Scaling of one-shot oscillation images with a reference data set

Kunio Hirata,[a] Eiki Yamashita,[a] Hiroshi Aoyama,[b] Kazumasa Muramoto,[c] Shinya Yoshikawa[c] and Tomitake Tsukihara[a]*

[a]Institute for Protein Research, Osaka University, Yamada-oka Suita, Osaka 565-0871, Japan, [b]Riken Harima Institute, Mikazuki Sayo, Hyogo 679-5148, Japan, and [c]Department of Life Science, Himeji Institute of Technology, Kamigohri Akoh, Hyogo 678-1297, Japan. E-mail: tsuki@protein.osaka-u.ac.jp

Combining a least-squares procedure with the program *MOSFLM*, we developed a program *SCLONE*, which processes diffraction images that do not contain serial oscillation images and may have a few or no full reflections. After each image was processed by *MOSFLM*, the partialities and structure amplitudes of the reflections were estimated using a least squares method to refine the scaling factor, the relative temperature factor, the mosaic spread, cell constants, and missetting angles for each independent image. The *SCLONE* calculation significantly improved the quality of the intensities from the reflections obtained by the initial *MOSFLM* calculation and crystal structural refinement confirmed the improvement. The *SCLONE* calculation indicated that the reflection of the present crystal had a rocking curve that was steeper at the middle of profile and more gradual at both ends of profile than that assumed in the program *MOSFLM*.

Keywords: X-ray diffraction; intensity data processing; scaling; partial reflection; mosaic spread.

1. Introduction

Irradiating with X-rays can induce structural changes to a crystalline state, such as cytochrome *c* peroxidase (Berglund *et al*., 2002). In our X-ray experiments for cytochrome *c* oxidase from a bovine heart at SPring-8, BL44XU, the visible absorption spectra were changed by X-ray irradiation. In spite of the absorption spectrum changes, the crystal preserved isomorphism, but exposure may have caused a local structural change. To prevent damage, the X-ray diffraction data must be collected using a short exposure time as possible. Each shot should be performed at a fresh position in the crystal and the oscillation range must be restricted to a small angle for intensities with a high signal-to-noise ratio. The diffraction data set will not contain serial oscillation images and each image may have a few or no full reflections. Schutt & Winkler (1977), Winkler, Schutt & Harrison (1979) and Rossmann *et al*. (1979) have developed programs for crystals with large unit cells that correct partially recorded intensities to their fully recorded equivalents. However these programs require that the number of partial reflections having fully recorded counterparts is sufficiently great. Analyzing the quantitatively reflecting range from various models, Greenhough and Helliwell (1982) have discussed estimation of the partiality of the reflection intensity. Several programs that process oscillation images are available worldwide. Among these, *MOSFLM* (Leslie, 1992) is a friendly program for both users and program developers. *MOSFLM* can process images recorded from mostly partial reflections. The cell constants, crystal orientation, and mosaic spread of each image are determined by an auto-indexing procedure and a subsequent procedure of "estimate mosaic spread" in *MOSFLM*. The parameters are used to estimate the integrated intensity and partiality of each reflection. However, the quality of the partiality obtained through these procedures is not enough to create a suitable intensity data set in our method for an accurate structural analysis. Thus, we combined a least-squares procedure with *MOSFLM* to develop a program, which refines the partiality of each reflection and scales the independently acquired oscillation images.

2. Methods

2.1. The program *SCLONE*

SCLONE was developed to refine the parameters that correct the observed intensity by using reflection files created by *MOSFLM*. *SCLONE* refines the scale factor, relative temperature factor, cell constants, missetting angles and mosaic spread of each independent image against a reference data set. These refined parameters are then used to determine the partiality of each reflection in *MOSFLM*. Whole procedures are schematically shown in Fig. 1. The corrected intensity is derived by the Equation (1) in the *SCLONE*.

$$I_{corr}(h) = \frac{K \exp\{2B[\sin\theta(h)]/\lambda\}}{P(h)} \text{ ---------- (1).}$$

