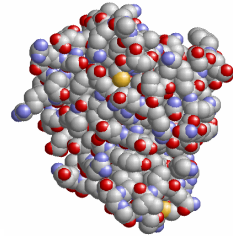
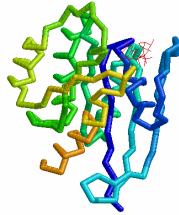


近畿大学・農学部・生命情報学

# 生体高分子の立体構造とその予測

Ras

```
MTEYKLVVVGAGGVGKSAL  
TIQLIQNHFVDEYDPTIED  
SYRKQVVIDGETCLLDILD  
TAGQEEYSAMRDQYMRGTG  
GFLCVFAINNTKSFEDIHQ  
YREQIKRVKDSDDVPMVLV  
GNKCDLAARTVESRQAQDL  
ARSYGIPYIETSATRQGV  
EDAFYTLVREIRQH
```



2007年5月22日(火)

奈良先端大・情報・蛋白質機能予測学講座

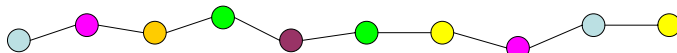
川端 猛

takawaba@is.naist.jp

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

## 生体高分子の立体構造

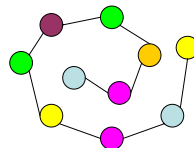
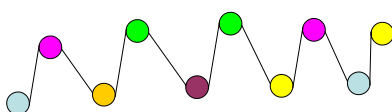
いくつかのユニットがひも状に一系列に並んだ高分子(ポリマー)



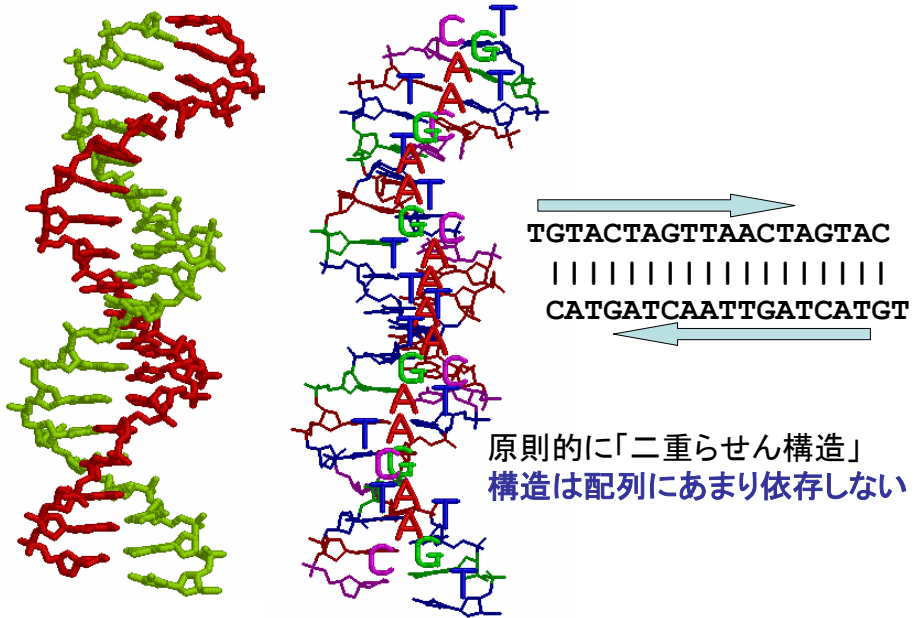
DNA、RNAは4種の塩基 [AT (U) GC]

タンパク質は20種類のアミノ酸 [AVFPMILDEKRSTYHCNQG]

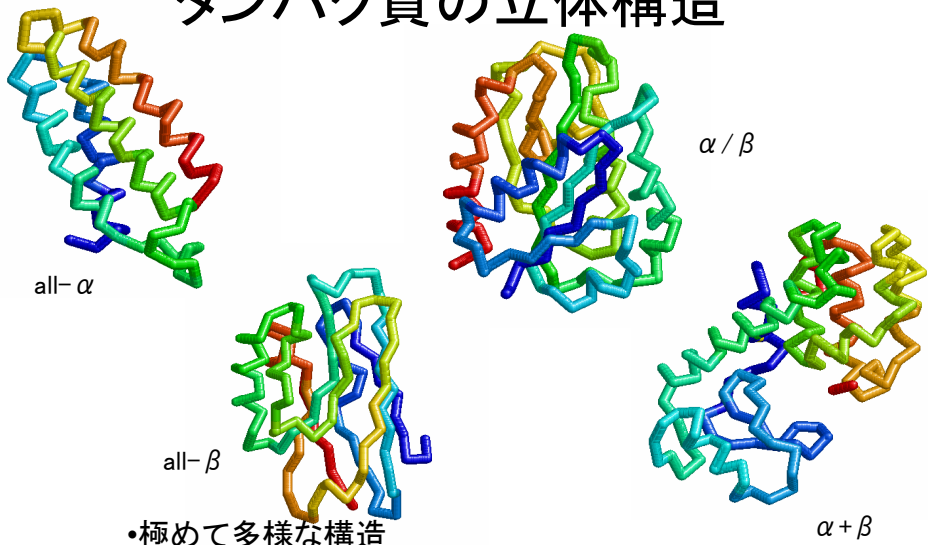
- ひもなので、とにかくフレキシブル
- 原理的にどんな形もとる。
- ユニットの並びや環境によってどんな形を好むかが決まる



# 核酸(DNA)の立体構造



# タンパク質の立体構造

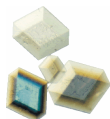


- 極めて多様な構造
- 構造はそのアミノ酸配列によって決定される
- 立体構造の多様性はその分子機能の多様性と関係

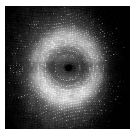
# 立体構造の決定法

## X線結晶解析

大量発現  
精製  
結晶化

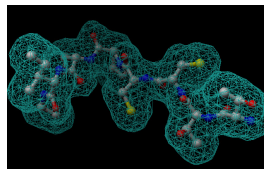


X線回折強度の測定



3次元  
電子密度マップ

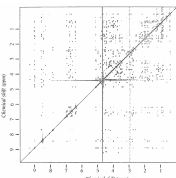
原子モデルの構築  
原子モデルの精密化



## NMR(核磁気共鳴法)

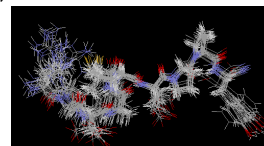
大量発現  
精製

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



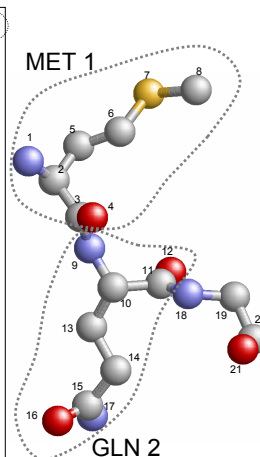
- (1)多数分子の平均構造の観察(精製が重要。大きな分子、複合体はより難しくなる)
- (2)発現・精製・結晶化のステップで、各タンパク質によって実験条件の調整が必須
- (3)実験データの収集・原子モデル構築の段階で計算機の支援が不可欠

# 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バンク

| 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 |

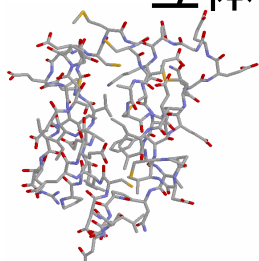


PDBは1971年に設立

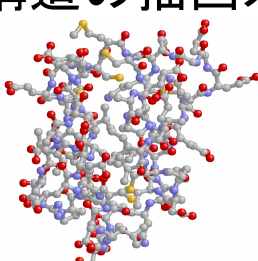
## 立体構造描画のフリーウェア・シェアウェア

| ソフト名                    | WebSite  | 対応機種                      | 特徴   |
|-------------------------|--|---------------------------|--|
| <b>RasMol</b>           | <a href="http://www.openrasmol.org">www.openrasmol.org</a>                       | Win Mac Linux<br>(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。                   |
| <b>Chime</b>            | <a href="http://www.mdlchime.com">www.mdlchime.com</a>                           | Win Mac<br>(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。                       |
| <b>Jmol</b>             | <a href="http://jmol.sourceforge.net">jmol.sourceforge.net</a>                   | Win Mac Linux<br>(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。 |
| <b>PyMOL</b>            | <a href="http://pymol.sourceforge.net">pymol.sourceforge.net</a>                 | Win Mac Linux<br>(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。  |
| <b>CHIMERA</b>          | <a href="http://www.cgl.ucsf.edu/chimera/">www.cgl.ucsf.edu/chimera/</a>         | Win Mac Linux<br>(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。                               |
| <b>VMD</b>              | <a href="http://www.ks.uiuc.edu/Research/vmd/">www.ks.uiuc.edu/Research/vmd/</a> | Win Mac Linux<br>(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。  |
| <b>Swiss PDB Viewer</b> | <a href="http://au.expasy.org/spdbv">au.expasy.org/spdbv</a>                     | Win Mac Linux<br>(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。  |

