MX Tutorial 2

Determination of the crystal structure of Hen Egg White Lysozyme by the Molecular Replacement method

Pedro Matias, Célia Romão, José Brito

In this tutorial, the crystal structure of Hen Egg White Lysozyme will be determined by the Molecular Replacement method, using as phasing model the 3D structure of the related protein from *Bos taurus* (cow).

ECS7 - 7th European Crystallographic School – July 12, 2022

Getting started

- Login to your account and start a shell terminal;

- Type the command:

  %cd ECS7/Lysozyme ↵

- Check that the tutorial files are present in this directory:

  % ls ↵

  2z2f-model.pdb – coordinates of the Lysozyme from cow

  alignment.aln – alignment between the sequence of both proteins (cow lysozyme on top)

  cu_Lyso_test1_1_0m.prp – logfile with the data processing statistics

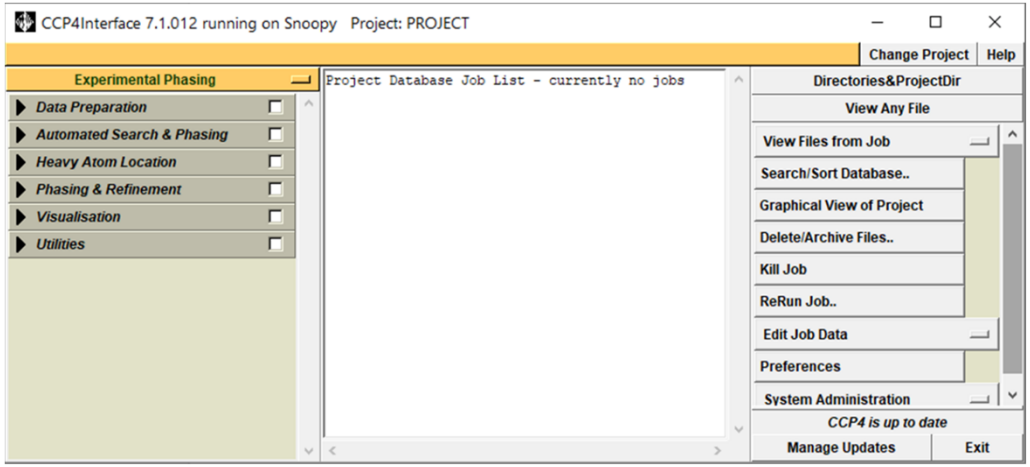
  cu_Lyso_test1_1_0m.sca – diffraction data

