A new approach to protein structure determination by X-ray diffraction methods: Structure analysis of *Flammulina Veltipes* agglutinin

A Doctoral Thesis

by

Masaki Yamamoto

Submitted to the Faculty

of Science, Osaka University

September, 1991

Approval

September, 1991

This thesis is approved
as to style and content
by

勝部幸輝
Member-in-chief

小高忠男
Member

小林雅通
Member

中村晃
Member

高木俊夫
Member

Acknowledgments

The present work has been performed under the direction of Professor Yukiteru Katsube of Institute for Protein Research, Osaka University.

The author would like to express his gratitude to Professor Yukiteru Katsube for his incessant guidance and encouragement throughout this investigation.

The author is deeply indebted to Associate Professor Yoshiki Matsuura and Dr. Yasuyuki Kitagawa, Institute for Protein Research, Osaka University for his continuing guidance and encouragement.

The author is deeply indebted to Dr. Mamoru Sato, Institute for Protein Research, Osaka University, for his guidance to development of Imaging Plate System.

The author is deeply indebted to Associate Professor Satoshi Fujii, Faculty of Pharmaceutical Science, Osaka University, for for his helpful advice to investigation of XPLOR.

The author is deeply grateful to Professor Youichi Hashimoto, Faculty of Science, Saitama University, for his helpful advice to Amino Acid Sequence of FVA.

The author is deeply grateful to Professor Noriyoshi Sakabe and Dr. Atsushi Nakagawa, Photon Factory, National Laboratory for High Energy Physics, for the helpful advice to experiments by use of Synchrotron Radiation.

Research, Osaka University and Dr. Yasuo Hata of Institute for Chemical Research, Kyoto University, for his helpful advice and encouragement. The author would like to thank all members in Kastube's laboratory for their kind assistances and encouragement.

Finally, the author thanks all presents for their incessant understanding and encouragement.

山本 雅貴

Masaki Yamamoto

September, 1991

Contents

General Introduction

Proteins play central role in all life processes, catalyzing biochemical reactions with remarkable specificity and serving as key structural elements in all cells and tissues. As enzymes, proteins catalyze both digestion of foodstuffs and construction of new macromolecules, and as antibody proteins they protect us from disease. Proteins are involved in every aspect of life: catalysis, structure, motion, recognition and regulation.

In recent decades, tremendous progress has been made in understanding protein structures and functions. Since, the amino acid sequence of insulin was reported in the early 1950s, many thousands of protein sequence have been known. Since, the three-dimensional structure of myoglobin was reported in 1960, hundreds of protein structure have been known. Recent advances in gene synthesis and genetic engineering have made it possible, in principle, to construct any desired amino acid sequence.

Our understanding of protein structure-function relationships is based firmly on X-ray crystallographic analysis of protein structure. The number of high resolution crystallographic solutions of protein structure has increased substantially as the result of the development of X-ray crystallography. This recent accumulation of structural data by protein crystallography has facilitated parallel advances in capabilities for theoretical predictions of some protein structures from their primary amino acid sequences. Such theoretical approaches are crude when applied to sizable amino acid sequences in cases

where no crystallographic data exist. However, theoretical approaches are relatively effective in predicting structural alterations resulting from changes in one or a few amino acid residues in the proteins of known structure.

Advances in molecular genetics and more precisely site specific mutagenesis are the other main element in a unique and powerful fusion of technologies. Predicting changes that will arise as a consequence of amino acid substitutions is one thing. Creating that mutant protein, obtaining an empirical solution of its structure, and testing for altered functional qualities are quite another.

Molecular genetics and structural biophysics are currently merging to form a hybrid field of protein engineering. In spite of progress in this field, predicting appropriate sequence changes is very difficult, and this is real challenge of protein engineering. There are two fundamental problems. First, we cannot predict how a protein will fold. Although a particular amino acid sequence of a protein determines its three-dimensional structure, it is difficult to predict the three-dimensional structure from the sequence. Second, we have known only a small part of the relationship between protein structure and function. To resolve these problems, we must know the three-dimensional structures of more proteins in atomic detail, and the X-ray crystallography has a great advantage for this purpose over other methods. In the protein engineering, therefore X-ray crystallography is the most basic and important technique.

The recent explosive growth in development and distribution of electronic computers has reduced the labor in the X-ray analysis of protein structures, and

the analysis may be used to solve many problems arising in the life-science. Protein Crystallography has advanced extremely rapidly in the last decade, but it still has difficulties and is time consuming. We think that structure determination should not be the end point of protein crystallography analysis. The successful solution of the structure forms the starting point for a whole range of experiments designed to elucidate a proteins biological function. Therefore, it is required to make the procedures faster and easier in such a way that biochemists who are not acquainted with the X-ray crystallography can use the technique. We have tried to routinize the process of determining the structure of protein. In this paper we developed new techniques of X-ray diffraction methods, which make possible the automatic determination of protein structure.

With these new techniques, we determined the structure of lectin-like protein FVA; *Flammulina veltipes* agglutinin[1,2], which agglutinates human erythrocytes. The hemagglutinating activity of FVA is not inhibited by either monosaccharides or simple oligosaccharides. Recently a lectin-like protein Ling Zhi-8 (LZ-8), whose sequence has 65 % homology with FVA, was isolated from *Ganoderma lucidium*[3]. LZ-8 has a mitogenic activity *in vitro* and immunomodulating activity *in vivo*. Since LZ-8 has considerable sequence similarity to the variable regions of immunoglobulin and LZ-8 has biological activities as an immunomodulator both *in vivo* and *in vitro*, one assumed that LZ-8 would belong to an immunoglobulin superfamily[4]. If this assumption is correct, the similarities of the sequence and the function of FVA to those of LZ-8

would suggest that FVA would also be classified as a member of an immunoglobulin superfamily. Information about the three-dimensional structures of FVA and LZ-8 is a key for deciding the relationship among FVA, LZ-8 and immunoglobulin.

Chapter 1 describes the hardware and software newly developed for the structure determination by X-ray diffraction and the development of the new equipment and methods for protein crystallography. Chapter 2 deals with the background of FVA. Chapter 3 is concerned with the structural determination of FVA. In Chapter 4, the structure of FVA is discussed in detail. In Chapter 5, the relationship between FVA and immunoglobulin is discussed in terms of the three-dimensional structure. Chapter 6 is for Summary and Conclusions.

Chapter-1

Development of new Protein Crystallographic Techniques

1-1. Introduction

The X-ray analysis of a protein structure usually proceeds in the following seven steps[5,6].

1. Growth of large, perfect crystal and preliminary characterization of the crystals and their diffraction pattern.
2. Preparation of heavy-atom derivatives.
3. Measurement and processing of diffraction data.
4. Calculation of phases for Fourier coefficients.
5. Interpretation of electron density maps and model building.
6. Refinement of molecular models.
7. Analysis of the structure.

The systematic appearance of this list leads to a misleading because in practice there is a frequent feed back between the various stages. Moreover the steps are not independent and success in a certain step depends on the success of the preceding steps. The interdependence of these steps makes the automatic analysis difficult. To make the structural analysis more routine, we need precise estimates of the errors involved in each step and their correlation

among the different steps. We mentioned the following five difficulties;

1) Difficulty for protein crystallization.

2) Difficulty of heavy-atom derivative crystals preparation.

3) Difficulty of the diffraction data collection.

4) Difficulty of the initial electron density map interpretation.

5) Difficulty of the protein structure refinement .

The difficulties 1) and 2) are mainly associated with the lack of theoretical backgrounds. The mechanism of protein crystallization is not fully understood for many cases, and presumably differs depending on samples as well as crystallization condition. The mechanism of heavy-atom binding is also not well understood coordination chemistry. Consequently, crystallization and the preparation of heavy-atom derivatives have been carried out by trial and error methods. The difficulties 3) and 5) are mainly due to the necessity of many times. The most important rate-limiting steps are those of the data collection and the structure refinement. Improvements of the data collection equipment and the refinement programs are the most straightforward way to be free from these two difficulties. The difficulty 4) comes from the fact that the isomorphous replacement methods become ineffective for crystals beyond a resolution of about 2.3 Å and the electron density maps with some uncertainty. This difficulty was reduced by the improvement of the model building program.

We introduced new equipment and methods to overcome these difficulties and to speed up the structure analysis. The first part of this chapter

describes new two-dimensional diffractometers, the second part describes a new model building system, and the third part describes a new molecular dynamics refinement program.

1-2. New equipment of data collection

Two types of the two-dimensional diffractometers were examined to speed up the data collection. One is the Imaging Plate system which has been constructed by combining an automated-oscillation camera and a Fuji Imaging Plate. The other is a commercial Siemens two-dimensional proportional counter system (Area Detector X-1000 system) designed by us for the protein structure analysis.

1-2-1. Imaging Plate system

The Imaging Plate system is an automated two-dimensional diffractometer based on the Fuji Imaging Plate (IP). The Imaging Plate system is an automated-oscillation camera with reusable film and an integrated film scanner. However, the inherent properties of the Imaging Plate compared to ordinary X-ray film make it a far superior medium from recording X-ray events[7].

Imaging plate

The Imaging Plates have a high dynamic range and a high sensitivity.

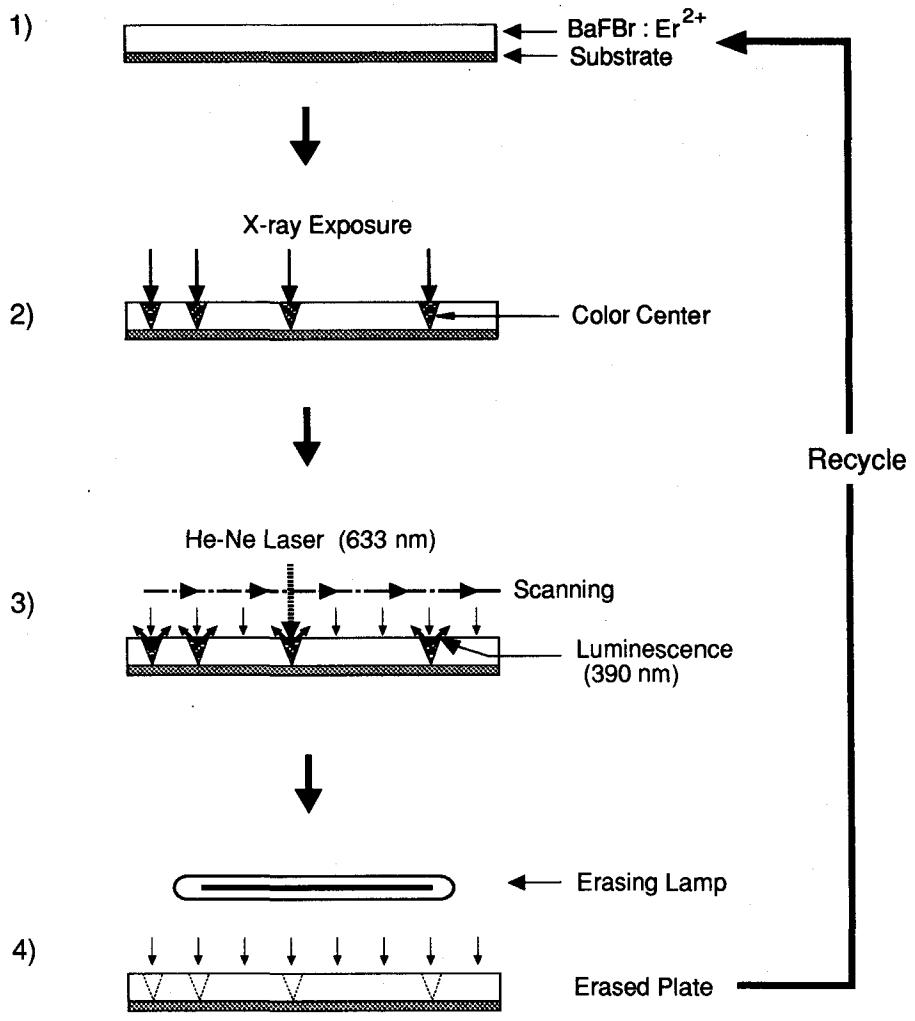


Figure 1-1 Mechanism of Imaging Plate

1) The active area of the Imaging Plate is composed of BaFBr:Eu^{2+}. 2) When X-rays strike the Imaging Plate Eu^{2+} is oxidized to Eu^{3+} with the free electron being trapped in a color center. 3) Measurement of the diffraction pattern is accomplished by scanning the Imaging Plate with a 633 nm He-Ne laser which triggers atomic fluorescence occurring at approximately 390 nm. This resulting luminescence is measured by a photomultiplier and has an intensity proportional to the number of absorbed X-ray. 4)The used Imaging Plate is erased by exposing it to visible light from a lamp, and reused.

The active area of the Fuji Imaging Plate is composed of a europium doped barium halide, $BaFBr:Eu^{2+}$. When X-rays strike the Imaging Plate Eu^{2+} is oxidized to Eu^{3+} with the free electron being trapped in a color center. In this way latent images of the diffraction pattern are stored on the Imaging Plate. Measurement of the diffraction pattern is accomplished by scanning the Imaging Plate with a 633 nm He-Ne laser beam which triggers photostimulated luminescence. The light from the laser causes an electron in a color center to reduce Eu^{3+} to Eu^{2+} accompanied with the atomic fluorescence occurring at approximately 400 nm. This resulting luminescence is measured by a photomultiplier and has an intensity proportional to the number of absorbed X-ray. The used Imaging Plate is erased by exposing it to visible light from a lamp, and reused. The recording and reading processes of Imaging Plate is sketched in Fig. 1-1.

System diagram

The mechanical configuration consists of a synchronized high speed shutter, a mirror-mirror focusing system, an ϕ axis goniostat, two imaging plates (200 x 200 mm), an erasing lamp and a laser/photomultiplier two-dimensional linear scanning system (Fig. 1-2).

Data processing

Data from two or more still data frames are examined to find the centroid

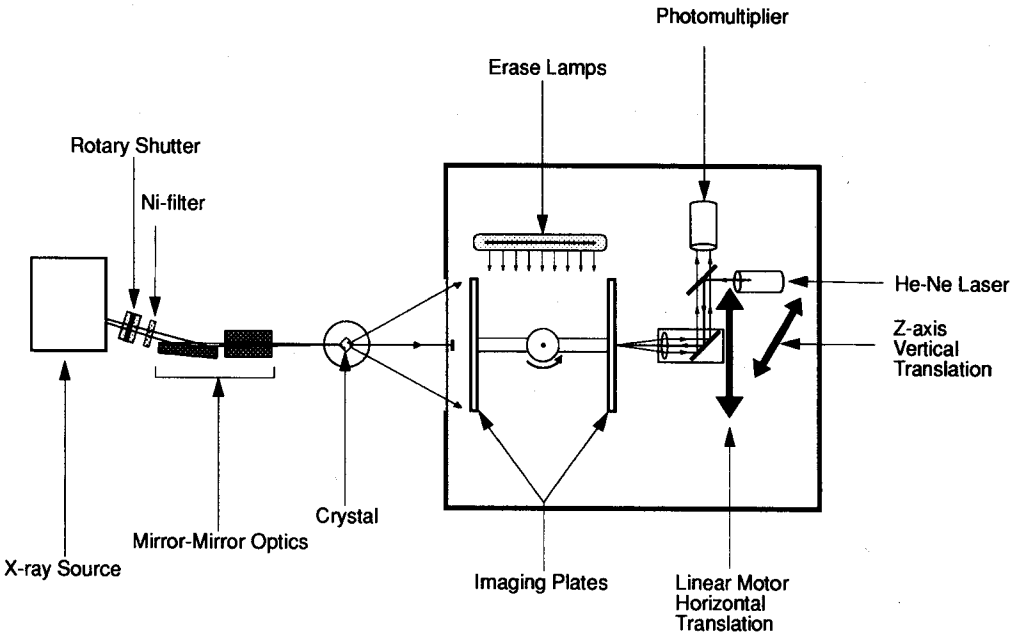


Figure 1-2 A schematic diagram of the Imaging Plate system

in the detector coordinates and the scanning angles of a set of bright spots. The (h, k, l) indices of the bright spots are determined by an auto-indexing technique, with some initial unit cell lengths and angles provided by the user. The indexed reflections are used to refine crystal unit cell and orientation parameters and the position of Imaging Plate.

After refinement of the camera parameter and indexing the reflections, the position of the reflections in X and Y are determined, and the intensities of those spots are calculated by profile fitting and by simple summation. The intensity values are corrected for Lorenz, and polarization, and then analyzed. The integration software is based on algorithms developed by Rossman[8,9].

After integration is complete all data are merged into one data set. Then scaling is then applied to the data. After a final scaling, the mean intensities of the scaled reflections are calculated and the mean structure amplitudes derived from them are generated. The scaling software is based on algorithms developed by Fox and Holmes[10].

Data collection by IP-system

As a first test the diffraction data of the native FVA crystal was collected by IP-system equipped with a rotating anode X-ray generator RIGAKU RU-200. X-ray generator was operated at 40 kV and 100 mA and Cu-Kα radiation was monochromatized by a Ni-filter and focused by the mirror-mirror optics. The setting parameters of the Imaging Plate system and the statistic data are summarized in Table 1-1. The crystal was mounted with the a^*-axis parallel to

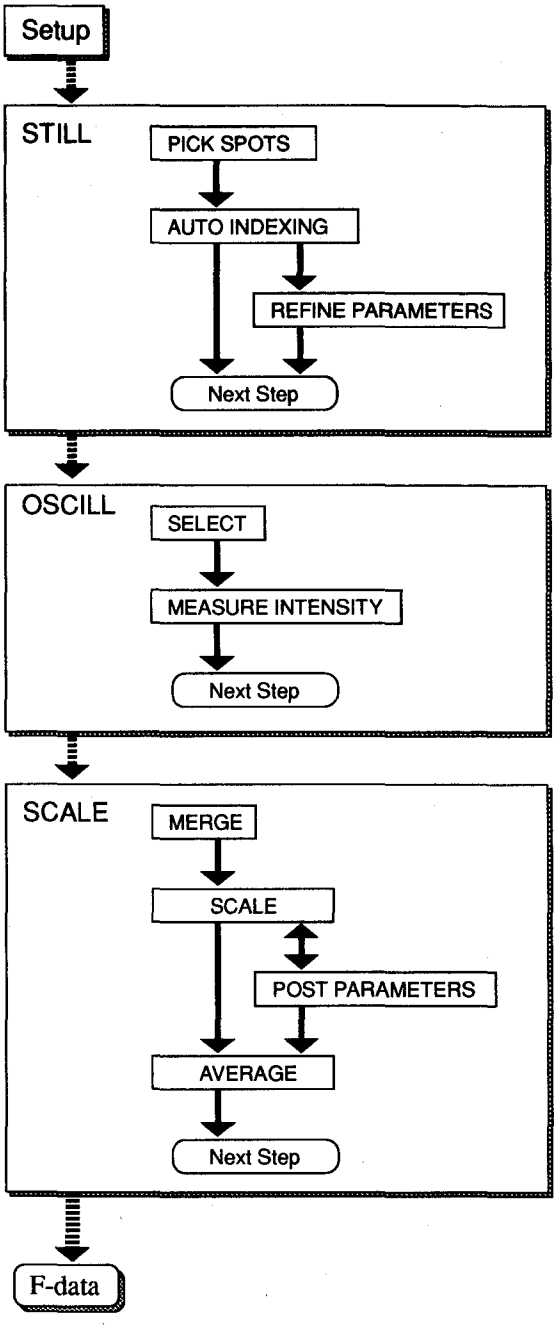


Figure 1-3 A schematic representation of the data processing with the Imaging Plate system

the oscillation axis of the goniostat. The crystal to Imaging Plate distance and the oscillation angle were determined from the maximum cell length. For FVA, the crystal to Imaging Plate distance was to 85 mm and the oscillation angle was set to 1.3 degree. The exposure time per one frame is 25 minutes, and 72 imaging plates were used to record all diffraction spots of the independent region in the reciprocal space. Data collection of the diffraction data at a resolution of 1.88 Å was completed with only one crystal within two days.

To determine the accuracy of the Imaging Plate system by the comparison of the anomalous difference, the data collection of the UO_2-derivative crystal was carried out. The setting parameters of the Imaging Plate system and the statistic data are summarized in Table 1-2. The crystal was mounted with the a^*-axis parallel to the oscillation axis of the goniostat, and all setting parameters were equal to the case of native crystal. The diffraction data of the UO_2-derivative crystal was processed up to a resolution of 2.5 Å. Data collection of the derivative data was also completed with only one crystal within two days. The mean change in structure factor between the native and the derivative data collected with the Imaging Plate system was 18.8 %, and the anomalous difference in structure factor was 5.0 %. The Harker sections of difference Patterson map and anomalous difference Patterson map at a resolution of 5.5 Å are shown in Fig.1-4. The map clearly shows the UO_2-site the same as that determined by the four-circle diffractometer data. This result suggests that the accuracy of the Imaging Plate system is enough to determine the anomalous dispersion effects.

Table 1-1 Summary of diffraction data collected by the IP-system

Detector	IP-system	
Sample crystal	FVA native	
Camera length	85.0	mm
Oscillation angle	1.3	degree
Number of imaging plates	72	
Exposure time	25	minutes
Resolution	1.88	Å
Number of recorded reflections	99867	
Number of unique reflections	20445	
Efficiency	81	%
R_{merge}(full)	5.24	%
R_{merge}(partial)	5.63	%
R_{merge}(total)	5.36	%

R_{merge} is defined as $R_{merge} = \sum \sum | I_i / G_i - \langle I \rangle | / \sum \langle I \rangle$, where G_i is the inverse scale factor.

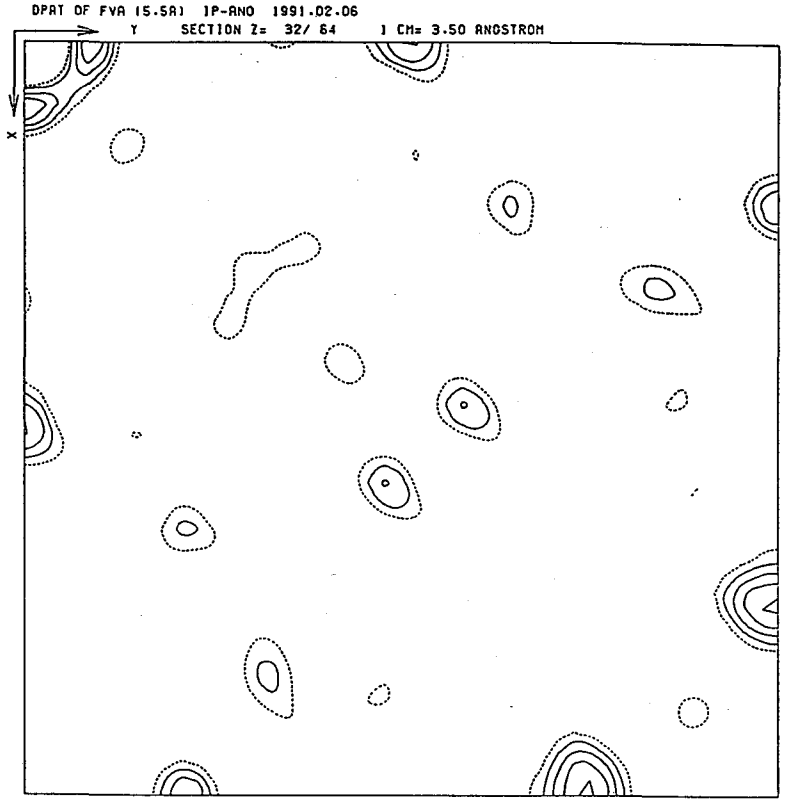
Table 1-2 Summary anomalous diffraction data on UO_2-derivatives.

