

平成23年4月20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

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72,000
- 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 shows the RCSB PDB website interface. At the top left is the logo for RCSB PDB (Protein Data Bank). To the right, it states 'A MEMBER OF THE PDB' and 'An Information Portal to Biological Macromolecular Structures'. Below this, it indicates the date and time: 'As of Tuesday Apr 05, 2011 at 5 PM PDT there are 72244 Structures' and provides links for 'PDB Statistics'. A search bar is located below the header, 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 introductory text about the PDB archive and the tools available. On the left side, there are several navigation menus: 'MyPDB' (with links for account management), 'Home' (with links for news, policies, and contact), 'Deposition' (with links for deposit services), 'Search' (with links for advanced search and sequence search), and 'Tools' (with links for downloading entries and ligands). On the right side, there are sections for 'Customize This Page' (with a 'New Features' toggle), 'RCSB PDB News' (with a 'Weekly | Quarterly | Yearly' filter), and a featured article titled 'J.P. Stevens HS Wins NJ Protein Modeling Event'. The featured article includes a photo of three students and a brief description of their achievement. Below the main content, there are sections for 'Featured Molecules' and 'Structural View of Biology' with various molecular models and a 'Molecule of the Month' section for 'Nanobodies'.

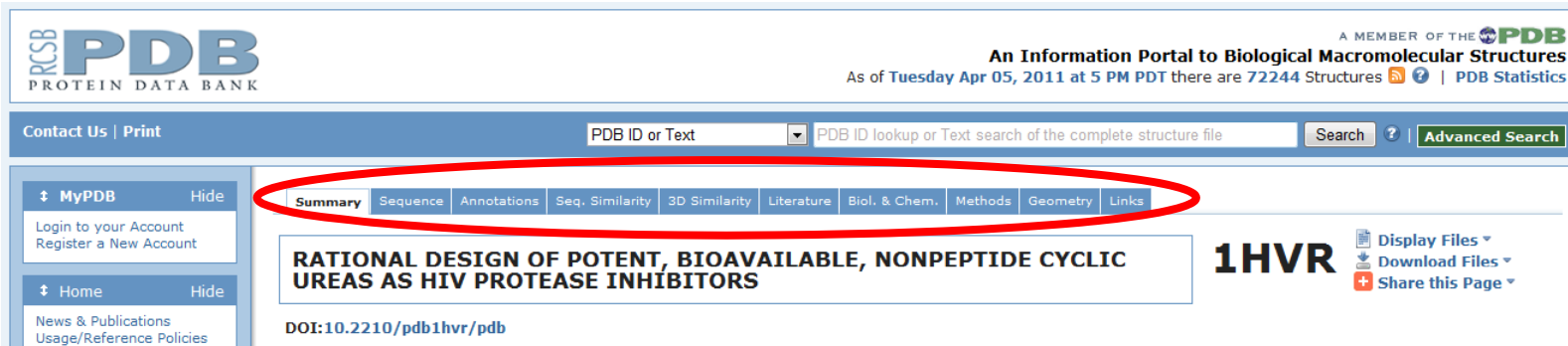
検索



- 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

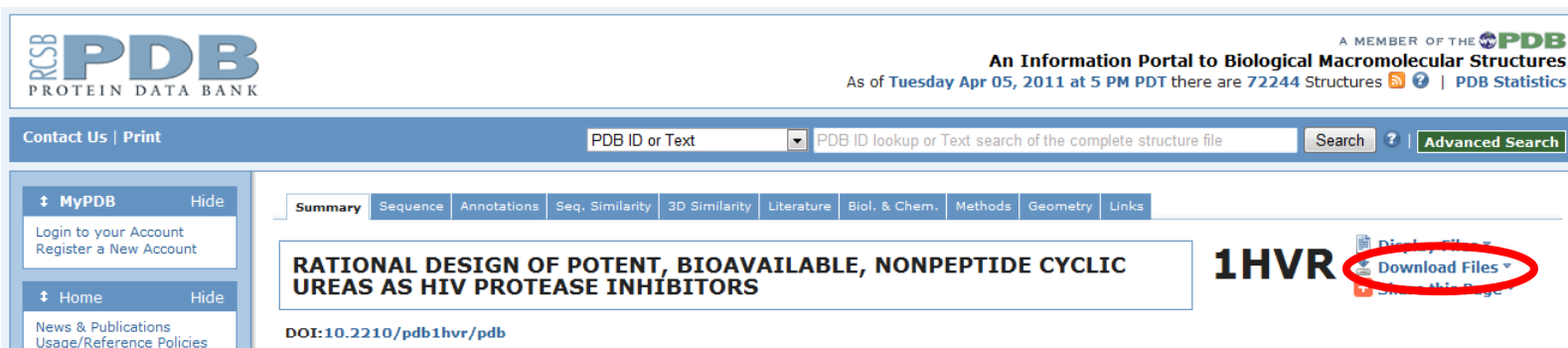


The screenshot shows the PDB website interface. At the top, there is a search bar with a dropdown menu set to 'PDB ID or Text' and a search button. Below the search bar, there is a navigation menu with tabs for 'Summary', 'Sequence', 'Annotations', 'Seq. Similarity', '3D Similarity', 'Literature', 'Biol. & 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 DOI '10.2210/pdb1hvr/pdb'. On the right side, there is a '1HVR' label and a 'Display Files' button.

