Development of the X-Ray Diffraction Data Collection System Using Imaging Plate

by

Min Yao

January 1995

Division of Biological Sciences,
Graduate School of Science,
Hokkaido University
Sapporo 060, Japan

Contents

Acknowledgement

I wish to express my gratitude to Professor Isao Tanaka who introduced me to the exciting field of protein crystallographly. All of my academic studies and whole degree dissertation were completed under the guidance of Professor Isao Tanaka. During this period I spent an academically profitable and yet very enjoyable time.

I would like to thank Professor Kunio Hikichi for his apposite comments and valuable advice.

I would also like to thank Professor Katsutoshi Nitta and Professor Mamoro Honma for contributing of protein samples and their valuable advice.

My sincere thanks are also given to all members of this laboratory for their help and encouragement.

I must also thank all my many other friends for their support and help.

Finally, I would like to thank my parents, and family for their understanding and support.

Thank you

Hokkaido University.

January 1995

Chapter 1 Introduction

1-1. Protein structure analysis

The processes of life depend basically on the structure and interactions of molecules in the living material. Thus knowledge of the structure and action of these molecules gives a clue for the understanding of life. Among these molecules (protein, nucleic acids, and carbohydrates), protein is the largest group and enzymes are the most diverse class of proteins. The protein consists of one or more polypeptide chains which constructed from 20 naturally occurring amino acid. The chains are folded into a three-dimensional structure and this structure determines the chemical and physical behavior and resulting function of the protein. To understand cellular process, knowledge of the three-dimensional structure of biological macromolecules is vital. Three techniques are widely used for structural determination of protein at atomic level: X-ray diffraction of crystal, nuclear magnetic resonance (NMR) and electron microscopy. Although NMR does not require crystal and provides more detailed information on the dynamics of a molecule in question, it can be used only for small protein with molecular weight less than about 20,000. The electron microscopy can only be used to study assembly of proteins. Only X-ray diffraction of crystal can dive the complicated three-dimensional structure of small as well as large (molecular weight up to at least 10^6 Da.) protein with a precision of about 0.1~0.2Å. Nowadays, with the computer which has been under a critical development, once a suitable crystal is available, a new structure of a protein or even a virus can be determined in a year or less under favorable circumstances. In the case where there is a closely related structure already known, then a new crystal structure may be obtained in as little as 1~2 week. However, we know about 50 times the number of protein sequences as three-dimensional protein structures (Bowie & Luthy et al.,1990). This

unbalanced development hinders progresses in many areas of biochemistry because a protein sequence has little meaning outside the context of its three-dimensional structure. Therefore, speedy analysis of X-ray protein crystal structure is necessary. Various problems (e.g. accurate data collection, search for best derivative, phase problem etc.) to face during structure determination have to be solved.

In general, the stages of determining protein X-ray crystal structure are :

crystallization of the pure sample;

determination of crystal unit cell parameters;

X-ray data collection to a given resolution;

solution of the crystallographic phase problem;

interpretation of the electron density map;

refinement of the molecular model against the observed data;

Since sparse matrix sampling of crystallization of protein has been published and even crystallization kit is commercially available, the probability of success of crystallization is increased. The high quality diffraction data is the fundamental base for determining protein structure. The crystals of protein have major effect on the properties of X-ray diffraction data which has following characteristics:

(1) The scattering intensity is relatively weak because of the atomic numbers of the elements constituting the crystal and the large fraction of solvent.

(2) The resolution is low because of the extensive flexibility and disorder in the crystal.

(3) The diffraction patterns are dense because unit cell is large.

(4) The crystals are radiation sensitive. Once started, the radiation damage continues even without further radiation.

All of the above properties demand rapid collection of X-ray diffraction data of the protein crystal. So, rapid automatic data collection system using sensitive

area-detector is necessity. For this reason, we have developed several automatic diffractometer systems for laboratory use using imaging plate and software systems (ELMS, WLEMS) for data processing in collaboration with MAC Science Co., Ltd.. Those diffractometers are: rotation camera type DIP-100, Weissenberg camera type IPD-WAS (developed at the Institute of Physical and Chemical Research) and Weissenberg camera type DIP-R300 with variable film-crystal distance. The software systems developed have following characteristics:

(1). automatic indexing function for speedy data collection.

This was developed for finding the orientation of crystal from reciprocal-space lattice vectors using double reflections method, under the condition that approximate unit-cell parameters are known. When reflection data are collected, determination of crystal orientation is necessary. Since crystals are radiation sensitive, rapid determination of crystal orientation is required. Using automatic indexing, the time of finding the orientation has been greatly reduced (section 3-2-1).

(2). eigenvalue filtering method for the parameter refinement.

In integrating intensity of collected reflection data, accurate parameters obtained by least-squares refinement are necessary. Because of correlation between parameters, this refinement often converges slowly, or even can not converge, or gives inaccurate-parameters. The eigenvalue filtering permits automatic identification and determination of those combinations of parameters to which residual is most sensitive. Here, we applied eigenvalue filtering to refine the camera-setting, unit-cell and crystal-orientation parameters (section 3-2-2).

(3). integration of reflections which are splitted due to Kα1 and Kα2.

Kα radiation from laboratory source consists of slightly different wavelength components Kα1 and Kα2. Such X-ray causes split in reflection spot. Ordinary, split can be ignored. But for high resolution reflection data, in the Weissenberg

photograph particularly, split of reflection spots recorded on the detector can not be ignored. For the integration of ELMS and WELMS systems this effect was considered using variable mask and improved profile-fitting (section 3-2-3).
(4). automatic alignment of crystal system using κ-goniometer with the help of computer graphic.

The DIP-R300 is equipped with κ-goniometer to orient crystals. But its movement is complicated and difficult to understand. So that, we use 3-D graphics software to align crystal and simulate its actual movement (section 3-2-4).

Now these diffractometers are in practical use in many laboratories.

In parallel to the development, I analyzed three protein crystal structures: pigeon egg-white lysozyme (section 4-1), dog milk lysozyme (section 4-2) and ACC deaminase (section 4-3). The structures of pigeon egg-white lysozyme and dog milk lysozyme were determined by a molecular replacement method using 3Å data collected by DIP-100 with ELMS software package. Crystallographic parameters of ACC deaminase were determined using low resolution reflections collected by DIP-100 with the ELMS software package. The structure of pigeon egg-white lysozyme was refined to R value of 20.3% at 2Å.

1-2. X-ray diffractometer in laboratory

There are various types of X-ray diffractometers suitable for laboratory use. In terms of goniometer on which crystal is mounted, χ-circle, κ-goniometer and fixed spindle axis type are available. In terms of camera type, precession, Weissenberg and rotation (oscillation) cameras are available. Moreover there are several different detectors.

Scintillation counter (zero-dimensional detector) has been used since the early days of X-ray diffraction. It gives very accurate result, but since it measures X-ray reflections sequentially it takes several weeks to collect a complete data set from protein crystal, so that it is being replaced by the much faster two-dimensional detector.

As two-dimensional detectors there are three kinds; X-ray film, imaging plate and electronic detector. The X-ray film is a classical detector for X-ray radiation. Although its spatial resolution is high, it is not used much anymore because of the availability of far more sensitive imaging plates and electronic area detectors. The imaging plates are used in the same manner as film, but have several advantages: the high sensitivity, the reusability and wide detection area with neither spatial distortion nor non-uniformity of response. They are very effective to develop an automatic diffractometer of camera type. The detailed introduction of the imaging plates will be descried in chapter 2. The electronic area detectors have three types: gas-filled, TV system and charge coupled device (CCD). Among those electronic area detectors, favorable CCD has attracted attention. CCD area detectors usually come with 512 x 512 pixels each of ~10μm x 10μm or up to 27μm x 27μm, and they have a high dynamic range, combined with excellent spatial resolution, low noise, and high maximum count rate. Since CCDs carry out the energy of the incident photon by electric charge on pixel, so they do not need reading-erasing time and can be used to real-time automatic diffractometer. However an optimal

design for X-ray work is not yet commercially available, and to make CDDs of the relatively large size and high sensitivity is necessary.

Reference

Bowie, J. U., Luthy, R. and Eisenberg, D. (1990). *Science* **253** 164-170

Chapter 2 A summary of developed diffractometer systems

In this chapter, a brief introduction of the diffractometer systems developed will be described.

2-1. Imaging plate area detector

The imaging plate (IP) is a plastic plate coated with fine photostimulable phosphor crystals (BaFBr:Eu2+) as X-ray sensitive medium (Amemiya & Miyahara, 1988). The phosphor crystal is able to store a fraction of energy of incident X-rays and release the stored energy as a luminescence when stimulated by a He-Ne laser light. The pixel size depends mainly on the reading system and is between 100μm x 100μm and 200μm x 200μm. A typical size of an IP is 200 x 400 mm.

Imaging plates are at least 10 times more sensitive than X-ray film and their dynamic range is very wide ($\geq 10^5$). The entire range from strong to weak reflections can therefore be collected with one exposure on a single plate. The plates can be erased by visible light and used repeatedly. X-ray diffraction data collection systems using IP as an area detector are now widely used in protein crystallography (Tanaka & Yao et al., 1990; Sakabe, 1991; Sato & Yamamoto et al., 1992). But imaging plates are similar to photographic film in the sense that they require a multi-step process: exposure as the first step and reading-erasing as the second step. Reading-erasing takes a few minutes (for example, about 8 minutes to read largest IP 520 x 440 mm). So IP can not be used as a real-time data collection system.

2-2. DIP-100 automatic diffraction data collection system

DIP-100 is an automatic diffraction data collection system for protein crystallography, that utilizes an imaging plate(IP) as the X-ray detector and a rotating-anode generator as the X-ray source (Tanaka & Yao, 1990). It was designed to work in a similar way to the conventional rotation camera model, but with full on-line mode. The view of DIP-100 hardware configuration with a schematic drawing is shown in Fig. 2-2.1, where A,B,C and D indicate the X-ray optical system, the goniometer, the rotational imaging plate disc, the reading and erasing system. The X-ray optical system consists of a rotating anode X-ray generator, a graphite monochromator and double-hole collimator. Two rotating anode (Mo, Cu) are provided as the X-ray source, so that we can choose more appropriate wavelength to the experiment. The reading system contains transferable reading arm with erasing lamp (erase any residual image on the IP by the illumination of the intense visible light), photo multiplier, amplifier and a He-Ne laser that is used in the excitation of the latent image and two photomultipliers. The IP, the reading head and photo multipliers are enclosed in a box to shield them from natural light.

DIP-100 adopts the normal-beam-with-flat-film geometry. A single disc-shaped IP with a radius of 100 mm is held perpendicular to the X-ray beam. The crystal spindle axis is in the horizontal plane and perpendicular to the X-ray beam. The crystal-plate distance is adjustable from 75 to 300 mm. Rotation or stationary patterns may be taken with this system. After an exposure to the X-ray beam, the IP rotates with an angular velocity of 390 rev min^{-1} whilst a reading head scans across the plate to measure the stimulated luminescence in a record-player-like manner. So the advantage of readout system is that readout mechanism becomes relatively simple. However a consequence of this mechanical simplicity is that a

correction for the reading data is required. In the reading step, the IP rotates and photons are counted at fixed time intervals. Each count in the (r, θ) polar coordinate system is added to the nearest four (X, Y) grid points on a Cartesian coordinate system. The Cartesian grid size is 125μm x 125μm with pixels in both X and Y. The number of contributors to each Cartesian point varies locally, and therefore the data require a local sampling number correction. Fig. 2-2.2 shows the correction curve for radius dependence. The intensity on IP, which was exposed uniformly to an Fe source mounted on the readout arm, was measured. The experimental curve was smoothed and used to correct for the r dependence of the intensity. The abrupt drop at the film center is the effect of the pre-erasing by laser beam at high sampling point density, and gives rise to an unusable area at the center of the IP of about 5 mm radius. In practice the correction factor for both sampling number and r dependence is calculated once for all grid points and used for data correction on pixel-by-pixel basis.

The DIP-100 is driven directly by a VME crate controlled by a Motorola 68000 processor, which is under the control of an NEC PC9801 personal computer (with 20 MB hard disk) through RS232C interface. The personal computer is connected to Sun SPARC-2 workstation by Ethernet protocol, and reflection data obtained by DIP-100 were transported to workstation. The data processing software ELMS has been developed for DIP-100 on the workstation. The detailed explanation of the software will be described in chapter 3.

The DIP-100 system has been proven to work successfully for several protein crystals: the Ca-binding pigeon egg-white lysozyme, the Ca-binding dog milk lysozyme, the ACC deaminase. The application of the system to three enzymes will be described in chapter 3.

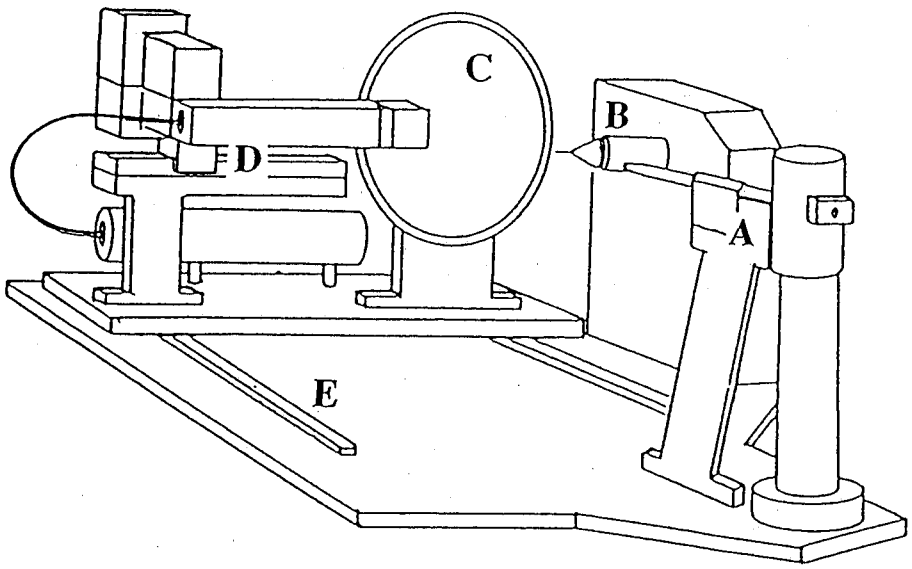


Fig. 2-2.1. A perspective view of the DIP-100 system with a schematic drawing showing A: graphite monochromator and double-hole collimator; B: crystal rotation axis and goniometer; C: imaging plate disc and servo motor; D: reading arm with erasing lamp, He-Ne laser and slide mechanism of reading arm; E: slide mechanism of detector unit.

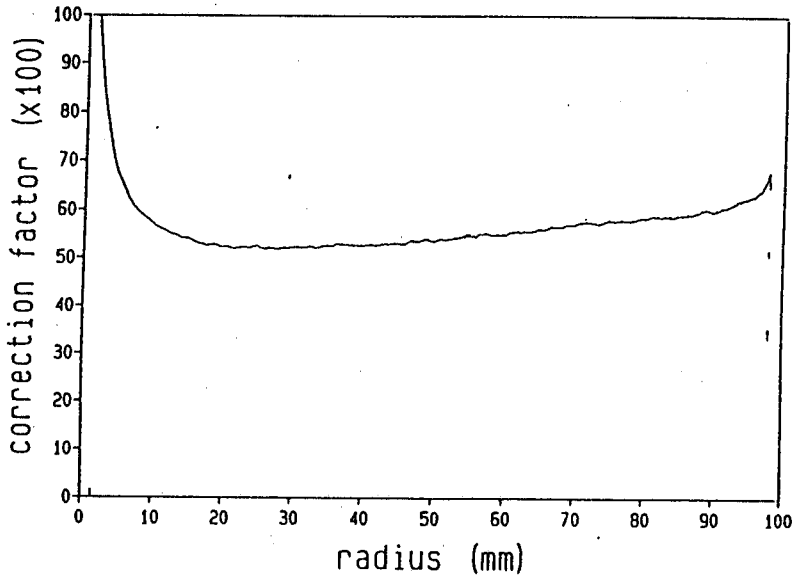


Fig. 2-2.2 The intensity correction curve for radius dependence. This curve was obtained by exposing IP to ^{55}Fe mounted on the laser arm, and by measuring the resulting image. The smoothed curve of the experimental correction curve was used for the data correction.

2-3. IPD-WAS automatic diffraction data collection system

In order to track the so-called "crystalline state reaction" pathway in the laboratory with a time resolution of about one hour, a new imaging plate diffractometer of Weissenberg camera type with adjustable multi-layer-line screen system (IPD-WAS) has been developed (Kamiya & Iwasaki, 1994). The IPD-WAS consists of five parts: the Weissenberg chamber including the X-ray optical system, the adjustable multi-layer-line screen system, the rotary mechanism for IP readout, the control system, and the software system. IPD-WAS was designed for the time-resolved X-ray crystallography of small molecules, with following characteristics:

(1). The adjustable multi-layer-line screen to decrease the background level.
(2). The two imaging plates to eliminate a time waste caused by a IP readout (total readout time is about 10 min. for one IP).

The control system of IPD-WAS contains main-computer (workstation Sun SPARC 1+), a electronic controller and sub-computer NEC PC9801 that were controlled by main-computer. The electronic controller controls movement of IPD-WAS by command sent from main-computer through RS232C interface. The sub-computer is used for displaying photograph patterns and as an output unit of reflection data. The software system consists of two parts and was developed on the workstation Sun SPARC 1+. One is a program package for the full-automatic data collection to control mainly the IPD-WAS hardware and it was written by MAC Science Co., Ltd.. We developed another program package for the data acquisition from the photographs, which is described in the chapter 3.

To collect the reflection data on the Weissenberg camera with the multi-layer-

line screen system, the axis alignment of sample crystal is required prior to the Weissenberg photographing. The axis alignment is based on one rotation photograph and one Weissenberg photograph. The method was explained in detail at section 3-2-1.

Since high-resolution data collection is required, $K_{\alpha 1}$ and $K_{\alpha 2}$ radiation can not be ignored. Then, we developed integration program for split spots. The method will be described in detail at section 3-2-3.

2-4. *DIP-R300 automatic data collection system*

When X-ray diffraction data from single crystal is collected, it is often necessary to bring the crystal axes into a standard orientation relative to the diffractometer axes; e.g. collecting the reflection data on the Weissenberg camera, collecting the reflections of the same area in order to search for heavy atom derivatives, or simultaneous recording of the reflections of the Friedel pair. On the DIP-100 and IPD-WAS systems, the crystal rotates about the goniometer-head axis. In these cases, the orientation of crystal was adjusted only by goniometer-head in which the largest adjustable angle is 15~20°. Therefore we have designed a new Weissenberg camera type imaging plate diffractometer DIP-R300, which includes the κ-goniometer mechanism. Film-crystal distance is also adjustable. The X-ray optical system is double mirrors for focusing. Readout is similar to IPD-WAS diffractometer, in which the pattern recorded on the cylindrical IP is read by a rotary mechanism. The DIP-R300 has four characteristics:

(1). using κ-goniometer

On the DIP-R300, the κ-goniometer is equipped, which rotates crystal in three axes while keeping the crystal at the center of the diffractometer. The κ-goniometer consists of three parts: ω block with a ω rotation axis, κ block with a κ rotation axis and ϕ block with a ϕ rotation axis (Fig. 2-4.1). The goniometer-head is mounted on the ϕ block which is supported by the κ block. The κ block can be rotated about the κ-axis which is carried by ω block. The ω block can be rotated about ω-axis, which is carried by the base of the diffractometer. The α is angle between κ-axis and ω-axis. Table 2-4.1 shows the relationship between the inclination angle of the crystal and alignment angle κ for several different α values. DIP-R300 was designed to α = 55°. Although κ-goniometer can

theoretically allow any orientation of crystal, the interference between κ-goniometer and collimator, or between crystal and κ-goniometer provides a blind region in the DIP-R300 system as shown at Fig. 2-4.2. The method of the axis alignment of the crystal by κ-goniometer is described in detail in section 3-2-4.

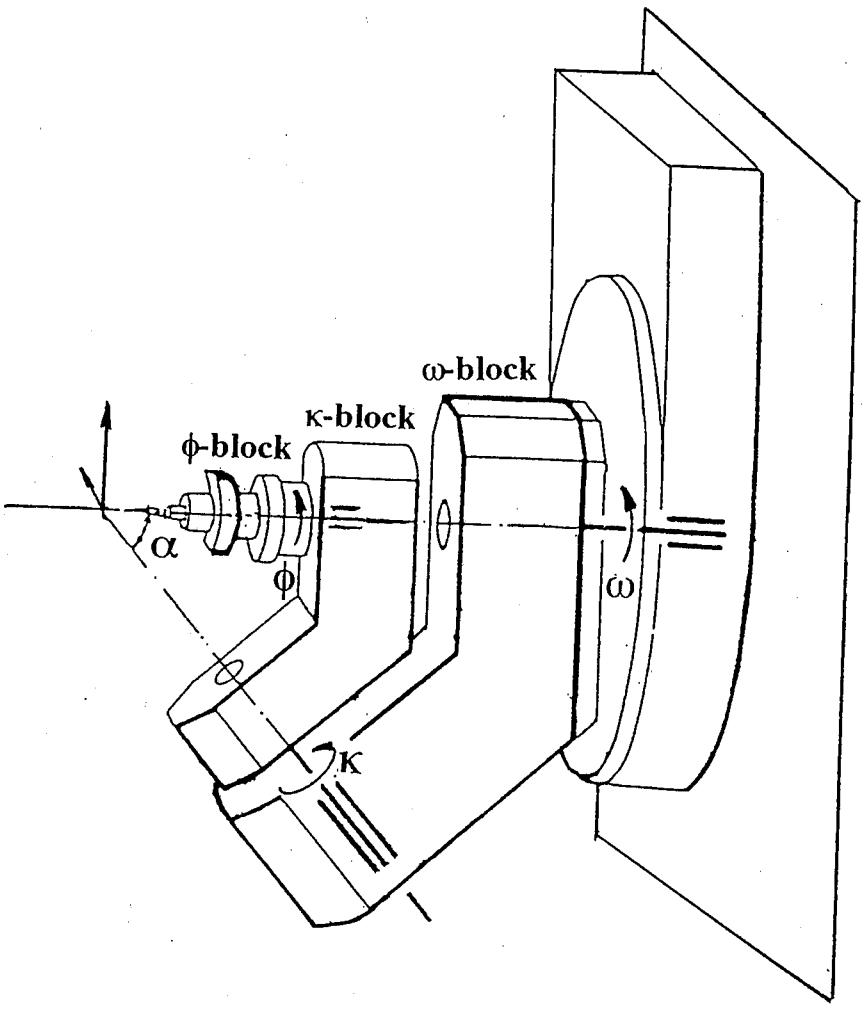


Fig. 2-4.1 The κ-goniometer of the DIP-R300.The figure shows all the angles ω, κ and ϕ set to 0°; α is fixed to 55°.

Table 2-4.1. The relation between the crystal inclination angle χ and alignment angle κ with different α values

χ(°)	$\alpha = 45^\circ$	$\alpha = 50^\circ$	$\alpha = 55^\circ$
		κ (°)	
10	14.16	13.07	**12.22**
20	28.43	26.20	**24.48**
30	42.94	39.49	**36.84**
40	57.85	53.04	**49.36**
45	65.53	59.94	**55.70**
50	73.41	66.97	62.12
60	90.00	81.49	75.24
70	108.42	96.97	88.89
80	130.75	114.09	103.39
90	180.00	34.76	119.36

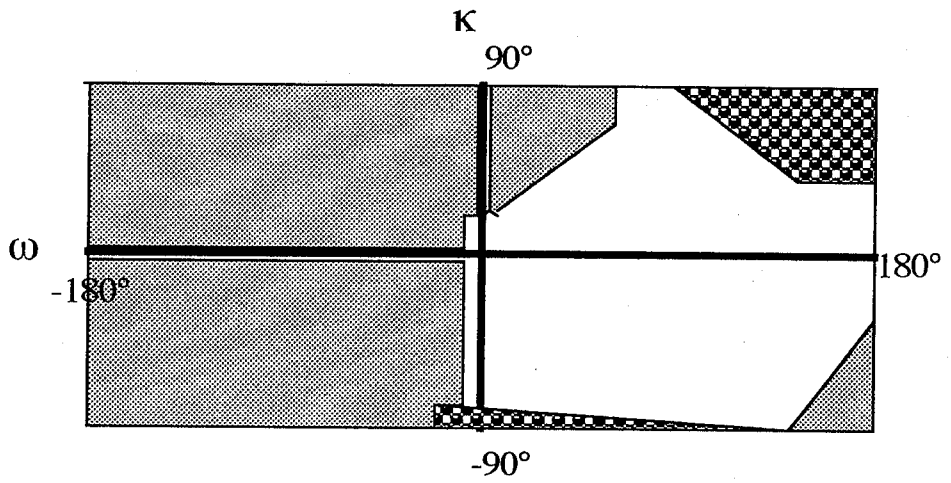


Fig. 2-4.2 The blind region of the DIP-R300 system.

: The blind region caused by κ-block of the DIP-R300.

: The region of crash between κ-goniometer and other part of the DIP-R300.

(2). variable distance between crystal and imaging plate.

The original Weissenberg camera has a cylindrical cassette. The axis of cylinder is coincident with the crystal rotation axis. The X-ray beam is normal to the center axis of the cylinder. The crystal is mounted on the point of intersect between the X-ray beam and the center axis of the cylinder. So the distance between crystal and detector (IP) is fixed to the radius of cylinder. The Weissenberg camera has some advantage over flat oscillation camera: (a) it can record high resolution reflections; (b) it can reduce the incidence of obliquity; (c) it can avoid overlapping of reflections on the detector. However, it is a fault that it has a fixed crystal-detector distance. In the DIP-R300 diffractometer, a cylindrical cassette is on a rail, so that variable crystal-detector distance (from 150 to 300 mm) is achieved. (Fig. 2-4.3). The adjustment of the distance is controlled by computer.

(3). large detector

The DIP-R300 has a very wide detection area (520mm x 440mm) achieved by using two commercially available imaging plates, which occupy about two-seventh of the full cylinder surface (Fig. 2-4.4). The radius of cylinder is 300 mm. Consequentially, the collection of high resolution data will be possible as shown in Table. 2-4.2.

Table 2-4.2. Resolution limit of DIP-R300 (λ=1.54Å)
in radial and axial direction of IP.

