

PRIOR - A computer program for calculation electron  
densities from analytical atomic form factors

User Manual

Lukáš Palatinus  
Sander van Smaalen

version: 05/01/2005

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Theory</b>	<b>3</b>
<b>3</b>	<b>Technical details</b>	<b>5</b>
3.1	Programming language and system requirements . . . . .	5
3.2	Execution . . . . .	5
<b>4</b>	<b>Specification of the input</b>	<b>6</b>
4.1	Types of input . . . . .	6
4.2	Definition of the atomic form factors . . . . .	6
4.3	Specification of keywords . . . . .	7
4.3.1	name: accuracy . . . . .	7
4.3.2	name: prioratoms - endprioratoms . . . . .	7
4.3.3	name: cell . . . . .	8
4.3.4	name: centers - endcenters . . . . .	8
4.3.5	name: normalize . . . . .	9
4.3.6	name: outputprior . . . . .	9
4.3.7	name: electrons . . . . .	9
4.3.8	name: progress . . . . .	9
4.3.9	name: symmetry - endsymmetry . . . . .	9
<b>5</b>	<b>Example of the input file</b>	<b>11</b>
<b>6</b>	<b>Description of the output</b>	<b>12</b>
<b>7</b>	<b>Practical issues</b>	<b>14</b>

# Chapter 1

## Introduction

The prior electron density (prior) is an essential part of the MaxEnt formalism. Only the MEM with a sufficiently informative prior can be used for studies of fine features of the electron density, like the atomic charges or the distribution of bonding electrons (De Vries et al., 1996; Papoular et al., 2002; Palatinus & van Smaalen, 2002). The best prior that can be obtained from standard structure refinements is the Independent Atom Model (IAM). In this model, the electron density is modelled as a superposition of electron densities of free, non-interacting atoms placed at their refined positions and convoluted with the refined thermal motion. An electron density constructed in this way is called procrystal electron density ( $\rho_{pro}$ ). Such a model describes the major part of the electron density very well, but it does not account fully for the effects of bonding on the electron density. Therefore, the difference density  $\rho_{MEM} - \rho_{pro}$  can be successfully used to determine the deformations of the electron density due to the bonding.

The easiest way to construct the procrystal prior is to calculate the structure factors corresponding to the model, using the well known formulas involving the atomic form factors and tensors of temperature parameters (Shmueli, 1996). The electron density can be then simply calculated as an inverse Fourier transform of the structure factors. Unfortunately, this method cannot be used but for the lightest elements. By heavier elements, the series termination errors become too pronounced and even calculation of all structure factors up to  $\sin(\theta)/\lambda \leq 6.0$  does not avoid termination ripples of such an amplitude that makes the electron density unusable as a prior in the MEM calculations. Therefore, an alternative method must be used. The method is based on evaluation of the electron density by means of an analytical Fourier transform, using the analytical approximation to the atomic form factors<sup>1</sup>.

---

<sup>1</sup>The method described here was inspired by personal communication with Robert Papoular and later published in Papoular et al. (2002).

## Chapter 2

# Theory

The atomic form factors are usually analytically approximated by a sum of exponentials:

$$f_a(\sin(\theta)/\lambda) = \sum_{i=1}^n a_i \exp\left(-b_i \frac{\sin^2(\theta)}{\lambda^2}\right) \quad (2.1)$$

The International Tables for Crystallography (Wilson & Prince, 1999) use  $n = 5$  and  $b_5 = 0$ , the atomic form factors published by Su & Coppens (1997) use  $n = 6$ .