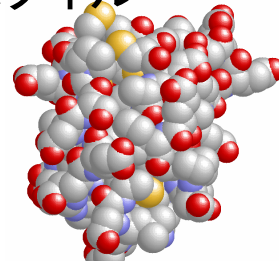
## 立体構造の描画スタイル



ワイアフレーム



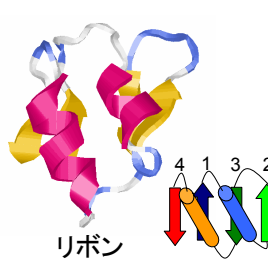
ボール&スティック



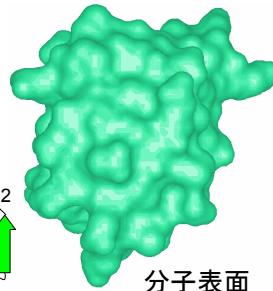
空間充填



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

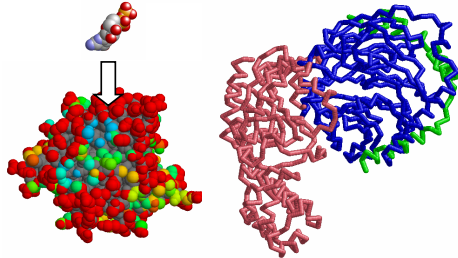
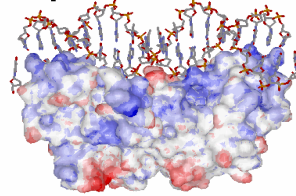
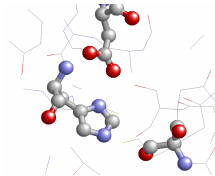
# 立体構造データからわかること

- 分子機能の理解が深まる

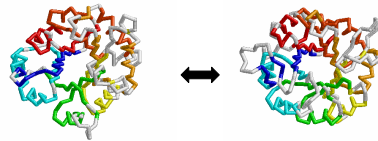
蛋白質の安定性のメカニズムの理解

他分子との結合のメカニズムの理解

酵素反応のメカニズムの理解

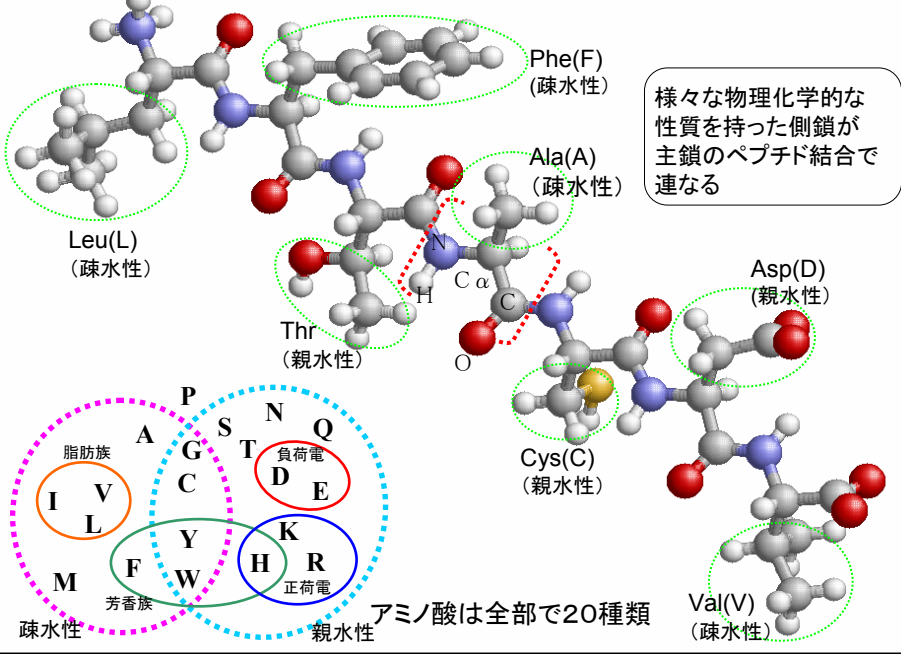


- 構造比較で進化がわかる



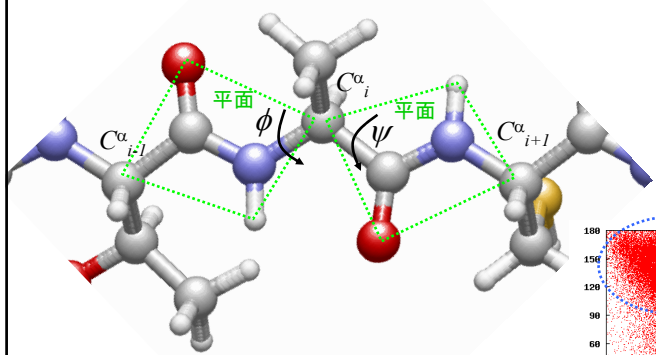
蛋白質の立体構造の  
基本的な要素と  
構造形成の原理

# 蛋白質：ペプチド結合したアミノ酸群

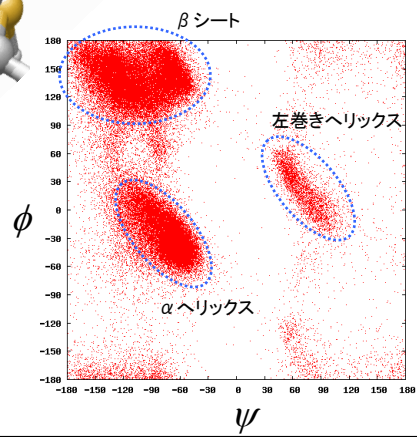


様々な物理化学的な性質を持った側鎖が主鎖のペプチド結合で連なる

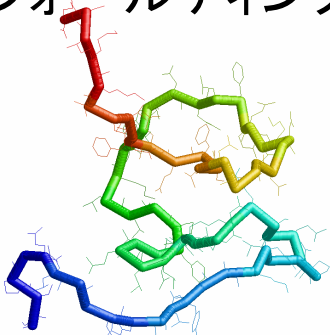
# ペプチド結合の平面性と $\phi$ 、 $\psi$ の定義



- ・主鎖の原子のXYZ座標は、 $\phi$ 、 $\psi$ の2つの角度でほとんど記述できる。
- ・ $\phi$ 、 $\psi$ の角度を2次元にプロットしたものをラマチャンドラン・プロット(Ramachandran plot)という。多くの領域は主鎖原子間の衝突により許されない。頻出領域は2次構造とよく対応。



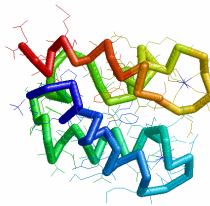
# フォールディング(折り畳み)という現象



変性状態(D)

非常に多種の構造の集合  
大きく広がっている

温度や変性剤濃度  
によって可逆に変化



天然状態(N)

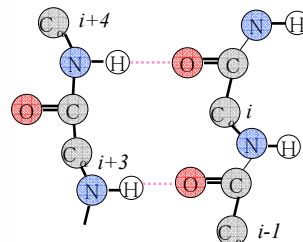
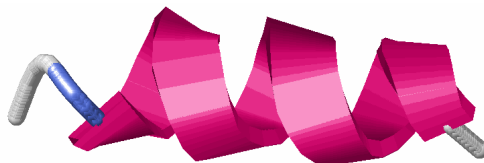
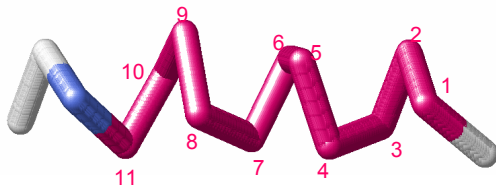
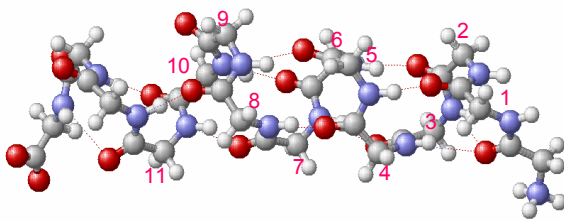
ほとんど唯一の構造  
小さくコンパクトに折りたたまっている

・折り畳みは、原則としてその蛋白質以外の分子の介助を必要としない

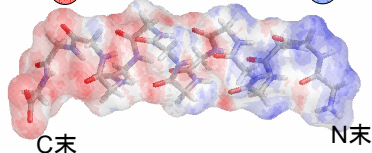
## アミノ酸配列の情報だけで、天然状態の立体構造が決定される

- ・どうやってコンパクトになるか? → (1) 主鎖の水素結合 (2) 側鎖間の疎水性相互作用
- ・あるアミノ酸配列がどうやって一つの構造を決めるのか???

# 主鎖の水素結合: $\alpha$ ヘリックス



- ・  $i$  番目のCOが  $i+4$  番目のNHと水素結合を形成
- ・ 3. 6残基が1周期の右巻きらせん
- ・ NH...O=Cが同じ向きに並ぶため、ヘリックス全体に電気双極子が発生

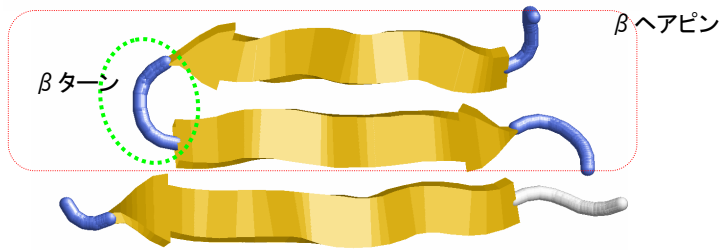
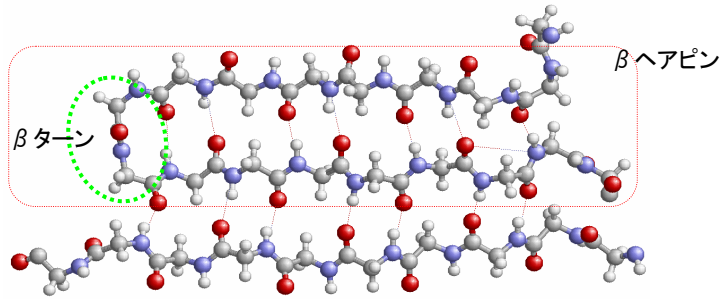


