東京大学大学院農学生命科学研究科
アグリバイオインフォマティクス人材養成プログラム
バイオインフォマティクスリテラシーI
平成19年5月28日(月)、31日(木)　＠農学部2号館化学第一講義室

立体構造予測　I [Web版]

フォールディング問題、構造分類、構造比較、相同性検索など

東京大学大学院農学生命科学研究科
アグリバイオインフォマティクス人材養成ユニット
特任助教

古田　忠臣

講義の予定

- 5月28日(月)、31日(木)
  - 構造データベース：PDB
  - 構造分類データベース：SCOP、CATH
  - 構造比較サーバー：CE、DALI/FSSP、VAST
  - 相同性検索：BLAST、PSI-BLAST、FASTA、CLUSTALW

構造類似性
配列類似性
1D検索

- 6月4日(月)、7日(木)

2D予測

  - 二次構造予測：PSIPRED、PHDsec、PREDETOR、NPS@

3D予測

  - 立体構造予測 ・・・参考： CASP
    - 比較モデリング法
      - ホモロジーモデリング：**MODELLER、SWISS-MODEL**
    - フォールド認識法：meta server (3D-Jury)
    - de novo / ab initio予測法： Robettaなど

分子生物学のセントラルドグマ

Webで顔写真を探して下さい。

- F.H.C.Crick 1958

F. Crick

遺伝的情報は以下の様に一方向的に伝達される

（複製）

DNA {A,T,G,C}

↓（転写）・・・メチル化

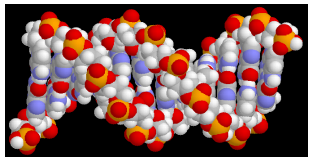
RNA {U,A,C,G}

↓（翻訳）・・・スプライシング

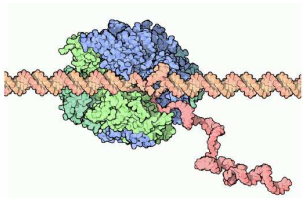
タンパク質 {A,C,D,,,Y}

（フォールディング問題）

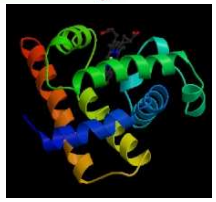
（レトロウィルスの逆転写酵素：RNA→DNA）



DNA
Watson & Crick, 1953
（ヒトゲノム：約30億塩基対、染色体：23対）
(32億5,400万bp)



（RNAポリメラーゼ）



Myoglobin
PDB:1MBN
Kendrew, 1960
（ヒト遺伝子：約3万(26,808)）
（リボソーム、tRNA）

F.H.C. Crick, *Symp. Soc. Exp. Biol.* **12**, 138-163 (1958), “On protein synthesis”

[URL] http://www.lif.kyoto-u.ac.jp/genomemap/

現在300種以上のゲノムが解読されている

Anfinsenのドグマ　1973

- タンパク質の天然構造は熱力学的に最も安定であり、「一次構造が決まれば立体構造も決まる」
  - 変性しても、元の生理的条件に戻すと再び折り畳まる
  - 現在では、巨大なタンパク質やミスフォールドしたタンパク質は分子シャペロンの助けを借りて折り畳まることが知られている
    ・・・β アミロイド形成　→　病気

Webで顔写真を探して下さい。

C. Anfinsen

$$dF = -SdT - PdV + \sum_i \mu_i dn_i$$

C.B. Anfinsen, *Science* **181**, 223-230 (1973), “Principles that govern the folding of protein chains ”

Levinthalのパラドックス　1969

Webで顔写真を探して下さい。

C. Levinthal

- ランダム探索では、天然構造への折り畳むまでに天文学的時間が掛かる
  - 各アミノ酸が3つのconformationsを持つとして、150残基のタンパク質の場合、可能はconformationsは3^{150}~10^{71}ある。一回のconformation変化に10^{-13}s掛かるとして、全探索には$10^{71} \times 10^{-13}\text{s} = 10^{58}\text{s} \sim 10^{50}$年掛かる。
  - しかし、実際のタンパク質は数ms程度で折れたたまる・・・ 10^{8}回程度の探索しかしていない（パラドックス）
  - 特定のFolding pathwaysがあるのではないか。

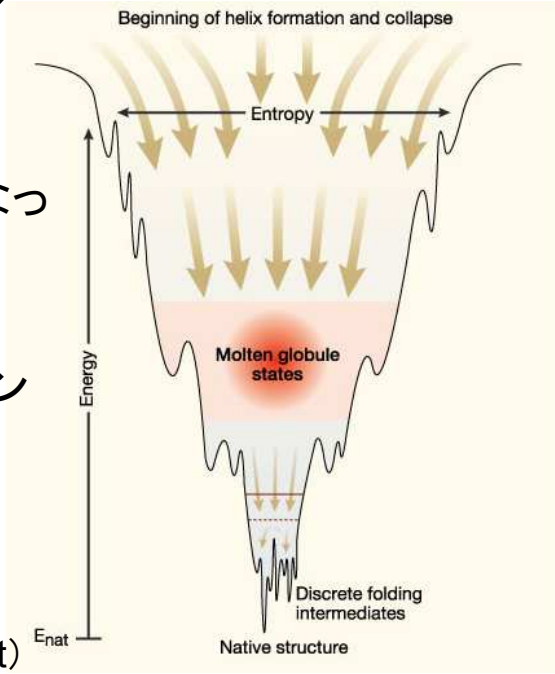
C. Levinthal, *Mössbaun Spectroscopy in Biological Systems Proceedings* **67**, 22-24 (1969), “How to fold graciously”

エネルギー・ランドスケープ理論

Webで顔写真を探して下さい。

N. Go　　P.G. Wolynes

- タンパク質のエネルギー地形はファネル状の形をとる
  - GoのConsistency Principle 1984
    - タンパク質は、天然状態で様々な相互作用が最適になっている（進化の所産）。
  - Wolynesらのフラストレーション最小原理 1987
    - タンパク質は、天然状態でエネルギーフラストレーションを最小にする様に設計されている。エネルギー地形はファネル状の形をとる。



(フォールディング問題：モルテングロビュール状態(K.Kuwajimaら)
遷移状態（H.A.Kramers）・・・φ値解析（A.Fersht）
2状態・3状態（H.A.Sheragaら））

N. Go, *Adv. Biophys.* **18**, 149-164 (1984), “The consistency principle in protein structure and pathways of folding”
J.D. Bryngelson, and P.G. Wolynes, *PNAS* **84**, 7524-7528 (1987), “Spin glasses and the statistical mechanics of protein forlding”

分子動力学の講義 → 分子モデリングと分子シミュレーション(寺田先生)
量子化学の講義 → 量子化学入門と分子軌道法(岩岡先生)

分子シミュレーションの簡単な歴史(手法の改良およびQMは除く)

1953 最初のMonte Carlo(MC)シミュレーション	Metropolisら
1957 最初のMolecular Dynamics(MD)シミュレーション	**Alder & Wainwright**
1969 液体(水)のMC	Barker & Watts
1971 液体(水)のMD	Rahman & Stillinger
1973 ヌクレオチドの真空中でのエネルギー最小化(GpC)	Stellmanら
(以下、タンパク質)	
1971 タンパク質のエネルギー最小化(Lysozyme)	Levitt & Lifson
1977 最初のタンパク質の真空中でのMD(BPTI, 10ps)	**McCammon, Gelin & Karplus**
1981 AMBER	Kollman, Caseら
1982 最初のタンパク質の溶液中でのMD(BPTI, 25ps)	van Gunsteren & Karplus
1983 CHARMM	Karplus, Brooksら
1990 GROMOS/GROMACS	van Gunsteren, Berendsenら
1996 NAMD	Schultenら
1997 世界最長時間!?のフォールディングのMD (villin headpiece(36aa, 4,000atoms), 1μs (2 months cpu time))	**Duan & Kollman**
2000 Folding@Home like SETI !?	Pandeら
2001 K^+チャンネルのMD (40,000 atoms, 38ns)	BerneÁche & Roux
2001 アクアポリンのMD (101,449 atoms, 10ns)	de Groot & Grubmüller
2002 F1-ATP synthaseのMD (183,674 atoms, 7ns (~400 months cpu time))	Böckmann & Grubmüller

