

平成22年4月16日

構造バイオインフォマティクス基礎

立体構造データベースと その利用

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

講義の予定

1. 4月16日(金)
担当: 寺田 透
内容: タンパク質立体構造データベースとその利用
2. 5月7日(金)
担当: 永田宏次
内容: X線結晶構造解析法
3. 5月26日(水)
担当: 伊藤 隆(首都大学東京)
内容: 核磁気共鳴(NMR)法
4. 6月11日(金)
担当: 光岡 薫(産業技術総合研究所)
内容: 電子顕微鏡による立体構造決定

本日の講義内容

- 立体構造データベースへのアクセス
 - 検索
 - データフォーマット
 - 立体構造データの可視化
 - 立体構造決定法
- 配列データベースとの連携
 - 配列データベースからのリンク
 - 配列からの検索
- 立体構造比較
- 階層的立体構造分類データベース
- 実習

立体構造データベース

- Protein Data Bank (PDB)
- タンパク質、核酸などの生体高分子の立体構造を収集、公開している世界で唯一のデータベース
- 2010年4月時点でのエントリ数は約64,500
- 主なWebサイト
 - 米国 : <http://www.rcsb.org/>
 - 欧州 : <http://www.ebi.ac.uk/pdbe/>
 - 日本 : <http://www.pdbj.org/>

データベースへのアクセス

- RCSBのサイト (<http://www.rcsb.org/>)

The screenshot shows the RCSB PDB website homepage. At the top, the logo for RCSB PDB (Protein Data Bank) is visible on the left, and the text "An Information Portal to Biological Macromolecular Structures" is centered. Below this, a search bar is present with a dropdown menu for "PDB ID or Text" and a "Search" button. The main content area is titled "A Resource for Studying Biological Macromolecules" and contains several sections: "Molecule of the Month" featuring Concanavalin A and Circular Permutation, and "Protein Structure Initiative Featured Molecule: Phytochrome". The left sidebar contains navigation links for Home, Deposition, Search, Tools, and Education. The right sidebar includes a "Customize This Page" section with options for New Features, Improved Ligand Search, and News, along with a "Share this Page" section with social media links.

RCSB PDB
PROTEIN DATA BANK

An Information Portal to Biological Macromolecular Structures
As of Tuesday Apr 06, 2010 at 5 PM PDT there are 64500 Structures | PDB Statistics

HELP | PRINT PDB ID or Text Search Advanced Search

MyPDB Login A MEMBER OF THE PDB

A Resource for Studying Biological Macromolecules

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the **wwPDB**, the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

[Hide Welcome Message](#)

Featured Molecules (Previous Features: MOH | PSI)

Molecule of the Month: Concanavalin A and Circular Permutation

Evolution is a great tinkerer: once a successful plan is found, it is used again and again, often with many changes and improvements. This is easily seen in the living things around us. Most mammals have four limbs, which have evolved into all manner of arms and legs, and even into flippers and wings. Most plants are covered with leaves, which range from spiny needles to huge jungle fronds. Looking at protein sequences and structures, we see the same diversity generated through variation.

[Full Article...](#)

Protein Structure Initiative Featured Molecule: Phytochrome

In two new NMR structures of an unusually small phytochrome, researchers have revealed how plants see light and shade. The structures reveal for the first time the complex motion of the chromophore after absorbing red light.

[Full Article](#) | [PSI Structural Genomics Knowledgebase](#)

New user? Try the browser [compatibility check](#) and information on [Getting Started](#) or the [narrated tutorial](#)

Customize This Page

New Features

Improved Ligand Search

Read more about the releases:

Website Release Archive:

News

Weekly | Quarterly | Yearly

Statement on Retraction of PDB Entries

2010-04-06
[Bookmark and Share RCSB PDB Webpages](#)

Share this Page

Email
Delicious
Stumble
Google

[Read More](#)

Previous Weekly News:

- [Advanced Search: Sequence Motif](#)
- [West Windsor Plainsboro High School North Wins New Jersey Science Olympiad](#)
- [Protein Modeling State Finals](#)
- [San Diego Science Festival Expo Day: March 27](#)
- [Exhibiting at NSTA](#)
- [Improved Ligand Searching](#)

FTP Archives

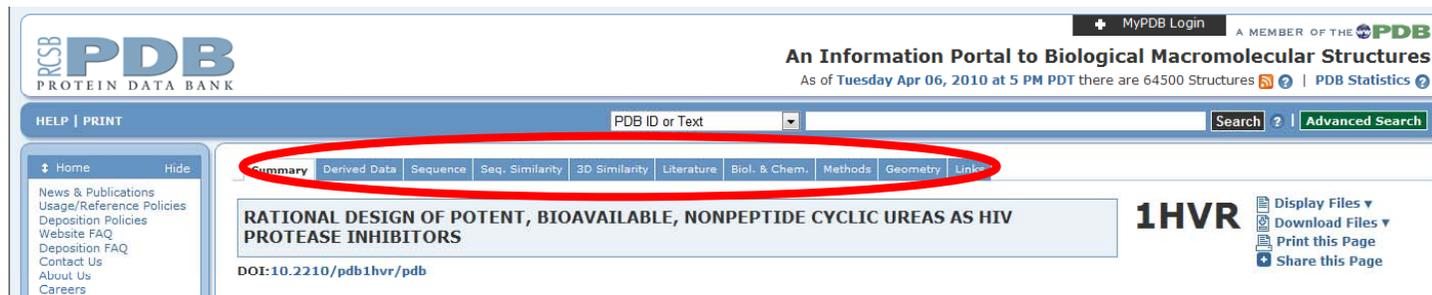
Current PDB FTP Archive:
<ftp.wwpdb.org>

検索

- キーワード
 - 立体構造データに対するテキスト検索
 - 例: “HIV Protease”, aquaporin, etc.
- PDB ID
 - 数字1文字と英数字3文字からなる、各立体構造データに固有のID
 - 例: 1HVR, 1J4N, etc.

検索結果の表示

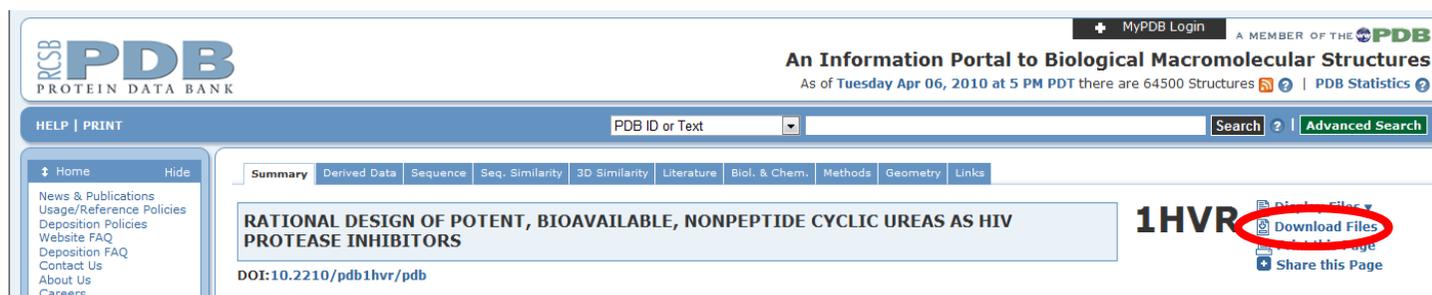
- 上部のタブをクリックして表示を切り替える



- Summary: 文献、組成
- Derived Data: 立体構造類似性に基づく分類
- Sequence: 配列データベースへのリンク、2次構造
- Biol. & Chem: 生物学・化学情報
- Methods: 立体構造決定方法に関する情報

