

X線結晶構造解析における構造バイオインフォマティクス

(1) 分子置換法を使ってタンパク質の立体構造を決定してみよう。

(2) Coot で分子モデルを電子密度に合わせてみよう。

東京大学 大学院農学生命科学研究科 応用生命化学専攻 食品工学研究室 永田 宏次

1. 背景と目的

Protein Data Bank (PDB) には 70,000 個以上のタンパク質立体構造が登録されている。この情報を利用して、すでにアミノ酸配列類似タンパク質の立体構造が報告されているタンパク質の X 線結晶構造解析を分子置換法により行う。分子置換法を用いれば、配列相同性 30%以上の類似タンパク質の立体構造情報をモデル(鋳型)として、たいいていの場合、目的タンパク質の立体構造解析が可能である。分子置換法で構造が解けない場合は、単波長・多波長異常分散法、重原子同型置換法等により構造解析を行う。

今回、X線結晶構造解析に用いるソフトウェアパッケージ CCP4 はフリーウェアで、Unix, Linux, Mac OSX, Windows で動くので、パソコンでも構造解析が可能である。

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

2. 流れ

目的タンパク質の選択—human S100A13

シグナルペプチドをもたないタンパク質の非古典的細胞外分泌に関わる
カルシウム結合タンパク質

発現系作成

発現・精製・結晶化

X線回折データ取得・処理

X線結晶構造解析(分子置換法、単波長・多波長異常分散法、重原子同型置換法)

構造精密化・確認・PDBへの登録の仕方の説明

3. 実習

提出課題1 以下の実習で最終的に作成した PDB ファイルを
aknagata@mail.ecc.u-tokyo.ac.jp に送ってください。

不完全な状態でもかまいません。

提出課題2 講義の感想を 10 行以上書いてください。

良かったことでも、厳しいご意見でもかまいません。

1. アグリバイオ講義 HP から、圧縮ファイル 110427.lzh をデスクトップにダウンロードし、100427.lzh のアイコンをダブルクリックして解凍する。□

<http://www.iu.a.u-tokyo.ac.jp/lectures/AG04/index.html>

デスクトップ上のフォルダ 110427 には、以下の 7 個のファイルが入っている。□

s100a13.seq	human S100A13 のアミノ酸配列ファイル (一文字表記)
s100a13.sca	human S100A13 の X 線回折データファイル (Denzo/HKL2000 フォーマット)
s100a13yobi.mtz	human S100A13 の X 線回折データファイル (予備) (CCP4 フォーマット)
3NXA_A.pdb	human S100A16 の原子座標ファイル
1XK4_C.pdb	human calgranulin B の原子座標ファイル
1XK4_C_molrep1_refmac2yobi.pdb	構造精密化途中の原子座標ファイル
s100a13_refmac2yobi.mtz	構造精密化途中の X 線回折データファイル

2. [この作業は時間節約のため、永田が実行するのを見るだけにしてください]
Blast を使って、PDB (すなわち立体構造情報が登録されているタンパク質) から human S100A13 にアミノ酸配列の類似したタンパク質を検索する。□

Choose the appropriate BLAST program and database:

blastp - query against the UniProt Knowledgebase (Swiss-Prot + TrEMBL)

Taxonomic groups (not available for PDB and translated EST):

select a database To restrict the search to a particular taxon, it is much faster to see the drop-down list on the left than to specify your own taxonomic subsection gives more accurate statistics.

or specify a taxonomic group Enter a species name, a TaxID or the latin name of a taxonomic group (".", "OX lines) to restrict your search to a particular taxon. You may enter a taxonomic group name (".", "OX lines). Example: Fungi; Homo sapiens.

or select a microbial proteome

Search only Swiss-Prot (curated sequences) Exclude fragment sequences

blastp - query against another protein database Please, supply an email address

tblastn - query against the six-frame translation of a nucleotide database

Taxonomic groups:

or select a microbial genome

Your email address:

If an e-mail address is provided, results will be automatically mailed back (recommended for tblastn searches).

or

S100A13 にアミノ酸配列相同性が高く、かつ立体構造情報が PDB に登録されているタンパク質のリストが出力される。

この中から、S100A13 の立体構造情報は除外する (S100A13 の結晶構造は未知と仮定して講義しているため)。□

また、NMR で決定された溶液構造は、結晶構造に比べて正確さと精密さで劣るので、分子置換法のモデルとして用いるには不向きである。ゆえに除外する。□

結果として、

9 個目の 3NXA-A (PDB entry: 3NXA の chain A) が最良のモデルと考えられる。

まずはこの座標をモデル (鋳型) として用いて分子置換を試みる。□

失敗したら、次の候補 1XK4-C をモデルとして用いる。

List of potentially matching sequences

Send selected sequences to Clustal W (multiple alignment) 実行 Select up to...

Include query sequence

Db	AC	Description	Score	E-value
<input type="checkbox"/>	pcdb	2KI4-C FGF1_HUMAN.. Chain C, Fgf1-S100a13 Complex Structure: Key Componen...	166	2e-42
<input type="checkbox"/>	pcdb	2KI4-B FGF1_HUMAN.. Chain B, Fgf1-S100a13 Complex Structure: Key Componen...	166	2e-42
<input type="checkbox"/>	pcdb	2H2K-A S10AD_HUMAN Chain A, Crystal Structure Analysis Of Human S100a13 ...	166	2e-42
<input type="checkbox"/>	pcdb	1YUT-A S10AD_HUMAN Chain A, Solution Structure Of Calcium-S100a13 (Minim...	166	2e-42
<input type="checkbox"/>	pcdb	1YUR-A S10AD_HUMAN Chain A, Solution Structure Of Apo-S100a13 (Minimized...	166	2e-42
<input type="checkbox"/>	pcdb	1YUR-B S10AD_HUMAN Chain B, Solution Structure Of Apo-S100a13 (Minimized...	164	7e-42
<input type="checkbox"/>	pcdb	2CXJ-A S10AD_MOUSE Chain A, 3d Solution Structure Of S100a13 >gi 1105910...	144	1e-35
<input type="checkbox"/>	pcdb	2KAX-A S10A5_HUMAN Chain A, Solution Structure And Dynamics Of S100a5 In...	59	5e-10
<input type="checkbox"/>	pcdb	3NXA-A S10AG_HUMAN Chain A, X-Ray Structure Of The Apo Form Of Human S10...	58	1e-09
<input type="checkbox"/>	pcdb	2LOU-A Chain A, Solution Structure Of Apo S100a16 >gi 309319...	58	1e-09
<input type="checkbox"/>	pcdb	1NSH-A S10AB_RABBIT Chain A, Solution Structure Of Rabbit Apo-S100a11 (19...	57	2e-09
<input type="checkbox"/>	pcdb	1XK4-C S10A8_HUMAN.. Chain C, Crystal Structure Of Human Calprotectin(S100...	57	3e-09
<input type="checkbox"/>	pcdb	1IRJ-A S10A9_HUMAN Chain A, Crystal Structure Of The Mrp14 Complexed Wit...	57	3e-09
<input type="checkbox"/>	pcdb	1ODB-A S10AC_HUMAN Chain A, The Crystal Structure Of Human S100a12 - Cop...	56	5e-09
<input type="checkbox"/>	pcdb	1E8A-A S10AC_HUMAN Chain A, The Three-Dimensional Structure Of Human S10...	56	5e-09
<input type="checkbox"/>	pcdb	3MOW-A S10A4_HUMAN Chain A, Structure Of S100a4 With Pcp >gi 295982436 p...	55	1e-08
<input type="checkbox"/>	pcdb	2WCE-A S10AC_HUMAN Chain A, Calcium-Free (Apo) S100a12 >gi 241913116 pdb...	55	1e-08
<input type="checkbox"/>	pcdb	3C1V-A S10A4_HUMAN Chain A, The 1.5 Å Crystal Structure Of Ca2+-Bound S1...	55	1e-08
<input type="checkbox"/>	pcdb	1M31-A S10A4_HUMAN Chain A, Three-Dimensional Solution Structure Of Apo-...	55	1e-08
<input type="checkbox"/>	pcdb	2JPT-A S10A1_BOVIN Chain A, Structural Changes Induced In Apo-S100a1 Pro...	54	1e-08

複数のペプチド鎖を含む場合は、似ているペプチド鎖だけの情報を抽出して、別名で保存する。例: 1xk4_C.pdb。□

ATOM	1434	N	LYS	C	4	9.892	70.055	167.750	1.00	50.82	N
ATOM	1435	CA	LYS	C	4	9.141	68.965	168.427	1.00	49.76	C
ATOM	1436	C	LYS	C	4	9.606	67.576	167.983	1.00	46.94	C
(途中省略)											
ATOM	2176	CD	GLU	C	92	33.783	49.554	166.930	1.00	44.92	C
ATOM	2177	OE1	GLU	C	92	34.739	48.784	167.199	1.00	45.60	O
ATOM	2178	OE2	GLU	C	92	33.105	50.143	167.813	1.00	45.73	O

アミノ酸配列のアラインメントをとると、以下の通り。配列相同性は 30%弱。

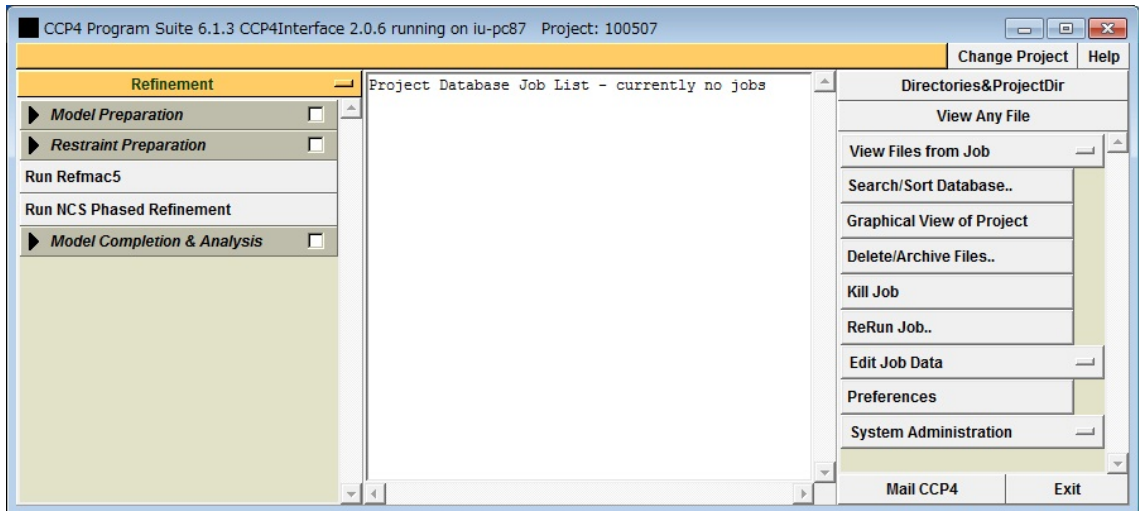
```

1      10     20     30     40     50     60     70     80     90    100    110 115
S100a13 MAREPLTELESITVYVTFFFRROEGRKQSLSYMEFKELVTOQLPHLLKDYGS---LDEKHKSLDYNQDSELKFMEYMRIGELAKEIRKKDLKIRKK
1XK4:C   TSKNSQLERNIETIINTHQYSVKLGHPDILNAGEFKELVYKQLQNFLLKENKNEKYIEITIMEDLDTRAKQLSFEFIMLMARLTHASHEKHEGDEGPGHHKPKLGEGTP
1XK4:A   MLTELEKALNSIDVYHKYSLIKGFHAYYRDLKLLLETESPOYIRKKA---DVMFKELDINTDGAIVNQEFLIIVIKNGYARHKSHIESHKE
Consensus .....$t#LE...i#t!!tZh,%...G...d.l...#fkelv...#lp...lkk.g.....#.nk.LD.W.D..l.F#E%.l....$.a.hkk.he...k.....

