

Using Molrep for fitting into the electron density

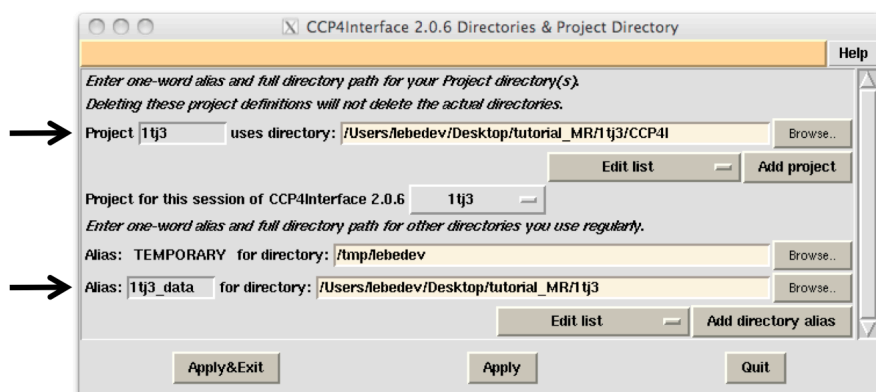
This tutorial demonstrates how to use Phased Translation Function (PTF), Spherically Averaged Phased Translation Function (SAPTF) and Phased Rotation Function (PRF) implemented in Molrep for fitting a model into the electron density maps. The X-ray data used in this tutorial are from the crystal of cyanobacterial sucrose-phosphatase (PDB code 1tj3). The template structure is from a crystal structure of the same enzyme but in different conformation (PDB code 1s2o). The tutorial presents a typical situation, in which a larger part of the structure is easily solved by Molecular Replacement (MR) using the Patterson search, the solved substructure is refined, and then smaller domain(s) are fitted into the electron density maps obtained after refinement. In many such cases the electron density for missing domains is quite poor and is not suitable for automatic or manual model building.

Another situation, when fitting of a model into the electron density can be useful is when only poor experimental maps and a low homology model are available. Even if these maps and model are by themselves useless for, respectively, de-novo model building and molecular replacement, the fitting of the model into the experimental maps could still be possible.

Next important implementation is the fitting of the model into the EM maps, and there is a separate Molrep tutorial on combination of EM and X-ray data, http://www.yesbl.york.ac.uk/~alexei/downloads/tutorial_EM.tar.gz

1. Tutorial data

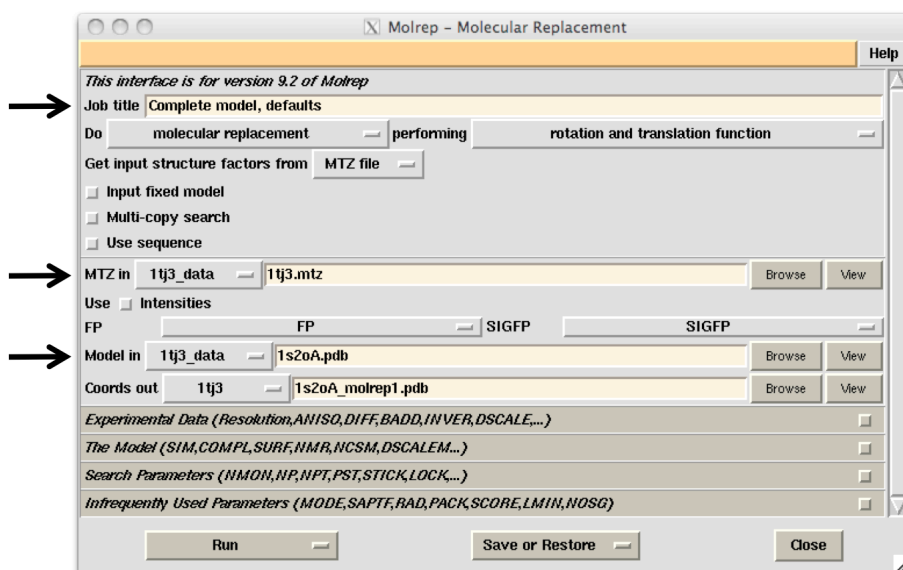
- Download the file **tutorial_MR.tar.gz** from http://www.yesbl.york.ac.uk/~alexei/downloads/tutorial_MR.tar.gz and unpack it in a directory **<path>**.
- Create the directory **<path>/tutorial_MR/1tj3/CCP4I**, where all the generated data will be stored
- Start CCP4I and open Directories and Project Directory window
- Define project **1tj3** with the project directory **<path>/tutorial_MR/1tj3/CCP4I**
- Define alias **1tj3_data** for the directory **<path>/tutorial_MR/1tj3**



2. Using complete model and default protocol

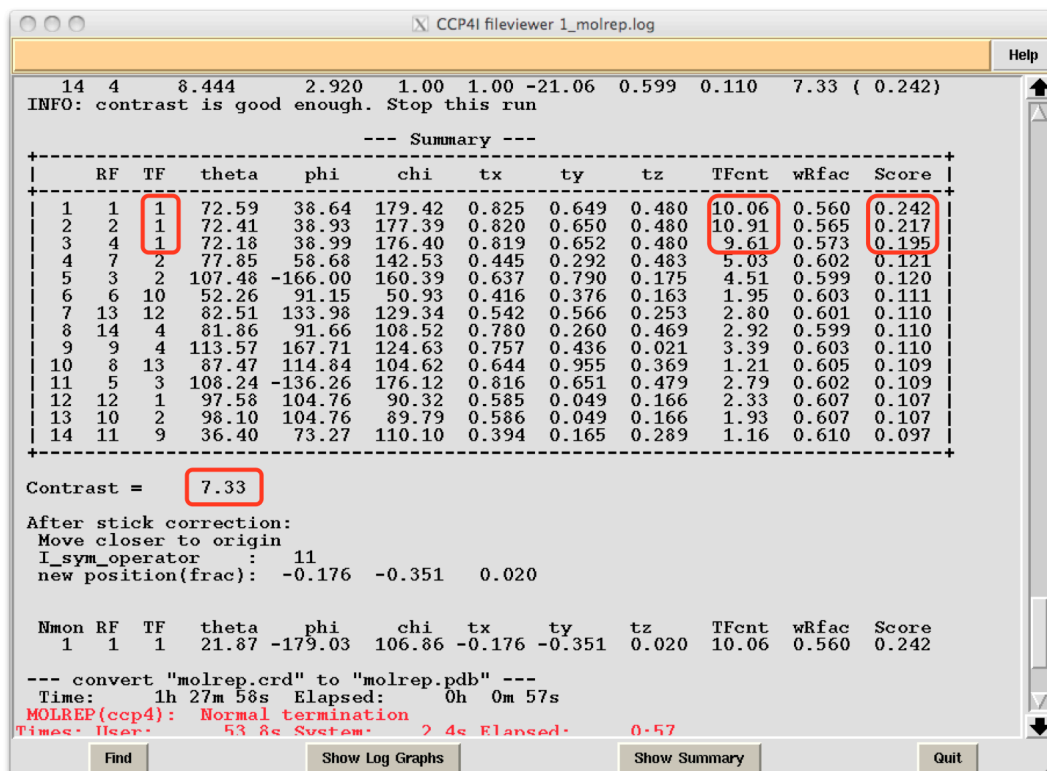
- Open Molrep task window (from the "Molecular Replacement" folder in the CCP4I task menu)
- Type in the title.
- Select **1tj3_data** in the menus next to "MTZ in" and "Model in" labels.
- Browse the input X-ray data, **1tj3.mtz** and model, **1s2oA.pdb** in **1tj3_data** directory.