- 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

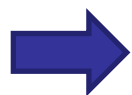
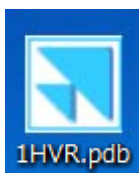


The screenshot shows the PDB website interface. At the top left is the PDB logo (RCSB Protein Data Bank). At the top right, it says 'A MEMBER OF THE PDB' and 'An Information Portal to Biological Macromolecular Structures'. Below this, it states 'As of Tuesday Apr 05, 2011 at 5 PM PDT there are 72244 Structures' and includes social media icons and a 'PDB Statistics' link. A search bar is present with a dropdown menu for 'PDB ID or Text' and a search button. On the left side, there are navigation links for 'MyPDB' (Login to your Account, Register a New Account) and 'Home' (News & Publications, Usage/Reference Policies). The main content area shows a title 'RATIONAL DESIGN OF POTENT, BIOAVAILABLE, NONPEPTIDE CYCLIC UREAS AS HIV PROTEASE INHIBITORS' and a DOI: 10.2210/pdb1hvr/pdb. To the right of the title, there is a '1HVR' label and a 'Download Files' button, which is circled in red. Below the 'Download Files' button, there are options for 'Download Files' and 'Download Images'.

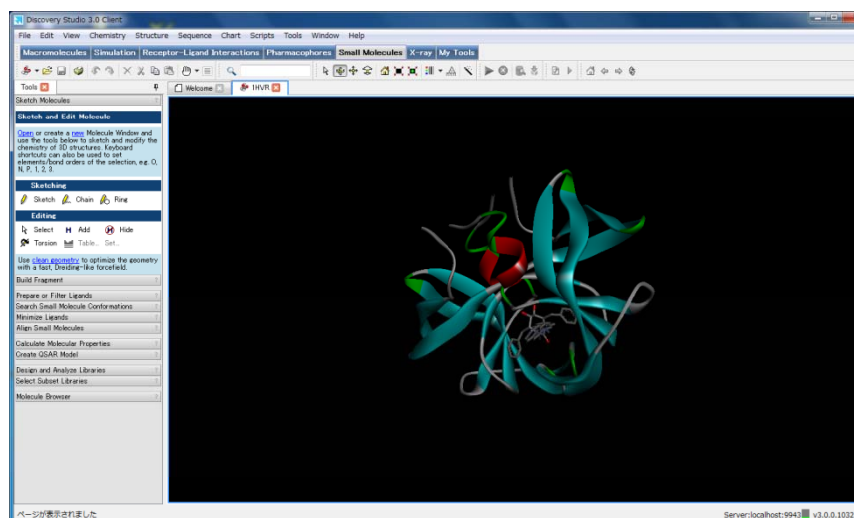
- 「PDB File (Text)」を選び、デスクトップの保存する

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. ファイルのアイコンをダブルクリック
4. Discovery Studio 3.0 Clientが起動






ダブルクリック




Molecule Windowの操作(1)

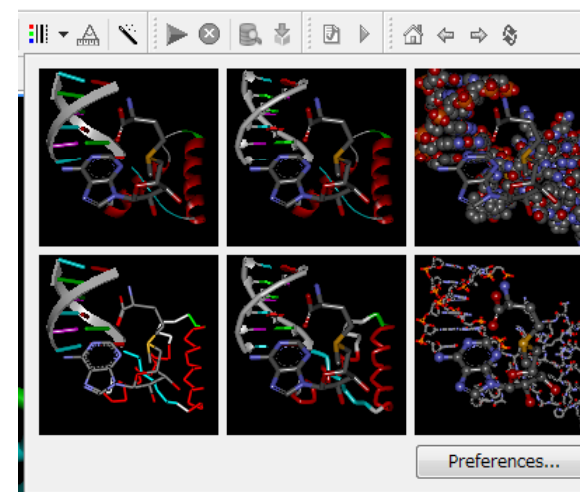
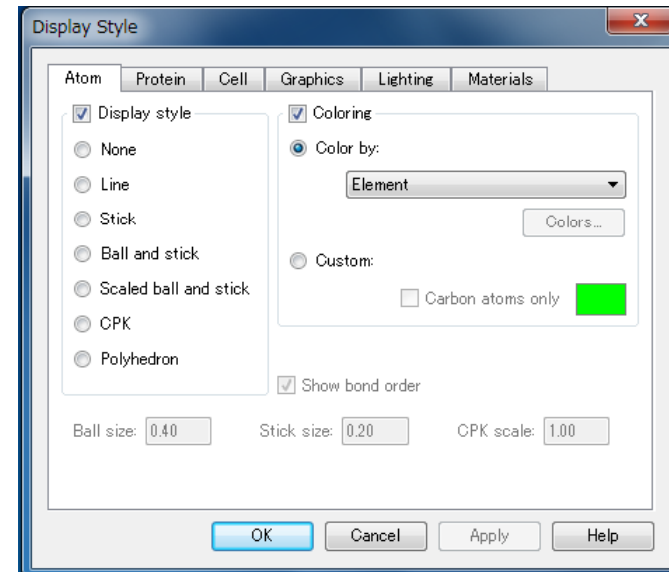
- Molecule Windowの中を左クリックしてアクティブにしてから以下の操作を行う
- 回転
 -  をクリックし、Molecule Windowの中で左ドラッグ
- 並進
 -  をクリックし、Molecule Windowの中で左ドラッグ
- ズーム
 -  をクリックし、Molecule Windowの中で左ドラッグ