スペースの都合上、
文献情報がないですが
必要でしたらご連絡下さい。

小タンパク質以外のフォールディングのMDは困難

タンパク質立体構造データベース: PDB – Protein Data Bank

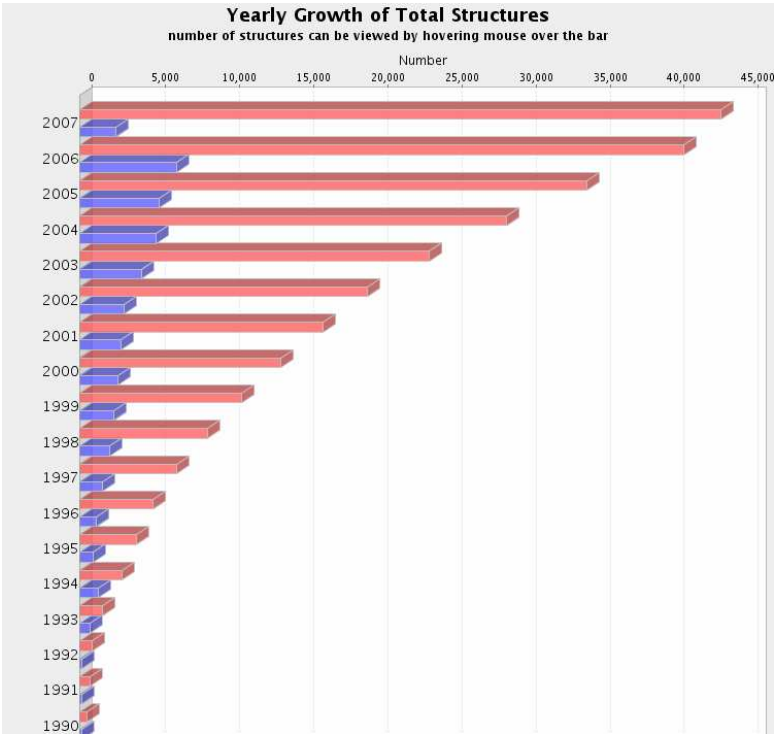
wwPDB [URL] http://www.wwpdb.org RCSB PDB [URL] http://www.rcsb.org/pdb/
MSD EBI [URL] http://www.ebi.ac.uk/msd/
PDBj [URL] http://www.pdbj.org/

- 現在(May/08/2007)、43,339構造が登録されている

PDB ID:4文字の英数字　ex) 1AB1

PDB Current Holdings Breakdown

		Molecule Type				
		Proteins	Nucleic Acids	Protein/NA Complexes	Other	**Total**
Exp. Method	X-ray	34259	964	1581	28	**36832**
	NMR	5375	760	129	7	**6271**
	Electron Microscopy	101	10	38	0	**149**
	Other	80	4	3	0	**87**
	Total	**39815**	**1738**	**1751**	**35**	**43339**



H.M. Berman, et al., *Nucleic Adids Res.* **28**, 235-242 (2000), “The Protein Data Bank”

Chothiaのフォールド数の見積もり

Webで顔写真を探して下さい。

C. Chothia

- Chothia 1992
  - タンパク質は約1000ファミリーしかない・・・今で言うフォールド数
    - 新規配列の1/3が既知配列と相同的
    - 既知配列の1/4が既知120ファミリーに属する
    - 統計的偏りがなければ、120×4×3～1,500ファミリー
  - 構造（フォールド）は配列よりもよく保存されている
    - 新規フォールドの割合は収束に向かっていない、、、CASP6 (vs. CASP5)
      CASP7では！？

C. Chothia, *Nature* **357**, 543-544 (1992), “One thousand families for the molecular biologist”

タンパク質立体構造分類データベース: SCOP – Structural Classification of Proteins

[URL] http://scop.mrc-lmb.cam.ac.uk/scop/

→ top of the hierarchy

スーパーファミリー: 機能・構造的特徴から恐らく共通の進化的起源
例）アクチン、ヒートショックタンパクのATPase、ヘキサキナーゼ

階層:

クラス、フォールド、スーパーファミリー、ファミリー

ex) sccs: a. 1. 1. 1

All alpha proteins. Globin-like. Globin-like. Truncated hemoglobin

- 2005年時点での27,599PDBが75,930ドメインに分割され、

  現在、約1,000（971）のフォールドが登録されている(1.71)

  Murzinが中心となり、

  人の手・目！？で分類されている

All alpha　All beta　a/b　a+b

Scop Classification Statistics

SCOP: Structural Classification of Proteins. **1.71** release
27599 PDB Entries (18 Jan 2005). 75930 Domains. 1 Literature Reference
(excluding nucleic acids and theoretical models)

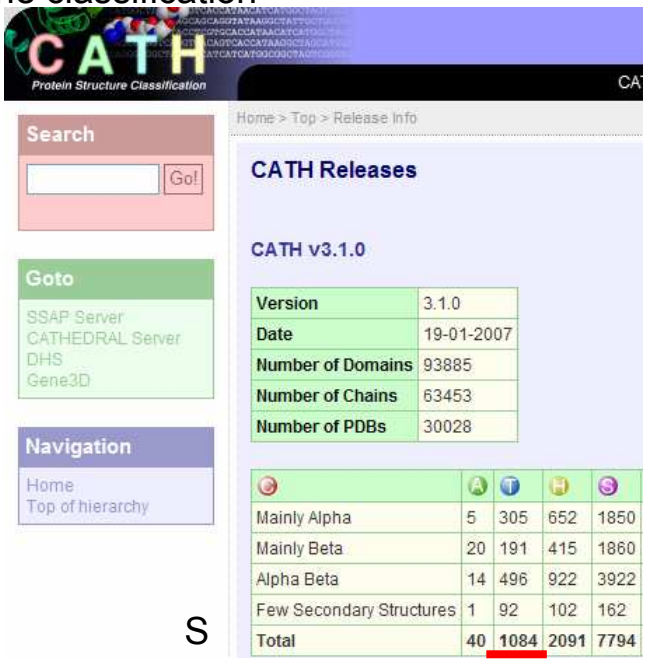
Class	Number of folds	Number of superfamilies	Number of families
All alpha proteins	226	392	645
All beta proteins	149	300	594
Alpha and beta proteins (a/b)	134	221	661
Alpha and beta proteins (a+b)	286	424	753
Multi-domain proteins	48	48	64
Membrane and cell surface proteins	49	90	101
Small proteins	79	114	186
Total	971	1589	3004

A.G. Murzin, et al., *J. Mol. Biol.*. **247**, 536-540 (1995),
“SCOP: a structural classification of proteins database for the investigation of sequences and structures”

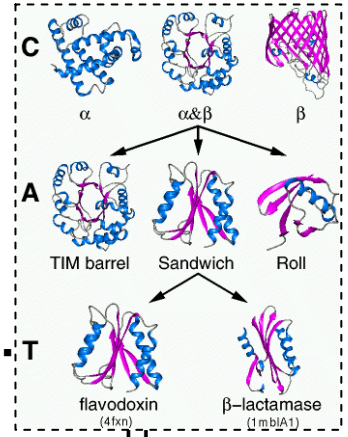
タンパク質立体構造分類データベース: CATH – Class, Architecture, Topology, Homologous superfamily

[URL] http://cathwww.biochem.ucl.ac.uk/latest/

→ Browse or search the classification



- 現在、1,084　topologies（folds）が登録されている (v3.1.0)
- かなり自動的に分類されるが、最後は人手



Fold・・・

階層　　C　　A　　T　　H　　S

クラス、アーキテクチャー、トポロジー、ホモロガススーパーファミリー、シークエンスファミリー

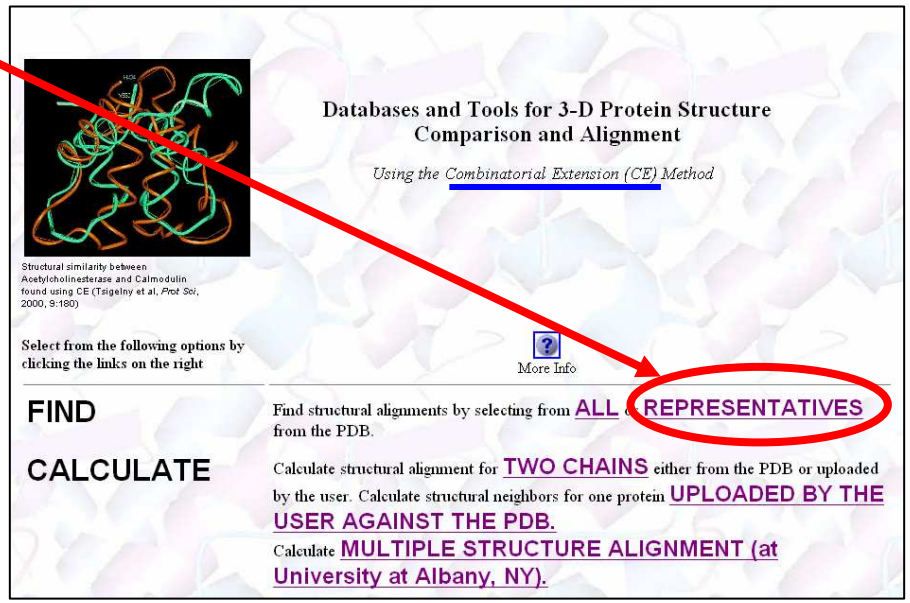
ex) CATH code: 1.　10.　8.　10.　1

1. Mainly Alpha
10. Orthogonal Bundle
8. Helicase,,,
10. DNA helicase RuvA subunit,,,
1. DNA helicase Ruv subunit,,,

C.A. Orengo, et al., *Structure* **5**, 1093-1108 (1997), “CATH – A Hierarchic Classification of Protein Domain Structures”

