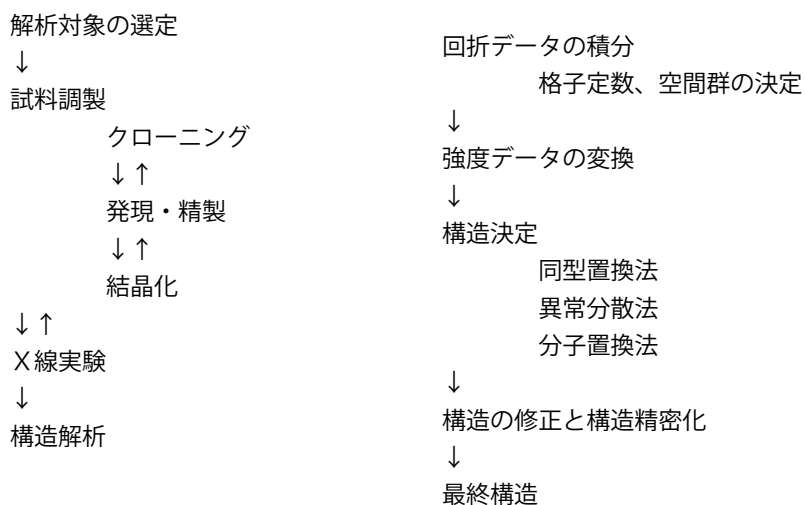


# TRF2-DNA 複合体の分子置換法による結晶構造解析

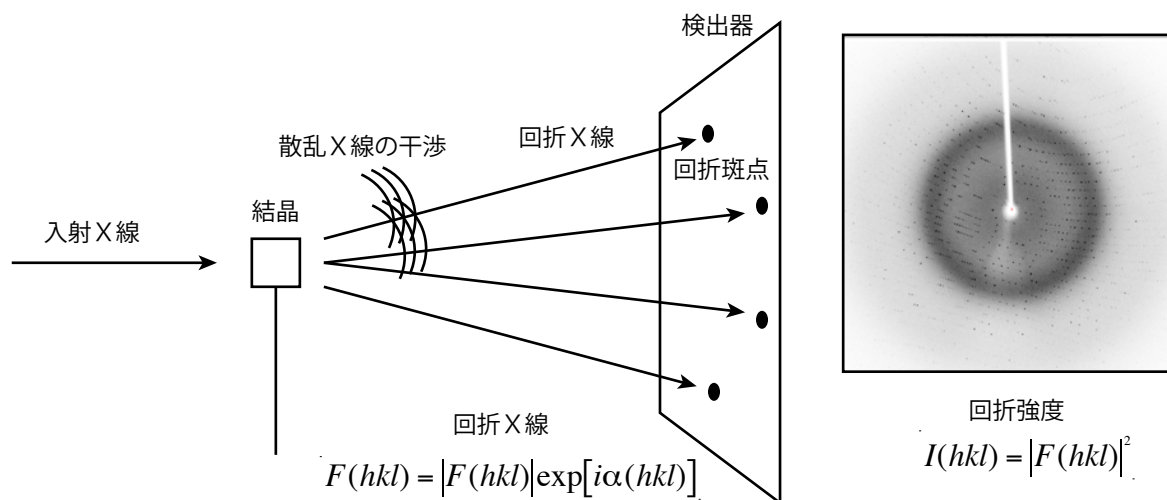
## X線結晶構造解析

X線結晶構造解析はタンパク質などの生体超分子の立体構造を決定する手法の1つで、NMR法と共に構造生物学における最も強力な構造解析法である。結晶構造解析の流れを下に示した。ここで最も重要な項目は、解析対象の選定はもちろんだが、実験的なところでは、結晶化を含めた試料調製である。遺伝子工学の進歩によって、我々は自在にタンパク質分子を作り替えることが可能になった。構造解析の過程の中で、我々は結晶化に適したタンパク質の試料調製に最も時間を費やしている。強力なX線が得られる放射光施設や測定技術、解析ソフトウェアや計算機の進歩により、良好な結晶が得られれば数日で最終構造を得ることも珍しくない。X線実験から構造解析のワークフローを下に示す。

### X線実験から構造解析の一般的なワークフロー

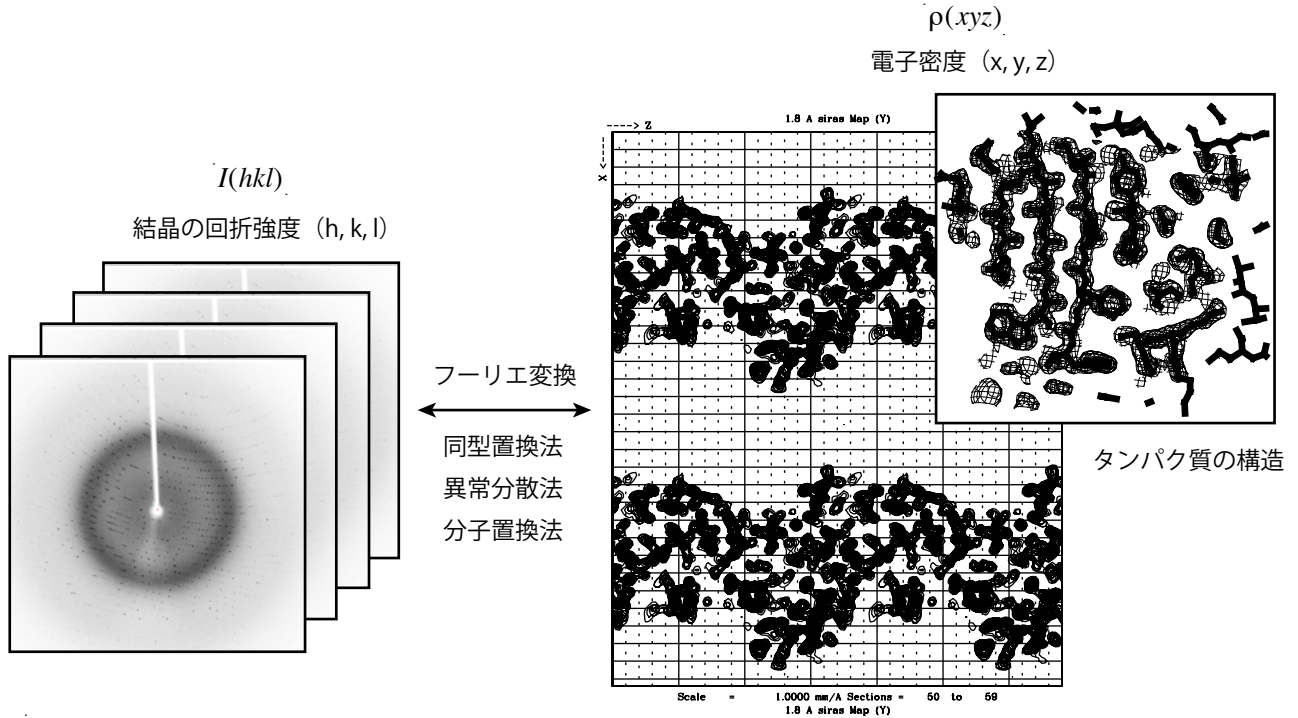


タンパク質の結晶にX線を照射すると、X線はタンパク質の電子と相互作用し散乱する。散乱X線は干渉作用によって、強め合ったり弱め合ったりすることで、特定の方向にだけ検出器上に斑点を示す。これを回折という。この回折X線はタンパク質の電子の情報を含んでおり、回折X線を解析することで、タンパク質の電子の分布（電子密度）を得ることができる。電子分布の中心には原子が存在すると考えられるので、電子密度が得られれば原子の位置が決まり、タンパク質の構造を原子レベルで決定できる。



文責：橋本 博

結晶にあらゆる方向からX線を照射し、電子の情報（電子密度） $\rho(xyz)$ を含む回折X線 $F(hkl)$ の強度 $I(hkl)$ を測定する。ここで、 $xyz$ は三次元の座標である。 $hkl$ はミラー指数といい、回折X線が回折の世界のどこから来たかを示すもので、回折の世界における座標のようなものである。これを理解するには逆格子の概念を導入する必要があり、本実習の範囲を超える。ここでは、結晶のX線像を解析するとタンパク質の電子密度が得られ、電子密度にアミノ酸などを当てはめていくとタンパク質の立体構造が決定できるという理解でよい。X詳細は結晶学の教科書を参照。



$$I(hkl) = |F(hkl)|^2$$

$$F(hkl) = |F(hkl)| \exp[i\alpha(hkl)]$$

$$= V \int_0^1 \int_0^1 \int_0^1 \rho(xyz) \exp[2\pi i(hx + ky + lz)] dx dy dz$$

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| \exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)]$$

本実習では、プログラム CCP4 と COOT を使って、テロメア DNA 結合タンパク質であるヒト TRF2 の DNA 結合ドメインと DNA との複合体のX線結晶構造解析を行う。

### TRF2 DNA 結合ドメインのアミノ酸配列

442            450            460            470            480            490            500  
 MTNITKKQKW TVEESEWVKA GVQKYGEGNW AAISKNYPFV NRTAVMIKDR WRTMKRLGMN

### 結晶化に用いた DNA 配列

5' CTCTAGGGTTAGGGTTAG 3'  
 3' AGATCCCAATCCCAATCT 5'

### タンパク質結晶構造解析の参考書

タンパク質のX線解析、佐藤 衛、共立出版  
 生命系のためのX線解析入門、D. Blow 著 平山令明 訳、化学同人  
 Principles of Protein X-ray Crystallography, J. Drenth, Springer

## CCP4I の初期設定

CCP4 Program Suite と COOT を用いて結晶構造解析を進めていく。CCP4 は X 線結晶構造解析に必要な様々なプログラムが実装されたプログラムパッケージである。個々のプログラムはスクリプトと呼ばれる入力ファイルを作成し、コマンドラインで実行するため、様々な使い方が可能であるという利点があるが、入力ファイルの作成やファイル操作など、CCP4 や Linux に関する多少の知識を必要とする。CCP4I は CCP4 のグラフィックユーザーインターフェイスであり、CCP4 に含まれるほとんどのプログラムを簡便なマウス操作で実行できる。

今回の実習では CCP4I を使用する。

CCP4 に関する情報は

<http://www.ccp4.ac.uk>

CCP4I はプロジェクトごとに指定されたディレクトリ（フォルダ）にほぼ全てのファイルを格納し、管理するため、はじめにプロジェクトのディレクトリを指定する必要がある。ここでは、kast という名前のプロジェクトを作成し、そのプロジェクトディレクトリをホームディレクトリ以下の kast という名前のディレクトリに指定する。

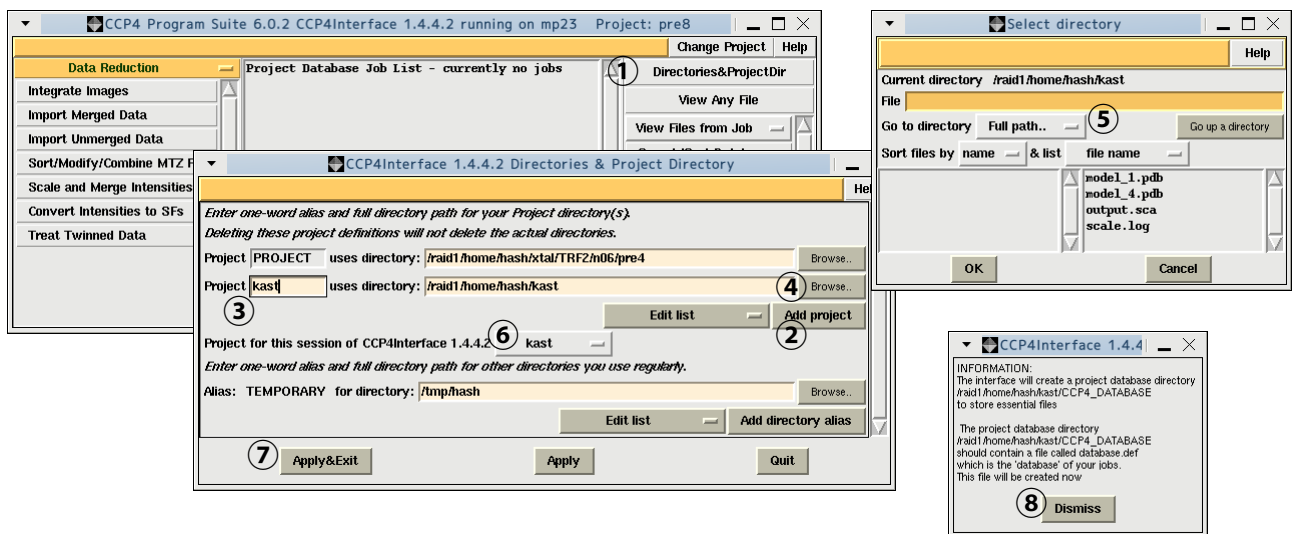
Linux にログインし、terminal を起動し、X 線のデータのあるディレクトリに移動する。