Molecule Windowの操作(2)

- 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

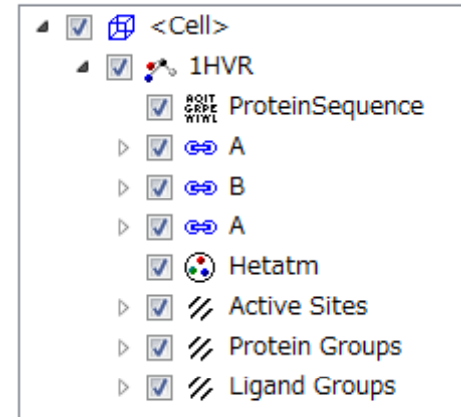
Display Styleの変更

- Display Styleアイコンをクリックすると、表示を変更するためのウィンドウが開く
- アイコンの横の▼をクリックすると、規定のスタイルが表示され、これを選択することでも変更できる



Hierarchyの表示

- メニューの「View」→「Hierarchy」を選択することでHierarchy表示・非表示を切り替えることができる
- ▷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

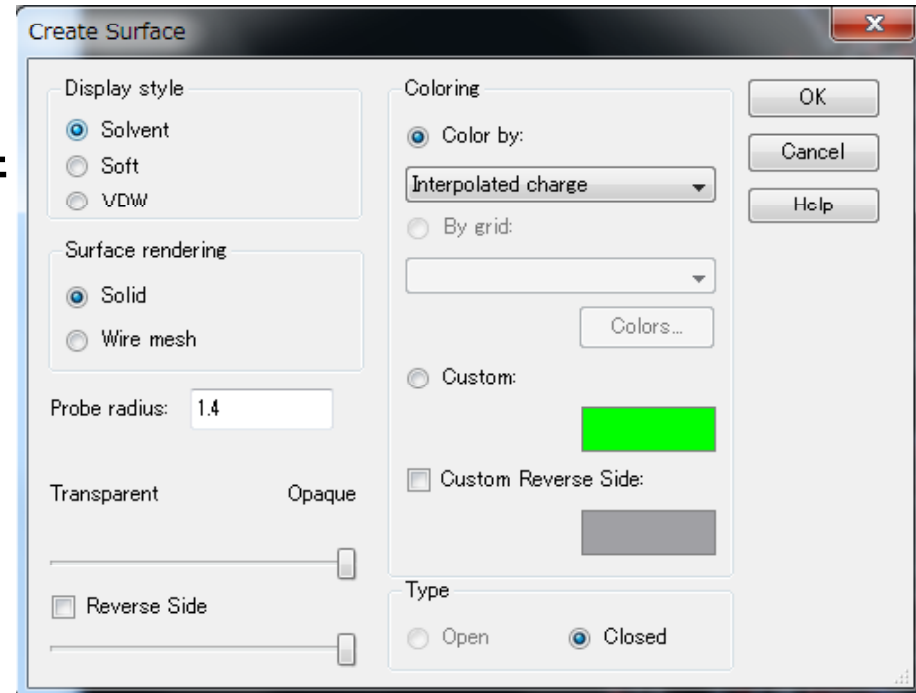


原子・残基・分子の選択

- Molecule Window上で原子をクリック→その原子が選択され、黄色でマークされる
- Molecule 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
- 何もないところをクリックすると選択を解除できる
- HierarchyではCtrlキーを、Molecule 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
- 選択した後右クリックで属性 (attribute) を表示できる

溶媒接触表面の表示

- 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
- 右のように設定し「OK」

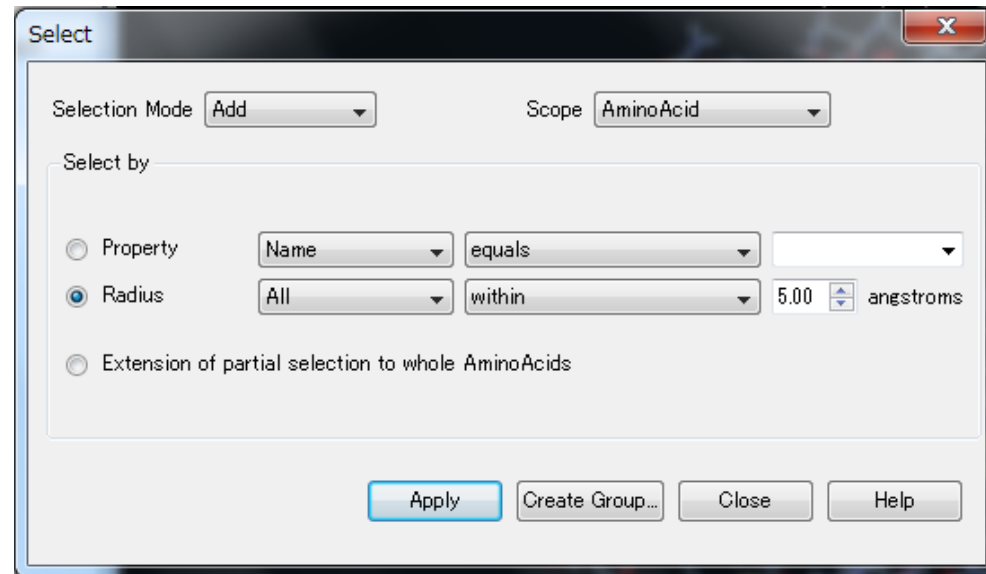


相互作用の検出(1)

