

平成23年5月11日

構造バイオインフォマティクス基礎

立体構造からの情報抽出

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

教育研究ユニット

寺田 透

立体構造データの使い方

- 個別のデータの利用
 - 変異体の設計
 - 立体構造予測→次回の講義で解説
- 複数のデータの比較
 - 類似性に基づく機能予測→立体構造比較
- データを網羅的に利用
 - 配列と立体構造の関係・傾向を明らかにする
→立体構造データの統計解析

立体構造データの活用事例(1)

The FASEB Journal • Research Communication

Acid-induced sweetness of neoculin is ascribed to its pH-dependent agonistic-antagonistic interaction with human sweet taste receptor

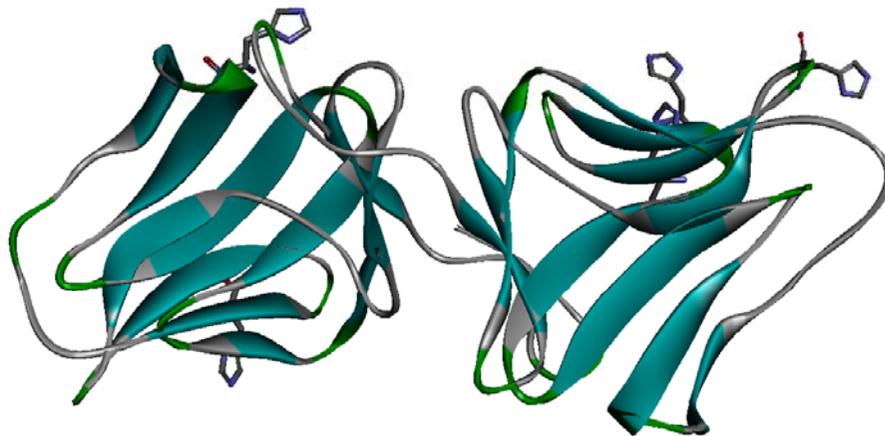
Ken-ichiro Nakajima,* Yuji Morita,* Ayako Koizumi,* Tomiko Asakura,*
Tohru Terada,*[†] Keisuke Ito,* Akiko Shimizu-Ibuka,[§] Jun-ichi Maruyama,[‡]
Katsuhiko Kitamoto,[‡] Takumi Misaka,* and Keiko Abe*^{†,1}

*Department of Applied Biological Chemistry, [†]Agricultural Bioinformatics Research Unit, and
[‡]Department of Applied Biotechnology, Graduate School of Agricultural and Life Sciences,
University of Tokyo, Tokyo, Japan; and [§]Department of Nutritional Science, Tokyo University of
Agriculture, Tokyo, Japan

Nakajima *et al.* *FASEB J.* **22**, 2323 (2008).

立体構造データの活用事例(1)

- Neoculin: 酸味を甘味に変える甘味修飾活性をもつ
- pHセンサーとして機能している残基を同定したい



PDB ID: 2D04

5つのHis残基を持つ
いずれも表面に露出している

分子内部に埋もれている
場合、別のアミノ酸に置換
すると立体構造を維持で
きなくなる場合がある

すべてAlaに置換

味覚修飾活性を失う

立体構造データの活用事例(2)

Biosci. Biotechnol. Biochem., **72** (12), 3237–3248, 2008



Alteration of the Substrate Specificity of the Angular Dioxygenase Carbazole 1,9a-Dioxygenase

Hiromasa UCHIMURA,^{1,*,**} Tadafumi HORISAKI,^{1,2,*} Takashi UMEDA,¹ Haruko NOGUCHI,^{1,3}
Yusuke USAMI,¹ Li LI,^{1,***} Tohru TERADA,^{3,****} Shugo NAKAMURA,⁴ Kentaro SHIMIZU,^{3,4}
Tetsuo TAKEMURA,² Hiroshi HABE,^{1,****} Kazuo FURIHATA,⁵ Toshio OMORI,^{1,****}
Hisakazu YAMANE,¹ and Hideaki NOJIRI^{1,3,†}

¹*Biotechnology Research Center, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan*

²*Department of Chemistry, Faculty of Science, Tokyo University of Science,*

1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan

³*Professional Program for Agricultural Bioinformatics, The University of Tokyo,*

1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

⁴*Department of Biotechnology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan*

⁵*Department of Applied Biological Chemistry, The University of Tokyo,*

1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

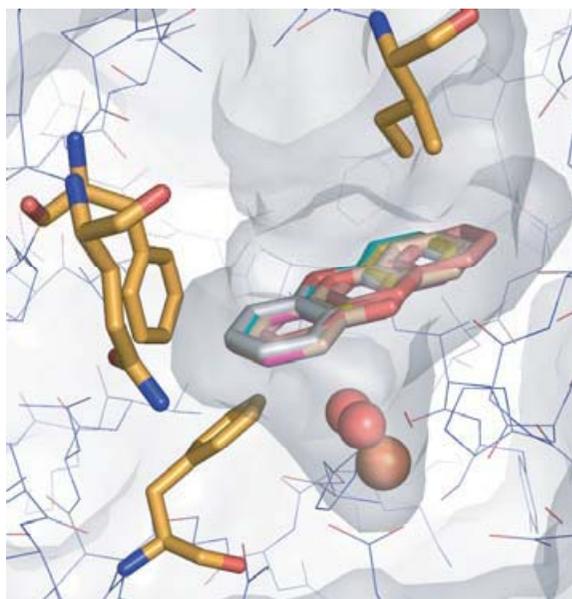
Received July 28, 2008; Accepted August 27, 2008; Online Publication, December 7, 2008

[doi:10.1271/bbb.80512]

Uchimura et al. *Biosci. Biotechnol. Biochem.* **72**, 3237 (2008).

立体構造データの活用事例(2)

- Carbazole 1,9a-dioxygenase (CARDO) は基質によって異なる位置に水酸基を付加する
- 基質特異性のメカニズムを明らかにしたい



結晶構造に基質をドッキングし、
基質と接触している残基を同定

これらの残基の変異を導入

反応特異性が変化することを確認

立体構造データの活用事例(3)



the **FEBS**
Journal

Molecular defect of isovaleryl-CoA dehydrogenase in the *skunk* mutant of silkworm, *Bombyx mori*

Kei Urano¹, Takaaki Daimon¹, Yutaka Banno², Kazuei Mita³, Tohru Terada⁴, Kentaro Shimizu^{4,5}, Susumu Katsuma¹ and Toru Shimada^{1,4}

1 Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, University of Tokyo, Japan

2 Institute of Genetic Resources, Graduate School of Bioresource and Bioenvironmental Science, Kyushu University, Fukuoka, Japan

