

平成24年4月19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19日(木)
担当: 寺田 透
内容: 立体構造データベースの利用と立体構造データの可視化
2. 4月26日(木)
担当: 永田宏次
内容: X線結晶構造解析による立体構造決定のインフォマティクス
3. 5月10日(木)
担当: 寺田 透
内容: 立体構造からの情報抽出
4. 5月17日(木)
担当: 清水謙多郎
内容: 立体構造のモデリング

本日の講義内容

- タンパク質立体構造データベース
 - 検索
 - データのダウンロード
- 立体構造データの可視化
- 立体構造データフォーマット
- 配列データベースとの連携
- 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
- 2012年4月時点でのエントリ数は約81,000
- 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 displays the RCSB PDB website interface. At the top left is the RCSB PDB logo (Protein Data Bank) and a search box containing 'PDB-101'. To the right, it states 'A MEMBER OF THE PDB' and 'An Information Portal to Biological Macromolecular Structures'. Below this, it shows the date and time: 'As of Tuesday Apr 10, 2012 at 5 PM PDT there are 80710 Structures | PDB Statistics'. A navigation bar includes 'All Categories', 'Author', 'Macromolecule', 'Sequence', and 'Ligand'. A search bar is labeled 'Search | All Categories:' and contains the text 'e.g., PDB ID, molecule name, author'. On the left side, there are navigation menus for 'Customize This Page', 'MyPDB' (with 'Login to your Account' and 'Register a New Account' options), and 'Home' (with links like 'News & Publications', 'Usage/Reference Policies', etc.). The main content area is titled 'Biological Macromolecular Resource' and features a 'Full Description' section. Below this is a 'Featured Molecules' section with a 'Hide' button and a 'List View of Archive By: Title | Date | Category' option. The featured molecule is 'Ras Protein', described as the 'Molecule of the Month'. The text states: 'Cells are constantly sending messages, discussing nutrient levels and growth rates with other cells, and also managing the internal needs of the cell. These messages need to be clear and strong, so that they can be heard over the busy bustle inside the crowded cytoplasm.' A 'Full Article' link is provided. On the right side, there are three panels: 'New Structures' (with 'Latest Release', 'New Structure Papers', and 'Search Unreleased Entries'), 'New Features' (with 'Links to DrugBank' and a 'Website Release Archive' dropdown), and 'RCSB PDB News' (with 'Weekly | Quarterly | Yearly' options and a 'Spring Newsletter' link for 2012-04-10).

検索

RCSB PDB PROTEIN DATA BANK

RCSB PDB-101

A MEMBER OF THE PDB

An Information Portal to Biological Macromolecular Structures

As of Tuesday Apr 10, 2012 at 5 PM PDT there are 80710 Structures | PDB Statistics | RSS | Help | Print

All Categories Author Macromolecule Sequence Ligand

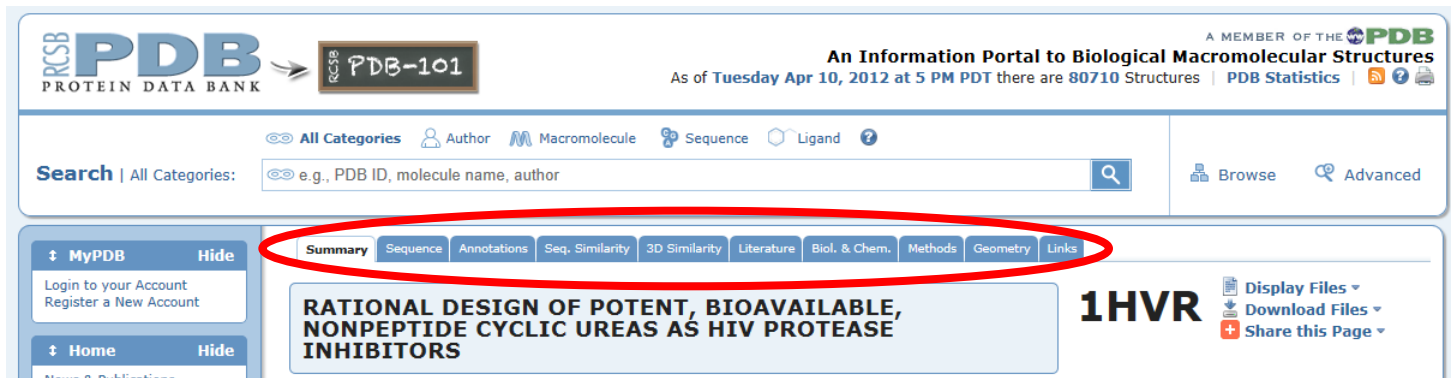
Search | All Categories: e.g., PDB ID, molecule name, author

Browse Advanced

- 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

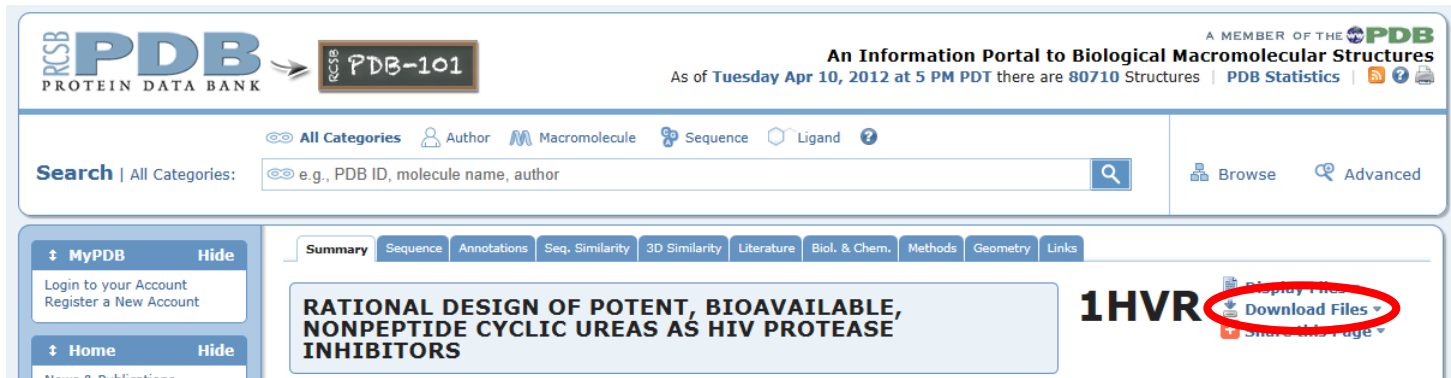


The screenshot shows the PDB website interface. At the top, there is a search bar with the text "Search | All Categories:" and a search button. Below the search bar, there is a navigation menu with tabs: "Summary", "Sequence", "Annotations", "Seq. Similarity", "3D Similarity", "Literature", "Biol. & Chem.", "Methods", "Geometry", and "Links". The "Summary" tab is highlighted with a red circle. The main content area displays the title "RATIONAL DESIGN OF POTENT, BIOAVAILABLE, NONPEPTIDE CYCLIC UREAS AS HIV PROTEASE INHIBITORS" and the PDB ID "1HVR". On the right side, there are options to "Display Files", "Download Files", and "Share this Page".