SCLONE minimizes

$$R2 = \sum_h \left[I_{ref}(h) - I_{corr}(h)\right]^2 \text{ ---------- (2).}$$

In these equations, $I_{ref}(h)$ is the reference intensity of reflection h, K and B are a relative scale factor and a relative temperature factor, respectively, for each image against a reference data set. A unique set of $I_{ref}(h)$ is obtained from the crystal isomorphous with the target crystals. Also, $I_{obs}(h)$ and $P(h)$ are the observed intensity and the estimated partiality of reflection h in the image, respectively. $P(h)$ is a function of the cell constants, the crystal orientation with missetting angles and the mosaic spread in *MOSFLM*. The observation equation (4) is

$$I_{ref}(h) - I_{corr}(h) = \sum_i \Delta p_i \left(\frac{\partial I_{corr}(h)}{\partial p_i}\right) \text{ - - - - - - - (4)}$$

where p_i represents the scale factor K, the relative temperature factor B, the mosaic spread m, cell constants a, b, c, α, β and γ, and missetting angles δy and δz defined in *MOSFLM*. The derivatives for K and B are as described in Equations (5) and (6),

$$\frac{\partial I_{corr}}{\partial K} = \frac{\exp\left\{2B\left[\sin\theta(h)/\lambda\right]^2\right\} I_{obs}(h)}{P(h)} \text{ ---------- (5)}$$

$$\frac{\partial I_{corr}}{\partial B} = \frac{2\left[\sin\theta(h)/\lambda\right]^2 K \exp\left\{2B\left[\sin\theta(h)/\lambda\right]^2\right\} I_{obs}(h)}{P(h)} \text{ - - (6)}$$

The derivatives for m, a, b, c, α, β, γ, δy and δz are estimated from numerical calculation, where $I_{obs}(h)$ and $P(h)$ are in the reflection files from the *MOSFLM* calculation. The derivative in term of p_i is approximated by Equation (7),

$$\frac{\partial I_{corr}(h,p_1,p_2,..,p_i,..)}{\partial p_i} = \frac{\left[I_{corr}(h,p_1,p_2,..,p_i+\delta p_i,..) - I_{corr}(h,p_1,p_2,..,p_i,..)\right]}{\delta p_i} \text{ ---(7).}$$

In this estimation, δp_i is set at one tenth of the estimated standard deviation of the target parameter of p_i. The appropriate damping factors are applied to the parameter shifts to avoid divergence. An R-factor is calculated by Equation (8),

$$R = \frac{\sum_h \sum_i \left| I_{ref}(h) - I_{corr}(h)_i \right|}{\sum_h \sum_i I_{ref}(h)} \text{----------(8)},$$

and the number of reflections obtained is output at each cycle of refinement.

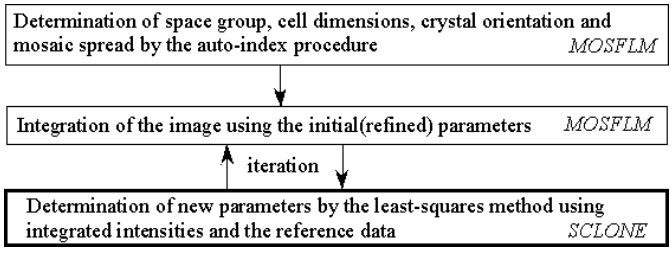

Figure 1

A flow-chart of the *SCLONE* refinement. Thin-line blocks and a heavy-line block represent programs of *MOSFLM* and *SCLONE*. *MOSFLM* evaluates the intensity of each reflection. *SCLONE* refines the scale factor, the B-factor, the cell dimensions, the missetting angles, and the mosaic spread and then, *MOSFLM* evaluates the intensity by using the refined parameters.

2.2. Test data

A crystal of oxidized cytochrome *c* oxidase from bovine heart with dimensions of 0.5 × 0.5 × 0.2 mm^3 was used to acquire an intensity data set. The oscillation angle and exposure time of each shot were 0.5 degrees and five seconds, respectively. The wavelength of the X-rays was 0.90 Å. A total of 200 oscillation images were recorded on a CCD detector, PX210, at BL44XU in the SPring-8. An auto-indexing procedure in *MOSFLM* determined the cell constants and orientation parameters for each image and then, the mosaic spread was optimized. The cell constants were refined and fixed in subsequent *MOSFLM* procedures. The observed intensity and the partiality of each reflection were evaluated by an integration procedure that refined the missetting angles and the mosaic spread of each image. Whole of these procedures was referred to the serial *MOSFLM* calculation. The intensity data was merged and evaluated by the program *SCALA* (Evans, 1993). The crystal data and the statistics of the intensity data were as follows: space group, $P2_12_12_1$; cell dimensions, a=184.01, b=207.45 and c=178.13 Å; resolution, 2.5 Å; number of observed reflections, 920,148; number of independent reflections, 233,485; completeness, 99.6 %; $R_{\rm merge} = \Sigma_h \Sigma_i \, | \, |F(h)_i|^2 - \langle |F(h)|^2 \rangle | \, / \, \Sigma_h \Sigma_i \, |F(h)_i|^2$, 0.110; and an averaged mosaic spread, 0.277°. The averaged intensities of the independent reflections were used as the reference intensities in subsequent *SCLONE* refinement.