ターミナルで ccp4i と入力し、リターンキーを押すと、CCP4I のメインウィンドウが表示される。

```
% cd kast
% ls
model_1.pdb model_4.pdb output.sca
```

(ここで、model\_1.pdb と model\_4.pdb は分子置換法での構造解析に使用するサーチモデル、output.sca は HKL2000 が出力した回折強度データである。)

- ①メインウィンドウ右上の <Directories & Project Dir> をクリックし、プロジェクトウィンドウを開く。
- ②プロジェクトウィンドウの <Add project> を左クリックし、新たにプロジェクトを作成する。
- ③プロジェクト名を kast と入力する。
- ④<Browse..> をクリックし、セレクトウインドウを開く。
- ⑤<Go to directory> のプルダウンメニューで <Full path..> を選択し、<OK> をクリックする。
- ⑥プロジェクト名を選択する。ここでは kast。
- ⑦<Apply & Exis> をクリック
- ⑧確認のウィンドウが表示される場合は、<Dismiss> をクリック。

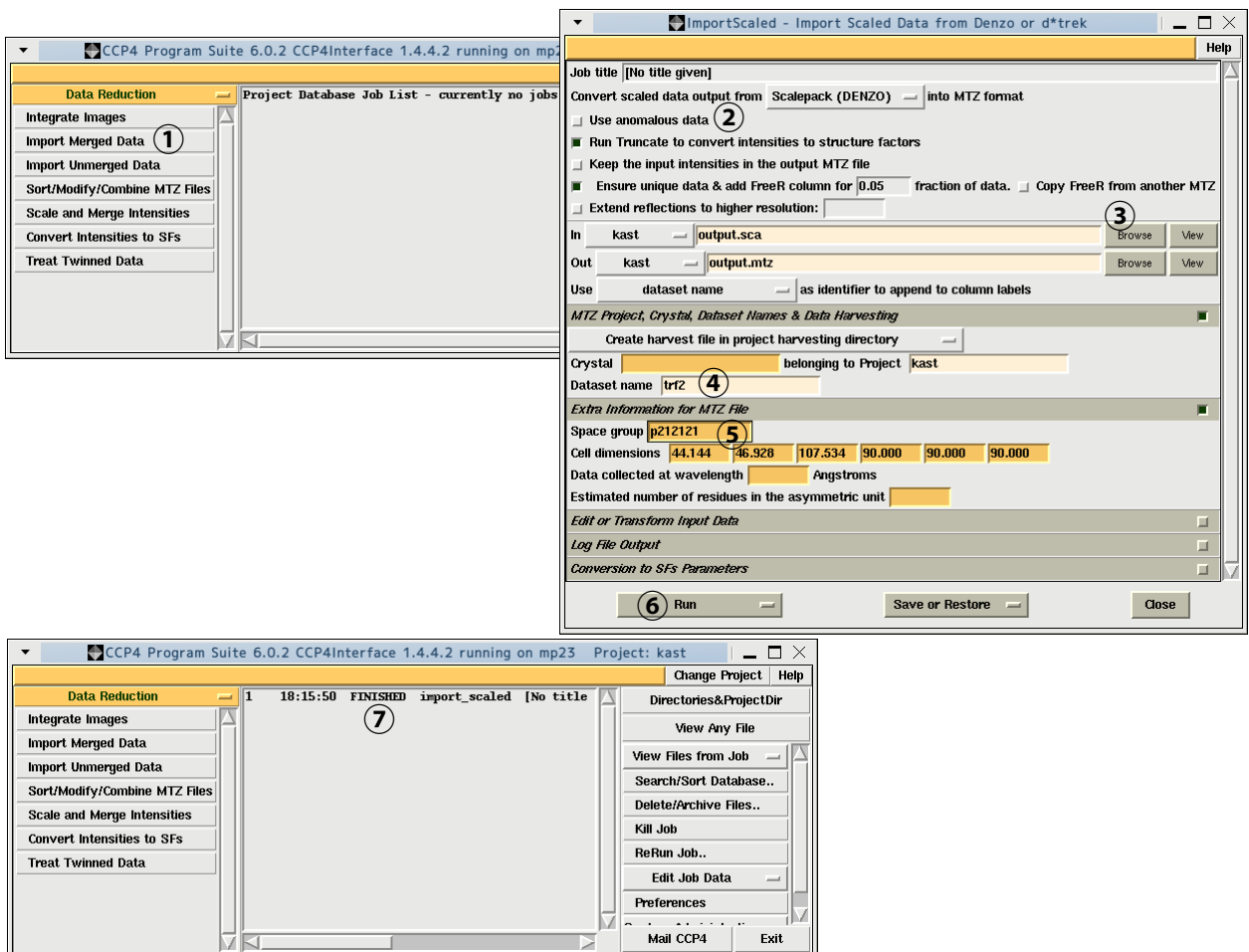


これで CCP4I の初期設定は終了。

## 強度データの変換

HKL2000 で出力された積分強度 output.sca を構造解析に使用するためにはデータ変換をする必要がある。

- ① CCP4I のメインウィンドウ左上のプルダウンメニューで <Data Reduction> を選択し、<Import Merged Data> を左クリックすると <Import Scaled> ウィンドウが開く。
- ② <Use anomalous data> 左のチェックを外す。  
(Anomalous data とは特殊な原子の入った結晶を特定の波長の X 線を使って観測できる異常分散効果であり、今回使用する X 線データは異常分散効果を含んでいない)
- ③ <In> 右の <Browse...> をクリックして、セレクトウインドウから output.sca を選択し、<OK> をクリックすると、<Out> の欄には自動で output.mtz と入力される。
- ④ <Dataset name> に trf2 と入力する。(名前は何でもよい)
- ⑤ 今回取り扱う TRF2-DNA 複合体結晶の空間群は P212121 なので、<Space group> に p212121 と入力する。
- ⑥ <Run><Run Now> を選択し実行する。
- ⑦ 実行するとメインウィンドウに Running と表示され、正常に終了すると、Finished と表示される。  
これで強度データ output.sca は、以後の構造解析に使用できる F データ output.mtz に変換された。





## 分子置換 ~1~

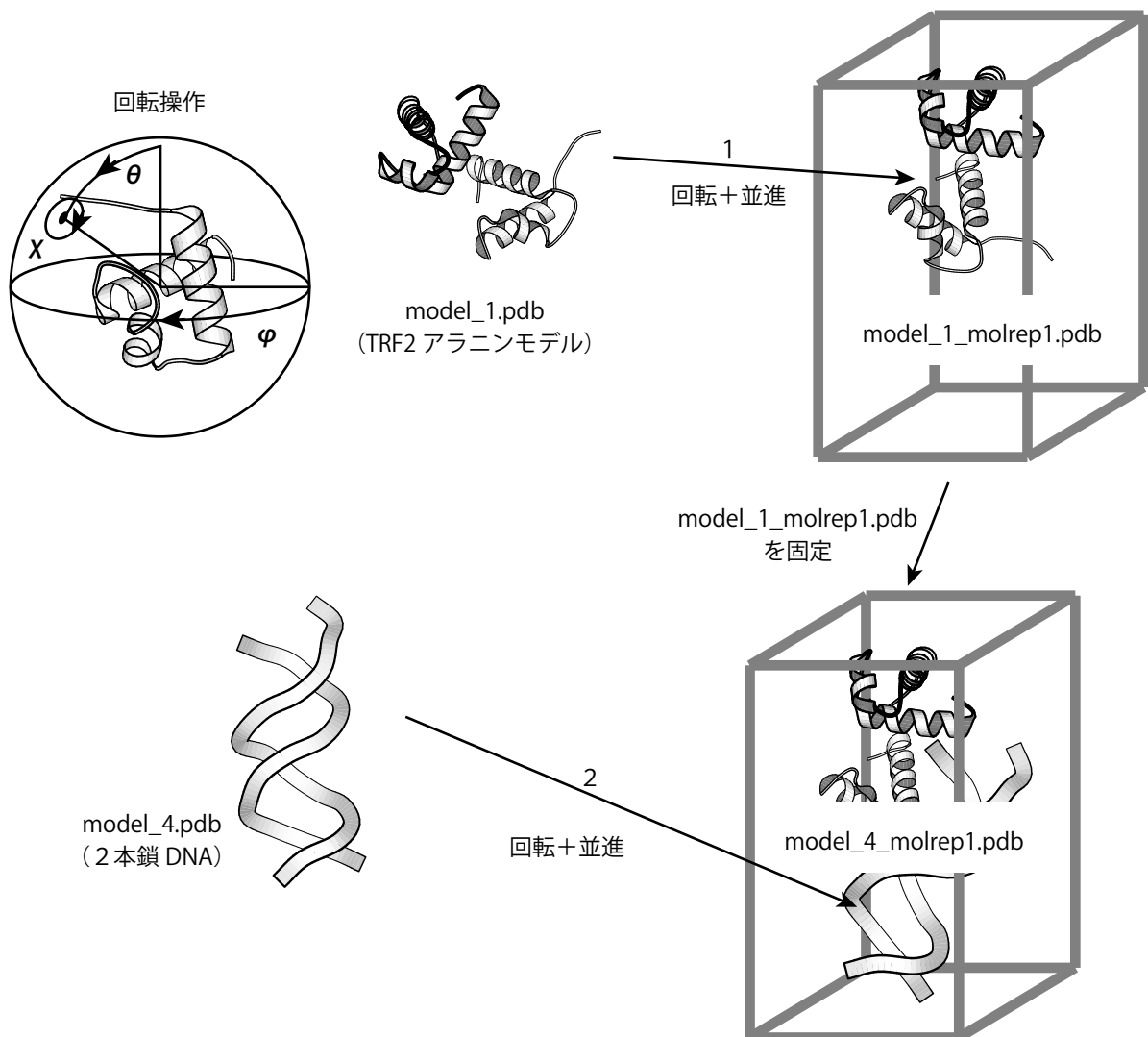
分子置換法はX線結晶構造解析における構造決定法の1つで、すでに目的タンパク質の類似構造（と思われる構造）があれば、それをモデル分子（サーチモデル）として、そのモデルに回転と並進の操作を行い、目的タンパク質の構造を決定することができる。雑な言い方をすれば、目的タンパク質と似ていると考えられる構造を使って、それを回転させたり並進させたりして、目的タンパク質の結晶格子中にどんな向きで収まっているかを決定することである。

分子置換法において、使用するサーチモデルの選択が重要で、単に類似構造をそのままモデル分子として使うだけでは構造決定できず、アミノ酸をすべてアラニンやセリンに置き換えたモデルを使用したり、配列アライメントなどの情報から、サーチモデルを改変して使用することもある。同じタンパク質であっても、構造変化している場合は、ドメインに分けたり、使用するサーチモデルに工夫が必要かもしれない。もちろん簡単に構造決定できる場合も多いが、非常に困難な場合もあり、お手軽に見えるが奥が深い構造決定法である。また、使用するプログラムを変えるとうまくいく場合もある。

分子置換法のプログラムはたくさん知られているが、今回は MOLREP を使用する。MOLREP は CCP4 Program Suite に含まれており、CCP4I から GUI 操作が可能である。入力項目が少ないため使いやすい。まずは MOLREP を使ってみて、うまくいかない場合は他のプログラムを試すのがよいだろう。

今回の実習では、TRF2 のアラニンモデル（全てのアミノ酸をアラニンに置換した構造）と 2 本鎖 DNA をサーチモデルとして、TRF2-DNA 複合体の構造を 2 段階の分子置換法で決定する。

1. まず、TRF2 アラニンモデルに対して回転と並進を行う。2. その後、動かした TRF2 アラニンモデルを固定し、それに対し（その隙間に）、DNA を回転と並進で当てはめる。



