

- [9] Dreyfuss, P., Keller, A. *J. Macromol. Sci.-Phys.* 1970, **B4**(4), 811
- [10] Fischer, E.W. *Pure Applied Chem* 1972, **31**, 113
- [11] Mandelkern, L. *Macromol.* 1969, **2**, 644
- [12] Ungar, G., Stejny, J., Keller, A., Bidd, I., Whiting, I.C. *Science* 1985, **229**, 386
- [13] Kovacs, A.J., Straupe, C. *Faraday Disc. Chem. Soc.* 1979, **68**, 225
- [14] Cheng, S.Z.D., Chen, J. *J. Polym. Sci. B Polym. Phys.* 1991, **29**, 311
- [15] Ryan, A.J., *et al.* Manuscript in preparation.
- [16] Barham, P.J., Keller, A. *J. Polym. Sci. B Polym. Phys.* 1989, **27**, 1029
- [17] Sadler, D.M. *Polymer Comm.* 1985, **26**, 204

## X-PLOR for Polycrystalline Fibre Diffraction

R.C.Denny<sup>1,2</sup>, M.Shotton<sup>3</sup> and V.T.Forsyth<sup>3</sup>

1 Biophysics Section, Blackett Laboratory, Imperial

College, London SW7 2BZ

2 CLRC Daresbury Laboratory, Warrington WA4 4AD.

3 Physics Department, Keele University, Keele,  
Staffordshire ST5 5BG

## Introduction

X-PLOR is a flexible software package for molecular dynamics, protein crystallography and NMR [1]. Among other things, the X-PLOR system allows a chosen target function to be used as an empirical X-ray energy term which can be added to the potential energy of a macromolecule. The techniques of energy minimization and simulated annealing can then be applied to the molecule to ensure reasonable stereochemistry and an optimal fit to the data. X-PLOR version 3.1 has certain limitations as far as fibre diffraction is concerned:

- i) It is not possible to connect the asymmetric units of crystallographic or strict non-crystallographic symmetry together with covalent bonds.
- ii) The diffraction data are treated as independent structure factors and cannot be input as composite intensities or overlapping Bessel function contributions.

A modified version of X-PLOR has addressed these problems for the case of structures with strict helical symmetry and continuous layer line intensity data and has already proved extremely useful in the study of filamentous viruses [2]. Once X-PLOR 3.1 has

been installed, these new and modified routines comprising FX-PLOR can be built into the system very easily. See

<http://www.molbio.vanderbilt.edu/fibre/software.html>.

The purpose of this work is to build upon FX-PLOR to provide a system suitable for both non-crystalline and polycrystalline fibres. Two examples are given below which demonstrate the power and flexibility of FX-PLOR using simulated fibre diffraction data.

## Modifications

In systems that give rise to fibre diffraction, molecules often span more than one asymmetric unit. For a molecule consisting of a simple chain of atoms this would involve defining an extra bond, two extra bond angles and three extra torsion angles to connect asymmetric units above and below the one in which the atomic coordinates of the molecule are defined. The symmetry linkage routines supplied with FX-PLOR cater for molecules where strict non-crystallographic (helical) symmetry applies. Polycrystalline fibre samples are more likely to have approximate helical symmetry and strict crystallographic symmetry. Therefore, the symmetry linkage routines in FX-PLOR required modification to allow the use of crystallographic symmetry and translations along a unit cell edge vector. This means that a continuous chain can be built passing from asymmetric unit to asymmetric unit or through the bottom and top of the unit cell.

Some of the crystallographic target functions available in X-PLOR have been adapted for polycrystalline fibre diffraction. The *F2F2* and *E2E2* target functions can be used with some modification to take account of reflection overlap. The *RESIDUAL* option has been more extensively modified to scale observed and calculated intensities together rather than structure factor moduli and to treat composite diffraction spots. These functions have been redefined for the case of a polycrystalline fibre as follows:

$$F2F2 = 1 - CC \left( \left\{ w_{hkl} I_{hkl}^{obs} \right\}, \left\{ w_{hkl} \sum_{h'k'l' \in M(hkl)} m_{h'k'l'} I_{h'k'l'}^{cal} \right\} \right) \quad (1)$$

$$E2E2 = 1 - CC \left( \left\{ \frac{w_{hkl} I_{hkl}^{obs}}{\langle w_{hkl} I_{hkl}^{obs} \rangle_{bin}} \right\}, \left\{ \frac{w_{hkl} \sum_{h'k'l' \in M(hkl)} m_{h'k'l'} I_{h'k'l'}^{cal}}{\langle w_{hkl} \sum_{h'k'l' \in M(hkl)} m_{h'k'l'} I_{h'k'l'}^{cal} \rangle_{bin}} \right\} \right) \quad (2)$$