Detector	IP-system	
Sample Crystal	FVA UO_2-derivative	
Camera length	85.0	mm
Oscillation angle	1.3	degree
Number of imaging plates	70	
Exposure time	25	minutes
Resolution	2.5	Å
Number of recorded reflections	57173	
Number of unique reflections	10087	
Efficiency	92	%
R_{merge}(full)	9.06	%
R_{merge}(partial)	9.58	%
R_{merge}(total)	9.18	%
R_F	25.67	%
R_{ano}	4.86	%

$$R_{merge} = \sum \sum \left| I_i / G_i - \langle I \rangle \right| / \sum \langle I \rangle$$

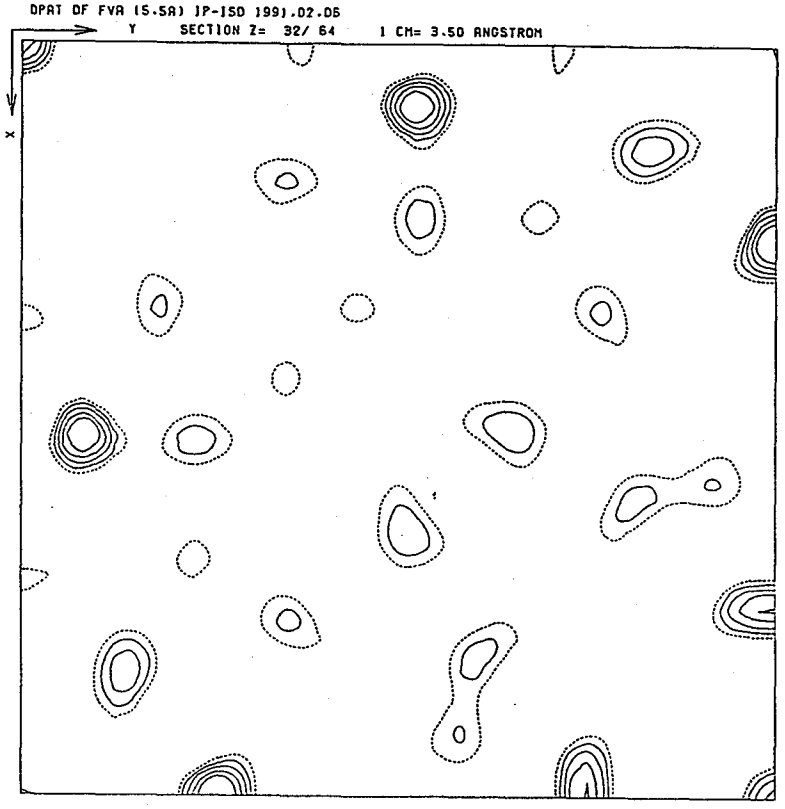
$$R_F = 2\sum \left| |F_{PH}| - |F_P| \right| / \sum \left(|F_{PH}| + |F_P| \right)$$

$$R_{ano} = 2\sum \left| \left| F_{PH}^{+} \right| - \left| F_{PH}^{-} \right| \right| / \sum \left(\left| F_{PH}^{+} \right| + \left| F_{PH}^{-} \right| \right)$$



Anomalous difference Patterson map

Figure 1-4 The Harker section (W=1/2) of anomalous difference Patterson map and difference Patterson map
Both maps were calculated by the diffraction data collected by the IP-system.



Difference Patterson map

Figure 1-4 continued

1-2-2. Area Detector system

The Siemens Area Detector[11] is an imaging proportional counter. This Area Detector, which is filled with pressurized xenon gas, is capable of determining the x and y position of X-rays which enter its imaging area. Individual xenon atoms in the detector are ionized by X-rays producing a shower of charged particles for each incident X-ray. The charged particles interact with the electrode assembly, which is formed as a multi-wire grid, to generate electrical signals indicative of the x-y position of the original X-ray. The active area of the wire plane is roughly circular with a diameter of 11.5 cm, and data are read out as a 512 x 512 pixel array of 16-bit numbers. Each pixel thus occupies an 0.2 x 0.2 mm square. Each data frame involves a small oscillation (5 to 15 minutes) about a vertical axis, and each succeeding frame is exactly one oscillation range further along than the preceding frame.

In Area Detector system the mechanical component system is controlled by the personal computer, and the diffraction data are collected by the program FRAMBO. Data processing is carried out by the program XENGEN[12], which is running on the frame processor, DEC VAXstation-3200.

Data collection by Area Detector system

As a first test the diffraction data of FVA was collected by the Area Detector system equipped with a rotating anode X-ray generator MAC-science MX-18. X-ray generator was operated at 40 kV and 80 mA and Cu-Kα radiation

Table 1-3 Summary of diffraction data collected by AD-system

Detector	SIEMENS X-1000 Area Detecter System	
Sample crystal	FVA native	
Camera length	120.0	mm
Oscillation angle	0.10	degree
Number of frames	1000	frames
Exposure time	45.0	seconds
Resolution	2.72	Å
Number of recorded reflections	24781	
Number of unique reflections	6527	
Rmerge	11.04	%

R_{merge} is defined as $R_{merge} = \sum | I_i - \langle I \rangle | / \sum \langle I \rangle$

was monochromatized by a Ni-filter. The setting parameters of the Area Detector and the statistic data are summarized in Table 1-3. The minimum crystal to detector distance was determined from the formula of $D_{min} = a_{max} / 0.8$, where a_{max} is the longest cell edge in angstroms. The longest cell edge is 97.7 in FVA, then the crystal to detector distance was set to 120 mm. The frame width, which is equivalent to the oscillation angle, was set to the one-third of the peak profile width. In this case, frame width was set to 0.1 degree. The exposure time per one frame is 45 seconds, and the number of total frame is 900 frames to collect an independent region of 90.0 degree. Data collection of the diffraction data at a resolution of 2.72 Å was completed with only one crystal in one day.

1-2-3. Comparison of the new detector

The comparison of the data collection among the new detector systems, the synchrotron radiation source system and the four-circle diffractometer is shown in Table 1-4. Two-dimensional detectors collect all data simultaneously on its detector. This is fast, but resolution is limited by the size of the detector. The conventional four-circle diffractometer collects individual diffraction, so it is slower than new detectors. Only one crystal was necessary for data collection with the two-dimensional detectors. However, the four-circle diffractometer needed at least 10 crystals. The total exposure time of the synchrotron radiation source system was much shorter than all other conventional X-ray source systems, because of the high brilliancy of the synchrotron X-ray source. The

Table1-4 Comparison of data collection

	Two-dimensional detector			
Detector	IP	AD	PF	AFC
Number of crystals	1	1	1	18
Oscillation angle (degree)	1.3	0.10	6.0	-
Number of frames	72	1000	20	-
Exposure time / frame (minute)	25.0	0.75	0.5	-
Total time for				
exposure (minute)	1800	750	10	-
data process (minute)	100	300	600	120
All (minute)	1900	1100	600	100000
All (day)	2	1	1	70
Resolution (Å)	1.88	2.72	2.0	2.5
Rmerge (%)	5.36	11.04	5.12	7.2
Total number of				
recorded reflections	99867	24781	56219	89542
unique reflections	20445	6527	15601	20334
Collected reflections / 1 hours	3000	1000	30000	100

IP : Data collected by Imaging Plate system.

AD : Data collected by Area Detector system.

PF : Data collected by Synchrotron radiation source.

AFC : Data collected by Four-circle diffractometer.

total exposure time of the IP-system and that of the Area Detector system are similar and are ten times shorter than that of the four-circle diffractometer. The collected reflections per 1 hour by the synchrotron radiation source system, IP-system, Area Detector system and the four-circle diffractometer are roughly 30000, 3000, 1000 and 100, respectively.

The comparison of the diffraction data among the new detector systems, the synchrotron radiation source system and the four-circle diffractometer system is shown on Table 1-5. The correlation R-value between the IP-system and the two conventional systems is lower than 10.0 %, and this value is similar to that between the synchrotron radiation source and the four-circle diffractometer. Because of this and the result of the anomalous diffraction data, the accuracy of the data is very high in the IP-system. The correlation R-value between the Area Detector system and two conventional systems is about 20.0 %, and this value is over twice of the value between the synchrotron radiation source and the four-circle diffractometer. In Area Detector system, the accuracy of the collected data is relatively low than other data collection system.

Both types of the two-dimensional detectors reduce the data collection time as well as the number of crystals to be used. The advantage of the IP-system is the accuracy of the collected data, and the weak point of the IP-system is that the data collection process is divided into two steps, the exposure and the reading out the diffraction pattern. The advantage in the Area Detector system is the real time data collection which means simultaneous exposure and reading the diffraction pattern, and the weak point of the Area Detector system is the low accuracy of the collected data.

Table1-5 Comparison of diffraction data for R_F values

Fo	F1		
Name	PF	AD	IP
AFC	7.67	18.04	9.93
PF	-	20.03	9.52
AD		-	22.40
IP			-

This table shows the mean change in structure factor; R_F.

AFC : Data collected by four-circle diffractometer.

PF : Data collected by synchrotron radiation source.

AD : Data collected by Area Detector system.

IP : Data collected by Imaging Plate system.

$$R_F = 2\sum ||F_1|-|F_0|| / \sum (|F_1|+|F_0|)$$

1-3. New program for model building

FRODO[13] is a new generation comprehensive molecular modeling and electron density fitting program. The comprehensive molecular graphics program FRODO, originally written by T. Alwyn Jones[13,14], was implemented on the Evans & Sutherland PS300-series by James W. Pflugrath and Mark A. Saper[15]. The graphical aspects of FRODO permit manipulation of the entire view such as rotation, intensity, scaling and windowing, and allow modeling of atomic groups and bonds such as bond making and breaking, translation of single atom, rotation and translation of groups of atoms and torsions. The user interacts with the display through dials and a data tablet.

We introduced FRODO in the new generation Evans & Sutherland PS390, a high performance computer graphics system, linked to a Micro-VAX II computer through the ethernet, a high speed computer network system. We first introduced PSFRODO ver. 6.4, and updated ver. 6.6 in our laboratory. In this environment all non-graphical aspects of the application program are performed by the host computer while all graphical operations executed in real time on the PS390. The molecular modeling program has no great requirement on the host computer because the graphical operations are limited on the PS390. The advantages over previous host-dependent graphics system are quite significant.

In the PS390 implementation of FRODO, all graphical functions, once initialized, are host independent. The host provides the initial coordinate data in angstroms. In addition, the PS390 communicates the structure change back

to the host only when explicitly requested. Thus host-PS390 communication is significantly reduced, while providing an extremely high degree of real time interactions with the display.

Procedure of FRODO

Program FRODO is divided into five major routines: SAM, CHAT, INTER, REFINE, and MOLCUL. Each of these routines serves a particular purpose:

Routine CHAT is used to set most of the program constants and data file names. All the parameters which determine what will be shown on the display screen are first set by entering commands on the terminal while in the CHAT routine before going on the interactive graphics.

Routine SAM is used to create, examine, and manipulate the atomic coordinate data file. This routine mediates between the internal and the external atomic coordinate by importing and exporting. In the case of an initial model, the atomic coordinate data file is made from the amino acid sequence.

Routine INTER is the interactive display mode. Manipulation of model on the computer graphic display is carried out by this routine. The model is displayed as a wire-frame model and the electron density map is displayed as a wire cage of desired contour level on the computer graphic display. The model is manipulated by the rotation of the dihedral-angles or by the translation and the rotation of a model fragment made by the breaking of bonds. Once in the interactive mode, most commands are entered using the PS390 data tablet, dials, and function keys rather than using the terminal.

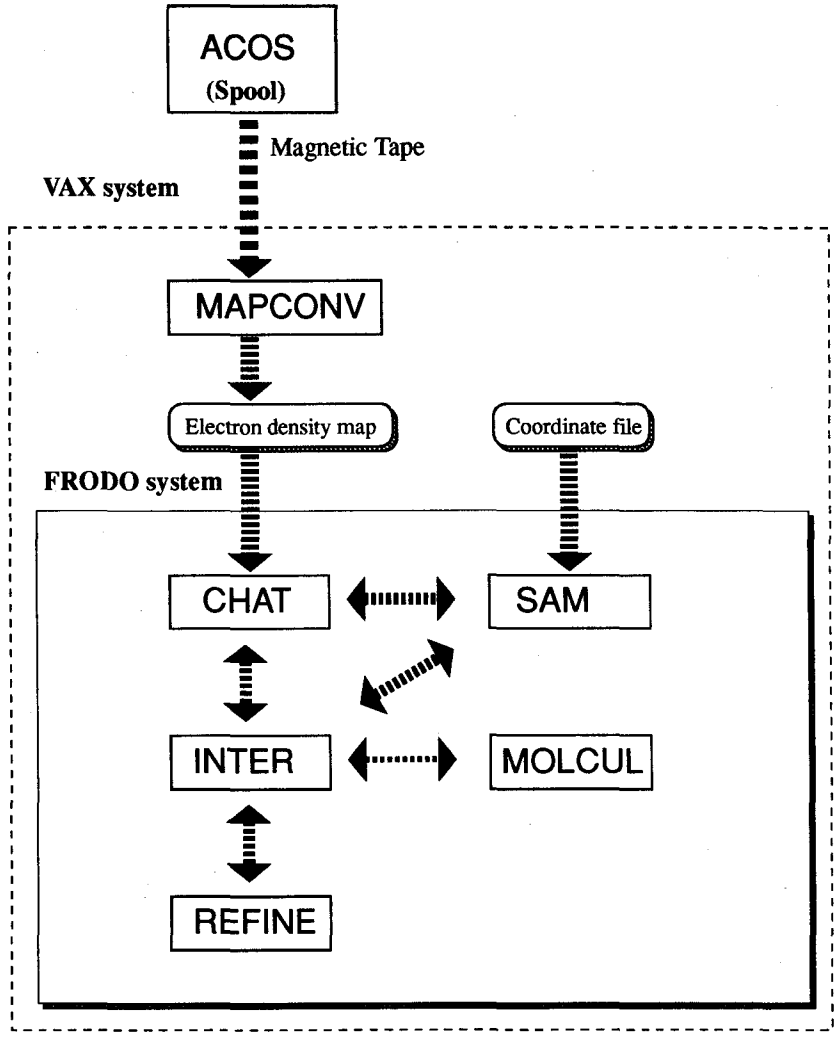


Figure 1-5 A schematic representation of the interactive model building program, FRODO

Routine REFINE is a model regularize routine[14]. This routine allows ideal bond lengths and angles to be applied to the atomic coordinates, usually after manual manipulation of the coordinates in the display has altered their geometry. This regularize function is very powerful. This function will be able to regularize the structure even if main-chain bonds are broken to greatly modify the structure.

Routine MOLCUL is used to generate static displays of the molecular background displays for the study of a molecular property or the presentation. For example back-bone models, molecular dot surface models and models colored for residue type or thermal motion.

The connection of five routines and auxiliary programs for importing information are shown in Fig. 1-5.

In this system, rotation and translation of the model at the computer graphics are more smooth and direct compared to the operation of the previous model building system, the MOLFITG and ACOS-S930 system. The regularize function of the mode is very powerful allowing faster and easier model building.

1-4. New program for refinement

XPLOR[16,17,18] is a general-purpose macromolecular refinement program that uses crystallographic diffraction or nuclear magnetic resonance data in combination with energy minimization[19] and molecular dynamics[20]. The program allows one to use simulated annealing[21] to overcome local energy minimum. Historically, the program has evolved from the CHARMM program[22], and developed by Axel T. Brunger at Yale University. Now, XPLOR has been adapted to the main frame computer ACOS-S930 and VAX-station system at Institute for Protein Research.

Molecular dynamics[20] is used to refine macromolecular structures by incorporating the difference between the observed crystallographic structure factor amplitude and that calculated from an assumed atomic model into the total energy of the system. The method has a radius of convergence that is larger than that of conventional restrained least-squares refinement.

Conventional refinement involves a series of steps, each of which consists of a few cycles of least-squares refinement with respect to the stereochemical and internal packing constraints or restrains, followed by rebuilding the model structure with interactive computer graphics. Finally, solvent molecules are included and alternative conformations for some atoms in the protein may be introduced. Conventional refinement is time-consuming, because the limited radius of convergence of least-squares algorithms necessitates the periodic examination of electron density maps which are computed with various combinations of F_{obs} and F_{calc} as amplitudes and with

phases calculated from the model structure. Restrained least-square refinement in general does not correct the positions of residues that are displaced by more than 1 Å. Also, the least-squares refinement process is easily trapped in a local minimum so that human intervention is required.

Simulated annealing[21], which makes use of molecular dynamics to explore the conformational space of the molecule, can help overcome the local-minimum problem.

Theoretical background of XPLOR

The XPLOR total energy can be grouped into two different classes:

$$E_{TOTAL} = E_{EMPIRICAL} + E_{EFFECTIVE}$$

$E_{EMPIRICAL}$ describes the energy of the molecules through an empirical energy function. $E_{EFFECTIVE}$ is comprised of restrained energy terms that use experimental information or other information.

$$E_{EFFECTIVE} = E_{XREF} + E_{NOE} + E_{HARM} + E_{CDIH} + E_{NCS}$$

E_{XREF} is the effective energy calculated by the X-ray diffraction data. E_{NOE} is the effective energy calculated by the NMR structure information. E_{HARM} is the effective energy calculated by the restraints between two coordinate sets; a main coordinate set and a reference coordinate set. E_{CDIH} is the effective energy calculated by the restraint dihedral angles. E_{NCS} is the effective energy calculated by the non-crystallographic symmetry information.

The foundation of the entire function consists of the separable internal coordinate and pairwise nonbonded interactions. A harmonics approximation is used to account for deformations in bond length and angles. Four-atom terms (ijkl) are related to the angle made by two planes where the first plane is defined by the atoms (ijk) and the second plane is defined by the atoms (jkl). This four-atoms term is also used to maintain both chirality about the tetrahedral atom and planarity of certain atoms with a quadratic distortion potential. In XPLOR, the empirical energy function has the form:

$$E_{EMPIRICAL} = E_{BONDS} + E_{ANGLES} + E_{DIHEDRAL} + E_{IMPROPERS} + E_{NONBONDED} + E_{SBOUNDS}$$

where

$$E_{BONDS} = k(r - r_0)^2$$

describe the covalent bonded energy,

$$E_{ANGLES} = k(\theta - \theta_0)^2$$

describe the bond angle energy,

$$E_{IMPROPER} = E_{DIHEDRAL} = \begin{cases} k(1 + \cos(n\phi + \sigma)) \text{ if } n > 0 \\ k(1 + \cos(\phi + \sigma)) \end{cases}$$

describes energy terms involving dihedral angles, chirality or planarity.

The nonbonded energy term $E_{NONBONDED}$ has the form:

$$E_{NONBONDED} = E_{HBONDS} + E_{VDW} + E_{ELEC} + E^{P}_{VDW} + E^{P}_{ELEC}$$

where E_{HBONDS} describes an explicit hydrogen bonding energy, E_{VDW} describes the non-symmetry-related van der waals energy, E_{ELEC} describes the non-symmetry-related electrostatic energy, E^{P}_{VDW} describes the symmetry-related van der waals energy, E^{P}_{ELEC} describes the symmetry-related electrostatic energy.

$E_{SBOUNDS}$ is a term describing mean field boundaries for solvent molecules.

Effective energy of E_{XRAY} is:

$$E_{XRAY} = E^{A}_{XRAY} + E^{P}_{XRAY}$$

The term E^{A}_{XRAY} consists of the weighted differences between observed ($|F_{obs}|$) and calculated ($|F_C|$) structure amplitudes,

$$E^{A}_{XRAY} = \frac{W_A}{N_A} \sum_{\vec{h}} W_{\vec{h}} [\, |F_{obs}(\vec{h})| - k\,|F_c(\vec{h})| \,]^2$$

where $\vec{h} = (h, k, l)$ is the index referring to the reciprocal lattice points of the crystal and the sum extends over all observed and selected reflections with indices $\vec{h}$. $F_C(\vec{h})$ represent the structure factors that are computed from the atomic model. The overall weight W_A relates E^{A}_{XRAY} to the other energy terms. The purpose of the normalization factor N_A is to make the weight W_A approximately independent of the resolution range during SA-refinement. N_A has been set to

$$N_A = \sum_{\vec{h}} W_{\vec{h}} \left| F_{obs}(\vec{h}) \right|^2$$

$W_{\vec{h}}$ are individual weights for each reflection $\vec{h}$.

The term E^P_{XRAY} describes phase information if available. In this case, this term is not used.

Molecular dynamics consists of solving Newton's equations of motion

$$m_i \frac{d^2 x_i(t)}{dt^2} = -\nabla_{x_i} E_{TOTAL}$$

where the index i runs through all free atoms and the gradient, $-\nabla_{x_i} E_{TOTAL}$, is derived from the energy function. Normally all atoms are free. The atomic masses, m_i, are defined through the topology-statement. The initial atomic coordinates are taken from the main coordinate set. The initial velocities can be assigned to a maxwellian distribution:

$$v_i = \left(\frac{m_i}{2\pi k_B T_I} \right)^{\frac{3}{2}} exp \left(\frac{-m_i \delta^2}{2 k_B T_I} \right)$$

where σ is a random number among 0 and 1. T_I is the specified initial temperature.

After completion of molecular dynamics, the main coordinate set contains the coordinates of the last step and the velocities of the last step are stored in the atom-properties.

Procedure of XPLOR

One cycle of the crystallographic refinement by XPLOR is divided into five subsequent steps. The protocol of XPLOR is summarized in Fig. 1-6.

The first step generates the protein structure information, which contains the amino acid sequence, the atom connection, the hydrogen atoms for the possible hydrogen bonds, topological information and electrostatistical information, for structure refinement.

The second step determines the weight of the relation between the empirical energy term, which is energy of protein structure, and the effective energy term, which concerns with the diffraction data in the crystallographic refinement.

The third step minimizes the total energy for searching the initial model for the molecular dynamics. In the molecular dynamics structure refinement the initial model is shaken in a high temperature environment, generally at 3000 K or higher. If the initial model contains the abnormal short non-bonded contact or collisions, the total energy of the model is rapidly dispersed in this high temperature environment. This energy minimization process is a very important preparatory step in the molecular dynamics structure refinement.

The fourth step carries out the molecular dynamics structure refinement. This step has two types of protocol; the Quenching protocol and the Slow-Cooling protocol.[23] In the Quenching protocol, first the initial model is shaken by the molecular dynamics, and the total energy minimized at the fixed high

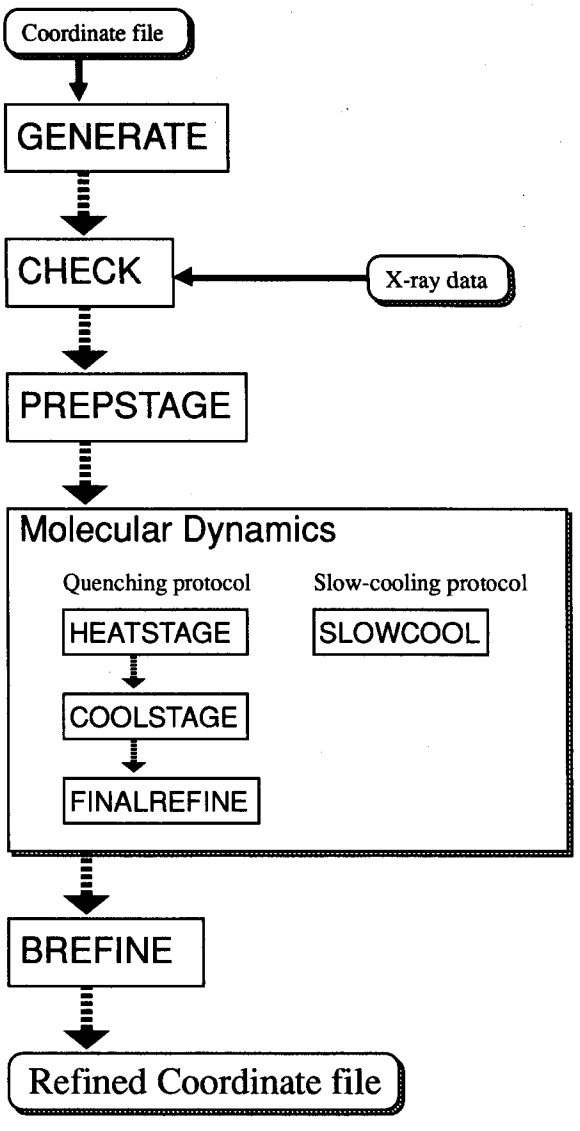


Figure 1-6 A schematic representation of the molecular dynamic refinement system, XPLOR

temperature environment, and second the resulting model is stepped from the high temperature to a room temperature environment like quenching. Finally the total energy of the model is minimized. In the Slow-Cooling protocol, the molecular dynamics of the model is started from the high temperature environment, and the temperature is gradually cooled to room temperature maintaining the molecular dynamics and the energy minimization. In general the Slow-Cooling protocol needs the much more computing time than the Quenching protocol, but the Slow-Cooling protocol provides a more stable refined structure than the Quenching protocol. Both protocols generate a final refined model.

The fifth step finally refines the individual temperature factors if desired.

Refinement by XPLOR

As a first test XPLOR was applied to cellulase[24], a protein of 221 amino acids, for the initial stage refinement. The initial structure for XPLOR was obtained from the coarsely refined model by the restrained least-square refinement methods. For cellulase, it is difficult to reduce the crystallographic R-value under to 35.0 % because of the stereochemical distortion in the restrained least-square refinement methods.

