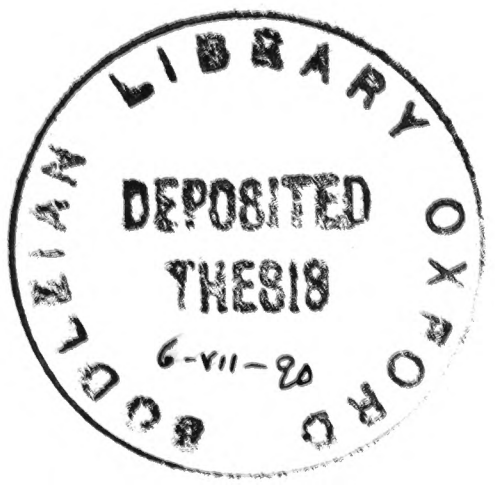
THE STRUCTURE AND SPECIFICITY OF IMMUNOGLOBULINS

P. D. Jeffrey

Thesis submitted for the degree of
Doctor of Philosophy in the University of Oxford



Laboratory of Molecular Biophysics
and Trinity College, Oxford

Trinity Term 1989

ABSTRACT

The Structure and Specificity of Immunoglobulins

Philip D. Jeffrey
Trinity College

D. Phil. Thesis
Trinity Term 1989

An investigation into the structures of the antigen-binding fragments (Fab) of mouse monoclonal immunoglobulins by X-ray crystallography, is presented. The family of immunoglobulins studied, Gloops 1–5, possess the ability to bind to both the peptide antigen and the parent protein: Hen Egg-white Lysozyme.

The Gloop1 and Gloop4 Fabs were generated by proteolytic cleavage of the antibody, and crystallisation trials yielded crystals of uncomplexed Fabs of both species. Attempts to grow Fab:antigen complex crystals proved unsuccessful. Partial purification of the heterogeneous Gloop4 Fab was found to be essential for the success of the crystallisations.

Data were collected on crystal forms of Gloop1, Gloop2 and Gloop4 on an area detector. The structures of three Gloop2 crystal forms were solved by the molecular replacement technique, using models consisting of Fab domain pairs. Two of the crystal forms were refined by molecular dynamics methods at maximum resolutions of 3.3Å and 2.8Å, and the third by rigid body methods alone at 3.5Å resolution.

Analysis of the Gloop2 structures between the crystal forms and with other Fabs indicated no atypical features in the domains. The elbow bend of the Gloop2 Fabs differed by up to 7°. The relative association of variable and constant domain pairs was also seen to vary between the Fabs. The changes in domain pairing caused significant differences in the relative disposition of the complementarity determining regions (CDRs), although there was no evidence for conformational change within individual CDRs. The Gloop2 combining site is dominated by a groove of approximate dimensions 12Åx9Åx7Å, containing many aromatic side-chains and a pair of glutamic acid residues in analogous positions to a those found in a Fab:Hen egg-white Lysozyme complex.

ACKNOWLEDGEMENTS

Firstly I would like to thank Professor Sir David Phillips for giving me the opportunity to work in the laboratory for the last three years, and for his continued encouragement. Thanks also to the SERC for monetary support, and to my parents for filling the gap between reality and the student grant.

Four people, above all others, have been crucial to the progress over the last three years: Garry Taylor as both my supervisor and computer manager; Bob Griest as an (almost) endless source of protein and crystals; Dave Stuart for his continued advice and help; Steven Sheriff for his invaluable assitance in the solution of Gloop2 and subsequent comments on refinement etc. (and for giving me a job at the end of it all). To these four, I am extremely grateful.

I am also grateful to the long-suffering inmates of T4 (and latterly S9), and to the Rees group. In particular I should like to thank Yvon, Liz, Robert, Ravi, Elspeth (Area Detector), Kate, Sally, Janet and Andrew. Thanks especially to Janet and Sally for spotting my Gloop2 sequence error. Many of the above have been kind enough to read parts (or all) of this thesis: I thank G.L.T., D.I.S., R.G., E.F.G. and D.B. for their helpful suggestions.

I would like to thank David Davies and his group at N.I.H. for making my short stay there so enjoyable and productive. I should also like to thank Paula Fitzgerald for supplying MERLOT, which made all the difference in the structure solutions. I am grateful to Dr. Roberto Poljak for supplying the D1.3 coordinates and to Eduardo Padlan and David Davies for supplying the HyHEL-10 coordinates, in both cases in advance of deposition.

Finally my thanks go to Stephen Lee for making an excellent (and fast !) job of the photographs.

ABBREVIATIONS AND SYMBOLS

ABBREVIATIONS

Amino Acids: Single and Triple Letter Codes

Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Asparagine	Asn	N	Methionine	Met	M
Aspartic Acid	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamine	Gln	Q	Serine	Ser	S
Glutamic Acid	Glu	E	Threonine	Thr	T
Glycine	Gly	G	Tryptophan	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

Other Abbreviations

A_{280}	Absorption at 280nm wavelength
ACS	Antibody combining site
CDR	Complementarity determining region
Fab	Antigen-binding fragment
Fc	Constant (or crystallisable) fragment
Fv	Variable region fragment
HEL	Hen Egg-white Lysozyme
HLA	Human Leucocyte Antigen
Ig	Immunoglobulin
MHC	Major Histocompatibility Complex

NA	Neuraminidase
NaEDTA	Sodium ethylene diamine tetra-acetate
PBS	Phosphate-buffered saline
PEG6000	Polyethylene Glycol 6000
pFc′	Pepsin constant region fragment
RCM	Reduced and carboxymethylated
RMS	Root mean square
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
w:v	Ratio of weight to volume (g/ml)
w:w	Ratio of weight to weight

SYMBOLS

$\mathbf{a}, \mathbf{b}, \mathbf{c}$	Real space unit cell vectors
$a, b, c, \alpha, \beta, \gamma$	Real space cell axis lengths and interaxial angles
$\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$	Reciprocal space unit cell vectors
$a^*, b^*, c^*, \alpha^*, \beta^*, \gamma^*$	Real space cell axis lengths and interaxial angles
B	Isotropic temperature factor
$\mathbf{F}_{obs}$, F_{obs}	Observed structure factor and it s amplitude
$\mathbf{F}_{calc}$, F_{calc}	Calculated structure factor and it s amplitude
f	Atomic scattering factor
h, k, l or $\mathbf{h}$	Miller indices
R	Crystallographic residual (defined in context)
$\mathbf{s}$	Scattering vector
V	Unit cell volume
w	Weighting parameter
$\mathbf{x}, \mathbf{y}, \mathbf{z}$	Cartesian coordinate axes

α	Phase angle
α, β, γ	Eulerian angles
ϕ, ψ, κ	Spherical polar angles
λ	Wavelength
σ	Standard deviation

UNITS

All units are S.I., except for the following:

Å	ångstrom unit, 1Å= 0.1nm
D	Dalton, 1D = 1 gram/mole

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 The Immune System

The immune system of higher animals can be divided into two components — the innate immune system and the adaptive immune system. The innate immune system (also present in lower animals) consists of agents of predetermined specificity, and is the first line of defence against infection. It consists of soluble components such as lysozyme, the complement system and the acute phase proteins, and cells such as the natural killer cells capable of recognising and killing virus-infected cells and phagocytes directed against specific bacteria.

The adaptive immune system has characteristic properties of memory, specificity and recognition of non-self. The cellular component of the adaptive immune system (lymphocytes) are subdivided into T cells and B cells. B cells are responsible for the secretion of the soluble component of the adaptive immune system — immunoglobulins. Immunoglobulins, also called antibodies, are soluble glycoproteins whose function is to bind to a target molecule (antigen). Antibodies also occur in a membrane-bound form as the B cell receptor. Antibodies have no enzymatic activity themselves but when bound to the antigen are in turn recognised by other components of the immune system (the complement system or cell-surface receptors of killer cells) which leads to the destruction of the antigen. Each antibody molecule is highly specific for a given antigen with the number of possible specificities being high — it has been estimated that more than 10^7 different antibody molecules can be produced by the immune system. T cells serve both a regulatory role (T-helper

and T-suppressor cells) in the production of antibody, and may also attack antigen directly (cytotoxic T cells).

When an infectious agent (antigen) is recognised by the immune system as non-self the production of antibodies is triggered. The immune response depends on whether the antigen has been previously encountered: the (secondary) immune response to "known" antigens is larger and more rapid than the (primary) immune response to new antigens. This property of memory is responsible for immunity against diseases, since the secondary immune response is rapid enough to eliminate the infection before symptoms appear. In principle almost any molecule can behave as an antigen. Small molecules (eg. dinitrophenol) do not produce immune responses by themselves unless linked to a "carrier" capable of triggering a response. Small molecules that are able to behave as antigens (antigenic) but do not induce antibodies (not immunogenic) are referred to as haptens. The suppression of the immune response toward self-antigens is critical to self-tolerance.

Immune responses to an antigen are in general polyclonal — many different antibodies are produced, each with different specificities toward the antigen. During the course of the immune response the relative populations of the different antibodies will change and new antibodies will appear. However each individual B cell only secretes one antibody clone. The following sections discuss antibody structure, recognition of antigen, and some applications of antibodies in therapy and research.

1.1.1 Antibody Structure

Porter (1959) proposed a four chain model for a monomeric antibody, consisting of two light chains of molecular weight $\approx$22kD and two heavy chains of molecular weight 50–77kD. The heavy chains are covalently linked to each other by disulphide bonds. The light chains are usually linked to the heavy chains by disulphide bonds,

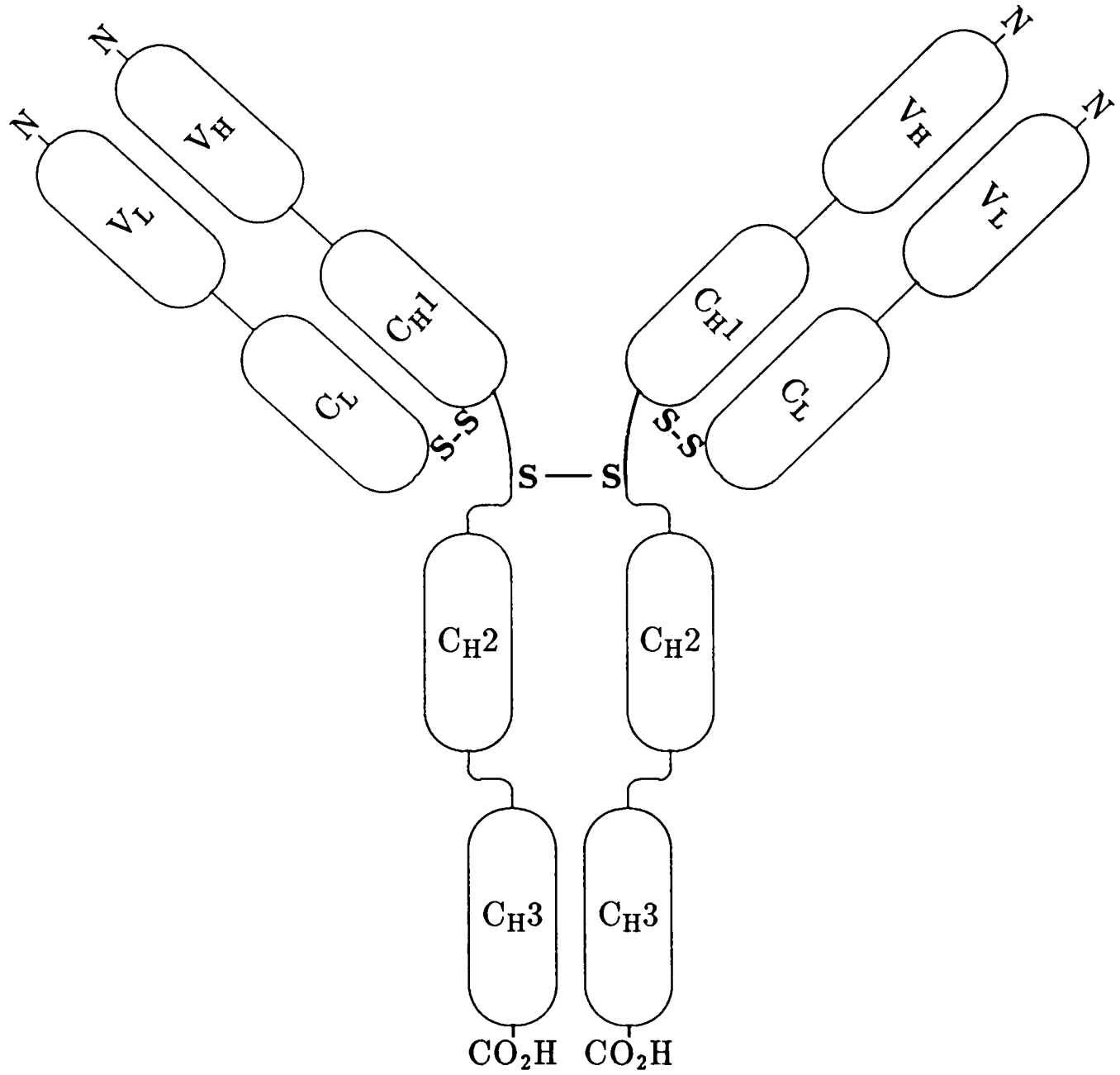


Figure 1: The Domain Structure of an IgG Monomer

but occasionally they may be covalently linked to each other.

Heavy chains may be one of five types: α, γ, δ, ϵ or μ, which define the antibody classes IgA, IgG, IgD, IgE and IgM respectively (Ig = immunoglobulin). Light chains may be one of two types: λ or κ, which are common to all antibody classes. Each antibody has identical light chains and identical heavy chains in the four-chain monomer. An analysis of the sequence of chains from IgG antibodies indicated that there were regions of significant homology within them. Edelman and Gall (1969) proposed that each of these homology regions was a domain, each fulfilling a specific function (Edelman, 1970). Domains were classified into two types according to their sequence homology: variable (V) domains occurring at the N-terminal end of the chains; constant (C) domains occurring at the other positions in the chain. Further analysis of the sequences of light and heavy chain variable domains highlighted regions of particularly high sequence variability — the hypervariable regions (Wu and Kabat, 1970; Kabat and Wu, 1971). It was hypothesised that these regions were involved in the recognition of antigen, and are alternatively known as the complementarity determining regions (CDRs). Kabat *et al.* (1987) divided antibody variable domains into framework and CDR regions according to sequence conservation. The domain structure of an IgG monomer is shown schematically in Figure 1. A brief description of the five antibody classes follows:

IgG: IgG accounts for the majority of immunoglobulin in human serum (70–75%). IgG is a monomer of molecular weight of $\approx$145kD, the γ heavy chains consisting of four domains: V_H, C_H1, C_H2 and C_H3. Electron micrographs of IgG molecules bound to a divalent hapten indicated that the antibodies had an approximate 'Y' shape, although no characteristic shape was seen when monovalent hapten was used (Valentine and Green, 1967). The angle between the arms of the 'Y' was seen to be variable — dimers, trimers, tetramers and pentamers were observed (Feinstein

and Beale, 1977). These electron micrographs confirmed that the site of antigen binding was within the variable domains, and indicated a region of flexibility (the 'hinge' region) between C_H1 and C_H2. There are four subclasses of human IgG (IgG1, IgG2, IgG3, IgG4) and also four in the mouse (IgG1, IgG2a, IgG2b, IgG3), differentiated by their heavy chain sequences. IgG is the only immunoglobulin able to cross the placental membrane, and is responsible for pre-natal immunity until the host immune system takes over.

IgA: IgA accounts for the 15–20% of the human immunoglobulin pool. IgA monomers have molecular weight of ≈160kD and have the same domain structure to the IgG monomer. In man IgA occurs mostly as the monomer, with smaller quantities of polymers which appear to be linked by intermolecular disulphide bonds. IgA dimers associate via the C-terminal ends of the heavy chains and contain the J chain polypeptide (molecular weight ≈15kD). Electron micrographs indicate that the connection between the IgA monomers in the dimer is flexible (Feinstein and Beale, 1977). In seromucous secretions IgA is the predominant immunoglobulin, occurring also as the secretory IgA (sIgA) dimer associated with the J chain and the secretory component proteins.

IgM: IgM occurs mostly as a pentamer in man, although the monomer and other polymers are also found. The μ heavy chains have five domains: V_L, C_H1, C_H2, C_H3 and C_H4 (one more than IgG or IgA), and lack the 'hinge' region of IgG and IgA. The pentamers are associated with a single J chain. Electron micrographs of IgM suggest that the C_H3 and C_H4 domains associate to form a rigid pentameric disc with 'Y' shaped projections at each vertex (Feinstein and Beale, 1977). These antigen-binding arms appear to be flexible, and are able to bind antigen both in plane and out of plane of the pentamer.

IgE: IgE occurs in trace amounts in the immunoglobulin pool. The exact role of

Immunoglobulin	IgG	IgA	IgM	IgD	IgE
Heavy Chain	γ	α	μ	δ	ϵ
Isotypes	1–4	1,2	–	–	–
Proportion of Serum Immunoglobulin (%)	70–75	15–20	10	< 1	trace
Molecular Weight of Heavy Chain (kD)	51	54	65	70	72
No. of Heavy Chain Domains	4	4	5	4	5
Hinge Region	+	+	–	+	–
Predominant Form	monomer	monomer	pentamer	monomer	monomer
Complement Fixation	+	–	++	–	–
Placental Transfer	+	–	–	–	–

Table 1: Properties of Human Immunoglobulins

IgE is uncertain, but it is associated with hypersensitivity diseases such as hayfever and asthma, triggering the degranulation of mast cells on contact with antigen. Like IgM, IgE has five heavy chain domains and no 'hinge' region, but unlike IgM it occurs only as the monomer.

IgD: IgD is present at less than 1% of the total immunoglobulin pool, but occurs in large quantities on the surface of circulating B cells. IgD has the same domain structure as IgG with an extended hinge region, which is particularly sensitive to proteolysis, and is heavily glycosylated. The biological role of IgD is uncertain.

A summary of the properties of these immunoglobulin classes is given in Table 1.

Immunoglobulins interface with other components of the immune system in order to trigger the destruction of the antigen. The first step in the lysis of an antigen by complement is the recognition of the antibody by the C1q component. C1q binds to the C_H2 domain of IgG, although the precise attachment site remains uncertain (Duncan and Winter, 1988). Another route to destruction of the antigen is the binding of the Fc portion of the antibody to cell-surface receptors of phagocytes and killer cells. The high-affinity Fc receptor (FcRI) binding site has been mapped to the C_H2 domain in IgG (Duncan *et al.*, 1988). These and the other functional sites of IgG have been reviewed by Burton (1985).

1.1.2 Generation of Antibody Diversity

The gene structure of immunoglobulins and the mechanisms of the generation of antibody diversity have been reviewed by Honjo (1983) and Tonegawa (1983). During B cell differentiation the genes encoding the antibody heavy and light chains are recombined to form the complete gene. Light chains are composed of three gene types: variable region genes (V), a short gene segment (J) and the constant region gene. Heavy chains have a similar structure, but a further gene segment (D) is

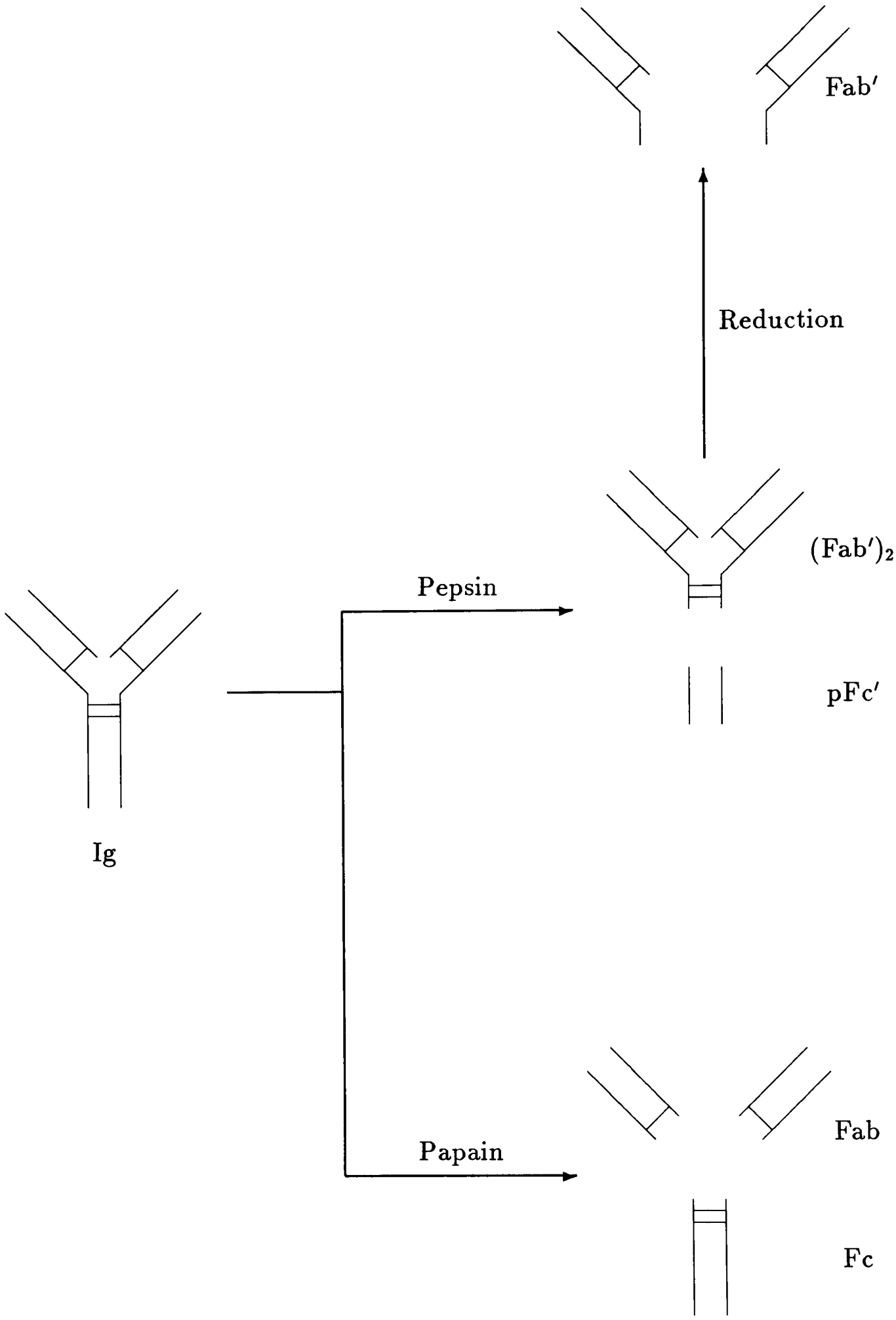


Figure 2: Schematic Representation of Antibody Proteolysis

present. During differentiation the genes recombine to produce V-J (light chain) V-D-J (heavy chain) variable region gene segments which are associated with the constant region genes. During maturation of the immune response a "class switch" may occur, where the constant region gene from one class is replaced by that from another. The VJ/VDJ genes remain intact and the antibodies retain the same antigen specificity.

Antibody diversity is generated by four mechanisms: (1) the presence of a large number of different V genes; (2) somatic recombination of V-J and V-D-J genes; (3) imprecision in the sites of V-J and V-D-J recombination and insertions at these sites; (4) somatic mutation at high rates within the CDRs. Tonegawa (1983) has estimated that $\approx 10^7$ different antibodies could be produced by the first three mechanisms alone. The diversity generated by the somatic mutation mechanism is illustrated by the alteration of the specificity of the S107 antibody from anti-phosphocholine to anti-DNA (Giusti *et al.*, 1987) by a single base change in V_H.

1.2 Three-Dimensional Structure

The flexibility of the antibody hinge region has prevented crystallisation of whole antibody in all but a few cases. Crystallographic studies have instead concentrated on proteolytic fragments of the antibody. Antibodies may be cleaved by papain (Porter, 1959), N-terminal of the inter-heavy chain disulphide bonds in the hinge region, into two antigen-binding fragments (Fabs) and the constant/crystallisable fragment (Fc). In both these fragments the chains are covalently linked. Alternatively, pepsin (Rossi *et al.*, 1969) cleaves C-terminal of the disulphide bonds to form the $F(ab')_2$ fragment (which may be converted to the Fab′ species by careful reduction of the disulphide bridges connecting the heavy chains) and the non-covalently linked pFc′ fragment. Both Fab and Fab′ contain V_L, V_H, C_L and C_H1 domains, the

Name	Fragment	Class	Species	Resolution (Å)	Reference
New	Fab′	IgG1,λ	Human	2.0	Saul *et al.*, 1978
Kol	Fab	IgG1,λ	Human	1.9	Marquart *et al.*, 1980
McPC603	Fab′	IgA,κ	Mouse	2.7	Satow *et al.*, 1986
J539	Fab	IgA,κ	Mouse	2.6	Suh *et al.*, 1986
S10/1	Fab	IgG,κ	Mouse	3.0	Colman and Webster, 1987
HED10	Fab	IgG2a,	Mouse	3.0	Cygler *et al.*, 1987
BV04-01	Fab	IgG2b,κ	Mouse	2.7	Herron *et al.*, 1987
JEL42	Fab	IgG,κ	Mouse	3.5	Prasad *et al.*, 1988
R19.9	Fab	IgG2b,κ	Mouse	2.8	Lascombe *et al.*, 1989
NQ10/12.5	Fab	IgG1,κ	Mouse	4.5	Alzari *et al.*, 1987
4-4-20	Fab	IgG2a,κ	Mouse	2.7	Herron *et al.*, 1989
D1.3	Fab	IgG1,κ	Mouse	2.5	Bentley *et al.*, 1989
D1.3	Fab:HEL	IgG1,κ	Mouse	2.8	Amit *et al.*, 1986
HyHEL-5	Fab:HEL	IgG1,κ	Mouse	2.5	Sheriff *et al.*, 1987
HyHEL-10	Fab:HEL	IgG,κ	Mouse	3.1	Davies *et al.*, 1988
NC41	Fab:NA-N9	IgG2a,κ	Mouse	2.9	Tulip *et al.*, 1989
NC10	Fab:NA-N9	IgG,κ	Mouse	3.0	Colman *et al.*, 1989
D1.3:E225	Fab:Fab	IgG1,κ, IgG2b,κ	Mouse Mouse	2.5	Bentley *et al.*, 1989
—	Fc	IgG	Human	2.9	Deisenhofer, 1981
—	Fc	IgG	Rabbit	2.7	Sutton and Phillips, 1983
—	pFc′	IgG1	Guinea Pig	3.1	Bryant *et al.*, 1985
Dob	IgG	IgG1,κ	Human	6.0	Sarma *et al.*, 1971
Kol	IgG	IgG1,λ	Human	3.0	Marquart *et al.*, 1980
Mcg	IgG	IgG1,λ	Human	6.5	Rajan *et al.*, 1983
Mcg	$(L)_2$	λ	Human	2.3	Edmundson *et al.*, 1975
Mcg/Weir	$(L)_2$	λ	Human	6.5	Edmundson *et al.*, 1984
Loc	$(L)_2$	λ	Human	3.0	Chang *et al.*, 1985
	$(L)_2$	λ	Human	2.8	Schiffer *et al.*, 1989
Rhe	$(V_L)_2$	λ	Human	1.9	Furey *et al.*, 1979
Rei	$(V_L)_2$	κ	Human	2.0	Epp *et al.*, 1975
Roy	$(V_L)_2$	κ	Human	3.0	Colman *et al.*, 1977
Au	$(V_L)_2$	κ	Human	2.2	Fehlhammer *et al.*, 1975

Fab:HEL — Fab-Hen Egg Lysozyme Complex
Fab:NA-N9 — Fab-Neuraminidase (subtype N9) Complex
Fab:Fab — Fab-Fab (idiotype/anti-idiotype) Complex

Table 2: Summary of Antibody Structures Determined By X-ray Crystallography

only difference being the length of the C-terminal end of the cleaved heavy chain. In the Fc fragment the C_H2 domain is intact, but in pFc′ this domain is further cleaved by pepsin. Figure 2 shows a schematic summary of the cleavage processes.

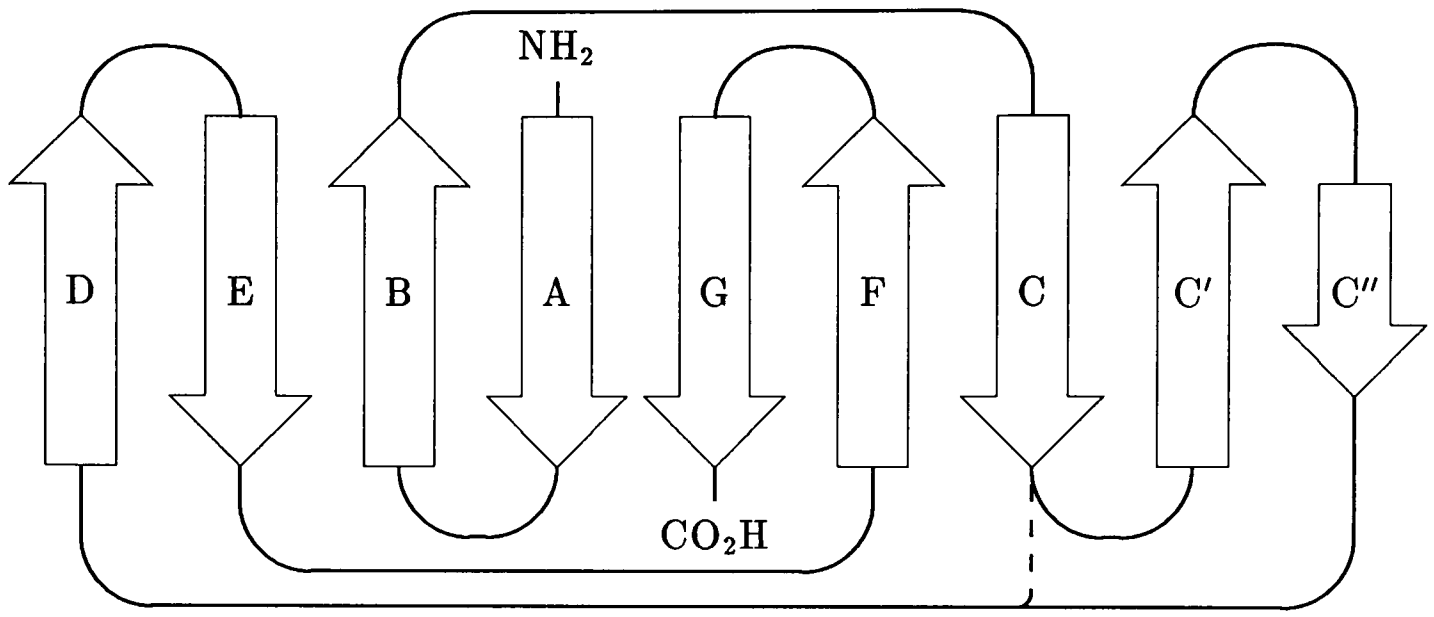
Examples of all these fragments have been crystallised. A summary of the crystallographic analyses of these fragments are given in the following sections, and are listed in Table 2.

1.2.1 Fab and Fab′ fragments

The first Fab/Fab′ structures were derived from human and mouse myeloma proteins which were available in a pure form in sufficient quantities for crystallographic work. The advent of hybridoma technology (Köhler and Milstein, 1975) has enabled the production of antibodies, with well-defined specificities to known antigens, to be produced in quantity. There are now eleven uncomplexed Fab and Fab′ fragments published at medium or high resolution. The structure of the Fab fragments has been elucidated in some detail, and has been the subject of a number of reviews (Colman, 1988; Alzari *et al.*, 1988; Davies and Metzger, 1983; Amzel and Poljak, 1979; Davies *et al.*, 1975).

The heavy and light chains of the Fab molecules are divided into the domains predicted by Edelman and Gall (1969). Both variable and constant domains consist of two twisted antiparallel β-sheets which form a β-sandwich structure. Variable and constant domains have related tertiary structures: constant regions have three and four stranded β-sheets (labelled A–G in Figure 3a) arranged in a 'greek-key' motif (Richardson, 1981); variable regions share this structure but have a further two short β-strands (C′ and C″) yielding a five-stranded β-sheet. The topology and tertiary structure of the two domains are illustrated in Figure 3.

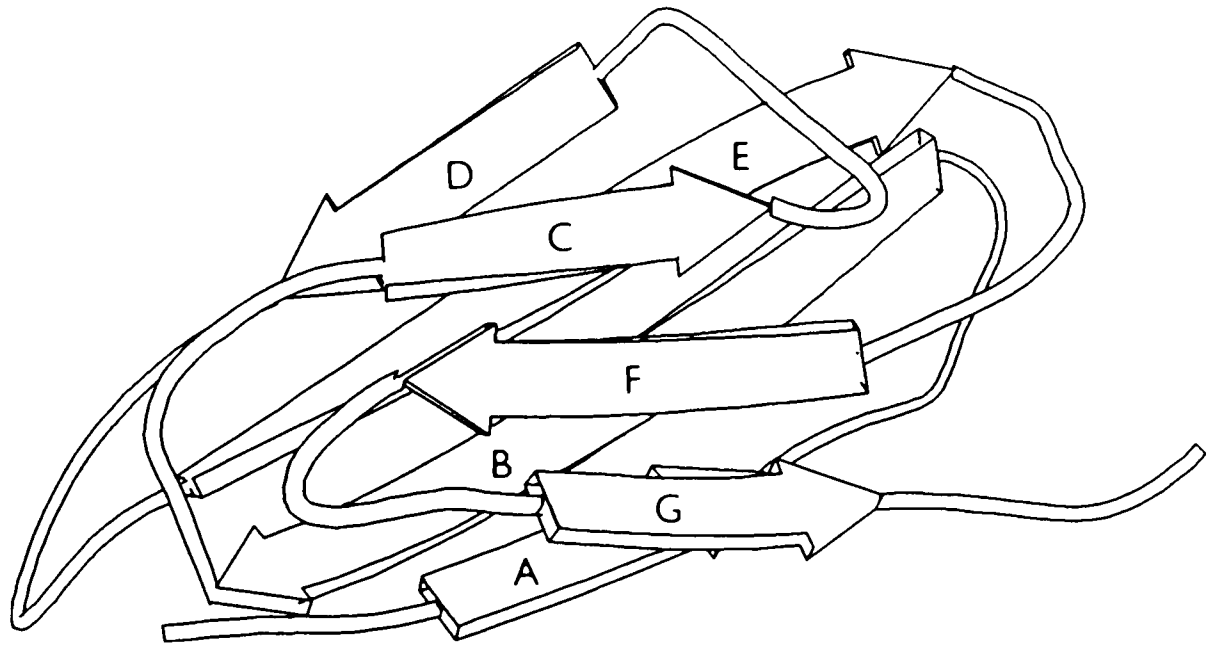
The two β-sheets within the domains are inclined at 30° to each other, as seen



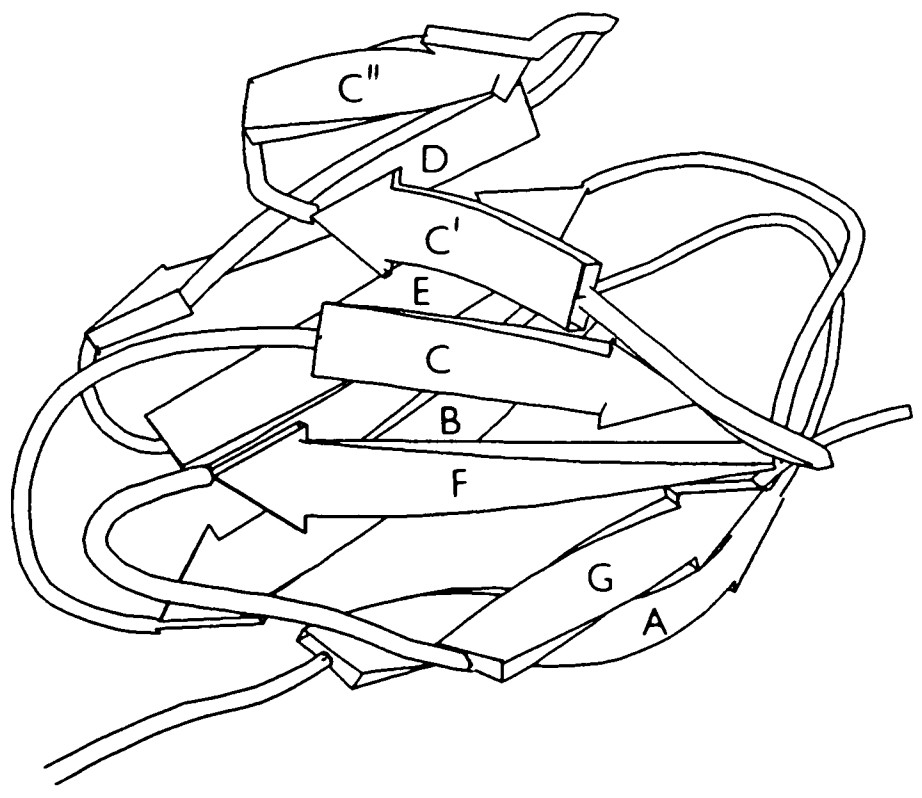
Constant domains: strands A–G excluding C′ and C″
Variable domains: strands A–G including C′ and C″

The domains form two separate β-sheets: Sheet 1: strands A, B, E, D
Sheet 2: strands G, F, C, (C′, C″)

Figure 3a: Immunoglobulin Domain Topology



Constant Domain



Variable Domain

Domains viewed from approximately equivalent positions.
β-strands labelled as in Figure 3a.

Figure 3b: Immunoglobulin Domain Structure

in other β-sheet structures (Chothia and Janin, 1981). A conserved disulphide bridge in each domain connects the two β sheets, although an antibody lacking the disulphide bridge in V_H has been reported with apparently no loss in function (Rudikoff and Pumphrey, 1986). An analysis of the domain structure of the variable and constant domains by Lesk and Chothia (1982) suggested that a central "pin" region, including the conserved disulphide bond, limited the conformational variability within each domain. Even with this limitation the relative orientation of the three- and four-stranded β-sheets was seen to vary by up to 18° between variable domains and by up to 10° between constant domains (Lesk and Chothia, 1982). The interactions between V_L and V_H, and between C_L and C_H1 are extensive (900Å^2 for each variable domain (Novotný *et al.*, 1983)). In contrast the interactions between V_L and C_L, and between V_H and C_H1 are quite sparse (e.g. Satow *et al.*, 1986).

The V_L and V_H domains associate via their five-stranded β-sheets (strands G, F, C, C′ and C″). The innermost β-sheets from V_L and V_H form a nine-stranded inter-domain β-barrell of radius ≈8.4Å, where the strands at the domain interface are inclined at ≈50° to each other (Novotný *et al.*, 1983). This association of β-sheets is unique to immunoglobulin variable domains (Chothia *et al.*, 1985). The V_L and V_H domains are related by an approximate diad axis running through the interface.

The CDRs are contained in the loops connecting the antiparallel β-strands. The V_L:V_H pairing brings the six CDRs into close proximity at the end of the domain distant from the V_L:C_L and V_H:C_H1 interfaces, and accessible to solvent. The CDRs participate in the domain-domain interface, and contribute ≈25% of the buried surface area on domain association (Colman, 1988). The disposition of the CDRs reflects the symmetry of the domain pairing, and an approximate diad relates each light chain CDR (L1, L2, L3) to its heavy chain counterpart (H1, H2, H3).

View approximately down pseudo-diad

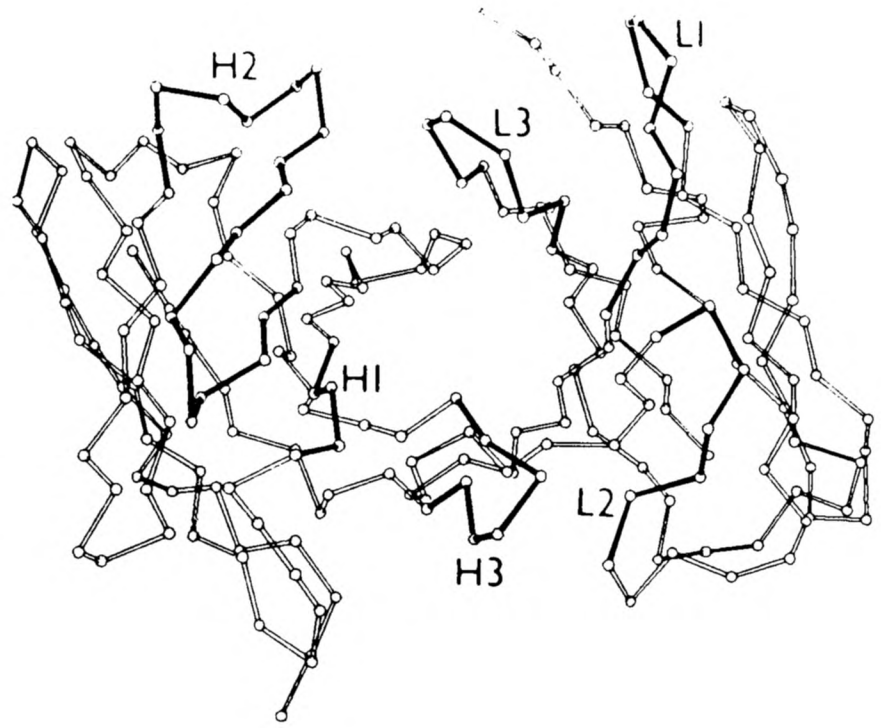
View approximately perpendicular to pseudo-diad

Variable Domains

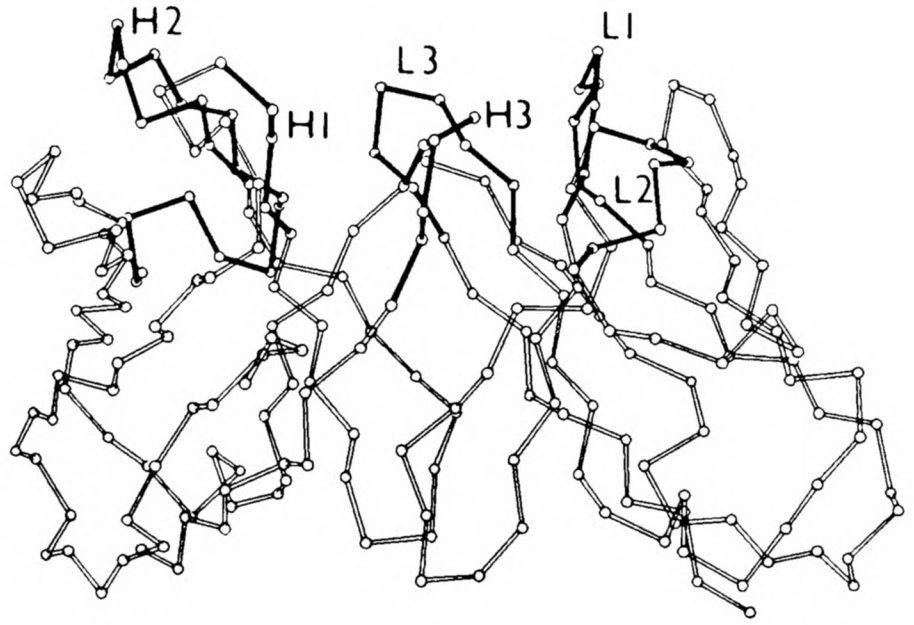
Constant Domains

Ribbon diagram derived from $C\alpha$ atoms taken from the HyHEL-5 coordinates (Sheriff *et al.*, 1987).

Figure 4a: Immunoglobulin Domain Association



View down pseudo-diad



View perpendicular to pseudo-diad

Fv $C\alpha$ atoms plotted from HyHEL-5 coordinates (Sheriff *et al.*, 1987)
Filled bonds — CDR residues
Open bonds — framework residues
CDR assignments from Kabat *et al.* (1987)

Figure 4b: CDR Positions within the Variable Domains

The C_L and C_H1 domains associate via their four-stranded β-sheets (strands A, B, E and D) such that the strands from the two domains are approximately perpendicular. This mode of association has been found in other protein structures (Chothia and Janin, 1982). The association, like that of V_L:V_H, also has an approximate two-fold symmetry to it. The association of V_L and V_H domains, and of C_L and C_H1 domains are shown in Figure 4a. The symmetry axes between C_L and C_H1, and between V_L and V_H, are generally not co-linear. The relative inclination of these axes describes the "elbow bend" of the Fab. The elbow bend is always in the same sense: V_H and C_H1 always rotate toward each other from the fully extended conformation (elbow angle = 180°). Elbow angles have been observed between 133° in McPC603 (Satow *et al.*, 1986) and the nearly fully-extended 178° of R19.9 (Lascombe *et al.*, 1989). For the JEL42 Fab (Prasad *et al.*, 1988) the two Fabs in the asymmetric unit had elbow angles differing by 7°, indicating that the elbow bend is a region of flexibility. Differences in the elbow angles have also been observed between the Fabs of HyHEL-5 (7°; Sheriff *et al.*, 1987) in different crystallographic environments, in D1.3 (34°; Bentley *et al.*, 1989) between the uncomplexed and HEL complex forms and between the Fab and intact IgG of Kol (8°; Matsushima *et al.*, 1978). Lesk and Chothia (1988) have shown that the region of contact between the V_H and C_H1 domains, conserved in sequence between antibodies, acts as a molecular "ball-and-socket" joint. This joint enables the separation of the variable and constant domains to be maintained and constrains the elbow bend to maintain the same sense.

The CDRs of the antibody structures show wide variability in sequence and length, as befits their function. Where the CDR is particularly long it may adopt no single conformation in the crystal structure (e.g. CDR L1 in McPC603 (Satow *et al.*, 1986)). Within this range of structural variation there appear to be families of

CDRs whose conformation is determined by the presence of certain "key residues" in the sequence (Chothia and Lesk, 1987; Kabat *et al.*, 1977). This observation has ramifications for the attempts to model antibody combining sites, as discussed in section 1.6.2. The disposition of the CDRs within the variable domains is shown in Figure 4b.

1.2.2 Fab–Antigen Complexes

Crystallographic structures are available for both Fab:hapten and Fab:protein complexes.

The binding of haptens to the Fab′ fragments of New (Saul *et al.*, 1978) and McPC603 (Satow *et al.*, 1986) have been investigated. Fab′ New is a human myeloma antibody of unknown specificity which was observed to bind a number of small molecules. The structure of Fab′ New with vitamin K_1OH bound to it was solved (Amzel *et al.*, 1974). The hapten bound in the depression between the six CDRs (the "hapten binding site") and interactions from both heavy and light chain CDRs were involved in binding the hapten. Similarly the structure of the anti-phosphocholine antibody McPC603 was solved with the phosphocholine hapten bound (Segal *et al.*, 1974) in the analogous binding site to that in Fab′ New. Interactions were less extensive between the antibody and this smaller antigen, but again included CDRs from both chains. Comparison with the structures of the uncomplexed Fab′s did not indicate any conformational changes within the CDRs in either case (Davies *et al.*, 1975).

There are now five Fab:antigen complexes whose three dimensional structures are known. Three of these — D1.3 (Amit *et al.*, 1986), HyHEL-5 (Sheriff *et al.*, 1987) and HyHEL-10 (Davies *et al.*, 1988) are complexed with Hen Egg-white Lysozyme (HEL). The other two complexes are with the influenza virus glycoprotein neu-

raminidase subtype N9 (NA-N9) — NC41 (Colman *et al.*, 1987; Tulip *et al.*, 1989) and NC10 (Colman *et al.*, 1989). In addition two complexes of escape mutants of neuraminidase have been obtained with the NC41 Fab, enabling investigation of the binding affinities of antigen mutants with antibody at the atomic level (Tulip *et al.*, 1989). The principal features of antibody:antigen interactions have been reviewed by Alzari *et al.* (1988), Davies *et al.* (1988), Colman (1988) and Mariuzza *et al.* (1987)), and are summarised below.

Conformational Change in Antigen: The structure of HEL has been determined at high resolution and in a number of crystal forms: tetragonal (Blake *et al.*, 1965), triclinic (Moult *et al.*, 1976), monoclinic (Rao *et al.*, 1983) and orthorhombic (Artymiuk *et al.*, 1982). Neuraminidase subtype N9 has been solved at 2.2Å resolution (Baker *et al.*, 1987). The structure of HEL in the D1.3 complex showed no larger differences to the uncomplexed forms of HEL than those observed between these uncomplexed forms (Amit *et al.*, 1986; Mariuzza *et al.*, 1987). Small differences were observed in side-chain conformation of antigen (including residues remote from the epitope) but were no larger than those observed between the different crystal forms of uncomplexed HEL. However Getzoff *et al.* (1988), in an analysis of the various crystal forms of HEL, have suggested that there is a small concerted change in the HEL structure in the D1.3 complex. HyHEL-5 and HyHEL-10 show small differences between the uncomplexed and complexed HEL structures: the $C\alpha$ of Pro70 moves by 1.7Å in HyHEL-5 (Sheriff *et al.*, 1987), whilst in HyHEL-10 the chain around Asp101 moved by up to 2.5Å (Davies *et al.*, 1988). The preliminary analysis of the NC41:NA-N9 complex suggested that there were deviations of >1.2Å for the residues 369–371 between the uncomplexed (subtype N2) and complexed (subtype N9) structure (Colman *et al.*, 1987). This suggestion has remained unconfirmed as the refinement of the uncomplexed neuraminidase N9

structure (Baker *et al.*, 1987) is incomplete (Tulip *et al.*, 1989).

Conformational Change in Antibody: The three-dimensional structure of the uncomplexed Fab is unknown for all five complexes. Conclusions about the extent of conformational change could only be made with reference to other antibody structures. The three anti-HEL antibodies showed no atypical structural features and there was no evidence for structural change upon antigen binding. The HyHEL-5 complex has been solved in two related crystal forms in which the elbow angle differs by 7° (Sheriff *et al.*, 1987). In the D1.3:HEL complex (Amit *et al.*, 1986) the D1.3 Fab has an elbow angle of 172°, in the uncomplexed form and the D1.3:E225 Fab:Fab complex (discussed below) the elbow angles were 138° and 140° respectively (Bentley *et al.*, 1989). In the NC10:NA-N9 neuraminidase complex the C_L:C_H1 domains are disordered, indicating flexibility in elbow bend in the complexed Fab (Colman *et al.*, 1989). These observations indicate that the suggestion of Huber *et al.* (1976) that the elbow bend could be a mechanism of allostery (where the liganded antibody would adopt an elbow angle of $\approx$120°) is not tenable. The pairing of the V_L and V_H domains in NC41 was originally reported as being atypical of other Fab fragments (Colman *et al.*, 1987). Further refinement of the NC41:NA-N9 complex has reduced these differences (Colman *et al.*, 1989), and Lascombe *et al.* (1989) have shown that the NC41 V_L:V_H pairing cannot be considered as an outlier. Colman (1988) has proposed the "interface adaptor hypothesis" in which the interactions between the CDRs (which contribute $\approx$ 1/4 of the V_L:V_H interface) modify the V_L:V_H packing between different antibodies.

Antigen-Antibody Interface: The NC10:NA-N9 complex (Colman *et al.*, 1989) is still at a preliminary state of refinement and has not been analysed in any detail. The other four complex structures all show a number of common features concerning the nature of the antigen:antibody interface. The interface is extensive: the buried

surface areas for antigen/Fab are 878Å^2/885Å^2 for NC41:NA-N9, 690Å^2/680Å^2 for D1.3, 750Å^2/745Å^2 for HyHEL-5 and 774Å^2/721Å^2 for HyHEL-10 (Tulip *et al.*, 1989). The Fab and antigen surfaces display a high degree of complementarity, with nearly all water molecules excluded from the interface (Davies *et al.*, 1988). The fit is not perfect however, and Mariuzza *et al.* (1987) have suggested that the cavities in the D1.3:HEL interface provides the possibility of D1.3 behaving as a heteroclitic antibody (one that binds to another antigen more strongly than the antigen to which it was reared). The D1.3:HEL interface is essentially flat, with the side-chain of Gln121 from HEL protruding into the "hapten binding pocket" formed between V_L and V_H (Amit *et al.*, 1986). The combining site of the HyHEL-5 Fab has a groove between CDRs L3 and H3 into which Arg45 and Arg68 from HEL protrude to make salt-bridges with Glu35H and Glu50H from HyHEL-5 (Sheriff *et al.*, 1987). The NC41:NA-N9 complex possesses an extensive groove between V_L and V_H which accommodates a ridge on the antigen (Tulip *et al.*, 1989).

Contacts Between Antibody and Antigen: In the three anti-HEL antibodies residues from all six CDRs make contact with the antigen. Additionally framework residues are also in contact with HEL: Thr30H and Tyr40L in D1.3 (Amit *et al.*, 1986) and Trp47H in HyHEL-5 (Sheriff *et al.*, 1987). Only five out of the six CDRs make contact with the neuraminidase in NC41:NA-N9 — CDR L1 is at least 5Å from the antigen (Tulip *et al.*, 1989). Three framework residues also make contact with antigen: Tyr49L, Thr28H and Thr30H. The preliminary refinement of the NC10:NA-N9 complex indicates that only four of the CDRs are in contact with the antigen in this complex (CDRs L1, L3, H2, H3) (Tulip *et al.*, 1989). The epitope recognised by the antibody is discontinuous in sequence in all five cases. The number of segments of polypeptide chain vary from 2 to 5, and in the case of NC10:NA-N9 also includes part of the carbohydrate from the neuraminidase (Tulip *et al.*, 1989;

Colman *et al.*, 1989). The number of antibody/antigen residues in contact with each other is 17/16 for D1.3 (Amit *et al.*, 1986), 17/14 for HyHEL-5 (Sheriff *et al.*, 1987), 19/17 for HyHEL-10 and 20/22 for NC41:NA-N9 (Tulip *et al.*, 1989). Single mutations in the antigen are sufficient to drastically reduce antigen binding: the mutation Arg68→Lys in the HyHEL-5 epitope of Turkey egg-white Lysozyme decreases the binding by a factor of 1000 (Smith-Gill *et al.*, 1982), and the mutation Gln121→His in the D1.3 epitope decreases the binding affinity by at least 450-fold (Amit *et al.*, 1986).

1.2.3 Fab–Fab Complexes

Recently the structure of an idiotope:anti-idiotope Fab:Fab complex has been reported (Bentley *et al.*, 1989). The Fab possessing the idiotope is the anti-HEL antibody D1.3, and the anti-idiotope antibody is E225. Anti-idiotypic antibodies interact with the variable regions of immunoglobulins and have been shown to modulate antigen binding affinity of certain antibodies (Sawutz *et al.*, 1987), and have been proposed to cause conformational change in the variable regions.

D1.3 and E225 interact via their CDRs, thus providing a clear explanation of E225's inhibition of antigen binding. All the E225 CDRs interact with D1.3 but only 5 CDRs from D1.3 interact with E225. The inter-domain two-fold axes of the two Fv's do not intersect and the interaction of E225 is primarily with the V_L of D1.3. The E225 paratope bears no similarity with the HEL epitope in tertiary structure and amino acid composition, but charge and surface similarity is indicated since anti-E225 sera contain antibodies with are both anti-E225 *and* anti-HEL.

1.2.4 Bence-Jones Proteins

Bence-Jones proteins are immunoglobulin light chains secreted in large quantities in the urine of patients with multiple myeloma. Typically the antigens of the parent antibody are unknown. The light chains, either as intact light chain or as the V_L fragment, crystallise as dimers ($(V_L)_2$ and L_2) The known structures of the Bence-Jones dimers are summarised in Table 2. The light chain and V_L dimers usually pair up in an analogous manner to the Fab fragments. There are, however, three proteins for which significant structural differences have been observed.

