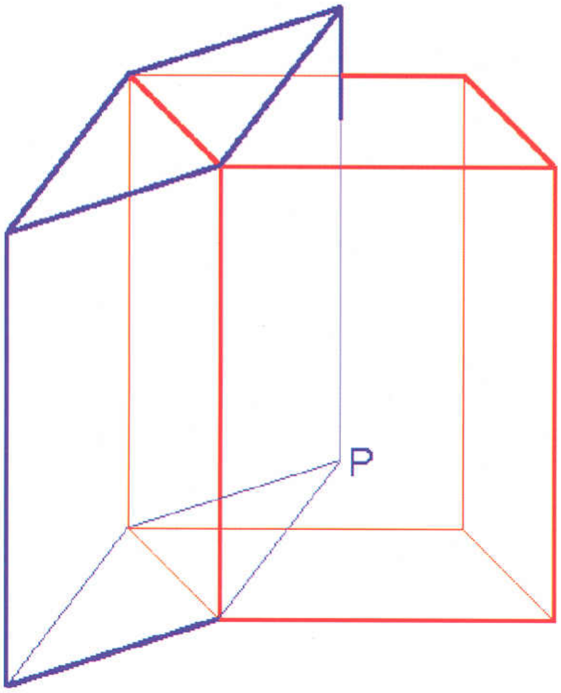

Figure 3: Relationship between the monoclinic unit cell (red) for cellulose Ib, and the triclinic unit cell (blue) for cellulose Ia. The point P has fractional coordinates (1/2,1/2,1/4) in the monoclinic system.

allomorphs. The small crystallite dimensions, inherent disorder, biphasic nature, and the presence of amorphous material mean the X-ray data from cellulose are low resolution and "murky." However, even under these conditions, X-ray fibre diffraction analysis has produced an enormous amount of essential structural information on this important material. Structural studies of cellulose demonstrate the truth of Arnott's assertion [17] that in the application of fibre diffraction "... with today's technology scrupulously applied, most gross errors are detectable."

References

[1] Meyer, K.H. and Misch, L., *Helv. Chim. Acta* (1937) **20**, 232-244.

[2] Nieduszynski, I. and Atkins, E.D.T., *Biochim. Biophys. Acta* (1970) **222**, 109-118.

[3] Sarko, A. and Muggli, R., *Macromolecules* (1974) **7**, 486-494.

[4] Gardner, K.H. and Blackwell, J., *Biopolymers* (1974) **13**, 1975-2001.

[5] Millane, R.P., in *Frontiers in Carbohydrate Research 2* (Ed. R. Chandrasekaran) 168-190 (Elsevier, London, 1992).

[6] Woodcock, C. and Sarko, A., *Macromolecules* (1980) **13**, 1183-1187.

[7] French, A.D. and Howley, P.S., in *Cellulose and Wood* (Ed C. Schuerch) 159-167 (Wiley, New York, 1989).

[8] Millane, R.P. and Narasaiah, T.V., in *Cellulose and Wood* (Ed C. Schuerch) 39-51 (Wiley, New York, 1989).

[9] Millane, R.P. and Narasaiah, T.V., *Polymer* (1989) **30**, 1763-1767.

[10] Aabloo, A., French, A.D., Mikelsaar, R.H. and Pertsin, A.J., *Cellulose* (1994) **1**, 161-168.

[11]Heiner, A.P., Sugiyama, J. and Teleman, O., *Carbohydr. Res.* (1995) **275**, 207-223.

[12] Kroon-Batenburg, L.M.J., Bouma, B. and Kroon, J., *Macromolecules* (1996) **29**, 5695-5699.

[13] Koyama, M., Helbert, W., Imai, T., Sugiyama, J. and Henrissat, B., *Proc. Natl. Acad. Sci. USA* (1997) **94**, 9091-9095.

[14] Finkenstadt, V.F. and Millane, R.P., *Macromolecules* (1998) **31**, 7776-7783.

[15] Sugiyama, J., Vuong, R. and Chanzy, H., *Macromolecules* **24** (1991) 4168-4175.

[16] Atalla, R.H. and van der Hart, D.L., *Science* (1984) **223**, 283-285.

[17] Arnott, S., in *Biopolimeri* (Associazione Italiana di Scienza e Tecnologia delle Macromolecole) Chapt. 23 (Italy, 1984).