3. Results

3.1. Evaluation of the *SCLONE* by using test data

Before the *SCLONE* refinement, the 200 images were re-processed by *MOSFLM* assuming each image was independent of the other images. The cell constants and orientation parameters of each image were determined by an auto-indexing procedure and then, the mosaic spread was optimized. By fixing these parameters the observed intensity of each reflection was evaluated together with its partiality, which is referred to as the initial *MOSFLM* calculation. All the missetting angles of the orientation parameters were zero at this stage.

The *SCLONE* calculation was performed for each of the 200 images. The relative scale factor, the relative temperature factor, the mosaic spread, cell constants and missetting angles for each image were refined to evaluate $I_{\rm corr}$ by the *SCLONE* refinement. The initial parameters of the *SCLONE* refinement were the same as the initial *MOSFLM* calculation. During the initial five cycles of the *SCLONE* refinement, only the scale factor and the temperature factor of each image were refined. After the scale factor and the temperature factor converged, all the other parameters were refined. The R-factors given by Equation (8) were calculated for the respective intensity data obtained by the *SCLONE* calculation and by the initial *MOSFLM* calculation. The common reflections between the two data sets with $P(h)$>0.3 were selected to estimate the R-factors. Except for one image, the R-factors for the *SCLONE* and the initial *MOSFLM* calculations ranged between 0.095 and 0.180 and between 0.122 and 0.282, respectively. The averaged R-factors of 200 images were 0.125 for the *SCLONE*, and 0.161 for the initial *MOSFLM*. Of the 200 images, 198 had smaller R-factors in the *SCLONE* refinements than in the initial *MOSFLM* calculation. The initial *MOSFLM* processing for seven images yielded high R-factors, which ranged between 0.200 and 0.282, while the *SCLONE* refinement of the same images resulted in R-factors between 0.122 and 0.160. *SCLONE* reduced the average R-factor value from 0.227 (from the initial *MOSFLM* calculation to 0.137. The R-factors were calculated for the common reflections with different ranges of partiality less than 1.0 to evaluate qualities of the intensities and the partialities of each image. For all partiality ranges, the R-factors for the images were less using *SCLONE* than initial *MOSFLM*, and their averaged values were 0.143 for *SCLONE*, and 0.179 for initial *MOSFLM* (Table 1).

The cell constants and mosaic spreads obtained by the *SCLONE*, the initial *MOSFLM* and the serial *MOSFLM* calculations are listed in Table 1. The average values of cell dimensions determined by three different methods were consistent with one another within their e.s.d.s, while the averaged mosaic spread by the *SCLONE* was significantly smaller than those by the initial *MOSFLM* and the serial *MOSFLM* calculations.

Table 1 Comparison of two intensity data obtained by the SCLONE and the initial MOSFLM calculations, and statistics of cell dimensions and mosaic spreads determined by three different calculations

	SCLONE	Initial *MOSFLM*	Serial *MOSFLM*
R-factor*			
Average	0.125	0.161	
Minimum	0.095	0.122	
Maximum	0.400	0.374	
$R_{\rm merge}$[#]	0.143	0.179	
Cell dimensions[$]			
a (Å)	184.26 (0.12)	184.28 (0.08)	184.01
b (Å)	207.63 (0.25)	207.70 (0.13)	207.45
c (Å)	178.25 (0.13)	178.25 (0.09)	178.13
Mosaic spread (°)*	0.17 (0.03)	0.24 (0.04)	0.28 (0.02)

*R-factor was calculated by Equation (8) in the text.

