Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Untangle, a tool for filtering overlapping diffraction patterns from multicrystals

Lieven Buts,* Minh-Hoa Dao-Thi, Lode Wyns and Remy Loris

Department of Ultrastructure, Vlaams Interuniversitair Instituut voor Biotechnologie, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium

Correspondence e-mail: lievbuts@vub.ac.be

Received 25 February 2004
Accepted 25 March 2004

Standard crystallographic data-processing protocols are based on single-crystal models; data from aggregates of multiple crystals with different orientations are difficult to process. In certain cases, it is possible to separately index the diffraction patterns from the dominant crystals in the aggregate. *Untangle* is a program designed to identify and eliminate overlapping spots from such patterns in order to improve data quality. The program has a Python core with a simple and highly portable graphical user interface, permitting visual verification of the process and interactive modification of the overlap threshold. The software is available under an open-source license.

1. Introduction

All commonly used programs for the processing of crystallographic diffraction data, including the *HKL* package (Otwinowski & Minor, 1997), *MOSFLM* (Leslie, 1992; Powell, 1999) and *d*Trek* (Pflugrath, 1999), are based on a model describing a single consistent diffraction pattern generated by a single ideal crystal lattice. All programs include a mosaicity model, with a mosaic spread that is typically between 0.1 and 1.0°.

It is not uncommon that the only available crystals are aggregates of single crystals with different orientations, giving rise to two or more distinct but partially overlapping diffraction patterns. This situation is distinct from merohedral twinning, in which an exact symmetry relates the constituent lattices and the diffraction patterns overlap completely. Often, there are many different orientations contributing with similar intensities, yielding a very complex pattern that cannot be interpreted. In certain situations, however, it is possible to separately index two or more dominant patterns, resulting in one data set for each subcrystal orientation. Standard data-reduction software is not aware of the presence of the other patterns; spots from the current orientation which overlap with spots from another orientation will integrate to incorrect values. Outlier detection will eliminate some of the incorrect values, but it cannot systematically identify the pattern of overlaps.

Here, we present a program that identifies and optionally eliminates overlapping reflections from different crystal orientations in order to improve data quality.

2. Implementation

The core of the program is a collection of Python objects that take care of reading the input files and identifying all reflections from different orientations that lie within a specified distance (the overlap threshold) of each other. Overlap detection is speeded up tenfold by binning the data in two dimensions for faster look-ups. There is a simple driver script to process data sets in batch mode, as well as a Tkinter-based graphical user interface (GUI) for interactive visualization and control. The latter functionality is very useful for assessing the situation and determining the optimal overlap threshold for the filtering operation. The threshold can thus be adjusted to find the best compromise between completeness and quality.

The program can read *HKL* package x files and *CCP*4 MTZ files produced by *MOSFLM* (including XDET and YDET columns). After processing, the program outputs reflection files from which the overlapping spots have been removed. Although finding an initial indexing solution may still require some creativity (such as selecting the right frame for autoindexing) and familiarity with the software, *Untangle* can assist in the indexing of the additional patterns by reading *XDisplayF*'s peaks.file and removing the peaks corresponding to orientations that have already been successfully indexed. Autoindexing should subsequently converge on another orientation.

A more advanced approach to this problem would be to attempt to separate the contributions to overlapping spots by profile-fitting in a manner analogous to the method used by the

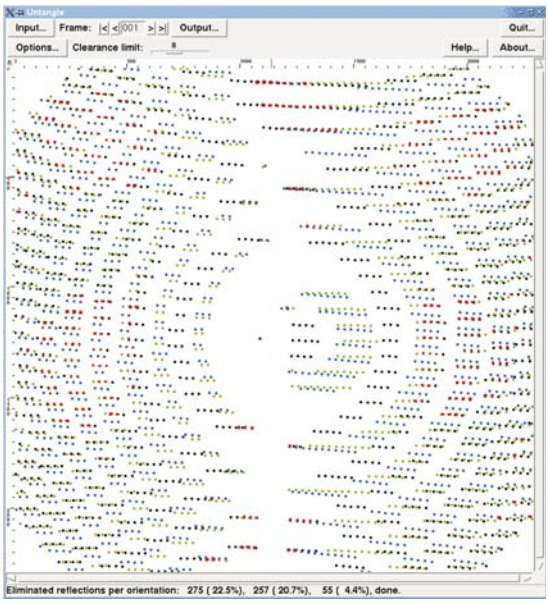

Figure 1
Screenshot of *Untangle* in action, displaying three overlapping diffraction patterns from the detector image in Fig. 2. The green and blue patterns correspond to crystals rotated by about 1° around the beam axis, resulting in a rotation of the pattern within the detector plane. The black pattern arises from a crystal rotated by about 3° about the spindle axis with respect to the other two.

PrOW program (Bourgeois, 1999) for the processing of Laue diffraction patterns and monochromatic data sets with poorly separated spots. This is a possible direction for future development.