C末

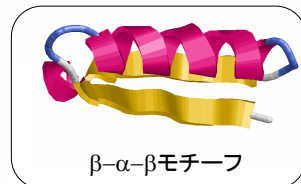
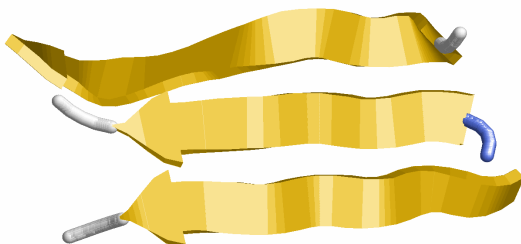
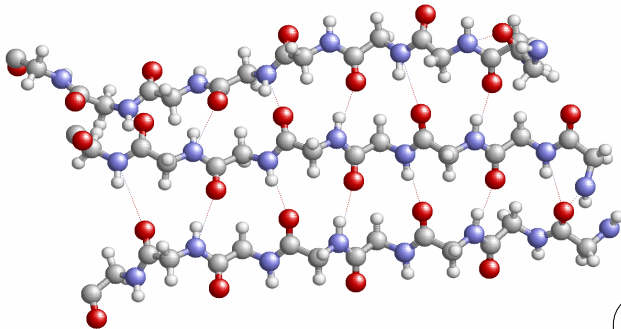
N末



# 主鎖の水素結合：逆平行 $\beta$ シート



# 主鎖の水素結合：平行 $\beta$ シート

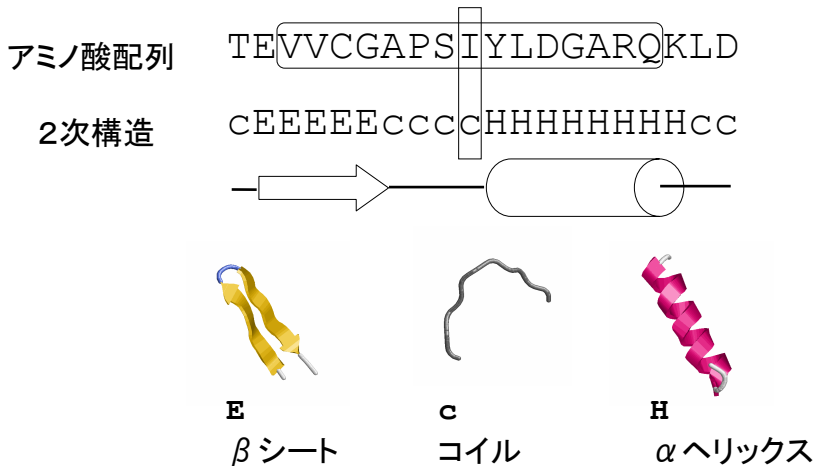


平行 $\beta$ シートが $\alpha$ ヘリックスで  
接続された構造モチーフ



## 2次構造予測

- アミノ酸配列から、2次構造を予測すること。



### (1)アミノ酸ごとに2次構造のなりやすさに差がある

Chou-Fasmanのパラメータ  $P_{Glu}(\alpha) = \frac{f(Glu \cap \alpha)}{f(\alpha)}$

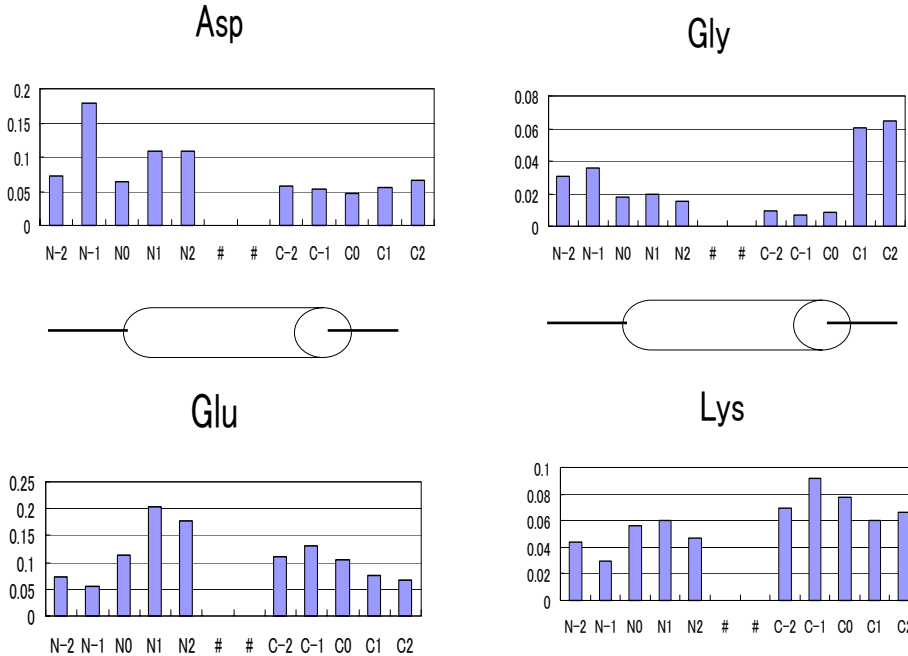
(scop1.59 30%list:3077chainsを用いて再計算した値)

| $\alpha$ ヘリックス | A | 1.42 | E | 1.37 | L | 1.33 | Q | 1.30 | M | 1.27 |
|----------------|---|------|---|------|---|------|---|------|---|------|
|                | R | 1.21 | K | 1.16 | W | 1.07 | I | 1.06 | F | 0.97 |
|                | Y | 0.96 | V | 0.92 | H | 0.87 | D | 0.82 | C | 0.80 |
|                | T | 0.77 | S | 0.77 | N | 0.74 | G | 0.43 | P | 0.42 |

| $\beta$ シート | V | 1.95 | I | 1.77 | F | 1.46 | Y | 1.45 | C | 1.30 |
|-------------|---|------|---|------|---|------|---|------|---|------|
|             | W | 1.29 | T | 1.20 | L | 1.11 | M | 1.01 | H | 0.99 |
|             | R | 0.90 | S | 0.83 | K | 0.79 | Q | 0.77 | A | 0.76 |
|             | E | 0.70 | G | 0.64 | N | 0.61 | D | 0.53 | P | 0.42 |

| コイル | P | 1.66 | G | 1.55 | N | 1.35 | D | 1.33 | S | 1.23 |
|-----|---|------|---|------|---|------|---|------|---|------|
|     | H | 1.09 | T | 1.07 | C | 1.01 | K | 0.98 | R | 0.90 |
|     | Q | 0.90 | E | 0.88 | Y | 0.84 | W | 0.83 | F | 0.82 |
|     | M | 0.81 | A | 0.81 | L | 0.72 | V | 0.65 | I | 0.62 |

(2)ヘリックスのN末端、C末端に出現しやすい残基がある



(3)2次構造ごとに疎水性パターンに特徴がある

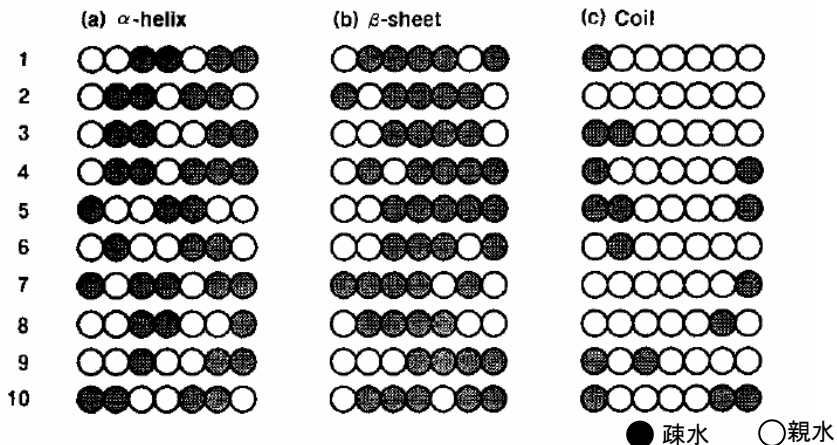
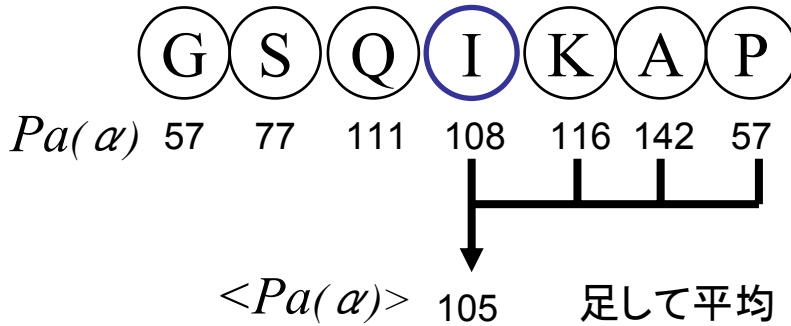


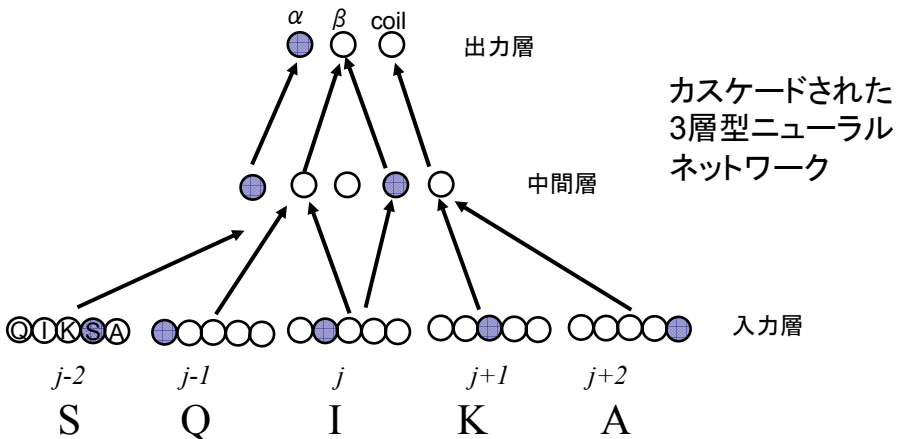
Fig. 9. Binary words that prefer the  $\alpha$  helix (a), the  $\beta$  sheet (b), and the coil (c). They are ranked by the value of the binary word information. The top 10 are shown.