3 Division of Insect Sciences, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan

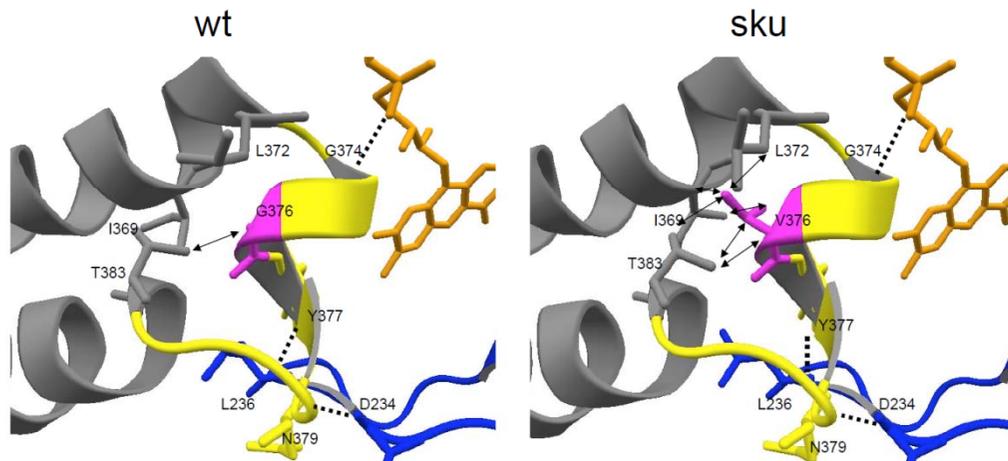
4 Agricultural Bioinformatics Research Unit, Graduate School of Agricultural and Life Sciences, University of Tokyo, Japan

5 Department of Biotechnology, Graduate School of Agricultural and Life Sciences, University of Tokyo, Japan

Urano *et al.* *FEBS J.* **277**, 4452 (2010).

立体構造データの活用事例(3)

- 致死的で異臭を伴うカイコの変異*sku*は、isovaleryl-CoA dehydrogenaseのGly376のValへの変異によって起こる
- この変異によってなぜ酵素の機能が損なわれるのか明らかにしたい



ヒトのホモログの立体構造をもとにカイコの野生型の立体構造をモデリング

Gly376をValに置換

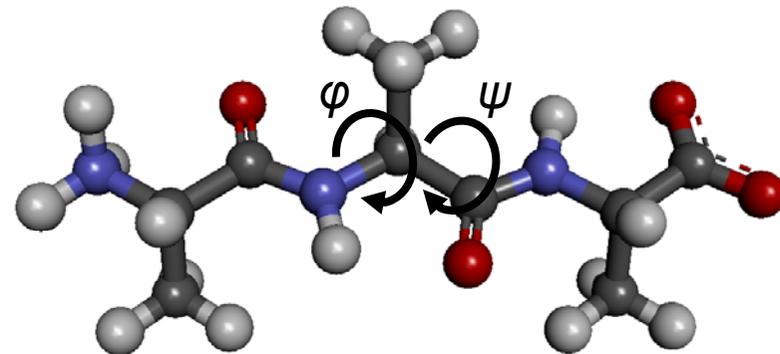
多量体構造や補酵素との相互作用が損なわれる可能性が示唆された

本日の講義内容

- 立体構造からのデータ抽出
 - 2次構造
 - 溶媒露出表面積
- 立体構造比較
 - 立体構造重ね合わせ
 - 立体構造分類データベース
- 立体構造データの統計解析
 - 非冗長立体構造データベース
 - 3D-1D法
- 実習

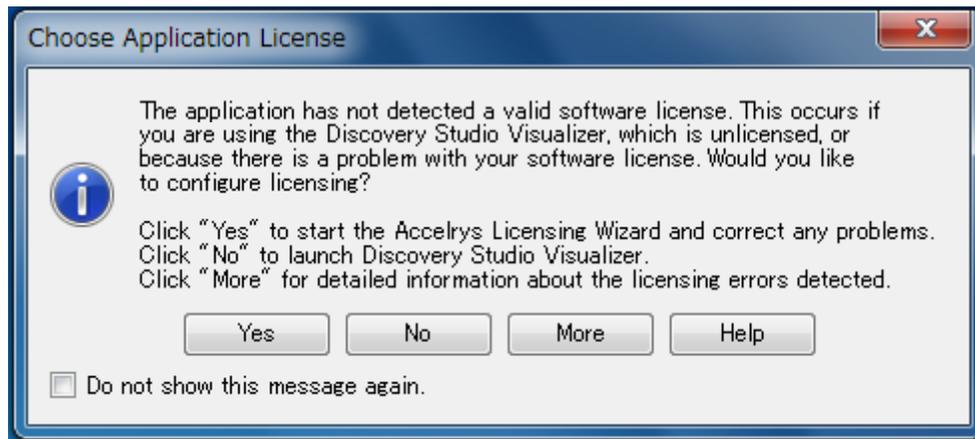
2次構造

- α helix
 - 残基 i の C=O と 残基 $i+4$ の N-H が 水素結合を形成
 - 主鎖の二面角 (φ, ψ) $\approx (-60^\circ, -60^\circ)$
- β sheet
 - 隣り合う β strand との間で 水素結合を形成
 - 主鎖の二面角 (φ, ψ) $\approx (-120^\circ, 120^\circ)$
 - 並行と逆並行がある



