X-ray Crystallographic Studies of Human Chorionic Gonadotropin

by

Deborah Clare Harris

A Thesis Submitted for the Degree of

Doctor of Philosophy

Department of Chemistry

University of Glasgow

December 1992

Summary

Human Chorionic Gonadotropin (hCG) is a glycoprotein hormone released by the trophoblast to stimulate progesterone production by the corpus luteum during the early stages of pregnancy. It exists as a heterodimer consisting of noncovalently bound α and β-subunits. The α-subunit is common with the other members of this glycoprotein family, the pituitary hormones Luteinizing Hormone, Follicle Stimulating Hormone and Thyroid Stimulating Hormone. The α-subunit consists of 92 amino acid residues including ten cysteine residues which form five disulphide bridges. There are two N-linked oligosaccharide moieties bound to Asn 56 and Asn 82. The β-subunit, which confers the specific biological role of each hormone, consists of 145 residues in hCG. There are twelve cysteine residues, conserved throughout the proteins, all involved in disulphide linkages. The β-subunit is also heavily glycosylated with two long chain N-linked oligosaccharides at Asn 13 and Asn 30 and four O-linked oligosaccharides at Ser 121, Ser 127, Ser 132 and Ser 138. From circular dichroic work, hCG consists largely of aperiodic structure, 25-30% β-sheet and 0-12% α-helix. There is a limited knowledge of the regions involved in the subunit interface and receptor binding areas, but many questions remain unanswered.

This thesis describes my work in attempting to elucidate the three-dimensional structure of human Chorionic Gonadotropin by X-ray crystallography. HCG was partially deglycosylated by treatment with anhydrous Hydrogen Fluoride to produce the species HF-hCG. The tertiary structure was largely unaffected as HF-hCG was still capable of binding to receptor and had a similar secondary structure content as measured by Circular Dichroism (Keutmann *et al.* (1983a)). Crystals were grown of these HF-hCG species using the hanging drop method. Optimisation of the conditions led to crystals with sharp hexagonal bipyramidal morphology growing up to 0.4mm in each direction. The crystals were characterized and found to be of the space group $P6_122$ or enantiomer with cell dimensions a = b = 88.68Å and c = 177.2Å.

The multiple isomorphous replacement method was employed for phase determination.

Many potential heavy atom derivatives were screened using either visual comparison of precession photographs or by collection of a low resolution dataset on a Xentronics area detector and calculation of differences and difference Pattersons. Several derivatives were discovered, but unfortunately they shared one of the three major heavy atom binding sites.

Data were collected on a Xentronics area detector, on film at station PX7.2 at the S.R.S., Daresbury and on the FAST at station PX9.6 at the S.R.S., Daresbury. Of the various datasets the film data were of the highest resolution, with diffraction observed to 2.5Å but problems with radiation damage meant that more than one crystal was required to collect a full dataset. The best data collected were on the Xentronics area detector where a whole dataset could be collected with a single crystal. However, the maximum diffraction observed using the rotating anode source was 3.0Å resolution. The data were processed using both the *XENGEN* and *XDS* packages, the latter giving far more accurate estimates of the intensities of the higher resolution data.

The first derivative solved was $K_2Pt(CN)_4$ using a combination of difference Patterson maps and direct method techniques. S.I.R. phases calculated from this derivative were used to phase difference Fouriers for other derivatives. The anomolous data were used to determine the correct hand, and the true space group was found to be $P6_522$. MIR electron density maps were calculated and the solvent and protein regions were clear. However, it was not possible to distinguish the separate subunits. Phase refinement and phase extension methods were employed in an attempt to improve and extend the higher resolution phases. Solvent flattening, histogram matching and maximum entropy techniques all produced improved maps in which β-sheet like density was apparent.

Attempts were made to fit the sequence to the electron density, but this was made difficult both by the quality of the maps, and the extensive network of disulphide linkages. Also the unequivocal assignments of these linkages had not been made. Segments of polyalanine chain were fitted where possible and the partial structure used in phase combination in a "bootstrap" attempt to solve the structure.

Acknowledgements

I would firstly like to thank my supervisor, Neil Isaacs, for his patience and help throughout the course of this project. Thankyou to the people at S.V.I.M.R.; Frank Morgan for introducing me to hCG, Ken and Genevieve for performing the HF-treatment, and the other members of the lab who made me so welcome there.

On to Glasgow, where I met up with the Scots sense of humour in the form of Big Gez. Thanks to all the group - Andy, Ben, Marjo, Neilg, Fran, Allison, Marie, David, Gerard, Paul, Jacqui *et al.*. Special thanks go to Paul Emsley for help with the figures, Jacqui Coll for help with the references, Ken Shankland for trying to demystify some of the mathematics and Ben Luisi for useful discussions and proof reading. Also all the unfortunate souls who "volunteered" to help me collect data at Daresbury, and scan the films at York.

Thankyou to the Dodsons - Guy for both allowing us the use of the film scanner, and for having faith in my abilities and Eleanor for her patient programming help over the years.

I am grateful to Dr. Canfield and Joyce Lustbader at Columbia University for supplying me with HF-treated protein and crystals over the latter part of this project. In addition I thank Joyce for her hospitality and generosity during my times in New York. Thankyou to Dr. Charles Carter for allowing me to visit his laboratory and for his close supervision while I used the maximum entropy programs. Also to Shibin Xiang who modified the programs as necessary.

This project received financial support from the Special Program of Research, Development and Research Training in Human Reproduction, World Health Organization and from the Medical Research Council.

Finally, cheers Bill for supplying flasks of tea and bearing with me while I completed this thesis.

For Yvonne and Ian, with love and thanks.

Declaration

This thesis is less than 100,000 words in length, exclusive of tables, maps and bibliographies. The work reported in this thesis was performed entirely by myself unless otherwise cited or acknowledged. Its contents have not been submitted for any other degree.

Signed

Deborah C. Harris

23rd December 1992

Contents

List of Tables

List of Figures

Abbreviations used

h	-	human
o	-	ovine
b	-	bovine
e	-	equine
hCG	-	human chorionic gonadotropin
LH	-	luteinizing hormone
TSH	-	thyroid stimulating hormone
FSH	-	follicle stimulating hormone
c.AMP	-	cyclic adenyl monophosphate
ATP	-	adenyl triphosphate
Mabs	-	Monoclonal antibodies
hGH	-	human growth hormone
HF	-	anhydrous hydrogen fluoride
TFMS	-	trifluoromethane sulphonic acid
HF-hCG	-	hygrogen fluoride treated hCG
Sialic acid	-	N-acetylneuraminic acid
Asialo-hCG	-	neuraminidase treated hCG
MPD	-	methyl 2,4-pentanediol
PEG	-	poly ethylene glycol
MIR	-	multiple isomorphous replacement method

Chapter 1

Introduction

Human Chorionic Gonadotropin was first isolated in 1927 by Ascheim and Zondek and is one of the earliest known hormones. A great deal of research has been directed at this glycoprotein and in this introduction I will attempt to summarize work on the structural aspects of hCG and, where relevant, the other glycoproteins belonging to the same family of hormones.

1.1 Biological Role of human Chorionic Gonadotropin

The action of many hormones occurs by stimulating formation of cyclic adenyl monophosphate (c.AMP) but each has a specific action because it stimulates this formation only in cells containing specific surface receptors for that hormone. The hormone binds to its receptor, activates the enzyme Adenylate Cyclase located in the cell membrane, and causes the conversion of adenyl triphosphate (ATP) to c.AMP. This then transmits and amplifies within the cell the chemical signal delivered by the hormone.

HCG is secreted only in physiologically effective levels during pregnancy and thus its main function is related to maintenance of pregnancy. However, immunoreactive hCG-like material has been discovered to be present in low levels in the body fluids and

tissues of normal individuals, but Yoshimoto *et al.* (1977) reported that the protein was not glycosylated and suggest that all normal tissues can synthesize the peptide portion of hCG, but that this has little or no biological function and is rapidly cleared.

The hCG present during pregnancy is produced by the trophoblast and is responsible in the very early stages for the maintenance of pregnancy. In a normal human menstrual cycle, luteinizing hormone (hLH) stimulates the progesterone production of the Corpus Luteum. If there is no fertilization of the ovum the Corpus Luteum fails, progesterone levels drop and menstruation occurs. However, in a conceptual cycle, stimulation of the Corpus Luteum is maintained by the hCG secreted by the early trophoblast, and progesterone levels remain high allowing pregnancy to proceed.

The level of hCG in the maternal blood has been measured throughout pregnancy, and is found to be highest in the first trimester, but it is still found at high levels during the second and last trimesters (Braunstein *et al.*, (1976)). This suggests a further role played by hCG beyond the stimulation of the Corpus Luteum. There is strong evidence that it is responsible for the stimulation and differentiation of the fetal testis and adrenal cortex (Jaffe *et al.* (1979)), and some evidence that it has immunosuppressive properties (Kaye & Jones, (1971), Adcock *et al.*, (1973)).

1.2 Pregnancy Vaccines

HCG is a prime candidate as an antigen for a vaccine to prevent pregnancy. If the activity of hCG produced by the trophoblast is neutralized, progesterone levels would fall as in a non-conceptual cycle and menstruation would occur apparently normally (Figure 1.1). The World Health Organization has set up a task force on vaccines for fertility regulation and since 1980 efforts have been directed at an attempt to develop a vaccine against the action of hCG. Obviously many problems have to be surmounted. The vaccine would have to be totally reversible, effective over a given period of time and very specifically active against hCG without disturbing the function

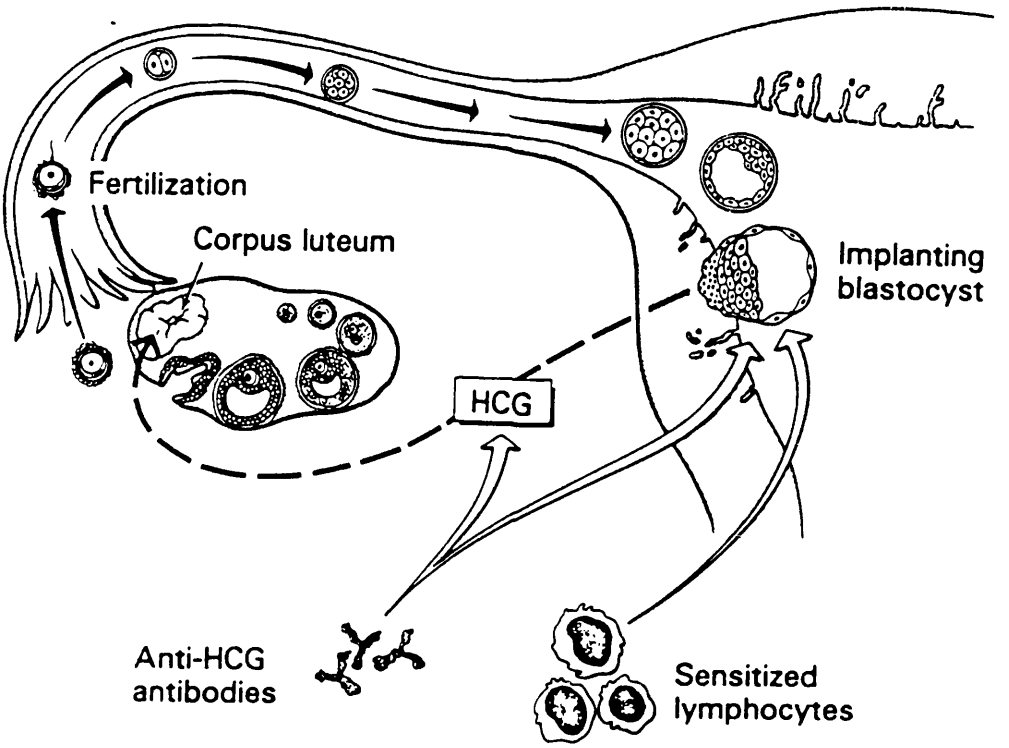


Figure 1.1: Possible modes of action of an anti-hCG vaccine.

Antibodies capable of neutralizing hCG may inhibit its luteotrophic action on the Corpus Luteum. Cytotoxic anti-hCG antibodies or sensitized lymphocytes may disrupt the early peri-implantation blastocyst.

of other hormones in the system. Also there is always the danger that other presently unknown functions of hCG could be affected. The problem of total specificity for hCG is compounded by the high structural similarity between it, the pituitary hormones LH, follicle stimulating hormone (FSH) and thyroid stimulating hormone (TSH). Much work has been conducted on the structural relationships between these hormones in an attempt to find a particular, distinct motif on hCG to use as an antigen. HCG, LH, FSH and TSH all exist as heterodimers consisting of two subunits (designated alpha and beta) bound by non-covalent forces. Knowledge of the three dimensional structures of these hormones would allow the determination of distinct antigenic regions.

1.3 Primary Structure of the Alpha Subunit

The alpha subunit of hCG was sequenced (Morgan *et al.*, (1975)) and was found to be identical to the 92 residue sequence for hLH-α (Sairam *et al.*, (1972)). The sequences of hTSH-α and hFSH-α are also identical to hLH-α. The sequences of these hormones from other sources have since been determined, and are found to be identical within a species. Figure 1.2 compares the sequence of human (h), bovine (b) and ovine (o) α-subunits. As can be seen, there is a high degree of homology between the species, with a 100% conservation of the ten cysteine residues. Positions of glycosylation are also indicated. Unless otherwise stated, throughout this thesis the numbering of residues in the α-subunit is consistent with that shown in Figure 1.2 i.e. relative to the bovine α-subunit. The genes coding for the α-subunit from a number of species have been isolated and sequenced. The human and bovine α-subunits exist as single genes consisting of four exons and three introns. As one exon codes for the presequence, the α-subunit can be discussed in terms of three domains, defined by the exons, as N-terminus to residue 10, residue 11 to residue 71 and 72 to C-terminus (Fiddes & Talmadge, (1984)). The largest homology between the bovine and human sequences is found in the second and third domains, with 75% and 80% identity respectively.

	1									10								
h	A	P	D	V	Q	(			)	D	C	P	E	C	T	L	Q	E
b,o	F	-	-	G	E	F	T	M	Q	G	-	-	-	-	K	-	K	-
e	F	-	-	G	E	-	-	-	-	B	-	-	-	-	K	-	R	-

		20										30						
h	N	P	F	F	S	Q	P	G	A	P	I	L	Q	C	M	G	C	C
b,o	-	K	Y	-	-	K	-	D	-	-	-	Y	-	-	-	-	-	-
e	-	K	Y	-	F	K	-	-	V	-	-	L	-	-	K	-	-	R

				40									50					
h	F	S	R	A	Y	P	T	P	L	R	S	K	K	T	M	L	V	Q
b,o	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	P
e	-	-	-	-	-	-	-	-	A	-	-	R	-	-	-	-	-	P

						60										70		
h	K	N*	V	T	S	E	S	T	C	C	V	A	K	S	Y	N	R	V
b,o	A	-	-	I	-	-	-	A	-	-	-	-	-	-	A	F	T	K
e	-	-	-	I	-	Z	-	-	-	-	-	-	-	A	F	I	-	-

								80										90
h	T	V	M	G	G	F	K	V	E	N*	H	T	A	C	H	C	S	T
b,o	-	-	-	-	N	V	R	-	-	-	-	-	E	-	-	-	-	-
e	-	-	-	-	G	I	-	L	Z	-	-	-	Z	-	Y	-	-	-

						96
h	C	Y	Y	H	K	S
b,o	-	-	-	-	-	-
e	-	-	H	-	-	I

Figure 1.2: The Amino Acid Sequences of the α-Subunits

* indicates location of carbohydrate. h - human ; b - bovine ; o - ovine ; e - equine

( ) indicates a deletion. - indicates the same residue as the bovine sequence.

1.4 Primary Structure of the Beta Subunit

The beta subunit determines the distinct functions of the hormones. Figure 1.3 compares the sequences of the human β-subunits of the four hormones. HLH contains 114 residues, while hCG contains 145 residues with a C-terminal extension rich in prolines and serines. There is a fair degree of homology between the four human β-subunits and also among different species. This is particularly high between β-hLH and the first 114 residues of β-hCG with an 85% identity. Again, the twelve cysteine positions are conserved in all the sequences for all the species so far determined. Positions of glycosylation are indicated.

The genes for the β-subunits of hCG, LH, FSH and TSH have been isolated and sequenced. There are seven genes which code for β-hCG (Boorstein *et al.*, (1982); Talmadge *et al.*, (1983)). At least two of these are expressed and contain three exons and two introns. The first intron is between residues -15 and -16 in the presequence and the second between residues 41 and 42. (Policastro *et al.*, (1983); Talmadge *et al.*, (1984)).

1.5 Secondary Structure

The amount of ordered structure in all the hormones is fairly low in aqueous solution. Circular dichroism experiments on the individual subunits predict a low α-helical content, in the region of 0-12%, a β-sheet content of 25-30% with the majority of the structure thought to be aperiodic in nature (Garnier *et al.*, (1975); Guidice & Pierce, (1978)), the best fit being 8% α-helix, 35% β-structure and 57% aperiodic. These findings are supported by prediction analyses (Chou *et al.* (1978); Lustbader *et al.* (1992)). Biochemical experiments have shown that the aperiodic structure is well defined as many residues are unavailable to enzymatic or chemical modification. This finding might be expected because of the large number of disulphide bridges discussed

	1									10								
hCG	S	K	E	P	L	R	P	R	C	R	P	I	N*	A	T	L	A	V
hLH	-	R	-	-	-	-	-	W	-	H	-	-	-	-	I	-	-	-
hTSH								F	-	I	-	T	E	Y	-	M	H	I
hFSH							N	S	-	E	L	T	-*	I	-	I	-	I

		20										30						
hCG	E	K	E	G	C	P	V	C	I	T	V	N*	T	T	I	C	A	G
hLH	-	-	-	-	-	-	-	-	-	-	-	-*	-	-	-	-	-	-
hTSH	-	R	R	E	-	A	Y	-	L	-	I	-*	-	-	-	-	-	-
hFSH	-	-	-	-	-	R	F	-	L	-	I	-*	-	-	W	-	-	-

				40										50				
hCG	Y	C	P	T	M	T	R	V	L	Q	G	V	L	P	A	L	P	Q
hLH	-	-	-	-	-	M	-	-	-	-	A	-	-	-	P	-	-	-
hTSH	-	-	M	-	R	D	I	N	G	K	L	F	-	-	K	Y	AL	S
hFSH	-	-	Y	-	R	D	L	-	Y	K	N	P	A	R	P	K	E	-

						60										70		
hCG	V	V	C	N	Y	R	D	V	R	F	E	S	I	R	L	P	G	C
hLH	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-
hTSH	QD	V	-	T	-	-	-	F	I	Y	R	T	V	E	I	-	-	-
hFSH	K	T	-	T	F	K	E	L	V	-	-	T	V	-	V	-	-	-

								80										90
hCG	P	R	G	V	N		V	V	S	Y	A	V	A	L	S	C	Q	C
hLH	-	-	-	-	D		-	-	-	F	P	-	-	-	-	-	R	-
hTSH	-	L	H	-	A	P	Y	F	-	-	P	-	-	-	-	-	K	-
hFSH	A	H	H	A	D	S	L	Y	T	-	P	-	-	T	Q	-	H	-

										100								
hCG	A	L	C	R	R	S	T	T	D	C	G	G	P	L	D	H	P	L
hLH	G	P	-	-	-	-	-	S	-	-	-	-	-	K	-	-	-	-
hTSH	G	K	-	N	T	D	Y	S	-	-	I	H	E	Q	I	K	T	N
hFSH	G	K	-	D	S	D	S	-	-	-	T	V	R	G	L	G	-	S

		110										120						
hCG	T	C	D	D	P	R	F	Q	D	S	S	S	S*	K	A	P	P	P
hLH	-	-	-	H	-	Q	{L	S	G	K	-	P	L}					
hTSH	Y	-	T	K	-	Q	K	S	Y	{L	V	G	F	S	V}			
hFSH	Y	-	S	F	G	E	M	K	Q	Y	P	T	A	L	S	Y		

				130										140					
hCG	S*	L	P	S	P	S*	R	L	P	G	P	S*	D	T	P	I	L	P	Q

Figure 1.3: The Amino Acid Sequences of the β-subunits

* indicates location of carbohydrate. Residues in braces are found in the gene but have not been found in pituitary extracts.

in detail in the next section.

Results of circular dichroism experiments on the native hormone show an increase in the amount of β-sheet present relative to the individual subunits (Giudice *et al.* (1978)). This is possibly due to changes in the conformation of one or both of the subunits, but is more likely β-sheet regions of the α and β-subunits come into contact forming a larger β-sheet.

Experiments conducted in non-aqueous solvents show a change in the proportions of secondary structure, particularly the region β38-57 which appears helical and also possibly a helical region β17-21. The secondary structure of the region β38-57 has been particularly well studied. Molecular modelling of LH β38-57 showed that this looped peptide can exist with either helical or turn secondary structure even without the disulphide linkage Cys38-Cys57 (Keutmann *et al.* (1987)). Circular Dichroism experiments on this peptide revealed the presence of amphipathic α-helical structure through the N-terminal region. In aqueous buffer the helical content is 4%, but in the helicogenic solvent trifluoroethanol, the helical content is 24% (Milius *et al.*, (1990)). This has also been observed with a disulphide free peptide. An extended polypeptide containing the β 38-57 sequence extended with Arginine and Glycine residues to enhance solubility and in which the cysteines were replaced by alanines was also studied and showed the loop structure is independent of the disulphide bridge. 2D-NMR work on these peptides also support the amphipathic nature of this region. (Keutmann *et al.*, (1992)). NOESY spectroscopy confirms the helical region to be at the N-terminal end, specifically involving residues 41-48 and β-turns in the C-terminal region.

1.6 Tertiary Structure - Disulphide Bridges

The cysteine positions in the α and β-subunits are highly conserved between the different gonadotrophins from all sources and it is highly probable that the disulphide

bridges are also identical. The actual assignment of the disulphide bonds in these hormones has been complicated by the fact that the cysteine residues are present as clusters (Figure 1.4). α-hCG has four distinct domains in cysteine residues at locations 11 and 14, 32, 35 and 36, 63 and 64 and 86, 88 and 91, while β-hCG has seven of the twelve cysteine residues concentrated in three regions of the polypeptide chain at positions 23 and 26, 34 and 38, 88, 90 and 93.

The conventional methods of disulphide bond assignment based on the isolation of individual cystinyl peptides is made difficult by this as there are relatively few bonds that are subject to chemical or proteolytic cleavage within the clusters. Also because of the large degree of glycosylation the subunits of hCG are resistant to hydrolysis by proteolytic enzymes. Conflicting results for the location of disulphide bonds in α-LH and β-LH have been reported using the enzymatic hydrolysis method (Table 1.1).

Cornell and Pierce (1974) assigned the disulphide linkages in bovine α-LH using partial reduction and alkylation, a method which depends on the differences in the reactivities of the various disulphide bonds. It involves partial reduction and *S*-[C-14]-carboxymethylation of the subunit, followed by complete reduction and *S*-carboxymethylation. The derivative is then fragmented and the specific radioactivity of each *S*-[C-14]-carboxymethylcysteine measured. Two radioactively-labelled cysteines having the same or close specific radioactivity are considered to be linked in the native hormone. Mise and Bahl (1980) used this same method to determine the linkages in α-hCG and arrived at the same conclusions. The assignment of bonds between Cys 11-39 and Cys 14-36 are widely accepted, but the other results are less definite as once these bonds are disrupted the remainder are broken in a cascade making confident assignment more difficult.

Table 1.1 summarizes the results of these experiments and the assignments made by Mise and Bahl are shown in Figure 1.4, with less well established linkages being represented by dashed lines. The assignments of the linkages in the β-subunits are less controversial. Results using enzymatic hydrolysis of ovine β-LH (Tsunasawa *et al.*, (1977)) are in complete agreement with those from Mise and Bahl (1981) using the

Enzymatic Hydrolysis		Partial Reduction & Alkylation	
pLH-α^a	oLH-α^b	bLH-α^c	hCG-α^d
11-88	11-64	11-35	11-35
14-64	14-86	14-36	14-36
32-35	32-88	32-64	32-64
36-91	35-63	63-91	63-91
63-86	36-91	86-88	86-88

Table 1.1: Assignments of Disulphide Linkages in the α-subunit*

* Adapted from Ryan et al (1987). Residues numbered relative to the bovine sequence.

[a] Combarnous and Hennen (1974); [b] Chung et al (1973)

[c] Cornell and Pierce (1974); [d] Mise and Bahl (1980)

partial reduction and alkylation method. However, the assignments Cys 26-110 and Cys 34-88 are not well established. Work on the disulphide bonds in the β-subunit is summarized in Table 1.2 and presented diagramatically in Figure 1.4.

1.7 Tertiary Structure - the Alpha Subunit

References to the amino acid sequence of the α-subunit use the numbering for the bovine sequence, consistent with Figure 1.2, unless otherwise stated.

1.7.1 Modification of Tyrosines

The relative reactivity of the Tyrosines have been deduced using a variety of chemical methods, including iodination and nitration (Cheng *et al.*, (1972); Hum *et al.*, (1974); Burleigh *et al.*, (1976); Liu *et al.*, (1976); Combarnous *et al.*, (1974)). Bovine LH Tyr α-21 (Phe α-17 in hCG), Tyr α-92 and 93 react in the intact molecule, Tyr α-41

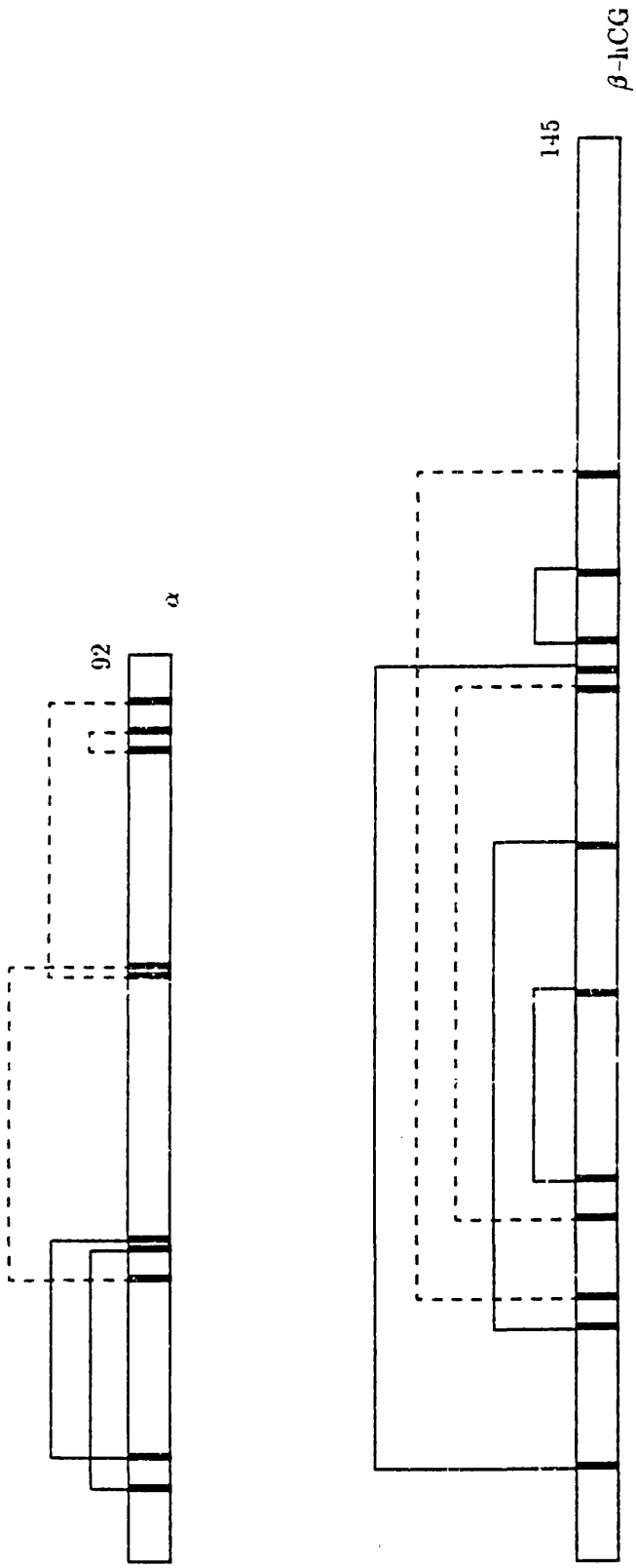


Figure 1.4: hCG Disulphide Bridges

Vertical bars represent cysteine residues. Linkages denoted by solid lines have been assigned. Those denoted by dashed lines are not well established.

Enzymatic Hydrolysis		Partial Reduction & Alkylation
oLH-β^a	oLH-β^b	hCG-β^c
9-38	9-90	9-90
23-72	23-72	23-72
26-110	26-110	26-110
34-90	34-88	34-88
57-88	38-57	38-57
93-100	93-100	93-100

Table 1.2: Assignments of Disulphide Linkages in the β-subunit*

* Adapted from Ryan et al. (1987)

[a] Chung et al (1975); [b] Tsunasawa et al. (1977)

[c] Mise and Bahl (1981)

reacts upon subunit dissociation and bovine LH Tyr α-30 (Leu α-26 in hCG) reacts only upon dissociation of the α and β subunits. When Tyr α-92 and 93 are nitrated there is a loss in receptor binding affinity suggesting that these residues are surface orientated and are located in the hormone-receptor binding region. The reactivity of Tyr α-21 in bovine LH indicates that this is also surface orientated. Tyr α-30 is protected from modification by the β-subunit and, in the primary sequence, is surrounded by hydrophobic residues which are invariant between species supporting the hypothesis that this region is at the α-β interface.

1.7.2 Modification of Lysines and Arginines

There are 10 Lys residues in the o,b α-subunit, 9 of which can be readily substituted with maleyl or citraconyl groups in the intact dimer, indicating these are all surface directed (Liu *et al.*, (1977); Sairam *et al.*, (1975)). Lys α-49, also present in hCG, only reacts in the isolated subunit inferring involvement in the α-β interface. This is

supported by the crosslinking experiments in which Lys α-49 can be covalently bound to Asp β-111 with retention of biological activity (Weare *et al.*, (1979)).

As with the lysines, there is a high degree of modification of arginine residues in the intact hormones indicating positioning at the surface. Modification of 8-10 arginines in LH (α and β) results in loss of activity relative to the native hormone suggesting there is a role played by the arginines in the action of the hormone. The relative effects of the different residues are unknown, but Arg α-39 and 46 are within an invariant region. Modification with 1,2-cyclohexanediol does not impair subunit recombination and so these residues are possibly important for hormone-receptor interaction (Sairam (1976)).

1.7.3 Modification of Methionines

Carboxymethylation of methionines in b-LH resulted in the preferential radio-labelling of Met α-33 and 8 (Val α-4 in hCG) which are perhaps surface orientated compared to Met α-51 and 75, which are more protected from alkylation. The methionine residues are not well conserved between species, but substitutions are of residues with alkyl side chains suggesting that they play a structural, size dependant role. The involvement in receptor binding of hCG Met α-33 is postulated on the basis of chemical modification experiments (Cheng, (1976))

1.8 Tertiary Structure - the Beta Subunit

Chemical modifications of the β-subunit seem to be more tolerated than in the α-subunit.

1.8.1 Modification of Tyrosines

All the tyrosines in β-hCG are protected from iodination and nitration in the native molecule, but can be modified in the free subunit suggesting that Tyr β-37, 59 and 82 are buried in the subunit interface. Tyr β-37 is part of the invariant C-A-G-Y-C sequence and Tyr β-59 and 82 are either conserved or subsituted with the homologous Phe residue suggesting the presence of an aromatic ring is the important characteristic.

1.8.2 Modification of Lysines and Arginines

The lysine residues are not well conserved between the various known sequence of the β-subunits and substitutions are not always with homologous residues. Acetylation and carbamylation of the lysines does not seriously impair subunit reassociation but does reduce the receptor binding affinity.

Modification of native oLH with 1,2-cyclohexanediol resulted in modification of 8 of the 11 Arg residues and similarly with o β-LH 7 of the 8 Arginines were affected. Thus the Arginines are postulated to be on exposed regions of the molecule and not be involved at the subunit-subunit interface.

1.9 Quaternary Structure

The unique β-subunit confers the biological specificity of the hormones, but the α subunit is also critical for receptor binding and biological activity; the individual subunits can bind to receptor, but with greatly reduced affinity relative to the intact heterodimer. The three dimensional structure of hCG has been probed using biochemical, mutagenic and immunochemical techniques. Surface directed amino acid residues have been identified, as have regions involved in the α-β subunit interface and receptor binding. These findings are discussed below.

1.9.1 Location of the alpha-beta Interface

Monoclonal antibodies raised against hCG react with topographic regions of the folded native hormone, involving residues which may be far apart in the primary sequence. They are thus useful in determining the three-dimensional surface structure of the protein. Inhibition studies using related peptides help determine the precise location of the topographic sites. Using this method, Bidart *et al.* (1988) have determined that the region α-41-45 is located at the α-β interface - as a monoclonal antibody raised to the free α-subunit was unable to recognize the site in the native hormone. This finding is supported by the experiments involving individual amino acids discussed above. Further support is given by the observation that α-subunits glycosylated at Thr 43 are unable to recombine with dissociated β-subunits (Parsons |it et al. (1984)). Enzymatic degradation of free α using trypsin resulted in the removal of the region between Arg 39 and Lys 55. This product was unable to combine with β-subunit, but still retained the ability to bind conformation dependent monoclonal antibody and anti-α subunit antisera.

Immunochemical mapping techniques have also been utilized by Troalan *et al.* (1988). A monoclonal antibody was raised against a 34-residue peptide analogous with hCG α-63-96. Inhibition experiments revealed that the binding site of the antibody was located at the C-terminal end and involved at least the last six residues. This antibody was unable to bind to native hormone indicating that this region is not accessible in the α-β dimer and, therefore, is probably also located at the subunit interface. The residues Lys 95, Tyr 92 and Tyr 93 are invariant between species, adding weight to the argument that this region plays an important role. Also, in the native hormone, the region is protected from the action of proteolytic enzymes. In apparent contradiction to this hypothesis, removal of the C-terminal five amino acids of the α-subunit does not prevent α-β dimerization. However, changes in conformation are observed and biological activity is lost. Combining these findings with the immunochemical work it is possible that this loss of activity is the indirect result of a disturbance of the effective interaction between the subunits, rather than direct involvement of the region binding

to receptor as postulated by Parsons and Pierce (1979).

Several regions in the β-subunit have been postulated to be located at the subunit interface. Pierce *et al.* (1981) discovered that the disulphide linkage 93-100 in bLH could be reduced in the free β-subunit, but not in the intact hormone unless denaturants were added. Also hCG β-Thr 97 can only be phosphorylated in the free subunit (Keutmann *et al.*, (1983b)). The C-terminal extension of hCG is not involved in subunit interaction as the α-hCG readily combines with the β-subunit of LH which lacks this region. Chen and Puett (1991) characterized two β-hCG mutants lacking residues 100-145 and 93-145. The former recombined with b-α to produce a biologically active mutant, whereas the latter was unable to bind to the α-subunit. The possible crosslinking of Asp β-111 to Lys α-49 also infers that this region is at or near the subunit interface. A monoclonal antibody developed by Bidart *et al.* (1987) binds to separate regions including peptides 1-8 and 82-92 of β-hCG. It inhibits the recombination of the subunits and so may be binding at or near the subunit interface. Similar results have been obtained by Lustbader *et al* (1992) who developed a monoclonal antibody which binds to β-hCG and to β-hCG lacking the C-terminal region but not to β-LH or to native hCG. The lack of binding to β-LH indicates that the epitopes for the antibody are the least homologous regions of the subunit, namely 8-10 and 82-92, and the lack of binding to hCG infers they are at or near the subunit interface.

1.9.2 Location of Receptor Binding Region

Receptor binding of the native hormone apparently directly involves both α and β subunits. Native hCG has been found to bind to ovarian membrane receptors with an affinity of 10^{-10}M. The individual α and β subunits bind with lower affinity (10^{-5}M). As the subunits recombine, they acquire the full biological activity. The relatively low binding affinities of the individual subunits suggest that conformational changes occur upon subunit combination or after receptor binding for a high affinity interaction.

Immunochemical work confirms that both subunits are directly involved in binding to

receptor in the native hormone. Five monoclonal antibodies (Mabs) were prepared by Moyle *et al.* (1982), their epitopes determined and their ability to inhibit the biologically activity of hCG measured. Three of the Mabs bind to regions on the β-subunit. One of these inhibits hCG binding to the receptor and is not exposed in the hormone-receptor complex. The other two antibodies are still able to bind to their epitopes after complex formation suggesting that a portion of the β-subunit remains exposed to the solvent. Proteolysis experiments indicate that these two Mabs bind to part of the core of the β-subunit. Other information on exposed regions of the β-subunit after receptor binding comes from the fact that hCG devoid of the 30 residue C-terminal region is still biological active so this area is presumably not involved in receptor binding (Matzuk *et al.* (1990b)).

Evidence for the direct involvement of the α-subunit in receptor complex formation comes from two monoclonal antibodies directed at part of it. One of these Mabs inhibits binding of hCG to receptor and prevents steroidogenesis. The other does not prevent hormone-receptor complex formation, but as it also cannot bind to the preformed complex its epitope is presumably at least partially shielded by the receptor.

Synthetic peptides have been utilised to determine exactly which residues are involved in hCG binding to receptor (Charlesworth *et al.* (1987); Keutmann *et al.*, (1987)). The receptor binding affinity of the peptides was measured by the inhibition of binding to ^{125}I-hCG to rat ovarian receptor. hCG was radiolabelled at the N-terminus. Four sites in hCG and hLH are postulated to be involved in receptor binding on the basis of these assays. These are α 29-45, α 80-96, β 38-57 and β 93-101. All these synthetic fragments are reported to bind to ovarian membrane receptors at concentrations of 10^{-4} - 10^{-6} M. The low affinities could be attributed to conformational differences between the peptides and the equivalent regions in the native hormone. It is postulated that the folding of the native hormone brings together these linear sequences to form a single topographic site which forms the receptor binding region of the hormone. Other non-sequential amino acids, undetectable by this synthetic peptide method, could also be involved.

The sequence α 29-45 is in an invariant region and contains many aromatic and basic residues. The effects of chemical modification of several of these amino acids have been reported (see section 1.7 above). Apparently contradictory conclusions have been drawn; for example Tyr-41 is thought to be present at the α-β interface, whereas Arg-39 and Arg-46 are not (Sairam (1976)). These findings should be considered in terms of the three dimensional nature of the hormone. Also the results of the immunochemical work suggest the region α 41-45 is critical for α-β recombination (Bidart *et al* (1988)). As both subunits have been found essential for full biological activity, it is reasonable to suggest that at least part of the receptor binding domain spans the two subunits near to the subunit interface. Thus perhaps it is not suprising that residues and regions implicated to be important in one of these roles also effects the viability of the other. Further work using synthetic peptides has revealed the relative importance of specific residues in this region (Reed *et al.*, (1991)). The core peptide required for inhibition of binding is α-34-48. Mutation of each residue in turn to alanine indicated that Arg-39 and Phe-37 are essential for activity and, to a lesser degree, Arg-46, Ser-47, Cys-35 and Cys-39. All these residues are absolutely conserved between species. Further mutagenic work would be neccessary before the roles of these residues in receptor binding can be determined.

The second region in the α-subunit, 88-96, is also invariant and contains the two Tyrosines which when modified reduce receptor binding affinity. Further support for it also being important in receptor binding comes from the work of Merz (1979). Removal of 2-5 residues at the C-terminal of the α-subunit results in almost the complete loss of receptor binding affinity after combination with the β-subunit.

Ward *et al.*, (1979) predicted that the loop between β Cys-93 and β Cys-100 was important in the biological activity of hCG and that the distribution of charge in this region is responsible for specificity of receptor interaction. In hCG and hLH this loop has a net positve charge. Keutmann *et al.* (1989) tested this hypothesis by measuring the relative binding affinities of this loop region and several analogues in which the net charge was changed by substitution of Arginines for Aspartates. It

was found that none of these analogues exhibited significant inhibition. The effect of overall charge is also highlighted by the lack of inhibition by hFSH which has a net negative charge in this loop. Also the importance of the disulphide linkage 93-100 was demonstrated. A synthetic peptide in which the Cysteines were replaced by Alanines was inactive, indicating the disulphide linkage is vital for the maintainance of the correct conformation for receptor binding. Other experiments apart from this synthetic peptide approach support the involvement of this loop in the biological activity of hCG. Phosphorylation of Thr-97 diminished receptor binding, and enzymatically cleaved β-hCG, including cleavages at 95/96 and 43/44, on recombination with α-subunit showed decreased binding ability (Birken *et al.*, (1987)). The proximity of this loop region to the α-β interface is demonstrated by the work of Chen and Puett (1991) on mutants, and by the observation that phosphorylation of Thr-97 is possible only in the isolated β-subunit.

The second loop region in the β-subunit implicated in receptor binding, 38-57, also has Cysteine residues at each end (Keutmann *et al.* (1988)). However, in this case the secondary structure of the loop is retained even when Cys-38 is mutated to Ala-38 as evidenced by CD measurements and full immunoreactivity to antibody recognizing the native peptide. 20% receptor binding activity remained. The amphipathic helical properties of this region support its role in receptor interaction, as it may be important in allowing the hormone to approach the lipid environment of the cell membrane. For receptor binding activity, the critical factors are the presence of the positive charge at Arg-43 (the analogues Asp-43 and Ala-43 exhibited no inhibition), neutral residues at 47 and 52 and an intact methionine at 41. Again this region is close to the α-β interface and includes residues which on chemical modification prevent subunit recombination.

The X-ray crystallographic structure of human growth hormone (hGH) complexed with its receptor (De Vos *et al.*, (1992)) has revealed that two receptor molecules bind to each hormone molecule. The same residues on the receptor are involved in the binding to two distinct domains on the hormone. It is possible that a similar interaction occurs when hCG binds to its receptor and that the receptor binding domains discussed above

are not necessarily spatially close.

1.10 Summary

Human Chorionic Gonadotropin is a member of the glycoprotein family which includes Luteinizing Hormone, Follicle Stimulating Hormone and Thyroid Stimulating Hormone. It consists of an α and a β subunit which are non-covalently bound. The α-subunit contains 92 residues, five disulphide bridges and has two sites of N-linked glycosylation. The β-subunit contains 145 residues, six disulphide bridges, two sites of N-linked glycosylation and four O-linked oligosaccharides in the proline and serine rich C-terminal region. CD experiments show it to contain very little helical structure and to consist mainly of β-sheet and aperiodic structure, the latter being well defined, presumably because of the disulphide linkages. The β-subunit is responsible for the biological specificity of hCG, but both subunits are involved in receptor binding. Four regions have been found to be important in this binding, two of which are on the α-subunit and two on the β-subunit. All regions are thought to be close to the α-β interface and probably form a single topographic binding domain.

The role of the carbohydrate is discussed in the next Chapter.

Chapter 2

Deglycosylation and Crystallization

2.1 Carbohydrate Content of the Hormones

As indicated in Chapter one, the glycoprotein hormones LH, FSH, TSH and CG all contain a large proportion of carbohydrate. hCG has the highest content of all of approximately 30-35% by weight. It contains four *N*-linked oligosaccharide chains bound to residues α56, α82, β13 and β30, found in the Asn-X-Thr sequence. Also there are four O-linked carbohydrate moieties in the serine rich carboxyterminal region at residues β121, β127, β132 and β138 (Kessler *et al.*, (1979)). The proposed structure of these chains is given in Figure 2.1. For comparison, the alpha subunits of ovine, porcine and bovine FSH and LH are glycosylated at Asn 56 and Asn 82. The β-FSH subunits are glycosylated at residues 13 and 30, whereas the β-LH subunits contain oligosaccharides bound either to residue 13 or 30, the latter being the case in b-LH. hTSH has *N*-linked oligosaccharides at residues α56, α82 and β30.

The carbohydrate structures for all the glycoproteins are similar to that in Figure 2.1. However, the pituitary hormones are prone to greater heterogeneity. The terminal residue may be *N*-acetyl neuraminic acid or sulphate, depending on the species and tissue of origin, *e.g.* oLH and bLH have sulphate as the terminal residue, whereas hLH has both present. The role of the carbohydrate is discussed in Section 2.5.

