

COOTを使ったモデル構築と構造精密化

担当 橋本

プログラムCOOTはPaul Emsley氏によって開発されたモデル構築ソフトである。入手法や設定方法は以下のサイトを参照。Windows版、OSX版もある。フリーソフトである。使いやすいGUIと強力なフィッティング機能には定評がある。
<http://www.yesbl.york.ac.uk/~emsley/cool/>

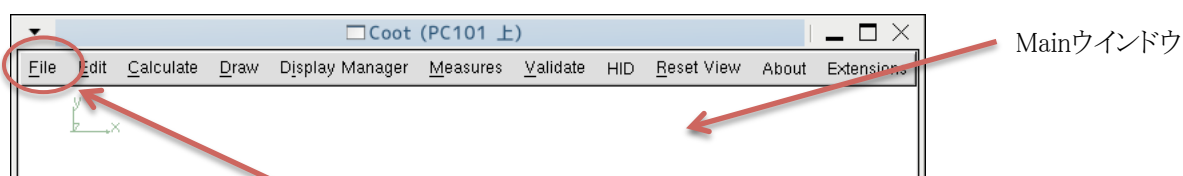


初期モデルの表示

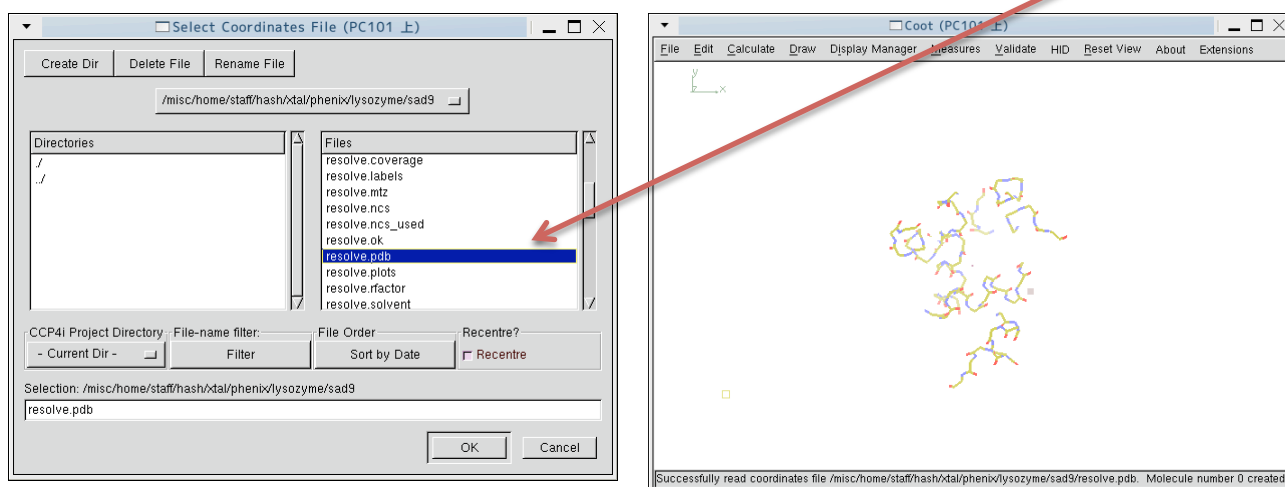
RESOLVEによるSAD法で得られた初期モデルを表示してみる。COOTの起動はコマンドプロンプトで

```
$ coot
```

と入力すればよい。COOTが起動すると、初めにMAINウィンドウが表示される。



Mainウィンドウの[File] => [Open Coordinates...]をクリックすると、選択ウィンドウが開くので、**resolve.pdb**を選択してOKをクリックするとRESOLVEが構築したリゾチーム分子の初期モデルが表示される。



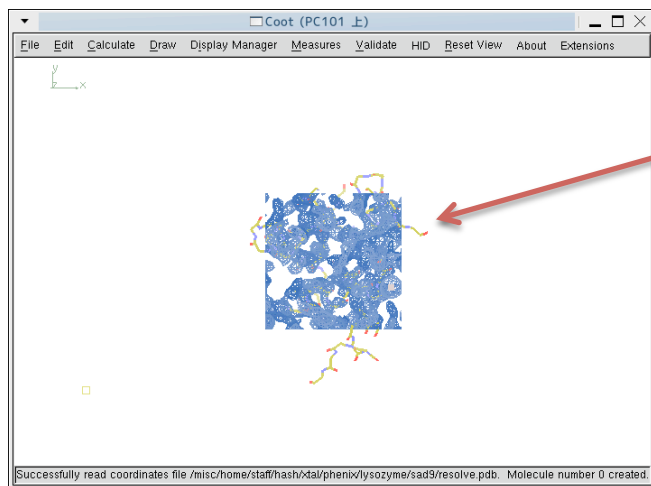
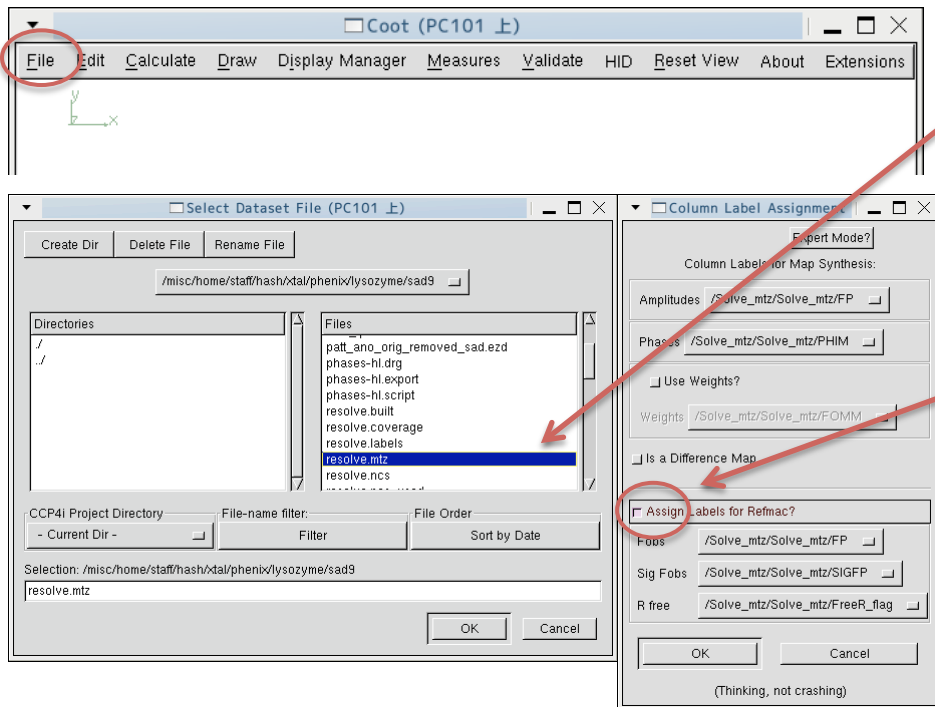
炭素原子は黄色、窒素原子は青、酸素原子は赤、硫黄原子は緑で表現されている。

マウスの操作法

左クリックでマウスを動かすと分子が回転する。
中クリックした原子が画面の中心になる。
右クリックでマウスを動かすと分子が拡大・縮小される。
Ctrlを押しながら左クリックでマウスを動かすと分子が並進する。
Ctrlを押しながら右クリックでマウスを動かすと画面の奥行きが変わる。
ホイールを回転させると電子密度のレベルが変わる。
原子をダブルクリックすると残基番号、残基の種類が表示される。
(シフト+シングルクリックでも同様)

初期電子密度の表示

Mainウインドウから[File] => [Open MTZ, mmCIF, fcf or phs...]をクリックし、選択ウインドウでlysozyme.mtzを選択するとラベルの選択ウインドウが表示される。