CCP13 Software Development

M.W. Shotton[1,2], R.C. Denny[1] and V.T. Forsyth[3]

1 CLRC Daresbury Laboratory, Daresbury, Warrington, Cheshire WA4 4AD, U.K.

2 Biophysics Section, Blackett Laboratory, Imperial College, London, U.K.

3 Institut Laue-Langevin, BP 156 F-38042, Grenoble Cedex 9, France.

There have been significant changes to two of the programs in the CCP13 suite. The CONV program has been replaced by XCONV, which is driven by an OSF/Motif-based graphical user interface (GUI). XCONV provides for the conversion of various image data files to BSL format and is aimed at being more user-friendly than CONV, especially in the case of multiple file processing. There have also been several changes to XFIX, including the *incorporation of background estimation techniques. All programs in the CCP13 suite have now been developed for use on LINUX operating systems in addition to SOLARIS, IRIX and OSF. All CCP13 software can be downloaded from the web pages at http://www.dl.ac.uk/SRS/CCP13.*

1 Introduction

1.1 XCONV

XCONV can be used to gain access to other programs in the CCP13 suite, such as XFIX, by providing for the conversion of image data files to BSL format. BSL format is the common input/output format for two-dimensional diffraction data used by CCP13 programs. The BSL file format is described in the BSL manual at http://www.dl.ac.uk /SRS/NCD/manual.bsl.html. XCONV replaces the CONV program, and has been made more user-friendly by the addition of an OSF/Motif-based GUI (see Figure 1) and by the incorporation of some additional features that were not available in CONV, including multiple file operations and user-defined data conversion profiles that can be saved and reused.

1.2 XFIX

XFIX has been designed for the preliminary processing of two-dimensional fibre diffraction data. Information such as the pattern centre, detector and fibre orientation can be estimated and refined, lattice points can be superimposed on the pattern and integrated slices or scans through the pattern can be plotted and fitted. It is now also possible to use XFIX to estimate the background component of the diffraction pattern by one of the following three techniques:

(i) Paul Langan's "roving window" method.
(ii) Calculation of a circularly symmetric background.
(iii) Calculation of a "smoothed" background through iterative low-pass filtering, based on the method of M.I. Ivanova and L. Makowski [1].

Methods (i) and (ii) were previously available within the LSQINT data-fitting program, while method (iii) is new to the CCP13 suite.

2 Data processing

2.1 Data conversion to BSL format using XCONV

The XCONV GUI is shown in Figure 1. It is split into two sections relating to the input image data file(s) and the output BSL file. The operation of the GUI is described in detail in the help pages that may be downloaded with the executable from the CCP13 web site at http://www.dl.ac.uk/SRS/CCP13 or, alternatively, can be viewed at http://www.dl.ac.uk/ SRS/CCP13/program/xconv.html. The main improvements over the command line-driven program CONV are discussed below.

2.1.1 Multiple input files

Data file(s) may be selected for input to XCONV by entering the path and filename directly in the text window provided or by using the File Selection Tool, which is opened by clicking on the "Browse" button. In order to convert a series of similar data files (i.e. collected on the same detector), whose filenames differ only in their run numbers, the wildcard characters "%" and "#" may be inserted in the filename. "%" marks the position in the filename where the run number is to be inserted. If the run number is to be written into a fixed number of positions in the filename, "#" is used to indicate the position of a digit.

For example, my_data_%.dat would expand to:
my_data_1.dat, my_data_2.dat, ... , my_data_10.dat

whereas my_data_###.dat would expand to:
my_data_001.dat, my_data_002.dat, ... , my_data_010.dat

The first and last run numbers and increment can be entered in the fields provided. The multiple input files are converted and written to consecutive frames of the output BSL data file.

2.1.2 Data types

The following input data types can be selected from the pull-down menu:

(i)	float	float (4 bytes)
(ii)	int	unsigned int (4 bytes)
(iii)	short	unsigned short int (2 bytes)
(iv)	char	unsigned char (1 byte)

(v) smar small MAR image plate (1200×1200 pixels)
(vi) bmar large MAR image plate (2000×2000 pixels)
(vii) fuji Fuji image plate
(viii) rax2 R-Axis II image plate
(ix) psci Photonics Science CCD

If the input file type corresponds to one of the primitive types (i) to (iv) above, then the user must also input the dimensions of the input data, the number of bytes to skip (e.g. header information) and the pixel aspect ratio. These are entered automatically in the case of data types (v) to (ix).