Distance(mm)	Radial (Å)	Axial(Å)
150	1.21	1.65
200	1.39	1.91
300	1.84	2.48

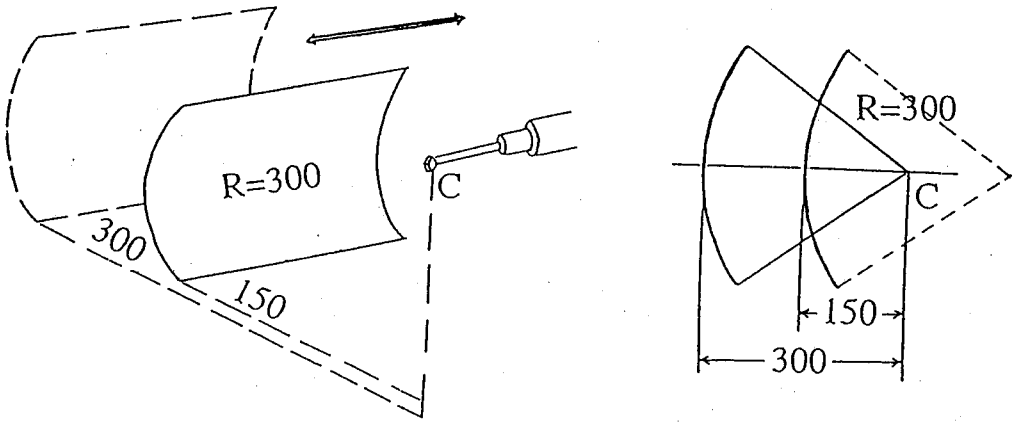


Fig. 2-4.3 The Weissenberg camera of variable distance between crystal and detector.

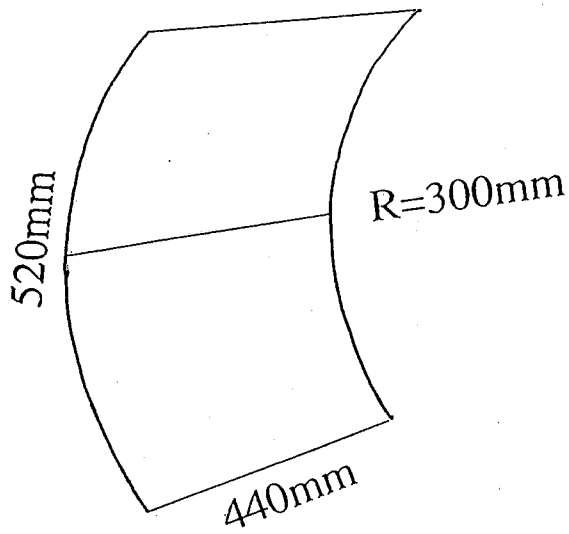


Fig. 2-4.4 The detector of DIP-R300 diffractometer.

(4). variable reading area on the detector

In the DIP-R300 diffractometer, the large imaging plate is used. The pixel size is designed as 100μm x 100μm. Readout time for one full IP is about 8 min. (number of pixel is 5200 x 4400, time for reading one pixel is 6 μsec, so total reading time is 5200 x 4400 x 3.5 x 6 μsec.). The IP data file size is 45MB. It may

be often the case that smaller IP is enough especially for small protein crystals. In order to eliminate a time and memory wasted by reading full IP size, the DIP-R300 detector was designed to have variable reading area. The size of active reading IP is controlled freely by computer.

The control system of DIP-R300 consists of a main computer IRIS workstation and an electronic controller. The IRIS workstation controls all of the DIP-R300 diffractometer mechanisms and integrates intensity for each reflection. The controller which controls camera movement, readout system and κ-goniometer, is under the control of the IRIS workstation through a network system with the Ethernet protocol. The software system of DIP-R300 was developed on the IRIS workstation and consists of two main programs XPRESS and WELMS. The XPRESS was designed to control DIP-R300 hardware for data collection by MAC Science Co., Ltd.. The WELMS is a program for reflection data processing which includes function to process IP with variable size, and a full automatic crystal alignment system using X11 library and IRIS graphics system (gl-library) with κ-goniometer (see section 3-2-4).

References

Amemiya, Y. & Miyahara, J. (1988). *Nature,* **336** 89-90.

Sakabe, N. (1991). *Nucl. Instrum. Methods,* **A303**, 448-463.

Sato, M., Yamamota, M., Imada, K., Katsube, Y., Tanaka, N. & Higashi, T. (1992). *J. Appl. Cryst.* **25**, 348-357

Tanaka, I., Yao, M., Suzuki, M., Hikichi, K., Matsumoto, T., Kozasa, M. & Katayama, C. (1990). *J. Appl. Cryst.* **23**, 334-339

Chapter 3 Development

In this chapter, the developed algorithm and programming of four methods which have been included in ELMS and WELMS software system will be described after brief description of whole software system and coordinate systems.

3-1. Whole software system

3-1-1. Coordinate system

Coordinate system of DIP-100

(1). The reciprocal-lattice coordinates (X,Y,Z) are described in an Cartesian laboratory coordinate system as shown in Fig. 3-1-1.1a. The origin is on the intersection of the rotation axis with the primary X-ray beam. X is parallel to the primary X-ray beam. Z is parallel to the rotation axis and Y is chosen such that it is perpendicular to both axes and form the right-handed Cartesian system. The ideal IP plate is at a distance D from the crystal and normal to the X axis.

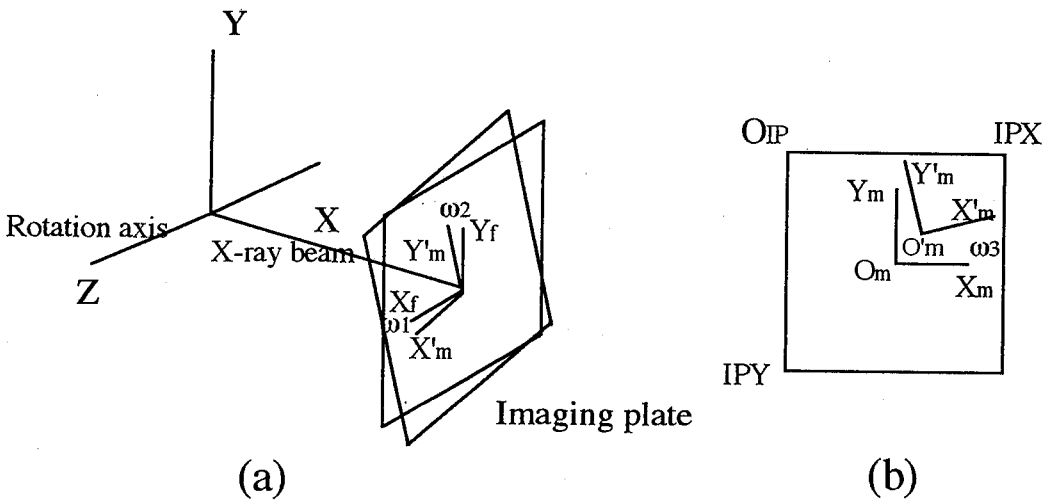


Fig. 3-1-1.1 Definition of the coordinate of DIP-100.

(2). The ideal detector coordinate system is defined on an ideal IP plane which is normal to the X axis. There are two coordinate systems(Fig. 3-1-1.1b). One is the pixel coordinates IPX, IPY where the optical density was recorded with its origin up-left corner of the IP. The intensity is first read onto the (r, θ) polar coordinate system and transferred to IP pixel coordinate by DIP-100 readout system. This coordinate system was used to display IP plane on graphics and calculate intensities. Another is Cartesian system (X_f, Y_f) where X_f is parallel to Z and Y_f to Y. Since the actual IP plane may be slightly inclined from this ideal plane, parameters ω1 and ω2 are introduced which represent twist and tilt angles of the IP plane respectively. The practical detector coordinate system is designated as (X_m, Y_m) where Xm is the intersection of the X-Y plane with the IP plane and Ym in the intersection of the X-Y plane with the IP plane. The vector from the center of the rotation(O_m) to the direct-beam position(O'_m) is X0,Y0 (Fig. 3-1-1.1b). The angle between X_m and X'_m is ω3. The relations between the ideal IP plane coordinate system (X_f, Y_f) and the IP measurement (X_m, Y_m), and IP pixel coordinate system (IPX, IPY) are

$$
\begin{aligned}
&IPX0 = (PX2+IPX1) / 2\\
&IPY0 = (PY2+IPY1) / 2\\
&X_m=(IPX-IPX0)*\{(IPX2-IPX1)/(XM2-XM1)\}\\
&Y_m=(IPY-IPY0)*\{(IPY2-IPY1)/(YM2-YM1)\}
\end{aligned}
\qquad (3\text{-}1\text{-}1.1)
$$

$$
\begin{aligned}
X_m=&X0+[D*X_f*\cos\omega2*\cos\omega3+D*X_f*\cos\omega1*\cos\theta*\cos\omega3\\
&- D*Y_f*\cos\omega1*\sin\theta*\cos\omega3)\\
&*(Xf*\sin\omega1*\cos\omega2+Y_f*\cos\omega1*\sin\omega2+D*\cos\omega1*\cos\omega2) -1]\\
Y_m=&Y0+[D*Xf*\cos\omega2*\sin\omega3+D*X_f*\cos\omega1*\sin\theta*\cos\omega3\\
&+ D*Y_f*\cos\omega1*\cos\theta*\sin\omega3)\\
&*(X_f*\sin\omega1*\cos\omega2+Y_f*\cos\omega1*\sin\omega2+D*\cos\omega1*\cos\omega2) - 1]
\end{aligned}
\qquad (3\text{-}1\text{-}1.2)
$$

Here $\cos\theta=\sin\omega1\sin\omega2$; IPX2-IPX1, IPY2-IPY1, XM2-XM1, YM2-YM1 are IP plane size in pixel and in mm, respectively.

Any reciprocal-lattice point **h**(h, k, l) is projected onto the laboratory Cartesian coordinate system by the relation

$$\mathbf{x}_b = \Phi\,\mathbf{U}\,\mathbf{B}\,\mathbf{h} \qquad (3\text{-}1\text{-}1.3)$$

Here, **B** is the orthogonalization matrix from the crystal coordinate system to the Cartesian coordinate system (Busing & Levy, 1967). **U** is the crystal orientation matrix which rotates the Cartesian coordinate system onto the laboratory coordinate system and is defined as

$$\mathbf{B} = \begin{pmatrix} a & b\cos\gamma & c\cos\beta \\ 0 & b\sin\gamma & -c\sin\beta\cos\alpha \\ 0 & 0 & 1/c \end{pmatrix} \qquad (3\text{-}1\text{-}1.4)$$

$$\mathbf{U} = \begin{pmatrix} \cos p3 & -\sin p3 & 0 \\ \sin p3 & \cos p3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos p2 & 0 & \sin p2 \\ 0 & 1 & 0 \\ -\sin p2 & 0 & \cos p2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos p1 & -\sin p1 \\ 0 & \sin p1 & \cos p1 \end{pmatrix}$$

$$= \begin{pmatrix} \cos p3\cos p2 & \cos p3\sin p2\sin p1-\sin p3\cos p1 & \cos p3\sin p2\cos p1+\sin p3\sin p1 \\ \sin p3\cos p2 & \sin p3\sin p2\sin p1+\cos p3\cos p1 & \sin p3\sin p2\cos p1-\cos p3\sin p1 \\ -\sin p2 & \cos p2\sin p1 & \cos p2\cos p1 \end{pmatrix} \qquad (3\text{-}1\text{-}1.5)$$

where p1, p2 and p3 are the rotation angle in the X, Y and Z axes respectively. Φ represents the rotation matrix around the spindle axis to swing (h, k, l) onto the Ewald sphere,

$$\Phi = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad (3\text{-}1\text{-}1.6)$$

The $\mathbf{x}_b(x_b, y_b, z_b)$ represents any reciprocal point on the Ewald sphere (Arndt & Wonacott, 1977). The relations between $\mathbf{x}_b(x_b, y_b, z_b)$ and the ideal IP plane system $\mathbf{x}_f(x_f, y_f)$ are from Fig. 3-1-1.2:

$$x_f = \frac{z_b}{1+x_b} D$$
$$y_f = \frac{y_b}{1+x_b} D \qquad (3\text{-}1\text{-}1.7)$$

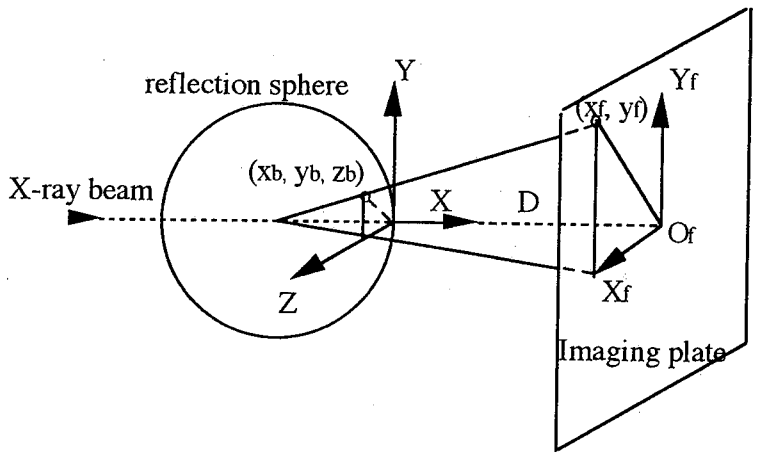


Fig. 3-1-1.2 Relation between reciprocal position vector $\mathbf{x}_b(x_b, y_b, z_b)$ of a reflection and its detector position coordinate $\mathbf{x}_f(x_f, y_f)$.

Coordinate system of Weissenberg type devices IPD-WAS and DIP-R300

The reciprocal-lattice coordinates (X, Y, Z) are described in an Cartesian laboratory coordinate system as same as DIP-100: X is parallel to the primary X-ray beam, Z to the rotation axis and Y is normal to the plane containing the X and

Y with right-handed system (Fig. 3-1-1.3a). The ideal IP coordinate system is different from DIP-100 because IP plate is arranged to be a cylindrical shape with radius R. The cylindrical detector moves synchronously with the crystal rotation in the direction parallel to the rotation axis (Fig. 3-1-1.3). The origin of pixel coordinates is at the down-right corner of IP plate, and IP pixel size is 3140 x 2400 for IPD-WAS and 5200 x 4400 for DIP-R300 (Fig. 3-1-1.3b). The cylindrical center axis is normal to the X axis, and the crystal is on this axis. The ideal detector coordinate (Xw, Yw) defined on the cylindrical detector plane in Fig. 3-1-1.3. The origin is in the center of IP plate. Xw is parallel to the X and Yw is in the radial direction of the cylinder and parallel to Y. The actual detector coordinate system (Xm, Ym) is defined on Fig. 3-1-1.3. The origin Ow(X0, Y0) is taken at the direct beam position when detector setting is zero. The angle between the Xw and Xm is ω3. The coordinate systems (Xw, Yw) and (Xm, Ym) are related by :

$$X_m = X_0 + X_w * \cos\omega 3 - Y_w * \sin\omega 3$$
$$Y_m = Y_0 + X_w * \sin\omega 3 + Y_w * \cos\omega 3 \qquad (3\text{-}1\text{-}1.8)$$

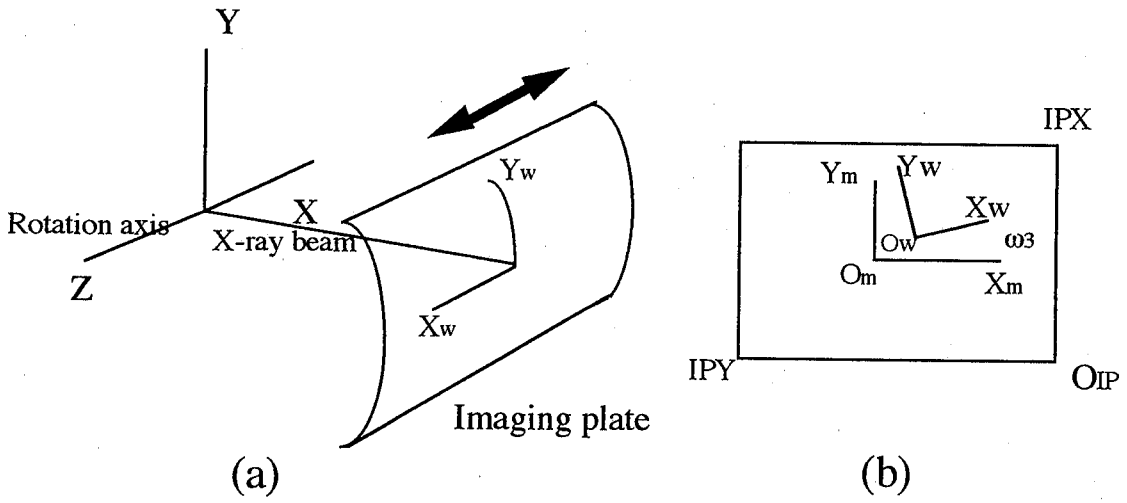


Fig. 3-1-1.3 Definition of the coordinate of IPD-WAS and DIP-R300.

The relations between the reciprocal-lattice point (h,k,l) and the Cartesian reciprocal-lattice coordinates (Xq, Yq, Zq) are as same as DIP-100 described at Eqs. (3-1-1.3)~(3-1-1.6). The ideal detector coordinate Xw,Yw are given by

$$X / Xq = Y / Yq = Z / Zq \tag{3-1-1.9}$$

$$X^2 + Y^2 + Z^2 = R^2 \tag{3-1-1.10}$$

$$\begin{aligned} Xw &= Z + (\phi - \phi_S)\, C \\ Yw &= R\alpha \\ \tan(\alpha) &= Y / X \end{aligned} \tag{3-1-1.11}$$

where ϕ_S is the starting angle of the oscillation and ϕ is the angle in which the reflection occurs. C is the coupling ratio of the crystal rotation and the detector translation. Although the DIP-R300 detector system is Weissenberg type, distance from crystal to IP plate is adjustable. Therefore, the ideal detector coordinates (Xw, Yw) are given by

$$\begin{aligned} &X / Xq = Y / Yq = Z / Zq \\ &(X + (R-D))^2 + Y^2 + Z^2 = R^2 \end{aligned} \tag{3-1-1.12}$$

$$\begin{aligned} Xw &= Z + (\phi - \phi_S)\, C \\ Yw &= R\alpha \\ \tan(\alpha) &= Y / (X+R-D) \end{aligned} \tag{3-1-1.13}$$

where R is cylindrical radius, and D is the distance between the crystal and IP plate (Fig. 3-1-1.4).

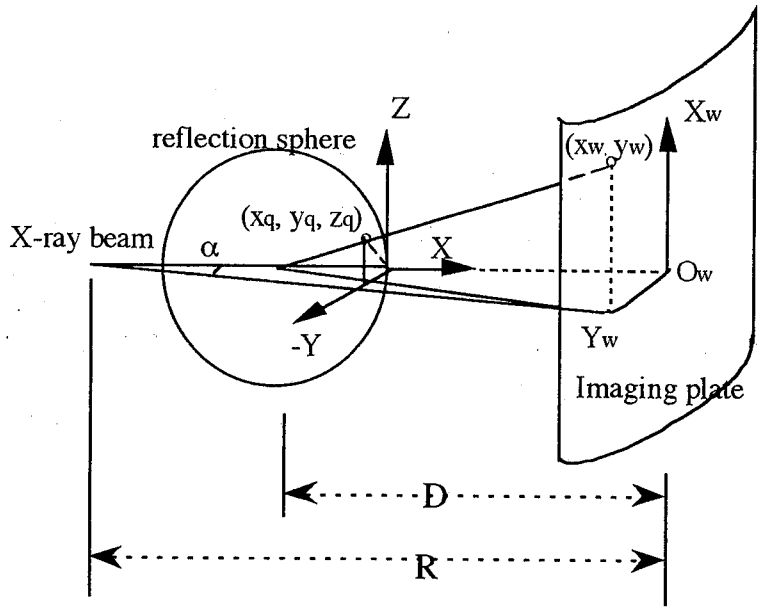


Fig. 3-1-1.4. Relation between reciprocal position vector $\mathbf{x}_q(x_q, y_q, z_q)$ of a reflection and its detector position coordinate $\mathbf{x}_w(x_w, y_w)$.

3-1-2. Whole program system

Fig. 3-1-2.1 shows the screen view of software system used to obtain intensity data from photographs. The whole program system is named ELMS for plate detector (DIP-100) and WELMS for Weissenberg detector (IPD-WAS, DIP-R300), respectively. The programs are written in FORTRAN and C for graphics routines with the X11 library, and execute on the workstation SUN or SGI. DIP-R300 system uses IRIS 3D graphics(gl) library for automatic crystal alignment. The system consists of 5 sub-programs and 14 functions, and 5 windows are used. Five sub-programs can be selected from main menu window by mouse and 14 functions from sub-menu (Fig. 3-1-2.1).

The LIST sub-program allows the user to display the contents of all the ASCII and binary data files produced by the program system.

The EDIT sub-program is used to display and modify control parameters for the program system.

The REFINE sub-program invokes automatic indexing (ATIX), parameter refinement and manual crystal orientation routines. These functions use the centroid coordinates of the spot from stationary photographs, excluding IDP-WAS. The algorithm of ATIX was described in section 3-2-1. The parameter refinement uses least-square method. In the ELMS, parameter refinement consists of three functions (CCLS, RSLS RGSLS). The CCLS function refine camera constants $\omega 1$, $\omega 2$, $\omega 3$, X0, Y0, μ, and D. Usually the first three parameters are close to zero and only D as read from the instrument is required input. The quantity minimized is the sum of the squared distance between observed and calculated reflection positions

$$M = \Sigma \{(Xmobs-Xmcal)^2 + (Ymobs-Ymcal)^2\} \qquad (3\text{-}1\text{-}2.1)$$

The RSLS and RGSLS function refine setting parameters p1, p2, p3 without and with unit cell dimensions, respectively. The quantity minimized is

$$M = \Sigma (1-Rcal)^2 \qquad (3\text{-}1\text{-}2.2)$$

where Rcal represents the distance of the each reciprocal-lattice point in the diffracting condition from the center of Ewald sphere (measured in dimension-less reciprocal-lattice units). In the WELMS, refinement program uses eigenvalue filtering (see section 3-2-2), and refines camera ($\omega 3$, μ, X0, Y0, D, R), setting (p1, p2, p3) parameters and unit cell dimensions.

The FRAM consists of five functions: (a) display the image on window; (b) background-level estimation and peaks search; (c) generates(GEN) all predicted positions of reflections within a given rotation range, display them on the window, and write them to file which is used in the integration function; (d) integrates(INT) the intensities at predicted position of the reflection in a file made with the GEN function. In our system, there are two methods of integration: mask and profile-fitting. The former integrates intensities in a mask defined by user. The latter uses a profile empirically determined for a given area in integrating intensities; (e) refines geometrical shift parameter about reflection pattern position (x_m, y_m); (f) indexing; The GEN and the INT function have been further developed to be able to process splitted spots caused by $K\alpha 1$, $K\alpha 2$ (see section 3-2-3).

The BATCH is used for processing data set automatically. LP correction is also made here.

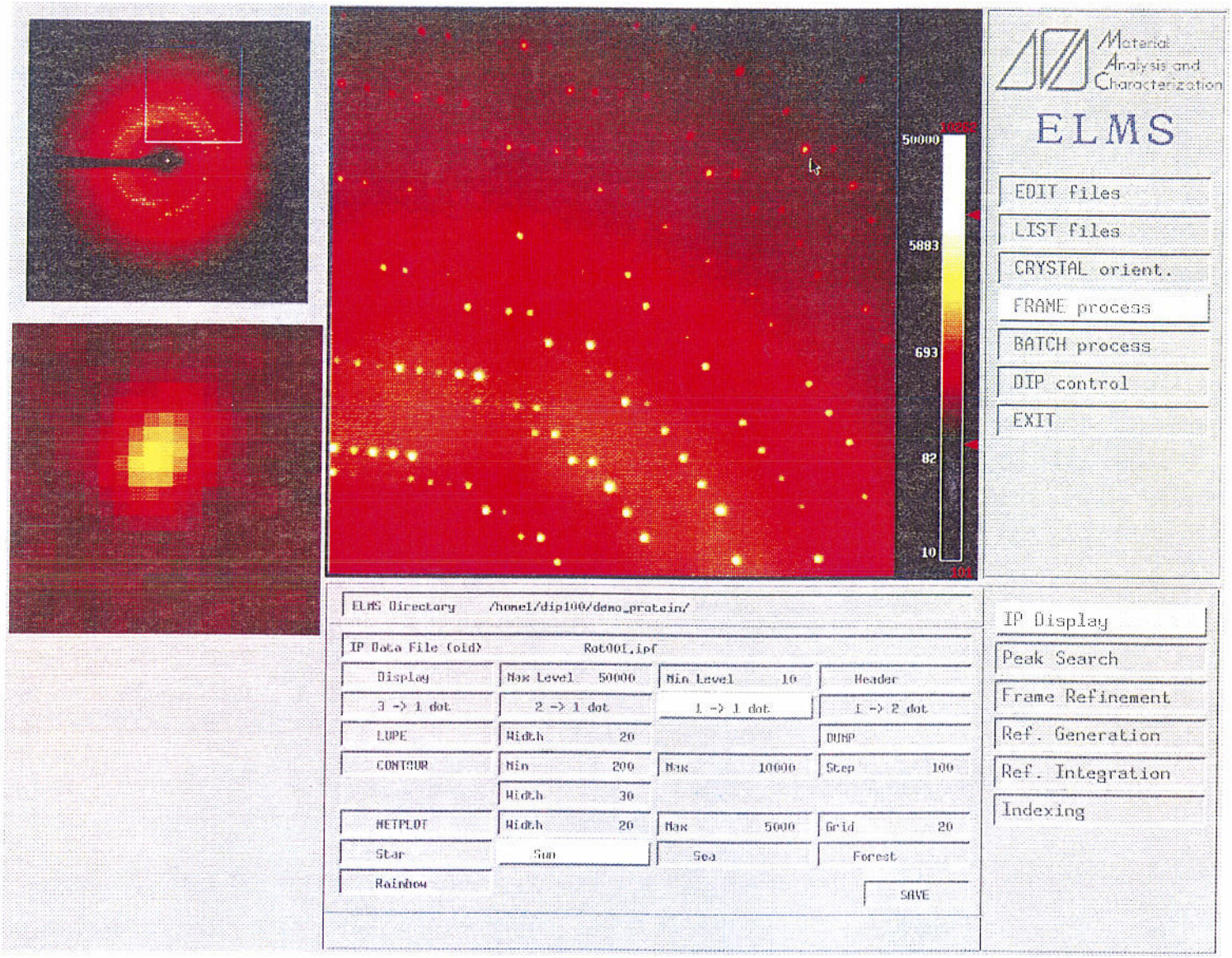


Fig 3-1-2.1 The screen view of ELMS system.