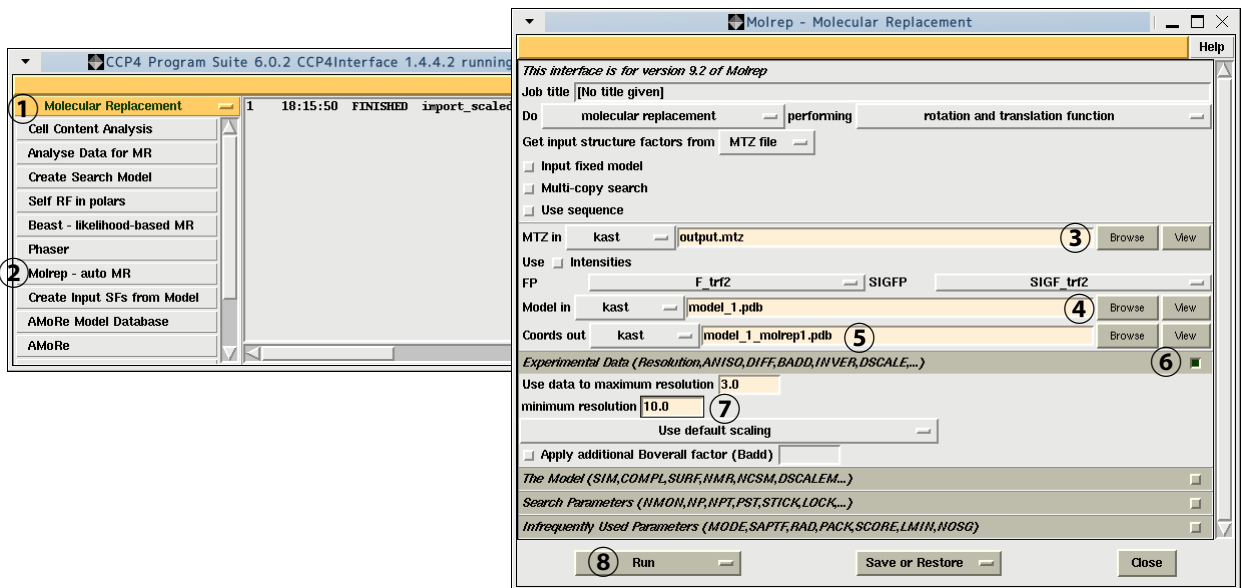
# 分子置換 ~ 2 ~

## 1. TRF2 アラニンモデルを使った分子置換

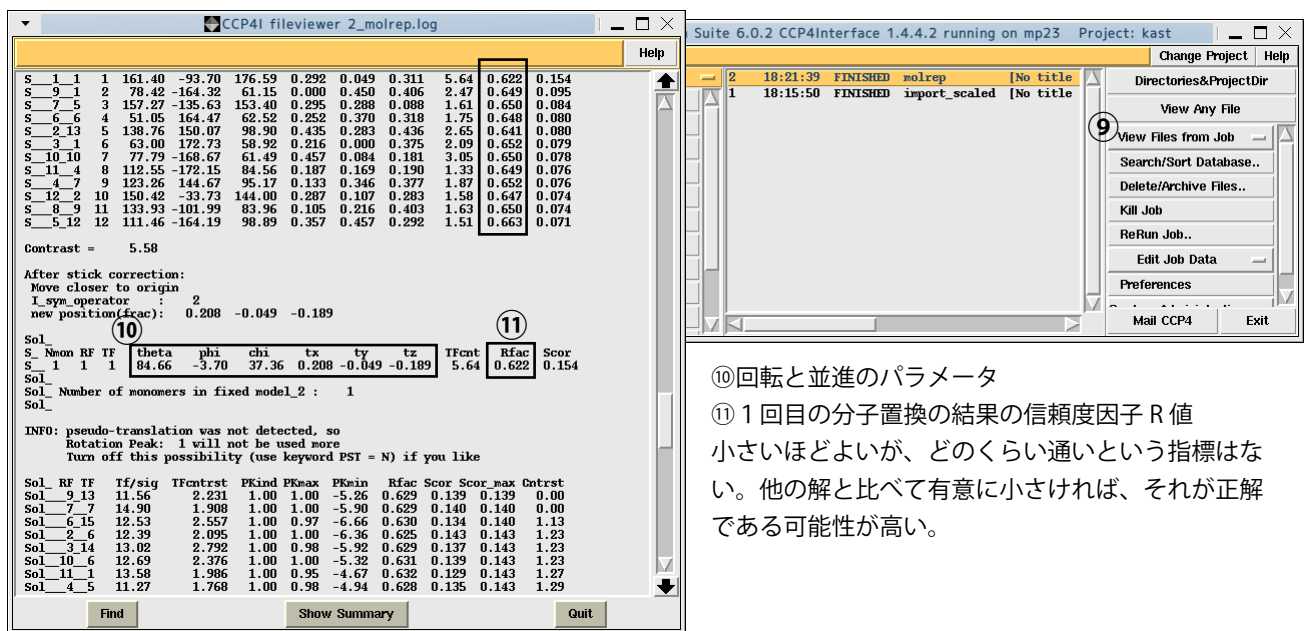
- ①メインウィンドウ左上のプルダウンメニューで <Molecular Replacement> を選択する。
- ②<Molrep auto MR> を左クリックし、Molrep ウィンドウを開く。
- ③<MTZ in> 右端 <Browse...> をクリックし、セレクトウインドウから output.mtz を選択し、<OK> をクリック。
- ④<Model in> 右端 <Browse...> をクリックし、セレクトウインドウから model\_1.pdb を選択し、<OK> をクリックする。
- ⑤自動で <Coords out> に model\_1\_molrep1.pdb が入力される。自動入力されない場合は、キーボードから入力する。
- ⑥<Experimental Data> の右端をチェックして、拡張項目を表示する。
- ⑦計算に使用する X 線データの分解能を入力する。分子置換法であれば、高分解能なデータは必要ない。今回は 3~10Å の範囲のデータを使用する。

<Use data to maximum resolution> は 3、minimum resolution に 10 と入力する。

- ⑧<Run><Run Now> を選択し、Molrep を実行する。



- ⑨終了するとメインウィンドウに Finished と表示されるので、左クリックで選択し（黄色に変わる）、左側の <View Files from job> のプルダウンメニューから <View Log File> を選択するとログが表示される。



⑩回転と並進のパラメータ

⑪ 1 回目の分子置換の結果の信頼度因子 R 値  
小さいほどよいが、どのくらい通いという指標はない。他の解と比べて有意に小さければ、それが正解である可能性が高い。

## 分子置換 ~ 3 ~

1 回目の分子置換が終了したら、つぎは動かした TRF2 アラニンモデルを固定して、DNA をサーチモデルとして分子置換を行う。

### 2. TRF2 アラニンモデルを固定し、DNA を使った分子置換

- ① Molrep ウィンドウで、<Input fixed model> にチェックを入れる。
- ② <MTZ in> で output.mtz を選択する。(既に入力されている場合は不要)
- ③ <Model in> で model\_4.pdb を選択する。
- ④ <Fix in> で model\_1\_molrep1.pdb を選択する。
- ⑤ <Coords out> には自動で model\_4\_molrep1.pdb が入力される。自動入力されない場合は、キーボードから入力する。
- ⑥、⑦ 分解能は 3~10 Å
- ⑧ <Run><Run Now> を選択し、Molrep を実行する。
- ⑨ 終了するとメインウィンドウに Finished と表示されるので、左側の <View Files from job> のプルダウンメニューから <View Log File> を選択しログを表示する。

The image shows two windows from the CCP4 software suite. The top window is 'Molrep - Molecular Replacement' and the bottom window is 'CCP4 fileviewer 3\_molrep.log'.

**Molrep - Molecular Replacement window:**

- Job title: [No title given]
- Do: molecular replacement
- Get input structure factors from: MTZ file
- Input fixed model (1)
- MTZ in: kast, output.mtz (2)
- Model in: kast, model\_4.pdb (3)
- Fixed in: kast, model\_1\_molrep1.pdb (4)
- Coords out: kast, model\_4\_molrep1.pdb (5)
- Use data to maximum resolution: 3.0 (6)
- minimum resolution: 10.0 (7)
- Run button (8)

**CCP4 fileviewer 3\_molrep.log window:**

```

Sol_10_5 8.857 3.121 1.00 1.00 -10.87 0.622 0.207 0.370 3.47
Sol_11_5 11.15 2.646 1.00 1.00 -9.26 0.625 0.217 0.370 3.57
Sol_12_1 9.755 2.675 1.00 1.00 -4.30 0.632 0.178 0.370 3.41
INFO: contrast is good enough. Stop this run

--- Summary ---
S_ RF TF theta phi chi tx ty tz TFcnt Rfac Scor
S_ 1_ 1 1 84.45 105.63 74.53 0.183 0.990 0.314 17.43 0.556 0.370
S_ 4_ 1 2 41.33 143.98 169.79 0.704 0.582 0.764 4.93 0.599 0.270
S_ 3_ 1 3 42.64 142.11 177.34 0.704 0.566 0.769 3.37 0.608 0.239
S_ 2_ 2 4 138.25 -34.73 177.43 0.187 0.905 0.263 2.52 0.623 0.219
S_ 11_ 5 5 46.39 137.04 155.20 0.676 0.573 0.731 2.65 0.625 0.217
S_ 5_ 5 6 142.62 -27.78 159.35 0.702 0.636 0.787 2.84 0.623 0.209
S_ 10_ 5 7 99.07 104.84 72.66 0.684 0.616 0.755 3.12 0.622 0.207
S_ 8_ 7 8 144.45 -21.93 150.40 0.778 0.463 0.653 1.42 0.634 0.202
S_ 6_ 9 9 143.68 -24.35 152.60 0.710 0.637 0.787 1.66 0.631 0.200
S_ 9_ 2 10 88.37 110.70 82.62 0.198 0.007 0.310 2.60 0.618 0.199
S_ 7_ 7 11 22.06 163.70 122.57 0.877 0.353 0.319 2.48 0.634 0.183
S_ 12_ 1 12 24.23 -179.46 134.58 0.308 0.076 0.796 2.67 0.632 0.178

Contrast = 3.41

After stick correction (more information in molrep.doc):
Move closer to molecule N : 1
Rotation to molecule N (polar angles) : 37.14 -164.37 173.28
Distance between centres (A) : 5.3
Orthogonal axis distance and parallel (A) : 3.5 24.1
l_syn_operator : 2
new position(frac): 0.317 0.010 -0.186

Sol_
S_ Nmon RF TF theta phi chi tx ty tz TFcnt Rfac Scor
S_ 1_ 1 1 84.26 15.63 173.28 0.317 0.010 -0.186 17.43 0.556 0.370
--- convert 'molrep.crd' to 'molrep.pdb' ---
Time: 18h 25m 24s Elapsed: 0h 1m 7s
MOLREP(ccp4)9.4.09: Normal termination
Times: User: 66.4s System: 0.7s Elapsed: 1:08
    
```

⑩回転と並進のパラメータ

⑪ 2 回目の分子置換の結果の信頼度因子 R 値

小さいほどよいが、どのくらい通いという指標はない。他の解と比べて有意に小さければ、それが正解である可能性が高い。また、1 回目と比べて R 値は小さくなるはずである。

新たにターミナルを開き、データディレクトリに移動し、rasmol 等で得られた構造を見てみるのも良い。

```
% rasmol model_4_molrep1.pdb
```

## 構造の修正と構造精密化 ～1～

ここまでで、分子置換法によって、TRF2-DNA 複合体の大まかな構造を得ることができた。しかし、似ていると考えられる構造をサーチモデルにしているため、正しい構造に近づけていく必要がある。そのために、構造の修正と構造精密化を行う。構造の修正はコンピューターグラフィックのプログラム COOT を使ってマニュアルで行う。構造精密化は精密化のプログラム REFMAC を実行し、計算機が行う。

### 構造精密化について

先に、構造精密化とは決定した構造を正しい構造に近づけていくことであると述べた。構造の正しさの指標はいくつかあるが、構造精密化では以下に示す信頼度因子 R 値を指標とする。

タンパク質の原子座標が与えられれば（タンパク質の構造があれば）、X線のデータを逆算できる（ $F_{calc}$ ）。理想的には、実験的に得られたX線データ（ $F_{obs}$ ）と得られた構造から逆算した  $F_{calc}$  は等しいため、信頼度因子  $R = 0\%$  である。しかし、実験には誤差があるため、実際には  $R = 0\%$  とはならない。データの質に依存するが、 $R = 20\%$  程度であれば良く精密化された（正しい）構造であるといえる。また、 $R$ -free 値は  $R$  値 +5% 程度であると良い。

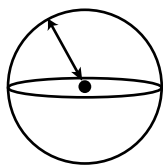
$$R = \frac{\sum_{hkl} \left| |F_{obs}| - |F_{calc}| \right|}{\sum_{hkl} |F_{obs}|} \times 100\% \quad R_{free} = \frac{\sum_{hkl \notin T} \left| |F_{obs}| - |F_{calc}| \right|}{\sum_{hkl \notin T} |F_{obs}|} \times 100\%$$

精密化するパラメータは原子座標 ( $x, y, z$ ) 以外に温度因子  $B$  がある。温度因子は原子の揺らぎを表現する量で以下の式で表現される。温度因子の単位は  $\text{\AA}^2$  である。

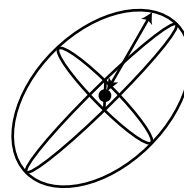
$$B = 8\pi^2 \overline{u^2}$$

たとえば温度因子が  $80 \text{\AA}^2$  である原子は、原子核を中心として平均的に  $1 \text{\AA}$  も動いていることになる。タンパク質の表面やループなどもともと運動性の高いところは温度因子が大きく、分子内部など運動性の低いところは温度因子が小さい。

