

天然構造周辺の本質的な自由度を知る

(1) 構造がポテンシャルエネルギー曲面の極小点に落ちます



(2) ポテンシャルの二次微分行列の固有値を求めることで、線形独立な振動の周波数と動く方向を求める

→天然構造付近の大まかな運動特性がわかる

### 単純な弾性体モデルによる基準振動解析

距離が9 Å以下のCa原子間にバネポテンシャルEを設定

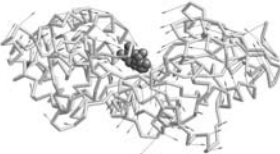
$$E_{ij} = \frac{k}{2} (D_{ij} - D_{ij}^{native})^2$$

Eの2次微分行列Hの固有値 $\omega^2$ 、固有ベクトル $v$ を求める

$$H = \frac{\partial^2 E}{\partial r_i^2 \partial r_j^2}$$

Triton (1996), Physical Review Letters, 77, 1905-1908

弾性体モデルによる基準振動解析の例(マルトース結合タンパク質 10mp)

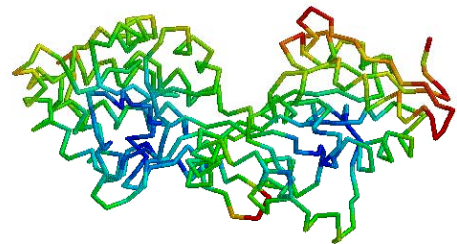


糖の結合していない構造(10mp)から結合していない構造(1anf)への構造変化

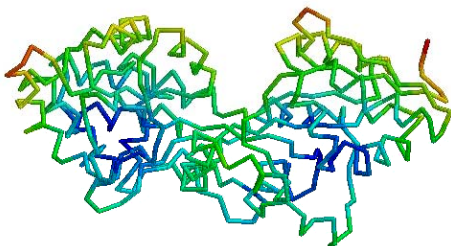


糖の結合していない構造(10mp)に基準振動解析を適用して得られた最低周波数のゆらぎ方向

### 糖の結合していない構造(10mp)から結合している構造(1anf)への構造変化



### 糖の結合していない構造(10mp)に基準振動解析を適用して得られた最低周波数のゆらぎ方向



### 参考図書

構造生物学一般について

- ・ Bluce Alberis他著、中村桂子、松原謙一監訳 「Essential 細胞生物学 原書第2版」 第2章、第4章、2005年、南江堂
- ・ C.Branden & J.Tooze (勝部幸輝ら訳)「タンパク質の構造入門」(2000)、ニュートンプレス
- ・ 松澤 洋編 「タンパク質工学の基礎」 2004年、東京化学同人
- ・ 後藤祐児、桑島邦博、谷澤克行 「タンパク質科学 - 構造・物性・機能 -」 2005、化学同人

構造バイオインフォマティクス

- ・ 郷通子・高橋健一 編集 「基礎と実習 バイオインフォマティクス」 2004年 共立出版
- ・ 藤博幸 編集「はじめてのバイオインフォマティクス」第2.2章 2006年 講談社
- ・ Arthur M. Lesk (高木淳一 訳)「ポストゲノム時代のタンパク質科学」2007年 化学同人
- ・ Arthur M. Lesk (岡崎康司・坊農秀雄 監訳)「バイオインフォマティクス基礎講義 一歩進んだ発想をみかため」 2003年 メディカル・サイエンス・インターナショナル
- ・ 美宅成樹・榎之 「バイオインフォマティクス」 2003年 東京化学同人 第6章、第7章
- ・ 中村春木・有坂文雄 編 「シリーズ・ニューバイオフィジックス1 タンパク質のかたちと物性」 1997年、共立出版
- ・ 後藤祐児、桑島邦博、谷澤克行 「タンパク質科学 - 構造・物性・機能 -」 2005年 化学同人 第2、4、5、6章