It is more convenient to express the atomic form factor as a function of the diffraction vector  $\vec{H}$  in reciprocal crystal coordinates:

$$f_a(\vec{H}) = \sum_{i=1}^n a_i \exp\left(\frac{b_i}{4} |\vec{H}|^2\right) = \sum_{i=1}^n a_i \exp\left(\frac{b_i}{4} \vec{H}^T \mathbf{G}^* \vec{H}\right) \quad (2.2)$$

Here  $\mathbf{G}^*$  denotes the reciprocal metric tensor  $G_{ij}^* = \vec{a}_i^* \cdot \vec{a}_j^*$

Convolution of the static atomic electron density with the anisotropic harmonic thermal motion of the atom is expressed in reciprocal space by multiplying the atomic form factor by corresponding displacement term that involves the displacement tensor  $\mathbf{U}$  (Shmueli, 1996). This yields the anisotropic dynamic atomic form factor  $d_a$ :

$$d_a(\vec{H}) = \sum_{i=1}^n a_i \exp\left(\frac{b_i}{4} \vec{H}^T \mathbf{G}^* \vec{H}\right) \exp(2\pi^2 \vec{H}^T \mathbf{A}^* \mathbf{U} \mathbf{A}^* \vec{H}) \quad (2.3)$$

The matrix  $\mathbf{A}^*$  is a diagonal matrix of the lengths of the reciprocal-space basic vectors:  $A_{ii}^* = |\vec{a}_i^*|^2$ ;  $A_{ij}^* = 0$ ,  $i \neq j$ .

After some reordering, Eq. 2.3 becomes:

$$d_a(\vec{H}) = \sum_{i=1}^n a_i \exp\left(-\vec{H}^T \underbrace{\left[\frac{b_i}{4} \mathbf{G}^* + 2\pi^2 \mathbf{A}^* \mathbf{U} \mathbf{A}^*\right]}_{\mathbf{M}_i} \vec{H}\right) = \sum_{i=1}^n a_i \exp\left(\vec{H}^T \mathbf{M}_i \vec{H}\right) \quad (2.4)$$

The dynamic electron density of an atom  $a$  is given by a Fourier transform of  $d_a$ . Since  $d_a$  is expressed as a three-dimensional Gaussian, its Fourier Transform can be calculated analytically. The result is:

$$\rho_a(\vec{r}) = |\mathbf{G}^*|^{\frac{1}{2}} \sum_{i=1}^n a_i \pi^{\frac{3}{2}} |\mathbf{M}_i^{-1}| \exp(-\pi^2 \vec{r}^T \mathbf{M}_i^{-1} \vec{r}) \quad (2.5)$$

The prefactor  $|\mathbf{G}^*|^{\frac{1}{2}}$  transforms  $\rho_a$  from units of the crystal coordinate system to  $e/\text{\AA}^3$ .

Eq. 2.5 provides a convenient way of calculating the electron density of an atom  $a$ . The atom is centered in the origin of the coordinate system and the density is a function of coefficients of the analytical approximation of the atomic form factors, and of the coefficients of the atomic tensor of displacement parameters. The formula does not involve any numerical Fourier Transform and therefore does not lead to series termination error. The analytical approximation becomes relatively inaccurate at high values of  $\sin(\theta)/\lambda$  (Wilson & Prince, 1999; Su & Coppens, 1997). However, these values influence mostly the density close to atomic nuclei. The experimental data provide almost no information about this electron density, and the exact form of the electron density in the vicinity of atomic nuclei is therefore not crucial for the fit to the experimental data.

## Chapter 3

# Technical details

The input to PRIOR is the standard crystallographic information in a format common with the program BAYMEM and the list of coordinates and thermal parameters of all independent atoms in the unit cell (see next chapter). The output is the procrystal electron density in m81 format of JANA2000 (Petříček et al., 2000). The prior density is calculated on the same grid that will be used for the MEM calculation. For a calculation with the same input data, but with different grid, a new prior on the new grid must be calculated. The input file for BAYMEM can be used as an input file for PRIOR, provided two specific keywords are added. These keywords are `outputprior filename`, and a pair of keywords `prioratoms -endprioratoms`. For the definition of all keywords accepted by PRIOR see next chapter.