- 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
- Molecule 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

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)

その他の可視化ソフトウェア

- 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>)

立体構造決定法

立体構造決定法	エントリ数	割合
X線結晶構造解析	62865	87.0
核磁気共鳴(NMR)	8828	12.2
電子顕微鏡	364	0.5
その他	187	0.3
合計	72244	100.0

- 全エントリ中87%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

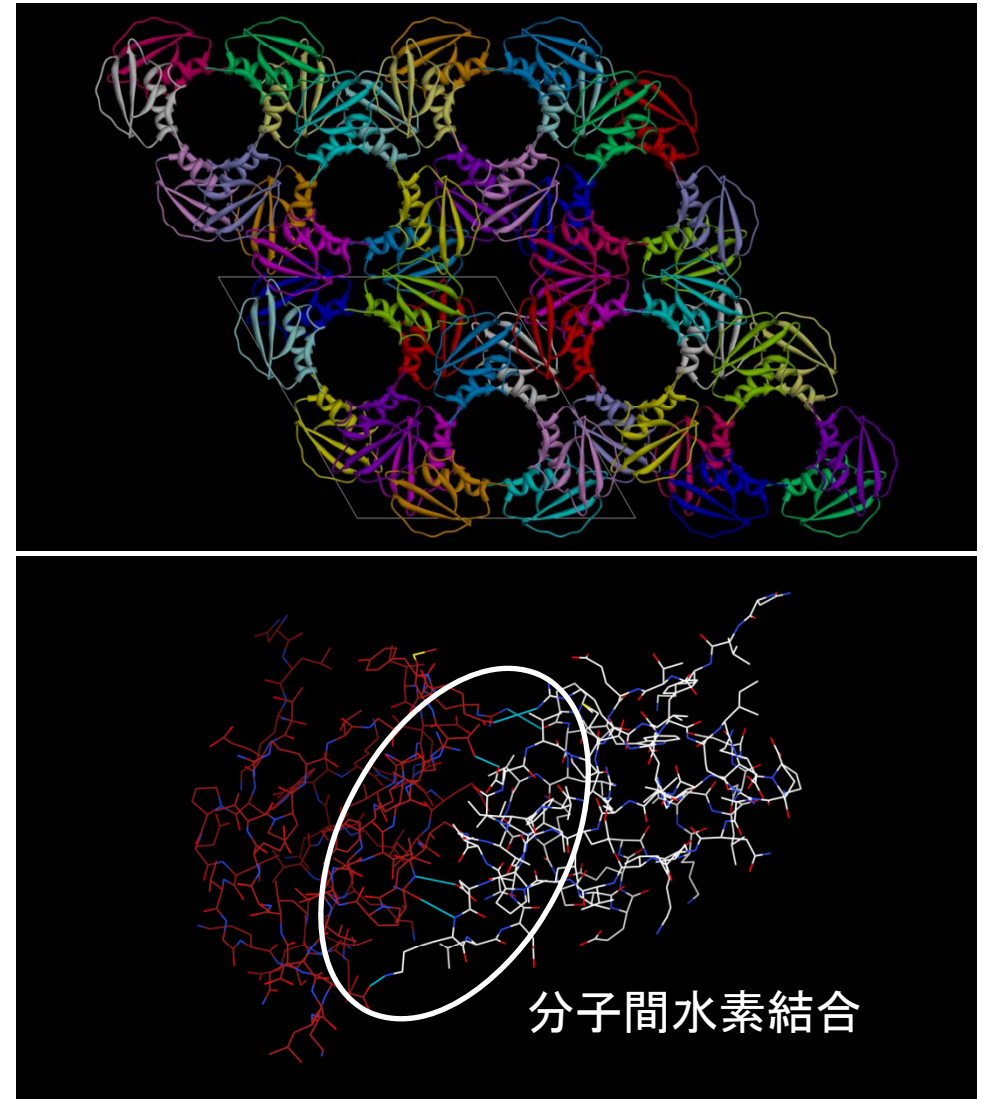
座標データに表れる違い

	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。

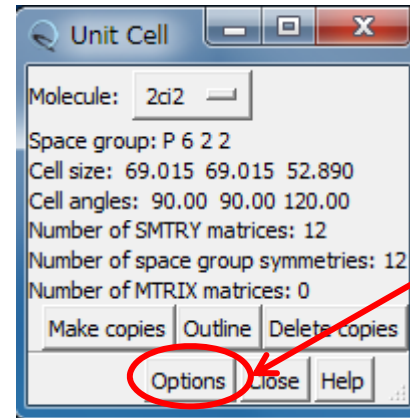
結晶構造の再現

- 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)
- Chimeraを用いると結晶構造を再現できる(右図は2CI2の例)

