STRUCTURAL STUDIES

ON

MANGANESE SUPEROXIDE DISMUTASE

Thesis submitted for the degree

of Doctor of Philosophy in the

University of Oxford.

October 1985

Michael William Parker

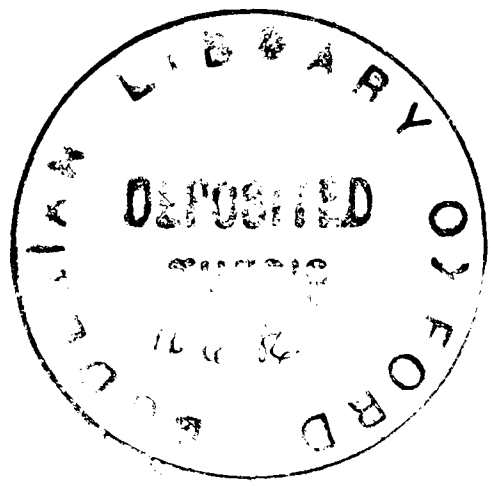
Laboratory of Molecular Biophysics

Department of Zoology

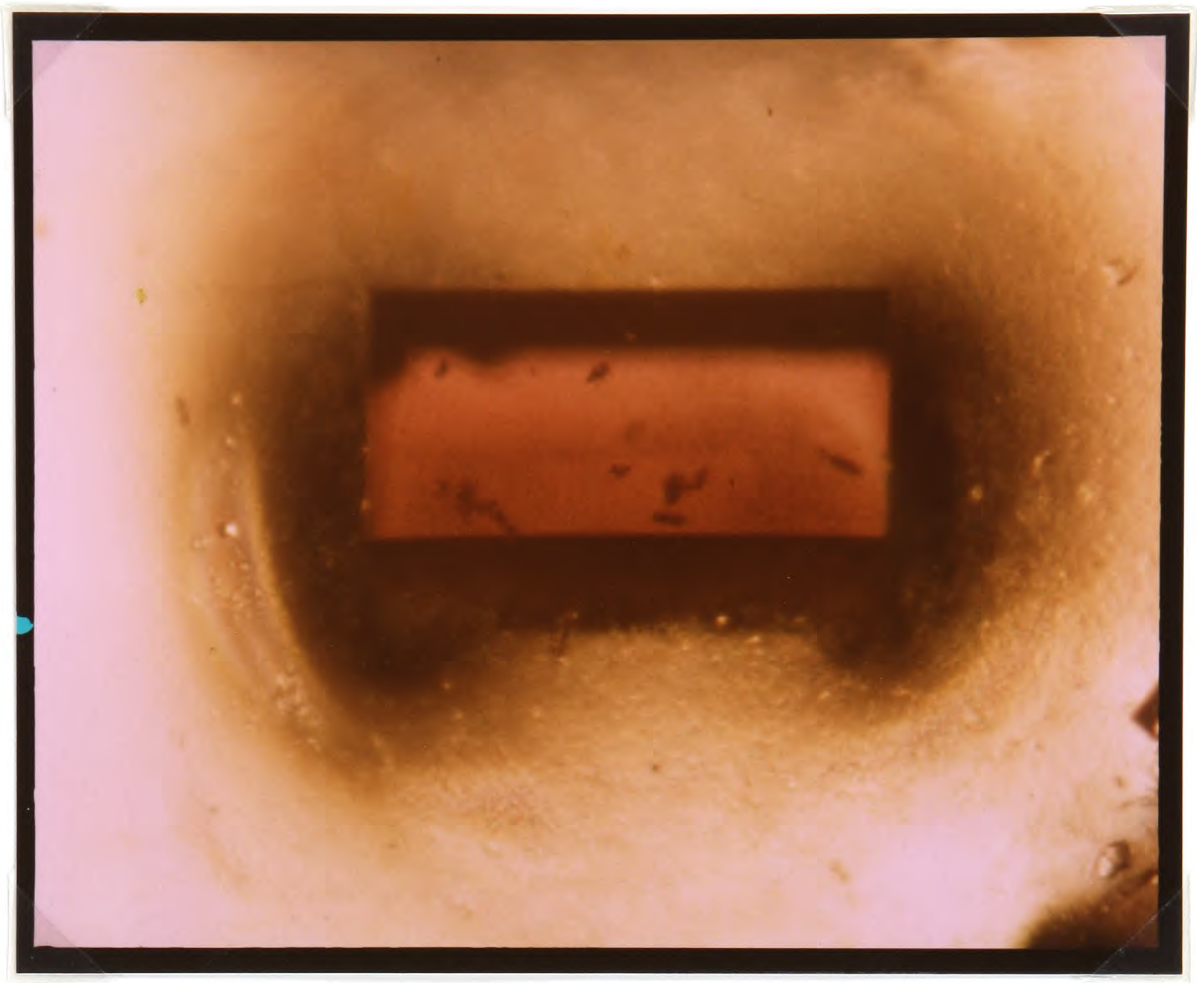
Oxford

Wolfson College

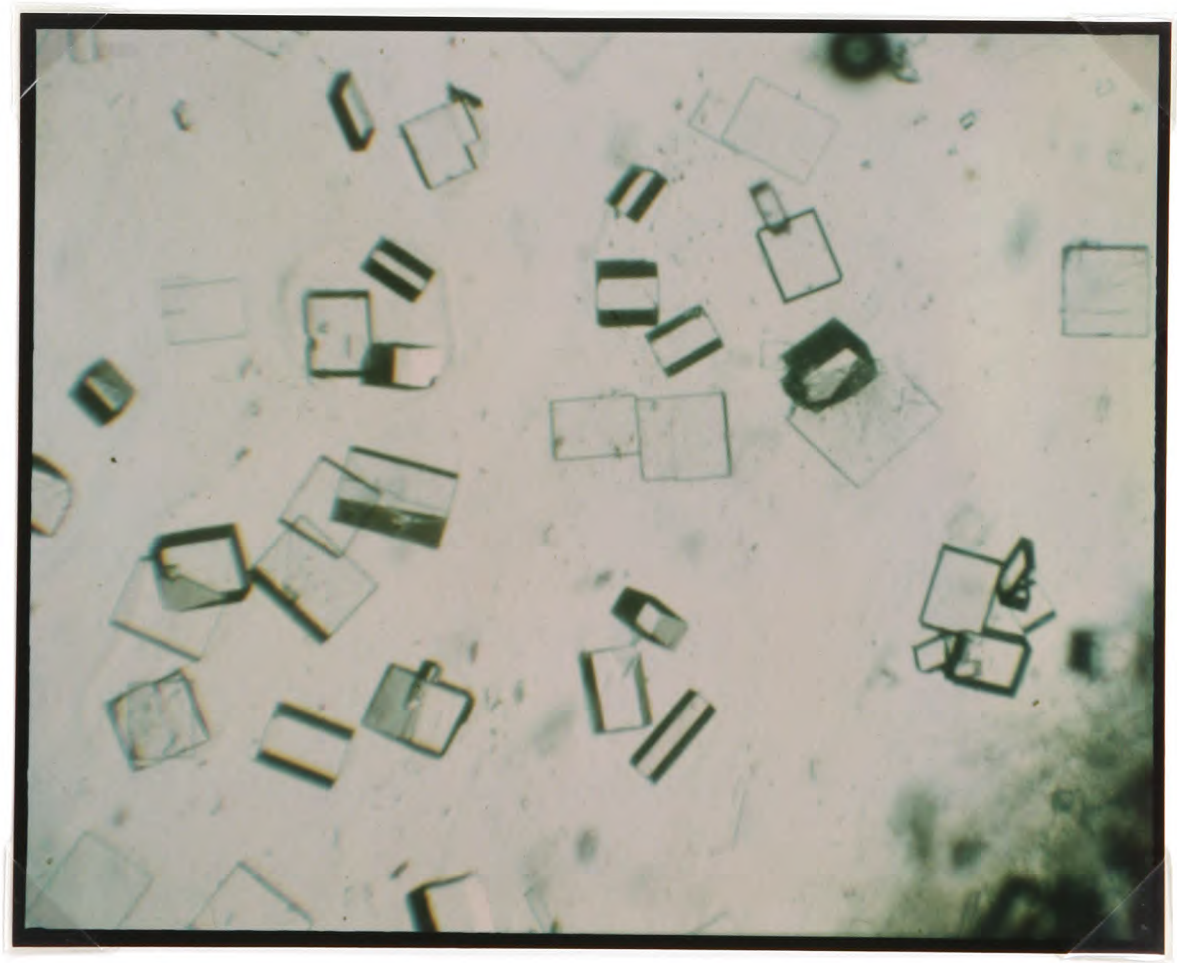
Oxford



Frontispiece



Crystals of manganese superoxide dismutase from Bacillus stearothermophilus.



To my family

ABSTRACT

STRUCTURAL STUDIES ON MANGANESE SUPEROXIDE DISMUTASE

Michael William Parker — Wolfson College,

D.Phil Thesis — Oxford

Michaelmas Term, 1985

Superoxide dismutases are widely distributed enzymes which catalyse the dismutation of superoxide radicals to dioxygen and hydrogen peroxide and are considered to be an important agent of an organism's defence against oxygen toxicity. The crystallization and low resolution structure determination of manganese superoxide dismutase (E.C. 1.15.1.1) from Bacillus stearothermophilus is described.

The enzyme crystallized in space group $P2_12_12$ with two monomers per asymmetric unit and cell dimensions of a=72.2Å, b=111.1Å and c=51.1Å. The crystals diffracted to beyond 2Å resolution but were fragile and prone to cell dimension changes. The cell dimension variability was overcome to some extent by crosslinking with glutaraldehyde.

An electron density map was calculated to 6Å resolution initially by the method of multiple isomorphous replacement using data obtained from six heavy atom derivatives. The final map was calculated from single isomorphous replacement data using a map modification procedure. The fitting of an alpha carbon model of iron superoxide dismutase into the map suggested the iron and manganese enzymes are structurally related. The position of the metal atoms in the model solved difference Patterson maps calculated from data collected from a manganese-free crystal and from anomalous dispersion data. The latter data were collected using synchrotron radiation tuned close to the manganese absorption edge. The low resolution map and the availability of 2.4Å resolution native data paves the way for higher resolution X-ray studies of the crystals.

A detailed analysis of amino acid sequences has been carried out on the various metal-containing superoxide dismutases. The results indicate that the enzymes can be classified according to their metal cofactor. The distribution and homology of the enzyme classes supports the endosymbiotic theory of the origin of cell organelles. The presence of the copper/zinc enzyme in Photobacterium leiognathi is shown to support the case for a eukaryote to prokaryote gene transfer.

Acknowledgements

I would like to record my appreciation and thanks to the following people for the part they have played in this work:

Sir David Phillips for giving me the opportunity to work in the Laboratory.

Dr. Colin Blake who supervised my studies.

Dr. Joe Bannister who provided much stimulation through collaborative work.

Dr. John Helliwell and coworkers for their help and continued interest in the anomalous scattering work.

Drs.' Sheila Gover, Gary Taylor and David Stuart for their help and expertise with computing.

Drs.' Peter Artymiuk, Jane Burridge and others in the F9 group for help and encouragement.

The technical staff, in particular John Marsh (diffractometers), Philip Keighley (X-ray generators), Lance Mangold (photography) and Anne Roper (diagrams).

And finally, to the Royal Commission for the Exhibition of 1851 for generous financial support.

CONTENTS

SYMBOLS AND ABBREVIATIONS

Roman Symbols

$\mathring{A}$	Ångstrom unit (=0.1 nanometres)
A	absorption
$\underline{a},\underline{b},\underline{c}$	real space cell dimensions
$\underline{a}^*,\underline{b}^*,\underline{c}^*$	reciprocal space cell dimensions
B	isotropic temperature factor
E	r.m.s. error in isomorphous scattering measurements
E'	r.m.s. error in anomalous scattering measurements
$\underline{F},F$	structure factor and its amplitude
F_P	native protein structure factor amplitude
F_{PH}^+,F_{PH}^-	derivative Friedel-related reflections
F_H	heavy-atom structure factor amplitude
F_{HLE}	heavy atom lower estimate of F_H
F_{HUE}	heavy atom upper estimate of F_H
f_j	atomic scattering factor for jth atom
f_o	atomic scattering factor for 'free' electrons
f',f''	real and imaginary part of anomalous scattering
I	intensity
k_{emp}	empirical anomalous ratio
k_{theor}	theoretical anomalous ratio ($=(f_o+f')/f''$)
m	figure of merit
R	reliability index
w	weighting term
$\langle\ \rangle$	arithmetric mean of quantity between brackets

Greek Symbols

α_{best}	best protein phase angle
α_{mp}	most probable protein phase angle
α_P	protein phase angle
α_H	heavy atom phase angle

α, β, γ	real space unit cell angles or Eulerian angles
Δ_{ano}	anomalous dispersion difference ($= F_{PH}^{+} - F_{PH}^{-}$)
Δ_{iso}	isomorphous difference ($= F_{PH} - F_{P}$)
ϵ_j	lack-of-closure error for jth derivative
ϵ'_j	anomalous lack-of-closure error
λ	wavelength
μ	precession angle
Θ	Bragg angle
Φ	rotation angle
Φ, Ψ, χ	spherical polar angles
$\Phi, \omega, \chi, \Theta, \sum$	five-circle diffractometer setting angles
ρ	electron density
σ	standard deviation

Abbreviations

ano	anomalous
atm	atmosphere
calc	calculated
cm	centimetres
eV, V	electron volts, volts
EXAFS	extended X-ray absorption fine structure (spectroscopy)
gpu	grid point unit
iso	isomorphous
Lp	Lorentz-polarisation (factor)
M	molar
MFID	mean fractional isomorphous difference
mg	milligrams
MIR	multiple isomorphous replacement
ml	millilitres

mm	millimetres
mM	millimolar
NHE	normal hydrogen electrode
nm	nanometres
nmr	nuclear magnetic resonance (spectroscopy)
obs	observed
req.	required
r.l.u.	reciprocal lattice unit
r.m.s	root mean square
sd	standard deviation
SIR	single isomorphous replacement
u.v.	ultra-violet (spectroscopy)
xs	excess

Chemical and Biochemical Names

ATP	adenosine 5'-triphosphate
BES	N,N-bis[2-hydroxyethyl]-2-aminoethane sulphonic acid
Cu/ZnSOD	copper/zinc superoxide dismutase
EDTA	ethylene diamine tetra-acetic acid, disodium salt
EMP	ethyl mercury phosphate
FeSOD	iron superoxide dismutase
MnSOD	manganese superoxide dismutase
-OAc	acetate
PCMBS	p-chloromercuribenzoate sulphonate
PEG	polyethylene glycol $[OH-(CH_2-CH_2-O-)_n-H]$
SOD	superoxide dismutase
Tris	2-amino-2-(hydroxymethyl)propane-1,3 diol

Crystallographic Notation

h, k, l	reciprocal lattice indices
u, v, w	Patterson space coordinates
x, y, z	real space coordinates

''During the last five years more papers have appeared on superoxide dismutase than on any other single enzyme.....''

Linus Pauling (1979) Trends Biochem. Sci., N270-271.

Reaction	Example
$O_2 \rightarrow H_2O$	(a) Energy Capture, Cytochrome Oxidase (Cytochrome chain reduction) (b) One Electron Reduction, Laccase (Copper chain reduction)
$O_2 \rightarrow O_2^{\cdot -}$	Reduction by NADH, Flavo-Enzyme
$O_2^{\cdot -} \rightarrow H_2O_2 + {}^{\cdot}OH$	Reduction by Metal Ions, Superoxide Dismutases (no other components)
$O_2 \rightarrow RO_2H$	Dioxygen Insertion, Dioxygenase
$O_2 \rightarrow ROH$	Hydroxylation, e.g., Cytochrome P-450, Mono-oxygenase (Requires subsidiary reductant such as NADH or α-ketoglutarate)

Table 1.1 Classification of reactions of oxygen in biological systems (from Williams, 1984).

Reaction Centre (X)	Example
Iron (a) Heme	Many: Cytochrome oxidase, Cytochrome P-450, Myoglobin
Iron (b) Non-heme	Many: Mono- and Dioxygenases, Hemerythrin, Superoxide Dismutase
Copper (a) Dimeric	Many: Oxidases, Laccase Hemocyanin, Superoxide Dismutase
(b) Monomeric	Amine Oxidases
Flavin (Pterin)	Oxygenases Superoxide production
Manganese	Oxygen release in chloroplasts Superoxide Dismutase

Note: A very different list of centers activate $O_2^{\cdot -}$ and H_2O_2.

Table 1.2 Reaction centres for dioxygen (from Williams, 1984).

CHAPTER 1

Superoxide Dismutases

1.1 Oxygen Toxicity

It has been hypothesised that life arose spontaneously three and half billion years ago from the reaction of basic chemicals stimulated by natural energetic processes such as ultraviolet radiation from the sun (Calvin, 1969). About one billion years later, blue-green algae had evolved and along with it appreciable amounts of oxygen derived from their photosynthetic activity. With the exception of a few anaerobes in oxygen-deficient environments, the oxygenation of the Earth's atmosphere constituted a stress that led to the development of organisms that could utilise oxygen. Up to eighteen times more energy in the form of ATP could be extracted by oxidation of glucose compared with anaerobic glycolysis.

The respiratory chain used to produce energy is based on the reduction of dioxygen to water. The reduction of ground state dioxygen normally follows an univalent pathway of reduction due to electronic spin restriction (Taube, 1965). A four-step reduction pathway is required and proceeds as follows:

$$X + O_2 \leftrightharpoons X.O_2 \xrightarrow{e} X.O_2^{\cdot -} \xrightarrow{} X(H_2O_2) \xrightarrow{e} (X).OH^{\cdot} \xrightarrow{} H_2O \quad (1)$$

In the course of the transport of dioxygen and its reduction to water, the carrier or catalyst X, can release these oxygen-derived intermediates. It has been conclusively demonstrated that all intermediates are produced in living organisms (Fridovich, 1979 and references therein). A summary of the reactions and reaction centres of dioxygen in biological systems is presented in tables 1.1 and 1.2.

The drawback for aerobic production of energy is that the partial

reduction products of dioxygen are potentially toxic. When generated in media surrounding cells in tissue culture, oxygen reduction products have been shown to lyse red blood cells, tumor cells, endothelial cells, fibroblasts and lymphoid cells (Simon et al.,1981 and references therein). The hydroxyl radical ($OH^{\bullet}$) reacts at almost diffusion-controlled rates ($k = 10^{8}-10^{10}$ $Lmol^{-1}s^{-1}$) with most biological substances (Anbar & Neta, 1967). The superoxide radical ($O_2^{-\bullet}$),produced in the first reduction step, is less reactive than $OH^{\bullet}$ but is known to react with protein sulfhydryls and polyunsaturated fats in lipids (Greenstock & Ruddock, 1978). It has been argued that $O_2^{-\bullet}$ has the potential to be more cytotoxic than $OH^{\bullet}$ because the latter being so reactive would be scavenged rapidly by reducing agents whereas the longer lived $O_2^{-\bullet}$ could diffuse to vulnerable parts of the cell (Fridovich, 1979). Hydrogen peroxide (H_2O_2) is more stable than either $O_2^{-\bullet}$ or $OH^{\bullet}$ but there is evidence for its toxicity at <u>in vivo</u> concentrations (for example, the death of human fibroblasts (Simon et al., 1981)). The toxicity of dioxygen may also express itself via the Haber-Weiss reaction in which the less reactive $O_2^{-\bullet}$ and H_2O_2 can react to form $OH^{\bullet}$ as follows:

$$O_2^{-\bullet} + H_2O_2 + H^{+} \longrightarrow H_2O + O_2 + OH^{\bullet} \qquad (2)$$

However this reaction is slow under normal physiological concentrations so that its <u>in vivo</u> significance has been considered in terms of possible metal-catalysed versions of reaction 2 (Rigo et al., 1977). For example, the Fenton reaction:

$$Fe(chelate)^{3+} + O_2^{-\bullet} \longrightarrow Fe(chelate)^{2+} + O_2 \qquad (3)$$

$$Fe(chelate)^{2+} + H_2O_2 \longrightarrow Fe(chelate)^{3+} + OH^{-} + OH^{\bullet} \qquad (4)$$

There are a number of other candidates for toxic oxygen species including

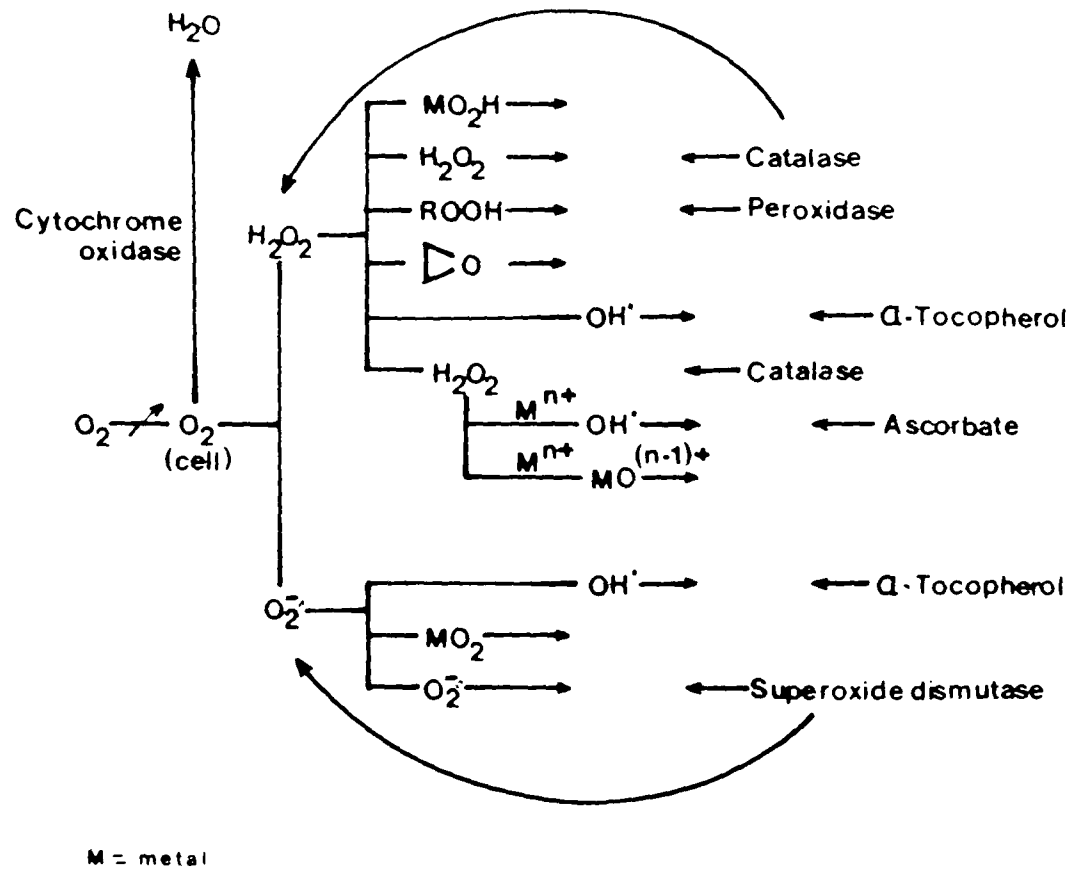


Figure 1.1 Flow diagram of the antioxidant defence screen (from Greenwood & Hill, 1982).

the conjugate acid of superoxide, the perhydroxyl radical (Gebicki & Bielski, 1981), singlet oxygen (Halliwell, 1981) and metal-oxygen complexes (Youngman, 1984). Indeed dioxygen itself has been proposed as the source of oxygen toxicity (Haugaard, 1968). Both amidophosphoribosyl transferase (involved in purine biosynthesis) and dihydroxyacid dehydratase (involved in branched chain amino acid biosynthesis) are inactivated by dioxygen *in vivo* (Itakura & Holmes, 1979; Brown & Yein, 1978). The source and nature of oxygen toxicity is a topic of heated debate (Fee, 1982; Halliwell, 1982a) but suffice it to say that antioxidant defences are required for superoxide and hydrogen peroxide, which if not toxic *in vivo* themselves, are considered to be precursers of toxic oxygen species.

The production of superoxide in aerobic organisms is controlled by the scavenging effects of superoxide dismutases by the overall reaction:

$$2O_2^{\cdot -} + 2H^+ \longrightarrow H_2O_2 + O_2 \qquad (5)$$

and the decomposition of hydrogen peroxide is carried out by catalase or peroxidases as follows:

$$H_2O_2 + AH_2 \longrightarrow 2H_2O + A \qquad (6)$$

where A is dioxygen in the case of catalase. These enzymes are thought to constitute the primary defence against oxygen toxicity. There exists a secondary defence screen of small nucleophilic molecules which include ascorbic acid, reduced glutathione, α-tocopherol (vitamin E), uric acid and β-carotene (Ames et al., 1981; Peto et al., 1981). The organisation of the antioxidant defence system against potentially toxic oxygen species is summarised in figure 1.1. There is considerable evidence of cooperation between the various constituents of the defence screen (Shimizu et al., 1984).

1.2 Superoxide Dismutase as an Antioxidant Defence Enzyme

Superoxide dismutase (SOD, E.C.1.15.1.1) was first isolated as a blue-green copper protein from bovine erythrocytes and liver by Mann and Keilin in 1939. The isolated proteins were named erythrocuprein and hepatacuprein respectively. It was not until 1969 that McCord and Fridovich assigned a function to erythrocuprein (McCord & Fridovich, 1969a). They observed that erythrocuprein was a competitive inhibitor of cytochrome c reduction by the superoxide-producing enzyme, xanthine oxidase. Two further dismutases were discovered by Fridovich and coworkers; a manganese dismutase (MnSOD) in 1970 and an iron dismutase (FeSOD) in 1973 (Keele et al., 1970; Yost & Fridovich, 1973).

The spontaneous rate of superoxide dismutation is fast ($2x10^5$ $M^{-1}sec^{-1}$ at pH 7.4 (Fridovich, 1975)) so why is an enzyme required to catalyse the reaction? Is the main function of SOD therefore the disproportionation of superoxide or is it incidental to some other function? The fact that superoxide is an important agent of oxygen toxicity and that SOD provides the first line of defence is supported by experimental evidence.

The first test of the theory came from a survey of organisms for SOD and catalase. It was reasoned that those organisms that utilise oxygen must maintain defences whereas those organisms that never encounter oxygen should lack a defence. Initial surveys demonstrated that aerobes contained more SOD and catalase than aerotolerant organisms (McCord et al., 1971). However later, anaerobes were found to contain SOD and this was explained by pointing out that anaerobes have widely different sensitivities to oxygen (O'Brien & Morris, 1971). A positive correlation was found between the degree of aerotolerance of some anaerobes and their content of SOD (Tally et al., 1977). The lack of SOD in some aerotolerant species was allied to the fact that they could not utilise oxygen at all (Gregory & Fridovich, 1974). The absence of SOD and catalase in the aerotolerant Lactobacillus

plantarum was disturbing until millimolar levels of Mn(II) were found in the organism (Archibald & Fridovich, 1981). When deprived of Mn(II), L.planterum exhibited oxygen-intolerance (Archibald & Fridovich, 1981).

The next line of evidence came from induction experiments. It was predicted that exposure to high levels of oxygen might induce higher levels of SOD. The exposure of Escherichia coli (Gregory et al.,1973), Streptococcus faecalis (Gregory & Fridovich, 1973a), Bacteroides fragilis (Privalle & Gregory, 1979) and numerous other organisms to high levels of oxygen led to increased intracellular levels of SOD and these levels correlated with increased resistance to oxygen toxicity. Higher levels of SOD also correlated with an increased resistance towards oxygen enhancement of the toxicity of the tumor antibiotic, streptonigrin to these organisms (Gregory & Fridovich, 1973b). Further, the concentration of SOD in E.coli could be controlled by changing culture conditions (Hassan & Fridovich, 1977a and b). Increased levels of SOD under constant oxygen pressure correlated with cell resistance towards lethality of hyperbaric oxygen (Hassan & Fridovich, 1977a). Intracellular levels of superoxide were increased artificially in E.coli cells by paraquat. This led to rapid induction of SOD but when the induction was inhibited by chloramphenicol the oxygen intolerance was returned . Removal of paraquat from the growth medium caused a sharp decline in the rate of SOD synthesis (Hassan & Fridovich, 1977c). Mutants of E.coli exhibiting a temperature-dependent decline in SOD concentration were shown to have a parallel temperature sensitivity towards tolerance of oxygen (Hassan & Fridovich, 1979). It has been shown only recently that insertional inactivation of the yeast MnSOD gene results in cells unable to survive exposure to high levels of oxygen (Marres et al., 1985).

Finally it should be noted that at physiological pH, all three types of SOD are more effective catalysts than are equivalent amounts of copper, manganese or iron salts respectively (Halliwell, 1981). No other copper

proteins are known to react catalytically with superoxide (Halliwell, 1982b) and despite an extensive search, the SOD enzymes have not been found to act catalytically on any substrate other than superoxide (Wardman, 1978; Halliwell, 1982b).

1.3 Clinical Aspects

The question of whether superoxide is an initiating or contributing factor in disease is part of the more general problem of the nature and pathogenicity of toxic oxygen discussed in section 1.1. Michelson and coworkers have put forward a concept of superoxide-related disease according to which both high and low levels of SOD can be pathogenic (Michelson, 1979). Their arguments imply that SOD strikes a balance between utilisation and elimination of superoxide. The requirements for balance can be understood readily if one accepts that free radicals actually play a role in living systems (Bannister, 1984). Oxygen-free radicals have been shown to play a role in a variety of physiological processes: For example, chemotaxis (Synderman & Goetzl, 1981; McCord & Roy, 1982), the inflammatory response (McCord, 1974; McCord & Roy, 1982), histamine release from mast cells (Ohmori et al, 1979), microsomal detoxification of chemicals (Cohen & Cedarbaum, 1979) and the bactericidal activity of granulocytes (Baehner et al., 1970, Johnston & Baehner, 1970; Baboir, 1982). Free radical modification of DNA may be necessary for heritable change (Witting, 1980).

The level of SOD in cells has been implicated in a wide variety of diseases (see Bannister, 1984 for a review). Well documented examples are haemolytic anaemia, neurological disorders, Down's syndrome, inflammatory and autoimmune diseases. Michelson and coworkers have been exploring potential clinical roles for SOD (Michelson, 1979). They have identified a number of factors that would play a key role in any pharmacological applications of SOD. These include circulation lifetime, penetration of cells, intracellular localisation of exogenous SOD and organ specificity. A

promising method of introduction to cells is the use of encapsulated SOD in liposomes which can be fused with cells and directed to specific organs by appropriate modifications of the surface of the liposomes (Michelson, 1979). Cu/ZnSOD (under the generic name Orgotein) has showed strong anti-inflammatory properties with almost no side effects in clinical trials (Huber & Menander-Huber, 1980). In particular, results have been encouraging for patients with rheumatoid arthritis and osteoarthritic complaints (Goebal et al., 1983). SOD has been implicated in a number of parasitic diseases; leishmaniasism (which affects 400,000 people a year), trypanosomiasis (LeTrant et al., 1983) and leprosy (Kusunose et al., 1980). The parasites from these diseases are thought to survive phagocytosis due to high levels of FeSOD. The inhibition of FeSOD may help in the cure of these diseases. SOD may play roles in both cancer and aging (for a review, see Florence, 1983). MnSOD levels in undifferentiated and cancer cells are lower than normal cells (Oberley & Buettner, 1979). Oberley and coworkers have proposed that antitumor agents that produce free radicals exert their cytotoxic effect through the radicals and because tumor cells have low levels of key antioxidant defences, the antitumor drugs preferentially kill tumor over normal cells (Oberley et al., 1983). A model of chemical carcinogenesis has been proposed involving MnSOD and gene amplification (Oberley & Oberley, 1984) but its relevance to human cancer conditions has been questioned (Halliwell & Gutteridge, 1984). It has been suggested that the mitochondria, where over 90% of oxygen consumed by mammals is utilised, may serve as a biological clock and that the maximum life span maybe a reflection of the rate of mitochondrial degradation (Miquel et al., 1980). Supporting evidence includes the association of rapid aging features with the decline of mitochondrial MnSOD activity in trisomy 21 (Down's syndrome). It is known that old people have a lower amount of antioxidant defences and that longer potential lifespan correlates with high ratios of SOD to superoxide in adult tissue of rodents and primates (Tolmasoff et

Organism	Photo-synthetic	Aer-obic	Superoxide dismutase			
			Cu/Zn	Mn and/or Fe	Mn	Fe
Prokaryotes						
Anaerobic bacteria	–	–	–			–
Chlorobiaceae (green sulphur bacteria)	+	–	–	+		
Chromatiaceae (purple sulphur bacteria)	+	–	–	+		
Rhodospirilliaceae (purple non-sulphur bacteria)	+	+	–		+	
Cyanophyta (blue-green algae)	+	+	–		+	+
Aerobic bacteria	–	+	–		+	–
Eukaryotes						
Rhodophyta (red algae)	+	+	–	+	+	
Chlorophyta (green algae)	+	+	–	+		
Charophyta	+	+	–	+		
Euglenophyta	+	+	–	+		+
Phaeophyta (brown algae)	+	+	–	+		
Chrosophyta (diatoms)	+	+	–	+		
Protozoa	–	+	–	+		
Fungi	–	+	+		+	+
Higher plants	+	+	+	+		
Higher animals	–	+	+		+	+

Table 1.3 The distribution of SODs in organisms (from Harris et al., 1980).

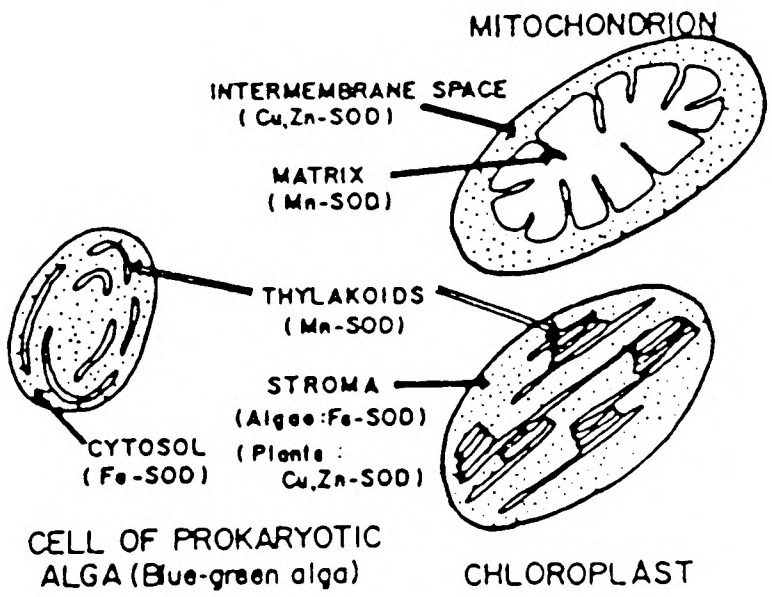


Figure 1.2 Subcellular distribution of SODs (from Asada et al., 1980).

al., 1980).

1.4 Distribution

The distribution of the three main types of SOD in organisms has been determined mainly by isolation and characterisation. However a combination of reactivities of the three enzymes to cyanide and hydrogen peroxide allows a determination of the prosthetic metal of the SOD in the crude extract. Cyanide inhibits Cu/ZnSOD whilst both Fe and Cu/ZnSOD are inhibited by hydrogen peroxide. MnSOD is insensitive to both reagents.

The distribution of SOD appears to reflect the phylogeny of organisms (see table 1.3). The Cu/Zn protein is considered to be found characteristically in the cytoplasm of all higher eukaryotic cells. The Mn protein is present in prokaryotes and in the mitochondria and chloroplasts of eukaryotic cells. The Fe protein is found normally in prokaryotes. Some prokaryotes contain both Fe and MnSODs (Yost and Fridovich, 1973). The Fe type appears to be preferentially located in the periplasmic space and the Mn type in the cell matrix of prokaryotic cells (Gregory et al., 1973). A number of exceptions to these distribution patterns have recently come to light. The Fe protein ,in addition to the other two, has been isolated from the mustard plant Brassica campestris (Salin & Bridges, 1980) and a survey of forty three plant families for the presence of FeSOD showed that it was present in three isolated families (Bridges & Salin, 1981). Recent experimental evidence suggests Cu/ZnSOD is also associated with lysosomes (Geller & Winge, 1982). Two bacterial species, Photobacterium leighathi and Caulobacter crescentus, have been shown to contain Cu/ZnSOD (Puget & Michelson, 1974; Steinman, 1982) although this form of the enzyme has been found to be absent from eukaryotic species more primitive than green algae such as protozoan and Euglena species (Asada et al., 1975). The reader is referred to a recent review for a detailed discussion of the distribution of SOD in living systems (Bannister and Rotilio, 1984).

There are a few interesting observations regarding the presence of Fe and MnSODs in early lifeforms. The presence of FeSOD in obligate photosynthetic anaerobes (species sensitive to oxygen pressures 0.2atm or less) is due probably to the bacteriostatic effect of oxygen rather than any bacteriocidal effect (Hewitt & Morris, 1975). The presence of MnSOD in hydrogenosomes of anaerobic ameobae is surprising because MnSOD tends to be found in more recently evolved prokaryotes (Lindmark & Miller, 1974). A few strains of Neisseria gonorrhoeae contain no SOD but are exceptionally rich in catalase which presumably compensates for SOD by disproportionating the hydrogen peroxide required in metal-catalysed reactions that produce hydroxyl radicals (see reaction 4, Norrod & Morse, 1979). Gram-positive and gram-negative bacteria cannot be distinguished on the basis of whether they contain Fe or MnSOD (Britton et al., 1978).

SOD occurs in high oxygen utilisation areas of mammals such as in the brain, liver, heart, erythrocytes and kidney (McCord & Fridovich, 1969b; Marklund, 1984). SOD is mainly intracellular with only traces in extracellular fluids (Marklund, 1984). FeSOD is found localised in the cell matrix of bacteria and in the cytosol of blue-green algae. MnSOD is localised in the thylakoids of chloroplasts of plants whilst FeSOD is found in the stroma of lower plants and Cu/ZnSOD in the stroma of higher plants. The localisation of MnSOD in algal lamellae maybe related to the Mn-requiring apparatus in the photosystem II of plants (see figure 1.2; Kanematsu & Asada, 1979 and references therein).

1.5 Evolutionary Aspects

The distribution of the three types of SOD in organisms is considered to be characteristic of the evolutionary stage of the organism and also of the cell organelle with which it is associated. If the distribution is linked to the evolutionary history of the organism, then the distribution of the various types of SODs have to be considered in the light of the

acquisition of a permanent defence against any form of toxicity arising from the increase of photosynthetic evolution of atmospheric oxygen.

The most primitive form of SOD is considered to be the Fe enzyme because of its presence in a range of anaerobic bacteria (Hewitt & Morris, 1975). The presence of SOD in photosynthetic anaerobes, sulphate reducers and fermentative anaerobes has been explained by the need to protect against low levels of oxygen and perhaps occasionally against higher levels of oxygen when these organisms move from one anaerobic niche to another (Hassan & Fridovich, 1977c). Thus for example, the level of FeSOD in the anaerobe, Bacteroides fragilis, increases in the presence of oxygen (Privalle & Gregory, 1979). Other workers have pointed out that superoxide may be generated from water in anaerobes by the action of ultraviolet light and thus require SOD (Kanematsu & Asada, 1978).

It is interesting to note that calculations suggest that both free Fe^{2+} and Mn^{2+} were available in abundant concentrations before the accumulation of oxygen in the biosphere but copper was tied up in the form of highly insoluble cuprous sulphides (Williams, 1976). One might speculate that this is the reason why Fe and MnSODs are the primitive forms and Cu/ZnSOD is a more recent adaption (Osterburg, 1974; Egami, 1975). Williams (1982) has argued that the presence of Cu/ZnSOD in the cytoplasm of eukaryotes was the result of a selective advantage over the Fe and MnSODs as eukaryotic cells evolved. The low redox potential of the cytoplasm ensured that Fe and Mn existed in their resting states as divalent metals which could readily dissociate from their proteins because of their weak binding constants. The evolution of Cu/ZnSOD held a selective advantage because DNA is of greater importance in longer-lived multicellular organisms and free Fe and Mn are mutagenic (El-Deiry et al., 1984).

Physico-chemical and amino acid sequencing studies suggest that Mn and FeSODs are related closely (Harris et al., 1980; Parker et al., 1984; see appendix 4). The synthesis of Fe and MnSOD depends solely upon the metal

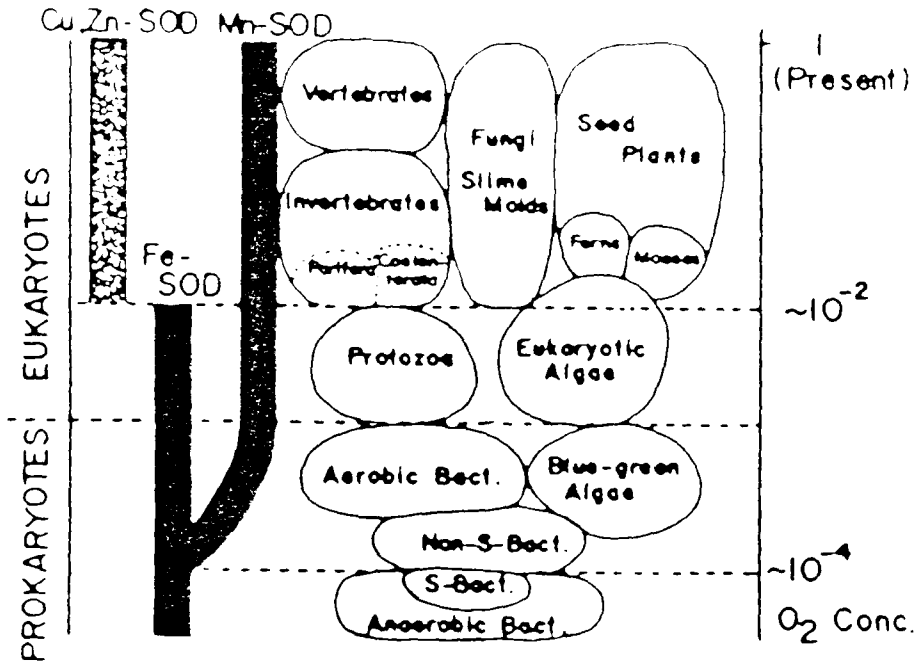


Figure 1.3 Phylogenic distributions of SODs in organisms at various stages of evolution (from Asada et al., 1980).

supplied in the growth medium of Propionibacterium shermanii (Meier et al., 1982) and it is has been shown that the metal in FeSOD from Bacteroides fragilis can be substituted by Mn without loss in activity (Gregory & Dapper, 1983). Nevertheless, it must be pointed out that these observations contrast with other studies which show that the Fe and Mn proteins in Plectonina boryania and E.coli can be distinguished antigenically (Okada et al., 1979; Fee, unpublished results) and metal-substituted FeSODs do not normally retain their activity. On the other hand, the Cu/Zn and Mn,Fe proteins have very different primary, secondary and tertiary structures so that these two types of SOD must have evolved independently from different ancestors (see figure 1.3, Parker et al., 1984).

The reason for the selection of Mn over Fe as the prosthetic group in mitochondrial SODs is unclear. Mitochondria are known to concentrate very efficiently both Fe and Mn (Williams, 1982). However, it has been observed that MnSOD is very resistant to hydroxyl radical damage in contrast to FeSOD and the production of significant amounts of hydroxyl radical in the mitochondria is quite likely since the organelle generates the necessary precursors, dioxygen and hydrogen peroxide (Bannister et al., 1983).

The presence of Mn and FeSODs in mitochondria and chloroplasts has been presented as evidence in favour of the endosymbiotic theory of mitochondrial and chloroplast origin (Steinman & Hill, 1973; Fridovich, 1974; Kwiatowski et al., 1985). The theory proposes that the present-day organelles of eukaryotic cells, such as mitochondria and chloroplasts, evolved from bacteria and blue-green algae, which existed as intracellular symbionts with primitive eukaryotic organisms. As the atmosphere of the primitive earth changed from anaerobic to aerobic conditions, these symbionts were to have been acquired as a means of confronting selection pressures that necessitated oxidative metabolic pathways (Sagan, 1967; Raven, 1970; Raff & Mahler, 1972; Doolittle, 1980). A common ancestral gene for the Fe and MnSODs has been proposed because of the similarity between

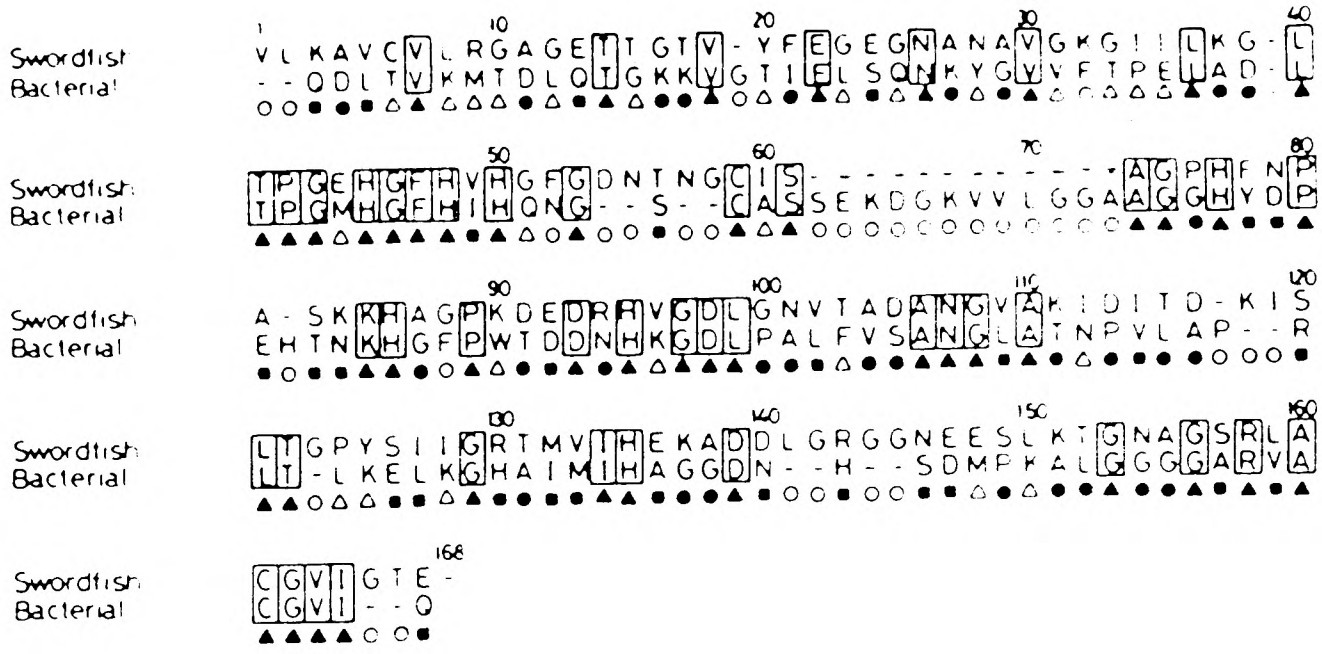


Figure 1.4 Amino acid sequences of swordfish liver (Rocha et al., 1984) and bacterial (Steffens et al., 1983) Cu/ZnSODs. Identical residues are boxed. Symbols refer to the natural substitution frequencies of amino acids as determined by McLachlan (1971): ▲, identity; ■, frequently found; ●, neutral; △, infrequent; ○, rarely found or no substitution (from Bannister & Parker, 1985; see appendix four for details).

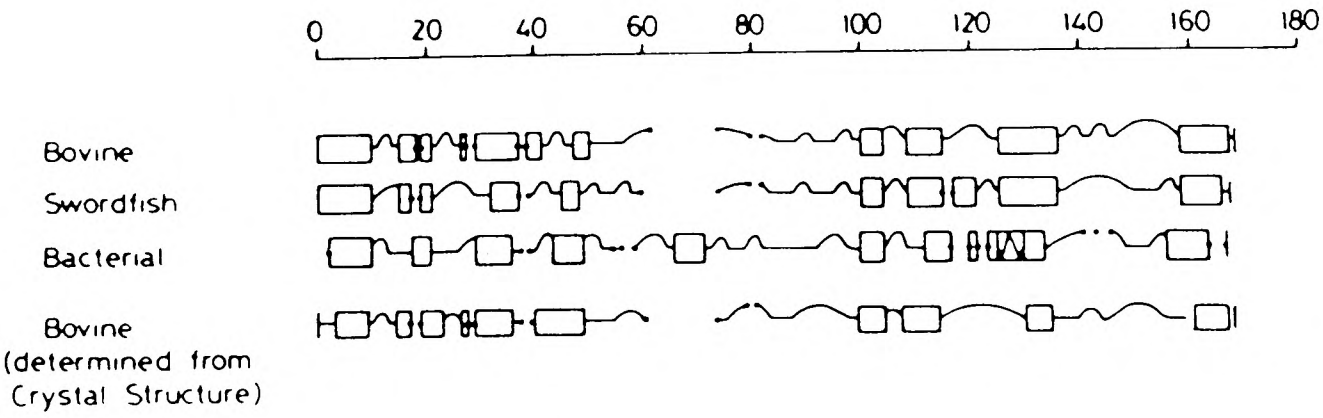


Figure 1.5 Comparison of predicted secondary structures of bovine erythrocyte, swordfish liver and bacterial Cu/Zn SODs. □ β-strand; ∩ β-turn (from Bannister & Parker, 1985; see appendix four for details).

the mitochondrial and bacterial Fe and MnSODs and their dissimilarity to the eukaryotic Cu/ZnSODs. The ancestral gene must have resided in the genome of the symbiotic bacteria inhabiting the primitive eukaryotic cell. Support for this conclusion has come from the observation that FeSOD in plant cells is located in the plasmid rather than the cytosol (Salin & Bridges, pers. commun.)

The evolution of SOD appears to approximate the evolution of a substantial portion of the bacterial genome on the basis of immunological studies (Bang et al., 1981). The presence of Cu/ZnSOD in Caulobacter crescentus is as yet an unexplained anomaly (Steinman, 1982). However,the presence of the enzyme in Photobacterium leiognathi has been considered to be a case of horizontal gene transfer from eukaryotes to prokaryotes (Martin & Fridovich, 1981; Bannister & Parker, 1985; see appendix 4). The bacterial species although found free living, is also a symbiont of the ponyfish (Haneda, 1950). The Cu/ZnSOD from this bacterium has a similar molecular weight, subunit size, metal content and catalytic activity as the eukaryotic enzymes. The single disulphide bridge, the seven metal ligands, and two residues shown to be important in the mechanism of the bovine erythrocyte enzyme are all found in the bacterial enzyme in the correct sequence order (figure 1.4). Also very similar secondary structure is predicted for the bacterial, swordfish (close relation of the ponyfish) liver and bovine erythrocyte enzymes (figure 1.5; Parker et al., 1984). Analysis of aligned sequences showed more than half of the bacterial polypeptide chain had closely related or identical residues with the swordfish sequence after deletions and insertions were excluded (figure 1.4; Bannister & Parker, 1985). Numerous species of marine photobacteria, closely related to P.leiognathi, have been examined and found to lack any detectable Cu/ZnSOD (Bang et al., 1978).

Recently it has been shown that SODs cannot be used as a satisfactory 'molecular evolutionary clock' because of greatly varying rates of amino

acid substitutions between the different sequences (Lee et al., 1985). The apparent rate of substitutions is much greater through the evolution of mammals than at earlier times. These observations have implications for the use of sequence data in reconstructing phylogenetic relationships and timing evolutionary events where it has been assumed that proteins evolve at a constant rate.

1.6 Chemical Characterisation

SODs have been characterised according to their metal content. Cu/ZnSOD is a homodimer of about 32,000 Daltons containing up to 1g atom of both Cu and Zn per subunit. (Recently a 135,000 Dalton glyco-protein containing four Cu atoms has been isolated from human lung tissue which may turn out to be a new type of SOD (Marklund, 1982, 1984 and 1985)). It is relatively stable in organic solvents and this property was used in its extraction from bovine erythrocytes in 1939 (Mann & Keilin, 1939). The protein is known to be stable to most proteases at neutral pH (Bannister, pers. commun.). The subunits of the protein were shown to be associated by strong non-covalent interactions (Abernethy et al., 1974). The isoelectric point of the enzyme varies between 4.6 and 6.8 depending on the species (Salin & Wilson, 1981). Both cyanide and hydrogen peroxide inhibit the enzyme (see table 1.6). The Cu-binding sites have been characterised by esr, nmr and uv spectroscopy (see Bannister & Rotilio, 1984; Fielden & Rotilio, 1984 for reviews). The metal atoms can both be removed at low pH in the presence of chaotropic agents. The activity of apoCu/ZnSOD can be almost fully restored (c.a. 95%) by reconstition of the Cu^{2+} ion and restored completely by the reconstition of both metal ions (Beem et al., 1974). The activity of the protein cannot be retained if the Cu^{2+} ion is replaced by other metal ions but replacement of the Zn^{2+} ion by Co^{3+}, Hg^{2+} or Cu^{2+} results in no loss of activity (Fee, 1973; Rotilio et al., 1973; Forman & Forman, 1973; Beem et al., 1974). On the basis of these and other results it has been concluded

Source	Metal content	Molecular weight
Escherichia coli B	1 g atom Fe	38,000
Plectonema boryanum	2	41,700
	1	36,500
Bacillum megaterium	1	40,000
Pseudomonas ovalis	1.4	40,000
Desulphovibrio desulfuricans	1.60	43,000
Photobacterium leiognathi	1.61	40,660
Chromatium vinosum	2	41,000
Chlorobium thiosulfatophilum	1.8	43,000
Thiobacillus denitrificans	1.35	43,000
Anacystis nidulans	1	37,000
Crithidia fasciculata	1.4	43,000
Strepthococcus mutans	2 g atoms Mn	39,500
Escherichia coli	1.6	39,500
Pleurotus olearius	2	76,000
Chicken liver mitochondria	2.3	80,000
Saccharomyces cerevisiae	4	96,000
Mycobacterium lepraemurium	1.29	45,000
Thermus aquaticus	2	80,000
Rat liver mitochondria	4	89,000
Bovine heart mitochondria	2	86,000
Pisum sativum	1	94,000
Paracoccus denitrificans	1.34–2	41,500

Table 1.4 Metal content of Fe and MnSODs from various sources (from Parker et al., 1984 and references therein; see appendix four).

Source	Metal content		Molecular weight
Mycobacterium phlei	1.7 g atoms Mn 1.2 g atoms Fe	0.7 g atoms Zn	80,000
Thermoplasma acidophilum	2.0 g atoms Fe	1.0 atoms Zn	82,000
Methanobacterium bryantii	2.7 g atoms Fe	1.7 g atoms Zn	91,000
Nocardia asteroides	1.2 g atoms Fe 1–2 g atoms Mn	1–2 g atoms Zn	100,000

Table 1.5 Sources of SODs which have atypical metal contents (from Parker et al., 1984 and references therein; see appendix four).

that the Zn^{2+} ion is not essential for catalysis but does play a structural role (Pantoliano et al., 1982).

The Fe and MnSODs have a subunit molecular weight of about 19-24,000 Daltons and can be isolated as dimers or tetramers of identical subunits. The metal content of the Fe and Mn enzymes varies between 1.0 and 2.0 grams atoms per dimer (see table 1.4). It is unclear whether this variation is due to metal loss during purification steps. A number of anomalies in the metal composition of some FeSODs have turned up; some also contain Zn and two examples contain Mn (table 1.5). Numerous epr studies of FeSOD from various sources have shown that the Fe exists in the high spin ferric form in a nearly rhombic environment (Yost & Fridovich, 1973; Lumsden & Hall, 1974; Asada et al., 1975). The distinct red-wine colour of MnSOD had suggested that the metal was in the tervalent state and this was latter confirmed by magnetic susceptibility and epr results (Keele et al., 1975; Fee et al., 1976). The optical spectra of FeSODs are characterised by a broad band near 350 nm that is attributed to a ligand-to-metal charge transfer band (Gaber et al., 1974). Most MnSODs have two absorption bands; one is around 450 - 480nm with molar extinction coefficients of 240 $M^{-1}cm^{-1}$ to 900 $M^{-1}cm^{-1}$ and the other is around 600nm (Ellman, 1959; Sato & Harris, 1977; Sato & Nakazawa, 1977). The Fe^{3+} and Mn^{3+} ions can be removed and replaced by other metal ions similar to the methods used with Cu/ZnSODs although the Fe and Mn proteins tend to be less stable than the Cu protein. Metal replacement studies have been performed on MnSOD from E.coli (Ose & Fridovich, 1976 and 1979), T.aquaticus (Sato & Harris, 1976) and B.stearothermophilus (Brock et al., 1976; Brock & Harris, 1977). Both apoprotein and reconstituted products were indistinguishable from the native protein as judged by gel filtration, polyacrylamide-gel electrophoresis and nmr. In all cases the reconstituted protein contained one trivalent ion (Mn,Fe,Co,Ni and Cu) per dimer but only the protein reconstituted with Mn showed any specific activity . Futhermore the

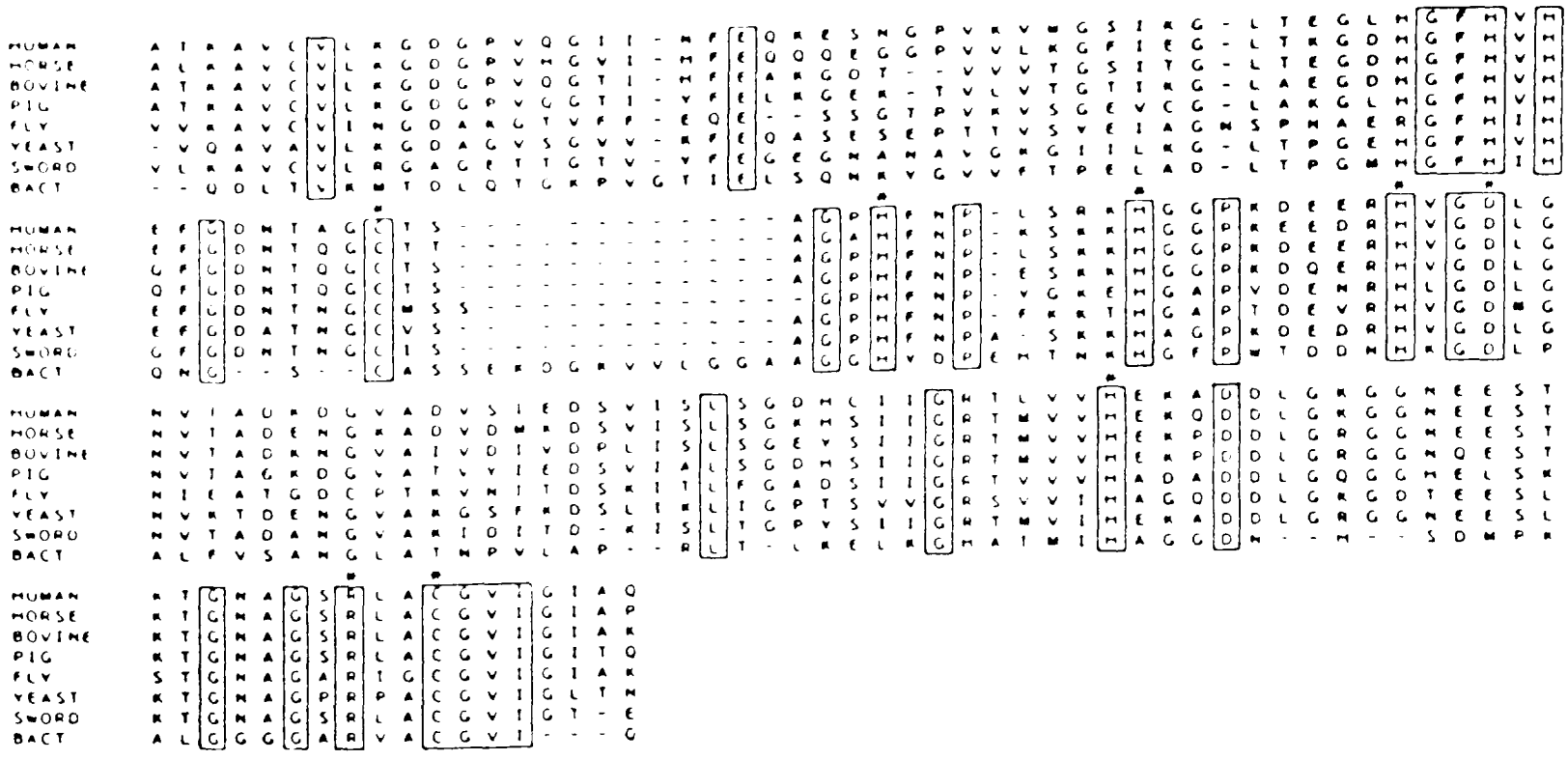
	Cyanide	Hydrogen Peroxide	Azide	Nitroprusside
Cu/Zn	inhibited	inactivated	partially inhibited	inhibited
Fe	no effect	inactivated	inhibited	inhibited
Mn	no effect	no effect	partially inhibited	inhibited

Table 1.6 Inhibition of SODs.

reconstituted Mn protein displayed full activity relative to the native enzyme. Epr studies of copper-substituted MnSOD support a tetrahedral arrangement of three histidine and one carboxylic residue as ligands to the metal (Barra et al., 1985).

Metal replacement studies on FeSOD from Ps.ovalis (Yamakura & Suzuki, 1980; Yamakura et al., 1980) and P.leiognathi (Puget et al., 1977) produced similar results except that the Mn-substituted enzyme was inactive and the Fe reconstituted enzyme regained full activity. The molar extinction coefficients of the Mn-substituted proteins at 480nm were a third to a sixteenth of the value for native MnSODs. Recently a FeSOD from B.fragilis has been reconstituted with Mn (1gram Mn per dimer) and shown to have full catalytic activity (Gregory & Dapper, 1983). Evidence from azide binding and kinetics of substrate dismutation have suggested that the Fe atoms in FeSODs act independently of one another (Fee et al., 1981a; Fee et al., 1981b). FeSOD can be inhibited by hydrogen peroxide, azide and fluoride (Misra & Fridovich, 1978; Fee et al., 1981a) whilst there are no known total inhibitors of MnSOD (see table 1.6). Nitroprusside cannot be regarded as a true inhibitor (table 1.6) because its mechanism of action involves competitive reaction with superoxide (Misra, 1984). The kinetics of azide binding indicates that azide inhibits FeSOD by competing with superoxide for an anion-binding site close but not on the iron site (Fee et al.,1981b). Early chemical modification studies have implicated histidine and tryptophan residues in the active site of FeSOD (Vanopdenbosch & Crichton, 1977). This has been supported by observations that hydrogen peroxide reacts with two tryptophan residues in inactivating FeSOD (Yamakura, 1984). Circular dichroism measurements suggest the Fe and MnSODs have about 35% helix and a little sheet (Anastasi et al., 1976; Yamakura, 1976; Sato & Nakazawa, 1978).

Most SOD assays are indirect because of the reactive nature of the substrate, ie. assays based on the fact that SOD decreases the rate of some



		Human	Horse	Bovine	Pig	Fly	Yeast	Swordfish	Bacterial
1	Human								
2	Horse	78							
3	Bovine	82	78						
4	Pig	82	75	84					
5	Fly	58	55	55	55				
6	Yeast	55	56	53	52	50			
7	Swordfish	66	66	71	67	54	53		
8	Bacterial	25	26	28	28	24	26	30	

F

Figure 1.6 Amino acid sequences of Cu/ZnSODs.

1. Human erythrocyte (Barra et al., 1980; Jabusch et al., 1980; Sherman et al., 1983).
2. Horse liver (Lerch & Ammer, 1981).
3. Bovine erythrocyte (Steinman et al., 1974).
4. Pig erythrocyte (Hering et al., 1985; Schinina et al., 1985).
5. Drosophila melanogaster (Lee et al., 1985).
6. Saccharomyces cerevisiae (Johansen et al., 1979; Steinman, 1980)
7. Swordfish liver (Rocha et al., 1984).
8. Photobacterium leiognathi (Steffens et al., 1983).

and Table 1.7 Sequence identities (%) between Cu/ZnSODs.

reaction that is mediated by superoxide. The original assay was based on the ability of SOD to inhibit the reduction of cytochrome c by superoxide which was generated by the action of xanthine oxidase on xanthine (McCord & Fridovich, 1969a). The disavantage of this method was the interference of the cytochrome c activity by cytochrome oxidase and peroxidase in crude extracts (Weisiger & Fridovich, 1973). Cytochrome c was replaced by other reagents such as nitroblue tetrazolium (Beauchamp & Fridovich, 1971). Other indirect assays have utilised the ability of SOD to inhibit the autoxidations of epinephrine (Misra & Fridovich, 1978), sulfite (McCord & Fridovich, 1969b), 6-hydroxydopamine (Heikkila & Cappat, 1976) and pyrogallol (Marklund & Marklund, 1974). There are now a few direct methods available but these suffer from the requirement of expensive instrumentation and a high degree of sample purity because the lifetime of superoxide is drastically shortened by the presence of metals. Examples are the use of esr (Misra & Fridovich,1978), pulse radiolysis (Fielden et al., 1974), polarography (Rigo et al., 1975), stop-flow spectrophotometry (McClune & Fee, 1978) and ^{19}F nmr (Rigo et al., 1979).

1.7 Structural Studies

The complete amino acid sequences of eight Cu/ZnSODs have been determined and are shown in figure 1.6. The enzymes from higher species have a blocked N-terminal group and tend to be more stable than the enzymes from lower sources. There is considerable variation in tyrosine and tryptophan content between the sequences. The three-dimensional structure of one, from bovine erythrocytes, has been determined to high resolution (refined at 2.0Å) (Tainer et al., 1982; Tainer et al., 1983). Each subunit is composed of eight antiparallel β-strands that form a flattened barrel plus three external loops of non-repetitive structure (figures 1.7 and 1.8). The Cu and Zn atoms are 6Å apart and the two Cu atoms of the dimer are 34Å apart. The barrel adopts a Greek key topology and the overall

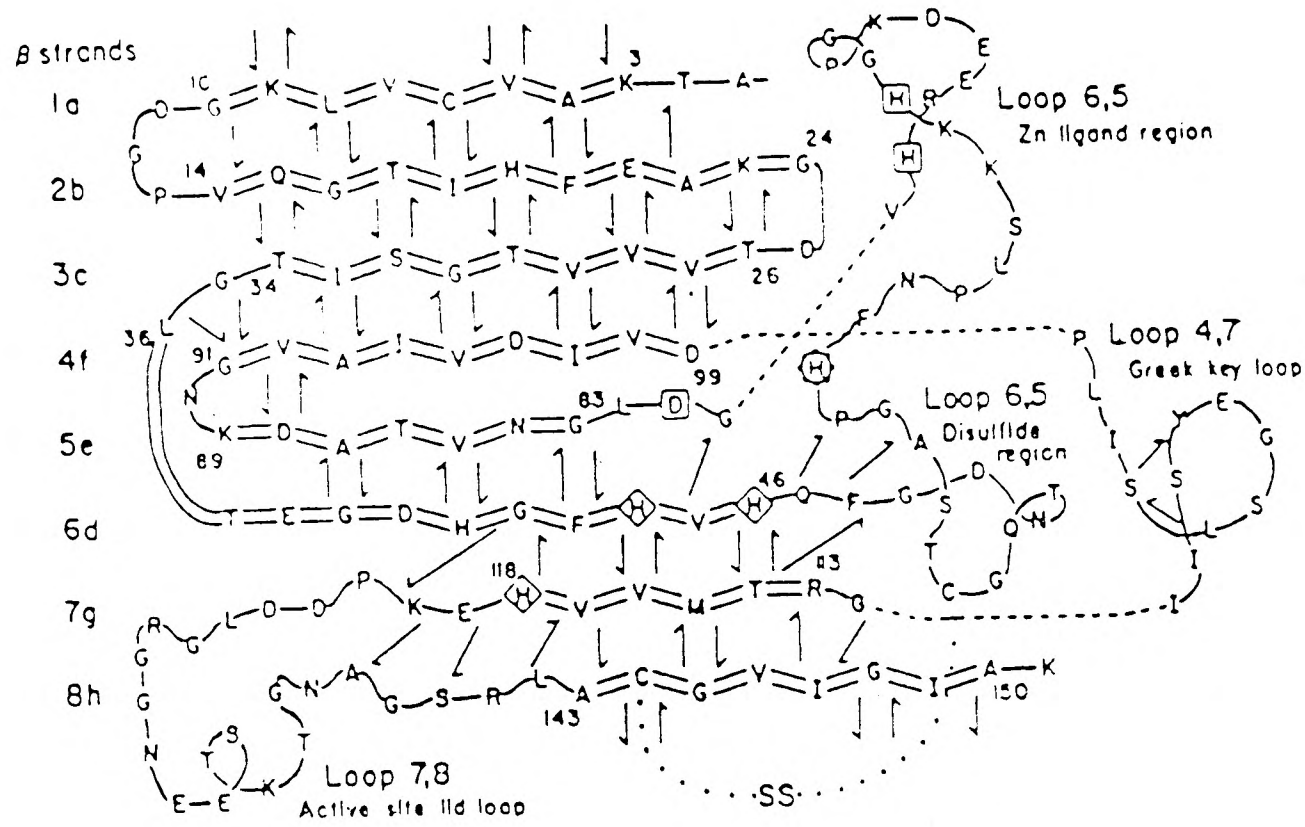


Figure 1.7 Schematic diagram of the X-ray crystal structure of bovine erythrocyte Cu/ZnSOD (from Tainer et al., 1982). β-strands are drawn as zigzag lines with residues in the 'valley' of the zigzag pointing into the interior of the β-barrel. Important residues are boxed.

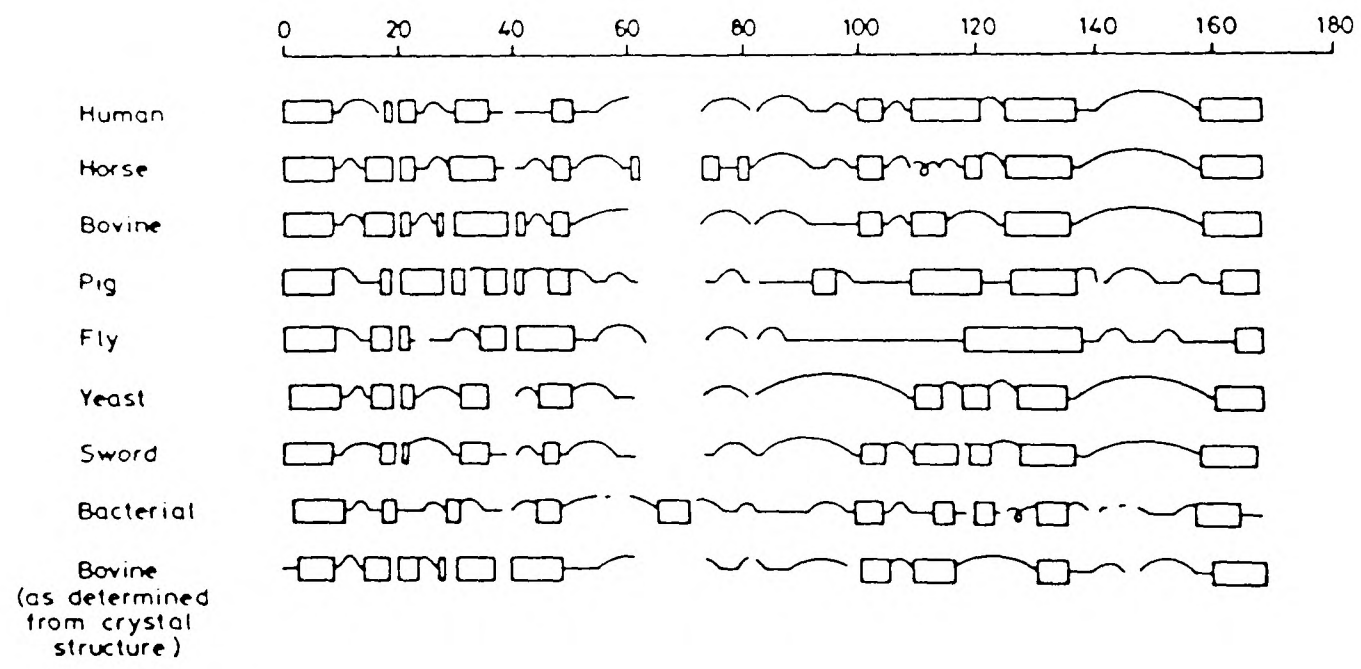


Figure 1.8 Comparison of the predicted secondary structures of Cu/ZnSODs. □ β-strand; ∿ α-helix; ⌒ β-turn (from Parker et al., 1985; see appendix four details).

folding is similar to the immunogloblin domain (Richardson et al., 1976). There is extensive contact area of mainly hydrophobic interactions at the dimer interface. The Cu atom has four histidine ligands situated in a square planar arrangement and the Zn atom has three histidine ligands and an aspartic acid ligand forming a tetrahedral arrangement of bonding about the Zn atom. The Cu atom is connected to the Zn atom via a histidine ligand bridge. The Cu atom is accessible to solvent whereas the Zn is buried completely.

All the sequenced enzymes in figure 1.6 have been found to have the same catalytic activity, metal content and subunit size (Parker et al., 1985 and references therein). The sequence alignment indicates that a number of residues determined previously from physico-chemical studies to be important for metal binding, activity and maintenence of tertiary structure in the bovine enzyme are conserved in the other enzymes (Tainer et al., 1982). These include all the histidine ligands to the metal ions (his 48, 50, 77, 86, 95 and 135), the aspartic acid (asp 98) ligand to the Zn, an arginine (arg 158) and aspartic acid (asp 139) considered essential for activity and the intrachain disulfide bridge formed by cysteine residues 59 and 161. A high proportion of glycine and proline residues are conserved. The high degree of sequence identity (table 1.7) and similar predicted secondary structures (figure 1.8) for the eukaryotic enzymes suggests that all the enzymes have similar three-dimensional structures. Despite the low degree of sequence identity between the eukaryotic Cu/ZnSODs and P.leiognathi Cu/ZnSOD, sequence analysis and model building studies predict the three-dimensional structure of the bacterial enzyme to be similar to the others (Bannister & Parker, 1985).

Complete sequences of MnSODs from four species have been published (see figure 1.9). There is a significant degree of sequence identity between them, independent of the source (table 1.8). Only the eukaryotic forms have cysteine. Seven glycine and three proline residues are conserved of which

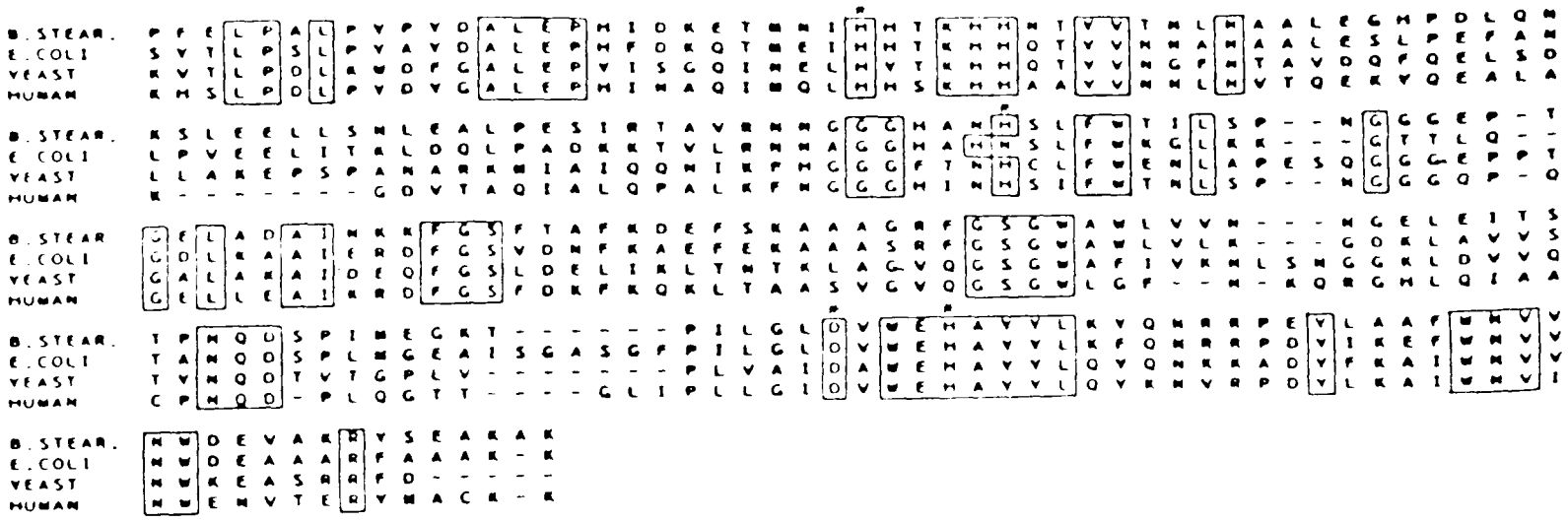


Figure 1.9 Amino acid sequences of MnSODs.

1. Bacillus stearothermophilus (Brock & Walker, 1980).
2. Eschericha coli (Steinman, 1978).
3. Saccharomyces cerevisiae (Ditlow et al., 1982).
4. Human liver (Barra et al., 1984).

		B. stearothermophilus	E. coli	Yeast
1	B. stearothermophilus			
2	E. coli	59		
3	Yeast	39	37	
4	Human liver	48	41	43

Table 1.8 Sequence identities (%) between MnSODs.

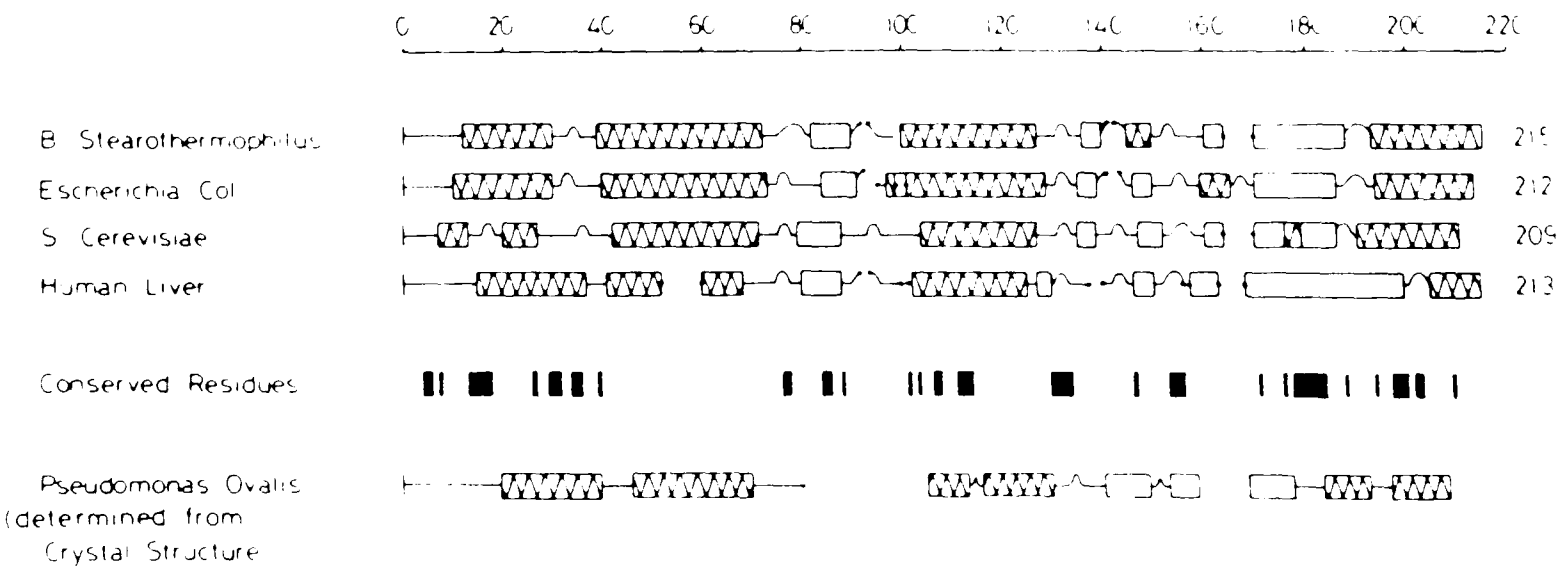


Figure 1.10 Comparison of the predicted secondary structures of MnSODs and the known secondary structure of FeSOD from Ps. ovalis. ▨ α-helix; □ β-strand; ∩ β-turn (from Parker et al., 1984; see appendix four for details).

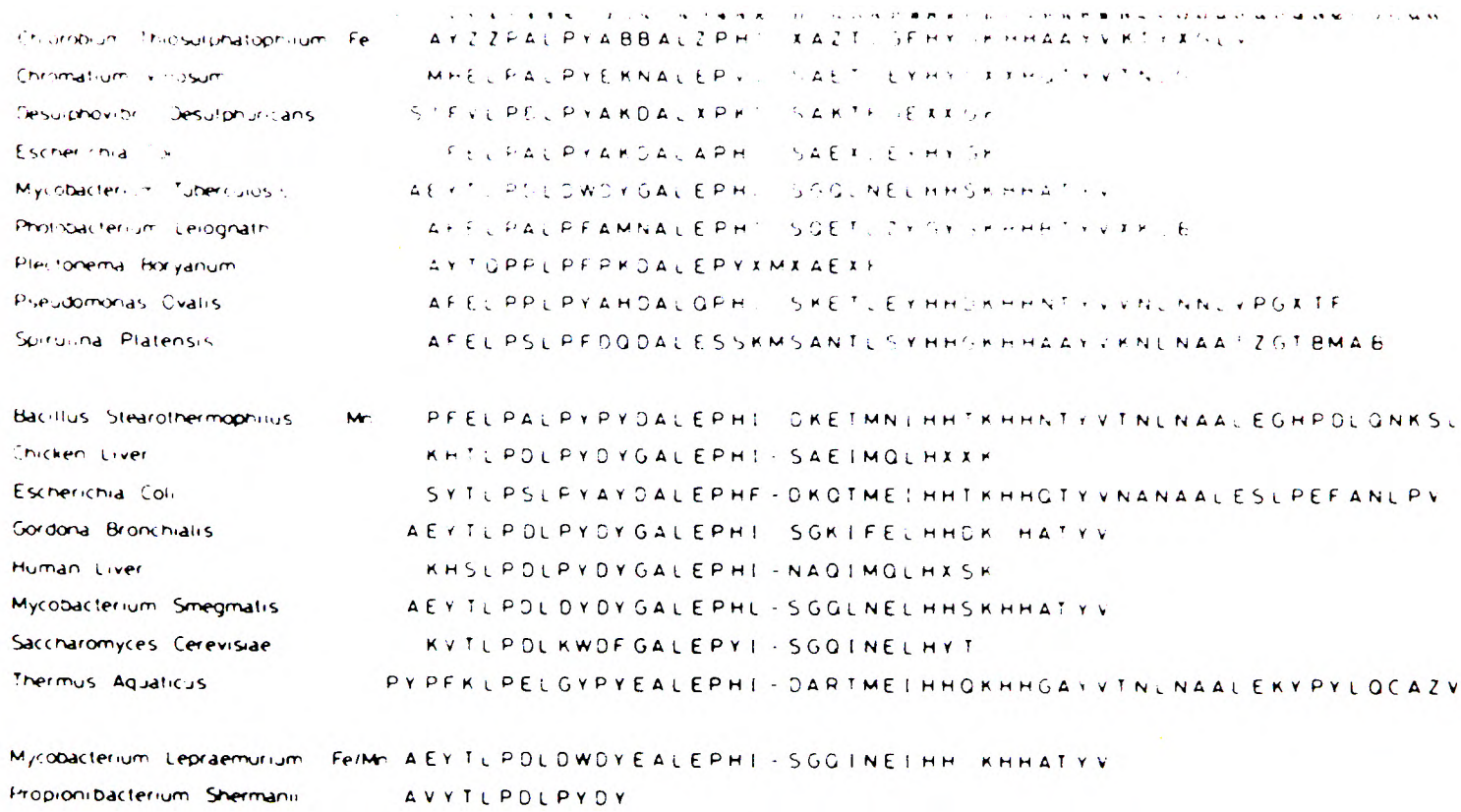
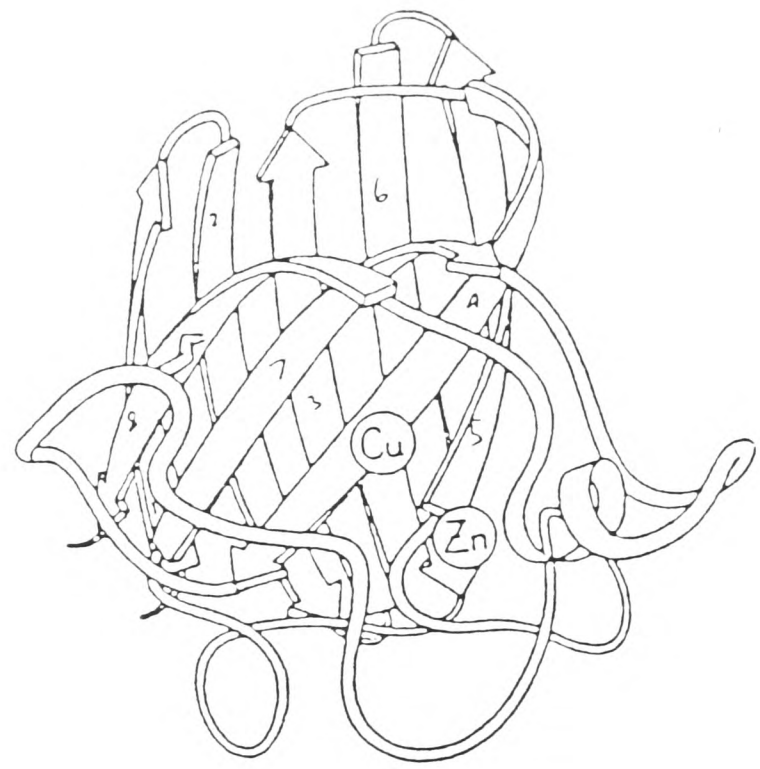


Figure 1.11 Comparison of the N-terminal sequences of Mn and FeSODs (from Parker et al., 1984 and references therein; see appendix four).

a



b

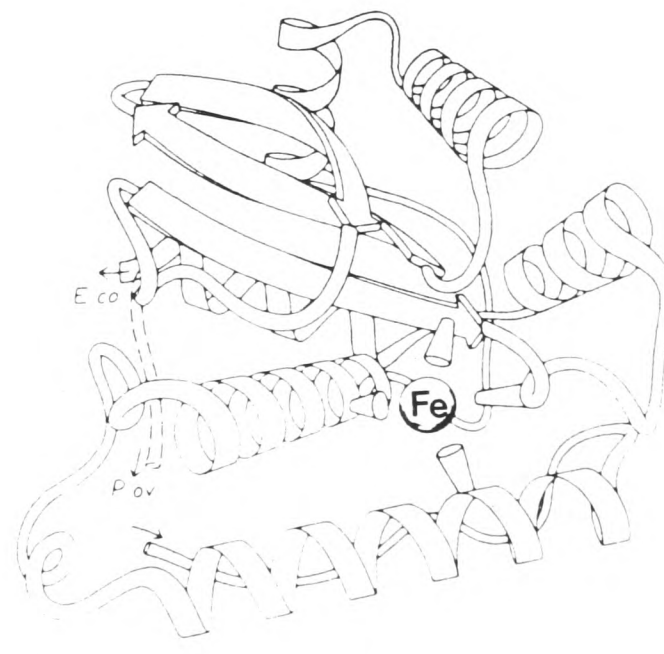


Figure 1.12 Schematic drawings of the polypeptide fold of

(a) the Cu/ZnSOD monomer from bovine erythrocyte

(from Tainer et al., 1982), and

(b) the FeSOD monomer from P. ovalis

(from Ringe et al., 1983).

some, if not all, must play a part in the maintenance of the protein's tertiary structure. The yeast mitochondrial MnSOD is synthesised as a larger proenzyme in the nuclear genome (Autor, 1982). The sequence of the gene shows that the proenzyme has an N-terminal extension of twenty seven amino acids and in common with other presequences of imported mitochondrial proteins, the presequence contains several basic but no negatively charged residues (Marres et al., 1985). It has been shown that the information in the presequence is sufficient to direct an attached cytosolic protein into the mitochondrial matrix which is then cleaved off by a protease (Hurt et al., 1985).

No Fe enzyme has yet been sequenced completely but on the basis of N-terminal sequence comparisons and secondary structure predictions the Fe and Mn proteins appear to be structural homologues (see figures 1.10 and 1.11, Harris et al., 1980; Parker et al., 1984). Crystal structures of FeSOD from Ps.ovalis and E.coli have been determined at medium resolution (to about 3Å) but the metal ligands are unknown due to the lack of complete sequence data (see figure 1.12, Ringe et al., 1983; Stallings et al., 1983). One ligand is probably histidine 26 on the basis of N-terminal sequence studies (figure 1.11). The subunit of FeSOD is triangular in shape with dimensions of 52 x 40 x 32Å. The dimer interface is supported by side chain contacts. The monomer is made up of six helices and three strands of antiparallel β-pleated sheet. The subunit can be visualised as consisting of two domains; the first N-terminal domain consists of two antiparallel crossing helices and the C-terminal domain is made up of a three layer structure having a mixed α/β secondary structure. Each domain provides two ligands to the metal and there is evidence for an iron-bound water. The Fe-Fe distance is 18Å. The overall folds of the proteins from P.ovalis and E.coli are very similar with the exception at the C-terminus where the conformation and chain length differ (figure 1.12).

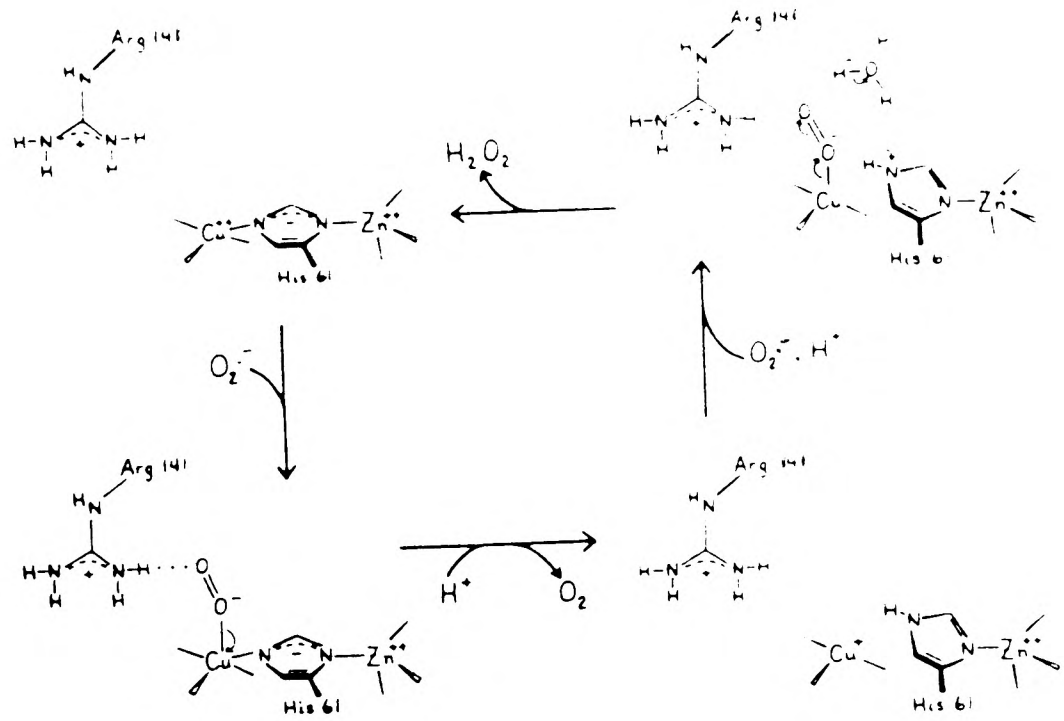


Figure 1.13 Schematic diagram of a proposed mechanism for superoxide dismutation by Cu/ZnSOD. Numbering scheme based on the bovine sequence. (From Tainer et al.,1983).

1.8 Mechanistic Studies

The catalytic metal, whether it is Cu,Fe or Mn, undergoes redox cycling in the dismutation reaction. At low superoxide concentrations, the catalysed reactions of all SODs has been shown to be first order with respect to substrate (Fielden et al., 1974; Lavelle et al., 1977; McAdam et al., 1977a; Fee et al., 1981a) in contrast to the second order kinetics of spontaneous decay. The Cu/ZnSOD is the most efficient scavenger of superoxide with the rate of the overall reaction (3.5×10^9 $M^{-1}s^{-1}$) close to the theoretical diffusion limit (Fielden et al., 1974).

The bovine erythrocyte Cu/Zn enzyme is the most studied form as far as the catalytic properties of SOD are concerned (Rotilo & Fielden, 1984). Pulse radiolysis studies (Fielden et al., 1974; Klug-Roth et al.,1973) have indicated a simple redox cycle for the Cu as follows:

$$SOD - Cu(II) + O_2^{\cdot -} \longrightarrow SOD'Cu(I) + O_2 \qquad (7)$$

$$SOD'Cu(I) + O_2^{\cdot -} \xrightarrow{2H^+} SOD-Cu(II) + H_2O_2 \qquad (8)$$

The rates of the two half reactions are identical to the rate of the overall reaction. The rate constant is independent of pH in the range 5 to 10 in contrast to free metal-catalysed dismutation (Klug et al., 1972). The proposed 'ping-pong' mechanism is enhanced by the slow reactivity of the bound Cu to other redox reagents; in particular H_2O_2 and O_2 (Rigo & Rotilo, 1980). Although the general features of the dismutation are well known, a detailed mechanism remains to be established firmly (Fielden & Rotilo, 1984). A mechanistic scheme has been proposed on the basis of all lines of investigation , shown in diagrammatic form in figure 1.13 (Tainer et al., 1983; Getzoff et al., 1983). An incoming $O_2^{\cdot -}$ displaces the Cu-bound water and hydrogen-bonds to a nearby arginine (arg 141). The bound $O_2^{\cdot -}$ reduces the Cu with simultaneous breaking of the bond between the bridging

histidine (his 61) and the Cu atom. Oxygen is released and his 61 becomes protonated. Binding of the second incoming $O_2^{-\cdot}$ to the Cu results in hydrogen bonds between the oxygen and both his 61 and arg 141. Electron transfer from Cu(I) together with proton transfers from his 61 and an active site water to the bound $O_2^{-\cdot}$ complete the reaction. The Zn enhances the catalysis by raising the pKa of his 61 to ensure its protonation at physiological pH and also positions his 61 at the correct hydrogen bonding distance for the second incoming $O_2^{-\cdot}$. Indirect evidence for a direct complex formation between superoxide and Cu has been obtained from saturation kinetics (Rigo et al., 1975) and ^{19}F nmr (Viglino et al., 1979). The breaking of the Cu-bridging histidine bond is supported by EXAFS studies of reduced Cu/ZnSOD (Blackburn et al., 1984) and the hydrogen bonding to arg 141 by phosphate binding studies (Freitan & Valentine, 1984) and quantum chemical simulations (Osman & Basch, 1984).

The first major mechanistic study of MnSOD was on the enzyme from E.coli (Pick et al., 1974). The kinetics appeared complex with a fast rate ($1.5X10^9$ $M^{-1}sec^{-1}$, pH7.9) for a slight excess of $O_2^{-\cdot}$ and a much slower rate (ca. $2X10^7$ $M^{-1}sec^{-1}$, pH 7.9) which occurs at a higher concentration of substrate. At least four valence states (I to IV) were proposed in the four oxidation/reduction steps and each step was characterised by a different rate constant. A pulse radiolysis study of MnSOD from B.stearothermophilus yielded a simpler model (McAdam et al., 1977a and b). The study supported a fast outersphere mechanism at low substrate concentrations ($5.6X10^8$ $M^{-1}s^{-1}$, pH 8.0) and a slower inner sphere mechanism at higher substrate concentrations ($4.8X10^7$ $M^{-1}s^{-1}$, pH 8.0). NMR studies supported an outersphere mechanism by which the substrate attaches itself to the Mn-bound water (Villafranca et al., 1974). The proposed reaction for the fast cycle was as follows:

$$SOD\text{-}Mn(III)\text{-}OH_2 + O_2^{-\cdot} \longrightarrow SOD\text{-}Mn(II)\text{-}OH_2 + O_2 \qquad (9)$$

$$SOD\text{-}Mn(II)\text{-}OH_2 + O_2^{\cdot-} \xrightarrow{2H^+} SOD\text{-}Mn(III)\text{-}OH_2 + H_2O_2 \qquad (10)$$

The activity of the enzyme was shown to decrease with increasing pH. Coulometric titrations have shown that there is a difference in behaviour of MnSOD from B.stearothermophilus and E.coli as far as their mid-point potentials are concerned (Lawerence & Sawyer, 1979). A study of MnSOD from Paracoccus denitrificans has shown that this protein exhibits saturation kinetics demonstrating that a Michaelis complex is formed in its reaction with superoxide. The pH-dependent kinetics implicated a tyrosine in the mechanism (Terech et al., 1983).

The general features of dismutation by FeSODs have been found to be similar to the bovine Cu/ZnSOD. The dismutation reaction of P.leioqnathi has been shown to be first order with respect to substrate concentration and the rate decreases with increasing pH (Lavelle et al., 1977). At pH 8.0 the rate constant was an order of magnitude less than that obtained for the bovine enzyme but like the bovine case showed no evidence of substrate saturation (Lavelle et al., 1977). The turnover rate constant of $3x10^8$ $M^{-1}s^{-1}$ at pH 7.4 for E. coli FeSOD was similar to the value for P.leioqnathi FeSOD but the E. coli enzyme exhibited saturation kinetics (Fee et al., 1981a). Results of potentiometric titrations of FeSODs from three bacterial sources are consistent with one electron being accepted per Fe(III) (Barette et al., 1983). Association of azide, spectral properties of Fe(III) and the Michaelis constant for $O_2^{\cdot-}$ are all affected by pH, consistant with an ionising group of pKa about 9 (Fee et al., 1981a and b). Proton relaxation measurements are consistent with this pKa belonging to a iron-bound water (Villafranca et al.,1974; Villafranca, 1976).

Despite the different types of kinetic behavior displayed by SODs, there are a number of similarities in terms of mechanism. All three types of SOD show a drop in activity with increasing ionic strength. Further, all

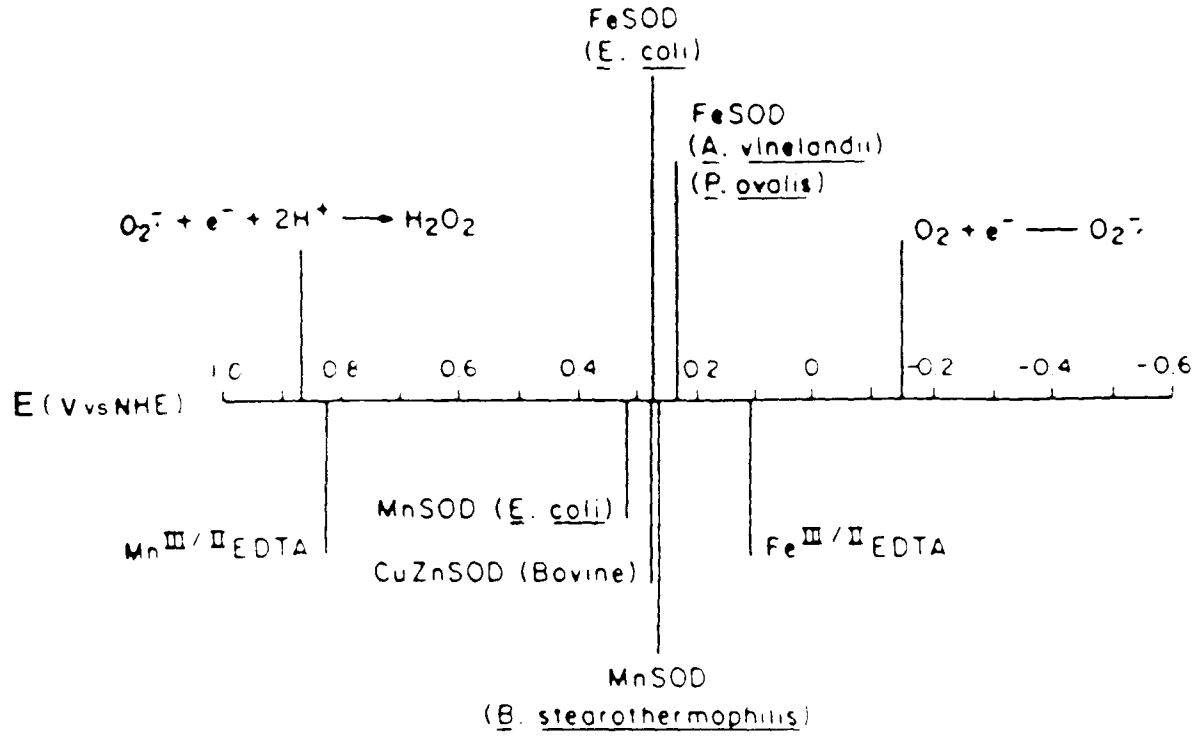


Figure 1.14 Formal reduction potentials ($E^{o\prime}$) for oxygen (unit molarity) and midpoint potentials (E_m) for SODs in aqueous solution at pH 7.0 (from Barrette et al., 1983).

three have low isoelectric points implying that they have a net negative charge at physiological pH. Chemical studies have indicated that the net charge around the active site is positive due to clusters of basic residues (Cocco et al., 1982; Cudd & Fridovich, 1982; Argese et al., 1984; Berovic et al.,1983). Electrostatic models based on the crystal structure of Cu/ZnSOD have been proposed which suggest a mechanism for guidance and charge stablisation of the substrate (Koppenol, 1981; Getzoff et al., 1983). Allison and McCammon (1985) have used dynamic simulations to show that the electrostatic interactions bias the substrate trajectories towards the active site of the protein and this substrate steering leads to a significant enhancement of the reaction rate.

A striking similarity in redox potentials exists for the three types of SOD (Barrette et al., 1983). The measured potentials fall in the range +0.23 to +0.40V at physiological pH (see figure 1.14). It has been estimated that the optimum potential required for catalysis is +0.36V vs NHE (Barrette et al., 1983). Thus the redox properties of SOD are well suited for $O_2^{-\cdot}$ dismutation which is a remarkable coincidence for a diverse set of metals and primary structures.

1.9 Aims of the Project

It was of timely importance that a three-dimensional structure determination of a MnSOD be pursued, given the background outlined in the preceeding sections of this chapter. The main aims of such a study would be twofold. Firstly, the structure of MnSOD would represent the first structure of a redox active manganese protein and so would provide useful information. Little is understood about the role of manganese and its associated proteins. It is known that manganese is required physiologically (Kemmerer et al., 1931; Orent & McCallum, 1931) and it is postulated to play an important role in control of biological systems (Williams, 1982). Although manganese binds to many enzymes *in vitro*, it does not bind

strongly so that its in vivo chemistry is difficult to verify (Scheuhammer & Cherian, 1985). At the start of the project the only high affinity non-redox manganese enzyme known was pyruvate carboxylase (Scrutton et al., 1966). Since then it has become apparent that manganese is an essential component of a number of proteins; these are phytohemaglutins, arginase, α-isopropylmalate synthase, plant acid phosphatase and pseudocatalase (Kono & Fridovich, 1983 and references therein). Of particular interest is the possible relationship of MnSOD to the manganese-containing protein of the photo-reaction centre (system II) which abstracts electrons from water to yeild oxygen (Yamamoto et al., 1984). In common with MnSOD, the photosystem cannot backreact with dioxygen.

The second aim of a crystallographic study would be an attempt to establish structural and functional relationships between the different types of SOD at the atomic level. The Cu/Zn and Mn,Fe proteins appear to be structurally very different yet their redox potentials are very similar (see section 1.8). The Fe and Mn proteins appear to be structural homologues yet their physico-chemical properties suggest differences at their active sites. The Cu/ZnSODS have basic residues clustered around their active site; structural analysis would confirm whether this was the case for the other SODs. A structure at atomic resolution would enable detailed proposals of the catalytic mechanism to be put forward and enable comparisons to be made with the Cu/ZnSOD mechanism. Finally, the crystal structure determinations of Fe and MnSODs were being pursued from a range of thermophilic sources; from mesophile (Yamakura et al., 1976) to extreme thermophile sources (Stallings et al., 1981). A structural comparison of the proteins from these sources would increase an understanding of the forces involved in the heat stabilty of proteins (see Argos et al., 1979 for a review).

The work presented in this thesis was carried out on the protein from the thermophilic bacterium, B.stearothermophilus. A structural study of

MnSOD from B.stearothermophilus offered a number of advantages. A complete amino acid sequence was known (Brock & Walker, 1980) and a great deal of physico-chemical work had been published on MnSOD from this source (see sections 1.7 and 1.8). Crystallization conditions for this source of protein had been published by two groups; one group had passed on the project to us (Bridgen et al., 1976) and the other group had apparently discontinued work on the project (Smit et al., 1977; D.Rice, pers. commun.).

At the start of the project (October 1981), the crystal structure determination of bovine Cu/ZnSOD was at an early stage (3Å resolution model by Richardson et al., 1975). There was enough N-terminal sequence information available to suggest that Fe and MnSODs could be structural homologues but structurally unrelated to Cu/ZnSOD. Crystal structure determinations of Fe and MnSODs were still at a preliminary crystallization and heavy atom search stage (see chapter three for a review). It was not until midway through the project that medium resolution structures of FeSODs started to appear in the literature (Ringe et al., 1983; Stallings et al., 1983).

CHAPTER 2

Theory and Methods

2.1 Introduction

The purpose of this chapter is to present a brief account of the theoretical background and experimental methods which formed a basis for the structural studies of MnSOD discussed in this thesis. A more detailed treatment of the subject can be found in, for example, Lipson & Cochran (1966) and Blundell & Johnson (1976).

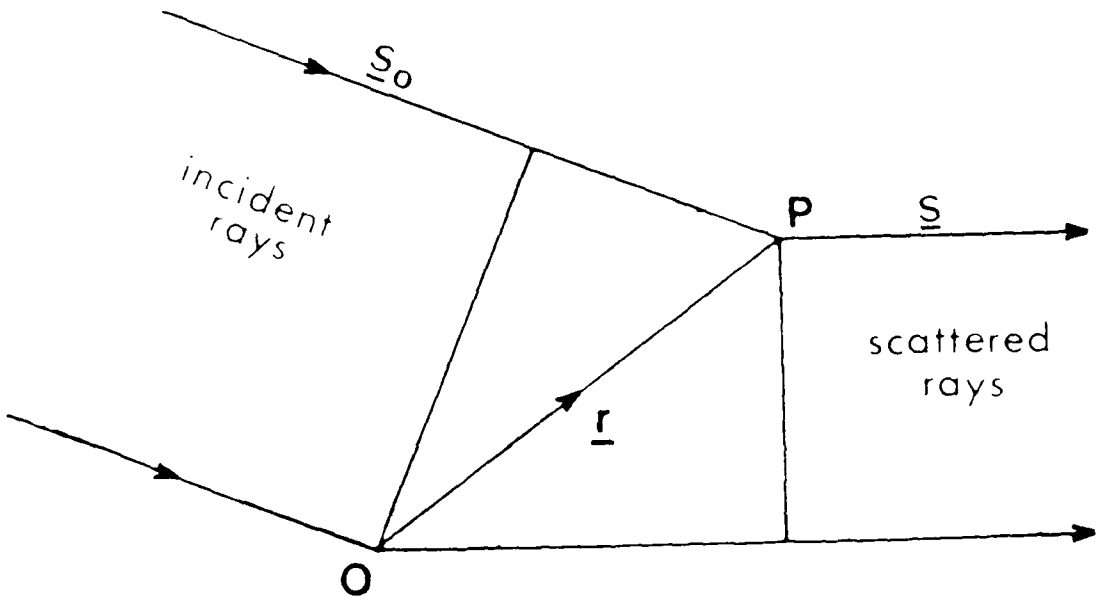
2.2 The Diffraction of X-rays

The fluctuating electromagnetic field of an incident X-ray wave causes a free electron to oscillate and so emit a secondary wave which has the same wavelength as the incident wave. The scattering from this interaction is said to be coherent because the emitted rays are all π out of phase with respect to the incident beam. This process can be modelled by the electromagnetic theory (as developed by J.J. Thomson) which states that the intensity of the scattered beam, $I(2\theta)$, can be related to the intensity of the incident beam, I_o, by the expression,

$$I(2\theta) = (ne^4/2r^2m_e{}^2c^4)(1+\cos^2 2\theta)I_o \qquad (2.1)$$

where n is the effective number of independent scattering electrons, e is the electronic charge, r is the distance from the scatterer, m_e is the electronic mass and c is the velocity of light. The $(1+\cos^2 2\theta)$ term allows for the partial polarisation of the emitted rays where 2θ is the scattering angle.

