



RETURNING MATERIALS:
Place in book drop to
remove this checkout from
your record. FINES will
be charged if book is
returned after the date
stamped below.

--	--	--

SEP 24 1955 10 15 AM

PART I: FORMATION, CRYSTALLIZATION, AND PRELIMINARY
CRYSTALLOGRAPHIC DATA ON THE TERNARY COMPLEX OF
 α -CHYMOTRYPSIN, β -TRYPSIN AND THE BOWMAN-BIRK
INHIBITOR.

PART II: X-RAY CRYSTALLOGRAPHIC STUDIES OF RIBONUCLEASE T₁.

By

James Richard Gaier

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

1983

ABSTRACT

PART I: FORMATION, CRYSTALLIZATION, AND PRELIMINARY CRYSTALLOGRAPHIC DATA ON THE TERNARY COMPLEX OF α -CHYMOTRYPSIN, β -TRYPSIN AND THE BOWMAN-BIRK COMPLEX.

PART II: X-RAY CRYSTALLOGRAPHIC STUDIES OF RIBONUCLEASE T₁.

By

James Richard Gaier

The ternary complex of the Bowman-Birk inhibitor with α -chymotrypsin and β -trypsin has been crystallized, and preliminary crystallographic information describing the crystals has been obtained. The crystals were grown from solutions containing 0.10 M sodium acetate and 10% PEG 4000 at pH 4.6. Three forms of the ternary complex have been characterized, all of which are triclinic (space group P1), and all have a single complex in the unit cell. The first form occurs at PEG concentrations below 12% and has cell constants which vary with pH and salt concentrations by ± 0.4 Å and $\pm 0.8^\circ$ from: $a = 51.5$ Å, $b = 56.1$ Å, $c = 55.7$ Å, $\alpha = 99.9^\circ$, $\beta = 105.3^\circ$, and $\gamma = 96.7^\circ$. The second form occurs at PEG concentrations above 15%, and has cell constants which only vary by ± 0.2 Å and $\pm 0.1^\circ$ from: $a = 50.7$ Å, $b = 56.1$ Å, $c = 52.1$ Å, $\alpha = 90.4^\circ$, $\beta = 112.1^\circ$, and $\gamma = 96.7^\circ$ from 15% to 25% PEG. The third form occurs at 25% PEG and has cell constants: $a = 52.0$ Å, $b = 56.3$ Å, $c = 46.7$, $\alpha = 98.8^\circ$, $\beta = 97.3^\circ$, and $\gamma = 97.1^\circ$. Crystals of the third form which are 1.0 mm in length scatter x-rays

to 2.2 Å resolution, whereas comparable crystals of the former two forms scatter to 2.8 Å resolution. A search for heavy atom derivatives was begun, but depletion of sufficiently pure Bowman-Birk inhibitor terminated the project.

Crystals of ribonuclease T₁ were studied by x-ray diffraction in an attempt to determine the tertiary structure of the protein. The crystals grown from 55% saturated ammonium sulfate in 0.125 M phosphate at pH 6.6, proved to be orthorhombic (space group P2₁2₁2₁) with cell constants: $a = 91.73 \pm 0.03$ Å, $b = 37.32 \pm 0.006$ Å, and $c = 77.78 \pm 0.03$ Å. Heavy atom searches were conducted at both pH 7.3 and pH 5.5 with limited success. Four promising derivatives at pH 5.5 were studied in detail in preparation for the multiple isomorphous replacement method of determining phase angles for a protein. Of these, only the mersalyl derivative proved to be usable. This forced the use of the iterative single isomorphous replacement method to determine phases to 3.5 Å resolution. The electron density maps synthesized using these phase angles suggest the position of dimers within the unit cell. Unfortunately, the quality of the map was insufficient to allow the tracing of the α-carbon backbone.

To my family and friends,
whose support made this work possible.

ACKNOWLEDGMENTS

I would first like to express my sincere appreciation to Dr. Alexander Tulinsky for his continued guidance and support throughout the course of this work.

I would also like to thank Dr. Bi-Cheng Wang for his assistance, and the use of his facilities during the phase refinement.

I would like to express my appreciation to Dr. Frederick G. Waltz, Jr., and to Dr. Irving E. Liener who kindly provided the proteins for this study.

Sincere thanks is also given to Dr. Donald Ward, and all of the members of Dr. Tulinsky's Laboratory for their many helpful discussion and general assistance.

Finally, I should like to thank the Upjohn Company, and the National Institutes of Health for their financial support throughout the course of this work.

TABLE OF CONTENTS

	PAGE
LIST OF TABLES.	vi
LIST OF FIGURES.	vii
INTRODUCTION.	1
I. OVERVIEW OF PROTEIN CRYSTALLOGRAPHY.	2
II. OVERVIEW OF EXPERIMENTAL TECHNIQUES.	4
III. OVERVIEW OF THE ELECTRON DENSITY MAP CALCULATION.	17
PART I: FORMATION, CRYSTALLIZATION, AND PRELIMINARY CRYSTALLOGRAPHIC DATA ON THE TERNARY COMPLEX OF α -CHYMOTRYPSIN, β -TRYPSIN, AND THE BOWMAN-BIRK INHIBITOR.	26
I. INTRODUCTION.	26
A. Proteases.	26
B. Protein Protease Inhibitors.	31
C. Bowman-Birk Inhibitor and Its Family.	36
II. EXPERIMENTAL METHODS.	44
A. Forming the Complex and Growing the Crystals.	44
B. Characterization of the Crystals and the Native State.	47
C. Verification of the Ternary Complex.	56
D. Heavy Atom Search.	59
PART II: X-RAY CRYSTALLOGRAPHIC STUDIES OF RIBONUCLEASE T ₁	61
I. INTRODUCTION.	61
II. EXPERIMENTAL.	70
A. Growing Crystals.	70
B. Preliminary Diffraction Studies.	74
C. Preparation of Heavy Atom Derivatives.	76
D. Intensity Data Collection.	84

	PAGE
III. STRUCTURE ANALYSIS.	.88
A. Determination of Heavy Atom Positions.	88
B. Difference Electron Density Methods.	.100
IV. PHASE DETERMINATION.	.112
A. The Iterative Single Isomorphous Replacement Method.	.112
B. The ISIR Phase Refinement of T_1 Using MRS.	.121
V. THE ELECTRON DENSITY.	137
A. The Electron Density Map.	137
B. Fitting the Model.	.141
VI. CONCLUDING REMARKS.	144
LIST OF REFERENCES.	147
APPENDIX.	152

LIST OF TABLES

TABLE		PAGE
1	Cell Constants of BBI Ternary Complex Crystals Under Various Conditions.	49
2	The Amino Acid Composition of T_1	.63
3	Number of Molecules per Asymmetric Unit with Corresponding Percent Mother Liquor.	75
4	T_1 Heavy Atoms.	.79
5	Native and Derivative Unit Cell Constants for T_1	.83
6	Some Quantities Related to Data Collection.	.87
7	Relative Scale and B Factors of Derivatives of T_1	90
8	Harker Vectors and Cross-vectors.	.92
9	Mercury Positions for MRS in Unit Cell Fractions.	.93
10	SHELX Refinement of the Mercury Atoms in MRS.	105
11	Progress of the ISIR Refinements.	127
12	Distribution of Figures of Merit for ISIR Refinements.	.134

LIST OF FIGURES

FIGURE	PAGE
1	Capillary Mounting of a Protein Crystal. 6
2	Four-circle Diffractometer Geometry. Twice the Bragg angle is 2θ , ϕ and χ are required to orient the crystal such that the diffracted beams are in the plane of the detector, and ω facilitates the initial location of those beams. Each of the four rotation axes passes through the crystal. 9
3	The Phase Ambiguity and Its Resolution. Possible solution vectors lie at the intersection of the circles. If two or more derivatives are used, the solution is unique. . . 22
4	The Specificities of Trypsin, Elastase, and Chymotrypsin. These are controlled by the geometry and charge of the specificity pocket of the enzyme. 29
5	Proposed Mechanism for Serine Protease Proteolysis (Ref. 12). 30
6	Protein Protease Inhibitor Relationships (Ref. 17). 32
7	Disulfide Linkages in Serine Protease Inhibitors. (Δ indicates active site location) (Ref. 17). 34
8	Amino Acid Composition and Sequence of BBI (Ref. 33). 38
9	Sequences of Members of the BBI Family. 40
10	Two Morphologies of BBI Ternary Complex Crystals. 48
11	Axial Intensities of BBI Ternary Complex Crystals with PEG Concentration. 51

FIGURE	PAGE
12	Axial Intensities of BBI Ternary Complex Crystals with pH. 53
13	Axial Intensities of Two Forms of BBI Ternary Complex. 55
14	SDS Gel Electrophoresis of BBI Ternary Complex and Its Components. Lane 1, bovine β -trypsin; Lane 2, bovine α -chymotrypsin (from top to bottom, chains B, C, A, respectively); Lane 3, BBI; Lane 4, ternary complex. 58
15	Amino Acid Sequence of Ribonuclease T_1 (Ref. 47). 64
16	Proposed Mechanism for T_1 Catalysis (Ref. 47). . . 66
17	Homologs of Ribonuclease T_1 (Ref. 64). 69
18	Mounting a Crystal in Saturated Ammonium Sulfate. This technique prevents the precipitation of ammonium sulfate near the protein crystal. 73
19	Axial Intensities of T_1 and Heavy Atom Derivatives. 77
20	Molecular Structure of Some Heavy Atom Reagents. . 78
21	Radial Distribution of $ F ^2$ for T_1 Derivatives. . 89
22	Harker Sections of MRS Difference Patterson Map. The map is contoured at a level corresponding to a vector between two 30 electron peaks, and incremented at the same level. 94
23	Harker Sections of PT Difference Patterson Map. The map is contoured at a level corresponding to a vector between two 30 electron peaks, and incremented at the same level. 95
24	Harker Sections of PTP Difference Patterson Map. This map is contoured at a level corresponding to a vector between two 30 electron peaks, and incremented at the same level. 97

25	Harker Sections of 8IG Difference Patterson Map. The map is contoured at a level corresponding to a vector between two 30 electron peaks, and incremented at the same level.	99
26	MRS Difference Electron Density Projections. This map is contoured in electron density levels starting at 30 electrons in increments of 5 electrons.	104
27	PT Difference Electron Density Projections. This map is contoured in electron density levels starting at 30 electrons and in increments of 5 electrons.	109
28	PTP Difference Electron Density Projections. This map is contoured in electron density levels starting at 30 electrons and in increments of 5 electrons.	110
29	8IG Difference Electron Density Projections. This map is contoured in electron density levels starting at 30 electrons and in increments of 5 electrons.	111
30	The Single Isomorphous Replacement Phase Ambiguity. In this construction α_h is the heavy atom phase angle and α is $(\alpha_h - \alpha_n)$ where α_n is the native protein phase angle. . .	114
31	A Section of ρ_j Defining the Molecular Envelope. The $y = 0$ section of the modified electron density map defining the molecular envelope used in the real space electron density filter in ISIR is shown.	119
32	An example of Reciprocal Space ISIR Phase Filtering.	122
33	A 2.3 Å Slice Through the Unit Cell Calculated with 35% Solvent (from $y = 0$).	130
34	Figure of Merit Distributions for ISIR Refinements. Numbers indicate percent of reflection with m greater than the value. . . .	133
35	Unit Cell Electron Density Greater than 5σ . The dimer is outlined.	140

INTRODUCTION

The technique of x-ray crystallography has revolutionized the study of protein chemistry and biochemistry in the thirty years of its application. It has only been within this time that the three-dimensional structures of some biological macromolecules have been elucidated, along with the principles that govern their structures. Determining the structure of a protein using crystallography is a very complex process, requiring years of effort even under the best circumstances, but the rewards are also great. No other technique can provide the wealth of structural detail, and few techniques rely on so few chemical assumptions.

The work described herein will trace the process that leads from a lyophilized protein powder to the electron density map of a protein. The first stages, those of producing large, reproducible crystals that diffract well, will be reported for a ternary protein complex consisting of two proteases and a protease inhibitor. The latter stages of the process, from analysis of the crystals through calculation of the electron density map, will be reported for a bacterial enzyme.

In order to put the process in perspective, and to set forth definitions, a generalized overview of protein crystallography will first be presented. This will be followed by two sections, each of which will begin with a short background of the proteins involved.

I. OVERVIEW OF PROTEIN X-RAY CRYSTALLOGRAPHY

Proteins are biological macromolecules which vary in size from a few thousand daltons to several million. They are linear polymers of 21 different α -amino acids linked together by a very stable (104 kcal/mole) amide bond referred to as a peptide bond. Each of the 21 amino acids has a different group on its α carbon referred to as its side chain. A long main chain (the peptide backbone) with many smaller side chains results. The complete covalent structure of a protein is known as its primary structure, and it is this primary structure that is thought ultimately to determine the three-dimensional structure of a protein. The principal characteristics of the primary structure are the sequence of amino acids, and the location of any crosslinks between parts of the main chain. These crosslinks occur because one of the amino acids, cysteine, contains a sulfhydryl group which, under certain conditions, may be oxidized along with a second one to form a disulfide linkage.

Thus, connections are made along the main chain between amino acids which are not adjacent in the sequence.

Due to the hydrophobic character of many of the side chains, the main chain of water soluble proteins bends back on itself and forms a globule, like a tiny oil drop. These globular proteins, unlike oil drops, possess characteristic structure. They also contain regions of high local order known as secondary structure. One basic configuration the main chain may assume is that of a helix. In a helix, the chain spirals and is stabilized by hydrogen bonds between the imino nitrogen of one amino acid with the carbonyl oxygen of another. For example, in the α -helix, hydrogen bonds are formed between the carbonyl of one amino acid, and the peptide nitrogen four residues away in the sequence. Another secondary structural feature, β -sheet, can be thought of as a two-fold helix, or a pleated structure. There are no hydrogen bonds formed along the strand, but strands are often found next to each other and many hydrogen bonds can form between the imino nitrogen of one strand and the carbonyl oxygen of another.

The overall folding pattern of a protein main chain along with the side chain stereochemistry is known as tertiary structure, and it is usually a connected mosaic of secondary structures. Determination of the folding pattern of the main chain is one of the principal

objectives in protein crystallography. Folding brings together side chains from different parts of the peptide chain, which in turn determines the protein's chemical and functional properties.

Quaternary structure describes the aggregation of protein molecules into small or large oligomeric structures; such aggregation can result in altered properties for the proteins. The oligomeric structures range from simple dimers to complex viruses (up to 2000 units).

II. OVERVIEW OF EXPERIMENTAL TECHNIQUES

Proteins, like other molecules, can be crystallized, but their large size and nonrigid character give protein crystals some nonclassical characteristics. Protein crystals are typically only about half protein, the remainder being mother liquor (the crystallization solution). Since proteins are so large, the interstitial spaces between them are large, and those spaces are filled by solvent, much of which is disordered. This, along with the nonrigidity of the protein, make the periodicity of the crystals imperfect, and thus the crystals are usually poor x-ray scatterers. The crystals are also mechanically very weak, and are susceptible to changes in pH, concentration of salts in the solvent, and to precipitant concentration.

Since the crystals contain so much solvent, they can dry and lose their crystallinity in the absence of solvent. It is important that the crystals be stable during x-ray intensity data collection. To maintain stable conditions, the crystals are sealed in thin walled (typically 0.03 mm thick) glass capillaries, along with a drop of mother liquor. The crystal is affixed to the side of the capillary (as shown in Figure 1), by a drop of mother liquor, and is held in place by its surface tension. A well mounted crystal will drift no more than 0.05° in any direction per day. The capillary is sealed with wax so that the mother liquor reservoir does not dry out. Epoxy glue is often spread over the outside of the wax as an additional precaution against the crystal drying out. The capillary is then mounted on a goniometer head, a platform set on two perpendicular arcs which can also be translated in three perpendicular directions. With this combination of arc and translation motion, the crystal can be very accurately aligned so that when it is mounted on an x-ray diffractometer, it can be placed in the middle of the x-ray beam in a well defined orientation.

As will be described later, it is necessary in protein x-ray crystallography to measure the diffraction intensities from several different crystals. In order to minimize the effects of systematic errors, it is desirable to mount each crystal in an identical

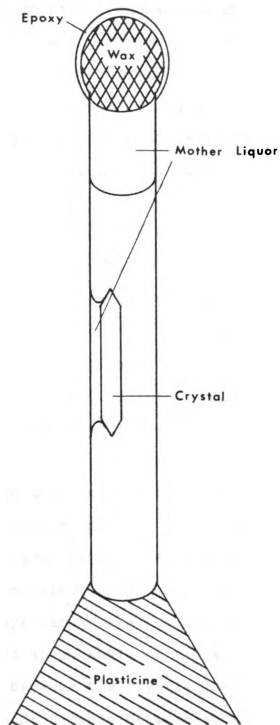


Figure 1: Capillary Mounting of a Protein Crystal.

configuration. This configuration should be easy to identify by the morphology of the crystal to minimize the handling of the crystals, and should contain a significant crystallographic direction nearly collinear with the capillary tube to facilitate the measurement of absorption of x-rays by the crystal.

Protein crystals, when handled carefully, give measureable x-ray diffraction intensities. As noted above, they do not diffract outstandingly well, and measurable intensities beyond 2 Å resolution are rather rare. Since the unit cell size is very large, the diffracted beams are angularly very close together. Thus, the detector of the diffractometer must be moved further from the crystal than it typically is in small molecule diffraction to resolve the diffracted beams. Helium is passed through the detector collimator to minimize absorption of the x-rays by the long pathlength of air.

The bulk of the work described within this paper is the result of measurements made on a four-circle x-ray diffractometer. X-rays, having a wavelength comparable to the molecular spacing within crystals, are diffracted by crystals. The x-rays strike the crystal at an angle, θ , and are scattered at an angle of reflection also of θ . If the spacing between the crystal lattice planes is d , then the path difference between an x-ray striking the first lattice plane, and another striking the one below it is $2d \sin \theta$. When θ is such that this distance equals

an integral number of wavelengths, constructive interference occurs, and a diffracted beam results in that direction.

This is Bragg's law:

$$n\lambda = 2d \sin \theta$$

The angle of the diffracted beam, with respect to the incident beam, is then 2θ , a variable dependent only upon the wavelength of radiation and the lattice spacings of the crystal. The angle of the detector of the four-circle diffractometer with respect to the beam is 2θ . The x-ray source, crystal, and detector are confined to the equatorial plane of the diffractometer, so two more angles are required to orient the crystal such that the diffracted beam also occurs within the equatorial plane. One of these angles, χ , lies in a plane perpendicular to the equatorial plane, and the other ϕ , lies perpendicular to the χ plane. A fourth angle, ω , added for convenience and in the equatorial plane, is for precise alignment of θ (Figure 2). This fourth degree of freedom greatly facilitates locating the reciprocal lattice initially, (i.e. -- the three-dimensional net of conditions satisfying Bragg's law for a particular crystal), but during automatic data collections ω is set equal to θ .

The electron density can be calculated from a quantity known as the structure factor (F), the amplitude of which is derived from the intensities measured in the diffraction experiment using the following relation:

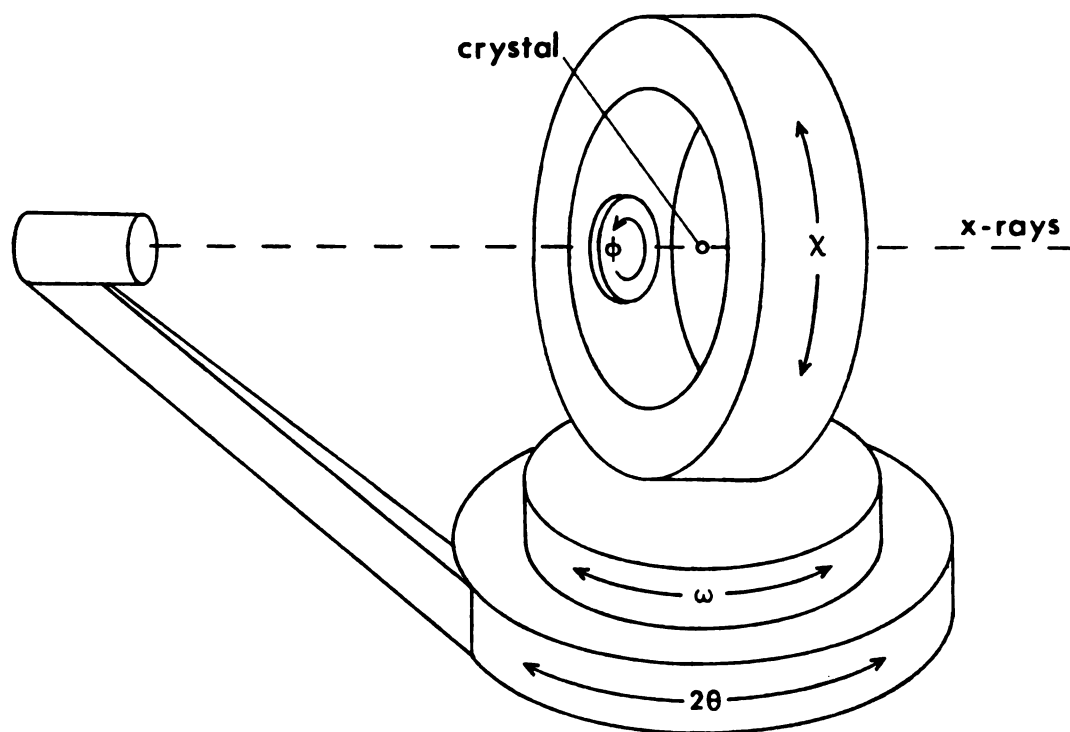


Figure 2: Four-circle Diffractometer Geometry. Twice the Bragg angle is 2θ , ϕ and χ are required to orient the crystal such that the diffracted beams are in the plane of the detector, and ω facilitates the initial location of those beams. Each of the four rotation axes passes through the crystal.

$$F^2 = (F)(F^*)$$

$$= I \text{ (lack of balance) (scale) (absorption) (decay) (LP) ,}$$

where F^* is the complex conjugate of F .

The intensity (I) is measured in counts per unit time, and is corrected for background. The intensities in this work were measured either on a Picker FACS-1 or on a Nicolet P3/F diffractometer. The integrated intensities for the reflections were approximated by measuring them with a wandering ω step scan. In this procedure the crystal is stepped through the ω angle in small increments (typically about 0.03°) through the highest part of the peak. If the crystal is not quite aligned, the diffractometer may take up to a specified number of steps (usually 3) beyond its usual number (usually 7). Only the steps with the highest intensities are used to calculate the total integrated intensity (usually 5 of 7).

The FACS-1 has been modified for protein data collection, and uses balanced nickel-cobalt filters to achieve monochromatic copper radiation. The nickel filter removes the K_β radiation from the copper x-ray spectrum because it has an absorption edge at approximately that wavelength (the copper K_β wavelength is $1.392 \text{ \AA}$, and the nickel absorption edge is at $1.488 \text{ \AA}$). The background is measured at the peak position, with a balanced cobalt

filter removing the K_{α} radiation (the copper K_{α} is at 1.542 Å, and the cobalt absorption edge is at 1.608 Å). The balance of the filters is not perfect and has a 2θ dependence, so a correction in the intensity of the background must be made due to the lack of balance between the nickel and cobalt filters.

The P3/F uses a graphite monochromator to achieve monochromatic copper radiation. This is a graphite crystal which diffracts the unfiltered x-rays at different Bragg angles depending on their wavelength. The monochromator is precisely aligned so that only the copper K_{α} wavelength of a given reflection passes through the collimator to the specimen crystal.

The background of the data sets from the P3/F was averaged in shells of 2θ . This results in more accurate measurements because the background is measured for only a short time on each reflection (typically one-fifth as long as the peak), and since the background intensity is low, the statistical fluctuations can significantly change low intensity reflection intensities. Averaging virtually eliminates the background source of error in the intensity measurement.

A scale factor must be applied to the intensity data so that data from several crystals can be compared. Among other things, the intensity is a function of crystal size, and so will vary from crystal to crystal. In the subsequent stages of the process this scale factor is

adjusted to put the structure factors on an absolute scale so that they can be converted to electrons per cubic Angstrom in the electron density map.

The crystal absorbs x-rays as well as scatters them, and this absorption will decrease the intensities. The absolute value of this absorption is not important for structural work since the data will be scaled, but it is important to know how the absorption varies with different orientations of the crystal. There is no easy way to deal with this, especially with protein crystals where both the capillary that the crystal is mounted in, and the mother liquor that affixes it to the capillary, also absorb the x-rays. In the present data sets the correction was made by the method of North, Phillips and Mathews (Ref. 1). The correction is a semiempirical one in which reflections at or near $\chi = 90^\circ$ are measured at various ϕ values. As the crystal is rotated about its diffraction vector, variations in the intensities about ϕ should be solely the result of differences in absorption. The relative values for absorption are then corrected for variations in path length of the beam through the crystal by the following equation:

$$\text{Absorption} = [A(\phi + \theta') + A(\phi - \theta')]/2 ,$$

where

$$\theta' = \arcsin[\lambda d \cos(\chi/2)] ,$$

where λ is the x-ray wavelength and d^* is the projection of the hkl vector on the equatorial plane.

There is often an additional dependence of the absorption on the height of the reflection above the equatorial plane, and its correction is approximated by using different absorption vs. ϕ tables for different heights.

The tertiary structure of proteins is maintained only through weak hydrogen bonds, ionic, and nonpolar interactions. The input of energy into such a system by the x-ray beam is sufficient to disturb these interactions and disrupt the structure of the protein crystal. The result is, that on exposure to x-rays, protein crystals tend to deteriorate and become poorer scatterers. Thus, the intensity of reflections measured at the end of a data collection tend to be smaller than those measured at the beginning. The decay is more severe for the higher resolution data, because smaller structural changes are more easily seen at higher resolution. The decay is corrected for in a linear way in different ranges of 2θ . Where the size of the crystal permitted, the x-ray flux was lowered to decrease the decay.

The amount of decay was monitored in a variety of ways. At intervals of 100 reflections, three check reflections were measured and their intensity was plotted as a function of either reflection number or exposure time. The corrected intensity was then taken to be:

$$I'(t) = I(t)[1/1-St]$$

where I' is the corrected intensity as a function of time (or reflection number), I is the measured intensity as a function of time, t is the time (or reflection number) and $S = -s/I(t=0)$ where s is the slope of the intensity versus time graph for the monitor reflections, and $I(t=0)$ is the intensity at time zero.

For the low resolution data, below $2\theta = 15^\circ$, decay was determined from the decrease in peak heights in Patterson maps (see below) calculated from the (h0l) zone which was measured both at the beginning and at the end of the data collection. The ratio of the peak heights calculated from the data measured before the data collection, and those calculated from data measured after the data collection, is used to obtain an intensity versus time line. The slope of this line is then used to calculate the correction.

A third monitor of the decay can be made with data taken on the P3/F diffractometer. The crystal is realigned at least daily either by the diffractometer sensing misalignment due to a drop in the monitor reflection intensities, or by manual intervention. When the alignment reflection array reflections are recentered, their intensities are recorded also. The sum of the intensities in the reflection array can be plotted against

time and the slope, s , can be calculated and the correction applied as above.

The LP factor is actually a combination of two different factors, the Lorentz factor, which is geometric and dependent on the measuring instrument, and the polarization factor.

The polarization factor arises because x-rays which have electric components parallel to the reflecting plane are reflected more efficiently than those with electric components perpendicular to the plane. The x-rays with electric components parallel are reflected with an intensity dependent only on the electron density. Those with electric components perpendicular have intensities which vary both with electron density and scattering angle, 2θ as follows:

$$I(\text{perp}) = I_0(\text{perp}) \cos^2(2\theta) .$$

But $I_0(\text{perp}) = I(\text{par})$ for a perfectly unpolarized beam, since statistically there will be as many perpendicular components as parallel. Hence, the total reflected intensity is

$$I(T) = I(\text{par}) + I(\text{perp}) = I(\text{par})[1 + \cos^2(2\theta)] .$$

In terms of the incident beam, since the parallel components will make up half of it,

$$I(T) = I_0(T)[1/2(1 + \cos^2(2\theta))] .$$

The correction is complicated in the case of the P3/F by the monochromator which is also crystalline and so subject to polarization effects. The polarization factor for a monochromated beam is:

$$\begin{aligned} & [f(1 + \cos(2\theta_m) \cos^2(2\theta_c))]/[1 + \cos^2(2\theta_m)] \\ & + [(1-f)(1 + \cos(2\theta_m) \cos^2(2\theta_c))]/[1 + \cos(2\theta_m)] \end{aligned}$$

where m and c refer to the monochromator and sample crystal, respectively, and f is related to mosaicity of the monochromator crystal (Ref. 2), (f is approximately 0.5 for the monochromator on the Nicolet P3/F).

The Lorentz factor arises when a crystal is moving, because the time required for a reflection to travel through the diffracting condition is not constant, but varies with its position in reciprocal space. With a stationary crystal measurement, such as was used in the data sets described herein, divergence of the beam with scattering angles has a similar effect. This varies with the measuring device, but for a four-circle diffractometer in the bisecting mode, such as both the diffractometers used,

$$L = \frac{1}{\sin 2\theta}.$$

So, for the FACS-1 diffractometer,

$$LP = \frac{1 + \cos^2(2\theta)}{2 \sin(2\theta)}$$

and for the P3/F,

$$LP = [1/\sin (2\theta)] \{ [f(1 + \cos^2 (2\theta_m) \cos^2 (2\theta_c)]/[1 + \cos^2 (2\theta_m)] \\ + [(1 - f)(1 + \cos (2\theta_m) \cos^2 (2\theta_c)]/[1 + \cos (2\theta_m)] \} .$$

In summary, the amplitude of the structure factor for each reflection can be calculated from accurate measurements of the intensity of the diffracted beams and application of appropriate corrections.

III. OVERVIEW OF THE ELECTRON DENSITY MAP CALCULATION

The structure factor itself is a measure of the scattering power of the unit cell contents, and is the sum of the scattering from each atom within the unit cell. Since each atom will scatter x-rays, the scattered x-rays will interfere with each other, the amount of constructive or destructive interference depending upon the relative positions of the atoms within the unit cell.

The form factor, f , is the ratio of ability of an atom to scatter x-rays to that of a single electron. The scattering of x-rays by the nucleus is small and usually ignored. So at zero scattering angle the form factor is equal to the number of electrons in the atom. The form factor decreases at higher scattering angles, and the exact form of this decrease depends upon the atomic model

used. Usually Hartree-Fock atoms are assumed to calculate the scattering.

If the scattering from each atom in the unit cell is represented by its form factor f , the total scattering from the unit cell is:

$$\vec{F} = \sum_j f(j) e^{2\pi i(\vec{s} \cdot \vec{r}_j)} \quad (1)$$

where $\vec{r}$ is the vector position of the scatterer (x, y, z) and $\vec{s}$ is the diffraction vector (h, k, l), such that $\vec{r} \cdot \vec{s} = (hx + ky + kz)$. The electron density of the unit cell is the Fourier transform of the structure factor given by:

$$\rho(\vec{r}) = \frac{1}{V} \int_{\substack{\text{V} \\ \text{diffraction} \\ \text{space}}} \vec{F}(\vec{s}) e^{-2\pi i(\vec{s} \cdot \vec{r})} dV ,$$

but $\vec{F}(\vec{s})$ is zero everywhere except where the Bragg condition holds, so the integral can be replaced by the sum:

$$\rho(\vec{r}) = \frac{1}{V} \sum_{\vec{s}} \vec{F}(\vec{s}) e^{-2\pi i(\vec{s} \cdot \vec{r})} .$$

A fundamental problem arises between the measurement of the beam intensities and the calculation of the electron density. The phase of the diffracted beam must be known to calculate the electron density, and it is dependent on the atomic positions, which are not known, but are to be derived from the electron density. This is the "phase problem" that is at the heart of x-ray crystallography. Direct methods for solving the phase problem, where phase

information is determined by conditions such as the electron density, are positive everywhere, are available, but these direct methods become less precise with increasing numbers of atoms with positions to be determined. Even for small proteins, the problem is considerably larger than has thus far been proven tractable by direct methods.

One method that has been useful in determining the positions of atoms within the unit cell is the use of Patterson maps. A Patterson map is a Fourier synthesis with coefficients $|F|^2$

$$P(uvw) = \sum_h \sum_k \sum_l |F_{(hkl)}|^2 e^{-2\pi i(hu + kv + lw)},$$

where u , v , and w are along unit cell edges. The effect of this synthesis is to produce a map showing the vectors between the atoms in the unit cell (Ref. 3). For small molecules these vectors can be deconvoluted to give atomic positions, but for proteins there are so many atoms that the peaks overlap and become uninterpretable. If, however, there are a few atoms in the unit cell with many more electrons than the remainder, their interatomic vectors will stand out above the background, and their positions can be determined.

The method of choice for solving the phase problem for proteins has generally been the multiple isomorphous replacement (MIR) method. Since protein crystals are about half mother liquor, heavy atom salts can be diffused into

the crystals. If these bind in a systematic way to the protein, they will produce changes in the intensity distribution of the diffracted beams. If the structure of the protein is not appreciably altered by the heavy atom, and if the crystal structure is the same, phase information can be obtained. Data are collected from the heavy atom derivative, and processed as described above to obtain $|F|$'s for the derivative. These are then scaled to the native protein, and a difference Patterson map calculated. This is like the Patterson map described above, except that the coefficients are $(|F_d| - |F_n|)^2$, where $|F_d|$ is the amplitude of the heavy atom derivative structure factor, and $|F_n|$ is the amplitude of the native structure factor. This removes the protein background from the Patterson map and, if the crystals are isomorphous, results in only displaying the vectors between heavy atoms $(|F_h| |F_h|)$, and between the heavy atoms and the protein $(|F_h| |F_p|)$. Since the heavy atom has many more electrons than the protein atoms, the first product will be much greater than the second, so the dominant peaks will be those between the heavy atoms. From these vectors, the heavy atom positions can be deduced.

If the coordinates of the heavy atoms are known, their phases are also known from Equation (1). If the derivative is exactly isomorphous, $\vec{F}_d = \vec{F}_n + \vec{F}_h$, where $\vec{F}_h$ is the structure factor due to the heavy atom. The amplitudes of both $\vec{F}_d$ and $\vec{F}_n$ are known, and both the

amplitude and phase of $\vec{F}_h$ is known. This can be conveniently represented by an Argand diagram (Ref. 4) as shown in Figure 3. If $|\vec{F}_n|$ is known, $\vec{F}_n$ must be a vector whose tip lies somewhere along the origin-centered circle. The vector $-\vec{F}_h$ is known, and drawn at its correct phase and amplitude with respect to the origin. A second circle is drawn with its origin at the tip of $-\vec{F}_h$, and radius $|\vec{F}_d|$. Where the circles cross $\vec{F}_n = \vec{F}_d - \vec{F}_h$, and the condition is satisfied.

Alternately, consider $\vec{F}_d$ as the vector sum of $\vec{F}_p$ and $\vec{F}_h$. From the law of cosines, if the phase angle of the native protein is α_n , and the phase angle of the heavy atom is α_h ,

$$(\alpha_h - \alpha_n) = \arccos \frac{|\vec{F}_d|^2 - |\vec{F}_h|^2 - |\vec{F}_n|^2}{2|\vec{F}_h||\vec{F}_n|} .$$

It is obvious that both of these constructions lead to an ambiguity in the phase angle. In the first, there are two points where the phase circles cross, and in the second, each value for arc cosine corresponds to two different values for the phase angle.

The ambiguity is removed by the addition of a third derivative, which will also give two different values for the phase angle, but in general, only one will satisfy the conditions for both derivatives.

While the MIR method is usually the method of choice in protein crystallography, forming isomorphous heavy

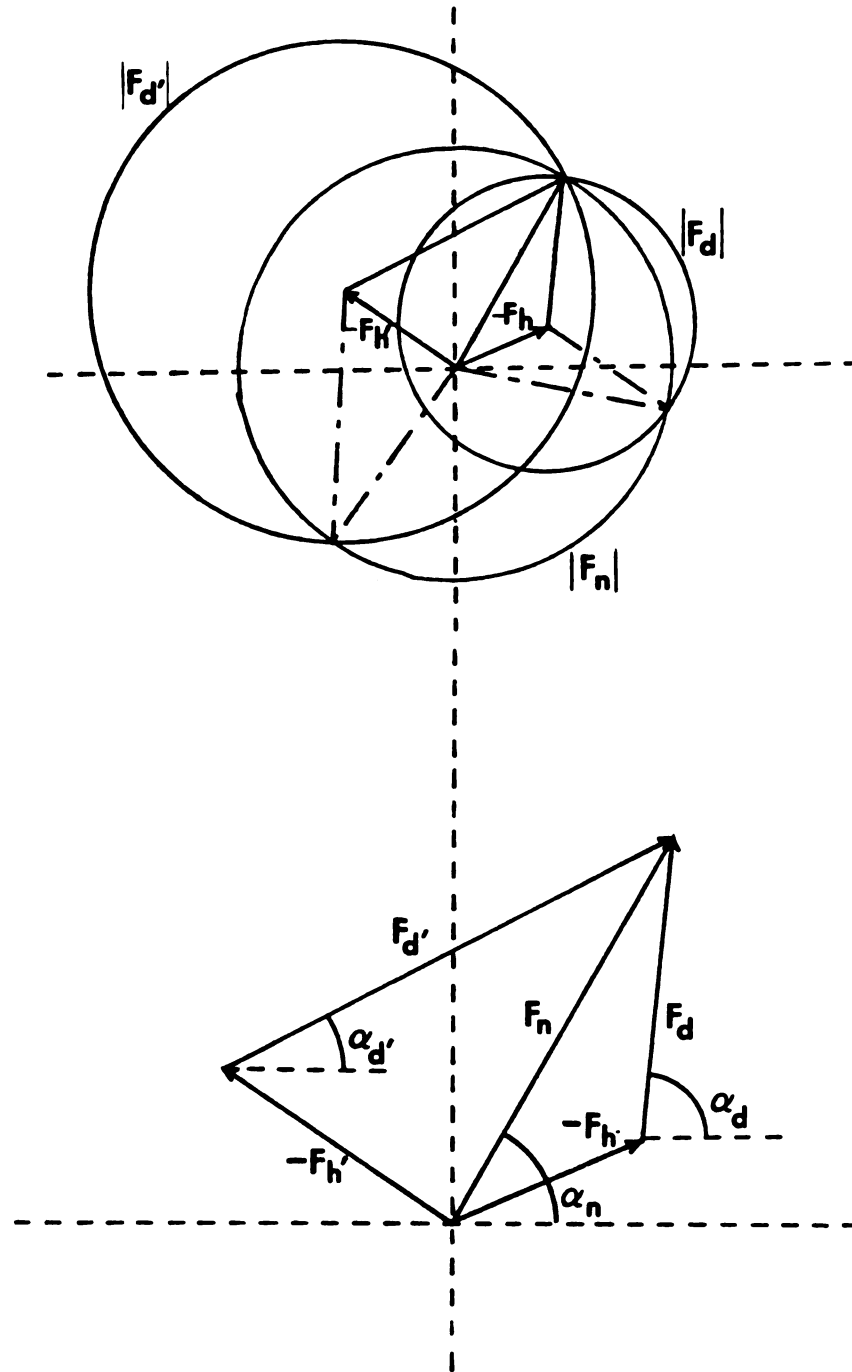


Figure 3: The Phase Ambiguity and Its Resolution. Possible solution vectors lie at the intersection of the circles. If two or more derivatives are used, the solution is unique.

atom derivatives in practice is often difficult. If, however, one can obtain just one heavy atom derivative, phase information can still be derived. The method is referred to as single isomorphous replacement (SIR) and the logic behind it is as follows. The heavy atom position yields two possible phase angles. If both phase angles are used to calculate a double phased Fourier, the true image of the electron density is generated plus an image from the false phase terms. If the false phases are randomly distributed, they will create no image, only a background. The result is a true electron density map superimposed on a rather higher background.

A variation in the SIR method has been developed by B.C. Wang (Ref. 5). In Wang's method a double phased Fourier is first calculated. Then the molecular envelope is defined by a density modification function which acts on the density to separate it into broad regions of high and low electron density. It is assumed that the protein will have a higher average electron density, so these broad regions of high density are considered protein, and those regions of low density are considered solvent. This map is then contoured such that the percentage of grid points in the map having density below a cut-off value is equal to the percentage of solvent in the crystal. This then defines the molecular envelope and creates a "density filter". The electron density map is then compared with the density filter. The electron

density in those regions outside the molecular envelope are given a constant value, and those within the molecular envelope are scaled upward. An inverse Fourier transform on the resulting map gives a set of phases which are then used to improve the double phased Fourier map. The probability of a phase angle being the correct one from the double phased Fourier, and the same quantity from the improved Fourier map, are multiplied together. In the best case this process performs a phase selection from the double phased Fourier, and the result of the "phase filtering" is a better map. This new map is then used instead of the double phased Fourier, and cycled through the process again. The resulting map is repeatedly cycled to convergence. This process is known as the iterative SIR method, or ISIR.

In practice, the choice of phase angles is not as clear as the constructions indicate. This is because of errors in the measurements, lack of isomorphism, errors in the exact position and occupancy of the heavy atoms, and the like. The result is that the phase circles do not intersect at one point, and the phase triangles do not neatly close. The phase angle used in this work is that which will give the "best" electron density map by the method of Blow and Crick (Ref. 6). "We define the 'best Fourier' as that Fourier transform which is expected to have the minimum mean square difference from the Fourier transform of the true $\vec{F}$'s when averaged over the whole

unit cell. The best Fourier is the Fourier transform obtained by using the centroid of the probability distribution $\vec{F}$ " (Ref. 6)

Once the electron density map is calculated, a model structure is fitted into the density. First, sections of the electron density are drawn on plexiglass sheets which are stacked to obtain a three-dimensional construction. This is done to discern molecular boundaries and to look for areas of helix, sheet, or other features known to be in the protein. The density map is then represented on an Evans and Sutherland PS300 vector graphics system, and a line model built within the three-dimensional "basket" contours of the electron density map.

FORMATION, CRYSTALLIZATION, AND PRELIMINARY
CRYSTALLOGRAPHIC DATA ON THE TERNARY COMPLEX
OF α -CHYMOTRYPSIN, β -TRYPSIN, AND THE
BOWMAN-BIRK INHIBITOR

I. INTRODUCTION

A. Proteases

Proteases are relatively small proteins (25,000 to 35,000 dalton) which cleave other proteins at peptide bonds. By the action of these proteases, a protein may undergo full proteolysis, where it is broken down to its constituent amino acids (as happens in digestion), or it may undergo limited proteolysis, where only a few of the peptide bonds are cleaved, (as in many of the control functions of the body).

All proteases studied to date fall into one of four mechanistic classes. Enzymes in each of these classes utilize a general acid-general base type of cleavage mechanism, but the side chains that act as the acid and base differ. Metallo proteases (such as carboxypeptidase A and B) have a metal center which acts as a Lewis acid in general acid-base catalysis (Ref. 7). Carboxyl or

acid proteases (such as pepsin) have a carboxyl group, from either an aspartic acid or glutamic acid (Ref. 8); sulfhydryl or thio proteases (such as papain) use a sulfhydryl group from cysteine (Ref. 9); and the serine proteases use the hydroxyl group of a serine.

Of the four groups, the serine proteases have been by far the most extensively studied. Three of these have had a particularly important role in understanding the structure and action of serine proteases; trypsin, chymotrypsin, and elastase. The three are highly homologous in amino acid sequence (Ref. 10), and in their three-dimensional structure (as determined by x-ray diffraction). The central differences among them lie in their specificity. Trypsin cleaves a peptide bond most efficiently on the carboxyl side of positively charged amino acids (lysine and arginine), elastase cleaves peptide bonds best on the carboxyl side of small amino acids (alanine and valine), and chymotrypsin favors the peptide bond on the carboxyl side of large amino acids (tryptophan, phenylalanine, tyrosine, leucine).

Analysis of the three-dimensional structures by x-ray diffraction has yielded reasons for such specificity. In chymotrypsin there is a "specificity pocket" next to the active site. The pocket is large, hydrophobic, and has a wide mouth. It is thus a low energy environment for a large, hydrophobic side-chain such as that of tryptophan, phenylalanine, tyrosine, or leucine (Ref. 11).

In trypsin, a very similar pocket occurs, but with an important difference. At the bottom of the pocket lies Asp 189 which, above about pH 5.0, is deprotonated. Thus, there is a negative charge at the bottom of the pocket. This makes it especially attractive for large side-chains with a positive charge, such as lysine or arginine (Ref. 12). Elastase also has this specificity pocket, but whereas both chymotrypsin and trypsin have glycine residues (which have no side chain) at the mouth of the pocket, elastase has Val 216 on one side, and Thr 226 on the other side. The pocket is thus effectively blocked to any but small side-chains, such as alanine and valine (Ref. 13; Figure 4).

While the specificity of these three enzymes differs, their catalytic mechanisms are believed to be the same (Refs. 12-14). The side-chain of the protein or peptide to be cleaved fits into the specificity pocket bringing its carbonyl carbon near Ser 195 (chymotrypsin numbering). The carbonyl oxygen is drawn into an "oxy-anion hole", made up of the imino nitrogens from Gly 193 and Ser 195. This distorts the carbonyl carbon into a pyramidal configuration making it susceptible to attack by the γ oxygen of Ser 195 (Ref. 15), which is given considerable nucleophilic character by its interaction with Asp 102 through His 57 as an intermediary (Ref. 16) (the "charge transfer system") (Figure 5). This results in the breaking of the peptide bond and the formation of an "acyl-enzyme"

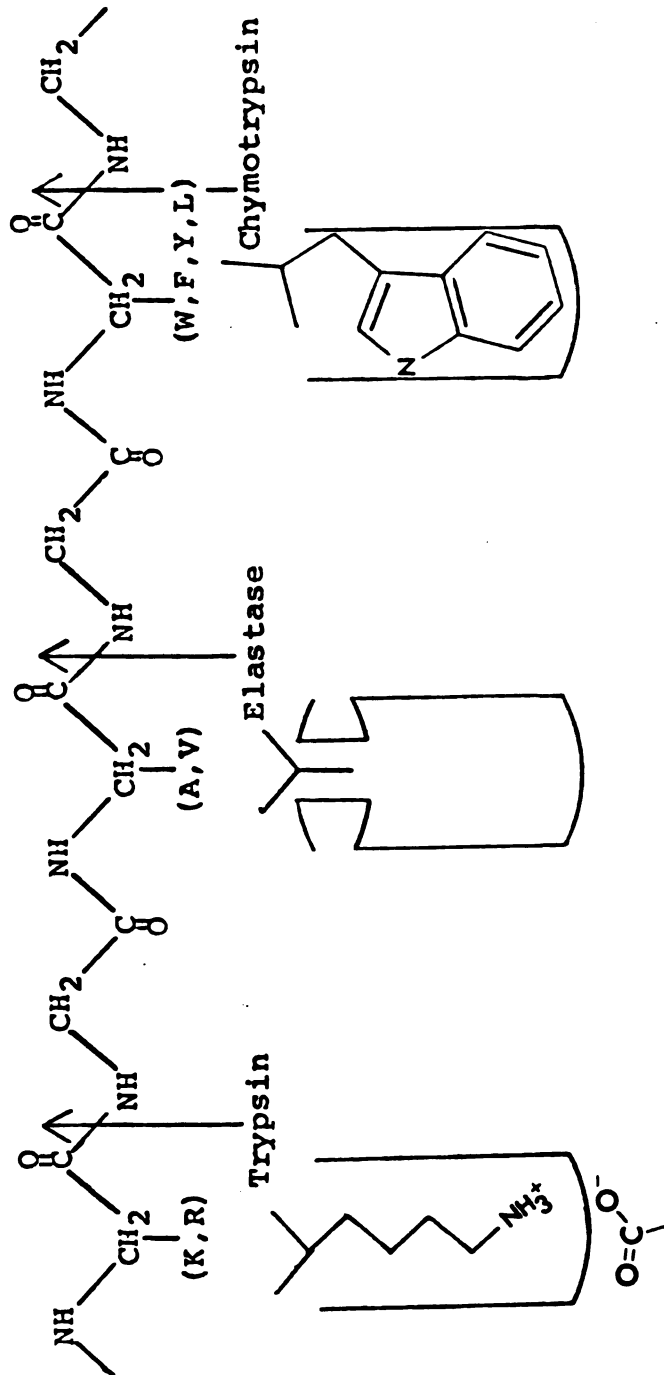


Figure 4: The Specificities of Trypsin, Elastase, and Chymotrypsin. These are controlled by the geometry and charge of the specificity pocket of the enzyme.