```

3. [お待たせしました。CCP4 を用いて、分子置換を行います。ここから皆さんに実行してもらいます]

まず、デスクトップ上の CCP4i アイコンをダブルクリックして CCP4i (CCP4Interface) を起動する。□



4. 作業ファイルを扱うディレクトリを設定する。

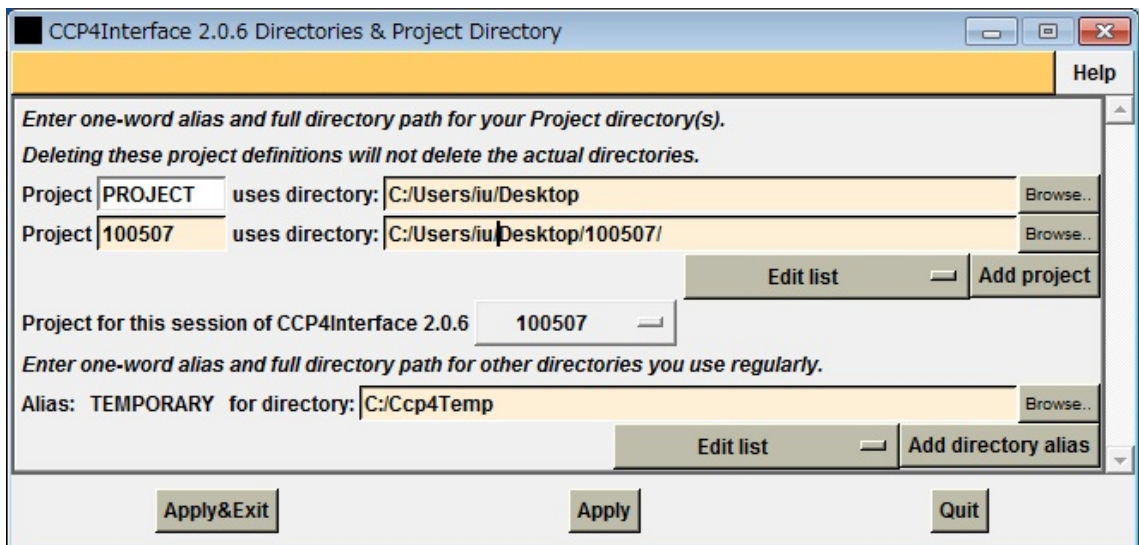
右上にある Directories&ProjectDir ボタンを押すと以下のウインドウが開く。□

Add project ボタンをクリックして、追加された空行に以下のように記入する。□

Project: 110427 uses directory: C:/Users/iu/Desktop/110427/

Project for this session of CCP4Interface 2.0.6 として 110427 を選択する。□

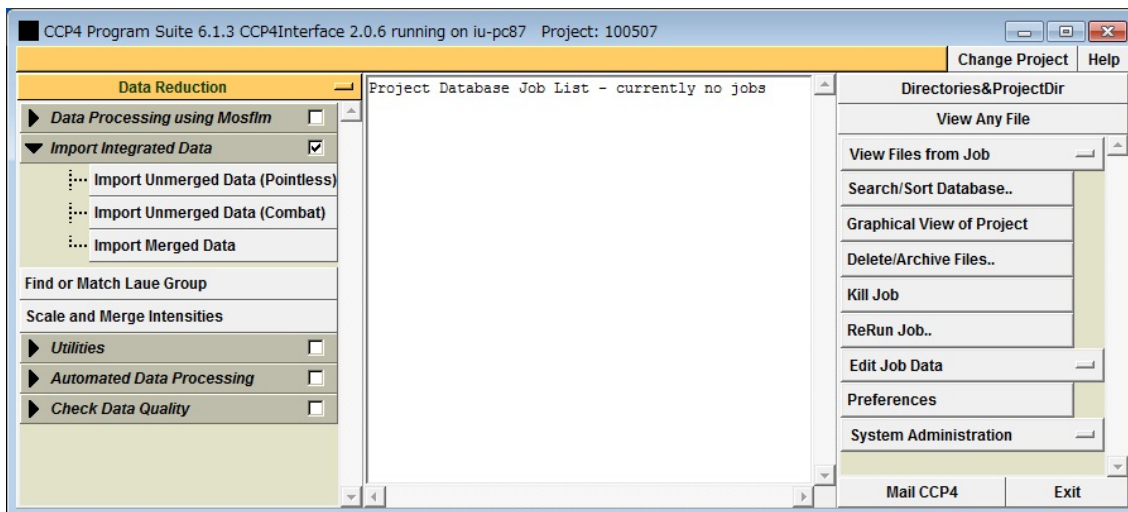
その後、Apply&Exit ボタンを押す。□



5. X線回折データのフォーマット変換 (Denzo/HKL2000 → CCP4) を行う。

左側の作業メニューの黄色いバーをクリックすると種々のメニューが現れる。□

Data Reduction → Import Integrated Data → Import Merged Data を選択すると ImportScaled のウインドウが開く。□



以下のようにチェックする。

Use anomalous data (異常分散データでないので、チェックをはずす)

Run Ctruncate to convert intensities to structure factors

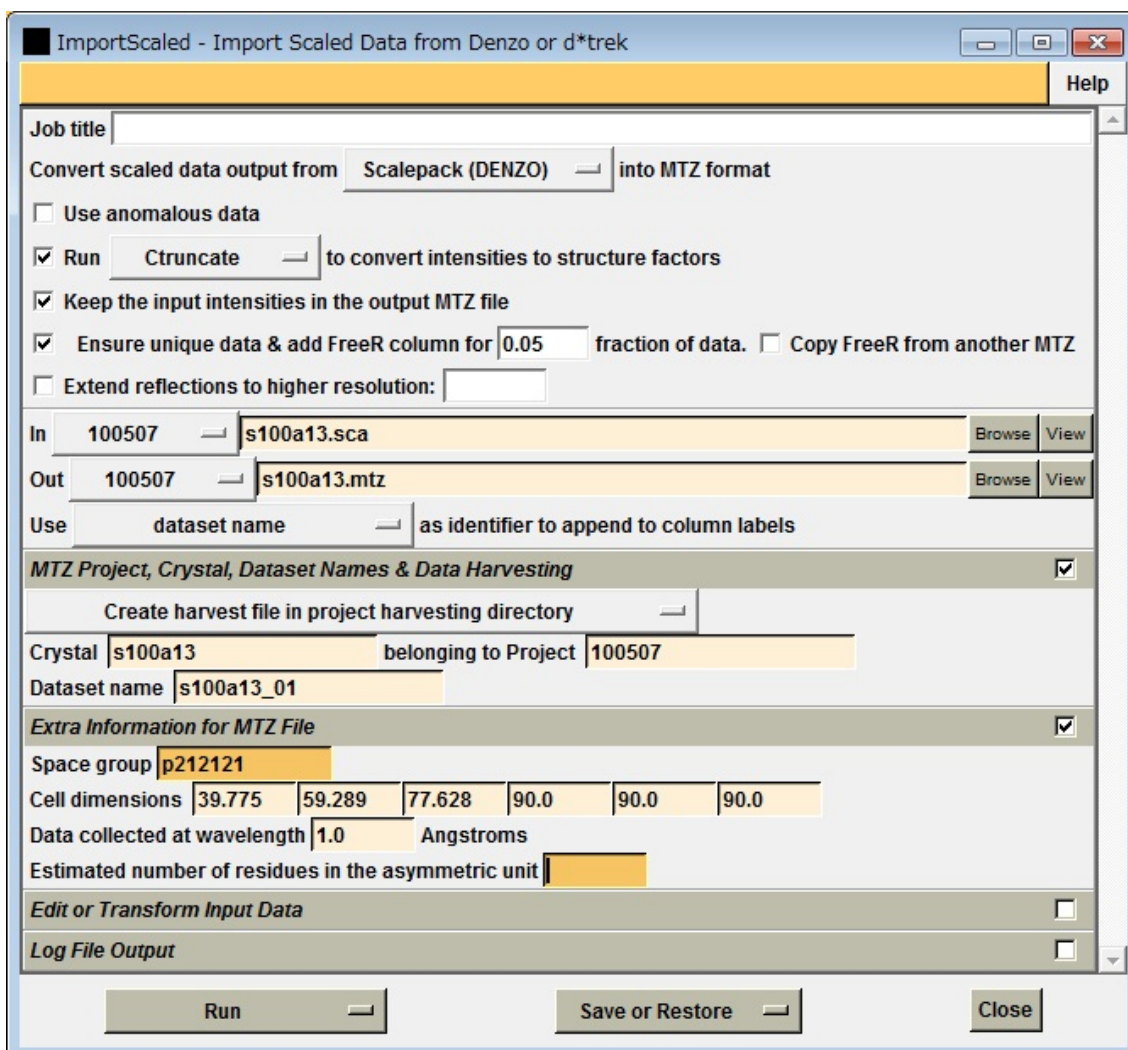
Keep the input intensities in the output MTZ file

Ensure unique data & add FreeR column for 0.05 fraction of data.

