

近畿大学・農学部・生命情報学

蛋白質の物理化学的性質と アミノ酸配列解析

2009年4月28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21年度「生命情報学&生命情報学実習」講義日程

2009.4.14

	講義	生命情報学	演習	生命情報学演習
4/7	川端1	配列決定とバイオインフォマティクス		
4/14	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データベースの使用法(BLAST)
4/21	川端3	マルチプルアライメントとその応用	中村	ChemOfficeを用いた計算化学演習
4/28	川端4	蛋白質の物理化学的性質と配列解析		
5/12	川端5	分子系統学基礎	中村	系統樹作成演習(ClustalX)
5/19	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(RasMol)
5/26	川端7	>>試験<<		
6/2	金谷1	ポストゲノム解析入門(トランスクリプトーム解析)		
6/9	金谷2	ポストゲノム解析入門(インタラクトーム解析)	金谷1	発現プロフィール解析演習
6/16	金谷3	ポストゲノム解析(統合解析)	金谷2	インタラクトーム・代謝物解析演習
6/23	金谷4	メタボローム解析(その1)		
6/30	金谷5	メタボローム解析(その2)		
7/7	金谷6	メタボローム解析(その3)		
7/14	金谷7	>>試験<<		

配列比較によるタンパク質の分類

Pfam : 蛋白質ファミリーのデータベース

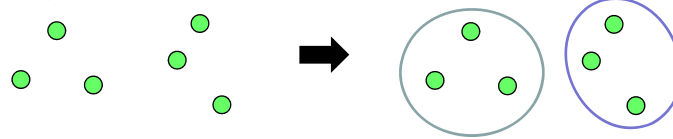
<http://pfam.sanger.ac.uk>
各蛋白質ファミリーのマルチプルアライメント、HMMなどを集めたデータベース

The screenshot displays the Pfam website interface. On the left, a table lists various protein families with columns for ID, Accession, Type, Number of Sequences (Seed and Full), Average length, Average fold, Average coverage, Hits, Change status, and Description. The 'zf-C2H2' family is highlighted. On the right, a detailed view of the 'Family: zf-C2H2 (PF00096)' is shown, including a summary, domain organization, alignments, and a 3D protein structure model.

ID	Accession	Type	Number of Sequences		Average length	Average fold	Average coverage	Hits	Change status	Description
			Seed	Full						
GP130	PF00516	Family	24	75199	375.5	54	87.21	✓	Changed	Envelope glycoprotein GP130
RYL_1	PF00078	Family	156	71335	185.3	67	39.98	✓	Changed	Reverse transcriptase (DNA-dependent DNA polymerase)
ABC_tran	PF00005	Domain	64	65707	184.2	26	37.89	✓	Changed	ABC transporter
Ark	PF00013	Repeat	1183	64674	30.6					
COX1	PF00115	Family	23	59394	232.7					
EVF	PF00077	Domain	30	54135	94.0					
LRB_1	PF00560	Repeat	2445	53866	22.7					
zLGN2	PF00096	Domain	196	52811	23.4					
Cytoschrom_B_H	PF00013	Repeat	8	49376	154.2					
WDR4	PF00490	Repeat	1893	45685	38.7					
TFS_1	PF00515	Repeat	562	37697	32.2					
Oxidored_d1	PF00261	Family	33	34594	313.3					

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。



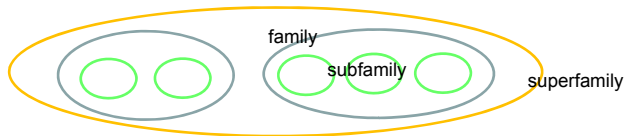
どのくらいの類似性で一つにまとめるかは様々な考えがあるが、少なくとも「**相同なタンパク質群**」を一つのグループにまとめるべき

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

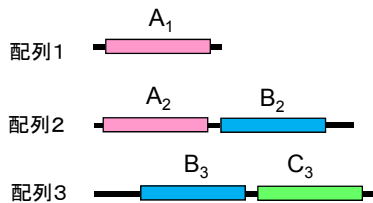
複数のファミリーがあつまった大きなグループ: **スーパーファミリー (superfamily, clan)**

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)**

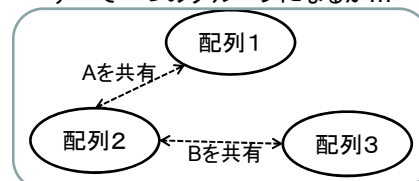


ドメイン単位による分類

配列の部分どうしだけが類似していた場合

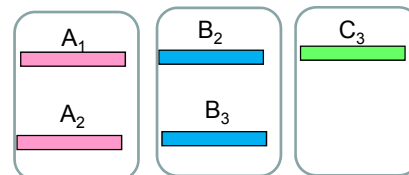
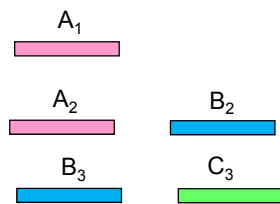


部分一致を重視すると、
すべて一つのグループになるが...