where  $M(hkl)$  is the set of indices belonging to a composite diffraction spot indexed as  $hkl$ ,  $m_{h'k'l'}$  is the multiplicity factor of reflection  $h'k'l'$  and  $w_{hkl}$  is the weight to be applied to the observation.  $CC$  is the correlation coefficient defined,

$$CC(X, Y) = \frac{\langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle}{\sqrt{\langle (x - \langle x \rangle)^2 \rangle \langle (y - \langle y \rangle)^2 \rangle}} \quad \forall x \in X, y \in Y \quad (3)$$

The *RESIDUAL* function is now given by,

$$RESIDUAL = \sum_{hkl} w_{hkl} \left( I_{hkl}^{obs} - k \sum_{h'k'l' \in M(hkl)} m_{h'k'l'} I_{h'k'l'}^{cal} \right)^2 \quad (4)$$

where the scale factor  $k$  is the least-squares solution for scaling intensities,

$$k = \frac{\sum_{hkl} w_{hkl} I_{hkl}^{obs} \sum_{h'k'l' \in M(hkl)} m_{h'k'l'} I_{h'k'l'}^{cal}}{\sum_{hkl} w_{hkl} \left( \sum_{h'k'l' \in M(hkl)} m_{h'k'l'} I_{h'k'l'}^{cal} \right)^2} \quad (5)$$

Where the term ‘‘R-factor’’ is used in the text it refers to the quantity,

$$R = \frac{\sum_{hkl} w_{hkl} \left| I_{hkl}^{obs} - k \sum_{h'k'l' \in M(hkl)} m_{h'k'l'} I_{h'k'l'}^{cal} \right|}{\sum_{hkl} w_{hkl} I_{hkl}^{obs}} \quad (6)$$

## Example 1: Refinement of a Perturbed Structure

### Data and Model Preparation

A model structure of the D form of poly d(A-T).poly d(A-T) similar to that of Arnott *et al.*, [3], was placed in a P1 unit cell with dimensions  $a = b = 17.2\text{\AA}$ ,  $c = 24.1\text{\AA}$ ,  $\alpha = \beta = 90^\circ$ ,  $\gamma = 94.5^\circ$ . These cell dimensions provide approximately a quarter of the number of unique measurements for fibre diffraction (471 diffraction spots) than for single-crystal data (1843 reflections) between  $10\text{\AA}$  and  $2\text{\AA}$  resolution. This choice of cell and alternative descriptions have been reported by Forsyth *et al.*, [4]. The model was energy minimized without X-ray or packing terms and a set of structure factors were calculated. These structure factors were saved to be used as simulated data for subsequent trials. The structure was then subject to a molecular dynamics simulation at 500K for 0.5ps followed by a final round of energy minimization in order to produce a perturbed model as a starting point for refinement. The simulated data were used to

provide two sets of input: one in which all reflections were treated as independent observations (SC dataset) and one in which systematic overlaps and some close lying reflections due to cylindrical averaging were processed to produce a composite observation (PF dataset).

### Energy Minimization and Simulated Annealing

The first test was to compare the results of energy minimization (EM) when the X-ray energy term was included with the single-crystal dataset and the polycrystal dataset. Separate trials were run with the *F2F2* target function and the *RESIDUAL* target function. A necessary initial stage is to determine the weight with which the X-ray energy is to added to the other energy terms. This was done using the method suggested by Brunger *et al* [5], where a short molecular dynamics simulation is performed on the initial model and the energy gradient at the end of the simulation is compared with the X-ray energy gradient. The weight of the X-ray energy is then chosen so as to make the gradients equal. Clearly, the subsequent use of a different target function requires that the weight be recalculated. Each observation was assigned unit weight. A maximum of 120 cycles was allowed for each refinement.

The same weight determined for the X-ray energy for energy minimization is appropriate for simulated annealing (SA). An initial energy minimization must be performed before SA can begin to decrease any large energy gradients in the initial model. At each temperature stage in the cooling schedule, a molecular dynamics simulation was run for 0.025ps. The temperature was then reduced by an increment (25K) and the process repeated until the temperature had reached 300K. An EM refinement, with an identical protocol to the one described above, was performed when the final temperature had been reached.

