University of Groningen

An optimal strategy for X-ray data collection on macromolecular crystals with position-sensitive detectors

Vicković, Ivan; Kalk, Kor H.; Drenth, Jan; Dijkstra, Bauke W.

Published in:
Journal of Applied Crystallography

DOI:
[10.1107/S0021889894004681](https://doi.org/10.1107/S0021889894004681)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1994

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Vicković, I., Kalk, K. H., Drenth, J., & Dijkstra, B. W. (1994). An optimal strategy for X-ray data collection on macromolecular crystals with position-sensitive detectors. *Journal of Applied Crystallography*, 27(5), 791-793. <https://doi.org/10.1107/S0021889894004681>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

An Optimal Strategy for X-ray Data Collection on Macromolecular Crystals with Position-Sensitive Detectors

BY IVAN VICKOVIĆ,* KOR H. KALK, JAN DRENTH AND BAUKE W. DIJKSTRA†

*Laboratory of Biophysical Chemistry and BIOSON Research Institute, Chemistry Department,
University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands*

(Received 4 February 1994; accepted 3 May 1994)

Abstract

X-ray data collection on macromolecular crystals is preferably done with minimum exposure time and high completeness. A Fortran procedure – *DCS* – has been written in the environment of the *MADNES* program to predict the completeness of data before the start of actual data collection. In addition, the program can check and refine the data-collection strategy and suggest optimal settings and rotation ranges for one or more crystals in different orientations that give highest completeness in minimum exposure time, thus extending the life of the crystal. The method has been tested on previously collected data as well as on new structures. Since the program has been in full use in this laboratory, the completeness of collected data has improved significantly.

Introduction

To collect highly complete diffraction data sets from protein crystals in an efficient way is sometimes a complicated problem, especially for low-symmetry space groups and area detectors with a small asymmetric detection area. Moreover, crystals are subject to radiation damage and may die before a complete set of data has been collected, necessitating the combination of data from several crystals. In the present article, we present an automated approach (*DCS*) to collect a complete set of diffraction data efficiently. In this approach, we try to minimize the data-collection time by minimizing both the number of different settings and the scan range used for each setting. Given a starting crystal orientation, oscillation axis, appropriate detector swing-out angle and crystal-to-detector distance, the program can predict the starting scan angle and scan range that give the highest completeness. This may be done for several user-defined starting settings or by searching completely automatically a user-defined region of reciprocal space. Subsequently, the program may scan through reciprocal space for a second (or third) setting that best covers the missing data. This approach has been developed in the environment of the *MADNES* program (Messerschmidt & Pflugrath, 1987), allowing an easy evaluation of forbidden regions where collisions may occur. *MADNES*

is a general program for data collection by area detectors. We use it to collect and process diffraction data from our FAST area detectors. In our approach, we retain the basic notation and concepts of *MADNES*.

Method

The *DCS* method is designed to predict the completeness of data in all space groups available for proteins. After pre-alignment of the crystal and auto-indexing, it allows a quantitative and qualitative evaluation of different reference zero positions and scan ranges. The reference zero position (see Fig. 1 of Messerschmidt & Pflugrath, 1987) is defined by the goniostat angles ω , κ and φ , the detector swing-out angle θ , the crystal-to-detector distance, the rotation axis, and the maximum resolution. It can be chosen arbitrarily but in many cases it is useful to orient one of the crystal axes parallel to the direct beam or along the rotation axis. Alternatively, we have excellent experience with aligning one of the crystal axes parallel to the diagonal of the detector. The program automatically suggests the optimum scan range and the best starting angle, either for a user-defined reference zero position or by scanning through reciprocal space, while taking into consideration the geometric limitations of the goniostat. The scan range is defined as the total rotation applied about one of the goniostat ω , κ or φ axes. To find the scan range and starting angle that give the highest completeness, several starting angles and scan ranges are analysed in a so-called super-range, which is the total rotation possible about an axis without collisions or other geometric limitations. Next, the best scan range found is narrowed to find an optimal range, which gives highly complete data in the narrowest possible scan range. In this procedure, completeness is calculated exactly by generating all possible unique reflections to the desired resolution that are not systematically absent.

The program can handle anomalous data. To obtain the best anomalous data, it is desirable to collect anomalous pairs within a very short time span of each other. Therefore, a symmetric setting is usually used, with one of the axes parallel to the rotation axis. The program is able to determine the completeness of, for example, *hkl* reflections separate from the completeness of *hkl* reflections.

* On leave from the Faculty of Science, University of Zagreb, Croatia.
† Author for correspondence.

Table 1. *Completeness of data as a function of starting scan angle*

The reference zero position was $\omega = 180.665^\circ$, $\kappa = 10.602^\circ$, $\varphi = -140.407^\circ$, the crystal-to-detector distance was 90.22 mm and the detector swing-out angle was -14.90° . ω was the scan axis and the scan range was 140° . This setting is such that the c^* axis is near the ω axis. Reflections were predicted to a maximum resolution of $3 \text{ \AA}$. These values are the same as those used for the original data collection. Completeness is calculated exactly by generating all possible unique reflections to the desired resolution that are not systematically absent. The reduced cell was obtained by normalizing the unit-cell volume to $10\,000 \text{ \AA}^3$, resulting in a division of the true cell axes by 3.17.

