

Processing fibre diffraction patterns from filamentous bacteriophage using the CCP13 suite of software

Liam C. Welsh

Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge.

The purpose of this article is to give a 'users' view of at least some of the CCP13 suite of software, and by illustrating how the suite is used in one particular line of work to give those who may not have used the suite some idea of what it can do.

The filamentous bacteriophages are flexible filaments about 60Å in diameter and 1-2µm long, with a shell of α -helical protein subunits surrounding a core of DNA. Fibre diffraction studies on the structure of filamentous bacteriophage first started about 30 years ago¹. One of the projects we are currently working on is the refinement of a model² of the Class I, fd strain, of filamentous bacteriophage against a 3Å fibre diffraction data set. A typical diffraction pattern from a magnetically aligned fibre of the Y21M mutant of fd² is shown in Fig 1. This pattern shows a mixture of Bragg sampled and continuous intensity on the equator and first two layer-lines, and continuous intensity elsewhere. If the low angle data in the regions of the pattern which exhibit Bragg sampling are required, then our approach has been to prepare a gel from an oriented fibre by adding water to a fibre in a capillary³. The inter-particle spacing is then large enough so that only continuous intensity is observed on the fibre pattern, however the increased disorientation in such a specimen means that only the low resolution data can be extracted from these patterns. These gel data are merged with the wide angle fibre data to form a complete data set.

We have been using the CCP13 suite of software in its current form to process our data for the past year. During this time we have collected data on film, Molecular Dynamics image plates, and on a MAR research image plate on station 7.2 of the SRS Daresbury. The recorded image of the diffraction pattern is first converted into BSL format (the CCP13 standard image format) using one of two programs: **CONV**, for scanned film or MAR data or **TIFF2BSL**, for MD data.

The next step is to calibrate the image using the program **FIX**. This program is useful in a number of ways. Firstly it allows the user to determine the specimen to film distance from a powder ring, the image centre, the image rotation, and the specimen tilt. Secondly once these parameters have been determined the reciprocal space co-ordinates of any Bragg reflexions may be measured. Furthermore we use **FIX** at each stage of the processing of the fibre data to simply display images output from other programs in the suite, as **FIX** permits up to ten images to be displayed simultaneously using either greyscale or false colour, and the contents of any window may be written out as a colour