The refinement of the cellulase was carried out at a resolution of 2.5 Å. The crystallographic R-factor was started from 35.0 % including one molecule of 1702 atoms. XPLOR first made 385 hydrogen atom coordinates for the possible hydrogen bonds at the GENERATE stage. To determine the weight for

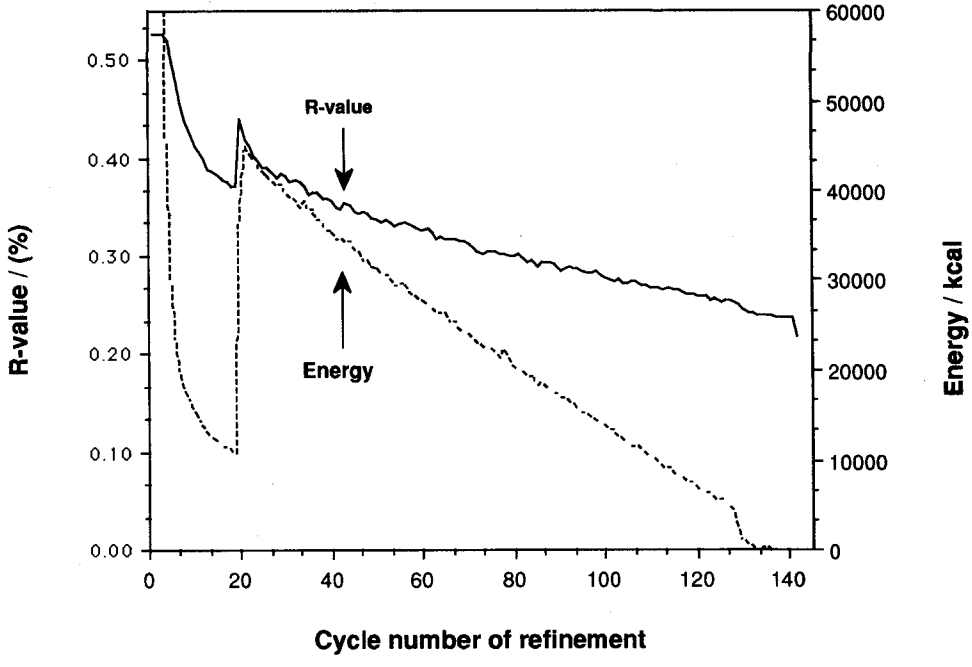


Figure 1-7 Refinement survey of cellulase by the molecular dynamic refinement. The molecular dynamic refinement was carried out by the Slow Cooling protocol. The R-value of the initial model was 52.6 %, and finally converged to 23.8 %.

the X-ray data, the CHECK stage was carried out. The refinement was carried out against the model of 2087 atoms and the X-ray data of 5954 reflections with $F/\sigma(F) > 2$. At PREPSTAGE the crystallographic R-factor and the total energy of the initial model were started from 52.6 % and 1.5×10^7 kcal, respectively. After 160 cycles of the energy minimization the crystallographic R-factor and the total energy were reduced to 37.0 % and 1.1×10^4 kcal, respectively. At the molecular dynamics refinement stage, one cycle of Slow-Cooling procedure was carried out. The starting temperature was 3000 K, the time step was 0.5 fsec. The number of steps for each temperature was 50 steps with a temperature step of –25 K, the reach a temperature of 300 K. With these conditions, total time and steps become 2.7 psec. and 5400 steps, respectively. Before the Slow-Cooling procedure the crystallographic R-factor and the total energy were 41.8 % and 2.6×10^4 kcal, and after the Slow-Cooling procedure and final refinement the crystallographic R-factor and the total energy were converged to 23.8 % and 81 kcal. Fig.1-7 shows the movement of the crystallographic R-factor and the total energy of the model at all stages of XPLOR refinement, and Fig.1-8 shows the movement of the crystallographic R-factor and the temperature at the Slow-Cooling procedure.

This calculation was carried out with DEC corporation VAX-station 3200 computer, which has the processing speed of 3.4 MIPS. Five steps; GENERATE, CHECK, PREPSATAGE, SLOWCOOL and BREFINE needed 0.3, 1.7, 14.0, 121.5 and 2.8 hours of C.P.U. time, respectively, and one cycle of the XPLOR refinement was completed in about 1 week. In the conventional refinement at least 100 cycles of the restrained least-square refinement

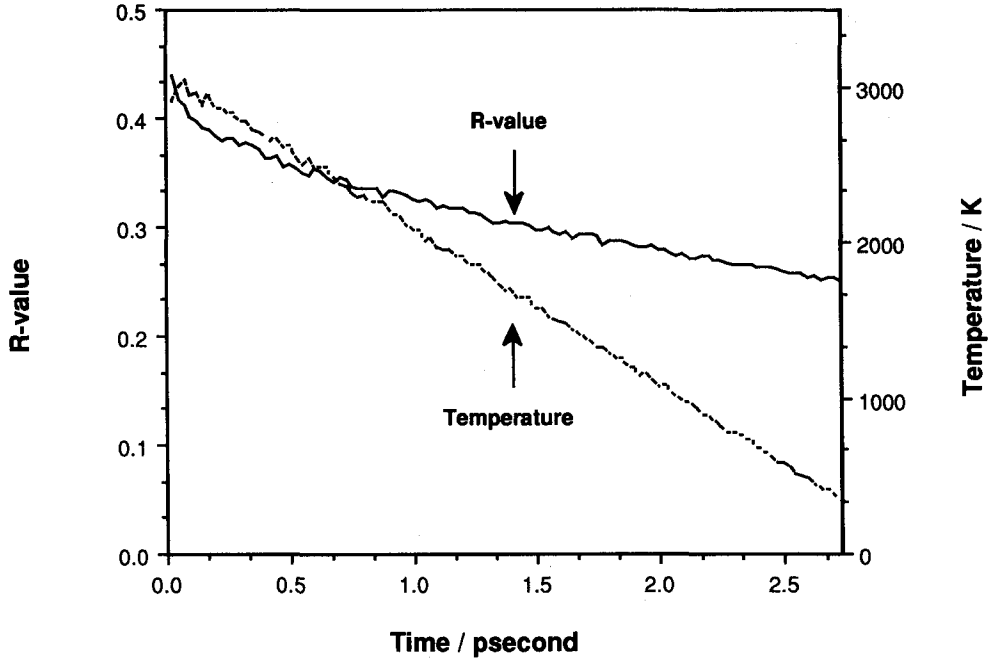


Figure 1-8 Time-course of the R-value of the cellulase model and the temperature. The molecular dynamic refinement was carried out by the Slow Cooling protocol. The refinement was started from the R-value 44.0 % and temperature 3000K, and after 2.7 pseconds molecular dynamics converged to the R-value 25.1 %, and temperature 300K.

calculation and several manual corrections of the model would be necessary; this refinement set would require several months. XPLOR could greatly speed up the refinement of the model structure. In this case, XPLOR easily overcame the local energy minimum without manual correction of the model. Thus XPLOR could semiautomatically refine the model structure.

As a second test, XPLOR was applied to FVA, a dimeric protein of 114 amino acids, for the final stage refinement. In order to determine the difference between the Slow-Cooling and the Quenching procedures, both procedures were carried out for the same initial model and contained the same peripheral stages. The initial structure for XPLOR was obtained from the finally refined model at a resolution of 2.0 Å by the restrained least-square refinement methods. The crystallographic R-factor was 19.8 % for 11604 reflections with $F/\sigma(F) > 3$ including two subunits of 1808 atoms and 159 water molecules.

At the heat stage of the Quenching procedure, heat stage temperature was 3000 K, the time step was 0.5 fsec., the number of steps was 5400 steps, and the total time was 2.7 psec.. At the cool stage of the Quenching procedure, the temperature was 300 K, the time step was 1.0 fsec., the number of steps was 250 steps, and the total time was 0.25 psec. After the Quenching refinement, the crystallographic R-factor converged to 19.5 % for 12476 reflections including 124 water molecules. Fig.1-9 shows the movement of the crystallographic R-factor and the temperature during the Quenching procedure. With the Slow-Cooling procedure, the first temperature was 3000 K, the time step was 0.5 fsec., the number of steps for each temperature was 50 steps, the

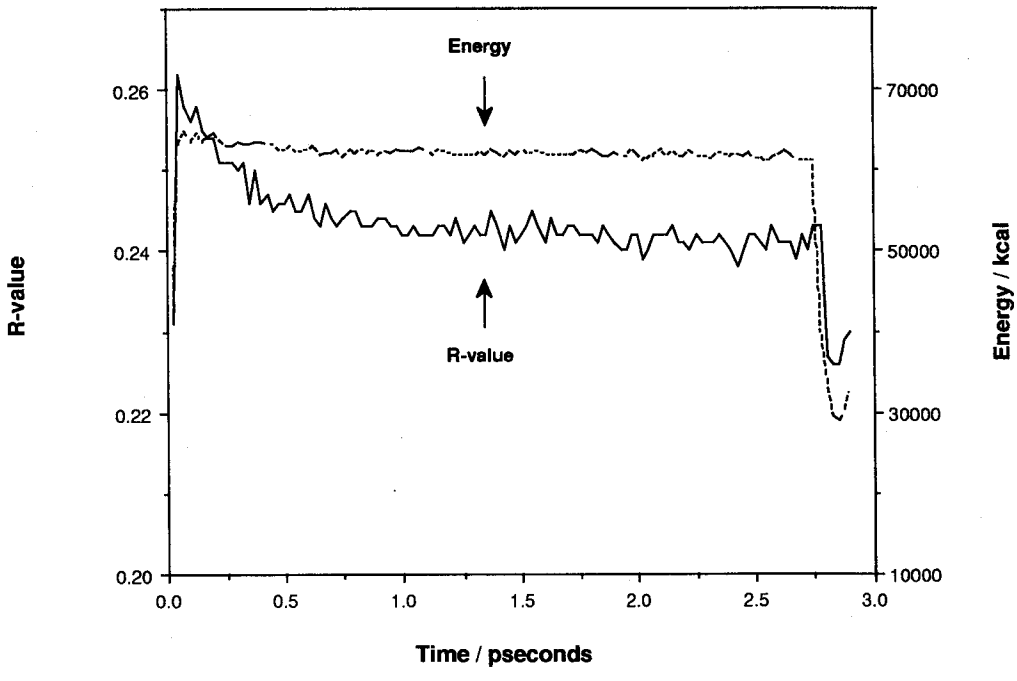


Figure 1-9 Time-course of the R-value and the total energy of the FVA model. The molecular dynamic refinement was carried out by the Quenching protocol. The refinement was started from the R-value 26.2% and temperature 2946 K, and after 2.7 pseconds molecular dynamics converged to the R-value of 23.0 %.

decrease temperature step was 25 K, and the final temperature was 300 K. In this condition the total time and the total steps became 2.7 psec. and 5400 steps, respectively. After the Slow-Cooling refinement, the crystallographic R-factor converged to 19.6 % for 12476 reflections including 124 water molecules. Fig.1-10 shows the movement of the crystallographic R-factor and the temperature at the Slow-Cooling procedure.

In both procedures, the finally refined models were very closely equivalent in structure. The root-means-square (r.m.s.) deviation between both models were 0.009 Å against all 1974 atoms. This result would suggest that the sufficient long heatstage of the Quenching procedure is equivalent to the Slow-Cooling procedure. Since the initial model was fully refined, these results are probably better than would be obtained in general case.

After the XPLOR refinement, the stereochemical deviation values of the model were made larger in appearance, but the distortions of the model were cured by the molecular dynamics. The crystallographic R-factor and the stereochemical deviation of the refined model were then reduced. In FVA after 15 cycle of PROLSQ for stereochemical refinement, the crystallographic R-factor converged to 18.5 % for 11604 reflections with $F/\sigma(F) > 3$ including 124 water molecules. The movements of stereochemical deviations before and after stereochemical refinement are summarized in Table 1-6.

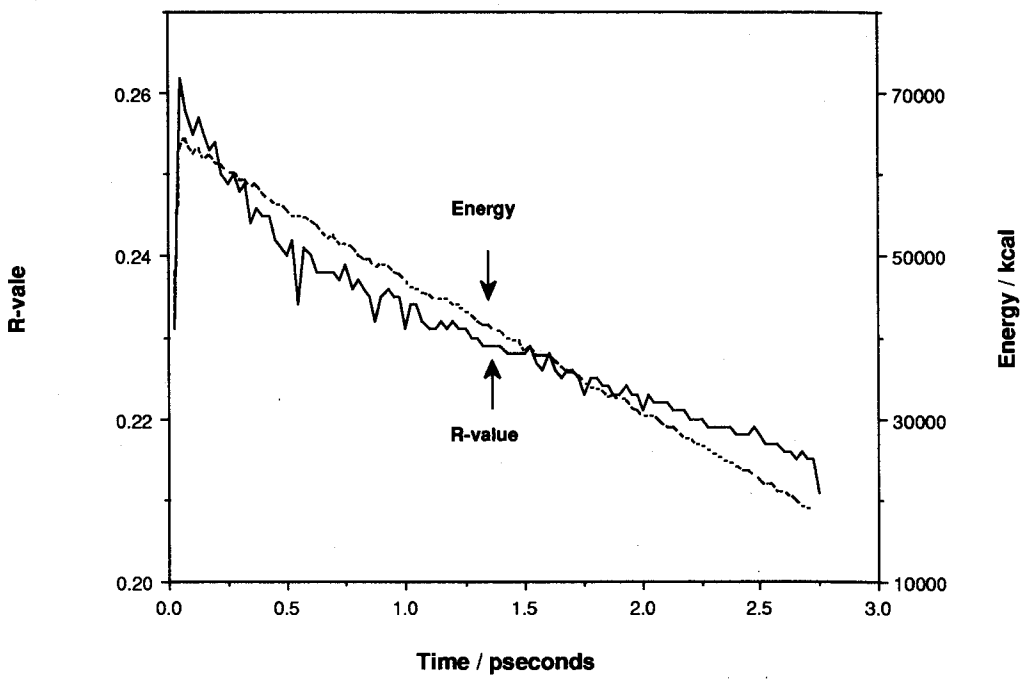


Figure 1-10 Time-course of the R-value and the total energy of the FVA model. The molecular dynamic refinement was carried out by the Slow Cooling protocol. The refinement was started from the R-value 30.3 % and temperature 3000K, and after 2.7 pseconds molecular dynamics converged to the R-value 24.4 %, and temperature 300K.

Table 1-6 Comparison of stereochemical deviations

Parameter	Output from XPLOR	rerefine by PROLSQ	
Bond distance			
1-2 bond	0.147	0.021	Å
1-3 angle	0.198	0.043	Å
1-4 planar	0.247	0.059	Å
Planes	0.041	0.017	Å
Chiral center	0.191	0.152	$Å^3$
van der Waals contacts			
Single torsion bonds	0.233	0.217	Å
Multiple torsion bonds	0.214	0.269	Å
Possible hydrogen bonds	0.145	0.222	Å
Torsion angles			
Planar (0,180)	7.1	2.7	degree
Staggered (60,180)	22.0	21.3	degree
Orthonormal (90)	28.1	25.7	degree
Noncrystallographic symmetry			
Tight position	0.036	0.073	Å
Medium position	0.143	0.214	Å
Weak position	0.668	0.503	Å
Tight thermal	1.790	0.478	$Å^2$
Medium thermal	2.008	1.956	$Å^2$
Weak thermal	2.295	2.742	$Å^2$
Temperature factors			
1-2 main chain	1.821	1.139	$Å^2$
1-3 main chain	2.863	1.978	$Å^2$
1-2 side chain	2.031	1.678	$Å^2$
1-3 side chain	3.048	2.458	$Å^2$
R-facter	19.1	18.7	%
reflections	11604	($F > 3\sigma(F)$)	

All parameters are using in the refinement program, PROLSQ.

Chapter-2

Flammulina veltipes agglutinin (FVA)

2-1. Lectin

Lectins generally has specific carbohydrate binding properties and exhibit hemagglutination or mitogenic activities toward lymphocytes. Carbohydrates are widely distributed on various cell surfaces, and lectins may mediate cell-cell interactions by their carbohydrate binding activities[25]. Lectins have took notice in recent years, because lectins behave like antibodies by interacting with specific carbohydrate sites on cell surfaces. Lectins have been isolated not only from plant sources, but also from widely diverse sources such as microorganisms, snails, invertebrates and some lower vertebrates. Lectins have been characterized by the molecular size and the properties such as specificity toward blood groups, effect for cells, antimitogenic effect, mitogenic effect, immunosuppressive effect, etc. Three-dimensional structures of several lectins, concanavalin A[26], pea lectin[27] and wheat germ agglutinin[28], have been determined by X-ray diffraction method. These do not have any structural resemblance, showing a further possibility of the structural divergence. The term lectin, has been frequently used to describe lectin-like protein which differ from the classical definition. Goldstein et al.[29], therefore, have defined a lectin as a sugar-binding protein or glycoprotein of non-immune origin which agglutinates cells and/or precipitates glycoconjugates; the specificity is defined

in term of the monosaccharides or simple oligosaccharides that inhibit lectin-induced agglutination reactions.

2-2. *Flammulina veltipes* agglutinin (FVA)

A lectin-like protein from a mushroom, *Flammulina veltipes*, was isolated by Tsuda[1] and characterized by Yatohgo et al.[2]. The lectin-like protein FVA, *Flammulina veltipes* agglutinin, agglutinates human erythrocytes regardless of ABO-blood type, and strongly agglutinates equine, porcine and rabbit erythrocytes. The hemagglutinating activity is not inhibited by either monosaccharides or simple oligosaccharides, but it is inhibited by asialofetuin or some glycoproteins. In solution, FVA is a dimer molecule which exhibits red blood cell agglutination. FVA subunit is a relatively small protein of 114 amino acids lacking cysteine, methionine, and histidine, and the approximate molecular weight of 12,700. FVA is an acidic protein with the pI value of 5.38. Recently, the amino acid sequence of FVA has been determined, and exhibits high homology with a lectin-like protein Ling Zhi-8 (LZ-8) isolated from another mushroom, *Ganederma lucidium*. Fig. 2-1 shows the sequences of FVA and LZ-8.

The lectin-like protein LZ-8, with mitogenic activity *in vitro* and immunomodulating activity *in vivo,* was isolated from *Ganoderma lucidium* by Kino et al.[3]. LZ-8 defined as immunomodulatory protein because of failure of monosaccharides and disaccharides to inhibit the agglutination reaction. From the facts that the amino acid sequence of LZ-8 has considerable similarity to

```
          1          10         20         30         40
FVA   Ac-SATSLTFQLAYLVKKIDFDYTPNWGRGTPSSYIDNLTFPK
LZ-8  Ac-SDTALIFRLAWDVKKLSFDYTPNWGRGNPNNFIDTVTFPK
         41         50         60         70         80
FVA      VLTDKKYSYRVVVNGSDLGVESNFAVTPSGGQTINFLQYN
LZ-8     VLTDKAYTYRVAVSGRNLGVKPSYAVESDGSQKVNFLEYN
         81         90        100        110   114
FVA      KGYGVADTKTIQVFVVIPDTGNSEEYIIAEWKKT
LZ-8     SGYGIADTNTIQVFVVDPDT--NNDFIIAQWN
```

Figure2-1 Complete amino acid sequence of FVA and LZ-8. FVA has 114 amino-acid residues, and contains no cysteine, methionine and histidine. N-terminus is blocked with acetylation.

that of the variable regions of immunoglobulin heavy chain and that LZ-8 has biological activities as an immunomodulator both *in vivo* and *in vitro*, Tanaka et al. has proposed that LZ-8 may be an ancestral protein of immunoglobulin superfamily[4]. FVA has a 65 % of sequence homology with LZ-8 and also has the highest sequence similarity of 24 % with immunoglobulin variable domain[30].

FVA has been classified as a lectin due to its hemagglutinating activity. However, the hemagglutinating activity of FVA is not inhibited by either monosaccharides or simple oligosaccharides, and FVA has shown some sequence similarity to immunoglobulins. Consequently, FVA may belong to a member of the immunoglobulin superfamily. The three-dimensional structure of FVA must play a key role to support this point of view. We carried out X-ray crystal analysis of FVA to elucidate the structural relationship with immunoglobulins and to discuss the evolutionary relationship with immunoglobulins and superfamilies.

2-3. Immunoglobulin and superfamily

Many domains of molecules which are involved in the cell surface recognition have been shown to exhibit sequence similarities with immunoglobulins[31]. These molecules are now classified as a member of the immunoglobulin superfamily with possible evolutionary relationships with immunoglobulin. On the other hand, most of the proteins or the glycoproteins that mediate cell-cell recognition or antigen recognition in the immune system

contain related structural elements, suggesting that the genes that encode them have a common evolutionary history. Included in this immunoglobulin superfamily are immunoglobulin, T cell receptors, MHC glycoproteins, the CD2, CD4 and CD8 cell-cell adhesion proteins, some of the polypeptide chains of the CD3 complex associated with T cell receptors, and the various Fc receptors on lymphocytes and other white blood cells. Each of these preceeding proteins contain one or more immunoglobulin-like domains. Each of these domains is typically about 100 amino acids in length and is thought to be folded into the characteristic sandwich-like structure made of two antiparallel β-sheets, usually stabilized by a conserved disulfide bond. Many of these molecules are dimer or higher oligomers in which immunoglobulin homology units of one chain interact with those in another.

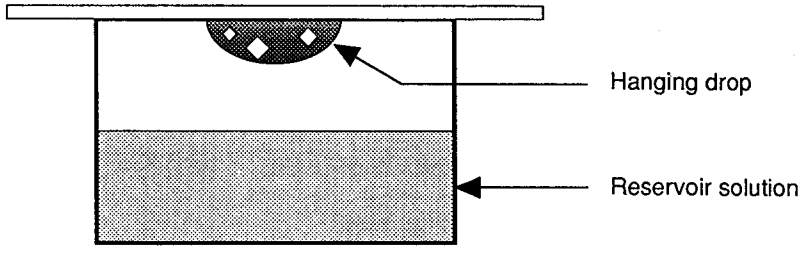
Each immunoglobulin homology unit is usually encoded by a separate exon, and it seems likely that the entire supergene family evolved from a gene coding for a single immunoglobulin homology unit which may have been involved in mediating cell-cell interactions. Since a Thy-1-like molecule has been isolated from the brain of squids, it is probable that such a primordial gene arose before vertebrates diverged from their invertebrate ancestors some 400 million years ago. New family members presumably arose by exon and gene duplications, and similar duplication events probably gave rise to the multiple gene segments that encode antibodies and T-cell receptors.

Chapter-3
Structure determination of FVA

3-1. Crystallization

Crystallization is the first important steps in X-ray structure analysis of protein. In this step, it is essential not only to crystallize the protein but also to grow the crystal to a substantial size.

FVA was purified at Hohnen Seiyu Co.Ltd. and purchased from Seikagaku Kogyo Co.Ltd.. The crystallization of FVA was carried out by the hanging drop vapor diffusion method[32]. Crystals were obtained from 1.8 M ammonium sulfate as a precipitate in 50 mM Phosphate buffer, 150 mM sodium chloride at pH 7.0. To improve the mosaic spread of the crystal, 10 mM galactose was added. The crystallization condition is summarized in Fig. 3-1. The shape of the crystal is a square thin pyramid, and shown in Fig. 3-2. The crystal grows up to an average size of 0.6 X 0.6 X 0.3 mm^3 in about one month. The crystals are stored in the solution containing 2.3 M ammonium sulfate, 50 mM phosphate buffer, 150 mM sodium chloride at pH 7.0. The crystallographic data were determined by the precession photographic method. The crystal belongs to a space group $P4_32_12$ with cell dimensions a = b = 97.7 Å, c = 62.2 Å. The crystallographic asymmetric unit contains two molecules. The crystallographic data of FVA are summarized in Table 3-1, and the precession photographs of (h k 0) plane and (h 0 l) plane are shown in Fig. 3-3.



