

近畿大学・農学部・生命情報学

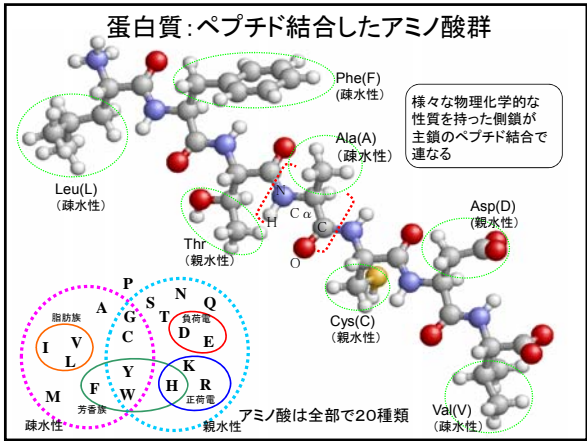
## タンパク質配列の分類 と機能推定

2008年5月27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>

## タンパク質の物理化学的性質を 利用した分類



### フォールディング(折り畳み)という現象

温度や変性剤濃度によって可逆に変化

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

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??

### タンパク質の構造・性質による大きな分類

		タンパク質の例	配列上の特徴
水溶性タンパク質 soluble protein	球状タンパク質 globular protein	多くの酵素、抗体、転写因子など	疎水性、親水性のアミノ酸がバランスよく含まれている。
	不定形タンパク質 Intrinsically disordered protein	真核生物の転写因子の非ドメイン領域など	親水性アミノ酸が多い。特にEPQSRKが多い。
膜タンパク質 membrane protein		レセプター、トランスポーターなど	疎水性アミノ酸が多く、連続する疎水性アミノ酸領域(膜貫通ヘリックス)が観察されることが多い。

### 球状タンパク質の疎水性アミノ酸

球状タンパク質では、分子内部に疎水性アミノ酸、分子表面に親水性アミノ酸が分布する傾向にある。

**疎水基：ACILMFVW**  
**親水基：RNDEQGHKPSY**

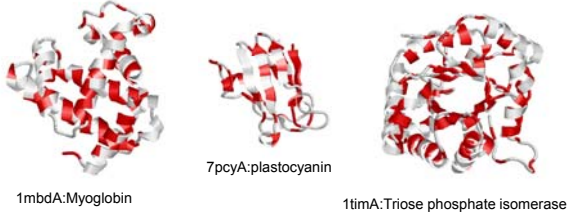
1mbdA:Myoglobin      7pcyA:plastocyanin      1timA:Triose phosphate isomerase

**疎水性相互作用(hydrophobic interaction)：**  
水分子と親和性の少ない非極性(non-polar、疎水性)基が水溶液中で互いに集まろうとする相互作用。電気や磁気と異なり符号はなく、疎水性どうしのは相手を問わず集合する。

## 球状タンパク質の疎水性アミノ酸

球状タンパク質では、分子内部に疎水性アミノ酸、分子表面に親水性アミノ酸が分布する傾向にある。

疎水基：ACILMFVW  
親水基：RNDEQGHKPSY



疎水性相互作用(hydrophobic interaction)：  
水分子と親和性の少ない非極性(non-polar、疎水性)基が水溶液中で互いに集まろうとする相互作用。電気や磁気と異なり符号はなく、疎水性どうしものは相手を問わず集合する。

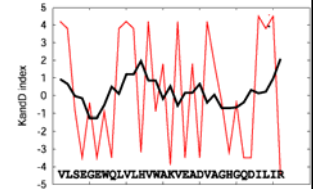
## 疎水性指標による内外予測

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

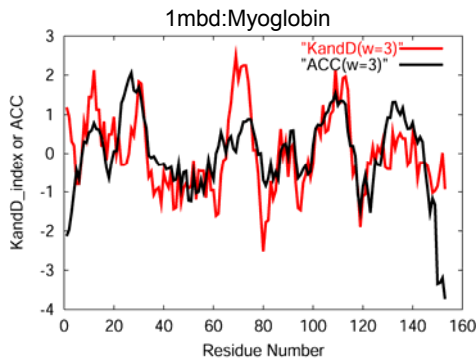
方法  
・配列を横軸にとつて、縦軸に対応する疎水性指標をプロット。  
・前後数残基でスムージング。