### 3.1 Programming language and system requirements

PRIOR is written in the programming language Fortran 90. It has been compiled and tested on two computers:

- Compaq AlphaStation ES40 with 500MHz 64-bit Alpha EV6 RISC processor and with Compaq Fortran Compiler V5.5-1877-48BBF
- Silicon Graphics Fuel with 500MHz IP35 MIPS R14000 processor and with MIPSPro Fortran compiler V7.4

The program obeys Fortran 90 standards and should be therefore compilable with any F90 compiler.

The program does not have any special system requirements. It does not use a graphical interface and the input can be edited with any plain text editor such as vi, nedit or emacs. The requirements for the system memory are comparable (and not higher) to those of the program BAYMEM. Thus, any computer used for computation of the electron density map with BAYMEM can be used for calculation of the prior electron density for that calculation by PRIOR.

### 3.2 Execution

The program is executed with command

```
prior input_file
```

where `input_file` is the full name of the ASCII file containing input parameters (see Section 4.3). There are no other command-line options.

# Chapter 4

## Specification of the input

### 4.1 Types of input

PRIOR requires two files as an input. The main input file contains specifications of the parameters of the electron density to be calculated. The input file is a free-format file based on keywords. Each keyword represents a specific parameter of the analysis and must be given a value.

Multiple spaces anywhere in the file are handled as a single space. The lines containing the character '#' in the first column are treated as comments and not interpreted. Blank lines anywhere in the input file are ignored. The length of the interpreted part of the line is 132 characters, every text exceeding this length is ignored.

Second input file contains the definition of the atomic form factors. Typically, this file is created only once and then used for all subsequent calculations with PRIOR. The file is described in the next section.

### 4.2 Definition of the atomic form factors

PRIOR calculates the density of the atom from the analytical approximation to the atomic form factors. The analytical approximation has the form:

$$f_a(\sin(\theta)/\lambda) = \sum_{i=1}^n a_i \exp\left(-b_i \frac{\sin^2(\theta)}{\lambda^2}\right) \quad (4.1)$$

The coefficients of the approximation must be listed in a separate data file. The entries in the file have this format in symbolic form:

Z	Atomic type					
	a1	a2	a3	a4	a5	a6
	b1	b2	b3	b4	b5	b6

The lines with coefficients are read in free format. The field **Atomic type** must always begin in the ninth column. Z is just a flag that is not interpreted by the program.

The maximum allowed number of exponentials ( $n$  in Eq. 4.1) is defined by the variable **MaxExpNr** in file **prior\_mod.f90**. **MaxExpNr** is currently 6.

The location of the file with the coefficients must be made available to the program. This is accomplished via an environment variable **PRIORDATA**, which must contain the full path to the file with coefficients.

A sample file with coefficients of neutral atoms and ions is supplied with PRIOR. Its name is **factors\_atomsions.fit**. It contains coefficients by Su & Coppens (1997) for atoms and ions with  $Z \leq 54$ , plus coefficients of a few other atoms from the International Tables for Crystallography (Wilson & Prince, 1999).

### 4.3 Specification of keywords

There are two basic types of keywords. The first type is followed by one or more values on the same line:

```
keyword value1 [value2 value3...]
```

The second type has the form:

```
begin_keyword  
line 1  
line 2  
...  
end_keyword
```

Each line may contain one or more values.

The name of the keyword of the first type is a single word without spaces. The name of the keyword of the second type is a pair of initial and final word (separated by a hyphen in the following text).

- Each value can be a constant of type real, integer or character. The type of the parameters and their allowed values are specified below. Alternative values are separated by slashes.
- The keywords are either compulsory or optional. The compulsory keywords must be specified for the analysis to proceed. The optional keywords can be omitted. If the optional keyword is omitted, the default value is used. The compulsory keywords are indicated by “compulsory keyword - no default” in the item “default”.
- description: Describes the function of the keyword, its influence on the output and relations to other keywords.

