

3-1-2 タンパク質の3次元構造を観る

概要

DNA やタンパク質など生体高分子の立体構造情報は、近年飛躍的に蓄積されてきています。ここでは、これらの分子構造について、無料ソフトを用いて表示・操作などをする際のガイドを、簡単なアニメーション作成等いくつかのテーマに絞ってまとめました。視覚的にそれらの知的資源を最大限に活用したり、研究の様々な場面でのプレゼンテーションに用いたりするために、少しでも参考になればと思います。

>1行アケル<

1. はじめに、Protein Data Bank (以下 PDB, <http://www.rcsb.org>)

生体高分子 (タンパク質, DNA, RNA, それらの複合体など) の3次元構造座標データは、このサイトで入手することができます。タンパク質というほど長くないインスリンなど、アミノ酸が数十残基のペプチドも含まれています。以前は理論計算モデルの登録も受け付けていましたが、現在は X 線・電子線結晶解析や核磁気共鳴 (以下 NMR) などの実験から得られたものに限定されています。英語のサイトはまだちょっと・・・, という方には、日本で管理されている PDBj (http://www.pdbj.org/index_j.html) からアクセスできます。

ダウンロードするファイルはいろいろな形式から選べますが、通常は PDB テキスト形式を選択します。すると、“英数字4文字.pdb” というファイル (以下 PDB ファイル) が保存でき、これを自分の PC 上で様々なソフトを使って利用するわけです。この英数字4文字コードは、PDB ID と呼ばれるもので、生体高分子の立体構造解析の論文にはこれを記載することが義務付けられています。一方 PDBj からダウンロードすると、同じ PDB ID のファイルが、pdb+PDB ID.ent という名前になっていますが、中身は PDB ID.pdb と同じです。

おそらく最も厄介な問題は、例えば単純に myoglobin (ミオグロビン) という名前で簡易検索 (PDB ID or keyword) すると 300 個ほどヒットするので、どれを見ればいいのかわからない、といったことかもしれません。これらの違いには、変異導入の有無・種類、タンパク質の環境、結合している成分など、様々なものがあります。欲しい構造の PDB ID が分からない場合は、とりあえず名前の簡易検索結果を構造解析の分解能で並べ替えて上から (高分解能側から) 見ていくのも一つの方法です (分解能での並べ替えは PDBj にはないようです)。単に構造をちょっと眺めたいだけであれば、特にファイルを手元に保存しなくとも、PDB や PDBj のサイト上で3次元表示させることもできます。

2. 使用するソフトとハード

2. 1 ソフト

PDB ファイルを読み込むことのできるフリーソフトには様々あり、PDB サイトの Molecular

執筆者 生命科学科 教授 岡田哲二 (tetsuji.okada@gakushuin.ac.jp)

生命科学科 助教 川元・尾崎洋子 (yoko.ozaki@gakushuin.ac.jp)

物理学科 学生 後藤孟平 (06041017@gakushuin.ac.jp)

Graphics Software Links には 30 以上がリストアップされています。ここでは構造解析を専門とされない方々向けに 2 つを取り上げます。

2. 1. 1 Swiss-PdbViewer 4.0.1 (以下 SPDBV, <http://spdbv.vital-it.ch/disclaim.html>)

Windows, Mac とも対応しています。上のアドレスにアクセスして免責事項に同意すると直接ダウンロードページに進みます。圧縮ファイル (.zip) を保存・解凍してできるホルダー内に実行ファイル (Windows では spdbv.exe) が作成されますので、特にインストールは必要ありません。

2. 1. 2 Discovery Studio Visualizer 2.5 (以下 DSV, <http://accelrys.com/products/discovery-studio/visualization/index.html>)

Mac には対応していません。上のアドレスにアクセスし、連絡先等を入力すると、メールでダウンロード情報が送られてきますので、プログラム本体 (約 120 MB) と ActiveX コントロール (約 25 MB) のセットアップファイルを使ってそれぞれインストールします。