Hanging drop	5 µl		Protein solution
	+ 5 µl		Reservoir solution
Reservoir solution	1.8	M	$(NH_4)_2SO_4$
	10.0	mM	Galactose
	50.0	mM	Phosphate buffer (pH 7.0)
	150.0	mM	NaCl
	0.03	%	NaN_3
Temperature	20 °c		

Figure 3-1 Crystallization of FVA
by the hanging drop vapor diffusion methods

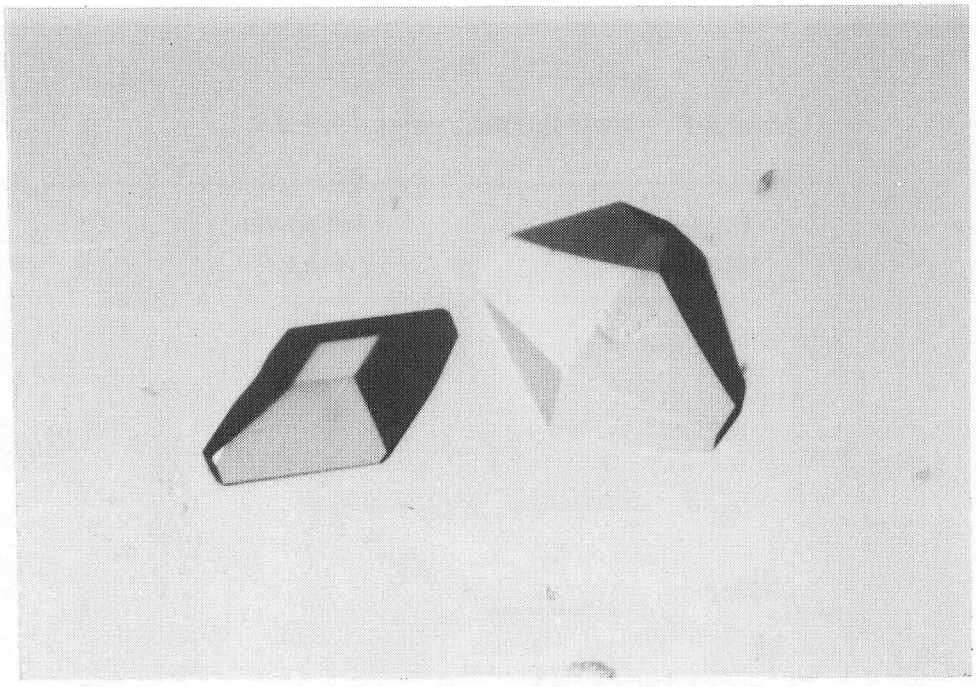
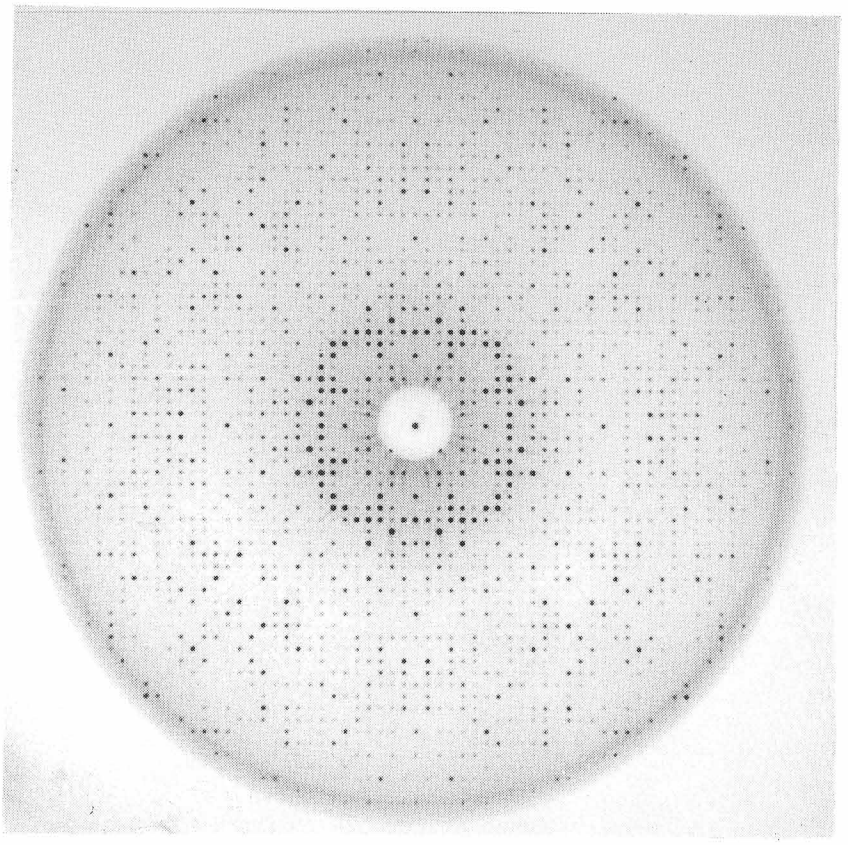


Figure 3-2 Crystal of FVA

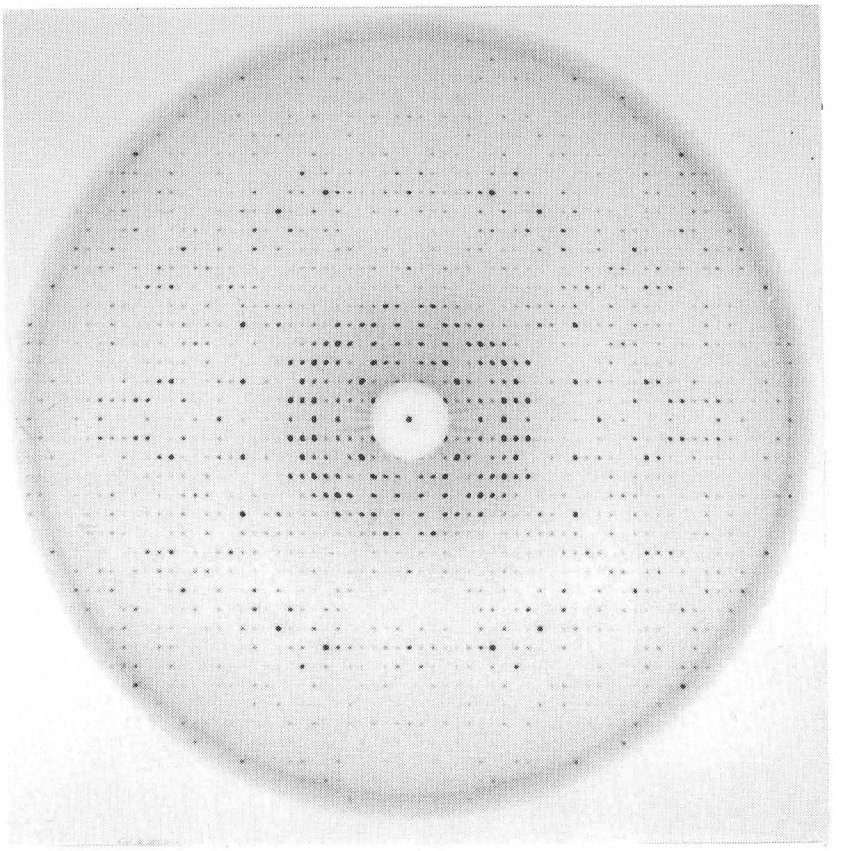
Table 3-1 Crystal parameters of FVA

Crystal system	Tetragonal	
Space group	$P4_32_12$	
Cell constants		
a =b=	97.7	Å
c=	62.2	Å
Volume		
V=	5.937×10^5	$Å^3$
Number of molecules in one unit cell		
Z=	16	
Volume per 1 dalton		
Vm=	2.52	$Å^3$/dalton
Volume of solvent		
Vsolv.=	46	%



(h k 0) plane

Figure 3-3 Precession photograph of an FVA crystal.
F = 100 mm, μ = 15°



(h 0 l) plane

Figure 3-3 continued

3-2. Heavy atom derivatives

In order to analyze the crystal structure of proteins, it is necessary to obtain heavy atom derivative crystals which are needed to determine the phases of structure-factor amplitudes.

Heavy atom derivatives were prepared by the conventional soaking method. Many kinds of heavy atom reagents were tested by varying the soaking conditions such as the concentration and soaking time. All the conditions used are listed in Table 3-2. The qualities of the soaked crystals were checked by the comparison of their diffraction patterns along the [1 0 0], [0 1 0], and [0 0 1] directions with those of the native crystals. If any change of diffraction patterns was observed, the diffraction data were collected at 5.0 Å resolution on a four-circle diffractometer, and the mean fractional isomorphous change in structure amplitudes and the difference Patterson map was calculated. Then three useful heavy atom derivatives were obtained. They were made by soaking the native crystals at pH 7.0 in a solution containing 2.3 M ammonium sulfate and the heavy atom compounds: 20 mM uranyl nitrate for 2 days, saturate phenylmercuric acetate for 3 days and 10 mM silver nitrate for 7 days. The soaking conditions are summarized in Table 3-3

3-3. Data Collection

In the present analysis of FVA, the intensities of the high resolution reflections are weak. Therefore, other strategies of data collection are necessary to collect the higher resolution data. In this case, the intensity data

Table 3-2 List of the compounds checked for heavy atom derivatives

Compounds	Concentration	Soaking time	R_F
$UO_2(NO_3)_2$†	20 mM	2 days	25.0 %
$UO_2(CH_3CO_2)_2$	20 mM	4 days	27.9 %
$K_3UO_2F_5$	20 mM	9 days	8.9 %
$SmCl_3$	5 mM	3 days	–
$Sm(NO_3)_3$	5 mM	2 days	–
K_2PtCl_4	1 mM	7 days	–
K_2PtCl_6	1 mM	7 days	–
$K_2Pt(NO_3)_4$	5 mM	7 days	–
C_2H_5HgCl	saturated	7 days	–
CH_3HgCl	5 mM	3 days	–
$HgCl_2$	10 mM	5 days	–
HgI_2	5 mM	3 days	9.2 %
E.M.T.S.	5 mM	3 days	–
P.M.A.†	saturated	3 days	11.8 %
Mercurochrome	5 mM	4 days	–
$AgNO_3$†	10 mM	4 days	23.7 %
Shotgun Solution‡		4 days	–
I_2 + KI	10 mM	4 days	–
CdI_2	5 mM	4 days	–
WO_3	saturated	4 days	–
K_2OsCl_6	saturated	3 days	–
$OsCl_3$	saturated	5 days	–

† The compound used for the MIR analysis.

‡ Shotgun Solution consists of 1mM $DyCl_3$, 1mM CsCl, 1mM $LuCl_3$, 1mM $PrCl_3$ and 1 mM Na_3IrCl_6.

$$R_F = 2\sum ||F_{PH}|-|F_P|| / \sum (|F_{PH}|+|F_P|)$$

E.M.T.S. : Ethyl Mercuric ThioSalicylate

P.M.A. : Phenyl Mercuric Acetate

Table 3-3 Soaking conditions of the three derivatives

	soaking	
componds	concentration	time
$UO_2(NO_3)_2$	20mM	2 day
P.M.A.	saturated	3 day
$AgNO_3$	10mM	1 week

All compounds were dissolved in 2.3 M $(NH_4)_2SO_4$ (pH 7.0) solvent with no buffer.

P.M.A. : Phenyl Mercuric Acetate

was collected by using three different methods. One was the four-circle diffractometer methods, another was the Weissenberg methods by using the synchrotron radiation source and Imaging plate, and the other was the oscillation methods by using the IP-system.

3-3-1. Four-circle diffractometer

The X-ray diffraction data for multiple isomorphous replacement analysis were collected by using a four-circle diffractometer RIGAKU AFC-5 equipped with a rotating anode X-ray generator RIGAKU RU-300. X-ray generator was operated at 40 kV and 300 mA and Cu-Kα radiation was monochromatized by a Ni-filter. Data collection was carried out in a room kept at 15 °C. The intensity of diffraction peak was counted by the ω-scan mode with scan speed for the range from 6.0 to 2.0 °/min, and the background of the peak was counted for the range from 3.0 to 6.0 seconds. All crystals for the data collection were used until the average intensity of three reference reflections decreased 10 % of the initial value due to radiation damage. The X-ray diffraction data of the native crystal was collected up to a resolution of 2.5 Å. The X-ray diffraction data of three heavy-atom derivative crystals, UO_2-derivative, Hg-derivative and Ag-derivative, were collected up to a resolution of 2.8 Å. To use the anomalous dispersion effects, Bijvoet-pair data of UO_2-derivative were collected. The statistical data were summarized in Table 3-4.

Table 3-4 Summary of the data collected by four-circle diffractometer

Crystal	Resolution Å	No. of unique reflections	R_{merge} %	R_F %
Native	2.5	15398	5.8	-
$UO_2(NO_3)_2$ ‡	2.8	15946	7.2	21.9
P.M.A	2.8	7445	6.8	10.8
$AgNO_3$	2.8	5333	4.6	23.5

‡$UO_2(NO_3)_2$ data was contains Bijvoet-pairs.
P.M.A. : Phenyl Mercuric Acetate

$$R_{merge} = \sum | F_i - \langle F \rangle | / \sum \langle F \rangle$$

$$R_F = 2\sum | |F_{PH}| - |F_P| | / \sum (|F_{PH}| + |F_P|)$$

3-3-2. Synchrotron radiation source

The native data were collected up to a resolution of 2.0 Å using the synchrotron radiation source at Photon Factory, National Laboratory for High Energy Physics, Tsukuba. Synchrotron radiation was used at a wavelength of 1.04 Å. The X-ray diffraction data were collected by using the Weissenberg camera with the camera radius of 430 mm[33,34]. The X-ray diffraction patterns were recorded on Fuji Film Imaging Plate with the size of 200 x 400 mm. Data collection was carried out in a room kept at 15 °C. The crystal was mounted with the *a*-axis* parallel to the spindle axis of the Weissenberg camera. Twenty-one imaging plates were used to record all diffraction spots of the independent region in the reciprocal space. Respective oscillation ranges were recorded with the oscillation angle of 6.0 °, the film slide of 4.0 mm and the oscillation speed of 4.0 °/min. The read-out of the data from the imaging plate was carried out by using Fuji Film BA-100 system. The read-out data were processed by the processing program WEIS[35] and scaled by the scaling program SCALE. The statistics of the data collection by the synchrotron radiation are summarized in Table 3-5.

3-3-3. Imaging plate system

The native data were also collected up to a resolution of 1.88 Å using the IP-system equipped with a rotating anode X-ray generator RIGAKU RU-200. X-ray generator was operated at 40 kV and 100 mA and Cu-Kα radiation was focused by the mirror-mirror optics. The details of the IP-system were described

Table 3-5 Summary of the data collected by syncrotron radiation source

Synchrotron	Photon factory	
Beam line	BL-6A$_2$	
Camera	Wissenberg camera for macromolecular crystallography	
Wave length	1.04	Å
Camera length	430	mm
Oscillation angle	6.0	degree
Film movement	4.0	mm
Number of imaging plates	21	
Exposure time	30.0	seconds
Resolution	2.0	Å
Number of recorded reflections	56219	
Number of unique reflections	15603	
R_{merge}	5.12	%

$$R_{merge} = \sum_h \sum_i | I_h - (I_{h,i} / G_i) | / \sum_h \sum_i I_h$$

in Chapter 1-2.

The crystal was mounted with the *a*-axis* parallel to the spindle axis of the oscillation camera. Seventy-two imaging plates were used to record all diffraction spots of the independent region in the reciprocal space. Respective oscillation ranges were recorded with the oscillation angle of 1.3 °, the camera length of 85 mm and the exposure time of 25 minutes. The statistics of the data collection by the IP-system are summarized in Table 3-6.

3-4. Phase determination

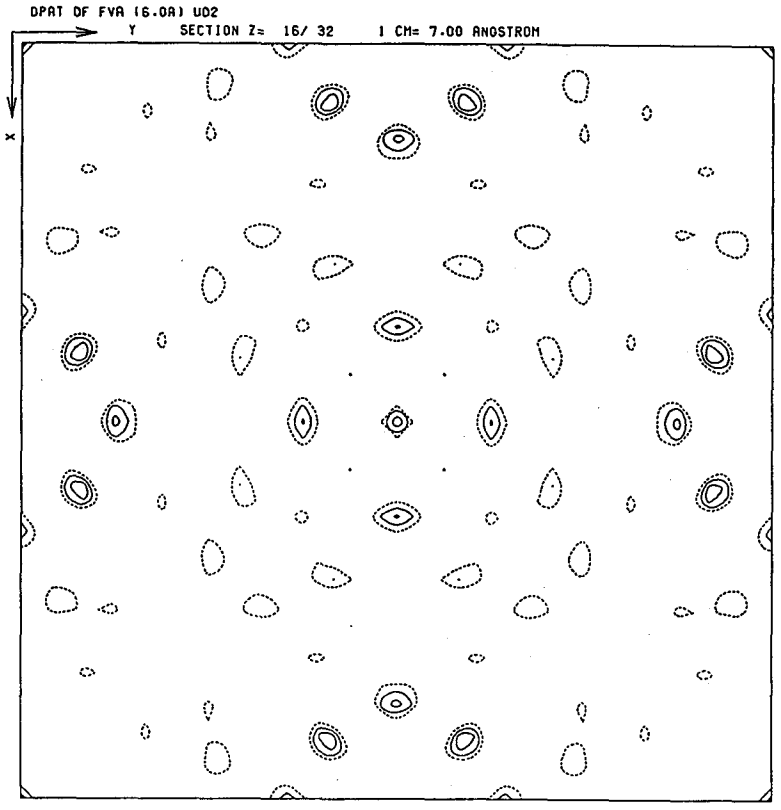
The phase determination has been carried out by the multiple isomorphous replacement (MIR) method at a resolution of 2.8 Å using the three heavy atom derivatives. Analysis of the MIR method was started with the determination of heavy-atom positions by using the data described in section 3-2 and 3-3. Difference Patterson function[36] is applied to locate heavy-atoms in protein crystals [37,38]. The coefficient of the difference Patterson function is $(F_{PH}-F_P)^2$, where F_{PH} and F_P are the structure factor amplitudes of heavy-atom derivative and native crystals, respectively.

The Harker sections of difference Patterson maps for the three derivatives at a resolution of 6.0 Å are shown in Fig. 3-4. One of the heavy-atom positions of UO_2-derivative was decided by inspection of the difference Patterson map. No other sites could be found at this stage, and later other two minor sites were found in the difference Fourier map calculated with the best phases[39] obtained from the calculation with first site of UO_2. The heavy atom positions in the Hg-

Table 3-6 Summary of data collected by IP-system

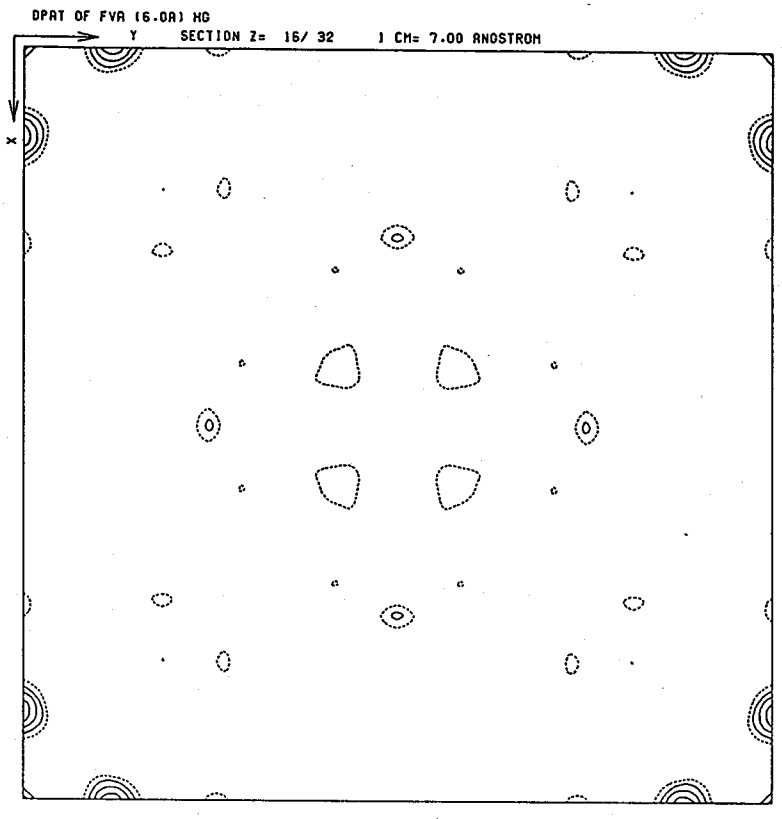
Detector	IP-system	
Camera length	85.0	mm
Oscillation angle	1.3	degree
Number of imaging plates	72	
Exposure time	25	minuts
Resolution	1.88	Å
No of recorded reflections	99867	
No. of unique reflections	20445	
Efficiency	81	%
R_{merge}(full)	5.24	%
R_{merge}(partial)	5.63	%
R_{merge}(total)	5.36	%

$$R_{merge} = \Sigma \Sigma \mid I_i / G_i - \langle I \rangle \mid / \Sigma \langle I \rangle$$



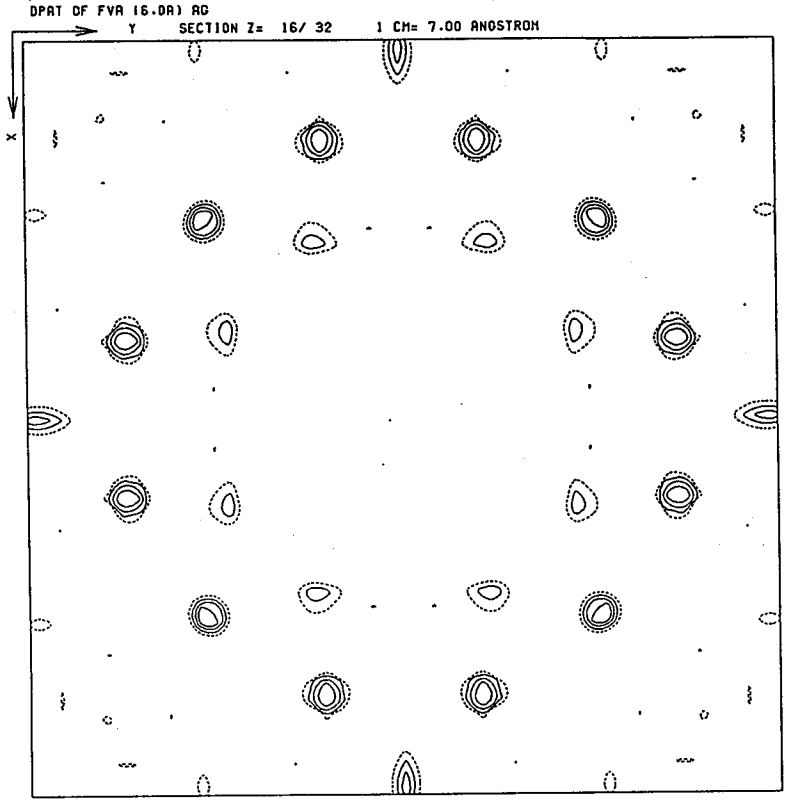
$UO_2(NO_3)_2$-derivative

Figure 3-4 The Harker section (W=1/2) of difference Patterson maps. All maps were calculated by the diffraction data collected by the four-circle diffractometer.



Phenyl Mercuric Acetate-derivative

Figure 3-4 Continued
The Harker section (W=1/2) of difference Patterson maps.



$AgNO_3$-derivative

Figure 3-4 Continued
The Harker section (W=1/2) of difference Patterson maps.

Table 3-7 Refined heavy atom parameters of FVA derivatives

Compound	site	x	y	z	Occupancy	B (Å^2)
$UO_2(NO_3)_2$	U-1	0.188	0.241	0.105	0.790	16.2
	U-2	0.517	0.009	0.081	0.491	25.9
	U-3	0.623	0.659	0.058	0.265	24.5
P.M.A.*	Hg-1	0.059	0.494	0.075	0.353	29.0
$AgNO_3$	Ag-1	0.437	0.198	0.246	0.768	11.2

x, y, z give the fractional coordinate in the unit cell. B gives the temperature factor.

* P.M.A. is Phenyl Mercuric Acetate.

derivative and Ag-derivative were decided by the difference Fourier maps. Atomic coordinates, occupancy and isotropic temperature factors of these heavy-atom positions were refined by the usual alternating least-square method.

The enantiomorph of the crystal, the space group of $P4_12_12$ or $P4_32_12$, was determined by using the anomalous dispersion effect from the uranyl derivative. The statistical data for the MIR phases calculated with the space group of $P4_32_12$ were better than that of $P4_12_12$. Finally the heavy atom parameters were refined at a resolution of 2.8 Å. The final heavy atom parameters are listed in Table 3-7. A mean figure of merit was 0.74 for 4691 reflections with $F/\sigma(F) > 3.0$ from a 15.0 to 2.8 Å resolution range.

3-5. Improvement of electron density map

Since the FVA crystal had two identical subunits in an asymmetric unit, averaging the electron density over the two subunits was applicable to enhance the signal-to-noise ratio of the protein density[40,41]. Generally, some identical subunits may differ only in orientation and position. It is possible to define an operation involving only rotation and translation. This means that it is necessary to determine subunit orientation and position for averaging the electron density.