  Lysozyme.pir - aminoacid sequence in 1-letter format

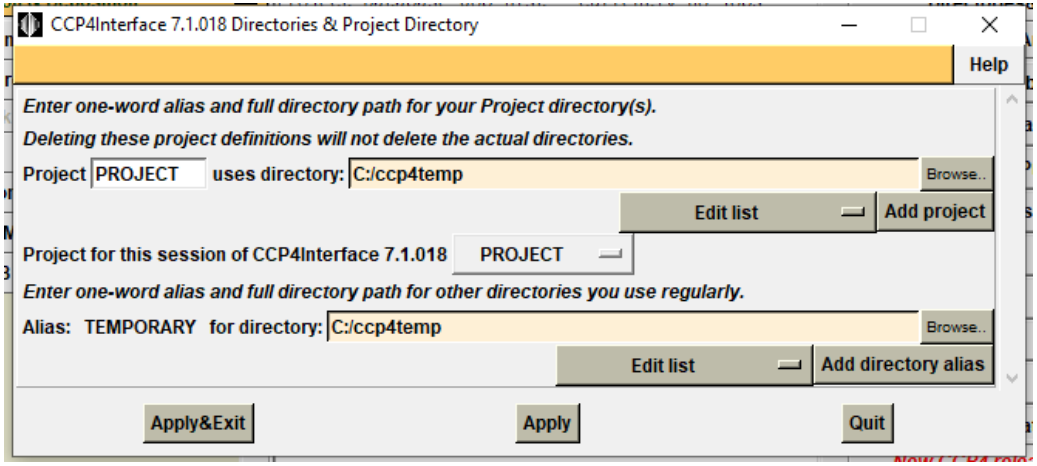
Start CCP4i and create a new project

- Start CCP4i

```
% ccp4i & ↵
```

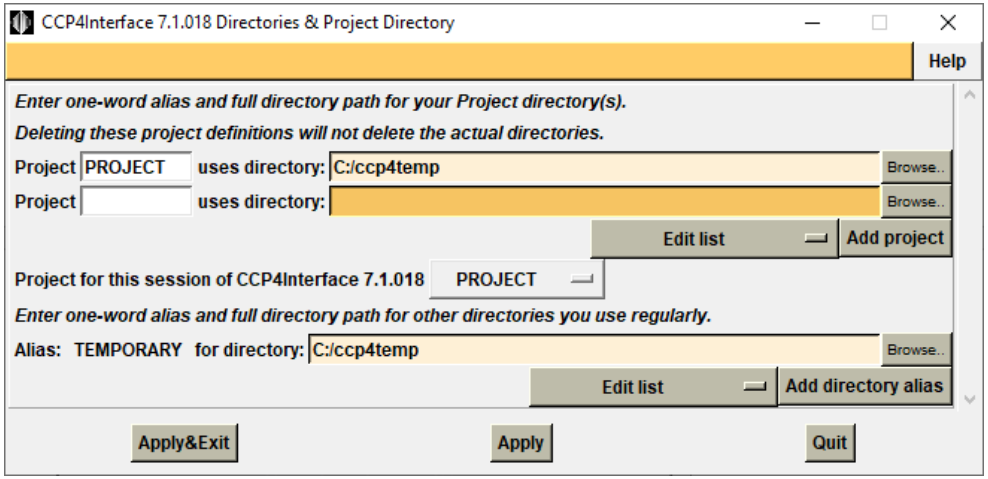


- Click on the **Directories&ProjectDir** button

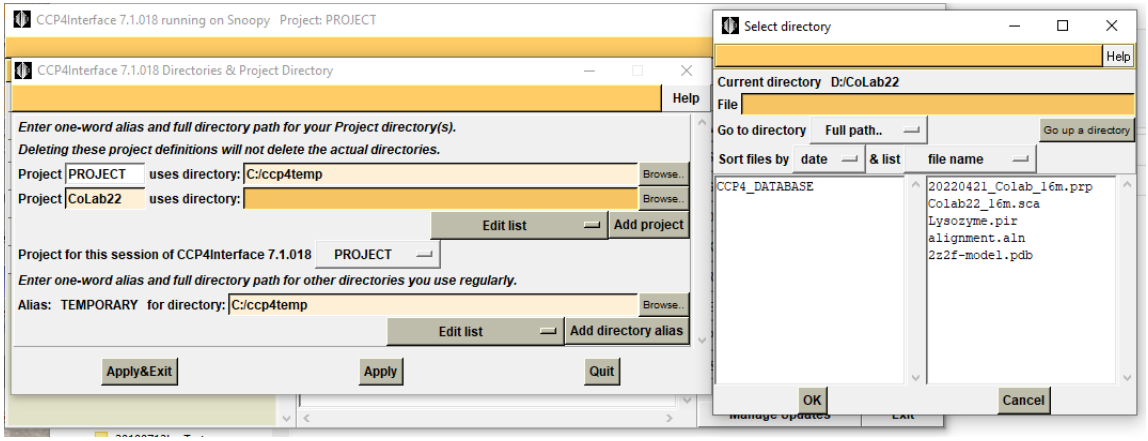


Start CCP4i and create a new project

- Click on the Add Project Button

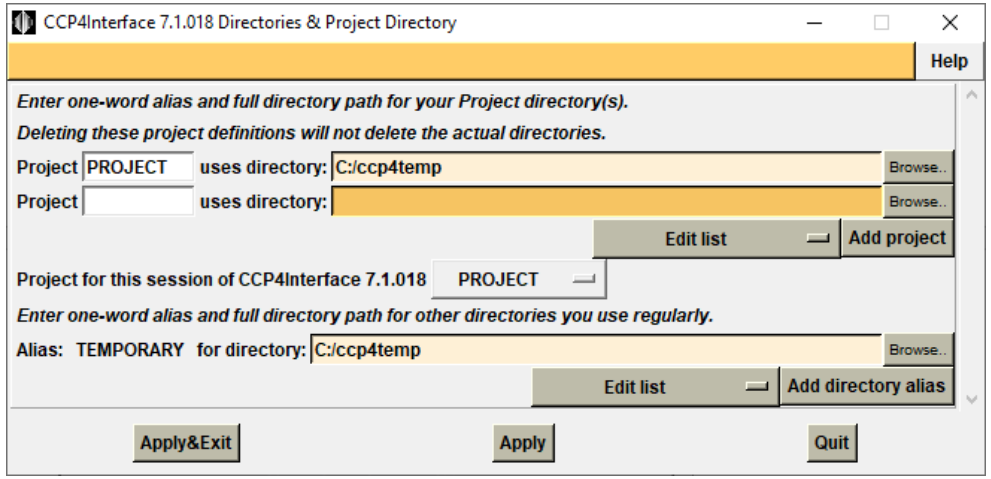


- In the Project box you can type a word to name the project (e.g., ECS7). Click on the **Browse** button near the empty box and select **Full Path** as shown. You should see the files listed. Click **OK** to close the directory selection window. Then, click on the **Project for this session** button, select the name you gave to this project and click **Apply&Exit**