(注) DSV は製品版の一部でもありますが、その購入を推奨するものではありません。SPDBV は、東海大学の岡本明弘先生により日本語のマニュアルが作成されています (<http://www2.ncc.u-tokai.ac.jp/okamoto/info/index.html>, ここには RasMol という汎用ソフトのマニュアルも掲載されています)

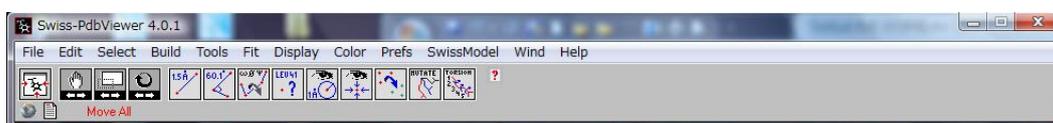


図 1. SPDBV ツールバー (Windows での表示)

2. 2 ハード

分子量の大きなタンパク質などの構造を 3 次元グラフィックスで操作する場合、PC の性能によっては動きが遅いなどの不都合が生じるかもしれません。参考までに、このガイドを作成する際に構造表示やアニメーション動作の確認などを行った PC の簡単なスペックを記載しておきます。

- (1) OS : Windows Vista Home Premium 32bit, プロセッサ : Intel Core 2 Duo Mobile Processor P8600 (2.4 GHz), メモリ : 3 GB, グラフィックス : NVIDIA GeForce 9300M GS
- (2) OS : Windows Vista Enterprise 32bit, プロセッサ : AMD Turion X2 Dual-Core Mobile Processor RM-70 (2.0 GHz), メモリ : 2 GB, グラフィックス : ATI Radeon HD3200

参考文献

[1] Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 18, 2714-2723.

3. PDB ファイルの基本

3. 1 ファイルの構成

PDB ファイルの中身はワードパッドなどテキスト表示・編集ソフトで見ることができます。あるいは、SPDBV を起動してツールバー (図 1) のメニューで File→Open PDB File としてファイルを読み込み、ツールバー左下の  マークをクリックすればテキスト表示されます。1 行目 (HEADER) から様々な付随情報が記されて (これらの座標情報前の行の全てをまとめて HEADER 部分ともいいます), 次に ATOM 行として実際の座標データが原子の数だけ続きます。1 つのポリペプチド鎖, ポリヌクレオチド鎖は, 末端の原子座標の次にある TER 行で区切られています。

各 ATOM 行は, 左から原子の通し番号, 原子タイプの分類, アミノ酸 (またはヌクレオチドなど) の種類, の次に ChainID が示されていて, インスリンのように 2 本のポリペプチド鎖でできている場合には, ここに A か B と示されています。2 本鎖 DNA の場合も同様です。続いてアミノ酸 (またはヌクレオチドなどの) 番号, 3 つの座標値, 占有率, 温度因子, と並んでいます。図 2 の例は, 1BRX.pdb という PDB ファイルの座標データ開始部分ですが, アミノ酸配列の 1 番目から 5 番目までの構造が欠けていて, 6 番目のアミノ酸であるグリシンから始まっていることがわかります。

ATOM	1	N	GLY	A	6	23.790	26.727	-16.219	1.00	37.42	N
ATOM	2	CA	GLY	A	6	23.038	27.522	-17.231	1.00	38.57	C
ATOM	3	C	GLY	A	6	23.069	28.991	-16.850	1.00	39.62	C
ATOM	4	O	GLY	A	6	23.376	29.871	-17.667	1.00	40.56	O
ATOM	5	N	ARG	A	7	22.681	29.259	-15.605	1.00	39.31	N

図 2. PDB ファイル中の座標データ

因みに ChainID は糖鎖などにも付けられますが, その場合には ATOM 行が全て終わったあとに HETATM 行としてまとめて書かれています。この HETATM 行は, アミノ酸・ヌクレオチド以外の成分全て (結合低分子, イオンなど) に用いられます。HETATM の内容がどのような成分か, ということは HEADER 部分に記載されています。また PDB サイトの各構造ファイルのメインページの下の方にも Ligand Chemical Component としてリストアップされているので確認できます。但し, 水分子も HOH という名前でも HETATM に含まれますが, リストには記載されません。