Copy FreeR from another MTZ

Extend reflections to higher resolution:

(次ページに続く)



入力ファイルとして、s100a13.sca を選択する。Browse ボタンを使うと楽。□
出力ファイル名が、勝手に指定される（拡張子が.mtz に変わっただけ）。

In 110427: s100a13.sca

Out 110427: s100a13.mtz

その他、入力が必要な項目は、Extra information for MTZ file の波長の値。有効数字を考慮して、1.0000 (Angstrom)と入力するが、勝手に 1.0 に変換される。□

Crystal と Dataset name の入力はしなくてもよいが、それぞれ S100A13、S100A13_01 と入力しておく。

Data collected at wavelength: 1.0 Angstroms

Run → Run Now ボタンを押して、フォーマット変換を実行すると、ファイル s100a13.mtz が作成される。□

Import Scaled - Import Scaled Data from Denzo or d*trek

Job title

Convert scaled data output from **Scalepack (DENZO)** into MTZ format

Use anomalous data

Run **Ctruncate** to convert intensities to structure factors

Keep the input intensities in the output MTZ file

Ensure unique data & add FreeR column for **0.05** fraction of data. Copy FreeR from another MTZ

Extend reflections to higher resolution:

In **100507** **s100a13.sca** Browse View

Out **100507** **s100a13.mtz** Browse View

Use **dataset name** as identifier to append to column labels

MTZ Project, Crystal, Dataset Names & Data Harvesting

Create harvest file in project harvesting directory

Crystal **s100a13** belonging to Project **100507**

Dataset name **s100a13_01**

Extra Information for MTZ File

Space group **p212121**

Cell dimensions **39.775** **59.289** **77.628** **90.0** **90.0** **90.0**

Data collected at wavelength **1.0** Angstroms

Estimated number of residues in the asymmetric unit **98**

Edit or Transform Input Data

Log File Output

Run Save or Restore Close

6. 分子置換法の準備として、非対称単位中の S100A13 分子数を見積もる。

(非対称単位 = 結晶中に現れる繰り返し構造の 1 つを取り出したもの。実際の結晶中ではこの構造がある法則にそって前後左右上下に繰り返されている)

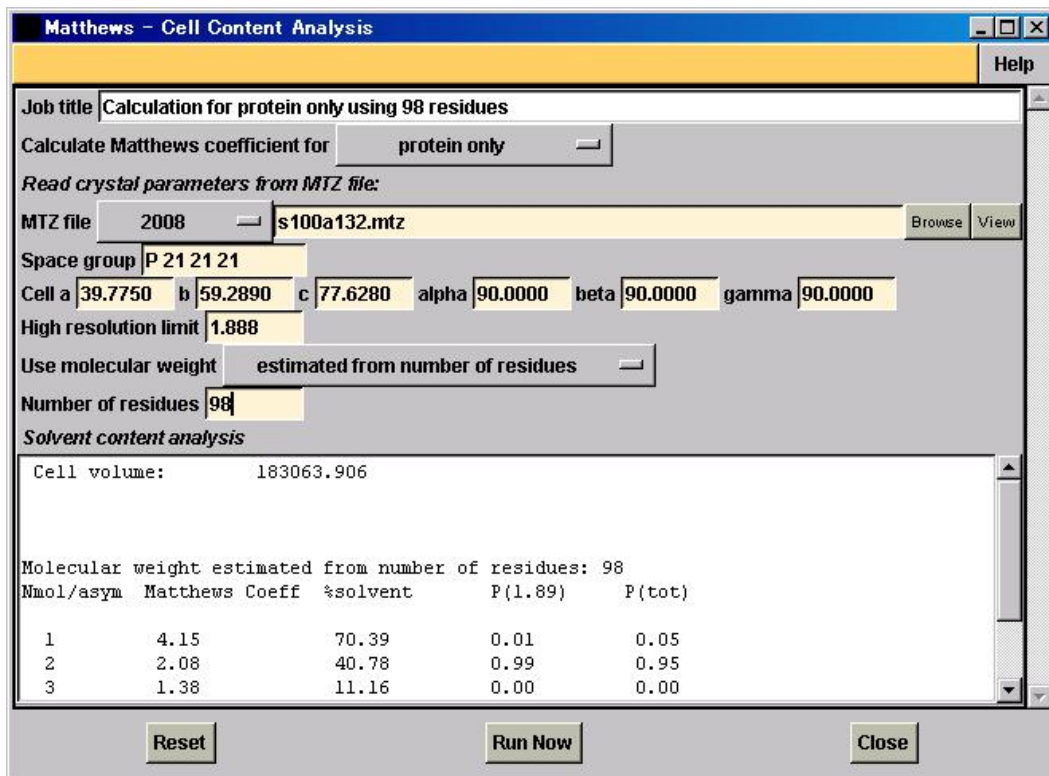
左側の作業メニューから、Molecular Replacement → Analysis → Cell Content Analysis を選択すると Matthews のウィンドウが開く。□

MTZ file として、s100a13.mtz を選択する。□

Use molecular weight: estimated from number of residues にして

Number of residues: 98 と入力する。□

Run Now ボタンを押すと、下の白い枠に、非対称単位中のタンパク質分子数、Matthews 係数、溶媒含有率、確率 (2 通り) が表示される。□



この場合、非対称単位中 S100A13 が 2 分子含まれると確定した。□

7. 非対称単位中の残基数 196 を入力し、Import Merged Data を再実行する。

左側の作業メニューから Data Reduction → Import Integrated Data → Import Merged Data を選択すると ImportScaled のウィンドウが開く。□

基本的に 5 と同じ設定だが、Extra information for MTZ file 中の

Estimated number of residues in the asymmetric unit に 196 と入力する。□

その後、Run → Run Now ボタンを押すと、すでに同じ名称の出力ファイルが存在するという警告メッセージが出るが、Continue ボタンを押して、上書きする。

これで、human S100A13 の X 線回折データファイル (Denzo 形式。s100a13.sca) の CCP 形式 (s100a13.mtz) への書式変換が完了した。

ImportScaled - Import Scaled Data from Denzo or d*trek

Help

Job title

Convert scaled data output from **Scalepack (DENZO)** into MTZ format

Use anomalous data

Run **Ctruncate** to convert intensities to structure factors

Keep the input intensities in the output MTZ file

Ensure unique data & add FreeR column for **0.05** fraction of data. Copy FreeR from another MTZ

Extend reflections to higher resolution:

In **100507** **s100a13.sca**

Out **100507** **s100a13.mtz**

Use **dataset name** as identifier to append to column labels

MTZ Project, Crystal, Dataset Names & Data Harvesting

Create harvest file in project harvesting directory

Crystal **s100a13** belonging to Project **100507**

Dataset name **s100a13_01**

Extra Information for MTZ File

Space group **p212121**

Cell dimensions **39.775** **59.289** **77.628** **90.0** **90.0** **90.0**

Data collected at wavelength **1.0** Angstroms

Estimated number of residues in the asymmetric unit **196**

Edit or Transform Input Data

Log File Output

8. Molrep を用いて分子置換を実行する。

作業メニューから Molecular Replacement → Model Generation → Run Molrep - auto MR を選択すると、Molrep のウィンドウが開く。□

以下のように設定する。□

Do: molecular replacement performing: rotation and translation function

Get structure factors from MTZ file

Input fixed model

Multi-copy search

Use sequence

入力ファイルは以下の3つ。□

MTZ in: 110427: s100a13.mtz (S100A13 の X 線回折データ)

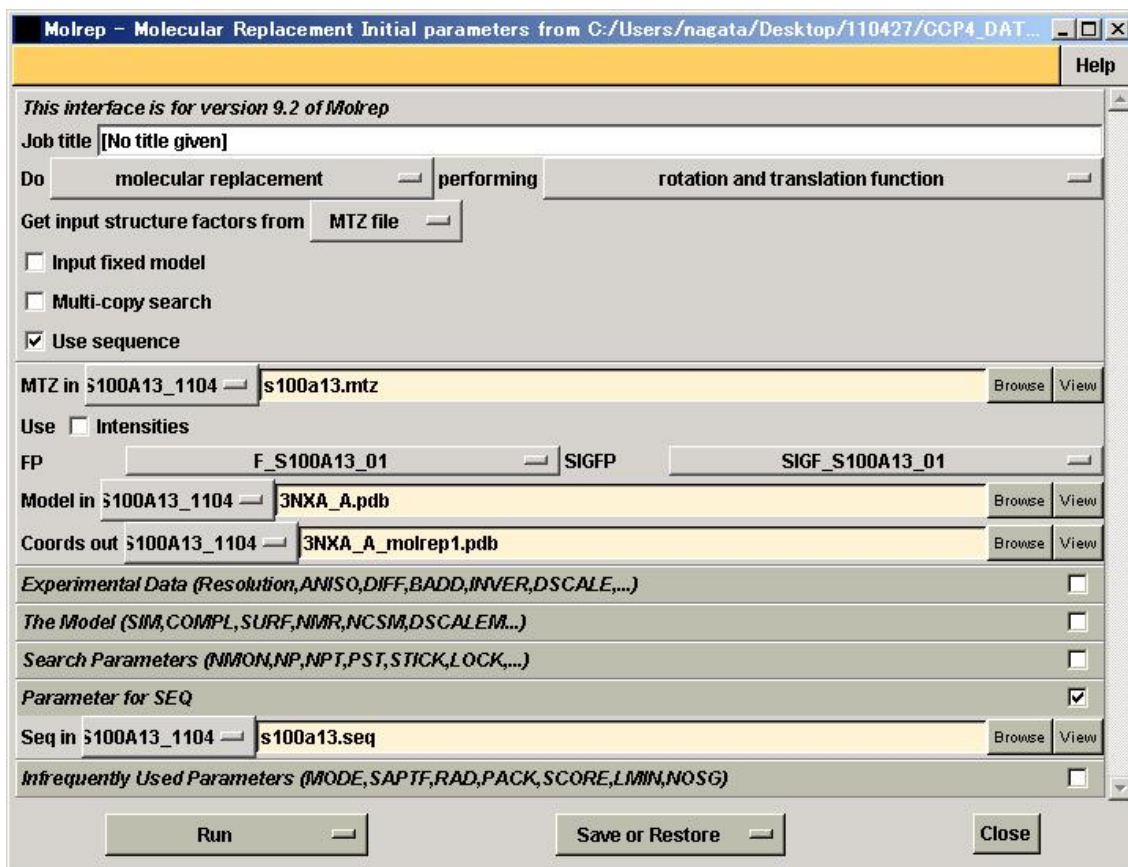
Model in: 110427: 3NXA_A.pdb (立体構造既知配列類似タンパク質 S100A16 単量体 (chain A) の原子座標ファイル)

Seq in: 110427: s100a13.seq (S100A13 のアミノ酸配列。一文字表記)