3-2. Algorithm and programming

3-2-1. Automatic indexing

3-2-1.1. Introduction

In collecting diffraction data, a complicated process of crystal alignment may seriously affect the amount of useful data obtainable within a limited time before crystal deteriorates, because alignment of crystals is a time-consuming process and many crystals of the protein are radiation sensitive. When crystal is once exposed, the radiation damage continues to process even in the absence of further radiation. ATIX (and WATIX) sub-program was developed for finding the orientation of crystal from reciprocal-space lattice vectors using double reflections method, under the condition that approximate unit-cell parameters are known. In this section, we describe two methods of orienting single crystals, which is especially useful for the application of the imaging plate system to protein crystals (DIP-100, DIP-R300) and small molecule crystals (IPD-WAS) .

3-2-1.2 General Method

The coordinates $\mathbf{x}_b(x_b, y_b, z_b)^T$ of any reflections are represented with respect to the laboratory Cartesian coordinate system from Eqs. (3-1-1.3) :

$$\mathbf{x}_b = \Phi\, \mathbf{x}_c \qquad (3\text{-}2\text{-}1.1)$$

and

$$\mathbf{x}_c = \mathbf{U}\, \mathbf{B}\, \mathbf{h} \qquad (3\text{-}2\text{-}1.2)$$

where the matrix **B** and **U** are given in Eqs. (3-1-1.4) and (3-1-1.5). The orientation of a crystal is conveniently represented by a 3x3 orientation matrix **U**. **h** is the vector of indices$(h, k, l)^T$. The matrix **U** can be given by

$$\mathbf{U} = \mathbf{x}_C (\mathbf{hB})^{-1} \qquad (3\text{-}2\text{-}1.3)$$

The problem of finding orientation of the crystals whose cell dimension are known is searching for the two non-linear base vectors in reciprocal space that can be correctly assigned to indices such (1, 0, 0), (0, 0, 1). So, low resolution reciprocal-space vectors are required. We can use the differences of the reciprocal-space vectors for this purpose.

$$\mathbf{x}_{cij} = \mathbf{x}_{ci} - \mathbf{x}_{cj} \qquad i < j \qquad i,j = 1,2\cdots\cdots n \qquad (3\text{-}2\text{-}1.4)$$

The low resolution vectors in reciprocal space are calculated up to the selected resolution and accumulated in a three-dimensional grid. The grid size is usually set to one-tenth of minimum unit cells in reciprocal space. Then they are sorted according to their length (Kabsch, 1988)).

The next step is to select sets of two base vectors $\mathbf{v}_i$ from $\mathbf{x}_{cij}$ by two criteria:

$$\begin{aligned} \text{frequency}(\mathbf{v}_i) &> NC && (NC > 1) \\ \frac{\mathbf{v}_1 * \mathbf{v}_2}{|\mathbf{v}_1|\,|\mathbf{v}_2|} &> \cos c1 && (c1 = 45 \text{ deg}) \end{aligned} \qquad (3\text{-}2\text{-}1.5)$$

Among the shortest vectors with maximum frequency, some sets of base vector are selected here, and are used to establish the orientation matrix **U**. Since the cell parameters are known, we can calculate the scattering vectors $\mathbf{h}_C$ in crystal Cartesian system of any reflection **h** by matrix **B** :

$$\begin{aligned} \mathbf{h}_{c1} &= \mathbf{B}\,\mathbf{h}_1 \\ \mathbf{h}_{c2} &= \mathbf{B}\,\mathbf{h}_2 \end{aligned} \qquad (3\text{-}2\text{-}1.6)$$

Using Eq. (3-2-1.3), the orientation matrix can be constructed that satisfies:

$$\mathbf{v}_1 = \mathbf{U}\,\mathbf{h}_{c1}$$
$$\mathbf{v}_2 = \mathbf{U}\,\mathbf{h}_{c1} \qquad (3\text{-}2\text{-}1.7)$$

However, because of the experimental errors in the measurements or uncertainties in the camera and cell parameters, it is not possible to find an orientation matrix **U** which exactly satisfies Eq. (3-2-1.7). Therefore, the orientation matrix is constructed by the double reflections method, as is done in four-circle diffractometry (Busing & levy, 1967) :

In the reciprocal-space coordinate system, define a right-handed orthogonal unit vector triple $\mathbf{t}_{v1}$, $\mathbf{t}_{v2}$, $\mathbf{t}_{v3}$ as $\mathbf{t}_{v1}$ is parallel to $\mathbf{v}_1$, $\mathbf{t}_{v2}$ is perpendicular to $\mathbf{v}_1$ and lies in the plane of $\mathbf{v}_1$ and $\mathbf{v}_2$, $\mathbf{t}_{v3}$ is perpendicular to this plane. In addition, define another such triple, $\mathbf{t}_{h1}$, $\mathbf{t}_{h2}$, $\mathbf{t}_{h3}$ in the crystal Cartesian coordinate system based on $\mathbf{h}_{c1}$ and $\mathbf{h}_{c2}$.

$$\mathbf{t}_{v1} = \frac{\mathbf{v}_1}{|\mathbf{v}_1|}\,, \quad \mathbf{t}_{v3} = \frac{\mathbf{v}_1 * \mathbf{v}_2}{|\mathbf{v}_1 * \mathbf{v}_2|}\,, \quad \mathbf{t}_{v2} = \mathbf{t}_{v1} * \mathbf{t}_{v3}$$
$$\mathbf{t}_{h1} = \frac{\mathbf{h}_{c1}}{|\mathbf{h}_{c1}|}\,, \quad \mathbf{t}_{h3} = \frac{\mathbf{h}_{c1} * \mathbf{h}_{c2}}{|\mathbf{h}_{c1} * \mathbf{h}_{c2}|}\,, \quad \mathbf{t}_{h2} = \mathbf{t}_{h1} * \mathbf{t}_{h3} \qquad (3\text{-}2\text{-}1.8)$$

Since these two unit vector triples can exactly be superimposed on each other, the equation (3-2-1.7) can be replaced by

$$\mathbf{t}_{vi} = \mathbf{U}\,\mathbf{t}_{hi} \qquad i = 1, 2, 3 \qquad (3\text{-}2\text{-}1.9)$$

Using matrix expression to represent these three vector equations, the orientation

matrix **U** is given by

$$\mathbf{T}_v = \mathbf{U}\,\mathbf{T}_{hc}$$

and

$$\mathbf{U} = \mathbf{T}_v\ (\mathbf{T}_{hc})^{-1} \qquad (3\text{-}2\text{-}1.10)$$

where $\mathbf{T}_v$ and $\mathbf{T}_{hc}$ are the matrices whose column vectors are $\mathbf{t}_{vi}$ and $\mathbf{t}_{hi}$ respectively. Since $\mathbf{T}_{hc}$ is orthogonal, Eq. (3-2-1.10) becomes

$$\mathbf{U} = \mathbf{T}_v\,\mathbf{T}_{hc}^{T} \quad (u_{ij} = t_{vik}\,t_{hcjk} \quad i, j, k = 1,2,3) \qquad (3\text{-}2\text{-}1.11)$$

from Eq. (3-1-1.4), we can obtain crystal orientation angles p1,p2,p3 as follows:

$$p1 = \arctan(\,u_{32} / u_{33}\,) = \arctan(\,(t_{v3k}\,t_{hc2k})/(\,t_{v3k}t_{hc3k})\,)$$
$$p2 = \arcsin(-u_{31}) = \arcsin(-t_{v3k}\,t_{hc1k})$$
$$p3 = \arctan(\,u_{21} / u_{11}) = \arctan(\,\,(t_{v2k}\,t_{hc1k})/\,(t_{v1k}\,t_{hc1k})\,) \qquad (3\text{-}2\text{-}1.12)$$

In principle, initial matrix **U** can be obtained from two best vectors $\mathbf{v}_i$ by Eq. (3-2-1.8) ~ (3-2-1.12). So, initial indices $\mathbf{h'}_i$ can be assigned to the vector $\mathbf{v}_i$ using initial matrix **U**.

$$\mathbf{h'}_i = (\,\mathbf{UB}\,)^{-1}\,\mathbf{v}_i \qquad (3\text{-}2\text{-}1.13)$$

The matrix **U** can be refined based on the fact that the indices of a reflection must be integers. In the actual procedure, refinement is carried out by two steps. The first step is refinement **U** using $\mathbf{x}_{ij}$ with Eq. 3-1-2.1. During the refinement those vectors are rejected from $\mathbf{x}_{ij}$ which do not satisfy the limiting condition of integer:

$$| \mathbf{h}_i - \mathbf{h'}_i | < \quad d \qquad (3\text{-}2\text{-}1.14)$$

where $\mathbf{h}_i$ is the nearest integer of $\mathbf{h'}_i$. The second step is the refinement for the matrix **U** using $\mathbf{x}_i$ instead of using $\mathbf{x}_{ij}$.

3-2-1.3. Automatic indexing sub-program ATIX on DIP-100 and DIP-R300

In the DIP-100 and DIP-R300, automatic indexing sub-program ATIX included in ELMS and WELMS software package is applied using two stationary photographs whose ϕ angles are separated by 90°. The reflection coordinates are obtained by peak-search function, and reciprocal-space vectors are given by Eq. (3-1-1.7) for DIP-100 and Eqs. (3-1-1.9), (3-1-1.12) and (3-1-1.13) for DIP-R300. The program consists of five steps and flow chart is shown at Fig. 3-2-1.1.

step1(CALPIJ): The vectors $\mathbf{x}_{ij}$ between reciprocal coordinates $\mathbf{x}_i$ of the reflections on stationary patterns are calculated. The shortest 100 vectors, whose frequency of occurrence is greater than a given value NC, are kept in a way mentioned above.

step2(GETIDX): All reciprocal-lattice vectors (up to 500) whose reciprocal length is within a given range are generated from the given cell parameters.

step3(ITANG): The shortest N vectors (usually $N \approx 20$) in step1 are used for this step. The triangles formed by any pair of vectors (which satisfy Eq. (3-2-1.5)) are compared with those formed by the calculated vectors in step2. The combinations for which all three edge-length differences are shorter than a given threshold are kept. The best 100 combinations in terms of sum of the differences

of edge-lengths are kept for the next step.

The flow chart of ATIX program

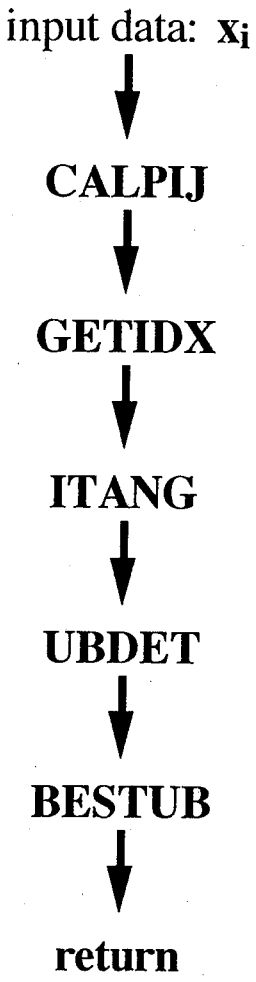


Fig. 3-2-1.1 The flow chart of the sub-program ATIX.

step4(UBDET): For all 100 combinations, the mis-orientation angles p1, p2 and p3 are calculated by the double-reflection method Eq. (3-2-1.8), (3-2-1.11), (3-2-1.12) and refined by RSLS using only the shortest N of the $\mathbf{x}_{ij}$. The results are sorted based on the number of rejected reflections (which do not satisfy Eq. (3-2-1.14)) and the sum of the absolute shift angle of p1, p2 and p3.

step5(BESTUB): The best solution in step4 is subject to the refinement using all 100 $\mathbf{x}_{ij}$ 'reflections' followed by one using all $\mathbf{x}_i$ reflections. The refinement terminates after repeated use of RSLS, CCLS, RSLS, CCLS and RSLS (RSLS and CCLS see chapter 3-1-2). An optional switch invokes the refinement of cell parameters.

3-2-1.4 Automatic indexing sub-program WATIX on IPD-WAS

A new technique to determine the orientation of a single crystal is presented at the IPD-WAS program system. In this section, we will describe this new technique for orienting single crystals. The technique requires a mechanism that allows transitional movement of the IP synchronized with rotational movement of the crystal around a spindle axis (so-called Weissenberg movement) with an adjustable film-translation to crystal-oscillation (T/O) gear ratio. The technique utilizes two Weissenberg photographs taken in the same oscillation range with different T/O gear ratios. Practically, one is a Weissenberg photograph and the other is simply an oscillation photograph. The technique is especially useful for single-crystal data collection systems using the imaging plate as an area detector when applied to small molecules.

In protein crystallography, still photographs or oscillation photographs with a narrow oscillation range, e.g. 1°, are often used to obtain, though in approximate form, the required information. However, this approach is not adequate for the

alignment of small molecule crystals, where only a few reflections are recorded on a still photograph, and a relatively wide oscillation range is adopted for efficient data collection.

Strategy

To determine the crystal orientation parameters, knowledge of the reciprocal coordinates of three non-planar vectors is necessary (Busing & Levy, 1967). In practical approaches a large number of reflections are used with the least-squares method. These data are straightforwardly obtained with a four-circle diffractometer where the rocking curve of each reflection can be recorded. In a system using IP, rocking curves cannot be obtained because the IP is an integration-type detector like X-ray photographic film, and information as to the time (or the crystal rotation angle ϕ around the spindle axis) at which the reflection occurs is not easily available.

In the present method, two orientation photographs are taken. They are screenless Weissenberg photographs taken in the same crystal oscillation range but with different T/O gear ratios. Identical reflection pairs recorded on different frames are detected and the ϕ values are determined from the positional differences of these reflections. Once the angle ϕ at which the reflection occurs is obtained, the reciprocal coordinate can be easily calculated from the ϕ value together with the reflection position on the IP (Arndt & Wonacott, 1977). In practical applications, a pair of orientation photographs covering about 10° of rotation is sufficient for determining the crystal orientation. However, a second set of orientation photographs, which are 90 degrees apart from the first, may be used to obtain more accurate orientation parameters.

Let (X,Y) be the position of a reflection recorded on the imaging plate, the origin of which is at the direct beam position. The X, Y axis is defined as the

section 3-1-1. X axis is parallel to the crystal spindle axis and the Weissenberg translation axis, Y axis is perpendicular to the X axis (Fig 3-2-1.2). (X',Y') is the position of a reflection recorded when there is no Weissenberg movement. (X1,Y1) and (X2,Y2) are the positions of the same reflection recorded when the T/O gear ratios are C1 and C2 mm/deg, respectively, (C2 > C1). Let ϕ_s and ϕ_e be the starting and ending angles of the oscillation for recording a particular frame, and ϕ be the rotation angle at which the reflection occurs. These quantities are related by the following equations.

$$X1 = X' + C1(\phi - \phi_s) \qquad (3\text{-}2\text{-}1.15)$$

$$X2 = X' + C2(\phi - \phi_s) \qquad (3\text{-}2\text{-}1.16)$$

From these equations we obtain

$$X2\text{-}X1 = (\phi - \phi_s)(C2\text{-}C1) \qquad (3\text{-}2\text{-}1.17)$$

or

$$\phi = (X2\text{-}X1) / (C2\text{-}C1) + \phi_s \qquad (3\text{-}2\text{-}1.18)$$

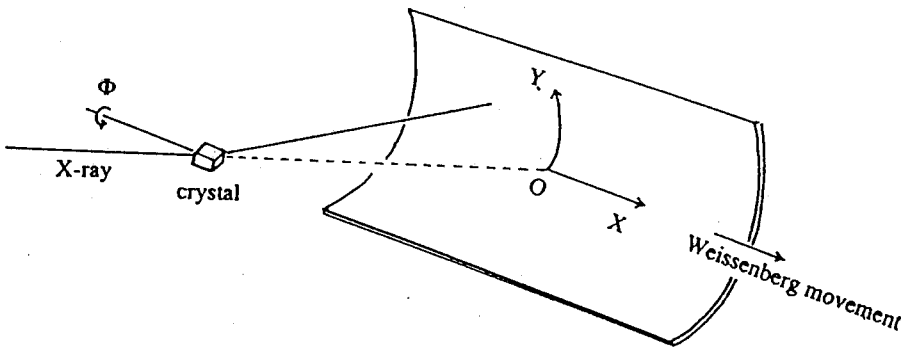


Fig. 3-2-1.2 Weissenberg geometry.

This last equation is used to calculate the ϕ value. The reciprocal coordinates of the reflection are obtained once ϕ is calculated (Arndt & Wonacott, 1977). These reciprocal coordinates can be fed into any kind of automatic indexing program (Kabsch, 1988, Higashi, 1990). Using Eq. (3-2-1.17) the conditions necessary for two reflections recorded on two photographs to be identical are given as follows.

$$-\Delta X < X2 - X1 < (\phi_e - \phi_s) (C2 - C1) + \Delta X \qquad (3\text{-}2\text{-}1.19)$$

$$-\Delta Y < Y2 - Y1 < \Delta Y \qquad (3\text{-}2\text{-}1.20)$$

$$|I1 - I2| < \Delta m(I1 + I2) / 2 \qquad (3\text{-}2\text{-}1.21)$$

Here, I1 and I2 are the observed intensities of the two reflections on the two orientation photographs, which should be the same within experimental error. Y1 and Y2 should also be the same since the T/O gear ratio has no effect on the Y coordinates. ΔX and ΔY are the allowance limits of the positional error and Δm is that for the intensity error.

Programming and result

The WATIX sub-program package in IPD-WAS system consists of two parts, GETFI and WATIX. The peak search sub-program searches in the two IP image files for reflections whose intensities are above a threshold and saves the (X,Y) coordinates of the centroid of the reflections and their integrated intensities. Each file is then sorted according to the (X,Y) coordinates. The GETFI reads the two peak search sub-program output files and searches for identical reflection pairs satisfying conditions of the Eqs. (3-2-1.19) ~ (3-2-1.21). The reflection pairs which uniquely satisfy all three conditions are saved. Their ϕ values are calculated from equation (3-2-1.18) and included in the output file together with the (X,Y) coordinates. Our automatic indexing program WATIX reads this file

and calculates the crystal orientation by a somewhat similar way to ATIX.

In usual applications, one of the two orientation photographs is simply an oscillation photograph (C1=0). The accuracy of the ϕ value determined depends on the accuracy of the positional difference which is generally within the error limit of 100μm. Therefore, when C2 is 1 mm/deg, the angular error is less than 0.14°, sufficiently accurate for crystal orientation purposes. One should note that reflections on the still photographs or narrow oscillation photographs, which are routinely used for orientation in protein crystallography, may have larger angular errors.

The ϕ values for all the reflections cannot be calculated because of the occasional overlap of the reflections. More than one reflection could be observed on the second frame in a region calculated from the reflection positions on the first frame using equations (3-2-1.19) and (3-2-1.20). A larger difference in the gear ratio results in a larger number of overlapping reflections. Table 3-2-1 shows the results of two test calculations for hypothetical small molecule crystals. In both cases, the film-crystal distance was 150mm and the X-ray wavelength was 0.7107Å. With five different crystal orientations chosen randomly, the reflection positions were calculated and fed into the GETFI program. In these calculations, the intensity of each reflection was set to a fixed value. For crystal No. 1, which has relatively small cell dimensions, a rotation range of 10° was adopted. When C2 was 1 mm/deg, almost all reflection pairs were identified and their ϕ values were calculated. When C2 was 5 mm/deg, about 20% of the reflections were lost because of the overlap. The estimated angular error 0.03° in the latter case may be too optimistic. In practice, the error in determining the reflection position limits the accuracy of the angle. For crystal No. 2 whose cell dimensions are larger, rotation by 3° provided about 1000 reflections of which more than 90% reflection pairs were uniquely determined when C2 was 1 mm/deg.

Table 3-2-1 Test calculation using two hypothetical crystals.

The calculation condition : Crystal film distance is 150mm; Film size is 314 mm x 300 mm; Wavelength is 0.7107Å; Resolution limit is 0.55Å; Allowance of error limit is ΔX=0.5mm and ΔY=0.5mm.

Crystal No. 1		
Cell dimension	a=5Å, b=6Å, c=7Å, α=90°, β=90°, γ=90°	
Rotation range	10°	
T/O gear ratio (C1=0mm/deg)	C2 = 1mm/deg	C2 = 5mm/deg
estimated averaged error angle	0.14°	0.03°
uniquely determined reflections (five different orientations)	205-211(98-100%)	156-187(74-89%)
Crystal No.2		
Cell dimension	a=10Å, b=15Å, c=20Å, α=90°, β=90°, γ=90°	
Rotation range	33°	
T/O gear ratio (C1=0mm/deg)	C2 = 1mm/deg	C2 = 5mm/deg
estimated averaged error angle	0.14°	0.03°
uniquely determined reflections (five different orientations)	880-968(90-100%)	642-941(66-85%)

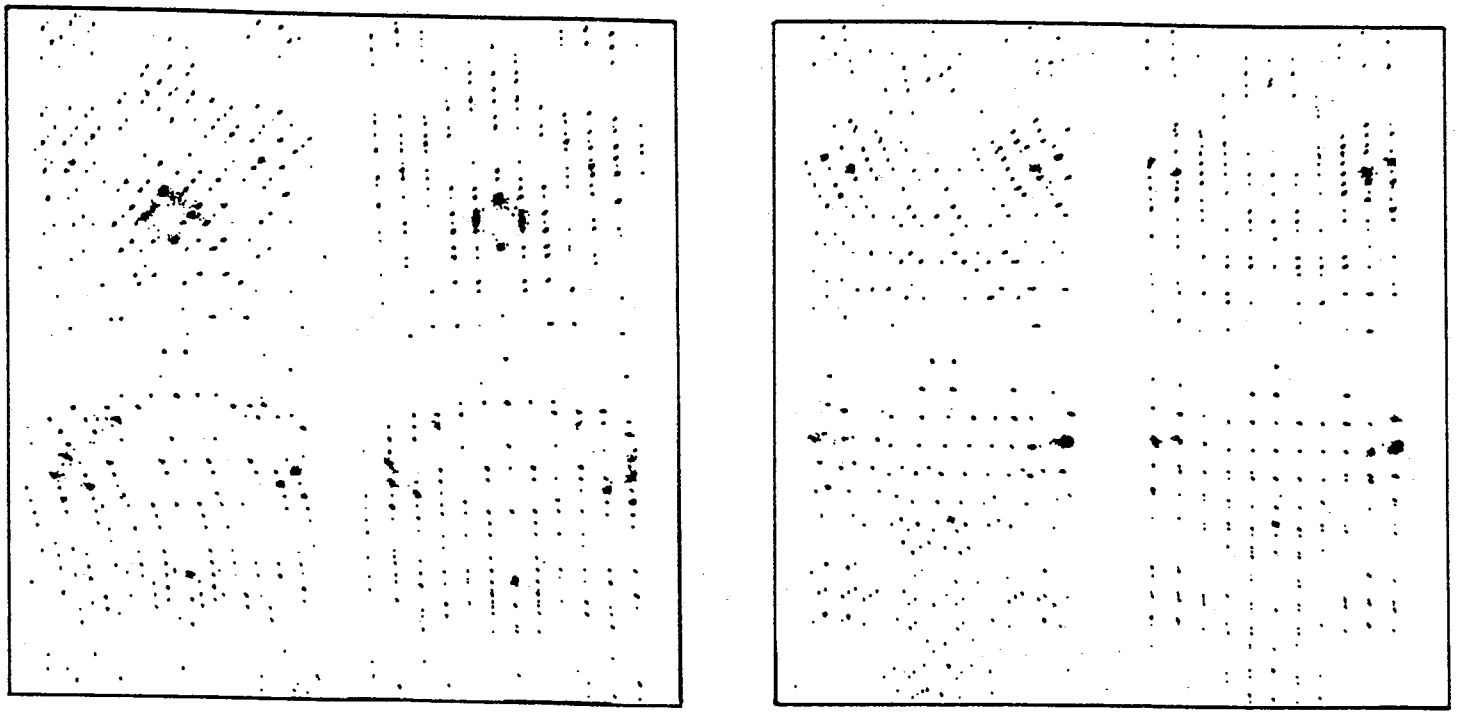


Fig. 3-2-1.3 Crystal orientation photographs taken by the IPD-WAS.
Sample : $[HRu_6C(CO)_{16}][\Phi_4P]$; cell dimensions: a=25.63Å, b=13.51Å, c=13.22Å, β=93.52°; space group $P2_1/n$. X-ray: Mo $K\alpha$ running at 45KV, 200mA, film-crystal distance 200mm, exposure time: 3.3min. Oscillation range : (a) 0 ~ 10° and (b) 90 ~ 100°. C1= 0 mm/deg (right half of each photo), C2= 1 mm/deg (left half of each photo). IP size : 314mm x 300 mm.

The technique has been applied to practical cases. In the first application, it was installed in the IPD-WAS system (Kamiya et al.) designed for studying phase transitions in the crystalline state. Since the IPD-WAS system has special screens to cover one half of the IP, two orientation photographs were taken together in one frame. Fig. 3-2-1.3 shows an example of the orientation photographs. In order to obtain accurate setting parameters, two sets of orientation photographs were taken; one at ϕ=0 ~10° (Fig. 3-2-1.3a) and the other at ϕ=90 ~ 100° (Fig. 3-2-1.3b). For both photographs, the right half is the oscillation and the left half is the Weissenberg photograph. Since the crystal (Chihara & Yamazaki, 1992) was aligned with its b axis almost parallel to the crystal spindle axis, the layer lines are visible in the right-hand halves of the photographs. Some of the strong identical reflections can be seen on both halves of the photograph.

The peak search program found 156 reflection positions from each half of photograph (Fig. 3-2-1.3a), from which 141 reflection pairs were identified. In photograph (Fig. 3-2-1.3b), 157 reflection positions were obtained, from which 129 pairs were identified. A total of 270 reflections were applied to the automatic indexing program. The setting parameters were determined and all the reflections were indexed. The r.m.s. residual of the ϕ values calculated from the obtained setting parameters was 0.061°.

The program has also been applied to a DIP-320, a Weissenberg-type IP diffractometer manufactured by Mac Science, and is used routinely for determining the setting orientation of small molecule crystals.

3-2-2. Application of eigenvalue filtering in the parameter refinement

3-2-2.1. Introduction

This chapter will describe an application of "eigenvalue filtering" for refining diffractometer parameters. In any diffractometer system with two dimensional detector, the observed reflections must be accurately indexed before their intensities are measured. The accurate indexing and the accurate integration depend on the precision of individual parameters such as unit-cell, crystal orientation and camera setting. The correct parameters can be obtained from initial values by least-squares refinement. However, the least-squares refinement of parameters often converge slowly, or even can not converge, or give inaccurate-parameters because of correlation between parameters. To get best parameters from observed data and expand converging range of refinement, a method of refinement is wanted that can automatically select those parameters or linear combinations of parameters to which the residual is most sensitive. The eigenvalues and eigenvectors of least-squares matrix provide the necessary information for automatic selection of the parameters to be refined. This technique has been applied to the refinement of protein structure in terms of conformational parameters by Diamond (Diamond, 1971), and the refinement of unit cell, crystal orientation and reflecting-range parameters using oscillation photography by Reeke (Reeke Jr, 1984). Here, we applied eigenvalue filtering to refine the camera-setting, unit-cell and crystal-orientation parameters.