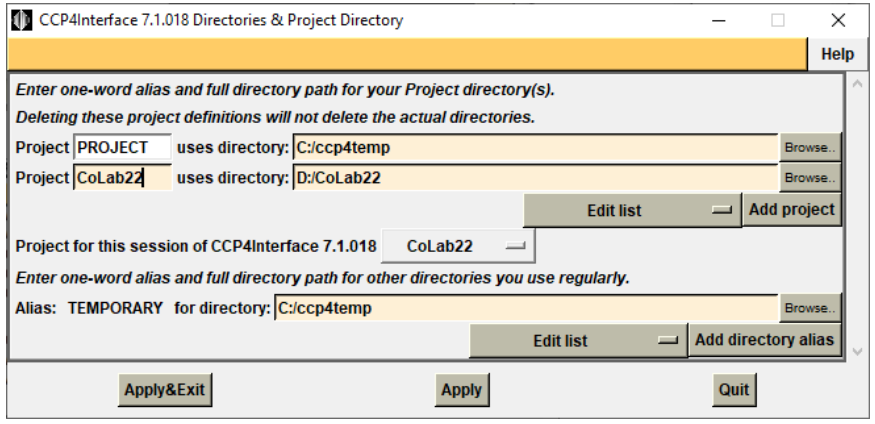


Start CCP4i and create a new project

- Click on the Add Project Button

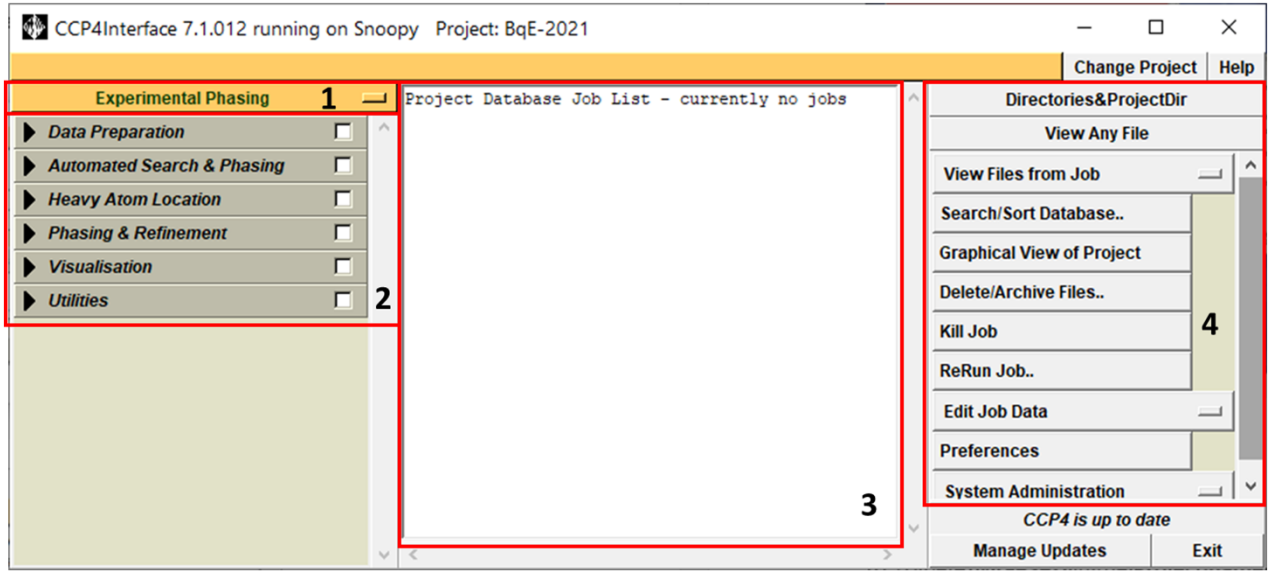


- In the Project box you can type a word to name the project (e.g., ECS7). Click on the **Browse** button near the empty box and select **Full Path** as shown. You should see the files listed. Click **OK** to close the directory selection window. Then, click on the **Project for this session** button, select the name you gave to this project and click **Apply&Exit**



Overview of the CCP4i interface

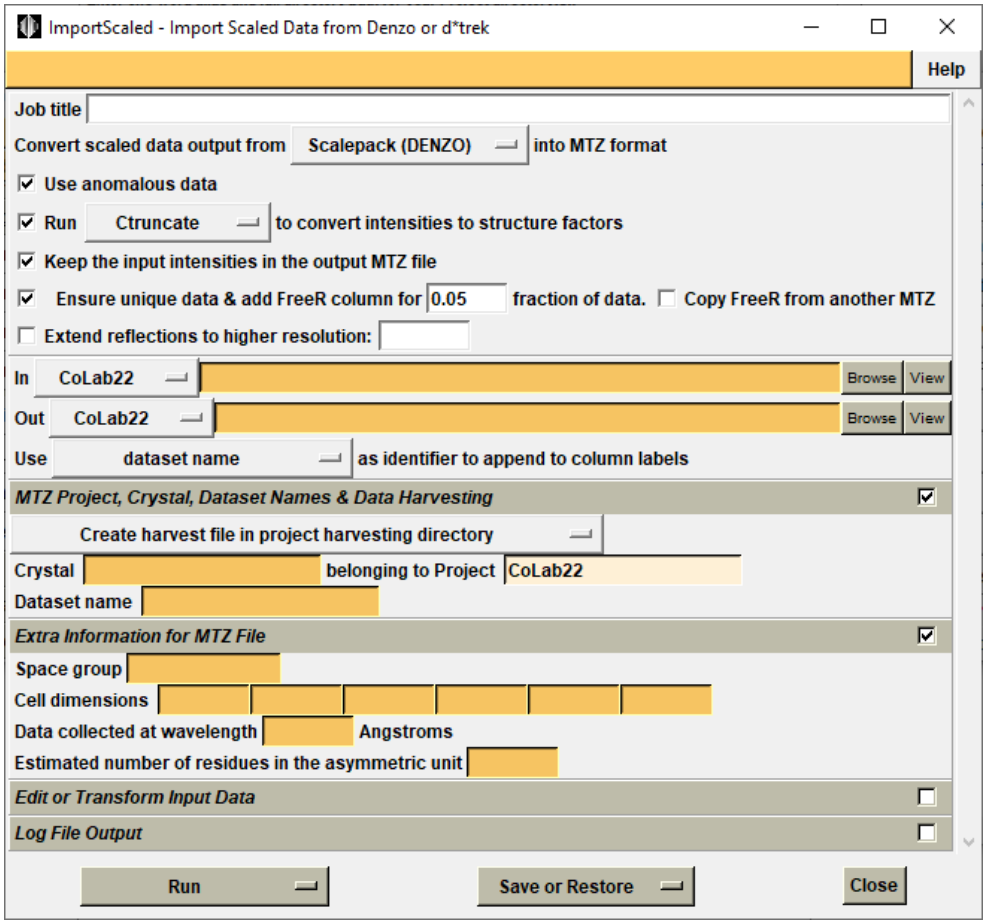
This is what the main window of the CCP4 GUI looks like:



1. *This button selects groups of programs;*
2. *Buttons in this window open a program list or select a program in the group;*
3. *This window lists all the run and running jobs in the current project;*
4. *This window contains buttons for acting on jobs and interface management;*

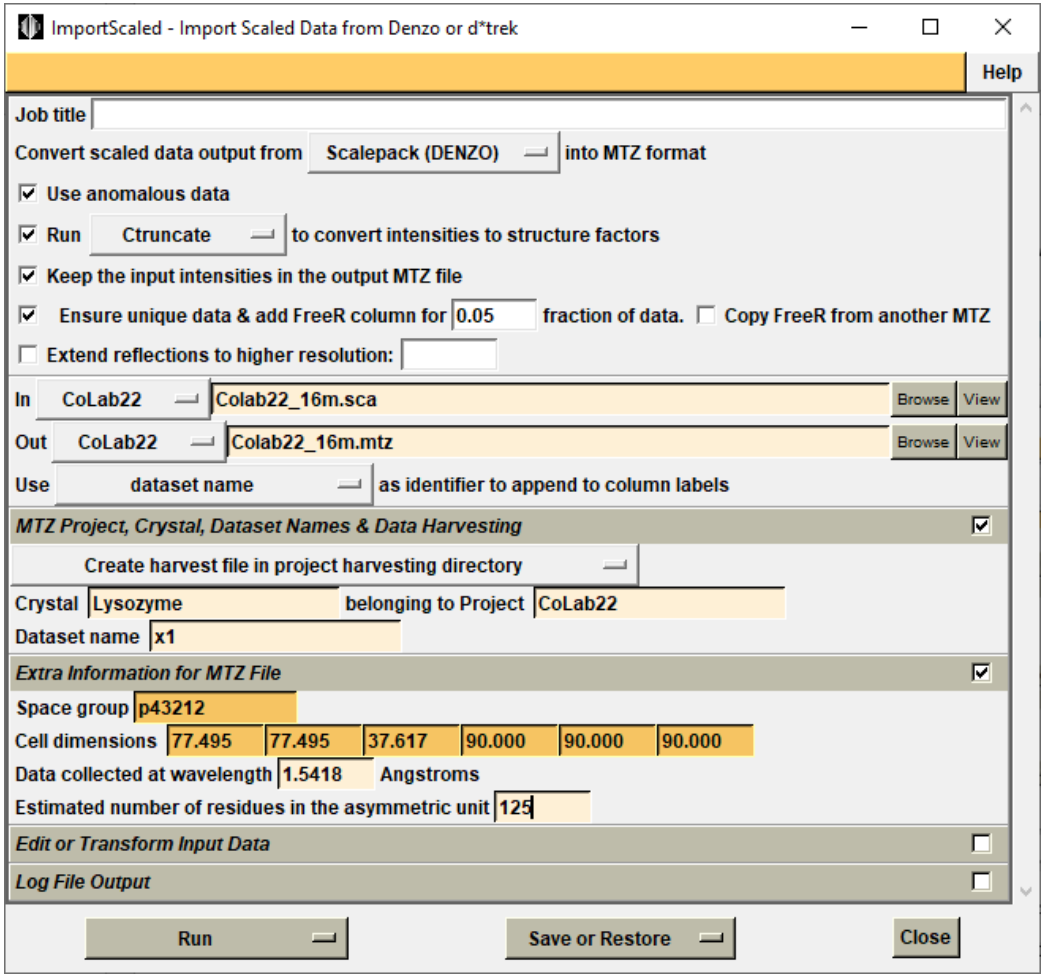
Importing the diffraction data into CCP4

Select the **Data Reduction and Analysis** group of programs in **1**, click on the **Import Integrated Data** subgroup and then select the **Import Merged Data** task:



Importing the diffraction data into CCP4

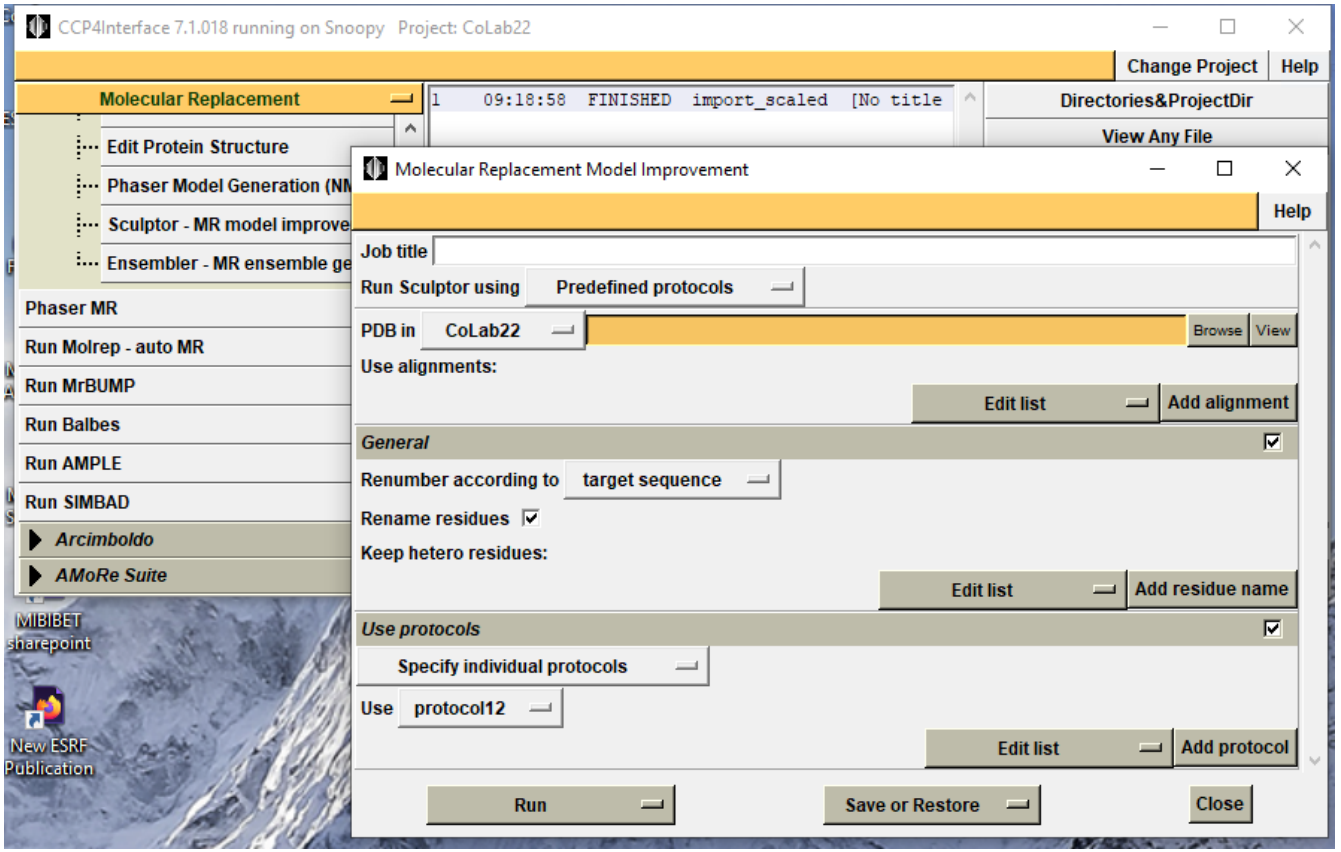
Load the **.sca** file and complete the remaining empty boxes as shown :



Click on the **Run** button, select **Run Now** from the drop-down menu and then click the **Close** button. When the program is **FINISHED** in the central GUI window you have now converted the .**sca** (text) file to the **.mtz** (binary) format used in CCP4 and most crystallographic programs.

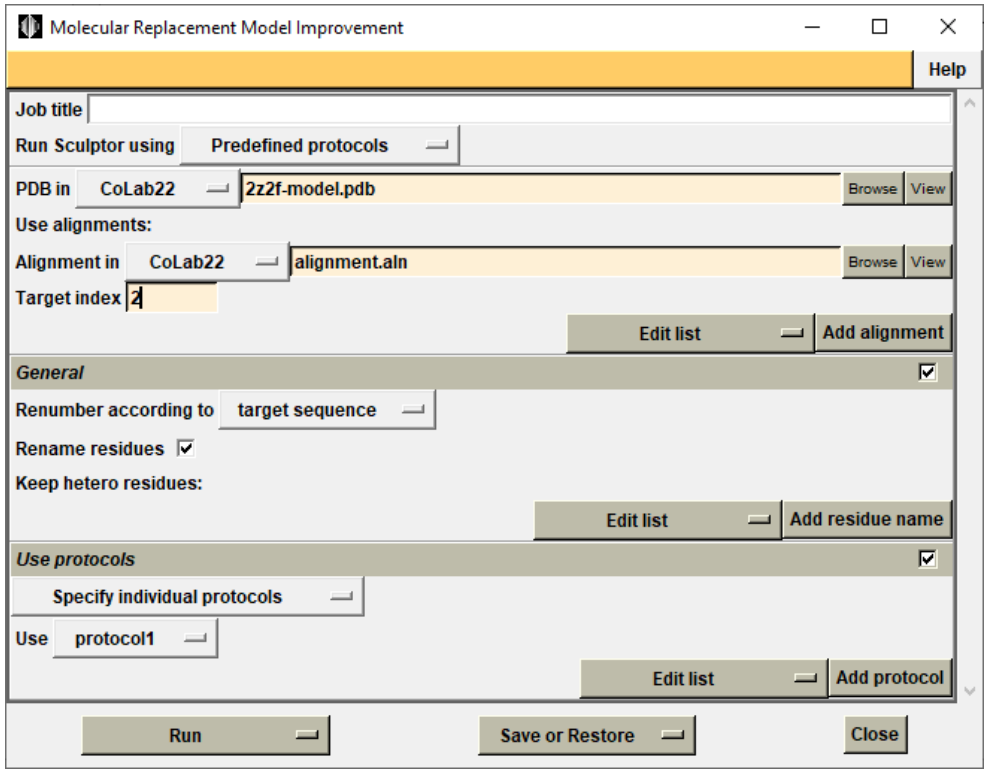
Preparing the bos taurus model for MR

Select the **Molecular Replacement** program group, click on the ***Model Generation*** line in **2** to expand its program list and then click on the **Sculptor** button:



Preparing the bos taurus model for MR

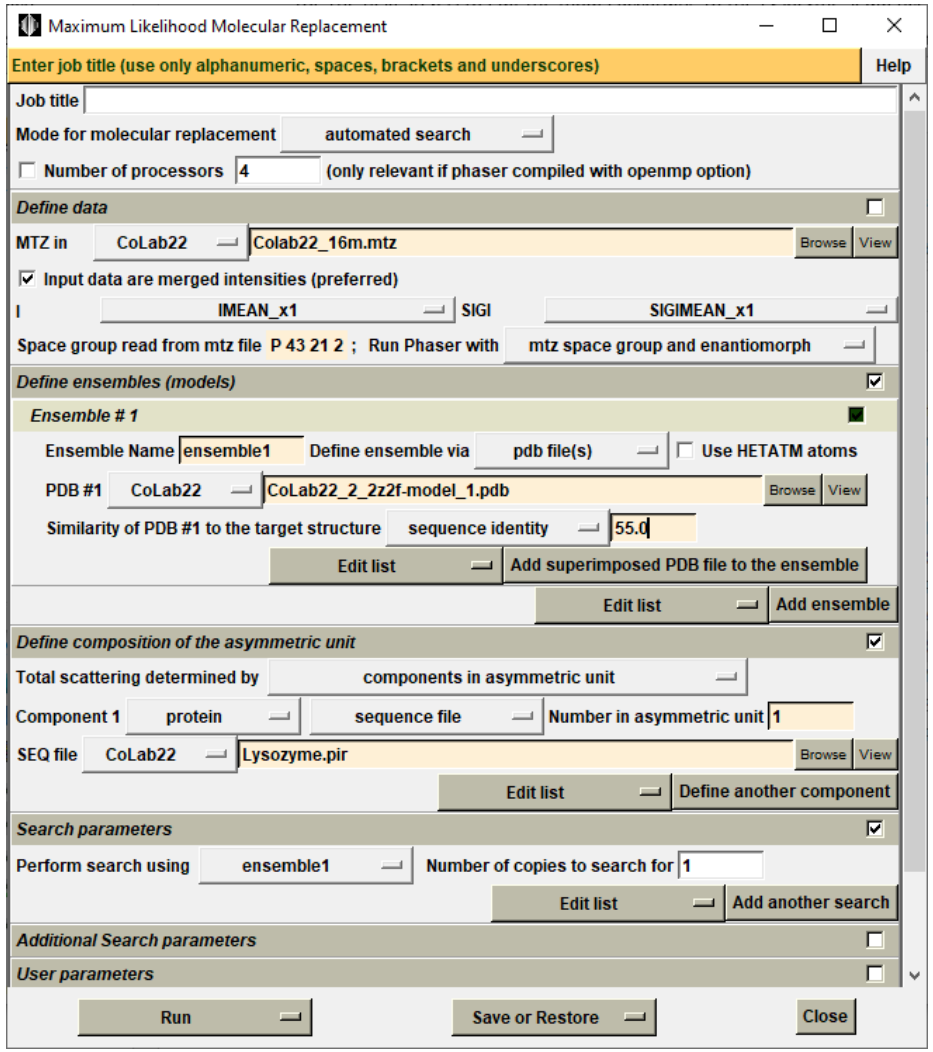
Select the **Molecular Replacement** program group, click on the ***Model Generation*** line in **2** to expand its program list and then click on the **Sculptor** button:



Load the **2z2f-model.pdb** file in the **PDB in** window, click on the **Add alignment** button and load the **alignment.aln** file in the **Alignment in** window. Use **Target Index 2** and select **Specify individual protocols** and **protocol1** to be used in the calculation. Click on the **Run** button and select **Run Now** from the drop-down menu and then click the **Close** button.

You should now have a **new PDB file** in your project directory which has been edited to remove the largest expected differences with your unknown 3D structure, based on the sequence alignment.

Running the MR calculations



Close the ***Model Generation*** program group and click on **Phaser MR** to open a new window. Load the **Colab22_16m.mtz** file into the **MTZ in** window; notice the program automatically selects the measured structure factor intensities and their estimated errors.

Load the PDB file from Sculptor into the **PDB#1** window and specify a **sequence identity** of **55.0**% (from the PDB file header). Load the sequence file **Lysozyme.pir** into the **SEQ file** window. Scroll down the job window and select **ensemble1** in the Search parameters pane.

Click on the **Run** button and select **Run Now** from the drop-down menu and then click the **Close** button.

Once the program is FINISHED, click on **View Files from Job** and then on **View Result of MR in Coot** to Load the output PDB and MTZ files in Coot and inspect the results. *Can you see any problems with the model and/or electron density?*

How do I know MR worked?

```
Top Peaks Without Clustering
Select peaks over 67.5% of top (i.e. 0.675*(top-mean)+mean)
   Also store peaks over 52.5% of top
There were 116 sites over 52.5% of top
116 peaks selected
The sites over 52.5% are:
#      Euler1 Euler2 Euler3    FSS    Z-score
1        64.4   73.0  168.4    100.000  6.05
2        64.2   72.5  167.0     98.670  5.97
3        64.0   72.0  165.6     97.471  5.90
#Sites = 116: output truncated to 3 sites
```

A Z-score **higher than 4** for the rotation function calculations.

```
Top Peaks Without Clustering
Select peaks over 67.5% of top (i.e. 0.675*(top-mean)+mean)
There were 8 sites over 67.5% of top
8 peaks selected
The sites over 67.5% are:
#      Frac X Frac Y Frac Z    FSS    Z-score
1       0.237  0.754  0.458     465.14 16.13
2       0.238  0.754  0.215     259.27  8.91
3       0.238  0.754  0.367     258.07  8.87
#Sites = 8: output truncated to 3 sites
```