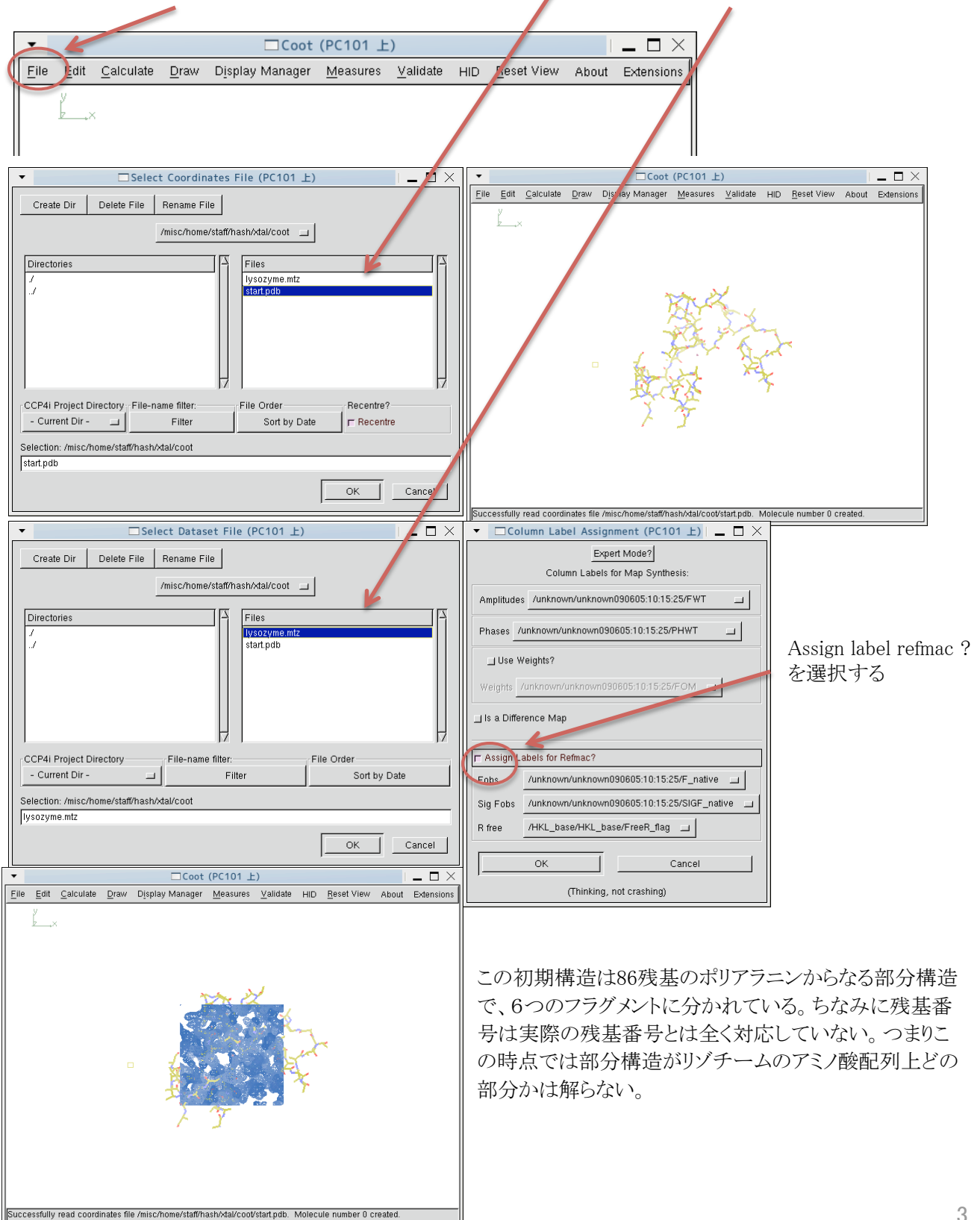


実習用の初期構造と電子密度の表示

今回SOLVE/RESOLVEによって得られた電子密度は分子の外形はくっきりと確認できるが、詳細な構造を構築するには時間を要する。そこで、今回の実習では比較的きれいな電子密度を用いて、リゾチームの部分構造からのモデリングを行う。

Mainウインドウの[File] => [Open Coordinates...]をクリック、start.pdbを選択。

Mainウインドウから[File] => [Open MTZ, mmCIF, fcf or phs...]をクリックし、lysozyme.mtzを選択。



The figure consists of several screenshots from the Coot software interface, illustrating the process of loading data and setting up for refinement:

- Top Left:** The main Coot window with the 'File' menu highlighted by a red circle and an arrow pointing to the 'Open Coordinates...' option.
- Middle Left (Top):** A 'Select Coordinates File' dialog box showing the file 'start.pdb' selected in the 'Files' list.
- Middle Left (Bottom):** A 'Select Dataset File' dialog box showing the file 'lysozyme.mtz' selected in the 'Files' list.
- Middle Right:** The main Coot window displaying a 3D molecular model of lysozyme in stick representation, overlaid on a blue electron density map.
- Bottom Left:** The main Coot window showing the same molecular model and density map, but with a different view or zoom level.
- Bottom Right:** A 'Column Label Assignment' dialog box. The 'Assign Labels for Refmac?' checkbox is checked and circled in red. A red arrow points from the text 'Assign label refmac? を選択する' to this checkbox. Other options like 'Use Weights?' and 'Is a Difference Map' are unchecked.

Assign label refmac ?
を選択する

この初期構造は86残基のポリアラニンからなる部分構造で、6つのフラグメントに分かれている。ちなみに残基番号は実際の残基番号とは全く対応していない。つまりこの時点では部分構造がリゾチームのアミノ酸配列上での部分かは解らない。

フラグメント1	1114~1129	16残基
フラグメント2	1144~1158	15残基
フラグメント3	1300~1314	15残基
フラグメント4	1422~1432	11残基
フラグメント5	1501~1513	13残基
フラグメント6	1623~1638	16残基

リゾチームのアミノ酸配列

> Lysozyme - chicken egg white - 129 aa, Molecular Weight 14 kDa

10	20	30	40	50
KVFGRCELAA	AMKRHGLDNY	RGYSLGNWVC	AAKFESNFNT	QATNRNTDGS
60	70	80	90	100
TDYGILQINS	RWWCNDGRTP	GSRNLCNIPC	SALLSSDITA	SVNCAKKIVS
110	120			
DGNMNAWVA	WRNRCKGTDV	QAWIRGCRL		

本実習の流れ

リゾチームの部分構造(ポリアラニンモデル) start.pdbを元に、これがリゾチーム分子のどの部分に相当するかを考え、正しいアミノ酸に置換する。



正しい残基番号に変更する。



構築されていない部分にアミノ酸を挿入し、リゾチーム分子を完成させる。



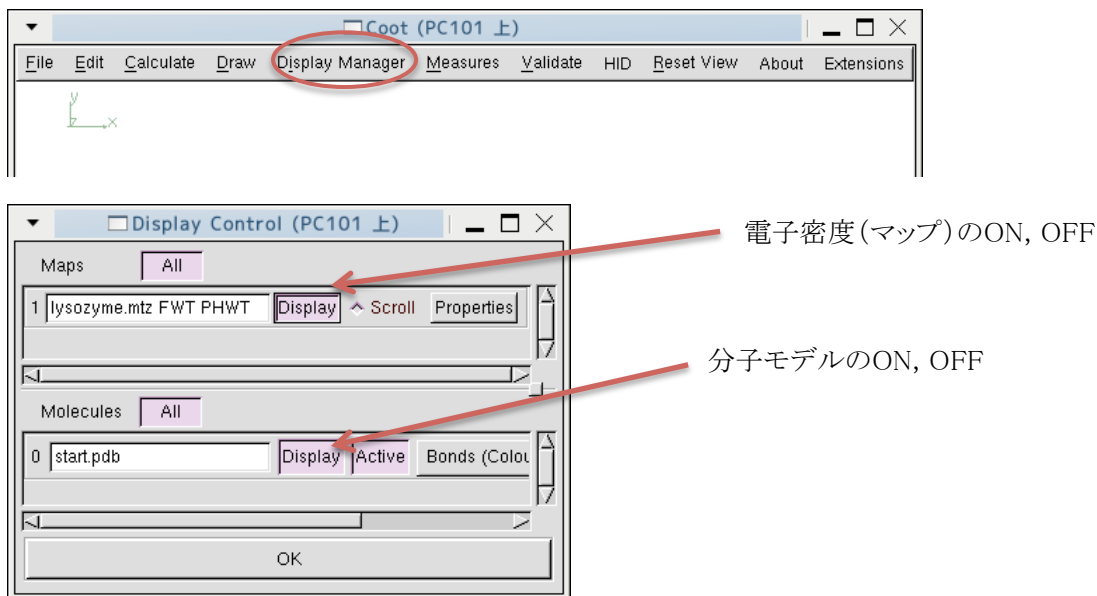
構造精密化を行い、水分子をアサイン。



再度構造精密化を行い、最終構造を得る。

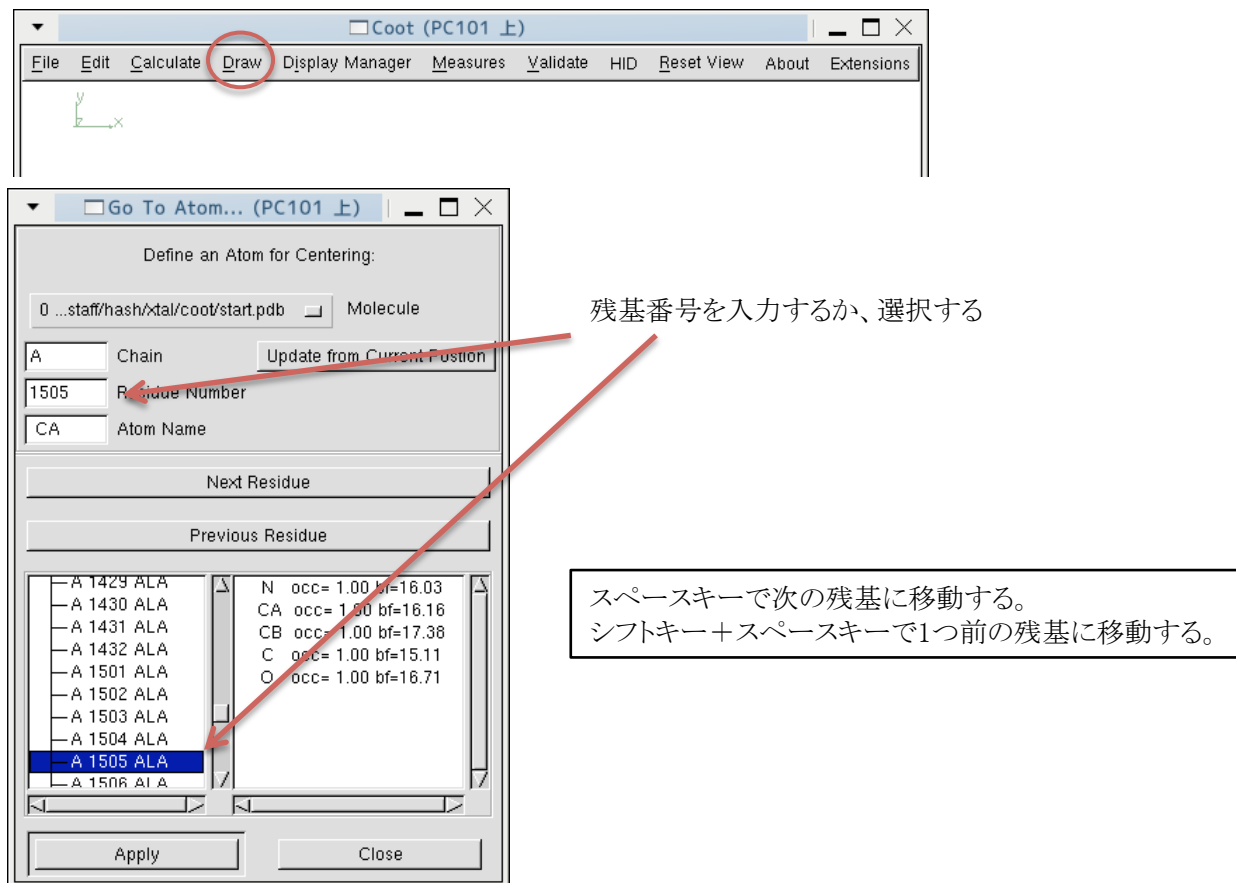
表示のONとOFF

Mainウインドウから[Display Manager]を選択し、Display Controlウインドウで、Displayボタンによって分子や電子密度の表示のON, OFFを行う。