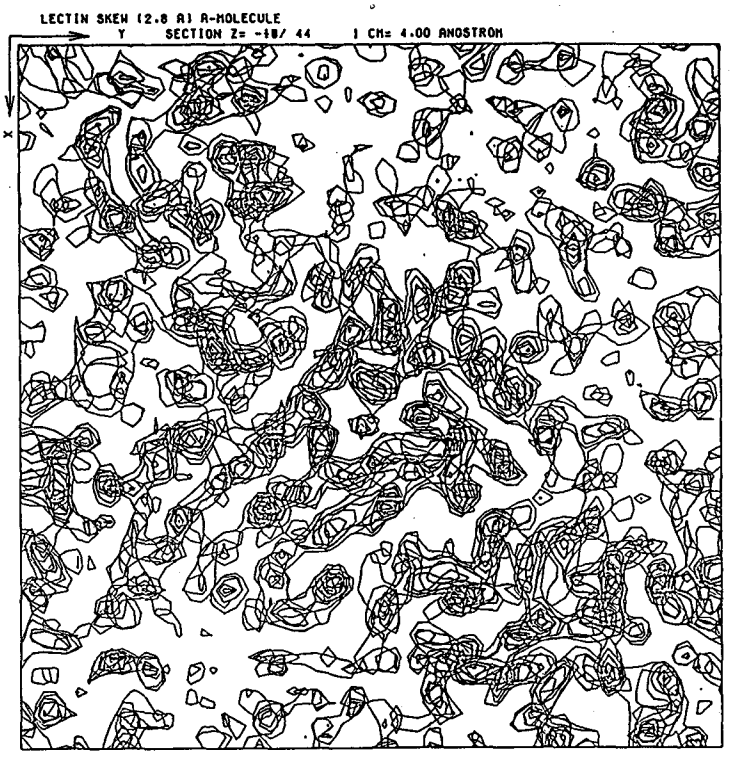
The best phase map showed clearly the two asymmetric molecules which are related each other by a non-crystallographic two-fold axis. To obtain a geometry of the non-crystallographic two-fold axis, the relations of the heavy-

Table 3-8 Refined rotation parameters between each subunit

Orign compound	Orign			Rotation			correlation coefficient
	X	Y	Z	θ	φ	ψ	
Ag-1	0.698	0.062	0.498	32.53	178.67	179.67	62.7 %
U-1	0.763	-0.188	0.392	57.55	0.19	181.20	61.8 %

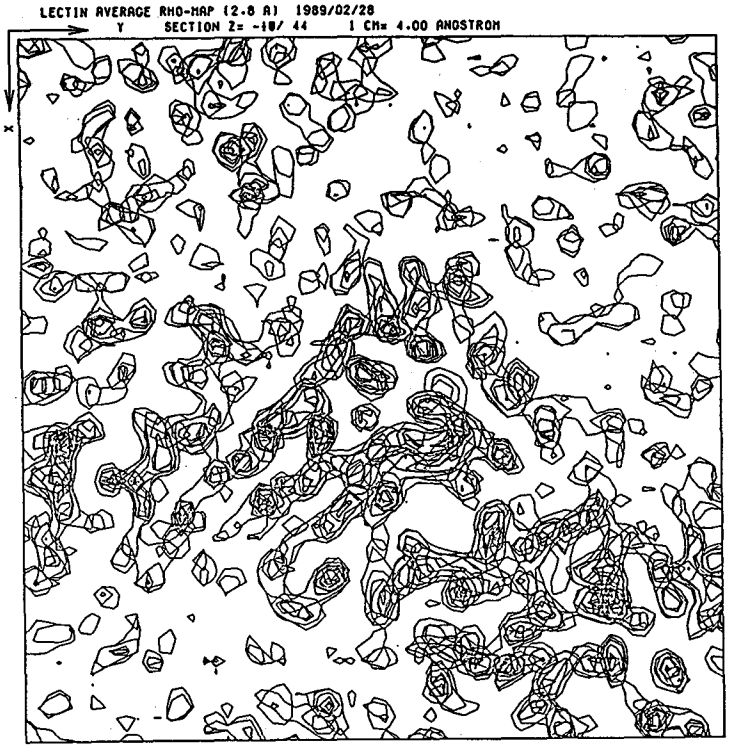
Orign; X, Y, and Z are given by the fractional coordinates in a unit cell.

Rotation; θ, φ, and ψ are given by Eulerian angles.



(a) The original electron density map

Figure 3-5 The improvement of the electron density map by averaging based on the local two-fold axis.
(a) The original electron density map caluculated at 2.8 Å resolution by the M.I.R. phase.
(b) The improved electron density map averaged by the local two fold axis.



(b) The improved electron density map

Figure 3-5 Continued

atom binding sites were investigated. The U-2 site and the U-3 site were equivalent in the electron density map. Furthermore the Ag-1 site and the U-1 site were placed at the midpoint of the U-2 and U-3 sites and passed through the non-crystallographic two-fold axis. Then the rotation and translation parameters were refined by the least-squares methods to obtain the maximal correlation among the electron densities of the two subunits. Table 3-8 summarizes these rotation and translation parameters at both origins of the Ag-1 site and the U-1 site. Finally the electron density map was averaged, based on the local two-fold axis through the Ag-site. The averaged map was cleared of the noise of the solvent region, and significantly improved from the initial unaveraged maps of respective molecules. Fig. 3-5 shows the improvement of electron density map.

3-6. Model building

The averaged electron density map at a resolution of 2.8 Å clearly showed a polypeptide chain folding and some of the side-chain orientations. Initially, the polypeptide chain folding was traced on a minimap made by stacking transparent sheets of electron density maps. Interpretation of the averaged electron density map was straight forward.

The model building procedure was carried out using the model building program FRODO[42], using the three-dimensional graphics display Evans and Sutherland PS390 linked to a Micro-VAX II computer. The detail of FRODO was

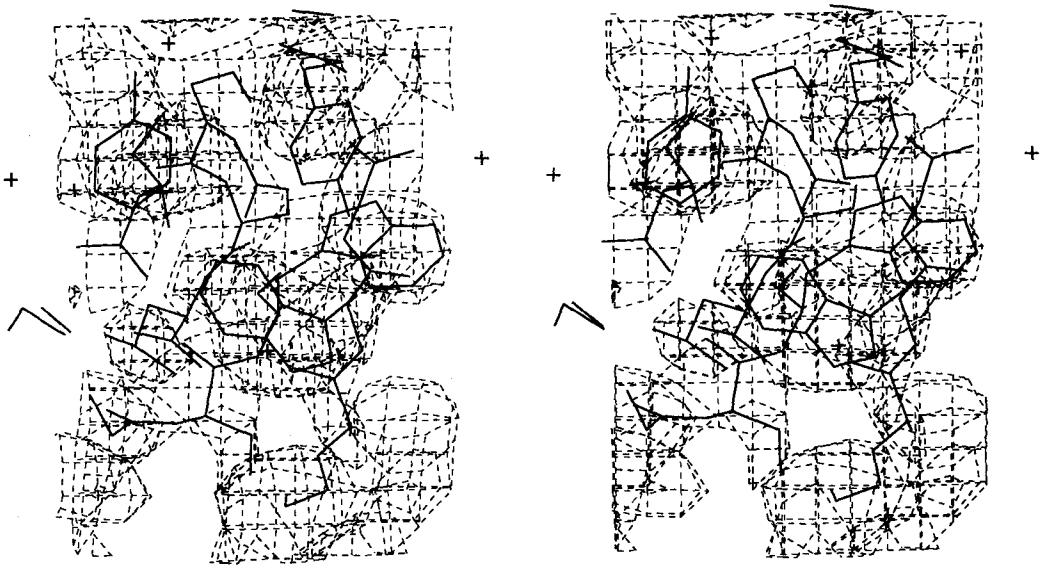


Figure 3-6 The initial model with the improved electron density map.

described at Chapter 1-3.

The model building procedure was straight forward as many secondary structure regions were found on the averaged electron density map. Only two loop regions, one loop from Thr-28 to Phe-34 and the other from Pro-98 to Ser-103, showed very weak electron density and these regions were difficult to assign atomic models. These regions were constructed considering geometries of backbone atoms. The initial model with the averaged MIR electron density map is shown in Fig. 3-6.

3-7. Refinement

A structure refinement was carried out with two programs. One is the Hendrickson and Konnert restrained least-squares program PROLSQ[43], the other is the general-purpose macromolecular refinement program XPLOR[16,23]. The former was widely used by the members of our laboratory, but the latter was recently adapted to the main frame computer ACOS-S930 and VAX-station 3200 computer, and used for the first time. The details of the program XPLOR are described in Chapter 1-4.

3-7-1. Theoretical background of PROLSQ

In this refinement, the least-square solution provides the shifts in the set of model parameters that minimize the sum of the weighted squares of differences between the observed and calculated values for both structure factors and

stereochemical parameters. The function that is minimized is

$$\Phi = \sum_{i=1}^{NREF} W_i(F_{Oi}-F_{Ci})^2 + \sum_{i=1}^{NBOND} W_i(D_{Oi}-D_{Ii})^2 + \sum_{i=1}^{NANG} W_i(D_{Oi}-D_{Ii})^2$$
$$+ \sum_{i=1}^{NPLN} W_i(D_{Oi}-D_{Ii})^2 + \sum_{i=1}^{NCHIRAL} W_i(V_{Oi}-V_{Ii})^2 + \sum_{i=1}^{NVDW} W_i(D_{Oi}-D_{Ii})^2$$
$$+ \sum_{i=1}^{NTHERM} W_i(B_{Oi}-B_{Ti})^2 + \sum_{i=1}^{NCRYST} W_i(|\vec{r}_{i,j}-\vec{r}_i|)^2 + \sum_{i=1}^{NCRYST} W_i(B_{i,j}-\langle B_i\rangle)^2$$

where the first term is the diffraction term and is the function usually minimized in the refinement of small molecules. The second and third terms minimize the deviation of bond distances and angle-defining distances from their ideal values. The forth term minimizes the deviation from planarity of groups of atoms expected to be planar. The fifth term minimizes the deviations from the ideal chiral volumes required for given chiral centers, thus ensuring the proper handedness. The sixth term prevents unreasonably short non-bonded contacts by minimizing the deviations from ideal contact distance; it is applied only to those contact distances which are deemed too short. The seventh term minimizes the unreasonable discrepancies between the thermal parameters of the neighboring atoms which are covalently bonded. The last two terms are useful for non-crystallographic symmetry restraints. The eighth term minimizes the deviations from the averaged structure over the two identical subunit. The last term imposes the restraint of thermal parameters for these similar atoms. All quantities subscripted with "i" are ideal values for the parameters as determined from crystal structure of small molecules. Quantities

subscripted with "o" refer to the current values of the parameters in the trial model. "w_i" denotes the weighting factor. It is obvious that the minimization of Φ with the proper weights should produce a structure that agrees well with the observed data and is consistent with almost everything known about the protein's structural chemistry.

3-7-2. Refinement survey

The refinement of the structure was mainly divided into three stages. First, the refinement was carried out against the data collected by the four-circle diffractometer at a resolution of 2.5 Å. Second, the refinement was carried out against the data collected by the synchrotron radiation source at a resolution of 2.0 Å. Third, the refinement was carried out against the data collected by the IP-system at a resolution of 1.88 Å and the final model of FVA was refined at a resolution of 1.88 Å.

In the first stage, the refinement was performed at the starting point using the data collected by the four-circle diffractometer at a resolution of 2.8 Å. The crystallographic R-factor for the initial model was 45.0 %, and after several cycles of PROLSQ and manual corrections using FRODO against the (2Fo - Fc) and calculated phase map, the crystallographic R-factor decreased to 25.0 %. At this stage the resolution was extended from 2.8 Å to 2.5 Å by introducing the individual temperature factors for all individual atoms of the two-asymmetric molecules. By adding 106 water molecules, the crystallographic R-factor

Table 3-9 r.m.s. deviations of the model at 2.0 Å resolution

Parameter	target	r.m.s. deviation	
Bond distance			
1-2 bond	0.020	0.021	Å
1-3 angle	0.030	0.040	Å
1-4 angle	0.050	0.052	Å
Planes	0.020	0.019	Å
Chiral center	0.150	0.190	$Å^3$
van der Waals contacts			
Single torsion bonds	0.400	0.239	Å
Multiple torsion bonds	0.400	0.326	Å
Possible hydrogen bonds	0.400	0.236	Å
Torsion angles			
Planar (0,180)	3.0	2.9	degree
Staggered (60,180)	15.0	25.4	degree
Orthonormal (90)	20.0	30.0	degree
Noncrystallographic symmetry			
Tight position	0.050	0.090	Å
Medium position	0.500	0.651	Å
Weak position	5.000	1.130	Å
Tight thermal	0.500	0.300	$Å^2$
Medium thermal	2.000	1.440	$Å^2$
Weak thermal	5.000	2.473	$Å^2$
Temperature factors			
1-2 main chain	1.000	1.045	$Å^2$
1-3 main chain	1.500	1.810	$Å^2$
1-2 side chain	1.500	1.620	$Å^2$
1-3 side chain	2.000	2.338	$Å^2$

All parameters are used in the refinement program, PROLSQ.

converged to 22.0 %.

In the second stage, the data that was collected at a resolution of 2.0 Å by using the synchrotron radiation was introduced. In this data-set, the refinement against the final model of first stage was started at a resolution of 2.5 Å by use of PROLSQ. The resolution was extended from 2.5 Å to 2.0 Å stepwise. After several cycles of the PROLSQ refinement and manual corrections using FRODO against the (2Fo - Fc) and calculated phase map, the crystallographic R-factor converged to 19.8 % for 11604 reflections with F/ σ(F) > 3 including 159 water molecules. The refinement program was switched from PROLSQ to XPLOR, and one cycle of Slow-Cooling protocol refinement was carried out. After the one cycle of XPLOR refinement, the crystallographic R-factor converged to 19.5 % for 12476 reflections including 124 water molecules. After 15 cycles of PROLSQ for stereochemical refinement, the crystallographic R-factor converged to 18.5 % for 11604 reflections with F/ σ(F) > 3 including 124 water molecules, and the r.m.s. deviations from the ideal geometry at this stage are given in Table 3-10.

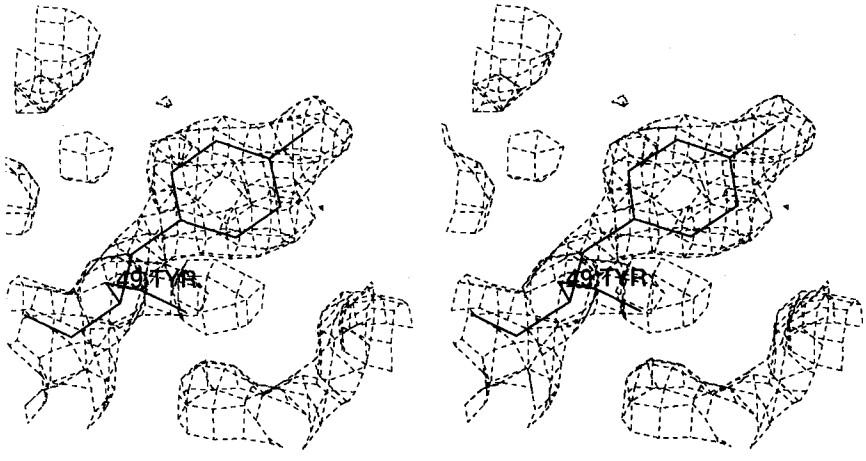
In the third stage for the final refinement at a resolution of 1.88 Å, data collected with the IP-system was introduced. One cycle of Slow-Cooling protocol refinement was carried out at a resolution of 1.88 Å. After the one cycle of XPLOR refinement, the crystallographic R-factor converged to 24.4 % for 18956 reflections including 124 water molecules. After 20 cycles of the PROLSQ refinement and manual corrections using FRODO against the (2Fo - Fc) and calculated phase map, the crystallographic R-factor converged to 19.9 % for 15752 reflections with F/ σ(F) > 3 including 129 water molecules.

Table 3-10 r.m.s. deviations of the final model (1.88 Å resolution)

Parameter	target	r.m.s. deviation	
Bond distance			
1-2 bond	0.020	0.021	Å
1-3 angle	0.030	0.038	Å
1-4 angle	0.050	0.050	Å
Planes	0.020	0.018	Å
Chiral center	0.150	0.152	$Å^3$
van der Waals contacts			
Single torsion bonds	0.400	0.209	Å
Multiple torsion bonds	0.400	0.254	Å
Possible hydrogen bonds	0.400	0.246	Å
Torsion angles			
Planar (0,180)	3.0	2.9	degree
Staggered (60,180)	15.0	21.3	degree
Orthonormal (90)	20.0	25.3	degree
Noncrystallographic symmetry			
Tight position	0.050	0.069	Å
Medium position	0.500	0.316	Å
Weak position	5.000	0.620	Å
Tight thermal	0.500	0.341	$Å^2$
Medium thermal	2.000	1.955	$Å^2$
Weak thermal	5.000	3.626	$Å^2$
Temperature factors			
1-2 main chain	1.000	1.334	$Å^2$
1-3 main chain	1.500	2.302	$Å^2$
1-2 side chain	1.500	1.876	$Å^2$
1-3 side chain	2.000	2.823	$Å^2$

All parameters are used in the refinement program, PROLSQ.

(a) Arg-26



(b) Tyr-49

Figure 3-7 The fitness of the final model for the 2Fo - Fc map at resolution of 1.88 Å at (a) Arg-26 and (b) Tyr-49. The map was obtained from calculated phase information (*R* -factor =0.199) with coefficient of 2Fo - Fc.

The final model has the r.m.s. deviations of 0.021 Å and 2.6° from ideal bond length and ideal bond angle, respectively. The target sigma and final r.m.s. deviations from the ideal geometry are given in Table 3-10. The r.m.s. deviations of the respective stereochemical restraint parameters are in good agreement with the target sigma values, so the final model has a stereochemical accuracy. Fig. 3-7 shows the fitness of the final model for the (2Fo - Fc) and calculated phase map.

3-8. Structural identity in subunit

There is a single FVA dimer in the asymmetric unit of the crystal. It is therefore necessary to show the structural identity between the two subunits in order to discuss the molecular structure. One subunit was superimposed on the other subunit by the least-squares methods as the deviation between the corresponding coordinates was minimized. The root-mean-square deviations between the each subunit are 0.248 Å for the 114 Cα atoms and 0.630 Å for all 904 atoms. The two independent subunits have almost the same structure with one another. The most disparate main-chain atoms are the N-terminal N-atom of Ser-1, which is 1.62 Å away from the other. The most disparate side-chain atoms are NH1-atom of Arg-50, which is 2.54 Å away from the other.

Chapter-4
Description of the structure

4-1. Structure of FVA

FVA is a dimeric molecule formed by two crystallographically independent subunits. Each subunit is a molecule of dimensions roughly 23 x 28 x 47 Å3. Fig.4-1 shows the atoms of one subunit, and Fig. 4-2 shows the α-carbon model of the subunit. Both subunit are almost equivalent in the conformation. The structure is predominantly characterized by a β-barrel structure composed by seven antiparallel β-strands.

The amino terminal nineteen residues (Ser-1 to Asp-19) protrudes from the barrel, and the first twelve residues (Ser-1 to Leu-12) fold into a short α-helix of three-turns followed by a β-strand of 5 residues (Lys-15 to Asp-19). The rest of the polypeptide chain, the 22th residue to the carbonyl terminal end, makes up the β-barrel structure comprising the seven antiparallel β-strands. The β-barrel of FVA subunit is composed of the immunoglobulin fold, and its folding pattern is more similar to that of the constant domain of an immunoglobulin than that of the variable domain. The schematic drawing of the FVA structure is shown in Fig.4-3.

The inside of the β-barrel is mainly composed of hydrophobic residues, such as Val-41,51,85,93, 95, Leu-36, 77 and Ile-33,74,91, 108. Therefore the β-barrel structure is stabilized by hydrophobic interaction. The two tryptophan

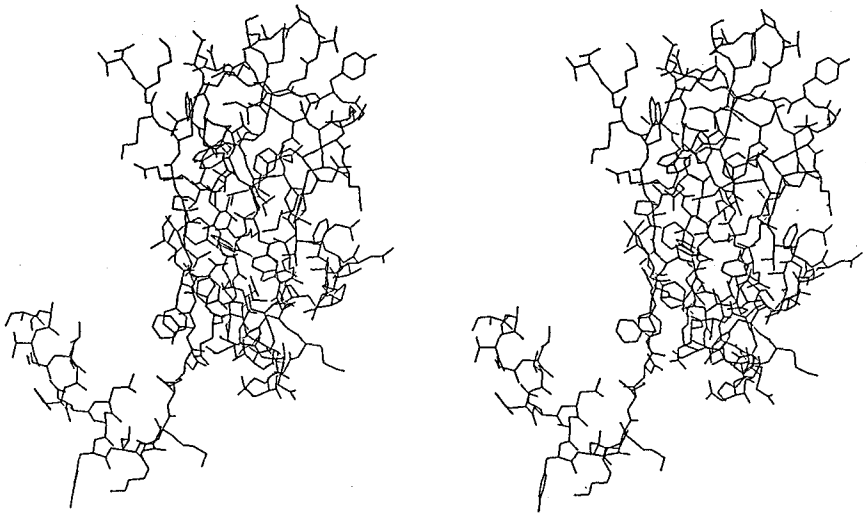


Figure 4-1 Stereo view of all atoms of FVA subunit.

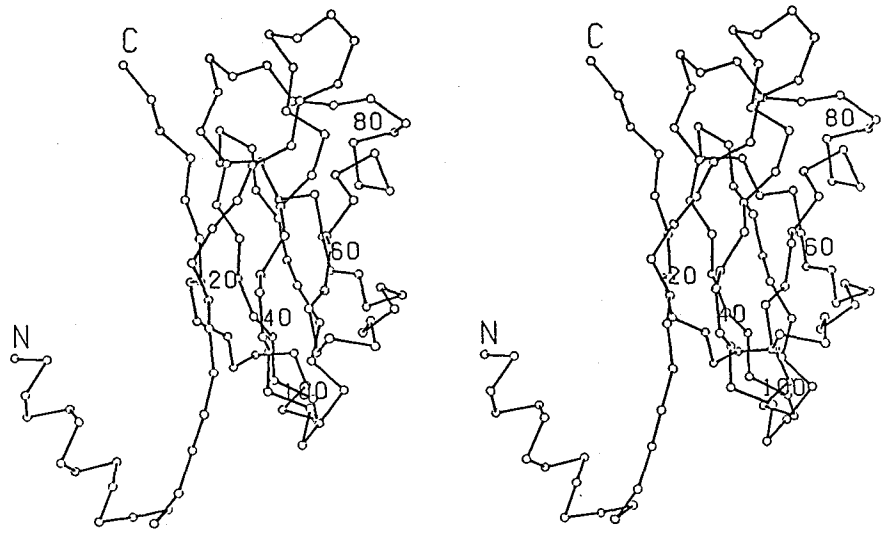


Figure 4-2 Stereo view of α-carbon backbone of FVA subunit. One molecule is mainly composed of seven-stranded β-barrel and N-terminal short α-helix.

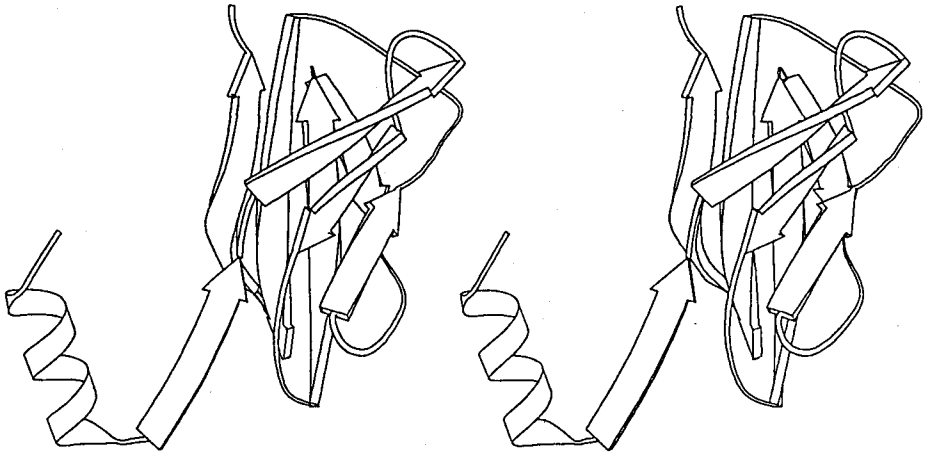


Figure 4-3 Stereo view of a schematic drawing. Arrows show β-sheets, ribbon helix shows α-helix.

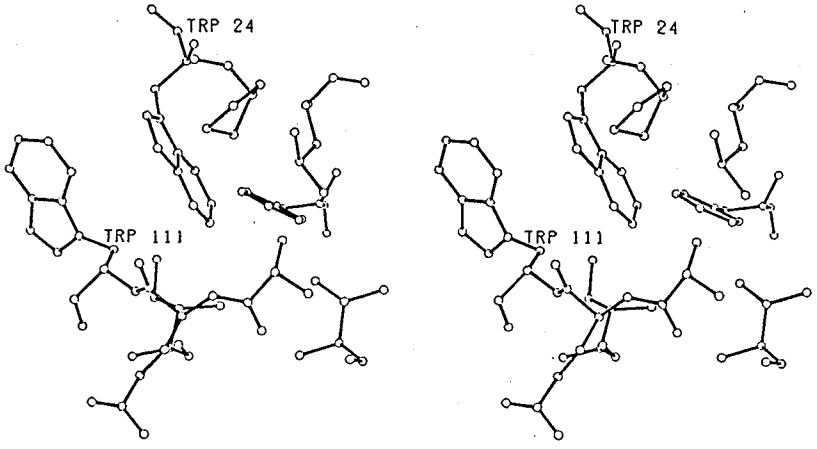


Figure 4-4 Stereo view of two tryptophans (Trp-24 and Trp-111). Each tryptophan is closely positioned near the surface of FVA, and has hydrophobic interactions with each other.

residues, Trp-24 in the first strand and Trp-111 in the seventh strand, are located near the surface of the subunit molecule and their side chains are in close contact by hydrophobic interaction as shown in Fig.4-4. This interaction might contribute in holding the β-barrel structure, because these tryptophans are placed at both ends of the β-barrel.