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



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

## 予測結果と実際の埋もれ度との比較



## 膜タンパク質

膜タンパク質：  
脂質二重膜に埋没して機能するタンパク質

性質  
(1) 疎水性な脂質尾部と相互作用するため、タンパク質表面には疎水性アミノ酸が多い。  
(2) 水に溶けにくく、凝集しやすい。  
(3) 多くは膜貫通ヘリックスを持つ  
(4) 膜貫通ヘリックスの長さは15-30アミノ酸ぐらい。アミノ酸配列からも連続した疎水性アミノ酸領域としてある程度、予測可能

## 不定形タンパク質 intrinsically disordered protein



親水性アミノ酸が多く、疎水性アミノ酸が少ないため、天然状態でもコンパクトに折たたまらない。

親水性アミノ酸の中でも、EPQSRK

球状タンパク質ドメインをつなぐ領域としてよく見られる。

真核生物に多く、原核生物には少ない。

転写因子に特に多くみられる。

生物学的な機能はよくわかっていないが、タンパク質間相互作用に重要な働きをするといわれている。

## タンパク質の構造・性質による大きな分類

		タンパク質の例	配列上の特徴
水溶性タンパク質 soluble protein	球状タンパク質 globular protein	多くの酵素、抗体、転写因子など	疎水性、親水性のアミノ酸がバランスよく含まれている。
	不定形タンパク質 Intrinsically disordered protein	真核生物の転写因子の非ドメイン領域など	親水性アミノ酸が多い。特にEPQSRKが多い。
膜タンパク質 membrane protein		レセプター、トランスポーターなど	疎水性アミノ酸が多く、連続する疎水性アミノ酸領域(膜貫通ヘリックス)が観察されることが多い。

アミノ酸疎水性による配列特徴解析: 学籍番号: \_\_\_\_\_ 氏名: \_\_\_\_\_

(1)以下の配列の疎水性アミノ酸 (ACFILMVW) を○で囲み、そのタンパク質のタイプを判定せよ。  
 (2)ヘリックスの周期的疎水パターン「XOOXOOX」、「XOOO×OO」、「×OO××OO」を口で囲め

>1mbdA (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)

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

>7pcyA (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)

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

>1bm1A (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)

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

>ZN428\_HUMAN (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)

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

生命情報学 (2008.5.27)

アミノ酸疎水性による配列特徴解析: 学籍番号: \_\_\_\_\_ 氏名: \_\_\_\_\_

(1)以下の配列の疎水性アミノ酸 (ACFILMVW) を○で囲み、そのタンパク質のタイプを判定せよ。  
 (2)ヘリックスの周期的疎水パターン「XOOXOOX」、「XOOO×OO」、「×OO××OO」を口で囲め

>1mbdA (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)

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

>7pcyA (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)

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

>1bm1A (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)

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

>ZN428\_HUMAN (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)

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

生命情報学 (2008.5.27)

アミノ酸疎水性による配列特徴解析: 学籍番号: \_\_\_\_\_ 氏名: \_\_\_\_\_

(1)以下の配列の疎水性アミノ酸 (ACFILMVW) を○で囲み、そのタンパク質のタイプを判定せよ。  
 (2)ヘリックスの周期的疎水パターン「XOOXOOX」、「XOOO×OO」、「×OO××OO」を口で囲め

>1mbdA (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)

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

>7pcyA (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)

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

>1bm1A (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)

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

>ZN428\_HUMAN (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)

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

生命情報学 (2008.5.27)

### 主鎖の水素結合: α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βシート

βヘアピン  
βターン

### 主鎖の水素結合: 平行βシート

β-α-βモチーフ  
平行βシートがαヘリックスで接続された構造モチーフ

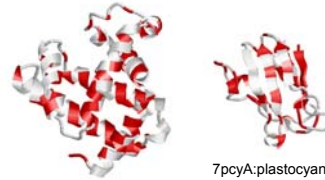
アミノ酸ごとに2次構造のなりやすさに差がある

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)