Discovery Studioの起動について

- Discovery Studio 3.0 Clientを起動すると、以下のようなウィンドウが現れる場合があります

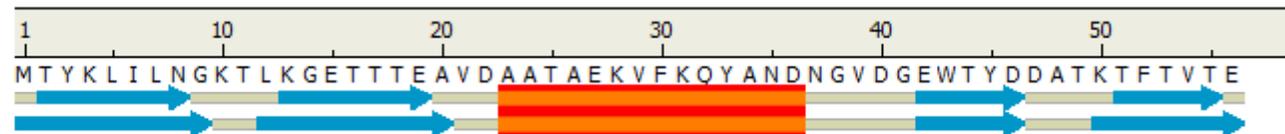


- この場合は「No」をクリックして下さい

2次構造の確認

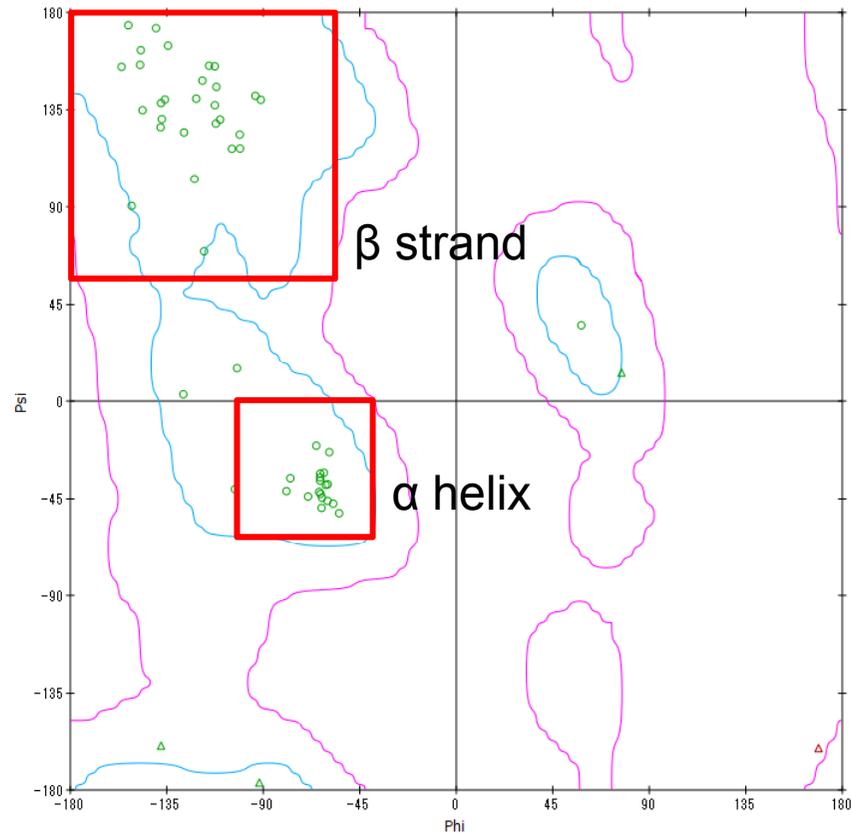
1. Discovery Studio 3.0 Clientで1PGAを開く
 - α helix、 β sheet(並行および逆並行)を確認
2. メニューの「Sequence」→「Show sequence」
3. Sequence WindowでDisplay Styleを開く
4. Secondary Structureタブを開き「Kabsch and Sander」と「PDB」をチェックし「OK」

1PGA
Kabsch and Sander
PDB



Ramachandran plot

5. Molecule Windowをactiveにしたのち、メニューの「Chart」→「Ramachandran plot」



2次構造データの由来(1)

- PDBのヘッダから
 - 立体構造を決定した人が記述

```
HELIX 1 1 ALA A 23 ASP A 36 1
SHEET 1 S1 4 LEU A 12 ALA A 20 0
SHEET 2 S1 4 MET A 1 GLY A 9 -1
SHEET 3 S1 4 LYS A 50 GLU A 56 1
SHEET 4 S1 4 GLU A 42 ASP A 46 -1
```

2次構造データの由来(2)

- 立体構造データから
 - KabschとSanderの方法(DSSP)*
 - <http://swift.cmbi.ru.nl/gv/dssp/>
 - DSSPの出力(H: α helix; E: β strand)

#	RESIDUE	AA	STRUCTURE	BP1	BP2	ACC	N-H-->O	O-->H-N	N-H-->O	O-->H-N	TCO	KAPPA	ALPHA	PHI	PSI		
1	1	A	M		0	0	116	0, 0.0	19, -2.8	0, 0.0	2, -0.5	0.000	360.0	360.0	360.0	151.3	
2	2	A	T	E	-A	19	0A	91	17, -0.2	2, -0.3	19, -0.1	17, -0.2	-0.900	360.0	-168.8	-101.0	123.5
3	3	A	Y	E	-A	18	0A	5	15, -2.8	15, -2.8	-2, -0.5	2, -0.3	-0.774	9.0	-143.9	-112.6	155.1
4	4	A	K	E	-Ab	17	51A	64	46, -1.7	48, -2.7	-2, -0.3	2, -0.5	-0.882	3.2	-156.2	-118.4	148.5
5	5	A	L	E	-Ab	16	52A	0	11, -3.4	11, -1.9	-2, -0.3	2, -0.6	-0.995	1.8	-161.2	-126.9	124.5
6	6	A	I	E	-Ab	15	53A	60	46, -2.8	48, -2.5	-2, -0.5	2, -0.6	-0.954	13.6	-159.1	-104.6	116.9
7	7	A	L	E	+Ab	14	54A	6	7, -2.2	7, -1.8	-2, -0.6	2, -0.4	-0.907	18.6	174.1	-100.8	117.1
8	8	A	N	E	+Ab	13	55A	75	46, -3.1	48, -2.4	-2, -0.6	5, -0.2	-0.695	26.5	149.4	-117.6	69.5
(中略)																	
23	23	A	A	H	> S+	0	0	25	-2, -0.3	4, -2.9	1, -0.2	5, -0.2	0.805	117.0	60.8	-63.5	-35.1
24	24	A	A	H	> S+	0	0	49	1, -0.2	4, -1.0	2, -0.2	-1, -0.2	0.859	109.0	44.1	-63.5	-33.4
25	25	A	T	H	> S+	0	0	60	-3, -0.5	4, -1.9	2, -0.2	3, -0.3	0.918	112.7	49.9	-79.2	-41.4
26	26	A	A	H	X S+	0	0	0	-4, -1.9	4, -2.8	1, -0.2	5, -0.3	0.912	107.5	57.1	-63.5	-36.6
27	27	A	E	H	X S+	0	0	61	-4, -2.9	4, -1.9	1, -0.2	-1, -0.2	0.855	107.8	46.7	-59.9	-38.3

*Kabsch & Sander, *Biopolymers* **22**, 2577 (1983).