出力ファイル名は自動で設定される。

Coords out: 110427: 3NXA_A_molrep1.pdb

Run → Run Now ボタンを押すと計算が始まる。□

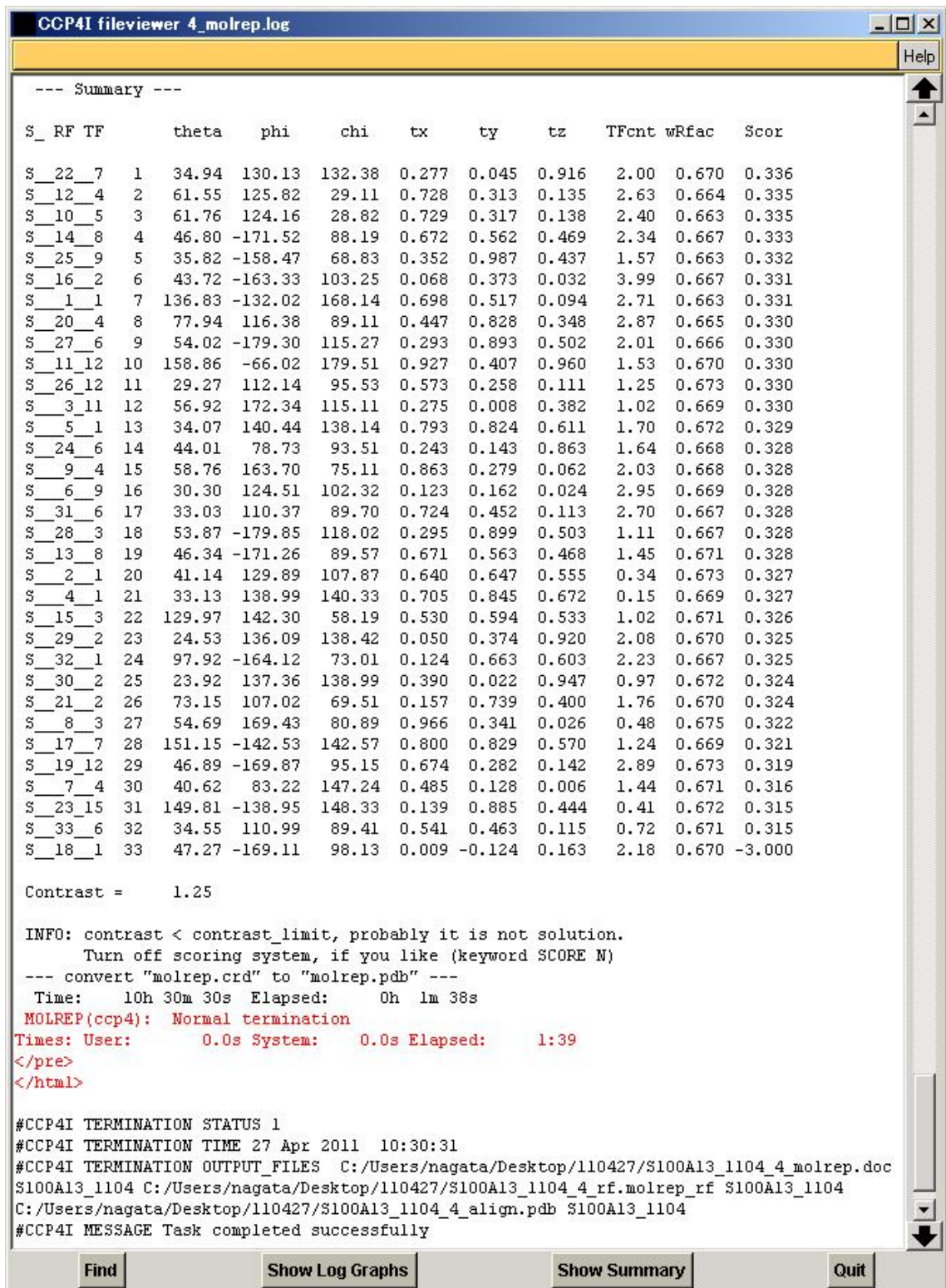


CCP4Interface の中央の作業ログで、molrep の行を選択した後、右側の View Files from Job ボタンをクリックし、プルダウンメニューの View Log File をクリックすると計算の過程を追うことができる。□

ログファイル最終行近くにある解 Summary の wRfac と Scor の値に注目すると、良い

解がないことがわかる。□

実際に、”contrast < contrast_limit, probably it is not solution”というメッセージが出ていることを確認する。□



```
CCP4I fileviewer 4_molrep.log
--- Summary ---
S_ RF TF      theta  phi    chi    tx    ty    tz    TFcnt wRfac  Scor
S_22_7  1  34.94 130.13 132.38 0.277 0.045 0.916  2.00 0.670 0.336
S_12_4  2  61.55 125.82  29.11 0.728 0.313 0.135  2.63 0.664 0.335
S_10_5  3  61.76 124.16  28.82 0.729 0.317 0.138  2.40 0.663 0.335
S_14_8  4  46.80 -171.52  88.19 0.672 0.562 0.469  2.34 0.667 0.333
S_25_9  5  35.82 -158.47  68.83 0.352 0.987 0.437  1.57 0.663 0.332
S_16_2  6  43.72 -163.33 103.25 0.068 0.373 0.032  3.99 0.667 0.331
S_1_1  7 136.83 -132.02 168.14 0.698 0.517 0.094  2.71 0.663 0.331
S_20_4  8  77.94  116.38  89.11 0.447 0.828 0.348  2.87 0.665 0.330
S_27_6  9  54.02 -179.30 115.27 0.293 0.893 0.502  2.01 0.666 0.330
S_11_12 10 158.86 -66.02 179.51 0.927 0.407 0.960  1.53 0.670 0.330
S_26_12 11  29.27 112.14  95.53 0.573 0.258 0.111  1.25 0.673 0.330
S_3_11 12  56.92 172.34 115.11 0.275 0.008 0.382  1.02 0.669 0.330
S_5_1 13  34.07 140.44 138.14 0.793 0.824 0.611  1.70 0.672 0.329
S_24_6 14  44.01  78.73  93.51 0.243 0.143 0.863  1.64 0.668 0.328
S_9_4 15  58.76 163.70  75.11 0.863 0.279 0.062  2.03 0.668 0.328
S_6_9 16  30.30 124.51 102.32 0.123 0.162 0.024  2.95 0.669 0.328
S_31_6 17  33.03 110.37  89.70 0.724 0.452 0.113  2.70 0.667 0.328
S_28_3 18  53.87 -179.85 118.02 0.295 0.899 0.503  1.11 0.667 0.328
S_13_8 19  46.34 -171.26  89.57 0.671 0.563 0.468  1.45 0.671 0.328
S_2_1 20  41.14 129.89 107.87 0.640 0.647 0.555  0.34 0.673 0.327
S_4_1 21  33.13 138.99 140.33 0.705 0.845 0.672  0.15 0.669 0.327
S_15_3 22 129.97 142.30  58.19 0.530 0.594 0.533  1.02 0.671 0.326
S_29_2 23  24.53 136.09 138.42 0.050 0.374 0.920  2.08 0.670 0.325
S_32_1 24  97.92 -164.12  73.01 0.124 0.663 0.603  2.23 0.667 0.325
S_30_2 25  23.92 137.36 138.99 0.390 0.022 0.947  0.97 0.672 0.324
S_21_2 26  73.15 107.02  69.51 0.157 0.739 0.400  1.76 0.670 0.324
S_8_3 27  54.69 169.43  80.89 0.966 0.341 0.026  0.48 0.675 0.322
S_17_7 28 151.15 -142.53 142.57 0.800 0.829 0.570  1.24 0.669 0.321
S_19_12 29  46.89 -169.87  95.15 0.674 0.282 0.142  2.89 0.673 0.319
S_7_4 30  40.62  83.22 147.24 0.485 0.128 0.006  1.44 0.671 0.316
S_23_15 31 149.81 -138.95 148.33 0.139 0.885 0.444  0.41 0.672 0.315
S_33_6 32  34.55 110.99  89.41 0.541 0.463 0.115  0.72 0.671 0.315
S_18_1 33  47.27 -169.11  98.13 0.009 -0.124 0.163  2.18 0.670 -3.000

Contrast = 1.25

INFO: contrast < contrast_limit, probably it is not solution.
      Turn off scoring system, if you like (keyword SCORE N)
--- convert "molrep.crd" to "molrep.pdb" ---
Time: 10h 30m 30s Elapsed: 0h 1m 38s
MOLREP(ccp4): Normal termination
Times: User: 0.0s System: 0.0s Elapsed: 1:39
</pre>
</html>

#CCP4I TERMINATION STATUS 1
#CCP4I TERMINATION TIME 27 Apr 2011 10:30:31
#CCP4I TERMINATION OUTPUT_FILES C:/Users/nagata/Desktop/110427/S100A13_1104_4_molrep.doc
S100A13_1104 C:/Users/nagata/Desktop/110427/S100A13_1104_4_rf.molrep_rf S100A13_1104
C:/Users/nagata/Desktop/110427/S100A13_1104_4_align.pdb S100A13_1104
#CCP4I MESSAGE Task completed successfully
```

3NXA_A は分子置換のための良い鋳型ではなかったと割り切って、次の鋳型を使って、分子置換を試みる。□

先に行った Blast 検索の結果、次に鋳型の候補となるものは、1XK4-C (PDB entry: 1XK4 の chain C。タンパク質名、calgranulin B)

9. 再度、Molrep を用いて分子置換を実行する。

作業メニューから Molecular Replacement → Model Generation → Run Molrep - auto MR を選択すると、Molrep のウィンドウが開く。□

以下のように設定する。□

Do: molecular replacement performing: rotation and translation function

Get structure factors from MTZ file

Input fixed model

Multi-copy search

Use sequence

入力ファイルは以下の 3 つ。□

MTZ in: 110427: s100a13.mtz

(S100A13 の X 線回折データ)

Model in: 110427: 1XK4_C.pdb

(立体構造既知配列類似タンパク質 calgranulin B 単量体 (chain C) の原子座標ファイル)

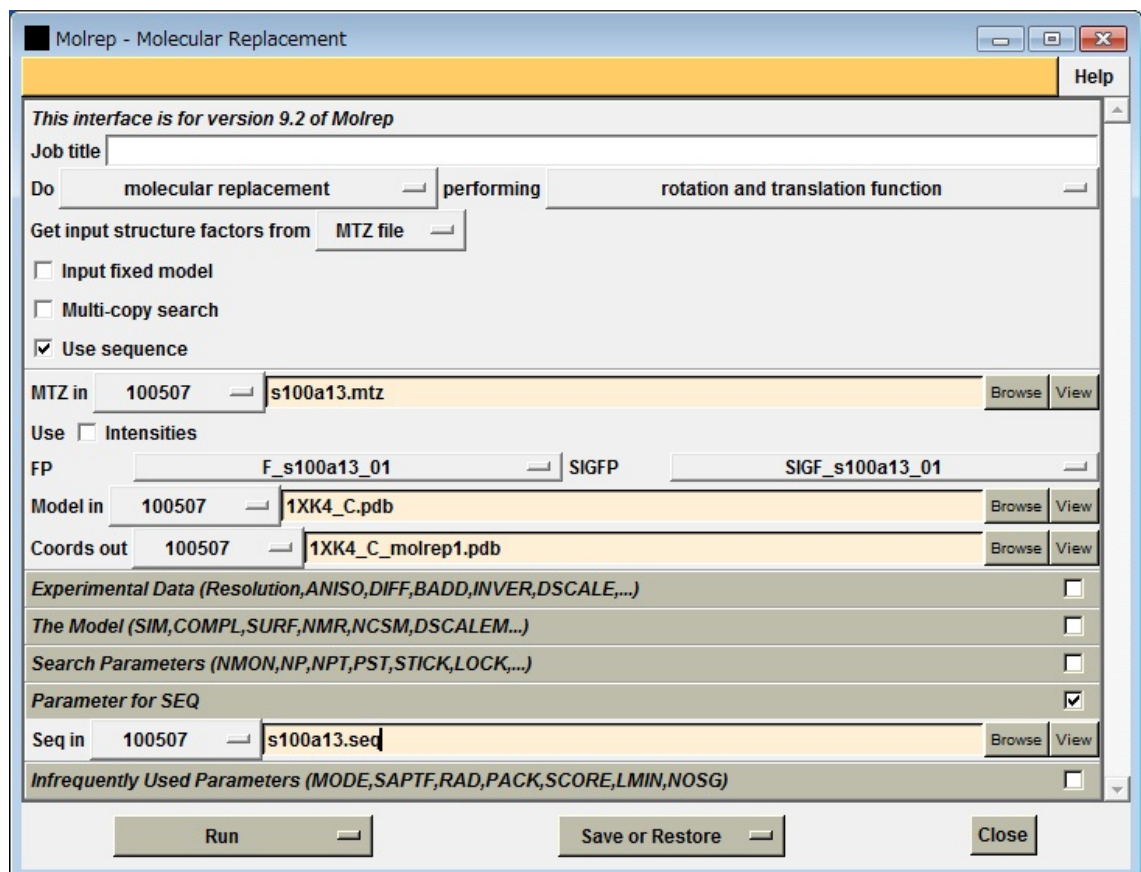
Seq in: 110427: s100a13.seq

(S100A13 のアミノ酸配列。一文字表記)