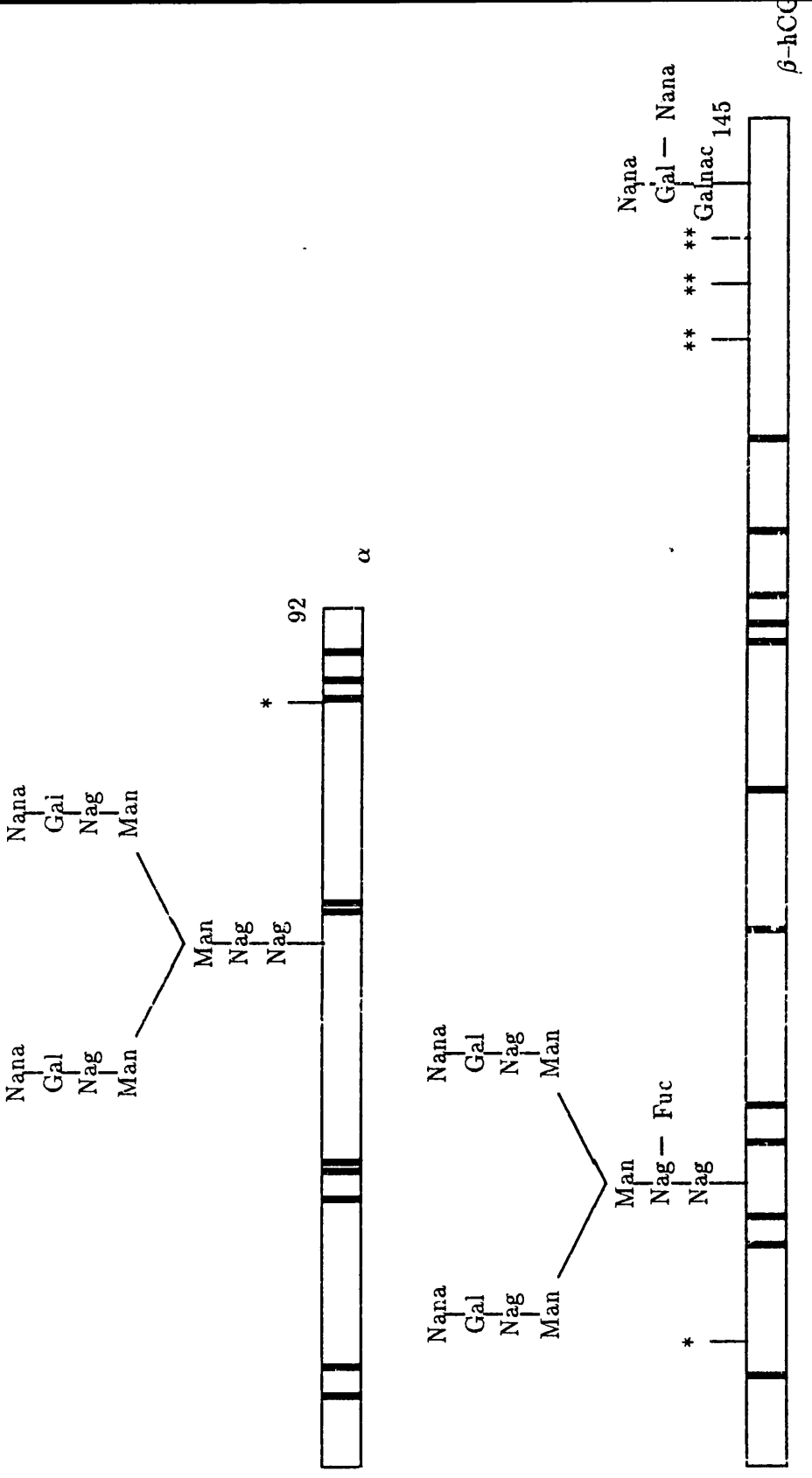


Figure 2.1: Carbohydrate Structure of hCG

Man - Mannose ; Nag - N-AcetylGlucosamine ; Gal - Galactose ; Fuc - Fucose

Galnac - N-AcetylGalactosamine ; Nana - N-acetylneuraminic Acid

* indicates position of oligosaccharide identical to that shown for the nearest neighbour.

** indicates postion of serine-linked oligosaccaride as shown.

Solid bars in the figure indicate positions of cysteine residues.

2.2 Problems with Crystallization - Carbohydrate Heterogeneity

Glycoproteins have commonly been found difficult to crystallize in a form suitable for X-ray diffraction studies. Attempts to grow crystals of hCG and other members of this family have met with no success so far. The only positive reports of crystallization experiments are from McPherson *et al.* (1986) using bTSH α and bLH α and Ryan *et al.*, (1987)) using αhCG. In all cases the crystals occurred as very fine needles emanating from a single nucleation site, none of which have proved useful for X-ray work.

The lack of crystallization is commonly attributed to the microheterogeneity of the carbohydrate moieties. It is presently believed that when a glycoprotein is biosynthesized the polypeptide chain and the carbohydrate units are constructed independently. The asparagine linked oligosaccharides are synthesized from oligomannose precursors. This addition occurs as the protein is translated and is temporally close to cleavage to the signal peptide making it a relatively precise event. These precursors are then acted upon by a series of glycosyltransferases and glycosidases. These enzymes act in a very specific manner to allow the building of complex carbohydrate structures; the sugars are transferred stepwise from the nucleotide derivative to the protein, with the product of one reaction becoming the determinant for the next. The degree of completion of glycosylation is thus dependent on the rate at which the nascent protein passes through the cell, and several variants are usually found to exist. It is thought that these variants probably prevent the growth of large crystals. One way to overcome this is to deglycosylate the glycoprotein. This can be achieved to varying degrees of completion and specificity depending on the method employed.

2.3 Enzymatic Deglycosylation

Enzymatic degradation of the carbohydrate moieties of the glycoproteins can be accomplished using exo- or endo-glycosidases. The latter can be used to degrade large carbohydrate units as they recognize oligosaccharides as substrates. Exo-glycosidases act in a very specific manner by removal of the terminal non-reducing sugar from the oligosaccharide and thus removal of a large portion of the carbohydrate can be achieved by sequential treatment with a series of glycosidases.

Partial deglycosylation of hCG has been carried out using this method by Moyle *et al.* (1975). This resulted in almost total removal of the sialic acid, 60% of galactose, 52% of *N*-Acetylgalactosamine and 20% of mannose *i.e.* approximately 60% of the total carbohydrate content. However, this method has several disadvantages; the long incubation periods neccessary (162 hours for four steps), the low yields obtained ($\sim$ 20% recovery) and the possibility of proteolytic cleavage.

Lustbader *et al.* (1989), following the procedure of Rosa *et al* (1984) successfully desialylated hCG using neuraminidase. This removed all the sialic acid, approximately 8.2% by weight of the total carbohydrate content. The effect on the biological activity of this treatment has been researched (Tsuruhara *et al*, (1972). Desialytion does not impair hormone binding *in vitro* to testicular LH receptor and steroidogenic activity is reduced by only 50%. However, *in vivo*, biological activity was reduced to 1% of native hCG, as a result of a decrease in the plasma half-life to 1 minute.

2.4 Chemical Deglycosylation

Chemical deglycosylation of hCG has been achieved using both anhydrous hydrogen fluoride (HF) and trifluoromethane sulphonic acid (TFMS). Deglycosylation using TFMS is slow and care must be taken to exclude water from the reaction vessels. Kalyan *et al.* (1983) report the successful use of TFMS to remove approximately

90% of the carbohydrate from the individual subunits, with only the last peptide linked hexosamine remaining attached. The same TFMS treatment of whole hCG also gave 90% deglycosylation, but dissociation of the subunit occurred. Degly-hCG could be obtained by recombination under changed conditions. Analysis of the deglycosylated product using several biochemical and immunological techniques indicated no change in secondary struture. In contrast to this, Ryan *et al.* (1987) have found that treatment of hCG with TFMS can generate fluorinated methionine residues.

Anhydrous hydrogen fluoride (HF) is less safe to use, but deglycosylation is rapid. Treatment with HF for one hour at 0°C in the presence of anisole scavenger cleaved more than 75% of the total carbohydrate content of hCG (Manjunath *et al.* (1982); Chen *et al* (1982)). The product was found to retain the quaternary structure of native hCG as evidenced by receptor binding and immunological activities. Keutmann *et al.*, (1983a) applied the same treatment to the individual subunits and found that 80% and 60% of carbohydrate was cleaved from the alpha and beta subunits respectively. Again, chemical studies showed the peptide portions were unaffected by the reaction. Circular Dichroism spectra were unchanged, see Table 2.1, and the deglycosylated subunits readily recombined.

Ryan *et al.* (1987) have reviewed the reports of deglycosylation of hCG and similar glycoproteins and suggest that the HF treatment is the most successful. Using data from Keutmann *et al.* (1983) they state that HF cleaves the *N*-linked oligosaccharides between the mannose and the penultimate *N*-Acetylglucosamine residues, but that the O-linked oligosaccharides remain intact except for the removal of the terminal sialic acid.

2.5 Biological Role of the Carbohydrate

The role of the carbohydrate content of these hormones was been widely studied. The oligosaccharides are important for *in vivo* stability and steroidogenic activity.

	Phosphate Buffer (0.05M)			90% Trifluoroethanol		
	α-helix	β-Structure	Random	α-Helix	β-Structure	Random
Form	(%)	(%)	(%)	(%)	(%)	(%)
βhCG	8	36	56	14	31	56
d.βhCG	8	36	56	15	28	57
αhCG	6	40	55	19	34	46
d.αhCG	4	41	54	20	35	46
hCG	9	36	54	12	38	50
d.hCG	10	35	55	12	37	51

Table 2.1: Circular Dichroic Analysis of hCG and HF-hCG[a]

[a] Ryan *et al.*, (1987)

The affects of chemical and enzymatic deglycosylation on tertiary structure, receptor binding and biological activity have already been discussed. The relative importance of the various *N*-linked oligosaccharide chains has been probed using site directed mutagenesis experiments (Matzuk *et al.* (1989)). The receptor binding affinity was little affected by the absence of all or any of the 4 N-linked oligosaccharides. The production of c.AMP and steroidogenesis was uneffected by the absence of the β-subunit carbohydrate or that bound to α Asn-82. In the mutant lacking carbohydrate at α Asn-56, both c.AMP and steroidogenesis responses were decreased. When the β-subunit mutants were combined with α-subunit without Asn-82 carbohydrate, it was found that β Asn-13 is the more important β *N*-linked oligosaccharide.

The mechanism by which the carbohydrate induces biological action is not understood. Receptor binding is independent of the carbohydrate moieties, so it is probable that these groups are located away from the receptor in the hormone-receptor complex. It is possible that a conformation change in the α and or β subunits is induced. Calvo and Ryan (1985) suggest that the *N*-linked oligosaccharides bind to a membrane lectin

which may be part of the receptor or nearby in the plasma membrane. Thotakura, *et al* (1990) report that the binding of hCG to highly purified receptor can be inhibited by a variety of glycopeptides, inferring that any lectin-like site must be located on the receptor.

The O-linked oligosaccharides in hCG are not thought to be involved in the biological activity of the hormone as *in vitro* a mutant lacking the C-terminal extension is capable of binding to receptor and inducing signal transduction. However, *in vivo* the truncated hCG is 3-fold less potent (Matzuk *et al.*, (1990a). The plasma half-life of native hCG is greater than that of the other glycoproteins which lack the C-terminal extension. This has been attributed to the increased negative charge due to the O-linked carbohydrate, and is supported by the fact that the more negatively charged (higher sialic acid content) forms of FSH also have longer half-lives.

2.6 Anhydrous Hydrogen Fluoride Treatment of hCG.

The method used to deglycosylate hCG in our laboratory was that of Manjunath and Sairam (1982), mentioned above. 90mg of commercial hormone (purchased from Organon, the Netherlands) was dried over phosphorus pentoxide in a dessicator under vacuum overnight. The dried protein was loaded into the reaction vessel of a hydrogen fluoride vacuum-line system (Peptide Institute Inc., Osaka, Japan), and the temperature reduced to -77°C using a dry ice/acetone bath. 10ml of hydrogen fluoride was condensed into the reaction vessel, and the reaction was allowed to proceed at 0°C for 45 minutes. The vessel was then opened to partial vacuum, and the hydrogen fluoride allowed to boil off over a period of about 30 minutes. Residual acid was removed by evacuation of the vessel using a rotary pump for at least 1 hour. The protein was brought to pH 6.0 using 0.2M sodium hydroxide, and rinsed from the reaction vessel with a minimum amount of distilled water.

The product was finally purified and analysed by Fast Protein Liquid Chromotography

Buffer Id.	Buffer Conc.	pH	Salt Conc.
Buffer A	25mM NaoAc	pH 4.9	
Buffer B	50mM NaoAC	pH 4.9	1M NaCl

Table 2.2: Buffers used for FPLC analysis of HF-hCG

Time (Mins.)	Variable	Conc.
0.0	Conc. %B	0.0
0.0	ml/min	1.00
5.0	Conc. %B	0.0
55.0	Conc. %B	100.0
60.0	Conc. %B	100.0
62.0	Conc. %B	0.0
70.0	Conc. %B	0.0

Table 2.3: Details of procedure followed during FPLC analysis of hCG

(Pharmacia) on a MONO-S column. The buffers used are listed in Table 2.2 and the experimental procedure in Table 2.3.

The result of this procedure is shown in Figure 2.2. A single peak is observed at 38% Buffer B concentration. A similar procedure was followed by Lustbader *et al.*, (1989). Table 2.4 gives their analysis of the resulting carbohydrate content of HF treated hCG. For comparison, their results of treatment with neuraminidase are included.

2.7 Crystallization

Crystallization trials of HF-hCG were conducted using the vapour diffusion method. 24 well plastic culture trays (Flow laboratories) were used as a number of trials could be set up using a minimum amount of protein and easy access was possible for inspection

	nmoles[b] Glycoprotein analysed	nmoles[c] Glc NH_2	nmoles[d] Gal NH_2	nmoles[e] Gal	nmoles[f] Man	nmoles[g] Sialic acid	% Carbohydrate
hCG	1.77	27.3(15.5)[h]	7.3(4.1)	17.9(10.1)	21.2(12)	27.6	33.9%
As-hCG[i]	1.66	25.4(15.3)	7.9(4.7)	20.6(12.4)	19.9(12)	-	25.7%
HF-hCG	1.65	12.9(7.8)	8.15(4.9)	9.1(5.5)	9.9(6)	1.96	17.1%

[a] Not including fucose

[b] Nano moles of glycoprotein analysed was calculated from nmoles of mannose (12 nmoles mannose/mole of hCG or As-hCG and 6 nmoles mannose/mole HF-hCG.

[c] Glc NH_2 = glucosamine; [d] Gal NH_2 = galactosamine; [e] Gal = galactose; [f] Man = Mannose; [g] Sialic acid = N-acetylneuraminic acid

[h] Parentheses contain nmoles normalized to 12nmole mannose for hCG and As-hCG and 6 nmoles mannose for HF-hCG.

[i] Asialo hCG is neuraminidase-treated hCG

Taken from Lustbader *et al.*, (1989)

Table 2.4: Carbohydrate Analysis of hCG, HF-hCG and Asialo-hCG.

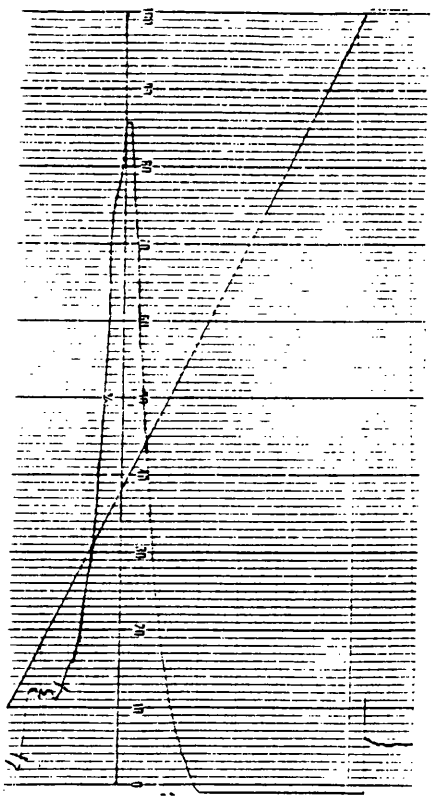


Figure 2.2: Analysis of HF treated hCG by FPLC

for crystals.

A 20mg/ml protein solution was made up in the chosen buffer system. Each well was filled with 1ml of buffer solution containing a known concentration of a precipitating agent. 4μl of this well solution was added to 4μl of protein solution and the drop was then suspended over the well from siliconized coverslips and sealed with vaccuum grease. A diagram of this is given in Figure 2.3. Crystallizing proteins is very much a trial and error process as there are so many variables, e.g. protein concentration, choice of precipitant and concentration, choice of buffer and concentration, temperature, drop size and pH. Table 2.5 gives the buffers and the precipitating agents used in attempting to crystallize hCG. As none of these combinations proved successful, the protein concentration was increased to 40mg/ml and the trials repeated. After three weeks crystals were found growing in solutions of 25-35% saturated Ammonium sulphate in sodium acetate buffer, pH 4.5. These crystals were small, reaching a maximum size of approximately 0.2mm in each direction, and of very poor morphology. However, when placed in an X-ray beam they diffracted to 9Å resolution.

The conditions were varied in an attempt to improve the crystal quality. The optimum set was found to be 36mg/ml protein solution in 30mM Sodium Acetate buffer, pH 4.2,

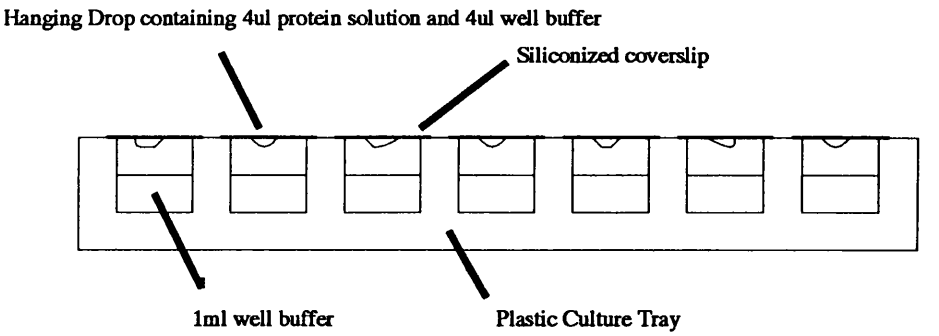


Figure 2.3: Crystallization Set-up

Buffer System	pH	Precipitating Agents
Bis-Tris[a]	6.0	Sat[d] Ammonium Sulphate
50mM Phosphate	5.5	Sat[d] Lithium Sulphate
50mM Phosphate	8.5	Sat[d] Sodium Chloride
30mM Sodium Acetate	4.5	Distilled Water
		Ethanol
		M.P.D.[b]
		PEG[c] 2K, 4K, 6K

Table 2.5: Buffer systems and precipitating agents used in the crystallization trials for HF-hCG.

[a] [bis(2-Hydroxyethyl)iminotris(Hydroxymethyl)methane]

[b] Methyl 2,4-Pentanediol

[c] Polyethylene Glycol

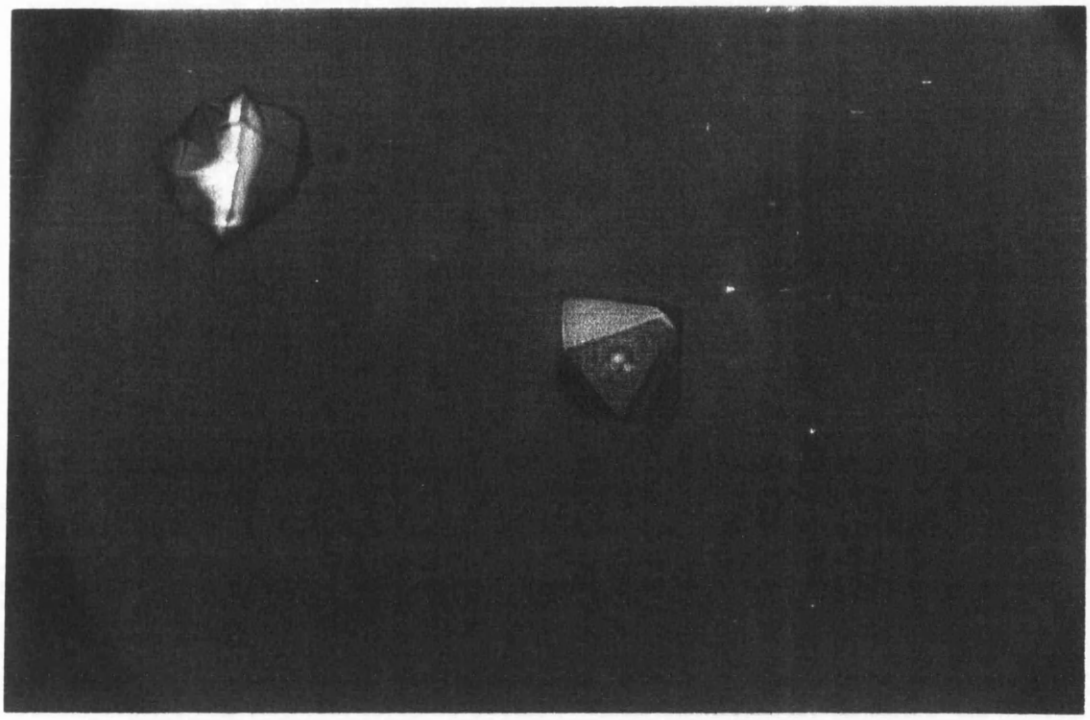


Figure 2.4: Crystals of HF-hCG

with a 30-36% saturated Ammonium Sulphate concentration. Under these conditions crystals appeared after three-four weeks, reaching full size in about two months. They grew as sharp hexagonal pyramids up to 0.6mm in each direction (Figure 2.4). The crystals were mounted in glass capillaries and a diffraction pattern recorded using a Huber precession camera on a Rigaku rotating anode X-ray generator. Diffraction extended beyond 3.5Å. (Harris *et al.* (1989)). Attempts to further improve the crystal quality using secondary precipitating agents proved unsuccessful. Low temperature (4°C) resulted in amorphous precipitate.

2.8 Crystal Characterization

As stated above, the HF-hCG crystals grew with a definite hexagonal habit. A crystal was mounted in a glass capillary on a Huber precession camera and aligned with the beam parallel to the presumed six fold axis. The resulting precession photograph, Figure 2.5, shows 6-*mm* symmetry in the *hk0* zone. An *h0l* zone precession photograph was also recorded, Figure 2.6. 2-*mm* symmetry is apparent, and systematic absences

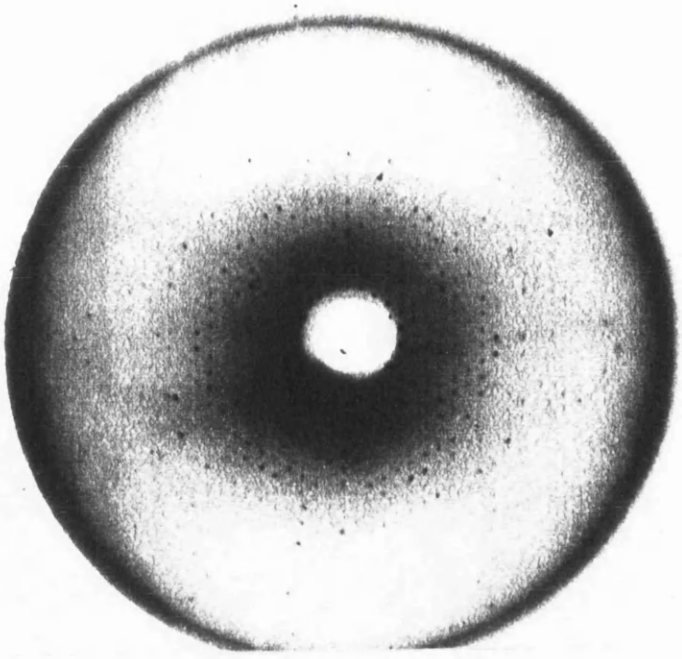


Figure 2.5: μ = 7° Precession Photograph - *hk0* zone. Crystal to film distance of 10cm.

are observable along the *00l* axis. Only reflections with l=6n are present, giving a space group of $P6_122$ or the enantiomer $P6_522$. The cell dimensions were measured from these photographs using a STOE device as a = b = 88.7Å and c = 177.2Å.

2.9 Asialo-hCG and HF-hCG crystals; a comparison

Lustbader *et al.*, (1989) subsequently used similar conditions to those developed in our laboratory to obtain crystals of neuraminidase treated hCG, the difference being the use of a Sodium Citrate rather than a Sodium Acetate buffer system. These crystals were found to have the same space group possibilities and very similar cell dimensions.

SDS gel electrophoretic Analysis of the HF-hCG crystals revealed a large degree of heterogeneity implying an absence of selectivity in the crystallization process. This degree of heterogeneity was not found in the asialo crystals. Lustbader *et al.*, (1989) report peptide bond cleavages in crystals of both species. HF-hCG was found to have undergone cleavage at the acid labile βAsp112-βPro113 and βAsp117-βSer118

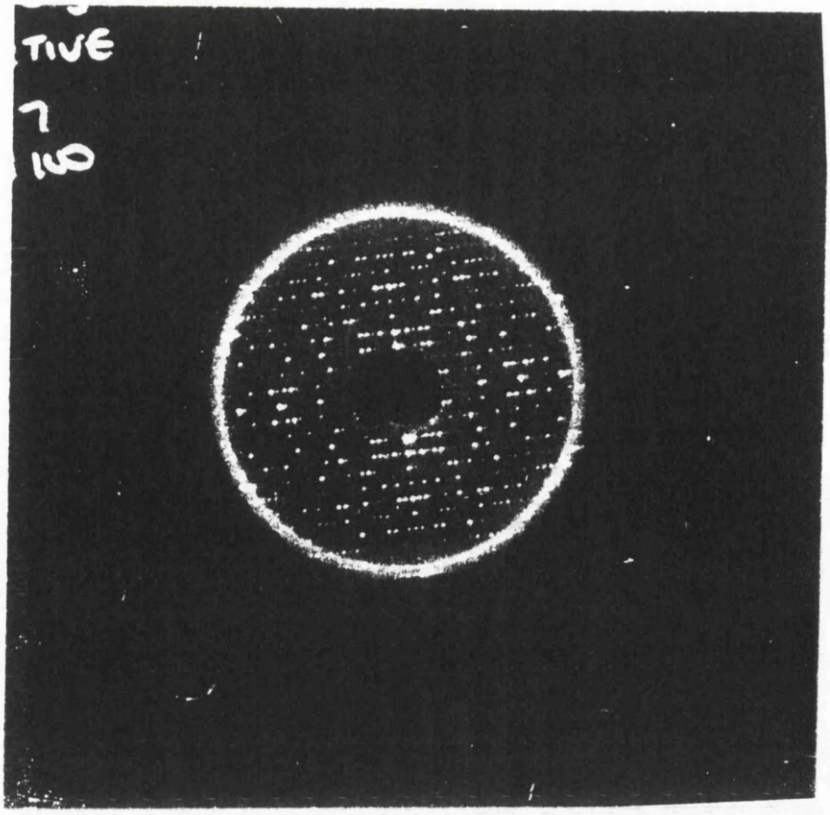


Figure 2.6: Precession Photograph - h0l zone. Crystal to film distance of 10cm.

positions. The degrees of cleavage are given as 45% and 20% on a molar basis respectively. The asialo-hCG only suffered 7% cleavage at the βAsp112-βPro113 position. It is possible that the harsh chemical treatment used to obtain HF-hCG increased the susceptibility of the polypeptide backbone to proteolytic cleavage. No cleavage was found in the starting material used for crystallization, so proteolysis occurs in the crystallization medium, a buffer at pH 4.2-4.5.

Both HF-hCG and asialo-hCG still contain a relatively large amount of carbohydrate and a degree of heterogeneity. This suggests that other factors may be responsible for the difficulty found in crystallizing glycoproteins in general. Isoelectric focussing experiments reveal that urinary hCG contains seven distinct species of different charges. However, asialo-hCG forms only one band at pH 9.5. Treatment with HF also removes the negatively charged sialic acid residues. It has been suggested that it is perhaps the heterogeneity in charge that is primarily responsible for problems with crystal growth of glycoproteins. This idea is supported by the improved crystals obtained from interleukin-2 receptor after removal of the terminal sialic acid residues (Lambert *et al.*, (1989)).

Chapter 3

Preparation of Heavy Atom Derivatives

3.1 Introduction

There are two main methods of phase determination for protein structures. These are the molecular replacement method suitable for those proteins which are closely homologous with a known structure, and the isomorphous replacement method when no such structure is available. As the three-dimensional structure of all the glycoproteins is unknown, the phase problem in this study was tackled using the isomorphous replacement approach.

The isomorphous method was first utilized for structure determination of phthalocyanines (Robertson & Woodward (1937)) and in 1954 the method was first used for protein structure determination (Green *et al.*, (1954)) to haemoglobin. For small molecules it is usual to actually substitute an atom with a similar, but heavier, atom. For proteins it is more common to introduce a heavy atom into the protein crystal. This is possible as protein crystals consist of a large percentage of solvent, usually in the order of 30-70%, and the heavy atom can be diffused into the solvent channels.

A good isomorphous derivative is produced when a heavy atom is introduced into a crystal with minimal disturbance of the protein conformation, orientation or crystal lattice. Ideally there should be a limited number of heavy atoms with high occupancies bound specifically to the protein, as this greatly simplifies the solution of the Pattersons. In reality binding is often found to be of a multiple site, low occupancy nature and small changes of the cell dimensions are usually observed. Crick and Magdoff (1956) estimated that a 0.5% change in cell dimensions or a 0.5° rotation of the molecules in the cell can cause a 15% change in intensities to 3Å resolution. These changes are dependent on resolution and so a good test for isomorphic binding of heavy atoms is to plot fractional isomorphous differences against resolution. The graph should not rise too steeply with resolution and the isomorphous differences should be in the order of 10-25%. The actual change in intensities for acentric reflections is given by

$$(2N_H/N_P)^{1/2}.(f_H/f_P)$$

where N_H heavy atoms of scattering factor f_H are bound to a molecule of N_P light atoms of scattering factor f_P (Crick & Magdoff (1956)).

3.2 Choice of Derivatives.

Heavy atom derivative preparation is still largely a matter of trial and error. However study of the primary amino acid sequence of a protein, if known, is advisable. A comparison of the amino acids available as prospective binding sites with the results of previous isomorphous replacement studies can narrow the prospective candidates for the trials. For example, if the protein contains a free cysteine residue then choice of a mercury compound is recommended. Also the mother liquor of the crystal is not neccessarily conducive to derivative preparation. High concentration salt-precipitating agents and phosphate buffers can react with many potential derivatives. If crystals are grown in these kind of conditions, where possible they should be transferred into the less reactive medium such as tris buffer or polyethylene glycol. Another consideration

is the pH of the crystallization liquor. If the crystals grow at acidic pH, then residues such as histidine are liable to be protonated and unavailable for specific heavy atom binding. Depending on crystal stability the pH can be altered to allow a certain number of residues to become available.

All these factors are very much dependent on the feasibility of transferring crystals from their mother liquor. Also such theoretical predictions about putative binding sites assume that the residues involved are orientated suitably in the crystal for heavy atom binding to be possible. Obviously this will not be known until the three dimensional struture is solved.

3.3 Derivative Preparation - Procedure

To find suitable heavy atom derivatives for hCG a number of experiments were performed. For each heavy atom chosen the concentration and length of the soak were varied. The majority of the trials to find derivatives for hCG have been conducted in the normal mother liquor. The crystals were fairly fragile and unstable in lower salt concentrations. A known mass of each heavy atom compound was made up in fresh 36% saturated ammonium sulphate in 30mM sodium acetate, pH 4.3. Stock solutions were made up ten times more concentrated than needed for the soak, and then diluted with liquor from the well to minimize disruption of the crystal by changes in salt concentration. A known amount of heavy atom solution, usually 40μl, was placed in a watch-glass and the crystal transferred into this using a needle. The soak experiment was then sealed in a petri-dish using silicon grease with 1ml of mother liquor in the dish to reduce evaporation. A diagram of this set-up is given in Figure 3.1. Each crystal was monitored at regular intervals during the soak experiments for signs of fragility, colour uptake or sinking. Soaking experiments are described in the following Tables. Table 3.1 details soaks involving gold and mercury compounds, Table 3.2 those using platinum compounds and Table 3.3 the remaining compounds. After a known time, the crystal was mounted in a capillary tube and immediately placed in the X-ray beam.

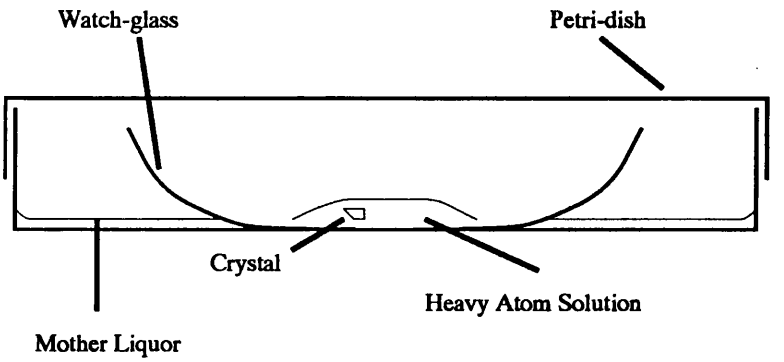


Figure 3.1: Heavy Atom Soaking Experiment

In some experiments, the crystals were back-soaked for a short period before being mounted to allow any unbound heavy atoms to diffuse out of the solvent channels.

For some trials crystals were successfully transferred into 30mM citrate buffer, pH 6.3, in the hope that some or all of the four histidines would be deprotonated at this pH. Crystals were first placed in the buffer and, once seen to be stable, were then transferred to the heavy atom solution made up in the citrate buffer, and the experiment conducted as described above. Table 3.4 gives details of these trials.

3.4 Screening of Derivatives.

During the course of this project, the putative heavy atom derivatives have been screened using several techniques. Initially, they were scanned using the precession method. Cell dimensions were measured from the photograph using a *STOE* device, and each photograph was inspected relative to the native precession photograph for changes in intensity. It was important that these changes be visible at high resolution, as changes solely at low resolution could be caused by unbound heavy atoms within the solvent channels. Once a Siemens area detector was installed in the laboratory the

No.	Heavy Atom Cmpd.	Soak Conditions			Observations
		Concn (mM)	Duration (Hrs)	Backsoak (Hrs)	
1	$Hg(OOCCH_3)_2$	2	29	0.25	Crystal developed dappled surface
2	$Hg(OOCCH_3)_2$	2	25	0.5	Surface cracks
3	$Hg(OOCCH_3)_2$	2	24	0.5	Some surface cracks
4	$Hg(OOCCH_3)_2$	1	40	1	Slight Cracking
5	$Hg(OOCCH_3)_2$	1	20	1	Small crystal. Poor surface. Cracking
6	$Hg(OOCCH_3)_2$	1	20	1	Poor surface. No bad cracks. Crystal fine
7	$Hg(OOCCH_3)_2$	5	24	3	Okay
8	$Hg(OOCCH_3)_2$	10	1	-	Immed. crack at edge. Rest okay
9	Thiomersal	Sat.	24	-	Surface cracks, worsening
10	Thiomersal	Sat.	17	0.5	'Crazy Paving' surface
11	Thiomersal	Sat.	3	-	Very bad cracking
12	$K_2Hg(SCN)_4$	Sat.	21	-	Small crystal. Okay
13	Phenylmercuri acetate	Sat.	24	-	Crystal disintegrated
14	Me_2Hg	-	-	-	Fine
15	Me_2Hg	-	-	-	Fine
16	$NaAuCl_4$	5	19	2	Surface cracks. Colour uptake
17	$NaAuCl_4$	4	18	-	Small crystal. Fine
18	$NaAuCl_4$	4	4	-	Colour uptake
19	$NaAuCl_4$	4	4	-	Colour uptake after 1hr. Fragile
20	$NaAuCl_4$	3	25	-	Surface cracks
21	$NaAuCl_4$	2	24	0.5	Cracks by edge. Colour uptake
22	$NaAuCl_4$	2	25	-	Surface cracks. Colour uptake
23	$NaAuCl_4$	2	24	-	Suface cracks. Colour uptake
24	$NaAuCl_4$	2	23	0.5	Some cracking. Yellow colour
25	$NaAuCl_4$	1	25	0.5	Surface cracking
26	$NaAuCl_4$	1	21	0.5	Fine
27	$KAu(CN)_2$	10	24	-	Bad cracking
28	$KAu(CN)_2$	10	3	-	Fine
29	$KAu(CN)_2$	10	2	-	Fine

Table 3.1: Heavy Atom Soaking Experiments - Mercury and Gold

No.	Heavy Atom Cmpd.	Soak Conditions			Observations
		Concn (mM)	Duration (Hrs)	Backsoak (Hrs)	
30	K_2PtCl_4	2.5	18	-	Fine
31	K_2PtCl_4	2.5	5	-	Crystal sunk
32	K_2PtCl_4	2.5	4	-	Fine
33	K_2PtCl_4	2.5	4	-	Fine
34	K_2PtCl_4	2.5	4	-	Slight cracking. Okay
35	K_2PtCl_4	2.5	2	-	Colour uptake
36	K_2PtCl_4	1.8	22	0.5	Okay
37	$K_2Pt(SCN)_4$	8	15	-	Deep cracks. Colour uptake
38	$K_2Pt(SCN)_4$	2.5	3.5	-	Deep cracks. Colour
39	$K_2Pt(SCN)_4$	0.2	6	-	Deep cracks. Colour
40	$K_2Pt(CN)_4$	1.8	24	0.5	Fine
41	$K_2Pt(CN)_4$	1.8	22	0.5	Fine
42	$K_2Pt(CN)_4$	1.8	20	0.5	Fine
43	$K_2Pt(CN)_4$	10	24	1	V. slight cracking. Fine
44	$K_2Pt(CN)_4$	10	24	-	Fine
45	$K_2Pt(CN)_4$	10	24	-	Fine
46	$K_2Pt(CN)_4$	10	2	-	Fine
47	$K_2Pt(NO_2)_4$	5	21	0.5	Fine
48	$K_2Pt(NO_2)_4$	5	21	0.5	Fine

Table 3.2: Heavy Atom Soaking Experiments - Platinum

No.	Heavy Atom Cmpd.	Soak Conditions			Observations
		Concn (mM)	Duration (Hrs)	Backsoak (Hrs)	
49	$UO_2(NO_3)_2$	2.5	4	-	Fine
50	$UO_2(NO_3)_2$	2.5	24	-	Slight cracking. Okay
51	$UO_2(NO_3)_2$	2.5	72	-	Slight cracking. Okay
52	$K_3Ir(CN)_6$	5	25	1	Fine
53	$K_3Ir(CN)_6$	5	25	48	Fine surface cracks. Okay
54	$K_3Ir(CN)_6$	5	24	-	Fine
55	$K_3Ir(CN)_6$	5	15	0.75	Slight surface cracks
56	$K_3Ir(CN)_6$	4.7	24	0.75	Fine. Lost on mounting
57	$K_3Ir(CN)_6$	4.7	24	0.75	Fine
58	$Lu_2(SO_4)_3$	5	5	-	Fine
59	$Gd(OAc)_3$	5	11	-	Fine

Table 3.3: Heavy Atom Soaking Experiments cont.

No.	Heavy Atom Cmpd.	Soak Conditions			Observations
		Concn (mM)	Duration (Hrs)	Backsoak (Hrs)	
60	$Hg(OOCCH_3)_2$	2	72	0.5	Some surface cracks
61	$Hg(OOCCH_3)_2$	2	17	-	Fine. Some surface cracks
62	Mercurochrome	Sat.	1	-	Severe cracking. Strong Colour
63	Mercurochrome	Sat.	0.5	-	Cracking. Less intense colour
64	PCMB	Sat.	0.5	-	Extreme cracking. Crystals destroyed
65	$[Pt(NH_3)_4](NO_3)_2$	2	144	-	Surface cracks
66	$[Pt(NH_3)_4](NO_3)_2$	5	24	-	Surface cracks
67	$KAu(CN)_2$	10	2	-	Fine. Crystal sunk
68	$AgNO_3$	20	48	-	Fine
69	$AgNO_3$	15	16	-	Crystal cracked
70	$AgNO_3$	10	2	-	Surface cracks
71	Na_2IrCl_6	5	4	-	Okay
72	$K_2Pd(Br)_4$	5	1	-	Colour uptake. Crystal fractured
73	$Gd(OAc)_3$	2	2.5	-	Fine
74	$Gd(OAc)_3$	2	21	-	Slight cracking. Okay

Table 3.4: Soaking Experiments conducted in Sodium Citrate Buffer, pH 6.3

preferred method of derivative testing became the collection of a low resolution (6Å) data set. Data could be collected within 24 hours and analysed relative to a native data set.

3.4.1 Precession Method

Potential derivatives were screened in-house using precession cameras mounted on an Elliott rotating anode generator. The generator was typically operated at 40kW/45mA. Crystal to film distance was fixed to 75 or 100mm. When data were collected on Kodak DEF-59 film, exposure times for stills were one to two hours depending on diffracting ability. A $\mu = 3°$ precession required four to six hours. These exposure times decreased when the CEA Reflex 25 film was used. Once a crystal was aligned, a higher angle precession photograph was taken. The preferred zone was *h0l* as it was much easier to align the crystals in this way. Tables 3.5, 3.6 and 3.7 give details of the heavy atom compounds screened using this method. As can be seen many of the experiments are inconclusive and one of the major problems with determining which of the derivatives was a good candidate was that of poor diffraction. It was extremely difficult to align the crystal and take a large angle precession photograph before the crystal died. This situation was improved dramatically by the acquisition of a finer collimator which resulted in a decrease in the level of background radiation and thus allowed the use of the more sensitive CEA Reflex 25 film in place of the Kodak DEF-59 film used previously. Some examples of the resulting precession photographs are presented. The native *h0l* zone photograph is repeated in Figure 3.2 for ease of comparison with the derivative photographs. Figure 3.3 shows the result of experiment no. 49, a Uranyl Nitrate soak. Diffraction extends to the edge, corresponding to a resolution of 4.6Å. Despite the weakness of reflections in Figure 3.3, it is possible to compare intensities with the native *h0l* zone. Some changes are apparent, *e.g.* the row of relections 4 0 3, 4 0 4, 4 0 5 and 4 0 6. For the uranyl soaked crystal these reflections are all of similar intensity whereas for the native it can be seen that 4 0 4 is very weak compared to the others. Also, for the native 1 0 16 and 1 0 17 are both quite strong but in Figure 3.3

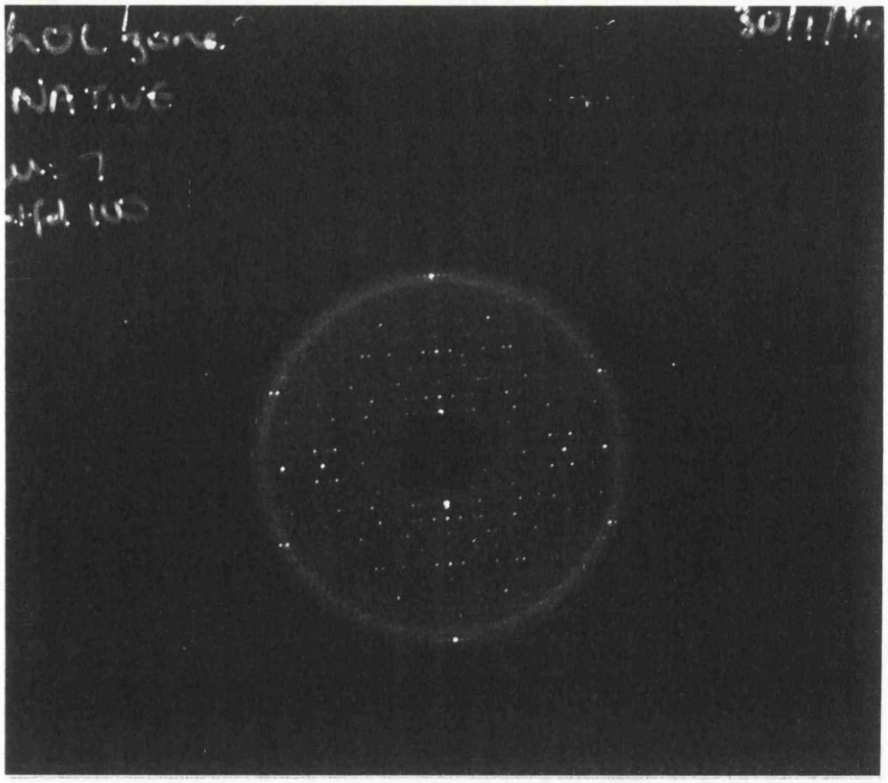


Figure 3.2: Native μ=7° precession; *h0l* zone. Crystal to film distance=100mm.

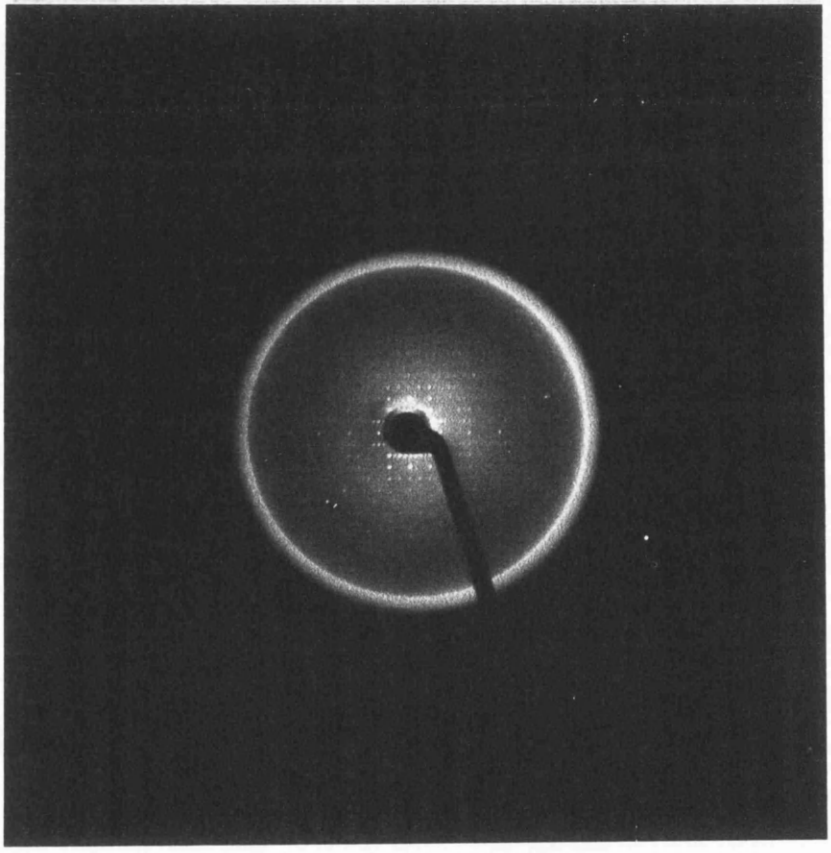


Figure 3.3: 3hr. 2.5mM $UO_2(NO_3)_2$ soaked hCG μ=10° precession; *h0l* zone. Crystal to film distance=100mm

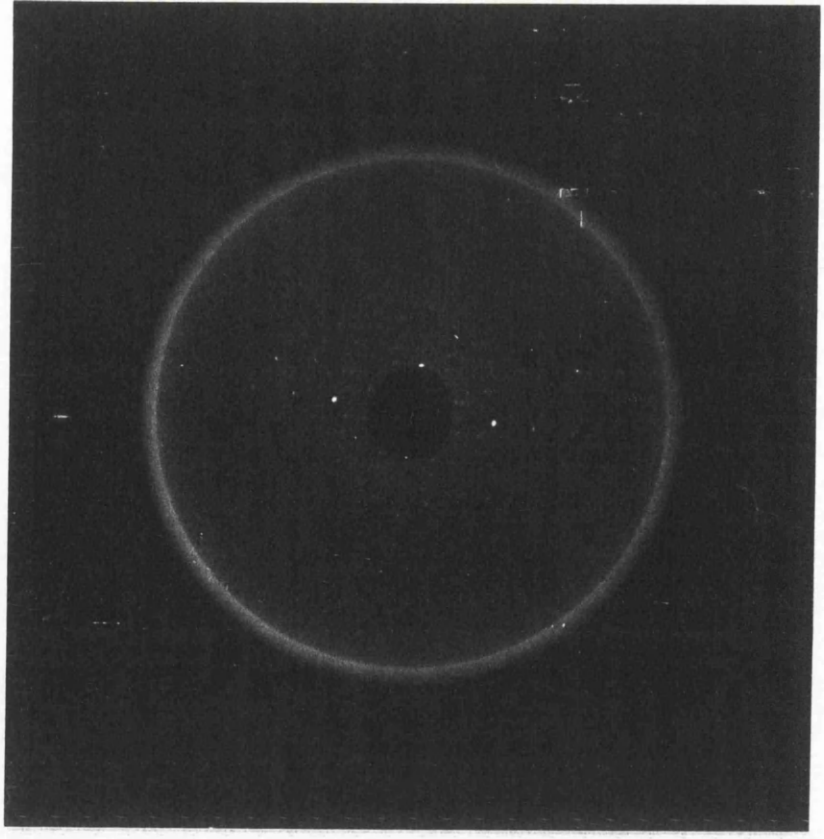


Figure 3.4: 4Hr. 4mM $NaAuCl_4$ soaked hCG μ=10°; *h0l* zone. Crystal to film distance = 100mm

1 0 17 is much weaker than 1 0 16.