### Results

Both the *F2F2* and the *RESIDUAL* target functions were employed in the EM trials. The rms deviation in atomic position was calculated between both the final models achieved by the refinements and the perturbed starting model ( $\Delta_s$ ) and between the refined models and the test model used to calculate the simulated data ( $\Delta_t$ ). The rms deviation between the test model and the starting model was  $1.32\text{\AA}$ . The results of the EM trials are summarized in table 1.

Data	Target	<i>R</i> start	<i>R</i> end	<i>CC</i> start	<i>CC</i> end	$\Delta t$	$\Delta s$
SC	<i>F2F2</i>	0.752	0.170	0.629	0.989	0.334	1.345
PF	<i>F2F2</i>	0.589	0.448	0.774	0.913	1.259	0.633
SC	<i>RESI</i>	0.752	0.157	0.629	0.991	0.296	1.295
PF	<i>RESI</i>	0.589	0.157	0.774	0.991	0.678	1.051

**Table 1:** The results of the energy minimization trials for the single-crystal (SC) and polycrystalline fibre (PF) datasets are shown. Start and end *R*-factors and correlation coefficients are given along with rms deviations in atomic position.  $\Delta s$  refers to the deviation between the perturbed starting model and the final model achieved in the refinement and  $\Delta t$  refers to the deviation between the final model and the test model.

Data	T start	<i>CC</i> end	<i>R</i> end	$\Delta t$	$\Delta s$
SC	4000	1.000	0.009	0.230	1.333
PF	4000	1.000	0.025	0.339	1.351

**Table 2:** The results of the simulated annealing trials are shown. T start refers to the initial temperature of the annealing schedule, other meanings are as for table 1.

The results for the SC data with the *F2F2* and the *RESIDUAL* target functions are similar, perhaps slightly favouring the use of the *RESIDUAL* function. The EM method was slightly less effective in the PF case, achieving some progress in 77 cycles of the minimization procedure with the *F2F2* target before the search was abandoned. Refinement of the PF data with the *RESIDUAL* target gave an identical *R*-factor and *CC* as in the SC case. However, the final rms deviation from the test model was slightly greater than for the SC case.

Only the *RESIDUAL* target was used for the SA trials on the basis that this target had allowed greater progress in the EM trials for the PF data. The results of these trials are summarized in table 2. Clearly, the result of the both trials show the efficacy of the method with an appropriate target function and annealing schedule.

### Example 2: Comparison of a Self-Rotation Search with a Rotation Search using a Perturbed Model

#### Data and Model Preparation

A test model and perturbed model were prepared in a similar way to that used in example 1, except that  $P2_1$  symmetry was imposed on the structures. A PF dataset was calculated from the test model with a resolution range of 25-2Å which was used as data in the subsequent rotation searches.