出力ファイル名は自動で設定される。

Coords out: 110427: 1XK4_C_molrep1.pdb

Run → Run Now ボタンを押すと計算が始まる。□



CCP4Interface の中央の作業ログで、molrep の行を選択した後、右側の View Files from Job ボタンをクリックし、プルダウンメニューの View Log File をクリックすると計算の過程を追うことができる。□

非対称単位中に S100A13 分子を 1 個置いたときの解。ログファイル最終行近くにある解 Sol_ の Rfac と Scor の値に注目すると、上位 2 個の値が良い。□

```

--- Summary ---

S_ RF TF      theta  phi    chi    tx    ty    tz    TFcnt wRfac  Scor
S__2_1  1  164.08 -133.11 162.46 0.105 0.238 0.145  4.15 0.636 0.402
S__1_1  2   92.63 -123.21  23.92 0.287 0.129 0.080  6.26 0.636 0.401
S__11_1  3  140.96  -78.66 122.72 0.239 0.279 0.223  2.12 0.661 0.352
S__10_3  4   43.16 -159.67  95.67 0.121 0.174 0.389  1.94 0.659 0.351
S__8_5   5   35.08  65.74 114.00 0.260 0.120 0.003  1.89 0.662 0.345
S__3_3   6   14.28  15.79 171.90 0.070 0.455 0.040  2.04 0.659 0.345
S__5_7   7  137.94 -173.54 171.33 0.156 0.117 0.467  2.21 0.671 0.338
S__7_1   8   13.24  31.51 160.90 0.059 0.431 0.047  2.40 0.666 0.337
S__9_11  9  142.26 134.89  53.80 0.171 0.325 0.361  2.21 0.658 0.337
S__4_1  10   44.08 -167.85 139.47 0.055 0.117 0.261  2.26 0.663 0.336
S__6_13 11  142.23 121.47  51.05 0.090 0.479 0.274  2.45 0.667 0.333

Contrast =      4.64

After stick correction:
Move closer to origin
I_sym_operator   :    1
new position(frac):  0.105  0.238  0.145

Sol_
S_ Nmon RF TF  theta  phi    chi    tx    ty    tz    TFcnt wRfac  Scor
S__1_2  2  1  164.08 -133.11 162.46 0.105 0.238 0.145  4.15 0.636 0.402

```

最上位の解を採用し (S100A13 分子を非対称単位中に 1 個置き)、2 個目の分子を置いたときの解。Rfac と Scor の値に注目すると、最上位の解 (さきほどの 2 位の解) が飛びぬけて良いので、これを採用する。□

2 個目の分子を置くと、1 個だけの時よりも、wRfac 値が下がり、Scor 値が上がる。□
このように Molrep を用いる分子置換法により、非対称単位中に S100A13 分子を 2 個置くことができた。□

結晶の最小構成単位の箱の中に 2 個のタンパク質分子を置く位置と向きを検討した結果、回折データと割とよく合う位置と向きが見つかった、と理解してください。□

CCP4I fileviewer 3_molrep.log

Help

--- Summary ---

S_ RF TF	theta	phi	chi	tx	ty	tz	TFcnt	wRfac	Scor	
S__1__1	1	92.63	-123.21	23.92	0.789	0.129	0.580	18.74	0.579	0.507
S__8__1	2	35.08	65.74	114.00	0.970	0.587	0.950	2.04	0.635	0.403
S__13_14	3	171.12	-162.05	141.11	0.843	0.611	0.021	1.90	0.631	0.401
S__3__2	4	14.28	15.79	171.90	0.025	0.095	0.637	3.32	0.632	0.401
S__12__4	5	68.87	88.34	38.24	0.218	0.877	0.033	1.75	0.634	0.400
S__11_10	6	140.96	-78.66	122.72	0.503	0.929	0.292	2.14	0.631	0.398
S__5__1	7	137.94	-173.54	171.33	0.918	0.633	0.974	2.68	0.638	0.398
S__10__9	8	43.16	-159.67	95.67	0.574	0.090	0.380	0.68	0.638	0.394
S__7__14	9	13.24	31.51	160.90	0.827	0.869	0.940	0.90	0.637	0.387
S__9__11	10	142.26	134.89	53.80	0.421	0.663	0.056	0.37	0.637	0.387
S__6__5	11	142.23	121.47	51.05	0.030	0.599	0.026	0.62	0.638	0.386
S__4__2	12	44.08	-167.85	139.47	0.614	0.216	0.399	-0.01	0.632	0.386
S__2__1	13	164.08	-133.11	162.46	0.105	0.238	0.145	4.15	0.636	-3.000

Contrast = 8.62

After stick correction (more information in molrep.doc):
 Move closer to molecule N : 1
 --- Dimer ---
 Rotation to molecule N (polar angles) : 74.15 -77.18 179.86
 Distance between centres (Å) : 23.9
 Orthogonal axis distance and parallel (Å) : 23.9 0.0
 I_sym_operator : 3
 new position(frac): 0.289 0.371 0.420

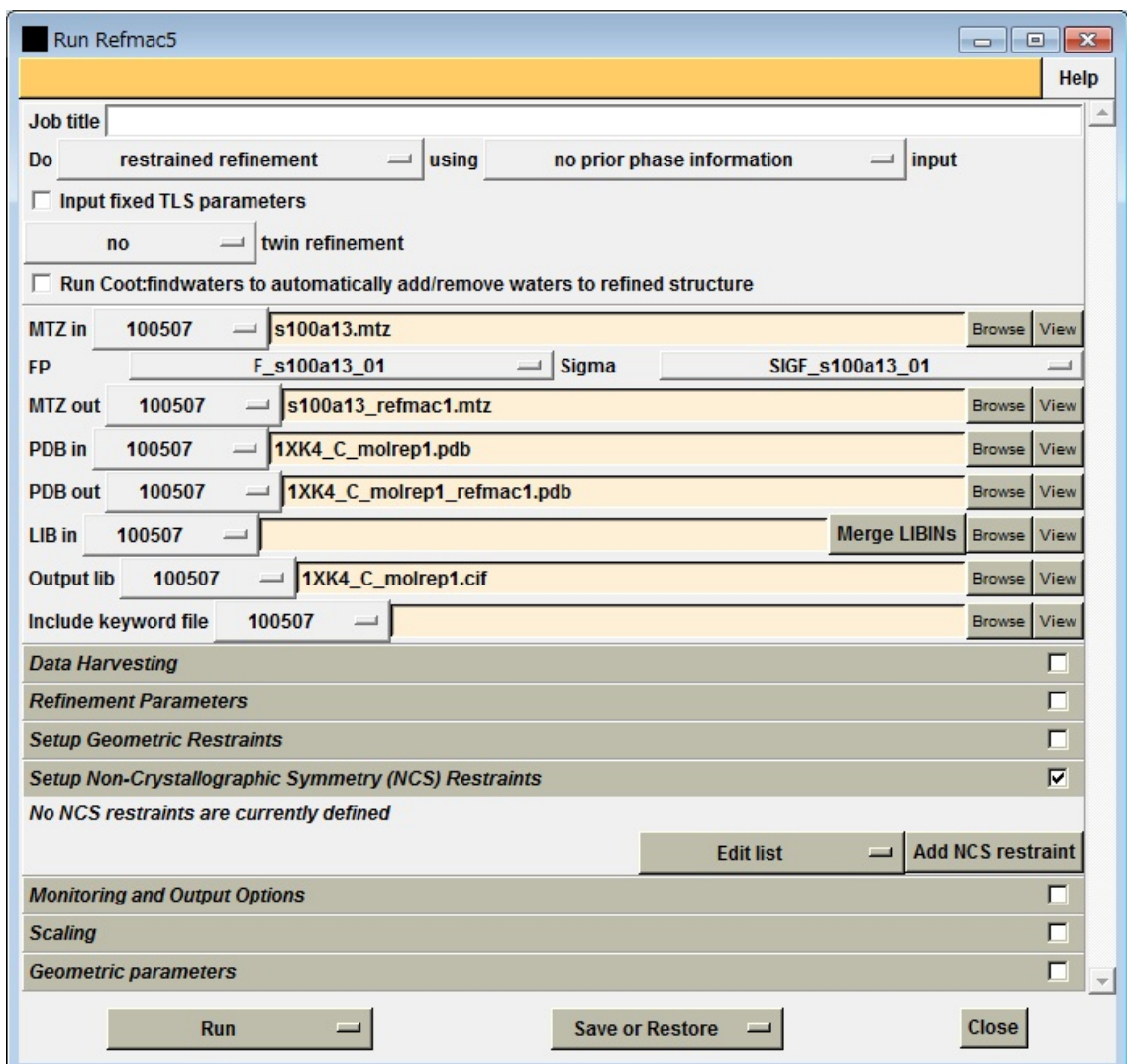
INFO: you can find this dimer in molrep_dimer.pdb

Sol_
 S_Nmon RF TF theta phi chi tx ty tz TFcnt wRfac Scor
 S__2__1 1 100.04 0.56 166.98 0.289 0.371 0.420 18.74 0.579 0.507