Figure 3.4 shows the result of experiment no. 26, a $NaAuCl_4$ soak. Diffraction is weak, but there are some obvious intensity changes, *e.g.* inversion of the relative intensities of the reflections 0 0 12 and 0 0 18 with respect to the native. Reflection 5 0 3 is absent on the native photograph, but is present for the soaked crystal. The precession photograph from a four hour, 2.5mM K_2PtCl_4 soak, experiment no. 40, is presented in Figure 3.5. The weakness of this photograph prevents analysis of the higher resolution reflections but changes are apparent for relections 3 0 3 and 3 0 4. These are clearly visible and of similar intensity to 3 0 5 in the native but absent for the platinum soaked crystal. As seen with the gold soak, the relative intensities of the 0 0 12 and 0 0 18 reflections are reversed when compared to native. The cell dimensions as measured from these films using the *STOE* device are given in Table 3.8, along with measurements made for other derivatives. The cell dimensions were measured over as many reflections as possible to increase the accuracy.

No.	Heavy Atom Cmpd.	Method	Lifetime in Beam (Hrs)	Comments
1	$Hg(OOCCH_3)_2$	Precession	7	Initially good diffraction
2	$Hg(OOCCH_3)_2$	Precession	4	Not aligned
3	$Hg(OOCCH_3)_2$	Precession	5.5	Poor diffraction. Not aligned
4	$Hg(OOCCH_3)_2$	Precession	-	No diffraction
5	$Hg(OOCCH_3)_2$	Precession	24	Good diffraction
6	$Hg(OOCCH_3)_2$	Precession	18	Good diffraction
7	$Hg(OOCCH_3)_2$	Oscillation		Data set collected
8	$Hg(OOCCH_3)_2$	Precession	16	3° precession
9	Thiomersal	Precession	-	No diffraction
11	Thiomersal	Precession	1	V. weak diffraction
12	$K_2Hg(SCN)_4$	Precession	-	No diffraction
14	Me_2Hg	Oscillation		Data set collected
15	Me_2Hg	Precession	15	Good diffraction. 3° precession
17	$NaAuCl_4$	Precession	30	7° precession
18	$NaAuCl_4$	Precession	10	3° precession
19	$NaAuCl_4$	Rotation		Weak diffraction. 19° collected
21	$NaAuCl_4$	Precession	8.5	Diffraction. Not aligned
22	$NaAuCl_4$	Precession	8	Not aligned
23	$NaAuCl_4$	Precession	13	Not Aligned
24	$NaAuCl_4$	Precession	10.5	3° precession
25	$NaAuCl_4$	Precession	-	No diffraction
26	$NaAuCl_4$	Precession	29	10° precession
27	$KAu(CN)_2$	-	-	Crystal too cracked for use
29	$KAu(CN)_2$	Rotation		Data set collected

Table 3.5: Heavy Atom Screening Experiments - Mercury and Gold

No.	Heavy Atom Cmpd.	Method	Lifetime in Beam (Hrs)	Comments
30	K_2PtCl_4	Precession	7	Reasonable diffraction
32	K_2PtCl_4	Precession	30	9° precession
33	K_2PtCl_4	Rotation		Weak diffraction
34	K_2PtCl_4	Precession	24	Not aligned
35	K_2PtCl_4	Rotation		Data set collected. Weak
36	K_2PtCl_4	Precession	-	No diffraction
37	$K_2Pt(SCN)_4$	Precession	-	No diffraction
38	$K_2Pt(SCN)_4$	Precession	-	No diffraction
39	$K_2Pt(SCN)_4$	Precession	-	No diffraction
41	$K_2Pt(CN)_4$	Precession	11	Good diffraction. Not aligned
42	$K_2Pt(CN)_4$	Precession	6	Reasonable diffraction
43	$K_2Pt(CN)_4$	Precession	9	3° precession
44	$K_2Pt(CN)_4$	Oscillation		Data set collected
45	$K_2Pt(CN)_4$	Rotation		Data set collected
47	$K_2Pt(NO_2)_4$	Precession	-	No diffraction
48	$K_2Pt(NO_2)_4$	Precession	-	No diffraction

Table 3.6: Heavy Atom Screening Experiments - Platinum

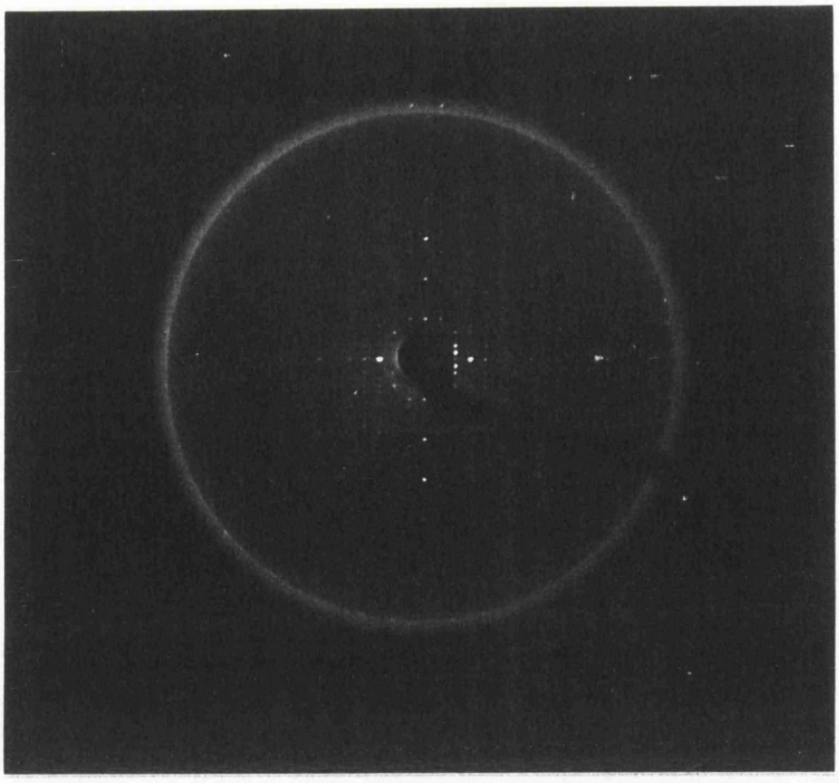


Figure 3.5: 2.5mM K_2PtCl_4 soaked hCG; *h0l* zone. Crystal to film distance 75mm

No.	Heavy Atom Cmpd.	Method	Lifetime in Beam (Hrs)	Comments
49	$UO_2(NO_3)_2$	Precession	22	10° precession
50	$UO_2(NO_3)_2$	Rotation		Diffraction to 5Å
51	$UO_2(NO_3)_2$	Rotation		Good diffraction. Detector down
52	$K_3Ir(CN)_6$	Precession	26	7° precession
53	$K_3Ir(CN)_6$	Precession	32	3° precession
54	$K_3Ir(CN)_6$	Precession	7	Weak diffraction. Not aligned
57	$K_3Ir(CN)_6$	Precession	12	Good diffraction. Not aligned
58	$Lu_2(SO_4)_3$	Rotation	-	Dataset collected
59	$Gd(OAc)_3$	Rotation	-	Dataset collected

Table 3.7: Heavy Atom Screening Experiments cont.

No.	Derivative	a (Å)	c (Å)
7	$Hg(OOCCH_3)_2$	88.9	174.0
23	$NaAuCl_4$	89.4	173.9
30	$UO_2(NO_3)_2$	89.0	174.3
36	$K_3Ir(CN)_6$	86.1	176.1
40	K_2PtCl_4	87.6	175.4

Table 3.8: Cell Dimensions of Derivatives measured from Precession photographs

Derivative	Soak Conditions			a	c
	Concn. (mM)	Duration (Hrs)	Backsoak (Hrs)	(Å)	(Å)
Me_2Hg	Sat.	-	-	88.21	176.70
$K_2Pt(CN)_4$	10	24	1	88.54	177.69
$NaAuCl_4$	4	4	-	89.84	175.72
$Hg(OOCCH_3)_2$	5	24	4	88.10	176.50
$UO_2(NO_3)_2$	2.5	24	-	88.78	177.10
K_2PtCl_4	2.5	4	-	Crystals	Died

Table 3.9: Putative derivatives collected at S.R.S., Daresbury

3.4.2 Derivatives collected at S.R.S., Daresbury

The results of the precession experiments, though not definitive, indicated the best putative derivatives for data collection. Three trips were made to S.R.S., Daresbury, Station PX7.2, to collect derivative data. Table 3.9 gives details of the soaks taken and the cell dimensions as determined by the auto-indexing programs in the MOSCO suite. The procedure for data collection and processing are given in the next Chapter.

3.4.3 Screening using Xentronics Area Detector

The screening of derivatives was made much easier after the acquisition of a Xentronics Area Detector. Low resolution data sets could be collected in a relatively short period, data processed, and scaled to a native set. Thus it was possible to obtain accurate cell dimensions and check for non-isomorphism. Table 3.10 and Tables 3.5 and 3.6 give details of heavy atom compounds screened in this way. The cell dimensions as determined by the auto-indexing program, *IDXREF*, are presented in Table 3.11. Data collection and processing methods are discussed in the next Chapter.

No.	Heavy Atom Cmpd.	Method	Comments
61	$Hg(OOCCH_3)_2$	Rotation	6Å data set collected
62	Mercurochrome	Rotation	Poor diffraction
63	Mercurochrome	Rotation	Poor diffraction
65	$[Pt(NH_3)_4](NO_3)_2$	Rotation	Very weak diffraction
66	$[Pt(NH_3)_4](NO_3)_2$	Rotation	Weak data set collected
67	$KAu(CN)_2$	Rotation	Good diffraction. 6Å data set collected
68	$AgNO_3$	Rotation	Died after first frame
69	$AgNO_3$	Rotation	No diffracion
70	$AgNO_3$	Rotation	6Å data set collected
71	Na_2IrCl_6	Rotation	Good diffraction, but crystal twinned
72	$K_2Pd(Br)_4$	Rotation	V. weak diffraction
73	$Gd(OAc)_3$	Rotation	Poor diffraction
74	$Gd(OAc)_3$	Rotation	Poor diffraction

Table 3.10: Heavy Atom Screening Experiments

No.	Derivative	a (Å)	c (Å)
24	$NaAuCl_4$	89.80	175.10
29	$KAu(CN)_2$	88.90	178.60
31	$UO_2(NO_3)_2$	88.96	176.80
55	$K_2Pt(CN)_4$	88.43	176.95
58	$Lu_2(SO_4)_3$	87.88	173.06
59	$Gd(OAc)_3$	89.20	179.20
59	$Hg(OOCCH_3)_2$	88.37	177.30
61	Mercurochrome	89.92	174.81
65	$KAu(CN)_2$	88.86	177.07
68	$AgNO_3$	89.01	176.89

Table 3.11: Cell Dimensions as determined by *IDXREF*

Chapter 4

Data Collection and Processing

4.1 Introduction

Data were collected using three different methods during the course of this project. Initially, the only option was to collect data sets on film at station PX 7.2., S.R.S., Daresbury. Several data sets were obtained in this way, but the limited amount of time available to us meant that collection was not truly optimized. HF-hCG crystals died more quickly in the Synchrotron beam than in a beam generated from a rotating anode, and so several crystals were needed for each dataset. Stills were taken prior to each data collection to align the crystals along c^* and to ensure minimum overlap of data.

The second option to become available was the Siemens area detector, installed in our laboratory in September 1990. This proved to be a far more efficient method of derivative screening and data collection. Far fewer crystals were used, no travel was involved, no scanning of films neccessary and data processing was straightforward. One limitation though, was the extent of diffraction on the rotating anode. The maximum observed diffraction was to 3.2Å resolution, compared to 2.5Å resolution at Daresbury. The Xentronics data may be the most convenient, and less demanding on the crystals, but any high resolution structure determination would require a Synchrotron source.

With this in mind, time was applied for on Station PX9.6 at the S.R.S., Daresbury. A total of three trips were made to use the FAST detector on this station. On two occasions the Synchrotron beam was down, and on the third the beam was unstable but useable. However, the crystals did not diffract as well as had been hoped. Radiation damage was severe, with crystals lasting a very short period of time, even when compared to lifetimes experienced on Station PX7.2. Approximately 30° of data were collected and processed by Dr. A. Littlejohn using the program *MADNES*. The dataset was found to be extremely weak and incomplete. Comparison with native datasets collected using other methods revealed the FAST data to be non-isomorphous. As the data was poor, of medium resolution only, and incomplete it was not used further.

In this Chapter I describe in turn the different methods used for data collection and processing, and the results obtained.

4.2 Film Data

4.2.1 Collection

Three independent trips were made to the S.E.R.C. Synchrotron source at Daresbury Laboratory to make use of the Synchrotron station PX 7.2. The radiation produced by the synchroton is of shorter wavelength than that available from standard laboratory generators and approximately a hundred fold more intense. Because of this, shorter exposure times can be used and so, in many cases, more data can be collected from each crystal before radiation damage becomes too great. Data were collected on film. A total of three films were placed in each cassette so that intensities of reflections saturated on film A could be measured from films B and C. In this way both intensities of higher and lower resolution data can be collected simultaneously. The beam for station PX 7.2 has a wavelength of 1.488Å, so oscillation ranges for the data were chosen accordingly.

For the native data sets, a crystal to film distance of 74mm was used with an oscillation range of 1° so that data were collected to the limit of the crystals' diffraction.

For the derivative sets, a crystal to film distance of 100mm and an oscillation range of 1.5° allowed data collection to 3Å resolution. On later visits, this range was increased to 1.8° as, when processing the first films, it was found that there was very little overlap of reflections. An increased range makes for more efficient data collection and processing.

4.2.2 Processing

All data sets collected on film at the S.E.R.C., Daresbury were processed using the programs written by A.J. Wonacott and F.J. Korber. Firstly, to get the data into digitized form, the films were scanned on the rotating drum densitometer kindly made available to us by G.G. Dodson at York University. All films were scanned on a 50μ raster.

In processing the data the first step is to determine the crystal orientation. There are two methods of achieving this.
i) Taking stills 90° apart and measuring the mis-set angles. These can then be refined, along with the crystal to film distance, camera parameters and the unit cell dimensions using the program *IDXREF*.
ii) Using the autoindexing programs *CENFIX* and *IMREFIX*, which take the coordinates of reflections on a range of photographs of known oscillation ranges, along with details of the cell dimensions, wavelength, crystal to film distance, and calculates the orientation matrix without any prior knowledge of the mis-set angles required. It was this last method that was mainly used for processing, though *IDXREF* was occasionally used to check certain parameters.

Once the orientation matrix is known the program *GMP* is run to create *"GENERT"* files. This calculates a complete list of reflections that will be observed for a given rotation range to a given resolution limit. GMP was useful for visually checking the

results of the autoindexing programs. There is a plot option which allows the predicted and the measured diffraction pattern to de displayed simultaneously on a Tektronix terminal.

The next stage is to determine the integrated intensity at a given position on the diffraction image. This was achieved using *MOSBA1* and *MOSBA2*. The former is an interactive program which allows final fine adjustment of the predicted pattern. Basically these programs carry out the following steps:

i) Locate the fiducial and film marks.

ii) Search for spots to refine the film/scanner transformation and correct for film distortion.

iii) Measure optical densities for all reflections on the film.

iv) Process the optical densities for all the spots to determine the integrated intensities and standard deviations.

4.2.3 Reduction and Scaling

Once *"GENERT"* files containing integrated intensities are obtained for all film packs, the next step is to merge together the reflections on each film within the pack and apply absorption corrections. This is done using *ABSCALE* which performs the following operations:

i) Applies oblique incidence absorption corrections.

ii) Calculates scale factors between films in the pack.

iii) Averages the scaled measurements, and analyzes the discrepancies from the mean.

iv) Applies Lorentz, polarisation and transmission corrections and writes the corrected intensities to an *MTZ* file.

v) Analyzes the agreement between symmetry related reflections within each film pack. It is at this stage that any mis-indexing of the data becomes apparent.

Individual *MTZ* files for each data set were merged and sorted using *SORTMTZ*, a program in the *CCP4* suite. The programs *ROTAVATA* and *AGROVATA* were then used for interpack scaling. The first program calculates the scale factors and temperature factors between overlapping batches of reflections. (Fox and Holmes (1966)). Scale factors are calculated relative to a chosen filmpack which is given a scale factor of 1.0. The latter program applies the scale factor, adds together partially recorded reflections, monitors and rejects bad reflections between repeated measurements or symmetry equivalents and averages them for output to an *MTZ* file.

4.2.4 Native Data

A total of four crystals were used to collect a native data set. The crystal to film distance was fixed at 74mm. With a 200 μm collimator the exposure time was 7.5 s/deg with 10 oscillations per degree. Approximately one carousel of data was collected from each crystal before radiation damage was too great. With an oscillation range of 1° degree, a maximum of 8° of data were collected per crystal. The native data set consisted of reflections measured from 26 filmpacks. The intrapack scaling resulted in data with R_{symm} values in the range 4.2% - 6.3%. A total of 16934 measurements including 4605 partials were used to obtain 7002 independent measurements with an overall R_{merge} of 8.0% to 3.0Å resolution. Diffraction extended to 2.5Å resolution with data 30% complete to this limit. However, the data were very weak beyond 3Å resolution and were not considered of high enough quality to be used. Details are given in Table 4.1. 76.9% of reflections have $I > 3\sigma(I)$.

4.2.5 Derivative Data Sets

In this Chapter datasets referred to as "derivative datasets" are those that are collected from crystals soaked in heavy atom solutions, rather than native crystals. The determination of the heavy atom positions is discussed in Chapter 5.

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	$\overline{I}/\sigma(\overline{I})$
9.45	435	210	66.4	2.1	0.041	0.041	16.5
6.70	1042	447	85.1	2.3	0.037	0.039	15.9
5.47	1459	584	88.0	2.5	0.049	0.043	13.0
4.74	1744	673	87.5	2.6	0.055	0.048	11.4
4.24	1879	713	82.7	2.6	0.059	0.051	11.3
3.87	2079	780	82.9	2.7	0.083	0.057	7.6
3.58	2198	869	85.5	2.5	0.119	0.063	5.7
3.35	1737	827	77.6	2.1	0.160	0.078	4.1
3.16	1592	837	71.9	1.9	0.244	0.084	2.7
3.00	1451	806	67.5	1.8	0.304	0.089	2.0
Total	15616	6746	79.9	2.4	-	0.083	7.8

Table 4.1: Interpack Scaling For HF-hCG Native Set

Dmin - Resolution limit of shell

Nmeas - Number of measurements within resolution shell used for data reduction

Nuniq - Number of unique reflections within resolution shell

% Poss. - Percentage of unique data after merging

Mult - Multiplicity of reflections

$$R_{fac} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_{h,j} >}$$

R_{cum} - R_{fac} up to this resolution range

4.2.6 Dimethyl mercury

Crystals were mounted in a glass capillary and a piece of filter paper soaked in Dimethyl Mercury placed in the tube which was quickly sealed. The data were collected approximately twenty four hours later. Two crystals were needed to collect twelve filmpacks of data. With a 1.5° degree oscillation range, 18° of data were collected. A total of 4831 reflections including 486 summed partials were merged with an overall R_{merge} of 10.0% to 3.5Å resolution. This gave a total of 3051 independent reflections, with 561 of these having anomalous measurements. 88% of the unique reflections have $I > 3\sigma$, but the anomalous data is very poor, with only 10.5% being recorded better than 3σ. Table 4.2 gives details of this data set.

4.2.7 $K_2Pt(CN)_4$ Derivative

The soaking conditions for this derivative are as for Experiment no. 44 in Table 3.2. Three crystals were used to collect 27 filmpacks. An oscillation range of 1.5° was employed. Crystals were aligned to maximize the percentage of data present by the use of stills. Diffraction extended to 3.1Å resolution, but data beyond 3.3Å resolution was very weak and so a 3.3Å cutoff was applied. Merging of 8030 reflections, including 324 partially recorded, gave an R_{merge} of 7.5% to 3.3Å resolution. The data set consists of 3100 independent reflections comprising 539 centric and 2561 acentric. Anomalous measurements were recorded for 321 reflections, but with only 5% having anomalous differences greater than 3σ.

4.2.8 $Hg(OOCCH_3)_2$ Derivative

Two crystals, soaked as described for Experiment no. 7, Table 3.1, were used to collect the $Hg(OOCCH_3)_2$ Derivative data set. An oscillation range of 1.5° degrees was used and reflections from thirteen filmpacks were included in the final data set. 3954

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	R_{anom}	$\overline{I}/\sigma(\overline{I})$
11.01	188	125	59.0	1.5	0.072	0.072	0.031	8.7
7.81	328	218	62.7	1.5	0.087	0.082	0.045	6.8
6.38	418	264	63.5	1.6	0.087	0.083	0.067	5.8
5.53	468	301	62.3	1.6	0.092	0.085	0.079	7.2
4.95	535	340	63.2	1.6	0.123	0.092	0.088	3.9
4.52	623	383	66.4	1.6	0.092	0.092	0.095	6.7
4.18	610	379	61.0	1.6	0.118	0.096	0.114	5.0
3.91	623	368	57.0	1.7	0.120	0.098	0.135	5.6
3.69	558	351	50.0	1.6	0.111	0.099	0.143	5.8
3.50	480	322	43.9	1.5	0.117	0.100	0.177	5.5
Total	4831	3051	57.6	1.6	-	0.100	0.103	5.5

Table 4.2: Interpack Scaling For Me_2Hg Derivative Set

Dmin - Resolution limit of shell

Nmeas - Number of measurements within the resolution shell used for data reduction

Nuniq - Number of unique reflections within the resolution shell

% Poss. - Percentage of unique data after merging

Mult - Multiplicity of reflections

$$R_{fac} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_{h,j} >}$$

R_{cum} - R_{fac} up to this resolution range

$$R_{anom} = \frac{\sum |< I+ > - < I- >|}{\sum (< I+ > + < I- >)}$$

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	R_{anom}	$\overline{I}/\sigma(\overline{I})$
10.38	461	173	72.9	2.7	0.048	0.048	0.048	12.1
7.36	818	287	73.5	2.9	0.050	0.050	0.041	12.0
6.02	943	330	67.4	2.9	0.062	0.054	0.045	10.5
5.21	1007	365	64.0	2.8	0.075	0.059	0.069	9.0
4.66	1213	426	67.1	2.8	0.073	0.063	0.047	9.3
4.26	1224	439	63.7	2.8	0.088	0.068	0.059	7.3
3.94	1022	413	55.0	2.5	0.094	0.072	0.073	6.9
3.69	677	321	40.5	2.1	0.107	0.074	0.049	5.7
3.48	425	211	24.9	2.0	0.104	0.074	0.097	6.6
3.30	240	135	15.0	1.8	0.133	0.075	0.075	5.0
Total	8030	3100	48.7	2.6	-	0.075	0.057	8.4

Table 4.3: Interpack Scaling For $K_2Pt(CN)_4$ Derivative Set

Dmin - Resolution limit of shell

Nmeas - Number of measurements within a resolution shell used for data reduction

Nuniq - Number of unique reflections within a resolution shell

% Poss. - Percentage of unique data after merging

Mult - Multiplicity of reflections

$$R_{fac} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_{h,j} >}$$

R_{cum} - R_{fac} up to this resolution range

$$R_{anom} = \frac{\sum |< I+ > - < I- >|}{\sum (< I+ > + < I- >)}$$

measurements, including 116 partials merged to give 2579 independent reflections with an R_{merge} of 8.6%. Table 4.4 gives details of the data set with respect to resolution.

4.2.9 $NaAuCl_4$ Derivative

Experiment no. 19 in Table 3.1 describes the soak conditions used for this derivative. Six crystals were taken for data collection, but due to poor diffraction data collected from only three crystals were finally used, with a maximum of four filmpacks collected from one crystal. In total ten filmpacks were used. After a resolution cutoff of 3.9Å was applied 4849 measurements were merged to give 2271 independent reflections with R_{merge} =10.7%. 82.1% of reflections are measured with $I > 3\sigma$. Table 4.5 gives further details.

4.2.10 $UO_2(NO_3)_2$ Derivative

Four crystals were soaked in a 2.5mM solution of the uranium compound as described in Experiment no. 49, Table 3.2. Diffraction was to high resolution, but only three filmpacks could be collected per crystal before radiation damage became too severe. 8475 measurements from the twelve filmpacks were merged, R_{merge}=10.8%, to produce a data set containing 3977 independent reflections with 1074 anomalous measurements. Relevent statistics are given in Table 4.6. As can be seen, a resolution cutoff of 3.5Å was applied; reflections were observed to 3.3Å, but were excluded from the data set as they were too weak and the data too incomplete.

4.2.11 K_2PtCl_4 Derivative

Crystals were soaked as described in Table 3.2, Experiment no. 35 and showed a good uptake of the red colour of the heavy atom solution. However, the crystals were fragile

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	R_{anom}	$\overline{I}/\sigma(\overline{I})$
10.38	253	165	66.6	1.5	0.046	0.046	0.034	16.1
7.36	419	271	66.8	1.5	0.055	0.051	0.053	10.8
6.02	420	267	54.0	1.6	0.075	0.057	0.061	8.3
5.21	492	306	53.0	1.6	0.069	0.060	0.073	9.5
4.66	601	356	55.7	1.7	0.071	0.064	0.067	9.5
4.26	575	374	53.7	1.5	0.101	0.071	0.071	6.0
3.94	487	321	43.8	1.5	0.146	0.079	0.083	2.7
3.69	379	265	33.5	1.4	0.118	0.081	0.103	4.7
3.48	220	166	19.6	1.3	0.134	0.083	0.129	4.3
3.30	108	88	10.3	1.2	0.163	0.083	0.094	3.7
Total	3954	2579	40.2	1.5	-	0.083	0.070	6.0

Table 4.4: Interpack Scaling For $Hg(OOCCH_3)_2$ Derivative Set

Dmin - Resolution limit of shell

Nmeas - Number of measurements within a resolution shell used for data reduction

Nuniq - Number of unique reflections within a resolution shell

% Poss. - Percentage of unique data after merging

Mult - Multiplicity of reflections

$$R_{fac} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_{h,j} >}$$

R_{cum} - R_{fac} up to this resolution range

$$R_{anom} = \frac{\sum |< I+ > - < I- >|}{\sum (< I+ > + < I- >)}$$

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	R_{anom}	$\overline{I}/\sigma(\overline{I})$
12.25	129	63	43.5	2.0	0.042	0.042	0.047	10.1
8.69	288	146	59.8	2.0	0.049	0.046	0.045	11.1
7.11	414	192	63.6	2.2	0.065	0.054	0.061	7.8
6.16	502	206	59.2	2.4	0.083	0.062	0.050	6.9
5.51	522	235	59.6	2.2	0.112	0.068	0.083	5.7
5.03	608	267	60.6	2.3	0.109	0.075	0.082	5.9
4.66	610	279	59.4	2.2	0.120	0.082	0.081	4.7
4.36	629	288	57.8	2.2	0.156	0.091	0.104	3.8
4.11	608	303	56.7	2.0	0.185	0.099	0.135	3.3
3.90	539	292	52.0	1.8	0.232	0.107	0.130	2.4
Total	4849	2271	57.7	2.1	-	0.107	0.078	5.3

Table 4.5: Interpack Scaling For $NaAuCl_4$ Derivative Set

Dmin - Resolution limit of shell

Nmeas - Number of measurements within a resolution shell used for data reduction

Nuniq - Number of unique reflections within a resolution shell

% Poss. - Percentage of unique data after merging

Mult - Multiplicity of reflections

$$R_{fac} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_{h,j} >}$$

R_{cum} - R_{fac} up to this resolution range

$$R_{anom} = \frac{\sum |< I+ > - < I- >|}{\sum (< I+ > + < I- >)}$$

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	R_{anom}	$\overline{I}/\sigma(\overline{I})$
11.01	240	124	62.6	1.9	0.045	0.045	0.036	10.8
7.81	535	265	81.5	2.0	0.058	0.054	0.039	8.1
6.38	713	328	80.1	2.2	0.063	0.057	0.044	9.4
5.53	839	357	74.7	2.4	0.092	0.065	0.055	6.9
4.95	961	409	76.2	2.3	0.092	0.070	0.074	6.6
4.52	1005	460	77.1	2.2	0.109	0.077	0.073	5.7
4.18	1098	495	76.9	2.2	0.155	0.085	0.118	4.1
3.91	1121	527	76.8	2.1	0.221	0.095	0.167	2.9
3.69	1021	510	70.5	2.0	0.262	0.102	0.219	2.5
3.50	942	502	66.3	1.9	0.323	0.108	0.256	1.8
Total	8475	3977	74.2	2.1	-	0.108	0.085	5.3

Table 4.6: Interpack Scaling For $UO_2(NO_3)_2$ Derivative Set

Dmin - Resolution limit of shell

Nmeas - Number of measurements within a resolution shell used for data reduction

Nuniq - Number of unique reflections within a resolution shell

% Poss. - Percentage of unique data after merging

Mult - Multiplicity of reflections

$$R_{fac} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_{h,j} >}$$

R_{cum} - R_{fac} up to this resolution range

$$R_{anom} = \frac{\sum |< I+ > - < I- >|}{\sum (< I+ > + < I- >)}$$

and did not travel well to Daresbury. Three crystals were exposed in the beam but diffraction was found to be extremely poor and it was not possible to collect data.

4.3 Xentronics Data

The preferred method of screening for derivatives was the collection of a low resolution data set on the laboratory Siemens area detector. Data were processed, scaled relative to the Xentronics data set and difference Pattersons calculated. If the Patterson map could be solved, the data collection step was repeated and data measured to the limit of diffraction for the heavy atom soaked crystal. This was an efficient method of screening as there was no need to align the crystal accurately (cf. the precession method). Also the crystals lasted far longer in the X-ray beam produced by the laboratory source compared to their lifetime in the Synchrotron beam so fewer crystals were necessary per data set and problems with merging data from more than one crystal were minimized.

4.3.1 Procedure for collection.

If the detector is to be successfully used for collection of X-ray diffraction data then the following criteria must be met:

i) Uniformity of sensitivity across the face of the detector. This is attained by application of the flood field correction. For this an ^{55}Fe source is used. The whole detector face is exposed for a time dependent upon the distance of the detector from the source. An on-line, photon-by- photon correction is applied.

To check this has been successful, the calibrated detector is exposed for a shorter period of time (300s). The image produced should be uniform across the image.

ii) Well defined spatial distortion. It is important that an accurate conversion between real space (cm) and detector space (pixels) is available. This is known as the brassplate correction. A brassplate with precisely drilled holes of well defined position is fixed over the detector face which is then irradiated using the ^{55}Fe source. Again the exposure

time is dependent in the detector distance. The recorded pixel positions are used to generate a lookup table for relating pixels to cm and vice versa. This correction is applied during parameter refinement and integration and thus its precision can be increased by reference to the diffraction data.

Before data collection is begun, it is also neccessary to determine the direct beam position on the detector face. A well centred small molecule crystal is used. This is rotated about an axis perpendicular to the beam direction for 360°. The direct beam position is then determined by inspection of symmetry equivalent reflections on the detector image.

4.3.2 Strategy of Data Collection using HF-hCG crystals.

HF-hCG crystallizes in the space group $P6_122$ or enantiomer, with cell dimensions a = b = 88.7Å and c = 177.2Å. There are two possible strategies for efficient data collection. Crystals could be mounted about c^*. For a complete data set this would involve collection of 30° of data, plus a small amount about a^* to fill in the blind region. The four-circle goniometer allows the accurate alignment of the crystals, as χ is not fixed as in a three-circle system. The inherent problem with this approach is the distance at which the detector must be set. It is primarily the spatial resolution of the detector which determines this distance. Howard *et al* (1987) gave an approximate empirical relationship defining the shortest possible crystal to detector distance, D

$$D \text{ (in cm)} = A_{max}/8$$

where A_{max} is the largest effective unit cell length in Å. Obviously other factors such as the mosaic spread of the crystal also have an effect.

Thus, as c^* = 177.2Å, a c^* mount dictates a distance of at least 220mm. At this position the exposure times required for each scan are long, due to absorption of the diffracted rays by the air. Use of a helium tube was considered, but at this distance absorption

by the mylar window counteracts the positive effect of the decreased absorption by the helium and little, if anything, is gained.

The second strategy is to mount the crystal about a^* . This would allow the detector to be positioned only 11cm from the crystal and thus a shorter exposure time per frame is possible. However, for a complete data set, 90 ° of data would be required. Approximately 75° could be collected at this distance before c^* came into the diffraction plane and reflection overlap occurred. At this point, the detector would have to be set back to 22cm, a new brassplate measured and the final 15° of data collected. Another problem with this strategy is it assumes a fairly accurate alignment about a^*. Depending on crystal morphology this could not always be guaranteed.

A third option, commonly employed by other Xentronics users, is a totally random mount and collection of 180° of data. This was not considered efficient for our crystals, as a distance of 22cm would still be required and therefore the long exposure times.

The strategy decided upon for data collection from the HF-hCG crystals was the first mentioned, the c^* mount. An added benefit of this is that equivalent reflections are measured on or about the same frame and so the accuracy of any measurement of anomalous data was maximized. In practise up to 60° of data were collected about this mount, as the *XENGEN* reduction programs rely heavily on the redundancy in the data. Better consistency in the scaling parameters is achieved for data sets with an average degree of redundancy of three or more (Derewenda *et al.* (1989)). Table 4.7 lists the exposure times for the flood fields and brassplate for the distances used in data collection.

For heavy atom screening, the crystals were soaked as described in Chapter 3. The mounted crystals were centred in the beam, and then aligned as near as possible along c^*. This usually resulted in a nonzero value of χ which, due to limitations in the software prohibited the use of a ϕ scan, and so a $-\omega$ scan was used. At a distance of 220mm and a 2θ angle of 0° the maximum resolution achievable was 5.5Å. This was considered sufficient for analysis of the putative derivative.

Distance (mm)	Flood field (Hrs.)	Brassplate (Hrs.)
130	0.75	1.00
220	2.00	3.50
250	3.00	5.00

Table 4.7: Exposure times for Flood fields and Brassplate

For the native set and previously screened derivatives, the 2θ angle was set to allow collection of data to the limit of diffraction of the crystal. If necessary another set (2θ = 0°) was also collected to fill in any gaps in the low resolution shells. The blind data were collected by a ϕ scan about a^* .

4.4 Data Processing of Xentronics Data using *XENGEN*

The program suite *XENGEN* (Howard *et al* (1987)) was used to process all the data sets.

The processing procedure is well documented in the Siemens *XENTRONICS* manual. Briefly, the detector face is calibrated and its active area defined. The orientation matrix of the crystal is obtained using the auto-indexing program, *REFINE*. This requires a number of spot centroids which are generated by *SPOTS* using contiguous frames. The user can specify the sigma value over which spots are accepted as useable. For hCG data sets, this was of the order 20-30σ. Spots are merged if they appear on more than one frame, and the three dimensional spot profiles are examined and a more precise value of the centroid is calculated. For *REFINE* to find good difference vectors for determination of the orientation matrix, the frames used must cover a reasonable rotation range. 15 to 20° were commonly used for our processing.

Once an acceptable orientation matrix is found, *REFINE* can be used to refine certain

parameters. It can individually refine the unit cell dimensions, crystal orientation angles, detector position and the rocking curve behaviour of the crystal. For processing HF-hCG data sets, once an orientation matrix was accepted, the cell dimensions were re-set to those expected for the native crystal. All parameters were fixed except the orientation angles which were refined linearly until changes were minimized. The parameters *XCEN* and *YCEN* which contain errors in the direct beam position were then refined in conjunction with the orientation angles. The detector parameters were released sequentially. Finally the cell dimensions were included for least squares refinement. The progress of refinement could be monitored by changes in the RMS error values for X, Y, Phi and index. For an acceptable solution, RMS errors in X and Y should be less than 0.5 of a pixel, PHI less than 0.3 of a frame and the index error less than 0.05.

Reflections are integrated using the program *INTEGRATE*. Further refinement of parameters is also performed. *INTEGRATE* performs three-dimensional profile fitting. Corrections are applied for Lorentz and polarization effects.

4.4.1 Data Reduction and Scaling

The program *REDUCE* creates a *MULTIREF* file which contains information on the crystal parameters and information on the data run. For each reflection there is a record containing h,k,l values, $\sin\theta/\lambda$ and the number of observations measured. The symmetry related observations are grouped together. *SCALEI* performs least squares minimization of any differences among intensities of symmetry related observations contained in the merged file. Two scale factors are calculated for each 5° of data; one for the top half of the detector and the other for the bottom half. For each batch the scale factor is defined as

$$K = g + As + Bs^2$$

where g, A and B are the parameters varied during refinement and s is sin θ/λ. Mean intensities are calculated from multiple measurements of equivalent reflections. The process is iterative and alternates between calculating mean intensities and calculating scaling parameters. This program is run in conjunction with *REJECT* which can be used to remove bad observations from the calculations. Data reduction and scaling was completed for HF-hCG using a scheme recommended by Dr. E. Garman. During all stages for all data sets, the Bijvoet pairs were treated as independent. The initial scale factors were determined using all data and one parameter (g) per shift. *REJECT* was used to remove very bad observations i.e those greater than 8*DEV where DEV is

$$DEV(i)= (I_i-(I))/\sigma(i)$$

I_i is the observed intensity for observation i with a standard deviation of sigma(i) and (I) is the calculated mean intensity. New scaling parameters were then calculated on these observations with one and then two parameters per shift. *REJECT* was re-run to remove those observations with deviations greater than five. *SCALEI* was run again on this smaller set with two parameters per shift and a final rejection cycle using the default cutoff value which is dependent on the number of observations present.

4.4.2 Processing native data

The native dataset processed using the *XENGEN* package was the result of a single scan from one crystal corresponding to crystal 1, scan 1 described in Section 4.5.1. 3033 independent reflections were merged from 6552 observations. Table 4.8 gives details. The second scan was also processed and and described in Table 4.8. This dataset was of a lower resolution and of poorer quality, presumably due to radiation damage. The exposure times were identical for the two scan runs (20 min/frame) but the $\overline{I}/\sigma(\overline{I})$ values are decreased for the second scan run, especially at the higher resolution range. In order to maximize the completeness of the native dataset, data from the two scan runs were merged together. As can be seen from Table 4.8, the data merged very poorly

as evidenced by the high R_{merge} values. However, it was this more complete, merged dataset that was initially used for the calculation of low resolution difference Patterson maps. The platinum derivative was solved using this native dataset, but no solution was discovered for the silver derivative. The native dataset obtained from the single first scan was then used. Scaling of the *XENGEN* processed derivatives to this resulted in lower mean isomorphous differences and an interpretable difference Patterson map for the Silver derivative. In the next Chapter, the datset referred to as Natxeng is the incomplete set collected on the first scan.

4.4.3 $AgNO_3$ Derivative data

A single crystal was used for the screening experiment no. 68 in Chapter 3. 4804 observations were merged to give 1219 unique reflections. Details are given in Table 4.9. As this was a screening experiment, the data were collected with 2θ = 358° and short exposure times. This resulted in reflections measured to the edge of the detector being weakly recorded which resulted in the poor R_{merge} values. However, the use of relatively short exposure times permitted the screening of a putative derivative by collection of a dataset of this quality in an overnight experiment. As discussed in the next Chapter, the data were sufficiently well measured to allow solution of a difference Patterson. A further $AgNO_3$ dataset was collected for higher resolution data. The soak was conducted in Sodium Citrate buffer, pH 4.4. The data shown in Table 4.9 gives the merging statistics for this dataset which contains 4214 independent reflections merged with an overall R_{merge} of 13.8% from 18196 measurements.

4.4.4 $K_2Pt(CN)_4$ Derivative Data

A dataset was collected from a single crystal soaked for 24 hours in 10mM $K_2Pt(CN)_4$ solution made up in Sodium Acetate buffer, pH 4.4. The data shown in Table 4.10 was obtained in a single scan and consists of 3686 reflections. Beyond 3.6Å resolution the

First Scan run

Dmin (Å)	Nmeas	Nuniq	% Poss	R_{merge}	$\overline{I}/\sigma(\overline{I})$
6.20	1646	818	73.2	0.047	28.7
4.92	1394	638	62.6	0.077	16.2
4.30	1231	537	53.2	0.081	15.0
3.90	1036	453	45.6	0.121	9.2
3.62	837	374	37.9	0.161	5.8
3.41	408	213	22.4	0.184	4.1
Total	6552	3033	49.9	0.080	16.2

2nd Scan run

Dmin (Å)	Nmeas	Nuniq	% Poss	R_{merge}	$\overline{I}/\sigma(\overline{I})$
7.27	1820	685	95.6	0.039	40.1
5.77	1709	628	95.7	0.075	20.2
5.04	1442	561	89.0	0.085	16.0
4.58	1251	518	83.1	0.083	16.8
4.25	955	452	74.0	0.099	11.5
4.00	384	229	36.8	0.165	5.9
Total	7561	3073	79.6	0.068	20.9

Merging of the two scan runs

Dmin (Å)	Nmeas	Nuniq	% Poss	R_{merge}	$\overline{I}/\sigma(\overline{I})$
6.23	5595	1083	98.1	0.051	47.4
4.94	4654	920	92.4	0.091	23.5
4.32	3736	833	89.0	0.102	20.4
3.92	1703	564	58.0	0.157	9.5
3.64	813	371	38.5	0.293	3.2
3.43	395	209	22.0	0.471	1.6
Total	16896	3980	66.6	0.081	24.3

Table 4.8: *XENGEN* processed Native data

Dmin - Resolution limit of shell

Nmeas - Number of measurements within a resolution shell used for data reduction

Nuniq - Number of unique reflections within a resolution shell

% Poss. - Percentage of unique data after merging

$$R_{merge} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_h >}$$

Crystal 1 - Screening Experiment

Dmin (Å)	Nmeas	Nuniq	% Poss	R_{merge}	$\overline{I}/\sigma(\overline{I})$
10.10	1198	220	75.8	0.046	50.3
8.01	1257	230	88.8	0.075	31.3
7.00	937	228	91.2	0.107	17.3
6.36	656	216	90.7	0.136	11.3
5.90	573	201	81.3	0.231	6.3
5.56	183	124	52.3	0.250	3.6
Total	4804	1219	80.1	0.077	21.6

Crystal 2 - Soak in Sodium Citrate, pH 4.4

Dmin (Å)	Nmeas	Nuniq	% Poss	R_{merge}	$\overline{I}/\sigma(\overline{I})$
6.32	5328	1006	95.2	0.078	27.9
5.01	4208	875	89.7	0.138	12.0
4.38	3406	814	86.1	0.154	10.5
3.98	2657	696	74.5	0.196	6.6
3.70	1863	554	59.8	0.285	3.6
3.48	734	269	29.1	0.316	1.9
Total	18196	4214	73.1	13.8	12.8

Table 4.9: Merging statistics for *XENGEN* processed $AgNO_3$ Derivative Data

Dmin - Resolution limit of shell

Nmeas - Number of measurements within a resolution shell used for data reduction

Nuniq - Number of unique reflections within a resolution shell

% Poss. - Percentage of unique data after merging

$$R_{merge} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_h >}$$

Dmin (Å)	Nmeas	Nuniq	% Poss	R_{merge}	$\overline{I}/\sigma(\overline{I})$
6.19	2636	1077	81.2	0.056	28.9
4.92	2351	765	75.7	0.084	17.1
4.30	1984	656	65.2	0.101	14.8
3.90	1730	592	60.0	0.195	6.6
3.62	1392	491	50.0	0.506	1.1
3.41	671	275	28.8	0.495	0.98
Total	10764	3686	60.9	0.101	14.6

Table 4.10: Merging statistics for *XENGEN* processed $K_2Pt(CN)_4$ Derivative Data

Dmin - Resolution limit of shell

Nmeas - Number of measurements within a reoslution shell used for data reduction

Nuniq - Number of unique reflections within a resolution shell

% Poss. - Percentage of unique data after merging

$$R_{merge} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_h >}$$

data is very poor. The reason for the dramatic decrease in the quality of the dataset at this resolution was unclear. When the same intensities were reprocessed using *XDS* this problem was not observed (Table 4.17).