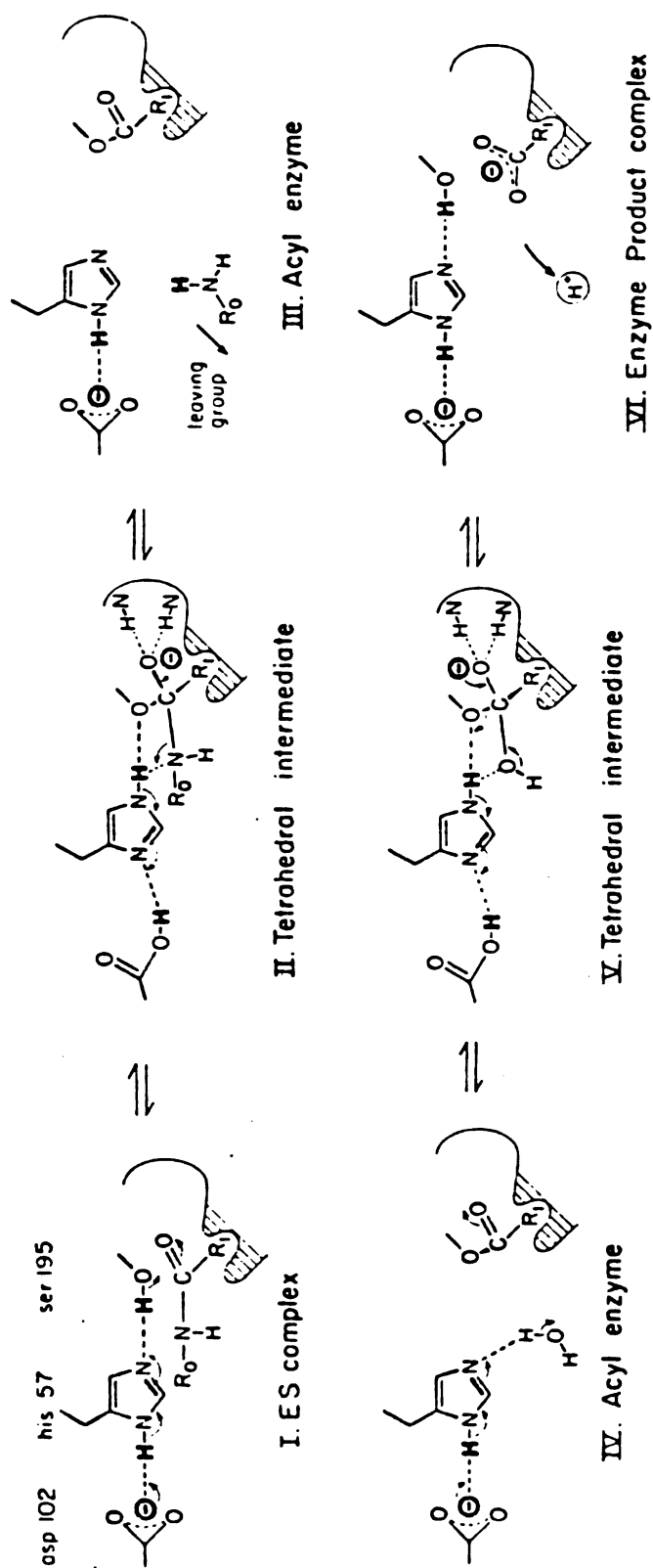


Figure 5: Proposed Mechanism for Serine Protease Proteolysis (Ref. 12).

intermediate. The amino fragment of the cleaved protein diffuses away from the enzyme. A water molecule then replaces it in the active site, and through a reversal of the above reaction the acyl-enzyme is cleaved, regenerating the enzyme and cleaving off the second peptide fragment.

B. Protein Protease Inhibitors

Protein protease inhibitors are found in all types of living organisms and in all sorts of tissues. Grossly, their function is to prevent unwanted proteolysis, but their exact physiological function is seldom well understood. They are often part of complex control mechanisms and are known to be involved in such diverse processes as digestion, blood coagulation, and fertilization. Often, the proteases they inhibit physiologically are not known; nevertheless, some patterns in their structure, function, and evolution have begun to emerge.

Perhaps the most comprehensive organization of these inhibitors has been the scheme described by Laskowski and Kato (Ref. 17), which is summarized in Figure 6. Laskowski and Kato identify two basic types of inhibitors; macroglobulins, and the class specific protease inhibitors.

Macroglobulins inhibit proteases from all four mechanistic classes. Inhibit is perhaps an improper term

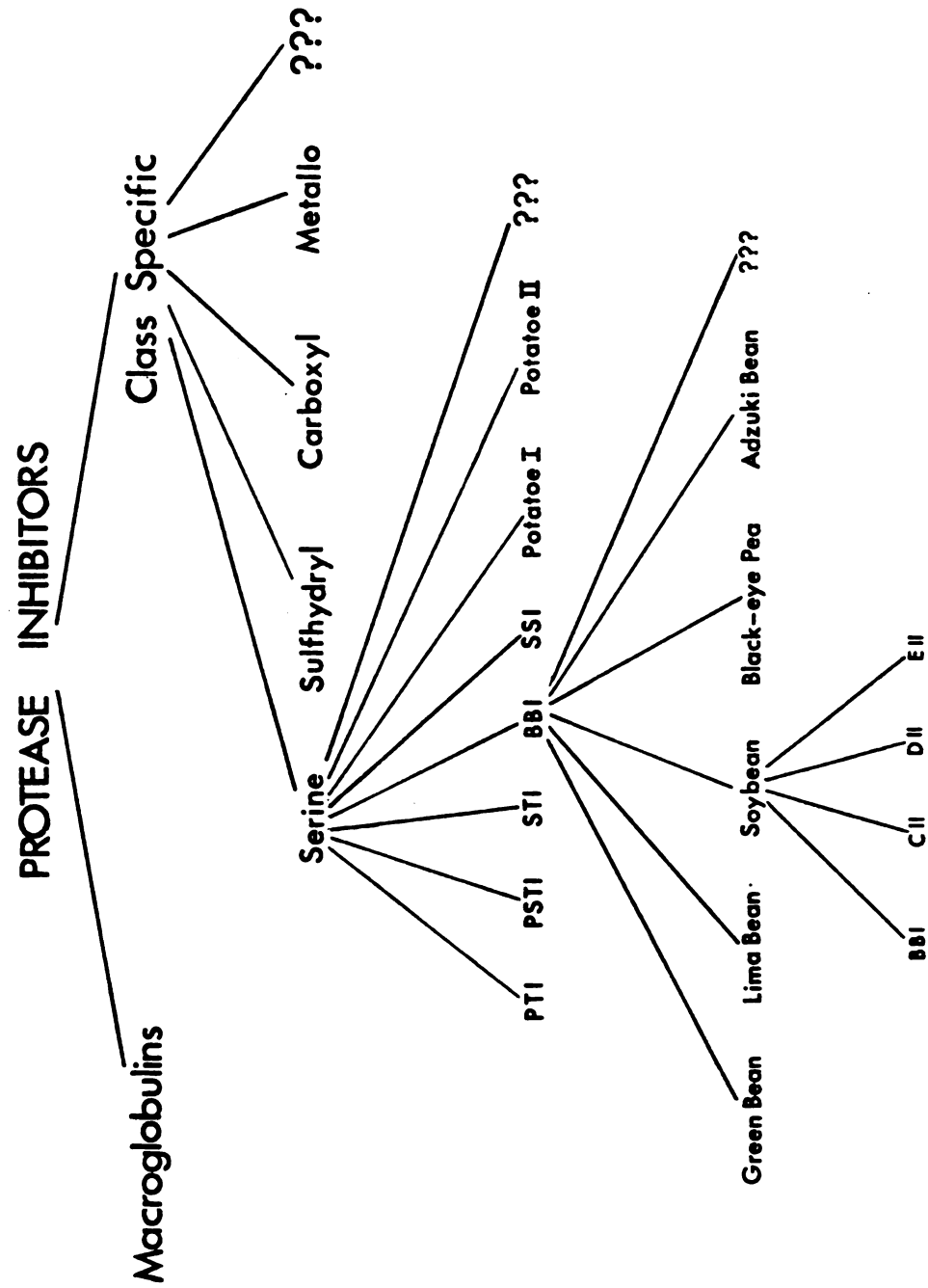


Figure 6: Protein Protease Inhibitor Relationships (Ref. 17).

here, because macroglobulins are not absolute inhibitors of proteases. Small substrates can still be cleaved by the "inhibited" enzyme, but large substrates, such as intact proteins, cannot. This suggests a steric inhibition. In addition, the inhibition is irreversible. As the name macroglobulins implies, the inhibitors are large molecules, with molecular weights of several hundred thousand. One mechanism proposed for their action is a trap mechanism. According to this model, the protease cleaves any one of several sensitive bonds on the macroglobulin. This causes the inhibitor to close around the enzyme, trapping it inside. The complex apparently leaves the enzyme functional, but sufficiently blocked to prevent its cleaving intact proteins. The prototype for this type of inhibitor is α -2-macroglobulin, found in blood (Ref. 18).

The class specific protease inhibitors differ fundamentally from the macroglobulins. As their name implies, they inhibit proteases from only one of the four mechanistic classes. In addition, their inhibition is absolute, and strictly competitive. Of these, only the serine protease inhibitors are well characterized.

Several families have emerged within the serine protease inhibitors. Members of the same family are closely homologous in their amino acid sequence, and probably also in their three-dimensional structure. They are most easily distinguished by comparing the pattern of their disulfide bridges (Figure 7). Many protein

Disulfide Linkages in Serine Protease Inhibitors

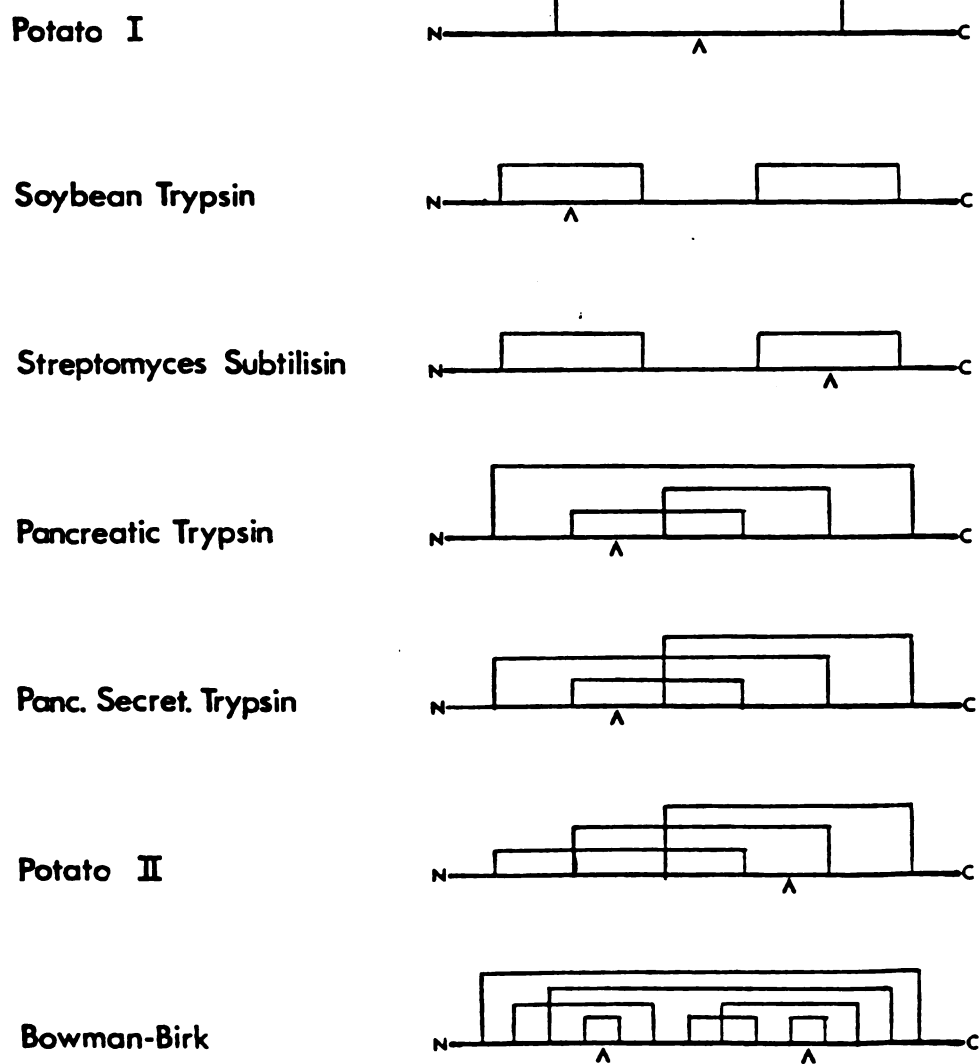


Figure 7: Disulfide Linkages in Serine Protease Inhibitors (λ indicates active site location) (Ref. 17).

protease inhibitors are "multi-headed", that is, they can inhibit more than one protease at the same time. Japanese quail ovoinhibitor, for example, has six "heads" (Ref. 19). Sequence studies suggest six domains, each of which is homologous to the pancreatic secretory trypsin inhibitor (PSTI). The Japanese quail ovoinhibitor is thus placed in the same family as PSTI. The only family elucidated at this point that has multiheadedness built into the disulfide structure is the Bowman-Birk inhibitor (BBI) family.

The three-dimensional structure of members of three of the inhibitor families, pancreatic trypsin inhibitor (PTI) (Ref. 20), soybean trypsin inhibitor (STI) (Ref. 21), and streptomyces subtilisin inhibitor (SSI) (Ref. 22), and their complexes with proteases (Refs. 15,23,24) have been determined by x-ray crystallography but there are no structural details yet about multi-headed inhibitors such as the object of these studies. However, progress has been reported on the crystallization of the BBI (Ref. 25), and the BBI-trypsin-chymotrypsin complex (Ref. 26).

PTI, STI, and SSI have very different three-dimensional structures, but the structure of the contact area between the inhibitor and the enzyme, which is very small, is essentially identical. All seem to have a classic "lock and key" type mechanism, with both the structure of the inhibitor and that of the enzyme remaining essentially unchanged on complex formation. Furthermore, circular

dichroism studies on the BBI complex with trypsin and chymotrypsin also suggest no changes on binding (Ref. 27). This appears to be a classic example of convergent evolution. Several different ancestral proteins varying in size and composition, converge on one active site structure, and one mechanism, to inhibit the same enzymes. Not only do trypsin and subtilisin converge to the same mechanism to cleave peptide bonds, but their inhibitors also converge to the same mechanism for inhibiting that reaction. SSI even forms complexes with and inhibits trypsin (Ref. 28).

C. Bowman-Birk Inhibitor and Its Family

Clearly the most complex of the inhibitors, in terms of the disulfide structure is the Bowman-Birk inhibitor family. Variants of this inhibitor are found in probably all legumes (plants such as peas and beans with pods containing seeds). The BBI itself is one of at least four found in soybeans.

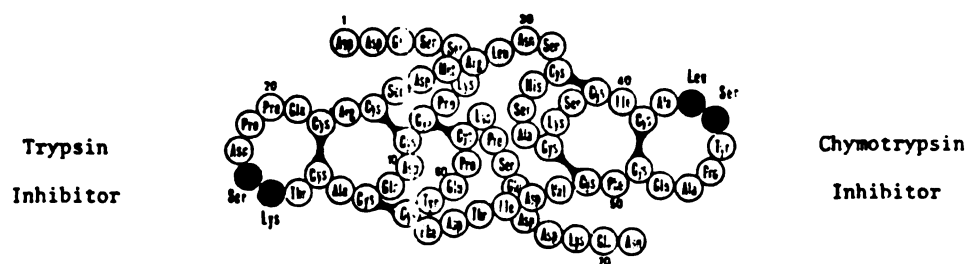
The Bowman-Birk inhibitor was first reported in 1946 by D.E. Bowman. He identified it as a second trypsin inhibitor (the first being STI) found in soybeans (Ref. 29), but did not further characterize it. This was accomplished about 15 years later by Y. Birk who found that it also inhibits chymotrypsin (Ref. 30). Birk, Gertler, and Khalef later showed that the inhibitor is

in fact divalent and can inhibit both enzymes simultaneously (Ref. 31).

The molecular weight and amino acid composition of BBI (Ref. 32,33) showed that it was a protein of low molecular weight (7975) which is crosslinked with seven disulfide bridges (Ref. 34). The specificity regions have been located in the sequence (Refs. 35-37) and occur in very similar nonapeptide loops. When these nonapeptide loops were synthesized, they were themselves inhibitors (Ref. 38).

The amino acid composition (Figure 8) is very atypical for a protein. Most outstanding is the fact that the molecule is 20% half-cystine. Clearly the disulfide bridges must be the dominant force holding the native conformation in place. This view is supported by the unusual stability of the protein towards heat and acid (Ref. 29). There is no evidence in circular dichroism studies suggesting either helices or pleated sheets. That such a structure would contain a large number of prolines (six) is not surprising, for a contorted structure is expected with so many disulfide bridges in a very small protein, but that it should contain no glycine whatsoever is most surprising. Clearly this represents some type of folding structure that has not yet been observed in proteins and hence another reason for this study.

Half of the amino acid sequence, including all of the disulfide bridges, is absolutely conserved in all of



Acidic

Aspartic acid (D)	11
Glutamic acid (E)	7

Basic

Arginine (R)	2
Histidine (H)	1
Lysine (K)	5

Polar

Half Cystine (C)	14
Serine (S)	9
Threonine (T)	2
Tyrosine (Y)	2

Hydrophobic

Alanine (A)	4
Isoleucine (I)	2
Leucine (L)	2
Methionine (M)	1
Phenylalanine (F)	2
Proline (P)	6
Valine (V)	1

71 Molecular wt. = 7975

Figure 8: Amino Acid Composition and Sequence of BBI (Ref. 33).

the members of the BBI family sequenced to date (Ref. 24; Figure 9). One quarter more of the sequence is essentially conserved, with only two different amino acids found in a given position. In addition, all have a highly acidic tail on the carboxyl end, though it varies in length. The amino end also has a tail which also varies in length and composition.

About 15% of the amino acids (eleven) are those with acidic side chains (aspartic acid and glutamic acid). Of these, four are located on the carboxyl tail and three are located on the amino tail. All four of the aspartate residues on the main part of the protein are absolutely conserved in all sequenced BBI family inhibitors, but interestingly, Glu 60 is replaced by positively charged lysine residues in about half of them. All of the inhibitors with a lysine at this position inhibit trypsin on the second domain, whereas some of those with glutamic acid inhibit trypsin, and others chymotrypsin on this domain.

About 11% (eight) of the amino acids on the BBI have basic side chains (lysine, arginine, and histidine). The only histidine in BBI (His 33) is absolutely conserved among all BBI type inhibitors. Other inhibitors in the BBI family contain as many as five additional histidines, usually associated with the amino or carboxyl tails. The lysines in this molecules are also conserved for the most part. One is located near the base of both the amino and

SEQUENCES OF BOWMAN-BIRK FAMILY INHIBITORS

BBI	DDESSKPCDDO	CACTKSNPPQCRCS	MDRLNSCHSACKSC	ICAI	SYP	QAQCF	VDIT	DFCYEP	CKPSE	DDKEN
	..	..	..	..	..	..	..	..	..	..
C-II	SDHSSDDDESSK	PCCDL	CMCTASMPPOCH	ADIRLNSCHSAC	DRCACTR	MPG	OCRC	LDTT	DFCYEP	CKSSDEDDDD
	..	..	..	..	..	..	..	..	..	..
D-II	SDQSSSYDDDEY	SKPCCD	LCMCTRSMPPQ	CSC	EDIRLNSCHSD	CKSMCTR	SOP	GOCRC	LDTN	DFCYKPKCKSRDD
	..	..	..	..	..	..	..	..	..	..
E-I	DEYSKPCCD	LCMCTRSMPP	QCS	EDIRLNSCHSD	CKSCMCTR	SOP	GOCRC	LDTN	DFCYKPKCKSRDD	
	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..
LBI-IV	SGHHEHSTDZPS	ZSSSKPCCB	HCACTKSIPPO	CRCTDLRLDSCHS	ACKSCICTLS	IP	QAQCV	CTBI	BTFCYEP	CKSSHSDDDNNN
	..	..	..	..	..	..	..	..	..	..
LBI-IV'	SGHHEHSTDZPS	ZSSSKPCCB	HCACTKSIPPO	CRCSDFRLDSCHS	ACKSCZCICTHS	AP	QAQCV	CTBI	BD	FCYEPCKSSHSDDDDD
	..	..	..	..	..	..	..	..	..	..
LBI-I	DZPSZSSSKPCC	BHCACTKSIPPO	CRCTDLRLDSCHS	ACKSCICTLS	IP	QAQCV	CTBI	BTFCYEP	CKSSHSDDDNNN	
	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..
GBI-II	BHBZHSSBZPS	ZSSPPCCB	ICVCTASIPPO	QIC	CTBVRLBSCHS	ACKSCMCTR	SM	PGKRC	CLBTT	BYCYKCKSBSGZBB
	..	..	..	..	..	..	..	..	..	..
GBI-II'	ZPSZSSPPCCB	ICVCTASIPPO	QIC	CTBVRLBSCHS	ACKSCMCTR	SM	PGKRC	CLBTT	BYCYKCKSBSGZBB	
	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..
ABI	EPSESSKPCDDQ	CCTKSMPPKRC	CSDIRLDSCHS	ACKSCACTY	SIP	AKCF	CTDIN	DFCYEP	CKSSR	DDDDWDN
	..	..	..	..	..	..	..	..	..	..

Figure 9: Sequences of Members of the BBI Family.

carboxyl tails (Lys 6 and Lys 63), and yet another is located near the end of the carboxy tail (Lys 69) in a position occupied by an acid group in most of the other family members. Most important, however, is Lys 16. This is the active site lysine to which trypsin binds (Ref. 35). Replacement of this lysine by alanine, as in green bean inhibitor II (GBI) or soybean inhibitor C-II results in an elastase inhibitor instead of a trypsin inhibitor. The BBI contains two arginines, Arg 28, which is conserved in all proteins in the family, and Arg 23. Arg 23 is conserved in all of the BBI family that inhibit chymotrypsin on the second domain, and glycine is substituted when the second domain inhibits trypsin.

Serine is an important amino acid in the BBI with seven of nine positions being conserved almost absolutely. The exceptions are Ser 65, located on the carboxy tail, and Ser 25 which is replaced by a variety of small side chains in different inhibitors. Of particular importance are Ser 17 and Ser 44. These are located on the carboxyl side of the active site scissile bond. Studies with model peptides indicate that peptides with a serine in this position are much better inhibitors than peptides with any amino acid in this position (Ref. 39). The best substitution is alanine, which suggests that size of the side chain is an important factor.

There are other polar amino acids that are conserved absolutely, which suggests their importance in either

structure or function. These are Thr 15, Tyr 59, and Gln 21. In addition, Asn 30 is conserved in all of the inhibitors in soybeans and lima beans, but found as aspartic acid in other beans.

Of the hydrophobic residues, only Leu 29 is conserved absolutely, but Ala 35, Ala 42, Ala 47, Ile 40, Ile 54, and Phe 57 have only very conservative substitutions. An important hydrophobic residue is Leu 43. This is the residue that binds to chymotrypsin. Chymotrypsin binding is retained if this is replaced by phenylalanine, as in the lima bean inhibitor LBI-IV', or by tyrosine, as in the adzuki bean inhibitor (ABI). If, however, it is replaced by an arginine, as in the soybean inhibitors C-II, D-II, and E-I, and the green bean inhibitors GBI-II and GBI-II', the inhibitor becomes a trypsin inhibitor at this head.

There is also absolute conservation of all fourteen of the cystine residues. If these disulfide bonds are cleaved, the molecule loses its activity (Ref. 40). These probably are essential to maintaining the three-dimensional structure of the molecule. Finally, four of the six prolines in the BBI molecule are conserved, and the two residues that are not conserved, Pro 61 and Pro 64, are located on the carboxy tail.

In solution the BBI exhibits quaternary structure. It self-associates into dimers and trimers (Ref. 41). However, there has been no evidence that the ternary

complex associates, and the fact that it crystallizes with only one complex per unit cell would also imply a lack of quaternary structure (Ref. 26).

From all of the foregoing, it can be seen that there are a number of reasons for interest in the tertiary structure of the trypsin-BBI-chyotrypsin complex. First, the structure of the Bowman-Birk inhibitor itself is of great interest. As a protein, it exhibits highly unusual primary structure, implying the same for its tertiary structure, which is probably based on an heretofore unknown folding scheme. As an inhibitor, the determination of its structure will elucidate an entire class of inhibitors, and possibly contribute to the overall picture of inhibitor evolution. As an inhibitor of serine proteases it will provide the first example of the two-fold inhibitor of enzymes, and will provide two more examples of enzyme-inhibitor complexes, perhaps clarifying more details of proteolytic cleavage. It will also provide a direct comparison of the action of trypsin and chymotrypsin on a protein substrate. The structure of the two domains of the Bowman-Birk inhibitor and their comparison may also answer questions about their evolution, and their three-dimensional structures may be compared to each other and to the other inhibited enzymes. One question to be answered is why, among all the members of the BBI family of inhibitors, chymotrypsin is only inhibited on the second head, and elastase only on the first head, when

trypsin can be inhibited on both. Questions about the role of amino acids away from the cleavage site in determining the specificity of the two domains may be answered, as well as the role of the amino and carboxyl tails. Finally, enzyme-inhibitor complexes are thought to be valid models of general protein-protein interactions, and the structure of the complex could shed light upon this important area.

II. EXPERIMENTAL METHODS

A. Forming the Complex and Growing the Crystals

The BBI was isolated and purified in Dr. Irvin E. Liener's laboratory, Department of Biochemistry, University of Minnesota. The inhibitor was isolated from unheated soybean meal and purified by successive chromatography on carboxy-methylcellulose and diethylaminoethylcellulose followed by preparative isoelectric focusing in the range of pH 3.5-5.5 using an LKB 2117 Multiphor apparatus. The resulting preparation gave a single band at MW 8000 when examined by sodium dodecylsulfate (SDS) gel electrophoresis.

The α -chymotrypsin and β -trypsin used to form the ternary complex were three times crystallized samples purchased from United States Biochemical Co.

Several different methods of crystallization were attempted. Initially, chymotrypsin conditions were tried (Ref. 42), since the ternary complex should resemble chymotrypsin dimer (the three-dimensional structure of trypsin is virtually identical to that of chymotrypsin, and the BBI should be very small with its low molecular weight and seven disulfide bridges). These conditions yielded only amorphous precipitates.

A technique which proved more successful was vapor diffusion using various polyethylene glycols (PEG) as the precipitant (Ref. 43). The crystallizations were carried out at pH 4.6 in 0.10 M sodium acetate buffer. Crystals were obtained with PEG 4000 (4000 is the average molecular weight) and PEG 6000, in concentrations varying from 8% to 50%, with protein concentrations varying from 1% to 2%. The best crystals were obtained with 10% PEG 4000 and 2% protein. This condition was optimized using the vapor diffusion method, but larger crystals grew when the batch method was employed. The solution also contained 0.02% sodium azide to inhibit bacterial and mold growth.

Care must be taken in the formation of the complex because two of the three proteins used are proteases, capable of autoproteolysis. Fortunately, the BBI, while strictly competitive, has very high binding constants for both enzymes (trypsin, $K_i = 2.8 \times 10^{-10}$; chymotrypsin, $K_i = 2.9 \times 10^{-8}$) (Ref. 44). Also, at low pH (below 6.0)

chymotrypsin dimerizes in such a way that it is no longer an effective protease (Ref. 45).

The buffer is made by combining 61.0 ml of 0.1 M acetic acid with 39.0 ml of 0.1 M sodium acetate. This makes 100 ml of a 0.1 M acetate solution with a pH of 4.4. In this is dissolved 10.0 grams of PEG 4000, which raises the pH to 4.6. If the pH is not quite 4.6, it is adjusted to that value using glacial acetic acid or concentrated sodium hydroxide, changing the volume as little as possible. To prevent biological contamination, 20 mg of sodium azide is added to the solution.

To form the complex, 1.7 mg of BBI is dissolved in 1.0 ml of buffer and precipitant solution. (If the protein is added to the buffer alone, and then the PEG's added, the protein precipitates immediately.) When it is fully dissolved, an equimolar amount of trypsin (5.1 mg) is added. Because of the high binding constant between BBI and trypsin, it is assumed that effectively all of the trypsin will be bound by the BBI. It has been reported that BBI comes to equilibrium with trypsin within 15 minutes (Ref. 44), so that much time was allowed to pass before chymotrypsin was added. An equimolar amount of chymotrypsin (5.4 mg) is added next, and since this is done at low pH, autoproteolysis of chymotrypsin, and proteolysis of the trypsin by chymotrypsin, and vice versa, is held to a minimum.

It was found that the protein-precipitant solution formed this way has many small pieces of dust and fiber in it which seem to hamper crystallization. To remove these, the solution is filtered through a metricell membrane with a pore size of 45 μm into a vial which has been cut to a height of 1.0 cm and cleaned very carefully. The final solution has a volume of 0.5 to 0.7 ml.

Crystals may appear in as short a time as 4 days, or as long a time as 6 weeks. The crystals grow in two different crystal habits. A prismatic morphology is usually observed, but triangular crystals also form (Figure 10). Crystals as large as 1.0 mm $\times$ 0.3 mm $\times$ 0.1 mm are common. The purity of the BBI seems to be critical in growing large crystals. Some batches of BBI yielded large crystals consistently, while others never produced crystals larger than 0.05 mm.

B. Characterization of the Crystals and the Native State

Once diffraction quality crystals were grown, it remained to characterize the crystals x-ray crystallographically and to establish native conditions. The crystals proved to be triclinic (space group P1) with a variety of cell constants depending on conditions as shown in Table 1. There exists at least three crystal forms, but they are indistinguishable from their morphology.

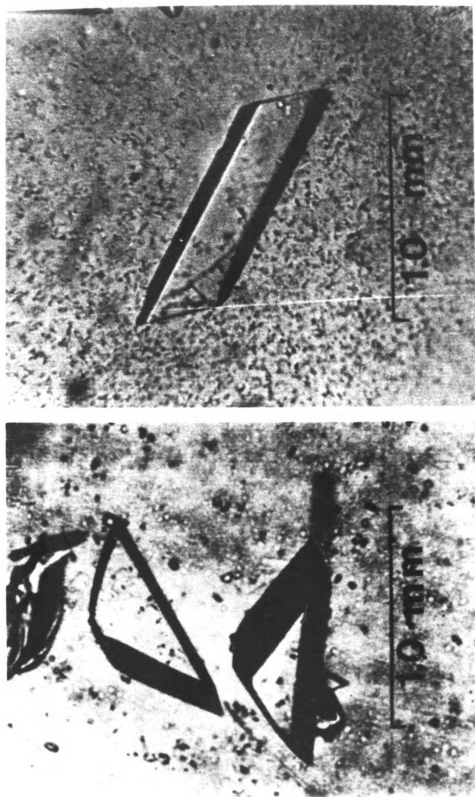


Figure 10. Two Morphologies of BBI Ternary Complex Crystals.

Table 1: Cell Constants of BBI Ternary Complex Crystals Under Various Conditions.

PEG	Buffer	pH	a in Å	b in Å	c in Å	α	β	γ	Volume
none	none	8.4	52.29	56.25	55.65	99.96°	105.72°	96.75°	154,800 Å ³
10%	0.01 M Tris	7.5	51.51	56.02	55.95	99.92°	105.06°	96.81°	153,100 Å ³
12%	0.10 M NaAc	4.6	51.97	56.16	55.53	99.52°	105.63°	96.78°	153,500 Å ³
15%	0.10 M NaAc	4.6	50.75	56.02	52.08	90.33°	112.14°	96.72°	136,100 Å ³
25%	0.10 M NaAc	4.6	50.77	56.21	52.18	90.45°	112.09°	96.77°	136,900 Å ³
25%	0.10 M NaAc	4.6	51.96	56.34	46.70	98.83°	97.29°	97.15°	134,500 Å ³
Gold Derivative									
25%	0.10 M NaAc	4.5	51.19	56.37	48.39	92.80°	114.08°	96.30°	126,700 Å ³
Platinum Derivative									
25%	0.10 M NaAc	4.5	51.11	57.17	48.76	93.16°	116.26°	93.27°	127,500 Å ³

Thus, care must be taken in choosing native conditions so that they are not at a pH, ionic strength, or precipitant concentration where small differences in any one of these variables will drastically alter the diffraction results.

Studies were first carried out to determine the effect of PEG concentration on the crystals. Crystals at pH 4.6 with 0.10 M sodium acetate were soaked in 12%, 15% and 25% PEG 4000. As can be seen in Figure 11, there were significant changes in the intensity distribution with PEG concentration. Table 1 reveals that while there are rather large changes in unit cell constants and volume between 12% and 15% PEG, those between 15% and 25% are minimal. The axial intensity distributions also differ much more between the 12% and 15% range than between 15% and 25% (Figure 11). It was decided on the basis of these experiments that the native state should have a PEG concentration either below 12% or above 15%, and that the PEG range between 12% and 15% should explicitly be avoided.

Limited studies were also carried out to determine the effect of pH on the crystals. As the pH increases, so do the binding constants of BBI for trypsin and chymotrypsin (Ref. 44). A pH of 7.5 was chosen to try to exploit the increased binding. A near neutral pH also avails the use of several heavy atom salts which are not soluble at the low pH of the acetate buffered crystals.

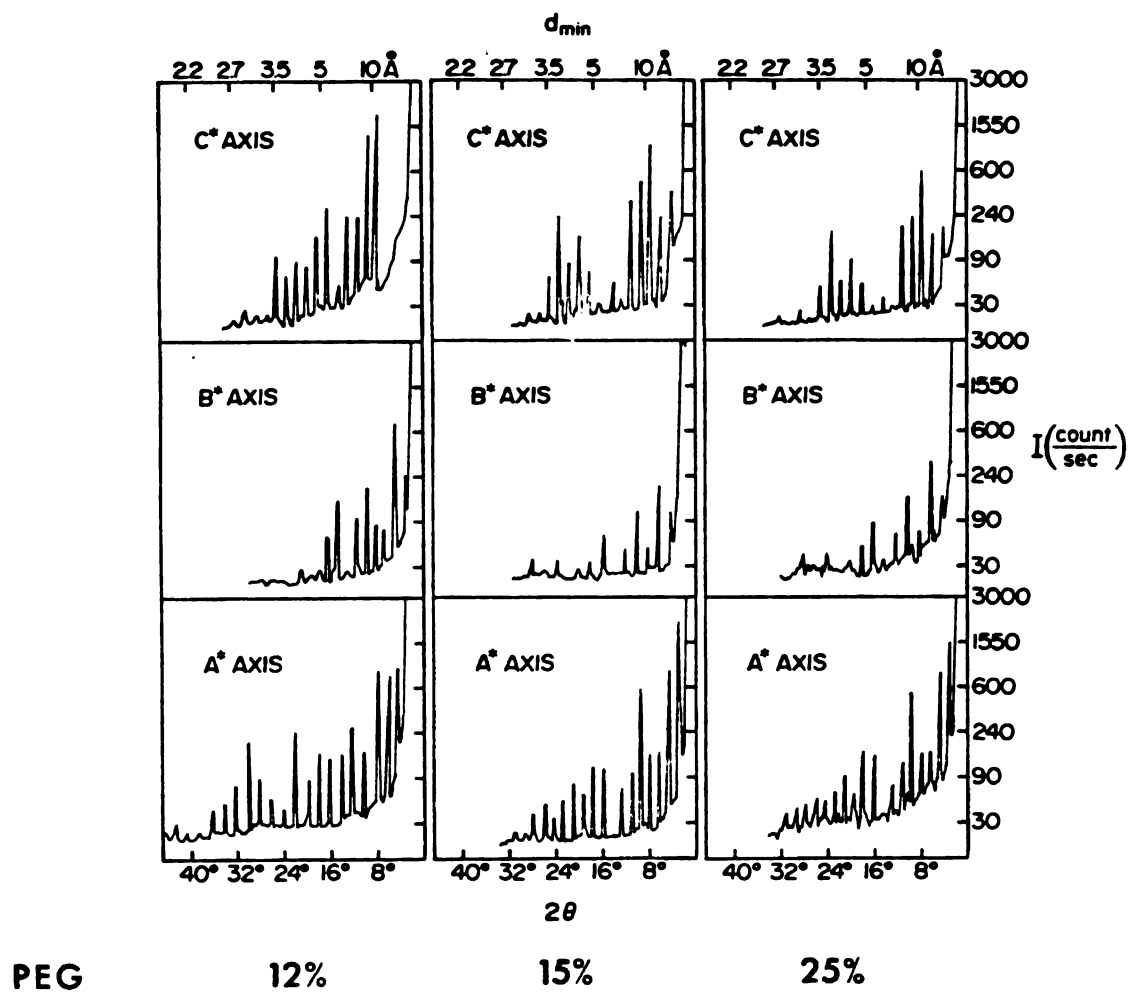


Figure 11: Axial Intensities of BBI Ternary Complex Crystals with PEG Concentration.

At pH 7.5, the buffer used was 0.010 M tris(hydroxymethyl)aminomethane (commonly known as tris). The PEG concentration at pH 7.5 could only be varied over a narrow range because at PEG concentrations higher than about 13% the crystals cracked and shattered, and at concentrations lower than about 8% the crystals dissolved.

It was found that buffer concentrations could only be lowered without the crystals dissolving, but then the precipitant concentration could be lowered in concert. Surprisingly, at the extreme, the crystals did not dissolve in distilled water, and the protein complex buffers the water at a pH of 8.4. Because there is not enough control over the pH in this case, and preparing heavy atom derivatives would necessarily increase the ionic strength and thus dissolve the crystals, it was decided not to use the distilled water crystals for intensity data collection, but their use greatly simplified the crystal density determination.

The 12% PEG sodium acetate crystals, the 10% PEG tris crystals, and the distilled water crystals offer a cross section of the effect of pH at low PEG concentration (Table 1; Figure 12). Even though there is a wide variation in buffer, ionic strength, and pH, the changes in cell constants are not as severe as varying the PEG concentration at constant buffer, ionic strength, and pH. However, the same is not true of the axial intensity distributions which show significant changes (Figure 12)

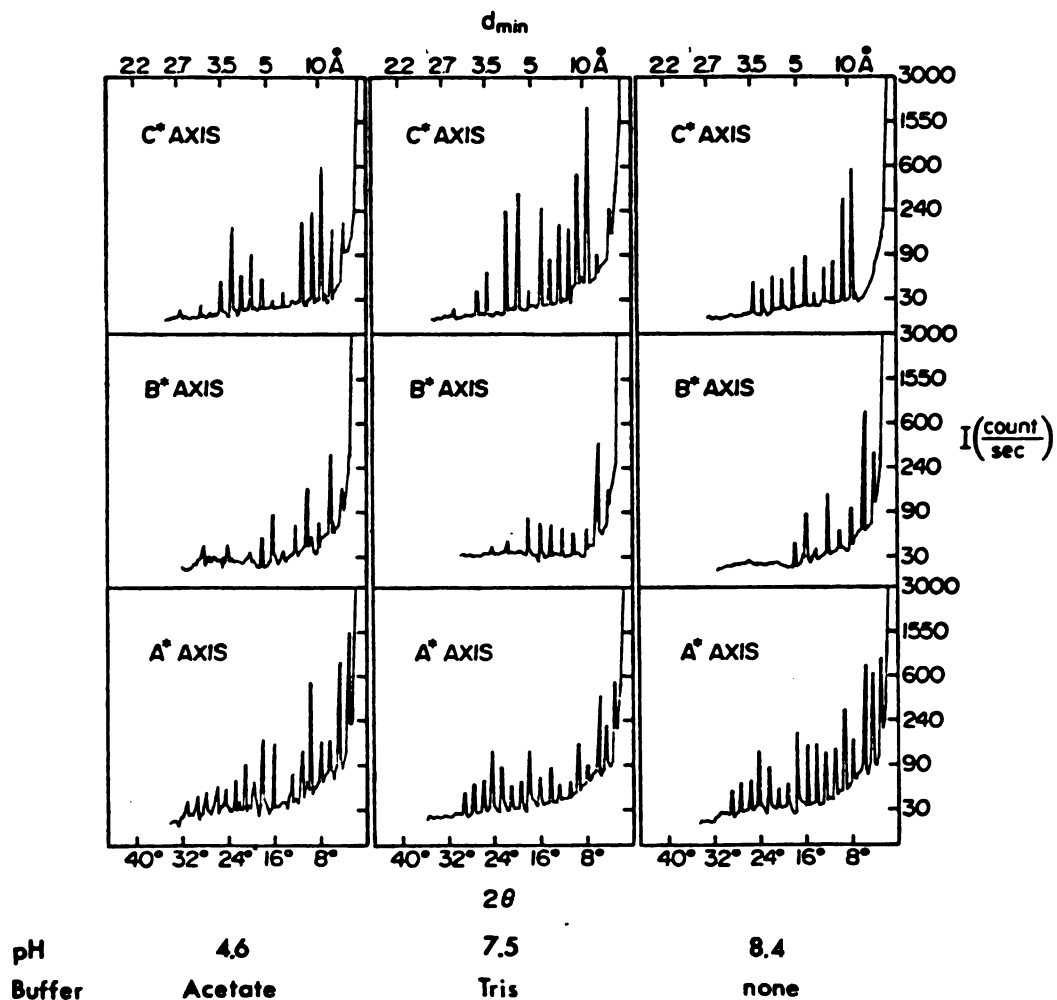


Figure 12: Axial Intensities of BBI Ternary Complex Crystals with pH.

that probably correspond to different pH conformers of the ternary complex.

Examination of the cell constants under these varying conditions clearly reveals two crystal forms, with the transition occurring somewhere between 12% and 15% PEG. There is, however, yet a third crystal form which coexists with the closer packed form at higher PEG concentrations. Specimens of both forms may exist in the same vial, so it seems unlikely that the third form is the result of a simple phase transition. As can be noted from Table 1 and Figure 13, this third form is contracted along the c axis, slightly decreased in volume, and is a much better scatterer of x-rays. While the intensity pattern of the second form extends to about 2.8 Å resolution, that of the third form extends to about 2.2 Å for comparably sized crystals (about 1.0 mm). Thus, it has been referred to as the high scattering form. Since it scatters x-rays considerably better than the other forms, it is this high scattering form that is preferred for study. Unfortunately, crystals of the high scattering form are a relatively low proportion of the larger crystals grown. Considering all of the foregoing, the native state of the BBI ternary complex has been taken to be pH 4.6, with 0.10 M sodium acetate and 25% PEG 4000 (preferably, the high scattering form).

The crystals were changed from growing conditions (10% PEG) to native storage conditions (25% PEG) by

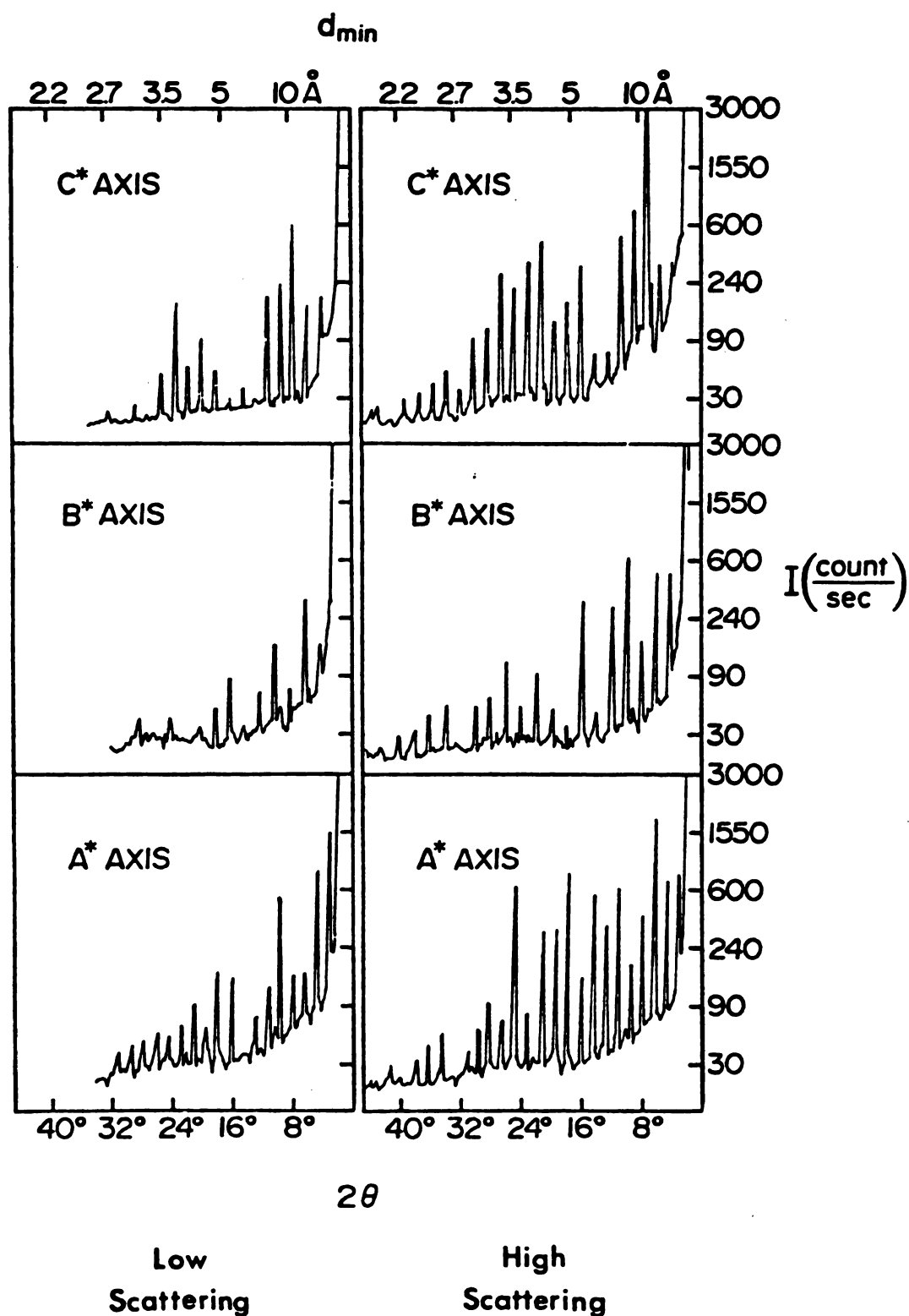


Figure 13: Axial Intensities of Two Forms of BBI Ternary Complex.

placing the open cut off vial containing the crystals and solution into a large (5 to 10 ml) sealed reservoir of the storage solution, and letting the solutions come to equilibrium by vapor diffusion. This process is completed in 5-7 days. The crystals can then be transferred to native storage solution conditions where they are stable indefinitely.

C. Verification of the Ternary Complex

Once crystals were grown it remained to be shown that the crystals were indeed the ternary complex. The large size of the unit cell is consistent with a host of possible combinations of the proteins in solution (ternary complex, chymotrypsin dimer, trypsin-chymotrypsin complex, etc.)

Evidence that the crystals are made up of the ternary complex was provided by SDS gel electrophoresis done in Dr. Leiner's lab. The crystals were first washed four times in acetate-PEG solution to remove all traces of free protein from solution. The crystals were then dissolved and treated with SDS and mercaptoethanol. The mercaptoethanol reduces the disulfide bonds, and the SDS is a detergent that denatures the proteins so that they migrate through the electric field according to their molecular weight. At the same time, a tube containing chymotrypsin, another containing trypsin, and another containing the BBI was subjected to the same treatment.

The results show (Figure 14) that the electrophoresis pattern of the crystals is the sum of the pattern of each of the three protein components.

Amino acid analysis was also performed in Dr. Leiner's lab with washed crystals and gave results consistent with the crystals containing equimolar amounts of trypsin, chymotrypsin, and the BBI.