特定の残基に移動する

Mainウインドウから[Draw] => [Go to Atom...]を選択し、Go To Atomウインドウで、移動したい残基番号を入力、あるいは選択し、Apply。その残基のC α 原子が画面の中心になる。



対称分子の表示

結晶中では分子が規則的に並んでいる。モデル構築している分子の周囲にも(結晶学的対称によって関係づけられた)同じ分子が存在している。隣の分子の電子密度を区別するためには、対称分子を表示するとわかりやすい。

Mainウィンドウで、[Draw] => [Cell & Symmetry...]を選択。

Symmetry by Moleculeをクリックすると対称分子の表示法を選択するウィンドウが現れる。

Display Near Chainsを選択しOK

結晶格子も表示するにはここをYes

最後にApplyをクリックすると対称分子が灰色で表示される。

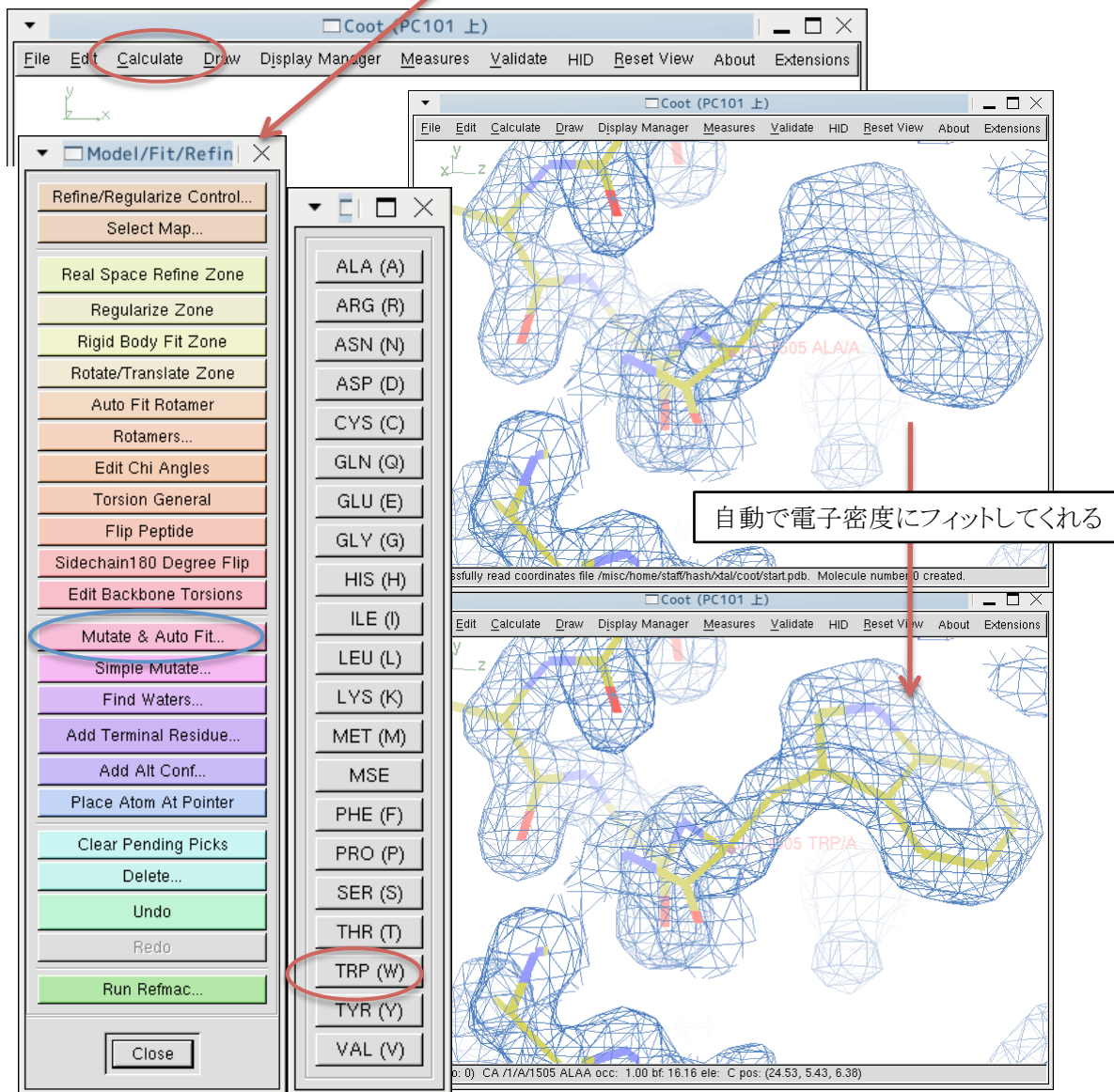
接近した電子密度に間違えてトレースしないように注意

アミノ酸残基の置換と電子密度へのフィッティング

Mainウインドウから[Calculate] => [Model/Fit/Refine]を選択し、Model/Fit/Refineウインドウを表示する。あるいは、F5キーでもこのMenuウインドウは現れる。

例として、1505に移動して、電子密度から正しいアミノ酸を推定し置換してみる。

Menuウインドウで、[Mutate&AutoFit...]を選択すると、マウスカーソルが十字になる。マップを選択するウインドウが現れる場合はOKを選択して、再び[Mutate&AutoFit...]を選択。変換したい残基の原子をクリック。アミノ酸を選択するウインドウが現れるので、正しいと思われるアミノ酸を選択すると、アミノ酸の置換と電子密度へのフィッティングを行ってくれる。



ヒント

特徴のある電子密度に注目する。芳香族残基は推測しやすい。

20種類のアミノ酸の構造を参考に側鎖を推測し、置換してみる。

間違っていればフィットしないので、正しいか間違っているかは比較的容易に判別できる。

通常、水素原子は無視する。

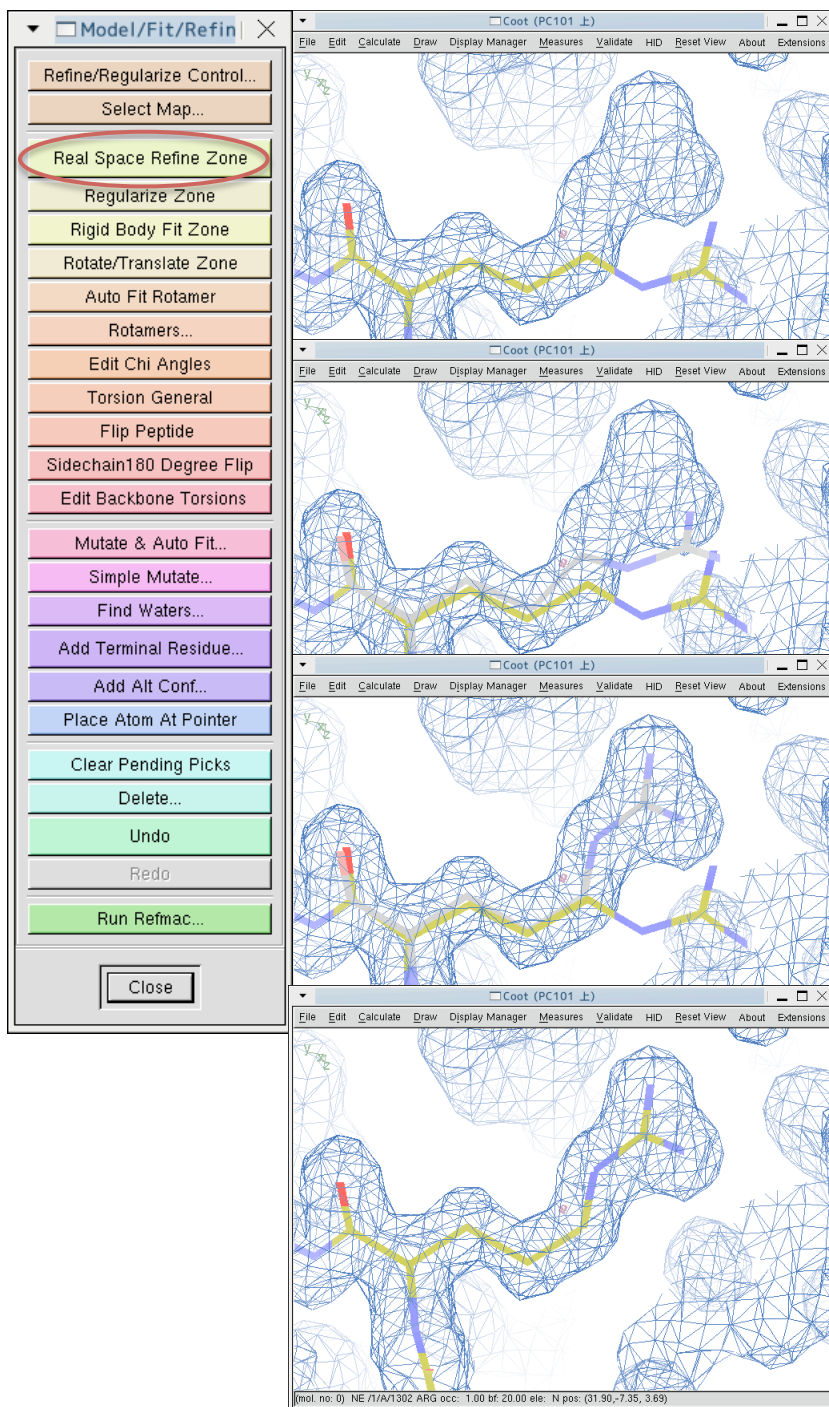
電子密度からだけでは、グルタミン酸とグルタミン、アスパラギン酸とアスパラギンを区別することは難しい。アミノ酸配列から推測するしかない。