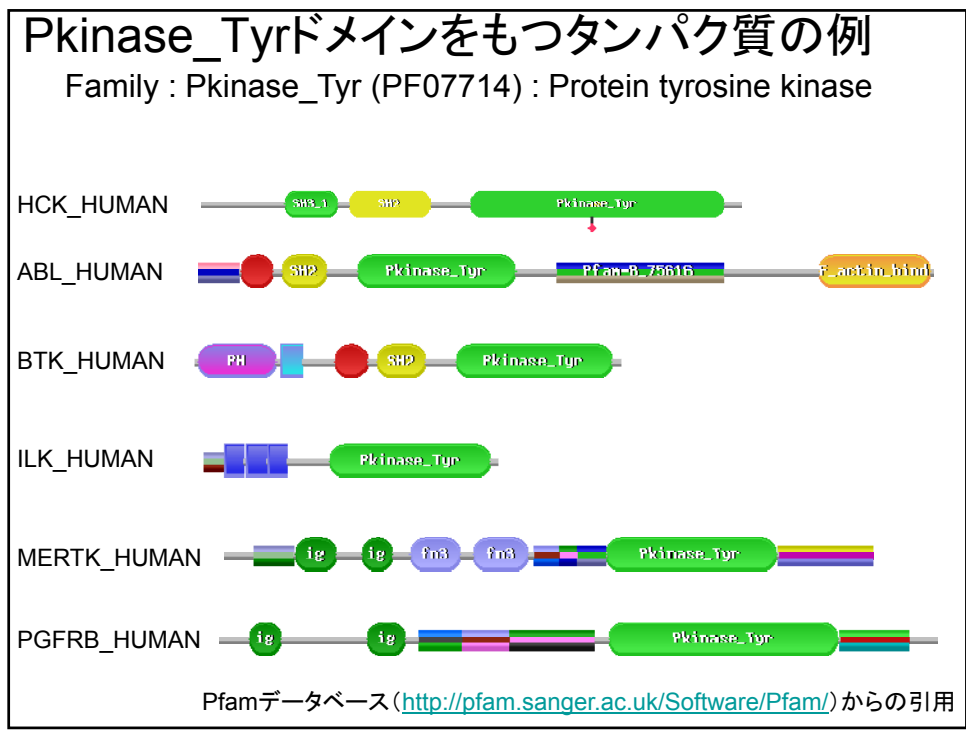


1と2が類似、2と3が類似していても
1と3は類似していない!

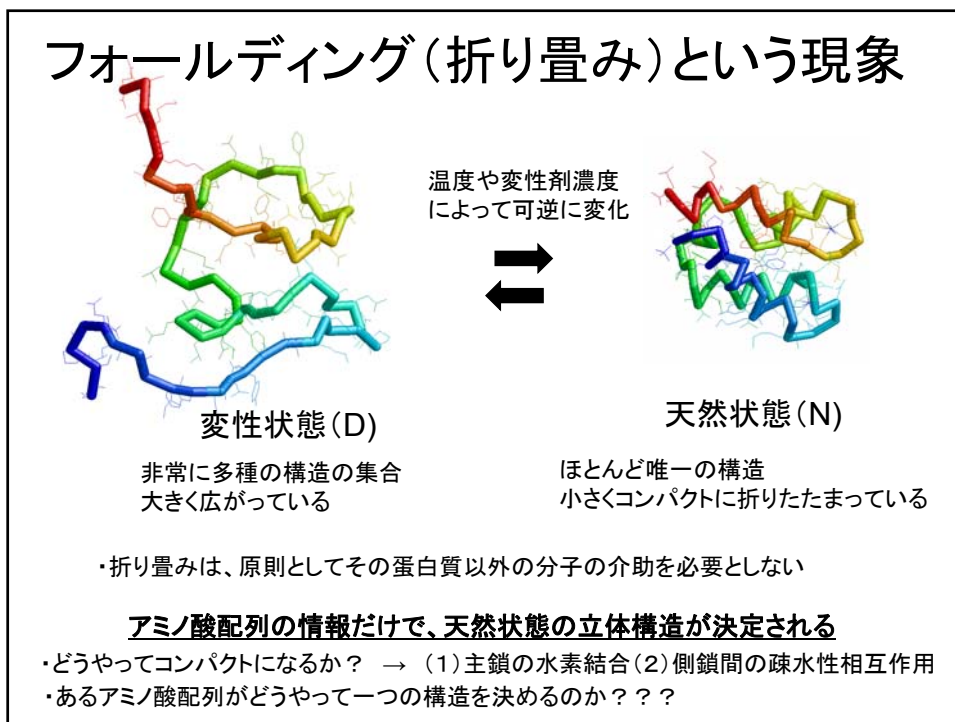
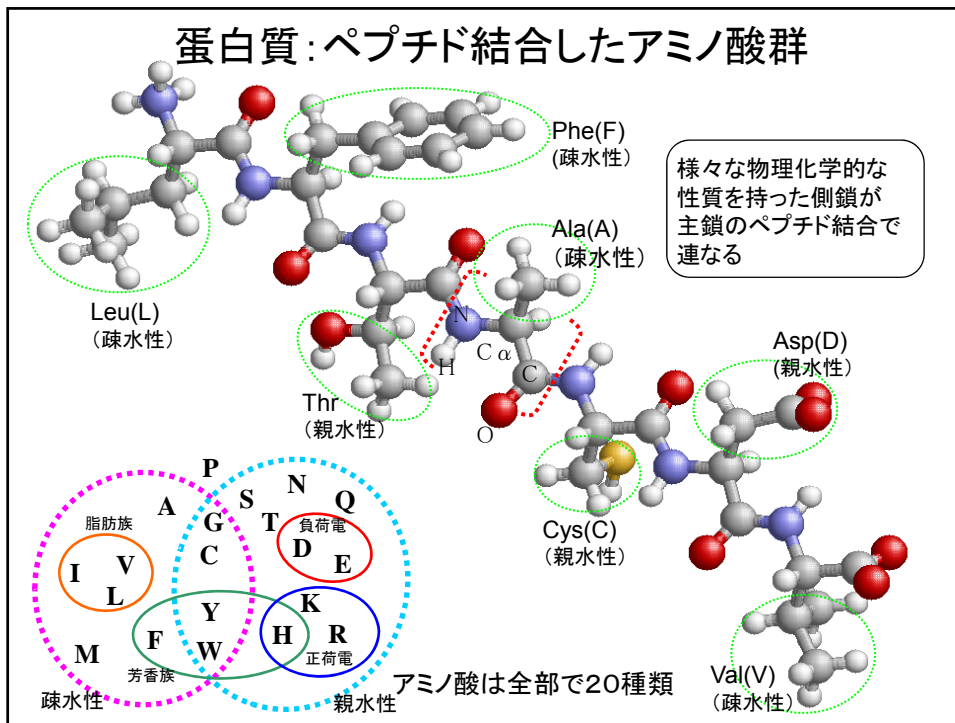
ドメインに分割すると



きれいに分類することができる!!