Evidence that the crystals are made up of the complex was also obtained by crystallographic means. Since the crystals are stable in distilled water for relatively long periods of time, the salt-free density of the crystals was obtained by measuring the density of crystals soaked in distilled water. Using a calibrated density gradient made of xylene and iodobenzene, the density of these crystals was measured to be 1.17 g/ml. If the partial specific gravity of the crystals as given by v_p , the molecular weight of the complex M , is given by

$$M = \frac{NV(d - 1.0)}{n(1 - v_p)}$$

where N is Avogadro's number, V is the unit cell volume, d is the salt-free density, and n is the number of complexes per unit cell. Assuming that the partial specific gravity for the complex is that of chymotrypsin (0.736 ml/g), and using the cell constants for the crystal in distilled water (Table 1), the molecular weight calculated for the unit cell is 59,000. This is within experimental error

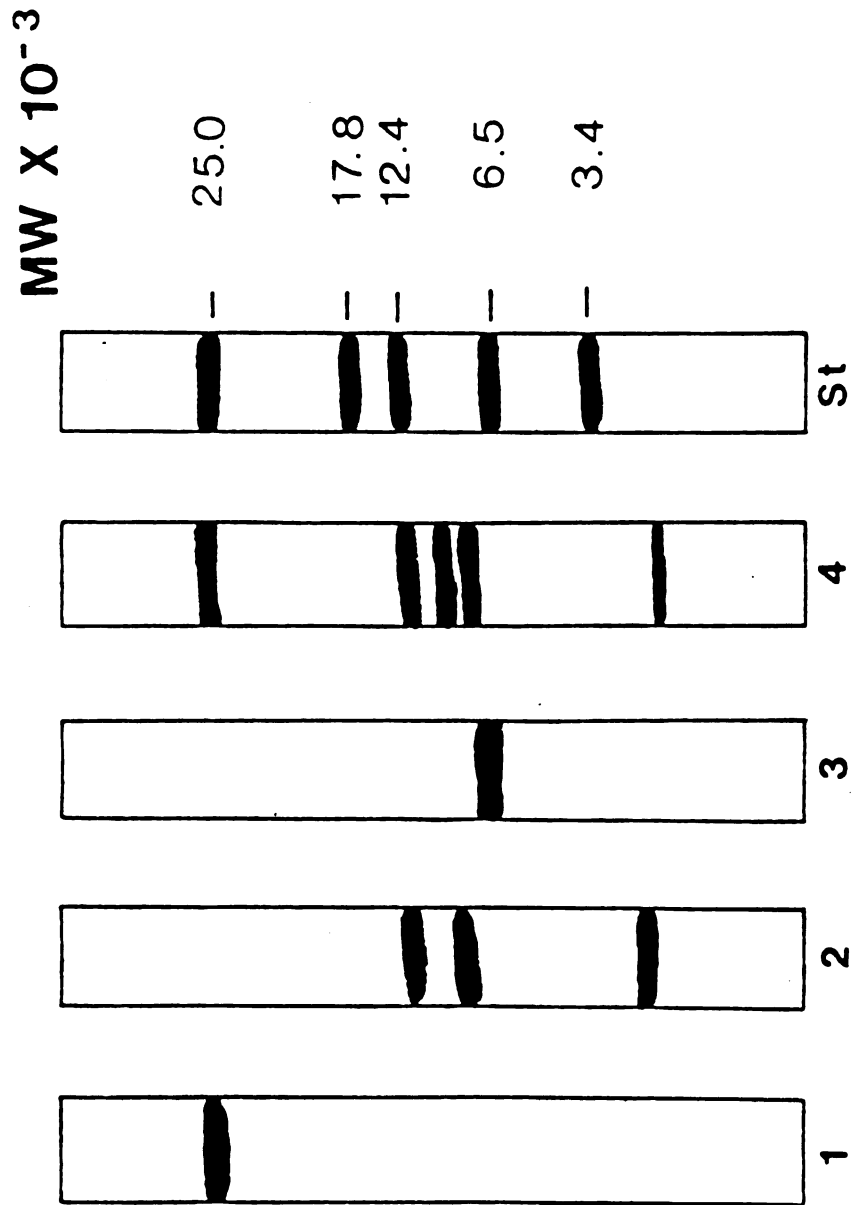


Figure 14: SDS Gel Electrophoresis of BBI Ternary Complex and Its Components. Lane 1, bovine β -trypsin; Lane 2, bovine α -chymotrypsin (from top to bottom, chains B, C, A, respectively); Lane 3, BBI; Lane 4, ternary complex.

of the weight calculated from sequence data of 57,550 per complex.

The density of the native high scattering crystals was found to be 1.26 g/ml, which implies crystals that are 58% protein and 42% mother liquor by weight. This is well within the normal range for protein crystals (Ref. 46). The higher density implies that the molecules are packed closer together in the native state than in distilled water and so is consistent with the fact that the native crystals scatter x-rays better.

D. Heavy Atom Search

A heavy atom search was started to determine a derivative suitable for the multiple isomorphous replacement method of phase determination. Since the protein crystals contain about 40% mother liquor, the salts were diffused through the crystals in the manner described in the introductory section of this report.

Two derivatives showed promise at the diffusion stage in that they did not shatter the crystals. The first reagent was potassium platinum (II) iodide (K_2PtI_4). The platinum compound was used to saturate a solution of 0.10 M sodium acetate, 25% PEG at pH 4.6. The pattern of the ternary complex crystals soaked in this solution showed considerable changes from the native pattern, but unfortunately, there were also rather large changes in the

cell constants (Table 1). These differences were considered too great for the derivative to be useful. The second reagent, potassium gold (III) iodide (KAuI_4) was soaked using the same conditions as the platinum. The results were similar, with large cell constant changes observed (Table 1).

At this point work had to be terminated due to a lack of usable crystals and a vanishing supply of BBI. As mentioned previously, the quality of the crystals is very dependent on the purity of the BBI. Although new batches of the BBI were received from Dr. Leiner, they apparently did not have the purity of the earlier samples. Although crystals were grown from these batches, none was large enough to be useful. Finally, since Dr. Liener's research efforts have taken new directions since we first received the original samples of BBI from him (eight years ago), when his stock of BBI became depleted, he ceased to be a source of material for us.

X-RAY CRYSTALLOGRAPHIC STUDIES OF RIBONUCLEASE T₁

I. INTRODUCTION

Nucleases are enzymes that cleave DNA and/or RNA into their constituent nucleic acids. They are analogous in their action on polynucleic acids to proteases in their action on proteins. One class of the nucleases is the ribonucleases, which act principally or exclusively on RNA. The ribonucleases are further subdivided into three major types, according to their mode of action. Phosphorylases catalyze the cleavage of RNA with the addition of phosphate. Phosphodiesterases catalyze a direct attack of water on the 3':5' phosphodiester bond, and require calcium or magnesium ion cofactors. Phosphotransferases use the 2'-hydroxyl group in an intramolecular attack at the adjacent phosphodiester bond, forming a nucleoside 2':3' cyclic phosphate intermediate (Ref. 47). The object of this study, ribonuclease T₁, is an endonuclease phosphotransferase, specific in cleaving RNA on the 3' side of guanosine residues.

Ribonuclease T₁ was first purified by Sato and Egami from the bacterium *Aspergillus Oryzae* found in

takadiastase (Ref. 48). It is the major component of two ribonucleases found in takadiastase, the other being referred to as ribonuclease T_2 . Because T_1 is very specific in its cleavage of RNA, it has found wide application in the study of the primary structure of nucleic acids.

Ribonuclease T_1 is a relatively small (11,127 dalton), thermostable protein. The amino acid composition of T_1 was determined in 1962 by Takahashi (Ref. 49; Table 2), who later determined the amino acid sequence (Ref. 50; Figure 15). The amino acid composition reveals a protein containing only a single lysine, a single arginine, and three histidines as basic groups, but six glutamic acids and five aspartic acids. The result is that T_1 is a very acidic protein, with an isoelectric point of about pH 3.5 (Ref. 51). It implies that at physiological pH it is strongly negatively charged, but yet it cleaves RNA, also strongly negatively charged in this pH range. T_1 contains no free sulfhydryls, and no methionines, but two disulfides. It is of note that the two disulfide bridges are located unusually near to each other (Figure 15). One makes a small nonapeptide loop at the amino end of the protein and the second connects the carboxyl end to that loop.

Chemical studies have revealed the importance and role of several of the amino acids in T_1 . The first residue shown to be essential for activity was Glu 58.

Table 2: The Amino Acid Composition of T₁.

Acidic	
Aspartic Acid (D)	15
Glutamic Acid (E)	10
Basic	
Arginine (R)	1
Histidine (H)	3
Lysine (K)	1
Polar	
Half Cystine (C)	4
Serine (S)	17
Threonine (T)	6
Tyrosine (Y)	8
Hydrophobic	
Alanine (A)	7
Isoleucine (I)	2
Methionine (M)	0
Phenylalanine (F)	4
Proline (P)	4
Tryptophan (W)	1
	104

Molecular Weight = 11,127

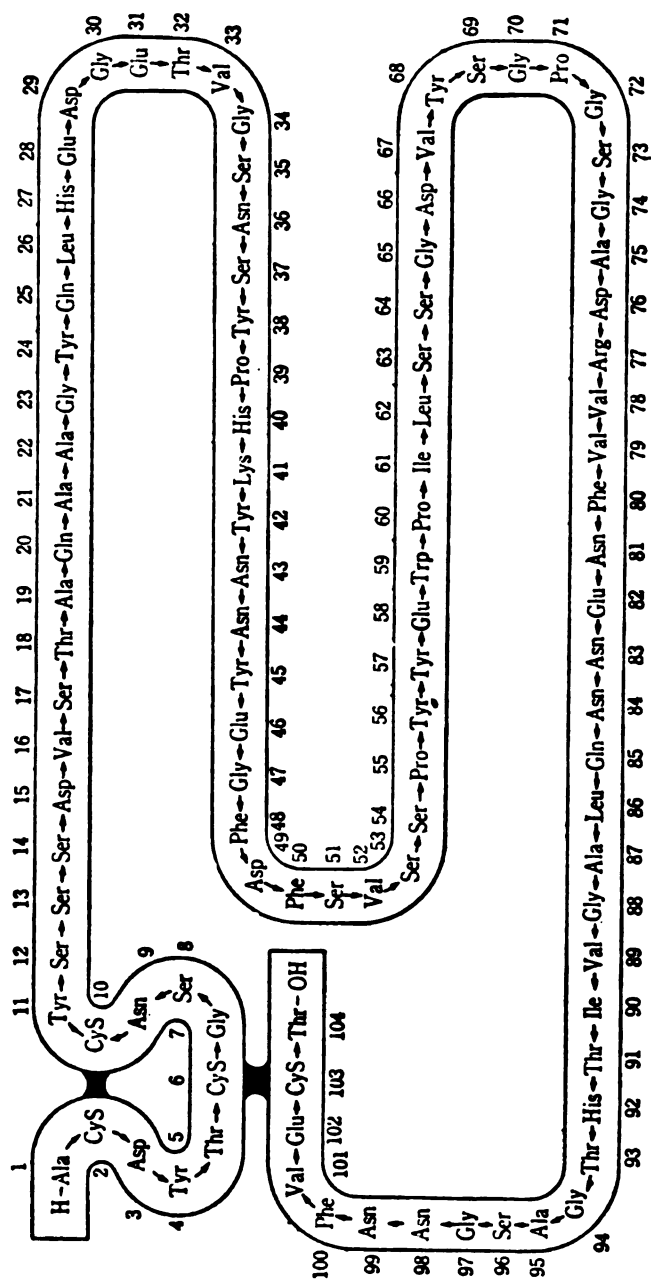


Figure 15: Amino Acid Sequence of Ribonuclease T₁ (Ref. 47).

When T_1 is treated with iodoacetate at a pH of 7.5 to 8.5, the enzyme is completely deactivated. Apparently two of the three histidines are essential (His 27 and His 92), with His 40 shown to have no catalytic role. Arg 77 has been shown to be essential for activity also, but the other two positive charges on the molecule (Lys 41 and the amino terminus) are not (Ref. 47). Finally, T_1 is rich in aromatic residues with eight tyrosines and a tryptophan, but after extensive study no catalytic role has been assigned to either of them.

The basic mechanism of T_1 is in many respects though to be analogous to that of ribonuclease A (Figure 16), except that a carboxyl group acts as the general base instead of an imidazole (Ref. 47). The 2'-hydroxyl of the RNA ribose is situated near a glutamate (Glu 58) and the phosphodiester oxygen to be cleaved is near a histidine (probably His 92). The glutamate abstracts the proton from the hydroxyl, and the nucleophilic oxygen attacks the phosphorus forming a five-membered ring with a pentavalent phosphorus. At the same time, the histidine gives up its proton to the 5'-ribose oxygen, forming the leaving group. The 5' fragment now leaves the active site and is replaced by a water molecule. The water molecule undergoes the reverse reaction that cleaved the 5' fragment, regenerating the active site residues and the 2'-hydroxyl (Figure 16).

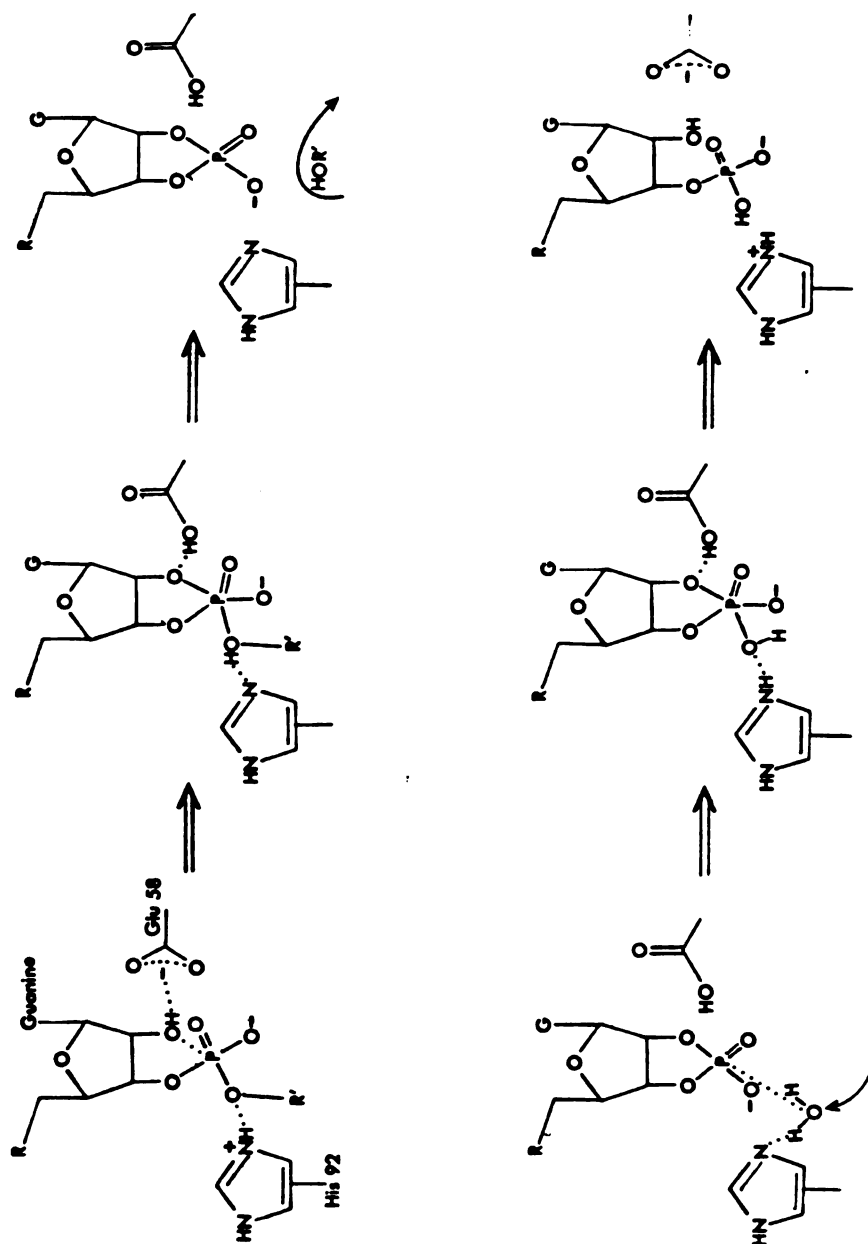


Figure 16: Proposed Mechanism for T_1 Catalysis (Ref. 47).

T_1 exhibits a number of interesting, and unique, features that should have broad significance for understanding catalytic mechanisms and the general nature of protein-nucleic acid interactions. For instance, subsites for nucleosides exist on the enzyme beyond the active site (Ref. 52). Although the enzyme is very specific in cleaving RNA at guanosine residues, it binds significantly, but does not cleave, adenosine substrates (Ref. 53). Specific subsite interactions also influence the catalytic mechanism significantly (Ref. 54). In comparison to ribonuclease A, the affinity of T_1 for bases and nucleosides is greater by a factor of ten (Refs. 55,56), and temperature jump studies indicate on-off rate constants an order of magnitude lower (Ref. 57). Whereas T_1 has specific recognition for the 2'-hydroxyl group, ribonuclease A does not (Ref. 58). The general base group in the catalytic site is a carboxyl group in T_1 (Glu 58) instead of an imidazole (His 12) as in ribonuclease A. The 5'-phosphate group interacts with a subsite in the latter (Ref. 59), and not in T_1 (Ref. 60). Finally, 6-thioguanosine containing substrates are only catalyzed in a limited way by T_1 (Ref. 61), whereas the corresponding 2- and 4-thiouridine containing substrates are readily cleaved by ribonuclease A (Ref. 62). A detailed structural analysis of T_1 should prove helpful in understanding all these differences.

There have been five other ribonucleases sequenced that appear to be significantly homologous to T_1 . They include ribonuclease U_2 , Ms, St, Bi, and barnase, and their aligned sequences are shown in Figure 17. Of these, T_1 has been by far the most extensively studied. The x-ray crystal structures of the 2'-guanosine monophosphate derivative of T_1 (Ref. 63), and of the native structures of St (Ref. 64) and barnase (Ref. 65) have also been reported. The tertiary structures of these three ribonucleases are grossly similar. Each has a long α -helix covered by four strands of anti-parallel β -pleated sheet, and their secondary structures belong to the $\alpha + \beta$ type of packing of Levitt and Chothia (Ref. 66).

From the foregoing, there are a number of reasons for interest in the tertiary structure of the native T_1 molecule. The structure of native T_1 can be compared to the 2'-guanosine monophosphate derivative which has been reported in order to study changes that occur on substrate binding. In addition, an 8-iodo-guanosine (8IG) derivative map can be studied, since it does not contain a bulky group at the catalytically critical guanosine 2' position. The 2'-guanosine monophosphate derivative was solved from a crystal form containing only one molecule per asymmetric unit, while the crystals described herein contain two, affording an opportunity to study protein-protein interactions in the crystal.

HOMOLOGS OF RIBONUCLEASE T₁

T ₁ --	ACDYTCGNCYSS	SDVST	AGAAQYGLHEDQETVGSNSYPHMYNNYEGDFSVSS	PVY	EMPILSSGDVYSG	POSGADRVVFNENW	QLAGVITHYGA	SCNMFVECT
St --	QAPCGDTSGFEVRLADLPPEATDTYELIEKQOPYPPEDOTVEENREGILPDCAE		OYYHEYTV	KTPSG	DDRCARRFVVQDGO	EYFYTEDHYE	SFRLTIWN	
B ₁ --	AVINTFDGVADYLIRYKRLPNDYITKSGASALOWASKODLAEVA	PQ	KSICGDVESNRECLPFAQS	RTKREADI	NYVSG	F	RNADRLVYSSDW	LIYKTTDHYA
B ₂ --	AGVINTFDGVADYLGTYHKLPNDYITKSEAGALCHWASKONLADVA	PQ	LSIGODIESNRECKLPQKSO	RTWREADI	NYTSG	F	RNSDRILYSSDW	LIYKTTDHYQ
U ₂ --	CNIPESINCGCNVYSN	DDINT	AIGGALDDVARPD	GDNYPHQVYD	EASDQITLCCGQSQSWSEFFLVYNGPYVSSRRNVVSGPDRVIYGTNTGEFCATVTHTGAASYDCFTOCS			
M _s --	ESCEYTCGTCYNS	SDVSA	AKAKQVSLYESQETI	DDYPHEYVDYEGDFDPVOT	SVY	EYPIMSDYDVYTG	CSPGADRVIFDCDD	ELAGVITHTGAGGDDFVACSSS

T ₁ --	T ₁ (Aspergillus oryzae)
St --	Ribonuclease St (Streptomyces erythraeus)
B ₁ --	Ribonuclease B ₁ (Bacillus intermedius)
B ₂ --	Barnase (Bacillus amyloliquefaciens)
U ₂ --	Ribonuclease U ₂ (Ustilago sphaerogena)
M _s --	Ribonuclease M _s (Aspergillus saitoi)

Figure 17: Homologs of Ribonuclease T₁ (Ref. 64).

Most importantly, however, the crystals are excellent x-ray scatterers and a 1.5 Å resolution structure should be obtainable.

II. EXPERIMENTAL

A. Growing Crystals

Ribonuclease T_1 was purified according to the method of Waltz and Hooverman (Ref. 67) in the laboratory of Dr. Frederick Waltz, Jr., Department of Chemistry, Kent State University. Crystals of ribonuclease T_1 are obtained using the method described by Martin, Tulinsky and Waltz (Ref. 68).

First, 4.0 mg of ribonuclease T_1 are dissolved in a 50 μ l drop of 0.25 M potassium phosphate solution adjusted to pH 7.3. To this is added 50 μ l of saturated ammonium sulfate that has been adjusted to pH 7.3. Then, more of the saturated ammonium sulfate solution is added drop by drop until a light precipitate forms and does not disappear on stirring. It should be noted that the pH of the growing solution resulting from the mixture of the pH 7.3 potassium phosphate and the pH 7.3 saturated ammonium sulfate is not 7.3 as reported (Ref. 68) but 6.6. The final concentration of ammonium sulfate is between 50% and 55% saturated, the buffer is about

0.125 M potassium phosphate, and the protein is about 4% by weight.

The solution is then seeded by seed crystals usually about 0.1 mm in the longest direction. The seed crystals are first soaked for about ten minutes in a solution of 50% saturated ammonium sulfate, 0.125 M potassium phosphate at pH 6.6 (Ref. 69). This is similar to the growing solution, except that there is no protein in it. The crystals are not stable in this solution, but slowly begin to dissolve (they completely dissolve in about 30 minutes). The procedure dissolves the outer layers from the crystal, providing a fresh growing surface. If the seeds are not first soaked in this solution, many small crystals usually result instead of a few large ones. In addition to the seeds increasing their size, several more crystals grow, often including two to five which grow as large as the seeds. The small crystals are used as seeds for subsequent crystallizations. The crystals grow to their maximum size in two to ten days. Crystals as large as 3.0 mm have been grown this way, but 1.0 mm is more usual.

Once the crystals attain a desired size, the conditions are changed to more stable ones. The hanging drops are placed over a well containing 0.05 M sodium acetate in saturated ammonium sulfate at pH 5.5. Since the ammonia and the acetic acid in the solutions are both volatile, the pH of the crystallization solution changes

slowly to 5.5, and water diffuses from the drop until the drop is also saturated in ammonium sulfate. Once the crystals have reached equilibrium (five to ten days) they may be transferred to the storage solution of 0.05 M sodium acetate and saturated ammonium sulfate at pH 5.5. This essentially removes the phosphate from solution. The crystals are stable at these conditions for indefinite periods of time, but care must be taken when handling them in their storage solution. Since the solution contains saturated ammonium sulfate, even brief exposure to the air will cause a precipitation of ammonium sulfate.

The saturated ammonium sulfate solution also causes special problems when mounting the crystal. If crystalline ammonium sulfate forms, its diffraction pattern may interfere with that of the protein. Glass x-ray diffraction capillary tubes were cut and then sealed so that the constant diameter part of the tube is 30 to 35 mm long (Figure 18). Then, leaving a space of approximately 10 mm from the sealed end of the capillary, all but about 5 mm of the rest of the capillary tube is filled with storage solution. A small gap of air is left, and then the rest of the tube is filled with storage solution (Figure 18). Ammonium sulfate precipitates out in the solution which is exposed to the air, but not in the main body of the solution. A T_1 crystal is then quickly transferred to the protected solution. The capillary is then tipped so that the

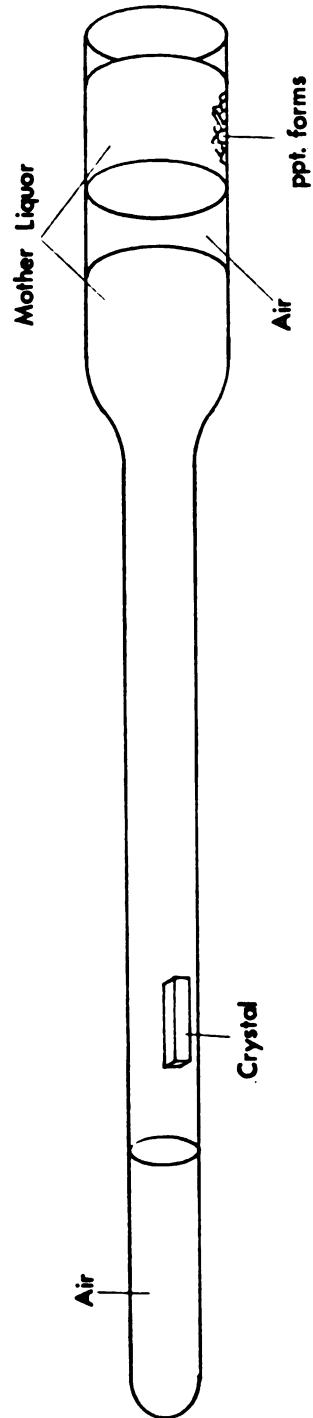


Figure 18: Mounting a Crystal in Saturated Ammonium Sulfate. This technique prevents the precipitation of ammonium sulfate near the protein crystal.

crystal falls to the bottom of the solution. From there it is poked through the surface into the air pocket using a hair or broom bristle and maneuvered into the desired orientation, which is with the long axis parallel to the capillary tube, and the crystal center about 5 mm from the sealed end of the tube. The solution is withdrawn from the capillary and excess mother liquor is quickly blotted from around the crystal with a strip of filter paper. A cap of mother liquor is placed about 10 mm above the crystal, and then another cap placed above it so that it does not dry out. The orientation of the crystal is then checked. If it is satisfactory, the capillary is cut between the two mother liquor caps and sealed with wax. The wax is further sealed with a thin film of epoxy glue after the capillary is mounted on the goniometer head.

B. Preliminary Diffraction Studies

Analysis of the crystals by precession and diffractometer methods reveals three mutually perpendicular axes of unequal length. The existence of mm symmetry in the reflection intensities of each of the $(hk0)$, $(h0l)$, and $(0kl)$ zones established the crystal system as orthorhombic. Each of the principal axes has systematic absences at the odd order reflections, further indicating three mutually perpendicular two-fold screw axes. The space group was

thus assigned $P2_12_12_1$, with cell constants $a = 91.73 \pm 0.03 \text{ \AA}$, $b = 37.320 \pm 0.006 \text{ \AA}$, and $c = 77.78 \pm 0.03 \text{ \AA}$.

Dr. P.D. Martin, of this laboratory, determined the salt free density of the crystals to be 1.23 g/ml. The volume of the unit cell is 266,300 $\text{\AA}^3$, so the unit cell mass is about 195,000 dalton, and the asymmetric unit mass is 48,700 dalton. Since the molecule has a mass of about 11,000 dalton, possibilities for the number of molecules per asymmetric unit are those summarized in Table 3. These suggested a slight ambiguity in the number

Table 3: Number of Molecules per Asymmetric Unit with Corresponding Percent Mother Liquor

<u>Number of Molecules</u>	<u>Percent Mother Liquor</u>
1	77.4%
2	54.8%
3	32.2%
4	9.7%

of molecules per asymmetric unit. Either the crystals contain two molecules per asymmetric unit and are 55% mother liquor, or they contain three molecules per asymmetric unit and are 32% mother liquor. Both compositions are well within the extremes observed for protein crystals, and the ambiguity was not fully resolved until later in the structure analysis.

Large crystals (1.2 mm) diffract x-rays to at least 1.5 Å resolution ($2\theta = 60^\circ$) and the crystals do not decay unduly fast in the x-ray beam. The distribution of intensities along two of the principal axes is shown in Figure 19.

C. Preparation of Heavy Atom Derivatives

Since the crystals contain an appreciable percentage of mother liquor, attempts to prepare heavy atom derivatives for use in the determination of phases followed conventional procedures. The heavy atom salt is dissolved in storage solution containing protein crystals. This solution is kept at the proper pH by vapor diffusion against a large well of storage solution throughout the time of the soaking period (2 to 10 weeks), much the same way that the pH is changed in going from growing to storage conditions.

An extensive search for heavy atom derivatives at pH 7.3 was carried out by Dr. P.D. Martin of this laboratory with only limited success (one derivative formation, mersalyl (Figure 20) of Table 4). This mersalyl derivative was interpreted to have a long substitution near a special position of $y = 1/4$. As a result, it was not particularly suitable for phasing since the mercury atom contributes only a real or imaginary term to the structure factors, and is never complex. Thus, it could not even

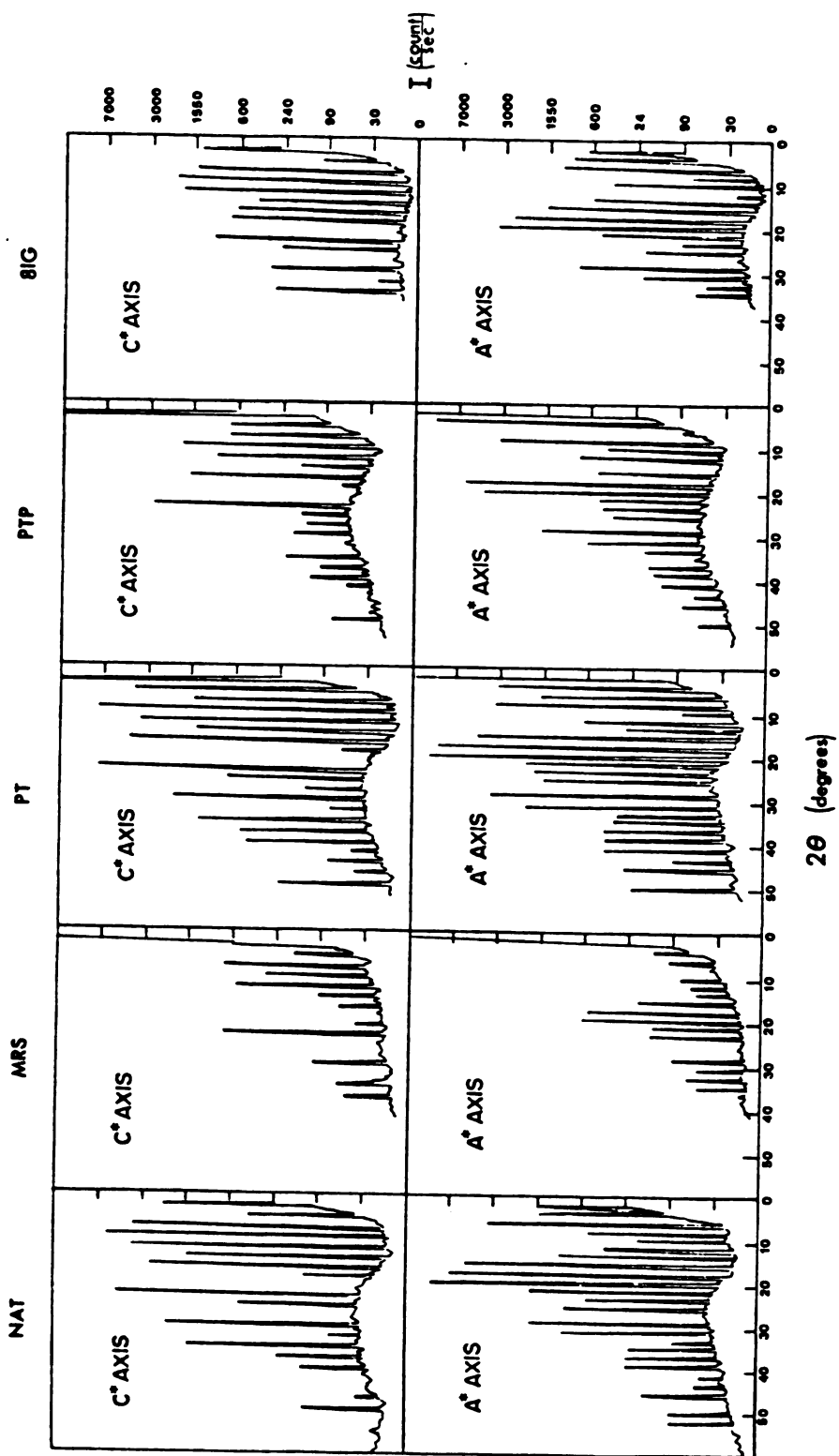
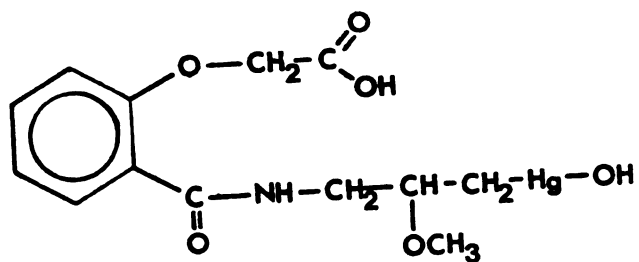
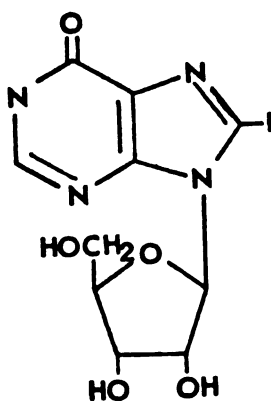


Figure 19: Axial Intensities of T_1 and Heavy Atom Derivatives.



Mersalyl Acid



8-Iodo-Guanosine

Figure 20: Molecular Structure of Some Heavy Atom Reagents.

Table 4: T₁ Heavy Atoms.

Heavy Atom Reagent	pH	Results
Thallium Sulfate	7.3	no intensity changes
Phenylmercury Acetate	7.3	no intensity changes
p-Chloromercuriphenyl Sulfonic Acid	7.3	no intensity changes
5-Mercurycytosine-5'-diphosphate	7.3	no intensity changes
Potassium Platinum(IV) Hexaiodide	7.3	no intensity changes
Potassium Platinum(II) Tetracyanide	7.3	no intensity changes
Ethylmercurithiosalicylic Acid	7.3	no intensity changes
Fluorecene Mercury Acetate	7.3	no intensity changes
Mercury(II) Acetate	7.3	no intensity changes
Lead(II) Acetate	7.3	no intensity changes
2-Hydroxyethanethiolate(2,2',2"-terpyridine) Platinum(II) Nitrate	7.3	no intensity changes
Samarium(III) Chloride	7.3	no intensity changes
Mercury Succinimide	7.3	no intensity changes
Mersalyl Acid	7.3	2 bound mercury atoms
Silver Nitrate	7.3	uninterpretable Patterson map
8-iodo-guanosine	7.3	uninterpretable Patterson map
Thallium Silver Sulfate	7.3	pattern same as silver nitrate
Platinum Diamino Dicyanide	7.3	uninterpretable Patterson map
Guanosine Triphosphate-γ-thiol + Mercury(II) Acetate	7.3	uninterpretable Patterson map
*Mersalyl Acid	7.3	uninterpretable Patterson map
Uranyl Nitrate	5.5	2 bound mercury atoms
*8-Iodo-guanosine	5.5	uninterpretable Patterson map
Mercury(II) Acetate	5.5	promising intensity changes
Silver Nitrate	5.5	no intensity changes
Mersalyl Acid + Silver Nitrate	5.5	uninterpretable Patterson map
Lanthanum Oxide	5.5	same as mersalyl acid
Phenyllead Triacetate	5.5	no intensity changes

Table 4 Continues.

Table 4 Continued.

Heavy Atom Reagent	pH	Results
Triphenyllead Acetate	5.5	no intensity changes
2-Hydroxyethanethiolate(2,2',2"-terpyridine) Platinum(II) Nitrate	5.5	intensity pattern changed during data collection
Chloro(2,2',2"-terpyridine) Platinum(II) Chloride	5.5	peaks in double special position
Thallium Nitrate	5.5	no intensity changes
Sodium Platinum(II) Tetraiodide	5.5	uninterpretable Patterson map
Platinum Ammonium Nitrobromide	5.5	uninterpretable Patterson map
*Potassium Platinum(IV) Hexathio- cyanate	5.5	promising intensity changes
*Mersalyl Acid + Platinum(IV) Hexathiocyanate	5.5	promising intensity changes

*These derivatives were studied in depth.

be used in a single isomorphous replacement method application because the double phased Fourier synthesis would create the correct electron density superimposed upon its mirror image.

A review of the literature reveals that in the past heavy atom searches have been more successful at lower pH (Ref. 46). Since this is also the region of optimal catalytic activity of T_1 , it was decided to lower the pH of the crystals and search for heavy atom derivatives. Ribonuclease T_1 has its activity maximum at pH 5.5 (Ref. 70), so this pH was chosen for the native state in the present heavy atom survey.

The intensity patterns along the axes show the native diffraction pattern does not appear to be substantially different at the two pH values, so large differences between pH conformers are not expected. The activity of the enzyme is not disrupted throughout this pH range (Ref. 70). Thus, the heavy atom search at pH 5.5 began with those heavy atom salts which showed the most promise at pH 7.3. The heavy atoms examined are summarized in Table 4. Preliminary low resolution studies suggest that the mersalyl derivative (which will be referred to as MRS), the 8IG (Figure 20), the potassium platinum (II) hexathiocyanate derivative (PT), and the potassium platinum (II) hexathiocyanate-mersalyl double derivative (PTP) were good possibilities for more detailed investigation. The 8IG could be extremely useful as it is also a

substrate analog. Thus, not only does it possess a heavy atom (iodine) which could be useful in determining phases, but it could possibly serve to locate the active site region of the T_1 molecule.

Mersalyl and platinum hexathiocyanate are both essentially insoluble in the storage solution. To prepare the derivatives, enough salt to make 1.0 mM solutions was weighted out and added to the storage solution, but no appreciable amount dissolved. In 1 to 2 weeks, the MRS crystals began to develop large cracks, and the PT crystals began to turn the gold color of the heavy atom salt. The 8-iodo-guanosine was only sparingly soluble, and enough was added to make a saturated solution. These crystals also cracked upon soaking. The PTP was formed by first adding enough platinum hexathiocyanate to the standard soaking solution to make it 1.0 mM, and then soaking a T_1 crystal for 4 weeks (as with PT). Then, enough solid mersalyl acid was added to the solution to make it 1.0 mM in mersalyl, and the T_1 crystal was allowed to soak for 2 more weeks. Other double derivatives listed in Table 4 were prepared similarly. Each of the four derivatives discussed above gives significant differences in axial intensity distributions from the native enzyme (Figure 19), but exhibits similar unit cell constants and volume (Table 5).

Table 5: Native and Derivative Unit Cell Constants for T₁

<u>Derivative</u>	<u>a</u>	<u>b</u>	<u>c</u>	<u>Volume</u>
NAT	91.73(0.03) Å	37.320(0.006) Å	77.78(0.03) Å	266,200 Å ³
MRS	91.23(0.07) Å	37.42(0.03) Å	77.77(0.05) Å	265,500 Å ³
PT	91.48(0.04) Å	37.42(0.01) Å	77.86(0.03) Å	266,500 Å ³
PTP	91.46(0.02) Å	37.42(0.01) Å	78.01(0.02) Å	267,000 Å ³
8IG	91.72(0.02) Å	37.29(0.01) Å	78.04(0.02) Å	266,900 Å ³

D. Intensity Data Collection

The intensity data for the native enzyme and the heavy atom derivatives (except for MRS) were measured with the Nicolet P3/F diffractometer. The Wyckoff omega step scan option was used, with the width of the scan determined by examining the (13, 0, 14) reflection, the largest in the data set, as a function of ω . The peak is counted in increments of 0.05° in ω until only the background intensity remained. The width of the scan was chosen such that it incorporated over half of the peak. The ω angle at which the background was measured was also determined from this peak profile. The scan speed was then selected so that each reflection was counted for about 35 seconds. The alignment of the crystal was monitored by measuring 3 reflections every 100 reflections (about once an hour). Two of the monitored reflections are located at high χ angles that are 90° apart in ϕ [(7,8,0) and (0,10,4)], and one at $\chi = 0^\circ$ (18,0,8). This gives reflections in virtually orthogonal directions so that motion or slippage in any direction is easily detected. When the intensity of any of these reflections dropped below 85% of its original intensity, the automatic realignment procedure was initiated. In the automatic realignment, 12 reflections, selected in such a way as to be distributed broadly in reciprocal space, are centered, and a new orientation matrix is calculated from their angles.

The absorption connection requires reflections be measured as they are rotated about their diffraction vector. In order for the measurement to be physically possible, the χ angle of the reflection must be near 90° . The crystals were mounted with the b^* axis at $\chi = 90^\circ$ (and arbitrary ϕ). Three (0k0) reflections at various 2θ values [(0,2,0), (0,6,0), (0,10,0)] were used to determine the absorption curves.

At the beginning and at the end of a three-dimensional intensity data collection, the intensities of the (h0l) zone at 6.0 Å resolution were measured. An (h0l) Patterson projection or electron density projection can then be calculated. For a given crystal, this permits an independent measure of crystal decay in these low order data. Peaks in either the Patterson projection or electron density projection can be examined and compared in the before and after maps, and from this a decay correction determined. A similar technique can be used to determine the scaling factor among different crystals.

The main body of the data was collected in shells of increasing 2θ . Intensity data were collected to $2\theta = 36^\circ$, or 2.5 Å resolution for the native (NAT), PT, and PTP derivatives. At the end of the native data collection, the crystal showed no decay, so the data to $2\theta = 25^\circ$ were remeasured. Since the R factor $(\sum ||F_1| - |F_2|| / \sum |F|)$, where $\sum |F| = \sum |F_1| = \sum |F_2|$) was 0.033, the average intensity

of the two data sets was used to calculate the native structure factors in that 2θ range. This averaging improves the standard error of the measurements by a factor of $(2)^{\frac{1}{2}}$. Having data measured well for the NAT data set is particularly important because each of the derivative data sets will be compared to it, and so it will be involved in all of the MIR calculations. It is also these data to which the final phases will be applied in order to generate the electron density map of T_1 . The 8IG data were collected to $2\theta = 32^\circ$ (2.8 Å resolution), and the MRS only to $2\theta = 25^\circ$ (3.5 Å resolution). Since the crystals are orthorhombic: $|F(hkl)| = |F(-hkl)| = |F(h-kl)| = |F(hk-l)| = |F(h-k-l)| = |F(-hk-l)| = |F(-h-k-l)|$; thus, only one-eighth of the reflecting sphere is independent and must be collected. Data collection parameters for each of the data sets collected for this study are summarized in Table 6.

The minimum intensities that are considered to be observed (as opposed to background) are derived from consideration of the average negative intensities after background averaging. This minimum (the unobserved limit) is chosen to be about 2.5 to 3 times the average negative intensity. Intensities less than this are considered to be unobserved and are given an intensity equal to half of the unobserved limit.

The scale to be applied when merging native and derivative structure factors is determined four ways.

Table 6: Some Quantities Related to Data Collection.

Derivative	$2\theta_{\max}$	Resolution	Voltage	Current	Reflections	Number Observed	Decay ^a	$R(3.5 \text{ \AA})$ ^b
NAT	36°	2.5 Å	50 KV	10 mA	13342	94.7%	0%	0.033 ^c
PT	36°	2.5 Å	50 KV	10 mA	9814	95.3%	2%	0.150
PTP	36°	2.5 Å	50 KV	20 mA	9797	85.9%	6%	0.283
8IG	32°	2.8 Å	50 KV	15 mA	7031	96.1%	0%	0.165
MRS ^d	25°	3.5 Å	40 KV	26 mA	3368	94.7%	3%	0.226
					43352			

^aDecay = $(I_o - I_f)/I_o$ where I_o is the average monitored reflection intensity before data collection and I_f is the average monitored reflection intensity after the data collection.

^b $R(3.5 \text{ \AA}) = \Sigma ||F_d| - |F_n| | / \Sigma |F_n|$.

^cFor NAT, $R(3.5 \text{ \AA}) = \Sigma ||F_{n_2}| - |F_{n_1}| | / \Sigma |F_n|$ where $\Sigma |F_n| = \Sigma |F_{n_1}| = \Sigma |F_{n_2}|$

^dThese data were collected with a Picker FACS-1 diffractometer.

First, the (h0l) Patterson projections for the derivative and for the native data are compared. A scale is determined to match the 15 largest peaks in the native Patterson map with the corresponding peaks in the derivative Patterson map. This scale is then applied to the derivative structure factors, and their radial distribution in 2θ is compared to that of the native. From this a scale and an overall difference thermal B factor is applied to make the radial distributions match as closely as possible. The difference thermal B factor corresponds to the difference in disorder between the crystals. Thirdly, the sum of the structure factors of the scaled derivative with B applied and the native are compared. Lastly, with MRS, after heavy atom refinement, but before the phasing calculations, the scale was varied to minimize the R factor between MRS and NAT. The radial distributions of $|F|^2$ of native T_1 and the derivatives are shown in Figure 21, and the corresponding scale and difference thermal B factors are listed in Table 7.

III. STRUCTURE ANALYSIS

A. Determination of Heavy Atom Positions

The initial determinations of the heavy atom positions were carried out using difference Patterson vector maps. The Fourier coefficients for this type

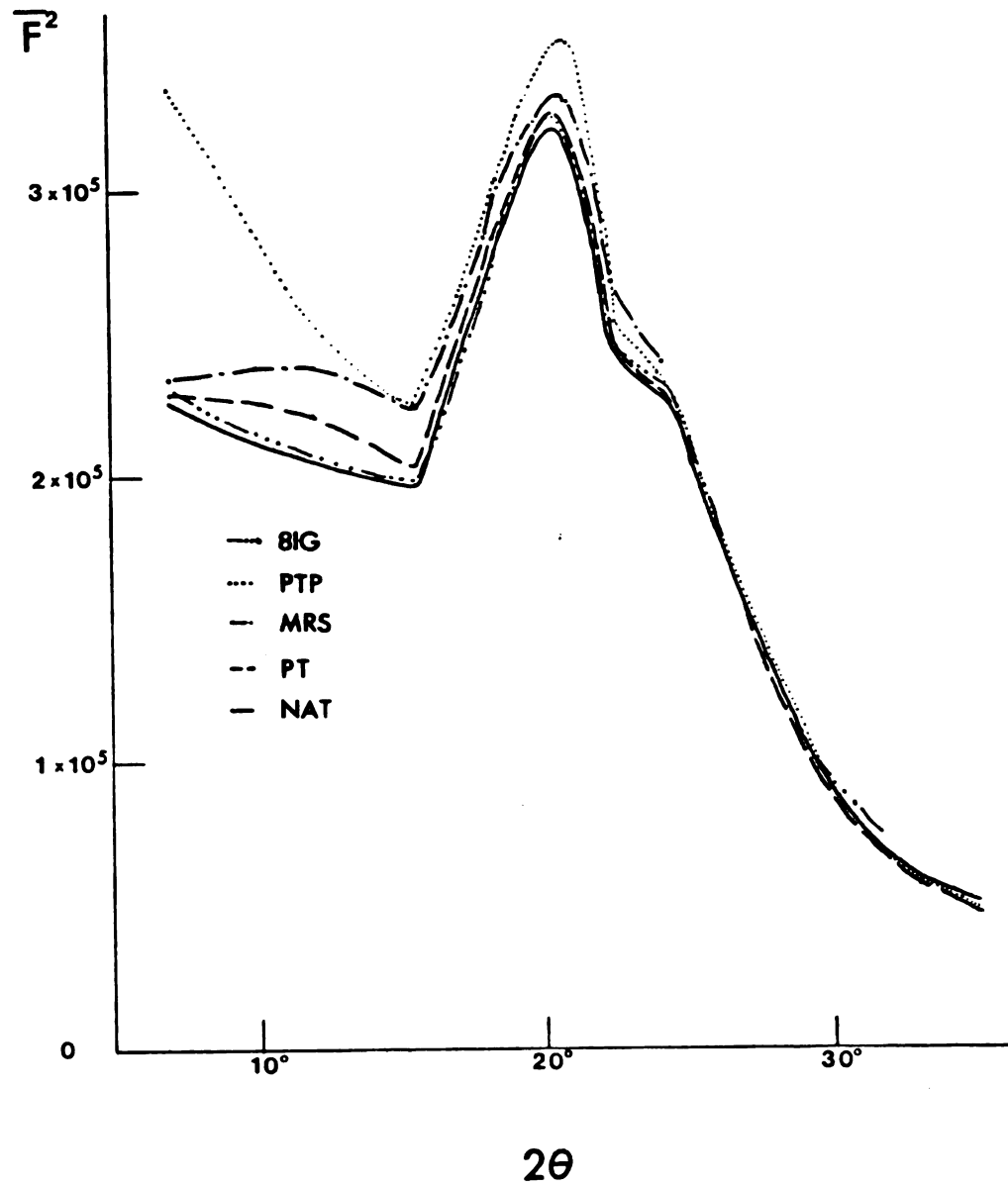


Figure 21: Radial Distribution of $|F|^2$ for T_1 Derivatives.

Table 7: Relative Scale and B Factors of Derivatives
of T_1 .^a

<u>Derivative</u>	<u>Scale</u>	<u>B</u>
NAT	1.00	0.0 Å ²
MRS ^b	6.51	2.5 Å ²
PT	0.83	0.0 Å ²
PTP	1.32	-4.0 Å ²
8IG	0.94	-2.0 Å ²

$$^a F_d = \text{scale}(|F_n|) \exp(-B(\sin \theta/\lambda)^2).$$

^b These data were measured on the Picker FACS-1 diffractometer which divides the intensities by 10. If these data were measured on the Nicolet P3/F the scale would be approximately 2.1.

of map are:

$$|\Delta F|^2 = ||F_d| - |F_n||^2 .$$

Peaks in the map indicate the interatomic vectors between the heavy atoms, with most of the effect of the protein (except for lack of isomorphism) subtracted out. The coordinate system used has as its axes Δx (u), Δy (v), and Δz (w). Two-dimensional projections of the Patterson map simplify the location of significant vectors by eliminating one coordinate, and in space groups containing a rotation or screw axis perpendicular to the projection, by having rotation satellite peaks (Ref. 71). In three-dimensions, certain sections of the Patterson, called Harker sections, must contain vectors due to the symmetry among asymmetric units in the unit cell (Ref. 72). These are particularly useful in determining heavy atom positions.

Initially, the two-dimensional centric zones [(hk0), (h0l), (hk0)] and the Harker sections ($u = \frac{1}{2}$, $v = \frac{1}{2}$, $w = \frac{1}{2}$) of the three-dimensional map are calculated. The relationships which generate the expected vector peaks in the space group $P2_12_12_1$ are given in Table 8. The positions of the heavy atoms are determined most easily from the projections and Harker sections, and then the three-dimensional Patterson is used to locate cross-vectors between the different positions.