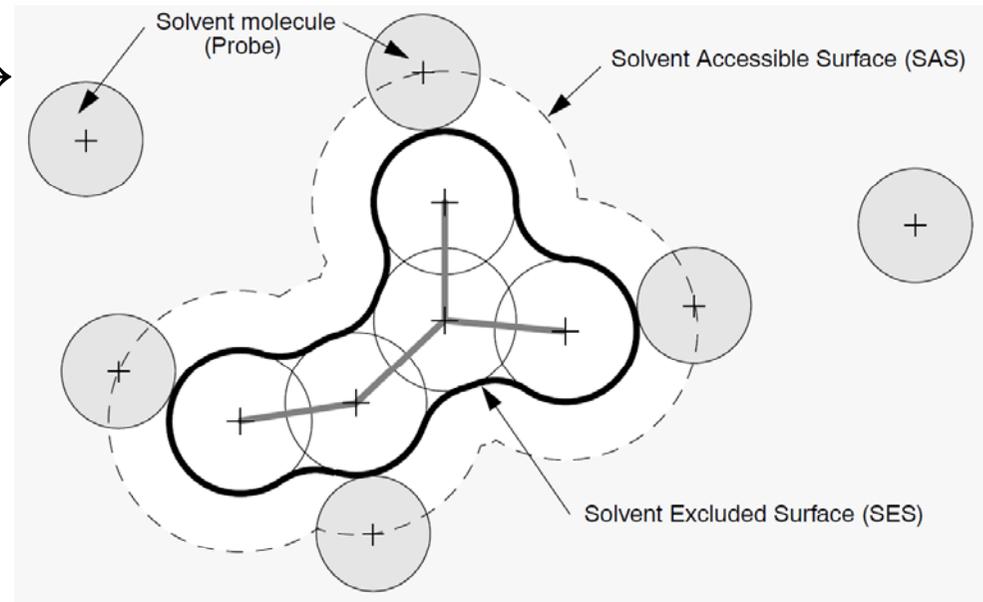
参考 : DSSPの実行

- DSSPCMBI.EXEをダウンロードし、デスクトップに保存
- スタートメニューから「すべてのプログラム」→「アクセサリ」→「コマンドプロンプト」を起動
- 以下を実行

```
C:¥Users¥iu>cd Desktop  
C:¥Users¥iu¥Desktop>DSSPCMBI.EXE 1PGA.pdb > 1PGA.txt
```
- 生成された1PGA.txtをダブルクリックして開く

溶媒露出表面積

- 分子に接触するように転がしたプローブ球の中心の軌跡→溶媒露出表面
- メニューの「Structure」→「Solvent Accessibility」で計算できる
- アミノ酸の大きさによって表面積が異なるので、Ala-X-Ala(またはGly-X-Gly)ペプチドにおけるそのアミノ酸の溶媒露出表面積で割った、溶媒露出度も指標として良く用いられる
- DSSPのACCカラムの値は溶媒露出表面積に相当



Sanner et al. *Biopolymers* **38**, 305 (1993)から引用

立体構造比較(1)

1. RCSBのホームページを開く
2. 左側のメニューのToolsにある「Compare Structure」をクリック
3. PDB1に「1WW9」、PDB2に「1NDO」
Comparison Methodに「jCE」を設定し
「Compare」

RCSB PDB Protein Comparison Tool

Calculate pairwise sequence or structure alignments.

Compare the following two proteins ⓘ

PDB1: Chain1: A (sequence: MANVDEAILKRVKGWAPYVDA)

PDB2: Chain2: A (sequence: MNYNKKILVSEGLSQKHLIHG)

jCE algorithm

If you are having problems launching Java Web Start applications, view our [troubleshooting Java Web Start page](#) for more help.

Alignment Options

Align custom files (Launches a Java Web Start application)

[Show Example](#) (4HHB.A vs 4HHB.B Java-FATCAT, rigid)

[More documentation](#) on the RCSB PDB protein structure alignments

Acknowledgements

[Click Here](#) to show the acknowledgements related to our structure comparison feature.

RCSB **PDB**
PROTEIN DATA BANK

Contact Us | Print

MyPDB Hide

Login to your Account
Register a New Account

Home Hide

News & Publications
Usage/Reference Policies
Deposition Policies
Website FAQ
Deposition FAQ
Contact Us
About Us
Careers
External Links
Sitemap
New Website Features

Deposition Hide

All Deposit Services
Electron Microscopy
X-ray | NMR
Validation Server
BioSync Beamline
Related Tools

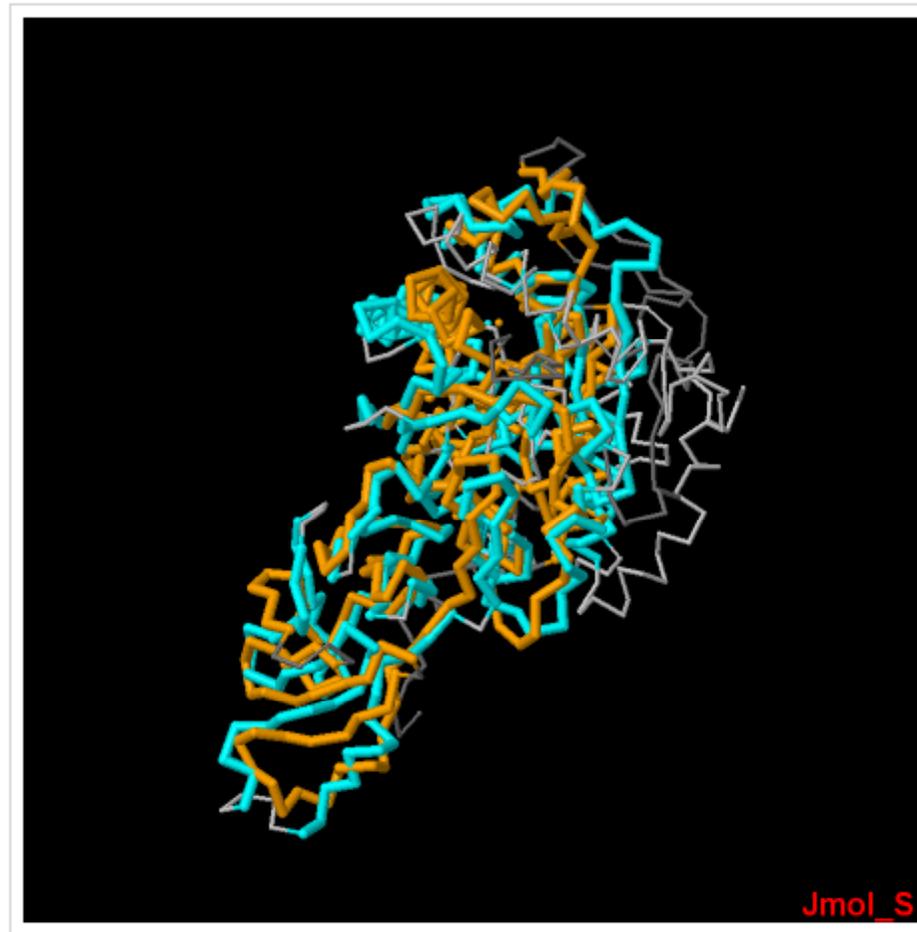
Search Hide

Advanced Search
Latest Release
New Structure Papers
Sequence Search
Chemical Components
Unreleased Entries
Browse Database
Histograms

Tools Hide

Download: Entries | Bands
Compare Structures
File Formats
Services: RESTful | SOAP
Widgets

立体構造比較(2)



立体構造比較(3)

- 類似性の指標

- Z-Score: 5.86 (この値が大きいほど偶然に類似する確率が低い)

>4.5	family level similarity
4.0-4.5	superfamily level similarities, strong function related similarities or strong recurring fold
3.7-4.0	twilight zone where some similarities of biological significance can be seen
<3.7	similarities of low significance, but still some biologically important similarities can be revealed, but interpretation normally requires additional evidence.