2.1.3 Saving user-defined data conversion profiles

The "Save profile" button may be used to save the contents of the following fields to an ascii file:

(i) File type
(ii) Input pixels and rasters
(iii) Bytes to skip
(iv) Pixel aspect ratio
(v) Dynamic range for Fuji image plates
(vi) Byte swapping on/off
(vii) Output pixels and rasters

This saved profile can then be reused at a later stage with the "Load profile" button.

2.2 Background subtraction using XFIX

There are three methods for estimating the background component of a diffraction pattern using XFIX:

(a) Paul Langan's "roving window" method.
(b) Calculation of a circularly symmetric background.
(c) Calculation of a "smoothed" background through iterative low-pass filtering, based on the method of M.I. Ivanova and L. Makowski [1].

These can be selected from the "process" menu by clicking on "background". All three methods require

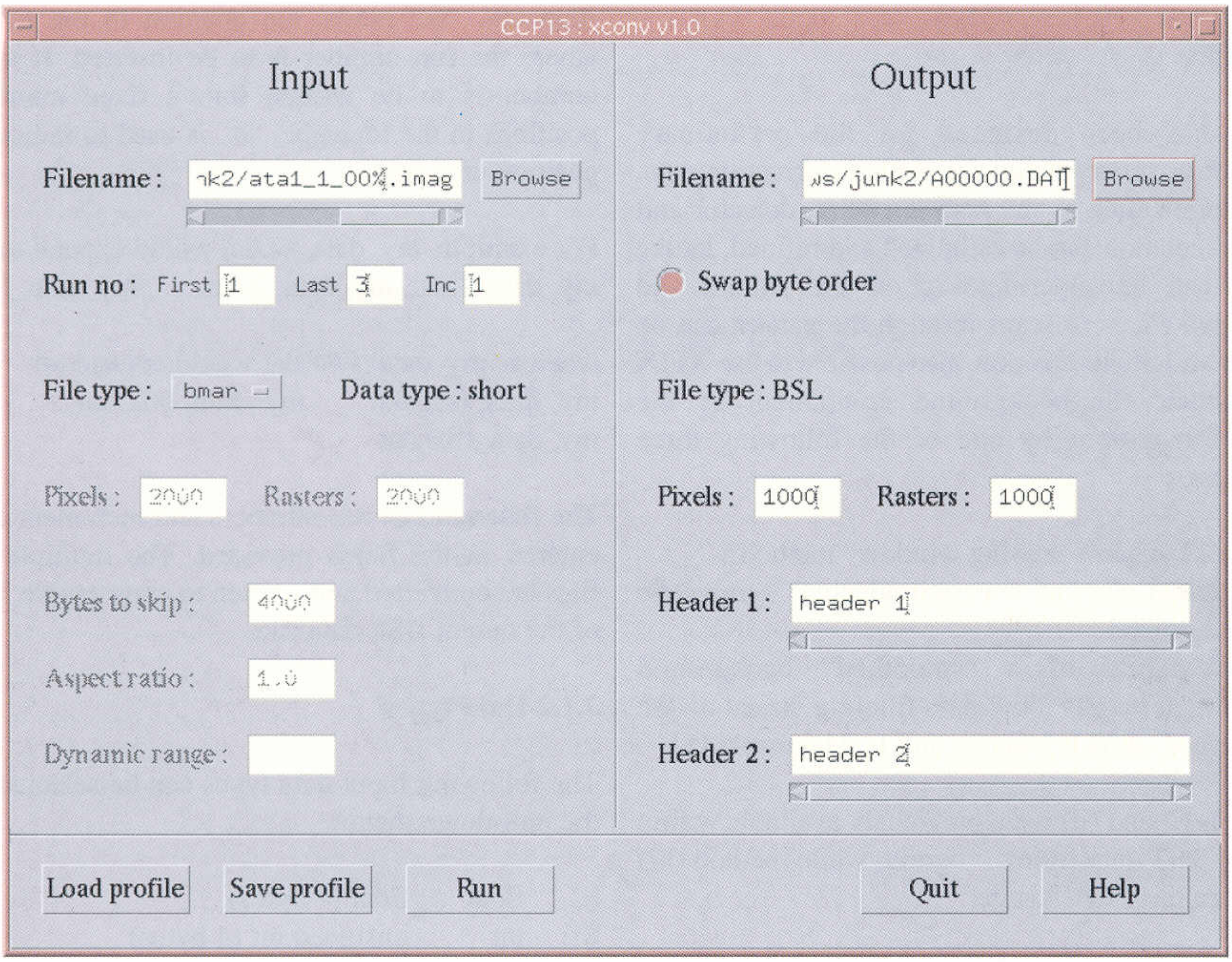

Figure 1: The XCONV GUI

the user to input the pattern centre and the extent of the pattern (minimum and maximum radii in pixel units) in relation to the centre. Pixel values outside these limits will be ignored in the background estimation. If the diffraction pattern is not circular and centred around a single point, then it is possible to discard values less than a user-input minimum so that, for example, all values less than or equal to 0.0 are ignored in the background calculation. By using the BSL program from the NCD software suite, it is possible to mask the areas of the image that are not of interest (for example by assigning a large negative value to unwanted pixels) and discarding these values in the background calculation.

The details of the different background estimation techniques are discussed below. In each case, once processing is completed, the user is prompted to view the calculated background. This will open a new XFIX interface with a 2-frame BSL file loaded. The first frame in this file is the estimated background and the next frame is the diffraction data with the estimated background subtracted. This provides a visual indication of the "goodness of fit" of the calculated background. The calculated background in frame 1 of the file can then be processed further using any of the three background estimation techniques. In this way, different background subtraction methods can be used consecutively if desired. Finally, the data minus background frame can be used with a fitting program such as LSQINT in order to measure the reflection intensities.

It should be noted that these methods of background estimation can be very time-consuming when used with a large image (e.g. 2000×2000 pixels). It may often be worthwhile to scale down the dimensions of the diffraction pattern using XCONV (e.g. to 200×200 pixels) to find the optimum parameters to be used in the process before operating on the full-size image.