NMR による構造の場合は複数モデルのアンサンブルとしてまとめられているので, ATOM 行の前に MODEL 行があつて番号が付けられており, 最後の原子座標の次の ENDMDL 行までが 1 つのモデルです。

3. 2 Biological Assembly (または Biological Unit)

結晶解析による PDB ファイルの場合には，“非対称単位 (Asymmetric Unit)” という部分が登録されるため、実際に機能する会合状態 (Biological Assembly, 以下 BA) との対応関係に注意が必要です. PDBj には見当たりませんが, PDB の方には, ダウンロード項目の中に, BA として作成されたものが用意されています. 但し, 圧縮ファイル (.gz) 形式のみですので解凍ツール・操作が必要です. BA に 2 種類選択肢が用意されている場合は, 非対称単位が BA を 2 つ含んでいることを意味します. この 2 つのモデルは必ずしも同一ではありませんので, できれば重ね合わせるなどの構造比較をしてから一方を選択した方がよいでしょう. (注) PDB サイトで BA とされているものが必ずしも正しくない場合もあります.

4. 構造表示と比較 (インスリンの場合)

まず一例として, インスリンの X 線解析モデル (PDB ID: 3e7z) を取り上げます. この PDB テキストファイル 3E7Z.pdb にはインスリンの BA (ポリペプチド鎖が 2 本) が 2 つ含まれています. 従って, BA の圧縮ファイルも 3E7Z.pdb1.gz と 3E7Z.pdb2.gz の 2 つあり, 各々を解凍したホルダー内から 3e7z.pdb1 と 3e7z.pdb2 が得られます. これらの名前を 3e7z1.pdb と 3e7z2.pdb などと付け替えて, 拡張子を .pdb にします.

ここで SPDBV を立ち上げて, まずメニューの Prefs→Loading Proteins に進み, 読み込み設定をします.

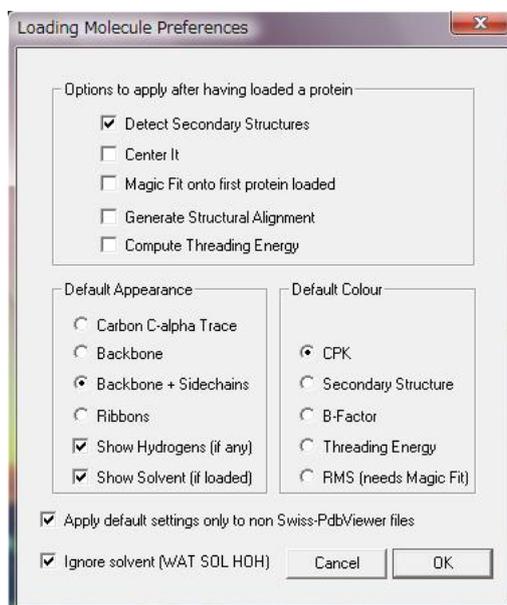


図 3. SPDBV の構造読み込み設定

図 3 のウィンドウで, Magic Fit onto…をチェックすると, 連続して読み込んだ 2 つの構造の重ね合わせを自動的に行ってくれます. Center It も有効にしておけば, 2 つの重ね合わせ

たモデルが画面の中心に表示されます。

また、Default Appearance や Colour を変えておけば、構造が見やすくなります。ここでは、Carbon C-alpha Trace と B-Factor を選んでおきます。これらは、ポリペプチド主鎖を α 炭素原子の繋がりのみとして表示し、各原子を温度因子で色分けするという意味です。