データのダウンロード

- 右上の「Download Files」から立体構造データをダウンロードできる



- 「PDB File (Text)」を選び、デスクトップの保存する

PDBフォーマット(1)

- PDBファイルをワードパッドを用いて開く
- フォントを「MS ゴシック」にすると見やすくなる
- 冒頭部分には、生体高分子の名前や由来、文献等のデータが記載されている

```
HEADER      HYDROLASE(ACID PROTEINASE)                14-FEB-94   1HVR      1HVR      2
COMPND      HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1) PROTEASE      1HVR      3
COMPND      2 COMPLEXED WITH XK263 OF DUPONT MERCK                    1HVR      4
SOURCE      BH102 HIV-1 EXPRESSED IN (ESCHERICHIA COLI)              1HVR      5
AUTHOR      C.-H.CHANG                                                    1HVR      6
REVDAT      2   15-MAY-95 1HVRA    1          REMARK                    1HVRA    1
REVDAT      1   26-JAN-95 1HVR     0                                     1HVR     7
JRNL        AUTH    P.LAM,P.JADHAV,C.EYERMANN,N.HODGE,Y.RU,L.BACHELER,  1HVR     8
JRNL        AUTH  2  J.MEEK,M.OTTO,M.RAYNER,Y.WONG,C.-H.CHANG,P.WEBER,  1HVR     9
JRNL        AUTH  3  D.JACKSON,T.SHARPE,S.ERICKSON-VIITANEN            1HVR    10
JRNL        TITL    RATIONAL DESIGN OF POTENT, BIOAVAILABLE,          1HVR    11
JRNL        TITL  2  NONPEPTIDE CYCLIC UREAS AS HIV PROTEASE INHIBITORS  1HVR    12
JRNL        REF     SCIENCE                V. 263    380 1994          1HVR    13
JRNL        REFN    ASTM SCIEAS   US ISSN 0036-8075                0038    1HVR    14
REMARK      1                                               1HVR    15
REMARK      2                                               1HVR    16
REMARK      2 RESOLUTION. 1.8  ANGSTROMS.                        1HVR    17
```

PDBフォーマット(2)

①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪		
ATOM	1	N	PRO	A	1	-12.735	38.918	31.287	1.00	39.83	1HVR	181
ATOM	2	CA	PRO	A	1	-12.709	39.097	29.830	1.00	39.29	1HVR	182
ATOM	3	C	PRO	A	1	-13.575	38.051	29.162	1.00	39.78	1HVR	183
ATOM	4	O	PRO	A	1	-14.097	37.126	29.753	1.00	38.67	1HVR	184
ATOM	5	CB	PRO	A	1	-11.243	39.010	29.398	1.00	37.79	1HVR	185
ATOM	6	CG	PRO	A	1	-10.636	38.128	30.469	1.00	38.69	1HVR	186
ATOM	7	CD	PRO	A	1	-11.368	38.593	31.729	1.00	37.10	1HVR	187
ATOM	8	1H	PRO	A	1	-13.142	39.756	31.758	0.00	15.00	1HVR	188
ATOM	9	2H	PRO	A	1	-13.429	38.158	31.502	0.00	15.00	1HVR	189
ATOM	10	N	GLN	A	2	-13.682	38.255	27.876	1.00	41.01	1HVR	190

①レコード名(標準アミノ酸はATOM、非標準はHETATM)

②原子番号

③原子名(主鎖アミド窒素:N、 α 炭素:CA、 β 炭素:CBなど)

④残基名(3文字表記)

⑤Chain ID

⑥残基番号(配列データベース中の番号に一致させる)

⑦⑧⑨それぞれ原子のx, y, z座標 [Å]

⑩occupancy(その原子の重み因子、通常は1.00)

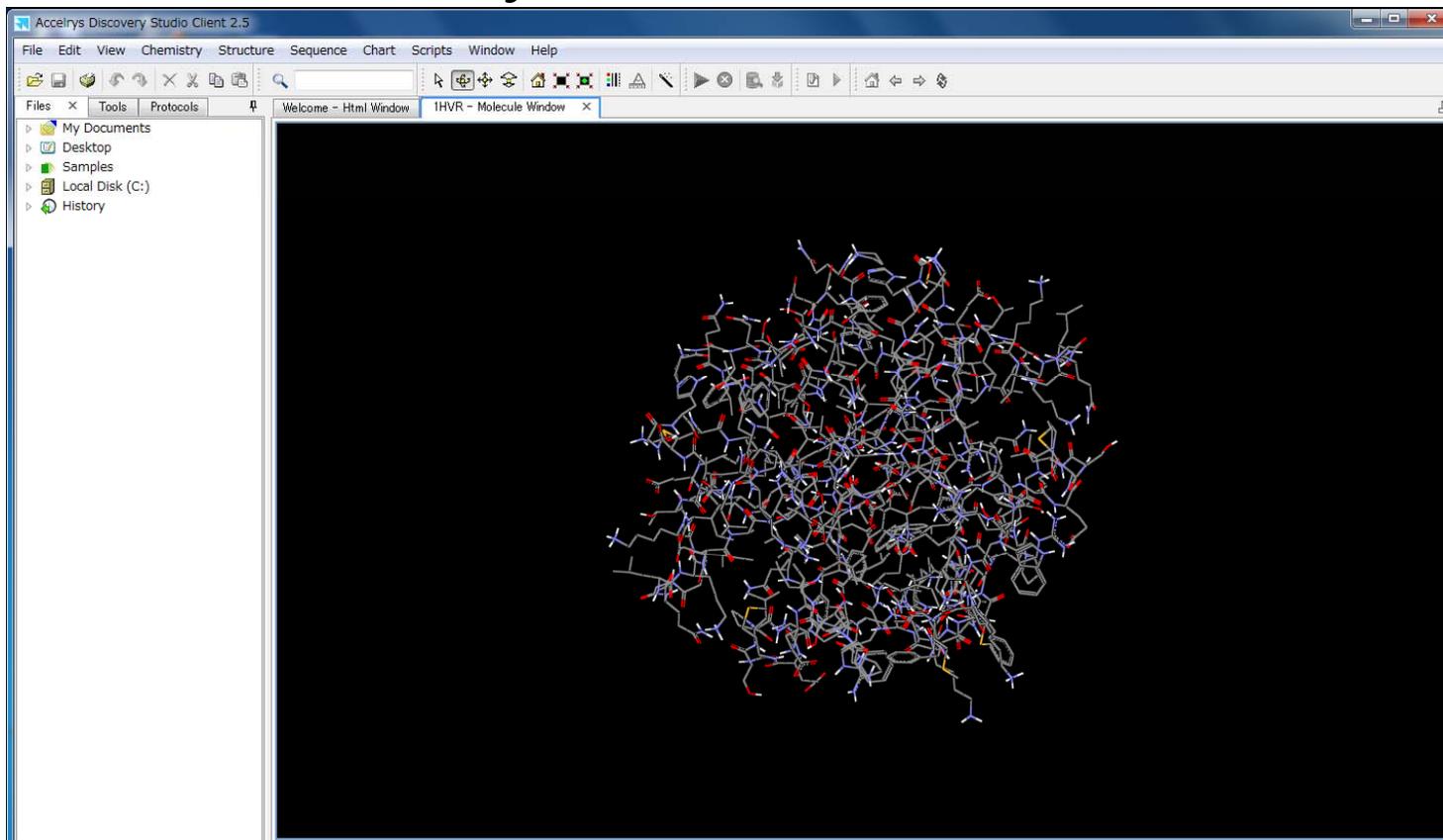
⑪温度因子B [Å²](X線結晶解析で決定されている場合のみ意味がある)