# Chou-Fasmanの方法(1974)



$\langle Pa(\alpha) \rangle$ 、 $\langle Pa(\beta) \rangle$ 、 $\langle Pa(c) \rangle$ のうち  
 最大のものを予測構造とする。  
 正答率は50~55%ぐらい。

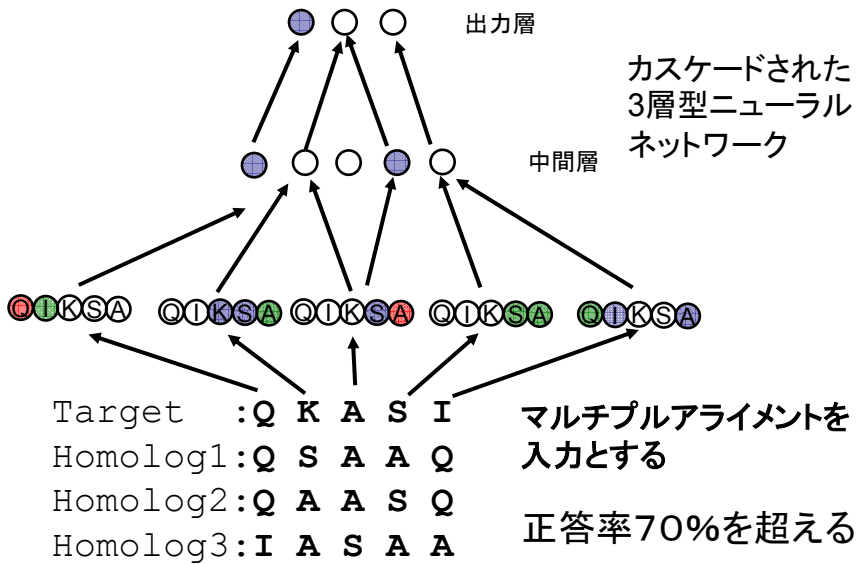
# ニューラルネットワークによる方法 (Qian & Sejnowski, 1988)



誤差逆伝播法(Back propagation method)で学習

← 非線形の回帰解析を最急降下法で解くこと

# Rost & Sander(1993)



<http://cubic.bioc.columbia.edu/predictprotein/>

## 2次構造予測の現状のまとめ

- Chou-Fasman等70年代に開発された方法の多くは正答率は**60%以下**
- ニューラルネットワークや改良GOR法を用いると**63~68%**ぐらいの正答率
- マルチプルアライメントを入力とする方法は正答率は**70%を超える**。現状では**76%**ぐらいまで到達。
- 一般に $\alpha$ ヘリックスに比べ、 $\beta$ シートの予測は困難

現状のベストの方法

マルチプルアライメントを入力とする ニューラルネットワーク

Blast, ClustalW,  
PSI-BLAST, HMMer

アンサンブル学習  
再帰ネットワーク  
K-NNとの組み合わせ



# 疎水性指標による内外予測

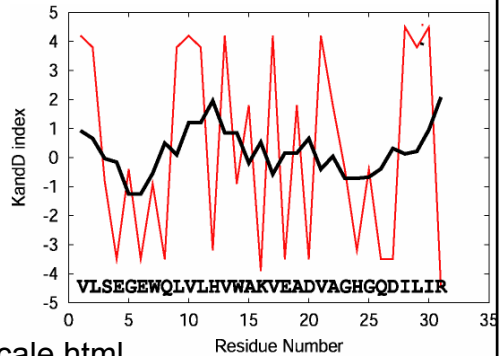
Kyte and Doolittle(1982)の疎水性指標

|   |      |   |      |   |      |   |      |   |      |
|---|------|---|------|---|------|---|------|---|------|
| I | 4.5  | V | 4.2  | L | 3.8  | F | 2.8  | C | 2.5  |
| M | 1.9  | A | 1.8  | G | -0.4 | T | -0.7 | S | -0.8 |
| W | -0.9 | Y | -1.3 | P | -1.6 | H | -3.2 | Q | -3.5 |
| N | -3.5 | E | -3.5 | D | -3.5 | K | -3.9 | R | -4.5 |

## 方法

- ・ 配列を横軸にとって、縦軸に対応する疎水性指標をプロット。
- ・ 前後数残基でスムージング。

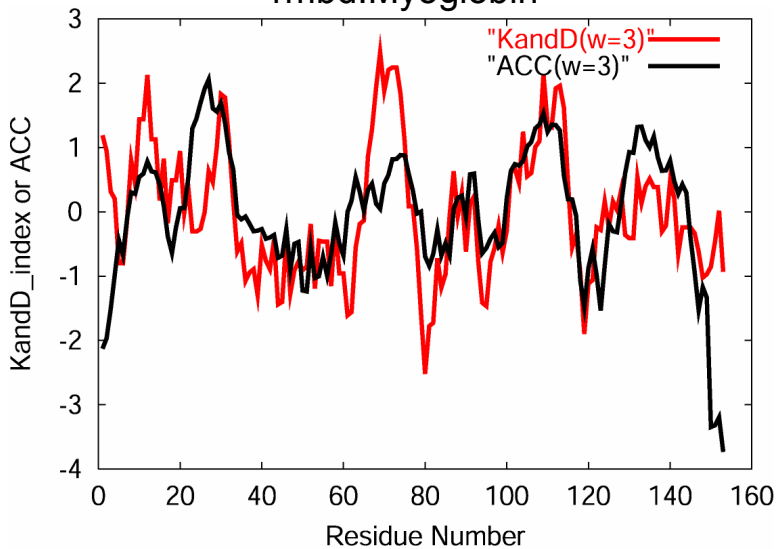
$$V[i] = \frac{1}{2w+1} \sum_{k=-w}^w KD\{Seq[i+k]\}$$



<http://kr.expasy.org/tools/protscale.html>

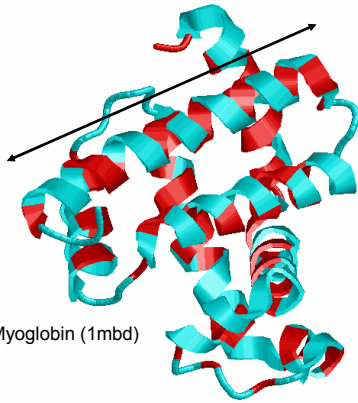
## 予測結果と実際の埋もれ度との比較

1mbd:Myoglobin

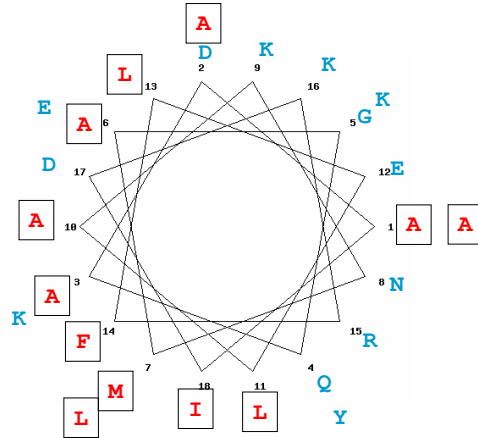


# 疎水性の車輪図(Helical Wheel)

最後のヘリックスの配列: ADAQGAMNKALELFRKDIAAKYKEL

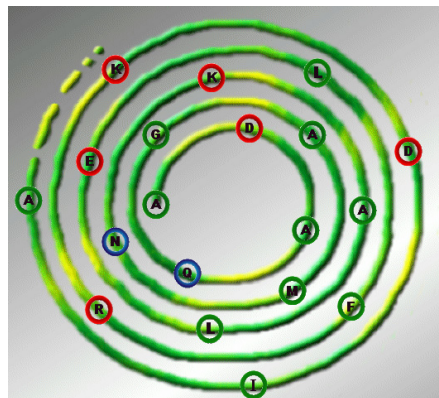
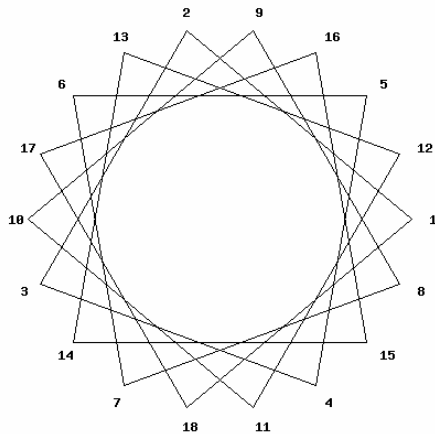


Myoglobin (1mbd)