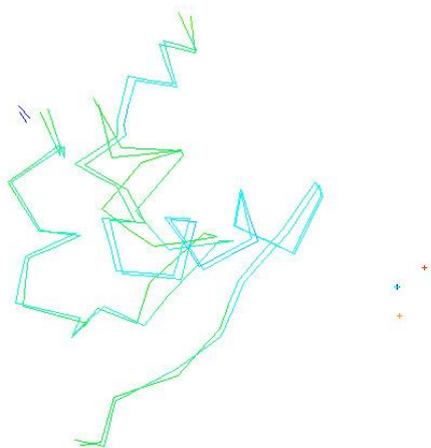


図4. 2つの構造の自動重ね合わせ結果

次に File→Open PDB File で 3e7z1.pdb と 3e7z2.pdb を続けて開くと、図4のようになり、微妙に構造が異なっているのがわかります（背景色は、Prefs→Colors の Background 設定で白くしてあります）。温度因子は構造の不確定さの目安となるもので、青色の度合いが強い（数値が小さい）方のモデルがベターといえます。また、必ずしも本来ポリペプチドを構成する全てのアミノ酸・原子が構造に含まれているとは限りませんので、欠けている部分が少ない方がよいでしょう。但し、よほど高分解能解析（1 Å程度）の構造でない限り、水素原子は通常含まれていません。

（注）3e7z には ATOM 行ごとに ANISOU 行という異方性温度因子の情報が付随していますが、色分けに使う温度因子は ATOM 行中の等方性温度因子を指しています。また、3e7z1.pdb と 3e7z2.pdb は各1つずつのモデルしか含まれていませんが、MODEL/ENDMDL でくくられているために、ファイルを開く際に複数モデルがあるかのようなメッセージが出ます。因みに、このメッセージは SPDBV で主に NMR による PDB ファイルに含まれている複数モデルを表示させるときに機能するものです。

5. 簡易ホモロジーモデリング

5. 1 アミノ酸配列の自動差し替え

次に SPDBV の機能を使って、自分が興味のあるタンパク質のアミノ酸配列を、相同性のある構造既知のタンパク質に当てはめ、簡単なホモロジーモデリングを行ってみます。まず、上記の要領で、Default Appearance を Backbone+Sidechains に戻し、側鎖も表示されるように

しておきます。次にモデル化したいタンパク質のアミノ酸配列ファイルを先に読み込みます。これは、メニューの **SwissModel**→**Load raw sequence from amino acids** を選択し、**FASTA** 形式ファイルを指定して行います。すると1本の棒状(ヘリックス)構造が表示されます(図5)。ここでは、プロテオロドプシンという海洋微生物に広く存在する光受容タンパク質のファイル **Q9F7P4.fas** を用いています(<http://us.expasy.org>などでアミノ酸配列ファイルを手に入ると、拡張子が **.fasta** となっていますので、**.fas** に変えておきます)。全体が赤くなっているのは、温度因子の情報がないためです。

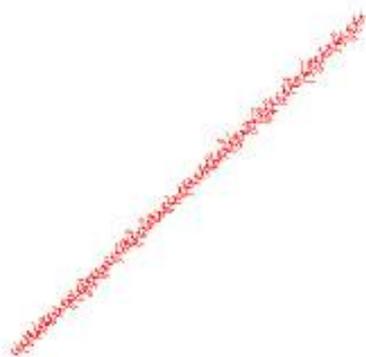


図5. 配列を読み込み後



図6. 構造読み込み後

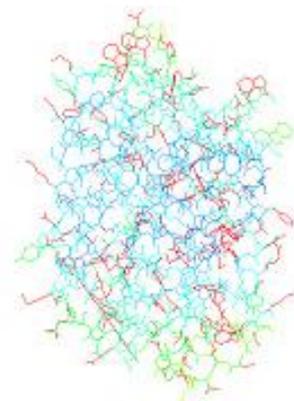


図7. 配列のモデル化後

次にこれと相同性のあるタンパク質で結晶構造が既知の光駆動プロトンポンプ、バクテリオロドプシンのファイル **1BRX.pdb** を読み込むと、図6の右上のように小さく表示されますが、温度因子が全体的に低いので青っぽくなっています。ここで、2つの配列情報を示す **Alignment** ウィンドウが表示されていなければ、メニューの **Wind**→**Alignment** を選択します。すると、横長のポップアップ画面が現れ、上の行に構造未知のタンパク質、下の行に構造既知のタンパク質のアミノ酸配列が1文字表記で示されます。