アミノ酸残基の置換と電子密度へのフィッティング(2)

[Mutate&AutoFit]でうまく側鎖が電子密度にフィット出来ない場合、あるいは主鎖がうまく電子密度にフィットしていない場合、[RealSpaceRefineZone]で側鎖や主鎖をフィッティングさせることも可能である。

[RealSpaceRefineZone]をクリックすると、マウскарソルが十字になるので、フィットさせたい残基間の原子を2つ選択する。同一残基内で2つ選択するとその残基のみでフィッティングを行う。主鎖のフィッティングを行うときは3残基くらいの範囲でフィッティングさせると良い。

また、残基を選択した状態(白い状態)では、マウスでドラッグすることで電子密度へのフィッティングが容易になる。



電子密度と構造がずれている

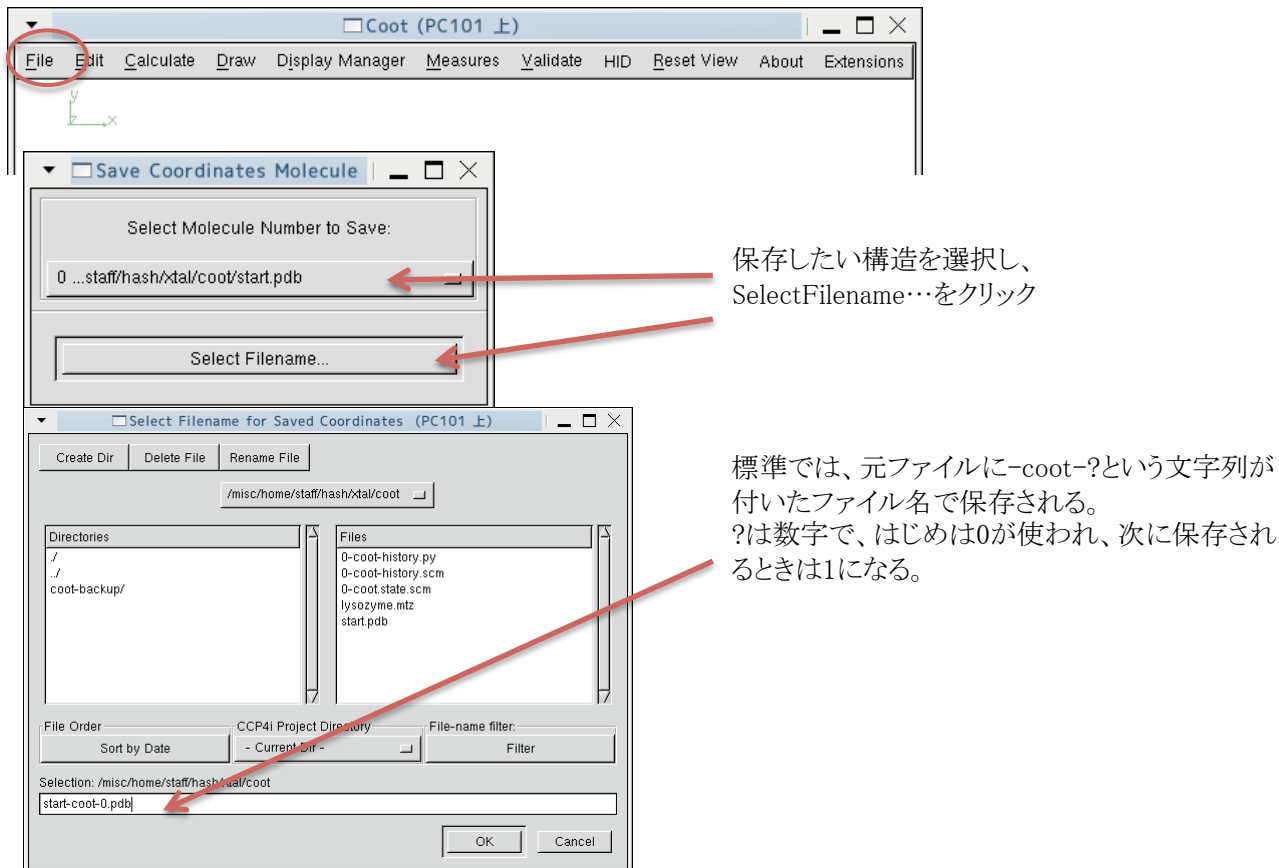
[RealSpaceRefineZone]を実行
大抵はこれでフィッティングされる

うまくいかない場合は、残基がアクティブな状態でマウスで電子密度にドラッグするとフィッティングされる。

構造データの保存

構築した構造をPDBフォーマットでファイルに保存する。モデリングソフトウェアがダウンすることもあるので、頻繁に保存することを勧める。

Mainウィンドウで、[File] => [SaveCoordinates...]を選択。



PDBフォーマットについて

ProteinDataBank (PDB) で使われている構造ファイルのフォーマット。原子座標が記述されている。

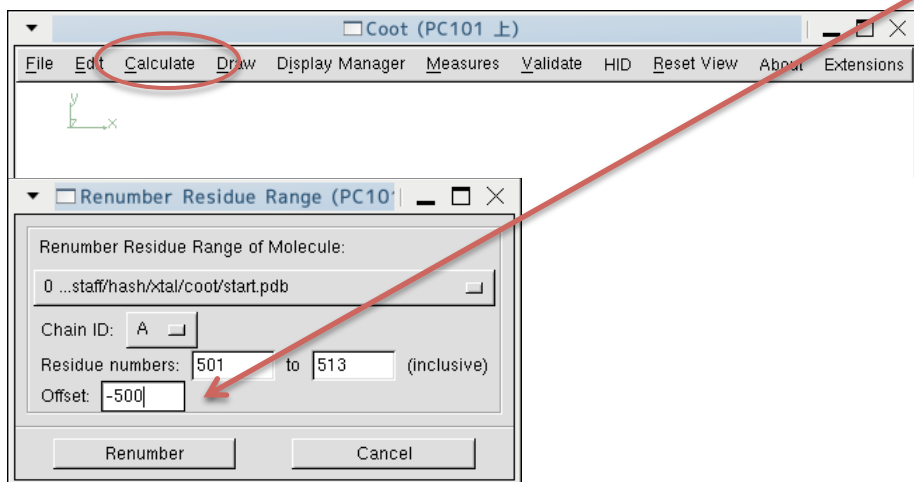
格子定数、空間群

CRYST1	79.168	79.168	37.995	90.00	90.00	90.00	P 43 21 2			
ATOM	1	N	LYS A	1	10.181	-3.246	0.782	1.00	19.88	N
ATOM	2	CA	LYS A	1	10.601	-2.321	-0.331	1.00	20.19	C
ATOM	4	CB	LYS A	1	10.178	-0.879	0.002	1.00	21.32	C
ATOM	7	CG	LYS A	1	10.629	0.144	-0.939	1.00	22.58	C
ATOM	10	CD	LYS A	1	10.105	1.498	-0.548	1.00	24.42	C
ATOM	13	CE	LYS A	1	10.713	2.513	-1.510	1.00	29.54	C
ATOM	16	NZ	LYS A	1	9.912	3.727	-1.535	1.00	32.57	N
ATOM	20	C	LYS A	1	12.087	-2.413	-0.528	1.00	19.42	C
ATOM	21	O	LYS A	1	12.847	-2.440	0.453	1.00	19.62	O