- Summary: 文献、組成
- Sequence: アミノ酸配列、2次構造
- Annotations: 立体構造分類、ファミリー分類
- Methods: 立体構造決定法

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



The screenshot shows the RCSB PDB website interface. At the top, the logo for RCSB PDB (Protein Data Bank) is visible, along with the text 'A MEMBER OF THE PDB'. Below the logo, there is a search bar and a navigation menu. The main content area displays the title 'RATIONAL DESIGN OF POTENT, BIOAVAILABLE, NONPEPTIDE CYCLIC UREAS AS HIV PROTEASE INHIBITORS' and the PDB ID '1HVR'. A red circle highlights the 'Download Files' button in the top right corner of the main content area.

- 「PDB File (Text)」を選び、デスクトップの保存する

立体構造データの可視化

1. PDB ID「1HVR」を検索し表示
2. ファイルをダウンロードし、デスクトップに保存
3. Chimera 1.5.2のアイコンをダブルクリックし起動
4. メニューの「File」→「Open」で1HVR.pdbを開く



UCSF Chimeraの操作(1)

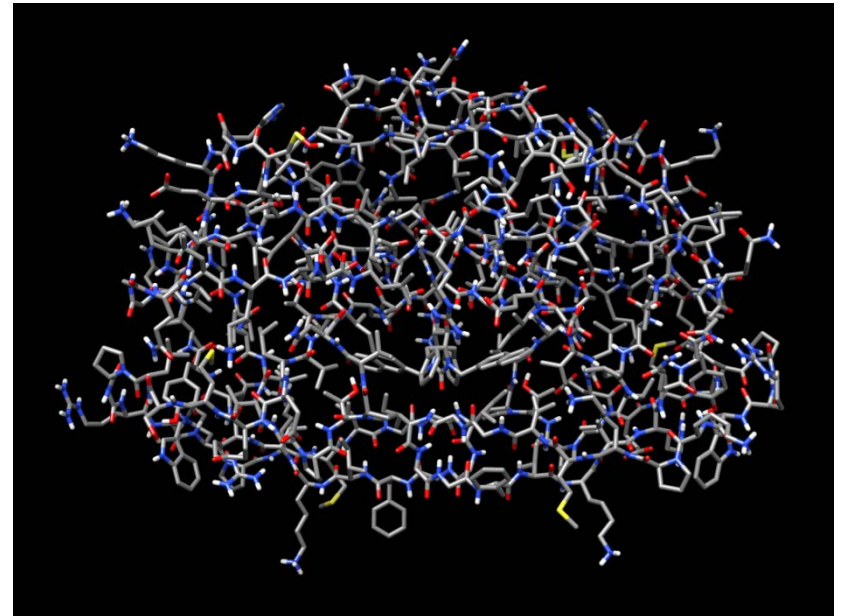
- 回転
 - マウスの左ボタンを押しながらドラッグ
- 並進
 - マウスのホイールを押しながらドラッグ
- ズーム
 - マウスの右ボタンを押しながらドラッグ
 - マウスのホイールを回転

UCSF Chimeraの操作(2)

- 選択(selection)
 - 「Ctrl」キーを押しながら左クリック
 - 選択を追加する時は、「Ctrl」と「Shift」キーを押しながら左クリック
 - 何もないところを「Ctrl」キーを押しながら左クリックすると解除
 - 「↑」キーで、選択範囲を原子→残基→チェーン→分子の順に拡大
- フォーカス
 - メニューの「Actions」→「Focus」で選択された原子を拡大表示する

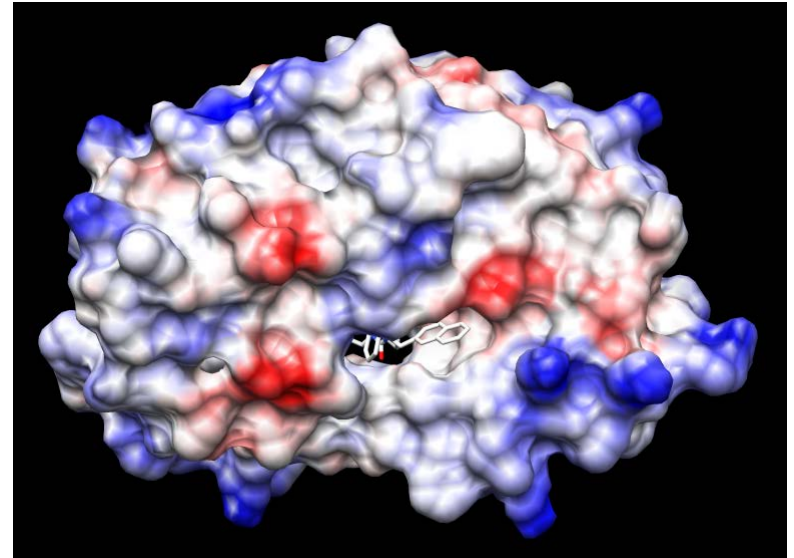
表示の変更(1)

- メニューの「Actions」を用いる
 1. 「Actions」→「Atoms/Bonds」→「show」
 2. 「Actions」→「Ribbon」→「hide」
 3. 「Actions」→「Color」→「by element」
- 選択している場合は、選択された原子の表示が変わる



表示の変更(2)