蛋白質の物理化学的性質と アミノ酸配列解析



アミノ酸の疎水性の分布による 蛋白質の特徴づけ

タンパク質の構造・性質による大きな分類

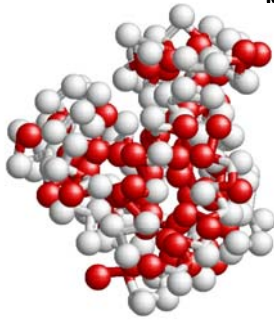
		タンパク質の例	配列上の特徴
水溶性 タンパク質 soluble protein	球状タンパク質 globular protein	多くの酵素、抗体、 転写因子など	疎水性、親水性のアミ ノ酸がバランスよく含 まれている。
	不定形タンパク質 Intrinsically disordered protein	真核生物の転写因 子の非ドメイン領域 など	親水性アミノ酸が多い。 特にEPQSRKが多い。
膜タンパク質 membrane protein		レセプター、トラン スポーターなど	疎水性アミノ酸が多く、 連続する疎水性アミノ 酸領域(膜貫通ヘリッ クス)が観察されること が多い。

球状タンパク質の疎水性アミノ酸

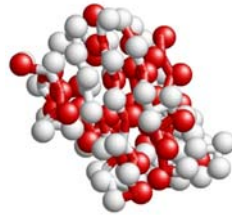
球状タンパク質では、分子内部に疎水性アミノ酸、分子表面に親水性アミノ酸が分布する傾向にある。

疎水基 : ACILMFVW

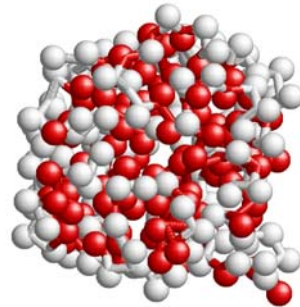
親水基 : RNDEQGHKPSTY



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 : RNDEQGHKPSTY



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。

疎水性指標による内外予測

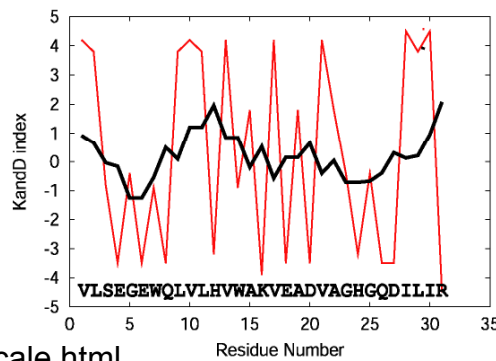
Kyte and Doolittle(1982)の疎水性指標

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。

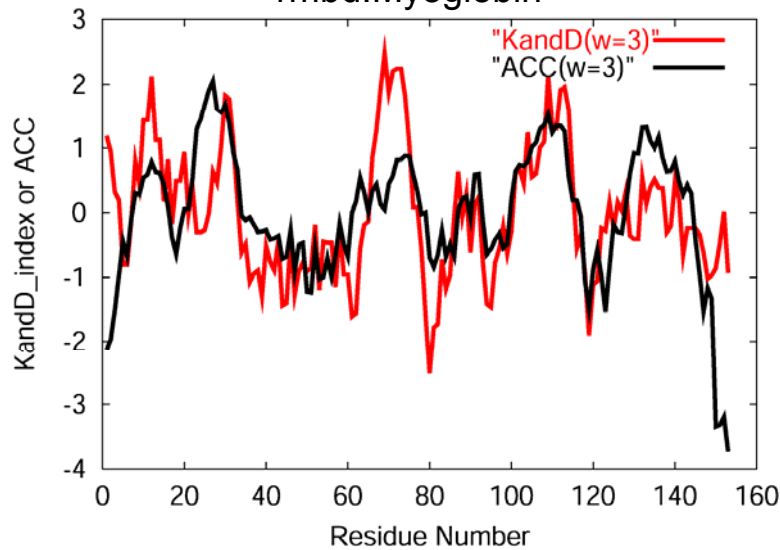
$$V[i] = \frac{1}{2w+1} \sum_{k=-w}^w KD\{Seq[i+k]\}$$