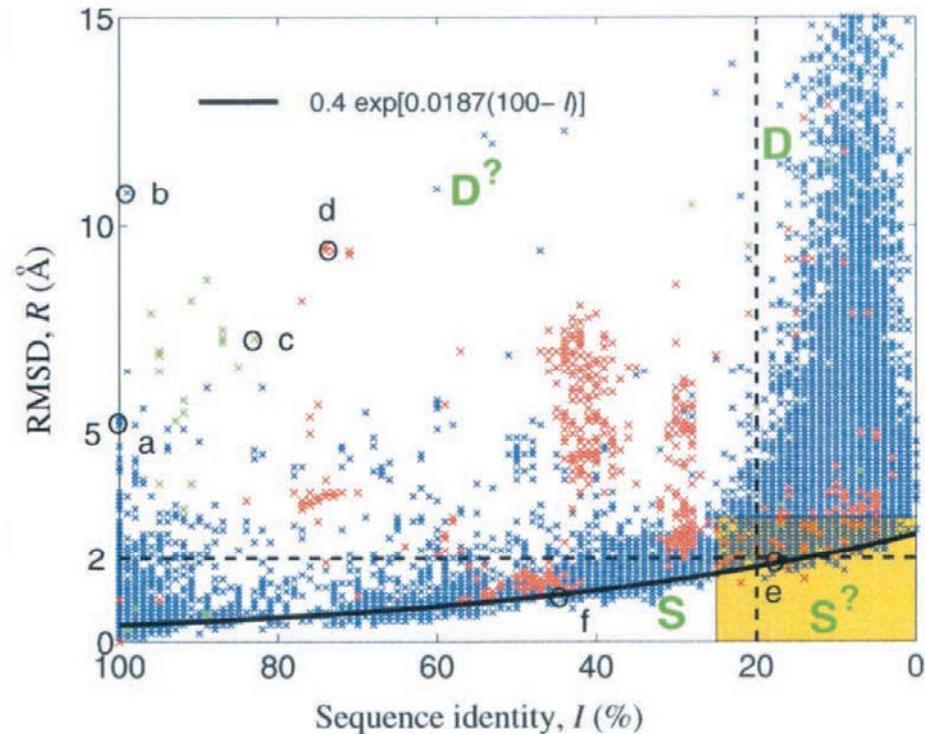
- RMSD: 4.34 Å

立体構造比較の利用

- 配列同士のアラインメントより、機能的に意味のある残基同士の対応付けが可能
- 立体構造の類似性の指標を用いて、立体構造をクエリとして、類似立体構造を検索できる
 - 主なサイト
 - VAST (<http://www.ncbi.nlm.nih.gov/Structure/VAST/>)
 - Dali (http://ekhidna.biocenter.helsinki.fi/dali_server/)
 - CE (<http://cl.sdsc.edu/ce.html>)
 - FATCAT (<http://fatcat.burnham.org/>)

配列類似性と立体構造類似性

- 立体構造比較により配列をアライン
- 一般に、配列一致度が高くなると、立体構造類似性も高くなる
- 配列一致度が20%を切ると、立体構造が大きく異なるケースが増えてくる



立体構造分類データベース

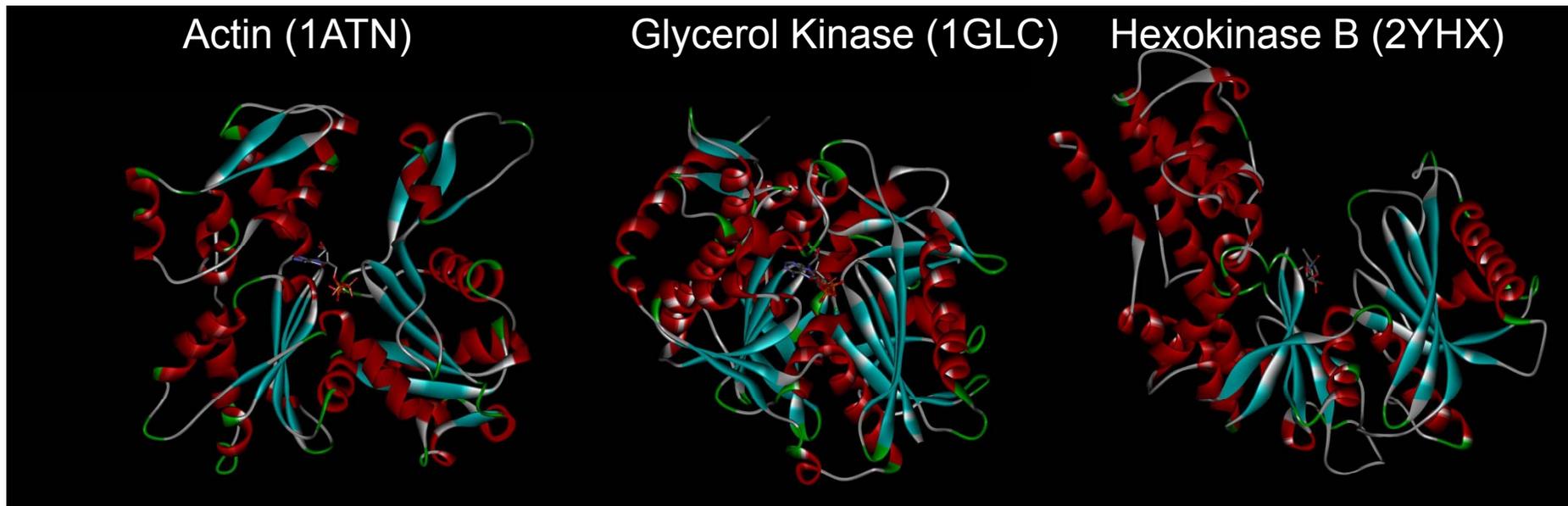
- 立体構造を立体構造類似性に基づいて階層的に分類したデータベース
 - SCOP (<http://scop.mrc-lmb.cam.ac.uk/scop/>)
主に人手による分類
Class → Fold → Superfamily → Family
 - CATH (<http://www.cathdb.info/>)
主に計算機による分類
Class → Architecture → Topology → Homologous Superfamily
- 例: 1HCTのSCOPにおける分類
 - Alpha and beta proteins (a+b) → SH2-like → SH2 domain → SH2 domain

SCOPにおける階層

- Family: 配列一致度が概ね30%以上あり、進化的な類縁関係が明らかなもの
- Superfamily: 配列一致度は低いが、立体構造と機能の類似性から、進化的な類縁関係が推定されるもの
- Fold: 主な2次構造の配置とトポロジーが同じだが、進化的な類縁関係はないもの
→ 構造類似性は特定のパッキングやトポロジーを好むタンパク質の物理化学的な性質に起因

Superfamilyの例

- Actin-like ATPase domain superfamily



- 配列相同性が低く、生物学的な機能は全く異なるが、立体構造は類似している

Kabsch & Holmes, *FASEB J.* **9**, 167 (1995).