The scattering in any particular direction can be defined by an amplitude and a phase. Consider the effect of a parallel beam of X-rays of



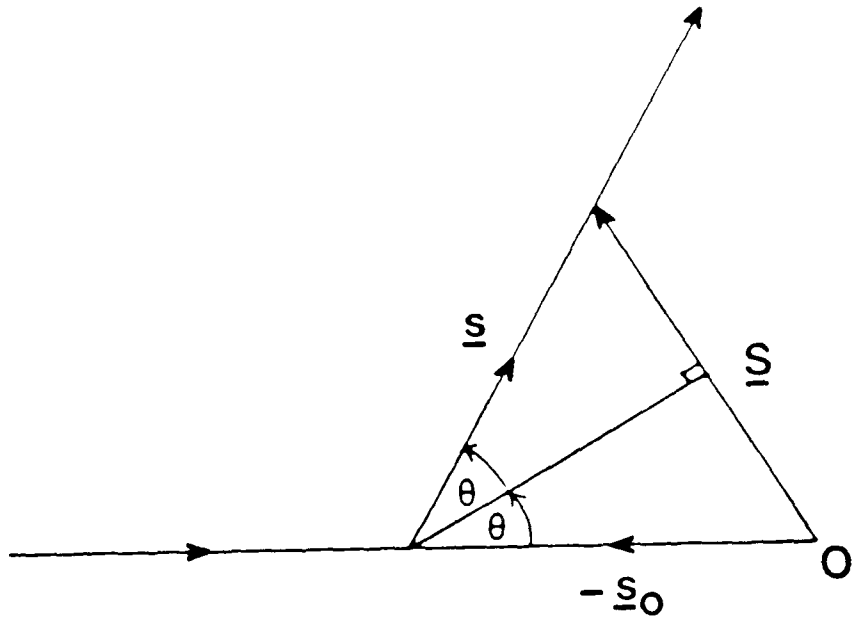


Figure 2.1 Scattering at a point P relative to an origin O and the relationship of the diffraction space vector $\underline{S}$ to the real space vectors $\underline{s}_o$ and $\underline{s}$.

wavelength, λ, falling onto a lattice of electrons. Referring to figure 2.1, the path difference between an X-ray scattered at some point P relative to that scattered by an electron at the origin is

$$\underline{r}.\underline{s}_o - \underline{r}.\underline{s} \qquad (2.2)$$

where $\underline{r}$ is the vector distance of P from the origin and $\underline{s}_o$ and $\underline{s}$ represent the vectors of the incident and scattered rays respectively. If we choose the modulus of $\underline{s}_o$ and $\underline{s}$ to be $1/\lambda$, then the phase difference is given by $2\pi\underline{r}.\underline{S}$ where $\underline{S}$ is the vector difference between the incident and scattered waves. The vector $\underline{S}$ is called the scattering vector and is used to describe position in diffraction ('reciprocal') space.

If we consider the lattice to be made up of atoms rather than electrons, we then must consider the total wave scattered by a partial volume of the atom, dV, and then sum up these individual contributions over the volume of the atom. Thus the total wave scattered by an atom at point P can be expressed as a function of the scattering vector,

$$\underline{F}(\underline{S}) = \int_{\text{volume}} \rho(r)\exp(2\pi i\underline{r}.\underline{S})dV \qquad (2.3)$$

where the amplitude is proportional to $\rho(r)dV$ and the phase is $2\pi\underline{r}.\underline{S}$. The function $\underline{F}(\underline{S})$ is referred to as the atomic scattering factor. If the electron density of an atom is assumed to be spherically symmetric, then it can be shown that the function is real and is denoted by the symbol f.

The total wave scattered by a molecule of N atoms can be derived from the vector addition of the atomic contributions to give

$$\underline{G}(\underline{S}) = \sum_{j=1}^{N} f_j\exp(2\pi i\underline{r}_j.\underline{S}) \qquad (2.4)$$

where the function $\underline{G}(\underline{S})$ is referred to as the molecular transform.

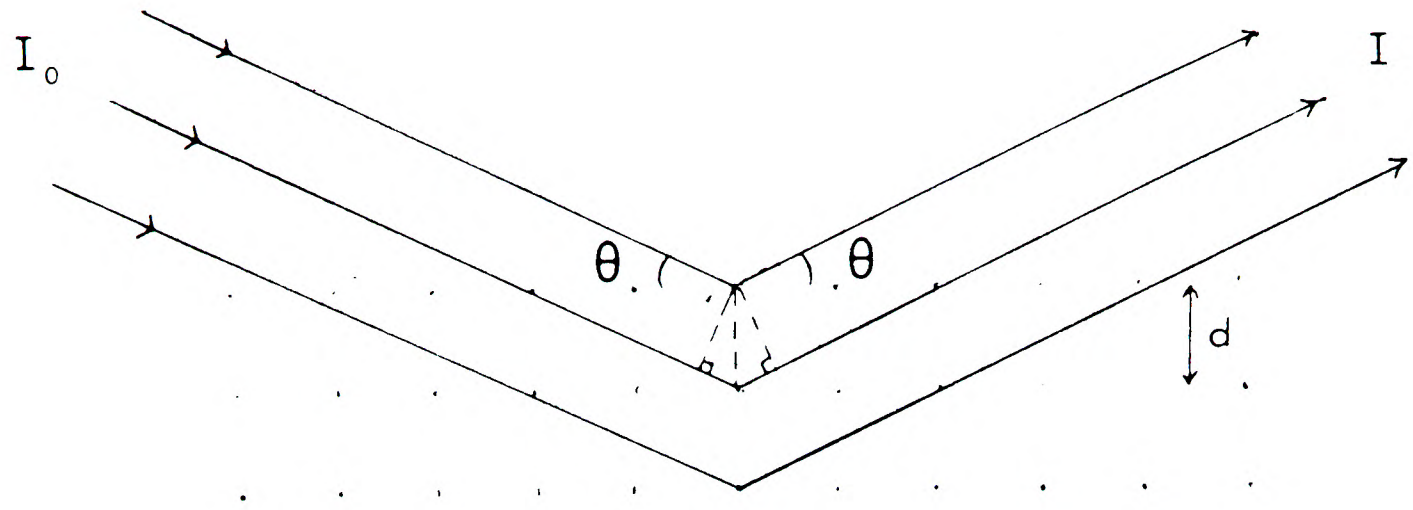


Figure 2.2 Bragg's law.

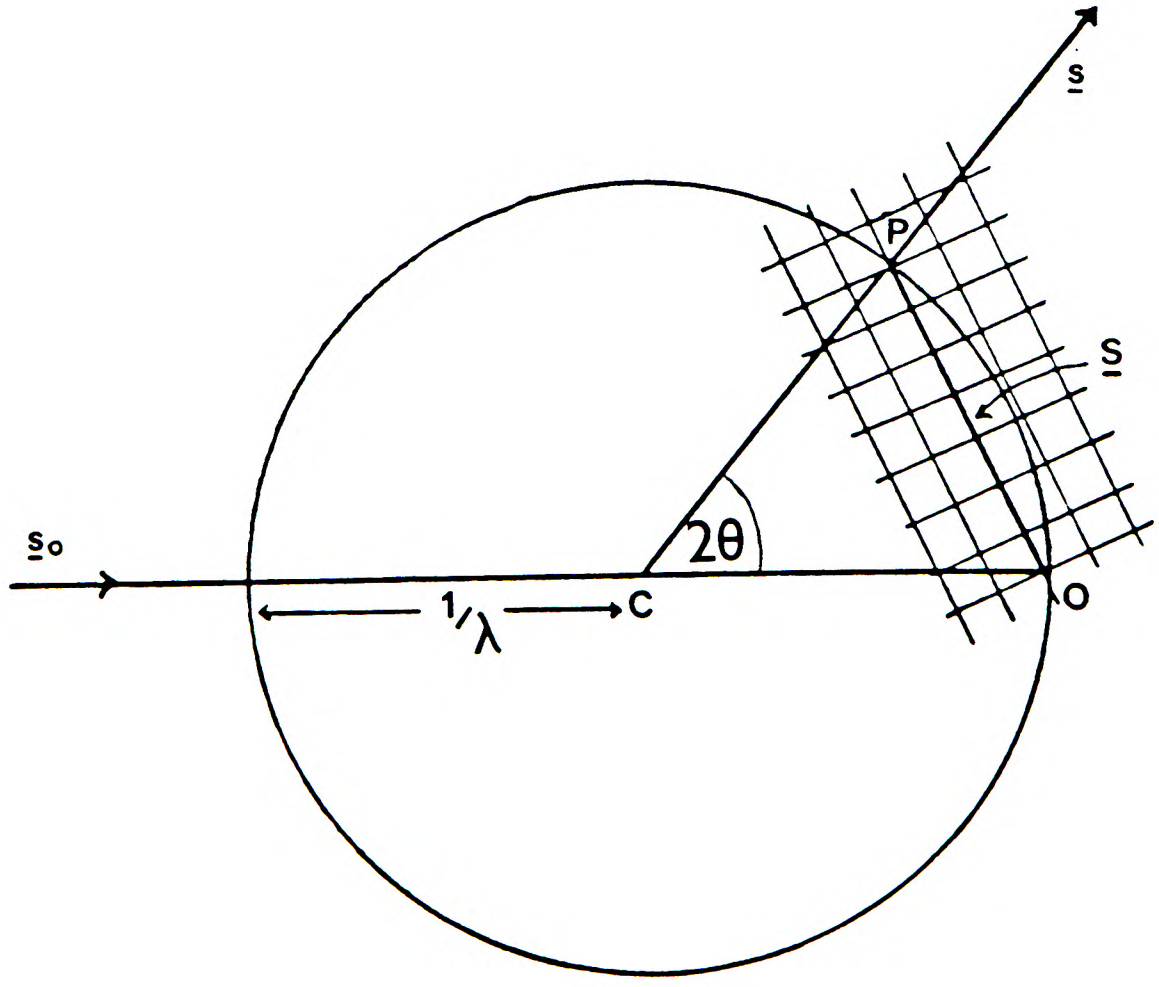


Figure 2.3 The Ewald Construction.

The total wave scattered from a single molecule is too weak to be observable but the scattering from a crystal can be observed on the basis that the waves scattered from a crystal's unit cells add up in a constructive fashion, ie. the phase difference in waves scattered by unit cells is some integral multiple of 2π. The Laue equations are an expression of this condition. If we define $\underline{r}$ in figure 2.1 in terms of the unit cell lattice vectors $\underline{a}$, $\underline{b}$ and $\underline{c}$, then the phase differences for a scattered beam of maximum intensity are

$$2\pi(\underline{a}.\underline{S}) = 2\pi h \; ; \; 2\pi(\underline{b}.\underline{S}) = 2\pi k \; ; \; 2\pi(\underline{c}.\underline{S}) = 2\pi l \qquad (2.5)$$

The integers h, k and l in the Laue equations are called Miller indices. Bragg (1913) showed that Miller indices could be identified with imaginary lattice planes of atoms in the crystal and that the scattering process could be visualised in terms of reflection from these planes. Thus, from figure 2.2, if d is the interplanar spacing, then for constructive interference

$$2d\sin\Theta = n\lambda \qquad (2.6)$$

where n is an integer and λ is the wavelength of radiation. Ewald (1921) proposed a simple geometric construction which describes Bragg's Law (and Laue's equations) in terms of the reciprocal lattice. In figure 2.3, a sphere is drawn with crystal C at its centre and diameter, $1/\lambda$. A reciprocal lattice is defined with origin at O and each lattice point defined by the Miller indices h, k, and l. X-rays will be diffracted in the direction of the diffracting position if P represents a reciprocal lattice point hkl. The reciprocal lattice vector is given by

$$\underline{S} = h\underline{a}^{*} + k\underline{b}^{*} + l\underline{c}^{*} \qquad (2.7)$$

where $\underline{a}^*$, $\underline{b}^*$ and $\underline{c}^*$ are the lattice constants of the diffraction lattice.

We can obtain the scattering equation for a crystal by combining the equation expressing the fractional coordinates of the jth atom in terms of the lattice vectors:

$$\underline{r}_j = \underline{a}x_j + \underline{b}y_j + \underline{c}z_j \qquad (2.8)$$

with the Laue equations,

$$\begin{aligned} \underline{r}_j.\underline{S} &= x_j\underline{a}.\underline{S} + y_j\underline{b}.\underline{S} + z_j\underline{c}.\underline{S} \\ &= hx_j + ky_j + lz_j \end{aligned} \qquad (2.9)$$

and by substitution into equation 2.4:

$$\underline{F}(hkl) = \sum_{j=1}^{N} f_j \exp 2\pi i(hx_j + ky_j + lz_j) \qquad (2.10)$$

This is the structure factor equation representing the molecular transform sampled at the reciprocal lattice points hkl.

In obtaining the equation for the atomic scattering factor (equation 2.3), a number of assumptions were made. The equation represents the scattering by an atom at rest. Debye (1914) showed that the thermal motion of atoms reduces the scattering intensity by a factor, $\exp(-B\sin^2\theta/\lambda^2)$. The constant B is called the temperature factor and is defined by

$$B = 8\pi^2\langle u^2\rangle \qquad (2.11)$$

where $\langle u^2\rangle$ is the mean square amplitude of atomic vibration. In practice, atoms probably tend to vibrate anisotropically and this behaviour is modelled by parameters describing an ellipsoid of vibration in real

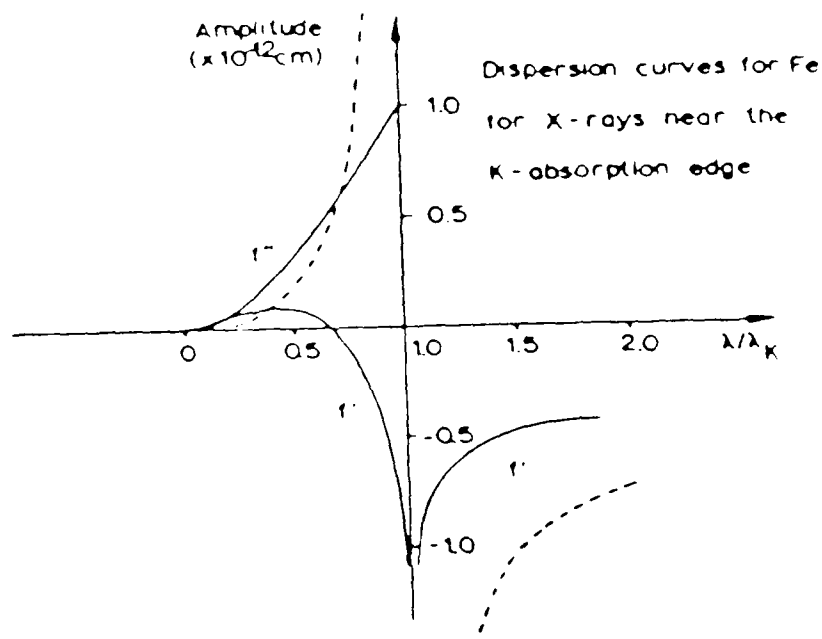


Figure 2.4 Behaviour of the real (f') and imaginary (f'') components of the scattering amplitude close to an absorption edge. The dashed curves give f' for a single harmonic oscillator. λ_K is the wavelength of the k-absorption edge (after Mossbauer, 1975).

space.

A further assumption made in the structure factor derivation was that electrons scatter freely. Normally the electron binding energy of an atom is small compared with the energy of an X-ray photon so that the scattering power of each electron is like that of a free electron. This is not true for the more strongly bound inner electrons so that the atomic scattering factor equation is expanded to include a dispersion correction,

$$f = f_o + f'(\lambda) + if''(\lambda) \qquad (2.12)$$

where f_o is the scattering factor for a free electron and f', f'' represent the real and imaginary components of the dispersion correction. As an absorption edge is approached, the threshold frequency above which an inner electron can be injected into the continuum is reached. Near the edge, f'' increases exponentially and there is a loss of real contribution to the structure factor (figure 2.4). The imaginary component, f'', lags $\pi/2$ out of phase with the primary beam and this phase shift means that Friedel's law (ie. $F(hkl) = F(\bar{h}\bar{k}\bar{l})$) no longer holds and the scattering is called anomalous.

2.3 The Phase Problem

The diffraction pattern corresponds to the Fourier transform of the crystal structure. Using the inverse property of Fourier transformation, the electron density ($\rho(\underline{x})$) of the crystal structure is described by the Fourier transform of the diffraction pattern. It can be shown that

$$\rho(\underline{x}) = (1/V) \sum_{\underline{h}=-\infty}^{\infty} \underline{F}(\underline{h})\exp(-2\pi i\underline{h}.\underline{x}) \qquad (2.13)$$

where $\underline{x}$ is the fractional coordinates of an atom in the unit cell, $\underline{h}$ is the reciprocal lattice vector (h,k,l) and V is the volume of the unit cell.

The structure factor $\underline{F}(\underline{h})$ is a complex value

$$\underline{F}(\underline{h}) = F(\underline{h})\exp[i\alpha(\underline{h})] \qquad (2.14)$$

Only the intensities (and thus the amplitude, $F(\underline{h})$) can be measured from the diffraction of a crystal. The basic problem in any crystal structure determination is to evaluate the phase component of the structure factor.

There are a number of approaches towards solving the phase problem. Small molecule structures can be solved using Patterson techniques (see section 2.4.2) or direct methods (whereby mathematical relationships are derived between phases and structure factor amplitudes). For a detailed treatment of the subject, the reader is recommended to read Lipson & Cochran (1966) and Gilmore (1982). Both methods suffer from an increasing complexity proportional to the number of atoms in the molecule; no protein has ever been solved by using these methods alone. Recently, a more general approach for solving the phase problem has been derived from information theory by Bricogne (1984), who regards traditional direct methods as a restricted version of his maximum entropy (minimum information) formulation. The most generally applicable method for solving protein structures is heavy atom multiple isomorphous replacement, which was used to solve the first protein structure (Kendrew et al., 1960). Anomalous scattering has often been used to settle phase ambiguities in the heavy atom approach but has proved sufficient in solving at least two small protein structures (Hendrickson & Teeter, 1981; Stuart et al., 1984). The availability of synchrotron radiation opens up the possibility of solving protein structures by anomalous dispersion using multiple wavelength techniques (Templeton et al., 1980; Greenhough & Helliwell, 1983; Kahn et al., 1985). In some cases structures are available which may be closely related to the protein under study so that molecular replacement techniques can be used (see section 2.5). In the study of MnSOD, both isomorphous and

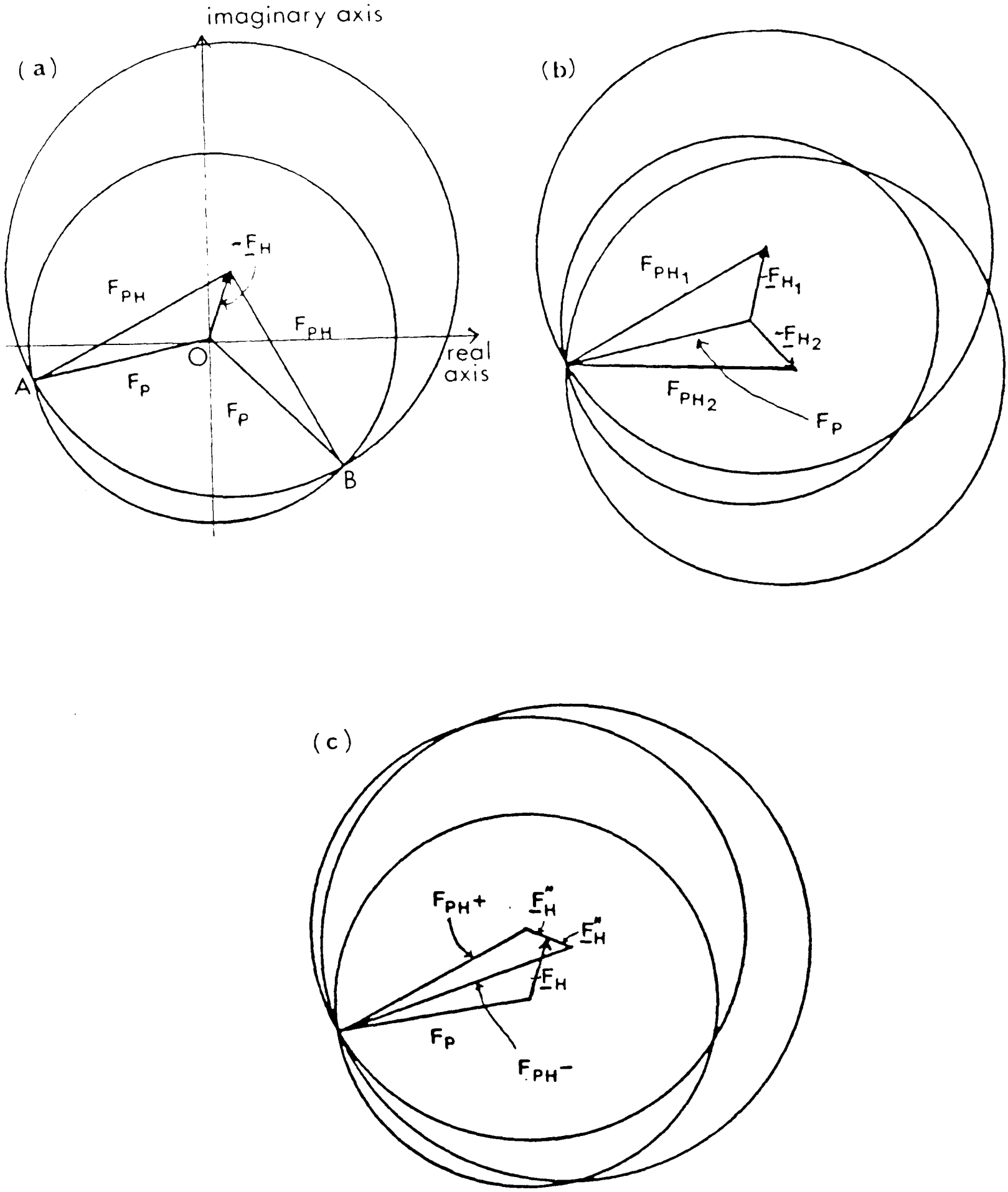


Figure 2.5 Harker constructions for phase calculation by the method of isomorphous replacement for

(a) a single isomorphous derivative (SIR),

(b) two independent derivatives (MIR) and

(c) one derivative with anomalous scattering (SIRAS).

molecular replacement methods were used and these methods are discussed in detail in the following sections.

2.4 Calculation of Phases by Isomorphous Replacement

2.4.1 Introduction

The solution of protein crystal structures by the method of isomorphous replacement requires the determination of positions, occupancies and thermal parameters of added heavy atoms. The method is illustrated by the Harker constructions shown in figure 2.5. By vector addition,

$$\underline{F}_{PH} = \underline{F}_{P} + \underline{F}_{H} \qquad (2.15)$$

If $\underline{F}_{H}$ is known, then two points of intersection made by circles of radius F_{PH} and F_{P} centred at each end of this vector, define the two possible vector solutions $\underline{OA}$ and $\underline{OB}$. This ambiguity can be resolved by incorporating the data from a second heavy atom derivative (figure 2.5b) or by use of anomalous scattering (figure 2.5c). The methods used to estimate $\underline{F}_{H}$ and thus the protein phases are described in the following sections.

2.4.2 The Patterson Function

Estimates of the heavy atom positions can be derived by direct methods or difference Patterson techniques. The method of choice is usually Patterson techniques if the interpretation is relatively straight forward. The Patterson function is defined as follows:

$$P(u,v,w) = \int_{\text{unit cell}} \rho(x,y,z)\rho(x+u,y+v,z+w)dxdydz \qquad (2.16)$$

where ρ is the electron density at some point (x,y,z) in the unit cell. The function will exhibit maxima where (x,y,z) and $(x+u,y+v,z+w)$ correspond to

atomic positions. It follows that the position of the maxima in a Patterson map at points (u,v,w) will represent interatomic vectors. The expression can be rewritten as follows:

$$P(u,v,w) = (1/V) \sum_{\underline{h}=-\infty}^{\infty} F(\underline{h})^2 \cos 2\pi(hu+kv+lw) \qquad (2.17)$$

which shows that the Patterson function requires only structure factor amplitudes for evaluation. A Patterson map will have the following properties; it is centrosymmetric, the peak heights are proportional to the product of the site occupancies of the heavy atoms contributing to the peak and if N is the number of atoms in the unit cell then there will be N(N-1) peaks in the unit cell excluding the origin peak (some of which may not be resolved). A Patterson function with coefficients $(F_{PH}-F_P)^2$ will give a vector map from which the positions of heavy atoms can be derived (Perutz, 1956; Blow, 1958).

2.4.3 Estimation of F_H

Once heavy atom positions have been determined, the next step is to estimate the heavy atom contribution $\underline{F}_H$ to the measured $\underline{F}_{PH}$ where

$$\underline{F}_H = \underline{F}_{PH} - \underline{F}_P \qquad (2.18)$$

This equation can be simplified for the case of centric data as follows:

$$F_H = | F_{PH} \pm F_P | \qquad (2.19)$$

The only ambiguity is in the sign; if the degree of heavy atom substitution is small then it is assumed that,

$$F_H = |F_{PH} - F_P| \qquad (2.20)$$

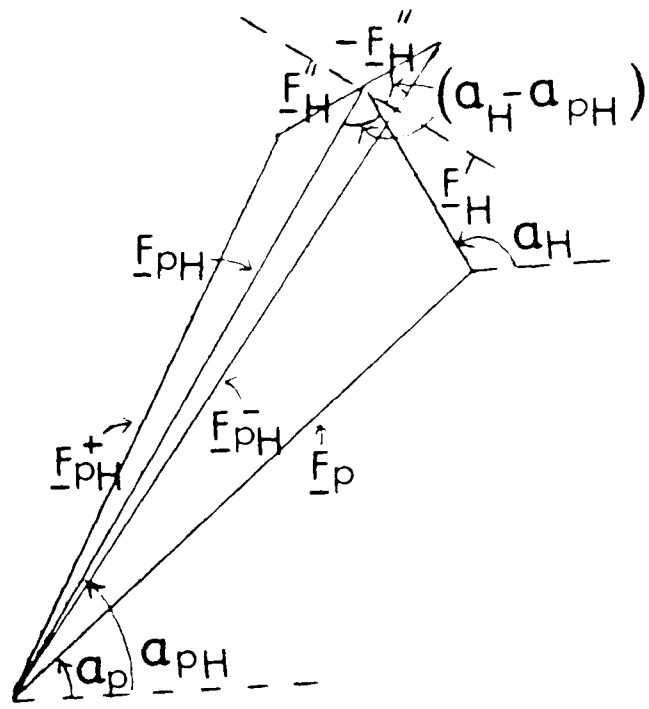


Figure 2.6 Vector diagram defining structure factor amplitudes and phases referred to in the text.

but if both F_{PH} and F_P are small, then the other sign may be correct; this is called 'crossing over'.

Consider the vector triangle in figure 2.6. By application of the cosine rule and assuming (α_{PH}-α_P) is small , the isomorphous difference is given by:

$$\Delta_{iso} = |F_{PH} - F_P| \cong F_H \cos(\alpha_H - \alpha_{PH}) \qquad (2.21)$$

and similarily for anomalous scattering differences:

$$\Delta_{ano} = |F_{PH}^{+} - F_{PH}^{-}| \cong (2/k) F_H \sin(\alpha_H - \alpha_{PH}) \qquad (2.22)$$

where k is defined as the ratio F_H'/F_H''. An estimate of F_H from a combination of isomorphous and anomalous differences has the advantage of complementarity because the differences are maximised for values of (α_H - α_{PH}) of 0 and 90° respectively. Singh & Ramaseshan (1966) have derived an expression for the combined difference:

$$F_H^2 = F_P^2 + F_M^2 \pm 2[F_M^2 F_P^2 - (k/4)^2 (\Delta I)^2]^{1/2} \qquad (2.23)$$

where

$$F_M^2 = (1/2)(F_{PH}^{+2} + F_{PH}^{-2})$$
$$\Delta I = (F_{PH}^{+2} - F_{PH}^{-2})$$

The sign ambiguity in equation 2.23 depends on the value of $\cos(\alpha_{PH} - \alpha_P)$. In most cases ($\alpha_{PH}$-$\alpha_P$) is acute so that the negative sign in front of the bracketed term in equation 2.23 is correct and F_H is called the heavy atom lower estimate or F_{HLE} for short. (When the sign is positive then F_H is called the heavy atom upper estimate or F_{HUE} for short). The parameter, k,

can be used as a relative weighting factor to take account of errors in the measured differences. Rather then use the theoretical expression of k, ie.

$$k_{theor} = (f_o + f')/f'' \qquad (2.24)$$

an empirical value can be determined from the data to take account of errors and non-isomorphism:

$$k_{emp} = (2\langle|F_{PH} - F_P|\rangle)/\langle|F_{PH}^{+} - F_{PH}^{-}|\rangle \qquad (2.25)$$

where the differences are averaged over all the acentric data (Matthews, 1966b). Tickle (1973) has shown that the expectation value of Δ^2, where Δ is the isomorphous or anomalous difference, is systematically overestimated:

$$E(\Delta^2) = [E(\Delta)]^2 + \sigma^2(\Delta) \qquad (2.26)$$

where σ is the standard deviation of the difference. Dodson et al., (1975) showed that the bias introduced by overestimating the value of F_{HLE} could be modelled reasonably well by the formula:

$$\text{bias}\ (F_{HLE}^2) = \sigma_{iso}^2 + (k/2)^2\sigma_{ano}^2 \qquad (2.27)$$

In summary, the errors involved in estimating F_{HLE} can be taken into account by using k_{emp} as a weighting factor or by calculating the bias and using (F_{HLE} - bias) as a corrected estimate instead.

2.4.4 The Refinement of Heavy Atom Sites

There are two distinct methods of refinement available which are based on minimising the lack-of-closure errors, ϵ_1 and ϵ_2, shown in the phase

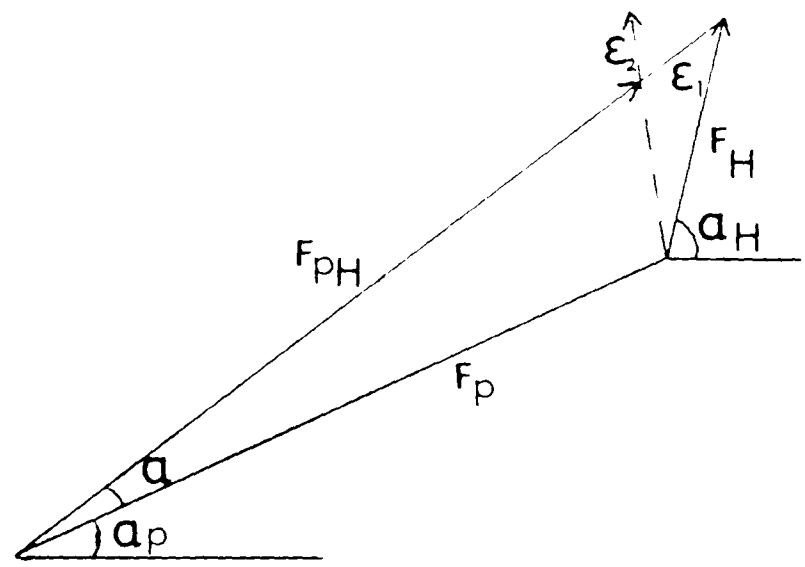


Figure 2.7 The phase triangle and definition of lack-of-closure errors.

triangle of figure 2.7. In F_{HLE} refinement, the summation $\Sigma w\epsilon_2^2$ is minimised where

$$\epsilon_2 = F_{H(obs)} - F_{H(calc)} \qquad (2.28)$$

and w is some chosen weighting factor. For the special case of centric reflections (where α=0 or π):

$$\begin{aligned}\epsilon_2 &= |F_{PH(obs)} \pm F_P| - F_{H(calc)} \\ &= F_{PH(obs)} - |F_P \pm F_{H(calc)}| \\ &= \epsilon_1 \end{aligned} \qquad (2.29)$$

will be true if the correct choice of sign is made.

The alternative method or refinement is 'phase refinement' where the summation, $\Sigma w\epsilon_1^2$, is minimised:

$$\begin{aligned}\epsilon_1 &= F_{PH(obs)} - F_{PH(calc)} \\ &= F_{PH(obs)} - |\underline{F}_P + \underline{F}_H| \end{aligned} \qquad (2.30)$$

The summation, $\Sigma w\epsilon_1^2$, is minimised with respect to the heavy atom parameters which define $\underline{F}_H$ and uses the most probable protein phase, α_P (see section 2.4.4). It is important to realise that $\underline{F}_P$ is not independent of the heavy atom parameters of a derivative which is included in the phase calculation (Blow & Matthews, 1973).

Phase refinement is the method of choice provided one good derivative has been solved and the other derivatives don't have sites in common with the derivative used in the phase calculation. Recently Bricogne (1982) has introduced a phase refinement scheme that will not be biased by derivatives having common sites. The next best choice of refinement is centric providing there are enough terms. F_{HLE} refinement requires good anomalous

data which may not be available. Unlike phase refinement, relative origins of heavy atom atom sites between derivatives cannot be determined by centric or F_{HLE} refinement because derivatives are treated independently. In these cases, the origin problem can be solved by searching Patterson maps for cross vector peaks or by cross-difference Fourier methods (see below).

The Oxford version of the F_{HLE} refinement method works via three steps. First an estimate of $F_{H(obs)}$ is made and reflections rejected for which the estimate is unreliable. Reflections are thrown out if the bias of F_{HLE} is greater than the F_{HLE}, if F_{HLE} is greater than the maximum F_H (estimated from the largest centric Δ_{iso}) in the appropriate resolution range and if F_{HUE} is less than the maximum F_H. The second rejection removes reflections with possible large spurious anomalous differences and the last rejection is a check on whether the upper estimate is plausible because the sign ambiguity is unresolved. The next step involves calculating a weighting factor (chosen to be equal to the inverse variance of the F_{HLE}) and applying it to the data. This has the effect of weighting up the more reliable centric terms and weighting down terms dominated by large (and possibly spurious) anomalous differences. Finally, partial derivatives are calculated with respect to all parameters. The derivative-to-native scale factor (K) can be refined using an iterative procedure of Kraut's method (Kraut et al., 1962; Arnone et al., 1971) in which K is evaluated from the following equation:

$$\sum_{hkl} \langle F_{HLE}^2 \rangle = \sum_{hkl} K \langle F_{PH}^2 \rangle - \sum_{hkl} \langle F_P^2 \rangle \qquad (2.31)$$

The progress of refinement can be monitored in several ways. The conventional R-factor:

$$R = \sum_{hkl} | F_{H(obs)} - F_{H(calc)} | / \sum_{hkl} F_{H(obs)} \qquad (2.32)$$

should decrease from its value of between 0.59 (acentric terms only) and 0.83 (centric terms only) for a totally incorrect structure (Wilson, 1950). Typical R-factors for refinement of heavy atom parameters in successful structure determinations have ranged between 45 and 65% for centric and between 30 and 45% for F_{HLE} refinement (Blundell & Johnson, 1976). The gradient of a plot of $F_{H(calc)}$ against median values of $F_{H(obs)}$ in a range of $F_{H(obs)}$ should approach unity for a correct solution. In practice, correlation values greater than 0.25 seem to be indicative of success.

Having refined parameters of major heavy atom sites derived from Patterson or direct methods, minor sites can be picked up from double difference maps with coefficients:

$$w|F_{H(obs)} - F_{H(calc)}|\exp(i\alpha_{H(calc)}) \qquad (2.33)$$

or by cross-phased double difference maps with coefficients:

$$w|F_{PH(obs)} - F_{PH(calc)}|\exp(i\alpha_{P}) \qquad (2.34)$$

where the phases, α_P, have been calculated from one or more different derivatives.

2.4.5 The Calculation of Phases

Blow and Crick (1959) have shown that protein phases can be expressed in terms of a Gaussian probability distribution function,

$$P_j(\alpha_p)d\alpha = \exp\,[-\Sigma_j\, e_j(\alpha_p)^2/2E_j^{\,2}]d\alpha \qquad (2.35)$$

where $P(\alpha)d\alpha$ is the relative probability that the phase angle α lies between α and $d\alpha$, ϵ_j $(= F_{PH(obs)} - F_{PH(calc)})$ is the 'lack-of-closure'

error, E_j^2 is the mean square error in F_{PH} for the jth derivative which can be estimated from centric data by

$$E^2 = \Sigma_{hkl}\, w(F_{PH} - |F_P \pm F_H|)^2/n \qquad (2.36)$$

where n is the number of observations used in calculating the value of E in heavy atom refinement. In the absence of centric data, it can be estimated from the rms difference between F_{HLE} and $F_{H(calc)}$. The lack-of-closure error takes into account errors in the intensity measurements, errors in heavy atom refinement parameters and lack of isomorphism. In this treatment all the errors are assumed to reside in $F_{PH(obs)}$. The corresponding anomalous 'lack-of-closure' error is defined as:

$$\begin{aligned}\epsilon_j'(\alpha_p) &= \Delta_{ano(obs)} - \Delta_{ano(calc)} \\ &= F_{PH}^{+} - F_{PH}^{-} - (2F_pF_H''/F_{PH})\sin(\alpha_p - \alpha_H) \qquad (2.37)\end{aligned}$$

(North, 1965; Matthews, 1966a). For more than one derivative, the total phase probability is given by the product of the individual probabilities. The 'most probable' phase, α_{mp}, corresponding to the highest value of $P(\alpha)$, is unsuitable for use because some reflections will have distinctly non-unimodal phase distributions. Blow and Crick (1959) showed that the centroid of the distribution gives the best phase (α_{best}).

The mean square error over the unit cell for any reflection is given by:

$$\langle \Delta\rho^2 \rangle = (1/V^2)(\underline{F}_s - \underline{F}_t)^2 \qquad (2.38)$$

where $\underline{F}_s$ is the coefficient used in the synthesis and $\underline{F}_t$ is the true value. If

$$\underline{F}_t = F\exp(i\alpha_p)$$

and $\underline{F}_t$ has a probability $P(\alpha_p)$ of having the phase angle α_p, then the mean square error can be written:

$$\langle \Delta\rho^2 \rangle = (1/V^2) \int_0^{2\pi} (\underline{F}_s - F\exp(i\alpha_p))^2 P(\alpha_p) d\alpha_p / \int_0^{2\pi} P(\alpha_p) d\alpha_p \qquad (2.39)$$

Blow and Crick (1959) showed that the integral has a minimum value when:

$$\underline{F}_s(best) = mF\exp(i\alpha_{best}) \qquad (2.40)$$

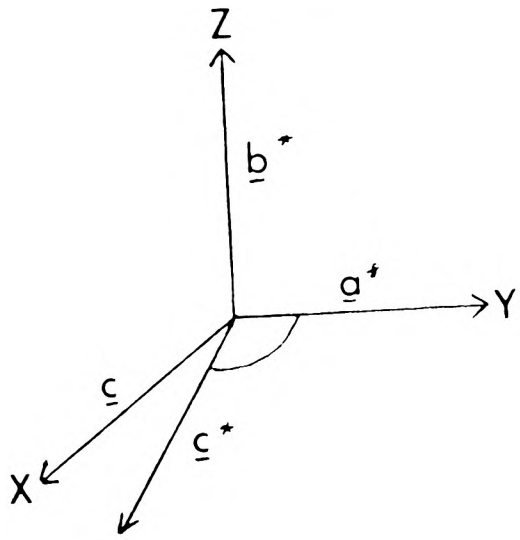
where

$$\underline{m} = \int_0^{2\pi} P(\alpha_p)\exp(i\alpha_p) d\alpha_p \; / \int_0^{2\pi} P(\alpha_p) d\alpha_p$$

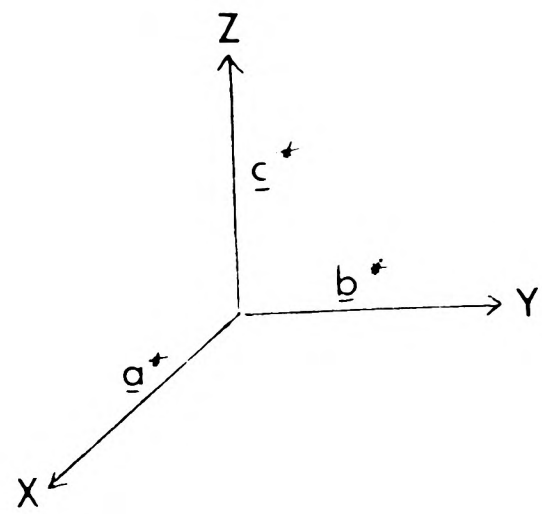
The minimum value of the integral occurs when the value of the phase angle corresponds to the centre of gravity of the probability distribution. The 'figure of merit', m , corresponds to the mean value of the cosine of the error in the phase angle for each reflection and acts as a weighting function. If the phase probability is sharp then m will have a value close to unity. The map known as the 'best Fourier' is calculated from coefficients mF_p and phases α_{best} and is expected to have the minimum mean square error from the true Fourier when averaged over the whole unit cell (Dickerson et al., 1961). (Note that that the value of m depends upon the estimate of E_j; if E_j is underestimated then m will be overestimated).

Phase information from a number of sources can be combined by multiplying their corresponding probability distributions (Rossmann & Blow, 1961). This is most conveniently done by expressing each probability function in the form:

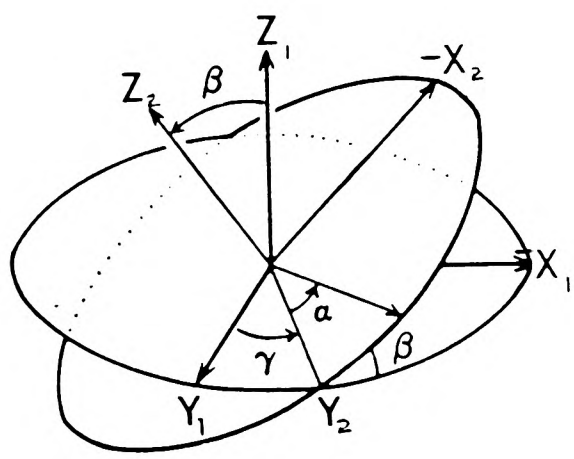
$$P_{ABCD}(\alpha) = \exp(A\cos\alpha + B\sin\alpha + C\cos 2\alpha + D\sin 2\alpha) \qquad (2.41)$$



monoclinic



orthorhombic



Euclidean space

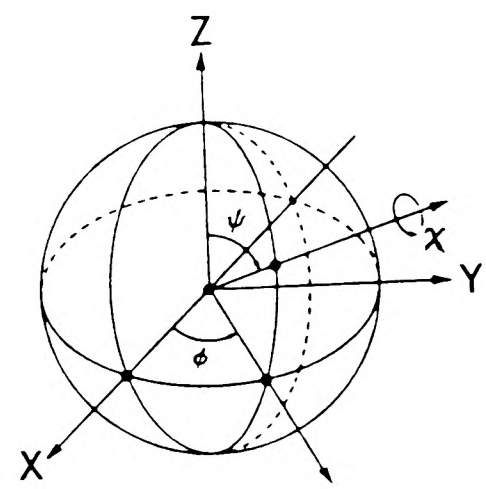
The rotated function P_1 is rotated by:

γ about $\underline{Z}_1$

β about $\underline{Y}_2$

α about $\underline{Z}_2$

clockwise looking away from the origin; in order.



Spherical polar space

The function P_1 is positioned by ψ and Φ rotations followed by a rotation of χ about the ψ, Φ vector.

$$R = \begin{bmatrix} \cos\alpha\cos\beta\cos\gamma-\sin\alpha\sin\gamma & -\cos\alpha\cos\beta\sin\gamma-\sin\alpha\cos\gamma & \cos\alpha\sin\beta \\ \sin\alpha\cos\beta\cos\gamma+\cos\alpha\sin\gamma & -\sin\alpha\cos\beta\sin\gamma+\cos\alpha\cos\gamma & \sin\alpha\sin\beta \\ -\sin\beta\cos\gamma & \sin\beta\sin\gamma & \cos\beta \end{bmatrix}$$

Figure 2.8 Rotation function conventions.

(neglecting normalisation) and adding together the coefficients in the exponent (Hendrickson & Lattman, 1970). The values of A, B, C and D can be calculated by fitting $P_{ABCD}(\alpha)$ to the probability distribution $P_{iso}(\alpha)$ (equation 2.35) using the least squares method of Hendrickson (1971).

2.5 Molecular Replacement

2.5.1 Introduction

When two or more identical molecules crystallize in the same asymmetric portion of a unit cell, then their relative orientation will be expressed in their diffraction pattern. Consider the case of two identical molecules, one of which has coordinates $\underline{X}_1$. Then the coordinates of the second molecule can be generated from the first by a linear transformation:

$$\underline{X}_2 = [R]\underline{X}_1 + \underline{T} \qquad (2.42)$$

where [R] is a rotation matrix and $\underline{T}$ is a translation vector.

2.5.2 Rotation Function

Rossmann and Blow (1962) showed that the rotation matrix could be solved from the analysis of Patterson-type functions in which intramolecular vectors were the predominant source of vectors. Their rotational correlation function can be written as follows:

$$F = \int P_o(\underline{X}_1)P_1([R]\underline{X}_1)dU \qquad (2.43)$$

where P_o is a stationary function and P_1 a function that is rotated by the matrix [R]. The function (expressed in terms of a chosen set of angular variables) is evaluated by integrating over the volume U. The function is

expected to take on a large value when the Patterson summations overlap with maximum coincidence.

Rossmann and Blow (1962) showed that the equation could be simplified for computational requirements by expanding each Patterson function as a Fourier series:

$$P_o(\underline{X}_1) = (1/V) \Sigma_{\underline{h}} F_{\underline{h}}^2 \exp(-2\pi i \underline{h}.\underline{X}_1) \qquad (2.44)$$
$$P_1(\underline{X}_2) = (1/V) \Sigma_{\underline{m}} F_{\underline{m}}^2 \exp(-2\pi i \underline{m}.\underline{X}_2)$$

and showed that the overlap function reduced to a double summation:

$$P = \Sigma_{\underline{m}} F_{\underline{m}}^2 (\Sigma_{\underline{h}} F_{\underline{h}}^2 G_{\underline{h},\underline{h}'}) \qquad (2.45)$$

where F_m and F_h are structure factor amplitudes corresponding to the Patterson functions P_1 and P_o, respectively. $G_{\underline{h},\underline{h}'}$ is an interference function whose magnitude depends on the reciprocal lattice vectors $\underline{h}$ and $\underline{h}'$, as well as the volume U over which the overlap function is integrated. The non-integral reciprocal lattice vector $\underline{h}'$ is given by:

$$\underline{h}' = \underline{m}[R] \qquad (2.46)$$

where the vector $\underline{h}'$ is the position of the reciprocal lattice point $\underline{m}$ after a rotation specified by the matrix [R]. It was shown, that if the volume U was chosen to be a sphere, then the interference function was determined entirely by the vector $\underline{H} = \underline{h} + \underline{h}'$ and that the interference function had a large value only if $\underline{H}$ was small. Thus the size of the calculation was reduced by only having to consider points $\underline{h}$ close to the non-integral points $-\underline{h}'$.

Crowther (1972) showed that the rotation function calculation could be speeded up significantly by expanding the Patterson functions in terms of

spherical harmonics. Besides being much faster than Rossmann's and Blow's method, it avoided various approximations made by the conventional method such as the need to interpolate density at unsampled reciprocal lattice points and having to truncate the interference function.

A number of practical points concerning the rotation function are worth considering. First, the rotation function is best calculated in terms of Eulerian angles because the symmetry of the rotation function is more readily expressed with them than with spherical polar angles (Tollin et al., 1966). Having evaluated a rotation function in terms of Euler angles, it is best to convert them to spherical polar angles as the result is more easily visualised with the latter. Because the overlap function is approximately proportional to F^4, then the sum of the products of many small intensities is outweighed by the contribution of a few large products. Thus only the highest terms (up to 1500 terms for Crowther's fast rotation program) need be used in calculating the rotation function. Peaks in rotation maps can be 'sharpened' in exactly the same way as a normal Fourier synthesis; by weighting down, or omitting the inner reflections of the reciprocal lattice. A shell of integration is used in Crowther's rotation function. The inner cutoff is designed to cut out very short range vectors which hold little useful information and the outer cutoff is chosen so as to maximise the ratio of intra- to intermolecular vectors in the calculation. Rotation functions can often be misleading because of accidental matching of a few very strong intensities which can dominate the map. For this reason, the rotation function is calculated for a number of resolution shells and consistent peaks looked for between the rotation maps.

2.5.3 Translation Function

Determination of the translation vector can be carried out by directly comparing observed and calculated structure factors or by matching

Patterson functions by correlation methods. In this work both approaches were used. The first was based on the residual-type translation function used by Cutfield et al., (1974) and described fully by Nixon and North (1976). In this method, the translational parameters are obtained by searching for a minimum R-factor ($= \sum_{hkl} F_{obs} - F_{calc} / \sum_{hkl} F_{obs}$) between observed and calculated structure factors obtained from a molecule with correct orientation (determined from rotation function) which is translated over the unit cell by the following method. The structure factor equation can be rewritten as:

$$\underline{F}(\underline{h}) = \sum_{i=0}^{natoms} \sum_{j=0}^{nsym} f_i[\exp 2\pi i \underline{h}([S_j](\underline{x}_i + \underline{\Delta}) + \underline{T}_j)] \qquad (2.47)$$

where natoms is the total number of atoms in the asymmetric unit, $[S_j]$ is a matrix of symmetry operations representing the rotational elements of the space group, nsym is the number of symmetry operations in the space group, $\underline{T}_j$ is a vector representing the translational elements of the space group, $\underline{x}_i+\underline{\Delta}$ are the coordinates of the molecule after a translation where $\underline{\Delta}$ is the coordinates of the centre of gravity of the rotated molecule. This equation can be rearranged as follows,

$$\underline{F}(\underline{h}) = \sum_{i=0}^{natoms} \sum_{j=0}^{nsym} f_i[\exp 2\pi i \underline{h}[S_j]\underline{x}_i][\exp 2\pi i \underline{h}.\underline{T}_j][\exp 2\pi i \underline{h}[S_j]\underline{\Delta}] \qquad (2.48)$$

If structure factors are calculated for the rotated model at the origin of the unit cell, then $\underline{\Delta}$ is zero. Thus,

$$\begin{aligned} \underline{F}(\underline{h}) &= \sum_{i=0}^{natoms} \sum_{j=0}^{nsym} f_i[\exp 2\pi i \underline{h}[S_j]\underline{x}_i][\exp 2\pi i \underline{h}.\underline{T}_j] \\ &= \Sigma\ \Sigma\ (\text{constant phase shift})[\exp 2\pi i \underline{h}[S_j]\underline{x}_i] \\ &= \underline{F}_{c_j}(\underline{h}) \end{aligned} \qquad (2.49)$$

Hence, by calculating complete P1 structure factors for a hemisphere of

reciprocal space, the nsym partial structure factors $\underline{F}_{c_j(\underline{h})}$ have already been calculated. The partial structure factors can be summed by vector addition and the correct translation vector found by searching the unit cell for the minimum R-factor between observed structure factors of the unknown crystal structure and calculated structure factors of the model at various values of $(\underline{x}_i + \underline{\Delta})$.

The other method used to search for the translation vector was based on Lifchitz's product-type correlation function:

$$T(\underline{t}) = N \Sigma_{\underline{h}} F_m(\underline{h})^2 \frac{\Sigma_{\underline{h}} F_{obs}(\underline{h})^2 F_{calc}(\underline{h},\underline{t})^2}{\Sigma_{\underline{h}} F_{obs}(\underline{h})^4 \Sigma_{\underline{h}} F_{calc}(\underline{h},\underline{t})^2} \qquad (2.50)$$

where N is the number of asymmetric units in the unit cell and $F_m(\underline{h})$ are structure factors of the isolated model at the origin of a P1 cell.

Lifchitz showed that the function could be considered as a product of two functions; one of which measures the degree of agreement between observed and calculated structure factors and the other, a measure of the degree of intermolecular overlap (Harada et al., 1981). A high value for the correlation function is expected when the observed and calculated structure factors agree and where the intermolecular overlap function is small.

2.5.4 Rigid Body Refinement

Having obtained an approximate rotation matrix and translation vector from molecular replacement methods (or from other sources), the position of the protein model can be refined using reciprocal space methods. In this work, the constrained-restrained parameter structure factor least squares program (CORELS) written by Sussmann was used (Sussmann et al., 1977).

There are a number of advantages to using CORELS for improving a molecular replacement solution. The method is designed to work with low resolution data and thus the radius of convergence is quite large (up to

5Å). Further, the internal geometry of the model can be refined by treating the model as a number of constrained groups. A constrained group might be a subunit, a domain, a secondary structure element and so on. Each constrained group can be refined as a function of translation ($\underline{t}$), rotation ($\underline{R}$), thermal parameters (B) and dihedral rotation angles (ψ). The refinement of the group parameters takes place whilst maintaining proper stereochemistry of bond distances, angles and van der Waals contacts for restrained atoms.

The quantity Q minimized in the least squares procedure is:

$$Q\ (\underline{t},\underline{R},\psi,B) = w_P DF + w_D DD + w_T DT \qquad (2.51)$$

where w_P, w_D and w_T are overall weights and

$$DF = \Sigma_{\underline{h}}\ w_{\underline{h}}(\ F_{\underline{h}(obs)} - k\ F_{\underline{h}(calc)}\ )^2$$

where k is an overall scale factor and

$$DD = \Sigma_d\ w_d(D_{d(obs)} - D_{d(calc)})^2$$

where the summation is over all subsidiary distance restraints, $D_{d(obs)}$ is the 'ideal' distance between specified pairs of atoms and $D_{d(calc)}$ is this distance calculated from the model and finally,

$$DT = \Sigma_i\ w_i \Sigma_j\ (X_T(i,j) - X(i,j))^2$$

where the summation is over i total target atoms, $X_T(i,j)$ is the axial coordinate (with three components j) of some target atom and X(i,j) are the coordinates of atoms in the model. The second term in equation 2.51 restrains the stereochemistry of the model and the third term restrains the

structure from moving away from a specified set of target coordinates.

The group derivatives with respect to the positional parameters are obtained by differentiating equation 2.51 and application of the chain rule. For constrained-restrained refinement, where P is a function of the group positional parameters, the normal matrix to be solved is:

$$[\underline{A}]\Delta\underline{p} = \underline{k} \qquad (2.52)$$

where $A_{nm} = (w_P \Sigma_{\underline{h}} w_{\underline{h}} (\delta F_{\underline{h}(calc)} / \delta P_n)(\delta F_{\underline{h}(calc)} / \delta P_m)$

$$+ (w_D \Sigma_d w_d (\delta D_{d(calc)} / \delta P_n)(\delta D_{d(calc)} / \delta P_m)$$

and $k_n = (w_P \Sigma_{\underline{h}} w_{\underline{h}} (F_{\underline{h}(obs)} - F_{\underline{h}(calc)})(\delta F_{\underline{h}(calc)} / \delta P_n))$

$$+ (w_D \Sigma_d w_d (D_{d(obs)} - D_{d(calc)})(\delta D_{d(calc)} / \delta P_n))$$

The least squares normal equations are solved by a conjugate gradient iterative method (Hestenes & Stiefel, 1952). However, because full matrix methods are computationally expensive, a 'sparse' matrix is solved. The 'sparse' matrix contains blocks of intragroup cross derivatives along the diagonal and off-diagonal blocks containing cross derivative terms between groups sharing at least one restraint condition. All other elements are assumed to be small and are set to zero. The 'sparse' matrix occupies only one to five per cent of the size of the corresponding full matrix.

The Oxford version of CORELS includes a solvent correction term. The atomic scattering factors are modified at low angle to allow for the effective displacement of bulk solvent by the protein atoms (which are modelled with Gaussian spheres) (Stuart & Leslie, unpublished).

2.6 Phase Improvement Methods

2.6.1 Introduction

There are a number of methods available by which the quality of a phase set can be improved. Phase improvement methods are based on the application of some chosen constraints in real or reciprocal space. Real space procedures are based on some prior knowledge of the electron density distribution. Such knowledge may concern the uniform character of the solvent (solvent flattening, eg. Hendrickson et al., 1975), the positive continuous electron density feature of the protein backbone (density modification, eg. Bhat & Blow, 1982) or the presence of non-crystallographic symmetry (symmetry averaging, eg. Bricogne, 1976). Traditional direct methods are examples of the use of reciprocal space constraints. The application of constraints in one space cannot be directly made in the other space because of their non-local nature in the other space. For example, experimentally derived structure factors put constraints on individual reflections and thus are localised in reciprocal space. In real space these constraints are non-local because they involve all the electron density simultaneously. Phase improvement methods are enhanced by linking constraints in real and reciprocal space and making use of the inverse Fourier transform relationship between the two spaces. The methods discussed here rely, as multiple isomorphous replacement and molecular replacement do, on some partial structure information being available to resolve phase ambiguities.

2.6.2 Single Isomorphous Replacement and Solvent Flattening

A Fourier map can be calculated from a single isomorphous derivative by taking the weighted mean of the two possible phase solutions. An electron density map calculated from a good single isomorphous derivative (SIR) can be regarded as a superposition of two Fourier maps; one being a correctly

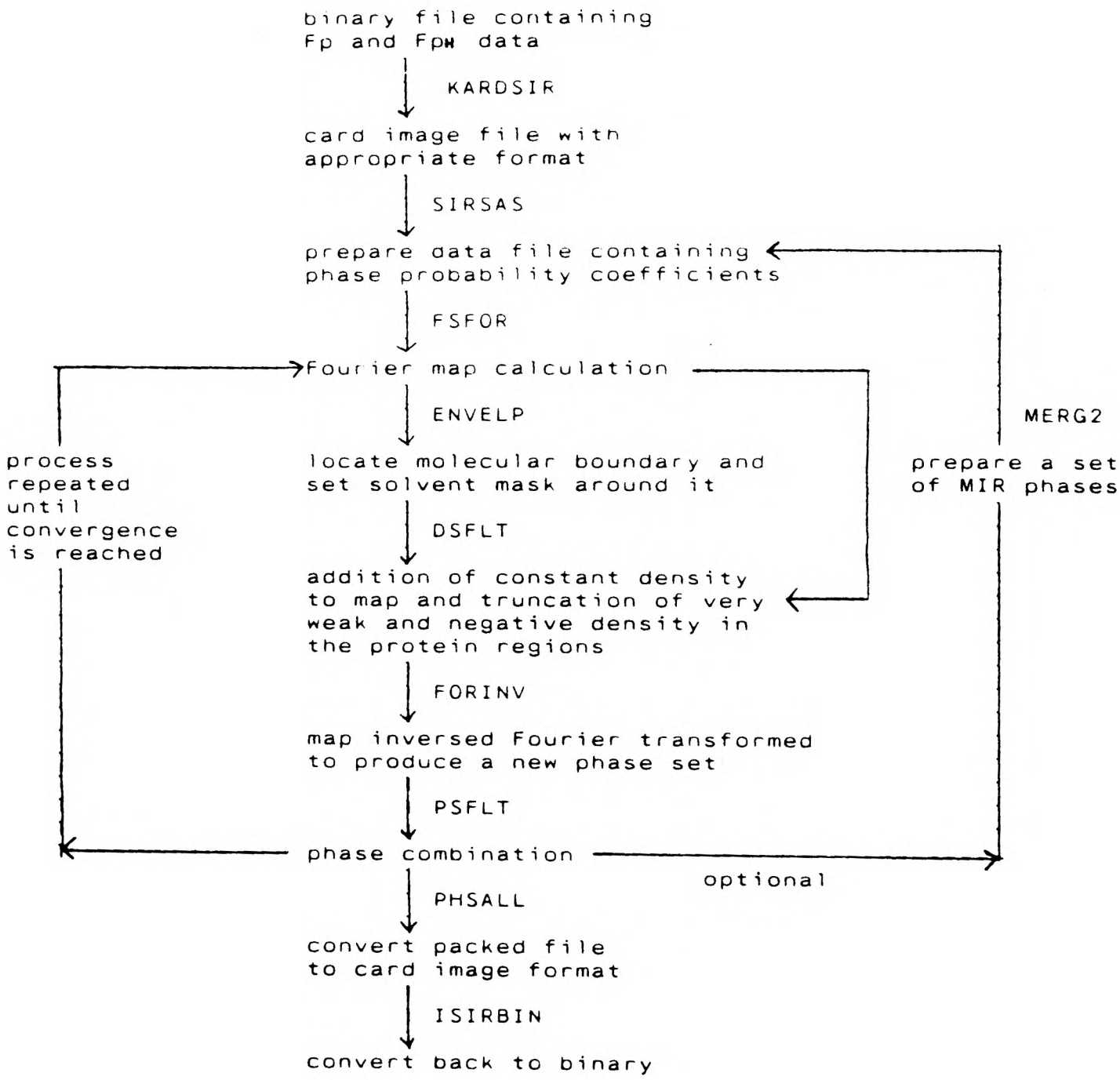


Figure 2.9 Flow diagram of steps taken to improve the quality of a protein map by the method of Wang (1985). Names of computer programs given in block letters.

phased protein electron density map, and the other being a map produced from random phase information (Blow & Rossmann, 1961). Providing the noise constituting the latter map is random and there exists a significant difference between the electron density levels of the protein and solvent, then it is possible to extract the gross image of the protein molecule from a map calculated with SIR (or SAS - single wavelength anomalous scattering) data alone.

Wang (1985) has developed a solvent flattening method by which the gross image of the protein molecule serves to break up the phase ambiguity. A flow diagram of the procedure is given in figure 2.9. In this method, a new map is constructed from an SIR map by recalculating density at each grid point which is proportional to the weighted sum of densities within a sphere of radius R from that grid point in the original map:

$$\rho_{new}(j) = k \sum_i w_i \rho_{old}(i) \qquad w_i = 1-r_{ij},\ r_{ij}<R \text{ and } \rho_{old}>0 \qquad (2.53)$$
$$w_i = 0,\ r_{ij}>R \text{ or } \rho_{old}=0$$

where r_{ij} is the distance between grid points, k is an arbitrary constant and ρ contains no F(000) term. The molecular boundary is revealed by defining some threshold density level, below which a solvent region (or mask) is assigned. The density level of the solvent regions is averaged and a constant level of density, ρ_c, is added to the entire map to reduce the noise level in the protein regions (and so partially compensate for the missing F(000) term). The value of ρ_c is calculated from:

$$(\rho_c + \langle\rho_{solv}\rangle)/(\rho_c + \rho_{max}) = S \qquad (2.54)$$

where $\langle\rho_{solv}\rangle$ is the average solvent density and ρ_{max} is the maximum protein density. The value of S is determined empirically by choosing a value that optimises the convergence of the procedure (see below). It is

interesting to note that the best values of S tend to be considerably less than the true value calculated from known crystal structures (Wang, 1985). The new map is Fourier inverted to obtain protein phases and the unimodal distributions of the phase set used to select the most likely mode of the original SIR distribution. The whole procedure is repeated until convergence is reached. The method has been extended to work with MIR information as well.

It is important to note that the method will not work if the data suffers from the effects of systematic experimental error or non-isomorphism because in order to correctly resolve the phase ambiguity, one of the phase choices (calculated by the SIR method) must be valid in the first place.

2.6.3 Symmetry Averaging

The use of non-crystallographic symmetry in phase improvement is well established. It has played an essential role in the determination of a number of virus structures at high resolution. Non-crystallographic symmetry was first proposed to be a source of phase information by Rossmann and Blow (1962). However, their expressions for applying constraints in reciprocal space were computationally expensive to implement. Bricogne (1974, 1976) simplified the problem by reformulating the expressions so that the constraints were made in real space.

The symmetry averaging procedure is started by computing an electron density map (from isomorphous phases or other sources) and applying the non-crystallographic symmetry to the density within (or between) the envelope(s) containing the structures of interest. Those parts of the map not included in the envelope are set to a constant value given by their average electron density (solvent averaging). The modified electron density map is Fourier transformed to provide refined phases suitable for starting a new cycle of iteration and the procedure is repeated until convergence.

2.7 Data Collection Methods

2.7.1 Introduction

In this section a brief overview of data collection methods is presented and is followed by a more detailed discussion of the methods used in the studies of MnSOD. Any consideration of data collection methods is best started off by a brief account of methods of X-ray production.

When a beam of high energy electrons are accelerated through a voltage in a vacuum to hit a target, a fraction of the energy is converted into X-rays whilst the remainder is dissipated as heat. For any particular target element, a critical minimum voltage is required to get emission. Conventional sources used in protein crystallography have copper target areas set into sealed-off tubes under high vacuum. Rotating anode instruments are becoming increasingly common due to higher fluxes and greater focussing properties than conventional tubes. The anode is rotated to diminish heating by presenting a large surface area to the electron flux but suffers from maintenance problems associated with maintaining the system at high vacuum.

Synchrotron sources of X-rays are proving popular due to their high photon flux and variable wavelength capabilities (Greenhough & Helliwell, 1983). Their high flux enables diffraction measurements of very small crystals and shortens data collection times considerably. With a synchrotron, electrons are accelerated by a linear accelerator and then injected into a circular ring. The electrons are constrained to move in a circular orbit by a series of bending magnets. The electrons are injected as equally spaced bundles and are constantly accelerated by pulses from a radiofrequency field. As the electrons approach the speed of light they emit a narrow cone (c.a. 0.3mrad) of radiation tangential to the instantaneous radial direction of flight.

The majority of sealed-off tubes and rotating anode sources in the lab are monochromatised by means of metal absorption filters. The radiation from the synchrotron source at Daresbury is focused and monochromatised by means of cross-curved mirrors and triangular-curved crystal monochromators. The crystal monochromators work by reflecting the direct beam through a single crystal set at the appropriate Bragg angle.

The intensity of a diffracted beam depends on a number of factors which are summarised in Darwin's formula:

$$E(\underline{h}) = (I_o/\omega)\lambda^3(e^4/m^2c^4)((1 + \cos^2 2\theta)/2)LA(V_x/V^2)\,(F(\underline{h}))^2 \qquad (2.55)$$

where the total energy (E) of a diffracted beam is expressed in terms of a number of familar constants (the wavelength, λ, the electronic charge, e, the electronic mass, m, and the velocity of light, c) and a number of correcting terms that must be applied to the incident intensity, I_o. The angular velocity term, ω, takes account of the fact that the reflecting position occurs over a small angular range due to crystal mosaicity (ie. imperfections in the crystal). The $(1 + \cos^2 2\theta)$ term corrects for the partial polarisation of the diffracted beam and the Lorentz factor, L, accounts for the relative time each reflection spends in the reflecting position. The crystal, of volume V_x and unit cell volume V also requires correction for absorption, A.

Three data collection methods were used in the structural studies of MnSOD and are discussed in the following sections. However, there are a number of advances in collection methods recently which are worth reviewing briefly. The first concerns the re-emergence of Laue photography as a method of data collection in protein crystallography. In a Laue experiment, a polychromatic source of X-rays are allowed to fall on a stationary crystal. Each Bragg reflection is satisfied by a different wavelength and may contain contributions from several reciprocal lattice points (harmonic

overlaps). The intensity of the Laue reflection is integrated over wavelength rather than angle. When used together with a synchrotron, the method offers a significant reduction in exposure times (from minutes to fractions of a second) over conventional methods. Further, the ability to record thousands of reflections simultaneously opens up the prospect of time-resolved kinetic experiments (for a review, see Moffat et al., 1984). How useful the Laue method will become is dependent on overcoming a number of data processing problems. Methods of deconvoluting reflection intensities from harmonic overlaps and normalising the intensity data set to account for wavelength-dependent variation need to be devised. Some progress has already been made with these problems (Helliwell, 1985).

Within the last few years there has been significant progress in the development of two dimensional position-sensitive detectors. These area detectors permit the simultaneous recording of a large number of reflections with the type of accuracy usually associated with single counter diffractometers. Present counters (proportional or scintillation type) of diffractometers are inefficient because they can measure very few reflections at any one time. Although film has excellent properties in terms of the ability to record large number of reflections simultaneously and high quantum efficiency, it suffers from poor signal-to-noise ratios. Area detectors are aimed to overcome these problems. At present there are three area detectors at an advanced stage of development for protein crystallography and these are discussed below.

The television detector, developed by Arndt and marketed by Enraf-Nonius, records the diffraction pattern of a crystal on a phosphor screen connected by fibre optics to an image intensifier (Arndt, 1984). The size of the screen and point-to-point resolution implies that a 3Å data set could be collected at a single detector setting on small unit cell-sized protein crystals (Harrison, 1984). The television detector can also cope with the high fluxes from synchrotron sources. The multiwire proportional

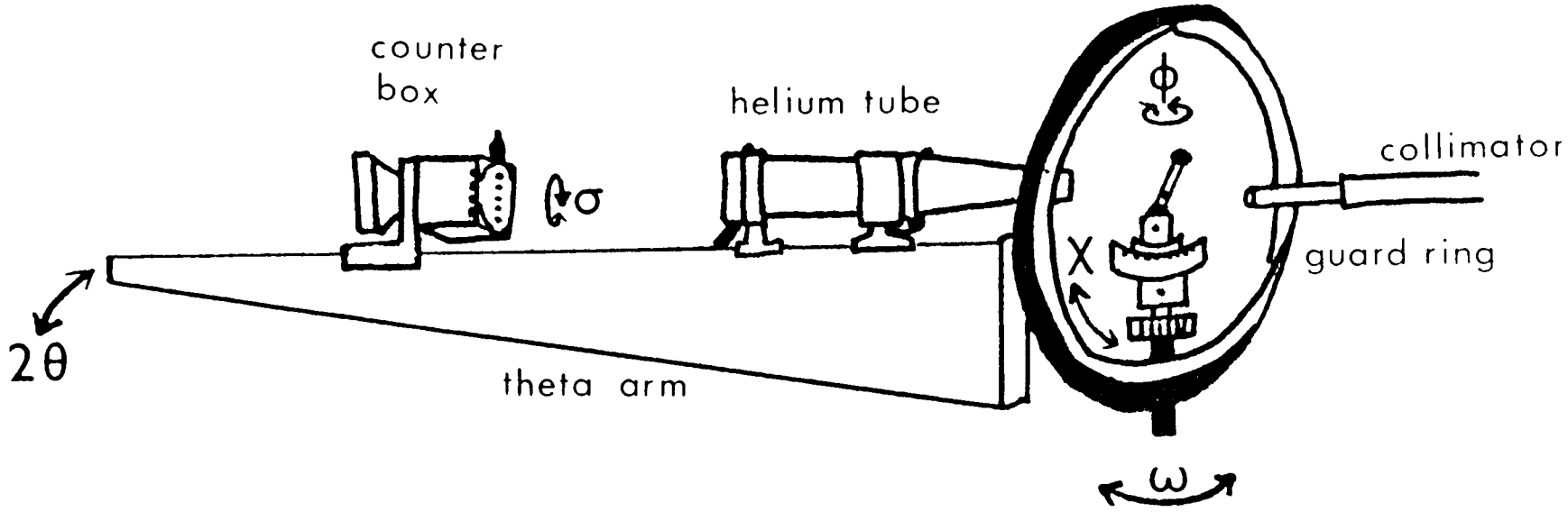


Figure 2.11 Schematic drawing of part of the five circle diffractometer.

counter, developed by Xuong at San Diego, consists of a xenon-filled ionisation chamber fitted onto a standard diffractometer (Cork et al., 1974 & 1975). It suffers from the requirement of large crystal-to-detector distances and multiple detector settings but a number of structures have been solved on it already (Matthews et al., 1977; Xuong et al., 1978; Xuong et al., 1985). Both problems have been overcome to some extent by the use of helium filled tubes to cut down air absorption of X-rays and the use of multiple arrays of detectors. The multiwire system developed by Burns (Xentronics, Massachusetts) has higher spatial resolution than Xuong's and has been fitted to a modified oscillation camera. The xenon-filled detectors are limited by dead-time and so detract from their use on a synchrotron source.

2.7.2 Diffractometry

This discussion will be limited to aspects of the five circle Oxford diffractometer which was used exclusively in the MnSOD work. For a complete account of the design, construction and operation of the five circle diffractometer the reader is referred to Banner et al., (1977). The five circle is a modified version of a Hilger and Watts Y230 four circle diffractometer. The major modification consists of an extension of the θ arm and the replacement of the old detector by a counter box of five xenon-filled side-window proportional detectors (see figure 2.11). The box is mounted vertically on the θ arm and is able to rotate about a horizontal axis defined by the optical centre of the instrument and the centre of the middle counter. (This axis is referred to as the σ axis). The box can be moved along the arm to allow for adjustment of the crystal-to-counter distance and can be centred on reflections by means of vertical and horizontal adjustments and the use of half-slits in front of the counters.

There are a number of ways used to reduce air absorption of X-rays before the rays reach the detectors. The space between crystal and counter

box is occupied by a plastic tube fitted with 'Mylar' windows containing helium at a pressure of about three pounds per square inch above atmospheric pressure. The window closest to the crystal contains a replaceable non-limiting aperature to reduce backstop scatter. A limiting aperture which can be inserted immediately before the counter windows and a non-limiting guard ring extending from collimator to crystal both act to cut down background scatter significantly.

The five circle diffractometer was designed for the quasi-simultaneous measurement of five reflections using equatorial geometry (ie. both incident beam and the diffracted beam that is directed to the centre counter lie in the same horizontal plane) and a modification of the flat-cone setting (ie. constrained to $\omega = 2\theta$). In this setting, the so-called optimum setting, the difference in the ω values of reflections in the central and outer counters ($\Delta\omega$) is minimised for a given quintuplet of reflections by an additional rotation ($\Delta\psi$), about the scattering vector for the central reflection (achieved by appropriate adjustments of Φ, χ, and ω). This has the effect of making pairs of reflections, symmetrically disposed about the centre counter reflection, pass through the Ewald sphere simultaneously. The flat-cone setting introduces a blind region of inaccessible reflections close to the mounting axis which can be measured by single counter geometry (see below) or by re-defining the collection axis (and moving the counter box) in five counter geometry.

The value of $\Delta\omega$ increases with decreasing resolution to the point where it is more efficient to collect data in single counter mode than having to scan over large angular (ω) ranges. Single counter data are collected using the symmetrical-A setting in which ω is constrained to equal θ. In this setting, the χ circle bisects the angle between the incident and diffracted beam.

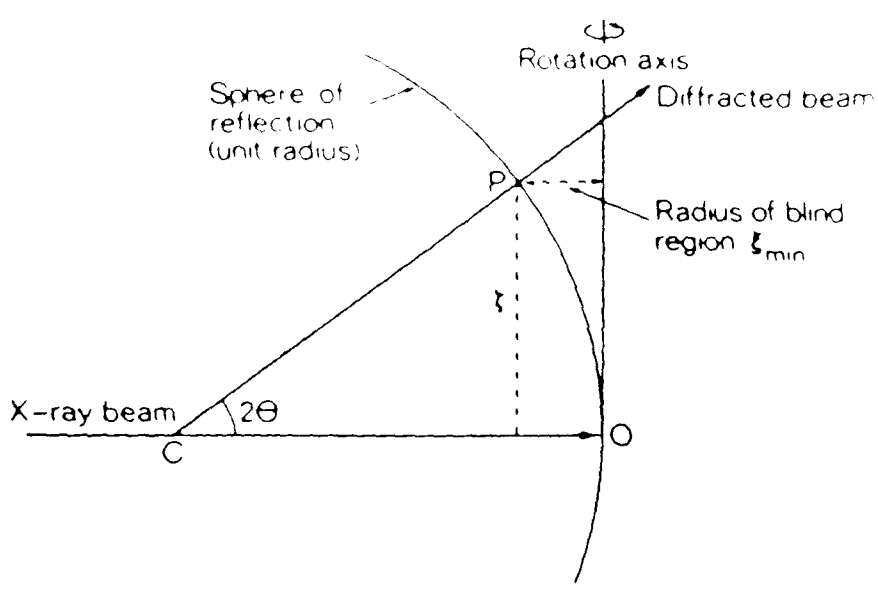


Figure 2.12 The blind region arising from normal beam geometry.

2.7.3 The Rotation Camera Method

This method has been described in great detail by Arndt and Wonacott (1977) and so will only be discussed briefly here. In this method, the crystal is rotated about an axis (the Φ axis) normal to the X-ray beam, in a series of angular steps. For each step, the crystal is continually oscillated. For any given oscillation range of the crystal, all the reflections passing through the Ewald sphere are recorded simultaneously by the detector. Because reciprocal lattice points subtend a finite angle (Δ) at the centre of the sphere of reflection, some reflections will not pass completely through the Ewald sphere and so will be only partially recorded for a given oscillation range. The angular range of any reflection is made up of various contributions as follows:

$$\Delta = \delta_c + \delta_f + \eta + \delta_\theta \qquad (2.56)$$

where δ_c is the angle subtended by the crystal at the X-ray focus, δ_f is the angle subtended by the X-ray tube at the crystal, η is the mosaic spread of the crystal and δ_θ is the dispersion spread of the wavelength. The mosaic spread takes account of crystal irregularities such as impurities, defects and lattice thermal motions. The size of the oscillation range is chosen so that the number of fully recorded reflections is maximized without introducing a significant number of overlapping reflections. Arndt (1968) showed that the maximum oscillation range, $\Delta\Phi_{max}$, can be calculated from the relationship:

$$\Delta\Phi_{max} \leq (p^*/d^*_{max}) - \Delta \qquad (2.57)$$

where d^*_{max} is the maximum resolution desired and p^* is the reciprocal lattice vector corresponding to the limiting axis for a given crystal mount. Another consideration in choosing the appropriate oscillation range is that the signal-to-noise ratio measured by the detector is inversely

Figure 2.13 Flow diagram of the steps involved in processing diffractometer data. The names of computer programs are given in block letters.

proportional to the size of the range.

The normal beam geometry of the rotation method gives rise to a blind region of inaccessible reflections (figure 2.12). Blind region data can be collected by rotating the crystal about a different axis. However, the percentage of reflections falling into the blind region is small; less than 3% at 2.5Å resolution with Cu Kα radiation.

In the work presented in this thesis, the detector was CEA Reflex 25 X-ray film. The rotation data were measured using an Arndt-Wonacott oscillation camera controlled by an Enraf-Nonius microprocessor. The films were all scanned using a Joyce-Loebl Scandig-3 rotating drum microdensitometer.

2.7.4 The Precession Camera

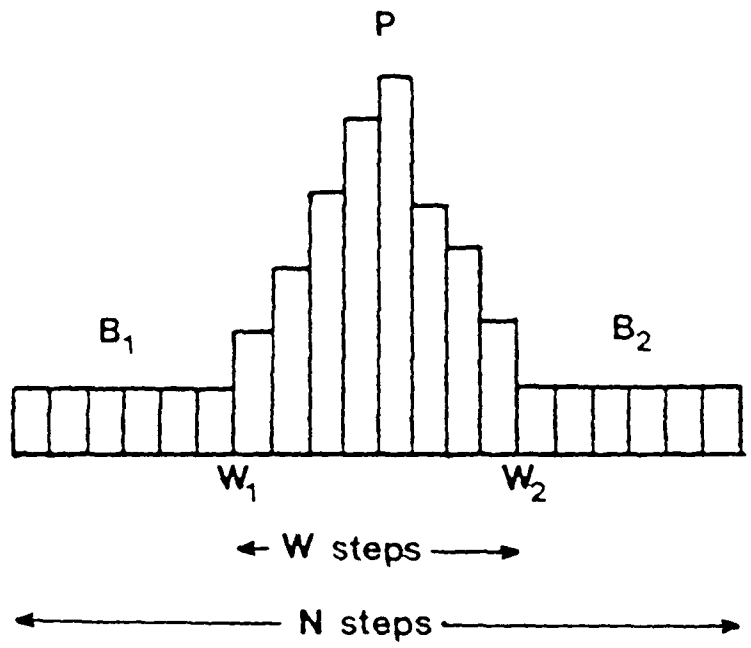
In the MnSOD studies, the precession camera was used to determine the space group of the crystals and to screen for potential heavy atom derivatives. Buerger (1942, 1964) introduced the camera as a means of recording an undistorted view of the reciprocal lattice. The use of annular screens allows for selected layers of reciprocal space to be recorded. The operation of the precession camera is covered in detail by Blundell and Johnson (1976).

2.8 Data Processing Methods

2.8.1 Processing of Diffractometer data

All diffractometer data collected during the course of this work was processed on the University's ICL 2988 computer using a suite of programs developed over the years in the Laboratory. A flow diagram of the program suite is presented in figure 2.13 and a full list of the programs together with their authors is presented in appendix three.

The first step is to transfer peak profile data from magnetic tape to



$$B_1 = \sum_{j=1}^{W_1-1} c_j \qquad P = \sum_{j=W_1}^{W_2} c_j \qquad B_2 = \sum_{j=W_2+1}^{N} c_j$$

$$I = P - [(B_1 + B_2)(N / (N - W))]$$

where I is the integrated intensity for peak P and

$\sigma^2 = P + [(B_1 + B_2)(N / (N - W))^2]$ where σ is the standard deviation of the intensity measurement.

Figure 2.14 Background-peak-background calculation of the integrated intensity of a measurement performed by the computer program PROPRINT.

the 2988 computer. This is achieved using the program, PROPRINT, which also calculates the integrated intensity for each measurement by a background-peak-background method as described in figure 2.14. The integration window is imposed on the central counter box peak by inspection and suitable window positions imposed on the other peaks by an appropriate shift of the central peak window value on the basis of the $\Delta\omega$ value used in data collection.

The next step is to correct the data for absorption. The correction relies on semi-empirical methods because of difficulties involved in estimating absorption contributions from scattering material surrounding the crystal. The present program suite allows for two types of correction; that described by North, Phillips and Mathews (1968) and by Huber and Kopfmann (Kopfmann & Huber, 1968; Huber & Kopfmann, 1969). Both rely on the fact that once a crystal is aligned so that a reciprocal lattice point is in the reflecting position, a rotation about the diffraction vector by an azimuthal angle Ψ will alter the total path length of X-rays through the crystal without violating the Bragg condition.

In the North, Phillips and Mathews (NPM) method, absorption curves are collected for strong mounting axis reflections by simply measuring intensity values for different values of Φ. Since mounting axis reflections have their diffraction vector lying parallel to the rotation axis, then variations in Ψ can be brought about by rotating on Φ. It is assumed in calculating the absorption factor A_Φ that

$$A_\Phi = (1/2)(A_p + A_s) \qquad (2.58)$$

where A_p and A_s are the absorption factors for the primary and secondary beams respectively. The absorption curves are used to correct data on the mounting axis and adjacent levels of reciprocal space.

A more sophisticated approach has been described by Huber and Kopfmann.

The method assumes that the transmission factor, T_{ps}, for a reflection with an incident path, p, and a diffracted path, s, may be approximated by:

$$T_{ps} = T_p.T_s \qquad (2.59)$$

where T_p and T_s are the partial transmission factors for the primary and secondary beams, respectively. A spherical transmission surface is calculated from intensity measurements of a series of reflections rotated about their reciprocal lattice vector. Each measurement gives an equation relating the true and observed intensity:

$$I(\Psi)_{obs} = I(\Psi)_{true}.T_p.T_s \qquad (2.60)$$

The equations can be reduced to a set of linear equations by taking logarithms and solved for the true intensity values. The transmission surface is constructed by dividing the surface into levels above and below the equatorial level and each level is divided into a number of evenly spaced grid points separated by some chosen angular distance α. The latitude furthest from the equator is designated as the critical level where there are less contributions from intensity measurements to the grid points. The contribution of any reflection to a given grid point is weighted by an inverse proportion to the angular distance between the beam vector and the grid point. The program STROGBLIT is used to choose suitable strong reflections for measurement so that the transmission surface is covered by a network of intersecting curves symmetrically displaced about the equator.

Next counter-to-counter scales are calculated by determining scale factors between batches of measurements of the same reflections measured in the different counters. The scales are calculated by the method of Fox and Holmes (1966) using the program SSM. These scales, absorption, Lorentz and

polarisation corrections are applied to the data using the program LPOPT5. Radiation damage is estimated from plots of intensity measurements versus time and applied using the program RADDAM.

The resolution shells of corrected data are ready to be sorted, scaled and merged together by the method of Fox and Holmes (1966) using the program SSM. Scales are deemed to be correct when shifts in the calculated values are less than one part in ten thousand. Equivalents can be merged as required and an unique set of intensities output. Negative intensities are treated by the program TRUNCATE. French and Wilson (1978) showed that negative intensity measurements do contain some useful information and developed a method of assigning positive values to them, assuming that the whole intensity set conforms to a normal Wilson distribution.

The program FRIEDANAL is used to analyse the agreement between intensity measurements of equivalent reflections. Scale factors are calculated between Friedel pairs as a function of volume of reciprocal space (in terms of cylindrical polar coordinates of level, Φ and Ψ^2) and the scales can be applied to the data using the program FRIEDCORR.

At this stage, it maybe desired to scale the data to another set of data; for example, to scale native and derivative data sets together. The program CAD combines data sets from different sources and can reduce data to a set of structure factor amplitudes. There are two types of scaling programs available in the computer program suite. The program ANSC analyses one set of data against another as a function of the Miller indices, resolution and in volumes of reciprocal space (as a function of cylindrical polar coordinates). Any trends can be corrected for, using ANSC or ANSCCORR. The other scaling program, LOCAL, is based on a method described by Matthews and Czerwinski (1975). Each reflection is assigned an individual scale based on the average relative scale of neighbouring reflections in reciprocal space. A sphere containing a specified number of reflections is centred on the reflection of interest and contributions from

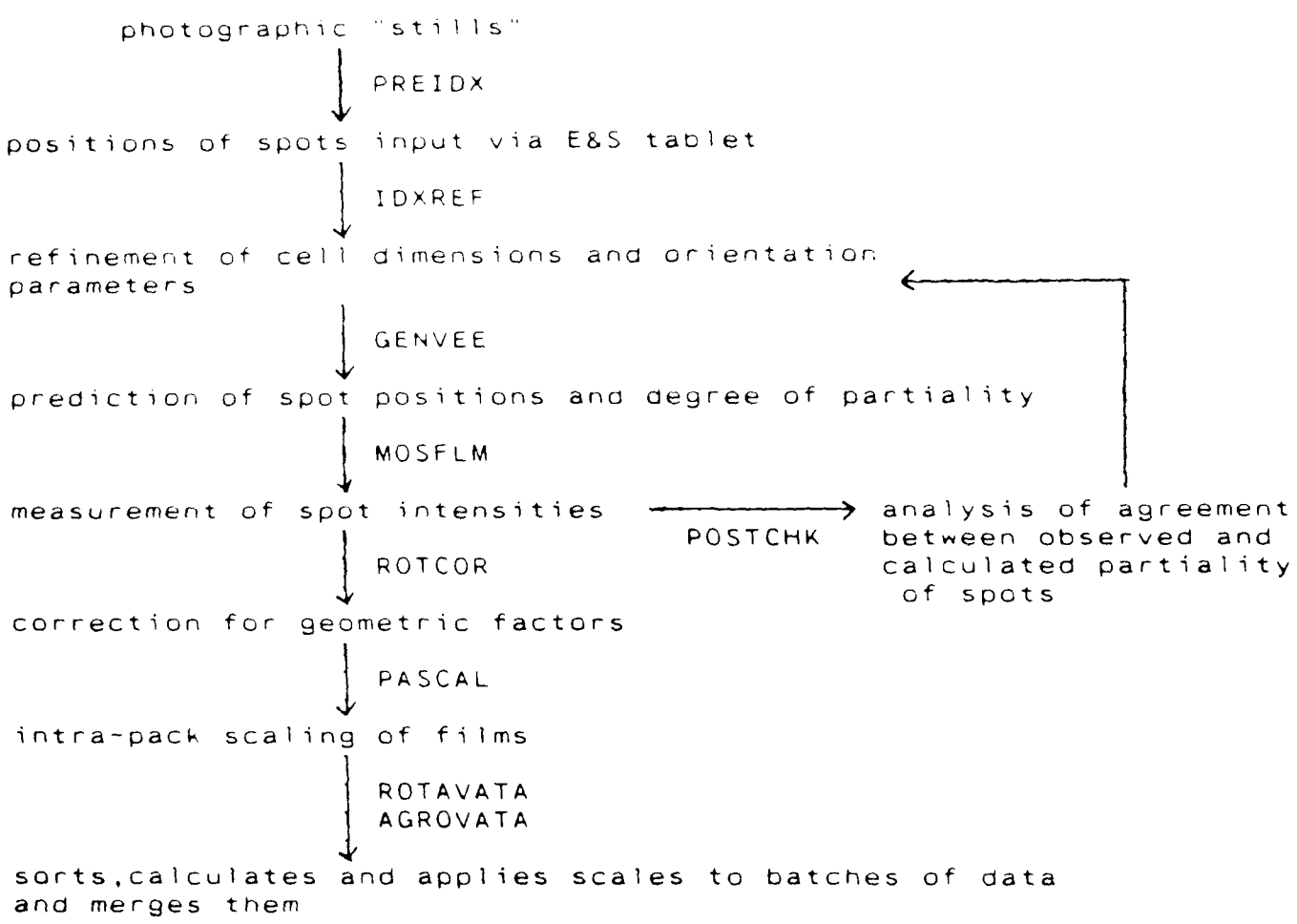


Figure 2.15 Flow diagram of steps taken to process oscillation film data. Names of computer programs are given in block letters.

other reflections within the sphere are given weights which decrease with their radial distance from the reflection in reciprocal space. The larger the sphere, the smaller will be the probable error in the local scale factor, but the less responsive will the scale factor be to local errors. The optimum local scaling area is chosen for when the average systematic variation in the local scale factor exceeds the average error in the estimation of the local scale factor by the greatest amount.

2.8.2 Rotation Photography

The suite of programs used to process the data in this work came from Daresbury Laboratory and based on the system developed at the MRC, Cambridge (Wonacott, 1980). A flow diagram of the program suite is shown in figure 2.15.

The first step is to convert the film data to digitised form for processing. This is done by measuring the optical density of the film by shining a light onto the film and measuring the ratio of the incident to transmitted light. The blackening of the film grains is proportional to the number of X-ray quanta absorbed by the film and so film is a reasonably linear detector. Nevertheless, there exists some non-linearity so that measured optical densities have to be corrected for using a 'look-up' table of response values.

The most important step in film processing is the determination of crystal missetting angles, lattice and camera constants. These parameters are used in indexing and estimating the degree of partiality of reflections on the film and are determined from the measurement of the coordinates of strong reflections from an appropriate set of still photographs. Also required are reasonably accurate estimates of cell dimensions, a rough knowledge of the crystal to camera orientation and a knowledge of the relationship between the fiducial spot positions (reference marks made at the corners of the films) and the true film centre (the fiducial

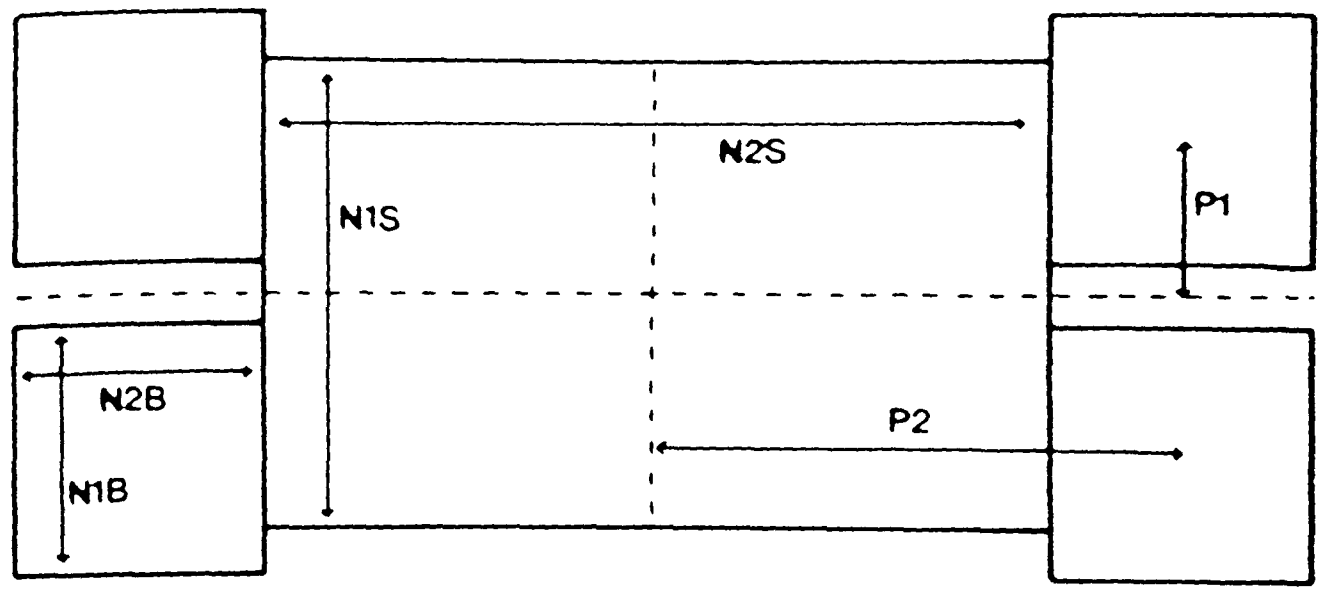


Figure 2.16 Spot and background window arrangement used in MOSFLM.

constants). The predicted and measured positions of the reflections on the stills are made to converge by refining the various parameters using the program IDXREF. Accuracy is judged from the root mean square residuals for spot positions and distances from the Ewald sphere, and from the standard deviations of refined parameters.

The spot prediction program GENVEE calculates spot positions on the basis of parameters calculated from IDXREF and the rotation angle position and range. For good predictions, the cell parameters must be correct to two parts in a thousand and the orientation angles within 0.02^{o} of the true values. As a precaution, the reflecting range of the crystal is usually set to a slightly larger value than its actual value to allow for errors in the lattice and orientation parameters. However, the more the reflecting range is overestimated, the number of predicted overlapping of spots increases and a partial bias is introduced. At this stage, predicted reflections are designated as unmeasurable if they overlap, never pass fully through the Ewald sphere or are outside the physical limits of the film.

The digitised spots are then ready for integration by the computer program MOSFLM. The first step is to refine the fiducial constants and crystal-to-film distance against the position of a few intense reflections from the strongest film in a film pack (the 'A' film). The initial search is made on spots situated close to the film centre and then is extended to include spots in outer portions of the film. The program sets a measurement box around each reflection and adjacent smaller boxes are arranged so as to sample the nearby background (figure 2.16). The box sizes are allowed to vary with the position of the spot on the film so as to allow for the effect of the oblique incidence of the X-ray beam on the film. The program calculates and displays averaged spot profiles so that box parameters can be checked and modified as necessary. The program takes account of film distortion by means of refining tilt, twist and bulge parameters. Two integration passes are made through the data. Between the first and the

second pass, the integration box may *be* moved towards the observed centre-of-gravity by some pre-set maximum shift. The final centre-of-gravity shifts between calculated and the true positions of spots on a film are calculated so that any systematic errors or large shifts can be identified. Spots are rejected if their observed and calculated positions vary too much or if they lie in areas of high background gradient. Standard deviations are calculated from background measurements taking into account film granularity.

The quality of the measurements can be checked by the program POSTCHK. The program compares the predicted and observed degree of partiality of reflections common between two packs of film. If necessary, POSTCHK can output a file of suitable reflections for input back into IDXREF for post-refinement of crystal parameters. POSTCHK is most useful when small crystal slippages have occured during data collection.

Having integrated the films, the data are corrected for oblique incidence of the X-ray beam, Lorentz and polarisation factors by the program ROTCOR. Then all film data within a pack are scaled together on the basis of common reflections by the program PASCAL. The scale, A, is calculated for each film in a pack by:

$$I_m/I_n = A^{\sec 2\theta} \qquad (2.61)$$

where I_m and I_n are the integrated intensities of a reflection on an adjacent film and the factor $\sec 2\theta$ accounts for the oblique incidence of the diffracted beam on the film.

The inter-pack scaling is performed in two stages. The program ROTAVATA calculates scale factors between overlapping batches of data by the method of Fox and Holmes (1966) and analyses the data in resolution and intensity ranges. The program AGROVATA applies the scale factors, adds together partially recorded reflections and monitors bad agreements between repeated

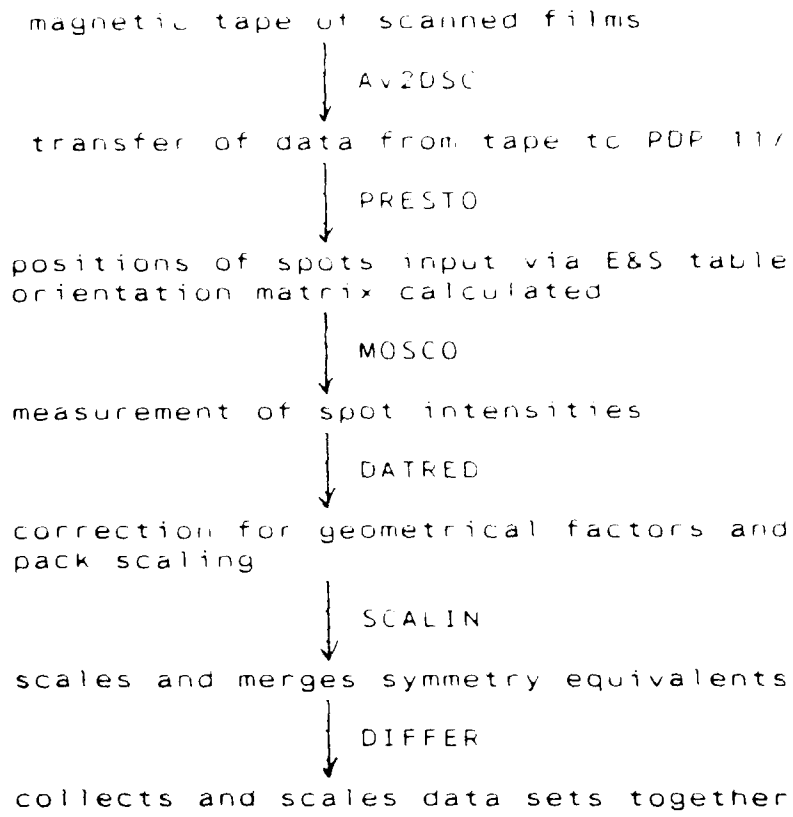


Figure 2.17 Flow diagram of steps taken to process precession film data. Names of computer programs are given in block letters.

and symmetry-related reflections (rejecting bad data) and averages them as required to produce a final set of structure factor amplitudes.

2.8.3 Precession Photography

All precession film processing was carried out using the Oxford PDP 11/70 computer suite of programs. A flow diagram of the programs used is presented in figure 2.17. After scanning the films by densitometry, the optical density bytes are transferred from tape onto disc using the program AV2DSC. The program PRESTO prepares a list of spots and their coordinates for measurement based on a knowledge of the space group, the cell dimensions of the crystal as well as the wavelength used. The scanned image of the film is displayed on an Evans & Sutherland PS 2 picture system and, with the aid of the graphics tablet and pen, the centre of gravity of a few strong spots are identified. The program then calculates spot positions and displays the spots against a reciprocal lattice grid derived from the calculations. The fit is considered to be good if the spots are within half a raster unit of the displayed grid.

The remaining steps in precession data processing are very similar to the steps outlined in section 2.8.2 for rotation photography data processing. The program MOSCO (an Oxford equivalent to MOSFLM) integrates intensities using an orientation matrix determined from PRESTO. The spots are profile-fitted (Rossmann, 1979) using a continually learnt profile algorithm (Diamond, 1969). The program DATRED performs Lorentz and polarisation corrections as well as intra-film scaling. The programs, SCALIN and DIFFER, were used to average equivalents, scale derivative and native data sets together and output the data as Patterson coefficients ready for input to the fast Fourier transform program FFT.

Class	Source	Space Group	Cell Dimensions (Å) a	b	c	β(°)	Subunits / Asym. Unit	Crystallization Conditions	Group	Reference
Mn	B.stearothermophilus	$P2_1$	50	70	69	111.2	2	$(NH_4)_2SO_4$, pH 8, dialysis	Harris MRC, Cambridge	Bridgen et al.(1976)
		$P2_12_12$	74.6	111.8	51.3		2	MPD, pH 6.3 or 7.8	Jansonius Univ. Basle Switzerland	Smit et al.(1977)
		$P2_12_12$	72.8	111.2	51.3		2	Zeppezauer		
		$P2_12_12_1$	78.0	112.4	50.5		2			
Mn	E.coli	P222	47	103	47.5		1	$(NH_4)_2SO_4$, pH 7.8, batch	Richardson Duke Univ. U.S.A.	Beem et al.(1976)
		$P222_1$	47	103	47.5		1			
		$C222_1$	101	108	180		4	PEG, pH 7.8, batch		
		$P2_12_12$	47	51	188		2			
Mn	Yeast mitochondrial	$P2_12_12$	63	115	125		4	MPD or PEG, pH 8.5, batch		
Mn	I.thermophilus	$P4_12_12$	147.5	147.5	55.9		4	$(NH_4)_2SO_4$, pH 5.7 or 7 hanging drop	Ludwig Univ. Michigan U.S.A.	Stallings et al.(1981)
Fe	E.coli	$P2_12_12_1$	81.8	75.2	71.3		2	$(NH_4)_2SO_4$ hanging drop, pH 7.5 vapour diffusion, pH 4.6		Powers et al.(1978)
Fe	R.ovalis	$P2_1$	81.9	49.0	61.9	106.8	2	$(NH_4)_2SO_4$, pH 4.8 vapour diffusion or batch	Petsko M.I.T. U.S.A.	Yamakura et al.(1976)

Table 3.1 Published crystal data for Fe and MnSODs.

CHAPTER 3

Crystallization and Search for Heavy Atom Derivatives

3.1 Background

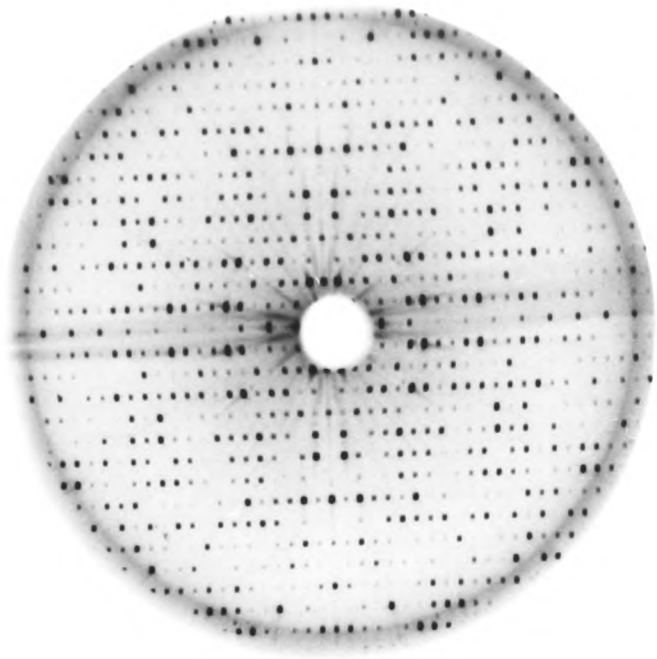
Crystallization and preliminary X-ray investigations have been reported for a number of Fe and MnSODs. Crystal data from these investigations have been compiled in table 3.1. It has been found that the enzyme can be crystallized readily under a variety of conditions.

The B.stearothermophilus MnSOD project was inherited from the Harris group (Cambridge) in 1980. Dr R. Cassels found the $P2_1$ crystal form of the enzyme unsuitable for further crystallographic work (pers. commun.). The crystals were very fragile and often twinned. This led Dr.'s P. Artymiuk and D. Rice to explore other crystal forms. By the start of the author's work, small crystals had been grown from polyethylene glycol (PEG) but had not been characterised.

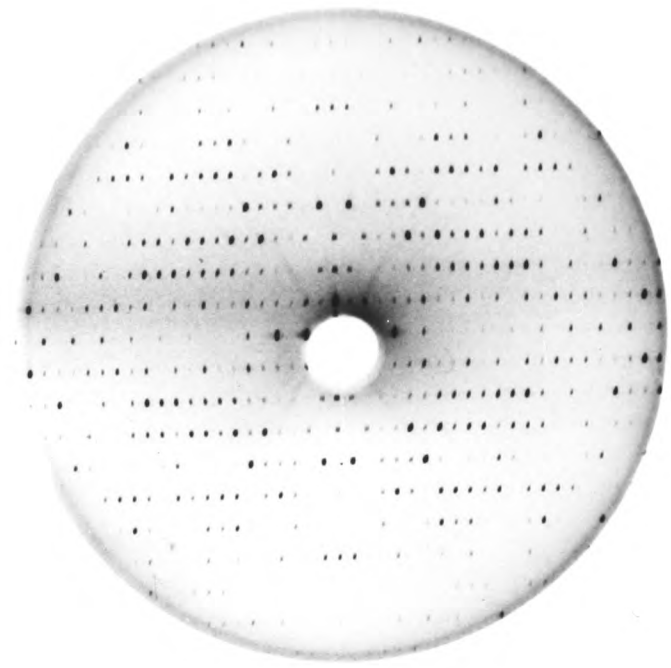
3.2 Isolation and Characterisation

A quantity of partially purified material was received as a kind gift from Dr. J.E. Walker (M.R.C. Lab. of Molecular Biology, Cambridge, U.K.). The method of isolation and purification of the enzyme from B.stearothermophilus cells has been described in the literature (Brock and Walker, 1980). The enzyme was stored in glycerol at $-20^{\circ}C$. The enzyme was further purified, as required, by chromatography on Sephadex G-75 (equilibrated and eluted with 50mM Tris-HCl, pH 7.8) in order to remove the glycerol and minor protein contaminants. Dr. Cassels had found that this step yielded sufficently pure protein as judged by gel electrophoresis. The final purification step was carried out by the kind efforts of Dr. J. Bannister.

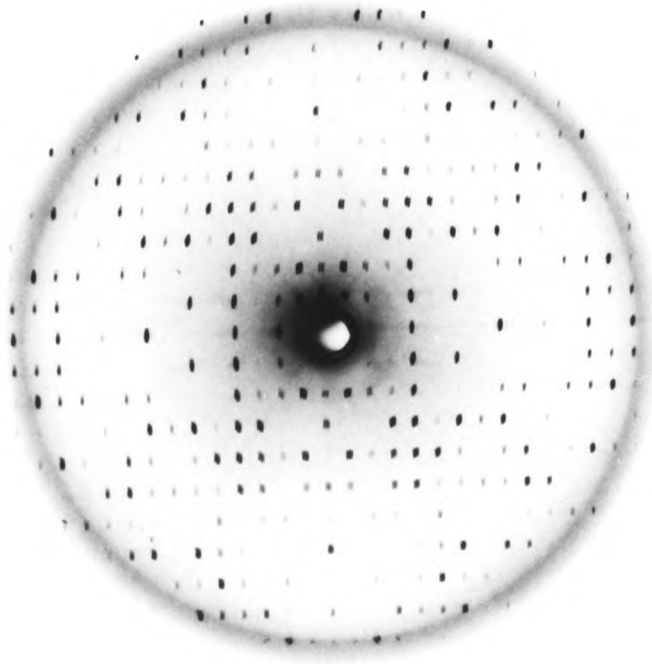
The concentration of the protein was estimated from the extinction



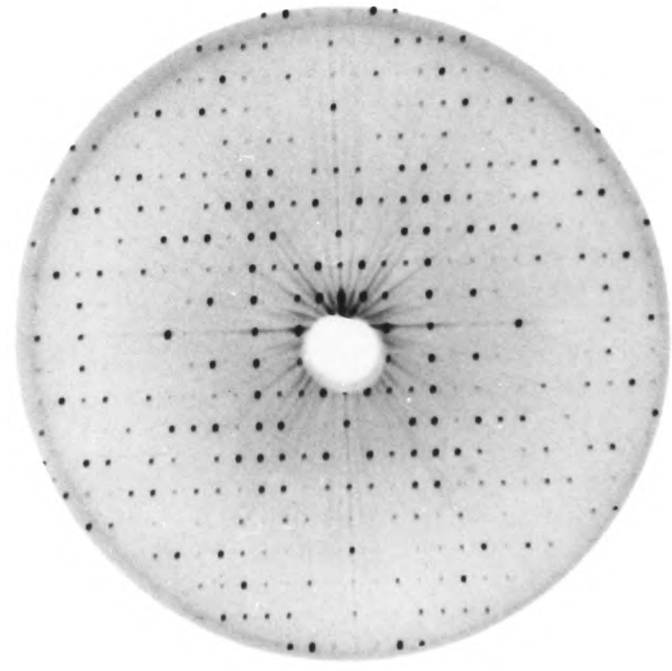
a



b



c



d

Figure 3.1 Precession photographs (μ=9°, crystal-to-film distance=100mm) of the three principal centric zones from MnSOD crystals, (a) hk0, (b) 0kl, (c) h0l with a=72.4Å and (d) h0l with a=74.2Å.

coefficient of the protein solution at 280nm (using a value of $A^{280}_{1\%}$ of 14.1 (Brock and Walker, 1980)). Dr. Cassels had measured the activity of the stock solution to be about 1000 units/mg by the method of McCord and Fridovich (1969a). This value compared favourably with that quoted in the literature (Brock and Walker, 1980).