両親媒性ヘリックス : 片側が疎水的、反対の側が疎水的になっているヘリックスのこと

周期的な疎水性 → ヘリックス構造を示唆

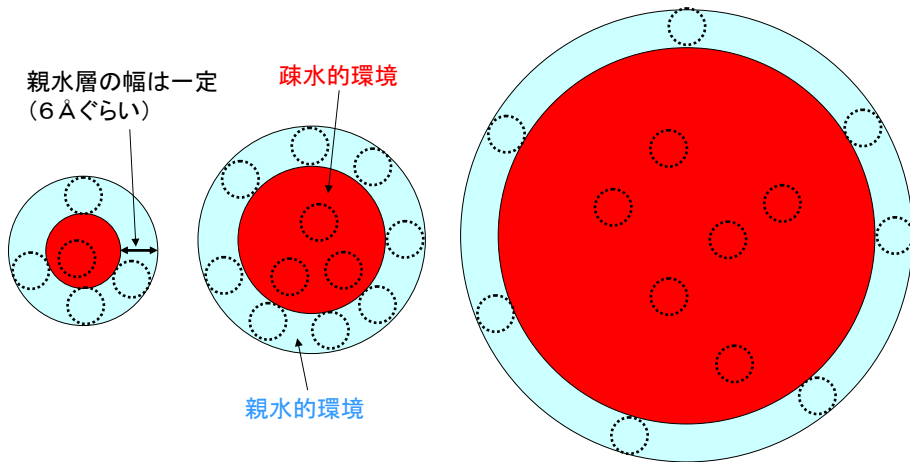


HelixWheel: <http://www.site.uottawa.ca/~turcotte/resources/HelixWheel/>

HelixDraw : <http://www.bioinf.man.ac.uk/~gibson/HelixDraw/helixdraw.html>



# タンパク質の大きさと疎水性の関係



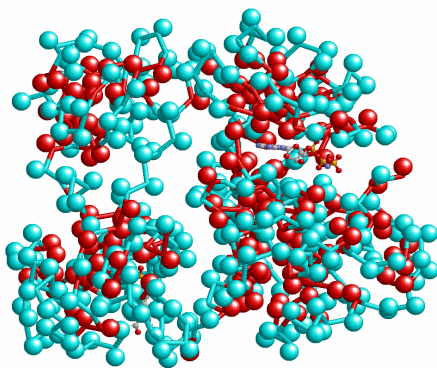
表面の親水層の幅はアミノ酸一つ分 (6 Åぐらい)

タンパク質が球形なら、アミノ酸数が増えるほど、疎水性のアミノ酸の比率が高くなってしまふ

疎水性のアミノ酸が多すぎると、折り畳まる前に凝集して沈殿してしまう危険性

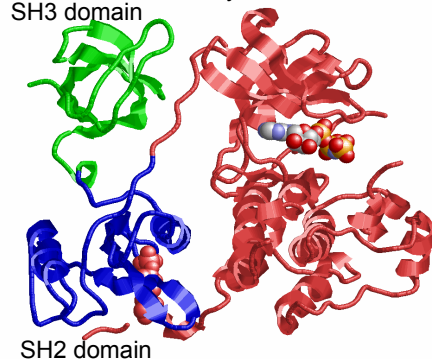
# ドメイン構造と疎水性の関係

Tyrosine protein kinase HCK (1ad5A, HCK\_HUMAN)