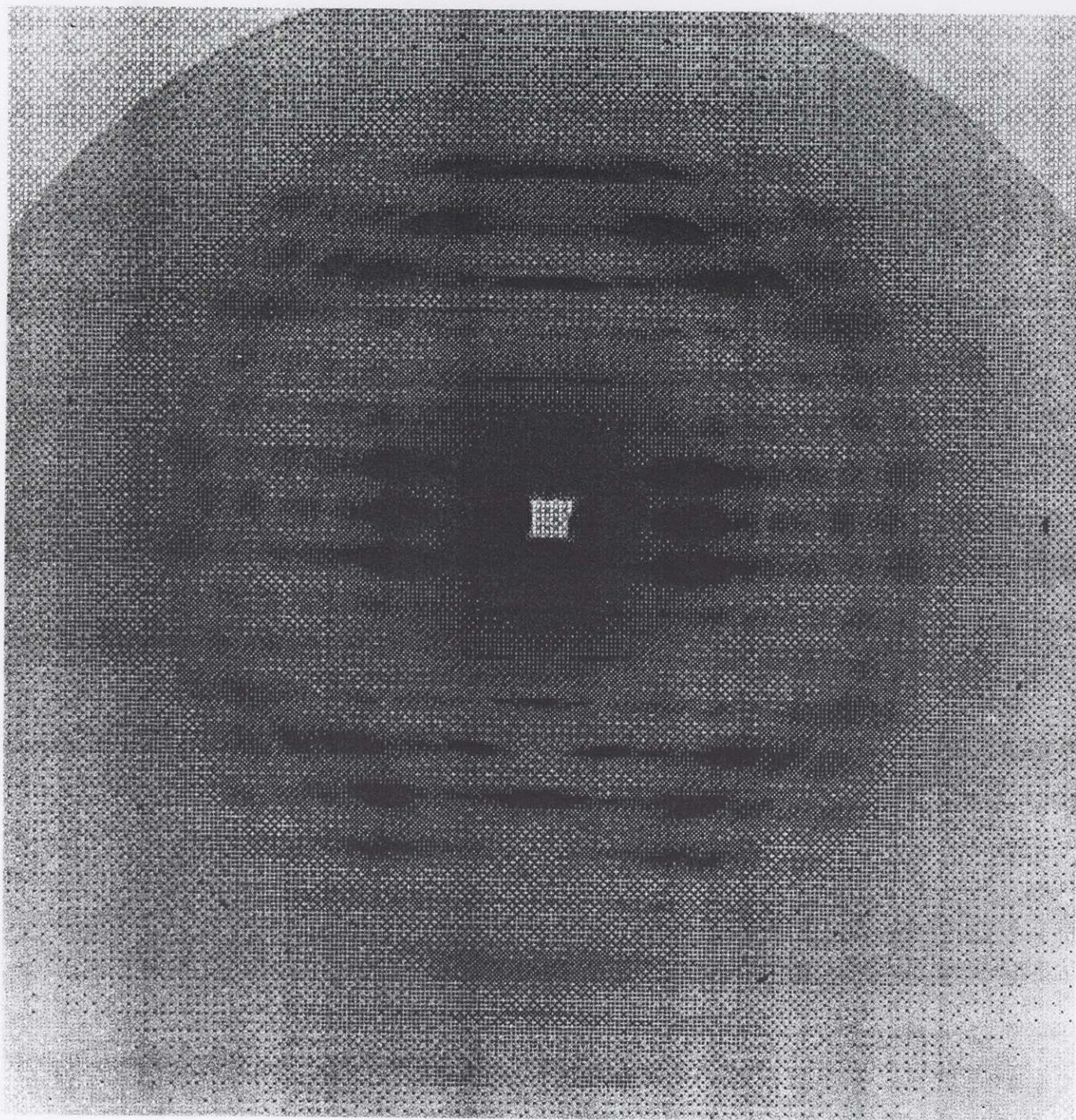


Figure 1.

X-ray diffraction pattern of the Y21M mutant of filamentous bacteriophage fd. The specimen was a fibre made by drying a concentrated solution of the phage in the bore of a 7T magnet. The diffraction pattern was recorded on film at station 7.2 of the SRS Daresbury using a wavelength of 1.488 Å. The fibre axis is vertical and the Si powder ring is at 3.14 Å.

postscript file. Alternatively the general image manipulation program **BSL** can be used to generate a contour plot of an image, and a utility, **GRID2PS**, used to produce hard copy. **FIX** also allows a user to look at the variation of the intensity in an image along a user defined line, and this feature has been useful in checking the integrated layer-line intensities that are finally obtained at the end of the data extraction process. **BSL** may also be used for this purpose.

Once a fibre diffraction image has been calibrated it can be mapped from film space to reciprocal space using **FTOREC**⁴, where it will be quadrant averaged by default. However if this is not desired the user may mask out any portion of the image using **BSL** prior to processing with **FTOREC**. A useful feature in **FTOREC** is the option to output an image corresponding to the standard deviations of the values in the reciprocal space bins, as this gives a good check on the accuracy of the calibration obtained in **FIX**. If for some reason the calibration is slightly incorrect then the only solution at the moment is to calibrate the image using **FIX** once more. However there are plans to implement some sort of refinement procedure for the image calibration parameters in **FTOREC** in the future.

After a satisfactory reciprocal space image has been obtained one may proceed to subtract a background from the image and fit the diffracted intensity using **LSQINT**⁴. This program can fit Bragg sampled data or continuous data, but not both together in a single image. In most of the recent work with filamentous bacteriophage we have been concerned with extracting only the continuous intensity from our diffraction patterns, and so we have used the maximum entropy peak fitting option in **LSQINT** (as opposed to the least squares option)⁴. This has proved to be quite straightforward to use and has yielded good data fairly quickly.