Table 8: Harker Vectors and Cross-vectors.

Equivalent Positions for Space Group $P2_1^2 2_1^2 2_1$

<u>x</u>	<u>y</u>	<u>z</u>
$1/2 - x$	$-y$	$1/2 + z$
$1/2 + x$	$1/2 - y$	$-z$
$-x$	$1/2 + y$	$1/2 - z$

Vectors Relating Space Group Symmetry
Related Positions for $P2_1^2 2_1^2 2_1$

<u>u</u>	<u>v</u>	<u>w</u>
$1/2 + 2x$	$2y$	$1/2$
$1/2$	$1/2 + 2y$	$2z$
$2x$	$1/2$	$1/2 + 2z$

Cross-vectors for Space Group $P2_1^2 2_1^2 2_1$

$(x_1 - x_2)$	$(y_1 - y_2)$	$(z_1 - z_2)$
$(x_1 + x_2) - 1/2$	$(y_1 + y_2)$	$(z_1 - z_2) - 1/2$
$(x_1 - x_2) - 1/2$	$(y_1 + y_2) - 1/2$	$(z_1 + z_2)$
$(x_1 + x_2)$	$(y_1 - y_2) - 1/2$	$(z_1 + z_2) - 1/2$

The MRS difference Patterson was the most straightforward to interpret. The Harker sections are shown in Figure 22. The peaks marked 1 are the self vectors due to a mercury. There are two vectors marked 1 in the $v = \frac{1}{2}$ section because the peak located at $w = \frac{1}{2}$ is the result of the Hg 1 position having a y coordinate of $\frac{1}{2}$. There is a second mercury position in this map at the position marked 2 (this peak went unnoticed in the pH 7.3 work) although its indication in the $w = \frac{1}{2}$ section is weak. In the three-dimensional Patterson map, the only other vectors present are those due to the Hg 1 — Hg 2 cross-vectors. This analysis gives the coordinates for the two mercury positions listed in Table 9.

Table 9: Mercury Positions for MRS in Unit Cell Fractions

	<u>x</u>	<u>y</u>	<u>z</u>
Hg 1	0.348	0.250	0.083
Hg 2	0.266	0.395	0.205

The PT difference Patterson map proved to be a surprise. Although the axial intensities show very substantial changes (Figure 19), and although the R factor between PT and NAT was over 15% (Table 6), the Patterson map (Figure 23) is nearly featureless. This is puzzling since there are heavy atom derivatives in

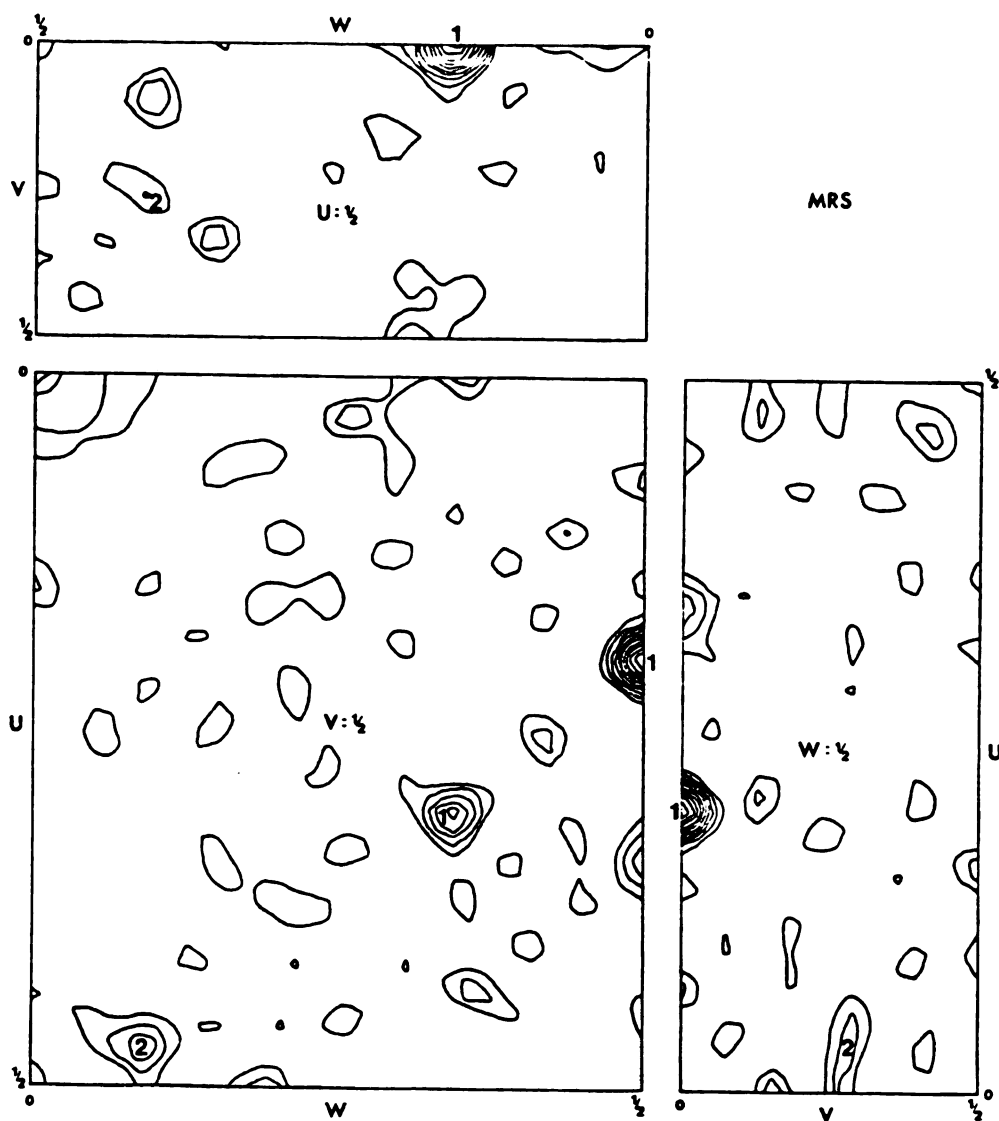


Figure 22: Harker Sections of MRS Difference Patterson Map. The map is contoured at a level corresponding to a vector between two 30 electron peaks, and incremented at the same level.

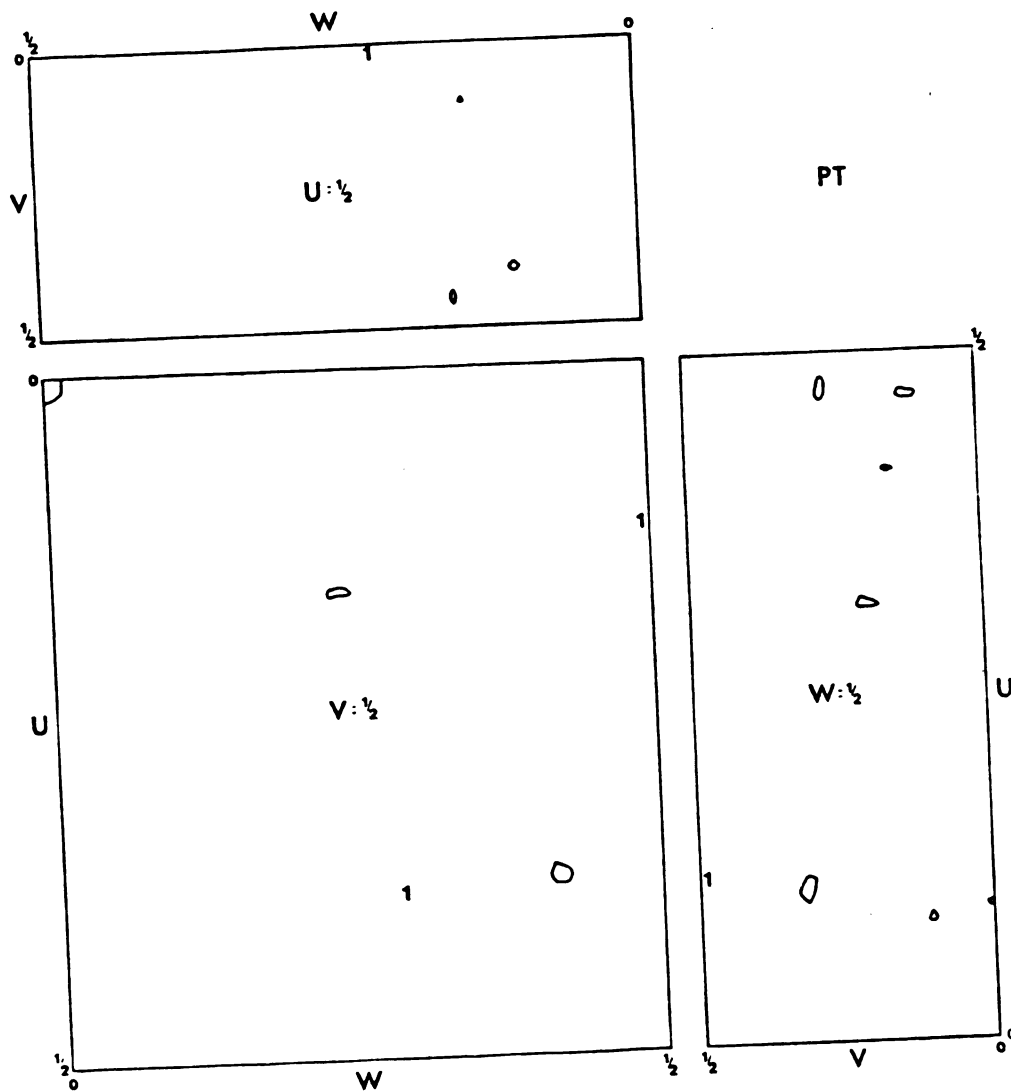


Figure 23: Harker Sections of PT Difference Patterson Map. The map is contoured at a level corresponding to a vector between two 30 electron peaks, and incremented at the same level.

this lab of other proteins which give well determined positions with R factors of 15%. There are two very weak peaks, but comparison with the MRS map (Figure 22) renders them insignificant. If Hg 1 is a full occupancy mercury atom (even though it is very difficult to get 100% occupancy with a soaking experiment), these peaks correspond to only 25 to 30 electrons (10 to 15 electrons is the observable limit). With a heavy atom contribution so small, even if there was a well determined position, which there is not, the phase angle (Figure 3) has a very $-\vec{F}_h$ side, and so it would be difficult to deduce accurate phase information. There are no positions in direct space which account for the few features seen in the PT Patterson map.

The PTP difference Patterson map is entirely different from the PT map. Although the axial intensity changes of PTP were of the same order as the PT (Figure 19), the R factor was the largest of the derivatives, 28% (Table 6), and the difference Patterson map is full of peaks (Figure 24). Conspicuously absent, however, are the large mercury peaks of MRS. Thus, mersalyl apparently did not systematically bind in this crystal. Additional changes in the intensity distribution must be due to the platinum hexathiocyanate reagent. This is supported by the fact that the PTP crystals did not crack on addition of mersalyl as the MRS crystals did. There are many broad, rambling peaks in the map, but no sharp,

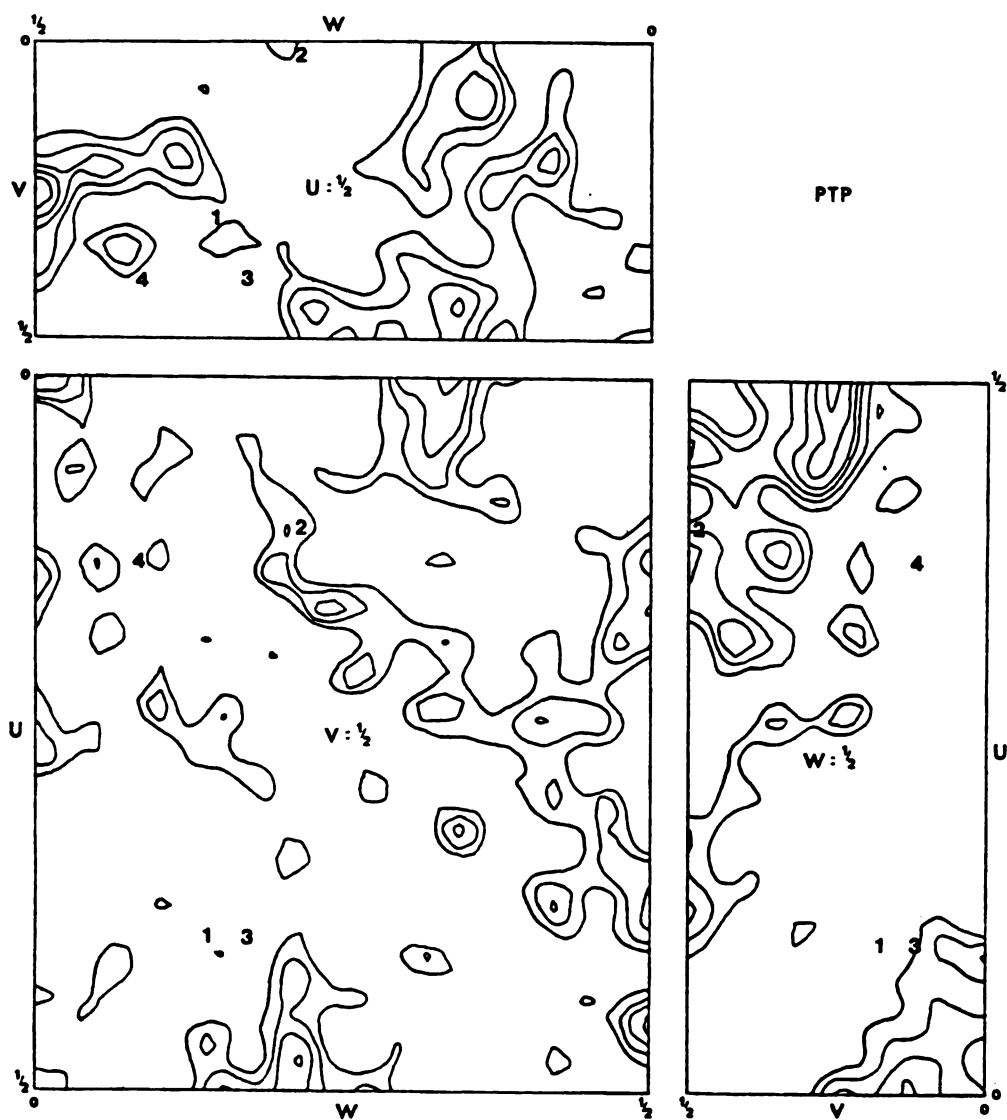


Figure 24: Harker Sections of PTP Difference Patterson Map. This map is contoured at a level corresponding to a vector between two 30 electron peaks, and incremented at the same level.

well defined peaks which can be related to a single heavy atom. Probably, the peaks are the result of disruption of the protein structure. If there are bound platinum atoms in the crystal, their vectors are masked by the apparent large lack of isomorphism between the PTP and NAT crystals.

The 8IG derivative was to have a key role in this structure determination: a) as a possible heavy atom derivative, and b) as a marker to locate the active site region of T_1 . It is known that 8-iodo-guanosine is a competitive inhibitor of T_1 in solution, so it could bind selectively to the active site. The iodine is electron-rich enough to aid in phasing the protein, and the (8IG - NAT) difference electron density map would locate the position of the nucleotide, and probably the active site. Besides the interest inherent in the active site, it would be helpful in establishing the position of the α -carbon chain along the sequence since the nucleotide is known to bind near Glu 58, Arg 77, and His 92 in the active site. Once again, the axial intensity distributions show large changes (Figure 19), and the R factor between 8IG and NAT is a respectable 0.165. Unfortunately, however, once again the difference Patterson map of 8IG is devoid of large peaks (Figure 25). None of the small features of the map could be correlated with consistent iodine positions, and none of the peaks came close

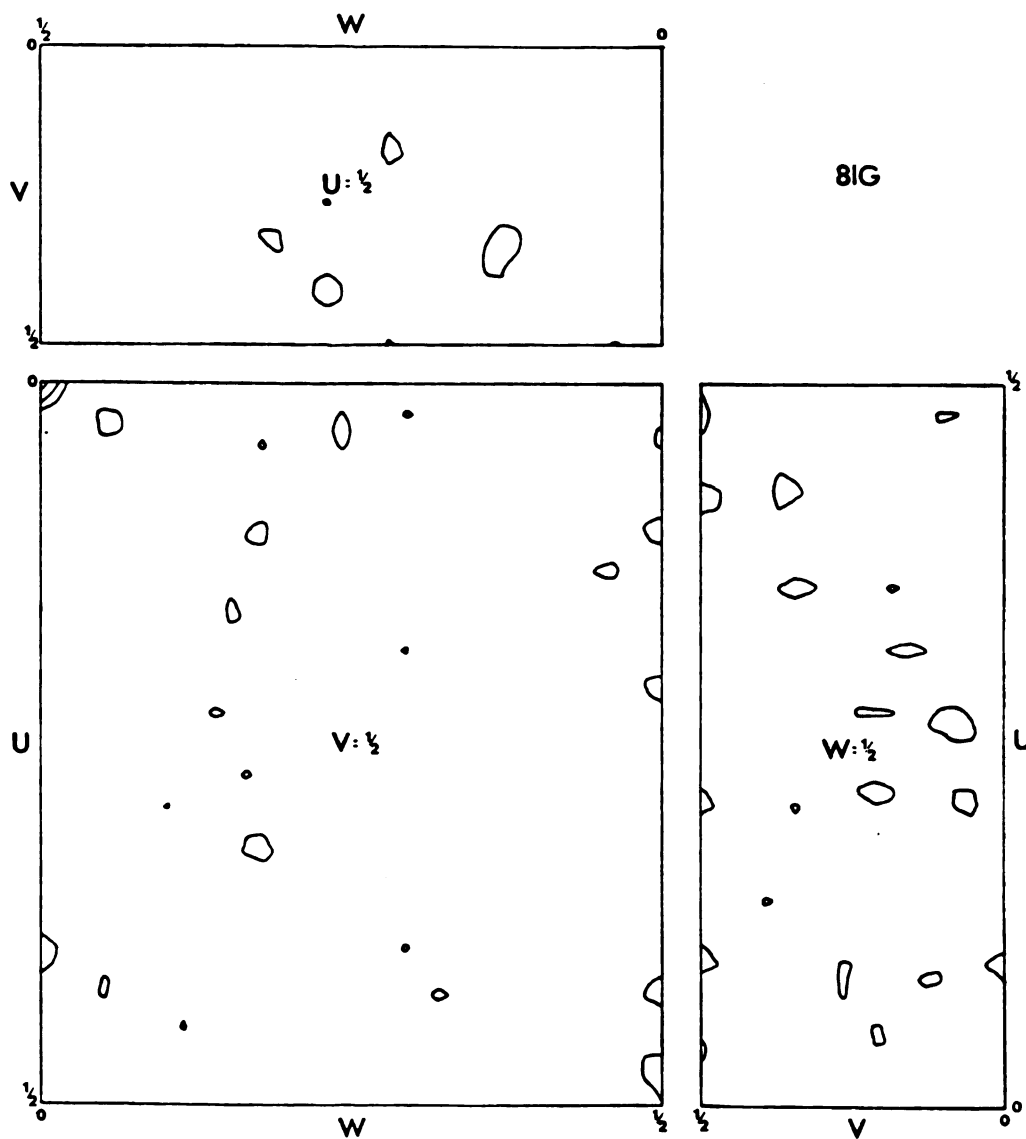


Figure 25: Harker Sections of 8IG Difference Patterson Map. The map is contoured at a level corresponding to a vector between two 30 electron peaks, and incremented at the same level.

to corresponding to vectors of 50 electron atoms representing iodine.

B. Difference Electron Density Methods

An attempt was made to use the MRS derivative to help interpret the PT, PTP, and 8IG difference Patterson maps using difference electron density methods. Reflections in the (hk0), (h0l), and (0kl) zones are centric in space group $P2_12_12_1$. This means that the structure factors will not be complex numbers, but will be entirely real (phase angle of 0° or 180°) or entirely imaginary (phase angle of 90° or 270°). Phase angle selection is thus greatly simplified, and a single heavy atom derivative will phase these reflections.

In order to determine the mercury positions better in the MRS derivative, it was decided to refine their positions, and occupancies using the full matrix least squares refinement program from SHELX76 (Ref. 73) crystal structure determination package. The quantities $||F_d| - |F_n||$ were used as the amplitudes of the observed structure factors ($|F_h|_o$). For the centric reflections the approximation is generally accurate, but fairly large experimental errors arise because $|F_h|_o$ is due to the difference between two large numbers (median $|F_d| = |F_n| = 300$) resulting in a relatively small one (median $|F_h|_o = 100$). The approximation is not generally correct for noncentric reflections since $\vec{F}_d$ and $\vec{F}_h$ are generally noncollinear.

Since $F_h = F_d - F_n$ is correct for the centric reflections, and because the centric reflections are able to be phased from only one heavy atom derivative, refinement began with only these reflections (there are 597 observable centric reflections at 3.5 Å resolution). One cycle of coordinate refinement, with the occupancies and thermal B factors held constant, was calculated first. Since the coordinates were measured from the Patterson maps, and the Patterson peaks are rather broad, it was felt that their positions should be improved first. The R factor ($R = \Sigma ||F_h|_o - |F_h|_c| / \Sigma |F_h|_o$) dropped slightly, from 0.516 to 0.511, and all of the coordinate shifts were insignificantly small. This was followed by a cycle of occupancy refinement, keeping the coordinates constant at the values calculated from the first cycle, with the B value constant. Experience has shown that the structure factors are more sensitive to occupancy than either coordinates (assuming the coordinates are nearly correct) or B factors. The occupancies of the mercury atoms increased from 0.80 for Hg 1 and 0.50 for Hg 2, to 0.866 and 0.578 respectively, decreasing the R factor to 0.504. Lastly, with the occupancies and coordinates held at their newly calculated values, a cycle of refinement on the thermal B values was calculated. The B values decreased appreciably, from 12.0 to 10.2 Å² for Hg 1, and from 15.3 to 9.7 Å² for Hg 2 (15% for Hg 1 and 37% for Hg 2), but the R factor

remained the same. Thus, the B factor did not have much effect on the calculations.

At this point acentric data were added to the calculation. To minimize assumptions, only acentric reflections with the largest differences were included. The new data set was composed of the centric reflections, plus the 30% of the acentric data with the largest differences (604 additional reflections). The addition of the acentric data lowered the R factor to 0.447. Again, the coordinates were refined for one cycle. The coordinates underwent additional small shifts which decreased the R factor to 0.445. A cycle of occupancy refinement was calculated next, and once again the occupancies increased (to 0.950 for Hg 1, and to 0.591 for Hg 2). Since the occupancy and the B factors are correlated, much of this shift is probably due to the decrease in the B factors in cycle 3 of the centric refinement. The change in occupancies dropped the R factor to 0.435. The B factors were again allowed to vary, and again they became smaller, dropping the R factor to 0.433. However, it was decided at this point that the B factors were unrealistically low compared to the native data Wilson behavior ($5.3 \text{ \AA}^2$ for Hg 1, and $7.62 \text{ \AA}^2$ for Hg 2), so they were raised to values more typical of the native protein crystal ($12.0 \text{ \AA}^2$ for Hg 1 and $15.3 \text{ \AA}^2$ for Hg 2). With the new B values, the occupancies were calculated for the final time giving a final R factor of 0.429. This value is

quite acceptable considering the aforementioned approximations. Undoubtedly, there are also some unaccounted contributions due to departures from isomorphism between the MRS derivative and the native protein crystal. The progress of the parameters of the refinement are summarized in Table 10.

In order to verify the mercury positions in yet another way, structure factors were calculated from each of the mercury positions alone, and compared with the (MRS - NAT) differences. When the Hg 1 position is calculated alone, the R factor is 0.489, and when the Hg 2 position alone is calculated, the R factor is 0.639 (the last 2 columns of Table 10). This confirms that, although Hg 1 is dominant in the derivative, Hg 2 also makes an important and significant contribution.

The difference electron density projections of MRS calculated using the phases derived from the parameters of the mercury atoms from the SHELX refinement are shown in Figure 26. The 2 mercury atoms are easily discernible, and their relative peak heights are consistent with the relative occupancies calculated. It can also be seen in this map that there are no more significant minor binding sites, and that lack of isomorphism is minimal.

Difference electron density projections were also calculated for each of the PT, PTP, and 8IG derivatives using phases deduced from the MRS refinement. The phases for the centric reflections of the native protein were

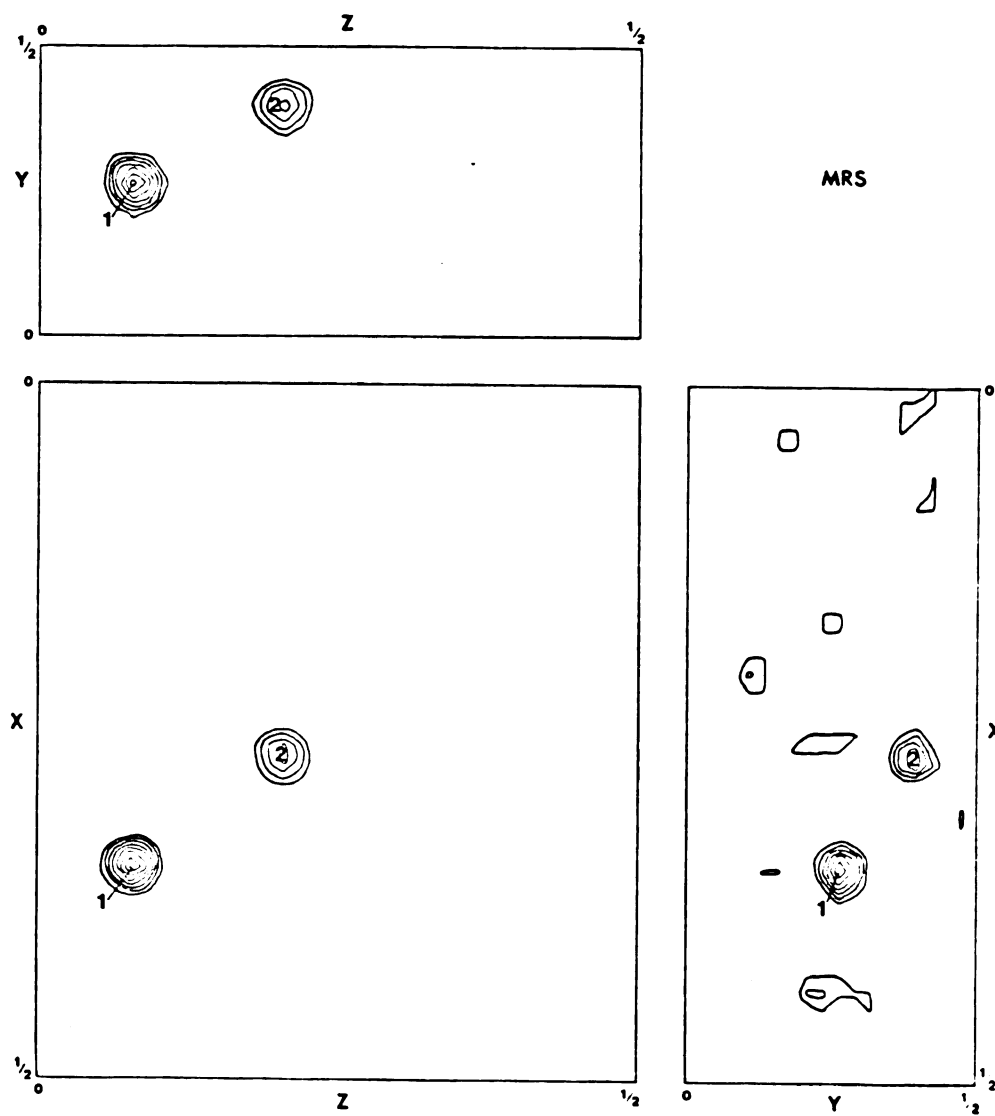


Figure 26: MRS Difference Electron Density Projections. This map is contoured in electron density levels starting at 30 electrons in increments of 5 electrons.

Table 10: SHELX Refinement of the Mercury Atoms in MRS

Cycle	0	1	2	3	4	5	6	7	Hg 1	Hg 2
Hg 1										
Occ	0.8		0.886			0.950		0.906	0.906	0.0
x	0.3466	0.3459			0.3454			0.3454		
y	0.2626	0.2620			0.2642			0.2642		
z	0.0809	0.0810			0.0801			0.0801		
B ($\text{\AA}^2$)	12.0			10.2			5.3	12.0		
Hg 2										
Occ	0.5		0.578			0.591		0.564	0.0	0.564
x	0.2676	0.2680			0.2683			0.2683		
y	0.3954	0.3951			0.3944			0.3944		
z	0.2046	0.2044			0.2048			0.2048		
B ($\text{\AA}^2$)	15.3			9.7			7.6	15.3		
R_{Cullis}^a										
	0.516	0.511	0.504	0.504	0.447					
R^b										
					0.445	0.435	0.433	0.429	0.489	0.639

$$^a R_{\text{Cullis}} = \sum ||F_d| - |F_n|| - |F_h| / \sum ||F_d| - |F_n|| \quad \text{for centric reflections (Ref. 74)}.$$

$^b R$ as above, but with all reflections.

inferred from the mercury contribution calculated with the refined heavy atom parameters of the SHELX refinement.

The derivative structure factors are the sum of the native and heavy atom structure factors in the centric case:

$$\vec{F}_d = \vec{F}_n + \vec{F}_h$$

The amplitude and phase of the mercury atom structure factor are known from Equation (1) because the positions are known from Patterson analysis, and the occupancies from heavy atom refinement. Since only the amplitudes of F_d and F_n are known,

$$\vec{F}_h = S_d |F_d| - S_n |F_n|$$

where S_d and S_n are the algebraic signs of the derivative and native structure factor, respectively. From the magnitudes of F_d and F_n , and from the sign and magnitude of F_h , the signs of the native and derivative structure factors, S_n and S_d , and hence their phase angles (since they are centric) can be deduced.

As a first approximation the signs or phase angles for the derivatives can be assumed to be the same as the native. This assumption is valid for centric reflections except possibly in those cases where the protein contribution is much smaller than the heavy atom contribution where the sign can change (the phase angle

changes by 180°). These are referred to as crossover reflections, and are only expected in a small percentage of the reflections (less than 10%). Difference electron density maps were calculated for the PT, PTP, and 8IG derivatives on the basis of native phases. Since the phase angle of the protein is independent of the heavy atom derivative used to calculate it (as long as it is a *bona fide* derivative), difference electron density maps using the correct protein phase angle will synthesize the projection of the heavy atoms for any given derivative. Since the space group of T_1 has three mutually perpendicular projections, each coordinate of the heavy atoms can be deduced twice.

The difference electron density projection of (PT - NAT) shows a few significant features (Figure 27), consistent with the Patterson map. Examination of the map revealed only a single position consistently present in all three projections (marked as 1 in Figure 27). When the self-vectors (those vectors arising due to crystal symmetry and found in the Harker sections) are examined in the difference Patterson map, they do not fall on any vector peaks (marked as 1 in Figure 23). Thus, there are no platinum heavy atom positions consistent with both the difference electron density projections, and the difference Patterson maps. As an additional check, phases were determined from the assumption of a 30 electron occupancy in this position, and applied to the (MRS - NAT)

differences. The mercury atoms did not appear in the difference electron density projections, indicating that the mercury positions and the platinum positions did not assign the same set of phases to the native protein. On the basis of the Patterson maps, MRS appears to be a much better derivative than PT, so it was concluded that PT is not a *bona fide* derivative.

The difference electron density projections of (PTP - NAT) (Figure 28), show many peaks, but none that is outstanding compared to the MRS difference electron projection (Figure 26). There are, however, four peaks present in all three projections (marked 1 through 4 in Figure 28). When the self-vectors of these positionally consistent electron density peaks are examined in the difference Patterson map, none of them falls on three independent vector peaks (marked 1 through 4 in Figure 24). Position 2 corresponds to the Patterson best, but its electron density is almost insignificantly small. Position 3, corresponding to the largest electron density peak and having promising coordinates, does not correspond well to vector positions. Thus, disappointingly once again, there are no heavy atom positions consistent with both the difference electron density projections, and the difference Patterson maps.

The difference electron density projections of (8IG - NAT) are essentially empty when contoured at the same level as the other projections (Figure 29). Since

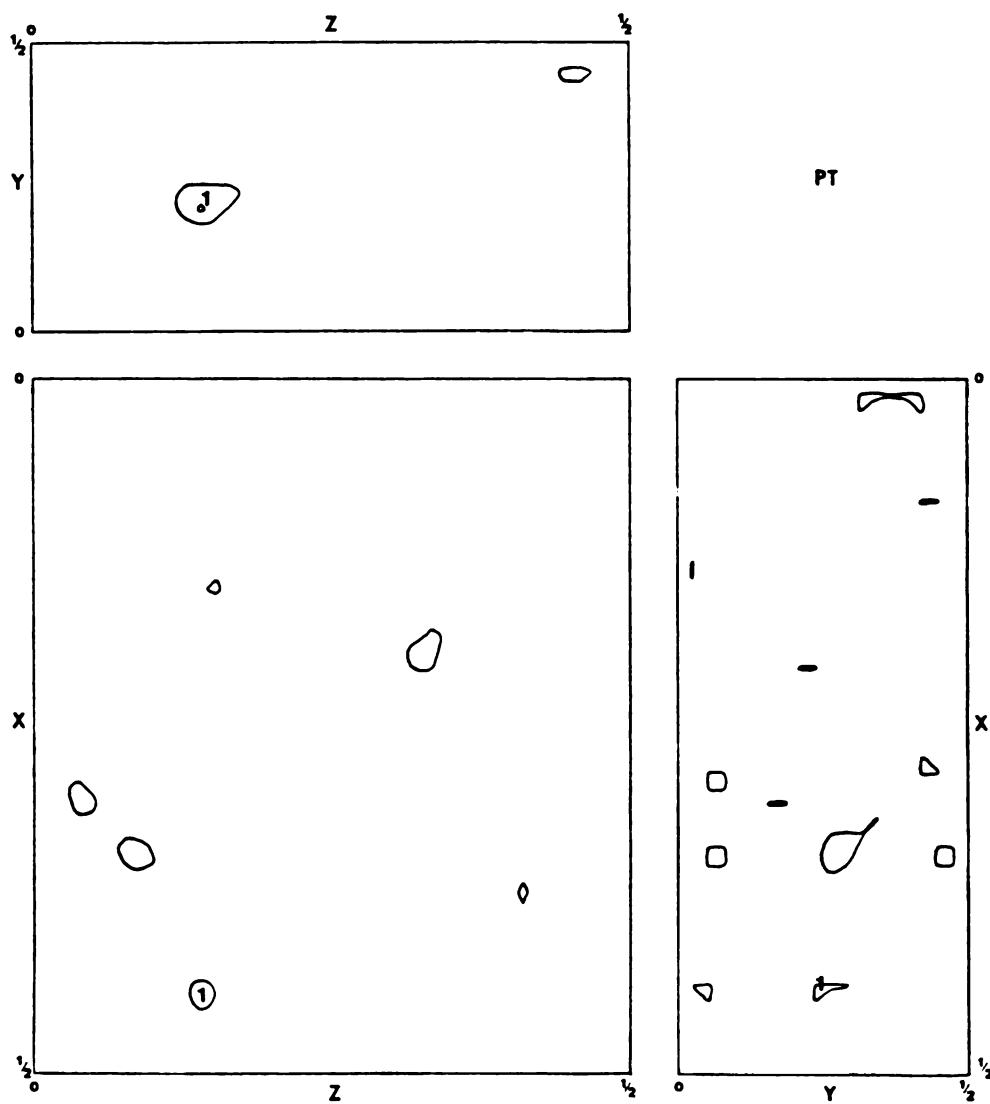


Figure 27: PT Difference Electron Density Projections. This map is contoured in electron density levels starting at 30 electrons and in increments of 5 electrons.

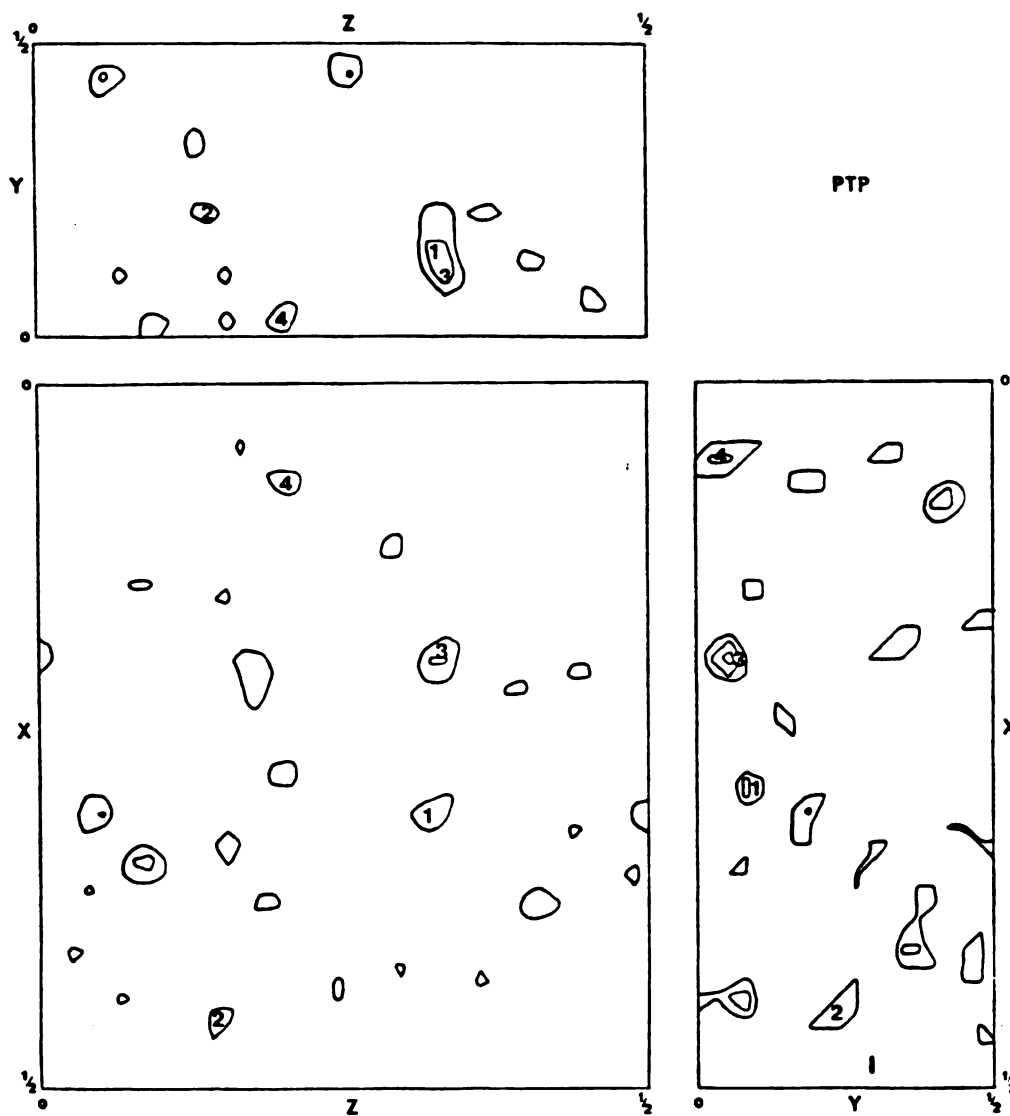


Figure 28: PTP Difference Electron Density Projections. This map is contoured in electron density levels starting at 30 electrons and in increments of 5 electrons.

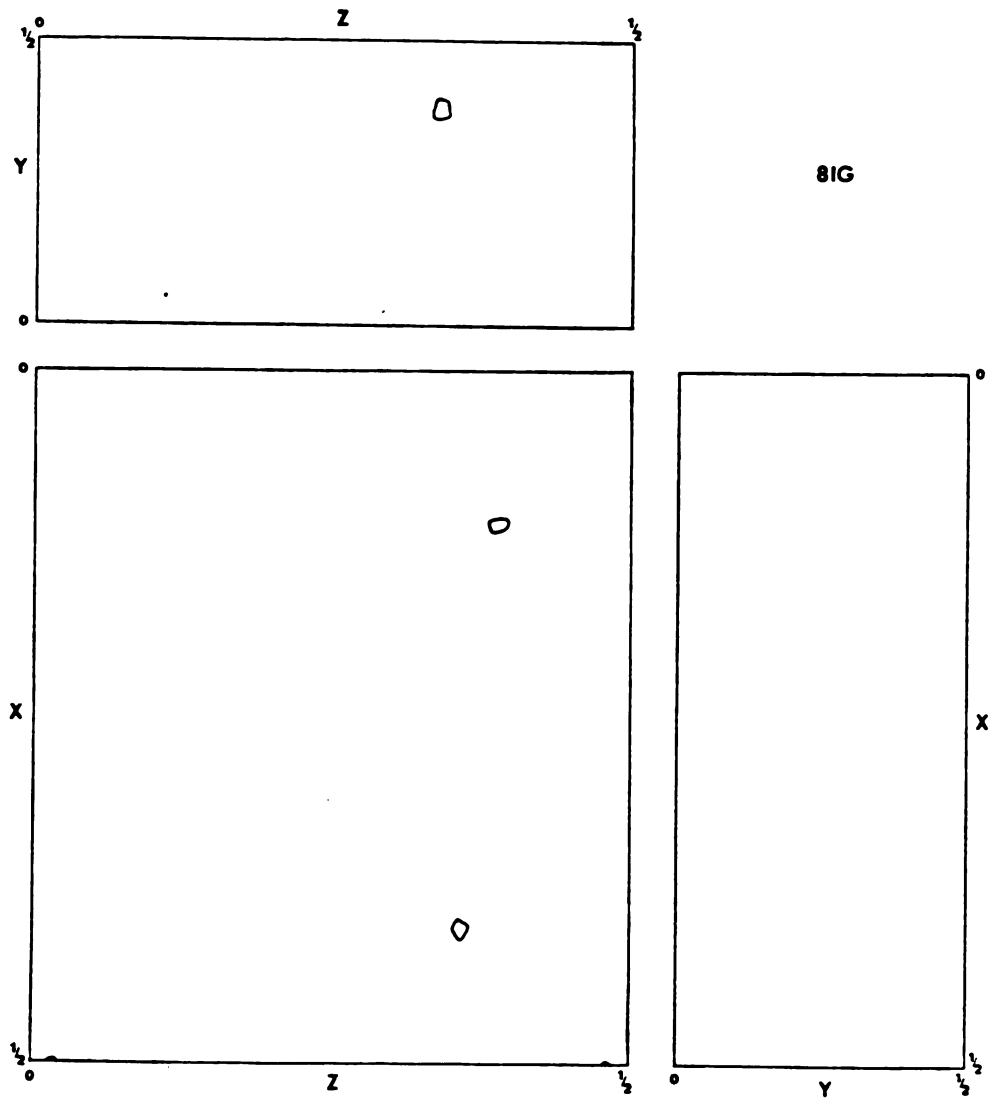


Figure 29: 8IG Difference Electron Density Projections. This map is contoured in electron density levels starting at 30 electrons and in increments of 5 electrons.

this derivative would be useful even if it is too weak to be used in phasing, it was also examined at lower contour levels to see if there was possibly a very low occupancy site, but none was found. Thus, 8IG has no iodine positions consistent with either the difference electron density projections or the difference Patterson maps, and no guanosine position could be located either.

In summary, of the four derivatives which show large changes in intensity, and significant changes in the structure factors (indicated by the R factors between the derivatives and native), only MRS and PTP show promising difference Patterson maps, and only the MRS derivative is suitable for phase determination. Consequently, MIR methods cannot be used for determining the phases of T_1 with the data presently available, and had to be abandoned.

IV. PHASE DETERMINATION

A. The Iterative Single Isomorphous Replacement Method

Since the PT, PTP, and the 8IG derivatives could not be used for phasing by MIR, and since heavy atom searches are very time consuming (the reported heavy atom searches were underway for four years) the iterative single isomorphous replacement method (ISIR) was used to determine protein phases from the MRS derivative alone. This work

was done under the supervision of Dr. B.C. Wang, who developed this variation of SIR at his facilities in the Biocrystallography Laboratory of the Veteran's Administration Medical Center in Pittsburgh, Pennsylvania. Although the method has not been utilized as much as MIR, nonetheless, it has produced a number of successful applications in protein structure determination (Refs. 75,76). Compared to the alternatives, ISIR appeared to be the most expedient recourse for attaining our original structural goals.

Wang describes ISIR in a series of seven steps (Ref. 77). The first step involves preparing a data file which contains the observed structure factor amplitudes of the native protein, and their corresponding figures of merit, centroid phases, and phase probability coefficients calculated from the SIR data. A double-phased Fourier calculated by the method of Rossmann and Blow (Ref. 78) is then synthesized, which is equivalent to the SIR method described by Kartha (Ref. 79). If the angle α is defined to be the difference between the native protein phase angle, α_n , and the heavy atom phase angle, α_h , then the two solutions given by the phase circles (Figure 30) are $(\alpha_h + \alpha)$ and $(\alpha_h - \alpha)$. One of these contributes to the true structure, and the other does not. When both phases are included in the calculation,

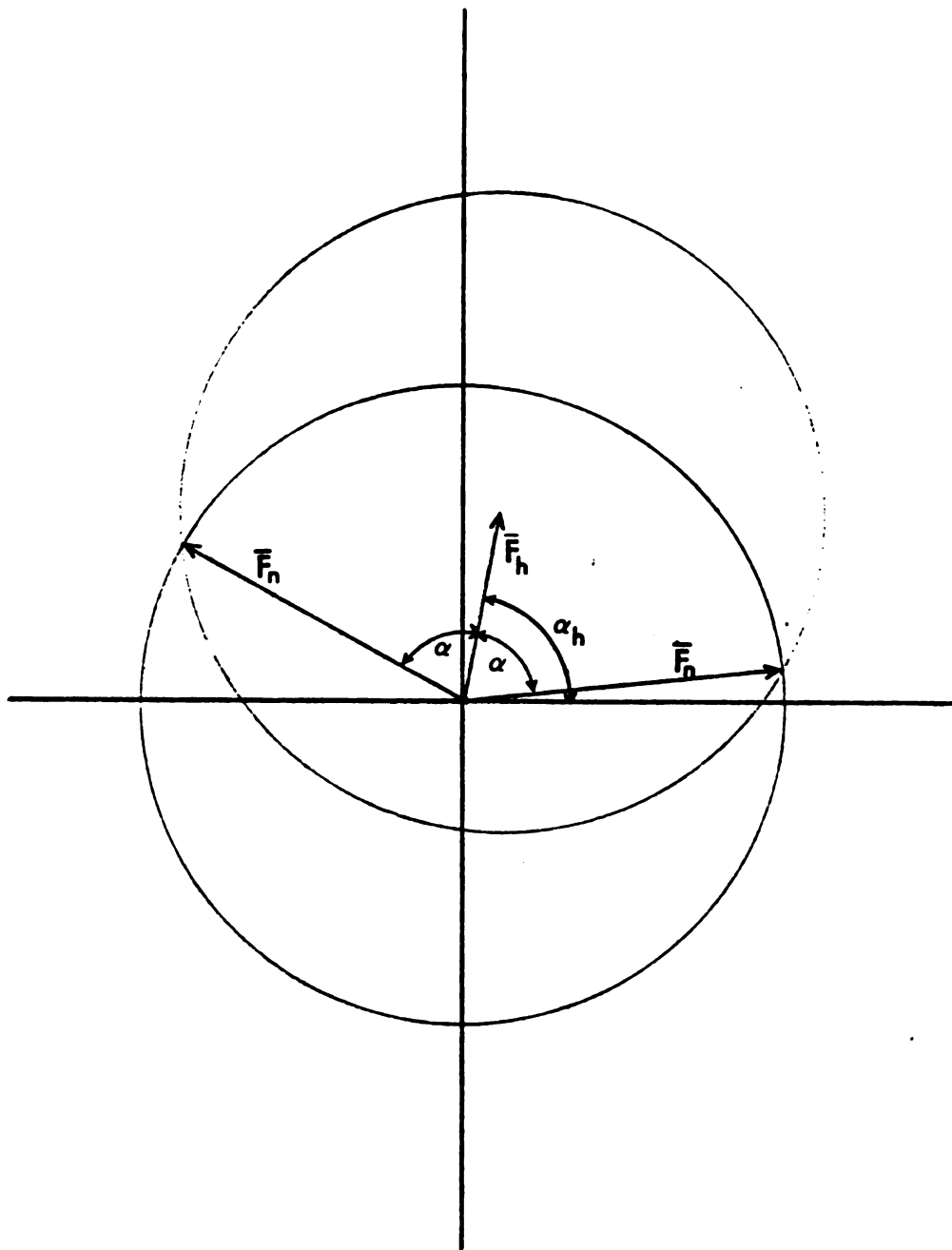


Figure 30: The Single Isomorphous Replacement Phase Ambiguity. In this construction α_h is the heavy atom phase angle and α is $(\alpha_h - \alpha_n)$ where α_n is the native protein phase angle.

$$\rho(\vec{r}) = \frac{1}{V} \sum_{\vec{s}} |F_n| \left[e^{i(\alpha_h + \alpha)} + e^{i(\alpha_h - \alpha)} \right] e^{-2\pi i(\vec{s} \cdot \vec{r})}$$

where $\vec{r}$ is the vector to coordinates (x,y,z) and $\vec{s}$ is the diffraction vector (h,k,l) such that $\vec{s} \cdot \vec{r} = hx + ky + lz$.