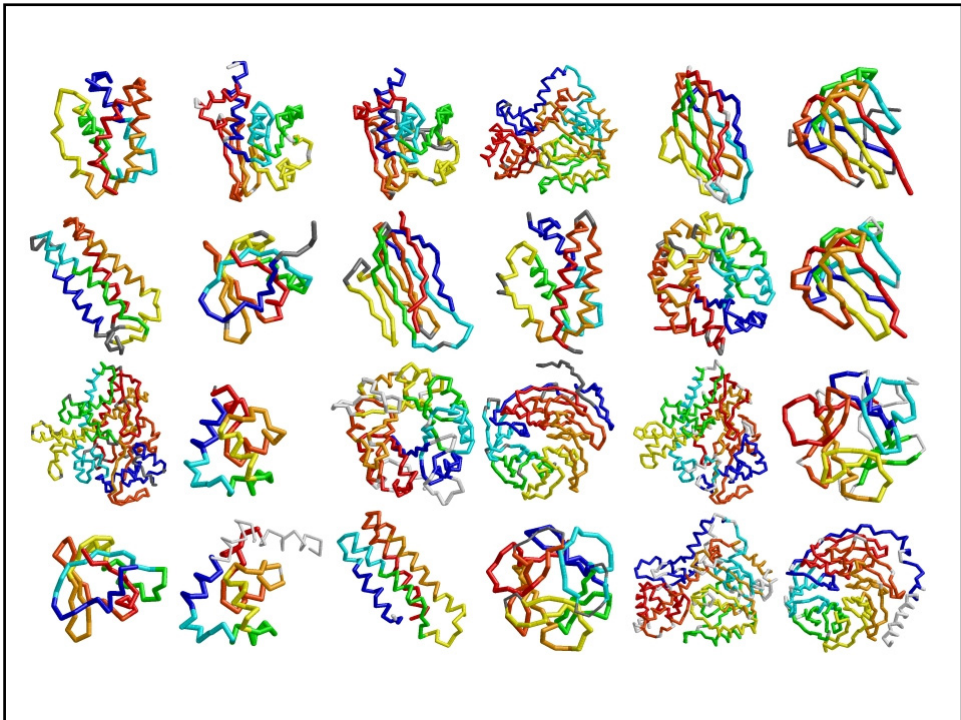
3つの分断された疎水性コア  
→ ドメイン構造

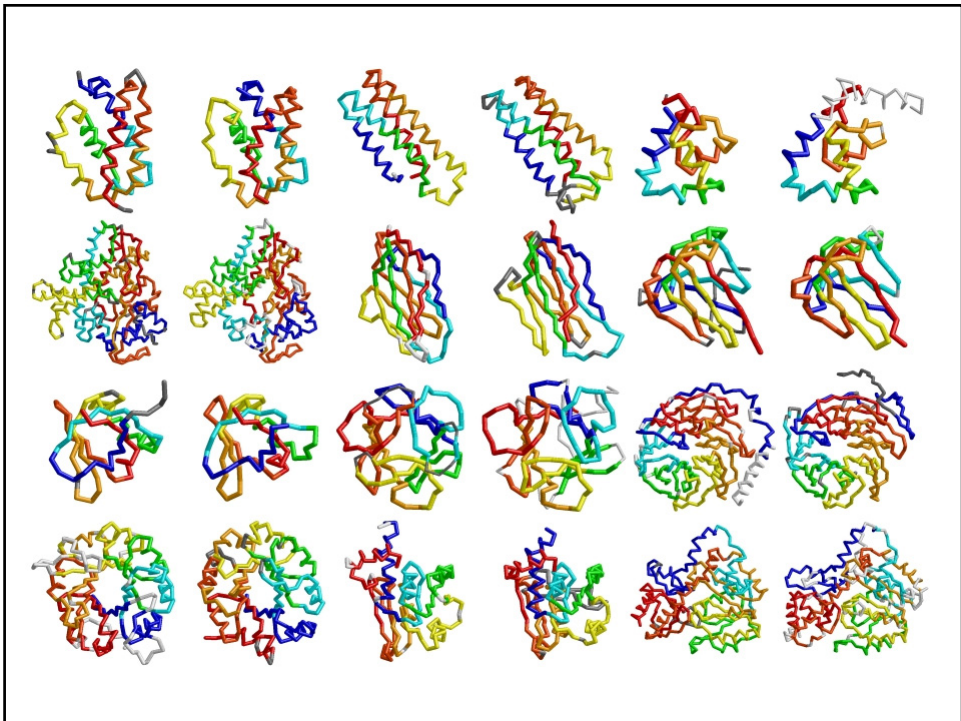
SH3 domain Tyrosine-kinase domain



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

# 立体構造の進化と 立体構造の比較分類

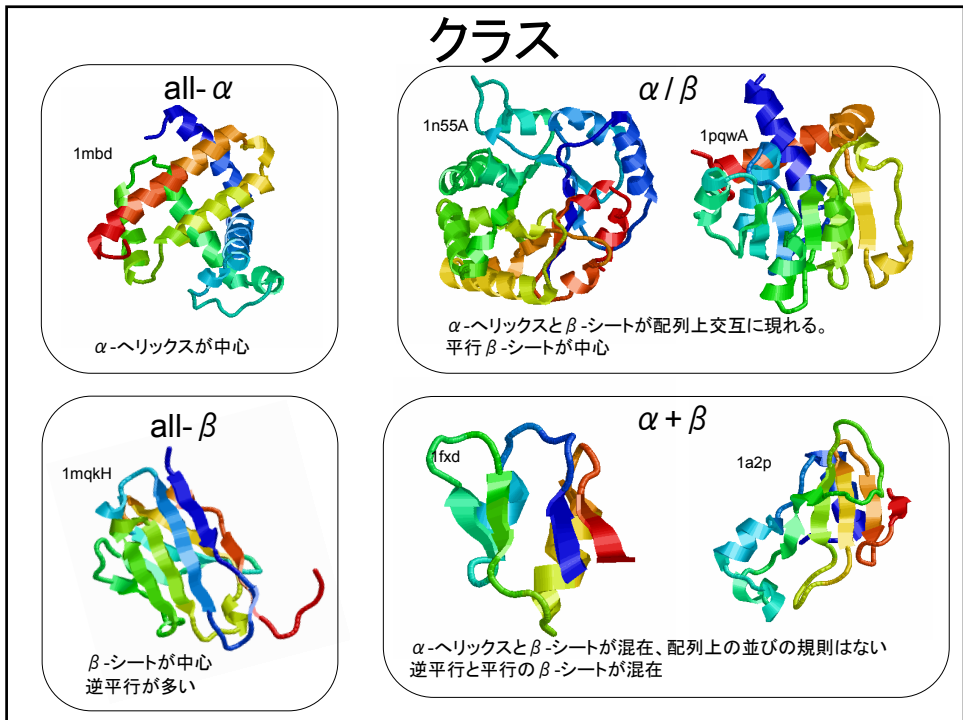
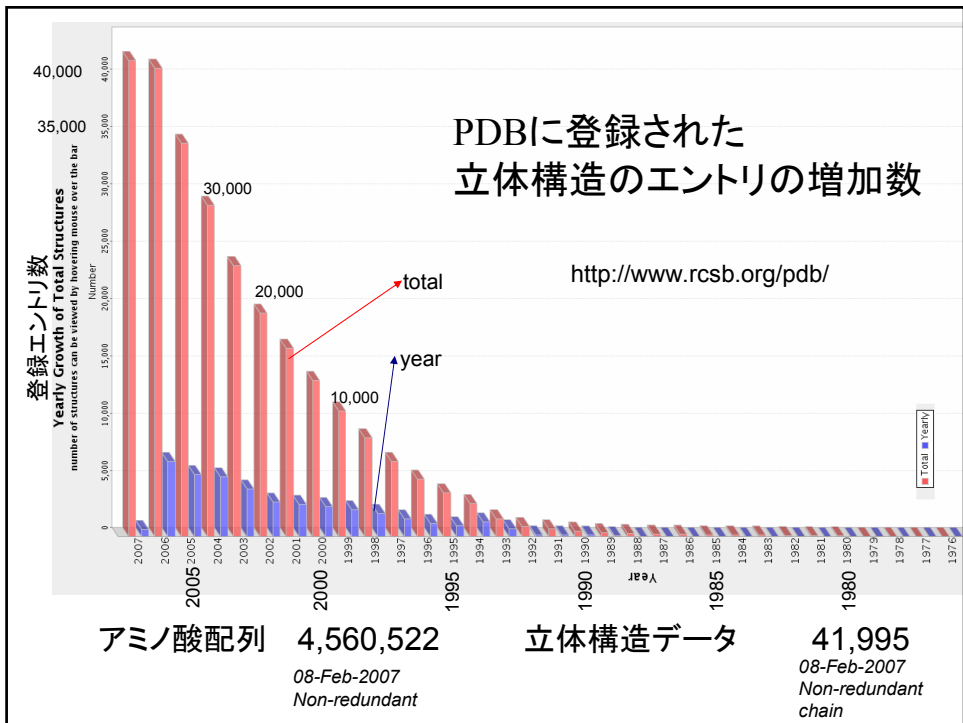




# タンパク質立体構造の分類と比較

## 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) 精確なアライメント
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



# 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

**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:**

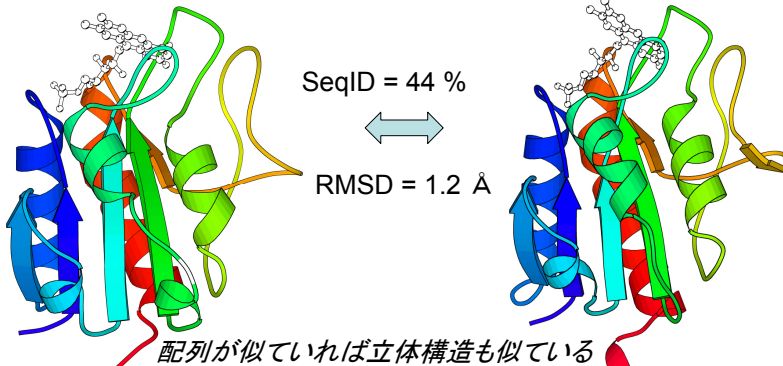
1. [1f4p](#) [59651]   
*complexed with fnn; mutant*  
1. [chain a](#) [59651]
2. [1f8q](#) [62744]   
*complexed with fnn*  
1. [chain a](#) [62744]
3. [1pe](#) [62756]   
*complexed with fnn; mutant*  
1. [chain a](#) [62756]
4. [1ak](#) [31152]   
*complexed with fnn; mutant*

## 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)



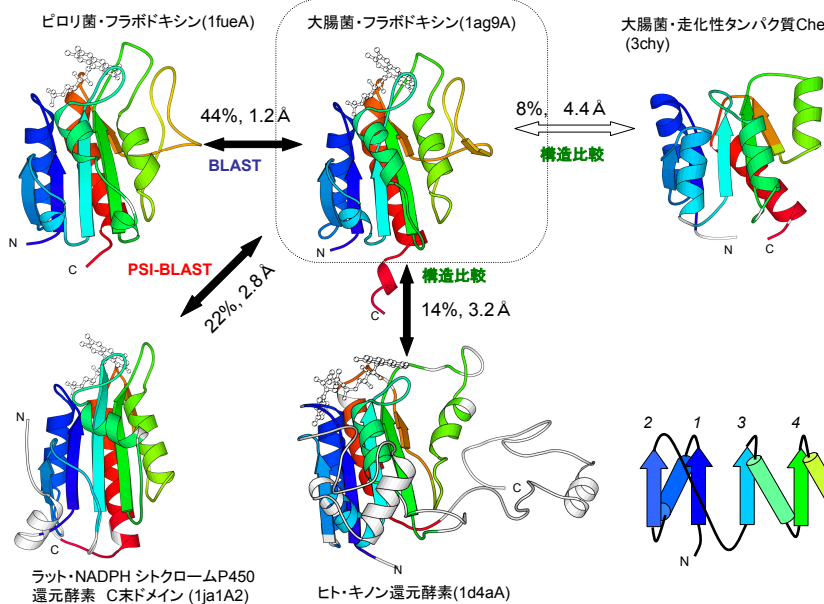
1fueA 2:GKIGIFFFGTDSGNAEAIAEKISKAI G--NAEVVDVAKASKEQFNNGFTKVLVAPTAGAGD:59  
\*\*\*\*\*

1ag9A 2:AITGIFFGSDTGNTENIAKMIQKQLGKDVADVHDI AKSKEDLEAYDILLGLIPTWYYGE:61

1fueA 60:LQTDWEDFLGTL EASD-FANKTIGLVGLGDQDTYSETFAEGIFHIYEKAK--AGKVVVGT:116  
\*\*\*\*\*

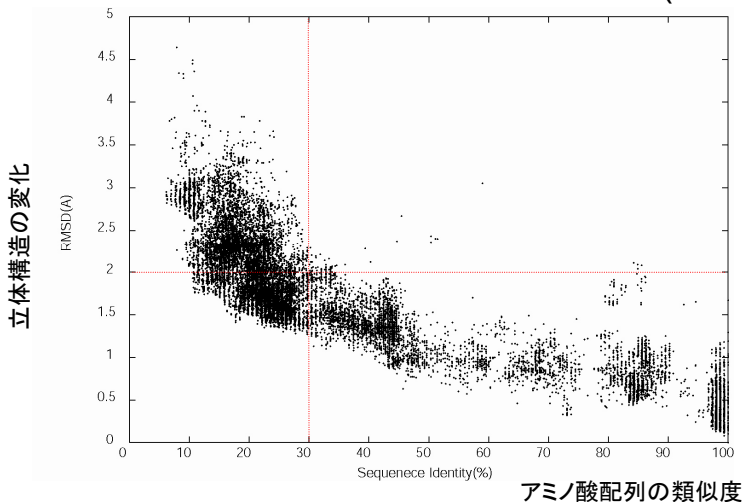
1ag9A 62:AQCWDDFFPTLE-EIDFNGKLVALFGCGDQEDYAEYFCDALGTIRDIIEPRGATIVGHW:120

# 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)



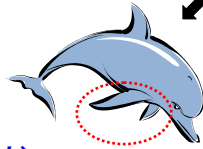
立体構造の変化はアミノ酸配列の変化と相関  
配列が30%以上一致していれば、RMSDは2 Å 以下



## 2つの構造の類似性

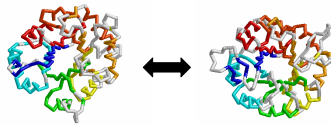
### ホモロジー (homology, 相同):

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。



チョウの羽



トリの翼

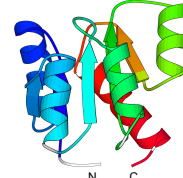
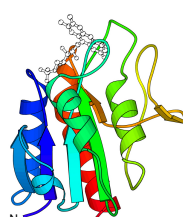
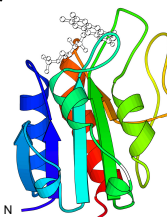


### "Flavodoxin-like" fold

ピロリ菌・フラボドキシ(1fueA)

大腸菌・フラボドキシ(1ag9A)

大腸菌・走行性タンパク質CheY (3chy)



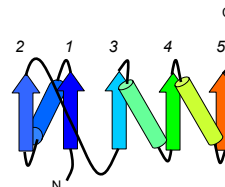
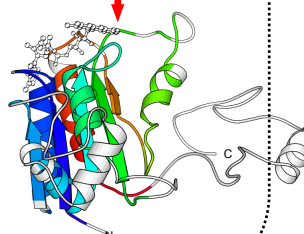
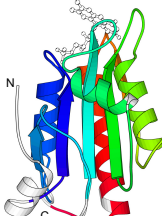
44%, 1.2 Å  
**Homology**

8%, 4.4 Å  
**Analogy**

22%, 2.8 Å  
**Homology**

14%, 3.2 Å  
**Homology**

"CheY-like" superfamily



ラット・NADPH シクロームP450  
還元酵素 C末ドメイン (1ia1A2)

ヒト・キノ還元酵素(1d4aA)

