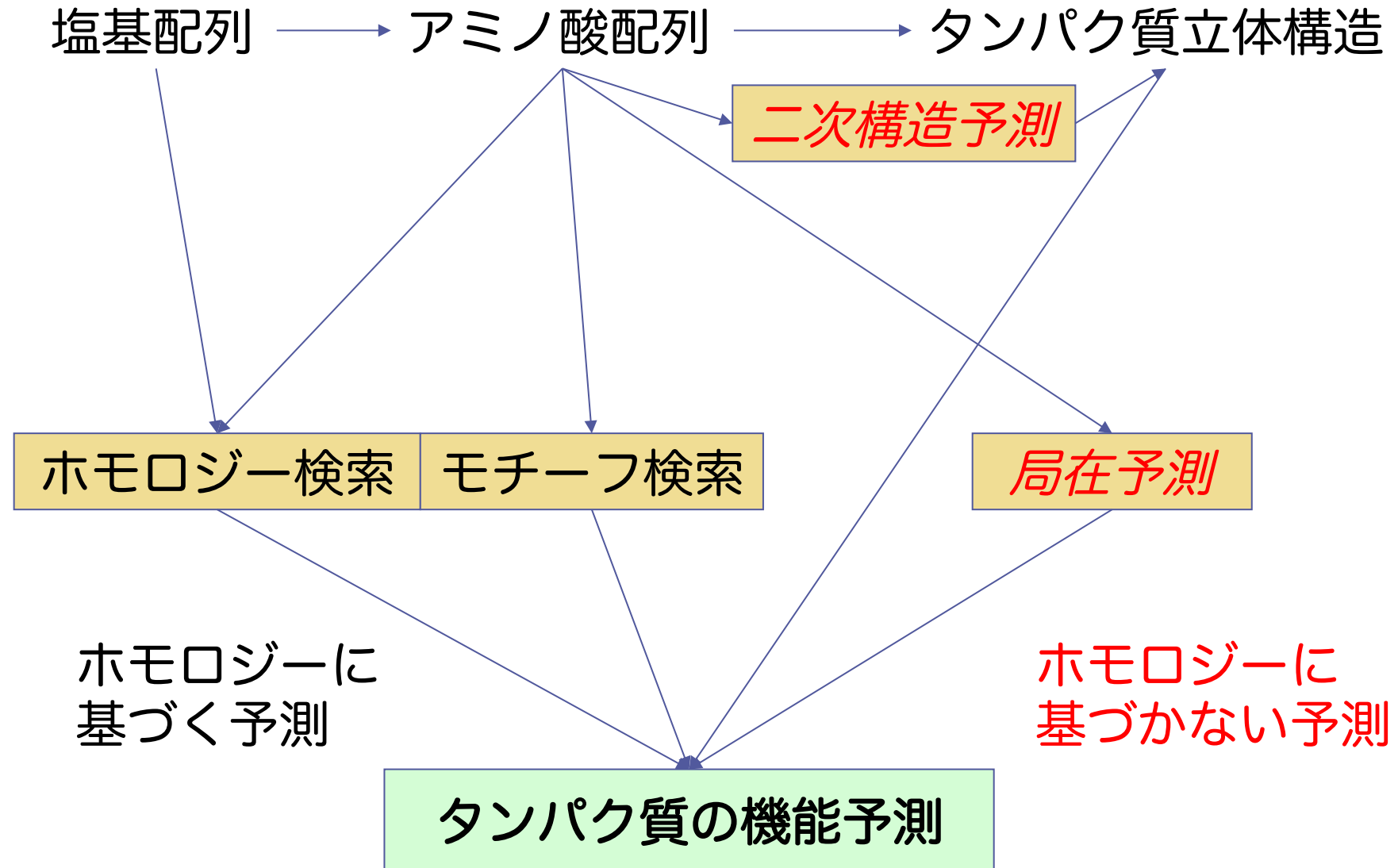


2016年度「バイオインフォマティクス」 講義予定表

1. 配列アライメント (五斗)
2. データベースサーチ (五斗)
3. 機能・構造予測 (五斗)
4. 配列モチーフ・隠れマルコフモデル (五斗)
5. アセンブリとマッピング (Blanc-Mathieu)
6. 分子進化 (緒方)
7. 分子系統解析 (緒方)
8. 分子系統解析【演習】 (Blanc-Mathieu)
9. 遺伝子予測 (緒方)
10. 機能アノテーション・比較ゲノム (緒方)
11. システムズバイオロジー (五斗)
12. メタゲノム解析 (緒方)
13. メタゲノム解析 (緒方)

<http://goto.kuicr.kyoto-u.ac.jp/lecture/bioinfo.html>

配列からの機能・構造予測



構造・機能予測 — 本日のトピック

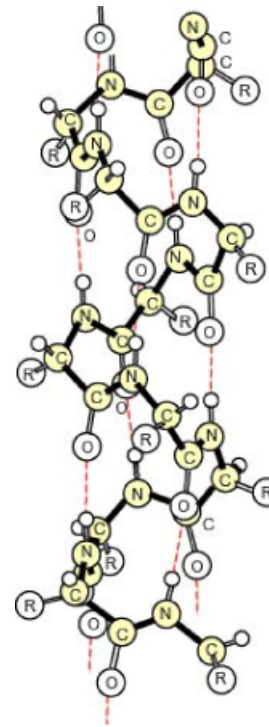
1. タンパク質の二次構造予測
2. タンパク質の局在予測
3. その他（発展）
 - タンパク質の立体構造予測
 - RNA二次構造予測

タンパク質の二次構造

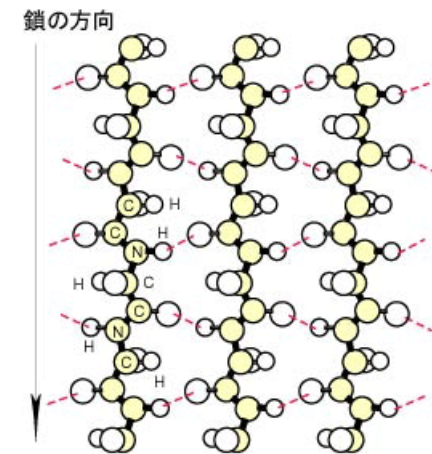
◆ 立体構造中に部分的に存在する主鎖の規則構造

- α ヘリックス
- β シート
- コイル・ループ

◆ C=O と N-H の水素結合が安定化に寄与



α ヘリックス



β シート

タンパク質の二次構造予測

- ◆ アミノ酸配列上の各残基にヘリックス、シート、コイル（ヘリックス、シート以外）のどれかを割り当てる手法
- ◆ なぜ二次構造予測をするのか？
 - ◆ 立体構造の構成要素となっている。
 - ◆ 膜貫通領域など、それ自身が構造・機能と直接関係している場合もある。
 - ◆ 配列のホモロジーがない場合でも、類似構造を取る場合がある。

タンパク質の二次構造予測 ～方法～

- ◆ ウィンドウサーチ
 - 既知の立体構造を元に、各アミノ酸残基が二次構造の要素（ヘリックス，シート，コイル）に現れる頻度を調べて統計解析
 - Chou-Fasman, GOR (Garnier, Osguthorpe, and Robson) 法など
- ◆ マルチプルアライメントとニューラルネットを用いる手法が主流
 - PHD, PSIPRED, Jpred
- ◆ サポートベクターマシン、決定木を用いたものもある
- ◆ 予測精度は70%程度

Chou-Fasman

Helical Residues ^b	P_α	β -Sheet Residues ^c	P_β
Glu ⁽⁻⁾	1.53	Met	1.67
Ala	1.45	Val	1.65
Leu	1.34	Ile	1.60
His ⁽⁺⁾	1.24	Cys	1.30
Met	1.20	Tyr	1.29
Gln	1.17	Phe	1.28
Trp	1.14	Gln	1.23
Val	1.14	Leu	1.22
Phe	1.12	Thr	1.20
Lys ⁽⁺⁾	1.07	Trp	1.19
Ile	1.00	Ala	0.97
Asp ⁽⁻⁾	0.98	Arg ⁽⁺⁾	0.90
Thr	0.82	Gly	0.81
Ser	0.79	Asp ⁽⁻⁾	0.80
Arg ⁽⁺⁾	0.79	Lys ⁽⁺⁾	0.74
Cys	0.77	Ser	0.72
Asn	0.73	His ⁽⁺⁾	0.71
Tyr	0.61	Asn	0.65
Pro	0.59	Pro	0.62
Gly	0.53	Glu ⁽⁻⁾	0.26

- 既知の立体構造から、各アミノ酸残基が α ヘリックスと β シートに現れる頻度 (P_α と P_β) を計算して分類。

- H: Strong former
- h: Former
- I: Weak former
- i: Indifferent
- b: Breaker
- B: Strong breaker

- 5~6残基のウィンドウで P_α と P_β を比較することによって α ヘリックスと β シートを予測。

Chou and Fasman (1974) J Mol Biol

タンパク質の二次構造予測

◆ 演習 3-1

- Chou-Fasman のインデックスを用いたウィンドウサーチで二次構造を予測してみよう。
- 例：ニワトリのトリオースリン酸イソメラーゼ
 - SWISS-PROT: TIPS_CHICK
- 5 残基ウィンドウの平均： $P\alpha$ の左端
 - $(1.20+1.45+0.59+0.79+1.07) / 5 = 1.02$
- まずは単純に $P\alpha$ の平均と $P\beta$ の平均を比較して、大きい値を持つ二次構造を割り当ててみよう。
- ちなみに実際の構造は

MAPRKFFVGGNWKMNQDKKSLGELIHTLNG

CCCCCEEEEECCCCCCHHHHHHHHHHHHHH

E:シート、H:ヘリックス、c:コイル

Chou-Fasman

- 実際には複数のヒューリスティクスを適用して α ヘリックス、 β シート、コイル（ループ）を割り当てている。
- ルール1： α ヘリックス
 - $\langle P\alpha \rangle \geq 1.03$ and $\langle P\alpha \rangle > \langle P\beta \rangle$ (1)
 - Proはヘリックス中には出現しない。
 - Pro, Asp, Glu は N末に、His, Lys, Arg は C末に出現することが多い。
 - ...
- ルール2： β シート
 - $\langle P\beta \rangle \geq 1.05$ and $\langle P\beta \rangle > \langle P\alpha \rangle$ (2)
 - Gluはめったにシート中に出現しない。
 - Proはシートの中央には出現しない。
 - TrpはほとんどがN末に出現し、C末に出現することはめったにない。
 - ...
- ルール1にもルール2にも当てはまらない：コイル