立体構造データの統計解析

- 立体構造データを網羅的に用いて、統計解析を行うことができる
 - 例: α helix、 β strandを取りやすいアミノ酸
- 立体構造データの偏りに注意する必要がある
 - 例えばhen egg lysozyme (PDB ID: 3AW6など)と100%配列が一致するエントリは283ある
- 類似した配列を除いたデータベースを使うと良い
 - PISCES (<http://dunbrack.fccc.edu/PISCES.php>) (配列一致度、分解能などを指定して冗長を排除したPDBのリストを作成することができる)

3D profile法

Research Article

A Method to Identify Protein Sequences That Fold into a Known Three-Dimensional Structure

JAMES U. BOWIE, ROLAND LÜTHY, DAVID EISENBERG

- 逆フォールディング問題「与えられた立体構造にどんなアミノ酸配列が適合するか」を解く

Bowie *et al.* *Science* **253**, 164 (1991).

タンパク質立体構造の一般的傾向

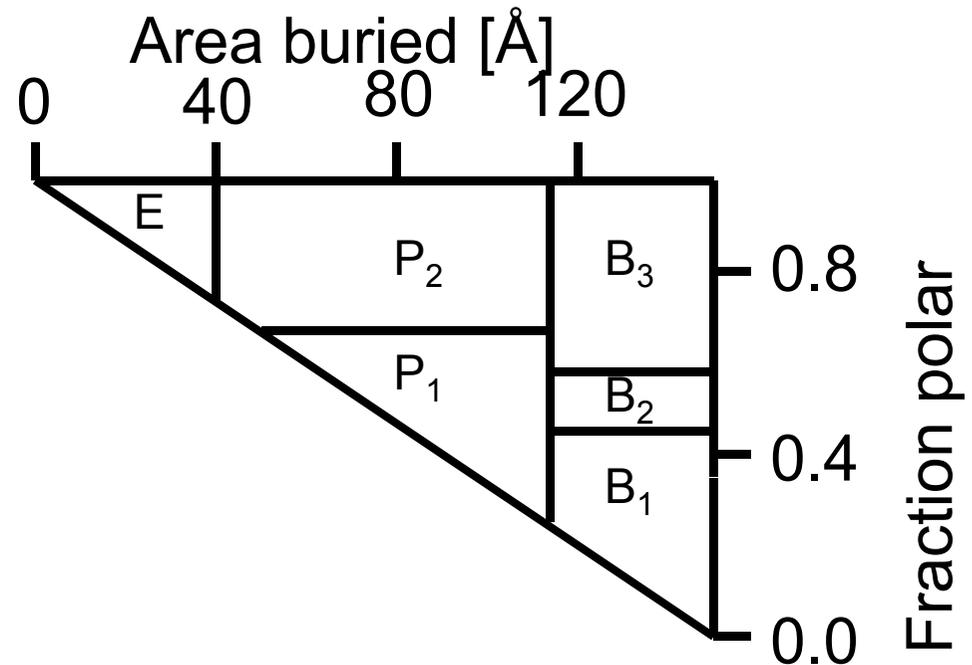
- 疎水アミノ酸はタンパク質の中心付近に、親水性アミノ酸はタンパク質表面に存在
- Ala、Glu、Leu、Metは α helixを形成しやすい
- Pro、Gly、Tyr、Serは α helixを形成しにくい



- 各残基が立体構造の中でおかれた「環境」に注目すると、アミノ酸のその環境への適合度合いに基づいて、立体構造に最も適合するアミノ酸配列を決めることができる

環境の定義

- 各残基の溶媒露出表面積を計算し、Gly-X-Glyペプチドの溶媒露出表面積との差をArea buriedとする
- 各残基の表面のうち、極性原子(溶媒を含む)に覆われている割合をFraction polarとする
- これに主鎖構造(α 、 β 、その他)を加えて18種類の環境を定義



3D-1D score

- 立体構造データベース中のタンパク質について、アミノ酸*i*の環境*j*における出現確率を計算

$$3D-1D \text{ score} = \ln \left[\frac{P(i|j)}{P(i)} \right]$$

Environment class	W	F	Y	L	I	V	M	A	G	P	C	T	S	Q	N	E	D	H	K	R
B ₁ α	1.00	1.32	0.18	1.27	1.17	0.66	1.26	-0.66	-2.53	-1.16	-0.73	-1.29	-2.73	-1.08	-1.93	-1.74	-1.97	-0.34	-1.82	-1.67
B ₁ β	1.17	0.85	0.07	1.13	1.47	1.09	0.55	-0.79	-2.02	-0.94	-0.22	-1.12	-2.91	-1.67	-1.42	-1.93	-2.56	-1.91	-2.69	-1.16
B ₁	1.05	1.45	0.17	1.10	1.11	1.02	0.98	-0.91	-1.92	0.26	-1.22	-1.53	-2.81	-1.17	-2.42	-2.52	-1.76	-1.12	-2.59	-2.16
B ₂ α	0.50	0.90	0.85	1.01	0.63	0.68	1.12	-0.69	-1.49	-2.21	-0.10	-1.50	-1.47	-0.23	-0.61	-0.71	-1.62	0.23	-0.78	0.06
B ₂ β	0.01	1.18	1.06	0.76	1.31	1.06	0.64	-1.55	-2.26	-0.49	-0.87	-2.27	-1.77	-1.22	-2.07	-1.07	-1.41	-0.77	-1.14	-0.20
B ₂	1.02	1.05	1.12	0.84	0.81	0.60	0.90	-0.66	-1.66	0.19	-0.05	-0.76	-1.17	-0.76	-0.66	-1.35	-1.28	0.46	-2.34	-0.80
B ₃ α	0.92	-0.03	0.58	0.15	0.04	-0.02	0.89	-0.57	-1.86	-0.68	-1.56	-0.57	-0.96	0.22	-0.06	0.08	-0.50	0.73	0.43	0.96
B ₃ β	0.75	0.81	1.30	0.18	0.54	0.56	-0.57	-0.93	-1.93	-0.34	-0.54	-0.44	-0.74	0.21	-0.24	-0.14	-0.86	0.82	-0.53	0.13
B ₃	1.07	0.70	1.13	0.35	-0.17	-0.03	0.23	-0.96	-0.98	-0.13	-1.20	-0.53	-0.54	0.05	0.04	-0.36	-1.05	1.01	0.10	0.66
P ₁ α	-1.35	-0.82	-0.59	-0.52	-0.24	0.10	-0.03	0.73	-0.49	-0.25	0.95	0.31	0.34	-0.14	-0.54	-0.17	-0.25	-0.52	-0.21	-0.28
P ₁ β	0.36	-0.49	0.17	-1.03	0.20	0.46	-0.27	0.64	-0.82	-0.55	1.49	0.93	0.33	-2.27	-1.32	-0.73	-1.07	-0.42	-1.21	-0.77
P ₁	-1.26	-1.20	-1.31	-0.62	-0.23	-0.01	-1.19	0.46	-0.24	0.66	1.35	0.56	0.49	-0.63	-0.13	-0.61	0.38	-1.12	-0.74	-1.29
P ₂ α	-1.14	-1.43	-0.79	-0.35	-0.54	-0.48	-0.45	0.06	-0.50	-0.26	-0.93	-0.05	-0.18	0.55	-0.05	0.56	0.28	0.06	0.61	0.50
P ₂ β	-0.79	-0.54	-0.84	-1.30	-0.33	0.13	-0.72	-0.55	-0.98	-1.29	-0.57	0.84	0.59	-0.08	-0.16	0.32	0.19	-0.87	0.59	0.10
P ₂	-0.82	-0.86	-0.51	-0.70	-1.09	-0.88	-0.89	-0.15	-0.40	0.44	-0.60	0.06	0.26	0.27	0.50	0.27	0.49	0.13	0.44	0.30
E α	-1.35	-2.20	-2.10	-1.58	-2.76	-1.10	-0.72	0.46	0.68	0.04	-0.44	-0.17	0.15	0.36	0.28	0.59	0.44	-0.19	0.13	-0.34
E β	0.64	-0.90	0.30	-1.66	-1.47	-1.74	-0.68	0.06	1.46	-0.96	-0.24	0.14	0.65	-0.19	-0.06	-0.16	-0.78	-0.83	-0.52	-0.49
E	-2.14	-1.90	-0.94	-1.19	-1.61	-0.91	-1.67	0.12	1.13	0.20	-0.46	0.12	0.32	-0.03	0.41	0.03	0.22	-0.25	-0.14	-0.32