4-2. Secondary structures of FVA

The dihedral angles (ϕ,ψ) of the polypeptide backbone are listed in Table 4-1; these angles were calculated from the refined atomic parameters. The Ramachandran plot[44] is also shown in Fig. 4-5.

The secondary structures defined by torsion angles are listed in Table 4-2. Fig.4-6 shows topological connectivities between the secondary structures.

The first β-strand in the barrel forms hydrogen bonds with the second β-strand. The loop between the first β-strand and the second β-strand is named Loop1-2, and so on. There are some ambiguities in the side chain conformations in Loop1-2, because of the diffused electron densities of this loop. This reflects that atoms in Loop1-2 have high temperature factors. A cis-proline-29 is located at the start point of Loop1-2. The second β-strand has the hydrogen bonds with the fifth β-strand. Loop2-3 is across the β-barrel, and the loop forms a long and irregular structure. The third β-strand has hydrogen bonds with the fourth β-strand. Loop3-4 forms a short turn. The fourth β-strand is a short distorted β-strand, but makes two hydrogen bonds with the third β-strand. Loop4-5 seems to be a distorted β-strand (Ala-65 to Thr-72). Loop5-6

Table 4-1 List of dihedral-angles

Ressidue		Φ	Ψ	Ressidue		Φ	Ψ
1	Ser	-56	-178	36	Leu	-127	132
2	Ala	-71	-30	37	Thr	-120	129
3	Thr	-66	-44	38	Phe	-107	128
4	Ser	-61	-51	39	Pro	-59	-28
5	Leu	-59	-46	40	Lys	-164	102
6	Thr	-52	-51	41	Val	-91	163
7	Phe	-74	-37	42	Leu	-78	136
8	Gln	-68	-34	43	Thr	-124	29
9	Leu	-68	-52	44	Asp	-80	-20
10	Ala	-54	-47	45	Lys	-142	173
11	Tyr	-58	-44	46	Lys	-98	86
12	Leu	-74	-42	47	Tyr	-75	153
13	Val	-114	-5	48	Ser	-120	169
14	Lys	66	-53	49	Tyr	-132	136
15	Lys	-135	151	50	Arg	-105	135
16	Ile	-167	160	51	Val	-123	145
17	Asp	-156	121	52	Val	-136	125
18	Phe	-152	168	53	Val	-109	118
19	Asp	-153	100	54	Asn	51	47
20	Tyr	-110	27	55	Gly	80	-11
21	Thr	-66	131	56	Ser	-86	114
22	Pro	-85	111	57	Asp	-72	113
23	Asn	-103	113	58	Leu	-90	8
24	Trp	-93	129	59	Gly	86	-169
25	Gly	-98	148	60	Val	-111	133
26	Arg	-103	142	61	Glu	-135	153
27	Gly	-101	-151	62	Ser	-138	1
28	Thr	-126	122	63	Asn	-79	64
29	Pro	-74	161	64	Phe	-124	121
30	Ser	-63	11	65	Ala	-62	139
31	Ser	-93	-2	66	Val	-89	125
32	Tyr	-104	137	67	Thr	-84	170
33	Ile	-103	121	68	Pro	-49	-40
34	Asp	-101	-26	69	Ser	-78	-3
35	Asn	-158	169	70	Gly	93	0

Table 4-1 List of dihedral-angles (continued)

Ressidue		Φ	Ψ
71	Gly	-79	172
72	Gln	-161	129
73	Thr	-105	123
74	Ile	-96	113
75	Asn	-88	108
76	Phe	-67	-26
77	Leu	-58	-30
78	Gln	-68	-19
79	Tyr	-103	-45
80	Asn	-90	46
81	Lys	48	36
82	Gly	85	11
83	Tyr	-129	147
84	Gly	-73	177
85	Val	-111	130
86	Ala	-57	128
87	Asn	-69	6
88	Thr	-91	-18
89	Lys	-91	167
90	Thr	-99	124
91	Ile	-114	127
92	Gln	-118	127
93	Val	-116	134
94	Phe	-114	142
95	Val	-92	128
96	Val	-105	136
97	Ile	-88	129
98	Pro	-50	150
99	Asp	72	-142
100	Thr	58	-11
101	Gly	67	29
102	Asn	60	25
103	Ser	-104	-10
104	Glu	-107	118
105	Glu	-154	159
106	Tyr	-138	141
107	Ile	-67	145
108	Ile	-136	-9
109	Ala	-162	150
110	Glu	-150	130
111	Trp	-109	116
112	Lys	-117	-55
113	Lys	74	151
114	Thr	-164	-99

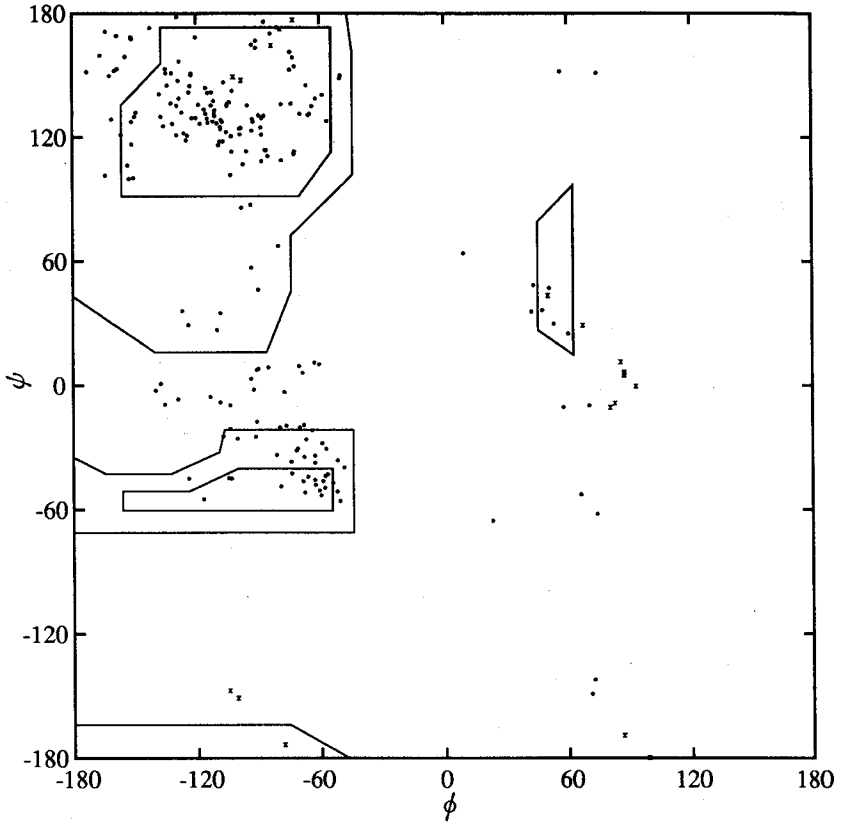


Figure 4-5 Ramachandran plot of FVA.

Table 4-2 Summary of FVA secondary structure

Name	Sequence No.			Length
Extended-region				
α1	2	-	14	3 turn-helix
β0	15	-	20	6
b-barrel				
β1	22	-	28	7
β2	33	-	39	7
β3	46	-	53	8
β4	56	-	61	6
β5	71	-	76	6
β6	90	-	98	9
β7	104	-	112	9

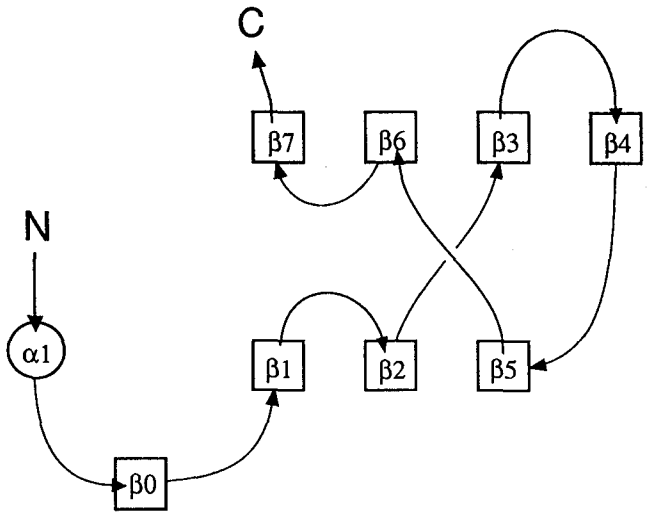


Figure 4-6 Topology of FVA. FVA is mainly composed of seven-stranded β-barrel structure (immunogulobulin fold), of which topology is essentially same with that of immunoglobulin constant domain. The β-barrel structure is divided into two β-sheets, which are three-stranded sheet (contains β1, β2 and β5) and four-stranded sheet (contains β3, β4, β6 and β7). In FVA fourth β-strand (β4) is hydrogen bonding with third strand (β3). N-terminal end additionally consists of short α-helix (α1) and β-strand (β0).

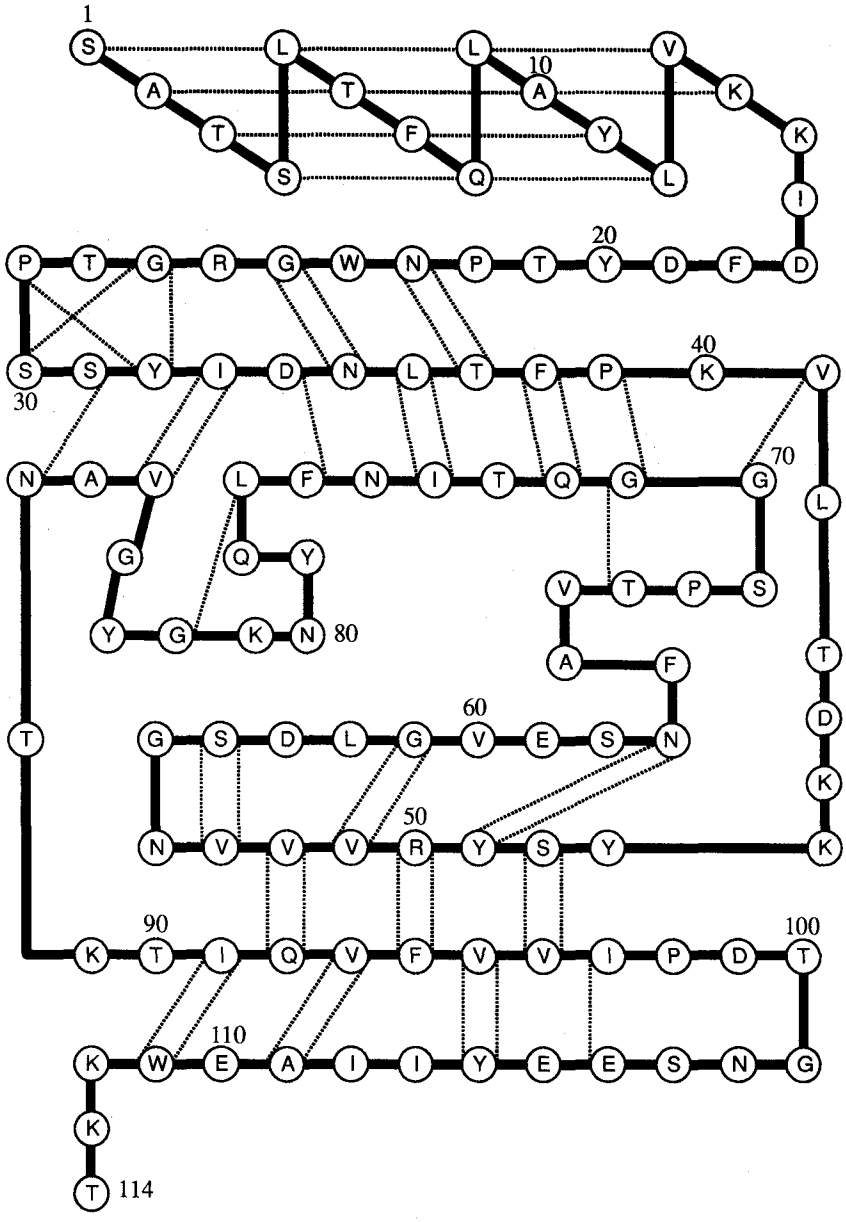


Figure 4-7 A schematic drawing of the main-chain to main-chain hydrogen bonds network of FVA.

Table. 4-3 List of main-chain to main-chain hydrogen-bond

Residue1			Residue2			d(Å)
1	Ser	O	5	Leu	N	3.21
2	Ala	O	6	Thr	N	2.94
3	Thr	O	7	Phe	N	3.19
4	Ser	O	8	Gln	N	2.97
5	Leu	O	9	Leu	N	3.01
6	Thr	O	10	Ala	N	2.88
7	Phe	O	11	Tyr	N	3.01
8	Gln	O	12	Leu	N	3.23
9	Leu	O	13	Val	N	2.79
10	Ala	O	14	Lys	N	2.84
23	Asn	N	37	Thr	O	3.12
23	Asn	O	37	Thr	N	2.75
25	Gly	N	35	Asn	O	2.82
25	Gly	O	35	Asn	N	3.66
27	Gly	N	32	Tyr	O	2.94
27	Gly	O	30	Ser	N	3.44
29	Pro	O	32	Tyr	N	3.19
31	Ser	O	87	Asn	N	2.83
33	Ile	N	85	Val	O	2.72
33	Ile	O	85	Val	N	2.56
34	Asp	O	76	Phe	N	2.59
36	Leu	N	74	Ile	O	2.89
36	Leu	O	74	Ile	N	2.84
38	Phe	N	72	Gln	O	2.94
38	Phe	O	72	Gln	N	3.10
39	Pro	O	71	Gly	N	3.39
41	Val	N	70	Gly	O	3.18
48	Ser	N	96	Val	O	3.31
48	Ser	O	96	Val	N	2.72
49	Tyr	N	61	Glu	O	2.97
49	Tyr	O	61	Glu	N	2.93

Residue1			Residue2			d(Å)
50	Arg	N	94	Phe	O	2.92
50	Arg	O	94	Phe	N	2.99
51	Val	N	59	Gly	O	2.88
51	Val	O	58	Gly	N	3.05
52	Val	N	92	Gln	O	3.03
52	Val	O	92	Gln	N	3.01
53	Val	N	56	Ser	O	3.04
53	Val	O	56	Ser	N	2.99
54	Asn	N	90	Thr	O	2.88
67	Thr	N	71	Gly	O	2.79
77	Leu	O	82	Gly	N	2.89
91	Ile	N	111	Trp	O	2.98
91	Ile	O	111	Trp	N	2.79
93	Val	N	109	Ala	O	2.84
93	Val	O	109	Ala	N	3.31
95	Val	N	106	Tyr	O	2.93
95	Val	O	106	Tyr	N	3.05
97	Ile	N	104	Glu	O	2.93

Table. 4-4 List of main-chain to side-chain hydrogen-bond

Residue1			Residue2			d(Å)
2	Ala	O	6	Thr	OG1	2.50
3	Thr	O	8	Gln	OE1	3.33
11	Tyr	O	14	Lys	NZ	2.81
21	Thr	N	19	Asp	OD1	2.91
21	Thr	O	23	Asn	OD1	3.09
22	Pro	O	24	Trp	NE1	3.96
24	Trp	O	28	Thr	OG1	3.60
25	Gly	N	35	Asn	OD1	3.31
27	Gly	N	34	Asp	OD1	3.87
27	Gly	N	34	Asp	OD2	3.61
30	Ser	O	26	Arg	NE	2.87
30	Ser	O	26	Arg	NH1	3.38
31	Ser	O	26	Arg	NH1	3.53
31	Ser	O	87	Asn	OD1	3.19
31	Ser	O	87	Asn	ND2	3.90
34	Asp	O	75	Asn	ND2	3.37
37	Thr	O	23	Asn	OD1	3.44
42	Leu	N	20	Tyr	OH	2.93
49	Tyr	O	79	Tyr	OH	2.63

Residue1			Residue2			d(Å)
51	Val	O	57	Asp	OD1	3.77
59	Gly	N	57	Asp	OD1	3.03
61	Glu	N	79	Tyr	OH	2.63
63	Asn	N	61	Glu	OE2	3.46
64	Phe	O	72	Gln	OE1	3.08
69	Ser	N	67	Thr	OG1	3.63
71	Gly	N	67	Thr	OG1	3.40
77	Leu	N	75	Asn	ND2	3.57
83	Tyr	O	80	Asn	ND2	2.41
84	Gly	O	80	Asn	ND2	2.66
87	Asn	O	111	Lys	NE1	2.90
89	Lys	O	111	Lys	NE1	3.82
90	Thr	N	54	Asn	ND2	2.96
100	Thr	O	103	Ser	OG	2.74

Table 4-5 List of side-chain to side-chain hydrogen-bond

Residue1			Residue2			d(Å)
4	Ser	OG	8	Gln	OE1	2.87
17	Asp	OD1	40	Lys	NZ	3.93
19	Asp	OD1	21	Thr	OG1	2.49
19	Asp	OD2	21	Thr	OG1	3.70
26	Arg	NH1	87	Asn	ND2	2.51
37	Thr	OG1	73	Thr	OG1	3.02
44	Asp	OD2	47	Tyr	OH	2.68
49	Tyr	OH	72	Gln	NE2	2.54
50	Arg	NH2	57	Asp	OD1	3.06
50	Arg	NE	105	Glu	OE1	2.82
50	Arg	NH1	105	Glu	OE2	2.98
54	Asn	ND2	90	Thr	OG1	3.88
61	Glu	OE1	78	Gln	OE1	3.30
67	Thr	OG1	69	Ser	OG	3.23
75	Asn	OD1	78	Gln	NE2	3.57
92	Gln	OE1	110	Glu	OE2	3.47
100	Thr	OG1	104	Glu	OE2	3.18
104	Glu	OE1	106	Tyr	OH	2.52
110	Glu	OE1	112	Lys	NZ	2.37

also is across the β-barrel, and the loop forms a long and irregular structure. The sixth β-strand has hydrogen bonds with the seventh β-strand. A hydrogen bond network is shown in Fig.4-7. Then the β-barrel is formed by the two facing β-sheets composed of a three stranded β-sheet; the first, second and fifth β-strands, and a four stranded β-sheet; the third, fourth, sixth and seventh. Main-chain to main-chain, side-chain to main-chain and side-chain to side-chain hydrogen bond pairs are listed in Table 4-3, Table 4-4, Table 4-5, respectively.

4-3. Temperature factors

Atomic temperature factors are also obtained after the refinement of the structure. Those values for 1937 atoms including water molecules range from 10.2 to 59.2 $Å^2$, and the average value is 28.5 $Å^2$. The temperature factors of the corresponding atoms in the two subunits are almost same. The highest temperature factor is found at the C-terminus atom in both subunits, and the atoms on the surface of the molecules have high temperature factors showing large displacements compared with those of the interior atoms. The interior atoms in the molecules are well-ordered with lower temperature factors. The temperature factors of Cα-atoms are shown in Fig. 4-8.

4-4. Dimer structure of FVA

Two subunits related by two local two-fold axes form dimeric contacts. One dimer contact is comparable to that of immunoglobulin H-L association as

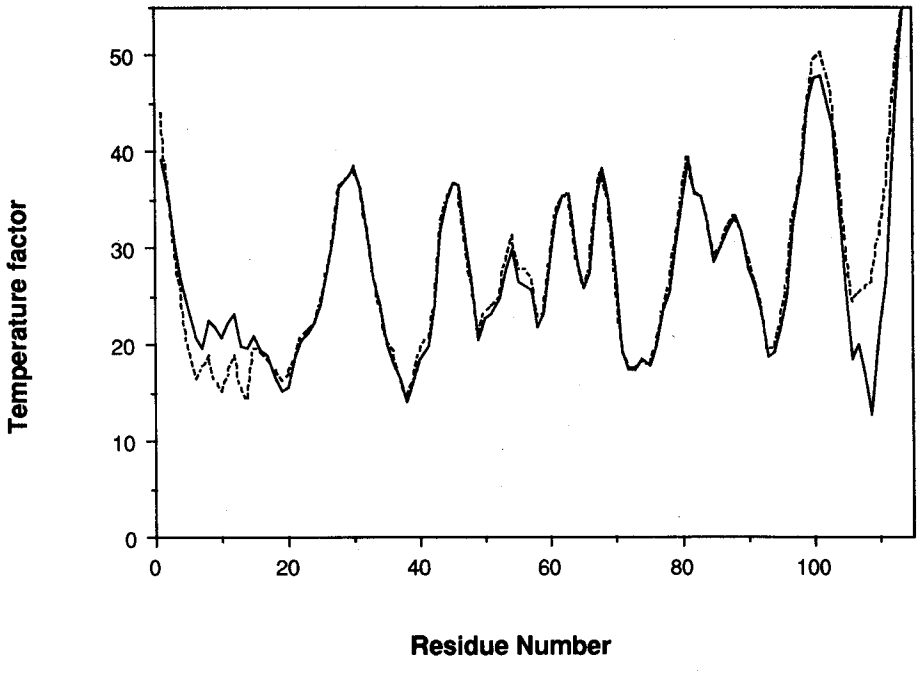


Figure 4-8 Temperature factor of the α-carbon atom of both subunits. The thick-line shows A-molecular, and the broken-line shows B-molecular.

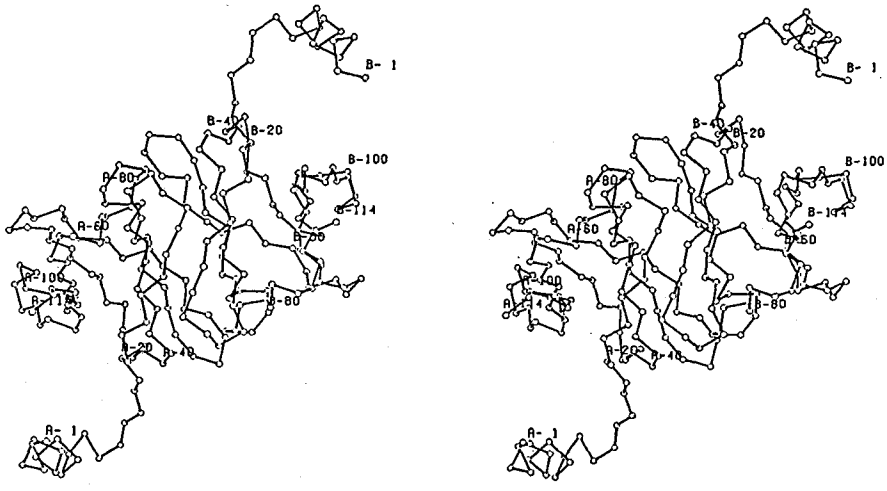


Figure 4-9 Stereo view of α-carbon backbone of FVA dimer looking down the local 2-fold axis. In this view two subunits contact at the β-sheet face.

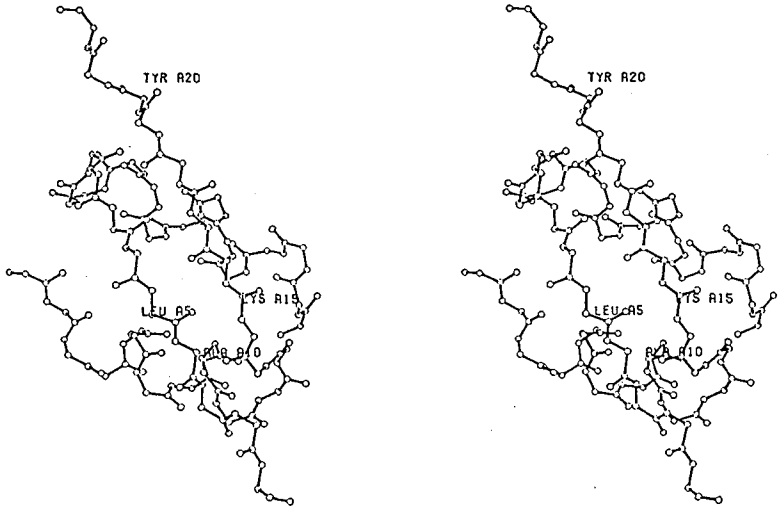


Figure 4-10 Main-chain of N-terminal dimer in the crystal. N-terminal 20 residues of two molecules are looking down the other local 2-fold axis. In this view, the two subunits contact with the β-strands and α-helixes which projected from β-barrel, and two β-strands of each molecule consist with one β-sheet.