3-2-2.2. Theory

Least-squares equations

The coordinate system of DIP-100, IPD-WAS and DIP-R300 are described in section 3-1-1. Here, let $\mathbf{x}_{mobs}$ $(x_{mobs}, y_{mobs},)^T$ be coordinates of actual observed reflections, $\mathbf{x}_{mcal}$ $(x_{mcal}, y_{mcal}, 1)^T$ be coordinates of the calculated

reflections (although $\mathbf{x}_{mcal}$ is two dimension, it is written here in three-dimension only for simple expression below). The coordinates of any reflections observed on IP can be transferred to reflection indices $\mathbf{h}_{obs}(h_{obs}, k_{obs}, l_{obs})^T$ by the relation:

$$\mathbf{h}_{obs} = (\Phi \mathbf{U} \mathbf{B})^{-1} \mathbf{C} \Omega \mathbf{x}_{mobs}, \qquad (3\text{-}2\text{-}2.1)$$

Where, **B** and **U** are the orthogonalization matrix from the crystal to the Cartesian coordinate system and the crystal orientation matrix given by Eq. (3-1-1.4) and (3-1-1.5), respectively. Φ is the rotation matrix around crystal spindle axis given by Eq. (3-1-1.6). Ω is idealization matrix from the actual IP(x_m, y_m, 1) to ideal IP coordinate system (x_w, y_w). From Eq. (3-1-1.8)

$$\Omega = \begin{pmatrix} \cos\omega 3 & \sin\omega 3 & -x0\cos\omega 3 - y0\sin\omega 3 \\ -\sin\omega 3 & \cos\omega 3 & x0\sin\omega 3 - y0\cos\omega 3 \end{pmatrix} \qquad (3\text{-}2\text{-}2.2)$$

C is not written here in matrix form but it is a conversion equation from (x_w, y_w) to Cartesian reciprocal-lattice coordinates (x_q, x_q, x_q) by Eqs. (3-1-1.9)~(3-1-1.11) for IPD-WAS, or Eqs. (3-1-1.9), (3-1-1.12) and (3-1-1.13) for DIP-R300.

Since the above calculation uses unit-cell parameters(a, b, c, α, β, γ), orientation parameters (p1, p2, p3), camera setting parameters (C, w3, x0, y0, D), $\mathbf{h}_{obs}$ may not be integer. But we can get ideal value $\mathbf{h}_{cal}$ from $\mathbf{h}_{obs}$ by taking nearest integer and can calculate reflection positions $\mathbf{x}_{mcal}$ from $\mathbf{h}_{cal}$ by inverted Eq. (3-2-2.1) with initial parameters. The final parameters are given by least-squares refinement. The quantity minimized is difference between observed and calculated reflection positions:

$$M = \Sigma \{(Xmobs-Xmcal)^2 + (Ymobs-Ymcal)^2\}$$

Representing the set of parameters refined in a particular run by q, calculated i reflection positions in k cycle express by Taylor expansion $f(q)_i$

$$f(q)_i = f(q^k)_i + \frac{\partial f(q)}{\partial q_1}\Delta q_1{}_{|q_1=q_1^k} + \frac{\partial f(q)}{\partial q_2}\Delta q_2{}_{|q_2=q_2^k} + \frac{\partial f(q)}{\partial q_3}\Delta q_3{}_{|q_3=q_3^k} \cdots \frac{\partial f(q)}{\partial q_n}\Delta q_n{}_{|q_n=q_n^k}$$

, and minimizing the least-squares problem in usual way. The partial derivatives $\frac{\partial f(q)}{\partial q_j}$ are straightforward but tedious to show, and will not be given here (the derivative may be obtained directly from Eq. (3-2-2.1)). To express least-squares problem by matrix form, let n be number of refined parameters, m be number of observed reflections, **b** be vector observed position of reflections, **q** be vector refined parameters, **f** be calculated reflection position vector and **A** be n x m matrix:

$$\mathbf{b} = (b_1, b_2, b_3, \ldots b_m)^T,$$

$$\mathbf{q} = (q_1, q_2, q_3, \ldots q_n)^T,$$

$$\mathbf{f} = (f(q^{(k)})_1, f(q^{(k)})_2, f(q^{(k)})_3, \ldots f(q^{(k)})_m)^T,$$

$$\mathbf{A} = \begin{pmatrix} \left(\frac{\partial f(q)}{\partial q_1}_{|q_1=q_1^k}\right)_1 & \left(\frac{\partial f(q)}{\partial q_2}_{|q_2=q_2^k}\right)_1 & \left(\frac{\partial f(q)}{\partial q_3}_{|q_3=q_3^k}\right)_1 & \cdots & \left(\frac{\partial f(q)}{\partial q_n}_{|q_n=q_n^k}\right)_1 \\ \left(\frac{\partial f(q)}{\partial q_1}_{|q_1=q_1^k}\right)_2 & \left(\frac{\partial f(q)}{\partial q_2}_{|q_2=q_2^k}\right)_2 & \left(\frac{\partial f(q)}{\partial q_3}_{|q_3=q_3^k}\right)_2 & \cdots & \left(\frac{\partial f(q)}{\partial q_n}_{|q_n=q_n^k}\right)_2 \\ \left(\frac{\partial f(q)}{\partial q_1}_{|q_1=q_1^k}\right)_3 & \left(\frac{\partial f(q)}{\partial q_2}_{|q_2=q_2^k}\right)_3 & \left(\frac{\partial f(q)}{\partial q_3}_{|q_3=q_3^k}\right)_3 & \cdots & \left(\frac{\partial f(q)}{\partial q_n}_{|q_n=q_n^k}\right)_3 \\ & & \ldots\ldots & & \\ \left(\frac{\partial f(q)}{\partial q_1}_{|q_1=q_1^k}\right)_m & \left(\frac{\partial f(q)}{\partial q_2}_{|q_2=q_2^k}\right)_m & \left(\frac{\partial f(q)}{\partial q_3}_{|q_3=q_3^k}\right)_m & \cdots & \left(\frac{\partial f(q)}{\partial q_n}_{|q_n=q_n^k}\right)_m \end{pmatrix} \quad (3\text{-}2\text{-}2.3)$$

So, an observation equation is given

$$\mathbf{b} = \mathbf{f} + \mathbf{A}\delta\mathbf{q} \quad (3\text{-}2\text{-}2.4)$$

By writing $\mathbf{t} = \mathbf{b} - \mathbf{f}$, we obtain a linear observation equation :

$$\mathbf{A}\delta\mathbf{q} = \mathbf{t} \qquad (3\text{-}2\text{-}2.5)$$

least-squares problem is

$$\mathbf{R}^2 = (\mathbf{A}\delta\mathbf{q} - \mathbf{t})^T (\mathbf{A}\delta\mathbf{q} - \mathbf{t})$$
$$= \delta\mathbf{q}^T \mathbf{A}^T \mathbf{A}\delta\mathbf{q} - 2\,\delta\mathbf{q}^T \mathbf{A}^T \mathbf{t} + \mathbf{t}^T \mathbf{t} \qquad (3\text{-}2\text{-}2.6)$$

$$\frac{\partial(\mathbf{R}^2)}{\partial\mathbf{q}} = 2\,\mathbf{A}^T \mathbf{A}\delta\mathbf{q} - 2\,\mathbf{A}^T \mathbf{t} = 0$$

$$\mathbf{A}^T \mathbf{A}\delta\mathbf{q} = \mathbf{A}^T \mathbf{t} \qquad (3\text{-}2\text{-}2.7)$$

By $\mathbf{N} = \mathbf{A}^T \mathbf{A}(\,n_X\,n\,)$, $\mathbf{c} = \mathbf{A}^T \mathbf{t}\,(\,1_X\,n)$, we can get the normal equations:

$$\mathbf{N}\,\delta\mathbf{q} = \mathbf{c} \qquad (3\text{-}2\text{-}2.8)$$

after k cycle refinement, calculated reflection positions are given by

$$\mathbf{q}^{(k)} = \mathbf{q}^{(k-1)} + \delta\mathbf{q}^{(k)} \qquad (3\text{-}2\text{-}2.9)$$
$$\mathbf{f}^{(k)} = \mathbf{f}^{(k-1)} + \mathbf{A}^{(k-1)}\delta\mathbf{q}^{(k)} \qquad (3\text{-}2\text{-}2.10)$$

In our system, refined parameters are (a, b, c, α, β, γ, p1, p2, p3, C, ω3, D, X0, Y0) for the DIP-R300 system and (a, b, c, α, β, γ, p1, p2, p3, C, w3, X0, Y0, μ) for the IPD-WAS system.

Eigenvalue filtering

In IPD-WAS and DIP-R300 systems, rotation angle ω3 between ideal detector and actual detector, crystal orientation parameter p1, p2 and p3 and angle κ

between X-ray beem and spindle axis are seriously correlated. In DIP-R300 system, unit cell dimensions are difficult to be refined accurately from spot positions because of variable distance between crystal and detector. For these reasons, we applied "eigenvalue filtering" to refine parameters in DIP-R300 and IPD-WAS systems. Let the eigenvalues and normalized eigenvectors of the matrix $\mathbf{N}(n \times n)$ of our least-squares normal equation be $\Lambda(\lambda_i)$ and $\mathbf{V}(\upsilon_i)$, respectively, for i=1,2, ... n.

$$\mathbf{N}\,\upsilon_i = \lambda_i \upsilon_i \qquad \text{(3-2-2-11)}$$
$$\mathbf{N}^{-1}\upsilon_i = 1/\lambda_i \upsilon_i$$

Since $\mathbf{N} = \mathbf{N}^T$, thus $\mathbf{V}^{-1} = \mathbf{V}^T$, and

$$\mathbf{V}^{-1}\mathbf{N}\,\mathbf{V} = \mathbf{V}^T\mathbf{N}\,\mathbf{V} = \Lambda \qquad \text{(3-2-2.12)}$$
$$\mathbf{V}^{-1}\mathbf{N}^{-1}\,\mathbf{V} = \mathbf{V}^T\mathbf{N}^{-1}\,\mathbf{V} = \Lambda^{-1}$$

which Λ and Λ^{-1} are equivalent to

$$\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & .. & \\ & & & \lambda_n \end{pmatrix}, \qquad \Lambda^{-1} = \begin{pmatrix} 1/\lambda_1 & & & \\ & 1/\lambda_2 & & \\ & & .. & \\ & & & 1/\lambda_n \end{pmatrix} \qquad \text{(3-2-2.13)}$$

By Eq. (3-2-2.12), Eq. (3-2-2.8) can reduces to

$$\delta\mathbf{q} = \mathbf{N}^{-1}\mathbf{c} = \mathbf{V}\,\mathbf{N}^{-1}\mathbf{V}^T\,\mathbf{c} \qquad \text{(3-2-2.15)}$$
$$\delta q_i = \sum_j^n \sum_i^n \frac{\upsilon_{j\,l}\upsilon_{j\,i}c_i}{\lambda_j} \qquad \text{(3-2-2.16)}$$

So that, the solution of our least-squares problem $\mathbf{N}\delta\mathbf{q} = \mathbf{c}$ can be obtained as superposition of eigenvalues and eigenvectors. Finding eigenvalues and eigenvectors of our least-squares matrix **N** but not directly solving the least-squares problem, can avoid the correlation between parameters, which frequently causes a failure of the refinement. From Eq. (2-2-2.16), it is clear that eigenvectors υ_i associated with large λi give the smallest changes in the parameters. The change in the residual during a cycle of refinement is dependent on large eigenvalue λ_i. In conclusion, eigenvectors υ_i associated with large λ_i give the largest on decrease in residual for the smallest changes in the parameters (see Appendix A).

3-2-2.3. Program features and results

The refinement program using "eigenvalue filtering" was included in WELMS system, which allows any combination of parameters. It consists of ordinal least-squares and eigenvalue filtering. The eigenvalue filtering contains calculating eigenvalue and eigenvector of least-squares matrix **N** by Jacobian method, sorting eigenvalue λ_i and selecting largest λ_i. The refined parameters shift by Eq. 3-2-2.16. The program can run by either manual intervention or automatic mode. On the manual refinement mode, user can select parameters to be refined. On automatic mode, parameters are selected automatically. The cycles of refinement, number of reflections used to refine, the error of indexing etc. are provided by user or by program, respectively. During refinement, reflections are rejected which do not satisfy the limiting condition of Eq. (3-2-1.14). The parameters are refined after analyzing eigenvalue.

Table 3-2-2.1 shows the result of refinement using eigenvalue filtering in which unit-cell, orientation, camera setting parameters in DIP-R300 are refined. The distance between crystal and detector has 1mm error at the first stage of the

Table 3-2-2.1 The part of refinement list using eigenvalue filtering.
Calculation conditions: Sample is chicken egg-white lysozyme(a=b=79.1Å c=39.7Å $P4_32_12$); Crystal detector distance is 150.456mm; Wavelength is 1.54178Å; Resolution limit is 1.8Å; Crystal detector distance D had been shifted intensionally by 1mm for program test.

		X0	Y0	ω3	D	cup	p1	p2	p3	a	b	c	α	β	γ	κ
N	EIGENVALUES	EIGENVECTORS														
1	378092.8	0.0047	0.0002	-0.6957	0.0000	0.0000	0.7002	0.1534	-0.0006	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0456
2	28109.82	0.0122	-0.0009	0.1148	0.0000	0.0000	-0.1058	0.9861	0.0001	0.0016	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0560
3	12598.16	0.0299	-0.0003	0.0256	0.0002	-0.0002	0.0765	0.0613	-0.0001	-0.0033	0.0000	0.0002	0.0000	0.0000	0.0000	0.9944
4	408.89	0.0063	-0.1565	-0.0468	-0.1336	-0.0002	-0.0458	-0.0008	0.0000	0.9762	0.0000	0.0156	0.0000	0.0000	0.0000	0.0078
5	210.38	-0.0084	-0.0305	0.7069	-0.0111	0.0002	0.7002	-0.0112	-0.0006	0.0609	0.0000	0.0049	0.0000	0.0000	0.0000	-0.0709
6	40.81	0.0292	0.9867	0.0148	-0.0182	-0.0003	0.0142	0.0000	0.0000	0.1566	0.0000	0.0167	0.0000	0.0000	0.0000	0.0015
7	26.05	0.9990	-0.0281	0.0069	-0.0040	-0.0025	0.0014	-0.0147	0.0000	-0.0109	0.0000	0.0009	0.0000	0.0000	0.0000	-0.0295
8	1.40	0.0047	0.0077	0.0035	0.6821	-0.0014	0.0035	-0.0003	0.0000	0.1064	0.0000	-0.7234	0.0000	0.0000	0.0000	-0.0002
9	1.00	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
10	1.00	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
11	1.00	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
12	1.00	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
13	0.01	0.0028	-0.0121	-0.0007	0.7185	-0.0197	-0.0006	-0.0002	0.0000	0.0853	0.0000	0.6899	0.0000	0.0000	0.0000	0.0000
14	0.00	0.0026	0.0000	-0.0002	0.0150	0.9998	-0.0001	0.0000	0.0000	0.0020	0.0000	0.0126	0.0000	0.0000	0.0000	0.0001

Largest 14 Eigenvalues are used
***** 1 CYCLE *****
TOTAL REFLECTIONS : 141
CHECKED REFLECTIONS : 70
USED REFLECTIONS : 50
REJECTED REFLECTIONS : 20

RMS RESIDUAL (TOTAL) : 0.40040
(XM,YM) :0.28023 0.28599

	X0	Y0	W3	CDIST	CUP	KAI
NEW :	1.521	-1.669	0.028	150.465	0.000	0.049
SHIFT:	0.0000	0.0000	0.0001	-0.9914	0.0000	0.0000

	E1	E2	E3
NEW :	-0.015	-0.573	0.296
SHIFT:	0.0001	0.0000	0.0008

CELL CONSTANTS

NEW :	31.355	31.355	200.289	90.000	90.000	120.000
SHIFT:	-0.0001	-0.0001	-0.0006	0.0000	0.0000	0.0000

refinement. For each eigenvector, large components are underlined. The eigenvector associated with the largest eigenvalue represents combination of ω3 and p1, which is the most sensitive for the residual. The eigenvectors associated with the next two largest eigenvalues represent p2 and κ. So the smallest changes were given for these parameters. The 13th largest eigenvalue 0.01 corresponds to refinement predominantly of the distance between crystal and detector.

3-2-3. The split problem in integrating intensity

3-2-3.1. Introduction

The another main pieces of hardware needed for the collection of X-ray diffraction data is X-ray source. In laboratory, X-ray generator uses a rotating anode tube. The sharp X-ray is produced by electron transitions between inner orbitals in the atom of the anode material. If electrons transfer from the L-shell to the K-shell, this process is emitted as X-ray radiation of specific wavelength called K_α. Because of the fine structure in the L-shell, K_α is split up in $K_{\alpha1}$ and $K_{\alpha2}$. Table 3-2-3.1 shows usual wavelength of copper and molybdenum. There is a small angle between two radiations with wavelength $K_{\alpha1}$ and $K_{\alpha2}$, after X-ray radiation containing two wavelengths is reflected by monochromator. The difference of wavelengths and angles of incident beam makes split of spots recorded on the detector. Usually, split caused by $K_{\alpha1}$ and $K_{\alpha2}$ can be ignored, and their weighted avarage K_α is used as wavelength of X-ray radiation. However, for high resolution reflection data collection, split of reflection spots recorded on the detector can not be ignored as shown at Table 3-2-3.2. So that, process of splitted spots was included into ELMS and WELMS system.

Table 3-2-3.1 the wavelength of copper and molybdenum anode

	Mo (λ)	Cu (λ)
$K_{\alpha1}$(Å)	0.70926	1.54056(1.54051)
$K_{\alpha2}$(Å)	0.71354	1.54439(1.54433)
K_α (Å)	0.71069	1.54184(1.54178)

$$\lambda_\alpha = \frac{2\lambda_1 + \lambda_2}{3}$$

Table 3-2-3.2 quantity (mm) splitted of reflection on IP of the IPD-WAS. collection condition: X-ray is Mo; camera radius is 200mm; $2\theta = 40°$; oscillation photograph. ($\Delta x1=x1-x$; $\Delta y1=y1-y$; $\Delta x2=x2-x$; $\Delta y2=y2-y$)

Kα (x, y)	$K_{\alpha 1}$ ($\Delta x1$, $\Delta y1$)	$K_{\alpha 2}$($\Delta x2$, $\Delta y2$)
-103.613 -108.819	0.165 0.014	-0.885 -0.741
103.613 -108.819	-0.150 0.028	0.690 -0.700
-103.613 108.819	0.165 -0.014	-0.885 0.741
102.613 108.819	-0.150 -0.028	0.690 0.700

3-2-3.2. Coordinate of positions of splitted reflections

In the DIP-100 and DIP-R300 systems, the method to calculate position of splitted reflections is simple. The camera setting and crystal orientation parameters are determined by coordinates (x, y) of reflection spot center with average wavelength λ. For each reflection, the center coordinates (x1, y1) and (x2, y2) of two position of reflections diffracted by incidence $K_{\alpha 1}$ and $K_{\alpha 2}$ are calculated by equations described at section 3-1-1 with $K_{\alpha 1}$ and $K_{\alpha 2}$, respectively. Then these coordinates are output into data file for integration, instead of coordinate (x, y) calculated with average wavelength λ.

In the IPD-WAS system, the method to calculate position of splitted reflections is more complicated than the DIP-100 and DIP-R300 systems. The camera setting and crystal orientation parameters are determined in the same way as DIP-100 and DIP-R300 system. For integration reflection coordinates (x, y) are calculated with average wavelength λ, and output into data file. Then center coordinates (x1, y1) and (x2, y2) of two position of reflections diffracted by incidence $K_{\alpha 1}$ and $K_{\alpha 2}$ are calculated by optical geometry from each coordinate (x, y). The calculation is straightforward but tedious to show here.

3-2-3.3. Integrating intensity of splitted reflections

As described at section 3-1-2, there are two methods for integration in our system. Bellowing, followings are these two methods when split of reflection due to $K_{\alpha 1}$ and $K_{\alpha 2}$ is considered.

Mask-integration

When mask is used for integration, it is fixed for one data set by assuming reflection-spots of a crystal have same shape. But, when incident X-ray is splitted due to $K_{\alpha 1}$ and $K_{\alpha 2}$ the spots at high resolution have different shape on one photograph (Fig. 3-2-3.2). So that, variable shape mask is used for each reflections. Before integration, two center coordinates (x1, y1) and (x2, y2) of each reflection corresponding to wavelength $K_{\alpha 1}$ and $K_{\alpha 2}$ are calculated. Then a new mask for each reflection is constructed using calculated two center coordinates (x1, y1) and (x2, y2) as shown at Fig. 3-2-3.1. The intensity of each reflection is integrated using individual mask.

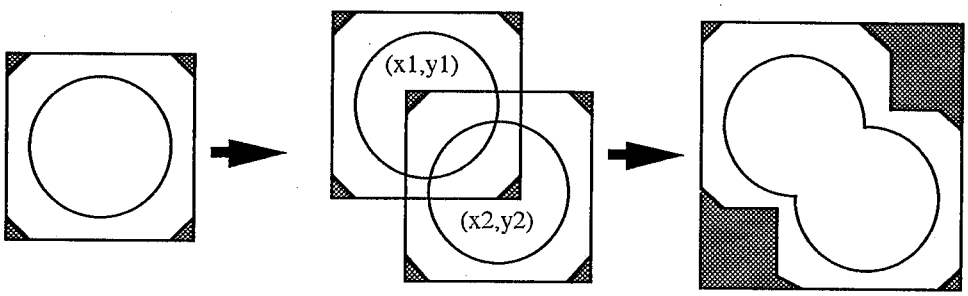


Fig. 3-2-3.1 Making split mask for integrating.

Integration by profile-fitting

Rossmann's method of integration by profile-fitting (Rossmann, 1979) divides detector to several regions and assumes reflection profiles are fixed for each region. The profile(P) of each region is calculated by minimization of

$$R = \sum_{i}^{n} \{J_i P - (\rho - \rho_b)\}^2 \qquad (3\text{-}2\text{-}3.1)$$

Here, J_i is intensity of ith reflection; n is number of reflection in the region; ρ is observed intensity, ρ_b is estimated value of background calculated by

$$\rho_b = ax + by + c \qquad (3\text{-}2\text{-}3.2)$$

where x, y is coordinate of pixel point of a reflection, and constants a, b and c are calculated with background region. But, this method is based on one wavelength λ and does not correspond to split problem. So that, we improved Rossmann's method for split problem. First, we suppose (1) two profiles caused by Kα1 and Kα2 have same shape; (2) the weight of those profiles caused by $K_{\alpha 1}$ and $K_{\alpha 2}$ are 2:1, same as the intesity of $K_{\alpha 1}$ and $K_{\alpha 2}$. Then whole profile is given by

$$\sum_{i}^{n} \left\langle J_i \{2P(x1, y1) + P(x2, y2)\} - (\rho - \rho_b) \right\rangle^2 \qquad (3\text{-}2\text{-}3.3)$$

However, since profile P is calculated at once, calculation time is very large: 2~3 min, for one profile P on the Sun-sparc workstation.

Fig 3-2-3.2 illustrates actual shape of reflection spot splitted and integrating mask calculated (Kamiya & Iwasaki, 1993). It is obvious that correct intensity of reflections may be obtained using each splitted mask than using single fixed mask for integration.

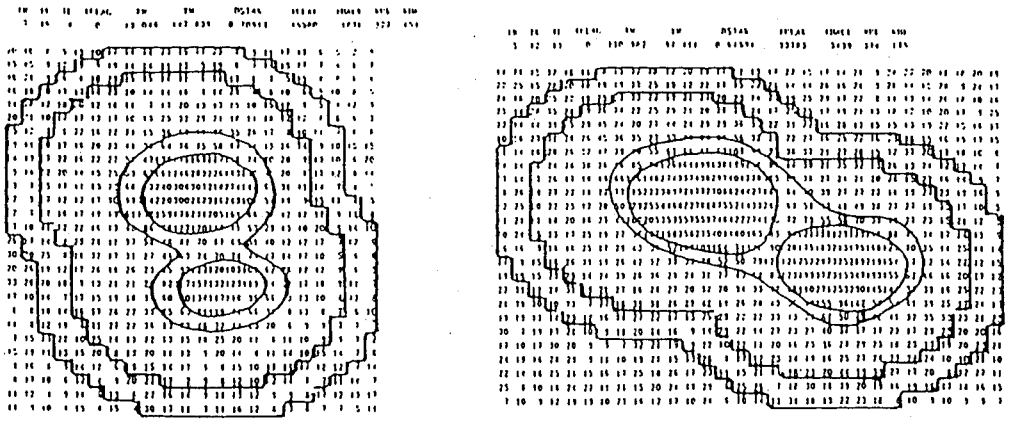


Fig. 3-2-3.2 The splitted spots recorded on Weissenberg photograph of IPD-WAS and its calculated mask. The right and left are spots recorded on the right and left of direct beam position, respectively. The notches lines are boundary lines of peak and background.

(Quoted from ＜分子性結晶の反応＞ 第５章 written by Kamiya & Iwasaki 1993)

3-2-4. Application of κ-goniometer for the automatic alignment with 3D graphics

3-2-4.1 . Introduction

For diffraction data collection of the protein crystal using area-detector, there are two alternative methods: data collection with the random orientation of the crystal and data collection after alignment of the crystal. The former has the advantage to collect independent reflections rapidly and to reduce blind region. As for rapid data collection, it is more ideal than the random orientation to estimate orientation of the crystal which allows rapid data collection. The latter method is appropriate for heavy-atom isomorphous replacement and MAD method in which photographs in the same region is taken to search for good heavy-atom derivatives, and simultaneous recording of the Friedel pair gives good anomalous difference Patterson map for analyzing new crystal structure of the protein. Moreover, alignment of the crystal is necessary for the Weissenberg photograph. The κ-goniometer has been designed as a machine which allows to adjust orientation of the crystal.

Since κ-goniometer uses polar angles to orient the crystal, the movement is complicated and difficult to understand. So that, an automatic crystal alignment system using the κ-goniometer was developed with IRIS 3D graphics. 3D graphics provides a useful and relatively inexpensive tools for visualizing complex systems. We can use 3-D graphics software not only to align crystal, but also to simulate actual movement of the κ-goniometer for learning single-crystal diffractometry. The orientation of crystal is determined by sub-program ATIX of the WELMS. The program uses two stationary photographs whose ϕ values are 90° apart.