3.3 Crystallization and Space Group Determination

The protein was crystallized readily by the batch method using approximate equivolume portions of SOD and PEG 6000 solutions as follows. The precipitant was made up of 10mg/ml PEG 6000 (50% w/v) in 50mM imidazole/HCl buffer, pH 7.5 - 8.0. Aliquots of PEG solution were added to the protein solutions (previously dialysed against 50mM imidazole/HCl buffer pH 7.5 - 8.0) in volume ratios of 0.9 to 1.2 that of the protein volume. The resultant cloudy suspension was vibrated to homogeneity and left to stand.

Dr. D. Rice had set up crystallization trials using protein concentrations ranging from 8 to 20 mg/ml and pH set at 6.0, 7.0 and 7.8. Under these conditions small block-shaped crystals (up to approximately 0.2mm x 0.2mm x 0.2mm in dimension) grew from the flocculent over a period of months. The author set up further crystallization trials using protein concentrations of 20, 30 and 40mg/ml at pH 7.8. The best crystals (up to approximately 2.0mm x 1.5mm x 1.0mm in linear dimensions) grew from protein concentrations of 40mg/ml and equivolume amount of precipitant in two to three weeks. The use of higher protein concentrations resulted in the rapid growth of large crystals but at the expense of a large amount of denatured protein and extreme fragility of the crystals.

The crystals grew as prisms with a wine-brown colouration (see frontispiece). The space group was determined from zonal precession photographs to be $P2_12_12$ (figure 3.1). There was no evidence of non-crystallographic symmetry from the molecular diad in the photographs.

Figure 3.2 Characteristic cracking pattern of MnSOD crystals.

Average unit cell dimensions measured from nine degree precession photos were $\underline{a}$=72.4Å, $\underline{b}$=111.3Å, and $\underline{c}$=51.3Å (but see section 3.6). The real axes pointed out of the crystal faces. When viewed down the $\underline{b}$ axis the crystals appeared colourless. If a needle axis was present in the crystal habit, it usually lay in the direction of the $\underline{c}$ axis. The crystals diffracted strongly to 2.0Å and appeared relatively insensitive to X-ray irradiation.

The unit cell dimensions corresponded to a volume of $4.13 \times 10^5 Å^3$ which, if the asymmetric unit was the dimer, gave a value for V_m (volume per unit mass) of $2.3 Å^3$/dalton. This value lies approximately in the centre of the normal range of values for proteins (Matthews, 1968). If the asymmetric unit contained either a monomer or two dimers, the corresponding values for V_m would have been well outside the range found normally for protein crystals (Matthews, 1968). The estimated solvent content of the crystals from the V_m value was 46%; a normal value found for protein crystals (Matthews, 1968).

3.4 Crystal Stability

The crystals were found to be very fragile. At least half the crystals grown in a pot had to be discarded due to twinning or subsequent fracturing on manipulation. A high proportion of crystals exhibited twinning in X-ray photos which was not always apparent from the crystal morphology. The crystals had a strong tendency to fracture and crack in planes parallel to the crystallographic $\underline{b}$-$\underline{c}$ plane as though composed of many layers of thin plates (figure 3.2). The nature of the crystal fragility was explored further because of these problems.

Experiments were performed to establish optimum conditions for the crystals with regard to temperature, pH and PEG concentration. The optimum conditions were ambient to 4°C, pH range 6 to 10 and a PEG concentration range equal to or slighty greater than the concentration in the mother liquor. A number of substitute mother liquors were tested because of

ASX	25
THR	11
SER	11
GLX	21
PRO	13
GLY	15
ALA	20
CYS	0
VAL	8
MET	2
ILE	9
LEU	19
TYR	8
PHE	8
HIS	9
LYS	12
ARG	6
TRP	6

Table 3.2 Amino acid composition of MnSOD from B.stearothermophilus (Brock & Walker, 1980).

crystal handling difficulties encountered with the highly viscous mother liquor. Crystals transferred to low concentrations of PEG solutions dissolved. Crystals survived for a short time only in high concentrations of ammonium sulphate solutions.

The fragile nature of the crystals appears to be a common feature of Mn and FeSOD crystals. Dr. R. Cassels had found the $P2_1$ form of MnSOD to be fragile and susceptible to twinning (pers. commun.). Petsko and coworkers have reported similar problems with FeSOD (Ringe et al.,1983).

3.5 Heavy Atom Derivative Search

A strategy was devised for the heavy derivative search on the basis of the amino acid composition of the protein given in table 3.2. The high content of carboxyl groups suggested 'hard' (electropositive and relatively unpolarisable) cations like uranyl might bind readily, although perhaps at many sites. The lack of sulphdryl groups appeared to limit the potential of mercurial derivatives. Nevertheless a number of other soft (readily polarisable and electronegative) ligands were available which might bind to mercurials and other 'soft' cations. A further class of reagent considered was those that might bind in hydrophobic crevices of the protein.

The crystals were mounted in 1.0mm diameter glass capillary tubes and soaked in heavy atom solution *in situ*. Heavy atom solutions were prepared by dissolving the appropriate reagent in artificial mother liquor (0.6g PEG/ml buffer) at concentrations between 1 and 10mM. A number of reagents could not be used because of insolubility or reaction with the mother liquor. These were $Hg_2(OAc)_2$, $UO_2(OAc)_2$, K_2PtCl_4, K_2PdCl_4, $Sm(OAc)_3$ and $Pt(NH_3)_2Cl_2$.

The results of the search are compiled in table 3.3. In general, crystals soaked in heavy atom reagents were very fragile and prone to cracking in *b*–*c* planes. Comparison of native and derivative diffraction patterns was complicated by cell dimension changes along the *a* axis of the

Crystal	Cell Dimensions (Å)		
	a	b	c
native	72.4	-	51.3
native	-	111.5	51.3
native	72.6	111.4	-
native	72.4	-	51.2
K_2HgI_4	-	111.1	51.3
$UO_2(NO_3)_2$	-	109.9	50.1
K_2HgI_4	74.9	-	51.3
native	72.4	-	51.2
native	74.2	-	51.4
native	72.2	-	51.1
native	-	111.2	51.2
K_2HgI_4	73.6	-	51.3
K_2HgI_4	-	111.7	51.2
$K_2Pt(CN)_4$	72.6	111.4	-
native	73.8	-	51.4
native	73.3	-	51.2
native	72.5	111.4	51.2
native	75.0	111.9	51.3
K_2HgI_4	74.0	-	51.0

Table 3.4 Cell dimension measurements of native and heavy atom derivative-soaked crystals measured from $\mu=9^\circ$ zonal precession photographs or diffractometer.

Derivative	Concentration (mM)	Soak Time	Cell Dimensions (Å) a	b	c	Observations
$K_3UO_2F_5$	10.1	3d				fresh solution, crystal disordered
$K_2Pt(CN)_4$	11.3	3.5d				crystal lost colour, dissolved on manipulation
	10.9	8d				crystal cracked
	5.0	8d				crystal disordered
	3.8	4d				crystal lost colour, disordered
	0.8	2d	72.6	111.4	-	many small intensity changes
$UO_2(NO_3)_2$	11.3	20hr				disordered
	5.6	42hr	-	109.9	50.1	large intensity changes
	3.7	2d				crystals too fragile for analysis
	1.1	42hr				native diffraction pattern
K_2HgI_4	10.5	3d 18hr	73.6	111.7	51.3	crystal crazed in planes, small intensity changes with native type 2
(xs KI req.)	8.0	3d 19hr	74.9	-	51.3	small intensity changes with native type 2
	5.0	4d 10hr	-	111.1	51.3	no change compared to native diffraction pattern
$AuNaCl_4$	9.9	3d 17hr				crystal split into planes
	2.0	49hr				disordered
$HgCl_2$	10.3	49hr	74	111		little change in 3^0 precession, split into planes on manipulation
(xs NaCl req.)						
Na_2PCMBS	10.0	-				crystals dissolve

Key: d=days, hr=hours.

Table 3.3 Conditions and observations in the heavy-atom derivative search.

derivative crystals. It was not until late into the search that a diffraction pattern was measured for a native crystal which had an a cell dimension of 74Å (see figure 3.1 (d)). The existence of a native 'type 2' pattern showed that the intensity changes observed from the diffraction of the K_2HgI_4 soaked crystals were, in large part, due to the cell dimension change that accompanied the soaks. At this stage the search was discontinued whilst ways of overcoming the cell dimension problem were investigated.

3.6 Cell Dimension Varibility

Cell dimension measurements of various native and heavy atom-soaked crystals have been compiled in chronological order of measurement in table 3.4. The native cell dimensions had appeared consistant before the heavy atom search had begun. The first sign of problems came from the K_2HgI_4 soaks. More recent measurements have shown that the a cell dimension for native crystals can have any value between 72 and 74Å.

The cell dimension variability appears to be related to the extent of crystal hydration (and the nature of the heavy atom, see chapter five). A data set measured by diffractometry (see chapter four) started with a crystal having a=74.5Å which decreased to 72.4Å overnight and was stable at this value for over a month. The diffracting power of the crystal was similar after the change, but a further change at the end of the data collection to a=67.0Å was accompanied by a limitation of diffraction to about 8Å. The total change in cell dimensions during data collection corresponded to a reduction in solvent content of 11%.

Calculations of the mean fractional isomorphous change (by the method of Crick and Magdoff, 1956) based on these cell dimension changes indicated a significant degree of non-isomorphism was introduced by the changes. The calculations indicated that intensity changes brought about by a cell dimension change of 2.1Å on the a axis were comparable to or greater than

the intensity changes that would occur on the addition of a single heavy atom to the protein, even at low resolution (6Å). The problem was further evidenced by the size of the intensity differences between the diffraction patterns of the two native types shown in figure 3.1. Small differences observed in 0kl precession photographs of the two native types suggested that differences between the two types were not solely due to cell dimension changes but might involve some small movements of the molecules in the unit cell. Cleary, the heavy atom approach of solving the phase problem was unsuitable unless some way could be found to reduce the size of the cell dimension changes. Such efforts are described in chapter five. Alternative methods of solving the phase problem were explored because of the difficulty in preparing isomorphous derivatives and these are discussed in the next chapter.

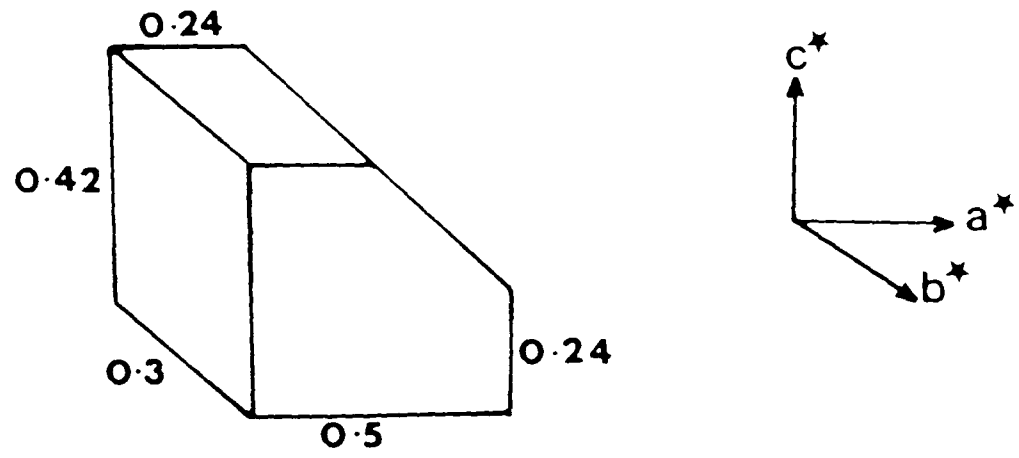


Figure 4.1 Morphology and dimensions (mm) of crystal used in high resolution data collection on diffractometer.

CHAPTER 4

The Phase Problem - Molecular Replacement

4.1 Introduction

The cell dimension variability discussed in the last chapter made the task of finding isomorphous heavy atom derivatives difficult. An alternative approach for solving the phase problem is discussed in this chapter. Midway though the heavy atom derivative search, discussed in the last chapter, a structure of FeSOD had been solved to medium resolution (Petsko and Ringe, pers. commun.). This opened the way to solving the MnSOD structure by molecular replacement procedures. The theory behind this approach has been discussed in chapter two. This chapter is started by discussing the collection and processing of a high resolution native data set. These data were used to solve the rotational relationship between the known FeSOD and unknown MnSOD crystal structures.

4.2 The Native Data Set

4.2.1 Data Collection

A decision was made to collect a high resolution data set by diffractometry rather than by photographic methods, chiefly, because of the former's inherent greater accuracy. Radiation damage and absorption corrections are calculated more readily by diffractometry and very low resolution reflections can be measured without the saturation problems encountered with film. Reference reflections are more readily monitored by diffractometry so that absorption changes, crystal slippage and radiation damage can be monitored thoughout data collection.

Over twenty crystals were scanned on the precession camera before one was judged suitable for high resolution data collection. The morphology and

Resolution Shell (Å)	Mount	Main/Blind Region	Counter Mode	Count Time (sec/step)	Scan Width (steps)	Peak Width (steps)	Delta Omega Max. (°)	Number of Reflections	Number of Intensities >3sd (%)
									-
6.0-10.0	a*	M	5-c	4.5	36	10	0.32	-	86
6.0-10.0	a*	M	5-c	4.5	36	10	0.32	1610	39
2.4-3.0	a*	M	5-c	9.0	24	10	0.11	7720	54
2.85 3.5	a*	M	5-c	6.0	26	10	0.12	3870	27
2.4 3.0	b*	B	5-c	11.0	22	10	0.04	520	47
2.85 3.5	b*	B	5-c	7.5	22	10	0.05	375	75
3.5 7.0	b*	B	5-c	5.0	26	10	0.10	565	83
6.0 10.0	b*	B	5-c	5.0	28	10	0.14	130	62
3.0 7.0	b*	B c:c	5-c	5.0	26	10	0.10	500	86
9.0-40.3	a*	M	1-c	2.0	18	10	-	592	79
3.3-7.0	a*	M	5-c	5.0	34	10	0.24	5690	37
2.4-3.0	a*	M rad	5-c	9.0	24	10	0.11	250	81
3.3 7.0	a*	M c:c	5-c	5.0	30	10	0.24	500	65
2.4-10.0	a*	M scale	5-c	9.0	36	10	0.32	525	

Key: c:c=counter-to-counter scales, rad=radiation damage block, scale=data block collected across shells of data.

Notes: 1) The step size used was 0.03°.

2) Radiation damage blocks of about 50 quintuplets were measured between each shell of data.

Table 4.1 Data collection parameters used in measuring the native data set on the five circle diffractometer. The table has been written in chronological order of data collection.

dimensions of the crystal used in data collection are shown in figure 4.1. X-ray photographs of larger crystals invariably showed split spots and occasional disorder.

The crystal was aligned and diffractometer set up as follows. The crystal was aligned accurately to within five minutes of arc by small angle precession photography and then attached to the goniometer spindle of the diffractometer. A 1mm diameter input collimator was used so as to ensure the crystal would be completely bathed in the X-ray beam at all orientations. A helium tube of length, 18.0 cm, was inserted between crystal and counter box and filled with helium to a pressure of 3 pounds per square inch above atmosphere. The tapered end of the tube was fitted with an aperture (5 mm internal diameter) to cut down air scatter of the beam (see section 2.7.2 for a full description). The initial crystal-to-counter distance was calculated to be 46.0 cm (for $\underline{a}^*$ mount) and then finely adjusted by centring a strong row of equivalent reflections (centred about the position of the axial reflection 0 13 0) in the counter box. The appropriate size of aperture for the counter box windows was determined from measuring the very strong 3 6 0 reflection with different aperture settings and allowing for crystal slippage by ensuring that a movement of $\chi = 0.15^\circ$ did not result in a reduction of the intensity. Fine adjustments of the goniometer arcs, crystal centring and instrument datum positions were made using the strong mounting axis reflection 8 0 0 (at a Bragg angle of 4.90°). The initial orientation matrix and cell constants were calculated from the position of the following strong axial reflections:

a^*	18 0 0	resolution = 4.0Å
b^*	0 36 0	resolution = 3.1Å
c^*	0 0 13	resolution = 3.9Å

The first step was to collect a low resolution data set (to 6Å) and to monitor the behaviour of the crystal in the X-ray beam before plans were made to extend the resolution of the data set. The calculated $\Delta\omega$ value for

Cell Axis	Before Shrinkage	After Shrinkage	Percentage Change	Final Values	Total Percentage Change
a	74.72	72.23	3.3	67.05	10.3
b	111.87	111.08	0.7	107.35	4.0
c	51.29	51.08	0.4	48.50	5.4

Table 4.2 Cell dimension changes (Å) measured on diffractometer during collection of the native data set.

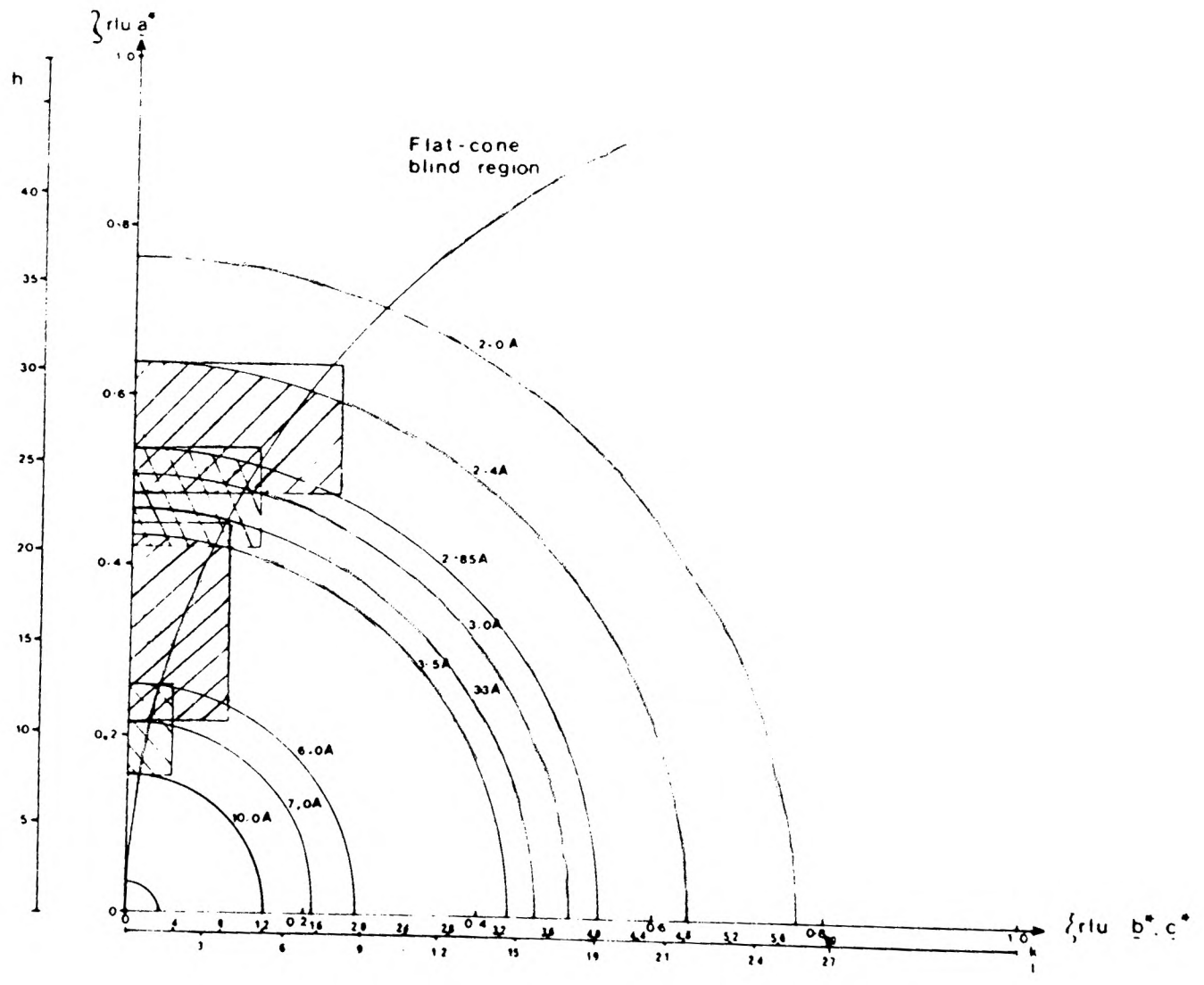


Figure 4.2 Diagram illustrating areas of reciprocal space swept out during data collection. The area of reciprocal space accessible to the diffractometer in flat-cone setting (with an a* crystal mount) is below the designated curve. The shaded boxes represent the blind region of reciprocal space which was collected with a b* mount.

a resolution shell of 6-10Å with an $\underline{a}^*$ mount was small enough to warrant the data being collected in the optimum setting with five counters. The measurement parameters used in the data collection are presented in table 4.1. Friedel pairs were measured in the hope of being able to determine the positions of the manganese atoms in the unit cell by means of an anomalous difference Patterson synthesis. The Friedel pairs were measured such that alternate rows were transversed in positive and negative parts of reciprocal space as follows:

0 0 L, 0 0 -L, 0 -1 -L, 0 1 L, 0 2 L, 0 -2 L, etc.....

were the hkl sequence was measured in zig-zag fashion so that each row or level was started as near as possible to the end of the previous one. The diffractometer was programmed to drive to the instrument zero datum positions after every hundred measurements and reference reflections measured as a check on crystal slippage, absorption changes and radiation damage. The reference reflections were chosen on the basis of being strong or axial; they were:

3 6 0 ; 8 0 0 Φ=0° ; 8 0 0 Φ=90°

During overnight automatic data collection, the crystal's cell dimensions changed, with the $\underline{a}$ cell dimension showing the greatest change of over 2Å (see table 4.2). The crystal still diffracted well, so the data collection was restarted from the beginning. A high resolution data collection was planned because the crystal was showing no signs of radiation damage. The crystal diffracted strongly to 2.4Å and weakly to 2.0Å. The collection strategy decided upon is outlined in figure 4.2 and table 4.1. Overlapping shells of roughly equivalent number of reflections were to be measured starting from the high resolution end and working inwards. The high resolution shells were collected first because data in these shells are more sensitive to radiation damage than lower resolution shells. The measuring parameters were determined from the measurement of strong reflections in each resolution shell. North, Phillips and Mathews

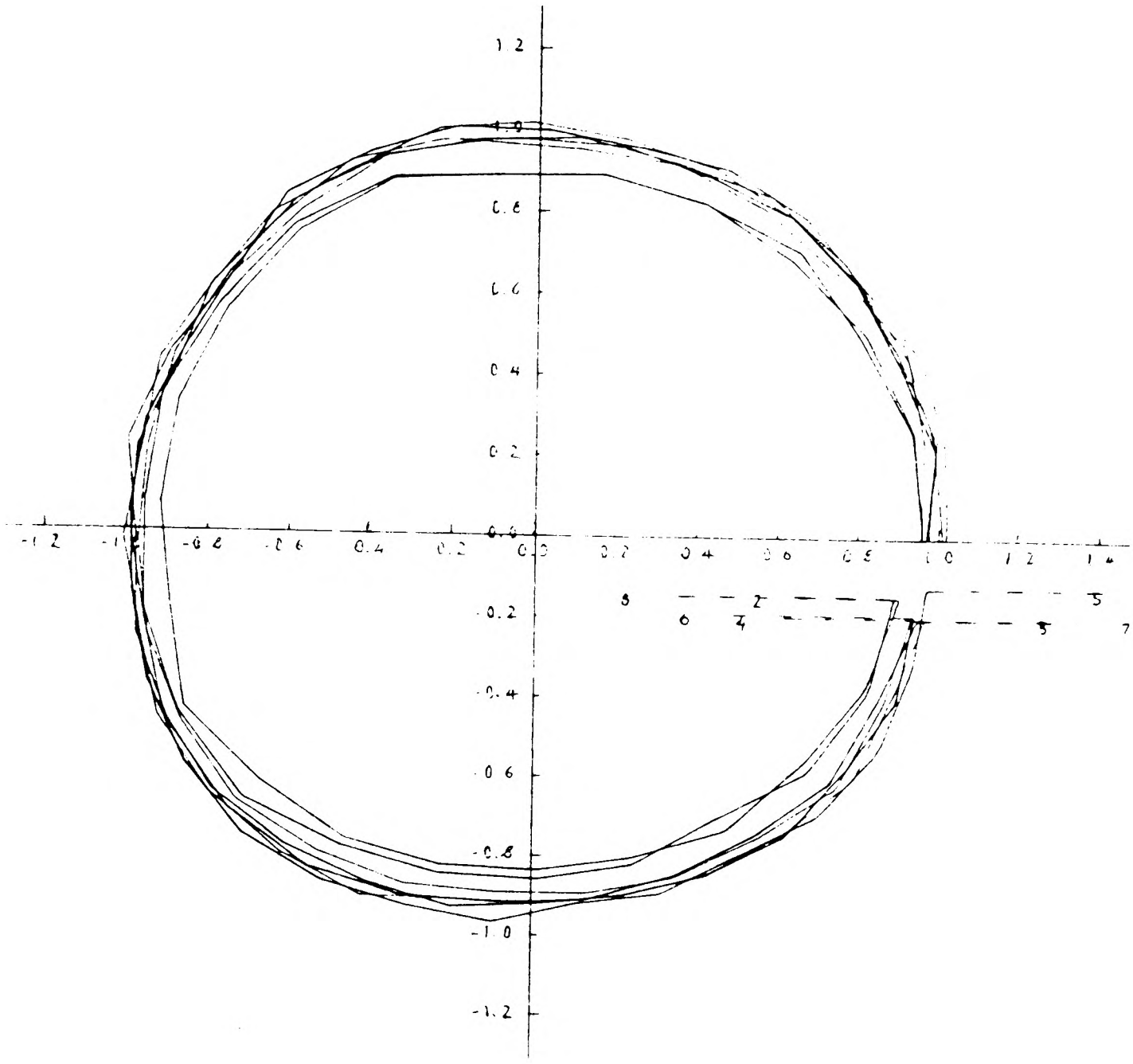
(i) North-Phillips-Mathews Absorption Curves

Reflection Family	Count Time (secs)	Scan Width (steps)	Peak Width (steps)	Step Size (^{0})	Number of Equivalents
4 0 0	3.0	22	12	0.03	2
18 0 0	3.0	22	12	0.03	2
8 0 0	0.5	22	12	0.03	2
14 0 0	7.0	22	12	0.03	2
22 0 0	17.0	22	12	0.03	2
26 0 0	20.0	22	12	0.03	2

(ii) Huber-Kopfmann Curves

Reflection Family	Count Time (secs)	Scan Width (steps)	Peak Width (steps)	Step Size (^{0})	Number of Equivalents
7 8 1	4.0	22	12	0.03	8
7 7 2	3.5	22	12	0.03	8
8 7 3	5.0	22	12	0.03	8
6 4 3	2.5	22	12	0.03	8
6 2 3	2.7	22	12	0.03	8
6 0 3	3.5	22	12	0.03	4
9 5 0	4.0	22	12	0.03	4
7 4 2	5.2	22	12	0.03	8
4 0 0	6.0	22	12	0.03	2
9 8 0	8.0	22	12	0.03	4

Table 4.3 Data collection parameters used to measure absorption curves for the native crystal on the diffractometer.



Angle between g.p.u = 12.6^{0}

Interpolation distance = 0.7 g.p.u.

Extrapolation distance = 0.7 g.p.u.

Reflections rejected if within 0.2 g.p.u. of critical level

Critical level defined as the fifth level above the equator

Normalised to 7 7 2 family of reflections

Total number of observations = 2592

Number of extrapolated points = 488

Number of interpolated points = 2104

Number of rejections = 0

Total number of grid points = 219

Absorption range = 0.95 - 1.19

Figure 4.3 **Transmission surface calculated at various levels of reciprocal space for the native crystal based on measurements of reflections by the Huber-Kopfmann method.**

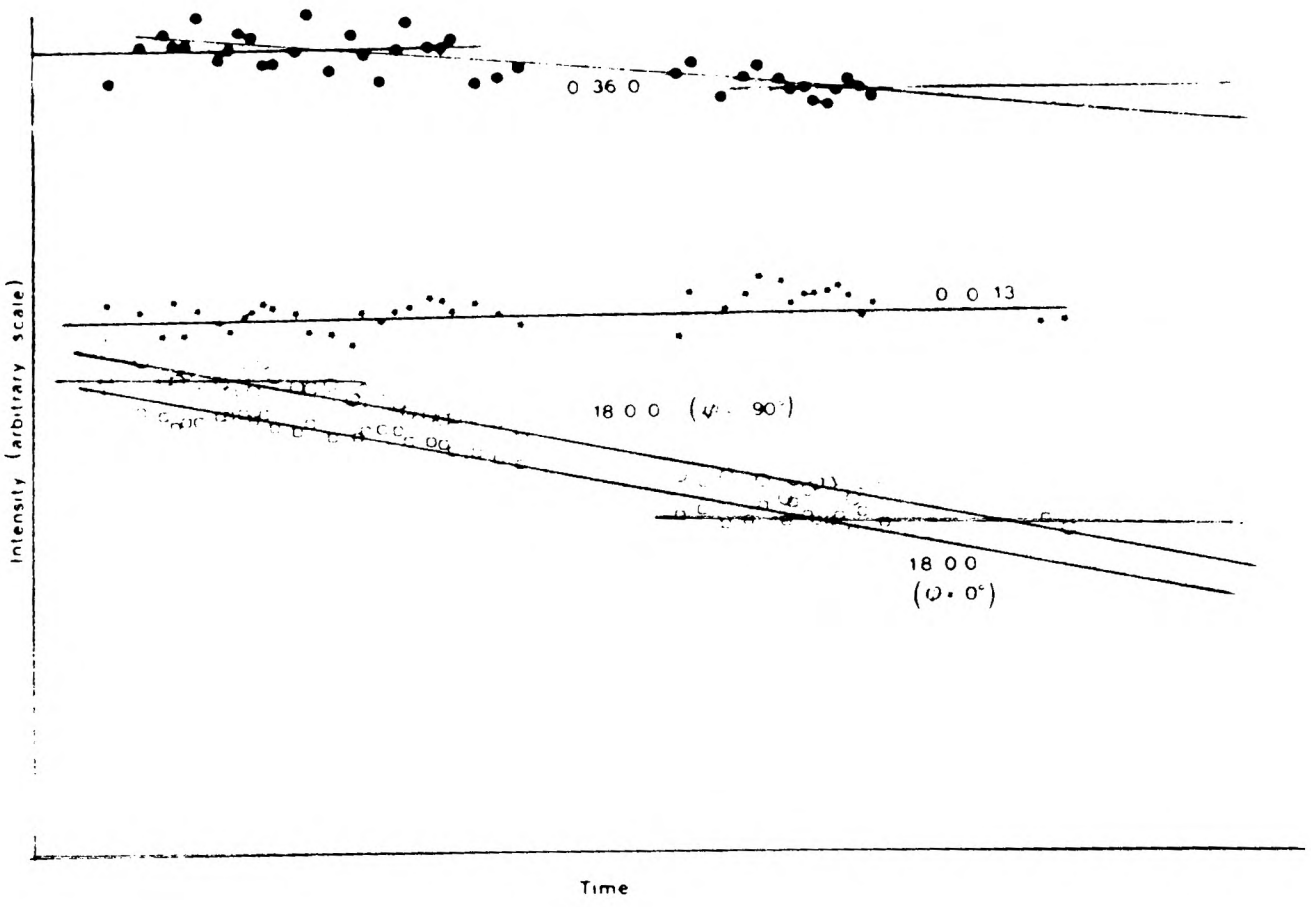


Figure 4.4 Intensity changes of reference reflections during data collection.

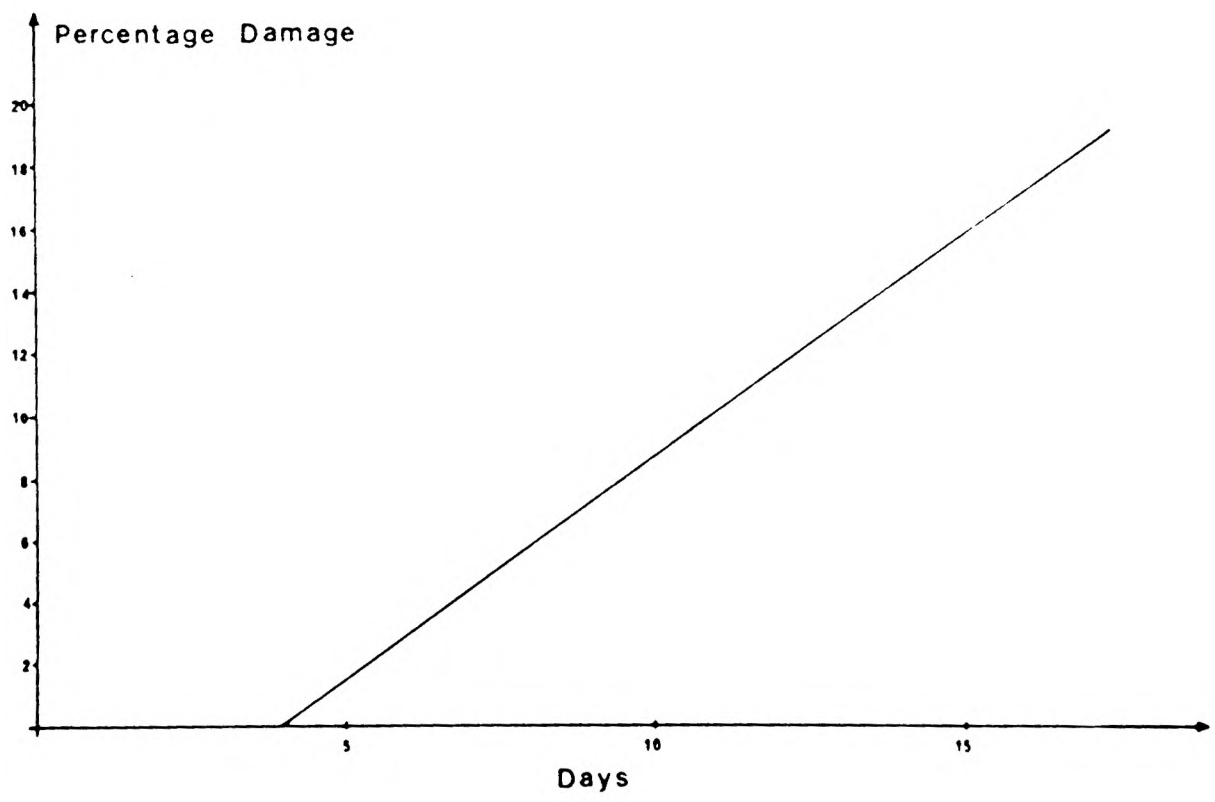


Figure 4.5 Radiation damage curve calculated from the repeated measurement of blocks of data during collection of the native data set.

(1968) absorption curves were measured before and after data collection so as to keep a record of the absorption behaviour of the crystal in case absorption changes occured during data collection. The following reference reflections were measured after every hundred measurements so as to monitor absorption changes, crystal slippage and radiation damage:

$$0\ 36\ 0,\ 0\ 0\ 13,\ 18\ 0\ 0\ \Phi=0^{\circ},\ 18\ 0\ 0\ \Phi=90^{\circ}$$

Reference blocks of fifty or more quintuplets were measured at the beginning and end of each shell of data for the purpose of using the data for the estimation of radiation damage corrections. A $\underline{b}^*$ mount was chosen to cover the blind region because of the smaller $\Delta\omega$ value required compared to the value for a $\underline{c}^*$ mount. This meant that the blind region could be collected in five counter mode with the helium tube left in place. Finally, Huber-Kopfmann absorption curves were measured and the details of these are presented in table 4.3.

The data collection progressed smoothly with only small adjustments of the goniometer arcs required to correct for crystal slippage. No absorption changes occured during collection. A plot of the reference intensities with time showed a slight fall off consistent with an overall radiation damage for the crystal of between fifteen and twenty five per cent (figure 4.4). After measurement of the final absorption curves, the crystal suffered a further dramatic change of cell dimensions accompanied by the loss of mother liquor in the capillary tube (see table 4.2).

4.2.2 Data Processing

The methods used in reducing the data to an unique set of structure factors has been described in chapter two. In this section brief comments will be made on the various steps employed in the data reduction.

The data were corrected for absorption using the Huber-Kopfmann data collected after the native data set had been measured. The transmission surface and parameters used to calculate it have been compiled in figure

Number of reflections processed = 21535

Overall radiation damage = 19%

Number of reflections rejected on merging = 101

Number of negative intensities = 920

Merging R-factor = 6.85%

Rfriedel = 3.5%

Maximum resolution = 2.3Å (5.4Å for anomalous data)

Minimum resolution = 38.9Å

Number of unique reflections = 16562

Key: Merging R-factor= $\sum_{hkl} \sum_{i=1}^{N} |I_i - \bar{I}| \; / \; \sum_{hkl} N\bar{I}$

where I is the value of the ith of N measurements of an intensity and $\bar{I}$ is the mean of all N measurements of that intensity.

Rfriedel = $\sum |F_{PH}^{+} - F_{PH}^{-}| \; / \; \sum \bar{F}_{PH}$

Note : A measurement was rejected on merging if,

(a) it was the smaller of a pair of estimates of a reflection, differing by more than four standard deviations from one another, or

(b) it was one of at least three estimates of a reflection and differed by more than three standard deviations from the mean intensity for that reflection.

Table 4.4 Summary of processing statistics for the native data set.

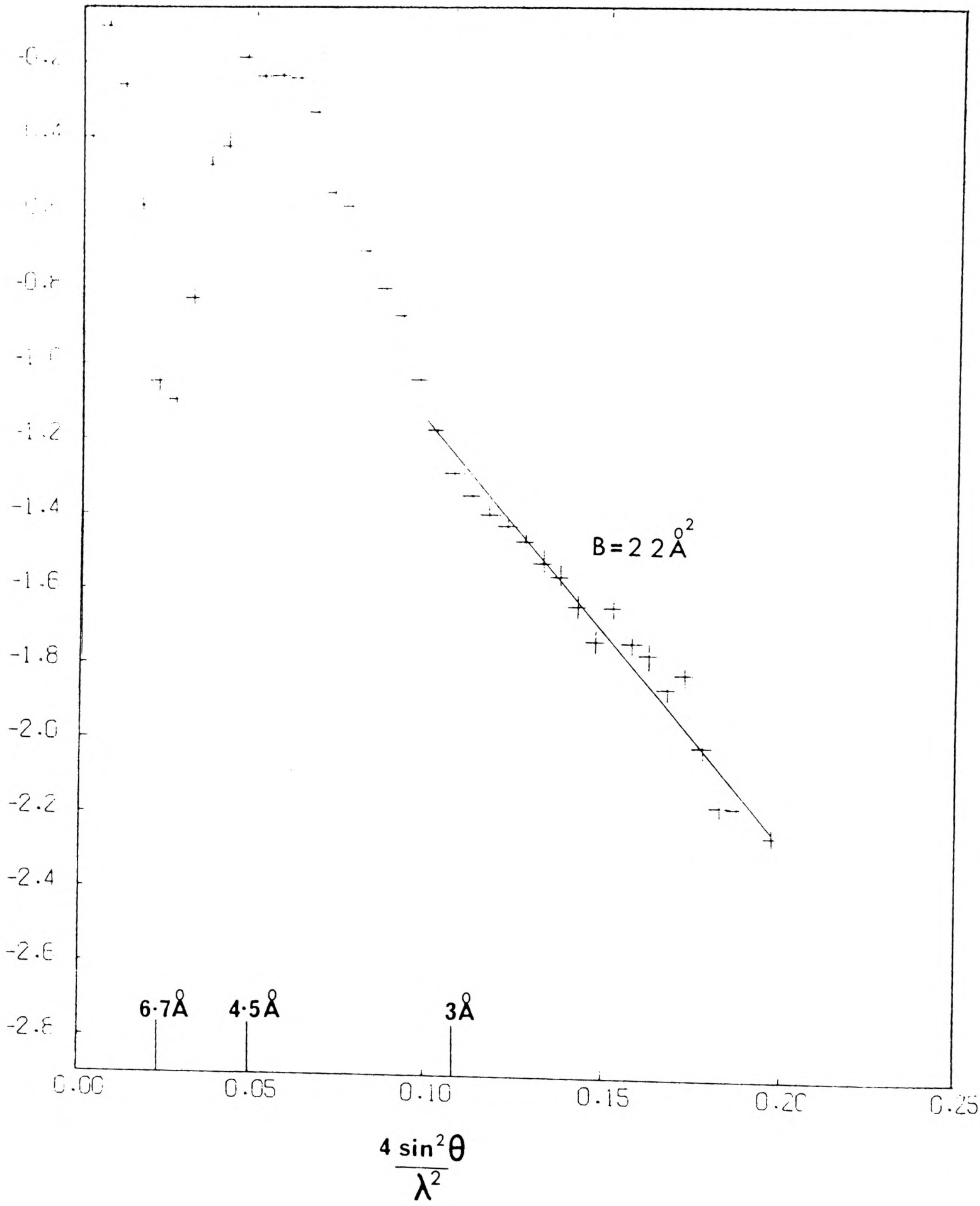


Figure 4.6 A Wilson plot for MnSOD. The value of the temperature factor B, determined from the best straight line for points between 3.2 and 2.3Å, was 22Å^2. The minimum and maximum in the low angle region occur at spacings corresponding to 6.7Å 4.5Å respectively.

4.3. The near uniform shape of the surface meant that the absorption of X-rays by the crystal was almost independent of the direction of the scattering vector.

The behaviour of the reference reflections during data collection suggested significant radiation damage of the crystal had taken place (figure 4.4). Analysis of blocks of reflections measured before and after measurement of data shells indicated that the damage could be approximated with a linear curve (figure 4.5). The estimated overall radiation damage was 19%.

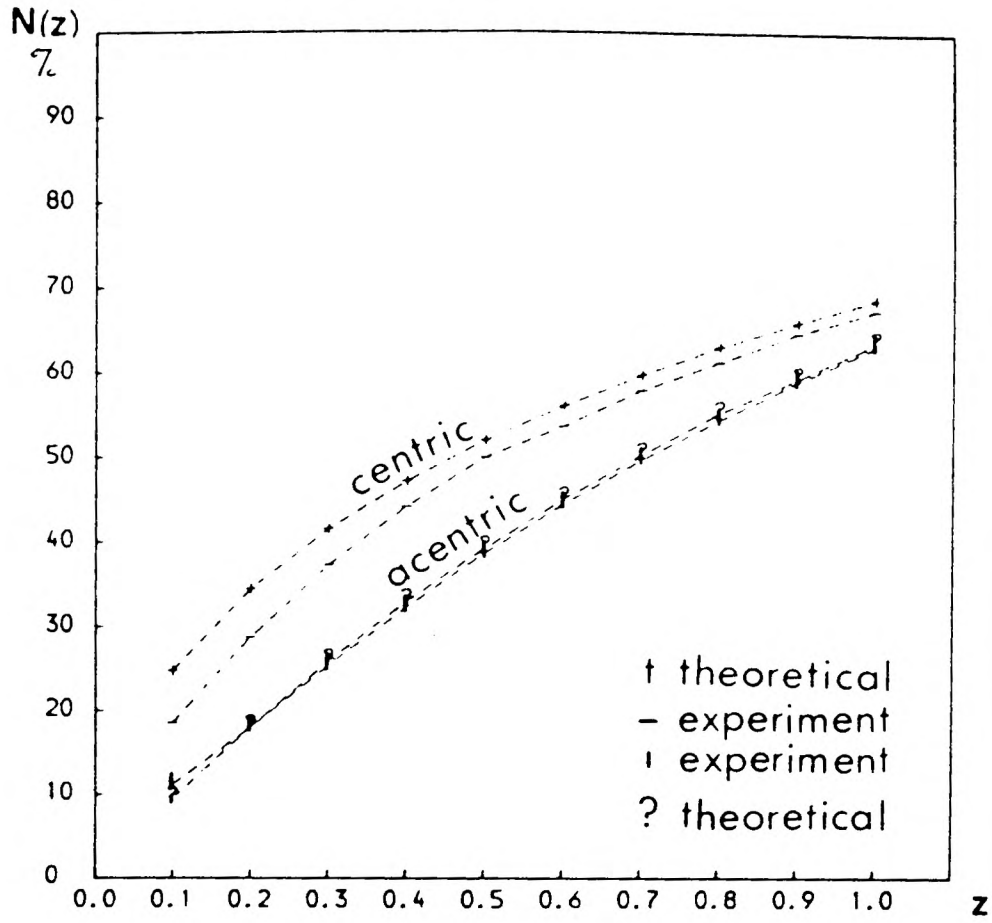
The data shells were corrected for Lp factors, absorption and radiation damage before being scaled and merged together. A summary of the processing statistics for the native data is given in table 4.4. Low R-factors for the merging of overlapping shells of data indicated the data were measured well. The R-factor between Friedel equivalents was very low indicating the measured anomalous contribution of the manganese was small. Patterson maps calculated with coefficients $(F_{PH}^{+}-F_{PH}^{-})^2$, were noisy and uninterpretable.

The approximate absolute scale of the data was calculated from a Wilson plot (figure 4.6). The best straight line was calculated from the higher angle data. The lower angle data suffered significant deviation from linearity due to the breakdown of the assumptions underlying Wilson's equation.

4.2.3 Data Analysis

The statistical distribution of the intensities of X-ray reflections in reciprocal space due to a centrosymmetric structure is characteristically different from that of a non-centrosymmetric structure (Wilson, 1949). This distinction applies to reflections within a crystallographic zone which may or may not have a centre of symmetry. The distribution of intensities in a centric zone is characterised by having a higher proportion of weak or absent reflections compared to acentric zones of data. In figure 4.7, the

(a)



(b)

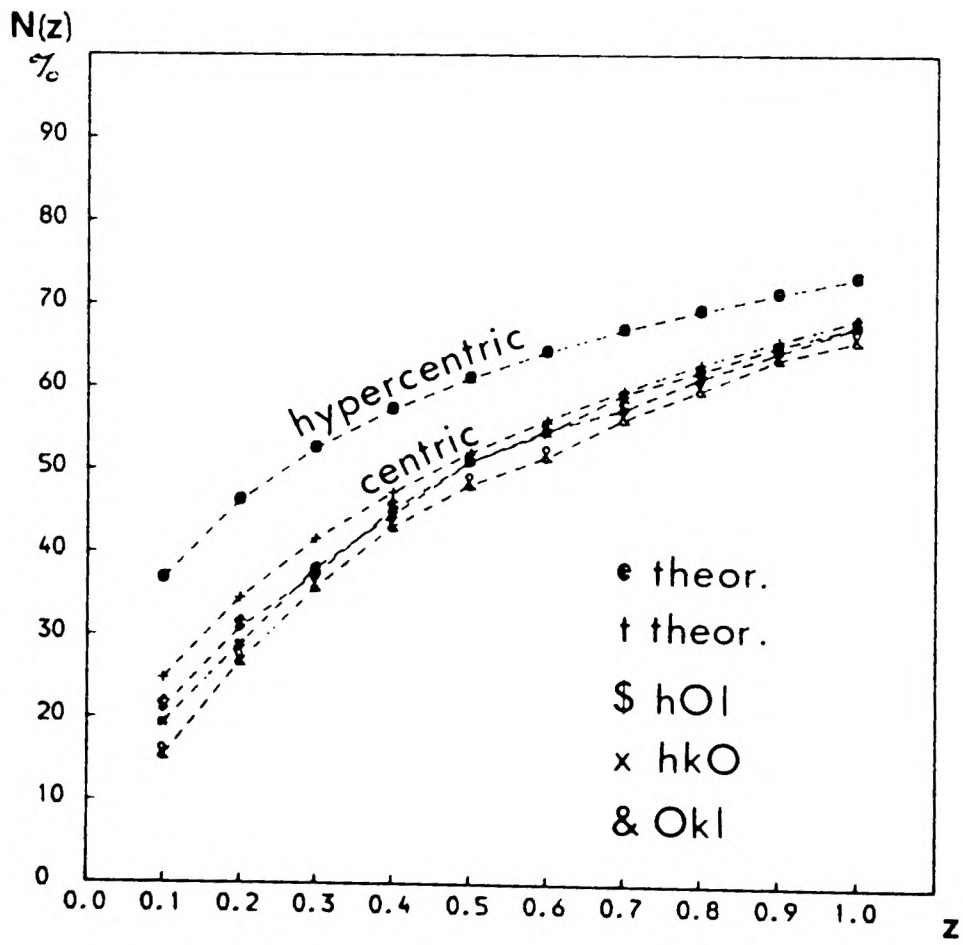


Figure 4.7 The observed intensity distributions within the principal zones of the native data; (a) for centric and acentric distributions (b) and for zonal centric distributions.

Notes: The number of reflections contributing to each zone of data is as follows:

hOl	498	hkO	979	Okl	772	centric	2201.

fraction of all reflections, N(z), which have intensities less than or equal to z times the average intensity is plotted as a function of z for the principal zones of native data.

In figure 4.7a, the acentric zonal data is shown to follow the expected distribution but the centric distribution is different. A closer look at the centric data (figure 4.7b) reveals a significant deviation from the expected trend at low z and the 0kl zonal distribution lies well under the expected distribution for all z. Deviations might result from errors, random or systematic, in the observed intensities (Rogers et al., 1955) and this may account for the observed h0l and hk0 deviations. But it appears that the deviation of the 0kl distribution from the expected trend is a real effect because the trend was also apparent in all the heavy atom derivative data sets collected (discussed in chapter five). The reason for the deviation was unclear.

4.3 The Rotation Function

4.3.1 Introduction

The molecular replacement work carried out on MnSOD is discussed in the following sections. The theory behind the approach has been discussed in chapter two (and so will not be repeated here).

At the time of processing the native data, it became known to us that a three-dimensional crystal structure of FeSOD had been solved to a resolution of 2.9Å by the method of multiple isomorphous replacement (G. Petsko and D. Ringe, pers. commun.). There was sufficient evidence available to suggest that Fe and MnSODs were structural homologues (see chapter one). This news provided a welcome alternative method of trying to solve the phase problem.

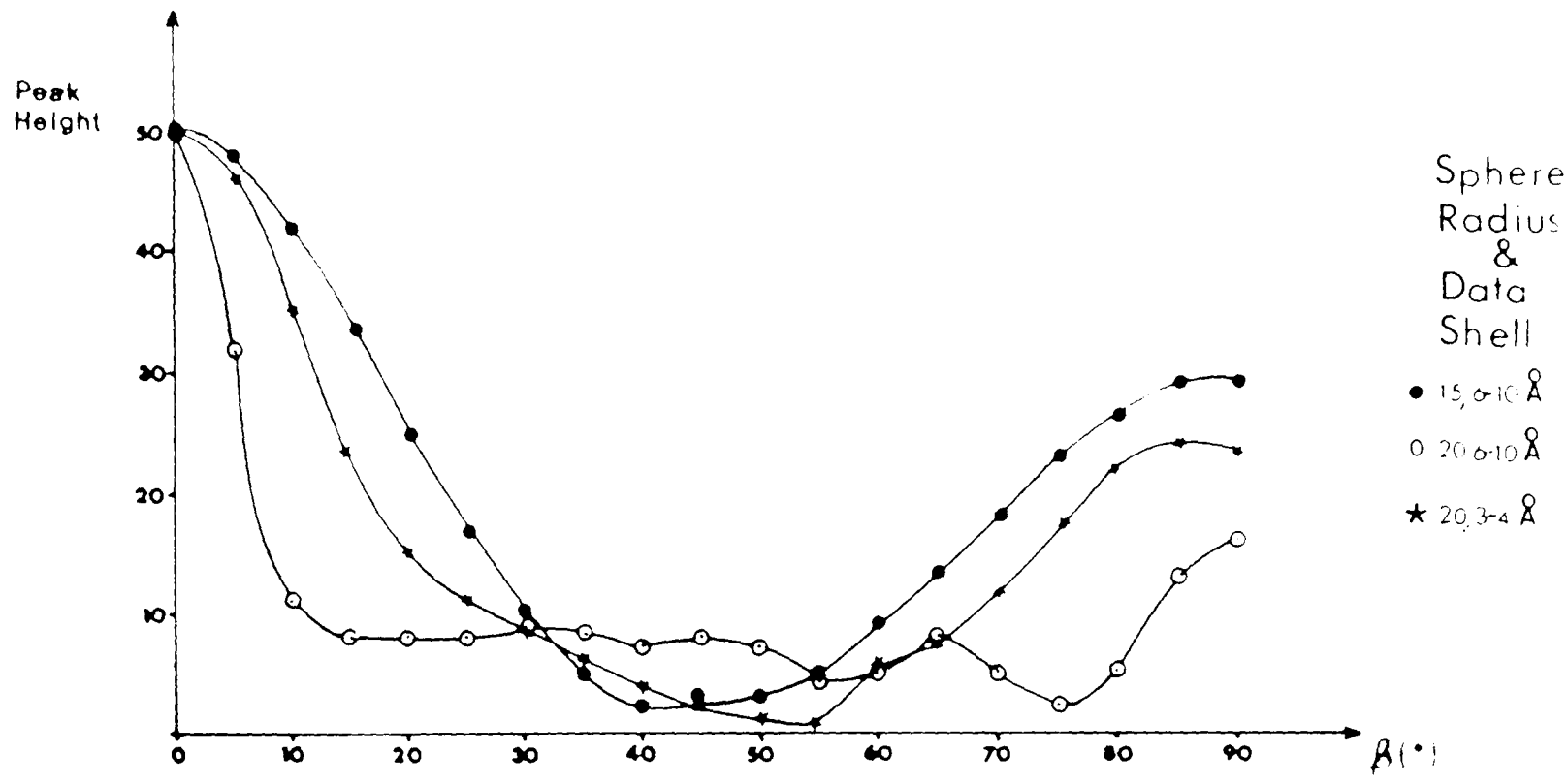


Figure 4.8 A line section through the self-rotation function R ($\alpha = 55^{\circ}$, β, $\gamma = 125^{\circ}$) calculated for FeSOD. Peak height of origin set to 50.

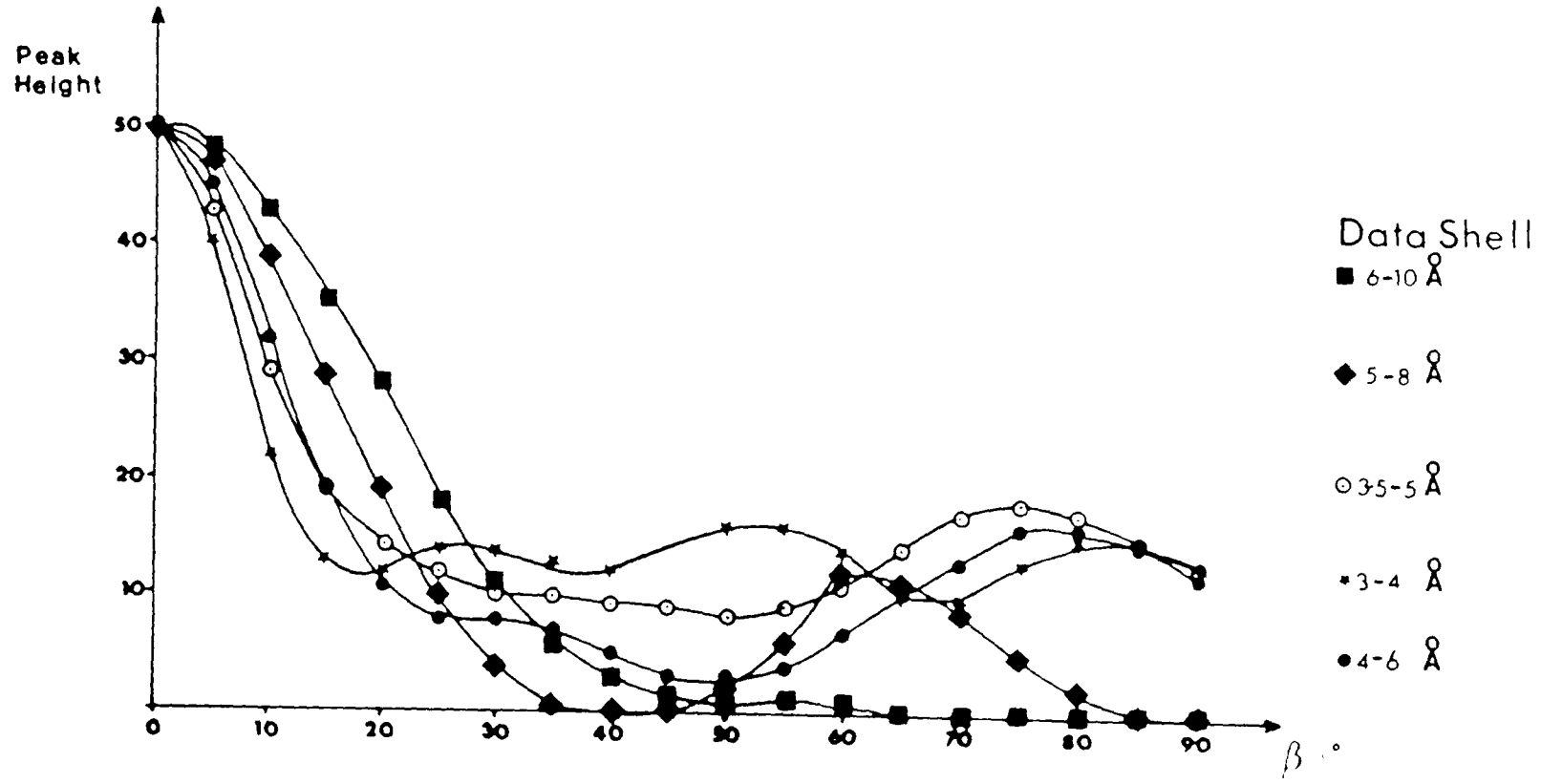


Figure 4.9 A line section through the self-rotation function R ($\alpha = 110^{\circ}$, β, $\gamma = 70^{\circ}$) calculated for MnSOD . Sphere size radius set to 15Å. Height of origin peak set to 50.

4.3.2 The Model

Professor G.A. Petsko and Dr. D. Ringe kindly sent monomer coordinates of a polyalanine model which had been built into the electron density map of the Pseudomonas ovalis FeSOD. The model was unrefined and poorly idealized. The appropriate transformation matrix to generate the dimer coordinates was included together with native structure factors to 2.9Å resolution. A number of workers have had success with the molecular replacement procedure when using polyalanine models (for example, Lattman et al., 1971; Navia et al., 1979; Vyas et al., 1983). In fact, working with polyalanine models might be advantageous in circumstances where there is low sequence homology between the unknown and model structures as may be the case here.

4.3.3 Self-Rotation Results

The native structure factors for both SODs were analysed for the presence of molecular diads using the fast rotation program FROTF (see chapter two). A search of the rotation surface for peaks with $\chi = 180^{\circ}$ represented a search for molecular diads. Peaks were sharpened by weighting down (using a B factor) or omitting inner reflections of the reciprocal lattice, both of which have the effect of reducing the dominance of the larger terms in the Patterson calculations. The search sphere was given an inner cutoff of 3.5Å radius so that short range vectors originating from solvent structure and other sources, which might contribute to the noise level, were eliminated. The optimal value for an outer cutoff radius of the sphere was estimated from the dimensions of the FeSOD monomer (52 X 40 X 32Å^3) to be between 15 and 20Å. The maps were calculated in five degree intervals on the Eulerian angles α, β and γ. The calculated range of angles explored depended on the rotational space group (Rao et al., 1980).

The self-rotation function yielded an unambiguous result for the FeSOD data. A single peak, with a height of between two and three fifths of the origin peak, was found on all maps calculated (figure 4.8). The peak height

was optimised with a sphere size of radius, 15Å. The peak broadened out if the sphere radius was set below this value. The orientation of the molecular diad calculated from the rotation function agreed with the known orientation from the M.I.R. results.

The results for MnSOD were less encouraging. Many maps were calulated (using a 15Å radius sphere) but no consistent peaks were found. The two most likely candidates were,

$$\alpha = 15^{\circ}, \beta = 40^{\circ}, \gamma = 15^{\circ}$$

which occurred in maps calculated with lower resolution terms (6-10 and 5-8Å resolution shells) and,

$$\alpha = 110^{\circ}, \beta = 80^{\circ}, \gamma = 70^{\circ}$$

which occurred in maps calculated with higher resolution terms (3.5-5, 3-4 and 5-8Å resolution shells, see figure 4.9). The peak-to-background ratio for each was always below two fifths of the height of the origin peak.

The results for FeSOD were impressive. The unconvincing results for the MnSOD data might have been due to tighter crystal packing which would introduce more noise into the maps due to the presence of a significant number of intermolecular vectors in the calculations.

4.3.4 Cross-Rotation Results

As a first step in determining the rotational relationship between the FeSOD and MnSOD, a cross-rotation function was performed on the native structure factors. All maps (calculated with data from 10-3, 10-6, 6-4 and 4-3Å resolution shells) yielded two self-consistent peaks of equal height at,

$$\alpha = 65^{\circ}, \beta = 55^{\circ}, \gamma = 35^{\circ} \text{ and}$$

$$\alpha = 160^{\circ}, \beta = 90^{\circ}, \gamma = 65^{\circ}.$$

These two peaks may have represented the two possible ways the FeSOD dimer could be rotated into alignment with the MnSOD molecule. These encouraging results meant the cross-rotation function might work despite the

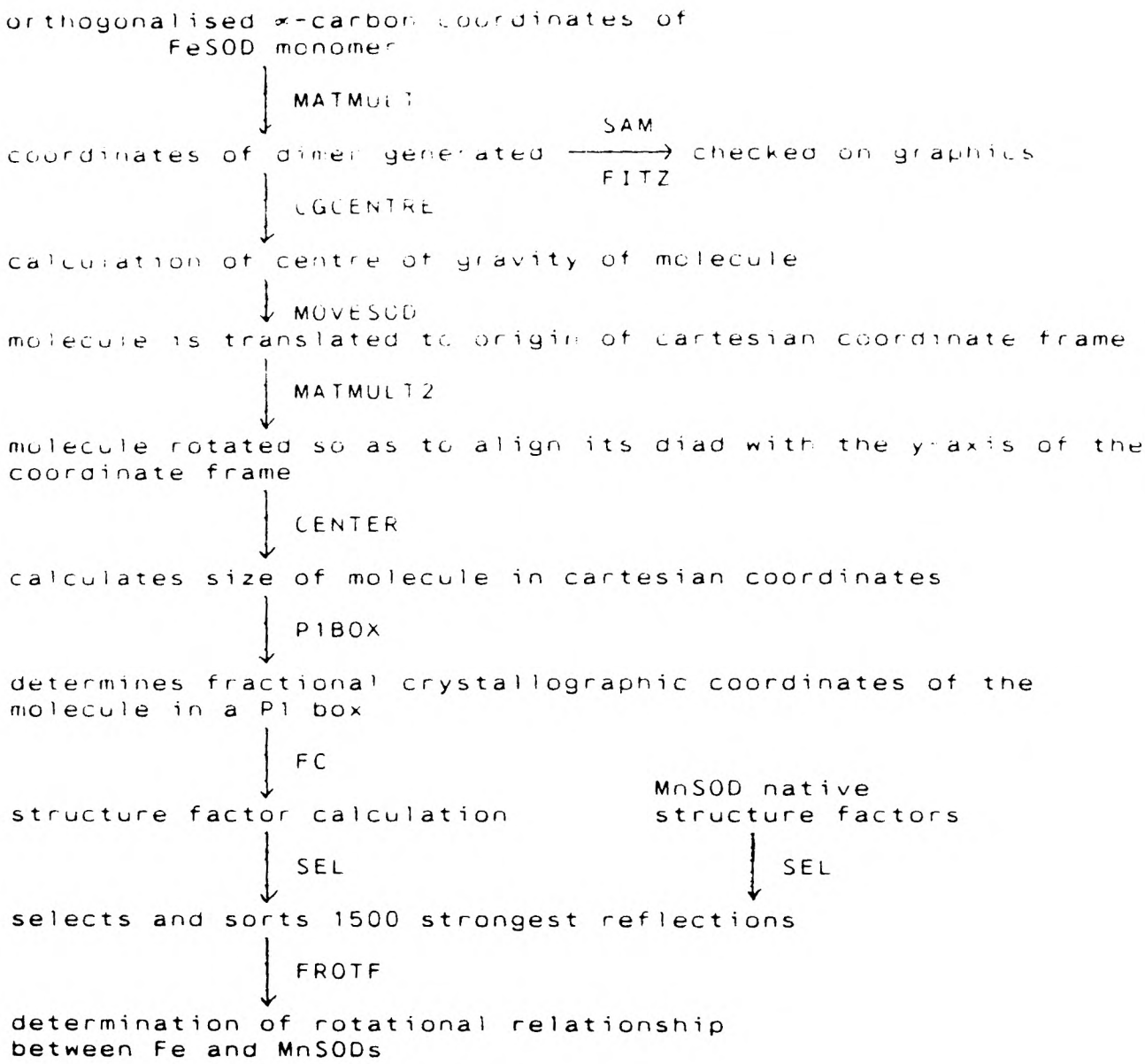


Figure 4.10 Flow diagram of computer programs used in the cross-rotation work.

disappointments of the MnSOD self-rotation results. Nevertheless, the maps were very noisy and the peaks were only just above the noise level. To improve the results, structure factors of the model were recalculated for the FeSOD dimer in a large P1 box. The reasons for this were to reduce the influence of intermolecular vectors of the model and thus reduce the noise level in the maps. The choice of using the dimer as the model rather than the monomer was based on the idea that the MnSOD monomers might be structurally asymmetric with respect to one another because of crystal contacts so that a dimer model might average out such effects. Also the dimer contains extra structural information over the monomer. One disadvantage of using the dimer model is that the spherical search probe used in the fast rotation program was not the optimal shape for the elliptically shaped dimer model. The use of a sphere of integration meant having to compromise between increasing the number of intramolecular vectors and decreasing the number of intermolecular vectors and in the case of the dimer shape, the noise level introduced into the maps must have been significant (as suggested by the self-rotation results).

The methods for preparing the model structure factors for the cross-rotation function are outlined in the flow diagram of figure 4.10. The molecular diad of the model was aligned with the y-axis of the cartesian coordinate frame (using the formulation of Hendrickson, 1979) so as to make interpretations of the final results easier. The box size (edge of 120Å) of the P1 unit cell was chosen to be one and a half times the largest molecular dimension of the FeSOD dimer to ensure the effects of intermolecular vectors were eliminated. Structure factors were calculated to 3.5Å using, in total, 1872 atoms including two iron atoms.

The superposition of the model Patterson onto the mmm Patterson symmetry of the unknown structure gave a rotation space group of $P2_1ab$ (Rao et al., 1980). An asymmetric unit of this space group, defined by the limits,

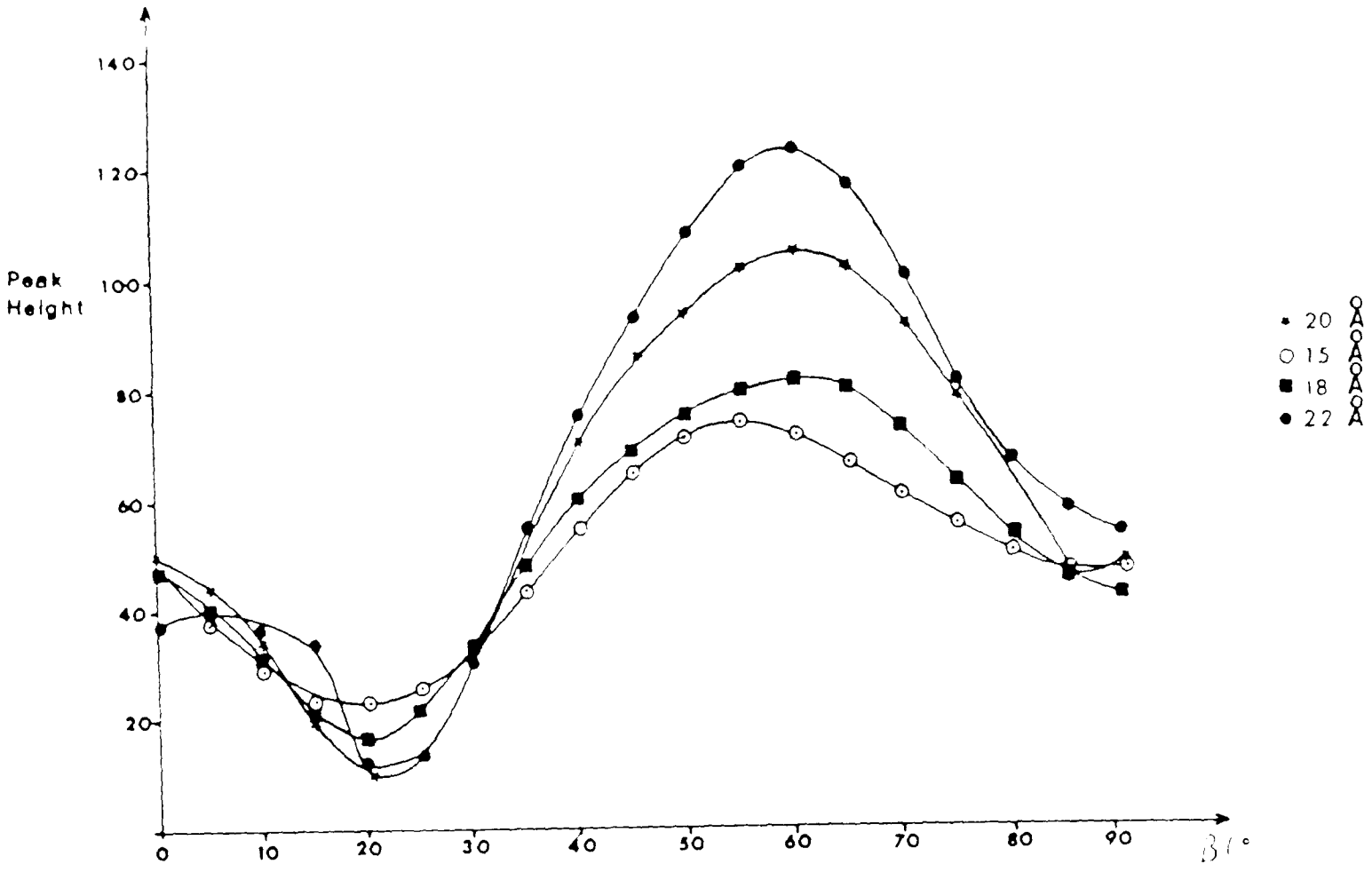


Figure 4.11 A line section through the cross-rotation function R ($\alpha = 5^{\circ}$, β, $\gamma = 170^{\circ}$) calculated at different radii of sphere. Peak height on an arbitrary scale.

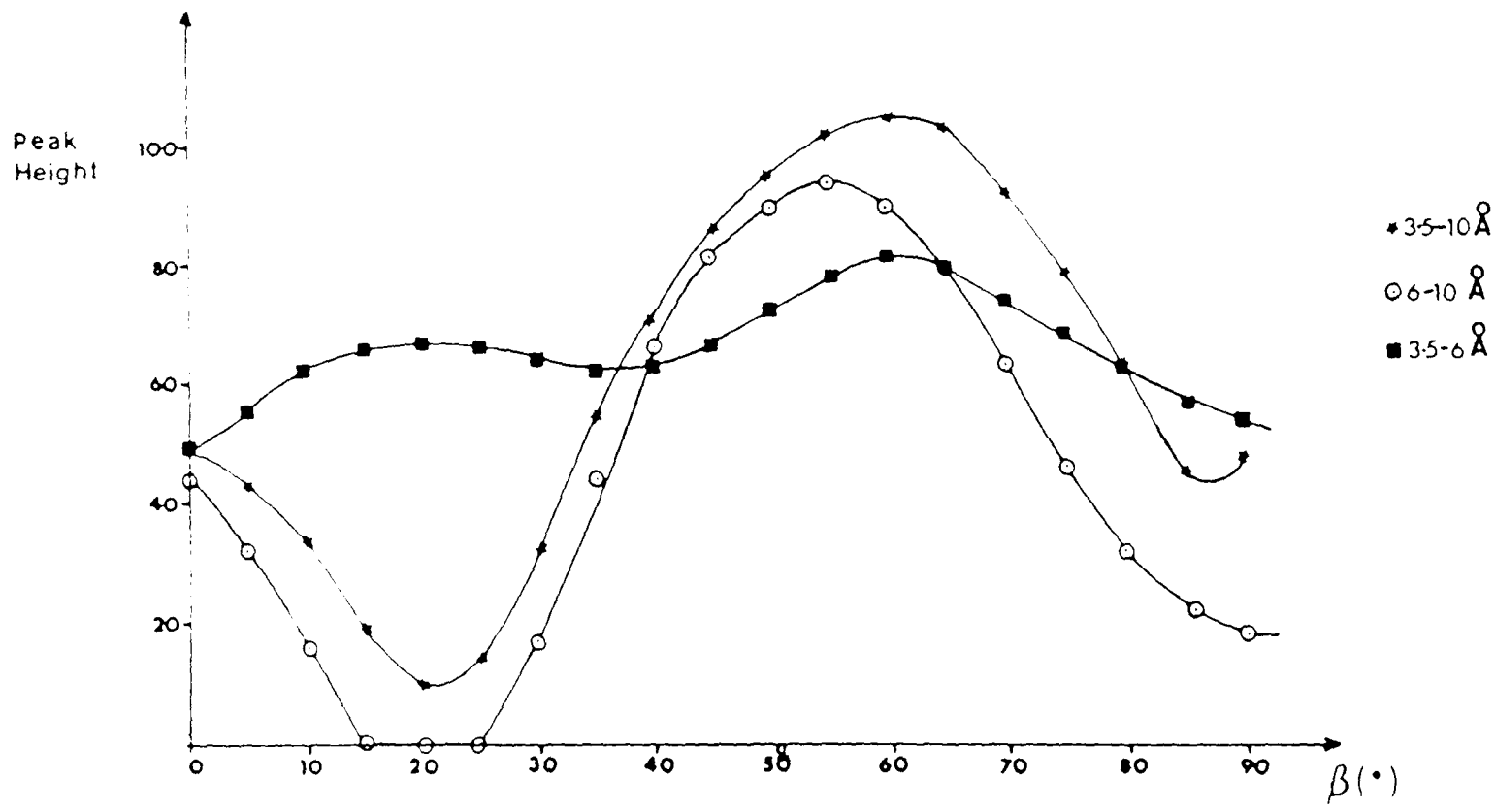


Figure 4.12 A line section through the cross-rotation function R ($\alpha = 5^{\circ}$, β, $\gamma = 170^{\circ}$) calculated for different resolution shells of data. Radius of sphere set to 20Å. Peak height set to an arbitrary scale.

$$0 \leq \alpha \leq \pi \ , \ 0 \leq \beta \leq \pi/2 \ , \ 0 \leq \gamma \leq 2\pi \ ,$$

was explored at five degree intervals on α, β and γ. The Patterson of the model was rotated onto the MnSOD Patterson using the computer program FROTF. All maps calculated (see figures 4.11 and 4.12) had two peaks consistently above the noise level and were the two highest features on all the maps. The best results were obtained with a sphere size of 20Å radius. The plots in figure 4.11 are deceptive because ,although the line section indicates a 22Å sphere gave the best peak-to-background ratio, the peak-to background ratio in the vicinity of the β=60° section was better for a 20Å radius sphere than for a 22Å radius sphere. It is interesting to note that the optimum sphere size for the ovelap of dimers was of the same magnitude for the self-rotation of the monomers. This sphere search size represented the best compromise between the inclusion of intramolecular and the exclusion of intermolecular vectors. This result suggests again that the MnSOD dimers pack closely in the crystal.

The two solutions were,

$$\alpha = 5^\circ \ , \ \beta = 60^\circ \ , \ \gamma = 170^\circ$$

$$\text{and} \quad \alpha = 175^\circ \ , \ \beta = 60^\circ \ , \ \gamma = 190^\circ$$

The error in each angle was estimated to be about 4° from the solutions obtained for different resolution ranges. The two solutions represented the two ways the model could be rotated into the unknown structure. This can be shown as follows. If R(x,y,z) is the rotation function required to rotate the model into alignment with the unknown structure, then there exists another function R(-x,y,-z) which will produce the same result (remembering the molecular diad of the model has been aligned with the y-axis of the cartesian frame of the P1 unit cell). It can be shown that, if α, β and γ are the Eulerian angles that solve for R(x,y,z), then by definition (from the equation in figure 2.8), $\pi+\alpha$, $\pi-\beta$ and $\pi-\gamma$ will solve for R(-x,y,-z). Further, the latter angles are related by symmetry to $\pi-\alpha$, β and $-\gamma$ (Tollin et al., 1966) so the two solutions above do represent the two ways that the

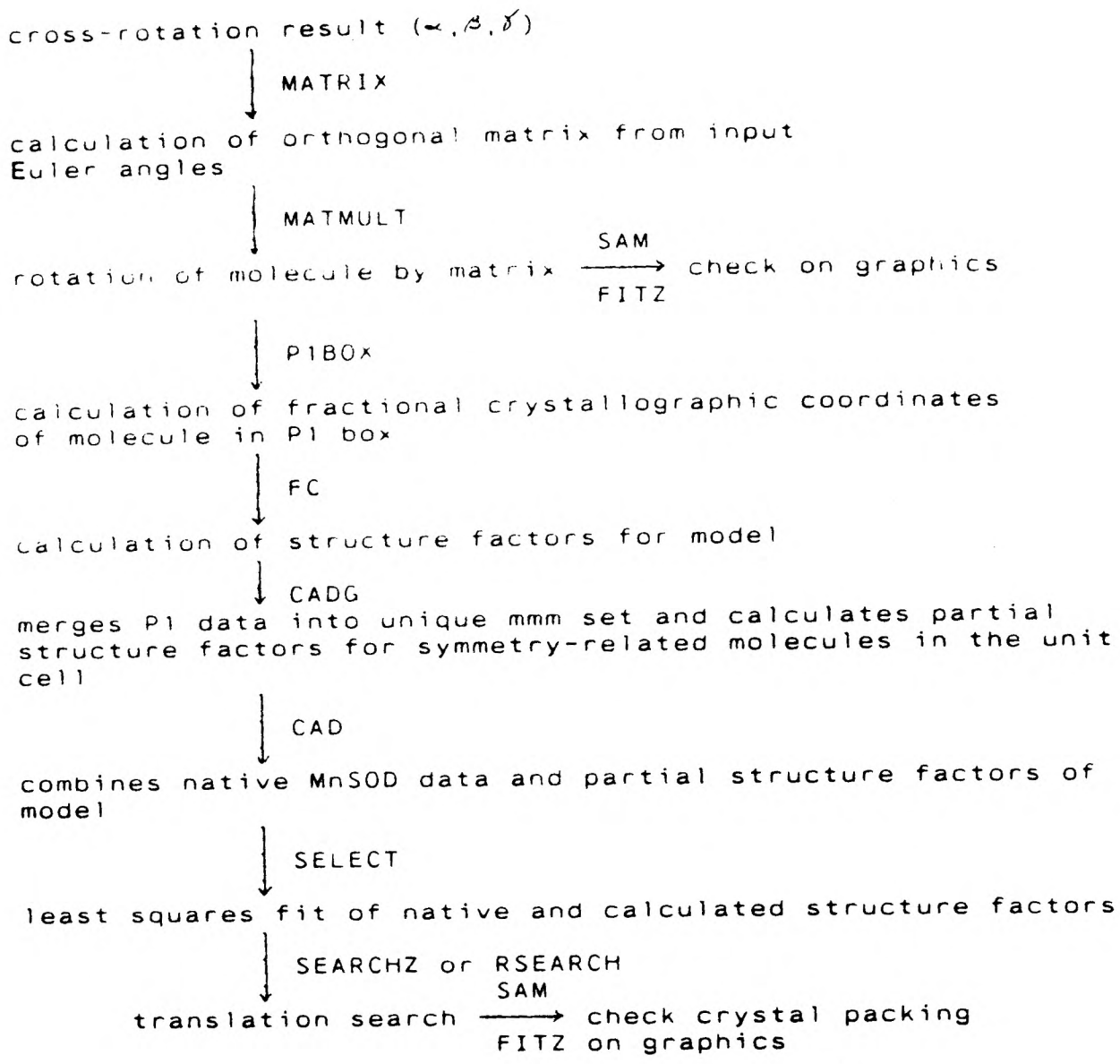


Figure 4.13 Flow diagram of computer programs used in the translation function work.

model can be brought into rotational alignment with the unknown MnSOD structure.

The Eulerian angles were converted to spherical polar angles so the result could be more easily visualised. The direction of the molecular diads was calculated using the program SEL (see appendix three; the convention defining the spherical polar angles has been given in chapter two). The direction of the molecular diad for the model was

$$\psi = 90^{\circ} \ , \ \Phi = 90^{\circ} \text{ and } \chi = 180^{\circ}$$

as expected, and the direction of the molecular diad for MnSOD was

$$\psi = 81^{\circ} \ , \ \Phi = -90^{\circ} \text{ and } \chi = 180^{\circ}.$$

This meant that the predicted molecular diad for MnSOD lay roughly parallel to the y-axis. If this result is correct, then the expected maximum for the molecular diad in the MnSOD self-rotation maps would have occurred under the origin peak.

4.4 The Translation Function

The methodology used for the translation search is outlined in the flow diagram in figure 4.13. The FeSOD model was rotated according to the average of the angles indicated by the rotation function work and centred about the origin of a cell with the same dimensions as the $P2_12_12$ cell of MnSOD. Structure factors for the model in the unit cell were calculated to a resolution of 3.5Å with an overall temperature factor of 10Å^2.

The translation search programs, RSEARCH and SEARCHZ, have been described in chapter two (and appendix three). A number of workers have found the best results from translation functions occurred when working with data between 7 and 3.5Å resolution (Dijkstra et al., 1982; Dijkstra et al., 1983; Vyas et al., 1983). In this work, the resolution range 7 to 3.5Å was explored. Complications from the exclusion of a model for the solvent structure are diminished by cutting out low resolution terms. A limit to medium resolution was imposed on the data because of the quality

Protein	Np	Na	$\frac{\langle \lvert \Delta F \rvert \rangle}{\langle Fp \rangle}$ (%)	$\frac{\langle Fa \rangle}{\langle Fp \rangle}$ (%)
-------	--	--	----	----
Crambin	400	6 S	1.4	29
Haemerythrin	1000	2 Fe	3.0	17
Hg myoglobin	1300	1 Hg	4.5	31
U lysozyme	1100	1 U	8.5	39
FeSOD	3000	1 Fe	1.3	7
MnSOD	3100	1 Mn	1.1	7
		4 S	0.5	9

Key: Np is the approximate number of non-hydrogen atoms and Na the number of anomalous scatterers in the protein. The third and fourth columns in the table are defined as follows:

$$\langle |\Delta F| \rangle / \langle F_p \rangle = \sqrt{2}\ (Na^{1/2} fa'')/ (Np^{1/2}\ Z_{eff})$$

and

$$\langle Fa \rangle / \langle Fp \rangle = (Na^{1/2} Za') / (Np^{1/2} Z_{eff})$$

where Z_{eff} is the effective atomic number taken as 6.7 and Za' is Za+fa'. Fp and Fa are structure factor magnitudes of the real contributions from the complete molecule and from the anomalous scatterers only, respectively. These equations come from the work of Crick & Magdoff (1956).

Figure 4.14 Expected diffraction ratios for various anomalous scatterers in proteins. The values for the first four proteins are taken from Hendrickson & Teeter (1981). All values have been calculated for zero scattering angle and CuKα wavelength.

of the model. Initial searches were conducted on a coarse grid of 2Å but this value was varied in the course of the work.

The results of the translation searches were disappointing. It was possible to find many solutions with conventional R-factors as low as 50% (from an average background value of around 60%) after the rejection of weak terms. All solutions were checked on the graphics and, without exception, all produced unfavourable packing arrangements of the molecules in the unit cell. Small rotations of the model from the mean solution of the rotation function did not lead to any convincing results with the translation function. Other workers have found that deviations of 2 or 3° in each of the Eulerian angles from the correct values can be tolerated in finding the correct solution from translation search functions (Harada et al., 1981; Dijkstra et al., 1983).

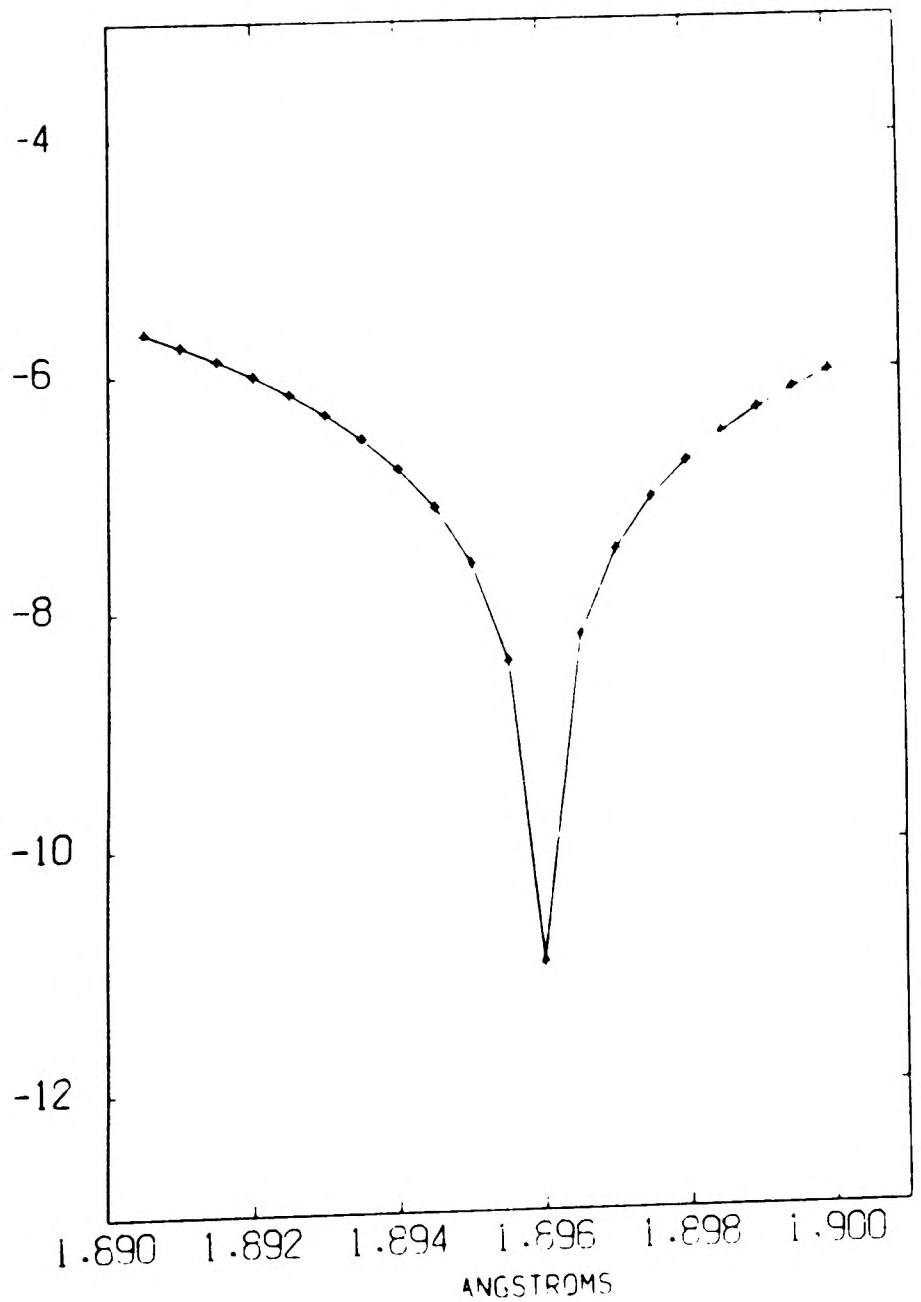
The successes and failures of the molecular replacement procedures are discussed at the end of this chapter. Attempts to solve the translation function by other techniques are discussed in the following sections.

4.5 Anomalous Scattering Experiments

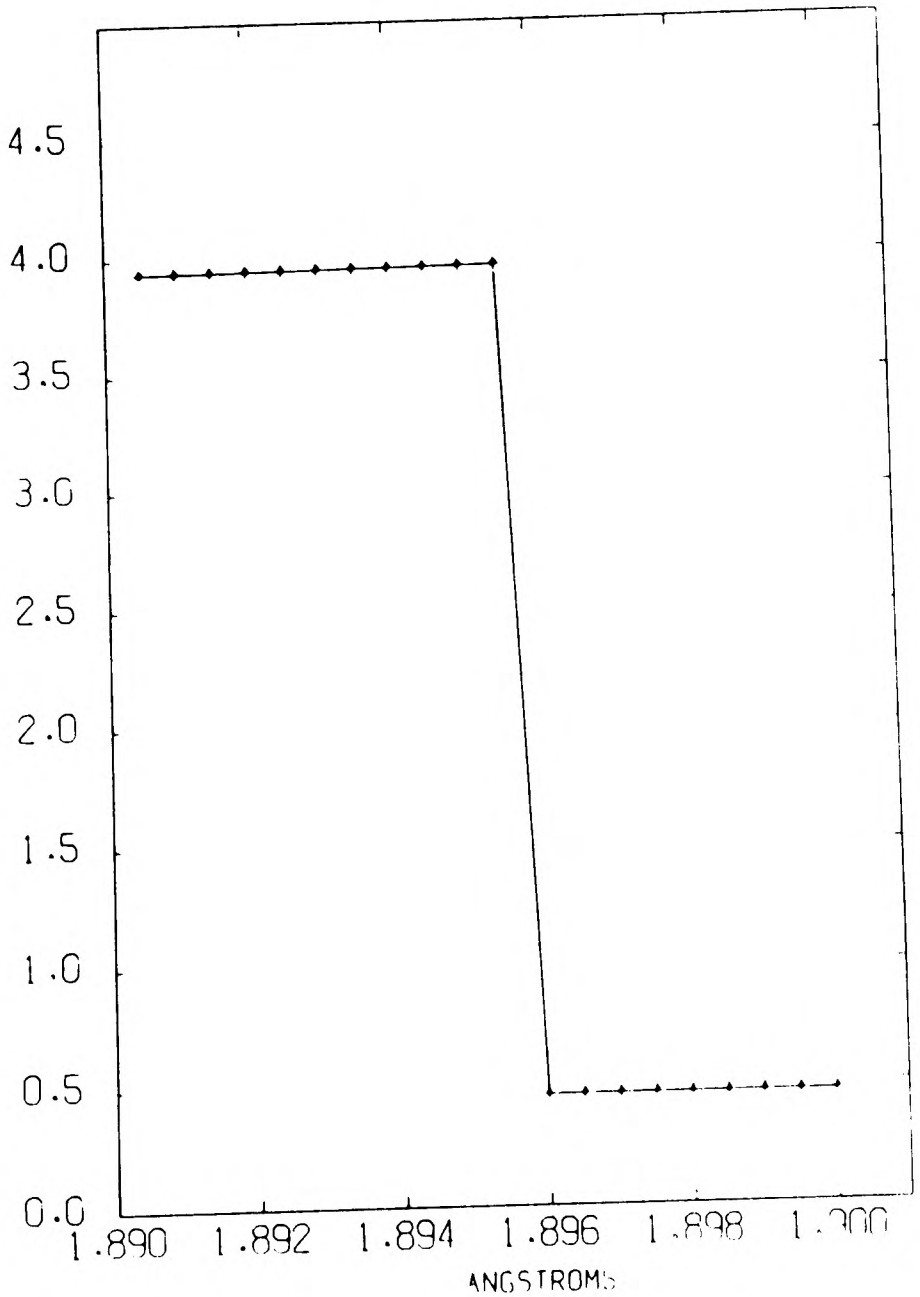
4.5.1 Introduction

Thoughts of finding an alternative method of solving the translation problem centred upon methods by which the position of the manganese atoms in the protein could be located. An attractive proposition was to utilise the anomalous scattering properties of the manganese and to solve the metal positions by anomalous difference Patterson techniques. The usefulness of anomalous difference Pattersons for locating the positions of heavy atoms in proteins was described as early as the 1960's (Rossmann, 1961; Matthews (1966b); Phillips (1966)). Unfortunately manganese is a light metal atom and a weak anomalous scatterer (see figure 4.14). Nevertheless, Hendrickson and Teeter (1981) had shown, that by careful measurement, the anomalous

f' V. WAVELENGTH



f'' V. WAVELENGTH



MU/RHO V WAVELENGTH

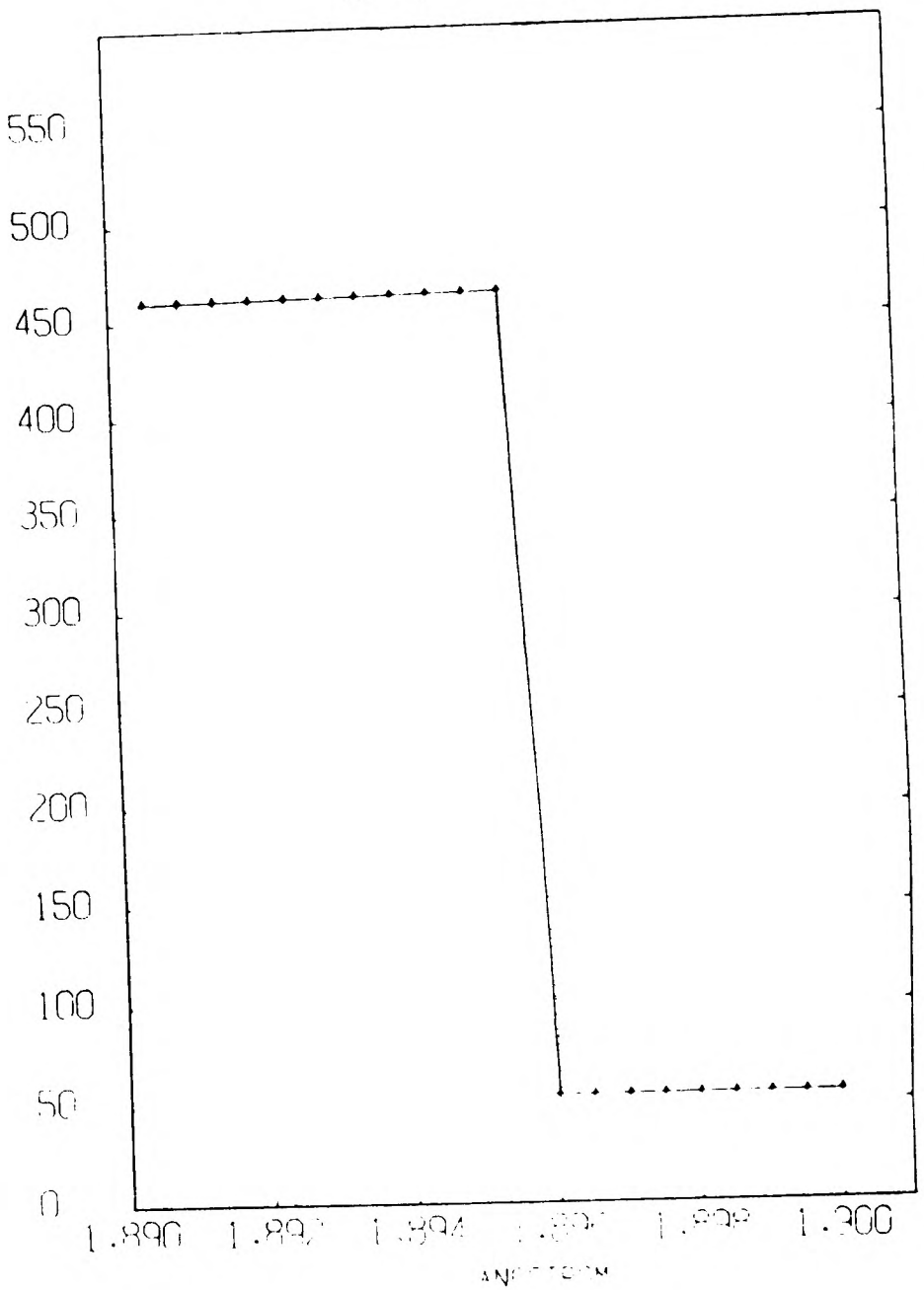


Figure 4.15 Behaviour of the dispersion (f') and absorption (f'') components of the anomalous scattering of manganese close to its absorption edge. Also shown is the variation of the absorption coefficient (mu/rho) as a function of wavelength close to the manganese edge. The values are calculated from photoelectric cross-section absorption data (Cromer & Liberman, 1981).

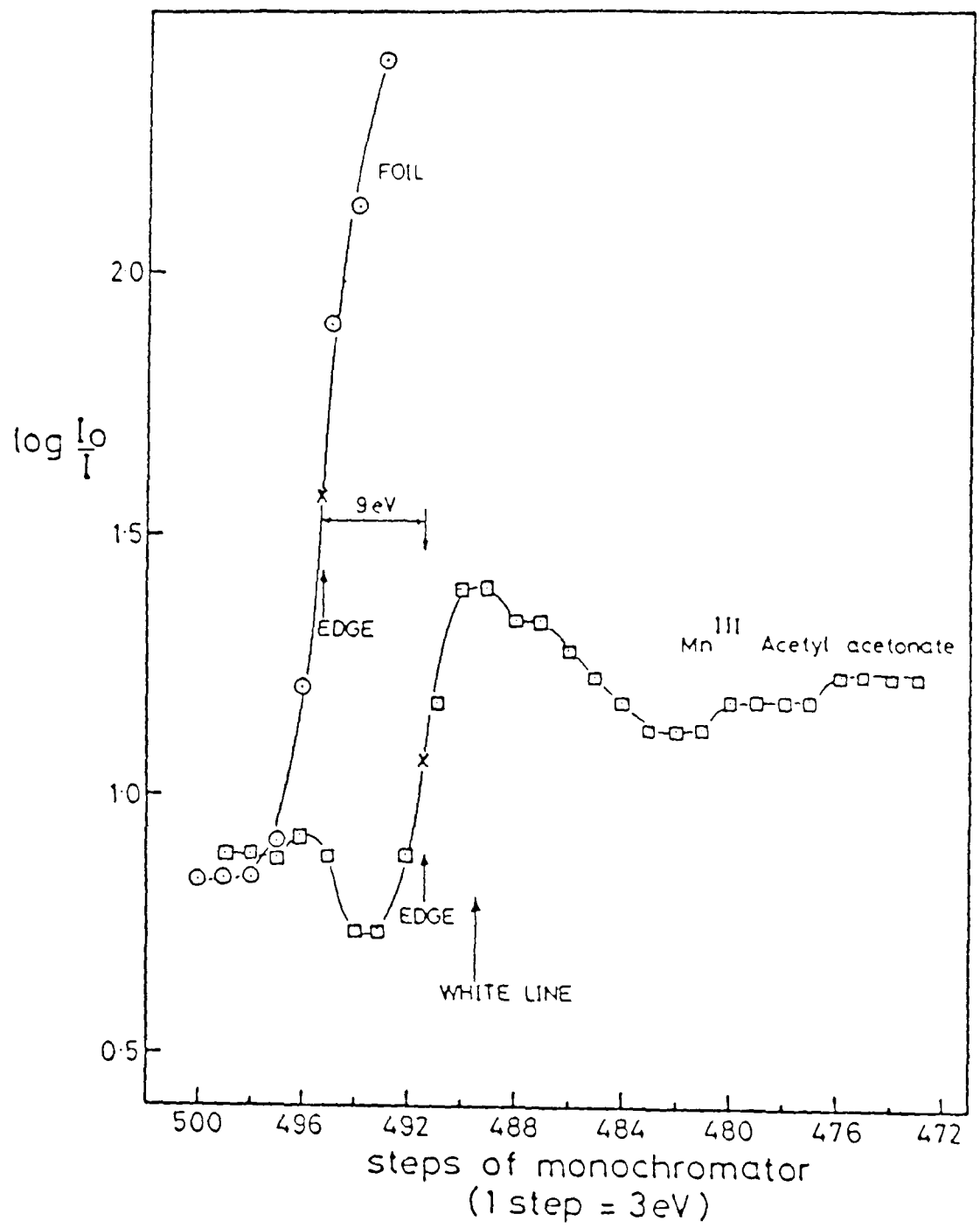


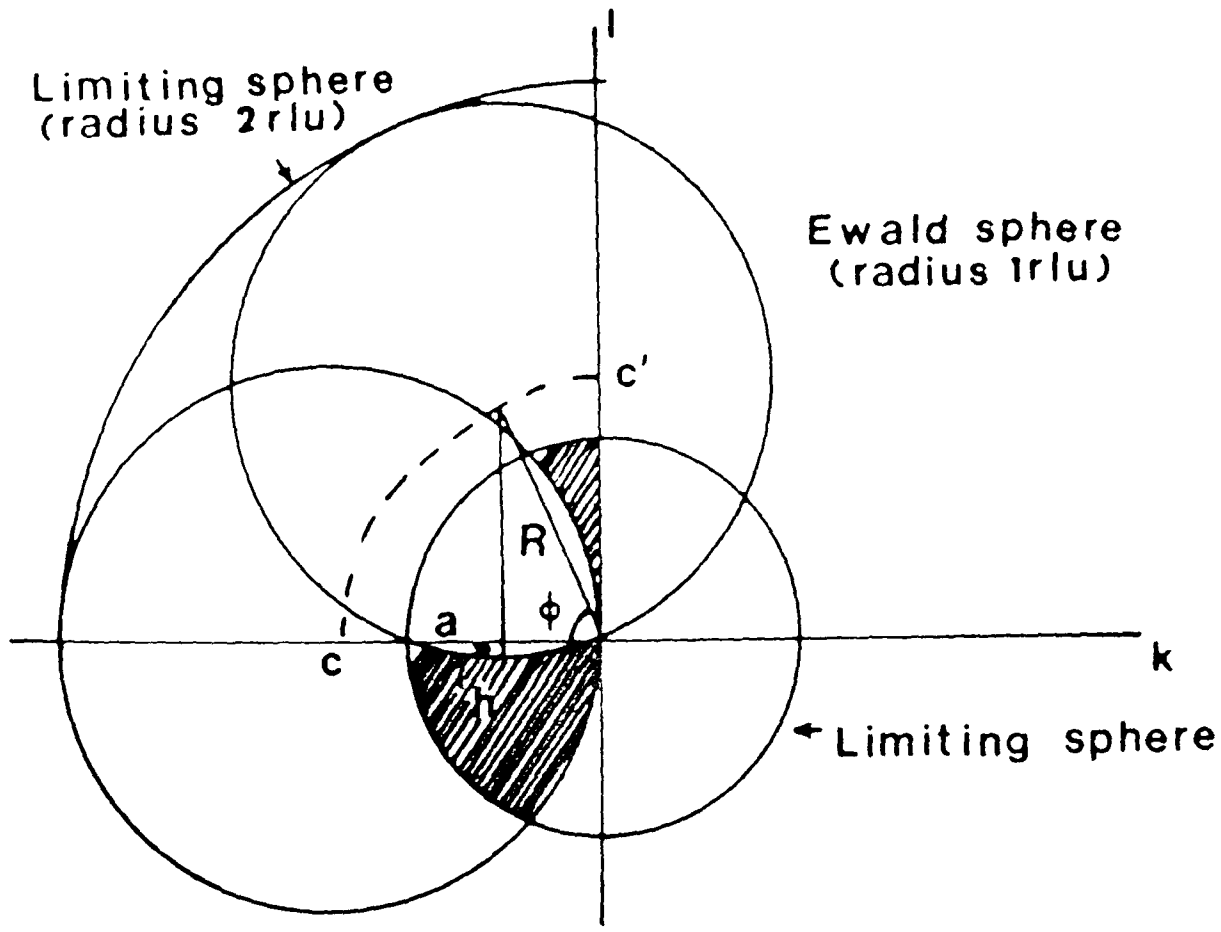
Figure 4.16 Energy scan across the manganese K edge for the compound, manganic acetyl acetonate, and for a manganese metal foil measured on the SRS PX station at Daresbury.

signal of weak scatterers could be measured and utilised. In their case, the Bijvoet differences arising from the anomalous scattering of sulphur from the protein crambin averaged only two per cent of the observed mean structure factor, yet sufficed to make a direct determination of the crystal structure feasible.

The first anomalous difference measurements of MnSOD were made on the in-house diffractometer. The resultant anomalous difference Patterson maps were very noisy and uninterpretable (see section 4.2). However, it was felt that the experiment could be successful if higher resolution data were collected close to the absorption edge of the manganese whereby the anomalous differences could be optimised. This experiment was made possible by the availability of the variable wavelength X-ray source from the synchrotron at Daresbury.

4.5.2 Data Collection

The first step was to determine the precise wavelength at which to carry out the experiment. As can be seen from figure 4.15, it was vital to make the measurements as close as possible to the short wavelength side of the edge. If measurements were made on the long wavelength side of the edge, almost no anomalous signal would be present. The wavelength required to optimise the imaginary dispersion component of the atomic scattering factor, f'', was determined by scanning the K absorption edge of a model compound using the EXAFS facilities at Daresbury. The compound chosen was manganic acetyl acetonate which had the right metal valence state and its octahedral coordination was thought to resemble the Mn environment in the protein. (The compound was relatively cheap, commonly available and its crystal structure known.) The white line was narrow (c.a. 10 eV) but its use essential to improve the measureability. To achieve the required spectal dispersion of about 3 eV (5×10^{-4}), a Si(III) 14.75° cut monochromator was used with a Guinier focussing distance of 1.9 metres. The



Volume of unique quadrant $Vq = 1/4 \times (4/3)\pi r^3$

$= \pi r^3/3$

and if $2a \leq \lambda$/resolution, then

Volume of Ewald sphere cap $Vc = (1/3)\pi h^2(3R-h)$

$= (\pi/3)h^2(3-h)$

since R = 1 rlu (reciprocal lattice unit).

Now $h = R-R\sin\Phi = 1 - \sin\Phi$ and so

$$Vc = (\pi/3)[2-3\sin\Phi+\sin^3\Phi].$$

For $\Phi = 70^\circ$, the percentage of main region data lost is about 2.3%.

Figure 4.17 Projection down a* rotation axis showing area of reciprocal space swept out (cross-hatched) as the crystal is rotating through about 70°. For 222 symmetry, one asymmetric unit of reciprocal space corresponds to one octant so that the equivalent of four asymmetric units are swept out in a full 90° rotation.

measured shift of 9 eV in the edge position relative to the results of a manganese foil agrees with work elsewhere (Greaves, pers. commun.; Salem et al., 1978). The absorption scan of the model compound at the PX7.2 station (Daresbury) is shown in figure 4.16. Note that the use of the manganese foil alone to establish the optimum wavelength (which was 1.890Å) would have led to virtually no anomalous signal at all.