GOR

- 各アミノ酸残基が α ヘリックス、 β シート、コイルに現れる頻度を残基の位置情報も考慮して計算。
- 17残基のウィンドウで α ヘリックス、 β シート、コイルへのなりやすさを予測。

Directional information measure for the α -helical conformation†

Amino acid residue	Residue position‡ (centinats)																
	$j - 8$	$j - 6$		$j - 4$		$j - 2$		j	$j + 2$	$j + 4$		$j + 6$		$j + 8$			
Gly	-5	-10	-15	-20	-30	-40	-50	-60	-86	-60	-50	-40	-30	-20	-15	-10	-5
Ala	5	10	15	20	30	40	50	60	65	60	50	40	30	20	15	10	5
Val	0	0	0	0	0	0	5	10	14	10	5	0	0	0	0	0	0
Leu	0	5	10	15	20	25	28	30	32	30	28	25	20	15	10	5	0
Ile	5	10	15	20	25	20	15	10	6	0	-10	-15	-20	-25	-20	-10	-5
Ser	0	-5	-10	-15	-20	-25	-30	-35	-39	-35	-30	-25	-20	-15	-10	-5	0
Thr	0	0	0	-5	-10	-15	-20	-25	-26	-25	-20	-15	-10	-5	0	0	0
Asp	0	-5	-10	-15	-20	-15	-10	0	5	10	15	20	20	20	15	10	5
Glu	0	0	0	0	10	20	60	70	78	78	78	78	78	70	60	40	20
Asn	0	0	0	0	-10	-20	-30	-40	-51	-40	-30	-20	-10	0	0	0	0
Gln	0	0	0	0	5	10	20	20	10	-10	-20	-20	-10	-5	0	0	0
Lys	20	40	50	55	60	60	50	30	23	10	5	0	0	0	0	0	0
His	10	20	30	40	50	50	50	30	12	-20	-10	0	0	0	0	0	0
Arg	0	0	0	0	0	0	0	0	-9	-15	-20	-30	-40	-50	-50	-30	-10
Phe	0	0	0	0	0	5	10	15	16	15	10	5	0	0	0	0	0
Tyr	-5	-10	-15	-20	-25	-30	-35	-40	-45	-40	-35	-30	-25	-20	-15	-10	-5
Trp	-10	-20	-40	-50	-50	-10	0	10	12	10	0	-10	-50	-50	-40	-20	-10
Cys	0	0	0	0	0	0	-5	-10	-13	-10	-5	0	0	0	0	0	0
Met	10	20	25	30	35	40	45	50	53	50	45	40	35	30	25	20	10
Pro	-10	-20	-40	-60	-80	-100	-120	-140	-77	-60	-30	-20	-10	0	0	0	0

Garnier, Osguthorpe and Robson (1978) J Mol Biol

PHD

profile generation
from a multiple
sequence alignment
(here: β -lactamase: 3bla)

protein	DSSP	aligned sequence	number of example: input
---------	------	---------------------	-----------------------------

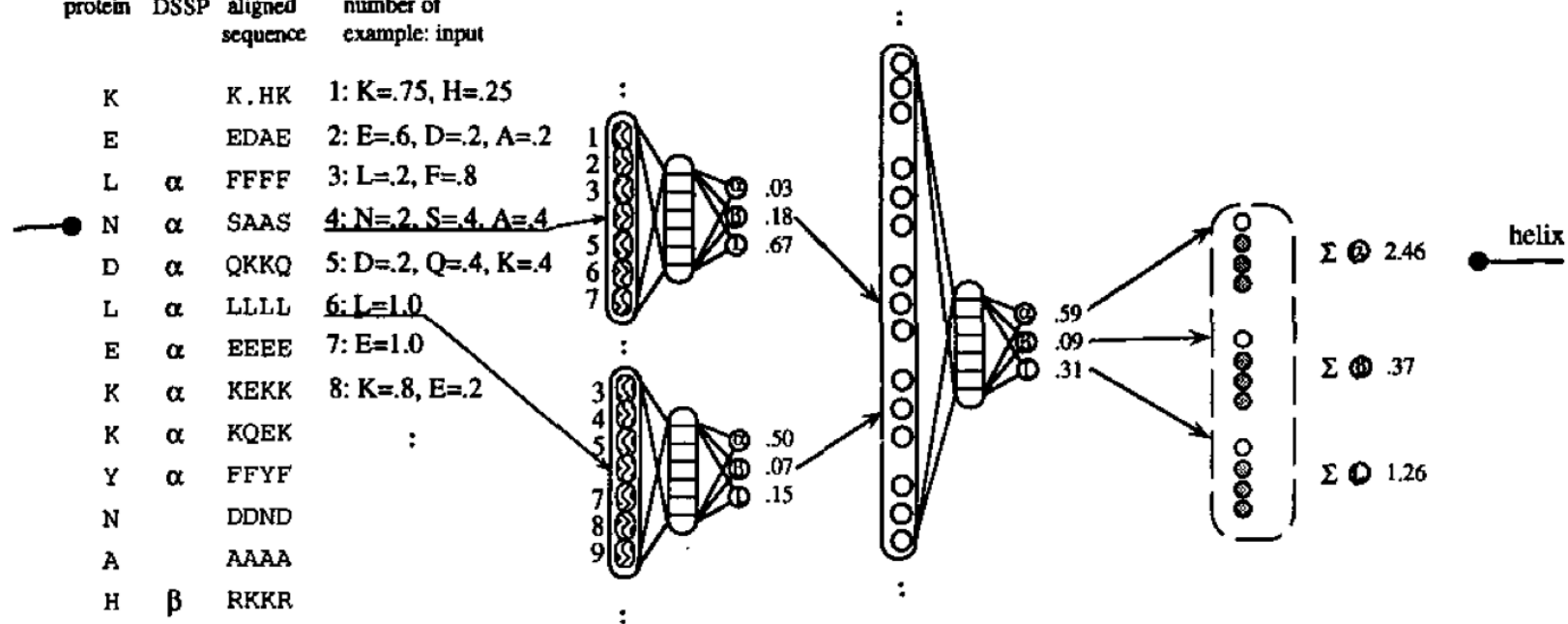
K		K . HK	1: K=.75, H=.25
E		EDAE	2: E=.6, D=.2, A=.2
L	α	FFFF	3: L=.2, F=.8
N	α	SAAS	4: N=.2, S=.4, A=.4
D	α	QKKQ	5: D=.2, Q=.4, K=.4
L	α	LLLL	6: L=1.0
E	α	EEEE	7: E=1.0
K	α	KEKK	8: K=.8, E=.2
K	α	KQEK	:
Y	α	FFYF	:
N		DDND	:
A		AAAA	:
H	β	RKKR	:
I	β	LLLL	:
G	β	GGGG	:
:	:	::::	:

first level:
sequence to structure
in: profiles,
out: units for
helix (α), strand (β)
and loop(L)

second level:
structure to structure
in: output of first
level, out: α , β , L

third level:
jury decision
in: output of
different networks
out: arithmetic
average for α , β , L

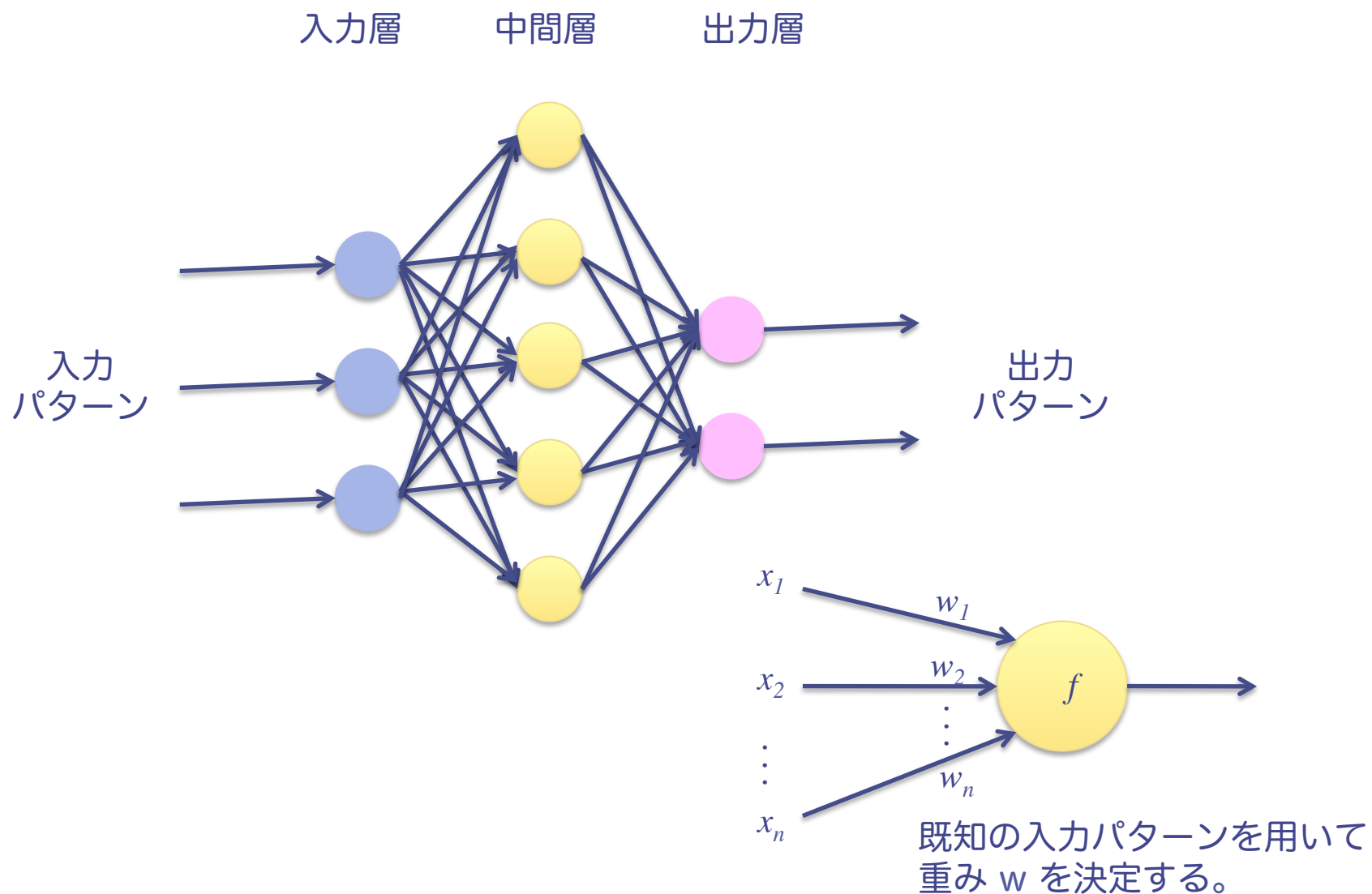
prediction
winner take all
(given here
for the N at
position 4)