For crystal alignment, there are three different devices used on the diffractometer:

(1) goniometer-head

In the diffractometer without goniometer, crystal spindle axis is fixed on the base of the diffractometer. The orientation of the crystal is adjusted by the goniometer-head which has two orthogonal axes as shown in Fig. (3-2-4.1). In ordinal goniometer-head, however the largest angle adjustable is as small as 15~20 degrees.

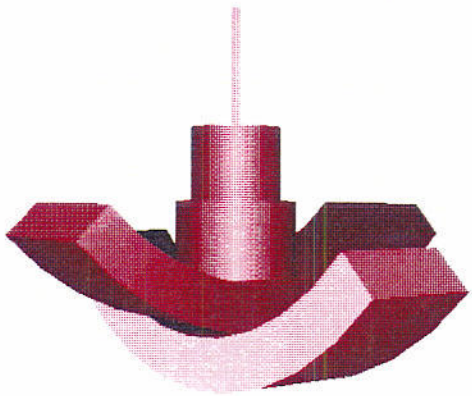
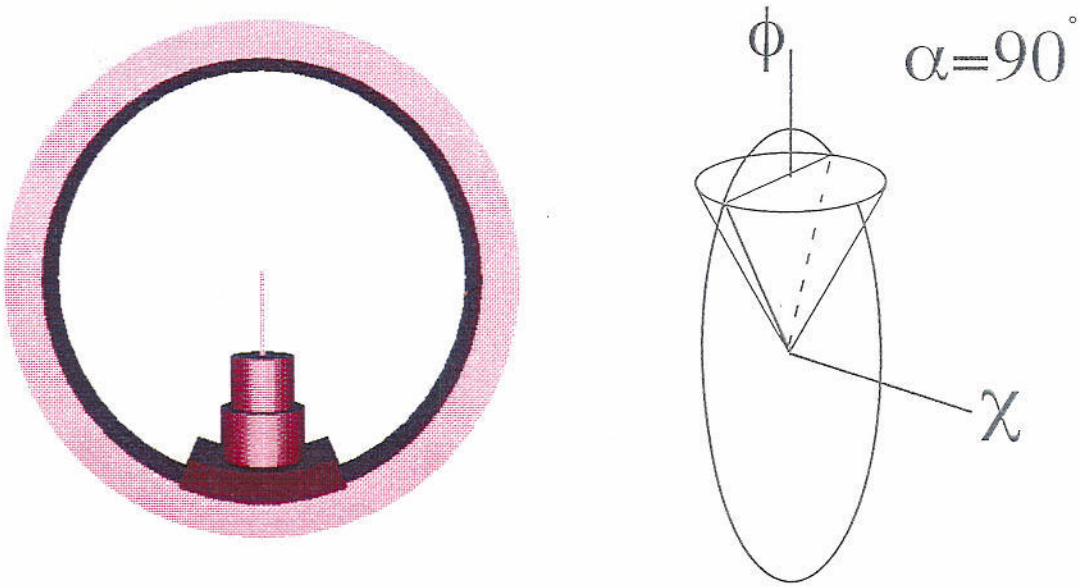


Fig. 3-2-4.1 The goniometer-head

(2) χ-circle goniometer

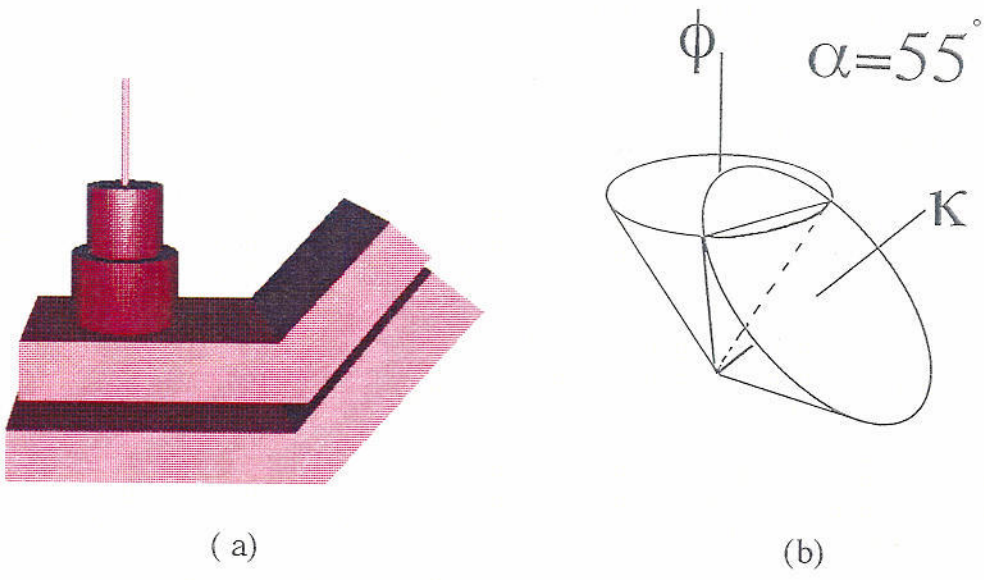
χ-circle goniometer is the most frequently used for diffraction data collection of small molecules. The crystal is mounted on the goniometer-head attached to the φ block. The φ block moves around the φ- circle and φ- χ assembly rotates as a whole about the vertical ω- axis which is the crystal oscillation axis. The crystal is oriented to any orientations by the three Eulerian angles φ, χ and ω (Fig. 3-2-4.2). However, χ-circle block is often an obstacle to mount an equipment such as cooling devices. It is especially nuisance for a system using area-detector.



(a) (b)

Fig. 3-2-4.2 The χ-circle.

(a) The χ- circle; (b) The χ-circle geometry.



(a) (b)

Fig. 3-2-4.3 The κ-goniometer.

(a) The κ-goniometer; (b) The κ-goniometer geometry.

(3) κ-goniometer

The κ-goniometer is shown in Fig. 2-4.1. It consists of ω block, κ block and ϕ block with ω, κ and ϕ rotation axes, respectively. The ω-axis is carried on base of diffractometer and crystal is rotated around this axis. The κ block is carried on ω block and κ-axis makes an angle α (In the DIP-R300, α is 55°) with ω- axis. The ϕ- axis is mounded on the κ block and carries the goniometer-head which in turn supports the crystal (see also section 2-4). All axes intersect at the center of the diffractometer. In this geometry, the orientation of the crystal is represent by three rotation angles ϕ, κ and ω (Fig. 3-2-4.3). Comparing with χ-circle, the κ-goniometer has small blind region, thus it is more appropriate to be used together with area-detector.

3-2-4.2. Strategy for crystal alignment

In DIP-R300 diffractometer, the mounted crystal has three degrees of rotational freedom. Three rotations are sufficient to bring any reflection vector to any orientation in diffractometer space. Before deriving expression for the alignment we describe coordinate system of DIP-R300.

An Cartesian laboratory coordinate system of DIP-R300 is given in Fig. 3-1-1.3. The X axis is defined as the direction of the incident X-ray beam. The Z axis is parallel to crystal rotation axis (ω), and Y axis completes a right-handed system. The ω is the angle of rotation about ω-axis, κ and ϕ are the angles made by the rotation about κ- and ϕ- axis respectively. When κ is zero, the ϕ-axis is along ω-axis. The origin is at the center of the diffractometer where all axes intersect.

For the crystal alignment, the crystal must be oriented so as to bring given reciprocal-lattice points to some orientation. Therefore, our problem of the crystal alignment is reduced to vector rotation problem: rotate given two reciprocal-

lattice points under present setting angles ω_0, κ_0 and ϕ_0. Our problems are to find ω, κ and ϕ angles to move the first vector to Z axis and second vector to the plane perpendicular to the X-ray beam.

Let $\mathbf{h}_1(h_1, k_1, l_1)$ and $\mathbf{h}_2(h_2, k_2, l_2)$ be the given two reciprocal-lattice points and **UB** is obtained from ATIX of WELMS as Eq. (3-1-1.3), using alignment photographs taken at ω_0, κ_0 and ϕ_0 angles. Two reciprocal-space vectors $\mathbf{x}_{c1}'$ and $\mathbf{x}_{c2}'$ in this setting angles are calculated as:

$$\begin{aligned} \mathbf{x}_{c1}' &= \mathbf{U}\,\mathbf{B}\,\mathbf{h}_1 \\ \mathbf{x}_{c2}' &= \mathbf{U}\,\mathbf{B}\,\mathbf{h}_2 \end{aligned} \qquad (3\text{-}2\text{-}4\text{-}1)$$

Since orientation matrix **U** was calculated under present setting angles $\omega = \omega_0$, $\kappa = \kappa_0$ and $\phi = \phi_0$, the vectors $\mathbf{x}_{c1}$ and $\mathbf{x}_{c2}$ under $\omega = \kappa = \phi = 0$ are calculated by the following equations.

$$\begin{aligned} \mathbf{x}_{c1} &= \Phi(-\phi_0)\,\mathbf{K}(-\kappa_0)\,\Omega(-\omega_0)\,\mathbf{U}\,\mathbf{B}\,\mathbf{h}_1 \\ \mathbf{x}_{c2} &= \Phi(-\phi_0)\,\mathbf{K}(-\kappa_0)\,\Omega(-\omega_0)\,\mathbf{U}\,\mathbf{B}\,\mathbf{h}_2 \end{aligned} \qquad (3\text{-}2\text{-}4.2)$$

where **K** is polar rotation matrix. In DIP-R300 coordinate system, **K** is given by

$$\mathbf{K} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos\kappa & -\sin\kappa & 0 \\ \sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix} \qquad (3\text{-}2\text{-}4.3)$$

where α (=55°) is angle between κ-axis and ω-axis. Φ and Ω are rotation matrix given by

$$\Phi = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad (3\text{-}2\text{-}4.4)$$

$$\Omega = \begin{pmatrix} \cos\omega & -\sin\omega & 0 \\ \sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad (3\text{-}2\text{-}4.5)$$

We now calculate ϕ and κ values to put the first vector $\mathbf{x}_{c1}$ to Z axis and ω value to put the second vector $\mathbf{x}_{c2}$ in the plane perpendicular to the X-ray beam. Fig. 3-2-4.4b is the view projecting traces which vector is rotated around ϕ-axis followed by κ-axis to the X-Y plane. Let Q be the position of the first vector $\mathbf{x}_{c1}$. It is our aim to put Q to position A or B by rotating ϕ_1 or ϕ_2 around ϕ-axis, then to put A or B to origin O by rotating κ_1 or κ_2 around κ-axis. From Fig 3-2-4.4a, the radius of the κ-circle is $d^*\sin\alpha$. The positions A and B are calculated as the intersection between the circle and ellipse given below.

$$x^2 + y^2 = r_o^2$$
$$\frac{x^2}{(d^*\sin\alpha)^2} + \frac{(y + d^*\sin\alpha\cos\alpha)^2}{(d^*\sin\alpha\cos\alpha)^2} = 1 \qquad (3\text{-}2\text{-}4.6)$$

where $d^* = \sqrt{x_c^2 + y_c^2 + z_c^2}$ and $r_0 = \sqrt{x_c^2 + y_c^2}$. From Eq. (3-2-4.6)

$$y = \frac{\cos\alpha}{\sin\alpha}(-d^* + |z_c|)$$
$$x = \pm\sqrt{x_c^2 + y_c^2 - y^2} \qquad (3\text{-}2\text{-}4.7)$$

So, that

$$\phi_{1,2} = \text{atan}\left(\frac{y}{x_{1,2}}\right) - \phi_o \qquad (3\text{-}2\text{-}4.8)$$

$$\kappa_{1,2} = \text{atan}\left(\frac{x_{1,2}}{\frac{y}{\cos\alpha} + d^* \sin\alpha} \right) \qquad (3\text{-}2\text{-}4.9)$$

Here underscript 1,2 corresponds to ± sign of equation (3-2-4.7). From the second vector $\mathbf{x}_{c2}$, we can obtain ω as follows (Fig. 3-2-4.4c).

$$\mathbf{x}_2 = \mathbf{K}(k_{1,2})\, \Phi(\phi_{1,2})\, \mathbf{x}_{c2} \qquad (3\text{-}2\text{-}4.10)$$

$$\omega_{1,2} = \text{atan}\left(\frac{x_{1,2}}{y} \right) \qquad (3\text{-}2\text{-}4.11)$$

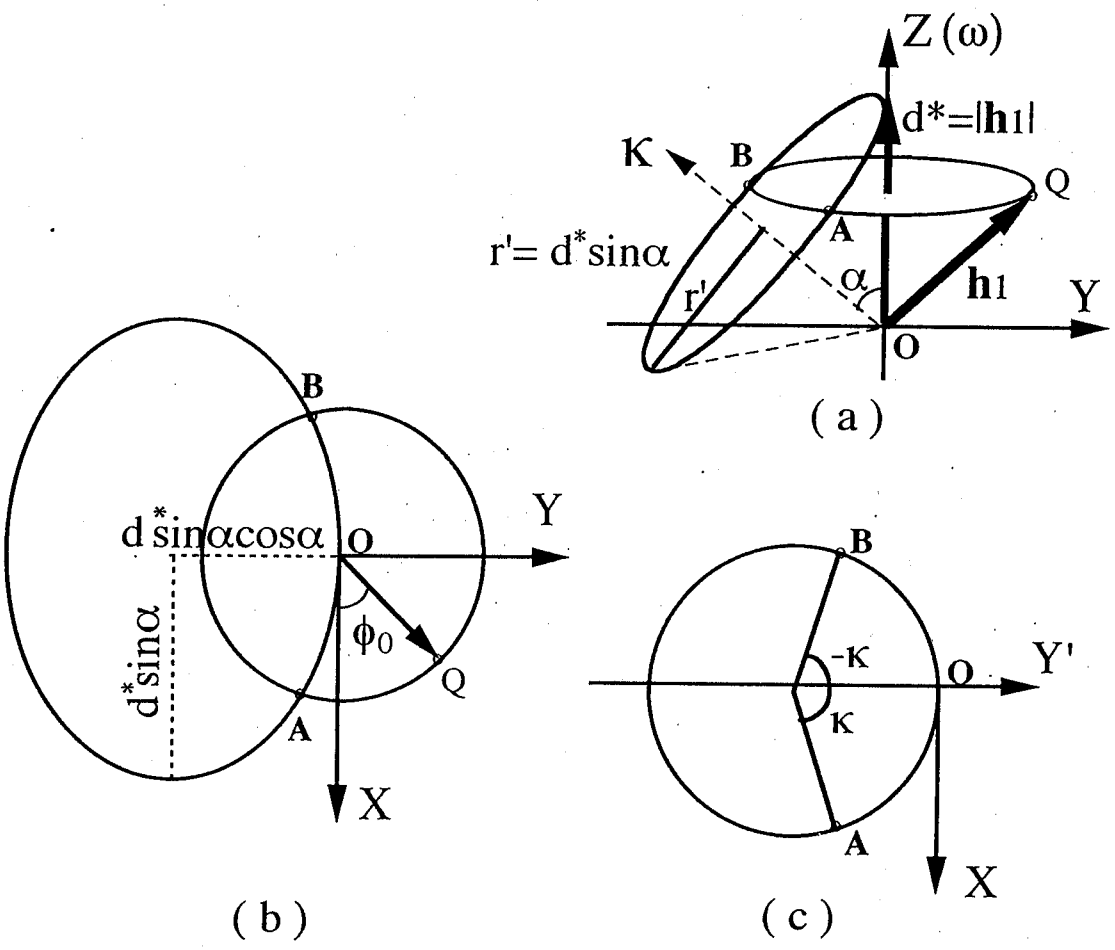


Fig. 3-2-4.4 Derivation of the angles ϕ and k from a reciprocal-space vector $\mathbf{x}_{c1}$. (a) The trace of a vector whose length is d^* and in Z axis, when it is rotated around κ-axis. (b) The projection of traces of vector OQ when it is rotated around ϕ-axis followed by κ-axis. (c) The κ-axis rotation viewed from κ-axis.

3-2-4.3. Program features

The sub-program (named DS) has been written in C-language and IRIS gl graphic library. Flow chart is shown in Fig 3-2-4.5. It was included in WELMS program system as a function. The program also includes more familiar χ-circle diffractometer and a zero-dimensional counter.

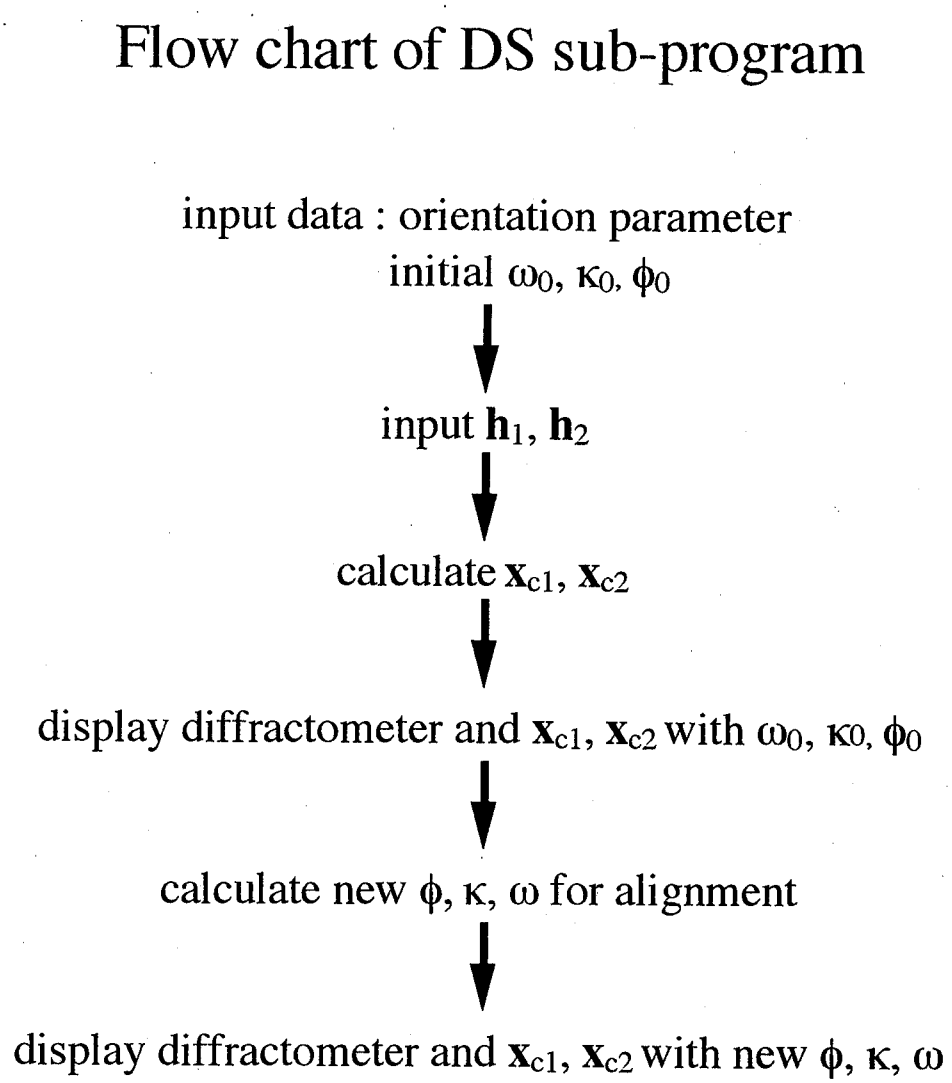


Fig. 3-2-4.5 The flow chart of DS program.

The program first displays the κ-goniometer on the screen with all the angles set to zero or the angles set to currently existing value. The shape of the κ-goniometer and detector, as shown in Fig. 3-2-4.6. The displayed model is constructed by three main functional blocks: ϕ-block, κ-block, and ω-block, each of which can be manually rotated using dials just like an actual device. The whole screen can also be rotated using the dials in order to choose the best angle to view the displayed model. The scattering vectors of any reflections can be displayed on the screen on request and move synchronously with the rotation of the axes of the diffractometer.

The crystal orientation matrix may be input in two ways. One is to input three mis-orientation angles (p1, p2, p3), the definition of which is described at section 3-2-1. The another is to get from parameters file of the WELMS. The three main functional blocks mentioned above can also be rotated by giving the angles using the keyboard. In contrast to the 5 degree step rotation using the dial, the rotational angles in this case can be given in any desired accuracy, and the model rotates smoothly to the given positions.

When the program is used for the crystal alignment, the two reflections are selected and the model rotates so as to make the first vector parallel to the ω axis and put the second vector in the plane perpendicular to the X-ray beam. This simulates the movement of the diffractometer to the starting position prior to rotating in the ω axis for the continuous data collection by the area detector, thus avoiding possible crash between κ-goniometer and other parts of the DIP-R300.

The program works on the Silicon Graphics IRIS INDIGO computer. Three kinds of input devices (mouse, keyboard and dials) are used for operating the program. Fig. 3-2-4.7 shows pattern of the result of aligning hen egg-white lysozyme crystal using this program, (a) is stationary photograph before alignment and (b) is one after alignment. The shape of the κ-goniometer after

alignment of Fig. 3-2-4.7 is shown at Fig. 3-2-4.8.

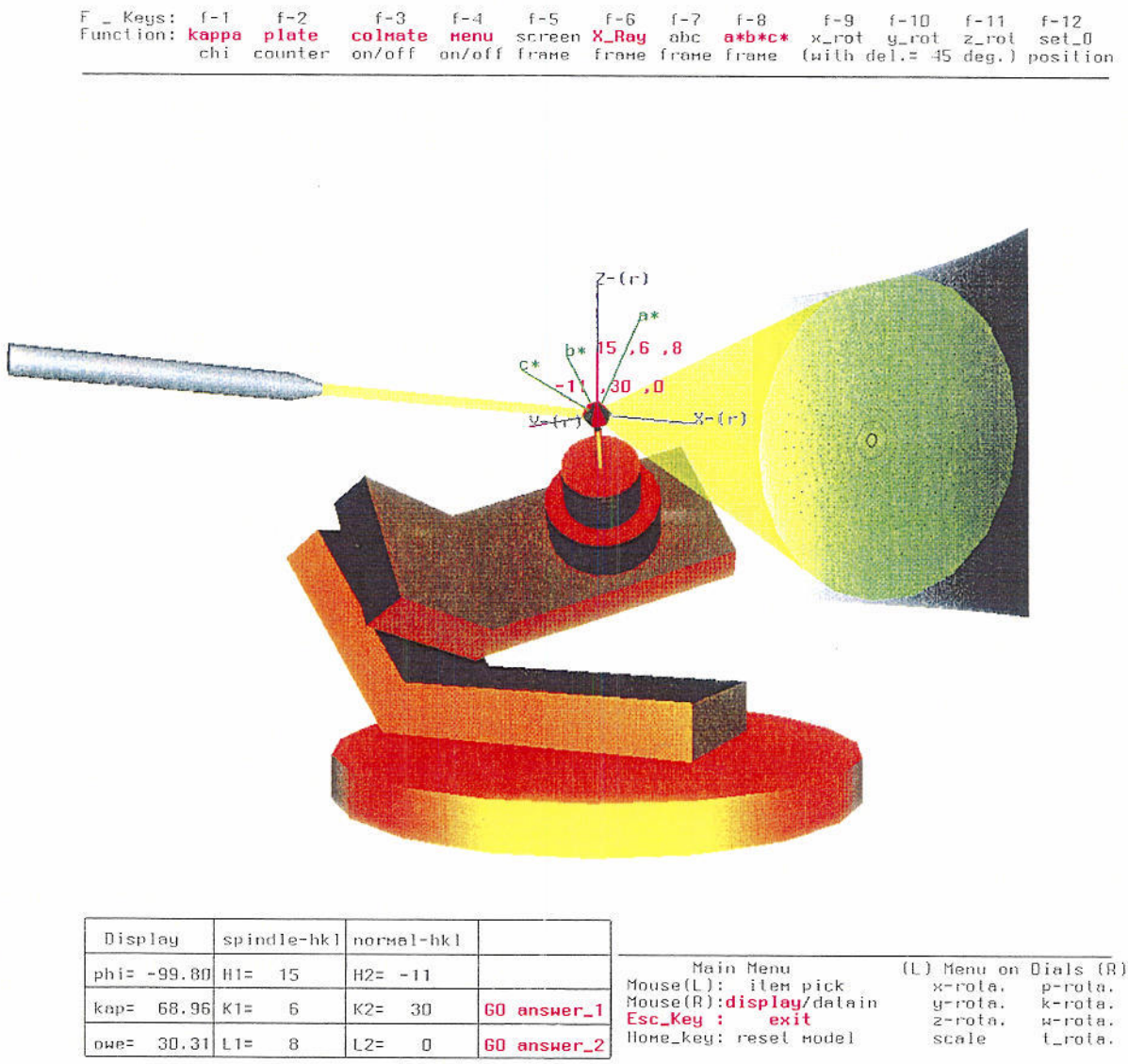


Fig. 3-2-4.6. The screen view examples of κ-goniometer with area-detector

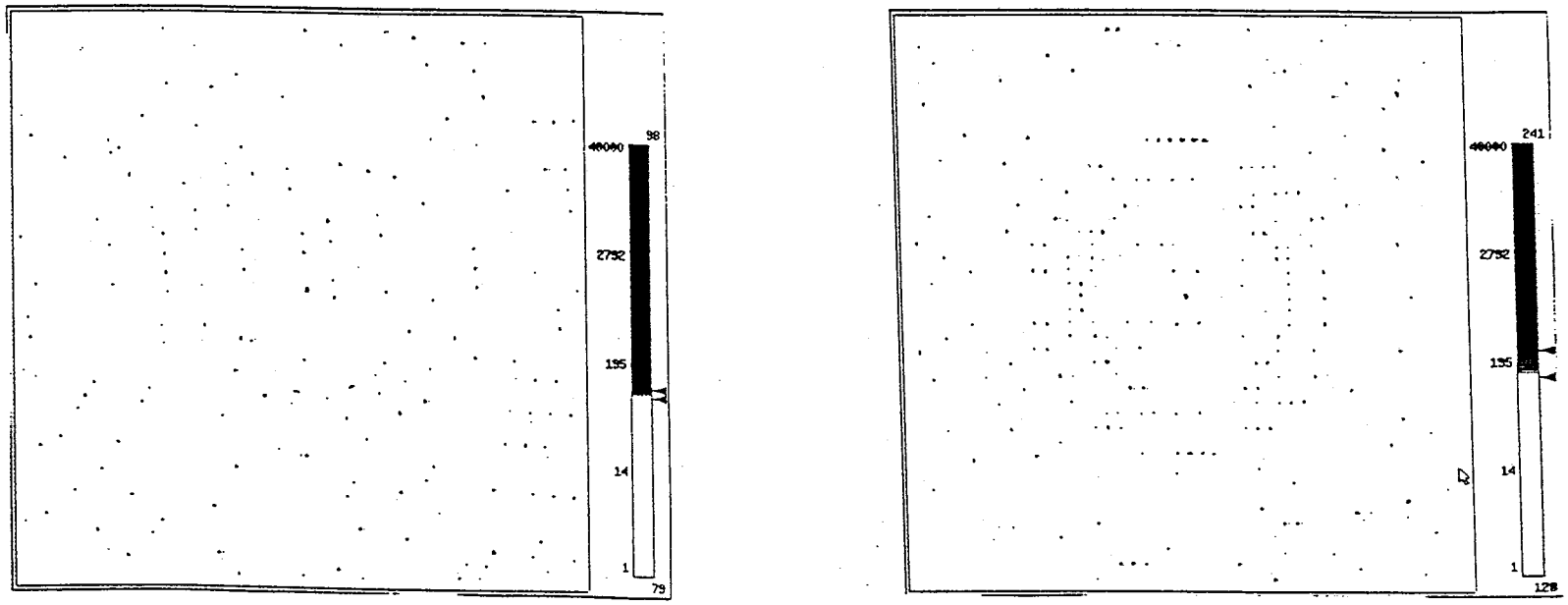


Fig. 3-2-4.7 Crystal alignment photographs taken by the DIP-R300.
(a) The photograph before alignment. (b) The photograph after alignment.
Sample: hen egg-white lysozyme. cell dimensions: a=b=79.1Å, c=37.9Å; space group: $P4_32_12$; X-ray: Cu $K\alpha$ running at 50KV, 90mA; film-crystal distance: 150mm; camera radius: 300mm; IP size: 320 x 240 mm; exposure time: 5min. Oscillation range 1°.