- Click mouse within the "Model in" text box and press tab button to set the default name for the output pdb-file, **1s2oA_molrep1.pdb**.
- Run



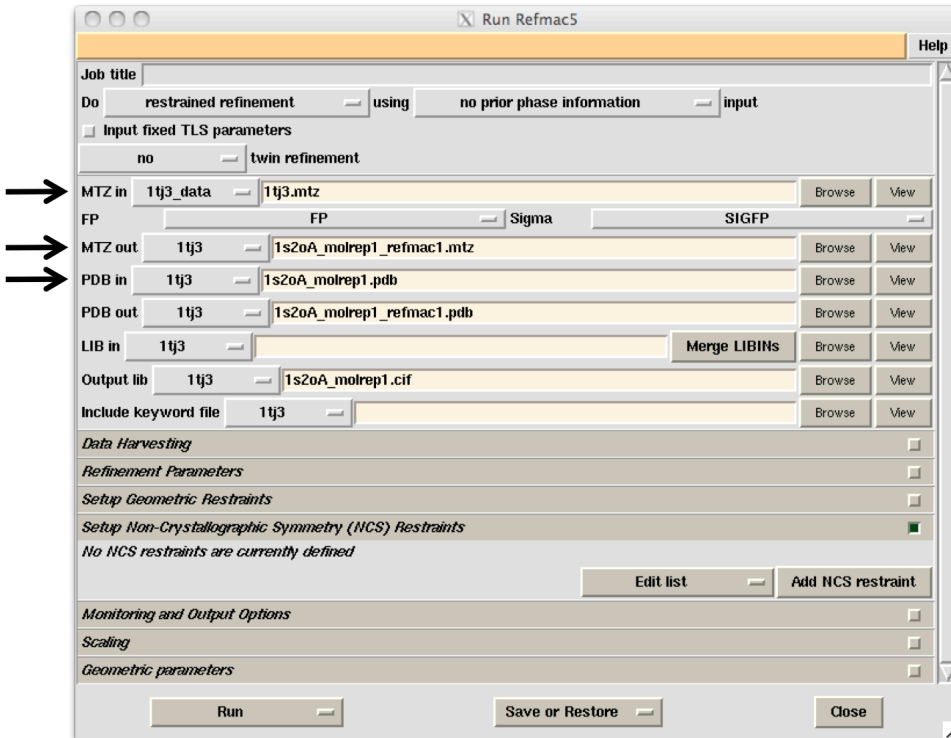
- In the main CCP4I window, select "View Files from Job > View log file". The image below shows a fragment of the log-file including summary table.

The first thing to look at is the contrast, a number shown just below the summary. Roughly, this number means the difference between the highest and the mean score expressed in terms of the e.s.d. units. In most cases without pseudo-translation, the values higher than 3 mean that a correct solution is found. In this case, there are three very close solutions shown in the first three lines of the summary table (compare *theta*, *phi*, *chi*, *tx*, *ty* and *tz* values for this solutions). These three have substantially higher score (correlation coefficient times packing function) and higher values of *TFcont* (contrast in the translation function) than other solution. Also important is the number in the column *TF*, where "1" means that, for this particular orientation, the TF solution with the highest score value is also the best in terms of the translation function.

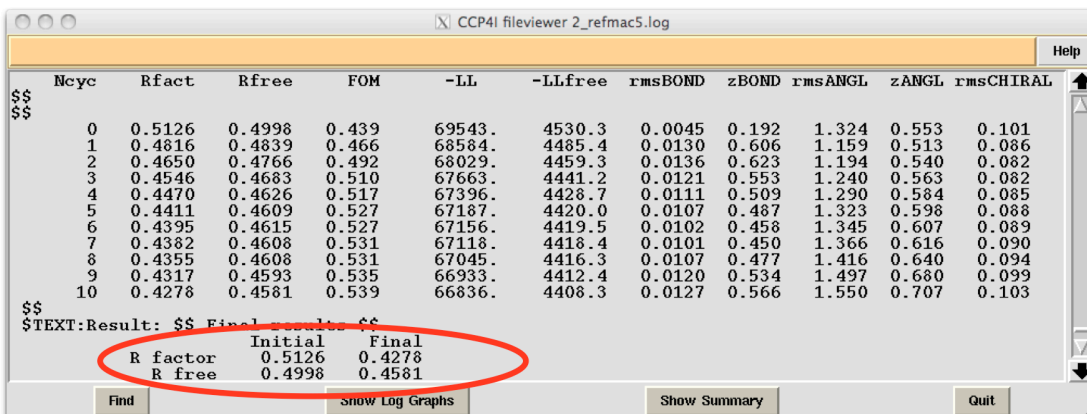


All these data suggest that the correct solution is found. However, a more important evidence is the behaviour of refinement and, of course, the electron density.