次に、メニューの **Fit**→**Magic Fit** を選択すると、赤い棒が消えて、右上の構造にモデル化された構造が重なって表示され、**Alignment** ウィンドウも配列合わせの計算結果を示すように変化します(図8)。Fileメニュー文字の下にあるセンタリングボタン  を押して構造を拡大し、回転させると図7のようになります。アミノ酸配列の異なる部位は赤色はみ出していることでわかります。

5. 2 モデルの修正

このようにして自動で行われた配列合わせを自分で修正したいときは、**Alignment** ウィンドウ内の動かしたいアミノ酸部位をクリックして、**Ctrl** キー+スペースキーでギャップが挿入でき、**Ctrl** キー+**backspace** キーでギャップを消せます。これと連動してモデル構造も変化します。

ここまでで作成されたモデル構造は、特にエネルギー的に最小化されているわけではなく、

単に既知の構造のアミノ酸配列を差し替えただけなので、随所に原子どうしがぶつかってしまっている部位があり得ます。これは、メニューの Select→Residues Making Clashes を選択すれば見つけることができます。また、Tools→Energy Minimization を実行すれば、側鎖の向きが動いて多少ぶつかり合いの数が減少します。



図 8. モデル化後に配列合わせされた状態のウィンドウ

(1BRX の配列が自動的に右にシフトして、右側の LA のように上下で一致する部位が多くなるように計算されます)

6. パワーポイントで構造を動かす

6. 1 DSV での構造ファイル作成

パワーポイントのコントロール機能を使って、生体高分子構造ファイルをプレゼンテーションの一部として自由に動かすことができますが、この目的には、DSV が適しています。DSV 本体とパワーポイント用の ActiveX コントロールをインストールすれば準備完了です。

まず DSV を起動してメインメニューの Edit→Preferences→Molecule Window→Graphics と進んで Background color を白に指定しましょう (デフォルト設定の黒が好みの場合は不要です)。次に File→Open で PDB ファイルを読み込みます。ここではミオグロビン (1MBN.pdb) を例にすると、図 9 の全原子表示になります。この構造表示画面内で右クリックするとポップアップメニューが現れ、Display Style を選択して、Atom-Display Style=None, Protein-Display Style=Flat ribbon, Coloring=Secondary Type, とすると、図 10 のようになります。このとき、色は 2 次構造で分けられていますが、ヘリックスやシート、ターンの同定方法は、Preferences→Molecule Window→Import を選択すれば確認・変更することができます。

ミオグロビンは酸素を結合するヘム分子をもっていますが、Atom-Display Style=None としたので、ヘムも消えてしまっています。そこで、メインメニューの View→Hierarchy を有効にすると、左枠として構造の階層表示が現れます。この中の、Hetatm の文字上をクリック選択して View→Display Style→Atom-Display Style=Stick とすると、図 11 のようになります。ここまでで編集を一旦止めて、File→Save As を選択して.msv 形式で保存します。

(注) この Hetatm にはヘムと結合した OH イオンも含まれていて、OH154, HEM155 という別々のユニットになっていますが、ここでは簡単のためにまとめて扱いました。