<b>αヘリックス</b>	<b>A</b> 1.42	<b>E</b> 1.37	<b>L</b> 1.33	<b>Q</b> 1.30	<b>M</b> 1.27
	<b>R</b> 1.21	<b>K</b> 1.16	<b>W</b> 1.07	<b>I</b> 1.06	<b>F</b> 0.97
	<b>Y</b> 0.96	<b>V</b> 0.92	<b>H</b> 0.87	<b>D</b> 0.82	<b>C</b> 0.80
	<b>T</b> 0.77	<b>S</b> 0.77	<b>N</b> 0.74	<b>G</b> 0.43	<b>P</b> 0.42
<b>βシート</b>	<b>V</b> 1.95	<b>I</b> 1.77	<b>F</b> 1.46	<b>Y</b> 1.45	<b>C</b> 1.30
	<b>W</b> 1.29	<b>T</b> 1.20	<b>L</b> 1.11	<b>M</b> 1.01	<b>H</b> 0.99
	<b>R</b> 0.90	<b>S</b> 0.83	<b>K</b> 0.79	<b>Q</b> 0.77	<b>A</b> 0.76
	<b>E</b> 0.70	<b>G</b> 0.64	<b>N</b> 0.61	<b>D</b> 0.53	<b>P</b> 0.42
<b>コイル</b>	<b>P</b> 1.66	<b>G</b> 1.55	<b>N</b> 1.35	<b>D</b> 1.33	<b>S</b> 1.23
	<b>H</b> 1.09	<b>T</b> 1.07	<b>C</b> 1.01	<b>K</b> 0.98	<b>R</b> 0.90
	<b>Q</b> 0.90	<b>E</b> 0.88	<b>Y</b> 0.84	<b>W</b> 0.83	<b>F</b> 0.82
	<b>M</b> 0.81	<b>A</b> 0.81	<b>L</b> 0.72	<b>V</b> 0.65	<b>I</b> 0.62

αヘリックスを好むアミノ酸上位5位までAEQLMを彩色すると、

>1mbdA (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)  
 V L S E G E W Q L V L H V W A K V E A D V A G H G Q D I L I R L F K S H P  
 E T L E K F D R F K H L K T E A E M K A S E D L K K H G V T V L T A L G A  
 I L K K K G H H E A E L K P L A Q S H A T K H K I P I K Y L E F I S E A I  
 I H V L H S R H P G D F G A D A Q G A M N K A L E L F R K D I A A K Y K E  
 L G Y Q G  
 >7pcyA (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)  
 A A I V K L G G D D G S L A F V P N N I T V G A G E S I E F I N N A G F P  
 H N I V F D E D A V P A G V D A D A I S A E D Y L N S K G Q T V V R K L T  
 T P G T Y G V Y C D P H S G A G M K M T I T V Q

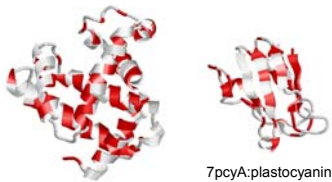


1mbdA:Myoglobin

7pcyA:plastocyanin

αヘリックスを好むアミノ酸上位5位までAEQLMを彩色すると、

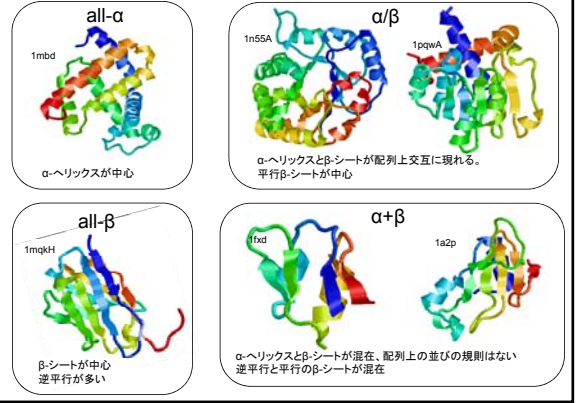
>1mbdA (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)  
 V L S E G E W Q L V L H V W A K V E A D V A G H G Q D I L I R L F K S H P  
 E T L E K F D R F K H L K T E A E M K A S E D L K K H G V T V L T A L G A  
 I L K K K G H H E A E L K P L A Q S H A T K H K I P I K Y L E F I S E A I  
 I H V L H S R H P G D F G A D A Q G A M N K A L E L F R K D I A A K Y K E  
 L G Y Q G  
 >7pcyA (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)  
 A A I V K L G G D D G S L A F V P N N I T V G A G E S I E F I N N A G F P  
 H N I V F D E D A V P A G V D A D A I S A E D Y L N S K G Q T V V R K L T  
 T P G T Y G V Y C D P H S G A G M K M T I T V Q



1mbdA:Myoglobin

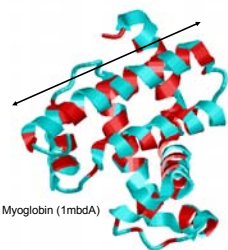
7pcyA:plastocyanin

球状タンパク質の構造クラス



疎水性の車輪図(Helical Wheel)