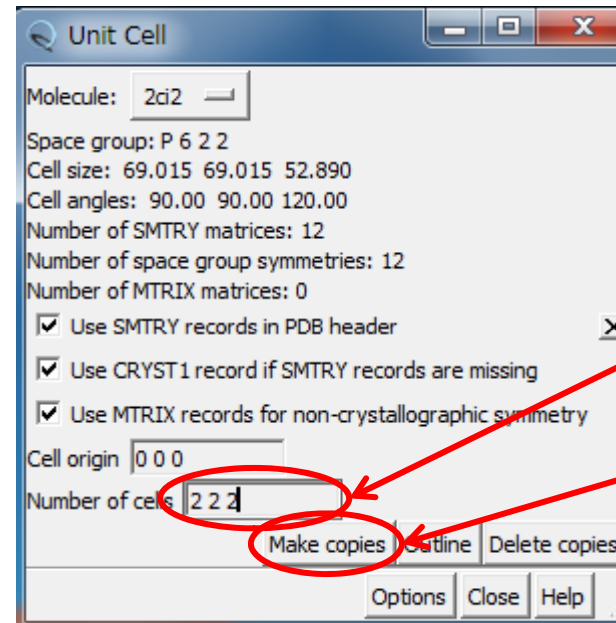


Chimeraの利用

- Chimera 1.5.2のアイコンをダブルクリック
- メニューの「File」→「Open...」からファイルを開くことができる
- メニューの「File」→「Fetch by ID...」でPDB IDを指定して開くこともできる
- 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