- In the task menu of the main ccp4i window select "Refinement > Run Refmac5".
- In the Refmac task window, select input files as shown in figure below
- To avoid confusions, change the default name of output mtz-file to the name consistent with the name of output pdb-file.



Inspect Refmac's log file. What we expect is the decrease in both R and R-free, similar to the one shown in figure below.

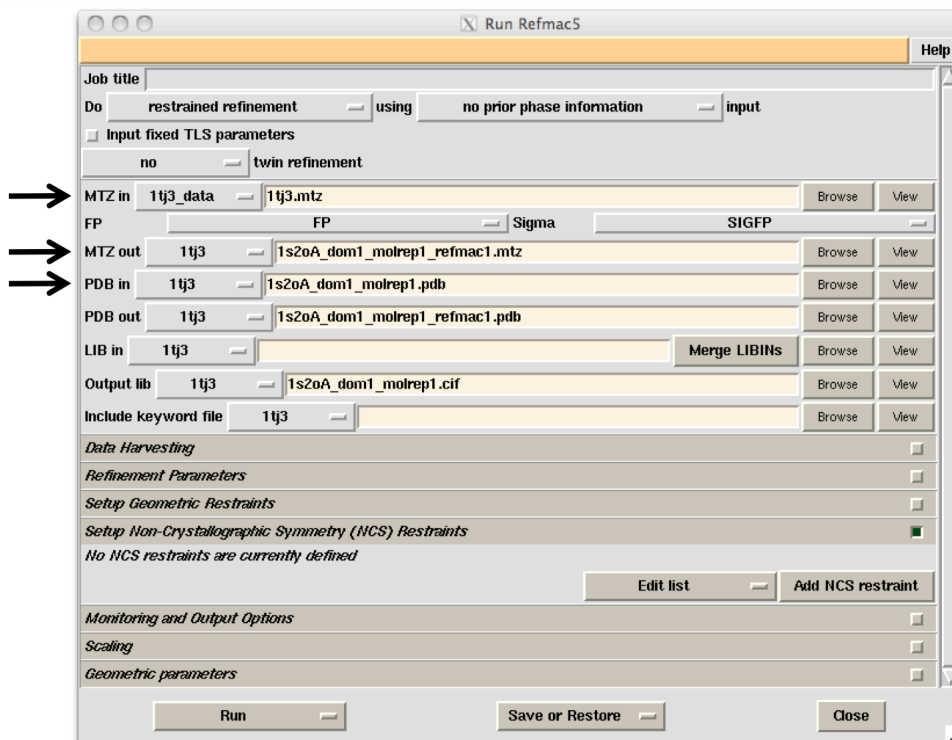
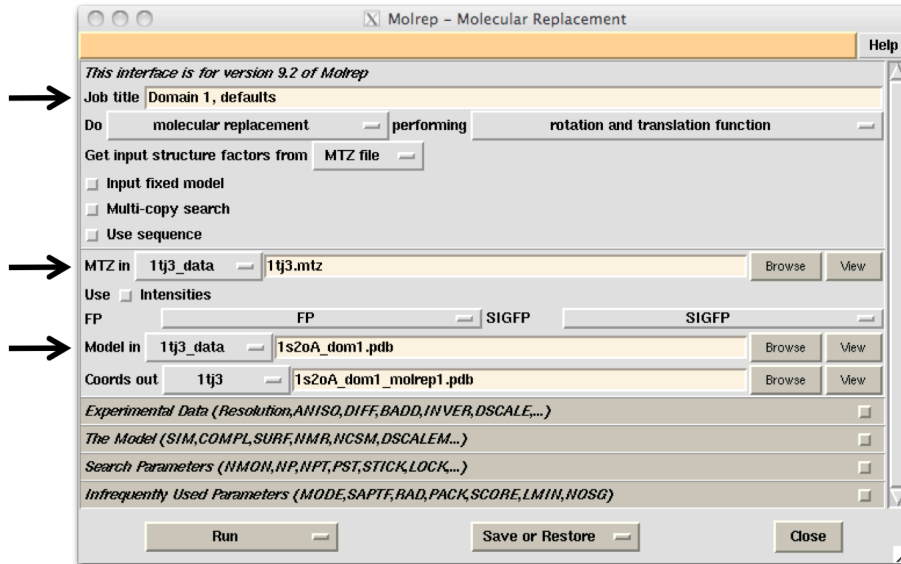


Use *Coot* to see that the large domain fits well the electron density. And this is not because of the "model bias", as suggested by a complete misfit of the smaller domain.

Therefore, there are no doubts at this point that the solution for the larger domain is correct. And there are numerous ways how to proceed. With this simple example, the auto-building or manual building should work, but we will examine Molrep's options that can be used for model completion in more complicated examples.