3層のニューラルネットワーク×2

Rost and Sander (1993) J Mol Biol

3層のニューラルネットワーク



PHD

◆ ニフトリのトリオースリン酸イソメラーゼ

■ SWISS-PROT: TIPS_CHICK をクエリに与えた結果

10 20 30 40 50 60 70
| | | | | | |
MAPRKFFVGGNWKMNGDKKSLGELIHTLNGAKLSADTEVVCGAPSIYLDFAEQKLDKIGVAAQNCYKVP
CCCceEEecCcccCCCCcchhhHhhhhhecCCcecCCeEecCCchhhHHHHHHHHhhhhchhhhhhhccC
KGAFTGEISPAMIKDIGAAWVILGHSERRHVFGESEDELIGQKVAHALAEGLGVIACIGEKLDEREAGITE
CCCCccccCHHHHHHHhhheeeccccceEecCcHHHHHHHHHHHHHHCCcEEEEeccCCCCcCchh
KVVFEQTKAIADNVKDWSKVVLAYEPVWAIGTGKTATPQQAQEVHEKLRGWLKSHVSDAVAQSTRIIYGG
hhhhHHHHHHHHHHhhcceEEEEEEeeEeeecCCCCCHHHHHHHHHHHHHHHHHHHHHhhceEEEEEE
SVTGGNCKELASQHDVDGFLVGGASLKPEFVDIINAKH
ccCCcChHHHHhCCCCceEEeCccCCchhHhhhhcCC

MAPRKFFVGGNWKMNGDKKSLGELIHTLNG
CCCCCEEEEEEECCCCCCHHHHHHHHHHHHHH
E:シート、H:ヘリックス、c:コイル

PSIPRED

Raw profile from PSI-BLAST Log File

Position-based scoring matrix used

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
-3	-4	-4	-4	-3	-4	-4	-4	-2	-1	-1	-4	-1	8	-5	-3	-3	0	2	-2	
0	-1	-1	3	-4	3	4	1	-1	-4	-4	0	-3	-4	-2	-1	-2	-4	-3	-3	
0	-1	2	1	-3	4	0	-1	-2	-4	-3	1	-2	-4	-2	2	0	-4	-3	-3	
-2	-3	-4	-5	-2	-3	-4	-6	-4	0	6	0	0	-1	-4	-3	-2	-4	-2	0	
0	-3	-1	-2	-3	0	-2	4	-3	-3	0	-2	-2	-4	-3	3	1	-4	-4	-3	
0	2	0	4	-4	1	2	1	-2	-4	-4	0	-3	-4	-3	1	-2	-5	-4	-4	
-1	5	3	-2	-4	-1	-1	1	-2	-1	-4	1	-3	-4	-3	1	-2	-5	-4	-4	
-2	-3	-4	-5	-3	-3	-4	-5	-4	3	4	-1	1	2	-4	-3	-2	-3	-1	0	
-2	3	2	-2	-4	2	1	-3	-2	-3	-3	1	1	-4	-3	2	1	-4	-3	-1	
0	2	3	1	-4	0	0	0	-2	-4	-4	1	-3	-4	-3	2	0	-5	-4	-4	
5	-3	-3	-2	-3	-3	-2	-3	1	-2	-3	-2	1	-3	0	1	-4	-2	0		
-1	-4	-5	-5	-3	-4	-4	-5	-4	3	3	-4	2	3	-5	-3	-2	5	-1	2	
0	3	3	0	-4	3	0	1	-2	-4	-4	1	-3	-4	-3	1	-1	-4	-3	-4	
-1	0	1	0	-4	1	-1	-1	-2	-4	-3	5	-2	0	-3	0	-2	-4	0	-3	
-2	-3	-1	-5	-3	-3	-4	-5	-4	3	4	0	4	2	-4	-3	-2	-3	-2	0	
0	3	0	-2	-3	-1	0	0	-2	0	0	1	0	-1	-3	2	0	-4	-3	0	
-1	1	3	-2	-4	0	-2	4	-2	-4	-4	0	-3	0	-3	0	0	-3	0	-4	

Window of 15 rows

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
	0.4	0.3	0.3	0.3	0.2	0.9	0.3	0.3	0.4	0.4	0.4	0.3	0.4	0.9	0.1	0.4	0.4	0.5	0.7	0.4
	0.3	0.2	0.3	0.8	0.4	0.3	0.7	0.1	0.6	0.2	0.4	0.3	0.5	0.2	0.1	0.4	0.8	0.2	0.3	0.2
	0.1	0.1	0.4	0.3	0.5	0.1	0.1	0.3	0.1	0.1	0.4	0.2	0.4	0.9	0.3	0.4	0.4	0.9	0.3	0.6
	0.6	0.3	0.3	0.1	0.3	0.5	0.5	0.2	0.1	0.4	0.4	0.3	0.6	0.9	0.1	0.5	0.1	0.5	0.7	0.4
.																				
.																				
.																				

15 x 20 scaled inputs to 1st network

1st Network
315 inputs
75 hidden units
3 outputs

Window of 15 x 3
outputs fed to 2nd
network

2nd Network
60 inputs
60 hidden units
3 outputs

Final 3-state
Prediction

60の入力：15×4
(二次構造+ウィンドウの
インデックス)