最後のヘリックスの配列: **A**D**A**Q**G**A**M**N**K**A**L**E**L**F**R**K**D**I**A**A**K**Y**K**E**L**



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  
 ヘリックスの1回転の周期が3.6アミノ酸なので、  
 3から4アミノ酸周期の疎水性 → ヘリックス構造を示唆

2次構造ごとに疎水性パターンに特徴がある

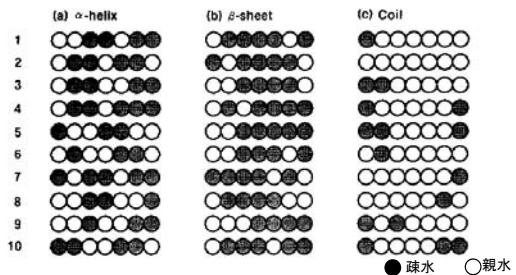


Fig. 9. Binary words that prefer the α helix (a), the β sheet (b), and the coil (c). They are ranked by the value of the binary word information. The top 10 are shown.

Kawabata, T. and Doi, J. (1997) "Improvement of Protein Secondary Structure Prediction Using Binary Word Encoding", Proteins, Vol 27, pp. 36-46



### 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

### 配列比較によるタンパク質の分類

### 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!!

### 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://www.sanger.ac.uk/Software/Pfam/>) からの引用

### モチーフ・プロフィールを用いた類似性

より大きなグループ(スーパーファミリー)にまとめようとした場合、弱い相同性をより正確に認識できる類似性を採用する必要

→ 近縁の配列群のマルチプルアライメントから、このファミリーの本質的な特徴を見出したい

```

5p21- MTEYKLVVVGAGGVGKSALTIQLIQNHVFVDEYDPTIEDSY
1ctqA MTEYKLVVVGAGGVGKSALTIQLIQNHVFVDEYDPTIEDSY
1c1yA MREYKLVVVGSGGVGKSALTVQFVQGFVEKYDPTIEDSY
1kao- MREYKVVVLGSGGVGKSALTVQFVGTGTFIEKYDPTIEDFY
1huqA --QFKLVLLGESAVGKSSLVLRVFKGQFHEYQESTIGAAF
1g16A ----KILLIGDSGVGKSCLLVRFVE---DKFNPID--DFK
1ek0A VTSIKLVLLGEAAVGKSSIVLRFVSNDFAEKNKEPTIGAAF
3rabA ---FKILIIQNSVVGKTSFLFRYADDSFTPAFVSTVGIDF
1mh1- ----KCVVVG DGAVGKTCLLISYTTNAPGGEIYPTVFDNY
2ngrA MQTIKCVVVG DGAVGKTCLLISYTTNKFPPSEYVPTVFDNY
1tx4B ----KLVI VGDGACGKTCLLIVNSKQDF---YVPTVFENY
  
```

サイトごとに保存の度合いに差がある。  
サイトごとにアミノ酸の出現傾向に差がある [AG]-x(4)-G-K-[ST]

## モチーフ解析

- ・正規表現風のパターンで、局所的な配列のパターンを表現。

PROSITE(<http://www.expasy.ch/prosite/>)が有名

### 1.進化的に保存している局所配列パターン

- ・マルチプルアライメント由来
- ・保存しているサイト→機能的に重要なサイト→活性部位

### 2.機能的な局所配列パターン

- ・リン酸化サイト、N-ミリスチル化サイトなど

## PROSITEのモチーフの記述法

(例)

ATP\_GTP\_A : [AG]-x(4)-G-K-[ST]

2FE2S FERREDOXIN:

C-{C}-{C}-[GA]-{C}-C-[GAST]-{CPDEKRHFYW}-C

ZINC\_FINGER\_C2H2\_1:

C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H

**x** : 任意のアミノ酸  
**x(n)** : n個の任意のアミノ酸  
**x(n,m)** : nからm個の任意のアミノ酸  
**[ACD]** : AかCかDのいずれかのアミノ酸  
**{ACD}** : AでもCでもDでもないアミノ酸

## PROSITEのモチーフの例(1)

ATP\_GTP\_A : [AG]-x(4)-G-K-[ST]