Table 4-6 List of hydrogen-bond at dimer-interface (β-barrel dimer)

(a) Dimer interface between the residues

Residue1 (A-subunit)			Residue2 (B-subunit)			d(Å)
24	Trp	O	28	Thr	OG1	3.60
26	Arg	N	26	Arg	O	2.86
26	Arg	O	26	Arg	N	2.97
28	Thr	OG1	24	Trp	O	3.58
35	Asn	ND2	35	Asn	ND2	3.48
69	Ser	OG	81	Lys	O	2.98
73	Thr	OG1	75	Asn	OD1	3.73
75	Asn	OD1	73	Thr	OG1	3.42
81	Lys	O	69	Ser	OG	3.44

(b) Dimer interface through the water molecule

Water Number	Residue1' (A-subunit)			d (Å)	Residue2' (B-subunit)			d (Å)
50	26	Arg	O	3.22	26	Arg	O	2.97
48	35	Asn	OD1	3.34	34	Asp	OD2	2.47
62	63	Asn	OD1	2.96	63	Asn	ND2	3.26
24	73	Thr	OG1	2.70	75	Asn	OD1	3.08
29	75	Asn	OD1	3.32	73	Thr	OG1	2.74
30	75	Asn	OD1	3.80	73	Thr	N	2.86
66	73	Thr	N	3.12	75	Asn	OD1	3.42

Table 4-7 List of hydrogen-bonds at dimer-interface (N-terminus β-sheet)

Residue1 (A-subunit)			Residue2 (B-subunit)			d(Å)
14	Lys	O	20	Tyr	N	2.93
16	Ile	N	18	Phe	O	3.15
16	Ile	O	18	Phe	N	2.91
18	Phe	N	16	Ile	O	2.95
18	Phe	O	16	Ile	N	3.41
20	Tyr	N	14	Lys	O	2.88

shown in Fig.4-9. The other contact is by the intermolecular β-sheet, which protrudes from the barrels of the different subunits in the dimer (Fig.4-10).

The former dimer contact is formed at each three-stranded β-sheet face. This is mainly composed of neutral residues, such as Asn-35, 75, and Thr-28, 73. The contact is rather loose and includes seven water molecules forming hydrogen bonds. Hydrogen bonds pairs in the dimer contacts are listed in Table 4-5.

The latter dimer contact is composed of each N-terminus region by the intermolecular β-sheet and dipole interaction of the each amino terminal α-helix. This β-sheet is consists of each β-strand from Lys-12 to Asn-20. All hydrogen bonds of this dimer contact are listed in Table 4-6.

4-5. Carbohydrate binding site

Initially we tried to make a complex crystal of FVA and carbohydrate by soaking and cocrystallization methods, but no complex crystal was obtained. Then we predicted the carbohydrate binding sites of FVA based on the three-dimensional structure. In many carbohydrate binding proteins, the aromatic residues i.e., Tyr, Phe, His or Trp, or acidic residues i.e., Glu or Asp, bind carbohydrates[45,46]. Distribution of these residues in the subunit suggests that, both sides of the β-barrel have the potential of carbohydrate binding. Tyr-32 and Tyr-83 at one side of the β-barrel are closely positioned and projected into solvent. At the other side of the β-barrel aromatic residues Tyr-20 and acidic residues, Asp-100, Glu-104, and Glu-105, lie close by and are projected into

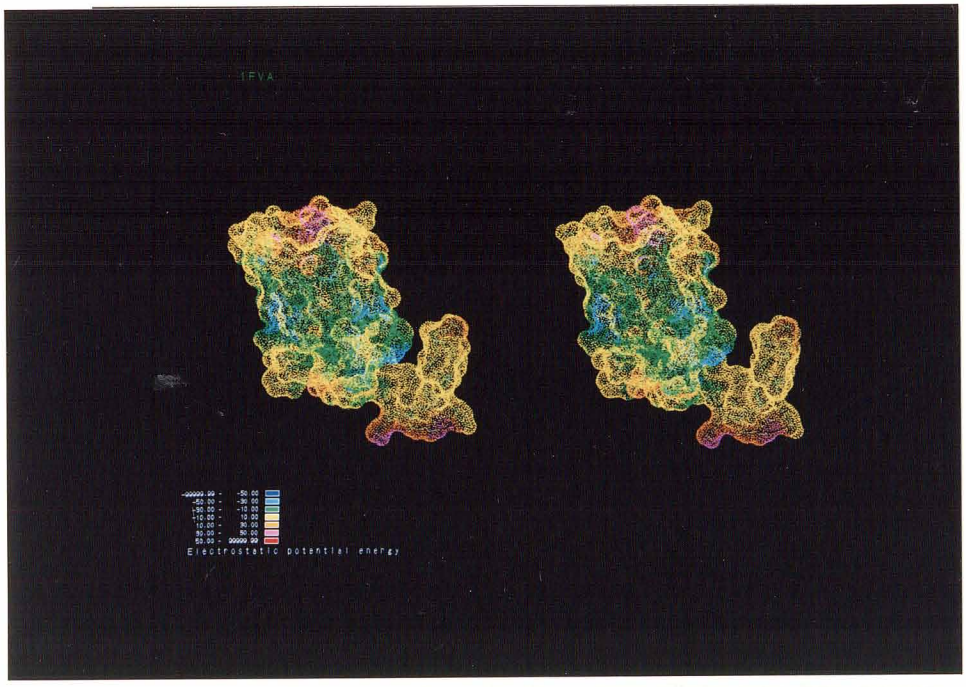


Figure 4-11 A dot surface of the electrostatic potential. The distribution of the electrostatic potential is shown with color. The positive potential is red and the negative is blue. Both sides of β-barrel have positve electrostatic potentials.

solvent. The electrostatic potential[47] of this region is shown in Fig.4-11.

4-6. Heavy atom binding sites

The main binding site of UO_2-ions is located near Oδ1 atom of Asp-17, and two minor binding sites of the UO_2-ions are located near Oε1 atom of Gln-92. The main site is located on the non-crystallographic two-fold axis, and two minor sites are related by the non-crystallographic two-fold axis.

Ag binding sites are located near Nδ2 atom of Asn-35 and O atom of Arg-26. The Ag site is located on the non-crystallographic two-fold axis.

Hg binding sites are located near NH1 atom of Arg-50 and Oε2 atom of Glu-105.

4-7. Water molecules

Ultimately, 129 water molecules are identified in the asymmetric unit. Most water molecules are contacted at the molecular surface by hydrogen bonds, and some water molecules are placed at the dimer interface which was described before. The water molecules were searched for on the (F_O-F_C) and calculated phase maps with computer graphics at each refinement cycle. Water molecules whose temperature factors were diverged over 50.0 $Å^2$ were rejected. The positional parameters and the temperature factors of water molecules were refined with the protein structure refinement.

Chapter-5
Comparison of FVA with immunoglobulin

5-1. Comparison of related proteins

One type of lectin, concanavalin A[26], has been revealed to be a β-structure rich protein. A main part of the structure of concanavalin A is composed of three β-sheets, which contain six, seven and five strands, respectively. The Cα-model is shown in Fig.5-1. The topology of the FVA structure is similar to that of immunoglobulins. The structure of FVA is quite different from those of concanavalin A and wheat germ agglutinin[28]. Each domain of immunoglobulin has a β-barrel structure; seven strands in the constant domain and nine strands in the variable domain. The Cα-model of Ig-G NEW is shown in Fig.5-2. Generally such a β-barrel structure is called immunoglobulin fold. The FVA barrel structure have the same topological connectivity as the corresponding strands in immunoglobulin constant domains. For a β-barrel with *n* strands there are $2^{n-1} \times (n-1)!$ possible topologies. If $n = 7$ this gives $2^6 \times 6! = 46080$ possibilities. A probability of structure matching the immunoglobulin constant domain is $1/46080 = 2.17 \times 10^{-5}$. The probability suggests the topological similarity between FVA and immunoglobulin is unusual if FVA is independent of immunoglobulin. In FVA molecule, the fourth β-strand forms hydrogen bonds with the third strands. In the immunoglobulin constant domain the fourth β-strand hydrogen-bonds with

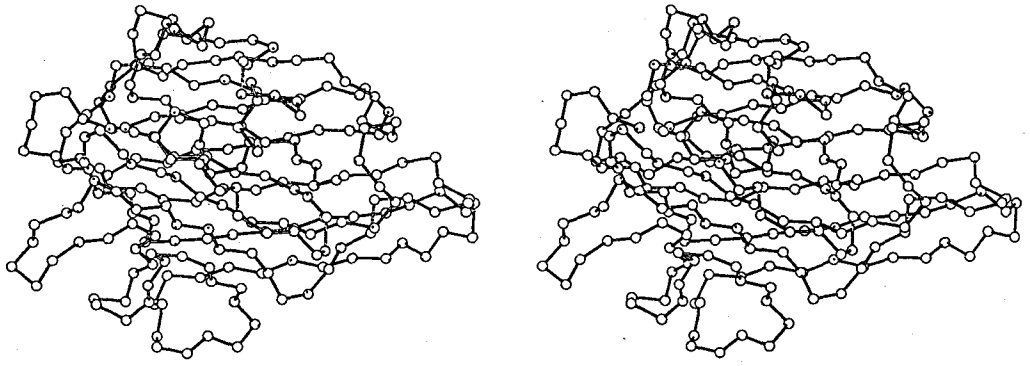


Figure 5-1 Stereo view of α-carbon backbone of concanavalin A. One molecule consists of two polypeptides chains; one chain is divided into two domains. One is a variable domain of nine-stranded β-barrel and the other is a constant domain of seven-stranded β-barrel.

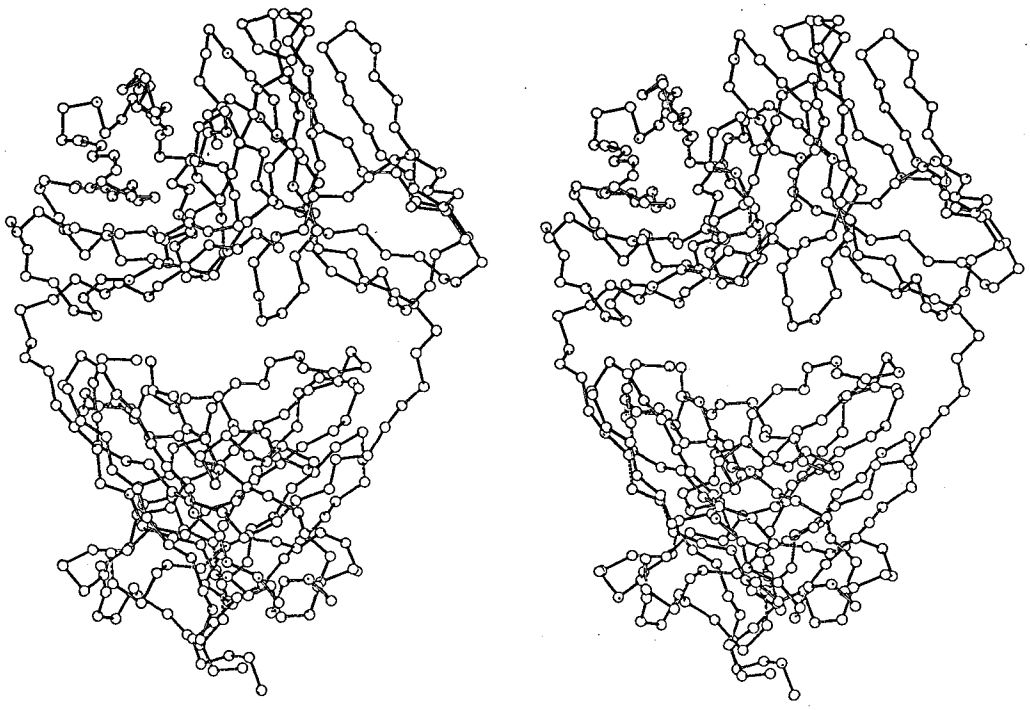


Figure 5-2 Stereo view of α-carbon backbone of 3FAB, Fab fragment of immunogulobulin NEW. One molecule consists of two polypeptides chains; one chain is divided into two domains. One is a variable domain of nine-stranded β-barrel and the other is a constant domain of seven-stranded β-barrel.

the fifth strands, but in the immunoglobulin variable domain, the fourth β-strand hydrogen-bonds with the third strands. Then FVA has a topological folding that can be recognized as immunoglobulin constant domain fold, but details of the folding are considerably divergent. Precedents for such a barrel structure of FVA have been found in domain-2 of papD[48] and domain-2 of CD4[49,50], and these molecules are classified as immunoglobulin superfamilies. The Cα-model of CD4 is shown in Fig.5-3. Topology of FVA and related proteins are shown in Fig. 5-4.

In order to determine whether the topological similarity between FVA and immunoglobulin domains reflects structural similarity, the structure of FVA is compared with those of immunoglobulin domains by means of the Rossman and Argos method[51]. All positional parameters of immunoglobulin domains were obtained from the Protein Data Bank at Brookhaven National Laboratory[52]. Table 5-1 lists all names of related proteins. Table 5-2 shows the similarity of tertiary structures between FVA and the related proteins. The result of comparison shows that an overall structure of FVA closely more resembles immunoglobulin structures than lectins. The percentage of the equivalent residues between FVA and related proteins, which shows the structural similarities of a whole molecule, are from 37.4 % to 75.0 %. The highest similarity observed at the constant domains is 75.0 % for Ig-G F19.9[58]. The highest similarity observed at the variable domains is 68.2 % for Ig-A McPC603[57]. However, the similarity of the domain-2 of CD4 has a very high value of 91.0 %. This result suggests that the topological structure of FVA is the same as domain-2 of CD4.

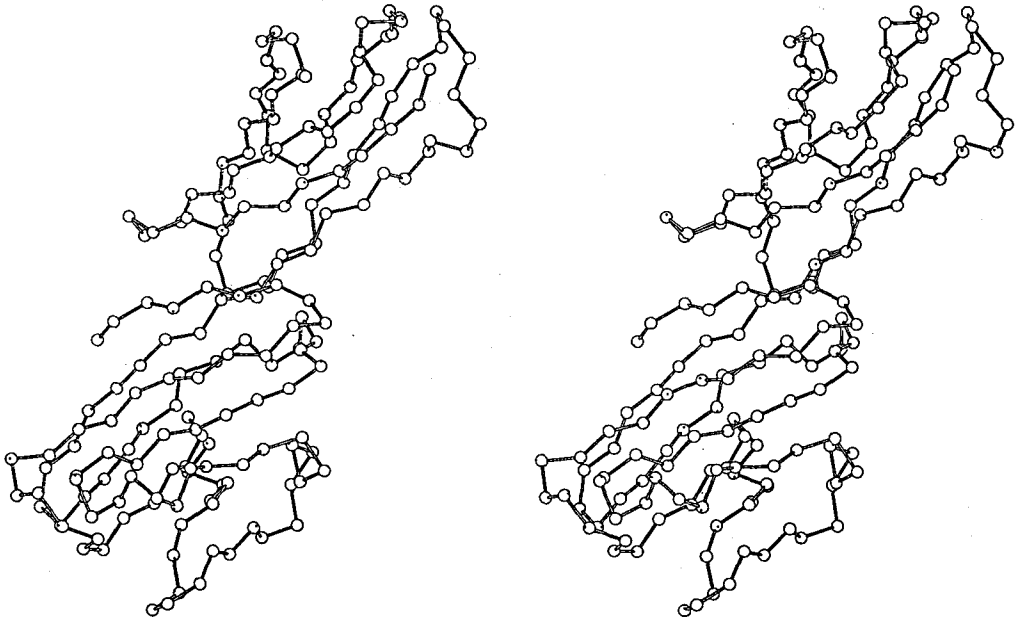


Figure 5-3 Stereo view of α-carbon backbone of CD4. One molecule is divided into two domains. One domain (upper side domain) is a nine-stranded β-barrel and the other domain (lower side domain) is a seven-stranded β-barrel.

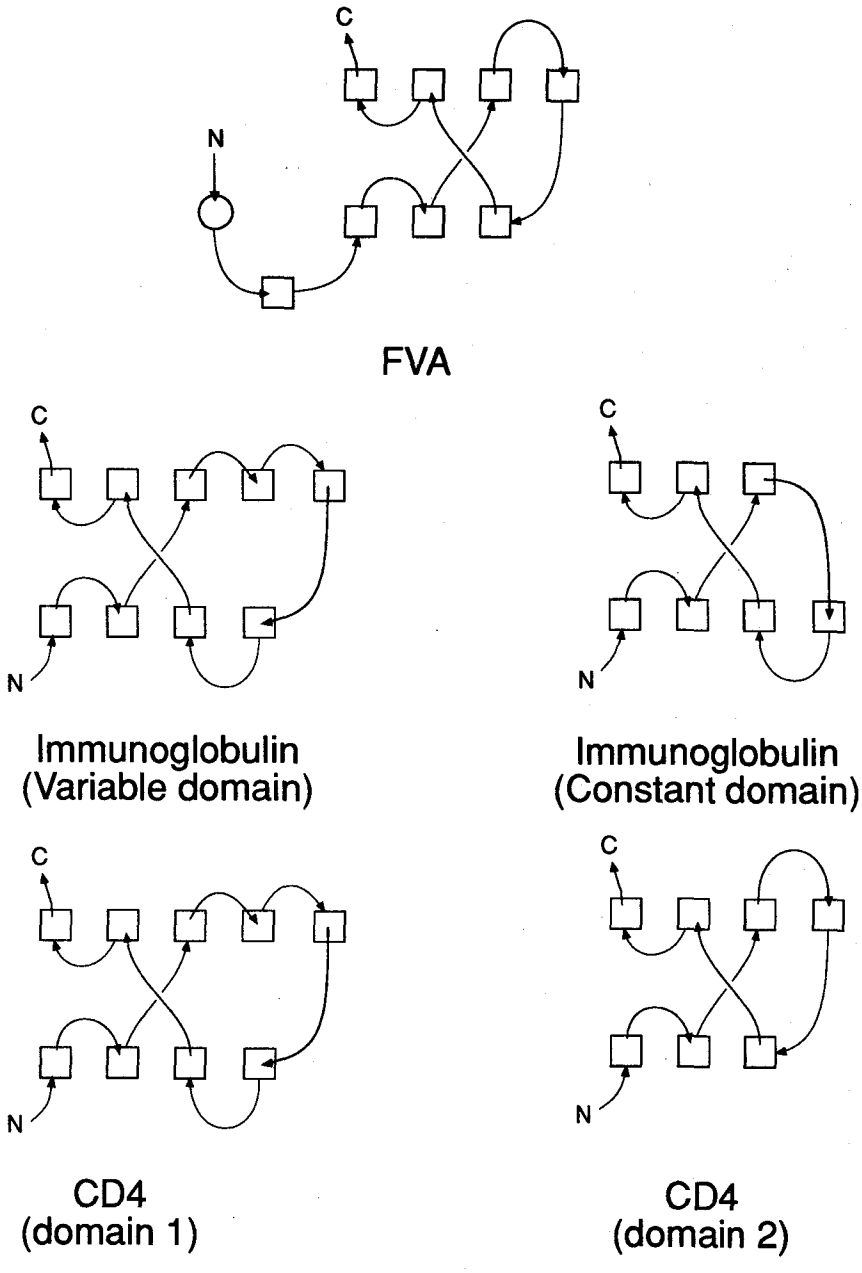


Figure 5-4 Comparison of the topology between FVA and immunoglobulin superfamily.
In this figure, a square means a β-strand, and a circle means a α-helix.

Table 5-1 List of IDENT code of the Protein Data Bank

PDB-ID	Protein name	Type*	Reference
3CNA	Concanavalin A	Lectin	(22)
3WGA	Wheat Germ Agglutinin	Lectin	(24)
3FAB	Ig-G Fab New	Ig	(53)
2FB4	Ig-G1 Fab(lambda) Kol	Ig	(54)
1MCG	Ig Bence-Jones dimer	Ig	(55)
2FBJ	Ig-A Fab(kappa) J539	Ig	(56)
2MCP	Ig-A Fab(kappa) McPC603	Ig	(57)
1F19	Ig-G Fab F19.9	Ig	(58)
1FC1	Ig-G Fc	Ig	(59)
1CD4	CD4 HIV binding fragment	Ig-superfamily	(49)
2CD4	CD4 N-terminal fragment	Ig-superfamily	(50)

* Ig : Immunoglobulin, Ig-superfamily : Immunoglobulin superfamily

Table 5-2 Structural similarity between FVA and related proteins

PDB-ID	Domain* type	Number of equivalent res.	Total residue	Similarity %	Whole similarity
Lectin					
3CNA		27	171	15.8	no
3WGA		35	233	15.0	no
Immunoglobulin					
1MCG	VL	48	115	41.7	no
	CL	40	117	37.4	yes
2FBJ	VL	66	120	55.0	yes
	CL	66	104	63.5	yes
	VH	71	120	59.2	yes
	CH	68	101	67.3	yes
2MCP	VL	75	110	68.2	yes
	CL	75	111	67.6	yes
	VH	48	120	40.0	yes
	CH	40	103	38.8	no
3FAB	VL	70	110	63.6	yes
	CL	76	105	72.4	yes
	VH	66	120	55.0	yes
	CH	58	101	57.4	yes

* VL : Variable domain of Light chain, CL : Constant domain of Light chain,
VH : Variable domain of Heavy chain, CH : Constant domain of Heavy chain.

Table 5-2 Structural similarity between FVA and related proteins (continued)

PDB-ID	Domain* type	Number of equivalent res.	Total residue	Similarity %	Whole similarity
2FB4	VL	78	120	65.0	yes
	CL	63	105	60.0	yes
	VH	42	120	35.0	no
	CH	37	120	31.1	no
1F19	VL	77	115	67.0	yes
	CL	78	104	75.0	yes
	VH	48	120	40.0	yes
	CH	74	101	73.3	yes
1FC1	CH	71	113	62.8	yes
	CH	69	94	73.4	yes
Immunoglobulin superfamily					
2CD4	D1	50	105	47.6	yes
	D2	76	83	91.6	yes
2CD4	D1	53	105	50.5	yes
	D2	79	86	91.9	yes

* D1 : Domain-1, D2 : Domain-2

To determine the deviations among FVA and structurally related proteins, the FVA molecule was superposed on the related proteins by the least-squares methods as the summation of deviations between the corresponding Cα atoms to minimize the distance. Table 5-3 summarizes the root-mean-square deviations between FVA and related proteins. The smallest root-mean-square deviation at the constant domains is 3.67 Å in Ig-G NEW. The smallest root-mean-square deviation at the variable domains is 4.04 Å also in Ig-G NEW[53]. The root-mean-square deviation from the CD4 domain-2 is 3.7 Å. In terms of the root-mean-square deviation, there is no major difference between immunoglobulins and CD4. All the deviation ranges from 3.67 Å to 5.38 Å. Fig. 5-5 shows the superimposed FVA and Ig-G NEW constant domain models, and Fig. 5-6 shows the superimposed FVA and CD4 models. In general, the constant domains show a little smaller deviation than the variable domains. Thus FVA is more similar to immunoglobulin constant domains than variable domains, and is most similar to CD4 molecules.

Such a structural difference from immunoglobulins may be attributed to the following three respects.

(1) FVA lacks the disulfide bond between the B and the F strand, which is usually found in immunoglobulin domains.

(2) The β-barrel structure of FVA can be seen as two layers of β-sheet, with the β-strands of β1, β2 and β5 in one layer and the strands of β3, β4, β6 and β7 in the other; the β4 strand is hydrogen-bonded to the β3 strand so as to join the β-sheet consisting of the strands of β3, β6 and β7, rather than being hydrogen-bonded to the β5 strand. In this respect, the

Table 5-3 r.m.s. deviation between FVA and related proteins

PDB-ID	Domain type	Homologus residue	r.m.s. deviation	Mean distance
Immunoglobulin				
1MCG	CL	40	3.76	3.51
2FBJ	VL	66	4.48	3.82
	CL	66	3.73	3.41
	VH	71	4.11	3.77
	CH	68	4.24	3.97
2MCP	VL	75	5.38	4.90
	CL	75	4.25	3.69
	VH	48	4.78	4.42
3FAB	VL	70	4.04	3.63
	CL	76	3.67	3.38
	VH	66	4.20	3.76
	CH	58	3.77	3.25
2FB4	VL	78	4.51	4.00
	CL	63	4.08	3.63
1F19	VL	77	4.67	4.22
	CL	78	4.14	3.69
	VH	48	3.91	3.57
	CH	74	3.91	3.30
1FC1	CH	71	3.94	3.37
	CH	69	3.86	3.56
Immunoglobulin superfamily				
1CD4	D1	50	4.27	3.90
	D2	76	3.72	3.31
2CD4	D1	53	4.16	3.79
	D2	79	3.75	3.36

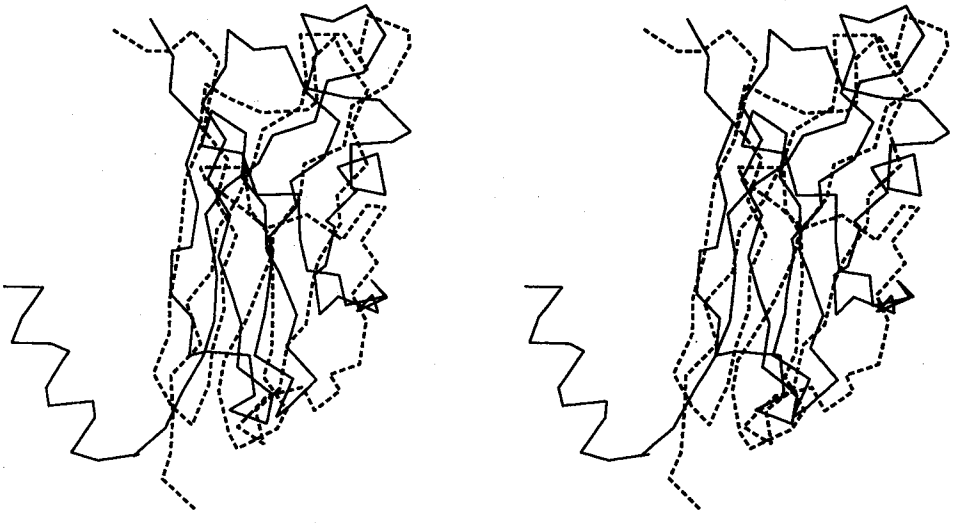


Figure 5-5 Comparison of the Cα backbone between FVA and Immunoglobulin constant domain. FVA is represented with thick line. Immunogloblin constant domain is the light chain constatnt domain of 3FAB, Ig-G NEW, and represented with broken line.