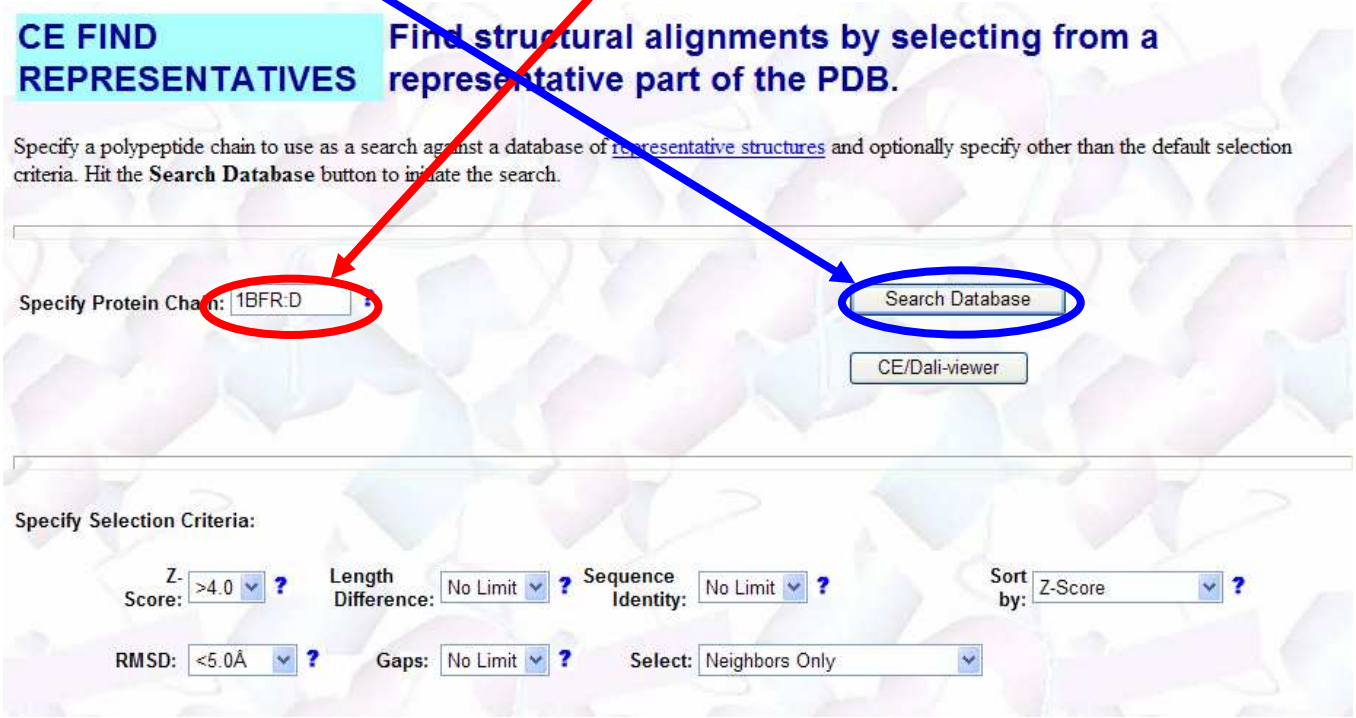
タンパク質立体構造比較サーバー1：CE

- CEサーバーで構造比較したアラインメントを得て、DS1.7で実際の構造を見てみましょう
- CEのホームページ[URL] http://cl.sdsc.edu/ を開く
- 「REPRESENTATIVES」をクリック

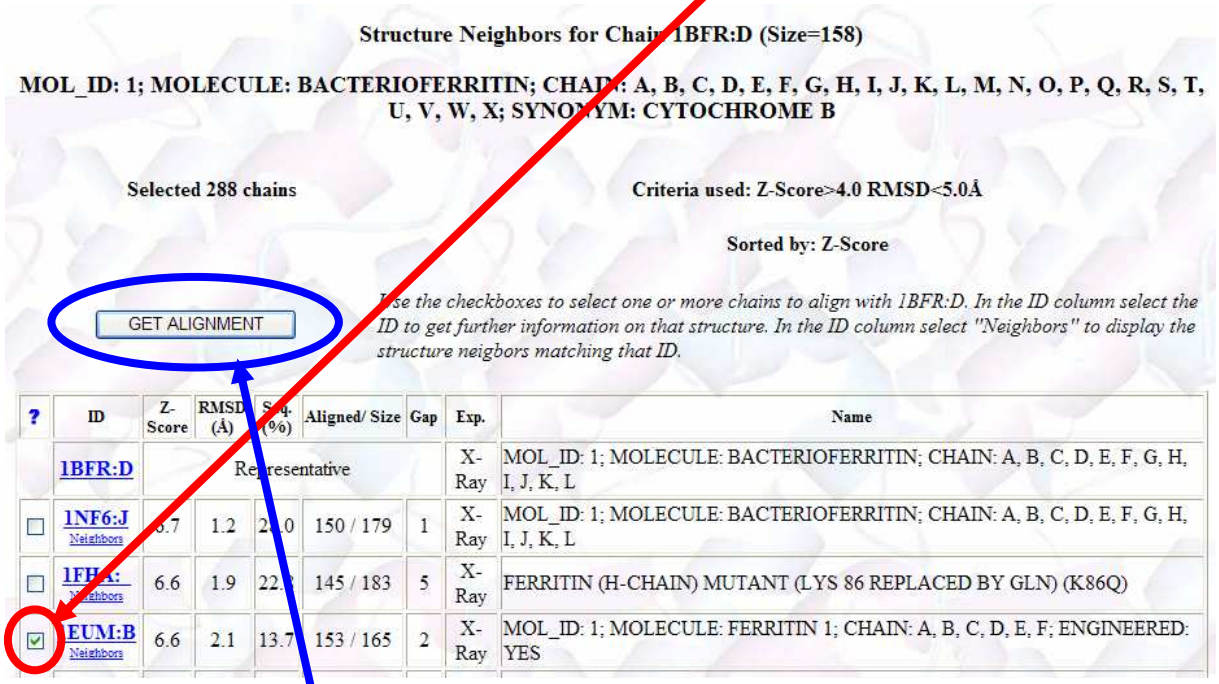


I.N. Shindyalov, P.E. Bourne, *Protein Engineering* **11**, 739-747 (1998), “Protein structure alignment by incremental combinatorial extension (CE) of the optimal path”

- Specify Protein Chainに「1BFR:D」を入力
- 「Search Database」をクリック



- Seq.(%)が13.7の「1EUM:B」をチェック（1NF6のチェックをはずす）
- 配列の一致度が13.7%なのに、RMSD=2.1 Å



Structure Neighbors for Chain 1BFR:D (Size=158)

MOL_ID: 1; MOLECULE: BACTERIOFERRITIN; CHAIN: A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X; SYNONYM: CYTOCHROME B

Selected 288 chains **Criteria used: Z-Score>4.0 RMSD<5.0Å**

Sorted by: Z-Score

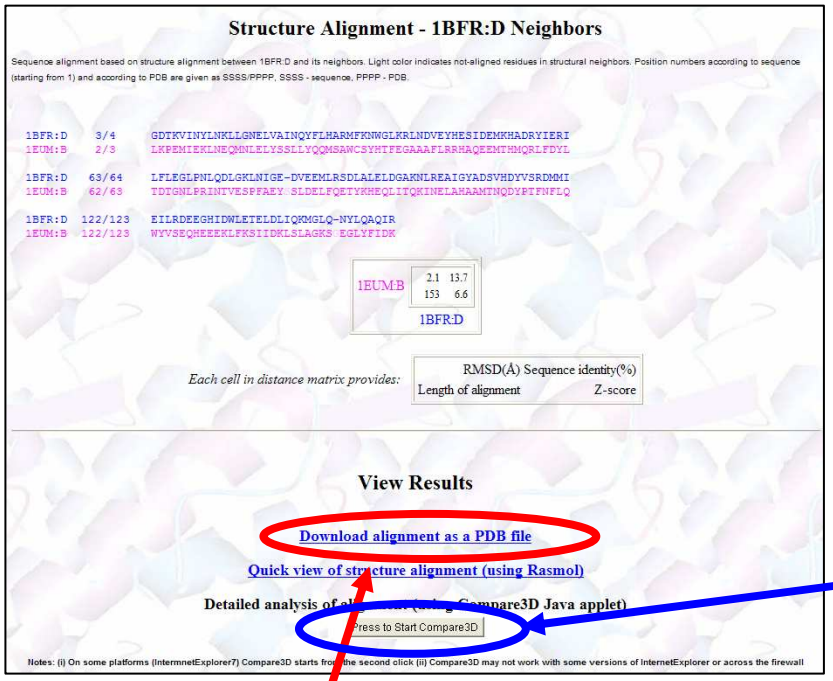
GET ALIGNMENT

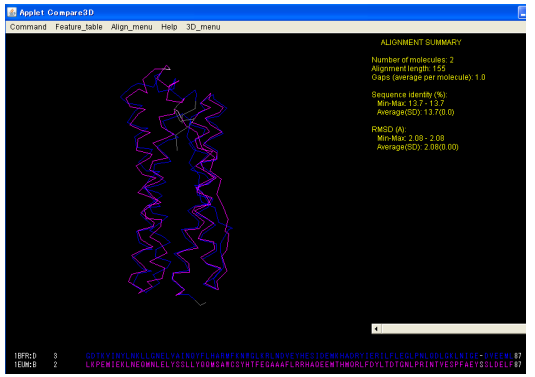
Use the checkboxes to select one or more chains to align with 1BFR:D. In the ID column select the ID to get further information on that structure. In the ID column select "Neighbors" to display the structure neigbors matching that ID.