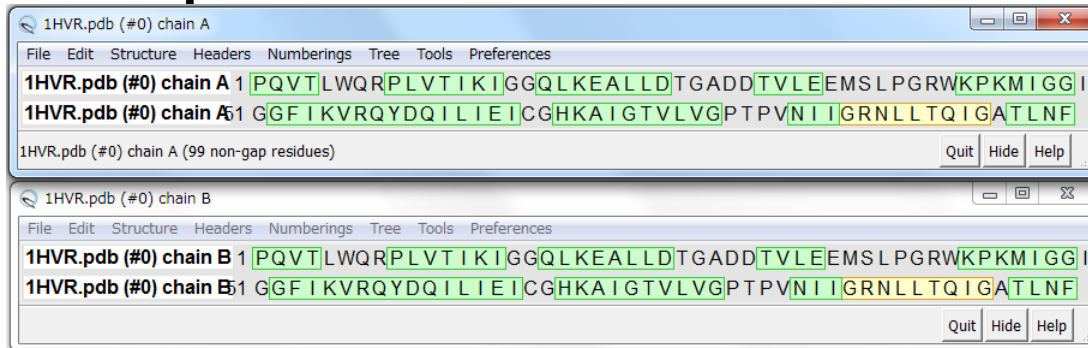
- 「Actions」→「Surface」→「show」で分子表面を表示
- 「Tools」→「Surface/Binding Analysis」→「Coulombic Surface Coloring」で静電ポテンシャルで色分け
(少し時間がかかる)



青: 正に帯電
赤: 負に帯電

配列の表示

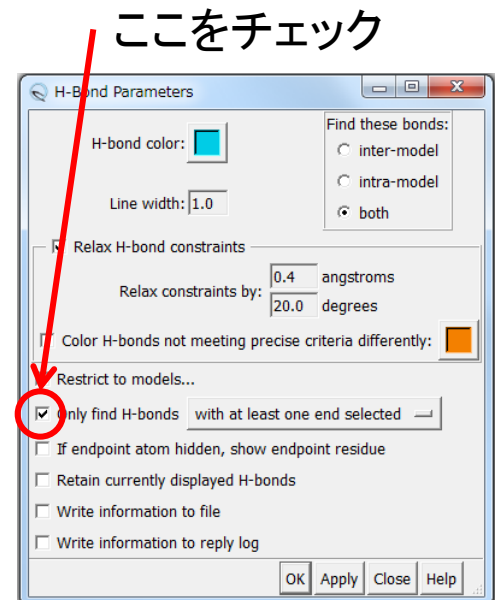
- メニューの「Tools」→「Sequence」→「Sequence」



- アミノ酸を選択すると、立体構造上でも選択される
 - マウスの左ドラッグで領域を選択
 - 「Shift」キーを押しながら左ドラッグで追加

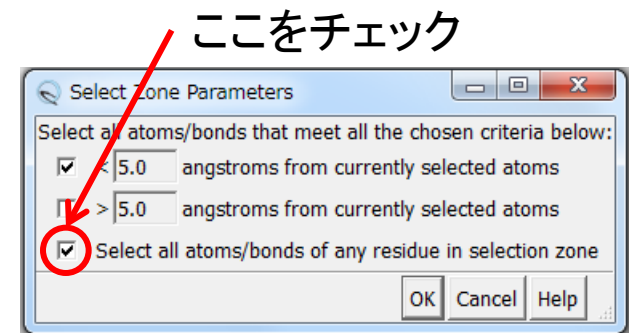
相互作用の検出(1)

- 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
- メニューの「Select」→「Residue」→「XK2」でリガンドを選択
- メニューの「Tools」→「Structure Analysis」→「FindHBond」を選択し右のように設定
→水素結合が青色の線で表示される



相互作用の検出(2)

- 疎水性相互作用は原子間距離で検出
- リガンドXK2を選択
- メニューの「Select」→「Zone」を選択し、右のように設定し「OK」→リガンドから5 Åにある残基が選択される
- 「Select」→「Name Selection」で選択範囲を保存して後で呼び出すことができる



データの保存

- メニューの「File」→「Save Session As」で作業状態を保存できる
- 保存した作業状態は、「File」→「Restore Session」で呼び出すことができる
- 画像は「File」→「Save Image」で保存できる
- 「File」→「Close Session」で立体構造データは閉じられ、初期状態に戻る

参考：可視化ソフトウェア入手先

- 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/>)
- Discovery Studio Visualizer
(<http://accelrys.com/products/discovery-studio/visualization-download.php>)

