

# **Lafire manual** (version 2.0)

(For version 1.0, January 1, 2005)

Min Yao

# Index

1. Introduction .....	3
2. Limitation of the current version.....	4
3. Installation .....	4
3-1. Installation of Lafire .....	4
3-2. Installation of Qt .....	5
4. Using Lafire .....	6
4-1. Data for Lafire .....	6
4-2. Preparing for executing Lafire.....	6
4-3. Executing Lafire (executing, stopping, restarting) .....	9
4-4. Output files of Lafire .....	9
4-5. How to read results of Lafire .....	11
5. Lafire contact .....	11
6. Simplified manual of Lafire_molview .....	12
7. Error message and fixing .....	13

# 1. Introduction

Lafire (Local-correlation-coefficient-based automatic fitting for refinement) is an automatic refinement system for protein crystallography. The system is designed to start from an initial model (approximate or partial model) and finish with the final structure including water molecules (Fig.1). Lafire uses the refinement programs CNS or REFMAC5.

At first step, Lafire confirms amino acids of the initial model based on the sequence file that may be provided by the program user, and displaces the incorrect amino acids of the model (assignment).

In the refinement loop, the program first builds the parts that are missing in the model or seriously misfitted to the electron density map calculated from the observed phases or to the  $2F_o-F_c$ ,  $F_o-F_c$  maps. Then, correlation coefficient is calculated for each residue, and if the correlation is not good enough, the corresponding residue is subjected to the refitting routine.

Lafire also writes out the files of Rfree/Rwork factors and Ramachandran plot for the current model in order to monitor the progress of refinement. The Rfree/Rwork factor plot and Ramachandran plot can be viewed easily using graphics program Lafire\_moleview.