The data collection strategy was based on the choice of rotation axis and the point group symmetry. The choice of the principal rotation axis depended on the most stable crystal mount. The most stable mount for a MnSOD crystal occurs when the longest crystal edge is aligned along the length of the capillary tube. This usually corresponds to an $\underline{a}^*$ or $\underline{c}^*$ mount. The maximum oscillation range per photograph to avoid spot overlap depends on the limiting axis for a given crystal mount (equation 2.57); in this case, the $\underline{b}^*$ axis. The maximium oscillation range for a MnSOD crystal, with the $\underline{b}^*$ axis pointing in the direction of the beam, was calculated to be 0.7° for a maximum resolution of 2.0Å spacings and a reflecting range for the crystal of 0.3°. In practice, the oscillation range was slightly overestimated to increase the number of fully recorded reflections at the expense of a small number of overlaps. For 222 symmetry, a total rotation range of 90° produces four equivalents of each reflection. However, it can be shown that only a small percentage of unique reflections are lost by rotating the crystal over a smaller range (figure 4.17).

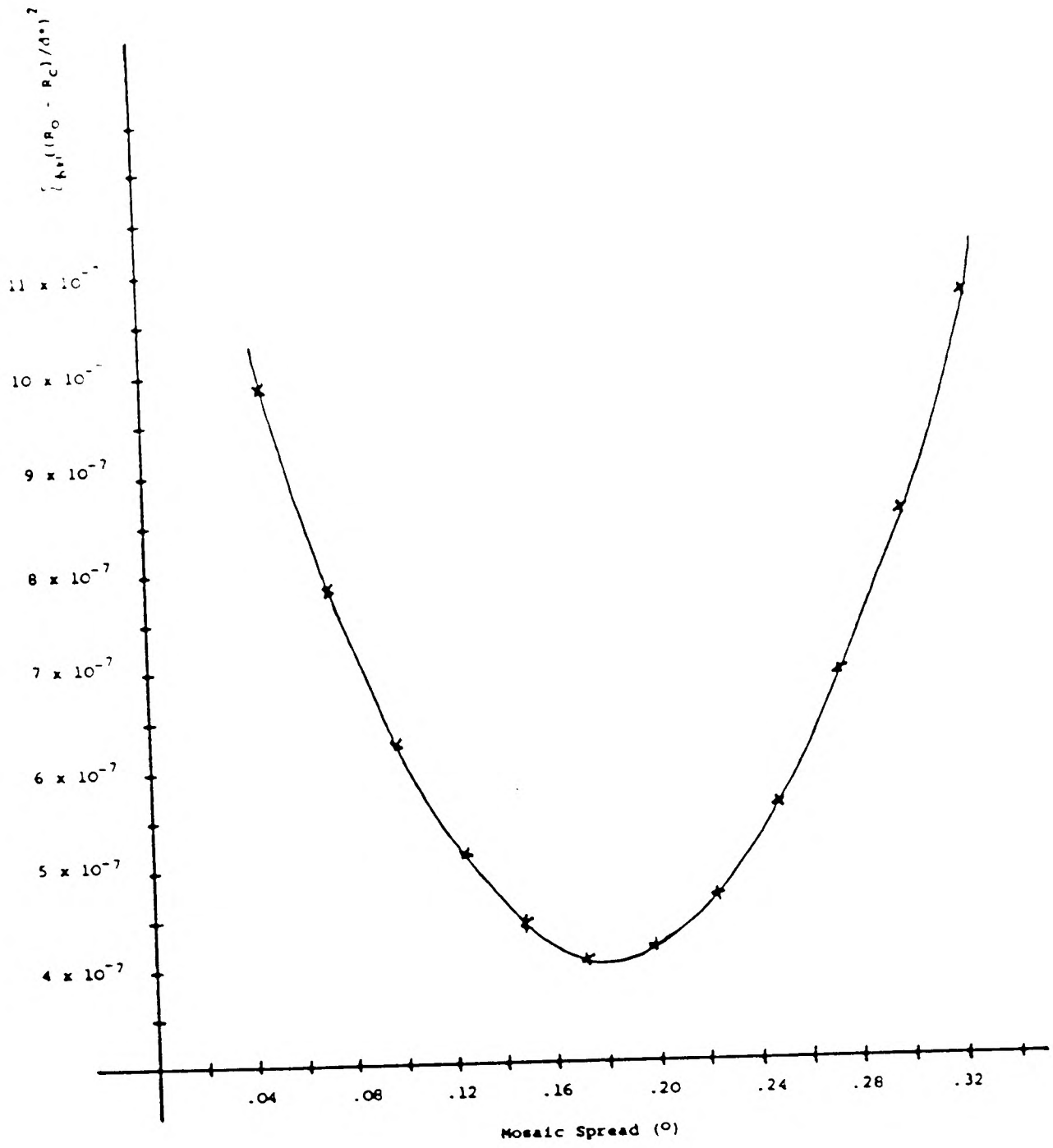
In all, two crystals were required for the data collection. The crystals were mounted with their longest edge running parallel to the mounting tube. Both crystals were quite large (about 1.3 X 1.0 X 0.5 mm^3 in linear dimensions) which, in theory, meant that fresh parts of the crystal could be exposed to the beam during the course of irradiation by translating the crystal through the beam. However, both crystals cracked in several places early on in the data collection and so limited the area of each crystal that could be exposed to the beam. The data were collected on

CEA film which had been loaded into V-shaped cassettes with two films per cassette. The crystals were cooled down to 5°C before irradiation so as to limit the effects of radiation damage. However, both crystals radiation damaged markedly during exposure to X-rays. Initial exposure times (200 seconds/degree of rotation) had to be doubled near the end of the lifetime of each crystal and the diffracting power of each crystal correspondly dropped from 2.0 to 2.3Å resolution. The severe radiation damage suffered by the crystal must have been due to the long wavelength (1.89Å) used because crystals suffered little damage at the shorter CuKα wavelength used in other experiments (see section 3.3). The thinnest parts of the crystal were exposed to X-rays first because of the significant absorption problems encountered with working at the long wavelength. The rotation angle was defined to be zero at this position. Both crystals were set as accurately as possible from stills to ensure that Friedel pairs of reflections were captured on the same film. The wavelength stability of the beam was checked before and after irradiation of each crystal and found to be constant.

4.5.3 Data Processing

The quality of the photographic data from the first crystal was poor; especially in the later packs of film where alternate rows of reflections had elongated spot shapes. The data from the first crystal had to be discarded.

The thirty four packs of film collected from the second crystal were processed using the computer program suite at Daresbury Laboratory. The theory and methods used have been discussed in chapter two. A number of observations are worth noting. First, the spots were small enough to warrant scanning the films with a fifty micron raster. It was apparent that the crystal had suffered quickly from radiation damage in the early exposures which was reflected in a steep rise in the value of the mosaic spread parameter required for good prediction of the spot pattern. Using a



			Conventional Geometry	SR Geometry
$\eta + \gamma$ (refined)			0.27°	-
η (refined)			-	0.185°
γ_H			-	0.15°
γ_V			-	0.013°
$(\delta\lambda/\lambda)_{conv}$			2.5×10^{-3}	0.77×10^{-3}
Prediction:	Pack 1	Full	492	671
		Partial	989	632
	Pack 2	Full	492	672
		Partial	984	613
Post refinement:	NREF		72	56
	In range		46	41
	Out of range		26	15
	$\sum_{hkl}((R_o - R_c)/d^*)^2$		5.2×10^{-7}	4.0×10^{-7}

Figure 4.18 and Table 4.5 The effect of refining mosaic spread on prediction parameters using conventional or synchrotron geometry in the refinement programs.

conventional geometry description of the mosaic spread (η) and beam crossfire (γ), and cycling through prediction, integration and post refinement for the first two packs led to a refined value of 0.27° for ($\eta+\gamma$). The refinement was carried out by varying the value of ($\eta+\gamma$) in repeated runs of the computer program, IDXREF, and plotting against the minimized residual,

$$\Sigma_{hkl}((R_o-R_c) / d^*)^2$$

where R_o and R_c are the observed and calculated distances of a lattice point from the centre of the Ewald sphere, respectively and d^* is the distance of the lattice point from the reciprocal lattice origin. Convergence was considered to be attained when two successive runs of prediction, integration and post refinement gave the same misorientation angles and the same minimum residual in IDXREF for the same value of ($\eta+\gamma$). The conventional reciprocal lattice volume radius is given by the equation,

$$\varepsilon_c = ((d^*\cos\theta)/2)\ [\ \eta + \gamma + (\Delta\lambda/\lambda)\tan\theta].$$

To allow for synchrotron geometry, the equation is modified to give

$$\varepsilon = 1/2\ [(\delta d^{*2} + z\gamma_H)^2 + \gamma_V^2 y^2]^{1/2} + ((d^{*2}\cos\theta)/2)[\eta + (\Delta\lambda/\lambda)\tan\theta]$$

where δ is the correlated component of the spectral dispersion, γ_H and γ_V are the horizontal and vertical crossfire angles, $\Delta\lambda/\lambda$ is the conventional type spectral dispersion and y and z are the crystal axis coordinates perpendicular to the beam with z the rotation axis (Greenhough & Helliwell, 1982). Using the correct geometry on the first two packs ($\gamma_H = 0.15^{\circ}$, $\gamma_V = 0.013^{\circ}$ and $(\Delta\lambda/\lambda)_{conv.} = 0.77 \times 10^{-3}$), a refined value of the mosaic spread was 0.185° (figure 4.18). The effect of the differences in the value of the mosaic spread obtained from conventional and synchrotron geometry can be seen in table 4.5. The prediction of partially and fully recorded spots differed dramatically between the two. Use of the conventional geometry led to too many partials being predicted towards the vertical axis and too few near the cusp. A simplified view is that the ($\eta+\gamma$) value of 0.27° from conventional processing was a compromise between the (η+0.013) needed in

Pack Number	Oscillation Range (^{0})	Misorientation Angles (^{0})	Total Spots Predicted	Mosaic Spread	% Fullies	% Partials	Saturated A	Saturated B	RMS R(IDXREF) (^{0})	R Intra (%)	Number of Scaling Rejections	Number of Negative Intensities	I >3sd (%)
1	30.0-31.0	0.29 -0.67 -0.56	1352	0.19	43	54	46	0	0.023	4.5	0	10	61
2	31.0-32.0	0.29 -0.67 -0.56	1339	0.19	44	53	48	1	0.043	6.6	2	12	61
3	32.0-33.0	0.30 -0.55 -0.48	1560	0.25	33	62	39	0	-	5.9	7	11	58
4	33.0-34.0	0.30 -0.55 -0.48	1564	0.25	31	63	69	6	-	5.8	5	9	60
5	34.0-35.0	0.29 -0.66 -0.49	1507	0.35	32	61	61	1	0.031	6.0	9	9	60
6	35.0-36.0	0.29 -0.66 -0.49	1449	0.35	29	65	42	1	0.034	5.9	1	11	56
7	36.0-37.0	0.29 -0.66 -0.46	1559	0.35	33	62	59	1	0.038	6.2	4	11	51
8	37.0-38.0	0.29 -0.66 -0.46	1524	0.35	31	64	44	0	0.046	11.8	35	16	42
9	38.0-39.0	0.32 -0.63 -0.48	1489	0.35	30	63	41	3	0.033	7.8	9	13	51
10	39.0-40.0	0.30 -0.65 -0.49	1530	0.35	31	62	44	0	0.038	8.3	9	14	51
11	40.0-41.2	0.30 -0.65 -0.49	1752	0.35	38	51	44	1	0.027	7.7	10	13	55
12	41.2-42.4	0.30 -0.65 -0.49	1660	0.35	38	50	41	1	0.027	6.8	1	11	56
13	42.4-43.6	0.33 -0.61 -0.50	1738	0.35	39	49	39	0	-	6.1	5	13	55
14	43.6-44.8	0.32 -0.63 -0.49	1722	0.35	39	49	62	1	0.026	7.2	5	13	52
15	44.8-46.0	0.32 -0.63 -0.49	1683	0.35	37	50	38	1	0.030	7.3	9	11	54
16	46.0-47.2	0.35 -0.60 -0.51	1742	0.35	39	50	24	0	0.036	6.7	7	11	53
17	47.2-48.4	0.36 -0.59 -0.49	1672	0.35	37	50	21	0	0.037	6.8	7	15	47
18	48.4-49.6	0.36 -0.59 -0.49	1731	0.35	41	48	17	0	-	6.9	5	15	47
19	49.6-51.0	0.36 -0.59 -0.49	1822	0.35	45	40	29	2	0.036	7.3	3	12	46
20	51.0-52.4	0.40 -0.55 -0.53	1691	0.35	45	43	31	0	0.040	7.6	6	11	51
21	52.4-53.8	0.41 -0.55 -0.53	1706	0.35	43	44	16	0	0.037	8.5	5	14	46
22	53.8-55.2	0.42 -0.54 -0.51	1799	0.40	39	44	16	0	-	7.4	8	15	41
23	55.2-56.6	0.42 -0.54 -0.57	1729	0.40	37	48	9	0	-	6.7	1	17	44

24	56.6-58.0	0.42	-0.54	-0.57	1786	0.40	38	46	11	1	-	6.8	2	15	43
25	58.0-59.4	0.45	-0.52	-0.53	1590	0.40	39	48	9	0	0.063	9.0	10	18	41
26	59.4-60.8	0.45	-0.52	-0.53	1562	0.40	39	48	18	0	-	7.4	4	14	47
27	60.8-62.4	0.45	-0.52	-0.52	1692	0.40	42	41	12	0	-	7.3	2	12	52
28	62.4-64.0	0.44	-0.53	-0.52	1545	0.40	41	44	16	0	0.038	9.0	11	12	47
29	64.0-65.6	0.44	-0.53	-0.52	1540	0.40	42	44	19	1	0.039	9.1	6	12	53
30	65.6-67.2	0.44	-0.53	-0.55	1537	0.40	43	41	34	7	-	8.7	3	13	50
31	67.2-68.8	0.45	-0.53	-0.58	1537	0.40	41	41	18	4	0.049	9.8	8	12	48
32	68.8-70.6	0.45	-0.53	-0.58	1674	0.40	44	37	40	6	0.061	9.5	4	10	59
33	70.6-72.4	0.45	-0.53	-0.57	1696	0.40	42	36	41	5	-	9.5	11	11	56
34	72.4-74.2	0.45	-0.53	-0.56	1688	0.40	41	36	65	2	-	9.3	4	9	64

Table 4.6 Film data parameters and intensity statistics.

Key: 1) There were two films, A and B, in each film pack.

2) Reflections were designated as 'fullies' or 'partials' depending on whether they were fully or partially recorded on the film.

3) R(IDXREF) = $\Sigma_{hkl}(Ro-Rc)/d^*)^2$ where Ro and Rc are the observed and calculated distances of a lattice point from the centre of the Ewald sphere, respectively and d^* is the distance of the lattice point from the reciprocal lattice origin. Values shown were obtained after post-refinement. Some values were not available at the time the table was compiled.

4) $R(Intra) = \sum_{hkl}\sum_{j=1,2} |\bar{I} - I_j| / \sum_{hkl}\sum_{j=1,2} \bar{I}$

where $\bar{I}$ is the weighted average value of the intensity of reflection hkl. The summation on j is over all films which hkl is observed and is performed only over reflections recorded on more than one film.

5) Reflections were rejected after intra-pack scaling if they were lying in unacceptably low background areas (for example, if background areas lay outside the exposed area of the film), had especially large Lorentz or polarisation factors or if their local scale factor deviated by more than three standard deviations from the overall scale between the two films.

Cell dimensions were a=72.39Å, b=110.75Å and c=50.97Å

Total number of reflections processed = 47519

Total number of fully recorded reflections = 20424

Total number of partially recorded reflections = 11047

Total number of reflections rejected on sd grounds = 956

Total number of independent reflections = 16056

Total number of negative intensities = 1186

Total number of centric data = 1182

Total number of reflections with anomalous measurements = 13290

Number of anomalous measurements from the same film pack = 10944

Number of reflections with intensties greater than 3sd = 63%

Number of anomalous differences greater than 3sd = 2.3%

Rsym = 6.9%

Rmerge = 8.9%

Table 4.7 Intensity statistics for the oscillation film data.

Key: 1) The criteria for rejecting reflections on standard deviation grounds was outlined in the note to table 4.4.

2) $Rsym = \sum_{hkl} |I^+ - I^-| / \sum_{hkl} \bar{I}$ where

$$\bar{I} = (I^+ + I^-) / 2$$

3) $$Rmerge = \sum_{hkl} \sum_{i=1}^{N} |I_i - \bar{I}| / \sum_{hkl} N\bar{I}$$

where I is the value of the ith of N measurements of an intensity and $\bar{I}$ is the mean of all N measurements of that intensity.

Rfriedel = 6.7%

Kemp = 2.6

Total number of negative deltas = 4858

Total number of positive deltas = 4774

Sum total of negative deltas = -218872

Sum total of positive deltas = 211674

Number of deltas greater than 5 sds = 28

Number of distribution outliers = 9

Number of terms where $|F_{PH} - F_P|$ was greater than 5 F_{PH} = 116

Percentage of data that passed ANOMSORT tests = 8%

Table 4.8 Anomalous data statistics after local scaling.

Key:

$$Rfriedel = \Sigma|F_{PH}^{+} - F_{PH}^{-}| / \Sigma \bar{F}_{PH}$$

$$K_{emp} = 2 \Sigma|F_{PH} - F_P| / \Sigma|F_{PH}^{+} - F_{PH}^{-}|$$

$$Delta = F^{+} - F^{-}$$

Distribution outliers were anomalous measurements which were five times the root mean square delta for the whole data set.

Reflections passed the ANOMSORT tests if

1) they had a non-zero delta,

2) they were not distribution outliers,

3) they had a delta greater than two times the standard deviation of the delta, and

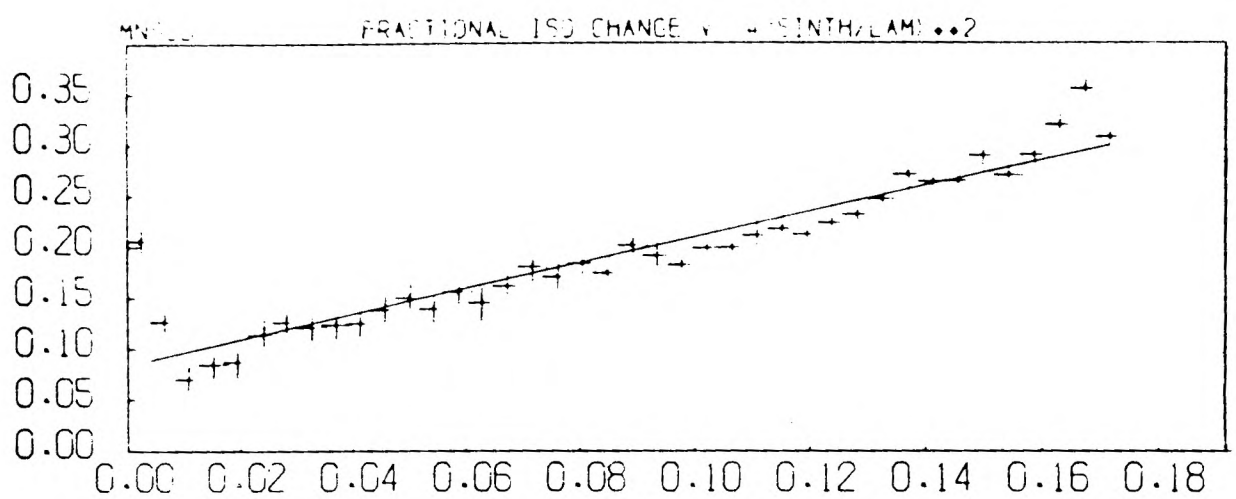
4) $|F_{Daresbury} - F_{diffractometer}| \leq 5 \times F_{Daresbury}$.

the vertical and the (η+0.15) needed towards the horizontal.

Details and statistics of the data processing are given in tables 4.6 and 4.7. The data processing did not proceed smoothly because of constant crystal slippage throughout the data collection as evidenced by the changes in the refined misorientation angles. The somewhat sudden appearance of split spots on films in latter packs was reflected by a sharp rise in the value of the mosaic spread required for good prediction. However, there seemed to be very little partial bias introduced by possible overestimation of the mosaic spread parameter. The mean fractional partial bias (= mean[$\langle I_F \rangle - I_p$]/mean$\langle I_F \rangle$, where I_F and I_p refer to intensities calculated from fully and partially recorded reflections, respectively) was less than 1% and no variations were apparent when the bias was analysed against intensity or resolution. The data were generally weaker than the native diffractometer data (see table 4.1) with the percentage of intensities greater than three standard deviations being 63% overall. Nevertheless, the merging R-factors for symmetry equivalents and for all the data were quite reasonable. Despite the absence of blind region data and some main region data, 73% of the total number of reflections to 2.4Å resolution had been measured from the oscillation film data.

4.5.4 Data Analysis

The Daresbury data were scaled to the native diffractometer data using the program ANSC. There were bad trends with the Miller indices and resolution before scaling. A program was written (ANSCCORR) to apply the three-dimensional scales output from ANSC to the Daresbury data. After the correction for primary absorption, the bad trends with the diffractometer data disappeared. However, the mean fractional isomorphous change between the two data sets increased linearily with resolution with an overall value of 18% (figure 4.19). This trend may have been due, in large part, to the significant radiation damage the crystal had suffered during data



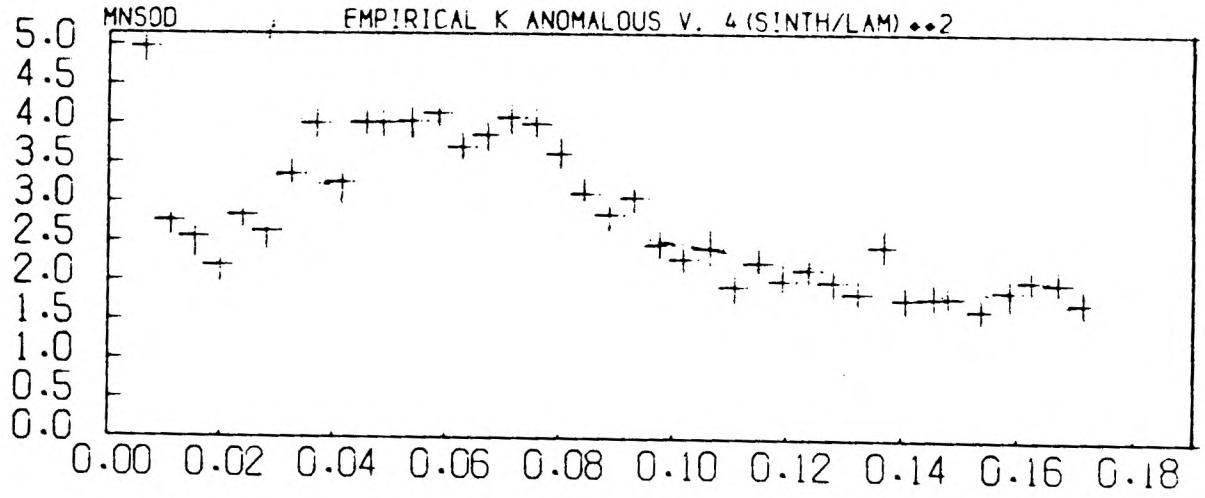


Figure 4.19 Plots of mean fractional isomorphous difference and k_{emp} as a function of resolution ($4\sin^2\theta/\lambda^2$) after scaling the Daresbury data to the diffractometer data.

collection. The Friedel-related reflections were corrected for residual systematic errors using the local scaling program LOCAL. Before scaling, the root mean square deviation in the local scale factor (0.057) was greater than its variance (0.033) so a local scaling correction was required. The correction was applied using an optimum sphere size of forty reflections (see chapter two).

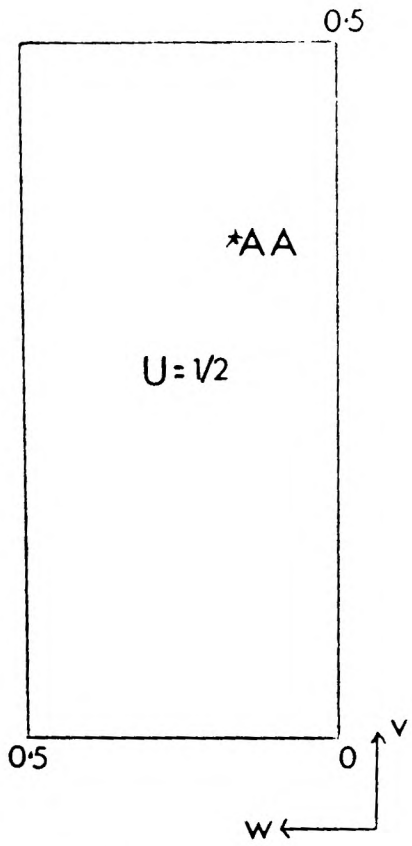
The anomalous measurements involved small differences and so it was important to eliminate erroneous measurements which might affect an anomalous difference Patterson. Hence a computer program was written (ANOMSORT) to analyse the anomalous measurements and throw out unreliable ones. Statistics calculated by the program are compiled in table 4.8. As can be seen from the table, very few measurements passed the rejection criteria. This was primarily due to the fact that the majority of anomalous differences were less than two standard deviations of their measurement and thus rejected on the basis of being unreliably measured. However, relaxing this criterion made little impact on the resultant Patterson maps. There were no significant trends of the root mean square anomalous differences with increasing size of intensity or with resolution although at higher resolution (greater than 3Å) the root mean square anomalous measurements were less than their corresponding root mean square standard deviations. The anomalous measurements were bias-corrected by the term,

$$\Delta_{ano} = \Delta_{ano}^3 / \Delta_{ano}^2 + (1.5 \text{ X } sd_{\Delta_{ano}})^2$$

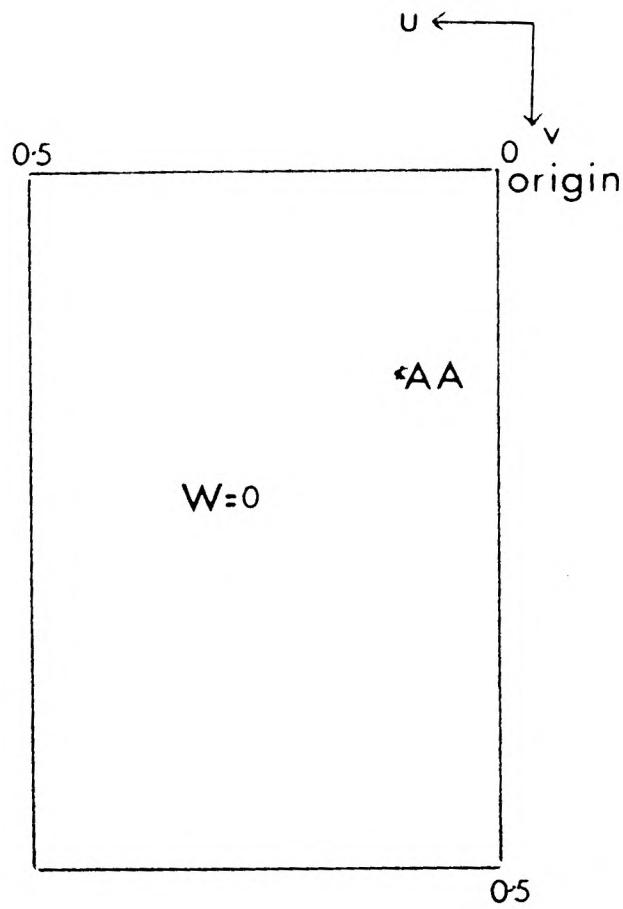
as suggested by Papiz (M. Papiz, pers. commun.).

4.5.5 Patterson Syntheses

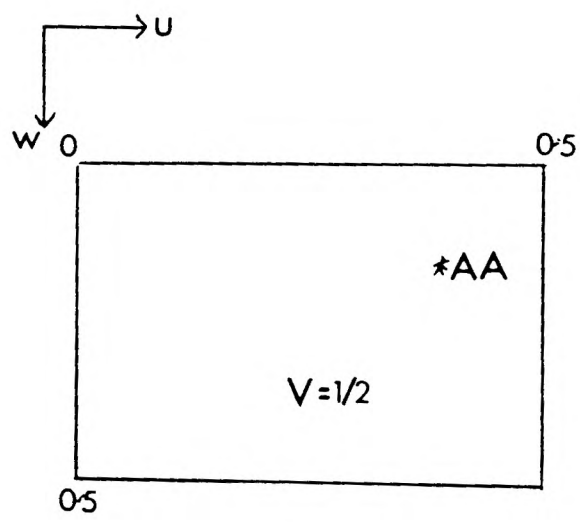
Anomalous difference Pattersons with coefficients $(F^+ - F^-)^2$ were calculated at different resolutions. The Harker sections of one such map are shown in figure 4.20. The maps tended to be very noisy when calculated with only low resolution terms (to 5Å) but quite flat when higher resolution terms were included. Nevertheless, the peak marked A in figure



$v = 1/2 - 2y$, $w = 2z$



$u = 2x$, $v = 2y$



$u = 1/2 - 2x$, $w = 2z$

Key: Harker section layout. The position of self-vectors (AA) for some heavy atom site A is also shown.

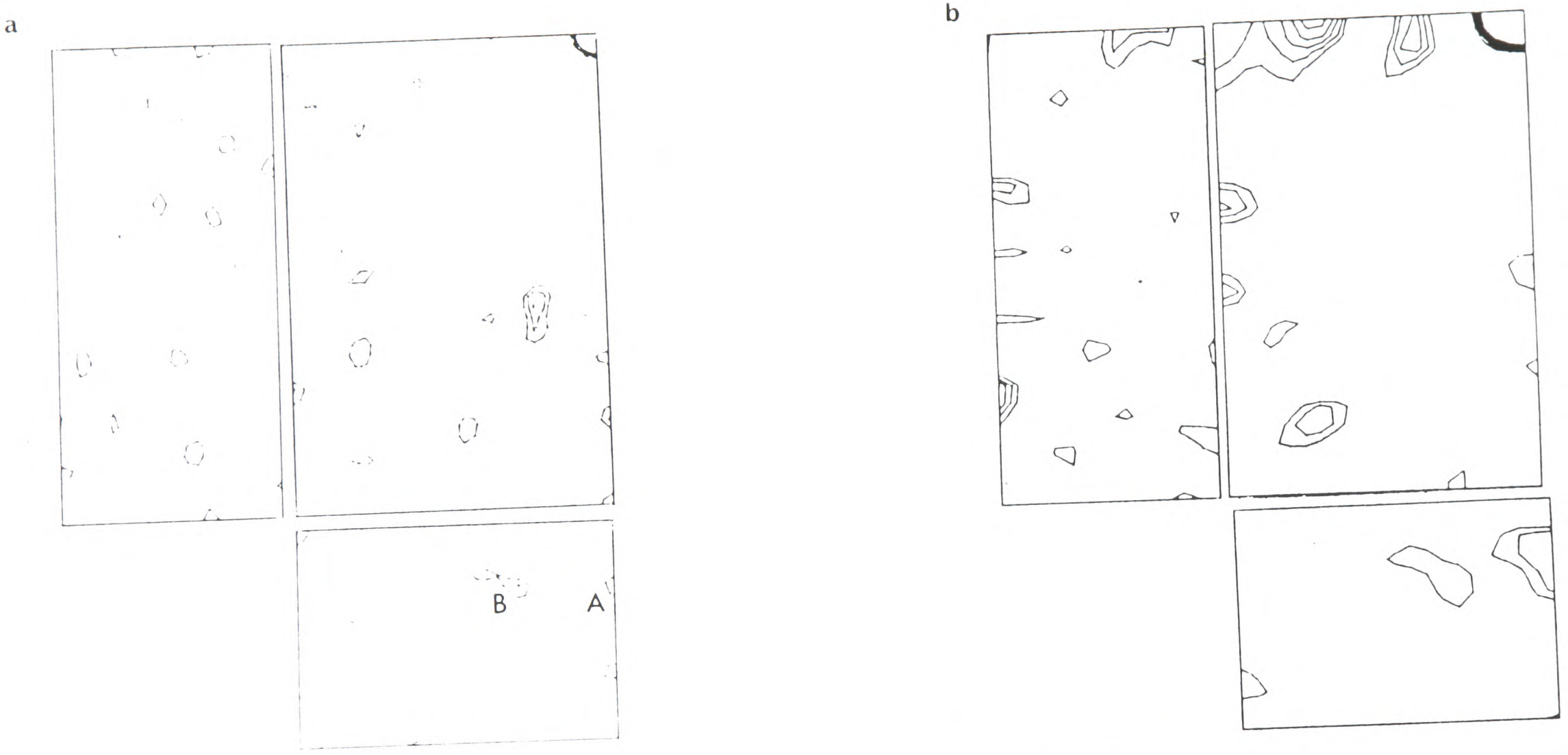


Figure 4.20 Harker sections (u=1/2, w=0 and v=1/2 from upper left going clockwise) of a) the anomalous difference Patterson calculated to 3.5Å resolution from the Daresbury data, and b) the isomorphous difference Patterson calculated to 6Å resolution from diffraction data of a suspected manganese-free crystal collected on the diffractometer. Both maps were contoured at equal but arbitrary intervals with the first level omitted.

4.20 was present in all the maps calculated and in most was the highest non-origin peak. The peak labelled B in figure 4.20 was also present as a positive feature in all maps calculated but tended to subside into the noise level in higher resolution maps. Despite these observations, no convincing solution of the Pattersons could be found.

The appearance of low resolution isomorphous difference Patterson maps calculated from diffraction data of a suspected manganese-free crystal (see next section) prompted a re-inspection of the anomalous difference maps. The Harker sections of the isomorphous difference map are shown in figure 4.20. The similarity of the v=1/2 Harker sections stands out. However, it was still not possible to solve either Patterson in an unambigous and convincing fashion.

4.6 Metal Replacement Experiments

4.6.1 Introduction

At this stage of the work, both conventional translation search functions and anomalous dispersion data had failed to help solve the translation problem. It was well known that the manganese atoms of the protein could be extracted and replaced by other light transistion metal atoms without any noticable conformational changes to the protein as judged by gel electrophoresis or nmr (see section 1.6). It was reasoned that removal of the manganese atoms to be replaced by heavier metal atoms could produce potential heavy atom derivatives. The translation problem might be solved by inspection of the resultant isomorphous difference Pattersons.

4.6.2 Manganese Extraction Studies

It was thought best to carry out the manganese extraction studies on the crystals rather than in solution for a number of reasons. The extraction and eventual metal replacement experiments might alter the

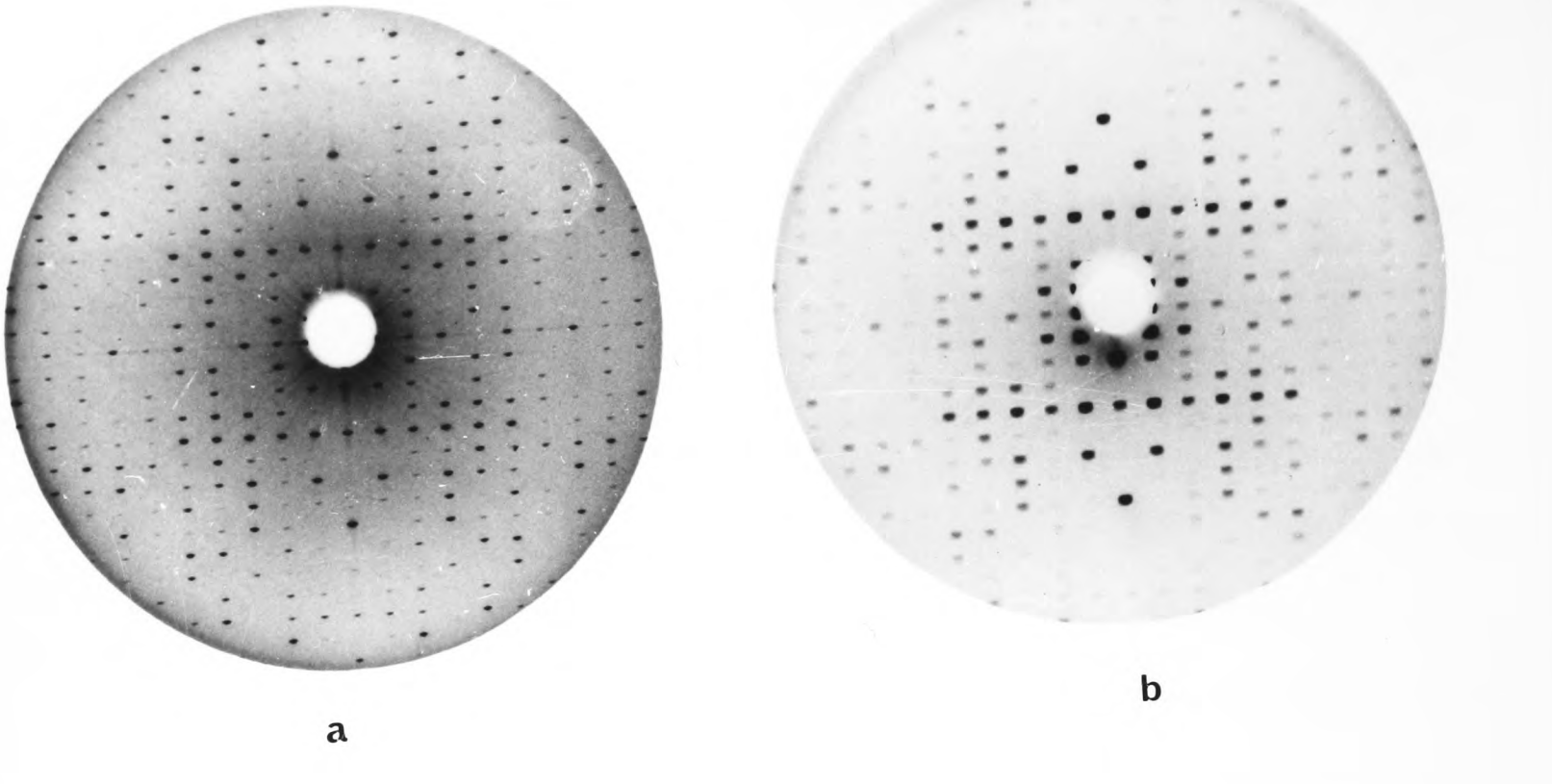


Figure 4.21 Precession photographs (μ=9°, crystal-to-film distance = 100 mm) of principal centric zones from crosslinked MnSOD crystals, (a) h0l and (c) 0kl of native, and (b) h0l and (d) 0kl of a suspected manganese-free crystal.

crystallization properties of the protein. There is some controversy concerning the nature of the apoprotein; in most people's hands, the protein almost completely unfolds upon metal extraction (J. Bannister, pers. commun.). It was hoped that the crystal forces between the protein molecules might prevent the unfolding.

It was known from metal extraction studies of MnSOD that, although the tervalent maganese ion is tightly bound to the protein, the divalent ion is quite labile. Thus it was planned to treat the crystals with sodium dithionite to reduce the metal in the presence of an excess of a metal chelating agent. Crystals were dropped into a solution of 10mM EDTA made up in artificial mother liquor (0.6g PEG/ml of buffer) and a pinch of dithionite added. When viewed under a microscope, the crystals lost their colour gradually, starting from the edges. Within five minutes, the crystals were completely colourless. Large crystals were prone to rapid cracking (along crystallographic b-c planes) upon addition of dithionite. Small crystals withstood the treatment.

At about this time, it was discovered that crystals could be crosslinked (see chapter five) so that all further experiments were performed on crosslinked crystals. Crosslinked crystals behaved in a similar fashion to non-crosslinked crystals when treated with dithionite except that they retained a straw yellow colour after treatment. The treated crystals diffracted well and showed small but significant intensity changes compared with the native diffraction pattern (figure 4.21).

The diffraction pattern of one of these crystals was measured to a resolution of 6Å on the five circle diffractometer. Relevant measuring parameters and statistics of the data collection and processing are compiled in table 4.9. Complete details of the methods used to collect and process diffractometer data have been discussed elsewhere in this thesis so will not be repeated here. A number of observations are worth pointing out. Surprisingly, the crystal suffered significant radiation damage of 28% for

Crystal mount = c*

Counter mode = single counter

Scan width = 22 steps

Peak width = 14 steps

Step size = 0.04°

Count time = 3.0 seconds/step

Number of reflections measured = 1233

Percentage of intensities greater than 3 sds = 76%

Maximum absorbance = 1.62

Radiation damage = 28%

Number of reflections rejected = 285

Number of independent reflections = 924

Mean fractional isomorphous difference = 8.8%

Table 4.9 Data collection parameters and processing statistics for the suspected manganese-free crystal.

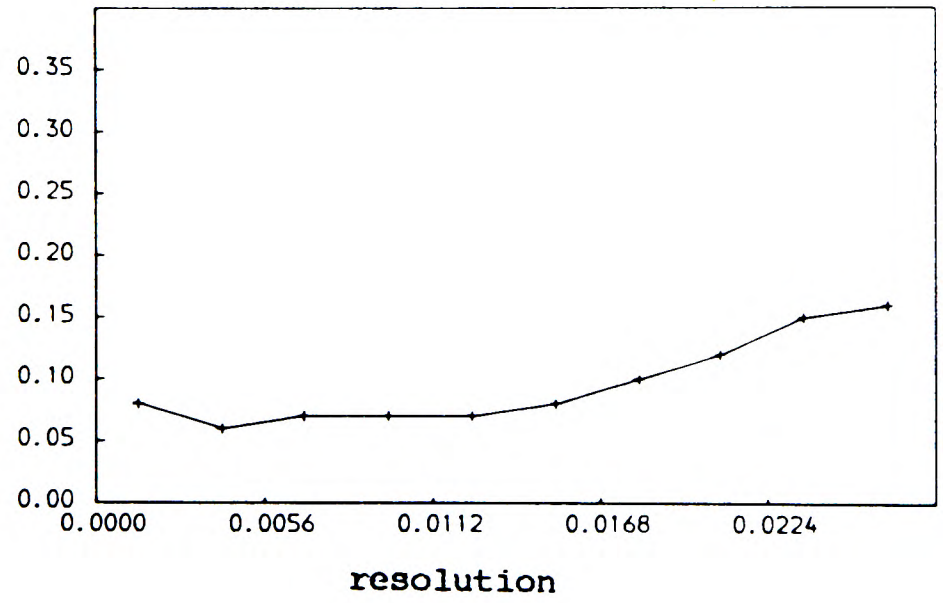


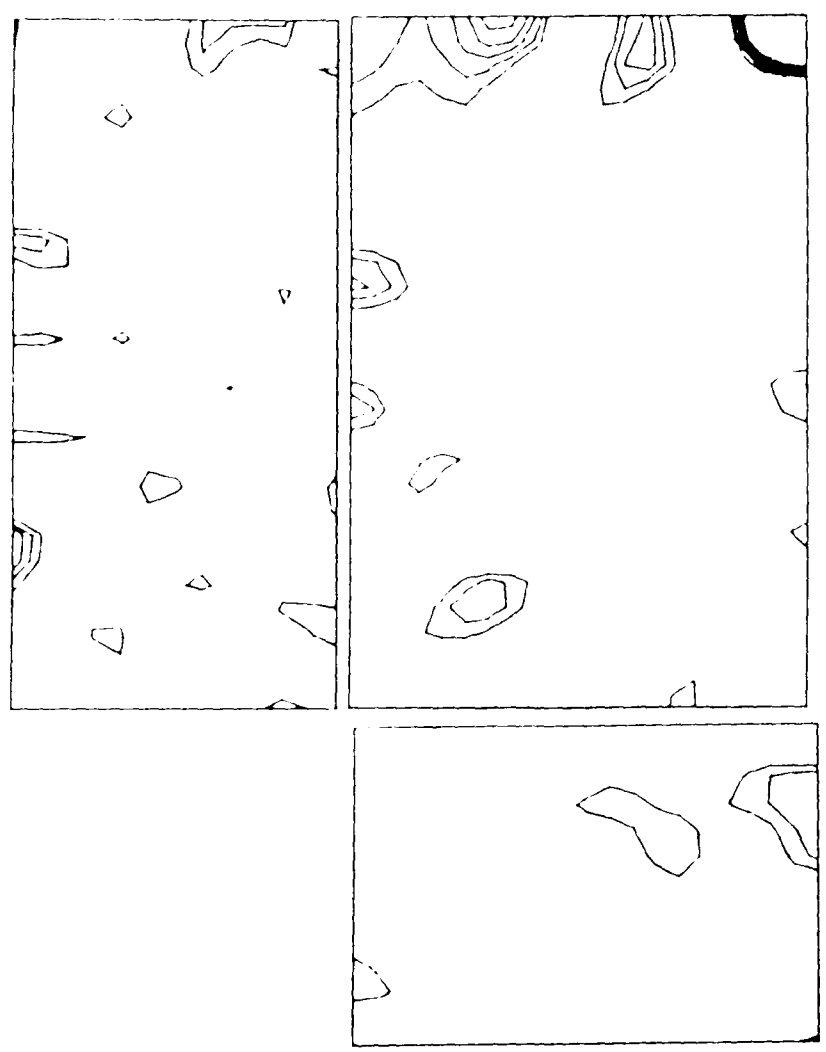
Figure 4.22 Mean fractional isomorphous difference as a function of resolution ($4\sin^2\theta/\lambda^2$) for the suspected manganese-free data set.

Key: m.f.i.d.= $\Sigma|F_{PH}-F_P|/\Sigma F_P$ where F_{PH} are the structure factors of the manganese-free data set and F_P are structure factors of the crosslinked native data set discussed in Chapter 5.

only two days exposure to X-rays. This observation maybe related to the fact that the intact protein carries out an oxygen radical scavenging role and the apoprotein is inactive. A large shell of reflection data, measured close to the end of data collection, had to be rejected because reference reflections showed an alarming drop in intensity. The reasons for this were unclear; it may have been due to crystal slippage or a change in cell dimensions. The intensity drop off was not noticed at the time because the computer teletype had broken down and a little later the diffractometer seized up. There was no evidence of absorption changes to the crystal during data collection (based on intensity measurements of reference reflections) so North, Phillips and Mathews (1968) absorption curves, measured before data collection commenced, were used to correct the data for absorption. The data were scaled to the crosslinked native data set (see section 5.3) using the program ANSC. The low value of the mean fractional isomorphous difference was as expected from the small intensity changes observed between native and derivative precession photographs (figure 4.21). There appeared to be a slight increase in mean fractional isomorphous difference with resolution (figure 4.22) which may have been due to the data collection problems discussed above or indicative of small scale conformational changes to the protein caused by removing the metal.

Isomorphous difference Pattersons, with coefficients $(F_{Mn} - F_{Mn\ free})^2$, were calculated to 8 and 6Å resolution. The Harker sections of both maps are presented in figure 4.23. (The full three-dimensional Patterson map is presented in appendix one). The higher resolution map was less noisy than the lower resolution map and so the decrease in isomorphism of the suspected maganese-free derivative with resolution did not appear to greatly effect the Patterson map. The interpretation of the maps was complicated by the presence of a peak on the edge of two Harker sections (labelled A in figure 4.23) and the absence of any significant features on non-Harker sections. Attempts to refine possible solutions to the Patterson

a



b

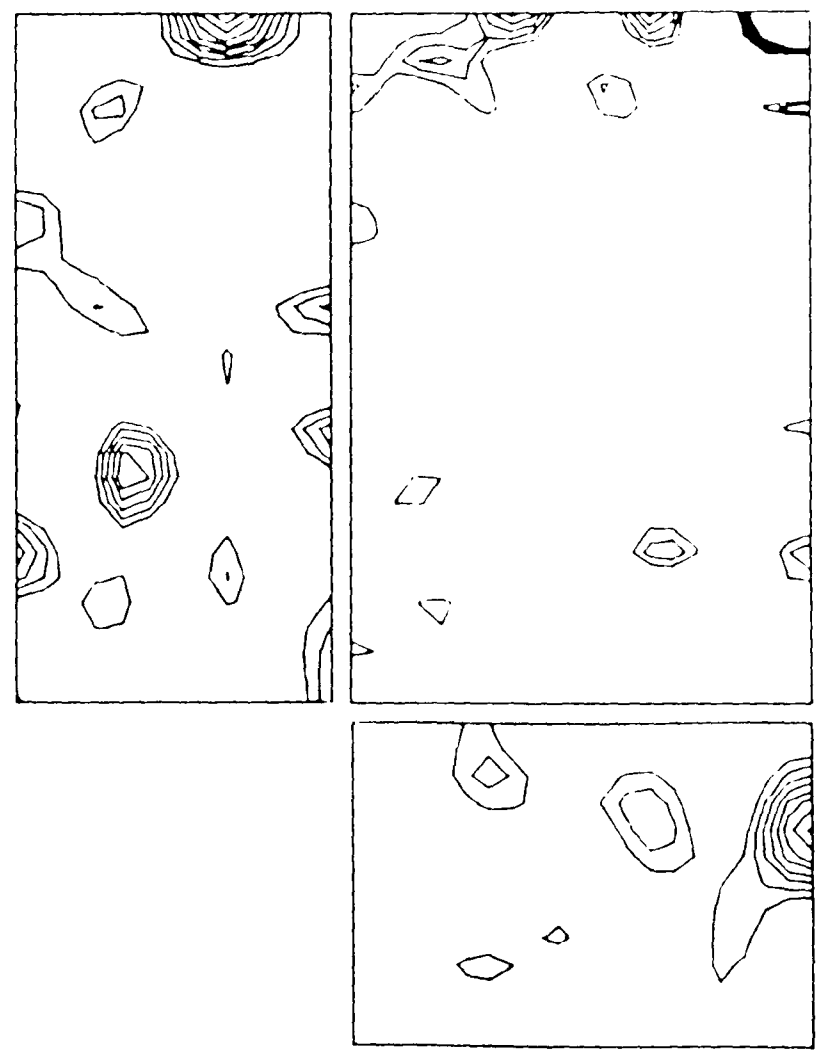


Figure 4.23 Harker sections (u=1/2, w=0 and v=1/2 from left going in a clockwise direction) of the difference Pattersons for the suspected manganese-free data set for terms to (a) 6Å and (b) 8Å spacings. The maps were contoured at equal but arbitrary intervals with the first level omitted.

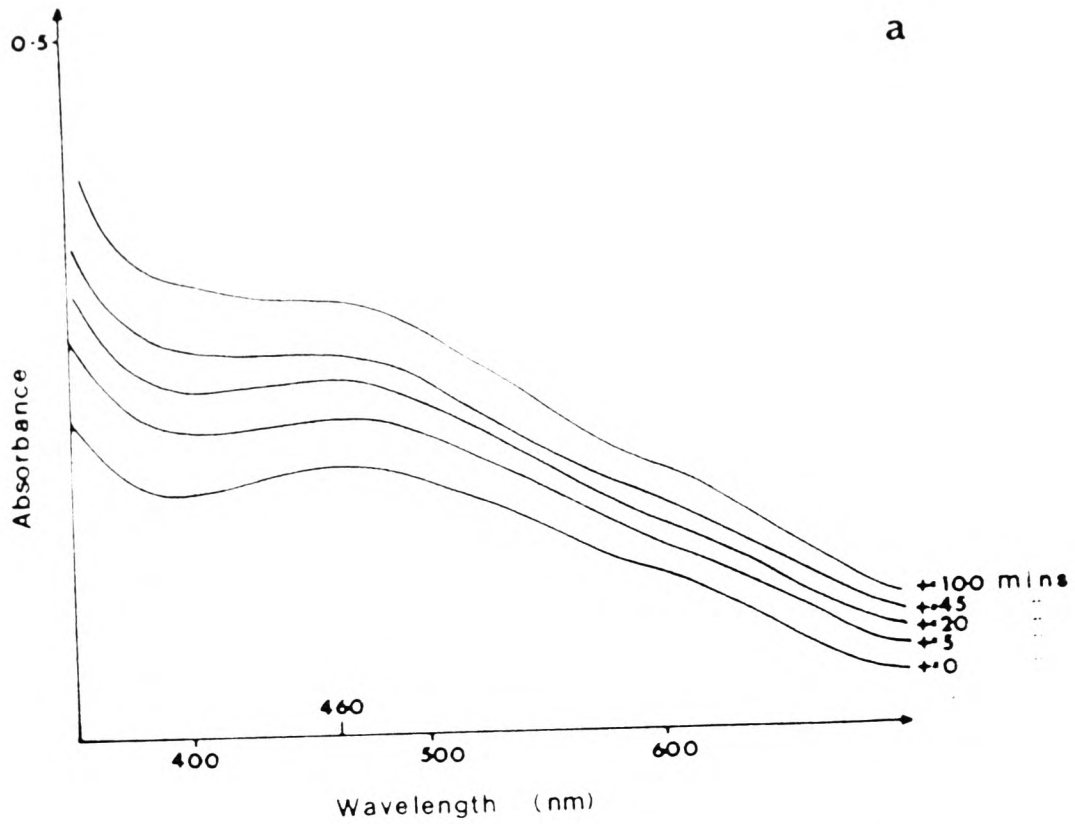
Key: see key to figure 4.20.

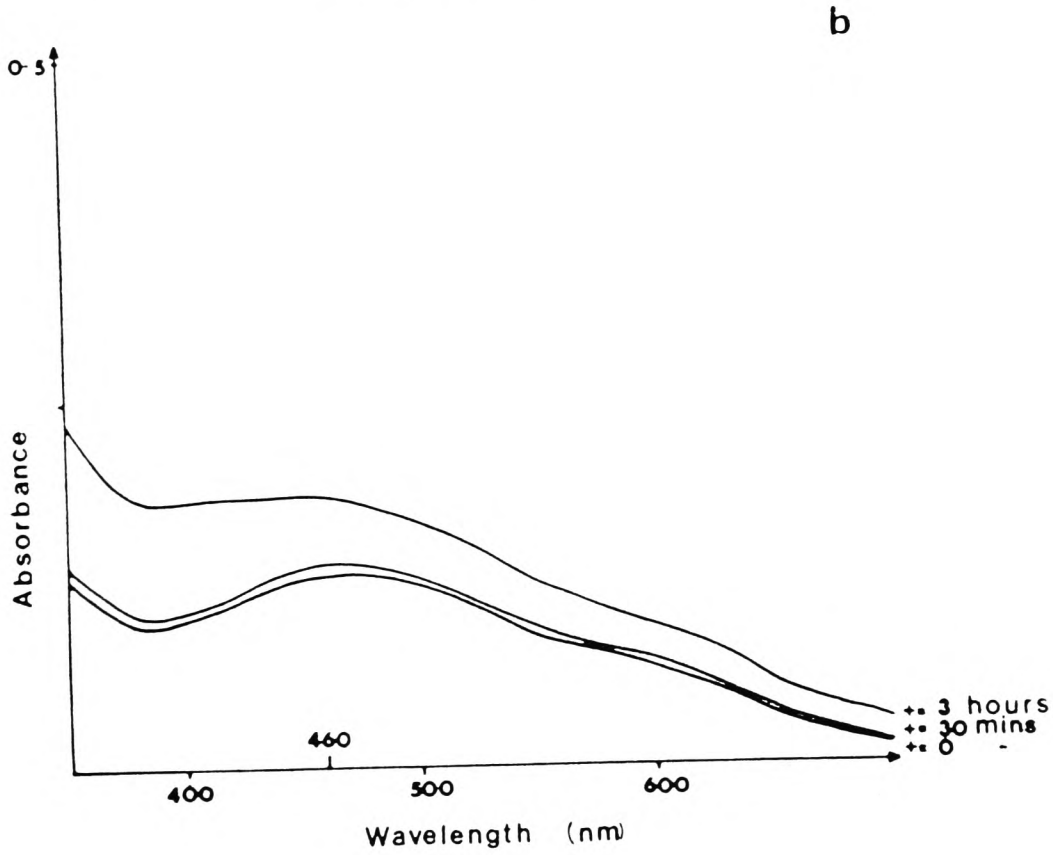
were inconclusive because almost any combination of sites yielded a low centric R-factor ($= \sum_{hkl} |F_{obs} - F_{calc.}| / \sum_{hkl} F_{obs}$) of between 45 and 50%. This was presumably because the very small intensity changes between the native and derivative data were readily modelled by a variety of solutions. However it was encouraging to note the similarities between these maps and the anomalous difference Patterson maps (figure 4.23) discussed in the previous section. It would appear that the manganese atoms were removed but the size of the resultant intensity changes were too small to produce interpretable Patterson maps.

4.6.3 Metal Replacement Experiments

The next step in the procedure was to attempt to prepare heavy atom derivatives of the suspected manganese-free crystals. It was hoped that isomorphous difference Patterson maps, calculated from diffraction data of metal-substituted SODs, might be easier to interpret than the noisy anomalous and isomorphous difference Patterson maps calculated to date. There was a number of published reports concerning the substitution of the manganese in MnSOD by other light metal atoms (see chapter one). No observations had been reported of heavy metal ions replacing the manganese. However, there was some crystallographic evidence that samarium and lead might bind close to or at the manganese-binding sites of the protein (see chapter five). These observations led to some collaborative work with Dr. J. Bannister of the Inorganic Chemistry Department.

It was decided that metal binding to the manganese sites of the protein was best followed by solution studies and a spectrophotometer. MnSOD has a visible absorption band at a wavelength of about 460 to 470nm due to d-d transistions from the manganese ion. Thus changes to the manganese binding site are readily noted by observing changes to this absorption band. The first experiments were designed to replace the manganese by competitive binding of other metal reagents as follows. A protein solution of 9mg/ml





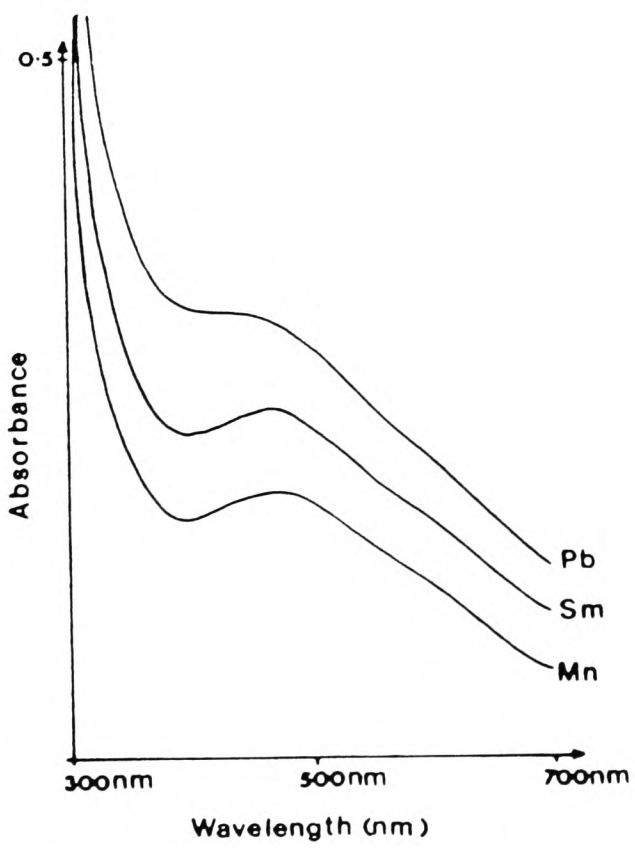


Figure 4.24 Absorption profiles of MnSOD solutions containing (a) lead or (b) samarium. The result after dialysing both solutions for twenty four hours is shown in (c). See text.

was equilibrated with 50 mM acetate buffer, pH 5.5, by dialysis for a week. Acetate salts of samarium and lead were made up to 5 mM in the same buffer. The absorption spectrum of the native protein was measured and microlitre quantities of the metal solutions were added slowly until an equimolar quantity had been added (one metal ion per protein molecule). The results of the additions are shown in figure 4.24. The addition of either metal caused a flattening of the absorption band on the short wavelength side of the absorption maximum with an overall increase in the height of the absorption band due to some protein coming out of solution. Both metal-containing solutions were incubated for twelve hours at 4 °C and then dialysed for twenty four hours to remove excess metal. The absorption profile of the protein solution that had samarium originally added to it, had reverted back to the native profile. The lead-added solution retained its flattened profile (see figure 4.24).

There was evidence from the colour of the solutions and of the metal-soaked crystals (see chapter five) as well as the appearance of the visible absorption band that some manganese was being retained by the protein. Hence the experiments were repeated with the lead solutions but this time the metal salt was added to a solution of apoprotein. The apoprotein was prepared by exhaustive dialysis against 10 mM EDTA in the presence of 8 M urea. The pH of the buffer was adjusted to pH 3.8 with 1 M acetic acid. Excess EDTA was removed by dialysis against 8 M urea at pH 5.5. The apoprotein was found to be unstable in the absence of urea. The lead-substituted protein was prepared by dialysis against 50 mM acetate, pH 5.0, containing urea and 10mM metal salt. Excess metal was removed by exhaustive dialysis against 1 mM urea. As a control, cobalt salt was added to one portion of apoprotein and found to bind, in agreement with the literature (Brock et al., 1976). To another portion, lead acetate was added. However, no lead was found to be bound to the protein, on the evidence of atomic absorption spectroscopy. This work suggests that lead does not bind very

strongly to the protein.

In hindsight, there appears to be a correlation between the size of the metal ion and its ability to bind to SOD. The ionic radii of the manganous and manganic ions are 0.80 and 0.66Å, respectively. The only metal ion, besides the manganous ion, that has been found to bind to the apoprotein and remain a stable metalloprotein in the absence of urea is cobalt which has an ionic radius of 0.63Å in its tervalent state. The reconstitution of apoprotein with manganous ion can only be done in the presence of urea. If the ionic radius is an important determining factor in metal binding, then it must be very sensitive. Apoprotein reconstituted with copper is only stable in the presence of urea (J. Bannister, pers. commun.). The cupric ionic radius is 0.72Å. However, the copper ion cannot take up a tervalent oxidation state, which may be a contributing factor towards the copper protein's instability. The ionic radii of samarium (III), lead (II) and lead (IV) are 0.96, 1.20 and 0.84Å respectively, and so are too big to fit the size criterion.

A final solution experiment was performed to see if the concentration of the manganese in the protein could be increased. The crystal structures of FeSOD clearly show that there are two metal binding sites per dimer (Ringe et al., 1983; Stallings et al., 1983) but the metal binding stoichiometry of Fe and MnSODs is usually about one metal per dimer. It has been postulated that some metal is lost during the protein purification steps (Williams, 1982). An increase in manganese content of the protein might improve the quality of the metal-free and anomalous difference Patterson maps discussed earlier in this chapter. The apoprotein was made, as described above, and excess manganese salt added to the solution. Atomic absorption spectroscopy indicated there was only one metal atom binding per dimer molecule. This suggests that the dimer displays negative cooperativity towards the binding of a second manganese atom.

4.7 Conclusions

This chapter has described attempts to solve the structure of MnSOD by molecular replacement techniques. A solution was obtained for the orientation of the unknown structure in its unit cell but its position remained unsolved. The failure of the translation function to work may have been due to the poor quality of the model, an incorrect solution to the rotation function or problems inherent with the method used.

	a	b	c
Crosslinked in PEG	73.0	111.3	51.1
Crosslinked and transferred to water	74.0	111.5	51.2
	74.5	-	51.5
	74.4	-	51.4
	74.7	-	51.6
	74.1	-	51.6
	74.0	-	51.2
Crosslinked for 25 days and then transferred to water	74.0	111.3	51.1

Table 5.1 Cell dimensions (Å) of crosslinked MnSOD crystals, measured from μ=9° zonal precession photographs.

CHAPTER 5

The Phase Problem - Isomorphous Replacement

5.1 Crosslinking

The difficulties met in trying to solve the translation function, discussed in the previous chapter, led to a reappraisal of isomorphous replacement as the method of choice for trying to solve the phase problem. As discussed in chapter three, the search for heavy atom derivatives was complicated by cell dimension variability. Petsko and Ringe (pers. commun.) had problems preparing heavy atom derivatives of FeSOD but solved them by crosslinking their crystals. It was decided to explore this approach with our crystals.

Crosslinking crystals is not without its disadvantages. It may reduce the number of heavy atom sites available and bring about changes in the conformation of the native protein. It was felt that the advantages outweighed the possible disadvantages. It was hoped that the cell dimensions of the crystals could be stabilised by crosslinking and, if the crystals could be transferred to water without dissolving, they would be much easier to handle and a greater range of heavy atom derivatives could be explored.

The crystals were lightly crosslinked by immersing them in 0.2% glutaraldehyde (made up in buffered 25% PEG solution) overnight at 4 $^{\circ}$C. The crosslinked crystals were found to diffract to the same resolution as native crystals and their cell dimensions gave consistent values with a maximum deviation about the average values of $\pm$0.4$\mathring{A}$ (after transferring to water, see table 5.1). The crystals could be transferred to water without any apparent ill effects as judged by their diffracting properties. They were stable to low pH and were found to be mechanically stronger than native crystals. The diffraction patterns of the crosslinked and native

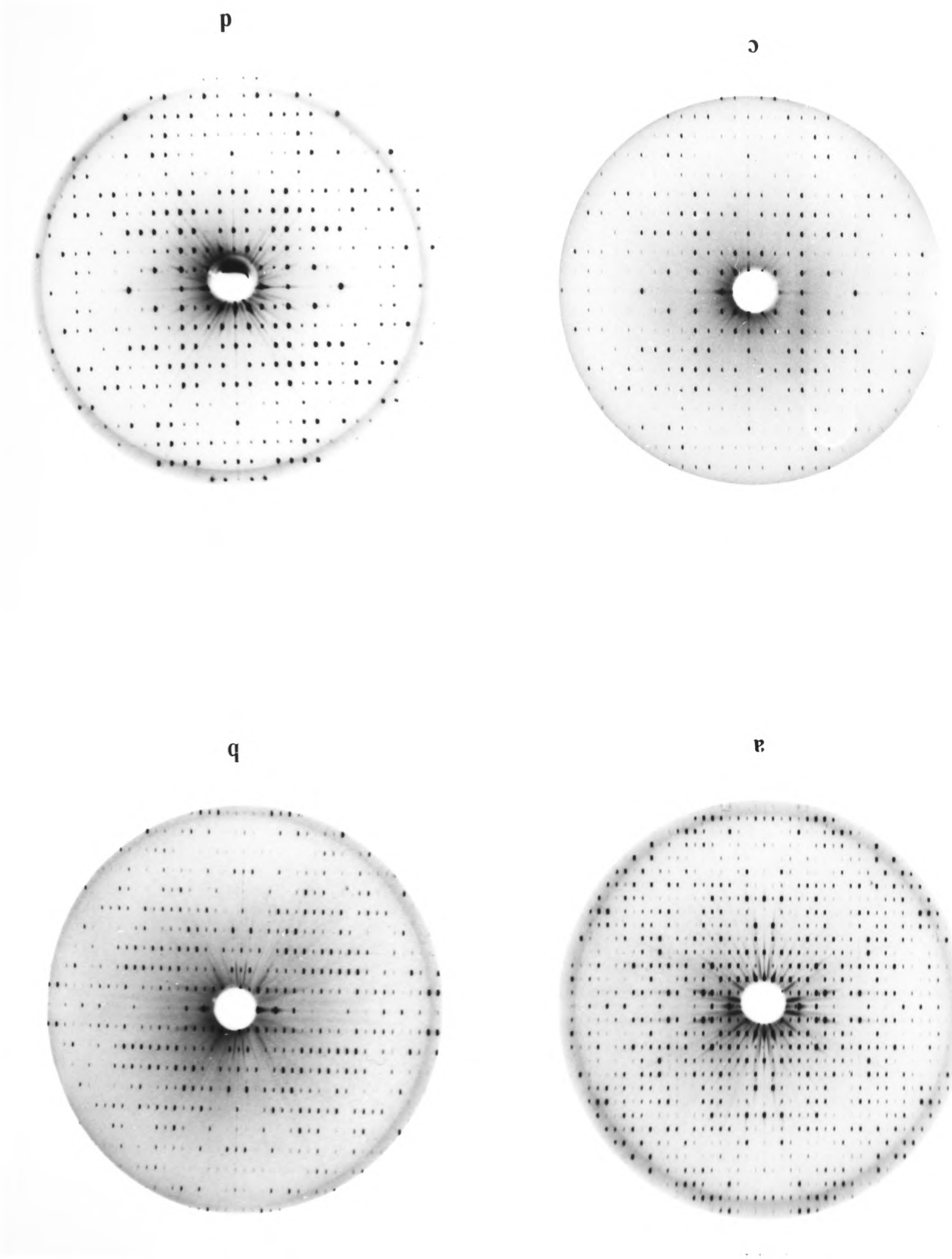


Figure 5.1 Precession photographs (μ=9°, crystal-to-film distance=100mm) of the three principal centric zones from crosslinked MnSOD crystals, (a) hk0, (b) 0kl, (c) h0l and (d) h0l after crosslinking crystals for 25 days.

type 2 crystals were similar indicating that there were no large conformational changes to the ordered parts of the crystalline protein by the treatment (compare figures 3.1 and 5.1). Unfortunately, more than 50% of crystals that were crosslinked were found useless for further study because of cracking. This failure rate was on the same scale to the loss encountered on handling non-crosslinked crystals and so it was assumed that the loss was mainly due to manipulation prior to the crosslinking treatment. A small percentage of crystals had to be rejected due an increase in diffraction spot mosaicity.

An experiment was performed to see what effect the crosslinking reagent had on the crystals over longer soak times. A batch of crystals were soaked in the crosslinking solution for 25 days at 4°C. The crystals still diffracted well and their cell dimensions were consistent with other crosslinked crystals (see table 5.1). But the diffraction pattern showed significant changes compared with the diffraction pattern from crosslinked crystals soaked for shorter times (figure 5.1 c and d). These changes might have been indicative of significant conformational changes to the protein, so long crosslinking soak times were considered undesirable.

5.2 Prepartion of Heavy Atom Derivatives

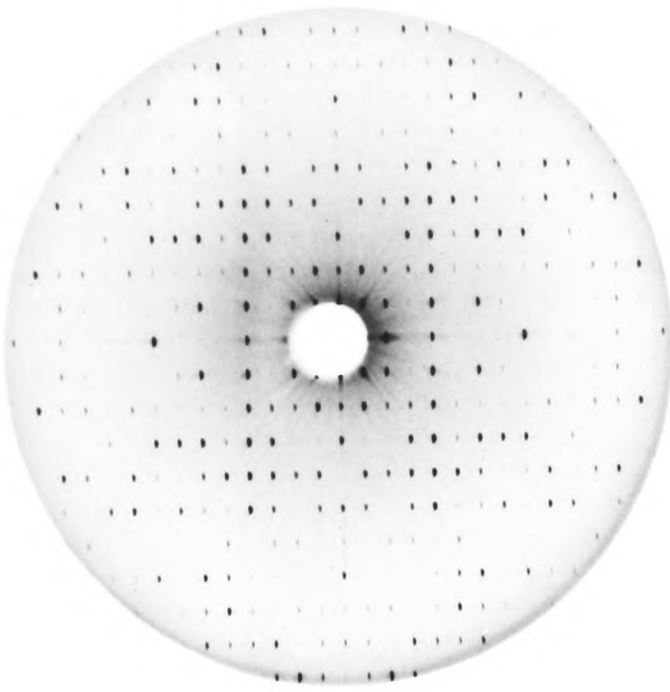
A summary of the heavy atom search is presented in table 5.2. The table has been divided into derivative types; first come the 'hard' cations, then those heavy atom reagents that tend to bind covalently to proteins and finally those reagents that tend to occupy hydrophobic crevices in proteins (the boundaries between the groups are rather arbitrary). In the course of the search it became apparent that the crystals could still crack and crumble, despite the crosslinking treatment, so the soaking conditions were kept on the mild side. Thus crystals were soaked in dilute heavy atom solutions for short periods of time directly in their mounting tubes. Further, all crystals were scanned on an X-ray camera before soaking so as

Table 5.2 Conditions and observations in the search for heavy atom derivatives of crosslinked crystals.

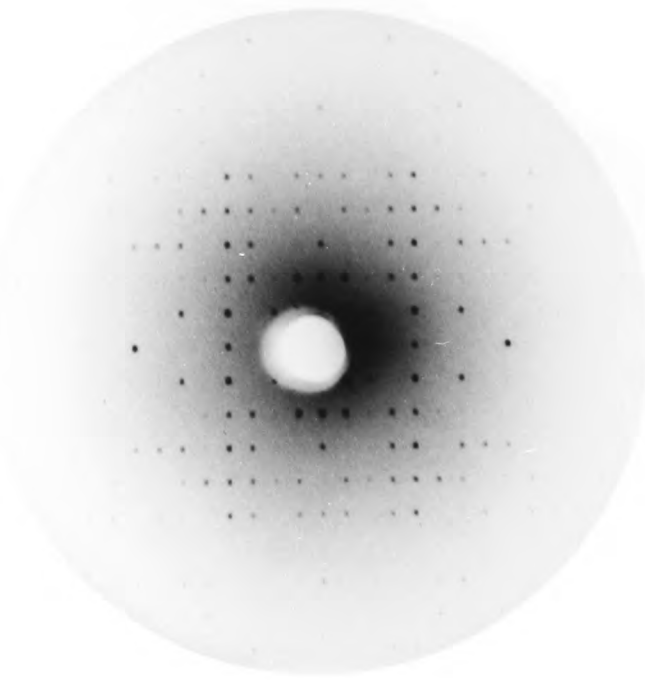
Heavy Atom Compound	Concentration (mM)	Soak Time	Cell Dimensions (Å) a	b	c	Observations
$UO_2(NO_3)_2$	19	24d				disordered
$UO_2(OAc)_2$	19	o				disordered
	10	o	73.6	-	51.1	large intensity changes, low angle only
	5	o	73.8	-	51.0	as for 10mM above
	1	o	73.8	-	51.3	large intensity changes over whole pattern
$K_3UO_2F_5$	20	o	73.8	111.4	-	mild intensity changes
$Pb(OAc)_2$	20	o	74.0	111.6	-	mild intensity changes
$Sm(OAc)_3$	100	o	73.8	-	51.3	mild intensity changes
	50	o	74.1	-	51.2	a few small intensity changes
			73.8	-	51.1	as above
	20	19d	74.6	112.1	51.3	as for 50mM above
$AgNO_3$	19	o	73.4	-	51.2	large intensity changes, disorder
K_2PtCl_6	9.3	o				crystal dark brown, disordered
	1	o	74.6	111.8	-	crystal dark brown but no intensity changes
$NaAuCl_4$	20	o				crystal brown, disordered
	9.9	9d				as for 20mM above
	2	o				crystal blackened at centre, disordered
	0.5	o	74.0	-	51.4	yellowish crystal, nice intensity changes
			74.3	111.6	51.4	purple-brown crystal, some disorder
K_2PtCl_4	20	o				disordered
	9.3	9d				black, disordered
	5	o	72.0	-	51.2	dark brown, large intensity changes
	2	o	72.4	111.6	-	large intensity changes
			73.7	-	51.3	as above
			74.5	-	51.2	few large intensity changes
	0.5	o	74.1	111.8	-	no intensity changes

K_2PtBr_4	21.9	o				disordered
	11	o				large intensity changes at low angle, disordered
	5.5	o	74.8	-	51.4	few small intensity changes, disordered
$PtCl_2(NH_3)_2$ (heating req.)	16.3	o	74.0	111.7	51.2	large intensity changes, disorder
$K_2Pt(CN)_4$	20	o				disordered
	15	o				no intensity changes, disorder
	10	8d				disordered
	8	8d	74.7	112.0	51.5	a few small intensity changes
	5,3,1	8d				high mosaicity
$KAu(CN)_2$	20.7					disordered
	10	o	74.7	112.4	-	disordered
			73.4	-	51.3	many small intensity changes, disordered
$Hg(OAc)_2$	20	o				disordered
	10	o				disordered
	5	o	73.8		51.3	large intensity changes at low angle only
K_2HgI_4	20	o				crystal split into planes
	10.2	8d				yellow crystal, cracks, disordered
	5	o	74.2		51.3	large intensity changes at low angle only
EMP	20	o				disordered
	10	2d				disordered
	5	o	74.7	111.2	-	large intensity changes at low angle but diffracts to 7Å only
	1	o	74.2	111.9	-	mild intensity changes
PCMBS	18.6	o				disordered
	9.3	o	75.2	112.5	-	large intensity changes at low angle, disordered
	2	o	75.0	112.6	-	mild intensity changes
SALY (xs NaCl req.)	20	o				disordered
	14	o	75.5	112.6	-	many small intensity changes

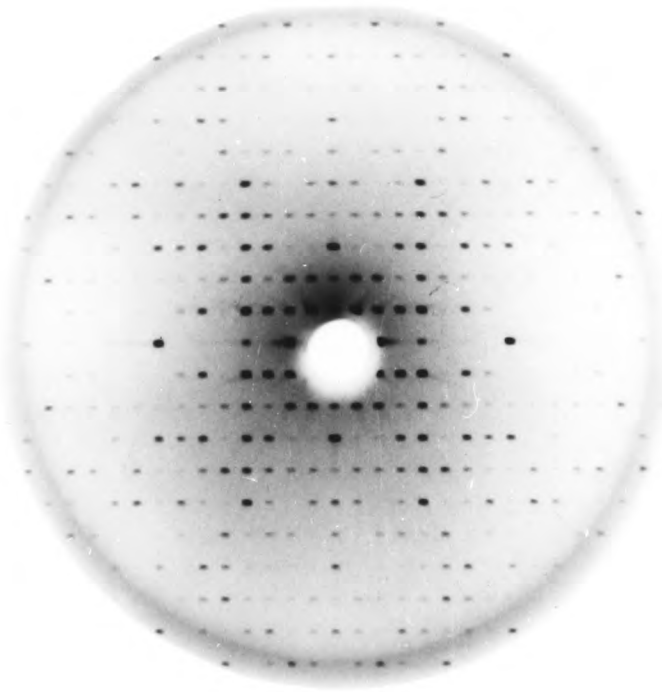
Key: d=days, o=overnight soak.



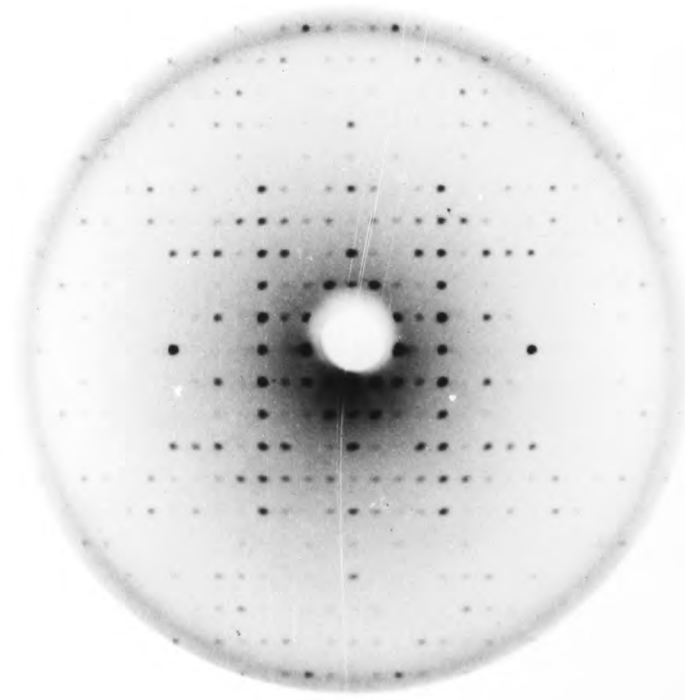
a



b

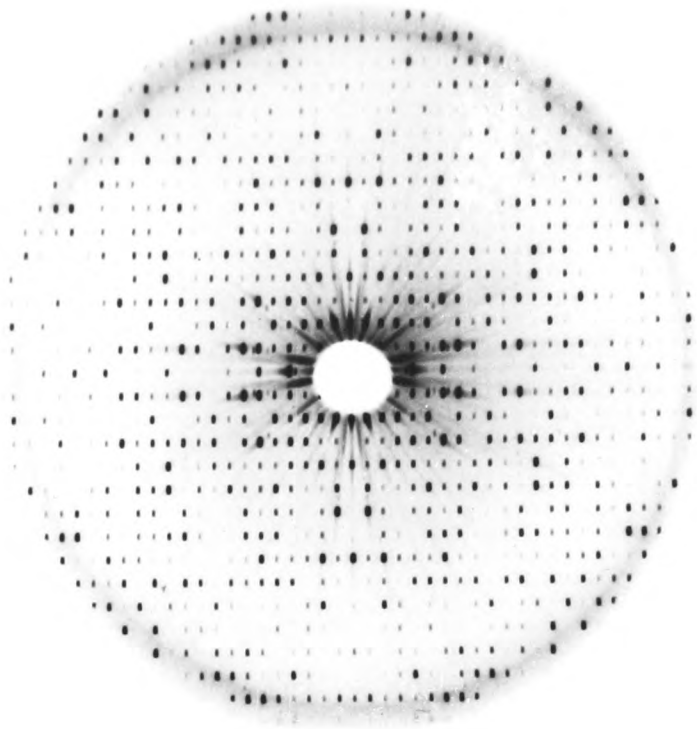


c

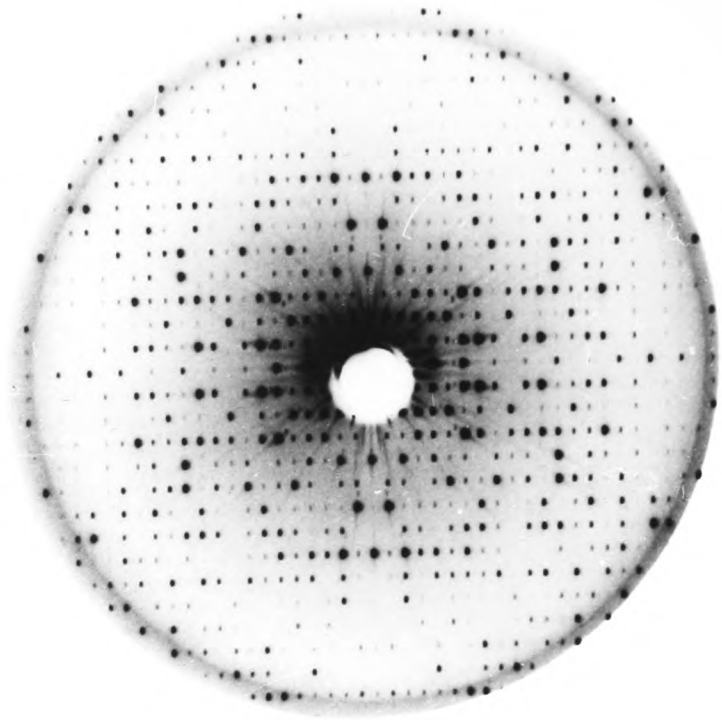


d

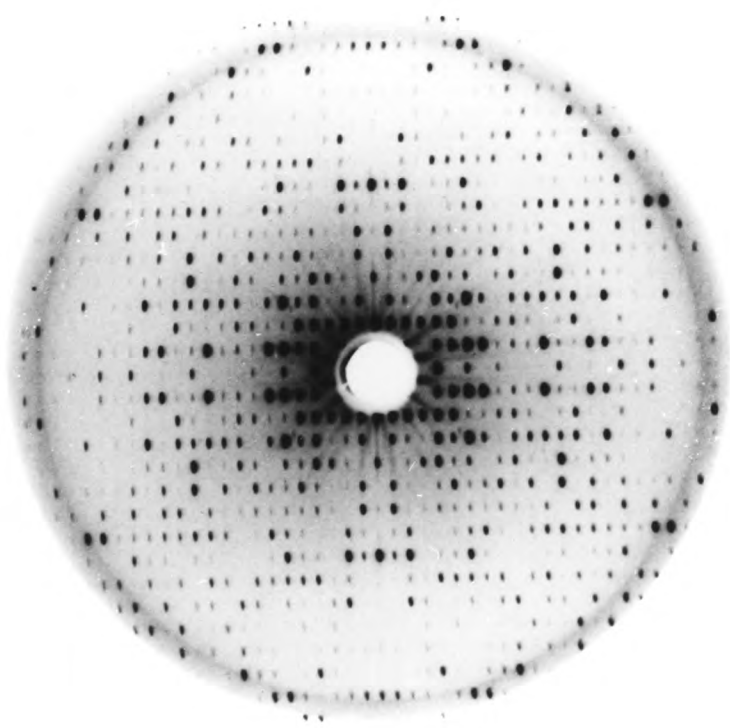
Figure 5.2 Precession photographs (μ=9°, crystal-to-film distance=100mm) of h0l zones measured from heavy atom-soaked crosslinked crystals, (a) native, (b) 0.5mM $AuNaCl_4$, (c) 1mM $UO_2(OAc)_2$ and (d) 100mM $Sm(OAc)_3$.



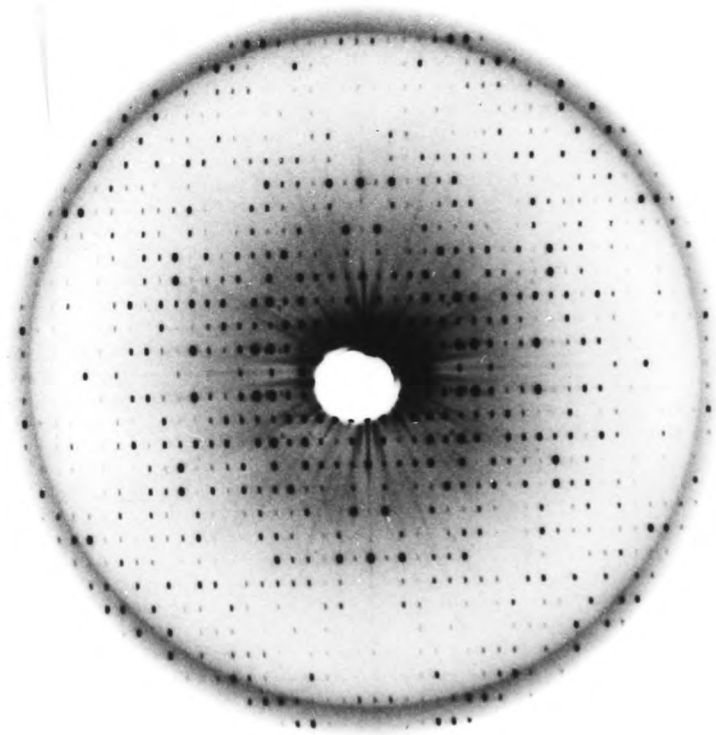
a



b



c



d

Figure 5.3 Precession photographs (μ=9°, crystal-to-film distance=100mm) of hk0 zones measured from heavy atom-soaked crosslinked crystals, (a) native, (b) 1mM EMP, (c) 20mM $K_3UO_2F_5$ and (d) 20mM $Pb(OAc)_2$.

not to waste time with substandard crystals. (It was common to find crystals which appeared single under the microscope but gave split spots in photographic stills). Generally no buffer was used so as to avoid reaction with or precipititation of heavy atom reagents, with the exception of reagents that might bind to histidine residues (where the optimal pH for binding is normally above 6.5). In these cases BES buffer set to between 6.5 to 7.0 pH units was used. (BES has been found to be unreactive to most heavy atom reagents (L. Johnson, pers. commun.)).

The majority of potentially useful derivatives were the 'hard' cation type of reagents. The other types of reagents tended to disorder the crystals. Large organomercurials resulted in significant cell dimension changes on the a axis. Perhaps the potential softer ligands of the protein were involved in manganese and glutaraldehyde binding and thus unavailable for heavy atom binding. Most of the heavy atom derivatives that proved useful in the structure determination of the FeSODs (ie. $K_2Pt(CN)_4$, EMP, $AgNO_3$, $Hg_2(OAc)_2$, $K_2Pt(NO_2)_4$) were tested (Ringe et al., 1983; Stallings et al., 1983). Only the EMP derivative gave promising intensity changes and thus pursued further. Precession photographs of the most promising derivatives are shown in figures 5.2 and 5.3.

5.3 Data Collection and Processing

All low resolution data sets were collected on the five circle diffractometer, principally because of its speed of data collection. The need to scan a large number of crystals for quality required the use of a precession camera for about a month before data collection began. In all, over one hundred and fifty crystals were scanned with only five or six crystals judged suitable for data collection. These crystals were soaked overnight in freshly made up heavy atom solutions as described in the previous section.

Data collection statistics have been compiled in table 5.3. Because of

Crystal	Concentration (mM)	Cell Dimensions (Å) a	b	c	Mount	Counter Mode	Scan Width (steps)	Peak Width (steps)	Main/Blind Region	Count Time (sec/step)	Number of Reflections	Intensities >3 sd
native no.1	-	74.8	112.1	51.3	c*	1-c	20	12	M	2.0	1240	70
native no.2	-	74.3	111.5	51.3	c*	1-c	24	14	M	1.0	1232	81
$UO_2(OAc)_2$	1	74.2	111.5	51.3	a*	5-c	32	12	M	10.0	2817	77
						1-c	20	12	B	10.0		
						1-c	32	12	M	10.0		
$AuNaCl_4$	0.5	74.2	111.7	51.4	a*	5-c	36	16	M	2.5	2805	70
						1-c	26	18	B	2.5		
						1-c	28	20	M	2.0		
$Sm(OAc)_3$	100	73.9	111.4	51.2	a*	5-c	32	12	M	3.5	2764	66
						1-c	20	12	B	3.5		
						1-c	22	14	M	3.0		
EMP	1	74.5	111.9	51.6	b*	1-c	20	12	M	3.0	2849	60
$K_3UO_2F_5$	20	73.7	111.5	51.3	a*	1-c	20	14	M	8.0	1147	

Key: 1-c = single counter geometry

5-c = five counter (optimal geometry)

sd = standard deviations

Notes: Step size was 0.04° and $\Delta\omega$ was 0.32° for all data.

Table 5.3 Data collection parameters and statistics for diffractometer data sets.

Crystal	Number of Reflections	Maximum Absorption	Radiation Damage (%)	Number of Independent Reflections	Number of Rejected Reflections	Number of Negative Intensities	Rmerge	MFID
native 1	1240	1.90	0	1176	0	72		
native 2	1232	1.41	0	1176	0			
$UO_2(OAc)_2$	2817	1.87	27	1205	1	50	4.3	0.18
$AuNaCl_4$	2805	1.87	13	1207	7	106	5.1	0.13
$Sm(OAc)_3$	2764	2.34	7	1204	9	79	4.9	0.16
EMP	2849	2.60	0	1219	0	199	5.9	0.21
$K_3UO_2F_5$	1147	2.32	15	1114	0	68	9.2	0.28

Key: $R_{merge} = 100 \sum_{hkl} \sum_{i=1}^{N} |I_i - \bar{I}| \; / \; \sum_{hkl} N\bar{I}$

where I is the value of the ith of N measurements of an intensity and

$\bar{I}$ is the mean of all N measurements of that intensity.

MFID = mean fractional isomorphous difference

$$= \Sigma |F_{PH} - F_P| \; / \; \Sigma \, F_P \; .$$

Table 5.4 Intensity statistics for native and heavy atom derivative data sets collected on diffractometer.

Derivative	Number of Friedel Pairs	Ratio of Deltas	Number of Delta minus	Number of Delta plus	Rfriedel	Kemp	Number of Delta >5sd	Rejectio Outliers
$UO_2(OAc)_2$	808	0.99-1.04	336	426	5.3	4.5	3	[illegible]
$AuNaCl_4$	772	0.99-1.01	383	374	5.2	2.6	[illegible]	[illegible]
$Sm(OAc)_3$	778	0.99-1.05	396	361	5.8	3.4	[illegible]	[illegible]
EMP	769	0.98-1.09	416	397	6.8	3.4	7	[illegible]

Key: Delta = $F_{PH}^{+} - F_{PH}^{-}$

Delta minus are those deltas less than zero and Delta plus those greater than zero

Rfriedel = $100 \sum |F_{PH}^{+} - F_{PH}^{-}| / \sum F_{PH}^{-}$

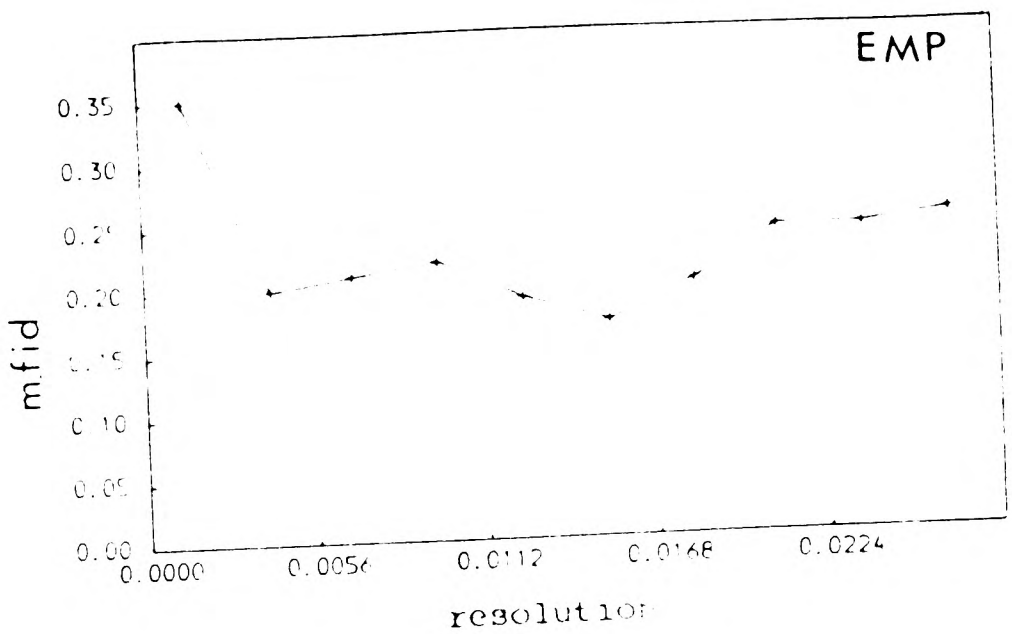
$K_{emp} = 2 \sum |F_{PH} - F_{P}|' \sum |F_{PH}^{+} - F_{PH}^{-}|$

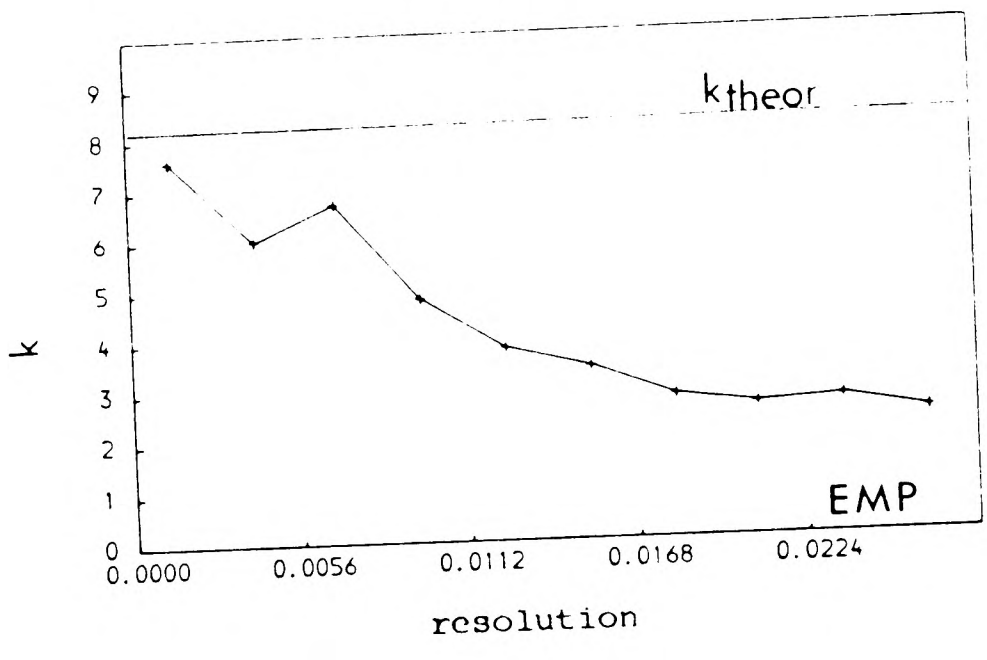
Rejection outliers - see text

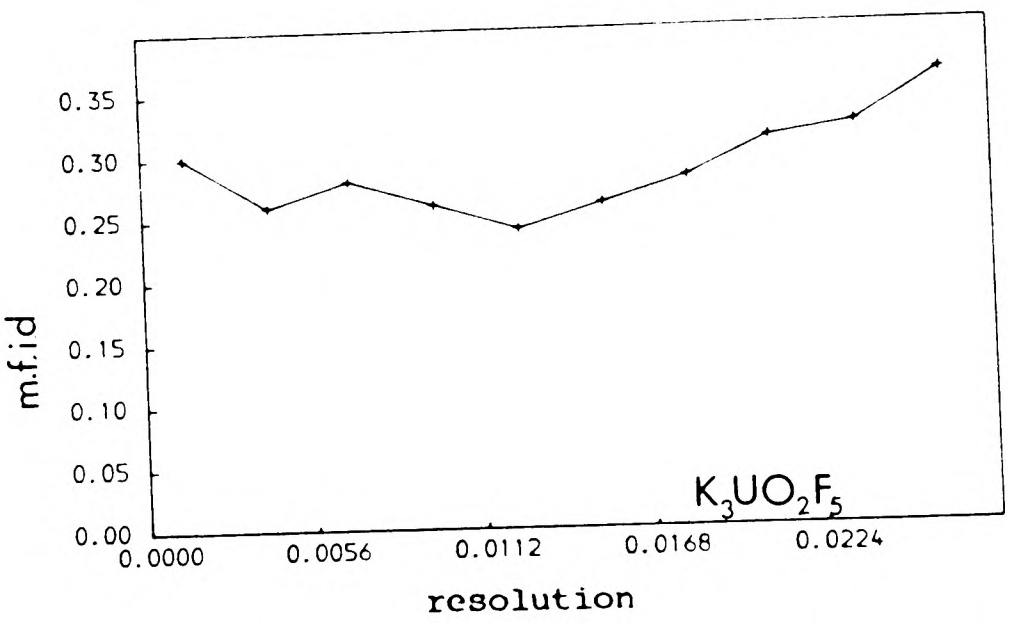
Table 5.5 Data statistics for the anomalous measurements in the heavy atom derivative data sets.

crystal fragility problems, no attempts were made to remount crystals with an optimum a* mount for data collection. The cell dimension values showed some variability but nothing like the variations encontered with non-crosslinked crystals (see chapter three). For all data sets, standard reference reflections were measured after every hundred measurements to check for crystal slippage, radiation damage and absorption changes. Blocks of reflections were measured before and after resolution shells of data for later use in estimating the radiation damage to the crystal. North, Phillips and Mathews (1968) absorption curves were measured before and after data collection on suitable strong axial reflections. Unfortunately, due to crystal quality and limited access to the diffractometer, the first native, the EMP and $K_3UO_2F_5$ derivative data sets were measured with shorter count times than would have been ideal. Full details of the methods used in collection and processing have been described in chapter two. Pertinent details of the processing have been compiled in tables 5.4 and 5.5.

The uranyl fluoride data had to be typed into computer filestore by hand because of difficulties met on reading the data off magnetic tape. This meant that the data had to be processed in ordinate analysis form rather than the more usual background-peak-background form (see section 2.7.2). The uranyl fluoride anomalous information was thought to be unreliable because of the weak measurements and thus was not typed in . It was noted that the $NaAuCl_4$-soaked crystal had turned from a bright yellow colour to a deep purple colour during data collection which was indicative of the deposition of colloidal gold in the crystal. Both uranyl derivatives appeared to have suffered significantly more radiation damage than the other derivatives, but this maybe simply due to the longer period of time that each uranyl crystal was on the diffractometer and exposed to X-rays. The merging R- factors for equivalent reflections in overlapping resolution shells were good except for the uranyl fluoride data where the intensity measurements were weak. The anomalous measurements showed only small trends







Key: $k_{theor} = (f_o + f') / f''$

$$k_{emp} = 2 \sum |F_{PH} - F_P| \; / \; \sum |F_{PH}^{+} - F_{PH}^{-}|$$

MFID = mean fractional isomorphous difference

$$= \sum |F_{PH} - F_P| \; / \; \sum F_P$$

Figure 5.4 (cntd.) Plots of mean fractional isomorphous difference (top and bottom diagrams) and kemp (middle diagram) as a function of resolution ($4\sin^2\theta/\lambda^2$).

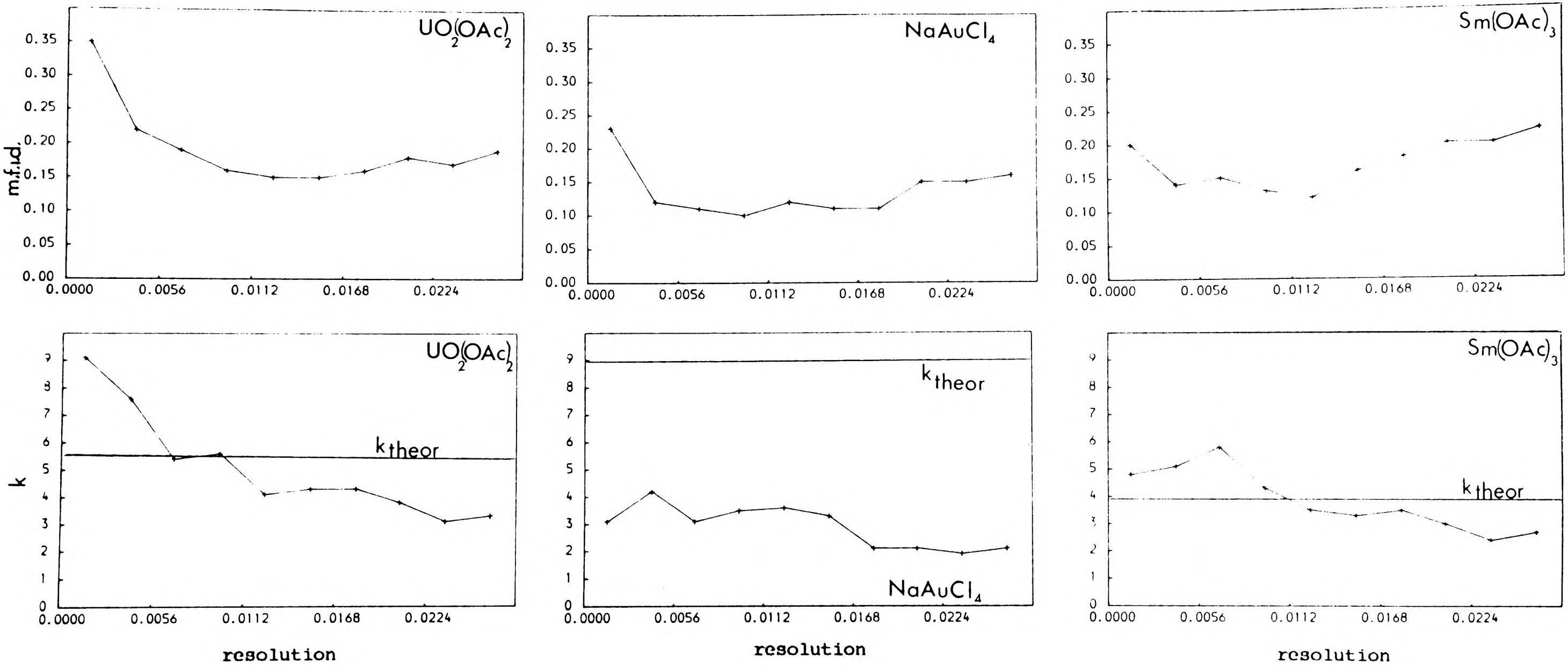


Figure 5.4 Plots of mean fractional isomorphous difference (top row) and kemp (bottom row) as a function of resolution ($4\sin^2\theta/\lambda^2$) for a number of heavy atom derivative data sets.

in positive and negative parts of reciprocal space but the trends were nevertheless corrected for, using the program FRIEDCORR (see appendix three). The Friedel R-factors were typical of values expected for reasonable anomalous measurements. The values of mean fractional isomorphous difference were of the order expected for useful isomorphous derivatives.

5.4 Data Analysis

All derivative data sets were scaled to the strong native data set (native no.2) using the program ANSC (see appendix three). All the data sets required only small secondary absorption corrections (on phi in reciprocal space) to make any small systematic trends on the Miller indices and resolution with respect to the native data set, vanish.

Figure 5.4 shows plots of the mean fractional isomorphous difference (MFID) and k_{emp} as a function of resolution for each derivative. The MFID plots were reasonably flat for the uranyl acetate and gold derivatives suggesting a good degree of isomorphism. (The large differences at low resolution were probably due to disordered heavy atom in the solvent channels of the crystal which greatly affect low order terms). The positive slopes in the MFID plots for the other derivatives are indicative of non-isomorphism. The anomalous differences were assessed by comparing the ratio between the observed isomorphous differences and the anomalous differences (k_{emp}) with the theoretical value for the ratio of the real to the imaginary part of the scattering factor ($k_{theor.}$), as a function of resolution. The better quality of measurement for the uranyl and samarium acetate derivatives was reflected in the closeness of the empirical and theoretical values. The increasing discrepancy of the theoretical and empirical values at higher resolution for all the plots reflected an overestimation of the anomalous differences, caused by random errors in the weaker higher angle data. The high values of k_{emp} at very low resolution

Crystal	Film in Pack	Number of Spots Measured	Number Saturated	Number Rejected	Number of Negative Intensities	Rintra-scale (%)	Number of Unique Reflections	Rmerge (%)	MFID
native	A	1129	113	38	49	4.2	238	2.2	
	B	814	15	23					
$Pb(OAc)_2$	A	1428	302	249	84	5.9	195	2.6	24.6
	B	975	84	192					
EMP	A	1239	105	103	74	4.5	232	4.6	24.7
	B	903	24	67					
$K_3UO_2F_5$	A	1129	71	73	80	4.8	234	5.5	30.8
	B	768	15	32					

Key: Rintrascale = $\sum_{hkl}\sum_{j=1,2} |\bar{I} - I_j| / \sum_{hkl}\sum_{j=1,2} \bar{I}$

where $\bar{I}$ is the weighted average value of the intensity of reflection hkl. The summation on j is over all films which hkl is observed and is performed only over reflections recorded on more than one film.

Rmerge = $\sum_{hkl}\sum_{i=1}^{N} |F_i - \bar{F}| / \sum \bar{F}$

MFID = mean fractional isomorphous change

$= 100 \sum |F_{PH} - F_P| / \sum F_P$

Note: spots were rejected in bad background areas of the film in the program, MOSCO.

Table 5.6 Intensity statistics for the precession film data.

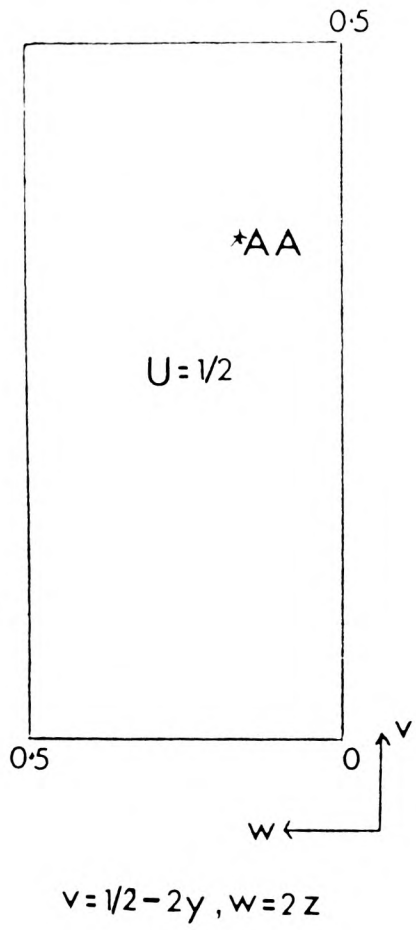
were partly due to problems of scaling the very low angle terms of the native and derivative data sets.