立体構造データの可視化

- 立体構造はPDBデータを可視化するソフトウェアを用いて表示できる
 - RasMol (<http://www.openrasmol.org/>)
 - PyMol (<http://www.pymol.org/>)
 - Swiss PDB Viewer (<http://spdbv.vital-it.ch/>)
 - UCSF Chimera
(<http://www.cgl.ucsf.edu/chimera/>)
- 本講義では、Discovery Studioを使用

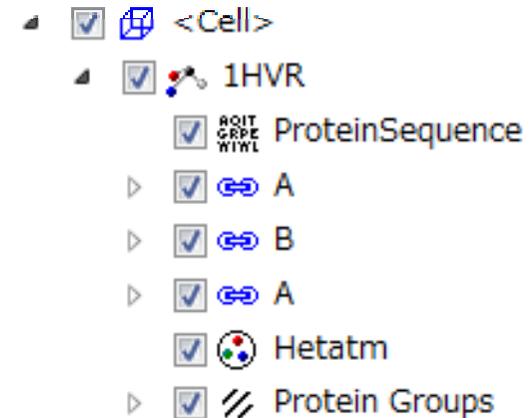
可視化ソフトウェアの起動

- ダウンロードしたファイルのアイコン  をダブルクリック→Discovery Studio Client 2.5が起動する



Hierarchy Windowの表示

- メニューの「View」→「Hierarchy」からHierarchy Windowを表示する
- ▷をクリックすると階層(チェーン・残基など)に属する残基・原子が展開して表示される
- チェックボックスのオン・オフで表示・非表示が切り替えられる



3D Windowの操作(1)

- 3D Windowの中を左クリックしてアクティブにしてから以下の操作を行う
- 回転
 - をクリックしてから3D Windowの中で左ドラッグ
- 並進
 - をクリックしてから3D Windowの中で左ドラッグ
- ズーム
 - をクリックしてから3D Windowの中で左ドラッグ

3D Windowの操作(2)

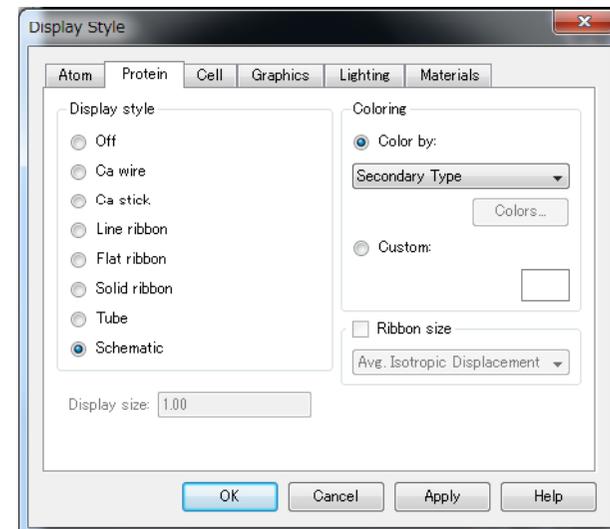
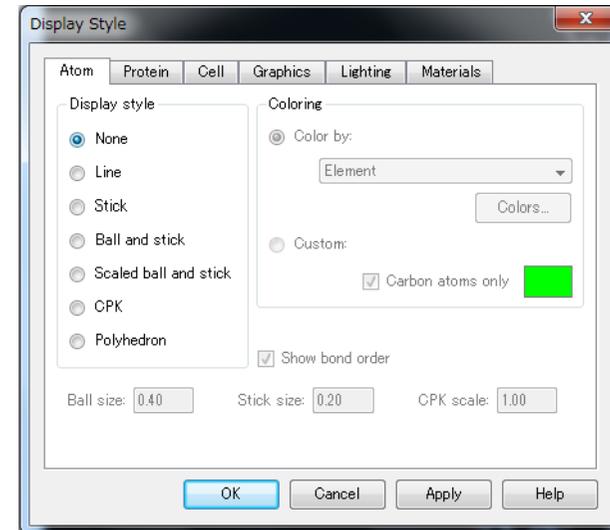
- 選択
 - 3D Window上で原子をクリック→その原子が選択され、黄色い四角でマークされる
 - 3D Window上で原子をダブルクリック→その原子を含む残基が選択され、マークされる
 - Hierarchy Windowでもチェーン、残基、原子、グループ (backboneなど) 単位で選択できる
 - 何も無いところをクリックすると選択を解除できる
 - Hierarchy WindowではCtrlキーを、3D WindowではShiftキーを押しながらクリックすると複数選択ができる
- 属性 (attribute)
 - 選択した後右クリックで表示できる

3D Windowの操作(3)

- Home 
 - 最初の向き、位置に戻す
- Fit to Screen 
 - (選択した)構造をWindowにフィットするように並進、拡大・縮小
- Center Structure 
 - (選択した)構造の中心がWindowの中心に来るように並進

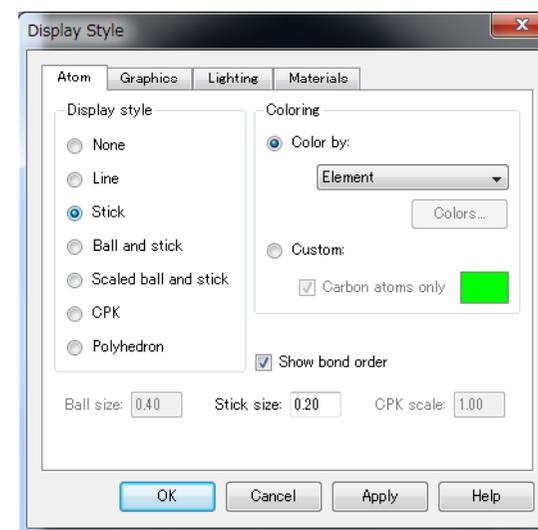
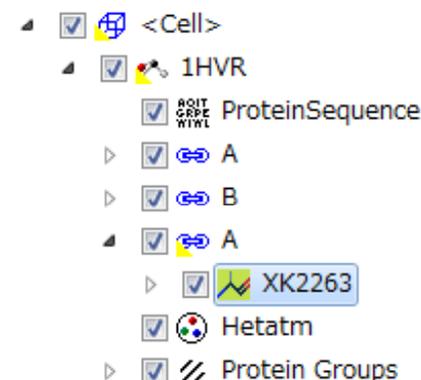
リボン表示