This can be shown to reduce to

$$\rho(\vec{r}) = \frac{1}{V} \sum_{\vec{s}} \left[2|F_n| e^{i\alpha_h} e^{-2\pi i(\vec{s} \cdot \vec{r})} \right] \cos \alpha$$

which is a $\cos \alpha$ weighted, heavy atom phased protein electron density map. Thus, the heavy atom phase angle is used at full weight when it is in phase with α_n , and is weighted -1.0 when it is exactly out-of-phase (equivalent to changing the quadrant of α_n). When α_h is 90° or 270° out-of-phase with α_n , its weight is zero.

The double phased Fourier can be thought of as two electron density maps superimposed on each other, one created by the true phases (a true image of the molecule) and one created by the false phases (random noise). If the false density map is totally random, ρ -false is essentially a constant added to ρ -true, but in the real case, ρ -false gives rise to some well defined spurious peaks which may make the interpretation of ρ -true difficult (Ref. 79).

In the third step, the ISIR technique attempts to increase the map to noise ratio by making phase selections using information known about the physical make up of

the crystal. Fourier methods have been used to determine crystal structures from partial structural information if the partial structure constitutes a significant portion of the scattering density. ISIR attempts to define the gross image of the molecule which is then upgraded, and which serves as the "partial structure" that can be effectively improved to become the total structure. It is therefore possible to obtain the total and best structural information which in turn points out which phase solutions of SIR are likely to be correct, ultimately overcoming the phase ambiguity.

The double-phased Fourier map is calculated with a grid spacing of about one-third of the nominal resolution, and it is used to calculate the protein-solvent boundary. The method used to accomplish this rests on the assumption that the average electron density of the protein is greater than the average electron of the solvent. Each grid point of the electron density map is evaluated in terms of its likelihood of being within the protein envelope. A grid point near high electron density is more likely to be within the protein envelope than one that is not. A new map is constructed such that the density at each grid point is proportional to the weighted sum of all positive density within a sphere of radius R from the corresponding point in the double-phased Fourier,

$$\rho_j = K \sum_i w_i \rho_i ,$$

where ρ_j is the density of the newly calculated map, ρ_i is the density of the double-phased Fourier map, K is a constant which will compensate for the lack of a $F(0,0,0)$ term in the Fourier sum, and w is a weight assigned to grid points of the double-phased Fourier map. The latter is estimated according to the following criteria:

if $r_{ij} < R$ and $\rho_i > 0$ then $w_i = (1 - r_{ij}/R)$

if $r_{ij} > R$ or $\rho_i < 0$ then $w_i = 0$

where r_{ij} is the distance between grid points ρ_j and any grid point ρ_i , and R is the radius of the sampling sphere. Test trials with known structures reveal that R chosen between 5 Å and 6 Å gives the fastest convergence with 3.5 Å resolution protein data (5.2 Å was used in this study). The errors in the map produced by the false phases, which are random in nature, will effectively sum to a constant and therefore have little effect on the new map. The new map is contoured in such a way that a percentage of the grid points equal to the percent mother liquor in the crystal fall below a threshold value. Those points are considered solvent, and the grid points which have density greater than the threshold value are considered protein. The resulting map has large regions of connected protein electron density surrounded by low solvent electron

density. A section of this map, contoured at 50% mother liquor, is shown in Figure 31. The blank areas represent protein regions, and the starred region indicates the solvent. The location of the protein is clearly visible in Figure 31, and closer examination reveals four areas of protein (marked as numbers 1 through 4), consistent with space group considerations.

In the fourth step of the ISIR method, this new map is used to filter the electron density in direct space. First, however, ρ^* , which may be regarded as the $F(0,0,0)$ contribution, must be calculated. It is defined by the ratio:

$$\frac{\rho^* + \langle \rho_{\text{solvent}} \rangle}{\rho^* + \rho_{\text{max}}} = S ,$$

where $\langle \rho_{\text{solvent}} \rangle$ and ρ_{max} are the average solvent electron density and the maximum electron density of the double-phased Fourier map, and S is a constant. Theoretically, S can be estimated from the solvent composition, overall thermal B factor, and resolution of the data once the phases are known. In practice S is chosen to be an empirical parameter adjusted to optimize convergence at a given resolution. There have not yet been a sufficient number of structures studied using ISIR for the value of S to be well characterized (Ref. 77). However, it has been investigated for Bence-Jones protein Rho crystals in 2.0 M ammonium sulfate and was optimized at 0.06 at

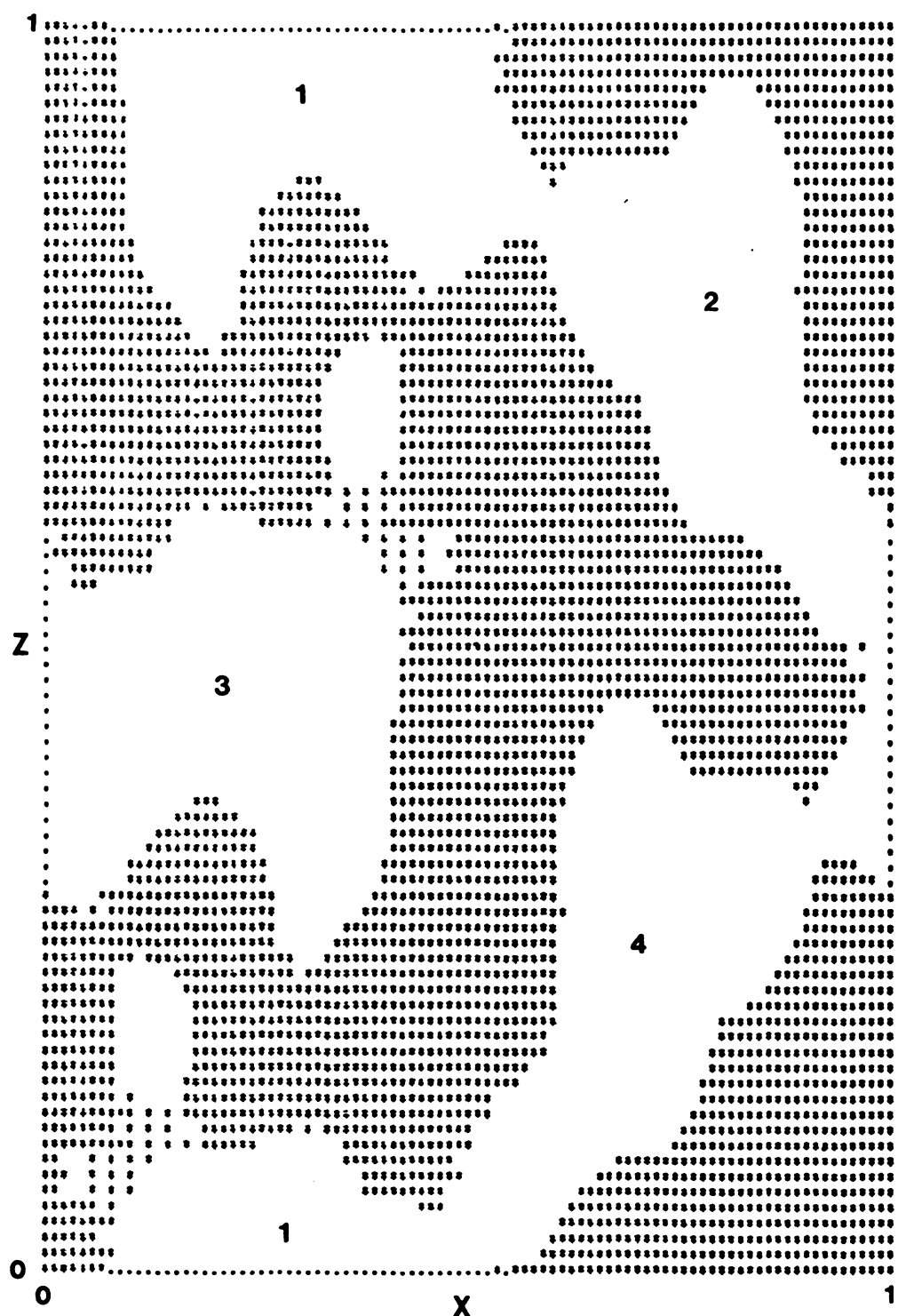


Figure 31: A Section of ρ_j Defining the Molecular Envelope. The $y=0$ section of the modified electron density map defining the molecular envelope used in the real space electron density filter in ISIR is shown.

3.0 Å resolution. If the true $F(0,0,0)$ term is included in the calculation, S is a larger number (0.17 for Rhe at 3.0 Å resolution). The direct space filtering treats the protein regions and the solvent regions differently. Once ρ^* is determined, it is added to the density within the protein region of the map, and any negative density remaining is considered an error and given a value of zero, whereas the solvent region is given a constant density of $(\rho^* + \langle \text{solvent} \rangle)$. It is this filtered map which contains the partial structure which is further refined and improved upon.

New structure factor amplitudes and phases are calculated in the fifth step by performing an inverse Fourier transform of the partial structure electron density map. In the sixth step, the phases are then filtered in reciprocal space. This is accomplished with a phase probability distribution of the structure factors determined from the inverse Fourier calculation. Bricogne's adaption (Ref. 80) of Sim's weighting scheme (Ref. 81) gives:

$$P_C(\alpha) \propto \exp \frac{2|F_O||F_C| \cos(\alpha - \alpha_C)}{\langle |F_O|^2 - |F_C|^2 \rangle} \quad (2)$$

where $P_C(\alpha)$ is the probability of the calculated phase angle (α_C) being a given phase angle (α) . The calculated phase angle probability distribution is then multiplied by the probability distribution from the double-phased

Fourier to produce a filtered probability phase distribution, $P_{\text{new}}(\alpha)$, (Figure 32):

$$P_{\text{new}}(\alpha) = P_{\text{O}}(\alpha)P_{\text{C}}(\alpha) .$$

The calculated phase probability distribution, $P_{\text{C}}(\alpha)$ is updated with every cycle of refinement, but the double-phased Fourier phase probability distribution, $P_{\text{O}}(\alpha)$, is not altered.

In the last step in ISIR, the new set of phases from the phase filter is then used to calculate another electron density map which is then substituted for the double-phased Fourier map, and the process is repeated. These iterations continue until convergence is attained. From a practical standpoint, a new density filter need not be calculated every cycle. The density filter changes very little from cycle to cycle; in addition, the calculation takes four to six times longer than the remainder of the calculations. Thus, the density filter is usually only calculated about once every four cycles.

B. The ISIR Phase Refinement of T_1 Using MRS

Structure factors were calculated for the mercury atoms of MRS, and before the ISIR refinement was begun, reflections for which $||F_{\text{d}}| - |F_{\text{n}}||/|F_{\text{hc}}|$ was greater than 3.0 and for which the difference divided by an average structure factor, $||F_{\text{d}}| - |F_{\text{n}}||/((|F_{\text{d}}| + |F_{\text{n}}|)/2)$, was

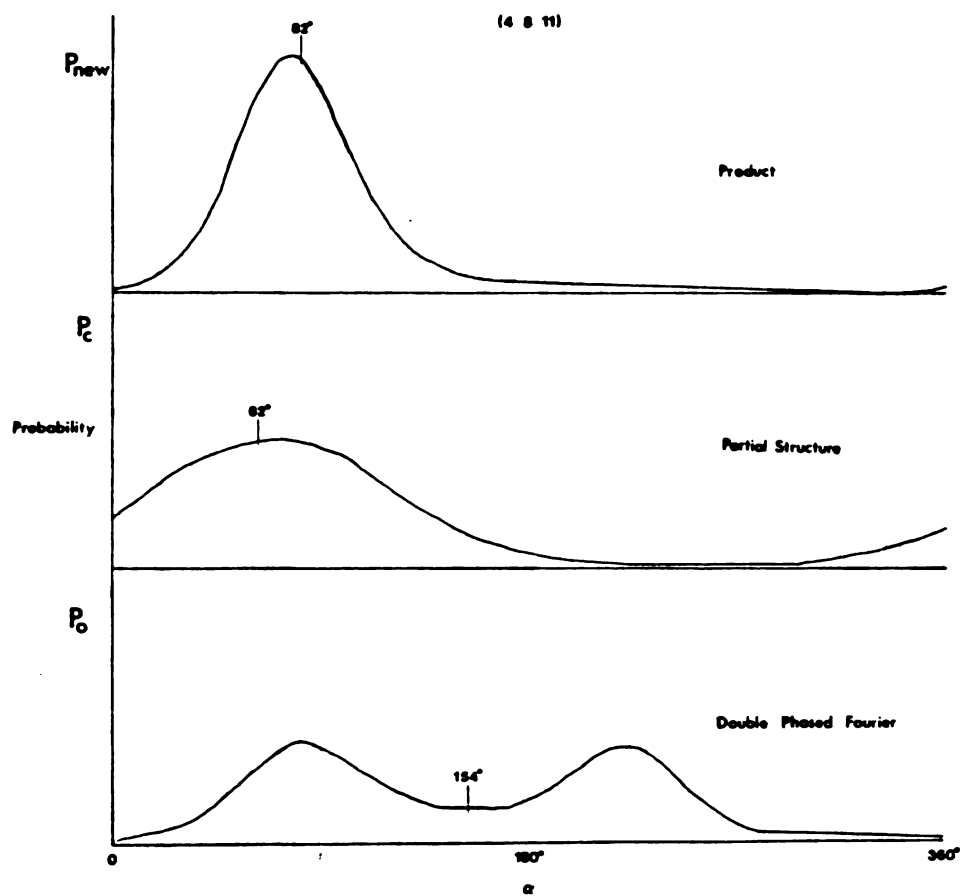


Figure 32: An Example of Reciprocal Space ISIR Phase Filtering.

greater than 0.30 were rejected from the calculation. The first condition rejects reflections for which the agreement between the observed differences and the calculated structure factors is poor. If the heavy atoms have well determined parameters, these reflections are likely to be in error. The second condition rejects reflections which have differences large with respect to the reflection itself. If these contain large structure factors, the latter will tend to dominate and possibly obscure the electron density map. If the reflections are small, it indicates that their error is large. In order for a structure factor to be rejected, it must have both a very large relative difference, and be in poor agreement with the calculated structure factor. Reflections with small differences (less than 20% of the median difference) were also initially left out of the calculation. The latter rejected 1038 of 3368 reflections, and 78 more were rejected by the aforementioned conditions, leaving 2252 reflections (67%) used to determine phases. The rejected structure factors were not used in the phase determination, but phases were assigned to these reflections later in the calculation based on the inverse Fourier transform of the direct space filtered electron density map.

Before the refinement, the Cullis R factor (R_C) using only the centric reflections, was 0.515, and the total R factor (same as above, but includes all 2252

reflections) was 0.509. These values compare favorably to those of proteins with which the ISIR method has been used successfully in the past, with 0.50 for R_c being a typical starting value at 3.5 Å resolution (Ref. 77).

A number of statistical parameters were used to follow the progress of the refinement (Table 11). The figure of merit, m , defined by,

$$m = \frac{\int_0^{2\pi} P(\alpha) e^{i\alpha} d\alpha}{\int_0^{2\pi} P(\alpha) d\alpha}$$

was considered one of the best indicators. For bimodal phase angle distributions, it describes how close the maxima are to each other, and for unimodal distributions it describes the degree of sharpness of the distribution. In both cases it is a measure of how well the phase angle is determined. The m in ISIR is not exactly the same quantity as it is in MIR. The shape of the distribution in ISIR is dependent upon Equation (2), and the factor of 2 in the numerator of that equation, while generally accepted, is entirely arbitrary. Thus, while m is a valuable quantity to monitor the progress of the phase refinement, it is not strictly comparable to m of MIR refinements. Both $\langle m \rangle_{\text{new}}$, which is the average final m , and $\langle m \rangle_{\text{cal}}$, which is the average m of the partial structure map, were monitored.

Another quantity which was monitored but which does not have its usual meaning is the residual, R_{map} , where

$$R_{\text{map}} = \Sigma ||F_o| - |F_i|| / \Sigma |F_o|$$

where $|F_o|$ is the observed structure factor amplitude, and $|F_i|$ is calculated by the inverse Fourier transform of the filtered electron density map. This quantity follows changes made to the structure as it passes through the density filter. Note that if there is no change in the structure, R_{map} is zero. The value of R_{map} is a function of percent solvent, because the solvent areas of the map are not permitted to change; hence, the change in R_{map} holds more significance than the value itself. The correlation coefficient,

$$\text{Cor} = \frac{|F_o| |F_c|}{(|F_o|^2 |F_c|^2)^{1/2}}$$

is another monitor of the agreement between the observed structure factors and those calculated during the ISIR process. Finally, change in the phase angle was an indicator of the convergence of the calculation.

The ISIR method requires some knowledge of the percent mother liquor of the crystal to determine the size of the molecular envelope, so the ambiguity in the number of T_1 molecules per asymmetric unit had to be considered again. If three molecules per asymmetric unit was assumed but there were actually two, then the result

would be that less solvent would be given a constant value. If, however, two molecules were assumed and there were actually three, then not only would the solvent be flattened, but also some of the protein. It was decided that the former was a more conservative risk, and more easily detected if in error.

The assumption was thus made that the crystals are composed of 35% mother liquor (Table 3) and a density filter was calculated. After one cycle of density and phase filtering, $\langle m \rangle$ increased from an initial value of 0.42 to 0.57, and the correlation coefficient was 0.859. After three more cycles of phase refinement with the same density filter, $\langle m \rangle$ had increased to 0.69, and the correlation coefficient to 0.960 (Table 11). Unlike the procedure whereby the mercury positions were refined, none of the heavy atom parameters (coordinates, occupancies, or thermal B factors) were refined during this refinement.

At this point a difference electron density map was calculated between MRS and NAT using the latest phases to search for additional minor occupancy mercury sites. The largest peak in the map was located at the Hg 1 coordinates, with the second largest peak (56% of the Hg 1 peak height) at the Hg 2 coordinates. The next highest peak was less than 10% of the Hg 1 peak, and it was located at a symmetry related position to Hg 1 (same x and z coordinates with $y = 1/8$ instead of $1/4$).

Table 11: Progress of the ISIR Refinements.

<u>Filter</u>	<u>Cycle</u>	<u>$\langle m \rangle_{\text{new}}$</u>	<u>$\langle m \rangle_{\text{cal}}$</u>	<u>R_{map}</u>	<u>Cor</u>	<u>$\Delta\alpha_n$</u>
3.5 Å — 35% Solvent						
1	1	0.57	0.48	0.433	0.863	29.8
1	4	0.69	0.64	0.253	0.960	4.0
2	1	0.57	0.47	0.434	0.862	29.0
2	4	0.69	0.64	0.257	0.958	4.2
3	1	0.57	0.67	0.433	0.863	29.8
3	8	0.72	0.67	0.221	0.969	1.9
3.5 Å — 50% Solvent						
1	1	0.54	0.45	0.507	0.809	34.6
1	4	0.62	0.54	0.390	0.893	12.9
2	1	0.54	0.44	0.508	0.807	35.4
2	4	0.69	0.63	0.299	0.938	4.8
3	1	0.54	0.45	0.507	0.809	34.6
3	8	0.72	0.66	0.263	0.954	2.0
5.0 Å — 50% Solvent						
1	1	0.54	0.45	0.472	0.846	33.9
1	4	0.67	0.61	0.293	0.950	5.2
2	1	0.56	0.46	0.466	0.857	35.1
2	4	0.68	0.62	0.284	0.953	5.2
3	1	0.56	0.46	0.466	0.857	35.1
3	8	0.71	0.65	0.247	0.965	1.6

Even if this is a true heavy atom binding site, the occupancy would only be about 7.5 electrons, too small to significantly affect the phasing of the protein. Once again the two mercury positions were confirmed, and the absence of additional significant mercury sites demonstrated.

The phase angles from the first four cycles were then used to calculate a second density filter. The second density filter had much sharper boundaries between the protein and the solvent. The observed data were again used to calculate a double-phased Fourier map, filtered with the new density filter. The new $\langle m \rangle$ was 0.57, the same as with the first density filter, but the correlation coefficient increased slightly to 0.862. Once again three more cycles of phase refinement were performed using the second density filter. The results were similar to the first four cycles with $\langle m \rangle$ of 0.69 and a correlation of 0.958.

The phase angles from these four cycles were used to calculate yet a third density filter. Once again, the protein-solvent boundaries were somewhat better defined, and small differences were noted among the three filters. The first cycle gave $\langle m \rangle$ of 0.57 and a correlation of 0.863. After eight cycles of phase refinement with this third density filter, $\langle m \rangle$ was 0.72, and the correlation was 0.969. The average phase shift between cycle 7 and cycle 8 was only 1.9° and taken to represent convergence.

An electron density map was calculated using these phases. It was obvious from an inspection of the map that the crystal contains considerably more than 35% mother liquor (Figure 33). This suggested strongly that there are not three protein molecules in the asymmetric unit. Therefore, the entire refinement was repeated using 50% for the solvent content of the crystals. This put tighter constraints on the molecular envelope and resulted in phase angle changes for many reflections. It was also decided to use the entire data set rather than eliminating those reflections with small differences since a small difference, especially between two large structure factors, could be a valid observation. Hence, of the 3368 reflections, only 132 (3.9%) were rejected because their difference was relatively large and did not agree well with calculated value. The remaining 3236 reflections had a $\langle m \rangle$ of 0.32 (considerably lower than the 0.42 of the first data set, however, the present data set had 44% more reflections). The R_c for this data was 0.575, which is somewhat higher than the 0.50 typically desired as a starting value. This value is probably a better indicator of the quality of the derivative, since typically all of the reflections are included in ISIR (Ref. 77).

A new density filter was calculated from the double phased Fourier map. Although the filter was calculated at the 50% level, it was very similar to that contoured at the 35% level. The location of the molecules was obvious and the same in both maps. After a cycle of

ISIR 35% Solvent

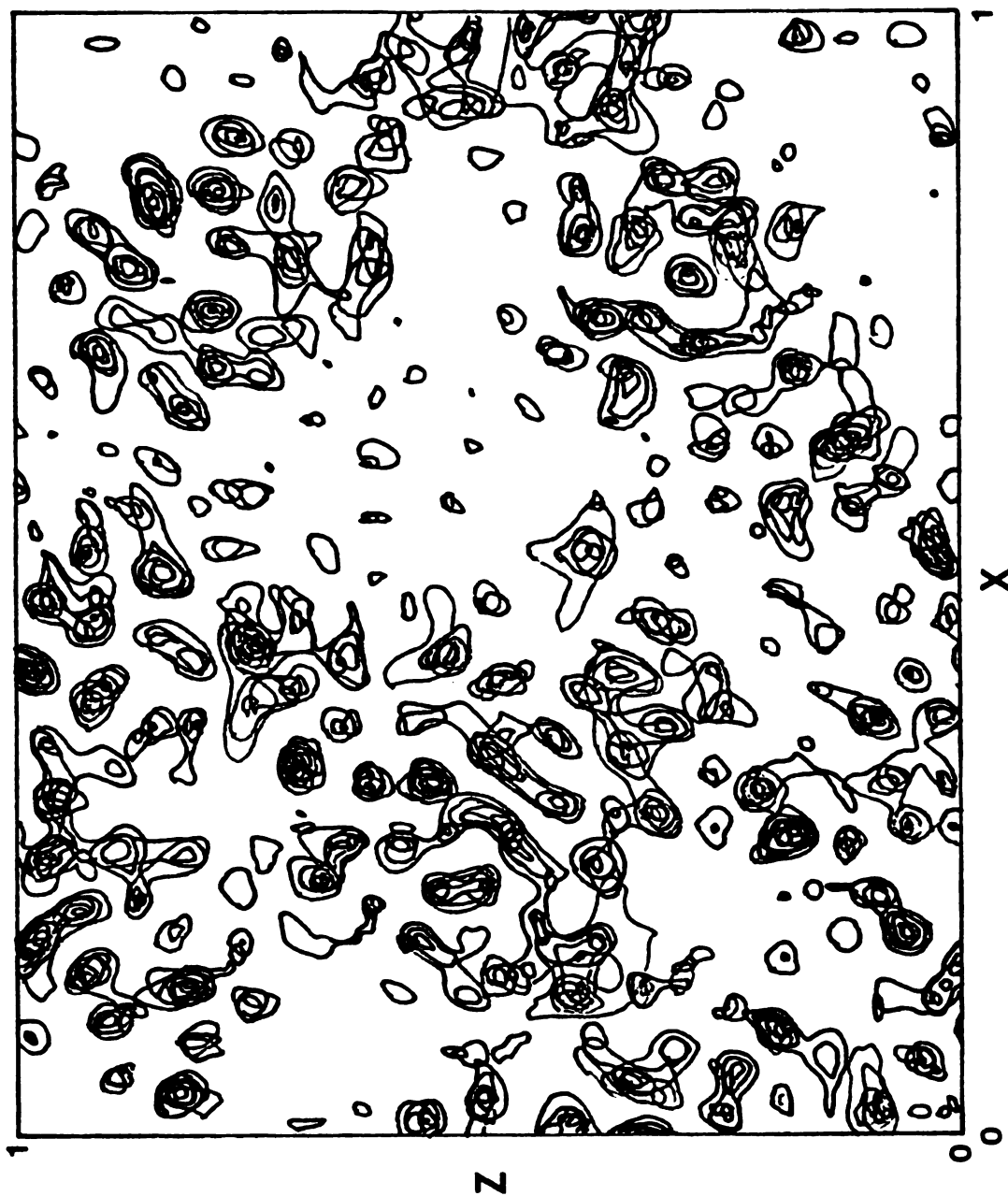


Figure 33: A 2.3 Å Slice Through the Unit Cell Calculated with 35% Solvent (from $y = 0$).

phase refinement, $\langle m \rangle$ was 0.53, compared with 0.57 for the 35% solvent structure, with a correlation of 0.801, compared with 0.859. These indicators were probably worse for the 50% solvent data, because it had more reflections. The extra reflections in the 50% solvent case also were those with small differences and so in general were probably more poorly determined. After three cycles of phase refinement $\langle m \rangle$ was 0.62 with a correlation of 0.893, still considerably inferior to the 35% solvent refinement.

A second density filter was calculated from the phases of the first four cycles and once again, it gave the molecules much sharper boundaries than the first filter. When the double-phased Fourier was filtered through both the density and phase filters, after one cycle $\langle m \rangle$ was 0.54 and the correlation was 0.807. Again three more cycles of phase refinement were calculated resulting in $\langle m \rangle$ of 0.69 and a correlation of 0.938. These compare favorably with the results from the 35% filter which had $\langle m \rangle$ of 0.69 and a correlation of 0.958 so that the second filter apparently improved the phases significantly.

The third density filter was calculated from the phases of the second four cycles and the observed data were again used to calculate a double-phased filtered Fourier. After eight cycles of phase refinement the

final $\langle m \rangle$ was 0.72 (the same as with the 35% solvent calculation) and the correlation was 0.954 (compared with 0.969 for 35% solvent). Since there are 44% more reflections in this second data set and since the solvent level most likely is considerably higher than 35%, more confidence is placed in the 50% solvent refinement (summarized in Table 11).

The distribution of m (Figure 34) reveals that 67% of the structure factors from this refinement have m above 0.70 (implying well determined phases), and only 7% of the structure factors have m below 0.20 (poor phases). It is puzzling to note that the centric reflections have $\langle m \rangle$ considerably worse than all reflections, with only 59% having m greater than 0.70, and 13% having m less than 0.20 (Table 12).

In an effort to verify and possibly obtain a more reliable set of phase angles, the refinement was repeated using only those data at 5.0 Å resolution. In this data set there were 2595 reflections, 78 of which were rejected because they had large differences which did not agree well with calculated values. The starting R_C for this data set was 0.506, however a more typical value for R_C for proteins that ISIR has been successful with at this resolution is 0.45 (Ref. 77). The $\langle m \rangle$ for these data before phase refinement was 0.42.

A density filter was calculated from the double-phased Fourier map as before. While the protein density

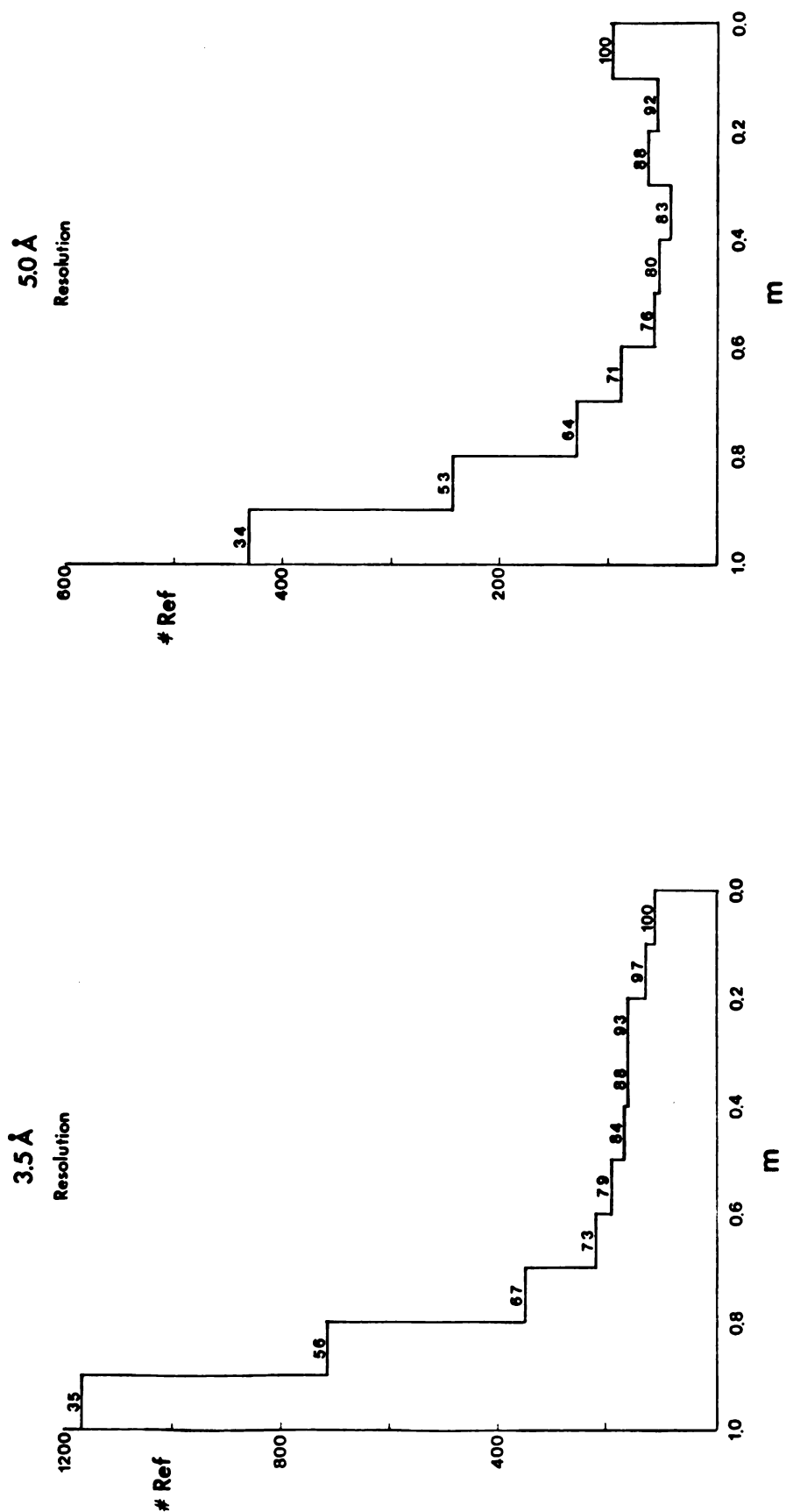


Figure 34: Figure of Merit Distributions for ISIR Refinements. Numbers indicate percent of reflection with m greater than the value.

Table 12: Distribution of Figures of Merit for ISIR Refinements.

$\langle m \rangle_{\text{new}}$	Number of Reflections (Percent of Total)			
	3.5 Å Refinement		5.0 Å Refinement	
	all	centric	all	centric
0-10	109 (3%)	48 (7%)	97 (8%)	45 (13%)
11-20	128 (4%)	44 (6%)	55 (4%)	24 (7%)
21-30	153 (4%)	42 (7%)	58 (5%)	19 (5%)
31-40	154 (4%)	38 (5%)	42 (3%)	15 (4%)
41-50	158 (5%)	28 (4%)	54 (4%)	21 (6%)
51-60	194 (6%)	39 (5%)	64 (5%)	14 (4%)
61-70	221 (7%)	48 (7%)	88 (7%)	20 (6%)
71-80	351 (11%)	71 (10%)	130 (10%)	30 (8%)
81-90	721 (21%)	122 (17%)	243 (19%)	49 (14%)
91-100	1177 (35%)	229 (32%)	433 (34%)	115 (33%)

was located in the same regions in this map as in those calculated with the 3.5 Å data, the shape of the regions was different, with the regions being somewhat "U" shaped instead of globular. After a cycle of phase refinement the $\langle m \rangle$ was 0.54 (compared to 0.53 for the 3.5 Å resolution data) and the correlation was 0.846 (compared to 0.801). This was followed by three cycles of phase refinement leading to $\langle m \rangle$ of 0.67, and a correlation of 0.946, considerably better than the 3.5 Å data (which had $\langle m \rangle$ of 0.62 and a correlation of 0.893) (Table 11).

A second density filter was calculated from the phases obtained from the first four cycles. The shape of the protein regions changed somewhat from the first one, with the center of the "U" filling in to give a more globular shape conforming closer to the 3.5 Å resolution density filter. When the double-phased Fourier map was calculated and filtered through both the density and phase filters, the $\langle m \rangle$ was 0.56 and the correlation 0.846. This is an improvement over both the first density filter and the 3.5 Å resolution calculation. Once again three more cycles of phase refinement were calculated resulting in $\langle m \rangle$ of 0.68 and a correlation of 0.953. The $\langle m \rangle$ at this point was actually slightly lower than the $\langle m \rangle$ of 0.69 for the 3.5 Å resolution data at this point.

The third density filter was calculated from the density map and the observed data were again used to calculate a double-phased Fourier to be filtered through

it. The third filter looked very similar to the second filter, with the protein regions being decidedly globular. After eight cycles of phase refinement the final $\langle m \rangle$ was 0.71, and the final correlation was 0.965 (Table 11).

The $\langle m \rangle$ did not improve over that of the 3.5 Å resolution data (0.72), which is surprising since one would expect the lower angle reflections, which are less sensitive to lack of isomorphism, and generally possess larger intensities, to be better determined than the higher angle reflections. The m distribution (Figure 34) reveals that the 3.5 Å resolution data give a higher percentage of well defined phases than the 5.0 Å resolution data set; moreover, the centric reflections are also more poorly determined than the acentric (Table 12) with only 59% of the centric reflections as opposed to 64% of all reflections, having m greater than 0.70 and 13% of the centric reflections, compared to 7% of all reflections having $\langle m \rangle$ less than 0.20. Since there are 29% more reflections in the 3.5 Å data set, and it extends to higher resolution, and yet statistical indications show the phase angles to be better determined (Table 11), more confidence is associated with the former refinement than the 5.0 Å data set. However, although the 3.5 Å refinement appears statistically to be better, the 5.0 Å refinement is nearly as good. Therefore it was decided to study the two maps in concert, anticipating

that map interpretation would be easier than with either one alone.

V. THE ELECTRON DENSITY

A. The Electron Density Map

"Best" electron density maps (Ref. 6) were calculated for both the 5.0 Å and the 3.5 Å resolution phase sets. The best map uses the centroid of the phase probability distribution for its phase, and weights reflections by their figures of merit. Thus,

$$\rho_{\text{best}} = \frac{1}{V} \sum_{\mathbf{s}} |F_{\mathbf{n}}| e^{i\alpha_{\text{best}}} e^{-2\pi i(\vec{\mathbf{s}} \cdot \vec{\mathbf{r}})}$$

Reflections with m less than 0.5 were omitted from the calculation since their phase angles were considered poorly determined. Inclusion of these reflections would increase the background, but probably would not add much useful information to the electron density map. The $F(0,0,0)$ term used was 44,000 electrons, corresponding to two protein molecules per asymmetric unit (8 per unit cell).

The mean square error in the best electron density map is given by

$$\langle \sigma_{\phi}^2 \rangle = \frac{1}{V^2} \sum_{\mathbf{s}} |F_{\mathbf{n}}|^2 (1 - m^2) \quad (3)$$

The standard error is the square root of Equation (3) which for the 5.0 Å electron density map was 0.0845 electrons/Å³, and for the 3.5 Å electron density map was 0.105 electrons/Å³.

The electron density was calculated in planes 1.02 Å apart normal to the short (37.32 Å) b crystals axis to make visualization of the molecules easier. The maps were contoured starting at 1.0σ, and at intervals of σ. The largest value of the electron density is 1.37 electrons/Å³ (13σ) in the 3.5 Å resolution map.

One asymmetric unit of the 3.5 Å resolution electron density map was transferred to acetate sheets which were sandwiched between one-eighth inch lucite sheets adding a third dimension and producing a minimap. This map may be found in the Appendix. Examination of the map reveals many peaks, 6σ to 9σ high, separated by distances of about 4.0 Å. Since the distance between peptide units is about 3.78 Å, these peaks were interpreted as being the peptides units. Markers were placed at the center of each of the peaks in the 3.5 Å map that were 6σ (0.630 electrons/Å³) or higher. It was assumed that most of these peaks would be due to the peptide units, thus marking the α-carbon chain, but there is not an exact correspondence. Some of 6σ peaks are probably due to sidechain density, and some of the peptides (i.e. Gly) probably have lower densities. There were 220 markers placed in the asymmetric unit, and there are 208

peptides in two T_1 molecules. The coordinates of these peaks were then measured to the nearest 0.5 Å by placing each of the acetate sheets on a grid. Each peak was then labeled with a number related to its coordinates.

An electron density map of the entire unit cell was now calculated and contoured at the 5σ level. This map shows where the peptides are in the unit cell (subject to the aforementioned limitations), and therefore makes it possible to visualize the arrangement of protein molecules within the cell. This map is shown in Figure 35, with the entire unit cell represented in two dimensions by overlaying the third dimension (y) from each section. The location of the dimer, which is obvious in three dimensions, is outlined in this map.

The electron density map of the entire unit cell was calculated and contoured at the 1σ level in such a way that it could be displayed on an Evans and Sutherland PS300 vector graphics display using the graphics program FRODO (Ref. 82). Due to memory limitations of the PS300 (1 Mbyte), and severe flickering on a display of more than 25,000 vectors, the entire unit cell electron density could not be displayed all at once, so cubes of up to 12 Å on a side were used. The vector graphics produces a three-dimensional view of the density which can be rotated, translated, expanded and otherwise manipulated to facilitate interpretation of the density map.

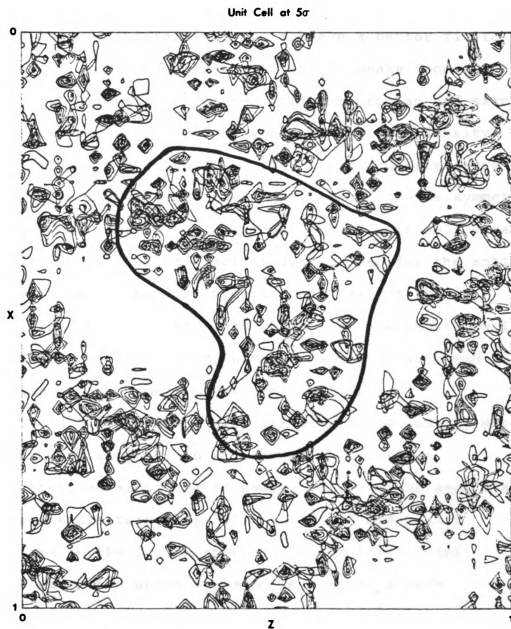


Figure 35: Unit Cell Electron Density Greater than 5 σ . The dimer is outlined.

B. Fitting the Model

Coordinates can also be entered into the FRODO program, and a molecular model may be built. This model can also be superimposed on the density, and the coordinates can be changed interactively to achieve a better fit to the density. The coordinates of the 6σ peaks from the minimap were therefore entered into the program and examined on the graphics system along with the electron density. After it had been established that the coordinates were well within the density, the coordinates from the symmetry related positions in the remainder of the unit cell were calculated. While the electron density map for the entire unit cell cannot be displayed at once, all of the coordinates entered (880) can be displayed without problems. The pattern of these coordinates once again confirmed the position of the dimers within the unit cell.

The 3.5 Å and 5.0 Å resolution electron density maps were studied together in an attempt to determine connectivities among the 6σ peaks. The minimap sections were stacked alternating 3.5 Å and 5.0 Å resolution densities on the lucite sheets. On the first trial, only those peaks which appeared in both maps were considered. In later trials, all of the 3.5 Å peaks were used with the 5.0 Å resolution map providing a guide to connectivities. To facilitate interpretation of the map the coordinates of the peak were listed as far as

possible, in an order such that those adjacent in the sequence of coordinates correspond to those connected by density. When such coordinates are entered into FRODO, lines can be drawn connecting the coordinates in the sequence, eventually forming a trace of the α -carbon backbone of the protein. Then, if some of the peaks can be matched with specific residues in the protein, the remainder of the peaks will be matched also, and the protein folding will be identified.

There are several regions where the path of the α -carbon chain seems unambiguous. It was felt that if secondary structural features could be identified, the folding of the protein might become more apparent. According to the α -carbon chain folding model for T_1 proposed by Heinemann and Saenger (Ref. 63), the protein should contain one α -helix of 4.5 turns (17 residues) and four β -pleated sheets of length 3, 5, 6, and 7 residues which make up a total of 37% of the molecule.

It was decided that the easiest feature to search for would be the α -helix, which would be 24 Å long, and crosses the entire length of the molecule. Helices are characterized by rods of large density at 5.0 Å resolution, and are discernable as helices at 3.5 Å resolution. No such regions were found in the minimap. FRODO was used to update the connectivities which were often difficult to define in the minimap (especially in the y direction

where the density flows from one map section to another). After an extensive search using FRODO, no regions were found that could be likely candidates for large helix, nor even smaller regions of helix that could be aligned and visualized as parts of a larger helix.

Since there are regions in the electron density that appear to have β -sheet structure, the next strategy was to attempt to match them with those proposed for T_1 by length and, to some extent, amino acid composition. Amino acid composition is tenuous at best at 3.5 Å resolution, but if there appears to be large side chains in the density, when the amino acid sequence would indicate a small one (such as a Gly or an Ala), there is cause for doubt. Once again, the density was searched using first the minimap, and then FRODO. While there were several chains of density that were 4 to 6 residues long, there were no regions that could unambiguously be identified as being a particular sequence of amino acids.

The third strategy was to identify the β -sheets by their gross relationship to each other, again using the proposed T_1 model. The model has the three sheets running roughly antiparallel, with the other sheet running approximately normal to the plane of the other three sheets. There are two stretches of density (20 Å long and 18 Å long) that appear to be β -sheets that are approximately perpendicular, but they could not be

unambiguously identified as corresponding to any of those proposed for T_1 by Heinemann and Saenger (Ref. 63).

In summary, there are many areas within the electron density map where the path of the α -carbon chain is quite clear (Appendix, section 13). However, there are also many places where the chain direction is quite ambiguous, and several of the 6σ peaks which at the 1σ level are not connected to any other electron density and should be ignored (Appendix; $y=7$, $x=29$, $z=6$). The quality of the map does not appear to be good enough to trace the chain unambiguously, nor to identify secondary structures of the T_1 model with confidence.

VI. CONCLUDING REMARKS

Protein x-ray crystallography, for all its aforementioned progress, is still not a routine procedure. Many protein crystallographic projects have failed somewhere along the process. Whether this project, or any other, is to result in a correct structure, cannot be known with certainty until the final electron density maps are interpreted. Even then, a chemically sensible structure does not always guarantee a correct one. There were indications early on that this structure would present difficulties. A great many heavy atom derivatives were examined, but with only one positive result. Indications of specific heavy atom binding, such as large intensity

changes, promised good derivatives, but the difference Patterson maps proved differently. And then, the difference Patterson map of the one good derivative appears to be clean and of high occupancy, yet the phase angles produced by it, as evidenced by the final electron density maps, appear to be poor. Using single isomorphous replacement (as opposed to MIR which utilizes much more data) probably was a contributing cause for this.

The structure of T_1 from these crystals probably will be solved, and hopefully the present work chronicles temporary setbacks only. A continued heavy atom search will probably yield other heavy atom derivatives, allowing MIR phase refinement. With the available data, other variations of SIR (Ref. 83), and single isomorphous anomalous scattering could be attempted (Ref. 84).

Another tactic would be to assume the T_1 model is correct, and use the molecular replacement method to assign phase angles (Ref. 85). The molecular replacement method assumes a model structure (in this case, the guanosine 2'-monophosphate derivative of T_1), and then rotates and translates the Patterson function of it (which requires no phases) into the Patterson function of the protein to be determined. Phases are then calculated from the model in its best fit of the relative orientations between the two molecules to calculate a starting structure. Unfortunately, this does not permit an independent check on the validity of the model. Complications would also

arise due to the oligomeric nature of the present crystal form.

Once phase angle assignments are made for reflections up to medium resolution (about 2.5 Å), other methods are available to extend the phases (Ref. 86) and utilize the data to as high a resolution as the crystals scatter (at least 1.5 Å), and thus to realize many or most of the goals of the project.

LIST OF REFERENCES

LIST OF REFERENCES

1. A.C.T. North, D.C. Phillips, F.S. Mathews (1968) Acta Cryst. A24, 351.
2. Data Collection Operation Manual (1982), Nicolet XRD Corp., Section E.15, 1.
3. A.L. Patterson (1934) Phys. Rev. 46, 372.
4. C. Bokhoven, J.C. Schoone, and J.M. Bijvoet (1951) Acta Cryst. 4, 275.
5. B.C. Wang (1983) Lecture Notes for School on Direct Methods and Macromolecular Crystallography, Buffalo, NY.
6. D.M. Blow and F.H. Crick (1959) Acta Cryst. 12, 794.
7. W.N. Lipscomb, G.N. Reeke, J.A. Hartsuck, and F.A. Quijcho (1970) Phil. Trans. Roy. Soc. Lond. B257, 177.
8. J.R. Knowles (1970) Phil. Trans Roy. Soc. Lond. B257, 135.
9. J. Drenth, J.N. Jansonius, R. Koekoek, L.A.A. Sluyterman, and B.G. Wolthers (1970) Phil. Trans. Roy. Soc. Lond. B257, 231.
10. B.S. Hartley (1970) Phil. Trans. Roy. Soc. Lond. B257, 77.
11. J.J. Birktoft, B.W. Mathews, and D.M. Blow (1969) Biochem. Biophys. Res. Commun. 36, 131.
12. R.M. Stroud, L.M. Kay, R.E. Dickerson (1974) J. Mol. Biol. 83, 185.
13. D.M. Shotton and H.C. Watson (1970) Phil. Trans. Roy. Soc. Lond. B257, 111.
14. D.M. Blow, J.J. Birktoft, and B.S. Hartley (1970) Nature 221, 146.

15. A. Ruhlmann, D. Kukla, P. Schwager, K. Bartels, and R. Huber (1973) J. Mol. Biol. 77, 417.
16. A.A. Kossiakoff and S.A. Spencer (1980) Nature 288, 414.
17. M. Laskowski and I. Kato (1980) Ann. Rev. Biochem. 49, 593.
18. F. Pochon, V. Favandon, and M. Tourbez-Perrin (1980) J. Biol. Chem. 256, 547.
19. E. Papamokos, E. Weber, W. Bode, R. Huber, M.W. Empie, I. Kato, and M. Laskowski (1982) J. Mol. Biol. 158, 515.
20. J. Deisenhofer and W. Steigemann (1974) Bayer-Symposium V, "Proteinase Inhibitors", Springer-Verlag, 484.
21. D.M. Blow, J. Janin, and R.M. Sweet (1974) Nature 249, 54.
22. Y. Mitsui, Y. Satow, Y. Watanabe, and Y. Iitaka (1979) J. Mol. Biol. 131, 697.
23. R.M. Sweet, H.T. Wright, J. Janin, C.H. Chothia, and D.M. Blow (1974) Biochem. 13, 4212.
24. S. Hirono, K. Torii, Y. Iitaka, and Y. Mitsui (1979) J. Mol. Biol. 131, 855.
25. C.H. Wei, S.P. Basu, and J.R. Einstein (1979) J. Biol. Chem. 254, 4892.
26. J.R. Gaier, A. Tulinsky, and I.E. Liener (1981) J. Biol. Chem. 256, 11417.
27. M.D. Jibson, Y. Birk, and T.A. Bewley (1981) Int. J. Pept. Prot. Res. 18, 26.
28. G. Robillard and R.G. Shulman (1974) J. Mol. Biol. 86, 519.
29. D.E. Bowman (1946) Proc. Soc. Exp. Biol. Med. 63, 547.
30. Y. Birk (1961) Biochim. Biophys. Acta 147, 378.
31. Y. Birk, A. Gertler, and F. Khalef (1967) Biochim. Biophys. Acta 147, 402.
32. V. Frattali (1969) J. Biol. Chem. 244, 274.