配列のアライメント

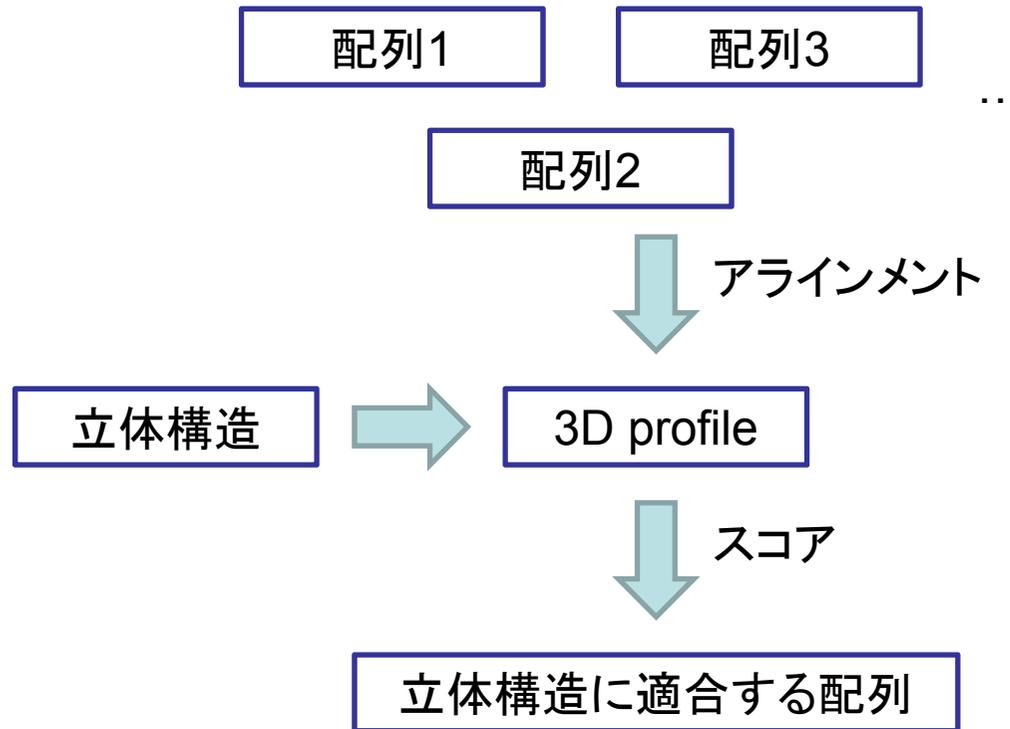
- Dynamic programmingを用いてアミノ酸配列を3D profileにアラインする
- 最も大きいスコアを与えた、配列がその立体構造に最も適合した配列といえる

$$\text{Total 3D - 1D score} = \sum_{ij} N_{ij} \ln \left[\frac{P(i|j)}{P(i)} \right]$$

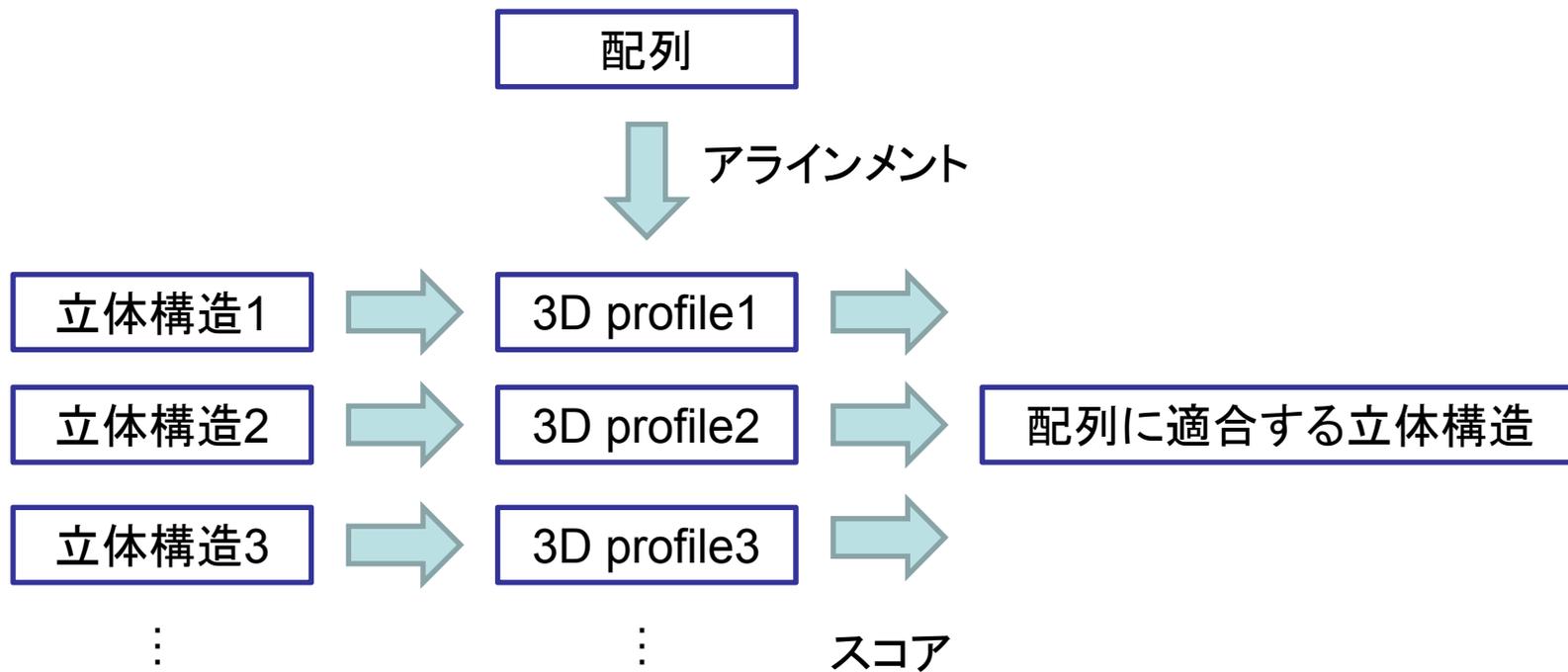
3D profileの応用(1)