- Display Styleアイコン  をクリック
- 「Atom」タブをクリック
- 「None」を選択
- 「Protein」タブをクリック
- 「Schematic」を選択
- 「Coloring」をColor by:
「Secondary Type」に設定
- 「OK」ボタンを押す



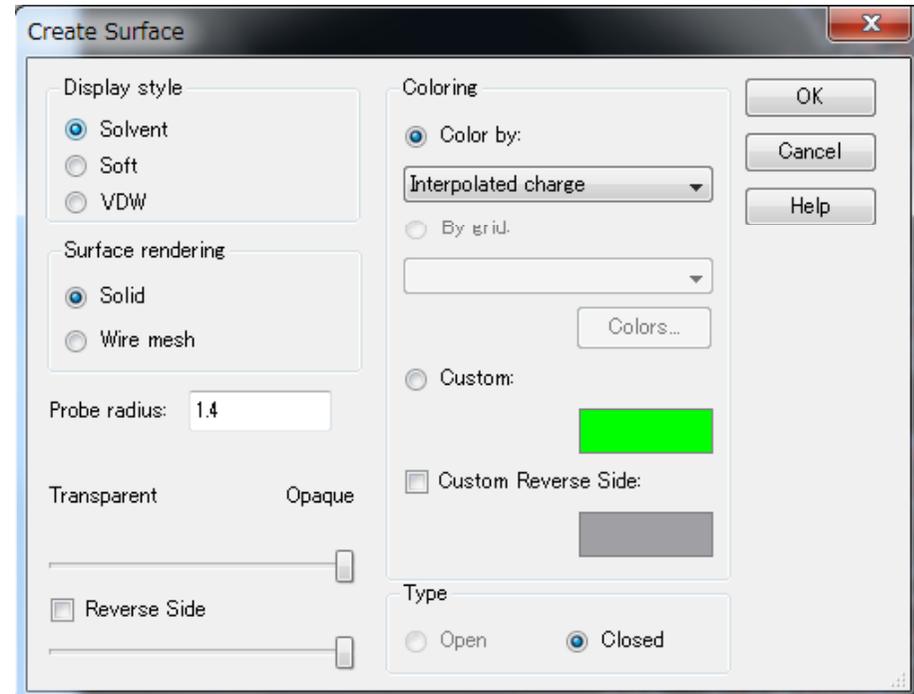
リガンドの表示

- Hierarchy Windowで2つ目のAの左側の▷マークをクリックして展開
- リガンド(XK2263)をクリックして選択
→黄色でマークされる
- Display Styleアイコンをクリック「Stick」を選択し「OK」



溶媒接触表面の表示

- Display Styleアイコン  をクリックし、「Atom」タブで「Line」を選択、「Protein」タブで「Off」を選択し「OK」→もとの表示に戻す
- メニューの「Chemistry」→「Charge」→「Calculate」で部分電荷を計算
- メニューの「Structure」→「Surface」→「Add」を選択
- 「Display Style」で「Solvent」を選択し「OK」→静電ポテンシャルで色分け

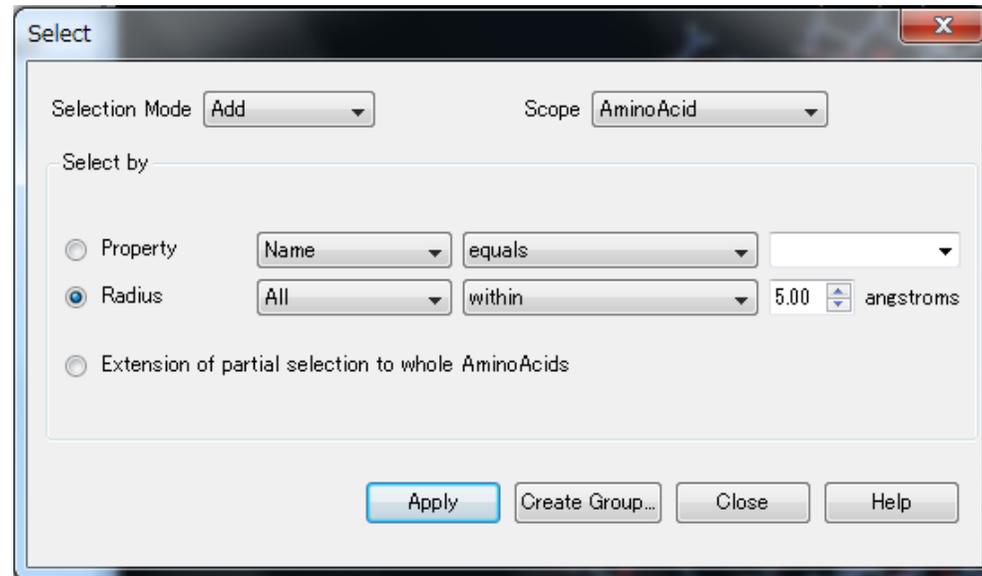


相互作用の検出(1)

- 水素結合の検出
 - X線結晶構造解析から得られた構造には水素原子の座標が含まれていないことが多いため、重原子間の距離で判定する
 - 水素結合を形成する重原子(窒素や酸素)間の距離は概ね2.8 Å ~ 3.5 Å
- Hierarchy WindowでリガンドXK2263を選択
- メニューの「Structure」→「Monitor」→「Intermolecular HBonds」を選択→分子間水素結合ペアが緑色の破線で表示される

相互作用の検出(2)

- 疎水性相互作用は原子間距離で検出
- Hierarchy WindowでリガンドXK2263を選択
- メニューの「Edit」→「Select」を選択
- Selection Modeを「Add」、Scopeを「AminoAcid」
- 「Radius」を選択し、右図のように設定
- 「Apply」を押す
→リガンドから5 Åにある残基が選択される



データの保存

- メニューの「File」→「Save As」でデータを保存できる
- 3D Windowをアクティブにして保存すると、独自形式 (*.dsv) やPDB形式 (*.pdb) の他、画像 (*.png) としても保存できる
- Sequence Windowをアクティブにして保存すると、Fasta形式 (*.fasta) 等で保存できる

立体構造決定法

- 主な立体構造決定法
 - X線結晶構造解析
 - 核磁気共鳴(NMR)法
 - 電子顕微鏡
- エントリ数の比較(平成22年4月現在)