33. S. Odani, T. Koide, and T. Ikenaka (1971) Proc. Jpn. Acad. 47, 621.
34. S. Odani and T. Ikenaka (1973) J. Biochem. (Tokyo) 74, 697.
35. D. Seidl and I.E. Liener (1971) Biochim. Biophys. Acta 251, 83.
36. D. Seidl and I.E. Liener (1972) Biochim. Biophys. Acta 258, 303.
37. S. Odani and T. Ikenaka (1972) J. Biochem. (Tokyo) 71, 839.
38. N. Nishino, H. Aoyagi, T. Kato, and N. Izumiya (1977) J. Biochem. (Tokyo) 82, 901.
39. S. Terada, K. Sato, T. Kato, and N. Izumiya (1978) FEBS. Lett. 90, 89.
40. J.M. Hogle and I.E. Liener (1973) Can. J. Biochem. 51, 1014.
41. J.B. Harry and R.F. Steiner (1969) Biochem. 8, 5060.
42. M.N. Liebman (1977) Ph.D. Thesis, Michigan State University.
43. A. McPherson (1977) Meth. Biochem. Anal. 23, 249.
44. D. Seidl and I.E. Liener (1972) J. Biol. Chem. 11, 3533.
45. K.C. Aune and S.N. Timasheff (1971) Biochem. 10, 1609.
46. T.L. Blundel and L.N. Johnson, "Protein Crystallography", Academic Press, London (1976).
47. K. Takahashi, "Proteins: Structure and Function", Vol. 1, Halsted Press, Tokyo (1972).
48. K. Sato and F. Egami (1957) J. Biochem. (Tokyo) 44, 753.
49. K. Takahashi (1962) J. Biochem. (Tokyo) 51, 95.
50. K. Takahashi (1965) J. Biol. Chem. 240, 4117.
51. K. Takahashi (1961) J. Biochem. (Tokyo) 49, 1.
52. F.G. Waltz and B. Terenna (1976) Biochem. 15, 2837.

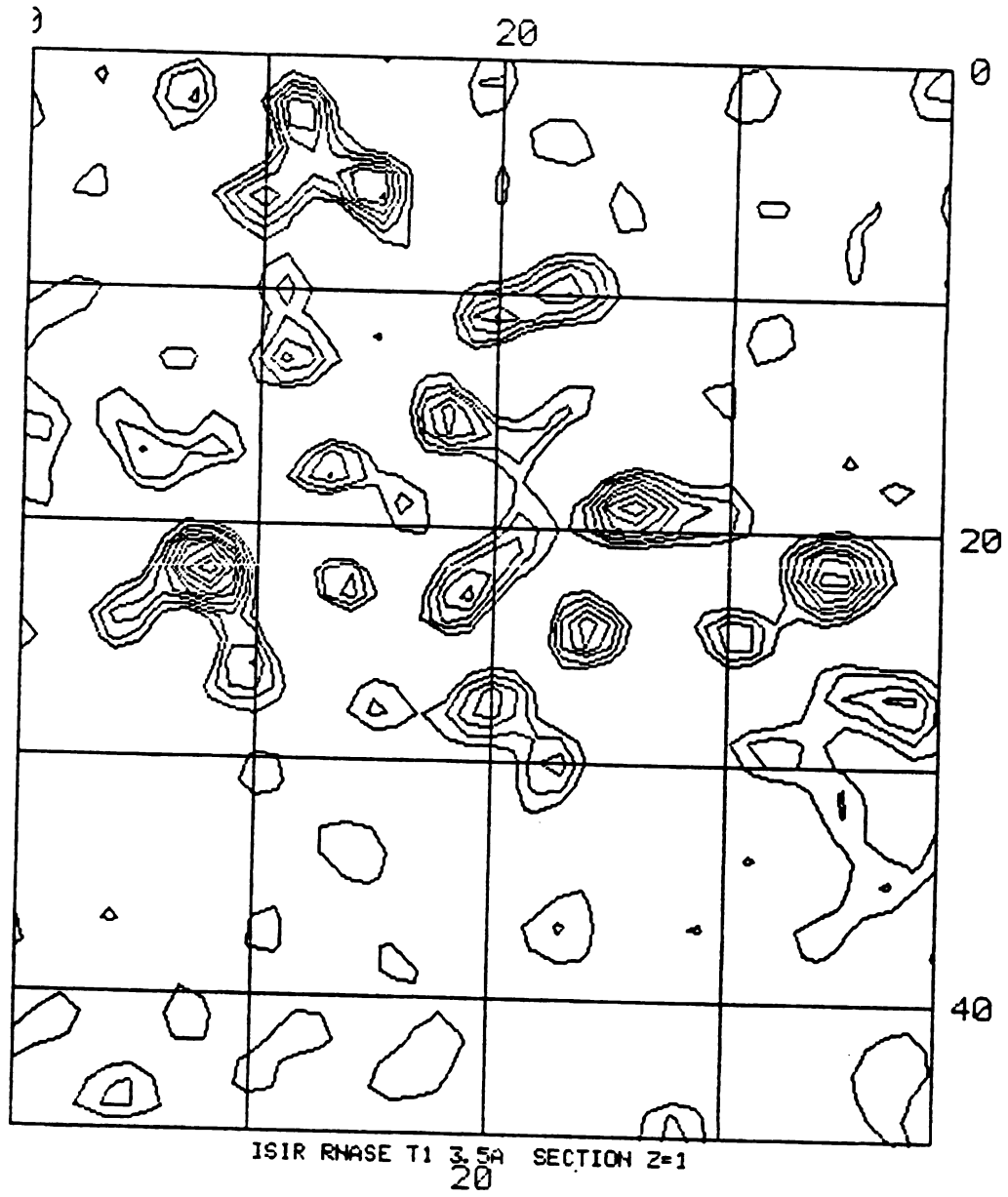
53. F.G. Waltz, H.L. Osterman, and C. Libertin (1979) Arch. Biochem. Biophys. 195, 95.
54. H.L. Osterman and F.G. Waltz (1979) Biochem. 18, 1984.
55. T. Oshima and K. Imahori: "Proteins, Structure and Function", Vol. 1, 333, Halsted Press, Tokyo (1972).
56. D.G. Anderson, G.G. Hammes, and F.G. Waltz (1968) Biochem. 7, 1637.
57. F.G. Waltz (private communication).
58. F.G. Waltz (1976) Biochem. 15, 4446.
59. J.R. Li and F.G. Waltz (1974) Arch. Bioch. Biophys. 161, 227.
60. F.G. Waltz and L.L. Hooverman (1973) Biochem. 12, 4846.
61. F. Eckstein, H.H. Schulz, H. Ruterjans, W. Haar, and W. Maurer (1972) Biochem. 11, 3507.
62. D.A. Usher, D.I. Richardson, F. Eckstein (1970), Nature 228, 663.
63. U. Heinemann and W. Saenger (1982) Nature 299, 27.
64. K.T. Nakamura, K. Iwahashi, Y. Yamamoto, Y. Iitaka, N. Yoshida, and Y. Mitsui (1982) Nature 299, 564.
65. Y. Mauguen, R.W. Hartley, E.J. Dodson, G.G. Dodson, G. Bricogne, C. Chothia, and A. Jack (1982) Nature 297, 162.
66. M. Levitt and C. Chothia (1976) Nature 261, 552.
67. F.G. Waltz and L.L. Hooverman (1973) Biochem. 12, 4846.
68. P.D. Martin, A. Tulinsky, and F.G. Waltz (1980) J. Mol. Biol. 136, 95.
69. C. Thaller, L.H. Weaver, G. Eichele, E. Wilson, R. Karlsson, J.N. Jansonius (1981) J. Mol. Biol. 147, 465.
70. S. Sato and F. Egami (1965) Biochem. Z. 342, 437.

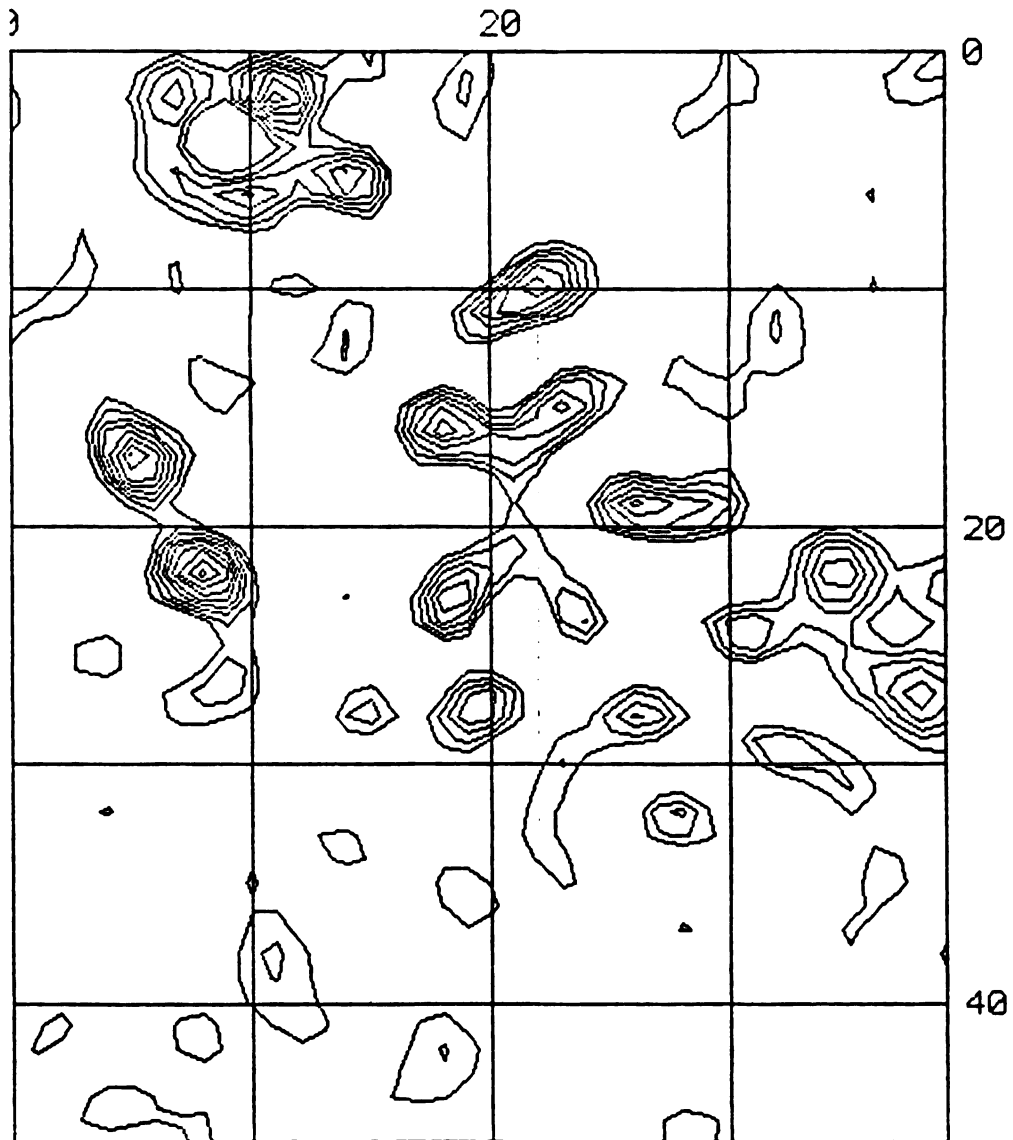
71. M.J. Burger, "Vector Space and Its Application in Crystal-Structure Determination", John Wiley & Sons, New York (1959).
72. D. Harker (1936) J. Chem. Phys. 4, 381.
73. G.M. Sheldrick (1976) SHELX76: A Program for Crystal Structure Determination, University of Cambridge, England.
74. A.F. Cullis, H. Muirhead, M.F. Perutz, M.G. Rossmann, and A.T.C. North (1961) Proc. Roy. Soc. Lond. A265, 15.
75. D.M. Blow and M.G. Rossmann (1961) Acta Cryst. 14, 1195.
76. I.A. Wilson, J.J. Skehel and D.C. Wiley (1981) Nature 289, 366.
77. B.C. Wang (private communication).
78. M.G. Rossmann and D.M. Blow (1961) Acta Cryst. 14, 641.
79. G. Kartha (1961) Acta Cryst. 14, 680.
80. G. Bricogne (1976) Acta Cryst. A32, 832.
81. G.A. Sim (1959) Acta Cryst. 12, 813.
82. T.A. Jones, "Computational Crystallography", Clarendon Press, Oxford, 303 (1982).
83. A. Tulinsky (1964) J. Am. Chem. Soc. 86, 5368.
84. R. Pepinsky and Y. Okaya (1956) Proc. Nat'l. Acad. Sci. USA 42, 286.
85. M.G. Rossmann, "The Molecular Replacement Method", Gordon and Breach, Science Publishers, Inc., New York (1972).
86. N.V. Raghavan and A. Tulinsky (1979) Acta Cryst. B35, 1776.

APPENDIX

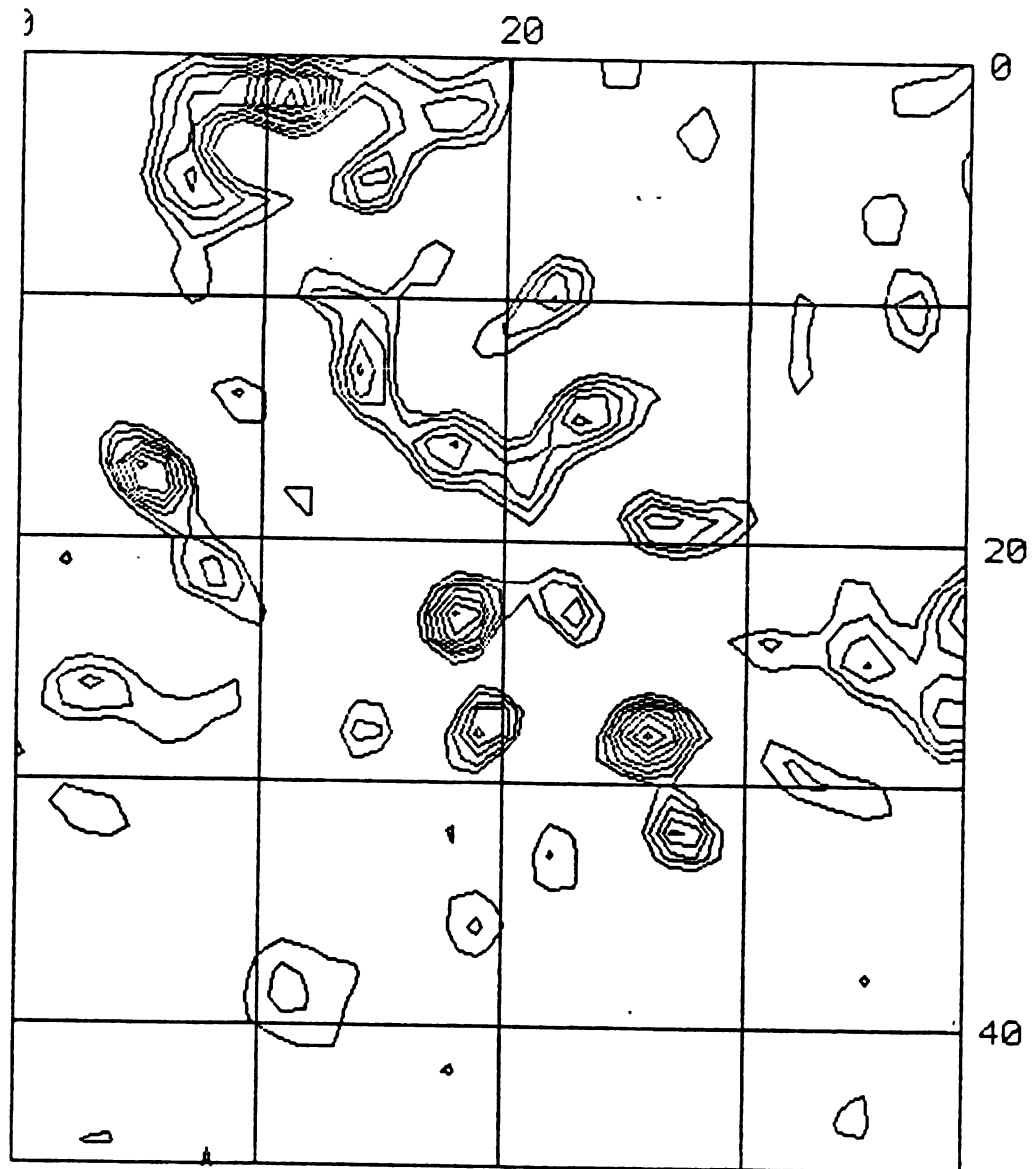
APPENDIX

The following pages show the electron density map calculated for Ribonuclease T₁ using the phases calculated from the ISIR method using the MRS derivative, at 3.5 Å resolution, assuming a solvent content of 50%. The map is contoured starting at σ (0.105 e/Å³) and in increments of σ . The x axis is drawn along the ordinate, and the z axis is drawn along the abscissa. Despite the labeling on the map, each section increments 1.02 Å along the y axis. Units shown are in Angstroms.