The Mcg light chain dimer crystallised in the trigonal form (Edmundson *et al.*, 1975) exhibited a structure similar to Fab structures. The orthorhombic form of Mcg differed from the trigonal form in that the elbow bend changed (Abola *et al.*, 1980), and there was also a difference in the conformation of the binding pocket (Edmundson *et al.*, 1984). The conformation of Mcg in solution adopted an elbow angle similar to these two forms (Schiffer *et al.*, 1982). Another light chain dimer, Loc, was originally crystallised (from high salt) in a form with atypical V_L:V_L pairing (Chang *et al.*, 1985). In this form the V_L:V_L two-fold axis was accompanied by a translation of one domain by 3.5Å with respect to the other, resulting in a convex "binding site". The elbow angle of this structure was only 97° (Chang *et al.*, 1985). More recently Loc has been crystallised from distilled water (Schiffer *et al.*, 1989; Chang *et al.*, 1987), in a form in which the V_L:V_L pairing closely resembled that of the Mcg light chain dimer (trigonal form). The association of the C_L domains in the two forms of Loc was identical. The third Bence-Jones dimer Rhe (Furey *et al.*, 1983), a $(V_L)_2$, was crystallised in a form which also had atypical V_L:V_L pairing, although dissimilar to that of Loc. Although the structure was solved from crystals grown at pH4.5, equivalent crystals were obtained at pH6.0 and no crystals exhibiting conventional pairing have been obtained (Wang *et al.*, 1979). Novotný

and Haber (1985) reported that Rhe had an overall repulsive electrostatic energy, unlike the Fab V_L:V_H regions of Fabs New, Kol and McPC603 and the V_L dimer Rei. It was suggested that such an abnormal V_L:V_L pairing could only be achieved under the high salt crystallisation conditions (Novotný and Haber, 1985).

The Mcg light chain dimer has an extensive cavity between the V_L domains which is reminiscent of the hapten binding pockets in Fab and Fab′ fragments, but more extensive than these (Edmundson *et al.*, 1975). Many aromatic side-chains line the cavity. Although the Mcg dimer does not have a biological role in binding antigen (the binding affinities for hapten are not shared between the light chain dimer and intact IgG1 (Firca *et al.*, 1978)), many studies have been performed to investigate the binding of haptens and small peptides to the Mcg dimer (reviewed in Edmundson *et al.*, 1984). The cavity contains three binding sites, two of which are contained in an outer pocket and the other one in a pocket deeper within the V_L:V_L domain pair. One hapten was seen to exhibit two different modes of binding: the initial binding site was in the outer pocket, but after a period of months the ligand progressed into the inner pocket. Extensive conformational changes were found to be induced by the binding of these haptens. The binding of N-formylated peptides to the binding site of Mcg has been suggested as a model for chemotactic peptides binding to the neutrophil receptor (Edmundson and Ely, 1985). Binding of an N-formylated dipeptide to Mcg indicated cis-trans isomerisation upon binding, whilst the tripeptide appeared to have a population of cis and trans isomers. In one case the binding of opioid peptides to Mcg caused major structural rearrangements which were not reversible on removal of the peptide (Edmundson *et al.*, 1987).

1.2.5 Fc Fragments

Structures of human (Deisenhofer, 1981) and rabbit (Sutton and Phillips, 1983) Fc and guinea pig pFc′ (Bryant *et al.*, 1985) fragments have been solved. In each case the C_H2 and C_H3 constant domains have the same structure as the C_H1 and C_L domains observed in Fab fragments and Bence-Jones proteins. The relationship between the two heavy chains is an approximate two-fold axis in all three cases, although the C_H2:C_H2 and C_H3:C_H3 approximate diads are not colinear in the human Fc (Deisenhofer, 1981). The pairing of the C_H3 domains is the same as the C_L:C_H1 pairing in Fab fragments but the C_H2 domains do not directly associate with each other. Instead the carbohydrate chains bound to the C_H2 domains form the interface in this region. This interface is less extensive than the C_H3 interface and the greater degree of disorder in the C_H2 domains in the Fc map was attributed to the weaker association of the C_H2 domains.

The human Fc fragment has also been solved as a complex with fragment B of the Staphylococcal Protein A (Deisenhofer, 1981). This fragment binds between the C_H2 and C_H3 domains, also making contact with the carbohydrate chains from C_H2.

1.2.6 Intact Antibody

In four cases crystals of whole IgG have been successfully grown. The crystal structure of IgG Kol (Marquart *et al.*, 1980) emphasised the flexibility of the hinge region: although the Fab arms were clearly visible in the electron density map the density became disordered at the hinge region and the density for the Fc portion was not interpretable. The diffraction data was consistent with the Fc portion occupying a small number of multiple conformations in the crystal (Marquart *et al.*, 1980). Further evidence of the flexibility of the hinge regions comes from fluorescence

studies of antibody in solution which suggested that the Fab arms are mobile with respect to the Fc (Hanson *et al.*, 1985). Mobility in the hinge region was also indicated in crystals of intact Ig ZIE, which were isomorphous with those of its $(Fab')_2$ fragment (Ely *et al.*, 1978), and of the RIV heavy chain protein (lacking V_H and C_H1 domains) (Mariuzza *et al.*, 1983) which was isomorphous with human Fc (Deisenhofer, 1981).

In two crystal structures the whole antibody was observable. In IgG Dob (Sarma *et al.*, 1971) a deletion of the hinge region reduced the mobility of the Fab arms with respect to the Fc. The structure adopted a "T" shape with the Fabs and Fc coplanar, and the Fab arms approximately anti-parallel to each other. The hinge region was also short in IgG Mcg (Rajan *et al.*, 1983), where the Fab regions were linked together by an interchain disulphide bond between the light chains (rather than between the light and heavy chains). Unlike the Dob IgG, the Fab arms in Mcg were not antiparallel. In all cases the domain structure of the Fab and Fc regions were similar to that observed in antibody fragments, within the limited resolution of the studies. Small angle X-ray scattering studies of human IgG (Gregory *et al.*, 1987) has suggested that in all four subtypes (IgG1–4) the Fabs and Fc are non-coplanar: IgG3 having an extended hinge region of ≈90Å; IgG1 and IgG3 having an extended structure with the Fab arms folding away from the Fc; and IgG2 and IgG4 having a compact structure with the Fab arms folding back toward the Fc.

Crystals of an antibody:antigen complex consisting of intact IgG and neuraminidase subtype N9 have been obtained (Tulloch *et al.*, 1985). Electron micrographs of these crystals, too small for X-ray diffraction analysis, suggested that a pair of tetrameric neuraminidase heads were cross-linked by four IgG molecules.

1.3 The Immunoglobulin Superfamily

In the same way that the domain structure of immunoglobulins was predicted from sequence homologies within the polypeptide chains (Edelman and Gall, 1969), comparisons of sequences between proteins has suggested that the immunoglobulin domains are not restricted to immunoglobulin molecules. The concept of the "immunoglobulin superfamily" has recently been reviewed (Williams and Barclay, 1988; Hunkapiller and Hood, 1988). Members of this superfamily are often associated with functions within the immune system: T cell receptors; members of the major histocompatability complex (MHC); T cell antigens; T cell adhesion molecules; immunoglobulin receptors; and immunoglobulins themselves. Other molecules, having no apparent link with the immune system, also have sequence homology: brain/lymphoid antigens Thy-1 and MRC OX-2; neural cell proteins (neural cell adhesion molecule, myelin associated glycoprotein, P_0 myelin); growth factor receptors (platelet-derived growth factor receptor; colony stimulating factor-1 receptor); and tumour antigens. Two proteins of the Epstein-Barr virus also appear to be members of the immunoglobulin superfamily (Wang *et al.*, 1989).

The members of the superfamily are subdivided into 3 subsets: the V-set, whose members show homology with the variable domains of immunoglobulins; and the C-1 and C-2 sets which are homologous to the constant domains of immunoglobulins. C-1 and C-2 set proteins are distinguished on the basis of characteristic conserved residues, with the E and F β-strands of the C2-set having a sequence pattern more similar to the V-set sequences (see Figure 3 for the strand labelling).

Evidence supporting the immunoglobulin superfamily has come from the three-dimensional structures of β_2-microglobulin (Becker and Reeke, 1985) and the human class I histocompatability antigen HLA-A2 (Bjorkman *et al.*, 1987a,b). The extra-

cellular part of HLA is divided into three domains ($\alpha_1, \alpha_2, \alpha_3$) and is non-covalently associated with β_2-microglobulin (β_2m). Recognition of cell-surface HLA by antibodies and cytotoxic T lymphocytes is responsible for transplant tissue rejection by the host immune system. Additionally HLA molecules play a direct role in the immune system as presenters of antigen peptides to T cell receptors.

The three-dimensional structures of isolated β_2m and HLA-A2 (with β_2m) showed conclusively that β_2m and the α_3 domain of HLA-A2 have the tertiary structure of immunoglobulin constant domains, as predicted (Becker and Reeke, 1985; Bjorkman *et al.*, 1987a). However the association of α_3 and β_2m in HLA-A2 was *not* the same as in Fab and Fc fragments of immunoglobulins: the screw axis relating the domains had a 146° rotation and 13Å translation and the β_2m made contact with all three α domains. Association of β_2m monomers within the crystal structure did not reproduce the C_L:C_H1 packing observed in Fabs. The α_1 and α_2 domains, which were not predicted to be part of the immunoglobulin superfamily, had identical tertiary structures and associated with each other to form a continuous antiparallel β-sheet. Two long helices ran along the top of the β-sheet forming a deep groove which appeared to be the site of antigen peptide binding (Bjorkman *et al.*, 1987b).

To date the HLA-A2/β_2m structures are the only experimental observations of the immunoglobulin superfamily. A model of the class II MHC molecules has been proposed on the basis of the HLA-A2 structure (Brown *et al.*, 1988). Another cell-surface protein, CD4, has acquired particular importance since it has been shown to be the molecule recognised by the human immunodeficiency virus HIV-I surface glycoprotein gp120. CD4 is predicted to have three immunoglobulin domains (two V-set, one C-2 set) and one domain of undetermined structure in the extracellular part of the chain.

Novotný *et al.* (1986c) have proposed a structure for the $\alpha\beta$ T cell receptor in which a Fab-like structure was suggested, with the antigen/MHC complex interacting with the equivalent region to the CDRs in Fab structures. Chothia *et al.* (1988) proposed a similar structure, and showed by sequence analysis that residues important for the maintainance of immunoglobulin domain structure and domain interface interactions were conserved in the $\alpha\beta$ T cell receptor (TCR) also. This sequence analysis also indicated that the TCR equivalent to CDRs L3 and H3 were the most variable in sequence, and a model for interaction with the antigen/MHC complex was proposed where the TCR "L3" and "H3" loops interacted with the processed antigen peptide and the TCR equivalents of the other four CDRs interacted with the two helices on the MHC molecule.

1.4 Antigenicity

The recognition of antigen by antibody is central to the action of the immune system. Correspondingly there has been much interest in the features on the antigen that lead to this recognition. An understanding of this process requires firstly that epitopes are accurately defined, and secondly that the features of these epitopes responsible for their recognition can be deconvoluted. Protein antigenicity has been reviewed in detail by Van Regenmortel (1989); Berzofsky (1985); and by Benjamin *et al.* (1985). The following sections on epitope mapping and antigenic structure are summaries of these discussions.

An epitope is defined as a region on the antigen surface which is bound by an antibody, irrespective of the antibody's origin. Getzoff *et al.* (1988) have identified two sets of residues within each epitope: contact residues, which are residues in the antigen in contact with the antibody; and critical residues are those which contribute the majority of the binding energy in the interaction. Getzoff *et al.* (1988) suggest

that these latter class of residues constitute the "antigenic determinant", which would be the epicentre of the interactions observed by footprinting methods of epitope mapping. Novotný *et al.* (1989) estimated the contribution of residues towards the binding energies of the D1.3:HEL HyHEL-5:HEL and McPC603:phosphocholine complexes using an empirical ΔG function incorporating solvent effects. Their results suggested that the lysozyme epitopes consisted of a small number of residues that contributed the bulk of the binding energy, and a larger number of 'passive' residues involved in surface complementarity.

The antigenicity of a region on the protein is a measure of its ability to be recognised by an antibody. A related property, immunogenicity, is a measure of the ability of the region to induce antibodies that bind to that region. Defined as such, immunogenicity is dependent on the genetic repertoire of the B and T cells of the host immune system and the suppression of antibodies due to self-tolerance. For example: HEL is not immunogenic in the chicken due to self-tolerance, but is immunogenic in the mouse. The properties of immunogenicity and antigenicity overlap, since epitopes are determined from antibodies obtained from a host immune response (dependent on immunogenicity), and any immunogenic site is necessarily antigenic.

1.4.1 Epitope Mapping

Benjamin *et al.* (1984) suggested that the surface of a protein is a continuum of overlapping potentially antigenic sites. All epitopes are considered to be conformational — they are dependent on the spatial arrangement of the chemical groups, not merely on the primary sequence of the epitope. Epitopes are conceptually divided into two types: "discontinuous", where the epitope is a topographically assembled site consisting of segments of polypeptide chain discontinuous in sequence;

and "continuous sites", which may also be assembled sites, but whose sequence is continuous (linear). Considerations of the area of interaction of antibodies with protein antigens (Barlow *et al.*, 1986) suggested that "discontinuous" epitopes are likely to be the rule, rather than the exception. The five Fab:protein complexes described above all have discontinuous epitopes, in which between 2 and 5 separate segments of polypeptide chain are in contact with the antibody (Tulip *et al.*, 1989).

The methods reported for epitope mapping have been reviewed by Van Regenmortel (1989) and by Getzoff *et al.* (1988). These are summarised in Table 3 and outlined in more detail below.

There are three physical methods available for mapping the epitopes of antigens: X-ray crystallography, electron microscopy and nuclear magnetic resonance (NMR). X-ray crystallography is time-consuming and applicable to comparatively few antibody:antigen interactions (if the current success rate in crystallising complexes is maintained). It does, however, provide a means of studying the epitope at near-atomic resolution. Boisset *et al.* (1988) reported the mapping of epitopes on hemocyanin by analysis of the electron micrographs of the antibody:antigen complexes. The NMR technique is applicable only to small antigens which have labile solution structures, such as peptides (J. C. Cheetham *et al.*, manuscript in preparation). This method is discussed in more detail in Chapter 7.

Two methods utilise the phenomenon of cross-reactivity between anti-protein antibodies and peptides (Atassi, 1984) representing fragments of the protein, and between anti-peptide antibodies and protein (Lerner, 1984). Both these methods map epitopes that are continuous and provide only a partial description of epitopes that are discontinuous in nature. Getzoff *et al.* (1987) have expanded the peptide/anti-protein cross-reactivity approach by replacing each residue in peptides that cross-reacted with anti-myohaemerythrin antibodies (replacement nets). The

1. Physical methods:	
X-ray Crystallography	Amit *et al.* (1986)
Electron Microscopy	Boisset *et al.* (1988)
N.M.R. of peptides	Cheetham *et al.* unpublished
2. Cross reactivity:	
Anti-protein antibodies with peptide	Atassi (1984)
Anti-peptide antibodies with protein	Lerner (1984)
3. Antigen variant binding affinities:	
Serological mapping (mutant antigens)	Benjamin *et al.* (1984)
Chemical modification of antigen	Cooper *et al.* (1987)
4. Inhibition by antibody binding:	
Antigen proteolysis	Jemmerson and Paterson (1986)
Antigen chemical modification	Burnens *et al.* (1987)

Table 3: Methods for Epitope Mapping

residues within the peptides were classified according to how their replacement affected antibody binding. Within these peptides, which corresponded to most antigenic segments of the myohaemerythrin sequence (Geysen *et al.*, 1987b) as judged by peptide/anti-protein antibody cross-reactivities, some residues were found to be completely replaceable.

Two related methods rely on identifying residues involved in the epitope by monitoring the binding of the antibody to variants of the antigen. Such variants may be related proteins with mutations in the desired area (serological mapping; Benjamin *et al.*, 1984) or generated from chemical modification of side-chains (Cooper *et al.*, 1987). The resolution of these techniques are limited by the availability of suitable variants within the epitope and are based on the presumption that only changes within the epitope will affect antibody binding. Al Modallal *et al.* (1982) found that mutations in Tobacco Mosaic Virus coat protein remote from the proposed epitope caused changes in the affinity of anti-peptide antibodies. However other studies, including structural ones, have concluded that in the majority of cases mutations cause only local changes in structure. The structure of an escape mutant of neuraminidase indicated that there were only local changes in the structure of the Lys→Glu mutant as compared to the native protein (Varghese *et al.*, 1988), and the structure of an escape mutant of haemagglutinin showed similar behaviour (Knossow *et al.*, 1984). An escape mutant of neuraminidase N9 subtype as a complex with the NC41 antibody (Tulip *et al.*, 1989) also indicated only local rearrangement.

The remaining methods compare the rates of modification of the protein antigen between the unbound form and when bound to antibody. Jemmerson and Paterson (1986) mapped an epitope on cytochrome c by analysing the results of the proteolysis of free and bound antigen. Burnens *et al.* (1987), also working on cytochrome c, determined an epitope by the relative rates of chemical modification of side-chains

between the complexed and uncomplexed protein.

These methods can be divided into two classes dependent on the type of epitope they will detect. The physical methods (NMR, X-ray crystallography) and the methods comparing the modification rates of the antigen are "footprinting" techniques — they determine (in principal) any residue that is buried on antibody binding. The other techniques are most sensitive to the critical residues in the epitope: those whose modification has a profound effect on the binding affinity. Moreover the antigenic response of a given species to the same antigen is not precisely reproducible, introducing an extra degree of variability.

1.4.2 Correlations with Antigenicity

Antigenicity has been correlated with structural properties of the antigen derived from the three-dimensional structures of proteins, and also with properties of the sequence.

One of the observed structural correlations with antigenicity is static accessibility (Novotný *et al.*, 1987a; Novotný *et al.*, 1986a,b; Novotný and Haber, 1986). In these analyses the accessibility of residues to a spherical probe, whose radius (10Å) represented the steric limitations of antibody binding to antigen, was measured. It was suggested that the antigenicity of lysozyme, myoglobin, cytochrome c and scorpion neurotoxin could be adequately accounted for using the static accessibility criterion alone. The antigenicity of antibody molecules was also predicted by this method and correlated with idiotypic, allotypic and isotypic determinants. This approach has recently been extended to predict epitopes on a computer model of the aspartic protease renin (Evin *et al.*, 1988). The absence of any major rearrangement in the antibody complexes solved by X-ray crystallography (discussed above) supports accessibility as a major factor in antigenicity. However cross-reactivity of

anti-peptide antibodies with intact protein suggest that epitopes may not be entirely limited to the accessible surface of antigens (Fieser *et al.*, 1987; Wilson *et al.*, 1984), and that some partially buried residues may be as antigenic as fully exposed ones (Getzoff *et al.*, 1988). Thornton *et al.* (1986) found that the antigenicity of myoglobin, HEL and myohaemerythrin correlated with regions that protruded most from the protein surface. Mariuzza *et al.* (1987) observed that the D1.3 epitope corresponded to local maxima in the accessibility plot of HEL, and Davies *et al.* (1988) found that accessibility also agreed with the epitopes of HyHEL-5 and HyHEL-10.

Another structural correlation with antigenicity is with mobility. Thermal factors are seen to correlate with continuous epitopes which can be mimicked with short peptides (Westhof *et al.*, 1984) and epitopes recognised by anti-peptide antibodies cross-reacting with intact protein (Tainer *et al.*, 1984). In a study of the antigenicity of myohaemerythrin as mapped by the binding of small peptides to anti-protein antibodies (Getzoff *et al.*, 1987; Geysen *et al.* 1987b) found that mobility gave the best correlation, with packing density, shape accessibility and surface exposure giving lesser correlations. There was no significant correlation between hydrophilicity and antigenicity. It was also found that the epitopes mapped had a generally convex shape and a preponderance of negatively-charged residues. Al Modallal *et al.* (1985) also found a correlation between the antigenicity of Tobacco Mosaic Virus coat protein, as mapped by short peptides, (6–8 residues) and mobility. Davies *et al.* (1988) reported that the mobility of the tetragonal HEL structure correlated less well than accessibility with the epitopes for the D1.3, HyHEL-5 and HyHEL-10 antibodies.

Other correlations with antigenicity have been proposed on the basis of: the greater frequency of hydrophilic residues in epitopes (Parker *et al.*, 1986; Hopp and

Woods, 1981); fractality of the protein surface (Lewis and Rees, 1985); sequence variation between homologous proteins (Krystek *et al.*, 1985); and the frequency of the occurrence of residues in epitopes with respect to their frequencies throughout proteins (Welling *et al.*, 1985).

The vigorous discussion between the proponents of mobility and accessibility correlations (for example: Novotný *et al.*, (1987b) and Geysen *et al.*, (1987a)) has done little to aid concensus on the nature of antigenicity in proteins. Van Regenmortel (1989) has pointed out that correlations with the antigenicity of a protein are "operationally determined" — dependent on the type of experiment used to determine the antigenicity. The presence of C-terminal groups on peptides has been reported as affecting antibody recognition (Gras-Masse *et al.*, 1986), and the cross-reactivity of antibodies has been demonstrated as dependant on the methodology of detecting the binding of antibody to antigen (Muller *et al.*, 1986).

Moreover many of the proposed factors are correlated with each other. Residues at or near the surface of an antigen (which, in the absence of major unfolding, are the only residues that an antibody is likely to interact with) will be on average more accessible, more mobile, less densely packed, more hydrophilic and subject to greater sequence variability than residues buried within the protein core.

1.5 Peptide Vaccines

Synthetic peptides have been shown to elicit antibodies that cross-react with the intact protein (Green *al.*, 1982). One of the major applications of this observation is in the development of synthetic peptide vaccines (reviewed by Steward and Howard (1987)). Synthetic peptides offer a number of advantages over conventional vaccines: they are generally more stable and easier to transport and unlike killed virus vaccines do not suffer from possible contamination by live virus (which has lead to at least

one outbreak of Foot and Mouth Disease (Bittle *et al.*, 1982)). Applications of peptide vaccines are still in the development stage, and have yet to progress into medical use, but potential vaccines have been developed against viral, toxin and parasite antigens.

Synthetic peptides have produced antibodies which include resistance in guinea pigs against Foot and Mouth Disease virus (Bittle *et al.*, 1982), and which increase the immune responses to Polio virus (Emini *et al.*, 1983) and Hepatitis B virus (Brown *et al.*, 1984). Peptides have also been exploited in order to elicit T cell immunity to the HIV surface glycoprotein gp120 in mice (Cease *et al.*, 1987). Immunisation with a peptide derived from diptheria toxin elicited antibodies which neutralised the effects of the toxin (Boquet *et al.*, 1982). Anti-toxin peptides have also been demonstrated for *E.coli* enterotoxin and cholera toxin (Steward and Howard, 1987). Two groups have demonstrated that synthetic peptides can confer immunity against malaria in monkeys and man (Patarroyo *et al.*, 1987 and 1988; Herrington *et al.*, 1987). Lastly, in a novel application of peptide vaccines, it has been shown that a peptide corresponding to the C-terminal residues of human chorionic gonadotropin reduced the fertility of baboons and may provide a vaccine-based approach to contraception (Stevens, 1986).

1.6 Modelling Of Antibody Structure

Given the long timescale of crystallographic studies, and the problems of crystallising whole immunoglobulin, there have been many attempts to model the structures of immunoglobulins based on the known structures and sequence homology. Antibody modelling can be divided into two types: modelling of antibody classes; and modelling of antibody combining sites. Both these areas are discussed below.

1.6.1 Modelling of Antibody Classes

Whole immunoglobulin structures are only available for hinge-deleted forms of IgG, and the Fc and pFc′ structures are also limited to IgG antibodies (see Table 2). Thus although the Fab fragments of IgG and IgA are relatively well characterised, and are presumed to be similar in the other antibody classes, the detailed structures of whole immunoglobulin of classes other than IgG remain unknown.

Pumphrey (1986a, 1986b) proposed speculative models for all immunoglobulin classes and human IgG subclasses, on the basis of computer model building, guided by the results of electron microscopy studies. Padlan *et al.* (1987) proposed a model for the Fc portion of IgE based on the Fc structure of IgG (Deisenhofer, 1981). In this model the $C_\epsilon 3$ domain was assumed to be similar to the $C_\gamma 2$ domain, and $C_\epsilon 2$ and $C_\epsilon 4$ domains similar to $C_\gamma 3$. Beale and Feinstein (1976) have proposed a model for IgM based on electron microscopy studies. In this model the $C_\mu 2$, $C_\mu 3$ and $C_\mu 4$ domains pair up as in the IgE model, with the $C_\mu 3$ domains separated from each other.

A series of related helical structures for the hinge region of human IgG1 was proposed by Renneboog-Squilbin (1972) on the basis of conformational energy calculations. Marquart *et al.* (1980) proposed a polyproline structure for the human IgG3 hinge based on the structure for the hinge region of Fab Kol and IgG Kol. Two models for the J chain have been proposed: Zikan *et al.* (1985) proposed an immunoglobulin-like antiparallel β-sheet bilayer based on sequence homology and circular dichroism data; Cann *et al.* (1982) proposed a two-domain structure, consisting of a domain with β-pleat structure and an α-helical domain, from structure prediction algorithms.

1.6.2 Modelling of Antibody Combining Sites

Several groups have attempted to model the conformation of the combining site of immunoglobulins. Rather than model the whole of the variable regions, these methods have used the (conserved) variable region frameworks from crystallographic structures as templates on which to build the CDRs.

The first attempts at modelling antibody combining sites sought to utilise the known CDR conformations available. The "maximum overlap" procedure used CDRs that shared the same (or nearest) length and maximum sequence homology to the modelled CDR. Sequence insertions and deletions were made so as to modify the local conformation only, and interactions within and between the CDRs were optimised. Models for the Fv regions of MoPC315 (Padlan *et al.*, 1976) and the Gloop antibodies (de la Paz *et al.*, 1986) were generated using this method. None of these models have yet been compared to experimentally determined structures, although the MoPC315 model was used as a starting point for the interpretation of data on the binding of dinitrophenol to the MoPC315 Fv fragment (Dwek *et al.*, 1977).

On a similar basis the structure of the J539 Fab was predicted using the McPC603 crystal structure (Satow *et al.*, 1986) rejoining the CDRs after appropriate deletions had been made (Feldmann *et al.*, 1981) and optimising favourable interactions within and between each CDR. Side-chains were modelled into positions that made no bad interatomic contacts. The initial manually-built model was subsequently modified by the incorporation of energy minimisation and local rebuilding (Mainhart *et al.*, 1984). A number of closely related Fv structures have been modelled from the initial J539 model (Pawlita *et al.*, 1981). None of these models have been compared with the crystallographic structures, although the structure of J539 is now available (Suh *et al*, 1986). Padlan and Kabat (1988) have proposed models

for two anti-$\alpha(1\rightarrow6)$dextran antibodies by a similar procedure, using the crystal structures of McPC603 (Satow *et al.*, 1986) and J539 (Suh *et al.*, 1986). The modelled combining sites were correlated with experimental data on the binding of the carbohydrate antigen.

Snow and Amzel (1986) reported a method which was able to model changes in the local conformation surrounding mutations in the CDRs between Fabs Kol and New. This method was an extension of the "minimum perturbation" procedure (Shih *et al.*, 1985) where conformational searches about dihedral angles were used to place side-chains due to mutations. The "coupled perturbation" procedure (Snow and Amzel, 1986) expanded the conformational search to include all residues which were within van der Waals contact distance of any conformation of the original or substituted residue.

It has been observed that there are only a limited number of conformations for each CDR from the currently available Fab structures (Chothia and Lesk, 1987; Kabat *et al.*, 1977). Such "canonical structures" have been the basis of an extension of the maximum overlap approach toward identifying the CDR conformation through the presence or absence of certain key residues within the CDR sequence. The canonical structure model for the antibody combining site has been applied to predict the conformations of the CDRs in the Fabs of D1.3 (Chothia *et al.*, 1986), HyHEL-5, HyHEL-10 and NC41 (C. Chothia *et al.*, unpublished results). This method is successful in predicting the local main-chain conformations of the CDRs with reasonable accuracy, although the position of the side-chains are subject to considerable variation in some cases. CDR H3 was difficult to model accurately using this method since no canonical structures have been observed for this CDR. The relative positioning of the CDRs were subject to greater variation, partly from differences in V_L:V_H pairing and also from packing influences from framework residues

and between the CDRs (C. Chothia *et al.*, unpublished results). The applicability of this approach is limited by the size of the canonical structure database and the ability to take into account these packing influences.

Three groups have approached the modelling of combining sites from an automated point of view, where subjective choices of conformation were replaced by evaluation of empirical conformational energies. The approaches have been either to perform a search of conformational space, or to perform energy minimisation of a large set of starting structures for each CDR.

Bruccoleri *et al.* (1988) reconstructed the combining sites of the McPC603 and HyHEL-5 antibodies using a conformational search procedure. The loops generated by the search were selected on the basis of minimum conformational energy, and if there was more than one loop with similar minimum energies, the loop with the smallest solvent-accessible area was chosen. This method was critically dependent on the order of loop construction as the conformational energy of the loop was modified by contacts with other CDRs. Unlike the canonical structure approach of Chothia and Lesk (1987), this method discards the known predispositions for CDR conformation. However this method has the advantage over the canonical structure method that side-chains should be placed with greater accuracy.

A protocol which combines both approaches has recently been developed (Martin *et al.*, 1989) and applied to the crystal structures of HyHEL-5 (Sheriff *et al.*, 1987) and Gloop2 (the structure presented in this thesis). Loops were picked from a database of protein structures on the basis of distance criteria derived for each CDR from other antibody crystal structures. A local search of the main-chain dihedrals was performed and then the side-chains were built by conformational searches. The potential used in the conformational search was modified to remove the attractive component of the van de Waals potential, and the electrostatic interactions, in or-

der to compensate for the absence of solvent molecules in the simulation. Each CDR was generated in the presence of the other CDRs from the crystal structure. The optimum loop conformations were selected on the basis of minimum conformational energy (with the modified potential) and minimum solvent accessibility of hydrophobic atoms. This method was able to reproduce the conformation of backbone and (most) side-chains in the CDRs of HyHEL-5 and Gloop2 (Martin *et al.*, 1989).

Fine *et al.* (1986) generated minimum energy conformations for the CDRs of McPC603 by energy minimising trial CDR structures. In order to eliminate the problem of energy minimisation being trapped in local energy minima a large number of random starting conformations for each CDR were generated followed by energy minimisation and/or molecular dynamics. Side-chains in each CDR were built on by a conformational search about the side-chain dihedral angles. This method predicted the conformation of the short CDR loops, especially in the presence of the remainder of the crystal structures, but was unsuccessful with the longer loops.

1.7 Meddling With Antibody Structures

The capacity of antibodies to bind specifically to antigens and to activate the immune system has led to numerous applications for monoclonal antibodies in therapeutic medicine, e.g. imaging of tumours with radiolabelled anti-tumour antibodies. Amongst the various uses of antibodies, three applications are discussed in more detail below, representing the potential use of antibodies in medicine and as versatile biochemical tools.

1.7.1 Hybrid Antibodies

Although the generation of monoclonal antibodies (Köhler and Milstein, 1975) in mice is a well-established technique, the generation of human monoclonal antibodies is limited by ethical and practical considerations. Although mouse antibodies have been used directly in therapy, they are subject to the host immune response as is any other antigen. In view of this, attempts have been made to couple the specificity generated in rodent systems with the frameworks from human antibodies by genetic engineering. The various approaches to antibody engineering have recently been reviewed by Morrison and Oi (1988).

The earliest attempts to engineer antibodies involved the attachment of mouse variable regions (either V_L and V_H or just V_H) onto the remainder of a human antibody (Morrison *et al*, 1984; Boulianne *et al.*, 1984; Neuberger *et al.*, 1985; Sun *et al.*, 1987; LoBuglio *et al.*, 1989). These "chimaeric" antibodies retained antigen-binding activity and exhibited reduced immunogenicity to the host immune system. In one therapeutic application, the variable regions from a mouse anti-tumour antibody were engineered into a human IgG1,κ antibody (LoBuglio *et al.*, 1989).

In an extension of this approach the CDRs from rodent antibodies have been transplanted into human antibody variable region frameworks. The transplantation of the heavy chain CDRs from a mouse antibody into a human antibody regenerated antigen-binding in association with the mouse light chain (Jones *et al.*, 1986). All six CDRs were transplanted from a rat into a human antibody resulting in a chimaeric antibody of therapeutic use as an anti-lymphocyte antibody (Riechmann *et al.*, 1988).

The generation of chimaeric antibodies has also been used to investigate the

effector functions of human immunoglobulins by combining mouse variable regions of known specificity with various human constant regions (Bruggermann *et al.*, 1987). The abilities of human IgG (isotypes 1–4), IgM, IgE and IgA2 to bind C1q and Protein A, and to trigger complement-mediated haemolysis and antibody-dependant cell-mediated cytotoxicity were investigated. It was suggested that the human isotype IgG1 was overall the most efficient at triggering these effects and would be of most therapeutic use.

Two types of single-chain Fv structures have been reported: Huston *et al.* (1988) reported a single chain Fv with the C-terminus of the V_H domain connected to the N-terminus of the V_L domain by a fifteen amino acid linker which retained digoxin binding (although the affinity was reduced by a factor of 6); Bird *et al.* (1988) reported three single-chain Fv proteins connected by a linker between the C-terminus of V_L and the N-terminus of V_H.

1.7.2 Immunoglobulins with Novel Functions

Immunoglobulin molecules have been constructed by chemical linkage and recombinant DNA techniques that combine the recognition function of the antibody with enzymatic or toxin activity. Neuberger *et al.* (1984) illustrated the potential of the technique by constructing a functional Fab-nuclease hybrid which retained antigen binding activity and showed nuclease activity. Vitetta and Uhr (1985) reviewed the progress in construction of immunotoxins which consisted of Fab arms conjugated to polypeptide chains from plant and bacterial toxins. More recently a single-chain Fv-toxin fusion protein has been reported (Chaudary *et al.*, 1989), generated by recombinant DNA techniques, which exhibited toxicity to cells exhibiting human interleukin-2 receptor.

The reverse case to immunotoxins are the so-called immunoadhesions. In this

case the effector functions of the antibody constant regions are used and the recognition function is derived from the non-antibody polypeptides. This approach has been used to create CD4-immunoglobulin hybrids (Capon *et al.*, 1989). CD4 is the cell-surface target molecule for the HIV-I virus glycoprotein gp120. CD4 has, as noted above, been proposed as a member of the immunoglobulin superfamily. It was reasoned that the packing of the CD4 antibody-like domains onto the remainder of the immunoglobulin would not significantly disturb either immunoglobulin or CD4 domain structure significantly. The CD4-antibody hybrids were designed to bind to HIV via the gp120 glycoprotein and for the effector functions of the antibody constant regions to trigger the destruction of the virus particle by other components of the immune system. The CD4-immunoadhesion had a longer serum lifetime than soluble CD4, and were more effective than antibodies to gp120 since they bind to a site which must be conserved for virus infectivity. The therapeutic use for these immunoadhesions remains to be established.

1.7.3 Abzymes

The production of antibodies with catalytic activity (abzymes) offers great potential for the development of reactions with predetermined specificity. Since antibodies can be made against most antigens, it remains only to introduce catalytic activity into these antibodies to generate the required abzyme. The many examples of abzymes to date indicate that this aim is realistically attainable.

The methods used to obtain abzymes have been to immunise with an analogue of the desired substrate, in order generate antibodies with specificity for substrate possessing the desired catalytic features (reviewed by Schultz, 1988; Lerner and Benkovic, 1988; Lerner and Tramontano, 1988 and 1987). The majority of catalytic antibodies have been generated using transition-state analogues. The rationale

for this approach is that enzymes lower the activation energy for a reaction by preferentially stabilising the binding of the transition state over the substrate. The extra binding energy reduces the energy of activation for the reaction. This approach has yielded a large number of catalytic antibodies. Other approaches have been to introduce chemical groups or cofactors into the combining sites of antibodies (Iverson and Lerner, 1989; Pollack *et al.*, 1988) or to immunise with groups designed to induce the required chemical features into the combining site (Shokat *et al.*, 1989). Benkovic *et al.* (1988) have reported a stereospecific amide formation on the basis of preferential binding of one isomer to the antibody and entropic effects.

Hilyard *et al.* (1989) have reported an attempt to engineer a proteolytic abzyme by a rational approach, where the crystal structure of an antibody is used to design antibody mutants possessing the desired features of a protease. This work, hampered by the lack of the antibody:antigen complex, has yet to demonstrate catalytic activity for the antibody mutants generated to date (K. L. Hilyard, unpublished results).

The first designed abzymes were reported by Tramontano *et al.* (1986) and Pollack *et al.* (1986), although it had been previously observed that anti-hapten antibodies were capable of catalysing the cleavage of the related ester (Kohen *et al.*, 1980). Subsequently many reactions have been reported (summarised in Shokat *et al.*, 1989). Most of these reactions catalysed have been of low activation energy, and occur at appreciable rates uncatalysed. The hydrolysis of a peptide bond is a much higher energy reaction. To date there have been only two reports of proteolytic abzymes: an abzyme with a metal complex cofactor (Iverson and Lerner, 1989); and the fortuitous discovery of an autoimmune antibody that had proteolytic activity (Paul *et al.*, 1989). Neither of these reactions were very fast: the cofactor abzyme had a turnover rate (k_{cat}) of $3.6\text{x}10^{-2}\text{min}^{-1}$ and the autoimmune

abzyme k_{cat} was 15.6min^{-1}. Rate enhancements of 10^7 have been found for esterase abzymes, approaching the rate enhancements exhibited by enzymes (Tramontano *et al.*, 1988).

1.8 The Gloop Antibodies

The antigenicity of a 'loop' region of HEL was reported by Arnon and Sela (1969), using a peptide corresponding to residues 64–83 of the HEL sequence. They obtained both anti-peptide antibodies which reacted with HEL and anti-HEL antibodies which bound to the peptide, although the anti-peptide antibodies were unable to inhibit the lysis of *M. lysodeikticus* by HEL. The anti-peptide antibodies bound to the the reduced and carboxymethylated (RCM) peptide with reduced affinity, and the anti-lysozyme antibodies showed no binding affinity for the reduced peptide. Arnon and Maron (1971) reported that antibodies against RCM forms of HEL and α-lactalbumin cross reacted with the 60–83 loop. Maron *et al.* (1971) reported anti-peptide (60–83) and anti-HEL antibodies that showed a drastic reduction in binding affinity for the RCM form of the peptide. Atassi (1978) proposed the "entire antigenic structure of lysozyme" was accounted for by three sites. Only one of residue from these sites is present in the loop peptides mentioned above. This view of HEL antigenicity has been critically reviewed by Benjamin *et al.* (1984). Moreover the epitope of the crystallographically determined HyHEL-5:HEL complex (Sheriff *et al.* does not lie within the sites proposed by Atassi (1978).

The Gloop antibodies are five mouse anti-peptide monoclonal antibodies raised against the 'loop' peptide of HEL (residues 57–84) selected for their cross-reactivity with HEL (Darsley and Rees, 1985a). The affinity of the Gloop antibodies for the loop peptide, the pep1 peptide (the loop peptide disulphide bonded to residues 91–108 from HEL), and a panel of variant lysozymes was investigated (Darsley and

Rees, 1985a). Binding of the RCM loop peptide, the refolded loop peptide and pep1 are all essentially identical, but the binding of intact HEL was ≈100 times weaker to these antibodies. None of these antibodies inhibit the lysis of *M.lysodeikticus* by HEL. The variable region sequences are available for all five Gloop antibodies (Darsley and Rees, 1985b), and the constant region sequences have been obtained for the Gloop2 sequence (Roberts, 1987).

The five Gloop antibodies fall three groups according to their serologically-mapped epitopes on the HEL antigen: Gloop1 (IgG2a,κ) and Gloop2 (IgG2b,κ) have essentially identical specificities, Gloop3 (IgG2a,κ) and Gloop4 (IgG2b,κ) also have similar epitopes, but these epitopes are distinct from those recognised by Gloop1 and Gloop2 (Darsley and Rees, 1985a). Gloop5 (IgG1,κ) has an epitope which was contained within the Gloop1/Gloop2 epitope. These five epitopes are contained within the region (65–70,72,74–79) in the HEL sequence. These mapped epitopes partially overlap the HyHEL-5 epitope (Sheriff *et al.*, 1987), including Arg68 (a critical residue in the binding of HyHEL-5 to HEL). The region 64–80 corresponds to a region accessible to a sphere of 10Å radius (Novotný *et al*, 1986), and is also a region of high "protrusion index" according to the calculations of Thornton *et al.*, 1986). The region around Pro70 contains one of the highest B factors in HEL and Human Lysozyme (Artymiuk *et al.*, 1979). Thus the epitopes of the Gloop antibodies correspond to a protruding, antibody-accessible and mobile part of the HEL structure.

Gloop2 has been the subject of a sustained effort to investigate the role of CDR residues in the binding of pep1 and HEL, by site-directed mutagenesis of the CDRs in the Gloop2 Fab fragment (Roberts and Rees, 1986; Roberts, 1987; Roberts *et al.*, 1987). The binding of the loop peptide to the Gloop1 Fab has been investigated by two-dimensional nuclear magnetic resonance techniques (J. C. Cheetham *et al.*,

manuscript in preparation). The results of both sets of experiments are discussed in detail with reference to the Gloop2 Fab structure in Chapter 7. In addition the Gloop2 model (and latterly, the crystal structure) has been the basis for the design of a proteolytic Gloop2 antibody by site-directed mutagenesis of the CDR residues (Hilyard *et al.*, 1989).

The structures of the Gloop2, Gloop4 and Gloop5 Fabs have been modelled, using the "maximum overlap" procedure described above (de la Paz *et al.*, 1986; Rees and de la Paz, 1986), and a model of a Gloop2:HEL complex was constructed. This model was the starting point for the design of mutations in the Gloop2 CDRs, and the results from these mutations have led to the reappraisal of the initial model (Roberts *et al.*, 1987). The Gloop2 structure has been the subject of a more advanced modelling exercise (Martin *et al.*, 1989) as discussed in detail above.

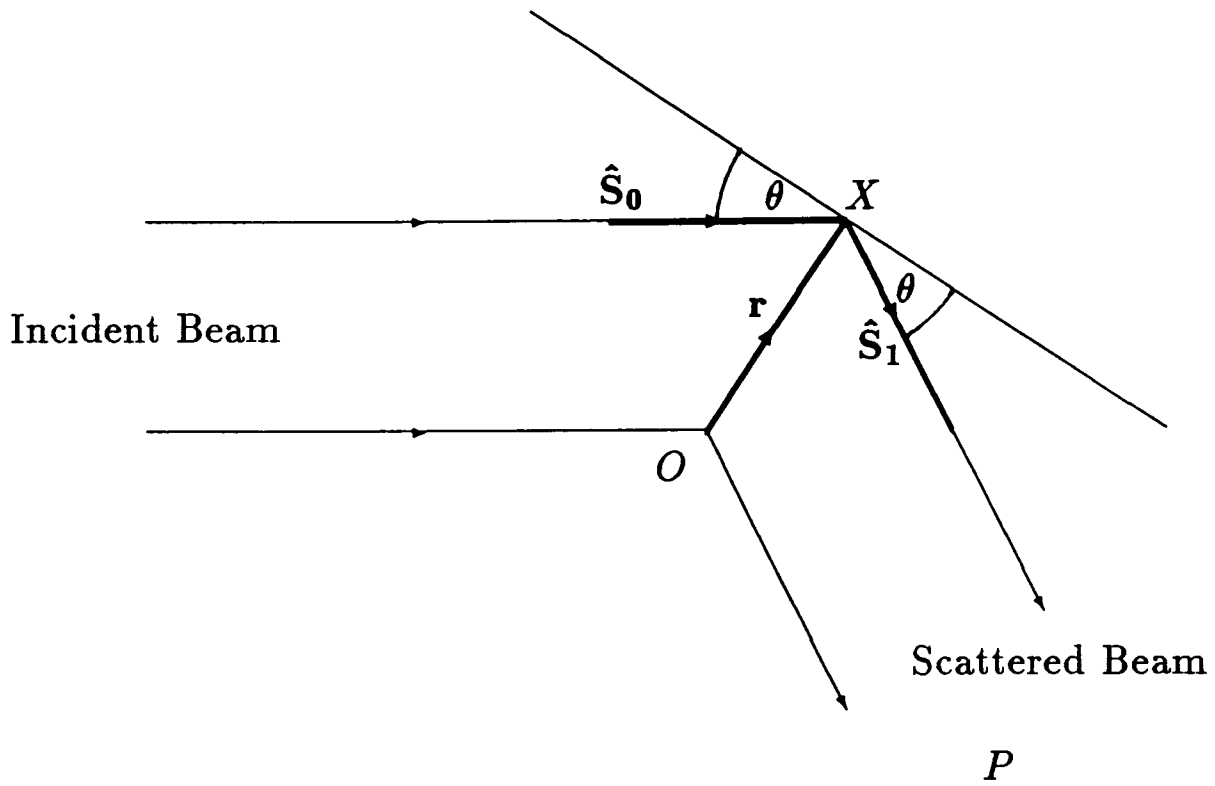
1.9 Aims of the project

The Gloop antibodies define a system which addresses a number of important issues concerning antibody structure and antigen recognition: the role of sequence differences in the CDRs in altering the fine specificities of the antibodies; the cross-reactivity with the loop peptide and HEL; the induction of conformational change in the (relatively mobile) loop peptide on binding to antibody. Structures of the Gloop antibodies and antibody complexes would be an important step toward understanding the structural basis to these questions. The structures of the Gloop Fabs are required as tests of the modelling processes, and to provide structures with which to interpret mutagenesis results. To this end the three-dimensional structures of the Gloop antibodies were required as their complexes with the loop peptide, with HEL, and also in their uncomplexed forms (to determine the role of conformational change in the antibody). Gloop3 was omitted from the crystallo-

graphic studies since it was the most difficult of the five antibodies to produce in sufficient quantities for crystallographic analysis.

The impossibility of even crystallising (let alone solving) the twelve desired structures (four uncomplexed Fabs and eight complexes) was acknowledged at the outset. However it appeared prudent to *try* crystallising all such species, subject to the availability of material, since it was not certain which of the Fabs and/or complexes would prove useful for structure determination. At the start of this project small crystals had been obtained of the uncomplexed Gloop1 and Gloop2 Fabs (R. E. Griest, unpublished results), which had been shown to diffract to at least moderate resolution.

This project continued from these intial promising results. Chapter 3 presents the results of preparation and crystallisation of the Gloop1 and Gloop4 Fabs. Data collected on these Fabs and the Gloop2 Fab are summarised in Chapter 4. The molecular replacement solution of Gloop2 (Chapter 5) and the refinement (Chapter 6) and analysis of this uncomplexed Fab (Chapter 7) are presented. Chapter 2 deals with elements of crystallographic theory relevant to the methods used within this thesis.



O, X	— origin and arbitary scattering point.
P	— observer.
$\hat{\mathbf{S}}_0, \hat{\mathbf{S}}_1$	— incident and scattered beam (unit) vectors.
$\mathbf{r}$	— vector giving position of X relative to O.

Figure 1: Scattering from a pair of points

CHAPTER 2

CRYSTALLOGRAPHIC THEORY

2.1 X-ray Diffraction

X-ray crystallographic techniques and theory appropriate to protein crystallography have been previously described by Blundell and Johnson (1976). Only those aspects of the theory and practice that pertain to the work in this thesis have been included in this chapter, omitting in particular the method of multiple isomorphous replacement.

2.1.1 Diffraction by a pair of points

Figure 1 illustrates the geometry of a diffraction event by a pair of points. The wave scattered at the point X has a phase shift of α relative to the wave scattered at the origin (O), as seen by an observer P:

$$\alpha = -\frac{2\pi}{\lambda}(\mathbf{r}.\hat{\mathbf{S}}_1 - \mathbf{r}.\hat{\mathbf{S}}_0)$$

where $\hat{\mathbf{S}}_0$ and $\hat{\mathbf{S}}_1$ are unit vectors along the direction of the incident and scattered beams respectively and $\mathbf{r}$ is the position vector of X. Defining the scattering vector $\mathbf{s}$ as: $\mathbf{s} = \mathbf{r}.\hat{\mathbf{S}}_0 - \mathbf{r}.\hat{\mathbf{S}}_1$, it follows that:

$$\alpha = 2\pi\mathbf{r}.\mathbf{s}$$

where

$$|\mathbf{s}| = \frac{2\sin\theta}{\lambda} \qquad (1)$$

The scattering from a volume dv at a point $\mathbf{r}$ for which the electron density is $\rho(\mathbf{r})$ is given by:

$$\rho(\mathbf{r}) \exp(2\pi i \mathbf{r}.\mathbf{s}) dv$$

and the scattering from the whole atom by:

$$f = \int_{\substack{\text{atomic} \\ \text{volume}}} \rho(\mathbf{r}) \exp(2\pi i \mathbf{r}.\mathbf{s}) dv \tag{2}$$

where f is the atomic scattering factor. For spherically symmetric electron density distributions f is a real value.

2.1.2 Diffraction from a Crystal

A crystal is made up of a regular assembly of the unit cell, described by three vectors **a**, **b** and **c** or alternatively by the lengths of these vectors and the inter-axial angles $(a, b, c, \alpha, \beta, \gamma)$. Unit cells are divided into seven crystal systems according to the highest symmetry present, and further into 32 crystal classes by the arrangement of symmetry elements with respect to each other.

The diffraction condition for a three-dimensional crystal is given by the Laue equations:

$$\begin{aligned} \mathbf{a}.\mathbf{s} &= h \\ \mathbf{b}.\mathbf{s} &= k \\ \mathbf{c}.\mathbf{s} &= l \end{aligned} \tag{3}$$

where h, k and l are integers. The triplet (hkl) defines a plane which intercepts the unit cell axes at $(a/h, b/k, c/l)$, and the scattering vector **s** is perpendicular to this plane.

The positions at which the diffraction conditions are satisfied form a lattice (the reciprocal lattice) whose unit cell dimensions are described by vectors $\mathbf{a}^*$, $\mathbf{b}^*$ and $\mathbf{c}^*$ with any point on the reciprocal lattice defined by the vector $h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$. The reciprocal lattice is related to the real space lattice by the equations:

$$\begin{aligned} \mathbf{a}.\mathbf{a}^* = 1, \mathbf{a}.\mathbf{b}^* = 0, \mathbf{a}.\mathbf{c}^* = 0 \\ \mathbf{b}.\mathbf{a}^* = 0, \mathbf{b}.\mathbf{b}^* = 1, \mathbf{b}.\mathbf{c}^* = 0 \\ \mathbf{c}.\mathbf{a}^* = 0, \mathbf{c}.\mathbf{b}^* = 0, \mathbf{c}.\mathbf{c}^* = 1 \end{aligned} \tag{4}$$

and it follows that $\mathbf{a}^*$ is perpendicular to the $\mathbf{b}.\mathbf{c}$ plane, $\mathbf{b}^*$ is perpendicular to $\mathbf{a}.\mathbf{c}$ and $\mathbf{c}^*$ perpendicular to $\mathbf{a}.\mathbf{b}$. By considering the form of the Laue diffraction conditions (3) and the above definition of the reciprocal lattice vectors (4), it is apparent that the scattering vector $\mathbf{s}$ fulfills the relation:

$$\mathbf{s} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \tag{5}$$

which corresponds to the vector from the origin of the reciprocal lattice to the lattice point described by (hkl). From equation (2) the scattering from N atoms in the unit cell is:

$$\mathbf{F}(\mathbf{s}) = \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{r}_j.\mathbf{s}) \tag{6}$$

and substituting in from equation (5):

$$\mathbf{F}(hkl) = \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{r}_j \cdot (h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*)) \tag{7}$$

The position vector $\mathbf{r}_j$ can be further expressed in terms of the unit cell vectors:

$$\mathbf{r}_j = x_j.\mathbf{a} + y_j.\mathbf{b} + z_j.\mathbf{c}$$

and inserting this into equation (7) and applying the relations from (4):

$$\mathbf{F}(hkl) = \sum_{j=1}^{N} f_j \exp(2\pi i(hx_j + ky_j + lz_j)) \tag{8}$$

Equation (8) is the structure factor equation. The structure factor $\mathbf{F}$ is a complex quantity which may be expressed in terms of it s magnitude (F) and the phase angle α or as the sum of sine and cosine terms:

$$\begin{aligned} \mathbf{F}(hkl) &= F(hkl)\exp(i\alpha(hkl)) \\ &= F(hkl)\cos(\alpha(hkl)) + iF(hkl)\sin(\alpha(hkl)) \end{aligned} \tag{9}$$

The structure factor $\mathbf{F}(hkl)$ is related to the intensity of the diffracted beam ($I(hkl)$) by:

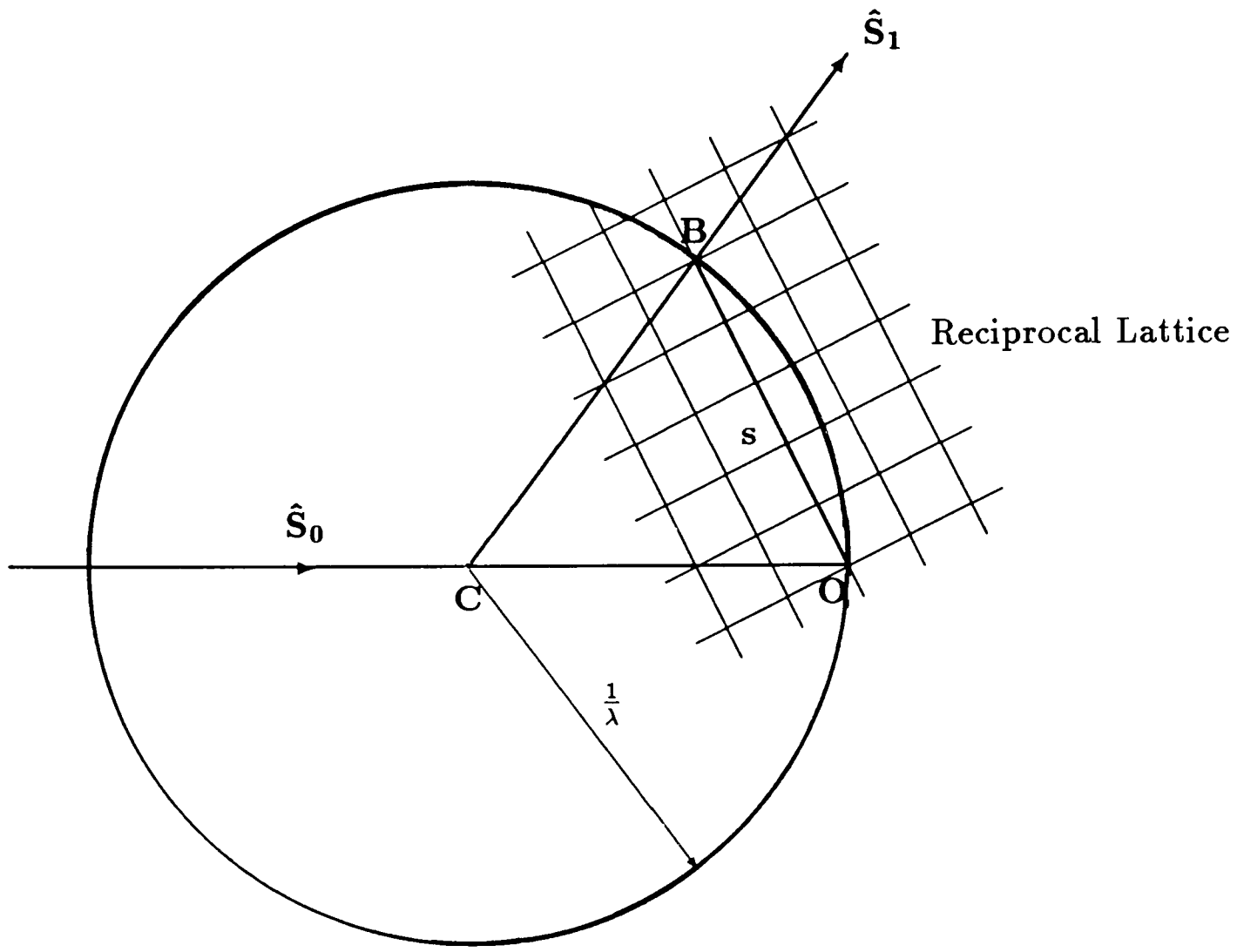
$$I(hkl) = \mathbf{F}(hkl)\cdot\mathbf{F}^*(hkl) = |\mathbf{F}(hkl)|^2 = F(hkl)^2$$

By considering the diffraction process as reflection from a series of parallel planes of atoms, Bragg expressed the requirement for diffraction as:

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

where 2θ is the angle between $\hat{\mathbf{S}}_\mathbf{0}$ and $\hat{\mathbf{S}}_\mathbf{1}$ as in Figure 1, λ is the wavelength of the incident wave and d is the spacing between adjacent planes. The integer n refers to the 'order' of the diffraction. The vector $\mathbf{s}$ is perpendicular to the reflecting planes and is related to d by:

$$|\mathbf{s}| = \frac{1}{d}$$



$\hat{\mathbf{S}}_0, \hat{\mathbf{S}}_1$ — incident and scattered beams, see Figure 1.
s — scattering vector (defined in text).
C — crystal (point of scattering).
O — origin of reciprocal lattice.
B — reciprocal lattice point in reflecting position.