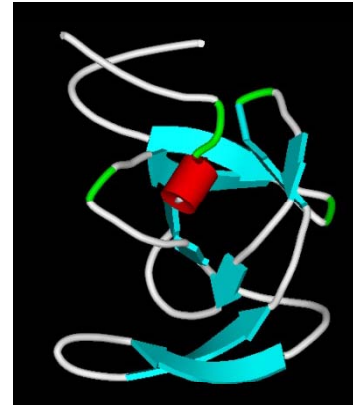
ここをクリックし
以下の表示に
する



変更
ここを
クリック

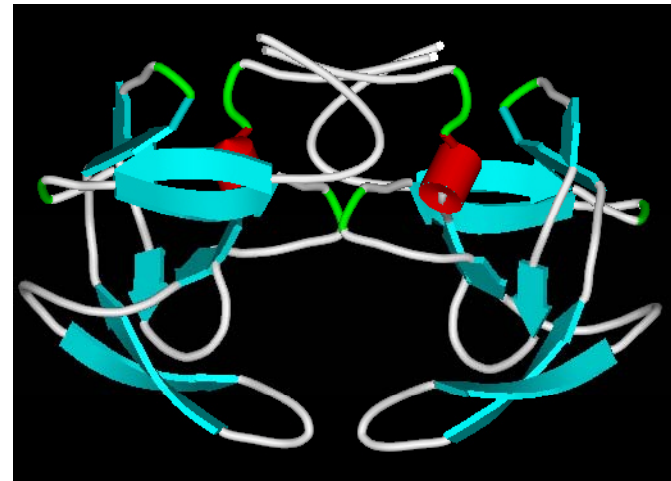
Biological unit

- 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



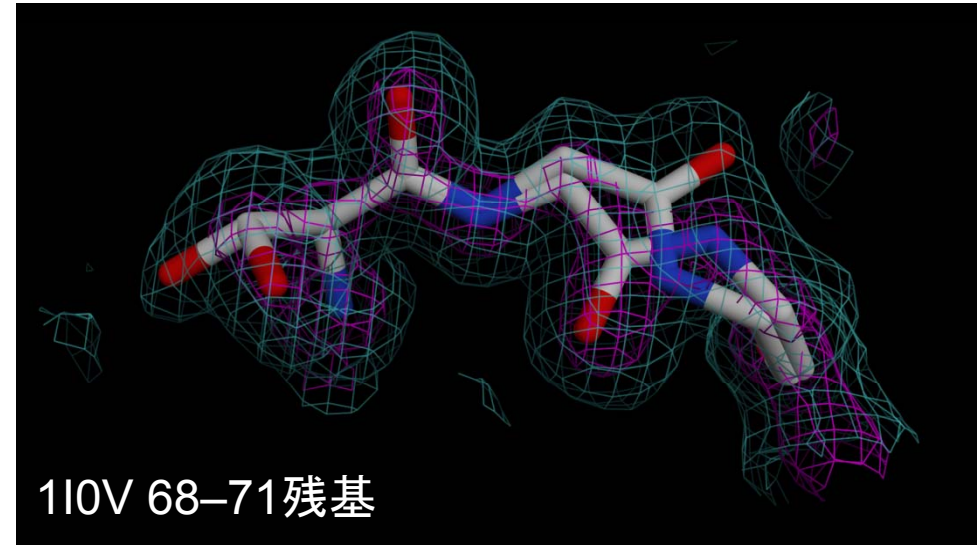
3PHVに登録されている座標

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



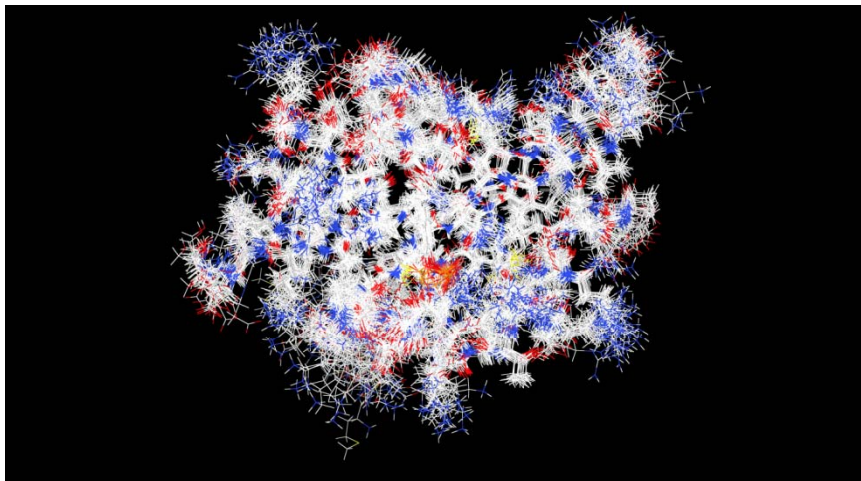
Alternate location
indicator



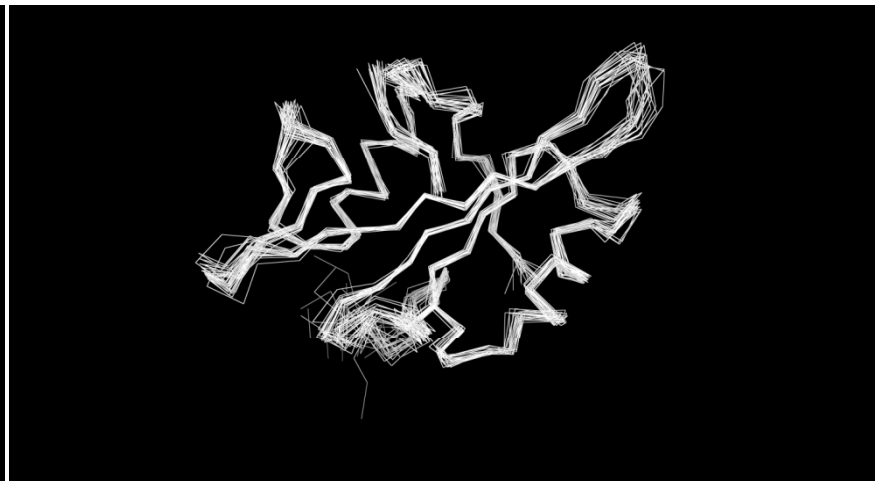
Occupancy

NMR構造

- Chimeraを使用してヒト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



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



「Action」→「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

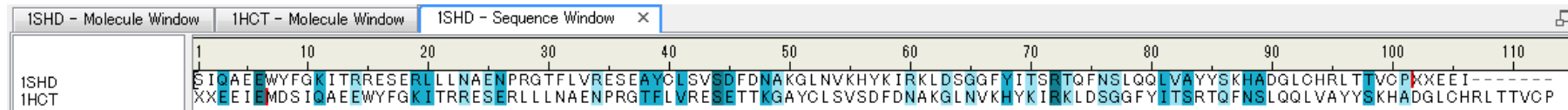
$\langle \mathbf{r}_i \rangle$: 平均構造の原子*i*の座標

構造の比較(1)

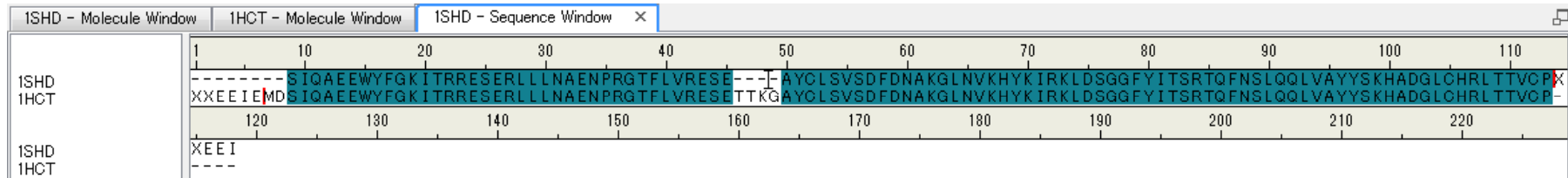
- ヒトSrc SH2ドメインと基質ペプチドの複合体についてX線構造とNMR構造を比較する
 1. Discovery Studio 3.0 Clientを起動し、「File」→「Open URL」を選択
 2. IDに1SHDと入力して「Open」
 3. 「File」→「Open URL」を選択
 4. IDに1HCTと入力して「Open」
 5. Hierarchyで1HCT_model_1をクリックして選択したのち、右クリックして「Copy」
 6. 1SHDのMolecule Windowで右クリックして「Paste」