315の入力：15×21 (アミノ酸残基+ウィンドウのインデックス)
3の出力：ヘリックス、シート、コイル

Jones (1999) J Mol Biol

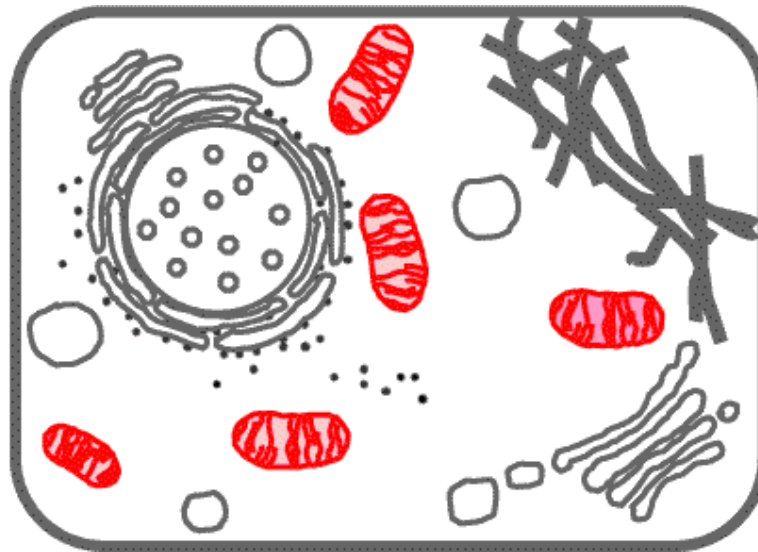
構造・機能予測 — 本日のトピック

1. タンパク質の二次構造予測
2. タンパク質の局在予測
3. その他（発展）
 - タンパク質の立体構造予測
 - RNA二次構造予測

細胞内局在予測

◆ アミノ酸配列から、そのタンパク質が細胞内のどの小器官で働くかを予測する手法

- 核
- 細胞質
- ミトコンドリア
- 葉緑体
- ゴルジ体
- 液胞
- 細胞膜



PLOC

◆ 機能が分からなくても、局在が分かれば機能の手がかりが得られることもある。

細胞内局在予測プログラム

◆ PSORT

- 配列中に既知のシグナルがあるかどうかを調べながら予測する決定木を用いた方法。
- <http://psort.hgc.jp/>

◆ iPSORT→Wolf PSORT

- PSORTに配列類似性の要素を導入した拡張。
- <http://ipsort.hgc.jp/>

◆ ESLPred, SubLoc

- SVM (Support Vector Machine) を用いた学習。

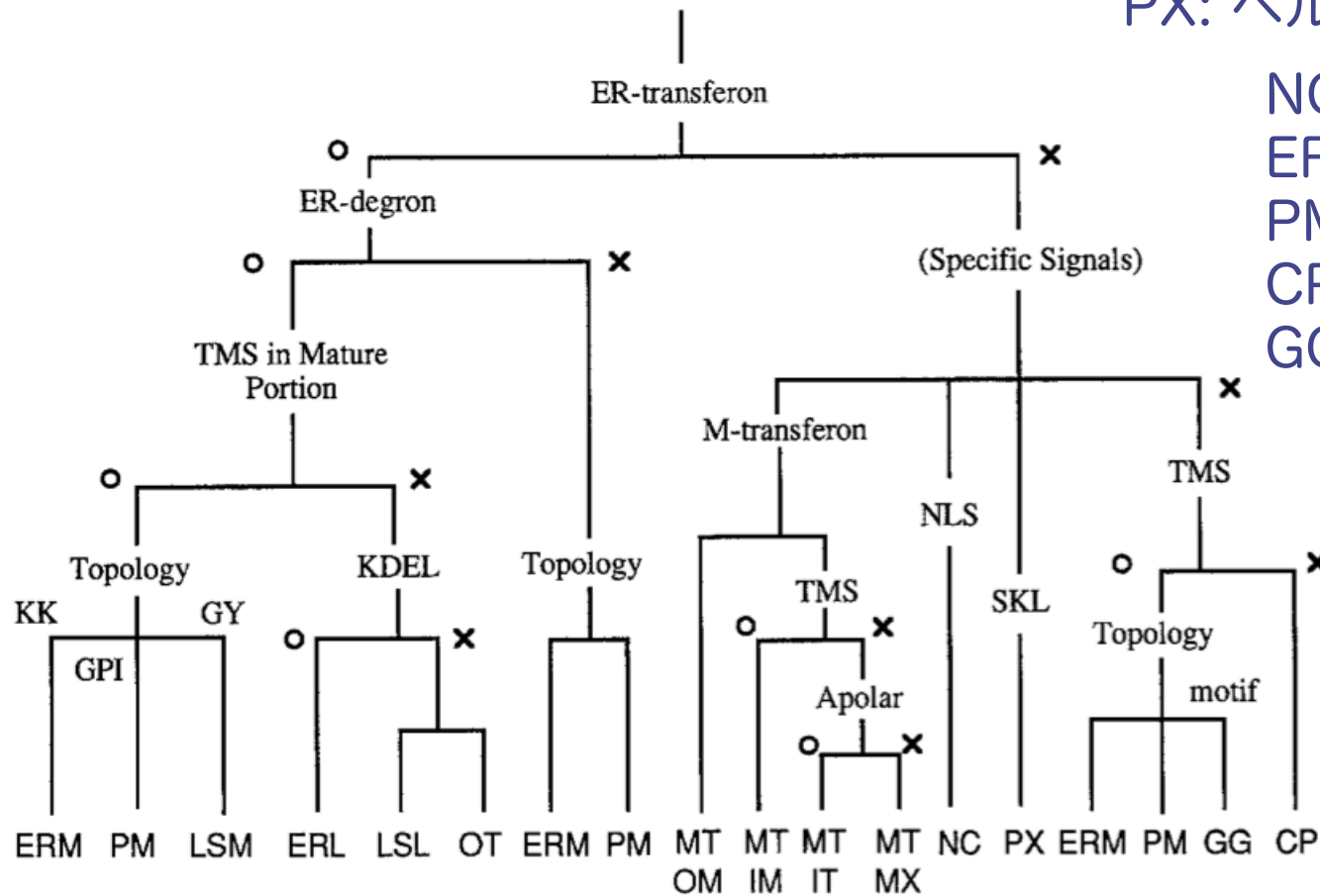
◆ SOSUI, TMHMM: 膜貫通部位予測

- SOSUI: ウィンドウベース。
- TMHMM: HMM によるモデル化。

PSORT の決定木

LS: リソソーム
MT: ミトコンドリア
PX: ペルオキシソーム

NC: 核
ER: 小胞体
PM: 細胞膜
CP: 細胞質
GG: ゴルジ体



Nakai et al. (1992) Genomics

Terminology for Various Sorting Signals^a

Name of a signal	Function
M-transferon	Translocation from cytosol into mitochondrial matrix (matrix targeting sequence)
M-degron	Recognition signal for the cleavage of M-transferons
M/IMS-transferon	Translocation from mitochondrial matrix (M) into intermembrane space (IMS)
Nu-transferon	Translocation from cytosol into nucleus (nuclear localization signal)
P-transferon	Translocation from cytosol into peroxisome
ER-transferon	Translocation from cytosol into endoplasmic reticulum (signal sequence)
ER-comparton	Retention in endoplasmic reticulum (for the lumen or membrane)
GPI-modon	Generation of GPI-anchor after the cleavage (also a degron?)
M6P-modon, G/L-comparton	Golgi-mediated generation of Man6P residues; may be identical with the signal for transfe from <i>trans</i> -Golgi to lysosomes
S-transferon	Translocation from cytosol into chloroplast stroma
S/T-transferon	Translocation from chloroplast stroma (S) into thylakoid space (T)

膜貫通部位予測

- ◆ 膜貫通領域
 - α ヘリックス
- ◆ 7～17残基程度の疎水性指標の平均値をプロット
- ◆ 平均値が高い部分が膜貫通領域と推定

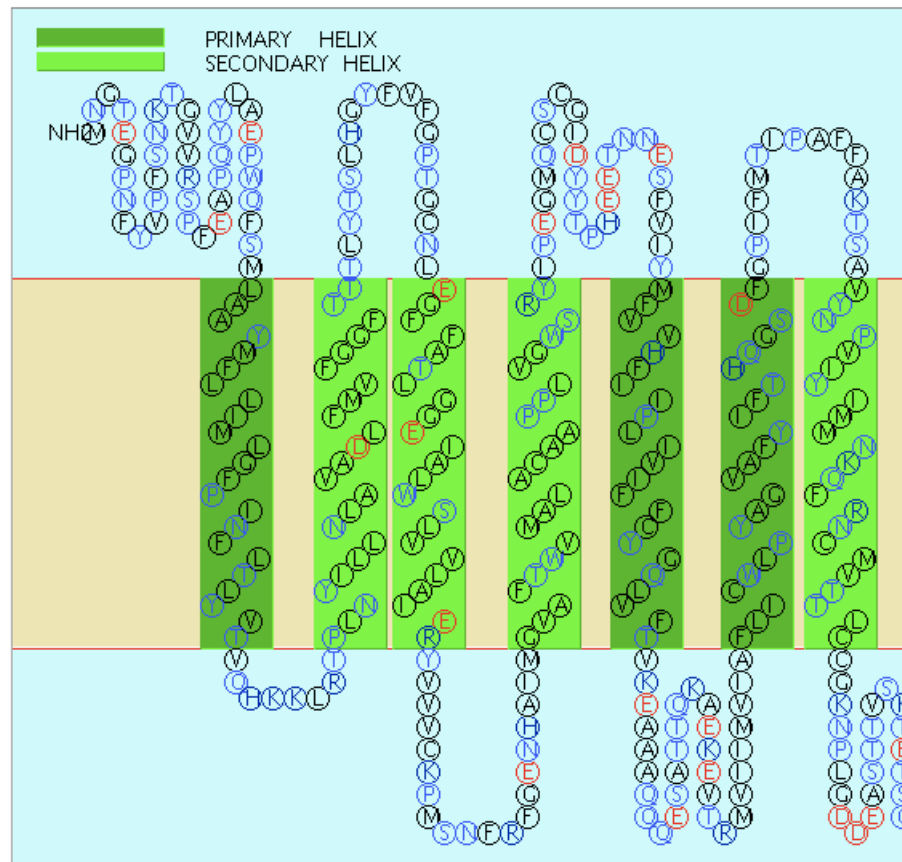
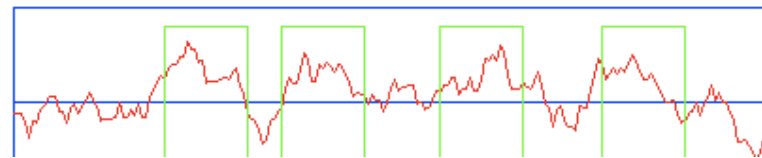
疎水性指標 (H) と両親媒性指標 (A, A')

Amino acid	A	A'	H
Lysine (K)	3.67	0	-3.9
Arginine (R)	2.45	0	-4.5
Histidine (H)	1.45	0	-3.2
Glutamic acid (E)	1.27	0	-3.5
Glutamine (Q)	1.25	0	-3.5
Aspartic acid (D)	0	0	-3.5
Asparagine (N)	0	0	-3.5
Trptophan (W)	0	6.93	-0.9
Tyrosine (Y)	0	5.06	-1.3
Serine (S)	0	0	-0.8
Threonine (T)	0	0	-0.7
Proline (P)	0	0	-1.6
Glycine (G)	0	0	-0.4
Alanine (A)	0	0	1.8
Methionine (M)	0	0	1.9
Cysteine (C)	0	0	2.5
Phenylalanine (A)	0	0	2.8
Leucine (L)	0	0	3.8
Valine (V)	0	0	4.2
Isoleucine (I)	0	0	4.5

Mitaku et al. (2002) Bioinformatics

膜貫通部位予測：SOSUI

[Hydropathy profile]

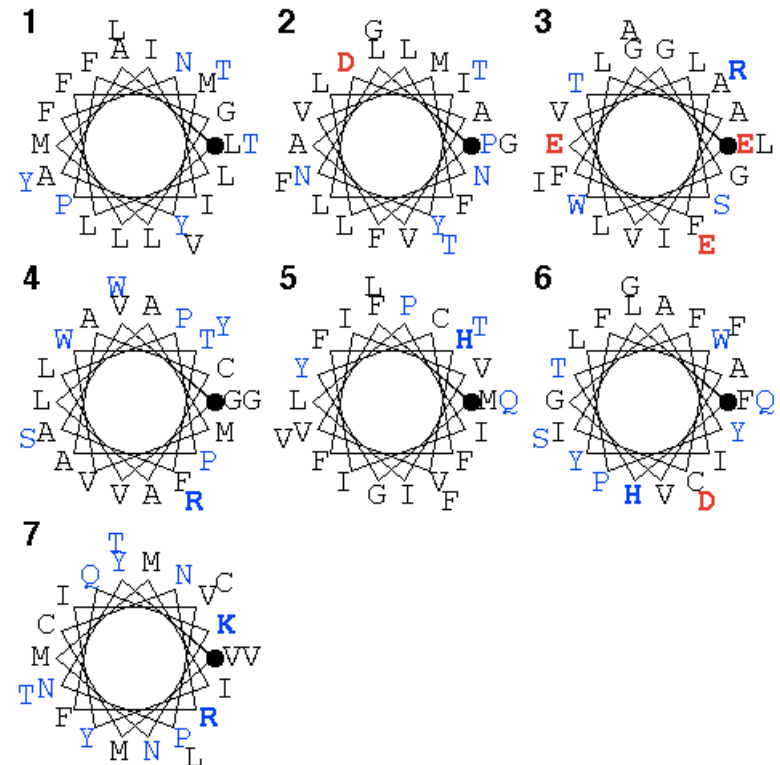


[Helical wheel diagram of predicted segments]

Hydrophobic residue: Black

Polar residue: Blue

Charged residue: Bold blue(+) Bold red(-)



膜貫通部位予測：SOSUI

SOSUI: Submit a protein sequence

[\[Sample Sequences\]](#) [\[References\]](#)

Enter a title or comment **SOSUI Result**

Rhodopsin

Enter your sequence with
(Minimum: 20 a.a., Maximum: 1000 a.a.)
Query title : Rhodopsin

PINFLTLY
VTVQHKKLRTPLNYILLNLAV
EGFFATLG
GEIALWSLVVLAIERVYVVK
VGWSRYIP
EGMQCSCGIDYYTPHEETNN
AQQES
ATTQKAEKEVTRMVIIMVIAF
AKTSAV
YNPVIYIMMNKQFRNCMVTT

Total length : 348 A. A.