Figure 2: The Ewald Construction

2.1.3 The Ewald Construction

Ewald (1921) proposed a geometric construction which summarises the geometry of the diffraction process from the point of view of both the Laue equations (3) and Bragg's Law. Figure 2 illustrate this construction as a two-dimensional section through this sphere. The crystal (**C**) lies at the centre of a sphere of radius $1/\lambda$. The origin of the reciprocal lattice (**O**) lies at a point where the incident beam (vector $\hat{\mathbf{S}}_\mathbf{0}$) intercepts the sphere after passing through the crystal. In the diffracting condition for a reflection denoted by the Miller indices (hkl) the reciprocal lattice point corresponding to these indices also lies on the surface of the Ewald sphere (**B**). The vector connecting the origin to this point is the scattering vector **s**. The angle $\mathbf{O\hat{C}B}$ is 2θ.

2.1.4 The Electron Density Equation

The structure factor equation (8) can be rewritten in terms of the electron density at any point in the unit cell. From equations (2) and (8):

$$\mathbf{F(s)} = \int_{\substack{\text{unit}\\ \text{cell}}} \rho(\mathbf{r}) \exp(2\pi i \mathbf{r}.\mathbf{s}) dv \tag{10}$$

The expression for $\rho(\mathbf{r})$ is obtained from the Fourier transform of the complete set of $\mathbf{F(s)}$:

$$\rho(\mathbf{r}) = \int_{\substack{\text{reciprocal}\\ \text{lattice}}} \mathbf{F(s)} \exp(-2\pi i \mathbf{r}.\mathbf{s}) dv^*$$

Since $\mathbf{F(s)}$ is non-zero only at integral values of h, k and l, the integral can be replaced by a summation, and by using the same manipulations as for equation (8):

$$\rho(xyz) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \mathbf{F}(hkl) \exp(-2\pi i(hx + ky + lz)) \tag{11}$$

or alternatively using equation (9):

$$\rho(xyz) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F(hkl) \exp(i\alpha(hkl)) \exp(-2\pi i(hx + ky + lz)) \quad (12)$$

The quantity recorded during data collection, $I(hkl)$, gives information on $F(hkl)$ only, and provides no information on the value of the phase $\alpha(hkl)$ which is essential to the calculation of the electron density map from the observed data (equation (12)). This 'phase problem' is the principal problem that has to be overcome in the determination of protein structure from diffraction data.

2.1.5 Thermal Motion

The above equations were obtained assuming static atoms undergoing no motion. In practice the atoms in a crystal will undergo thermal motion, and for an atom undergoing isotropic motion the scattering from an atom is decreased by a factor of $\exp(-B(\sin^2\theta/\lambda^2))$, where B is often referred to as the B factor. The 'B factor' is given by $B = 8\pi^2\overline{u}^2/3$ where $\overline{u}$ is the mean atomic displacement. Incorporation of thermal motion effects into equation (8) gives:

$$\mathbf{F}(hkl) = \sum_{j=1}^{N} f_j \exp(-B_j \frac{\sin^2\theta}{\lambda^2}) \exp(2\pi i(hx_j + ky_j + lz_j))$$

2.2 Molecular Replacement

The 'molecular replacement' technique encompasses a range of methods for structure determination which yield phase information from the relative position and orientation of two or more crystallographically independent but structurally-related molecules.

One application of the molecular replacement technique derives phase information and an initial model for the structure from determining the optimum superimposition of a known structure (or structures) onto the unknown structure. This six-dimensional problem is conventionally split up into it's rotation and translation elements, where first the rotation and then the translation is determined. Methods for the determination of both elements are described in the following sections.

The second aspect of the molecular replacement technique does not require the a model for the undetermined structure. Phase constraints derived from multiple copies of a molecule in the asymmetric unit were derived in direct space by Bricogne (1974) who formally showed them to be equivalent to the reciprocal space relations which came from the earlier work of Rossmann and Blow (1962, 1963). The use of this 'averaging' method has been valuable in the determination of virus structures (Acharya *et al.*, 1989; Luo *et al.*, 1987; Arnold and Rossmann, 1988; Arnold and Rossmann, 1986, and references therein) and also for other proteins (e.g. Jones *et al.*, 1989; Gaykema *et al.*, 1984; Argos *et al.*, 1975).

2.2.1 Rotation Functions

The rotation function relies on the properties of the Patterson function (Patterson, 1934):

$$P(\mathbf{u}) = \int_v \rho(\mathbf{x}).\rho(\mathbf{x}+\mathbf{u})dv$$

which may also be expressed in terms of the structure factor amplitudes:

$$P(\mathbf{u}) = \frac{1}{V}\sum_{\mathbf{h}} F(\mathbf{h})^2 \exp(-2\pi i \mathbf{h}.\mathbf{u})$$

and by Friedel's Law $F(\mathbf{h}) = F(-\mathbf{h})$:

$$P(\mathbf{u}) = \frac{2}{V}\sum_{\mathbf{h}} F(\mathbf{h})^2 \cos(2\pi(hu+kv+lw))$$

Where the vectors $\mathbf{u} = (u, v, w)$, $\mathbf{x} = (x, y, z)$, and $\mathbf{h} = (h, k, l)$ are the vectors in Patterson, real and reciprocal space respectively. The function $P(\mathbf{u})$ has significant values when the vector $\mathbf{u}$ corresponds to an inter-atomic vector. The Patterson function of an isolated molecule consists of a set of maxima, centered at the origin, whose positions correspond to inter-atomic vectors within the molecule (self-vectors). Introduction of the molecule into a unit cell introduces symmetry-related sets of self-vectors into the Patterson and also generates inter-molecular vectors (cross-vectors).

Rossmann and Blow (1962) proposed the rotation function to be the product of "stationary" and "rotated" Patterson functions (P_0 and P_1 respectively) evaluated within the volume U:

$$R(\Omega) = \int_U P_0(\mathbf{x}).P_1(\Omega\mathbf{x})d\mathbf{x} \tag{13}$$

where Ω is a rotation matrix, and U the volume of integration.

This function may be reformulated by expanding the Pattersons as their Fourier series (Rossmann and Blow, 1962). For an integration volume symmetric about the origin:

$$R(\Omega) = \frac{U}{V^3} \sum_{\mathbf{h}} (F_{\mathbf{h}}^2 \{ \sum_{\mathbf{p}} F_{\mathbf{p}}^2 G_{\mathbf{h},\mathbf{p}} \})$$

Where $\mathbf{h}$ and $\mathbf{p}$ are the reflection indices corresponding to Pattersons P_0 and P_1 respectively. The interference function G is the Fourier Transform of the integration volume. For a spherical integration volume $G_{\mathbf{h},\mathbf{p}}$ is large only where $\mathbf{h} \approx -\mathbf{p}\Omega$. Tollin and Rossmann (1966) found that a set of 27 symmetrically disposed points in $\mathbf{h}$ around $-\mathbf{p}\Omega$ was satisfactory for the calculation of $G_{\mathbf{h},\mathbf{p}}$.

The rotation function may be used in two ways. If P_0 and P_1 are calculated from the same data then $R(\Omega)$ is the *self*-rotation function where maxima in R

indicate that Ω is a symmetry operation which may reflect crystallographic or non-crystallographic symmetry. Where P_0 is calculated from observed data and P_1 is calculated from a model, $R(\Omega)$ is the *cross*-rotation function where peaks in R may correspond to the superimposition of the model onto the unknown structure on the basis of optimal matching of self-vectors. The presence of cross-vectors in P_0 may result in spurious matches between P_0 and P_1.

Lattman and Love (1970) showed that the cross-rotation function could be expressed directly in terms of reciprocal space quantities:

$$R(\Omega) = \sum_{\mathbf{h}} F_{obs}(\mathbf{h})^2 F_{model}(\mathbf{h}\Omega)^2$$

which has no integration volume definitions required if the model is an isolated molecule, and for which the value of $|\mathbf{F}_{model}(\mathbf{h}\Omega)|^2$ for non-integral values of $\mathbf{h}\Omega$ is obtained by linear interpolation.

Crowther (1972) expanded the Pattersons in the rotation function in terms of spherical harmonics, enabling the use of Fast Fourier Transforms in the calculation, considerably decreasing the execution time of the program. Crowther's algorithm introduces additional limitations however: the volume of integration is constrained to be spherical; and the maximum resolution permitted is coupled to the integration sphere radius by the number of Bessel function orders supplied.

Tanaka (1977) modified Crowther's Fast Rotation function to give a rotation function parameterised in spherical polar angles (ϕ, ψ, κ) rather than eulerian angles (α, β, γ). The definitions of (α, β, γ) and (ϕ, ψ, κ) are given in Chapter 4. In Tanaka's rotation function the value of $R(\phi, \psi, \kappa)$ can be extracted from an $R(\alpha, \beta, \gamma)$ map where P_0 is rotated by $\Omega(\alpha, \beta, 0)$ and P_1 by $\Omega(\alpha, \beta, \gamma)$. In such a

map the rotation angles are related by

$$(\alpha, \beta, \gamma) = (\phi + 90, \psi, \kappa)$$

using the angle conventions of Tanaka (1977) and Crowther (1972). This form of the self-rotation function is especially valuable for visualising non-crystallographic symmetry axes.

Huber (1965) evaluated the rotation function in direct space, rather than reciprocal space, by representing the Patterson functions P_0 and P_1 as lists of local maxima and performing the summation directly as in equation (13).

The disadvantage with the cross-rotation functions as expressed above is that the model matches with only one set of the crystallographically related self-vectors that occur in the P_0. Recently Yeates (1989) proposed a modified cross-rotation function which takes into account these symmetry-related sets of self-vectors. The modified function (N) simplifies to a function of the self-rotation function (Q) and the cross-rotation function (C):

$$N(\alpha, \beta, \gamma) = \frac{C(\alpha, \beta, \gamma)}{[\sum_{k=1}^{n} Q(\alpha'_k, \beta'_k, \gamma'_k)]^{1/2}}$$

where (α, β, γ) are related to $(\alpha'_k, \beta'_k, \gamma'_k)$ by:

$$\Omega(\alpha'_k, \beta'_k, \gamma'_k) = \Omega(\alpha, \beta, \gamma)^{-1} S_k * \Omega(\alpha, \beta, \gamma)$$

S_k is the rotation component of the k^{th} symmetry operation.

2.2.2 Translation Functions

The most obvious method of determining the position of a correctly oriented fragment in the unit cell is the crystallographic residual search where the residual:

$$R(\mathbf{t}) = \sum |F_{obs} - kF_{calc}(\mathbf{t})| / \sum F_{obs}$$

is evaluated for all positions ($\mathbf{t}$) of the model fragment within the asymmetric portion of the unit cell. This function is sensitive to the value of the scale factor k, whereas the related correlation coefficient is independent of the value of k:

$$CC(F) = \frac{\sum (F_{obs}^2 - \langle F_{obs}^2 \rangle) * (F_{calc}(\mathbf{t})^2 - \langle F_{calc}(\mathbf{t})^2 \rangle)}{\sqrt{\sum ((F_{obs}^2 - \langle F_{obs}^2 \rangle)^2 * (F_{calc}(\mathbf{t})^2 - \langle F_{calc}(\mathbf{t})^2 \rangle)^2}}$$

where the scale factor k has been omitted, (Fujinaga and Read, 1987). A number of other residual functions have been proposed (Nixon and North, 1976), but these will not be considered here.

Crowther and Blow (1967) derived three functions which gave information on the position of a correctly-oriented model relative to a local origin by evaluating the agreement between a set of calculated cross-vectors with the observed Patterson. The T function, defined as:

$$T(\mathbf{t}) = \int_v P_{01}(\mathbf{u}, \mathbf{t}) P(\mathbf{u}) d\mathbf{u}$$

has a large positive value where the calculated Patterson P_{01} between any two symmetry-related molecules optimally matches the observed Patterson P (i.e. $\mathbf{t}$ describes the molecular position). This function may be expressed in terms of the structure factors of observed and calculated data by:

$$T(\mathbf{t}) = \sum_{\mathbf{h}} F_{obs}(\mathbf{h})^2 \mathbf{F}_M(\mathbf{h}) \mathbf{F}_M^*(\mathbf{hA}) \exp(-2\pi i \mathbf{h}.\mathbf{t})$$

Where $\mathbf{A}$ is the rotational component of the symmetry operation, and $\mathbf{F}_M$ are

the calculated structure factors for the model relative to the local origin. $\mathbf{F}_M^*$ is the complex conjugate of $\mathbf{F}_M$. The T function may be modified by removal of intra-molecular (self) vectors from the native Patterson yielding the T_1 function:

$$T_1(\mathbf{t}) = \sum_{\mathbf{h}} \left(F_{obs}(\mathbf{h})^2 - \sum_{i=0}^{n-1} F_M(\mathbf{hA}_i)^2 \right) * \mathbf{F}_M(\mathbf{h})\mathbf{F}_M^*(\mathbf{hA}) \exp(-2\pi i \mathbf{h}.\mathbf{t})$$

where $\mathbf{A}_i$ is the rotational component of the i^{th} symmetry operation for the space group. The T and T_1 functions compare the cross vectors between any pair of symmetry-related molecules with the cross vectors in the observed Patterson between *all* molecules. The T_2 function allows matching of all cross-vectors between the calculated and observed Pattersons:

$$T_2(\mathbf{t}) = \int_v (P(\mathbf{u}) - P_M(\mathbf{u})).P_{calc}(\mathbf{u}, \mathbf{t}) d\mathbf{u}$$

where P_M is the Patterson of the model calculated with respect to the local origin (i.e. the self-vectors due to the model) and $P_{calc}(\mathbf{u}, \mathbf{t})$ is the Patterson of the model at a position $\mathbf{t}$ in the unit cell.

Harada and co-workers (1981) defined the TO/O translation function where the function TO has the form:

$$TO(\mathbf{t}) = \sum_{\mathbf{h}} F_{obs}^2 |\mathbf{F}_{calc}(\mathbf{h}, \mathbf{t})|^2 / \sum_{\mathbf{h}} F_{obs}^4(\mathbf{h})$$

(corresponds to the T_2 function above), whilst the O element function has the form:

$$O(\mathbf{t}) = \sum_{\mathbf{h}} |\mathbf{F}_{calc}(\mathbf{h}, \mathbf{t})|^2 / N \sum_{\mathbf{h}} \mathbf{F}_M(\mathbf{h})|^2$$

where N is the number of asymmetric units. $O(\mathbf{t})$ lies in the range 1–2, where $O(\mathbf{t}) = 1$ corresponds to no overlap between the symmetry related molecules and

$O(\mathbf{t}) = 2$ corresponds to complete overlap between them. The TO/O function is therefore given by:

$$\frac{TO(\mathbf{t})}{O(\mathbf{t})} = N \sum_{\mathbf{h}} F_M(\mathbf{h})^2 \frac{\sum_{\mathbf{h}} F_{obs}^2 |\mathbf{F}_{calc}(\mathbf{h},\mathbf{t})|^2}{\sum_{\mathbf{h}} F_{obs}^4 \sum_{\mathbf{h}} |\mathbf{F}_{calc}(\mathbf{h},\mathbf{t})|^2}$$

which introduces both the matching between all cross-vectors between the calculated and observed Pattersons and also a packing term.

Other translation functions have been described by Tollin (1966), Hendrickson and Ward (1976), Langs (1985) and Cygler and Anderson (1988b).

2.3 Refinement

The refinement of the atomic positions for protein structures suffers from the formally under-determined nature of the problem for data whose diffraction limit is ≈ 2.7Å (Hendrickson and Konnert, 1980), the resolution limit depending on the proportion of solvent in the unit cell. Only in rare cases for which high resolution data is available is it possible to refine a protein structure using free atoms (Watenpaugh *et al.*, 1973). Two approaches have been used to alleviate this situation: (1) to increase the number of observations; or (2) to decrease the number of variables. In case (1) stereochemical restraints derived from quantities observed in small-molecule structures are used to increase the effective number of observations (Waser, 1963; Konnert 1976; Hendrickson and Konnert, 1980; Hendrickson, 1985; Jack and Levitt, 1978). In case (2) stereochemical features are constrained to these pre-determined values, which serves to reduce the number of variables (Diamond, 1971; Sussman *et al.*, 1977). Outlines of a number of these procedures are given below.

More recently, with the increased availability of supercomputers, the incorporation of refinement using molecular dynamics techniques to increase the radius of

convergence of the refinement method has proved valuable (Brünger *et al.*, 1987; Fujinaga *et al.*, 1989). A molecular dynamics simulation of the protein in the unit cell is performed with an additional term due to the crystallographic residual in the potential function. The kinetic energy provided by the molecular dynamics simulation enables local energy minima (common in such non-linear systems) to be escaped and the effective radius of convergence of the refinement to be increased.

2.3.1 Crystallographic Least-squares

The conventional crystallographic least-squares method, as exemplified by the program PROLSQ (Hendrickson and Konnert, 1980; Hendrickson, 1985), seeks to minimise a function Φ by least-squares methods. Φ consists of the sum of functions that are themselves the sum of squared residuals of observed and calculated stereochemical quantities and the crystallographic residual:

$$\Phi = \sum_i \phi_i$$

The crystallographic residual term has the form:

$$\phi = \sum_{\mathbf{h}} \frac{1}{\sigma^2(\mathbf{h})}(F_{obs}(\mathbf{h}) - F_{calc}(\mathbf{h}))^2$$

Where $\sigma(\mathbf{h})$ is the standard deviation for the observation. Stereochemical terms for bond distances, bond angles, planarity, torsion angles and chiral centres have the general quadratic form:

$$\phi = \sum_n \frac{1}{\sigma_n^2}(X_{ideal,n} - X_{model,n})^2$$

In planar groups the atoms are restrained to the least-squares plane of the group. Chiral centres are restrained to the correct chiral volumes, and torsion angles are restrained to the nearest of the possible local minima around the rotation axis.

Non-bonded contacts are restrained according to the function:

$$\phi = \sum_m \frac{1}{\sigma^4(m)}(D_{min,m} - D_{model,m})^4$$

evaluated only where $D_{model,m} < D_{min,m}$. Non-crystallographic symmetry restraints are incorporated by restraining coordinates and thermal parameters for each copy of the object to the mean of the k copies (suitably rotated and translated):

$$\phi = \sum_i \frac{1}{\sigma^2(i)} \sum_k (X_{i,k} - \overline{X}_{i,k})^2$$

where X is either the (x, y, z) vector for the atom or the B factor for each atom. The variation of isotropic thermal parameters at j pairs of atoms connected by a bond or via an intervening atom can also be restrained by the function:

$$\phi = \sum_j \frac{1}{\sigma^2(j)}(B_{origin,j} - B_{target,j})^2$$

There are equivalent functions to resist excessive shifts in the model, or to restrain the occupancy factor for a given ligand to a mean value.

During refinement one seeks to minimise Φ, where Φ can be written as the sum of the squares of the deviations between M observations a_i and the corresponding value g_i, calculated from a model consisting of N parameters x_j:

$$\Phi = \sum_{i=1}^{M} w_i[g_i(x_1, x_2, \ldots, x_N) - a_i]^2$$

Where w_i is the weight for each observation. These N observational equations are non-linear in x_j. The equations are linearised about an approximate solution

$(x = x_0)$ by a Taylor expansion truncated at the first order term:

$$\Phi = \sum_{i=1}^{M} w_i[(g_i)_{x=x_0} - a_i + \sum_{j=1}^{N}(\frac{\partial g_i}{\partial x_j})_{x=x_0}\Delta x_j]^2$$

At the minimum the derivative of Φ with respect to x_j will be zero, and the resulting normal equations are:

$$2\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{i=1}^{M} w_i(\frac{\partial g_i}{\partial x_j})(\frac{\partial g_i}{\partial x_k})\Delta x_j + 2\sum_{j=1}^{N}\sum_{i=1}^{M} w_i(g_i - a_i)(\frac{\partial g_i}{\partial x_j}) = 0$$

which is more compactly represented as:

$$A\mathbf{x} - \mathbf{b} = 0$$

Where A is the $N * N$ normal matrix and $\mathbf{x}$ is the parameter shift vector. The value of Δx_j is obtained by evaluating $A^{-1}\mathbf{b}$. The conjugate gradients method for solving this problem has been described by Hendrickson and Konnert (1980).

Jack and Levitt (1978) achieved a similar form of refinement by using an energy function to describe the stereochemical restraints:

$$\Phi = \sum_{\text{bonds}} \frac{1}{2}K_b(b - b_0)^2 + \sum_{\substack{\text{bond}\\\text{angles}}} \frac{1}{2}K_\tau(\tau - \tau_0)^2 + \sum_{\substack{\text{dihedral}\\\text{angles}}} \frac{1}{2}K_\theta\{1 + \cos(n\theta - \delta)\}$$

$$+ \sum_{\substack{\text{non-bonded}\\\text{pairs}}} \varepsilon_{ij}\{(r_{ij}^0/r_{ij})^{12} - 2(r_{ij}^0/r_{ij})^6\}$$

which is similar to the stereochemical functions outlined above. The crystallographic residual is included as an 'energy' term and the refinement is performed by minimising the energy of the system. This energy minimisation is analogous to the least-squares refinement procedure in PROLSQ. The X-PLOR package (Brünger *et al.*, 1987) possesses a similar energy minimisation refinement method, whose details and differences from the method of Jack and Levitt (1978) have been discussed by Brünger *et al.* (1989).

2.3.2 Rigid Body Refinement

Diamond (1971) described a refinement procedure in which the number of variables was reduced by rigid body movement of groups by dihedral angles and by local variation in bond lengths and angles. The principal limitation in this method is that the quantity minimised:

$$\int_v (\rho_{obs} - \rho_{calc})^2 dV \equiv \frac{1}{V} \sum_{\mathbf{h}} (\mathbf{F}(\mathbf{h})_{obs} - \mathbf{F}(\mathbf{h})_{calc})^2$$

is tied to a particular set of phases ($\alpha(\mathbf{h})$).

Sussmann *et al.* (1977) developed the program CORELS using a refinement method of a number of constrained groups linked by distance restraints. Like the real space refinement of Diamond (1971) this method has a large radius of convergence and reduces the number of parameters in the refinement by the incorporation of rigid bodies.

The quantity minimised is the sum of three terms:

$$\Phi = w_F \phi_{DF} + w_D \phi_{DD} + w_T \phi_{DT}$$

where ϕ_{DF} is the term due to the structure factors:

$$\phi_{DF} = \sum_{\mathbf{h}} w_{\mathbf{h}} (F(h)_{obs} - F(h)_{calc})^2$$

the term ϕ_{DD} is due to the distance restraints:

$$\phi_{DD} = \sum_{d} w_d (D(d)_{obs} - D(d)_{calc})^2$$

and the term ϕ_{DT} restrains the structure to a 'target' structure:

$$\phi_{DT} = \sum_{i} w_i \sum_{j} (X_{i,j}^{TARGET} - X_{i,j})^2$$

The weight on this latter term (w_T) is set to zero in X-ray refinements.

The shifts on each of the groups are determined in the same manner as the shifts to the atoms in PROLSQ (Hendrickson and Konnert, 1980) given above. CORELS differs from the Diamond real space refinement in that it minimises the difference between the amplitudes of the structure factors and is not tied to a set of phases. However real space refinement has a larger radius of convergence and the inclusion of experimentally derived phase information may have advantages over a refinement method which does not restrain phases.

2.3.3 Molecular Dynamics Refinement

Molecular dynamics refinement consists of solving Newton's equations of motion for every atom under the influence of a potential that is composed of empirical stereochemical terms and the observational restraints. In the absence of the latter term the molecule would undergo a 'true' molecular dynamics simulation, subject to the limitations of the potential function. The addition of the observational restraints (F_{obs}, experimentally observed phases, non-crystallographic symmetry etc.) distorts the simulation so the molecule will no longer undergo motion representative of the 'free' molecule.

Nevertheless the extra energy introduced into the system by the assignment of velocities to each atom enables local energy minima to be traversed and conformational space to be explored more extensively. The description below refers to the X-PLOR program (Brünger *et al.*, 1987) but the principals are generally applicable to the MDXREF program (Fujinaga *et al.*, 1989).

Newton's equation of motion expressed in terms of the X-PLOR energy function is:

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

where the velocities give the 'temperature' of the system by Boltzmann's Law.

Kinetic energy enters into the system by the release of stereochemical strain and increased agreement with experimental data and this extra energy must be removed to maintain a constant temperature. The velocities may be periodically rescaled, or alternatively restrained by coupling to a heat bath. In this latter case the dynamics equation becomes:

$$m_i \frac{d^2 x_i}{dt^2}(t) = -\nabla_{x_i} E_{TOTAL} - b(T_0/T - 1) m_i \frac{dx_i}{dt}(t)$$

where b is the friction coefficient, and T_0 and T are the target and current temperatures respectively. Velocity rescaling was used in the conventional molecular dynamics simulations used in refinement, and the temperature coupling method was used in the 'slow' cooling molecular dynamics protocols (Chapter 6).

2.4 The X-PLOR Energy Function

The form of the energy function is critical to the performance of the energy minimisation and molecular dynamics simulations. In the following sections the X-PLOR energy function is outlined in more detail, omitting those facets not applicable to the method used in this study. The forms of the empirical and effective energy functions given below were used in the energy minimisation and molecular dynamics refinements with X-PLOR (Chapter 6).

The energy function is composed of two terms:

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

where the empirical term derives from the stereochemical restraints on the molecule and the effective term derives from the imposition of experimental restraints.

2.4.1 The Empirical Energy Term

The empirical term is made up of the following stereochemical terms:

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

where the covalent bond energy is given by

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

the bond angle energy is given by

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

the energy function for dihedrals and impropers (chirality and planarity) is given by

$$E_{DIHEDRAL} = E_{IMPROPER} = k(1 + \cos(n\phi + \delta)) \qquad \text{if } n > 0$$

or

$$= k(\phi - \delta)^2 \qquad \text{if } n = 0$$

Where n describes the periodicity of the dihedral, δ is the ideal dihedral and ϕ is the calculated dihedral. Impropers consist of terms for chirality and planarity which have harmonic restraints.

The non-bonded term has the form:

$$E_{NONBONDED} = E_{VDW} + E_{ELEC} + E^P_{VDW} + E^P_{ELEC}$$

where E^P_{VDW} and E^P_{ELEC} are equivalent functions to E_{VDW} and E_{ELEC} but between symmetry-related molecules.

The functions for E_{ELEC} and E_{VDW} are given by

$$E_{ELEC} = \sum_{i<j} f_{ELEC}(R_{ij}) + e_{14} \sum_{(i,j)\in\{1-4\}} f_{ELEC}(R_{ij})$$

$$E_{VDW} = \sum_{i<j} f_{VDW}(R_{ij}) + \sum_{(i,j)\in\{1-4\}} f_{VDW}(R_{ij})$$

where

$$f_{ELEC}(R_{ij}) = Q_i Q_j \frac{C}{\epsilon_0 R}(1 - \frac{R^2}{R_{off}^2})^2$$

$$f_{VDW}(R_{ij}) = (\frac{A}{R^{12}} - \frac{B}{R^6}) * SWITCH(R, R_{on}, R_{off})$$

The energy expression ignores all 1–2 and 1–3 interactions and the value of e_{14} modifies the 1–4 interaction energies. To save evaluating interactions between *all* atom pairs the VDW and $ELEC$ functions are evaluated fully for distances $R < R_{on}$ and reduced smoothly from R_{on} to reach zero at R_{off}. This is achieved by the $(1 - R^2/R_{off}^2)^2$ term in f_{ELEC} and by the $SWITCH$ function in f_{VDW}.

In the energy function used in this study the hydrogen-bond energy is contained within the electrostatic and van der Waals terms, and not explicitly expressed as an empirical energy term (as with some options).

2.4.2 The Effective Energy Term

The effective term consists of terms from the experimental observations:

$$E_{EFFECTIVE} = E_{XRAY} + E_{NCS}$$

The $XRAY$ term has contributions from the amplitude and phase:

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

where

$$E^A_{XRAY} = WA/NA \sum_{\mathbf{h}} W_{\mathbf{h}}[F_{obs}(\mathbf{h}) - kF_{calc}(\mathbf{h})]^2$$

and

$$E^P_{XRAY} = WP/NP \sum_{\mathbf{h}} W_{\mathbf{h}} S\{(\phi_{obs}\mathbf{h} - \phi_{calc}\mathbf{h}), \arccos(m(\mathbf{h}))\}$$

WA and WP are the user-set weights on the structure factor amplitudes and phases, and $W_{\mathbf{h}}$ the weights on individual observations. $NA = \sum_{\mathbf{h}} W_{\mathbf{h}} F_{obs}(\mathbf{h})^2$ and NP is the number of phases specified, and k is the scale factor determined such that E^A_{XRAY} is a minimum. The function S which determines the phase restraints is a square-well function and $m(\mathbf{h})$ is the figure of merit for each phase. $\mathbf{F}_{calc}$ is evaluated using a specially adapted Fast Fourier Transform algorithm for supercomputers (Brünger, 1989).

The non-crystallographic term restrains the individual segments to the mean structure of all the (rotated and translated) segments and is the same method as used in PROLSQ (Hendrickson and Konnert, 1980):

$$E_{NCS} = \sum w(x - \overline{x})^2$$

and the B factors are restrained by

$$(B - \overline{B})^2/\sigma^2_B$$

where $\overline{x}$ and $\overline{B}$ are the mean values from the N equivalent copies.

CHAPTER 3

PREPARATION AND CRYSTALLISATION OF FAB FRAGMENTS

3.1 Preparation of Fab Fragments

The preparation of the Gloop hybridoma cell lines has been previously described (Darsley and Rees, 1985a) and will not be repeated here. Preparation of purified antibody by Dr. R. E. Griest (unpublished results) used the following protocol: hybridoma cells secreting the Gloop antibodies were grown as ascites tumours in F1 hybrid SWRxBalb/c mice primed with pristane. The extracted ascites fluid was precipitated by 50% saturated ammonium sulphate followed by ion-exchange chromatography on diethylaminoethyl-Sephadex (Harrison and Mage, 1967). The purified IgG was digested by papain according to the method of Porter (1959) in order to generate the Fab fragments.

The cleavage conditions had already been determined for the Gloop1 and Gloop2 Fabs prior to the start of this project (R. E. Griest, unpublished results). The following sections give an account of the procedures used to obtain Gloop4 and Gloop1 Fab and the subsequent analysis of the samples.

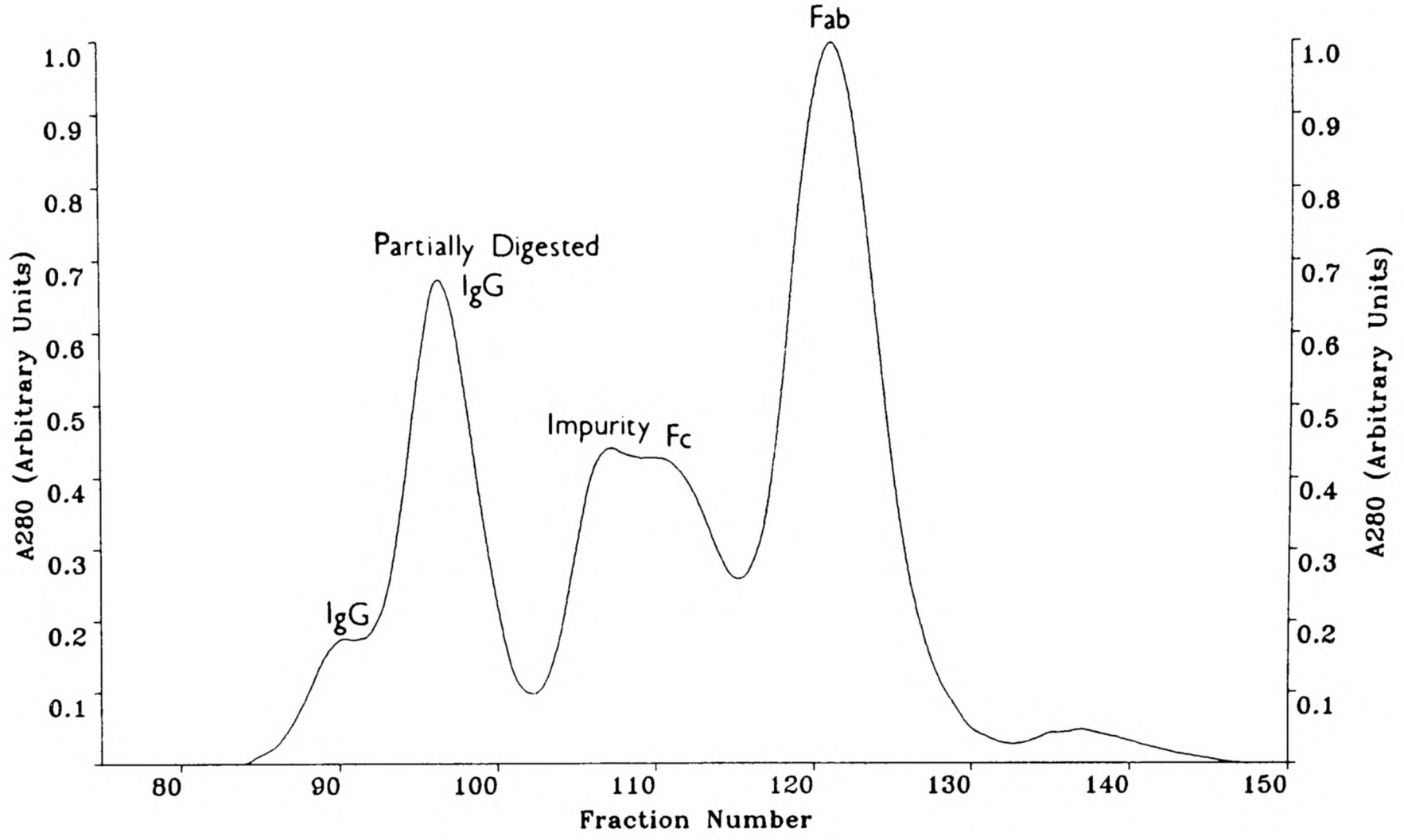
3.1.1 Preparation of Gloop4 Fab Fragments

The Gloop4 IgG used in this study had not been prepared by Dr. R. E. Griest, and had been stored at $-20°C$ for at least two years. Analysis by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) under reducing and non-reducing conditions (7.5% and 12% (w:v) polyacrylamide minigels with β-mercaptoethanol as the reducing agent) indicated that the majority of the protein was antibody.

The concentration of the protein was determined as approximately 1.8mg/ml by absorption at 280nm (A_{280}); ≈50mg of IgG was available for digestion. The conditions for Gloop4 digestion had not been previously determined and a small-scale trial digestion was performed to determine these conditions, based on the conditions determined for digestion of the Gloop2 antibody (R. E. Griest, unpublished results). A total of 2.9mg of Gloop4 IgG was used in this trial digestion. Mercuri-papain was activated by incubation in 20mM phosphate-buffered saline (PBS; Na_2HPO_4/NaH_2PO_4 and 0.1M NaCl) at pH7.0 for 15 minutes at 37°C in the presence of 1.0mg/ml sodium ethylene diamine tetracetate (NaEDTA) and 1.3mg/ml cysteine. Aliquots of this 'activated' papain mixture were added to the protein sample in 20mM PBS and 1.0mg/ml NaEDTA at pH7.0 and incubated at 37°C for up to 8 hours. Samples were taken before the addition of papain and at two-hourly intervals thereafter to monitor the progress of digestion. Papain:Ig ratios of 1:200, 1:500, 1:1000 and 1:2000 (w:w) were investigated in the presence of 0.23mg/ml or 1.3mg/ml cysteine. Reactions were quenched by adding iodoacetamide to a concentration of 3mg/ml and incubating the mixture at 4°C for at least 30 minutes.

Analysis of the results by SDS-PAGE showed that the digestion was essentially complete within 2 hours at a papain:Ig ratio of 1:500 (w:w). There was little difference between the two cysteine concentrations. The gels showed that there was a ≈5% (by weight) of impurity of approximate molecular weight 60kD and some residual antibody apparently resistant to papain cleavage even under the most extreme of the conditions tried. The remainder of the samples not used in SDS-PAGE analysis were put to one side to be pooled with the results of the full-scale digestion to conserve protein.

The rest of the protein was digested using the following conditions: 1:500



Solid Line — A_{280} Trace (Zeroed on Buffer)
2ml Fraction Volumes

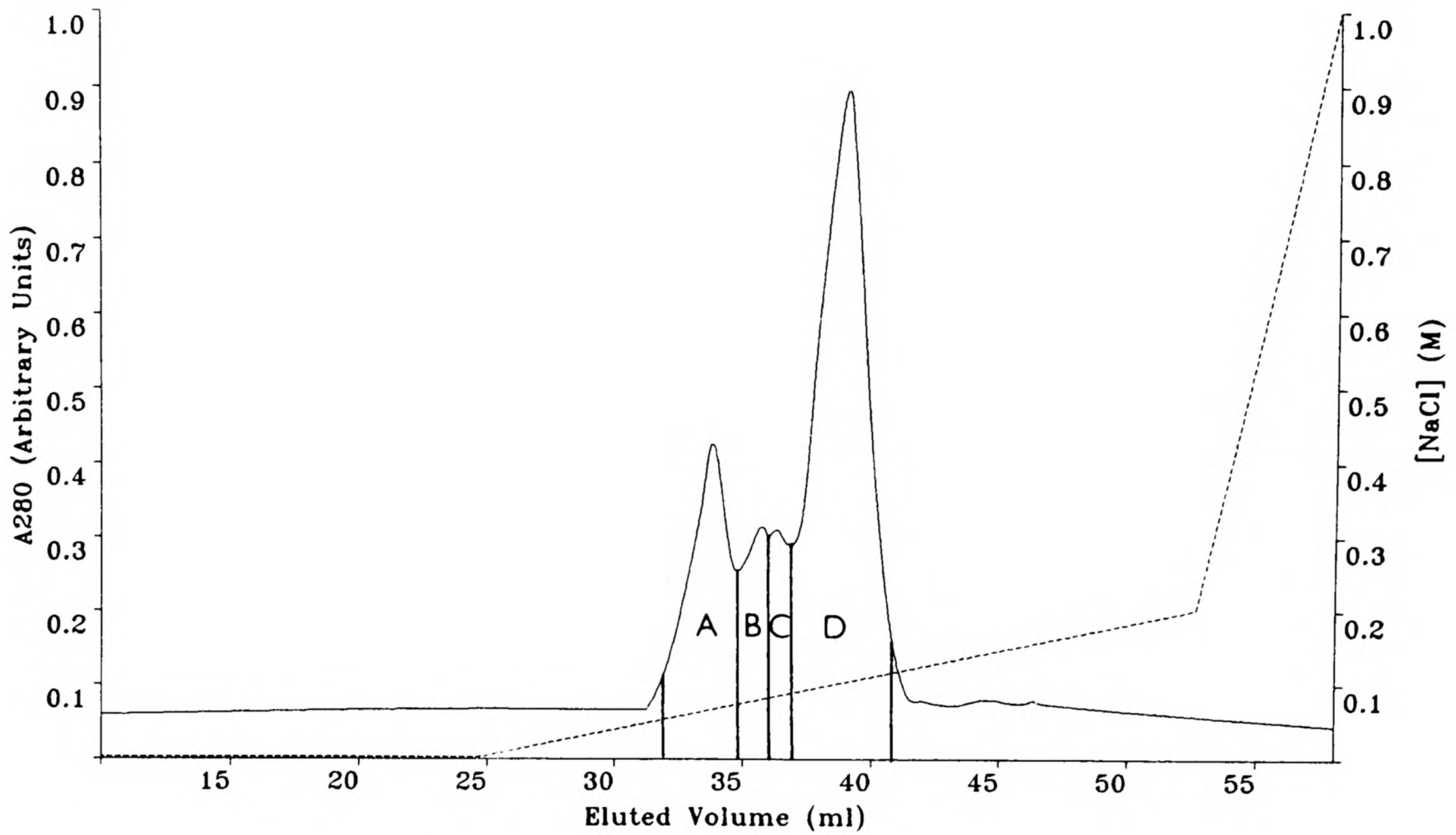
Figure 1: Gloop4 Digest Separation — S200 Column Trace

papain:Ig (w:w), 0.23mg/ml cysteine in 10mM PBS (0.1M NaCl) at pH7.0 with 1.0mg/ml NaEDTA. A second addition of 'activated' papain after 1 hour ensured that the digestion was as complete as possible and that papain autolysis was not responsible for the end of the digestion. The digest (including the samples from the trial digestion) was concentrated using a Diaflo PM-30 membrane (Amicon) with a 30kD cutoff under 40psi of nitrogen gas to a final volume of ≈5ml.

The concentrated sample was applied to a 700mmx26mm (length x diameter) Sephacryl S-200 (Pharmacia) sizing column in order to separate the Fab fragments from Fc, impurity and intact or partially digested IgG. The sample was eluted overnight in 20mM PBS at pH7.5. The protein content of the elute was monitored by A_{280} and a reproduction of this A_{280} trace is given in Figure 1. Analysis by SDS-PAGE of each of the peaks in the trace led to the assignment of the peaks as indicated in the figure. The trace showed a large quantity of partially-digested IgG, although there was hardly any remaining IgG. The Fab peak was well-separated from the Fc peak but had a trailing edge which may have been due to dissociated light chain or heavy chain fragments.

The bulk of the Fab peak showed no trace of contamination by other protein as judged by Coomassie-stained SDS-PAGE. The concentration of the pooled Fab peak was estimated to be 1.25 mg/ml by A_{280}, giving a total of $\approx$ 20mg Fab. The protein was concentrated by passing the solution through the Diaflo PM-30 membrane under 40psi of nitrogen to a final concentration of $\approx$ 10mg/ml.

Failure of the crystallisations at 1.25mg/ml and 10mg/ml concentrations (see below) led to the analysis of the Gloop4 Fab sample by ion-exchange chromatography (FPLC system, Pharmacia). Ion-exchange chromatography using a Mono-Q (Pharmacia) column with an 0.0–1.0M NaCl salt gradient in a 20mM Tris/HCl buffer at pH7.5 showed that the sample contained many different ionic species,



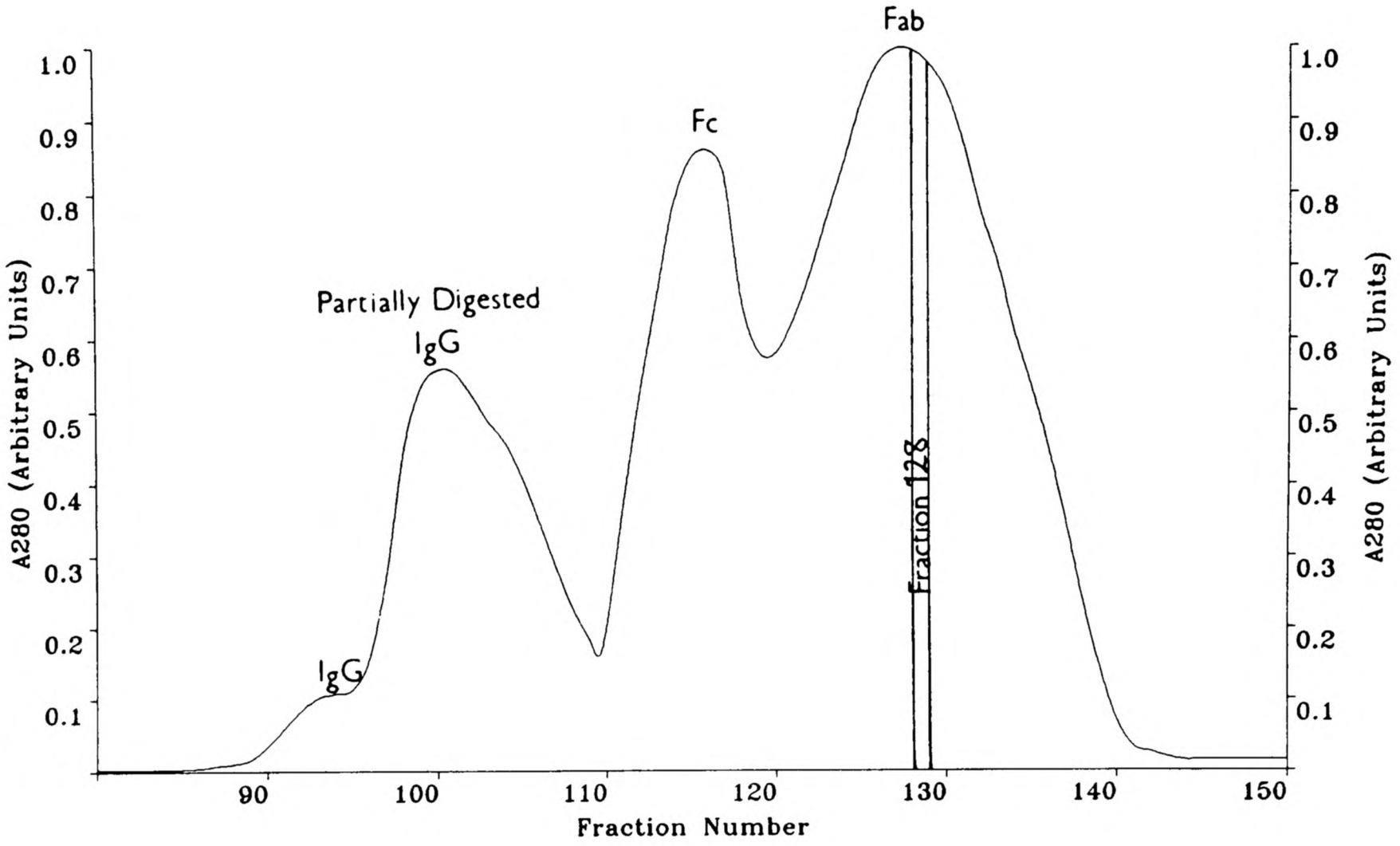
Vertical bars indicate pooling of eluted protein into samples A, B, C, D
Solid Line — A_{280} Trace (Zeroed on Buffer)
Broken Line — 0.0–1.0M Salt Gradient

Figure 2: Gloop4 Heterogeneity Analysis — Ion-exchange Column Trace

grouped into three main peaks (Figure 2), despite running as a single band on non-reduced polyacrylamide gels. The remaining protein was separated into four samples (A, B, C and D in Figure 2) and concentrated using Centricons (Amicon) with a 30kD cutoff. It was not possible to concentrate three of the samples to the desired 10mg/ml concentration due to the limited quantity of material. The Fab concentration of the samples A, B, C and D was estimated to be 4.0, 2.7, 3.6 and 10mg/ml respectively after concentration, on the basis of the volume of sample produced (neglecting loss of Fab on the filter). The total quantity of Fab remaining was approximately 0.40mg, 0.27mg, 0.36mg and 1.0mg for samples A, B, C and D respectively.

3.1.2 Preparation of Gloop1 Fab fragments

Previous work by Dr. R. E. Griest (unpublished results) had determined the best conditions for the Gloop1 Fab digestion. 300mg of purified Gloop1 IgG was digested at a papain:Ig (w:w) ratio of 1:1000 in 10mM PBS (0.1M NaCl) at pH8.0, in the presence of 1.3mg/ml cysteine and 1.0mg/ml NaEDTA. The reaction was stopped by adding iodoacetamide to 2.0mg/ml and incubating at 4°C for 30 minutes. The reaction mix was run on the S-200 gel filtration column and a reproduction of the A_{280} trace is given in Figure 3, where the assignments are based on SDS-PAGE analysis of the peak contents. As with the Gloop4 digestion above, the Fab was separated from the other constituents of the digest, although the resolution was not quite as good. Although the Fc peak appeared to merge into the Fab peak the SDS-PAGE results indicated that only the first few fractions were affected, and most of the Fab was without contaminant. The total amount of Fab in the peak was estimated to be $\approx$ 150mg from the trace, at a mean concentration of 7mg/ml. The sample used in the crystallisation trials (see below) was fraction 128 from the top of the Fab peak at $\approx$12mg/ml (fraction marked in Figure 3), and required no



Solid Line — A_{280} Trace (Zeroed on Buffer)
2ml Fraction Volumes

Figure 3: Gloop1 Digest Separation — S200 Column Trace

further concentration before use.

The multiple ionic species encountered with the Gloop4 Fab sample prompted the investigation of the Gloop1 Fab sample. Gloop1 was applied to a Mono-Q column in 20mM Tris/HCl buffer at pH9.4 and eluted by a salt gradient of 0.0–1.0M NaCl. Although in this case the Gloop1 sample *had* yielded crystals (see below) without further purification, the sample was again shown to be heterogeneous. A reproduction of the A_{280} trace of Gloop1 is shown in Figure 4.

3.2 Crystallisation of Fab Fragments

Prior to the start of this project approximate crystallisation conditions had been determined for Gloop2 and Gloop1 using hanging drop vapour diffusion (Wlodawer and Hodgson, 1975). Small crystals which were sensitive to radiation damage had been obtained for both Gloop1 and Gloop2 (R. E. Griest, unpublished results). The precipitant used was polyethylene glycol 6000 (Sigma Chemical Company) (PEG6000) at a concentration of $\approx$ 20% (w/v) in the presence of 20mM Na_2HPO_4/NaH_2PO_4 buffer at pH values in the range 6–8, and 0.0–0.3M NaCl. No data had been collected or space groups determined for these crystals, although the Gloop2 crystals had been shown to diffract to at least 4Å on a precession camera (B. J. Sutton, unpublished results).

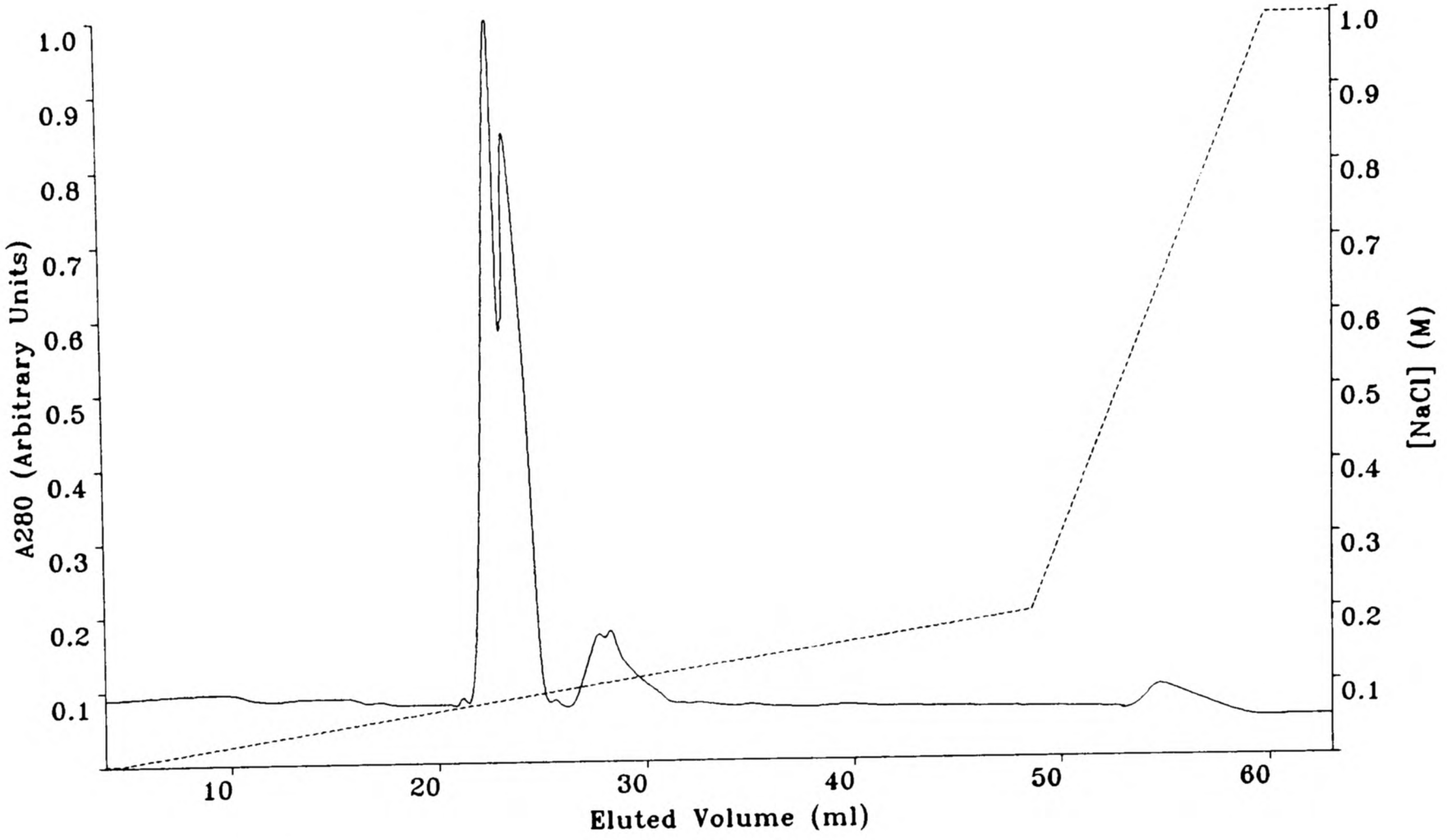
3.2.1 Crystallisation of Gloop4 Fabs

Crystallisation solutions used in the trials had the following compositions:

PEG6000 at 40% (w:v)

80mM Na_3PO_4 buffer at pH 6–8 in 0.5 pH steps.

0.0–1.2M NaCl in 0.4M steps.



Solid Line — A_{280} Trace (Zeroed on Buffer)
Broken Line — 0.0–1.0M Salt Gradient

Figure 4: Gloop1 Heterogeneity Analysis — Ion-exchange Column Trace

The pH of the solutions were adjusted such that the required pH was attained after four-fold dilution with distilled water. Some of the high-salt solutions formed two immiscible phases, which merged together on heating in a water bath at 40°C, but re-separated on returning to room temperature. Crude analysis of the contents through AgCl precipitation, flame tests and evaporation showed that the denser phase was phosphate- and salt-rich, and the less dense phase was PEG6000-rich. However both phases apparently contained all three components. In these cases the less-dense PEG6000-rich phase was used for the crystallisations. An explanation for this phenomenon may be that PEG6000 competes with the phosphate and salt for water so well at 40% (w:v) that the phase separation becomes thermodynamically favourable.