[#] $R_{\rm merge}$s defined in the text were calculated for 539,540 reflections whose partialities were 0.30 ≤P<1.00 in both cases of the *SCLONE* and the initial *MOSFLM* calculations.

[$]The figures except for the cell dimensions for the serial *MOSFLM* were their averaged values of 200 images. The cell dimensions in the serial *MOSFLM* calculations were determined by using three sets of two neighboring images. The Figures in parentheses are their e.s.d.s.

SCALA merged, scaled, and averaged the obtained intensity data from the 200 images by the *SCLONE* calculation. Reflections with $I/\sigma_I>0$ were included in these calculations. Table 2 lists the statistics of intensity data for the *SCLONE* and the serial *MOSFLM* calculations.

3.2. Structural refinement with the structure factor from *SCLONE*

The quality of *SCLONE* was evaluated by comparing the structural refinements using the structure factors obtained from *SCLONE* and serial *MOSFLM* calculation. First, the crystal structure was determined for the serial *MOSFLM* data at 2.5 Å resolution by the molecular replacement method with the reference model previously determined at 1.8 Å resolution (unpublished data). The program *CNS* (Brunger *et al*., 1998) was applied for the molecular replacement analysis and the structural refinement. Default values were used for the restraining parameters except for the bond energy of S^{δ}-C^{ε} of the methionine, which was replaced by 500.0 Kcal mol^{-1} $Å^{-2}$ (Odoko *et al*., 2001). Five percent of total reflections were reserved for a R-free estimation to monitor progress of the refinement. Another structure refinement was performed for the structure factors from the *SCLONE* calculation. Both refinements with the *SCLONE* and the serial *MOSFLM* data sets used the same reflection indices for R-free estimations. Table 3 lists the statistics of the structural refinements. The refined structures with the *SCLONE* data and with the serial *MOSFLM* data had similar quality.

Table 2 Statistics of the $|F(h)|^2/P(h)$ by the SCLONE, $|F(h)|^2$ from the serial *MOSFLM* calculations.

d_{max} - d_{min}	Number of reflections		Completeness	Multiplicity	R_{merge}
	measured	independent	(%)		
SCLONE calculation (P>0.3)					
50.0- 2.50	1,005,186	232,434	99.0	4.3	0.120
SCLONE calculation (1.0>P>0.3)					
50.0- 2.50	574,269	210,113	90.0	2.7	0.121
Serial *MOSFLM*					
50.0- 2.50	920,148	233,485	99.6	3.9	0.110

Table 3 Structure refinements with two data sets from the *SCLONE* and the serial *MOSFLM* calculations.

Data set	Free-R	R	Number of Reflections	r.m.s.d.* bond (Å)	angle (°)
SCLONE	0.260	0.213	232,434	0.013	1.87
Serial *MOSFLM*	0.261	0.207	233,485	0.012	1.85

*r.m.s.d. is root mean square deviation from the ideal value of each parameter.

4. Discussions

The corrected intensities from the *SCLONE* calculation significantly improved R-factor from Equation (8) and R_{merge} compared to the results than those from the initial *MOSFLM* calculation (Table 1). The statistics of observed structure factors and structural refinements indicated that the *SCLONE* calculation was almost equivalent to the serial *MOSFLM* calculation (Tables 2 and 3), which implies that *SCLONE* is available to process independent images that mostly contain partial reflections.

The mosaic spreads determined by *SCLONE* were significantly smaller than those determined by the initial *MOSFLM* and the serial *MOSFLM* calculations. In *MOSFLM* the partiality is given by the function

$$P = (1-\cos\pi q_{\varphi})/2 \qquad (9)$$

where q_{φ} is a function of the mosaic spread (Winkler, Schutt & Harrison, 1979). In the *SCLONE* refinement I_{corr} is evaluated from Equation (1) containing the partiality P, and the minimizing function of the least-squares procedure is given by Equation (2). The *SCLONE* calculation refines the mosaic spread through the function defining the partiality. While the initial *MOSFLM* calculation estimates the mosaic spread by analyzing a cumulative intensity of each image. In order to understand the difference between the two mosaic spreads determined by the *SCLONE* and by the initial *MOSFLM* calculations, the function defining the partiality in *MOSFLM* was replaced by a hypothetical function