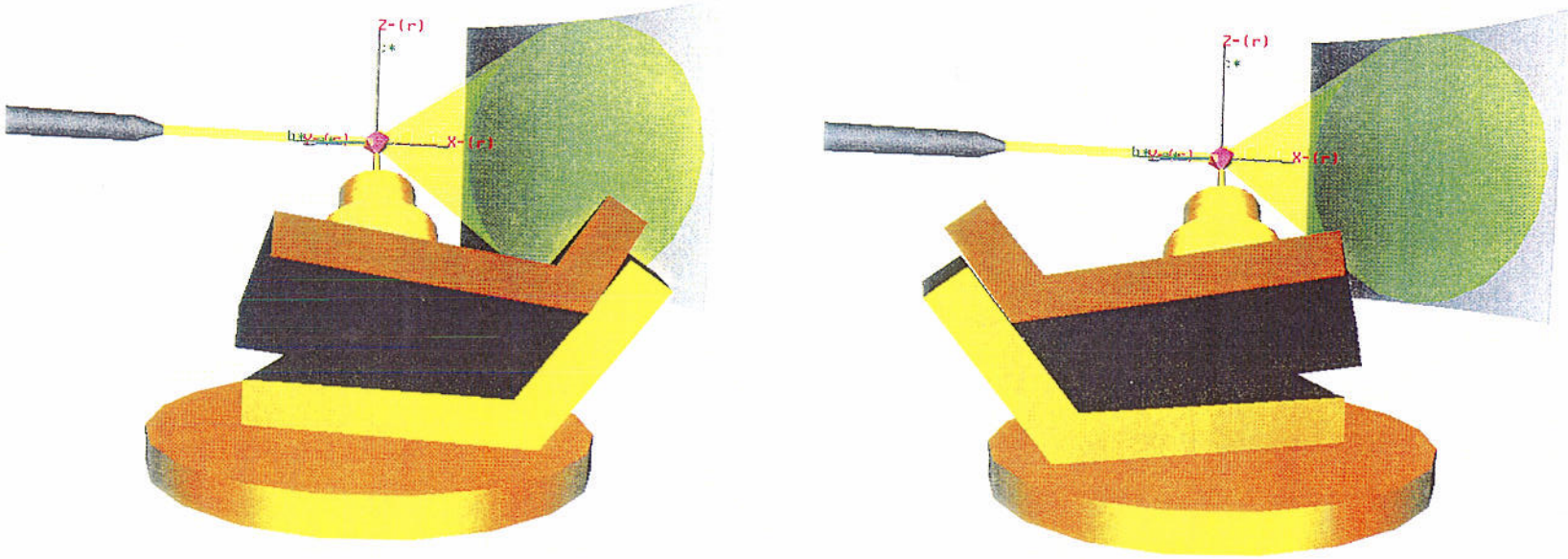


Fig. 3-2-4.8 Screen view of κ-goniometer (DIP-R300) after alignment.
Two figures correspond to two answers.

References

Amemiya, Y. & Miyahara, J. (1988). *Nature (London),* **336**, 89-90

Arndt, U.W., & Willis T.M. (1966) *Single Crystal Diffractometry.* Cambridge at the University Press

Arndt, U. W. & Wonacott, A. J. (1977). *The Rotation Method in Crystallograhy.* Amsterdam: North-holland.

Busing, W. R. & Levy, H. A. (1967). *Acta Cryst.* **22**, 457-464

Chihara, T. & Yamazaki, H. (1992). *J. Cluster Sci.* **3**, 489-497

Diamond, R., (1971). *Acta Cryst.* **A27**, 436-452

Higashi, T. (1990). *J. Appl. Cryst.* **23**, 253-257

Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 67-71

KamiyaA, N. et al., (1994). *J. Appl. Cryst.* 1994 Submitted

Kamiya, N., Iwasaki, H., (1993) ＜分子性結晶の反応＞第５章 51-60

Krause, K. L. & Phillips Jr, G. N. (1992). *J. Appl. Cryst.* **25**, 146-154

Miyahara, J., Takahashi, K., Amemiya, Y., Kamiya, N. & Satow, Y. (1986) *Nucl. Instrum. Methods* **A246**, 572-578

Reeke Jr G. N.,(1984). *J. Appl. Cryst.* **17**, 238-243

Rossmann M. G., (1979) *J. Appl. Cryst.* **12** 225-238

Sakabe, N. (1991). *Nucl. Instrum. Methods* **A303**, 448-463

Sato, M., Yamamoto, M., Imada, K., Katsube, Y., Tanaka, N., & HigashiI, T. (1992). *J. Appl. Cryst.* **25**, 348-357

Tanaka,I., & Yao, M., (1993) ＜分子性結晶の反応＞第６章 61-68

Tanaka, I., Yao, M., Suzuki, M., Hikichi, K., Matsumoto, T., Kozasa, M. & Katayama, C. (1990). *J. Appl. Cryst.* **23**, 334-339

Tucker, P. A. (1990). *Proc. of CCP4 study weekend on Accuracy and Reliability of Macromolecule Crystal Structures,*. 26-27 *Jamuary* 1990, *Daresbury Laboratory*. SERC Daresbury Laboratory, Warrington. England

Zheng, C. D., Yao, M., & Tanaka, I. (1995) *J. Appl. Cryst.* in press

Chapter 4 Application

In this chapter, X-ray structure analysis of pigeon egg-white lysozyme, dog milk lysozyme and ACC deaminase will be described.

4-1. X-ray structure analysis of Pigeon egg-white lysozyme

4-1.1. Introduction

The lysozymes constitute a group of enzymes with the common property of lytic activity towards bacterial cell walls. They are widely distributed in animal and vegetable tissues, and are present in a large amount in avian egg-white. This group of proteins has been classified into three different types (type c, type g and phage type) based on their molecular properties and protein sources (Gruetter & Weaver, et al., 1983). Pigeon egg-white lysozyme has been reported to be a type c lysozyme (Gavilanes & Gonzalez et al., 1982). It consists of a single polypeptide chain of 127 amino acid residues with a molecular weight of about 14,500. The amino acid sequence is 47% homologous with that of chicken egg-white lysozyme. The lack of immunological cross-reactivity against other type c lysozyme antisera (Prager & Wilson et al., 1974) suggests some degree of structural difference. Another important difference between pigeon lysozyme and other type c lysozymes is that pigeon lysozyme binds one molar calcium ions (Nitta & Tsuge et al., 1988). In this respect, pigeon lysozyme is more similar to α-lactalbumin, an evolutionally related protein whose enzymatic role as a specific modifier that converts galactosyltransferase to lactose synthase is totally different from that of lysozymes (Brew & Campbell, 1967; Stuart & Acharya et al., 1986). The amino acid sequence homology between pigeon lysozyme and α-lactalbumin is 37%. Thus, elucidation of the three-dimensional structure of pigeon lysozyme is important as to evolutional studies (Nitta & Sugai , 1989) on type c lysozymes

and α-lactalbumin.

4-1.2. Crystallization and Data Collection

The crystals were grown from 2.5-2.8M ammonium sulphate and 20mM ammonium acetate buffer(pH 3.7-3.9) with 1mM $CaCl_2$ by means of hanging drop vapor diffusion technique (McPherson,1982). The crystals belong to space group $P2_12_12_1$ with unit cell dimensions a=34.2Å, b=34.8Å, c=99.4Å. There are 4 molecules in the unit cell with Vm=2.05$Å^3$/Da and solvent volume fraction to be 40%.

X-ray intensity data(up to 2.7Å resolution) used in molecular replacement were collected on an automated oscillation camera system with DIP-100. Diffraction intensity data recorded on the imaging plate were integrated by the program ELMS(Tanaka & Yao et al., 1990)(Table4-1.1). The high-resolution diffraction data from four native crystals were collected with a Weissenberg camera especially designed for macromolecules using synchrotron radiation at Photon Factory(Sakabe 1991), and processed by program WEIS(Higashi, 1989). The total data-set contained 60,416 observations corresponding to 8176 unique reflections(95.5% completion) in the resolution range 100 to 2.0Å. The final merging R-factor was 10.5%.

Table 4-1.1 The condition of data collection on DIP-100.

wavelength	1.5478Å
X-ray power	50KV 250mA
crystal-IP distance	110mm
resolution limit	2.7Å
total number of IP	46
total rotation angle	115°
merging R	7.46%
unique reflections	3054 (93% completion)

4-1.3. Structure Refinements

The structure of pigeon egg-white lysozyme was determined by a molecular replacement method using atomic coordinates of the chicken lysozyme as a starting model, without solvent molecule (Yao &Tanaka et al., 1992). Refinement of the pigeon egg-white lysozyme structure was carried out in 15 rounds, each consists of model-fitting to electron density maps followed by the simulated annealing method refinement using X-PLOR(Bruenger, 1990) program (Table 4-1.2). Fitting the structure to electron density maps with coefficients of (2Fo-Fc) was performed on an Evans and Sutherland PS-390 computer graphics system on-lined to a VAX using FRODO (Jones, 1978) program package. Both electron density maps with coefficient (2Fo-Fc) and difference maps with coefficient (F_o-F_c) were calculated every round by FFT of CCP4 program package. All atoms were given an occupancy of 1.0 and initial thermal factor (B-factor) of 20.0Å^2. At the first two rounds each XPLOR model was refined to improve its stereo chemistry using the program PROLSQ (Hendrickson & Konnert, 1980). From 2nd round, all subsequent rounds of refinement were based on the merged data set collected by synchrotron radiation at Photon Factory.

In the first three rounds, only low-resolution data to 3Å were included in the refinement. Then high-resolution data were gradually added. When the 6th round was completed, solvent molecules which make reasonable hydrogen bonds with either the protein atoms or other solvent molecules were searched from the density in the (Fo-Fc)exp(iα_c) map by CCP4 program. The highest positive density peaks in the (Fo-Fc)exp(iα_c) map were located as water molecules. Then they were checked using FRODO. The final refined model consists of the protein with 81 solvent molecules. Crystallographic R-factor was 20.3% for 2Å data.

Table 4-1.1 Summary of refinement by X-PLOR.(O: exection)

cycle	1	2	3	4	5	6	7	8
resolution(Å)	10~2.7	10~2.5	10~2.3	7~2.3	7~2.0	7~2.0	7~2.0	7~2.0
reflection	F>3σ	F>3σ	F>3σ	F>3σ	F>3σ	F>3σ	F>3σ	F>3σ
waters						63	69	69
prep stage	O						O	O
slowcool	3000K	3000K	3000K	3000K	3000K	3000K	3000K	3000K
final stage	O	O	O	O	O	O	O	O
B-refinemet	O	O	O	O	O	O	O	O
R-factor(%)	23.6	24.0	24.9	23.8	25.9	22.8	23.1	22.7

cycle	9	10	11	12	13	14	15
resolution(Å)	7~2.0	7~2.0	7~2.0	7~2.0	7-2.0	7~2.0	7~2.0
reflection	F>3σ	F>3σ	F>3σ	F>3σ	F>3σ	F>3σ	F>3σ
waters	69	59	87	80	71	75	81
prep stage	O	O	O	O	O	O	O
slowcool	3000K	3000K	3000K	3000K	3000K	3000K	3000K
final stage	O	O	O	O	O	O	O
B-refinemet	O	O	O	O	O	O	O
R-factor(%)	22.5	22.3	21.8	21.5	20.9	20.7	20.3

4-1.3. Results and discussion

The refined structure of pigeon egg-white lysozyme is given in Fig. 4-1.1 with 81 waters. The r. m. s. deviations from ideality for bond, dihedral and improper angles are summarized in Table 4-1.3. The positional errors of the final structure were determined by plotting the R-factor as a function of resolution using Luzzati (1952) plot. Fig. 4-1.2 is such plot of the refined pigeon structure. It shows errors are 0.25Å.

Table 4-1.3 Summary of the deviation statistics

Refinement cycles	14
Resolution (Å)	7~2
Reflections(no.)	7968 ($>3\sigma$)
r. m. s. deviations of	
Bond-lengths(Å)	0.002
Bond-angles(deg.)	3.378
Impropers(deg.)	1.683
Dihedrals(deg.)	25.157

Ramachandran plot for final structure is shown in Fig. 4-1.3. Seven non-glycine residues have a left-handed α-helical conformation. This number is same as model structure. Among seven residues, three residues (Gln57, Asn74 and Asn77) are identical and two residues (Val21 and Tyr38) occur at the same place for pigeon and chicken lysozyme. But other two residues (Asn49 and Arg103) occur at different place and are at the surface loop regions(45~52; 100~104) which connect secondary structural units (Yao & Tanaka et al., 1992).

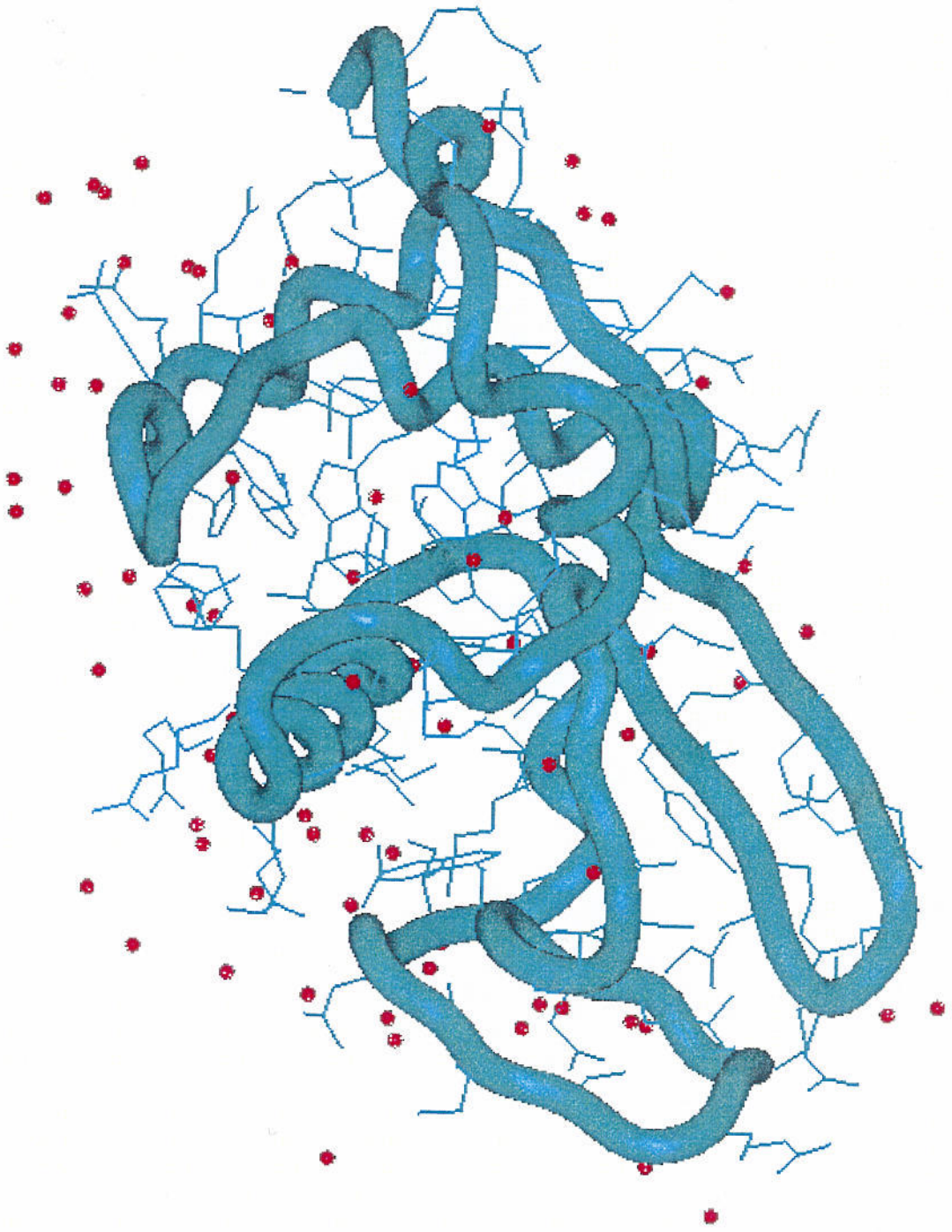


Fig. 4-1.1 The drawing of structure of refined pigeon egg-white lysozyme produced by program MOL-GRAPH. The red points are 81 water molecules. The main-chain was drawn by blue tube.

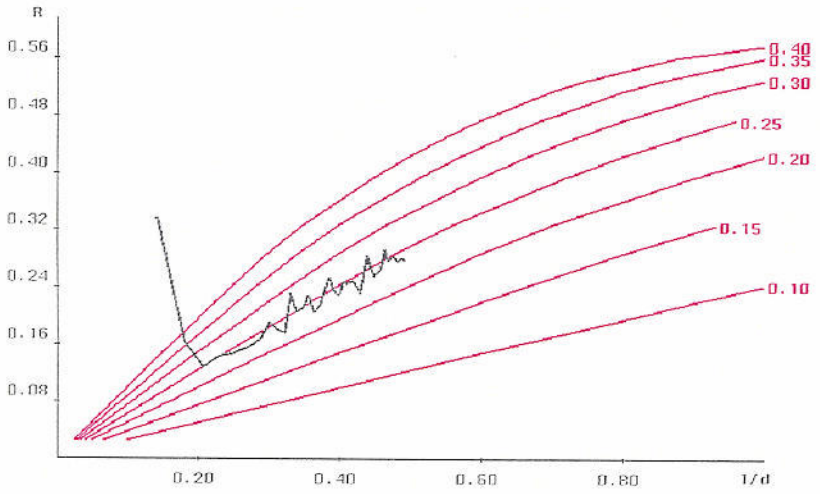


Fig. 4-1.2 Luzzati (1952) plot of the refined structure of pigeon egg-white lysozyme. The R-factor has been calculated from all reflections of $F > 3\sigma$.

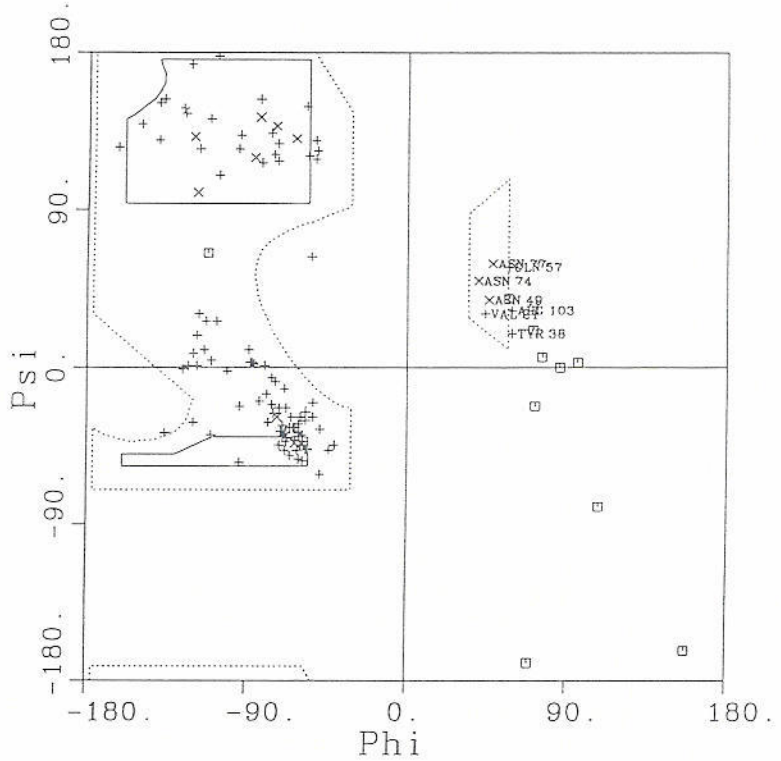


Fig. 4-1.3. Ramachandran plot of main-chain torsion angles for structure of refined pigeon egg-white lysozyme. (□)Glycine; (x)Asparagine; (+)other residues.

The temperature factors of main-chain atom of pigeon and model are plotted in Fig. 4-1.4. The average temperature factor for all main-chain atoms is higher than model. The highest B-factors are associated with the poor density on the surface of the protein.

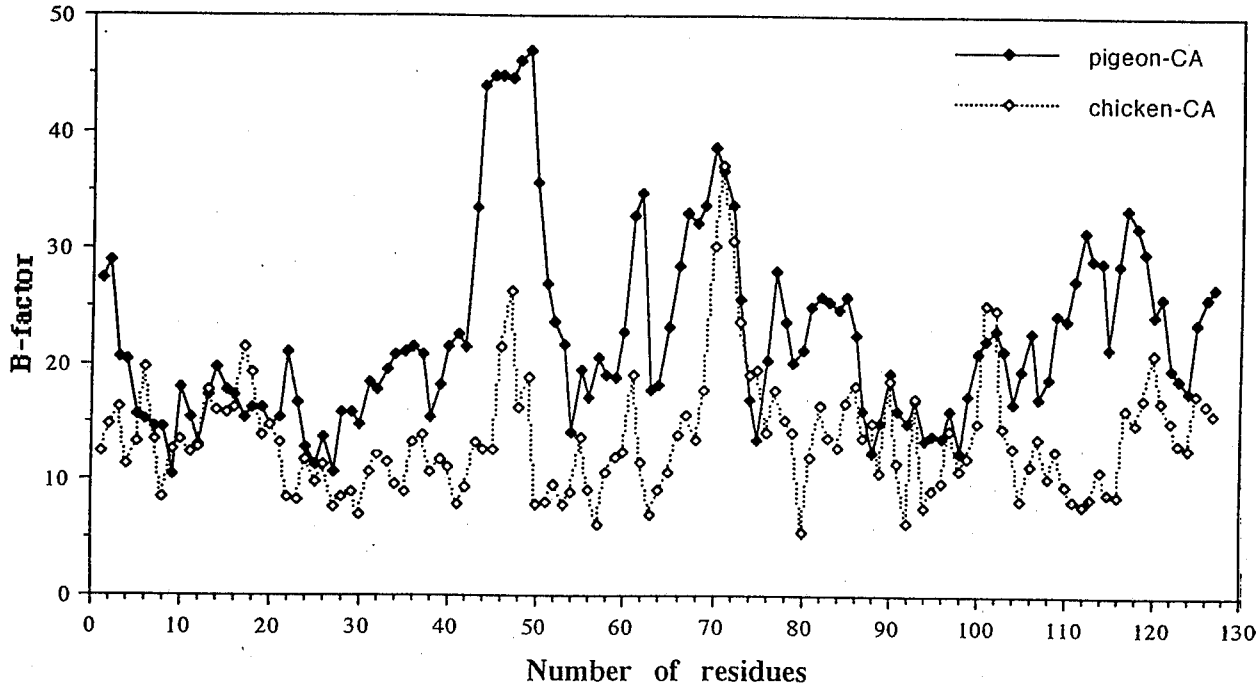


Fig. 4-1.4 Plot of average B-factor for main-chain of pigeon and chicken. The dotted line is average B-factor of chicken.

Fig. 4-1.5 shows a detailed structure of calcium binding loop which are found in the structure of the pigeon egg-white lysozyme. In the (Fo-Fc)exp(iα_c) map of pigeon egg-white lysozyme, there are three positive peaks in tne region near loop, but they are not high enough for calcium, nor bonded to the atoms for calcium binding. So they were assigned as water molecules (H_2O1, H_2O2 and H_2O3). Table 4-1.5 shows distances between waters and residues of loop region. Although in solution the protein was shown to bind a molar of calcium ions, we conclude that the protein does not bind calcium ions in crystalline state in the present condition.

Table 4-1.5 Distances(Å) of water molecules and protein atoms in the calcium binding loop.