等方性温度因子は、原子が全方向一様に（球状に）揺らいでいると仮定している。一方、異方性温度因子は原子の揺らぎは一樣でなく、ある特定の方向に偏っている（楕円状）と見なしており、より現実に近いと考えられる。高解像度の結晶構造解析では非等方性温度因子を導入して構造精密化できる。



等方的な揺らぎ



異方的な揺らぎ

## 構造の修正と構造精密化 ～2～

マニュアルでの構造修正を行う前に、分子置換で得られた直後の構造を少し精密化する。この場合、原子の1つ1つの構造を精密化するのではなく、タンパク質分子を塊として考え、精密化を行う。これを Rigid body refinement という。REFMAC (CCP4) を使って構造精密化を行う。

- ① CCP4I のメインウィンドウ左上のプルダウンメニューから <Refinement> を選択
- ② <Run Refmac5> を左クリック
- ③ Refmac5 ウィンドウで <Do><rigid body refinement> を選択する。
- ④ <MTZ in> で output.mtz を選択
- ⑤ <MTZ out> は output\_refmac1.mtz が自動入力される。自動入力されない場合はキーボードで入力する。
- ⑥ <PDB in> は 2 回目の分子置換の解、model\_4\_molrep1.pdb を選択する。
- ⑦ <PDB out> は model\_4\_molrep1\_refmac1.pdb が自動入力される。
- ⑧、⑨ 構造精密化に使用するデータの分解能を入力する。今回は 20 - 2 Å の範囲で行う。
- ⑩ <Run><Run Now> で実行。終了したらログで R 値を確認する。

The screenshot displays the CCP4 Program Suite interface. The main window is titled "Run Refmac5" and shows the configuration for a rigid body refinement job. The "Do" dropdown is set to "rigid body refinement". The "MTZ in" field is set to "output.mtz" and "MTZ out" is set to "output\_refmac1.mtz". The "PDB in" field is set to "model\_4\_molrep1.pdb" and "PDB out" is set to "model\_4\_molrep1\_refmac1.pdb". The "Refinement Parameters" section shows "Do 20 cycles of maximum likelihood rigid body refinement" and "Resolution range from minimum 20.0 to 2.0". The "Use weighting term" is set to 0.3. The "Run" button is highlighted with a circled 10.

The "CCP4I fileviewer 4\_refmac5.log" window shows the output of the refinement. The overall R factor is 0.4247, which is circled with a circled 11. The log also shows the resolution limits (20.000 and 2.000) and the number of used reflections (14763).

Resolution limits	20.000	2.000
Number of used reflections	= 14763	
Percentage observed	= 99.2928	
Percentage of free reflections	= 5.2682	
Overall R factor	= 0.4247	
Free R factor	= 0.4242	
Overall weighted R factor	= 0.4415	
Free weighted R factor	= 0.4445	
Overall correlation coefficient	= 0.6499	
Free correlation coefficient	= 0.6604	
Cricksbanks DPI for coordinate error	= 0.2493	
DPI based on free R factor	= 0.2490	
Overall figure of merit	= 0.5764	
M. based su of positional parameters	= 0.2664	
M. based su of thermal parameters	= 9.6753	

- ⑪ Rigid body refinement 後の信頼度因子 R 値もちろん小さいほど良い。  
この時点で R と R-free は約 42%。これから構造の修正と構造精密化を繰り返す、R 値を下げていく。

## 構造の修正と構造精密化 ～3～

次はいよいよプログラム COOT で電子密度を見ながら構造の修正を行う。

COOT は CCP4 Program Suite と関連したプログラムで、CCP4I 同様に GUI が充実しており非常に使いやすい。また、電子密度への構造のフィッティング機能が強力であるため、広く使われているプログラムである。

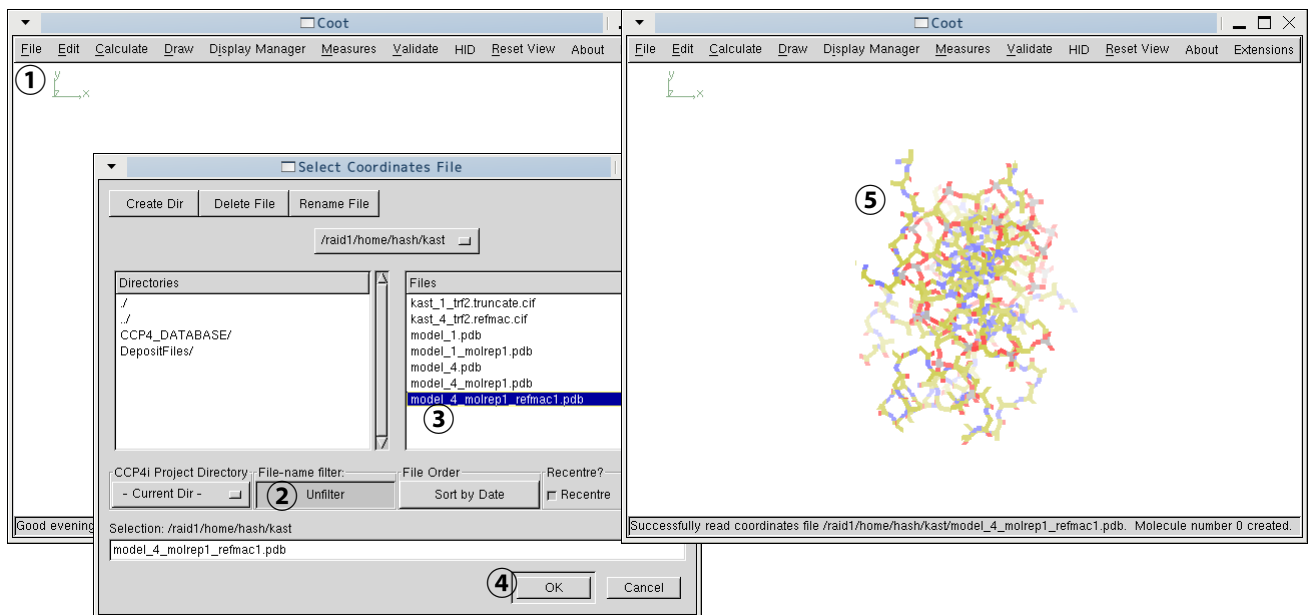
<http://www.yasbl.york.ac.uk/emsley/coot/>

ターミナルで、データディレクトリに移動し、COOT を起動する。

```
% coot
```

と入力すると COOT のメインウィンドウが起動する。

- ①メニューの <File> から <Open Coordinates...> を選択する。
- ②セレクトウインドウで <Filter> をクリック。
- ③先ほど REFMAC で精密化された model\_4\_molrep1\_refmac1.pdb を選択する。
- ④<OK> をクリックするとメインウィンドウに構造が表示される⑤。



- ⑥再びメニュー <File> から <Auto Open Mtz...> を選択する。
- ⑦セレクトウインドウで <Filter> をクリック。
- ⑧REFMAC の精密化で出力された X 線データ（電子密度の情報が含まれている） output\_refmac1.mtz を選択する。
- ⑨<OK> をクリックするとメインウィンドウの構造に電子密度が重なって表示される⑩。





# 構造の修正と構造精密化 ～4～ COOTの基本操作1

## マウス操作



左ボタンをクリックしながらマウスを動かすと分子が回転する。

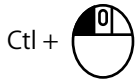


右ボタンをクリックしながらマウスを動かすと分子が拡大・縮小する。

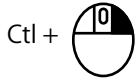


中ボタンで原子をクリックすると、クリックした原子が画面の中心に移動する。

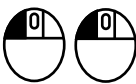
ホイールを回転させると電子密度のレベル(濃さ)が変わる。



Ctrl + 左ボタンを押しながらマウスを動かすと分子が並進する。

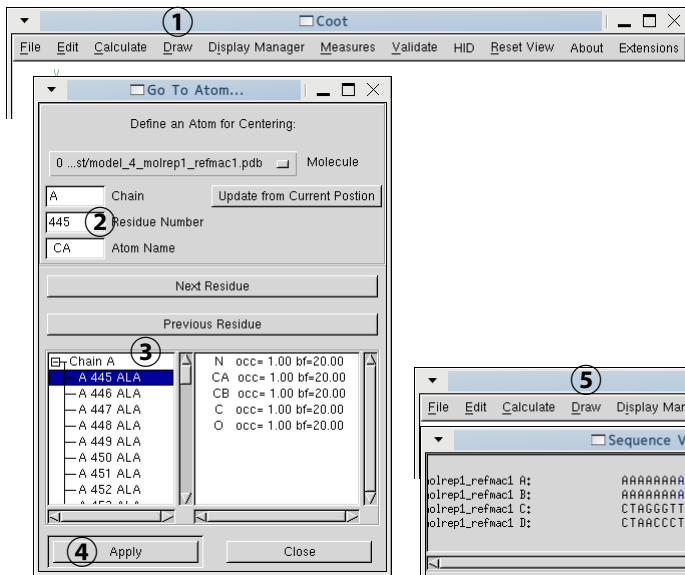


Ctrl + 右ボタンを押しながらマウスを動かすと画面の奥行きが変わる。



原子を左ボタンでダブルクリックすると残基番号、残基名が表示される。  
(Shift + 左ボタンのシングルクリックでも同様)

## 特定の残基に移動する



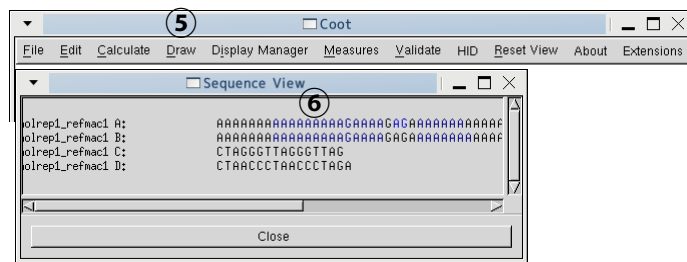
①メインウィンドウのメニューから <Draw><Go to Atom..> を選択。

②セレクトウインドウで、移動したい残基番号を入力するか、③一覧から移動したい残基を選択。

④<Apply> をクリックすると、その残基が画面の中央に移動する。

Space キーで次の残基に移動する。

Shift + Space キーで1つ前の残基に移動する。

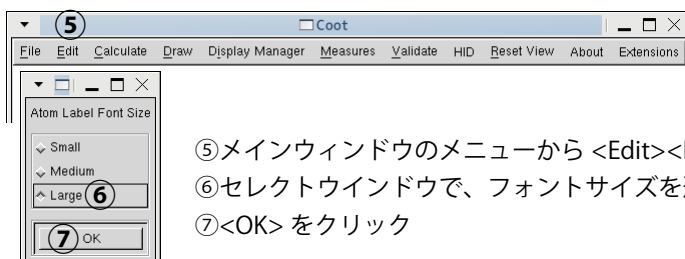


あるいは、

⑤メインウィンドウのメニューから <Draw><Sequence View> で選択し、配列を表示

⑥配列にマウスカーソルを合わせると残基番号が表示されるので、クリックするとその残基に移動できる。

## 画面の文字を大きくする



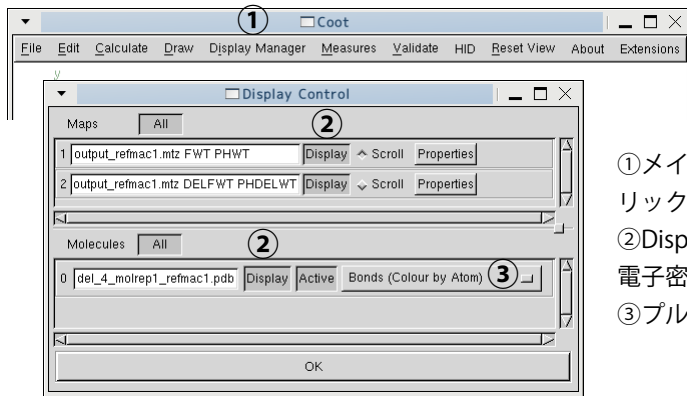
⑤メインウィンドウのメニューから <Edit><Font Size..> を選択。

⑥セレクトウインドウで、フォントサイズを選択。

⑦<OK> をクリック

## 構造の修正と構造精密化 ～5～ COOTの基本操作2

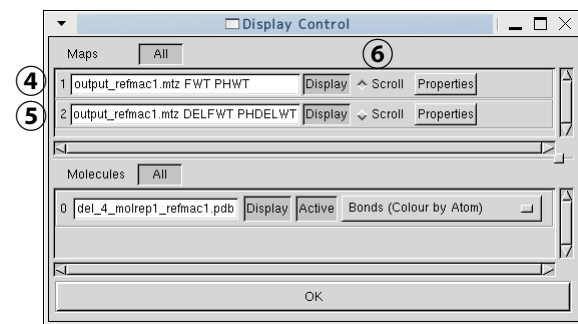
### 表示のONとOFF



- ①メインウィンドウのメニューで <Display Manager> をクリック。
- ②Display control ウィンドウの <Display> ボタンで、分子や電子密度の表示をON, OFFできる。
- ③プルダウンメニューで分子の表示方法を変更できる。  
通常は <Bonds (Colour by Atom)>