} 第2回以降
の講義で解
説

実験の種類	エントリ数	割合 [%]
X線結晶構造解析	55,836	86.4
NMR	8,336	12.9
電子顕微鏡	288	0.4
複数	19	0.0
その他	144	0.2
合計	64,623	100.0

立体構造データ利用上の注意

X線結晶構造解析

- 精度の指標: 分解能
- 立体構造データの特徴
 - 立体構造データに水素原子が含まれない
 - 運動性が高い領域で原子座標が欠失していることがある
- 結晶学的な最小単位の立体構造が登録されている
- 結晶中で隣接分子と非生理的な相互作用をすることがある

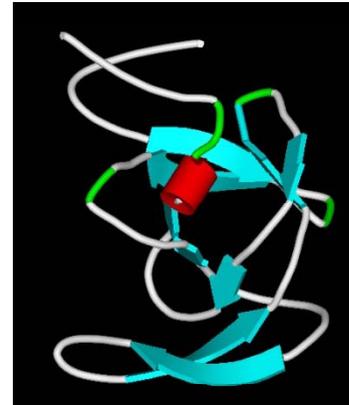
NMR

- 精度の指標: RMSD*
- 立体構造データの特徴
 - 実験データを満たす構造の集まりで表現
 - 実験データが少ない領域で立体構造がばらつく
- 実験で使用した分子の立体構造が登録されている
- 溶液中で測定するため、非生理的な相互作用をすることはない

*RMSD: root mean square deviation (根平均2乗変位)

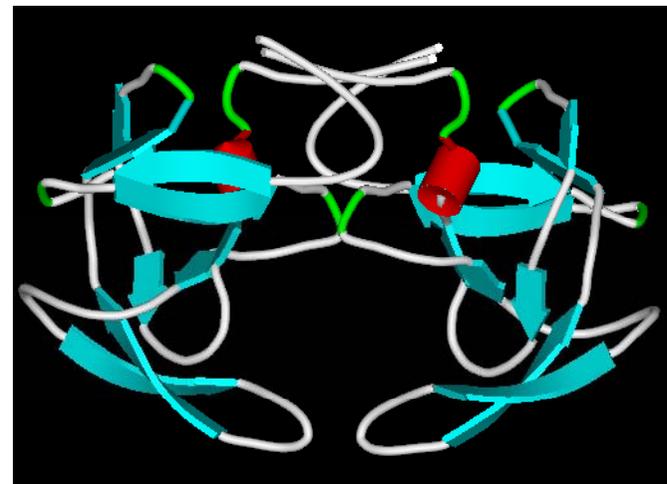
Biological unit

- 生物学的に機能する最小限の分子構成を biological unit と呼ぶ
- X線結晶構造解析では、登録されている立体構造が biological unit とは異なる場合がある
- RCSBのサイトでは biological unit の座標がダウンロードできる