Average of hydrophobicity : 0.473851

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

ウィンドウベースで疎水性の指標を
計算する。

To execute the query, press

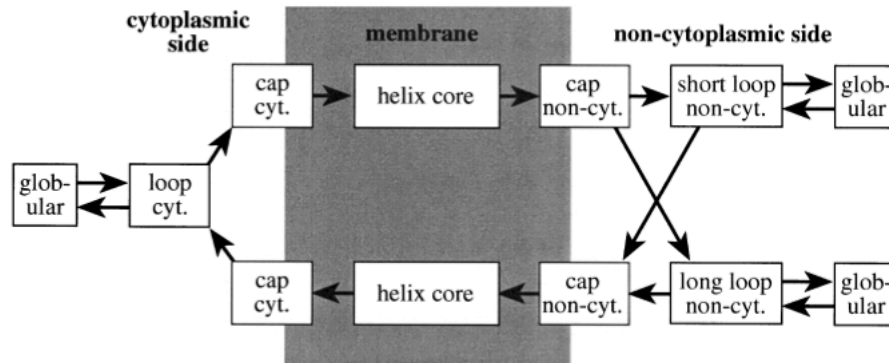
To clear the form, press

**This amino acid sequence is of a MEMBRANE PROTEIN
which have 7 transmembrane helices.**

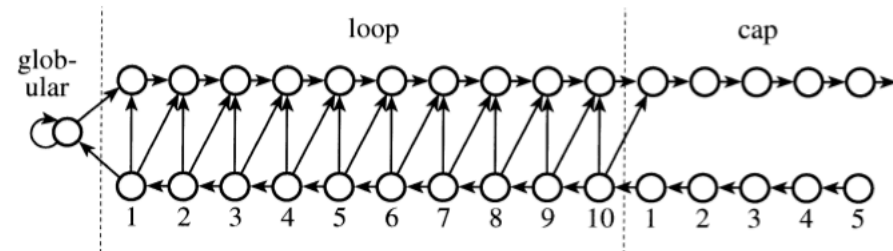
No.	N terminal	transmembrane region	C terminal	type	length
1	40	LAAYMFLLIMLGFPINFLTLYVT	62	PRIMARY	23
2	71	PLNYILLNLAVADLFMVFGGFTT	93	SECONDARY	23
3	113	EGFFATLGGEIALWSLVVLAIER	135	SECONDARY	23
4	156	GVAFTWVMALACAAPPLVGWSRY	178	SECONDARY	23
5	207	MFVVHFIPLIVIFFCYGQLVFT	229	PRIMARY	23
6	261	FLICWLPHYAGVAFYIFTHQGSDF	283	PRIMARY	23
7	300	VYNPVIYIMMNKQFRNCMVTTLC	322	SECONDARY	23

膜貫通部位予測：TMHMM

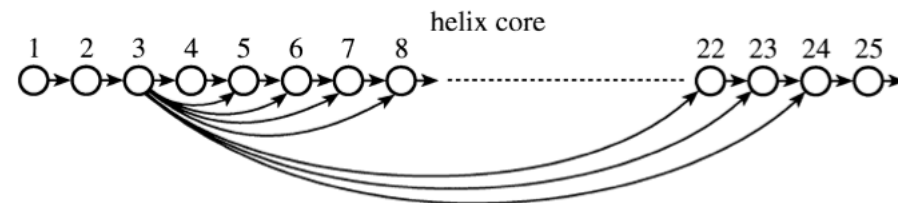
(a)



(b)



(c)



Krogh et al. (2001) J Mol Biol

構造・機能予測 — 本日のトピック

1. タンパク質の二次構造予測
2. タンパク質の局在予測
3. その他（発展）
 - タンパク質の立体構造予測
 - RNA二次構造予測

立体構造の予測

◆ ホモロジーモデリング

- Homology modeling / Comparative modeling
- 配列が似ていれば構造も似ている
- 立体構造既知のタンパク質で配列ホモロジーがあるものから構造をモデリング

◆ スレッディング

- Threading / 3D-1D / Fold recognition
- 立体構造ライブラリー
- 配列・構造適合性評価関数（ポテンシャル関数）

◆ アブイニシオ法

- 構造エネルギーがエネルギーが最小になるような原子の配置を計算する

フラグメント・アセンブリ法

方法

- ◆ 数残基から十数残基の断片構造をプロフィール比較法などを用いて既知構造データベースから取得
→ 各断片配列ごとにいくつかの候補を選ぶ
- ◆ フラグメントをつなぎ合わせることでより全体構造を予測。つなぎ合わせる際には分子動力学法などによるエネルギー最適化などを行う
- ◆ Univ. Washington の Baker らが開発
- ◆ 現時点では最強の方法と考えられている

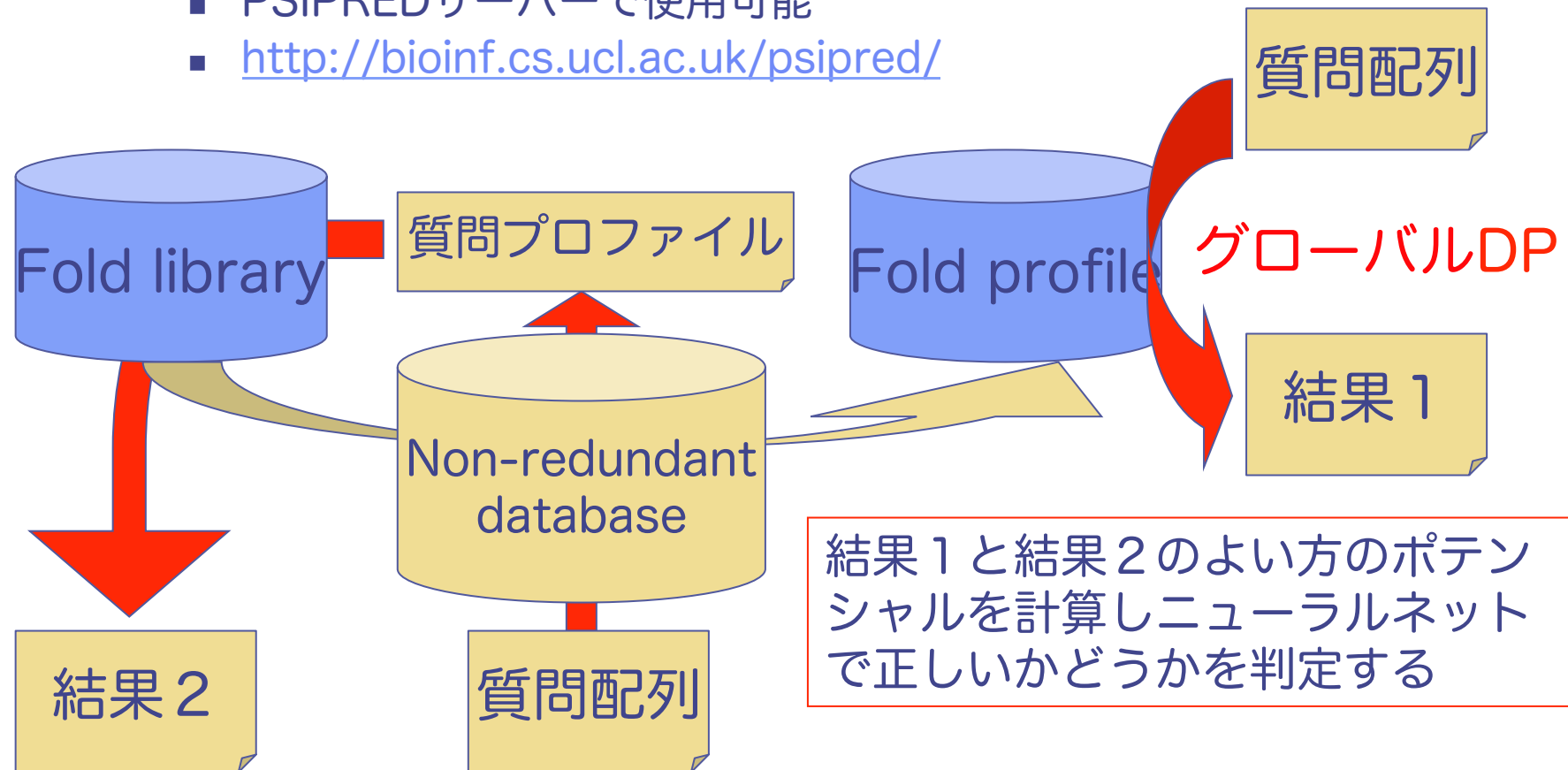
立体構造予測におけるブレークスルー

- ◆ スレッディング法の発明 (Eisenberg et al., 1991)
 - 構造既知の配列と類似性が無い配列の構造予測
- ◆ PSI-BLASTの開発 (Altschul et al, 1997)
 - プロファイルに基づくマルチプルアライメントの繰り返し実行によるスレッディング
- ◆ David Baker による ab initio 予測 (1997)
 - 統計情報+シミュレーション

立体構造予測システムの実例

◆ GenTHREADER

- ホモロジーモデリングとスレッディングを組み合わせる手法
- PSIPREDサーバーで使用可能
- <http://bioinf.cs.ucl.ac.uk/psipred/>



アブイニシオ法

物理的原理に基づく方法

- ◆ エネルギー最小化、もしくは、微分方程式を（数値的に）解く、などの物理的原理に基づく方法
- ◆ 主として分子動力学法(Molecular Dynamics)
- ◆ 数十残基程度であれば、実際のタンパク質やペプチドと似た構造を推定可能（なことがある）
- ◆ 構造の最適化や安定性の解析には実用的
⇒ ホモロジーモデリング
- ◆ 超並列計算機の利用、専用計算機の開発

アブイニシオ法

タンパク質の構造エネルギー

$$E = E_b + E_\theta + E_\phi + E_{vdw} + E_{el}$$

$E_b = \sum K_b (r - r_0)^2$: 共有結合長ポテンシャル

$E_\theta = \sum K_\theta (\theta - \theta_0)^2$: 共有結合角のポテンシャル

$E_\phi = \sum K_\phi [1 + \cos(n\phi - \psi)]$: 二面角のポテンシャル

$E_{vdw} = \sum_{i < j} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right]$: ファン・デル・ワールスカ