5.5 Precession Film Data

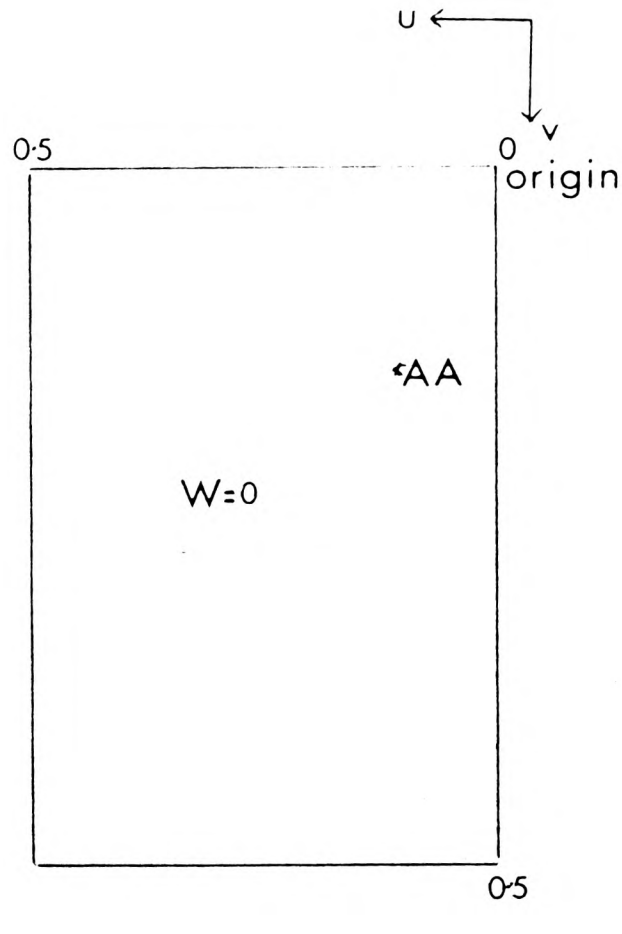
A number of precession films were scanned as an adjunct to the three dimensional diffractometer data sets. These data sets were to play an important role in the solutions of the isomorphous difference Pattersons calculated from the diffractometer data. The only films scanned were projection photographs of the most populated centric zone, the hk0 projections. These photographs are shown in figure 5.3. The data processing proceeded smoothly with the exception of the lead derivative where there were a large number of reflections rejected because of high background areas on the film. Details of the processing are compiled in table 5.6. The final merging R-factors were good and the mean fractional isomorphism differences compared favourably with those obtained from processing of the diffractometer data sets (see table 5.4).

5.6 Patterson Syntheses

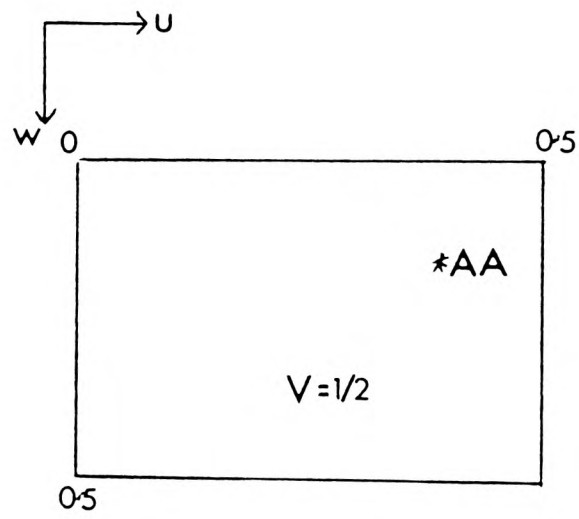
Isomorphous difference Patterson maps were calculated from syntheses with coefficients $(F_{PH}-F_P)^2$ and anomalous difference Pattersons from syntheses with coefficients $(F_{PH}^{+}-F_{PH}^{-})^2$. The maps were calculated using all terms to 6Å spacings and are presented in figure 5.5. Maps calculated with terms less than 20Å omitted, were very similar to maps with all terms included with one exception. The gold derivative map was almost featureless without the very low order terms, suggesting that the ''all terms'' Patterson map represented the effects of solvent movement or the deposition of disordered gold in the crystal. Combined difference maps with coefficients F_{HLE}^2 were very similar to the isomorphous difference maps. The anomalous difference maps were much noisier than their corresponding isomorphous difference maps and had few peaks in common, with the exception



$v = 1/2 - 2y$, $w = 2z$

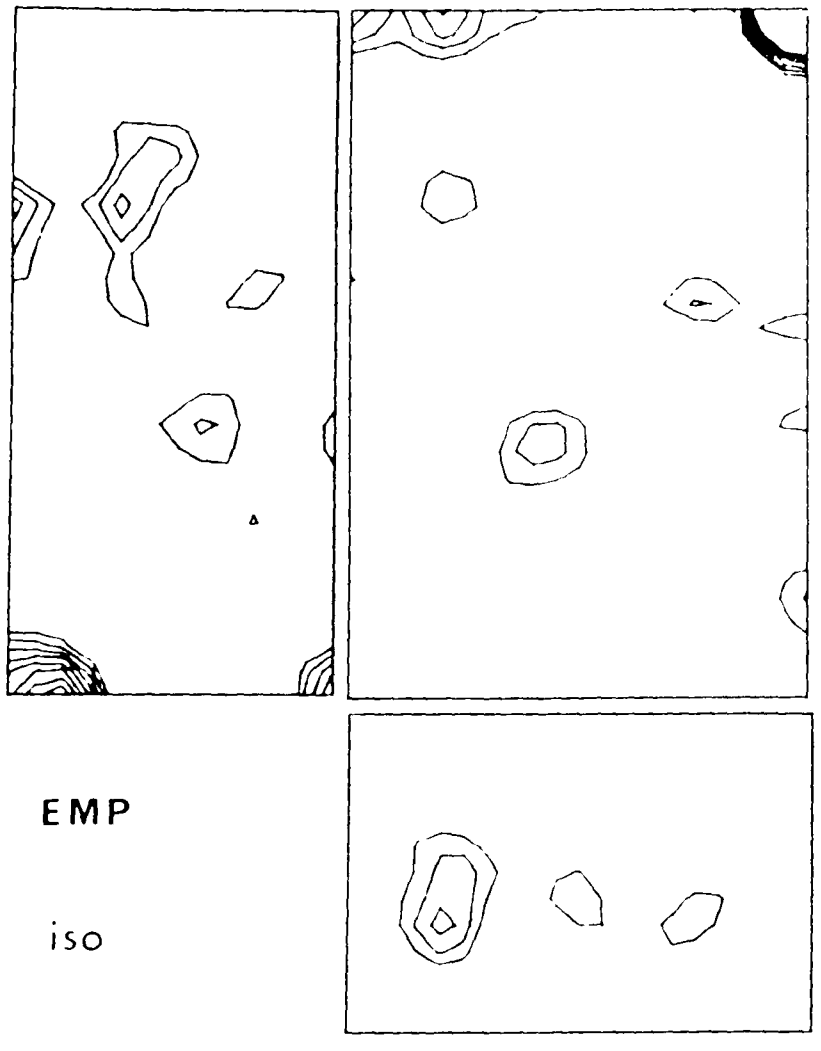


$u = 2x$, $v = 2y$



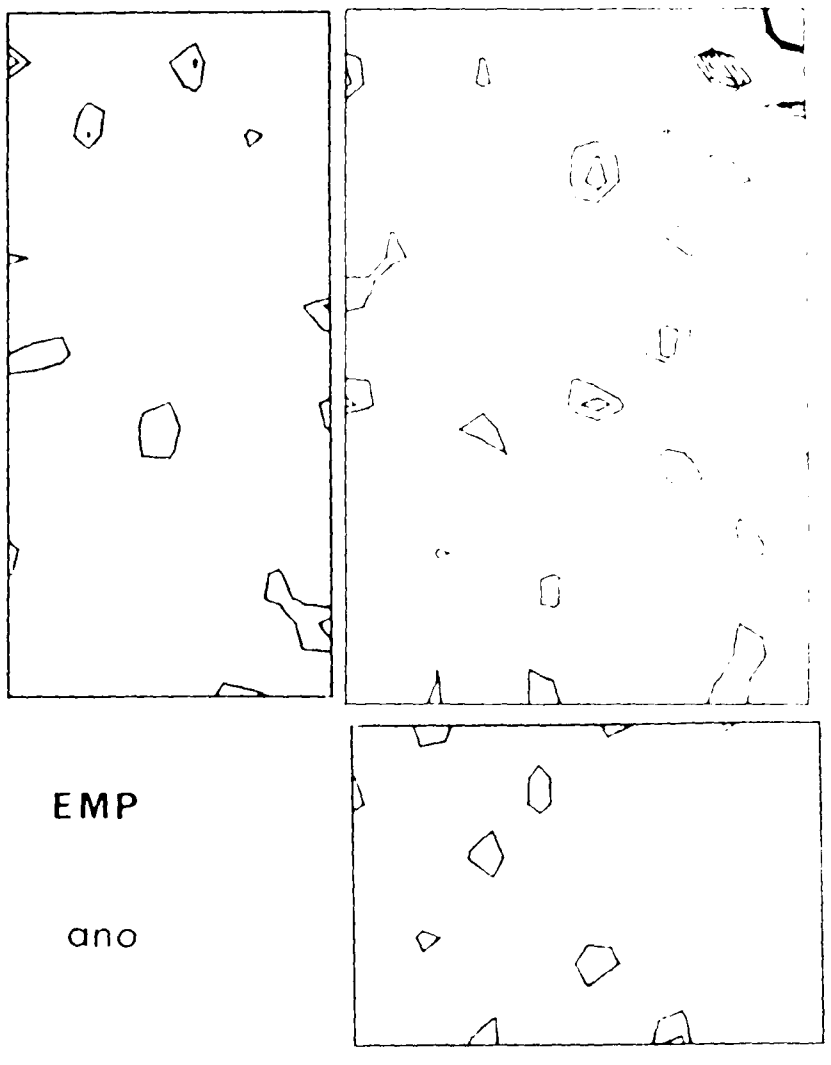
$u = 1/2 - 2x$, $w = 2z$

Key: Harker section layout. The position of self-vectors (AA) for some heavy atom site A is also shown.



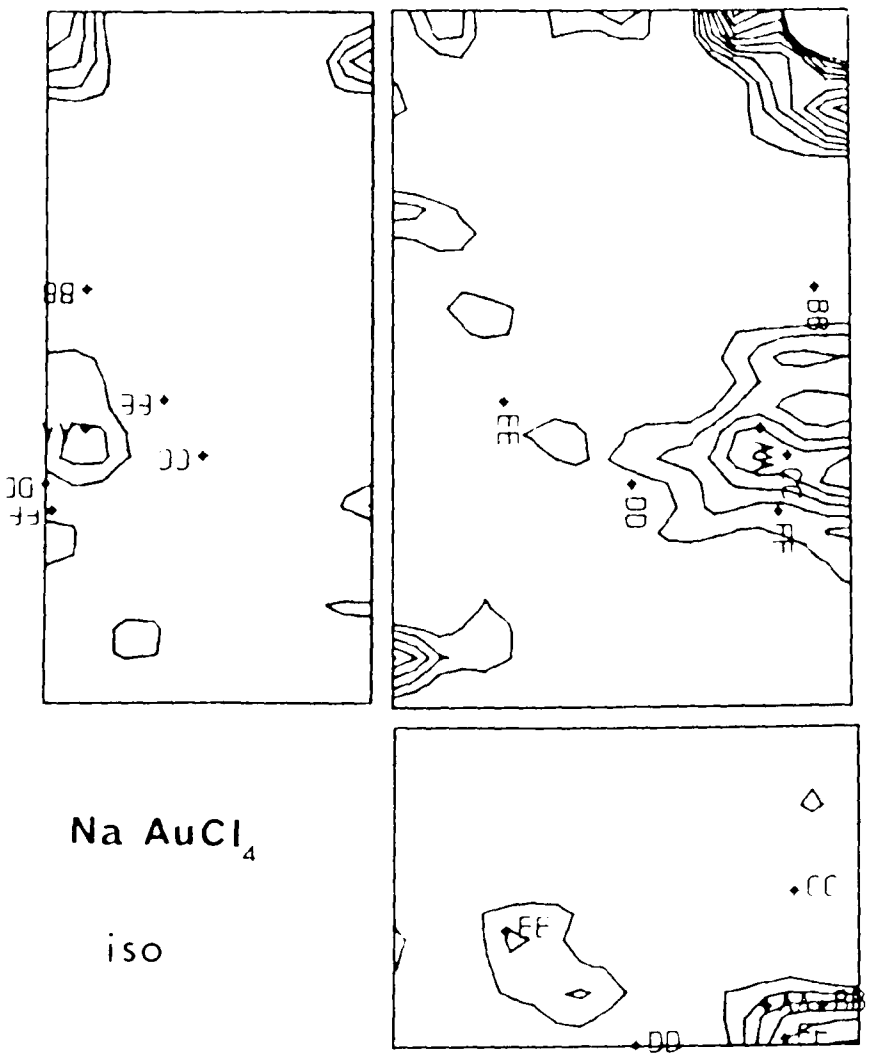
EMP

iso



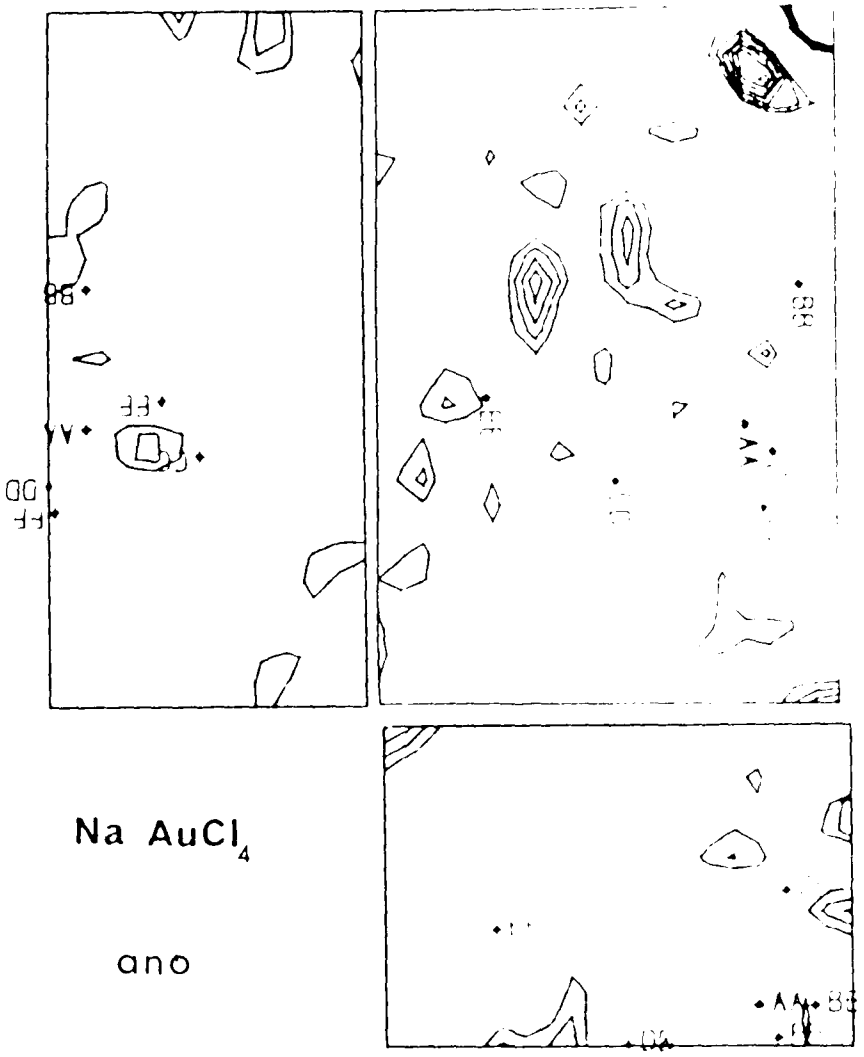
EMP

ano



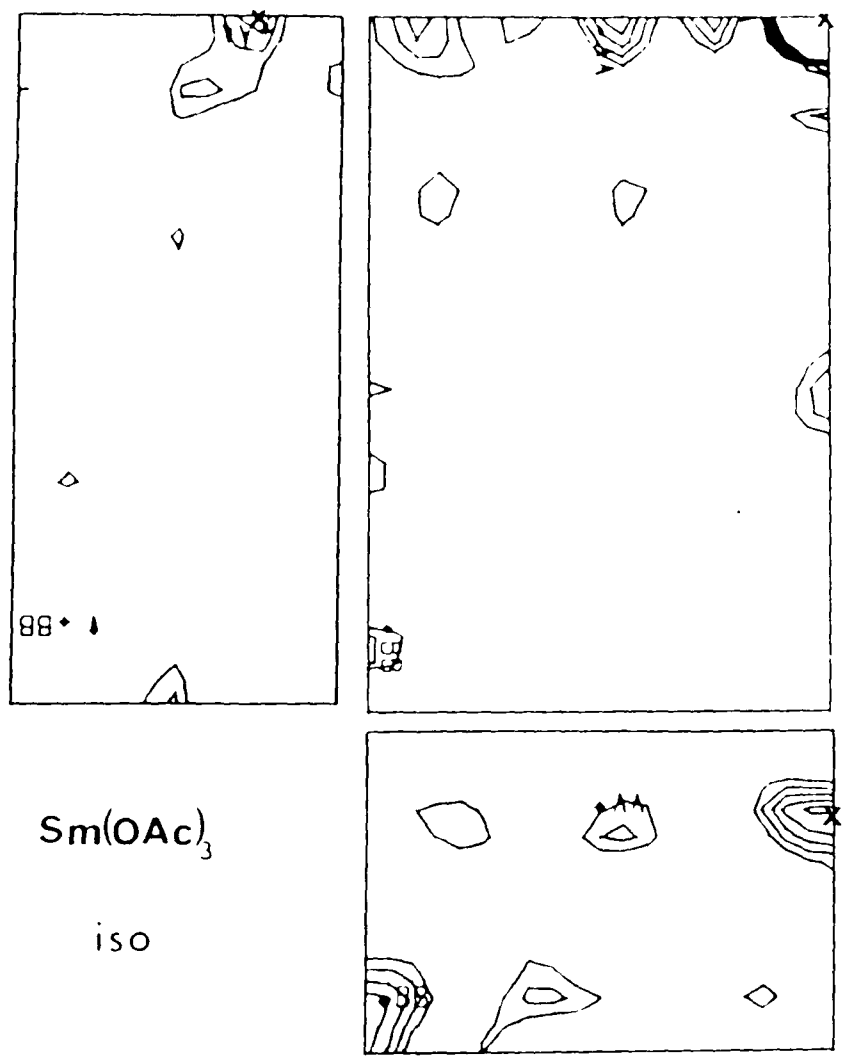
Na $AuCl_4$

iso



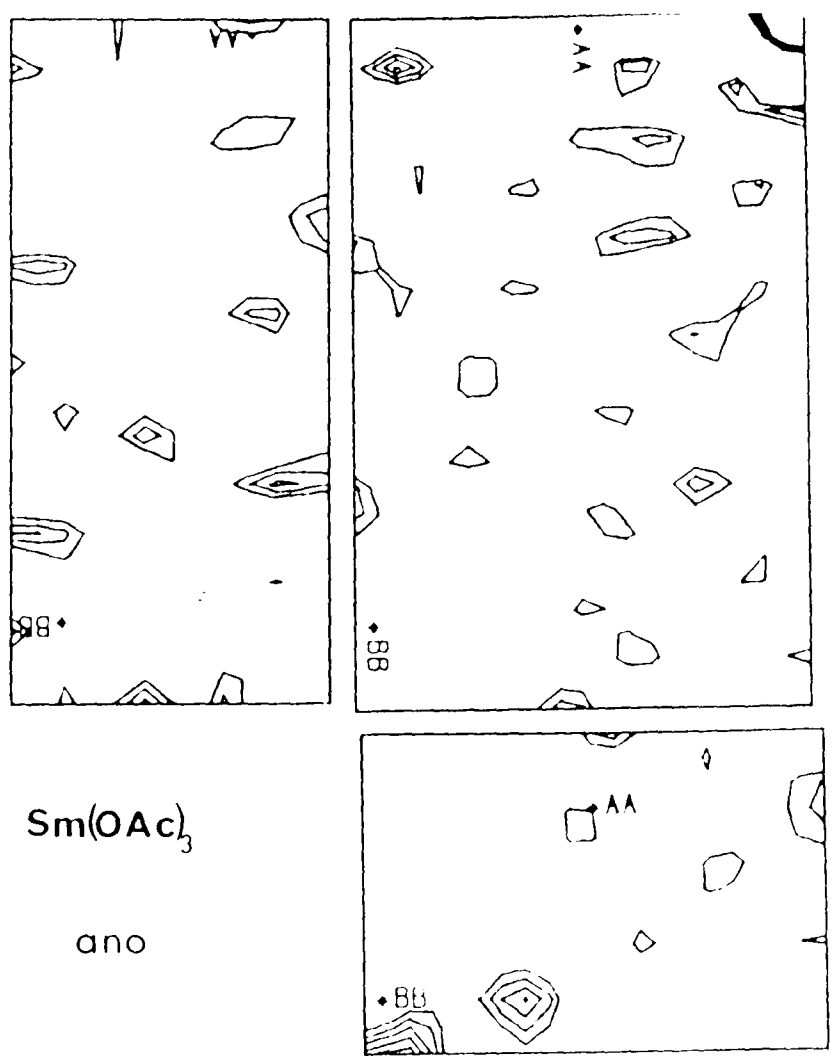
Na $AuCl_4$

ano



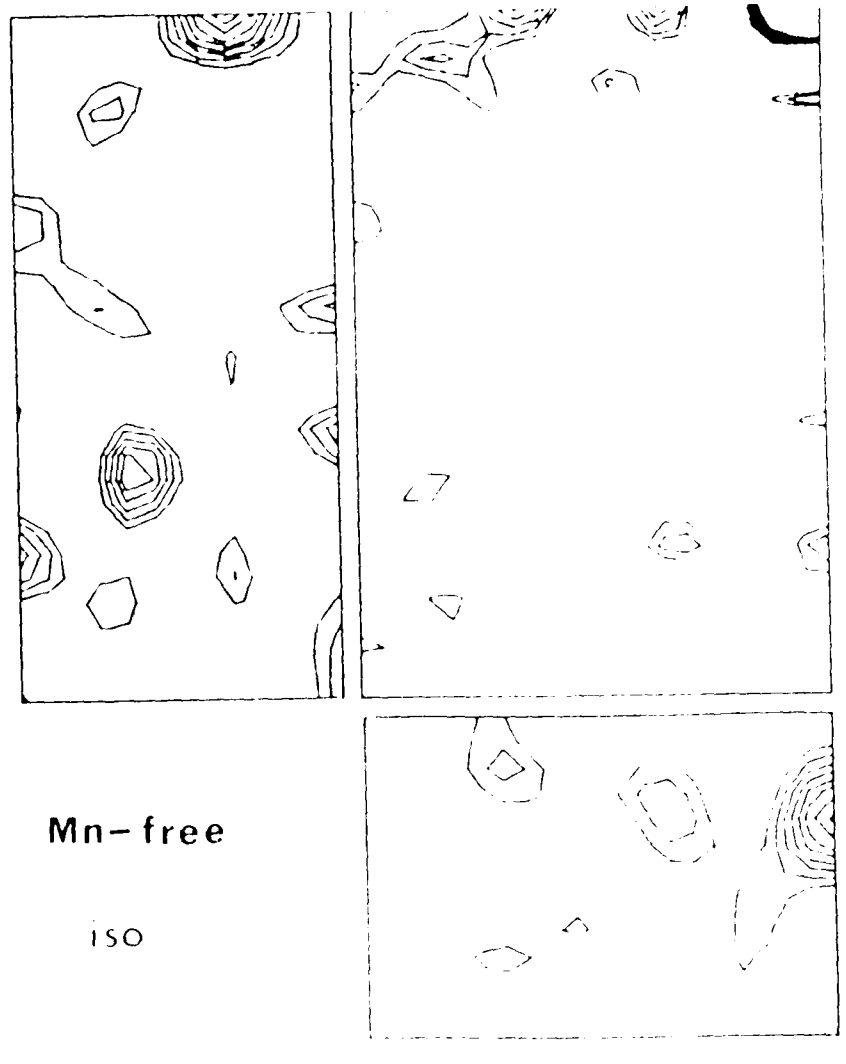
$Sm(OAc)_3$

iso



$Sm(OAc)_3$

ano



Mn-free

iso

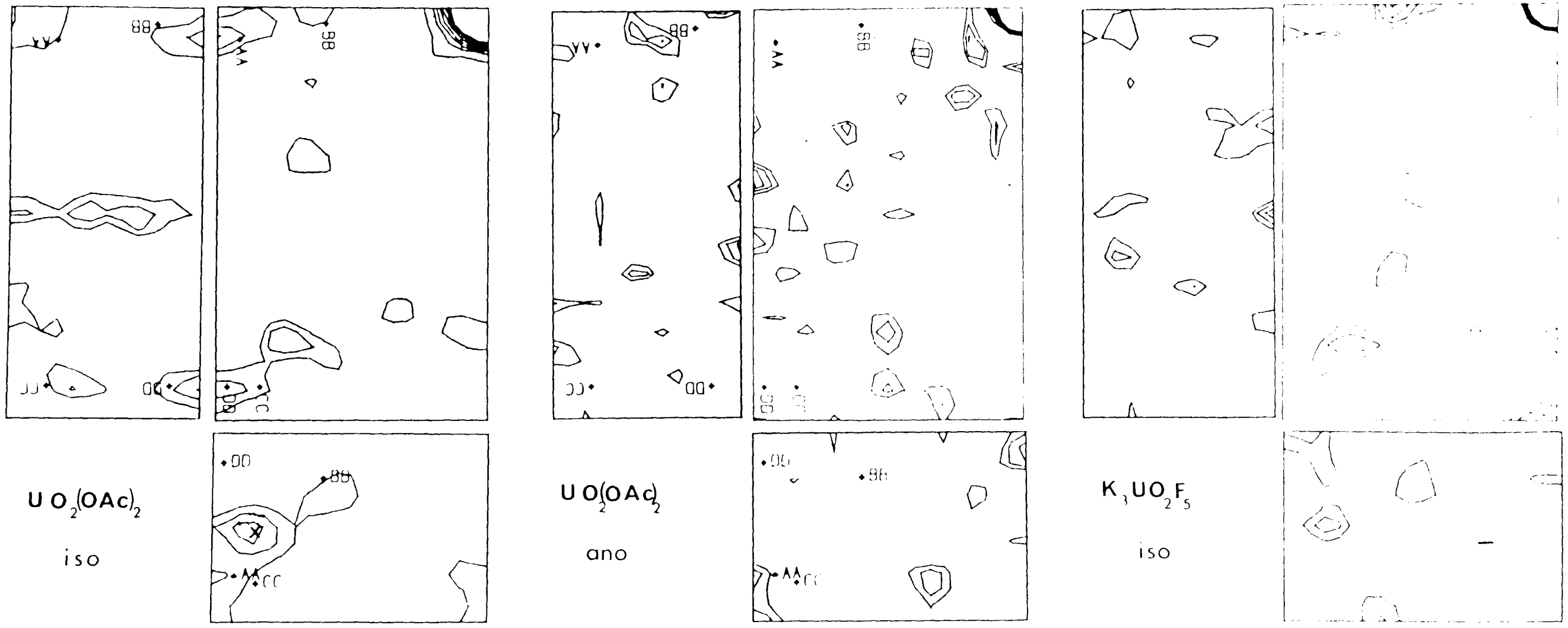
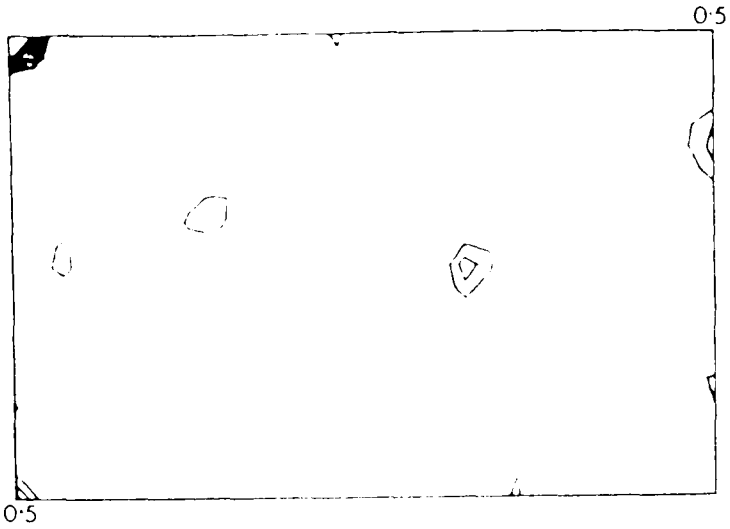


Figure 5.5 Harker sections (u= 1/2, w= 0 and v= 1/2 from upper left going clockwise) of the isomorphous and anomalous difference Pattersons for the heavy atom derivatives measured on the diffractometer. The maps were all contoured on the same equal but arbitrary intervals with the first level omitted. Positions of self-vectors derived from the final, refined heavy atom sites are shown.

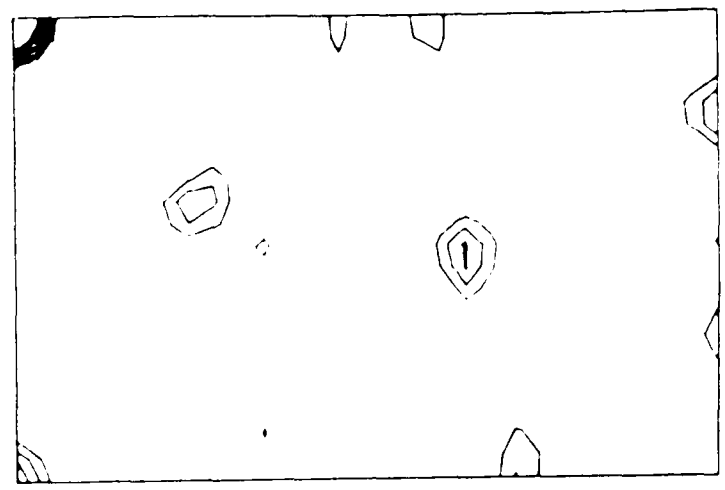
Figure 5.6 Isomorphous difference projection Patterson maps calculated from hk0 centric data. The maps were contoured at the same equal but arbitrary levels with the lowest level omitted. The origin peak is at the top left corner of each map. Also shown are the self- and cross-vectors derived from the final, refined positions of the heavy atom sites. The left hand column of maps (a, c and e) were calculated from precession film data and the right from diffractometer data.

(a) EMP, (b) EMP, (c) $Pb(OAc)_2$, (d) $Sm(OAc)_3$, (e) $K_3UO_2F_5$, (f) $K_3UO_2F_5$ and (g) $UO_2(OAc)_2$.

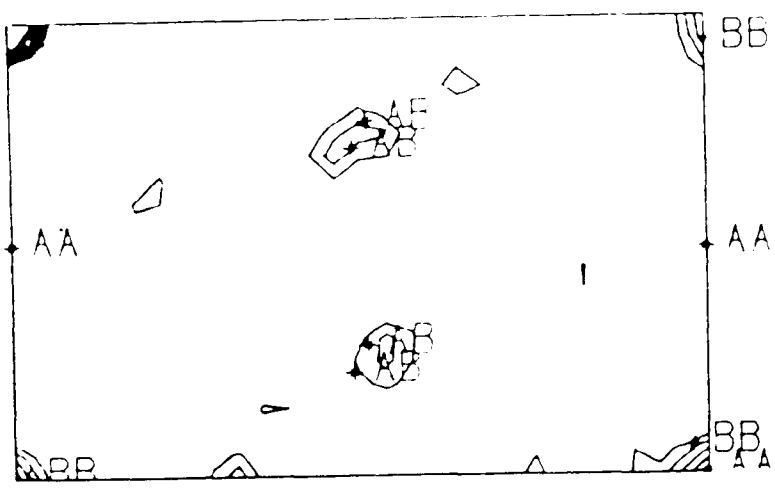
a


V →
0·5
0·5
U
↓

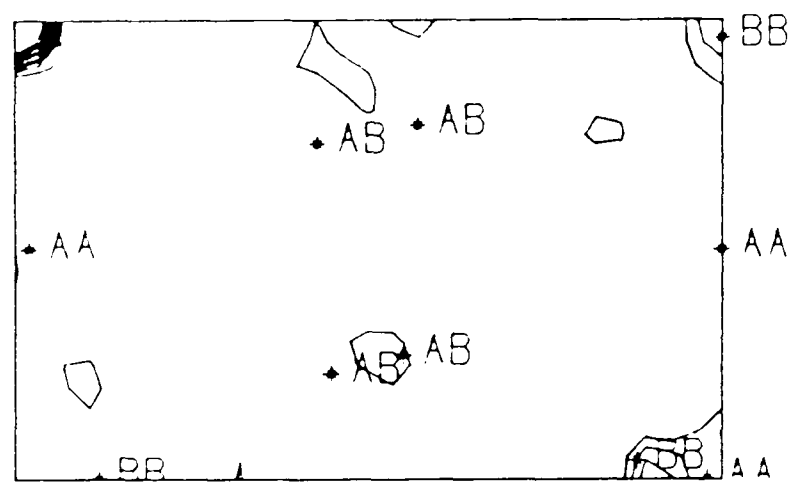
b



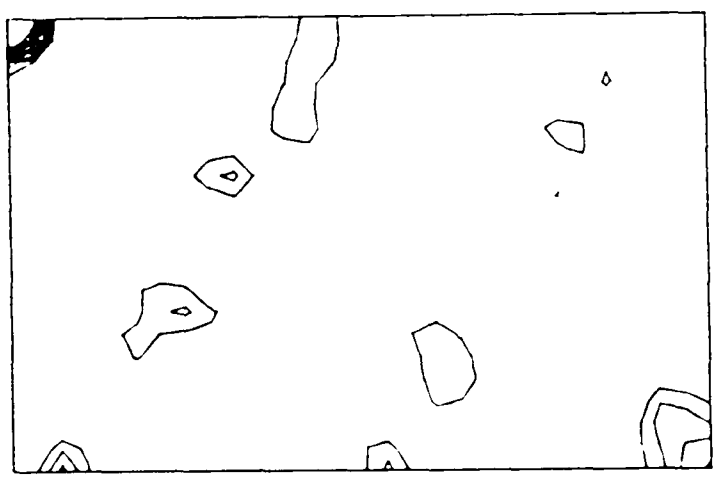
c


BB
AA
AA
BB
BB
AA

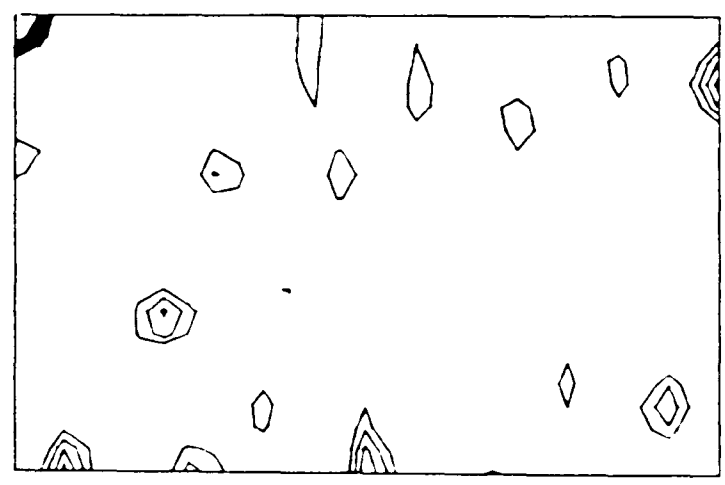
d


BB
AB
AB
AA
AA
AB
AB
BB
AA

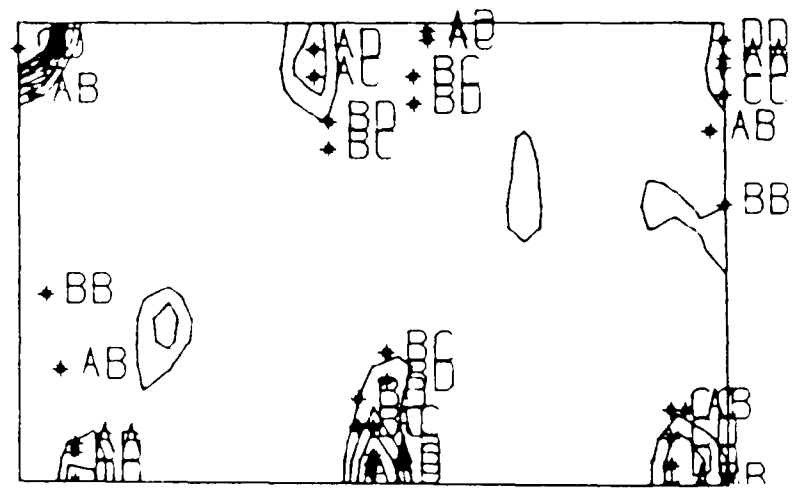
e



f



g


AB
BB
AB
BB
AB
BB
BB
AB
BB

of the samarium maps.

The isomorphous difference maps can be classed into three categories on the basis of their Harker sections. These are the gold, EMP and hard cation (uranyl and samarium) derivative maps. This classification is not surprising when one considers the chemistry of the heavy atom reagents. Finally, note the similarities in the samarium and suspected manganese-free Harker sections. The similarities suggest that some of the samarium and manganese bind at the same places on the protein. Some manganese was probably still present in the samarium-soaked crystal because of its colouration. So it was possible that some of the samarium might be occupying vacant manganese binding sites on the protein (since it is known that only 50% of the manganese sites are occupied in the native protein).

The projection isomorphous difference maps calculated from hk0 centric data (from precession film and diffractometer data) are presented in figure 5.6. Note the similarities in the maps calculated from precession film and diffractometer data and also the similarities between the maps of the hard cation derivatives.

5.7 Heavy Atom Refinement and Phasing

5.7.1 Introduction

In this section, attempts to solve the positions of heavy atom sites from the Patterson maps are discussed. For convenience, the final consensus vector positions of the heavy atom sites in the Harker sections of the Pattersons are shown in figures 5.5 and 5.6, and will be referred to thoughout the discussion. The methods used in deriving and refining the heavy atom sites have already been discussed (see section 2.4). There was a sufficient number of centric reflections to allow centric refinement of heavy atom parameters and thus avoid bias resulting from poorly measured anomalous differences in F_{HLE} refinement. Temperature factors were set to

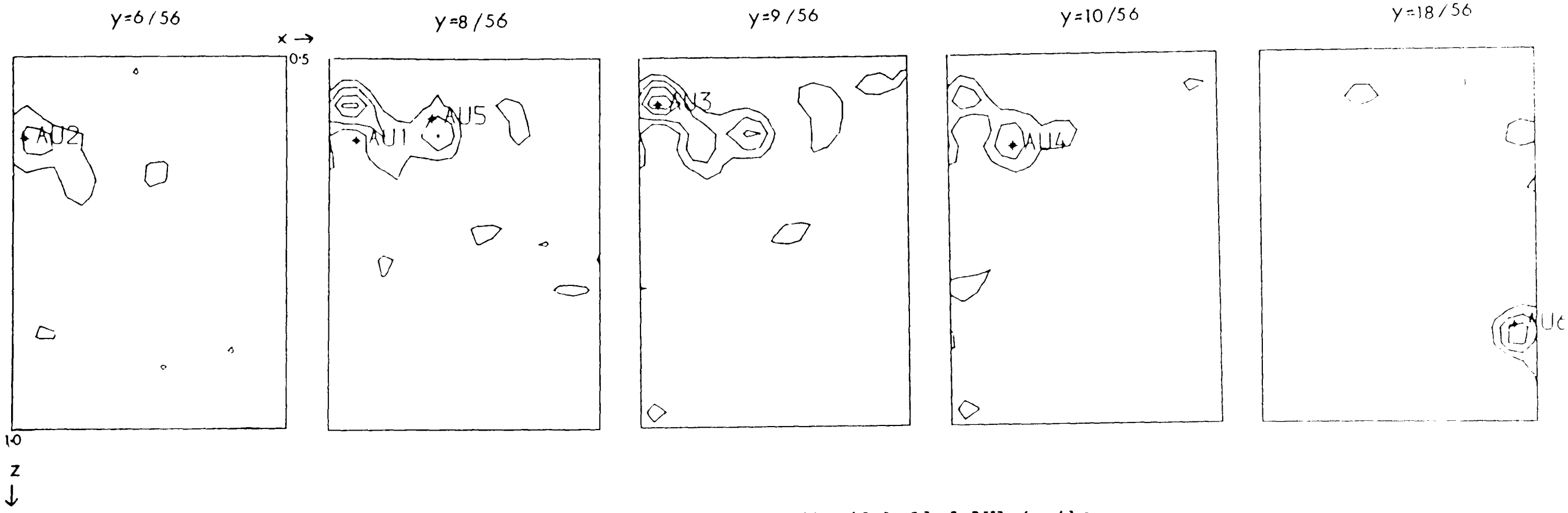


Figure 5.7 Selected y sections of the one site (labelled AU1 in the diagram) gold difference Fourier map calculated with phases from the single site gold F_{HLE} refinement described in the text. The final, refined positions of all gold sites are shown. The map was contoured at equal but arbitrary intervals with the lowest level omitted. Site AU6 is related to the cluster feature of the other sections by symmetry.

$40\mathring{A}^2$ and not refined. Refinements were carried out using Kraut scales (Kraut et al., 1962; Arnone et al., 1971) as the overall scale factors but the scales were allowed to refine together with the heavy atom parameters. In cases where F_{HLE} refinement was used, k_{emp} was used to weight down the contribution of the anomalous difference measurements in the calculation of F_{HLE} as described in section 2.4. The use of k_{theor} or bias-corrected k_{emp} and k_{theor} weighting values made little difference to the uranyl and samarium derivatives. The ratio of centric to acentric $\langle F_{HLE} \rangle$ was closest to ideal (=0.9, Wilson, 1949) for the gold derivative when k_{emp} values were used.

5.7.2 Gold Derivative

The first Patterson to be calculated was from the gold derivative. The Patterson map was interpreted in terms of two sites by inspection (sites A and B, figure 5.5) but centric refinement of the position of these sites proved disappointing. The R-factors for either and both sites was high (c.a. 70%) and difference maps suggested the site might be a complex cluster. The gold precession photo showed little change *from* the native photo at higher resolution suggesting the low order changes were due to a contrast effect by the metal in the solvent channels of the protein. This hypothesis was backed by the appearance of the gold Patterson discussed in the last section. The sites were refined by F_{HLE} methods, because the projection difference maps were difficult to interpret, and three-dimensional difference maps calculated. Six sites in a clustered arrangement were apparent in the difference map calculated with phases from the site A F_{HLE} refinement (figure 5.7). A lot of effort was spent attempting to model the feature using a variety of sites and refining the feature as one or two sites with isotropic and anisotropic temperature factors. There was no satisfactory model from this work as judged by R-factors.

5.7.3 Samarium Derivative

The uranyl and samarium acetate Patterson maps were calculated during the modelling work on the gold derivative. The uranyl map looked complex but the samarium looked simple to solve (figure 5.5). The vector solutions to the most obvious site (marked with a cross) landed on a peak positioned on the edge of two Harker sections as well as the origin peak. Attempts to refine this site and find other sites by difference maps proved unsuccessful. Similarily, attempts to refine site A by itself and in combination with the crossed site led nowhere. R-factors were dissapointing (R_C of about 70%; see table 5.7 for definition) and peaks in the one or two site difference Fourier maps did not solve the Patterson. There was an ambiguity in the v (Patterson space) coordinate for site B in the Patterson (figure 5.5) because of the presence of two peaks with the right u coordinate in the w=0 Harker section. Refinements of various interpretations of this site with the other possible solutions yielded no results. F_{HLE} difference maps were unhelpful. A lot of work was spent on trying to solve the origin problem between the possible sites by trial-and-error methods. The absence of significant peaks on the non-Harker sections of the samarium Patterson meant the origin problem could not be solved by searching for suitable cross-vector peaks in the Patterson map.

There appeared a close relationship between the samarium and suspected manganese-free difference Pattersons (figure 5.5). However, the manganese-free Patterson could not be solved (this work was discussed in the last chapter). Attempts to solve the samarium Patterson by using phase sets calculated from possible solutions to the maganese-free data were unsuccessful.

5.7.4 Uranyl Derivative

Although the uranyl acetate Patterson looked complex it was possible to

solve it by inspection in terms of three sites (sites A, B and C) which took account of all the major features on the map (including cross-vectors). (The initial vector position of site C in the v=1/2 Harker is marked with a cross; the vector position refined away from the cross and was replaced by cross-vector positions. See the complete difference Patterson maps in appendix one). The solution was confirmed by using phases calculated from a model for the gold Patterson (see table 5.7). These phases were sufficient to build up a solution for the uranyl Patterson by difference and double difference map techniques. However it was dissappointing to find that the phases calculated from the resultant uranyl sites were not good enough to reproduce the gold cluster sites in difference maps.

Phase sets calculated from both the gold and uranyl solutions produced noisy samarium difference maps with bad 'ghosting' from the input sites. In hindsight, the uranyl phase set was producing real samarium sites, but at the time these features were attributed to the ghosts of the uranyl sites used in the phasing. A close look at the uranyl and samarium Patterson maps indicated that some sites could well be common and thus would explain why samarium difference maps from uranyl phases were dominated by 'ghosting' from the uranyl sites.

5.7.5 EMP and Uranyl Fluoride Derivatives

At this stage, Patterson maps of the EMP and uranyl fluoride came available. The uranyl fluoride map was similar to the uranyl acetate map but of poorer quality due to weaker data. The uranyl fluoride data was refined by taking the solution of the uranyl acetate Patterson but the refinements statistics were not as good as the uranyl acetate ones. The signal-to-noise ratio of the EMP Patterson map was low, making it difficult to solve the Patterson by inspection. Attempts to refine possible sites led nowhere. Difference maps of the EMP derivative using phases calculated from

the uranyl or gold derivatives yielded sites which did not solve the Patterson. Eventually, the EMP data were discarded because nothing could be made of the data.

5.7.6 Projection Data

At this stage, solutions had been obtained for the gold and uranyl acetate Pattersons which accounted for the major features on both maps. But refinement statistics suggested the gold was a poor derivative and the uranyl solution might be improved. Difference maps of the uranyl derivative calculated with gold phases were dominated by two peaks which solved the uranyl Patterson but the broadness of the peaks suggested that they represented at least two or three sites each. It was hoped that information from hk0 projection data might help improve the solutions. The quality of the anomalous data for all derivatives was poor as judged from anomalous difference maps and various statistics, so projection data was important information. Further it was possible to incorporate new information from precession data of hk0 zones (see section 5.5). The lead acetate hk0 projection Patterson appeared immediately soluble, subject to an ambiguity in the number and origin relationships of sites. A two site solution accounted for all features on the map (figure 5.6c). The centric R-factor for the lead solution was dissappointing (c.a. 65%) and double difference maps yielded no further sites. Phase sets calculated from the lead solution produced noisy maps of the other derivatives but confirmed the interpretation of the uranyl acetate Patterson map.

It was decided at this stage to refine the hk0 data using the best sites of all derivatives and to combine the phases so as to pick up minor sites. The phasing was good enough to produce the gold cluster feature when the gold sites were left out of the phasing calculation. After an exhaustive round of refinement and phasing a final interpretation of each derivative was obtained.

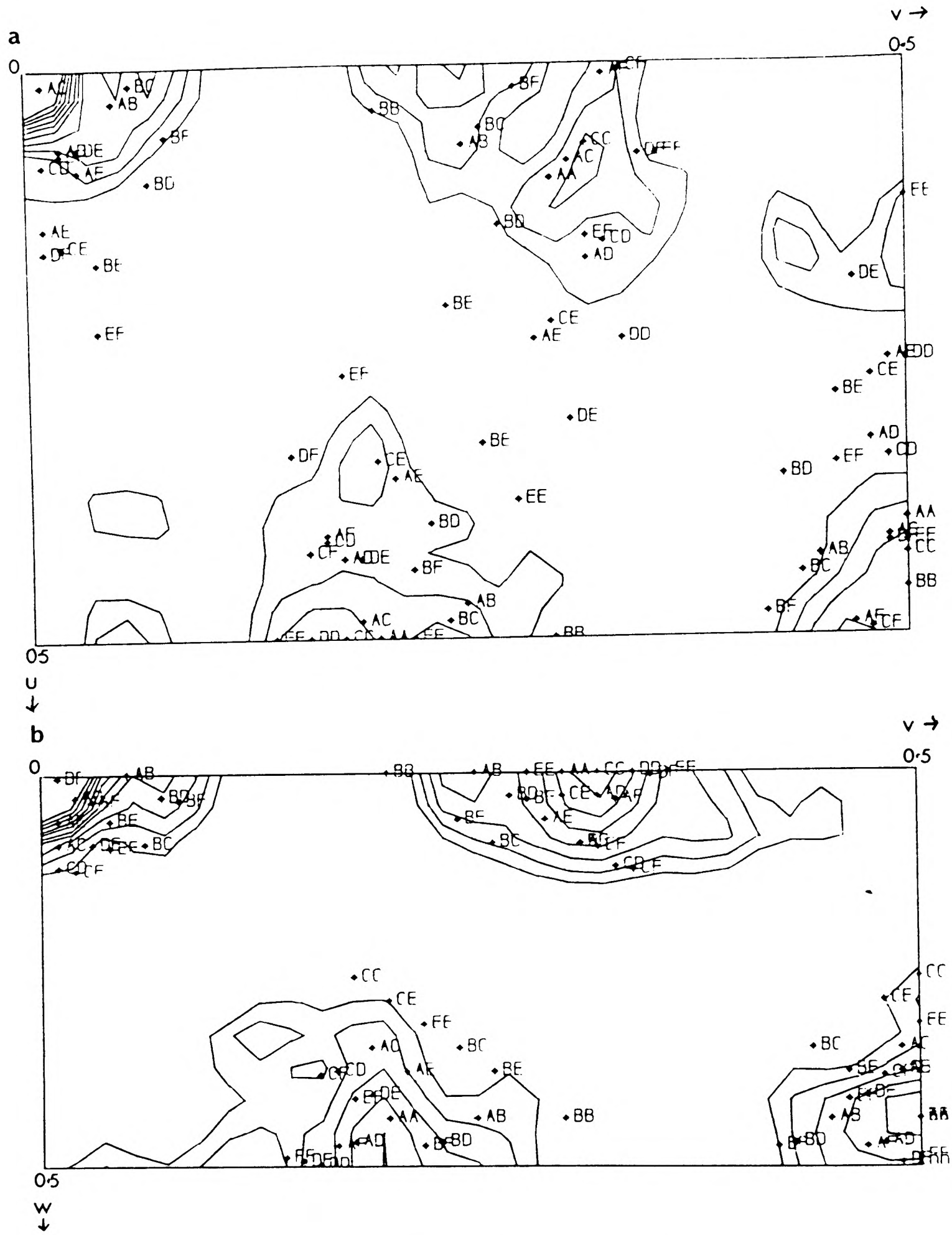


Figure 5.8 Projection isomorphous difference Patterson maps of the gold derivative showing self- and cross-vectors from the final, refined heavy atom sites. The maps were contoured on the same equal but arbitrary intervals. (a) hk0 projection and (b) 0kl projection.

However, the z coordinate of the sites were either uncertain or unknown. The z coordinate of the main gold site was thought to be correct (derived from F_{HLE} difference maps, see figure 5.8), so this was used as a starting point for cross difference maps of the other derivatives. The solutions were cross-checked in three-dimensional difference and double difference maps using phase sets calculated from various combinations of solutions to the different Pattersons. During these procedures it became obvious that the uranyl fluoride had to be discarded due to poor quality data.

The analysis of the projection data gave a clearer indication of the positions of sites for the 'cationic' derivatives. The interpretation of the gold remained unchanged from that obtained from the F_{HLE} work.

5.7.7 Final Refinement Runs

The final interpretation of the gold Patterson came from the F_{HLE} work discussed in section 5.7.2. The phase sets calculated from solutions of the uranyl and samarium derivatives were only good enough to yield the major site (site A) of the gold cluster feature. The phase sets calculated from the gold model and projection data clearly indicated that the 'cationic' derivatives (lead, samarium, uranyl fluoride and uranyl acetate) all bind at two distinct places on the protein in the asymmetric unit of the unit cell. For the lead and samarium derivatives, this corresponded to two sites. But analysis of difference and double difference maps of the uranyl derivatives suggested that each position was occupied by two or three heavy atoms. In the end, a four site interpretation of the uranyl acetate was settled for on the basis of refinement statistics. Phase refinement of the heavy atom derivatives was not attempted because of the poor quality of the gold derivative and the similarity of the 'cationic' derivatives. Final double difference maps of each derivative (using phases calculated from the gold, samarium and uranyl acetate solutions) were flat, confirming the

Gold — Rc=55.3%, c.g.=0.3, k=0.98

Site	Occupancy	X	Y	Z	B
1	0.38	0.05	0.15	0.22	40
2	0.24	0.02	0.10	0.22	40
3	0.17	0.04	0.16	0.13	40
4	0.16	0.12	0.17	0.25	40
5	0.10	0.19	0.14	0.16	40
6	0.19	0.46	0.32	0.75	40

Uranyl — Rc=53.9%, c.g.=0.3, k=0.96

Site	Occupancy	X	Y	Z	B
1	0.40	0.27	0.02	0.69	40
2	0.36	0.35	0.01	0.44	40
3	0.40	0.21	0.23	0.80	40
4	0.27	0.24	0.23	0.96	40
5	-	0.23	0.24	0.95	40
6	-	0.28	0.21	0.85	40

Samarium — Rc=63.5%, c.g.=0.16, k=0.99

Site	Occupancy	X	Y	Z	B
1	0.55	0.38	0.01	0.44	40
2	0.46	0.24	0.22	0.79	40

Lead — Rc=65.0%, c.g.=0.19, k=0.99

Site	Occupancy	X	Y	Z	B
1	0.45	0.38	0.0	-	40
2	0.79	0.24	0.25	-	40

Table 5.7 Final heavy atom parameters from centric refinement.

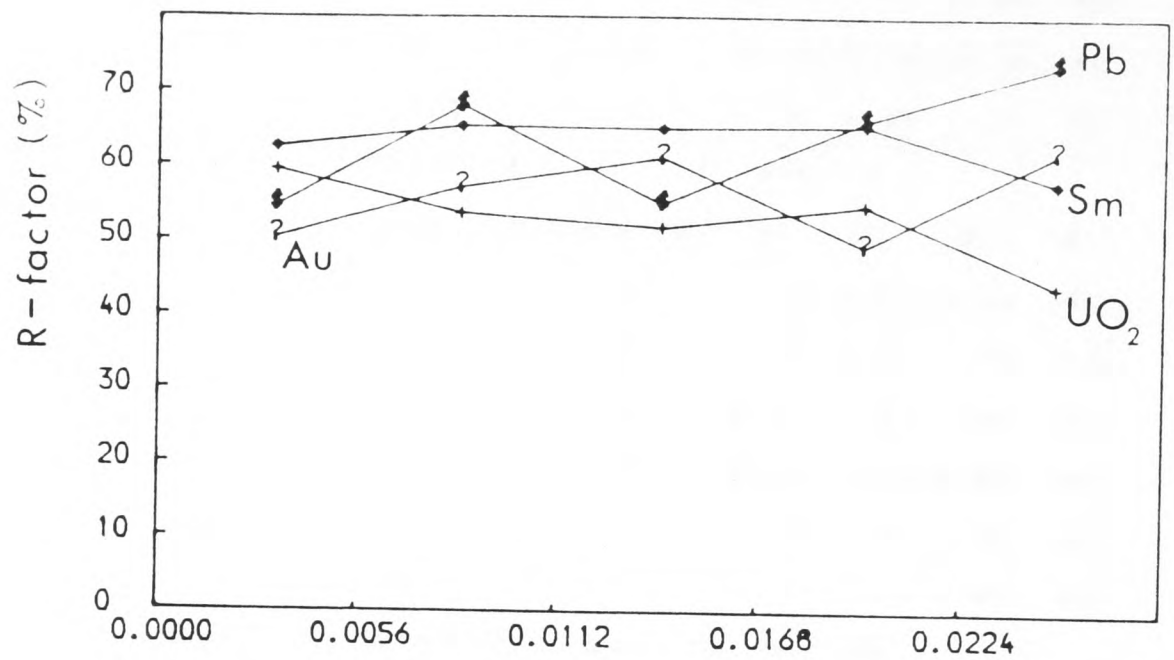
Key: $Rc = \sum | F_{H(obs)} - F_{H(calc)} | / \sum F_{H(obs)}$

c.g. = the slope of the plot of rms $F_{H(calc)}$ for different $F_{H(obs)}$ ranges against median $F_{H(obs)}$ for that range.

k = overall scale factor.

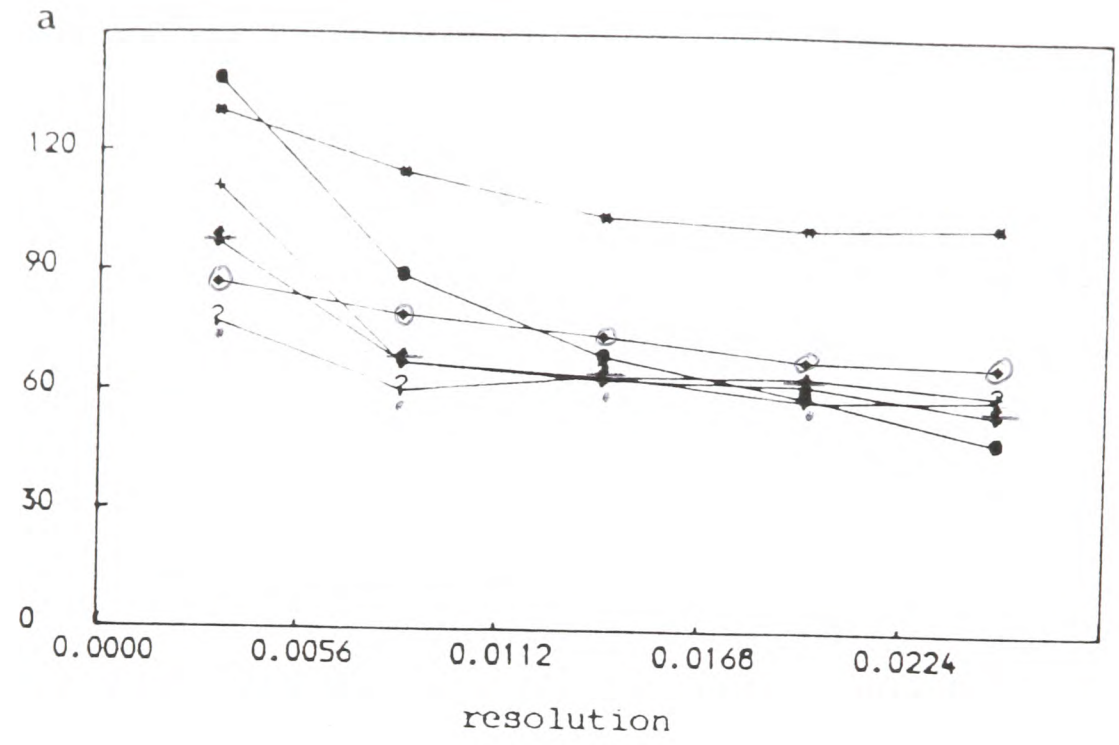
B = isotropic temperature factor.

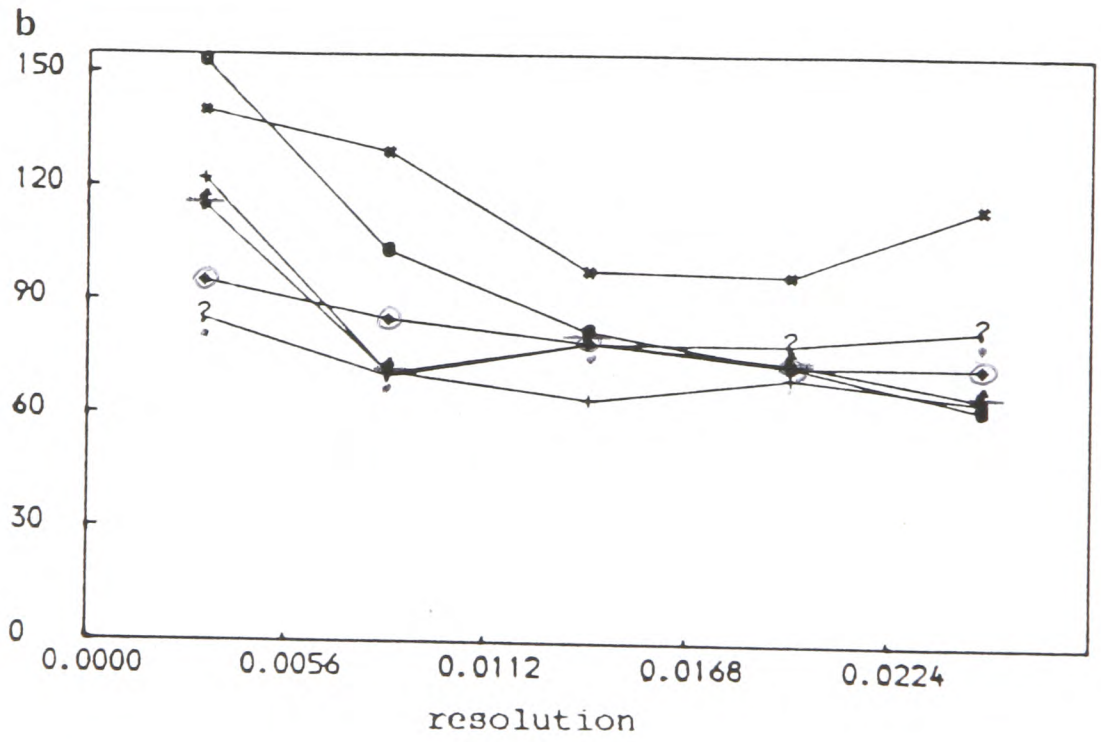
Notes: For six uranyl sites the centric R-factor was 54.8% and c.g.= 0.40.



Key : R-factor = $\sum |F_{Hobs} - F_{Hcalc}| / \sum F_{Hobs}$.

Figure 5.9 R-factor plot against resolution ($4\sin^2\theta/\lambda^2$) obtained from centric refinement runs.





	F_{Hcalc}	E_{mp}
$UO_2(OAc)_2$	x	▪
$NaAuCl_4$	†	?
$Sm(OAc)_3$	❀	$

Figure 5.10 Plots of calculated heavy atom contribution and most probable lack-of-closure errors as function of resolution ($4\sin^2\theta/\lambda^2$) for, (a) all data and (b) centric terms only.

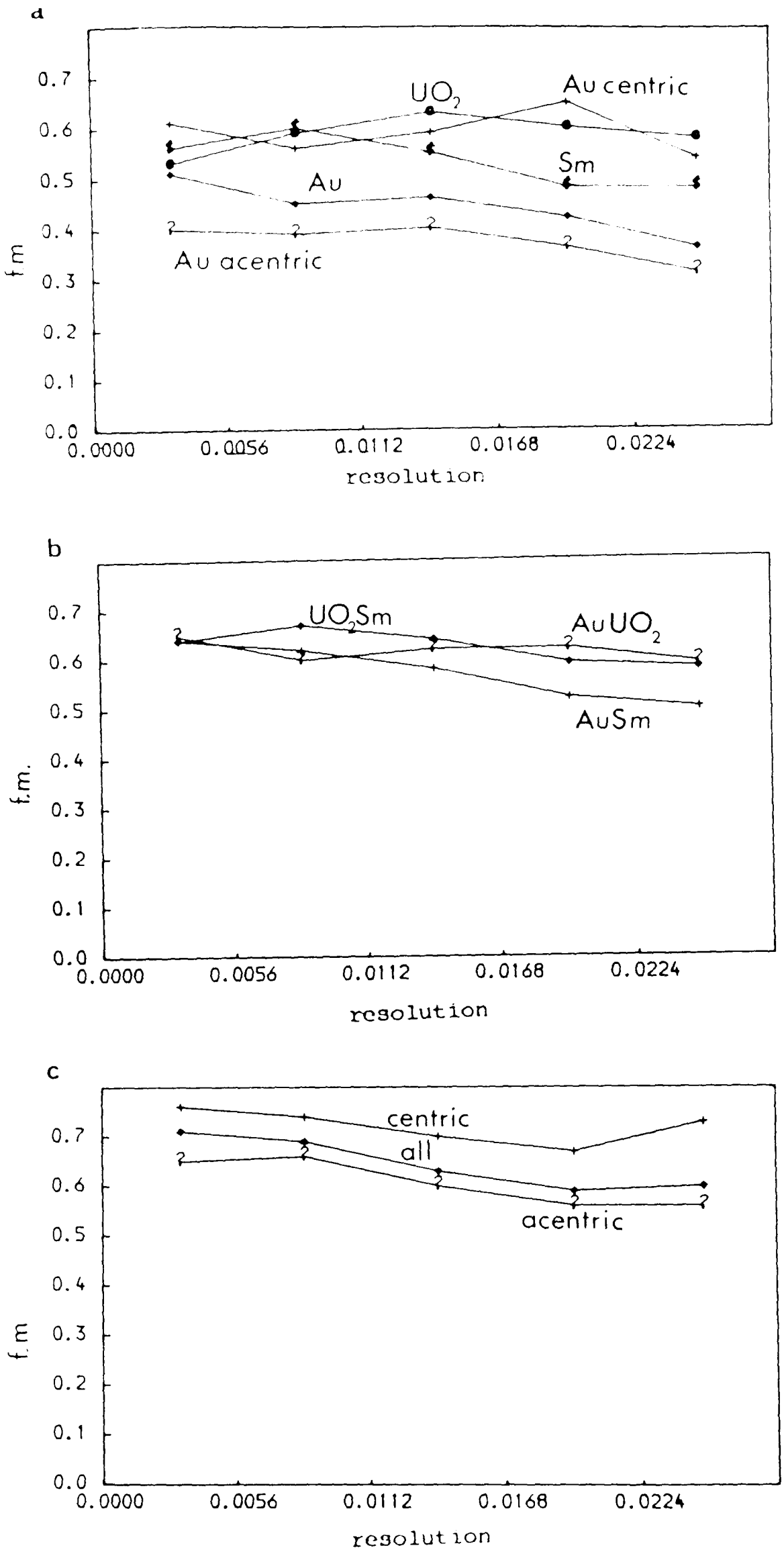


Figure 5.11 Plots of figures-of-merit versus resolution ($4\sin^2\theta/\lambda^2$) for, (a) single, (b) double and (c) three derivative protein maps. The anomalous contribution of each derivative was included in the phasing runs.

Gold Sites	Uranyl Sites	Samarium Sites	Hand	Figure of Merit
-	4	-	R	0.588
-	-	2	R	0.523
6	-	-	R	0.431
-	4	2	R	0.618
6	4	-	R	0.614
6	-	2	R	0.561
6	4	2	R	0.631
6	4	2	A	0.630
6	6	2	R	0.629

Table 5.8 Figure-of-merit values for various phase sets. Anomalous information was included for all derivatives in the phase calculation.

Key: R = retarded phases

A = advanced phases

final interpretations.

The final, refined heavy atom parameters are compiled in table 5.7. On the basis of statistics presented in the table, the uranyl four site solution appeared to be the most satisfactory solution of all the derivatives. This point has been emphasised in figure 5.9 were the centric R-factors have been plotted as a function of resolution. The final fit of self- and cross-vectors to the three-dimensional Pattersons are shown in appendix one.

5.8 The Electron Density Maps

The final parameters from the heavy atom refinement were used to calculate the protein phase angles and figures-of-merit by the method of Blow and Crick (1959). The lack-of-closure errors were taken from the heavy atom centric refinements and phase runs repeated until the input and output root mean square lack-of-closure errors agreed. Phasing statistics are presented in figures 5.10 to 5.11 and table 5.8. The major points to note from the statistics are the poor phasing quality of the gold derivative, the superior quality of the phases produced from the uranyl derivative over the others and the deterioration of the phasing power of all the derivatives with increasing resolution. It was not possible to distinguish the absolute configuration of the heavy atoms with confidence on the basis of figure-of-merit values. A comparison of heavy atom difference maps calculated on both hands did not solve the enantiomorph problem either.

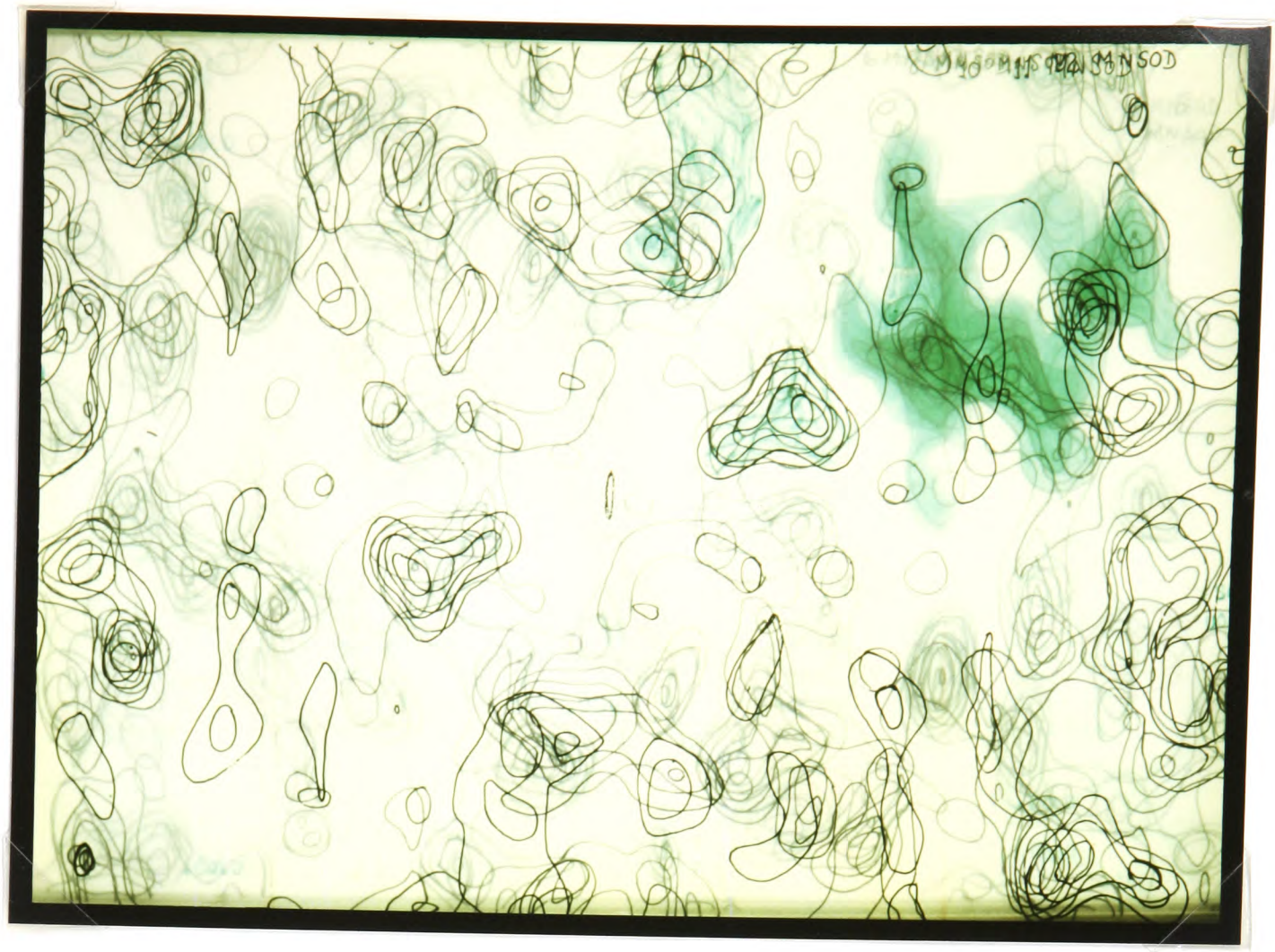
Electron density maps were calculated throughout the course of refinement but were difficult to interpret. The electron density maps were sampled in 40 intervals along x, 56 along y and 24 along z, corresponding to intervals of approximately 2Å along a, b and c. The z axis was chosen as the sectioning axis because it was the shortest. The maps were contoured at equal but arbitrary intervals. No attempt was made to calculate the maps on an absolute scale.

The final parameters obtained from centric refinement of the gold, samarium and uranyl derivatives were used to calculate two sets of native phases, one with anomalous advancing, and one retarding the phases. Electron density maps calculated from both sets of phases are shown in figures 5.12 and 5.13. The map calculated with anomalous-retarding phases had some encouraging features. Extensive solvent regions were apparent but a clear molecular boundary could not be traced without ambiguity. It was possible to define an unique portion of density in the map (coloured in green, figure 5.12) which had the right dimensions of an FeSOD monomer. The internal detail of the proposed monomer density was uninterpretable with the exception of a couple of rod-like dense features which might have been helices. The anomalous-advanced map was more noisy and it was more difficult to define solvent boundaries and assign possible protein features in the map. Attempts to improve the phases are discussed in the next chapter and so a detailed discussion of the map will not be made here.

5.9 Discussion

It should be apparent from the discussion in the preceeding sections that the heavy atom refinement did not progress smoothly because of poorly isomorphous derivatives. Analysis of single derivative phase sets showed, that on average, only 60% of the centric phases agreed. It was found possible to refine false sites for all the derivatives. Derivative cross-phased maps were always noisy and sometimes produced features that did not solve the Patterson. Ghosting of sites used in the phase calculations was common. All these observations are indicators of the poor quality of the derivatives. It was only by the incorporation of many different sources of information that convincing solutions for the derivatives were found. The final refinement R-factors and the quality of the final electron density maps further confirm the poor degree of isomorphism of the derivatives.

Despite a systematic survey of potential derivatives (section 5.2), not



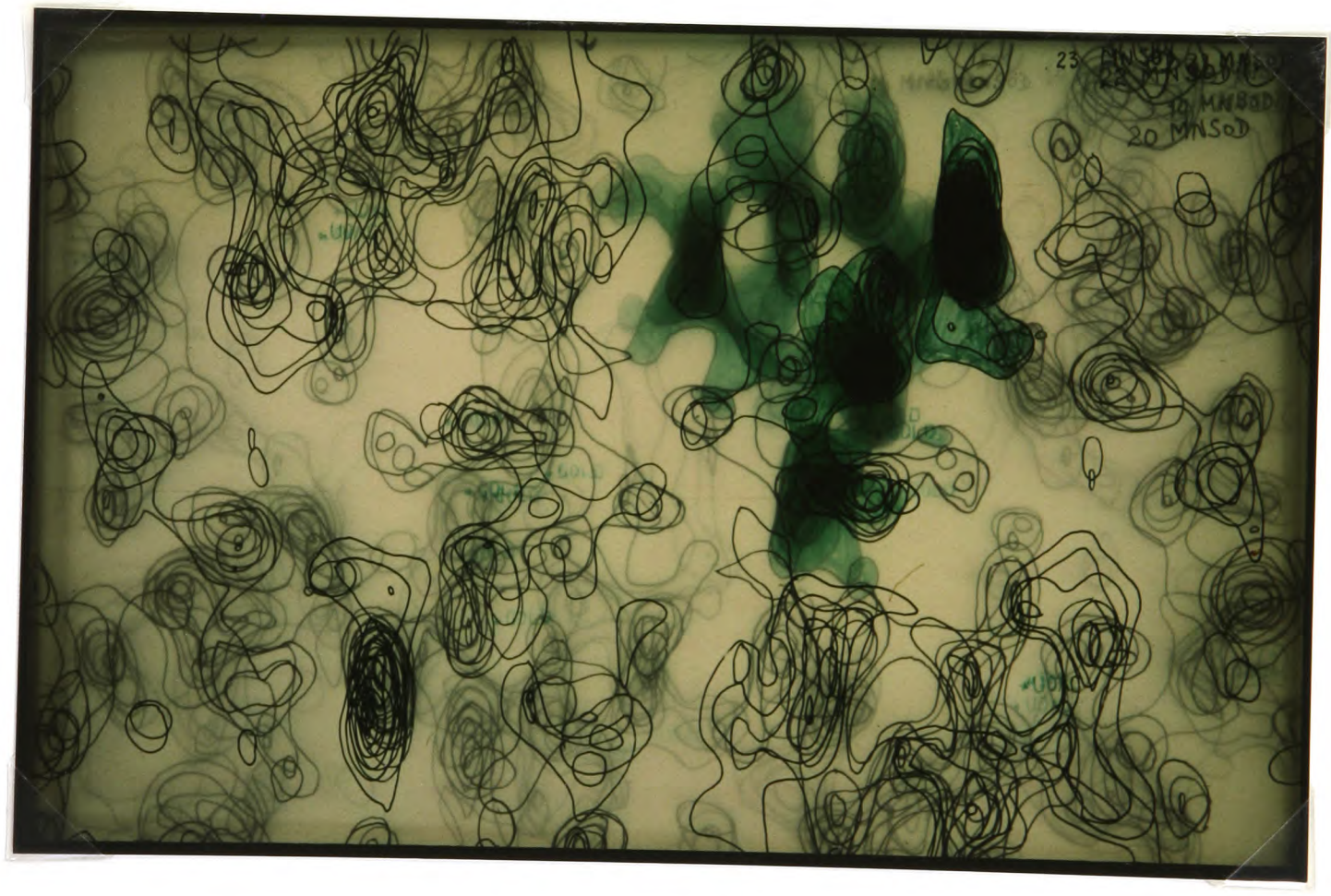
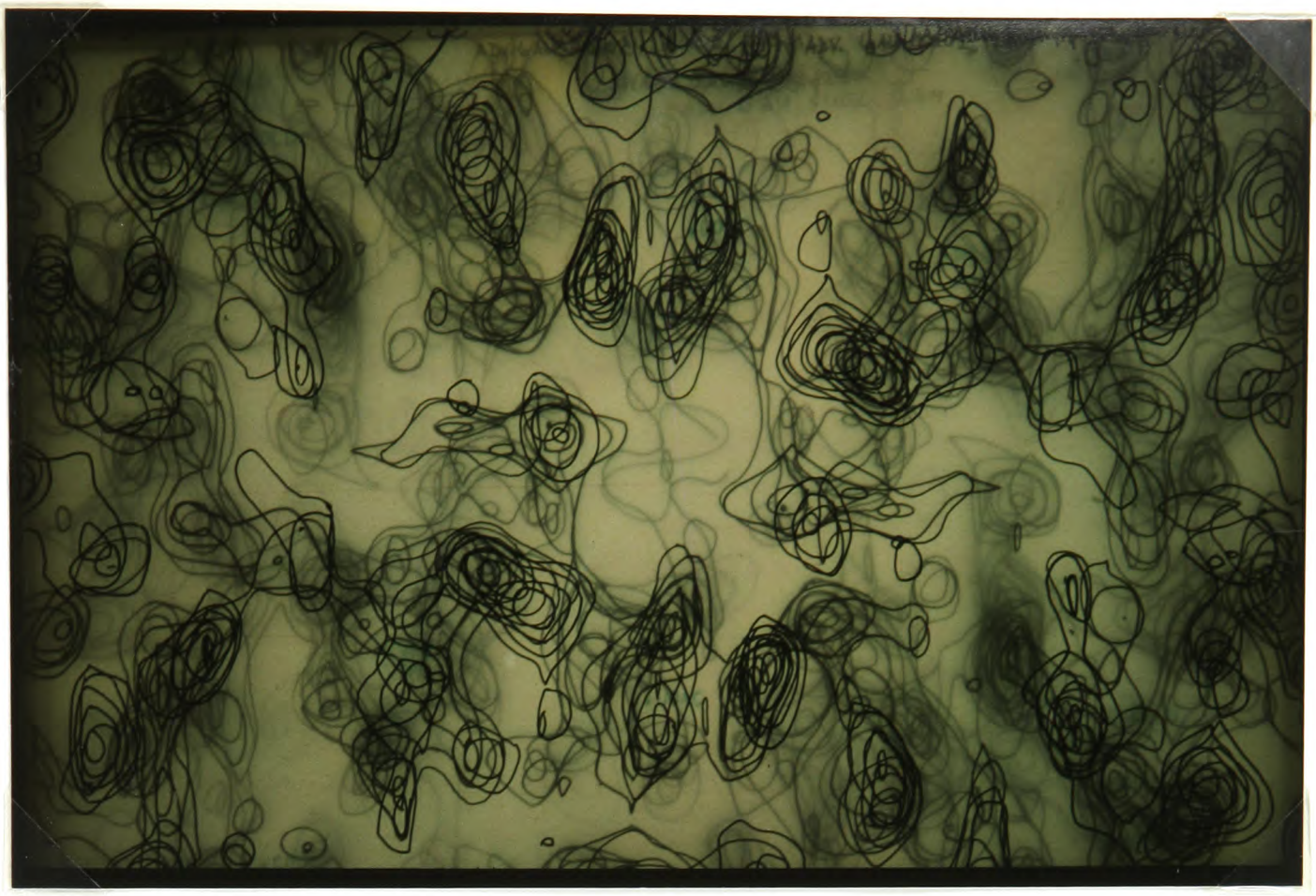


Figure 5.12 Electron density map calculated with 'anomalous-retarded' phases (see text). The map was calculated to 6Å resolution for a whole unit cell and contoured at the same equal but arbitrary intervals. The top map stack is sectioned from 0 to 1/2 and the bottom from 1/2 to 1 along the z (parallel to c) sectioning axis of the unit cell.



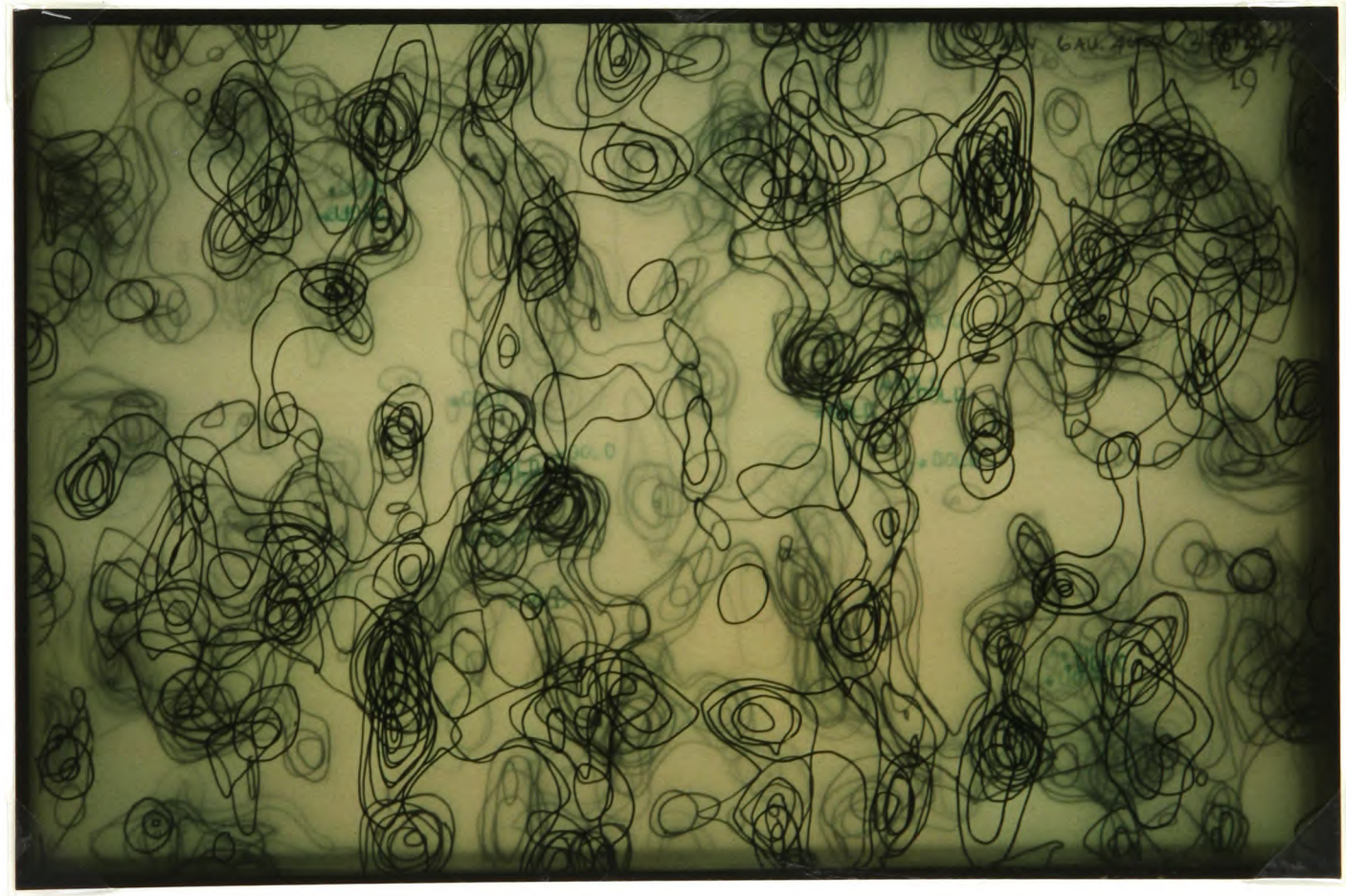


Figure 5.13 Electron density map calculated with 'anomalous-advanced' phases (see text). See caption to figure 5.12 for details.

one proved useful as a good isomorphous derivative. This finding was in accord with the observations of Jansonius's group in Switzerland who were working on the $P2_12_12_1$ form of the enzyme (section 3.1). They had explored over a hundred derivatives and found all to be badly isomorphous with respect to the native (Jansonius, pers. commun.). Methods of improving the phasing are described in the next chapter and a full discussion of the problems encountered with solving the phase problem is left till the final chapter.

CHAPTER 6

Phase Improvement and Extension

6.1 Density Modification

The calculation of the final electron density maps, discussed in the last chapter, presented a number of problems. It was clear from the phasing statistics that the uranyl derivative was contributing most of the useful phase information towards the map and it was difficult to assess the relative phasing contributions of the other derivatives. It was evident that all three derivatives were suffering from a large degree of non-isomorphism because of their heavy atom refinement behaviour and cross-phasing capabilities. Maps calculated from single derivative data were very noisy but were probably being adversely affected by the unreliable phasing power of the weakly measured anomalous scattering data. Maps calculated from two derivative combinations tended to retain a significant proportion of the promising features seen in the final three derivative map and thus suggested that each derivative was contributing some useful phase information.

The map modification procedure developed by Wang (1985), and discussed in chapter two, presented a way of assessing the isomorphous contribution of each derivative towards the phasing process. This is because maps could be calculated from single derivative data (single isomorphous replacement, SIR) without the need to resort to anomalous scattering in order to solve the phase ambiguity. Although the Wang procedure cannot resolve the phase ambiguity correctly if the intensity differences between native and derivative data are due to a significant degree of non-isomorphism, it was hoped that a least one derivative was sufficiently isomorphous to enable an unambigous determination of the molecular boundary in an SIR map. The presence of relatively featureless regions in the various electron density

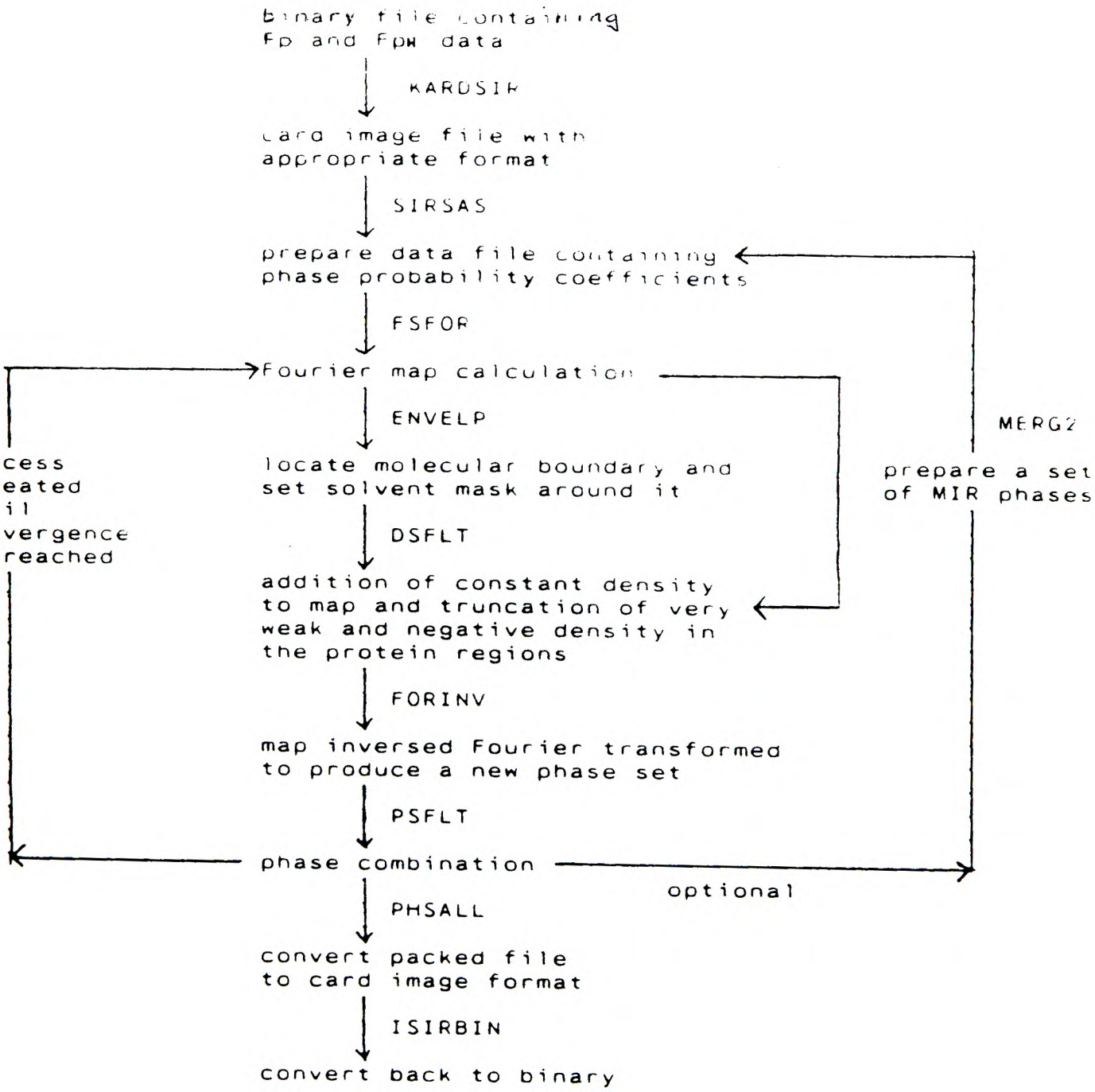


Figure 6.1 Flow diagram of steps taken to improve the quality of a protein map by the method of Wang (1985). Names of computer programs given in block letters.

maps suggested that solvent information could be extracted and used to resolve the phase ambiguity inherent in the SIR data.

A flow diagram of Wang's map modification procedure is presented in figure 6.1 and the theory behind the approach has been described in chapter two. The first step in the procedure was to calculate phase probability coefficients from the SIR data (Hendrickson & Lattman, 1970) of each derivative using the heavy atom parameters listed in table 5.7. The 'best' Fourier map (Blow & Crick, 1959) was calculated for each derivative with the F(000) term set to zero. Each map was then modified by setting negative regions to zero and calculating a new map as described in section 2.6.2 (using an averaging sphere of 6Å radius and making two passes through a given map). A solvent mask was then calculated, assuming the solvent content of the crystal to be 45% (see section 3.3), as described in section 2.6.2. The density in each map was rescaled to account for the omission of the F(000) term, remaining negative regions were set to zero and the solvent regions were set to an averaged value. Each map was transformed to produce structure factor amplitudes and phases which were then combined with the SIR phase information using Bricogne's adaption (1976) of Sim's weighting function (1959):

$$P_{calc}(\Phi) = \exp\left[2\, F_{obs}\, F_{Wang} \cos(\Phi - \Phi_{Wang}) \,/\, \langle F_{obs}^{2} - F_{Wang}^{2} \rangle\right]$$

The new phases were then used to recalculate maps and the filtering processes repeated using the same solvent mask until convergence (which was attained in four to eight cycles of iteration). The stability of the envelope determination was ensured by recalculating the solvent mask a further two times using phases output from the final round of iteration obtained with the previous mask. This iteration procedure has been found to be effective for other data (Wang, 1985).

The convergence of the iteration procedure was monitored by calculating average phase shifts, figures-of-merit, conventional R-factors and

	$UO_2(OAc)_2$	$AuNaCl_4$	$Sm(OAc)_3$	Three Derivatives	$UO_2(OAc)_2$
number of reflection pairs	1036	1062	1033	1103	1036
number of rejections	69	41	66	-	69
solvent content (%)	45	45	45	45	40
average phase shift (^{0})	2.4	1.8	1.8	1.8	1.7
overall figure-of-merit	0.71	0.72	0.73	0.72	0.74
R-factor	0.19	0.21	0.19	0.24	0.17
correlation coefficient	0.98	0.98	0.98	0.96	0.98

Table 6.1 Phasing statistics after the final round of iteration of Wang's map modification procedure.

Key: R-factor = $\Sigma_{hkl} | F_{obs} - F_{calc} | \; / \; \Sigma_{hkl} F_{obs}$

correlation coefficient = $\langle F_{obs} F_{calc} \rangle \; / \; (\langle F_{obs} F_{obs} \rangle \langle F_{calc} F_{calc} \rangle)^{1/2}$

Note: Reflections were rejected if:

(a) $\Delta_{iso} > 3F_{H(calc)}$ or

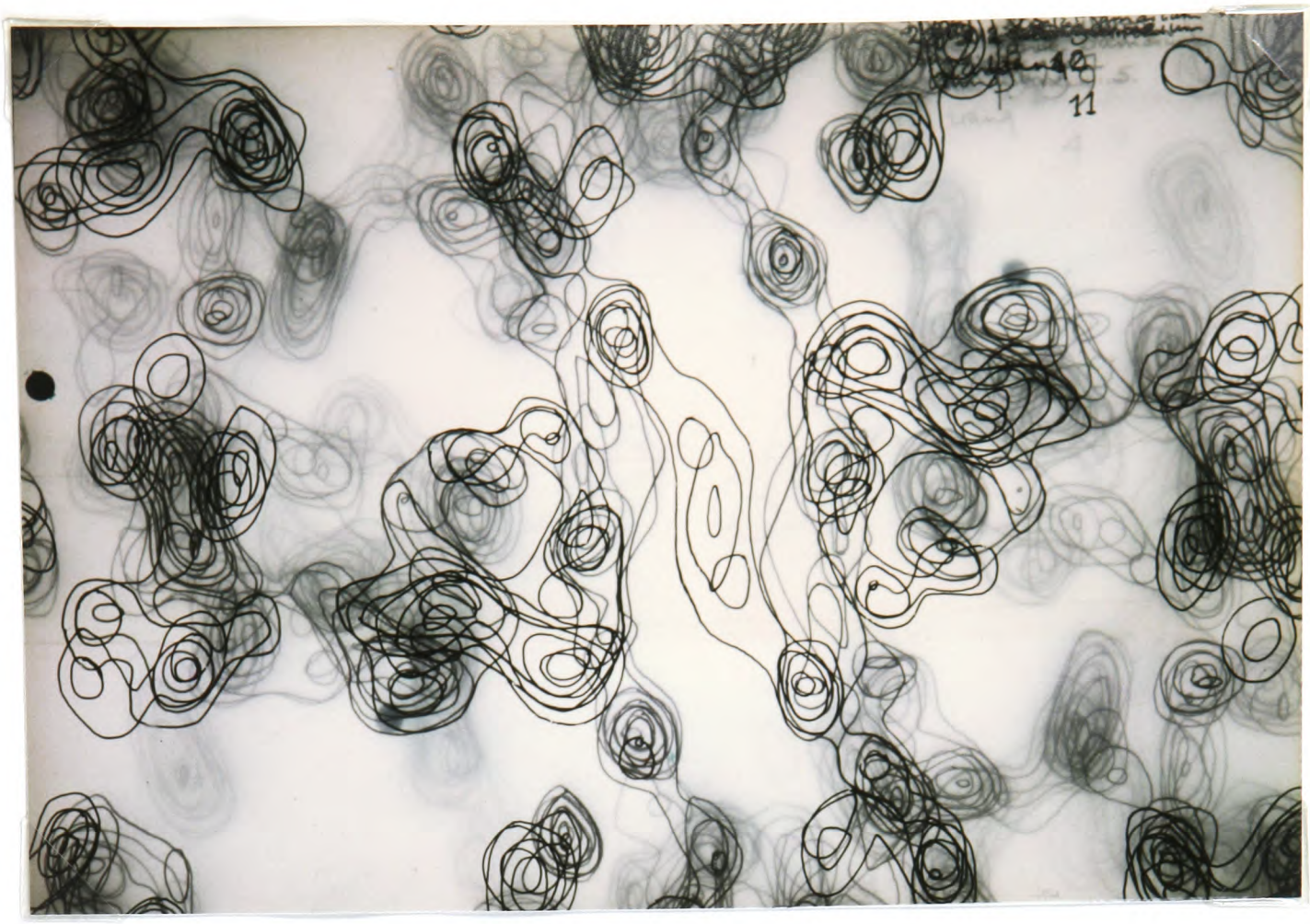
(b) $\Delta_{iso} > 0.3($ average of F_P and F_{PH} $)$.

correlation coefficients after each iteration step. During the course of iteration the average phase shifts were reduced from about 40° to less than 3° and the overall figures-of-merit increased from about 0.2 to 0.7. Final statistics from the map modification procedure are compiled in table 6.1. Experience suggests that R-factors of about 20% represent good agreement between calculated and observed structure factor amplitudes. Values above 30% give cause for concern (Huber, 1985). The values presented in table 6.1 indicate that the calculated structure factors agree tolerably well with the observed data.

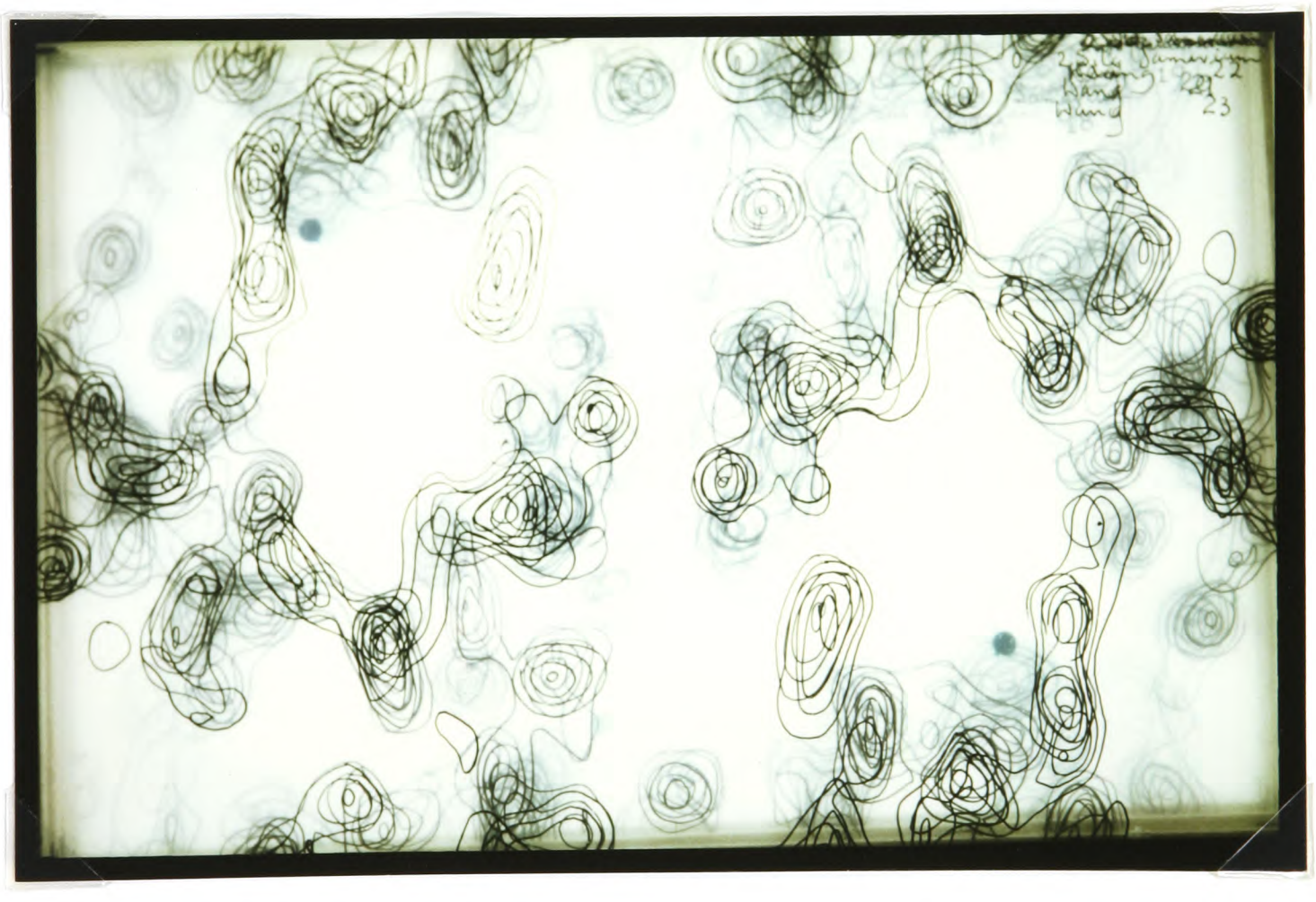
Electron density maps were calculated for each derivative as described in chapter five. Photographs of the calculated maps are shown in figure 6.2. The uranyl map was immediately interpretable in terms of two unique monomers in the asymmetric unit with the molecular diads lying in coincidence with crystallographic diads parallel to the z (stacking) axis (figures 6.2a and b). The proposed monomer density features were shown to be of the right shape and size by overlaying an approximately scaled alpha carbon plot of an FeSOD dimer (figure 6.2o). SIR maps calculated from the samarium (figures 6.2c and d) and gold (figures 6.2e and f) data were dissappointing. The maps were characterised by stringy density features and compact density globules which were too small to be interpretable as protein monomers. The presence of the gold cluster feature within a prominant dense portion of the map (figure 6.2f) was another discouraging feature of the maps. The three derivative isomorphous map (with anomalous data excluded from the phase calculation) was also modified by the Wang procedure. The resultant map was very similar to the original map. The lack of improvement was not surprising because the original map had significant portions of featureless regions so the solvent flattening algorithm did not act as a very powerful constraint in the phase calculation. Further it was apparent that the map was being adversely affected by the poor quality of the samarium and gold derivatives as evidenced by the quality of the SIR

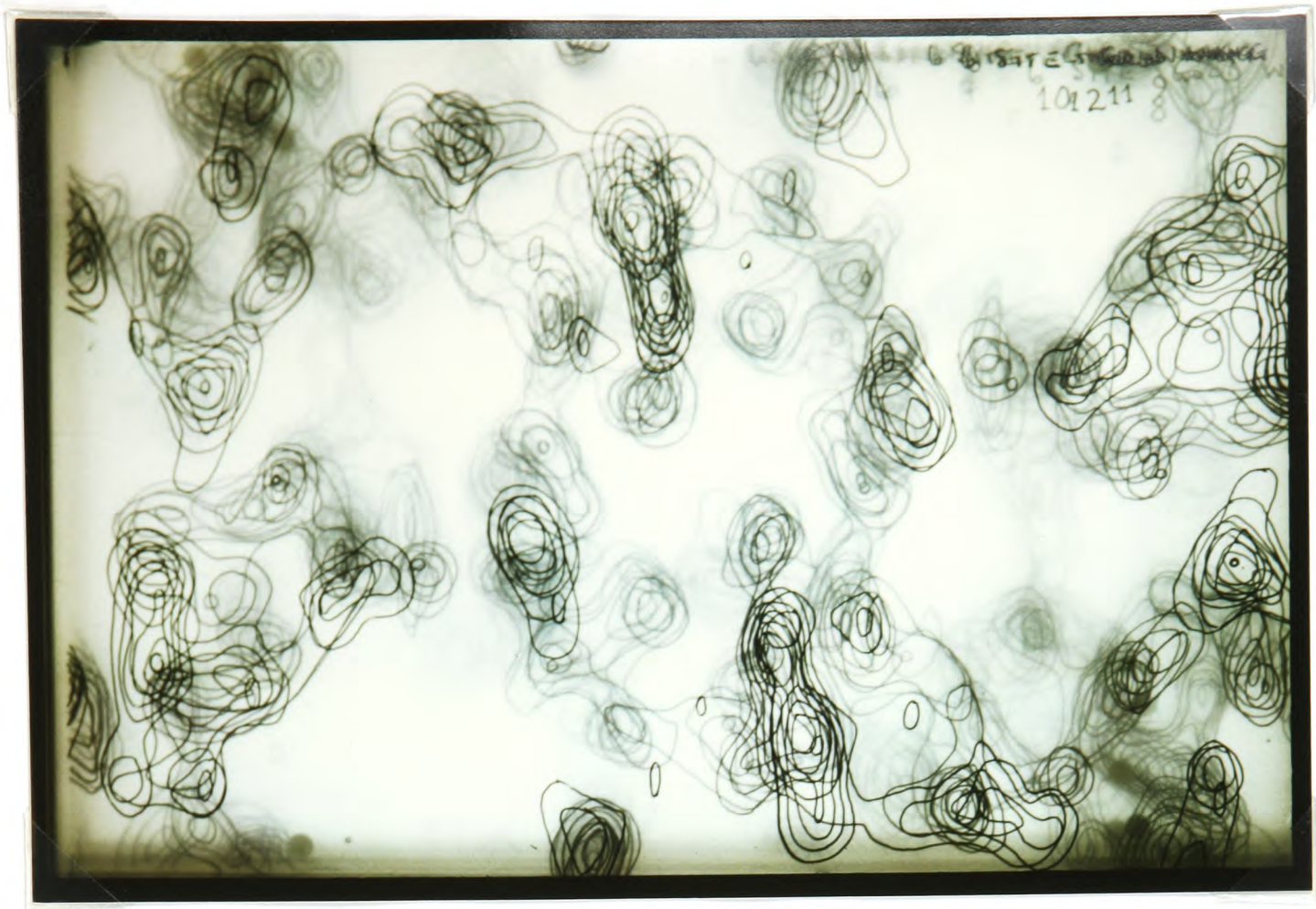
(a), (b) Map calculated from uranyl acetate SIR data.

Figure 6.2 Electron density maps calculated by the Wang map modification procedure. The maps were calculated to 6Å resolution for a whole unit cell and contoured at the same equal but arbitrary intervals. The map stacks are presented in two parts: 0 to 12/24 and 12/24 to 1 of the unit cell edge along the z (parallel to c) sectioning axis. The origin of the maps is at the top left hand corner (on section zero) and the longer map edge corresponds to the direction of the y (parallel to b) axis. Heavy atom positions are marked with dots.