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
TITLE   RATIONAL DESIGN OF POTENT, BIOAVAILABLE, NONPEPTIDE CYCLIC
TITLE   2 UREAS AS HIV PROTEASE INHIBITORS
COMPND  MOL_ID: 1;
COMPND  2 MOLECULE: HIV-1 PROTEASE;
COMPND  3 CHAIN: A, B;
COMPND  4 ENGINEERED: YES
SOURCE  MOL_ID: 1;
SOURCE  2 ORGANISM_SCIENTIFIC: HUMAN IMMUNODEFICIENCY VIRUS 1;
SOURCE  3 ORGANISM_TAXID: 11676;
SOURCE  4 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE  5 EXPRESSION_SYSTEM_TAXID: 562
KEYWDS  HYDROLASE(ACID PROTEINASE)
EXPDTA  X-RAY DIFFRACTION
AUTHOR  C.-H.CHANG
```

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	N
ATOM	2	CA	PRO	A	1	-12.709	39.097	29.830	1.00	39.29	C
ATOM	3	C	PRO	A	1	-13.575	38.051	29.162	1.00	39.78	C
ATOM	4	O	PRO	A	1	-14.097	37.126	29.753	1.00	38.67	O
ATOM	5	CB	PRO	A	1	-11.243	39.010	29.398	1.00	37.79	C
ATOM	6	CG	PRO	A	1	-10.636	38.128	30.469	1.00	38.69	C
ATOM	7	CD	PRO	A	1	-11.368	38.593	31.729	1.00	37.10	C
ATOM	8	H2	PRO	A	1	-13.142	39.756	31.758	0.00	15.00	H
ATOM	9	H3	PRO	A	1	-13.429	38.158	31.502	0.00	15.00	H
ATOM	10	N	GLN	A	2	-13.682	38.255	27.876	1.00	41.01	N

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)

立体構造決定法

立体構造決定法	エントリー数	割合
X線結晶構造解析	70714	87.6
核磁気共鳴(NMR)	9362	11.6
電子顕微鏡	422	0.5
その他	212	0.3
合計	80710	100.0

- 全エントリー中約88%がX線結晶構造解析法により、立体構造が決定されている。
- 残りのほとんどは核磁気共鳴法
- X線結晶構造解析法については、次回解説

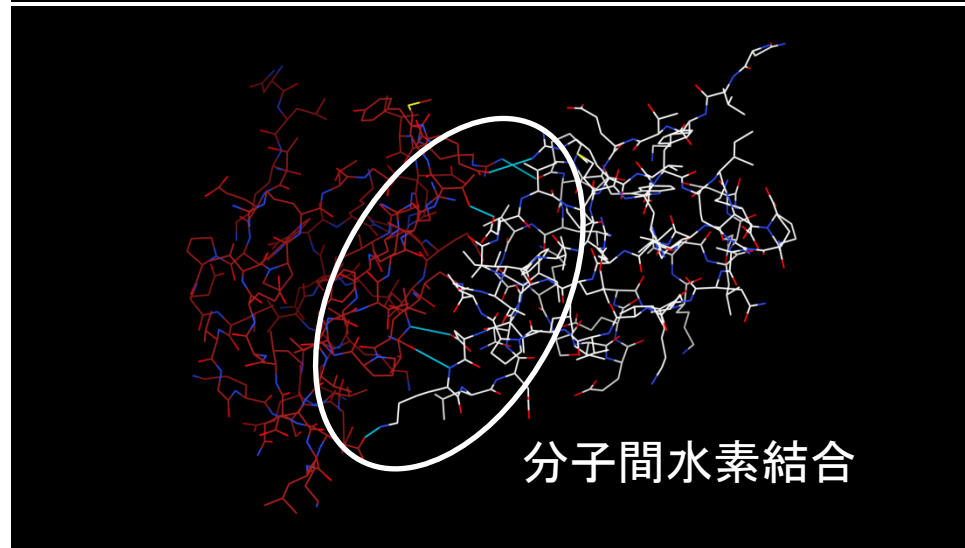
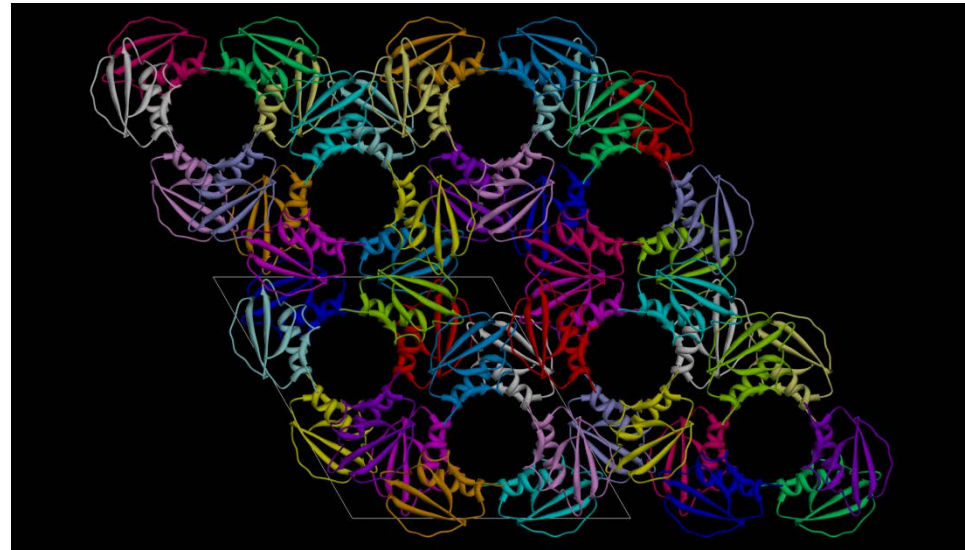
座標データに表れる違い

	X線結晶構造解析法	核磁気共鳴法