A Z-score **higher than 8** for the translation function calculations.

```
Packing Table
-------------
Solutions accepted if pairwise clashes less than 10 % of trace atoms
#in   #out Clash-% Symm TF-SET  ROT TFpk#        TF     TFZ    SpaceGroup
1     Top1  0.428  --       1     1    1        222.42 16.13  P 43 21 2
2     2     9.519  --       1     1    2         10.43  8.69  P 43 21 2
3     3     1.604  --       1     1    3         -8.82  8.87  P 43 21 2

3 accepted of 3 solutions
   3 pack of 3 accepted solutions
```

Very few or no **packing clashes**.

```
Refinement Table (Sorted)
-------------------------
Refinement to full resolution
#out =#out #in =T (Start LLG Rval TFZ) (Refined LLG Rval TFZ==) SpaceGroup  Cntrst
Top1 ---   1           494.7 50.8 n/a         494.8 50.8 24.8   P 43 21 2    n/a
```

The **R-value** is not very informative because there is no scaling between F_o and F_c

How do I know MR worked?

```
   $TEXT:MR Result: $$ Baubles Markup $$

** SINGLE solution

** Solution written to SOL file:  /home/users/matias/ECS7/Lysozyme/ECS7_Lyso_3.sol

** Solution written to PDB file:  /home/users/matias/ECS7/Lysozyme/ECS7_Lyso_3.1.pdb
** Solution written to MTZ file:  /home/users/matias/ECS7/Lysozyme/ECS7_Lyso_3.1.mtz
   Solution annotation (history):
   SOLU SET  RFZ=6.1 TFZ=16.1 PAK=0 LLG=275 TFZ==16.5 LLG=495 TFZ==24.8 PAK=1 LLG=495 TFZ==24.8
   SOLU SPAC P 43 21 2
   SOLU 6DIM ENSE ensemble1 EULER  151.4   71.6  168.4 FRAC -0.25 -0.25  0.22 BFAC -0.26 #TFZ==24.8
   SOLU ENSEMBLE ensemble1 VRMS DELTA +0.0143 #RMSD  0.76 #VRMS  0.77
```

This is also a good sign.

```
Number LLG Z-Score
$$ loggraph $$
1      20.00  6.13
2       5.18  4.24
3       4.02  4.10
$$
```

Rotation function: 20.00

```
$$
Number LLG Z-Score
$$ loggraph $$
1     222.42 16.13
2      10.43  8.69
3      -8.82  8.87
$$
```

Translation function: 222.42

```
$$
Number final-LLG initial-LLG final-R initial-R
$$ loggraph $$
1     274.60     222.42 39.21 41.49
2      83.50      10.43 47.60 49.12
3      76.13      -8.82 48.34 50.23
$$
```

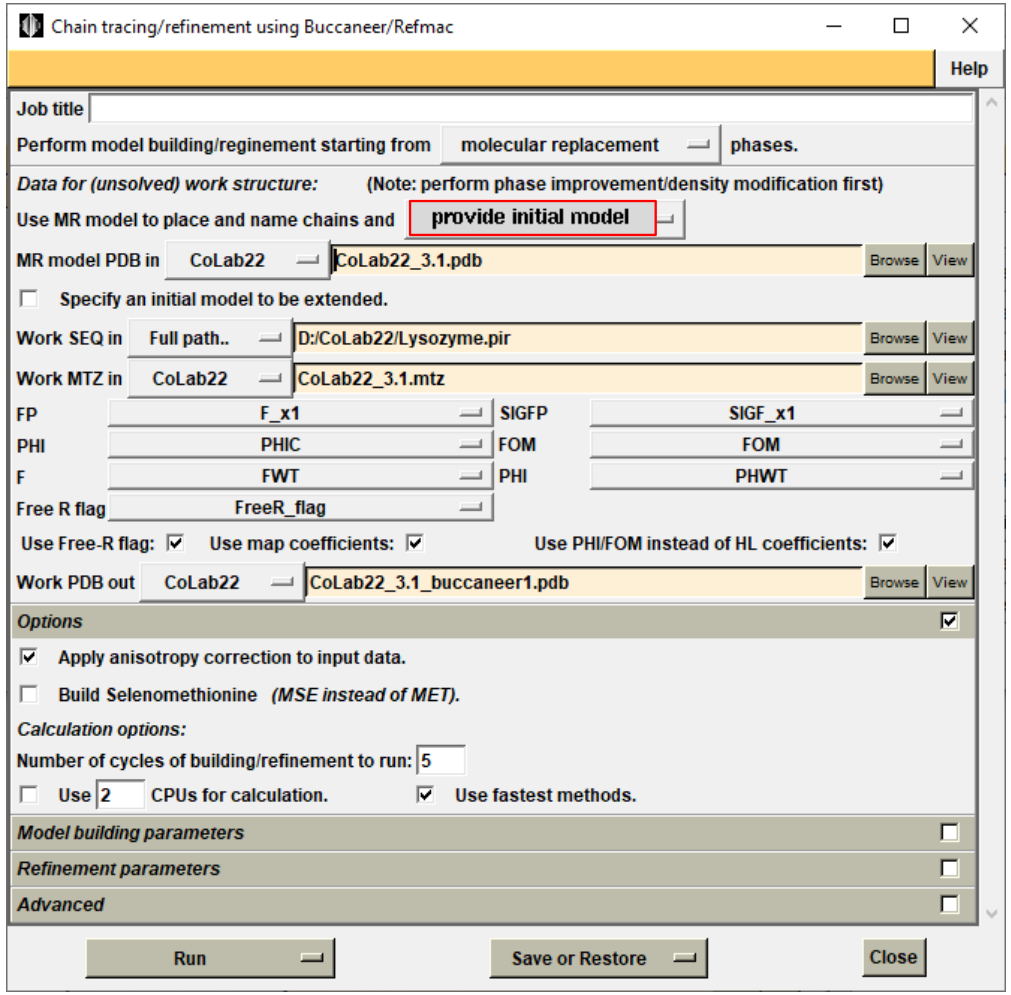
Refinement to 3.8 Å resolution: 276.40

```
$$
Number final-LLG initial-LLG final-R initial-R
$$ loggraph $$
1     494.81     494.72 50.82 50.82
$$
```