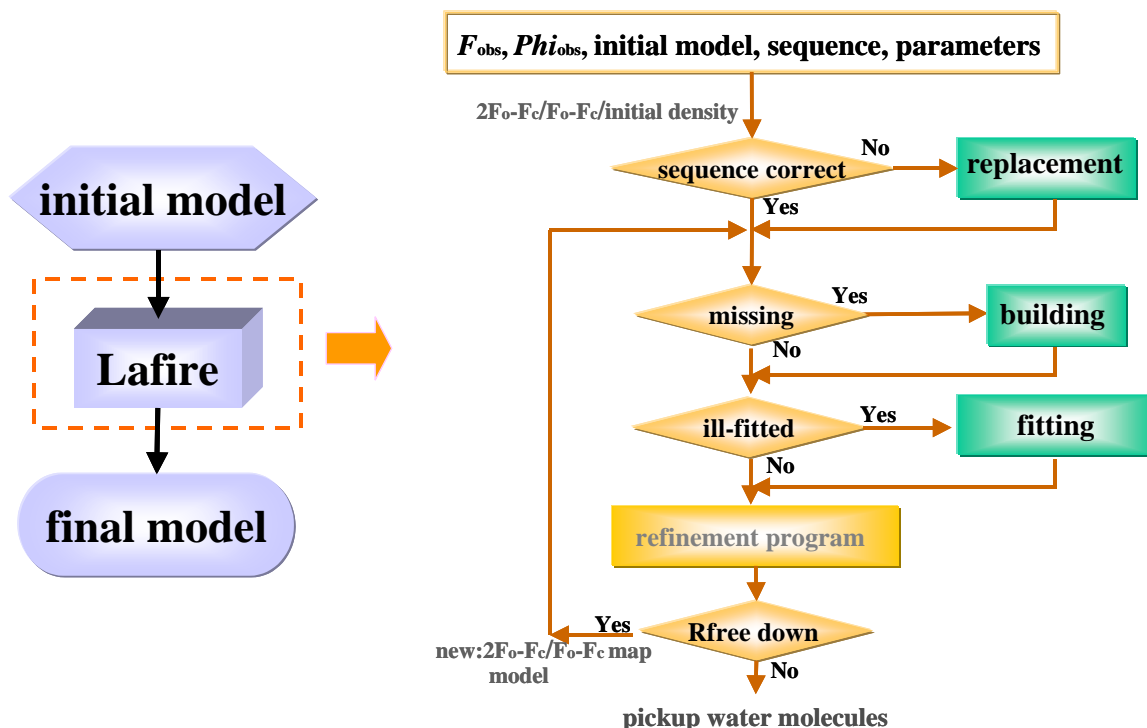


Figure 1. Automatic refinement using Lafire

## 2. Limitation of the current version

- 1). Since Lafire can not recognize amino acid residues at present, the residue number in the initial model must be consistent with that of the sequence file.
- 2). The maximum number of molecules in the asymmetric unit is 26. The PDB file of the initial model has to have the name of sigid or chain-name for CNS or REFMAC5 program, respectively. In the case of using CNS, the names of sigid must be the order of PROA, PROB, PROC, ..., PROZ, or the chain-names must be the order of A, B, C, ..., Z for using REFMAC5. In the case of that there are N complex molecules in the asymmetric unit, you have to add segid or chain-name by the order of inside of a complex at first, then copies. For example, there are two complexes in an asymmetric unit, and one complex molecule involves three protein molecules (hetero-trimer). For CNS, the sigid of one hetero-trimer have to be PROA, PROB PROC, and another one have to be PROD, PRO, PROF. PROA and PROD, PROB and PROE and PROC and PROF are copies of same molecule, respectively. Likewise, for REFMAC5, the chain-name of one hetero-trimer have to be A, B, C, and another one have to be D, E, F. A and D, B and E, and C and F are copies of same molecule, respectively.
- 3). Lafire does not support for fitting or building ligands or nucleic acids at present.
- 4). Lafire does not support for two-conformation of side-chain.
- 5). Lafire does not support for non-Pro cis-peptide.
- 6). Lafire can only use the refinement program CNS or REFMAC5.
- 7). Lafire only supports SGI (IRIS6.5 and after) and Linux (RedHat 9.0, Mandrake1.0).
- 8). The graphics program Lafire\_moleview is used only for displaying the results of Lafire, and it was made using OpenGL and Qt. Lafire\_moleview has been tested for SGI (IRIS6.5).

## 3. Installation

Lafire uses CCP4 program package and CNS program (If you want to use CNS program for refinement). Therefore, before installing Lafire, you should install CCP4 and CNS.

### 3-1. Installation of Lafire

#### 3-1-1. Uncompressing Lafire

For example, if you want to install the Lafire at /home/xtal, please put Lafire.tar.gz into /home/xtal directory, then use “gunzip” and “tar” commands as below to uncompress the Lafire:

```
gunzip Lafire.tar.gz  
tar xf Lafire.tar
```

(or **gunzip Lafire.tar.gz | tar xf -** )

After uncompressing Lafire.tar.gz, the directory Lafire is made and four directories: **bin**, **client**, **lib**, **example** and one file of **configure** will be made at /home/xtal/Lafire directory.

### 3-1-2. *Configuring the system for using Lafire*

For checking environment required by Lafire, execute **configure** at first. The CCP4, CNS, Qt are required by Lafire (Qt is only required for using graphic program Lafire\_moleview).

```
./configure
```

### 3-1-3. *Setting environment parameters for executing Lafire*

Before executing Lafire, you have to set path and environment parameters and put them into .cshrc file. For example, if Lafire is installed under /home/xtal directory, please do as below:

```
setenv LAFIRE_DIR /home/xtal/Lafire (for Linux: export LAFIRE_DIR=/home/xtal/Lafire)  
set path=($path /home/xtal/Lafire/bin)  
alias cns_solve cnx_solve ( only for using CNX )
```

## 3-2. *Installation of Qt library (only for using graphics program Lafire\_moleview )*

Qt is a graphical user interface library for C++. It is free software developed by Troll Tech of Norway. Normally, Qt is installed into /usr/local on the Linux RedHat9.0 (or newer than 9.0), while SGI machines don't have it in default situation. If you want to use graphic program Lafire\_moleview, and there is not Qt library, please download and install Qt as following steps.