サンプルの状態	結晶(分子間接触あり)	溶液
分子量の上限	なし	200残基程度まで
水素原子	座標データに含まれない	座標データに含まれる
欠失原子	あり	通常なし
モデル数	通常1つ(部分的に複数)	複数
精度の指標	分解能 ^注	モデル構造のばらつき
原子の分布	温度因子	モデル構造のばらつき

注: X線結晶構造解析法ではどれだけ回折像を用いたかによって決まる分解能 (resolution) が全体の精度の指標。2.0~2.5 Åが普通、1.5 Å以下だと高分解能。

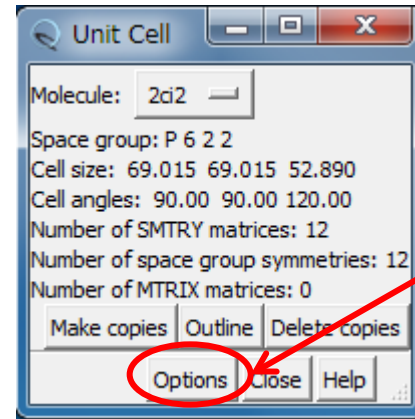
結晶構造の再現(1)

- 結晶中では、タンパク質分子が規則正しく並んでいる
- PDBに登録されている座標は、繰り返しの最小単位(非対称単位)
- 隣接したタンパク質分子間で相互作用していることがわかる

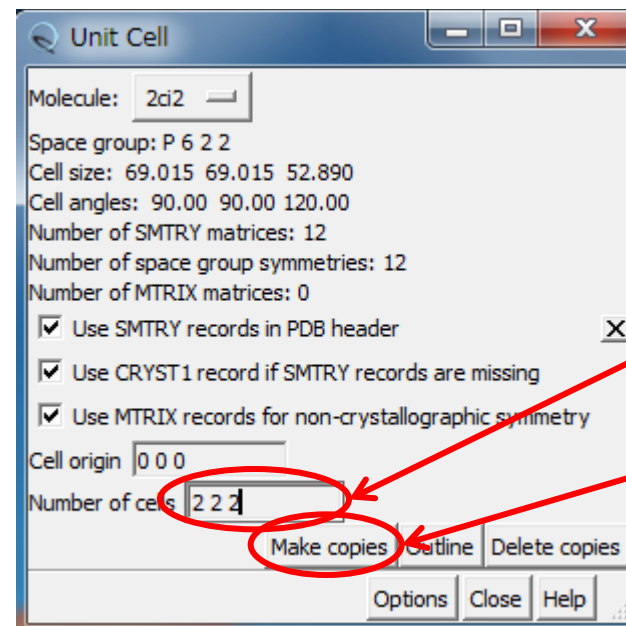


結晶構造の再現(2)

- メニューの「File」→「Fetch by ID」を選択し、PDB IDに2CI2を指定して「Fetch」
- 結晶構造の再現には、「Tools」→「High-Order Structure」→「Unit Cell」で右図のように指定する



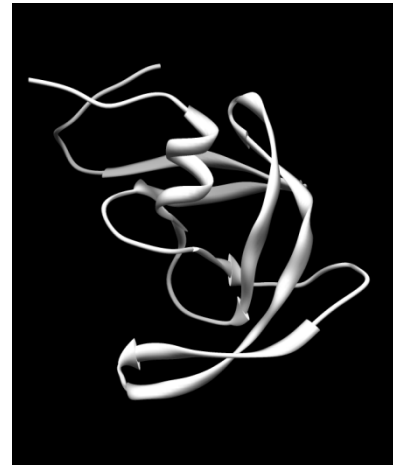
ここをクリックし
以下の表示に
する



変更
ここを
クリック

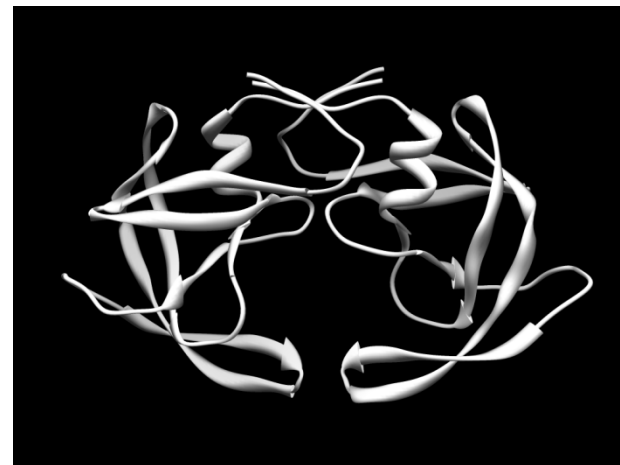
Biological unit

- 生物学的に機能する最小限の分子構成を biological unit と呼ぶ
- 分子の対称性が、結晶の対称性と偶然一致すると、非対称単位には多量体の一部しか含まれない場合がある
- RCSBのサイトでは biological unit の座標がダウンロードできる



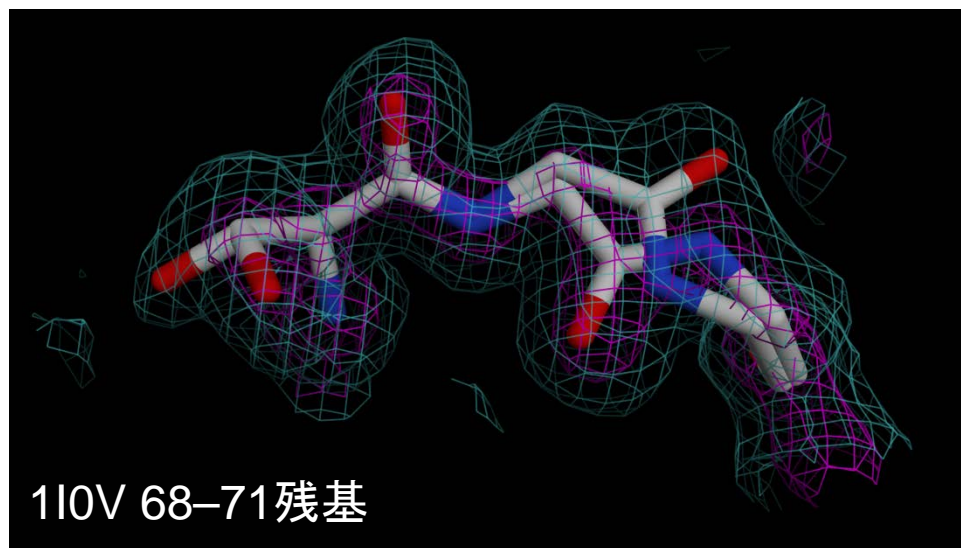
3PHVに登録されている座標

Biological unitの座標



Alternative conformation

- 結晶には異なるコンフォメーションを持つ複数の構造が含まれる可能性がある
- このような場合、X線回折データから得られる電子密度図では、それらの構造が存在割合に応じて複数見えることになる
- PDBファイルでは、occupancyに1より小さい重みを与え、同じ名前の原子を複数の座標で表す
- この時、原子名の残基名の間(17文字目)に、コンフォメーションを区別するIDを記入する



ATOM	548	N	AGLY	A	70	8.699	28.734	14.638	0.75	16.18
ATOM	549	N	BGLY	A	70	8.755	28.829	14.563	0.25	15.67
ATOM	550	CA	AGLY	A	70	9.857	29.561	14.390	0.75	16.18
ATOM	551	CA	BGLY	A	70	9.772	29.792	14.136	0.25	18.27
ATOM	552	C	AGLY	A	70	10.666	29.621	15.667	0.75	11.67
ATOM	553	C	BGLY	A	70	11.119	30.042	14.811	0.25	15.91