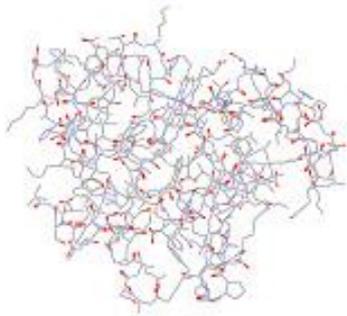


図9. 全原子表示

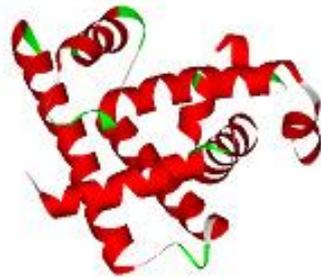


図10. 二次構造表示

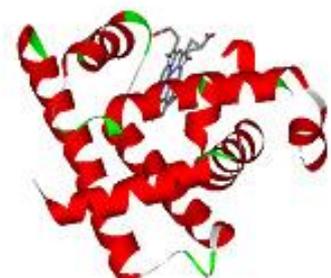


図11. 二次構造+ヘム

6. 2 PowerPointでの設定

次にPowerPointですが、DSVのActiveXコントロールがインストールされていれば、以下の手順で.msv ファイルを読み込みます。

(1)画像を挿入したいスライドを表示する

(2) [コントロールの選択] ボタン  (金槌とスパナがクロスした絵柄) をクリックする

*このボタンが表示されていないときは、PowerPoint のオプション→ユーザー設定→コマンドの選択、と進んで「すべてのコマンド」の中から [コントロールの選択] をクイックアクセスツールバーに追加して下さい。

(3)使用可能なコントロールのリストが表示されるので [Accelrys DSVisualizer Control R1] を選択する

(4)カーソルが+になるので、マウスで矩形領域を作成します (始点で左クリック→押しながらマウスを動かし、適当な形にしたらクリック解除)

(5)矩形領域の内部で右クリックし、[プロパティ] を選びます

(6) [プロパティ] ウィンドウが開くので、[Source] 行の右端のボタンをクリックして、.msv ファイルを指定します。

以上で、図11の画像がスライド上に現れますから、プレゼンテーションモードに移り、まず構造の近くで1回クリックして、その領域をアクティブにします。あとはクリックしたままマウスを動かせば回転しますし、マウスのローラーを回せば拡大・縮小、ローラーを押したままマウスを動かせば並進します。簡単な自動回転のアニメーションも、右クリックして現れるメニューで Spin を選ぶだけで実行されます。あるいは(5)と同様に編集モードで、[プロパティ] → [Spin] 行を True にしておけば、スライド開始時から回り始めます。次のスライドに進みたいときはカーソルを画像から十分離れた位置でクリックします。

7. 構造変化のアニメーション

7. 1 2つの構造ファイルの交互表示

タンパク質の働きをイメージするために、やはり形が変わる様子を観てみたいと思うのではないのでしょうか。PDBに登録されている構造などから生体高分子の動きを作成したアニメーションを集めたサイトもあります (<http://www.molmovdb.org/>)。ここでは DSV を使って、がん遺伝子 *ras* がコードする Ras タンパク質の GTP 結合型 (1LF0.pdb) と GDP 結合型 (1LF5.pdb) という 2 状態間の構造変化をアニメーションにして観てみます。因みに、この Ras タンパク質は、GTP を結合しているときに別のタンパク質への活性を示しますが、自らの触媒機能により GTP を GDP に加水分解することで活性を失います。

まず 1 つめの構造ファイル 1LF0.pdb を File→Open で開いて、次に File→Insert From で、2 つめのファイル 1LF5.pdb を指定します。このとき、必ずしも 2 つめの構造が 1 つめに重なって表示されるとは限らず、ウィンドウからはみ出ているかもしれません。そこで、2 つの構造を Structure→Superimpose→By Residue でフィッティングさせます。あとは、Structure→Animation→Create を選択するだけです。動作の実行には、Structure→Animation→Play とするか、ツールバーの再生ボタン ▶ を押せば、2 つの構造が交互に表示されて、活性型—不活性型の構造変化を観ることができます。

7. 2 仮想的な中間状態を含めた構造変化の連続表示

更にいわゆる *morphing* という操作を行って、2 つの構造の中間状態を補えば、より連続的な変化として動かすこともできます。この *morphing* には、上記の molmovdb サイト内の Morph Server を利用できます。Ras のようにポリペプチド鎖が 1 本の場合は、Single chain server を選択して、作りたいフレーム数 (初期値 8, 最大 30 まで) と 2 つの PDB ファイル (または PDB ID) を指定すれば自動的にファイルの作成が開始され、計算が終了するとメールで知らせてくれます。