2.2.1 Roving window method

The roving window background subtraction method of Paul Langan estimates the background by moving a window (of size input by the user) across the collected data. The pixel values within this window are sorted by value and those in the user-selected range are taken as background (except pixel values lying outside the pattern extents or specified by the user as values to discard). The average pixel value within this range is then assigned to be the estimated background at the centre of the window. Finally, a smoothing spline under tension is fitted to fill in the gaps between window centres.

2.2.2 Circularly-symmetric background

A circularly-symmetric background can be formed by the radial binning of pixel values followed by averaging those pixel values lying within a specified range. This average value is assigned to the particular radius and a smoothing spline under tension is fitted to yield the estimated background

2.2.3 Smoothed background - Iterative low-pass filtering

This method of background subtraction is based on that of M.I. Ivanova and L. Makowski [1]. It takes advantage of the fact that the background of a fibre diffraction pattern is typically composed of lower spatial frequencies than the diffraction maxima. Hence an iterative low-pass filter can be applied to separate the two components on the basis of their frequencies. The estimated background depends on the frequency limit of the filter in both X and Y.

The application of the low pass filter is achieved by the convolution of the observed diffraction pattern with a box car or gaussian function (the smoothing function) having an average value of unity taken over the number of pixels it occupies and a value of zero elsewhere. This function is convoluted with the real data at each pixel within the pattern limits (excluding pixels that are specified by the user as to be discarded). The result of applying this filter is an overestimated background, containing some intensity from the diffracted maxima (e.g. Figure 2(b)). This overestimated background is then subtracted from the real data to leave the diffraction maxima whose intensities are now underestimated (e.g. Figure 2(c)). Positive pixel values in the image resulting from this subtraction are then subtracted from the original data to yield the next estimate of the background. This is then convoluted with the smoothing function and the procedure is repeated, with images of the original data minus estimated background providing a visual indication of the "goodness of fit" of the estimated background. The results of applying the iterative low-pass filter to high-angle fibre diffraction data collected from the D conformation of DNA are shown in Figure 2.