### 3-2-1. *Download Qt*

Download Qt from Web site <http://www.trolltech.com/download/qt/x11.html>, and uncompress it under /usr/local directory.

### 3-2-2. *Setting environment parameters for compiling Qt*

```
setenv QTDIR /usr/local/qt  
setenv PATH ${PATH}:${QTDIR/bin}  
setenv MANPATH ${MANPATH}:${QTDIR/doc/man}  
setenv LD_LIBRARY_PATH ${LD_LIBRARY_PATH}:${QTDIR/lib}  
setenv QMAKESPEC irix-g++  
(in the case of Linux, use "export" as "setenv" )
```

### 3-2-3. *Compiling Qt*

Configure system and compile Qt.

```
./configure -qt-gif -system-libpng -system-zlib -system-libjpeg  
./make
```

You may also see Web site for installing Qt  
<http://paison.hp.infoseek.co.jp/paison/qt/qtinstall.html>

## 4. Using Lafire

Lafire requires a big size of memory. Normally 0.5-1MB is used dependent on the size of the sample. In the case of using NCS restraint for refinement, you have to modify refinement program command files that are in Task\_files directory to give NCS restraint:

For using CNS: rigid.inp, ncs.inp have to be modified to give NCS restraint

For using REFMAC5: tls.inp, .cif, and define the NCS in all files of refmac\_\*.inp.

If your model involves ligands or nucleic acids, and you want to refine them, you have to define and make the topology and energy parameter files, and put them into directory of Task\_files (see 4-2). You also need to define the name of topology and parameter files in the file of generate.inp (in directory of Task\_files) and cns\_configure file for using CNS, and in refmac\_\*.inp (in directory of Task\_files) for using REFMAC5.

### 4-1. Data for Lafire

#### 1). Initial model (PDB file)

Residue number in the initial model file must be consistent with that of sequence file. The initial model (PDB file) has to have chain-name (A, B, C...). In the case of using refinement program CNS, the initial model (PDB) file should have segid name (PROA, PROB, PROC...). Numbering system should be same for all independent chains.

#### 2). mtz file with Fobs, Rfree-flag, and phases used for building initial model. (The phases are only for using initial map, except for MR method)