Crystallisation trials were carried out in 10μl hanging drops on siliconised glass cover slips equilibrated over 0.5ml wells in 24-well tissue culture plates (Flow Laboratories Inc.). The seal was made air-tight by Vaseline. The hanging drops consisted of 7.5μl protein sample plus 2.5μl of crystallisation solution (i.e. 10% (w:v) PEG6000 in the drop). The wells consisted of 0.31ml of crystallisation solution diluted to 0.5ml by distilled water (i.e. 25% (w:v) PEG6000 in the well). The effective pH and salt concentrations in the drop were modified due to the protein sample buffer. The volume of crystallisation solution added to the drop was minimised in order to dilute the protein as little as possible. Use of PEG6000 as the precipitant limited the maximum concentration of the crystallisation solutions, as PEG6000 at more than $\approx$40% (w:v) becomes very viscous and difficult to handle.

The crystallisation trials were set up over the full range of 20 conditions (pH6.0–8.0 in 0.5pH steps and 0.0–0.3M NaCl in 0.1M steps) for the 1.25mg/ml and 10mg/ml Gloop4 Fab samples at room temperature ($\approx$ 16°C in the crystal growing room). After 6 months no crystals had formed in the 1.25mg/ml Fab drops, nor

was there any precipitate, suggesting that the protein concentration was too low for these crystallisation conditions. In the 10mg/ml Fab drops, however, there was precipitate, but no crystals had grown after 6 months.

After the Gloop4 Fab was partially purified by ion-exchange chromatography the resulting samples (A, B, C and D) were subjected to crystallisation trials under the same conditions as the previous samples. Crystals formed in drops with samples C (3.6mg/ml) and D (10mg/ml) after about two weeks. The A and B samples produced no crystals after six months. Crystals grew from drops containing sample C at pH6.0 and pH6.5 at 0.0M and 0.1M NaCl as thin plates up to 1mm in length or shorter but thicker needles of maximum dimension 0.2mmx0.02mmx0.02mm. These crystals were too small to be of use in X-ray diffraction experiments. The sample D crystals originally grew at 16°C, then re-dissolved due to temperature fluctuations in the crystal growing room, and subsequently re-grew when transferred to 20°C. The crystals grew from pH6.5 and pH7.0 with 0.0M and 0.1M salt to a maximum dimension of 0.5mmx0.3mmx0.15mm as rectangular cross-sectioned needles. These crystals were used in the data collection discussed in Chapter 4.

In addition to the uncomplexed Fab crystallisation trials, co-crystallisation of Fab:HEL and Fab:loop peptide complexes was attempted. Equimolar quantities of antigen and Fab from samples A and D were incubated overnight at 4°C and subjected to crystallisation trials using the same conditions as given above. No crystals were obtained with sample A, but sample D gave crystals with both HEL and loop peptide at pH6.0 and 7.0 in 0.0M or 0.1M salt. However all the crystals were in the form of very thin plates or clusters of very thin needles ("sea-urchins") and were unsuitable for X-ray analysis.

By the end of these crystallisation trials all of the Fab samples had been used in crystallisation trials. Since no more Gloop4 antibody has been prepared to date

it has not proved possible to repeat the crystallisation of the uncomplexed Fab on a larger scale.

Gloop4 crystals appeared to be quite stable in the hanging drops, but it was apparent that the crystals had deteriorated after four months from their diffraction patterns. In future therefore, it may be worth trying to stabilise the crystals in a higher concentration of PEG6000 for long-term use.

3.2.2 Crystallisation of Gloop1 Fabs

Crystallisation trials on the Gloop1 Fabs used the same procedure as the Gloop4 crystallisation trials. Crystals were grown *before* the FPLC analysis of the Gloop1 sample indicated that the Fab was heterogeneous.

Crystallisations were set up using Gloop1 Fab sample at 12mg/ml in 20mM PBS buffer (i.e. the S-200 elution buffer), with 7.5μl of protein sample and 2.5μl crystallisation solution mixed together to form the 10μl hanging drop (10% (w:v) PEG6000). For the wells, 0.25ml of crystallisation solution was diluted to 0.5ml by distilled water (20% (w:v) PEG6000 in the wells). Crystals grew readily from these drops between pH6–7.5 at 0.1M or 0.2M salt. The largest crystals (pH6.0/0.1M NaCl, pH7.0/0.2M NaCl) grew as ellipsoids with no clearly defined faces up to 1mm in the largest dimension and $\approx$ 0.3mm in diameter. Smaller crystals were obtained at pH6.5/0.1M NaCl and pH7.5/0.1M NaCl which grew to $\approx$ 0.5mm long.

The lifetime of the crystal in the drop appeared to be inversely proportional to it's size — the largest crystals lasted barely two weeks before the crystals showed signs of serious deterioration (cracked surfaces and eventually total disintegration) whereas the smallest crystals (from pH7.0/0.1M NaCl, up to 0.2mm long) lasted for over two months before any sign of deterioration. The crystals from pH6.5/0.1M NaCl were used in data collection, since the largest crystals had deteriorated by the

time that data collection had commenced.

Not only did the crystals deteriorate rapidly, but the crystallisations did not appear to be readily reproducible despite using the same solutions, samples and methods. Temperature may have been a factor here, as the temperature stability of the crystal growing room was quite poor. It is possible, however, that the Fab sample changed over time: Gloop1 samples prepared for NMR experiments on the binding of peptide to the Fab were initially a single peak by FPLC analysis, but analysis of the samples after two months showed the presence of a second peak (R. E. Griest, unpublished results). It may prove necessary to purify the sample each time before crystallisation attempts, in order to obtain reproducible crystals.

Co-crystallisations with HEL and loop peptide were set up with the Fab incubated with equimolar quantities of antigen overnight at 4°C and crystallised using the same conditions as the Gloop1 Fab, except that the well concentrations were as used for Gloop4 (25% (w:v) PEG6000). The same sample of Gloop1 was used as had proved successful with the uncomplexed Fab crystallisations. Gloop1-loop crystallisations gave semi-crystalline 'blobs' at pH7–8 with 0.1–0.3M NaCl, but no identifiable individual crystals were formed and the blobs were not separable into individual crystals. Gloop1-HEL crystallisations gave a few crystals which diffracted very strongly and which turned out to be the tetragonal form of HEL as judged by the cell dimensions and diffraction pattern. No other crystals were obtained.

3.3 Discussion

Gloop4 is not unique in its requirement for purification. Bott *et al.* (1982) used isoelectric focussed gel electrophoresis to purify the J539 Fab fragment and thereby obtained better crystals, and Cygler *et al.* (1987) also used this method to purify the HED10 Fab fragment. All crystals used for data collection (Chapter 4), with the

exception of the Gloop1 Fabs, were obtained from samples purified by ion-exchange crystallography.

Using PEG6000 as a precipitant it was possible to grow crystals of Gloop1 and Gloop4 uncomplexed Fabs, and work by Dr. R. E. Griest showed even better results under these conditions with the Gloop2 Fab. Complex crystals, with either loop peptide or HEL, did not form under these conditions and it would be prudent to try other conditions for the crystallisation of these complexes (ammonium sulphate, sodium phosphate, sodium citrate). The current lack of Gloop4 has stopped all investigation into the structure of this Fab. The NMR experiments (see Chapter 1) were a major competitor for the Gloop1 protein. Since the Gloop2 Fab crystallised readily and reproducibly, and was of more interest for the interpretation of mutagenesis results (Chapter 1), the subsequent work concentrated on the Gloop2 Fab as a source of crystals for X-ray work. However the Gloop1 and Gloop4 crystallisation conditions are now well-established (although they would benefit from further refinement), yielding crystals of a suitable size for X-ray diffraction studies and provide the first step toward the structural studies of this family of homologous antibodies.

Gloop2 Fab crystals used in this project were grown over a range of pH (6–8) and salt concentrations (0.0–0.2M) by Dr. R. E. Griest and the results of the crystallisations are briefly described here. Gloop2 Fab crystals were grown from 10μl hanging drops containing 10–15% (w:v) PEG6000 equilibrated over wells containing 18–20% PEG6000. Buffer and salt conditions were as for the Gloop1 and Gloop4 crystallisations. Crystals grew as long needles with a rhomboid cross-section up to 1.5mm x 0.25mm x 0.15mm from pH6.0–8.0 at 0.0M and 0.1M NaCl. The largest crystals were often intergrown and required manual cleavage before crystal mounting. The shelf life of the crystals was in excess of four months.

Gloop5 Fab′ fragments have been prepared by pepsin digestion of the IgG1, which was particularly resilient to papain digestion (R. E. Griest, unpublished results). The Fab′ has been crystallised from PEG6000 solutions, and data collected on the area detector (whose setup is discussed in Chapter 4) indicated that the space group was $P3_121$ or enantiomorph ($a = 112.4$Å, $c = 102.0$Å), with a diffraction maximum of 3.5Å (R. E. Griest, unpublished results). Rotation function studies are underway on this Fab, but no solution has been forthcoming to date (R. E. Griest and P. D. Jeffrey, unpublished results).

CHAPTER 4

DATA COLLECTION AND PROCESSING

4.1 Data Collection Using the Area Detector

All data used in this project were collected on the in-house Nicolet-Xentronics area detector equipped with a 3-circle goniostat. Two detectors were used in the course of the data collection presented below: the second differed from the first by an internal hardware modification which allowed an increased incident photon flux to be used and reduced the sensitivity of the detector response to vibration. The detector was mounted on a Rigaku RU200H rotating anode generator giving CuK_α ($\lambda = 1.54$Å) radiation. Unwanted X-ray wavelengths were removed with a graphite monochromator. Modifications had been made to the commercially supplied system by extension of the collimator and the backstop, to minimise the path of the direct beam in air (D. I. Stuart and E. F. Garman, unpublished results): (1) the distance between the first pin-hole in the collimator and the second pin-hole was extended from 3.2cm to 4.2cm, and distance between the collimator exit aperture and the crystal was reduced to ≈4mm; (2) a new backstop was made to minimise the path of the direct beam from the crystal to the backstop whilst keeping the backstop shadow as small as possible. These hardware modifications led to a significant decrease (60%) in background scatter (which accounts for the majority of the photon flux over the detector surface), although some restrictions were found in the mounting of crystals on goniometer heads due to the decreased clearance between backstop and collimator. All data were collected with an 1.0mm pin-hole nearest the X-ray source and an 0.5mm pin-hole nearest to the crystal.

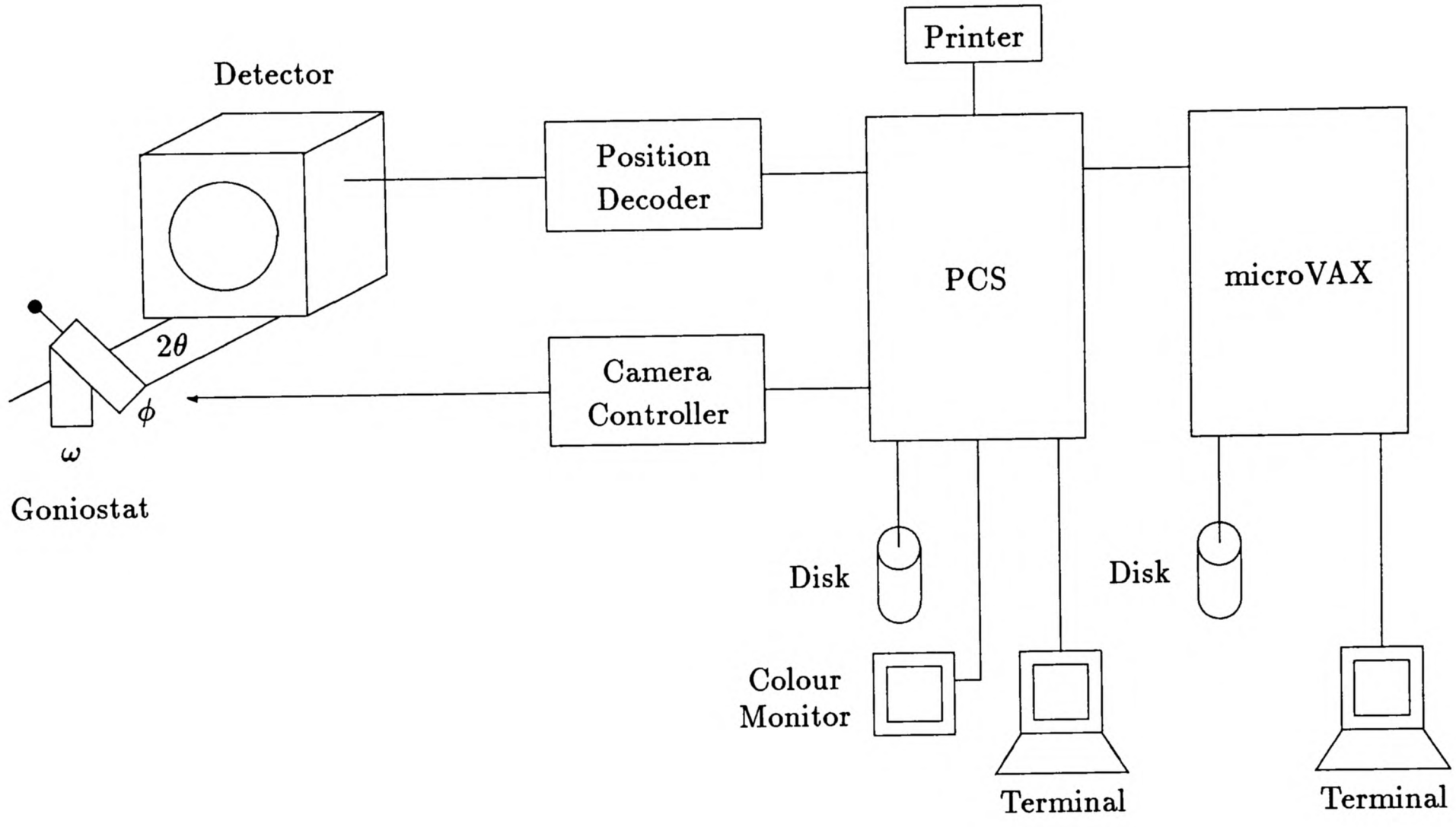


Figure 1: Schematic Representation of the Area Detector Setup

4.1.1 Data Collection Hardware and Software

The principles of operation of the Nicolet-Xentronics area detector, developed by R. Burns (Durbin *et al.*, 1986) are similar to that of the multiwire area X-ray diffractometer at San Diego (Hamlin, 1985), and will be briefly described below. The setup used in data collection is very similar to that described by Blum and co-workers at Harvard (Blum *et al.*, 1987), and the software for the control of the goniometer and for data collection is derived from the Harvard package (Durbin *et al.*, 1986). A Motorola 68010-based PCS computer, running an UNIX operating system controls the goniostat and detector via the camera controller and position decoding microprocessors. A colour monitor, connected to the PCS, enables inspection of data as they are collected and hard copies are possible using a dot-matrix printer. The PCS is connected to a MicroVAX computer by an Ethernet link. A schematic description of the various components to the system is given in Figure 1.

The 3-circle goniometer head has the ϕ axis fixed at 45° to the vertical ω axis (i.e. $\chi = 45°$) which serves as the rotation axis for the oscillation camera. The detector is mounted on a moveable 2θ arm whose axis is coincident with the ω axis. The limited aperture size of the detector (diameter 11.5cm) requires the use of non-zero 2θ angles for collection of higher resolution data, which can also necessitate an increase in the total ω range to be traversed (dependent on space group symmetry and crystal setting).

X-ray photons cause ionisation of xenon gas (at 4 atmospheres) in the detector chamber. The high pressure of the gas ensures that ionisation events occur just behind the 1mm concave beryllium window, inhibiting spot spreading due to parallax. The concave nature of the beryllium window also helps to minimise parallax by ensuring the electric field lines are approximately parallel to the incident X-ray beams. Electrons produced by the ionisation events are accelerated towards the

anode wires, at ≈5kV above ground, causing secondary ionisation events and amplifying the signal. The cathodes consist of two parallel planes of wires — one with the wires running vertical and the other horizontal, enabling the determination of the position of the ionisation events in two dimensions. The signals produced by the xenon cations as they reach the cathode wires are passed to the position decoder which decodes the signals as the (X, Y) coordinate in the anode planes with an effective resolution of 0.2mm. The data are stored as a 512x512 pixel array (each pixel 0.2mm on an edge) in memory and written to disk at the end of the exposure. Efficiency of data collection depends on the count rate in two ways: (1) the decoder can store a second event that arrives whilst it is decoding the first, but loses the third and subsequent events; (2) the occurrence of high photon flux causes spots to spread out, reducing the effective resolution of the detector. The '% late' parameter reported by the PCS is a direct measure of the first effect — it represents the number of events arriving at the PCS whilst the previous one is being decoded, as a proportion of the total counts recorded. The proportion of events lost due to multiple event arrival is approximately the square root of this value (Blum *et al.*, 1987) and usually lies in the range 0–10%. The '% late' is also a monitor of the second effect since it is also a function of photon flux. The first detector had an effective data recording rate limit of 29kHz which is equivalent to ≈30% lates. The second detector had an upper limit of 43kHz (≈40% lates).

To eliminate local heterogeneity in detector response due to irregularities in the wire spacing, a correction is performed using a radioactive ^{55}Fe X-ray source. A 'flood field' correction is recorded with the detector being exposed to the ^{55}Fe X-radiation. The counts from each pixel are locally re-binned to give a smooth response over the detector surface using the program CORRECT (Blum *et al.*, 1987). This correction does not compensate for any non-local sensitivity differences

in the detector response. The correction is applied on-line whilst data are collected and thus recorded frames have already been corrected in this manner. An additional image is required before data collection — the 'brass plate' image is collected using the ^{55}Fe source: a brass shield which has holes drilled at precise intervals is placed over the detector face. The distribution of the spots in the brass plate image in comparison with their known position on the brass plate enables a correction for geometric distortion to be made.

Data are collected as series of consecutive non-overlapping 'frames', with the crystal oscillating about the vertical ω axis over a small angular range. The oscillation ranges used are smaller than that used in conventional oscillation photography — each diffraction point is thus divided into a number of slices in ω according to the angular width of a frame ($\Delta\omega$), from which the three-dimensional profile of the reflection can be reconstructed. The small angular range also increases the signal-to-noise ratio of the reflections. The crystal is oscillated about the ω axis, rather than simply rotated, to minimise errors due to fluctuations in the X-ray beam and because the opening of the shutter and the ω axis motors are not synchronised (Blum *et al.*, 1987).

Each frame is stored as a 512x512 byte array with appropriate header information on exposure time etc. Pixels which contain more than 255 counts have their true count stored in the header and are flagged as 'overflows'. The number of overflows on a frame thus depends on the number of diffraction maxima and their strength. Frames are transferred from memory either to the PCS hard disk and subsequently to a microVax via an Ethernet using the program GETXEN (G. L. Taylor, unpublished results), or directly to the microVax disk using a TCP/IP protocol.

4.1.2 Data Collection Protocol

Crystals were mounted in thin-walled 0.7 or 1.0mm diameter glass capillary tubes, sealed with vacuum wax, in the conventional manner. The longest dimension of the crystals were aligned with the tube axis. Capillary tubes were mounted onto goniometer heads in one of two ways: (1) mounted at 45° to the ϕ axis on an angle bracket (c.f. Derewenda and Helliwell, 1989), allowing the tube to be parallel or perpendicular to the rotation axis during data collection; or (2) such that they were parallel to the ϕ axis and therefore inclined at 45° to the ω rotation axis during data collection. This first mounting method was preferred initially since the the variation in absorption was minimised, but the second method was adopted after vibrations from the generator introduced a high risk of crystal slippage if the tube was vertical.

The crystal to detector distance (XTDD) was determined by the necessity to resolve adjacent spots. The minimum XTDD in cm is given approximately by the maximum dimension of the primitive cell divided by 8 (Howard *et al.*, 1987). For the Gloop2 primitive cells (Table 2) this corresponded to $\approx$10.5cm, at which distance the resolution at the detector edge was $\approx$3.0Å. However the near-circular projection of the aperture onto the pixel array means that only a very small proportion of the highest resolution reflections will ever be collected at non-zero 2θ angles and so the 'effective' maximum resolution is a little lower than this. To compensate for this effect the detector was moved on the 2θ axis such that it recorded slightly higher resolution data at the edge of the detector than that desired in data collection. The first Gloop2 dataset, and the Gloop1 and Gloop4 datasets were collected using a XTDD which was greater than that required by the cell dimensions, because at that stage the cell dimensions were not established and so the setting of the XTDD erred on the side of caution.

Frames were collected at 0.25° or 0.20° per frame ($\Delta\omega$). The greater of the two frame sizes was preferred due to the large angular range that needed to be traversed for the space groups listed below (especially $P1$). The mosaic spread of the crystals (due to the angular spread of the orientation of the crystalline blocks that make up the crystal) also made the use of the lesser frame size unnecessary, as the spots profiles were sufficiently spread out in space to be sample adequately by the larger frame size. Exposures varied according to the strength of diffraction of the crystal and the power of the generator — for the largest crystals with the first detector the generator could only be run at 40kV 50mA because the count rate at the short XTDD used was too great to permit full power (60kV 90mA). Usually the generator was run at 50kV 60mA. Exposures were adjusted to be a compromise between collecting strong data and the need to collect large angular ranges. Exposures of 150–600 seconds/frame were used, usually giving about 100 overflows/frame. The number of overflows and the appearance of frames on the colour monitor were used as a primitive monitoring technique for crystal decay. Checks for crystal twinning and disorder in the diffraction pattern were made using the program VIMVS (G. L. Taylor, unpublished results) enabling the superposition of sequential frames to be inspected on a microVAX 3000 workstation screen.

Data were collected at room temperature ($\approx$ 16°C). Cooling was not available when the Gloop1 and Gloop4 datasets were collected, and cooling of Gloop2 crystals to $\approx$ 7°C using a Peltier cooling device (E. F. Garman, unpublished results) appeared to disorder them.

In the case of Gloop1 and Gloop4 datasets an error in the setting of the motor speeds for the camera controller led to some frames being collected more than once and other frames being missed out altogether. The ω axis *had* traversed the correct range in each case, but in an uneven fashion. This error, which grossly distorted the

Fab	Space Group	No. of Crystals	Exposure (sec./frame)	$\Delta\omega$ (°)	2θ (°)	XTDD (mm)	Omega Range (°)	Code
Gloop1	$P1$	1	200	0.25	14.0	140	295	G1P1
Gloop2	$P1$	1	100–300	0.25	0.0	170	160	G2P1A
Gloop2	$P1$	1	400	0.25	7.0	115	85	G2P1B
Gloop2	$P1$	1	200–300	0.25	0.0	105	155	G2P1C
Gloop2	$P1$	2	300–500	0.25	7.0	103	248	G2P1D
Gloop2	$P1$	1	300	0.25	7.0	103	130	G2P1E
Gloop2	$P1$	1	400–500	0.25	7.0	103	55	G2P1F
Gloop2	$P2_1$ 'A'	1	200–300	0.20	0.0	105	95	G2P2_1G
Gloop2	$P2_1$ 'A'	1	300–400	0.25	7.0	115	140	G2P2_1H
Gloop2	$P2_1$ 'B'	1	200	0.25	5.0	120	46	G2P2_1I
Gloop4	$C2$	1	200	0.25	14.0	140	210	G4C2

Space group – see Table 2 and discussion in text.
$\Delta\omega$ – frame width in degrees.
2θ – detector offset in degrees.
XTDD – crystal to detector distance.
Omega range – total amount of data collected in degrees.

Table 1: Data Collection Details

spot profiles, rendered the data of little use except for space group determination and was particularly disappointing in view of the considerable amount of data involved and the long lifetime of the crystals in the beam (more than 3 days). Subsequent data collected on Gloop1 was unprocessable due to crystal twinning, whilst the limited supply of Gloop4 crystals and the effects of aging rendered subsequent data collection on this Fab impractical. Thus the only datasets for Gloop1 and Gloop4 that are listed in Table 1 are the data mentioned above.

Over 60 crystals of Gloop2 were examined in order to collect the datasets listed in Table 1. Intergrowing of the Gloop2 needles proved to be a major limiting factor in finding appropriately-sized crystals. Many crystals did not diffract beyond 4.0Å and were discarded, although diffraction to 2.6Å was observed for the very best crystals. Space groups and cell dimensions were determined at the data processing stage and all data were collected on the assumption that the crystal form was $P1$. In particular the $P2_1$ 'A' form of Gloop2 was morphologically identical to the $P1$ form and had a similar cell, and the monoclinic nature of those crystals was only discovered during data processing. With the benefit of hindsight the $P2_1$ 'A' form crystals grew at the low pH (5.0–7.0) end of the conditions under which crystallisation occurs, and were smaller and more fragile than the $P1$ crystals. Only one crystal of the Gloop2 $P2_1$ 'B' crystal form has ever been grown. This crystal had a large mosaic spread and was very radiation sensitive. The short lifetime of this single crystal in the X-ray beam only permitted a partial dataset to be collected.

4.2 Data Processing

All data were processed using the XENGEN suite of programs (Howard *et al.*, 1987) with versions 1.2 and 1.3 of the software, with the exception of the first data set collected (G2P1A in Table 1) which was processed with a pre-release version of

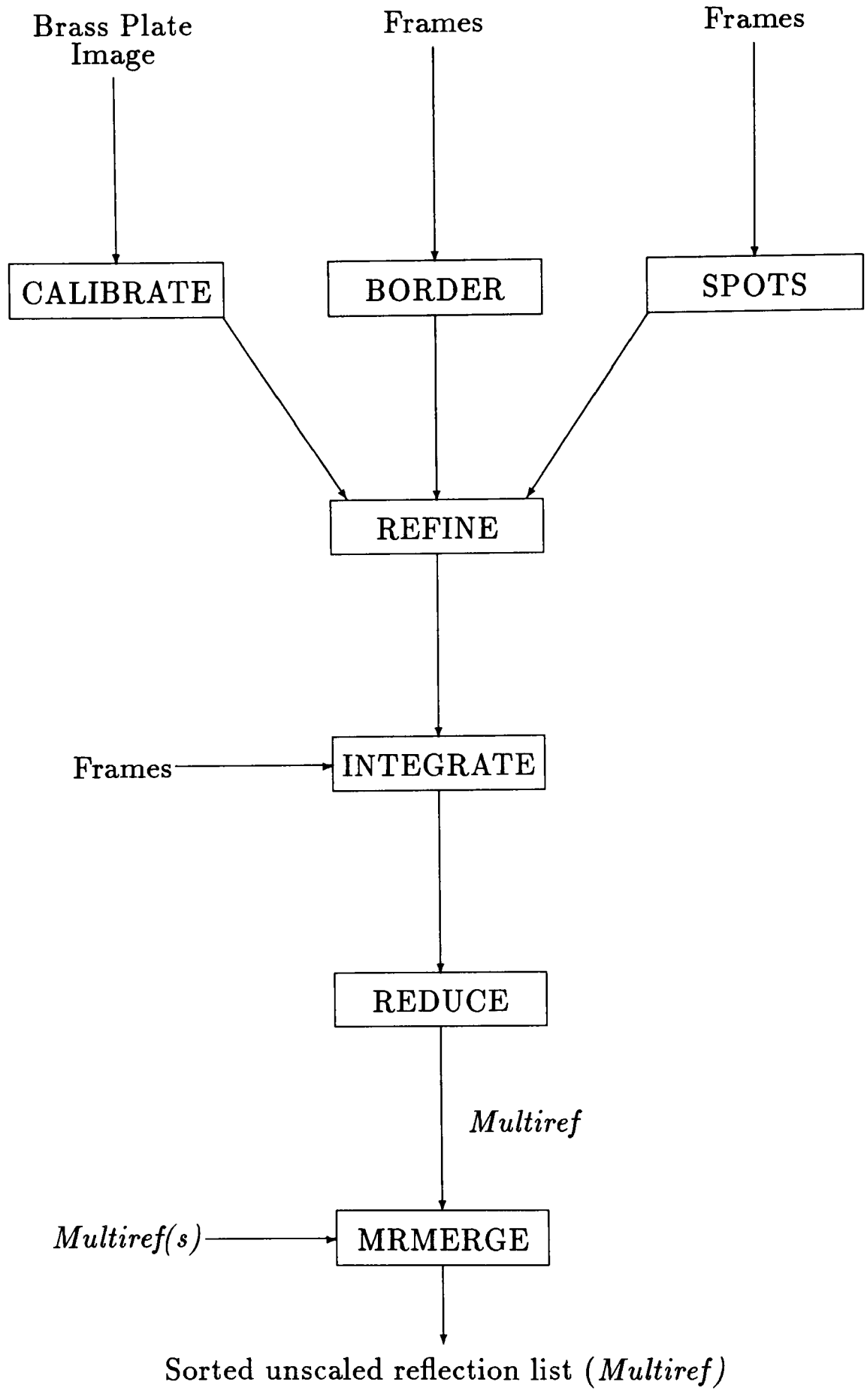


Figure 2: Data Processing Protocol — Production of unscaled intensities

the software which will not be described here, and whose only major difference was the manual definition of two-dimensional spot masks for use in the integration step.

4.2.1 Data Processing Protocol

A flow diagram for the data processing protocol used to produce unscaled integrated intensities is given in Figure 2. Output from three independent programs is required before the cell dimensions and the orientation matrix can be determined for each crystal.

CALIBRATE determines the distortion of the diffraction pattern, arising from the projection of ionisation events occurring near the concave beryllium/gas interface onto the planar recording wires, from the brass plate image. The program produces a lookup table to interconvert between pixels and the laboratory coordinate system. BORDER determines the 'inactive' pixels in the pixel array by calculating the mean count for each pixel from a range of frames (typically 15–40). If the mean count for a pixel is less than 20% of the mean count for the whole array the pixel is flagged as inactive. The circular front aperture of the detector introduces inactive regions at the corners of the pixel array, and the presence of the backstop on the image will introduce a further inactive region. SPOTS produces a list of bright spots (where the 'brightness' cutoff is defined as a multiple of σ for counts over the whole detector), from a contiguous range of frames. The centroids of the spots in (X, Y, ω) are output in pixel and frame units, and also a set of spot masks is produced from each of 12 regions on the detector surface to act as first approximations to the spot profiles for use in INTEGRATE.

The output from these three programs are brought together in REFINE which obtains the cell parameters and misorientation angles from an auto-indexing algorithm (which determines the shortest three representative non-coplanar difference

vectors in the diffraction pattern), and refines both these and the detector parameters for crystal-to-film distance, 2θ angle, the direct beam position, the 'tilt' (rotation from expected orientation of the detector about an axis normal to it's face), and the beam divergence (including crystal mosaicity) in the X and Y directions. REFINE seeks to minimise the residual:

$$R = \sqrt{\sum_j (\omega_j^{obs} - \omega_j^{calc})^2}$$

When the refinement of the crystal and detector parameters is near convergence it is possible to "remap" the brass plate calibration on the basis of the refined data. Each fiducial point j from the brass plate image is moved in (X, Y) by an amount $(\langle \delta x \rangle_j, \langle \delta y \rangle_j)$ given by:

$$\langle \delta x \rangle_j = \sum_i w(i) * (X_{obs}(i) - X_{calc}(i)) / \sum_i w(i)$$

where X_{obs} and X_{calc} are the observed and calculated positions of a number of spots in the vicinity of the fiducial point. The weighting parameter $w(i)$ decreases as the spot becomes more distant from the fiducial point. This new set of fiducial points forms the basis of the formation of a new lookup table.

Once the discrepancy between observed and calculated spot centroids has been minimised by variation of the above parameters, this description of detector configuration and crystal orientation is passed to INTEGRATE which uses these values to predict spot positions over a number of frames. INTEGRATE has the option of further refining these parameters against the centroids of spots observed during the course of integration. Spots intensities are estimated by both simple summation and by learnt profile fitting methods after the intensity has been corrected for background scatter. The background count for each pixel is estimated as the mean

of all counts for that pixel from recent frames for which the pixel is not part of a reflection. For any pixel (X, Y) in frame f which is not part of a reflection the updated background count $B(X, Y, f)$ is given by:

$$B(X, Y, f) = (15/16) * B(X, Y, f - 1) + (1/16) * Z(X, Y, f)$$

where $Z(X, Y, f)$ is the count for that pixel in the frame. Reflection intensities are corrected for Lorentz and polarisation effects (defined below), and their centroids are output in the (X, Y, ω) coordinate system.

The recorded intensities are corrected for the polarisation of the incident beam (itself polarised by the graphite monochromator) by the diffraction process, and for the amount of time each reflection spends in the diffracting position (Lorentz correction). The form of the Lorentz correction depends on the geometry of the diffraction event. If the incident and diffracted lie in a plane perpendicular to the axis of crystal rotation the Lorentz correction is a factor of $\sin 2\theta$ applied to the recorded intensities.

REDUCE performs the indexing of observations output from INTEGRATE and groups duplicate observations of the same reflection together. REDUCE excludes observations if any part of them lies on an inactive area of the detector (defined by BORDER) or whose intensity estimates from the two methods used in INTEGRATE differ too greatly. Observations are rejected if:

$$|I(profile) - I(summed)| > 10 * \sqrt{\sigma^2(profile) + \sigma^2(summed)}$$

The resultant *Multiref* file is a self-contained file which does not require output from any previous program for subsequent scaling procedures. *Multiref* files from more than one run, and between different crystals, are merged together using MRMERGE.

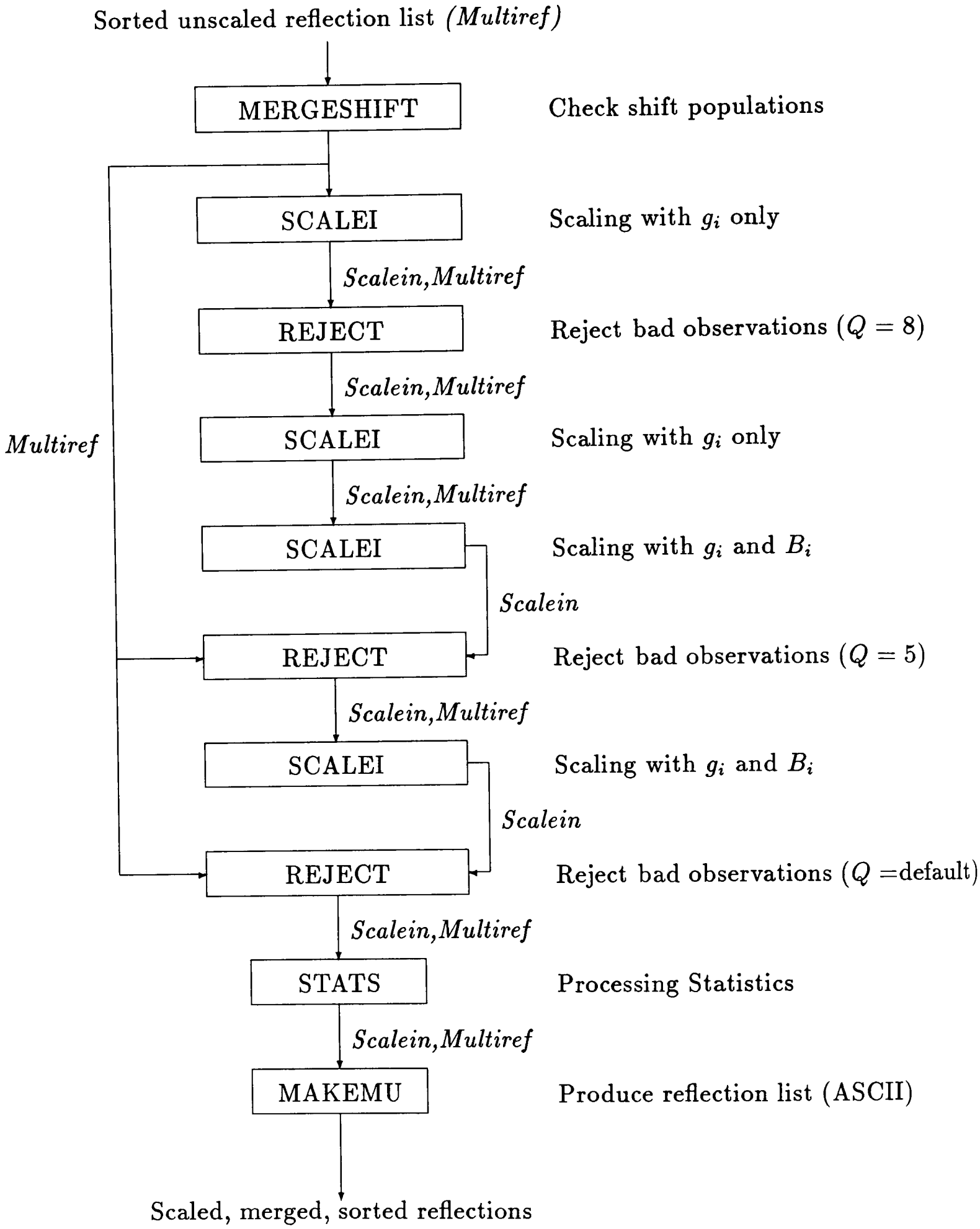


Figure 3: Data Processing Protocol — Scaling of data

The currently used protocol for scaling and merging observations is shown in Figure 3. REDUCEd and MRMERGEd observations are assigned scale-shifts, where a contiguous set of frames is split up into $\approx 5°$ bins and each bin is further divided into separate shifts for the top and bottom of the detector. MERGESHIFT checks that each shift is sufficiently populated by observations so that the scaling is well-behaved. Any under-populated shifts are merged into neighbouring shifts. SCALEI minimises the residual:

$$\sum_{ij}[w_{ij}(\langle I\rangle_j - I_{corr,ij})]^2$$

between an observation ($I_{corr,ij}$) and the mean intensity calculated from all the observations of a reflection ($\langle I\rangle_j$), by the determination of up to three parameters (g_i, A_i, B_i) for each shift:

$$I_{raw} = I_{corr} * (g_i + A_i * stl + B_i * stl^2)$$

where $stl = \sin\theta/\lambda$ and I_{raw} is the observed intensity. The weighting parameter w_{ij} is $[\sigma(I_{ij})]^{-1}$ if there is an observation of the reflection in the ith shift, and zero otherwise. The value of g_i is set to unity for the first shift. The scaling protocol does not use the A_i parameter since in all cases the data have insufficient redundancy to justify its use (Howard, 1988). REJECT deletes observations of multiply-observed reflections that differ by more than $Q * \sigma(I)$ from the mean value of I determined from the observations, where Q is a user-supplied value. STATS performs statistical analysis of the scaled data. MAKEMU outputs the observations as merged intensities or structure factor amplitudes. In the latter case reflections which for which $I \leq 0$ have $F = 0$ and $\sigma(F) = 9999$ in the output file.

In the protocol, the first scaling run used only the g_i parameter for scaling, and only the worst outliers in the observation list were rejected ($Q = 8$). An improved

estimate for the scaling parameters was then obtained by scaling first with one parameter (g_i), and then with two parameters (g_i, B_i). (Derewenda and Helliwell (1989) used the (g_i, A_i) parameters when scaling with two parameters). This new scaling was then applied to the original *Multiref* file and a more severe cutoff applied ($Q = 5$) in REJECT. The scaling was redetermined with two parameters and the scaling applied once again to the original *Multiref* file. The final rejection was performed using the program's default value for Q (3.5–4.5). The resultant file was converted to an ASCII list of structure factor amplitudes with the dependence on $\sin\theta/\lambda$ maintained.

4.2.2 Space Group Determination

The auto-indexing algorithm in REFINE gave the primitive form of the space group for each crystal. For the Gloop2 crystal forms, the triclinic and monoclinic primitive cells were immediately apparent. The $P2_1$ crystal forms were initially processed as $P1$ and the presence of the crystallographic two-fold screw axes determined by inspection of the reflection list. The data were then reprocessed in the correct space group.

Auto-indexing for the Gloop4 crystals gave a $P1$ cell with dimensions a = 39Å, b = 72Å, c = 84Å, $\alpha = 91°$, $\beta = 106°$, and $\gamma = 76°$ and the data were processed as such. It was subsequently pointed out that this cell could be reindexed on the basis of a centered monoclinic cell (A. Howard, pers. commun.). The program LATTICE (Taylor, 1989a) was used to view the $P1$ reciprocal lattice weighted by the values of the structure factors within FRODO (Jones, 1985). The crystallographic two-fold was readily apparent despite the poor quality of the data, and the new cell was assigned as $C2$ with approximate cell dimensions a = 100Å, b = 40Å, c = 140Å, $\beta = 130°$. It was then possible to reprocess the data on the basis of the new cell,

Fab	Space Group	Cell Dimensions						Dmax (Å)	Code
		a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)		
Gloop1	$P1$	55.3	55.1	70.9	87.0	96.1	75.2	2.6	G1P1
Gloop2	$P1$	50.0	84.5	69.5	106.6	104.7	95.1	2.6	G2P1A
Gloop2	$P1$	49.7	83.6	68.8	106.5	104.6	94.9		G2P1B
Gloop2	$P1$	49.8	83.6	69.1	106.2	104.6	94.9		G2P1C
Gloop2	$P1$	49.1	82.8	68.4	106.9	104.3	94.3		G2P1D
Gloop2	$P1$	49.8	83.6	69.1	106.2	104.6	94.9		G2P1E
Gloop2	$P1$	49.8	83.5	69.3	106.6	104.5	95.1		G2P1F
Gloop2	$P2_1$ 'A'	52.0	75.1	61.0	90.0	105.5	90.0	2.9	G2P2$_1$G
Gloop2	$P2_1$ 'A'	50.2	74.4	60.6	90.0	104.0	90.0		G2P2$_1$H
Gloop2	$P2_1$ 'B'	60.8	67.7	68.3	90.0	112.4	90.0	3.1	G2P2$_1$I
Gloop4	$C2$	104.4	39.3	138.3	90.0	129.0	90.0	3.0	G4C2

Dmax estimated from processing statistics.

Table 2: Space Groups and Cell Dimensions for the Datasets

Fab	Space Group	Unit Cell Volume (V) (Å^3)	Fabs per Cell (N)	V/N (Å^3)	Reference
McPC603	$P6_3$	1388117	6	231352	Satow *et al.* (1986)
Kol	$P2_12_12_1$	435035	4	108759	Matsushima *et al.* (1978)
J539	$P2_12_12_1$	525060	4	131265	Navia *et al.* (1979)
New	$C2$	510443	4	127610	Avey *et al.* (1968)
Extrapolations to Gloop Fabs:					
Gloop1	$P1$	206599	2	103299	
Gloop1	$C2$	415496	4	103874	(alternative cell)
Gloop2	$P1$	257264	2	128362	
Gloop2	$P2_1$ 'A'	229553	2	114776	
Gloop2	$P2_1$ 'B'	262769	2	131384	
Gloop4	$C2$	440979	4	110244	

Table 3: Packing of Fabs in the Unit Cells

limiting the auto-indexing search to cell dimensions near to this estimate. Refined cell dimensions for this $C2$ cell are given in Table 2.

The Gloop1 data were more troublesome to process than the Gloop4 and it would appear that these data were more severely affected than the Gloop4 data by the $\Delta\omega$ error. The auto-indexing solution gave a $P1$ cell with a = 55.3Å, b = 55.1Å, c = 70.9Å, $\alpha = 87.0°$, $\beta = 96.1°$ and $\gamma = 75.2°$. The program LATTICE was used to investigate the possibility of indexing the cell on the basis of a centered lattice. The structure factor amplitudes were sufficiently in error that there was no clear indication of a two-fold axis, although it was apparent that the lattice could be indexed as pseudo-$C2$ with a cell of a = 67Å, b = 88Å, c = 71Å, $\beta = 97°$. All data processing statistics are based on the $P1$ cell assignment.

The limited angular range of data used to produce spot centroids for REFINE was not the best choice for determination of accurate cell dimensions, particularly in the $P1$ cells. In order to increase the reliability of the cell dimensions obtained from REFINE, bright spots from small angular ranges of data separated by $\approx 90°$ from each other were merged (program CENMERGE) and the resultant spot list used to generate the 'best' cell parameters.

A list of the space groups for the crystal forms of Gloop1 (as a $P1$ cell), Gloop2 and Gloop4 are given in Table 2.

The assignment of the expected number of Fabs per asymmetric unit for all the crystal forms was straightforward. Table 3 lists the cell volumes and number of Fabs per cell for the four uncomplexed Fabs in the Protein Databank. The typical packing densities of the Fab crystals can be evaluated by dividing the unit cell volume by the number of Fabs in the cell. The lower limit of $\approx 110000Å^3$ gives an estimate of the maximum packing density for a Fab in a unit cell. The

volume/Fab for McPC603 is particularly high (231352Å^3) due to the packing in the unit cell, Padlan *et al.* (1983) estimated that there was ≈70% solvent content in the McPC603 unit cell.

The Gloop2 $P2_1$ 'A' crystal form necessarily had $2n$ Fabs per cell, where n was an integer, and it was immediately apparent that the only reasonable assignment was 2 Fabs per cell. Using these arguments it was concluded that the $P2_1$ and $C2$ crystal forms of Gloop1, Gloop2 and Gloop4 each contained one Fab per asymmetric unit. The Gloop2 $P1$ cell could contain either one Fab/cell which was loosely packed, or two Fabs/cell. The latter estimate was favoured on the grounds that the $P1$ crystals are morphologically identical to the $P2_1$ 'A' form crystals and have a similar unit cell shape.

On the basis of these estimates, the values V_M and solvent content (Matthews, 1968) are as given below, assuming the Fabs have a molecular weight of 50kD and the partial specific volume of the protein is 0.74cm^3/g:

Gloop1 $P1$, Z=2	$V_M = 2.07$Å^3/dalton	41% solvent
Gloop2 $P1$, Z=2	$V_M = 2.57$Å^3/dalton	52% solvent
Gloop2 $P2_1$ 'A', Z=2	$V_M = 2.30$Å^3/dalton	46% solvent
Gloop2 $P2_1$ 'B', Z=2	$V_M = 2.63$Å^3/dalton	53% solvent
Gloop4 $C2$, Z=4	$V_M = 2.20$Å^3/dalton	44% solvent

These values lie toward the centre of the observed range of V_M for other protein crystals (Matthews, 1968).

Fab	Space Group	Dmax (Å)	Number of Observations	Number of Unique Refls.	Completeness (%)	$R_{sym}(I)$ (%)	$R_{sym}(F)$ (%)	Rejection Rate (%)	$\langle I/\sigma(I)\rangle$	Code
Gloop1	$P1$	2.6	26817	19448	69.7	15.44	10.25	20.4	15.3	G1P1
Gloop2	$P1$	4.9	7070	3573	66.4	3.54	2.40	2.6	48.0	G2P1A
Gloop2	$P1$	2.9	10745	9000	36.0	4.52	3.15	1.2	19.3	G2P1B
Gloop2	$P1$	3.2	21109	12606	70.1	5.29	3.78	10.1	29.9	G2P1C
Gloop2	$P1$	2.6	38575	20792	64.9	7.31	6.00	9.2	22.9	G2P1D
Gloop2	$P1$	2.7	20570	16489	52.4	4.49	3.20	2.9	25.2	G2P1E
Gloop2	$P1$	2.7	9437	7877	25.5	6.66	4.47	6.3	23.3	G2P1F
Gloop2	$P2_1$ 'A'	3.3	11643	6081	86.3	5.64	4.28	16.5	27.3	G2P2$_1$G
Gloop2	$P2_1$ 'A'	2.9	14018	7263	71.5	7.02	6.66	11.7	25.2	G2P2$_1$H
Gloop2	$P2_1$ 'B'	3.1	7032	4026	43.0	5.93	4.42	17.5	26.7	G2P2$_1$I
Gloop4	$C2$	3.0	14154	7682	83.5	11.64	11.25	1.5	6.0	G4C2
Merging:										
Gloop2	$P1$	2.6	60084	21729	65.7	7.42	5.98	18.1	23.8	G2P1C + D
Gloop2	$P1$	2.6	59145	23474	72.2	7.23	5.95	15.7	22.5	G2P1D + E

Dmax = Maximum resolution of reflections in scaled reflection list.
No. of unique reflections after merging Bijvoets.
$R_{sym}(I) = \sum_i |I - \langle I\rangle| / \sum_i I$, same function for $R_{sym}(F)$
Completeness = (No. of unique reflections)/(No. of reflections theoretically observable ∞–Dmax).
Rejection Rate = (No. of observations rejected)/(No. of observations collected).
Statistics include all reflections for which $F \geq 0$.

Table 4: Data Processing Statistics

4.2.3 Data Processing Results

Processing statistics on the data collected (Table 1) are given in Table 4. The $\Delta\omega$ errors for the Gloop1 and Gloop4 datasets introduced large errors in the integration procedure and the scaling runs were badly behaved. In particular the scaling protocol for rejecting bad duplicate observations of the same reflection threw out the majority of the duplicate observations for Gloop1. The values of R_{sym} (defined in Table 4) are likely to be underestimates of the true error levels in the Gloop1 and Gloop4 datasets due to the paucity of duplicate observations used in scaling and the high rejection rate. These datasets will not be considered further, but at least provided information on the space group and cell dimensions.

The Gloop2 $P1$ datasets suffered from both the low symmetry of the space group, and non-isomorphism between crystals. The isomorphism of the different datasets was checked by merging the *Multiref* files for the datasets together using the scaling protocol outlined above. Non-isomorphism manifested itself as a significantly higher R_{sym} for the combined dataset than for either of the datasets by themselves. In this manner various combinations were tried in merging other datasets with the major G2P1D and G2P1E datasets. In fact G2P1D and G2P1E were apparently isomorphous with each other, as were G2P1C and G2P1D. However merging either G2P1C with G2P1E or all three datasets together indicated non-isomorphism and the overall $R_{sym}(I)$ in the latter case was 10.37%. A possible explanation of this phenomenon is that G2P1D is intermediate in some manner to G2P1C and G2P1E and these two are sufficiently different to be significantly non-isomorphous. G2P1B did not merge successfully with these datasets. The mean fractional isomorphous differences (MFIDs) were evaluated using the program DIFFER (Appendix 2), where the datasets (which had been internally scaled using the procedure given above) were scaled together using a six-parameter anisotropic B

Datasets	G2P1B	G2P1C	G2P1D	G2P1E
G2P1B	—	14.3	15.5	22.6
G2P1C	5446	—	14.1	15.5
G2P1D	7094	11222	—	14.0
G2P1E	7649	8722	12562	—

Mean fractional differences calculated by the program DIFFER (Appendix 2), using 6-parameter anisotropic B factor scaling.

All data with $F > 0$ used for the comparison.
Upper right triangle: MFID (%)
Lower left triangle: Number of reflections in comparison

$$\text{MFID} = \sum_{\mathbf{h}} |F_1(\mathbf{h}) - F_2(\mathbf{h})| / \sum_{\mathbf{h}} |F_1(\mathbf{h})|$$

Table 5: Mean Fractional Isomorphous Difference in Gloop2 $P1$ Datasets

factor and a scale factor (Stuart *et al.*, 1979). These MFIDs are summarised in Table 5. These numbers will contain contributions from non-isomorphism between the data and also from bad observations in the datasets which were not rejected due to the lack of duplicate observations. The differences due to both effects will increase with increasing resolutions since the proportion of duplicate observations decreases with resolution in the datasets collected with non-zero 2θ angles. The MFIDs echo the results from scaling the datasets together within XENGEN, with G2P1C and G2P1E being less isomorphous with each other than they are with the G2P1D data. A more detailed breakdown of processing statistics for G2P1C, D and E is given in Table 6, for the two merged datasets in Table 7.

A significant source of non-isomorphism may have been slight differences in the concentrations and pH of the crystallisation media, since in the case of the G2P1D dataset data from crystals from the *same* hanging drop merged together adequately. To investigate whether it was possible to induce the crystals to be isomorphous, some crystals were soaked overnight in a standard mother liquor before data collection. The deterioration in the soaked crystals relative to their expected quality suggested that Gloop2 was sensitive to this treatment. Since cooling the crystals did not appear to to be useful for the reduction of crystal damage (see above) it may be necessary to use the flash-freezing technique currently being developed for the Oxford area detector (E. F. Garman, unpublished results) in order to collect a more complete dataset from a single crystal without the associated problems of significant radiation damage.

Non-isomorphism was also significant in the $P2_1$ 'A' crystal form. Merging of these two datasets produced a significantly higher R_{sym} than for the individual datasets ($R_{sym}(I) = 16.29\%$, as compared to 5.64% and 7.02% for the unmerged datasets). The MFIDs calculated by the program DIFFER (details as for the $P1$

data) was 23.9% on all $F > 0$ in common between the two datasets. More detailed processing statistics on both these datasets and for the G2P2$_1$I dataset are presented in Table 8.

4.3 Discussion

The area detector has proved a very useful tool in the determination of the space groups of the Fab crystals and the collection of viable native data. It has been shown that a high proportion of the unique data to a moderate resolution ($\approx$ 3.0Å) can be collected off a single crystal, even for the $P1$ datasets where over 180° of data needed to be collected for a complete dataset.

The data collection method employed for all the datasets has been that of random orientation. For the Gloop2 $P1$ space group this is no particular disadvantage unless the data are intended to fill gaps in another dataset. However for the monoclinic datasets this is clearly not the ideal strategy, and data collection efficiency would be improved if the crystallographic two-fold axes ($\mathbf{b}^*$) were placed perpendicular to the vertical rotation axis. The small size of the primitive cells of all the crystal forms studied was helpful in that a small XTDD can be used, and only comparatively small 2θ offsets were required to collect data to the diffraction limit of the crystal. In this way at least some data were being collected both sides of the backstop simultaneously, improving the redundancy of the data.

Derewenda and Helliwell (1989) found that the Xentronics area detector response varied with both time and position. The time variation appeared to be due to the influence of vibrations, and affected pairs of adjacent wires. The spatial variation was approximately $\pm 1\%$ in both vertical and horizontal directions. More significantly it was found that changing the scaling procedure within the XENGEN programs (Howard *et al.*, 1987) resulted in improvements in the data: (1) rejection

of reflections for which σ was significantly less than the mean σ for all observations of the reflection; (2) dividing the 5° bins into left and right halves of the detector rather than top and bottom. The first method corrects for deficiencies in the BORDER program which appeared to leave spots too near to the inactive areas. The second method presumes that the effects of absorption etc. are greater in the horizontal plane than in the vertical one. It is not clear that this would be the case for crystals mounted parallel to the ϕ axis with an offset detector ($2\theta \neq 0$). It does seem, however, that there is room for improvement over the scaling regime provided for in SCALEI.

(a) Dataset G2P1C

Resolution Range (Å)	$R_{sym}(I)$ (%)	$R_{sym}(F)$ (%)	$\langle I/\sigma(I)\rangle$	Completeness (%)
∞–5.9	3.59	2.49	51.5	67.3
5.9–4.7	4.07	2.83	47.1	78.6
4.7–4.1	4.84	3.33	42.7	81.7
4.1–3.7	7.00	4.79	27.6	82.7
3.7–3.4	8.81	6.17	20.8	73.7
3.4–3.2	12.41	9.09	13.4	36.1
Overall:				
∞–3.2	5.29	3.78	29.9	70.1

(b) Dataset G2P1D

Resolution Range (Å)	$R_{sym}(I)$ (%)	$R_{sym}(F)$ (%)	$\langle I/\sigma(I)\rangle$	Completeness (%)
∞–4.8	3.84	2.88	62.4	80.9
4.8–3.8	6.03	4.51	38.8	85.3
3.8–3.3	11.21	8.21	17.3	76.7
3.3–3.0	18.05	13.32	9.0	68.0
3.0–2.8	26.03	18.45	4.6	59.0
2.8–2.6	37.62	28.72	2.9	16.6
Overall:				
∞–2.6	7.42	5.98	23.8	65.7

(c) Dataset G2P1E

Resolution Range (Å)	$R_{sym}(I)$ (%)	$R_{sym}(F)$ (%)	$\langle I/\sigma(I)\rangle$	Completeness (%)
∞–4.8	3.95	2.74	54.7	66.7
4.8–3.8	5.56	4.10	35.3	68.6
3.8–3.4	6.26	6.05	16.4	60.9
3.4–3.1	1.81*	1.17*	9.2	54.9
3.1–2.8	3.40*	2.06*	5.3	45.4
2.8–2.7	8.39*	4.29*	4.0	15.9
Overall:				
∞–2.7	4.49	3.20	25.2	52.4

Notation as in Table 4

Table 6: Details on Datasets — Gloop2 $P1$

* There are only a few duplicate observations in these ranges.