4.4.5 $KAu(CN)_2$ Derivative Data

Two distinct datasets were collected using this derivative. The first was a low resolution screening set described in Chapter 3, experiment no. 67. The dataset consists of 1248 reflections merged from 3796 observations. The second was an attempt to collect a higher resolution dataset from a crystal soaked in a gold solution at pH 4.4. As can be seen from Table 4.11 the dataset is weak beyond 4Å resolution. To 3.2Å 10410

measurements were merged to give 4034 unique reflections.

4.4.6 $Hg(OOCCH_3)_2$ Derivative Data

A single crystal was used to collect the dataset detailed in Table 4.12. The soaking conditions are described in Chapter 3, experiment no. 61. 1566 independent reflections were merged from 5400 measurements with an overall R_{merge} of 4.6%.

4.4.7 $[Pt(NH_3)_4](NO_3)_2$ Derivative Data

A dataset was collected from a crystal soaked in this derivative (experiment no. 66). Diffraction extended to 3.4Å resolution but data beyond 4Å was very weak. To 4Å 2751 independent reflections were merged from 8218 observations with an overall R_{merge} of 6.3%. Details are given in Table 4.13.

4.5 Processing of Xentronics Data using *XDS*

The *XDS* package written by Dr. W. Kabsch was used to process all data collected on the Xentronics, including that previously processed using the *XENGEN* package. *XDS* is a collection of programs which allow processing with minimal interaction by the user. An input data file is created giving data collection details. A calibration is performed using the brass plate in a manner similar to that in *XENGEN*. The program *COLSPOT* collects a number of strong reflections from a user-defined number of frames and writes these to *SPOT*.xds. These are used by the auto-indexing program *IDXREF* (Kabsch, W. (1988)) to determine the orientation matrix for the crystal. *COLPROF* then collects a series of three-dimensional reflection profiles for the whole scan range and these are used by *PROFIT* to estimate the reflection intensities. *CORRECT* then makes corrections for decay, absorption and detector surface sensitivity. As these

Crystal 1 - low resolution screening experiment

Dmin (Å)	Nmeas	Nuniq	% Poss	R_{merge}	$\overline{I}/\sigma(\overline{I})$
10.07	933	267	90.6	0.062	47.0
7.99	994	249	96.1	0.079	31.6
6.98	764	234	92.1	0.085	19.7
6.34	518	206	85.1	0.127	11.8
5.89	445	188	76.4	0.178	7.5
5.54	142	104	43.7	0.209	5.5
Total	3796	1248	81.3	0.081	23.6

Crystal 2 - higher resolution set

Dmin (Å)	Nmeas	Nuniq	% Poss	R_{merge}	$\overline{I}/\sigma(\overline{I})$
5.70	2585	955	67.9	0.071	16.9
4.53	2250	840	65.0	0.120	8.7
3.95	1878	760	59.7	0.162	4.4
3.59	1701	683	54.5	0.331	1.4
3.34	1338	523	42.2	0.456	0.5
3.20	658	273	31.5	0.582	0.1
Total	10410	4034	54.0	0.125	6.7

Table 4.11: Merging statistics for *XENGEN* processed $KAu(CN)_2$ Derivative Data

Dmin - Resolution limit of shell

Nmeas - Number of measurements within a resolution shell used for data reduction

Nuniq - Number of unique reflections within a resolution shell

% Poss. - Percentage of unique data after merging

$$R_{merge} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_h >}$$

Dmin (Å)	Nmeas	Nuniq	% Poss	R_{merge}	$\overline{I}/\sigma(\overline{I})$
9.42	1558	311	88.8	0.036	47.0
7.48	1333	360	99.0	0.044	30.5
6.54	896	293	97.0	0.061	16.2
5.95	815	286	95.3	0.075	11.4
5.51	615	247	87.9	0.085	9.8
5.19	183	157	53.2	0.091	7.8
Total	5400	1598	87.1	0.046	22.3

Table 4.12: Merging statistics for *XENGEN* processed $Hg(OOCCH_3)_2$ Derivative Data

Dmin - Resolution limit of shell

Nmeas - Number of measurements within a resolution shell used for data reduction

Nuniq - Number of unique reflections within a resolution shell

% Poss. - Percentage of unique data after merging

$$R_{merge} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_h >}$$

Dmin (Å)	Nmeas	Nuniq	% Poss	R_{merge}	$\bar{I}/\sigma(\bar{I})$
7.27	1610	576	80.5	0.050	34.1
5.76	1663	534	81.6	0.076	19.1
5.04	1444	472	75.4	0.087	16.2
4.58	1250	419	67.6	0.086	17.2
4.25	1157	379	62.8	0.110	13.2
4.00	1094	371	60.1	0.182	7.7
Total	8218	2751	71.7	0.063	19.1

Table 4.13: Merging statistics for *XENGEN* processed $[Pt(NH_3)_4](NO_3)_2$ Derivative Data

Dmin - Resolution limit of shell

Nmeas - Number of measurements used for data reduction

% Poss. - Percentage of unique data after merging

$$R_{merge} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_h >}$$

profiles are learned for all the strong data the program is able to compensate for crystal slippage. Finally *GLOREF* uses all the observed reflections and performs a global refinement of the collection parameters. All data sets were processed in P_1, and re-indexed to $P6_122$ after integration and profile fitting. The *XDS*.hkl files containing the unmerged data were converted into *MTZ* files using the program *ROTAVATPREP* and merged using the *CCP4* suite programs *ROTAVATA/AGROVATA*. A measurement was rejected if its standard deviation was three or more times from the mean value for that reflection. These mean intensities were then converted to mean amplitudes by the program *TRUNCATE* and truncated using a procedure based on Bayesian statistics (French, G.S. and Wilson, K.S. (1978)). Amplitudes are calculated using the prior knowledge of Wilson's distributions for acentric or centric data and the mean intensity and standard deviation values. The F's output are all positive and follow Wilson's distribution (Wilson (1949)). The truncation procedure has little effect on reflections with $I > 3\sigma(I)$, but significantly improves the weaker data. Wilson scaling is applied so the output data is on an absolute scale.

4.5.1 Processing of native data

Two crystals were used to collect native data on the Xentronics. Crystal 1. was mounted on c^* and crystal 2. was an approximate a^* mount. The procedure for data collection was

XTL1a : xtdd=250mm, 2θ=14°, 0.3° ω scan, ω_{init}=345°, 123 frames.
XTL1b : xtdd=250mm, 2θ=10°, 0.3° ϕ scan, ϕ_{init}=0°, 300 frames.
XTL2a : xtdd=130mm, 2θ=8°, 0.3° ϕ scan, ϕ_{init}=230°, 320 frames.
XTL2b : xtdd=130mm, 2θ=8°, 0.3° ϕ scan, ϕ_{init}=5°, 320 frames.

Table 4.14 and Table 4.15 give details of the results of merging reflections within scan runs for crystals 1 and 2. The dataset collected from cystal 1 can be compared to the dataset obtained from the same data and processed using itXENGEN. The *XDS*

processed dataset is of superior quality, but is less complete at higher resolution *i.e.* more reflections have been rejected. The dataset collected from Scan 1. consists of 2784 independent reflections with an overall R_{merge} of 5.3%. As this dataset was incomplete, a second crystal was used to collect the datasets described in Table 4.15. As this set was considered complete, the data from the two crystals was not merged.

4.5.2 $AgNO_3$ Derivatives

Two Silver Nitrate soaked HF-hCG data sets were collected. The first was a low resolution set collected in screening experiment no. 68. from a crystal soaked for 2 hours in 10mM $AgNO_3$ in 30mM Sodium Citrate Buffer, pH 6.3. The second was a similar soak, but the pH of the Citrate Buffer was 4.4. In both cases crystals were approximately aligned along c^*. Two scans were made for each data set -

Crystal 1a : xtdd=220mm, 2θ=358°, -0.3° ω scan, ω_{init}=10°, χ=0°, 106 frames.
Crystal 1b : xtdd=220mm, 2θ=358°, -0.3° ω scan, ω_{init}=10°, χ=90°, 80 frames.
Crystal 2a : xtdd=220mm, 2θ=348°, -0.3° ω scan, ω_{init}=20°, χ=255°, 150 frames.
Crystal 2b : xtdd=220mm, 2θ=348°, -0.3° ω scan, ω_{init}=20°, χ=345°, 71 frames.

Data for the two scan runs were merged for each crystal. Table 4.16 gives details of the datasets with respect to resolution. 1096 independent reflections were obtained from the first crystal with an overall R_{merge} of 5.4% for 4523 used measurements. The second set of data consists of 3664 independent reflections merged with R_{merge} of 9.7% from 11660 measurements. 2352 reflections have anomalous measurements.

4.5.3 $K_2Pt(CN)_4$ Derivative

The dataset described in Table 4.17 was collected from a single scan using one crystal

Scan 1

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	$\overline{I}/\sigma(\overline{I})$
11.01	263	150	71.9	1.8	0.016	0.016	40.4
7.81	542	279	83.2	1.9	0.021	0.019	33.5
6.38	641	301	76.0	2.1	0.033	0.023	20.3
5.53	726	332	70.9	2.2	0.046	0.029	16.0
4.95	742	341	65.3	2.2	0.051	0.033	14.8
4.52	721	330	57.0	2.2	0.054	0.037	14.2
4.18	718	324	51.5	2.2	0.075	0.042	9.7
3.91	681	304	45.4	2.2	0.113	0.048	6.8
3.69	533	265	26.5	2.0	0.152	0.052	5.0
3.50	158	158	15.8	1.8	0.132	0.053	5.3
Total	5850	2784	52.5	2.1	-	0.053	13.7

Scan 2

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	$\overline{I}/\sigma(\overline{I})$
11.01	678	196	88.2	3.5	0.031	0.031	20.1
7.81	1449	362	100.0	4.0	0.037	0.035	16.0
6.38	1691	427	97.4	4.0	0.051	0.039	13.0
5.53	1780	476	95.0	3.7	0.070	0.045	10.3
4.95	1728	493	88.8	3.5	0.073	0.049	10.1
4.52	1547	498	82.4	3.1	0.071	0.053	10.2
4.18	856	350	54.0	2.4	0.088	0.055	8.4
Total	9768	2802	84.4	3.5	-	0.055	13.5

Table 4.14: *XDS* processed Native Data - Crystal 1 merging statistics

Dmin - Resolution limit of shell

Nmeas - Number of measurements within a resolution shell used for data reduction

Nuniq - Number of unique reflections within a resolution shell

% Poss. - Percentage of unique data after merging

Mult - Multiplicity of reflections

$$R_{fac} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_{h,j} >}$$

R_{cum} - R_{fac} up to this resolution range

Scan 1

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	$\overline{I}/\sigma(\overline{I})$
11.01	1333	200	90.7	6.7	0.034	0.034	18.6
7.81	2730	347	98.4	7.9	0.041	0.039	15.4
6.38	2672	418	96.3	6.4	0.056	0.043	12.2
5.53	2526	478	95.3	5.3	0.070	0.047	10.5
4.95	2577	501	90.2	5.1	0.067	0.050	11.0
4.52	2712	541	89.4	5.0	0.067	0.053	11.0
4.18	2824	580	88.3	4.9	0.089	0.058	8.6
3.91	2843	614	88.2	4.6	0.128	0.064	5.9
3.69	2812	631	85.8	4.5	0.177	0.070	4.4
3.50	2765	659	85.4	4.2	0.235	0.077	3.4
Total	25794	4969	89.7	5.2	-	0.077	9.4

Scan 2

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	$\overline{I}/\sigma(\overline{I})$
11.01	1328	182	84.8	7.3	0.033	0.033	17.4
7.81	2712	320	91.9	8.5	0.041	0.038	15.1
6.38	2695	391	90.8	6.9	0.052	0.041	13.3
5.53	2541	467	93.3	5.4	0.071	0.046	10.5
4.95	2567	519	93.4	4.9	0.070	0.049	10.6
4.52	2715	575	94.5	4.7	0.069	0.053	10.6
4.18	2805	625	94.6	4.5	0.097	0.058	7.8
3.91	2859	666	95.0	4.3	0.142	0.064	5.5
3.69	2809	700	94.4	4.0	0.204	0.070	3.8
3.50	2802	741	95.4	3.8	0.277	0.077	2.8
Total	25833	5186	93.7	5.0	-	0.077	9.1

Table 4.15: *XDS* processed Native Data - Crystal 2 merging statistics

$$R_{fac} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_{h,j} >}$$

R_{cum} - R_{fac} up to this resolution range

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	R_{anom}	$\overline{I}/\sigma(\overline{I})$
18.07	165	37	66.2	4.5	0.022	0.022	0.026	27.4
12.88	410	72	85.9	5.7	0.033	0.029	0.022	21.6
10.55	619	98	94.1	6.3	0.037	0.033	0.024	17.6
9.15	624	106	92.2	5.9	0.052	0.038	0.033	12.6
8.19	665	121	89.9	5.5	0.057	0.041	0.036	12.5
7.48	564	137	98.3	4.1	0.073	0.044	0.045	9.7
6.93	466	143	93.6	3.3	0.096	0.047	0.059	7.5
6.48	475	150	94.0	3.2	0.138	0.050	0.077	5.5
6.11	393	144	83.6	2.7	0.169	0.053	0.101	4.3
5.80	142	88	50.5	1.6	0.176	0.054	0.226	4.0
Total	4523	1096	84.7	4.1	-	0.054	0.045	12.8
Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	R_{anom}	$\overline{I}/\sigma(\overline{I})$
11.01	577	186	86.1	3.1	0.034	0.034	0.033	10.2
7.81	1134	316	92.8	3.6	0.039	0.037	0.033	18.2
6.38	1389	379	89.0	3.7	0.057	0.044	0.047	12.5
5.53	1520	439	88.4	3.5	0.091	0.054	0.061	8.1
4.95	1508	484	88.3	3.1	0.096	0.062	0.064	7.5
4.52	1513	500	83.4	3.0	0.098	0.069	0.066	7.2
4.18	1495	491	75.0	3.0	0.128	0.077	0.078	5.7
3.91	1357	446	64.0	3.0	0.202	0.087	0.131	3.6
3.69	1022	345	46.5	3.0	0.278	0.096	0.178	2.6
3.50	145	78	9.9	1.9	0.332	0.097	0.237	2.3
Total	11660	3664	66.0	3.2	-	0.097	0.073	6.8

Table 4.16: *XDS* processed $AgNO_3$ Derivative Data - merging statistics

A) Crystal 1 - Soak in 30mM Sodium Citrate buffer, pH 6.3

B) Crystal 2 - Soak in 30mM Sodium Citrate buffer, pH 4.4

$$R_{fac} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_{h,j} >}$$

R_{cum} - R_{fac} up to this resolution range

$$R_{anom} - \frac{\sum |< I+ > - < I- >|}{\sum (< I+ > + < I- >)}$$

xtdd=250mm, 2θ=346°, -0.3° ω scan, ω_{init}=30°, 200 frames.

9453 measurements were merged to produce 3169 independent reflections with an overall R_{merge} of 6.0%.

4.5.4 $KAu(CN)_2$ Derivative Sets

Two datasets were collected for this derivative. The first described is a low resolution screening experiment (No. 67, Table 3.4) in which a crystal was soaked in the gold solution made up in Sodium Citrate buffer, pH 6.3. Three scans were merged to give the data set shown in Table 4.18 -

Crystal 1a : xtdd=220mm, 2θ=358°, -0.3° ω scan, ω_{init}=22°, 113 frames.
Crystal 1b : xtdd=220mm, 2θ=358°, -0.3° ω scan, ω_{init}=5°, 32 frames.
Crystal 1c : xtdd=220mm, 2θ=351°, -0.3° ω scan, ω_{init}=20°, 43 frames.

Scan 1b is a continuation of scan 1a which was interrupted due to problems with the cooling system. This recurred again after the 32 frames in the second scan. The data were processed and found to be a derivative so the crystal was replaced on the detector in the hope of collecting higher resolution data. However as diffraction only extended to 5.5Å resolution the crystal was removed after a short period. Table 4.18 gives details of this data set. 1282 independent reflections were obtained with an R_{merge} of 4.7% using 4483 observations. The second experiment was an attempt at collecting a higher resolution dataset. The crystal used was grown in Sodium Acetate buffer, pH 4.2. To minimize disruption to the crystal the gold soak was conducted in this buffer. Again the crystal was mounted approximately along c^*;

Crystal 2a : xtdd=220mm, 2θ=345°, -0.3° ω scan, ω_{init}=20°, χ=90°, 105 frames.
Crystal 2b : xtdd=220mm, 2θ=345°, -0.3° ω scan, ω_{init}=5°, χ=0°, 50 frames.

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	R_{anom}	$\overline{I}/\sigma(\overline{I})$
11.01	403	148	69.1	2.7	0.020	0.020	0.039	33.4
7.81	887	296	85.8	3.0	0.023	0.022	0.040	31.2
6.38	1083	350	81.9	3.1	0.041	0.028	0.044	17.7
5.53	1184	381	76.6	3.1	0.054	0.034	0.059	14.1
4.95	1205	391	70.7	3.1	0.055	0.038	0.052	13.7
4.52	1206	389	64.5	3.1	0.062	0.043	0.050	11.9
4.18	1191	394	59.8	3.0	0.086	0.048	0.064	8.8
3.91	1110	378	54.3	2.9	0.131	0.054	0.098	5.9
3.69	882	314	42.6	2.8	0.158	0.059	0.120	4.8
3.50	302	128	16.1	2.4	0.120	0.060	0.175	6.2
Total	9453	3169	56.9	3.0	-	0.060	0.061	12.1

Table 4.17: *XDS* processed $K_2Pt(CN)_4$ Derivative Data - merging statistics

Dmin - Resolution limit of shell

Nmeas - Number of measurements within resolution shell used for data reduction

Nuniq - Number of unique reflections within a resolution shell

% Poss. - Percentage of unique data after merging

Mult - Multiplicity of reflections

$$R_{fac} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_{h,j} >}$$

R_{cum} - R_{fac} up to this resolution range

$$R_{anom} = \frac{\sum |< I+ > - < I- >|}{\sum (< I+ > + < I- >)}$$

The dataset described in Table 4.18 consists of 3285 independents with an overall R_{merge} of 9.1%.

4.5.5 $Hg(OOCCH_3)_2$ Derivative Data

Two separate datasets were collected for this derivative. The first is detailed in Chapter 3, experiment no. 61. The crystal was mounted about c^* and two scan ranges collected. The second dataset was collected from a crystal soaked in 5mM $Hg(OOCCH_3)_2$ in Sodium Acetate buffer, pH 4.3, for 24 hours. The crystal was mounted without backsoaking and aligned along a^* . Initially diffraction was good, but the crystal died overnight and, during processing the data was found to be quite weak.

Crystal 1a : xtdd=220mm, 2θ=357°, -0.3° ω scan, ω_{init}=21°, ϕ=93°, 150 frames.
Crystal 1b : xtdd=220mm, 2θ=357°, -0.3° ω scan, ω_{init}=10°, ϕ=0°, 40 frames.
Crystal 2a : xtdd=130mm, 2θ=355°, 0.3° ϕ scan, ϕ_{init}=0°, 280 frames.

The first data set consists of 1374 independent reflections merged from 4631 measurements with overall R_{merge}= 3.5%. The second set is much weaker and incomplete, with R_{merge} of 10.0% for 3957 measurements of 1608 independents. Details for both are given in Table 4.19.

4.5.6 $[Pt(NH_3)_4](NO_3)_2$ Derivative Data

The data set described in Table 4.20 is the result of screening experiment no. 64 in Chapter 3. Diffraction was poor, but a data set to 5.8Å was collected with

Crystal 1a : xtdd=220mm, 2θ=358°, 0.3° ϕ scan, ϕ_{init}=0°, 225 frames.

The dataset consists of 952 unique reflections, see Table 4.20.

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	R_{anom}	$\overline{I}/\sigma(\overline{I})$
17.16	198	57	86.4	3.5	0.028	0.028	0.042	21.1
12.22	433	100	96.4	4.3	0.034	0.032	0.043	17.3
10.01	573	122	99.8	4.7	0.035	0.033	0.046	19.4
8.68	693	133	95.1	5.2	0.042	0.036	0.052	16.0
7.77	682	154	100.0	4.4	0.052	0.039	0.050	13.3
7.09	527	154	93.3	3.45	0.063	0.041	0.051	10.6
6.57	529	169	96.8	3.1	0.082	0.043	0.072	8.9
6.15	461	158	86.0	2.9	0.116	0.046	0.067	6.3
5.80	263	134	70.4	2.0	0.155	0.047	0.120	4.9
5.50	124	101	51.9	1.2	-	0.047	0.182	0.0
Total	4483	1282	84.9	3.5	-	0.047	0.053	14.2
Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	R_{anom}	$\overline{I}/\sigma(\overline{I})$
11.01	279	138	67.5	2.0	0.029	0.029	0.046	24.4
7.81	655	259	77.4	2.5	0.036	0.034	0.045	15.5
6.38	806	282	67.6	2.9	0.058	0.042	0.050	11.6
5.53	918	328	67.1	2.8	0.070	0.049	0.063	9.6
4.95	971	373	68.6	2.6	0.079	0.055	0.063	9.1
4.52	992	394	65.0	2.5	0.081	0.060	0.070	9.0
4.18	973	392	59.1	2.5	0.116	0.066	0.101	6.5
3.91	1002	389	55.1	2.6	0.193	0.075	0.153	4.0
3.69	939	377	50.3	2.5	0.252	0.083	0.180	3.0
3.50	897	353	44.7	2.5	0.317	0.091	0.245	2.4
Total	8432	3285	59.6	2.6	-	0.091	0.091	7.7

Table 4.18: *XDS* processed $KAu(CN)_2$ Derivative Data - merging statistics

A) Crystal 1 - Soak in Sodium Citrate buffer, pH 6.3

B) Crystal 2 - soak in Sodium Acetate buffer, pH 4.4

$$R_{fac} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_{h,j} >}$$

$$R_{anom} = \frac{\sum |< I+ > - < I- >|}{\sum (< I+ > + < I- >)}$$

R_{cum} - R_{fac} up to this resolution range

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	R_{anom}	$\overline{I}/\sigma(\overline{I})$
17.16	211	57	85.5	3.7	0.014	0.014	0.017	49.4
12.22	517	102	98.0	5.1	0.021	0.018	0.019	31.2
10.01	671	126	100.0	5.3	0.022	0.020	0.020	30.4
8.68	616	137	96.6	4.5	0.027	0.022	0.029	25.0
7.77	563	163	100.0	3.5	0.029	0.023	0.035	24.7
7.09	516	168	97.1	3.1	0.035	0.024	0.036	20.3
6.57	508	181	99.9	2.8	0.047	0.025	0.052	15.0
6.15	501	182	93.8	2.8	0.058	0.027	0.064	11.7
5.80	425	186	93.3	2.3	0.071	0.028	0.082	10.9
5.50	103	72	36.6	1.4	0.057	0.028	0.034	16.1
Total	4631	1374	88.2	3.4	-	0.028	0.035	24.5
Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	R_{anom}	$\overline{I}/\sigma(\overline{I})$
11.01	184	60	27.7	3.1	0.036	0.036	0.040	17.9
7.81	432	121	34.4	3.6	0.046	0.043	0.050	15.2
6.38	517	151	35.4	3.4	0.084	0.055	0.093	8.7
5.53	534	183	36.6	2.9	0.125	0.066	0.107	5.8
4.95	435	173	31.1	2.5	0.138	0.073	0.094	5.2
4.52	432	190	31.4	2.3	0.152	0.080	0.113	4.8
4.18	411	198	29.6	2.1	0.172	0.086	0.147	4.0
3.91	433	208	30.0	2.1	0.250	0.094	0.243	2.9
3.61	367	188	25.9	2.0	0.252	0.098	0.268	2.9
3.50	212	136	17.5	1.6	0.269	0.100	0.363	2.8
Total	3957	1608	29.0	2.5	-	0.100	0.134	7.0

Table 4.19: *XDS* processed $Hg(OOCCH_3)_2$ Derivative Data - merging statistics

A) Crystal 1 - Soak in Sodium Citrate buffer, pH 6.3

B) Crystal 2 - Soak in Sodium Acetate buffer, pH 4.2

$$R_{fac} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_{h,j} >}$$

$$R_{anom} = \frac{\sum |< I+ > - < I- >|}{\sum (< I+ > + < I- >)}$$

R_{cum} - R_{fac} up to this resolution range

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	$\bar{I}/\sigma(\bar{I})$
18.07	177	40	68.2	4.4	0.030	0.030	18.6
12.88	456	75	85.2	6.1	0.035	0.033	18.7
10.55	642	92	86.1	7.0	0.040	0.036	15.9
9.15	733	101	84.7	7.3	0.046	0.039	15.2
8.19	733	111	80.9	6.6	0.060	0.043	11.9
7.48	644	113	77.6	5.7	0.072	0.046	9.8
6.93	564	115	73.3	4.9	0.097	0.050	7.6
6.48	542	120	73.2	4.5	0.112	0.053	6.5
6.11	383	131	75.0	2.9	0.122	0.055	6.5
5.80	71	54	29.7	1.3	0.118	0.055	6.4
Total	4945	952	70.9	5.2	-	0.055	12.3

Table 4.20: *XDS* processed $[Pt(NH_3)_4](NO_3)_2$ Derivative Data - merging statistics

Dmin - Resolution limit of shell

Nmeas - Number of measurements within a resolution shell used for data reduction

Nuniq - Number of unique reflections within a resolution shell

% Poss. - Percentage of unique data after merging

Mult - Multiplicity of reflections

$$R_{fac} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_{h,j} >}$$

R_{cum} - R_{fac} up to this resolution range

4.5.7 K_2PtCl_4 Derivative Data

Experiment no. 42 from Chapter 3 resulted in the dataset described in Table 4.21. One crystal was used, mounted along c^* -

Crystal 1 : xtdd=220mm, 2θ=355°, 0.3° ϕ scan, ϕ_{init}=20°, 167 frames.

Similar problems were encountered with this soak as were found with the K_2PtCl_4 soaked crystals taken to Daresbury. Two crystals were soaked in a 2.5mM solution for four hours, mounted and exposed to X-rays. A strong colour uptake implied that heavy atoms were present in the crystal. However, diffraction in both cases was very poor with the crystals dying after a short period of time. The 2 hour soak produced a crystal also of more intense colour than the surrounding soak medium, and diffraction was relatively better. The crystal died after 170 frames, but it was possible to process the data, described in Table 4.21, integrating only over the first 100 frames. However it was apparent from the very high R_{fac} values that the data is extremely poor and the data was not used.

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	$\overline{I}/\sigma(\overline{I})$
13.32	255	98	77.9	2.6	0.057	0.057	11.1
9.46	337	175	87.0	1.9	0.185	0.069	3.8
7.74	402	231	94.0	1.7	0.397	0.075	1.8
6.71	425	255	90.0	1.7	0.852	0.081	0.7
6.00	389	261	83.4	1.5	0.684	0.086	1.1
Total	1808	1020	87.4	1.8	-	0.086	7.2

Table 4.21: *XDS* processed K_2PtCl_4 Derivative Data - merging statistics

Dmin - Resolution limit of shell

Nmeas - Number of measurements within a resolution shell used for data reduction

Nuniq - Number of unique reflections within a resolution shell

% Poss. - Percentage of unique data after merging

Mult - Multiplicity of reflections

$$R_{fac} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_{h,j} >}$$

R_{cum} - R_{fac} up to this resolution range

4.5.8 $Lu_2(SO_4)_3$ Derivative Data

Data were collected from a single crystal mounted along a^* -

Crystal 1 : xtdd=130mm, 2θ=354°, 0.15° ϕ scan, ϕ_{init}=90°, 330 frames.

The resulting dataset is shown in Table 4.22. 5818 measurements were merged (R_{merge}=10.8%) to give 1685 independent reflections.

4.5.9 $UO_2(NO_3)_2$ Derivative Data

A crystal was soaked in a 2.5mM solution of $UO_2(NO_3)_2$ made up in Sodium Acetate buffer, pH 4.2 for four hours. The crystal was aligned along a^* -

Crystal 1a : xtdd=130mm, 2θ=355°, 0.25° ϕ scan, ϕ_{init}=0°, 400 frames.

Diffraction appeared strong with reflections observed at 3.3Å. However problems were encountered during processing. The data were split into thirty frame batches for input to *ROTAVATA/AGROVATA*. Of the 400 frames collected only the first 180 frames were used to produce the dataset shown in Table 4.23 as beyond this the R_{merge} values became intolerably high. This is similar to the problems encountered while processing the uranium film data. Another dataset collected from a crystal soaked in exactly the same solution, but for 30 hours, was too weak to process.

4.6 Comparison of Native Datasets

During the course of this project several different native sets have been used. The first was that collected on the Xentronics area detector and processed using the *XENGEN* package. This was merged with the native set collected on film (Natf) to give a more

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	$\overline{I}/\sigma(\overline{I})$
12.85	236	67	48.6	3.5	0.039	0.039	16.7
8.92	506	131	59.1	3.9	0.052	0.048	13.8
7.29	657	170	60.3	3.9	0.072	0.056	10.4
6.32	754	197	61.5	3.8	0.103	0.066	7.0
5.65	702	209	57.7	3.4	0.146	0.075	4.9
5.16	577	194	49.2	3.0	0.126	0.081	5.9
4.78	633	186	43.7	3.4	0.124	0.087	5.9
4.47	573	171	37.6	3.4	0.140	0.092	5.4
4.22	595	182	37.4	3.3	0.182	0.099	4.0
4.00	585	178	46.7	3.5	0.246	0.108	2.8
Total	5818	1685	46.7	3.5	-	0.108	6.6

Table 4.22: *XDS* processed $Lu_2(SO_4)_3$ Derivative Data - merging statistics

Dmin - Resolution limit of shell

Nmeas - Number of measurements within a resolution shell used for data reduction

Nuniq - Number of unique reflections within a resolution shell

% Poss. - Percentage of unique data after merging

Mult - Multiplicity of reflections

$$R_{fac} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_{h,j} >}$$

R_{cum} - R_{fac} up to this resolution range

Dmin (Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{fac}	R_{cum}	R_{anom}	$\overline{I}/\sigma(\overline{I})$
11.01	841	105	48.0	8.0	0.045	0.045	0.028	15.0
7.81	1790	191	55.2	9.4	0.068	0.059	0.037	10.6
6.38	2332	244	56.6	9.6	0.137	0.080	0.052	5.5
5.53	2363	287	57.5	8.2	0.183	0.098	0.080	4.1
4.95	1913	318	57.1	6.0	0.157	0.107	0.094	4.7
4.52	1855	365	59.5	5.1	0.187	0.119	0.109	3.9
4.18	1803	392	59.2	4.6	0.225	0.130	0.145	3.3
3.91	1754	426	60.6	4.1	0.323	0.144	0.207	2.3
3.69	1596	456	61.8	3.5	0.393	0.155	0.295	1.9
3.50	701	326	42.1	2.2	0.349	0.157	0.397	1.9
Total	16948	3110	56.1	5.4	-	0.157	0.123	4.5

Table 4.23: *XDS* processed $UO_2(NO_3)_2$ Derivative Data - merging statistics

Dmin - Resolution limit of shell

Nmeas - Number of measurements within a resolution shell used for data reduction

Nuniq - Number of unique reflections within a resolution shell

% Poss. - Percentage of unique data after merging

Mult - Multiplicity of reflections

$$R_{fac} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_{h,j} >}$$

R_{cum} - R_{fac} up to this resolution range

$$R_{anom} = \frac{\sum |< I+ > - < I- >|}{\sum (< I+ > + < I- >)}$$

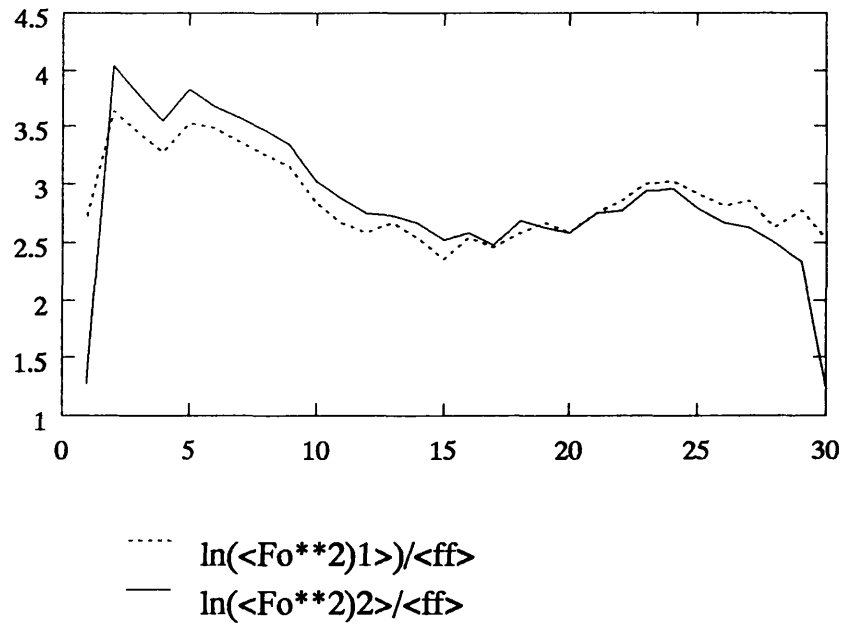


Figure 4.1: Wilson Plots for Natxeng native dataset

complete, higher resolution set for use in phase improvement. Table 4.24 gives the merging statistics for this combined native set. Later the set collected on the Xentronics and processed using *XDS* (that referred to as Crystal2,Scan1 in Table 4.25.) was used as this was more complete and from a single crystal. The effect of this new dataset can be seen from the results in Chapter 6. I shall refer to the first set as Natxeng, the second as Natcomb and the third as Natxds for the remainder of this thesis. The Wilson plots for the various sets are presented in Figures 4.1 - 4.3.

The theory of Wilson (Wilson (1949)) states that assuming the atoms are randomly distributed about the unit cell then

$$<ff>=scale<F_{obs}^2>exp(-2Bsin^2\theta/\lambda^2)$$

where <ff> is the mean squared atomic form factor summed over all atoms in the unit cell. Thus a plot of $ln<F_{obs}^2>/<ff>)$ v. $2sin^2\theta/\lambda^2$ allows the determination of the absolute scale and temperature factor of the data by fitting a least squares line through this plot.

Dmin (Å)	Nmeas	% Poss.	Mult.	R_{fac}	R_{cum}	I/sigI
9.45	514	89.8	1.7	0.049	0.049	12.1
6.70	1069	88.4	2.3	0.048	0.048	12.2
5.47	1426	90.3	2.4	0.057	0.052	11.1
4.74	1758	84.9	2.5	0.063	0.056	9.9
4.24	1662	85.7	2.3	0.069	0.059	8.9
3.87	1854	84.1	2.4	0.093	0.065	7.1
3.58	1962	86.1	2.3	0.133	0.072	5.1
3.35	1737	77.6	2.1	0.160	0.078	4.1
3.16	1592	71.9	1.9	0.244	0.084	2.7
3.00	1451	67.5	1.8	0.304	0.089	2.0
Total	14845	80.5	2.2	-	0.089	7.1

Table 4.24: Merging statistics for Natxeng with Natf

Dmin - Resolution limit of shell

Nmeas - Number of measurements used for data reduction

% Poss. - Percentage of unique data after merging

Mult - Multiplicity of reflections

$$R_{fac} = \frac{\sum_j \sum_h | I_{h,j} - < I_h > |}{\sum_j \sum_h < I_{h,j} >}$$

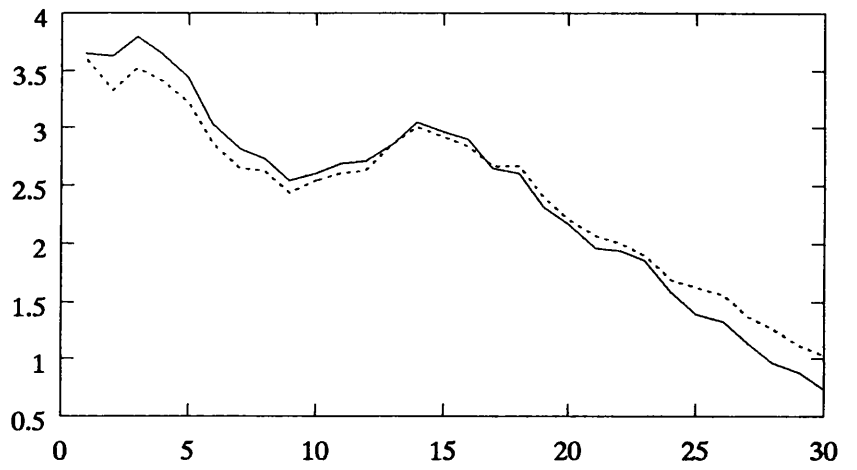


Figure 4.2: Wilson Plots for Natcomb native dataset

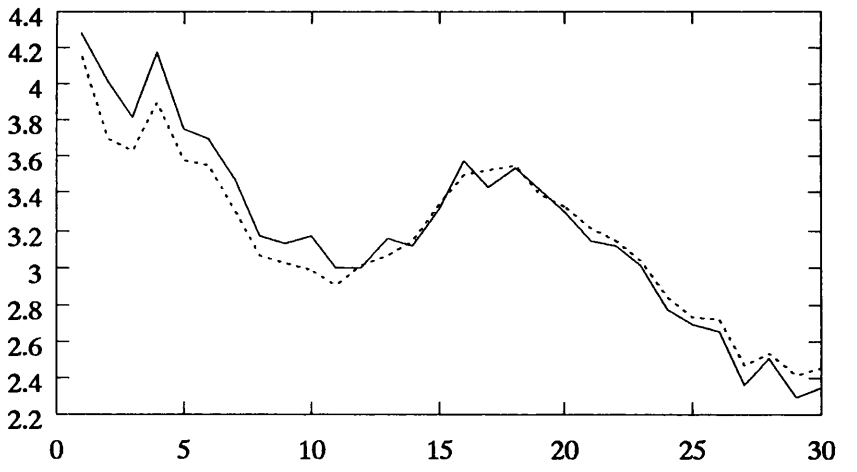


Figure 4.3: Wilson Plots for Natxds native dataset

I.D.	Source	Processing Package	Crystal Id.	Temperature Factor	Scale
Natf	Film	*MOSCO*	-	36.91	0.01680
Natxeng	Xentronics	*XENGEN*	Xtl1,S1	26.10	0.4038
Natxds	Xentronics	*XDS*	Xtl1,S1	33.36	0.00952
Natxds2	Xentronics	*XDS*	Xtl1,S2	23.82	0.04005
Natxds3	Xentronics	*XDS*	Xtl2,S1	28.28	0.02351
Natxds4	Xentronics	*XDS*	Xtl2,S2	35.49	0.02544
Natcomb	Natf+Natxeng	-	-	57.35	0.01469

Table 4.25: Absolute Scale and Temperature Factors for Native datasets

Two plots are presented for each dataset. That labelled ln(<(Fo**2)1>)/<ff> uses the average $<F_{obs}>$ derived from the number of observed reflections in a hemisphere of reciprocal space. If the reflections that were not measured were all weak then $<F_{obs}^2>$ is more accurately estimated if the average is derived using the number of possible reflections. This is represented in the figures by the plot labelled ln(<(Fo**2)2>)/<ff>. Table 4.25 gives the results of the Wilson scaling.

4.7 Summary

During the course of this project data were collected in several different ways. Sources included a Rigaku rotating anode and a Synchrotron beam at PX7.2 and PX9.6, S.R.S., Daresbury. Diffraction was observed to a higher resolution at the synchrotron, but radiation damage was more severe than witnessed on the laboratory source. The datasets collected on the Xentronics area detector were more complete than those collected on film at PX7.2 and it was these datasets that were primarily used to determine the phases. The Xentronics data were processed using both *XENGEN* and *XDS*. The datasets processed using *XDS* were of superior quality, particularly at higher resolution.

Chapter 5

Determination of Phases

This Chapter describes the use of the various datasets, described in the previous Chapter, to solve the phases for HF-hCG. The methods used and results obtained are discussed in chronological order to show how the quality of the phases improved as a number of procedures were tried. The first heavy atom derivatives were solved using the Xentronics data processed using *XENGEN*. These initial MIR phases were improved using solvent flattening and maximum entropy techniques. At this stage, the data were processed with *XDS* which was found to improve its quality, and all subsequent work was conducted using these datasets.

5.1 Scaling of Data

All data were anisotropically scaled using the program *SCALEIT*. A derivative to native scaling function is calculated using F^2 and then applied to F_{PH}. The scaling factor used is of the form

$$K = C * \exp(-2(h^2B_{11} + k^2B_{22} + l^2 B_{33} + 2hkB_{12} + 2hlB_{13} + 2kl\ B_{23}))$$

where B terms are anisotropic temperature factors. The initial scale factor is calculated using

$$K_{init} = \sqrt{(\frac{\sum F_P^2}{\sum F_{PH}^2})}$$

and then refined using a modification of the Fox and Holmes (1966) method. The difference between the measured intensity and the mean intensity is minimized for all measurements of all reflections. The function

$$\sum_h \sum_i w(h_i[I(h_i) - G_i I_h]^2)$$

is minimized with respect to all parameters

$$G1 = 1.0$$
$$G2 = (1/C)\exp(2\vec{h}B_h)$$

The degree of heavy atom substitution for each derivative is determined from the mean fractional difference in the structure factor amplitudes. However, this is also effected by non-isomorphism. The best method to evaluate substitution is to plot

$$\frac{\sum |F_P - F_{PH}|}{\sum F_P}$$

versus $\sin^2\theta/\lambda^2$. If there are no experimental errors then this plot should decrease with increasing resolution for the perfectly isomorphous case. In practise the plot increases with resolution because of random errors and lack of isomorphism arising from heavy atoms perturbing the native structure. However, it is still useful for evaluating derivatives as a dramatic increase in $\triangle$F/F at higher resolution indicates non-isomorphism. The absolute value of $\triangle$F/F can be misleading; large differences

do not always indicate a high degree of substitution and can instead be due to non-isomorphism or problems in scaling. Similarly a derivative with small values of $\Delta F/F$ may be isomorphous and still contain useful phase information.

Certain heavy metals will absorb X-rays and result in anomalous diffraction effects, manifested as a non-equivalence in reflections related by Friedel symmetry. This signal is generally small but can be measured if intensity measurements of Friedel partners are carefully collected. This anomalous data can also be used to determine the degree of substitution. These measurements usually have larger experimental errors but have the advantage that they are independent of the degree of isomorphism.

Data from derivative and native crystals were collected with film or with a Xentronics area detector. The derivative datasets collected from film or area detector were scaled with the corresponding native dataset. The film data were collected at Station PX7.2, S.R.S., Daresbury and were anisotropically scaled against the 3.0Å resolution film native dataset (Table 5.1). Plots of $\Delta F/F$ are given for all these derivatives in Figure 5.1. The greatest isomorphous differences are observed for the $UO_2(NO_3)_2$ and $NaAuCl_4$ derivatives. During data collection the crystals soaked with these compounds suffered much more severe radiation damage than witnessed with the other derivatives.

The first datasets to be collected on the Xentronics were scaled against Natxeng and the results are summarized in Table 5.2. Figure 5.2 shows the plot of $\Delta F/F$ v. resolution for these putative derivatives. The $K_2Pt(CN)_4$ derivative shows a large degree of isomorphism at higher resolution. This will reflect the poor quality of this data beyond 3.9Å resolution as processed using *XENGEN*. The results of anisotropic scaling of the potential derivatives to the native set Natxds are given in Table 5.3. Plots of $\Delta F/F$ v. resolution are given in Figure 5.3 and 5.4 for derivatives collected on the Xentronics and on film respectively. Due to the limited resolution of the data used, these plots are of limited use for determining the degree of non-isomorphism. The derivative mercury acetate was found to contain a single bound heavy atom, whereas no heavy atom binding in the dimethyl mercury derivative was discovered. In Figure 5.1 these derivatives have very similar profiles with a low isomorphous difference. Figure 5.3

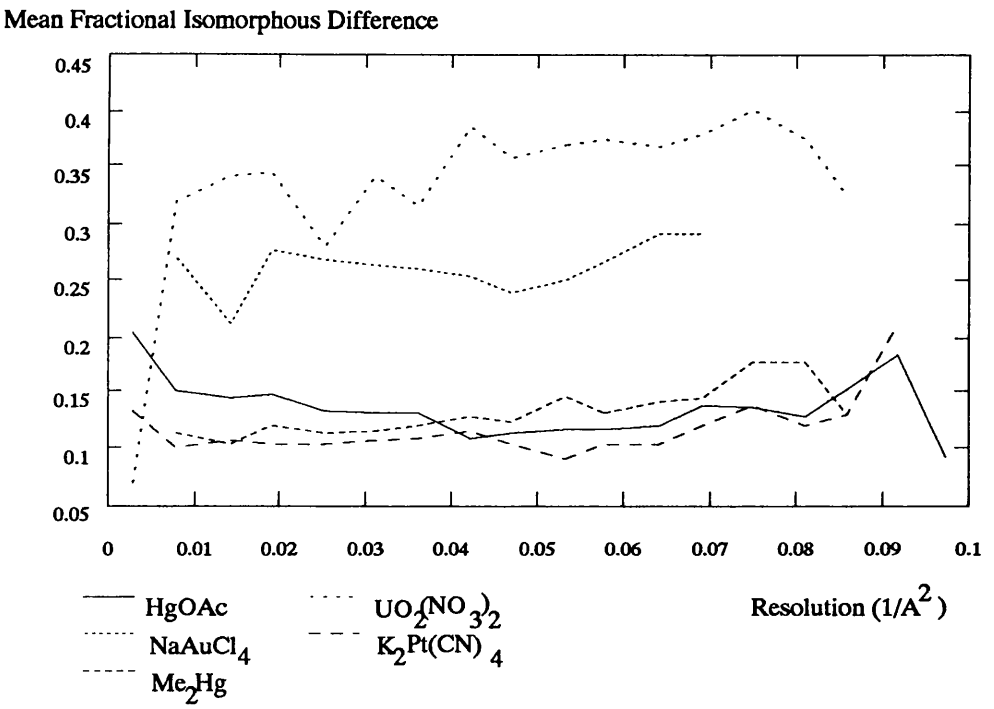


Figure 5.1: Plots of Mean Isomorphous Difference v. Resolution. *MOSCO* processed derivatives scaled relative to the Natfilm native dataset.