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

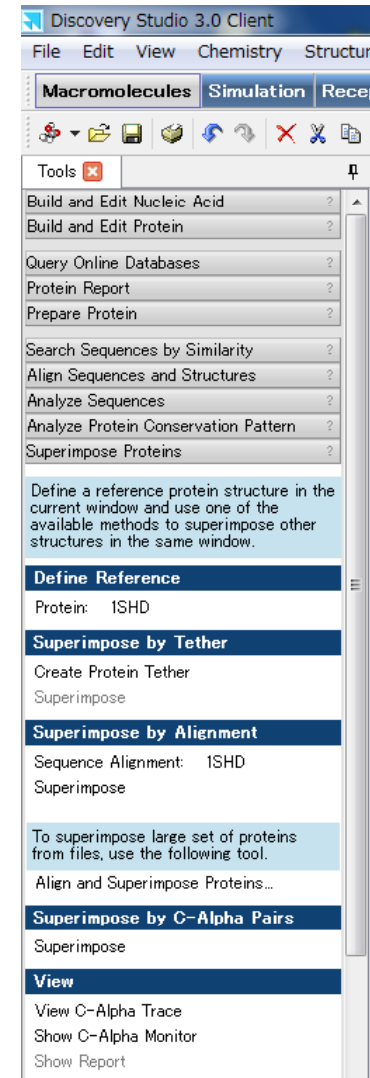


↑
1SHDの先頭に8つギャップを挿入

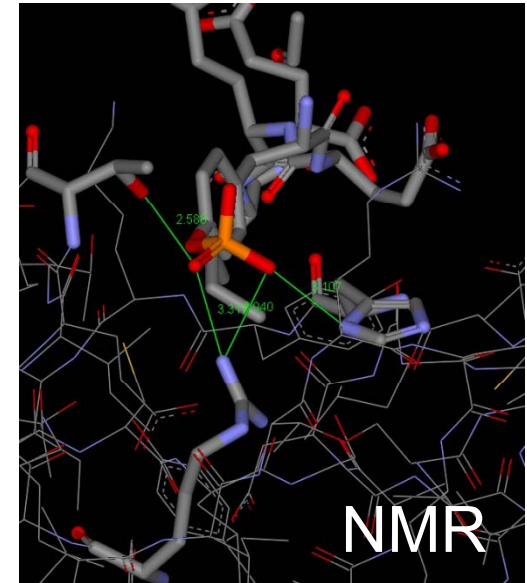
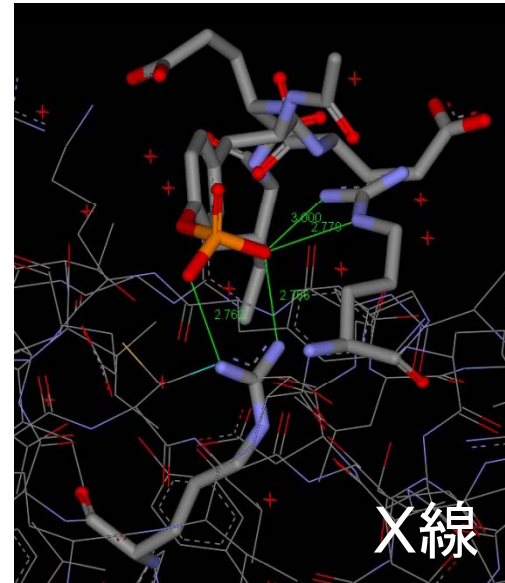
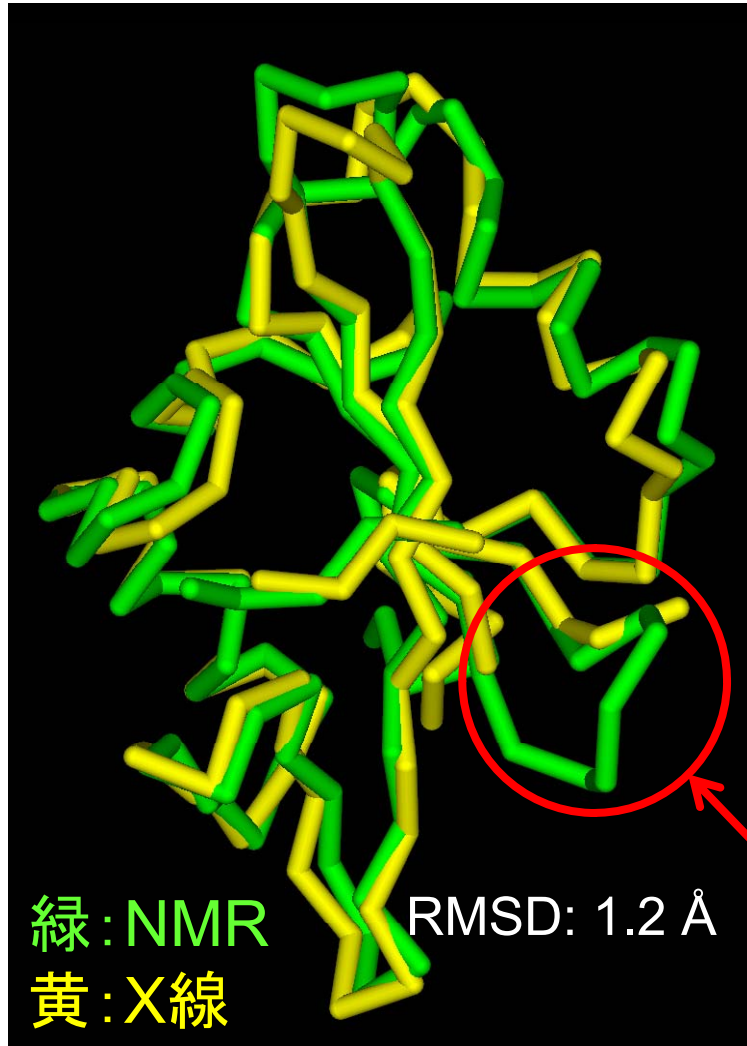
↑
1SHDのGLU181の後に4つギャップを挿入