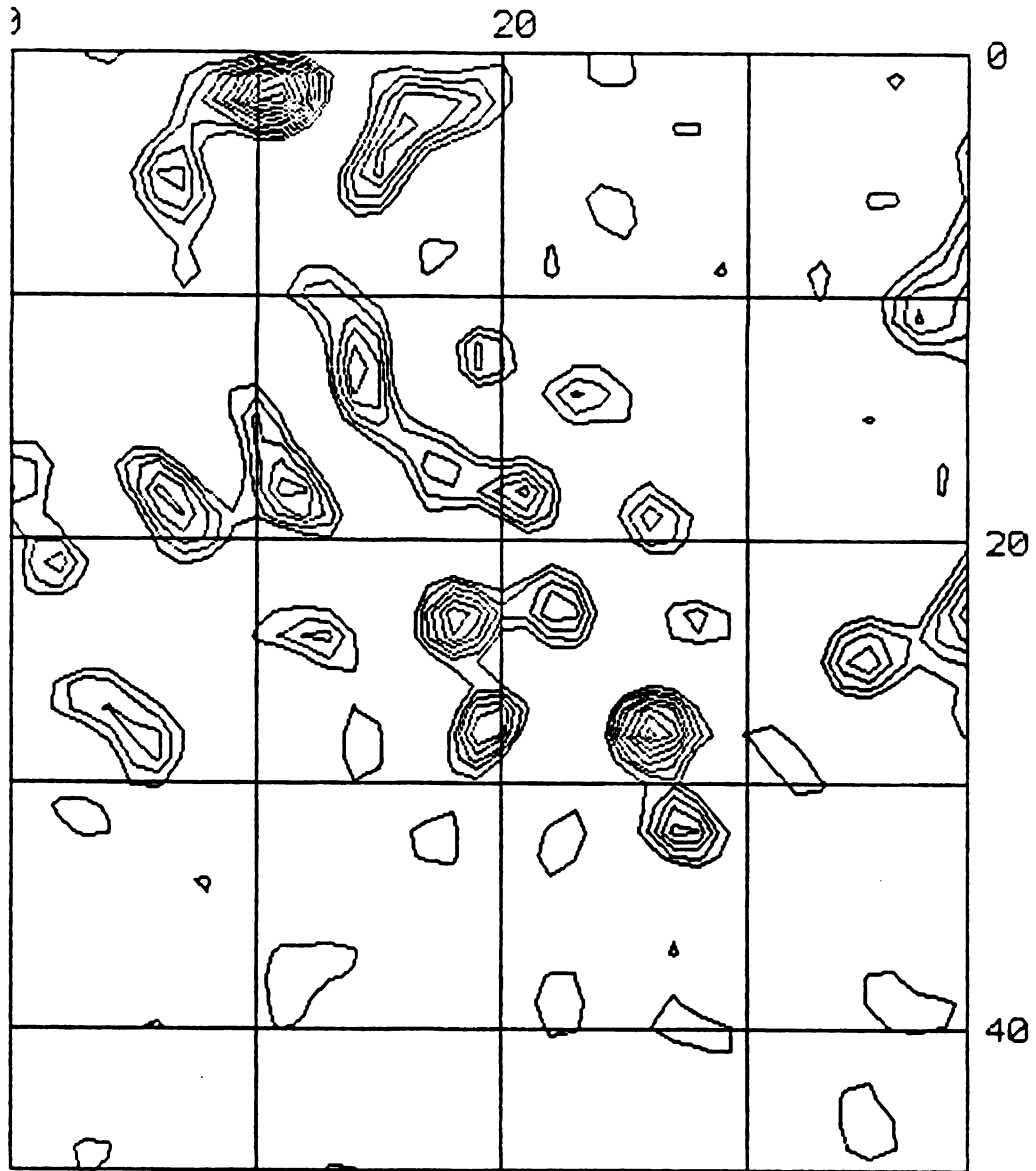




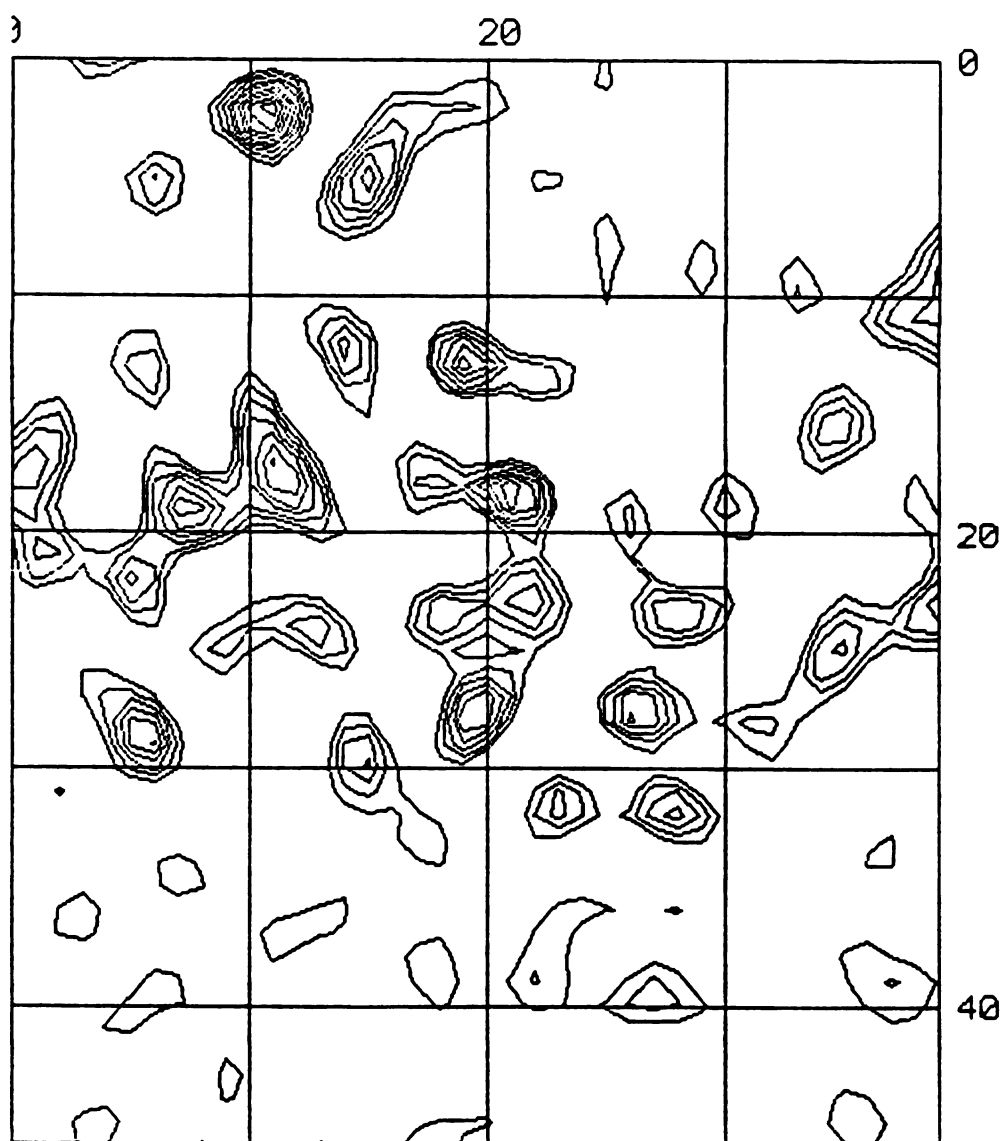
ISIR RHASE T1 3.5A SECTION 2=2
20



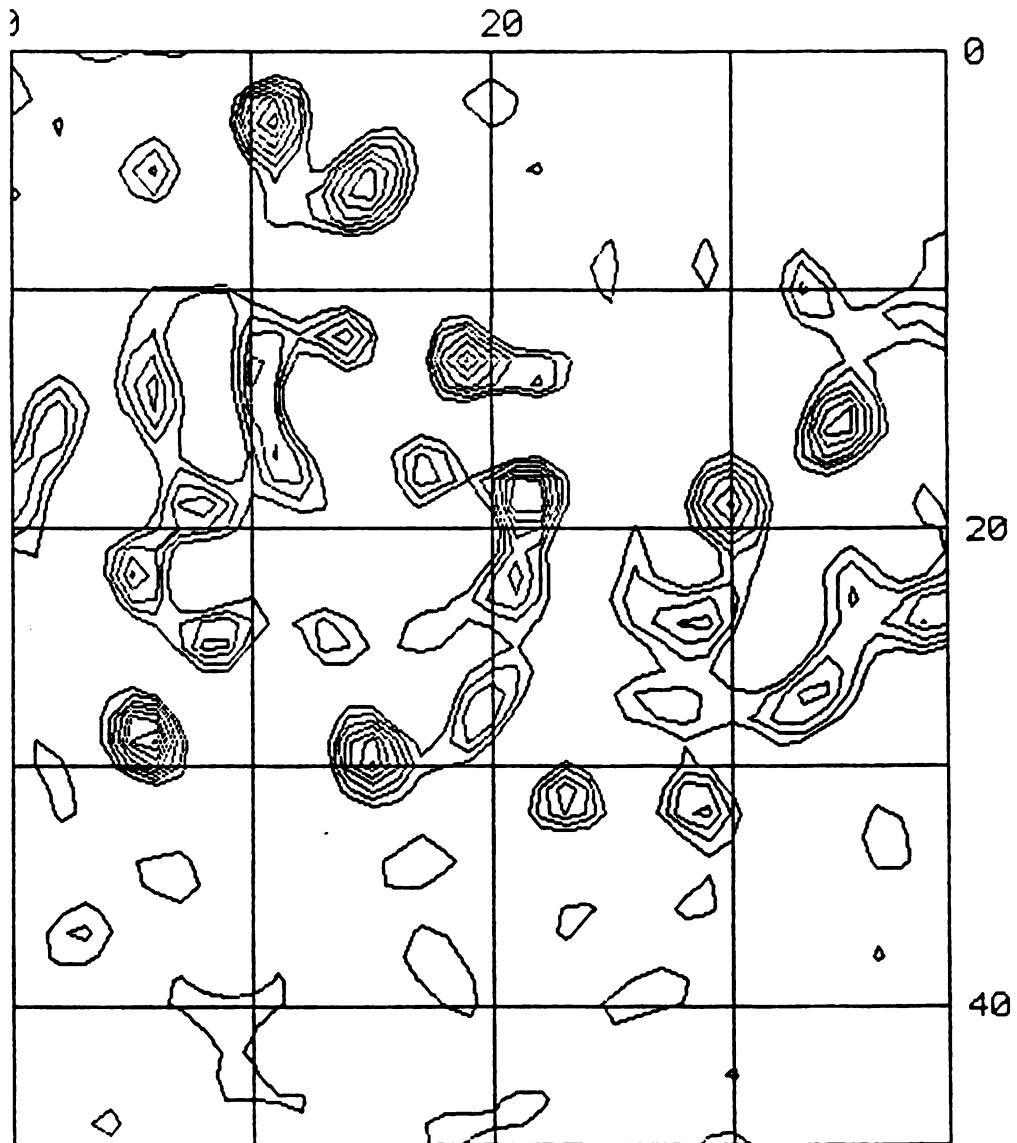
ISIR RMSE T1 3.54 SECTION Z=3
20



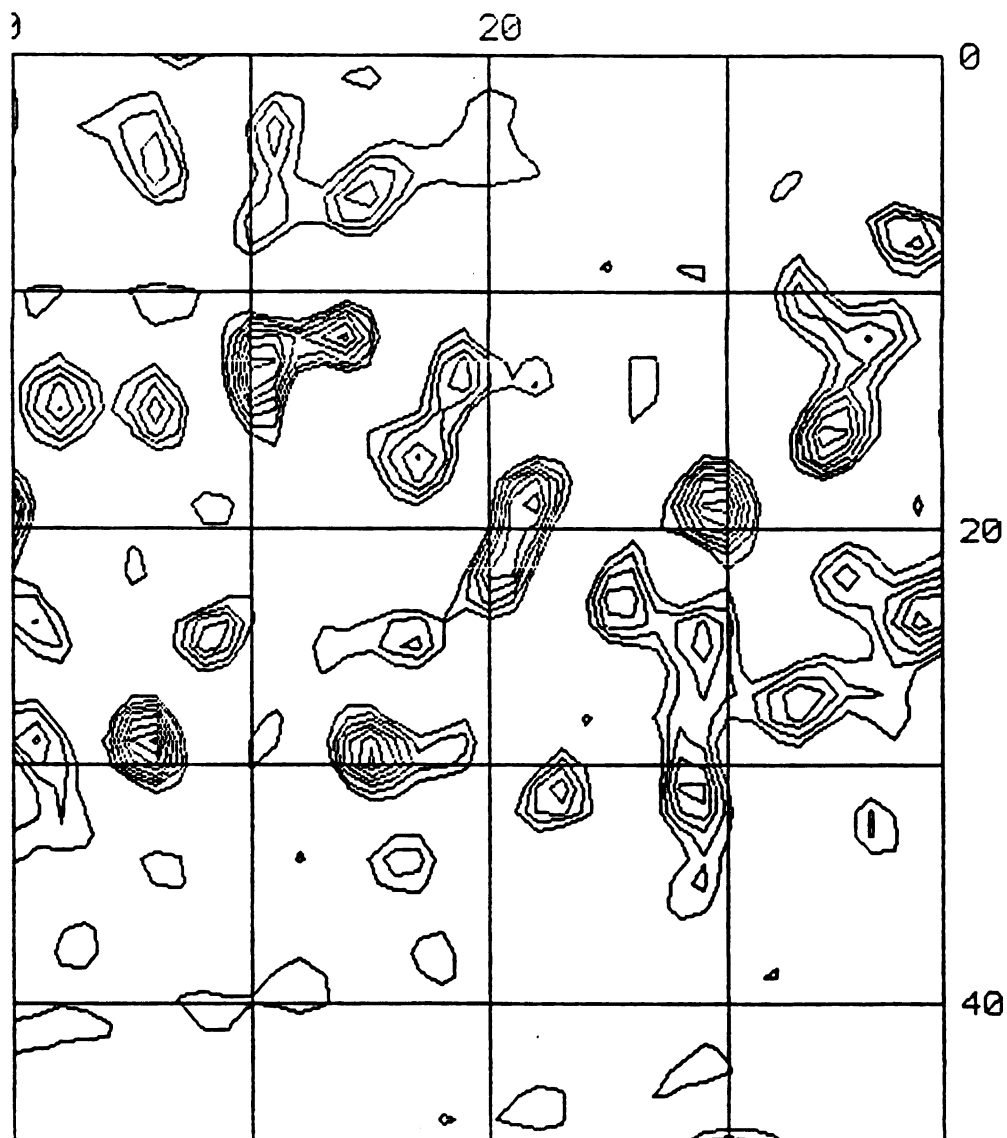
ISIR RNASE T1 3.5A SECTION Z=4
20



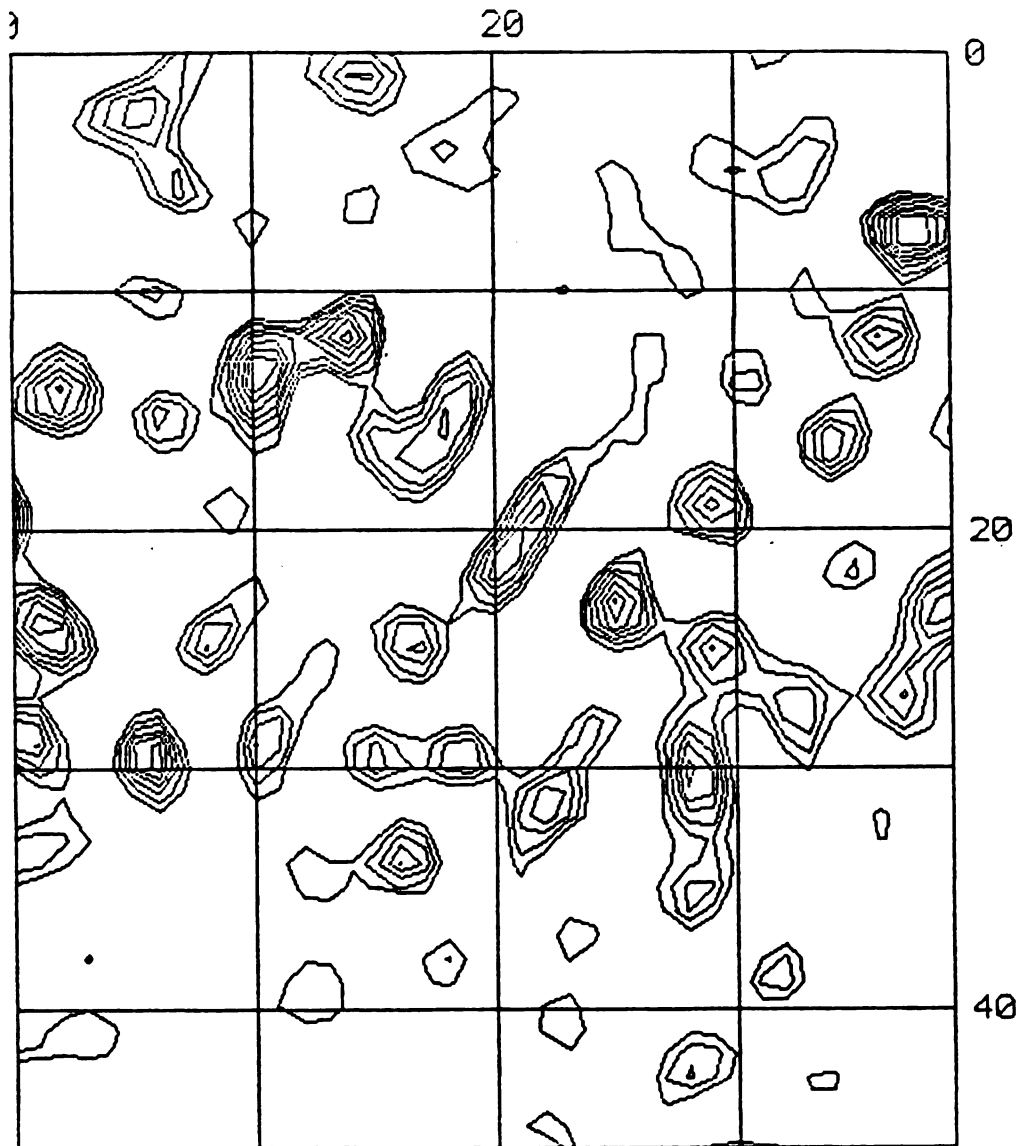
ISIR RNASE T1 3.5A SECTION 2-5
20



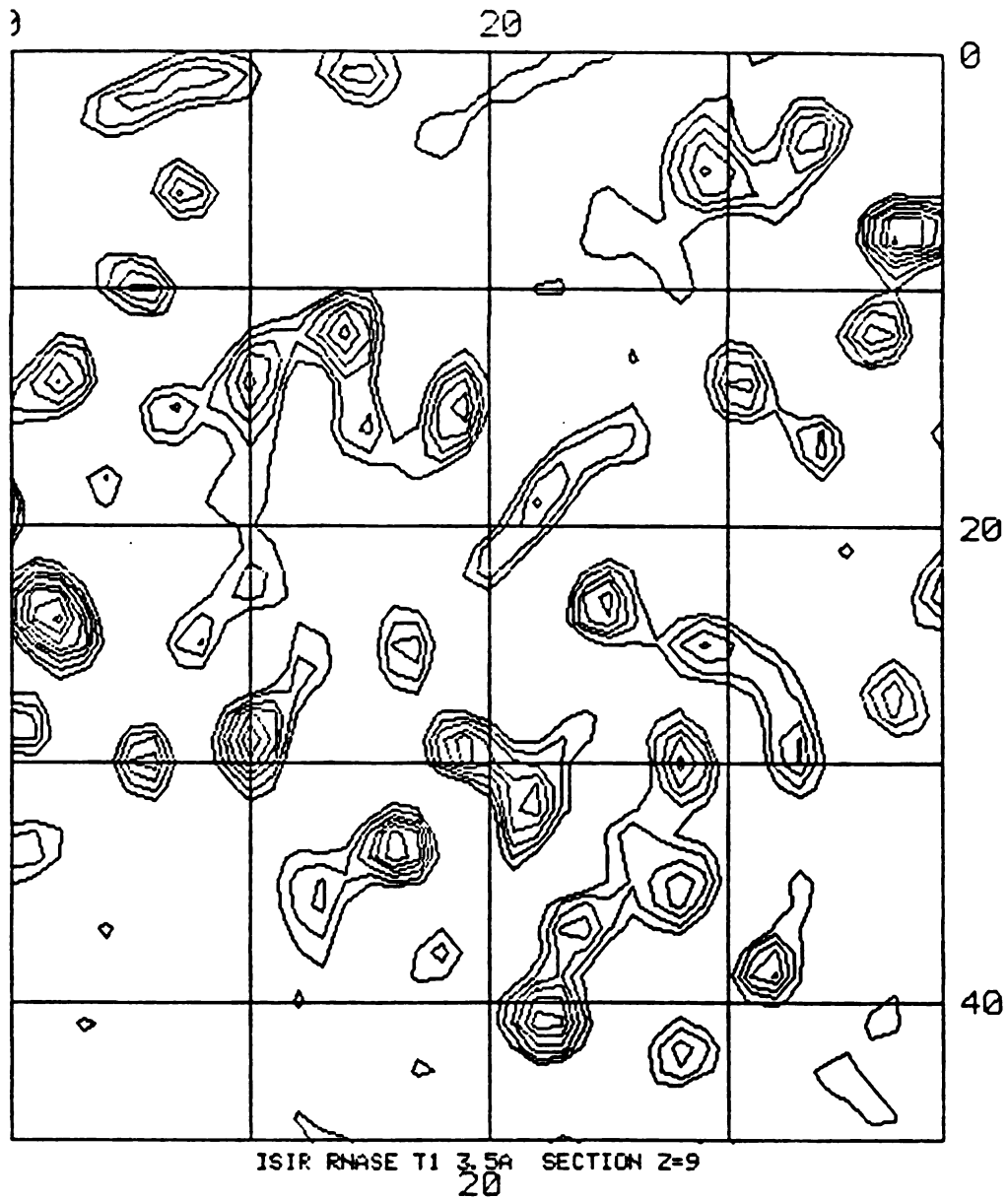
ISIR RMSE T1 3.5A SECTION Z=6
20

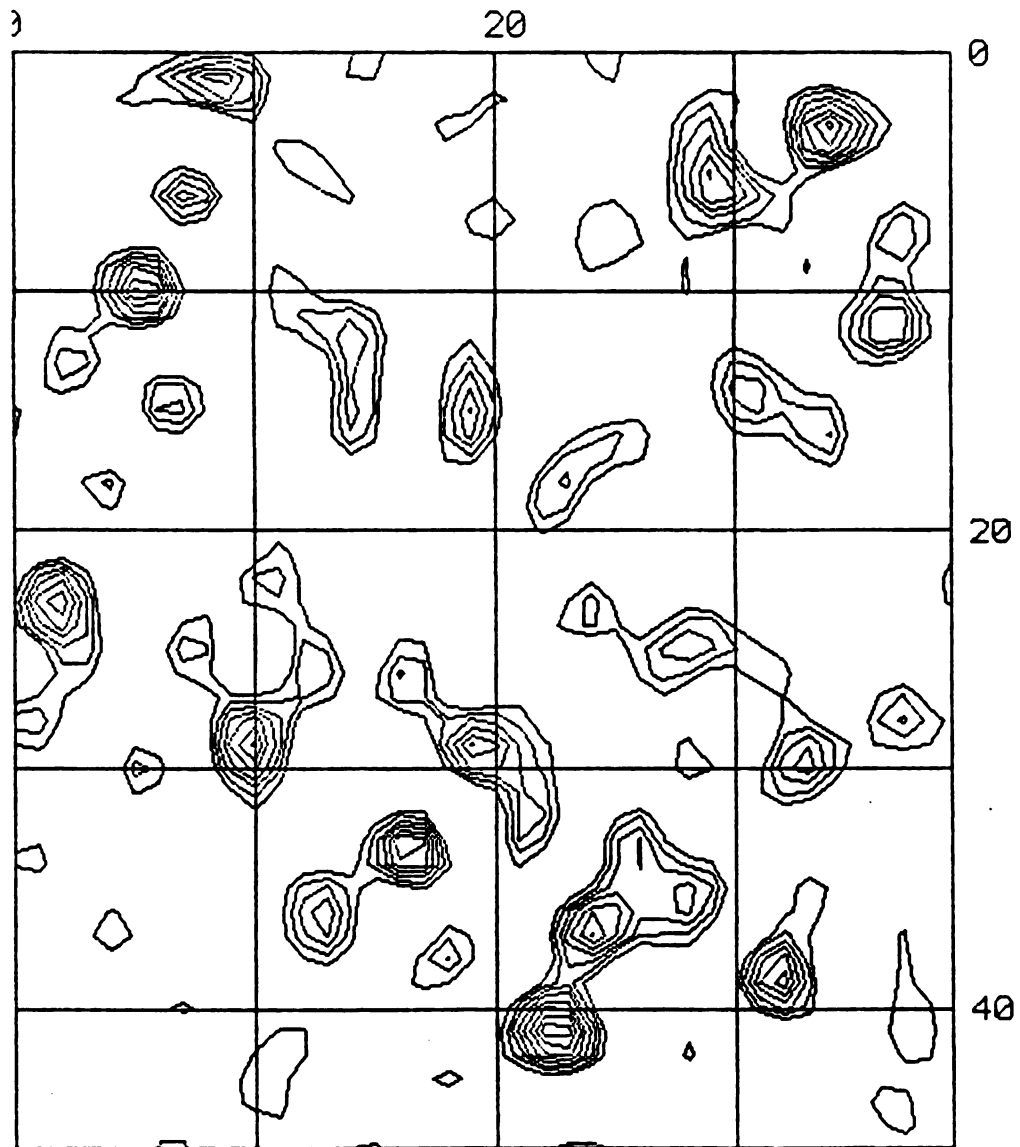


ISIR RNASE T1 3.5A SECTION Z=7
20

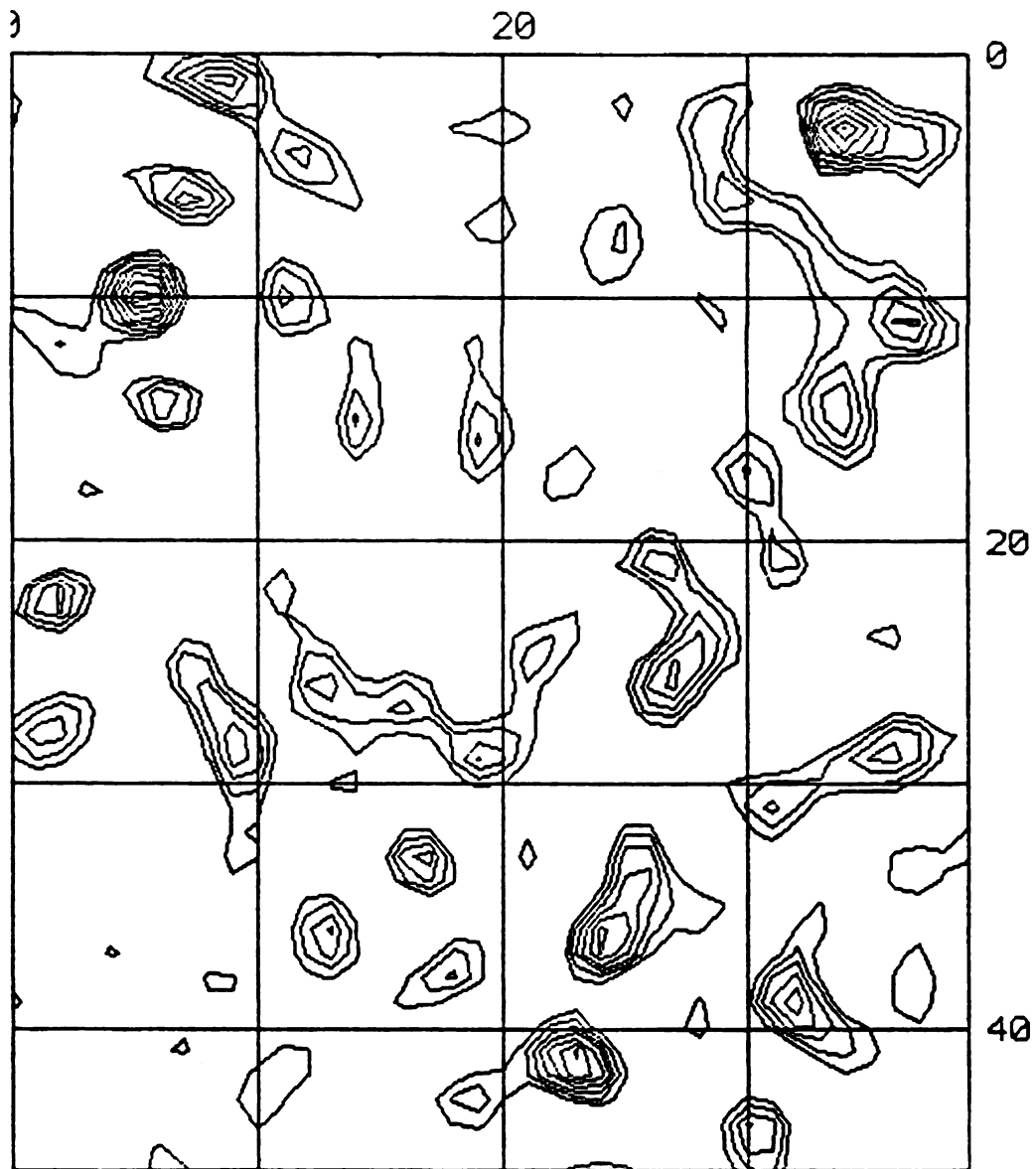


ISIR RHASE T1 3.5A SECTION Z=8
20

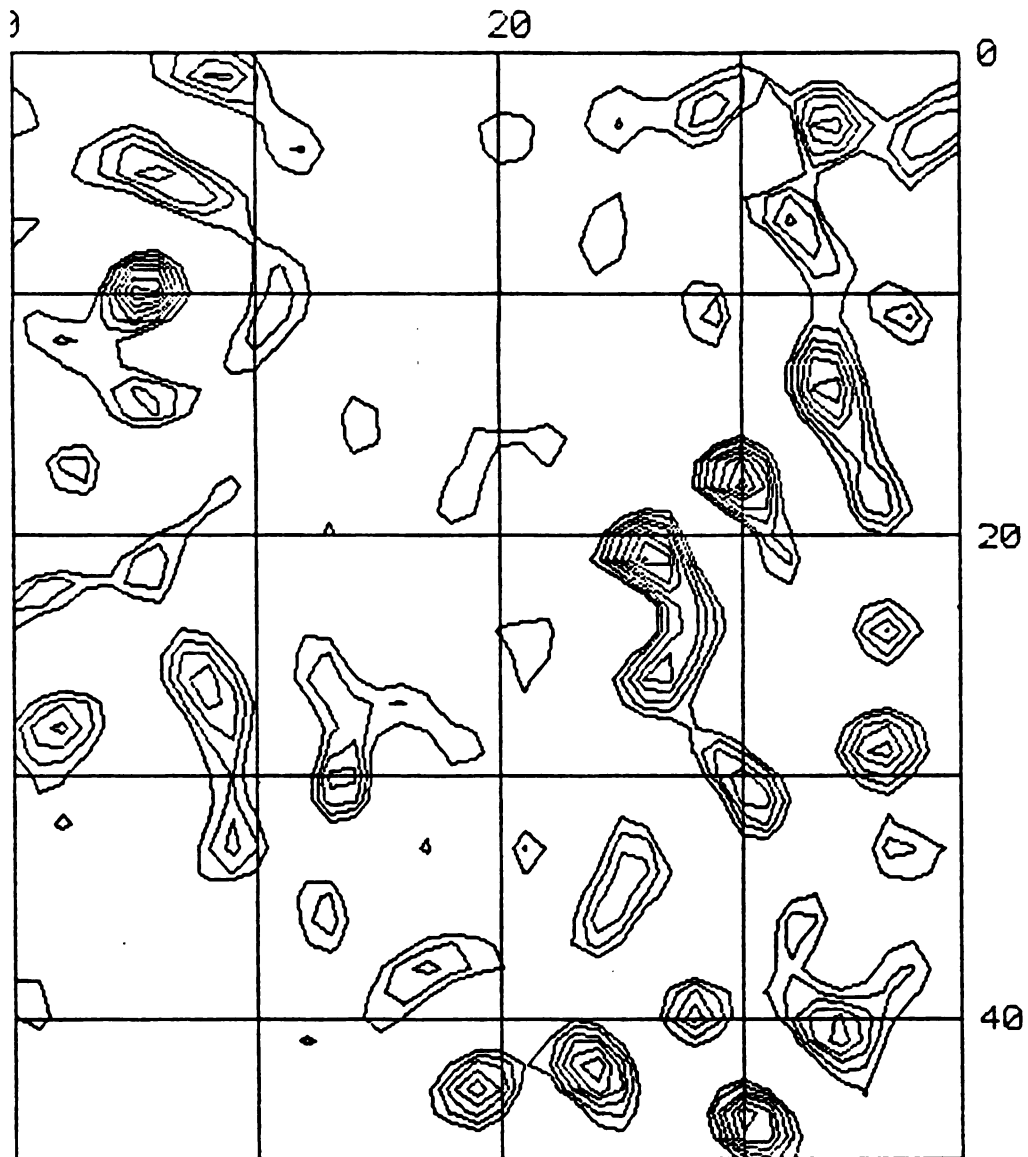




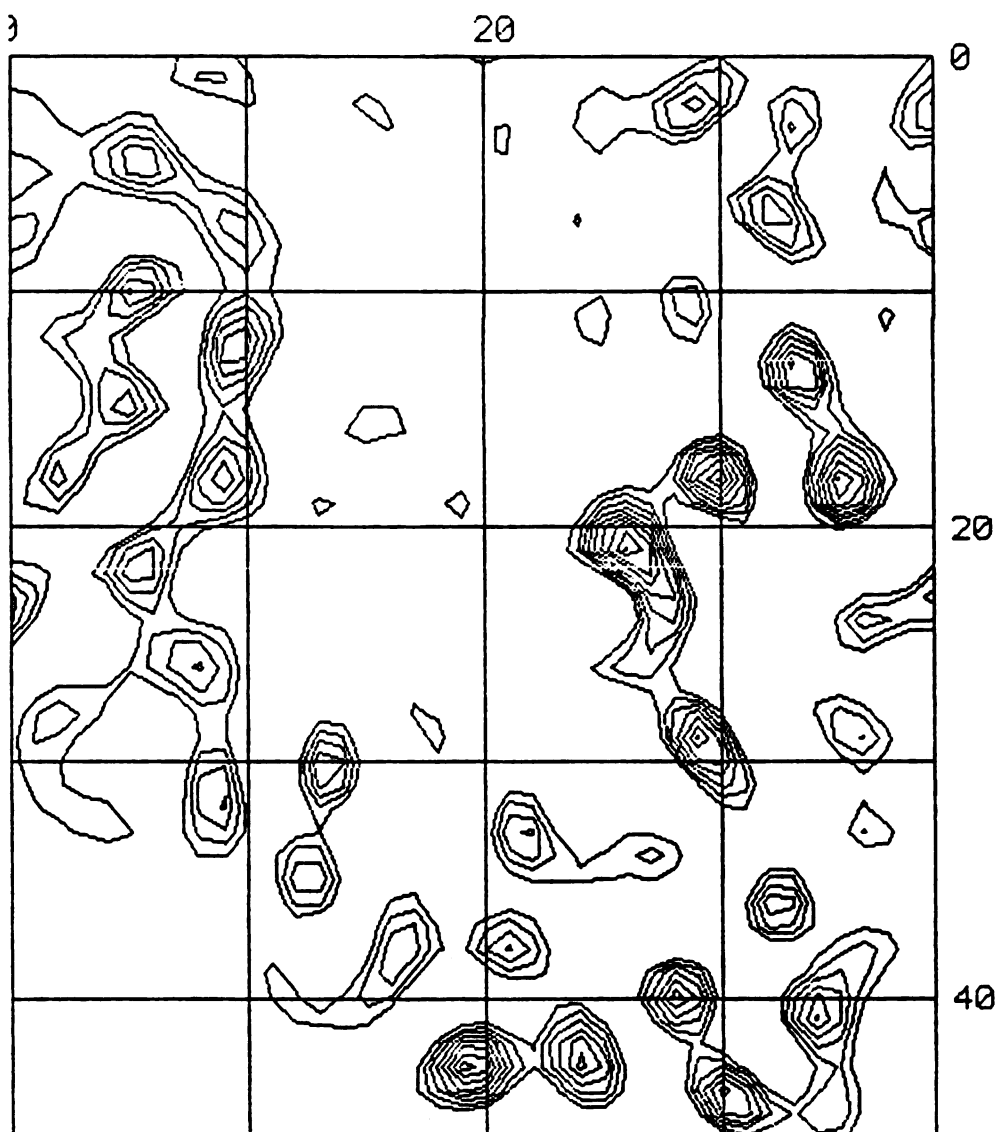
ISIR RNASE T1 3.5A SECTION Z=10
20



ISIR RNASE T1 3.5A SECTION Z=11
20

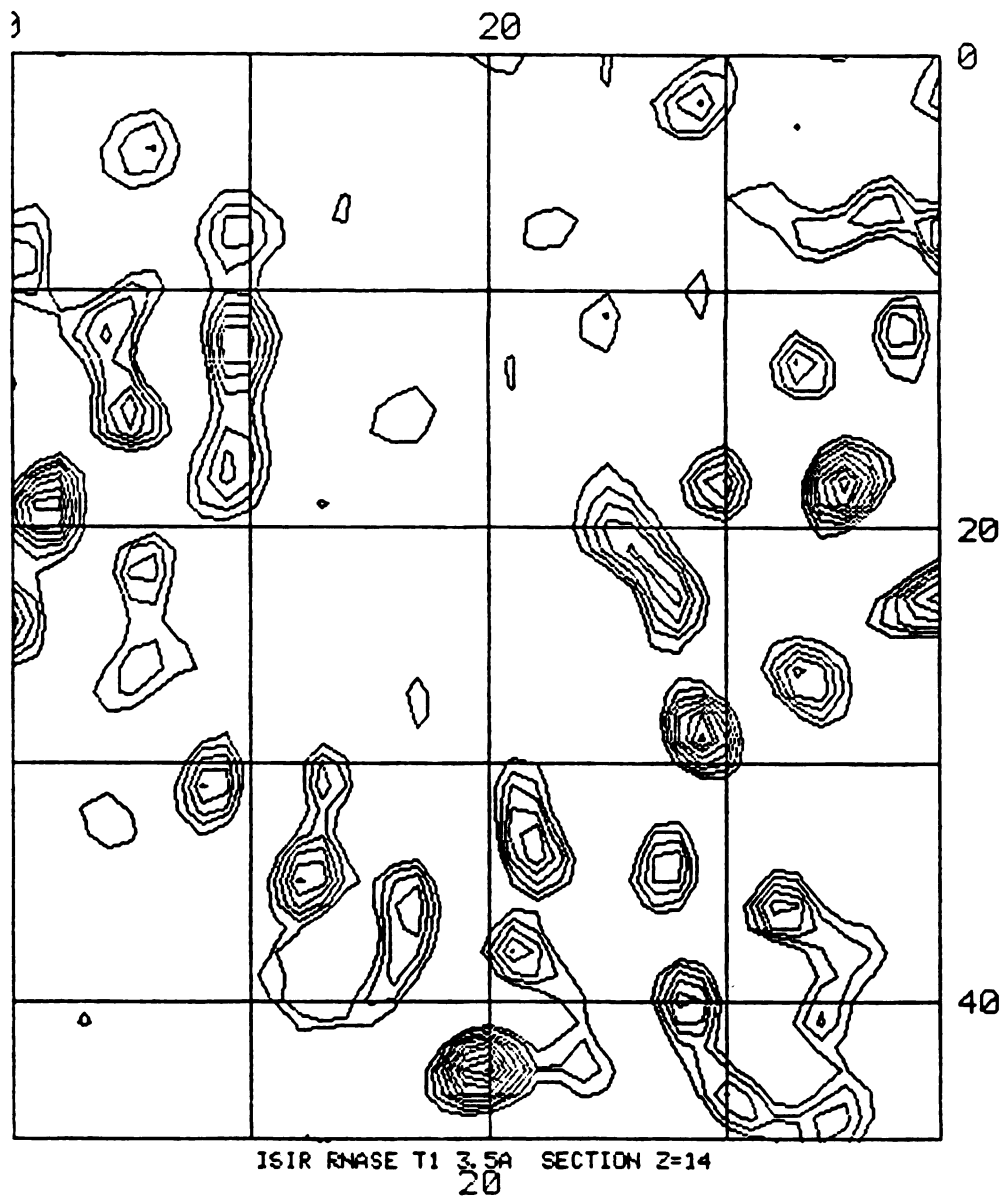


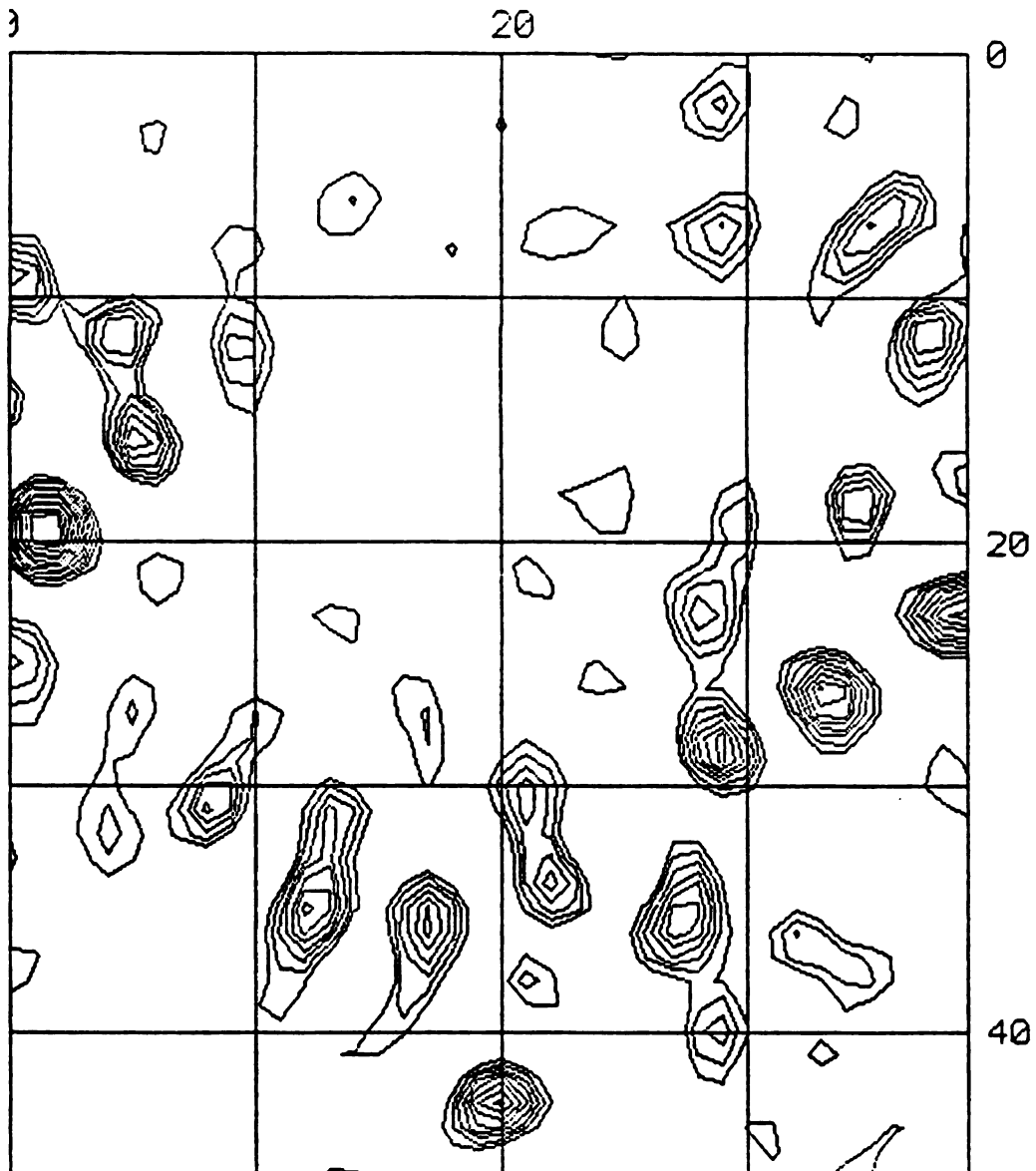
ISIR RNASE T1 3.5A SECTION Z=12
20



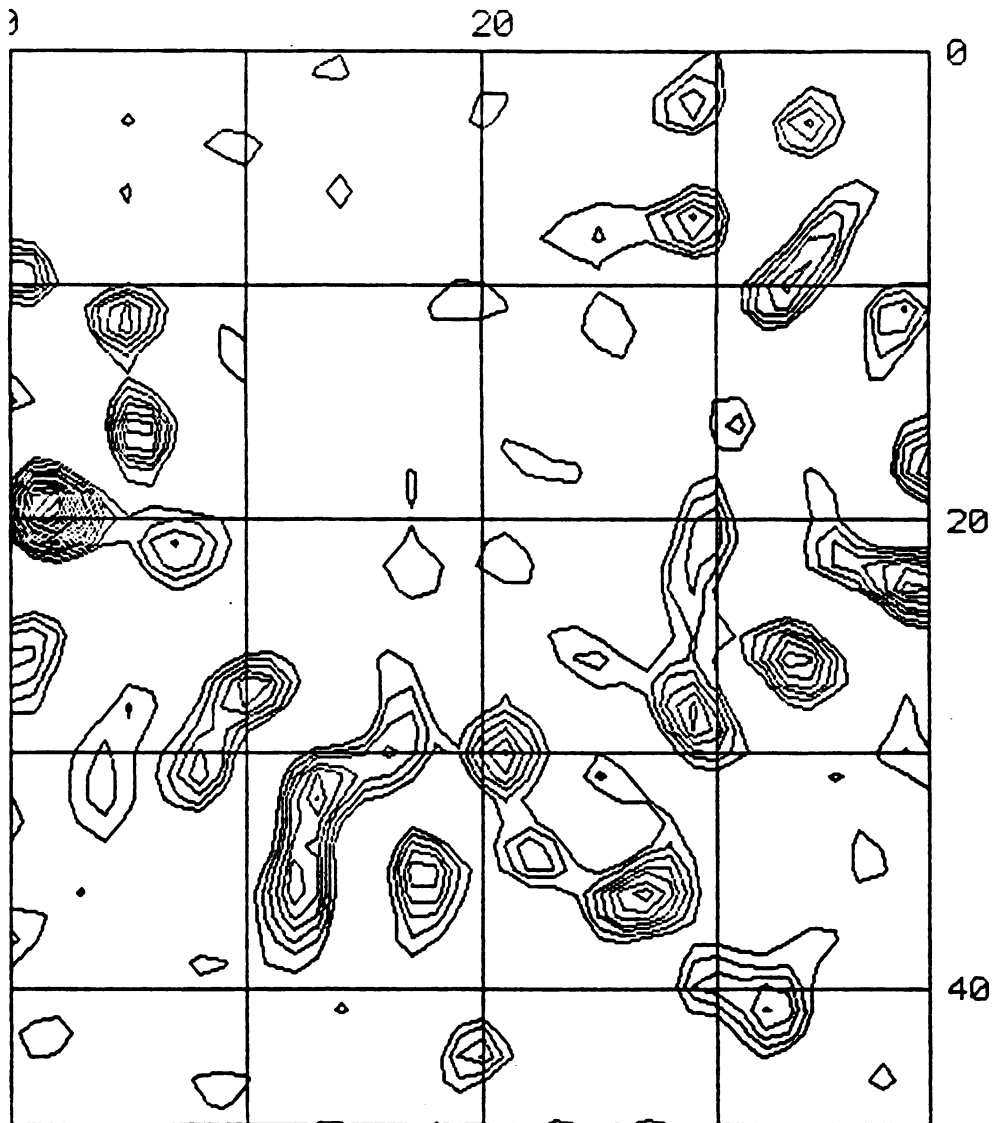
ISIR RHASE T1 3.5A SECTION Z=13

20

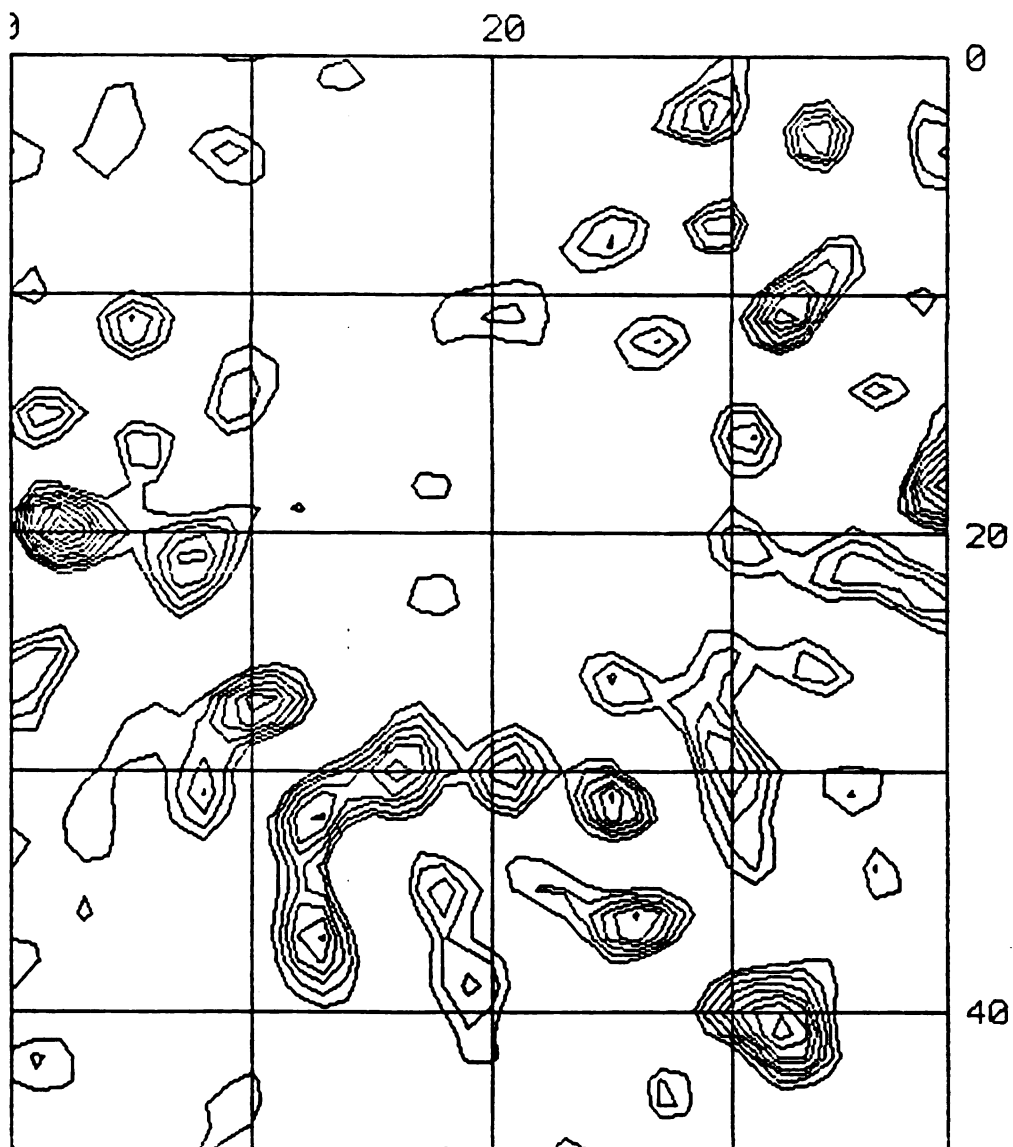


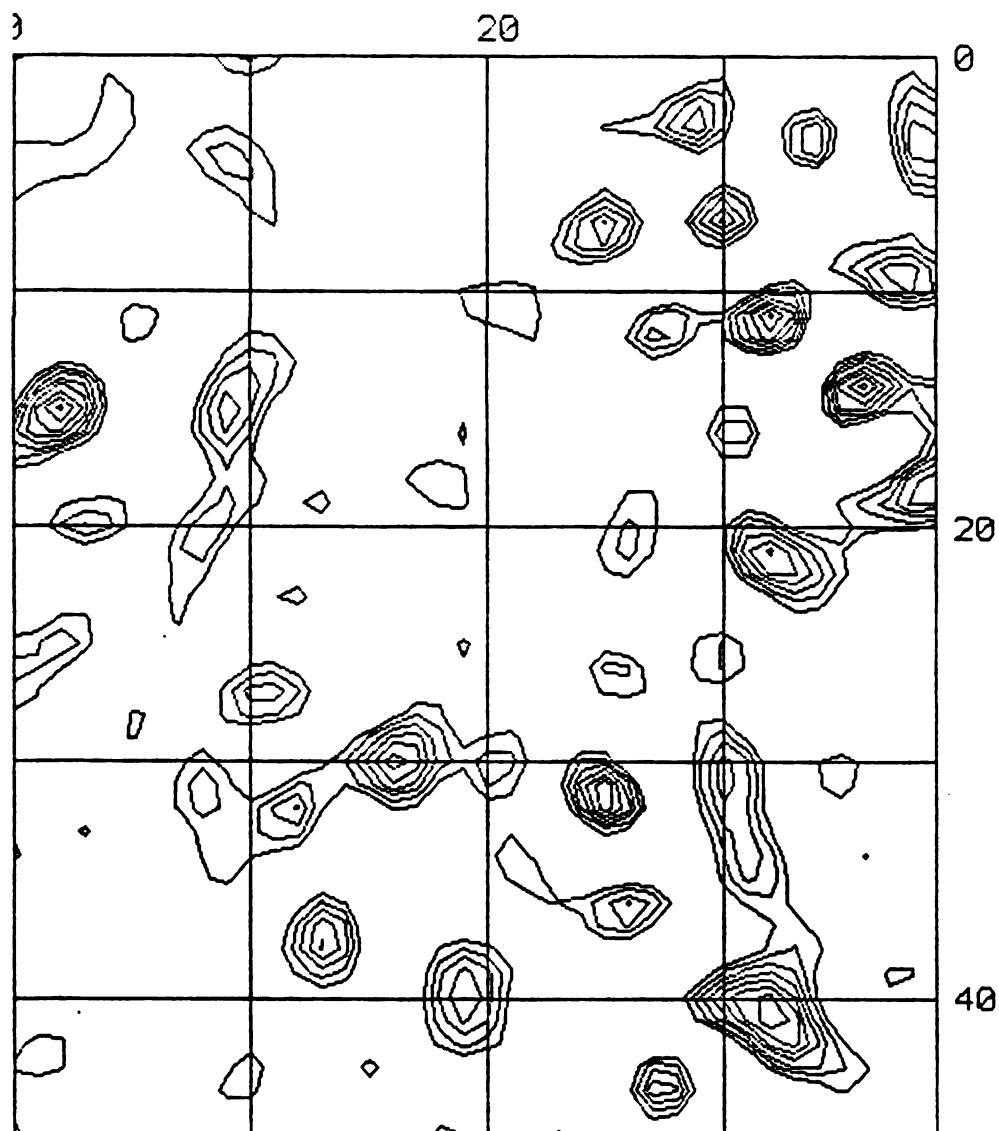


ISIR RHASE T1 3.5A SECTION Z=15
20

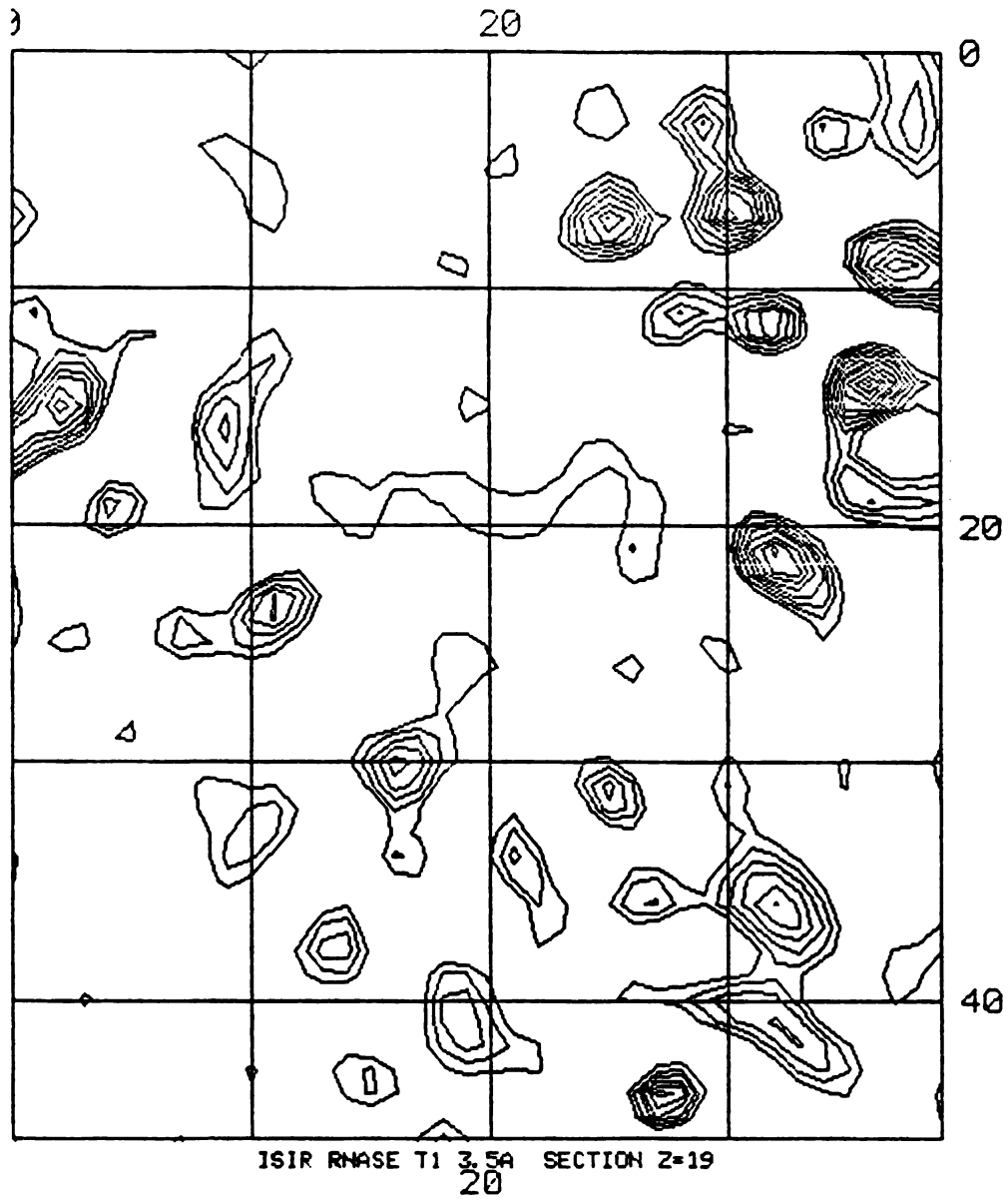


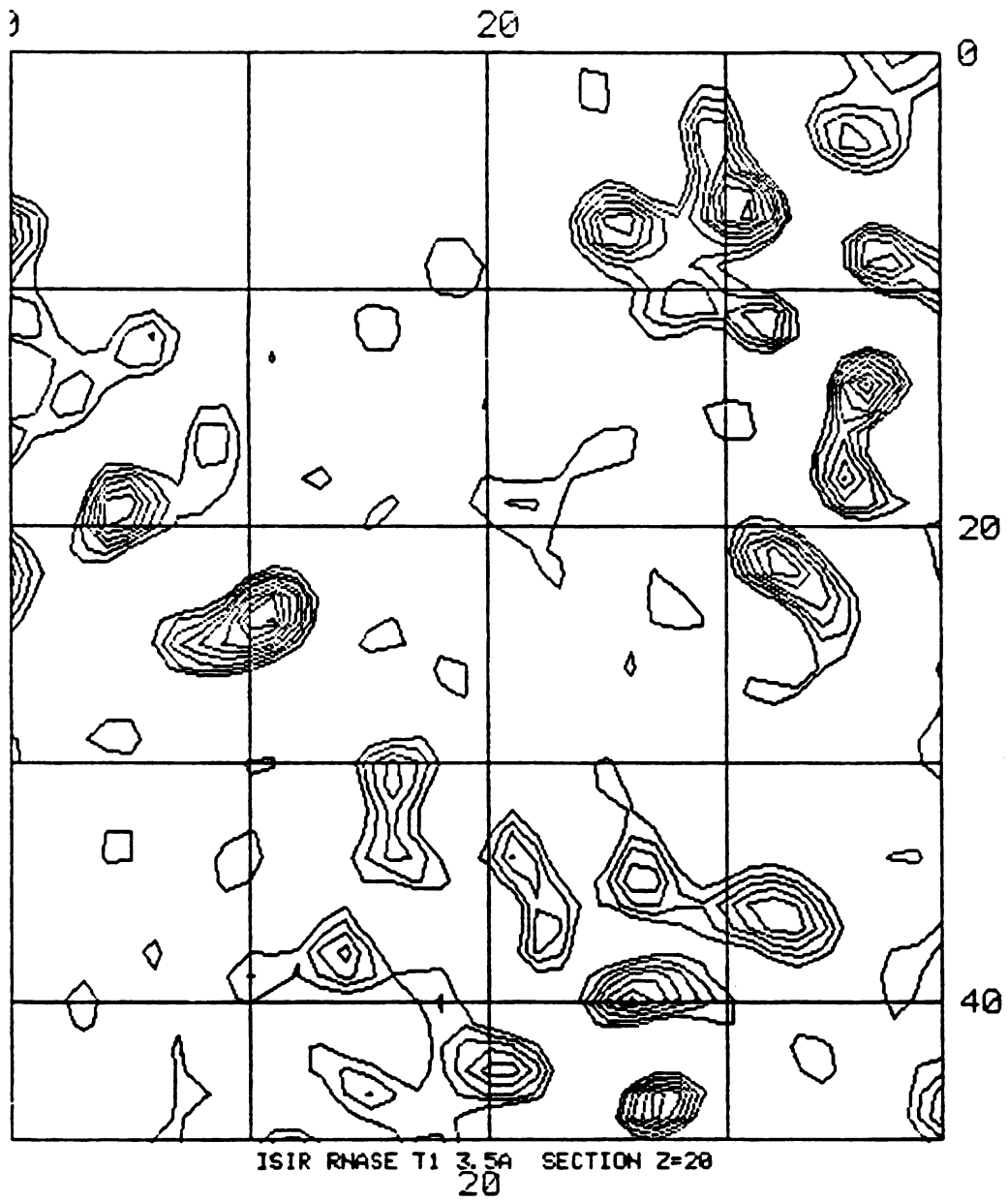
ISIR RHASE T1 3.5A SECTION 2-16
20

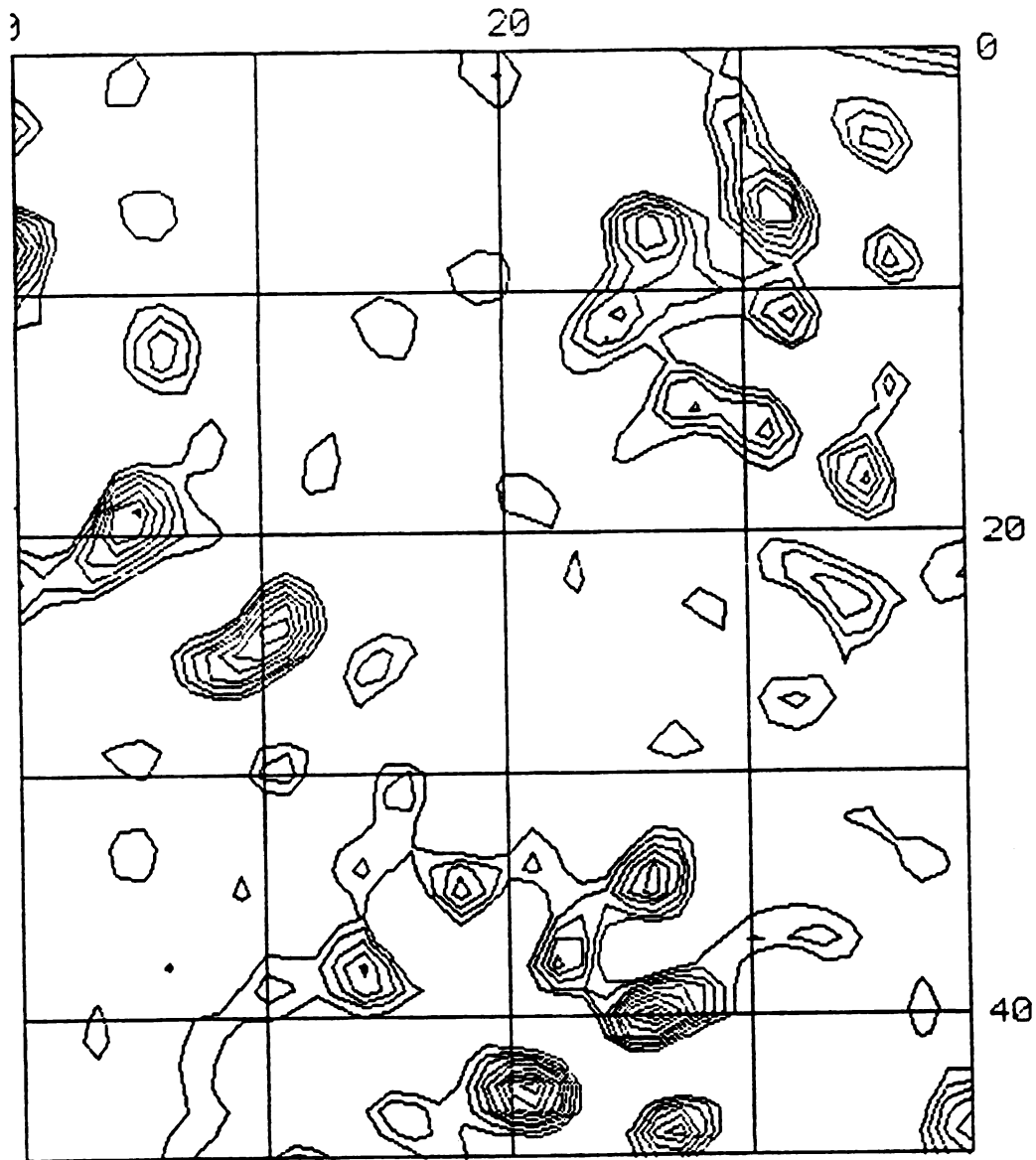