ATOM	554	O	AGLY	A	70	10.224	29.152	16.720	0.75	15.70
ATOM	555	O	BGLY	A	70	12.083	30.400	14.131	0.25	22.24



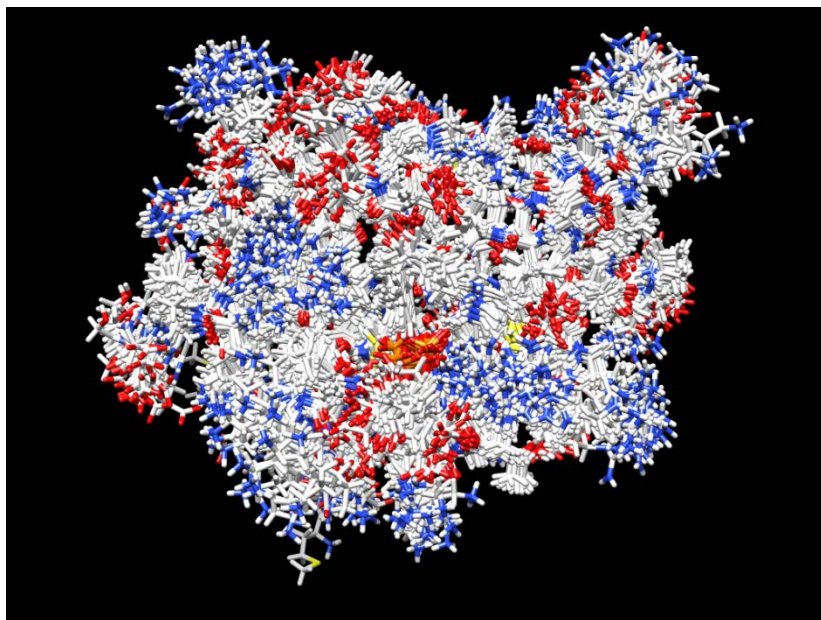
Alternate location
indicator



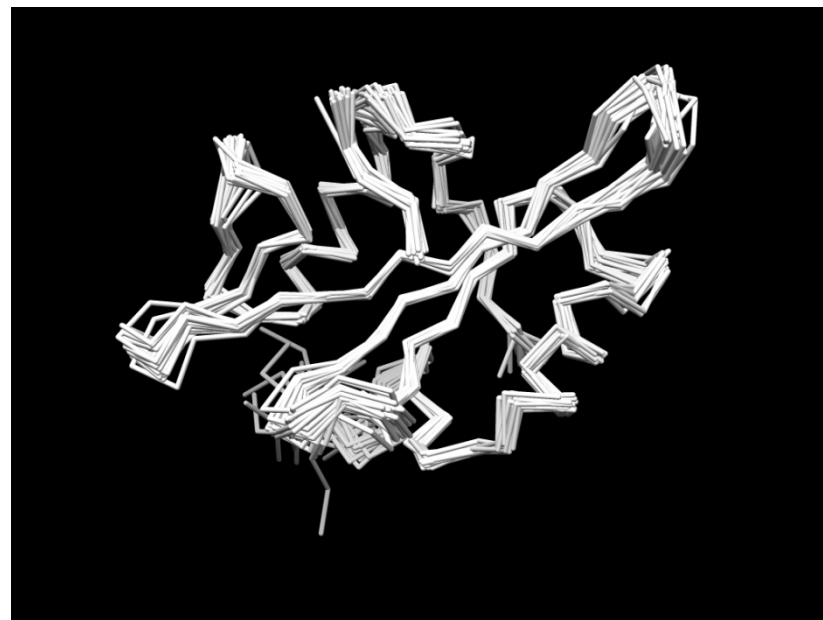
Occupancy
26

NMR構造

- ヒトSrc SH2ドメインと基質ペプチドの複合体の立体構造、1HCTを開く



「Actions」→「Atoms/Bonds」→「show」
「Actions」→「Ribbon」→「hide」



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

NMR構造の特徴

- 立体構造が複数のモデルの重ね合わせで表現される
- モデル構造のばらつきを精度の指標とする
 - モデル構造の平均構造からのばらつき
RMSD (root-mean-square deviation)

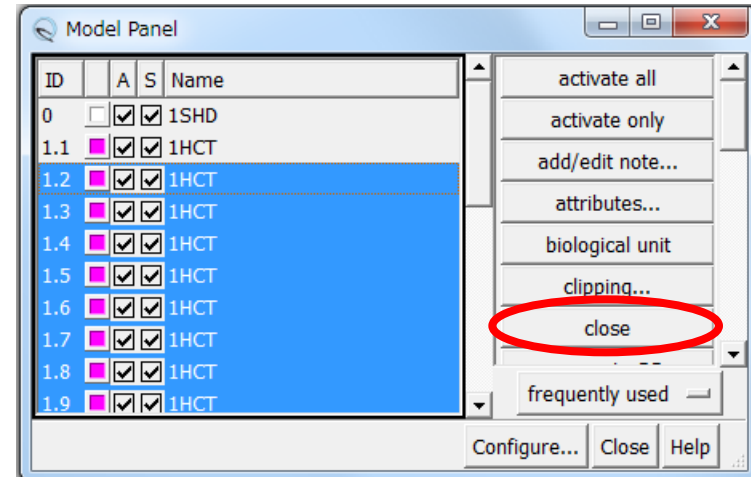
$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N |\mathbf{r}_i - \langle \mathbf{r}_i \rangle|^2}$$

\mathbf{r}_i : 原子*i*の座標

$\langle \mathbf{r}_i \rangle$: 平均構造の原子*i*の座標

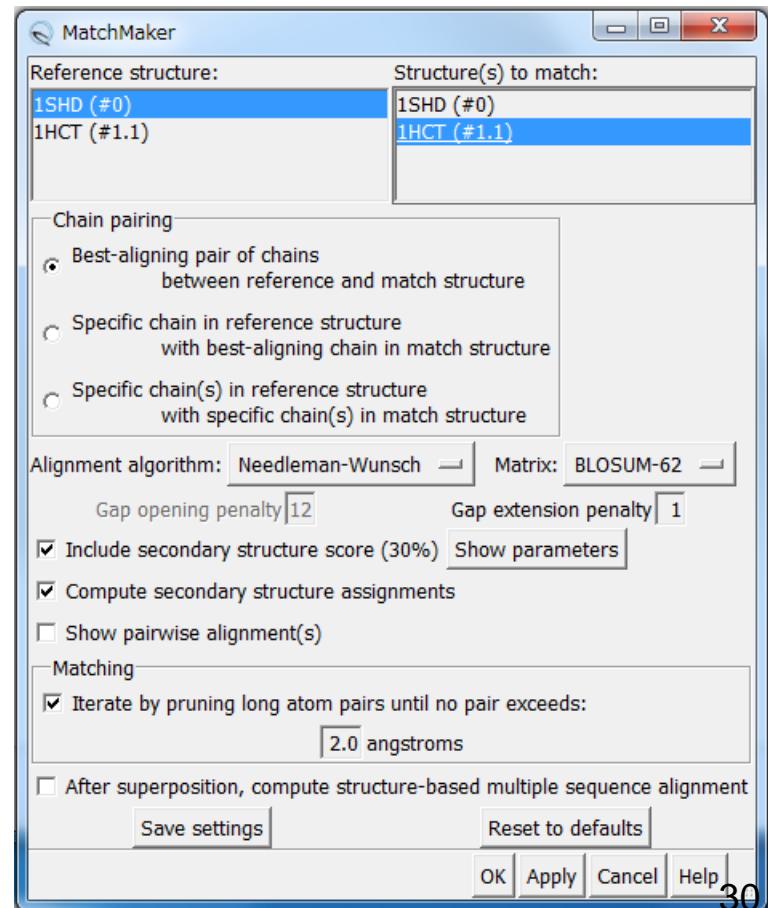
構造の比較(1)

- ヒトSrc SH2ドメインと基質ペプチドの複合体についてX線構造とNMR構造を比較する
1. 「File」→「Fetch by ID」で1SHDを開く
 2. 同様に1HCTを開く
 3. 「Favorite」→「Model Panel」を開く
 4. 右図のようにID 1.2を左クリックしたのち、「Shift」キーを押しながらID 1.23を左クリック
 5. 「close」をクリックしてこれらを閉じる

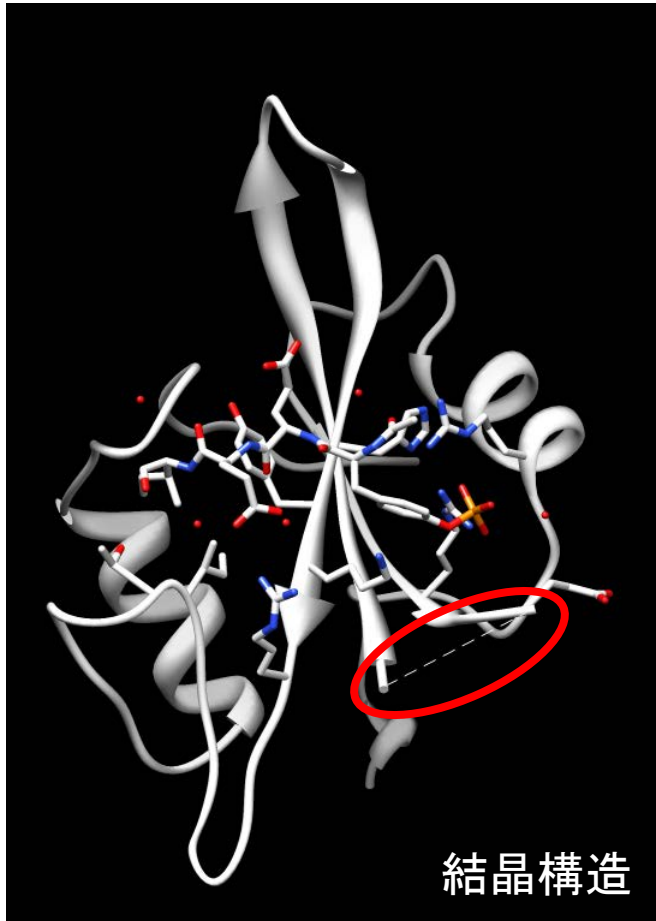


構造の比較(2)

6. 「Tools」→「Structure Comparison」→「MatchMaker」を選択
7. 右図のように設定し「OK」
8. ChimeraのWindowの下部に重ね合わせに使われた残基数(90残基)とRMSD(0.933 Å)が表示される

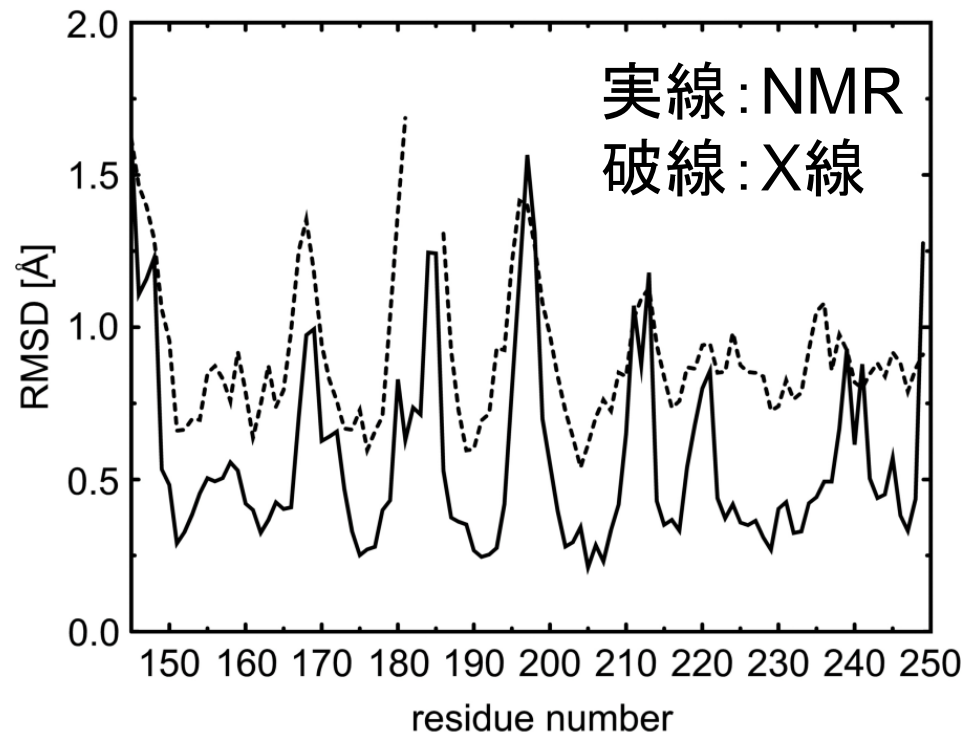


構造の比較(3)



構造の比較(4)

- NMR構造については、各モデルのC α 原子の平均構造からのずれの平均値(RMSD)
- X線構造では温度因子から換算
 - $B = 8/3\pi (\Delta r)^2$
 - $B = 30$ で $\Delta r = 1.07 \text{ \AA}$
- 温度因子が大きい残基は、NMRでも構造のばらつきが大きい傾向



参考: その他の立体構造データベース

- Nucleic Acid Database