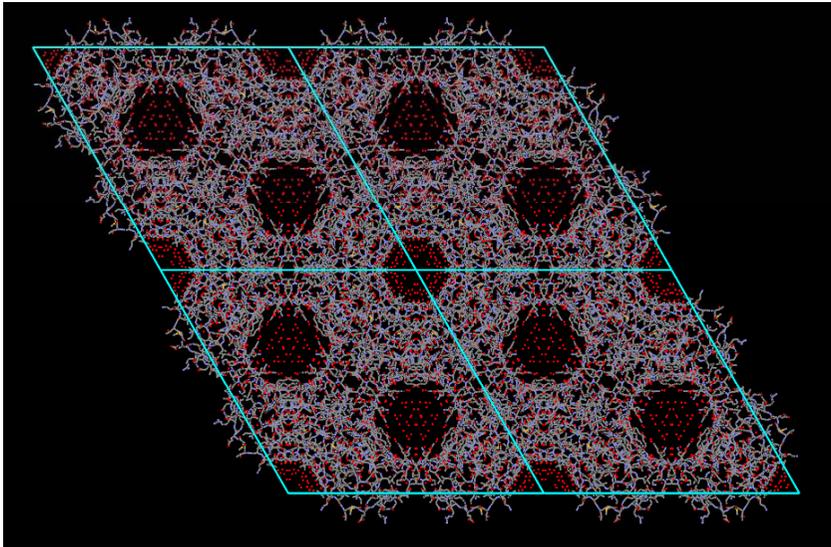


3PHVに登録されている座標

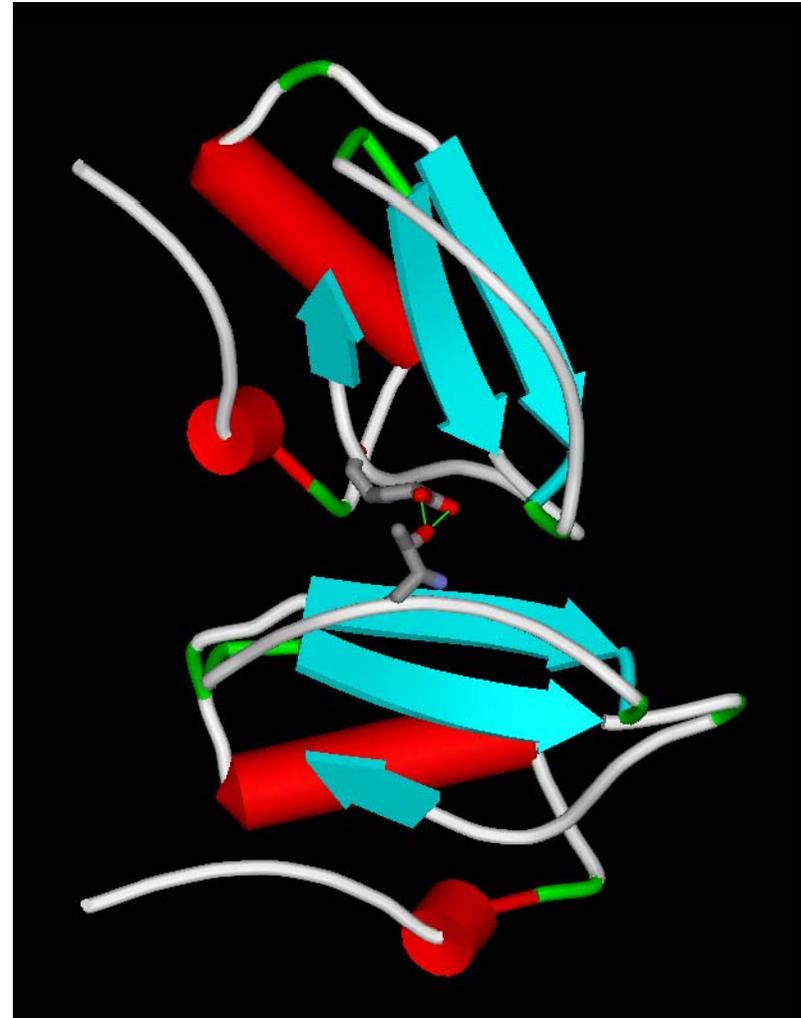
Biological unitの座標



隣接分子との接触の例

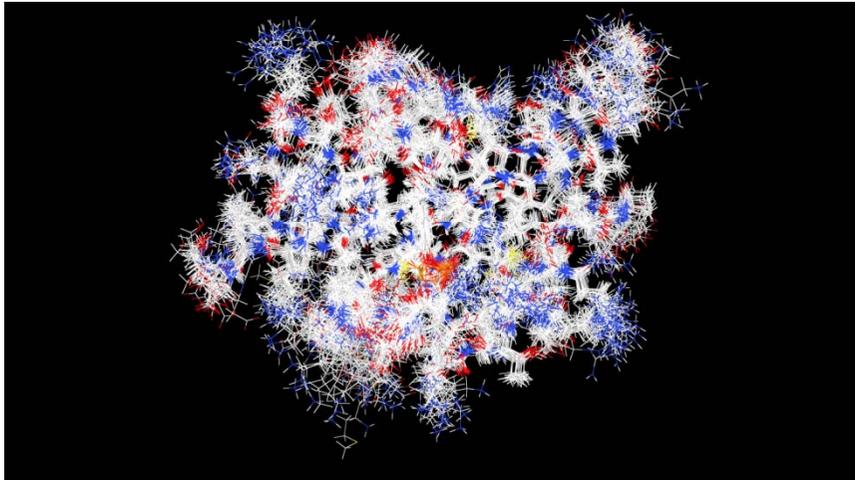


Chymotrypsin Inhibitor 2
(2CI2)の例
1単位格子に12分子
隣接分子間で水素結合を
形成している

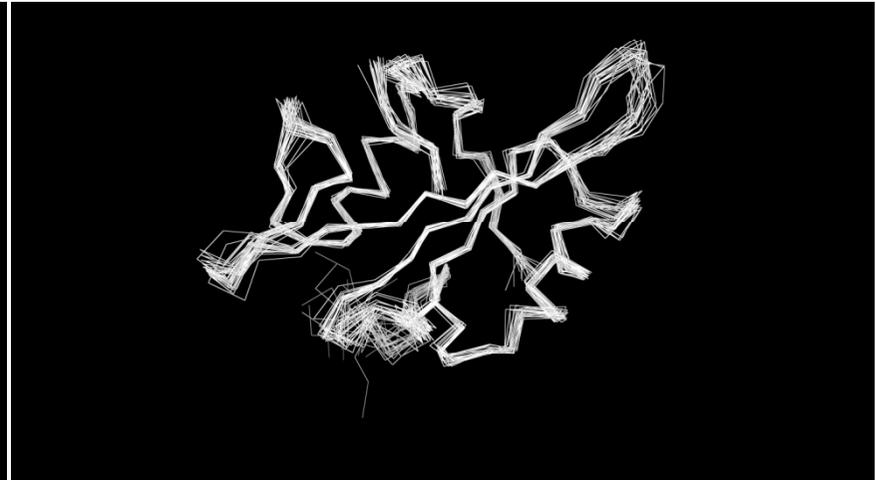


NMR構造の例

- UCSF Chimeraを使用
- メニューの「File」→「Fetch by ID」を選択し、PDB IDとして1HCTを指定して「Fetch」



「Action」→「Atoms/Bonds」→「show」
「Action」→「Atoms/Bonds」→「wire」
「Action」→「Ribbon」→「hide」



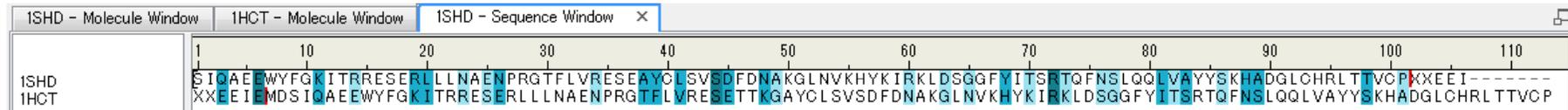
「Action」→「Atoms/Bonds」→
「backbone only」→「chain trace」

構造の比較(1)

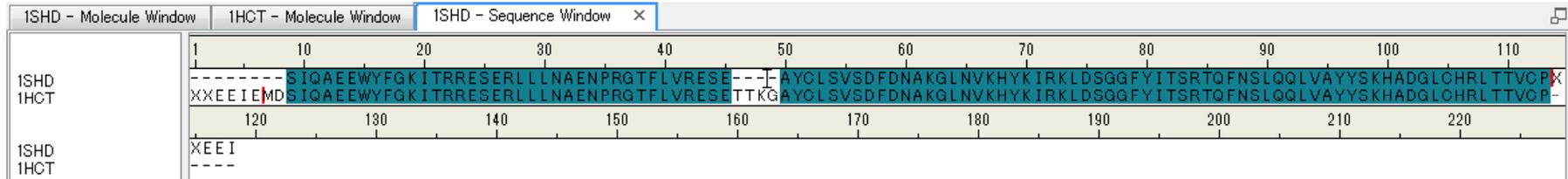
- ヒトSrc SH2ドメインと基質ペプチドの複合体はX線結晶構造解析もなされている
 1. 「File」→「Open URL」を選択
 2. PDB IDを選び、1SHDと入力して「Open」
 3. 「File」→「Open URL」を選択
 4. PDB IDを選び、1HCTと入力して「Open」
 5. Hierarchy windowで1HCT_model_1をクリックして選択したのち、右クリックして「Copy」
 6. 1SHDの3D Windowで右クリックして「Paste」

構造の比較(2)

7. 「Sequence」→「Show Sequence」を選択



8. Spaceキーを押してギャップを挿入し、アライメントを以下のように修正する

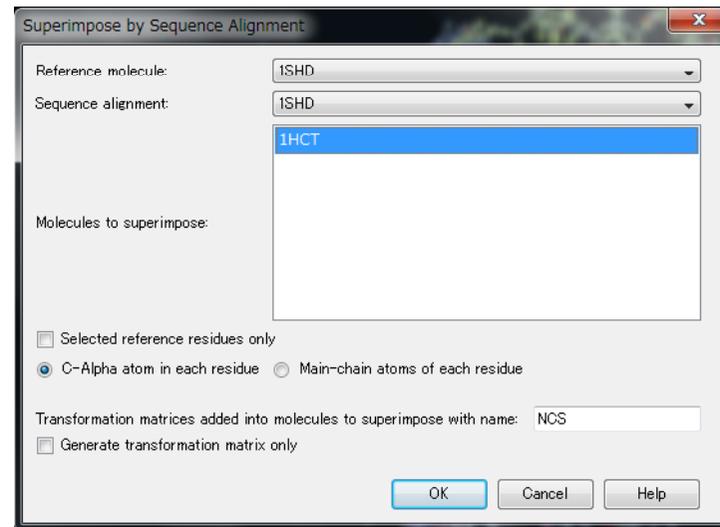


1SHDの先頭に7つギャップを挿入

1SHDのGLU181の後に4つギャップを挿入

構造の比較(3)