?	ID	Z-Score	RMSD (Å)	Seq. (%)	Aligned/ Size	Gap	Exp.	Name
	1BFR:D	Representative					X-Ray	MOL_ID: 1; MOLECULE: BACTERIOFERRITIN; CHAIN: A, B, C, D, E, F, G, H, I, J, K, L
☐	1NF6:J Neighbors	6.7	1.2	21.0	150 / 179	1	X-Ray	MOL_ID: 1; MOLECULE: BACTERIOFERRITIN; CHAIN: A, B, C, D, E, F, G, H, I, J, K, L
☐	1FHA: Neighbors	6.6	1.9	22.8	145 / 183	5	X-Ray	FERRITIN (H-CHAIN) MUTANT (LYS 86 REPLACED BY GLN) (K86Q)
☑	1EUM:B Neighbors	6.6	2.1	13.7	153 / 165	2	X-Ray	MOL_ID: 1; MOLECULE: FERRITIN 1; CHAIN: A, B, C, D, E, F; ENGINEERED: YES

- どれほど似ているか構造を見るために、ページ上の「GET ALIGNMENT」をクリックしアラインメントを得ましょう

- 構造比較したアラインメントが得られます。



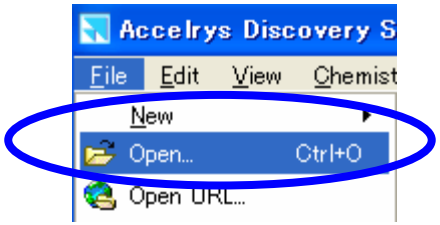


Press to Start Compare 3Dを
クリックすると、Java Appletで
比較した構造を見ることも出来ます。

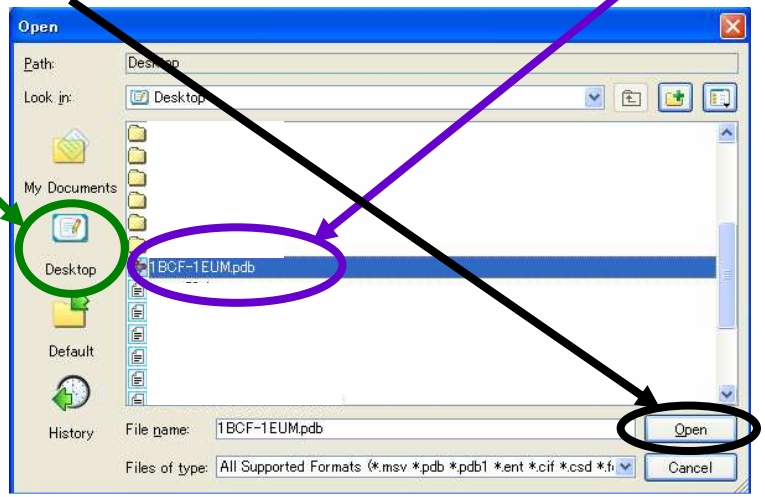
- 構造をダウンロードする
  - 「Download alignment as a PDB file」を右クリック
  - →「対象をファイルに保存」
  - デスクトップに「リテラシーI」などのフォルダを作成し、ファイル名「1BFR-1EUM.pdb」、ファイルの種類「すべてのファイル」で保存

- 「Discovery Studio 1.7」を起動（ダブルクリック）

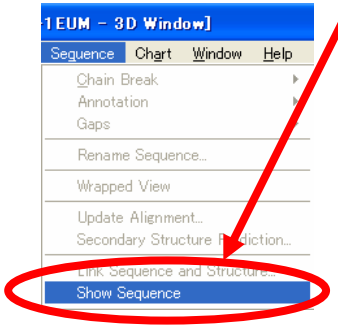
- 「File」メニュー→「Open…」を選択



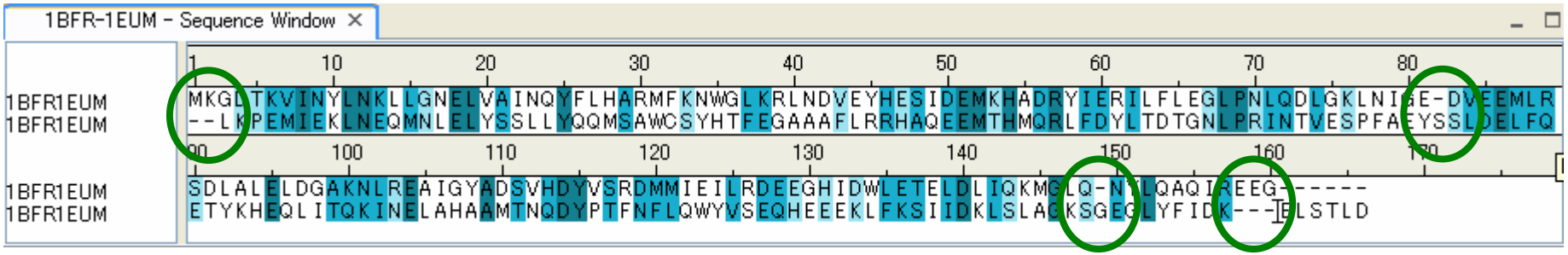
- Desktopをクリックし「リテラシーI」フォルダから、ファイル「1BCF-1EUM.pdb」を選択し、「Open」をクリック



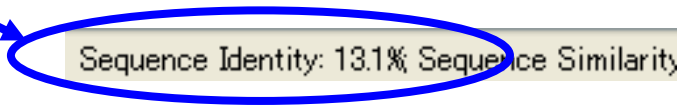
- 「Sequence」メニュー→「Show Sequence」を選択し、配列を表示



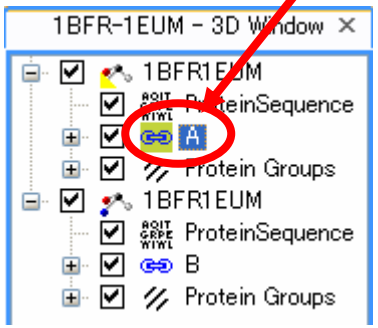
- カーソルをGapの部分にして、「Space」を入力し、先程のアラインメントに合わせてみましょう



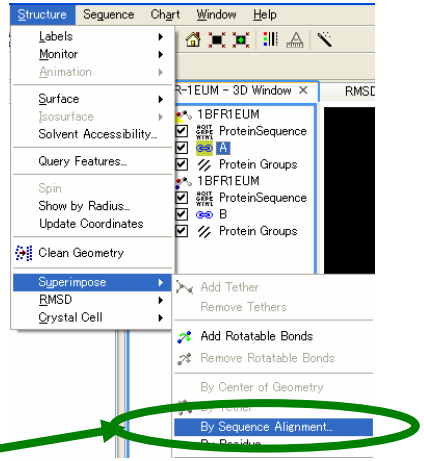
濃い緑が一致した残基を表しています。
Windowの右下に配列の一致度13.1%が表示されています



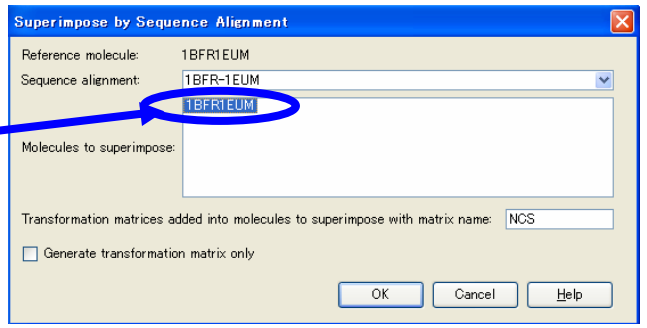
- 既に構造アラインメントされていますが、今編集したアラインメントを基に2つの構造を重ね合わせ(superimpose)、RMSDを計算してみましょう
  - 「3D Window」で1BFRのA chainを選択



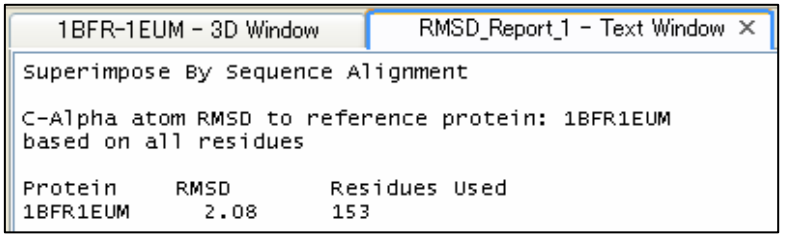
  - 「Structure」
    →「Superimpose」
    →「By Sequence Alignment」を選択