The ability of **LSQINT** to refine the unit cell and profile (disorientation and particle coherence length etc.) parameters together during the fitting is useful, but obviously can be slow if an attempt to refine many parameters together is made. The best policy is to get as close to the true profile parameters as possible by eye using the 'nofit' option (where only the calculated profiles are output and no intensity fitting takes place), prior to calculating a fit to the data. Of the four background subtraction methods available in **LSQINT**⁴ the most satisfactory for our work seems to be the Paul Langan roving window method. This method seems very robust and has the advantage that it does not require any information about the sampling point profiles and so can be used reliably even if one is not yet entirely happy with the estimates of the profile parameters. Alternatively it is perfectly possible for a user to remove a background from an image with some other program, if desired, before processing with **LSQINT**, and there are plans to introduce an interactive background subtraction facility into **FIX**.

The quality of a given fit to an **FTOREC** image may be judged from several different data output by **LSQINT**. Firstly an image corresponding to the calculated background is output and this can be inspected using **FIX**, a second image is also output consisting of the fitted intensity profiles and background image together, and this may be compared using **FIX** with the input **FTOREC** image. **LSQINT** also

calculates an R-factor for the fit as well as such quantities as the sum of the background and fitted peak pixels and the sum of the observed pixels minus background. Our experience suggests that the R-factor for a satisfactory fit will be somewhere in the region of 0.1 to 0.4 depending on the number of profiles being fitted. The R-factor is calculated over the entire FTOREC input image regardless of whether LSQINT is fitting the whole image or not, so large R-factors can arise if a portion of an image is not being fitted for some reason. Lastly the output $F(hkl)$ or $F(Rl)$ (for continuous data) may be plotted and compared with line plots from the raw image data in FIX.

The only difficulty encountered in LSQINT has been with over fitting the input image, resulting in noisy or even spurious intensity profiles. However this problem can be circumvented by using the 'sigma' option for specifying the standard deviations of the input image pixel values, and so weighting the fit. Since the standard deviations are not usually known they are calculated from the pixel values using: $sd = \min + \text{fact} * \sqrt{I}$, where min and fact are user defined. The choice of these parameters involves some guesswork, but in practice it has been possible to find values for these parameters which result in an accurate fit to the data and in smooth intensity profiles.

A further feature of LSQINT which has proved very useful in our work is the ability to easily simulate a fibre pattern from a set of calculated $F(hkl)$ or $F(Rl)$ (for continuous intensity). The user must supply an ASCII text file with columns containing h,k,l,R, and I, then LSQINT will output an image of a single quadrant of the simulated pattern for given lattice and smearing parameters. One may produce a full fibre pattern using the program **EXPAND**, which takes a BSL format image of a quadrant of a fibre pattern and mirrors it about the equator and meridian, as it is sometimes easier to see what is happening on the equator from a full pattern. Finally it should be noted that LSQINT can deal with diffraction patterns from fibres in which the c axis of the crystallites is not parallel with the fibre axis, as in fibres of the synthetic polymer PET⁵, and in fibres of wild-type fd filamentous bacteriophage².

So, the CCP13 suite as it stands, allows one to rapidly extract quantitative intensity data from a fibre pattern obtained from most commonly used area detectors, and also to simulate a fibre pattern from a set of calculated $F(hkl)$. Currently, what a user does with the set of observed $F(hkl)$ thus obtained or how a user calculates a set of $F(hkl)$ from a model, is outside of the scope of the software suite.

References

- D. A. Marvin (1966) *J. Mol. Biol.* **15**, 8.
- D. A. Marvin, R. D. Hale, C. Nave & M. Helmer Citterich (1994) *J. Mol. Biol.* **235**, 260.
- M. F. Symmons, L. C. Welsh, C. Nave, D. A. Marvin & R. N. Perham (1995) *J. Mol. Biol.* **245**, 86.
- R. Denny (1993) *The CCP13 Newsletter* No. 2
- H. Tadokoro (1979) *Structure of Crystalline Polymers* p162. John Wiley & Sons, Inc.