### 電子密度の調節

REFMAC は電子密度の情報を含んだX線データを出力する (output\_refmac1.mtz)。これを用いて COOT は2種類の電子密度を表示する。1つは FWT PHWT (2FoFc マップ)④、もう一つは DELFWT PHDELWT (FoFc マップ)⑤である。



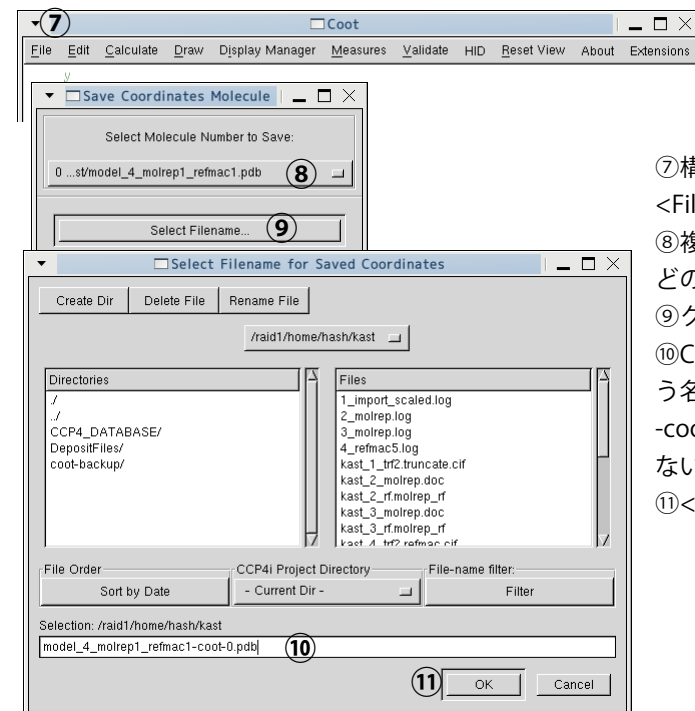
ここでは、2FoFc マップは青色で、FoFc マップは緑色で表示されている。

⑥電子密度の <Scroll> を選択し、マウスのホイールを回転させると電子密度の濃さを調節できる。



通常、2FoFc (青) は 1.0 程度、FoFc (緑) は 2.0 ~ 3.0 程度で表示するのがよい。

### 構造を保存する

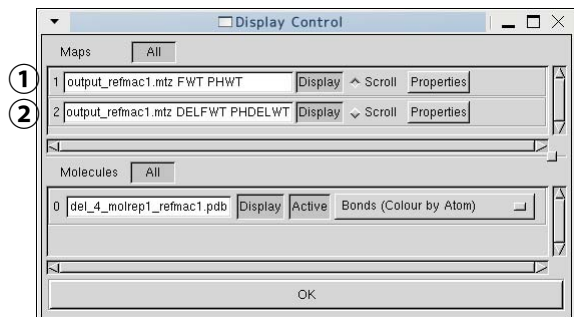


- ⑦構造を保存するには、メインウィンドウのメニューで <File><Save Coordinates ...> を選択。
- ⑧複数の構造 (PDB ファイル) を読み込んでいる場合は、どの構造を保存するかをプルダウンメニューで選択
- ⑨クリック
- ⑩COOT ではオリジナルのファイル名の後ろに -coot-0 という名前で PDB ファイルを作成する。次に保存する際には、-coot-1 になり、数字が増えていくので、上書きする心配がない。
- ⑪<OK> をクリック

## 構造の修正と構造精密化 ～6～

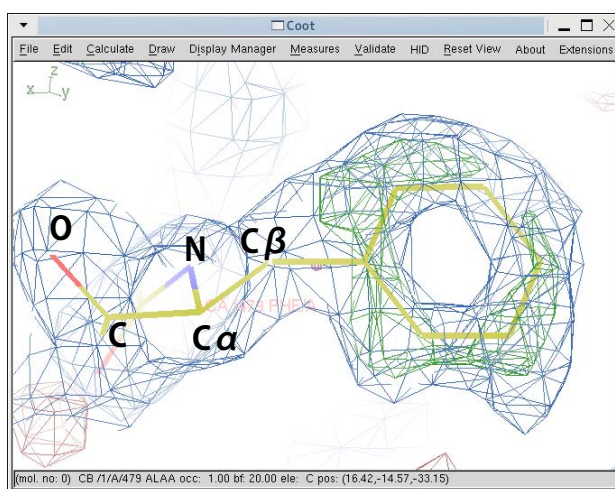
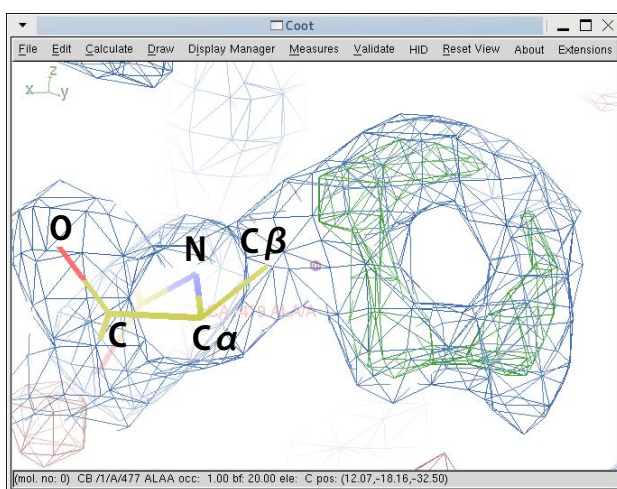
### 電子密度について

REFMAC は電子密度の情報を含んだX線データを出力する (output\_refmac1.mtz)。これを用いて COOT は2種類の電子密度を表示する。1つは FWT PHWT (2FoFc マップ)①、もう一つは DELFWT PHDELWT (FoFc マップ)②である。



テキストでは 2FoFc マップは青色で、FoFc マップは緑色で表示されている。

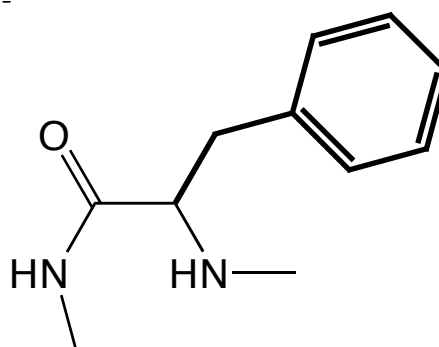
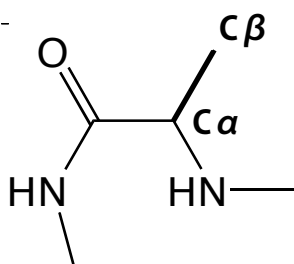
詳細は本実習の範囲を超えるが、大まかには、青い電子密度とタンパク質の構造は重なっているはずであり、青い電子密度があるのに構造がない、かつ緑の電子密度が現れている部分には何か原子を当てはめるべきである (下図参照)。



黄色が炭素原子、青色が窒素原子、赤色が酸素原子 (緑色が硫黄原子)

Cβの先にも青色の電子密度と緑色の電子密度があるので、Cβの先に何か原子が存在することを示している。

電子密度が環状であることから、側鎖にベンゼン環をもつフェニルアラニンであることが分かる。



最終的には 2FoFc マップ (青色の電子密度) とタンパク質の構造が重なり、FoFc マップ (緑色の電子密度) が無くなるように構造を精密化していくが、実際は FoFc が 2~3 の濃さのレベルで全くなくなることはない。

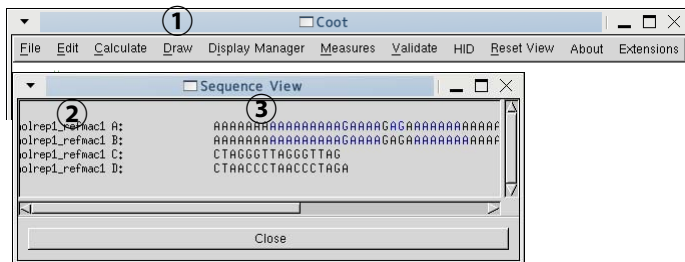
# 構造の修正と構造精密化 ~7~

## ポリアラニンモデルから TRF2 へ

ここからが本題である。

本実習では分子置換のサーチモデルとして、TRF2 アラニンモデルを使用したため、本来のアミノ酸に置き換えていく作業が必要になる。2本鎖DNA1分子に対して、TRF2は2分子結合している。現在の構造のアミノ酸配列を確認してみる。

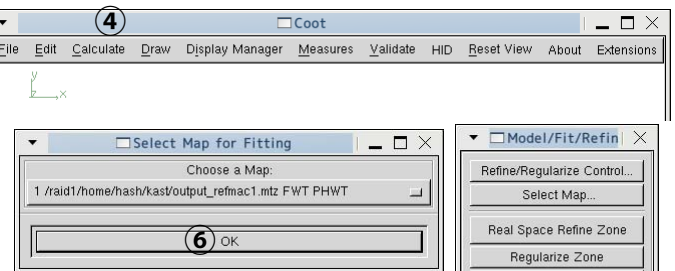
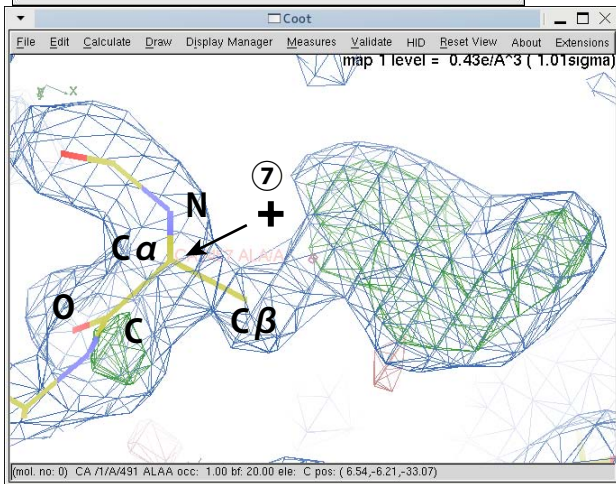
①メインウィンドウのメニューから <Draw><Sequence View> で選択すると配列を確認できる。



②読み込まれている構造の chain (ポリペプチド鎖やオリゴヌクレオチド鎖) ごとに配列が表示される。ここでは TRF2 は A-chain と B-chain であり、DNA は C-chain と D-chain が 2 本鎖を作っている。TRF2 の A-chain と B-chain はポリアラニン構造であることが確認できる (一部グリシン)。

③配列にマウスカーソルを合わせると残基番号が表示されるので、クリックするとその残基が画面の中央になる。

例えば、A-chain の 457 に移動する。電子密度の濃さは、2FoFc (map1, 青) と FoFc (map2, 緑) はそれぞれ、1 と 2.5 位が良いだろう。



TRF2 のアミノ酸配列を確認すると、本来 457 は W (トリプトファン) である。

このアラニンをトリプトファンに置換するには、

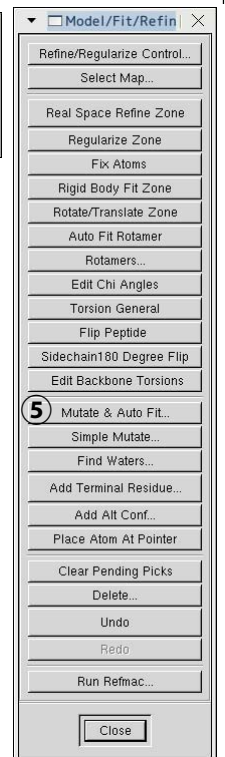
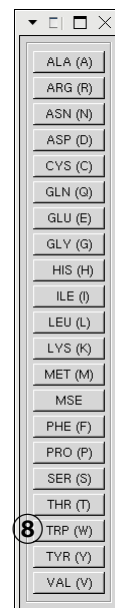
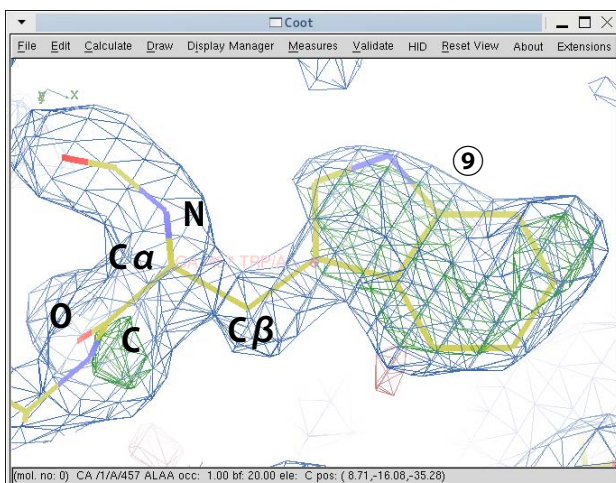
④メニューから <Calculate><Model/Fit/Refine..> を選択、Refine ウィンドウを表示。