**x** : 任意のアミノ酸  
**x(n)** : n個の任意のアミノ酸  
**x(n,m)** : nからm個の任意のアミノ酸  
**[ACD]** : AかCかDのいずれかのアミノ酸  
**{ACD}** : AでもCでもDでもないアミノ酸

>5p21-

```
MTEYKLVVVGAGGVGKSALTIQLIQNHFVD
EYDPTIEDSYRKQVVIDGETCLLDILDITAG
QEEYSAMRDQYMRGTGEGLFCVFAINNTKSF
EDIHQYREQIKRVKDSDDVPMVLVGNKCDL
AARTVESRQAQDLARSYGIPYIETSAKTRQ
GVEDAFYTLVREIRQH
```

## PROSITEのモチーフの例(1)

ATP\_GTP\_A : [AG]-x(4)-G-K-[ST]

**x** : 任意のアミノ酸  
**x(n)** : n個の任意のアミノ酸  
**x(n,m)** : nからm個の任意のアミノ酸  
**[ACD]** : AかCかDのいずれかのアミノ酸  
**{ACD}** : AでもCでもDでもないアミノ酸

>5p21-

```
MTEYKLVVVGAGGVGKSALTIQLIQNHFVD
EYDPTIEDSYRKQVVIDGETCLLDILDITAG
QEEYSAMRDQYMRGTGEGLFCVFAINNTKSF
EDIHQYREQIKRVKDSDDVPMVLVGNKCDL
AARTVESRQAQDLARSYGIPYIETSAKTRQ
GVEDAFYTLVREIRQH
```

## PROSITEのモチーフの例(2)

ZINC\_FINGER\_C2H2\_1:

C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H

**x** : 任意のアミノ酸  
**x(n)** : n個の任意のアミノ酸  
**x(n,m)** : nからm個の任意のアミノ酸  
**[ACD]** : AかCかDのいずれかのアミノ酸  
**{ACD}** : AでもCでもDでもないアミノ酸

>ZN428\_HUMAN (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)

```
MTETREPAETGGYASLEEDDEDLSPGPEHSSDSEYTL
SEPDSEEEEDDEEEEEEETDDPEYDPPGYKVKQRLGGG
RGGPSRRAPRAAQPAPQPCQLCGRSPLGEAPPPTPPC
RLCCPATAPQEAAPAEGRALGEEEEEPPRAGEGRPAAG
REEEEEEDEEGTYHCTECEDSFDNLGELHGHFMLHAR
GEV
```

## PROSITEのモチーフの例(2)

ZINC\_FINGER\_C2H2\_1:

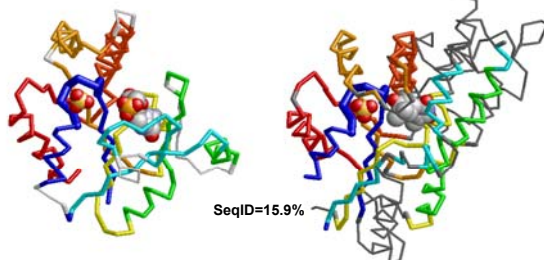
C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H

**x** : 任意のアミノ酸  
**x(n)** : n個の任意のアミノ酸  
**x(n,m)** : nからm個の任意のアミノ酸  
**[ACD]** : AかCかDのいずれかのアミノ酸  
**{ACD}** : AでもCでもDでもないアミノ酸

>ZN428\_HUMAN (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)

```
MTETREPAETGGYASLEEDDEDLSPGPEHSSDSEYTL
SEPDSEEEEDDEEEEEEETDDPEYDPPGYKVKQRLGGG
RGGPSRRAPRAAQPAPQPCQLCGRSPLGEAPPPTPPC
RLCCPATAPQEAAPAEGRALGEEEEEPPRAGEGRPAAG
REEEEEEDEEGTYHCTECEDSFDNLGELHGHFMLHAR
GEV
```

P-loopモチーフ: [AG]-x(4)-G-K-[ST] の立体構造



SeqID=15.9%

1gky:Guanilate Kinase  
(8-15:GPSGTGKS)

1e2kA:Thymidine Kinase  
(56-63:GPHGMGKT)