Both the data-processing core and the GUI are highly portable and run on a wide variety of computer platforms. The code is licensed under the GNU General Public License and is available from http://ultr.vub.ac.be/untangle/, with installation and usage instructions.

3. Case study

Figs. 1 and 2 show three related $P2_12_12_1$ diffraction patterns from crystals of a complex of the bacterial toxin CcdB and a gyrase fragment (our unpublished results). In spite of extensive trials, it was not possible to obtain single crystals for this complex and it was decided to collect data from one of the better aggregates showing three dominant diffraction patterns. The three orientations were separately indexed using *DENZO*. Overlapping reflections from different orientations which are filtered out by the program are marked in red. Overlaps occur in a number of concentrated areas distributed over the entire detector surface.

Table 1 shows the effect of overlap elimination on the data set. Removing the reflections that are affected by overlaps reduces the completeness of the individual data sets by 5–10%, but since the different orientations lose different spots, the overall completeness of the combined data only falls from 100 to 96.3%, with the average redundancy falling from 36 to 30. The combination of data from multiple crystals partially compensates for the loss of redundancy owing to rejected spots. The fitting statistics indicate a useful improvement of the data quality. A further illustration is provided by a number of reflections which should be systematically absent in $P2_12_12_1$ but have high $I/\sigma(I)$ after processing the raw data with *SCALEPACK*. The filtering operation correctly rejects these data points, as it does in the rest of the data set. For instance, the $I/\sigma(I)$ for $h = 0, k = 15, l = 0$ falls from 10.2 in the raw data to 0.9 in the filtered data.

In conclusion, *Untangle*'s specific elimination of erroneous values owing to pattern overlaps appears to act synergistically with the rejection criteria already present in *SCALEPACK*.

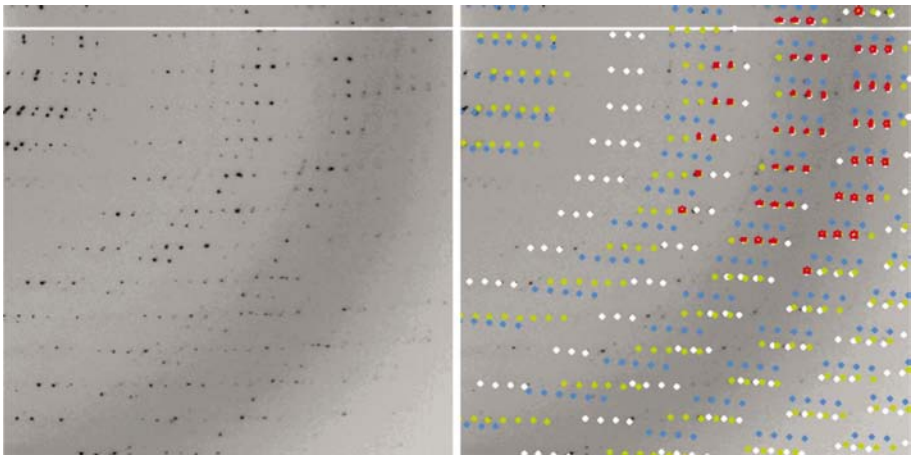

Figure 2
Detector image corresponding to the data in Fig. 1. The left panel shows a section of the detector image. The right panel shows the same section with the different pattern predictions overlaid. White markers correspond to the black pattern in Fig. 1.

Table 1
Comparison of data quality before and after *Untangle* filtering with an overlap threshold of four pixels.

χ^2, fitting χ^2; $R_{\rm merge}$, linear merging error; $N_{\rm reject}$, number of reflections flagged for rejection by *SCALEPACK*; C, completeness; $N_{\rm filt}$, number of spots eliminated during overlap filtering. There are 12 351 unique reflections for the unit cell in the resolution range 25–2.8 Å.

		Unfiltered				Filtered				
	$N_{\rm total}$	$N_{\rm reject}$	$R_{\rm merge}$	χ^2	C (%)	$N_{\rm filt}$	$N_{\rm reject}$	$R_{\rm merge}$	χ^2	C (%)
Crystal 1	149226	1706	0.103	2.2	99.8	24417 (16.3%)	107	0.074	1.1	88.7
Crystal 2	149200	4213	0.162	3.7	99.9	22386 (15.0%)	129	0.092	1.2	90.7
Crystal 3	149012	1830	0.181	3.4	97.5	18542 (12.4%)	82	0.097	0.9	91.8
Combined	447438	14351	0.200	4.4	100.0	65345 (14.6%)	360	0.108	1.3	96.3

The authors acknowledge the use of beamline ID14-1 at the ESRF in Grenoble. We thank Savvas Savvides, Dominique Maes and Klaas Decanniere for helpful discussions and Jan Ghyssaert for interface testing.

References

Bourgeois, D. (1999). *Acta Cryst.* D**55**, 1733–1741.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EAMCB Newsl. Protein Crystallogr.* **26**, 27–33.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pflugrath, J. (1999). *Acta Cryst.* D**55**, 1718–1725.
Powell, H. R. (1999). *Acta Cryst.* D**55**, 1690-1695.