--- convert "molrep.crd" to "molrep.pdb" ---

Find Show Log Graphs Show Summary Quit

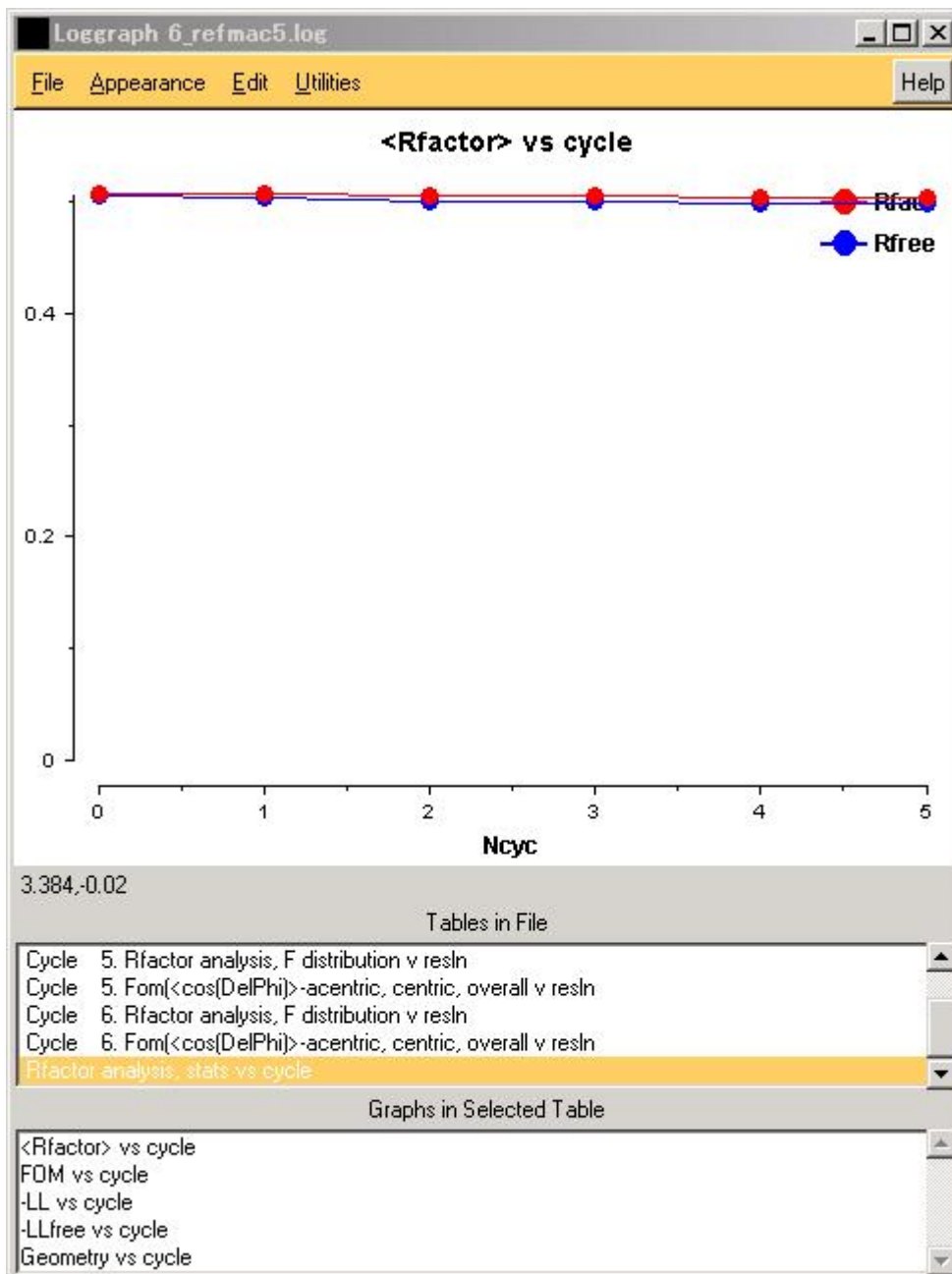
10. Refmac を用いて、まず rigid body 構造精密化 (分子の向きの微調整) を行う。
 作業メニューから Refinement → Run Refmac5 を選択すると、Run Refmac5 のウィンドウが開く。□
 以下のように設定する。□
 Do: rigid body refinement using: no prior phase information
 Input fixed TLS parameters
 no twin refinement
 入力ファイルは以下の 2 つ。□
 MTZ in: 110427: s100a13.mtz (S100A13 の X 線回折データ)
 PDB in: 110427: 1XK4_C_molrep1.pdb (分子置換法で得られた原子座標ファイル。
 詳細説明 : calgranulin B の側鎖を S100A13 のものに置換し
 非対称単位の中で X 線回折データに合うように位置と向きを調整した
 2 分子)
 出力ファイル名は自動で設定される。
 MTZ out: 110427: s100a13_refmac1.mtz
 PDB out: 110427: 1XK4_C_molrep1_refmac1.pdb
 Refinement Parameters で refinement のサイクル数を 20 から 5 に減らしても良い。
 Run → Run Now ボタンを押すと計算が始まる。□



CCP4Interface の中央の作業ログで、refmac5 の行を選択した後、右側の View Files from Job ボタンをクリックし、プルダウンメニューの View Log File をクリックすると計算の過程を追うことができる。□

Job が FINISHED になった後、View Log Graph で構造精密化の過程を視覚的に追えて分かりやすい。□

View Files from Job → View Log Graphs でロググラフを開き、Tables in File → Rfactor analysis, stats vs cycle を選択する。Graphs in Selected Table → <Rfactor> vs cycle で R factor (X 線回折データと立体構造とのずれの指標。小さい値ほど望ましい) が微減したことを確認できる。□



1 1. Refmac を用いて、次に restrained 構造精密化を行う。

作業メニューから Refinement → Run Refmac5 を選択すると、Run Refmac5 のウィンドウが開く。□

以下のように設定する。□

Do: restrained refinement using: no prior phase information

Input fixed TLS parameters

no twin refinement

入力ファイルは以下の 2 つ。□

MTZ in: 110427: s100a13_refmac1.mtz

(rigid body 構造精密化後の X 線回折データ)

PDB in: 110427: 1XK4_C_molrep1_refmac1.pdb

(rigid body 構造精密化後の原子座標)

出力ファイル名は自動で設定される。

MTZ out: 110427: s100a13_refmac2.mtz

PDB out: 110427: 1XK4_C_molrep1_refmac2.pdb

Refinement Parameters で refinement のサイクル数を 10 から 50 に増やした方がよい。

Run → Run Now ボタンを押すと計算が始まる。□

Run Refmac5

Job title

Do restrained refinement using no prior phase information input

Input fixed TLS parameters

no twin refinement

Run Coot:findwaters to automatically add/remove waters to refined structure

MTZ in 100507 s100a13_refmac1.mtz Browse View

FP F_s100a13_01 Sigma SIGF_s100a13_01

MTZ out 100507 s100a13_refmac2.mtz Browse View

PDB in 100507 1XK4_C_molrep1_refmac1.pdb Browse View

PDB out 100507 1XK4_C_molrep1_refmac2.pdb Browse View

LIB in 100507 Merge LIBINS Browse View

Output lib 100507 1XK4_C_molrep1_refmac1.cif Browse View

Include keyword file 100507 Browse View

Data Harvesting

Refinement Parameters

Do 50 cycles of maximum likelihood restrained refinement

Use hydrogen atoms: use if present in file and output to coordinate file

Resolution range from minimum 47.118 to 1.888

Use automatic weighting Use experimental sigmas to weight Xray terms

Refine isotropic temperature factors

Exclude data with freeR label FreeR_flag with value of 0

Use the free set of reflections for fitting the SigmaA estimate

Setup Geometric Restraints

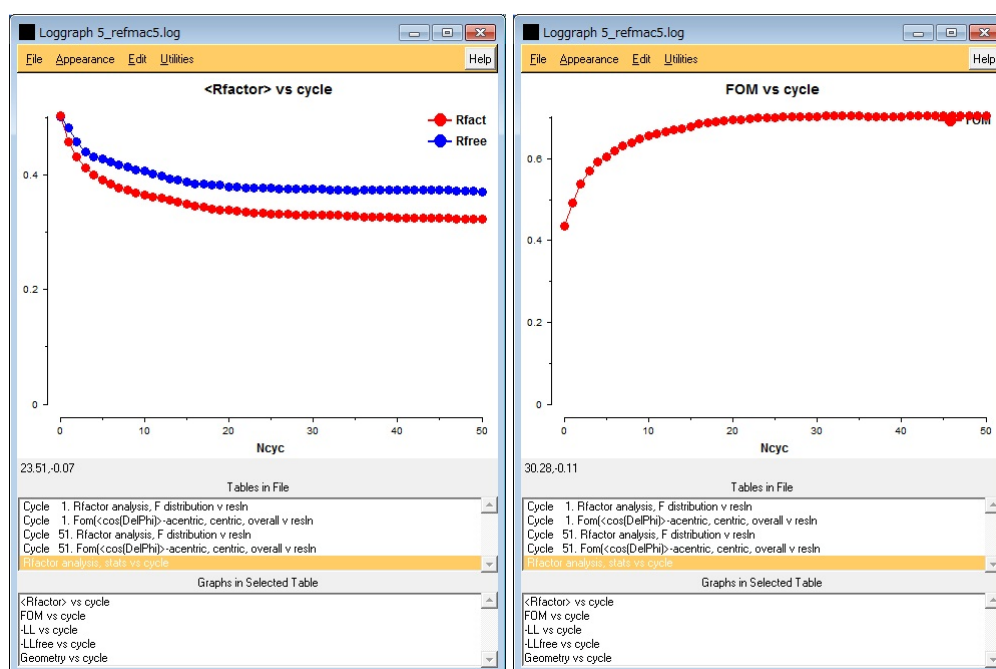
Setup Non-Crystallographic Symmetry (NCS) Restraints

Run Save or Restore Close

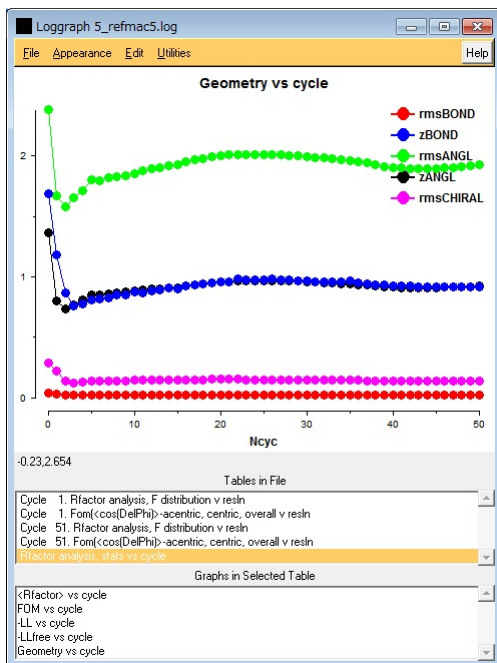
CCP4Interface の中央の作業ログで、2 回目の `refmac5` の行を選択した後、右側の View Files from Job ボタンをクリックし、プルダウンメニューの View Log File をクリックすると計算の過程を追うことができる。□

Job が FINISHED になった後、View Log Graph で構造精密化の過程を視覚的に追えて分かりやすい。□

View Files from Job → View Log Graphs でロググラフを開き、Tables in File → Rfactor analysis, stats vs cycle を選択する。Graphs in Selected Table → <Rfactor> vs cycle でサイクル毎に R factor (X 線回折データと立体構造とのずれの指標。小さい値ほど望ましい) が低下していく様子を確認できる。Graphs in Selected Table → <Rfactor> vs cycle でサイクル毎に FOM vs cycle でサイクル毎に FOM (位相の確からしさの指標。大きい値ほど望ましい) が向上していく様子を確認できる。□



Graphs in Selected Table → Geometry vs cycle で rmsBOND, rmsANGLE, rmsCHIRAL (それぞれ結合長、結合角、不斉性における理想値からのずれ。小さい値ほど良い) が低下傾向にあればなお良い。