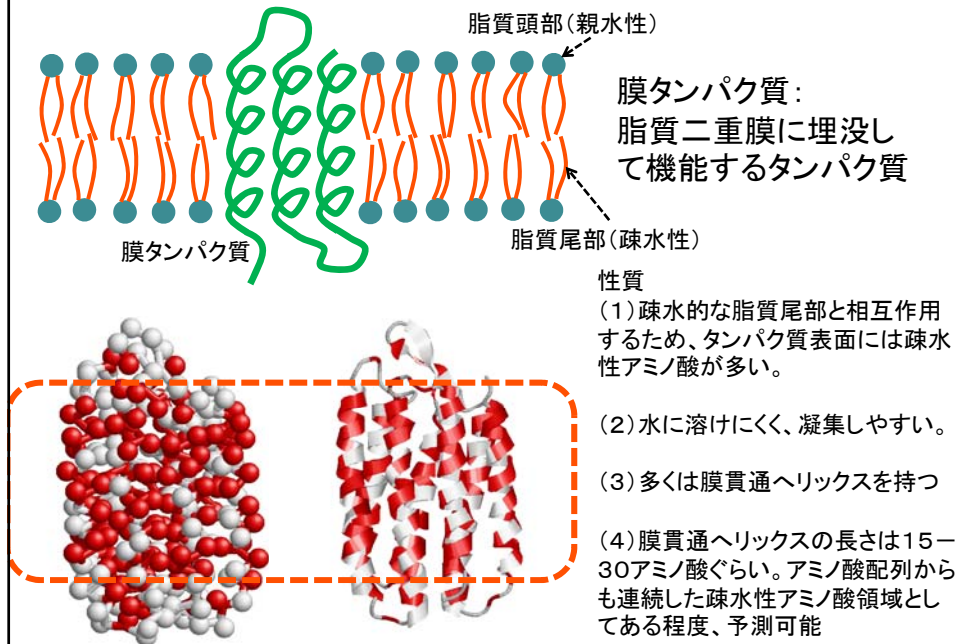
<http://kr.expasy.org/tools/protscale.html>

予測結果と実際の埋もれ度との比較

1mbd:Myoglobin

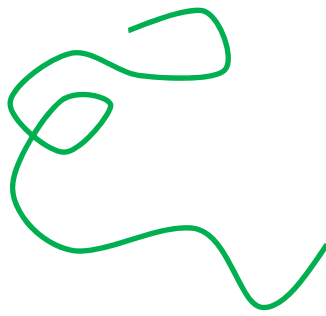


膜タンパク質



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「X○○X○○X」、「×○○○×○○」、「×○○××○○」を□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 W
 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 G 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 F
 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 D 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 E P P R A G E G R P A G
 R 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

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「X〇〇X〇〇X」, 「x〇〇〇x〇〇」, 「x〇〇x x〇〇」を□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 I 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 E 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 I Y F I 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 W
 A R Y A D W I F T T P I L L L D L A L L V D A D Q G T I L A L V G A D G 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 I F F
 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 I 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 D 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 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 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

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「X〇〇X〇〇X」, 「x〇〇〇x〇〇」, 「x〇〇x x〇〇」を□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 I 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 E 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 I Y F I 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 W
 A R Y A D W I F T T P I L L L D L A L L V D A D Q G T I L A L V G A D G 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 I F F
 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 I 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 D 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 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 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

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「X〇〇X〇〇X」, 「×〇〇〇×〇〇」, 「×〇〇××〇〇」を□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 I 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 E 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 L 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 I G T A L M G L G T I Y F I V K G M G V S D P D A K K F Y
 A T T T L V P A T A F T M Y L S M L I G Y G L T M V P F G G E Q N P I Y W
 A R Y A D W I F T T P L L L L D L A L L V D A D Q G T I L A L V G A D G I
 M T G T G L V G A L T K V Y S Y R F V W W A T S T A A M L Y I L Y V I F F
 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 I 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 I D V S A K V G F G L I L L R S R

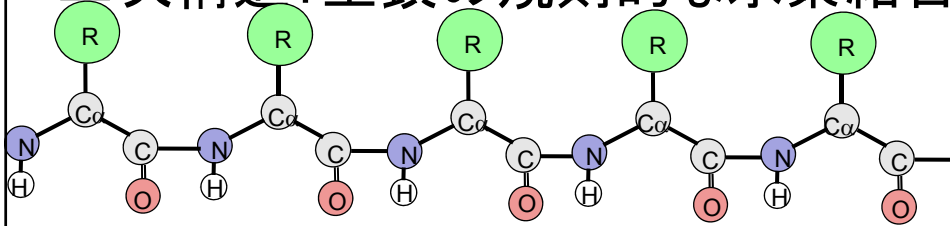
>ZN428_HUMAN (球状タンパク質・膜タンパク質・不定形タンパク質)(どれか一つを選択)
 M T E T R E P A E T G G Y A S L E E D D E D L S P G P E H S S D S E Y T I
 S E P D S E E E E E E E E E T T D D P E Y D P G Y K V K Q R L G G G
 R G G P S R R A P R A A Q P P A Q P C Q L C G R S P L G E A P P G T P P C
 R L C C P A T A P Q E A P A P E 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 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

アミノ酸配列から膜貫通ヘリックスを予測するサーバ

SOSUIサーバ
<http://bp.nuap.nagoya-u.ac.jp/sosui/>

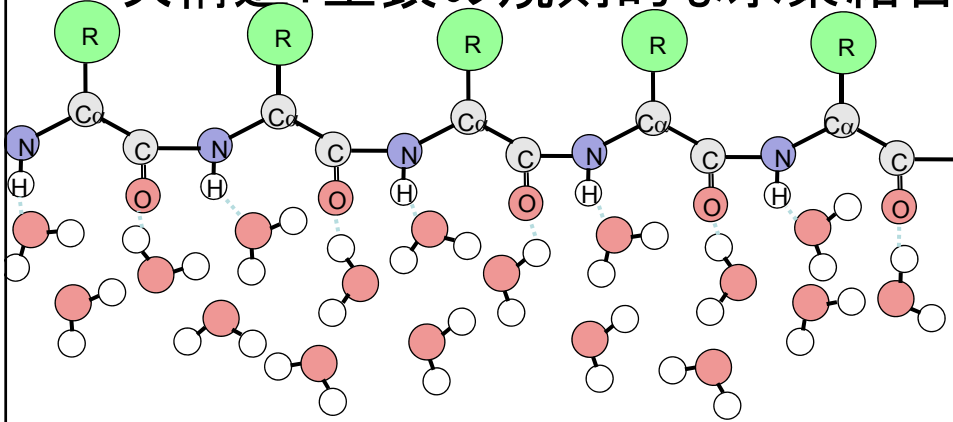
蛋白質の二次構造 (α ヘリックス、 β シート)