3. Solution for the large domain using default protocol

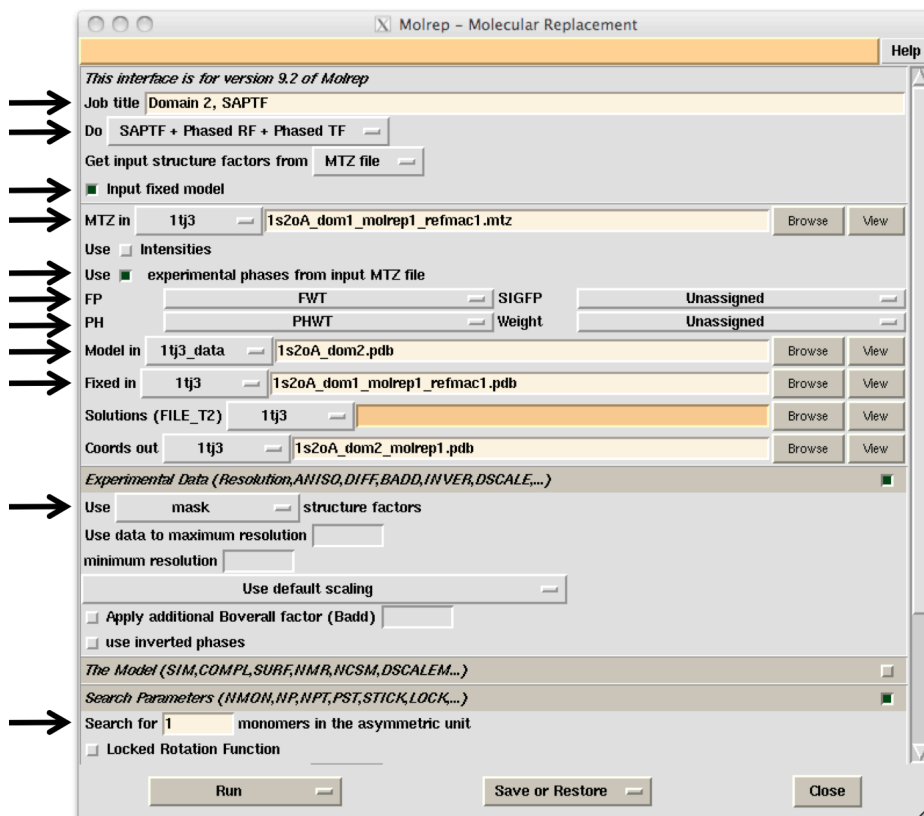
The separate pdb-files for two domains are already prepared. (Files **1s2oA_dom1.pdb** and **1s2oA_dom2.pdb** in *1tj3_data* directory contain large and small domains, respectively). Therefore simply repeat the procedure from the previous section but with the search model **1s2oA_dom1.pdb**, as shown in figures below. Also, change the default name for output mtz-file in refmac task window to **1s2oA_dom1_molrep1_refmac1.mtz**.



At this point you can examine the refined model and the electron density, files **1s2oA_dom1_molrep1_refmac1.pdb** and **1s2oA_dom1_molrep1_refmac1.mtz**.

4. Solution for the small domain using SAPTF and phased RF and TF

Open Molrep task window and select file names and options as shown in the next figure.



Run the job. The log file should be similar to one shown in the next figure. With this protocol, Molrep performs the following steps:

- calculates electron density map using FWT and PHIWT columns from **1s2oA_dom1_molrep1_refmac1.mtz**. This is 2Fo-Fc type map from the previous refmac run.
- calculates a mask from **1s2oA_dom1_molrep1_refmac1.pdb** and sets to zero the density within the mask

Note that masking the density makes sense even if Fo-Fc maps are used, as this removes noise in the map from an imperfect model

- calculates the Spherically Averaged Phased Translation Function (SAPTF), in which the target density is the masked 2Fo-Fc map and the template density is the spherically averaged density from the model (Table "peaksearch / SAPTF /" in the log file). The SAPTF is fast but noisy and it is necessary to examine many peaks from this function in subsequent steps of the protocol

The default number of SAPTF peaks is 100, and this may be too few in other examples. This parameter can be redefined in the text field "Nptd ...", located in the folder "Infrequently Used Parameters" of the Molrep task window.

- calculates the Phased Rotation Function (PRF) for each SAPTF peak using spherical harmonics up to $L=20$ only (Table "Quick search orientation by PRF"). This step is for fast initial filtering of the SAPTF peaks. By default, 30 best peaks are passed to the next step. This number can be redefined in the text field "Search for ... peaks in rotation map ...", located in the folder "Search Parameters" of the Molrep task window.
- For each SAPTF peak that passed the initial filter, Molrep calculates the PRF using more spherical harmonics, with the actual number of harmonics depending on the resolution limit (Table "Phased Rotation function").

At this point both positions and orientations of potential solutions are determined. However, neither SAPTF not PRF is a good score for selection of the best solution. Moreover, the position of molecule estimated with the SAPTF is less accurate than that obtained with the Phased Translation Function (PTF) for a known orientation.