⑤<Mutate & Auto Fit...> をクリック。

⑥電子密度の確認、<OK> をクリック。

⑦+字カーソルで、何れかの原子をクリック。

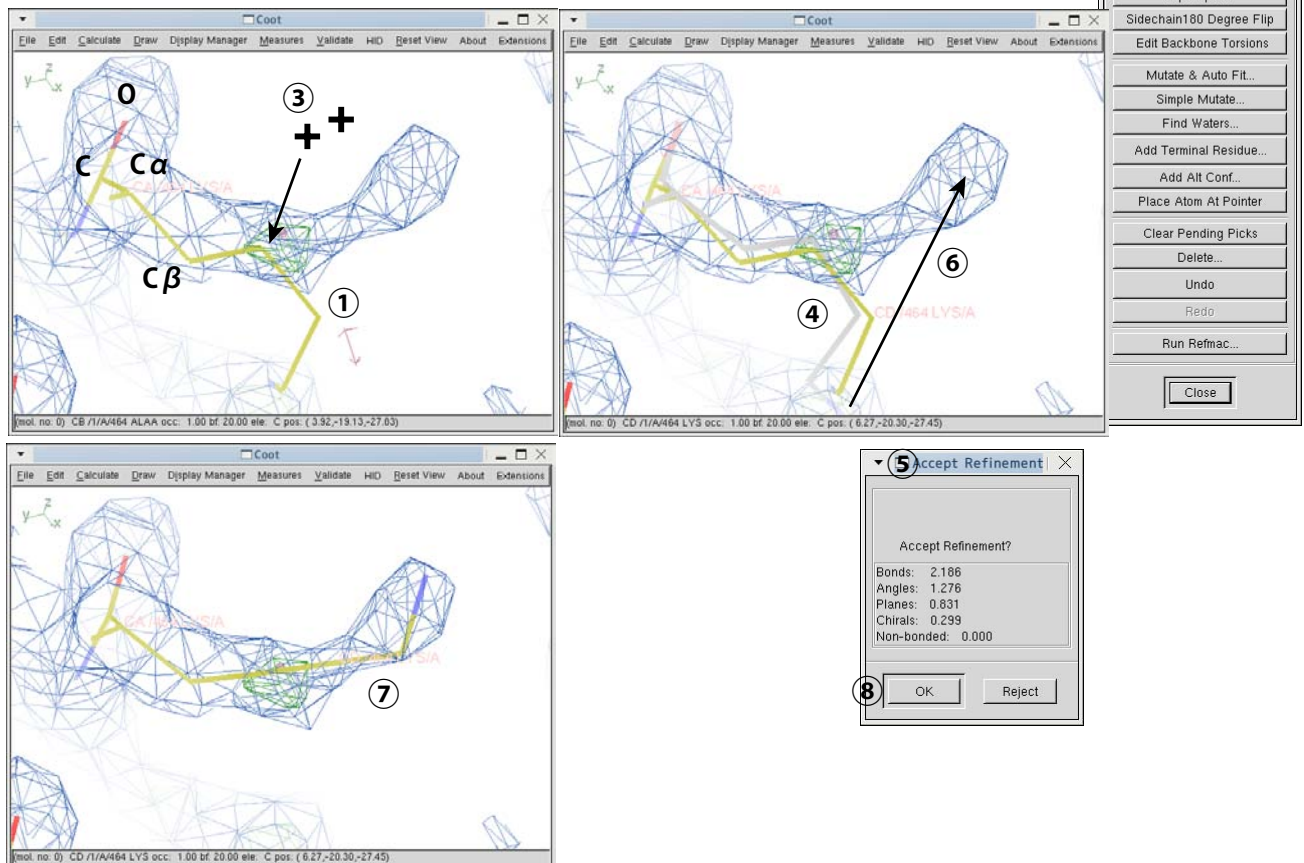
⑧アミノ酸 <TRP (W)> を選択すると、トリプトファンに置換される⑨。



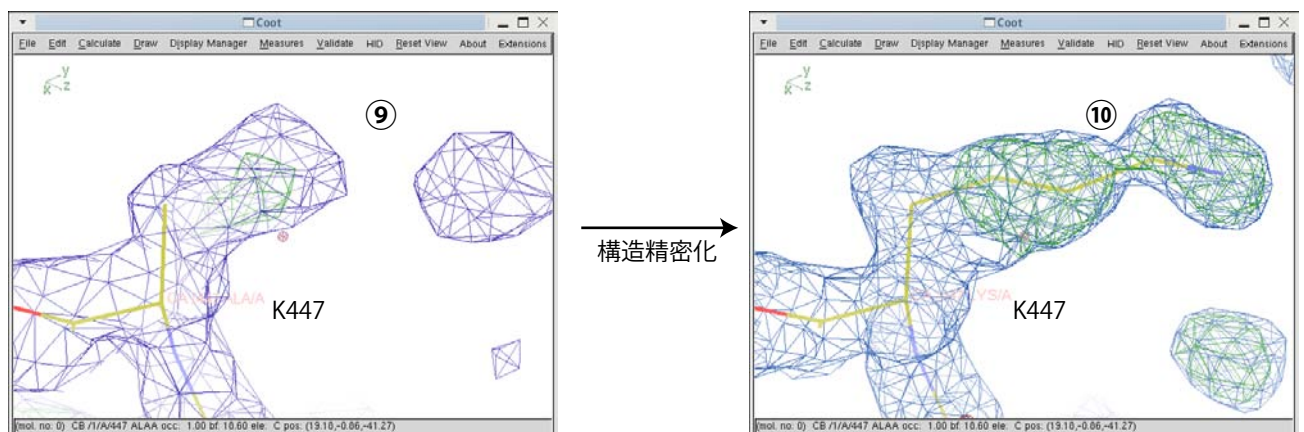


## 構造の修正と構造精密化 ～8～

- ①側鎖の電子密度はしっかり見えているのに、うまく構造ははまらない場合がある。そのときは、
- ②Refine ウィンドウから <Real Space Refine Zone> をクリック。
- ③十字カーソルで、何れかの原子をダブルクリック
- ④電子密度へのフィッティングが行われ、目的の残基がグレーになる。
- ⑤フィッティング結果のウィンドウが現れるが、じゃまなので横にずらす。
- ⑥グレーの残基中の何れかの原子をマウスで電子密度へドラッグする。
- ⑦ドラッグした方向にフィッティングされる。
- ⑧うまく電子密度にフィットできれば <OK> をクリック。

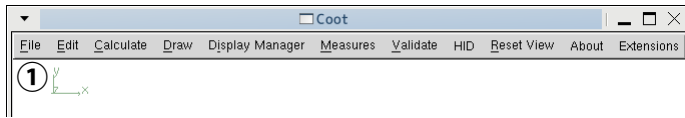


このようにして、アラニンモデルをTRF2 本来のアミノ酸に置き換えていくが、アミノ酸によっては側鎖の電子密度が曖昧で、本来のアミノ酸が当てはまらない様なケースがある。その場合は、とりあえずアラニンのままにしておく。次の構造精密化の後、側鎖の電子密度が確認できるようであれば、アミノ酸を置き換える。下の図は、447 番目のリジンである。分子置換直後は電子密度が曖昧だが⑨、構造精密化後にはしっかりと側鎖の電子密度が確認できる⑩。



## 構造の修正と構造精密化 ～9～

一通り電子密度を確認し、アミノ酸を置き換えたら、構造を保存する。構造を保存したら、COOT を終了する。  
(COOT での作業中、こまめに保存すると良い)



①COOT のメインウィンドウのメニュー、<File><Exit>

この修正した構造をプログラム REFMAC を用いて構造精密化を行う。

②CCP4I のメインウィンドウ左上のプルダウンメニューで <Refinement> を選択し、<Run Refmac> をクリック。

③左上 <Do> のプルダウンメニューで <restrained refinement> を選択。(一般的な構造精密化)

④<MTZ in> は output.mtz

⑤<MTZ out> に自動で output\_refmac1.mtz と入力される場合は、1 回目の精密化の結果が上書きされないよう、キーボードで output\_refmac2.mtz と変更する。(1 を 2 に変えるだけ)

⑥<PDB in> は COOT で最後に保存した構造ファイルを選択。この場合は、model\_4\_molrep1\_refmac1-coot-0.pdb

⑦<PDB out> は model\_4\_molrep1\_refmac2.pdb と自動入力される。自動入力されない場合はキーボードから変更する。

⑧精密化に使用するデータの分解能は 20~2Å とする。

⑨<Run><Run Now> で実行

⑩終了したらログを確認する。

⑪分子置換直後の REFMAC では R/R-free 共に約 42% だったが、今回は R/R-free = 30/34% であり、構造の修正によって大きく改善されている。

The screenshot shows the CCP4 Program Suite interface. The 'Run Refmac5' dialog box is open, showing various settings for the refinement process. The 'Do' dropdown is set to 'restrained refinement'. The 'MTZ in' is 'kast' and 'MTZ out' is 'output\_refmac2.mtz'. The 'PDB in' is 'model\_4\_molrep1\_refmac1-coot-0.pdb' and 'PDB out' is 'model\_4\_molrep1\_refmac2.pdb'. The 'Resolution range' is set from 20.0 to 2.0. The 'Run' button is highlighted with a red circle '9'. The 'CCP4I fileviewer 5\_refmac5.log' window is open, showing the log output. The log displays various statistics, including the Overall R factor (0.3058) and Overall weighted R factor (0.3472), both highlighted with red circles '11'. The log also shows the percentage of used reflections (14763) and the percentage observed (99.2928).

Resolution (Å)	Number of used reflections	Percentage observed	Percentage of free reflections	Overall R factor	Free R factor
0.1324	692	0.687	107	0.599	0.799
0.1448	723	0.684	113	0.549	0.836
0.1572	764	0.704	95	0.599	0.859
0.1696	794	0.736	119	0.601	0.913
0.1819	813	0.713	109	0.571	0.922
0.1943	839	0.761	103	0.601	0.942
0.2067	883	0.771	114	0.663	0.997
0.2191	897	0.706	105	0.601	1.002
0.2314	906	0.730	112	0.676	1.018
0.2438	962	0.706	110	0.615	1.072