原子の種類 残基の種類 残基番号 原子座標(x, y, z) 占有率 温度因子

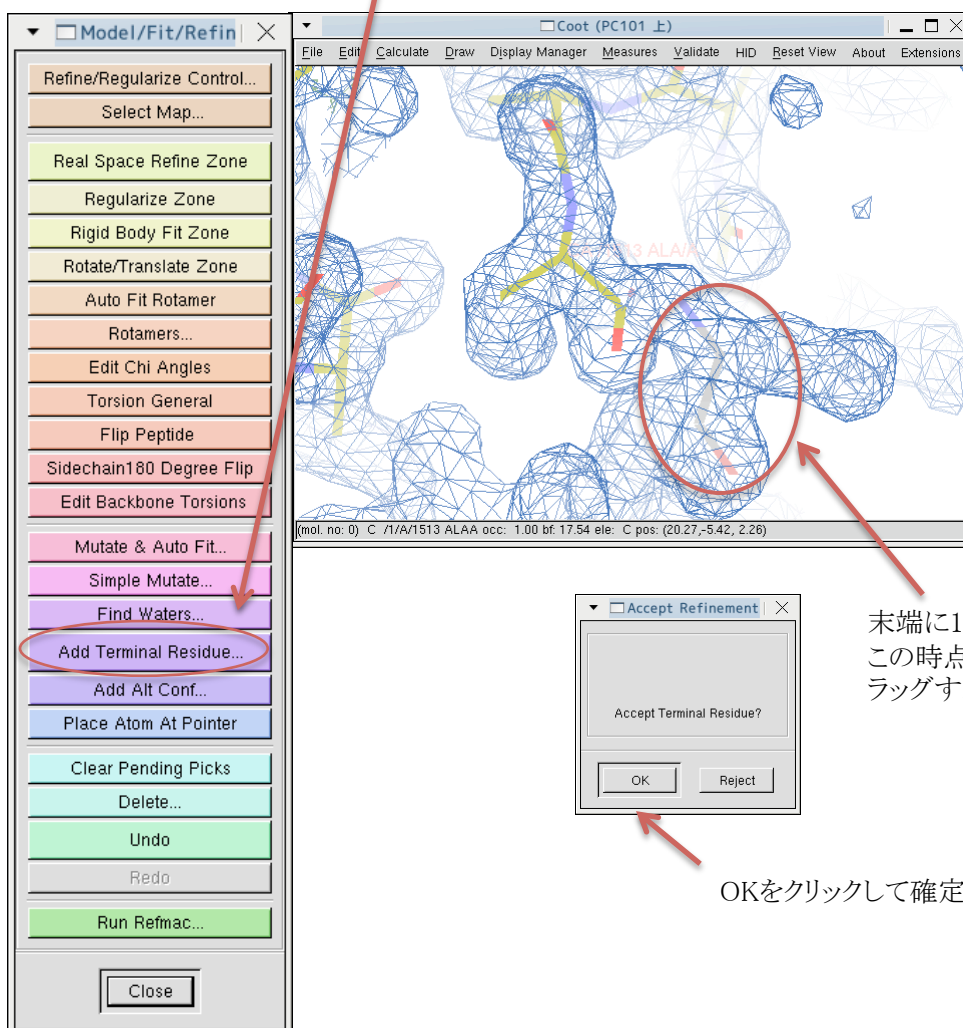
残基番号の変更

フラグメントがリゾチーム分子のどこに相当するかがわかったら、正しいアミノ酸番号に変更する。Mainウィンドウで、[Calculate] => [RenumberResidues...]を選択。例えば、501-513のフラグメントを1番から番号をふりたいときにはOffsetには-500を入力する。



アミノ酸の挿入

Model/Fit/Refineウィンドウで、[AddTerminalResidue...]をクリックするとマウスカーソルが十字なるので、末端の残基の原子をクリックすると1残基のアラニンが挿入される。C末端に挿入すれば残基番号が1つ増え、N末に挿入すれば残基番号が1つ小さくなる。挿入後、電子密度をもとに[Mutate&AutoFit]で正しいアミノ酸に置換し、電子密度にフィットさせる。

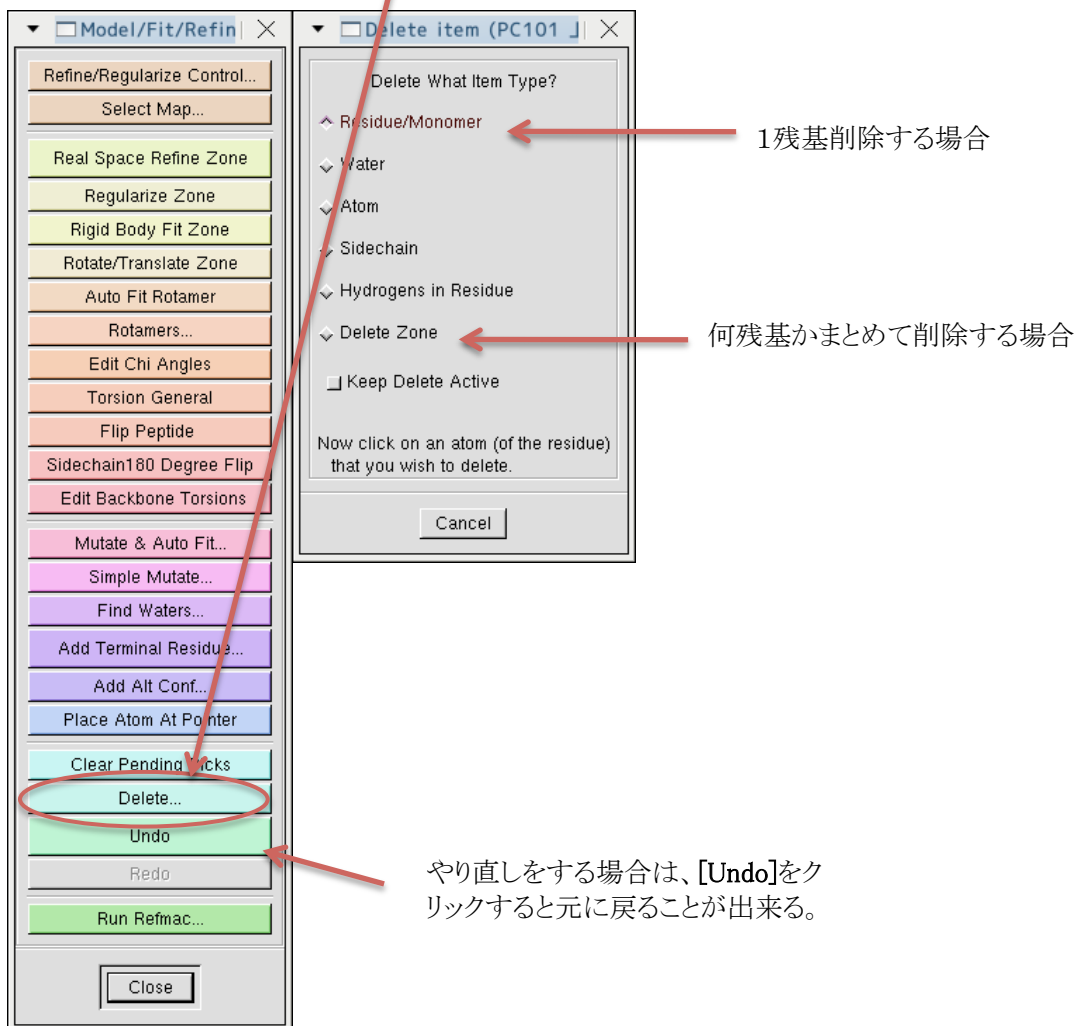


末端に1アミノ酸が挿入される。この時点では白いので、マウスでドラッグすれば、変更可能である。

OKをクリックして確定

残基の削除

Model/Fit/Refineウィンドウで、[Delete]をクリックすると、何を削除するかを選ぶウィンドウが現れる。選択すると、マウスカーソルが十字になるので、削除したいものをクリックすれば削除される。



構造精密化

リゾチーム分子が構築できたら、構造精密化を行う。本実習ではREFMAC (CCP4)を使用する。Model/Fit/Refineウインドウで、**[RunRefmac]**をクリックし、現れたウインドウで**[RunRefmac]**をクリックするとREFMACが実行される。COOTを起動しているターミナルで、計算の進行状況をモニタできる。

The screenshot displays the COOT software interface during a refinement process. On the left is the 'Model/Fit/Refin' menu, with 'Run Refmac...' highlighted. The top window, 'Run Refmac (PC101 上)', shows the 'Run Refmac' button circled in red. The main window, 'Coot (PC101 上)', shows a 3D model of a protein structure with electron density maps. Annotations indicate that yellow structures are pre-refinement and green structures are post-refinement. A purple map is identified as a $2m|Fo|-D|Fc$ map, and a green map as an $m|Fo|-D|Fc$ map. The bottom window, 'Display Control (PC101 上)', lists the maps and molecules, with 'Run Refmac...' circled in red. The 'Maps' list includes: 1 lysozyme.mtz FWT PHWT, 3 start-coot-0_refmac0.mtz FWT PHWT, and 4 start-coot-0_refmac0.mtz DELFWT PHDELWT. The 'Molecules' list includes: 0 start-coot-0.pdb and 2 start-coot-0_refmac0.pdb.