- 立体構造は配列より保存されやすいことを利用した、遠縁のホモログの検出
 - Actinとheat shock proteinは配列相同性はほとんどないが、立体構造は類似している
 - Actinの3D profileを用いると、actin自身の配列に次いでheat shock proteinが高いスコアを示す
- 様々な立体構造について3D profileを計算しておき、与えられた配列に適合する立体構造を探すこともできる→フォールド認識

3D profileの応用(2)



3D profileの応用(3)



3D profileの応用(4)

- 予測構造の評価

- 予測に使用した配列と、予測構造の適合

度を予測構造の精度の指標とする

- 3D-1D scoreの和を用いて、予測構造同士の精度の比較ができるほか、各残基の3D-1D scoreを用いて局所構造の精度を評価できる

- Verify3D

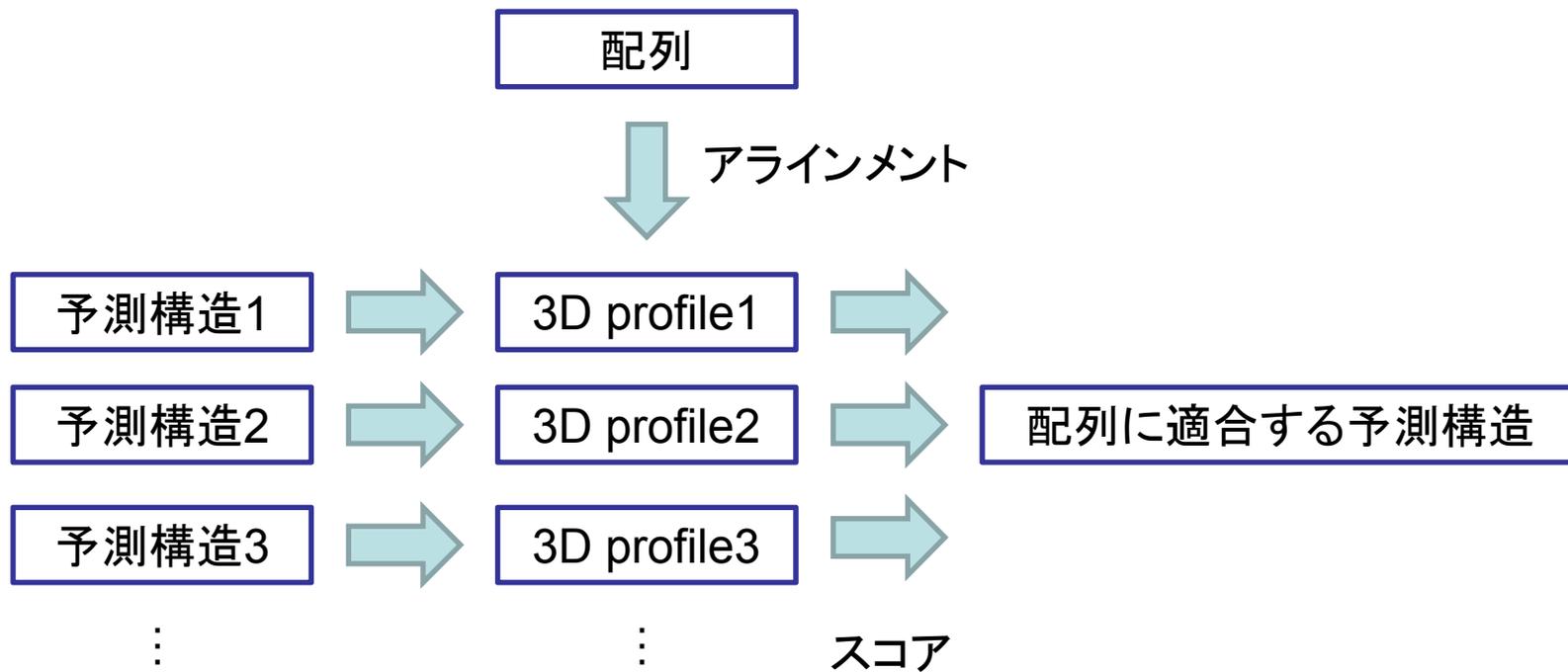
(http://nihserver.mbi.ucla.edu/Verify_3D/)

Assessment of protein models with three-dimensional profiles

Roland Lütthy*, James U. Bowie† & David Eisenberg†

The Molecular Biology Institute and Department of Chemistry and Biochemistry, UCLA, Los Angeles, California 90024-1570, USA

3D profileの応用(5)



実習課題

- Actin (PDB ID: 1ATN)、heat shock protein (PDB ID: 3HSC)について以下に答えよ
 1. これらのタンパク質の立体構造はどの程度類似しているか、RCSBの立体構造比較ツールを用いて調べよ。2つの立体構造を重ね合わせた図、Z-score、RMSD、Align-len、Identityを報告せよ。
 2. これらのタンパク質のアミノ酸配列はどの程度類似しているか調べよ。E-valueを報告せよ。
 3. これらのタンパク質はSCOPのどのfold、super family、familyに属しているか報告せよ。

課題の提出

- 項目1の2つの立体構造を重ね合わせた図についてはpng形式で保存し、メールに添付すること
(Jmolの画面で右クリック→「File」→「Export PNG image」)
- 項目1のZ-score、RMSD、Align-len、Identity、項目2、3はメールの本文に記載すること
- メールは寺田宛 (tterada@iu.a.u-tokyo.ac.jp) に送ること
- その際、件名は「構造実習」とし、本文に氏名と学生証番号を必ず明記すること