  - Windowが開くので
    →「1BCF1EUM」を選択し、
    →「OK」をクリック

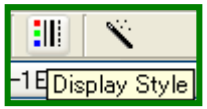


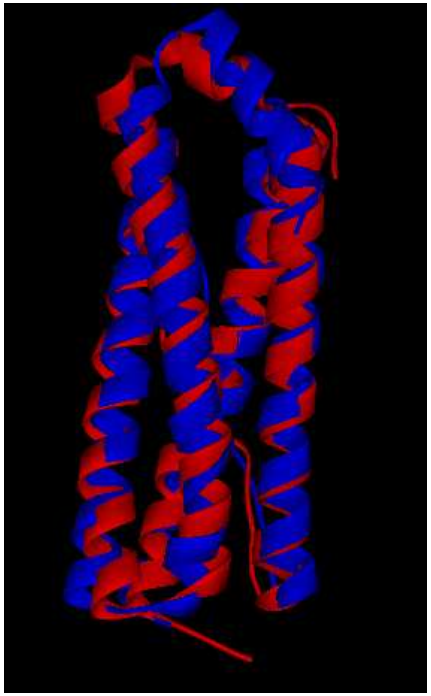
- 「Text Window」が表示され、
  - 153残基を使って、
  - RMSD＝2.08と計算されました。



- 同時に「3D Window」の構造は重ねあわされています。

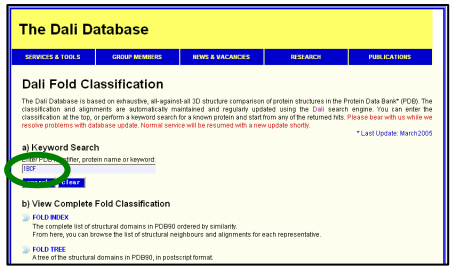
- では、どれくらい似た構造か、表示（Display Style）を変えてみてみましょう。
  - A chain(1BFR:D)を選択し、
    →「Ctrl」＋「D」（ショートカット）・・・ボタンをクリックと同じ
    - Atom: None
    - Protein: Solid ribbon
      - Coloring－Custom:赤
  - B chain(1EUM:B)を選択し、
    →「Ctrl」＋「D」（ショートカット）
    - Atom: None
    - Protein: Solid ribbon
      - Coloring－Custom:青

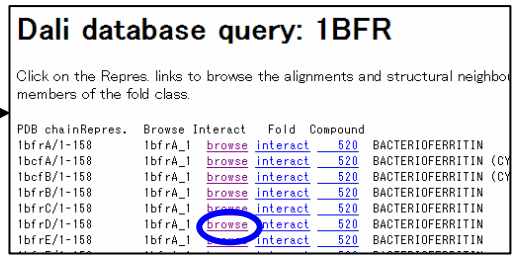


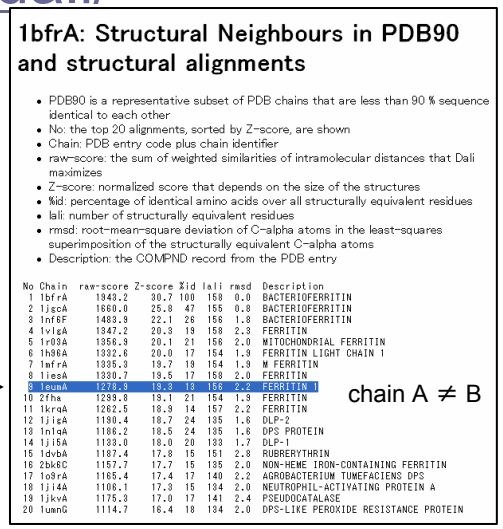


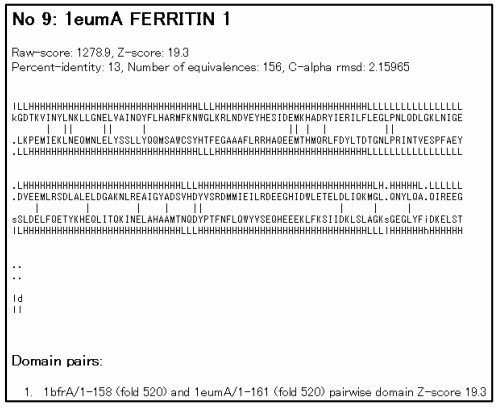
タンパク質立体構造比較サーバー2: DALI / FSSP

- [URL] http://ekhidna.biocenter.helsinki.fi/dali/
  - →Keyword Search: 1BFR → (1bfrD/1-158)browse







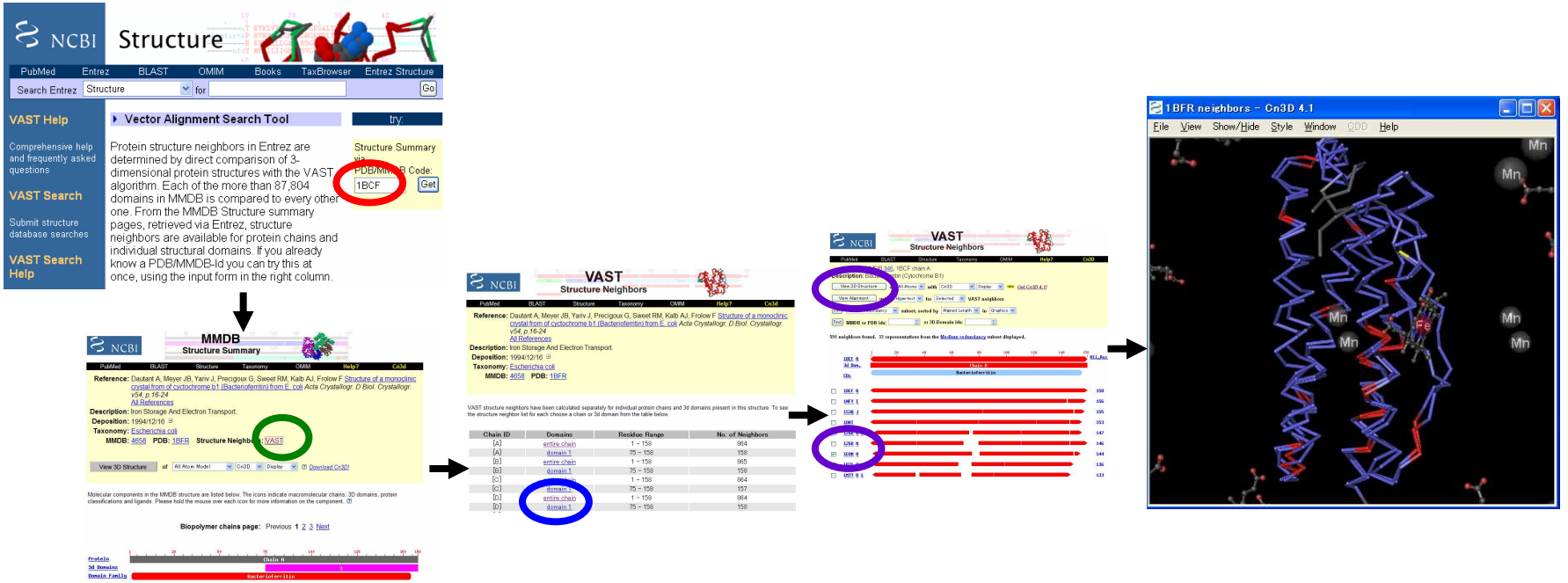


DALIサーバー(http://www.ebi.ac.uk/dali/)では
PDBにない、新しく自分が決めた構造をデータベースと
比較したり、2つの構造をアラインメントしたりもできます。
（CEやVASTでももちろんできます）

L. Holm, C. Sander, *J. Mol. Biol.* **233**, 123-138 (1993), “Protein structure comparison by alignment of distance matrices”

タンパク質立体構造比較サーバー3: VAST

- [URL] http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml
  - →try: 1BFR → VAST→[D]entire chain→1EUM Aをチェック (Cn3D 4.1をインストール)→View 3D Alignment



T. Madej, J.F. Gibrat, S.H. Bryant, *Proteins* **23**, 356-369 (1995), “Threading a database of protein cores”

タンパク質のフォールド空間へのマッピング

- Kimらは、SCOPの498フォールドをDALIのスコアを基に計量行列を作り、クラスター解析することにより、フォールド空間へマッピングしました。

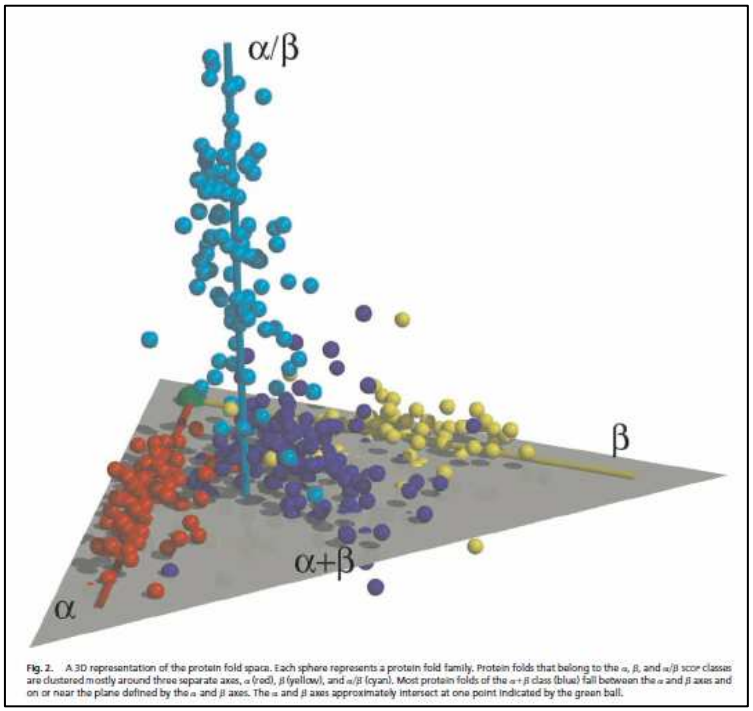


Fig. 2. A 3D representation of the protein fold space. Each sphere represents a protein fold family. Protein folds that belong to the α, β, and α/β SCOP classes are clustered mostly around three separate axes, α (red), β (yellow), and α/β (cyan). Most protein folds of the α+β class (blue) fall between the α and β axes and on or near the plane defined by the α and β axes. The α and β axes approximately intersect at one point indicated by the green ball.