メール中のリンク先へ行くと、molmovdb サイト内で動いている様子が表示されますが、画面右下部分にデータへのリンク "Download PDB coordinate file" もあり、movie.pdb.gz という圧縮ファイルとして入手できます。これを解凍すると、movie.pdb ホルダー中に temp.txt というファイルがありますので、拡張子を .pdb に変更します。このファイルには、NMR の構造ファイルと同様に複数 (用いた構造 1 + 指定したフレーム数の中間構造 + 用いた構造 2) のモデルが MODEL/ENDMDL を区切りとして収められています。

これを DSV で読み込む前に、Edit→Preferences→Molecule Window→Import と進んで、Create animation from multiple molecules をチェックしておきます。あとはファイルを開いて、上記のように Play とするか再生ボタンを押せば動作します。

(注) この Morph Server では、現在 HETATM の処理が DISABLED になっているので、GTP と GDP は含まないアニメーションになります。因みに、上記のようにして作成したファイルは、molmovdb サイト内の Motion Databases→Movie Gallery に残るので誰でも観ることができます。

以上のように、2つの別々の PDB ファイルを用いた交互表示、または1つの PDB ファイル中の複数モデルを用いた連続表示のアニメーションを DSV で作成してから、.msv ファイルとして保存すれば、パワーポイント中でこれらのアニメーションを実行することもできます。基本的な手順は前節6で説明したのと同様で、違いは [プロパティ] で [Animate] 行を True にすることだけです。あとは、[AnimationPeriod] 行の数値を変えることで動きのスピードが調節できます。交互表示には初期値の 1 0 0 0 で問題ないですが、連続表示では遅くなるので 1 0 0 くらいに下げることがあるかもしれません。

参考文献

- [1] W.G. Krebs, M. Gerstein, The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. (2000) Nucleic Acids Res 28: 1665-75.

8. ステレオ図の作り方

SPDBV に表示されている構造をウィンドウの中心に置いておけば、ステレオ画像として保存することができます。メニューの File→Save→Image(Stereo)を選択するだけです。但し、左右の画像の距離が必ずしも適切になっていないので、別途画像処理ソフトで編集する必要があるでしょう。図12には、NMR によるインスリンの構造モデル10個のアンサンブルを示しています (2HIU.pdb)。眠い時のような感じで眺めていると、左右の図がそれぞれ2つに分かれて、内側の2つが重なったときに立体的に見えることでしょう。未経験の方は一度お試し下さい。

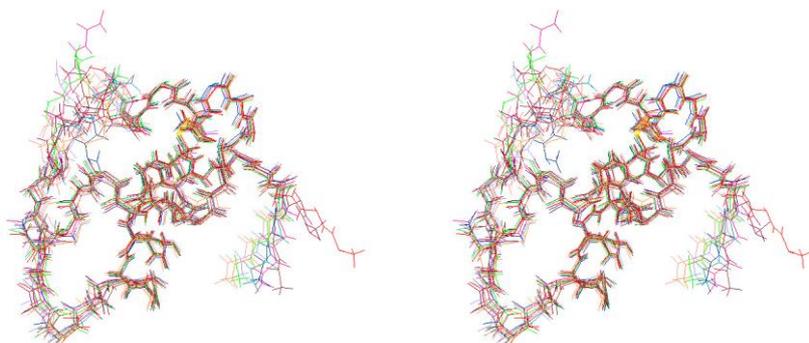


図12. NMR によるインスリン主鎖構造アンサンブルのステレオ図

9. おわりに

以上、一般的に興味をもって頂けそうなテーマを取り上げて、生体高分子を立体的に体感するためのガイドと致しました。Mac ユーザーには本章の6割くらいしか参考にならないかと思いますが、どうぞご容赦下さい。