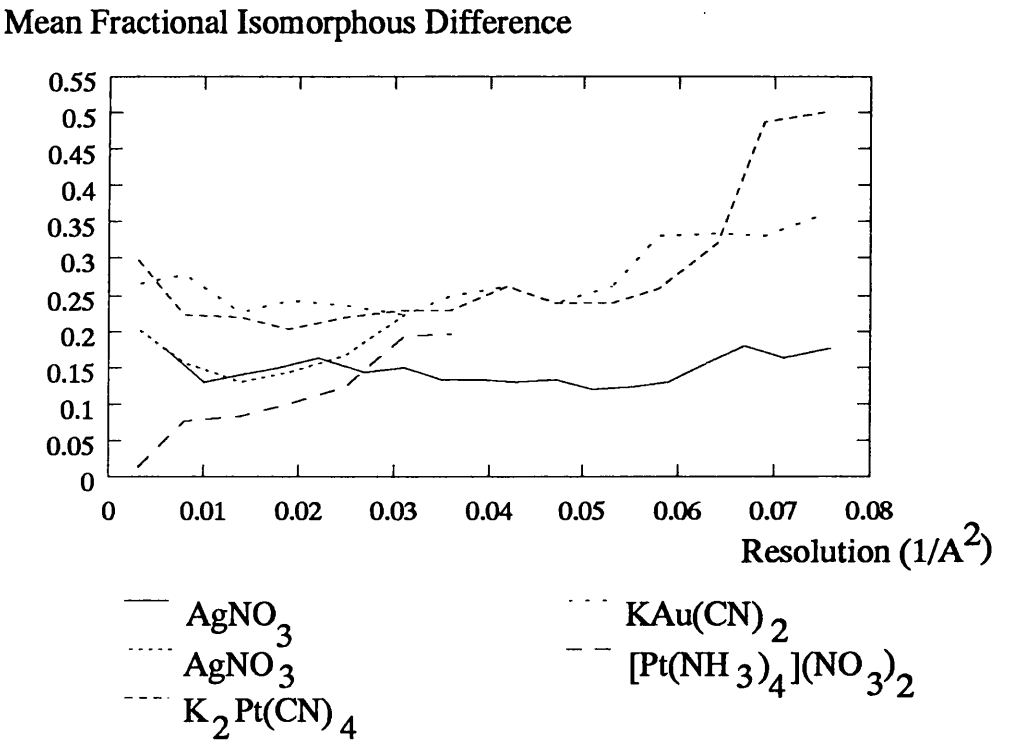
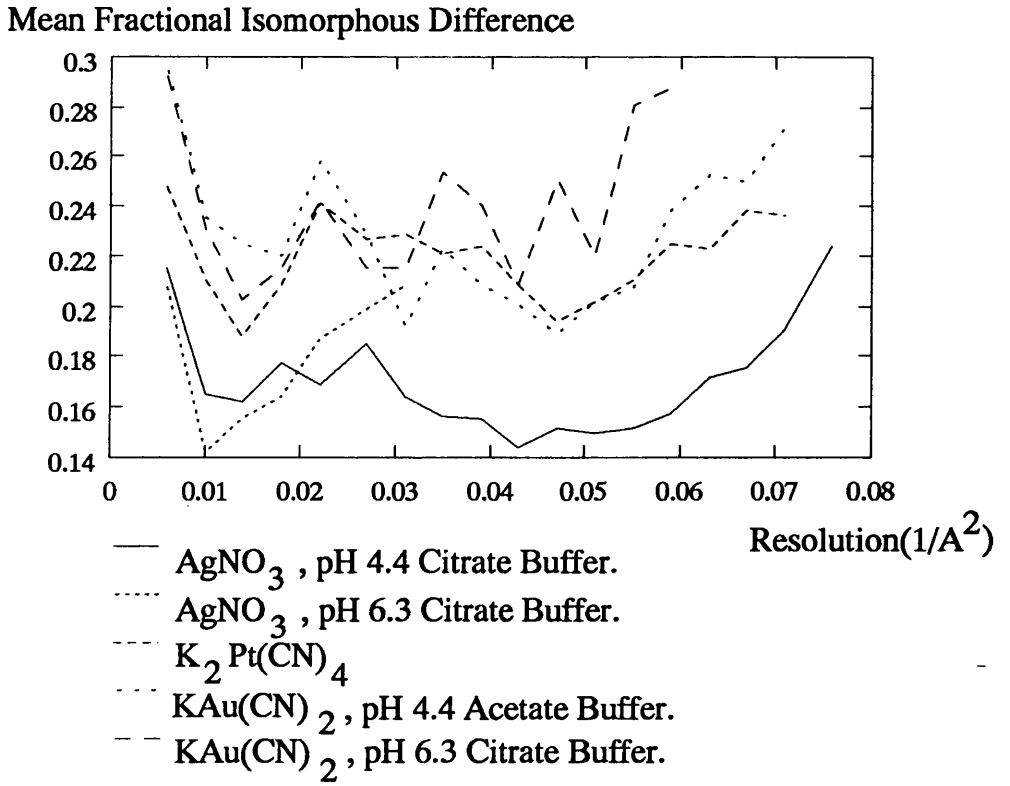


Figure 5.2: Plots of Mean Isomorphous Difference v. Resolution *XENGEN* processed data scaled relative to the Natxeng native dataset.



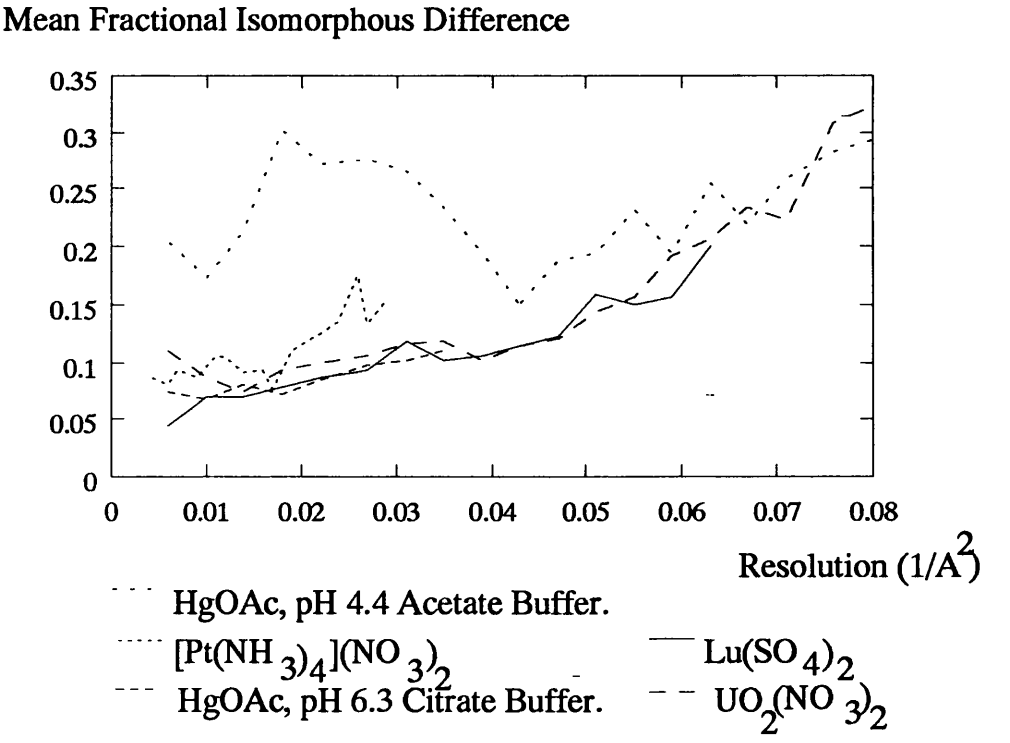


Figure 5.3: Plots of Mean Isomorphous Difference v. Resolution. Derivatives are processed using *XDS* and scaled relative to the Natxds native dataset.

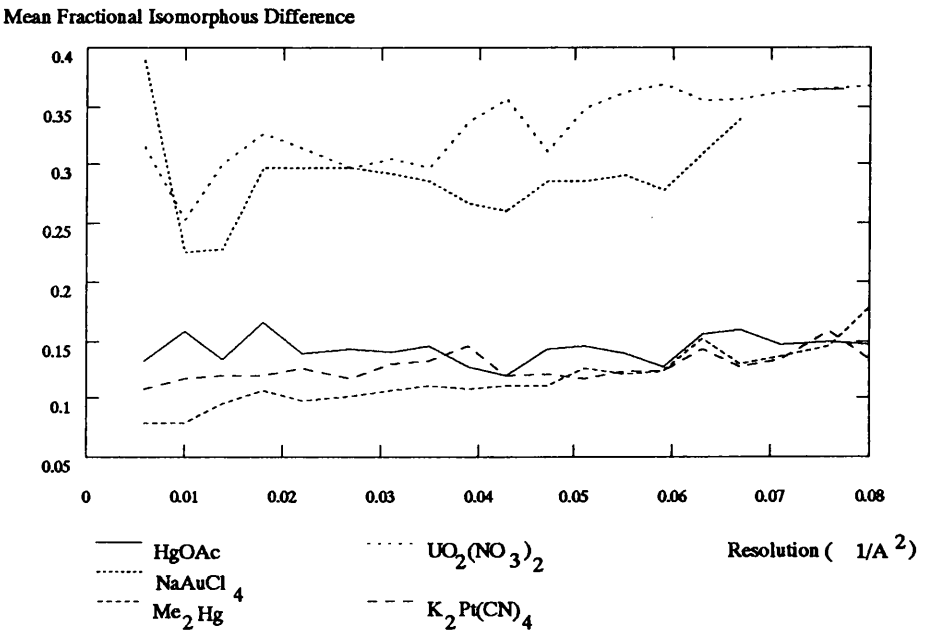


Figure 5.4: Plots of Mean Isomorphous Difference v. Resolution *MOSCO* processed derivatives scaled relative to Natxds native dataset

Derivative	Resoln (Å)	No. Common Reflns.	Scale Factor	Temp. Factor	R_{iso}	N_{anom}	K_{emp}
$K_2Pt(CN)_4$	3.2	2649	0.9617	1.082	0.110	283	3.21
Me_2Hg	3.5	2430	0.9616	-0.267	0.130	449	1.70
$Hg(OOCCH_3)_2$	3.5	2177	0.9146	2.762	0.129	461	3.22
$UO_2(NO_3)_2$	3.5	2728	1.2179	-2.376	0.356	872	5.97
$NaAuCl_4$	4.0	1697	1.1195	-7.791	0.260	581	4.29

Table 5.1: Scaling Statistics for Natfilm to *MOSCO* processed Derivatives

$$K_{emp} = 2\left(\frac{\overline{(F_{PH} - F_P)^2}}{\overline{(F_{PH}(+) - F_{PH}(-))^2}}\right)^{1/2}$$

showing the plots for data processed using *XDS* can be compared to Figure 5.2. in which the data has been scaled by *XENGEN*. The latter plots indicate a much greater lack of isomorphism than those using the *XDS* processed data (note the relative scales of the plots). It has been found that *XENGEN* systematically overestimates the weak reflections. This may be due overestimation at the measurement stage or because the scaling is poor (Dr. P. Moody, private communication). The difference between the datasets is especially pronounced for the $K_2Pt(CN)_4$ derivative at the higher resolution. (Table 4.10).

5.2 Phase Determination

The theory of phase determination, treatment of errors and phase refinement is discussed fully in several texts, for example Blundell and Johnson (1976). Here I present a brief summary of the theory and the practical applications.

If a heavy atom compound is introduced isomorphously into a protein crystal, then it s contribution to the diffraction pattern is given by

Derivative	Resoln (Å)	No. Common Reflns.	Scale Factor	Temp. Factor	R_{iso}	N_{anom}	K_{emp}
$K_2Pt(CN)_4$	3.5	2160	1.0823	-3.3517	0.259	2335	7.04
$AgNO_3$	5.5	830	1.0098	0.1531	0.165	-	-
$KAu(CN)_2$	3.5	1538	0.9900	-1.3595	0.268	1785	6.36
$(Pt(NH_3)_4)(NO_3)_2$	5.7	838	0.9977	0.5856	0.116	-	-

Table 5.2: Scaling Statistics for Natxeng to *XENGEN* processed Derivatives

Derivative	Source	Resln (Å)	No. Common Reflns.	Scale Factor	Temp. Factor	R_{iso}	N_{anom}	K_{emp}
$K_2Pt(CN)_4$	A.D.	3.5	2781	0.9551	-0.5968	0.221	2391	4.70
$KAu(CN)_2$	A.D.	3.5	1974	1.1175	-6.1885	0.232	1481	4.34
$KAu(CN)_2$	A.D.	4.0	1301	1.1677	-14.8949	0.235	725	4.99
$AgNO_3$	A.D.	3.5	2410	0.9871	-1.1440	0.166	2090	3.57
$AgNO_3$	A.D.	5.7	877	0.9861	-8.1435	0.176	-	-
$(Pt(NH_3)_4)(NO_3)_2$	A.D.	5.7	827	1.4250	-7.3419	0.109	-	-
$Lu_2(SO_4)_3$	A.D.	4.0	1274	1.0437	1.7939	0.107	-	-
$Hg(OOCCH_3)_2$	A.D.	5.5	1160	1.3391	-3.8065	0.083	601	2.92
$Hg(OOCCH_3)_2$	A.D.	3.5	894	1.1448	0.6204	0.227	894	3.41
$UO_2(NO_3)_2$	A.D.	3.5	1882	1.1436	1.1323	0.158	1920	2.35
$K_2Pt(CN)_4$	Film	3.5	2814	0.9881	0.3162	0.126	273	3.77
$Hg(OOCCH_3)_2$	Film	3.5	2216	0.9402	1.7522	0.142	537	3.57
$NaAuCl_4$	Film	3.9	1959	1.1373	-7.3430	0.284	708	4.55
$UO_2(NO_3)_2$	Film	3.5	2971	1.2339	-5.7423	0.334	1061	4.97
Me_2Hg	Film	3.5	2608	0.9790	-1.0179	0.118	548	1.74

Table 5.3: Scaling Statistics for Natxds to *XDS* Derivatives

A.D. - Xentronics area detector

$$\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$$

where $\mathbf{F}_P$ represents the structure factor of the native protein, $\mathbf{F}_{PH}$ that of the derivative and $\mathbf{F}_H$ is the structure factor of the heavy atom contribution.

Once the position of the heavy atom in the unit cell is known, the phase for the protein can be determined from the simultaneous equations

$$F_{PH}^2 = F_P^2 + F_H^2 + F_P F_H \cos(\alpha_P - \alpha_H)$$

as F_P and F_{PH} can be measured experimentally and F_H and α_H can be calculated. This can be represented diagramatically as in Figure 5.5. As shown in Figure 5.5 there are two possible phase solutions which satisfy the geometrical conditions. This ambiguity can be resolved either by a second derivative or by anomalous scattering. In reality the phase triangles rarely close as indicated because of errors, primarily due to inaccuracy in the measurement and scaling of F_{PH} and F_P and errors in $\mathbf{F}_H$ due to non-isomorphism. To gain an estimate of the error Blow & Crick (1959) showed that an acceptable approximation was to consider that the lack of closure error resulted from F_{PH}, Figure 5.6. The lack of closure error is defined as

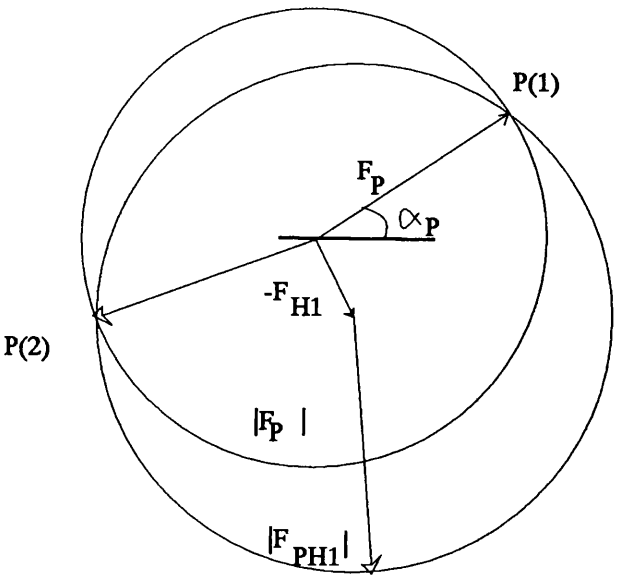
$$\varepsilon = (F_{PH}\text{obs} - F_{PH}\text{calc})$$

If a Gaussian distribution of errors is assumed, then the probability for j^{th} derivative that α_P is correct is given by

$$P_j(\alpha) = N exp[\frac{-\epsilon_j(\alpha)^2}{2E_j^2}]$$

where N is a normalization factor, $\varepsilon(\alpha)$ is the lack of closure of the phase triangle along F_{PH} at protein phase angle α, and E_j is the standard deviation of the distribution. In combining the results of several derivatives, there are two possible phase angles from $P(\alpha)$:

i) the most probable phase - Max($P(\alpha)$)



P(1) and P(2) represent the two possible values for the phase.

Figure 5.5: Ideal Phase Circles for a single derivative

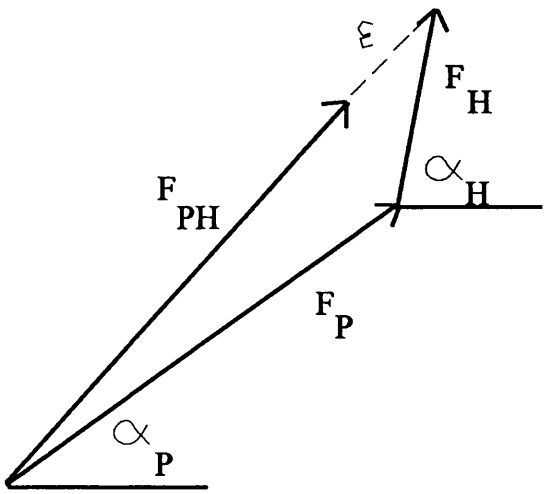


Figure 5.6: Effect of errors on the phase triangle

ii) the centroid or best phase given by the centroid of the probability distribution (Blow & Crick (1959)). Use of this results in an electron density map with the least square error.

$$F_{best} = m|F_P|e^{i\alpha_{best}}$$

where m is the figure of merit dependent on the sharpness of the phase probability and is equivalent to a weighting function:

$$\frac{\int_{\alpha=0}^{2\pi} e^{i\alpha} P(\alpha) d\alpha}{\int_{\alpha=0}^{2\pi} P(\alpha) d\alpha}$$

5.2.1 Determination of Heavy Atom Positions

From the above theory, in order to determine the phases it is necessary to obtain a model for the vector F_H which is dependent upon the knowledge of the position of the heavy atoms in the cell. There are several possible approaches to obtaining this information, some of which are discussed below.

5.2.2 The Patterson Method

The most general approach which requires no previous information is use of the Patterson method (Patterson, (1934)). A Difference Patterson obtained using the coefficients (F_{PH}^2 - F_P^2) would result in a vector map containing peaks corresponding to heavy atom-heavy atom vectors and also heavy atom-protein vectors and would therefore be extremely difficult to interpret. A better method is to use the coefficients $(F_{PH} - F_P)^2$. This expands to give

(a)

$$4F_{PH}^2 sin^4[(\alpha_P - \alpha_{PH})/2]$$

(b)

$$+F_H^2 cos^2(\alpha_{PH} - \alpha_H)$$

(c)

$$-4F_P F_H sin^2[(\alpha_P - \alpha_{PH})/2] cos(\alpha_{PH} - \alpha_H)$$

If $F_H << F_P$ the terms (a) and (c) will generate background noise, and term (b) will result in a map containing the expected vectors at about half weight (Blundell & Johnson (1976)). The magnitude of the differences between native and derivative amplitudes is not generally equal to the heavy atom structure factors due to scaling and measurement errors as well as lack of isomorphism. These effects contribute noise to the Patterson. Difficulties can also arise when there are several sites with low occupancies, high symmetry or the protein size is large resulting in a Patterson which can be very difficult to interpret.

Alternatively a difference Patterson with coefficients $(F_{PH}(+) - F_{PH}(-))^2$ can be calculated from anomalous data (Rossman (1961)). This results in a map ideally containing positive peaks of half weight at the end of vectors relating anomalous scatterers. As the observations used are measured from the same crystal problems with systematic error are minimized, no scaling is neccessary and isomorphism is not a factor, these maps are potentially very useful. However, the difficulty in accurately measuring very small differences results in the anomalous difference Pattersons generally being much more noisy than the respective isomorphous difference Pattersons. Also, the Patterson function can be dominated by the largest terms. If these have associated large errors, then the maps will be impossible to interpret. Great care must be taken to exclude any spuriously large differences from the Patterson calculation. (This equally applies to the calculation of isomorphous difference Pattersons). Given this, anomalous difference Pattersons are extremely useful especially when used in conjunction with the isomorphous map for checking solutions.

5.2.3 Direct Methods

The location of a larger number of heavy atoms by difference Patterson becomes increasingly difficult and the use of Direct Methods can help in solving the problem. The theory of Direct Methods relies heavily on the postulate that the unit cell contains a random distribution of scatterers which is not strictly the case, especially if isomorphous differences are used. In this case, assuming no errors, the number of scatterers is equivalent to the number of heavy atom sites. The higher the symmetry and the greater the number of sites, the more realistic is the use of Wilson statistics. A knowledge of the number of heavy atoms is also required for the probabilty expressions. Despite these theoretical problems, Direct Methods have been successfully used to locate heavy atom positions (Steitz (1968); Nagai (1991)). Often several equally probable solutions are produced and these can be used in conjunction with the difference Pattersons to give the correct solution.

5.2.4 Difference Fouriers

Once the heavy atom positions are known for one derivative it is then possible to use this information to phase difference Fouriers for other potential derivatives. The difference Fourier is calculated with coefficients

$$m(F_{PH} - F_P)\exp(i\alpha_P)$$

One potential problem with this method is the fact that Fourier syntheses tend to be dominated by phases more than by amplitudes; if the α_P are too biased with respect to one derivative, "ghost" peaks corresponding to the positions of its sites tend to appear in the map and can be erroneously interpreted as minor sites. It is therefore wise to check the solutions produced from this method with the difference Pattersons.

5.2.5 Refinement of Phases

Once the heavy atom positions have been determined, the parameters that describe them are optimized by minimization of the squared differences of the calculated and observed values. This is not strictly correct, as all equations for the calculated values (F_H) depend on other observable quantities ($|F_{PH}|$ or $|F_P|$). Refinement and phasing are correlated as the lack of closure (ϵ) can be minimized either by changing the parameters defining F_H or the phase (α_P) (Figure 5.6).

The method of phase refinement used during this project and implemented in the *CCP4* program *PHARE* was that of alternate phase and refinement. The residual minimized is

$$R = \sum_h w_h(|\mathbf{F_{PH}}|_{obs} - |\mathbf{F_{PH}}|_{calc})^2$$

or

$$R = \sum_h w_h(|\mathbf{F_{PH}}|_{obs} - k_{rel}|\mathbf{F_P} + \mathbf{F_H}|)^2$$

where

$$\mathbf{F_P} = |\mathbf{F_P}|_{obs} e^{(i\alpha_{mostprobable})}$$

The phases are held fixed and the above residual is minimized with respect to the heavy atom parameters. These refined parameters are then used to determine new phases and the process repeated iteratively until convergence is achieved. An inherent problem with this procedure is the bias imposed on the phases by incorrect heavy atom parameters.

Implimentation of maximum likelihood refinement in the program *MLPHARE* (Ottwinowski (1991)) has reduced this bias. In standard least squares heavy atom refinement methods only one phase is considered per reflection. With *MLPHARE* method a likelihood function, based on the Blow and Crick estimate of errors, can be used to produce a probablility distribution for the phase calculated from the heavy atom parameters. The lack of closure residual is weighted by the probability of a phase being correct.

x	y	z	-y	x-y	1/3+z	y-x	-x	2/3+z
-x	-y	1/2+z	y	y-x	5/6+z	x-y	x	1/6+z
y	x	1/3-z	-x	y-x	2/3-z	x-y	-y	-z
-y	-x	5/6-z	x	x-y	1/6-z	y-x	y	1/2-z

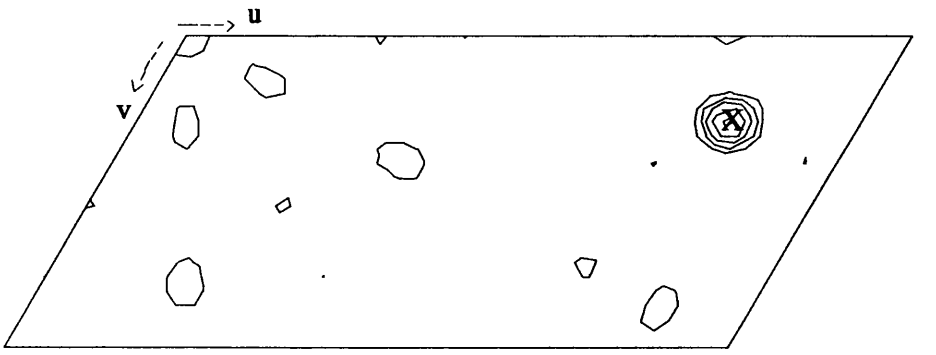
Table 5.4: Equivalent Positions for $P6_122$

5.3 Solution of Derivatives ; *XENGEN* processed data

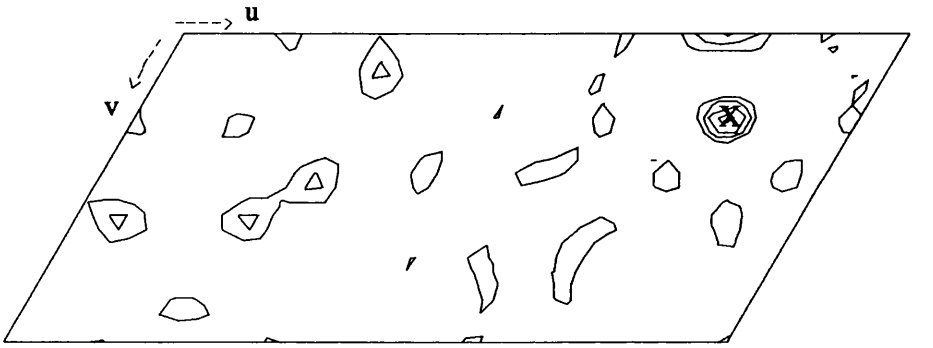
All the results discussed in this section are determined from the datasets processed using *XENGEN* as listed in Table 5.2. Low resolution isomorphous difference Pattersons were calculated for all derivatives using the program *FFT89KW* in the *CCP4* suite. Data where F is less than 3σ(F) were excluded from the calculation along with unusually large differences as indicated by the program *SCALEIT*. Due to limitations in the FFT program, the data were expanded and the maps calculated in P1. Table 5.4 gives the equivalent positions for space group $P6_122$ and Table 5.5 lists the resulting vectors for the Harker planes (u,v,1/6), (u,v,1/3), (u,v,1/2), (0,v,w) and (u,0,w). From these sections it is possible to solve the heavy atom position from the heavy atom - heavy atom vectors.

5.3.1 Solution of $K_2Pt(CN)_4$

The resulting Pattersons were difficult to interpret because of the high symmetry. For $K_2Pt(CN)_4$ the *x* and *y* co-ordinates were determined from both the isomorphous and anomalous Pattersons, Figure 5.7. However, it was difficult to determine the z co-ordinate so Direct Methods were then employed. The package *MITHRIL* written by Dr. C. Gilmore was first used, and the resulting solutions checked using *MULTAN* written by Dr. P. Main. These were compared with the difference Pattersons. For direct methods, the number of expected heavy atoms in the unit cell was varied from 12 to



Isomorphous Difference Patterson



Anomalous Difference Patterson

Figure 5.7: $w = 1/6$ Harker section for $K_2Pt(CN)_4$ Difference Pattersons. The map is calculated at 5Å resolution and contours are at 1σ intervals.

(u,v,1/6)	(u,v,1/3)	(u,v,1/2)	(0,v,w)	(u,0,w)
±(x-y,x)	±(2x-y,y+x)	±(2x,2y)	±(2y-x,2z-1/6)	±(2x-y,2z-1/2)
±(y,y-x)	±(x+y,2y-x)	±(2y,2x-2y)	±(2x-y,2z-1/2)	±(y-x,2z+1/6)
±(x,y)	±(2y-x,y-2x)	±(2y-2x,-2x)	±(x-y,2z+1/6	±(2y-x,2z-1/6)
±(y-x,y)	±(2y-x,x+y)	±(2y,2x)	±(x-2y,2z-1/6)	±(y-2x,2z-1/2)
±(y,x)	±(y+x,2x-y)	±(2x,2y-2x)	±(y-2x,2z-1/2)	±(y+x,2z+1/6)
±(x,y-x)	±(2x-y,x-2y)	±(2x-2y,-2y)	±(x+y,2z+1/6)	±(x-2y,2z-1/6)

Table 5.5: Patterson Vectors on the Harker Planes

48. In practice this had little effect on the most probable solution. All small differences (F_{PH} - F_P) < 20 and the strongest reflections with (F_{PH} - F_P) > 450 were removed. The structure factor amplitudes were converted to normalised structure factors (E's) and triplets were generated using the top 220E's. The most probable solution had an ABSFOM of 0.5 and a residual of 37.56%. ABSFOM is a measure of the internal consistency among Σ_2 triplet relationships and would ideally be 1.0. The residual is analogous to a conventional R value. The top three peak heights are listed in Table 5.6. Peak 1 gave similar *x* and *y* co-ordinates to those determined from the Pattersons. A *z* co-ordinate of 0.25 resulted in the peaks on the Harkers coinciding with the major peaks already used for other vectors and gave a solution that was very close to a special position. This position was refined using the *CCP4* program *PHARENEW* using data to 5Å resolution. The initial isomorphous and anomalous occupancies were set to 1.0 and the site refined for 10 cycles. This resulted in an overall figure of merit of 0.29 and 0.50 for acentric and centric reflections respectively. The final occupancies were 1.10 and 0.70 for isomorphous and anomalous respectively. The phases calculated from this site were then used to calculate difference Fouriers for the remaining derivatives.

No.	Co-ordinate x	y	z	Peak Height
1	0.5448	0.0909	0.2500	1732
2	0.5456	0.0935	0.5886	542
3	0.5260	0.0062	0.5225	383

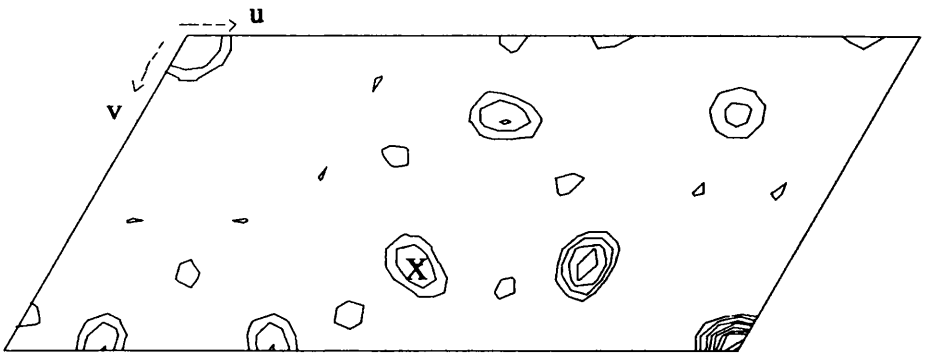
Table 5.6: Solution using *MITHRIL* for $K_2Pt(CN)_4$

5.3.2 Solution of $AgNO_3$ Derivative

A 5Å resolution difference Fourier with coefficients (F_{Ag} - $F_{Natxeng}$) was calculated using the S.I.R. phases calculated from the Platinum derivative. The maximum peak was found to occur at 0.0909, 0.3333, 0.0833. with a height of 6.3σ. Peaks above 3σ were listed and that corresponding to the Platinum site was not amongst them. The result from this difference Fourier was checked against the isomorphous difference Patterson for the silver (Figure 5.8), and the position was found to account for most of the major peaks. The two large peaks present in the z = 1/6 section shown in the figure do not have corresponding peaks on the other Harker sections. The silver site was refined with *PHARENEW* and these S.I.R. phases then used to calculate a Platinum difference Fourier. The maximum peak, 8.3σ, was found at 0.5455, 0.0909, 0.2500 *i.e.* the original Platinum site. There were no "ghost" silver sites among the peaks greater than 3σ.

5.3.3 Calculation of the first MIR Phases

PHARENEW was used to calculate the first MIR phases for HF-HCG using the two sites described above. After 10 cycles of refinement the heavy atom positions were stable with values



Isomorphous Difference Patterson

Figure 5.8: w = 1/6 Harker Section for 6Å $AgNO_3$ Difference Patterson. Contours at 1σ intervals.

Platinum - x,y,z : (0.544, 0.092, 0.238)

Isomorphous Occupancy 1.47, Anomalous Occupancy 0.24

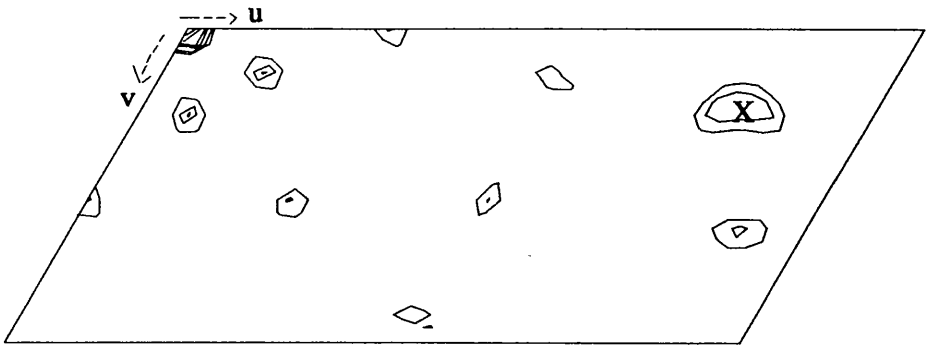
Silver - x,y,z : (0.081, 0.328, 0.085)

Isomorphous Occupancy 1.35

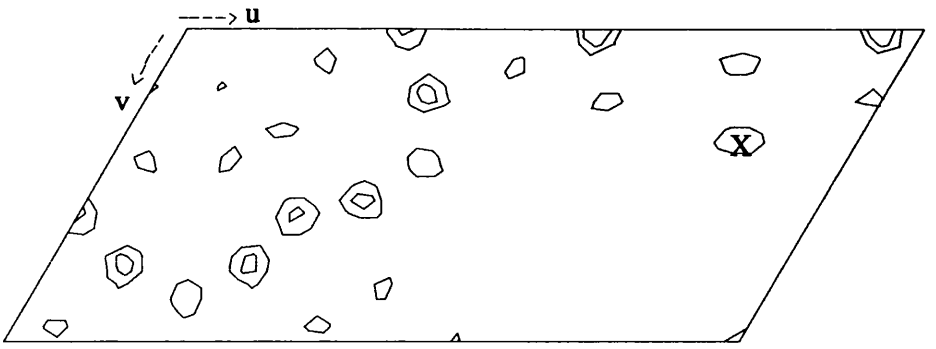
These refined positions gave phases with an overall Figure of Merit of 0.411 and 0.671 for the Acentric and Centric reflections, respectively. The Figure of Merit is analysed against resolution for the Acentric and Centric terms in Table 5.7.

Resoln (Å)	14.55	11.43	9.41	8.00	6.96	6.15	5.52	5.00	Total
No. Acentric Reflns	21	50	76	114	153	191	231	239	1075
F.O.M.	0.623	0.532	0.496	0.545	0.511	0.452	0.330	0.256	0.411
No. Centric Reflns	26	30	40	29	28	29	28	30	240
F.O.M.	0.792	0.675	0.781	0.750	0.585	0.722	0.566	0.469	0.671

Table 5.7: Figure of Merit Analysis for first MIR phases



Isomorphous Difference Patterson

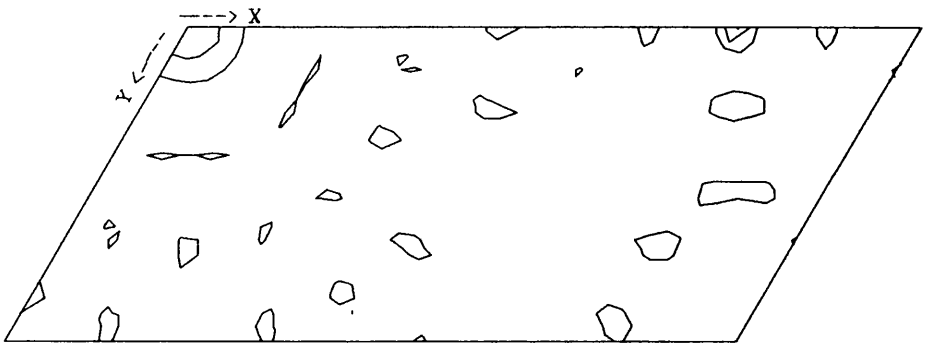


Anomalous Difference Patterson

Figure 5.9: w = 1/6 Harker Section for 6Å $KAu(CN)_2$ Difference Pattersons. Contours at 1σ intervals.

5.3.4 Solution of $KAu(CN)_2$ Derivative

The MIR phases were used to phase a Fourier with coefficients (F_{Au} - $F_{Natxeng}$). The maximum peak was located at the same position as for the Platinum derivative. This was not suprising, as the difference Pattersons for the gold data, Figure 5.9, have similar features to the Platinum ones. However, to check that the result from the difference Fourier was not a "ghost" peak from the Platinum site's contribution to the phases, another difference Fourier was calculated using the S.I.R. phases from the Silver site. The maximum peak in this map, 6.8σ, corresponded to the Platinum site, confirming



Isomorphous Difference Patterson

Figure 5.10: w = 1/6 Harker Section for 6Å $[Pt(NH_3)_4](NO_3)_2$ Difference Patterson. Contours at 1σ intervals.

that the Gold data did contain bound heavy atom.

5.3.5 Analysis of $[Pt(NH_3)_4](NO_3)_2$ Derivative Data

The MIR phases calculated using the Silver and Platinum derivatives were used to calculate a difference Fourier for the derivative soaked with a second Platinum bearing salt. In this case the top peaks corresponded to the Silver and Platinum sites, and no new sites were found. This was perhaps expected as this data has a very low isomorphous difference and the difference Patterson, Figure 5.10, shows no distinct peaks.

5.3.6 Determination of Handedness

All the Fouriers and phase refinements discussed above were calculated assuming the true enantiomer was $P6_122$. As both the Platinum and Gold derivatives had associated anomalous data it was possible to check this assumption using the refinement program *PHARENEW*. In the first case the spacegroup was defined as $P6_122$ and the sites used

Spacegroup	Deriv.	Iso. Occ.	Ano. Occ.	Overall F.O.M. (Acen)
	Pt	1.50	-0.45	
$P6_122$	Ag	1.33	-	0.4691
	Au	1.47	-0.39	
	Pt	1.50	0.45	
$P6_522$	Ag	1.33	-	0.4691
	Au	1.47	0.39	

Table 5.8: Determination of Handedness using 5Å MIR phases

as given above *i.e.* x,y,z. For the second case, the spacegroup was defined as $P6_522$ and the sites used were -x,-y,-z. The occupancy of all isomorphous sites were initially given as 1.0, and the anomalous sites as 0.0. Ten cycles of refinement were calculated and the results in Table 5.8 obtained. From this the true spacegroup was determined to be $P6_522$. Table 5.9 gives the variation of the overall Figure of Merit with resolution. These initial results were corroborated later in the project using the *XDS* processed data. Phases were refined for a single derivative in each spacegroup, with and without anomalous data. These were used to phase difference Fouriers using an independent derivative and the relative peak sizes used to check this assignment of $P6_522$ as the true space group. The statistics for these MIR phases are given in Table 5.10

5.4 Analysis of the *MOSCO* processed data

The phases calculated using the Xentronics data were used to phase difference Fouriers for the film data.

As might be expected, the MIR phased difference Fourier with coefficients (F_{Pt} - $F_{Natfilm}$) resulted in a map with the major peak at the same site for the xentronics

Resoln (Å)	14.55	11.43	9.41	8.00	6.96	6.15	5.52	5.00	Total
No. Acentric Reflns	21	50	76	115	153	191	238	249	1093
F.O.M.	0.667	0.650	0.572	0.615	0.565	0.503	0.375	0.321	0.469
No. Centric Reflns	26	30	41	29	28	29	31	33	247
F.O.M.	0.810	0.690	0.834	0.816	0.672	0.727	0.645	0.525	0.716

Table 5.9: Figure of Merit Analysis for Pt, Ag and Au calculated phases

	$K_2Pt(CN)_4$		$AgNO_3$		$KAu(CN)_2$	
	Acentric	Centric	Acentric	Centric	Acentric	Centric
α_{diff}	82.84	90.78	91.53	77.63	84.56	87.34
Mean Iso. Diff.	224.7	336.0	186.1	226.9	251.0	389.8
Mean Closure Error	142.4	193.4	151.3	170.2	179.2	285.2
Mean Phasing Power	1.87	1.55	1.07	0.88	1.44	0.94
Cullis R	-	0.76	-	0.87	-	0.86
Mean Anom. Diff.	52.47	-	-	-	60.62	-

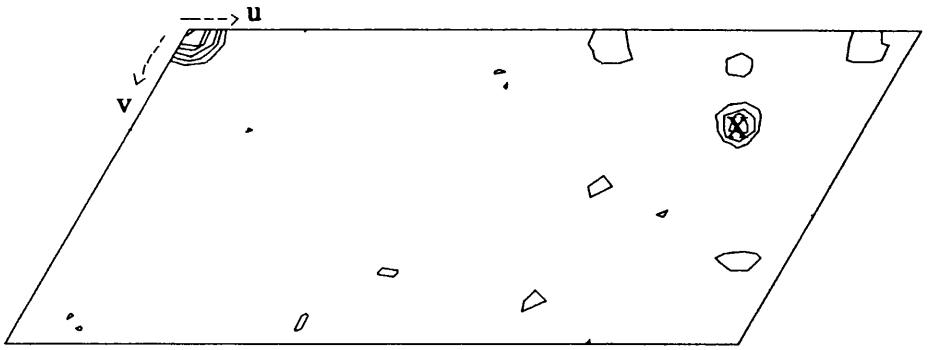
Table 5.10: Phasing Statistics for *XENGEN* data

$$\alpha_{diff} = \frac{\sum \mid \alpha_P - \alpha_H \mid}{N}$$

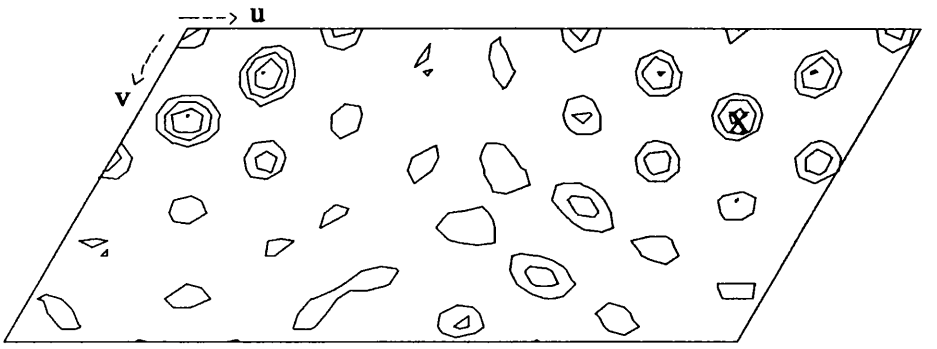
$$MeanClosureError = \frac{\sum (F_{PH} - F_{PHcalc})_2}{N}$$

$$CullisR = \frac{\sum \mid (F_{PH} \pm F_P) - F_H \mid}{\sum \mid F_{PH} - F_P \mid}$$

$$PhasingPower = \frac{rms(|F_H|_{calc})}{rms(\epsilon)}$$



Isomorphous Difference Patterson



Anomalous Difference Patterson

Figure 5.11: w = 1/6 Harker Section for 5Å $K_2Pt(CN)_4$ Difference Pattersons. Contours at 1σ intervals.

Platinum dataset. Figure 5.11 shows the z=1/6 Harker section for the difference Pattersons which show the same features as that in Figure 5.7. The anomalous map calculated using the film data is less clear than that for the Xentronics data. The Xentronics data was collected from a single crystal mounted along c* which allowed the accurate measurement of the anomalous data. The film data was collected from three separate crystals which were only approximately aligned along c*. The Isomorphous Difference Pattersons for the other derivative datasets collected on film are presented in Figure 5.12. The results of the difference Fouriers are given in Table 5.11. The peaks in the Pattersons were inconsistent with the top heavy atom site for each of these

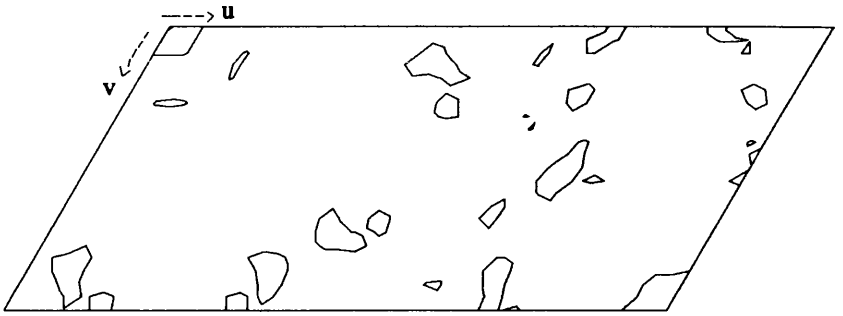
Derivative	x	y	z	Pk.Ht. (σ)
$K_2Pt(CN)_4$	0.5454	0.0909	0.2500	25.5
Me_2Hg	0.5454	0.0909	0.2500	4.8
$Hg(OOCCH_3)_2$	0.5412	0.5636	0.0985	9.0
$UO_2(NO_3)_2$	0.5454	0.0909	0.2500	12.5
$NaAuCl_4$	0.2445	0.3266	0.08545	11.6

Table 5.11: Top peaks found in difference Fouriers for Film data phased with the 5Å resolution MIR phases calculated from the Xentronics data.