#### 3). Sequence file (one character)

#### 4). Second structure information (only for the case that you want to use the information of second structure)

**/users/phoIDH/Lafire**

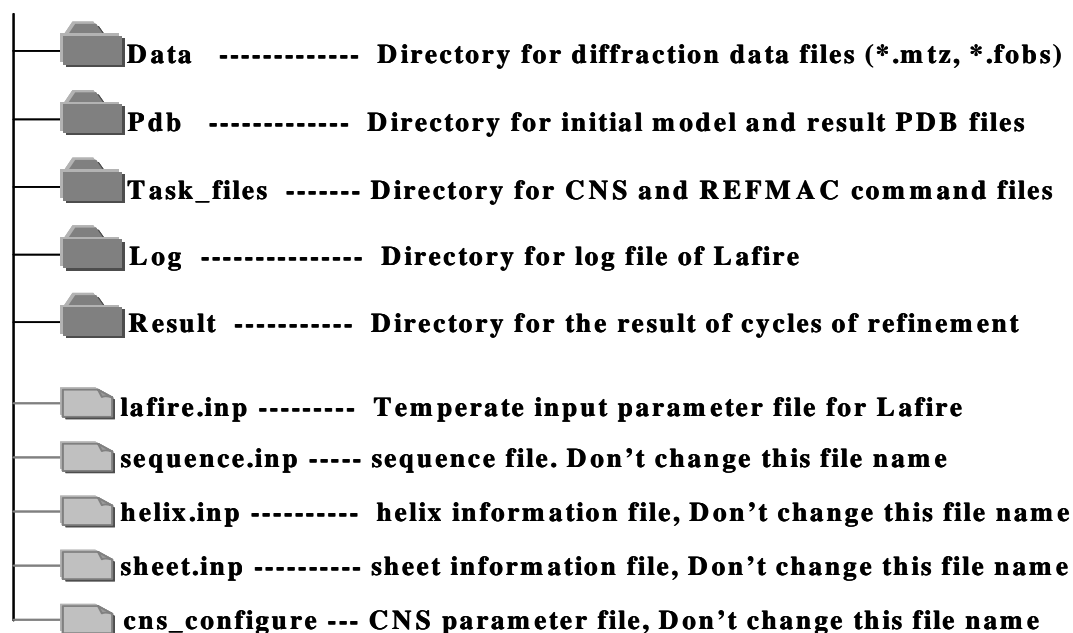


Figure 2. Lafire working directory

## 4-2. Preparing for executing Lafire

### 4-2-1. Setting your sample directory for Lafire

First, use “Lafire\_setup” to setup your working directory. For example, when /user/phoIDH/Lafire is the working directory (\$working-directory), then use command as

```
mkdir /users/phoIDH/Lafire  
Lafire_setup /users/phoIDH/Lafire
```

After running Lafire\_setup, the 5 directories (Data, Pdb, Task\_files, Log, and Result) and 5 files (lafire.inp, sequence.inp, helix.inp, sheet.inp, and cns\_configure) will be made (Figure2). If you want to run example that is provided in Lafire program package, use command as

### **Lafire\_setup /users/phoIDH/Lafire example**

Then all data files of the example for running Lafire are setup under /user/phoIDH/Lafire directory. If you want to run Lafire using example data, you should modify directory name in phoIDH.inp by following (4-2-8), then run Lafire by following (4-3-1).

### 4-2-2. Preparing Diffraction data files

You should put diffraction data file (\*.mtz ) into directory, \$working-directory /Data.

(For using CNS, the diffraction data file will be created automatically under the directory of \$working-directory /Data.)

### 4-2-3. Preparing PDB file for initial model

You should make a PDB file for initial under the directory that you define in Lafie parameter file (see 4-2-8). Residue number of the initial model in PDB file must be same with that you give in sequence file (sequence.inp) (see 4-2-4). You also have to give the segid for using CNS program or chain-name for using REFMAC5. The roles of segid and chain-name are described in 2(2) section.

### 4-2-4. Preparing Sequence file

The following is the format of sequence file (sequence.inp). The maximum number of characters per 1 line is 120.

```
-----  
2                               (number of different chains in the asymmetric unit)*  
>chain1                          (the name of the chain)  
MSRKLGVVPYESVRDSLVRDQDYIGSKEAKIKFVERE      (sequence)  
KHEVKVKGIGVFPNPNYIRVIMRAFIADVNESVRDSL      (sequence)  
LARLEAPFRARLRGTGYFPNEGA                    (sequence)  
  
>chain2                          (second chain name, if it exists)  
RDELAPYGVMRKSIPGWDPFKPHITLARRKAPAPRVPPVL    (sequence)  
FGLEWPVEGFALVRSELKPKGA                      (sequence)  
-----
```

\* If there are N molecules in the asymmetric unit, and they have same sequence, this parameter is 1.

#### 4-2-5. Preparing CNS parameters (only for using CNS program)

You should modify cns\_configure file for using CNS.

```
-----  
(===>) high_res=2.05; (high resolution limit used for refinement)  
(===>) atom_select=(known and not (hydrogen)); (define selected atoms for refinement)  
(===>) ncs_infile="PARAM:ncs_main_chain.inp"; (NCS-restraint file name)*  
(===>) parameter_infile_3="PARAM:atp.param"; (user-defined ligands parameter file name)**  
-----
```

\* ncs-restraint file should be made under Task\_files directory. "PARAM:" must be written.

\*\* user-defined parameter and topology files should be made in Task\_files directory. "PARAM:" must be written. You should modify generate.inp for topology files.

#### 4-2-6 Modifying generate file (only for using CNS, and there are more than 12 molecules in asym. unit)

If program CNS is used, and number of molecules in the asymmetric unit is more than 12, you need to modify the generate.inp file under \$working-directory/Task\_files directory, to give the input PDB file names. The input PDB file names must be "WORKD: 'protein\_prefix'\_PROA.pdb", "WORKD: 'protein\_prefix'\_PROB.pdb", "WORKD: 'protein\_prefix'\_PROC.pdb"... "WORKD: 'protein\_prefix'\_PROZ.pdb". The 'protein\_prefix' is the name of sample, and it is same with that given in Lafire parameter file (see 4-2-8).