$E_{el} = \sum_{i < j} \frac{q_i q_j}{\epsilon r_{ij}}$: 静電的相互作用

立体構造の予測（まとめ）

◆ ホモロジーモデリング

- 立体構造既知のタンパク質で配列ホモロジーがあるものから構造をモデリング

◆ スレッディング

- 配列・構造適合性評価関数（ポテンシャル関数）

◆ アブイニシオ法

- 構造エネルギーがエネルギーが最小になるような原子の配置を計算する

フォールド以外の構造の特徴

◆ 基質結合部位

◆ 触媒部位

◆ アロステリックな構造変化

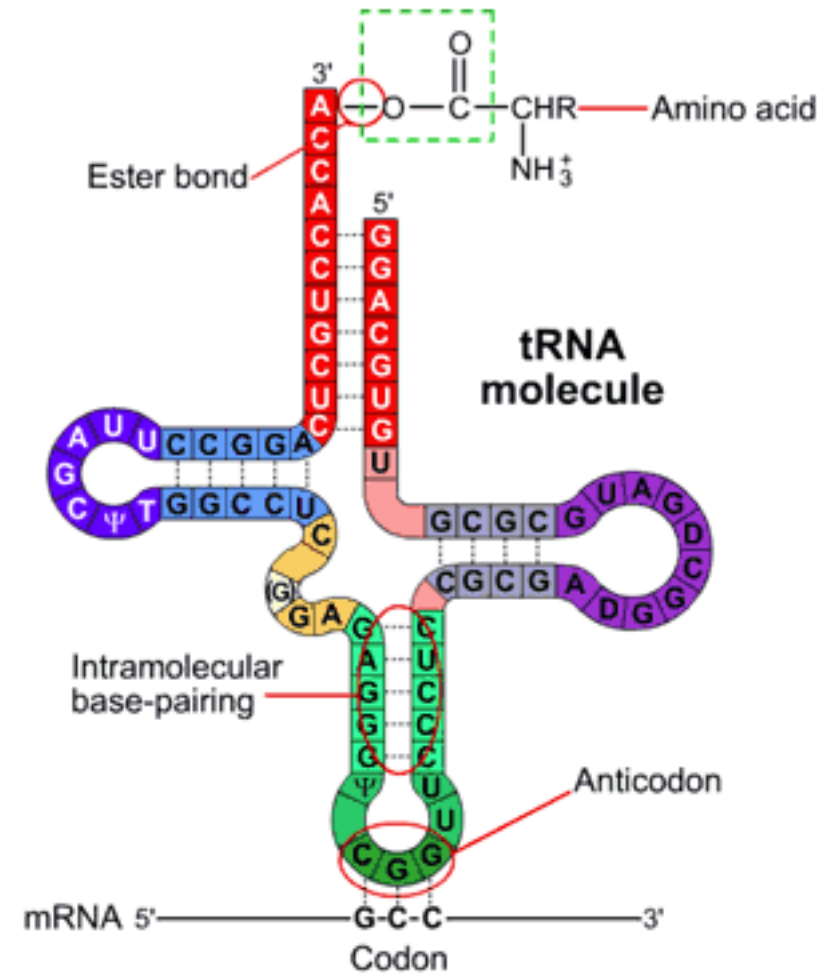
立体構造から機能まで予測するためには、
これらの特徴も考慮する必要がある

構造・機能予測 — 本日のトピック

1. タンパク質の二次構造予測
2. タンパク質の局在予測
3. その他（発展）
 - タンパク質の立体構造予測
 - RNA二次構造予測

DNA, RNA の二次構造

- ◆ DNAのパリンドローム十字構造
- ◆ RNAのステムループなど

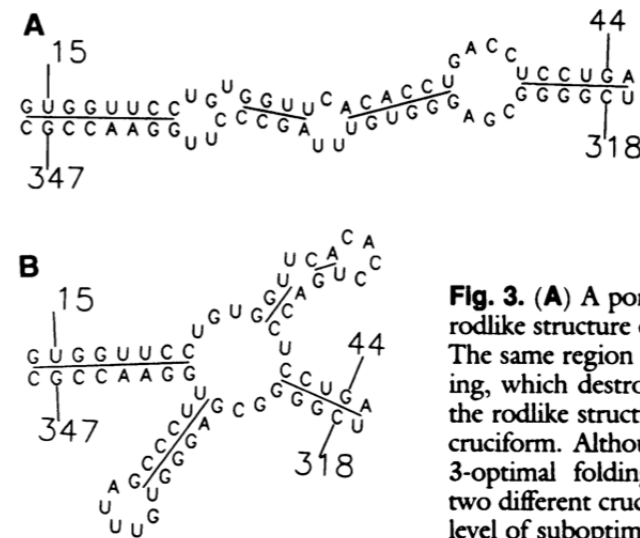
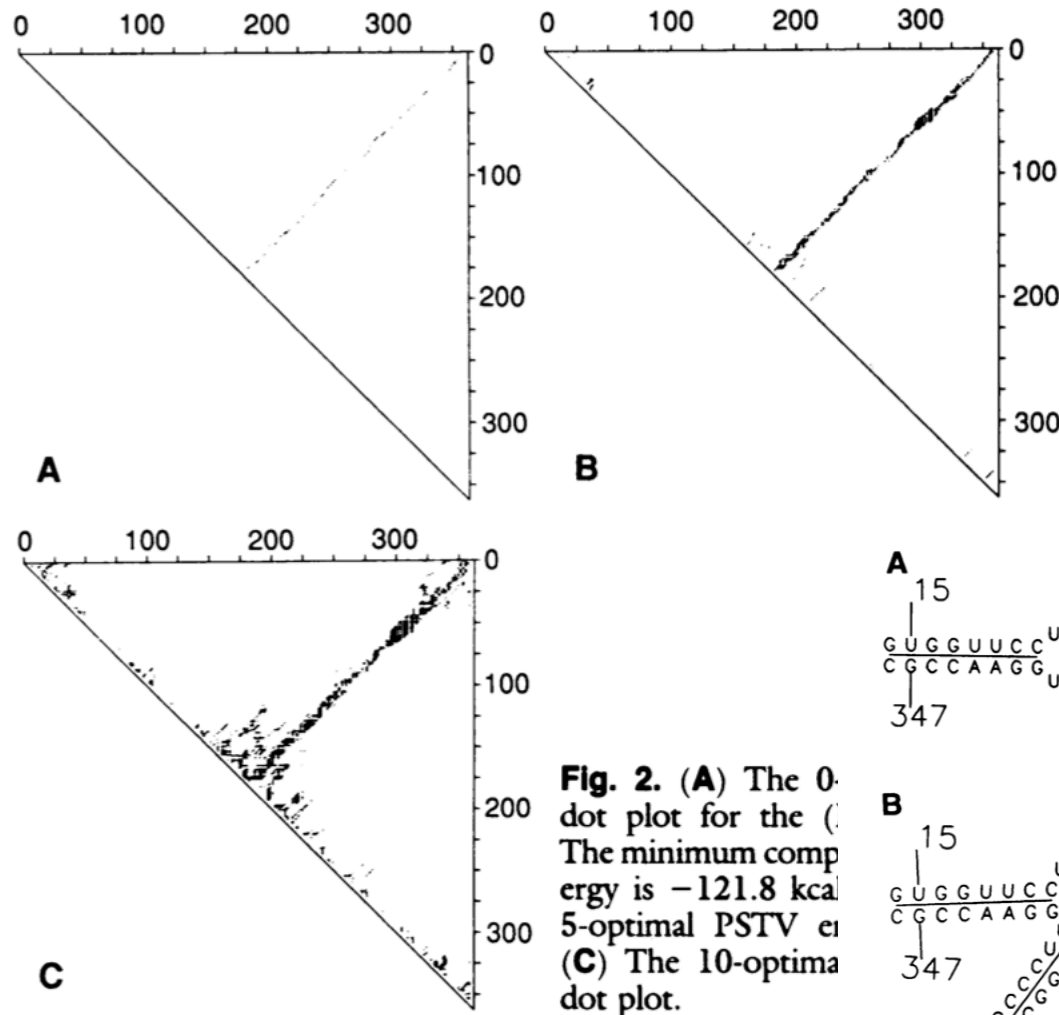


http://www.wiley.com/college/boyer/0470003790/structure/tRNA/trna_intro.htm

RNA の二次構造予測

- ◆ ステム構造と各ループ構造の安定性の評価値を合計したエネルギー関数を最小化する問題
- ◆ 配列の最適化アライメント
 - DP 法：配列の自分自身との最適アライメントを探す
 - ニューラルネット：ステムの候補をつくりその最適組合せを探す
 - 遺伝的アルゴリズム
 - 文脈自由文法

RNA の二次構造予測 (DP法)

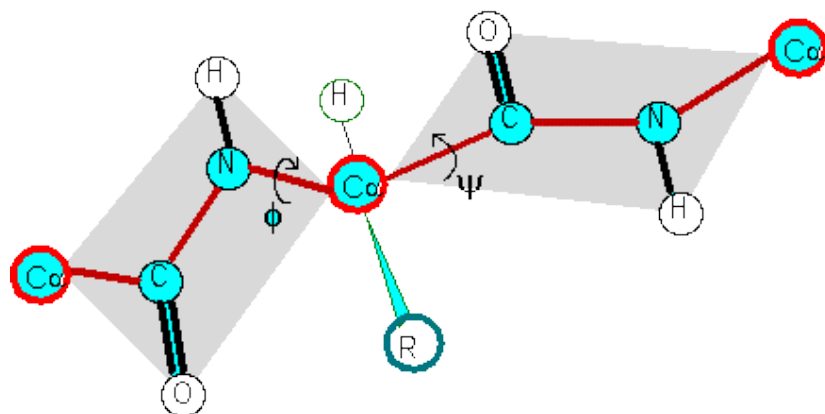


Zuker (1989) Science

構造・機能予測 — 本日のトピック

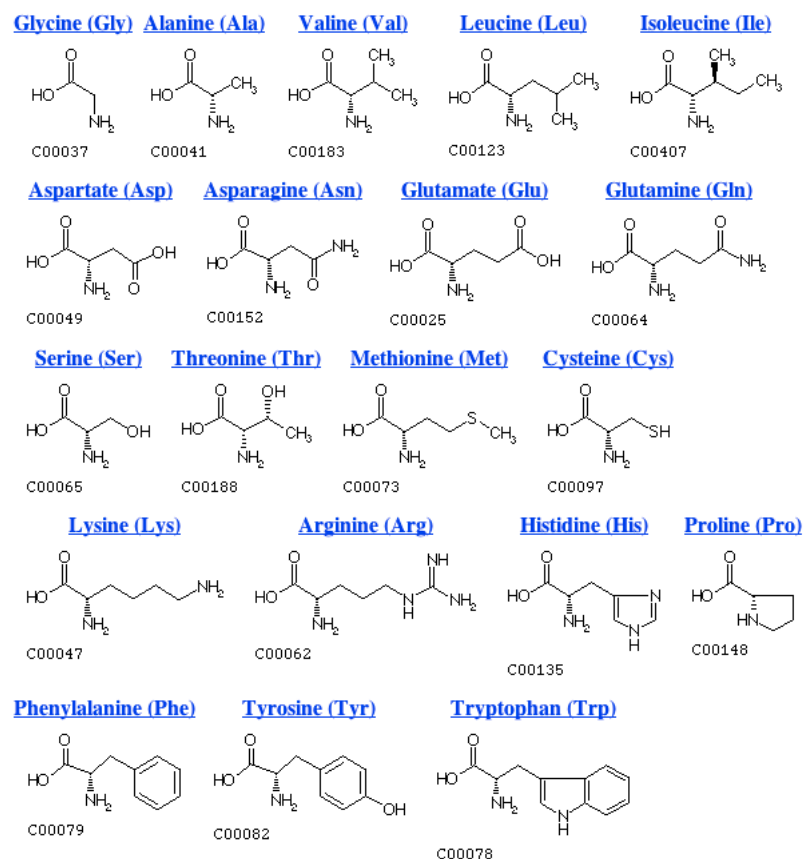
1. タンパク質の二次構造予測
2. タンパク質の局在予測
3. その他（発展）
 - タンパク質の立体構造予測
 - RNA二次構造予測
4. 補足資料