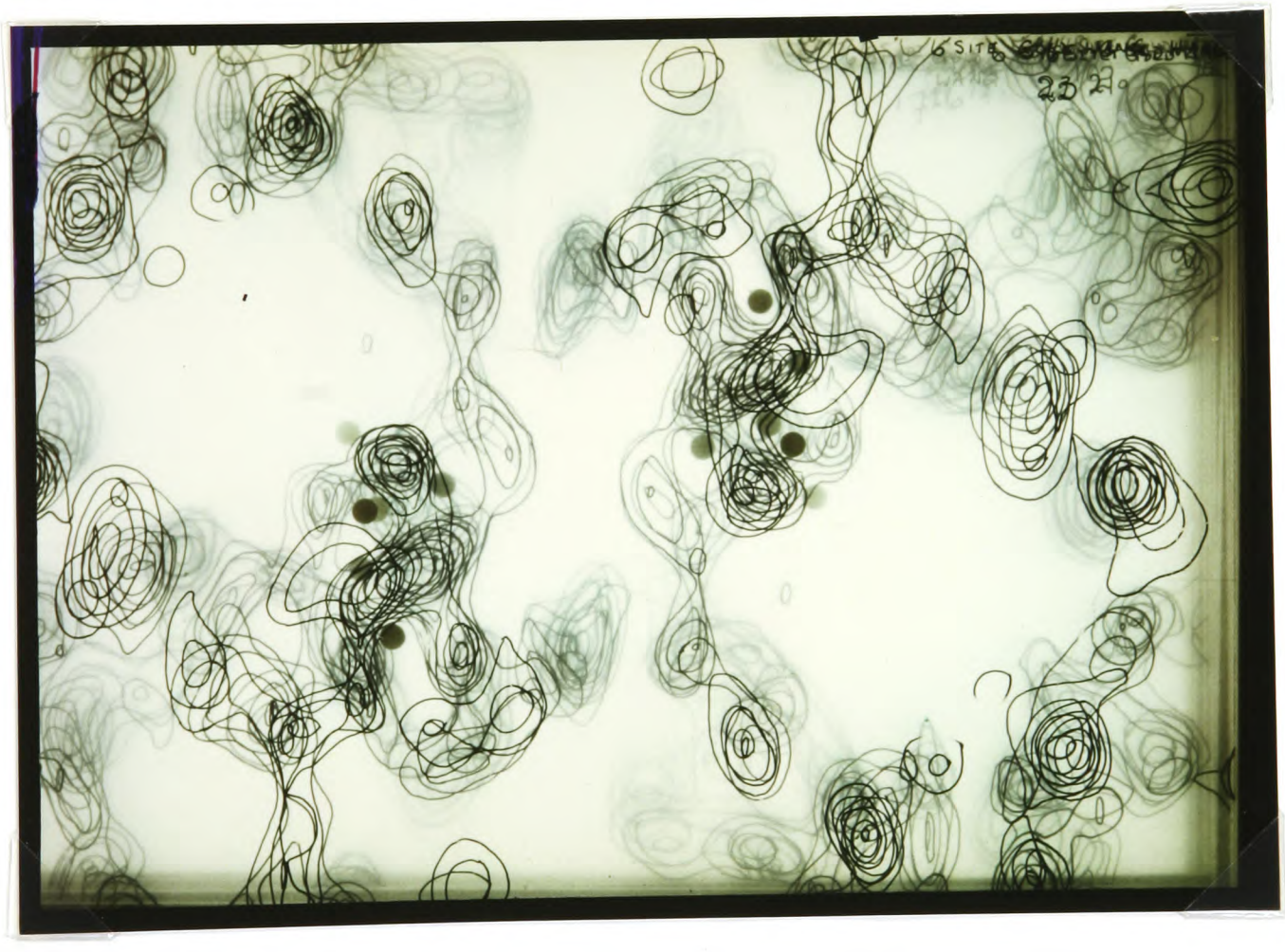


(c), (d) Map calculated from samarium acetate SIR data.





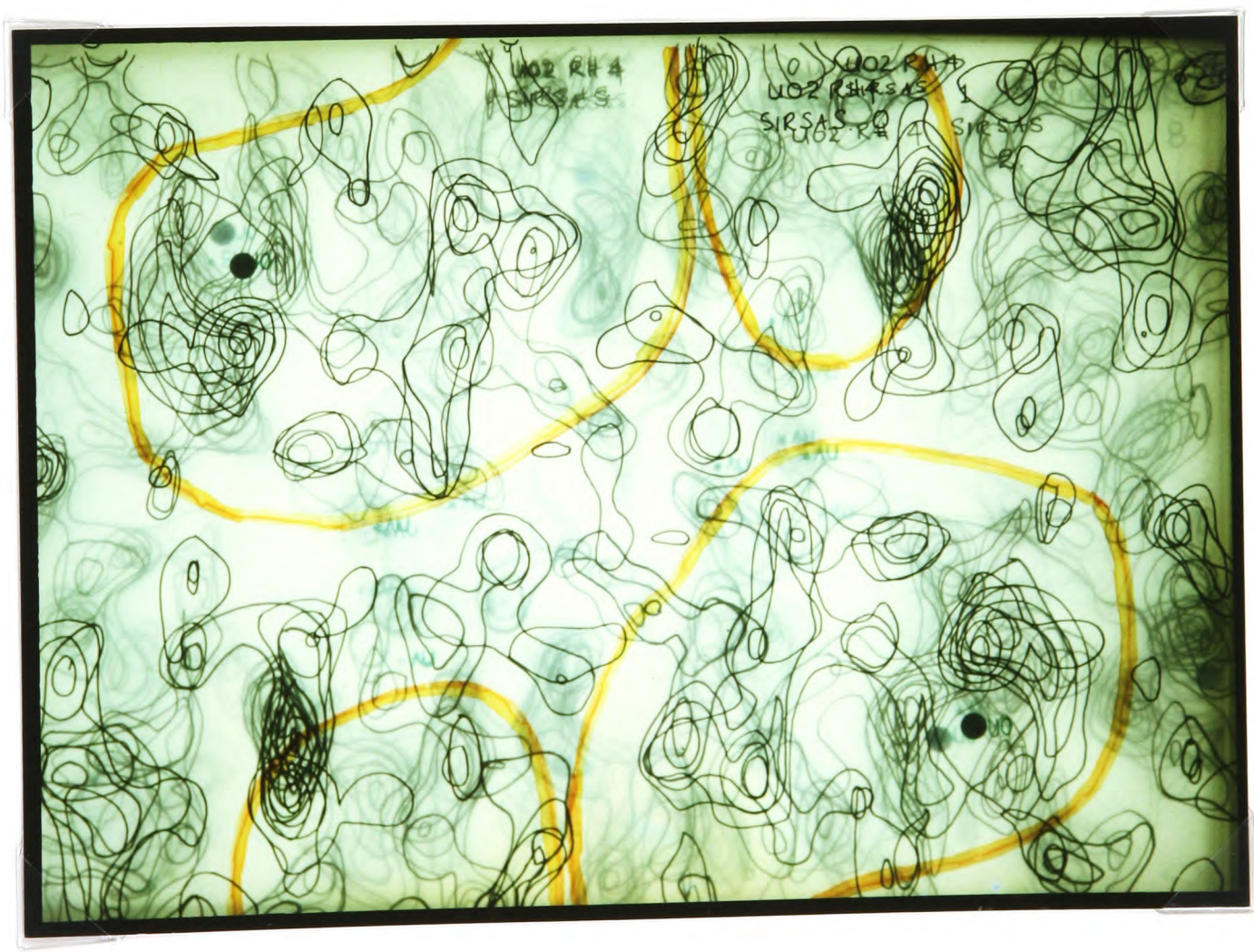
(e), (f) Map calculated from gold chloride SIR data.

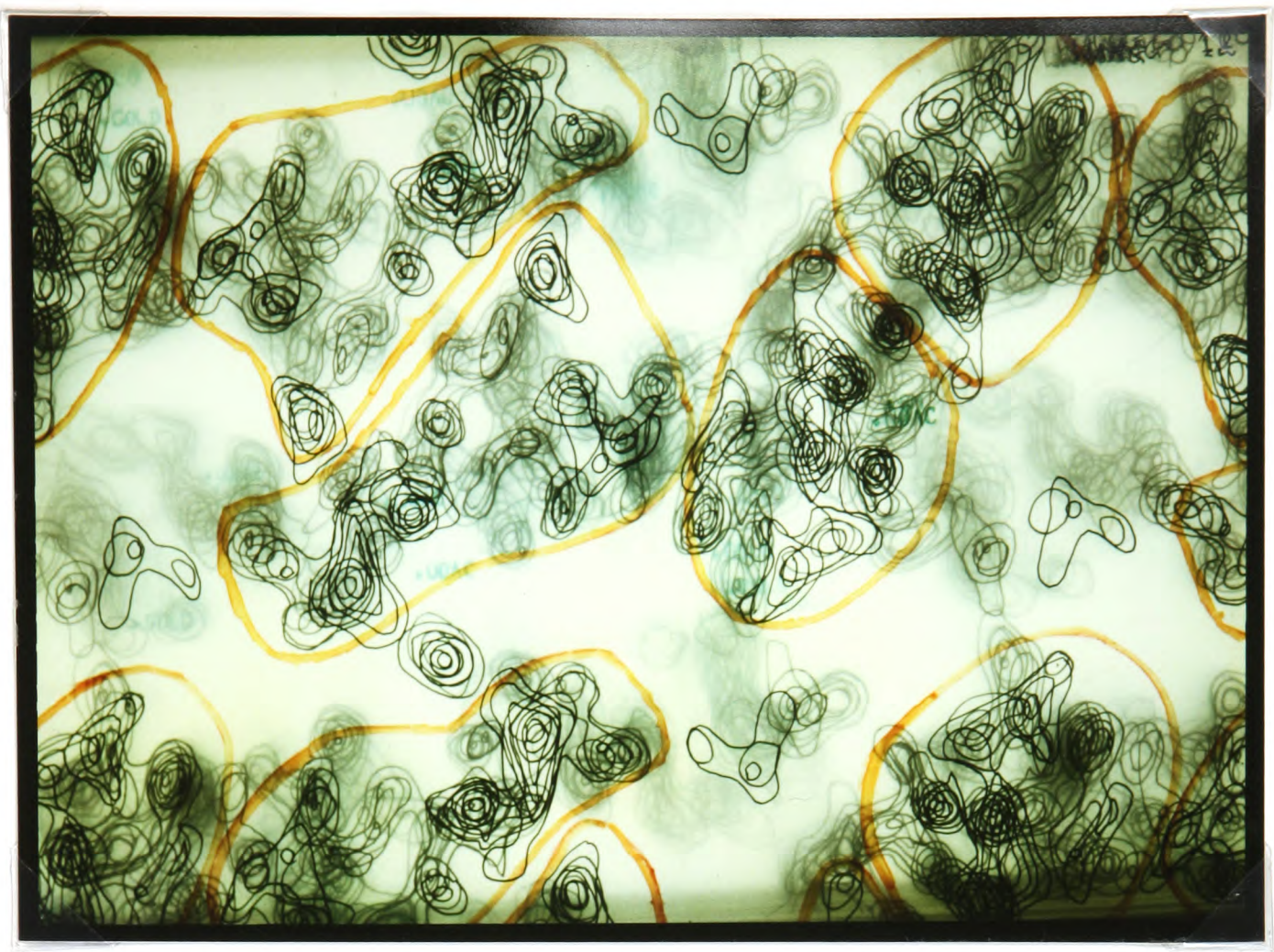


(g), (h) Map calculated from uranyl acetate SIR data - solvent content set to 40% of unit cell.

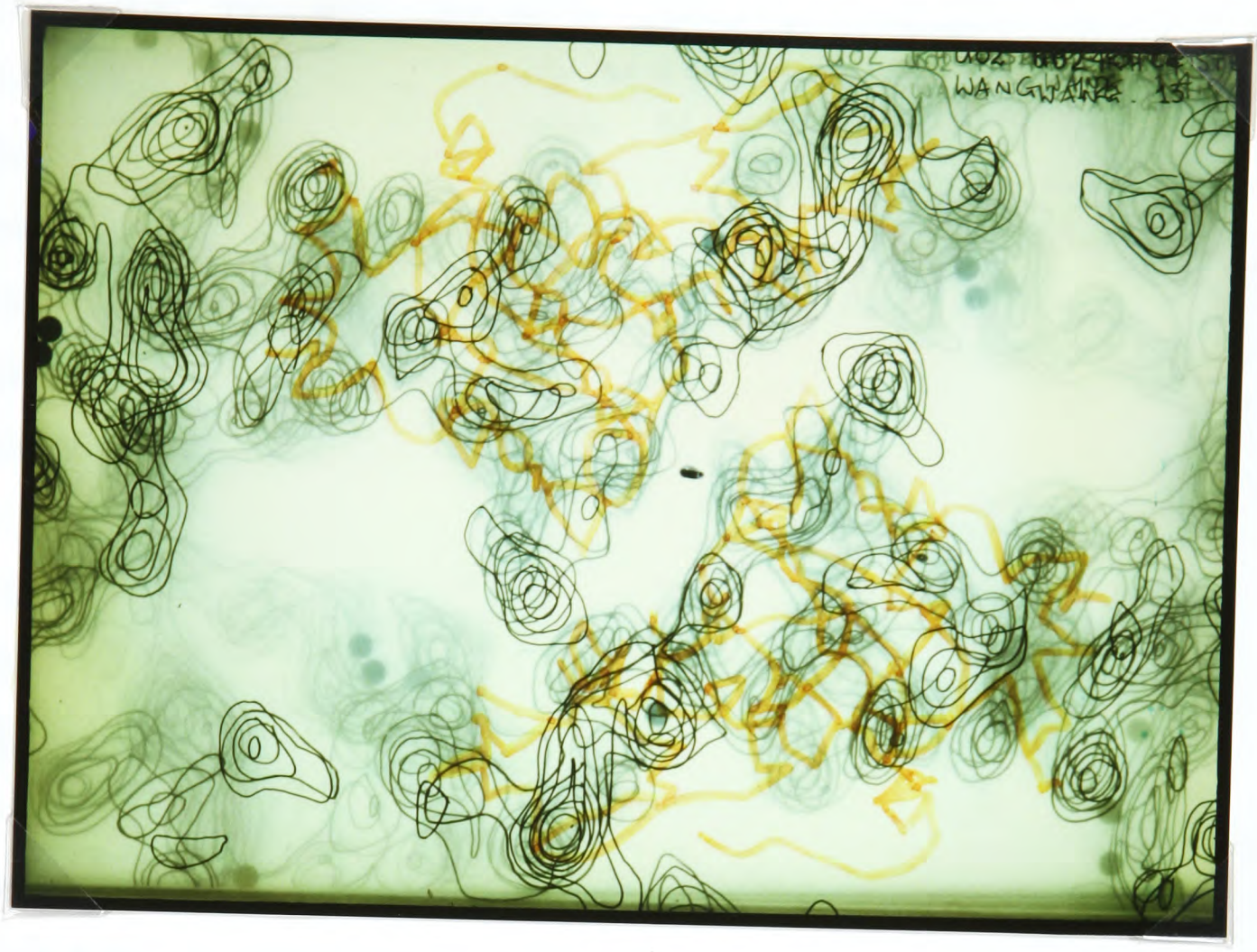
(i), (j) Unmodified map calculated from uranyl acetate data using the 'anomalous-retarded' phase set to resolve the phase ambiguity.

(k), (l) Unmodified map calculated from uranyl acetate data using the 'anomalous advanced' phase set to resolve the phase ambiguity.





(m), (n) Map calculated from uranyl acetate SIR data - calculated from 0 to 1.5 unit cells in the x and y map axis directions.



(o) Map calculated from uranyl acetate SIR data – with a projection plot of the alpha carbon model of FeSOD overlayed on the map.

Wang maps.

There was a fear that the solvent content of the crystal had been overestimated. This would result in truncation of protein density at the protein-solvent interface. However, the larger the volume that can be associated with the solvent regions, the more powerful the constraint. Rayment (1983) showed the gains made in phase improvement by using as large a solvent fraction as possible made it well worthwhile setting the solvent content as high as reasonably likely. Nevertheless, the uranyl map was recalculated with a lower solvent content of 40% (figure 6.2g and h) as a test. The maps calculated using different solvent contents were very similar; the molecular boundaries were almost identical and only a few density continuities had changed. No attempt was made to include anomalous scattering information towards the calculation of the Wang maps. Conventional electron density maps calculated using both isomorphous and anomalous data from the uranyl derivative were noisy and of poor quality (figure 6.2i, j, k, and l). However, the map calculated with ''anomalous-retarded'' phases gave clearer solvent regions (in the same places as the final uranyl Wang map) than the ''anomalous-advanced'' maps. It was shown later (section 6.2) that the former map was indeed the one calculated on the correct hand.

The phases from the modified uranyl maps were used to calculate difference Fourier maps of the other derivatives. The major site of the gold derivative was correctly predicted. However, interpretation of maps from the other cationic derivatives was complicated by bad ''ghosting'' from the uranyl heavy atom sites. The EMP map was very noisy and possible sites were found not to agree with Patterson maps. A double difference map of the uranyl derivative was flat.

The uranyl Wang map was recalculated from 0 to 1.5 unit cells on x and y (corresponding to the a and b cell edges respectively) and from 0 to 1 along the stacking axis (figures 6.2m and n). A number of promising

features were apparent. The molecules were packed in a layer structure with long solvent channels running parallel to the b axis and perpendicular to the a axis. The observations of cell dimension variability along the a axis and crystal cracking in planes parallel to the b–c crystallographic plane (see chapter three and five) could be explained in terms of the packing seen in the map. The positioning of the gold derivative in a solvent channel between two protein monomers agreed with proposals made in chapter five (sections 5.2 and 5.6) where it was thought that the measured intensity differences between the native and gold derivatives were dominated by scattering contributions from the presence of colloidal gold in the solvent channels of the crystal.

The fit of an approximately scaled alpha carbon plot of FeSOD overlayed on the final uranyl Wang map is shown in figure 6.2o. In general, the fit was very convincing but there were discrepancies. For example, the N-terminal tail of the model projected into the solvent regions. This discrepancy might be ascribed to a disordered tail in the MnSOD crystal or perhaps a real structural difference between the model and MnSOD molecule. For ease of discussion, the protein density shown in figure 6.2o will be referred to as the density of dimer 1 and the other unique portion of density in the unit cell referred to as the density belonging to dimer 2.

6.2 Fitting the Model to the Electron Density Map

Initial attempts to fit the coordinates of the FeSOD model to the new electron density map were made using an Evans & Sutherland PS 2 graphics system. The program FITZ was used to display chosen portions of the final uranyl Wang map (recalculated on a course grid with only one contour level because of computer memory limitations) and the model moved as a rigid body so as to superimpose it on the map. It became apparent from the asymmetric shapes of the model and the dimer 1 density envelope that the map had been calculated on the wrong hand. This was rectified and the fit of the model

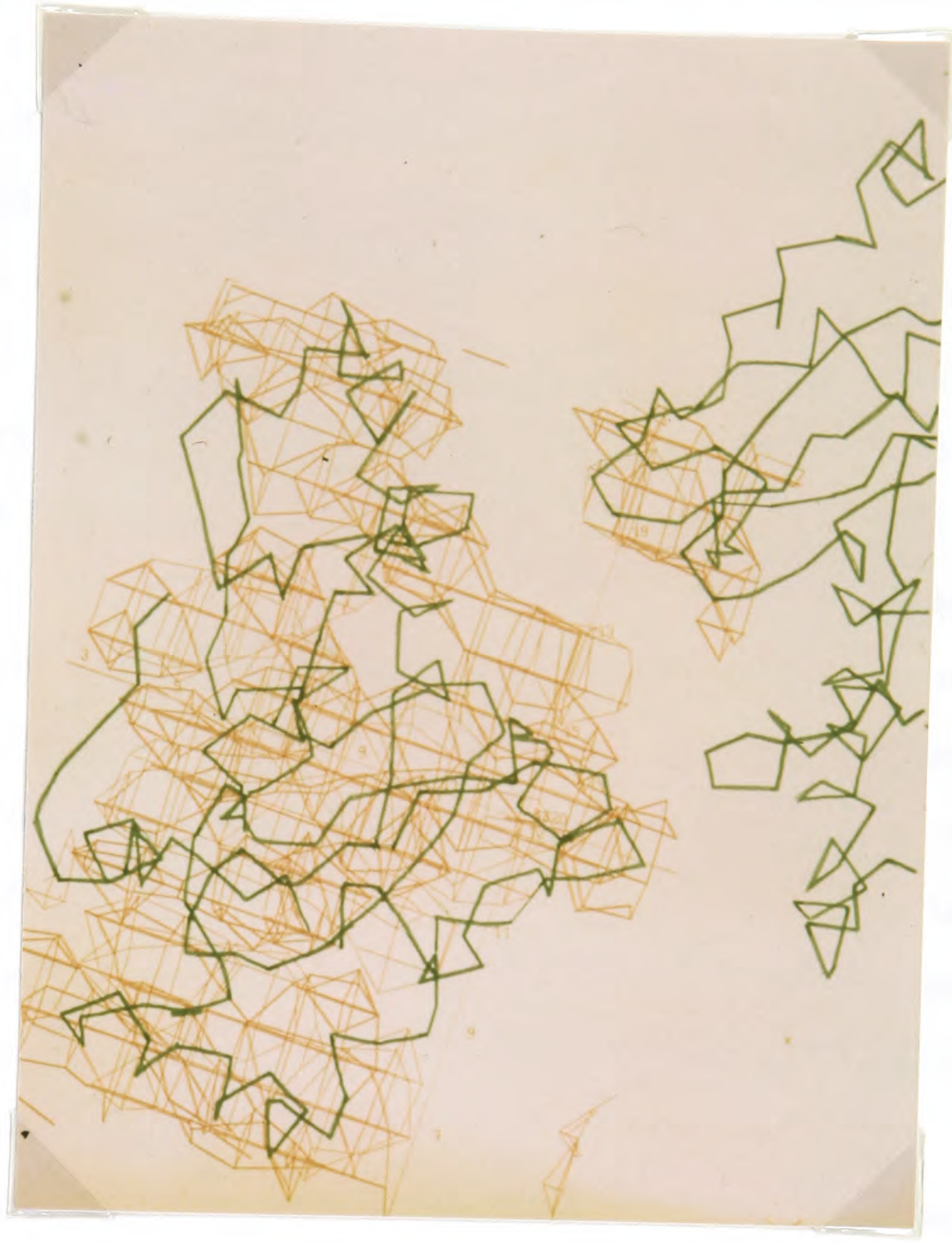


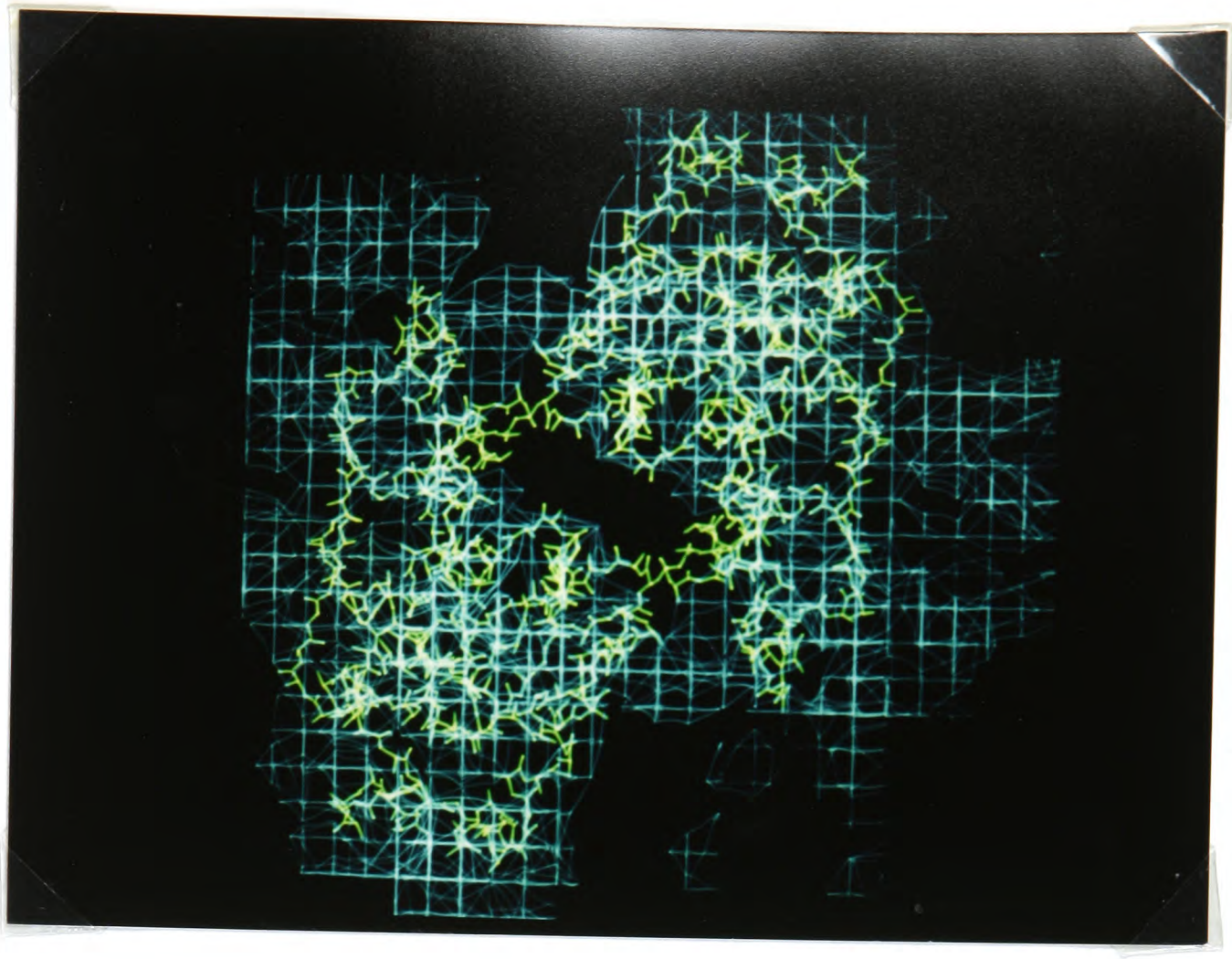
Figure 6.3 Fit of the alpha carbon model of FeSOD to the density of dimer 1 in the Wang map calculated from the uranyl acetate SIR data. (Generated on the PS 2 graphics screen using the program FITZ).

to the new map is shown in figure 6.3. The overall fit of the model was convincing but again it was apparent that the internal detail of the map was poor. This was not surprising considering the problems of poor isomorphism that had been encountered with the heavy atom derivatives (see chapter five).

At this stage an Evans & Sutherland PS 300 colour graphics system became available. This system offered facilities for displaying a lot more vectors than the PS 2 and the colour coding options proved useful in the model fitting efforts. Again, the fit of dimer 1 to the map was convincing; the monomer density was of the right shape and size and the Mn-Mn distances proved to be the same as found for the Fe atoms in FeSOD. Orthogonal views of the fit of dimer 1 to the map are shown in figure 6.4. Attempts to fit dimer 2 to the map were straightforward in the x and y directions but ambiguous in the z direction because of tight crystal packing with symmetry equivalent molecules of dimer 1. This problem was overcome by using the symmetry options of the programs SAM and FRODO to generate and display the appropriate symmetry equivalent molecules of dimer 1 (figure 6.5) and then fitting dimer 2 into the remaining density. Orthogonal views of the dimer 2 fit are shown in figure 6.6.

The overall interpretation of the map was convincing but there were a few worrying features such as the projection of some backbone atoms of the model into the solvent regions (for example, figure 6.6a). This problem might be attributed to real structural differences between the model and the MnSOD molecule or perhaps a partially incorrect envelope determination of a map calculated from a poorly isomorphous derivative. It was decided to improve the fit of the model to the map by reciprocal space rigid body refinement and thus avoid the problems associated with the poor quality of the protein phase set.

a



b

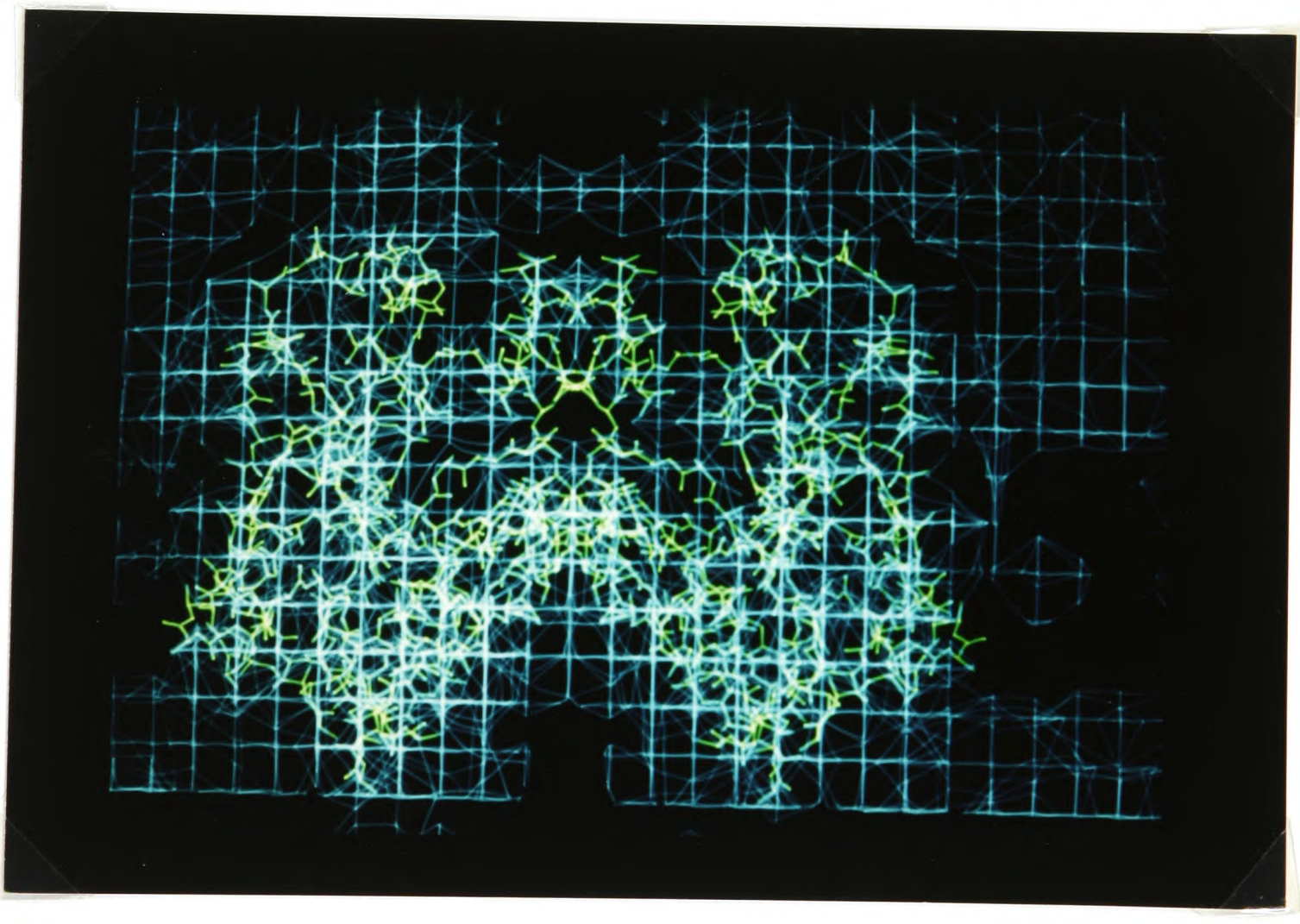


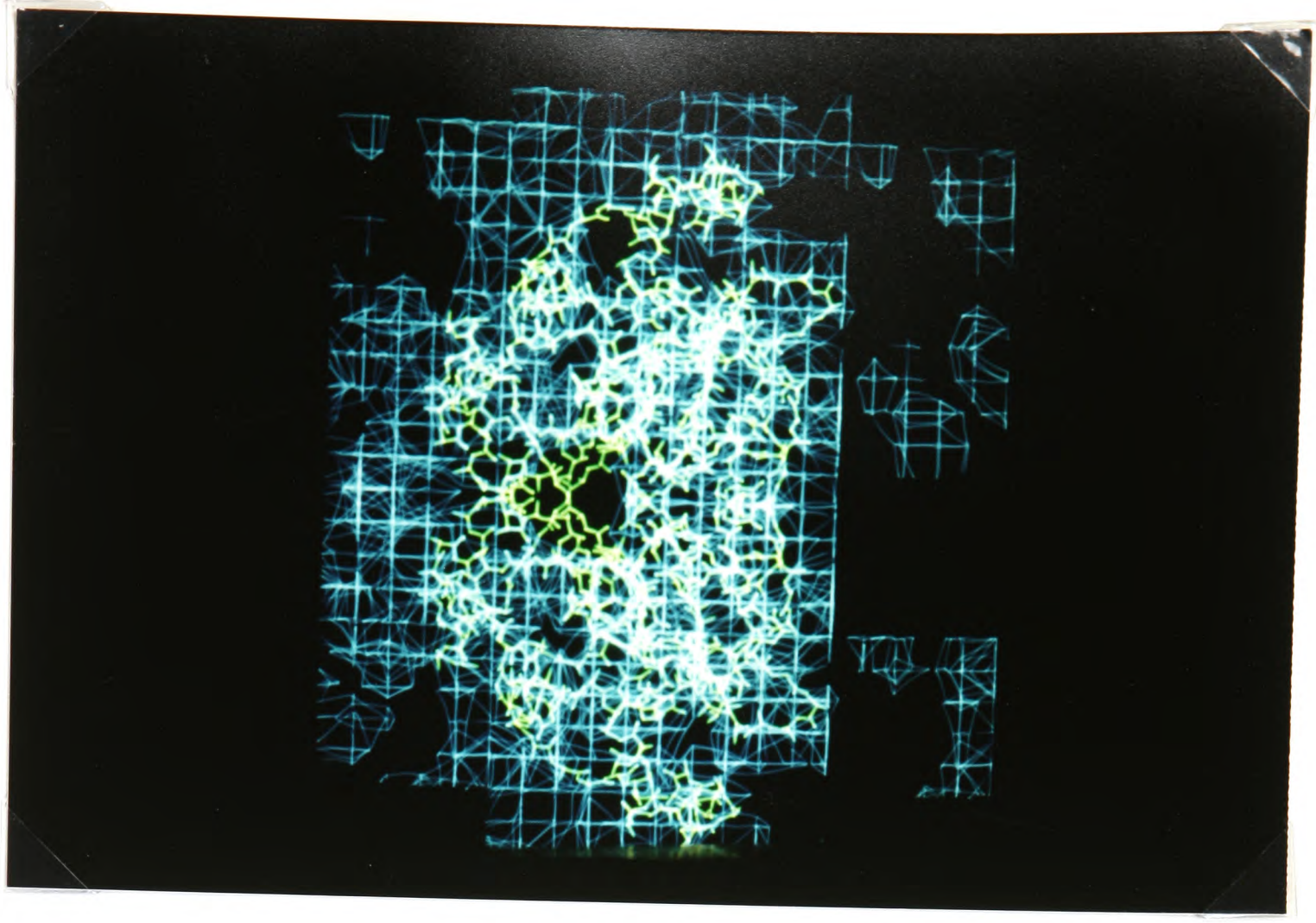
Figure 6.4 Fit of the FeSOD polyalanine model to the Wang map calculated from uranyl acetate SIR data.

(a) dimer 1, z view

(b) dimer 1, y view

(c) dimer 1, x view

C



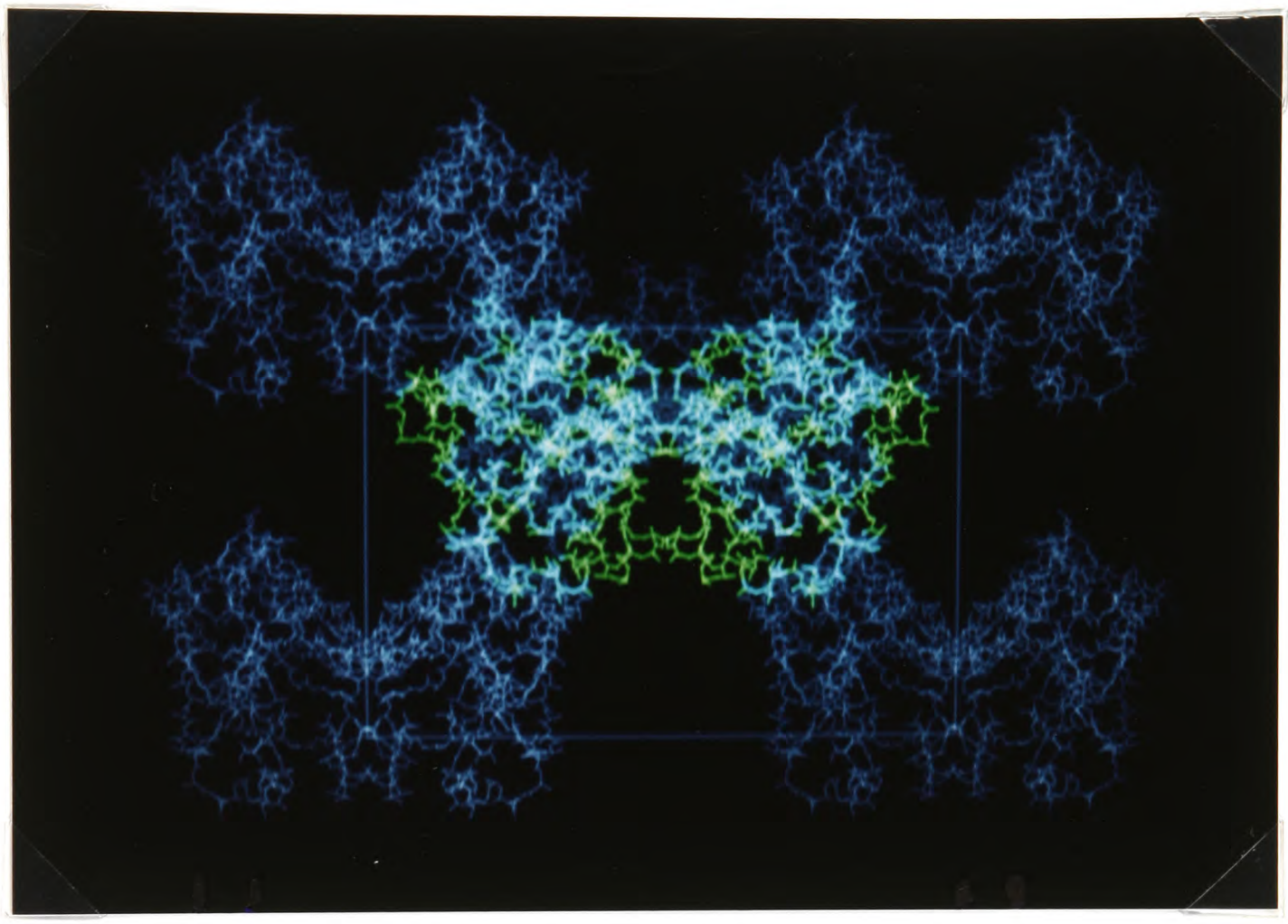
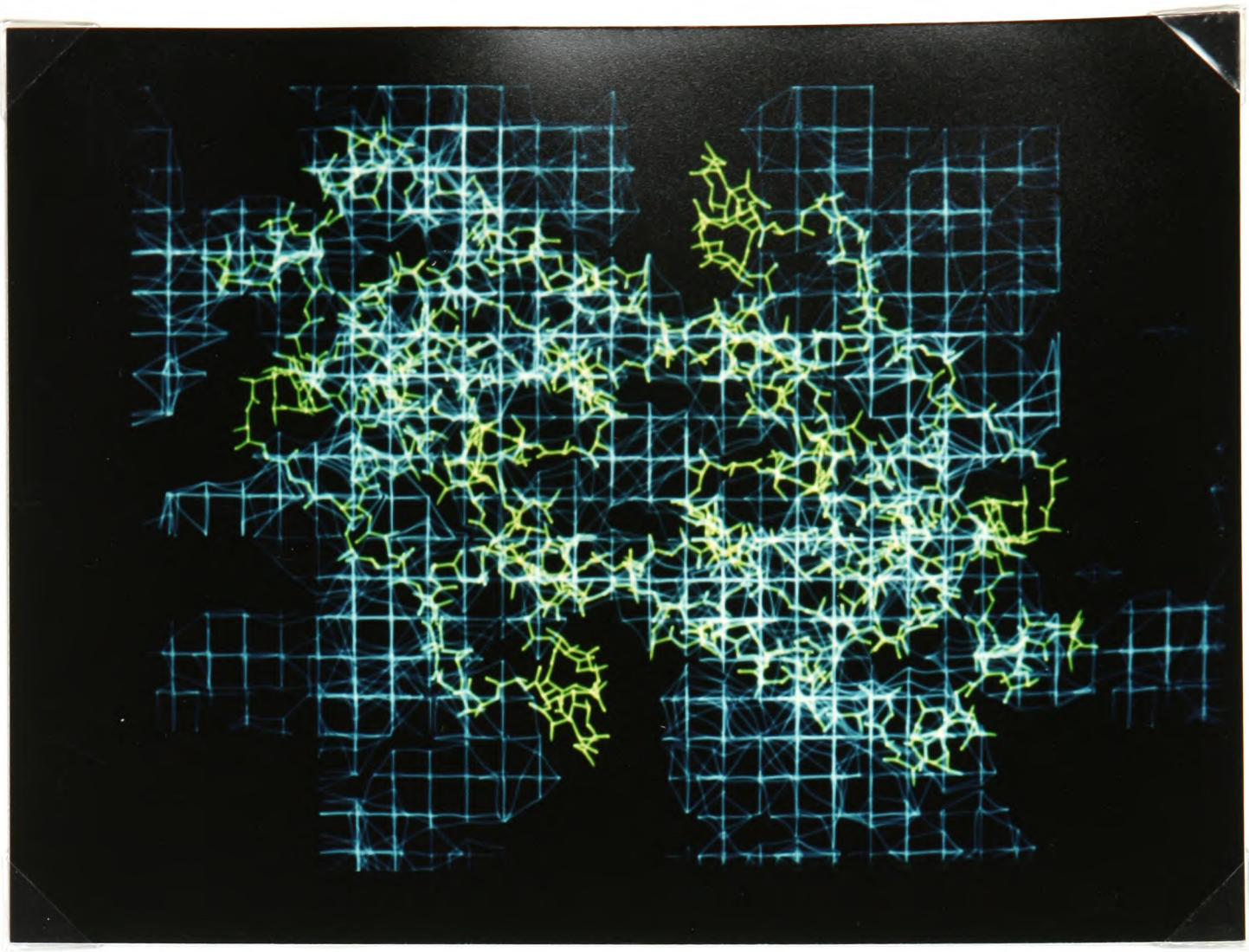


Figure 6.5 Packing of dimer 2 molecule (centre) with symmetry equivalent molecules of dimer 1.

a



b

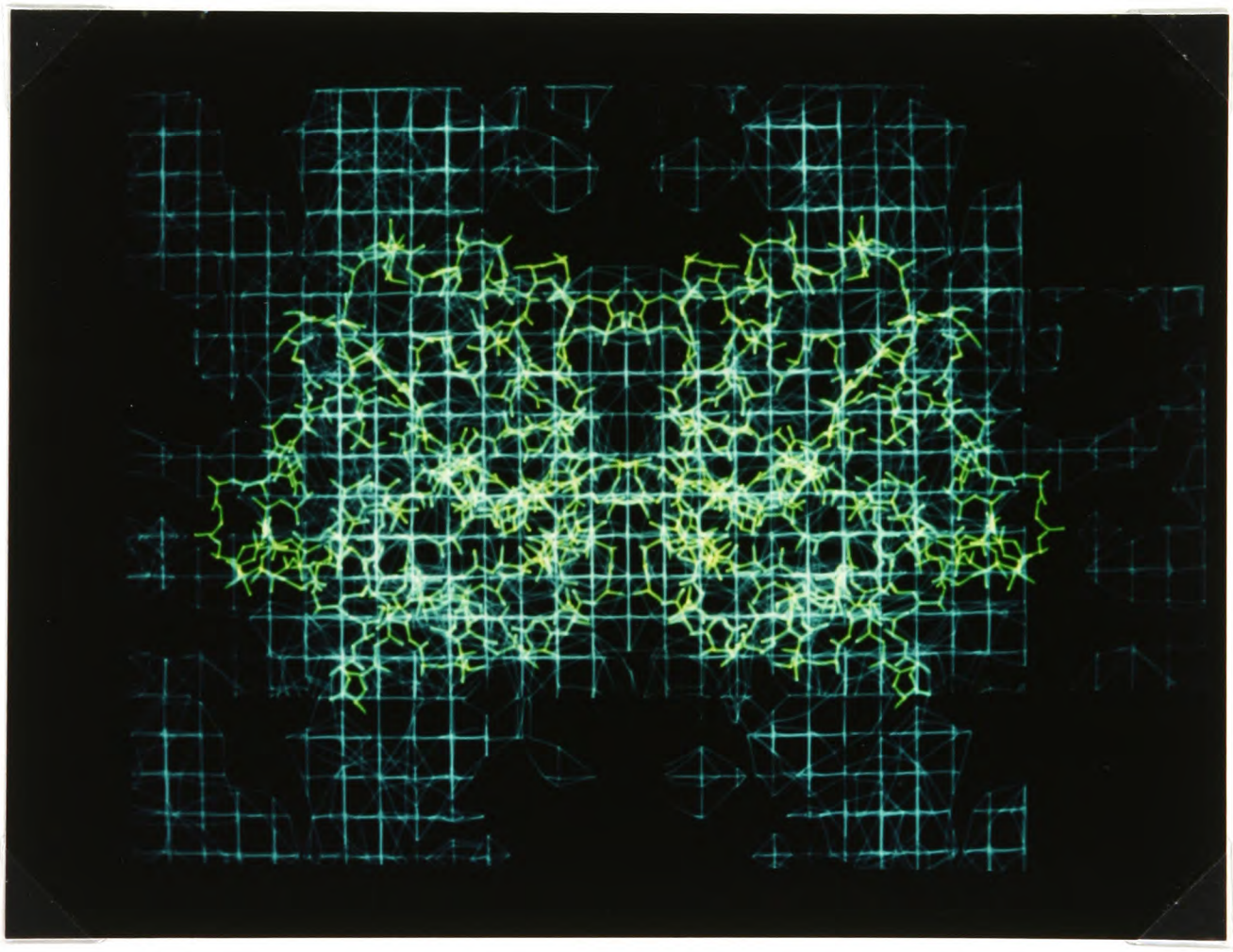


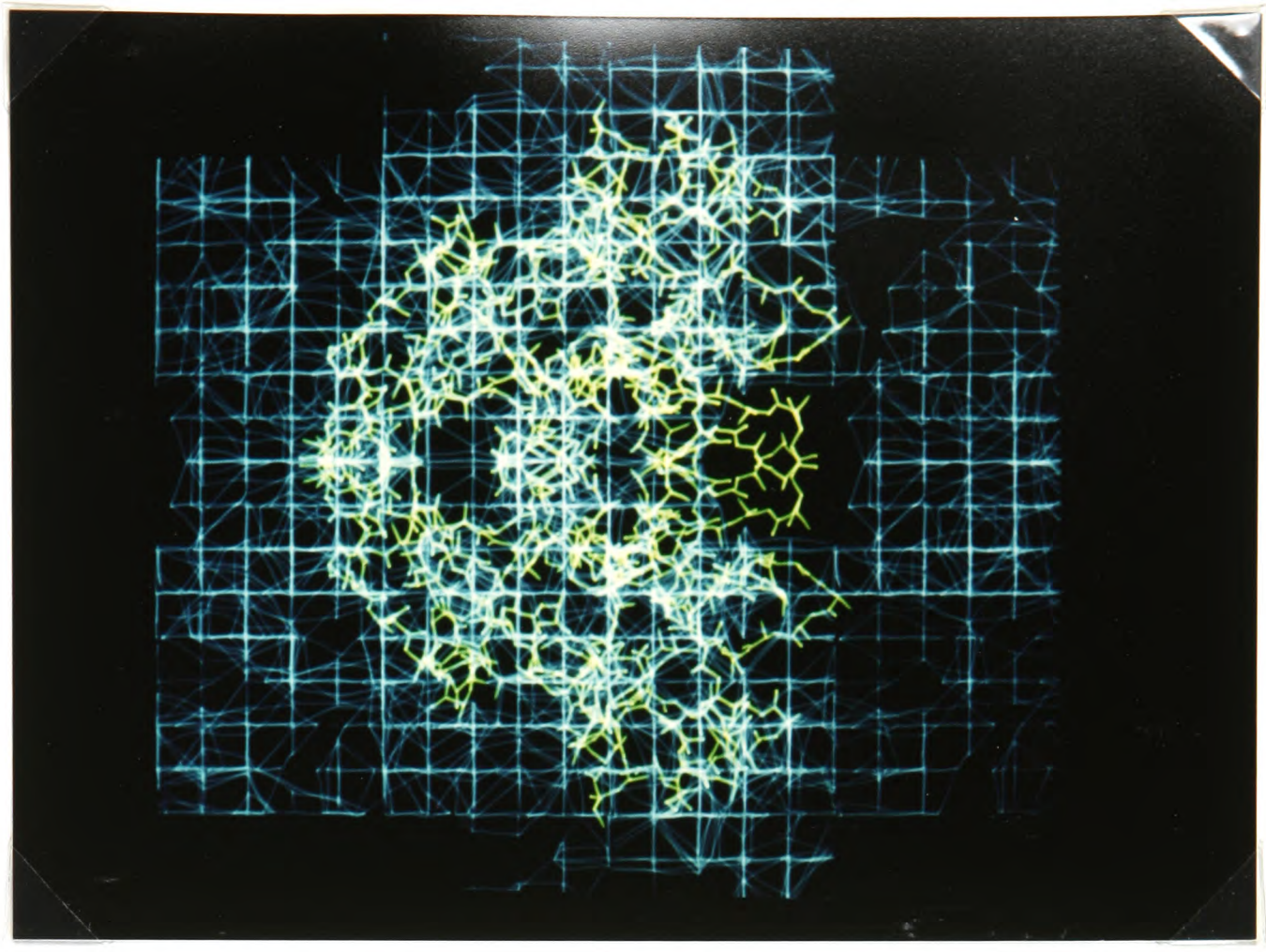
Figure 6.6 Fit of the FeSOD polyalanine model to the Wang map calculated from uranyl acetate SIR data.

(a) dimer 2, z view

(b) dimer 2, y view

(c) dimer 2, x view

C



6.3 Rigid Body Refinement

The graphics fit of the model into the density map was refined using the rigid body refinement program CORELS (see section 2.5.3). Native structure factor amplitude data were converted to card image and reformatted using the program BINTCI. The coordinate files of the two dimer molecules were edited and merged to produce a single file containing coordinates of the two unique monomers. The program PQRXYZ was used to convert the coordinates to fractional crystallographic form and the program PRECOR run to prepare input files for CORELS.

The model was divided into two constrained groups; each being the unique monomer. A file was generated by PRECOR that contained a list of non-bonded contact distances out to 4Å around each atom. These contact distances were minimised during the course of refinement by use of appropriately weighted restraints. The scattering factors were corrected for solvent effects (as described in section 2.5.3) using a solvent density value of 0.28 electrons/$Å^3$ and a temperature factor of $100Å^2$.

An initial round of five refinement cycles were performed in order to calculate an overall scale factor between the observed and calculated structure factor data. The work on the graphics had suggested that the fit of the dimers in the *x*-*y* plane was reasonable but there was some ambiguity regarding the fit in the z direction. Thus the first few rounds of refinement were performed allowing only movements in the z direction. General rigid body rotations and translations were allowed to refine in latter rounds.

Many test runs were performed in which different parameters were allowed to refine and different data shells were used. Refinement runs were started with data from low resolution shells to ensure the radius of convergence was large (about 3 to 4Å). Even with very low resolution data shells (30-12.5Å) the high observation-to-parameter ratio ensured the solution to the normal equations was overdetermined. Refinement statistics from the most satisfactory run are compiled in table 6.2. The overall R-

Round	Resolution Shell (A)	Q	Number of cycles	Overall Scale	centre-of-gravity translation x	y	z	centre-of-gravity rotation ϕ	θ	ρ	R	c.c.
Dimer 1												
1	30 - 10	16	20	0.92	0.34	0.43	0.83	0.90	-0.17	-5.96	0.51	0.23
2	30 - 8	34	10	0.84	0.34	0.43	0.82	-0.41	1.29	-4.56	0.53	0.20
3	30 - 7	52	10	0.83	0.33	0.42	0.81	-1.14	3.19	-6.39	0.52	0.24
4	30 - 6	85	11	0.79	0.33	0.42	0.80	-1.07	2.58	-8.74	0.53	0.26
Dimer 2												
1	30 - 10	16	20	0.92	0.31	0.03	0.73	5.51	3.18	-6.55	0.51	0.23
2	30 - 8	34	10	0.84	0.33	0.03	0.73	4.82	3.94	-9.72	0.53	0.20
3	30 - 7	52	10	0.83	0.33	0.04	0.73	1.66	5.71	-11.14	0.52	0.24
4	30 - 6	85	11	0.79	0.34	0.03	0.73	2.71	5.30	-10.95	0.53	0.26

Table 6.3 Refinement statistics - native type 1.

Key: Q = (number of observations) / (number of parameters)

$R = \Sigma_{hkl} | F_{obs} - F_{calc} | / \Sigma_{hkl} F_{obs}$

c.c. = correlation coefficient

$= \langle F_{obs} F_{calc} \rangle / (\langle F_{obs} F_{obs} \rangle \langle F_{calc} F_{calc} \rangle)^{1/2}$

translations in fractions of unit cell edges and rotations as a function of Euler angles.

Note: Starting model position was:

dimer 1 $x = 0.33$, $y = 0.41$, $z = 0.80$, $\Phi = 0^\circ$, $\theta = 0^\circ$ and $\rho = 0^\circ$

dimer 2 $x = 0.29$, $y = 0.03$, $z = 0.70$, $\Phi = 0^\circ$, $\theta = 0^\circ$ and $\rho = 0^\circ$

Round	Resolution Shell (Å)	Q	Number of cycles	Overall Scale	centre-of-gravity translation x	y	z	centre-of-gravity rotation Ø	Θ	Φ	R	C.C.
Dimer 1												
1	30 - 10	21	10	1.09	0.33	0.41	0.82	-3.52	6.14	-3.12	0.58	0.12
2	30 - 8	40	10	0.96	0.33	0.41	0.82	-1.53	5.44	-1.75	0.58	0.05
3	30 - 7	60	10	0.94	0.32	0.41	0.81	-2.32	4.20	-0.67	0.58	0.11
4	30 - 6	90	11	0.90	0.32	0.42	0.82	-2.35	4.32	0.10	0.59	0.17
Dimer 2												
1	30 - 10	21	10	1.09	0.27	0.03	0.73	3.13	6.18	8.14	0.58	0.12
2	30 - 8	40	10	0.96	0.28	0.02	0.74	4.40	6.76	5.23	0.58	0.05
3	30 - 7	60	10	0.94	0.28	0.02	0.74	6.56	9.05	3.83	0.58	0.11
4	30 - 6	90	11	0.90	0.28	0.02	0.74	6.15	9.03	3.96	0.59	0.17

Table 6.2 Refinement statistics - native type 2.

Key: Q = (number of observations) / (number of parameters)

$R = \Sigma_{hkl} | F_{obs} - F_{calc} | / \Sigma_{hkl} F_{obs}$

c.c. = correlation coefficient

$= \langle F_{obs} F_{calc} \rangle / (\langle F_{obs} F_{obs} \rangle \langle F_{calc} F_{calc} \rangle)^{1/2}$

translations in fractions of unit cell edges and rotations as a function of Euler angles.

Note: Starting model position was:

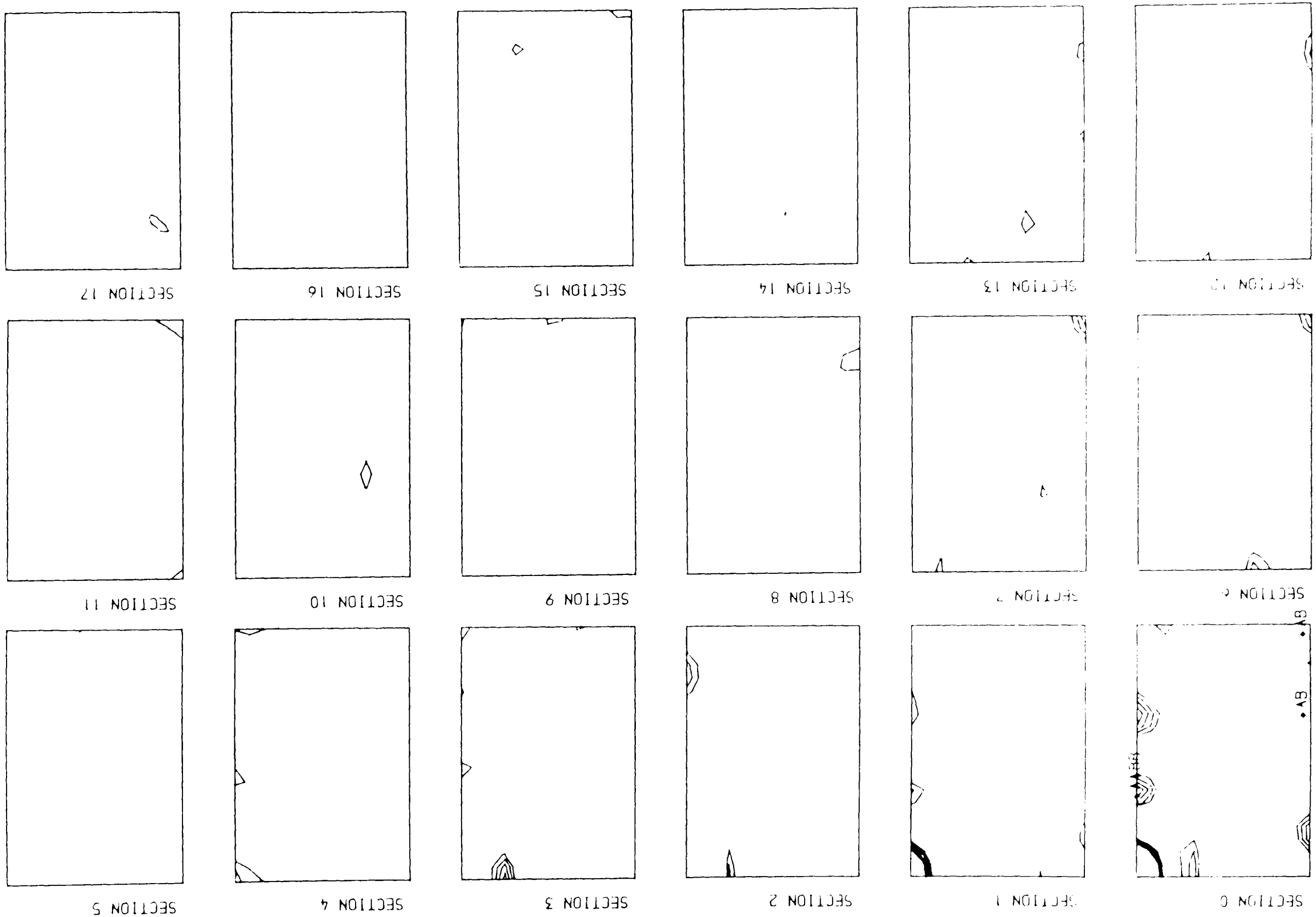
dimer 1 $x = 0.33$, $y = 0.41$, $z = 0.80$, $\Phi = 0^\circ$, $\theta = 0^\circ$ and $\rho = 0^\circ$

dimer 2 $x = 0.29$, $y = 0.03$, $z = 0.70$, $\Phi = 0^\circ$, $\theta = 0^\circ$ and $\rho = 0^\circ$

factors and correlation coefficients were disappointing. However, the final R-factor was lower than the random R-factor of 66% expected for the particular combination of centric and acentric reflections in space group $P2_12_12$. The R-factor value is the same as Sussmann (1980) obtained for refinement of yeast initiator transfer RNA using 6Å data before the RNA model was divided into a number of constrained groups. The refined position of the FeSOD model gave a slightly better packing fit compared with the original model when examined on the graphics screen. The root mean square shift of the alpha carbon coordinates from the input coordinates was about 2Å for the dimer 1 and nearly 4Å for the dimer 2.

The model was also refined against the native type 1 data (see chapter four). It was hoped that the molecular mechanism of the type change could be determined from this analysis. Statistics of the refinement run are compiled in table 6.3. The statistics were better than those obtained with the other native type but were not good enough to feel confident that a true minimum had been found (D. Stuart, pers. commun.) and so a detailed analysis of packing changes between the native types was not carried out. The root mean square shift of the alpha carbon coordinates from the input coordinates was 2.3Å for dimer 1 and 4.0Å for dimer 2.

A final word needs to be said about the quality of the model and its effect on the R-factor. The FeSOD model consisted of a polyalanine trace of 186 residues. The two iron atoms had to be deleted from the model for the purpose of refinement because of programming problems. In contrast, the unknown protein is known to consist of 215 residues and up to two manganese atoms. Thus, although the Fe and MnSODs were thought to be structurally related (see chapter one and seven), a significant proportion of scattering matter is absent from the model used in these studies (approximately 46% of the main and side chain non-hydrogen atoms were absent). Further, the model was unrefined and poorly idealised. The R-factors calculated in this work were outside the range of values (40 to 50%, Sussman (1980)) expected


SECTION 17
SECTION 16
SECTION 15
SECTION 14
SECTION 13
SECTION 12
SECTION 11
SECTION 10
SECTION 9
SECTION 8
SECTION 7
SECTION 6
SECTION 5
SECTION 4
SECTION 3
SECTION 2
SECTION 1
SECTION 0

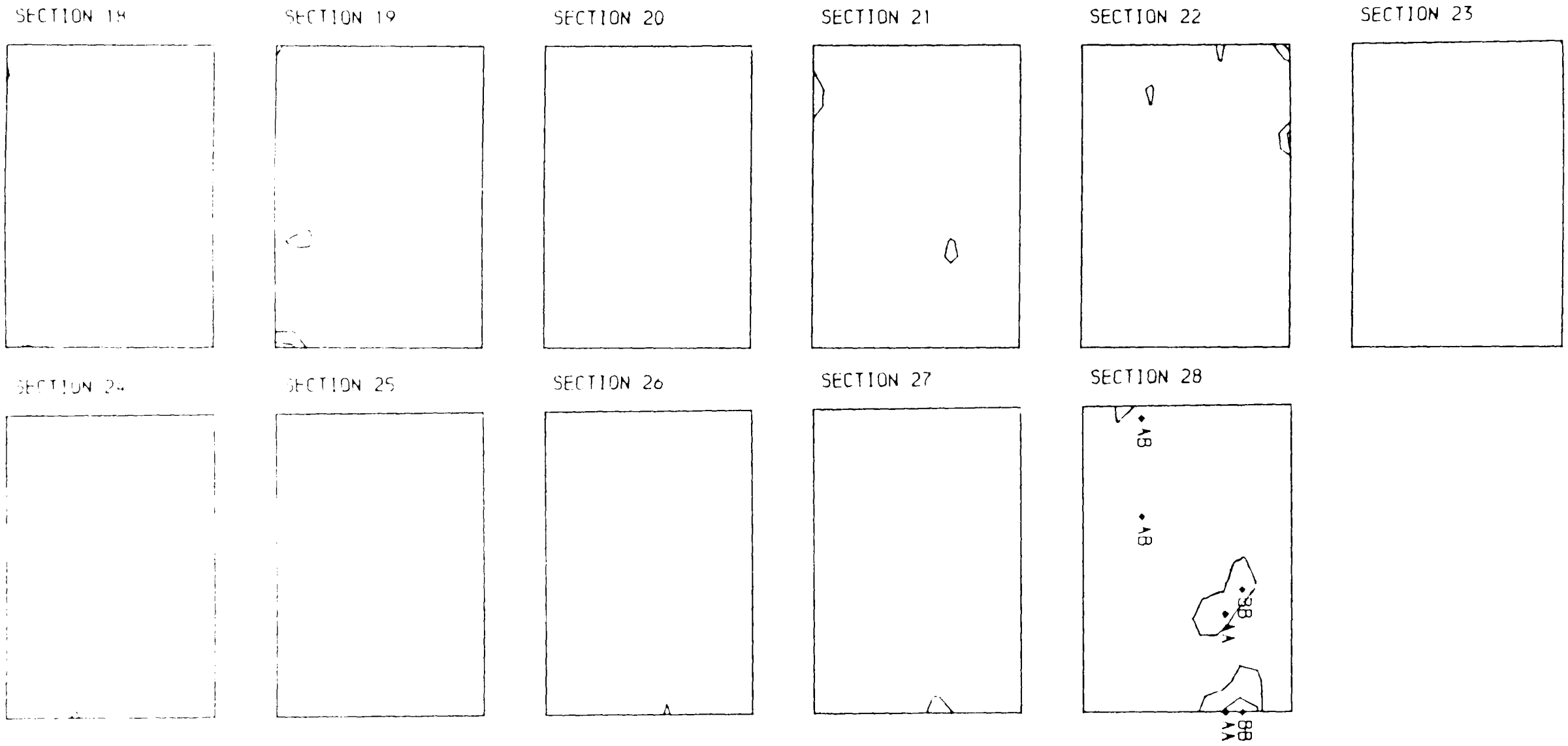


Figure 6.7 Sections of the isomorphous difference Patterson calculated from native and manganese-free data. The map was calculated to 6Å resolution and sectioned on the v axis (section interval of b/56). The long and short edge of each section correspond to the u and w axes, respectively. The map was calculated from 0 to 1/2 on each axis and contoured at equal but arbitrary intervals with the first level omitted. The origin is at the top left hand corner of section 0. The positions of self- (AA and BB) and cross-vectors (AB) derived from the positions of the metal sites in the model are also shown.

for rigid body refinement using a good model. Nixon and North (1976) have discussed the effect of incomplete and inaccurate models on reciprocal space residuals. It is clear that the shortcomings of the model were a contributing factor towards the rather high value of the R-factors calculated in these studies.

6.4 The Model and the Manganese-free Patterson

If the model is a good representation to the crystal structure of MnSOD then the iron atoms of the model might be in closely similar positions to the manganese atoms in MnSOD. In fact there is firm experimental evidence that the metal atoms occupy the same positions in the two different proteins (Stallings et al., 1984). It follows that the position of the metal atoms in the model should solve the manganese-free Patterson discussed at the end of chapter four. The most obvious interpretations of the Patterson are of two manganese sites within the asymmetric unit of the unit cell, both of which must lie within a 10Å radius of the crystallographic diads that lie in the direction of the z axis of the unit cell. It follows that the diad of the model must lie on the crystallographic diads because of symmetry considerations.

It turns out that the metal positions of the model solve the Patterson in a fairly convincing fashion (figure 6.7). It can be shown that most of the self-vector peaks (those on u=1/2 and w=0) should be twice the peak height of the cross-vector peaks because of special positions and symmetry considerations and this provides a reason why the cross-vector sites fall into the noise level of the map. The unassigned peaks on section 0 and 3 might be due to the effect of missing terms in the 6Å data set (see chapter four) or signify other changes that occurred during the metal extraction procedures.

It was realized that a better Patterson calculated at higher resolution would define the position of the metal sites and thus the protein molecules

more accurately. However, larger crystals than used previously on the diffractometer were required for good quality higher resolution data. The crosslinking conditions described in chapter five were not satisfactory for these purposes because larger crystals dissolved when transferred from mother liquour to an aqueous environment. Thus attempts were made to crosslink large crystals by modifying the crosslinking conditions. Crystals were soaked either overnight at room temperature or for two days at 4°C using 0.2% glutaraldehyde in mother liquour. Either method resulted in large crosslinked crystals that could be transferred to water without dissolving. The crystals were treated in the usual way to extract the metal (see chapter four) and their diffraction pattern measured by precession camera. The crystals were found to be highly radiation-sensitive and weakly diffracting; so much so that they proved useless for data collection purposes. It was noted that the native crystals crosslinked under the abovementioned conditions were well behaved (radiation resistant and strongly diffracting) but a lot more dithionite was required to extract the metal from them compared with crystals crosslinked under milder conditions or not at all. It was quite possible that under the new crosslinking conditions the dithionite was damaging the protein and a milder reducing agent might prove more successful. Dithionite has been shown to damage Cu/ZnSOD but borohydride seems to work well in reducing the copper for metal extraction purposes without causing any corresponding damage to the protein (Viglino et al., 1985). An alternative method to produce suitable crystals could be to extract the metal from large crystals before carrying out the crosslinking treatment.

6.5 Double Derivative Experiments

It was reasoned that the origin of the poor isomorphism of the heavy atom derivatives might be small cell dimension changes with respect to the native crystals and associated movements of the molecules within the unit

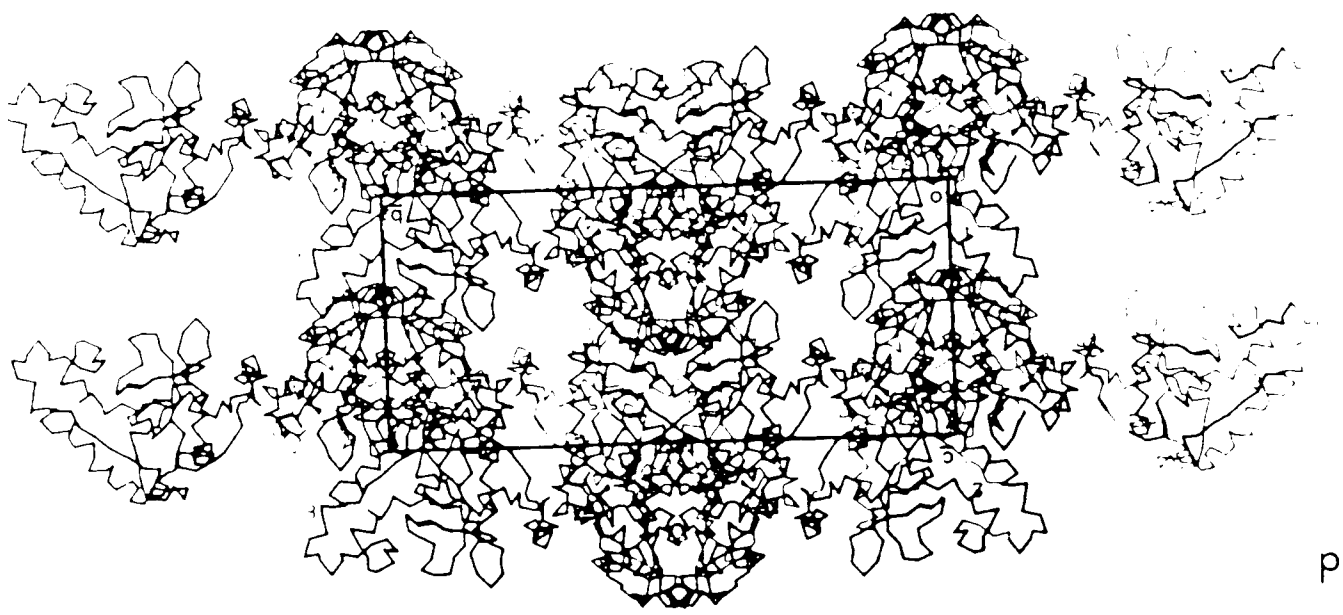
cell. It was hoped that the addition of some heavy atom reagent to the crystals might stabilise them and that such heavy atom-soaked crystals might be used as 'native' crystals for the purpose of finding new heavy atom derivatives. However, there appeared to be only one good candidate; the hard cation reagents which appeared to bind at the same sites in the unit cell. The problem with using these reagents was that they were the only reasonable derivatives available despite an extensive heavy atom derivative search (see chapter five).

A closer look at earlier results (table 5.2) suggested that organomercurial reagents might be suitable for the task. When these reagents were soaked into crosslinked crystals the cell dimensions changed so much that they were discounted as possible candidates for good derivatives. However, it was possible that the large organomercurial molecule might be locking the molecules into the one place and restricting their movement within the unit cell.

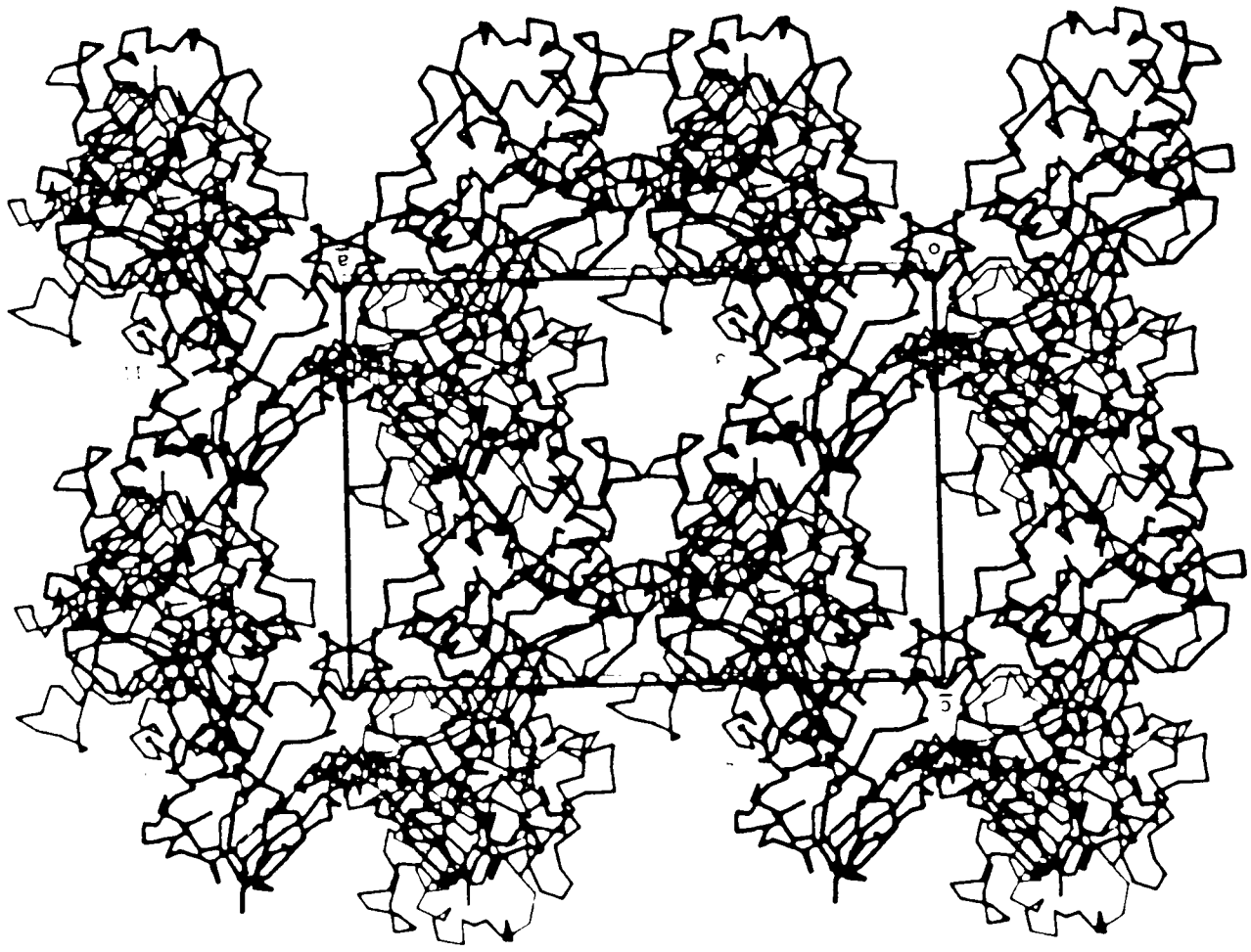
PCMBS appeared to be the best reagent to use on the basis of intensity changes in precession photographs. Crosslinked crystals were presoaked in 2mM PCMBS (in BES buffer pH 7.5) and their diffraction pattern measured. Intensity and cell dimension changes indicated that PCMBS had been bound to the crystals. The crystals were then soaked in 5mM uranyl acetate and the diffraction pattern of the crystals recorded. The crystals had suffered an a cell dimension change from 75.0 to 73.4Å with accompaning large changes in the diffraction patterns compared to the presoaked PCMBS crystals. The experiment was repeated because of the possibility that the PCMBS had been soaked out when the uranyl acetate was soaked in. Native crosslinked crystals were soaked in a solution containing 5mM uranyl acetate and 3mM PCMBS. The crystals were found to have an a cell dimension of 73.9Å and diffracted to 7Å only.

A final word needs to be said about the crosslinking conditions. Crystals that were crosslinked with glutaraldehyde overnight at room

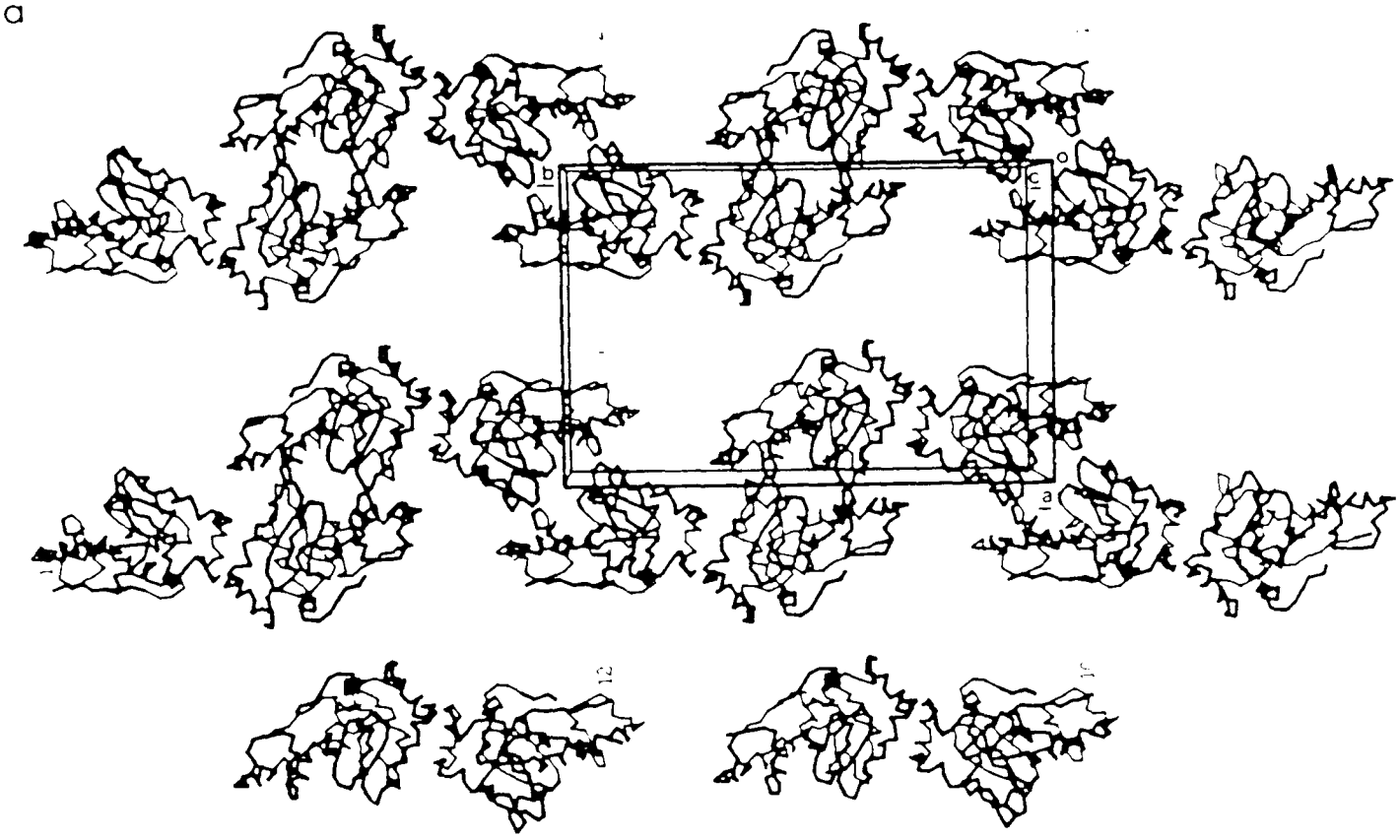
temperature had an a cell dimension of 75Å and diffracted to 4.5Å resolution only in the a cell direction even though they diffracted to greater than 2.5Å resolution along the b and c cell directions. Addition of uranyl acetate to the crystals caused the a cell dimension to change to between 73 and 74Å. Clearly, crystals soaked with glutaraldehyde under too harsh conditions would not be useful for higher resolution studies. Attempts to strike a medium between the old and new crosslinking conditions are in progress. It should be noted that crosslinking with glutaraldehyde does not inhibit cell dimension changes and does not significantly reduce the susceptibility of crystals to crack in planes normal to the a axis.



p



a



b

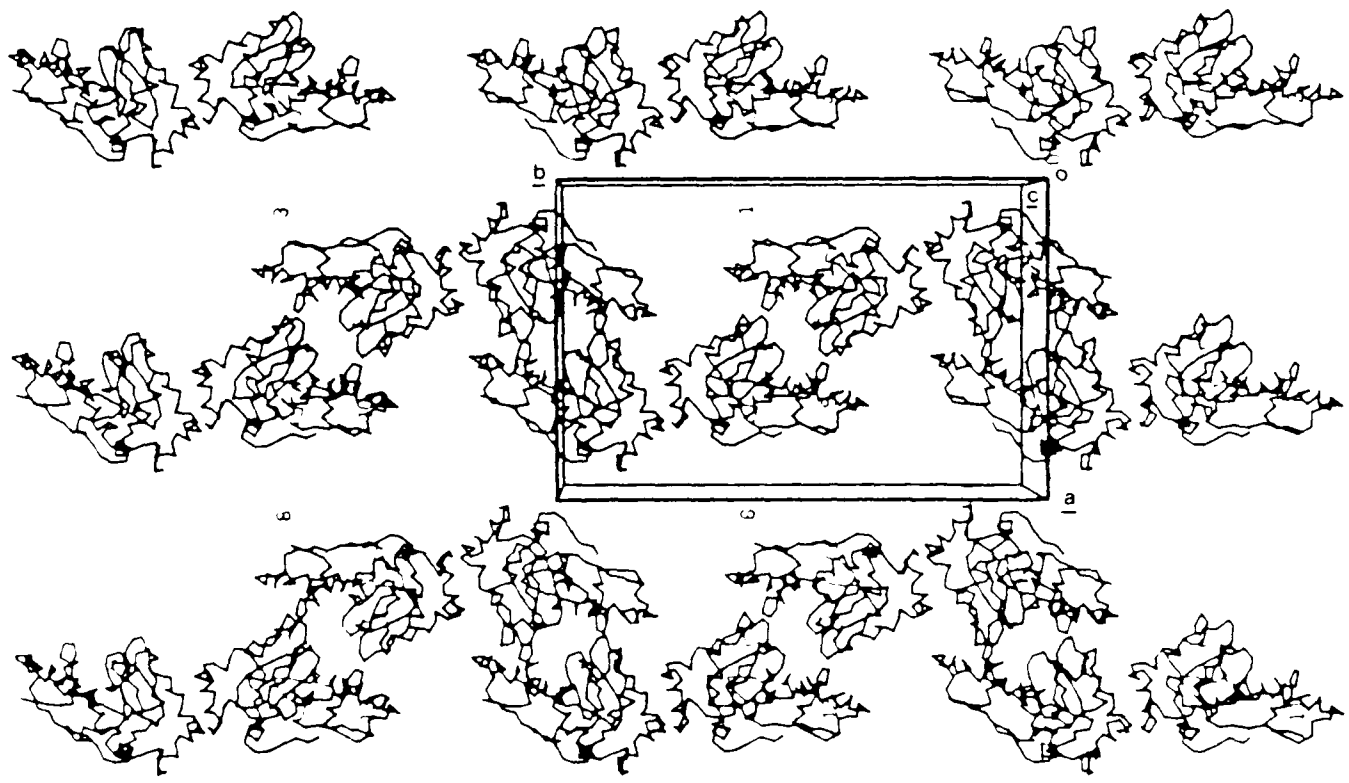


Figure 7.1 Packing diagram of MnSOD molecules as viewed down

(a) the crystallographic c axis, $0 < z < 0.5$,

(b) the crystallographic c axis, $0.5 < z < 1.0$,

(c) the crystallographic b axis, $-0.5 < y < 0.5$,

(d) the crystallographic a axis, $-0.5 < x < 0.5$.

The diagrams were generated by the computer program FITZ.

CHAPTER 7

Conclusions

7.1 The Crystal Structure of MnSOD

7.1.1 Introduction

The low resolution model of the crystal structure of MnSOD discussed in the last chapter can be used to explain various observations made of the crystals throughout this thesis.

7.1.2 Crystal Packing

The molecular origin of the crystal fragility problem discussed in chapter three can be explained by the model. It was observed that the crystals were prone to cracking in planes normal to the a axis and that there was a large degree of variability in the size of the a cell edge. The packing diagrams shown in figure 7.1 demonstrate that the dimer molecules pack in a layer fashion with long solvent channels (up to 20Å in width) running normal to the a axis. Changes in the a axis cell dimension could be readily accommodated because of loose packing between the molecules in the direction of the a axis.

It is interesting to note that the packing arrangements between the two known orthorhombic crystal forms of MnSOD ($P2_12_12$ and $P2_12_12_1$) can not be related by simple addition or removal of symmetry elements (Blake & Swan, 1971). The possibility of a space group other than $P2_12_12$ for the crystals studied in this work was discounted on the basis of high resolution precession photographs (μ=18°) and diffractometer measurements as well as analysis of Patterson maps.

7.1.3 Molecular Replacement

The absence of any consistent maxima in the self-rotation function maps

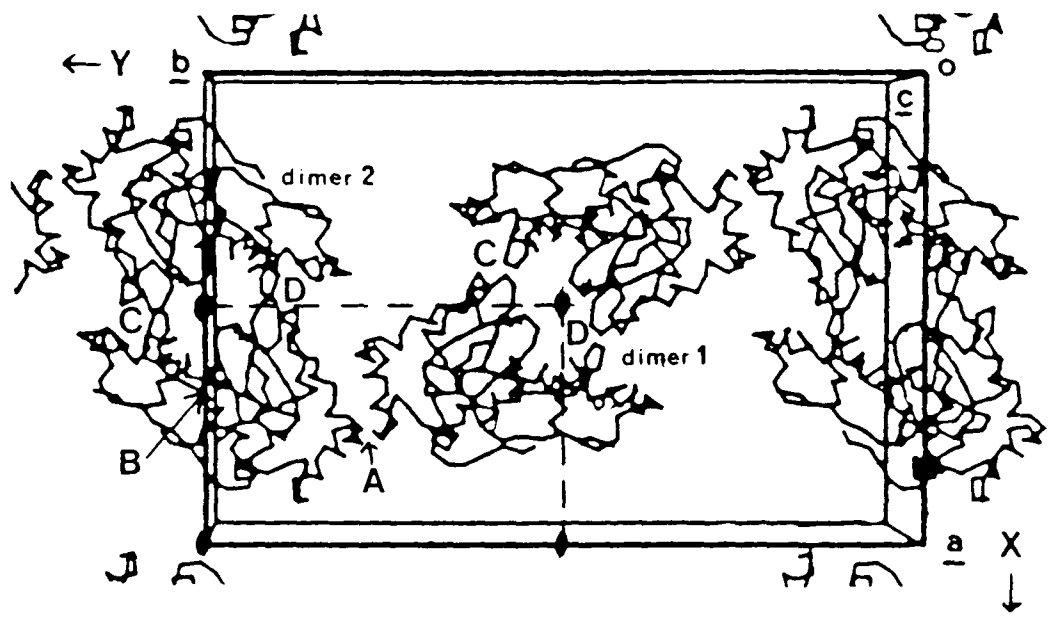


Figure 7.2 Exploded view of a portion of figure 7.1b showing the packing of dimer molecules in the unit cell. An asymmetric unit of the cell is indicated by the dashed line.

of MnSOD (see section 4.3.3) can be explained in terms of the model. The monomers of any one dimer are related by the crystal symmetry and so peaks representing the rotation of one into the other monomer are not expected. The rotational relationship of monomers from dimers not related by the crystal symmetry results in peaks that lie under the origin peak of the degenerate $\beta=\pi$ section. (This can be readily seen by inspection of figure 7.2 using the rotation function convention defined in figure 2.8). The peaks representing the rotational relationship between the FeSOD model and the MnSOD dimers in the cross-rotation maps were present but just above the noise level and consistently below the highest maxima. It has been noted by others that the highest maxima in rotation maps are not always the correct ones (Hol, 1985). The poor contrast in the cross-rotation maps was probably due to the poor quality of the model and the absence of side chains.

It should be noted at this stage that an alternative packing model (model II) had been considered during the model building efforts described in chapter six. In this model the position of dimer 1 was left unchanged but dimer 2 was rotated by 180° about an axis parallel to the x axis so that its molecular diad remained in alignment with the crystallographic diad (refer to figure 7.1a and 7.2). The alternative fit of dimer 2 to the electron density was more convincing (refer to figure 6.6); however there was some bad packing contacts between the dimers and the manganese positions of dimer 2 did not solve the manganese-free and anomalous Pattersons. The expected peaks for the self-rotation function, using this model, would fall under the origin peak of the degenerate $\beta=0$ section. It was interesting to note the presence of a fortuitous diad, roughly aligned with the y axis, that related monomers of different dimers. The presence of this 'pseudo-dimer' would explain the consistent maxima observed in the cross-rotation function maps (see section 4.3.4). If the alternative model is indeed correct then it is not clear why the stronger maxima in the cross-rotation maps were related to the 'pseudo-dimers' rather than the

real dimers. The methods outlined in section 7.2 will resolve which is the correct model. However, most of the evidence points towards model I being correct (for example, better packing and explainable heavy atom binding patterns) and it is this model that is referred to in the rest of the chapter.

The positions of the metal sites in the model fit the manganese anomalous (chapter four) and manganese-free (chapter four and six) Pattersons reasonably well. It is interesting to note that the self-vector peaks present in the v=1/2 Harker section of each map (figure 4.20) are in slightly different positions. The manganese anomalous data were measured from a native type 1 crystal ($\underline{a}$ = 72.4Å) whereas the manganese-free data were measured from a native type 2 crystal ($\underline{a}$ = 74.3Å). These results show that the molecules move no more than a couple of Ångstroms within the unit cell when the crystal changes from one type to another. Further analysis shows that most of the movement must occur along the molecular diads which lie in the z direction of the unit cell.

7.1.4 Heavy Atom Binding

The model fits in well with observations made of the gold derivative (see chapter five). It was suggested that the gold derivative had deposited itself into a solvent channel and the model clearly shows this to be the case (figure 6.2n).

It is not surprising that evidence of the non-crystallographic symmetry was not apparent from the isomorphous difference Pattersons and difference Fourier syntheses in view of the fact that two monomers rather than the dimer occupy the asymmetric unit of the unit cell. It was observed that the cation-type derivatives bind at two positions in the asymmetric unit of the unit cell (see section 5.7.7). One of the positions (corresponding to uranyl sites 3 and 4, samarium site 2 and lead site 2) is situated between dimers 1 and 2 (marked A in figure 7.2) and the other (corresponding to

uranyl sites 1 and 2, samarium site 1 and lead site 1) is situated on the surface of dimer 2 close to the monomer-monomer interface (marked B in figure 7.2). The poor isomorphism of the cation derivatives may be due to movements of the dimers in response to heavy atom binding at position A (figure 7.2). The position on dimer 1, equivalent to position B on dimer 2, appears to be inaccessible to heavy atoms because of tight crystal packing. It is interesting to note that the position B cation-binding site is close to the manganese site (about 6Å away) and this may have implications for the observation that most metal-substituted SODs are inactive (see chapter one).

There was no serious disturbance of the electron density at the heavy atom site positions; there was a small peak at position B and a small hole at position A suggesting the heavy atom occupancy estimations were reasonable but could be improved slightly.

7.1.5 The Electron Density Map and the Model

The quality of the final map was poor yet contains sufficient information to make a number of observations. The information attainable from a low resolution (6Å) map is rather little (Blundell & Johnson, 1976). An impression of the general shape and subunit arrangement of a protein may be made and helices might be apparent as rods of electron density. However, it is usually not possible to identify an unique path for the polypeptide chain; some parts of chain will be visible but there are usually many breaks, especially near the protein surface.

The density for the dimers is roughly ellipsoidal and the monomers have dimensions of about 50 X 40 X 30Å^3. The monomers are rather globular and there is no evidence for a distinct domain structure. It is interesting to note that the first eight N-terminal residues of both dimers do not fit into density suggesting that these residues are disordered in the crystal. The majority of atoms fit into density and a couple of helices fit into

rod-like density features. Two of the densest features in the map occur at the manganese positions of the model. There is definite evidence in the dimer 1 density that a number of residues need to be inserted between the first two helices of the model as was found for T.thermophilus MnSOD (see section 7.1.6; Stallings et al., 1984). The lack of density along the molecular diad of dimer 1 was an appealing feature of the map; however the appearance of density along the dimer 2 diad indicated there were residual phase errors.

The monomer-monomer contacts are the same for the two dimers; two contacts per monomer which are labelled C and D in figure 7.2. The electron density that bridges the two monomers in dimer 1 suggests some rebuilding is necessary (figure 6.4a). The contact residues of dimer 2 fit nicely into bridges of density (figure 6.6a). The rather few contacts between the monomers has been noted for other SODs (Ringe et al., 1983; Stallings et al., 1984).

There is only one contact position between the dimers along the y direction of the unit cell (marked A in figure 7.2). The dimer contacts in the z direction appear to be more extensive; it is interesting to note that the residues involve in stabilizing the tetramer of T.thermophilus MnSOD (Stallings et al., 1984) are probably also involved in dimer interactions here. The proposed dimer contacts are supported by the electron density map in which bridges of electron density are seen to join dimers at the appropriate positions given by the model.

7.1.6 Other Crystallographic Studies

A number of other groups were actively pursuing crystal structure determinations of MnSODs during the course of the work presented in this thesis. The first group to come to attention was Jansonius's group at the Biozentrum, University of Basel, Switzerland (Smit et al., 1977). This group had apparently given up on the project by the time the work at Oxford

had started (D. Rice, pers. commun.) because of the failure to find good isomorphous derivatives despite a search of over one hundred possible derivatives (R. Karlsson & J.N. Jansonius, pers. commun.). It was interesting to note that they also had cell dimension variability and crystal fragility problems. They renewed their efforts on obtaining coordinates and structure factors of an FeSOD (Ringe et al., 1983). They made little progress using the same molecular replacement procedures used in this work; the lack of progress being attributed to the poor quality of the FeSOD model (R. Karlsson, pers. commun.). Progress awaited the development of new reciprocal space methods which did not require the use of model coordinates (Karlsson, 1985). The phases obtained from the new procedures were good enough to confirm previosly determined heavy atom positions in five derivatives. It was possible to calculate an electron density map to 2.8Å resolution which revealed various promising features (Karlsson et al., 1984) but proved of too poor quality for a satisfactory interpretation (Karlsson, 1985).

The group who appear to have made most progress is Ludwig's group, University of Michigan, U.S.A. (Stallings et al., 1981). They have been working on a tetrameric MnSOD which apparently yields a stable crystal form. This suggests that the molecular packing arrangement in the unit cell is tighter for the tetramer than the dimer. Ludwig's group have determined the crystal structure to 4.4Å resolution by local averaging of an electron density map calculated by isomorphous replacement (Stallings et al., 1984). The heavy atom reagents used successfully in their work (K_2PtCl_4 and an organomercurial) did not yield promising derivatives in this work (see chapter five). Their results show that the spatial arrangement of the principal secondary structural features of FeSOD is approximately conserved in their structure and the metal ions of the two SODs are situated in equivalent locations within the protein molecules. However, a major difference occurs in the region between the first and second helices of the

FeSOD model and the experimental density where the MnSOD chain is lengthened by an insertion. The insertion is apparently responsible for stabilizing the tetramer by appropriate interchain contacts.

The only other group known to be working on MnSOD is the Richardson's group, Duke University, U.S.A. (Beem et al., 1976; Tainer et al., 1982). However, there has been no news of their progress to date.

The low resolution result obtained in this work fits in with the results of Ludwig's group and the predictions of other workers (see chapter one) that the Fe and MnSODs are structural homologues.

7.2 Future Studies

The crystallographic study presented in this thesis has described the methods by which a low resolution (6Å) electron density map of MnSOD has been obtained and lays a foundation upon which a higher resolution study may follow. Such a study may follow a number of courses.

(a) The clearest course to higher resolution involves fitting model coordinates to the low resolution map and extending the phases by molecular replacement. This technique has been successfully used in this laboratory for determining the structure of tortoise egg-white lysozyme to 1.6Å resolution from a fitting of hen egg-white lysozyme coordinates to a 6Å resolution map (Blake et al., 1983). However, a good model is required for the procedure to work (Hol, 1985) and it is expected that a better model will become available within the near future (G. Petsko, pers. commun.; Stallings et al., 1984; Xuong et al., 1985). The fit of the model to the map will be made by either improving the quality of the map by symmetry averaging (in progress) and fitting on the graphics or by reciprocal space refinement procedures (Sussman et al., 1977).

(b) It is planned to continue the heavy atom derivative search using, as a basis, the results of earlier work as outlined in table 5.2. It is clear that the three-dimensional uranyl fluoride data set should be

Protein	Np	Na	$\frac{\langle \Delta F_i \rangle}{\langle F_p \rangle}$ (%)	$\frac{\langle F_a \rangle}{\langle F_p \rangle}$ (%)
Crambin [1]	400	6 S	1.4	29
DPinsulin [2]		1 Ca	5	36
MnSOD	3100	1 Mn	1.1	7
		2 S 0	10	39

Table 7.1 Expected diffraction ratios for various anomalous scatterers in proteins. All values have been calculated for zero scattering angle and Cu Kα wavelength. See key to figure 4.14 for details.

References: 1) Hendrickson & Teeter, 1981

2) Stuart et al., 1984

recollected. In addition, there are a couple of other derivatives worth looking at in further detail ($PtCl_2(NH_3)_2$ and PCMBS). It would appear from results so far that hard cation-type reagents are unlikely to be of much further use so that efforts will be concentrated on searching for other types of derivatives. The possibility of co-crystallizing heavy atoms with protein will also be explored.

(c) An alternative course would be to utilise the latest advances in resolved anomalous phasing (RAP) techniques. It has been shown that anomalous scattering measurements alone are sufficient to solve protein structures in favourable cases. The phase information from anomalous scattering measured at a single wavelength is inherently ambiguous and must be combined with other phase information to be resolved. This information might come from assumming the phases for the total structure are close to those of the anomalously scattering atom structure (Peerdeman & Bijvoet, 1956; Ramachandran & Raman, 1956), from a knowledge of the partial structure of the anomalously scattering atoms derived from anomalous difference Pattersons (Smith & Hendrickson, 1981; Hendrickson & Teeter, 1981; Stuart et al., 1984), from consideration of the chemical composition of the protein (Karle, 1985), from solvent leveling (Wang, 1985), direct methods (Cassarano et al., 1982; Karle, 1984; Woolfson, 1984) or multiple wavelength data (Bartunik, 1978; Karle, 1980; Woolfson, 1984; Kahn et al., 1985).

The manganese in SOD appears to be too weak a scatterer to yield useful results based on calculation (table 7.1) and experiment (see chapter four). However, the real and imaginary scattering contributions of uranium in uranyl-soaked MnSOD crystals are large enough to make a RAP experiment, even at Cu Kα wavelength, feasible. Uranyl fluoride appears to have a better potential for such an experiment than uranyl acetate on the basis of intensity changes at higher resolution in precession photographs (figures 5.2 and 5.3).

Preparations are under way on two fronts. Uranyl-soaked crystals are being prepared for single wavelength diffraction measurement by diffractometry. It is planned to measure anomalous dispersion data to 2.4Å resolution and use at least one of the above mentioned methods to solve the structure of the protein.

Preliminary measurements of uranyl-soaked crystals are being made on the FAST detector system at the SRS facility, Daresbury, in collaboration with Dr. John Helliwell. The plans are to collect multiwavelength data about the L_3 absorption edge of uranium (λ=0.72Å). Although the imaginary scattering contribution at this wavelength is actually less than at Cu Kα wavelength, there are a number of decided advantages to using wavelengths about this edge. The variation of the real component of the anomalous scattering about the edge can be utilised, there is almost no crystal absorption at these wavelengths and expected to be little radiation damage. The X-ray flux from the Wiggler station is optimised at these wavelengths.

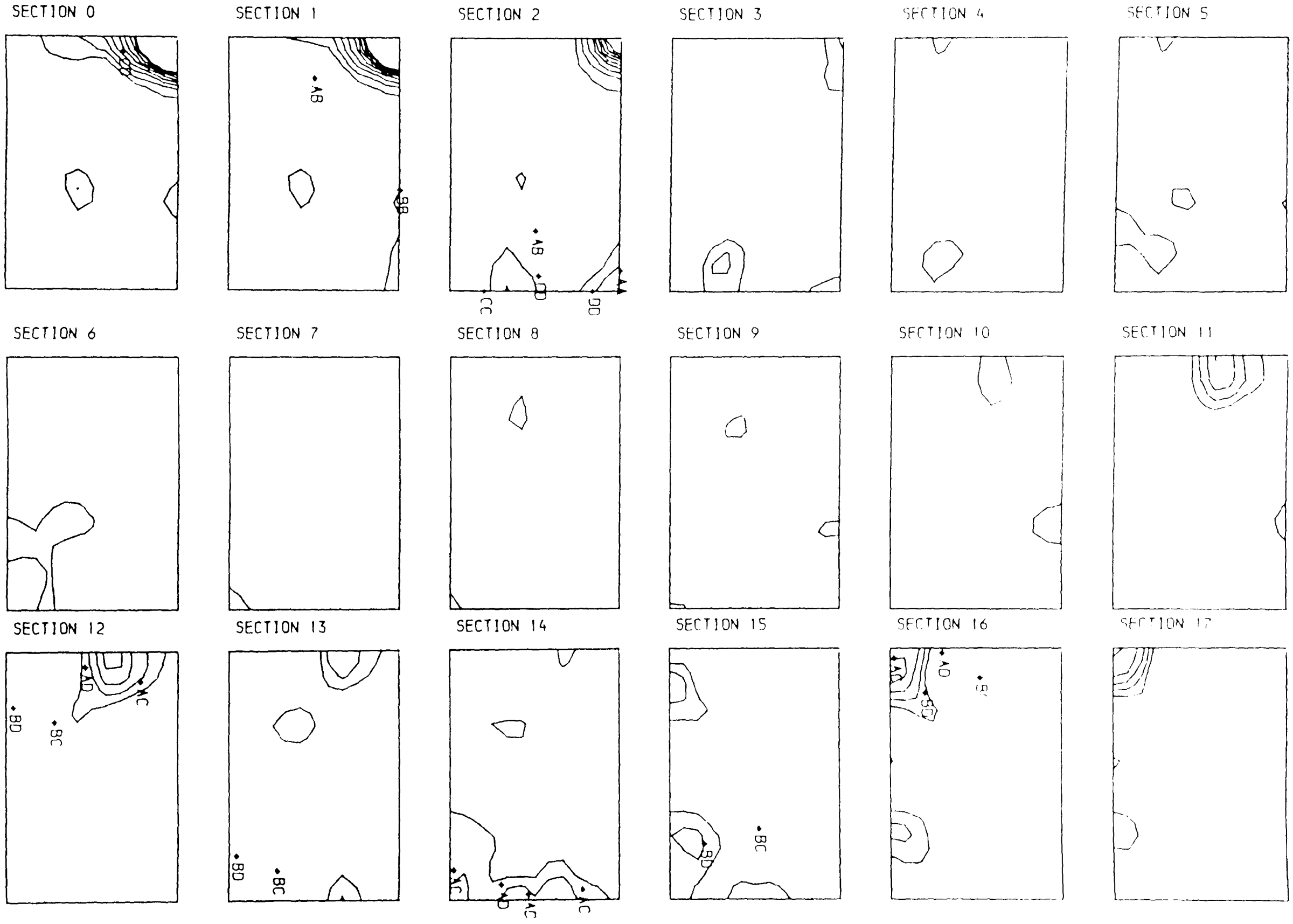
The eventual aim of a high resolution crystallographic study would be the definition of mechanistic and structural similarities and differences between the SODs which may also clarify the evolutionary relationships among the different classes of dismutases. This work would represent the first high resolution X-ray analysis of an enzyme containing a redox-active manganese ion; a group of enzymes of increasing biological interest.

Appendix 1

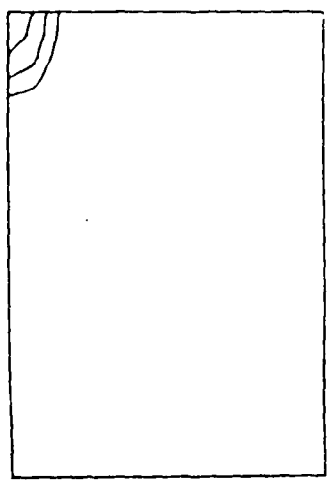
Patterson Maps

Sections of the isomorphous (iso) and anomalous (ano) difference Pattersons with the positions of self- and cross-vectors derived from the final, refined, heavy atom sites. All maps were calculated to 6Å resolution and sectioned on the v axis (section interval of b/56). The long and short edge of each section correspond to the u and w axes, respectively. Each map was calculated from 0 to 1/2 on each axis and contoured at the same equal but arbitrary intervals with the origin at the top left hand corner of section 0.