#### 4.3.1 name: accuracy

- value: positive real number
- default:  $10^{-6}$
- description: Defines the sensitivity of the calculation to small contributions of distant atoms to the density at each pixel. Approximately, the density below the value of **accuracy** is considered to be negligible and atoms contributing this or less density are ignored. The default value is safe for vast majority of cases, however, under special circumstances (very large and anisotropic displacement parameters) the value must be decreased. Higher value (up to  $10^{-5}$ ) can be often used with good results while considerably speeding up the calculation.

#### 4.3.2 name: prioratoms - endprioratoms

- value: contains a specification of the atoms to be inserted in the density either in free format or in the m40 format of the crystallographic computing system JANA2000 (Petříček et al., 2000).
- default: compulsory keyword - no default
- description: the free-format version has the form:



```

prioratoms
name1 type occupancy x y z U11 U22 U33 U12 U13 U23
name2 type occupancy x y z U11 U22 U33 U12 U13 U23
.
.
.
endprioatoms

```

Each line defines one atom. The name is an arbitrary identifier without spaces. Type is the symbol of the atomic type and is used to locate the parameters of the form factors of the atom in the form-factors file (Section 4.2). Occupancy includes the site multiplicity of the atom. For example, if the multiplicity of the position of an atom is 4, and the position is fully occupied, then the value of occupancy must be 0.25. Next values are the fractional coordinates of the atom and the harmonic displacement parameters.

Alternatively, the atoms can be defined in the format of the m40 file of JANA2000. The atomic part of m40 can be simply copied and pasted. Neither molecules nor anharmonic displacement parameters are supported in PRIOR.

In the line following the line with keyword `atoms`, the atomic types must be defined. The definition has the form:

`elements` *list of the elements*

Each atomic type has an associated number in JANA2000, which occurs after the name of the atom in the m40 file. The order, in which the atomic types occur after the word `elements` must correspond to the numbers assigned to the atoms of each type in JANA2000.

An example of the keyword `prioratoms` - `endprioratoms` in the m40 format is given below:

```

prioratoms
elements Nb S
S2b      2  2    0.500000 0.000000 0.506836 0.299940    000 0 0 0
  0.039839 0.006494 0.005365 0.000000 0.000000-0.000736
Nb1a     1  2    0.500000 0.000000 0.000000 0.426885    000 0 0 0
  0.040992 0.008020 0.006940 0.000000 0.000000 0.000350
endprioratoms

```

Note that the atom S2b is of type S, although it appears first. This is because the associated number following its name is 2, thus assigning to it the second atomic type in the list.

For the precise format specification of the m40 format see the documentation of JANA2000.

### 4.3.3 name: cell

- value:  $a b c \alpha \beta \gamma$
- default: compulsory keyword - no default
- description: Lattice parameters of the structure.

### 4.3.4 name: centers - endcenters

- value: Each line contains one centering vector.
- default: 0 0 0

- description: Defines the centering vectors of the space group. The components can be given both as fractions and as fractional number.

#### 4.3.5 name: normalize

- value: background/factor/none
- default: background
- description: The calculated prior will not be perfectly normalized due to the numerical errors and inaccuracy of the analytical approximation to the form factors. In vast majority of cases the raw calculated density contains slightly less electrons than appropriate. If **normalize background** is specified, these missing electrons are added in a form of a uniform background, i.e. a small constant is added to the density of each pixel to make the density normalized. With **normalize factor** each pixel is multiplied by a constant factor to achieve normalization. With **normalize none** the output density is not normalized.

#### 4.3.6 name: outputprior

- value: a valid filename of a non-existing file
- default: compulsory keyword - no default
- description: Specifies the filename of the output density.

#### 4.3.7 name: electrons

- value: real number
- default: compulsory keyword - no default
- description: Gives the number of electrons in the unit cell.

#### 4.3.8 name: progress

- value: yes/no
- default: yes
- description: With **progress no** the progress of calculation is not indicated. Useful if the standard output is redirected to a file — the long list of numbers indicating the progress is avoided.