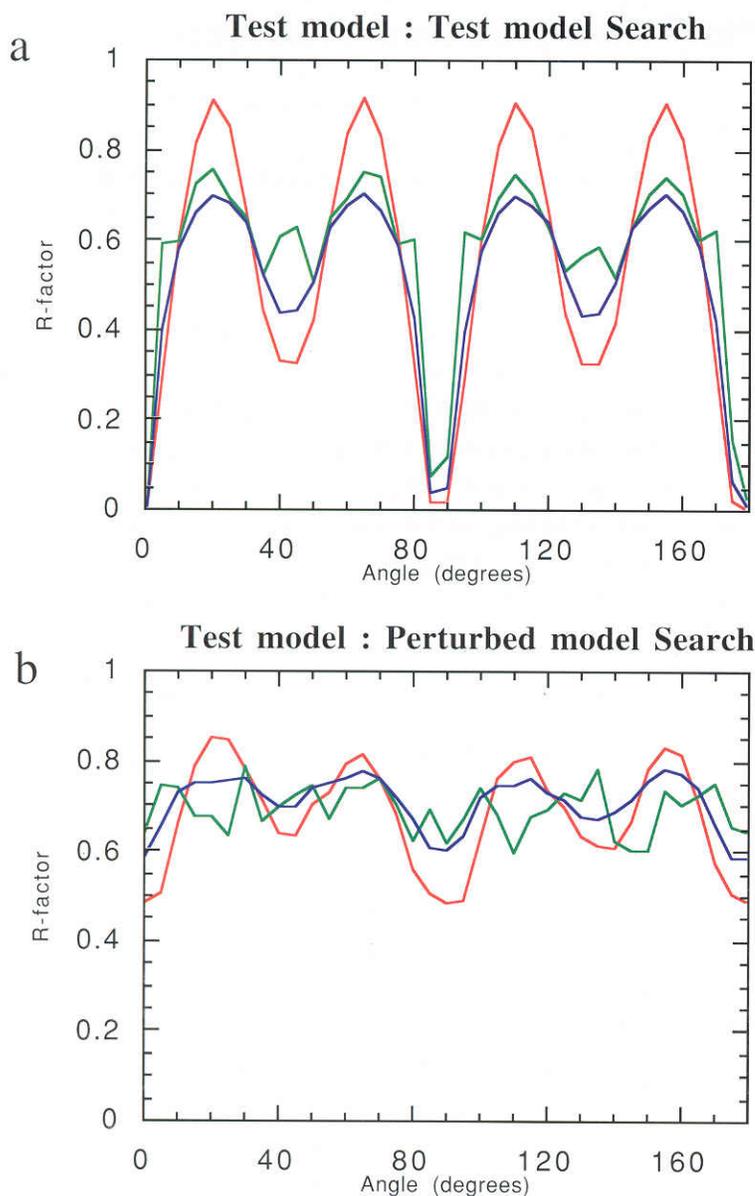
#### Rotation Searches

It is a simple task to set up a loop in the X-PLOR language to rotate the atomic coordinates by some angular increment about a specified axis. It is then possible to monitor the *R*-factor as a function of resolution for each rotation. Two searches were performed, both over 180° at 5° increments, one being a self-rotation search using the test model coordinates and the other being a cross-rotation search using the perturbed model coordinates.

#### Results

The self-rotation search shows the maximum *R*-factor that can be achieved by a simple rotation of the atomic coordinates in the unit cell. The results of this search are shown in figure 1(a): the *R*-factors are shown for resolution ranges 25.0-2.0Å, 25.0-4.0Å and 2.09-2.0Å. It is clear that the 25.0-4.0Å *R*-factor shows the greatest contrast between molecular orientations. The four-fold helical symmetry of the test model and the closeness of the cell to being tetragonal is shown by the near perfect repetition of the curves at 90° intervals. Deviations from this repeat are most easily discernible in the 2.09-2.0Å curve.

The rms deviation in atomic coordinates between the test model and the perturbed model was 0.91Å and the corresponding *R*-factor (at 0° rotation) was 0.587 using the full resolution range. Figure 1(b) shows the results of the cross-rotation search. Again, the 25.0-4.0Å *R*-factor shows the greatest contrast with molecular orientation while the 2.09-2.0Å *R*-factor



**Figure 1:** This shows the results of the rotation searches. R-factors are shown for data in three resolution ranges: 25-4Å (red), 25-2Å (blue), 2.09-2Å (green). Graph (a) shows the R-factor curves for the self-rotation search for the test model, graph (b) shows the curves for the cross-rotation search between the test model and the perturbed model.

indicates very little correlation to the low resolution curve or to the corresponding self-rotation curve.

### Conclusions

X-PLOR provides a very flexible language for performing calculations using atomic coordinates and X-ray diffraction data. The extension of this package to cope with fibre diffraction applications should enable researchers to identify more rapidly a strategy suited to dealing with their data processing requirements.

Pannu & Read [6] have implemented two maximum likelihood target functions in X-PLOR, increasing the radius of convergence of refinement. Bricogne [7] has formulated a likelihood function for overlapping intensity data. Implementation of this function within FX-PLOR may prove helpful where there is a significant amount of diffracting material in the unit cell not represented in the model.

### References

- [1] Brünger, A.T., Kuriyan, J & Karplus, M. (1987). *Science* **235**, 458-460
- [2] Wang, H. & Stubbs, G (1993) *Acta Cryst.* **A49**, 504-513
- [3] Arnott, S., Chandrasekaran, R., Puigjaner, L.C., Walker, J.K., Hall, I.H., Birdsall, D.L. & Ratliff, R.L. (1983). *Nucleic Acids Research* **11**, 1457-1474
- [4] Forsyth, V.T., Mahendrasingam, A., Langan, P., Pigram, W.J., Stevens, E., Al-Hayalee, Y., Bellamy, K.A., Greenall, R.J., Mason, S.A. & Fuller, W. (1990). *Inst. Phys. Conf. Ser.* **101**, 237
- [5] Brünger, A.T., Karplus, M. & Petsko, G.A. (1989). *Acta Cryst.* **A45**, 50-61
- [6] Pannu, N.S. & Read, R.J. (1996). *Acta Cryst.* **A52**, 659-668
- [7] Bricogne, G. (1991). *Acta Cryst.* **A47**, 803-829