#### 4-2-7. Preparing second structure information files (only for you want to use it)

If the resolution of data is lower than 3.0Å, use of second structure information is suggested. Presently, Lafire does not make second structure information automatically. Therefore you have to modify the file of helix.inp and sheet.inp by yourself. The helix.inp and sheet.inp are in working directory, and the formats of them are:

[the first residue number in helix1]:[ the last residue number in helix1] and [the first residue number in helix2]: [the last residue number in helix2] and [... ]:[... ] and [... ]:[... ] and .....

[the first residue number in sheet1]:[the last residue number in sheet1] and [the first residue number in sheet2]:[the last residue number in sheet2] and [... ]:[... ] and [... ]:[... ] and .....

#### 4-2-8. Preparing parameter file of Lafire

Please modify the italic part for your sample in Lafire.inp.

```
-----  
work_dir = /users/proIDH/Lafire (the name of working directory)  
dir_prefix = cycle_ (prefix part of directory name for refinement cycle)  
init_pdb = /users/ proIDH/Lafire /Pdb/phoIDH_init.pdb (the name of initial model file)  
protein_prefix = phoIDH_ (the name of sample. It is used as a prefix of file name)  
resolution = 2.05 (high resolution limit)  
MR = 0 (the method used for solving structure. 0:not MR; 1:MR)  
complex = 0 (the flag of complex.0:monomer/homo-oligomer;1:hetero-oligomer)  
nmole = 4 (number of molecules/complex in asymmetry unit)**  
jump = 0 (job control flag for Lafire. 0: default  
1:start from refinement step (1)  
2:start from refinement step (2)  
3:start from pickup water step)**  
MSPP = 0 (the flag for building strategy. 0: use multi-level strategy, 1: don't use (fast processing))  
CNS = 1 (the flag for using refinement program. 0:REFMAC5; 1: CNS)  
annealing = 1 (refinement control flag, only for CNS. 1:use SA, 0:don't use SA)  
mtz_in = resolve_remote (the name of mtz file without extension .mtz)  
labo= remote (label of Fobs, sigma(Fobs) in mtz file for using refinement)***
```

labi = *P* (label of Fobs, sigma(Fobs) in mtz file. This Fobs is used for building initial model)\*\*\*  
phi = *M* (label of phases, only for using phases information)\*\*\*\*

- 
- \* In the case of complex, this parameter is number of complexes in the asymmetric unit
  - \*\* There are two cases for starting from refinement step. The refinement step (1) (jump=1) is that the program starts from refinement based on finished fitting and building by Lafire. In this case the cycle\_0 ( $\{\text{dir\_prefix}\}_0$ ) directory is necessary. The refinement step (2) (jump=2) is to refine directly without fitting and building by Lafire.
  - \*\*\* The name of label does not include F and SIGF. For example, if the label of F, and sigma(F) are “FP”, and “SIGFP” in mtz file, the name of label is P; the label of F, and sigma(F) are FPromote, and SIGFPromote, the name of label is Promote. Please don’t use F and SIGF as the label name of F, and sigma(F) in mtz file.
  - \*\*\*\* The name of label does not include PHI. For example, if the label of phases is PHIM in mtz file, the name of label is M. If the label of phases is PHIdm in mtz file, the name of label is dm. Please don’t use PHI as the label name of phases in mtz file. The labels of HL(Hendrickson and Lattman )coefficient, should be same with PHI. In the above example, HL labels are HLAM, HLBM, HLCM, and HLDM.