- Therefore, there is one more step, in which the PTF is calculated and the preliminary estimates of positions are replaced by the closest peaks of the PTF. If the correction shift is larger than a threshold, which equals the upper resolution limit of the X-ray data, the potential solution is rejected. (See table "phased translation function", where the rejected candidate solutions are indicated with "No peaks, Dlim<Dist".) In a typical case such as this one, the most of the candidate solution are rejected following this criterion.
- The solution with the best score is saved in the output PDB-file, which also contains the fixed model that has been used for masking the density.

```

CCP4I fileviewer 5_molrep.log
-- Spherically Averaged Phased Translation Function (SAPTF) --
* Plotfile: /tmp/lebedev/1tj3_5_molrep_molrep_saptf.ps
INFO: you can find list of saptf_peaks in molrep_saptf.tab
use keyword FILE_T2

--- peaksearch / SAPTF / ---
(expected model positions)
+-----+
| Xfrac  Yfrac  Zfrac  Dens  Dens/sigma |
+-----+
| 1  0.297  0.354  0.056  0.2414  6.06 |
| 2  0.344  0.241  0.036  0.2387  5.99 |
| 3  0.334  0.194  0.018  0.2363  5.93 |
| 4  0.268  0.162  0.032  0.2151  5.40 |
| 5  0.295  0.120  0.021  0.2146  5.39 |
| 6  0.252  0.194  0.017  0.2003  5.03 |
| 7  0.278  0.256  0.028  0.1994  5.01 |
| 8  0.305  0.339  0.038  0.1950  4.89 |
+-----+

--- Quick search orientation by PRF ---
--- PRF (Lmin= 2 Lmax= 20 Nptd= 100)
+-----+
| Isaptf  theta  phi  chi  shx  shy  shz  PRf  SAPTF |
+-----+
| 1  7  102.48 -61.45 146.90 0.278 0.256 0.028 149.1 0.1994 |
| 2  6  111.35 -53.73 135.02 0.252 0.194 0.017 139.0 0.2003 |
| 3  8  93.74 -154.37 147.06 0.305 0.339 0.038 95.70 0.1950 |
| 4  3  106.38 46.24 164.27 0.334 0.194 0.018 90.47 0.2363 |
| 5  10 84.37 97.98 126.08 0.315 0.153 0.002 89.28 0.1862 |
+-----+

Radius of integration : 11.33
--- Phased Rotation function ---
+-----+
| theta  phi  chi  shx  shy  shz  Rf  SAPTF |
+-----+
| 1  102.58 -61.10 146.72 0.278 0.256 0.028 135.2 0.1994 |
| 2  111.47 -53.70 134.89 0.252 0.194 0.017 130.1 0.2003 |
| 3  115.98 -162.83 162.96 0.023 0.683 0.006 94.73 0.1493 |
| 4  56.36 -144.28 166.34 0.359 0.420 0.018 90.97 0.1622 |
| 5  93.78 -154.38 146.83 0.305 0.339 0.038 87.82 0.1950 |
+-----+

--- phased translation function ---
Resolution          : 268.75  2.90
NCS model           : 0
Radius of integration : 11.33
Corr_for_fixed_model: 0.033
Time_elapsed:      0h 5m 45s  Remained: 0h 2m 25s

RF TF  Tf/sig  TFentrst  PFind  PF  PFmin  wRfac  Score  Cntrst  (for)
1  1  56.50  0.000  1.00  1.00 -23.37 0.669 0.164 0.00 (0.164)
2  12 14.10  0.000  1.00  1.00 -24.07 0.682 0.039 0.00 (0.164)
3  0  --- No peaks Dlim<Dist ---
4  0  --- No peaks Dlim<Dist ---
30 0  --- No peaks Dlim<Dist ---

--- Summary ---
+-----+
| RF  TF  theta  phi  chi  tx  ty  tz  TFcnt  wRfac  Score |
+-----+
| 1  1  0  102.58 -61.10 146.72 0.291 0.256 0.022 0.00 0.669 0.164 |
| 2  2  0  111.47 -53.70 134.89 0.339 0.234 0.021 0.00 0.682 0.039 |
| 3  10 0  161.74 98.13 123.16 0.000 0.000 0.000 -2.00 0.000 -2.000 |
| 4  11 0  106.50 46.22 164.16 0.000 0.000 0.000 -2.00 0.000 -2.000 |
+-----+

Nmon RF  TF  theta  phi  chi  tx  ty  tz  TFcnt  wRfac  Score
1  1  1  102.58 -61.10 146.72 0.291 0.256 0.022 0.00 0.669 0.164

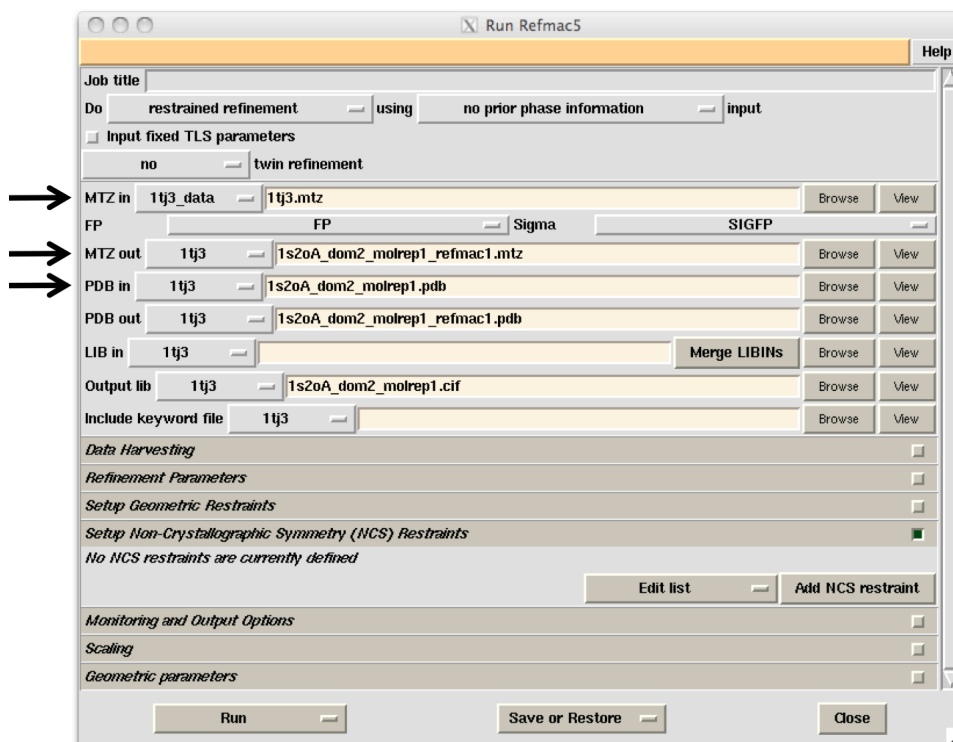
--- convert "molrep.crd" to "molrep.pdb" ---
Time: 2h 0m 32s Elapsed: 0h 7m 13s
MOLREP(ccp4): Normal termination
Times: User: 417.5s System: 8.5s Elapsed: 7:14

Find Show Log Graphs Show Summary Quit

```

In this figure and in similar figures below, arrows trace the "history" of the selected solution

- Refine the MR model as usual



As follows from low values of R-factors, the second domain is placed correctly.

Ncyc	Rfact	Rfree	FOM	-LL	-LLfree	rmsBOND	zBOND	rmsANGL	zANGL	rmsCHIRAL
0	0.3361	0.3534	0.708	64243.	4227.5	0.0101	0.443	1.378	0.620	0.101
1	0.2851	0.3214	0.754	62909.	4173.3	0.0153	0.681	1.470	0.685	0.101
2	0.2682	0.3121	0.771	62462.	4149.1	0.0150	0.659	1.551	0.723	0.103
3	0.2574	0.3047	0.782	62137.	4133.0	0.0144	0.629	1.549	0.722	0.104
4	0.2502	0.2992	0.791	61879.	4118.6	0.0146	0.638	1.564	0.731	0.106
5	0.2442	0.2945	0.799	61686.	4106.3	0.0147	0.643	1.576	0.739	0.107
6	0.2397	0.2887	0.806	61521.	4098.6	0.0148	0.649	1.596	0.750	0.108
7	0.2363	0.2845	0.811	61388.	4090.3	0.0151	0.661	1.616	0.761	0.109
8	0.2326	0.2807	0.816	61262.	4081.5	0.0152	0.668	1.639	0.773	0.110
9	0.2303	0.2773	0.819	61172.	4075.6	0.0155	0.677	1.655	0.781	0.111
10	0.2286	0.2738	0.821	61128.	4071.0	0.0155	0.679	1.665	0.786	0.111

Final results:

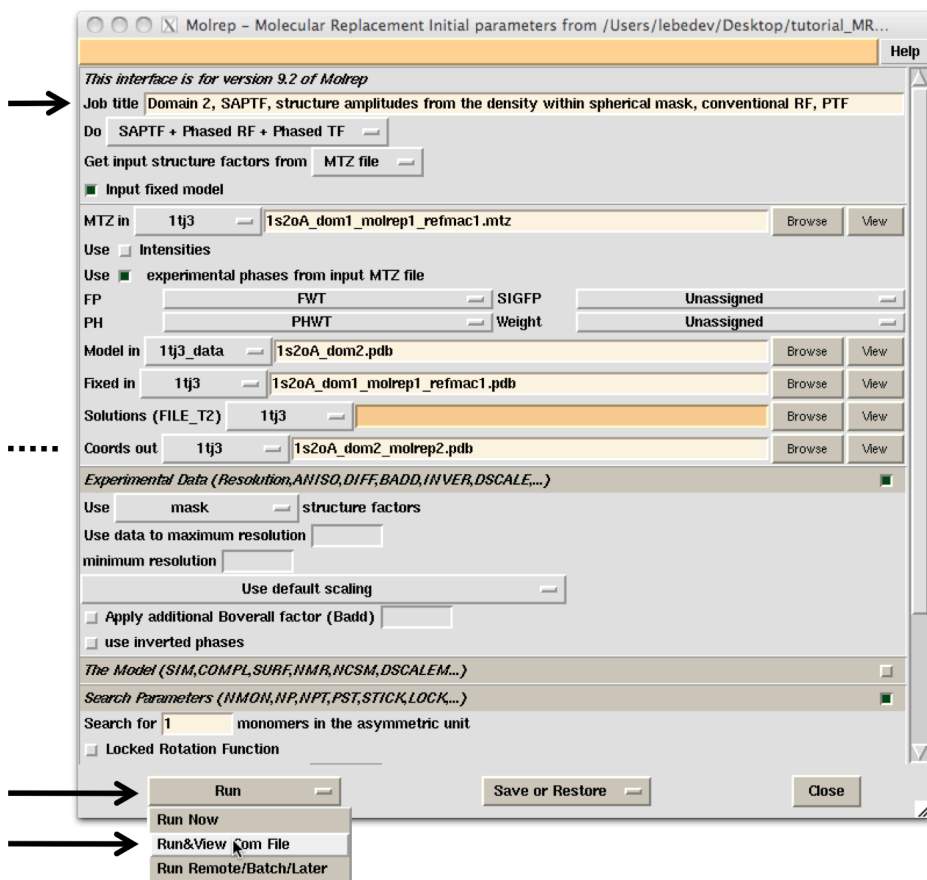
	Initial	Final
R factor	0.3361	0.2286
R free	0.3534	0.2738

5. Solution for the smaller domain using SAPTF, local RF and phased TF

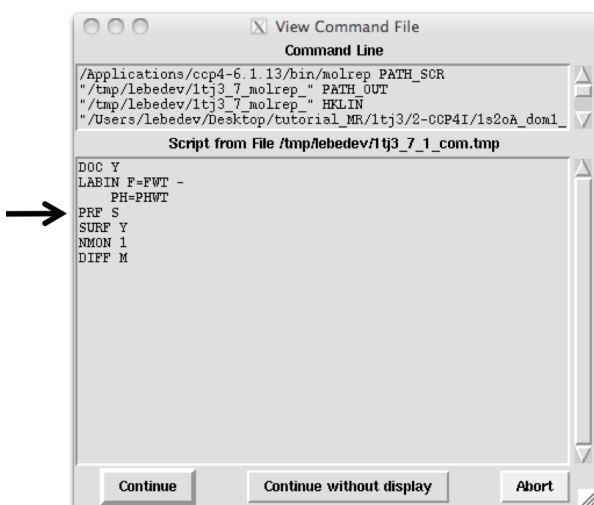
The difference with the previous approach is that the local RF substitutes for the phased RF. For each potential model position defined by a SAPTF peak, the electron density beyond a sphere centred at this position is set to zero, and the structure amplitudes from such a map are used as a target for the conventional RF. The motivation is that the phased RF is quite sensitive to any disagreement between target and template centres, but the SAPTF estimate of the template position is not very accurate. The price for using local RF instead of PRF is that the phase information is disregarded and contrast is lower.

This protocol is invoked similarly to the previous example but requires manual editing of the command script, as shown below.

- Reopen the previous Molrep task window. (Click on the Job No 5 "Domain 2, ..." in the job panel of the main CCP4I window and then press the "Rerun Job " button)
- Fill in the title
- Click within the "Model in" text box and press the tab button to set the output file name
- Select "Run > Run & View Com File"



- In the lower part of the "View Com File" window, substitute "PRF S" for "PRF Y"
- and press "Continue" button



With this approach, the SAPTF peaks located close to each other (e.g. the solution peak and its satellites) give close values of local RF and are all passed to the next, PTF step. This is to remember when analysing the molrep log-file shown below. Note that the table "peak similarity" may help