Refmac5 を用いた restrained refinement の結果、

R factor は 32%、free R factor は 37%まで下がった。

(構造精密化計算に使用する回折データは全体の 95%。

残り 5%の回折データの R factor を free R factor と呼ぶ。

free R factor は、構造精密化が正しく進んでいるか否かの客観的な指標になる)

FOM は 70%まで上がった。

rmsBOND, rmsANGLE, rmsCHIRA のいずれも初期値より下がった。□

ということで、すべての点において望ましい構造精密化をすることができた。

1 2. さらに構造精密化を進めるために、Coot を用いて、視覚的に、分子モデルを電子密度に合わせていく。□

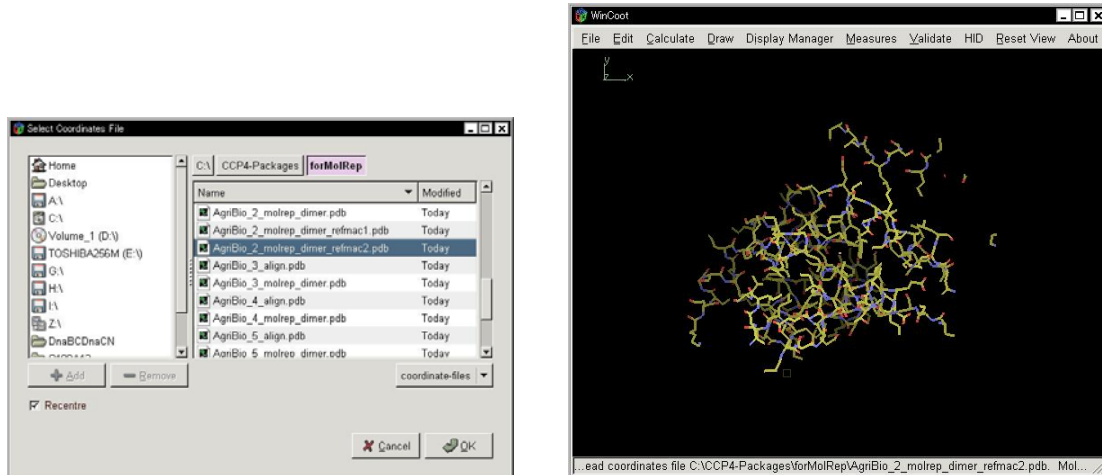
Coot Tutorial で Coot の使い方を一通り説明した後、Run Refmac5 の View from Job → Output files ..の PDB ファイルと MTZ ファイルを使って、立体構造モデルを電子密度に合わせて行きます。

Coot (クロガモ=鳥) アイコンをダブルクリックして、Coot を起動。Close。No。

まず、構造精密化した原子座標ファイルを開きます。

WinCoot: File → Open Coordinates....

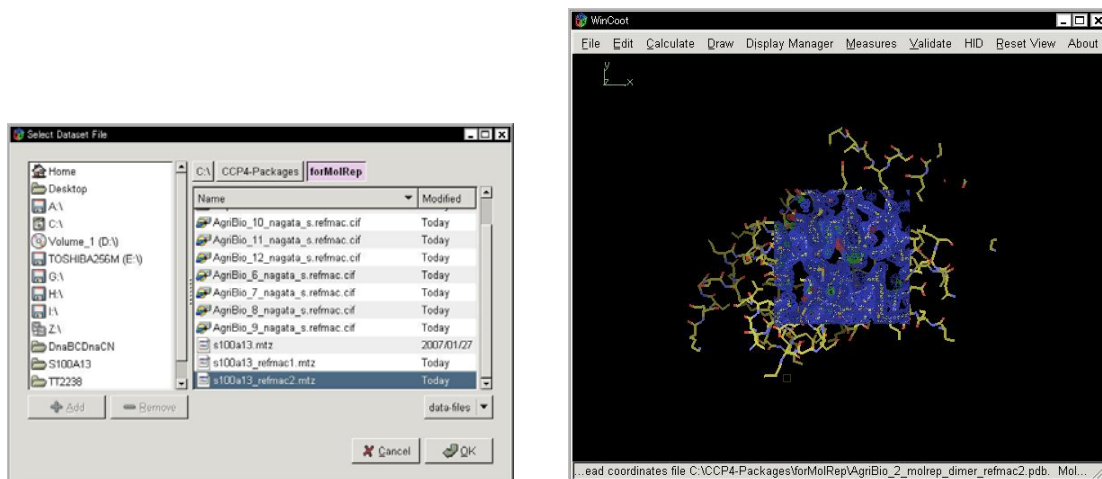
Select Coordinates File: 1XK4_C_molrep1_refmac2.pdb → OK。



次に、精密化した X 線回折データファイルを開きます。

WinCoot: File → Auto Open MTZ....

Select Dataset File: s100a13_refmac2.mtz → OK。



電子密度マップの表示領域を半径 20 Å に設定します。

WinCoot: Edit → Map Parameters....

Global map properties window: Map Radius: 20.0 Angstroem → OK。

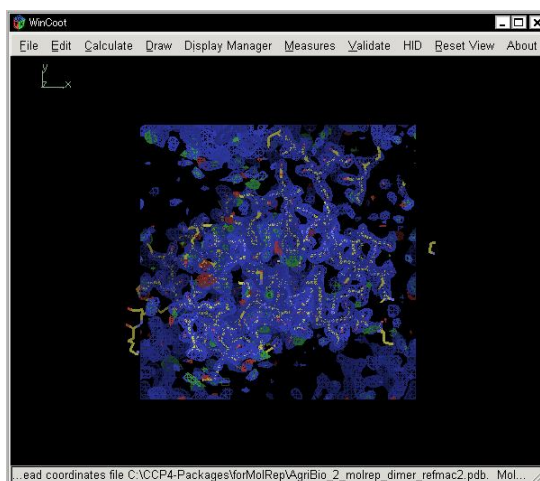
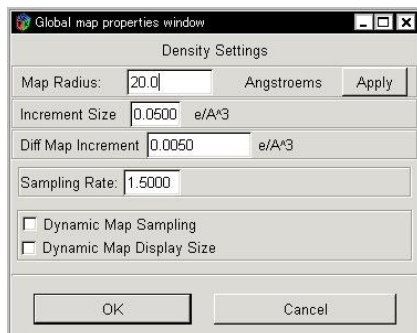
電子密度のうち、

青は、 $2F_o - F_c$ マップと呼び、電子の存在位置を示します。

赤と緑は、 $F_o - F_c$ マップのそれぞれ正と負を示し、

本来電子密度がないはずなのに構造が置かれている場所が 赤
本来電子密度があるはずなのに構造が置かれていない場所が 緑
で示されています。

この赤と緑の電子密度が現れている場所は、構造を修正する必要があるので、N 末端から順に手動で修正して行きます。

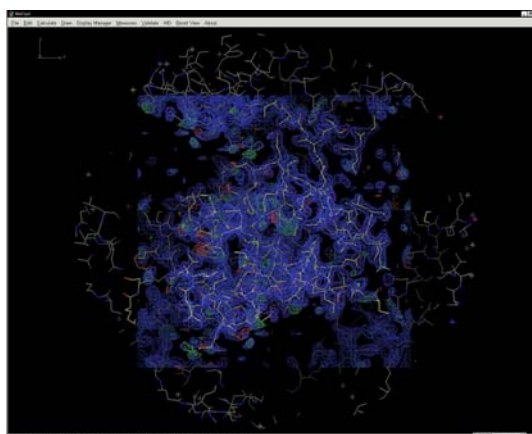
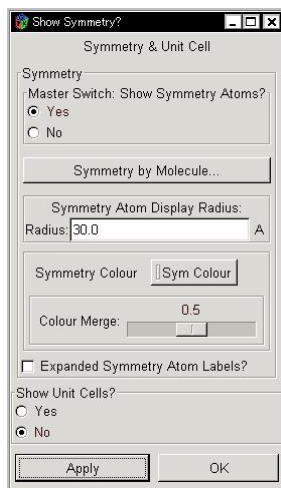


注目している 2 分子の他に、結晶格子中の隣の 2 分子についても半径 30 \AA 以内のものは表示するように設定します。

WinCoot: Draw → Cell & Symmetry...

Symmetry/Master Switch: Show Symmetry Atoms? → Yes。

Symmetry Atom Display Radius: 30 \AA → OK。

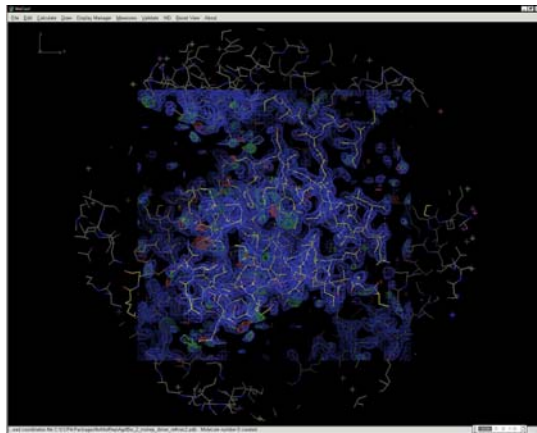
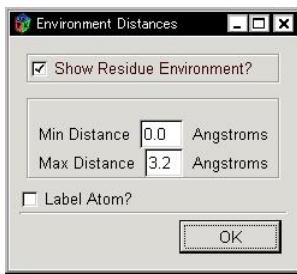


注目している原子のその周囲の原子との距離を表示するように設定します。

WinCoot: Measures → Environment Distances...

Environment Distances: ■ Show Residue Environment?

■ Label Atom? → OK。

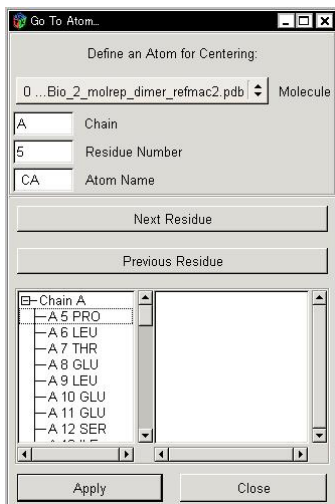


WinCoot: Draw → Go To Atom....