REFMAC による構造精密化が終了したら、再び COOT で構造の修正を行う。



## 構造の修正と構造精密化 ~ 10 ~

データディレクトリのターミナルで COOT を起動する。

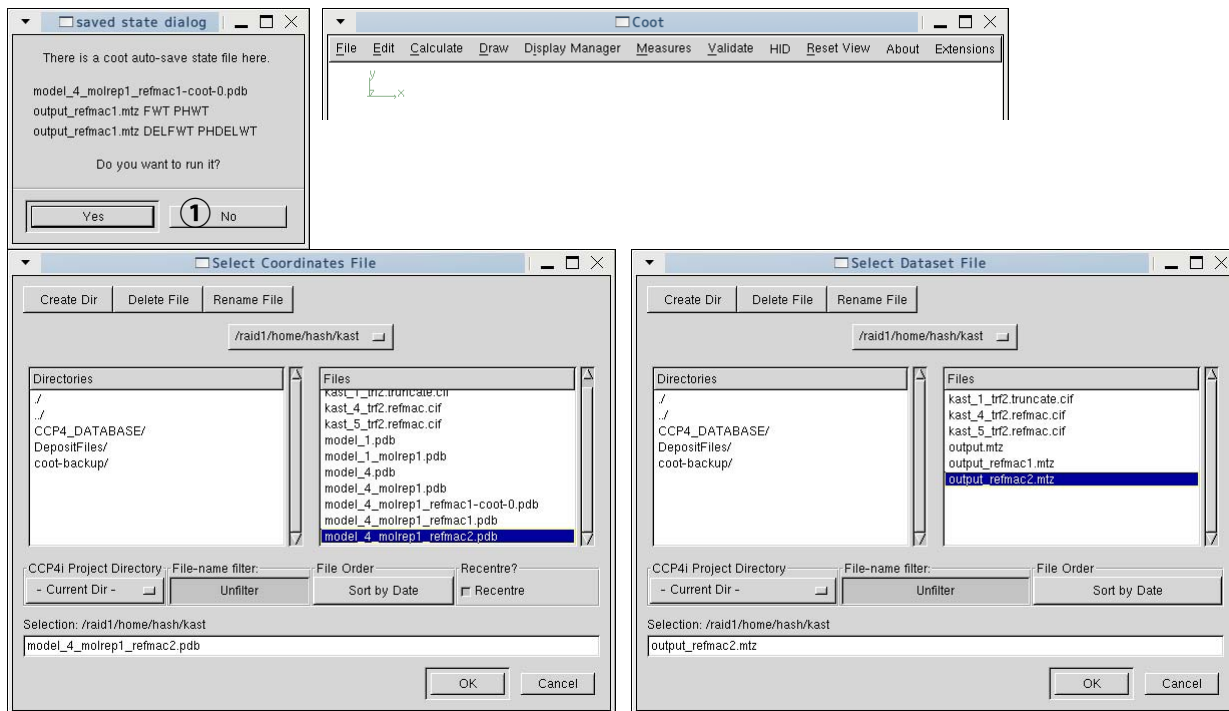
% coot

① COOT を起動すると、前回の保存した状態を復元するかと聞かれるので、通常は <No>。

REFMAC によって精密化された構造と電子密度を読み込む。

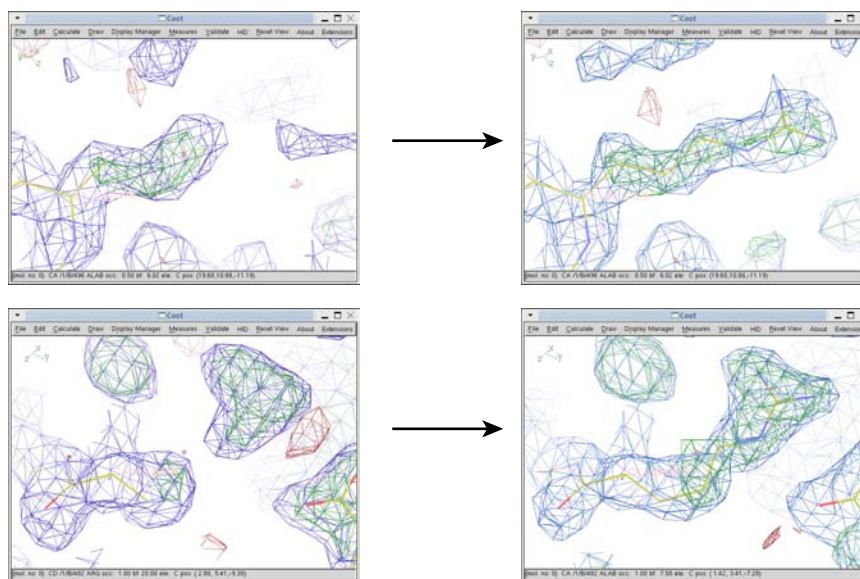
メインウィンドウのメニュー <File><Open Coordinates...>, <File><Auto Open MTZ...>

ここでは、model\_4\_molrep1\_refmac2.pdb と output\_refmac2.mtz



構造と電子密度を読み込んだら、電子密度の濃さを調節する。メインウィンドウのメニュー <Display Manager>

COOT の画面で Space キーを押すと、はじめの残基に移動するので、先と同様に電子密度とアミノ酸配列を見ながら、構造の修正を行う。前回は、曖昧だった電子密度が構造精密化によって改善されているかもしれない。



一通り修正したら、構造を保存し、COOT を終了する。

## 構造の修正と構造精密化 ~ 1 1 ~

再度 REFMAC で構造精密化を行う。今回は最高分解能で精密化する。

- ① CCP4I のメインウィンドウ左上のプルダウンメニューで <Refinement> を選択し、<Run Refmac> をクリック。
- ② 左上 <Do> のプルダウンメニューで <restrained refinement> を選択。(一般的な構造精密化)
- ③ <MTZ in> は output.mtz
- ④ <MTZ out> これまでの精密化の結果が上書きされないよう、キーボードで output\_refmac3.mtz と変更する。
- ⑤ <PDB in> は COOT で最後に保存した構造ファイルを選択。この場合は、model\_4\_molrep1\_refmac2-coot-0.pdb
- ⑥ <PDB out> は model\_4\_molrep1\_refmac3.pdb と自動入力される。自動入力されない場合はキーボードから変更する。
- ⑦ 今回、精密化に使用するデータの分解能は 20~1.25 Å とする。
- ⑧ <Run><Run Now> で実行
- ⑨ 終了したらログを確認する。
- ⑩ 前回の精密化では R/R-free = 30/34% であったが、今回は R/R-free = 28/31% であり、構造の修正と分解能を上げたことにより大きく改善されている。

The screenshot displays the CCP4 Program Suite interface. The main window shows the 'Run Refmac5' dialog box with the following settings:

- Job title: Rigid body refinement
- Do: **restrained refinement** (circled 2)
- MTZ in: **output.mtz** (circled 3)
- MTZ out: **output\_refmac3.mtz** (circled 4)
- PDB in: **model\_4\_molrep1\_refmac2-coot-0.pdb** (circled 5)
- PDB out: **model\_4\_molrep1\_refmac3.pdb** (circled 6)
- Resolution range: 20.0 to **1.25** (circled 7)
- Buttons: **Run** (circled 8), Save or Restore, Close

An overlaid log window titled 'CCP4I fileviewer 8\_refmac5.log' shows the following statistics:

Resolution limits	=	20.000	1.250
Number of used reflections	=	57976	
Percentage observed	=	97.4857	
Percentage of free reflections	=	5.0601	
Overall R factor	=	<b>0.2894</b>	(circled 10)
Free R factor	=	0.3112	
Overall weighted R factor	=	0.2848	
Free weighted R factor	=	0.3076	
Overall correlation coefficient	=	0.9025	
Free correlation coefficient	=	0.8908	
Cricksbanks DPI for coordinate error=	=	0.0639	
DPI based on free R factor	=	0.0649	
Overall figure of merit	=	0.7773	
M based su of positional parameters	=	0.0406	
M based su of thermal parameters	=	0.8951	

Additional log output includes: Time in seconds: CPU = 187.34, Elapsed = 187.00. A table at the bottom shows Rfactor analysis stats vs cycle.

精密化後の構造を COOT を使って修正する。

COOT を起動すると、前回の保存した状態を復元するかと聞かれるので、通常は <No>。

REFMAC によって精密化された構造と電子密度を読み込む。

メインウィンドウのメニュー <File><Open Coordinates...>, <File><Auto Open MTZ...>

ここでは、model\_4\_molrep1\_refmac3.pdb と output\_refmac3.mtz

構造と電子密度を読み込んだら、電子密度の濃さを調節する。メインウィンドウのメニュー <Display Manager>

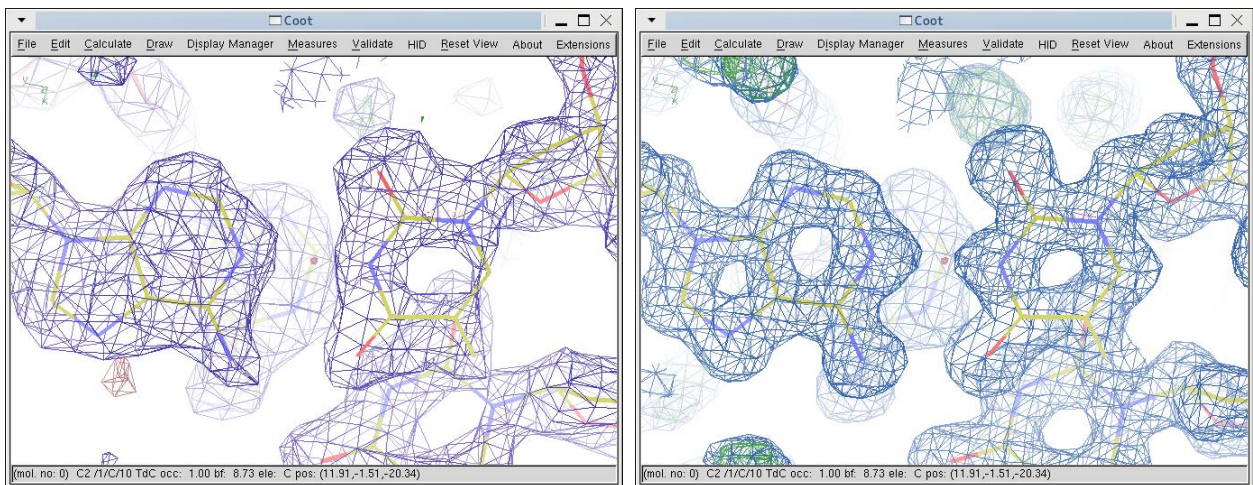
COOT の画面で Space キーを押すと、はじめの残基に移動するので、先と同様に電子密度とアミノ酸配列を見ながら、構造の修正を行う。前回は、曖昧だった電子密度が構造精密化によって改善されているかもしれない。

また、分解能を 2.0 Å から最高分解能 (ここでは 1.25 Å) に上げているので、分解能による電子密度の違いを見ることが出来る。



## 構造の修正と構造精密化 ～12～

最高分解能で精密化を行い、いよいよ構造修正の最終段階である。分解能を上げたことにより、電子密度の見え方が随分異なる。左下は 2.0Å 分解能の電子密度で、右下は 1.25Å 分解能の電子密度である。DNA の AT ペアを表示している。右下ではより原子 1 つ 1 つが識別できるほど高い解像度で構造が精密化されている。



これまで同様、COOT で一通り構造を修正する。その過程で、球状の電子密度がたくさん確認できたはずである①。これはタンパク質や DNA に結合した水和水である。高分解能の構造解析では水和水の構造までも決定できる。

COOT は電子密度から自動で水分子をピックできる。

- ② Refine ウィンドウ <Find Waters ...> をクリック。
- ③ 対象とする電子密度を選択。DELFWT PHDELWT (緑色の電子密度 FoFc) を使う。
- ④ 電子密度のレベル(濃さ)を指定する。FoFc なので、2.0~3.0 程度で良いだろう。
- ⑤ <Find Waters> をクリック
- ⑥ 自動で水分子をアサインしていくれる。
- ⑦ 水分子にしては大きな電子密度があると教えてくれるが、今回は無視する。<Dismiss> をクリック。

構造を保存して、COOT を終了する。

## 構造の修正と構造精密化 ～13～

構造精密化の最終段階では水分子を含めた構造精密化を行う。

- ①REFMAC ウィンドウ左上 <Do> のプルダウンメニューで <restrained refinement> を選択。(一般的な構造精密化)
- ②<MTZ in> は output.mtz
- ③<MTZ out> これまでの精密化の結果が上書きされないよう、キーボードで output\_refmac4.mtz と変更する。
- ④<PDB in> は COOT で最後に保存した構造ファイルを選択。この場合は、model\_4\_molrep1\_refmac3-coot-0.pdb
- ⑤<PDB out> は model\_4\_molrep1\_refmac4.pdb と自動入力される。自動入力されない場合はキーボードから変更する。
- ⑥精密化に使用するデータの分解能は 20~1.25 Å とする。
- ⑦今回の精密化では、異方性温度因子を導入する。
- ⑧<Run><Run Now> で実行
- ⑨終了したらログを確認する。
- ⑩前回の精密化では R/R-free = 28/31% であったが、今回は R/R-free = 19/24% であり、水分子の追加と異方性温度因子の導入によって、十分精密化されている。

The screenshot shows the Run Refmac5 window on the left and the CCP4I fileviewer on the right. The Run Refmac5 window has several sections with numbered callouts:

- Job title:** Rigid body refinement
- Do:** ① restrained refinement (selected), using no prior phase information, input
- Input fixed TLS parameters:** unchecked
- MTZ in:** kast, ② output.mtz
- MTZ out:** kast, ③ output\_refmac4.mtz
- PDB in:** kast, ④ model\_4\_molrep1\_refmac3-coot-0.pdb
- PDB out:** kast, ⑤ model\_4\_molrep1\_refmac4.pdb
- Library:** kast
- Refinement Parameters:** Do 10 cycles of maximum likelihood restrained refinement. Resolution range from minimum 20.0 to 1.25 (⑥). Refine ⑦ anisotropic and temperature factors.
- Buttons:** ⑧ Run, Save or Restore, Close

The CCP4I fileviewer shows the following statistics:

Resolution limits	=	20.000	1.250
Number of used reflections	=	57976	
Percentage observed	=	97.4857	
Percentage of free reflections	=	5.0601	
Overall R factor	=	0.1994	⑩
Free R factor	=	0.2422	
Overall weighted R factor	=	0.1941	
Free weighted R factor	=	0.2447	
Overall correlation coefficient	=	0.9438	
Free correlation coefficient	=	0.9185	
Cruickshanks DPI for coordinate error	=	0.0582	
DPI based on free R factor	=	0.0584	
Overall figure of merit	=	0.8641	
ML based su of positional parameters	=	0.0298	
ML based su of thermal parameters	=	1.4879	

Time in seconds: CPU = 210.24, Elapsed = 211.00

\*\*\*\* Things for loggraph, R factor and others vs cycle \*\*\*\*

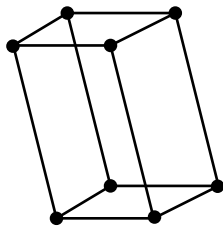
TABLE: Rfactor analysis, stats vs cycle :  
\$RAPHIS:<rfactor> vs cycle :N:1,2,3:  
:FOM vs cycle :N:1,4:  
:LL vs cycle :N:1,5:  
:LLfree vs cycle :N:1,6:  
:Geometry vs cycle:N:1,7,8,9,10,11:

構造を評価するには、精密化の信頼度因子である R 値以外に、立体化学的な検証が必要であり、本来は Ramachandran プロットや理想的なペプチド結合からのずれなどもチェックしながら構造の修正を進めていくが、今回の実習では割愛した。最後に精密化した構造を COOT で表示し、タンパク質と DNA がどのような相互作用で結合しているかを観察して終了する。

お疲れ様でした。

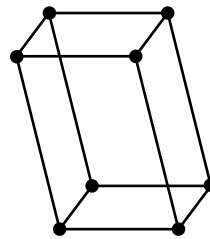
# 補足資料

## 結晶格子 (ブラベ格子)

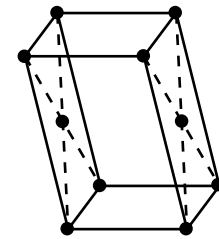


P

Triclinic  
No Symmetry  
 $a \neq b \neq c$   
 $\alpha \neq \beta \neq \gamma \neq 90^\circ$

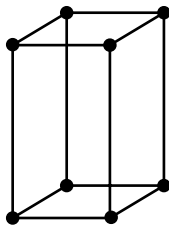


P

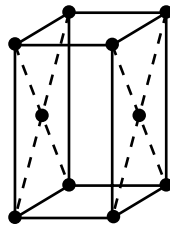


C

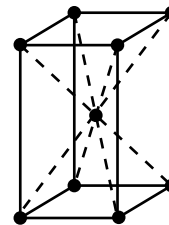
Monoclinic  
2-fold axis along b  
 $a \neq b \neq c$   
 $\alpha = \gamma = 90^\circ, \beta \neq 90^\circ$



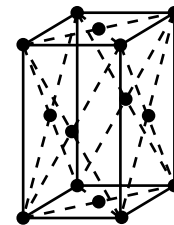
P



C

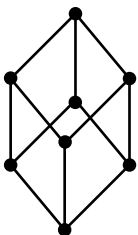


I



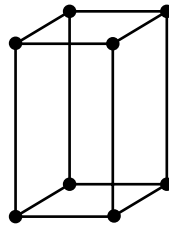
F

Orthorhombic  
Three mutually perpendicular 2-fold axes  
 $a \neq b \neq c$   
 $\alpha = \beta = \gamma = 90^\circ$

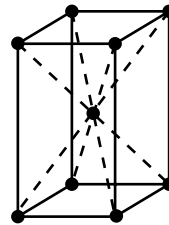


R

Rhombohedral  
3-fold axis along one body diagonal  
 $a = b = c$   
 $\alpha = \beta = \gamma \neq 90^\circ$

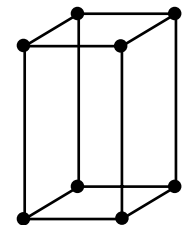


P



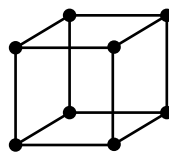
I

Tetragonal  
4-fold axis along c  
 $a = b \neq c$   
 $\alpha = \beta = \gamma = 90^\circ$

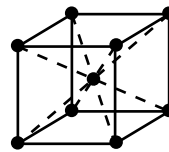


P

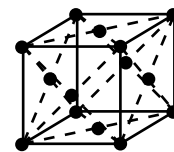
Hexagonal  
6-fold axis along c  
 $a = b \neq c$   
 $\alpha = \beta = 90^\circ, \gamma = 120^\circ$



P



I



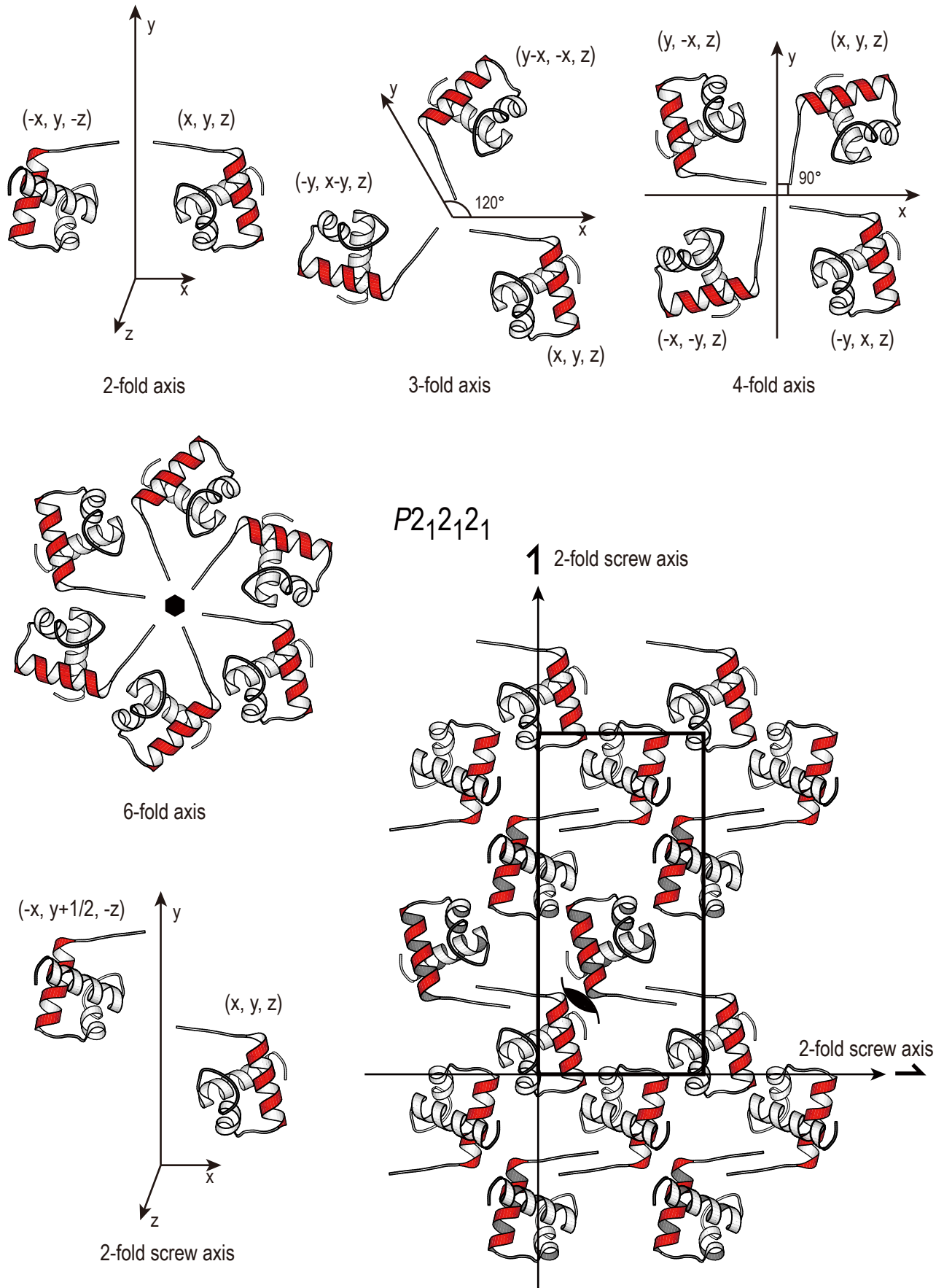
F

Cubic  
3-fold axes along all body diagonals  
 $a = b = c$   
 $\alpha = \beta = \gamma = 90^\circ$



# 補足資料

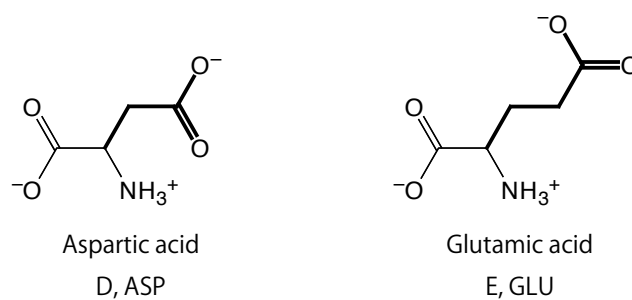
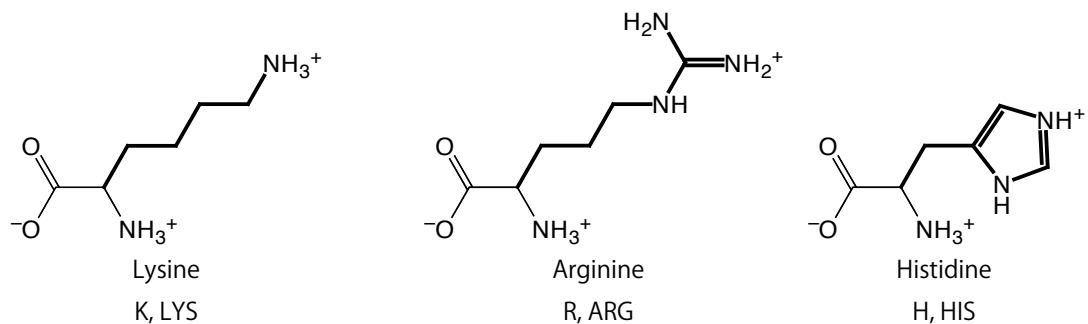
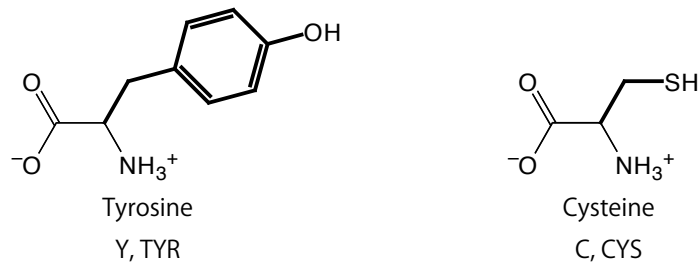
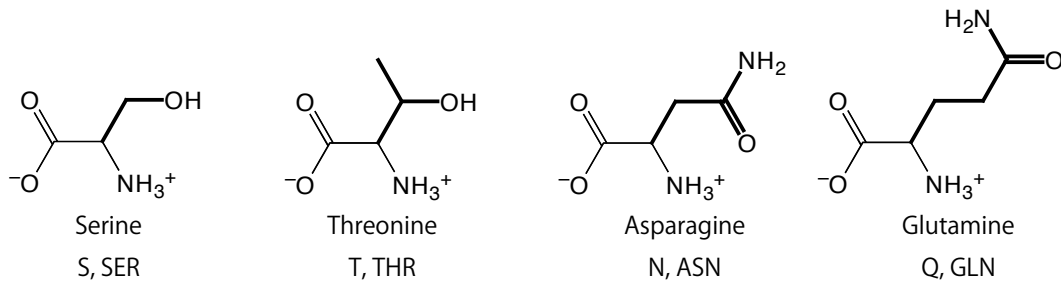
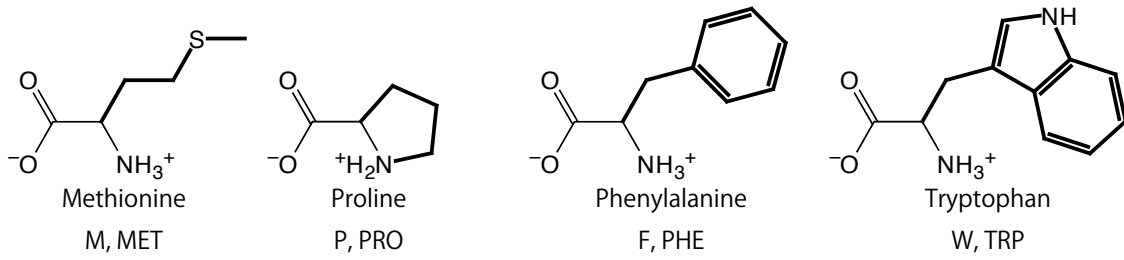
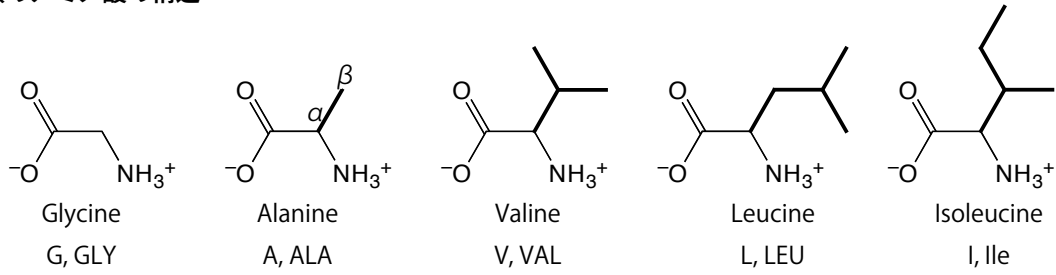
## 結晶の回転対称





## 補足資料

### 20 種類のアミノ酸の構造



# 補足資料

## DNA の構造