二次構造：主鎖の規則的な水素結合



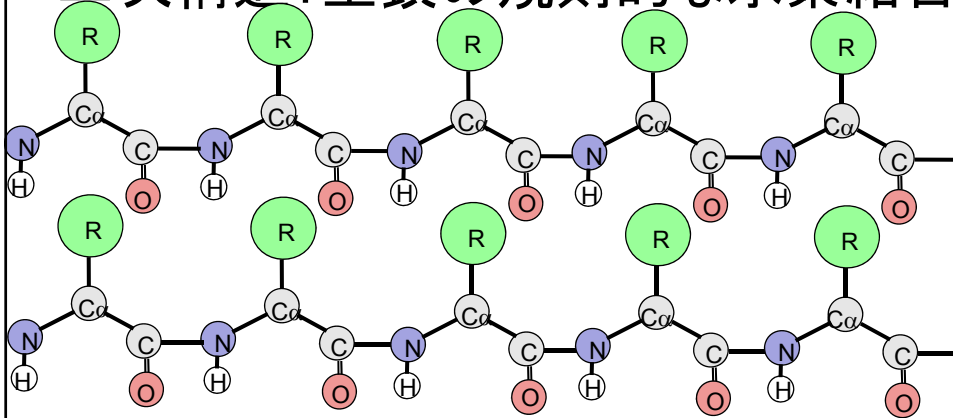
・主鎖はNH基とCO基の繰り返しであるため、極めて親水的

二次構造：主鎖の規則的な水素結合



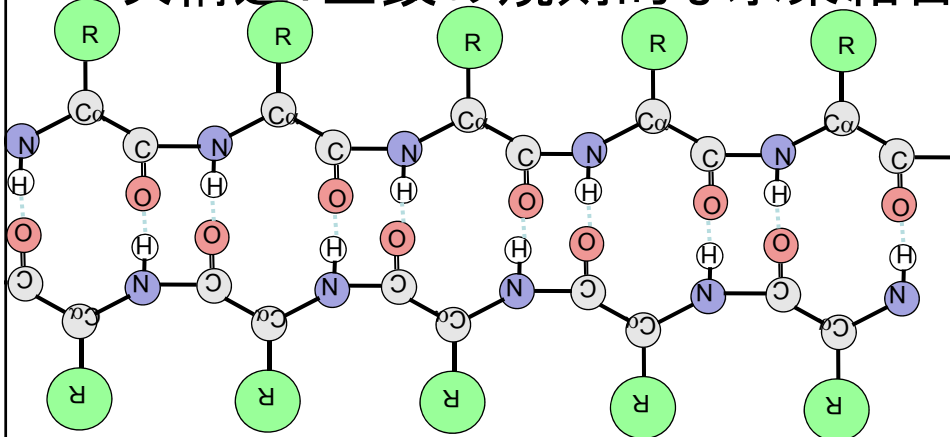
- ・主鎖はNH基とCO基の繰り返しであるため、極めて親水的
→ 水分子が水素結合しやすい

二次構造：主鎖の規則的な水素結合



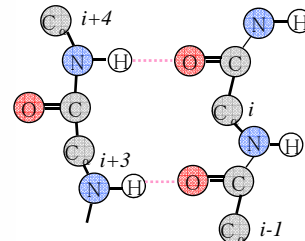
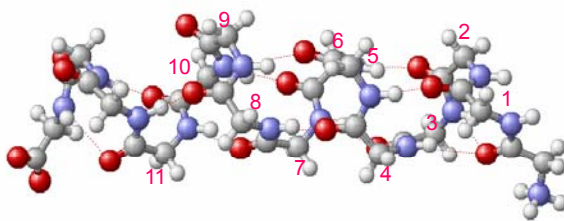
- ・主鎖はNH基とCO基の繰り返しであるため、極めて親水的
→ 水分子が水素結合しやすい
- ・折りたたまれ、水が排除されると水素結合は切断される
→ エネルギー的に不利

二次構造：主鎖の規則的な水素結合

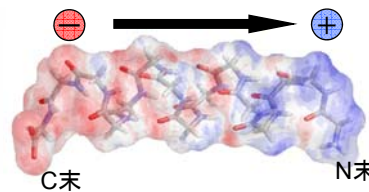


- ・主鎖はNH基とCO基の繰り返しであるため、極めて親水的
→ 水分子が水素結合しやすい
- ・折りたたまれ、水が排除されると水素結合は切断される
→ エネルギー的に不利
- ・蛋白質内で新たな水素結合を作る必要性 → 二次構造の生成