(a) Dataset G2P1C + G2P1D

Resolution Range (Å)	$R_{sym}(I)$ (%)	$R_{sym}(F)$ (%)	$\langle I/\sigma(I)\rangle$	Completeness (%)
∞–4.8	4.10	3.09	69.7	84.6
4.8–3.8	6.25	4.69	46.5	89.8
3.8–3.3	11.88	8.66	21.7	83.9
3.3–3.0	21.86	15.99	8.4	67.9
3.0–2.8	28.76	20.70	4.3	55.3
2.8–2.6	44.18	30.89	2.6	9.7
Overall:				
∞–2.6	7.42	5.98	23.8	65.7

(b) Dataset G2P1D + G2P1E

Resolution Range (Å)	$R_{sym}(I)$ (%)	$R_{sym}(F)$ (%)	$\langle I/\sigma(I)\rangle$	Completeness (%)
∞–4.8	3.93	3.02	71.8	83.3
4.8–3.8	5.88	4.42	44.2	88.5
3.8–3.3	10.11	7.52	19.8	85.4
3.3–3.0	15.51	11.43	10.5	79.4
3.0–2.8	23.30	16.80	5.3	74.1
2.8–2.6	27.89	20.61	3.2	19.9
Overall:				
∞–2.6	7.23	5.95	22.5	72.2

Notation as in Table 4

Table 7: Details on Datasets — Gloop2 $P1$ Merged Datasets

(a) Dataset G2P2$_1$G

Resolution Range (Å)	$R_{sym}(I)$ (%)	$R_{sym}(F)$ (%)	$\langle I/\sigma(I)\rangle$	Completeness (%)
∞–6.0	3.85	2.68	48.3	94.3
6.0–4.8	4.83	3.22	42.7	98.1
4.8–4.2	4.83	3.56	41.1	98.8
4.2–3.8	7.07	5.52	26.4	99.0
3.8–3.5	7.96	6.35	20.3	83.3
3.5–3.3	8.73	6.93	14.6	43.4
Overall:				
∞–3.3	5.64	4.28	27.3	86.3

(b) Dataset G2P2$_1$H

Resolution Range (Å)	$R_{sym}(I)$ (%)	$R_{sym}(F)$ (%)	$\langle I/\sigma(I)\rangle$	Completeness (%)
∞–5.2	3.63	3.00	65.5	88.6
5.2–4.1	5.12	4.02	45.4	90.8
4.1–3.6	9.74	8.02	23.1	82.1
3.6–3.3	18.33	16.64	12.8	73.1
3.3–3.1	34.06	29.63	5.5	66.4
3.1–2.9	39.05	29.87	2.7	27.7
Overall:				
∞–2.9	7.02	6.66	25.2	71.5

(c) Dataset G2P2$_1$I

Resolution Range (Å)	$R_{sym}(I)$ (%)	$R_{sym}(F)$ (%)	$\langle I/\sigma(I)\rangle$	Completeness (%)
∞–5.6	7.69	5.01	48.3	46.6
5.6–4.5	4.10	2.86	49.0	50.9
4.5–3.9	4.61	3.46	32.9	51.7
3.9–3.6	9.48	7.23	16.8	47.7
3.6–3.3	10.93	9.59	10.1	43.8
3.3–3.1	31.40	21.38	6.7	17.1
Overall:				
∞–3.3	5.93	4.42	26.7	43.0

Notation as in Table 4

Table 8: Details on Datasets — Gloop2 $P2_1$ forms 'A' and 'B'

CHAPTER 5

MOLECULAR REPLACEMENT

5.1 Molecular Replacement — Limits and Successes

The essential requirement of the success of rotation and translation function searches is a sufficiently similar model for the undetermined structure. Huber (1985) has shown that eight structures solved by molecular replacement techniques have root mean square (RMS) deviations on $C\alpha$ atoms of 0.5–1.0Å between the models and the refined structures, although the structure of pepsinogen (James and Sielecki, 1986) was solved with a model that had an RMS deviation of 1.68Å to the refined structure (Fujinaga and Read, 1987) in a $C2$ space group. However in high-symmetry space groups, the superimposition of many sets of self-vectors in the native Patterson may require a structure that is more homologous than in the equivalent case in lower symmetry space groups (Reynolds *et al.*, 1985).

In cases where structural changes are expected, or where there is more than one molecule in the asymmetric unit, it may be necessary to resort to models which correspond to less than the whole asymmetric unit. It has been shown that it is possible to solve structures using models that are $\approx \frac{1}{4}$ of the protein scattering matter in the asymmetric unit (Barford and Johnson, 1989; Fitzgerald, 1988b) although there has been more success when using models that are $\approx \frac{1}{2}$ of the asymmetric unit (e.g. Fitzgerald and Love, 1979; Derewenda *et al.*, 1981; Evans *et al.*, 1985).

The principal variables in the rotation function calculation are the choices of resolution shell and integration sphere radius. In the case of the Fast Rotation

Function (Crowther, 1972) these two factors are coupled according to the maximum order of Bessel functions supplied with the program. There would appear to be no consensus on the resolution range requirements: both narrow ranges (Driessen and White, 1985) and wider ranges (Schierbeek *et al.*, 1985; Dijkstra *et al.*, 1982; Derewenda *et al.*, 1981) have been used. Blow (1985) advised against using data with a resolution lower than 8Å. The optimum integration radius is a function of the shape of the molecule and the crystal packing in the individual case — loosely packed crystals may permit a larger integration radius than for the same molecule in a closely packed system. Generally the integration radii lie in the range 50–80% of the mean molecular dimension (Blow, 1985; Cygler and Anderson, 1988a). In the case of despentapeptide insulin the requirement for a spherical integration volume led to a great dependence of the rotation function on the integration sphere radius (Ru-Chang *et al.*, 1983).

Rotation function calculations may be further manipulated in a number of ways. The use of normalised structure factor amplitudes has a beneficial effect by putting the model and observed structure factors onto the same scale and by 'sharpening' the structure factors by increasing the contribution from high resolution reflections. Since the rotation function varies as the fourth power of $F(\mathbf{h})$ the inclusion of strong reflections is critical, and the weakest reflections may be omitted (Schierbeek *et al.*, 1985). This method is often used with the Rossmann and Blow rotation function to decrease the execution time (Tollin and Rossmann, 1966). Origin removal (Lattman and Love, 1970) is necessary where a unit cell dimension is sufficiently small that a second origin peak would be included within the integration sphere, producing spurious correlations. Origin removal may also be useful in the general case where the 'ripples' in the Patterson function from the origin peak due to series termination errors may affect rotation function results.

Translation functions are dependent on the accurate orientation of the model prior to the search (Dodson, 1985). The required angular accuracy is a function of the size of the molecule and the maximum resolution used in the searches. An analysis of the performance of the translation functions of Crowther and Blow (1967), Harada *et al.* (1981), crystallographic residual and correlation coefficient searches by Fujinaga and Read (1987) did not give a clear indication of the best translation function although the residual search performed the worst. However Tickle (1985) reported that the TO/O function (Harada *et al.*, 1981; Chapter 2) performed better than the Crowther and BlowT_2 function (Crowther and Blow, 1967; Chapter 2), and found that normalised structure factors scaled in shells were beneficial to the performance of these functions.

Four Fab structures have been solved by the molecular replacement method: HED10 (Cygler *et al.*, 1987) and JEL42 (Prasad *et al.*, 1988) as uncomplexed Fabs; HyHEL-5 (Sheriff *et al.*, 1987) and HyHEL-10 (Davies *et al.*, 1988) as Fab:antigen complexes. These results demonstrated the potential of the molecular replacement approach to Fab structures. In particular the behaviour of the HED10 solution to variations in the parameters used in the calculations has been investigated (Cygler and Anderson, 1988a and 1988b) and are outlined in the following two paragraphs.

Two different types of model were used: intact Fab structures; or fragments corresponding to C_L:C_H1 and V_L:V_H domain pairs. The rotation function results using whole Fab proved uninterpretable although there were correct features present. The rotation function from the C_L:C_H1 domain gave a clear solution and even produced the highest peak with only the $C\alpha$ atoms in place. The best results were obtained using a 24Å integration sphere radius and 10–4Å data. The V_L:V_H also produced a clear solution, with the best conditions with 15–4Å data and a 24Å integration sphere radius. Again $C\alpha$ atoms were seen to be sufficient. Narrow shells of data

Program	Functionality
MERLOT:	
PROCOR	Convert coordinates to MERLOT format
CENTOM	Move centre of gravity of coordinates to origin
STRFAC	Calculate structure factors
PROREF	Convert reflection list to MERLOT format
HARMCO	Calculate spherical harmonic coefficients
CROSUM	Crowther rotation function
WRITTF	Prepare "continuous" transforms for LATSUM, TRNSUM
LATSUM	Lattman rotation function
TRNSUM	Translation function
MAPOUT	Map analysis
BRUTE	Crystallographic residual and correlation coefficient search
SCATT	Prepare structure factors for BRUTE

Table 1: Functionalities of Molecular Replacement Programs

produced the correct result in some cases, but in general broad resolution ranges performed better. The omission of CDRs from the V_L:V_H model did not improve the rotation function results significantly. Omission or inclusion of weak reflections did not appear to make much difference to the rotation function results.

The three translation function algorithms investigated were the correlation coefficient search of Fujinaga and Read (1987), the T and T_1 translation functions of Crowther and Blow (1967) and a density correlation procedure which used an electron density map phased from a correctly oriented fragment in a $P1$ symmetry cell of the correct size (Cygler and Anderson, 1988b). In all three cases the solutions of the translations of the C_L:C_H1 and V_L:V_H domain pairs were clear. The phases from a correctly oriented fragment in the $P1$ cell were also shown to be capable of indicating the solution of a heavy-atom derivative.

5.2 Programs Used

The MERLOT program suite (Fitzgerald, 1988a) was used for rotation function and translation function calculations and the program BRUTE (Fujinaga and Read, 1987) used for crystallographic residual and correlation coefficient translation searches. A maximum of 30 Bessel orders are available for the Crowther (1972) Fast Rotation Function in MERLOT.

In subsequent discussions the programs used for each calculation are indicated within square brackets: [...]. Table 1 provides a list of the functions of these programs (in the case of MERLOT these are really subroutines). The program SCATT, used for the preparation of native data for BRUTE was supplied as part of the PROLSQ protein refinement programs (Appendix 2). The MERLOT subroutine MAPOUT was used to analyse all maps from MERLOT and BRUTE.

Real space cell axes : $\mathbf{a}, \mathbf{b}, \mathbf{c}$
Reciprocal cell axes : $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$
Cartesian axes : $\mathbf{x}, \mathbf{y}, \mathbf{z}$

NCODE 1 — $\mathbf{a}$ parallel to $\mathbf{x}$, $\mathbf{c}^*$ parallel to $\mathbf{z}$, $\mathbf{b}$ in $\mathbf{xy}$ plane.
NCODE 3 — $\mathbf{c}$ parallel to $\mathbf{x}$, $\mathbf{b}^*$ parallel to $\mathbf{z}$, $\mathbf{a}$ in $\mathbf{xy}$ plane.
NCODE 5 — $\mathbf{a}^*$ parallel to $\mathbf{x}$, $\mathbf{c}$ parallel to $\mathbf{z}$, $\mathbf{b}^*$ in $\mathbf{xy}$ plane.

NCODE 3 — used by MERLOT (Fitzgerald, 1988b) for **b**-unique space groups.
NCODE 5 — used by MERLOT (Fitzgerald, 1988b) for **c**-unique space groups.

Table 2: Definition of Orthogonalisation Systems

Eulerian Angles

Definition of Crowther (1972) for angles (α, β, γ).

1st rotation : α about **z**
2nd rotation : β about the new **y**
3rd rotation : γ about the new **z**

Rotation matrix (pre-multiplication matrix applied to coordinates of rotating model):

$$\begin{pmatrix} -\sin\alpha * \sin\gamma & \cos\alpha * \sin\gamma & -\sin\beta * \cos\gamma \\ +\cos\alpha * \cos\beta * \cos\gamma & +\sin\alpha * \cos\beta * \cos\gamma & \\ -\sin\alpha * \cos\gamma & \cos\alpha * \cos\gamma & \sin\beta * \sin\gamma \\ -\cos\alpha * \cos\beta * \sin\gamma & -\sin\alpha * \cos\beta * \sin\gamma & \\ \cos\alpha * \sin\beta & \sin\alpha * \sin\beta & \cos\beta \end{pmatrix}$$

Spherical Polar Angles

Definition of Tanaka (1977) for angles (ϕ, ψ, κ).

Rotation by κ about an axis defined by the direction cosines (l, m, n).

Rotation matrix (pre-multiplication matrix applied to coordinates of rotating model):

$$\begin{pmatrix} l^2 + (m^2 + n^2)\cos\kappa & lm(1-\cos\kappa) - n\sin\kappa & nl(1-\cos\kappa) + m\sin\kappa \\ lm(1-\cos\kappa) + n\sin\kappa & m^2 + (l^2 + n^2)\cos\kappa & mn(1-\cos\kappa) - l\sin\kappa \\ nl(1-\cos\kappa) - m\sin\kappa & mn(1-\cos\kappa) + l\sin\kappa & n^2 + (l^2 + m^2)\cos\kappa \end{pmatrix}$$

where the direction cosines l, m, n are given by:

$$\begin{aligned} l &= \sin\psi * \cos\phi \\ m &= \sin\psi * \sin\phi \\ n &= \cos\psi \end{aligned}$$

In both cases positive rotations are anticlockwise when viewed down the rotation axis toward the origin.

Table 3: Definitions of Rotation Matrices and Angles

The definitions of the orthogonalisation and angle systems used throughout the discussions are given in Tables 2 and 3. The eulerian angles (α, β, γ) correspond to those defined by Crowther (1972), and the polar angle convention is that of Tanaka (1977). MERLOT adopts different orthogonalisations for **c**-unique and **b**-unique space groups such that the highest symmetry axis of the space group lies along the cartesian **z** axis. The orthogonalisation systems (NCODES) are defined in Table 2.

Small modifications have been made to MERLOT and BRUTE as implemented in Oxford and are listed in Appendix 2. These affect input/output and orthogonalisation systems only and do not otherwise alter the calculations.

5.3 Choice of Models

The principal problem in constructing a model of the Fab fragment is the flexibility at the elbow bend. The same Fab is capable of adopting different elbow bends in different crystallographic environments (Sheriff *et al.*, 1987; Prasad *et al.*, 1988). Therefore the relative association of V_L:V_H and C_L:C_H1 domain pairs is impossible to predict *a priori*.

Two approaches to this problem are: using whole Fab; or using the V_L:V_H and C_L:C_H1 domain pairs as models. There have been successes with both approaches — the structures of HyHEL-5 (Sheriff *et al.*, 1987) and HED10 (Cygler *et al.*, 1987) were solved using domain pair models, whilst both of the JEL42 Fabs in the unit cell were detected using the intact HED10 Fab as the model (Prasad *et al.*, 1988). Both approaches have attendant difficulties. The use of whole Fab requires either that the unknown structure must have a similar elbow angle to a previously determined structure, or that it is possible to model the elbow bend in order to generate the correct relative orientation of the domains. The domain pair approach involves the use of probes that are at best half the protein scattering matter in the asymmetric

		Residue Range and Chain Label			
File*	Code	V_L	V_H	C_L	C_H1
2HFL	HY5	1L–104L	1H–114H	109L–212L	120H–213H
1MCP	MCP	1L–112L	1H–121H	117L–220L	128H–222H
1FBJ	J539	1L–105L	1H–117H	110L–213L	122H–218H
1FB4	KOL	1L–107L	1H–116H	113L–214L	123H–221H
3FAB	NEW	1L–107L	1H–116H	114L–214L	121H–220H

* Name of file in Protein Databank (Bernstein *et al.*, 1977).
Chain labels and residue numbers as given in the Protein Databank file.

Table 4: Definitions of domain pair models

unit, so that the signal/noise ratio in the rotation function may be degraded relative to that of the intact Fab model.

The domain pair approach was chosen for this study because of the ease of using the different Fab structures available from the Protein Databank depositions (Bernstein *et al.*, 1977). Domain pair models were generated after inspection of the Fab structures, where the domain boundaries were arbitrarily chosen so as to exclude the peptide chain in the vicinity of the elbow bend. These domain definitions are given in Table 4. The CDRs were included in the V_L:V_H models, since the presence of CDRs had been seen to make little difference to rotation function results (Cygler and Anderson, 1988a). The models shall be referred to by their appropriate three or four letter code, as given in Table 4, in subsequent discussions (HY5, MCP, KOL, NEW, J539).

The variation in elbow bend between different Fabs can be modelled by a pseudo rotation axis at the interface of the V_L:V_H and C_L:C_H1 domain pairs. If this axis is placed parallel to the **z** axis of the coordinate system then peaks due to the V_L:V_H and C_L:C_H1 models corresponding to the *same* Fab should have approximately the same values of α and β, with the value of γ differing if the elbow angle is different from that of the standard model (Sheriff *et al.*, 1987). It may be necessary to consider symmetry related values of (α, β, γ) to obtain this relation if γ lies near to the limits of the angular range sampled.

To achieve this orientation the HyHEL-5 Fab was manipulated manually using FRODO (Jones, 1985), to align the elbow bend axis approximately along the **z** axis, and the other domains were superimposed on the corresponding HyHEL-5 domains using the program SHP (Appendix 2).

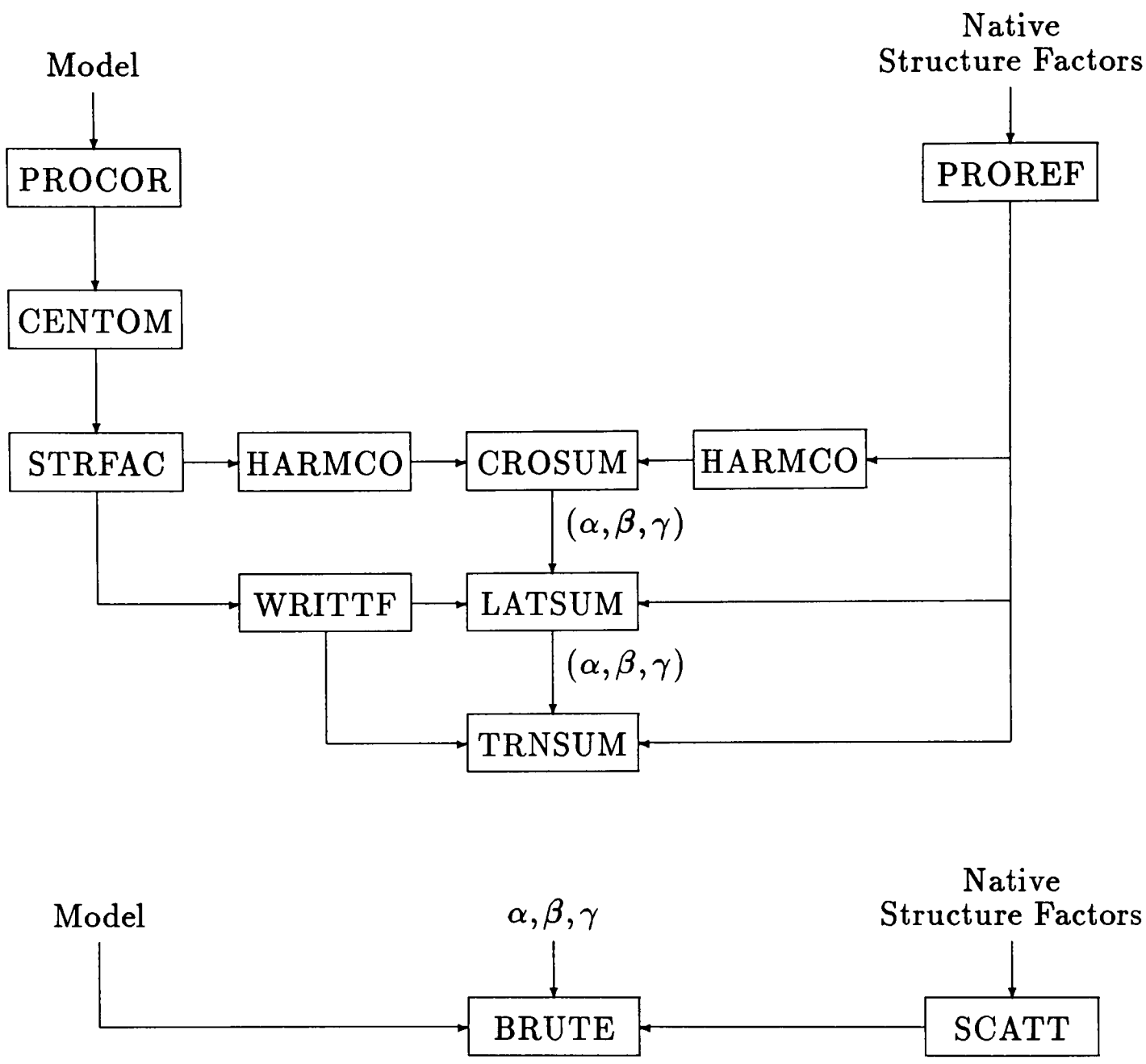


Figure 1: Molecular Replacement Protocol

5.4 Solution of Gloop2 $P2_1$ 'A' Form

A flow diagram for the general molecular replacement protocol used is given in Figure 1. Structure factors were generated for each model centered in a $P1$ cell whose dimensions were sufficient to exclude cross vectors at the maximum integration radius to be used [PROCOR, CENTOM, STRFAC]. Harmonic coefficients were generated for a chosen set of parameters from calculated and observed structure factors [PROREF, HARMCO]. Rotation functions using the method of Crowther (1972) were performed and the maps analysed [CROSUM]. Structure factors for the model were generated in a large cell [STRFAC] corresponding to $\approx$ 3 times the maximum molecular dimension, and 'continuous' transforms were generated [WRITTF]. The initial eulerian angles were refined, by successively finer local grid searches, to the nearest degree using the rotation function of Lattman and Love (1970) [LATSUM]. These angles were then used as input to the translation function [TRNSUM] and crystallographic residual search [BRUTE]. Final Fab models were assembled using BRUTE by initial placement of the C_L:C_H1 model and searching for the placement of the V_L:V_H model over the whole unit cell.

5.4.1 Rotation and Translation Functions

The dataset used for the structure solution was G2P2_1F (see Chapter 4), being preferred over the G2P2_1G dataset on the basis of greater completeness of the data at the resolutions used in the study (up to 4Å).

Structure factors were generated for models centered at the origin of a $P1$ cell with dimensions $a = b = c = 90$Å, $\alpha = \beta = \gamma = 90°$ with an overall B factor of 17Å^2. The cell dimensions corresponded to the largest vector in the domain pair models ($\approx$60Å) plus the maximum integration sphere radius (27Å). Harmonic coefficients were generated for resolution shells of 15–4Å, 10–4Å, 8–4Å and 6–4Å

(a) Effects of resolution range and integration sphere for best model

	Sphere		
Shell†	20Å	24Å	27Å
15-4Å	6.04	6.61	6.27
	1.61	1.64	1.54
10-4Å	7.42	8.04	7.09
	1.76	1.77	1.70
8-4Å	7.29	8.09	7.09
	1.76	1.79	1.54
6-4Å	6.92	7.74	6.26
	1.65	1.74	1.41

Model: C_L:C_H1 from HY5
Upper value: RMS of correct peak
Lower value: Signal/Noise* of correct peak
Optimum integration sphere/resolution shell: 24Å/8–4Å
† Actual values of Dmax: 4.0022Å for integration spheres of 20Å and 24Å.
4.5000Å for integration sphere of 27Å.

(b) Different models at optimum conditions

Model	RMS of Correct Peak	Signal/Noise* of Correct Peak
HY5	8.09	1.79
MCP	5.34	1.24
J539	5.00	1.19
KOL	5.52	1.28
NEW	4.36	1.10

Integration sphere/resolution shell: 24Å/8–4Å
* Signal/Noise = (Height of correct peak)/(Height of top noise peak)
All rotation functions calculated with origin removal
For definition of model labels see Table 2

Table 5: Rotation Function Studies on Gloop2 $P2_1$'A' form — C_L:C_H1 model

(a) Effects of resolution range and integration sphere for best model

	Sphere		
Shell†	20Å	24Å	27Å
15-4Å	4.64	5.37	5.25
	0.93	1.13	1.26
10-4Å	5.48	5.98	5.43
	0.97	1.09	1.09
8-4Å	5.27	5.78	5.05
	0.97	1.09	1.08
6-4Å	5.27	5.65	4.85
	0.98	1.05	1.05

Model: V_L:V_H from MCP
Upper value: RMS of correct peak
Lower value: Signal/Noise* of correct peak
Optimum integration sphere/resolution shell: 27Å/15–4Å
† Actual values of Dmax: 4.0022Å for integration spheres of 20Å and 24Å.
4.5000Å for integration sphere of 27Å.

(b) Different models at optimum conditions

Model	RMS of Correct Peak	Signal/Noise* of Correct Peak
NEW	4.36	1.10
HY5	4.10	1.01
MCP	5.25	1.26
J539	4.04	1.02‡
KOL	3.52	0.91
NEW	3.86	1.03‡

Integration sphere/resolution shell: 27Å/15–4Å
* Signal/Noise = (Height of correct peak)/(Height of top noise peak)
‡ (α, β, γ) differ significantly from correct values.
All rotation functions calculated with origin removal.
For definition of model labels see Table 2

Table 6: Rotation Function Studies on Gloop2 $P2_1$'A' form — V_L:V_H model

at integration sphere radii of 20Å, 24Å and 27Å (in the case of the 27Å integration sphere the maximum resolution available is 4.5Å since only 30 Bessel function orders are supplied in MERLOT) with both sets of structure factors modified by origin removal (Lattman and Love, 1970). No cutoffs on intensity value or $I/\sigma(I)$ were applied to either calculated or observed data, in accordance with the procedure for origin removal (Lattman, 1985), although all $F = 0$ were omitted from the data used. Cross rotation functions were performed for each model at each resolution shell/integration sphere combination over the asymmetric unit (α = 0–90° in 2.5° steps, β = 0–180° in 5.0° steps, γ = 0–180° in 5.0° steps).

The best results, as judged by the peak height expressed as multiples of the root mean square (RMS) of the map and as the ratio of the top peak to the the next highest peak, were at 24Å/8–4Å (integration sphere radius/resolution range) for the C_L:C_H1 model and 27Å/15–4Å for the V_L:V_H model. The C_L:C_H1 model showed the best RMS and best signal/noise ratio under the same set of conditions and the choice of the optimum set of parameters was obvious. The V_L:V_H model showed the best RMS at 24Å/10–4Å and the best signal/noise ratio at 27Å/15–4Å. The latter conditions were chosen as the signal/noise ratio indicated better discrimination between the top peak and the next noise peak. The best models were the HY5 C_L:C_H1 and the MCP V_L:V_H. Tables 5 and 6 give results for these models upon variation of resolution shell and sphere radius, and for the panel of models at the optimum parameters. Over the range of resolution shells and sphere radii examined the correct HY5 C_L:C_H1 peak was always the highest peak, and in the cases where the correct MCP V_L:V_H peak was not the top peak it was always the next highest, with the top peak then corresponding to the mapping of $V_L \Longleftrightarrow V_H$.

The results of the cross rotation functions were unambiguous: the peaks were well separated from the noise level and had similar α and β angles, indicating that

the peaks from the two models corresponded to the same Fab (Table 7). The similarity of γ in the two cases suggested that the elbow angle of this Fab was very similar to that of HyHEL-5. On the basis of the rotation function results the C_L:C_H1 model was considerably superior to that of the V_L:V_H model. The angles (α, β, γ) were refined using LATSUM to the values given in Table 7, using 8–4Å data for C_L:C_H1 models and 15–4Å data for V_L:V_H models (as used with the Crowther rotation functions).

For each domain pair model the translations in the **a.c** plane were determinable, with the translation along **b** arbitrary. Once one domain pair model was fixed along **b**, then the relative position of the other model in the **b** direction also became determinable. Two methods were used to determine the position of the models in the **a.b** plane: the T_1 translation function of Crowther and Blow (1967) using data in the 10–4Å range and an 0.02 fractional (translation along unit cell axes) grid [TRNSUM]; and the crystallographic residual and correlation coefficient search using data in the 5–4Å range (as recommended in the program write-up) and a 1.0Å grid [BRUTE]. In both cases the translational asymmetric unit was 0–0.5 along **a** and **c** (fractional units) for individual models. Once the position of the C_L:C_H1 domain pair was determined in **a** and **c** the translation in **b** was arbitrarily set to 0. A search over the whole of the unit cell using the crystallographic residual search determined the relative translation of the V_L:V_H pair with respect to the C_L:C_H1 [BRUTE]. This result also agreed with the translations along **a** and **c** determined from the search in the **a.c** plane for the V_L:V_H model.

A summary of the molecular replacement solution of this crystal form of Gloop2 is given in Table 7. The number of reflections used in the calculations are given in Table 10.

(a) Cross Rotation Function Results

Model	Crowther					Lattman		
	α (°)	β (°)	γ (°)	RMS	S/N	α (°)	β (°)	γ (°)
C_L:C_H1	135.0	50.0	355.0	8.04	1.79	134.0	49.0	354.0
V_L:V_H	130.0	50.0	355.0	5.25	1.26	127.0	52.0	354.0

(b) T_1 Translation Function Results

Model	Eulerian Angles			Fractional Translations				
	α (°)	β (°)	γ (°)	FA	FB	FC	RMS	S/N
C_L:C_H1	134.0	49.0	354.0	0.16	0.00	0.46	5.69	> 1.54
V_L:V_H	127.0	52.0	354.0	0.16	0.00	0.01	4.06	1.42

(c) Crystallographic Residual and Correlation Coefficient Searches

Model	Eulerian Angles			Crystallographic Residual						Correlation Coefficient					
	α (°)	β (°)	γ (°)	FA	FB	FC	Residual (%)	RMS	S/N	FA	FB	FC	C. Coeff. (%)	RMS	S/N
C_L:C_H1	134.0	49.0	354.0	0.15	0.00	0.46	49.3	5.77	1.03	0.15	0.00	0.46	33.3	5.45	1.27
V_L:V_H	127.0	52.0	354.0	0.15	0.00	0.02	50.8	3.32	1.01	0.15	0.00	0.00	23.0	3.58	1.16
Fixed: C_L:C_H1 at FA = 0.15, FB = 0.00, FC = 0.46															
V_L:V_H	127.0	52.0	354.0	0.65	0.05	0.02	44.7	10.16	1.07	0.65	0.05	0.02	43.8	7.57	1.22

(α, β, γ) — Eulerian angles.
FA, FB, FC — fractional translations along unit cell axes.
S/N — Signal/Noise Ratio, as defined in Table 3.

Table 7: Molecular Replacement Solution of Gloop2 $P2_1$ 'A' form

5.5 Solution of Gloop2 P1 Form

The solution of the P1 crystal form structure proceeded in an analogous manner to that of the $P2_1$ 'A' form. Since the P1 crystals were morphologically related to the $P2_1$ 'A' crystals (Chapter 4), the same integration spheres, resolution ranges and models were used as for the $P2_1$ 'A' form on the assumption that the packing was similar in the two crystal forms.

5.5.1 Rotation and Translation Functions

The dataset used for the calculations was the G2P1C dataset (Chapter 4) on the basis of greater completeness over the resolution range employed of the datasets collected at that time (the G2P1D and G2P1E datasets were collected *after* the structure was solved).

Self rotation functions were calculated to ascertain the presence or absence of non-crystallographic symmetry between the two Fabs in the unit cell. The rotation range used was (ϕ = 0–180° in 5.0° steps, ψ = 0–180° in 5.0° steps, κ = 0–180° in 5.0° steps). Optimum conditions were found to be an integration sphere of 24Å and a resolution range of 8–4Å. The only significant peak in the whole map lay on the $\kappa = 180°$ section at $\phi = 125°$, $\psi = 65°$ (next highest peak on the $\kappa = 180°$ section was at 70% of the height of this peak). Without origin removal this feature was at the noise level.

Cross rotation function results for the C_L:C_H1 models gave two peaks, at (α, β, γ) = $(165°, 90°, 65°)$ and $(315°, 80°, 200°)$, which were clearly separated from the noise peaks [CROSUM]. Although the V_L:V_H model gave a much noisier rotation function, the top two peaks — $(310°, 75°, 185°)$ and $(165°, 90°, 45°)$ — showed the required relation in α and β to the C_L:C_H1 peaks [CROSUM]. Both pairs of angles had the angular relation $(\phi, \psi, \kappa) \approx (235°, 65°, 180°)$ [ROTSYM]. MERLOT uses a different

(a) Cross Rotation Function Results

Model	Crowther					Lattman		
	α (°)	β (°)	γ (°)	RMS	S/N	α (°)	β (°)	γ (°)
C_L:C_H1	165.0	90.0	65.0	7.20	1.66	166.0	89.0	63.0
C_L:C_H1	315.0	80.0	200.0	5.98	1.38	313.0	79.0	199.0
V_L:V_H	310.0	75.0	185.0	5.18	1.09	311.0	76.0	182.0
V_L:V_H	165.0	90.0	45.0	4.99	1.05	167.0	87.0	45.0

(b) Crystallographic Residual and Correlation Coefficient Searches

Model	Eulerian Angles			Crystallographic Residual						Correlation Coefficient					
	α (°)	β (°)	γ (°)	FA	FB	FC	Residual (%)	RMS	S/N	FA	FB	FC	C. Coeff. (%)	RMS	S/N
C_L:C_H1	166.0	89.0	63.0	0.00	0.00	0.00	51.7	—	—	0.00	0.00	0.00	17.2	—	—
Fixed: C_L:C_H1 at FA = 0.00, FB = 0.00, FC = 0.00															
C_L:C_H1	313.0	79.0	199.0	0.90	0.39	0.22	48.4	6.86	1.02	0.90	0.39	0.22	33.3	7.65	1.22
Fixed: C_L:C_H1 at (0.00, 0.00, 0.00) and (0.90, 0.39, 0.22)															
V_L:V_H	167.0	87.0	45.0	0.02	0.26	0.65	46.0	9.63	1.07	0.02	0.26	0.65	35.8	9.20	1.26
Fixed: C_L:C_H1 at (0.00, 0.00, 0.00) and (0.90, 0.39, 0.22), V_L:V_H at (0.02, 0.26, 0.65)															
V_L:V_H	311.0	76.0	182.0	0.50	0.74	0.56	43.7	10.26	1.04	0.50	0.74	0.56	43.0	10.12	1.23

Notation as in Table 7

Table 8: Molecular Replacement Solution of Gloop2 $P1$ form

polar angle convention in ROTSYM to that in CROSUM, with the appropriate relation being

$$(180 - \phi, \psi, \kappa)_{ROTSYM} = (\phi, \psi, \kappa)_{CROSUM}$$

(Fitzgerald, 1988b). Thus the relation between the two top peaks in both cross rotation functions corresponded to the top peak in the self-rotation function. The eulerian angles from the Crowther rotation function were then defined to the nearest degree using the Lattman rotation function [LATSUM] with 15–4Å data for the V_L:V_H model and 8–4Å data for the C_L:C_H1 model.

For the first domain pair model the translation in the $P1$ cell is arbitrary. Consequently the first of the two C_L:C_H1 models was centered at the origin. The subsequent models were added in by searching over the whole unit cell with the crystallographic residual search [BRUTE] in the presence of models whose translations were already defined, using data in the range 5–4Å and a 1.0Å grid. The placement of the second C_L:C_H1 domain pair, and then the two V_L:V_H pairs by this procedure was straightforward (Table 8).

A summary of the molecular replacement solution is given in Table 8. The number of reflections used in the calculations are given in Table 10.

5.6 Solution of Gloop2 $P2_1$ 'B' Form

The structure of the $P2_1$ 'B' form of Gloop2 was determined in the same way as the $P2_1$ 'A' form, with the exception that the models used were from the partially refined coordinates of the $P2_1$ 'A' form (cycle 12 model; see Chapter 6).

(a) Cross Rotation Function Results

	Crowther					Lattman		
Model	α (°)	β (°)	γ (°)	RMS	S/N	α (°)	β (°)	γ (°)
C_L:C_H1	92.5	45.0	175.0	6.65	1.67	93.5	47.0	172.0
V_L:V_H	90.0	50.0	190.0	5.29	1.27	90.0	51.0	193.0

(b) T_1 Translation Function Results

	Eulerian Angles			Fractional Translations				
Model	α (°)	β (°)	γ (°)	FA	FB	FC	RMS	S/N
C_L:C_H1	93.5	47.0	172.0	0.26	0.00	0.45	2.83	1.08
V_L:V_H	90.0	51.0	193.0	0.30	0.00	0.48	5.27	> 1.54

(c) Crystallographic Residual and Correlation Coefficient Searches

	Eulerian Angles			Crystallographic Residual						Correlation Coefficient					
Model	α (°)	β (°)	γ (°)	FA	FB	FC	Residual (%)	RMS	S/N	FA	FB	FC	C. Coeff. (%)	RMS	S/N
C_L:C_H1	93.5	47.0	172.0	0.26	0.00	0.45	48.6	4.92	1.02	0.24	0.00	0.44	32.9	5.11	1.24
V_L:V_H	90.0	51.0	193.0	0.31	0.00	0.48	48.7	4.67	1.02	0.31	0.00	0.48	34.9	5.56	1.16
Fixed: C_L:C_H1 at FA = 0.26, FB = 0.00, FC = 0.45															
V_L:V_H	90.0	51.0	193.0	0.29	0.96	0.98	39.0	14.73	1.14	0.30	0.96	0.98	61.7	11.69	1.50

Notation as in Table 7

Table 9: Molecular Replacement Solution of Gloop2 $P2_1$ 'B' form

5.6.1 Rotation and Translation Functions

The dataset used for these calculations was the only one available for this space group, G2P2_1H. All data with $F > 0$ was used in the calculations.

Cross rotation functions used the conditions that had proved successful in the $P1$ and $P2_1$ 'A' form cases above (24Å/8–4Å for C_L:C_H1 and 27Å/15–4Å for V_L:V_H). Searches over the angular range (α = 0–90° in 2.5° steps, β = 0–180° in 5.0° steps, γ = 0–180° in 5.0° steps) showed well-resolved single peaks for both the C_L:C_H1 and V_L:V_H models, which also had similar values of α and β. The incompleteness of the data, which would be expected to degrade the significance of peaks in the rotation function due to distortion of the Patterson function, appeared to have been offset by the expected closeness of the models to the undetermined structure. The eulerian angles for both models were determined to the nearest degree using LATSUM with the same data ranges as used in CROSUM.

Determination of the translation of each model in the **a.b** plane using the T_1 translation function and 10–4Å data with an 0.02 fractional grid gave quite poor results [TRNSUM], probably due to the incompleteness of the data used. However the equivalent crystallographic residual and correlation coefficient searches using 5–4Å data gave very clear results, since these methods were not affected by the incompleteness of the data [BRUTE]. The Fab was assembled using the same method as for the $P2_1$ 'A' structure and exhibited a far better agreement with the observed data than for the other solutions, again indicating the quality of the starting models.

A summary of the molecular replacement solution of this crystal form is given in Table 9. The number of reflections used in the calculations are given in Table 10.

5.7 Discussion

The molecular replacement method has shown itself to be well-suited to the structural solution of the Gloop2 crystal forms. The solutions were aided by two effects: (1) the availability models with high sequence homology; (2) the low symmetry of the space groups.

The C_L:C_H1 from HyHEL-5 — identical with Gloop2 in C_L and with 92% sequence identity in C_H1 — showed itself to be an excellent model. However it is apparent that at the optimum set of parameters for this model *all* the C_L:C_H1 models used would have been sufficient, perhaps reflecting the high degree of structural conservation in C_L:C_H1 domains. In sequence homology there is little to choose between the V_L:V_H from McPC603 and HyHEL-5: the V_L/V_H sequence identities with Gloop2 are 63%/39% for McPC603 and 56%/60% for HyHEl-5 and yet the former is by far the better model in the rotation function. It is probable that differences in domain pairings and CDR conformation are contributory factors to this. As with the study of HED10 (Cygler and Anderson, 1988a and 1988b) the two domain pair models have different optimum parameter requirements reflecting their structural differences, and the degree of similarity with the unknown structure.

The choice of models used in these structure determinations was chronologically determined. The solutions of the $P2_1$ 'A' and $P1$ forms were determined almost simultaneously, the former during a three week visit to the laboratory of Dr. David Davies at the National Institutes of Health, Bethesda, MD., under the guidance of Dr. Steven Sheriff, and the latter upon return to Oxford after modification of the BRUTE program (Chapter 2). Therefore refined coordinates for the Gloop2 structure were not available for the $P1$ structure solution. When the $P2_1$ 'B' form structure came to be solved, partially-refined coordinates with CDRs in place were

available from the $P2_1$ 'A' form refinement, and these were the obvious choice for models.

In all cases the rotation function was searching for agreement of a set of self-vectors from a model with $\approx \frac{1}{4}$ of the self-vectors in the native Patterson. Therefore although the models were only $\approx \frac{1}{4}$ of the asymmetric unit in the $P1$ cell the proportion of the unit cell was the same and the expected signal/noise ratio should have been similar. This was reflected in the C_L:C_H1 domain pair rotation function in $P1$ where there were only two significant peaks, but the V_L:V_H rotation function performed less well. The lesser completeness of the $P1$ data, and possible differences in the local packing in the cell may have been contributory factors. It was gratifying, however, to see how the correlation coefficient search gave unambiguous results in the assembly of the $P1$ cell Fabs with BRUTE.

The molecular replacement solutions were straightforward, and it did not prove necessary to use any more complicated techniques in order to solve these structures, other than the routine application of origin removal. In higher symmetry space groups it may be necessary to use additional techniques to interpret rotation function results. Some approaches might be: deletion of CDRs; 'smearing out' of CDRs using high thermal factors; modelling of framework sequence changes; modelling of CDR backbone conformation; use of normalised structure factors; omission of the weakest reflections.

It is interesting to note that the rotation function using the whole HyHEL-5 structure gave a very clear peak in the cross rotation function with the $P2_1$ 'A' form (results not shown), and it is likely that this model would have proved sufficient for the structural solution. However this model would not have been sufficient for the structural solution of the other two crystal forms, where the elbow angle is different to that of HyHEL-5.

The RMS differences between the refined models of the Gloop2 $P2_1$ 'A' and $P1$ forms (see Chapter 6) and the V_L:V_H and C_L:C_H1 models were evaluated. All $C\alpha$ atoms in the Gloop2 structure (including CDRs) were matched with a $C\alpha$ atom in the model. The RMS fits for the C_L:C_H1 domains lay in the range 1.18–1.28Å and the RMS fits for the V_L:V_H models lay in the range 1.29–1.43Å.

A final check of the three structural solutions by inspection of the crystal packing on a graphics device did not reveal any significant interpenetration, and maintained the high confidence in the structural solutions. The correlation coefficients for the solutions of the $P2_1$ 'A' and $P1$ forms of Gloop2 (44% and 43% respectively on 5–4Å data, see Tables 7 and 8) was comparable to that reported for the HED10 Fab fragment prior to the refinement of the domain positions (45% on 6–4Å data; Cygler *et al.*, 1987).

Program	Resolution (Å)	Number of Reflections			
		$P1$	$P2_1$ 'A'	$P2_1$ 'B'	Model
CROSUM	15–4.5	5039	2562	1460	16223
	8–4.0	6571	3267	1937	20733
LATSUM	15–4.5	5039	2562	1460	132592
	8–4.0	6571	3267	1937	169162
TRNSUM	10–4.0	7028	3382	2030	181170
BRUTE	5–4.0	3742	1812	1479†	N/D

† Resolution range 6–4.0Å

Table 10: Number of Reflections Used in Molecular Replacement Calculations

CHAPTER 6

REFINEMENT OF GLOOP2 STRUCTURES

6.1 Refinement Strategy

The first step in the refinement of the Gloop2 structures was to optimise the orientation and position of the models from the molecular replacement solutions by rigid body refinement of the intact domain pairs. Derewenda (1989) has emphasised the need to minimise the systematic errors introduced into molecular replacement models by errors in placement. A method for optimising the orientation based on conventional least-squares techniques was proposed (Derewenda, 1989). The programs CORELS (Sussman *et al.*, 1977) and X-PLOR (Brünger *et al.*, 1987) were used for rigid body refinement of the Gloop2 models. After optimising the positions of the molecular replacement models, individual domains were refined as rigid bodies, since domain pairing was one source of structural variability (Lascombe *et al.*, 1989) that was not necessarily allowed for by the models used in the molecular replacement solution.

The sequence was then changed to conform with the Gloop2 sequence (Appendix 1) and the sequence-modified models were used as starting points for simulated annealing refinement. In order not to bias the CDR conformations toward those existing in the starting models, the CDRs were deleted from the models before the start of least-squares refinement and subsequently added in on the basis of unattributed electron density.

Extensive use was made of the molecular dynamics technique to refine the Gloop2 structures, in preference to least-squares refinement techniques. There are

two similar approaches to this method: the X-PLOR package (Brünger *et al.*, 1987) uses molecular dynamics simulations at high temperatures (500–9000K) (Taylor, 1989b; Kuriyan *et al.*, 1989; Brünger *et al.*, 1989; Brünger, 1988a) in order to sample conformational space more widely. Increased stereochemical restraints are used to prevent chiral interconversion at $C\alpha$ atoms, buckling of aromatic rings and *cis–trans* peptide bond transitions (except for proline where *cis* peptide bonds are stereochemically acceptable) (Brünger *et al.*, 1989). The MDXREF addition to the GROMOS package (Fujinaga *et al.*, 1989) uses conventional potentials and molecular dynamics simulations at 300K, relying on the use of lower resolution data with gradual inclusion of higher resolution terms to give a greater radius of convergence, rather than the use of raised temperatures.

The X-PLOR package, implemented on the Convex C210 by Dr. G. L. Taylor, was used in the case of Gloop2. The approach taken was to perform many refinement cycles, usually including molecular dynamics runs, with only modest changes to the model between each cycle.

The parameters adopted for the calculation of energies were those suggested by the X-PLOR manual (Brünger, 1988b). All structure factor calculations used the FFT algorithm of Brünger (1989), with structure factors and gradients re-calculated if any atom moved by more than 0.05Å during energy minimisation or 0.2Å during molecular dynamics. Non-bonded interactions were calculated when less than 7.5Å, with $R_{off} = 6.5$Å, $R_{on} = 6.0$Å and the 1–4 interaction parameter $e_{14} = 0.4$ (see Chapter 2 for a description of these parameters). Electrostatic interactions were calculated using a constant dielectric, with the value of the dielectric constant in vacuum. Polar hydrogen atoms were explicitly represented in the structure, and apolar hydrogens were accounted for by using 'extended atoms'. The parameters describing the topology of the structure and the potential were taken from the files

PARAM19x.PRO and TOPH19x.PRO.

6.2 Refinement of Gloop2 $P2_1$ 'A' Form

The refinement of the Gloop2 $P2_1$ 'A' form used the G2P2$_1$G dataset (Chapter 4) with which the structure was solved. The resolution limit of this data was 3.3Å. All data with $F > 0$ were used in the refinements. The refinement of this crystal form was severely under-determined, with the stereochemical restraints acting as essential extra observations. At 3.3Å the peptide planes were not resolved, making the assignment of these difficult. The use of well-refined closely-related structures for the molecular replacement solution (Chapter 4) assisted the refinement of the structure, since most of the peptide plane orientations in the framework regions of the antibody would be expected to be conserved.

6.2.1 Rigid Body Refinement

The molecular replacement solution model (Chapter 5) was output from MERLOT as fractional crystallographic coordinates and the V_L:V_H translated to construct the intact Fab. All B factors were set arbitrarily to 20Å^2.

The program CORELS (Sussman *et al.*, 1977) was used for the rigid body refinement, using solvent correction with a solvent density of 0.29 electrons/Å^3 and a B factor of 70Å^2. This value for the solvent density was chosen on the assumption that the PEG6000 precipitant was excluded from the crystal and that the expected solvent density for a dilute aqueous solution would be $\approx$0.30 electrons/Å^3 after the absence of hydrogens in the protein model was taken into account. Modest variation in the value of the solvent correction did not appear to affect the residual significantly.

Rigid Bodies	Resolution Range (Å)	Initial Residual (%)	Initial Corr. Coeff. (%)	Final Residual (%)	Final Corr. Coeff. (%)	No. of Reflections	No. of Cycles
2	10.0–8.0	53.0	21.7	46.0	40.0	230	10
2	10.0–7.0	47.9	37.1	45.8	45.8	466	10
2	10.0–6.0	49.8	32.3	46.8	36.8	883	16
2	8.0–6.0	46.9	36.8	46.9	37.0	653	12
2	8.0–5.5	47.3	35.1	46.4	38.0	989	20
2	8.0–5.0	45.4	43.3	44.9	44.3	1467	20
2	8.0–4.5	44.5	47.7	43.8	50.6	2193	6
2	8.0–4.0	44.7	45.4	44.3	46.0	3299	20
CDRs removed							
2	8.0–3.8	47.5	36.8	45.3	43.5	3908	20
4	8.0–3.5	52.4	39.9	45.5	42.2	4904	20
V_H from HyHEL-5 inserted							
4	8.0–3.5	47.6	35.6	43.8	45.9	4904	20

Corr. Coeff. — Correlation Coefficient =

$$\frac{N * \sum F_{obs} * F_{calc} - \sum F_{obs} * \sum F_{calc}}{\sqrt{(N * \sum F_{obs}^2 - \sum F_{obs} * \sum F_{obs}) * (N * \sum F_{calc}^2 - \sum F_{calc} * \sum F_{calc})}}$$

N = number of reflections summed over.
Residual = $\sum |F_{obs} - F_{calc}| / \sum F_{obs}$

Table 1: Rigid body refinement of Gloop2 $P2_1$ 'A' form using CORELS

Domain	Domain Pairs		Individual Domains	
	Angle (°)	Distance (Å)	Angle (°)	Distance (Å)
C_L	1.65	0.36	2.45	0.83
C_H1			2.78	1.59
V_L	2.65	0.42	6.15	0.51
V_H			8.18	1.84

Shifts with respect to molecular replacement model:
Angle - Absolute value of angular shift as rotation about unspecified axis.
Distance - Distance moved by centre of gravity.
Domain Pairs - model after refinement with domain pairs as rigid bodies.
Individual domains - model after refinement with V_H replaced (see text).

Table 2: Domain Shifts During Rigid-Body Refinement of $P2_1$ 'A' form

For the first stage of refinement the model was divided into the two fragments: V_L:V_H and C_L:C_H1: with the position of the C_L:C_H1 along **b** kept fixed. Initial attempts to refine these domain pairs using data at >10.0Å resulted in large angular shifts which decreased on the addition of higher resolution data (results not shown). This effect was attributed to the large contribution to the structure factors from bulk solvent scattering at low resolutions.

The refinement was restarted using data no lower than 10.0Å, initially using 10.0–8.0Å data. Upon convergence of the refinement at each current resolution range the maximum resolution was extended gradually, up to a maximum of 4.0Å. Once the 10.0–6.0Å refinement had converged, the 10.0–8.0Å shell was omitted from subsequent refinements, since this shell also seemed affected by solvent scattering. Upon convergence of the refinement using data in the range 8.0–4.0Å, the CDRs were deleted from the model. After refinement of the CDR-less model against 8.0–3.8Å data, the model was divided into four domains (V_L, V_H, C_L, C_H1). Further refinement brought little change to the orientation and position of these domains. On the basis of superior sequence homology (36% identity with McPC603, 60% identity with HyHEL-5, sequences in Appendix 1) the current V_H model from McPC603 (Satow *et al.*, 1986) was replaced by the V_H domain from HyHEL-5 (Sheriff *et al.*, 1987) (with the CDRs omitted). Refinement against 8.0–3.5Å data gave a lower residual and a higher correlation coefficient indicating the success of this substitution.

The progress of the refinement using CORELS is summarised in Table 1. The shifts of each domain with respect to the molecular replacement models during the course of refinement are given in Table 2.

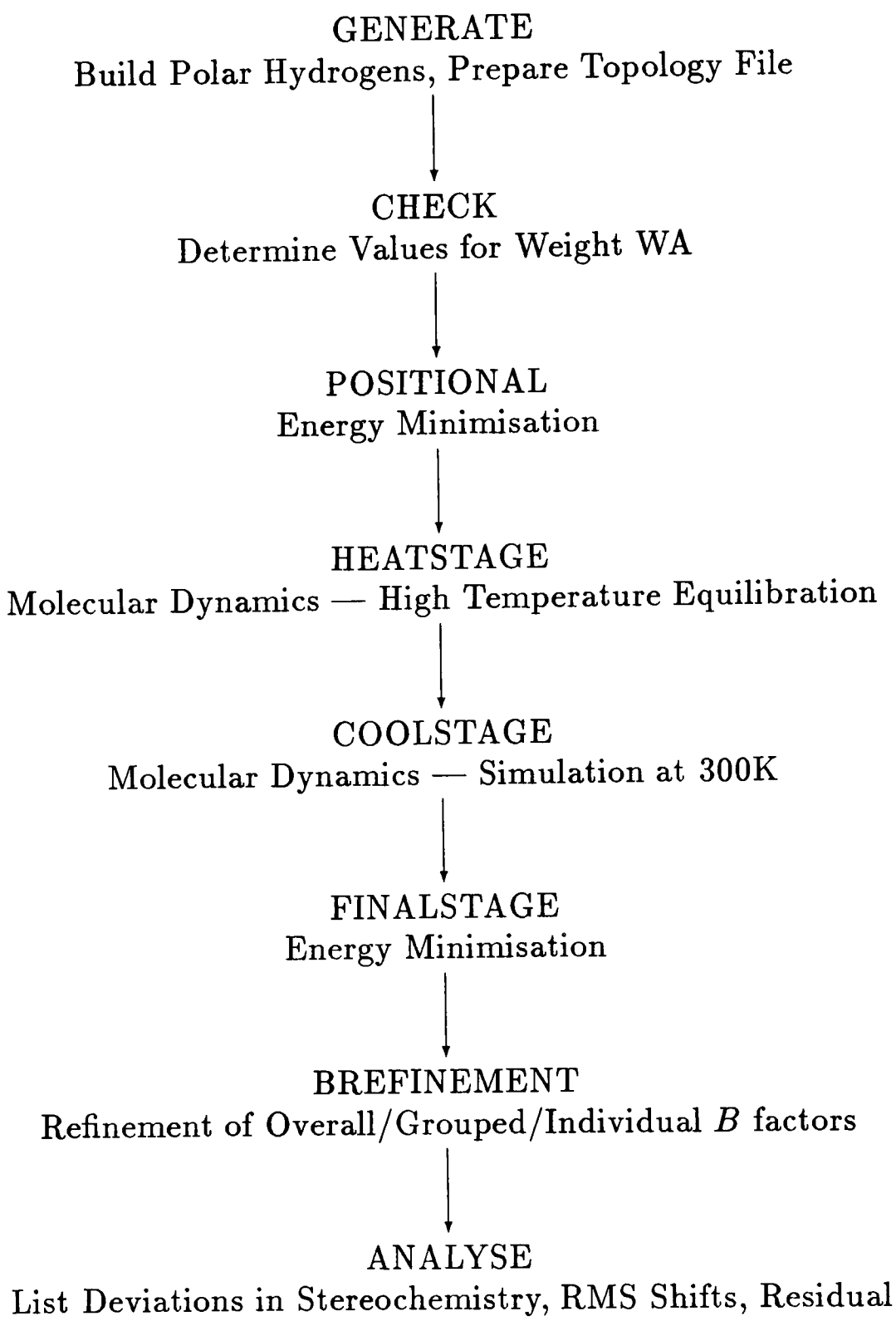


Figure 1: Molecular Dynamics Refinement Protocol

6.2.2 Refinement by Simulated Annealing

The model at the end of rigid body refinement contained many incorrect residues with respect to the Gloop2 sequence (Appendix 1). Sequence changes were made automatically using FRODO (Jones, 1985) without inspection of the density at these points.