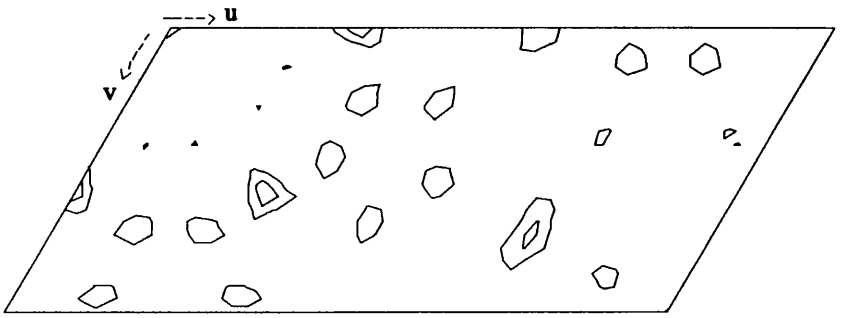
derivatives from the difference Fourier and so were assumed to be "ghost" peaks from the input heavy atom sites.

5.5 Determination of higher resolution phases

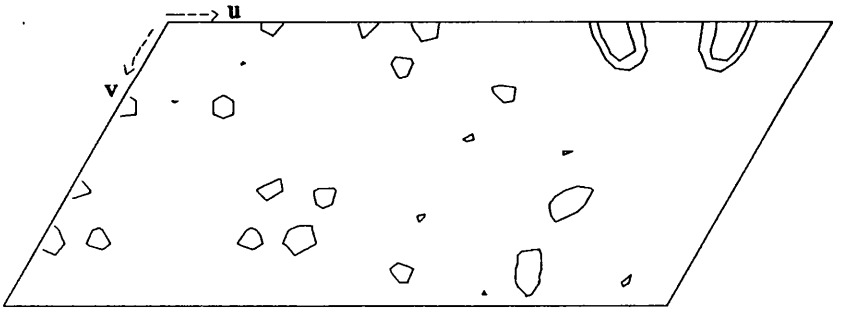
Once the Silver derivative was solved, a higher resolution dataset was collected using the Xentronics area detector. Further difference Fouriers were calculated and some minor sites were discovered. These were included in the phase refinement. To allow the use of phase extension techniques, the *XENGEN* processed derivatives were scaled to the combined native dataset Natcomb, as described in section 4.6. This extended the data to 3Å resolution. Despite the data being incomplete and the higher resolution Platinum data being poorly measured, it was decided to use the datasets to their resolution limits *i.e.* 3.5Å for the Platinum and Gold sets and 5.5Å and 3.5Å for the Silver sets. The sites were refined using the program *PHARE* and the results of the refinement are given in Table 5.12.



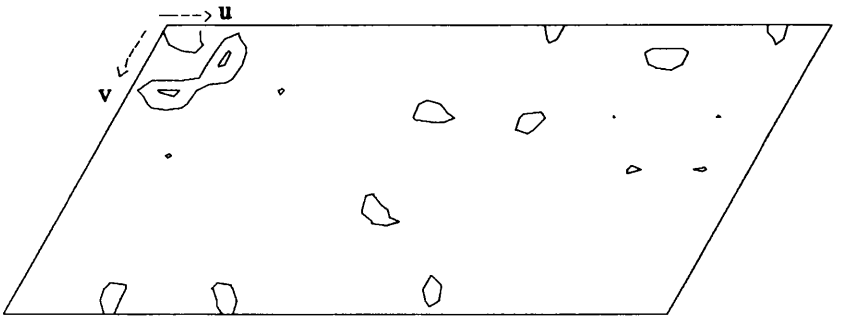
Isomorphous Difference Patterson for Me2Hg



Isomorphous Difference Patterson for NaAuCl4



Isomorphous Difference Patterson for HgOAc



Isomorphous Difference Patterson for UO2(NO3)2

Figure 5.12: w = 1/6 Harker Sections calculated using Film datasets.

Contours at 1σ intervals.

Dmin (Å)	Derivative	Co-ordinates	Iso. Occ.	Anom. Occ.	R_{cullis}	Phasing Power	α_{diff}
3.5	Platinum	(-0.5393, -0.0922, -0.2371)	3.0	2.4	0.65	2.2	90.9
3.5	Gold	(-0.5412, -0.0982, -0.2389)	2.7	0.6	0.83	1.3	96.7
5.5	Silver	(-0.0802, -0.3289, -0.0822)	3.1		0.64	2.2	91.4
		(0.7760, 0.1467, 0.2494)	1.0				
3.5	Silver	(-0.0802, -0.3289, -0.0822)	2.1	0.8	0.68	1.7	90.9
		(0.4465, 0.4625, 0.0772)	2.5	1.6			
		(0.4556, 0.4902, 0.0106)	1.1	1.4			

Table 5.12: Results of heavy atom site refinement with *XENGEN* processed datasets

$$\alpha_{diff} = \frac{\sum \mid \alpha_P - \alpha_H \mid}{N}$$

$$CullisR = \frac{\sum \mid (F_{PH} \pm F_P) - F_H \mid}{\sum \mid F_{PH} - F_P \mid}$$

$$PhasingPower = \frac{rms(|F_H|_{calc})}{rms(\epsilon)}$$

5.6 Phase improvement

The electron density map calculated using these phases was not interpretable. Two phase improvement methods were used in an attempt to improve the quality of the map.

5.6.1 Solvent Flattening

It is possible to improve the isomorphous phases using the technique of solvent flattening. A molecular envelope is determined and the region outside this closed boundary this is designated solvent. The electron density of the solvent is set to a constant value and a positivity constraint is applied to the protein density. This modified map is Fourier transformed, the phases combined with the original MIR phases and a new electron density map calculated. This is then used as the starting map and the procedure is repeated iteratively. The solvent flattening programs in the *CCP4* suite use the algorithm of B.C. Wang (1985) which allows automatic envelope determination (Leslie (1988)).

A 5Å resolution MIR map was calculated using the phases described in Table 5.12 and truncated to remove negative densities. Structure factors were calculated from this map and weighted structure factors determined using the weighting function $(1 - (r_i/R)^2)$ where r is the distance of the grid point *i* from the centre of the sphere of radius R. The radius used for HF-hCG was 9Å. These weighted structure factors were used to determine the averaged map in which the electron density at each grid point is replaced by the weighted average of all density around that point within 9Å. The molecular envelope was determined from this averaged map.

A 3.5Å resolution MIR map was calculated and the molecular envelope applied to flatten this map. Structure factors were calculated and scaled relative to the observed structure factors. The calculated phases were combined with the MIR phases and a new map computed. This was then used as the initial map for the next cycle. The process

Resln	No.	Mean F.O.M.		Abs. Phase Change	
Å		Old	New	Old	Calc
11.1	185	0.804	0.902	16.9	26.9
7.8	349	0.800	0.903	10.6	23.7
6.4	426	0.670	0.764	14.3	27.8
5.5	496	0.694	0.847	17.8	27.1
4.9	542	0.648	0.844	18.2	23.5
4.5	587	0.620	0.700	23.4	21.3
4.2	604	0.538	0.798	33.2	21.7
3.9	547	0.383	0.748	44.3	12.9
3.7	571	0.230	0.711	59.2	8.6
3.5	611	0.133	0.694	70.4	4.66

Table 5.13: Results of solvent flattening the *XENGEN* processed data

was repeated iteratively until convergence. The final phases were used to calculate a 5Å resolution map and a new envelope determined. This was continued until no further improvement in the electron density was observed. A total of 5 envelopes were applied to 5 iterative cycles to produce the final solvent flattened map. Table 5.13 gives details of the results. The electron density map is discussed in the next chapter.

5.6.2 Maximum Entropy

A method for improving protein phases using the theory of maximum entropy is being developed by Dr. C. Carter. The program *MICE* (Maximum Entropy in a Crystallographic Environment) is a modified version of the program used for small molecule structure determination written by Dr. C. Gilmore (Bricogne (1984); Bricogne (1990)).

The maximum entropy method is based upon the theory of Bayesian statistics. Bayes'

Theorem states that, given two events E and F then

$$\underset{\text{posterior}}{p(E|F)} \propto \underset{\text{prior}}{p(E)} \quad \underset{\text{likelihood}}{p(F|E)}$$

where p denotes probability. p(E), the prior model, represents the current knowledge. The likelihood represents what is learned from consulting the experimental data, F, and the resulting posterior probability reflects modifications to the prior due to consideration of the data. This can then be used as the new prior and the process repeated iteratively. It is essential that the prior model be as unbiased as possible, *i.e.* no assumptions are made about what is unknown. This can be achieved by using maximum entropy theory. The entropy, S, of a probability distribution p , is given by

$$S = - \sum_i p_i \log p_i$$

This should be a maximum subject to the constraints of reproducing the prior model. The prior consists of information such as well determined phases and the protein envelope. The application of this theory to the improvement of phases for proteins is discussed below. For the method to be used the prerequisites are

i) Good MIR phases for the basis set

ii) An accurate envelope

iii) A knowledge of the unit cell constants

iv) An estimate of the electron density of the contents

v) Amplitudes scaled as near as possible to absolute scale.

This maximum entropy approach has been successfully used to phase cytidine deaminase. (Carter, D.C., private communication). The structure had previously been determined and the envelope was calculated from a (2Fobs - Fc) map. Reflections with resulting phases were combined with the MIR phases, the process repeated, and the R-factor fell to 25%. In the light of these results it was hoped that the method could be applied to hCG. Reasonable phases were available to 3.5Å resolution and native structure factors were measured to 3.0Å resolution.

In order to make use of this technique I worked in Dr. C. Carter's laboraratory at the University of North Carolina for ten days. The MIR phases used were those described in Section 5.5 and the final envelope calculated using the *CCP4* solvent flattening programs was used as the envelope to be applied in *MICE*. The basis set {H} used included all reflections whose phases had a F.O.M. of 0.7 or greater. All other reflections were assigned to the nonbasis set{K}.

The native amplitude data are scaled using the program *MITHRIL* (Gilmore (1984)). *MICE* uses unitary structure factors ($U_{000} = 1.0$). The amplitudes are scaled then divided by $\sqrt{N}$ where N is the number of atoms in the structure to give unitary structure factors. The{H}set was used as a constraints in the generation of an exponential model for the distribution of atoms, q^{ME}.

$$q^{ME}(\underline{x}) = \exp \sum_{\underline{h} \in H} \zeta_{\underline{h}} \exp 2\pi i \underline{h}.\underline{x}$$

An ω map is calculated using the ζ's from the basis set as coefficients. An initial approximation to $q^{ME}(x)$ is calculated from the normalized exponential of the ω map. It is at this stage that the prior information *i.e.* the envelope constraint is applied within this normalization factor. This inverse Fourier of this q map is taken to give the $U_h{}^{ME}$'s. A difference Fourier with coefficients ($U_h{}^{obs}$ - $U_h{}^{ME}$) is then calculated and from this shifts to be applied to the ζ's are calculated. These are applied and the process repeated iteratively to produce the q^{ME} map which reproduces what is known *i.e.* the apriori information and is maximally noncommital with respect to any unknown structure factors. There are several measures of how successful a series of cycles were. These are

i) L Likelihood gain which should increase to a maximum. This is effectively a figure of merit and is dependent upon the agreement between $|U^{ME}|$ and $|U^{obs}|$.

ii) Goodness of fit (χ^2) which should be close to 1.

iii) Maximum entropy should be as high as possible.

iv) R-factor which gives a good general indication of the solution.

Taken in conjunction, these parameters give a reasonable indication of the progress

of refinement. For example, a solution having a high Likelihood but low maximum entropy is unlikely to be correct. Refinement was considered complete when the Likelihood reached a maximum.

Once a *MICE* run had converged, the phases were combined with the original MIR phases. Using the structure factors for the envelope determined using the *CCP4* suite, convergence was achieved in 45 cycles. 466 unique reflections were used in the basis set to give R = 52.8% and a likelihood of 2492. The statistics for phase combination are given in Table 5.14. The phases were used to calculate a map which was displayed and compared to the solvent flattened and MIR maps using *FRODO* (Jones (1978). Figure 5.13 shows a single section of the *MICE* map and the MIR map. The former looks a lot cleaner than the MIR map as would be expected from the solvent flattening. However, when compared to the *CCP4* solvent flattened map there was little apparent improvement. From the figure, it can be seen that the envelope is not optimal. Despite this, the maximum entropy approach looked promising as regions defined as solvent but which appear to be protein have remained in the final map. The definition of the envelope was found to be critical to the method. Forty cycles of *MICE* were run using no envelope to test the effect on the map. The statistics for the run looked promising, but this was due to the use of too few constraints rather than a true improvement in the phases. The resulting electron density map was totally uninterpretable. Errors in the solvent region had built up to strong density and it was impossible even to define the solvent:protein boundary. At this stage it was decided to attempt to redefine the solvent boundary by hand. The map resulting from the *MICE* run using the *CCP4* envelope could not be used for this as the old mask was already part of the constraints used within *MICE* to determine these phases. The original MIR map was used for the mask determination which was drawn section-by-section using a program written by Dr. J. Pflugraph. The new envelope was more conservative, allowing 57.5% protein *c.f.* 48% protein for the previous envelope. Convergence was reached after only 28 cycles and the phasing statistics for this run are given in Table 5.15. The combined phases were used to calculate a new starting map for re-input into *MICE*. After 24 cycles the R-factor was 28.66% and the likelihood 2766. These phases were recombined with the MIR

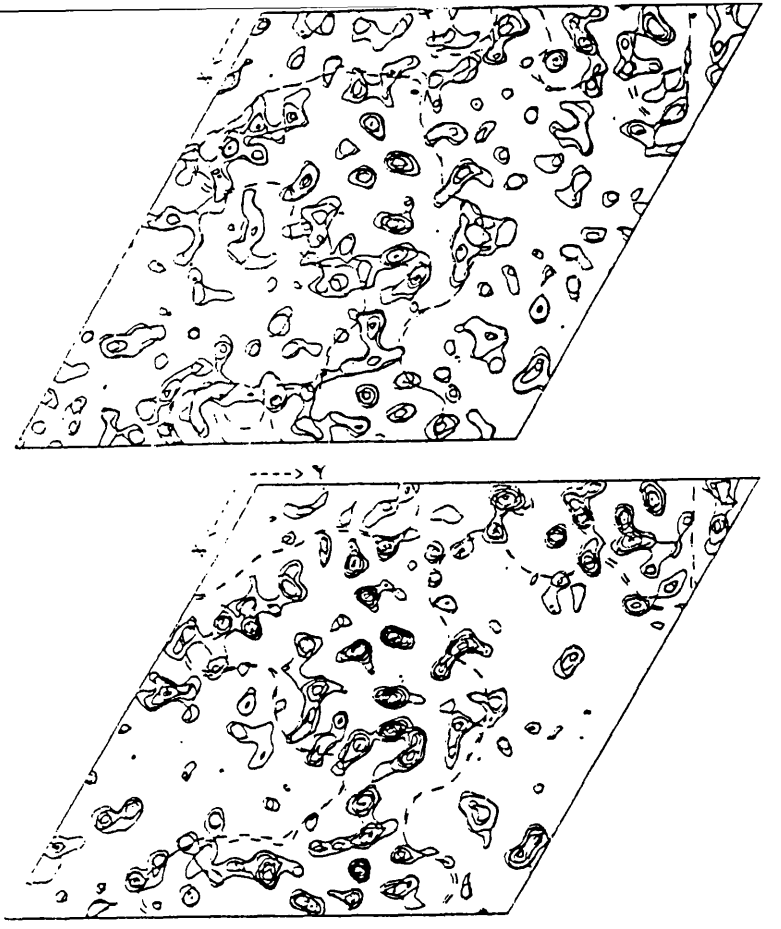


Figure 5.13: Equivalent sections of the MIR electron density map and the map calculated with MIR phases combined with *MICE* phases. Note also the solvent boundary.

phases and another electron density map calculated. These electron density maps again showed no marked improvement over the *CCP4* solvent flattened one. Comparisons of the maps are made in the next Chapter. Table 5.16 summarises the results of the different refinements.

5.7 XDS Data

The maps described above were studied and attempts were made to fit Cα traces. Once it became apparent that the maps were not of a high enough quality to allow the solution of the protein structure, the search for another derivative was resumed. Also, it was hoped that higher resolution data could be collected for those derivatives already solved. At this point the data processing package *XDS* became available so new datasets were processed using this, and previously collected datasets were re-processed as discussed

	No. Refins.	<shift>	<old FOM>	<new FOM>
Basis Set	2107	5.73	0.878	0.934
Extrap. Set	5067	75.50	0.138	0.551
Total	7174	53.01	0.355	0.664

Table 5.14: Phase combination statistics for MIR and *MICE* phases calculated using the *CCP4* envelope

	No. Refins.	<shift>	<old FOM>	<new FOM>
Basis Set	2107	4.64	0.878	0.939
Extrap. Set	5067	75.53	0.138	0.517
Total	7174	54.71	0.355	0.641

Table 5.15: Phase combination statistics for MIR and *MICE* phases calculated using the interactive envelope

in Chapter 4. Study of the relevant Tables in the previous Chapter reveal an obvious improvement in the quality of the higher resolution data (*i.e.* better than 4Å) resolution. Difference Pattersons were calculated and the w = 1/6 sections are displayed in Figures 5.14 and 5.15 for comparison with those calculated using *XENGEN* processed data. In general there seems to be an improvement in the datasets overall. Identical sites to those given in Table 5.12 were input for refinement in the program *MLPHARE*. After nine cycles of refinement the sites were stable with values given in Table 5.17. The temperature factors of all sites were fixed at 25Å^2. The variation of the Figure of Merit and phasing statistics are given in Table 5.18. These phases were used to calculate difference Fouriers for the derivatives included in the refinement. As can be seen from Figures 5.16 and 5.17, the maps are convincing.

The putative derivatives collected on film were scaled to the new native *XDS* processed set, Natxds, and difference Fouriers calculated using these new phases. Table 5.19 gives the results of these. These solutions were checked using the difference Pattersons

With Wang/Leslie envelope

Cycle	Crystallographic R-Factor	Likelihood*	χ^2
Initial	72.25	912.6	7.75
Final	42.9	2492	4.00

With manually built envelope

Cycle	Crystallographic R-Factor	Likelihood*	χ^2
Initial	66.6	730.5	6.43
Final	31.4	2769	2.21

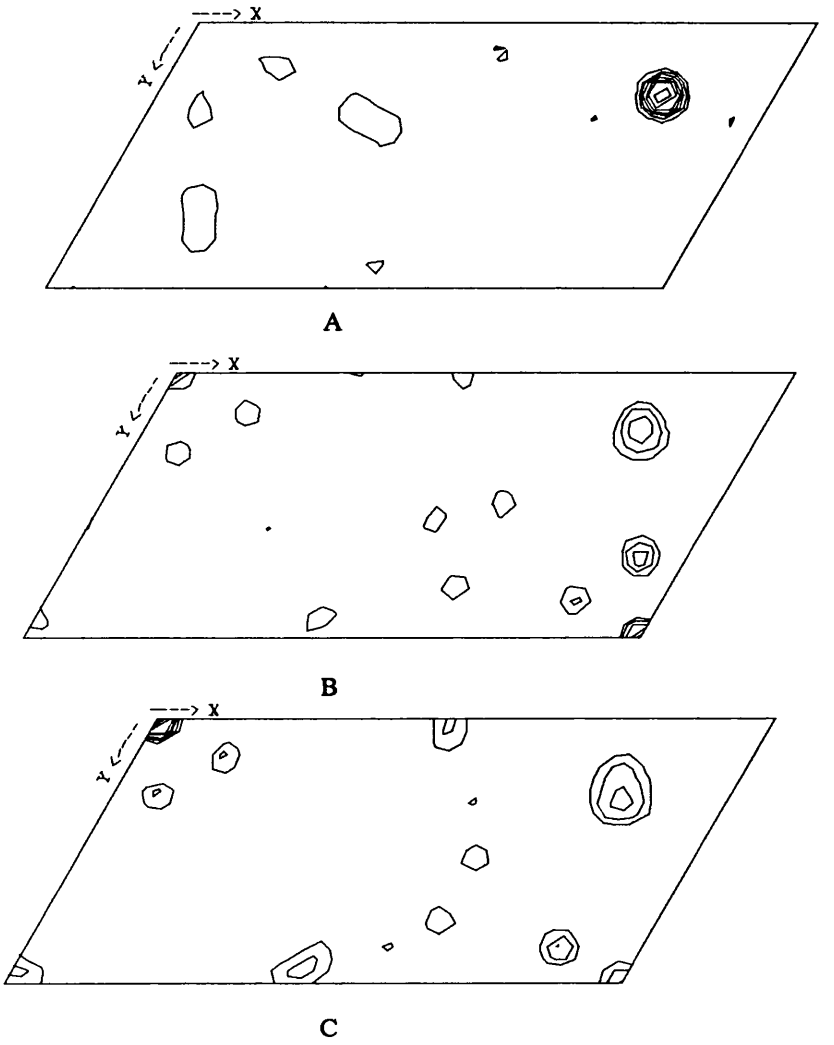
"MICE" phases and MIR phases combined

Cycle	Crystallographic R-Factor	Likelihood*	χ^2
Initial	31.74	2593	2.11
Final	28.66	2766	1.81

Table 5.16: Results of Maximum Entropy density modification
* Log Likelihood gain

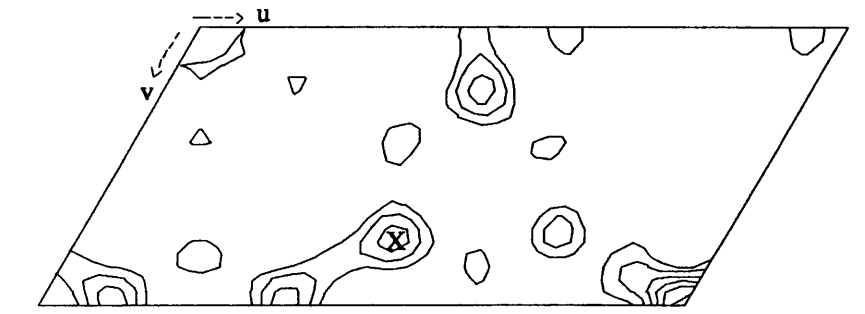
Dmin	Derivative	Co-ordinates	Occupancy	Anom. Occ.
3.5Å	Platinum	(-0.543, -0.091, -0.238)	1.05	0.95
5.5Å	Silver	(-0.081, -0.328, -0.085)	1.12	
		(-0.781, 0.156, 0.254)	0.38	
3.5Å	Silver	(-0.081, -0.328, -0.085)	0.65	0.32
		(0.446, 0.462, 0.076)	0.67	0.52
		(0.459, 0.487, 0.013)	0.34	0.25
5.5Å	Gold	(-0.542, -0.091, -0.238)	0.94	1.28
3.5Å	Gold	(-0.542, -0.091, -0.238)	0.94	0.93

Table 5.17: Results of heavy atom site refinement with *XDS* processed datasets

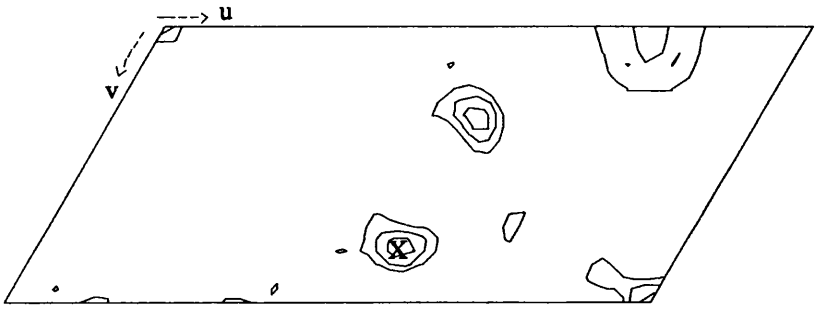


5.0A Isomorphous Difference Pattersons for
A) Platinum , B) Low resn Gold and C) Higher resn Gold XDS processed
Derivatives. All contoured at 1 sigma intervals.
Native dataset used was NATXDS

Figure 5.14: 5Å resolution Isomorphous Difference Pattersons using *XDS* processed data.



A



B

Low resolution Isomorphous Difference Pattersons using
XDS processed derivative datasets for
A) Low resn. Silver and B) Higher resn. Silver
Both are contoured with 1 sigma intervals.

Figure 5.15: Isomorphous Difference Pattersons using *XDS* processed data.

Dmin (Å)	17.8	11.2	8.2	6.5	5.3	4.5	3.9	3.5	Total
No. Reflns.	12	73	167	321	481	678	735	646	3113
F.O.M.	0.735	0.843	0.782	0.703	0.534	0.432	0.378	0.289	0.4632
No. Reflns.	31	74	116	141	167	171	173	110	983
F.O.M.	0.703	0.843	0.871	0.782	0.653	0.555	0.538	0.497	0.658

Derivative	Resln (Å)	α_{diff}	Isomorphous Difference	Lack of Closure	Phasing Power	Cullis R	Anomolous Difference
$K_2Pt(CN)_4$	3.5	78.4	171.8	108.3	1.4	0.63	61.8
		83.5	249.2	140.8	1.2	0.57	
$AgNO_3$	5.5	87.6	143.5	95.3	1.5	0.66	-
		88.4	208.9	127.7	1.1	0.61	
$AgNO_3$	3.5	84.1	130.1	103.4	1.0	0.79	54.9
		85.9	194.4	147.9	0.7	0.76	
$KAu(CN)_2$	4.5	78.4	193.4	142.9	1.1	0.74	71.3
		90.7	274.7	199.3	0.8	0.73	
$KAu(CN)_2$	3.5	74.2	172.6	134.1	1.0	0.78	68.4
		85.4	259.9	200.5	0.7	0.77	

Table 5.18: Analysis of phasing statistics for 3.5Å resolution phases using *XDS* processed data.

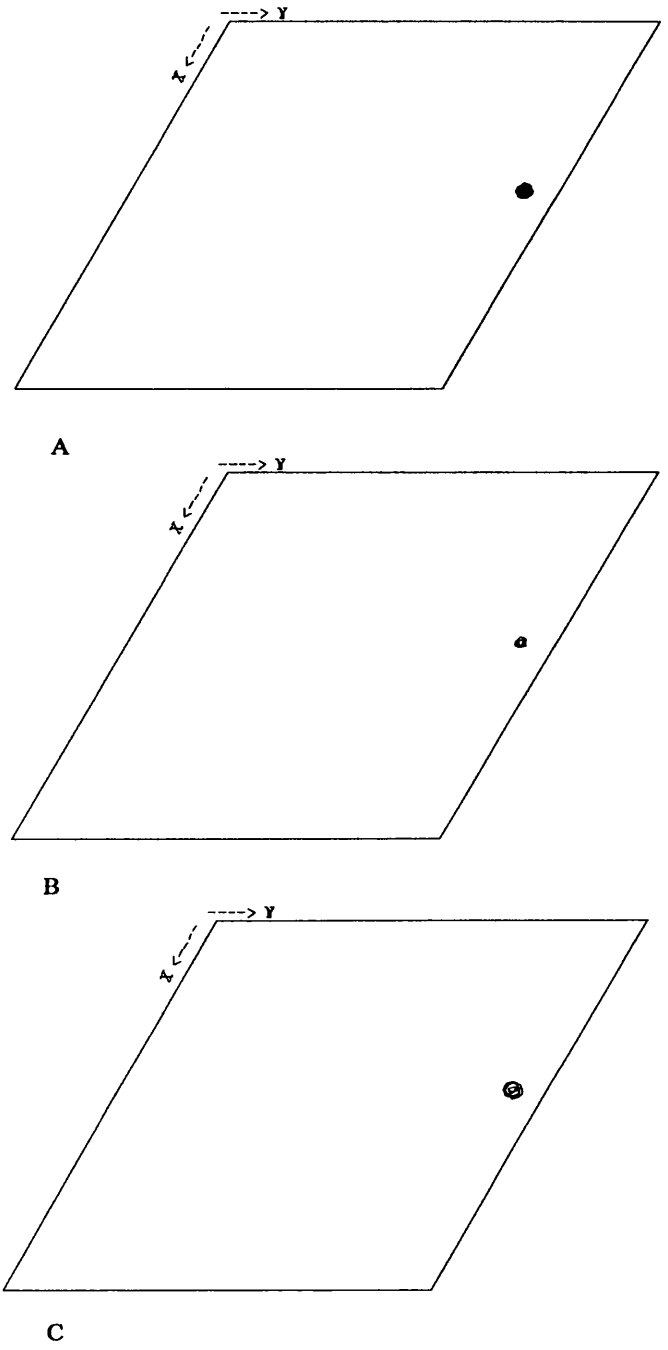
The upper half gives the variation of Figure of Merit with resolution for acentric and centric terms respectively. The lower half gives further phasing statistics for acentric and centric terms respectively for each derivative.

$$\alpha_{diff} = \frac{\sum \mid \alpha_P - \alpha_H \mid}{N}$$

$$MeanClosureError = \frac{\sum (F_{PH} - F_{PHcalc})_2}{N}$$

$$CullisR = \frac{\sum \mid (F_{PH} \pm F_P) - F_H \mid}{\sum \mid F_{PH} - F_P \mid}$$

$$PhasingPower = \frac{rms(\mid F_H \mid_{calc})}{rms(\epsilon)}$$



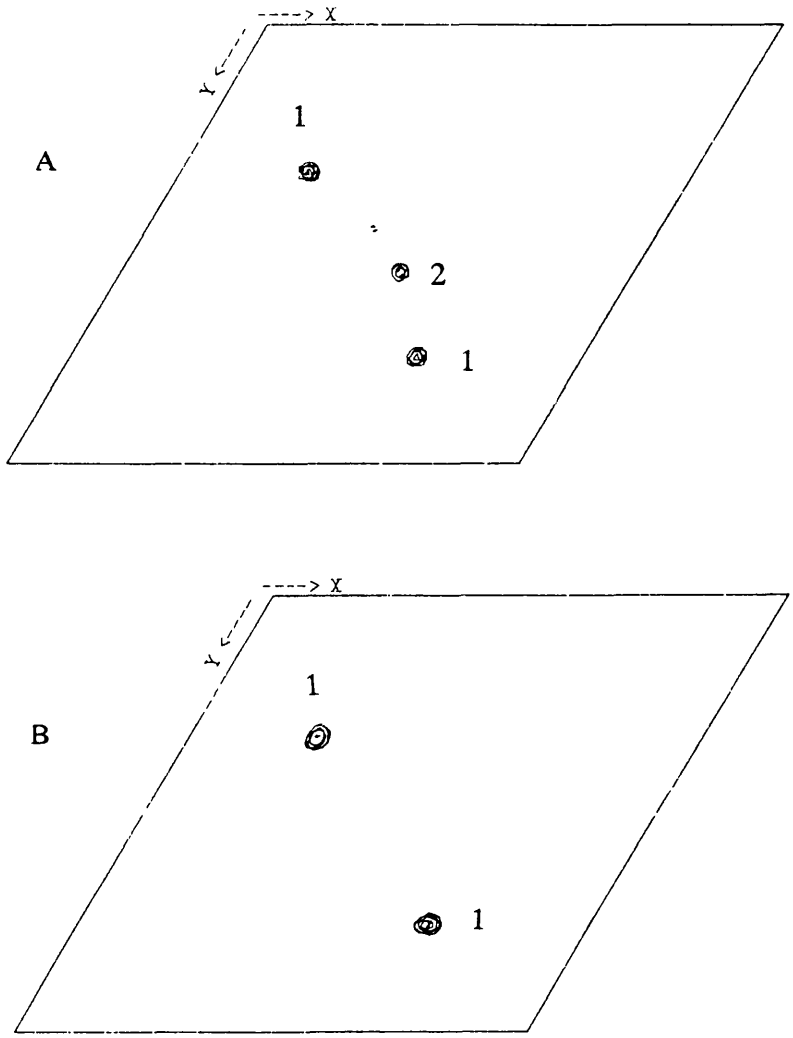
Difference Fouriers contoured from 5 rms mean density in steps of 2 rms

A) 3.5A map for Platinum derivative.

B) 3.5A map for Gold derivative. (pH 4.4 soak)

C) 5.5A map for Gold derivative (pH 6.3 soak)

Figure 5.16: *XDS* difference Fouriers for K_2PtCN_4 and K_2AuCN_2 derivatives.



Difference Fouriers contoured from 4σ in steps of $2\ \sigma$.

A) 3.5Å resolution map for Silver derivative (pH 4.4).

The two major sites are labelled 1 & 2.

B) 5.5Å resolution map for Silver derivative (pH 6.3).

The major site is labelled 1.

Figure 5.17: *XDS* difference Fouriers for $AgNO_3$ Derivatives

(Figure 5.18). In most cases the peaks on the Harker sections did not fully satisfy the results of the difference Fouriers. When peaks were present on at least two of the Harker sections, the site was included for refinement in *MLPHARE*. Datasets collected from similar heavy atom soaks, *e.g.* for the $K_2Pt(CN)_4$ derivative datasets, were not merged, but treated as independent derivatives. It was thought that this should minimize problems with merging datasets from different sources collected from different crystals with possibly different occupancies of heavy atoms bound. Several cycles of refinement were performed. Any sites which did not refine were removed and the process repeated. Once all sites were stable the resulting phases where used in the calculation of difference Fouriers for all derivatives. Any new sites were input and the process repeated until all sites were accounted for in the difference Fouriers. The resulting sites are given in Table 5.20. The final set of phases have an overall figure of merit of 0.492 and 0.704 for acentric and centric terms respectively. The resulting electron density map calculated using these phases is discussed in the next chapter.

5.7.1 Improvement of Phases - use of *SQUASH*

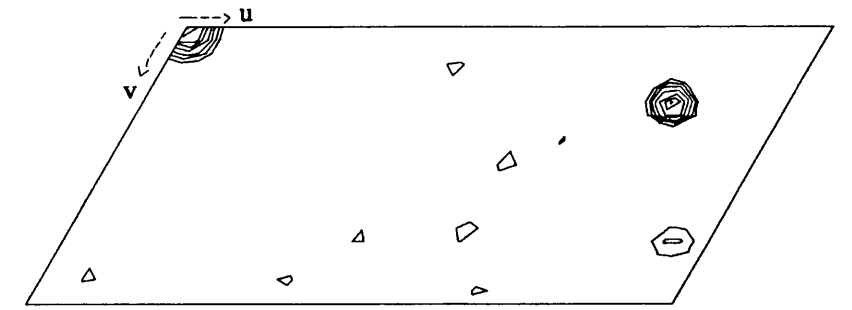
The map calculated from the phases discussed in the previous section looked promising but was still very difficult to interpret. To improve these phases the program *SQUASH* was utilised. The solvent flattening and histogram matching options in this package were used to improve and extend the phases for hCG. The program requires a set of structure factors on an absolute scale and estimated phases and figures of merit for a subset of the reflections. An envelope is determined from the electron density map and applied in the same way as for the standard solvent flattening described above.

The histogram matching method, applied to protein phase inprovement by Dr. P. Main and colleagues (Zhang *et al* (1988)), involves the matching of the test histogram to an expected histogram. The histogram is the probability distribution of electron density values at the grid points at which the map is calculated and the expected histogram is found from experimental values taken from refined protein structures. The similarity

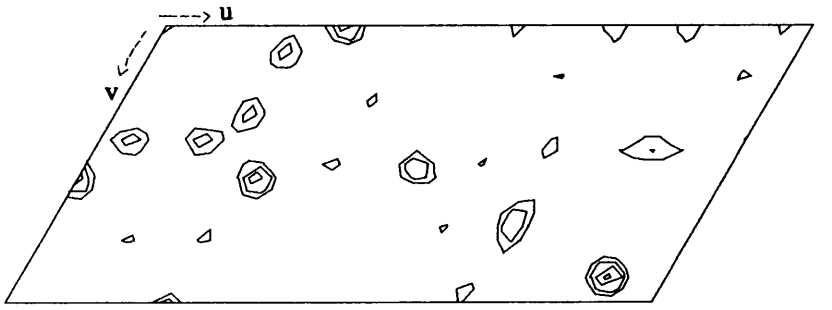
Derivative	x	y	z	Pk. Ht.
$Hg(OOCCH_3)_2$	0.442	0.460	0.073	6.2
	0.446	0.536	0.118	6.2
$UO_2(NO_3)_2$	0.455	0.544	0.0833	8.4*
	0.381	0.131	0.093	5.8
	0.264	0.922	0.124	4.6
	0.544	0.086	0.146	4.3
	0.366	0.046	0.107	3.9
	0.947	0.610	0.117	3.9
	0.340	0.312	0.092	3.8
$KAuCl_4$	0.244	0.322	0.089	7.2
	0.287	0.144	0.032	4.2
	0.513	0.467	0.148	4.0
$K_2Pt(CN)_4$	0.455	0.544	0.083	10.3*
	0.463	0.555	0.049	4.4
	0.455	0.269	0.083	4.4
	0.579	0.611	0.073	4.2
	0.541	0.557	0.076	4.2

Table 5.19: Top peaks on difference Fouriers phased using Xentronics data Fouriers calculated using Film derivative datasets and Natxds native data set. Phases are those derived using the sites in Table 5.17 for the *XDS* processed derivatives. Peak heights are given as a multiple of the rms density.

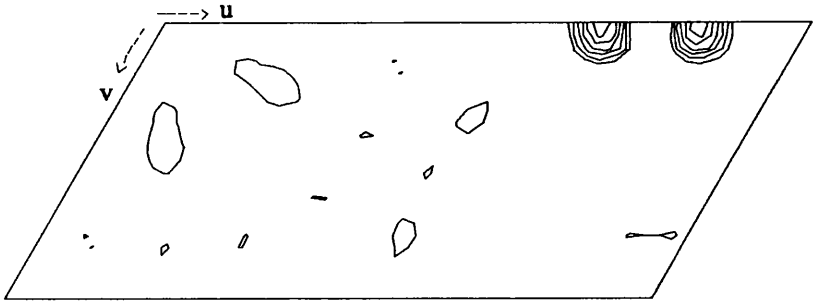
* indicates a site equivalent to that for the major Platinum site for the Xentronics dataset.



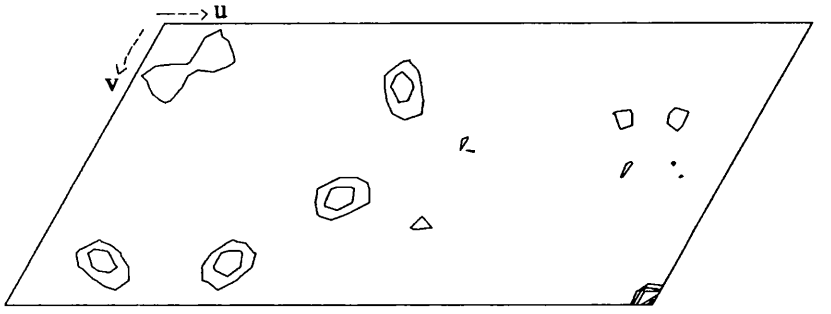
A



B



C



D

Isomorphous Difference Pattersons calculated with coefficients $(F_{PH} - Fnatxds)^2$

A) Platinum , B) Gold Chloride, C) Mercury Acetate and D) Uranium

Figure 5.18: 5Å resolution Isomorphous Difference Pattersons for film data v. Natxds

Dmin (Å)	Derivative	Source	Co-ordinates	Occupancy	Anomolous Occupancy	Phasing Power Centric	Acentric
3.5	$K_2Pt(CN)_4$	Xen	-0.542, -0.091, -0.237	1.04	1.01	1.1	1.4
			0.428, 0.472, 0.061	0.10	0.07		
4.5	$K_2Au(CN)_2$	Xen	-0.542, -0.091, -0.238	0.93	1.26	0.8	1.0
3.5	$K_2Au(CN)_2$	Xen	-0.542, -0.093, -0.238	0.90	0.94	0.7	1.0
5.0	$AgNO_3$	Xen	0.248, 0.327, 0.085	1.05		1.0	1.5
			0.456, 0.461, 0.079	0.28			
			0.463, 0.494, 0.014	0.26			
3.5	$AgNO_3$	Xen	0.246, 0.328, 0.085	0.66	0.36	0.7	1.1
			0.445, 0.461, 0.074	0.72	0.57		
			0.458, 0.488, 0.013	0.33	0.28		
3.5	$Hg(OOCCH_3)_2$	Film	0.444, 0.461, 0.068	0.59	0.48	0.7	0.9
3.5	$K_2Pt(CN)_4$	Film	-0.545, -0.095, -0.241	0.41	0.54	0.6	0.6
3.5	$KAuCl_4$	Film	0.242, 0.323, 0.085	1.00	0.67	0.7	0.8
3.5	$UO_2(NO_3)_2$	Film	-0.535, -0.087, -0.237	0.47		0.4	0.5
			0.370, 0.135, 0.115	0.16			
			0.254, 0.928, 0.129	0.25			
			0.360, 0.044, 0.106	0.19			
			0.955, 0.618, 0.115	0.15			
			0.334, 0.319, 0.099	0.30			

Table 5.20: Results of heavy atom site refinement using datasets collected on Film and on Xentronics.

Xen - Xentronics Area Detector.

Film - Film data collected on station PX7.2, S.R.S., Daresbury.

$$PhasingPower = \frac{rms(|F_H|_{calc})}{rms(\epsilon)}$$

of the test to the model protein was found to be unimportant; the density distribution is only dependent upon the fact that the density results from atoms at characteristic distances apart with particular stereochemistry. The basic procedure is to calculate the test histogram and a model histogram at the same resolution (if phase extension is being attempted, the model histogram is calculated at the higher resolution.) These histograms are then divided into equal segments and the test model scaled so that the probability distribution is the same as that of the model one. The solvent flattening and Histogram matching are carried out automatically within *SQUASH*. An envelope is determined to a user-specified resolution, the solvent region is truncated and the density within the envelope modified by histogram matching. Structure factors are calculated from the revised map and their Sim weights are calculated (Sim (1960)). The new phases are combined with the MIR phases and the resulting new map used as the starting point for the next cycle. This is repeated iteratively until the process has converged.

For HF-hCG the best results were found using an envelope determined from a 5Å resolution MIR map. Attempts to use a higher (3.5Å) resolution initial map were unsuccessful - the mean Figure of Merit to this resolution is low and as this is used to calculate the initial weighted map, the resulting protein density was very weak. Phases were extended to 3.5Å resolution. The statistics of this phase improvement method are given in Table 5.21.

5.7.2 Solvent Flattening

Several groups have successfully used the solvent flattening technique to improve poor MIR phases sufficiently to allow an interpretable map to be calculated. Initially the mask is calculated from a low resolution electron density map, applied to the MIR map until convergence is achieved. The initial resolution is increased by a small amount (*e.g.* 0.0022Å $^{-1}$) and the process repeated. This procedure was applied to phase improvement for HF-hCG. An initial electron density map was calculated to

Resln (Å)	No. Reflns.	<FOM_{obs}>	<FOM_{cmb}>	<PHI_{cmb} - PHI_{obs}>
100.00	269	0.866	0.916	4.1
8.45	452	0.829	0.887	2.9
6.45	567	0.699	0.795	8.2
5.42	599	0.597	0.734	9.5
4.71	663	0.565	0.685	8.2
4.26	720	0.474	0.598	10.3
3.92	772	0.359	0.474	12.8
3.65	785	0.245	0.335	18.8
3.50	136	0.182	0.258	16.0

Table 5.21: Results of phase improvement and extension for *XDS* MIR phases using *SQUASH*.

The envelope was calculated from a 5Å resolution map. Histogram matching was performed using R3 2Zn pig Insulin for the model histogram. Phases were extended to 3.5Å.

4.5Å resolution, an averaged map produced using an averaging radius of 9Å and an envelope determined. This was applied to the MIR map, phases calculated, combined with the MIR phases and the combined phases used to produce a new starting map. This was repeated for 6 cycles, and the resulting map then used to determine a new envelope. The resolution of the averaged map used to calculate the envelope was increased in steps of 0.001Å $^{-1}$ after each iterative cycle.

The results of this phase improvement procedure can be seen in Table 5.22. The figure of merit improved from 0.57 to 0.80 after the 180 cycles. The resulting electron density calculated to 3.5Å using the combined phases was far superior to the original MIR map. There were also slight changes in the connectivity when compared to the map calculated using the phases determined by the program *SQUASH*. These points are discussed in greater detail in the next Chapter.

This approach seemed very promising so it was decided to attempt to extend the phases beyond 3.5Å resolution. As this was the limit of resolution of the *XDS* processed native dataset, it was necessary to use the native set collected on film. This was merged with the Natxds dataset (R_{merge} = 9.3%) to give a native set with overall completeness of 86% (67.5% between 3.16Å and 3.00Å). The initial electron density map was calculated at 4.0Å resolution and the phases incremently extended in steps of 0.001Å $^{-1}$ to 3.0Å resolution. The resulting map was compared with that obtained using the Natxds datset, and is discussed in the next Chapter. The phase shifts are given in Table 5.23.

5.8 Summary

This chapter has described the results obtained using a variety of datasets. The first electron density maps were calculated using data collected on the Xentronics and processed using *XENGEN*. The native set was incomplete so to allow the use of phase extension techniques it was combined with the 3.0Å resolution native dataset collected

Resln ($Å^{-1}$)	A <PH_{cmb} - PH_{mir}>	B <PH_{cmb} - PH_{mir}>
0.0000-0.0082	16.5	17.3
0.0082-0.0163	22.6	19.8
0.0163-0.0245	26.9	32.0
0.0245-0.0327	35.2	40.8
0.0327-0.0408	41.9	51.3
0.0408-0.0490	47.7	53.3
0.0490-0.0571	42.4	55.6
0.0571-0.0653	62.9	60.5
0.0653-0.0735	65.1	70.3
0.0735-0.0816	70.0	76.9
Overall	46.3	51.2

<PH_{cmb} - PH_{mir}> = Average Phase difference between combined and MIR phase

Table 5.22: Average Phase shifts after solvent flattening

MIR phases used as described in Section 5.XX.

A - Natxds used as native dataset.