### 4-3. Executing Lafire (*Executing, Stopping, Restarting*)

#### 4-3-1. Executing Lafire

**Lafire** [Parameter file name] &

For example, if the parameter file (see 4-2-8) name is phoIDH.inp, execute command as below:

**Lafire phoIDH.inp &**

#### 4-3-2. Stopping Lafire

If you want to stop the job of Lafire, execute kill\_Lafire as below

**kill\_Lafire** [Parameter file name]

For example, if the parameter file name (see 4-2-8) is phoIDH.inp, execute command as below:

**kill\_Lafire phoIDH.inp**

#### 4-3-3. Restarting Lafire

If Lafire stops with error, or you force to stop Lafire, and restart it, execute **clear\_Lafire** command first. Then execute Lafire again.

**clear\_Lafire** [Parameter file name]

For example, if the parameter file (see 4-2-8) name is phoIDH.inp, execute command as below:

**clear\_Lafire phoIDH.inp**

### 4-4. Output file of Lafire

There are two kinds of output files in Lafire, job monitor files (\*.log) and result files.

Job monitor files

auto.log and job.log are job monitor files. They are used only for debugging and checking Lafire

program. Normally, user does not need to read them. If problem occurred, and you could not resolve it, please send them to author.

Result files

**‘protein\_prefix’\_seq.seq**

This file is made under working directory, and it is only for Lafire program.

**‘protein\_prefix’\_mut.out**

The list of amino acids replaced (based on sequence file). It is saved under \$working-directory/Log directory.

**‘protein\_prefix’\_built.out**

The building list, and it is saved under \$working-directory/Log directory.

**‘protein\_prefix’\_rfree.out**

This is a list of Rfree factors during refinement, and it is saved under \$working-directory/Log directory. There are 4 columns in the file:

First column : Rfree factor

Second column: R factor

Third column: flag of refinement steps

0: auto-building and fitting step.

1: conjugate-gradient minimization refinement step.

2: B-factor refinement step (ML refinement step in REFMAC5)

3: simulated-annealing step

4: final step of small cycle before pickup water molecules

5: final step of pickup water molecules cycle

Fourth column: the file name in the refinement steps

**‘protein\_prefix’\_Lafire.pdb**

The best model before pickup water molecules, and it is saved under \$working-directory/Pdb directory.

**‘protein\_prefix’\_Lafire\_water.pdb**

The final model including water molecules, and it is saved under \$working-directory/Pdb directory.

**‘protein\_prefix’\_Lafire\_1.ps**

The ps file of Ramachandran plot of the best model before pickup water molecules, and it is saved under \$working-directory/Pdb directory

**‘protein\_prefix’\_Lafire\_water\_1.ps**

The ps file of Ramachandran plot of the final model (after pickup water molecules), and it is saved under \$working-directory/Pdb directory.

**‘protein\_prefix’\_Lafire\_2fo-fc.map** (CCP4-format)

The 2fo-fc map of the best model before pickup water molecules, and it is saved under

\$working-directory/Pdb directory.

**‘protein\_prefix’\_Lafire\_fo-fc.map** (CCP4-format)

The fo-fc map of the best model before pickup water molecules, and it is saved under \$working-directory/Pdb directory.

**‘protein\_prefix’\_Lafire\_water\_2fo-fc.map** (CCP4-format)

The 2fo-fc map of the final model including water molecules, and it is saved under \$working-directory/Pdb directory.

**‘protein\_prefix’\_Lafire\_water\_fo-fc.map** (CCP4-format)

The fo-fc map of the final model including water molecules, and it is saved under \$working-directory/Pdb directory.

#### **4-5. How to read the result of Lafire**

Since program Lafire consists of multi programs, the progress of refinement may not be affected by errors caused by fitting or building process (for example, Segmentation fault (core dumped) error). Therefore, you should check the output file of **‘protein\_prefix’\_rfree.out** (see 4-6-1) to make sure that refinement is done well.

##### **4-6-1. Checking Rfree/R factors**

The changes of Rfree/R factor during refinement using Lafire, are listed in output file of **‘protein\_prefix’\_rfree.out** under \$working-directory/Log directory (see 4-4). If Rfree/R factors are set to 0 in the file, there is something wrong, and Lafire does not work well. If graphic program Lafire\_moleview is installed on your system, you can show Rfree/R plot easily (see 6), otherwise Microsoft Excel can be used to plot the Rfree/R factors.