 - <http://ndbserver.rutgers.edu/>
 - 核酸に特化した検索が可能
- PDBePISA
 - http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html
 - 予測多量体構造 (Biological Unit) のデータベース
- PiQSi
 - http://supfam.mrc-lmb.cam.ac.uk/elevy/piqsi/piqsi_home.cgi
 - 実験的に検証された多量体構造のデータベース
- Orientations of Proteins in Membranes (OPM) database
 - <http://opm.phar.umich.edu/>
 - 膜タンパク質の脂質2重膜に対する配向の予測
- ModBase
 - <http://modbase.compbio.ucsf.edu/>
 - 予測立体構造のデータベース

配列データベースとの連携

- 配列データベースへのリンク
 - RCSBの検索結果のSequenceタブ
- 配列データベースからのリンク
- 配列からの検索

配列データベースからのリンク

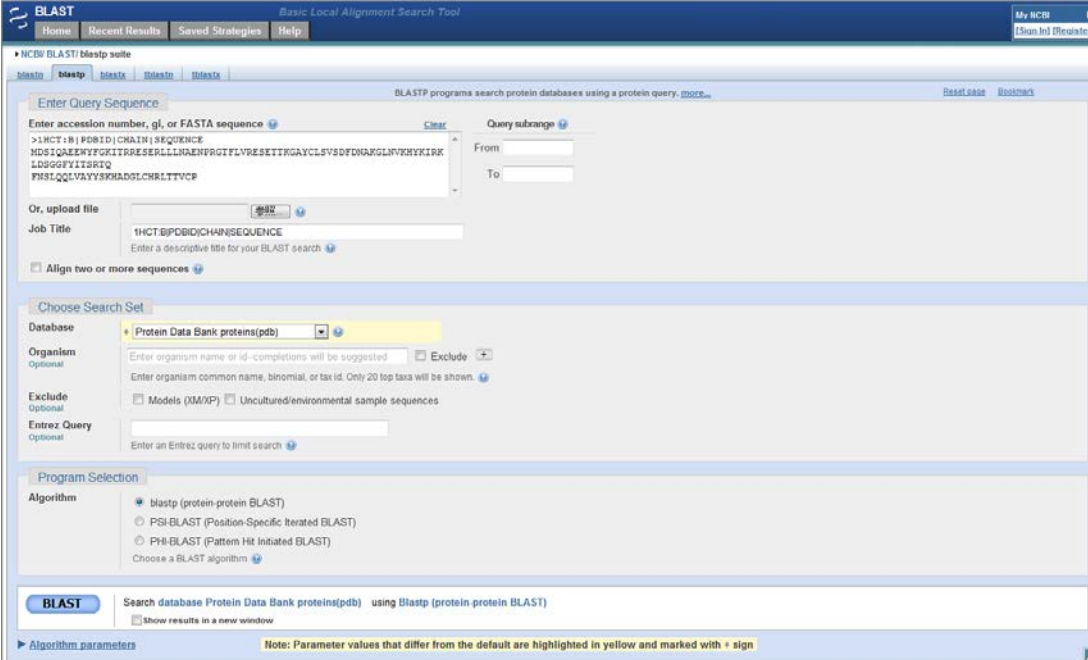
1. タンパク質配列データベースUniProt (<http://www.uniprot.org/>)を開く
2. QueryにSRC_HUMANと入力し「Search」
3. 検索結果の下のほうに、“3D structure databases”のセクションがあり、1HCTや1SHDが現れていることを確認すること

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. 講義のページで1HCT_B.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 'Choose Search Set' section is highlighted, and a red arrow points to the 'Database' dropdown menu, which is set to 'Protein Data Bank proteins (pdb)'. The 'Algorithm' section is set to 'blastp (protein-protein BLAST)'. The 'BLAST' button is visible at the bottom of the form.

BLAST Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help My NCBI [Sign In | Register]

NCBI BLAST/blastp suite

blastp blastx blastn blastz

Enter Query Sequence

BLASTP programs search protein databases using a protein query. [more...](#) [Reset base](#) [Bookmarks](#)

Enter accession number, gi, or FASTA sequence Clear Query subrange From To