タンパク質の主鎖と側鎖



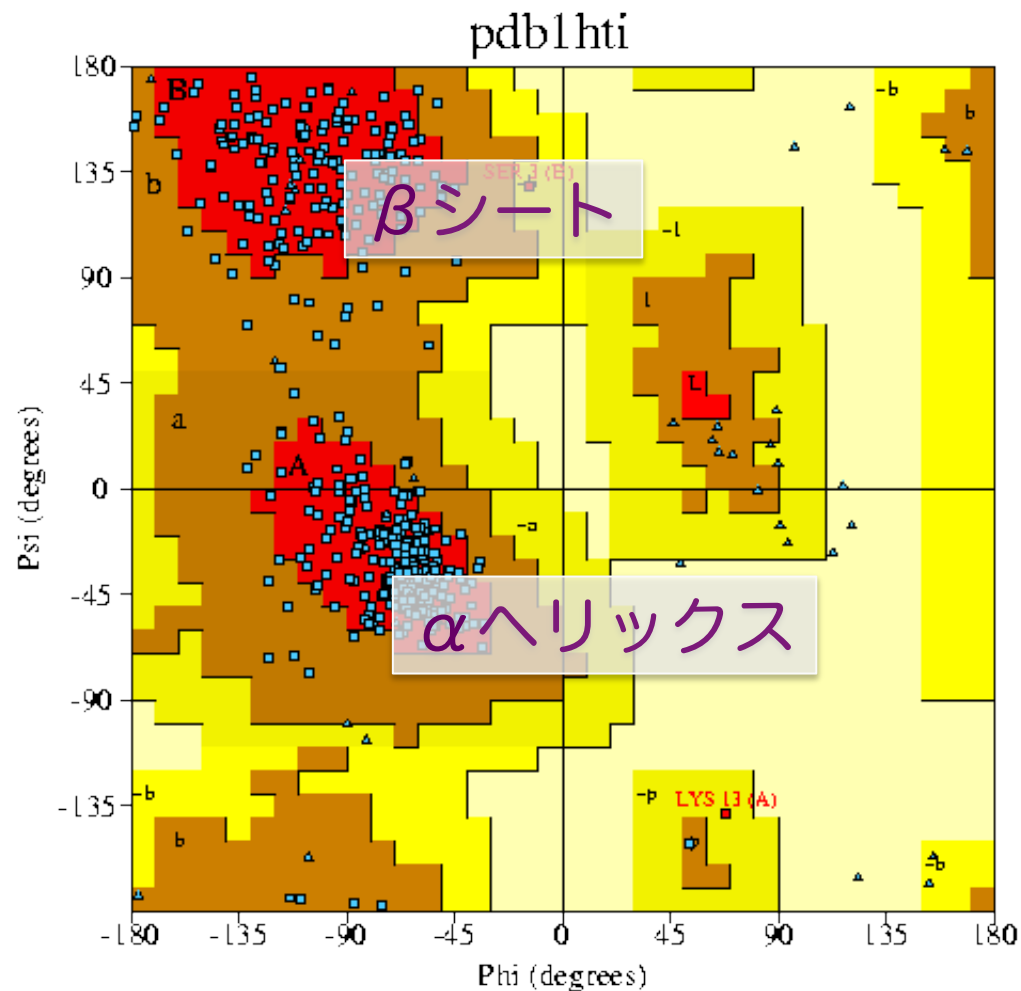
主鎖：-C α -C-N-C α -C-N-
ペプチド結合

側鎖：R (アミノ酸の種類)



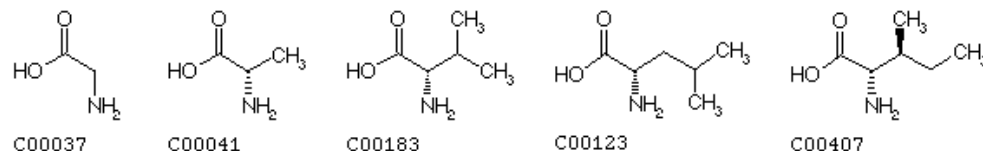
ラマチャンドランプロット

アミノ酸残基がとりうる二面角と二次構造の関係

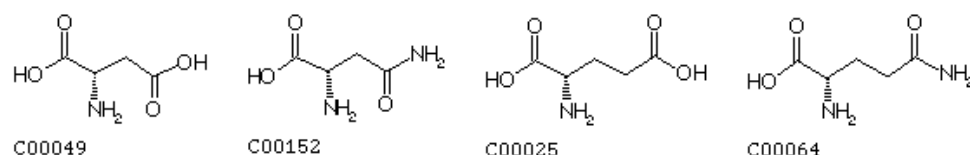


アミノ酸側鎖の性質

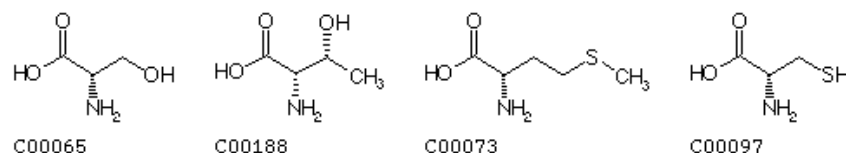
Glycine (Gly) Alanine (Ala) Valine (Val) Leucine (Leu) Isoleucine (Ile)



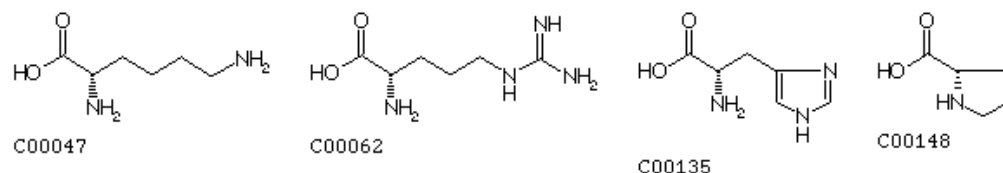
Aspartate (Asp) Asparagine (Asn) Glutamate (Glu) Glutamine (Gln)



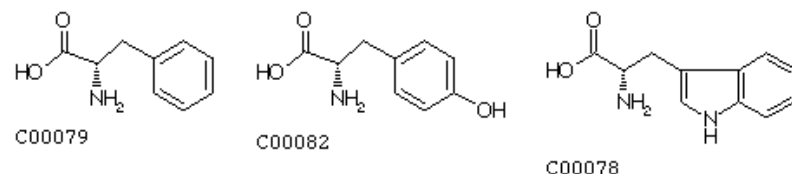
Serine (Ser) Threonine (Thr) Methionine (Met) Cysteine (Cys)



Lysine (Lys) Arginine (Arg) Histidine (His) Proline (Pro)



Phenylalanine (Phe) Tyrosine (Tyr) Tryptophan (Trp)



◆ 疎水性残基

- Ala, Val, Phe, Pro, Met, Ile, Leu, Trp, Gly

◆ 電荷

- 酸性：Asp, Glu
- 塩基性：Lys, Arg, His

◆ 極性

- Ser, Thr, Tyr, Cys, Asn, Gln,

◆ 共有結合

- Cys

タンパク質立体構造とバイオインフォマティクス

◆ 立体構造データベース

- X線結晶解析やNMRによる3次元座標データ
- 立体構造表示

◆ 立体構造アライメント

- 立体構造の類似性を調べる
- 共通フォールドパターン（構造モチーフ）を探す
- フォールドパターンによるタンパク質の分類

◆ 一次配列から立体構造予測

- 配列の類似度から：ホモロジーモデリング
- 配列と構造の適合性から：スレッディング
- 物理化学的原理に基づいて：アブイニシオ法

立体構造データベース

◆ PDB: Protein Data Bank

- X線結晶解析やNMRによる立体構造データ
 - ◆ 各原子の3次元座標
 - ◆ 2次構造：ヘリックス, シート
 - ◆ 解像度, 文献情報
- ルトガーズ大学などを中心としたコンソーシアム
 - ◆ <http://www.rcsb.org/>
- 日本でのミラーサイトと独自のXML化
 - ◆ <http://www.pdbj.org/>
- ゲノムネットでのキーワード検索
 - ◆ http://www.genome.jp/dbget-bin/www_bfind?pdb