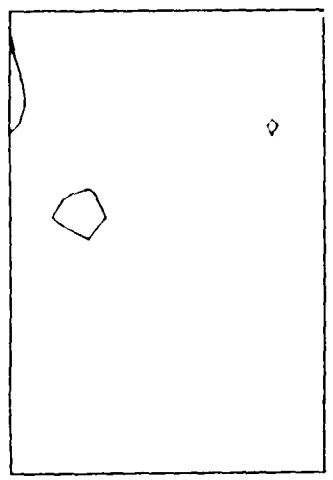
(a) uranyl acetate (iso)



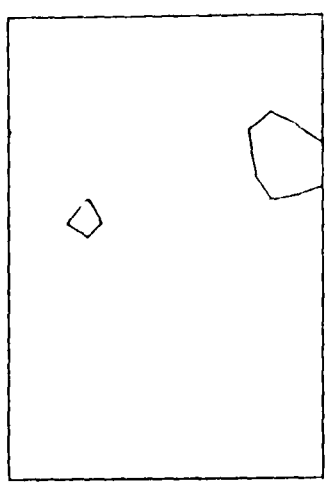
SECTION 18



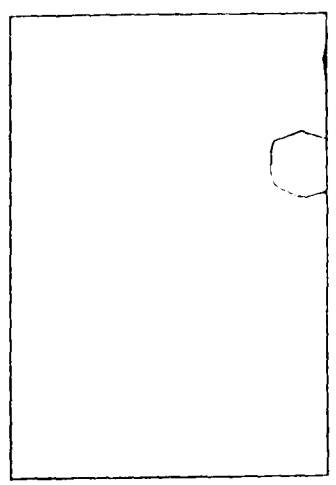
SECTION 19



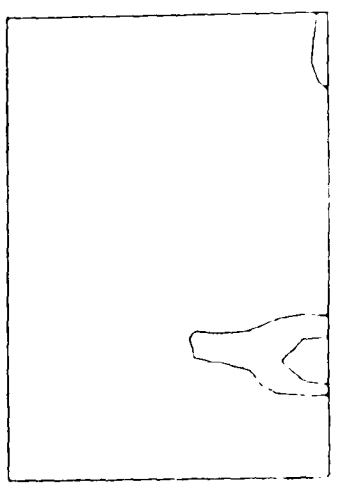
SECTION 20



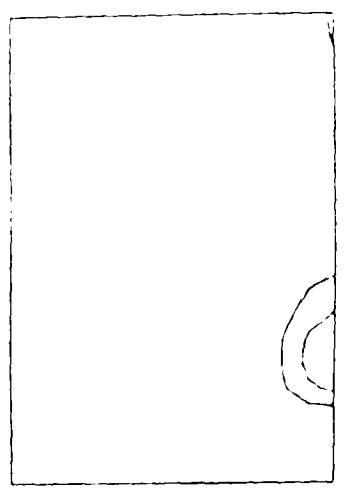
SECTION 21



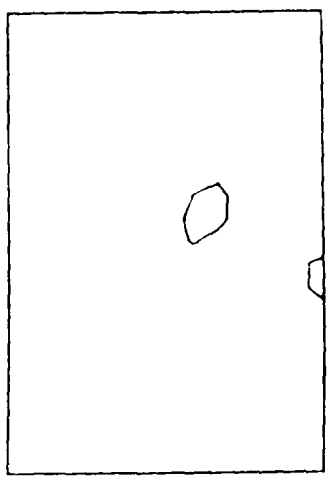
SECTION 22



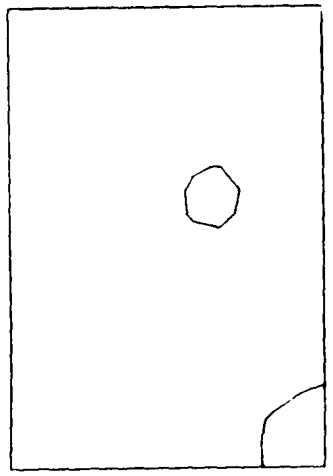
SECTION 23



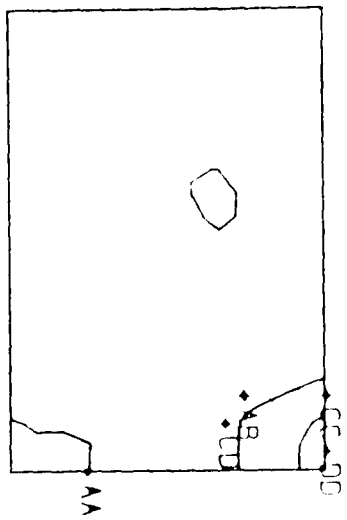
SECTION 24



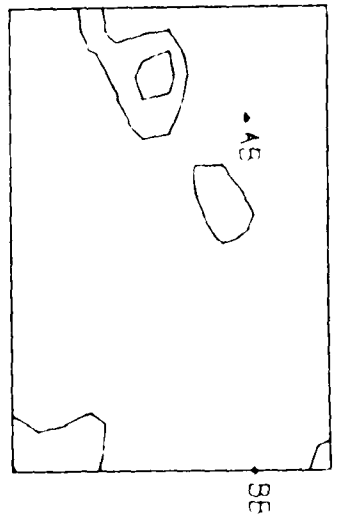
SECTION 25



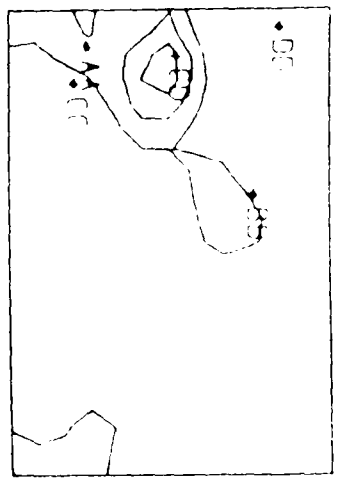
SECTION 26

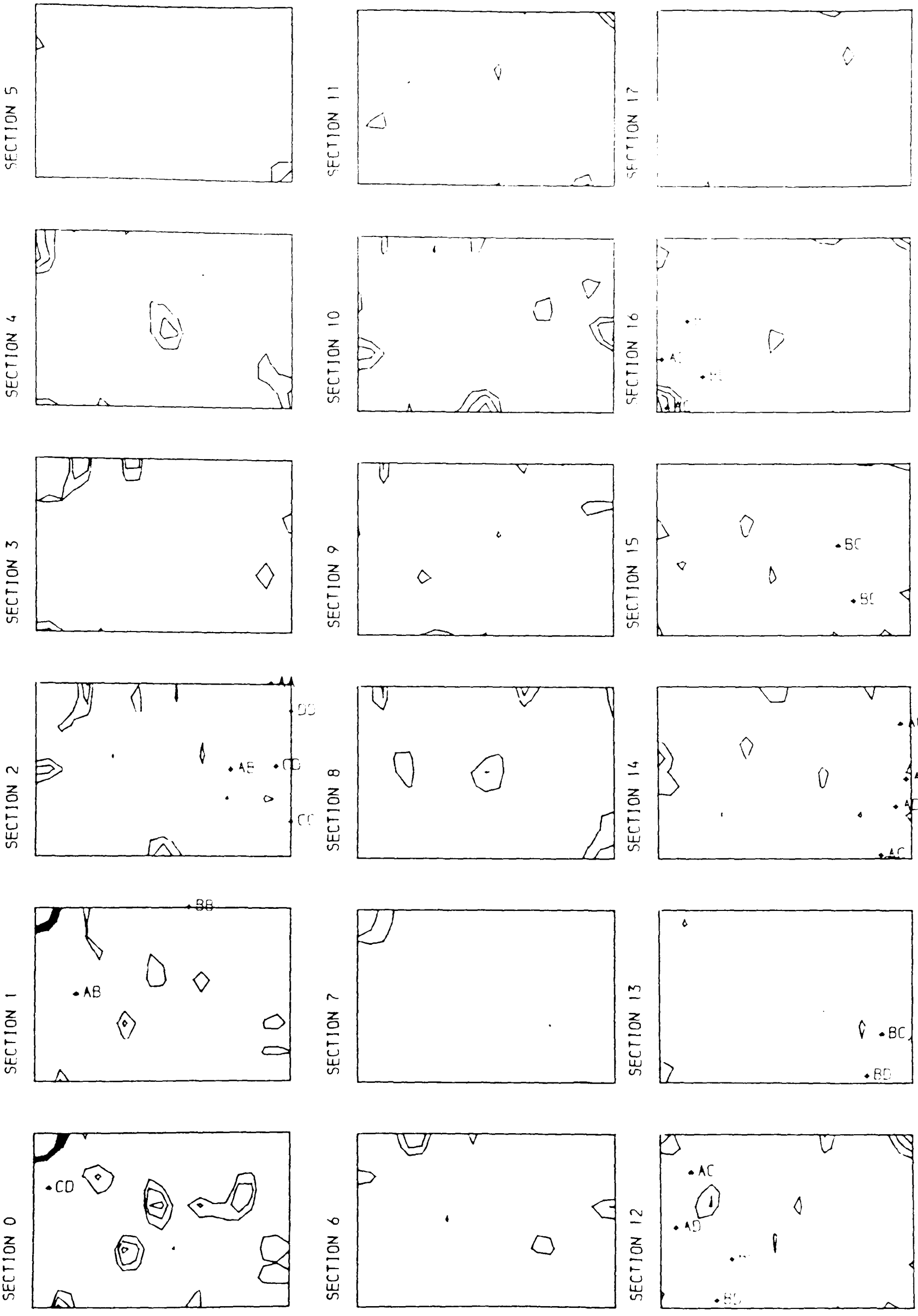


SECTION 27



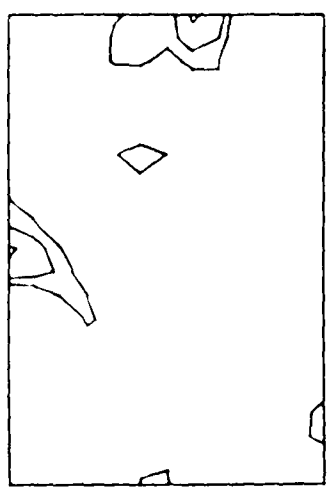
SECTION 28



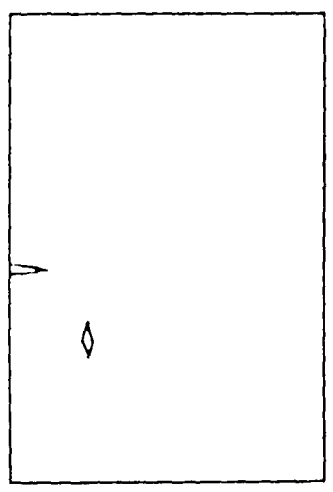


(b) uranyl acetate (ano)

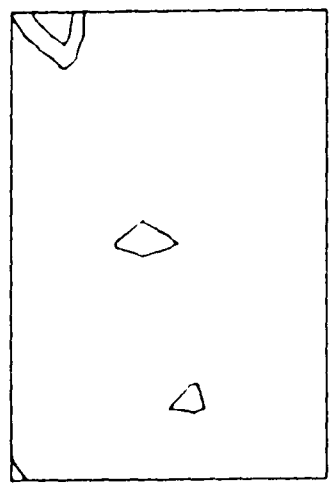
SECTION 18



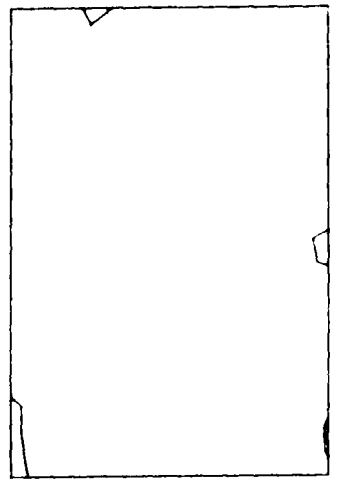
SECTION 19



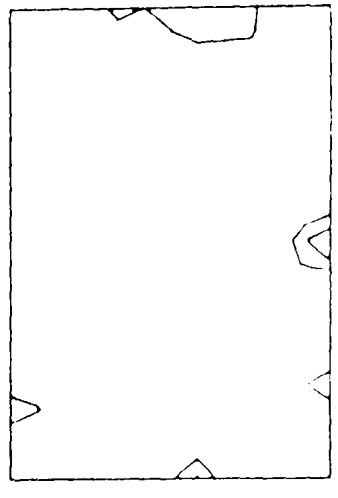
SECTION 20



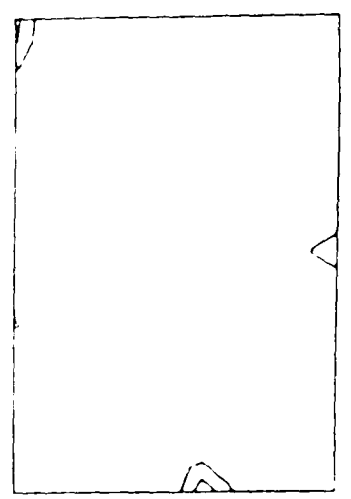
SECTION 21



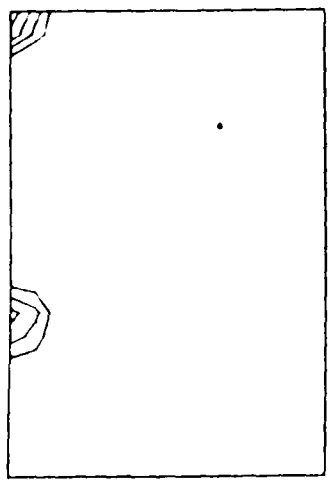
SECTION 22



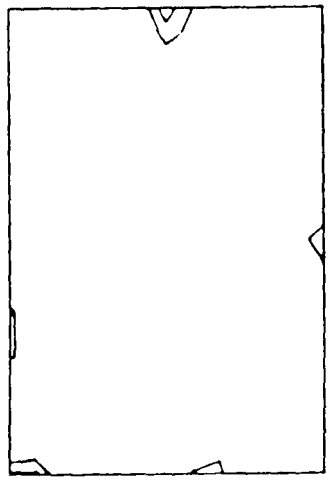
SECTION 23



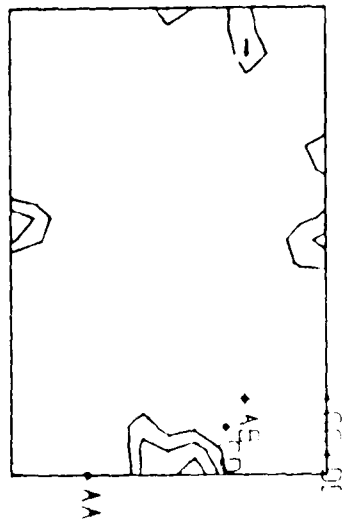
SECTION 24



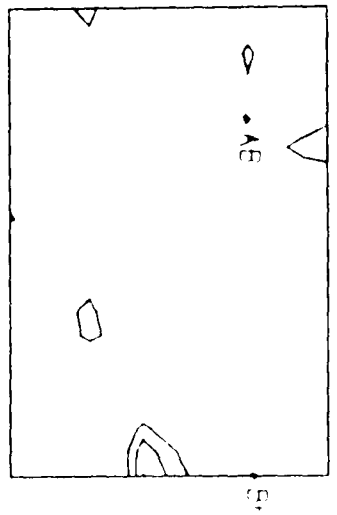
SECTION 25



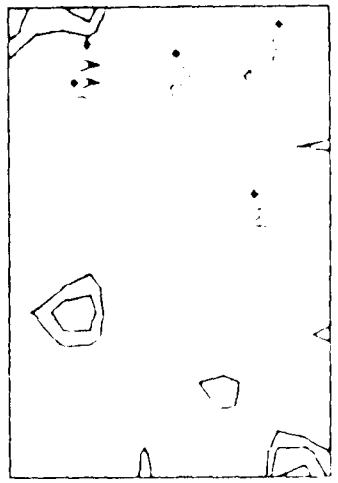
SECTION 26

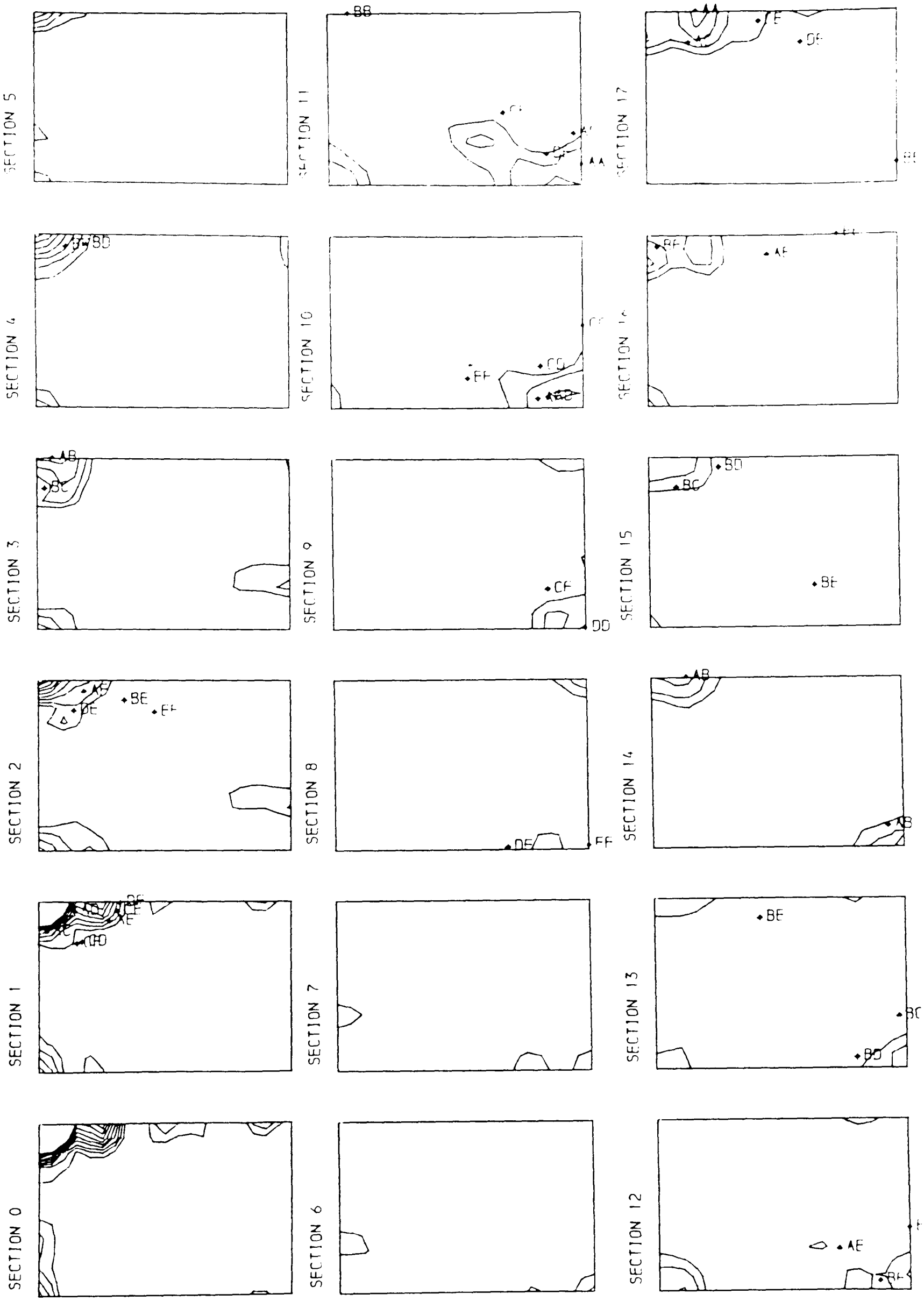


SECTION 27



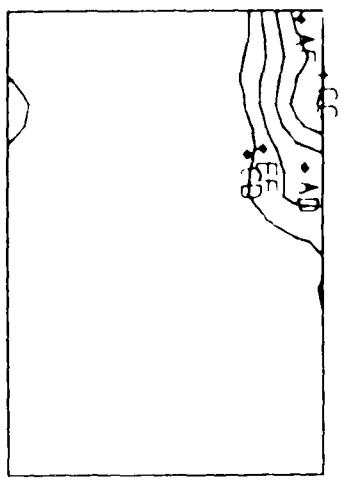
SECTION 28



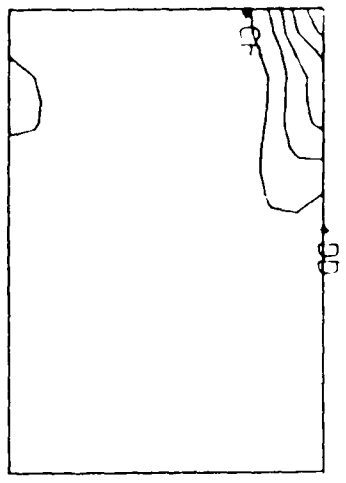


(c) gold chloride (iso)

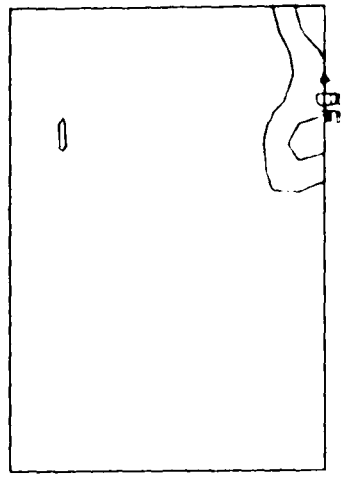
SECTION 18



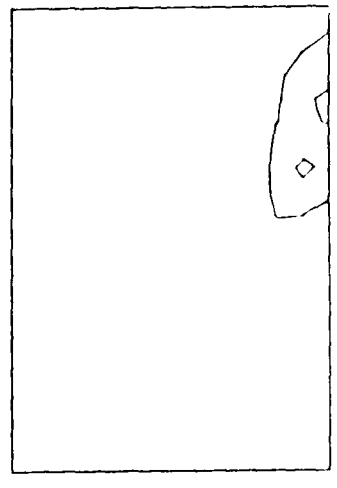
SECTION 19



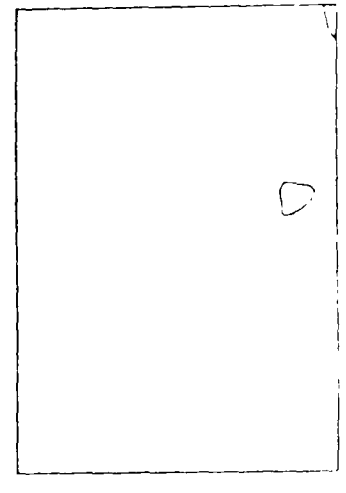
SECTION 20



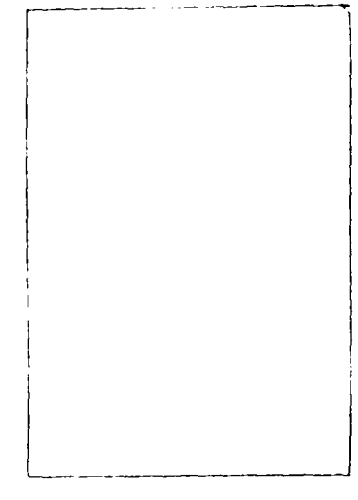
SECTION 21



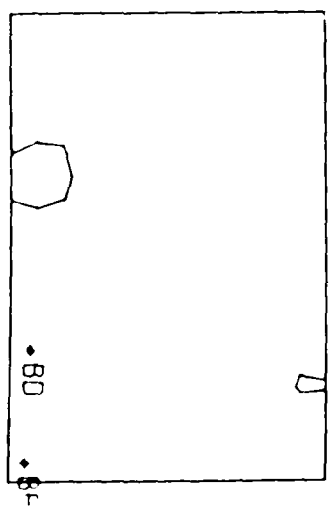
SECTION 22



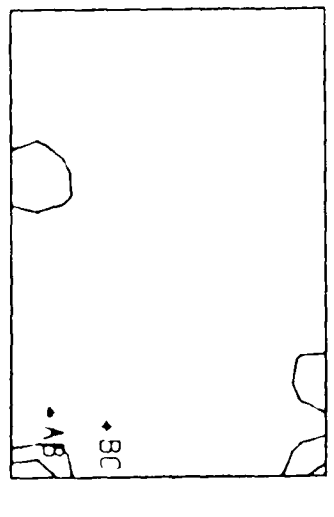
SECTION 23



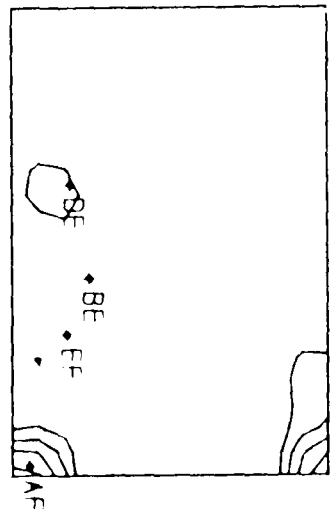
SECTION 24



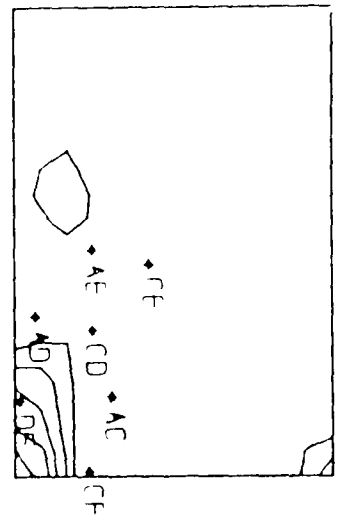
SECTION 25



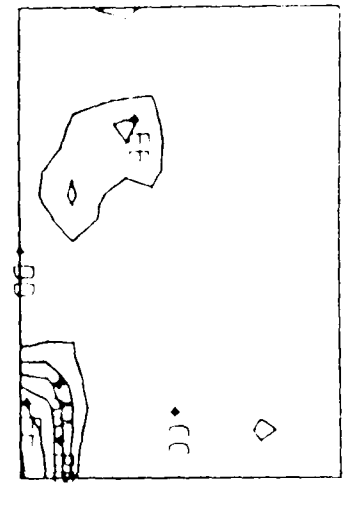
SECTION 26

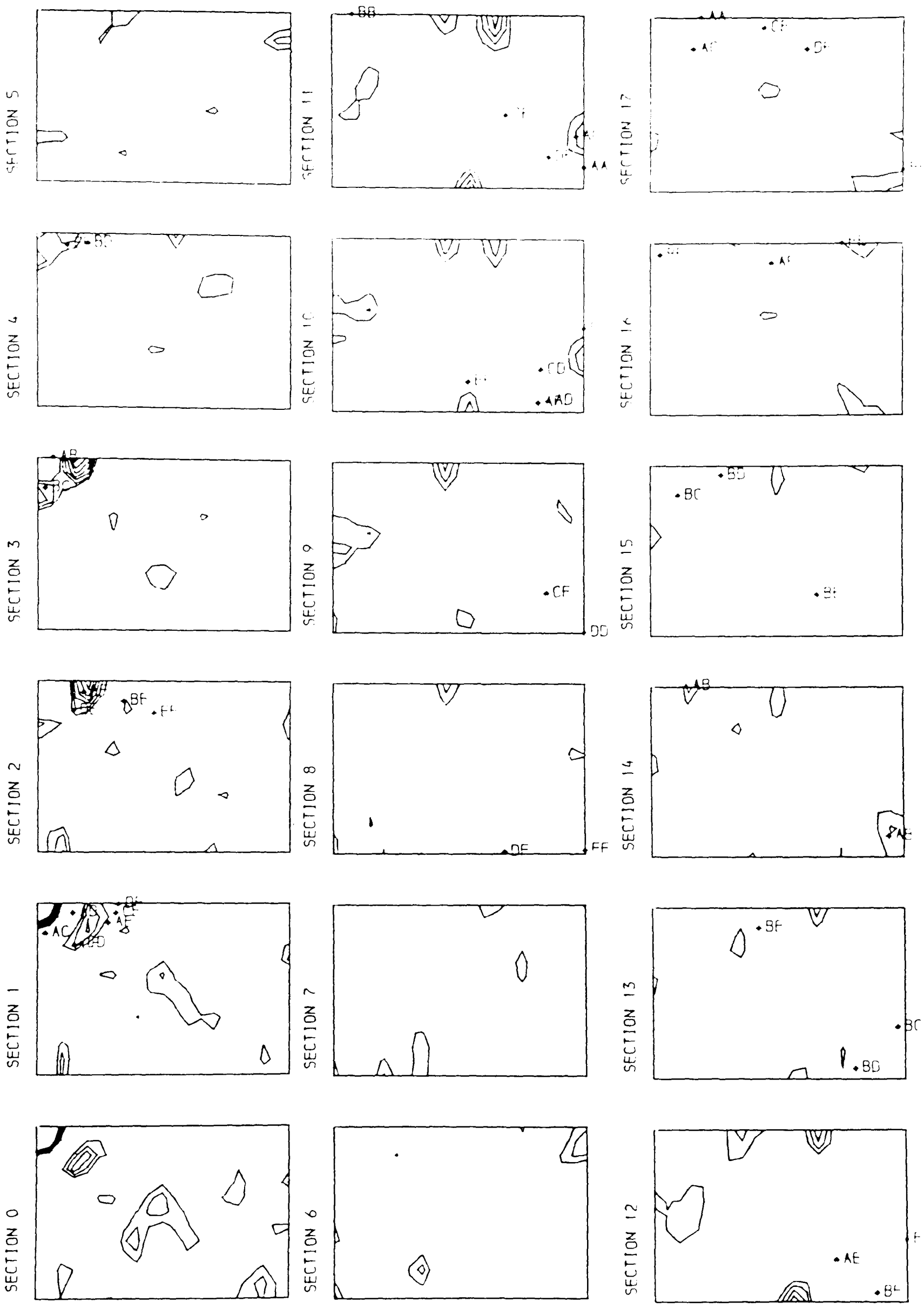


SECTION 27



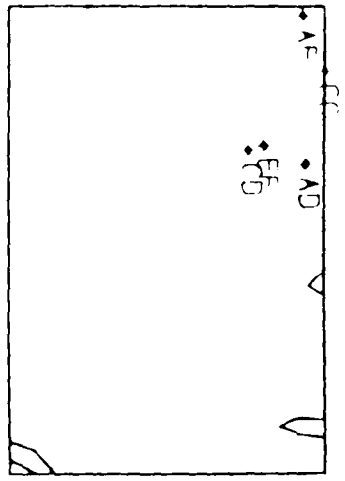
SECTION 28



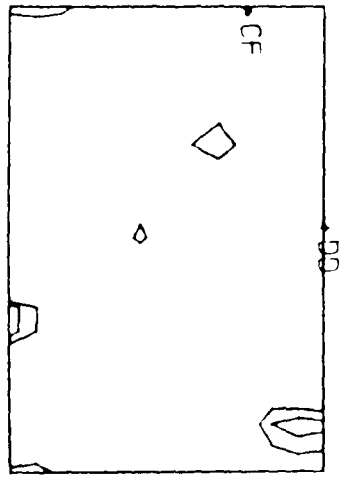


(d) gold chloride (ano)

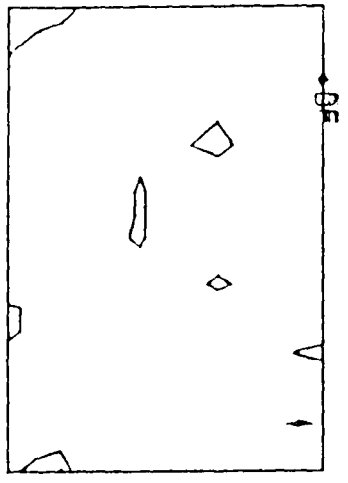
SECTION 18



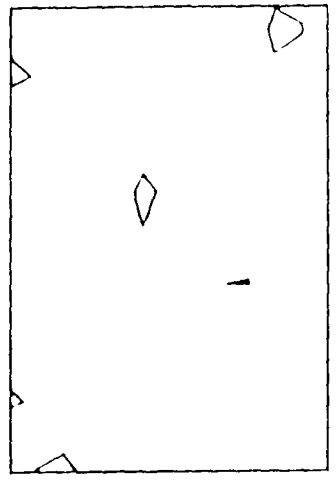
SECTION 19



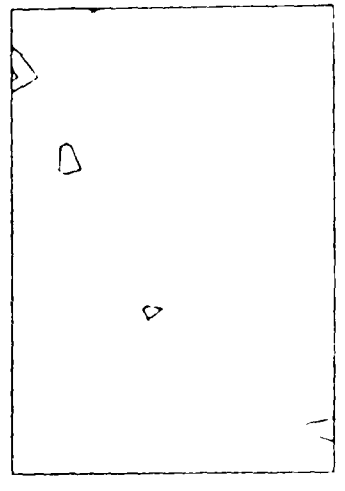
SECTION 20



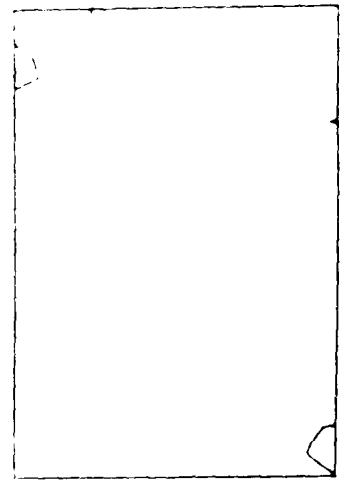
SECTION 21



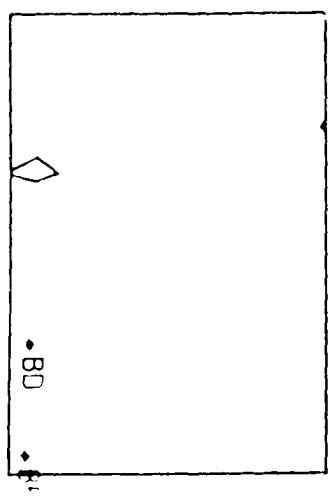
SECTION 22



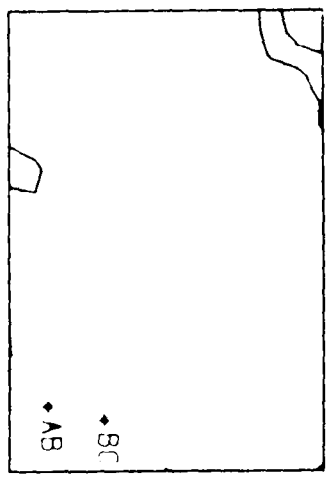
SECTION 23



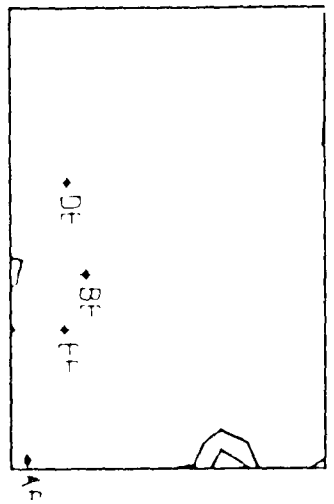
SECTION 24



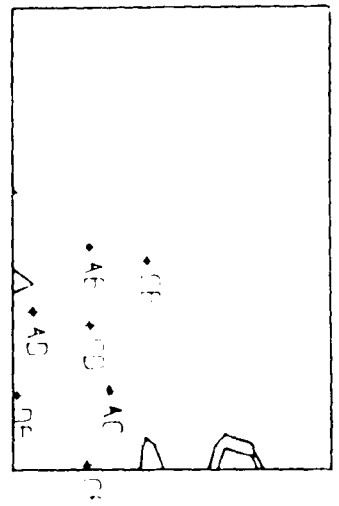
SECTION 25



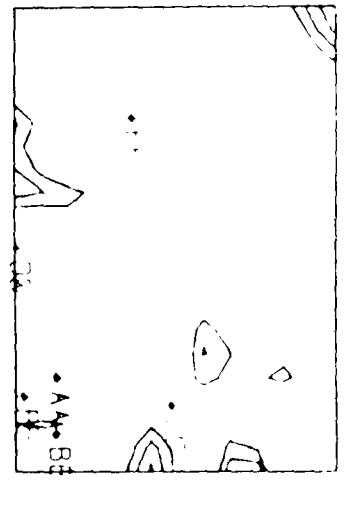
SECTION 26



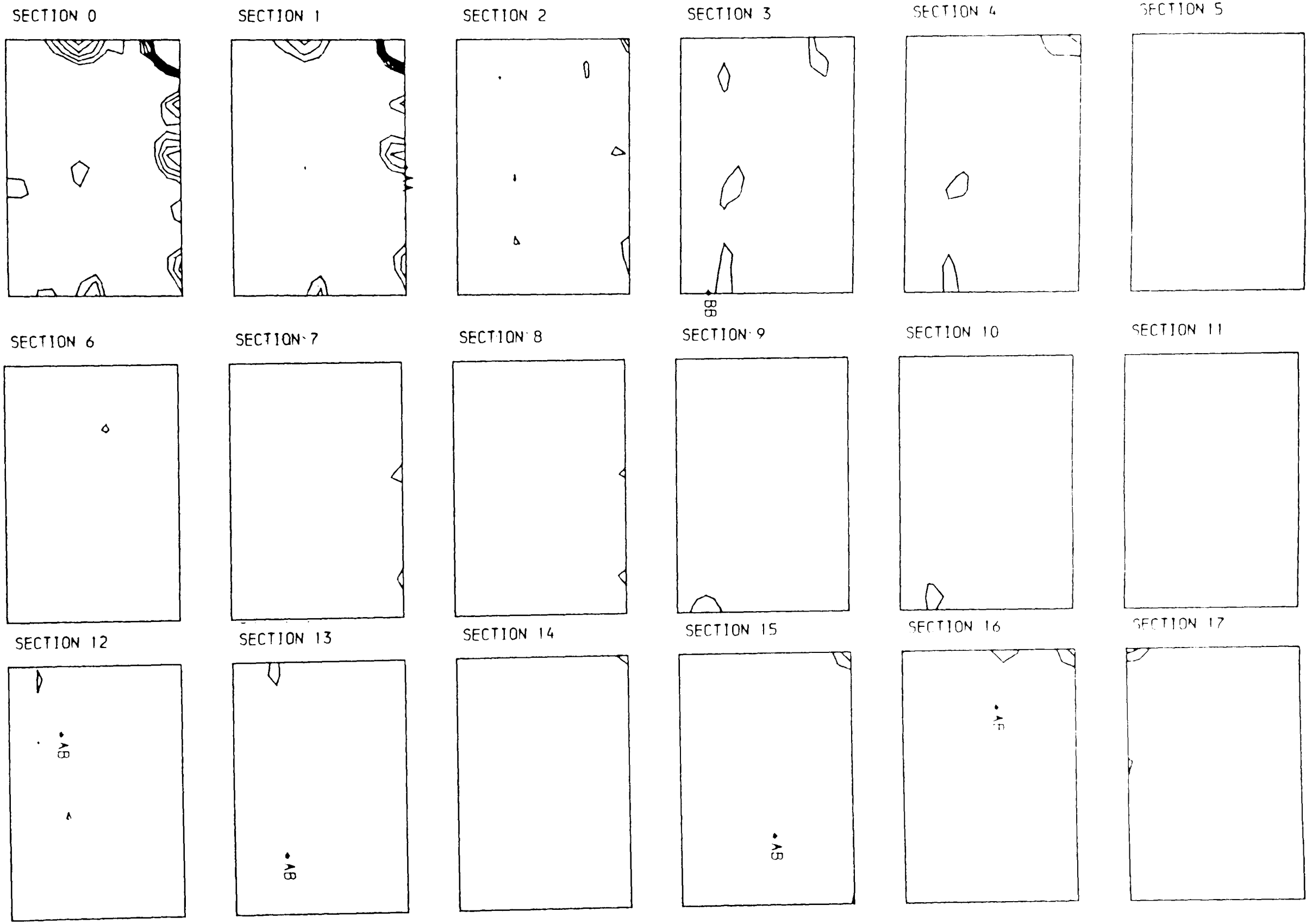
SECTION 27



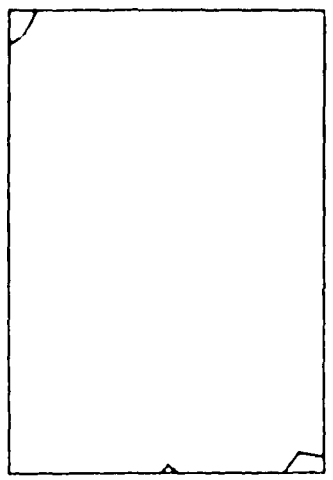
SECTION 28



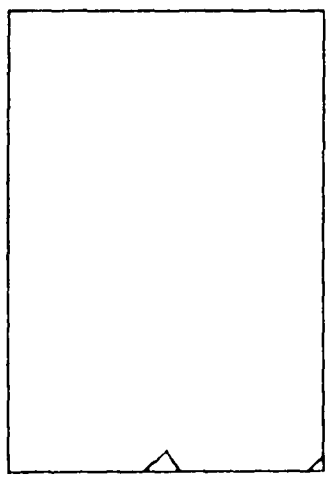
(e) samarium acetate (1so)



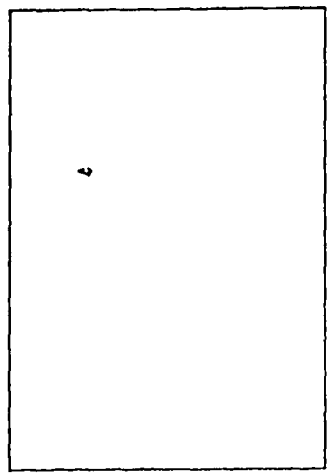
SECTION 18



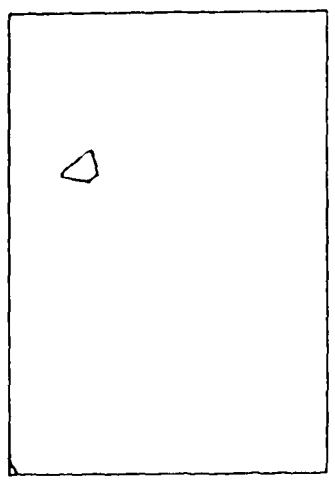
SECTION 19



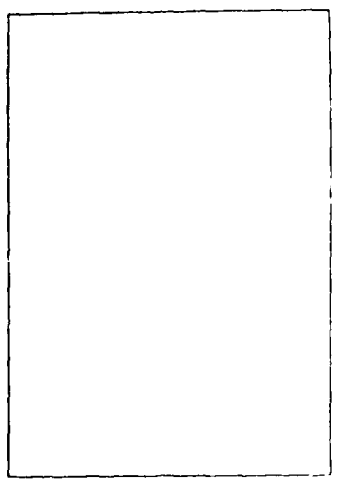
SECTION 20



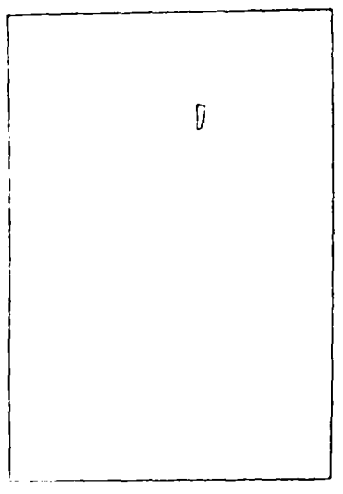
SECTION 21



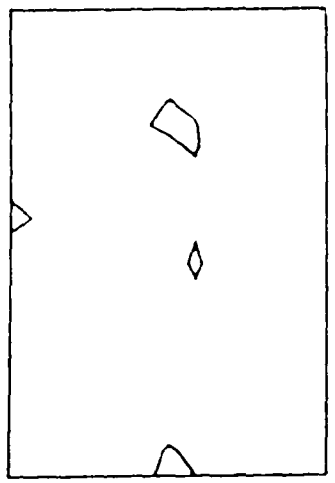
SECTION 22



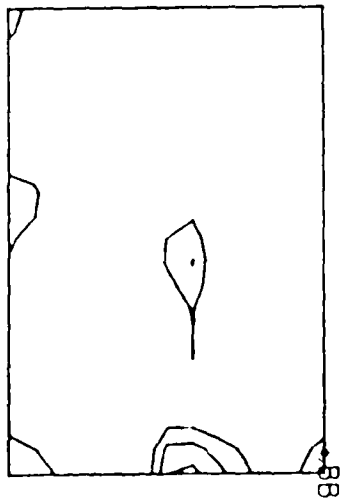
SECTION 23



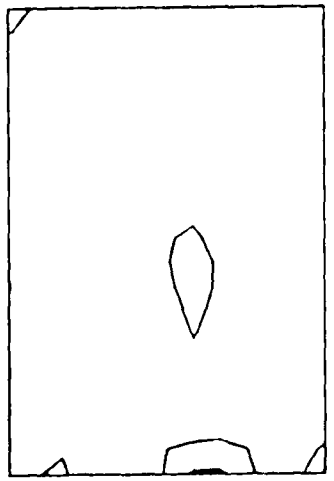
SECTION 24



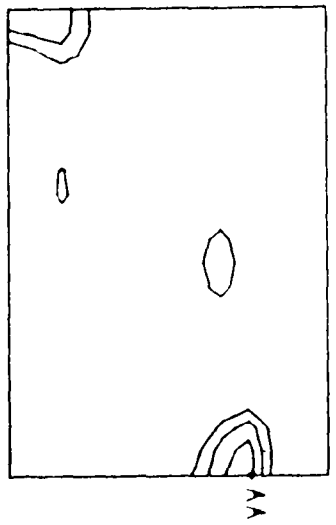
SECTION 25



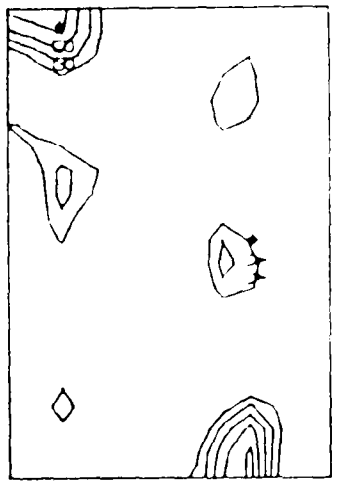
SECTION 26



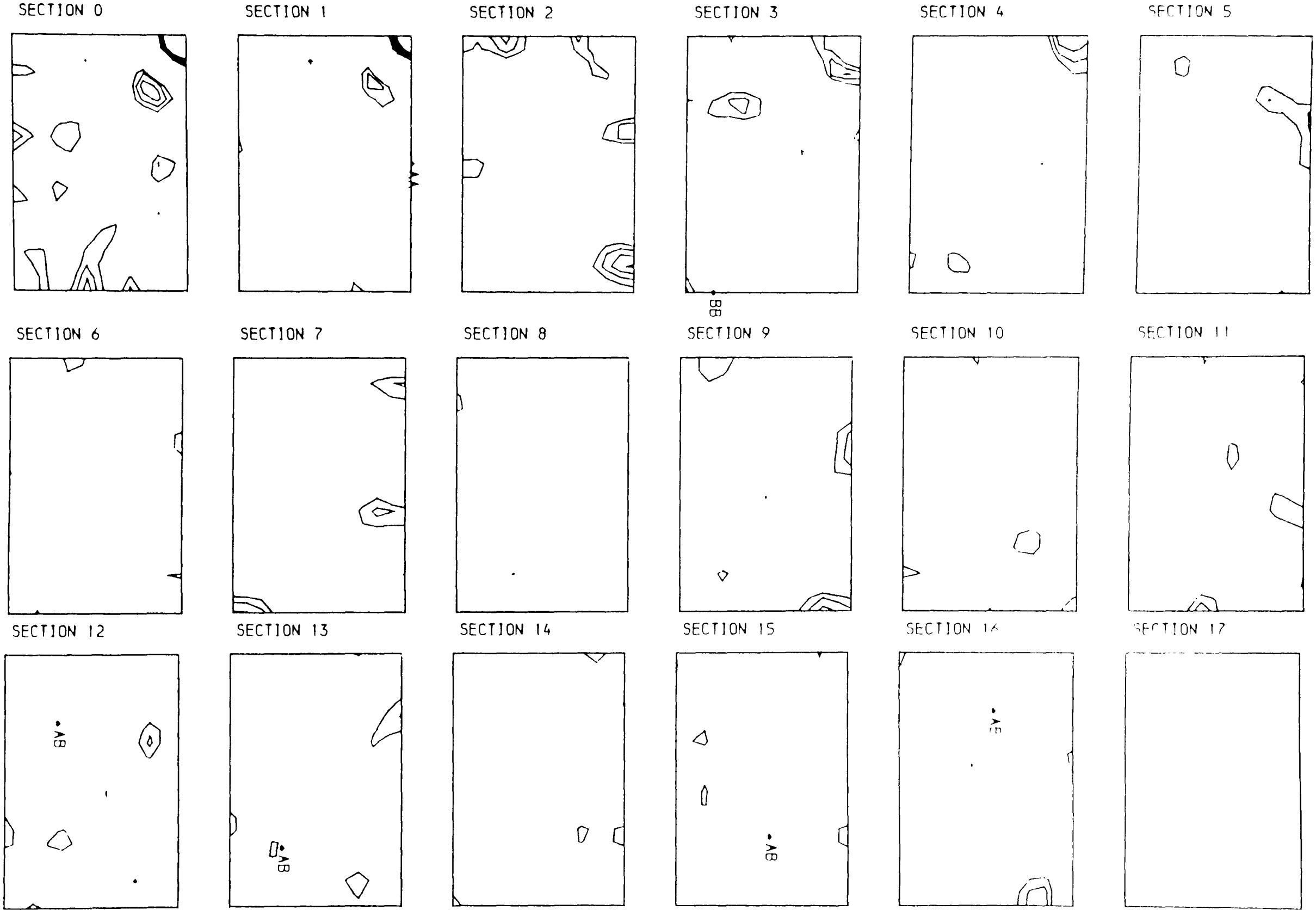
SECTION 27



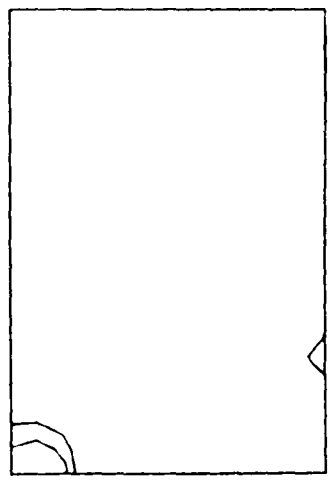
SECTION 28



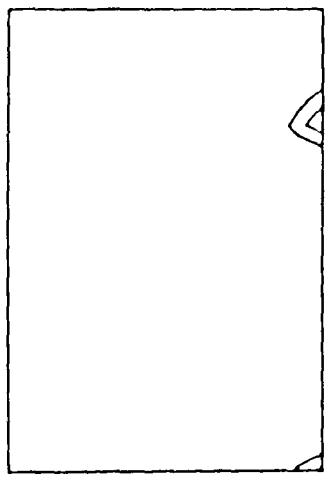
(f) samarium acetate (ano)



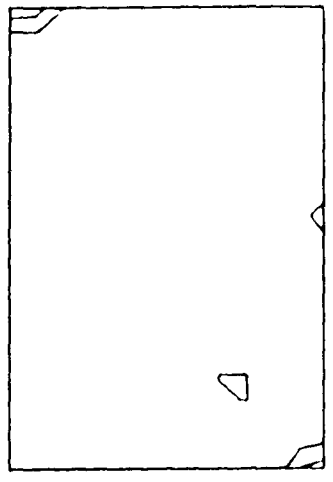
SECTION 18



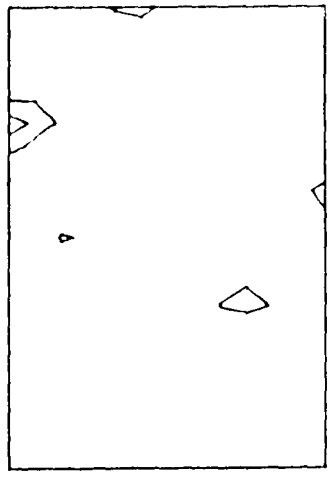
SECTION 19



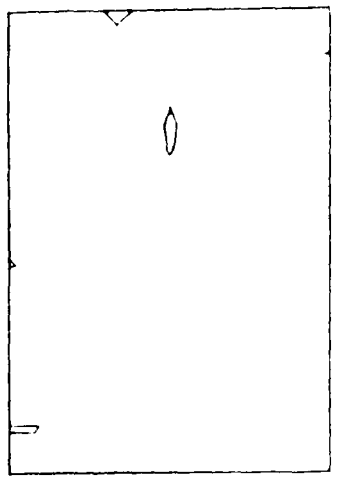
SECTION 20



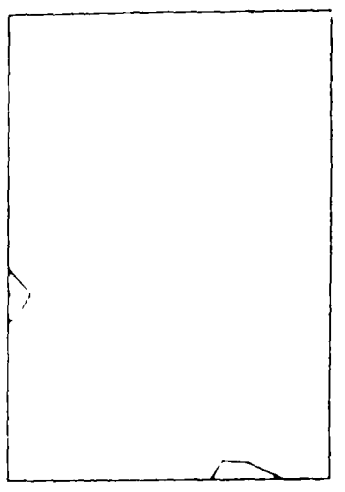
SECTION 21



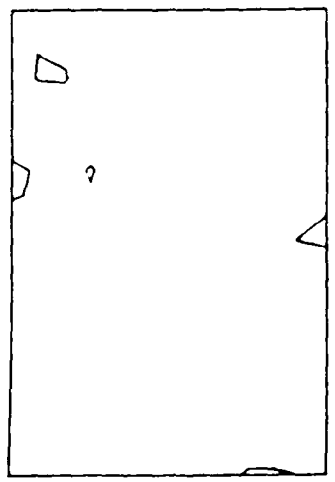
SECTION 22



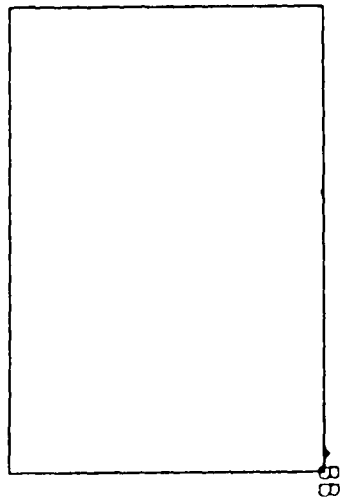
SECTION 23



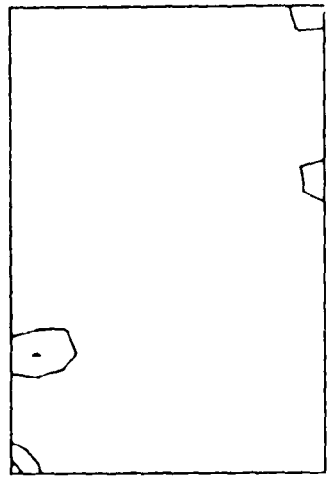
SECTION 24



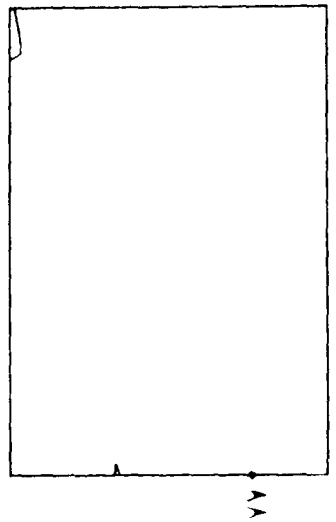
SECTION 25



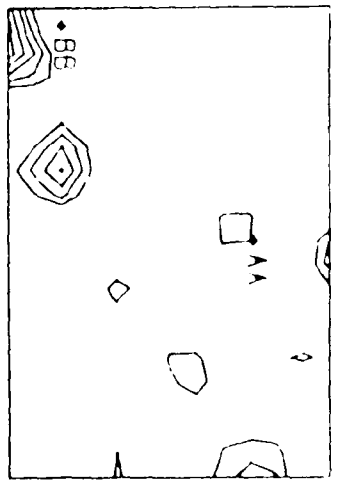
SECTION 26



SECTION 27

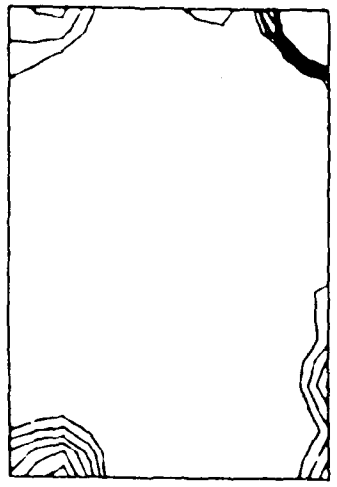


SECTION 28

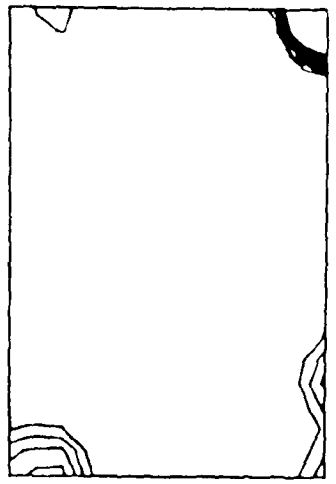


(g) EMP (ISO)

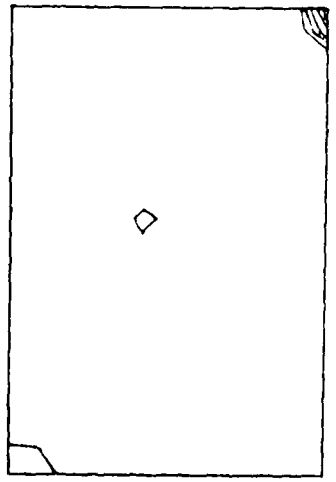
SECTION 0



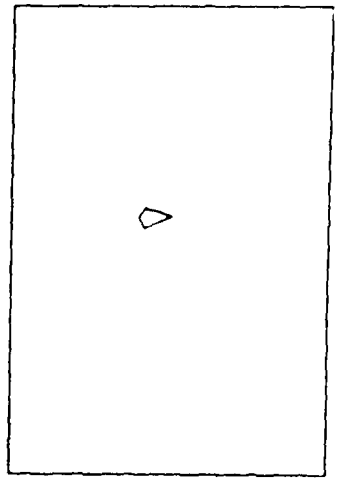
SECTION 1



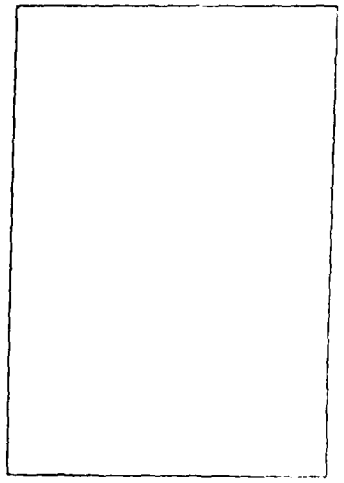
SECTION 2



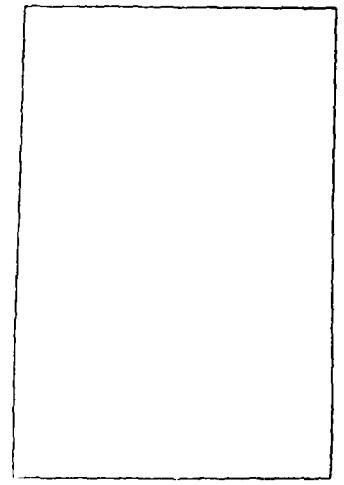
SECTION 3



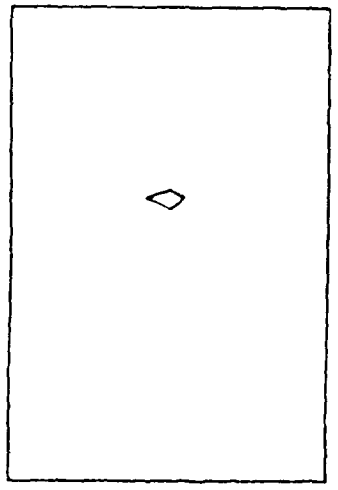
SECTION 4



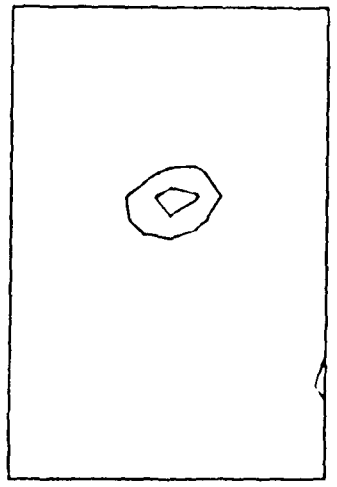
SECTION 5



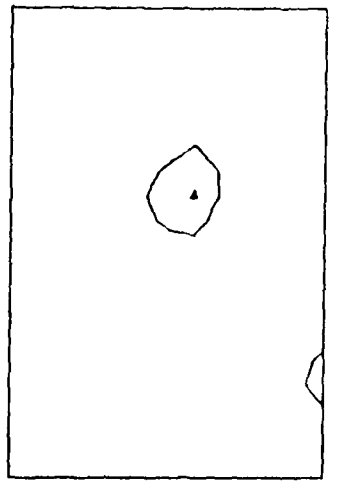
SECTION 6



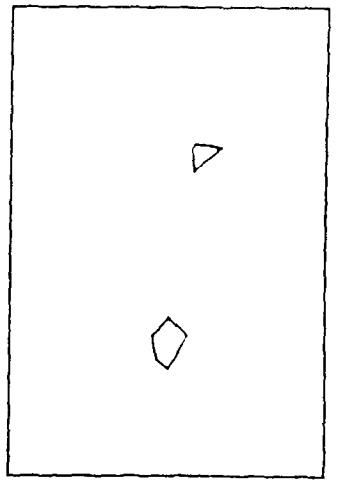
SECTION 7



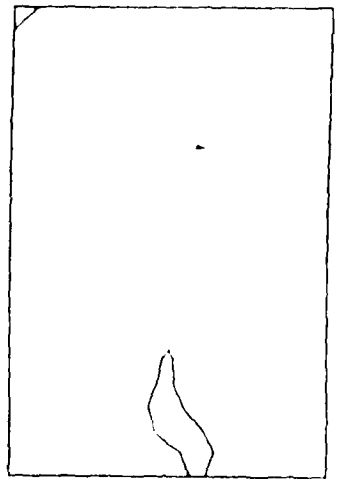
SECTION 8



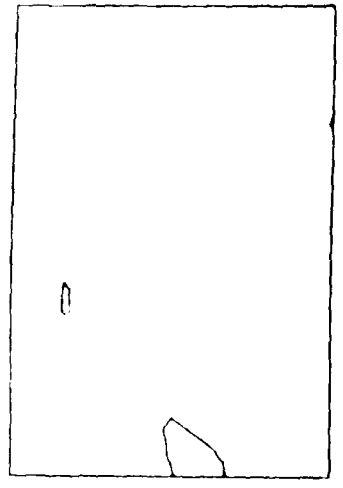
SECTION 9



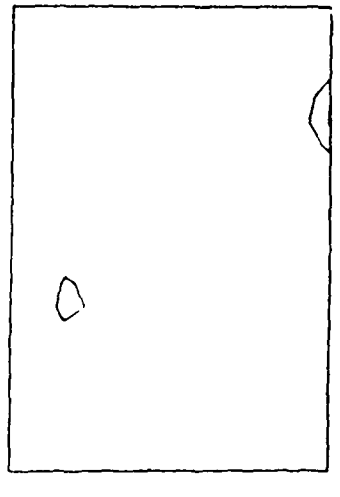
SECTION 10



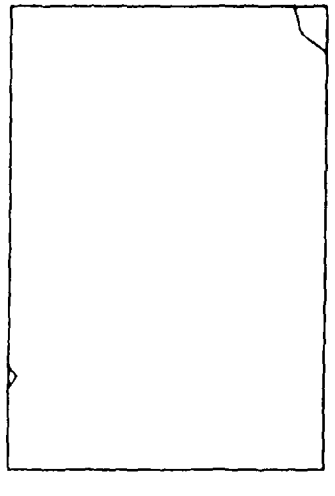
SECTION 11



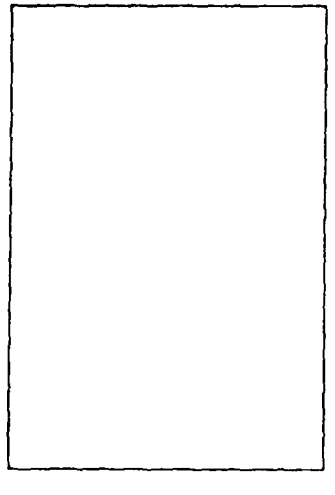
SECTION 12



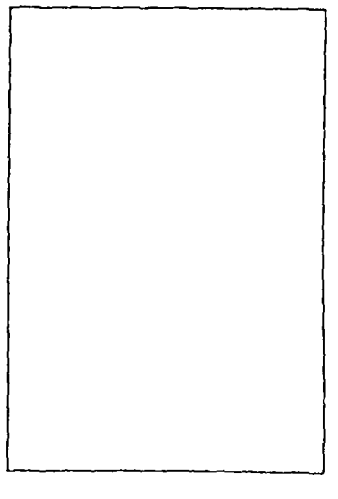
SECTION 13



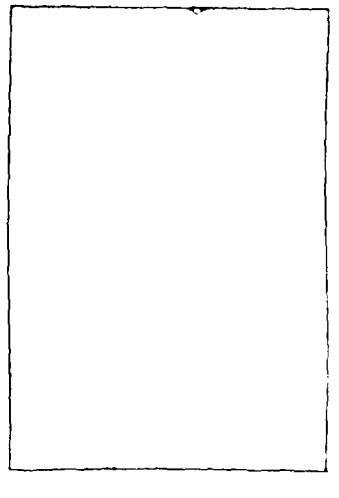
SECTION 14



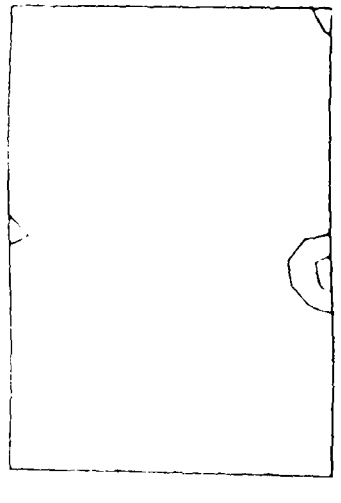
SECTION 15



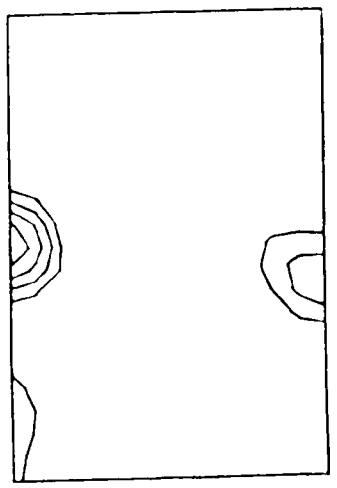
SECTION 16



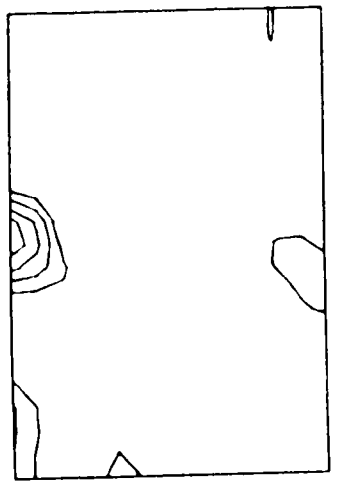
SECTION 17



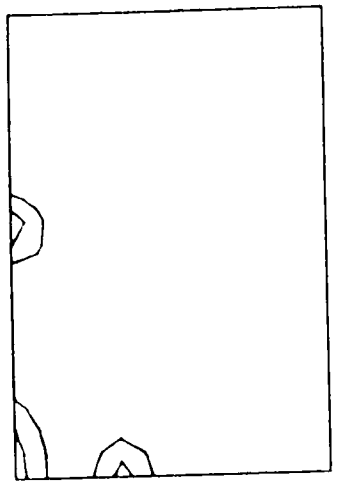
SECTION 18



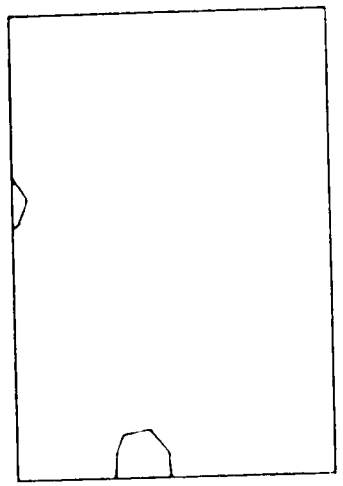
SECTION 19



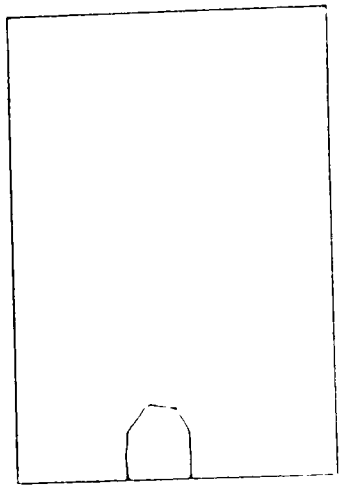
SECTION 20



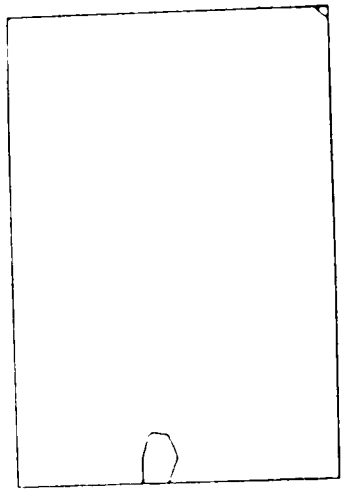
SECTION 21



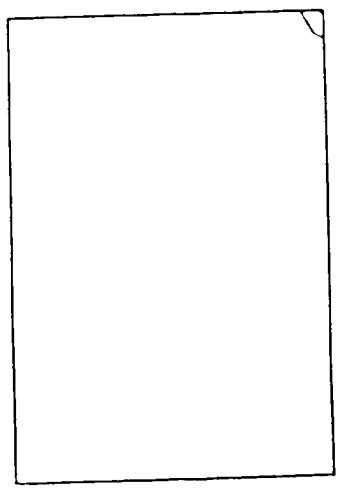
SECTION 22



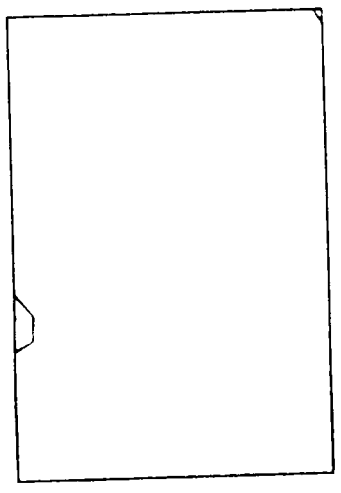
SECTION 23



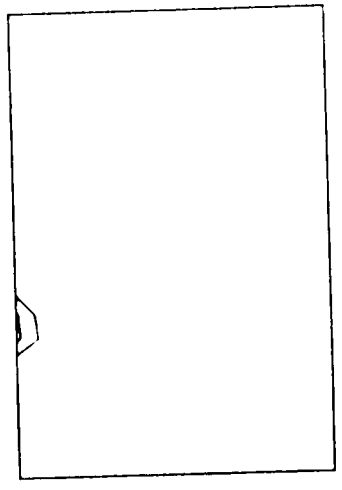
SECTION 24



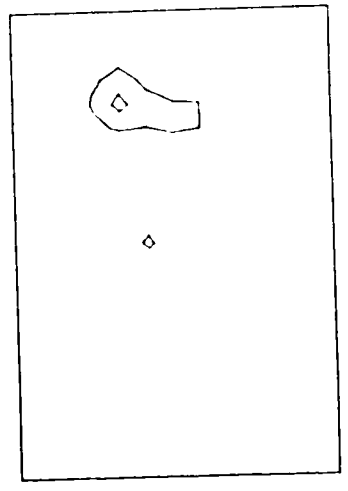
SECTION 25



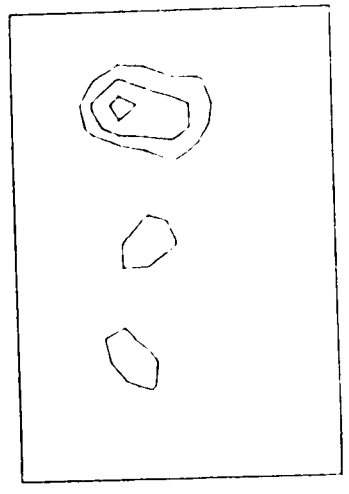
SECTION 26



SECTION 27

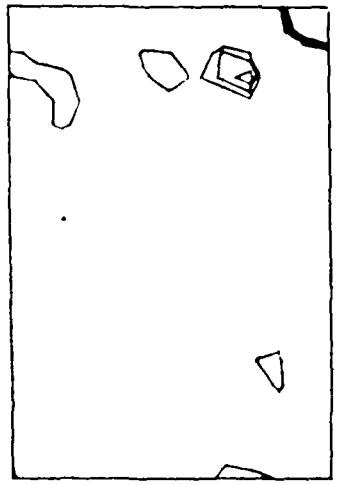


SECTION 28

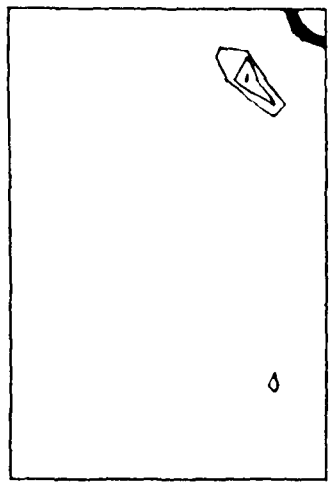


(h) EMF (emo)

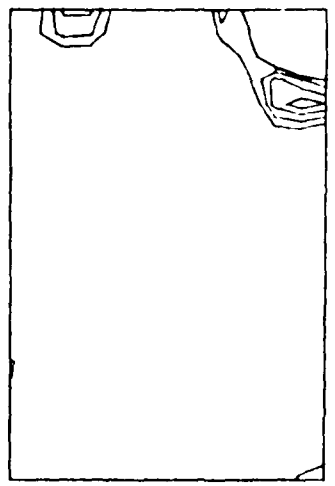
SECTION 0



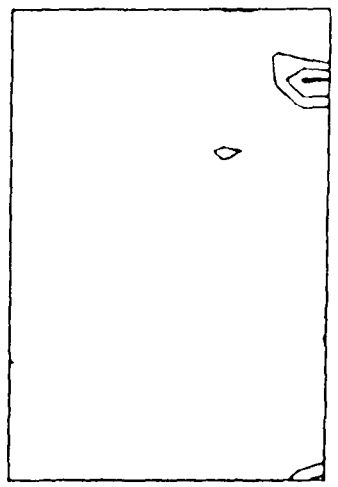
SECTION 1



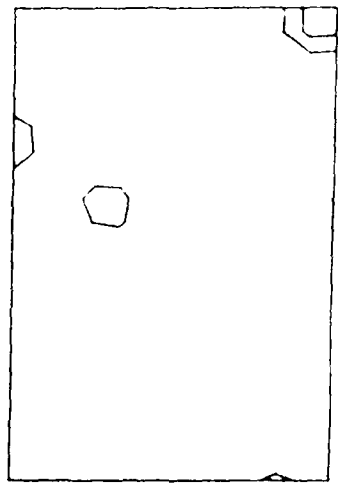
SECTION 2



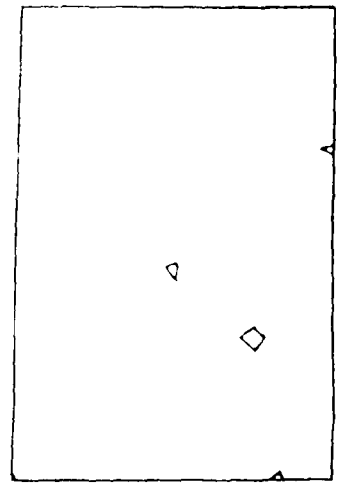
SECTION 3



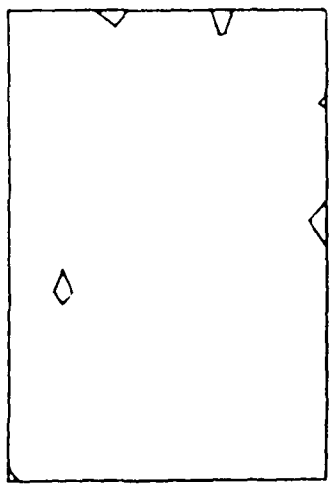
SECTION 4



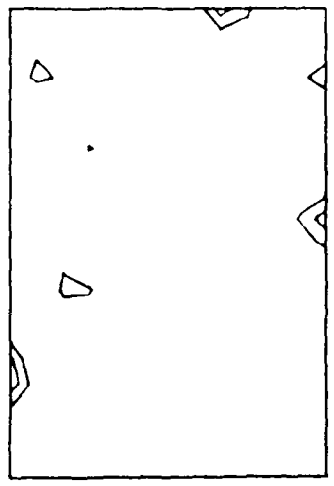
SECTION 5



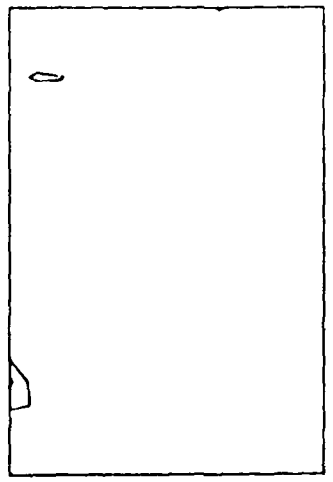
SECTION 6



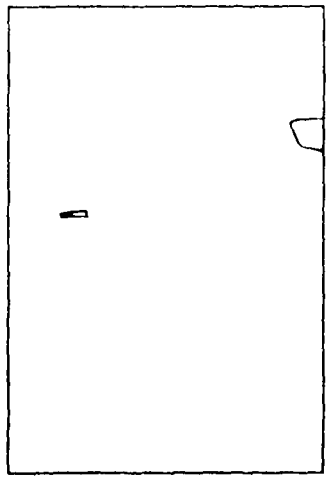
SECTION 7



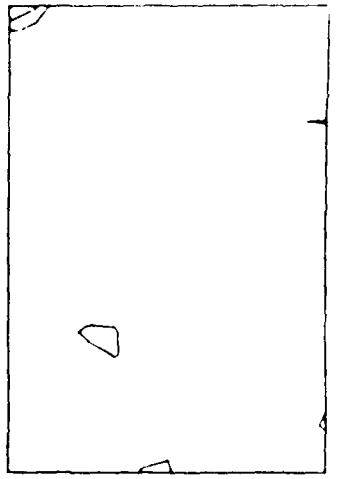
SECTION 8



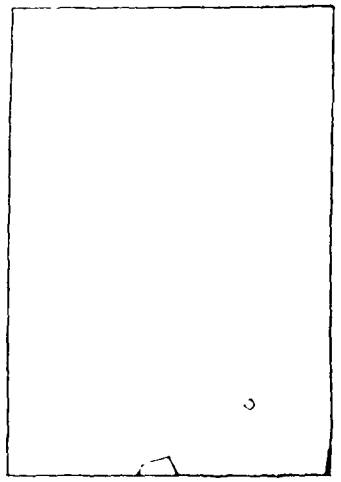
SECTION 9



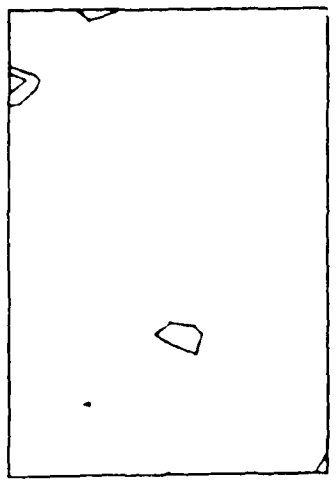
SECTION 10



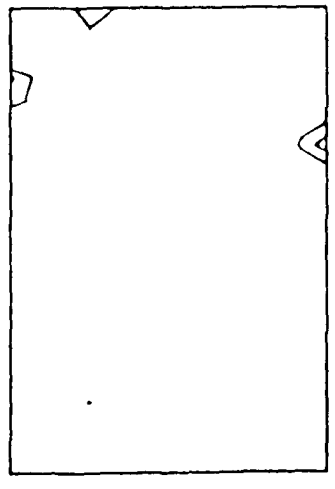
SECTION 11



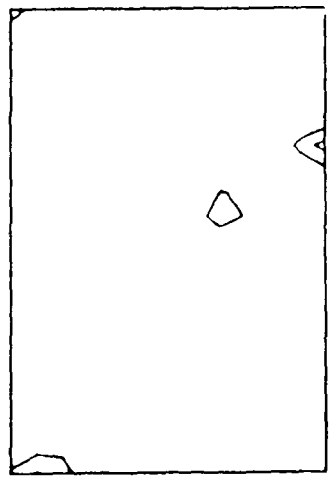
SECTION 12



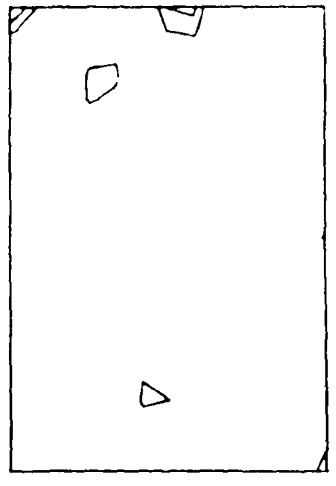
SECTION 13



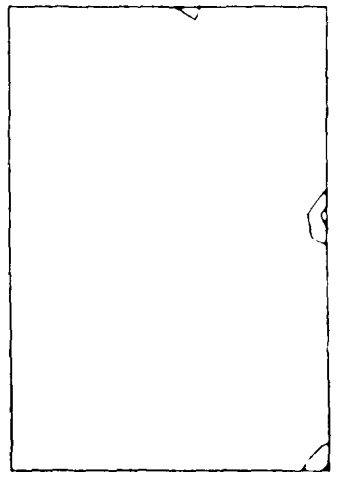
SECTION 14



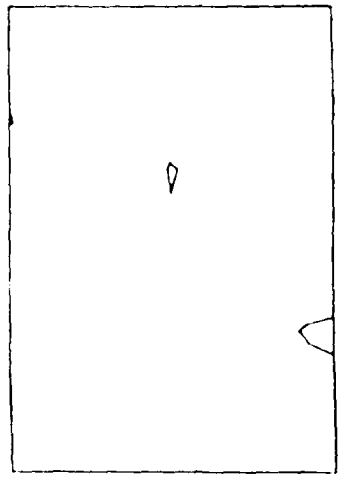
SECTION 15



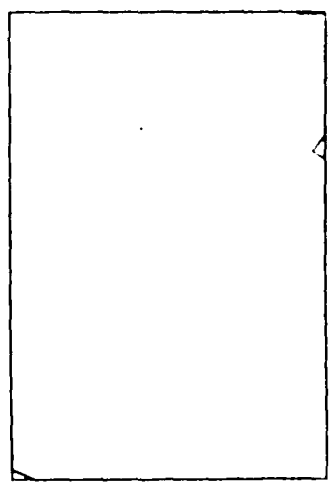
SECTION 16



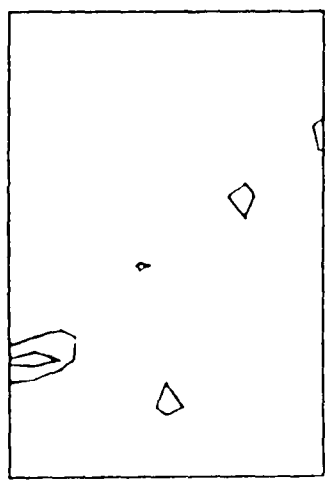
SECTION 17



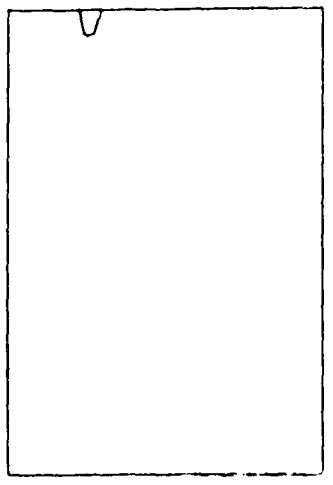
SECTION 18



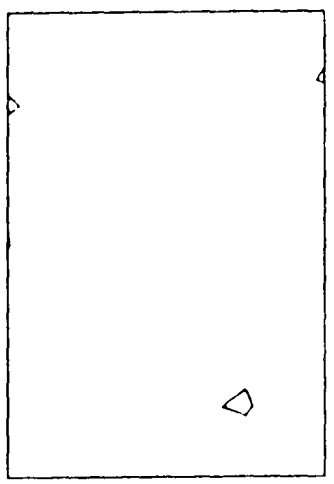
SECTION 19



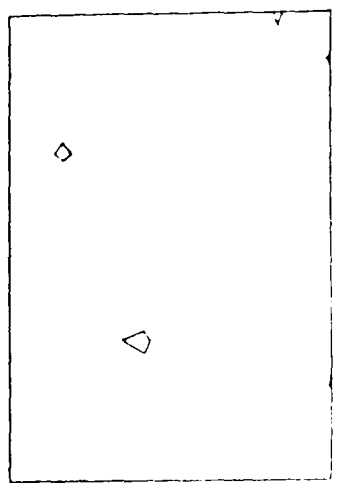
SECTION 20



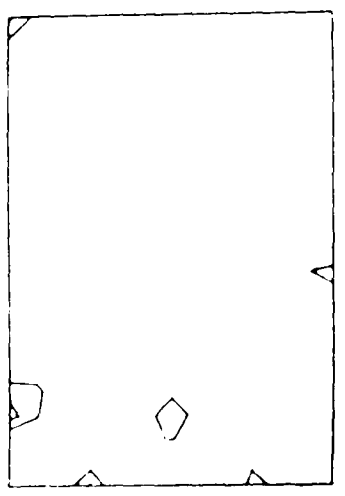
SECTION 21



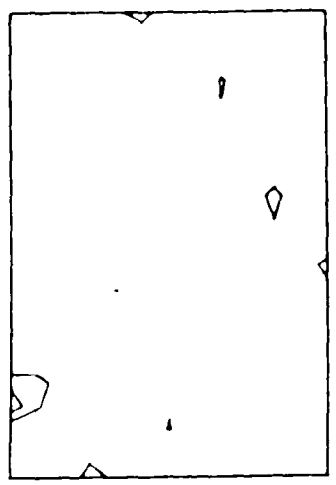
SECTION 22



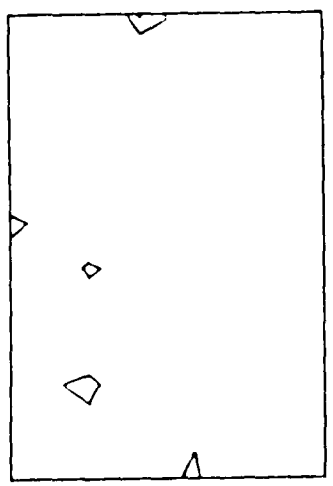
SECTION 23



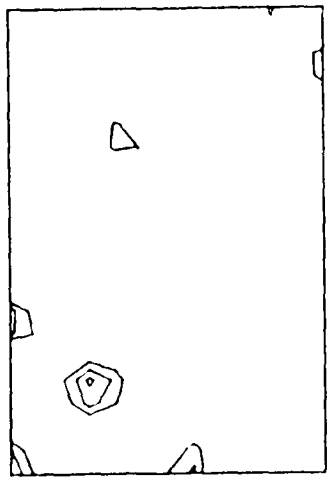
SECTION 24



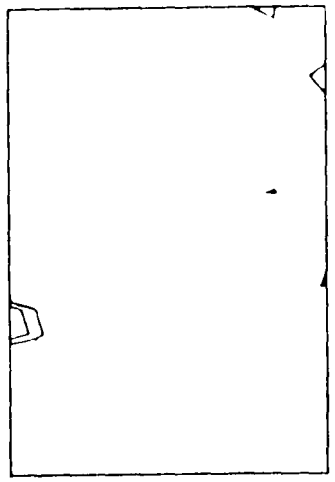
SECTION 25



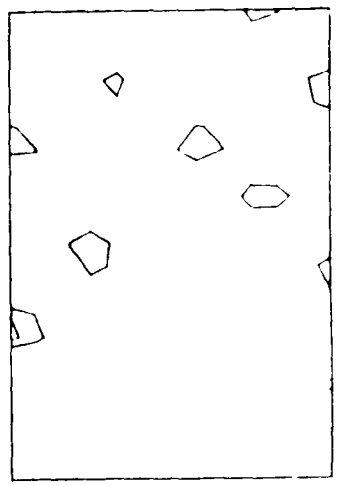
SECTION 26

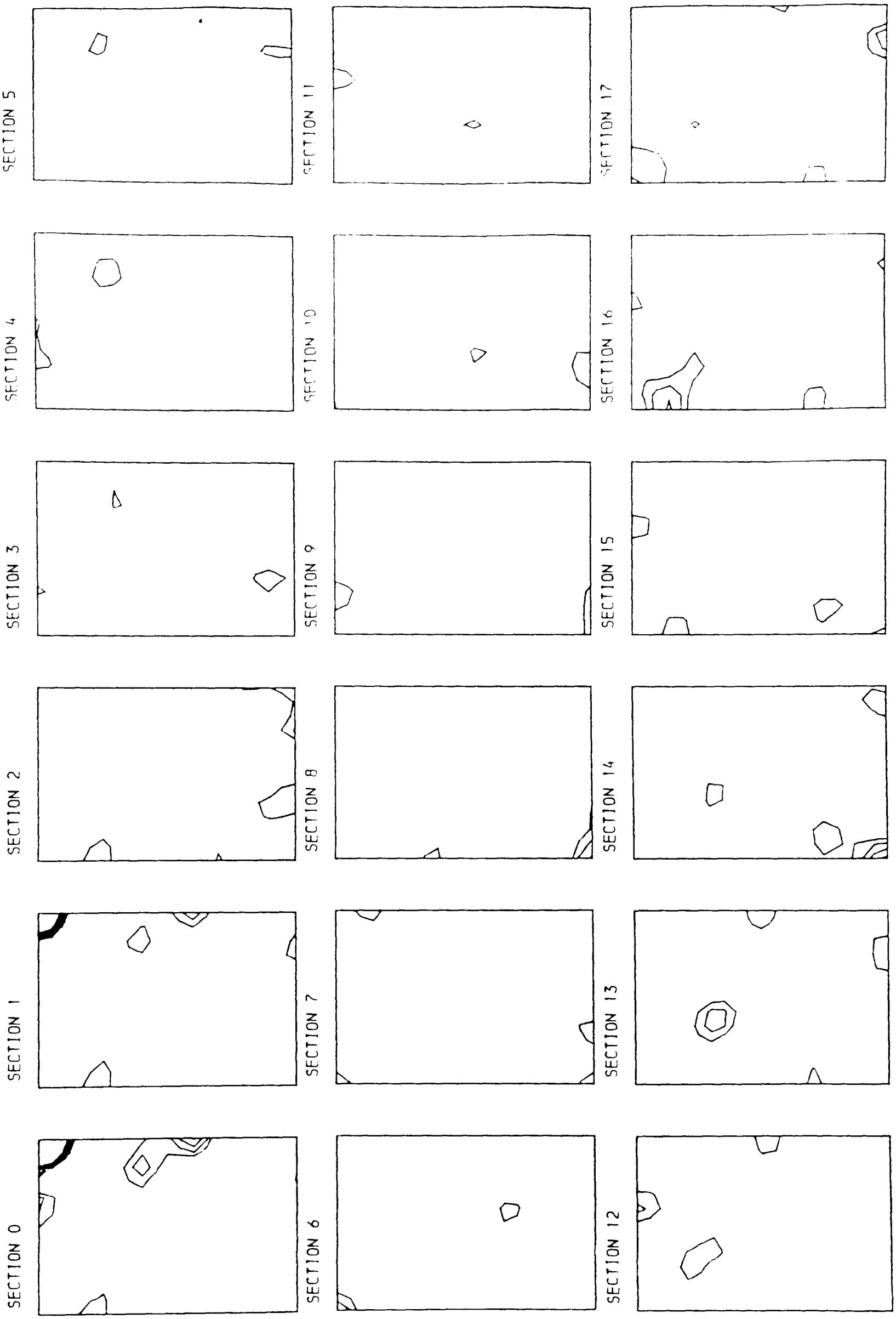


SECTION 27

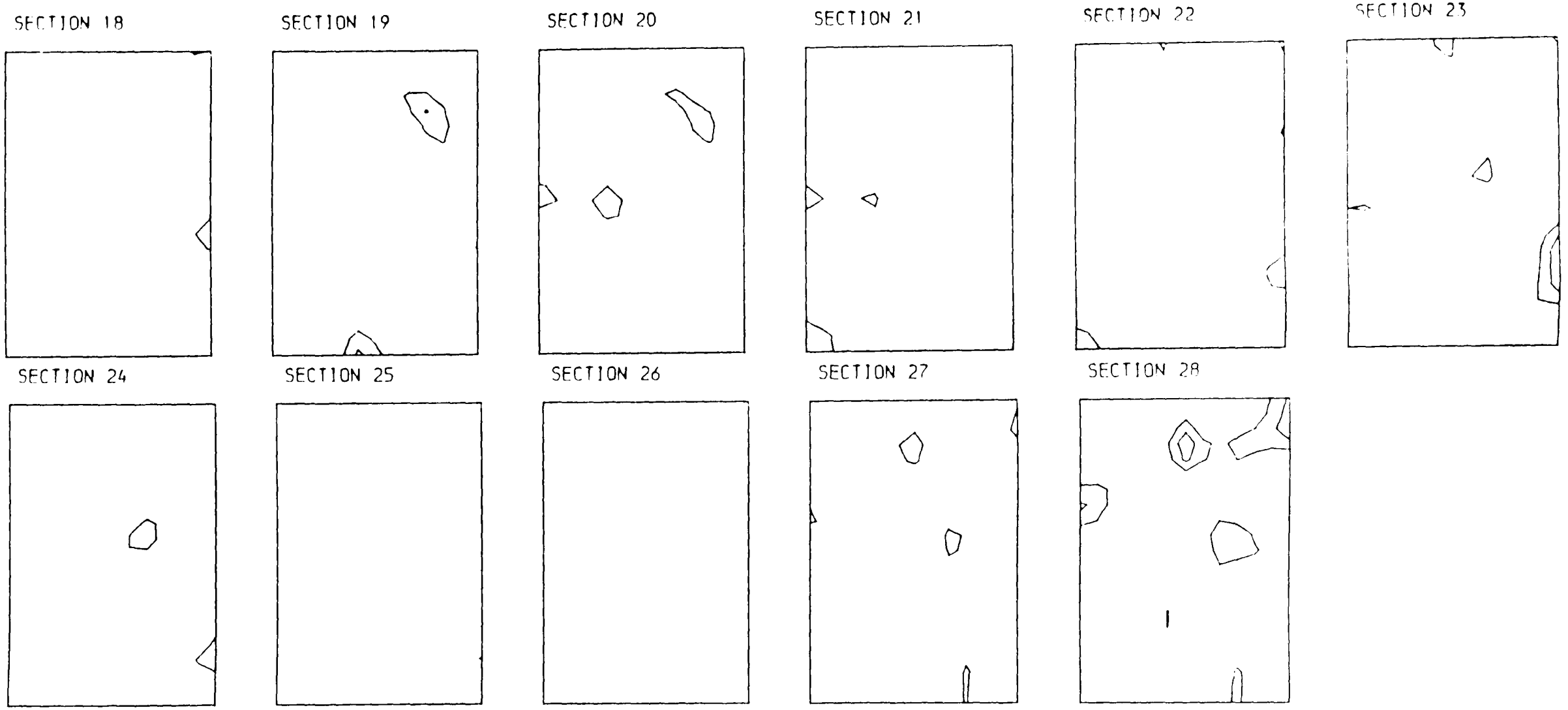


SECTION 28

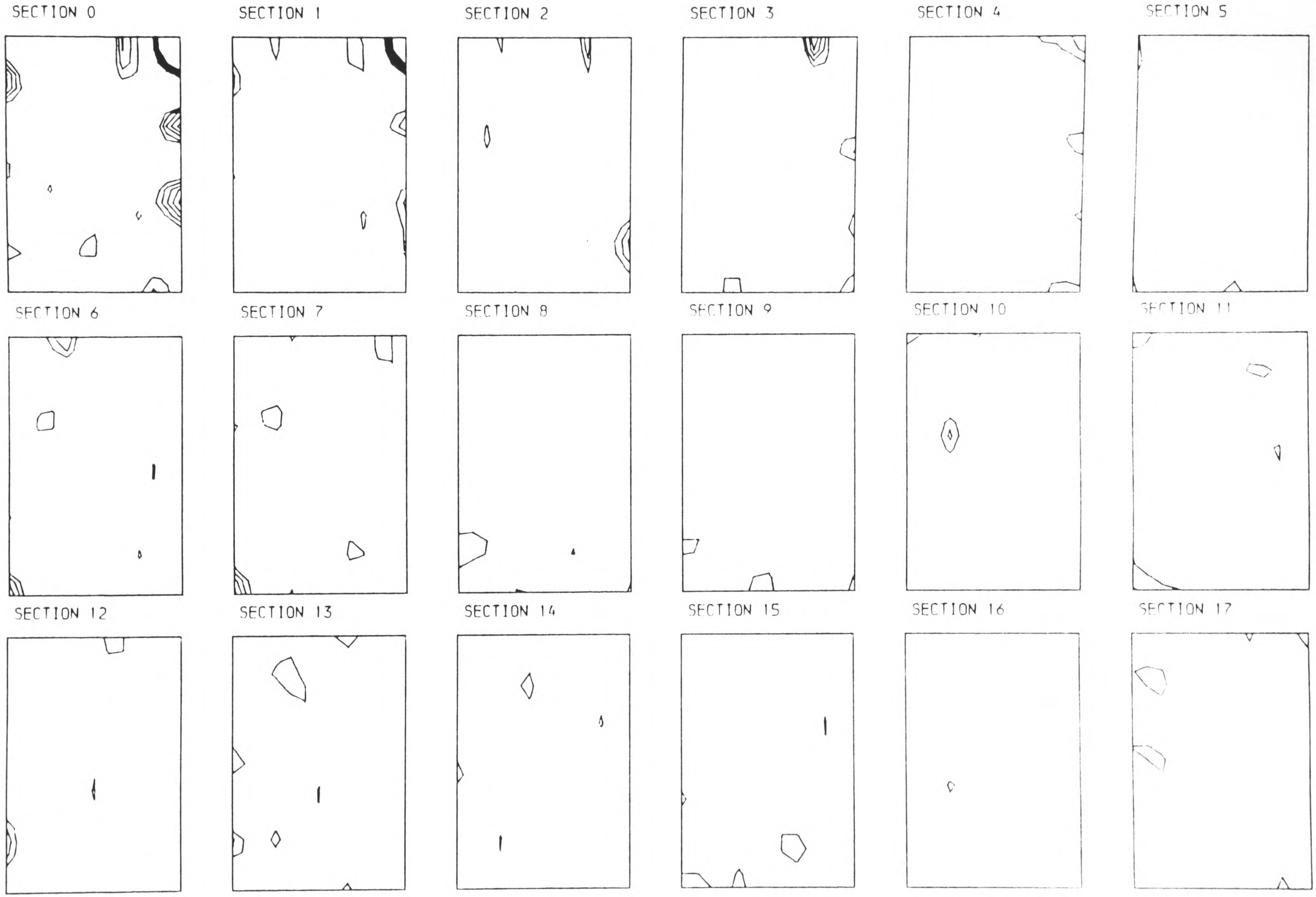




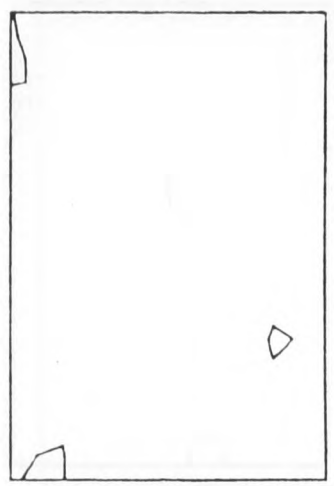
(i) uranyl fluoride (iso)


SECTION 18
SECTION 19
SECTION 20
SECTION 21
SECTION 22
SECTION 23
SECTION 24
SECTION 25
SECTION 26
SECTION 27
SECTION 28

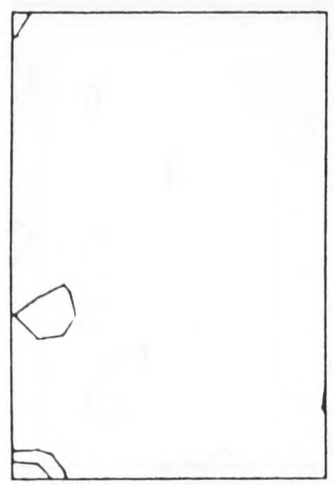
(j) manganese free (iso)



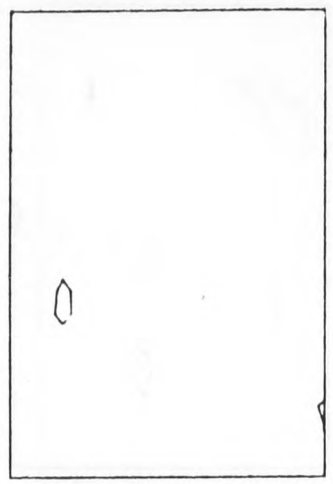
SECTION 18



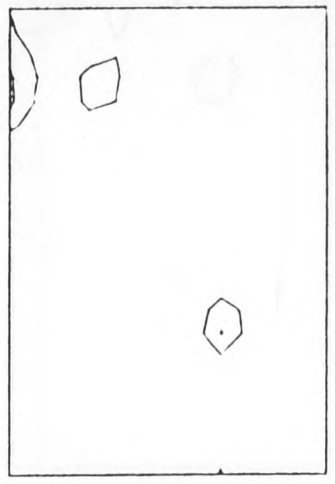
SECTION 19



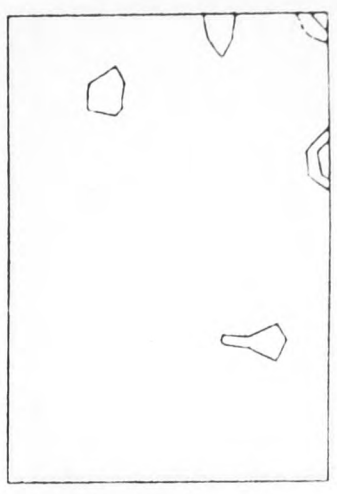
SECTION 20



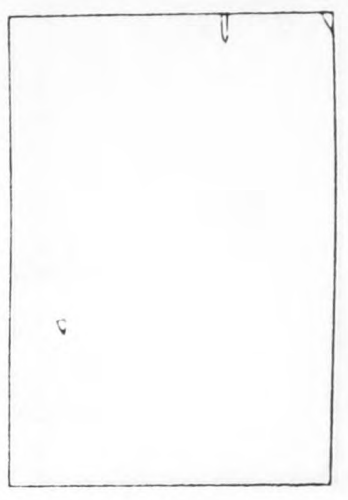
SECTION 21



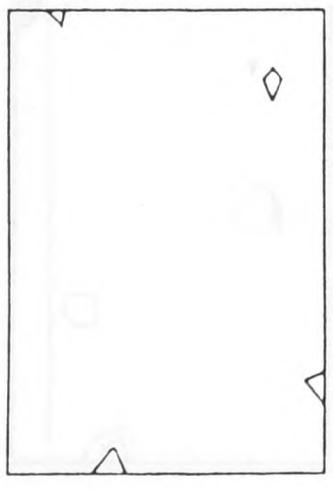
SECTION 22



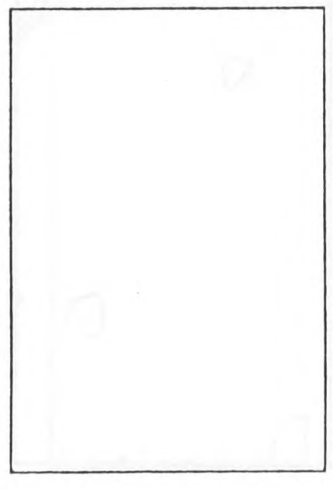
SECTION 23



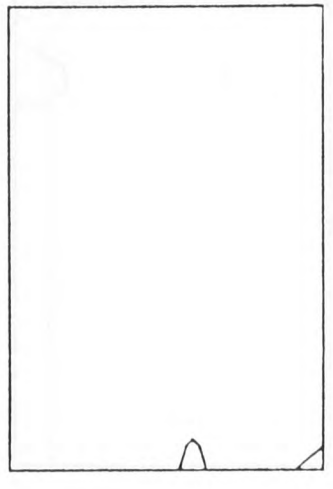
SECTION 24



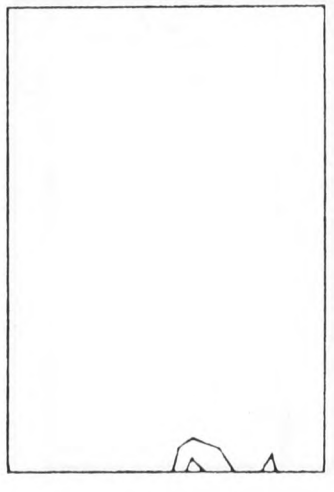
SECTION 25



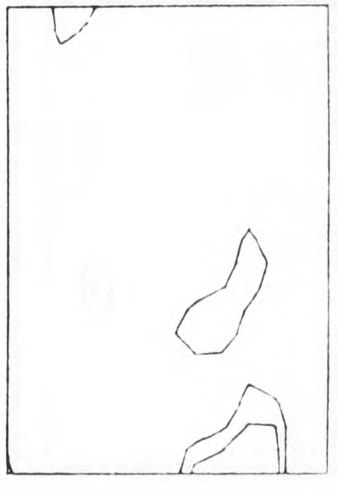
SECTION 26



SECTION 27



SECTION 28



Appendix 2

Microfiche Listings

Structure factors for:

(1) native type 1 to 2.4Å resolution,

(2) native type 2 to 6Å resolution and

derivative 1 uranyl acetate

derivative 2 gold chloride

derivative 3 samarium acetate

derivative 4 manganese-free

derivative 5 EMP

derivative 6 uranyl fluoride

Format:

hkl s F_p σ_p F_{der1} σ_{der1} Δ_{der1} σ_{der1} F_{der2} σ_{der2} Δ_{der2}
F_{der3} σ_{der3} Δ_{der3} σ_{der3} F_{der4} σ_{der4} F_{der5} σ_{der5}
Δ_{der5} σ_{der5} F_{der6} σ_{der6} fm α_p F_{calc} α_{calc}

where hkl are the reflection indices

s is $4\sin^2\theta/\lambda^2$

F is the structure factor amplitude of a reflection

p is native

der is derivative

σ is standard deviation associated with measurement preceeding it

Δ is anomalous difference

fm is figure-of-merit

α is calculated native phase (°)

F_{calc}, α_{calc} are the calculated heavy atom structure factor amplitudes and phase angles for the uranyl acetate derivative.

All data measured by diffractometry.

Accompanying material
stored separately.
Search OLIS for shelfmark.

2

1

Appendix 3

Computer Programs

Appendix 3

Computer Programs

The following computer programs were used in the course of the work. The authors of all programs are credited; uncredited programs were written by me.

1. Data Reduction

1.1 Diffractometry

All programs on the Oxford University's ICL 2988 computer.

PROPRINT (W.C.A. Pulford; modified by P.J. Artymiuk) Reads intensity profiles from magnetic tape onto disc.

STROCBLIT (P.J. Artymiuk) Selects appropriate reflection families for Huber-Kopfmann absorption measurements.

HKABS (R. Huber & G. Kopfmann; modified by D.W. Rice & S. Gover) Calculates an absorption surface from an input set of absorption curves and performs an agreement analysis on the surface.

LPOPT5 (P.R. Evans & D.W. Rice, modified by P.J. Artymiuk) Applies Lorentz and polarisation corrections for diffractometer data collected in various settings. Applies counter-to-counter scales correction and North, Phillips and Mathews or Huber-Kopfmann absorption corrections.

RADDAM (P.J. Artymiuk) Applies input radiation damage correction to data from a specified resolution shell.

SSM (E.J. Dodson; modified by K.S. Wilson, E.A. Stura & S. Gover) Applies input scales, sorts equivalent reflections, calculates batch scale factors by the method of Fox & Holmes (1966) and merges reflections for output.

FRIEDANAL (K.S. Wilson & P.R. Evans) Analyses agreement between

intensities of equivalent and Friedel-related reflections.

FRIEDCORR (K.S. Wilson & P.R. Evans) Corrects data for secondary absorption using scales output from FRIEDANAL.

TRUNCATE (S. French & K.S. Wilson) Truncates reduced intensities using prior knowledge of Wilson's (all positive) distribution for acentric and centric terms.

1.2 Oscillation Photography

All programs on the VAX 11/750 at S.R.S., Daresbury except ROTAVATA and AGROVATA which are on the NAS AS/7000 at S.R.S, Daresbury and PREIDX which is on the PDP 11/70 at Oxford.

PREIDX (D.I. Stuart) Input of reflections for determination of crystal orientation via the Evans & Sutherland tablet.

IDXREF (J. Nyborg et al., 1975) Refines camera constants, crystal unit cell dimensions and orientation parameters on the basis of a set of partially recorded reflections.

GENVEE (A.J. Wonacott) Spot prediction from crystal and camera constants as well as orientation parameters.

MOSFLM (J. Nyborg & A.J. Wonacott, 1977; A.J. Wonacott, S.Dockerill & P. Brick) Integration of spot intensities.

POSTCHK (A.J. Wonacott) Finds complementary partials on adjacent films and selects a subset to be data for a post-refinement run of IDXREF.

ROTCOR (A.J. Wonacott) Calculates absorption, oblique incidence and Lp corrections and applies them.

PASCAL (A.J. Wonacott) Scales films belonging to a pack together.

ROTAVATA (J.M. Smith & A.J. Wonacott) Determines relative scales and temperature factors for overlapping batches of data using a modification of the Fox & Holmes method (1966).

AGROVATA (J.M. Smith & A.J. Wonacott) Merges and scales data sets. Calculates statistics.

1.3 Precession Photography

All programs on the in-house PDP 11/70 computer. Programs implemented by D.I. Stuart and R. Todd.

AV2DSC (D.I. Stuart) Copies optical density bytes written on tape by microdensitometer onto disc.

PRESTO (J. Westman) Performs equivalent functions to IDXREF and GENVEE as discussed above. Spot positions are input via the Evans & Sutherland tablet.

MOSCO (J. Nyborg & A.J. Wonacott, 1977; rewritten by M. Levine, D. Stammers and D.I. Stuart) Oxford version of MOSFLM (see above).

DATRED (M. Levine) Applies Lp correction and performs intra-pack scaling.

SCALIN (D.I. Stuart) Averages symmetry-related reflections, evaluates anomalous dispersion differences, scales and merges data sets.

2. Data Analysis

The following programs are on the University's ICL 2988 computer.

CAD (K.S. Wilson, E.J. Dodson, R. Pickersgill) Collects together items associated with a particular reflection, from various sources into a single file.

ANSC (A.C. Bloomer & P.R. Evans; extended by S.Gover) Calculates scale factors between two data sets and analyses them as a function of several variables. Applies scales and calculates isomorphous differences and k_{emp}.

RMS Calculates r.m.s. anomalous dispersion difference for a data set for input to ANOMSORT.

ANSCCORR Applies three dimensional scales output from ANSC to data.

ANOMSORT Calculates statistics and rejects unreliable anomalous dispersion data.

ISOSORT — As for ANOMSORT but looks at isomorphous differences.

LOCAL — (I.J. Tickle) Matthews & Czerwinski (1975) local scaling method for two intensity data sets.

WILSON — (S. Gover) Calculates Wilson plot and fits the plot by linear regression.

FFT — (L. Ten Eyck, E.J. Dodson, J. Fail, J. Kirz, P.R. Evans, & W.C.A. Pulford) Fast Fourier synthesis which uses Cooley-Tukey algorithm.