The protocol used for the simulated annealing refinement (program X-PLOR; Brünger *et al.*, 1987) followed the recommendations in the manual (Brünger, 1988b) and is given in Figure 1. The only exception to this was that the $C\alpha$ atoms were not restrained during the first energy minimisation step ('POSITIONAL' in Figure 1). Versions 1.3 and 1.5 of X-PLOR were used for the refinements. Polar hydrogens were explicitly built onto the model at the 'generate' stage and a description of the molecular topology was formed. The value of the weight for the structure factors (WA) was obtained from the 'check' procedure, which comprised 100 steps of energy minimisation (no X-ray term) and 50fs of molecular dynamics at 300K. WA was estimated by comparison between the gradients of the potential energy ($E_{EMPIRICAL}$) and the X-ray energy ($E_{EFFECTIVE}$) at the end of the short dynamics run (see Chapter 2 for a description of $E_{EMPIRICAL}$ and $E_{EFFECTIVE}$). The first refinement stage was 200 steps of energy minimisation refinement ('positional'), followed molecular dynamics simulation at high temperature (1500–4000K) for 500 steps of 1.0fs or 1000 steps at 0.5fs. The coordinates were input to the 'coolstage' of 200 steps of 1.0fs at 300K, followed by a further 200 steps energy minimisation and B factor refinement.

This protocol was applied to the coordinates after rigid body refinement and sequence changes, except that the B factor was maintained at 20Å^2 for all atoms in the model. The first refinement cycle was particularly dramatic, and is described in detail below.

The initial residual for data in the 8–3.5Å range was 45.4%, higher than for the model output from CORELS, due to the sequence changes and the absence of solvent correction. The value of WA (161540) from the 'check' procedure was adopted for the subsequent steps. 200 steps of energy minimisation refinement using 8–3.5Å data brought the residual down to 28.2%. A molecular dynamics simulation of 1000 steps of 0.5fs at 2000K ('heatstage'), was followed by 250 steps of 1.0fs at 300K ('coolstage'). The residual at the end of this procedure was 25.2%. A further 200 steps of energy minimisation refinement brought the residual down to 22.5%. No B factor was refined. The root mean square deviations on backbone and side-chain atoms between the initial model and that after cycle 1 were 1.01Å and 1.62Å respectively.

Inspection of $(2F_{obs} - F_{calc}), \alpha_{calc}$ Rayment-weighted (Rayment, 1983) electron density maps calculated for the model after cycle 1, showed density in the regions expected for the CDRs and for the peptide chain at the elbow bend (which was not included in the molecular replacement model). The method used for calculation of the electron density maps for this and subsequent models is given in Appendix 3. CDRs and elbow peptides were built in as alanine/glycine backbones (glycine where there was glycine in the Gloop2 sequence, otherwise alanine) to the available density. An inspection of all the sequence changes was made and these were rebuilt as necessary.

The resolution used in the refinement of this ALA/GLY-CDR model was extended to 3.3Å. The refinement used the same protocol as for cycle 1, and brought the residual down, although the stereochemistry became slightly worse. The $(2F_{obs} - F_{calc}), \alpha_{calc}$ electron density maps for the model after cycle 2 showed clear density for a number of side-chains in the CDRs and elbow peptides, and these were built in. Modest movements in the CDRs and elbow peptide backbones had accommodated

some inaccurate building after cycle 1, but the overall conformation was maintained in each case.

Re-evaluation of the value of WA for this model using the 'check' procedure gave a larger value than that used for the previous refinement cycles (350000 rather than 207100 and 161540). When this value was adopted the stereochemistry of the model deteriorated during the refinement cycle. In order to maintain the quality of the stereochemistry the value of WA was set to 210000 (similar to the values used for previous cycles) for subsequent refinement cycles.

Refinement proceeded, adding in side-chains in the CDRs and elbows as they became visible in the electron density maps. The whole structure was inspected and rebuilt where necessary after cycles 1, 4 and 8 on the basis of $(2F_{obs} - F_{calc}), \alpha_{calc}$ omit maps (map calculated from a model in which 22 residues had been deleted). The stereochemistry of the model was tightened at cycle 5 with WA being set to 120000 for all subsequent cycles. An overall B factor was refined at cycles 5 and 12 on the basis of data in the range 5–3.3Å. This adopted unreasonably low values of 4.91Å^2 and 3.40Å^2 at these two refinement stages, possibly because the trend in F_{obs} with $\sin^2\theta/\lambda^2$ for such a narrow resolution shell is insufficient to determine the B factor for the structure accurately.

The four C-terminal residues of C_L, omitted after cycle 1 due to poor density, were built in after cycle 5 and the inter-chain disulphide bond formed. Rebuilding was hampered by the limited resolution of the electron density maps, which rarely enabled the peptide plane to be resolved, and made the interpretation of side-chain density often difficult. The rebuild after cycle 11 was guided by the comparison with the structure of the $P1$ form at 2.8Å (see below) which led to a number of modifications which would have been more difficult to detect using the electron density map alone. In particular arginine side-chains were often indistinguishable

Cycle	Resolution Range (Å)	Number of Atoms†	Number of Reflections	Initial Residual (%)	Final Residual (%)	RMS Deviations Bond Length (Å)	RMS Deviations Bond Angle (°)	MD	Weighting WA
1	8.0–3.5	2902	4904	45.4	22.5	0.023	5.05	2000K	161540
2	8.0–3.3	3119	5423	29.4	21.3	0.025	5.10	2000K	200710
3a	8.0–3.3	3192	5423	23.4	18.5	0.033	6.00	4000K	350000
3	8.0–3.3	3192	5423	18.0	19.3	0.025	5.38	NONE	210000
4	8.0–3.3	3169	5423	21.3	19.9	0.025	5.05	4000K	216000
5	8.0–3.3	3169	5423	19.9	21.1	0.018	4.40	NONE	120000
6	8.0–3.3	3224	5423	22.9	21.2	0.017	4.13	4000K	120000
7	8.0–3.3	3233	5423	22.7	20.5	0.016	4.14	2000K	120000
8	8.0–3.3	3241	5423	21.3	20.6	0.017	4.12	2000K	120000
9	8.0–3.3	3241	5423	23.3	20.5	0.016	4.09	2000K	120000
10	8.0–3.3	3241	5423	20.5	20.1	0.017	3.63	2000K*	120000
11	8.0–3.3	3241	5423	21.0	19.7	0.017	3.58	NONE	120000
12	8.0–3.3	3241	5423	20.9	19.5	0.017	3.59	NONE	120000
13	8.0–3.3	3241	5423	20.2	19.7	0.017	3.59	NONE	120000
14	8.0–3.3	3241	5423	21.5	19.2	0.017	3.52	1500K*	120000

MD - temperature of molecular dynamics simulation.

* 'Slow' cooling molecular dynamics run.

† Number of non-hydrogen atoms.

Residual defined in Table 1

For details see discussion in text.

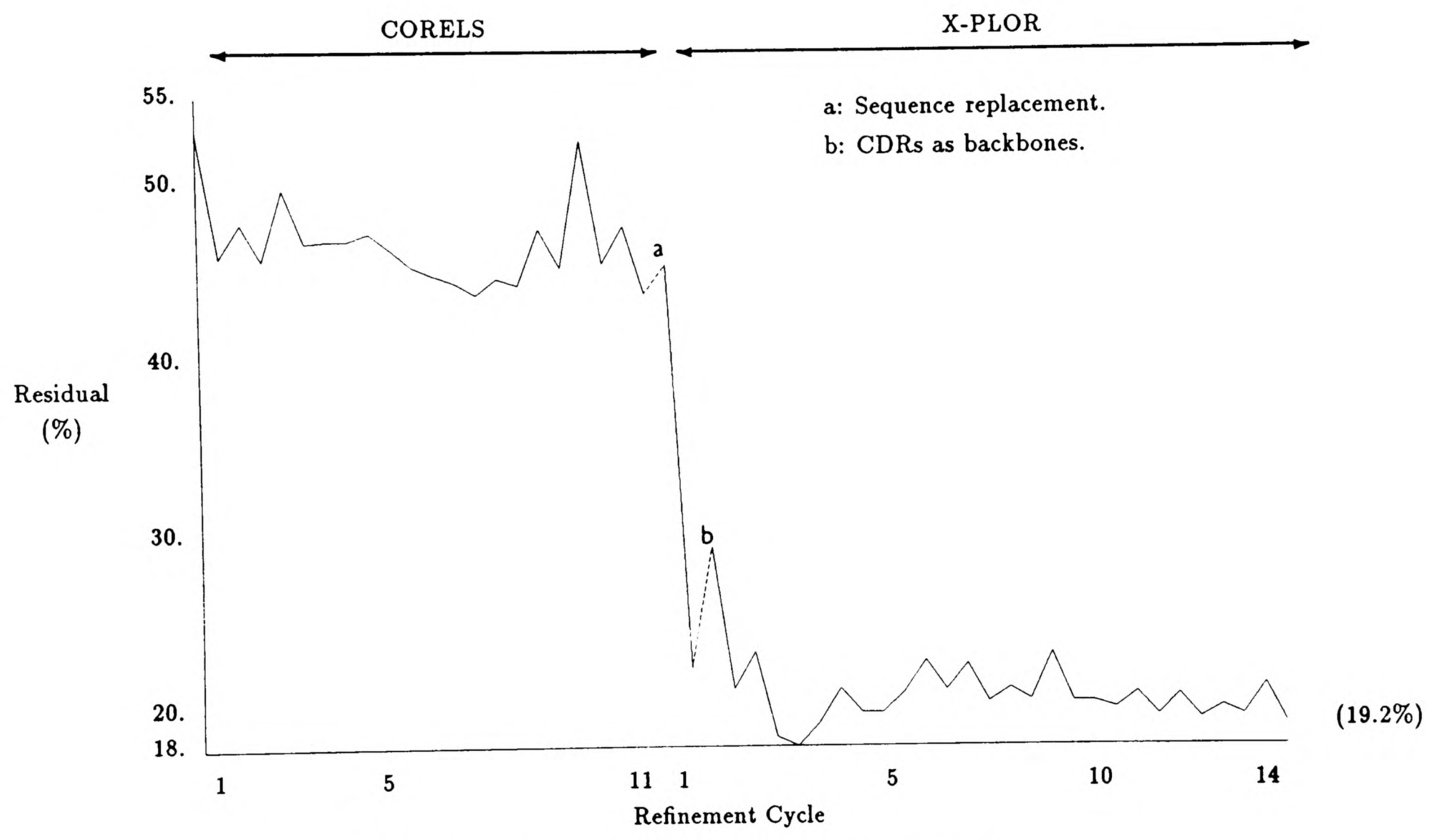
Table 3: Simulated Annealing Refinement of Gloop2 $P2_1$ 'A' form

from main-chain density in less well phased areas. This led to the initial incorrect placing of the C-terminus of C_L which was resolved by the requirement of disulphide bond formation, but more seriously led to the misplacement of the loop 64L–71L (sequential numbering) in V_L which was rebuilt in an incorrect conformation after it was mistakenly deleted before the first X-PLOR cycle. Comparison with the higher resolution $P1$ structure (see below) led to the rebuilding of this loop after cycle 13 and gave a decrease in the residual upon refinement.

The progress of refinement is shown in Table 3. In the last stages of refinement the 'slow' cooling option was used in preference to the 'heatstage' and 'coolstage' steps, where the molecular dynamics simulation was started off at a high temperature, and reduced every 50 steps to a final temperature of 300K. Two modifications were made to the potential function after cycle 9 on the advice of Dr. W. Weiss (pers. commun.). The parameter file PARAM19X.PRO was altered by changes in bond angle force constants:

$C\alpha$–C–N	60kcal/mol from 20kcal/mol
C–$C\alpha$–N	70kcal/mol from 45kcal/mol
C–N–H_{AMIDE}	50kcal/mol from 30kcal/mol
$C\alpha$–N–H_{AMIDE}	70kcal/mol from 35kcal/mol

X-PLOR had previously buried much of the stereochemical distortion into the peptide groups, but with these new parameters the distortions were buried preferentially in the C_α–C_β bonds. Clearly neither of these potentials are ideally balanced, but at least in the latter case the peptide planes are better behaved. The other modification was to turn off the electrostatic term in the potential, which had resulted in some polar side-chains on the protein surface looping back to bind to the surface, and in some cases side-chains had been pulled partially out of density to make elec-



Residuals plotted at beginning and end of each refinement step.
Broken lines indicate major revision of model.

Figure 2: Gloop2 $P2_1$ 'A' form - Residual Versus Refinement Step

Resolution	8.0–3.3Å
No. of reflections	5043
No. of atoms	3241
Residual	19.2%
Overall B-factor	3.40Å^2
RMS deviations from ideality:	
RMS Δbond length	0.017Å
RMS Δbond angle	3.521°
RMS Δdihedral	28.648°
RMS Δimpropers	1.457°

Residual defined in Table 1.

Table 4: Status of Gloop2 $P2_1$ 'A' form Model After Cycle 14

trostatically favourable interactions. Although turning off electrostatic terms also removed hydrogen-bonding, this was not seen to adversely affect the conformation in regions where there was extensive hydrogen bonding (e.g. β-sheets).

The current model at the end of cycle 14 has a residual of 19.2% on all $F > 0$ in the range 8-3.3Å, with all residues in place according to the Gloop2 sequence in Appendix 1. No solvent molecules have been built in, although density is visible for a few of them, and no solvent correction has been applied. The variation of the residual throughout refinement is shown in Figure 2, and the current refinement status is shown in Table 4.

6.2.3 Refinement of Gloop2 $P2_1$ 'A' Form, G2P2$_1$H Dataset

In order to investigate the differences between the two non-isomorphous Gloop2 $P2_1$ 'A' form structures, the cycle 14 model from the G2P2$_1$G dataset was refined against the G2P2$_1$H dataset, using all data with $F > 0$ in the range 8–3.3Å. The value of WA adopted for refinement was the same as for the G2P2$_1$G dataset (120000), since the number of parameters and observations were approximately the same in both cases, and the two sets of F_{obs} were on approximately the same absolute scale.

To allow for any major movements between the crystal forms, the first step was to use rigid body refinement of individual domains against 5–3.3Å data with van de Waals and electrostatic interaction energies omitted. The initial residual of 40.3% decreased to 30.8% after 20 steps. 200 steps of energy minimisation refinement, using data in the range 8–3.3Å, brought the residual down from 31.1% to 21.6%. The molecular dynamics simulation used the 'slow' cooling method, from 2000K to 300K with 50 steps of 0.5fs per 25K decrement in temperature. The residual at the end of this procedure was 21.1%. A further 200 steps of energy minimisation brought

Domain	Cycle 1		Cycle 2	
	Angle (°)	Distance (Å)	Angle (°)	Distance (Å)
V_L	4.43	0.70	3.11	0.72
C_L	3.72	0.94	4.74	1.07
V_H	5.54	0.36	5.77	0.56
C_H1	1.66	0.39	2.28	0.33

Shifts with respect to initial model:
Angle - Absolute value of angular shift as rotation about unspecified axis.
Distance - Distance moved by centre of gravity.
Cycle 1 - molecular replacement model updated with $P2_1$ 'A' domains.
Cycle 2 - Cycle 1 model updated with $P1$ domains.

Table 5: Domain Shifts During Rigid-Body Refinement of $P2_1$ 'A' form, G2P2$_1$H Dataset

Resolution	8.0–3.3Å
No. of reflections	4906
No. of atoms	3241
Residual	19.1%
Overall B-factor	1.23Å^2
RMS deviations from ideality:	
RMS Δbond length	0.017Å
RMS Δbond angle	3.533°
RMS Δdihedral	28.324°
RMS Δimpropers	1.451°

Residual defined in Table 1.

Table 6: Status of Gloop2 $P2_1$ 'A' form model, G2P2$_1$H dataset

the residual down to 19.2% which decreased to 19.1% after refinement of an overall B factor (B factor refined on the basis of 5–3.3Å data but residual on the basis of 8–3.3Å data). The overall B factor, in common with the G2P2$_1$G refinement, adopted an unreasonably low value (1.23Å^2). The RMS shifts on backbone and side-chains with respect to the initial model were 0.91Å and 1.31Å respectively (excluding hydrogens).

A summary of the movements of the four domains under rigid body refinement are given in Table 5. The refinement status of the current model at 3.3Å is given in Table 6.

6.3 Refinement of Gloop2 $P1$ Form

The refinement strategy for the Gloop2 $P1$ crystal form followed the same strategy as the $P2_1$ 'A' form. The G2P1C dataset, with which the structure was solved, was used for the rigid body refinement and the first X-PLOR cycle. After the higher resolution datasets G2P1D and G2P1E had been collected, the merged G2P1C+G2P1D dataset was used for subsequent refinements. This dataset, although slightly inferior to the G2P1D+G2P1E merged dataset, was used because it was consistent with the G2P1C dataset (i.e. it had data in common with it). Latterly, as the model improved, the G2P1D+G2P1E dataset was used, as it was more complete and had a better trend in R_{sym} than the G2P1C+G2P1D dataset (Chapter 4). All data with $F > 0$ were used in the refinements.

6.3.1 Rigid Body Refinement

Rigid body refinement with CORELS used the G2P1C dataset. The molecular replacement solution was output as fractional coordinates and Fv fragments translated to assemble the Fabs, with the B factors set arbitrarily to 20Å^2. The solvent

Rigid Bodies	Resolution Range (Å)	Initial Residual (%)	Initial Corr. Coeff. (%)	Final Residual (%)	Final Corr. Coeff. (%)	No. of Reflections	No. of Cycles
CDRs removed							
4	8.0–6.0	45.5	34.6	43.3	38.7	1263	15
4	8.0–5.0	45.5	39.8	43.6	44.8	2871	15
4	8.0–4.0	44.4	43.9	42.8	45.8	6645	15
4	8.0–3.5	46.5	40.5	44.2	40.6	10093	15
V_H from HyHEL-5 inserted							
8	8.0–3.5	48.1	33.2	40.3	50.8	10093	15

Corr. Coeff. — Correlation Coefficient =

$$\frac{N * \sum F_{obs} * F_{calc} - \sum F_{obs} * \sum F_{calc}}{\sqrt{(N * \sum F_{obs}^2 - \sum F_{obs} * \sum F_{obs}) * (N * \sum F_{calc}^2 - \sum F_{calc} * \sum F_{calc})}}$$

N = number of reflections summed over.
Residual = $\sum |F_{obs} - F_{calc}| / \sum F_{obs}$

Table 7: Rigid body refinement of Gloop2 $P1$ form using CORELS

Domain	Domain Pairs Angle (°)	Domain Pairs Distance (Å)	Individual Domains Angle (°)	Individual Domains Distance (Å)
C_L	0.60	0.00	1.34	1.22
C_H1			1.68	1.38
C_L	1.41	1.04	2.26	1.35
C_H1			1.34	1.75
V_L	2.30	0.25	3.53	1.02
V_H			3.11	0.73
V_L	1.64	0.89	1.82	1.59
V_H			3.34	0.45

Shifts with respect to molecular replacement model:
Angle - Absolute value of angular shift as rotation about unspecified axis.
Distance - Distance moved by centre of gravity.
Domain Pairs - model after refinement with domain pairs as rigid bodies.
Individual domains - model after refinement with V_H replaced (see text).

Table 8: Domain Shifts During Rigid-Body Refinement of $P1$ form

correction was set to 0.29 electrons/Å^3 with a B factor of 70Å^2. The CDRs were deleted, and this initial model was refined as four rigid domain pairs against 8–6Å data with the first C_L:C_H1 fixed at the origin. As the refinement converged for each resolution shell the maximum resolution was extended, up to a maximum of 3.5Å. The two McPC603 V_H domains in the model were then replaced with HyHEL-5 V_H domains with the CDRs deleted (as in the $P2_1$ 'A' case) and the model divided into the eight domains. These were refined as rigid bodies against 8–3.5Å data until convergence.

A summary of this refinement is given in Table 7, and a summary of the shifts of each domain during refinement in Table 8.

6.3.2 Refinement by Simulated Annealing

Sequence changes were then made to the model to convert it to the Gloop2 sequence (Appendix 1). Simulated annealing refinement, against 8–3.5Å data from the G2P1C dataset, was seen to behave similarly to the $P2_1$ 'A' form refinement. The initial residual rapidly decreased to 28.9% after 200 cycles of energy minimisation. A molecular dynamics simulation for 1000 steps of 0.5fs at 2000K, followed by 250 steps of 0.5fs at 300K, lowered the residual to 23.7%. After another 200 steps of energy minimisation the residual was 22.9%. No B factor was refined for the model at this stage (B factors maintained at 20Å^2). The stereochemistry of the model was comparable to that of the $P2_1$ 'A' form at the same stage (RMS deviation from ideality on bond lengths were 0.026Å for $P1$ and 0.023Å for $P2_1$ 'A'). The RMS differences on backbone and side-chains with respect to the sequence-replaced model was 1.22Å and 1.80Å respectively.

Inspection of $(2F_{obs} - F_{calc}), \alpha_{calc}$ electron density maps after cycle 1 at the CDR regions showed poor, disconnected density which was not readily interpretable. The

refinement of the model was suspended until higher resolution data became available in the form of the G2P1C+G2P1D merged dataset.

Upon the switch to the new dataset, the resolution range used in refinement was extended to 8–2.8Å. Refinement of the cycle 1 model at 2.8Å used a value for WA of 473730 and the stereochemistry was seen to relax as the residual was brought down from 34.2% to 26.0%. The value of WA was set to 250000 and the refinement repeated, and individual B factors were refined (cycle 3). The restraints on B factors were $\sigma = 3.0$ on 1–2 atom pairs and $\sigma = 4.0$ on 1–3 atom pairs. These restraints were deliberately weak in order to identify regions of chain that were badly positioned, by their high thermal parameters. The stereochemistry of the model was greatly improved (RMS deviations from ideality on bond lengths dropped from 0.031Å to 0.019Å).

The $(2F_{obs} - F_{calc}), \alpha_{calc}$ maps from this model after cycle 3 showed improved density for the CDRs, which were built into unattributed electron density as alanine/glycine backbones using the $P2_1$ 'A' form coordinates as an aid to density interpretation. Comparison of the two Fabs in the unit cell at the end of the next cycle gave RMS fits of 1.2–1.8Å on backbone plus $C\beta$ atoms for each domain. In order to reduce these deviations the model was "averaged" by applying tight non-crystallographic symmetry restraints between all atoms in the model. The restraints were applied on a domain-by-domain basis, allowing variation in domain association to be maintained whilst restraining the domain structures to be similar. Inspection of the structures showed that the elbow angles in the two Fabs were *not* identical (evaluated in Chapter 7), and the non-crystallographic symmetry was not exact. The weights used to restrain symmetry-related positions and B factors were 300kcal/mol and $\sigma_B = 1.0$ respectively.

After refinement of the model with the non-crystallographic restraints applied,

the stereochemistry of the model improved, and the residual rose by 0.4% to 25.0% after B factor refinement (cycle 5). Density was apparent for some CDR side-chains in the $(2F_{obs} - F_{calc}), \alpha_{calc}$ maps and were built into this density *only* if they were observed in both Fab copies to maintain the symmetry. The four C-terminal residues on both C_L domains were deleted since there was no ordered density for them. More CDR side-chains were built in cycles 6–9 and the elbow regions became better defined in density. Elbow peptides were built in as alanine/glycine backbones after cycle 9 and some elbow peptide side-chains built in at cycle 10 where there was unattributed density. The light chain elbows were much clearer than the heavy chain elbows and it was not possible to build a continuous length of chain for the latter until a later stage.

The non-crystallographic symmetry restraints were reduced at cycle 10 and removed altogether at cycle 11. It was apparent that the averaging procedure had had detrimental effects on the C_H1 domain loops at the end of the domain remote from the C_L:C_H1-V_L:V_H interface, indicated by a substantial increase in the mean B factor for these loops. A compromise structure was created from the model after cycle 5 with the additional CDR and elbow peptide structure from the model after cycle 11. This "hybrid" structure was refined using the "slow" cool protocol, using the revised potential (as with the $P2_1$ 'A' form refinement) with a reduced value of WA (220000) to maintain good stereochemistry. The model at the end of this 12th cycle was rebuilt where necessary on the basis of $(2F_{obs} - F_{calc}), \alpha_{calc}$ omit maps where 42 residues were deleted from the model at a time. Non-crystallographic symmetry was re-imposed (50kcal/mol on position, $\sigma_B = 1.5$) on $C\alpha$ atoms in the β-sheet regions (defined manually on the graphics) during cycle 13, and the electrostatic term in the energy function was omitted. The model at the end of cycle 13 was rebuilt with reference to the second copy of each domain in the unit

Cycle	Resolution Range (Å)	Number of Atoms†	Number of Reflections	Initial Residual (%)	Final Residual (%)	RMS Deviations Bond Length (Å)	RMS Deviations Bond Angle (°)	MD	Weighting WA
1	8.0–3.5	5840	10093	48.1	22.9	0.026	5.39	2000K	345280
G2P1C+G2P1D dataset									
2	8.0–2.8	5840	19079	34.2	26.0	0.031	5.72	4000K	473730
3	8.0–2.8	5832	19079	26.0	26.1	0.019	4.44	2000K	250000
4	8.0–2.8	6068	19079	28.8	24.6	0.017	4.12	4000K	250000
5	8.0–2.8	6068	19079	24.6	25.0	0.016	3.72	2000K	250000
6	8.0–2.8	6156	19079	26.6	24.7	0.015	3.54	2000K	250000
7	8.0–2.8	6196	19079	24.9	24.2	0.013	3.38	2000K	250000
8	8.0–2.8	6280	19079	24.4	23.7	0.013	3.38	2000K	250000
9	8.0–2.8	6312	19079	23.6	23.1	0.013	3.39	2000K	250000
10	8.0–2.8	6312	19079	23.1	23.0	0.013	3.41	2000K	250000
11	8.0–2.8	6312	19079	23.0	22.1	0.014	3.59	2000K	220000
12	8.0–2.8	6312	19079	25.3	22.6	0.014	3.38	2000K*	220000
13	8.0–2.8	6324	19079	24.4	22.4	0.015	3.29	NONE	220000
14	8.0–2.8	6338	19079	24.3	21.9	0.014	3.15	NONE	220000
15	8.0–2.8	6372	19079	22.3	21.8	0.014	3.14	2000K*	220000
16	8.0–2.8	6372	19079	21.9	21.4	0.014	3.10	2000K*	220000
Using re-evaluated cell dimensions:									
17	8.0–2.8	6372	18893	26.2	21.5	0.014	3.10	3000K*	220000
18	8.0–2.8	6343	18893	22.3	21.6	0.014	3.10	NONE	220000
19	8.0–2.7	6343	19513	21.9	21.9	0.014	3.10	NONE	220000
20	8.0–2.8	6343	18893	21.3	21.1	0.015	3.12	NONE	260000
G2P1D+G2P1E dataset									
21	8.0–2.8	6343	20177	23.1	22.0	0.016	3.24	3000K*	260000
22	8.0–2.8	6343	20177	22.0	22.7	0.015	3.12	3000K*	260000
23	8.0–2.8	6343	20177	22.7	21.5	0.016	3.15	2000K*	260000
24	8.0–2.8	6343	20177	22.4	21.2	0.015	3.08	1500K*	260000

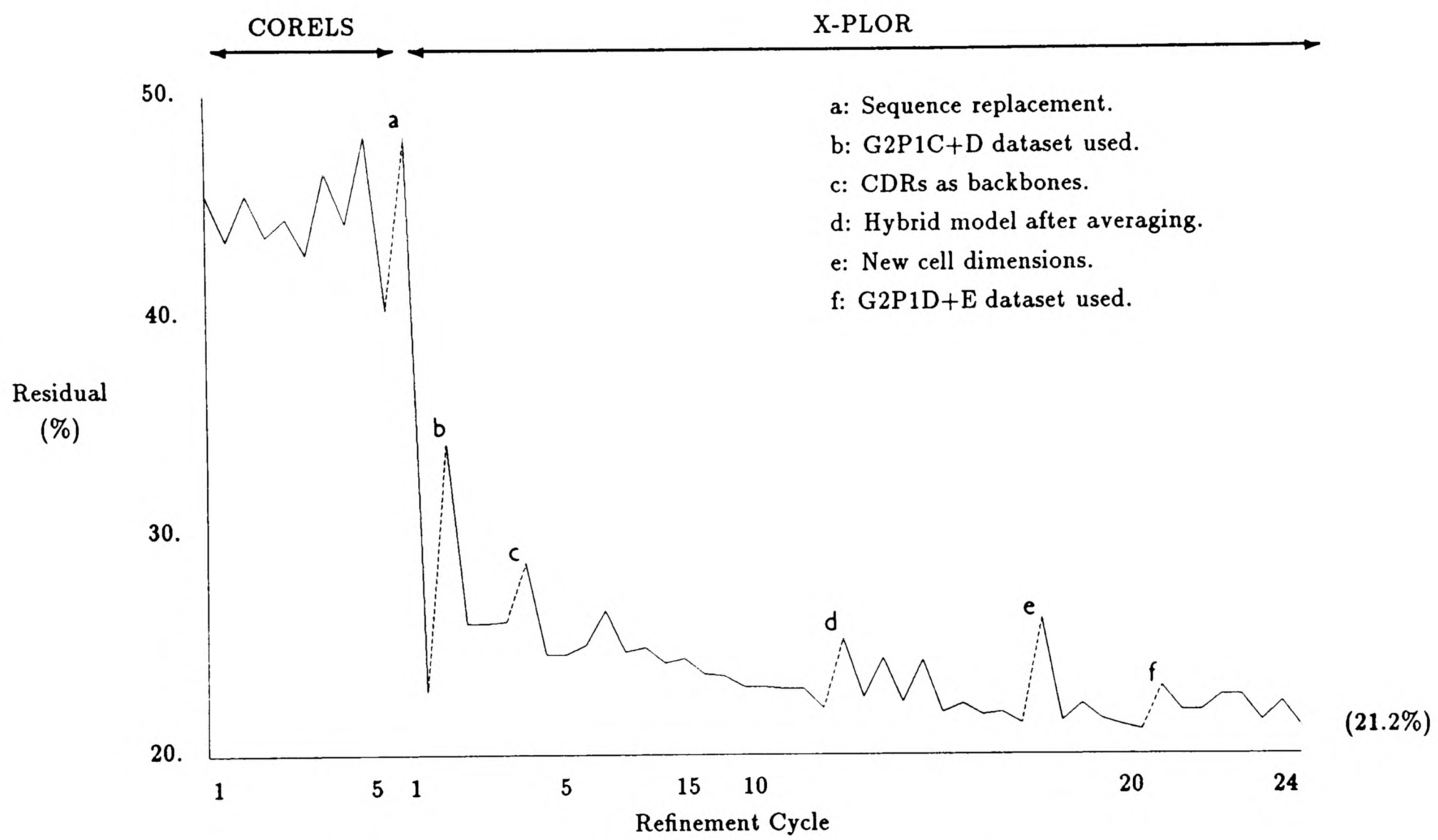
Notation as for Table 3.

Table 9: Simulated Annealing Refinement of Gloop2 $P1$ form

cell. Minor modifications and a reduction in the non-crystallographic symmetry weight to 25kcal/mol and $\sigma_B = 2.0$ on $C\alpha$, brought the residual to 21.4% after cycle 16.

Inspection of the model at this point showed that the C_H1 loops mentioned above had generally high B factors (35–60Å^2) and some regions packed against symmetry-related regions with low B factors (usually < 30Å^2). $(2F_{obs} - F_{calc}), \alpha_{calc}$ and $(F_{obs} - F_{calc}), \alpha_{calc}$ maps calculated with these loops omitted showed little density on which to base chain retracing. The regions of B factor correlated to the same residues on the two C_H1 domain copies, although they had not been restrained to do so. These disordered regions made it impossible to build in the four C-terminal residues of C_L and form the inter-chain disulphide bridge since the C_H1 cysteine was disordered.

At this point the cell dimensions of the G2P1D dataset were re-evaluated. A spot centroid list was constructed from a large angular range of frames from the G2P1D dataset and the cell parameters re-refined (see Chapter 4 for a description of the software and procedures). On the basis of this refinement the new $P1$ cell for the G2P1C+G2P1D dataset was $a = 49.4$Å, $b = 83.3$Å, $c = 68.4$Å, $\alpha = 106.4°$, $\beta = 104.8°$ and $\gamma = 94.9°$, a little different to the previous cell dimensions ($a = 49.8$Å, $b = 83.6$Å, $c = 69.1$Å, $\alpha = 106.2°$, $\beta = 104.6°$, $\gamma = 94.9°$). The model after cycle 16 was refined on the basis of these new cell dimensions and resulted in essentially the same residual (21.5%) with small changes in atomic positions. No improvement in the C_H1 loops was observed. The new cell also decreased the number of reflections in the 8.0–2.8Å range slightly (to 18893 from 19079). Nevertheless the new cell was adopted since it was likely that this was a better estimate of the true cell dimensions than the previous cell, as it had been refined over a larger angular range of data (128°, rather than small angular ranges at 0° and 90°).



Residuals plotted at beginning and end of each refinement step.
Broken lines indicate major revision of model.

Figure 3: Gloop2 *P*1 form - Residual Versus Refinement Step

The N- and C-termini were rebuilt during cycles 18–20, and the value of WA increased to 260000. At the end of cycle 20 the residual was 21.1%. The dataset used in refinement was then changed to G2P1D+G2P1E (Chapter 4), using the revised cell dimensions and over the same resolution range (8–2.8Å). Non-crystallographic symmetry (25kcal/mol, $\sigma_B = 2.0$) was maintained on $C\alpha$ atoms only. The initial residual for this dataset was 2.0% higher than for the G2P1C+G2P1D data at 23.1%, probably due to the increased proportion of higher-resolution reflections in the data (see Chapter 4). Energy minimisation followed by B factor refinement brought the residual down to 22.0% (cycle 21).

In cycles 22 and 23 the non-crystallographic symmetry was imposed on the β-sheet regions of the Fabs: restraints of 100kcal/mol on position and $\sigma_B = 2.0$ were applied to all atoms in the β-sheet regions for one cycle (including molecular dynamics), and then the restraints on the side-chains were released to allow surface residues to deviate according to crystal packing forces. The restraints were maintained on main chain (C,N,O,$C\alpha$) atoms and also on proline $C\delta$ atoms since some proline ring 'puckers' were seen to be ill-behaved. At the end of this procedure the residual had dropped by 0.5%. The model after cycle 23 was modified by addition of the four residues at the C-terminus of C_H1 for the second Fab, for which ordered density was observed, with reference to the conformation in the $P2_1$ structure (cycle 14 model). Both heavy chain elbow regions were rebuilt, with the chain trace of the second one changed to conform to the first. Refinement of this model brought a modest improvement in the residual, and encouragingly the four C-terminal residues of C_H1 maintained reasonable B factors (cycle 24). Omit map density at the heavy chain elbow regions confirmed the correctness of the rebuild there.

The model after cycle 24 is the current model. Not all the atoms are included

Cycles	Restraints		Atoms affected
	Position (kcal/mol)	σ_B	
4–8	300.0	1.00	All except hydrogen
9	300.0	1.00	$C\alpha, N, C, CO$
	200.0	1.50	All sidechain except hydrogen
10	150.0	1.50	$C\alpha, N, C, CO$
	100.0	2.00	All sidechain except hydrogen
11–12	NONE		
13–14	50.0	1.50	β-sheet $C\alpha$
15–21	25.0	2.00	β-sheet $C\alpha$
22	100.0	2.00	All β-sheet except hygrogen
23–24	100.0	2.00	β-sheet $C\alpha, N, C, CO$ and PRO $C\delta$

Table 10: Application of Non-Crystallographic Symmetry Restraints in the Refinement of the $P1$ form of Gloop2

Resolution	8.0–2.8Å
No. of reflections	20177
No. of atoms	6394
Residual	21.2%
RMS deviations from ideality:	
RMS Δbond length	0.015Å
RMS Δbond angle	3.083°
RMS Δdihedral	28.709°
RMS Δimpropers	1.336°

Residual defined in Table 1.

Table 11: Status of Gloop2 $P1$ form Model After Cycle 24

$|2F_{obs} - F_{calc}|, \alpha_{calc}$ electron density map.
Contoured at 1σ above zero, no $F(000)$ term.
All data in range 8.0–2.8Å with $F > 0$ included in calculation.
All protein atoms from current (cycle 24) model included in calculation
Glu50H–Wat1H O–O distance 2.42Å.
Tyr94L–Wat1H O–O distance 3.28Å.

Figure 4: A Putative Solvent Molecule in the Gloop2 $P1$ Cell

into the structure due to disordered/uninterpretable density. The two N-terminal residues of V_H and the two C-terminal residues of C_H1 are missing from both Fabs, and the four C-terminal residues of C_L are missing from the first Fab. The regions of high B factor on the C_H1 domain remain a problem area: the density provides no obvious alternative to the current chain tracing and yet the omit map density for the current chain trace is unconvincing. These regions, some of which participate in crystal contacts, are in need of further attention. Solvent molecules have not been incorporated in the model. Figure 4 shows a putative water site, with a water placed manually at the centre of the density to show the interactions with the neighbouring residues. Limited adjustment of Glu50H would optimise hydrogen bonding with the water, and the hydroxyl groups of Thr35H and Tyr94L. The density is suggestive of a bifurcated hydrogen bond between the Glu50H carboxylic oxygen, Wat1H and the Tyr94L *OH* with hydrogen bond distances in the range observed in other protein structures (2.5–3.5Å, (Baker and Hubbard, 1984)). This density was obtained from the first of the two Fabs in the $P1$ cell. In the second Fab this water position is occupied by the NH_3^+ group of a lysine side-chain from a symmetry-related molecule.

Table 9 gives a summary of the refinement cycles outlined above, which is also shown in Figure 3. Table 10 summarises the use of non-crystallographic symmetry and the weights applied, and Table 11 gives the current status of the model after cycle 24.

6.4 Refinement of Gloop2 $P2_1$ 'B' Form

The incompleteness and limited resolution of the G2P2$_1$I dataset made the use of stereochemically-restrained refinement on this crystal form untenable. Only rigid body refinement was used, with the assumption that the domain structure would be the same as the other Gloop2 crystal forms. The program X-PLOR (Brünger *et al.*

Domain	$P2_1$ 'A' Domains		$P1$ Domains	
	Angle (°)	Distance (Å)	Angle (°)	Distance (Å)
C_L	4.43	0.70	4.10	0.72
C_H1	3.72	0.94	4.74	1.07
V_L	5.54	0.35	5.77	0.55
V_H	1.66	0.39	2.18	0.33

Shifts with respect to molecular replacement model:
Angle - Absolute value of angular shift as rotation about unspecified axis.
Distance - Distance moved by centre of gravity.

Table 12: Domain Shifts During Rigid-Body Refinement of $P2_1$ 'B' form

Resolution	8.0–3.3Å
No. of reflections	3450
No. of atoms	3169
Residual	36.6%
Overall B-factor	33.9Å^2
RMS deviations from ideality:	
RMS Δbond length	0.016Å
RMS Δbond angle	3.036°
RMS Δdihedral	28.814°
RMS Δimpropers	1.378°

Residual defined in Table 1.

Table 13: Status of Gloop2 $P2_1$ 'B' form model

was used, 1987) rather than CORELS (Sussman *et al.*, 1977) for this refinement. All data with $F > 0$ were used in the refinement.

The model, solved using partially refined $P2_1$ 'A' coordinates (model after cycle 12), was output from MERLOT as fractional coordinates and updated with the most recent $P2_1$ 'A' form coordinates (model after cycle 14). The initial residual on all $F > 0$ in the range 5–3.3Å was 45.9%. Fifty cycles of rigid body refinement, with domains as rigid bodies, lowered the residual to 37.7% on 5–3.3Å data (38.5% on 8–3.3Å data). Refinement of an overall B factor for the model on the basis of 5–3.3Å data brought the residual to 37.6% (38.4% on 8–3.3Å data). The overall B factor was 27.1Å^2 (in contrast to the low overall B factors of the $P2_1$ 'A' refinements). A summary of the angle and translation shifts applied to the model are given in Table 12 for each domain.

The model was amended by superimposing individual domains from the first Fab in the $P1$ form (model after cycle 24) onto the resultant structure and updating the coordinates accordingly (the four C-terminal C_L residues, the two C-terminal C_H1 residues and the two N-terminal residues of V_H were kept from the previous model, being absent from the $P1$ Fab model used). The overall B factor was set to 20Å^2. Another 50 cycles of rigid body refinement and refinement of an overall B factor brought the residual down from 37.0% to 35.4% on data in the range 5–3.3Å (36.6% on 8–3.3Å data). This model is the current model.

The modified model ($P1$ Fab domains) was favoured because of three reasons: (1) the $P1$ structure was the most highly refined structure; (2) the elbow bends in the $P1$ and $P2_1$ 'B' Fabs were quite similar (evaluated in Chapter 7) which may limit conformational changes in the elbow region due to elbow bend differences; and (3) the residual was lower. A summary of the current status of this model is given in Table 13.

6.5 Discussion

The simulated annealing refinement method has shown itself to be excellent in refining initial models of the Gloop2 Fabs derived from the molecular replacement solutions. The comparative accuracy of the molecular replacement models is testified by the residuals at the end of a single cycle of energy minimisation refinement, and the addition of the molecular dynamics step gave the extra radius of convergence needed to refine the model further.

Figures 5–12 compare the models at stages in the refinement process with the current models, for all four cells described above. The comparison between the molecular replacement solutions and the current models indicate that the C_L:C_H1 was often better placed than the V_L:V_H. In all cases the rigid body refinements moved the model closer toward the current model. Ramachandran plots are shown in Figures 13–15 for the $P2_1$ 'A' and $P1$ structures. The occurrence of many non-glycine residues in 'forbidden' areas of $\phi\psi$ space for the $P2_1$ 'A' forms shows that there is still room for improvement in these models. The $P1$ structure fares better in this respect due to the higher resolution of the refinement and non-crystallographic symmetry restraints, but there are still some anomalies to be addressed. Comparison with existing Fab structures with better-behaved $\phi\psi$ values may assist with the resolution of these anomalies.

Although solvent peaks in the difference maps of the $P2_1$ 'A' and $P1$ crystal forms *are* apparent, there is little justification for placing them in the absence of higher resolution data. An improvement to the residual could be obtained by applying a solvent correction to the F_{calc} within X-PLOR since the lower resolution terms are clearly affected by solvent scattering (see Figures 1–3 in Chapter 7). The solvent correction implemented in X-PLOR subtracts electron density from

every point on the electron density map, setting all points whose resultant electron density is negative to zero (Brünger, 1988b). However this algorithm is currently inoperational (A. Brünger, pers. commun.) in v1.5. Instead the method of Phillips (1980) could be used, where all points outside the molecular boundary are set to the mean solvent level and all other points to zero. A set of $\mathbf{F}_{calc}$s due to bulk solvent scattering are calculated from this map. A high B factor is applied to the solvent density in order to smooth out the sharp edges in the map. Another approach would be to use atomic scattering factors modified for bulk solvent effects (Bryant *et al.*, 1985; Fraser *et al.*, 1978). The corrected atomic scattering factor $ff(\mathbf{s})$ is related to the conventional atomic scattering factor $f(s)$ by:

$$ff(\mathbf{s}) = f(\mathbf{s}) - v\rho[\exp(-4\pi v^{2/3}s^2)]$$

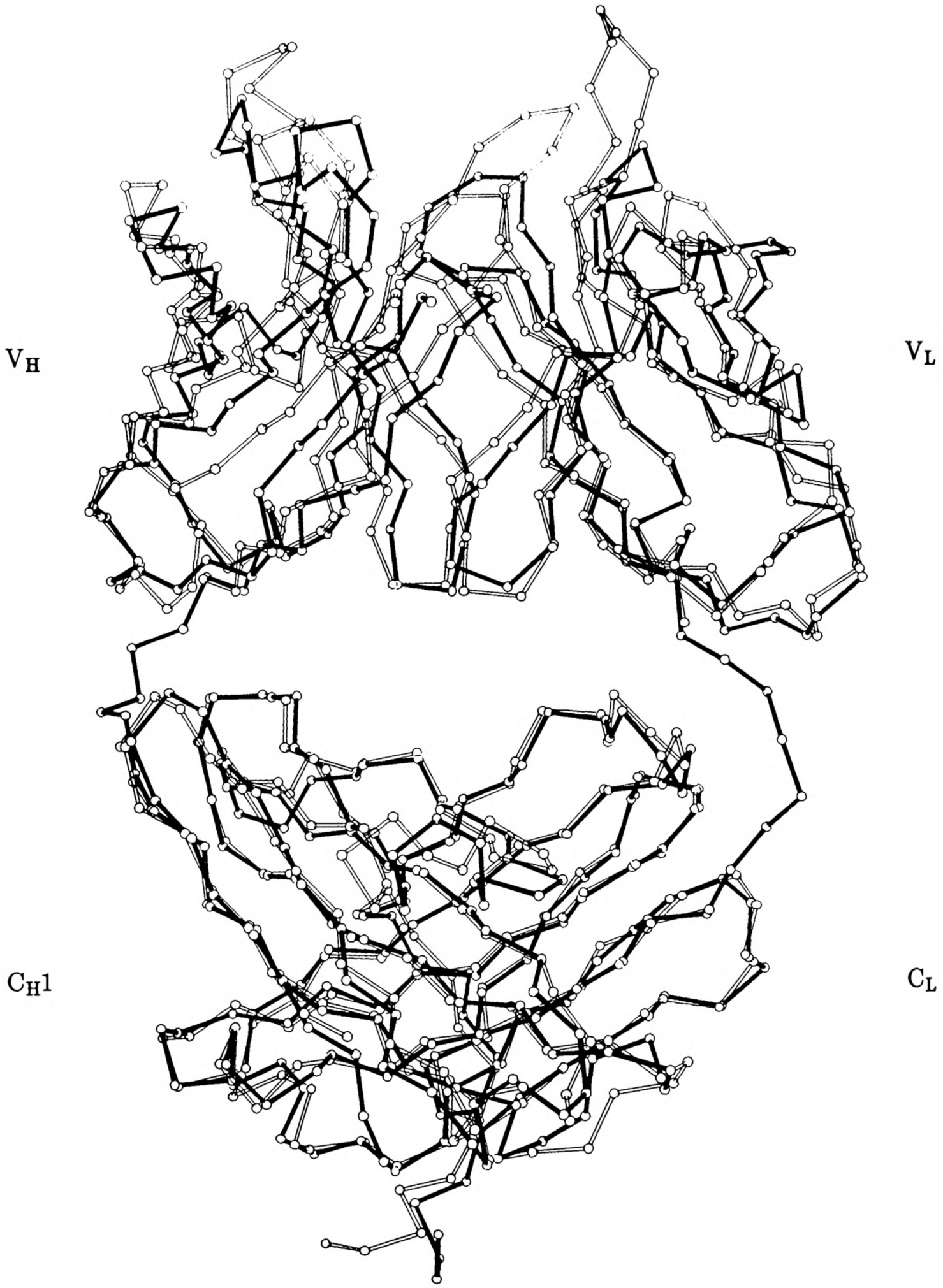
v is the mean volume of solvent displaced by a solvent atom and ρ is the mean electron density of the bulk solvent. This latter solvent correction takes no account of the shape of the bulk solvent volume within the crystal.

The molecular dynamics aspect of the refinement has been used more frequently in the refinement than could be justified purely from the results of it s application. In many cases, where the structure was mostly correct already, the molecular dynamics refinement step converged to essentially the same model as at the end of the first energy minimisation step. This in itself is reassuring, showing that the molecular dynamics conditions used in these refinements did not appear to distort the model. At no point during any of the dynamics simulations used did the residual exceed 50%, where it has been reported that the structure can degrade (Kuriyan *et al.*, 1989; Brünger *et al*, 1989).

The approach of turning off all electrostatic interactions during refinement is wasteful of the information on hydrogen bonding that would otherwise be included.

Kuriyan and coworkers (1989) have used a more subtle method, where the charges of Lys, Arg, Asp and Glu are set to zero may prove more effective. It is also clear that although molecular dynamics is a powerful tool in protein structure refinement it cannot correct all the errors in the structure where gross conformational changes are required (Taylor, 1989; Kuriyan *et al.*, 1989). This behaviour has been illustrated in these studies by the case of the incorrectly placed loop (64L–71L) in the $P2_1$ 'A' structure.

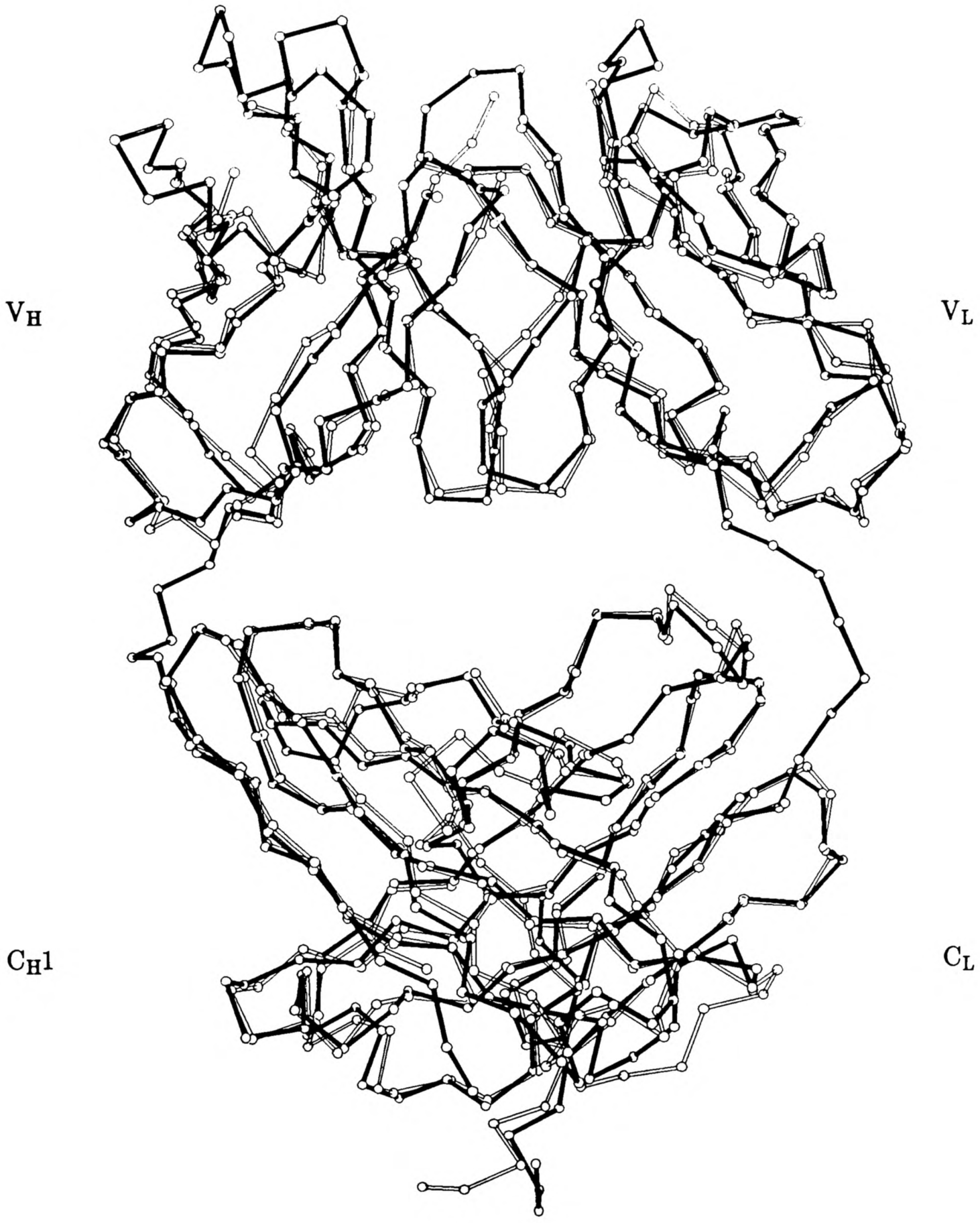
The refinement of the $P1$ form has not been handled in the most efficient manner possible. Averaging of the $P1$ structure was handled incautiously in the first instance, and most of the benefits of this process were discarded when the hybrid model was created. A better approach to utilising the non-crystallographic symmetry would be to apply strict symmetry restraints at the start of the refinement, and subsequently relaxing these during refinement. The under-determined nature of all these refinements could be largely eliminated by co-refining all the four models (two $P2_1$ 'A', $P1$ and $P2_1$ 'B') simultaneously with non-crystallographic symmetry restraints applied between the models in the different cells. This would require modification of the X-PLOR program. A related method has been used in the refinement of Ascorbate Oxidase from initial isomorphous replacement maps (Messerschmidt *et al.*, 1989).