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Molecule Windowをアクティブにし、
メニューの下にある「Macromolecules」ボタン
をクリック→Toolsタブの表示が変わる
11. Toolsタブにある「Superimpose Proteins」を
クリックし、右のように表示されることを確認
12. Superimpose by Alignmentの項目にある、
「Superimpose」をクリック
13. Molecule Windowでは1HCTの構造が1SHD
に重ねあわされている
14. Viewの項目にあるShow Reportをクリックす
るとRMSDが表示される



構造の比較(4)

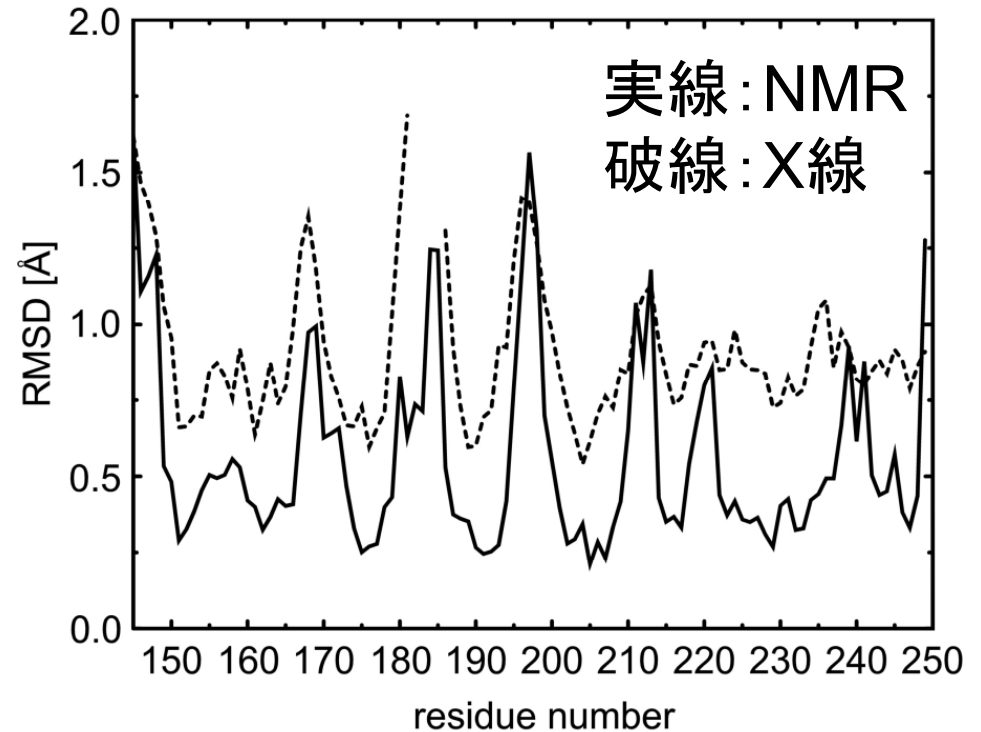


主鎖構造はよく似ているが、リン酸基と相互作用している残基が異なる

X線構造では、182~185残基の構造がないことに注意

構造の比較(5)

- 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



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.pdbe.org/Pisa>
 - 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. 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. 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 'Enter Query Sequence' section contains a FASTA sequence: >lhct:B|PDBID|CHAIN|SEQUENCE HSD1QAEWIFGSLTRRESRELLLNENFRDTFLVRESETTGGAYCLSVSDFDWAQGLNWKYKIRK LDDGQFYVTSRTQ FNSLQQLVAYYSKIDAGLCHRITVCP. The 'Job Title' is 'lhct:B|PDBID|CHAIN|SEQUENCE'. In the 'Choose Search Set' section, the 'Database' dropdown is set to 'Protein Data Bank proteins (pdb)'. The 'Program Selection' section has 'blastp (protein-protein BLAST)' selected. At the bottom, the 'BLAST' button is highlighted, and the search parameters are displayed: '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'.

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. Discovery Studio 3.0 ClientでPDBファイルを開き、同等なポリペプチドが複数含まれている場合は、1つを残してそれ以外を削除せよ
3. タンパク質をSolid 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
4. 全体像と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
(Chimeraでは「File」→「Save Image...」で画像を保存できる)

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