黄色の構造が精密化前、緑の構造が精密化後

紫のマップが $2m|Fo|-D|Fc$ マップ
緑のマップが $m|Fo|-D|Fc$ マップ

精密化後の $2m|Fo|-D|Fc$ マップ
精密化後の $m|Fo|-D|Fc$ マップ

精密化後の構造

精密化前後の電子密度と構造が読み込まれている。混乱を避けるために、一度cootを終了する。精密化後の構造はディレクトリcoot-refmac/の中にある。精密化後の構造の信頼性を表す統計値はPDBファイルに書き込まれているので、gedit等で開いてみる。

構造精密化(2)

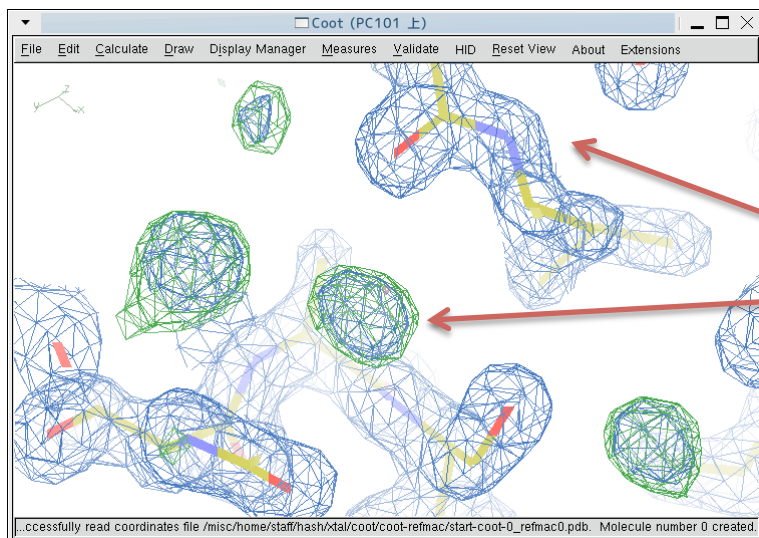
精密化後の構造ファイル

```
.....  
REMARK 3 DATA USED IN REFINEMENT.  
REMARK 3 RESOLUTION RANGE HIGH (ANGSTROMS) : 1.60  
REMARK 3 RESOLUTION RANGE LOW (ANGSTROMS) : 19.01  
REMARK 3 DATA CUTOFF (SIGMA (F)) : NONE  
REMARK 3 COMPLETENESS FOR RANGE (%) : 100.00  
REMARK 3 NUMBER OF REFLECTIONS : 15593  
REMARK 3  
REMARK 3 FIT TO DATA USED IN REFINEMENT.  
REMARK 3 CROSS-VALIDATION METHOD : THROUGHOUT  
REMARK 3 FREE R VALUE TEST SET SELECTION : RANDOM  
REMARK 3 R VALUE (WORKING + TEST SET) : 0.15960  
REMARK 3 R VALUE (WORKING SET) : 0.15872  
REMARK 3 FREE R VALUE : 0.17658  
.....  
  
REMARK 3 RMS DEVIATIONS FROM IDEAL VALUES COUNT RMS WEIGHT  
REMARK 3 BOND LENGTHS REFINED ATOMS (A): 1025 ; 0.020 ; 0.021  
REMARK 3 BOND LENGTHS OTHERS (A): 708 ; 0.003 ; 0.020  
REMARK 3 BOND ANGLES REFINED ATOMS (DEGREES): 1389 ; 1.760 ; 1.903
```

統計値を確認したら、ファイルを閉じて、再びcootを起動する。

先と同様に、構造は Main ウィンドウの[OpenCoordinates...], 電子密度は[OpenMTZ, mmCIF, fcf or phs...]から読み込む。

読み込むファイルは、_refmac0が付いたファイルである。この例では、PDBファイルはstart-coot-0_refmac0.pdb、MTZファイルはstart-coot-0_refmac0.mtzである。



青のマップが $2m|Fo|-D|Fc|$ マップ
緑のマップが $m|Fo|-D|Fc|$ マップ
この緑の電子密度は水分子であると考えられる。

電子密度について

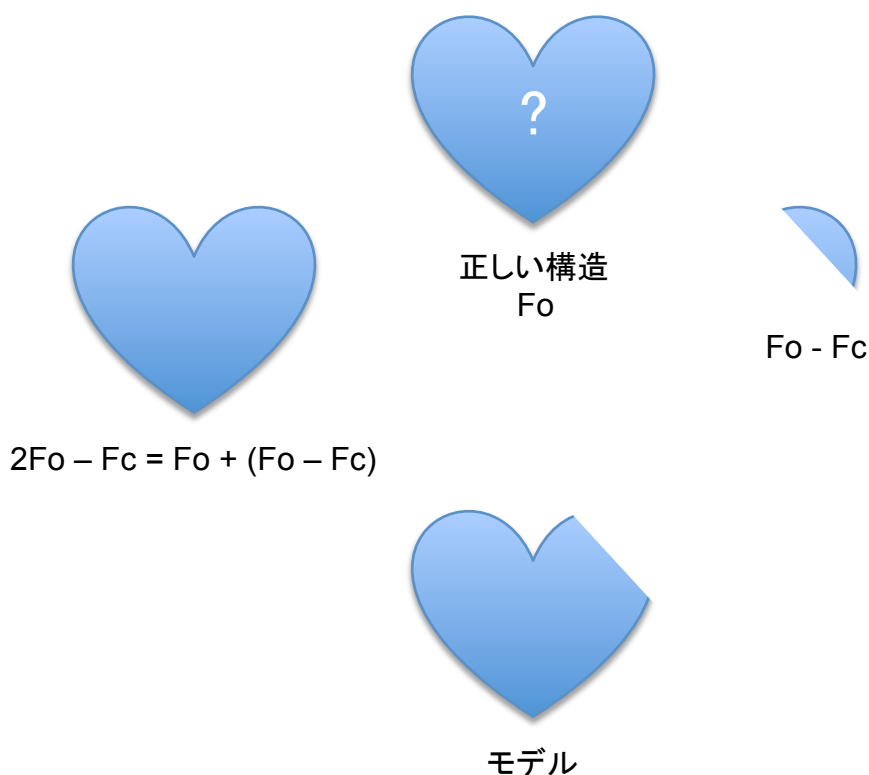
よく使われる電子密度は $2|F_o|-|F_c|$ を係数とした電子密度と $|F_o|-|F_c|$ を係数とした電子密度である。本実習で精密化後に用いている電子密度は $2m|F_o|-D|F_c|$ を係数としたもの、 $m|F_o|-D|F_c|$ を係数とした電子密度であり、厳密には $2|F_o|-|F_c|$ 、 $|F_o|-|F_c|$ とは異なる。その詳細は本実習の範囲を超えるが、これらの電子密度はモデルバイアスの少ない電子密度である。

2Fo-Fcマップ

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} (2|F_{obs}| - |F_{calc}|) \exp[-2\pi i(hx + hy + lz) + i\alpha_{calc}]$$

Fo-Fcマップ

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} (|F_{obs}| - |F_{calc}|) \exp[-2\pi i(hx + hy + lz) + i\alpha_{calc}]$$



モデルが完全に正しいなら、 $F_o - F_c$ はゼロであり、 R 値 = 0%である。分子モデルに間違いがあれば、 $|F_o| - |F_c|$ を係数として電子密度を計算すれば、間違っている部分が浮き上がる。したがって、 $F_o - F_c$ マップを元にモデルを修正することが出来る。

水分子のアサイン

高分解能データでは水分子の電子密度が観測される。電子密度から自動で水分子をアサインする。
Model/Fit/Refineウインドウの[Find Waters...]をクリックすると、電子密度を選択するウインドウが現れるので、[Find Waters]をクリック。

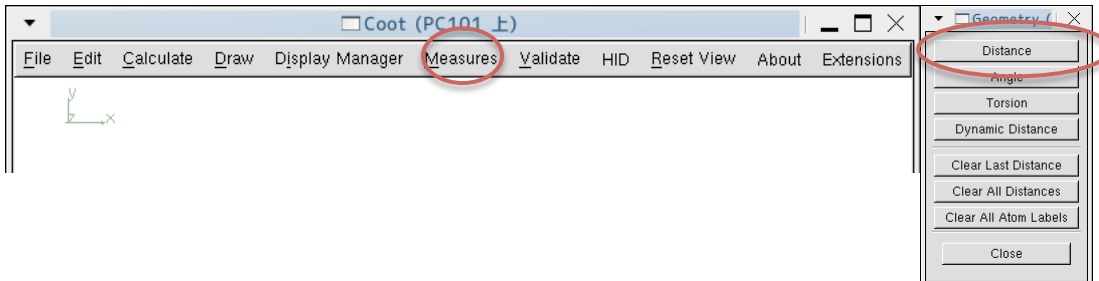
The image displays the Coot software interface for water molecule assignment. On the left is the 'Model/Fit/Refine' window with a sidebar of tools; the 'Find Waters...' button is circled in red. The top right shows the 'find waters dialog' window, also with the 'Find Waters' button circled in red. The main area shows the 'Coot' window with a 3D protein model and electron density mesh. A blue arrow points from the 'Find Waters...' button to the 3D view. Below, another screenshot shows the 3D view with water molecules assigned, marked with red 'X's, and a blue arrow points to them with the text '水分子がアサインされる'.