立体構造の表現

◆ テキスト表現

- 3次元座標をテキストで表現
 - ◆ フラットファイルやXML
 - ◆ 解析用

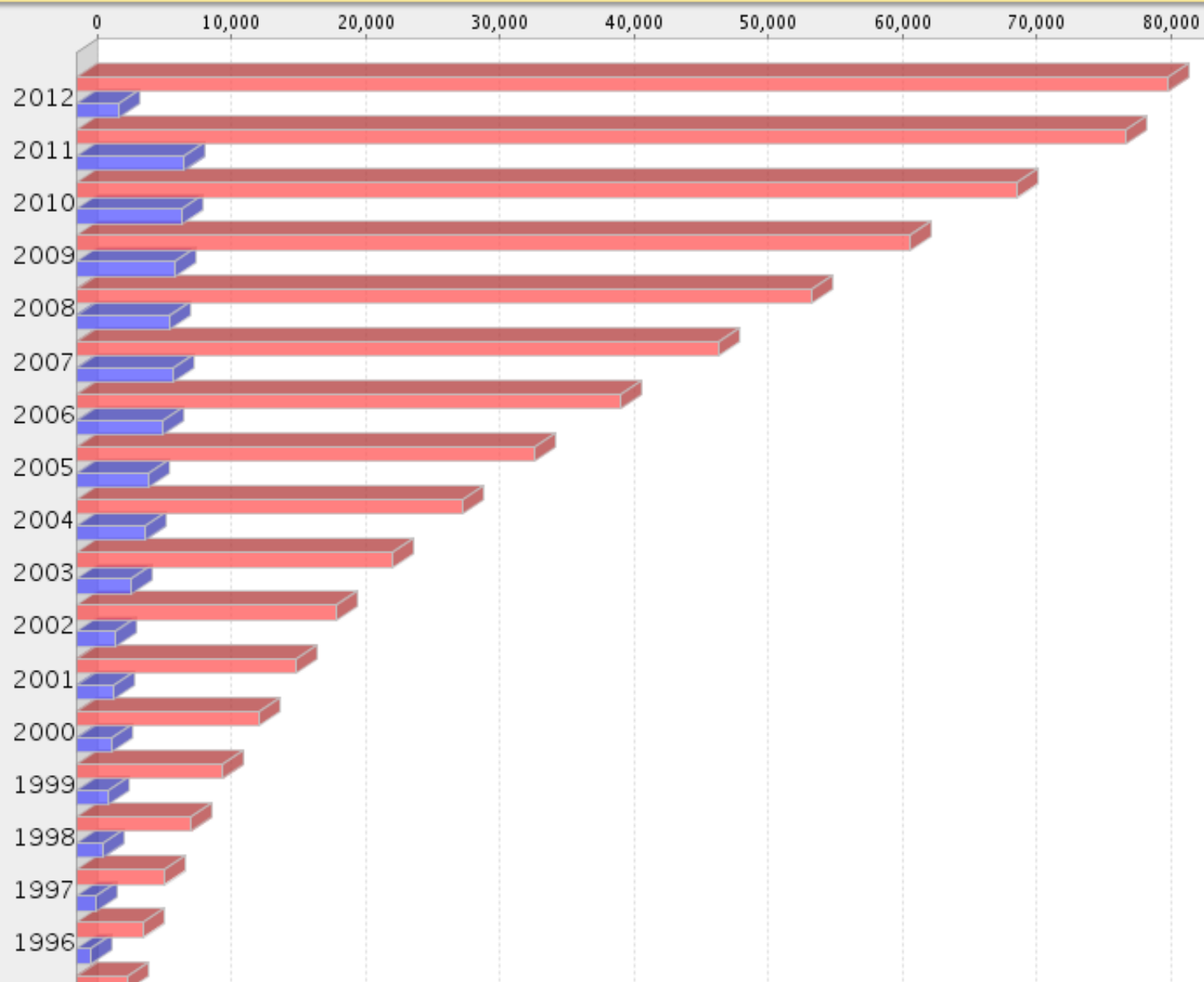
◆ グラフィック表示

- PDBでのイメージ
- 表示ツールを用いる方法
 - ◆ RasMol、Jmol、Chime、Protein Explorer、VRMLなど



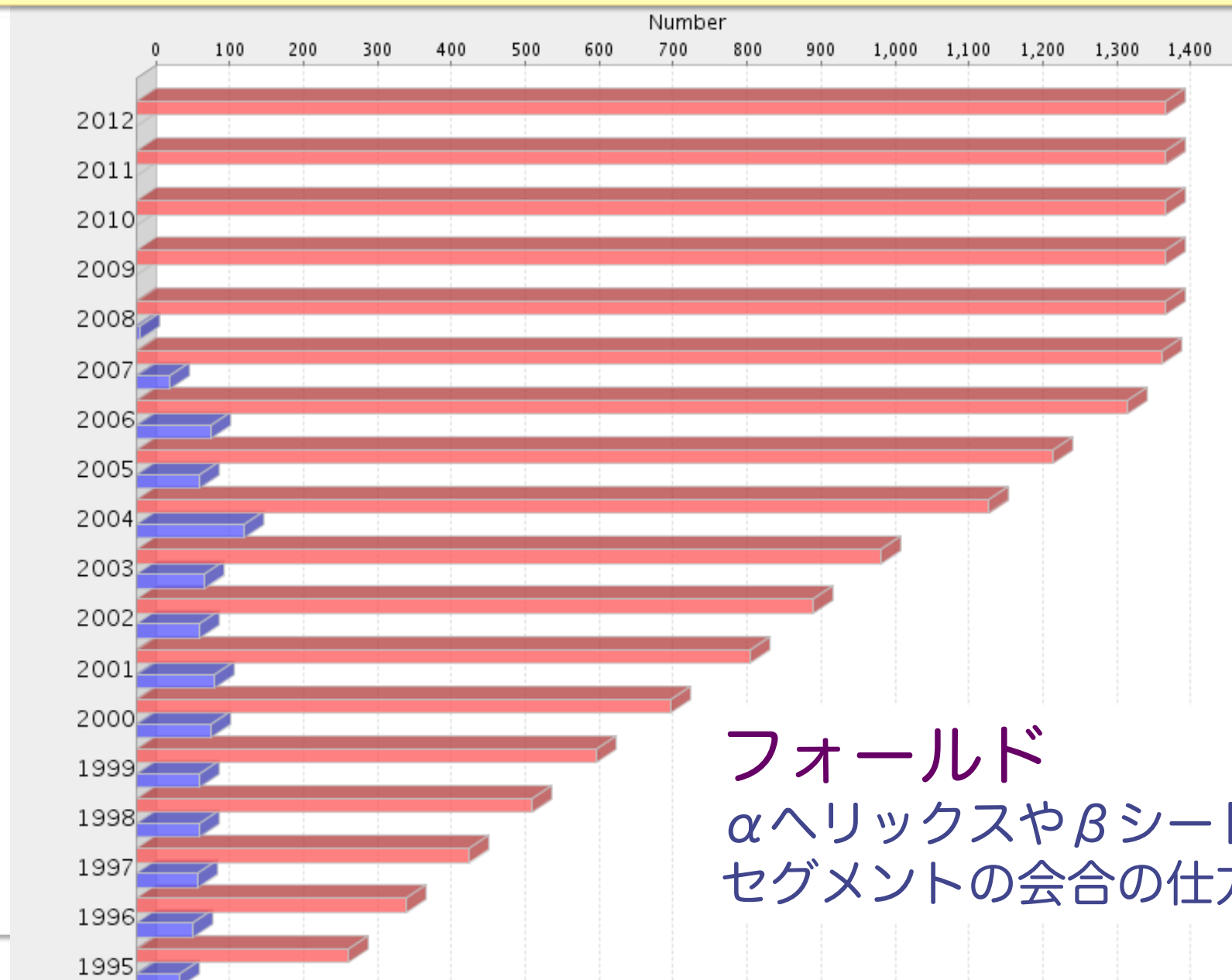
ミオグロビンの例

PDB エントリー数の推移



<http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100> より

PDB フォールド数の推移



フォールド
 α ヘリックスや β シートの
セグメントの会合の仕方

フォールド

- ◆ α や β のセグメントの会合の仕方
 - ヘリックスループヘリックス
 - β ヘアピン
 - フォーヘリックスバンドル
 - グリークキー
 - β バレル
 - β / α (TIM) バレル

タンパク質の分類と立体構造

◆役割による分類

- 構造タンパク質
- 機能性タンパク質

◆形状と環境による分類

- 球状タンパク質 (globular protein)
- 繊維状タンパク質 (fibrous protein)
- 膜タンパク質 (membrane protein)

球状・膜タンパク質の特徴

◆球状タンパク質

- 内部に疎水性 (hydrophobic) アミノ酸
- その周りを親水性 (hydrophilic) アミノ酸

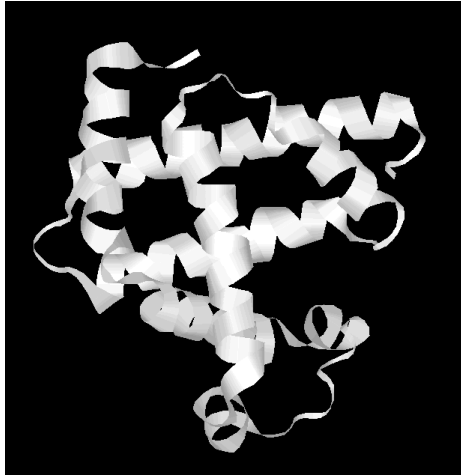
◆膜タンパク質

- 内部に親水性アミノ酸 (チャネルなど)
- 外部に疎水性アミノ酸 (脂質と接触)
- 両親媒ヘリックス (amphipathic helix)
 - ◆ 3.6残基ごとに疎水性・親水性が周期的に出現

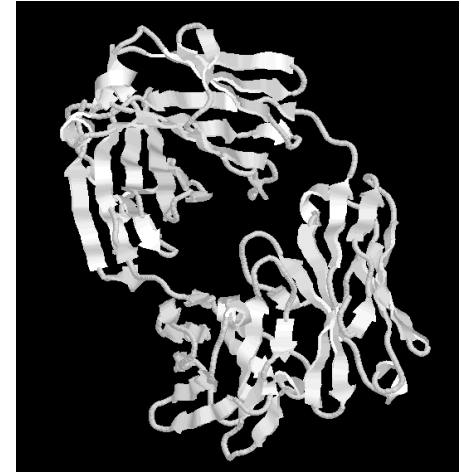
二次構造の組成による球状タンパク質の分類

- ◆ All α 型タンパク質
 - 主に α ヘリックスからなる
- ◆ All β 型タンパク質
 - 主に β シートからなる
- ◆ α/β 型タンパク質
 - α ヘリックスと β シートが入り交じった構成
- ◆ $\alpha+\beta$ 型タンパク質
 - α ヘリックスと β シートが分離した構成
- ◆ ランダム
 - 二次構造を特に持たない

二次構造の組成による球状タンパク質の分類



All α 型の例
ミオグロビン
PDB:1MBN



All β 型の例
免疫グロブリン
PDB:7FAB



α/β 型の例
トリオースリン酸
イソメラーゼ
PDB:1TIM



$\alpha + \beta$ 型の例
リボヌクレアーゼA
PDB:7RSA

フォールド分類データベース

◆ フォールド分類の必要性

- 立体構造と機能との間には密接な関係
- 配列が似ていなくても構造類似のタンパク質が多数存在

◆ SCOP: Structural Classification of Proteins

- <http://scop.mrc-lmb.cam.ac.uk/scop/>
- クラス：二次構造の組成に基づく分類
- フォールド：構造の類似性
- スーパーファミリー：進化的類縁性
- ファミリー：明らかな進化的類縁性

◆ CATH: <http://www.cathdb.info/>

- クラス(C)、アーキテクチャ(A)、トポロジー (T)、スーパーファミリー (H)

◆ Dali Domain Dictionary: 現在は PDBeFold

- <http://www.ebi.ac.uk/msd-srv/ssm/>

立体構造アライメント

- ◆ 差がなるべく小さくなるように2つの構造を並べる
 - 考え方は配列アライメントと同じ。
 - 必ずしも配列類似性や進化的な関係が保存されているとは限らない。
 - 構造が類似でも2次構造の順序が保存されているとは限らない。
- ◆ 2次構造情報をベクトル表現して比較
 - ◆ (position, length, direction)
- ◆ Double Dynamic Programming
- ◆ アミノ酸間の対応関係が分かっている場合には
RMSD (root mean square deviation) を用いて最適な重ね合わせを定義する。

関連 URL

◆ 立体構造データベース (PDB)

- <http://www.rcsb.org/>
- http://www.genome.jp/dbget-bin/www_bfind?pdb

◆ フォールド分類

- SCOP: <http://scop.mrc-lmb.cam.ac.uk/scop/>
- CATH: <http://www.cathdb.info/>

◆ 構造予測

- ExPasy リンク集: <http://www.expasy.ch/tools/#tertiary>
- 立体構造予測コンテストCAFASPにおける自動サーバー
 - ◆ <http://www.cs.bgu.ac.il/~dfischer/CAFASP3/servers.html>

◆ 立体構造表示

- Jmol: <http://jmol.sourceforge.net/>