B - Natxds combined with Natfilm.

on film at PX7.2, Daresbury. This dataset was then used for the solvent flattening and maximum entropy techniques described in Section 5.6. The Xentronics data was reprocessed using the *XDS* package. Another more complete native set was collected and used for all the work described in Section 5.7.

Chapter 6

Electron Density maps

6.1 Introduction

This Chapter contains the electron density maps calculated using the various phase sets described in Chapter 5. The electron density maps were displayed on an Evans & Sutherland graphics system using the program *O* (Jones *et al.* (1991)). Before discussing the maps, I will outline some of the difficulties in determining this structure using maps of limited quality and resolution.

The mother liquor of the HF-hCG crystals has a high salt concentration. The difference in density between the protein and solvent regions is far less than that found in, for example, a helical structure crystallized from polyethylene glycol. From the reported CD experiments the protein structure contains a large degree of aperiodic structure and some β-sheet. The lack of helical structure makes it significantly more difficult to check that the correct handedness of the protein has been assigned. α helices are right handed and provide a useful test for the handednes pf a structure solution. Handedness can also be discerned from the orientation of side chains, but this requires adequate resolution. The direction of the chain is difficult to establish from a 3.5Å resolution map of this quality (Lange (1992)).

A large number of disulphide linkages in the protein make tracing the protein mainchain initially more difficult. Once one or two disulphide positions have been assigned it should be much easier to determine the direction and course of the mainchain. However, if the disulphide linkages are not all known, this again is less straightforward.

The problem in defining an accurate protein envelope and of the low difference in density of the protein and solvent regions is probably exacerbated by residual carbohydrate still bound after the HF treatment. The mean fractional isomorphous difference (MFID) between the Asialo-hCG and HF-hCG, both collected on film at PX7.2, is 8.0%, *i.e.* of the order of the R_{merge} values for the datasets. One would expect an increase in carbohydrate content to increase the MFID between the two datasets, so as the Asialo-hCG contains approximately 9% more carbohydrate than the HF-hCG, a large proportion of the oligosaccharide moieties are presumably very disordered. This was further supported by the calculation of a difference Fourier with coefficients (F_{Asialo} - F_{HF-hCG}). There were no clear regions of density indicating the position of the carbohydrate. No account was taken of the possibilty of disordered carbohydrate when the envelope was determined and the effect this would have on the density distribution in the crystal. Furthermore there is the possibility that the 32 residues in the C-terminal region of the protein has been cleaved during crystal growth in up to 40% of the molecules. It may be that the fragments are held in an ordered position in the crystal lattice.

Several electron density maps have been calculated during the course of this project, and are discussed in turn in this chapter.

6.2 Use of *XENGEN* processed data

All maps discussed in this section used the data collected on the Xentronics area detector and processed using the package *XENGEN* (Howard *et al.*, (1987)). The native structure factor amplitudes used were those collected from a single crystal as

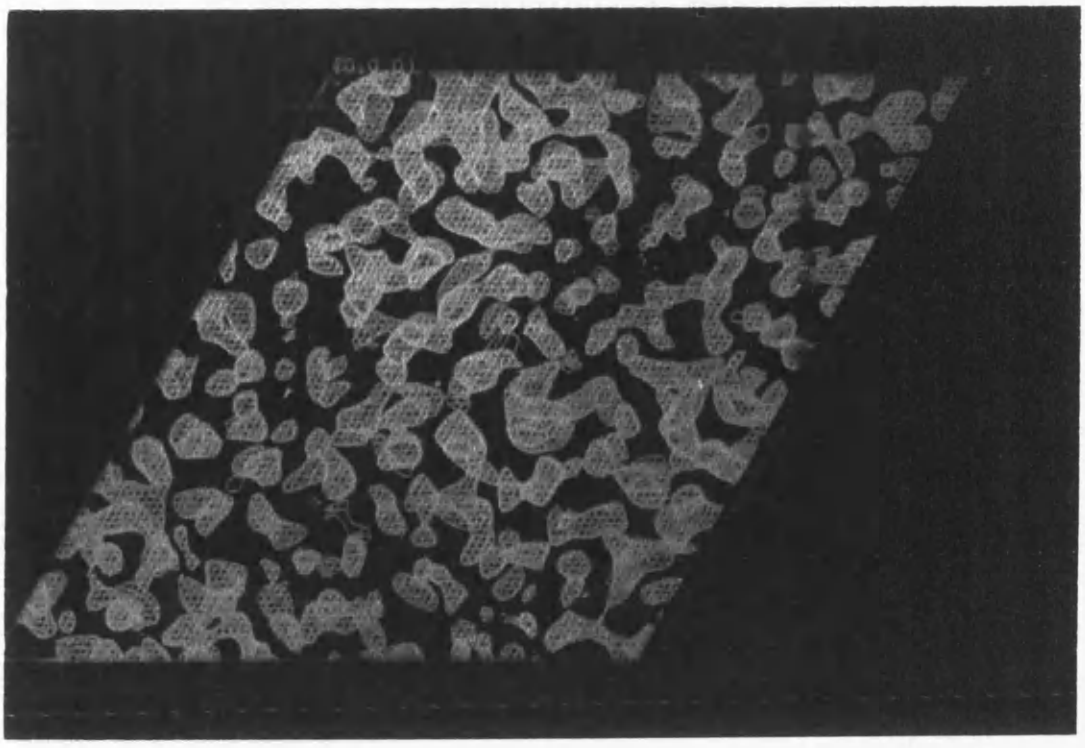


Figure 6.1: An asymmetric unit of initial 5Å electron density map as viewed down the *z*-axis.

described in Section 4.4.2.

6.2.1 Initial MIR map

The first electron density map calculated for HF-hCG was calculated using the phases described in section 5.3.6. The map was initially calculated to 5.0Å resolution. Figure 6.1 shows an asymmetric unit ($0 < x < 1$, $0 < y < 1$, $0 < z < 1/12$) of the map, viewed down the *z*-axis. It is possible to discern protein and solvent regions in the map. Figure 6.2 shows 3 asymmetric units of the same map with the *z*-axis vertical. Again, there are obvious regions of protein and solvent, but it is difficult to determine the protein - solvent boundary along the *z* -axis.

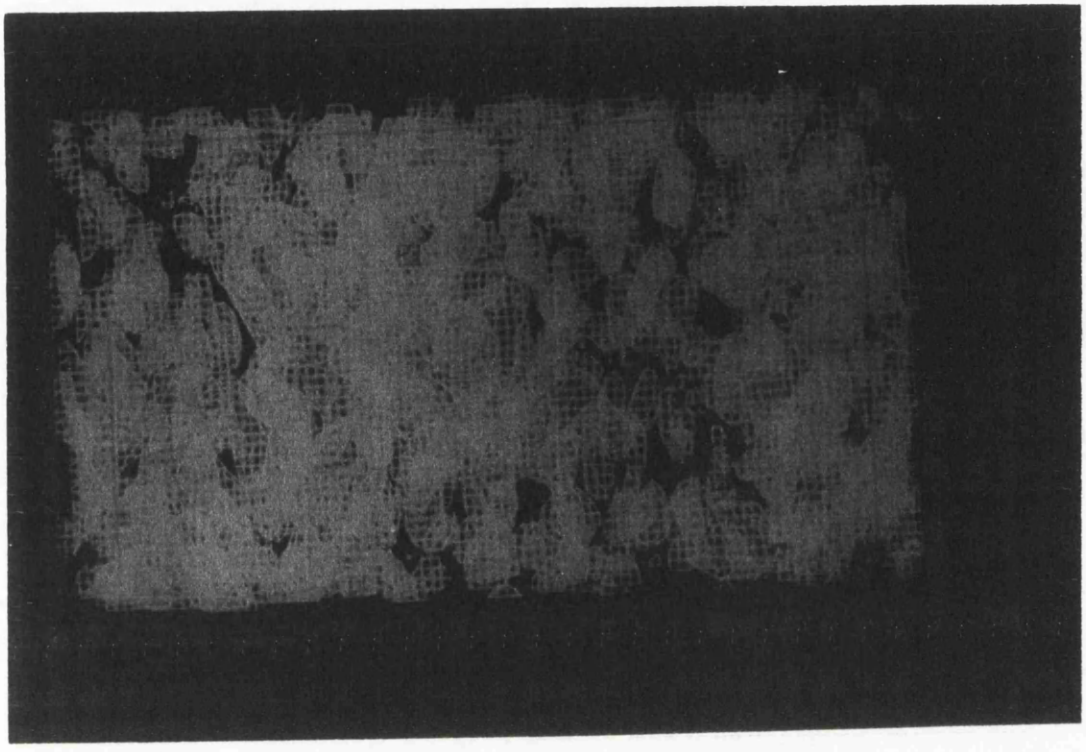


Figure 6.2: View of the 5Å electron density map - the *z*-axis is vertical in the plane of the paper.

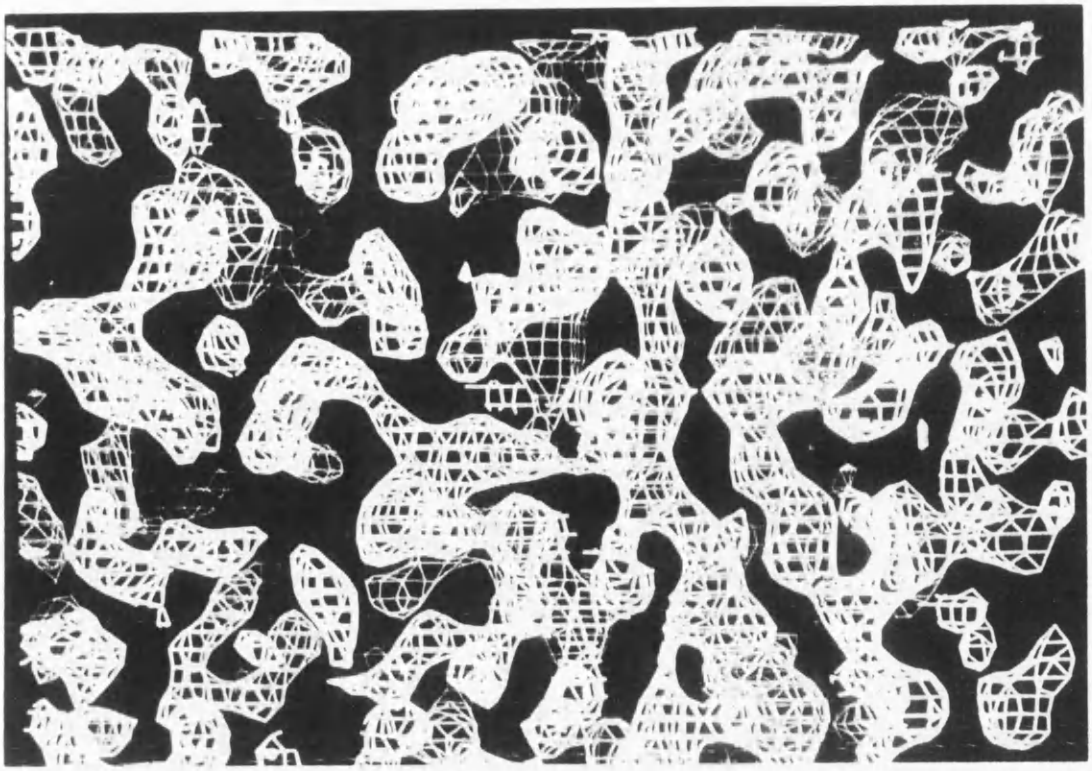


Figure 6.3: A portion of the 3.5Å resolution electron density map calculated using *XENTRONICS* processed data. The map is contoured at the 1σ level.

6.2.2 Initial 3.5Å resolution MIR map

The derivatives collected on the Xentronics detector and processed using the package *XENGEN* were used to calculate a 3.5Å resolution MIR map. A portion of this map is shown in Figure 6.3. It was not of a high enough quality to allow interpretation.

6.2.3 Solvent flattened electron density map

Figure 6.4 shows a section of the results of the solvent flattening procedure described in Section 5.6.1. The MIR phases used to determine the envelope were those used to calculate the map shown in Figure 6.3. The envelope was determined to 5Å resolution and applied to a 3.5Å resolution map. As expected, there is an obvious improvement in the relative densities of the solvent and protein regions. The protein-solvent boundary is more easily discernible. There is also some improvement in the connectivity in the

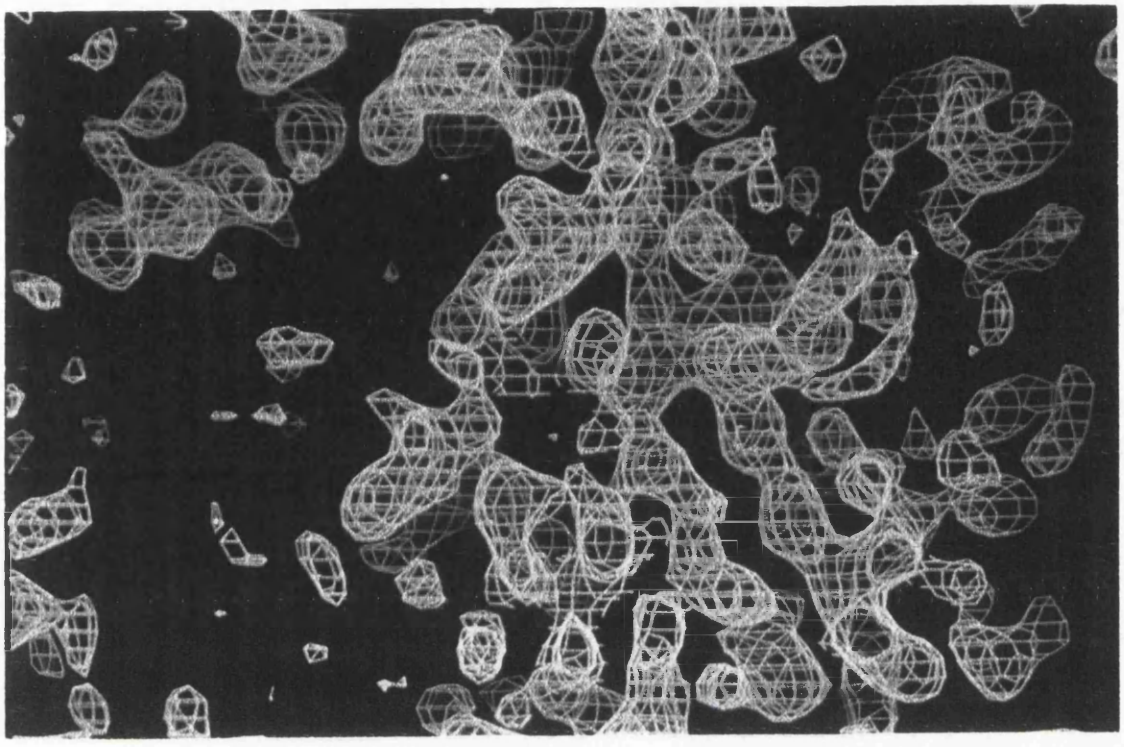


Figure 6.4: A portion of the initial solvent flattened 3.5Å map.
To allow direct comparison between maps, the view is identical to that for Figure 6.3.

protein region compared to the original MIR map. Attempts were made to build into this solvent flattened map. Portions of polypeptide chains were fitted to the better regions of the map (Figure 6.5). In total it was possible to fit 131 residues. The calculated phases from this model were combined with the phases from the solvent flattening procedure using the program *SIGMAA* (Read (1986)). The mean figure of merit for the partial structure was 0.43. A least squares fit to a plot of ln(SigmaA) v. $\sin^2\theta/\lambda^2$ yielded a co-ordinate error of 0.74ÅÅ new map was calculated and inspected. These phases were used to calculate a new electron density map. There were slight changes to the side chain density, but no change to the overall connectivity of the map. The poor regions of thei solvent flattened map which had not been built into were not improved in the combined map. No further building to this map was attempted.

6.2.4 *MICE* electron density map

The maximum entropy phase improvement technique as implemented in *MICE* was used at this stage of the project to improve the MIR phases obtained from the *XENGEN* processed data. Figure 6.6 shows a portion of the resulting electron density map phased with combined phases from those calculated using *MICE* and the original MIR phases. There is a slight improvement in the map, but again it was not possible to interpret it.

6.3 Use of *XDS* processed datasets

All the electron density maps discussed in this section use the data processed with the *XDS* package and phases calculated as described in Section 5.7. The MIR phases obtained from *MLPHARE* were refined using the program *SQUASH* and the solvent flattening programs within the *CCP4* suite. Phase combination with partial structure was also attempted. The MIR electron density map calculated with the *XDS* processed data was much improved on the MIR map from the Xentronics data shown in Figure 6.3. There is a greater degree of connectivity and an improvement in side chain density.

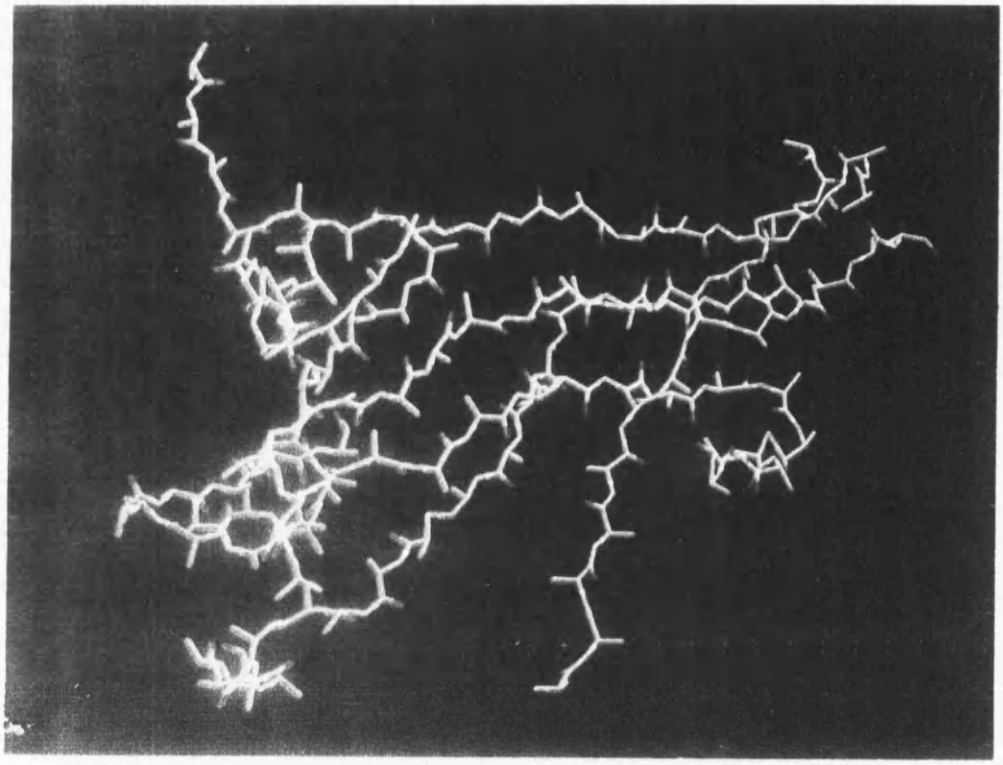


Figure 6.5: Polypeptide chain built into solvent flattened map and used for phase calculation and combination.

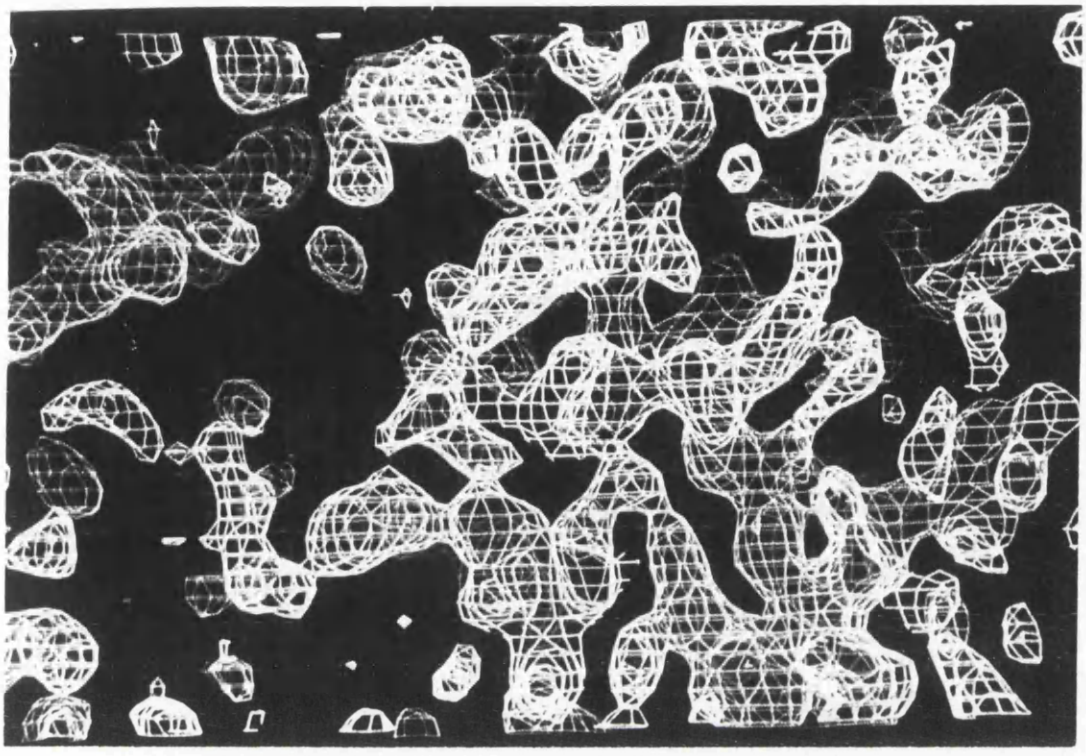


Figure 6.6: Portion of the map calculated with phases resulting from use of *MICE*. The view is identical to that used for Figures 6.3 and 6.4. The calculated phases were combined with the MIR phases to produce this 3.5Å resolution electron density map.

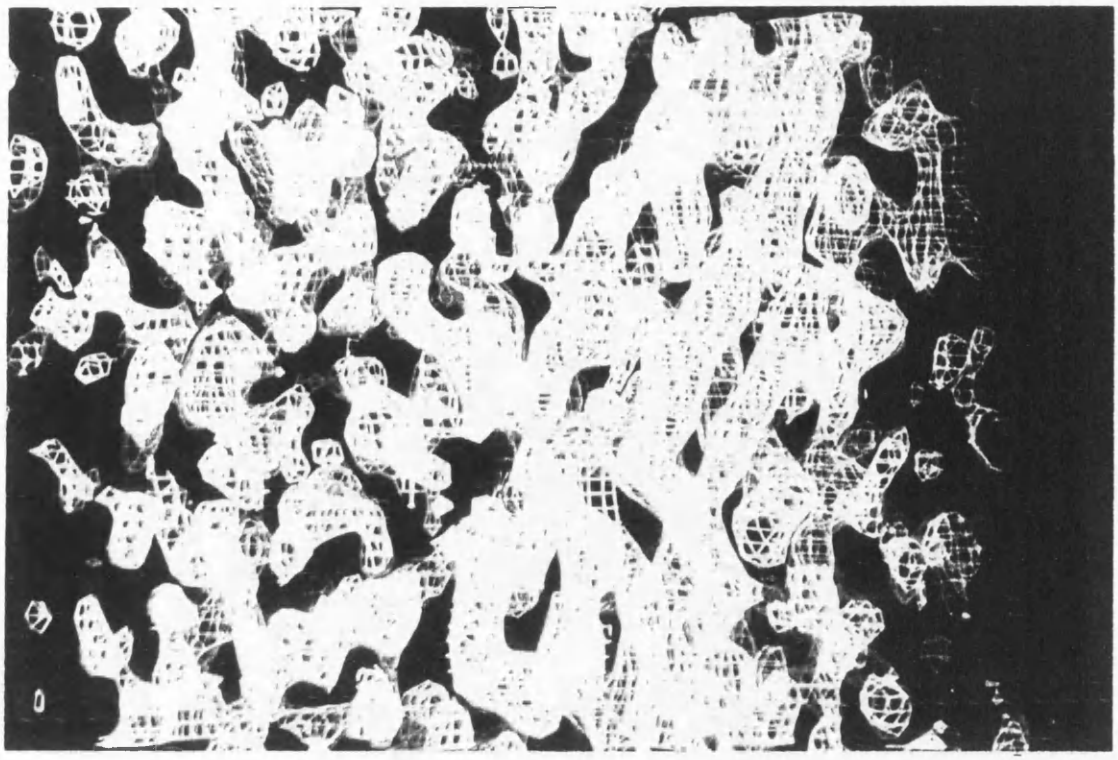


Figure 6.7: Comparison of MIR maps

The map shown in brown is calculated with the native dataset from Xtl1,Sc1 and that in blue with data collected from Xtl2,Sc1. Both are calculated with identical phases and contoured at 1σ levels. The view is of the same region of density as shown in previous Figures.

There is more β-sheet like density present in the map (see Figure 6.10).

6.3.1 3.5Å resolution map calculated using *XDS* processed data.

All the data collected on the Xentronics was reprocessed as described in Chapter 4 using the *XDS* program. The initial map was calculated using the native data set from the first scan run (Xtl1 Sc1). Subsequently, another more complete native dataset was collected (Xtl2, Sc1) and used for the coefficients of a new electron density map. Figure 6.7 shows a portion of the maps, calculated with identical MIR phases as described in Section 5.7. The map calculated using the native set from Xtl1,Sc1 is shown in brown

and the map using structure factors from the dataset Natxen (Xtl2,Sc1) is shown in blue. Both maps are contoured at 1 σ level. There are slight, but significant differences. The major features of the maps are the same, but that calculated with the less complete native set (shown in brown) shows poorer connectivity which would make tracing more difficult.

6.3.2 Improved phases using *SQUASH*

The MIR phases were used to calculate an envelope as described in Section 5.7.1. The calculated phases from *SQUASH* were used to calculate the map also shown in Figure 6.8. This can be compared with the MIR map shown below. The *SQUASH* map is much less noisy, but it was still difficult to determine the molecular boundary along the *z* axis. The option *BONES* within the program *O* which skeletonizes electron density was used to try to trace a peptide chain in the molecular asymmetric unit. Figure 6.9 shows one such asymmetric unit surrounded by its symmetry equivalents. As can be seen, there are three main areas where it is difficult to determine the molecular boundary. The β-sheet like density is improved in this map relative to the MIR map (Figure 6.10).

6.3.3 Fitting of primary sequence

An attempt was made to fit the primary sequence to the *SQUASH* electron density map. The disulphide linkages were used as a starting point. Even though the full disulphide assignments have not been made, the known linkages limit the possibilities for the relative positioning of portions of the sequence. Several regions of the map showed some "H" like density (Figure 6.11). The density for the disulphide linkage should be greater than that for the polypeptide backbone, so a map contoured at 3σ was calculated and superimposed on the 1σ map. Also, use was made of the anomalous native data that had been collected. It was hoped that some features due to the sulphur atoms might be present. Anomalous differences were very poorly measured, but maps

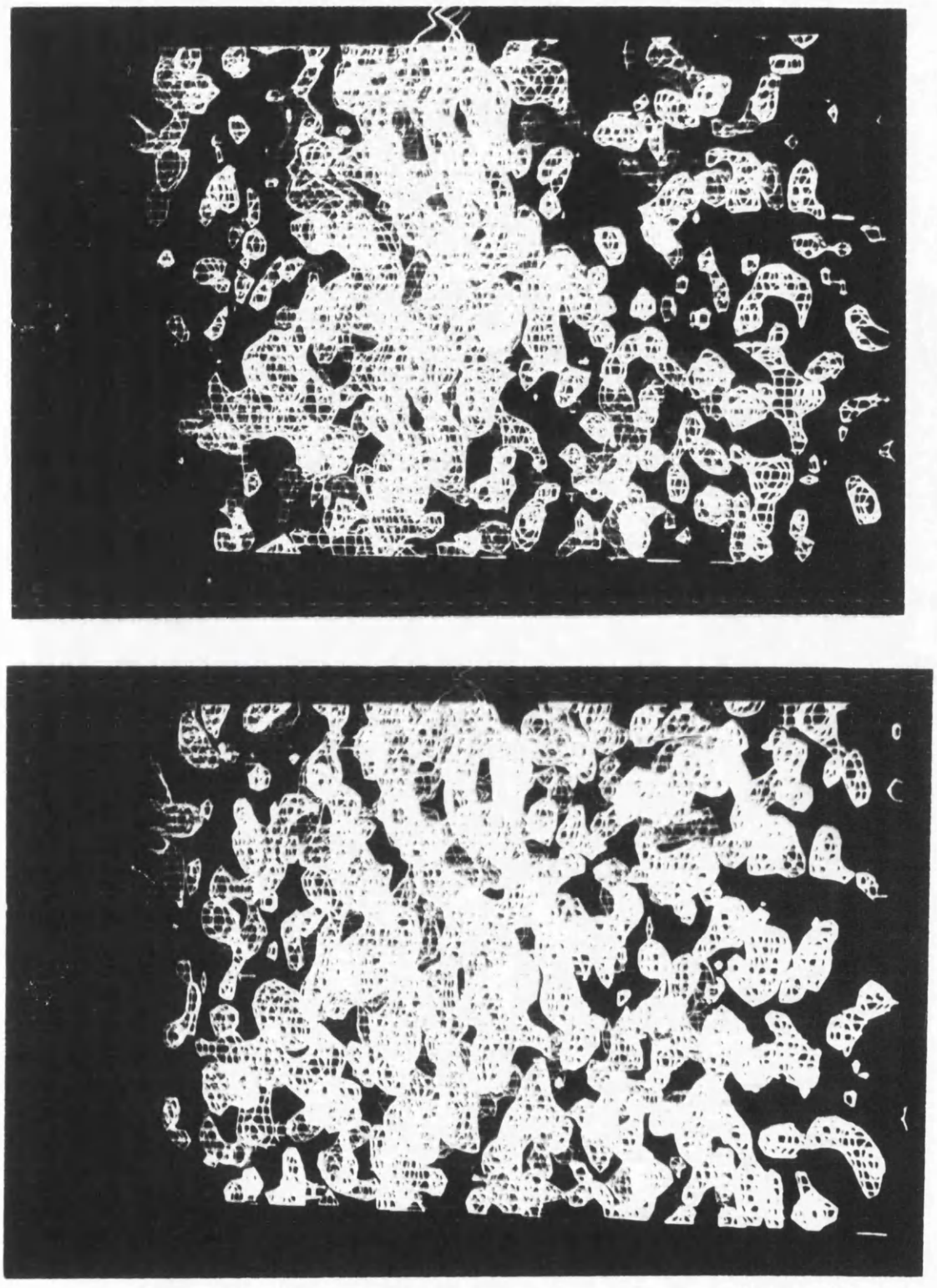


Figure 6.8: Comparison *SQUASH* and MIR phased maps

The electron density map shown in brown was calculated using phases calculated using the program *SQUASH*. The lower photograph shows an identical region of the map calculated using MIR phases. Both are contoured at 1σ levels.

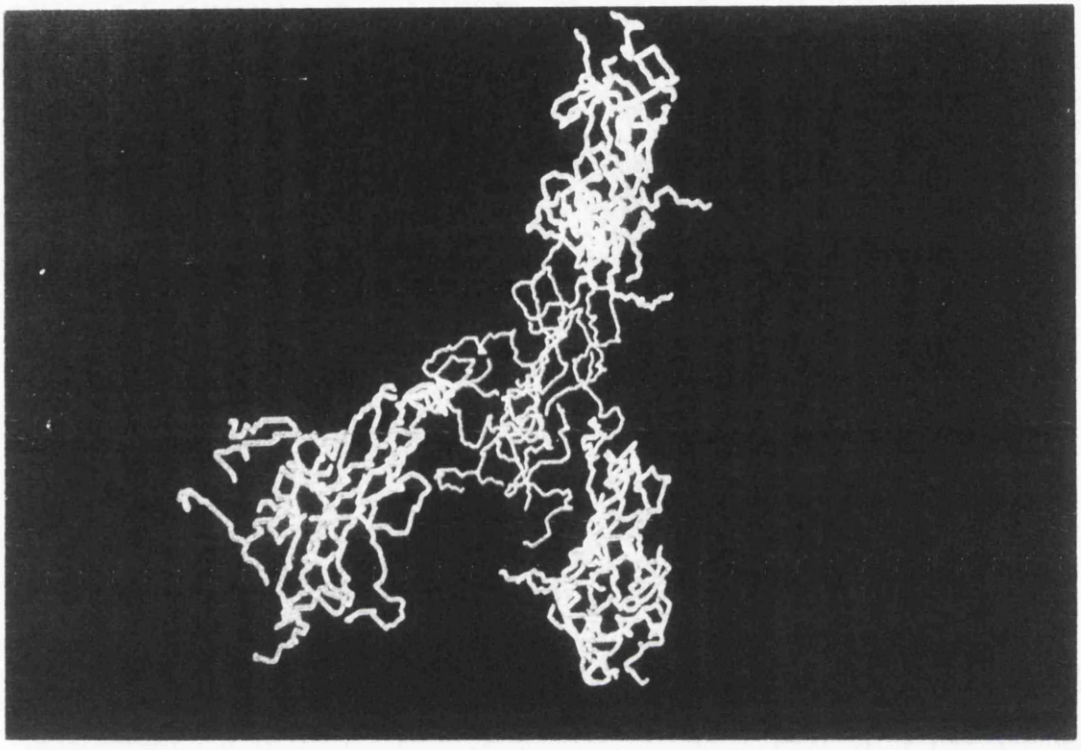


Figure 6.9: Skeletonized electron density

BONES structure from the electron density map calculated using the phases from *SQUASH*. The blue skeletons are symmetry equivalents of the pink skeleton.

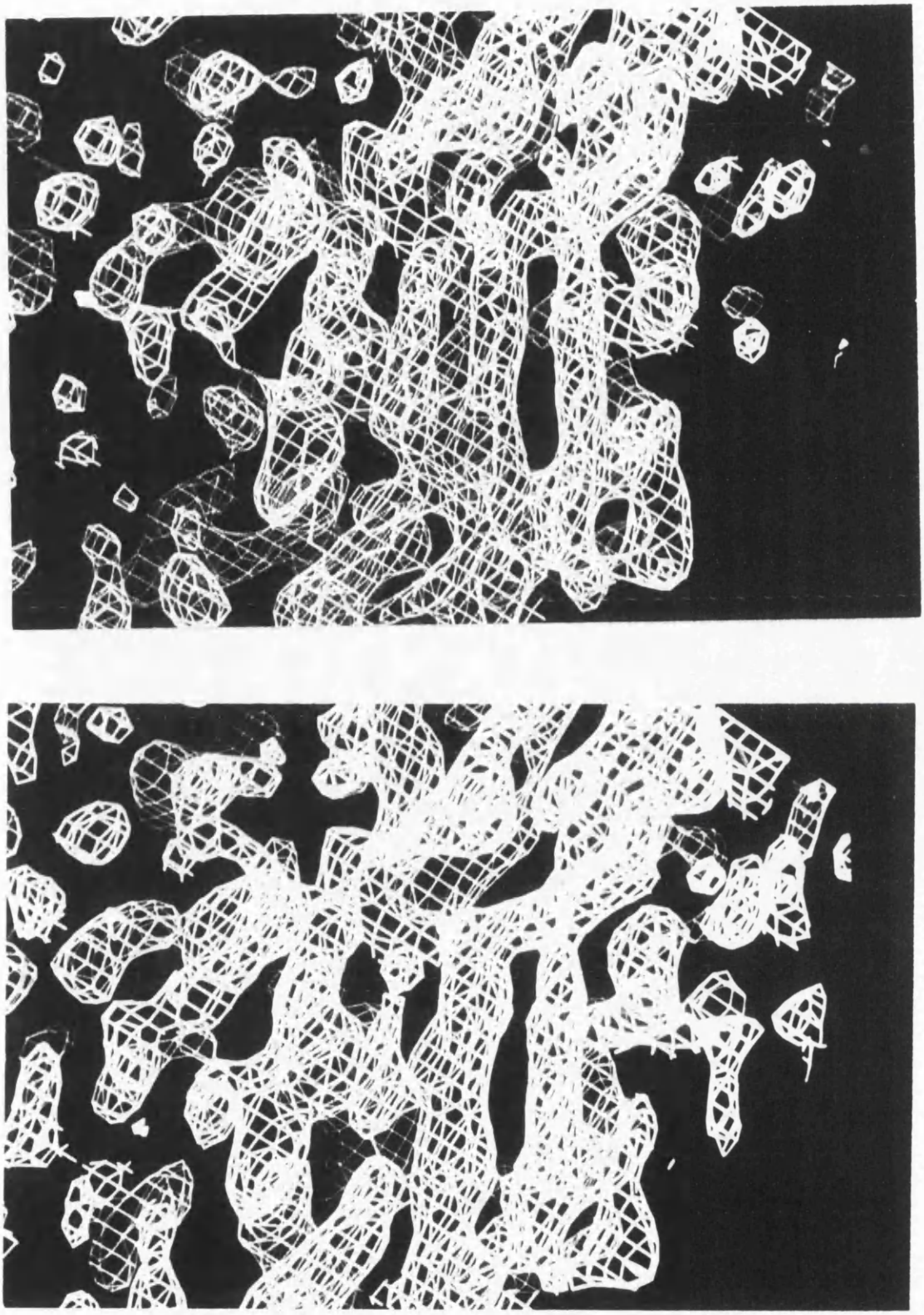


Figure 6.10: Detail of *SQUASH* and MIR maps

The map in brown is that calculated with phases from *SQUASH*. That in yellow is the MIR map. An improvement in the connectivity is apparent after solvent flattening and histogram matching. Maps contoured at 1σ levels.

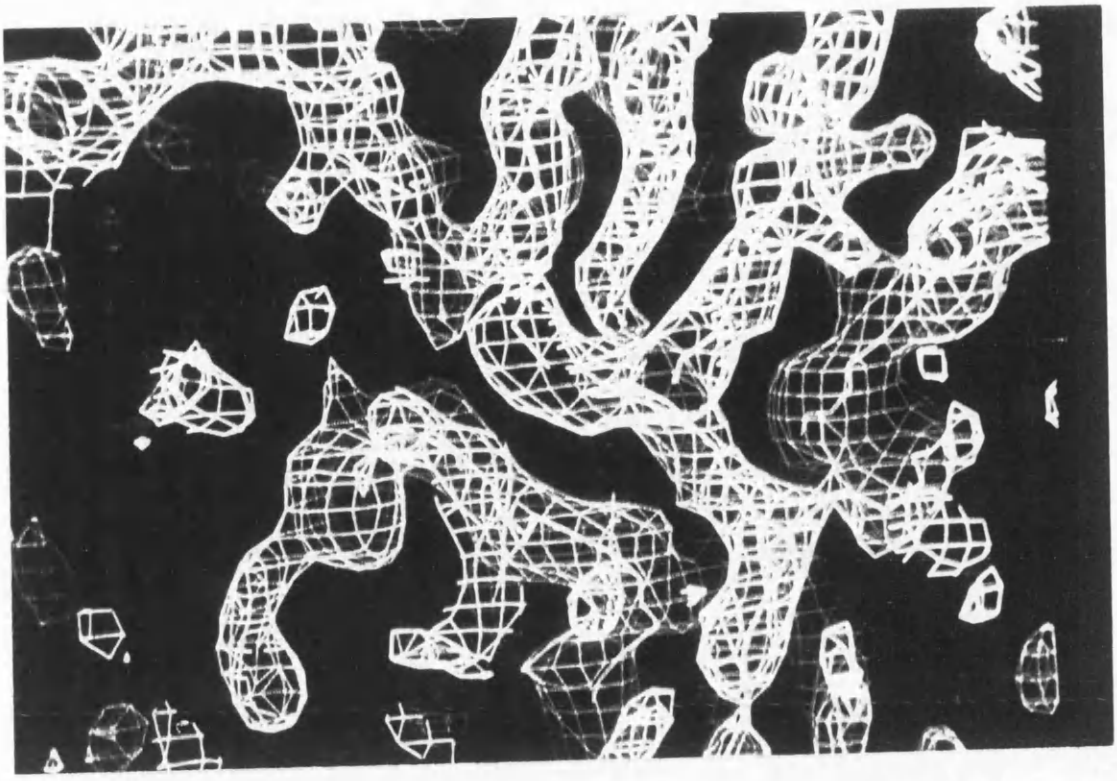


Figure 6.11: Detail of the 3.5Å resolution *SQUASH* map

were calculated from both the *XDS* processed dataset and that collected on film. In this way it was hoped that true peaks due to anomalous differences would coincide and the others were assumed to be noise. Care was taken to reject statistical outliers from the scaling in the calculation of the Fouriers.

In two regions of the map, the maximum isomorphous peaks, peaks on both anomalous maps and "H" connectivity in the *SQUASH* map coincided. One of these positions was midway along a long strand of continuous density with no further ambiguity in the density. Analysis of the primary sequence (Figures 1.2 and 1.3) showed two portions of the sequence to be the most likely candidates including Cys β-57 or Cys β-72. The direction of the chain was not obvious, so the four possible sequences were fitted to the map using the *LEGO* option in *O*. Val β-44 - Ser β-66, Pro β-70 - Val β-48, Tyr β-59 - Ser β-81 and Ala β-85 - Arg β-63 were built in and the side chains judged for quality of fit. The β-44-66 and β-85-63 builds were rejected. Large sidechains did not correspond to the density and, for other sidechains, density remained unaccounted for. The β-70-48 portion was rejected because of steric problems fitting Tyr β-59.

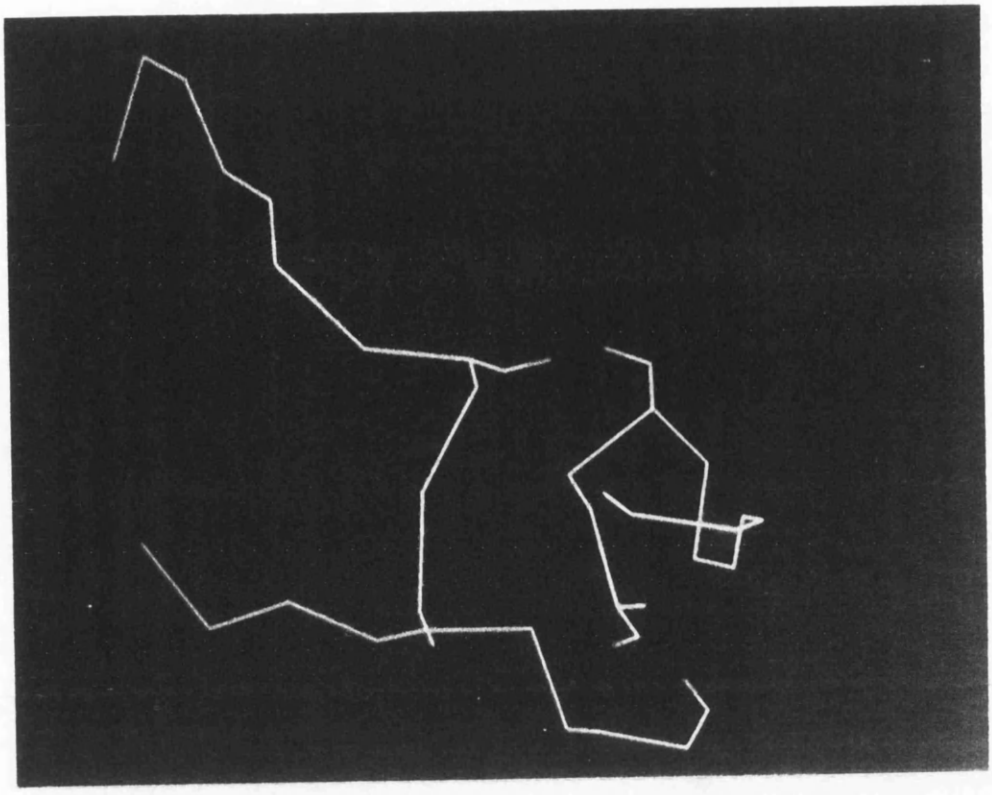


Figure 6.12: Portion of hCG sequence fitted to the *SQUASH* map

The remaining chain β-59-82 was accepted as the most likely candidate and used as a basis to continue the fitting. Cys β-72 has been unequivocally assigned to be linked to Cys β-23 (Mise (1981)). The portion of the sequence around Cys β-23 was fitted (in both directions) to the connected density. In one direction, Cys β-26 was found to be positioned at the only other part of the map with good anomalous peaks. This residue is known to be linked to Cys β-110 (Figure 6.12). The fitting of residues β-111 - 100 followed. Unfortunately, at this stage problems were encountered. It was impossible to fit the sequence in such a way that Cys β-100 was positioned to bind to Cys β-93, a well established linkage. Further, Asp β-111 was on the edge of the molecule, but not near the density thought to be the α-subunit. hCG retains biological activity when Asp β-111 is cross-linked to Lys α-49, and so the model was thought to be unsatisfactory.

6.3.4 Fitting of polyalanine chain

As the primary sequence could not be fitted to the present best map, portions of polyalanine chain were fitted to the map. The side chain density was not good enough to determine unequivocally the direction of the chain even in the better regions of the map, but residues were fitted where possible. A total of 118 alanines were fitted, *i.e.* approximately 25% of the expected total protein atoms (Figure 6.13). Phases were calculated from this partial structure using the program *SIGMAA*. This model had an rms co-ordinate error of 1.17Å and a mean figure of merit of 0.27. These calculated phases were combined with the solvent flattened phases. An electron density map was calculated using the combined phases for use as a new starting map for further interpretation. However, due to the limited fraction of structure included for phase calculation and combination, the combined map was little different from the *SQUASH* map. In order to make use of partial structure it will be necessary to build a larger percentage of the protein into the map, make more use of side chain density and improve the geometry of the model by refinement.

6.3.5 Solvent flattened maps

The MIR phases calculated using *XDS* processed data were improved using the solvent flattening programs in the *CCP4* suite as described in Section 5.7.2. The resulting map was very similar to that produced by the use of *SQUASH*. The small stepwise increase in the higher resolution limit used in the calculation of the envelope and maps did not result in a markedly improved map. Figure 6.14 shows a portion of this map which can be compared directly to the MIR and *SQUASH* maps in Figure 6.10. The use of solvent flattening and phase extension to 3.0Å resolution also did not produce a more interpretable map than any of the previous maps.