ファイルを保存し、[RunRefmac]で再び構造精密化を行い、最終構造とする。
 R 値、 R_{free} 値、理想的な結合距離、結合角からのRMSDをチェックする。
ペプチド結合の幾何はRamachandran Plotでチェックする。

その他

距離を測る

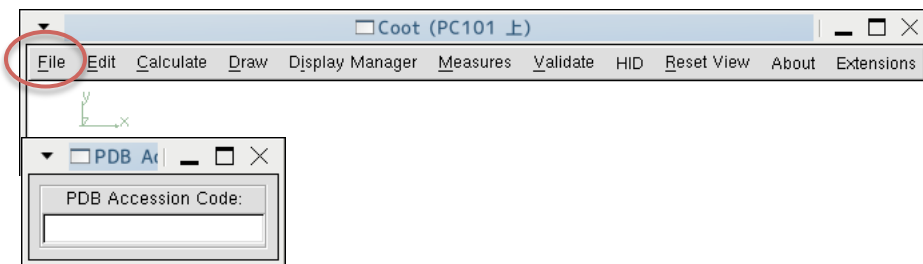
Mainウインドウから、[Measures]を選択すると、距離、角度を選択するウインドウが現れる。[Distance]をクリックすると、マウスカーソルが十字になるので、計りたい原子間距離の2原子をクリックする。



プロテインデータバンク(PDB)から構造をダウンロードし表示する

Mainウインドウから、[File] => [Get PDB using Accession Code...]を選択し、現れたウインドウにPDBコードを入力すると、データベースから構造をダウンロードし、表示される。

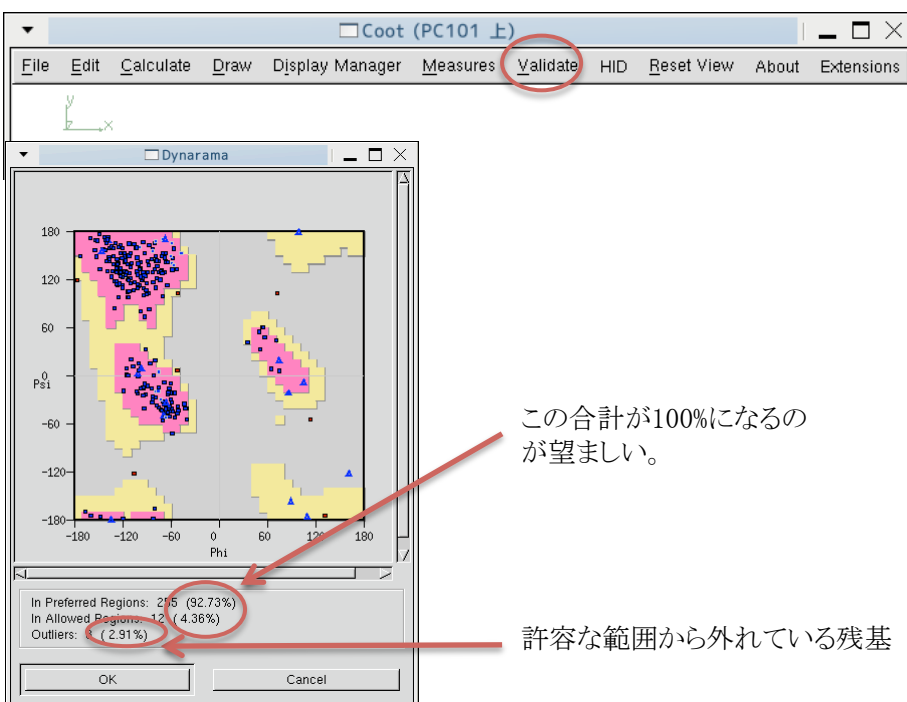
[File] => [Get PDB & Map using EDS...]を使うと、構造と電子密度が表示される。



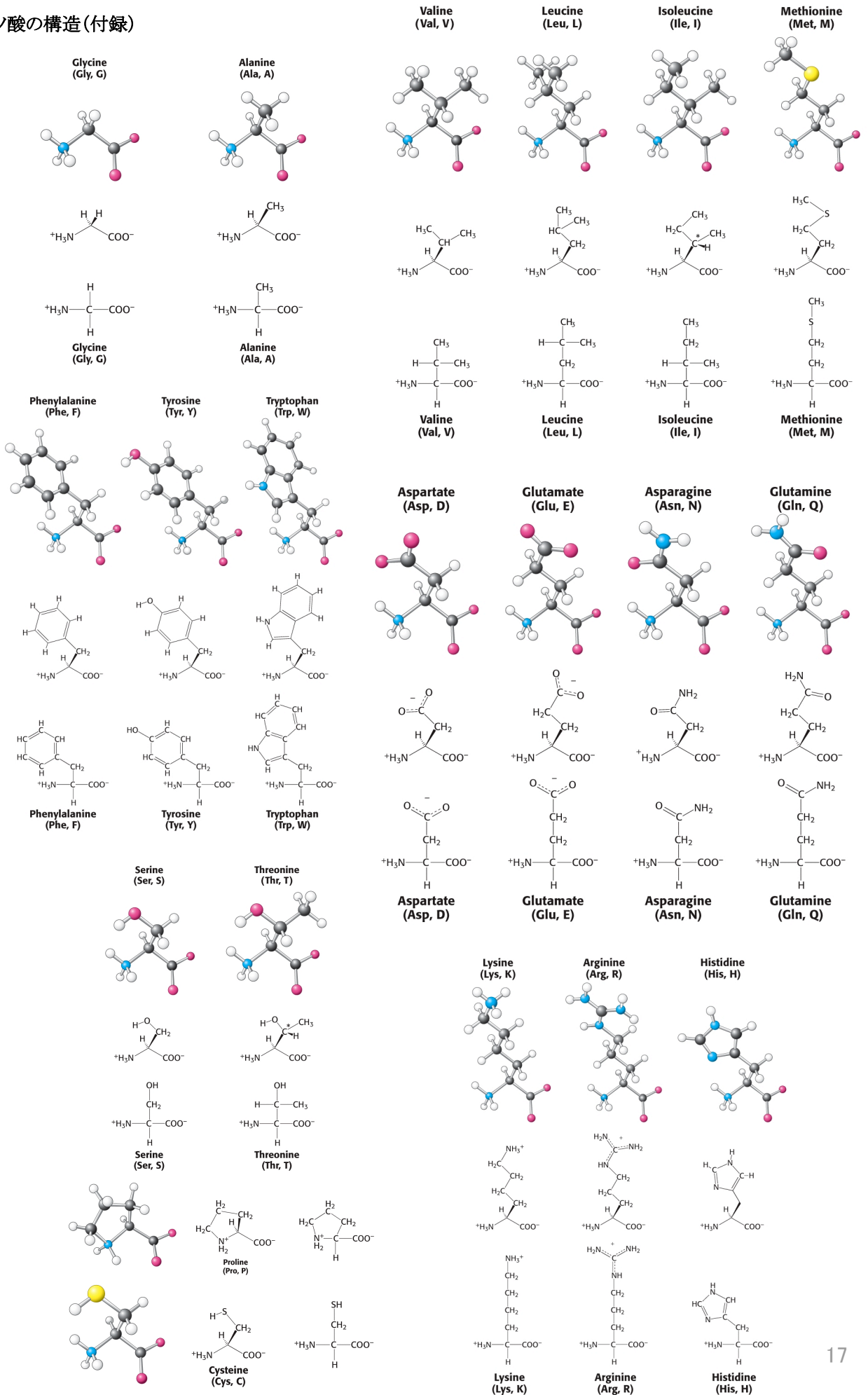
Ramachandran Plot

ペプチド結合の幾何をチェックするためにRamachandran Plotを利用する。

Mainウインドウから、[Validate] => [Ramachandran Plot]で、構造を選択する。



アミノ酸の構造(付録)

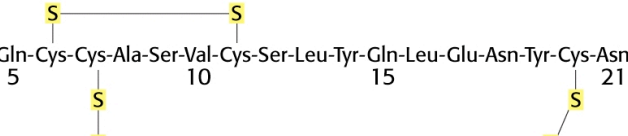


タンパク質の構造(付録)

一次構造

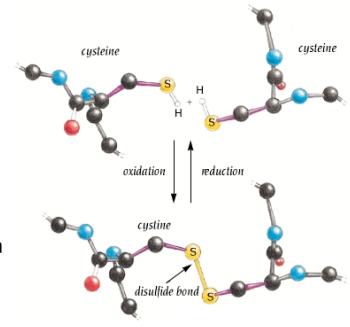
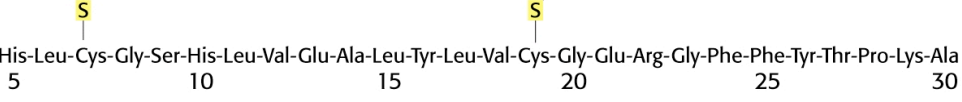
A chain

Gly-Ile-Val-Glu-Gln-Cys-Cys-Ala-Ser-Val-Cys-Ser-Leu-Tyr-Gln-Leu-Glu-Asn-Tyr-Cys-Asn

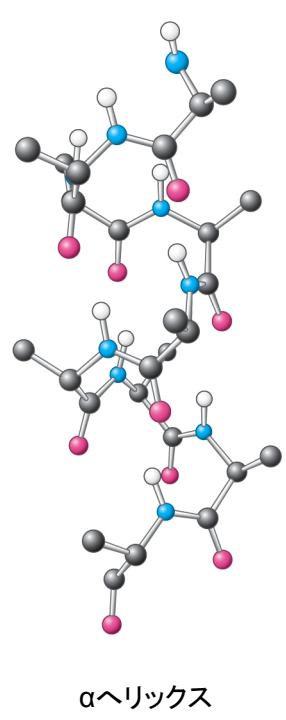


B chain

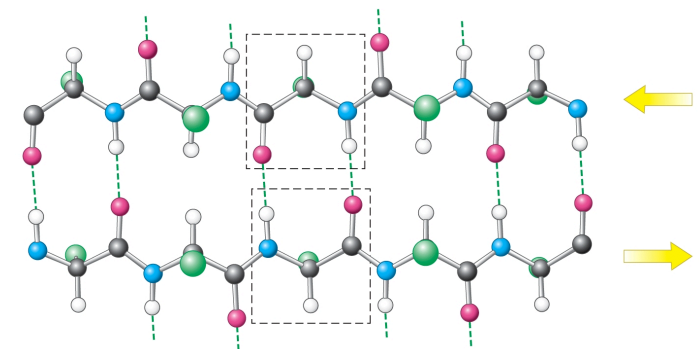
Phe-Val-Asn-Gln-His-Leu-Cys-Gly-Ser-His-Leu-Val-Glu-Ala-Leu-Tyr-Leu-Val-Cys-Gly-Glu-Arg-Gly-Phe-Phe-Tyr-Thr-Pro-Lys-Ala