Webで顔写真を探して下さい。

S.-H. Kim

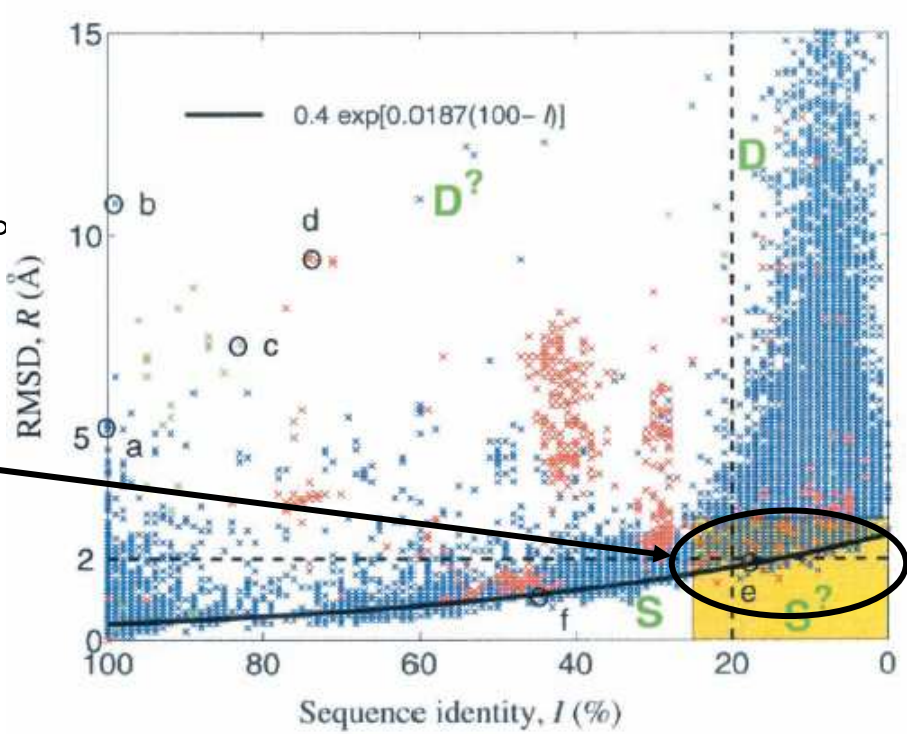
J. Hou, G.E. Sims, C. Zhang, S.-H. Kim, *PNAS* **100**, 2386-2390 (2003), “A global representation of the protein fold space”

配列類似性と構造類似性の関係

- アラインした53,383タンパク質ペアの RMSD vs. 配列一致度

- 配列一致度が20%以上のものはほとんど（×）RMSD＜3Åに入っています。
  - カルモジュリン（×）、イミュノグロブリン（×）を除く

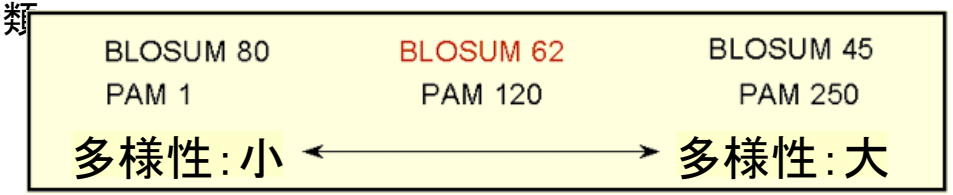
- 配列一致度が20%以下でもRMSD＜3Åのものも多くある
  →配列より構造の方が保存されている



H.H. Gan et al., *Biophys. J.* **83**, 2781-2791 (2002), “Analysis of Protein Sequence/Structure Similarity Relationships”

相同性検索の簡単な歴史

- 1970 **ドットマトリックス** Gibbs-McIntyre・・・フィルタリングし一致配列を可視化
  - ダイナミックプログラミング（DP、動的計画法）・・・最適なアラインメントを検索
- 1970 **Needlemen-Wunschのアルゴリズム**・・・グローバルDP
- 1978 **PAM行列** Dayhoffら・・・進化率(受け入れられた変異率)に基づくアミノ酸置換行列
  - PAM1・・・平均1%の配列上の位置が置換する時間
- 1981 **Smith-Watermanのアルゴリズム**・・・ローカルDP
- 1985 **FASTA** Lipman & Pearson
- 1990 **BLAST** Altschul et al.
- 1992 **BLOSUM行列** Henikoffら・・・ファミリーのアラインメントに基づくアミノ酸置換行列
  - BLOSUM62・・・62%類似の配列を分類
- 1997 **PSI-BLAST** Altschul et al.



A.J. Gibbs, G.A. McIntyre, *Eur. J. Biochem.* **16**, 1-11 (1970), “The diagram, a method for comparing sequences”
S.B. Needleman, C.D. Wunsch, *J. Mol. Biol.* **48**, 443-453 (1970),
“A general method applicable to the search for similarities in the amino acid sequence of two proteins”
M.O. Dayhoff et al., In *Allas of Protein Sequence and Structure*, Chap. **22** (1978), “A model of evolutionary change in proteins”
T.F. Smith, M.S. Waterman, *J. Mol. Biol.* **147**, 195-197 (1981), “Identification of common molecular subsequences”
D.J. Lipman, W.R. Pearson, *Science* **227**, 1435-1441 (1985), “Rapid and sensitive protein similarity searches”
S.F. Altschul, et al., *J. Mol. Biol.*. **215**, 403-410 (1990), “Basic local alignment search tool”
S. Henikoff, J.G. Henikoff, *PNAS* **89**, 10915-10919 (1992), “Amino acid substitution matrices from protein blocks”
S.F. Altschul, et al., *Nucleic Adids Res.*. **25**, 3389-3402 (1997),
“Gapped BLAST and PSI-BLAST: a new generation of protein database search programs”