Go To Atom...: +-Chain A → A 5 PRO → Apply → Close。

右ドラッグ (左から右へ) で、指定したアミノ酸残基を中心に拡大する。

右ドラッグ (右から左へ) で、指定したアミノ酸残基を中心に縮小する。



左ドラッグ (上↔下、左↔右) で指定したアミノ酸残基を中心に回転する。

スペースバーを押すと次のアミノ酸残基に移動する。

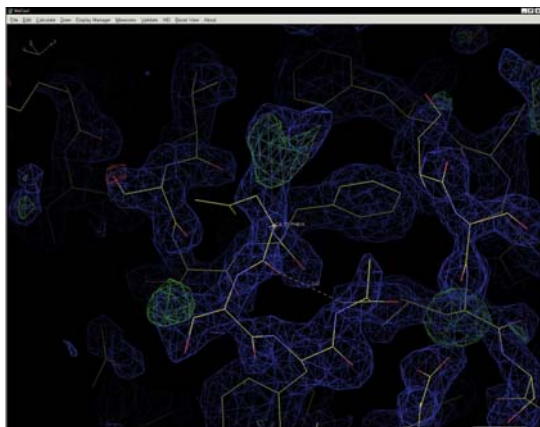
Shift + スペースバーを押すと前のアミノ酸残基に移動する。

ここまでが、Coot の使用法の簡単な説明です。

スペースバーを何回も押して、21 PHE/A まで移動してください。

もし、行きすぎた時は、Shift + スペースバーを押して 21 PHE/A まで戻ってください。

21 PHE/A は、分子モデルの側鎖と電子密度とが合っていないし、側鎖の構造が Leu になっています。これを修正します。



WinCoot: Calculate → Model/Fit/Refine...

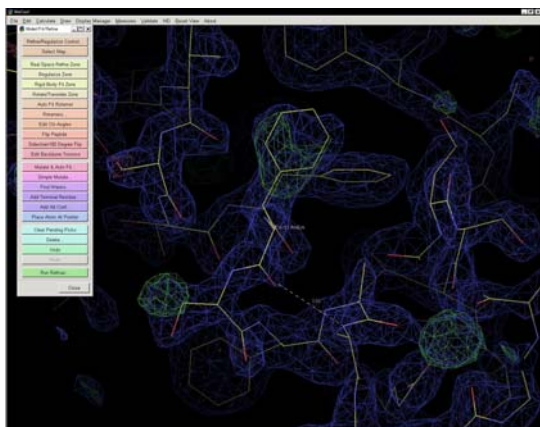
Model/Fit/Refine: Mutate & Auto Fit...

Choose a Map: 1 s100a13_refmac2.mtz FWT PHWT を選択し、OK。

Model/Fit/Refine: Mutate & Auto Fit...

WinCoot: CA/21 PHE/A 原子をクリック。

Resi...: PHE (F)。



側鎖の構造が Phe に修正されて、かつ、分子モデルの側鎖と電子密度とが合いましたか？
合ったことを確認してください。

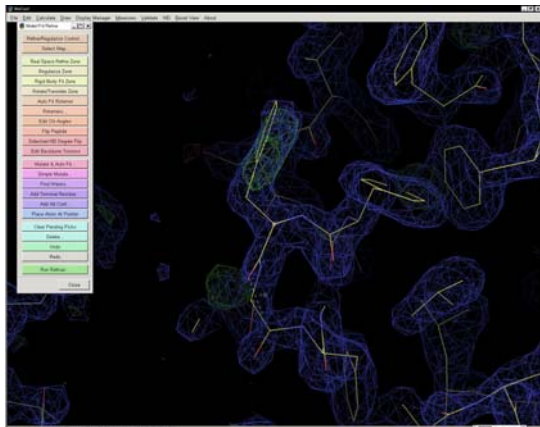
今の方法は簡単過ぎるので、別の方法で合わせてみましょう。

Model/Fit/Refine: Undo を 2 回クリックして、分子モデルの側鎖を元に戻します。

Model/Fit/Refine: Simple Mutate...

WinCoot: CA/21 PHE/A 原子をクリック。

Resi...: PHE (F)。



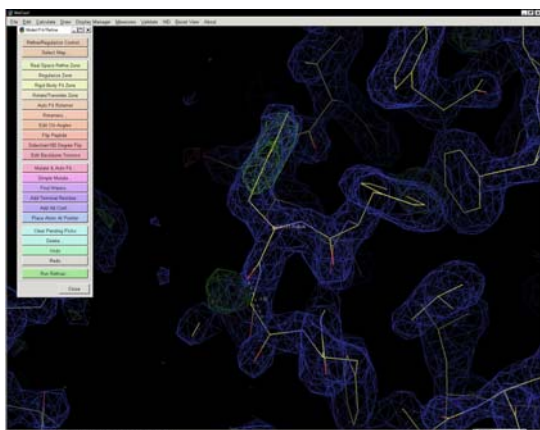
さきほどと違って、分子モデルと電子密度とが微妙にずれています。

Model/Fit/Refine: Real Space Refine Zone。

WinCoot: 21 PHE/A の任意の原子をダブルクリック。

補正後の座標（白で表示される）が補正前の座標（黄色）よりも電子密度に合っていたら、受理する。

Accept Refinement?: Accept。

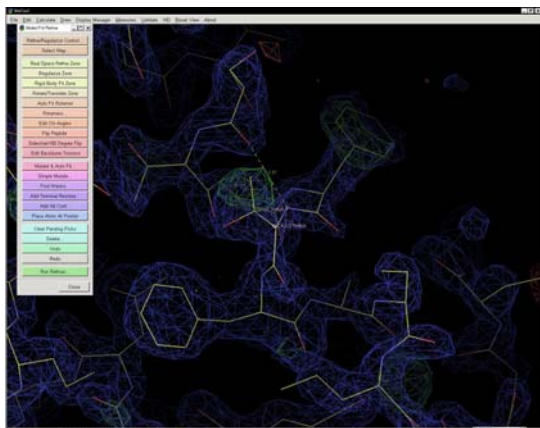


分子モデルと電子密度とが完全に合いました。

スペースバーを押して、次のアミノ酸残基 22 THR/A に進みます。

課題： 22 THR/A の分子モデルの側鎖と電子密度とが合っていません。

これを合わせてください (5 分間 自分でやってみる。質問は受け付けます)。



このようにして、N 末端から C 末端まで、すべてのアミノ酸残基の分子モデルと電子密度とを合わせていきます。

教育的配慮から、N 末端から C 末端まで、1 残基ずつ確認・修正していくべきなのですが、手動で全部やっていると時間も必要で疲れますので、自動で修正する方法も紹介します。

まず、分子置換後の座標は側鎖が不完全なものがありますので、それを修正します。

WinCoot: Extensions → All Molecule → [Post MR] Fill Partial Residues → *****.pdb
→ OK

つぎに、タンパク質の構造を電子密度に自動でフィットさせます。chain A の N 末端から C 末端まで、その後、chain B の N 末端から C 末端まで、1 残基ずつ順次修正してくれます。

2 通りの方法があります。

前者は "Fit Protein using Rotamer Search" です。これは、各アミノ酸側鎖がとりやすい構造が数通りずつ知られているので、その構造 (rotamers) の中から電子密度に一番合うものを選択する方法です。

WinCoot: Extensions → All Molecule → Fit Protein → *****.pdb → OK

後者は "Fit Protein using Real-Space Refinement" です。これは、電子密度に合うように構造を微調整する方法です。

WinCoot: Extensions → All Molecule → Stepped Refine → *****.pdb → OK

上記 2 通りの方法を、この順番で実行しても良いと思います。

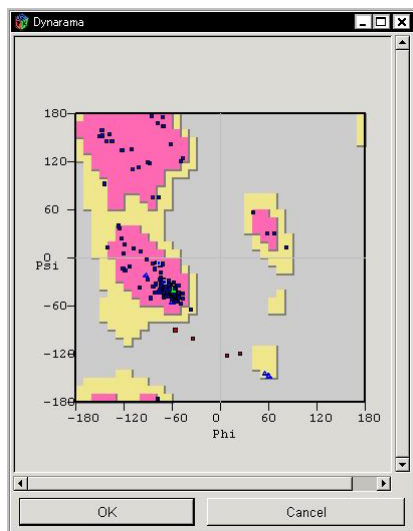
原子座標をできるだけ電子密度に合わせたら、次に、WinCoot: Validate の種々のメニューを使って、立体構造の不適切な箇所を見つけ出し、修正して行きます。

まず、Ramachandran プロットで、主鎖の二面角 (ϕ , ψ) の分布が適切かどうか調べます。不適切な残基は修正します。

WinCoot: Validate → Ramachandran Plot → *****.pdb

Dynarama: Ramachandran Plot (Phi-Psi Plot)で Disallowed Region にあるアミノ酸残基 ■ にカーソルを合わせると、そのアミノ酸残基を表示する。

87 ILE A, 88 ARG A, 8 GLU B, 88 ARG B の 4 残基。いずれもペプチド鎖末端付近のアミノ酸残基なので、修正が難しい。

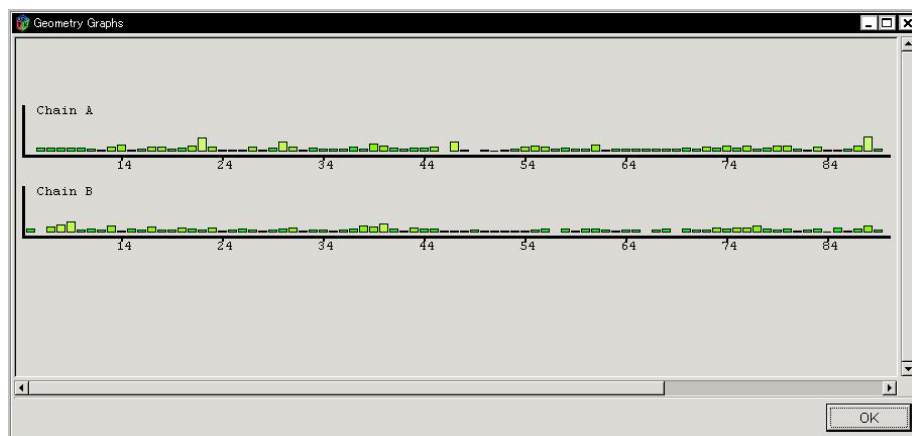


次に、結合長、結合角などの化学構造が不適切な残基を見つけて、修正します。

WinCoot: Validate → Geometry analysis → *****.pdb

Geometry Graphs: 各アミノ酸残基の理想の geometry からのずれが表示されている。

赤いアミノ酸残基があれば、そのバーをクリックし、その残基の分子モデルを修正する。



同様に、Peptide omega analysis、Temp. fact. variance analysis、Rotamer analysis を行います。

すべての項目について **validate** された分子モデルが得られたら、ファイルに保存。

WinCoot: File → Save Coordinates...

Save Coordinates Molecule Selector: Save Molecule Number to Save: 0 *****.pdb →
Select Filename...

Save Filename for Saved Coordinates: Name: デフォルトのまま (*****-coot-0.pdb) →
Save in folder: CCP4 で指定したフォルダ → OK

修正された分子モデルを使って、**Refmac5** によりさらに構造精密化すると、**R factor** および **free R** の値が以前より小さくなっている (改善されている) はずです。

その後、小さな電子密度にリガンドや水分子を当てはめ、**Refmac5** で精密化し、最終構造を求めることで、立体構造解析が完了します。

そして、得られた原子座標ファイルと X 線回折データファイルとを **Protein Data Bank** に登録します。

以上。

提出課題 1 以下の実習で最終的に作成した PDB ファイルを
aknagata@mail.ecc.u-tokyo.ac.jp に送ってください。
不完全な状態でもかまいません。

提出課題 2 講義の感想を 10 行以上書いてください。
良かったことでも、厳しいご意見でもかまいません。