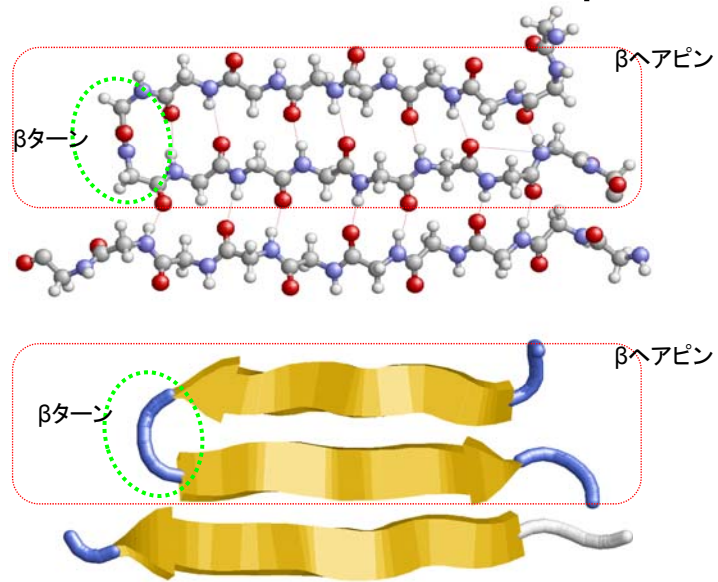
主鎖の水素結合： α ヘリックス



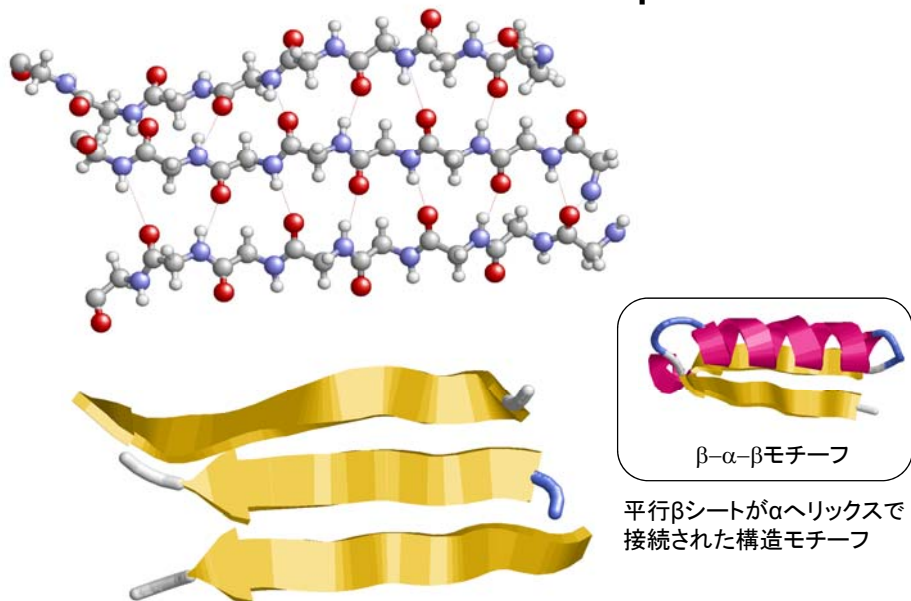
- ・ i 番目のCOが $i+4$ 番目のNHと水素結合を形成
- ・ 3.6残基が1周期の右巻きらせん
- ・ NH...O=Cが同じ向きに並ぶため、ヘリックス全体に電気双極子が発生



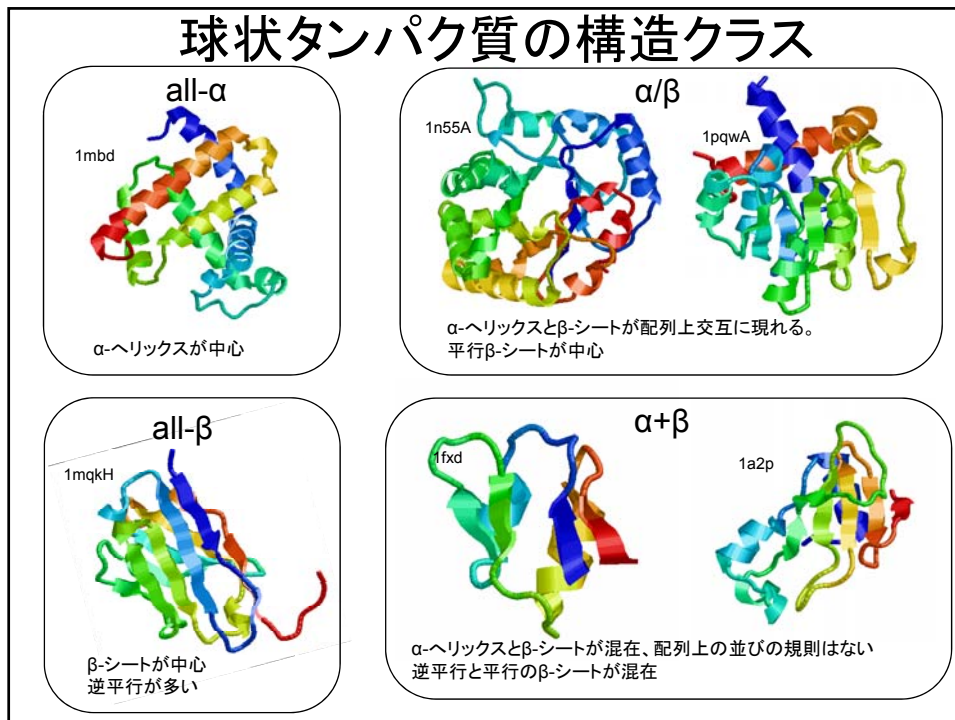
主鎖の水素結合：逆平行βシート



主鎖の水素結合：平行βシート



球状タンパク質の構造クラス



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)

α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

β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

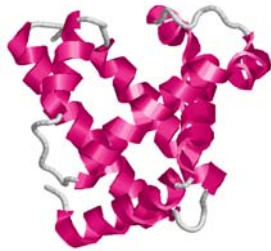
(3) α ヘリックスを好むアミノ酸AEQLMを○で囲みなさい。この二つがall- α かall- β のどちらかだとしたら、どちらに属すると考えられるか、○をつけて答えよ。

>1mbdA (all- α クラス、 all- β クラス) (どちらか一つを選択)

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 (all- α クラス、 all- β クラス) (どちらか一つを選択)

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



all- α



all- β

生命情報学 (2009.4.28)