相同性検索：BLAST, PSI-BLAST

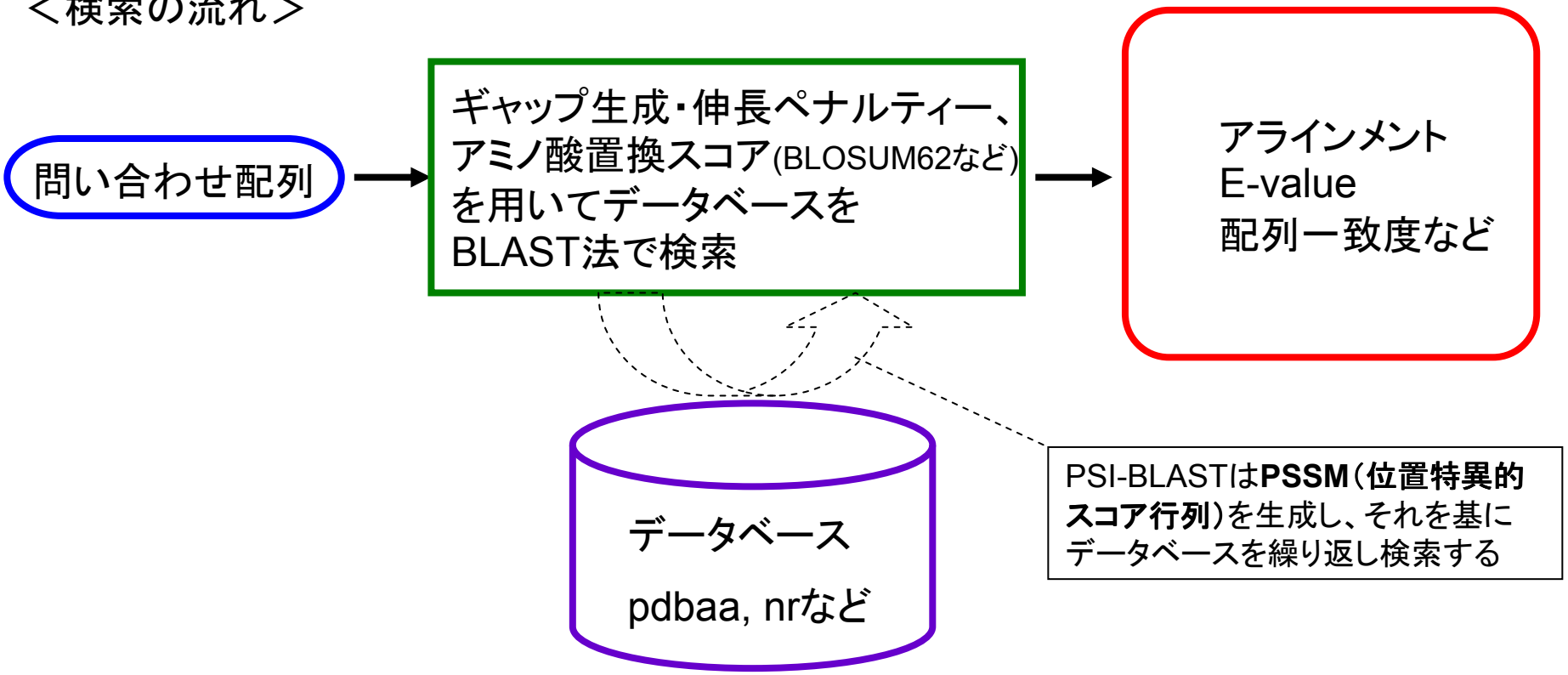
Webで顔写真を探して下さい。

S. Altscul

[URL] http://www.ncbi.nlm.nih.gov/BLAST/

- 1990/1997 Altschulらは、Smith-Waterman のアルゴリズムを改良し、局所的な類似部分配列を高速検索する手法を開発した

＜検索の流れ＞

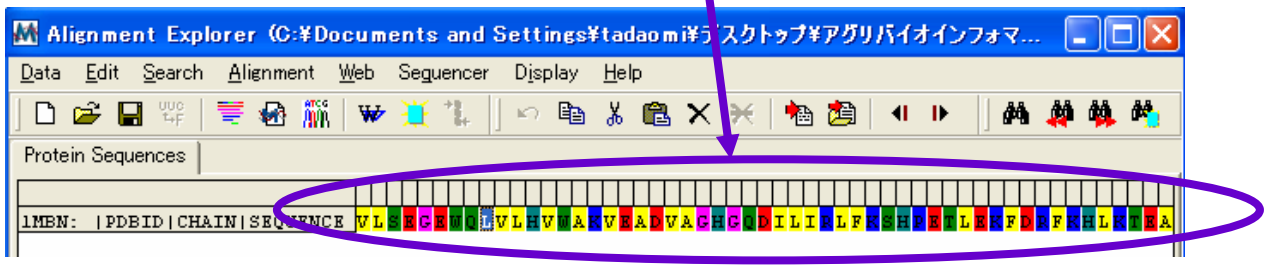


PDBサイトで配列を得る

- PDBサイト([URL] http://www.rcsb.org)を開き、「1MBN」を入力、検索(SEARCH)

- PDBの情報が表示されるので、「FASTA Sequence」をクリックし、ファイル名「1MBN.fasta」で保存

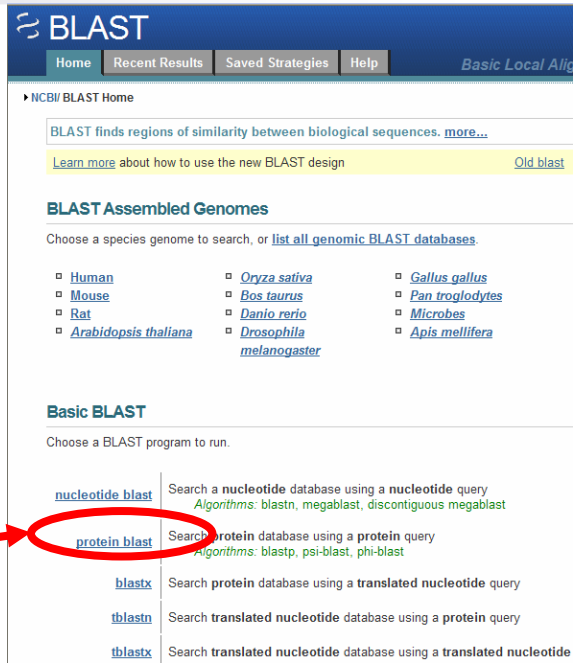
- ファイル「1MBN.fasta」をダブルクリックで開き、Ctrl+A、Ctrl+Cでfasta形式の配列をコピー



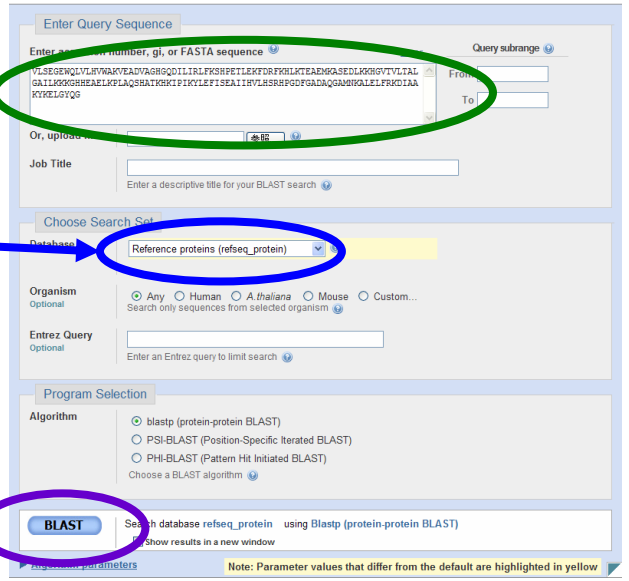
開いたMEGAの2つのウィンドウは閉じる

NCBIサイトでblastp検索

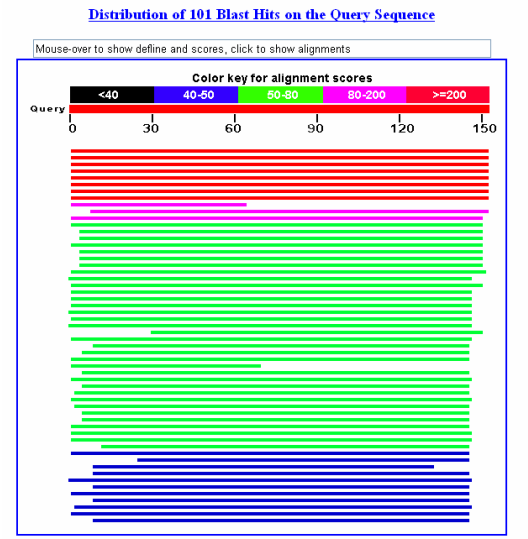
- NCBIのBLASTサイト([URL] http://www.ncbi.nlm.nih.gov/BLAST/)を開き、Basic BLASTの項目の中からprotein BLASTをクリック

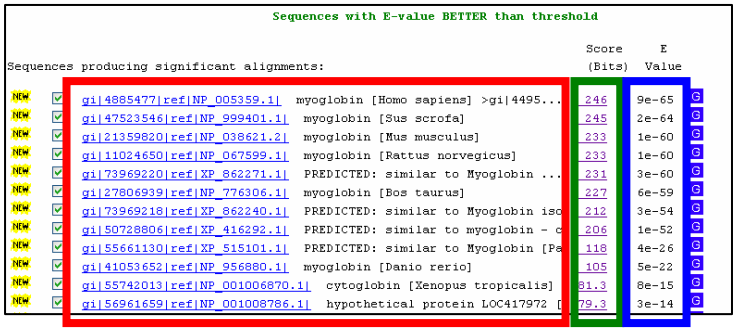


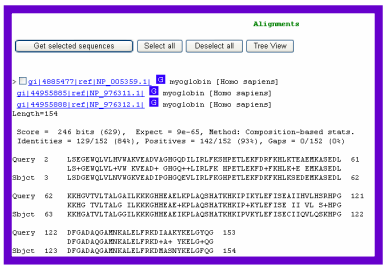
- FASTA sequenceボックスに「1MBNの配列」を貼り付け、Databaseから「Reference proteins」を選択し、「BLAST!」をクリック



- 検索結果、スコアが色分けされて表示され、下の方にヒットしたgi、スコア、E-value、さらに下の方にアラインメントなどが表示されます。





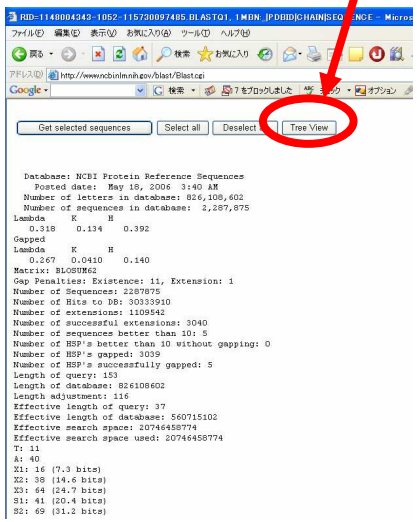


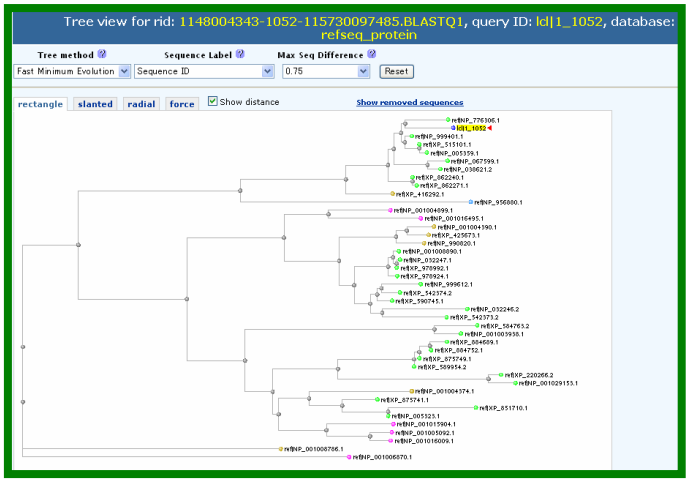
Homo sapiens（ヒト）やSus scrofa（イノシシ）, Mus musculus（ハツカネズミ）などのmyoglobinそしてXenopus tropicalis（アフリカツメガエル）のcytoglobinがヒットしています。