##### **4-6-2. Checking situation of building**

The file of **‘protein\_prefix’\_built.out** under \$working-directory/Log directory (see 4-4) shows the situation of building including replaced residues (based on sequence file), the number of residues built in every cycle, missing residues in final model.

##### **4-6-3. Checking final models**

The final models, the models before (**‘protein\_prefix’\_Lafire.pdb**) and after (**‘protein\_prefix’\_Lafire\_water.pdb**) pickup water molecules, are given under \$working-directory/Pdb directory (see 4-4). The 2Fo-Fc and Fo-Fc maps for both final models are also given in the same directory. You should check them using graphic programs such as O, Turbo-Frodo.

##### **4-6-4. Checking Ramachandran plots of final models**

If graphic program Lafire\_moleview can be installed on your system, it is easy to show Ramachandran plot of final models (see 6). The PS files of Ramachandran plot for both final models (see 4-6-3) are also calculated using PROCHECK.

## 5. Lafire contact

If you have any problem, question, or suggestion, please contact to [lafire@castor.sci.hokudai.ac.jp](mailto:lafire@castor.sci.hokudai.ac.jp)

## 6. Simplified manual of Lafire\_molview

Lafire\_molview is a graphic program for monitoring refinement. Unlike programs of O, Turbo-Frodo, or Quanta, this program can not be used for modifying the model, and it only displays 3/5/7 residues in main-windows with keeping a good view. The density maps are also be displayed by sphere of an R radius around an atom, not box.

Lafire\_molview supports only SGI (IRIS6.5).

### 6-1. Executing Lafire\_molview

**Lafire\_molview &**

### 6-2. Reading PDB file

There is a menu of [File] at the upper left of Lafire\_molview's window. You just choose [Open PDB file] in menu of [File] for reading PDB file. Lafire\_molview can read maximum 5 of PDB files.

### 6-3. Rfree/R plot

There is a menu of [Window] at the upper right of Lafire\_molview's window. You just choose [Rfactor] to open a sub-window of "Rfactor", and read '**protein\_prefix'\_rfree.out**'. Then Rfree/R factor plots will be displayed on sub-window.

### 6-4. Ramachandran plot

First, you need to read target PDB file by choosing [Open PDB file] from the menu of [File] (see 6-2). Then choose [Ramachandran Plot] from the menu of [Window] (see 6-3) to open a sub-window of "Ramachandran". After using button of [molecule] at the lower left in sub-window to choose target that you have read in, you just push button of [Draw Now] at the lower right in sub-window to display the Ramachandran plot. The blue marks are residues that are shown on main window.

## 7. Error message and fixing

The Lafire checks the data prepared by user for executing program. If the data do not satisfy conditions required by Lafire, error message and error cord will be given.

### 7-1. Error message and fixing data

#### 7-1-1. Error cord 1

The PDB file of initial model is not present under ~/Pdb directory, or you have given a wrong file name in parameter file (Lafire.inp) of Lafire (see 4-2-8).

#### 7-1-2. Error cord 2

The diffraction data (mtz file) is not present under ~/Data directory, or you have given a wrong file name in parameter file (Lafire.inp) of Lafire (see 4-2-8).

#### 7-1-3. Error cord 3

The labels (Fobs or Phase) defined in \*.inp file (Lafire.inp) are not present in mtz file.

#### 7-1-4. Error cord 4

The segid (for using CNS) or chain name (for REFMAC5) in PDB file of initial model is disaccord with requirement of Lafire (see 2-(2)). (The segid or chain name may not be given; segid was not 4 characters; segid or chain name is given in different way).

#### 7-1-5 Error cord 5 (only for nmol >1)

Numbering system is not same for molecules in the asymmetric unit (see 4-1-(1)).

#### 7-1-6. Error cord 6 (only for jump=1)

The parameter of "jump" in parameter file of Lafire was set to 1, Lafire will run from refinement step based on the model coordinates that have been built and fitted by Lafire. However, the first cycle directory (cycle\_0) could not be found (see 4-2-8).