		H_2O1	H_2O2
Lys82	O	3.184	4.387
Asp85	OD1	5.065	5.253
Asn87	O	3.099	3.106
Asp90	OD1	6.437	3.669
Asp91	OD1	2.663	4.062
H_2O3		3.600	7.112
H_2O2		4.036	

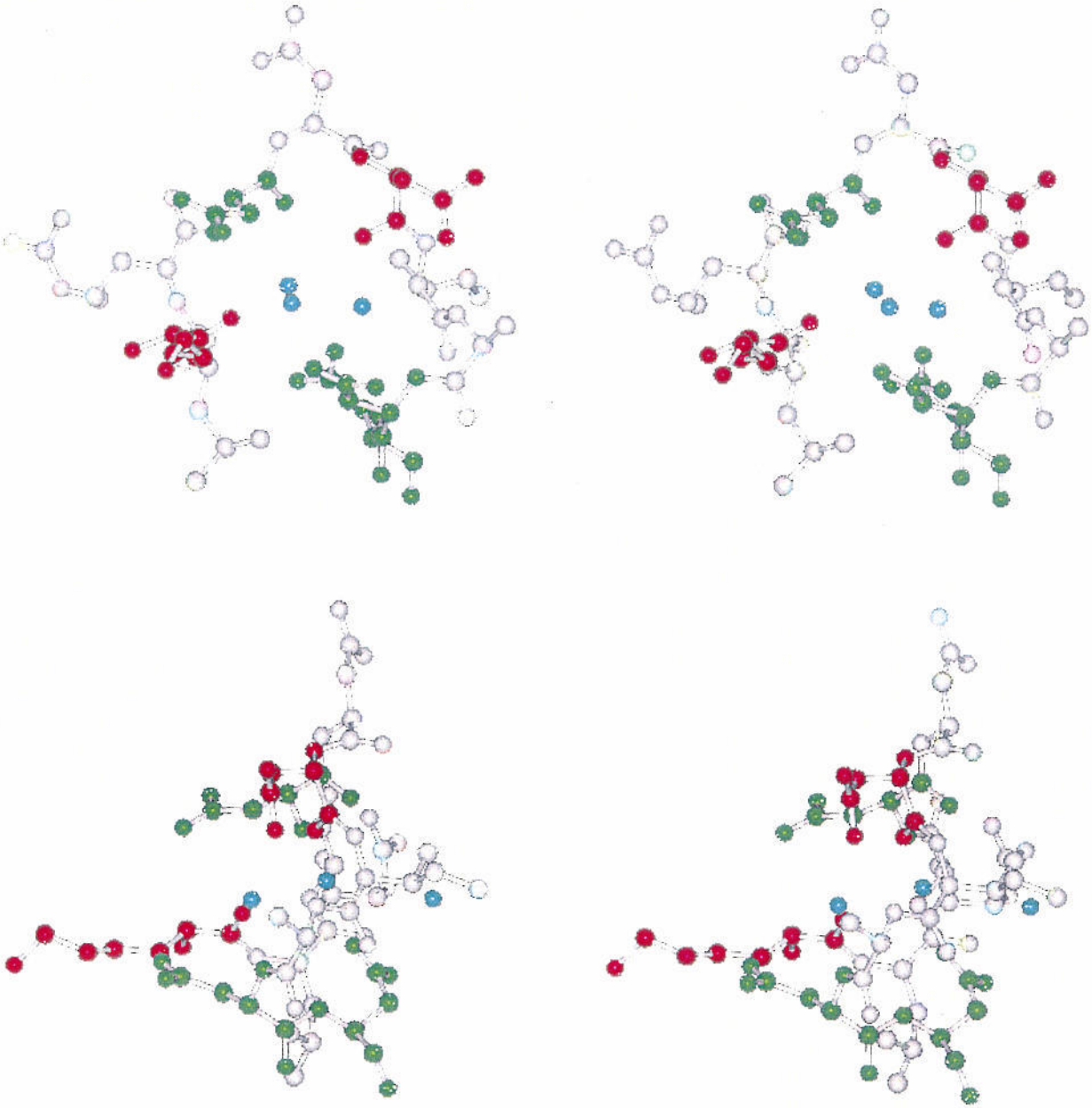


Fig. 4-1.5 Stereo drawing of refined structure of calcium binding loop of pigeon egg-white lysozyme. The red residues are Lys82 and Asn87. The green residues are Asp85, Asp90 and Asp91. The light blue points are water molecules H_2O1, H_2O2 and H_2O3.

4-2. X-ray structure analysis of Dog milk lysozyme

4-2.1. Introduction

In the previous section, a structure analysis of calcium-binding lysozyme pigeon egg-white lysozyme has be described. The structure obtained was calcium-nonbinding form. This was attributed to low pH of the crystallization condition. Here, structure analysis of another calcium-binding lysozyme from dog milk lysozyme (type c) will be described. Dog milk lysozyme consists of a single polypeptide chain of 130 amino acid residues with a molecular weight of about 14,600. It belongs to the c-type lysozyme. Dog milk lysozyme has a calcium-binding site (two backbone carbonyl groups of residues Lys-83 and Asn-88, three carboxyl groups of Asp-86, Asp-91 and Aps-92), and binds one molar of calcium ions. Thus dog milk lysozyme is similar to pigeon egg-white lysozyme and equine milk lysozyme, which are the calcium-binding c-type lysozymes. In this section, we will describe an X-ray structure analysis of dog milk lysozyme by a molecular replacement technique.

4-2.2. Crystallization and data collection

Crystallization of dog lysozyme was carried out by the hanging-drop vapor diffusion technique. The protein solution was prepared by redissolving lyophilized protein sample in 30mM phosphate (pH6.0) to a concentration of 20mg/ml. Droplets containing 5μl of protein solution were mixed with 5μl of reservoir solution and were equilibrated against 1ml reservoir solution at 18°C. Crystals of dog lysozyme were grown from a reservoir solution consisting of 2.1M ammonium sulphate and 1.5 mM $CaCl_2$ in 30mM phosphate buffer (pH5.8 ~ 6.4). The crystals reached maximum size(1.2mm x 0.6mm x 0.1mm) after one month (Fig 4-2.1). The crystal belongs to trigonal space group $P3_112$ or its

enantiomorph $P3_212$. The unit cell dimensions are a=b=31.6Å, and c=200.3Å. There are six molecules with molecular weight of 14,600 in the unit cell; Vm is calculated to be 1.97 $Å^3$/Da, and the solvent volume fraction is 38% (Matthews 1968).

X-ray intensity data up to 2.5Å resolution were collected on a diffraction imaging processor, DIP-100. Diffracted intensities recorded on imaging plate (IP) were evaluated using the ELMS program (Tanaka & Yao et al., 1990). Since the crystal has a large cell dimension in one axis and it is difficult to mount this axis as rotation axis, so high resolution data collection was severe on DIP-100. The data collection condition was shown at Table 4-2.1. A total of 16,741 measurements were merged to give 4,233 (98.6% complete) unique reflections with a merging R-factor 6.5% .

Table 4-2.1 The condition of data collection on DIP-100

wavelengh	1.5478Å
X-ray power	50KV 90mA
crystal-IP distance	130mm
resolution limit	2.5Å
rotation axis	a (or b)
total number of IP	107
total rotation angle	90°

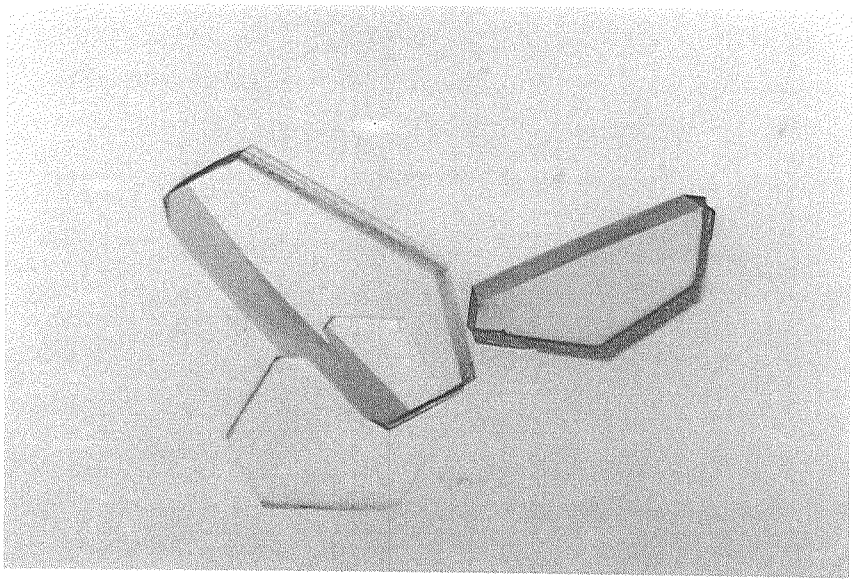


Fig 4-2.1 The crystal of dog milk lysozyme.

4-2.3. Structure analysis

The structure of dog lysozyme was determined by a molecular replacement method using program X-PLOR (Bruenger, 1990) with 10 to 4Å resolution reflections. All of the atomic coordinates of human lysozyme were used as the search model because it shows high sequence homology (50.8%) with no deletion nor insertion. The Patterson correlation refinement (PC-refinement) was performed after the conventional rotation search. The peak which gives the largest correlation coefficient (0.098) was selected and submitted to the translation search. Rigid-body refinement was used to refine the rotation and translation parameters. Considering the fact that this crystal belongs to trigonal space group

and c axis is 6 times large than others, we may assume that the molecules stack in c direction. Therefore, it is difficult to calcualte translation parameters correctly and to distingush space group $P3_112$ and $P3_212$. Figs. 4-2.3 and 4-2.4 are the result of PC, rigid-body and SA refinement of about 200 answers of translation function. A solution which gave minimum R-free value (46.2%) at the SA refinement was chosen (Fig. 4-2.4). The 66 amino acids differing from the human milk lysozyme were changed to the correct ones with model-building graphics system FRODO (Jones, 1978). Refinement converged at an R value of 23.0% (using 80% of the value suggested by X-PLOR for the weight of X-ray term) for data from 10Å to 3Å. The molecular packing of the main-chain is shown in Fig. 4-2.2. The 2.0Å structure refinement is currently in progress using program X-PLOR with the help of the FRODO graphics system.

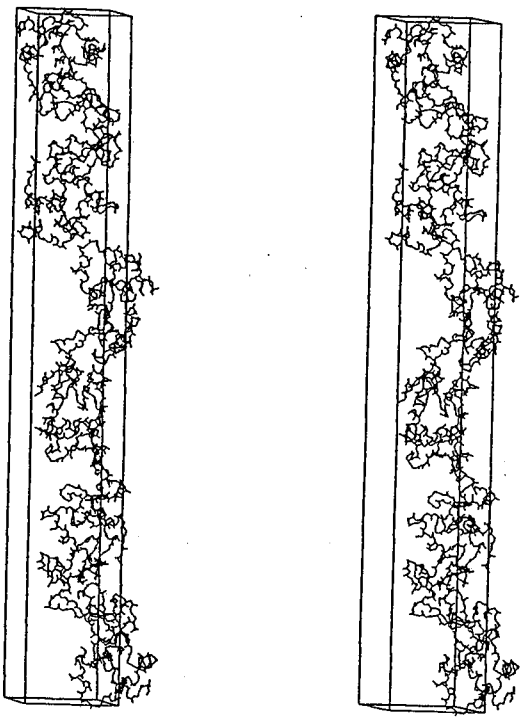


Fig. 4-2.2 Stereo drawing of the molecular packing of main-chain of dog milk lysozyme in the trigonal crystal.

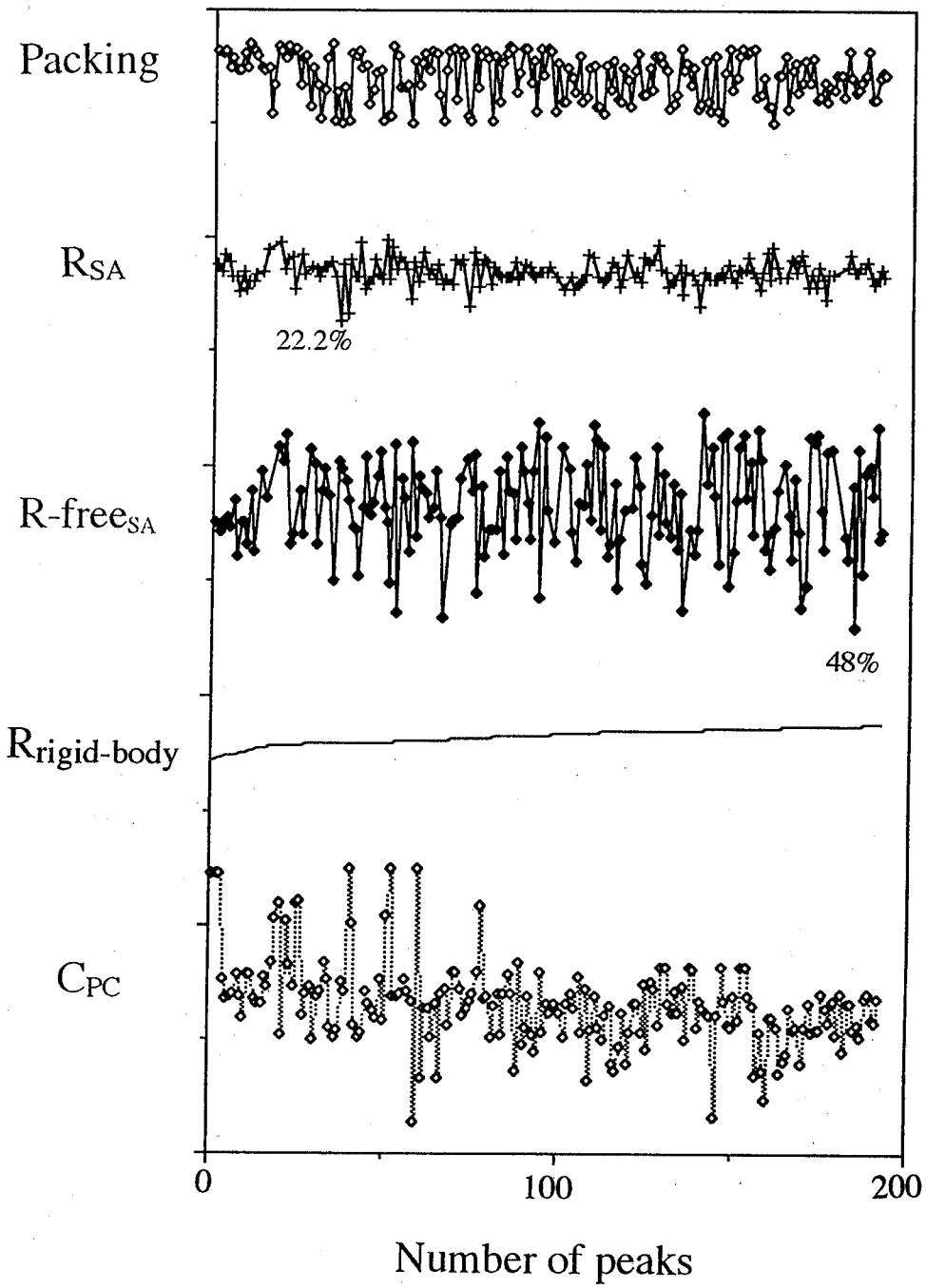


Fig. 4-2.3 The result of translation function and PC, rigid-body, SA refinement with space-group *P3112*. To calculate translation function 10~5Å, 10~4Å and 10~3Å resolution reflections were used. For PC and rigid-body refinement 10~4Å data were used. For SA refinement 10~3Å data were used.

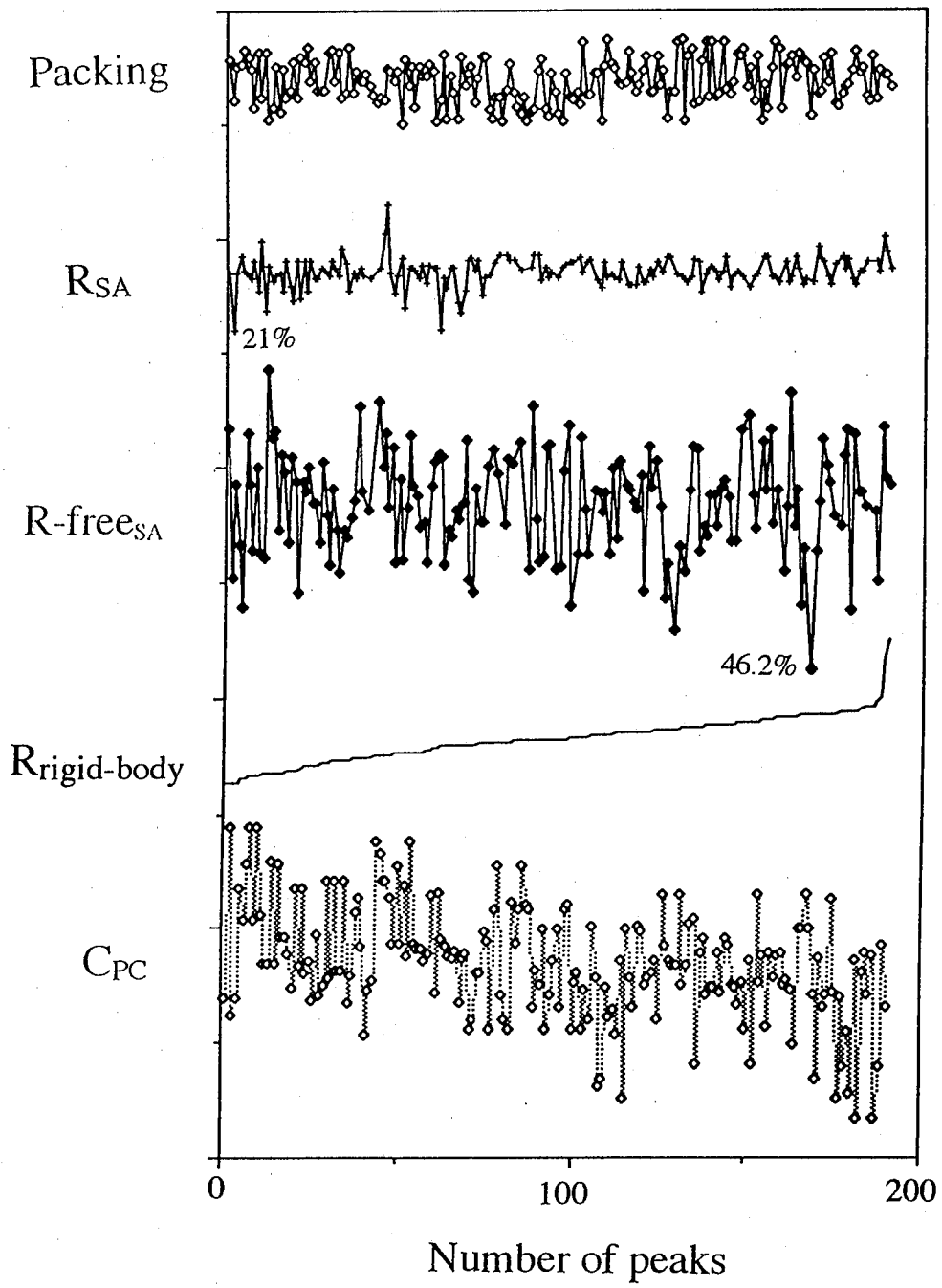


Fig. 4-2.4 The result of translation function and PC, rigid-body, SA refinement with space-group *P3₂12*. The calculation condition is given in Fig. 4-2.3.

4-3. Preliminary X-ray structure analysis of ACC Deaminase

4-3.1. Introduction

1-Aminocyclopropane-1-carboxylic acid(ACC) is a cyclic amino acid isolated from several plant tissues including pears and apples(Burroughs, 1957) (Fig. 4-3.1). It is now regarded as a key intermediate in the biosynthesis of ethylene, a plant hormone that affects diverse growing and developmental processes, including fruit ripening, leaf and flower senescence, and stress responses (Adams & Yang, 1979) (Fig. 4-3.2). Ethylene biosynthesis starts with S-adenosylation of methionine to give S-adenosylmethionine followed by a cyclopropane ring closing to form ACC, which is then oxidatively cleaved to give ethylene. Chemical inhibitors of ethylene biosynthesis can block ripening in fruit and senescence in flowers of many plant species. ACC deaminase isolated from a soil bacterium or from yeast catalyzes the cleavage of ACC to α-ketobutyrate and ammonia (Honma & Shimomura, 1978) by a cyclopropane ring opening reaction (Fig. 4-3.3). The introduction of this enzyme into plants as an inhibitor of ethylene biosynthesis by gene technology has been proved to be useful in plant physiology; it provides a way to regulate ACC levels and ethylene biosynthesis (Sheehy & Honma et al., 1991; Klee & Hayford et al., 1991).

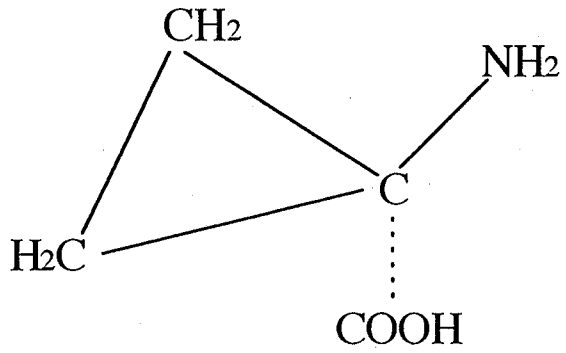


Fig. 4-3.1. 1-Aminocyclopropane-1-carboxylic acid(ACC) .

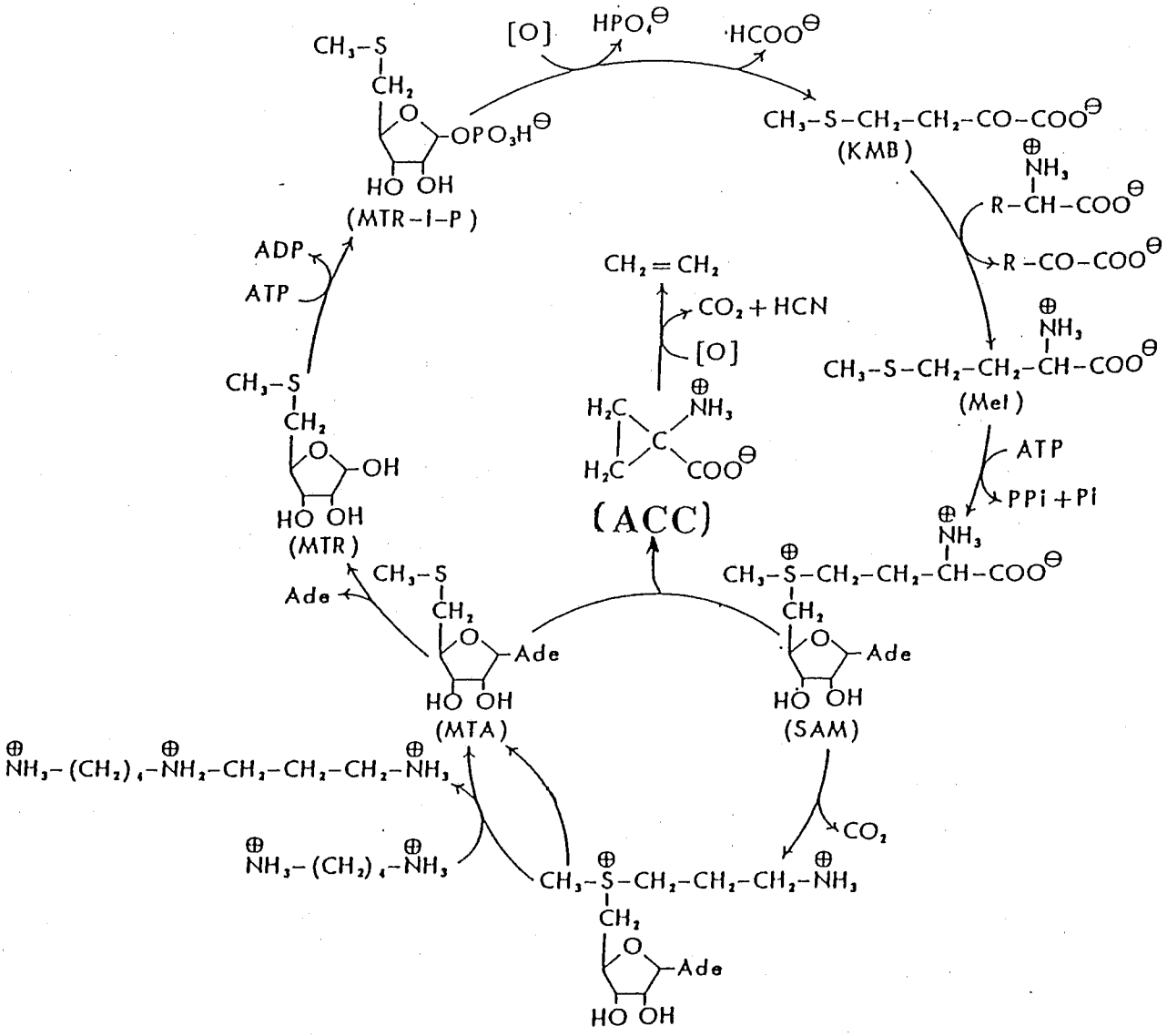


Fig. 4-3.2 Biosynthesis of ethylene in the plant.