Open Bonds — Molecular Replacement Model

Filled Bonds — Model after cycle 14 of X-PLOR

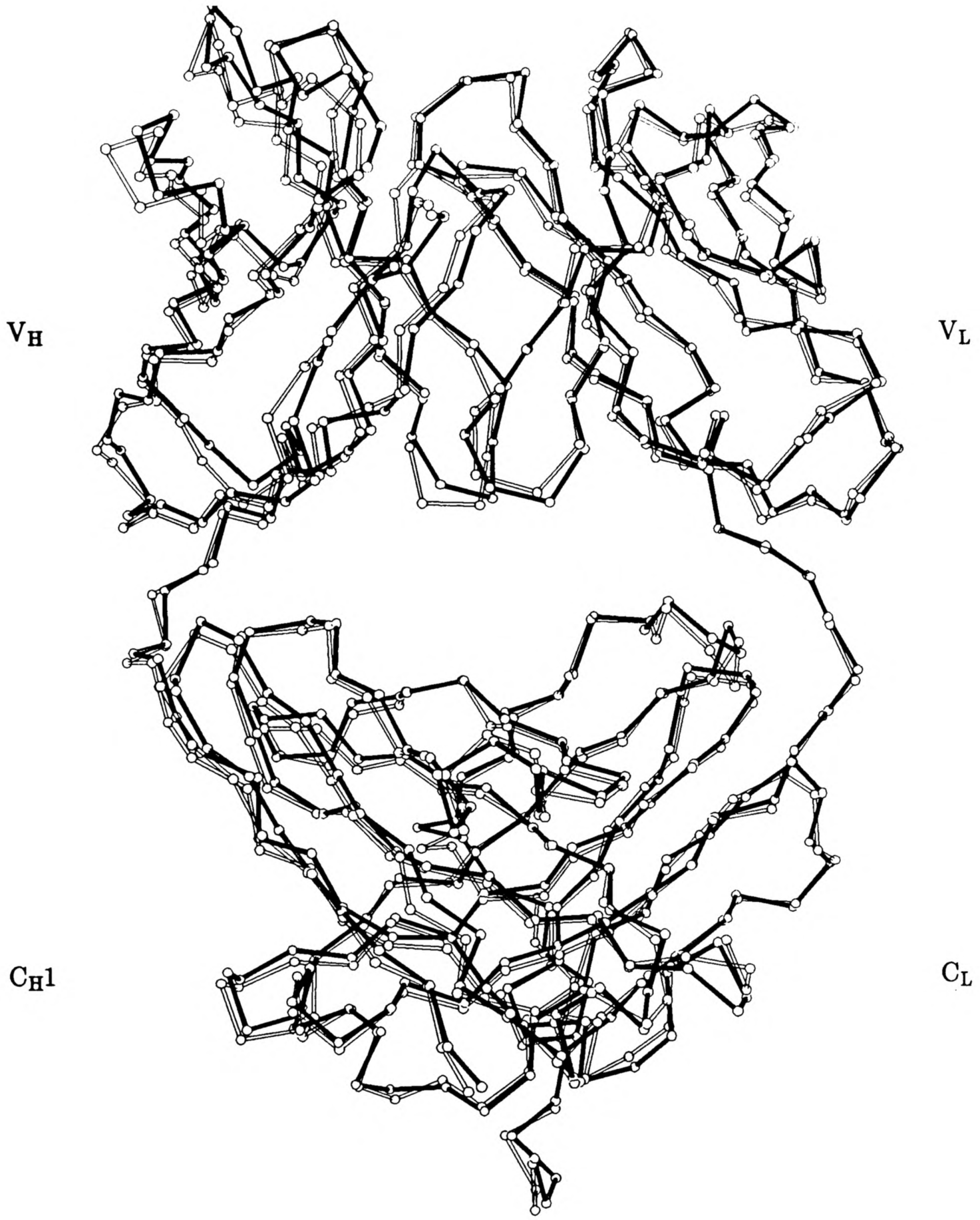
Figure 5: Comparison of Gloop2 $P2_1$ 'A' form (G2P2$_1$G Dataset) Models



Open Bonds — Model after Rigid Body Refinement (CORELS)

Filled Bonds — Model after cycle 14 of X-PLOR

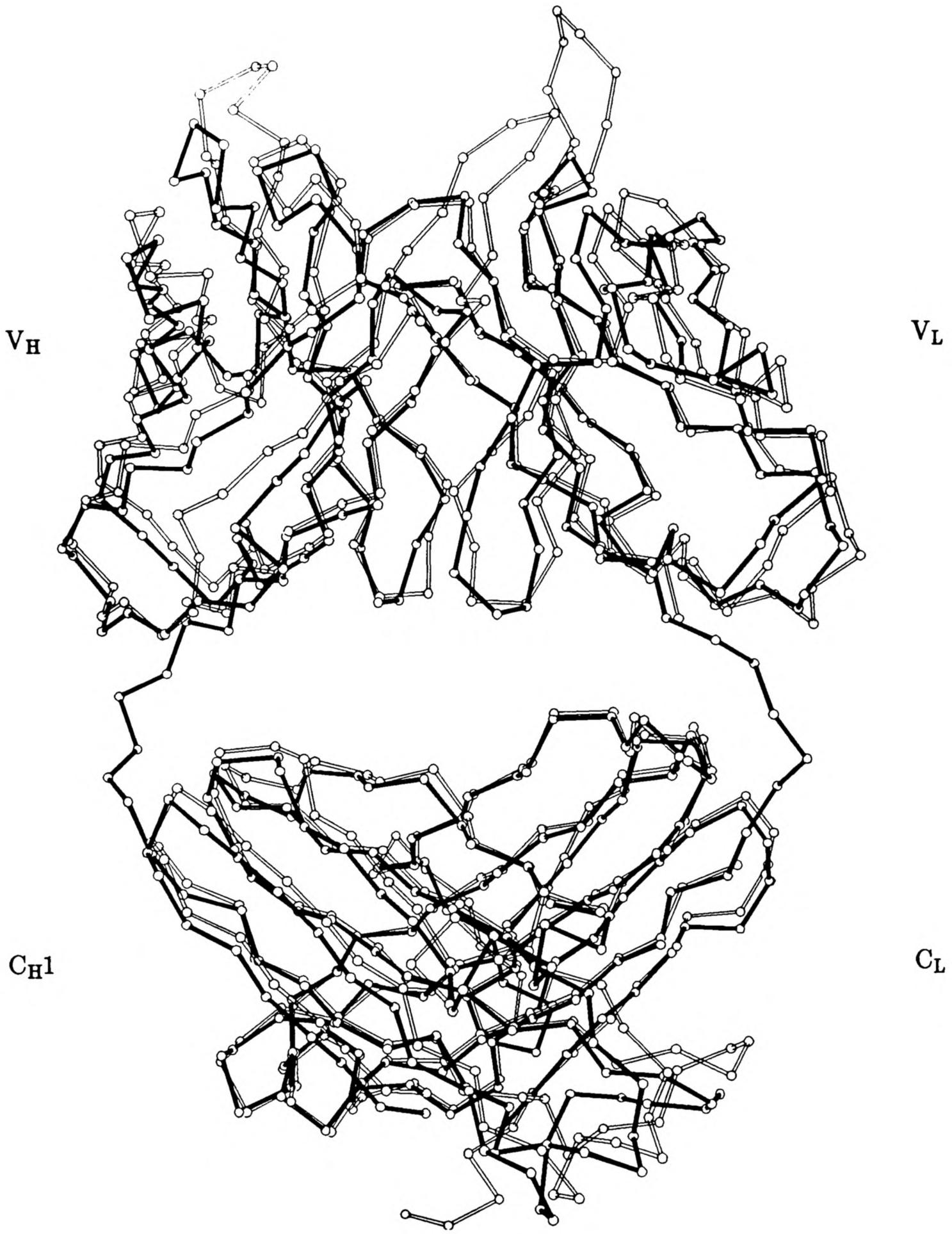
Figure 6: Comparison of Gloop2 $P2_1$ 'A' form (G2P2$_1$G Dataset) Models



Open Bonds — G2P2$_1$G Dataset Model

Filled Bonds — Model after one cycle of X-PLOR

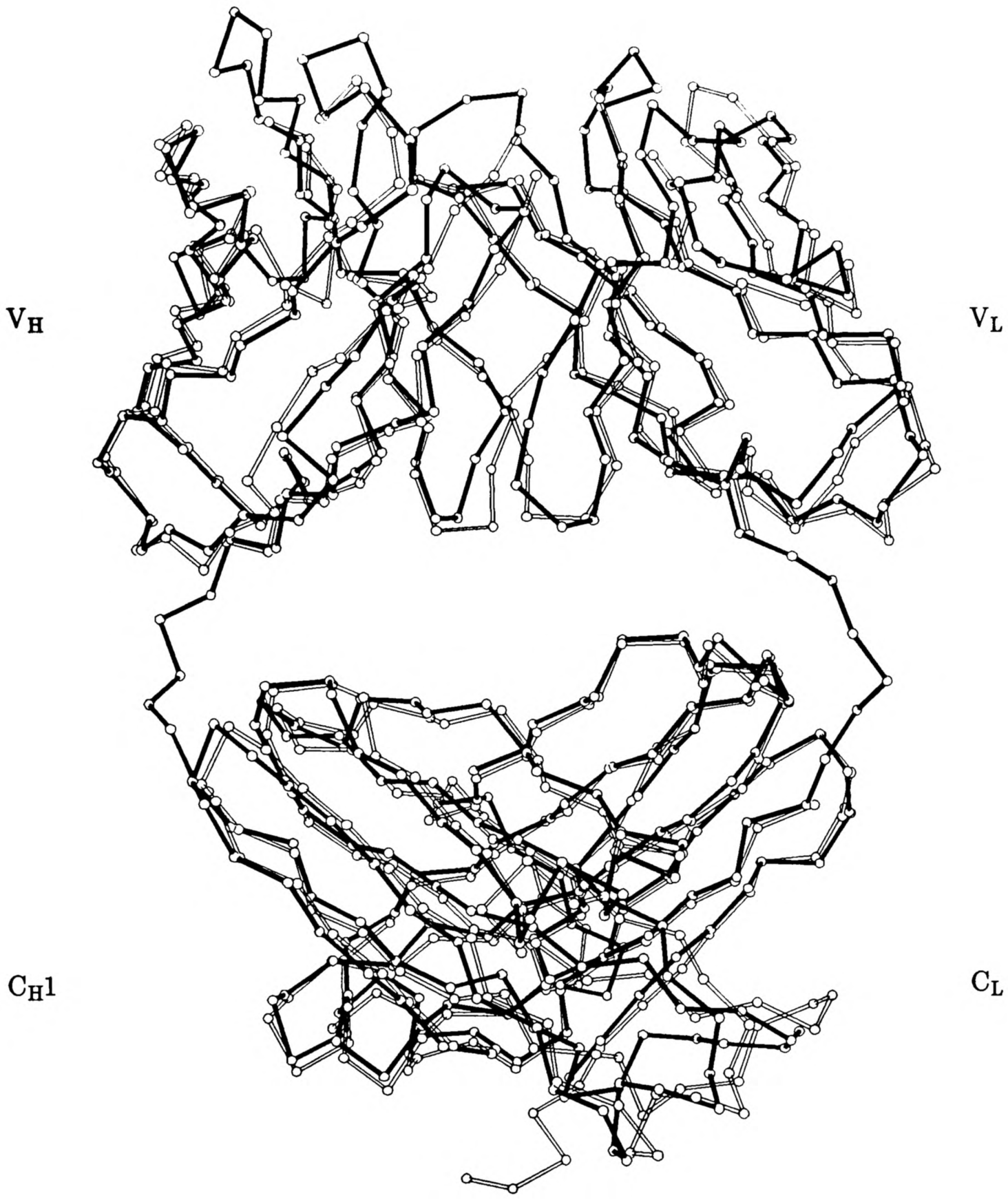
Figure 7: Comparison of Gloop2 $P2_1$ 'A' form (G2P2$_1$H Dataset) Models



Open Bonds — Molecular Replacement Model

Filled Bonds — Model after cycle 24 of X-PLOR

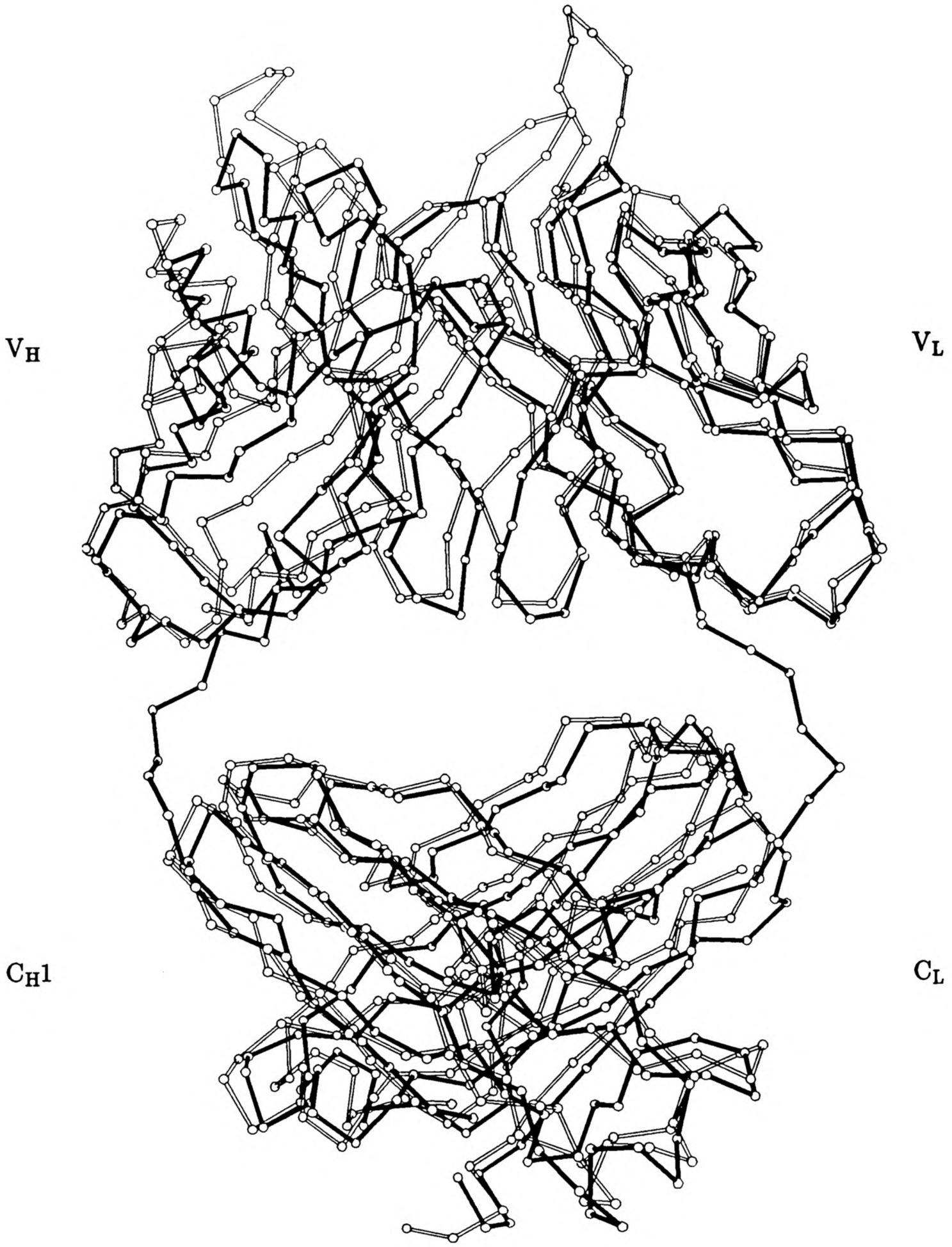
Figure 8: Comparison of Gloop2 $P1$ form (first Fab) Models



Open Bonds — Model after Rigid Body Refinement (CORELS)

Filled Bonds — Model after cycle 24 of X-PLOR

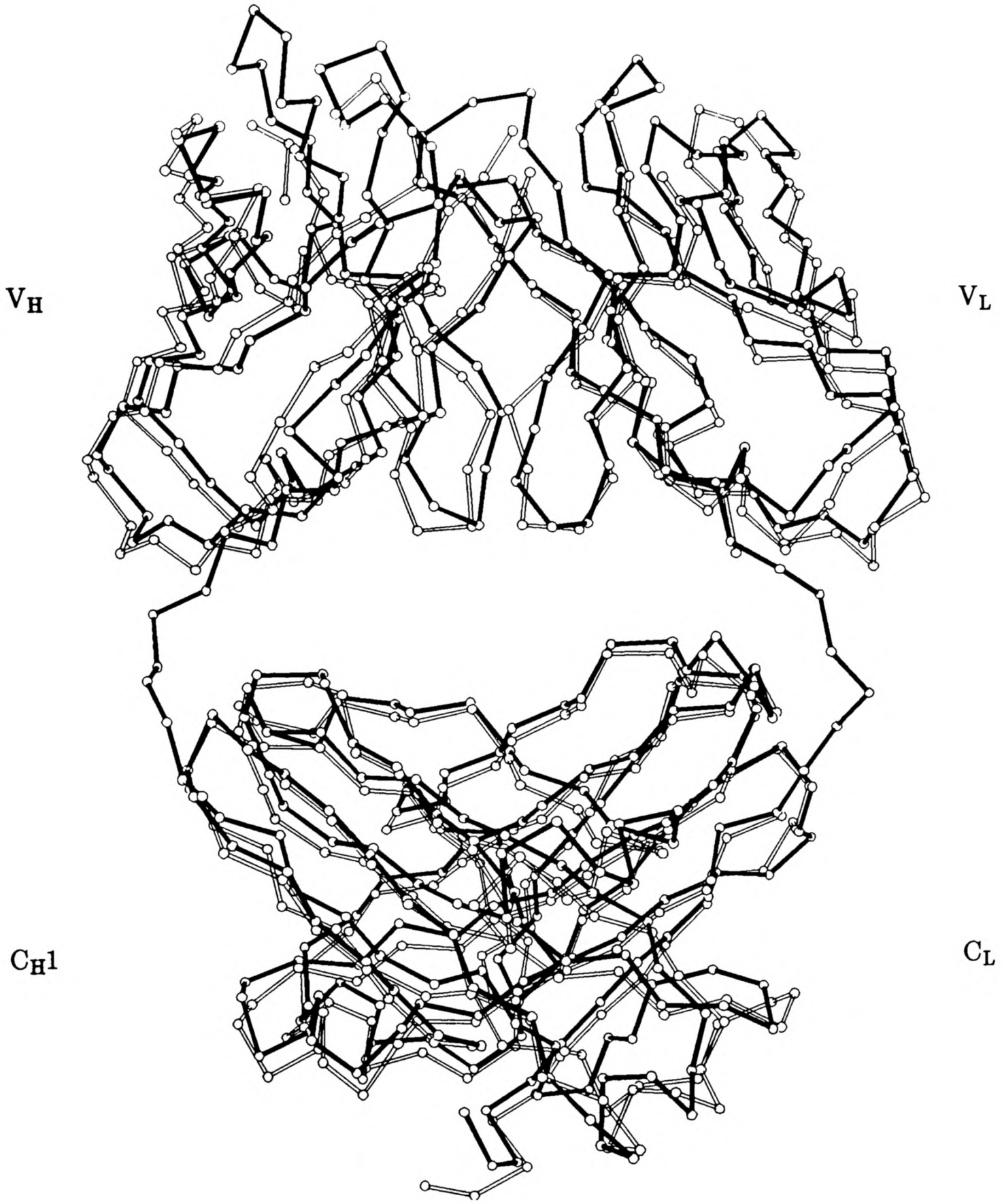
Figure 9: Comparison of Gloop2 $P1$ form (first Fab) Models



Open Bonds — Molecular Replacement Model

Filled Bonds — Model after cycle 24 of X-PLOR

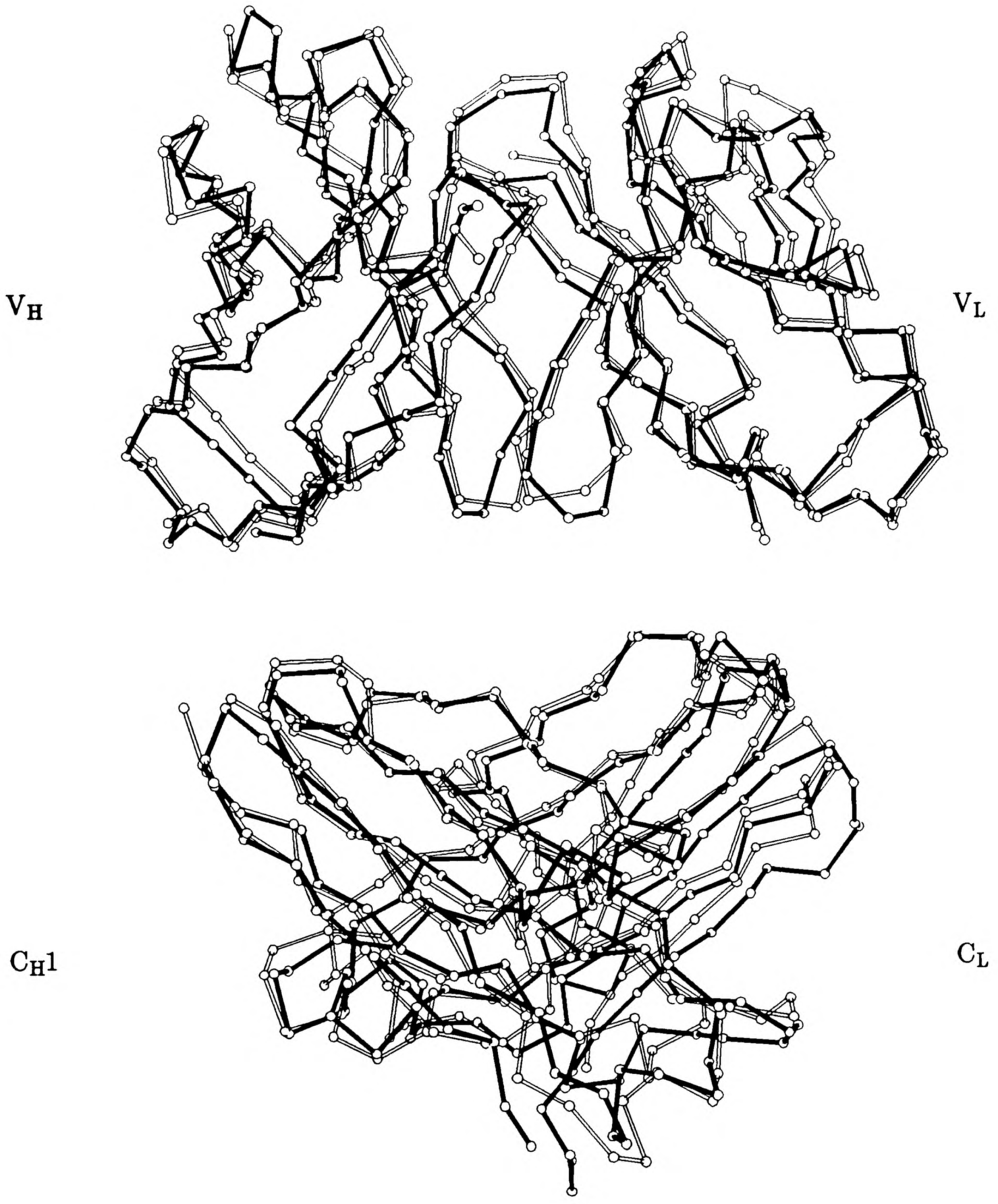
Figure 10: Comparison of Gloop2 $P1$ form (second Fab) Models



Open Bonds — Model after Rigid Body Refinement (CORELS)

Filled Bonds — Model after cycle 24 of X-PLOR

Figure 11: Comparison of Gloop2 $P1$ form (second Fab) Models



Open Bonds — Molecular Replacement Model

Filled Bonds — $P1$ Domain Model after X-PLOR

Figure 12: Comparison of Gloop2 $P2_1$ 'B' form Models

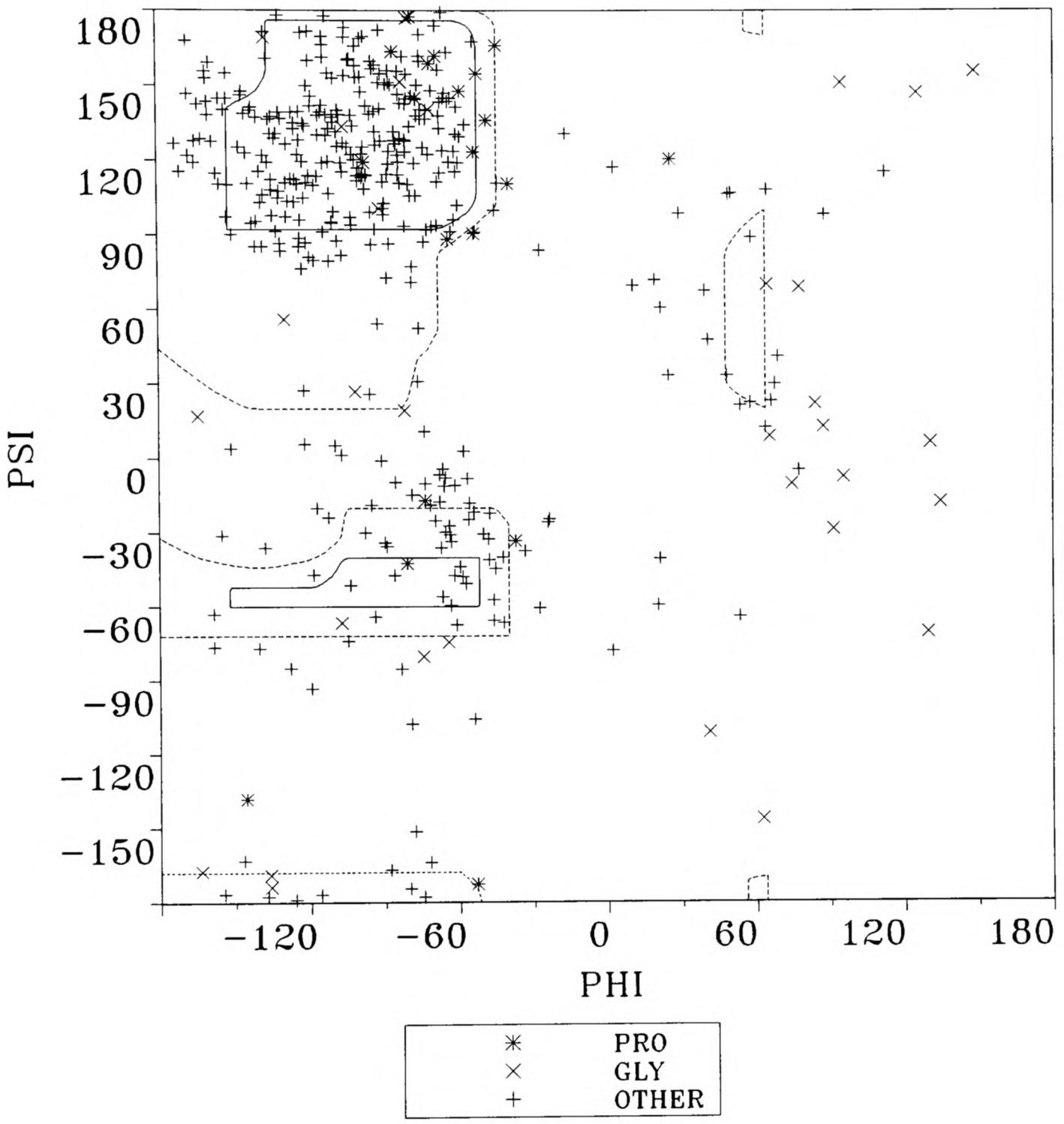


Figure 13: Ramachandran Plot of All Residues in the $P2_1$ 'A' form Fab, G2P2$_1$G Dataset

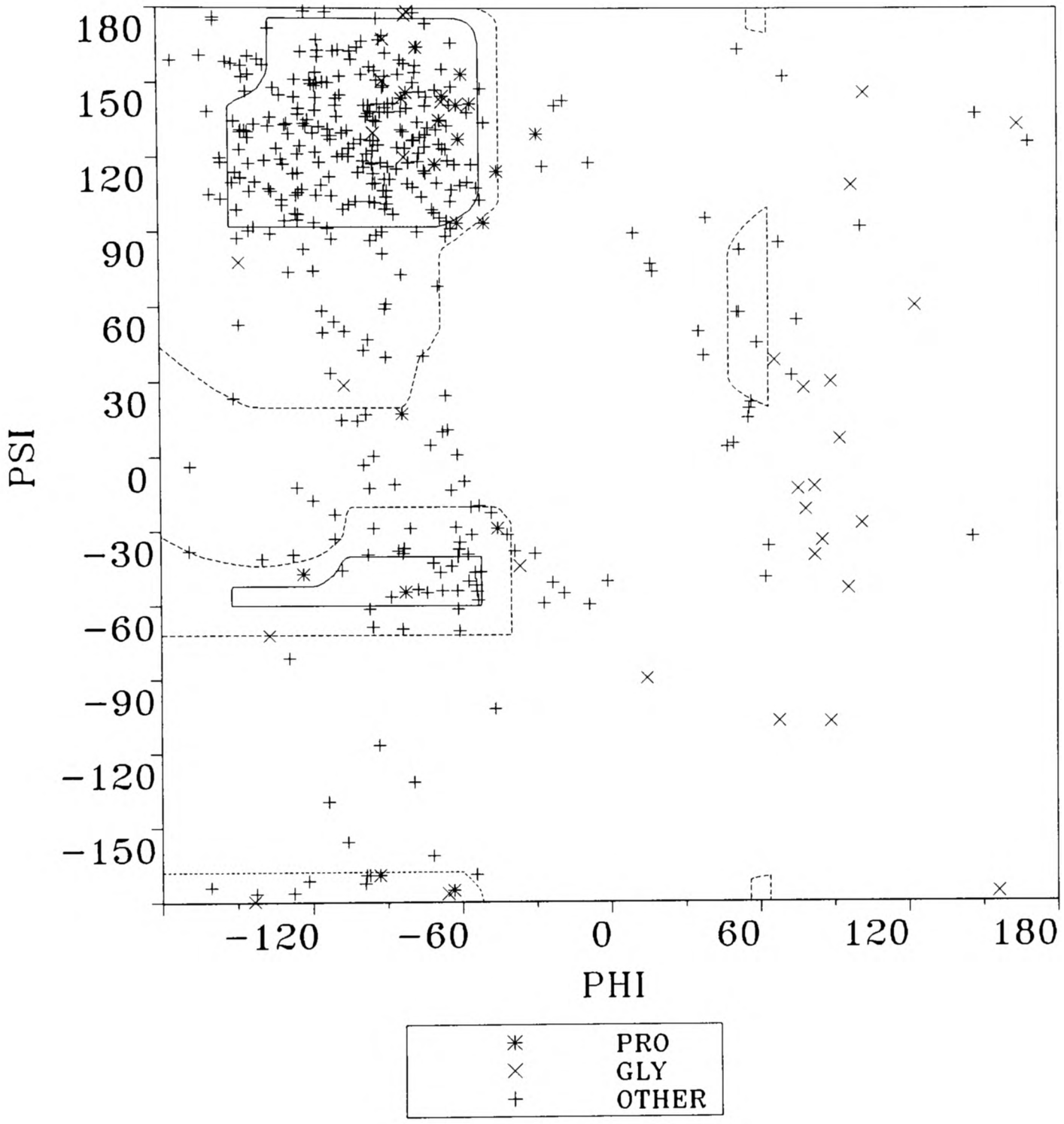


Figure 14: Ramachandran Plot of All Residues in the $P2_1$ 'A' form Fab, G2P2$_1$H Dataset

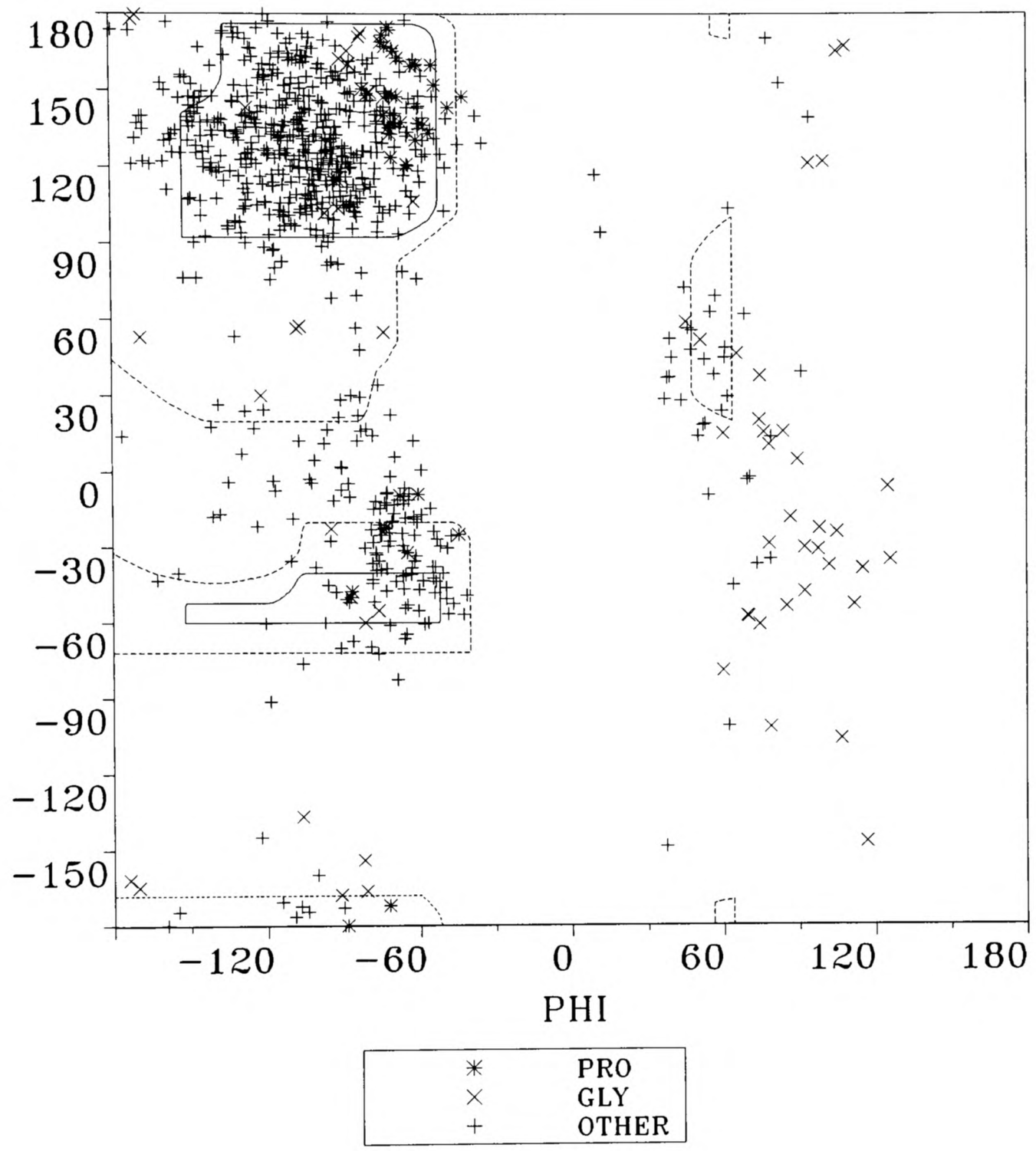


Figure 15: Ramachandran Plot of All Residues in the $P1$ form Fabs

CHAPTER 7

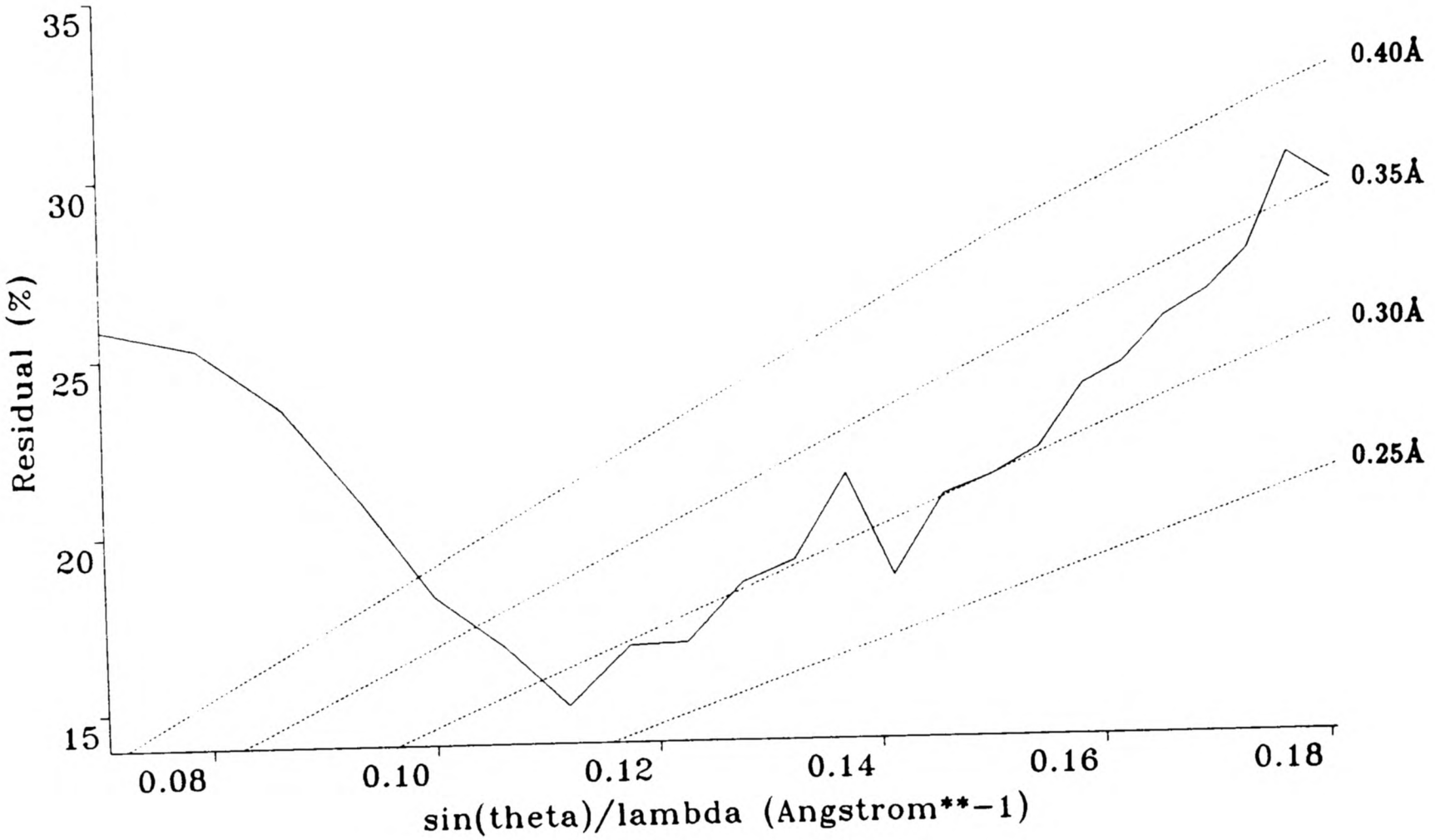
STRUCTURAL ANALYSIS

7.1 Estimation of Errors

Luzatti (1952) determined a relation between the residual as a function of $\sin\theta/\lambda$ and the mean coordinate error $\overline{\Delta r}$ in a structure. In deriving these relations the assumptions were made that the only contribution to the residual were the errors in position of the coordinates, and that these errors were normally distributed. This relation was used to estimate the mean coordinate error for the Gloop2 $P2_1$ 'A' and $P1$ cells.

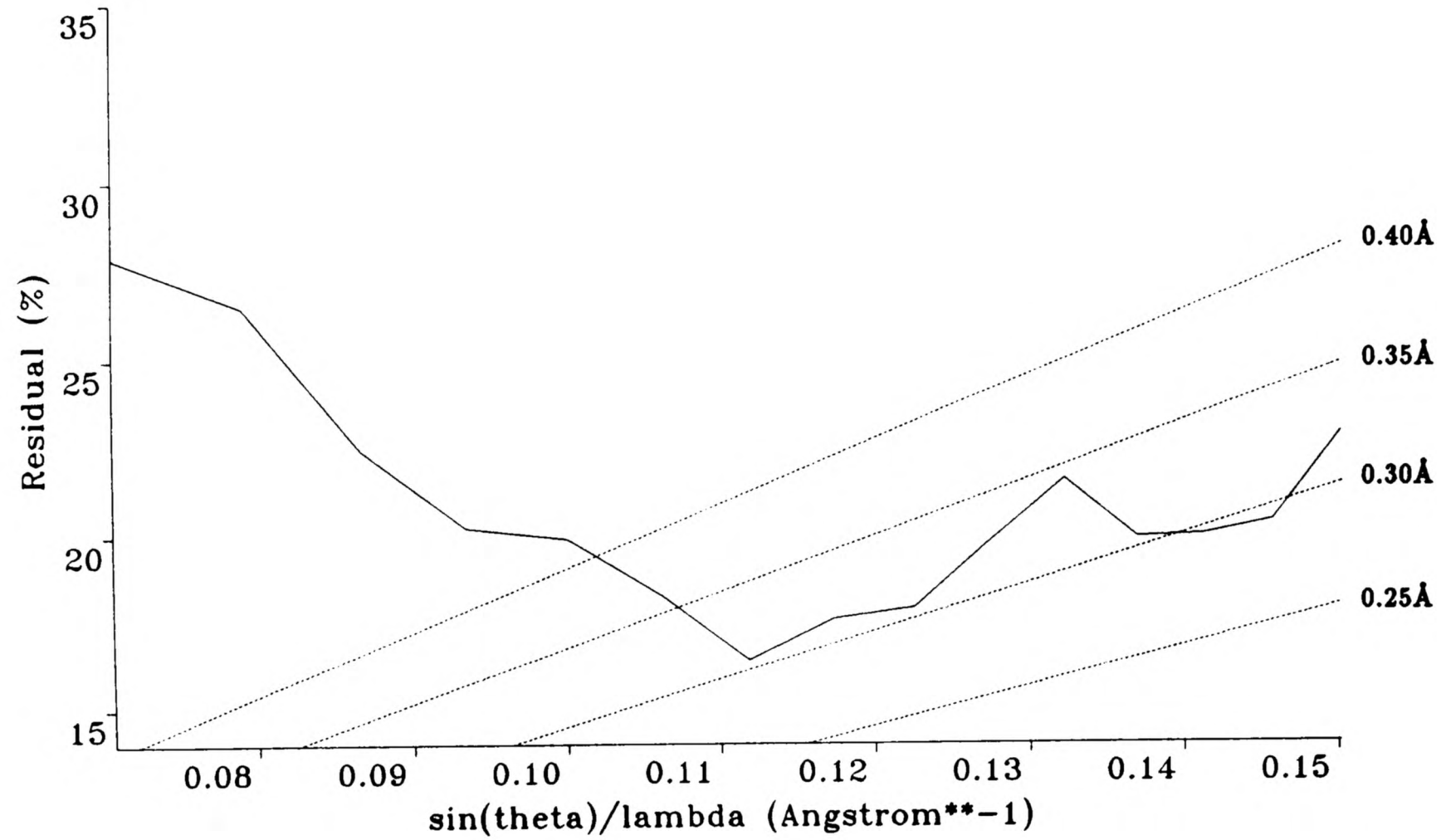
The residual $R = \sum |F_{obs} - F_{calc}| / \sum F_{obs}$ was calculated in shells of $\sin\theta/\lambda$ for all data with $F > 0$ for the $P1$ crystal form and the two $P2_1$ 'A' form cells. The residual was plotted against $\sin\theta/\lambda$ in Figures 1–3 along with the curves corresponding to the expected variation in R with $\sin\theta/\lambda$ for a selected range of $\overline{\Delta r}$ values. On the basis of these plots the mean coordinate error for the all three crystal forms was in the range 0.30–0.35Å. The increased trend in the $P1$ Luzatti plot (Figure 1) may be attributable to unreliability in the high resolution measurements. This value is comparable to the value (0.3Å) estimated for the refined structure of McPC603 at 2.7Å (Satow *et al.*, 1986) and for the HyHEL-5:HEL complex at 2.54Å (Sheriff *et al.*, 1987) for which an 0.40Å error was estimated. Both errors were estimated by the method of Luzatti (1952).

The Luzatti function is expected to indicate the upper bound of the mean error in the coordinates. Kuriyan *et al.* (1986) have demonstrated that limitations in the modelling of motion in proteins by isotropic B values will contribute significantly



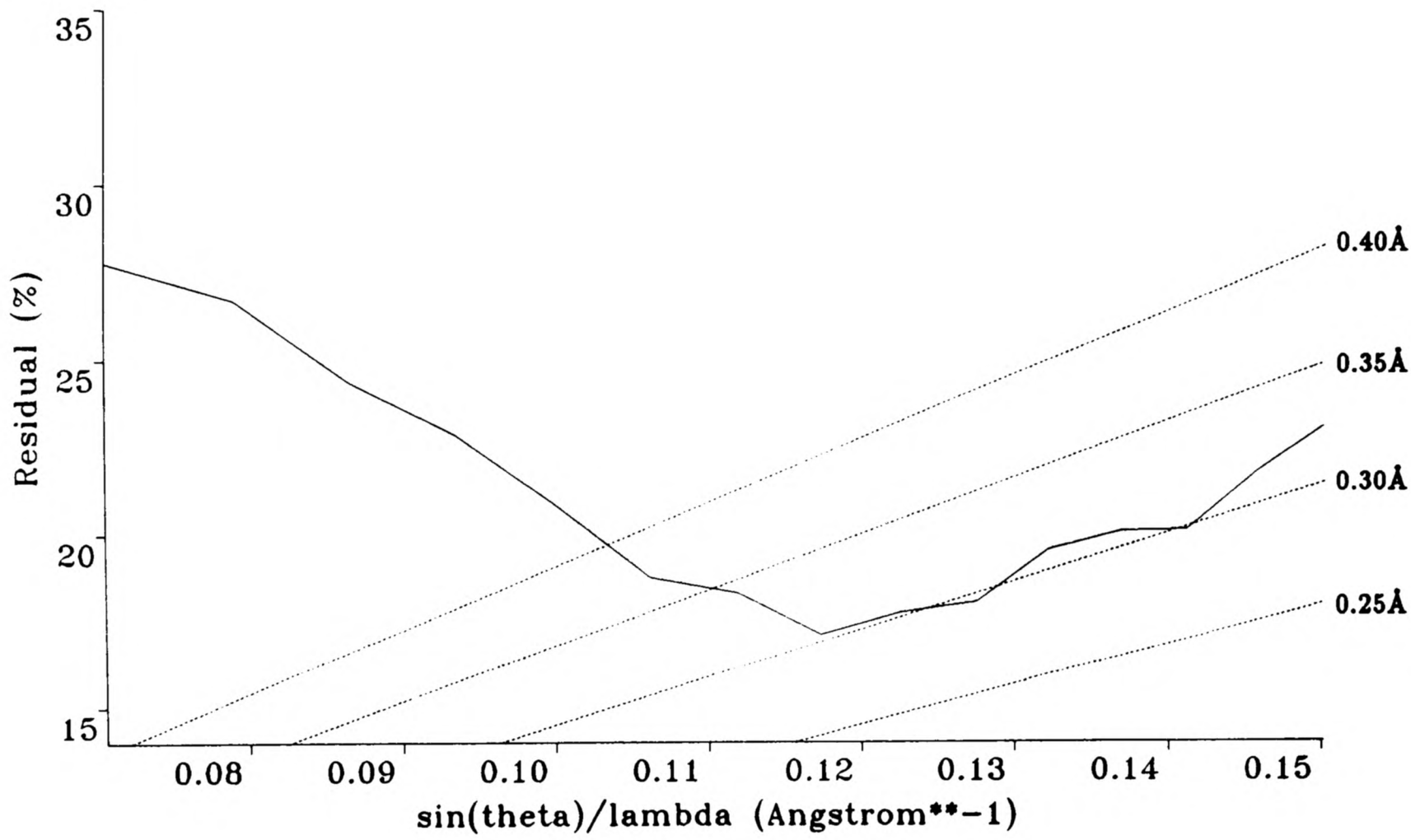
Solid line — residual calculated in shells of $\sin\theta/\lambda$ on all $F > 0$ for current model.
Broken lines — predicted residual as a function of $\sin\theta/\lambda$ for specified values of $\overline{\Delta r}$, from Luzatti (1952).
Residual $= \sum |F_{obs} - F_{calc}| / \sum F_{obs}$.

Figure 1: Luzatti Plot for the Gloop2 $P1$ Cell at 2.8Å



Solid line — residual calculated in shells of $\sin\theta/\lambda$ on all $F > 0$ for current model.
Broken lines — predicted residual as a function of $\sin\theta/\lambda$ for specified values of $\overline{\Delta r}$, from Luzatti (1952).
Residual $= \sum |F_{obs} - F_{calc}| / \sum F_{obs}$.

Figure 2: Luzatti Plot for the Gloop2 $P2_1$ 'A' Cell (G2P2$_1$G Dataset) at 3.3Å



Solid line — residual calculated in shells of $\sin\theta/\lambda$ on all $F > 0$ for current model.
Broken lines — predicted residual as a function of $\sin\theta/\lambda$ for specified values of $\overline{\Delta r}$, from Luzatti (1952).
Residual $= \sum |F_{obs} - F_{calc}| / \sum F_{obs}$.

Figure 3: Luzatti Plot for the Gloop2 $P2_1$ 'A' Cell (G2P2$_1$H Dataset) at 3.3Å

(a) Domain Definitions

Domain	Residues
V_L	1L–105L
C_L	111L–214L
V_H	1H–111H
C_H1	117H–212H
Elbow-L	106L–110L
Elbow-H	112H–116H

(b) CDR Definitions

CDR	Residues	Sequence
L1	24L–34L	R A S Q E I S G Y L S
L2	50L–56L	A A S T L D S
L3	89L–97L	L Q Y L S Y P L T
H1	31H–35H	T F G I T
H2	50H–65H	E I F P G N S K T Y Y A E R F K
H3	99H–102H	E I R Y

Chain numbering sequential from N-terminus.
Sequence given in the conventional single letter code.
Full Gloop2 sequence in Appendix 1.

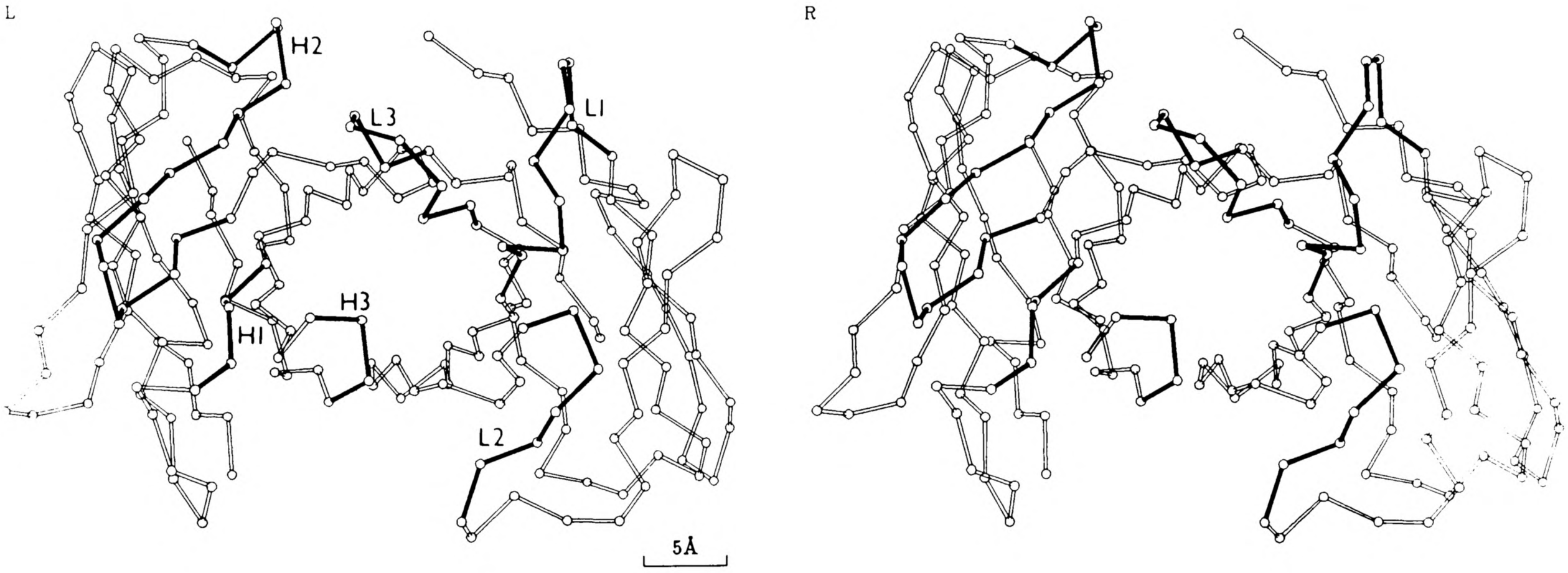
Table 1: Definition of Domains and CDRs

to the residual. An alternative method of estimating the errors in the Gloop2 structures, is by the comparison of the structures between the cells. This method is complicated by the influence of crystal contacts on deviations between the crystal structures. Nevertheless inspection of Figures 11–14 (deviations between the Gloop2 domains) indicate that the mean error level (ignoring the largest deviations) is less than 0.5Å. Similar plots comparing the P1a Fab with domains from Rei (Epp *et al.*, 1975) and HyHEL-5 (Sheriff *et al.*, 1987) show similar behaviour.

7.2 Gloop2 Fab Structures

The two $P2_1$ 'A' form cells, the $P2_1$ 'B' form cell and the $P1$ cell gave five crystallographically distinct structures of the Gloop2 Fab fragment. These five Fabs are subsequently referred to as Fabs $P2_1A1$ ($P2_1$ 'A' form, $G2P2_1G$ dataset), $P2_1A2$ ($P2_1$ 'A' form, $G2P2_1H$ dataset), $P2_1B$ ($P2_1$ 'B' form), P1a and P1b ($P1$ form). Four of these Fabs have been refined by molecular dynamics methods (Chapter 6) and it is possible to draw some conclusions about the variation in the peptide chain conformations between the four copies. The $P2_1B$ Fab has been refined as four rigid bodies (Chapter 6) and is only considered in terms of the domain associations in the following analyses.

All residues are sequentially numbered from the N-terminus of each chain with an H or L suffix for the heavy and light chains respectively. In cases where residues were omitted from the Fabs ($P1$ form, see Chapter 6), the numbering conforms to that for the intact chains. The residues associated with the four domains and the CDRs are listed in Table 1. The definitions of CDR lengths and positions are taken from Kabat *et al.* (1987) (rather than from Novotný *et al.* (1983) or Chothia and Lesk (1987)), although the numbering system used here is not the same as that of Kabat *et al.*



Open bonds — framework $C\alpha$s.
Solid bonds — CDR $C\alpha$s.
V_L and V_H domains looking down into the combining site from above.

Figure 4: Stereo View of the Gloop2 Combining Site

All atoms in the CDRs (defined in Table 1) are colour coded as follows:

Blue	— CDR L1
Red	— CDR L2
Green	— CDR L3
Purple	— CDR H1
Orange	— CDR H2
Yellow	— CDR H3

View down into the combining site from the direction of the antigen.

Figure 5: Space-filling Representation of the Gloop2 Combining Site

7.2.1 The Combining Site

Figure 4 shows a stereo perspective view of the antibody combining site (ACS) of the P1a Fab looking approximately down the inter-domain pseudo two-fold axis from the direction of the expected position of the antigen toward the V_L:V_H interface. The CDRs are shown as $C\alpha$ backbones in solid lines, the framework (i.e. non-CDR $C\alpha$s) in open lines. Only V_L and V_H domains are included in the figure.

From Figure 4 the approximate two-fold symmetry of the CDRs is immediately apparent. The CDR H3 is clearly very short (it is the shortest H3 of all solved Fab structures), consisting of only four residues. The shortness of CDR H3, and the fact that it packs down towards the bottom of the combining site, causes a truncated groove to be formed at the centre of the ACS. The groove runs over CDR H3 and is terminated by CDR L3. The groove is shown more clearly in the space-filling image in Figure 5 where the CDR atoms are colour-coded. The groove is ≈12Å long, ≈9Å wide and up to 7Å deep. It is lined predominantly by side-chains, although there are a few contributions from main-chain atoms. A list of the residues that line the groove is given in Table 2, and these are discussed in more detail below.