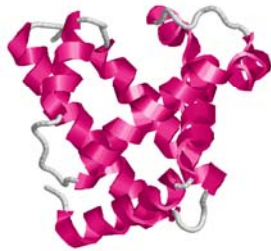
(3) α ヘリックスを好むアミノ酸AEQLMを○で囲みなさい。この二つがall- α かall- β のどちらかだとしたら、どちらに属すると考えられるか、○をつけて答えよ。

>1mbdA (all- α クラス、 all- β クラス) (どちらか一つを選択)

V L S E G 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 (all- α クラス、 all- β クラス) (どちらか一つを選択)

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



all- α



all- β

生命情報学 (2009.4.28)

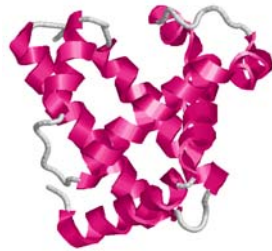
(3) α ヘリックスを好むアミノ酸AEQLMを○で囲みなさい。この二つがall- α かall- β のどちらかだとしたら、どちらに属すると思われるか、○をつけて答えよ。

>1mbdA (all- α クラス、all- β クラス) (どちらか一つを選択)

V L S **E** G **E** W **Q** **L** V **I** 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 (all- α クラス、all- β クラス) (どちらか一つを選択)

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**



all- α

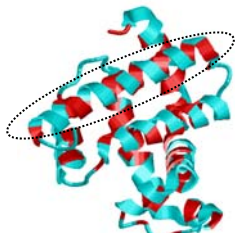


all- β

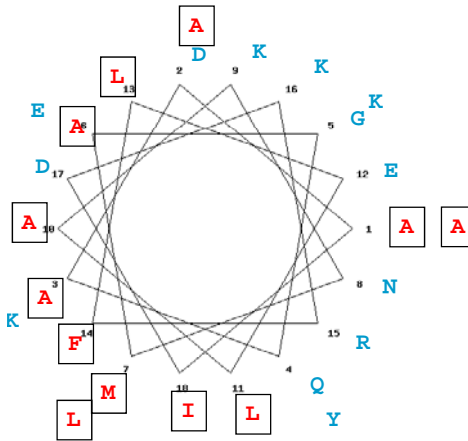
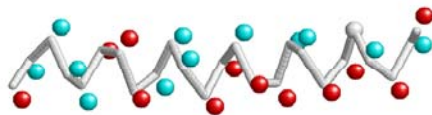
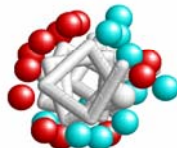
生命情報学 (2009.4.28)

疎水性の車輪図(Helical Wheel)

1mbdAの最後のヘリックスの配列: **ADAQGAMNKALELFRKDI AAKYKEL**



Myoglobin (1mbdA)



両親媒性ヘリックス : 片側が疎水的、反対の側が親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

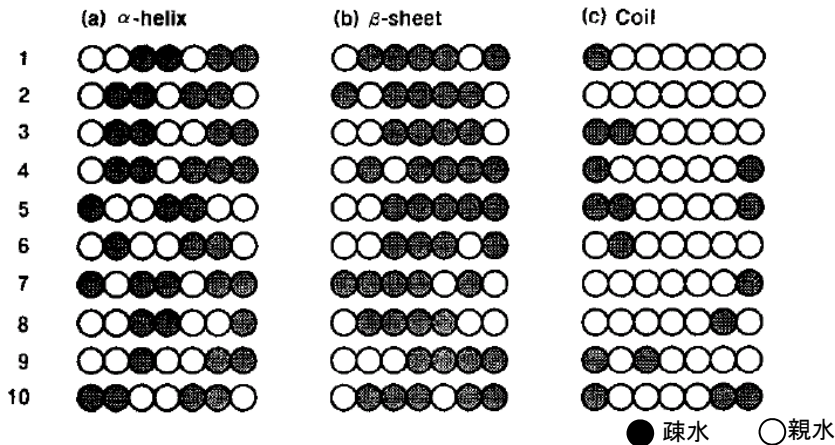


Fig. 9. Binary words that prefer the α helix (a), the β sheet (b), and the coil (c). They are ranked by the value of the binary word information. The top 10 are shown.

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

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」, 「xOOOxOO」, 「xOOxOO」を□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 I 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 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 W
 A R Y A D W I F T T P L L L L D L A L L V D A D Q G T I L A L V G A D G 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 F
 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 I 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 D 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 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 P A G
 R 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

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「X○○X○○X」, 「×○○○×○○」, 「×○○××○○」を□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)

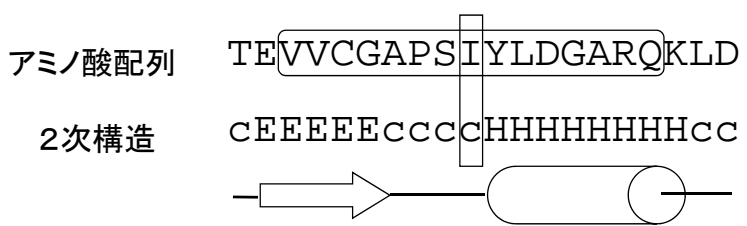
R P E W I W L A L G T A L M G L G T I Y F I V K G M G V S D P D A K K F Y
 A I T T L V P A I A F T M Y L S M L L G Y G L T M V P F G G E Q N P I Y W
 M R Y A D W I F T T P I L L L D I A L L V D A D Q G T I L A L V G A D G I
 M I G T G L V G A L T K V Y S Y R F V W W A I S T A A M L Y I L Y V I F F
 G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

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。



E
βシート



c
コイル



H
αヘリックス

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 α ヘリックスに比べ、 β 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