系統樹を表示

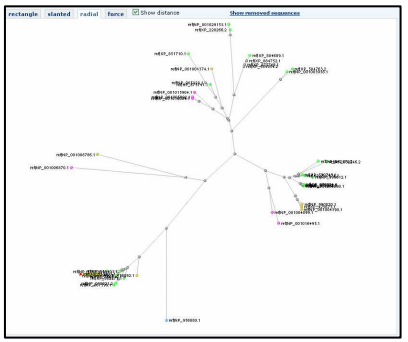
- 一番下で、「Distance tree of results」をクリックすると系統樹も描けます。

  （実際は、いくつか見たい種、遺伝子などを選択（チェック）し表示すると進化的な関係が得られます）





rectangle表示

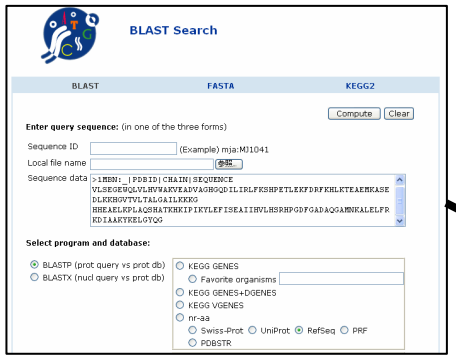


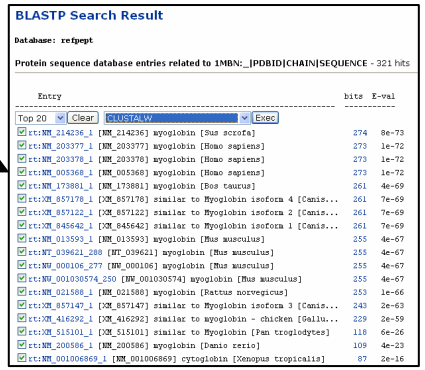
radial表示

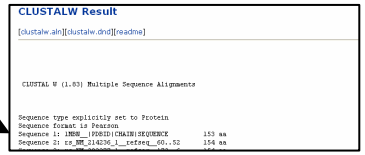
→　バイオスタティスティクス基礎論（西田先生ら）
　　生物配列統計学（岸野先生ら）

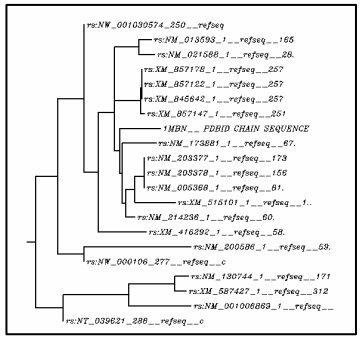
同様な検索は多くのサイトで提供されています

- 例） Genome Net [URL] http://blast.genome.jp









「配列」を貼り付け、
「RefSeq」を選択し、
「Compute」をクリック

「FASTA」検索もできる

「Top 20」など選択し、
「CLUSTALW」を選択し、
「Exec」をクリック

一番下で
「N-J Tree with branch length」を選択し、
「Exec」をクリック

BLAST, FASTA　　ペアワイズシークエンスアラインメント
CLUSTALW, PSI-BLAST　マルチプルシークエンスアラインメント

D. Higgins et al., *Nucleic Acids Res.* **22**, 4673-4680 (1994), “CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice”

バイオインフォマティクスの有用なリンク

- 統合サイトなど
  - Entrez/NCBI [URL] http://www.ncbi.nlm.nih.gov/gquery/
    - GenBank, PubMed, BLAST,,,Structure(VAST), CDD
  - Services/EMBL-EBI [URL] http://www.ebi.ac.uk/services/
    - <Tookbox> BLAST, FASTA, InterProScan, CLUSTALW, DALI,,,
  - GenomeNet/Kyoto Univ. [URL] http://www.genome.jp/
    - KEGG, KEGG2(PATHWAY,,,),,,BLAST, FASTA, CLUSTALW
    - DBGET Database Links [URL] http://www.genome.jp/dbget/dbget.links.html
      - GenBank, EMBL, PubMed,,,PDB, Prosite, Pfam, Blocks, ProDom, PRINTS
  - DDBJ/NIG [URL] http://www.ddbj.nig.ac.jp/Welcome-e.html
    - <Search&Analysis> BLAST, FASTA, ClustalW,,,
  - ExPASy [URL] http://www.expasy.ch/
    - PROSITE,,,
    - <Site Map> http://www.expasy.ch/sitemap.html
      - SWISS-MODEL,,,
  - ANGIS [URL] http://www.angis.org.au/
    - <Links> [URL] http://www.angis.org.au/links.shtml
  - Bio-mirror [URL] http://bio-mirror.net/

参考までに、
バイオインフォマティクスのWeb上での勉強に
（少し古いかも知れませんが）

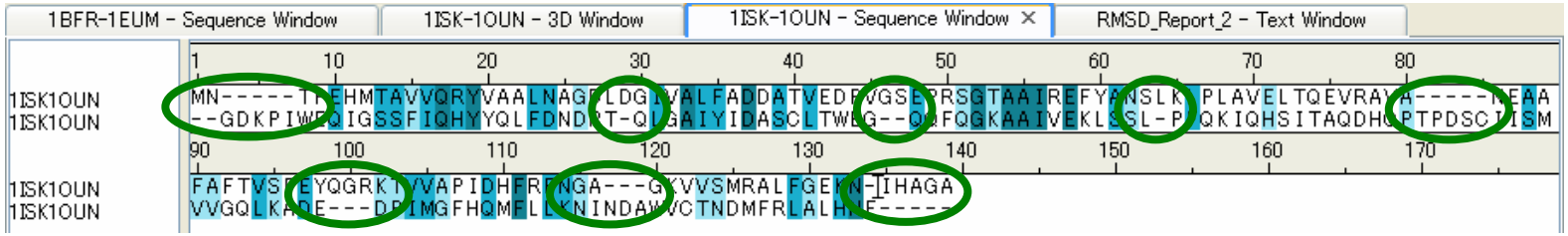
- JST Webラーニングプラザ
  - http://weblearningplaza.jst.go.jp/
    - 分野・映像から選ぶ→ライフサイエンス

- GenomeNet バイオインフォマティクス入門コース
  - http://www.genome.jp/Japanese/lect/course.html
    - 昔の京大での講義

【課題1】 構造比較 （構造類似性）

1ISKの（1OUNとの）構造比較をして、結果をPowerPointにまとめよ

1. CEサイト（[URL] http://cl.sdsc.edu/）で「1ISK:A」の類似構造を検索し、「1OUN:A」との構造アラインメントを得て、そのPDBをダウンロードする
2. Discovery Studio 1.7で、ダウンロードしたPDBを表示し、アラインメントを合わせ、RMSDを計算し、構造を色を変えたSolid Ribbon表示にする
3. 配列のアラインメント、構造の画像をPowerPointに貼り、配列一致度、RMSD、アラインメントされた残基数などを記述する



【課題2】 相同性検索 （配列類似性）

1UB4の相同性検索を行い、結果をPowerPointにまとめよ

1. PDBサイト（[URL] http://www.rcsb.org ）で「1UB4」を検索し、FASTA形式の配列を保存する
2. 保存したファイルを開き、chain Aをコピーする
3. NCBI([URL] http://www.ncbi.nlm.nih.gov/blast/)かGenome Net([URL] http://blast.genome.jp/)のBLASTサイトでその配列をrefseqデータベースを用いて相同性検索する
    1. [オプション] 系統樹を描く
4. 課題の配列はどんなファミリーに属するか、検索された相同性のあるいくつか配列に関して、どの様な生物種の何という遺伝子・タンパク質か、そしてそのE-value、アラインメントなどを含め記述する
    1. [オプション] 系統樹から進化的な関係を述べる

（参考）タカラバイオ [URL] http://www.takara-bio.co.jp/news/2005/08/17.htm
[URL] http://www.takara-bio.co.jp/news/2006/05/15-5.htm
Nature Japan [URL] http://www.natureasia.com/japan/jobs/tokusyu/050908-2.php

＜課題の提出＞

- 上記、【課題１】、【課題２】をPowerPointで2ページの1つのファイルにまとめる
- PowerPointファイルを添付し、Subject（件名）を「課題：構造比較・相同性検索」とし、本文に学籍番号、講義用ID、氏名等記入し、E-mailで以下のメールアドレスへ送信する
  - E-mail address: tadaomi@iu.a.u-tokyo.ac.jp