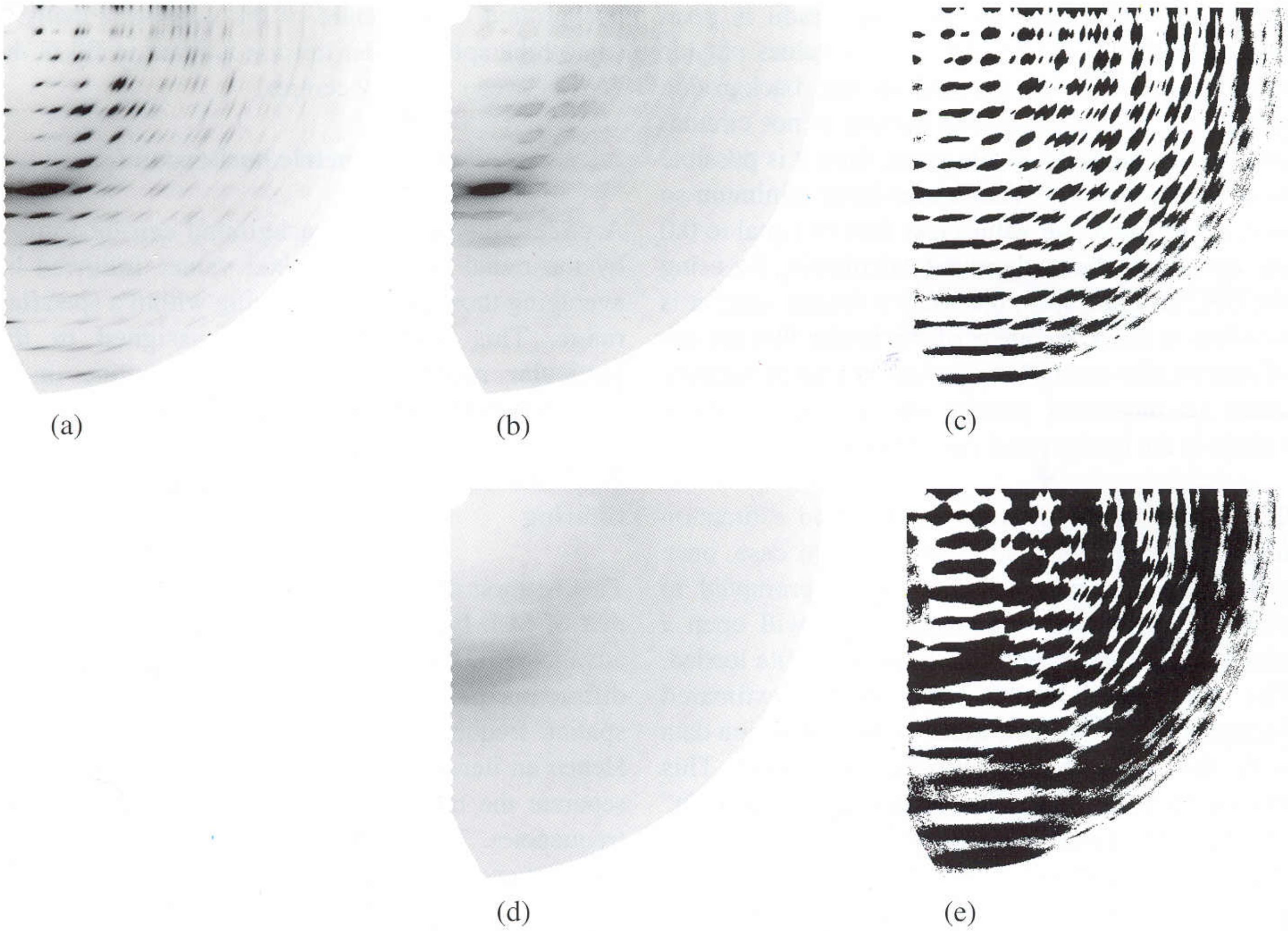

Figure 2: **(a)** High-angle fibre diffraction data collected from the D conformation of DNA (displayed at threshold levels of 2000, 0). **(b)** The estimated background following one cycle of low-pass filtering (display thresholds 2000, 0). **(c)** is the pattern (a) minus the estimated background (b) (display thresholds 1, 0). **(d)** The estimated background following five cycles of iterative low-pass filtering (display thresholds 2000, 0). **(e)** shows pattern (a) minus estimated background (d) (display thresholds 1, 0).

There are problems that exist with this method of background estimation at the edge of the pattern where information cannot be obtained for the convolution. This can lead to edge effects (typically an underestimation of the background), which become more pronounced with each cycle of filtering. In this case, it is possible to import a background calculated by another method to be used as the final background estimate at the edges of the pattern. This is then left as constant throughout each cycle of filtering. The backgrounds calculated by the different methods are then merged by fitting a smoothing spline under tension.

Once the background estimation process has been completed, the user will be prompted to view the estimated background and original pattern minus estimated background as described above. This will open a new XFIX interface. Once the estimated background has been inspected, the user may then choose to perform more iterations of filtering. Hence it is possible to view the estimated background after each iteration if desired.

References

[1] M.I. Ivanova and L. Makowski, *Acta Cryst.* (1998) **A54**, 626-631.