>|HCT:|H|PDB:|D|CHAIN|SEQUENCE
HDS:|QAEWFGGITRRESERLLNENRGTFLVRESEITRGAYCLVSDFDWARGLNKHYKIKR
LDGGGFYITRFD
FHSIQQLVAYYSKHADGLCHRLLTVCP

Or, upload file

Job Title
1HCT:BPGRDICHANSEQUENCE
Enter a descriptive title for your BLAST search.

Align two or more sequences

Choose Search Set

Database

Organism
Optional
Enter organism name or id-completions will be suggested Exclude

Exclude
Optional
 Models (MM/XP) Uncultured/environmental sample sequences

Entrez Query
Optional
Enter an Entrez query to limit search

Program Selection

Algorithm

blastp (protein-protein BLAST)
 PSI-BLAST (Position-Specific Iterated BLAST)
 PHI-BLAST (Pattern Hit Initiated BLAST)
Choose a BLAST algorithm

Search database Protein Data Bank proteins(pdb) using Blastp (protein-protein BLAST)
 show results in a new window

[Algorithm parameters](#) Note: Parameter values that differ from the default are highlighted in yellow and marked with + sign

実習課題1

1. インフルエンザ治療薬oseltamivir(商品名: Tamiflu)と結合したneuraminidaseの立体構造を検索しPDBファイルをダウンロードせよ
2. ChimeraでPDBファイルを開き、タンパク質をRibbon表示した後、薬剤と薬剤から5 Å以内にあるタンパク質の残基をstick表示せよ
3. 結合部位の拡大図をpng形式で保存せよ

実習課題2

1. 講義のページで、kagai.fastaを表示し、この配列をもつタンパク質の立体構造データを検索せよ
2. ChimeraでPDB IDを指定して表示せよ
3. 全体像をpng形式で保存せよ

課題の提出

- 課題1の全体像と結合部位の拡大図をPowerPointのスライドに貼り付け、PDB IDと立体構造決定の方法を記入せよ
- 同じPowerPointファイルの別のスライドに課題2の全体像を貼り付け、PDB IDとタンパク質名、立体構造決定の方法を記入せよ
- PowerPointファイルはメールに添付して寺田宛 (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