The following programs were implemented by D.I. Stuart onto the PDP 11/70 at Oxford.

DIFFER — (P. Shaw, D.I. Stuart, J. Westman) Scales two data sets together and calculates statistics.

EXPAND — (R. Fisher) Converts a crystallographic asymmetric unit of data into a hemisphere of data in space group P1.

FFT — (R. Fisher, D.I. Stuart) Fast Fourier transform program.

3. Refinement and Phase Programs

3.1 General

The following programs are on the University's ICL 2988 computer.

REFINE — (W.R.Busing et al., 1962; P. Main, modified by K.S. Wilson, E.J. Dodson, S. Gover) Least squares refinement of heavy atom positions using F_{HLE} or phase information.

FHLEKR — (G. Bentley, P.R. Evans, E.J. Dodson, K.S. Wilson) Uses procedure of Arnone et al. (1971) to calculate derivative-to-native scale factor by iterations of Kraut's method (Kraut et al., 1962) and outputs F_{HLE} and F_{HLE} (bias corrected) coefficients.

PHASE — (H. Muirhead, G. Bentley, M.Lewis, M.J. Geisow, S. Gover) Calculation of isomorphous phases by the method of Blow & Crick (1959).

PHASEANAL — (E.J. Dodson) Analyses agreement between two phase sets.

PATTPKS (K.S. Wilson; modified by author) Calculates self- and cross-vectors from a set of atomic positions. Space group general.

REALPKS Calculates symmetry-related atomic positions for input to CONTOUR program. $P2_12_12$ specific.

3.2 The Wang Suite of Programs

The following programs were written by B-C Wang (1985) and have been implemented on the University VAX 11/780 computer.

SIRSAS Prepares SIR or SAS phase probability distributions, figures-of-merit, etc on master data file.

FSFOR Reciprocal to real space Fourier transformation.

ENVELP Determines protein-solvent boundary from electron density map.

DSFLT Real space filtering program.

FORINV Real to reciprocal space fast Fourier inversion program.

PSFLT Reciprocal space filtering and phase extension.

PHSALL Outputs card image data file.

MERG2 Scales and merges data and combines phases.

The following programs are on the University's ICL 2988 computer.

KARDSIR (B. Samaroui) Converts binary files on ICL 2988 to card image for use by the Wang programs.

ISIRBIN (B. Samaroui) Converts Wang card image output file to binary for use on ICL 2988.

3.3 Rigid Body Refinement Programs

The following programs are on the in-house Vax.

PQRXYZ (D. Stuart) Converts protein data bank format to format suitable for CORELS.

PRECOR (Sussman et al., 1977) Prepares various input files for CORELS.

CORELS (Sussman et al., 1977) Rigid body refinement program.

COR2PDB (D. Stuart) Converts coordinates output by CORELS into protein data bank format.

4. Molecular Replacement Programs

The following programs are on the University's ICL 2988 computer.

MATMULT2 Generates dimer coordinates from monomer coordinates and the appropriate transformation matrix.

MATMULT Matrix multiplication of coordinate sets.

MOVESOD Translates coordinate sets relative to a chosen origin.

NAGSUP (J.M. Thornton) Calculates transformation matrix between two sets of coordinates.

CGCENTRE Calculates centre-of-gravity of molecular coordinates.

CENTER (S. MOWBRAY) Locates maximum and minimum coordinates with respect to the cartesian coordinate frame.

P1BOX (unknown, modified by author) Orthogonalises coordinates and puts them into a P1 box with centre-of-gravity at the origin of the box.

FC (unknown, modified by author) P1 structure factor calculation.

SEL (E.A. Stura) Selects 1500 strongest reflections for input to FROTF.

FROTF (E.J. Dodson, implemented by G.Taylor) Crowther's fast rotation function.

MATRIX Converts Euler angles to orthogonal 3X3 matrix.

CADG (G. Taylor) Merges P1 data into unique mmm set and calculates partial structure factors for symmetry-related molecules in the unit cell.

SELECT (P.R. Evans, implemented by G.Taylor) Least squares fit of native and calculated structure factors.

SEARCHZ (E.J. Dodson, implemented by G. Taylor) Translation search using the product-type correlation function of Lifchitz

(Harada et al., 1981).

RSEARCH (E.J. Dodson, implemented by G. Taylor) Translation search using a conventional R-factor calculation.

The following program is on the PDP 11/70.

SEL (E.A. Stura, D.I. Stuart) Converts Euler angles to spherical polar angles.

5. Graphics Programs

The first four below-mentioned programs are on the ICL 2988 whilst the rest are on the PDP 11/70 or the in-house Vax.

CONTOUR (S.H. Banyard, P.E. Nixon, P.R. Evans, K.W. Snape, S.J. Oatley) Contours maps from FFT output.

NORTH Reformats coordinate file for input to SDRAW.

SDRAW (A.C.T. North, S.J. Oatley) Draws atomic coordinates on contoured map sections.

GRAPH (S.J. Oatley) Multi-purpose plotting program.

FORIN (D.I. Stuart, W.C.A. Pulford) Reads magnetic tape written from ICL 2988.

CONVERT (D.I. Stuart) Converts serial access file to direct access file.

MAO (D.I. Stuart) Allows extension of an asymmetric unit of a Fourier map using symmetry operators.

CONTOR (A. Jones) Three dimensional contouring program.

SAM (A. Jones) Creates and manipulates coordinate data sets for input to FRODO.

FRODO (A. Jones) Graphics package used in conjuction with the Evans & Sutherland picture system that allows a protein model to be built into an electron density map.

CAFITZ (G. Taylor) Prepares an alpha carbon coordinate file for input to FITZ.

FITZ (G. Taylor) Graphics package to allow manipulation and comparison of molecular structures.

PSP (G. Taylor) Transfers images from Evans & Sutherland picture system screen to Benson pen plotter.

6. Miscellaneous Programs

All the following programs can be found on the University's ICL 2988 computer.

CROSSEC (D.T. Cromer, implemented by S. Gover) Interpolates cross-section data in order to compute real and imaginary dispersion corrections for X-ray atomic scattering factors. Also calculates mass absorption coefficients based on photoelectric contributions.

AACOMP Calulates amino acid composition from sequence.

ANTIGEN Prediction of antigenic determinants by the method of Hopp & Woods (1981).

BACKTRANS Converts protein sequence to ambiguous DNA sequence.

BAZCOD (P.J. Artymiuk, modified by author) Calculates best alignment of two sequences by mean base change per codon method.

HOMOLOGY Calculates degree of residue identity between two sequences.

PHOMOLOGY Similar to HOMOLOGY but takes into account residue similarity according to the similarity matrix of McLachlan (1971).

TREE Calculates mean base change per codon and other statistics for two sequences.

ROBSON (Garnier et al., (1978), implemented by M.J.E. Sternberg) Secondary structure prediction.

LIM (E. Eliopolus; implemented by M.J.E. Sternberg) Secondary structure prediction by the method of Lim (1974).

CHOU1, CHOU2 (M.J.E. Sternberg & B. Robson) Secondary structure prediction by the method of Chou & Fasman (1974).

Appendix 4

Reprints

Redacted article "Chemical aspects of the structure, function and evolution of superoxide dismutases" is available at:
http://dx.doi.org/10.1016/S0020-1693(00)81854-X

Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 149–152, January 1985
Evolution

The presence of a copper/zinc superoxide dismutase in the bacterium *Photobacterium leiognathi*: A likely case of gene transfer from eukaryotes to prokaryotes

(amino acid mutation/secondary structure/model building)

JOE V. BANNISTER AND MICHAEL W. PARKER

Inorganic Chemistry Laboratory and Laboratory of Molecular Biophysics, University of Oxford, Oxford OX1 3QR, England

Communicated by Irwin Fridovich, September 4, 1984

ABSTRACT **The free-living bacterium *Photobacterium leiognathi* is also known to be a symbiont of ponyfish. The presence of a copper/zinc superoxide dismutase in *P. leiognathi* has been considered to be a case of gene transfer from eukaryotes to prokaryotes because this form of superoxide dismutase is normally present only in higher eukaryotic species. However, the amino acid sequence of the enzyme from the bacterium exhibited low identities (25–30%) with the same enzyme from eukaryotes. When amino acid mutations are taken into consideration, the weighted sequence similarity increases significantly; furthermore, the bacterial enzyme has the same active site residues and similar predicted secondary structure as the eukaryotic enzymes. The possibility of convergence is ruled out and the case of divergence is considered unlikely because of the observed phylogenetic distribution of the enzyme. This indicates that the presence of the copper/zinc superoxide dismutase in *P. leiognathi* can indeed be considered a case of gene transfer from eukaryotic species to prokaryotic species.**

The first line of defense against the toxicity of active oxygen species is removal of the superoxide anion radical [O_2^-] (1, 2). The enzymes called superoxide dismutases (EC 1.15.1.1), discovered by McCord and Fridovich (3), scavenge the superoxide radical [$2O_2^- + 2H^+ \rightarrow H_2O_2 + O_2$]. These enzymes are a group of metalloproteins containing either copper/zinc, manganese, or iron as the prosthetic group. The copper/zinc protein is now considered to be characteristically found in all higher eukaryotic cells. A report indicating its absence in adipocytes (4) has been shown to be in error (5). The manganese protein is present in prokaryotes and in the mitochondria of eukaryotic cells, while the iron protein is usually present in prokaryotes. A number of exceptions to these observations have recently come to light. The iron enzyme has, in addition to the copper/zinc and manganese enzymes, been isolated from the mustard plant *Brassica campestris* (6), and a survey of 43 plant families for the presence of the iron enzyme showed that it is present in three isolated families (7). Two bacterial species, *Photobacterium leiognathi* (8) and *Caulobacter crescentus* (9), have been shown to contain the copper/zinc superoxide dismutase although this form of the enzyme has been found to be absent from eukaryotic species more primitive than green algae such as protozoans and *Euglena* species (10).

The presence of a copper/zinc superoxide dismutase in *C. crescentus* is as yet an unexplained anomaly. However, the presence of this enzyme in *P. leiognathi* has been considered to be a case of gene transfer from eukaryotic species to prokaryotic species (11, 12). This is because this bacterial species, although found free-living, is also a symbiont of ponyfish (13). Using the amino acid sequences of the copper/zinc superoxide dismutases isolated from *P. leiognathi* (14) and from the swordfish *Xiphias gladius* liver (15), we compare the two sequences. The swordfish liver enzyme has been shown to be closely related to the unsequenced enzyme of the bacterial host, the ponyfish (12). Considering amino acid mutations, predicted secondary structure, and model building, a case for convergent evolution is shown to be unlikely. On the basis of phylogenetic distribution the case for the bacterial enzyme evolving prior to the prokaryote–eukaryote split is considered also unlikely. The most plausible case is natural gene transfer from the host fish to the symbiotic bacterium.

MATERIALS AND METHODS

Complete amino acid sequences of copper/zinc superoxide dismutases from *P. leiognathi* (14) swordfish liver, (15) bovine erythrocyte (16), horse liver (17), and yeast (18) are known. The sequence alignment of the eukaryotic enzymes is taken from the literature (19), and the alignment of the bacterial to the other enzymes was carried out by maximizing identity between the sequences together with the use of gap penalties (penalty of −2 identities). For ease of comparison, the corresponding residue numbers in the bovine erythrocyte enzyme sequence have been included here.

Two methods were used to calculate the degree of sequence homology between the higher eukaryotic and *P. leiognathi* enzymes. The first method calculates the percentage of perfect identity between two aligned sequences. The second takes into account mutations in the gene coding for the proteins. This was carried out according to McLachlan (20). The method is based on frequences of observed substitutions in families of homologous proteins, and the percentage similarity between the superoxide dismutases was determined from McLachlan's frequency matrix. For example, identities in the sequences gained a score of 8 or 9 out of 9 depending on the residue. A score of 3 was average and a score of 0 was given for gaps. Common gaps in the aligned sequences were excluded from the calculation. This method is referred to here as a measure of the weighted degree of similarity.

Secondary structures were calculated using algorithms that predict conformation from amino acid sequence. Predictions were carried out according to Garnier *et al.* (21), Lim (22), and Chou and Fasman (23). An overall combined prediction was derived from the individual methods based on the requirement that at least two out of the three methods agreed for each of the structural elements predicted.

Model building was carried out using an interactive computer graphics system (Evans and Sutherland picture system II) equipped with the FRODO program (24). The program was used to delete, insert, and replace residues and also allowed for manual rotations about bonds so as to eliminate bad contacts. A model was built for the bacterial enzyme using the coordinates for the bovine erythrocyte copper/zinc

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

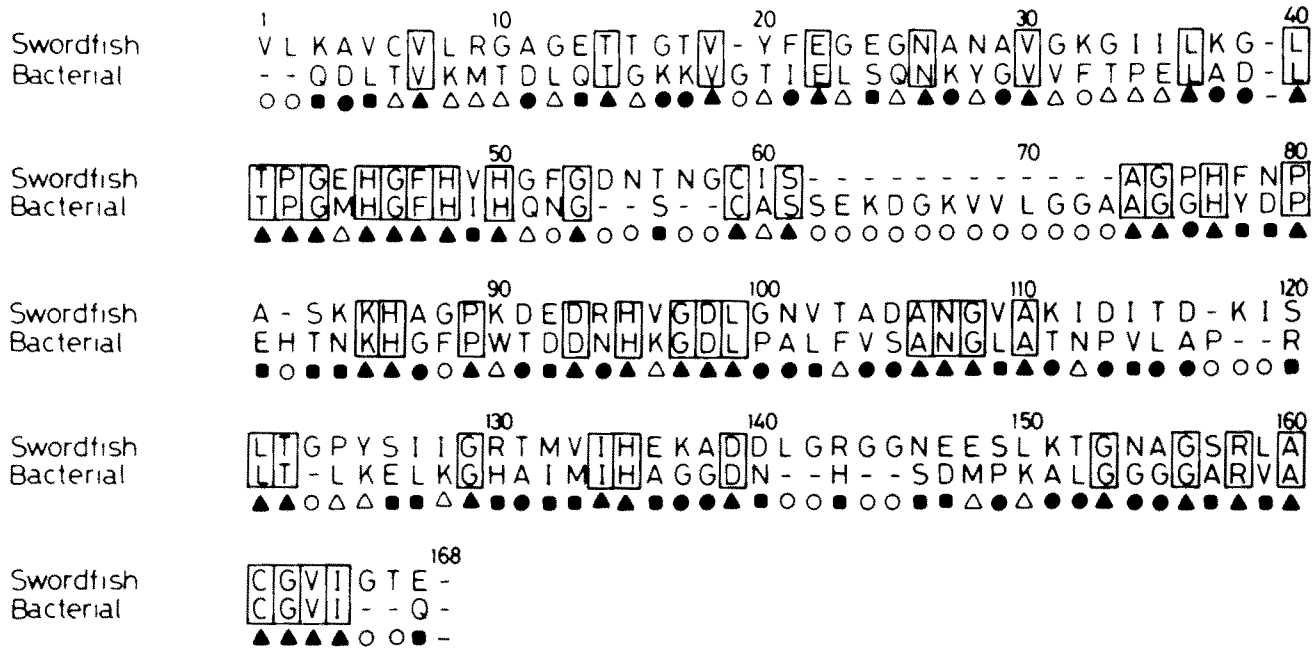


FIG. 1. Amino acid sequences of swordfish liver and bacterial (*P. leiognathi*) copper/zinc superoxide dismutases. Identical residues are boxed. Symbols refer to the frequency of substitutions determined according to McLachlan (20): ▲, identity; ■, frequently found substitution; ●, neutral substitution; △, infrequent substitution; ○, rarely found or no substitution. Common gaps are due to alignment to other eukaryotic enzymes (not shown).

superoxide dismutase available from the Brookhaven Protein Data Bank (25).

RESULTS AND DISCUSSION

Copper/zinc superoxide dismutase isolated from swordfish liver (26) and from *P. leiognathi* (12) has been found to have the same catalytic activity, metal content, and subunit size as the bovine erythrocyte enzyme. The amino acid sequences for the bacterial and swordfish enzymes are shown in Fig. 1. The alignment indicates that a number of residues previously determined to be important for metal binding, enzymatic activity, and maintenance of tertiary structure in the bovine enzyme (14, 15, 18, 27) are conserved in the bacterial protein as well as in all the other sequenced eukaryotic enzymes (19). These are His-48 (bovine 44), -50 (bovine 46), -77 (bovine 61), and -135 (bovine 118) as ligands for the copper; His-77 (bovine 61), -86 (bovine 69), and -95 (bovine 78), and Asp-98 (bovine 81) as ligands for the zinc; and Arg-158 (bovine 141) and Asp-139 (bovine 122), which are considered to be essential for activity and the intrachain disulfide bridge formed by Cys-59 and -161 (bovine 55 and 144). Out of 49 glycine residues in the bacterial and swordfish enzymes, 20 are conserved while, out of 13 proline residues, 6 are conserved.

The degree of identity and weighted similarity for the sequences are given in Fig. 2. The amino acid sequences of bacterial and swordfish liver enzymes indicate little identity relative to identities between the eukaryotic enzymes. However, when mutations are taken into account long stretches

	Bacterial	Swordfish	Bovine	Horse	Human	Yeast
Bacterial		44	44	44	43	43
Swordfish	30		73	72	72	63
Bovine	28	71		78	80	65
Horse	26	66	78		79	66
Human	25	66	82	78		66
Yeast	26	53	53	56	55	

FIG. 2. Sequence similarities between copper/zinc superoxide dismutases. The left-hand side of the diagonal shows percentage identity figures and the right-hand side shows weighted similarities in percentages.

of closely related sequences are obtained (Fig. 1) and the weighted similarity between the two enzymes rises to 44%. This corresponds to 78 identical or closely related (above average substitution) residues out of 136 residues excluding deletions and insertions.

The predicted secondary structures for the bovine erythrocyte, swordfish liver, and bacterial enzymes are presented in Fig. 3. The predicted secondary structure for the bovine erythrocyte enzyme agrees well with the structure obtained from x-ray data (27) and the prediction for the swordfish liver enzymes is similar to the bovine erythrocyte enzyme structure. The bacterial enzyme, despite having a low identity with the two enzymes, also has, with two exceptions, a similar secondary structure. The first exception is the 12-residue insertion, residues 62–73 (after bovine 57) in the bacterial sequence, which is, in part, predicted to form a β-strand. The methods of Garnier *et al.* (21) and Chou and Fasman (23) predict an α-helix between residues 125 and 129 (bovine 108–112), whereas Lim (22) predicts nothing and a turn is present in this region for the bovine enzyme.

The subunit structure of bovine erythrocyte copper/zinc superoxide determined from crystal structure has been found to be comprised of eight antiparallel β-strands that form a flattened cylinder plus three external loops (27). After establishing close similarity between the primary and secondary structures of the bovine erythrocyte, swordfish liver, and *P. leiognathi* enzymes, an attempt was made to see whether the bacterial enzyme could be built into the known bovine erythrocyte tertiary structure without serious disruption to the fold.

A schematic diagram of the model obtained for the bacterial enzyme is presented in Fig. 4. The model was built taking into consideration the positioning of essential and conserved residues (including closely related residues) and secondary structure predictions. All the prediction methods indicated a strand for the 12-residue insertion (residues 62–73) in the bacterial enzyme. However, despite exhaustive attempts a strand could not be built in the model so as to interact appropriately with nearby strands. The insertion was successfully built into the structure as a loop. The predicted α-helix also proved to be an unreliable prediction. The overall fold of the model could not have been retained if the α-helix had not been built as a loop of random coil. The loop positions the seventh predicted β-strand into favorable hydrogen-bonding contacts with strands six and eight and completes the fold of the β-barrel. This strand also positions two conserved resi-

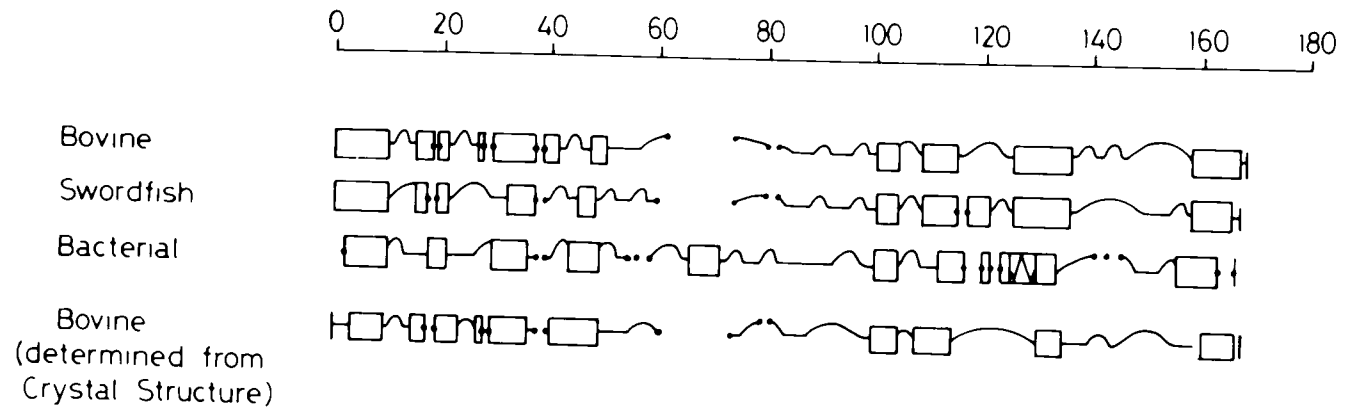


FIG. 3. Comparison of predicted secondary structures of bovine erythrocyte, swordfish liver, and bacterial (*P. leiognathi*) copper/zinc superoxide dismutases. □, β-Strand; ▨, α-helix; ∿, β-turn.

dues (His-135 and Asp-139), considered essential for activity in the bovine enzyme, near the active site. These cases demonstrate one of the problems encountered in secondary structure prediction methods. These methods do not take into account possible nonlocal interactions. The *P. leiognathi* sequence could be fitted into the fold of the bovine erythrocyte enzyme without serious disruption to the fold. No obvious steric problems were encountered as a result of the introduction of larger residues for smaller residues and all residue side chains pointing into the interior of the β-barrel were uncharged.

The presence of a copper/zinc superoxide dismutase in a symbiotic bacterial species initially appeared to be a case of direct gene transfer. However, although gene transfer among prokaryotes is a well-documented phenomenon, as is transfer from prokaryotic to eukaryotic species (28), gene transfer from eukaryotic to prokaryotic species has never been clearly proven. The amino acid sequence of *P. leiognathi* copper/zinc superoxide dismutase turned out to show low identity with bovine erythrocyte, human erythrocyte, horse liver, and yeast copper/zinc superoxide dismutases. Therefore, in 1983, the hypothesis of gene transfer was abandoned and an independent evolutionary line was suggested for the presence of the copper/zinc enzyme in a bacterial species (14). This hypothesis continued to be plausible when the sequence for the first fish copper/zinc superoxide dismutase, from swordfish liver, was determined (15) and also showed very low identity with the bacterial enzyme.

The present investigation reexamines the possibility of a direct evolutionary relationship and of gene transfer for the presence of copper/zinc superoxide dismutase in *P. leiognathi*. The possibility of convergent evolution between lowly related proteins is always difficult to exclude. However, no example in which convergent evolution has led to close similarities in structure and sequence is known. The oxygen-binding proteins—hemoglobin, hemerythrin, and hemocyanin—are examples of convergent evolution, but their primary and tertiary structures are totally unrelated.

The following evidence establishes convincingly that the *P. leiognathi* copper/zinc superoxide dismutase is directly related to the eukaryotic copper/zinc enzymes: (*i*) similar molecular weight, subunit size, metal content, and catalytic activity; (*ii*) the single disulfide bridge, the seven metal ligands, and the two other residues shown to be important in the mechanism of the bovine erythrocyte enzyme are all found in the bacterial enzyme, in the correct sequence order; (*iii*) analysis of the aligned sequences shows that more than half of the bacterial sequence has closely related or identical residues in common with the swordfish liver enzyme after deletions and insertions are excluded; and (*iv*) very similar secondary structure is predicted for bacterial, swordfish liver, and bovine erythrocyte enzymes. The case for a diver-

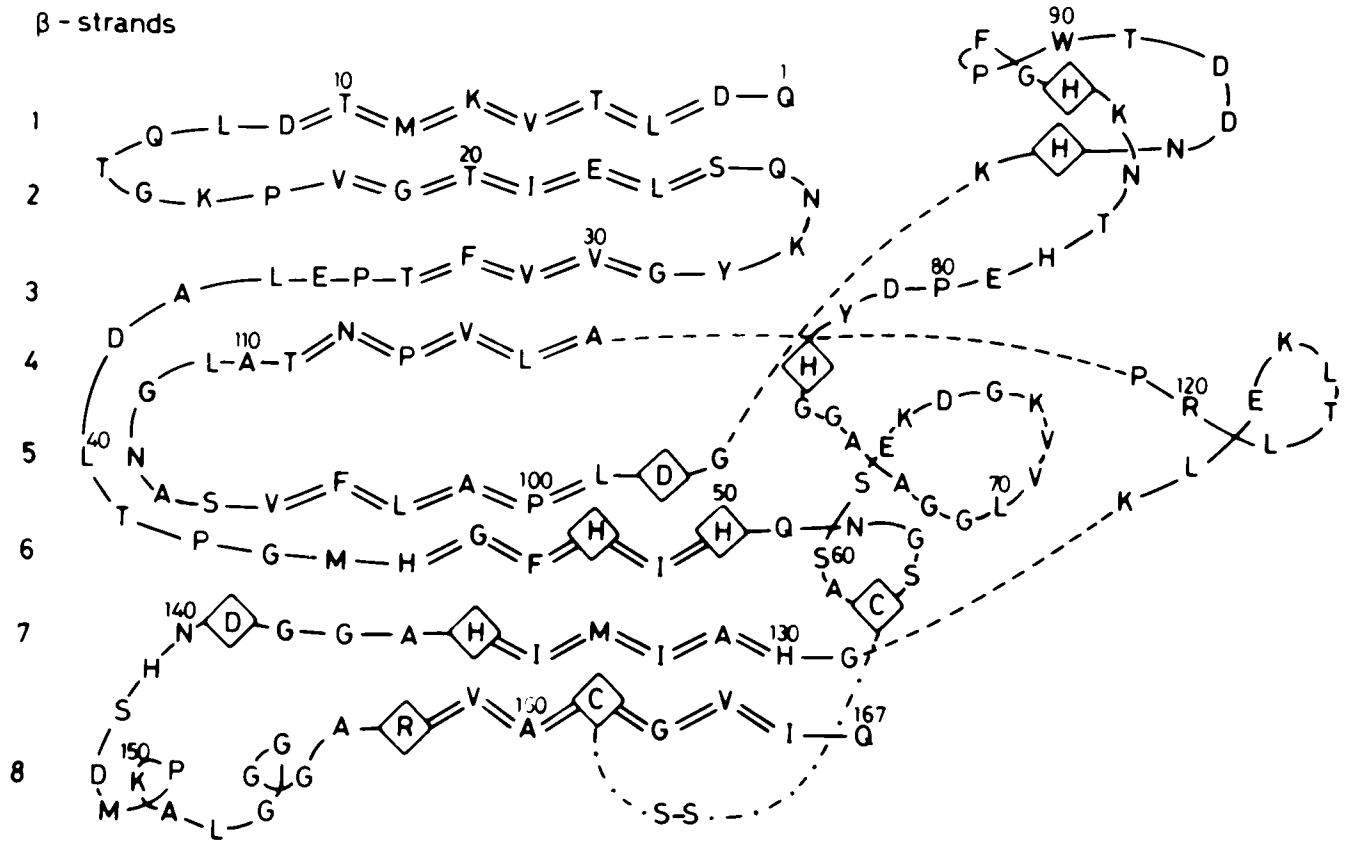


FIG. 4. Schematic topological diagram of bacterial (*P. leiognathi*) copper/zinc superoxide dismutase based on model building, by analogy with the bovine enzyme (27). β-Strands are drawn as zigzag lines with residues in the "valley" of the zigzag pointing into the interior of the β-barrel. Essential residues are boxed. Numbering is as in Fig. 1.

gent evolutionary relationship between the eukaryotic enzymes and the bacterial enzyme is not supported by the phylogenetic distribution of the enzymes. The copper/zinc superoxide dismutases have not been found in primitive eukaryotic species or in other bacteria [with one as yet unexplained exception (9)], including closely related marine bacterial species. The only simple explanation that remains is therefore gene transfer between the host ponyfish and the symbiotic bacterium *P. leiognathi*. This, however, leaves unexplained why the degree of identity between the bacterial enzyme sequence and the eukaryotic enzyme sequences is low. The presence of iron superoxide dismutase in the bacterial species implies that the introduction of the copper/zinc enzyme was not essential for the survival of the bacterium so that the introduced protein was able to accept mutation changes more readily in its new environment. The amino acid composition of the host ponyfish copper/zinc superoxide dismutase is much closer to that of the *P. leiognathi* enzyme than is the composition of the swordfish liver enzyme (12). The amino acid sequence of the ponyfish enzyme would be expected to show an even higher homology with the bacterial enzyme, as historical evidence of a gene transfer from eukaryotes to prokaryotes.

We thank Profs. I. Fridovich and R. J. P. Williams for helpful advice and criticism. M.W.P. thanks Prof. Sir D. C. Phillips and Dr. C. C. F. Blake for much encouragement. This work was supported by the Medical Research Council (J.V.B.) and the Royal Commission for the Exhibition of 1851 (M.W.P.).

1. Bannister, J. V., Bannister, W. H., Hill, H. A. O. & Thornalley, P. J. (1982) *Life Chem. Rep.* **1,** 49–53.
2. Fridovich, I. (1972) *Acc. Chem. Res.* **5,** 321–326.
3. McCord, J. M. & Fridovich, I. (1969) *J. Biol. Chem.* **244,** 6049–6055.
4. Peeters-Joris, C., Vandervoorde, A. M. & Bandheim, P. (1975) *Biochem. J.* **150,** 31–39.
5. Bannister, W. H. & Bannister, J. V. (1982) *FEBS Lett.* **142,** 42–43.
6. Salin, M. L. & Bridges, S. M. (1980) in *Chemical and Biochemical Aspects of Superoxide and Superoxide Dismutase*, eds. Bannister, J. V. & Hill, H. A. O. (Elsevier/North-Holland, New York), pp. 176–284.
7. Bridges, S. M. & Salin, M. L. (1981) *Plant Physiol.* **68,** 275–278.
8. Puget, K. & Michelson, A. M. (1974) *Biochem. Biophys. Res. Commun.* **58,** 830–838.
9. Steinman, H. M. (1982) *J. Biol. Chem.* **157,** 10283–10292.
10. Asada, K., Kanematsu, S., Okaka, S. & Hayakawa, T. (1980) in *Chemical and Biochemical Aspects of Superoxide and Superoxide Dismutase*, eds. Bannister, J. V. & Hill, H. A. O. (Elsevier/North-Holland, New York), pp. 136–153.
11. Fridovich, I. (1975) *Annu. Rev. Biochem.* **44,** 147–159.
12. Martin, J. P. & Fridovich, I. (1981) *J. Biol. Chem.* **256,** 6080–6089.
13. Haneda, Y. (1950) *Pac. Sci.* **4,** 214–227.
14. Steffens, G. J., Bannister, J. V., Bannister, W. H., Flohe, L., Gunzler, W. A., Kim, S.-M. A. & Otting, F. (1983) *Hoppe-Seyler's Z. Physiol. Chem.* **364,** 675–690.
15. Rocha, H., Bannister, W. H. & Bannister, J. V. (1984) *Eur. J. Biochem.*, in press.
16. Steinman, H. M., Naik, V. R., Abernethy, J. L. & Hill, R. L. (1974) *J. Biol. Chem.* **249,** 7325–7338.
17. Lerch, K. & Ammer, D. (1981) *J. Biol. Chem.* **256,** 11545–11551.
18. Steinman, H. M. (1980) *J. Biol. Chem.* **255,** 6758–6762.
19. Parker, M. W., Schinina, M. E., Bossa, F. & Bannister, J. V. (1984) *Inorg. Chim. Acta* **91,** 307–317.
20. McLachlan, A. D. (1971) *J. Mol. Biol.* **61,** 408–424.
21. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120,** 97–120.
22. Lim, V. I. (1974) *J. Mol. Biol.* **88,** 857–872; 873–894.
23. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13,** 211–222; 222–245.
24. Jones, T. A. (1978) *J. Appl. Crystallogr.* **11,** 268–272.
25. Tainer, J. A., Getzoff, E. D., Richardson, J. S. & Richardson, D. C. (1980) in *2SOD:Cu,Zn Superoxide Dismutase Complete Atomic Co-ordinates*, eds. Richardson, D. C. & Richardson, J. S. (Brookhaven Protein Structure Data Bank, New York).
26. Bannister, J. V., Anastasi, H. & Bannister, W. H. (1977) *Comp. Biochem. Physiol. B* **56,** 235–238.
27. Tainer, J. A., Getzoff, E. D., Beem, K. M., Richardson, J. S. & Richardson, D. C. (1982) *J. Mol. Biol.* **160,** 181–217.
28. Drummond, M. (1979) *Nature (London)* **281,** 343–347.

Fish to Bacterium Gene Transfer

Although natural gene transfer from prokaryotes to eukaryotes is well known, and includes, for instance, the insertion into plant cells of plasmid fragments by *Agrobacterium tumefaciens* in the formation of crown gall disease, examples of the reverse transfer have remained uncertain at best. One putative such case that has been debated for some years concerns the enzyme copper/zinc superoxide dismutase in *Photobacterium leiognathi*, the gene for which has been suggested to have derived from ponyfish, with which the bacterium has a close symbiotic relationship (*1*). Rigorous analysis of the primary, secondary, and tertiary structure of the *P. leiognathi* enzyme now supports very strongly the idea of natural transfer for the origin of the bacterial gene, according to Oxford University researchers Joe Bannister and Michael Parker (*2*).

Superoxide dismutase scavenges the superoxide radical $O_2^{\cdot-}$, and therefore forms an important line of defense against this highly toxic species. The enzyme is a metalloprotein: in higher eukaryotes it contains copper and zinc, in some prokaryotes and mitochondria of eukaryotes the metal is manganese, while iron is usually present in other prokaryotes. This neat pattern was ruffled somewhat by the discovery that iron superoxide dismutase is to be found in three plant families—Gingkoaceae, Nymphaeaceae, and Cruciferae—out of a total of 43 surveyed, which observation has been interpreted to imply three separate gene transfers from bacteria to the plants. The presence of the solidly eukaryotic copper/zinc enzyme in *P. leiognathi* was again inferred to be the result of gene transfer, this time from eukaryote to prokaryote. The only other known example of a bacterium containing this enzyme—*Caulobacter crescentus*—presents a mystery as, unlike *P. leiognathi*, this organism does not regularly consort with eukaryotic organisms.

Photobacterium leiognathi, a light-producing bacterium, colonizes the luminescent organ of ponyfish, thus providing an intimacy that might permit natural gene transfer. The demonstration in 1981 that the amino acid composition of the copper/zinc enzyme from host and symbiont are much closer than either is to that of the swordfish seemed to confirm this contingency (*3*). Elucidation of the bacterial enzyme's amino acid sequence in 1983, however, revealed that it had little direct identity with several eukaryotic comparisons, and the idea of natural gene transfer fell into disfavor (*4*).

In spite of this, Bannister decided to pursue the issue further and, with Parker, embarked on an in-depth structural analysis. It turns out that although direct identity is in the region of 25 percent or so, when constraints on mutation are taken into account a figure for weighted sequence similarity of 44 percent is obtained, which is much more indicative of a homologous relationship. In spite of the differences in primary structure, the *P. leiognathi* enzyme yields a secondary structure prediction closely similar to that from eukaryotes. Studies on the crystal structure of the bovine enzyme (the only one for which such information is available) shows the tertiary structure to be comprised of eight antiparallel β-strands that form a flattened cylinder plus three external loops. The question was, could the features of the bacterial enzyme be "squeezed" into a comparable three-dimensional configuration, using computerized model building? Again, although some modification was necessary, there was an impressive conservation of overall structure.

Bannister and Parker point out that although convergent evolution in terms of function in the biochemical realm is common, as in many of the oxygen-binding proteins for example, convergence on structure is not. The structural similarities they see between the bacterial and eukaryotic enzymes in this case, they conclude, strongly indicate, if not prove, homology.—ROGER LEWIN

References

1. R. Lewin, *Science* 217, 42 (1982).
2. J. V. Bannister and M. W. Parker, *Proc. Natl. Acad. Sci. U.S.A.* 82, 149 (1985).
3. J. P. Martin and I. Fridovich, *J. Biol. Chem.* 256, 6080 (1981).
4. G. J. Steffens *et al.*, *Hoppe-Seyler's Z. Physiol. Chem.* 364, 675 (1983).

to be published
1985

STRUCTURAL AND EVOLUTIONARY RELATIONSHIPS BETWEEN THE EUKARYOTIC SUPEROXIDE DISMUTASES

M.W. PARKER[1], F. BOSSA[2], D. BARRA[2], W.H. BANNISTER[3] and J.V. BANNISTER[4]
[1] Laboratory of Molecular Biophysics, University of Oxford, Oxford, England; [2] Institute of Biological Chemistry and CNR Centre for Molecular Biology, University of Rome, Rome, Italy; [3] Nuffield Department of Clinical Biochemistry, University of Oxford, Radcliffe Infirmary, Oxford, England and [4] Inorganic Chemistry Laboratory, University of Oxford, Oxford, England.

GENERAL INTRODUCTION

The enzyme superoxide dismutase (SOD) is unique in that it exists as three isoenzyme forms whose similarity is that they catalyse the same dismutation reaction: $2O_2^- + 2H^+ \rightarrow H_2O_2 + O_2$. The main important difference between the three isozymes is that they contain different metals as their functional prosthetic group (1). Eukaryotic species have two forms of the enzyme which in general contain either Cu and Zn or Mn. Two bacterial species, Photobacterium leiognathi and Caulobacter crescentus also contain the Cu/Zn enzyme (2,3). The eukaryotic SODs were found to be compartmentalised inside the cell. Intact mitochondria contain the MnSOD in the matrix space and the Cu/Zn enzyme previously thought to be exclusively localised in the cytosol was found in the intermembrane space. In rat liver the Cu/ZnSOD was also found to be lysosomal with the amount varying with the nutritional state of the animal (4). Also an examination of the subcellular distribution of the MnSOD in human liver revealed that the enzyme is not exclusively localised in the mitochondria (5). Both the Cu/Zn and the MnSOD are synthesised by the nuclear genome (6) and it is possible that the MnSOD is present in the cytosol as an active proenzyme before it is transported into the mitochondrion. Autor (7) observed that the MnSOD is synthesised as a precursor protein with an extra peptide of about 2000 daltons/subunit. However, no evidence has as yet been presented as to whether the precursor protein is an active enzyme. It is also unclear as to why mitochondria selected Mn rather than Fe as the prosthetic group metal. Prokaryotic species contain either an FeSOD or the MnSOD or both. Mitochondria are known to concentrate very efficiently both Fe and Mn (8) and if these organelles evolved from free living species then either form of the enzyme could be present. However, the MnSOD was found to be very resistant to hydroxyl radical damage in contrast to the FeSOD (9). The production of hydroxyl radicals by mitochondria via a metal catalysed Haber-Weiss reaction is a possibility since the organelle produces both O_2^- and H_2O_2. The presence of MnSOD in mitochondria is therefore an adaptation against this possibility.

INTERRELATIONSHIPS BETWEEN THE Cu/Zn SUPEROXIDE DISMUTASES

The complete primary structure has been determined for eight SODs (10-17) (Fig.1). Gaps and insertions are included to optimise sequence homologies. Only human, horse, bovine and pig SOD have an acetylated N-terminal residue. The blocked N-terminus appears to be restricted to mammals. Lower species have an exposed N-terminus. A free NH_2 terminal residue has also recently been demonstrated in the bullfrog indicating that reptiles probably also have an exposed N-terminus (18). The alignments presented show that the metal binding

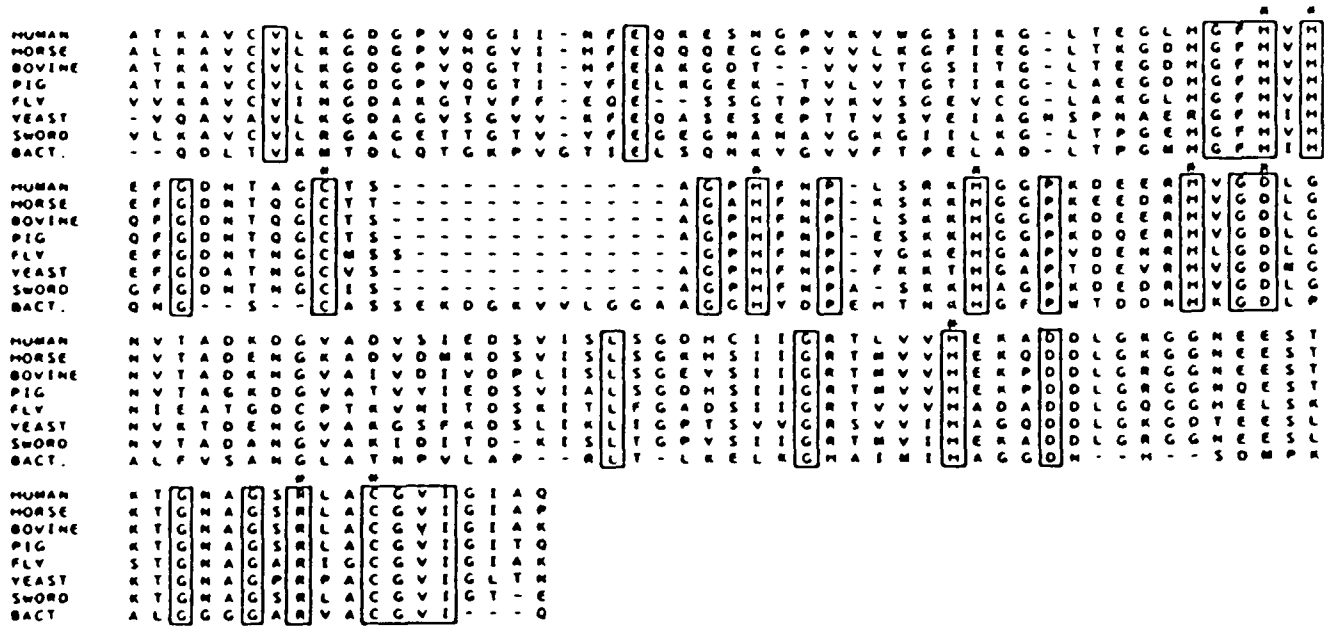


Fig. 1. Amino Acid Sequences of Cu/Zn Superoxide Dismutases.

sites are conserved in all the enzymes. These are His48, 50, 77 and 135 as ligands for the Cu and His77, 86, 95 and Asp88 as ligands for the Zn. Also conserved is Arg158 which is considered essential for facilitating the flow of O_2^- to the active site Cu (19) and the intrachain disulphide bridge formed by Cys59 and 161. Only 18 other residues are conserved in all the eight sequences and of these 8 are Gly residues which are connected with the three dimensional tight packing of the molecule. A high degree of homology is observed between the eukaryotic Cu/ZnSODs. The lowest homology is 50% between yeast and fruitfly and the highest is 84% between pig and bovine enzymes. The bacterial enzymes exhibits a very low degree of homology with the other enzymes (Table 1). The Cu/ZnSOD in P. leiognathi was proposed to have arisen in this species as a result of gene transfer from the ponyfish Leiognathus on which the bacterium is symbiotic (20). The homology with the swordfish enzyme is however only 30%. When mutations are taken into account long stretches of closely related sequences are obtained and the weighted similarity between the two enzymes rises to 44% (21). This corresponds to 78 identical or closely related residues out of 136 residues excluding deletions and insertions.

The predicted secondary structure for all the eight Cu/ZnSODs indicates considerable identity with the secondary structure determined for the bovine enzyme

Table 1. Sequence Homologies of Cu/Zn Superoxide Dismutase

	Human	Horse	Bovine	Pig	Fly	Yeast	Swordfish	Bacterial
Human								
Horse	78							
Bovine	82	78						
Pig	82	75	84					
Fly	58	55	55	55				
Yeast	55	56	53	52	50			
Swordfish	66	66	71	67	54	53		
Bacterial	25	26	28	28	24	26	30	

Values given as percentage of sequence homologies

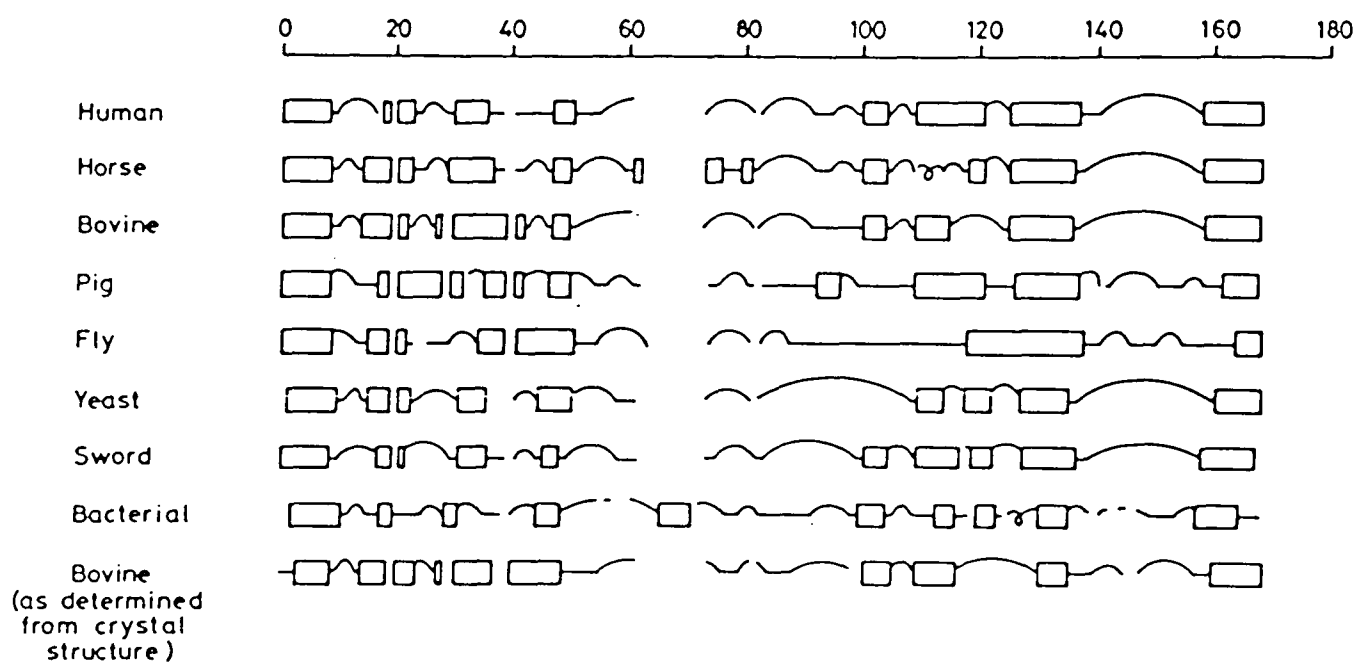


Fig. 2. Predicted Secondary Structure for Cu/Zn Superoxide Dismutases.

from the crystal structure (Fig.2). The bacterial enzyme despite having a low homology also has a similar secondary structure. The twelve residue insertion, residues 62-73, was also predicted to form a β-strand. However when an attempt was made to fit the bacterial sequence into the known bovine tertiary structure the strand could not be built in the model so as to interact appropriately with nearby strands. The insertion was built into the structure as a loop (21). The bacterial Cu/ZnSOD sequence could be fitted into the bovine structure without any serious disruptions to the fold. This evidence strongly indicates that gene transfer from the eukaryotic ponyfish to the prokaryotic P. leiognathi has taken place. The evidence for gene transfer is reinforced when predicted DNA sequences for the swordfish and bacterial SODs are compared using the known DNA sequence for the human enzyme (22) as the base sequence for the comparison. The identities obtained when deletions and insertions are excluded are given in Table 2.

Table 2 Identities Obtained from Comparison of Predicted DNA Sequences

	Human/Swordfish	Human/Bacteria	Swordfish/Bacteria
Nos. of residues	150	137	136
Nos. of identities	102 (68%)	42 (30%)	49 (36%)
Nos. of base change	26	44	47
Total nos. of residues (identities ≠ one base change)	128 (85.3%)	86 (62.8%)	96 (70.6%)

When the swordfish and bacterial enzymes are compared the number of identities are found to be 49 or 36%. However, when one base mutations are considered the number of identities rise to 96 or 70.6%.

INTERRELATIONSHIPS BETWEEN THE Mn SUPEROXIDE DISMUTASES

Compared with the Cu/Zn SODs very little work has been carried out on the MnSODs. The amino acid sequence has been determined for only four species (23-26) (Fig.3). The crystal structure of the enzyme from Thermus thermophilus

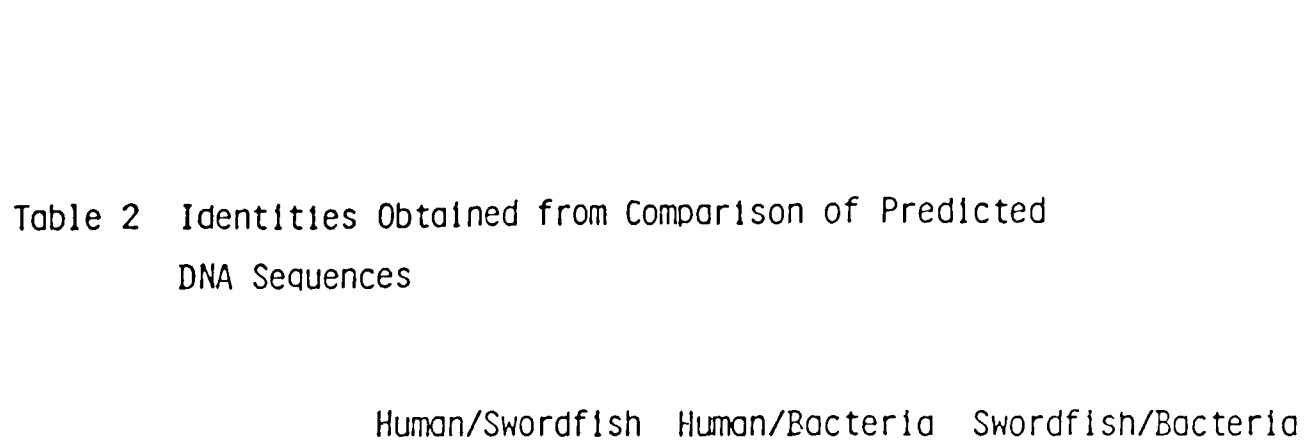


Fig. 3. Primary Structure of Mn Superoxide Dismutases

at 2.4Å has been determined (27) and the ligands to the metal are His, His, Asp and His which when combined with the sequence of the MnSOD from B. stearothermophilus corresponds to positions 26, 81, 163 and 167. In the aligned sequences presented in Fig.3 these correspond to His26, His81, Asp175 and His179. Histidine 81 in E. coli appears in position 80 but this appears to be an error in the sequence. The ligands to the Mn are also present in the Fe SOD (28) and there are strong indications that the Mn and Fe proteins are structural homologs (27). The predicted secondary structures of the MnSODs also

show a great resemblance to the FeSOD structure (1). The four MnSODs exhibit varying degrees of sequence homolysis (Table 3) which are not as high as those between the vertebrate Cu/ZnSODs.

Table 3 Sequence Homologies of Mn Superoxide Dismutase

	B. stearothermophilus	*E. coli*	Yeast	Human liver
B. stearothermophilus				
E. coli	59			
Yeast	39	37		
Human liver	48	41	43	

INTERRELATIONSHIPS BETWEEN THE Cu/Zn AND Mn SUPEROXIDE DISMUTASES

The synthesis of Cu/Zn and MnSOD is carried out by different genes. In humans, the gene for the Cu/ZnSOD is located on chromosome 21 (29) whilst the gene for the MnSOD is located on chromosome 6 (30). This indicates that the two proteins will have a different primary structure and a comparison of the amino acid sequence of human Cu/Zn and MnSODs shows this to be the case (Fig.4).

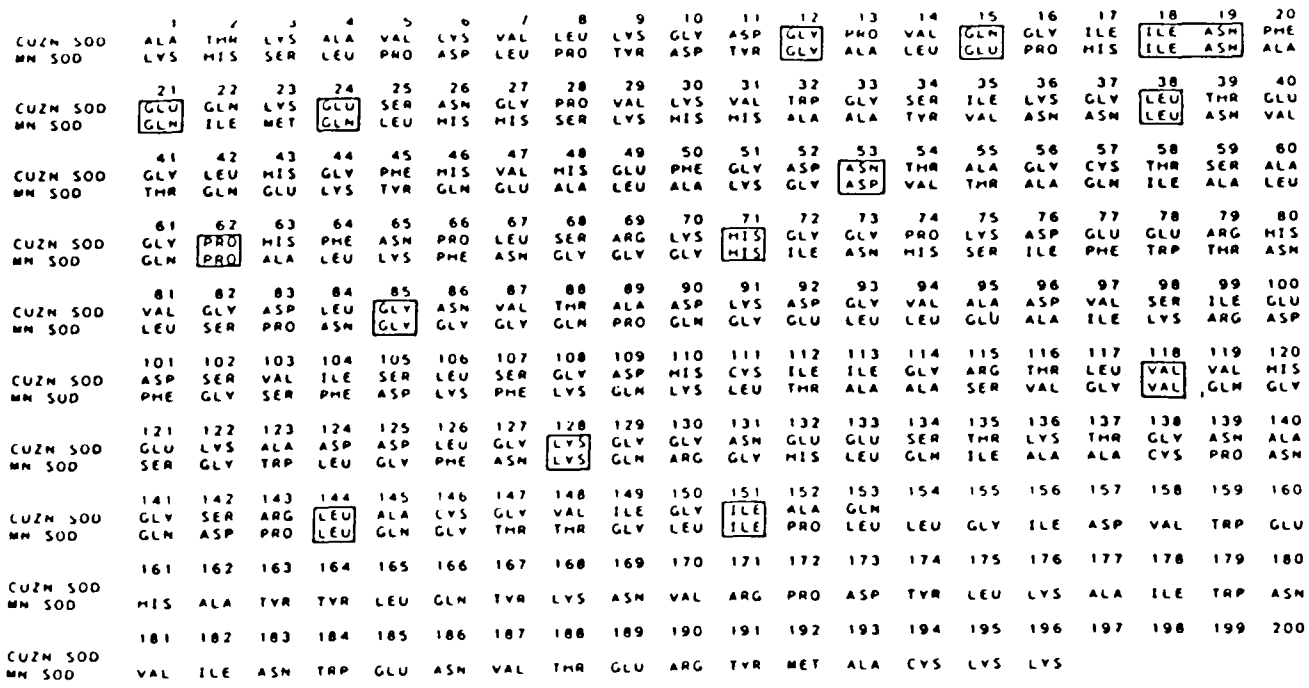


Fig. 4. Comparison of the Primary Structure of Human Cu/Zn and MnSODs.

When the two isozymes are aligned from the N-terminal and there are only 15 identical residues, however when a computer alignment is carried out, 43 residues are found to align (Fig.5) and 10 of these are Gly residues. A comparison of the ligands to the Cu, Zn and Mn is given in Table 4.

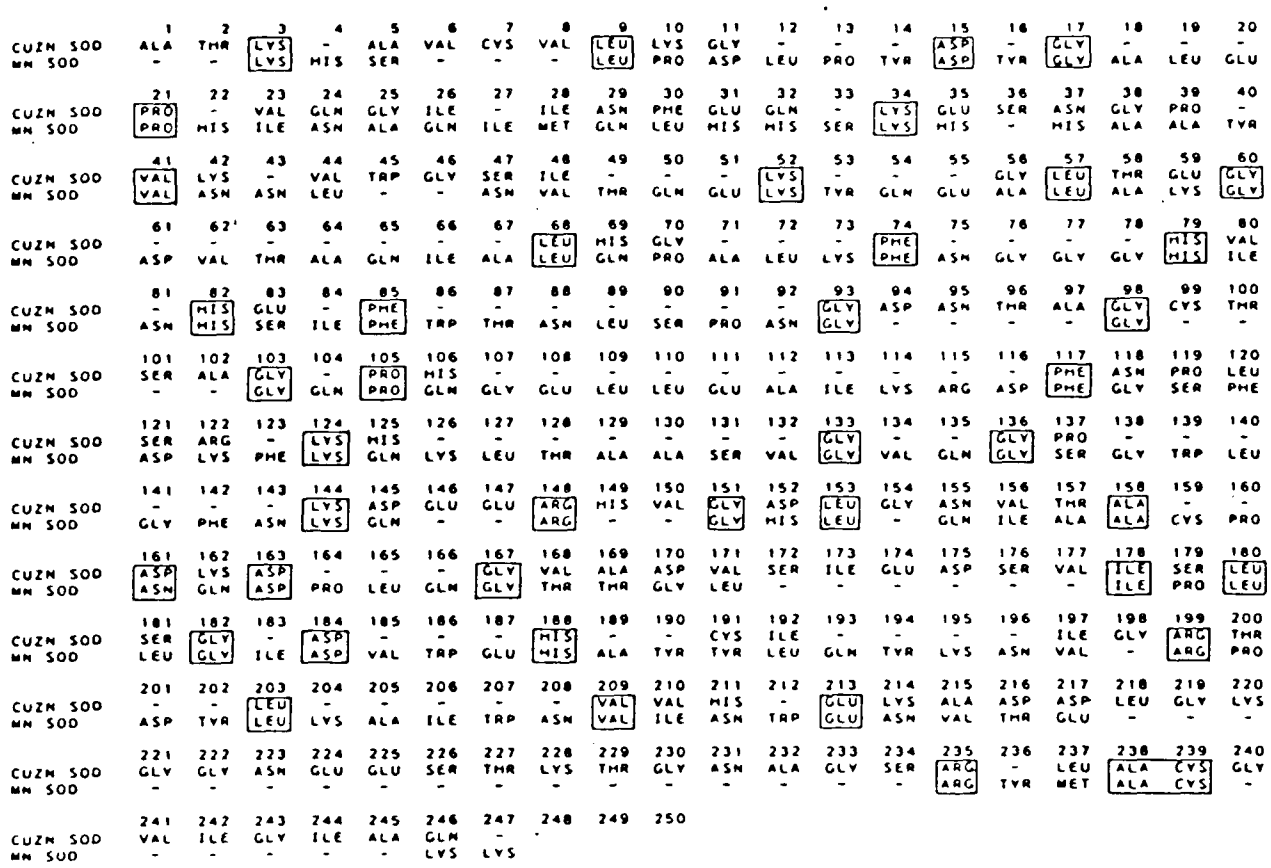


Fig. 5. Computer Alignment of the Primary Structure of Human Cu/Zn and MnSODs.

Table 4 Comparison of the Ligands to the Cu, Zn and Mn

Copper	Zinc	Manganese
His 46 → 79 (His)	His 63 → 106 (His)	His 26 → 31 (Glu)
His 48 → 82 (His)	His 71 → 125 (Gln)	His 74 → 82 (His)
His 63 → 106 (His)	His 80 → 149 (—)	Asp 157 → 184 (Asp)
His 120 → 210 (Asn)	Asp 83 → 152 (His)	His 161 → 188 (His)

The corresponding amino acid is given in brackets

Histidine 82 is a ligand to both the Cu (His48 in the human Cu/ZnSOD) and the Mn (His72 in the human MnSOD). However, most interesting is the fact that the four histidine ligands in the Cu/ZnSOD align with three histidines and one aspartate in the MnSOD with correct sequence order. Three histidines and one aspartate

eukaryotes has been a puzzle. This posed the question whether introns evolved at the time of the prokaryote/eukaryote split or at some much earlier time. It would appear that the Cu/ZnSOD evolved at the time of the prokaryote/eukaryote split on the basis of its known phylogenetic distribution. The gene structure of human Cu/ZnSOD is composed of five exons and four introns (22). Each exon codes for a compact structural unit, three code for a ββ structure. A possible proposition might be that the protein evolved from gene shuffling and fusion of five different exons. Such a mechanism would account for the rapid evolution of the Cu/ZnSOD gene that would be required if it was developed late in evolutionary time. There is no evidence from a search of distant sequence relationships that the protein evolved from some other single protein.

REFERENCES

1. Parker MW, Schinina ME, Bossa F, Bannister JV (1984) Inorg Chim Acta 91:307-317

2. Puget K, Michelson AM (1974) Biochem Biophys Res Commun 58:830-838

3. Steinman HM (1982) J Biol Chem 257:8945-8952

4. Geller BL, Winge DR (1982) J Biol Chem 257:8945-8952

5. McCord JM, Boyle JA, Day EA, Rizzolo LJ, Salin ML (1977) In: Michelson AM, McCord JM, Fridovich I (eds) Superoxide and Superoxide Dismutase. Academic Press, New York, pp 129-138

6. Weisiger RA, Fridovich I (1973) J Biol Chem 248:3582-3592

7. Autor AP (1982) J Biol Chem 257:2713-2718

8. Williams RJP (1982) FEBS Lett 140:3-10

9. Bannister WH, Bannister JV, Searle AJF (1983) In: Cohen G, Greenwald R (eds) Oxy-Radicals and their Scavenging Systems: Molecular Aspects. Elsevier Biomedical, New York, pp 387-392

10. Barra D, Martini F, Bannister JV, Schinina ME, Rotilio G, Bannister WH, Bossa F (1980) FEBS Lett 121:215-218

11. Lerch K, Ammer D (1981) J Biol Chem 256:11545-11551

12. Steinman HM, Naik VR, Abernethy JL, Hill RL (1974) J Biol Chem 249:7376-7386

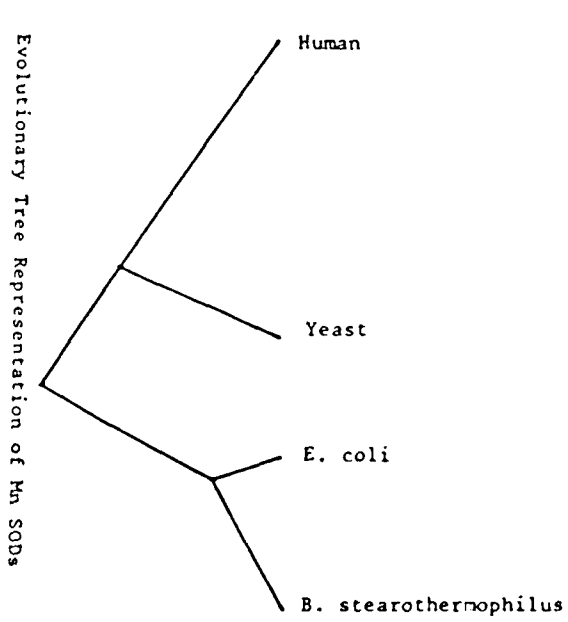
13. Schinina ME, Barra D, Simmaco M, Bassa F, Rotilio G (1985) FEBS Lett (In press)

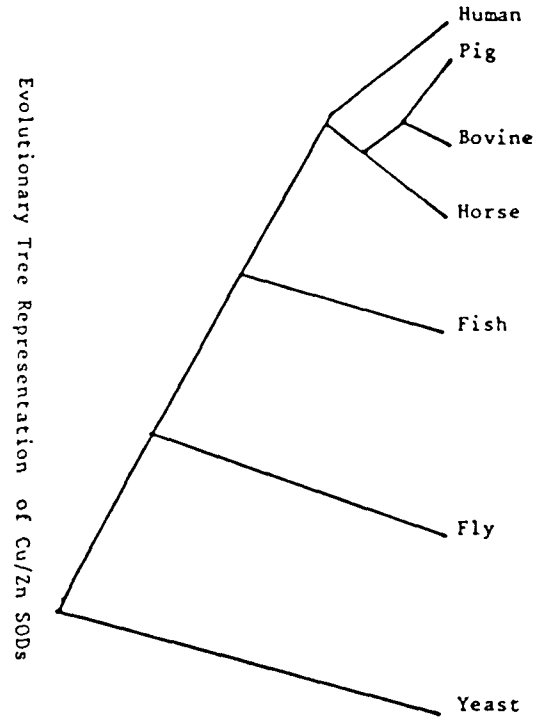
14. Lee JM, Friedman DJ, Ayala FJ (1985) Proc Natl Acad Sci USA 82:824-828

are the ligands to the Mn (27) and recently the Cu has been found to displace the Mn from the MnSOD of B. stearothermophilus (31). The computer aligned homology of 43 residues out of 247 represents 17% homology. However when the predicted DNA sequence for the MnSOD is aligned with the known DNA sequence of the Cu/ZnSOD it appears that 46 residues have undergone a one base change. When this is taken into consideration the number of identical residues rises to 89 and the homology to 36%.

EVOLUTIONARY RELATIONSHIPS BETWEEN THE Cu/Zn and MnSODs

The topologies of the phylogenetic relationships between the SODs are shown in Fig.6. The topologies were calculated by the "present-day" ancestor approach (32). No attempt was made to assign branch length to the trees because of the inaccuracies involved in determining aligned nucleotide sequences from amino acid sequences. Analysis of the data produced topologies which were in agreement with the cytochrome c evolutionary tree and trees derived from classical taxonomy and paleontology. However the rate of evolutionary changes in terms of amino acid and nucleic acid substitution varied greatly between the organisms as recently reported (14).





There has been considered interest in recent years concerning the presence of both coding (exons) and non coding (introns) regions in genes. Gilbert (33) put forward the hypothesis that exons code for functional protein units. This was extended by Blake (34) who pointed out that exons might code for structural units as well. The observation of split gene structures led to the hypothesis that new proteins could evolve from ancestral proteins by gene shuffling (33,34). However the observation of the split gene structures solely in

15. Steinman HM (1980) J Biol Chem 255:6758-6765

16. Rocha H, Bannister WH, Bannister JV (1984) Eur J Biochem 145:477-484

17. Steffens GJ, Bannister JV, Bannister WH, Flohe L, Gunzler WA, Kim SA, Otting F (1983) Z Physiol Chem 364:675-690

18. Abe Y, Okazaki T, Shukuya R, Furuta H (1984) Comp Biochem Physiol 7718:125-130

19. Maeinowski D, Fridovich I (1979) Biochemistry 18:237-242

20. Martin JP, Fridovich I (1981) J Biol Chem 256:6020-6089

21. Bannister JV, Parker MW (1985) Proc Natl Acad Sci USA 82:149-152

22. Levanon D, Lieman-Hurwitz J, Dafni N, Wigderson M, Sherman L, Bernstein Y, Laver-Rudich Z, Danciger E, Stein O, Groner Y (1985) The EMBO Journal 4:77-84

23. Brock CJ, Walker JE (1980) Biochemistry 19:2873-2882

24. Steinman HM (1980) J Biol Chem 255:6758-6765

25. Ditlow C, Johansen JT, Martin BM, Ivendsen IB (1982) Carlsberg Res Commun 47:81-91

26. Barra D, Schinina ME, Simmaco M, Bannister JV, Bannister WH, Rotilio G, Bossa F (1984) J Biol Chem 259:12595-12601

27. Staelings WC, Pattridge KA, Strong RK, Ludwig ML (1984) J Bio Chem 259:10695-10699

28. Barra D, Schinina ME, Bossa F, Bannister JV (1985) FEBS Lett 179:329-331

29. Tan YH, Tischfield J, Ruddle FH (1973) J Exp Med 137:317-330

30. Creagan R, Tischfield J, Riciutti F, Ruddle FH (1973) Humangenetik 20:203-209

31. Bannister JV, Desideri A, Rotilio G (1985) FEBS Lett (In press)

32. Blanken RL, Klotz LC, Hinnebusch AG (1982) J Mol Evol 19:9-19

33. Gilbert W (1978) Nature 271:501

34. Blake CCF (1978) Nature 273:267

References

Abernathy, J.L., Steinman, H.M. and Hill, R.L. (1974) J. Biol. Chem. 249, 7339-7347

Allison, S.A. and McCammon, J.A. (1985) J. Phys. Chem. 89, 1072-1074

Ames, B.N., Cathcart, R., Schwiers, E. and Hochstein, P. (1981) Proc. Natl. Acad. Sci. USA 78, 6858-6862

Anastasi, A., Bannister, J.V. and Bannister, W.H. (1976) Int. J. Biochem. 7, 541-546

Anbar, M. and Neta, P. (1967) Int. J. Appl. Radiat. Isotopes 18, 493-523

Archibald, F.S. and Fridovich, I. (1981) J. Bacteriol. 145, 442-451

Argese, E., Rigo, A., Viglino, P., Orsega, E., Marmocchi, F., Cocco, D. and Rotilio, G. (1984) Biochim. Biophys. Acta 787, 205-207

Argos, P., Rossmann, M.G., Grau, U.M., Zuber, H., Frank, G. and Tratschin, J.D. (1979) Biochem. 18, 5698-5703

Arndt, U.W. (1968) Acta Cryst. B24, 1355-1357

Arndt, U.W. (1984) in 'Methods and Applications in Crystallographic Computing', (Hall, S.R. and Ashida, T., eds.), Clarendon Press, Oxford, pp 1-18

Arndt, U.W., Champness, J.N., Phizackerley, R.P. and Wonacott, A.J (1973) J. Appl. Cryst. 6, 457-463

Arndt, U.W. and Wonacott, A.J. (1977) 'The Rotation Method in Crystallography', North-Holland, Amsterdam

Arnone, A., Bier, C.J., Cotton, F.A., Day, V.W., Hazen, E.E., Richardson, D.C., Richardson, J.S. and Yonath, A. (1971) J. Biol. Chem. 246, 2302-2316

Asada, K., Kanematsu,S., Okada, S. and Hayakawa, T. (1980) in 'Chemical and Biochemical Aspects of Superovide and Superoxide Dismutase', (Bannister, J.V. and Hill, H.A.O., eds.) Elsevier/North Holland, New York, pp 136-153

Asada, K., Kanematsu, S., Takahashi, M. and Kona, Y. (1976) in 'Iron and Copper Proteins', (Yasunobu, K.T., Mower, H.F. and Hayaishi, O., eds.) Plenum Publishing Corporation, New York, pp 551-562

Asada, K., Yoshikawa, K., Takahashi, M., Maeda, Y. and Enmanji, K. (1975) J. Biol. Chem. 250, 2801-2807

Autor, A.P. (1982) J. Biol. Chem. 257, 2713-2718

Baboir, B.M. (1982) Can. J. Physiol. Pharmacol. 60, 1353-1358

Baehner, R.L., Nathan, D.G. and Karnovsky, M.L. (1970) J. Clin. Invest. 49, 865-870

Bang, S.S., Baumann, L., Woolkalis, M.J. and Baumann, P. (1981) Arch. Microbiol. 130, 111-120

Bang, S.S., Woolkalis, M.J. and Baumann, P. (1978) Curr. Microbiol. 1, 371-376

Banner, D.W., Evans, P.R., Marsh, D.J. and Phillips, D.C. (1977) J. Appl. Cryst. 10, 45-51

Bannister, W.H., Bannister, J.V. and Searle, A.J.F. (1983) in 'Oxy-Radicals and their Scavenging Systems: Molecular Aspects', (Cohen, G. and Greenwald, R., eds.) Elsevier Biomedical, New York, pp 387-392

Bannister, J.V. and Parker, M.W. (1985) Proc. Natl. Acad. Sci. USA 82, 149-152

Bannister, J.V. and Rotilio, G. (1984) in 'The Biology and Chemistry of Active Oxygen', (Bannister, J.V. and Bannister, W.H., eds.), Elsevier/North Holland, New York, pp 146-190.

Bannister, W.H. (1984) in 'The Biology and Chemistry of Active Oxygen', (Bannister, J.V. and Bannister, W.H., eds.) Elseveir/North-Holland, New York, pp 208-238.

Barra, D., Martini, F., Schinina, M.E., Rotilio, G., Bannister, W.H. and Bossa, F. (1980) FEBS Lett. 120, 53-56

Barra, D., Schinina, M.E., Bossa, F. and Bannister, J.V. (1985) FEBS Lett. 179, 329-331

Barra, D., Schinina, M.E., Simmaco, M., Bannister, J.V., Bannister, W.H., Rotilio, G. and Bossa, F. (1984) J. Biol. Chem. 259, 12595-12601

Barrette, W.C., Sawyer, D.T., Fee, J.A. and Asada, K. (1983) Biochem. 22,

624-627

Bartunik, H-D (1978) Acta Cryst. A34, 747-750

Beauchamp, C.O. and Fridovich, I. (1971) Anal. Biochem. 44, 276-287

Beem, K.M., Rich, W.E. and Rajagopalan, K.V. (1974) J. Biol. Chem. 249, 7298-7305

Beem, K.M., Richardson, J.S. and Richardson, D.C. (1976) J. Mol. Biol. 105, 327-332

Berovic, J., Tillman, T., Cudd, A. and Fridovich, I. (1983) Arch. Biophys. Biochem. 221, 329-332

Bhat, T.N. and Blow, D.M. (1982) Acta Cryst. A38, 21-29

Bilderback, D.H., Moffat, K. and Szebenyi, D.M.E. (1984) Nucl. Instr. and Meth. 222, 245-251

Blackburn, N.J., Hasnain, S.S., Binsted, N., Diakun, G.P., Garner, C.D. and Knowles, P.F. (1984) Biochem J. 219, 985-990

Blake, C.C.F., Pulford, W.C.A. and Artymiuk, P.J. (1983) J. Mol. Biol. 167, 693-723

Blake, C.C.F. and Swan, I.D.A. (1971) J. Mol. Biol. 61, 217-224

Blow, D.M. (1958) Proc. Roy. Soc. Lond. Ser. A247, 302-336

Blow, D.M. and Crick, F.H.C. (1959) Acta Cryst. 12, 794-802

Blow, D.M. and Matthews, B.W. (1973) Acta Cryst. A29, 56-62

Blow, D.M. and Rossmann, M.G. (1961) Acta Cryst. 14, 1195-1202

(also correction (1962) 15, 1060)

Bragg, W.L. (1913) Proc. Camb. Phil. Soc. 17, 43-57

Blundell, T.L. and Johnson, L.N. (1976) 'Protein Crystallography', Academic Press, London

Bricogne, G. (1974) Acta Cryst. A30, 395-405

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

Bricogne, G. (1982) in 'Computational Crystallography', (Sayre, D., ed.) Clarendon Press, Oxford, pp 223-230

Bricogne, G. (1984) Acta Cryst. A40, 410-445

Bridgen, J., Harris, J.I. and Kolb, E. (1976) J. Mol. Biol. 105, 333-335

Bridges, S.M. and Salin, M.L. (1981) Plant Physiol. 68, 275-278

Britton, L., Malinowsksi, D.P. and Fridovich, I. (1978) J. Bacteriol. 134, 224-236

Brock, C.J. and Harris, J.I. (1977) Biochem. Soc. Trans. 5, 1537-1539

Brock, C.J., Harris, J.I. and Sato, S. (1976) J. Mol. Biol., 107, 175-178

Brock, C.J. and Walker, J.E. (1980) Biochem. 19, 2873-2882

Brown, O.R. and Yein, F. (1978) Biochem. Biophys. Res. Commun. 85, 1219-1224

Buerger, M.J. (1942) 'X-ray Crystallography', John Wiley, New York

Buerger, M.J. (1964) 'The Precession Camera', John Wiley, New York

Busing, W.R., Martin, K.O. and Levy, H.A. (1962) 'ORFLS, A Fortran Crystallographic Least-Squares Program', ORNL-TM-305, Oak Ridge National Laboratory, Oak Ridge, Tennessee

Calvin, M. (1969) 'Chemical Evolution', Oxford University Press, London.

Cascarano, G., Giacovazzo, C., Peerdeman, A.F. and Kroon, J. (1982) Acta Cryst. A38, 710-717

Chou, P.Y. and Fasman, G.D. (1974) Biochem. 13, 211-222; 222-245

Cocco, D., Rossi, L., Barra, D., Bossa, F. and Rotilio, G. (1982) FEBS Lett. 150, 303-306

Cohen, G. and Cedarbaum, A.I. (1979) Science 204, 66-68

Cork, C., Fehr, D., Hamlin, R., Vernon, W., Xuong, N.H. and Perez-Mendez, V. (1974) J. Appl. Cryst. 7, 319-323

Cork, C., Hamlin, R., Vernon, W. and Xuong, N.H. (1975) Acta Cryst. A31, 702-703

Crick, F.H.C. and Magdoff, B.S. (1956) Acta Cryst. 9, 901-908

Cromer, D.T. and Libermann, D.A. (1970) J. Chem. Phys. 53, 1891-1898

Cromer, D.T. and Libermann, D.A. (1981) Acta Cryst. A37, 267-268

Crowther, R.A. (1972) in 'The Molecular Replacement Method. A Collection of Papers on the Use of Non-crystallographic Symmetry', (Rossmann, M.G., ed.),

Gordon & Breach, New York, pp 173-178

Cudd, A. and Fridovich, I. (1982) J. Biol. Chem. 257, 11443-11447

Cutfield, J.F., Cutfield, S.M., Dodson, E.J., Dodson, G.G. and Sabesan, M. (1974) J. Mol. Biol. 87, 23-30

Debye, P. (1914) Annl. Phys. 43, 49-95

Diamond, R. (1969) Acta Cryst. A25, 43-45

Dickerson, R.E., Kendrew, J.C. and Strandberg, B.E. (1961) Acta Cryst. 14, 1188-1195

Dijkstra, B.W., van Nes , G.J.H., Kalk, K.H., Brandenburg, N.P., Hol, W.G.J. and Drenth, J. (1982) Acta Cryst B38, 793-799

Dijkstra, B.W. Renetseder, R., Kalk, K.H., Hol, W.G.J. and Drenth, J. (1983) J. Mol. Biol. 168, 163-179

Ditlow, C., Johansen, J.T., Martin, B.M. and Ivendsen, I.B. (1982) Carlsberg Res. Commun. 47, 81-91

Dodson, E.J. (1976) in 'Crystallographic Computing Techniques' (Ahmed, F.R., ed.), Munksgaard, Copenhagen, pp 204-211

Dodson, E.J., Evans, P.R. and French, S. (1975) in 'Anomalous Scattering', (Ramaseshan, S. and Abramsons, S.C., eds.) Munksgaard, Copenhagen, pp 423-435

Doolittle, W.F. (1980) Trends Biochem. Sci. 5, 146-149

Egami, F. (1975) J. Biochem. 77, 1165-1169

El-Deiry, N., Downey, K.M. and So, A.G. (1984) Proc. Natl. Acad. Sci. USA 81, 7378-7382

Ellman, G.L. (1959) Arch. Bioch. Biophys., 82, 70-77

Ewald, P.P. (1921) Z. Kristollagr. Miner. 56, 148-150

Fee, J.A. (1973) J. Biol. Chem. 248, 4229-4234

Fee, J.A. (1982) Trends Biochem. Sci. 7, 84-86

Fee, J.A., McClune, G.J., Lees, A.C., Zidovetzki, R. and Recht, I. (1981a) Isr. J. Chem. 21, 54-58

Fee, J.A., McClune, G.J., O'Neill, P.O. and Fielden, E.M. (1981b) Biochem.

Biophys. Res. Commun. 100, 377-384

Fee, J.A., Shapiro, E.R. and Moss, T.H. (1976) J. Biol. Chem. 251, 6157-6159

Fielden, E.M., Roberts, P.B., Bray, R.C., Lowe, D.J., Nauter, G.N., Rotilio, R. and Calabrese, L. (1974) Biochem. J. 139, 49-60

Fielden, E.M. and Rotilio, G. (1984) in 'The Copper Proteins', (Lontie, R., ed.) CRC Press, Inc. Florida U.S.A., vol. 2, pp 27-62

Florence, T.M. (1983) Chemistry in Australia 50, 166-174

Forman, H.J. and Fridovich, I. (1973) J. Biol. Chem. 248, 2645-2649

Forman, H.L. and Forman, H.J. (1973) Biochem. J. 137, 32-44

Fox, G.C. and Holmes, K.C. (1966) Acta Cryst. 20, 886-891

Freitas, D.M. and Valentine, J.S. (1984) Biochem. 23, 2079-2082

French, S. and Wilson, K. (1978) Acta Cryst. A34, 517-525

Fridovich, I. (1974) Life Sci. 14, 819-826

Fridovich, I. (1975) Ann. Rev. Biochem. 44, 147-159

Fridovich, I. (1979) Ciba Foundation Symposium No. 65, Excerpta Medica, New York

Gaber, B.P., Miskowski, V. and Spiro, T.G. (1974) J. Amer. Chem. Soc. 96, 6868-6876

Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97-120

Gebicki, J.M. and Bielski, B.H.J. (1981) J. Amer. Chem. Soc. 103, 7020-7072

Geller, B.L. and Winge, D.R. (1982) J. Biol. Chem. 257, 8945-8952

Getzoff, E.D., Tainer, J.A., Weiner, P.K., Kollman, P.A., Richardson, J.S. and Richardson, D.C. (1983) Nature 306, 287-290

Gilmore, C.J. (1982) in 'Computational Crystallography' (Sayre, D. ed.) pp 126-140, Clarendon Press, Oxford

Goebal, K-M, Herbert, V. and Stuzik, U. (1983) in 'Oxy-radicals and their Scavenger Systems', (Greenwald, R.A. and Cohen, G., eds.) Elsevier/North Holland, Amsterdam, vol. 2: Cellular and Medical Aspects, pp255-263

Greenhough, T.J. and Helliwell, J.R. (1982) J. Appl. Cryst. 15, 338-351

Greenhough, T.J. and Helliwell, J.R. (1983) Prog. Biophys. Mol. Biol. 41, 67-123

Greenstock, C.L. and Ruddock, G.W. (1978) Photochem. Photobiol. 28, 877-882

Greenwood, C. and Hill, H.A.O. (1982) Chem. in Brit. 18, 194-196

Gregory, E.M. and Dapper, C.H. (1983) Arch. Biochem. Biophys. 220, 293-300

Gregory, E.M. and Fridovich, I. (1973a) J. Bacteriol. 114, 543-548

Gregory, E.M. and Fridovich, I. (1973b) J. Bacteriol. 114, 1193-1197

Gregory, E.M. and Fridovich, I. (1974) J. Bacteriol. 117, 166-169

Gregory, E.M., Yost, F.J. and Fridovich, I. (1973) J. Bacteriol. 115, 987-991

Halliwell, B. (1981) in 'Age Pigments', (Sohal, R.S. ed.) Elsevier/North Holland, Amsterdam, pp 1-62

Halliwell, B. (1982a) Trends Biochem. Sci. 7, 270-272

Halliwell, B. (1982b) in 'The Copper Proteins', (Lontie, R., ed.) CRC Press, Inc. Florida U.S.A., vol. 2, pp 63-102

Halliwell, B. and Gutteridge, J.M.C. (1984) Biochem J. 219, 1-14

Haneda, Y. (1950) Pac. Sci. 4, 214-227

Harada, Y., Lifchitz, A., Berthou, J. and Jolles, P. (1981) Acta Cryst. A37, 398-406

Harris, J.I., Auffret, A.D., Northrop, F.D. and Walker, J.E. (1980) Eur. J. Biochem. 106, 297-303

Harrison, S.C. (1984) Nature 309, 408

Hassan, H.M. and Fridovich, I. (1977a) J. Bacteriol. 130, 805-811

Hassan, H.M. and Fridovich, I. (1977b) J. Bacteriol. 132, 505-510

Hassan, H.M. and Fridovich, I. (1977c) J. Biol. Chem. 252, 7667-7672

Hassan, H.M. and Fridovich, I. (1979) Rev. Infect. Dis. 1, 357-367

Haugaard, N. (1968) Physiol. Rev. 48, 311-373

Heikkila, R.E. and Cappat, F. (1976) Anal. Biochem. 75, 356-362

Helliwell, J.R. (1985) in 'Information Quarterly for Protein

Crystallography' CCP4 newsletter, July, no. 15

Hendrickson, W.A. (1971) Acta Cryst. B27, 1472-1473

Hendrickson, W.A. (1979) Acta Cryst. A35, 158-163

Hendrickson, W.A., Klippenstein, G.L. and Ward, K.B. (1975) Proc. Natl. Acad. Sci. USA 72, 2160-2164

Hendrickson, W.A. and Lattman, E.A. (1970) Acta Cryst. B26, 136-143

Hendrickson, W.A. and Teeter, M.M. (1981) Nature 290, 107-113

Hering, K., Kim, S-M.A., Michelson, A.M., Otting, F., Puget, K., Steffens, G.J. and Flohe, L. (1985) Biol. Chem. Hoppe-Seyler 366, 435-445

Hestenes, M.R. and Stiefel, E. (1952) J. Res. Natl. Bur. Stand. 49, 409-436

Hewitt, J. and Morris, J.G. (1975) FEBS Lett. 55, 282-285

Hol, W.G.J. (1985) at the 'Molecular Replacement Conference', SERC, Daresbury Laboratory, 15-16 February

Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3824-3828

Huber, R. and Kopfmann, G. (1969) Acta Cryst. A25, 143-152

Huber, W. and Menander-Huber, K.B. (1980) Clinics in Rheumatic Diseases 6, 465-470

Hurt, E.C., Pesold-Hurt, B., Suda, K., Oppliger, W. and Schatz, G. (1985) EMBO J. 4, 2061-2068

Itakura, M. and Holmes, E.W. (1979) J. Biol. Chem. 254, 333-338

Jabusch, J.R., Farb, D.L., Kerschensteiner, D.A. and Deutsch, H.F. (1980) Biochem. 19, 2310-2316

Johansen, J.T., Overballe-Petersen, C., Martin, B., Haseman, V. and Svendsen,I. (1979) Carlsberg Res. Commun. 44, 201-217

Johnston, R.B. and Baehner, R.L. (1970) Blood 35, 350-355

Kahn, R., Fourme, R., Bosshard, R., Chiadmi, M., Risler, J.L., Dideberg, O. and Wery, J.P. (1985) FEBS Lett. 179, 133-142

Kanematsu, S. and Asada, K. (1978) Arch. Biochem. Biophys. 185, 473-482

Kanematsu, S. and Asada, K. (1979) Arch. Biochem. Biophys. 195, 535-545

Karle, J. (1980) Int. J. Quantum. Chem. 7, 357-367

Karle, J. (1984) Acta Cryst. A40, 4-11

Karle, J. (1985) Acta Cryst. A41, 387-394

Karlsson, R. (1985) at the 'Molecular Replacement Conference', SERC, Daresbury Laboratory, 15-16 February

Karlsson, R., Weaver, L.H., Smit, J.D.G., Pauptit, R.A. and Jansonius, J.W. (1984) in abstracts to the '8th International Biophysics Congress', 29 July - 4 August, Bristol, p 133

Keele, B.B., Giovagnoli, C. and Rotilio, G. (1975) Physiol. Chem. Physics 7, 1-6

Keele, B.B., McCord, J.M. and Fridovich, I. (1970) J. Biol. Chem. 245, 6176-6181

Kemmerer, A.R., Elvehjem, C.A. and Hart, E.B. (1931) J. Biol. Chem. 92, 623-630

Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D.R., Phillips, D.C. and Shore, V.C. (1960) Nature 185, 422-427

Kirby, T.W., Lancaster, J.R. and Fridovich, I. (1981) Arch. Biochem. Biophys. 210, 140-148

Klug, D., Rabani, J. and Fridovich, I. (1972) J. Biol. Chem. 247, 4839-4842

Klug-Roth, D., Fridovich, I. and Rabani, J. (1973) J. Amer. Chem. Soc. 95, 2782-2790

Kono, Y. and Fridovich, I. (1983) J. Biol. Chem. 258, 6015-6019

Kopfmann, G. and Huber, R. (1968) Acta Cryst. A24, 348-351

Koppenol, W.H. (1981) in 'Oxygen and Oxy-Radicals in Chemistry and Biology', (Powers, E.L. and Rodgers, M.A.J., eds.), Academic Press, New York, pp 671-676

Kraut, J., Sieker, L.C., High, D.F. and Freer, S.T. (1962) Proc. Natl. Acad. Sci. USA 48, 1417-1424

Kusunose, E., Kusunose, M., Ichihara, K. and Izumi, S. (1980) J. Gen. Appl. Microbiol. 26, 369-372

Kwiatowski, J., Safianowska, A. and Kaniuga, Z. (1985) Eur. J. Biochem.

146, 459–466

Lattman, E.E., Nockolds, C.E., Kretsinger, R.H. and Love, W.E. (1971) J. Mol. Biol. 60, 271–277

Lavelle, F., McAdam, M.E., Fielden, E.M. and Roberts, P.B. (1977) Biochem J. 161, 3–11

Lawerence, G.D. and Sawyer, D.T. (1979) Biochem. 18, 3045–3050

Lee, Y.M., Friedman, D.J. and Ayala, F.J. (1985) Proc. Natl. Acad. Sci. USA 82, 824–828

Lerch, K. and Ammer, D. (1981) J. Biol. Chem. 256, 11545–11551

LeTrant, N., Meshnick, S.R., Kitchner, K., Eaton, J.W. and Cerami, A. (1983) J. Biol. Chem. 258, 125–130

Lim, V.I. (1974) J. Mol. Biol. 88, 857–872; 873–894

Lindmark, D.G. and Miller, M. (1974) J. Biol. Chem. 249, 4634–4637

Lipson, H. and Cochran, W. (1966) 'The Determination of Crystal Structures', (Bragg, L. ed.) G. Bell & Sons Ltd., London

Lumsden, J. and Hall, D.O. (1974) Biochem. Biophys. Res. Commun. 58, 35–41

McAdam, M.E., Fox, R.A., Lavelle, F. and Fielden, E.M. (1977a) Biochem. J. 165, 71–80

McAdam, M.E., Lavelle, F., Fox, R.A. and Fielden, E.M. (1977b) Biochem. J. 165, 81–87

McClune, G.J. and Fee, J.A. (1978) Biophys. J. 24, 65–69

McCord, J.M. (1974) Science 174, 529–531

McCord, J.M. and Fridovich, I. (1969a) J. Biol. Chem. 244, 6049–6055

McCord, J.M. and Fridovich, I. (1969b) J. Biol. Chem. 244, 6056–6063

McCord, J.M., Keele, B.B. and Fridovich, I. (1971) Proc. Natl. Acad. Sci. 68, 1024–1027

McCord, J.M. and Roy, R.S. (1982) Can. J. Physiol. Pharmacol. 60, 224–226

McLachlan, A.D. (1971) J. Mol. Biol. 61, 409–424

Mann, T. and Keilin, D. (1939) Proc. Roy. Soc. Lond. Ser. B126, 303–315

Marklund, S.L. (1982) Proc. Natl. Acad. Sci. USA 79, 7634–7638

Marklund, S.L. (1984) Biochem. J. 220, 269-272

Marklund, S.L. (1985) FEBS Lett. 184, 237-239

Marklund, S.L. and Marklund, G. (1974) Eur. J. Biochem. 47, 469-474

Marres, C.A.M., van Loon, A.P.G.M., Oudshoorn, P., van Steeg, H., Grivell, L.A. and Slater, E.C. (1985) Eur. J. Biochem. 147, 153-161

Martin, J.P. and Fridovich, I. (1981) J. Biol. Chem. 256, 6080-6089

Matthews, B.W. (1966a) Acta Cryst. 20, 82-86

Matthews, B.W. (1966b) Acta Cryst. 20, 230-239

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

Matthews, B.W. and Czerwinski, E.W. (1975) Acta Cryst. A31, 480-487

Matthews, K.A., Alden, J.T., Bolin, J.T., Freer, S.T., Hamlin, R., Xuong, N., Kraut, J., Poe, M., Williams, M. and Hoogsteen, K.V. (1977) Science 197, 452-455

Meier, B., Barra, D., Bossa, F., Calabrese, L. and Rotilio, G. (1982) J. Biol. Chem. 257, 13977-13980

Michelson, A.M. (1979) in 'Metalloproteins: Structure, Molecular Function and Clinical Aspects', (Weser, U., ed.) Georg Thieme Verlag Stuttgart, New York, pp 88-116.

Miquel, J., Economos, A.C., Fleming, J. and Johnson, J.E. (1980) Exp. Gerontol. 15, 575-581

Misra, H.P. (1984) J. Biol. Chem. 259, 12678-12684

Misra, H.P. and Fridovich, I. (1978) Arch. Biochem. Biophys. 189, 317-322

Moffat, K., Szebenyi, D. and Biderback, D. (1984) Science 223, 1423-1425

Mota de Freitas, D. and Valentine, J.S. (1984) Biochem. 23, 2079-2082

Navia, M.A., Segal, D.M., Padlan, E.A., Davies, D.R., Rao, N., Rudikoff, S. and Potter, M. (1979) Proc. Natl. Acad. Sci. USA 76, 4071-4074

Nixon, P.E. and North, A.C.T. (1976) Acta Cryst. A32, 320-325

Norrod, P. and Morse, S.A. (1979) Biochem. Biophys. Res. Commun. 90, 1287-1294

North, A.C.T. (1965) Acta Cryst. 18, 212-216

North, A.C.T., Phillips, D.C. and Mathews, F.S. (1968) Acta Cryst. A24, 351-359

Nyborg, J. and Wonacott, A.J. (1977) in 'The Rotation Method in Crystallography' (Arndt, U.W. and Wonacott, A.J., eds), North Holland, Amsterdam, p 145-147

Nyborg, J., Wonacott, A.J., Thierry, J-C and Champness, J.N. (1975) 'IDXREF Program', unpublished notes

O'Brien, R.W. and Morris, J.G. (1971) J. Gen. Microbiol. 68, 307-318

Oberley, L.W. and Buettner, G.R. (1979) Cancer Res. 39, 1141-1149

Oberley, L.W. (1983) Biochem. Biophys. Res. Commun. 113, 883-888

Oberley, L.W. and Oberley, T.D. (1984) J. Theor. Biol. 106, 403-422

Ohmori, H., Komoriya, K., Azuma, A., Kurozumi, S. and Hashimoto, Y. (1979) Biochem. Pharmacol. 28, 333-334

Okada, S., Kanematsu, S. and Asada, K. (1979) FEBS Lett. 103, 106-110

Orent, E.R. and McCallum, E.V. (1931) J. Biol. Chem. 92, 651-678

Ose, D.E. and Fridovich, I. (1976) J. Biol. Chem. 251, 1217-1218

Ose, D.E. and Fridovich, I. (1979) Arch. Biochem. Biophys. 194, 360-364

Osman, R. and Basch, H. (1984) J. Amer. Chem. Soc. 106, 5710-5714

Osterberg, R. (1974) Nature 249, 382-383

Pantoliano, M.W., Valentine, J.S., Burger, A.R. and Lippard, S.J. (1982) J. Inorg. Biochem. 17, 325-341

Parker, M.W., Bossa, F., Barra, D., Bannister, W.H. and Bannister, J.V. (1985) in 'Abstracts of the Fourth International Conference on Superoxide and Superoxide Dismutase', Rome, 1st - 6th September, and to be published

Parker, M.W., Schinina, M.E., Bossa, F. and Bannister, J.V. (1984) Inorg. Chim. Acta 91, 307-317

Peerdeman, A.F. and Bijvoet, J.M. (1956) Acta Cryst. 9, 1012-1015

Perutz, M.F. (1956) Acta Cryst. 9, 867-873

Peto, R., Doll, R., Buckley, J.D. and Sporn, M.B. (1981) Nature 290, 201-208

Phillips, D.C. (1964) J. Sci. Instru. 41, 123-129

Phillips, D.C. (1966) Adv. Res. Diff. Meth. 2, 75-140

Pick, M., Rabani, J., Yost, F. and Fridovich, I. (1974) J. Amer. Chem. Soc. 96, 7329-7333

Powers, T.B., Slykhouse, T.O., Fee, J.A. and Ludwig, M.L. (1978) J. Mol. Biol. 123, 689-690

Privalle, C.T. and Gregory, E.M. (1979) J. Bacteriol. 138, 139-145

Puget, K., Lavelle, F. and Michelson, A.M. (1977) in 'Superoxide and Superoxide Dismutases', (Michelson, A.M., McCord, J.M. and Fridovich, I. eds.) Academic Press, New York, pp 139-150.

Puget, K. and Michelson, A.M. (1974) Biochem. Biophys. Res. Commun. 58, 830-834

Raff, R.A. and Mahler, H.R. (1972) Science 177, 575-582

Ramachandran, G.N. and Raman, S. (1956) Curr. Sci. 25, 348-351

Rao, S.N., Jih, J-H and Hartsuck, J.A. (1980) Acta Cryst. A36, 878-884

Raven, P.H. (1970) Science 169, 641-646

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

Richardson, J.S., Thomas, K.A., Rubin, B.H. and Richardson, D.C. (1975) Proc. Natl. Acad. Sci. USA 72, 1349-1353

Rigo, A., Viglino, P. and Rotilio, G. (1975) Anal. Biochem. 68, 1-8

Rigo, A., Stevanato, R., Finazzi-Agro, A. and Rotilio, G. (1977) FEBS Lett. 80, 130-132

Rigo, A., Viglino, P. and Rotilio, G. (1975) Biochem. Biophys. Res. Commun. 63, 1013-1018

Rigo, A., Viglino, P., Argese, E., Terenzi, M. and Rotilio, G. (1979) J. Biol. Chem. 254, 1759-1760

Rigo, A. and Rotilio, G. (1980) in 'Chemical and Biochemical Aspects of Superoxide and Superoxide Dismutase', (Bannister, J.V. and Hill, H.A.O., eds.) Elsevier/North Holland, pp 56-64

Ringe, D., Petsko, G.A., Yamakura, F., Suzuki, K. and Ohmori, D. (1983) Proc. Natl. Acad. Sci. USA 80, 3879-3883

Rocha, H., Bannister, W.H. and Bannister, J.V. (1984) Eur. J. Biochem. 145, 477-484

Rogers, D., Stanley, E. and Wilson, A.J.C. (1955) Acta Cryst. 8, 383-393

Rossmann, M.G. (1961) Acta Cryst. 14, 383-388

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

Rossmann, M.G. and Blow, D.M. (1961) Acta Cryst. 14, 641-647

Rossmann, M.G. and Blow, D.M. (1962) Acta Cryst. 15, 24-31

Rotilio, G., Calabrese, L. and Coleman, J.E. (1973) J. Biol. Chem. 248, 3855-3859

Rotilio, G. and Fielden, E.M. (1984) in 'The Biology and Chemistry of Active Oxygen', (Bannister, J.V. and Bannister, W.H., eds.) Elsevier/North Holland, New York, pp 65-78

Sagan, L. (1967) J. Theor. Biol. 14, 225-274

Salem, S.I., Chang, C.N., Lee, P.L. and Severson, V. (1978) J. Phys. Chem. 11, 4085-4090

Salin, M.L. and Bridges, S.M. (1980) in 'Chemical and Biochemical Aspects of Superoxide and Superoxide Dismutases', (Bannister, J.V. and Hill, H.A.O., eds.) Elsevier/North Holland, New York, pp 176-284

Salin, M.L. and Wilson, W.W. (1981) Molec. Cell. Biochem. 36, 157-161

Sato, S. and Harris, J.I. (1977) Eur. J. Biochem. 73, 373-381

Sato, S. and Nakazawa, K. (1978) J. Biochem. (Tokyo) 83, 1165-1171

Scheuhammer, A.M. and Cherian, M.G. (1985) Biochim. Biophys. Acta 840, 163-169

Schinina, M.E., Barra, D., Simmaco, M., Bossa, F. and Rotilio, G. (1985) FEBS Lett. 186, 267-270

Scrutton, M.C., Utter, M.F. and Mildvan, A.S. (1966) J. Biol. Chem. 241, 3480-3487

Searcy, K.B. and Searcy, D.G. (1981) Biophys. Biochem. Acta 670, 39-46

Sherman, L., Dafni, J., Lieman-Hurwitz, J. and Groner, Y. (1983) Proc. Natl. Acad. Sci. USA 80, 5465-5469

Shimizu, N., Kobayashi, K. and Hayashi, K. (1984) J. Biol. Chem. 259, 4414-4418

Sim, G.A. (1959) Acta Cryst. 12, 813-815

Simon, R.H., Scoggin, C.H. and Patterson, (1981) J. Biol. Chem. 256, 7181-7186

Singh, A.K. and Ramaseshan, S. (1966) Acta Cryst. 21, 279-280

Smit, J.D.G., Pulver-Sladek, J. and Jansonius, J.N. (1977) J. Mol. Biol. 112, 491-494

Smith, J.L. and Hendrickson, W.A. (1981) Acta Cryst. A37, C-11

Stallings, W.C., Pattridge, K.A., Powers, T.B., Fee, J.A. and Ludwig, M.L. (1981) J. Biol. Chem. 256, 5857-5859

Stallings, W.C., Pattridge, K.A., Strong, R.K. and Ludwig, M.L. (1984) J. Biol. Chem. 259, 10695-10699

Stallings, W.C., Powers, T.B., Pattridge, K.A., Fee, J.A. and Ludwig, M.L. (1983) Proc. Natl. Acad. Sci. USA 80, 3884-3888

Steffens, G.J., Bannister, J.V., Bannister, W.H., Flohe, L., Gunzler, W.A., Kim, S.A. and Otting, F. (1983) Hoppe-Seylers Z. Physiol. Chem. 364, 675-690

Steinman, H.M. (1978) J. Biol. Chem. 253, 8708-8720

Steinman, H.M. (1980) J. Biol. Chem. 255, 6758-6765

Steinman, H.M. (1982) J. Biol. Chem. 257, 10283-10292

Steinman, H.M. and Hill, R.L. (1973) Proc. Natl. Acad. Sci. USA 70, 3725-3829

Steinman, H.M., Naik, V.R., Abernethy, J.L. and Hill, R.L. (1974) J. Biol. Chem. 249, 7376-7386

Strahs, G. and Kraut, J. (1968) J. Mol. Biol. 35, 503-512

Stuart, D., Jinbi, D., Meizhen, L., Junming, Y., Todd, R., Jiayao, L., Jiahuai, W. and Dongcai, L. (1984) Scientia Sinica Ser B27, 572-582

Sussman, J.L., Holbrook, S.R., Church, G.M. and Kim, S-H (1977) Acta Cryst. A33, 800-804

Sussman, J.L. (1980) in 'Refinement of Protein Structures', Proceedings of the Daresbury Study Weekend, 15-16 November, (compiled by Machin, P.A., Campbell, J.W. and Elder, M.) No. DL/SCI/R16 SERC 1981

Synderman, R. and Goetzl, E.J. (1981) Science 213, 830-837

Tainer, J.A., Getzoff, E.D., Beem, K.M., Richardson, J.S. and Richardson, D.C. (1982) J. Mol. Biol. 160, 181-217

Tainer, J.A., Getzoff, E.D., Richardson, J.S. and Richardson, D.C. (1983) Nature 306, 284-286

Tally, F.P., Goldin, H.R., Jacobus, N.V. and Gorbach, S.L. (1977) Infect. Immun. 16, 20-25

Taube, H. (1965) 'Oxygen: Chemistry, Structure and Excited States', Little Brown, Boston.

Templeton, D.H., Templeton, L.K., Phillips, J.C. and Hodgson, K.O. (1980) Acta Cryst. A36, 436-442

Terech, A., Pucheault, J. and Ferradini, C. (1983) Biochem. Biophys. Res. Commun. 115, 114-120

Tickle, I.J. (1973) unpublished results

Tollin, P., Main, P. and Rossmann, M.G. (1966) Acta Cryst. 20, 404-407

Tolmasoff, J.M., Ono, T. and Cutler, R.G. (1980) Proc. Natl. Acad. Sci. USA 77, 2777-2781

Vanopdenbosch, B. and Crichton, R.R. (1977) FEBS Lett. 82, 42-46

Viglino, P.,Rigo, A., Stevenato, R., Ranieri, A., Rotilio, G. and Calabrese, L. (1979) J. Magn. Res. 34, 265-274

Viglino, P., Scarpa, M., Cocco, D. and Rigo, A. (1985) Biochem. J. 229, 87-90

Villafranca, J.J. (1976) FEBS Lett. 62, 230-232

Villafranca, J.J., Yost, F.J. and Fridovich, I. (1974) J. Biol. Chem. 249, 3532-3536

Vyas, N.K., Vyas, M.N. and Quiocho, F.A. (1983) Proc. Natl. Acad. Sci. USA 80, 1792-1796

Wang, B.C. (1985) in 'Diffraction Methods for Biological Macromolecules', (Wyckoff, H.W., Hirs, C.H.W. and Timasheff, S.W. eds.) Academic Press, New York (in press)

Wardman, P. (1978) in 'Radiation Biology and Chemistry: Research Developments', (Edwards, H.E., Navaratnam, S., Parsons, B.J. and Phillips, G.O., eds.) Elsevier/North Holland, Amsterdam, pp 189-196

Weisiger, R.A. and Fridovich, I. (1973) J. Biol. Chem. 248, 3582-3592

Williams, D.R. (1976) 'An Introduction to Bioinorganic Chemistry', C.C. Thomas, Springfield, Illinois.

Williams, R.J.P. (1982) FEBS Lett. 140, 3-10

Williams, R.J.P. (1984) in 'The Biology and Chemistry of Active Oxygen', (Bannister, J.V. and Bannister, W.H., eds.) Elsevier/North Holland, New York, pp 1-16.

Wilson, A.J.C. (1949) Acta Cryst. 2, 318-321

Wilson, A.J.C. (1950) Acta Cryst. 3, 397-399

Witting, L.A. (1980) in 'Free Radicals in Biology' vol. IV, (Pryor, W.A., ed.) Academic Press, New York, pp 295-355

Woolfson, M.M. (1984) Acta Cryst. A40, 32-34

Wonacott, A.J. (1980) 'A Suite of Programs for On-line Evaluation and Analysis of Integrated Intensities on Small Angle Rotation/Oscillation Photographs', user manual

Xuong, N.H., Freer, S.T., Hamlin, R., Nielsen, C. and Vernon, W. (1978) Acta Cryst. A34, 289-296

Xuong, N.H., Sullivan, D., Nielsen, C. and Hamlin, R. (1985) Acta Cryst. B41, 267-269

Yamakura, F.(1976) Biochim. Biophys. Acta 422, 280-294

Yamakura, F. (1984) Biochem. Biophys. Res. Commun. 122, 635-641

Yamakura, F. and Suzuki, K. (1980) J. Biochem. (Tokyo) 88, 191-196

Yamakura, F., Suzuki, K. and Mitsui, Y. (1976) J. Biol. Chem. 251, 4792-4793

Yamakura, F., Suzuki, K., Petsko, G.A. and Tsernoglou, D. in (1980) 'Chemical and Biochemical Aspects of Superoxide and Superoxide Dismutase', (Bannister, J.V. and Hill, H.A.O, eds.), North Holland, New York, pp 242-253

Yamamoto, Y., Shinkai, H., Isogai, Y, Matsuura, K. and Nishimura, M. (1984) FEBS Lett. 175, 429-438

Yost, F.J. and Fridovich, I. (1973) J. Biol. Chem. 248, 4905-4908

Youngman, R.J. (1984) Trends Biochem. Sci. 9, 280-283

Corrigenda

Symbols and Abbreviations

The following should be added:

Saly salyrganic acid

epr electron paramagnetic resonance (spectroscopy)

Chapter 2

page 38 line 19 the mean square error is in the electron density

page 43 line 15 $\underline{x}_i$ and $\underline{\Delta}$ are vectors which give the coordinates...

page 48 line 15 r_{ij} and R are in terms of fractional coordinates

page 49 line 4 the new phase selected by the Wang procedure is calculated by phase combination of the original SIR phase distribution and the phase distribution calculated from the Fourier inverted map as described on pages 39 and 112.

Chapter 4

figure 4.4 The best straight line was drawn through the data points but it was found that the intensity changes at the beginning and near the end of data collection could be modelled by lines of zero gradient as illustrated for the reflections 0 36 0 and 18 0 0.

table 4.7 For the purpose of the table, independent reflections refer to I_{mean}.

page 85 line 14 The volume swept out by a reciprocal lattice point passing through the Ewald sphere can be modelled by a sphere of radius ϵ_c.

Chapter 5

page 98 line 18 A reference to ordinate analysis is Watson et al. (1970).

References

The following references need to be inserted:

Rayment, I. (1983) Acta Cryst. A39, 102-116

Watson, H.C., Shotton, D.M., Cox, J.M. and Muirhead, H. (1970) Nature 225, 806-811