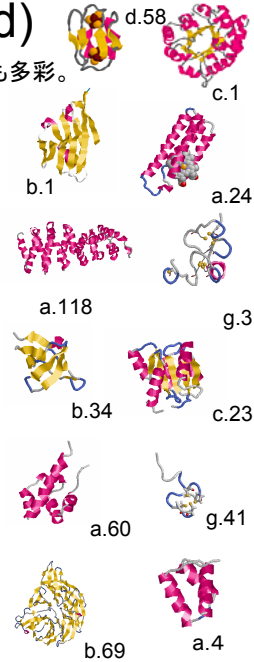
"Flavoproteins" superfamily



# 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 (SCOP1.71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                      | 51              |
| c.1   | TIM beta/alpha-barrel                | 32              |
| b.1   | Immunoglobulin-like beta sandwich    | 25              |
| a.24  | Four-helical up-and-down bundle      | 24              |
| a.118 | Alpha-alpha superhelix               | 20              |
| g.3   | Knottins                             | 19              |
| b.34  | SH3-like barrel                      | 15              |
| c.23  | Flavodoxin-like                      | 15              |
| a.60  | SAM(sterile alpha motif) domain-like | 14              |
| g.41  | Rubredoxin-like                      | 14              |
| 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

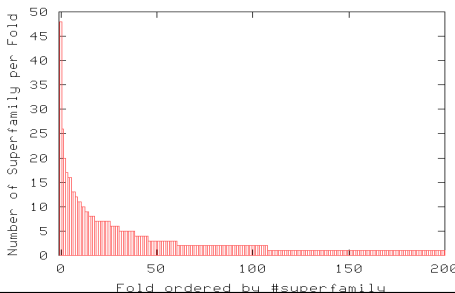
Chothiaは、生物界には約 **1000** のファミリーが存在すると推定  
(Nature. 1992年, 357,543-544)

SCOP 1.69 (Jun 2005)

フォールドの数 : 945  
スーパーファミリーの数 : 1539  
ファミリーの数 : 2845

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)

フォールドあたりのスーパーファミリーの数の分布



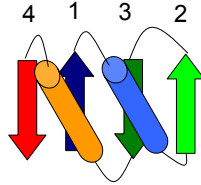
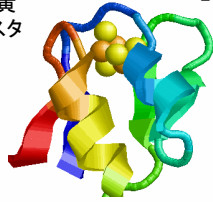
ごく少数のフォールドが多数のスーパーファミリーをかかえる

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

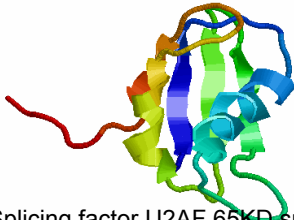
大多数の蛋白質はメジャーな100程度のフォールドに含まれるが、残りの蛋白質のフォールドは多様

スーパーフォールドの例  
Ferredoxin-like(d.58)  
[44 superfamily]

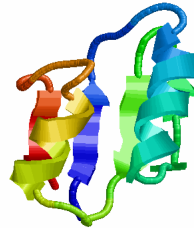
鉄硫黄  
クラスター



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

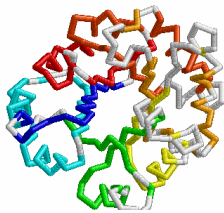


Splicing factor U2AF 65KD subunit(1u2fA, d.58.7)  
RNA-binding domain, RBD

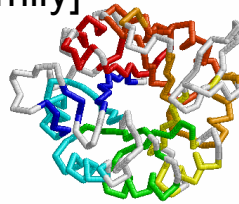


CheY binding domain of CheA  
(1eayC,d.58.24)

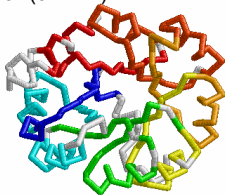
スーパーフォールドの例  
TIM beta/alpha barrel (c.1)  
[26 superfamily]



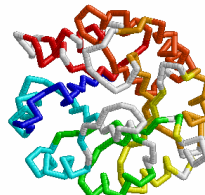
Triosephosphate isomerase  
1n55A(c.1.1.1)



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



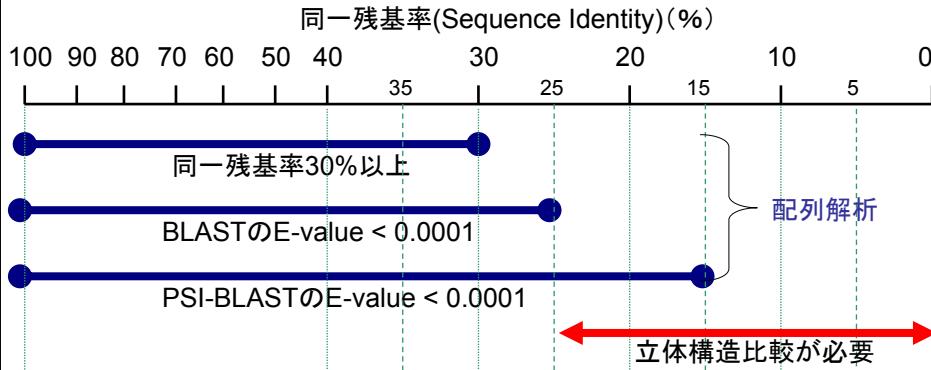
KHG/KDPG aldolase 1euaA(c.1.10.1)



D-ribulose-5-phosphate 3-epimerase 1h1yA(c.1.2.1)

全般に解糖系の酵素が多い。基質、酵素反応は極めて多彩。

# 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こ高い (PSI-BLASTでヒットする, SeqID>=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

# 立体構造比較プログラム

4万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

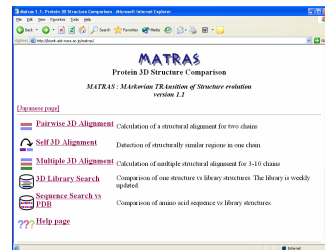
1990年ごろから多くの構造比較プログラムが開発されている

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://biunit.naist.jp/matras/>)

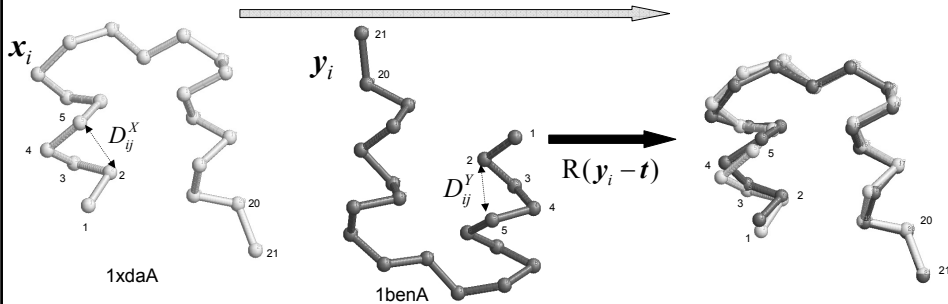


## 構造比較プログラム開発における2つの問題

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



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

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  
 $t$  : 二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

## アライメント

どうやって対応する箇所を見つけるか？

配列のアライメントの場合

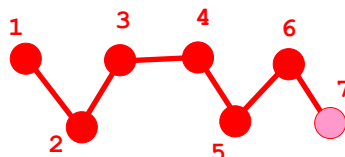
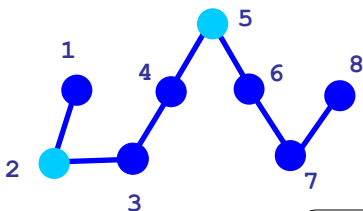
ABCDEF  
CDE

ABCDEF  
--CDE--

BCDEF  
ABEEFG

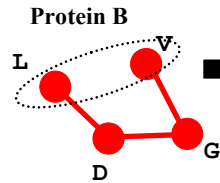
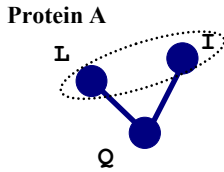
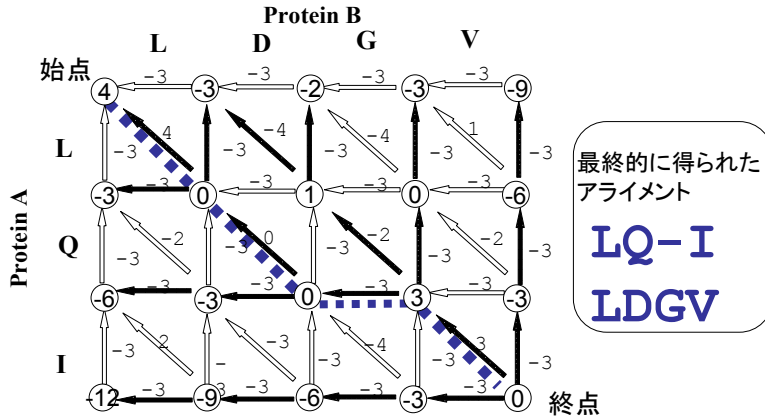
-BCDEF-  
AB-EEFG