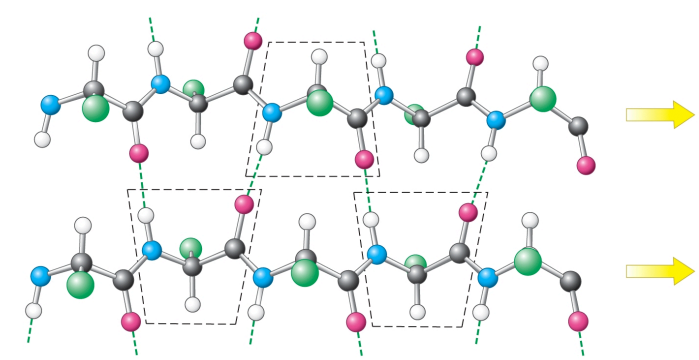
二次構造



α ヘリックス



β シート(逆平行)

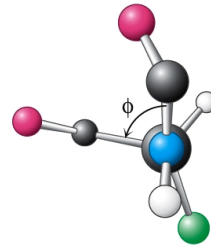
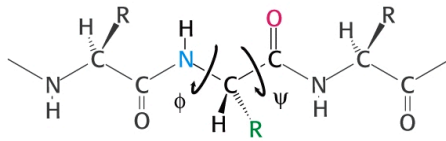
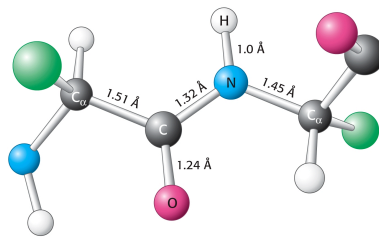


β シート(平行)

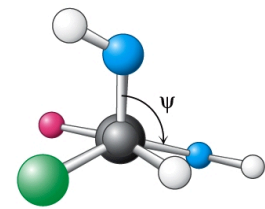
三次構造



ペプチド結合の幾何(付録)

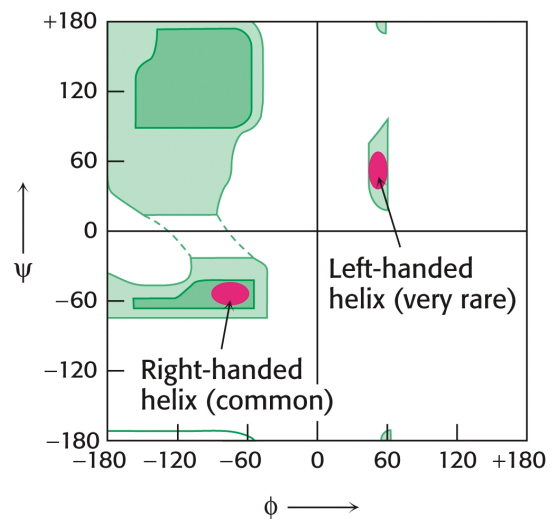
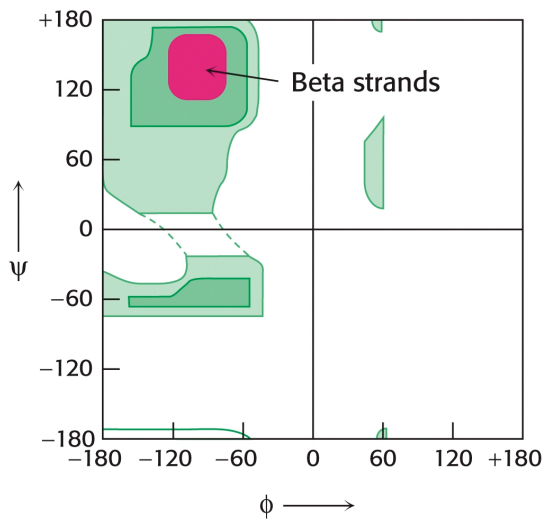
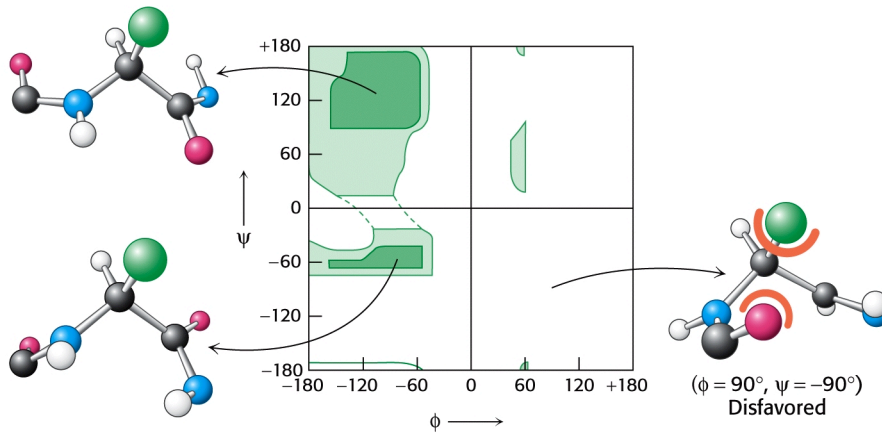


$\phi = -80^\circ$

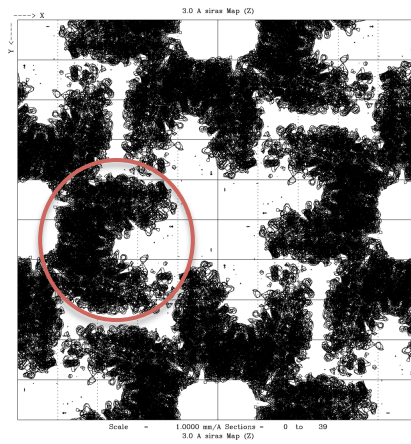
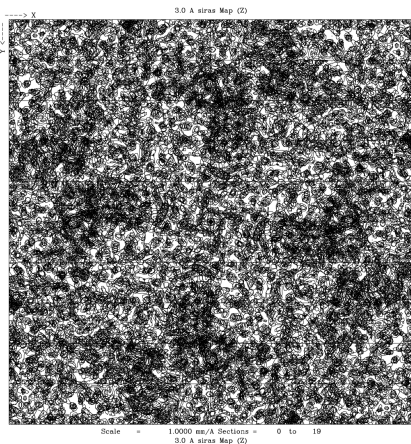


$\psi = +85^\circ$

Ramachandran Plot

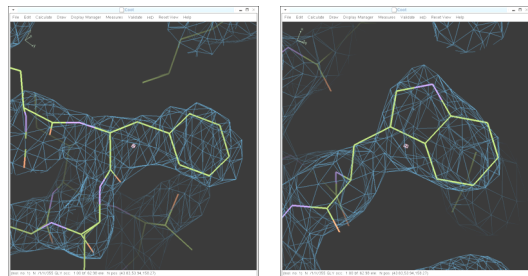
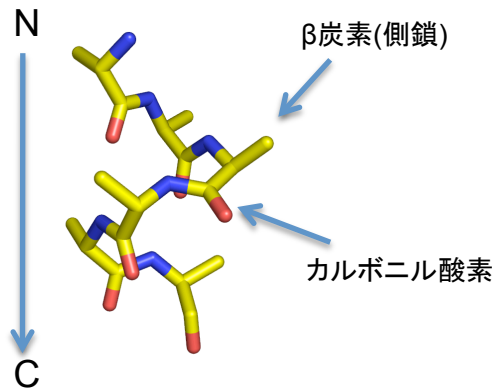
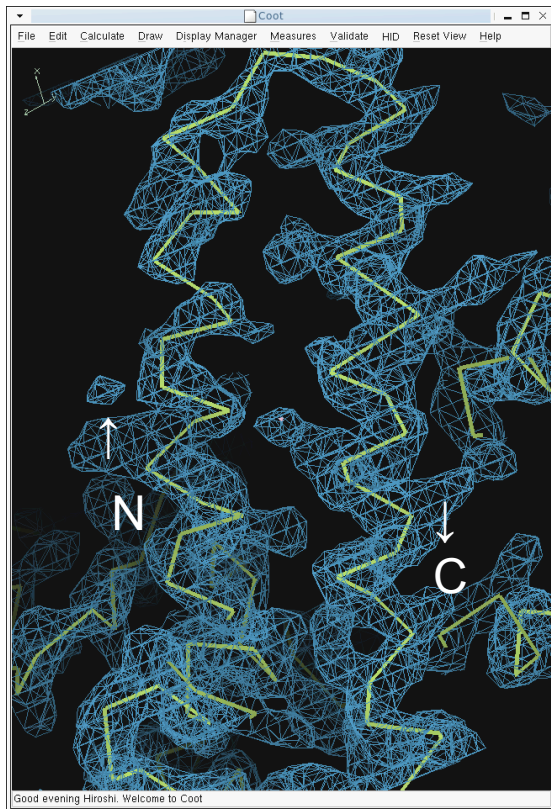


電子密度の解釈について(付録)



溶媒平滑化
分子対称を利用した平均化

分子の境界を把握する



- 2次構造を見つける
 - αヘリックス、βシート
- 主鎖の方向を決める
 - 側鎖の向き、カルボニル酸素
- 側鎖の特徴
 - 芳香族残基
- 2次構造予測の利用