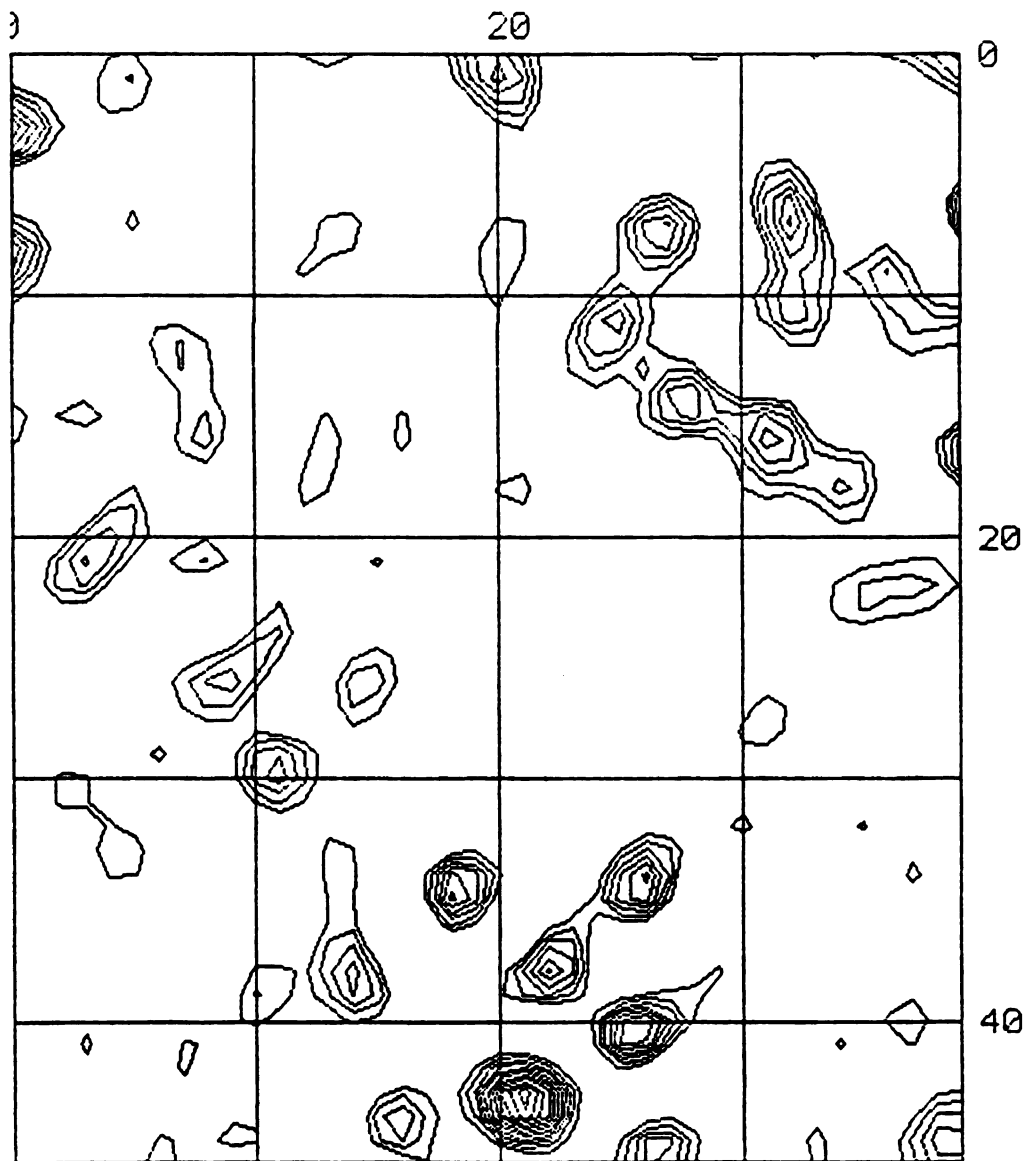
ISIR RHASE T1 3.5A SECTION Z=18
20



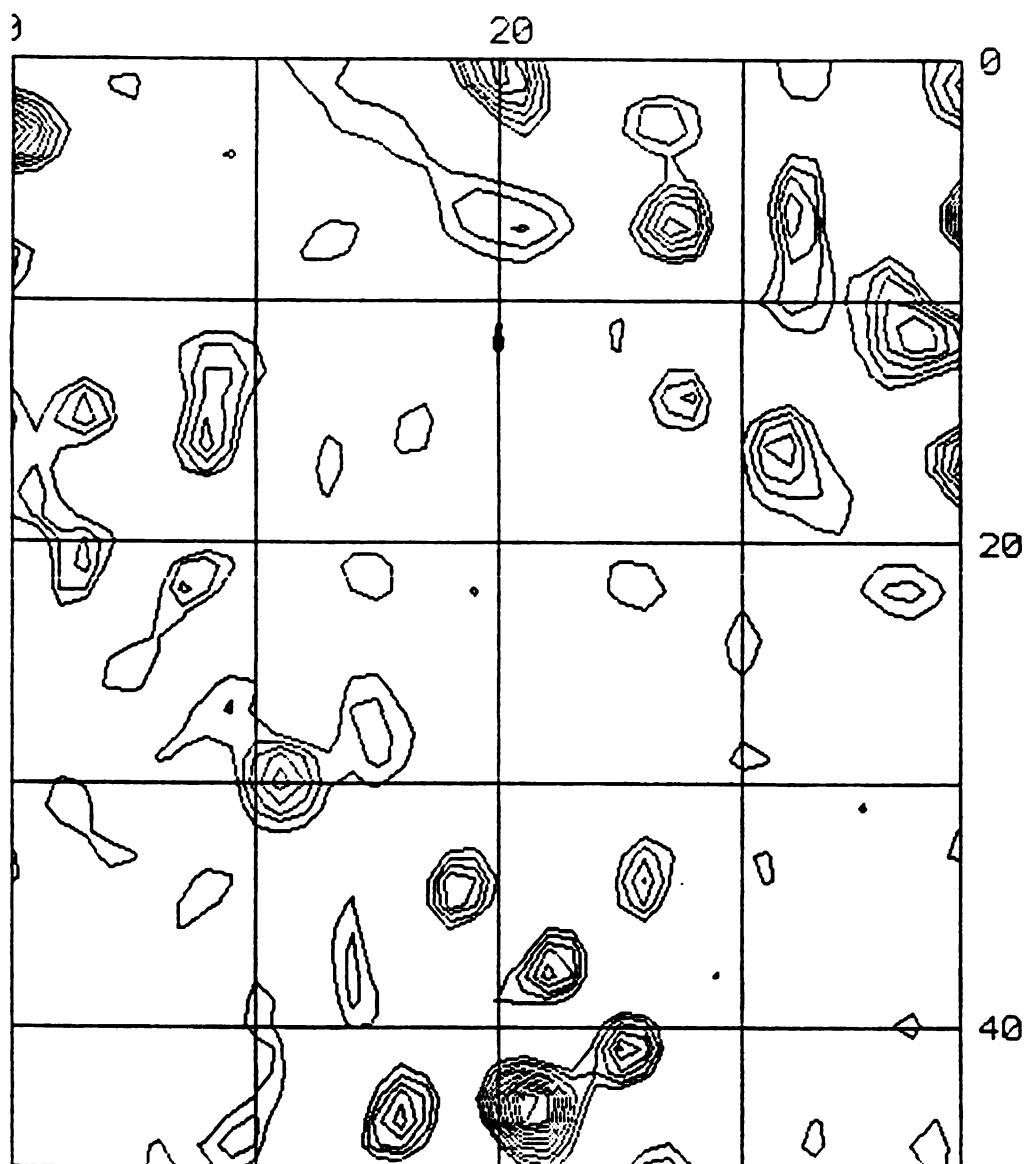




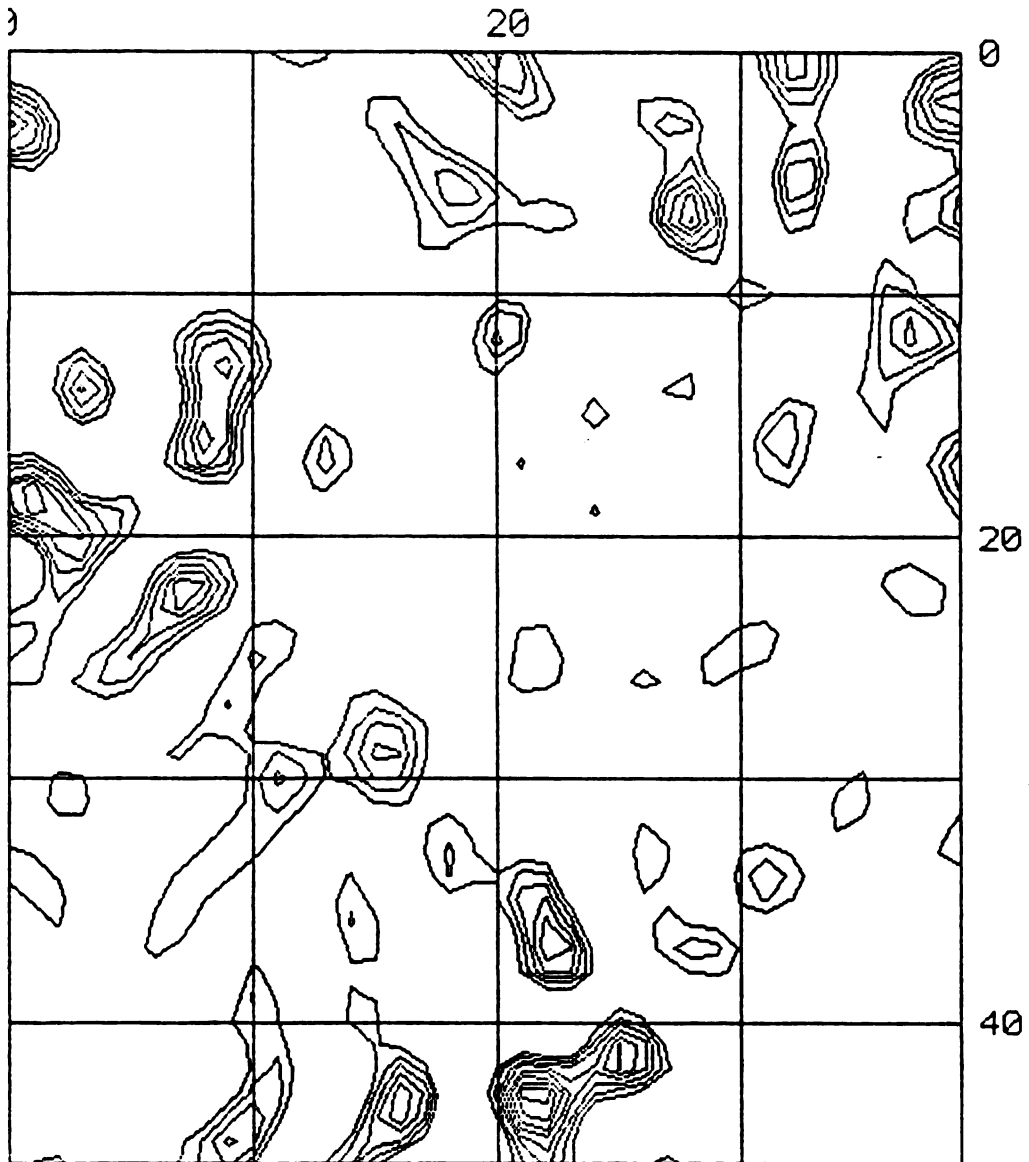
ISIR RNASE T1 3.5A SECTION Z=21
20



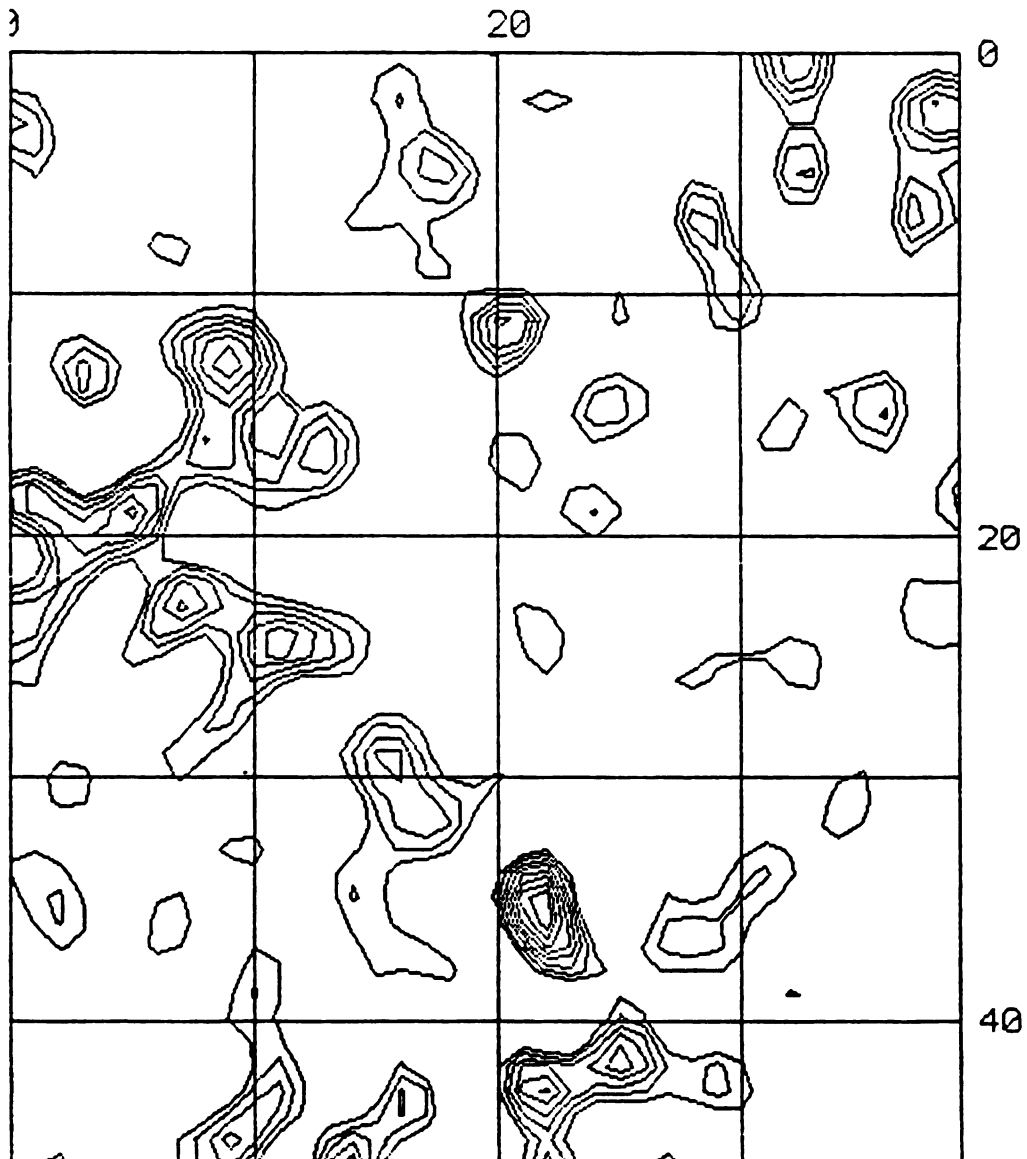
ISIR RNASE T1 3.5A SECTION 2-22
20



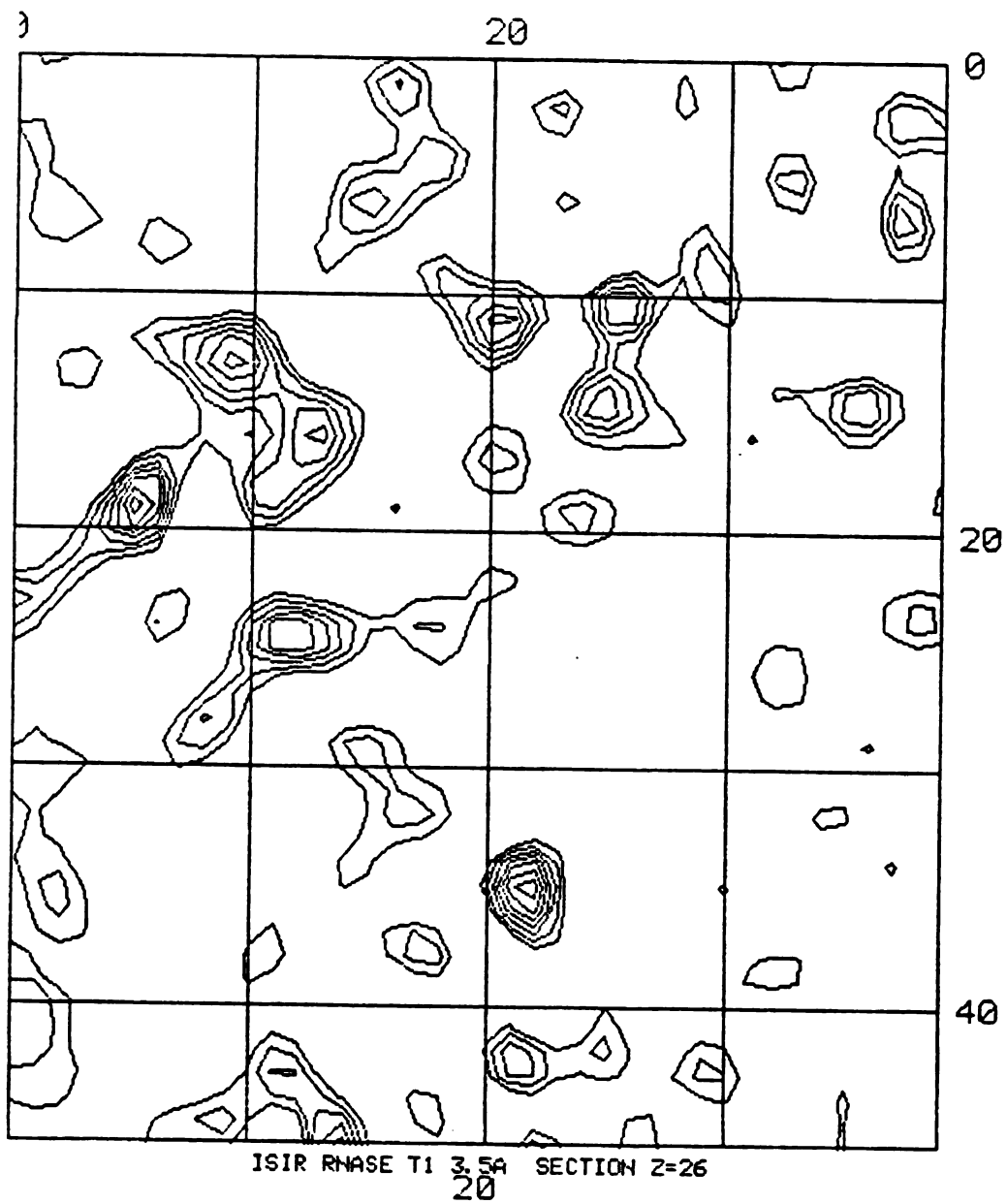
ISIR RHASE T1 3.5A SECTION Z=23
20

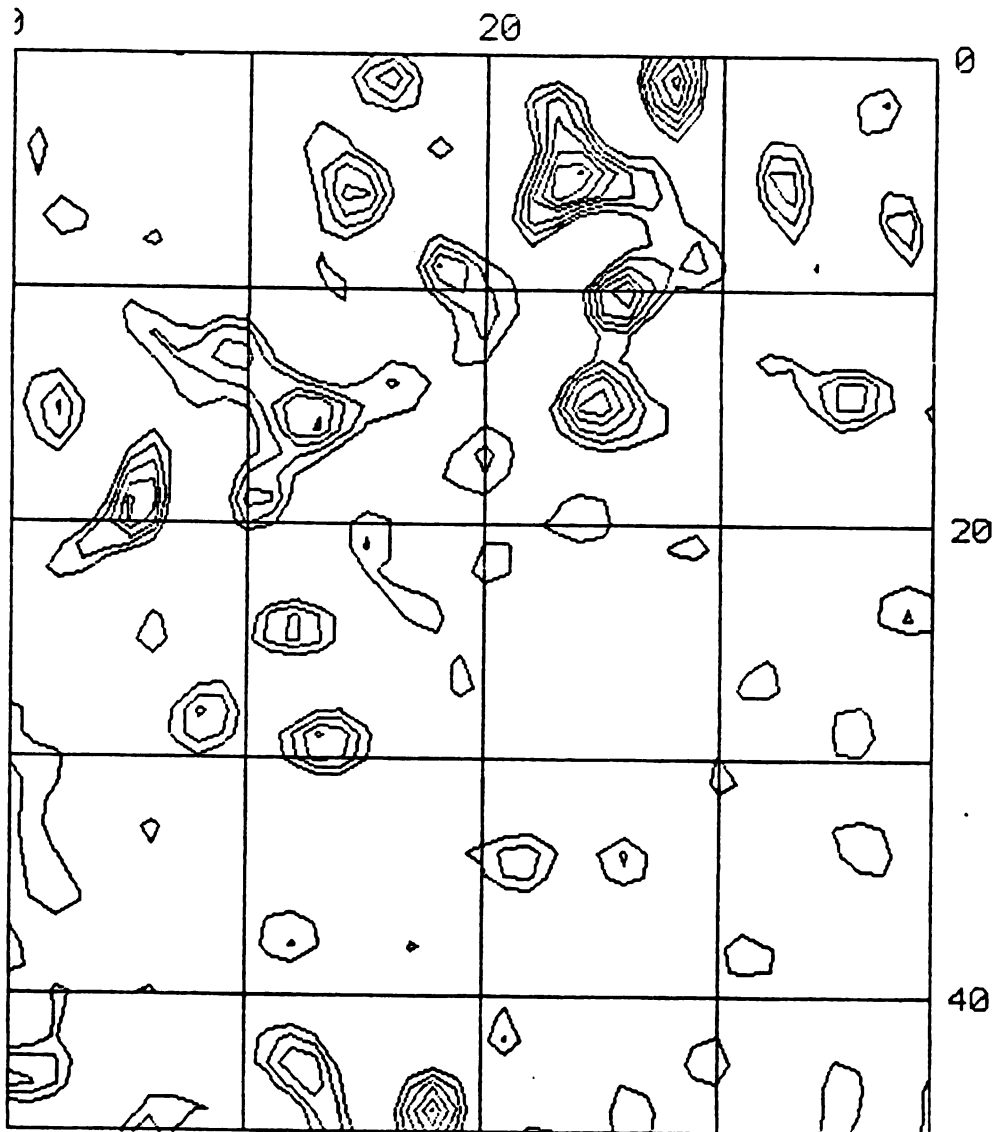


ISIR RNASE T1 3.5A SECTION 2=24
20

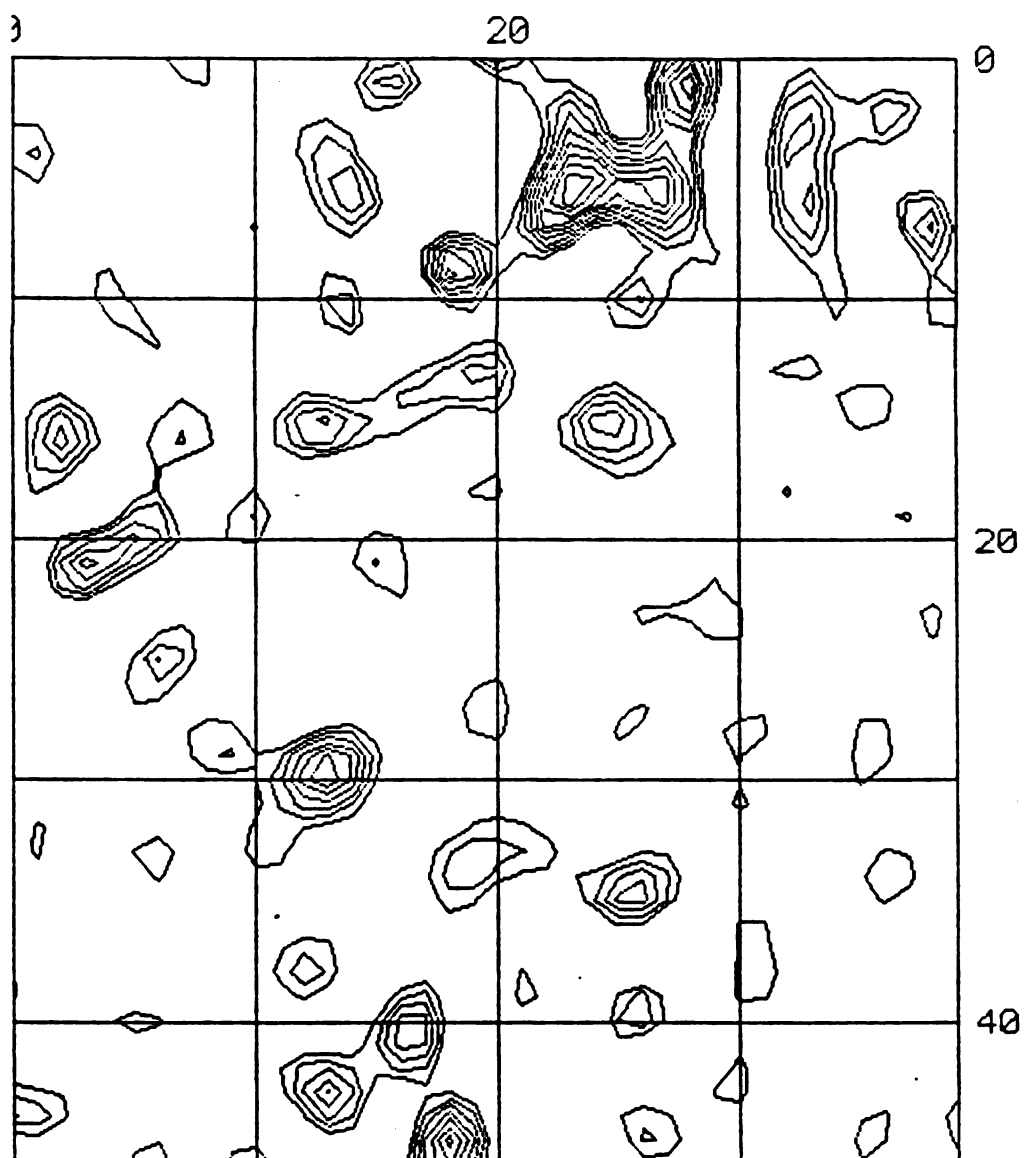


ISIR RNASE T1 3.5A SECTION Z=25
20

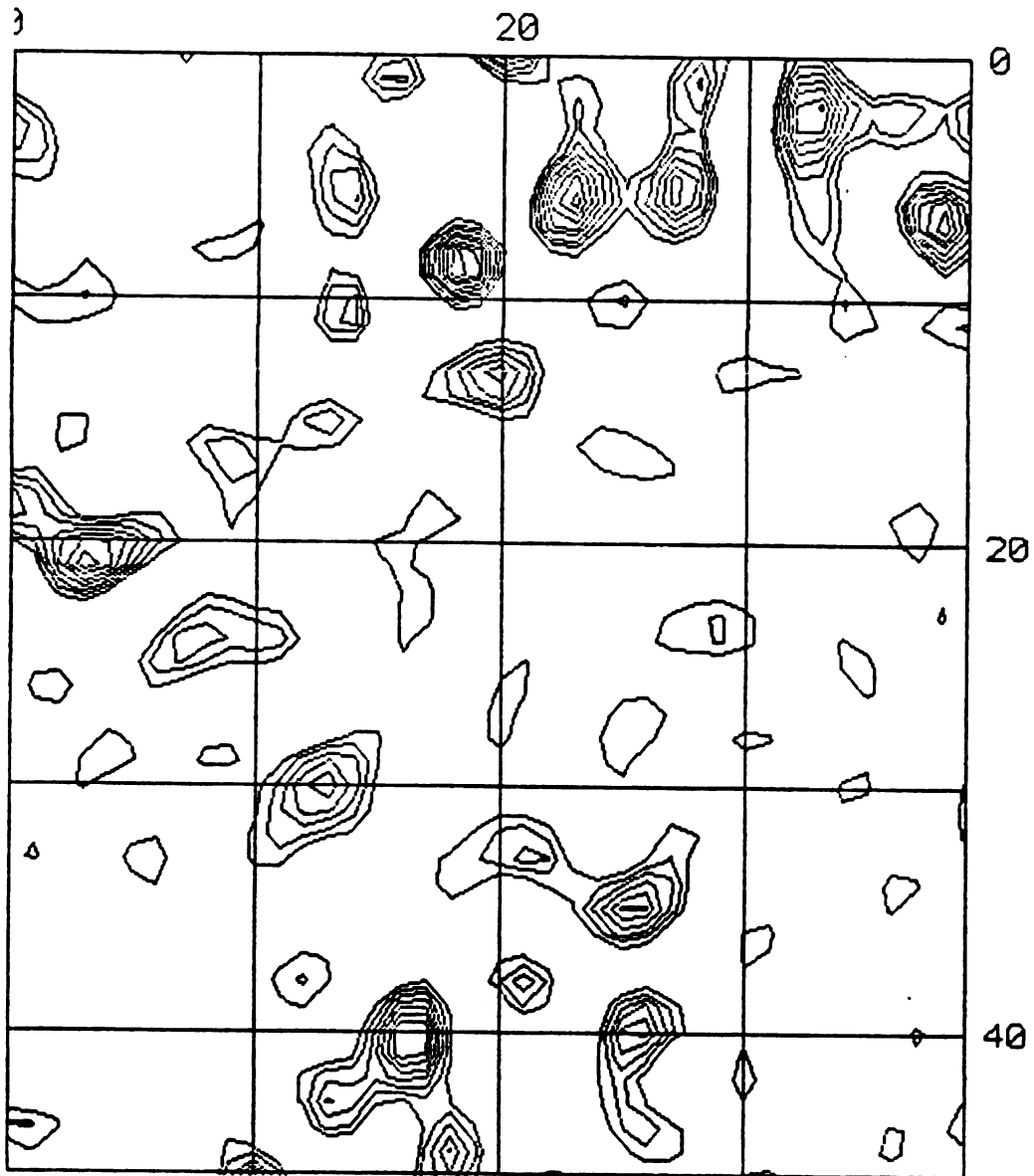


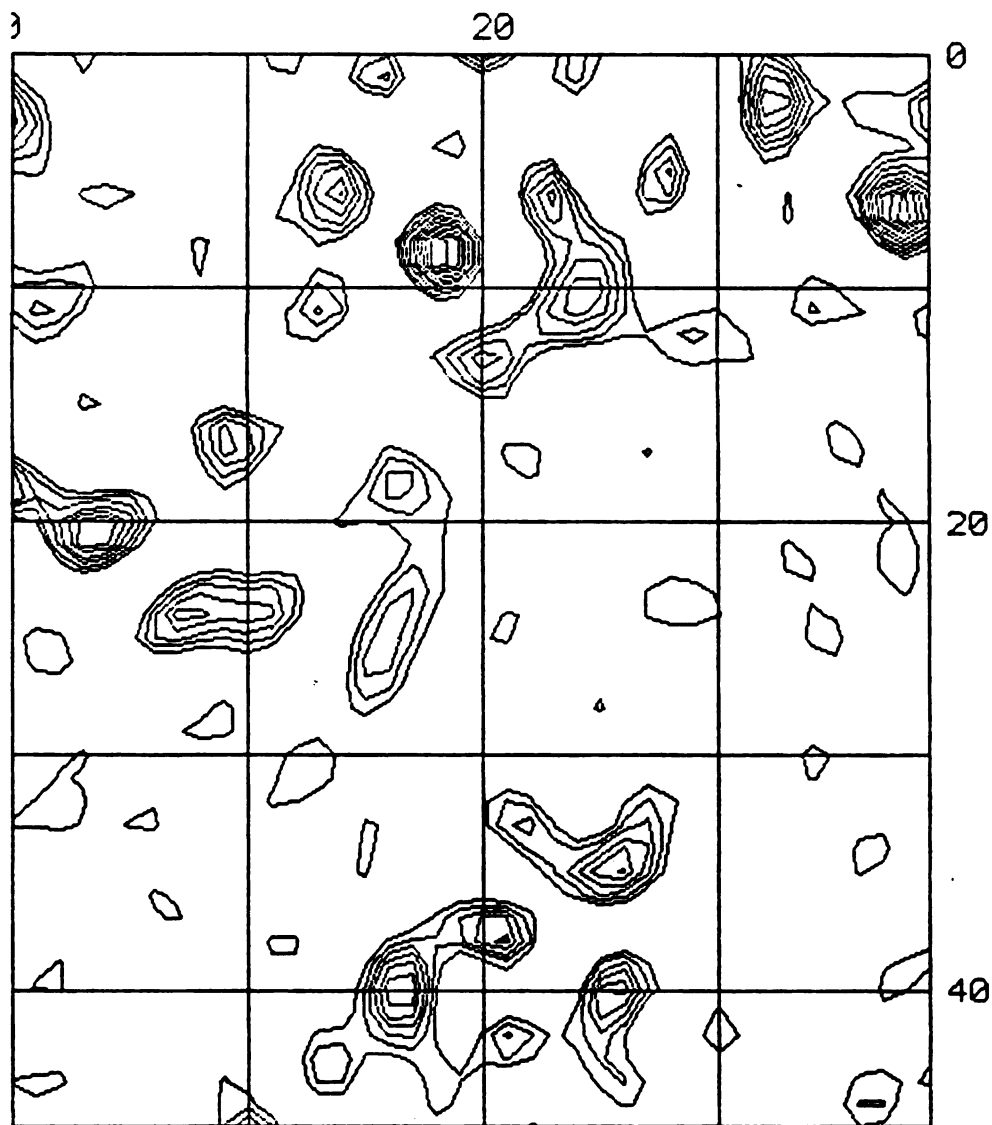


ISIR RNASE T1 3.5A SECTION 2-27
20

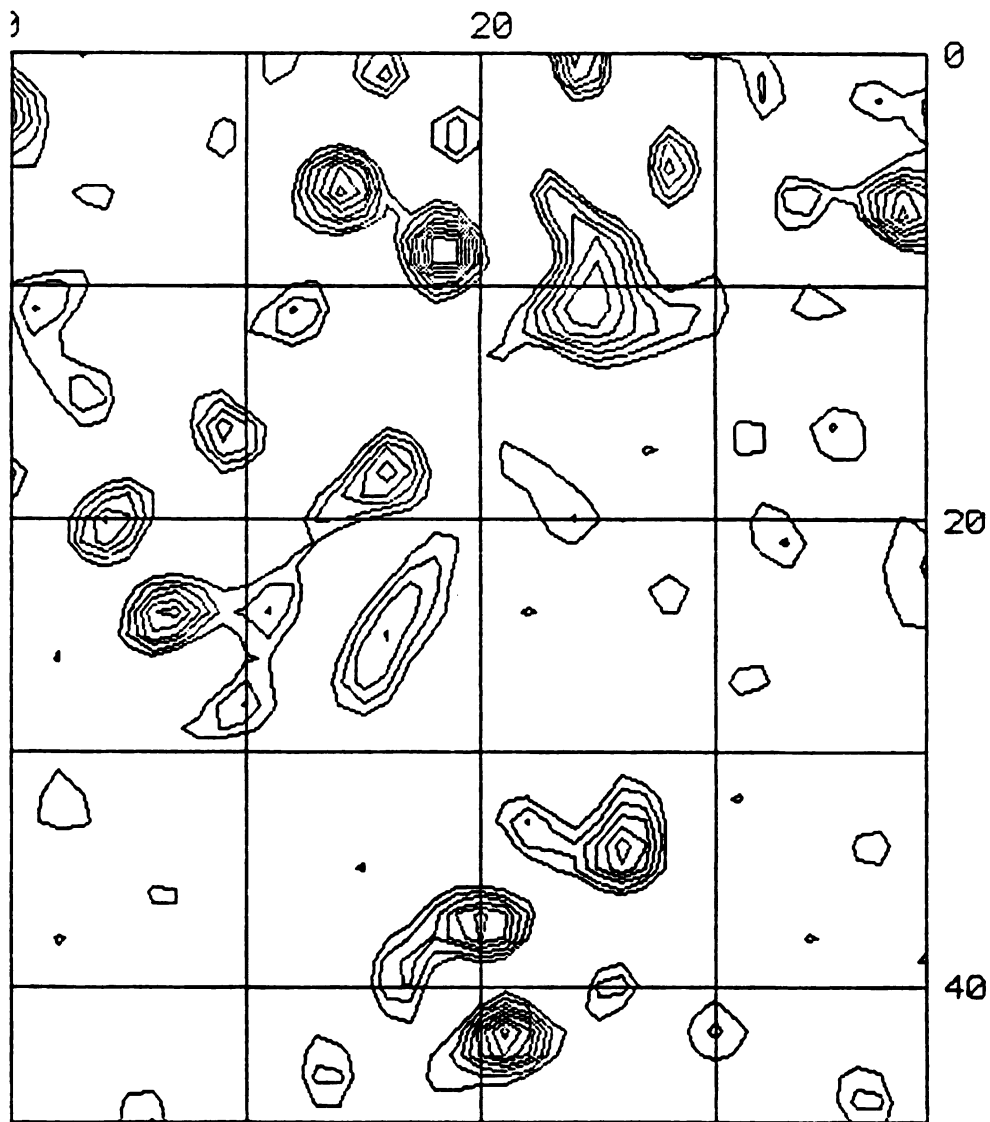


ISIR RHASE T1 3.5A SECTION Z=28
20

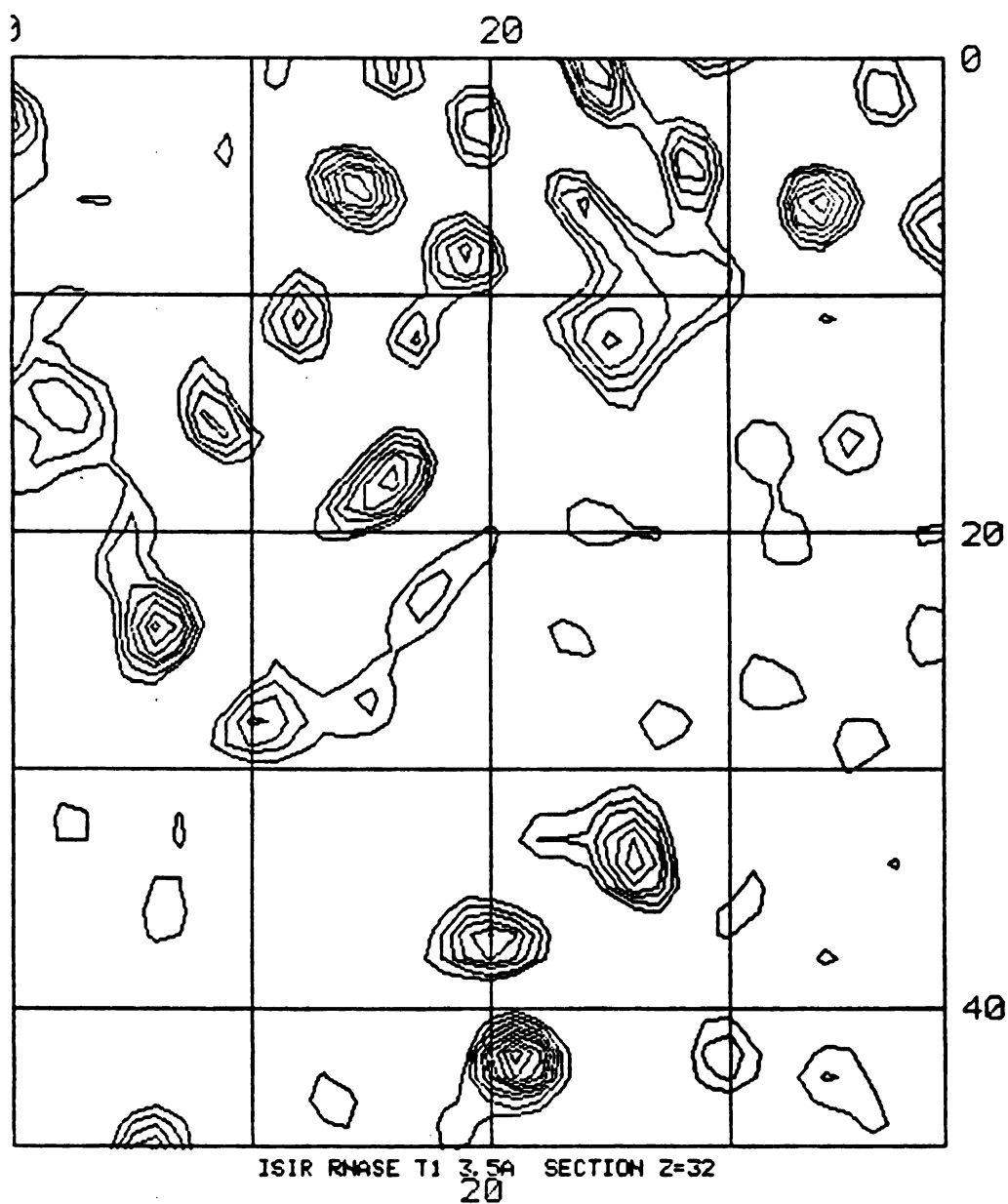


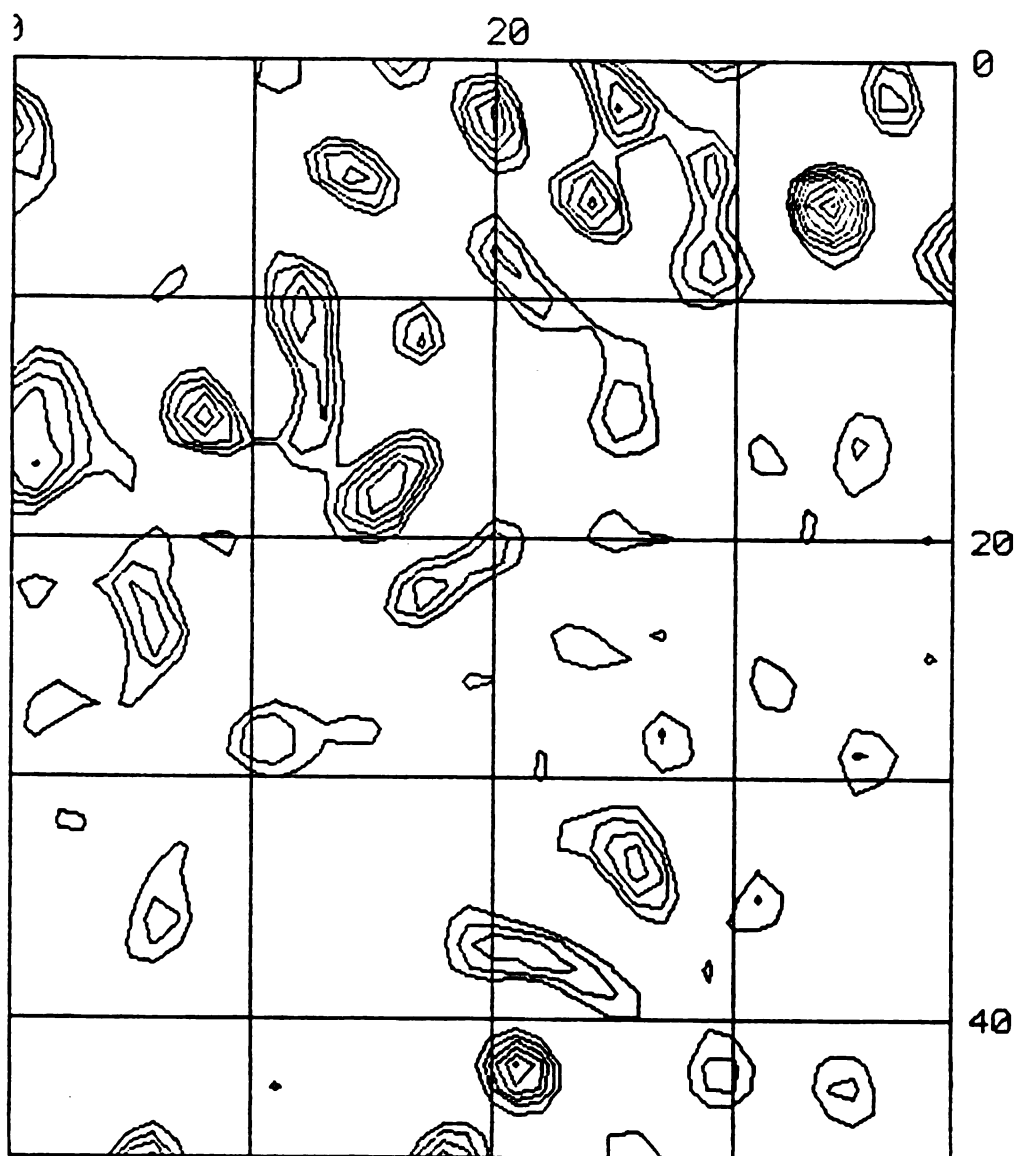


ISIR RHASE T1 3.5A SECTION 2=30
20

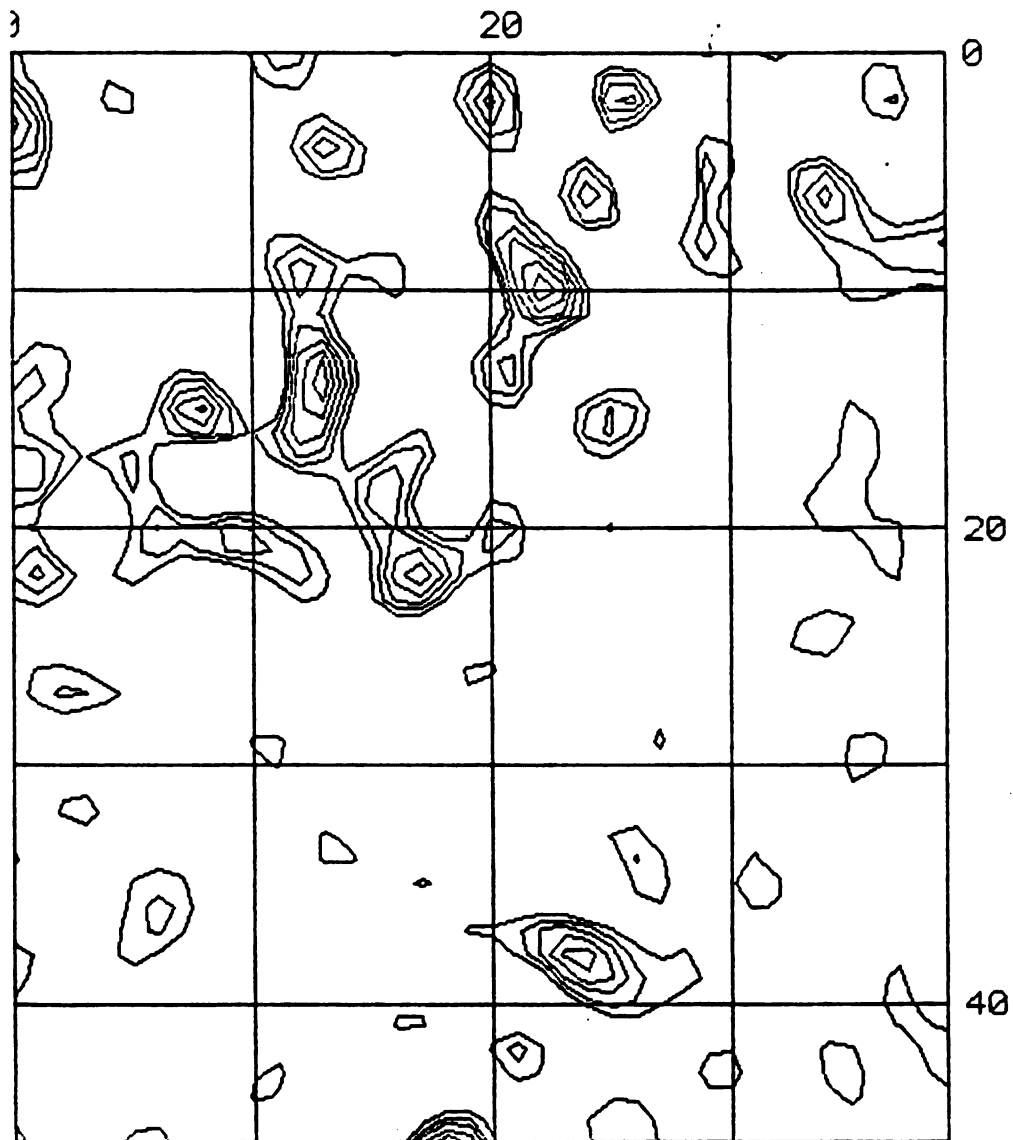


ISIR RHASE T1 3.5A SECTION Z=31
20

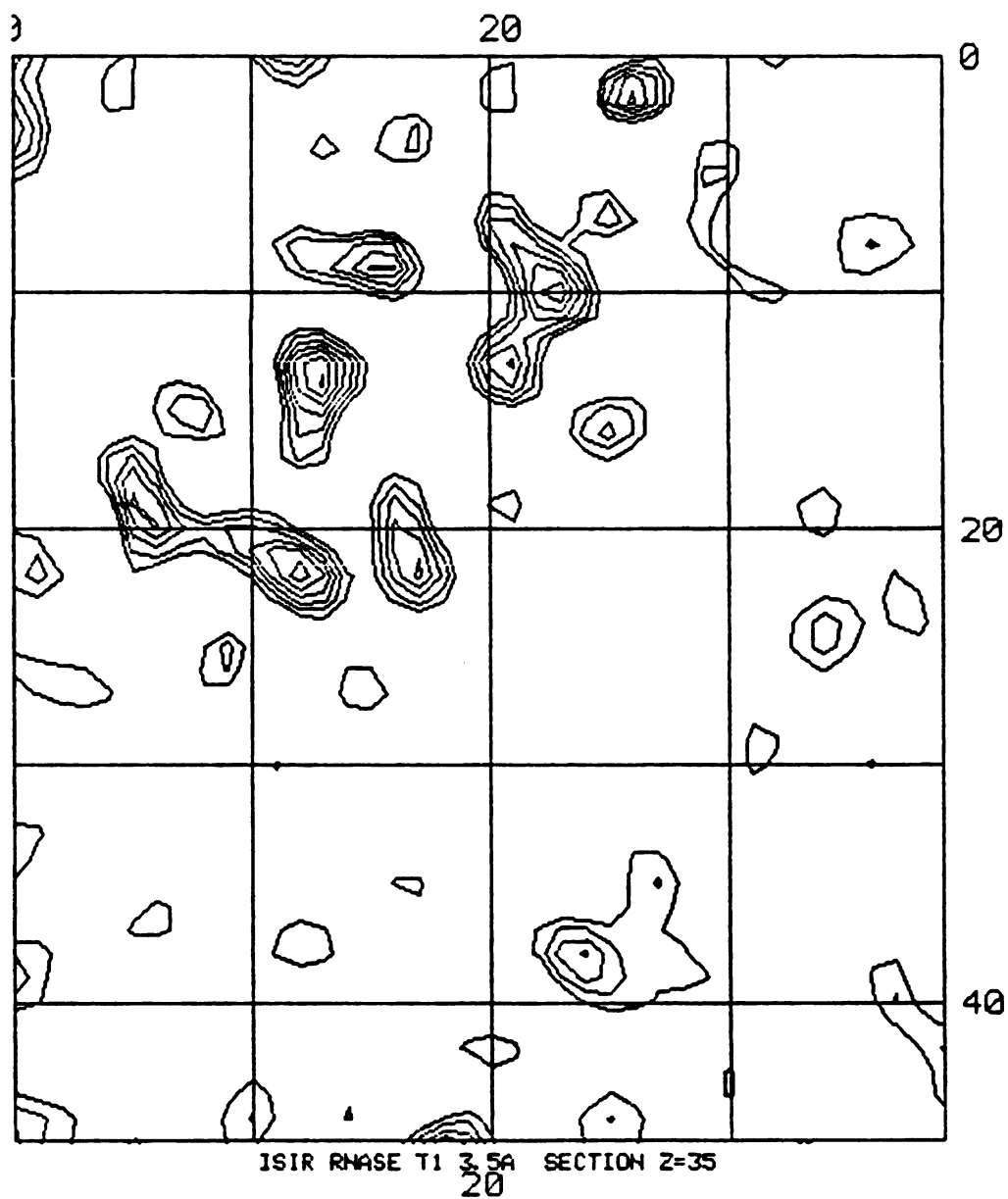


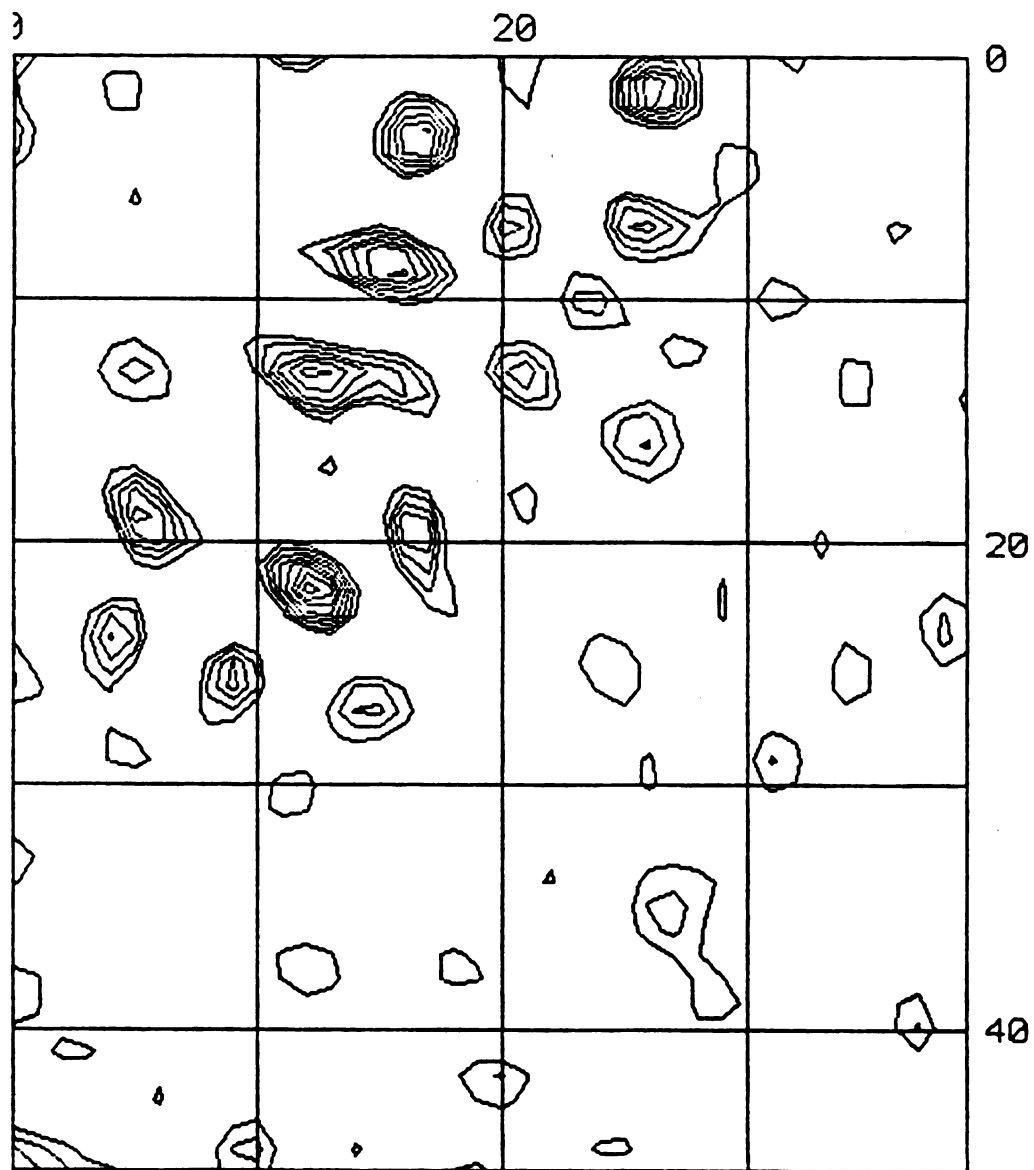


ISIR RHASE T1 3.5A SECTION Z=33
20

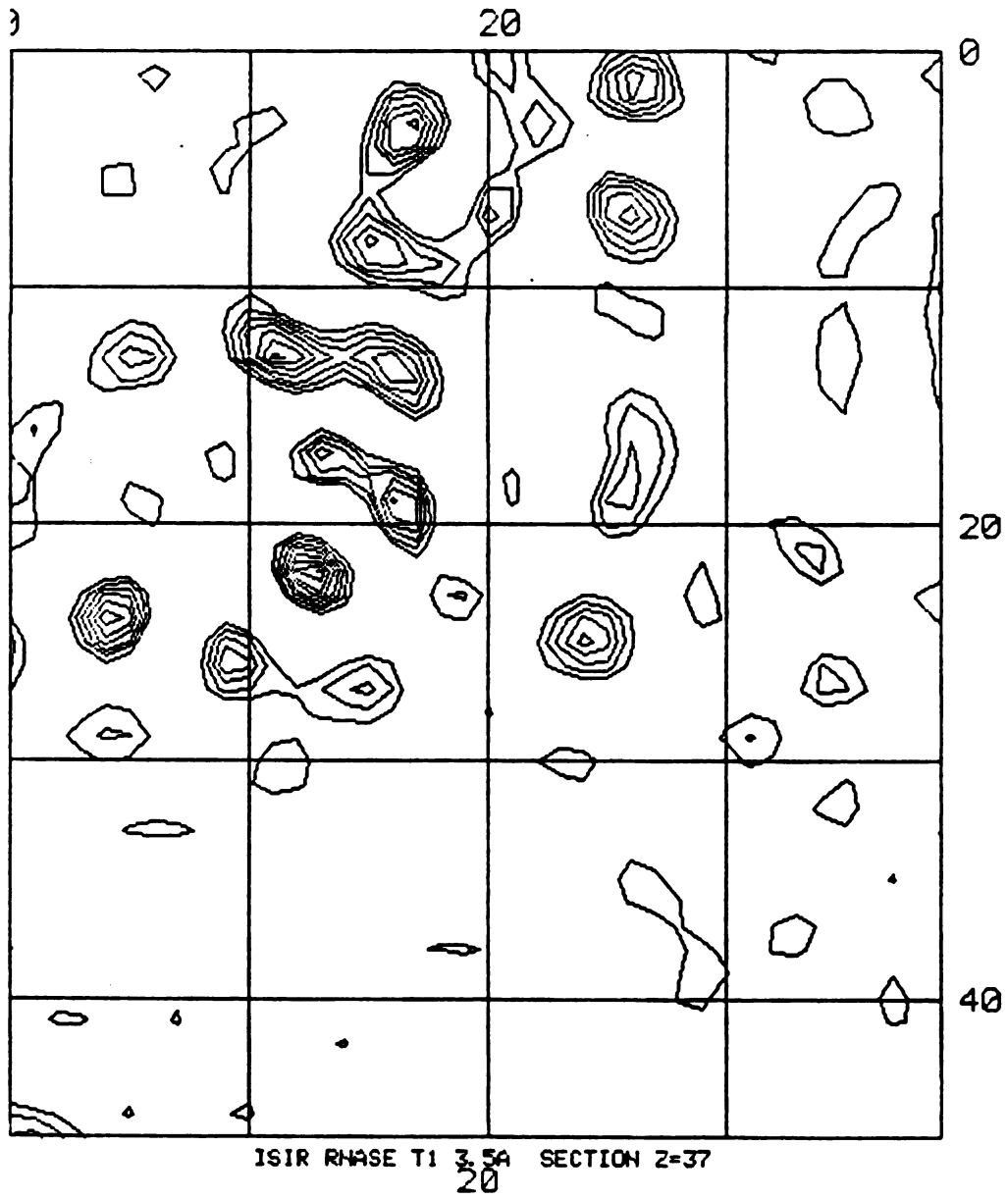


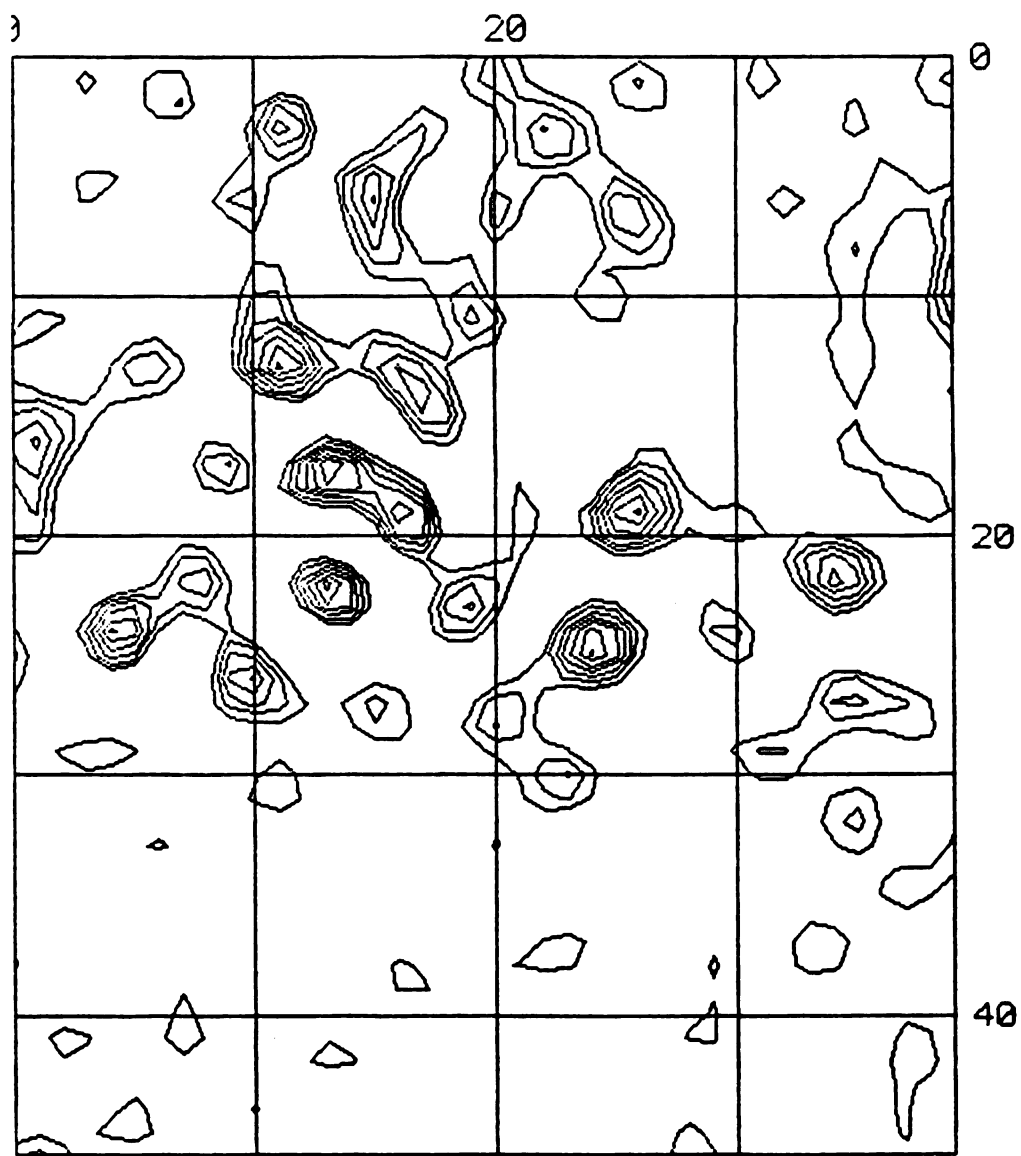
ISIR RHASE T1 3.5A SECTION Z=34
20





ISIR RNASE T1 3.5A SECTION Z=36
20





ISIR RNASE T1 3.5A SECTION Z=38
20