2次構造予測の例

Single : 1つの配列を入力とするニューラルネットワーク [Q3=66.6%]

Profile : 複数の配列を入力とするニューラルネットワーク [Q3=83.5%]

```
>ln55A [c.1.1.1] ISOMERASE A: - - - TRIOSEPHOSPHATE ISOMERASE
AAseq : AKQPPIAAANWKCNGTTASIEKLVQVFNEHTISHDVQCVVAPTFFVHIPLVQAKLRNPKYV
Single : ccccccccccccccccHHHHHHHHHHHHccccccEEEEccccccccHHHHcccccccH
Profile: ccccEEEEcHccccHHHHHHHHHHHHccccccccEEEEccHHHHHHHHHHHHccccE
Observe: cccccEEEEccccHHHHHHHHHHHHccccccccEEEEccccHHHHHHHHccccEE

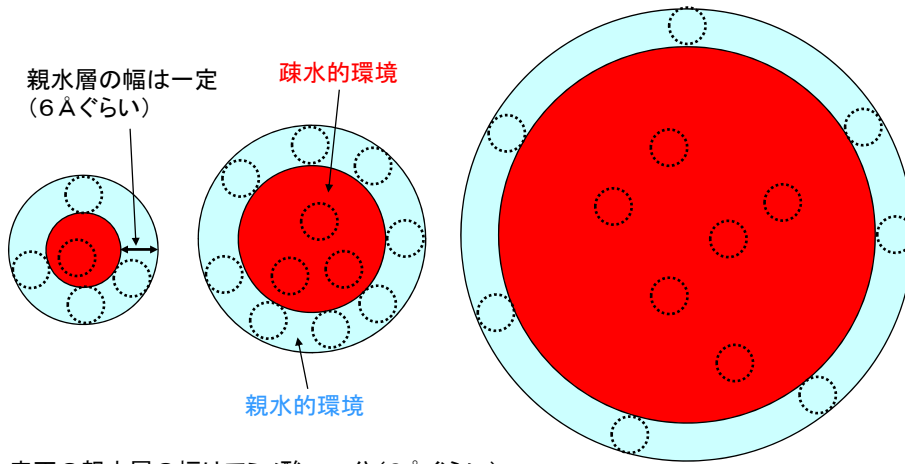
AAseq : ISAQNAIAKSGAFTGEVSMPIKDIGVHWVILGHSERRTTYGETDEIVAQKVSEACKQGF
Single : ccccccccccccccccHHHHHHHHHHHHccccccEEEEccccccccHHHHcccccccH
Profile: EEcccccccccccccccHHHHHHccccEEEEcccccccccccHHHHHHHHHHHHccc
Observe: EEEcccccccccccccccHHHHHHccEEEEccHHHHHHccccHHHHHHHHHHHHccc

AAseq : MVIACIGETLQOREANQTAKVVLSTSAIAAKLTKDAWNQVVLAYEPVWAIPTGKQVATPE
Single : EEEEEHHHHHHHHHHccccHEHHEHccHHHHHHHHHHHHHHHHccccccEEcccccccHH
Profile: cEEEEccccHHccccHHHHHHHHHHHHHHccccccccEEEEcccccccccccccccHH
Observe: EEEEEcccHHHHccccHHHHHHHHHHHHHHccccccccEEEEcccccccccccccccHH

AAseq : QAQEVHLLLRKQVSENIQTDVAALRLRILYGGSVNAANAATLYAKPDINGFLVGGASLKPE
Single : HHHHHHHHHHHHHHHccccHHHHHHHHHHccccHHHHHHHHcccccccccccccccccccH
Profile: HHHHHHHHHHHHHHHHHccccccccEEEEccccHHHHHHHHHHcccccccccccccccHH
Observe: HHHHHHHHHHHHHHHHHccccHHHHccccEEEEccccccccHHHHHHccccEEEEccccccc

AAseq : FRDIIDATR
```


タンパク質の大きさと疎水性の関係



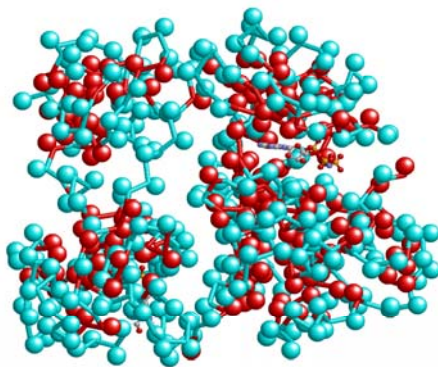
表面の親水層の幅はアミノ酸一つ分(6 Åぐらい)

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

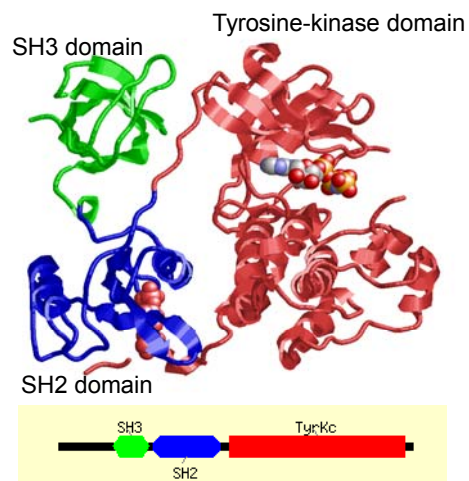
疎水性のアミノ酸が多すぎると、折り畳まる前に凝集して沈殿してしまう危険性

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



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

参考図書

構造生物学一般について

- Bruce 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.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
- 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
- 郷通子、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くために」(2003)、メディカル・サイエンス・インターナショナル
- 美宅成樹・榊佳之「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・榊佳之「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)