9. Sequence windowで1SHDを選択
(配列の背景が黒色になる)
10. 1SDHの3D Windowをアクティブにし、
「Structure」→「Superimpose」→
「By Sequence Alignment」を選択
11. 「Molecules to superimpose」のリストから1HCT
を選び、「OK」
12. RMSDが表示され、
3D Windowでは1HCT
の構造が1SHDに重ね
あわされている



配列データベースからのリンク

1. 代表的なタンパク質配列データベース
Swiss-Prot (<http://www.expasy.ch/sprot/>)
を開く
2. 右上のテキストボックスに“SRC_HUMAN”と
入力し「Go」
3. 検索結果の下のほうに、“3D structure
database”のセクションがあり、1HCTや
1SHDが現れていることを確認すること

配列からの検索(1)

1. NCBI BLASTのサイトにアクセス
(<http://blast.ncbi.nlm.nih.gov/Blast.cgi>)
2. Basic Blastにある「Protein BLAST」をクリック
3. 新しいタブで講義のページを開き、1HCT.fasta
をクリック
4. 右クリックして、「すべて選択」を選んだあと、再
び右クリックして、「コピー」
5. BLASTのページの「Enter accession
number, gi, or FASTA sequence」のテキスト
エリアの中で右クリックし、「貼り付け」

配列からの検索(2)

6. Choose Search SetのDatabaseを「Protein Data Bank proteins (pdb)」に設定
7. BLASTをクリック



The screenshot shows the NCBI BLAST web interface. The 'Enter Query Sequence' section contains a text area with a protein sequence: >lhct:B|PDBID|CHAIN|SEQUENCE
HSD1QAEWIFGHTTRRESRELLLNENFRDTFLVRESETTGGAYCLSVSDFDNAKGLNKKYKIRK
LDSGQFYVTSRTQ
FNSLQQLVAYYSKIDAGLCHRITVCP. Below this is a 'Job Title' field containing 'lhct:BPDBIDCHAINSEQUENCE'. The 'Choose Search Set' section has a dropdown menu for 'Database' set to 'Protein Data Bank proteins (pdb)'. The 'Program Selection' section has 'blastp (protein-protein BLAST)' selected. At the bottom, a 'BLAST' button is visible, and the search parameters are summarized as 'Search database Protein Data Bank proteins(pdb) using Blastp (protein-protein BLAST)'. A note at the bottom states: 'Note: Parameter values that differ from the default are highlighted in yellow and marked with + sign'.

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: MNYNKNILVSEGLSQKHLIHG)

jCE algorithm - Java Web Start

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

HELP | PRINT

↓ Home Hide

News & Publications
Usage/Reference Policies
Deposition Policies
Website FAQ
Deposition FAQ
Contact Us
About Us
Careers
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
Latest Publications
Sequence Search
Chemical Components
Unreleased Entries
Browse Database
Histograms

↓ Tools Hide

File Downloads
FTP Services
File Formats
Services: RESTful | SOAP
Web Services
Compare Structures

↓ Education Hide

Understanding PDB Data
Molecule of the Month
Educational Resources

立体構造比較(2)

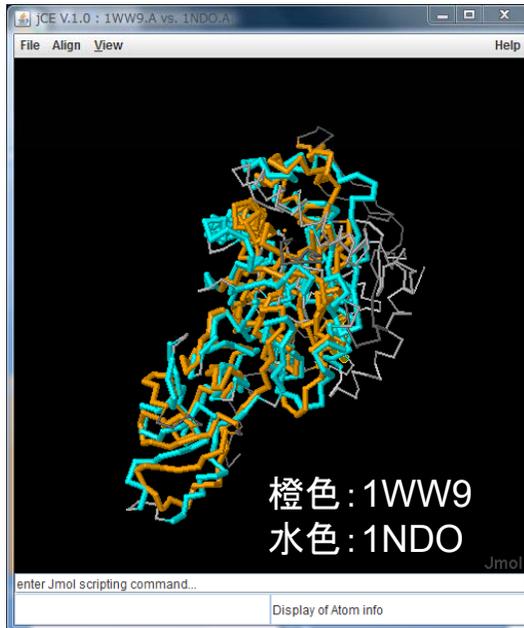
4. 以下をクリックして、jCEを起動

Java Web Start

The comparison should automatically start using **Java Web Start**. If it does not, please click here:

Launch Structure Alignment for 1WW9.A vs. 1NDO.A.

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



```
EQR:300Len1:389Len2:447score:1023.75Z-score:5.86RMSD:4.34Sim1:77%Sim2:67%
13:A  KGW-----APYV---DAKLG----FRNHWYPVMFSKEINE- GEPKTLKLLGENLLVNRID- GKLY 63:A
6:A  KILVSESGLSQKHLIHGDEELFQHELKTI FARNWLELTHDSLIPAPGDYVTAKMGI DEVI VSRQNDGSI R 75:A

64:A  CLKDRCLHRGVQLSVK- - -VECKTKSTITCWYHAWTYRWE DGVLCDI LTNP- - - TSAQI GROKCLKTY- P 125:A
76:A  AFLNVCRRHRGKTLVSV EAGNAK- - - GFVCSYHGWGEGSN- GELQSVPEK DLYGESLNKKCLGLKEVAR 140:A

126:A VQEA KGCVF IY LGD GPPPLARDT- - PPNFLD- - - DDMEI LGK- - - NQI I KS- - NWRLAVENG- FDP SHI 184:A
141:A VESEHGFI YGC- EDQEAPPLMDYLGDAAWYLEP MFKHS GGLELV GPPGKVV I KANWKAPAENE VGDAYHV 209:A

184:A - YI HKDSI LVKDN DLALPL GFAP GDRKQQT RVV DDDVV GRKG VYDL I GEHGV PVF EGTI GGEVV REGAY 253:A
209:A GWHASSLRSGESI FSSLA-----GNAAL 233:A

254:A GEKI VANDI SI WLP GVLKVPFPNP-----DMMQFEW- - 285:A
234:A PPEGA- GLQMTSKY GSGMGVLWDGYS GVHSADLVP ELMAF GGA KQERLNKEI GDVRARI YRSHLNCTVFP 299:A

285:A -----YVPI DE- - - NTHYYFQTL GKPCANDEERKKYE QEFESKWKPMAL EGF- - - NNDDI WAR 337:A
299:A NNSMLTCSGVFKVWNPI DANTTEVWTYALV-----EKDMPEDLKRRLADSVQRTFGPAGFWE SDDN 363:A

338:A EAMVDFYADDKGWVNEI LFESEEA-----I VAWRKLASEHNQ 374:A
364:A DNMETASQNGKKYQSRDSDLLSNLGF GEDVY GDAVYPGVV GKS AI GETSYRGFYRAYQAHVS 425:A

allg pos:0 [LYS]13:A (K) : [LYS]6:A (K)
```

上段: 1WW9、下段: 1NDO

立体構造比較(3)

- 類似性の指標

- Z-Score: 5.86 (偶然に類似する確率; 大きいほど低い)

>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 (立体構造のずれ)

立体構造比較(4)