立体構造のアライメントの場合



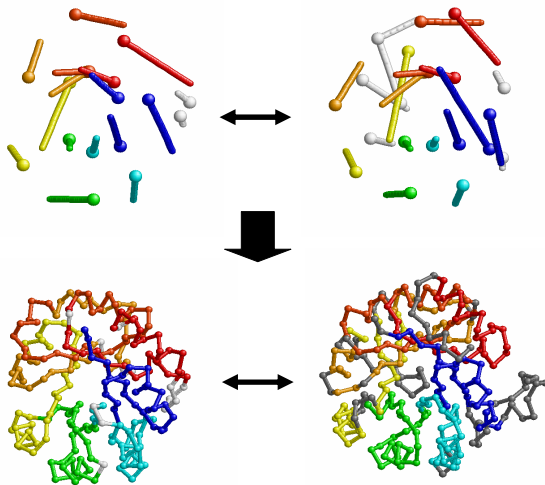
12345678-  
1-23-4567

# 動的計画法による配列アライメント



アミノ酸間距離のスコアはアミノ酸ペアの対応がわからないとスコアが計算できないので、動的計画法は使えない。

# 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



# 構造ゲノム科学: Structural Genomics

high throughputで大量の構造を決定

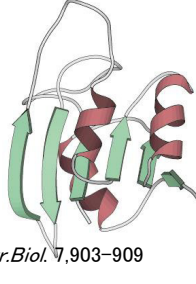
- ・自然界の基本構造（フォールド）の列挙
- ・ゲノム中の機能未知のタンパク質の構造を解く → 機能推定
- ・生物学的に重要な機能を担当するタンパク質を集中的に解く
- ・製薬ターゲットになるタンパク質を集中的に解く

MTH1615

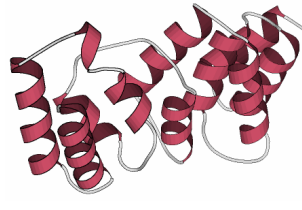


Christendat, D. et al. *Nature Str. Biol.* 7, 903-909

MTH1175



ST1625

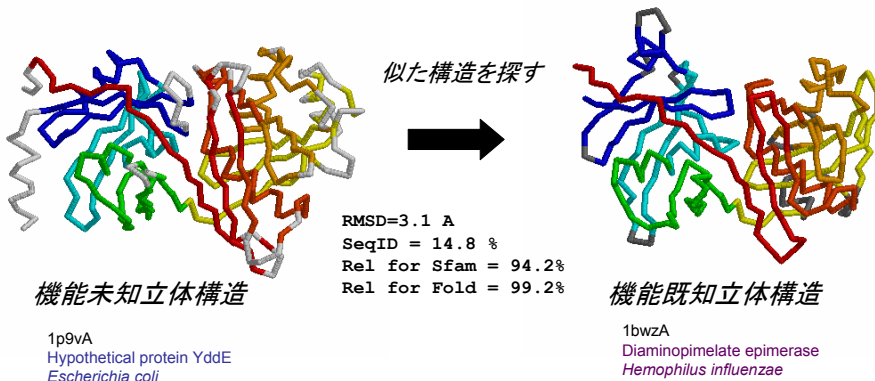


Yoneda, K. et al. (1VDU, 2004)

日本でも理研を中心に、当初から構造ゲノム科学を牽引  
2002年から「タンパク3000」プロジェクトが開始  
<http://www.mext-life.jp/protein/>

## 「形」の比較による弱い相同性認識

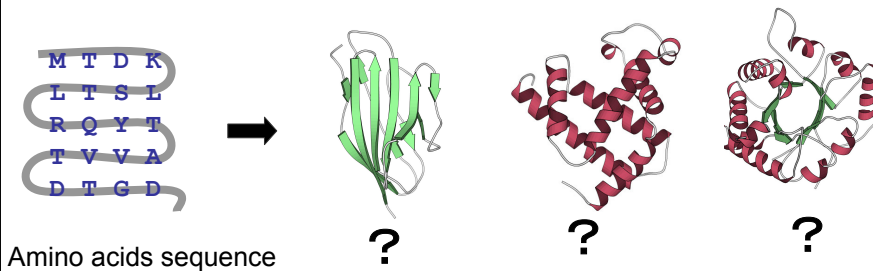
機能未知の立体構造データから機能を予測するには？



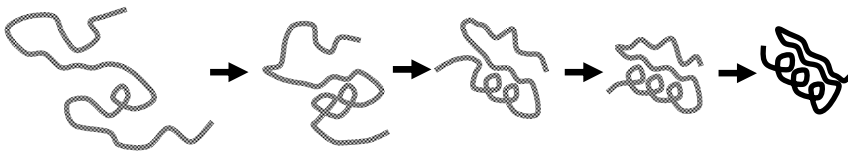
論理の流れ

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

# 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

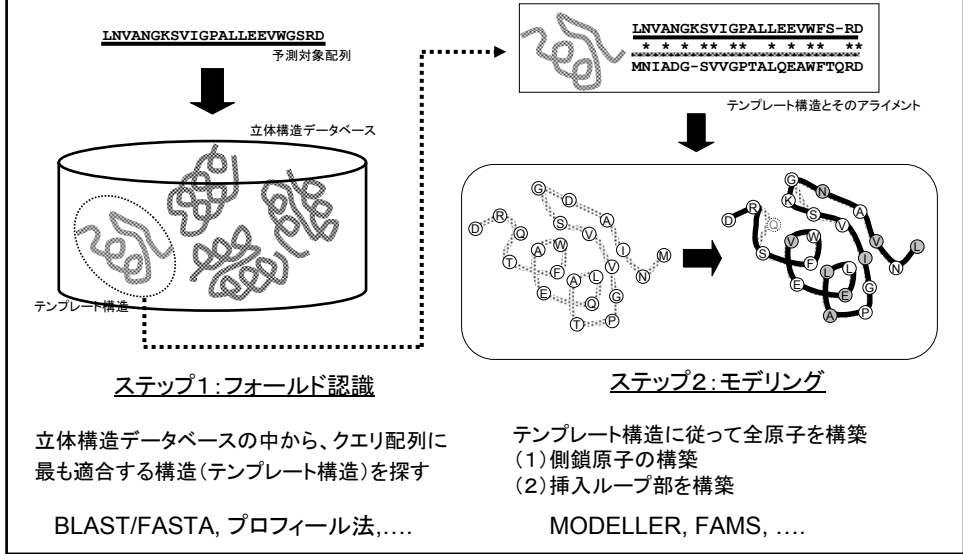
Baker グループ (U.Washington) <http://depts.washington.edu/bakerpg/newindex.html>

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。

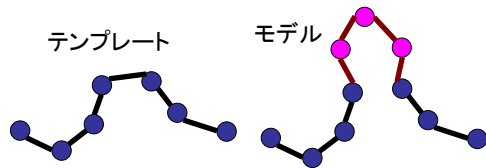


## モデリング

テンプレート構造を元にした全原子の構築  
(MODELLER, FAMS)

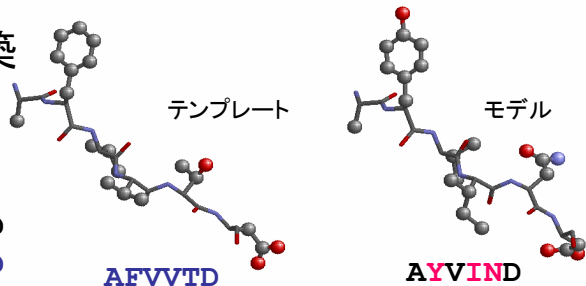
### (1) ループの構築

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



### (2) 側鎖原子の構築

Sequence A**YV**IND  
Structure AFV**V**TD

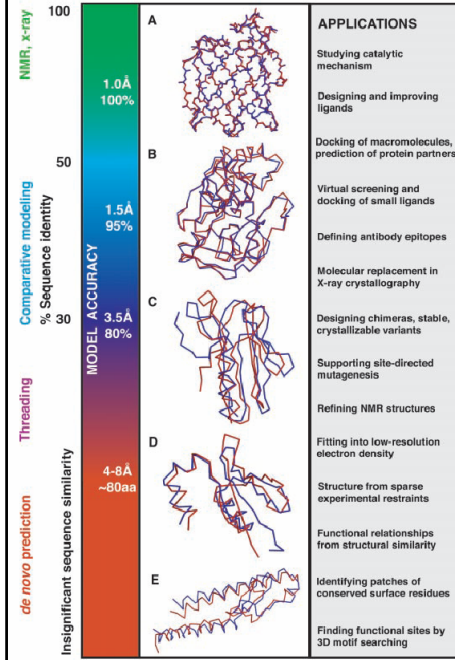


MODELLER : <http://www.salilab.org/modeller/modeller.html>

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

# モデリングした構造の精度と用途

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

[分子置換法による精密化]

SeqID = 30 %

部位特異的置換のサポート

[NMRの精密化]

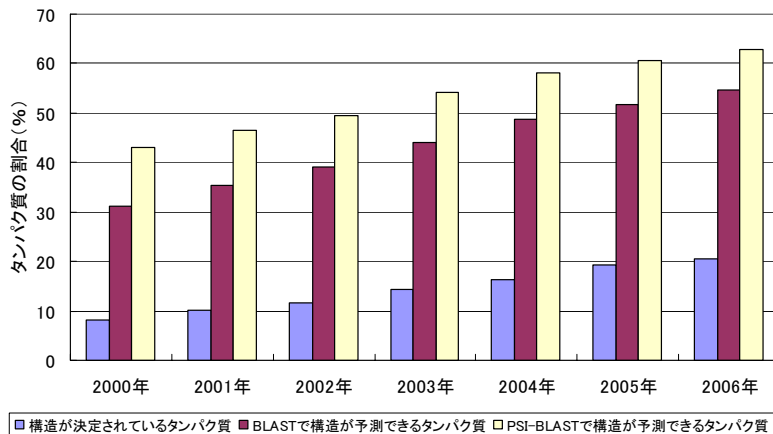
[電顕等の粗い電子密度へのフィット]

*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。

# 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

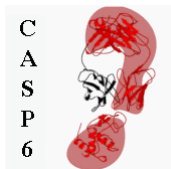
- John Moultらが主催で1994年に開始。2年おきに開催。
- 2004年にCASP6が開催。世界各国から 224チーム、65サーバが参加。
- 予測の手法・難易度によっていくつかのカテゴリに分けて審査

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)

## 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くために」 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・榊佳之 「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