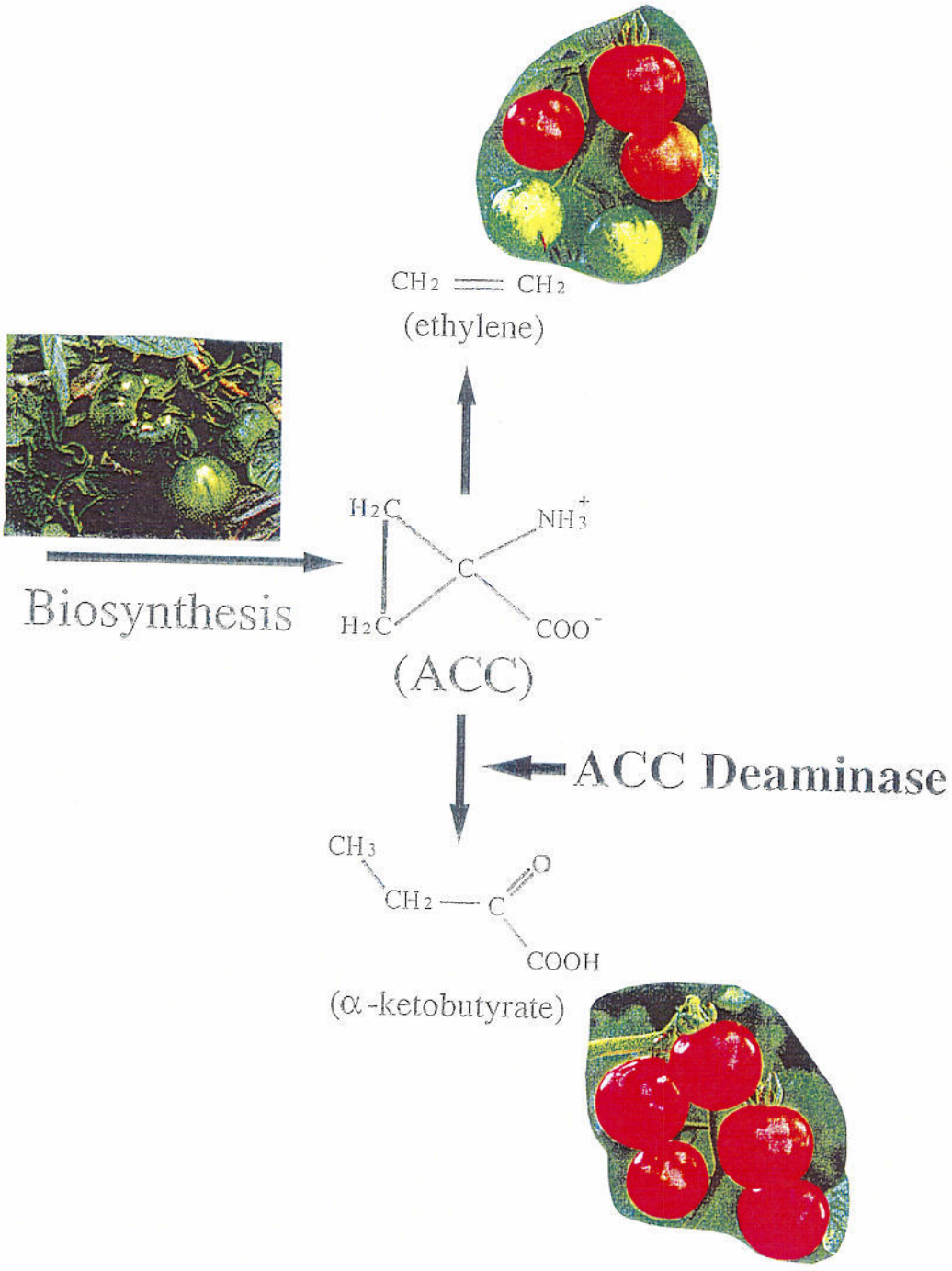


Fig. 4-3.3 Control of ethylene level by ACC deaminase

ACC deaminase purified from a bacterium *Pseudomonas sp.* has an estimated molecular weight of 110,000 and is composed of three identical subunits. Each consists of a single polypeptide chain of 338 amino acid residues with a molecular weight of 36,670 and contains tightly bound pyridoxal 5'-phosphate(PLP) as a cofactor(Honma & Shimomua 1978). The chemical modification studies in which the enzyme activity is inhibited by several sulfhydryl-modifying reagents (tetrathionate, *p*-chloromercuribenzoic acid, 5,5'-dithiobis(2-nitrobenzoic acid), N-ethylmaleimide, etc.) indicate that one of the six sulfhydryl groups in the subunit is in the active site, or close to it (Honma, 1985). ACC deaminase behaves like a PLP-dependent γ-elimination enzyme on the way to cleavage ACC to form α-ketobutyrate. Unlike the ordinal substrate for the γ-elimination , in which α-hydrogen atoms are present, the substrate ACC in this case does not contain α-hydrogen atoms(Fig. 4-3.4). It is considered that a cleavage of the C_α-C_β bond in the cyclopropane ring leads to the same result as do the eliminations of both γ- substituent and α-hydrogen in the γ-elimination of the other amino acids (Honma & Shimomura, 1978; Walsh & Pascal et al., 1981). A three dimensional structure is inevitable for a more detailed explanation of enzyme specificity.

4-3.2. Crystallization

ACC deaminase from the bacterium *Pseudomonas sp.* was purified as previously described (Honma & Shimomura, 1978). The vapor diffusion method in hanging-drops and sparse matrix sampling were used to search for a variety of crystallization conditions of ACC deaminase. The protein sample was dissolved in 50 mM K-phosphate (pH 7.5) at a concentration of 10 to 15 mg/ml. Droplets containing 5μl of protein solution and 5μl of reservoir solution were equilibrated against 1ml reservoir solution at 18°C. Crystals were grown over a pH range from

pH 6.0 to 7.6 with several precipitants: MPD(2-Methyl-2,4-pentanediol), PEG(polyethylene glycol)1500, PEG4000, PEG8000, and ammonium sulfate. The crystals grown in ammonium sulfate tended to aggregate and thus are unsuitable for structural investigation(Fig. 4-3.4(d)). The best crystals of ACC deaminase suitable for X-ray diffraction experiments were obtained at pH 6.0 to 6.4 with 19% to 25% MPD. The crystals appeared in a week and grew up to 0.8mm x 0.3mm x 0.3mm within one month(Fig. 4-3.4(a)). Crystals grown in PEG also have enough size for diffraction studies(Fig. 4-3.4(b),(c)). However, their cell dimensions are similar to the crystals grown in MPD and the crystals diffract poorly, thus are less suitable for X-ray diffraction studies.

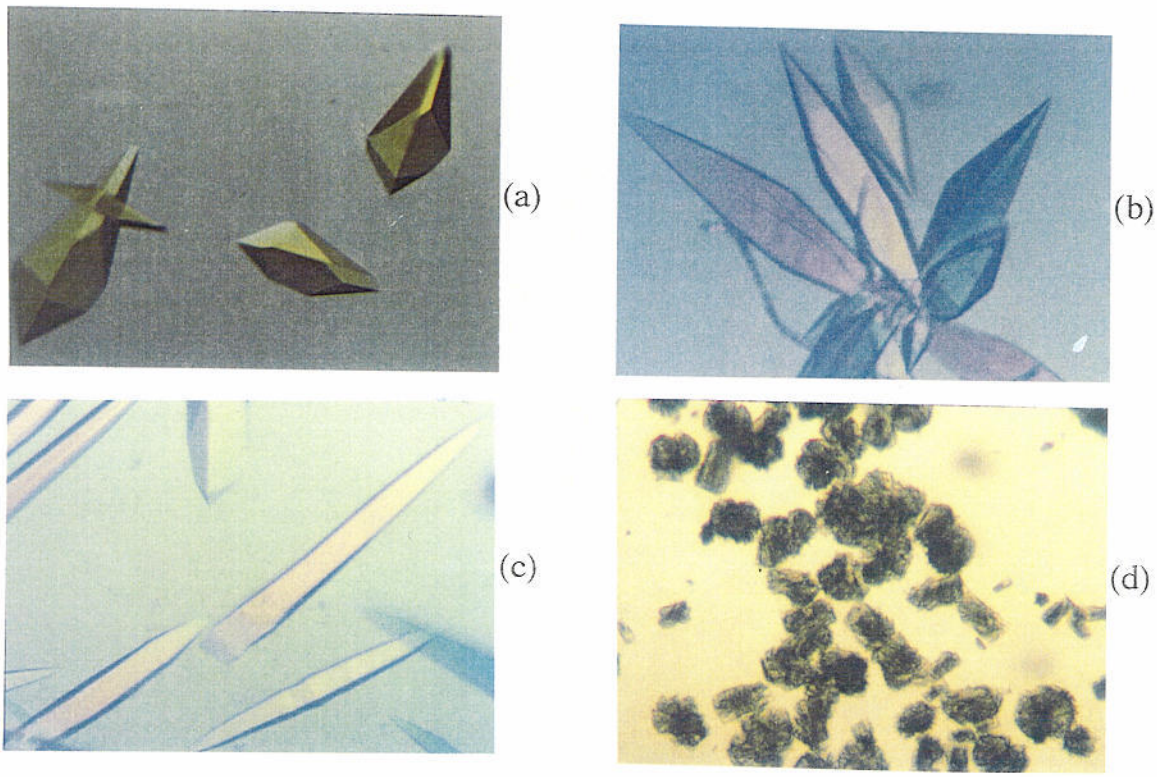


Fig. 4-3.4 Crystals of ACC deaminase. (a) Crystals grown with 21% ~ 23% MPD(pH6.0~6.6). (d) Crystals grown with 20% PEG4000, 10% 2-Propanol, 0.1M Hepes. (c) Crystals grown with 15% PEG1500, 0.2Mg formate. (d) Crystal grown with 0.4M ammonium sulfate(pH 7.4~7.6)

Since the crystal has a very large cell dimension in one axis, it was difficult to determine cell dimensions by an ordinal precession photograph. Crystallographic parameters were determined using low resolution reflections collected by a MAC Science diffraction image processor, DIP-100 (CuKα radiation from a rotating-anode X-ray generator operated at 50 kV and 90 mA) with the ELMS software package(Tanaka & Yao et al., 1990). The unit cell dimensions are a=70.0 Å, b=70.0 Å, and c=355.0 Å. Though the crystals belong to the orthorhombic space group $P2_12_12_1$, the systematic absence of 00l reflections for $l \neq 4n$ was observed at low resolution up to 20 Å, which suggests a pseudo-tetragonal arrangement of molecules in the unit cell. Crystal and solvent densities were 1.239 g/cm^3 and 1.059 g/cm^3, respectively. If it is assumed that an asymmetric unit contains two trimers of a molecular weight of 110,000, then Vm was calculated to be 1.99 Å^3/Da; the solvent volume fraction is 39%(Matthews 1968). This is consistent with the value (38%) obtained from density measurements (see Appendix B).

4-3.3. Data collection and molecular packing by self-rotation function

X-ray diffraction data from two native crystals were collected to 3.5 Å resolution by a Weissenberg camera (radius = 429.7mm) designed for macromolecules using synchrotron radiation at the Photon Factory (Sakabe 1991) and processed by the WEIS software (Higashi 1989) (UNIX version) on SGI workstation. A total of 93,237 observations were merged to give 20,375 unique reflections with a completeness of 88.6% (between 50 to 3.5 Å), and the average agreement of the intensities of the symmetry-related reflection (the merging R-factor) is 8.83%. The self-rotation function was calculated by PROTEIN (Steigemann 1974) using data in various resolution ranges and different values for the Patterson integration radius. Spherical polar angles were defined in an orthorhombic lattice as follows: Ψ, inclination versus b-axis; Φ, azimuthal angle

between the a-axis and the projection of the rotation axis on the a-c plane. The non-origin peaks corresponding to non-crystallographic axes occurred on the K=90° (four-fold), K=120° (three-fold) and K=180° (two-fold) sections, which were unaffected by the resolution range of data or integration radius used. On the K=120° section, there are two same height peaks at Ψ=35° , Φ=90° and Ψ=90° , Φ=35°, that correspond to two three-fold axes in the a-c plane and b-c plane related by a 90° rotation about an axis parallel to the crystallographic c-axis (Fig. 4-3.7a). At Ψ=90°, Φ=90° and K=90° , the peak occurs at a relative height of 80% that of the crystallographic two-fold axes; this peak corresponds to a four-fold rotation axis along the direction of the c-axis (Fig. 4-3.7b). On the K=180° section there is a peak at Ψ=45°, Φ=0° (Fig. 4-3.6), whose height is almost the same as the peak on the K=90° section. This arrangement of four-, two- and three-fold axes suggests a pseudo-tetragonal packing of trimer molecules, each of which contains C_3 symmetry, in the crystal.

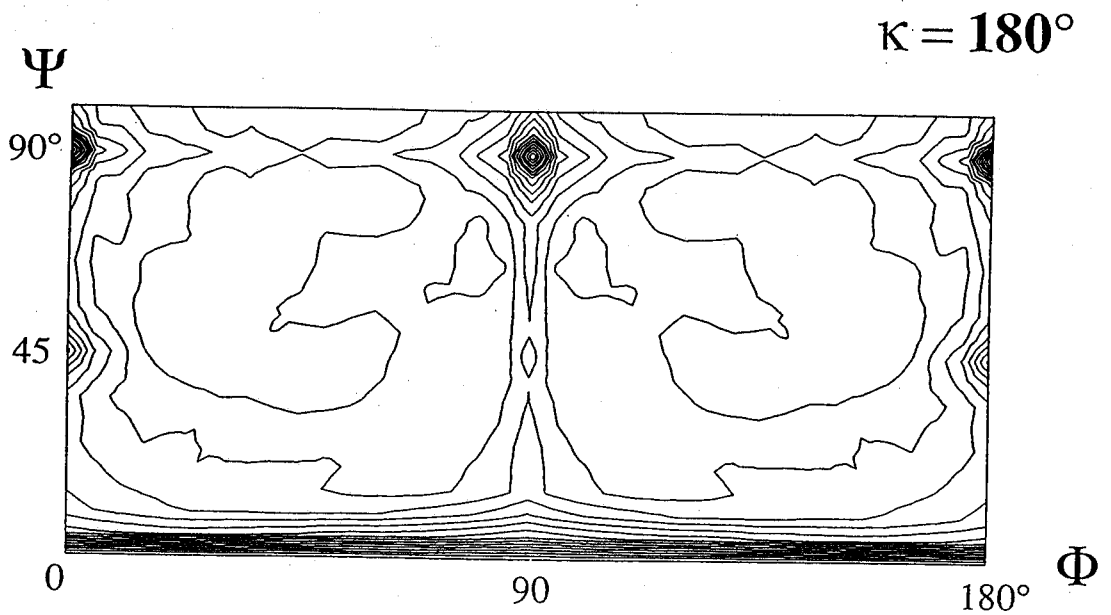


Fig. 4-3.6 The self rotation function calculated in a polar coordinate system with data in the range 10 Å - 4 Å and integrated within a radius of 20 Å. The plot of section K=180° indicates a non-crystallographic two-fold axis on the a-c plane and b-c plane.

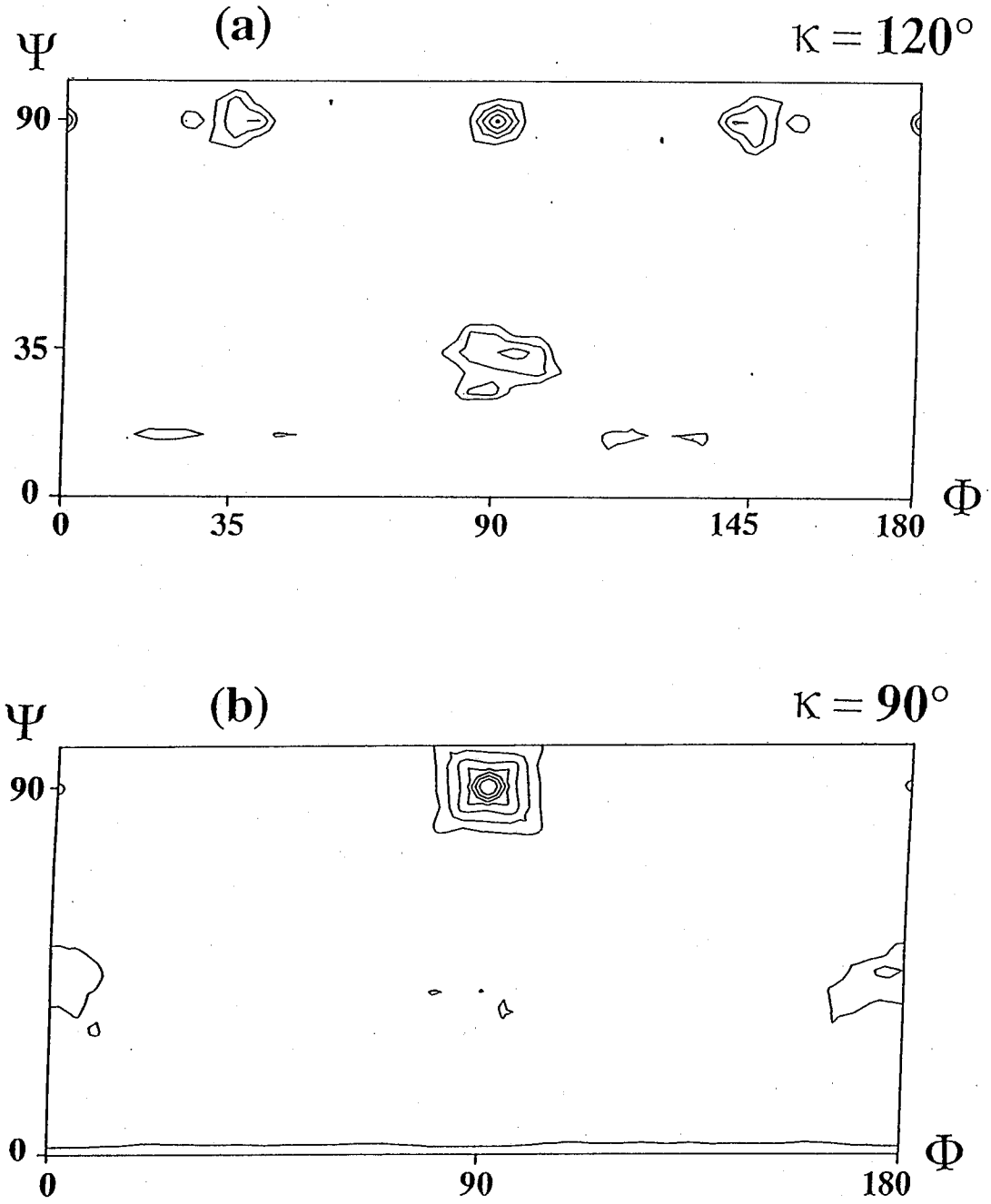


Fig. 4-3.7 The self rotation function calculated in a polar coordinate system with data in the range 10 Å - 4 Å and integrated within a radius of 20 Å. (a)The plot of section K=120° provides evidence for two local three-fold axes related by a 90° rotation about an axis parallel to the c-axis. (b) The plot of section K=90° indicates a non-crystallographic four-fold axis parallel to the c-axis.

4-3.4. Search of the heavy-atom derivatives

Heavy-atom derivatives of ACC deaminase were prepared by soaking. Approximately 12 different ionic heavy-atoms compounds with different concentration were tried and soaking time. The crystal is not suitable for soaking: K_2PtCl_4 gave a distortion of crystal structure and Laue group changed to 4/mmm, uranium gave a super-lattice (a=140.0Å b=140.0Å c=335.5Å) for a long-time in the ammonumin acetate buffer. Although ACC deaminase contains six free cysteines, none of the reactive mercurials did produce a useful derivative. The six heavy-atom derivatives with three ionic compounds provided changes on the X-ray diffraction pattern and diffraction data were collected on the Weissenberg camera for macromolecule (radius = 429.7mm) using synchrotron radiation in Photon Factory (Sakabe, 1991) . The data were processed using program package WEIS (Higashi, 1989) on SGI workstation and derivative diffraction data were scaled to the native using ANSC program. The results of data processing are given in Table 4-3.1, K_2OsCl_6-2 derivative gave the best change in diffraction. The search for heavy-atom positions is in progress from Patterson maps calculated using isomorphous differences and anomalous differences by PROTEIN program (Steigemann ,1974).

Table 4-3.1. Summary of data processing

	native	K_2OsCl_6-1	K_2OsCl_6-2	$K_3UO_2F_5$-1	$K_3UO_2F_5$-2	$K_2Pt(NO_2)_4$
conditions		0.5mM x 24h	0.25mM x 6day	1mM x 27h	1mM x 27h	0.1mM x 48h
resolution(Å)	3.5	3.5	3.5	3.5	4.0	4.0
wave length(Å)	1.00	1.00	1.00	1.00	1.488	1.00
unique reflections	19916	34147	35033	25848	19740	21069
merge-R(%)	6.90	7.10	8.95	7.82	6.81	8.66
Kemp		4.65	4.05	9.17	3.438	5.95
RMSAD		69.52	94.55	54.873	88.68	96.16
MIFD (%)		16.37	21.76	28.29	15.00	28.01
MID		112.8	148.2	202.6	102	209.8

Kemp : overall empirical K anomalous
RMSAD : root mean square anomalous difference
MFID : mean fractional isomorphous difference
MID : mean isomorphous difference

References

Adams, D. O., & Yang, S. F. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 170-174

Brew K., & Campbell P. N. (1967) *Biochem. J.* **102**, 258-269

Bruenger, A.T. (1990) *X-PLOR Manual, Version 2.1*, Yale Univ., New Haven, U.S.A

Burroughs, L. F. (1957) *Nature* **179**, 360-361

Gavilanes J. G., Gonzalez de Buitrago G., Martinez del Pozo A., Perez-Castells R., & Podriguez R. (1982) *Int. J. Peptide Prot. Res.* **20**, 238-245

Gruetter M. G., Weaver L.H., & Matthews B.W. (1983) *Nature* **303**, 828-830

Hendrickson W.A. and Konnert J.H. (1980) *Computing in Crystallography,* R.Diamond, S.Ramaseshan and K.Venkatesan, eds, The Indian Academy of Science, Bangalore, 13.01-13.23

Higashi, T. (1989) *J. Appl. Cryst.* **22**, 9-18

Honma, M. (1985) *Agric. Biol. Chem.* **49**, 567-571

Honma, M., & Shimomura, T. (1978) *Agric. Biol. Chem.* **42**, 1825-1831

Jones, T.A. (1978) *J. Appl. Cryst.* **11**, 268-272

Klee, H. J., Hayford, M. B., Kretzmer, K. A., Barry, G. F., and Kishore, G. M., (1991) *The Plant Cell* **3**, 1187-1193

McPherson A. (1982) *The Preparation and Analysis of Protein Crystals,* pp.82-159, John Wiley & Sons, New York

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491-497

Nitta K. and Sugai S. (1989) *Eur. J. Biochem.* **182**, 111-118

Nitta K., Tsuge H., Shimazaki K., & Sugai S. (1988) *Biol. Chem. Hoppe-Seyler* **369**, 671-675

Prager E. M., Wilson A. C., & Arnheim N. (1974) *J. Biol. Chem.* **249**, 7295-7297

Sakabe, N. (1991) *Nucl. Instrum. Methods* **303**, 448-463

Sheehy, R. E., Honma, M., Yamada, M., Sasaki, T., Martineau, B., & Hiatt, W. R. (1991) *J. Bacteriol.* **173**, 5260-5265

Steigemann, W. (1974) Ph.D. Thesis. Technische Univ. Muenchen, Germany

Stuart D. I., Acharya K. R., Walker N. P. C., Smith S. G., Lewis M., & Phillips D. C. (1986) *Nature* **324**, 84-87

Tanaka, I., Yao, M., Suzuki, M., Hikichi, K., Matsumoto, T., Kozasa, M., & Katayama, C. (1990) *J. Appl. Cryst.* **23**, 334-339

Walsh, C., Pascal, R. A., Jr., Johnston, M., Raines, R., Dikshit, D. Krantz, A., & Honma, M., (1981) *Biochemistry* **20**, 7509-7519

Appendix A

Derivation of the relation ships between residual and eigenvalue.

In the Eq. (3-2-2.5), the residual ε was omitted. A complete equations is given by

$$\mathbf{t} = \mathbf{A}\delta\mathbf{q} + \boldsymbol{\varepsilon} \tag{A1}$$

where, $\boldsymbol{\varepsilon}$ is residual vector. The residual $\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$ is minimized when $\boldsymbol{\varepsilon}$ is orthogonal to every column of **A**:

$$\mathbf{A}^T\boldsymbol{\varepsilon} = 0 \tag{A2}$$

Let $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}_0$ when the solution of Eq. (A1) is $\delta\mathbf{q}_0$, then at some value $\delta\mathbf{q}$ in $\delta\mathbf{q}$ space near $\delta\mathbf{q}_0$, we may write

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}_o + \delta\boldsymbol{\varepsilon} \qquad \delta\mathbf{q} = \delta\mathbf{q}_0 + \delta(\delta\mathbf{q}) \tag{A3}$$

and from Eq. (A1)

$$\delta\boldsymbol{\varepsilon} = -\mathbf{A}\delta(\delta\mathbf{q}) \tag{A4}$$

From Eqs. (A2) and (A4), the residual associated with the value $\delta\mathbf{q}$ is

$$\begin{aligned}
\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} &= (\boldsymbol{\varepsilon}_o + \delta\boldsymbol{\varepsilon})^T(\boldsymbol{\varepsilon}_o + \delta\boldsymbol{\varepsilon}) \\
&= \boldsymbol{\varepsilon}_o^T\boldsymbol{\varepsilon}_o - \delta(\delta\mathbf{q})^T\mathbf{A}^T\boldsymbol{\varepsilon} - \delta(\delta\mathbf{q})^T\mathbf{A}^T\boldsymbol{\varepsilon}_o + \delta\boldsymbol{\varepsilon}_o^T\delta\boldsymbol{\varepsilon} \\
&= \boldsymbol{\varepsilon}_o^T\boldsymbol{\varepsilon}_o + \delta(\delta\mathbf{q})^T\mathbf{A}^T\mathbf{A}\,\delta(\delta\mathbf{q}) \\
&= \boldsymbol{\varepsilon}_o^T\boldsymbol{\varepsilon}_o + \delta(\delta\mathbf{q})^T\mathbf{N}\,\delta(\delta\mathbf{q})
\end{aligned} \tag{A5}$$

Here, $\mathbf{N} = \mathbf{A}^T \mathbf{A}$. and in $\delta\mathbf{q}$ space

$$\varepsilon^T\varepsilon - \varepsilon_o{}^T\varepsilon_o = \delta(\delta\mathbf{q})^T \mathbf{N}\, \delta(\delta\mathbf{q}) \qquad \text{(A6)}$$

is a quantity of changed residual when $\delta\mathbf{q}$ is shifted by $\delta(\delta\mathbf{q})$ from $\delta\mathbf{q}_0$. Using eigenvectors $\mathbf{V}$ of the matrix $\mathbf{N}$, $\delta\mu = \mathbf{V}^T\delta(\delta\mathbf{q})$ and Eq. (3-2-2.12), Eq. (A6) reduces to

$$\varepsilon^T\varepsilon - \varepsilon_o{}^T\varepsilon_o = \delta\mu^T \Lambda\, \delta\mu = \sum_i^n \lambda_i\, \delta\mu_i^2 \qquad \text{(A7)}$$

Here, μ_i named eigenshift are components of $\delta(\delta\mathbf{q})$ in the directions of eigenvectors of $\mathbf{A}^T\mathbf{A}$. Evidently the contribution of the ith eigenshift μ_i to decrement in residual is $\lambda_i\mu_i$. So, $\varepsilon^T\varepsilon - \varepsilon_o{}^T\varepsilon_o$ of the change in the residual during a cycle of refinement is proportional to large eigenvalue λ_i. In conclusion, eigenvector υ_i associated with large λ_i gives the largest decrease in residual for the smallest changes in the parameters.

Appendix B

Calculation of molecular number of ACC deaminase in unit cell from densities.

Let define constants as:

V	:	unit cell volume.
Vp	:	unit cell volume occupied by protein
Vs	:	unit cell volume occupied by solvent
Dc	:	crystal density
Ds	:	solvent density
Dp	:	density of protein in crystal

We assume $Dp = (1/0.74)\ g/cm^3$.

Crystal and solvent density measured were:

$$Dc = 1.239\ g/cm^3$$

$$Ds = 1.059\ g/cm^3$$

from $V = Vp + Vs$ and $Dc = (Vp/V)Dp + (Vs/V)Ds$,

we can get Vp as

$$Vp = V \{ (Dc - Ds) / (Dp - Ds) \} = 0.62V$$

Since

$$Dp = Mp\ Z / (N\ Vp) \quad \text{or} \quad Z = N\ Vp\ Dp / Mp$$

where N is Avogadro number, Mp is molecular weight of protein and Z is

molecular number of protein in unit cell,

we calculate

$$Z = \{ (6.02 \times 10^{23}) \times (0.62 \times 70 \times 70 \times 355 \times 10^{-24}) \times (1/0.74) \}/11000$$

$$= 7.9796 = 8$$