- 3.でComparison methodを「blast2seq」にすると、配列アラインメントが表示される

```
blast2seq
RMSD: 6.11 Å
1WW9: 41 GEPKTLKLLGENLLVNRI-DGKLYCLKDRCLHRGVQLSVKVECKTKSTITCWHAWTYRW 99
          G+ T K+ + ++V+R DG + + C HRG L V VE C YH W +
1NDO: 52 GDYVTAKMGIDEVIVSRQNDGSIRAFNLNCRHRGKTL-VSVEAGNAKGFVCSYHGWF-- 108

1WW9: 100 EDGVLCDILTNP TSAQIGRQKL-----KTPVQEAKGCVFIY-LGDGDPPPL----- 145
          G ++ + P + + L + V+ G FIY D + PPL
1NDO: 109 --GSNGELQSVPF EKDL YGESLNKKCLGLKEVARVESFHG--FIYGCFDQEAPPLMDYLG 164

1WW9: 146 --ARDTPPNFL-DDDMEILGK--NQIIKSNWRLAVEN 177
          A P F +E++G +IK+NW+ EN
1NDO: 165 DAAWYLEPMPFKHSGGLELVGPPGKVVIKANWKAPAEN 201

CE
RMSD: 4.34 Å
1WW9: 13 KGW-----APYV---DAKLG----FRNHWYPVPMFSKEINE-GEPKTLKLLGENLLVNRID-GKLY
          ||| |||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1NDO: 6 KILVSEGLSQKHLIHGDEELFQHELKTI FARNWLF LTHDSLIPAPGDYVTAKMGIDEVIVSRQNDGSIR

1WW9: 64 CLKDRCLHRGVQLSVK---VECKTKSTITCWHAWTYRWEDGVLCDILTNP---TSAQIGRQKLKTY-P
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1NDO: 76 AFLNVC RHRGKTLVSVEAGNAK----GFVCSYHGWF GSN-GELQSVPF EKDL YGESLNKKCLGLKEVAR

1WW9: 126 VQEA KGCVFIY LGDGDPPPLARDT--PPNFLD---DDMEILGK---NQIIKS--NWRLAVENG-FDPSHI
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1NDO: 141 VESFHGFIYGC-FDQEAPPLMDYLG DAAWYLEPMPFKHSGGLELVGPPGKVVIKANWKAPAENFVGDAYHV
```

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/>)

階層的立体構造分類データベース

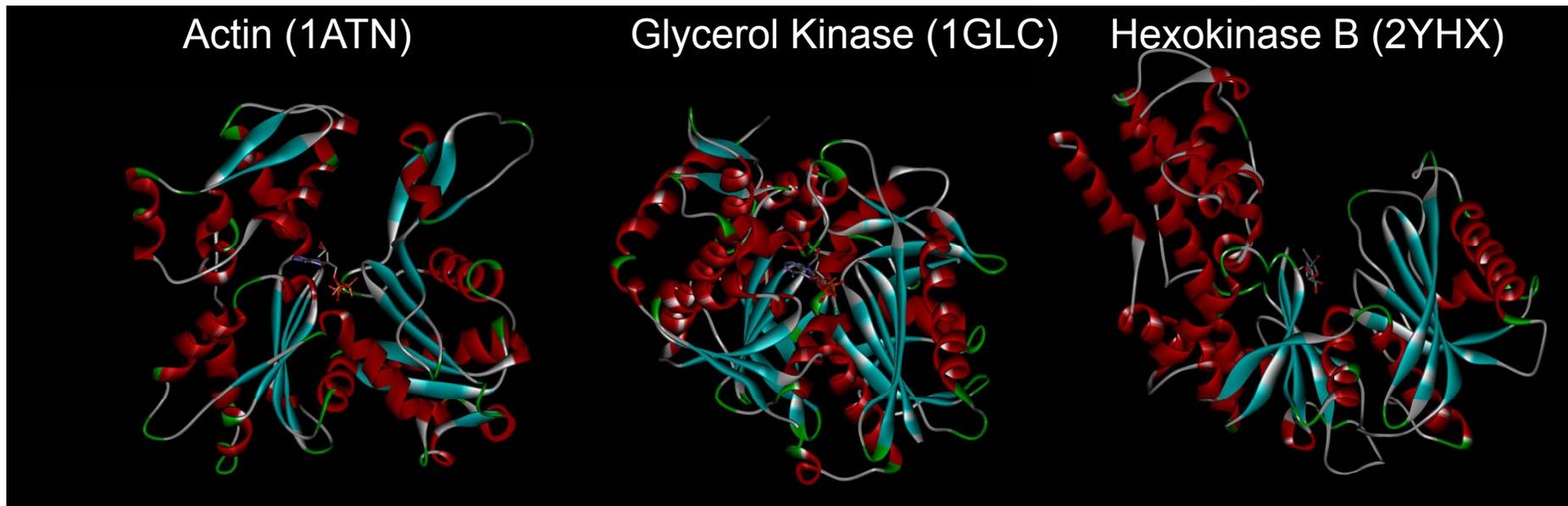
- 立体構造を立体構造類似性に基づいて階層的に分類したデータベース
 - SCOP (<http://scop.mrc-lmb.cam.ac.uk/scop/>)
主に人手による分類
Class → Fold → Superfamily → Family
 - CATH (<http://www.cathdb.info/>)
主に計算機による分類
Class → Architecture → Topology → Homologous Superfamily
- 例: 1HCTのSCOPにおける分類
 - Alpha and beta proteins → SH2-like → SH2 domain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

SCOPにおける階層

- Family: 配列相同性が概ね30%以上あり、進化的な類縁関係が明らかなもの
- Superfamily: 配列相同性は低いが、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).

実習課題

- 講義のページのkagai.fastaの配列を持つタンパク質について、以下に答えよ
 1. このタンパク質の名前は何か
 2. このタンパク質の立体構造のPDB IDおよび立体構造決定に使用した方法を答えよ
 3. Discovery Studioを用いて、タンパク質を「Solid ribbon」、水分子を「CPK」として立体構造を表示せよ
 4. このタンパク質はいくつのポリペプチドで機能していると考えられるか？
 5. このタンパク質と進化的に類縁関係にあると考えられるタンパク質のうち、立体構造が決定されているタンパク質を2つ以上挙げ、このタンパク質と、配列および立体構造がどの程度類似しているか答えよ

課題の提出

- 課題の回答はメールで寺田宛 tterada@iu.a.u-tokyo.ac.jpに提出すること
- メールのは件名は「構造バイオインフォマティクス基礎課題」とし、本文に、氏名、所属、学生証番号(学生のみ)、課題1, 2, 4, 5の回答を記載し、3の画像(PNG形式)を添付すること
- メールの送信には、アカウント発行元のWeb mail等を用いること。その際、メールアドレスやパスワードがブラウザに残らないように注意すること。(インターネットエクスプローラーのメニューから、「セーフティ」→「閲覧の履歴の削除」を開き、「フォームデータ」、「パスワード」をチェックして「削除」)