understand what is happening. In this particular case 26 top local-RF solutions are all from closely located SAPTF peaks and eventually result in very similar solutions of PTF at the final step of the procedure.

```

CCP4I fileviewer 7_molrep.log
Help

-- Spherically Averaged Phased Translation Function (SAPTF) --

--- peaksearch / SAPTF / ---
(expected model positions)
-----
| Xfrac  Yfrac  Zfrac  Dens  Dens/sigma |
|-----|
| 1  0.297  0.354  0.056  0.2414  6.06 |
| 2  0.344  0.241  0.036  0.2387  5.99 |
| 3  0.334  0.194  0.018  0.2363  5.93 |
|-----|

--- Quick search orientation by PRF ---
--- PRF (Lmin= 2 Lmax= 20 Nptd= 100)
-----
| Isaptf  theta  phi  chi  shx  shy  shz  PRf  SAPTF |
|-----|
| 1  7  102.48 -61.45 146.90 0.278 0.256 0.028 149.1 0.1994 |
| 2  6  111.35 -53.73 135.02 0.252 0.194 0.017 139.0 0.2003 |
| 3  8  93.74 -154.37 147.06 0.305 0.339 0.038 95.70 0.1950 |
| 4  3  106.38 46.24 164.27 0.334 0.194 0.018 90.47 0.2363 |
| 5 10  84.37 97.98 126.08 0.315 0.153 0.002 89.28 0.1862 |
| 6  2  106.91 72.04 116.33 0.344 0.241 0.036 88.61 0.2387 |
| 7  4  77.00 -122.50 124.05 0.268 0.162 0.032 82.86 0.2151 |
|-----|

--- Search orientation ---
--- Rotation function (for modified map) ---
Radius of integration : 22.66

| theta  phi  chi  alpha  beta  gamma  Rf  SAPTF |
|-----|
| 1  102.53 -59.59 147.92 173.37 139.50 112.56 0.2030E+05 0.1994 |
| 2  102.90 -60.19 147.86 172.03 139.00 112.41 0.1768E+05 0.2003 |
| 3  102.88 -60.04 147.82 172.27 139.00 112.36 0.2004E+05 0.1950 |
| 4  103.01 -60.07 149.35 170.54 140.00 110.68 0.1493E+05 0.2363 |
| 5  103.16 -60.31 148.28 170.98 139.00 111.60 7999. 0.1862 |
| 6  102.56 -59.76 147.97 173.10 139.50 112.62 0.2050E+05 0.2387 |
| 7  102.92 -60.15 147.90 171.99 139.00 112.28 0.1939E+05 0.2151 |
|-----|

--- after sort ---
-----
| theta  phi  chi  shx  shy  shz  Rf  SAPTF |
|-----|
| 1  102.88 -59.68 147.83 0.223 0.225 0.056 0.2196E+05 0.1354 |
| 2  102.71 -59.63 147.56 0.247 0.201 0.042 0.2093E+05 0.1563 |
| 3  102.52 -59.67 147.91 0.310 0.263 0.037 0.2067E+05 0.1616 |
| 4  102.56 -59.76 147.97 0.344 0.241 0.036 0.2050E+05 0.2387 |
| 5  102.53 -59.61 147.92 0.373 0.349 0.028 0.2050E+05 0.1599 |
|-----|

---RF peak similarity ---
-----
| 1 | 9 9 9 9 9 9 9 9 8 9 9 9 8 9 9 9 9 9 9 9 9 9 8 8 8 8 . . . . |
| 26 | 8 8 8 8 8 8 8 9 9 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 9 9 9 . . . . |
| 27 | . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9 . . . . |
| 28 | . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9 . . . . |
| 29 | . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9 . . . . |
| 30 | . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9 . . . . |
|-----|

Time: 10h 35m 49s Elapsed: 0h 5m 28s

--- phased translation function ---

RF TF Tf/sig TFentrst PFind PF PFmin wRfac Score Cntrst (for)
|-----|
| 1 1 59.68 0.000 1.00 1.00 -22.15 0.668 0.168 0.00 ( 0.168) |
| 2 1 59.64 0.000 1.00 1.00 -22.12 0.668 0.168 0.00 ( 0.168) |
| 3 1 59.37 0.000 1.00 1.00 -21.93 0.667 0.169 0.00 ( 0.169) |
| 4 1 59.35 0.000 1.00 1.00 -21.98 0.668 0.169 0.00 ( 0.169) |
| 5 1 59.40 0.000 1.00 1.00 -21.90 0.667 0.168 0.00 ( 0.169) |
|-----|

26 6 14.19 0.000 1.00 1.00 -22.72 0.672 0.041 0.00 ( 0.169)
27 0 --- No peaks Dlim<Dist ---
28 0 --- No peaks Dlim<Dist ---
29 0 --- No peaks Dlim<Dist ---
30 0 --- No peaks Dlim<Dist ---

--- Summary ---
-----
| RF TF theta phi chi tx ty tz TFent wRfac Score |
|-----|
| 1 4 0 102.56 -59.76 147.97 0.292 0.255 0.022 0.00 0.668 0.169 |
| 2 22 0 102.27 -59.85 147.52 0.292 0.255 0.022 0.00 0.667 0.169 |
| 3 9 0 102.26 -60.08 147.52 0.292 0.255 0.022 0.00 0.667 0.169 |
|-----|

Nmon RF TF theta phi chi tx ty tz TFent wRfac Score
|-----|
| 1 4 1 102.56 -59.76 147.97 0.292 0.255 0.022 0.00 0.668 0.169 |
|-----|

--- convert "molrep.crd" to "molrep.pdb" ---
Time: 10h 37m 19s Elapsed: 0h 6m 58s
MOLREP(ccp4): Normal termination

Find Show Log Graphs Show Summary Quit

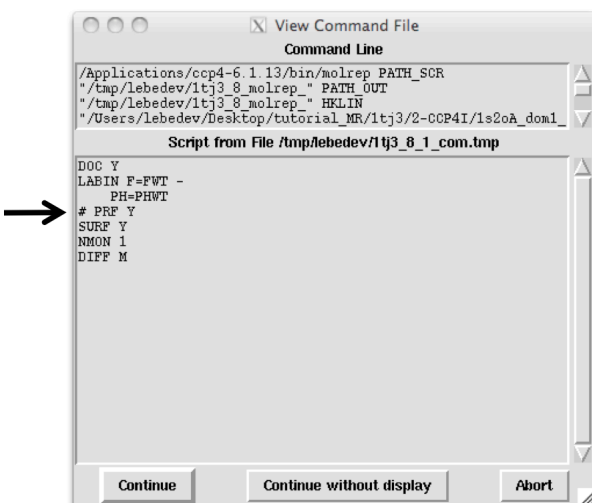
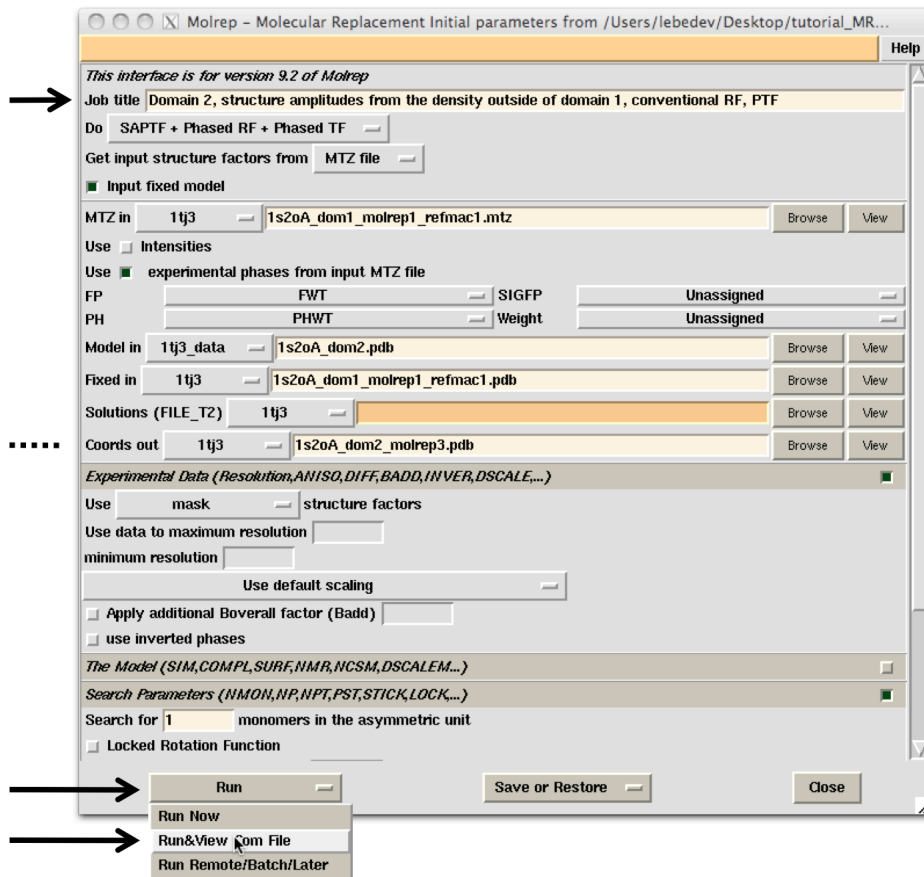
```