$$P = q_{\varphi} - (\sin 2\pi q_{\varphi})/2\pi \qquad (10)$$

which is steeper around q_{φ} = 0.5 than Equation (9), and P = 0, 0.5 and 1.0 at q_{φ} = 0, 0.5, and 1.0, respectively, as Equation (9). When two rocking curves corresponding to Equations (9) and (10) with a same size of mosaic spread are compared to each other, the latter is steeper at the middle of profile and more gradual at both ends of profile than the former. The *SCLONE* refinement and the initial *MOSFLM* calculation were performed for selected 50 images by using the modified program *MOSFLM*. The selected images were 1st-10th, 41st-50th, 81st-90th, 121st-130th and 161st-170th of the test data. The mosaic spread of each image by the *SCLONE* refinement was larger than the original calculation by about 0.04°, while that by the initial *MOSFLM* calculation did not significantly change from the original ones. The R-factors from Equation (8) for the *SCLONE* and the initial *MOSFLM* calculations by using the modified *MOSFLM* program were almost identical to those for the *SCLONE* and the initial *MOSFLM* calculations, respectively, by using the original *MOSFLM*. The *SCLONE* calculations by using two different *MOSFLM* programs resulted in an almost equal partiality of each reflection by adjusting the mosaic spreads, while the initial *MOSFLM* calculation made no significant difference both in the partiality and in the mosaic spread whether the equations (9) or (10) were used in the program *MOSFLM*. Thus the initial *MOSFLM* and the serial *MOSFLM* calculations must result in adequate mosaic spreads close to true one whether the original or the modified *MOSFLM* programs were used. Since the *SCLONE* calculation by using the equation (10) effected in the mosaic spread closer to that of the initial or the serial *MOSFLM* calculations than the *SCLONE* calculation by using the equation (9), the equation (10) must be a better equation relating the mosaic spread and the partiality than the equation (9) for the present crystal. Thus the reflection of the present crystal has a rocking curve which is steeper at the middle of profile and more gradual at both ends of profile than that corresponding to Equation (9) used in the program *MOSFLM*.

This work was supported in part by grants from the New Energy and Industrial Technology Development Organization (NEDO), Japan, and by 21st Century COE Program of Japan to T.T. and Center for Renewable Energy and Sustainable Technology (CREST), Japan to S.Y.

References

Berglund, G.I., Carlson, G.H., Smith, A.T., Szoke, H., Henrikden, A. & Hajdu, J. (2002). The Catalytic Pathway of Horseradish Peroxidase at High Resolution. *Nature*, **417**, 463-468.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., & Warren, G. L. (1998). Crystallography and NMR System: A New Software for Macromolecular Structure Determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-21.

Evans, P.R. (1993). *Proceedings of the CCP4 Study Weekend on Data Collection & Processing*. pp.114-122. Daresbury Laboratory, Warrington: UK.

Greenhough, T. J. & Helliwell, J. R. (1982). Oscillation Camera Data Processing: Reflection Range and Prediction of Partiality. I. Conventional X-ray Sources. *J. Appl. Cryst.* (1982). **15**, 338-351.

Leslie, A.G.W. (1992). *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography*. Vol 26, Daresbury Laboratory, Warrington: UK.

Odoko, M., Yao, M., Yamashita, E., Nakashima, R., Hirata, K., Aoyama, H., Muramoto, K., Shinzawa-Itoh, K., Yoshikawa, S. & Tsukihara, T. (2001). Optimization of the energy constant of the methionine S^{δ}-C^{ε} bond for X-PlOR refinement of protein structure. *J. Appl. Cryst*. **34**, 80-81.

Rossmann, M.G., Leslie, A.G.W., Sherin, A.M. & Tsukihara, T. (1979). Processing and Post-refinement of Oscillation Camera Data. *J. Appl. Cryst.* **12**, 570-581.

Schutt, C. E. & Winkler, F. K. (1977). The Rotation Method in Crystallography, edited by U. W. Arndt & A. J. Wonacott, pp. 173-186. Amsterdam: North-Holland.

Winkler, F. K., Schutt, C. E. & Harrison, S. C. (1979). The Oscillation Method for Crystals with Very Large Unit Cells. *Acta Cryst.* **A35**, 901-911.

Received 30 April 2003 • Accepted 15 October 2003