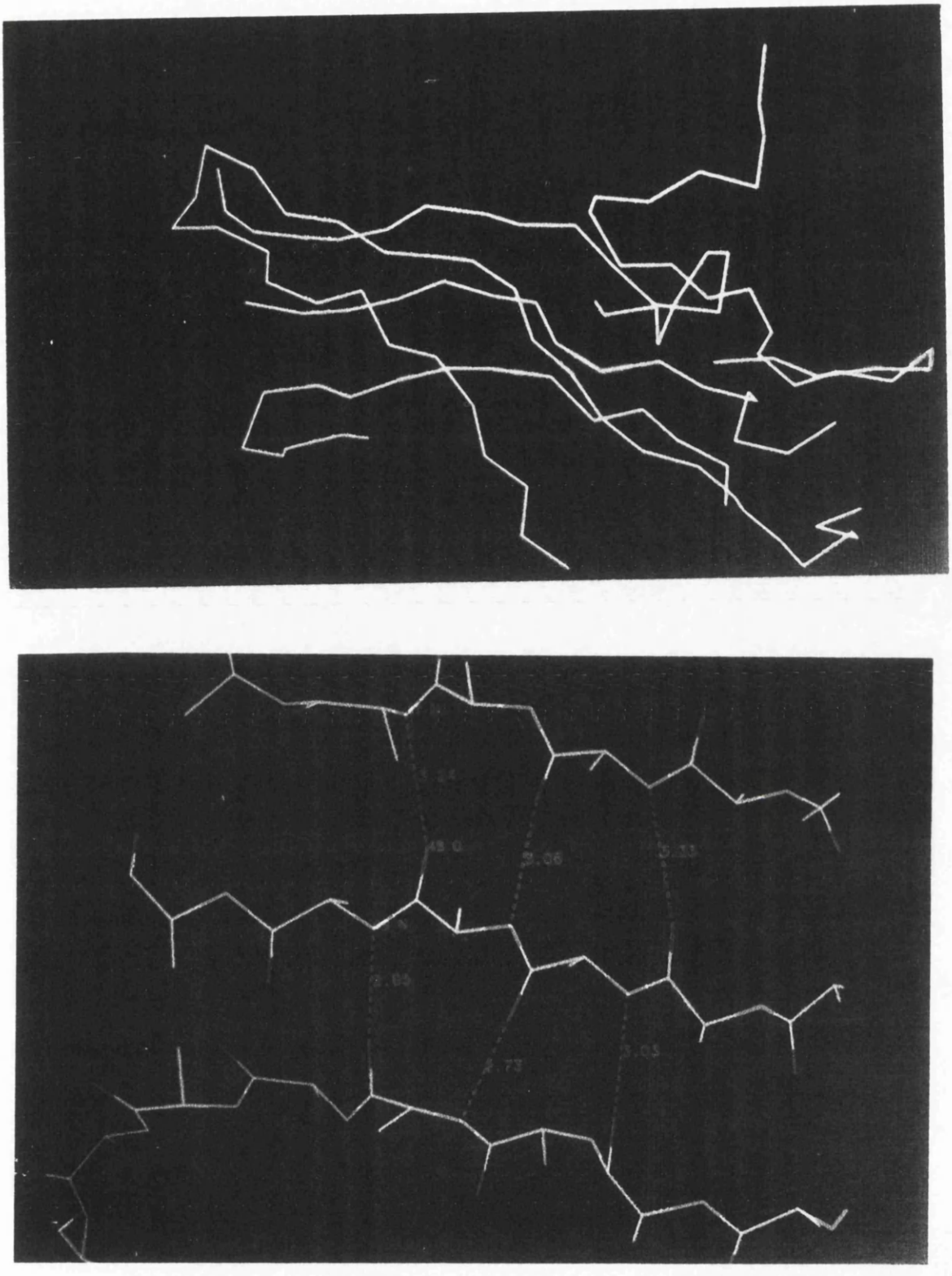


Figure 6.13: Polyalanine model from the *SQUASH* map.

The upper photograph shows a Cα trace of all the partial structure fitted. The lower shows detail of the β-sheet portion of the model.

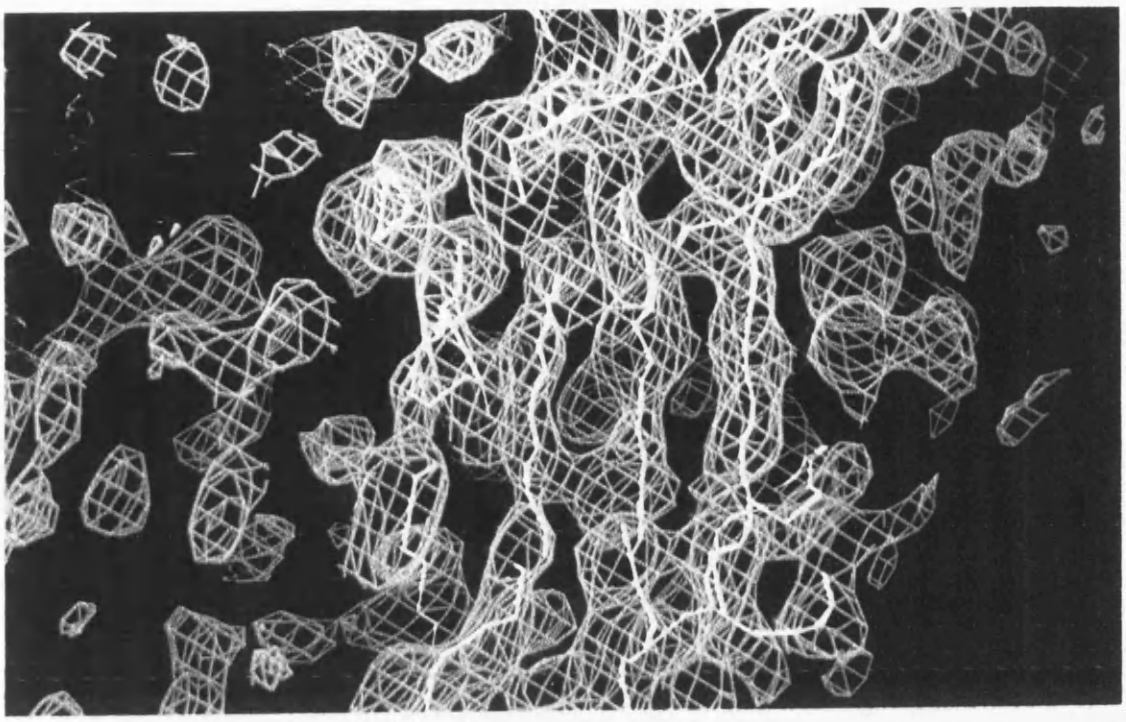


Figure 6.14: phases from *CCP4* solvent flattening - 3.5Å resolution map

6.4 Summary

In this Chapter I have shown similar portions of the electron density maps from the various datasets to allow comparison of the quality of each map. Maps resulting from phase improvement techniques have also been included. The first 3.5Å resolution map calculated using *XENGEN* processed data and the three derivatives given in Table 5.10 showed definite regions of protein and solvent and was much improved by solvent flattening and maximum entropy techniques. The connectivity in the protein region was improved, but full structure determination was not possible using any of these electron density maps. At this stage, *XDS* was used to reprocess the data. More native and derivative data were collected and also processed using this program. The resulting MIR map was superior to any of the preceding maps. This was due to both the improvement in the quality of the higher resloution data, and the inclusion of more useful heavy atom derivatives. Once phase improvement methods were applied to the native phases, the resulting maps showed clear regions of β-sheet structure. Attempts

to interpret this map directly were made but were unsuccessful due to many regions of poor density and ambiguity in the connectivity of the density. If the folding motif were known, interpretation would have perhaps been possible. The rat CD4 structure has been solved from a 3.5Å resolution MIR map which contained good electron density for the β-sheet regions with clear density for the bulky side chains (Lange (1992)). However, in that map some connectivity was not clear and the loops in general were disordered. It was possible to interpret the map because the folding motif (in this case the immunoglobin fold) was known and could be used as a model for interperetation of the map. As this was not possible for hCG, a polyalanine chain was fitted to the electron density map. The partial structure was used to calculate phases for recombination with those used to calculate the map. Only small changes in the map was apparent. The quality of the calculated phases could be improved by refinement of the partial structure and the fitting of more structure but this was not attempted. In the next Chapter I discuss what I consider the best paths to take to allow the calculation of an interpretable electron density map.

Chapter 7

Discussion

In this chapter I intend to summarize the reasons for this project, the results obtained and ideas for future work to enable the full structural determination of human Chorionic Gonadotropin.

Chapter 1 reviews the current understanding of the structure of hCG and related glycoprotein hormones. The primary sequence had been determined for several species and a high degree of homology established both between the hormones and between the different species. The secondary structure had been probed using circular dichroism and the tertiary and quaternary structure have been studied using a variety of techniques. Several interesting structural details remain unanswered, *e.g.* the unequivocal assignment of the 11 disulphide linkages had not been possible because of the lack of proteolytic or chemical cleavage sites within the clusters of cystine residues. Much work had been carried out to identify the regions involved at the α-β subunit interface and receptor binding. Progress had been made in understanding these regions, but some of the results are apparently contradictory. In order to fully understand the role played by specific residues and regions, it is essential to have a full knowledge of the three dimensional structure of hCG. This would also reveal the which cysteines are linked in the disulphide bridges.

The prerequisite for an X-ray crystallographic study of a protein is the availability of

suitable crystals. Crystallization of fully glycosylated hCG had been attempted for many years with no success. This reason for this failure was thought to be due to the heterogeneity of the carbohydrate. The effect of treatment of hCG with anhydrous hydrogen fluoride resulted in the removal of approximately half the carbohydrate. The protein backbone was uneffected as evidenced by circular dichroism and receptor binding studies (Keutmann *et al.* (1983)) A more detailed analysis of the remaining carbohydrate revealed that the O-linked oligosaccharides were still intact, but that the N-linked oligosaccharides had been cleaved at the mannose residues (Ryan *et al.* (1987)). As it was known that the tertiary structure was not disrupted by this treatment, our laboratory attempted to remove the oligosaccharide units by treatment with anhydrous hydrogen fluoride as described in Chapter 2. It was thought that removal of a large proportion of the carbohydrate might allow the growth of crystals suitable for X-ray crystallographic studies, and allow the determination of the three dimensional structure of the protein portion of hCG.

The resulting protein, HF-hCG, was used in crystallization trials. We were able to crystallize this protein using the standard hanging drop method. Trials were attempted using a number of buffers and precipitating agents. Success was achieved using Sodium Acetate at a pH close to the isoelectric point of the protein. The only unusual factor was the high concentration of protein used: 35-40mg/ml. Crystals grew over a period of six to eight weeks and initially were of very poor morphology. Diffraction was observed to 9Å resolution. Trays set up three months earlier using these conditions with a 20mg/ml protein solution were inspected and very small crystals were found. No further growth occurred. The conditions were optimised by varying the protein concentration and pH of the buffer and the resulting crystals grew as sharp hexagonal bipyramids within two months, to a maximum size of approximately 0.4mm in each direction. Diffraction was observed to 3.5Å on a Rigaku rotating anode. The crystals were characterized using precession photography and found to have the space group $P6_122$ or enantiomer with cell parameters a = b = 88.68Å and c = 177.2Å (Harris *et al*, (1989)).

The question of carbohydrate heterogeneity was considered. Analysis of protein from crystals on an SDS gel revealed a degree of heterogeneity which obviously had not prevented crystal growth. Attempts were made to crystallize intact hCG using these conditions, but non-crystalline precipitate formed. At this time a group at Columbia University, New York used our conditions and obtained crystals of Asialo-hCG, a species produced on removal of the terminal sialic acid residues by the specific action of the exo-glycosidase, neuraminidase (Lustbader *et al.*, (1989)). The crystals had identical morphology and characteristics to the HF-hCG crystals. Lustbader *et al.* undertook a more detailed analysis of the two forms. They discovered that in the HF-hCG crystals some cleavage of the protein backbone occurred during crystal growth at the acid labile regions βAsp112-Pro113 and βAsp117-Ser118. This has been reported at 45% and 20% respectively. There is the possiblity that this is caused by some residual HF, but as 7% cleavage was also found in Asialo-hCG crystals at β-Asp112-Pro113, it is also possible that it is a result of the acidic crystallisation conditions (pH 4.2-4.5). As both these species still contain a large degree of carbohydrate and the HF-hCG remains heterogeneous, it is unlikely that either of these factors is preventing the crystallization of intact hCG. It is likely that the overall charge of the glycoprotein is responsible. It is the removal of the negatively charged sialic acid residues, either specifically using neuraminidase, or non-specifically using anhydrous Hydrogen Fluoride, that allows the growth of crystals suitable for X-ray crystallographic studies.

Once the crystallization conditions had been optimised, the next step in the structure determination was the solution of the phase problem. As the three dimensional structures of the other glycoproteins are unknown, this could only be achieved using the multiple isomorphous replacement method. The screening of putative derivatives was initially conducted using the precession method, but this was superseded after the acquisition of a Xentonics area detector. As can be seen in Chapter 3, the precession technique was unsatisfactory. Crystals seldom lasted long enough in the beam to allow a μ = 10° precession photograph to be taken, so many of the experiments were inconclusive. However some useful results were obtained - promising changes were observed for derivatives using K_2PtCl_4, $KAuCl_4$ and $UO_2(NO_3)_2$ although the resolution of the

diffraction photographs did not allow study of higher resolution reflections to categorically determine specific binding of these heavy atoms to the protein. Also during this period we were able to discount many potential heavy atom compounds because of lack of solubility in the mother liquor, severe cracking of the crystal upon entry to the heavy atom solution or because of unsatisfactory diffraction.

As little progress had been made by this point, it was decided to try to use another buffer system for the soaks. The medium in which the crystals grow is far from perfect for successful heavy atom derivatives. The salt concentration is high (30-36% Ammonium Sulphate) and the pH range is very specific (pH 4.2-4.5). Attempts were made to transfer the crystals to a lower salt concentration but they proved too fragile to allow either an immediate transfer, or even a slow decrease in the salt concentration. Attempts to transfer the crystals to solutions of PEG also failed. Results of previous MIR experiments (Blundell & Johnson (1976)) were considered and the amino acid sequence was studied to determine if any particular heavy atom compound may be particularly suitable. One of the most commonly successful combinations, mercury binding to free cystine, was not a viable option as all the cysteine residues in hCG are involved in disulphide linkages. However, hCG does contain four histidine residues. At pH 4.4 these would most likely be protonated and unavailable as heavy atom binding sites. As Sodium Acetate is not a high pH buffer, another buffer system had to be found. The group at Columbia University had grown crystals in Sodium Citrate Buffer, pH 4.3-4.5, so this was our favoured choice. We discovered it was possible to transfer crystals directly to the Citrate Buffer, pH 6.3. Some crystals suffered slight cracking, but generally few deleterious effects were observed. Several heavy atom compounds were tried at this pH, but primarily silver and mercury salts as these had been reported in the literature to bind to histidine (Blundell & Johnson, (1976)).

Before the Xentronics area detector was installed, the available method of data collection was at the S.R.S., Daresbury. Datasets of derivatives that appeared promising were collected on film on station PX7.2. The data collected in this manner were satisfactory but, because no definite derivatives had been discovered, some datasets were collected

which either proved to be non-isomorphous with native, or to have no heavy atom bound. The crystals did not diffract for as long in the Synchrotron beam as they had in a beam from a rotating anode, which meant that it was necessary to use more than one crystal per dataset. However, diffraction was to higher resolution than had been witnessed on the laboratory source. The native crystals diffracted to 2.5Å resolution and the heavy atom soaked crystals to a maximum of 3.0Å.

When the Xentronics area detector became available the heavy atom derivatives were recollected and the screening experiments continued. This was far more convenient as crystals could be soaked, mounted and used immediately. Diffraction was to a lower resolution, typically 3.5 - 4.0Å resolution, but a whole dataset could be collected from a single crystal and the data processed quickly.

The first MIR map of HF-hCG was calculated using data collected on the Xentronics and processed using *XENGEN*. The 5Å map clearly showed solvent and protein regions. The derivatives used were $K_2Pt(CN)_4$, $KAu(CN)_2$ which shared the same binding site, and an $AgNO_3$ soak which had been conducted at pH 6.3 in Sodium Citrate buffer. The silver and gold derivatives were recollected to the limit of diffraction and 3.5Å resolution phases calculated. This map again showed clear regions of protein, but the connectivity was poor and Cα tracing was not attempted. Analysis of the data with resolution as shown in the Tables in Chapter 4 reveals that the higher resolution data suffered from high R_{merge} values and this is reflected in the Figure of Merit analysis given in Chapter 5.

Phase improvement methods were employed to try to increase the quality of the higher resolution phases. The solvent flattening technique was utilised and this resulted in an improved, though still uninterpretable map. At this point I was given the opportunity to try the use of the maximum entropy approach to improve and extend the phases. To maximise the use of the data available, the *XENGEN* processed native data was combined with the native dataset collected on film and a more complete 3.0Å resolution native set obtained. It was hoped that the maximum entropy approach would allow the extension of the phases to this resolution. This work was carried out with Dr.

Charles Carter (North Carolina University). Several problems were encountered, the greatest of which was the inaccuracy of the envelope determined automatically using the algorithm of A. Leslie implemented in the *CCP4* suite of programs. A new envelope was determined manually and the results from this looked promising. Unfortunately due to lack of time the technique was not fully exploited. After trying these phase improvement methods without being able to produce an interpretable map, it was decided to collect higher resolution data.

Three trips were made to make use of the FAST detector on station PX9.6 at the S.R.S., Daresbury. Similar problems to those encountered at station PX7.2 were experienced. On two occasions there were severe problems with the beam and no data collection was possible. On the third, data collection was attempted but the crystals suffered very severe radiation damage. Only a few degrees of data could be collected per crystal. Also diffraction was not to a significantly higher resolution than that achieved in-house. This could have been due to problems in transporting crystals. The data were processed using the program *MADNES* but found to be very poor and incomplete. The FAST dataset was scaled to a Xentronics native dataset and found to be non-isomorphous and unuseable.

The data collected on the Xentronics were reprocessed using the package *XDS*. The quality of the higher resolution data was much improved on that processed using *XENGEN*. A new native dataset was collected which was complete to 3.5Å resolution. The importance of the quality of the native dataset is vividly illustrated in Chapter 6. Despite the limitations of resolution, this Xentronics data gave the best MIR map so far. The phases were improved using solvent flattening and histogram matching methods and the resulting map showed β-sheet structure. Although it was not possible to trace a continuous chain through the map, a Cα trace was fitted to the better regions and a polyalanine chain was built in.

The attempts to fit the sequence to the map were unsuccessful. Full interpretation of the map was dependent upon assigning disulphide positions. "H"-like connectivity in the map were studied in conjunction with use of native anomalous maps. Two regions of

the map showed disulphide like density connectivity which also coincided with peaks from the anomalous maps and these were used as the starting point for interpretation. The position and size of side chain density for neighbouring residues were studied. Several possibilities were built into the map. One of the ambiguities in connectivity was in the middle of a β strand. The amino acid sequence, bridge assignments and secondary structure prediction results were used to narrow the choice. Two particular cysteine residues seemed particularly likely. The four possible models (the extra parameters being due to the direction of chain) were fitted using the auto build option in *O*. Of these two were immediately discounted because of poor fit, in particular the positioning of β Tyr-59. Once the first strand was fitted, and the disulphide assignment β Cys 23 - 72 assumed correct, then the region around Cys 23 could be fitted. This in turn led to the disulphide Cys 26 - 110 and to the next portion to be fitted. At all times the results of the biochemical work were considered. A total of 35 residues were fitted before sufficient problems were encountered to suggest this interpretation was incorrect.

As the map was apparently not of a high enough quality for direct interpretation, a polyalanine chain was fitted to the density. A total of 118 alanines could be fitted, *i.e.* approximately 25% of the expected polypeptide chain. This model was used to produce calculated phases which were combined with the solvent flattened phases.

Future Work.

In order to fully determine the structure of hCG several possible options are available. The list below gives what I consider to have been the major causes of problems in this structure determination, and the possible solutions.

Shortage of Crystals.

In the later stages of this project there has been a shortage of crystals which has prevented repetition of the data collection experiments. In some cases ($AgNO_3$ and $KAu(CN)_2$) the buffer system used to collect higher resolution data was different from that used for the screening experiments. This was to minimize the possibility of losing crystals during transfer from one pH to another. Ideally the soaks should have been conducted in the same buffer system. This shortage has also prevented experimentation

with cryogenic techniques which are expensive in terms of crystal usage.

Poor diffraction.

Despite the fact that diffraction has been observed to 2.5Å resolution for a native crystal, the maximum resolution used during this project has been 3Å. Data of higher resolution were poor and incomplete. The problem with severe radiation damage occuring in a Synchrotron beam and preventing collection of as much data being collected per crystal compared to a rotating anode source is unusual. However, no definitive tests were carried out (*e.g.* comparison of degrees of data collected to a given resolution from a native crystal grown a particular batch of protein). It is possible that it is the intensity of the beam rather than the wavelength that increases the problems. Use of the image plate on station PX9.5 may prove useful. The beam for this station is attenuated relative to that for PX9.6. Also the sensitivity of the image plate compared to film or an area detector would allow more accurate measurement of the weaker reflections. Use of an image plate detector on a laboratory rotating anode source may also allow the collection of better quality higher resolution data. Regarding the problems we have encountered at Daresbury, it would be worthwhile setting up crystallisation trials at the laboratory in case the fragility of the crystals is exacerbated by the journey.

Heavy atom derivatives.

Despite the large number of heavy atom compounds used to determine the final set of MIR phases described in this thesis, there are only three major positions of heavy atom binding. All of these share very similar *z* co-ordinates and one of them is very close to a special position. It would almost certainly be worthwhile continuing screening for further derivatives. Other buffer systems could also be tested for suitability. Lustbader and co-workers are presently working on development of a seleno-met derivative. This would offer two immediate advantages; the use of multiple anomalous dispersion methods and the knowledge of the position of the selenium in the sequence - there are only four methionine residues in hCG.

Carbohydrate Content.

It is likely that the carbohydrate is contributing to the disorder in the crystals and

preventing better diffraction. Also it may be responsible for the difficulty in obtaining a satisfactory envelope for the phase improvement methods. It is possible that crystallization of recombinant hCG without the carbohydrate moieties would result in better ordered crystals, better diffraction and a more distinct density distribution in the cell. This would also provide a more consistent source of proteins. Jones *et al*, (1992) report the use of a lectin resistant cell line of CHO-K1 cells to produce recombinant rat CD2. Instead of the complex oligosaccharides found in rat CD2 grown from standard CHO cells, glycosylation was limited to the oligomannose precursor. The majority of this was removed using endoglycosidase H - only the N-linked N-Acetyl galactosamine residue remained - and crystals obtained.

To summarize, I feel confident that this structure will soon be fully determined. We succeeded in overcoming the problems of crystallization by removal of a large proportion of the carbohydrate. Crystallization conditions were discovered and optimized to allow the growth of sharp hexagonal bipyramidal crystals. The crystals were characterized and found to diffract to a maximum of 2.5Å resolution. After many trials, five heavy atom derivatives have been solved. Datasets for native and derivative sets were collected and allowed the calculation of a 3.5Å resolution MIR electron density map. Phase improvement methods were employed to increase the quality of the electron density. However, the limited phasing power of the higher resolution phases and both the quality and completeness of the derivative datasets have prevented the calculation of a map of sufficient quality for a full structure determination. I feel use of some of suggestions made above regarding the improvement of the quality of data will allow such a map to be calculated. It is also possible that the fitting of partial structure to the current electron density followed by phase calculation and combination will improve the map sufficiently for a full structure determination.

References

1. Adcock, E.W., Teasdale, F., August, C.S., Cox, S., Meschia, G., Battaglia, F.C. & Naughton, M.A. (1973) Human chorionic gonadotropin: its possible role in maternal lymphocyte suppression. *Science* **181**, 845-847.

2. Ascheim & Zondek (1927) Hypophysenvorderlappenhormon und ovarialhormon im Harn von Schwangeren. *Klin. Wochenschr.* **6**, 1322-1324.

3. Bidart, J.M., Troalen, F., Bousfield, G.R., Birken, S. & Bellet, D.H. (1988) Antigenic determinants on human choriogonadotropin alpha-subunit: Characterization of topographic sites recognized by monoclonal antibodies. *J. Biol. Chem.* **263**, 10364-10369.

4. Bidart, J.M., Troalen, F., Salesse, R., Bousfield, G.R., Bohuon, C.J. & Bellet, D.H. (1987) Topographic antigenic determinants recognized by Mabs on the beta subunit of human chorionic gonadotropin. *J. Biol. Chem.* **262**, 8551-8556.

5. Birken, B., Gawinowicz-Kolks, M.A., Amr, S., Nisula, B. & Puett, D. (1987) Structural and functional studies of the tryptic core of the human chorionic gonadotropin beta-subunit. *Endocrinology* **121**, 657-666.

6. Blow, D.M. & Crick, F.H.C. (1959) The Treatment of errors in the Isomorphous Replacement Method. *Acta Cryst* **12**, 794-802.

7. Blundell, T.L. & Johnson, L.M. (1976). *Protein Crystallography* Academic Press, London.

8. Boorstein, W.R., Vamvakopoulus, N.C. & Fiddes, J.C. (1982) Human chorionic-gonadotropin beta-subunit is encoded by at least eight genes arranged in tandem and inverted pairs. *Nature.* **300**, 419-422.

9. Braunstein, G.D., Rasor, J., Adler, D., Danzer, H. & Wade. M.E. (1976) Serum hCG levels throughout normal pregnancy. *Am. J. Obst.* **126**, 678.

10. Bricogne, G. (1984) Maximum entropy and the foundations of direct methods. *Acta Cryst.* **A40**, 410-445.

11. Bricogne, G & Gilmore, C.J. (1990) A multisolution method of phase determination by combined maximization of entropy and likelihood. I. Theory, Algorithms and Strategy. *Acta Cryst.* **A46**, 284-297.

12. Burleigh, B.D., Liu, W,-K. & Ward, D.N. (1976) Reaction of tetranitromethane with lutropin, oxytocin and vasopressin. *J. Biol. Chem.* **251**, 308-315.

13. Calvo, F.O. & Ryan, R.J. (1985) Inhibition of adenylyl cyclase activity in rat corpora luteal tissue by glycopeptides of human Chorionic-Gonadotropin and the alpha-subunit of human Chorionic-Gonadotropin. *Biochemistry* **24**, 1953-1959.

14. Charlesworth, M.C., McCormick D.J., Madden, B. & Ryan, R.J. (1987) Inhibition of Human Choriotropin Binding to Receptor by Human Choriotropin α Peptides. *J. Biol. Chem.* **262**, 13409-13416.

15. Chen, F. & Puett, D. (1991) Delineation via site-directed mutagenesis of the carboxyl-terminal region of human choriogonadotropin β required for subunit assembly and biological activity. *J. Biol. Chem.* **266**, 6904-6908.

16. Chen, H.C., Shimohigashi, Y., Dufau, M.L. and Catt, K.J. (1982) Characterization and Biological Properties of Chemically Deglycosylated Human Chorionic Gonadotropin. *J. Biol. Chem.* **257**, 14446-14452.

17. Cheng, K.W. (1976) Carboxymethylation of methionine residues in bovine pituitary luteinizing hormone and its subunits. Effects of binding activity with receptor sites and interactions between subunits. *Biochem. J.* **159**, 71-77.

18. Cheng, K.W. & Pierce, J.G. (1972) The reaction of tetranitromethane with pituitary, luteinizing and thyroid-stimulating hormones. *J. Biol. Chem* **247** , 7163-7172.

19. Chou, P.Y. & Fasman, G.D. (1978) Prediction of the secondary structure of proteins from their amino acid sequence *Adv. Enzymology* **47**, 45-148.

20. Chung, D., Sairam, M.R. & Li, C.H. (1973) The Primary structure of ovine interstitial cell-stimulating hormone III : Disulfide bridges of the α-subunit. *Arch. Biochem. Biophys.* **159**, 678-682.

21. Chung, D., Sairam, M.R. & Li, C.H. (1975) The Primary structure of ovine interstitial cell-stimulating hormone : Disulfide bridges of the β-subunit. *Peptide Protein Res.* **7**, 487-493.

22. Combarnous, Y. & Hennen, G. (1974) The disulphide bridges of porcine luteinizing hormone α subunit. *Biochem. Soc. Trans.* **2**, 915-917.

23. Combarnous, Y. & Maghuin-Rogister, G. (1974) Luteinizing hormone II. Relative reactivities of tyrosyl residues of the porcine hormones towards iodination. *Eur. J. Biochem.* **42**, 13-19.

24. Cornell, J.S. & Pierce, J.G. (1974) Studies on the Disulfide bonds of Glycoprotein Hormones- locations in the α-chain based on partial reductions and formation of ^{14}C-labeled *S*-Carboxymethyl derivatives. *J. Biol. Chem.* **249**, 4166-4180.

25. Crick, F.H.C., & Magdoff, B.S. (1956) The Theory of the Method of Isomorphous Replacement for Protein Crystals. I. *Acta Cryst.* **9** 901-908.

26. Derewenda, Z. & Helliwell, J.R. (1989) Calibration tests and use of a Nicolet/Xentronics imaging proportional chamber mounted on a conventional source for protein crystallography. *J. Appl. Cryst.* **22**, 123-137.

27. De Vos, A.M., Ultsch, M., Kossiakoff, A.A. (1992) Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* **255**, 306-312.

28. Fiddes, J.C. & Talmadge, K. (1984) *J. Biol. Chem.* **258**, 11492-. Structure, expression, and evolution of the genes for the human glycoprotein hormones. *Rec. Prog. Horm. Res.* **40**, 43-78.

29. Fox, G.C. & Holmes, K.C. (1966). An alternative method of solving the layer scaling equations of Hamilton, Rollet and Sparks. *Acta Cryst.* **20**, 886-891.

30. Garnier, J., Pernollet, J.C., Tertrin-Clary, C., Salesse, R., Casteing, M., Barnavon, M., De La Llosa, P. & Jutisz, M. (1975) Conformational studies of ovine leutropin and its native and chemically modified subunits by circular dichroism and ultra violet absorption spectroscopy. *Eur. J. Biochem.* **53**, 243-254.

31. Gilmore, C.J. (1984) MITHRIL - an integrated direct methods computer program. *J. Appl. Cryst.* **17**, 42-46.

32. Giudice, L.C., & Pierce, J.G. (1978) Gylcoprotein hormones: Some aspects of studies of secondary and tertiary structure. McKerns, Structure and function of the gonadotropins. Plenum Press, N.Y. 81-110.

33. Green, D.W., Ingram, V.M. & Perutz, M.F. (1954). The structure of haemoglobin. *Proc. Roy. Soc.* **A225**, 287-295.

34. Harris, D.C., Machin, K.J., Evin, G.M., Morgan, F.J. & Isaacs, N.W. (1989) Preliminary X-ray diffraction analysis of human Chorionic Gonadotropin. *J. Biol. Chem.* **264**, 6705-6706.

35. Holmes, K.C. & Blow, D.M. (1966). *The use of X-ray diffraction in the study of protein and nucleic acid structure*, Interscience, New York.

36. Howard, A.J., Gilliland, G.L., Finzel, B.C., Poulos, T.L., Ohlendorf, D.H. & Salemme, F.R. (1987). The use of an imaging proportional counter in macromolecular crystallography. *J. Appl. Cryst.*, **20**, 383-387.

37. Hum, V.G., Knifel, J.E. & Mori, K.E. (1974) Human chorionic gonadotropin reaction with tetranitromethane. *Biochem.* **13**, 2359-2364.

38. Jaffe (1979) Fetal Differentiation. In Serio, M. (Ed.) *Animal models in human reproduction* (N.Y., Pergamon Press)

39. Jones, E.Y., Davis, S.J., Williams, A.F., Harlos, K & Stuart, D.I. (1992) Crystal-structure at 2.8-Angstrom resolution of a soluble form of the cell-adhesion molecule CD2. *Nature* **360**, 232-239.

40. Jones, T.A. (1978) A graphics model building and refinement system for macro-molecules. *J. Appl. Cryst.* **11**, 268-272.

41. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991) Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Cryst.* **47**, 110-119.

42. Kabsch, W. (1988). Automatic indexing of rotation diffraction patterns. *J. Appl. Cryst.* **21**, 67-71.

43. Kalyan, N.K. & Bahl, O.P. (1983) Role of carbohydrate in human Chorionic-Gonadotropin - effect of deglycosylation on the subunit interaction and on its invitro and invivo biological properties. *J. Biol. Chem.* **258**, 67-74.

44. Kaye & Jones (1971) The effect of hCG on *in vivo* lymphocyte transmission. *Amer. J. Obstet. Gynec.* **116**, 39-42.

45. Kessler, M.J., Molakal, S., Reddy, T., Ramesh, H. Shah & Bahl, O.P. (1979) Structure of N-glycosidic carbohydrate units of human chorionic gonadotropin. *J. Biol. Chem.* **254**, 7901-7905.

46. Keutmann, H.T., Charlesworth, M.C., Kitzmann, K., Mason, K.A., Johnson, L. & Ryan, R.J. (1988) Primary and secondary structural determinants in the receptor binding sequence β-(38-57) from human luteinizing hormone. *Biochem.* **27**, 8939-8944.

47. Keutmann, H.T., Charlesworth, M.C., Mason, K.A., Ostrea, T., Johnson, L. & Ryan, R.J. (1987) A receptor-binding region in human choriogonadotropin/lutropin β-subunit. *Proc. Natl. Acad. Sci. (USA)* **84**, 2038-2041.

48. Keutmann, H.T., Hua, Q-X. & Weiss, M.A. (1992) Structure of a receptor binding fragment from human luteinizing hormone beta-subunit determined by [^{1}H] and [^{15}N] N.M.R. spectroscopy. *Molec. Endocrinology* (in press).

49. Keutmann, H.T., Mason, K.A., Kitzmann, K. & Ryan, R.J. (1989) Role of β(93-100) determinant loop sequence in receptor binding and biological activity of human leutropin and choriogonadotropin. *Molec. Endocrinology* **3**, 526-531.

50. Keutmann, H.T., McIlroy, P.J., Bergert, E.R. & Ryan, R.J. (1983a) Chemically deglycosylated human Chorionic-Gonadotropin subunits - characterization and biological properties. *Biochem.* **22**, 3067-3072.

51. Keutmann, H.T., Ratanabanangkoon, K,. Pierce, M.W., Kitzman, K. & Ryan. R.J. (1983b) Phosphorylation of human choriogonadotropin: stoichiometry and sites of phosphate incorporation. *J. Biol. Chem.* **258**, 14521-14526.

52. Krystek Jr. S.R., Dias, J.A. & Andersen, T.T. (1991) Identification of subunit contact sites on the alpha-subunit of luteinizing hormone. *Biochem* **30**, 1858-1864.

53. Lambert G., Stura E.A. & Wilson I.A. (1989) Crystallization and Preliminary X-ray Diffraction Studies of a Complex between Interleukin-2 and a Soluble Form of the p55 Component of the High Affinity Interleukin-2 Receptor. *J. Biol. Chem.* **264**, 12730-12736.

54. Lange, G. (1992) Structural studies on CD4, a glycosylated receptor protein oral presentation at *B.C.A. group winter meeting.*

55. Leslie, A.G.W. (1988) A reciprocal space algorithm for calculating molecular envelope using the algorithm of B.C. Wang. *Proceedings of the study weekend, Daresbury Laboratory* 25-31.

56. Liu, W.-K., & Ward, D.N. (1976) Effect of selective nitration of ovine lutropin on the subunit association and biological activity of the hormone. *J. Biol. Chem.* **1976**, 316-319.

57. Liu, W.-K., Yang, K.-P., Nakagawa, Y. and Ward, D.N. (1977) The role of the amino group in subunit association and receptor site interaction for ovine luteinizing hormone as studied by acylation. *J. Biol. Chem.* **249**, 5544-5549.

58. Lustbader, J.W., Birken, S., Pillegi, N.F., Gawinowicz Kolks, M.A., Pollak, S., Cuff, M.E., Yang, W., Hendrickson, W.A. & Canfield, R.E. (1989) Crystallization and characterization of human Chorionic Gonadotropin in chemically deglycosylated and enzymatically desialylated states. *Biochem.* **28**, 9239-9243.

59. Lustbader, J.W., Yarmush, D.L., Puett, D., Birken, S. & Canfield, R.E. (1992) Application of a computer modelling program to predict and to visualize the structure of the glycoprotein hormone human chorionic gonadotropin. *Proteins: Structure, Function and Genetics* (submitted).

60. McPherson, A., Koszelak, S., Axelrod, H., Day, J., Williams, R., Robinson, L., McGrath, M. & Cascio, D. (1986) An experiment regarding crystallization of soluble-proteins in the presence of beta-octyl glucoside. *J. Biol. Chem.* **261**, 1969-1975.

61. Manjunath, P. and Sairam, M.R. (1982) Biochemical, Biological, and Immunological Properties of Chemically Deglycosylated Human Choriogonadotropin. *J. Biol. Chem.* **257**, 7109-7115.

62. Matzuk, M.M., Hsueh A.J.W., Lapolt P., Tsafriri A., Keene J.L., Boime I. (1990a) The biological role of the carboxyl-terminal extension of human chorionic-gonadotropin beta-subunit. *Endocrinology* **126**, 376-383.

63. Matzuk, M.M., Hsueh, A.J.W., Lapolt, P., Tsafriri, A., Keene, J.L. & Boime, I. (1990b) Use of Mabs to subunits of human chorionic gonadotropin to examine the orientation of the hormone in its complex with receptor. *Endocrinology* **126**, 376-383.

64. Matzuk, M.M., Keene, J.L. & Boime, I. (1989) Site specificity of the chorionic gonadotropin *N*-linked oligosaccharides in signal transduction. *J. Biol. Chem.* **264**, 2409-2414.

65. Merz, W.E., & Doerner, M. (1979) Studies of the specific role of the subunits of choriogonadotropin for biological, immunological and physical properties of the hormone. Digestion of choriogonadotropin and its isolated subunits with serine carboxypeptidase. *Hoppe-Seylers Zeitshrift fuer Physiologie und Chemie* **360**, 1783-1797.

66. Milius, R.P., Keutmann, H.T. & Ryan, R.J. (1990) Molecular modeling of residues 38-57 of the β subunit of human luteinizing hormone. *Molec. Endocrinology* **4**, 859-868.

67. Mise, T. & Bahl, O.P. (1980) Assignment of disuplhide bonds in the alpha subunit of human chorionic gonadotropin. *J. Biol. Chem.* **255**, 8516-8522.

68. Mise, T. & Bahl, O.P. (1981) Assignment of the disulphide linkages in the beta subunit of human chorionic gonadotropin. *J. Biol. Chem.* **256**, 6587-92.

69. Morgan, F.J., Birken, S. & Canfield, R.E. (1975) The amino acid sequence of human chorionic gonadotropin: The alpha-subunit and beta-subunit. *J. Biol. Chem.* **250**, 5247-5258. *Mol. Cell. Biochem.*

70. Morris, J.C., Jiang, N-S., Charlesworth, M.C., McCormick, D.J. & Ryan, R.J. (1988) Effects of synthetic α-subunit peptides on thryoid stimulating hormone interactions with its receptor. *Endocrinology* **123**, 456-462.

71. Moyle, W.R., Bahl, O.P. & Marz, L. (1975) Role of the carbohydrate of Human Chorionic Gonadotropin in the Mechanism of Hormone Action. *J. Biol. Chem.* **250**, 9163-9169.

72. Moyle, W.R., Ehrlich, P.H. & Canfield R.E. (1982) Localization of residues that confer antibody binding specificity using human choriogonadotropin/leutropin β chimeras and mutants. *Proc. Natl. Acad. Sci (USA)* **79**, 2245-2249.

73. Nagai, K., Oubridge, C., Jessen, T.H., Li, J., Evans, P.R. (1990) Crystal-structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein-A *Nature* **348**, 515-520.

74. Parsons, T.F. & Pierce, J.G. (1979) Biologically active covalently cross-linked glycoprotein hormones and the effects of modification of the COOH-terminal region of their α subunits. *J. Biol. Chem.* **254**, 6010-6015.

75. Parsons, T.F. & Pierce, J.G. (1984) Free α-like material Removal of its O-linked oligosaccharide permits combination with lutropin-β. *J. Biol. Chem.* **259**, 2662-2666.

76. Patterson, A.L. (1934). Fourier series method for the determination of contacts and interatomic distances in crystals. *Phys Rev.* **46**, 372-376. *Endocrinology* **128**, 1209-1217.

77. Pierce, J.G. & Parsons, T.F. (1981) Glycoprotein Hormones; Structure and function. *Annu. Rev. Biochem.* **50**, 465-495.

78. Polisastro, P., Ovitt, C.E., Hoshina, M., Fukuoka, H., Boothby, M.R. & Boime, I. (1983) The beta-subunit of human chorionic-gonadotropin is encoded by multiple genes. *J. Biol. Chem.* **258**, 1492-1499.

79. Read,R.J. (1986) Improved Fourier coefficients for maps using phases from partial structure with errors. *Acta Cryst.* **A42**, 140-149.

80. Reed, D.K., Ryan, R.J. & McCormick, D.J. (1991) Residues in the α subunit of human choriotropin that are important for interaction with the lutropin receptor. *J. Biol. Chem.* **266**, 14251-14255.

81. Robertson, J.M. & Woodward, I. (1937) An X-ray study of the pthalocyanines. Part III, Quantitive structure determination of nickel pthalocyanine. *J. Chem. Soc.* **219**, 219-230.

82. Rosa, C., Amr, S., Birken, S., Wehmann, R. and Nisula, B. (1984) Effect of desialyation of Human Chorionic Gonadotropin on its metabolic- clearance rate in humans *J. Clin. Endocr. Metab.* **59**, 1215-1219.

83. Rossman M.G. (1961) The positions of Anomalous Scatterers in Protein Crystals. *Acta Cryst.* **14**, 383-388.

84. Ryan, R.J., Keutmann, H.T., Charlesworth, M.C., McCormick, D.J., Milius, R.P., Calvo, F.O. & Vutyavanich, T. (1987) *Recent Progress in Hormone Research* J.H. Clark, ed. Academic Press, **43**, 383-429.

85. Sairam, M.R. (1976) Role of arginine residues in ovine lutropin: reversible modification by 1,2-cyclohexanedione. *Arch. Biocem. Biophys.* **176**, 197-201.

86. Sairam, M.R., & Li, C.H. (1975) Reaction of ovine interstitial cell stimulating hormone with citraconic and maleic anhydrides. *Arch. Biochem. Biophys.* **167**, 534-537.

87. Sairam, M.R., Papkoff, H. & Li, C.H. (1972) Human Pituitary Interstitail Cell Stimulating Hormone: Primary Structure of the α subunit. *Biochem. Biophys. Res. Comm.* **48**, 530-537.

88. Sim, G.A. (1960) A note on the heavy atom method. *Acta Cryst.* **13**, 511-512.

89. Steitz T.A. (1968) A new method of locating heavy atoms bound to protein crystals. *Acta Cryst* **B24** 504-507.

90. Talmadge, K., Boorstein, W.R. & Fiddes, J.C. (1983) The human genome contains 7 genes for the beta-subunit of chorionic gonadotropin but only one gene for the beta-subunit of luteinizing hormone. *DNA* **2**, 281-289.

91. Talmadge, K., Vamvakopoulus, N.C. & Fiddes, J.C. (1984) Evolution of the genes for the beta-subunits of human chorionic gonadotropin and luteinizing hormone *Nature* **307**, 37-40.

92. Thotakura, N.R., Weintraub, B.D. & Bahl, O.P. (1990) The role of carbohydrate in human choriogonadotropin (hCG) action - effects of N-linked carbohydrate chains from hCG and other glycoproteins on hormonal activity. *Mol. and Cell. Endo.*, **70** 263-272.

93. Troalen, F., Bellet, D.H., Ghillani, P., Puisieux, A., Bohoun, C.J. & Bidart, J.M. (1988) Antigenic determinants on human choriogonadotropin alpha-subunit: Immunochemical mapping by a monoclonal antipeptide antibody. *J. Biol. Chem.* **263**, 10370-10376.

94. Tsunasawa, S., Liu, W.K., Burleigh, B.D. & Ward, D.N. (1977) Studies of disulfide bond location in ovine lutropin β subunit. *Biochim Biophys. Acta* **492**, 340-356.

95. Tsuruhara, T., Van Hall, E.V., Dufau. M.L. & Catt, K.J. (1972) Ovarian binding of intact and desialylated hcg *in vivo* and *in vitro Endocrinology* **91**, 463-469.

96. Wang, B.C. (1985) *Diffraction Methods for Biological Macromolecules* Wyckoff, H. ed., Ac. Press, New York.

97. Ward, D.N. & Moore, W.T. (1979) Comparative study of mammalian glycoprotein hormones. In: Alexander, N.J. (ed.) *Animal models for research on contraception and fertility* Harper & Row. 151-164.

98. Weare, J.A. & Reichert, L.E., Jr. (1979) Studies with carbodiimide-cross-linked derivatives of bovine lutopin. II. Location of the cross-link and implication for interaction with the receptors in testes. *J. Biol. Chem.* **254**, 6972-79.

99. Weiner, R.S., Dias, J.A. & Anderson, T.T. (1991) Epitope mapping of human follicle stimulating hormone-α using monoclonal antibody 3A identifies a potential receptor binding sequence. *Endocrinology* **128**, 1485-1494.

100. Wilson, A.J.C. (1949) The probability distribution of X-ray intensities. *Acta Cryst.* **2**, 318-321.

101. Yoshimoto, Y., Notfsen, A.A. & Odell, W.D. (1977) HCG like substance in non-endocrine tissues of normal subjects. *Science* **197**, 575.

102. Zhang, K.Y.J. & Main, P. (1988) Histogram matching as a density modification technique for phase refinement and extension of protein molecules. *Proceedings of the Daresbury Study weekend* 57-64.