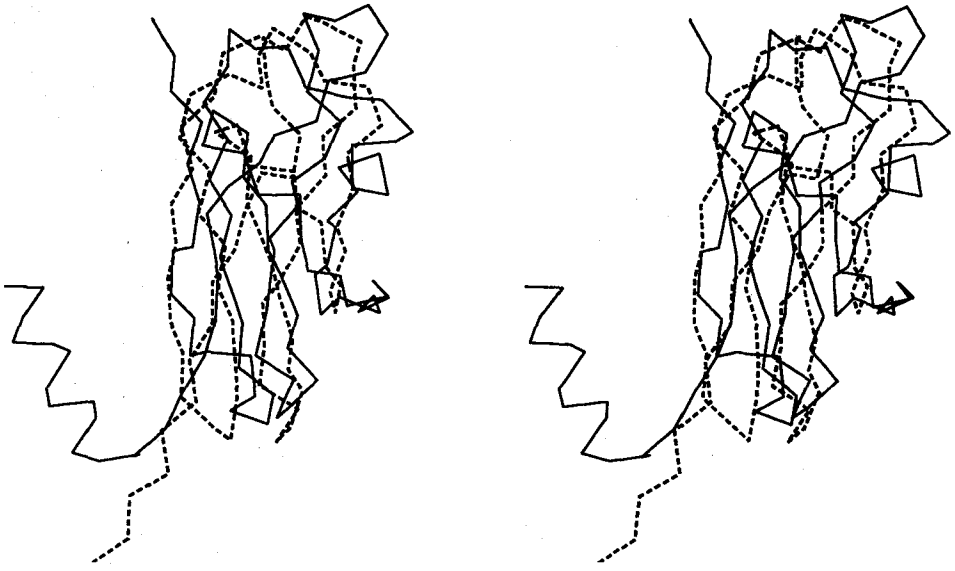


Figure 5-6 Comparison of the Cα backbone between FVA and CD4. FVA is represented with thick line and CD4 is represented with broken line.

barrel structure of FVA resembles those of immunoglobulin variable domains, CD4 domain-2 and papD domain-2. Then the names of β-strands from β1 to β7 are configured A,B,C,C', E, F, and G, respectively.

(3) The size of the barrel of FVA is small compared with that of the immunoglobulin constant domains, and the comparison of sequences whose alignment was based on three-dimensional structure show almost no similarity. In FVA, the total number of β-barrel composing residues is 93, and the average length of one β-strand is 7.4, compared with 105 and 9.1 in McPC603[57], respectively.

5-2. Comparison of dimer associations with immunoglobulins

The two asymmetric molecules are related by two different local two-fold axes in the crystal. One type of dimer association is comparable to an immunoglobulin H-L association. The mode of interactions of dimer interface is different between FVA and immunoglobulins. In immunoglobulin the interface is mainly composed of hydrophobic residues, in contrast in FVA the interface is mainly composed of neutral residues.

The geometrical differences from immunoglobulins are the angles formed by the direction of β-sheets in a domain and inter-domain association, as defined in Fig.5-7[60,61]. Those angles are listed in Table 5-4 which compares FVA molecules with both the constant and the variable domains of the four typical immunoglobulins. The γ1 angle of 45° in the β-barrel of FVA is significantly large compared with immunoglobulins, 25° in constant domains

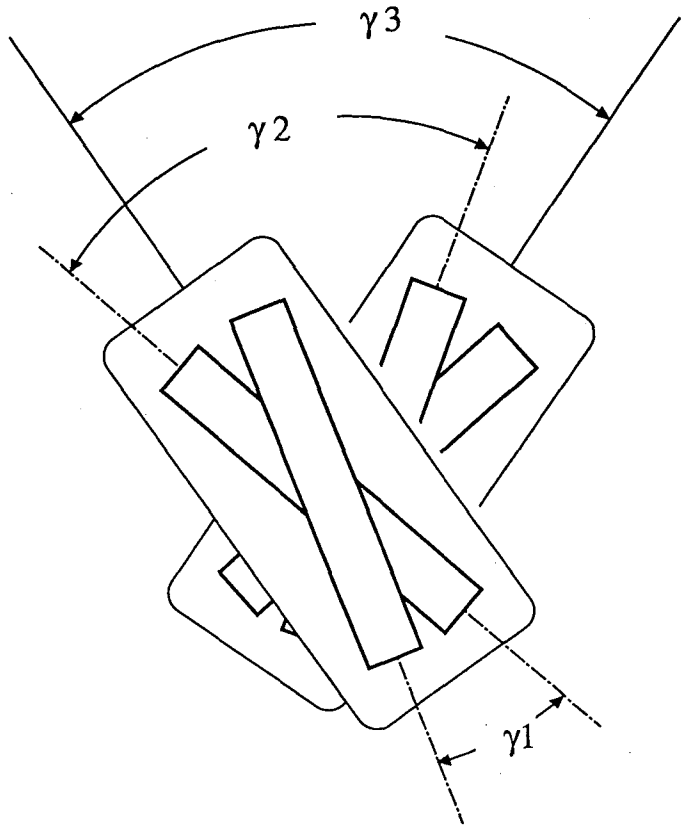


Figure 5-7 Schematic drawing of the β-barrel domain association. The large rectangle represents a β-barrel and the small rectangle represents a β-strand. The angles are defined as follows; γ1 is the crossover angle between the β-sheets facing inside of a domain, γ2 is the crossover angle between the β-strands inside of the domain-domain interface and γ3 is the crossover angle between the axes of domains.

Table 5-4 Dimer association angles in FVA and Immunoglobulin

	γ1		γ2		γ3	
PDB-ID	c*	v**	c*	v**	c*	v**
FVA	45		27		108	
3FAB	24	35	64	53	92	88
2FBJ	26	35	55	43	99	78
2MCP	25	34	57	44	98	78
2FB4	23	37	65	43	92	80

γ1, γ2 and γ3 are defined at Fig. 5-7, and measured using the computer graphic system. The units are degrees.

*C : Constant doamin, **V : Variable domain.

and 35° in variable domains. The γ2 angle of 27° in FVA, which is the angle formed between facing β-sheets of the domain interface, is smaller than that of immunoglobulins. The angle γ3 of 108° in FVA, which is defined as a domain axis crossover angle, is larger than immunoglobulins. These angles are approximately conserved among immunoglobulins, however FVA shows slight differences from them.

Such a dimer association difference from immunoglobulins may be attributed to the following two aspects.

(1) The average length of one β-strand is short and the size of the barrel of FVA is small, compared with that of immunoglobulins.

(2) These results are consequences of the difference of the composing amino acid residues and intervening bound water molecules that causes the difference of dimer interface between FVA and immunoglobulin.

Chapter-6
Summary and Conclusions

6-1. New Protein Crystallographic Techniques

Now we need to make structure determination faster and easier. We have tried to routinize the process of determining the structure of protein and make it possible to determine a protein structure automatically.

Apart from the crystallization and preparation of heavy-atom derivative crystals, the difficulties of structure determination are mainly derived from the necessity of many times. Especially, the most serious rate-limiting steps are the steps of data collection and structure refinement. Improvement of the data collection equipment and the refinement programs are the most straightforward way to resolve these two difficulties. Still more the difficulties of the model building was reduced by the improvement of the model building program.

We have developed a new two-dimensional diffractometer system (IP-system) based on the new X-ray recording medium (Fuji film, Imaging Plate), and introduced the Siemens Area Detector system. Both types of the two-dimensional detectors, IP-system and Area Detector system, cut down the data collection time of the diffraction data and reduce number of crystals to be used. Those diffractometers make it possible to collect all diffraction data by one crystal in a few days.

We introduced two new programs, FRODO for molecular model building and XPLOR for model refinement. The new model building system, FRODO,

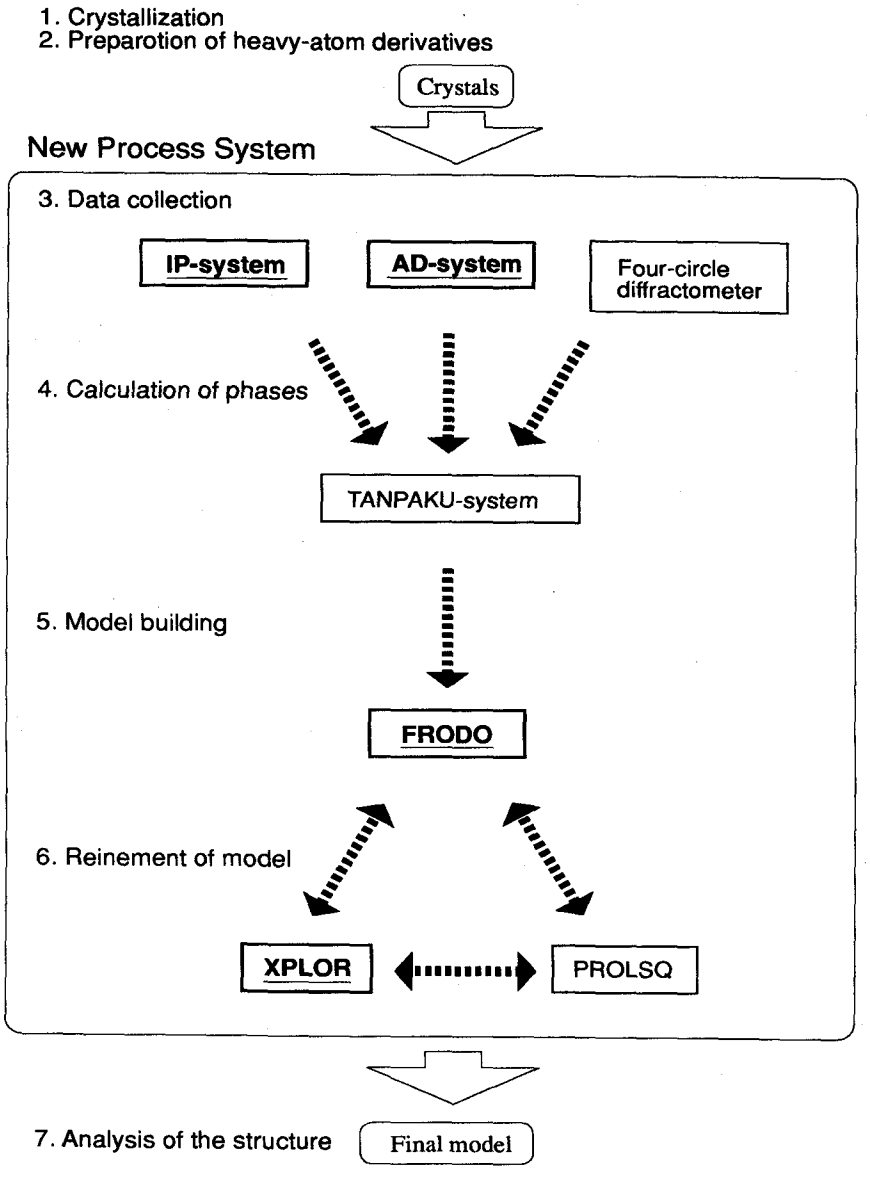


Figure 6-1 A schematic representation of the newly developed Protein Crystallographic system. New equipments and methods are shown by the under-line.

has a good interface to build a molecular model and highly regularizes the distorted structure. FRODO could reduce the difficulties of the model building.

The new molecular dynamics refinement program, XPLOR, easily overcame the local energy minimum of the distorted model without manual correction. Thus XPLOR could semiautomatically refine the model structure. XPLOR could greatly speed up the refinement of the model structure.

New equipment and methods reduce the difficulties of each structure determination steps respectively and totally speed up the structure analysis. We combined new equipment and methods with the former processing programs. The process of determining a protein structure is almost routinized and closely connects the respective steps of structure determination. The new process of determining a protein structure is summarized in Fig. 6-1. The new process could reduce the time required for the protein structure analysis by half and make it considerably easier to determine a protein structure. However the new process have not completely made possible the automatic determination of protein structure.

6-2. Structure analysis of FVA

The result of the structure analysis suggests that the FVA molecule is mainly composed of a seven stranded β-barrel structure and is categorized as a new type of lectin-like protein by the three-dimensional structure. The FVA molecule has some sequence similarities with immunoglobulins in the sequence analysis. Highest sequence similarity has been shown to be 24

%[30]. In particular, the sequence from residue 10 to 30 has shown high similarity. Nevertheless the FVA molecule contains no cysteine residues, which are highly conserved in immunoglobulins. The early results of the amino acid sequence comparison between FVA and Immunoglobulin based on only the amino acids sequence showed some similarities, however, with the progress of the structure analysis the sequence similarity was significantly decreased when the three-dimensional structure was taken into account.

The possible reason of this result is that some sequence similarity between FVA and immunoglobulin could not be understood under some considerations. Since the evolutional speed of immunoglobulin sequence is very high[62], no sequence similarity would be expected because of the long period of time since the juncture between mushrooms and animals. However, the possibility of the sequence similarity between FVA and immunoglobulin could not be negated, regardless of considering the conserved topology of the β-barrels and the sequence homology through the whole chains. The N-terminal 20 residues of FVA may have been excluded from the β-barrel during the long evolutional period.

During evolution, the amino acid sequence of a protein is much more variable and changes more rapidly than its tertiary structure[63]. In long evolutional period, the amino acid sequences of proteins derived from a common precursor may alter to the point that they are no longer demonstrably homologous. The ability to make meaningful comparisons between such distantly related proteins must come from structural similarity. On the other hand, structural homology in the absence of sequence homology might be

attributed to convergent rather than divergent evolution. A common fold might be dictated by folding requirement. The FVA molecule has a structure similarity with immunoglobulin constant domains and CD4 molecules. The topological structure of FVA is the same as domain-2 of CD4.

One possible interpretation of immunoglobulin like β-barrel in FVA is a result of conservation in spite of a long period of evolutional time. This means that the structure similarity is caused by a divergent evolution as seen among immunoglobulins and their superfamilies. Another possible interpretation of immunoglobulin like β-barrel in FVA is merely a result of a structural stability of such topological β-barrel. In this respect the structural similarity is caused by a convergent evolution as seen in Cu,Zn superoxide dismutase[64].

However, functionally lectins have some immunomodulating properties, which would suggest that they may interact with some receptors which are involved in the immunological reaction processes and may be a member of immunoglobulin superfamily. The dimer association of FVA is comparable to an immunoglobulin H-L association. Then we would conclude that FVA is a member of immunoglobulin superfamily and has an evolutional relationship with them.

References

1. Tsuda, M. (1979) *J. Biochem.* ,**86**,1463-1468
2. Yatohgo,T., Nakata,M., Tsumuraya,Y.,Hashimoto,Y., and Yamamoto,S. (1988) *Agric. Biol. Chem.*,**52**,1485-1493
3. Kino, K., Yamashita, A., Yamaoka,K., Watanabe, J., Tanaka, S., Ko,K., Shimizu, K., and Tsunoo, H. (1989) *J. Biol. Chem.* ,**264**,472-478
4. Tanaka, S., Ko, K., Kino, K., Tsuchiya, K., Yamashita, A., Murasugi, A., Sakuma,S., and Tsunoo, H. (1989) *J. Biol. Chem.* ,**264**,16372-16377
5. Blundell, T.L., and Johnson, L.N. (1976) in *Protein Crystallography*, Academic Press, London
6. Stout, G.H., and Jensen, L.H. (1989) in *X-ray Structure Determination*, John Wiley & Son, New York
7. Miyahara, J., Takahashi, K., Amemiya, Y., Kamiya, N. and Satow, Y., (1986) *Nucl. Instar. Methods* , **A246**, 572-578
8. Rossmann, M.G. (1979) *J. Appl. Cryst.*, **12**, 225
9. Rossmann, M.G., Lslie, A.G.W., Abdel-Meguid, S.S., and Tsukihara, T. (1979) *J. Appl. Cryst.* , **12**, 570
10. Fox, G.C. and Holmes, K.C. (1966) *Acta. Cryst.*, **20**, 886-891
11. Blum, M., Metcalf, P. Harrison, S.C., and Wiley, D.C. (1987) *J. Appl. Cryst.*, **20**,235-242
12. Howard, A.J., Gilliland, G.L., Finzel, B.C., Poulos, T.L., Ohlendorf, D.H., and Salemme, F.R. (1987) *J. Appl. Cryst.*, **20**,383-387

13. Jones, T.A. (1982) in *Computational Crystallography*, Clarendon Press, Oxford. pp. 303

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

15. Pflugrath, J.W., Saper, M.A., and Quiocho, F.A. (1984) in *Methods and Applications in Crystallographic Computing*, Clarendon Press, Oxford. pp. 407

16. Brunger, A.T., Kuriyan, J., and Karplus, M. (1987) *Science* , **235**, 458-460

17. Brunger, A.T. (1988) *J. Mol. Biol.* , **203**, 803-816

18. Brunger, A.T., Karplus, M., and Petsko, G.A. (1989) *Acta. Cryst.* , **A45**, 42-50

19. Jack, A., and Levitt, M. (1983) *Acta. Cryst.*, **34**, 931

20. Karplus, M., and McCammon, J.A. (1981) *CRC Crit. Rev. Biochem.*, **9**,293-349

21. Kirpatrick, S., Gelatt, C.D., and Vecchi, M.P. (1983) *Science*, **220**, 671-680

22. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. (1983) *J. Comput. Chem.*, **4**,187-217

23. Brunger, A.T., Krukowski, A., and Erickson, J. (1990) *Acta. Cryst*, **A46**, 585-593

24. Natori, K. (1991) in *A Master thesis*, Faculty of Science, Osaka University, Japan

25. Nathman, S., and Halina, L. (1989) *Science* , **246**, 227-246

26. Karl, D.H., and Cliton, F.A. (1972) *Biochem* ,**11**,4910-4919

27. Christine, S.W. (1977) *J. Mol. Biol.* ,**111**,439-457

28. Howard, E.,Elizabeth, H.P., Suguna, K., Eswara, S, and Suddath, F.L. (1986) *J. Biol. Chem.* ,**261**,16518-16527

29. Goldstein, I.J., Hughes, R.C., Monsigny, M., Osawa, T., and Sharon, N. (1980) *Nature*, **285**,66

30. Yamada, M. (1990) in *A Master thesis*, Faculty of Science, Saitama University, Japan

31. Alan, F. W. (1987) *Immunology Today* , **8**,298-303

32. Hirano, S., Matsuura, Y., Kusunoki, M., Kitagawa, Y., and Katsube, Y. (1987) *J. Biochem.* ,**102**,445-446

33. Sakabe, N. (1983) *J. Appl. Cryst.*,**16**,542-547

34. Sakabe, N., Nakagawa, A., Sakabe, K., Watanabe, N., Kondo, H., and Shimomura, M. (1989) *Rev. Sci, Instrum.* , **60(7)**, 2440-2441

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

36. Patterson, A.L. (1934) *Phys. Rev.* , **46**, 372

37. Perutz, M.F. (1956) *Acta Cryst.* , **9**, 867

38. Blow, D.M. (1958) *Proc. Roy. Soc.* , **A247**, 302

39. Blow, D.M., and Crick, F.H.C. (1959) *Acta Cryst.* , **12**, 794

40. Bricogne, G. (1976) *Acta. Cryst.* , **A32**, 832-847

41. Kusunoki, M. (1980) in *A doctoral thesis*, Faculty of Science, Osaka University, Japan

42. Jones,T.A. (1985) *Methods Enzymol.* ,**115**, 157-171

43. Hendrickson, W.A. (1985) *Methods Enzymol.* ,**115**, 252-270

44. Ramakrishman, C., and Ramachandran, G.N. (1965) *Biophys. J.* , **5**,

909-933

45. Reeke Jr., G.N., and Becker, J.W. (1988) *Current Topics in Microbiology and Immunology* , **139**, 35-53

46. Johnson, L.N., Cheethman, J., McLaughlin, P.J., Acharya, K.R., Barford, D., and Phillips, D.C. (1988) *Current Topics in Microbiology and Immunology* , **139**, 81-129

47. Conolly, M. (1983) *J. Appl. Cryst.* , **16**, 548-558

48. Holmgren, A. and Branden, C.I. (1989) *Nature* , **342**, 248-251

49. Wang, J., Yan, Y., Garrett, T.P.J., Liu, J., Rodgers, D.W., Garlick, R.L., Tarr, G.E., Husain, Y., Reinherz, E.L., and Harrison, S.C. (1990) *Nature*, **348**, 411-418

50. Ryu, S., Kwong, P.D., Truneh, A., Porter, T.G., Arthos, J., Rosenberg,M., Dai, X., Xuong, N., Axel, R., Sweet, W.R., and Hendrickson, W.A. (1990) *Nature*, **348**, 419-425

51. Rossman, M.G., and Argos, P. (1976) *J. Mol. Biol.*, **105**, 75-95

52. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F.Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol.*, **112**, 535

53. Saul, F.A., Amzel, L. M., and Poljak,R.J. (1978) *J. Biol. Chem.* ,**253**,585

54. Marquart, M., Disenhofer, J., Huber, R., and Palm, W. (1980) *J. Mol. Biol.* **141**,369

55. Ely, K.R., Firca,J.R., Williams, K.J., Abola,E.E., Fenton, J.M., Schiffer,M., Panagiotopoulos, N.C., and Edmundson, A.B. (1978) *Biochem.*, **17**, 158

56. Suh, S.W., Bhat, T.N., Navia, M.A., Cohen, G.H., Rao, D.N., Rudikoff,S.,

and Davies, D.R. (1986) *Proteins. Struct., Funct.* , **1**, 74

57. Padlan, E.A., Cohen, G.H., and Davies, D.R. (1985) *Ann. Immunol.*, **C136**, 271

58. Lascombe, M.B., Alzari, P.M., Boulot, G., Saludjian, P., Tougard, P. Berek, C., Haba, S., Rosen, E.M., Nisonoff, A., and Poljak, R.J. (1989) *Proc. Natl. Acad. Sci. U.S.A.* , **86**, 607

59. Deisenhofer, J., Colman, P.M., Epp, O., and Huber, R. (1976) *Z. Phys.* , **357**, 1421

60. Cyrus, C., and Joel, J. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 4146-4150

61. Jiri, N., Robert, B., John, N., David, M., Edogar, H., and Martin,K. (1983) *J. Biol. Chem.* , **258**,14433-14437

62. Dayhoff, M.O. (1976) in *ATLAS of PROTEIN SEQUENCE and STRUCTURE*, **5**, 1-8

63. Grutter, M.G., Weaver, L.H., and Matthews, B.W. (1983) *Nature.*, **303**, 828-831

64. Richardson, J.S., Richardson, D.C., Thomas, K.A., Silverton, E.W., and Davies, D.R. (1976) *J. Mol. Biol.*, **102**, 221-235

List of Publications

(1) Development of a High Speed Data Collection System with an Imaging Plate in Consideration of Large Unit Cell Crystals
M.Sato, M.Yamamoto, K.Imada, Y.Katsube, N.Tanaka , and T.Higashi
(1991) *J. Appl. Cryst.*, **23**, (in press)

(2) High Resolution Data Collection of the Lectin from *Flammulina veltipes* using Weissenberg Camera System
M.Yamamoto, Y.Kitagawa, H.Ago, M.Kusunoki, Y.Matsuura, and Y.Katsube
(1990) *Photon Factory Activity Report*, **8**, 86

(3) X-ray Crystal Structure Analysis of *Flammulina veltipes* agglutinin
M.Yamamoto
(1991) *Tanpakushitu Kogaku Kiso Kenkyu Senta Dayori*, **12**, 33-39
Resarch Center for Protein Engineering, Institute for Protein Research, Osaka University, Japan

(4) The three-dimensional structure of *Flammulina Veltipes* agglutinin contains an immunoglobulin fold
M.Yamamoto, Y.Kitagawa, Y.Matsuura, Y.Katsube, and Y.Hashimoto
(in preparation)