#### 4.3.9 name: symmetry - endsymmetry

- value: each line contains definition of one symmetry operator
- default: compulsory keyword - no default
- description: This keyword contains the complete definition of the symmetry with exception of the centering vectors. Each line contains one symmetry operator. The format of the symmetry operators corresponds to the conventions used in the International Tables for Crystallography. If the symmetry operator is  $\{R|\tau\}$ , then the format of each entry is:

$$\tau_1 + R_{11}x_1 + R_{12}x_2 + R_{13}x_3 \quad \tau_2 + R_{21}x_1 + R_{22}x_2 + R_{23}x_3 \quad \dots$$

Example - n-glide perpendicular to b:

$1/2+x1 -x2 1/2+x3$

The translation components can be given either as fractions or decimal numbers.

## Chapter 5

# Example of the input file

This is an example of a typical input file for PRIOR. The result will be a prior electron density of oxalic acid dihydrate. The keyword `centers - endcenters` can be omitted in this case, where just the trivial centering vector is present, and has been added here just for illustration.

```
voxel 64 32 128
cell 6.1005 3.4999 11.9554 90. 105.781 90.
electrons 132
dimension 3
outputprior oxa_acid.m81

normalize background
accuracy 1.0E-6

centers
0 0 0
endcenters

symmetry
      x1      x2      x3
1/2-x1 1/2+x2 1/2-x3
      -x1      -x2      -x3
1/2+x1 1/2-x2 1/2+x3
endsymmetry

prioratoms
C1 C 1.000000 -0.044888  0.058625  0.051932  0.011064  0.012765  0.008718  0.002100  0.002882  0.000212
O1 O 1.000000  0.085182 -0.056346  0.150097  0.014725  0.020599  0.008216  0.005266  0.002952  0.000562
O2 O 1.000000 -0.221511  0.242500  0.036240  0.013328  0.019064  0.010130  0.005865  0.003696  0.000808
O3 O 1.000000 -0.048691  0.131377  0.321196  0.013390  0.019565  0.009611 -0.002958  0.004229 -0.001131
H1 H 1.000000  0.023487  0.021450  0.222777  0.022332  0.027018  0.017827  0.002872  0.005043  0.000198
H2 H 1.000000  0.078443  0.196622  0.387411  0.023278  0.034582  0.020062 -0.006285  0.000518 -0.004821
H3 H 1.000000 -0.141792 -0.045394  0.350385  0.026889  0.032439  0.030014 -0.009085  0.012404  0.002196
endprioratoms
```

## Chapter 6

# Description of the output

The principal output of PRIOR in the electron density in the m81 format of JANA2000<sup>1</sup>. No other output file is produced, all information is written to standard output. If a run of PRIOR in batch mode is needed (or the text written to the standard output should be preserved for whatever reason), the output can be redirected to a file by running PRIOR with this command:

```
$ prior inputfile > logfile_name
```

Setting `progress no` can be used in the input file to avoid listing of the numbers showing the progress of the calculation.

With the example from the previous chapter, PRIOR will write out this type of information to the standard output:

Reading input file

```
Taking atom C1
Atomic type: C
The site multiplicity of the atom is 1
Computing its electron density
Maximum density = 53.89803
Number of electrons for this atom = 6.00000
The "true" charge from the atomic form factors = 6.00000
Generating all equivalent positions of the atom
4/ 4
Taking atom O1
Atomic type: O
The site multiplicity of the atom is 1
Computing its electron density
Maximum density = 58.30443
Number of electrons for this atom = 7.99988
The "true" charge from the atomic form factors = 7.99988
Generating all equivalent positions of the atom
4/ 4
.
.
.
Taking atom H3
Atomic type: H
The site multiplicity of the atom is 1
Computing its electron density
```