Compare this solution with the previous one in *Coot*.

6. Solution for the small domain using "masked" RF and phased TF

Here, similarly to the "standard" protocol, the orientation and position of the search model are found in two steps in this order. One difference is that RF is calculated using structure amplitudes from the map, in which the area occupied by the known partial model is masked out. Another difference is that position is found by phased TF.

To try this method, open the previous molrep job and proceed as shown below. The difference from the previous example is that PRF keyword in the "View Com File" window needs to be removed or commented out.



The log-file from this job will be similar to the one shown below.

```

CCP4I fileviewer 8_molrep.log
Help

Keyword DIFF : M
--- Remove fixed model from Fobs by mask ---
Space group : P 65 2 2
No: 179 Sett: 1
Cell: 68.750 68.750 268.750 90.00 90.00 120.00
Centre of model2 /fract/ : 0.110 0.345 -0.037

--- Rotation function ---
Radius of gyration : 11.33

--- Peaks of Rotation Function ---
+-----+-----+-----+-----+-----+
| theta  phi    chi  Rf/sigma |
+-----+-----+-----+-----+
| 1  140.69 161.96 58.84 3.84 |
| 2  106.02 136.54 92.28 3.81 |
| 3   71.96  93.25 70.40 3.57 |
| 4  100.43 164.51 172.97 3.57 |
| 27  93.99 166.16 165.54 2.85 |
| 28 144.11 124.00  72.86 2.85 |
| 29  94.21 -126.17 179.76 2.84 |
+-----+-----+-----+-----+

--- phased translation function ---
Resolution : 268.75 2.90
NCS model : 3
Corr_for_fixed_model: 0.034
Time_elapsed: 0h 0m 13s Remained: 0h 4m 0s

RF TF  Tf/sig  TFentrst  PFind  PF  PFmin  wRfac  Score  Cntrst (for)
1 1  16.61  2.845  1.00  1.00 -27.49 0.682 0.047 0.00 (0.047)
2 10 14.61  1.577  1.00  1.00 -32.36 0.682 0.044 0.00 (0.047)
3 3  15.15  3.739  1.00  0.96 -28.37 0.675 0.051 0.00 (0.051)
4 1  18.73  1.945  1.00  1.00 -52.16 0.675 0.049 0.00 (0.051)
27 15 12.42  2.315  1.00  1.00 -52.38 0.683 0.048 0.00 (0.054)
28 1  62.43 31.731  1.00  1.00 -32.67 0.668 0.174 0.00 (0.174)
29 11 13.08  2.502  1.00  1.00 -43.54 0.676 0.046 0.00 (0.174)

--- Summary ---
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| RF  TF  theta  phi    chi  tx  ty  tz  TFent  wRfac  Score |
+-----+-----+-----+-----+-----+-----+-----+-----+
| 1  28  1 144.11 124.00 72.86 0.957 0.254 0.477 31.73 0.668 0.174 |
| 2  19  9 175.88 -176.36 144.25 0.221 0.931 0.190 3.97 0.675 0.054 |
| 3  31  7 175.99 172.59 144.75 0.221 0.932 0.189 3.37 0.677 0.053 |
| 4  47  7  77.86  87.89 111.76 0.686 0.814 0.530 4.41 0.678 0.053 |
| 5  17  7  90.64 101.78 172.54 0.223 0.280 0.643 3.31 0.677 0.052 |
+-----+-----+-----+-----+-----+-----+-----+-----+
| Nmon RF  TF  theta  phi    chi  tx  ty  tz  TFent  wRfac  Score |
+-----+-----+-----+-----+-----+-----+-----+-----+
| 1  28  1 101.73 -59.12 146.45 0.298 0.254 0.023 31.73 0.668 0.174 |
+-----+-----+-----+-----+-----+-----+-----+-----+

--- convert "molrep.crd" to "molrep.pdb" ---
Time: 11h 1m 24s Elapsed: 0h 2m 38s
MOLREP(ccp4): Normal termination

Find Show Log Graphs Show Summary Quit

```

In a simple example as this one, all three approaches work. However, it is difficult to predict what will happen in a more difficult case. In general, several techniques are worthy trying in parallel, if the solution was not found in an automatic manner.