- ・ P-loopモチーフは、スクレオチドのリン酸基結合サイトに対応
- ・ モチーフ以外の領域も、立体構造は似ている

ProSiteモチーフの問題点

False positiveが多く、ファミリの認識能力は高くない。

[AG]-x(4)-G-K-[ST]

```

5p21- MTEYKLVVVVGAGGVGKSAL
1ctqA MTEYKLVVVVGAGGVGKSAL
1c1yA MREYKLVVVGSGGVGKSAL
1kao- MREYKVVVLGSGGVGKSAL
1huqA --QFKLVLLGESAVGKSSL
1g16A ----KILLIGDSGVGKSL
1ek0A VTSIKLVLLGEAAVGKSSI
3raba ---FKLLIIGNSSVGTSTF
1mh1- ----KCVVVDGAVGKTC
2ngrA MQTIKVVVDGAVGKTC
1tx4B ----KLIVVDGACGKTC
1i2mA --QFKLVLDGDTGKTTTF
2efgA -RLRNIGIAAHIDAGKTTT
    
```

1. パターンの表現能力の限界
2. 客観的にパターンを生成するのが難しい。
3. もっと大域的な領域も淡く似ているはず

プロフィール法

マルチプルアライメントからサイトごとのスコア行列を作成。  
これに対して動的計画法等を用いて配列をアライメント。

サイトごとのスコア行列

↓  
プロフィール(Profile)  
位置特異的スコア行列

(PSSM; Position Specific Score Matrix)

	1	2	3	4	5	6	..
A	3	-1	-3	-4	6	-4	..
Q	0	3	-1	-2	-4	0	..
H	-3	-3	-4	11	-4	4	..
:	:	:	:	:	:	:	..
V	-4	-2	-1	-6	-2	-4	..

HMMer

マルチプルアライメントを入力とする。隠れマルコフモデル(HMM)を使用しているため、表現力はPSI-BLASTより高いはずだが、計算速度は遅い。PfamはHMMerを採用している。

PSI-BLAST

BLASTの拡張版。反復的にデータベース検索を行うことで、厚いマルチプルアライメントを生成する。

Site of query sequence

	1	2	3	4	5	6	7	8	9	..
query	A	Q	S	H	A	T	K	H	K	..
homolog1	A	N	S	H	A	T	K	H	K	..
homolog2	S	G	K	H	A	K	S	F	Q	..
homolog3	A	R	K	H	G	E	-	L	L	..
homolog4	S	D	L	H	A	H	-	L	R	..
homolog5	S	D	L	H	A	H	K	F	R	..

マルチプルアライメント

$$S(His, Ath) = \log \frac{P(His / Ath)}{P(His)}$$

Sites of query sequence

	1	2	3	4	5	6	7	8	9	..
A	3	-1	-3	-4	6	-4	-3	-4	-4	..
Q	0	3	-1	-2	-4	0	0	-4	0	..
G	-2	-1	-5	-5	-1	-4	-2	-6	-5	..
H	-3	-3	-4	11	-4	4	-3	6	6	..
I	-5	-3	-1	-6	0	-4	-2	-1	-5	..
:	:	:	:	:	:	:	:	:	:	..
V	-4	-2	-1	-6	-2	-4	-4	-2	-5	..

プロフィール  
(Score Table)

Pfam : 蛋白質ファミリのデータベース

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

各蛋白質のファミリの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

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」(2005)、第2章、第4章、南江堂
- ・ 松澤洋 (編集)「タンパク質工学の基礎」(2004)、東京化学同人
- ・ C.Branden & J.Toozee (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
- ・ G.A.Petsko & D.Ringe (横山茂之監訳)「タンパク質の構造と機能 ゲノム時代のアプローチ」(2005)、メディカル・サイエンス・インターナショナル
- ・ Arthur M.Lesk (高木淳一訳)「ポストゲノム時代のタンパク質科学 -構造・機能・ゲノミクス」(2007)化学同人

構造バイオインフォマティクス

- ・ 藤博幸 編 「はじめてのバイオインフォマティクス」(2006) 講談社サイエンティフィック
- ・ 郷道子、高橋健一 編 「基礎と実習 バイオインフォマティクス」(2004)、共立出版
- ・ 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に」(2003)、メディカル・サイエンス・インターナショナル
- ・ 美宅成樹・楠佳之「バイオインフォマティクス」(2003)、第6、7章、東京化学同人
- ・ 日本バイオインフォマティクス学会編「バイオインフォマティクス事典」第10章、共立出版(2006)

配列の分類・モチーフ・プロフィール法

- ・ 美宅成樹・楠佳之「バイオインフォマティクス」(2003)、第3、4章、東京化学同人
- ・ 日本バイオインフォマティクス学会編「バイオインフォマティクス事典」第7、8章、共立出版(2006)