---

<sup>1</sup>Additional formats can be added upon users' request.

```
Maximum density =      0.83953
Number of electrons for this atom =      0.99989
The "true" charge from the atomic form factors =      0.99989
Generating all equivalent positions of the atom
  4/ 4
Number of electrons given in the input file:  132.00000
Number of electrons in the map:  131.99723
Adding      0.00277 electrons as a flat background.
Final number of electrons:  132.00000
Writing resulting density to file oxa_acid.m81
Total time of the calculation [s]:  6.00
```

After reading the input, the density of each atom is computed individually. The type of the atom is identified from the atomic symbol in the input file, and the electron density is calculated using the corresponding coefficients of the analytical approximation to the atomic form factors. When the calculation is finished, basic information about the density is printed — the maximum calculated density and the total number of electrons. For checking purposes, the total number of electrons as defined by the coefficients of the form factors is also printed. For the sake of comparison between the two values, the former does not include possible partial occupancy of the atom and should be therefore always equal to the theoretical value following from the form factors.

When all atoms are processed, the total numbers of electrons present in the map, in the input file and after the normalization are printed, as well as the name of the output density file and the total time needed for the calculation.

## Chapter 7

# Practical issues

If the input is defined properly, the program usually proceeds without any problems. Occasionally, the total charge of the atoms differs from the charge expected from the form factors. This can have two reasons:

- Too coarse grid: The electron density of heavy atoms is extremely peaked around the nucleus. If the grid is not fine enough, this peaked region is not sufficiently covered by points of the grid and the total charge can be substantially influenced by the resulting inaccuracies. The only solution to this problem is to select finer grid division. In practice, this problem usually occurs only with very small or zero displacement parameters. Almost all realistic displacement parameters smear out the density sufficiently to smooth the peak at the nucleus.
- Too low accuracy: For majority of cases, the default accuracy  $10^{-7}$  is sufficient. However, for extremely large temperature factors the accuracy must be increased.

Occasionally, a large mismatch is found at the end between the expected number of electrons and the true number in the map. If the mismatch exceeds small fraction of electron corresponding to the inaccuracy of the form factors and is not caused by the effects described above, then the most likely reason is some kind of discrepancy in the occupancy of the atoms and the total number of electrons defined in the input file.

# Bibliography

- De Vries, R. Y., Briels, W. J. & Feil, D. (1996), ‘Critical analysis of non-nuclear electron-density maxima and the maximum entropy method’, *Phys. Rev. Lett.* **77**, 1719–1722.
- Palatinus, L. & van Smaalen, S. (2002), ‘The Generalized F-constraint in the Maximum Entropy Method — a study on simulated data’, *Acta Crystallogr. A* **58**, 559–567.
- Papoular, R. J., Collin, G., Colson, D. & Viallet, V. (2002), Direct imaging of fractional oxygen  $O_\delta$  in Hg-based High- $T_c$  superconductors, *in* B. Fry, ed., ‘Proceedings of the 21<sup>st</sup> Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering’, American Institute of Physics, Melville, NY.
- Petříček, V., Dušek, M. & Palatinus, L. (2000), *The crystallographic computing system Jana2000*, Institute of Physics, Praha, Czech Republic.
- Shmueli, U., ed. (1996), *International Tables for Crystallography, vol.B*, Kluwer Academic Publishers, Dordrecht.
- Su, Z. & Coppens, P. (1997), ‘Relativistic X-ray Elastic Scattering Factors for Neutral Atoms  $Z = 1-54$  from Multiconfiguration Dirac-Fock Wavefunctions in the  $0-12 \text{ \AA}^{-1} \sin(\theta)/\lambda$  Range, and Six-Gaussian Analytical Expressions in the  $0-6 \text{ \AA}^{-1}$  Range’, *Acta Crystallogr. A* **53**, 749–762.
- Wilson, A. J. C. & Prince, E., eds (1999), *International Tables for Crystallography, vol.C*, Kluwer Academic Publishers, Dordrecht.