Starting scan angle ($^\circ$)	Completeness (%)	
	True cell	Reduced cell
-138.0	68.38	61.06
-130.0	75.37	65.93
-120.0	82.30	77.43
-110.0	84.17	82.74
-100.0	83.95	82.74
-90.0	83.86	82.74
-80.0	83.71	82.74
-70.0	80.59	80.53
-60.0	72.34	74.34
-50.0	69.69	70.80
-40.0	77.77	76.11
-33.0	82.57	80.53

As with other area-detector diffractometers, the FAST system (Arndt & Thomas, 1985) suffers from the 'cusp problem', caused by the inability to collect diffraction data from a region in reciprocal space close to the rotation axis in a single setting. Especially for space groups with low symmetry, such as triclinic or monoclinic space groups, and for data collected with detectors with a small detection area, high completeness is difficult to achieve. To collect the missing data, a new setting has to be found. Therefore, the program can simulate data collection for a number of second settings, taking into account the data collected at the previous setting(s), and a new optimum scan range and starting scan angle for the missing data are obtained. This procedure may be repeated several times. It can also be used to add data to an existing incomplete data set, for instance from a crystal that has prematurely died. Although such a subsequent setting can be estimated by experience, our procedure automatically investigates possible reference zero positions for this setting by rotating the reciprocal space in the laboratory coordinate system (*e.g.* in steps of 10 to 20° ; see example given later) and searching for that portion of reciprocal space where the majority of the missing reflections can be found. This allows for an efficient collection of missing data. The procedure has been developed as a Fortran subroutine to the *MADNES* program (Messerschmidt & Pflugrath, 1987). As input, the routine requires the crystal-orientation matrix, cell dimensions, space group, detector parameters and resolution. Other data-collection programs that can provide this information could, in principle, use this routine. However, for an optimal result the routine needs the versatility of a κ or three- or four-circle goniostat to scan reciprocal space efficiently.

Table 2. *Completeness of data as a function of scan range*

The reference zero position is the same as in Table 1. The starting scan angle was -110° . The reduced cell was obtained by normalizing the unit-cell volume to $10\,000 \text{ \AA}^3$, resulting in a division of the true cell axes by 3.17.

Scan range ($^\circ$)	Completeness (%)	
	True cell	Reduced cell
140.0	84.17	82.74
135.0	84.00	82.74
130.0	83.95	82.74
125.0	83.83	82.74
120.0	83.84	82.74
115.0	83.69	82.74
110.0	83.60	82.74
105.0	82.89	82.74
100.0	80.35	80.53
95.0	76.14	76.11
90.0	71.60	74.34
85.0	67.26	67.70
80.0	62.82	65.04

Results

Data for bovine rhodanese, the structure of which was solved in our laboratory (Kooystra, Kalk & Hol, 1988), were used to test the procedure to predict the completeness. The crystals belong to monoclinic space group *C2* with cell dimensions $a = 156.0$, $b = 49.0$, $c = 42.3 \text{ \AA}$ and $\beta = 98.6^\circ$. The original data set for this structure was only 68% complete. Using the same reference zero position as in the original data collection, we analysed the completeness as a function of starting scan angle (Table 1) and scan range (Table 2). The speed of prediction can be improved significantly by using a smaller unit cell (for example with a normalized volume of $10\,000 \text{ \AA}^3$). This produces somewhat less precise but still very usable results. The original and reduced cell show similar behaviour (Tables 1 and 2) but the CPU time was reduced from 590 to 120 s on a VAX3200. From Table 1, it can also be seen that for a scan range of 140° the best results are obtained for a starting scan angle of between -110 and -80° , with a completeness of over 80%. The original data collection started at -138° , giving only 68% completeness. A complete scanning of reciprocal space to find the best starting setting took 237 min (reduced cell; 20° steps in ROTY, ROTZ, scan range and starting scan angle, resulting in a starting setting with even 90% completeness).

Next, the scan range was shortened, -110° being taken as the starting scan angle. From Table 2, it is clear that the scan range can be reduced to 105° without a significant reduction of completeness. The original and reduced cell produced similar results.

From these results, it is obvious that for this crystal a high completeness cannot be achieved in one setting. Therefore, a second setting has to be found to increase the completeness. This was done by scanning through a range of settings (Table 3) and by calculating for each of them the fraction of missing reflections that can be recovered. From Table 3, which lists the ten best

Table 3. *The ten best second settings*

Reciprocal space was searched by rotating about the laboratory z and y axes (ROTZ and ROTY, respectively) in steps of 10° . The starting scan angle was varied in steps of 20° . ω_2 , κ_2 and φ_2 define different reference zero positions for the second setting. Their values are determined by the ROTZ and ROTY settings. The crystal-to-detector distance, the detector swing-out angle and the scan axis are the same as in Table 1. The original cell dimensions were used in this analysis. The last column indicates the percentage of reflections missing after data collection in the first setting that could be recovered in the second setting.