Refinement to 1.68 Å resolution: 494.81

The LLG (Log-likelihood gain) **must increase** during the calculations

Automated model editing and rebuilding with buccaneer

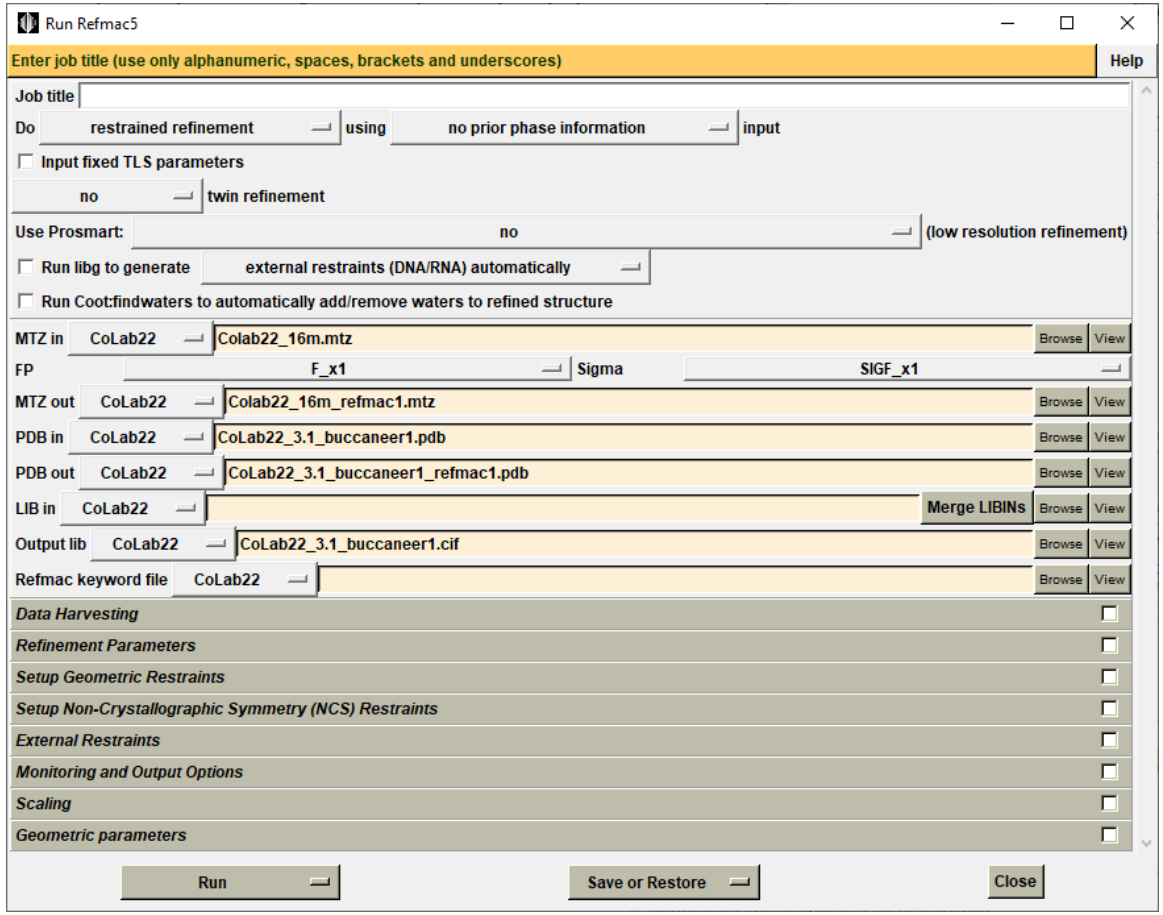


The next step is to edit the model according to the Lysozyme sequence. Remember, only **55%** of the aminoacids are identical between the two structures. This can be done manually in Coot or in a more automated way using **Buccaneer**.

Select the **Model Building** program group, click on the **Buccaneer** button and fill out the data as shown (be sure to select the **Molecular Replacement** mode in the first line):

Click on the **Run** button and select **Run Now** from the drop-down menu and then click the **Close** button. When the program is FINISHED you can inspect the results in Coot.

Automated model editing and rebuilding with buccaneer



To make life easier, let us first do a round of refinement using **Refmac5**. Select the **Refinement** program group, click on the **Run Refmac5** button and fill out the data as shown (you only need to supply the datafiles for the “MTZ in” and “PDB in” boxes):

Click on the **Run** button and select **Run Now** from the drop-down menu and then click the **Close** button. When the program is FINISHED you can inspect the results in Coot. Click on **View Files from Job** and then on **View result of refinement in Coot**.

NEXT STEPS – correct the model in Coot (today) and continue the refinement in Refmac5 (tomorrow)