	ROTZ ($^\circ$)	ROTY ($^\circ$)	ω_2 ($^\circ$)	κ_2 ($^\circ$)	φ_2 ($^\circ$)	Starting scan angle ($^\circ$)	Scan range ($^\circ$)	Coverage of missing reflections (%)
1	80.00	40.00	124.82	130.56	-147.34	-30.00	80.00	98.65
2	80.00	40.00	124.82	130.56	-147.34	-30.00	70.00	97.84
3	80.00	30.00	125.30	130.31	-157.20	-30.00	80.00	97.57
4	80.00	30.00	125.30	130.31	-157.20	-30.00	70.00	96.76
5	80.00	20.00	125.72	130.20	-167.14	-30.00	80.00	96.49
6	70.00	40.00	127.41	115.10	-137.73	-30.00	80.00	95.95
7	80.00	20.00	125.72	130.20	-167.14	-30.00	70.00	95.77
8	70.00	30.00	130.29	113.05	-146.70	-30.00	80.00	93.06
9	60.00	40.00	127.55	101.92	-129.36	-10.00	70.00	92.79
10	70.00	40.00	127.41	115.10	-137.73	-30.00	70.00	91.98

second settings, it can be seen that one second setting is sufficient to collect nearly all the missing data. Table 4 shows the combined completeness of both settings as a function of the scan range in the second setting. Thus, data at nearly 100% completeness can be collected in two settings. In the 80° scan of the best second setting, it was possible to collect 1095 missing unique reflections and 3012 unique reflections that were already present in the data from the first setting. This gives sufficient overlap to scale and merge the data from the two settings. In total, 10751 measurements were present for 6472 unique reflections, resulting in an average redundancy of 1.66.

Discussion

We have developed a method – *DCS* – to predict the completeness of diffraction data to be collected on an area detector. Especially for detectors with a small asymmetric detection area, such as the FAST area detector, this procedure ensures the collection of complete data sets as efficiently as possible. The method is also suitable for data collection for missing reflections from incomplete data sets of other crystals and may therefore be of general use. The program is similar in philosophy to the *STRAT* program (Zhang & Matthews, 1993), which also uses a numerical approach to solve the data-collection strategy problem. The *STRAT* program has been written as a separate auxiliary program for a Xuong–Hamlin area-detector system; we have incorporated our procedure as a subroutine in the *MADNES* program (Messerschmidt & Pflugrath, 1987). This allows for an automatic transfer of crystal-orientation and detector parameters to the procedure, thus limiting errors in user input. In addition, it is easy to take account of the geometrical limitations of the goniostat and to avoid regions where collisions might occur. User input is interactive using the *MADNES* user interface.

Other procedures to aid in efficient data collection have been published. Xuong, Nielsen, Hamlin & Anderson (1985) proposed a complicated two-dimensional data-collection scheme for the Xuong–Hamlin detector (Hamlin, 1985). *RSPACE*

Table 4. *Total completeness of the data as a function of the scan range in the second setting*

The reference zero position is setting 1 from Table 3. The starting scan angle was -30° in all cases. The completeness is the combined completenesses of the first and second settings.

Scan range ($^\circ$)	70	65	60	55
Completeness (%)	99.63	99.63	99.60	99.43
Scan range ($^\circ$)	50	45	40	35
Completeness (%)	98.83	97.07	94.73	92.48

(Harris, Fitzgibbon & Hage, 1989) and *LATTICEPATCH* (Klinger & Kretsinger, 1989) are computer graphics programs for visualizing the relationship between reciprocal space and data-collection hardware, thereby helping crystallographers to collect their data efficiently. These procedures are used separately from the data-collection software, in contrast to our *DCS* routine.

The program and manual are available from the authors without cost by electronic mail. Address requests to bauke@rugch2.chem.rug.nl.

This work was supported in part by an EMBO short-term fellowship to IV. We thank Professor Wim Hol for stimulating discussions.

References

- ARNDT, U. W. & THOMAS, D. J. (1985). *Crystallographic Computing 3: Data Collection, Structure Determination, Proteins, and Databases*, edited by G. M. SHELDON, C. KRÜGER & R. GODDARD, pp. 43–51. Oxford: Clarendon Press.
- HAMLIN, R. (1985). *Methods Enzymol.* **114**, 416–452.
- HARRIS, M. R., FITZGIBBON, M. & HAGE, F. (1989). *J. Appl. Cryst.* **22**, 624–627.
- KLINGER, A. L. & KRETSINGER, R. H. (1989). *J. Appl. Cryst.* **22**, 287–293.
- KOOYSTR, P. J. U., KALK, K. H. & HOL, W. G. J. (1988). *Eur. J. Biochem.* **177**, 345–349.
- MESSERSCHMIDT, A. & PFLUGRATH, J. W. (1987). *J. Appl. Cryst.* **20**, 306–315.
- XUONG, N. G., NIELSEN, C., HAMLIN, R. & ANDERSON, D. (1985). *J. Appl. Cryst.* **18**, 342–350.
- ZHANG, X.-J. & MATTHEWS, B. W. (1993). *J. Appl. Cryst.* **26**, 457–462.