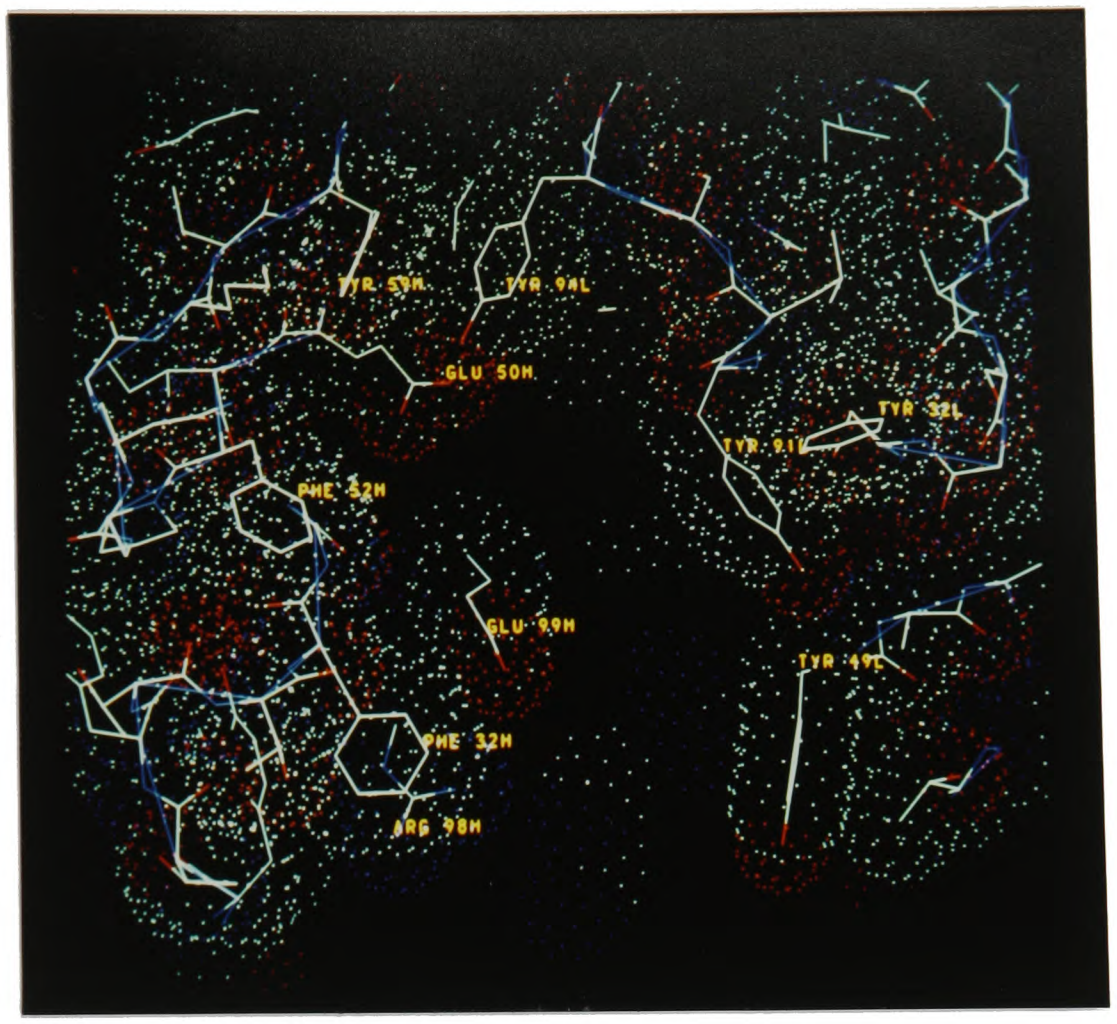
Four of the six CDRs are involved in the formation of the groove. CDR L1 is almost completely excluded from the groove by CDR L3 except for the side-chain of Tyr32L, whilst CDR L2 performs a secondary packing role. Four framework residues are involved in the groove — Tyr49L (the residue preceding CDR L2), Arg98H (the residue preceding CDR H3), Trp47H and Arg46L.

Figures 6 and 7 show the ACS in a little more detail and the most important residues are labelled. The V_L side of the groove is dominated by the side-chains of Tyr91L (CDR L3) and Tyr49L (V_L framework) with Tyr32L (CDR L1) packing against them. Ala50L from CDR L2 packs against the back of Tyr49L. Two carbonyl

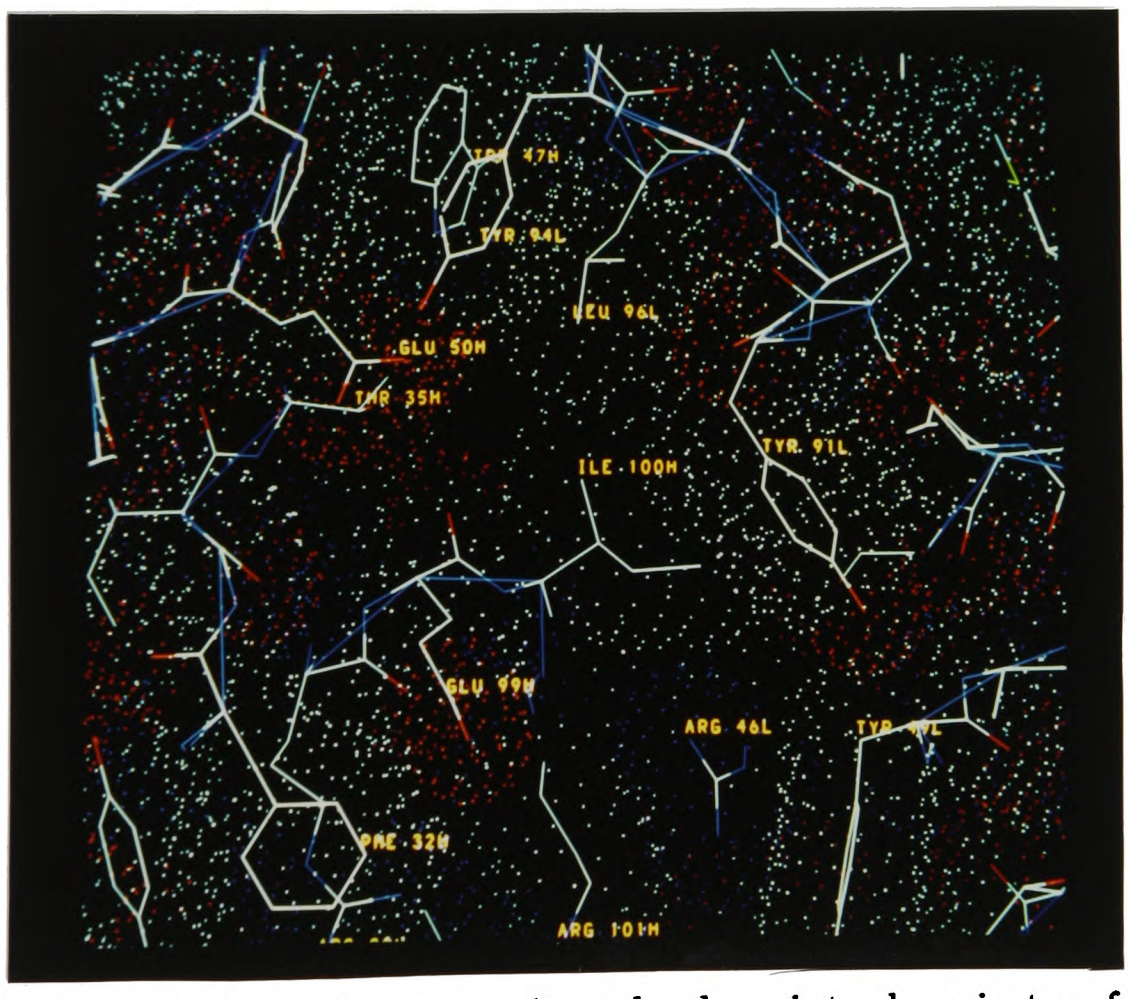
Segment	Residue		
Residues from CDRs			
L1	Tyr	32L	S/C
L2	Ala	50L	S/C
	Asp	55L	S/C
L3	Leu	89L	S/C
	Tyr	91L	S/C + M/C
	Leu	92L	M/C
	Tyr	94L	S/C + M/C
	Leu	96L	S/C
H1	Phe	32H	S/C
	Gly	33H	M/C
	Thr	35H	S/C
H2	Glu	50H	S/C
	Phe	52H	S/C
	Tyr	59H	S/C
H3	Glu	99H	S/C
	Ile	100H	S/C
	Arg	101H	S/C
Residues from Framework			
V_L	Arg	46L	S/C
	Tyr	49L	S/C
V_H	Trp	47H	S/C
	Arg	98H	S/C

S/C — Contribution from side-chain atoms.
M/C— Contribution from main-chain atoms.

Table 2: Residues Involved in the Gloop2 Combining Site Groove



(a) View of combining site approximately along interdomain two-fold, top of groove



(b) View of combining site approximately along interdomain two-fold, base of groove

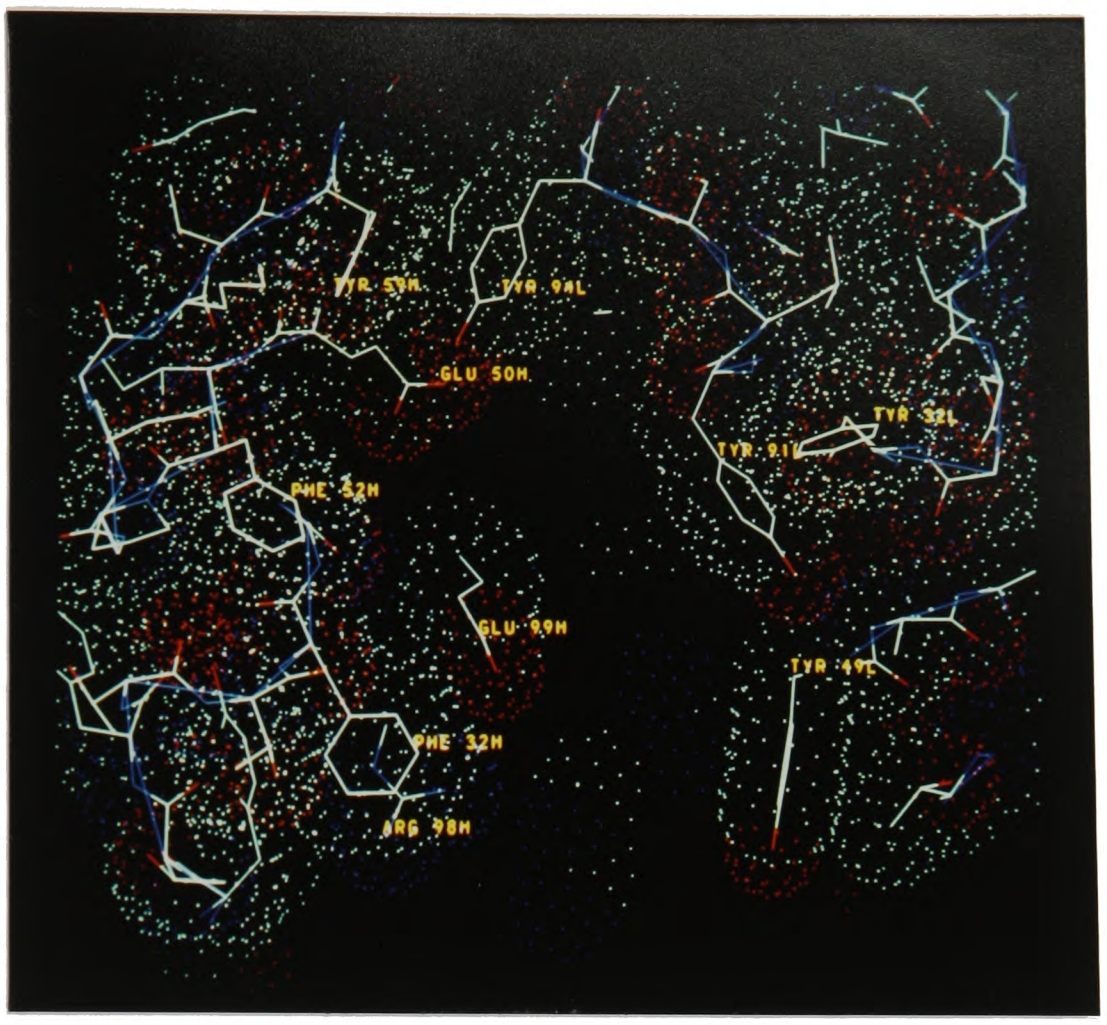
Van de Waals surface at 100% of the atomic radius.

Figure 6: Views of the Gloop2 Combining Site (P1a Fab)

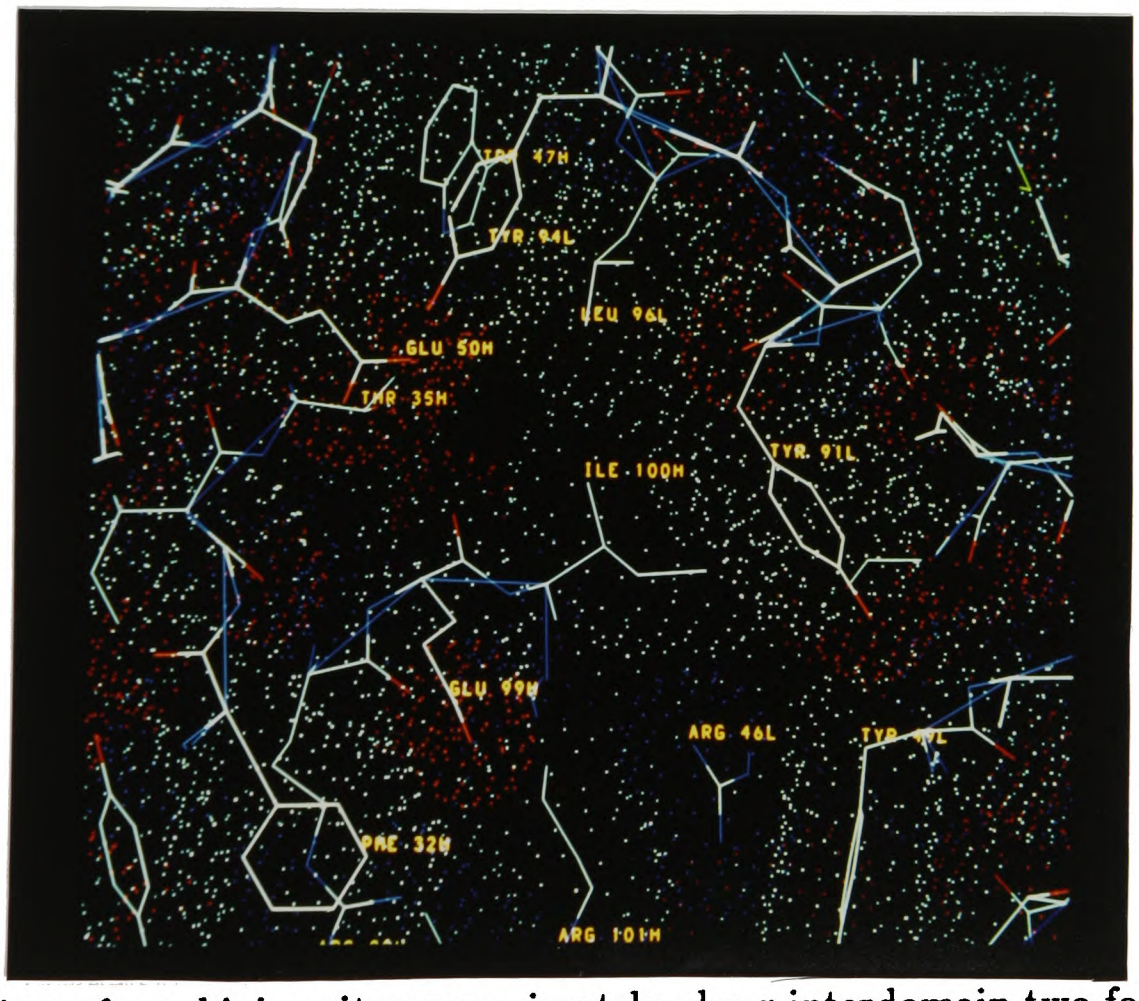
Segment	Residue		
Residues from CDRs			
L1	Tyr	32L	S/C
L2	Ala	50L	S/C
	Asp	55L	S/C
L3	Leu	89L	S/C
	Tyr	91L	S/C + M/C
	Leu	92L	M/C
	Tyr	94L	S/C + M/C
	Leu	96L	S/C
H1	Phe	32H	S/C
	Gly	33H	M/C
	Thr	35H	S/C
H2	Glu	50H	S/C
	Phe	52H	S/C
	Tyr	59H	S/C
H3	Glu	99H	S/C
	Ile	100H	S/C
	Arg	101H	S/C
Residues from Framework			
V_L	Arg	46L	S/C
	Tyr	49L	S/C
V_H	Trp	47H	S/C
	Arg	98H	S/C

S/C — Contribution from side-chain atoms.
M/C— Contribution from main-chain atoms.

Table 2: Residues Involved in the Gloop2 Combining Site Groove



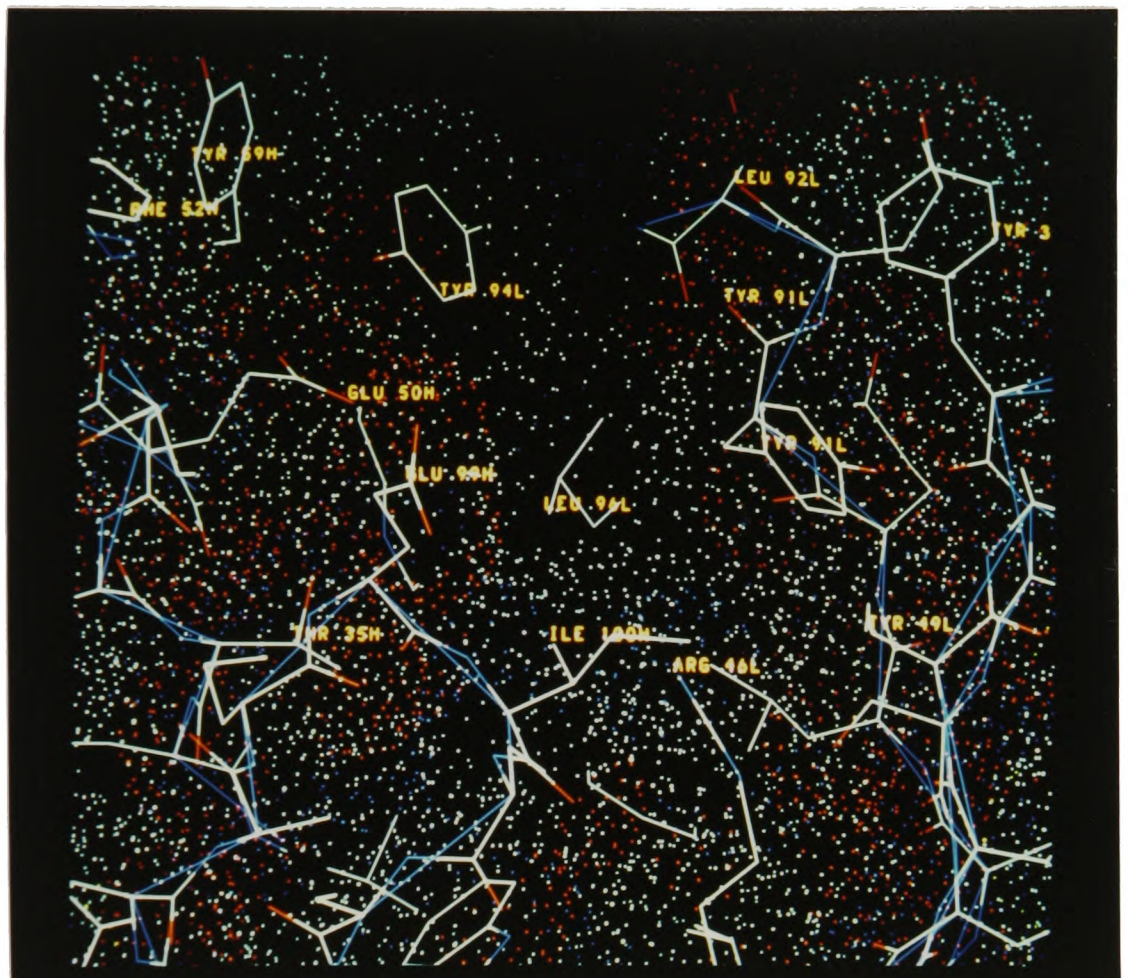
(a) View of combining site approximately along interdomain two-fold, top of groove



(b) View of combining site approximately along interdomain two-fold, base of groove

Van de Waals surface at 100% of the atomic radius.

Figure 6: Views of the Gloop2 Combining Site (P1a Fab)



View of combining site along the groove from H3 toward L3

Van de Waals surface at 100% atomic radius.

Figure 7: Views of the Gloop2 Combining Site (P1a Fab)

oxygens from CDR L3 — Tyr91L and Leu92L, also contribute to this side of the groove. The back of the groove is lined by the side-chains of Tyr94L and Leu96L and the amide nitrogen of Tyr96L, all from CDR L3. There is also a small contribution from the aromatic ring nitrogen of the side-chain of Trp47H (V_H framework). The V_H side of the groove is dominated by the two glutamic acid residues: Glu50H (CDR H2) and Glu99H (CDR H3). Three aromatic side-chains — Phe32H (CDR H1), Phe52H and Tyr59H (both CDR H2), the amide nitrogen from Gly33H (CDR H1) and the side-chain of Arg98H (V_H framework) make up the remaining residues on this side. The base of the groove is filled at the back by the side-chain of Ile100H (CDR H3) with minor contributions from Thr35H (CDR H1) and Leu89L (CDR L3). The base of the groove at the front is notable for a cluster of arginine side-chains: Arg46L (V_L framework) and Arg101H (CDR H3) pack close to each other and are partially compensated by interactions with the side-chain of Asp55L (CDR L2). Arg98H (V_H framework), also in this area, completes this cluster of positively-charged residues.

The concentration of aromatic side-chains lining the groove is quite clear in Figure 6, and agrees with the observation that aromatic side-chains (particularly tyrosine) are found at higher frequencies in the CDRs than would be expected from their frequencies in sequences from other proteins. To date there is no explanation for their frequent occurrence in the ACS.

The groove in Gloop2 is an extension of the hapten-binding pocket, and is caused by the conformation of CDR H3. The Fab HyHEL-5 (Sheriff *et al.*, 1987) has a small groove in this region, except that it is terminated at *both* ends — by CDRs L3 and H3. A fully-formed groove is present in the NC41:neuraminidase complex however (Tulip *et al.*, 1989), spanning the combining site, caused by CDRs L3 and H3 packing back against the V_L and V_H domains respectively. Thus the groove in

Gloop2 is not necessarily a reflection of it's peptide-binding function (Darsley and Rees, 1985a). The nature of Gloop2's interaction with peptide and HEL, and in particular the possible significance of the presence of Glu50H and Glu99H in the ACS is discussed in more detail in section 7.3.

7.2.2 Elbow Bends

The elbow bend of a Fab is defined as the relative inclination of the pseudo-diads relating V_L to V_H and C_L to C_H1. Colman *et al.* (1976) used five residues which were located in an approximate pairwise manner across the inter-domain two-fold axes, to define these axes. The definitions of the inter-domain pseudo-diads for the purposes of this work included (in principle) the $C\alpha$s of all residues in the domains with the exception of the CDRs.

The method used to obtain superimpositions of structurally-related domains is given here in some detail as it is also relevant to the work on domain pairing given in the following section. The program SHP (Appendix 2) was used in order to determine the optimum superimpositions of V_L onto V_H and C_L onto C_H1 based on $C\alpha$ atoms only. The 'probability of equivalence' (P_{IJ}) of the $C\alpha$ of residue I in molecule 1 and residue J in molecule 2 is given by:

$$P_{IJ} = \exp(-\frac{d_{IJ}^2}{E_1^2}) * \exp(\frac{S_{IJ}^2}{E_2^2})$$

(Rossmann and Argos, 1976), where d_{IJ}^2 is the square of the distance between I and J. S_{IJ}^2 is given by:

$$S_{IJ}^2 = |\mathbf{d}_{IJ} - \mathbf{d}_{I+1,J+1}|^2 - |\mathbf{d}_{IJ} - \mathbf{d}_{I-1,J-1}|^2$$

where $\mathbf{d}_{IJ}$ is the vector between the $C\alpha$s of residues I and J. The function S_{IJ}^2 is a measure of the similarity of the local conformation of the two molecules and is

Fab	Calculated Elbow Angle (°)	Reported Elbow Angle (°)	Reference
$P2_1A1$	151.9	—	
$P2_1A2$	152.0	—	
P1a	169.7	—	
P1b	166.6	—	
$P2_1B$	167.3	—	
HyHEL-5	161.1	161	Sheriff *et al.*, 1987
McPC603	131.2	133	Satow *et al.*, 1986
J539	143.1	145	Suh *et al.*, 1986
KOL	164.8	165	Matsushima *et al.*, 1978

Method of calculating elbow angles given in text.
Reported elbow angle taken from reference cited.

Table 3: Elbow Angles of the Gloop2 Fabs

independent of $|\mathbf{d}_{IJ}|$. The values of E_1 and E_2 provide the relative weighting of the two terms. SHP determines a set of equivalent residues by determining the best path through the two-dimensional matrix of all pairwise equivalences between the $C\alpha$ atoms in the two molecules (Stuart, 1979). Only this set of equivalent residues are used to determine the superimposition.

Superimpositions between the heavy and light chain domains were determined with $E_1 = 1.0$ and $E_2 = 10.0$. The correctness of the equivalences between $C\alpha$ atoms were checked by observing the relative positions of the structurally invariant cysteine residues in each domain. The matrix determined for the superimposition was deconvoluted into the rotation axis (of unit length) and rotation angle κ. The relative inclination of the axes (the elbow angle) was obtained from the dot product of the rotation axis vectors from V_L:V_H and C_L:C_H1. The elbow angles for the five Gloop2 Fabs are given in Table 3, as well as those for the five Fabs deposited in the Protein Databank (Bernstein *et al.*, 1977) evaluated by the same method. These angles correlate well with the reported elbow angles in the literature (also given in Table 3).

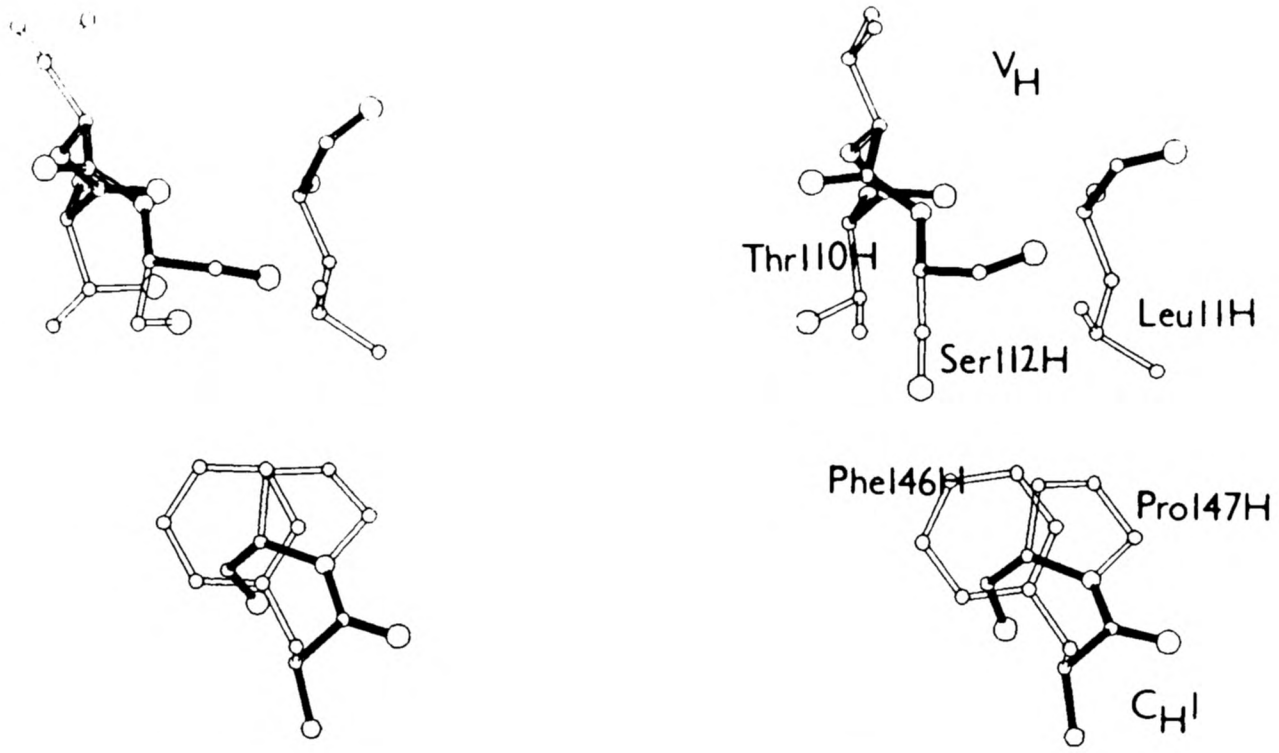
There is marked variation between the elbow angles of the various Gloop2 Fabs with a maximum difference of 15°. This agrees with the observations that the elbow angles of the HyHEL-5 and JEL42 Fabs can differ by 7° between crystallographically distinct Fabs (Sheriff *et al.*, 1987; Prasad *et al.*, 1988), and that the elbow angle changes by 8° between the structures of the Kol Fab and the intact IgG (Matsushima *et al.*, 1978). In what is the most dramatic example yet of elbow angle differences, the elbow angles of the D1.3 Fab in the uncomplexed (138°) and D1.3:E225 Fab complex (140°) were more than 30° different from the D1.3 elbow angle in the D1.3:HEL complex (Bentley *et al.*, 1989). Although the molecular replacement solution of the $P2_1$ 'A' form of Gloop2 suggested that the elbow angle was similar

to that of HyHEL-5 (Chapter 4) the refined structure of the $P2_1A1$ Fab has an elbow bend 9° different to that of HyHEL-5 (152° for $P2_1A1$ for Gloop2 and 161° for HyHEL-5; see Table 3).

Lesk and Chothia (1988) proposed a 'ball-and-socket' feature at the heavy chain elbow bend on the basis of sequence conservation and the structures of five Fabs from the Protein Databank (Bernstein *et al.*, 1977). This feature is also present in Gloop2 (Figure 8) involving the residues Leu11H, Thr110H, Ser112H, Phe146H and Pro147H. In this figure the residues are shown in the same orientation with respect to the V_H domain for each Fab. There is some variation of the exact orientation of the side-chains, which is consistent with the modest resolution of the structures, but the pattern observed by Lesk and Chothia (1988) was observed. With the more extended elbow angles (Fabs P1a and P1b) Phe146H interacts predominantly with Leu11H, as the elbow angle becomes more acute (Fabs $P2_1A1$ and $P2_1A2$) the Phe side-chain moves away from Leu11H and is partially replaced by the side-chain of Pro147H.

7.2.3 Domain Pairing

Considered initially to be a structural invariant (Amzel and Poljak, 1979), the issue of V_L:V_H and C_L:C_H1 pairing became of particular interest when Colman *et al.* (1987) reported an atypical V_L:V_H pairing in the NC41:neuraminidase complex and suggested that the pairing may have changed on complex formation. In the absence of the uncomplexed NC41 Fab this suggestion remains unproven, and Lascombe *et al.* (1989) have shown that the NC41 pairing is not as atypical as had originally been suggested. Two Bence-Jones proteins have been crystallised with abnormal V_L:V_L pairings: Rhe (Furey *et al.*, 1979) and Loc (Chang *et al.*, 1985). In the latter case the same protein crystallised from distilled water has a more typical V_L-



Residues viewed from same orientation with respect to the V_H domain in each case.

Figure 8: The 'Ball and Socket' Residues in the Gloop2 Elbow Bend

V_L pairing (Schiffer *et al.*, 1989). Since the V_L:V_L pairs are different in detail from V_L:V_H pairs it is not certain whether these observed conformational changes are not limited to Bence-Jones dimers. Conformational change in the Fv has been proposed to explain changes in idiotopes and paratopes on binding antigen or anti-idiotypic antibodies (Stevens *et al.*, 1988).

In this section the variability of V_L-V_H and C_L-C_H1 interactions are analysed. Two methods were used to parameterise the pairing interactions. The first method was that of Cygler and Anderson (1988a), where the rotation component of the superimposition was deconvoluted into a rotation (κ) about an axis, and the translation expressed as the component of the translation vector along the rotation axis. The second method was taken from Lascombe *et al.* (1989), where the differences in domain pairing were expressed as the distance and rotation required to superimpose one heavy chain domain onto another.

In this analysis the aim was to determine the parameters of *domain pairing* rather than the inter-domain contacts. Consequently the superimpositions between domains were made on the basis of all 'equivalent' residues in the domains, with the exception of the CDRs in the V_L and V_H domains, which were omitted from the analysis since they are subject to greater variability (in structure and sequence) than any other part of the molecule. The basis of 'equivalence' was described in the previous section. Equivalent $C\alpha$ atoms between domains were obtained using SHP (Appendix 2) and the final superimposition was performed using the program ASH (Appendix 2) on the basis of these equivalenced $C\alpha$s. This semi-automatic procedure enabled loops which had different conformations between the compared domains to be omitted from the comparison, and insertions and deletions in the sequence to be allowed for (for instance residue 10L is deleted in V_λ domains, Kabat *et al.* (1987)). The details of the application of the two methods are described below.

Fab	SUPERIMPOSITION									
	C_H1 onto C_L					V_H onto V_L				
	κ (°)	T_a (Å)	Δ (°)	RMS (Å)	No. of $C\alpha$	κ (°)	T_a (Å)	Δ (°)	RMS (Å)	No. of $C\alpha$
Gloop2 P1a	168.8	1.74	1.82	0.99	64	176.7	0.64	2.42	0.85	70
Gloop2 P1b	168.8	1.47	1.90	1.01	65	176.9	0.72	3.21	0.88	68
Gloop2 $P2_1A1$	167.0	1.73	1.11	1.06	70	175.5	0.88	3.23	1.05	67
Gloop2 $P2_1A2$	165.2	1.77	0.63	1.16	70	175.2	0.86	3.26	0.97	67
Gloop2 $P2_1B$	166.5	1.98	3.00	0.93	64	177.4	0.79	3.09	1.01	71
HyHEL-5	168.1	1.83	0.00	0.99	76	171.4	−0.19	0.00	0.83	69
McPC603	171.4	2.86	4.19	1.14	61	173.4	0.14	2.90	1.02	67
J539	173.8	2.13	2.94	1.03	69	168.1	−0.07	4.62	0.89	66
Kol	169.7	3.60	6.44	1.08	73	164.3	0.03	1.51	1.11	61
New	171.7	3.45	4.85	1.00	74	167.5	−0.30	2.25	1.09	63

κ = rotation about axis.
T_a = translation along axis.
Δ = inclination of rotation axis to that of HyHEL-5 when V_L/C_L domains are superimposed.
No. of $C\alpha$ = number of equivalenced $C\alpha$ atoms between domains.
RMS on equivalenced $C\alpha$ atoms.
See text for full details of superimposition (method 1).

Table 4: Domain Pairing Parameters for Fab Structures

FIRST METHOD: The superimposition of V_H or C_H1 onto V_L or C_L was determined using SHP and ASH, as described above. The rotation matrix was deconvoluted into the rotation κ about a rotation axis, and the translation (T) was expressed as the component of the translation vector along this rotation axis (T_a). The positive direction of the axis was chosen to run from the C-terminal ends of the V_L:V_H and C_L:C_H1 toward the N-terminal end. The sign of the axis-component of T was preserved to indicate the direction of the screw element to the axis. A positive rotation angle κ was defined as a clockwise rotation when viewed from the origin out along the positive axis direction (same convention as used in Chapter 4). The lengths and positions of the domains used in the comparison is as given in Table 4, Chapter 5 (Gloop2 domain definitions in Table 1, this chapter). The light chain domains of the domain pairs of each model were superimposed (using SHP) onto the light chain domain of HyHEL-5 (Sheriff *et al.*, 1987). The rotation axes of the HyHEL-5 V_L-V_H and the C_L-C_H1 domain pairs were taken as the standard to which all other axes were compared. The angle of inclination (Δ) of each rotation axis with the HyHEL-5 standard was determined. The set of parameters derived by this method were the rotation angle κ, the axial translation T_a and the inclination of the axis Δ to that of HyHEL-5 after superimposition of V_L or C_L domains. These parameters are presented in Table 4 along with the number of equivalences between the heavy and light chain domains and the RMS fit on the $C\alpha$ atoms based on these equivalences.

SECOND METHOD: The light chain domains were superimposed on the corresponding light chain domain of HyHEL-5, as with the first method. All possible pairwise superimpositions were analysed. Superimpositions between the heavy chain domains were obtained by the combination of SHP and ASH, as for the first method. The differences in domain pairings were given by two parameters: (a) the

	P1a	P1b	$P2_1A1$	$P2_1A2$	$P2_1B$	HyHEL-5	McPC603	J539	Kol	New
P1a	–	3.03	2.99	2.66	1.85	6.42	5.08	12.25	11.10	13.54
P1b	0.34	–	3.55	3.28	1.86	7.76	3.36	10.54	10.56	14.00
$P2_1A1$	0.66	1.09	–	0.57	3.33	6.80	4.32	11.67	10.64	13.20
$P2_1A2$	0.73	0.84	0.33	–	3.00	6.94	4.20	11.88	11.03	13.18
$P2_1B$	0.53	0.24	1.02	0.83	–	7.93	4.96	12.28	11.60	14.54
HyHEL-5	2.04	2.18	2.64	2.53	2.29	–	6.15	10.23	6.85	7.88
McPC603	1.69	1.89	1.95	1.89	1.95	0.99	–	7.91	7.24	10.53
J539	2.58	2.73	2.84	2.89	2.81	0.85	1.05	–	4.14	7.29
Kol	2.57	2.77	2.67	2.74	2.94	1.35	1.21	1.10	–	3.88
New	2.81	3.03	2.78	2.78	3.13	1.65	1.37	1.20	0.40	–

Upper right trangle — Rotation angle κ (°) required to superimpose the domains.
Lower left trangle — Distance moved in order to superimpose the domains (no rotation).
See text for full details of superimposition (method 2)

Table 5: Domain Pairing In V_L:V_H Domains

length of the translation vector which optimally superimposed the two heavy chain domains without rotation; and (b) the rotation required to optimally superimpose the two domains (translation applied in the determination of this rotation but ignored since it is a function of the position of the coordinate origin). The results of these superimpositions are given in Tables 5 and 6.

The results of both methods compare qualitatively with the analyses of Cygler and Anderson (1988a) and Lascombe *et al.* (1989). The values obtained are not the same due to the differences in the definition of the superimpositions. The results from the first method show that C_L and C_H1 domains have similar orientations but the translational element to the screw axis is subject to considerable variation. The V_L and V_H domains show variability in both orientation *and* axis translation. The direction of the rotation axis was more constant with respect to the light chain domain in C_L:C_H1 than in V_L:V_H. The second method shows domain pairing is at least partially a function of domain types: the IgG-κ, IgA-κ and IgG-λ Fabs show similarity within each type and often greater differences between types. However the Fv of McPC603 (Satow *et al.*, 1986) has a greater similarity with the Gloop2 Fv's than HyHEL-5 (Sheriff *et al.*, 1987) has with Gloop2, which provides an explanation as to why the McPC603 Fv was a better model for the molecular replacement solution of Gloop2 (a combination of sequence homology and similar quaternary structure, see Chapters 5 and 6). Similar dependence on the domain pairing was suggested for the differences in the behaviour of the Fv of Kol and McPC603 in the cross-rotation function with the HED10 structure (Cygler and Anderson, 1988a).

It is also apparent that the different domain pairs of the *same* protein have small differences. This has not been noted previously for other Fab structures. The largest difference is between the Fv of the P1b Fab and the Fv of the $P2_1A1$ Fab. A $C\alpha$ trace is shown in Figure 9, where the V_L domains have been superimposed

	P1a	P1b	$P2_1A1$	$P2_1A2$	$P2_1B$	HyHEL-5	McPC603	J539	Kol	New
P1a	–	1.37	4.20	2.49	3.48	1.40	8.71	8.79	11.27	11.43
P1b	0.31	–	4.45	4.20	4.20	2.09	7.56	8.67	12.39	12.81
$P2_1A1$	0.62	0.55	–	2.03	5.71	2.07	5.86	6.02	12.09	11.10
$P2_1A2$	0.37	0.54	0.48	–	4.51	1.58	6.71	7.95	11.34	10.43
$P2_1B$	0.81	1.05	1.07	0.56	–	3.45	10.81	11.24	13.86	14.33
HyHEL-5	0.61	0.62	0.13	0.46	1.06	–	8.08	8.18	11.09	10.09
McPC603	0.81	0.51	0.56	0.89	1.31	0.72	–	2.51	13.66	9.95
J539	1.02	0.86	0.81	1.14	1.71	0.79	0.44	–	14.50	11.24
Kol	1.76	1.95	1.89	1.62	0.93	2.21	2.34	2.74	–	4.18
New	1.18	1.44	1.39	1.18	0.28	1.60	1.86	2.29	0.57	–

Upper right trangle — Rotation angle κ (°) required to superimpose the domains.
Lower left trangle — Distance moved in order to superimpose the domains (no rotation).
See text for full details of superimposition (method 2)

Table 6: Domain Pairing in C_L:C_H1 Domains

(a) View approximately down the interdomain two-fold axis

(b) View apporoximately normal to the interdomain axis

V_L:V_H domains from Fab P1b shown in red as a $C\alpha$ trace.
V_L:V_H domains from Fab $P2_1A1$ shown in yellow as a $C\alpha$ trace.
The domain pairs were superimposed on the V_L domain (to the right in both (a) and (b)) on the basis of all $C\alpha$s from residues 3H–111H.

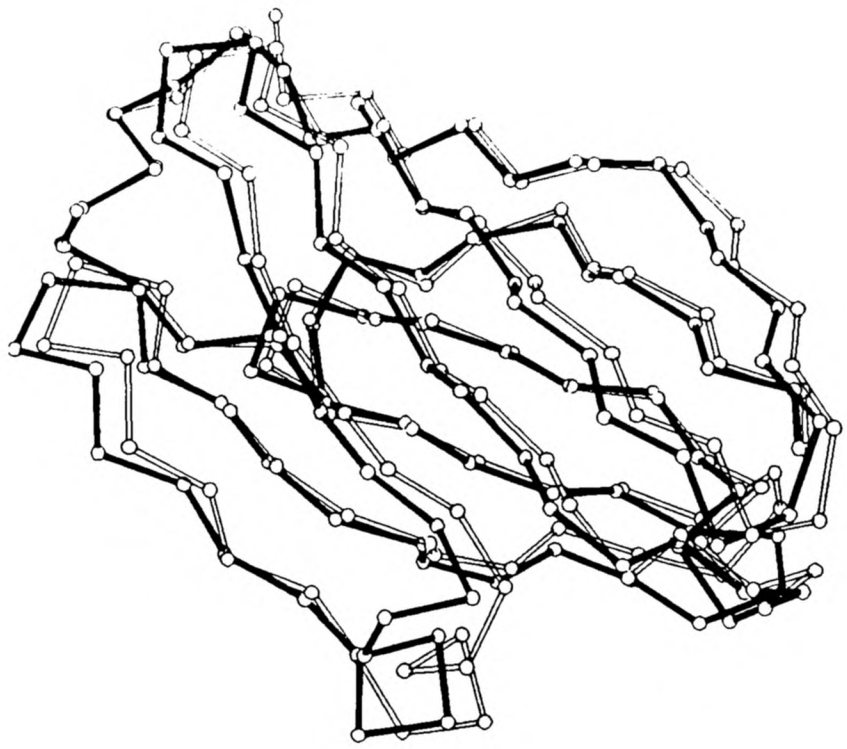
Figure 9: V_L:V_H Pairing Differences in Gloop2

on the basis of all $C\alpha$s in V_L. The movement is visibly significant: the V_L:V_H interface region scarcely moves, but there is greater movement at the CDRs (e.g. the distance between the $C\alpha$ atoms of Ser56H in CDR H2 in P1b and $P2_1A1$ is 2.37Å). Such a difference in the relative disposition of the CDRs could be expected to make a difference in the energy of interaction with antigen.

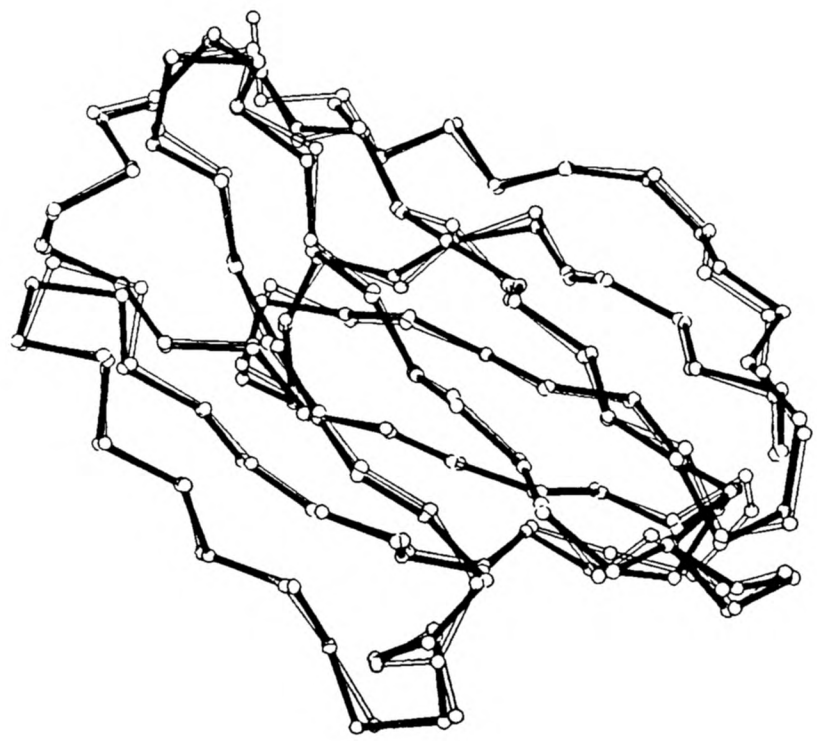
It is by no means certain that these conclusions could be extended to the solution structures. The elbow angle is clearly influenced by crystal packing in the Gloop2 Fabs and it is reasonable to assume that crystal packing can affect domain pairing either directly, through lattice contacts, or indirectly through influencing the elbow angle. There is a suggestion from the values in Tables 5 and 6 that the domain pairing is partly influenced by the elbow angles since the $P2_1B$ Fab (elbow angle 167.3°) is more similar to the P1a and P1b Fabs (elbow angles 169.7° and 166.6° respectively) than it is to either the $P2_1A1$ or $P2_1A2$ Fabs (elbow angles 151.9° and 152.0° respectively).

The recently reported differences in domain pairing between the refined NC41 Fv (Colman *et al.*, 1989) and the Fv's of Kol (Marquart *et al.*, 1980), McPC603 (Satow *et al.*, 1986) and J539 (Suh *et al.*, 1986) are lower than those originally reported (Colman *et al.*, 1987). Since Colman *et al.* (1989) do not calculate the domain pairing in the same way as used above it is not possible to make direct comparisons with the reported values and those of Gloop2. However the NC41 comparison (Colman *et al.*, 1989) does not include an IgG-κ Fab fragment (NC41 is an IgG2b,κ Fab (Colman *et al.*, 1987)) and the difference between NC41 and McPC603 is only 5.5° (Colman *et al.*, 1989). There appears to be no evidence for a significant change in domain pairing on antigen binding.

The resolution of the Gloop2 Fab structures is only modest, and in order to illustrate that this domain pairing change is NOT the result of some error in re-



(a) V_H from hybrid model (see text) and current $P2_1A1$ model, before refinement



(b) V_H from hybrid model (see text) and current $P2_1A1$ model, after refinement

5Å

V_H domain from Fab $P2_1A1$ shown in open lines as a $C\alpha$ trace (current model).
V_H domain from Fab P1a shown in solid lines as a $C\alpha$ trace.
See the text for a description of the initial model and the refinement steps.

Figure 10: Refinement of the $P2_1A1$–P1a Model to Illustrate Domain Pairing Differences

finement the Fv in the $P2_1A1$ Fab was replaced by the Fv from the P1a Fab by superimposition on all $C\alpha$ atoms in the V_L domain. The resultant Fab was then refined in the $P2_1$ 'A' cell against 8.0–3.3Å data (using the techniques discussed previously in Chapter 6). The initial residual was 35.1% on all data with $F > 0$, with the overall B factor set to that of the refined $P2_1A1$ Fab after cycle 14 (3.40Å^2, Chapter 6). Rigid body refinement of the V_H domain against 5.0–3.3Å data brought the residual down to 32.9% (on 8.0–3.3Å data) and 200 steps of energy minimisation rapidly brought the residual down to 22.8% on 8–3.3Å data. The weight on the structure factors used in the refinement was the same as that used in the final refinement of the $P2_1A1$ cell (WA=120000). The addition of a molecular dynamics simulation, using the slow cooling method, from 1000–300K followed by a further 200 steps of energy minimisation brought the residual to a 20.1%. The stereochemistry of this final model was comparable to the stereochemistry of the $P2_1A1$ model: RMS deviations from ideality on bond lengths, bond angles, dihedrals and improp-ers were 0.017Å, 3.590°, 28.727° and 1.501° respectively. The equivalent values for the $P2_1A1$ Fab were 0.017Å, 3.521°, 28.648° and 1.457° (Chapter 6).

The 'before and after' views of the V_H of the hybrid Fab in Figure 10 in comparison with the refined V_H from the $P2_1A1$ Fab show clearly that the V_L:V_H pairing had reproduced itself. These results indicate that the domain pairing changes seen between the Gloop2 Fabs were not a result of refinement error but a reflection of real structural differences.

7.2.4 Comparisons Between Gloop2 Structures

The previous sections have illustrated the flexibility of the quaternary structure of the Fabs by rigid body movement of whole domains in the Gloop2 structures, and it was shown that that this structural variability is representative of other Fab

(a) V_L Domain

	P1a	P1b	$P2_1A1$	$P2_1A2$
P1a	–	0.40	0.83	0.71
P1b	0.44	–	0.79	0.68
$P2_1A1$	0.92	0.85	–	0.67
$P2_1A2$	0.85	0.77	0.77	–

(b) C_L Domain

	P1a	P1b	$P2_1A1$	$P2_1A2$
P1a	–	0.92	0.92	0.80
P1b	1.00	–	0.83	0.88
$P2_1A1$	1.03	0.96	–	0.62
$P2_1A2$	0.98	1.06	0.74	–

(c) V_H Domain

	P1a	P1b	$P2_1A1$	$P2_1A2$
P1a	–	0.40	0.77	0.73
P1b	0.47	–	0.68	0.65
$P2_1A1$	0.88	0.81	–	0.59
$P2_1A2$	0.82	0.76	0.73	–

(d) C_H1 Domain

	P1a	P1b	$P2_1A1$	$P2_1A2$
P1a	–	1.28	1.50	1.59
P1b	1.38	–	1.38	1.39
$P2_1A1$	1.58	1.51	–	0.74
$P2_1A2$	1.69	1.45	0.90	–

Above diagonal — RMS on $C\alpha$ atoms.
Below diagonal — RMS on Backbone plus $C\beta$ atoms.
Residue ranges for comparison: V_L 1L–105L.
C_L 111L–214L.
V_H 3H–111H.
C_H1 117H–210H.
All values in Angstrom units.

Table 7: Comparison of Gloop2 Domains

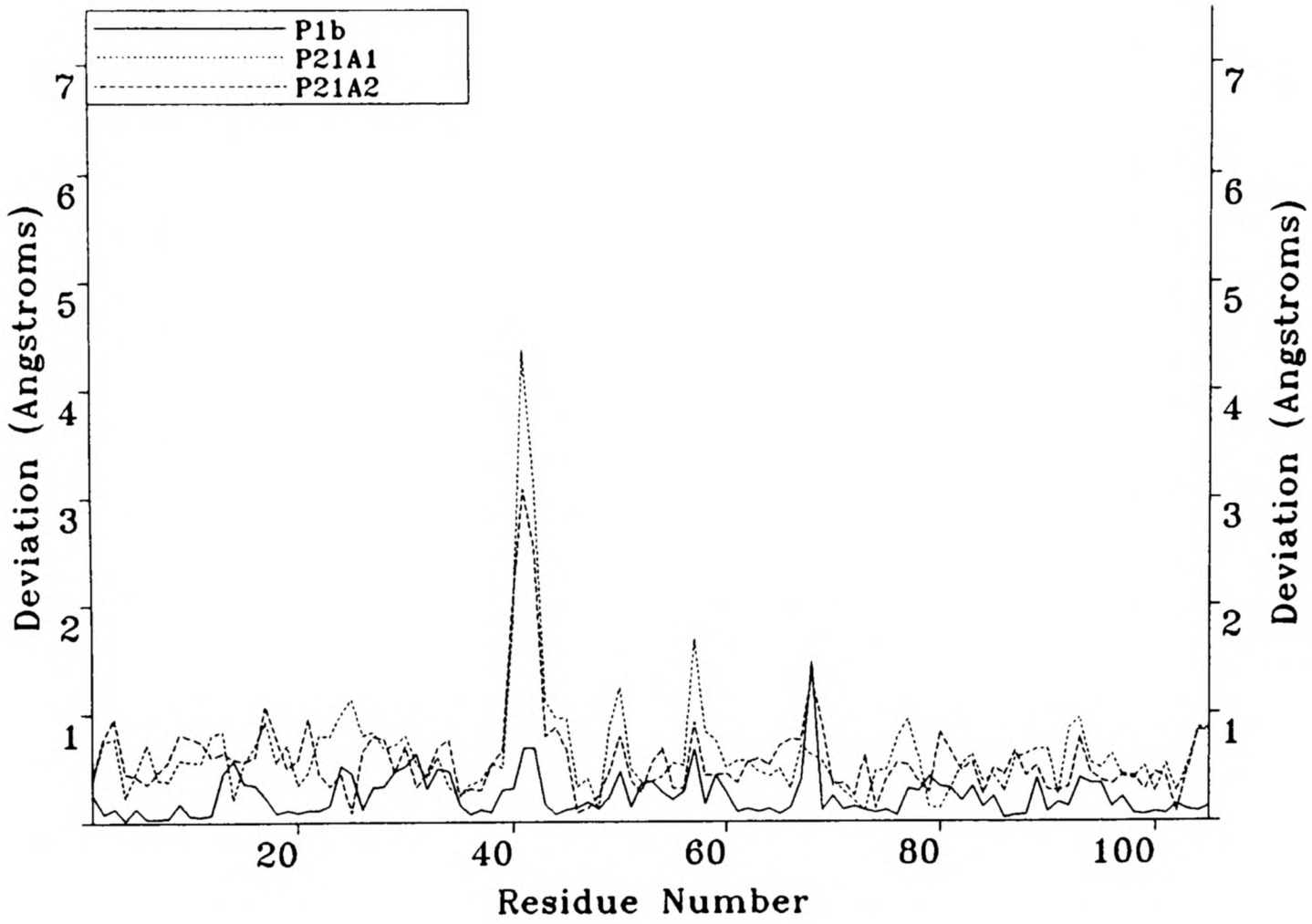
structures. In this section the local conformational changes within domains in the Gloop2 crystal forms are analysed.

Conformational changes are expected between the Fabs from the crystal packing forces in the different crystal environments of the Fabs. However the difference in elbow angle and domain pair associations may also have structural ramifications in the conformation of the polypeptide chain. Of particular interest is the conformation of the CDRs in the crystal forms, as it remains to be demonstrated whether these regions are capable of undergoing conformational change on binding antigen, as has been suggested.

The program ASH (Appendix 2) was used to analyse the structural homology between the Gloop2 Fabs based on the superimposition of $C\alpha$ atoms or backbone atoms plus $C\beta$ (i.e. $C\alpha, C, O, N, C\beta$) in each residue. Table 7 gives the RMS values for the best fits between domains with the CDRs included. The superimpositions on all $C\alpha$ atoms in each domain with respect to the P1a Fab are also shown in Figures 11–14 by domains. The principal differences are summarised below for each domain:

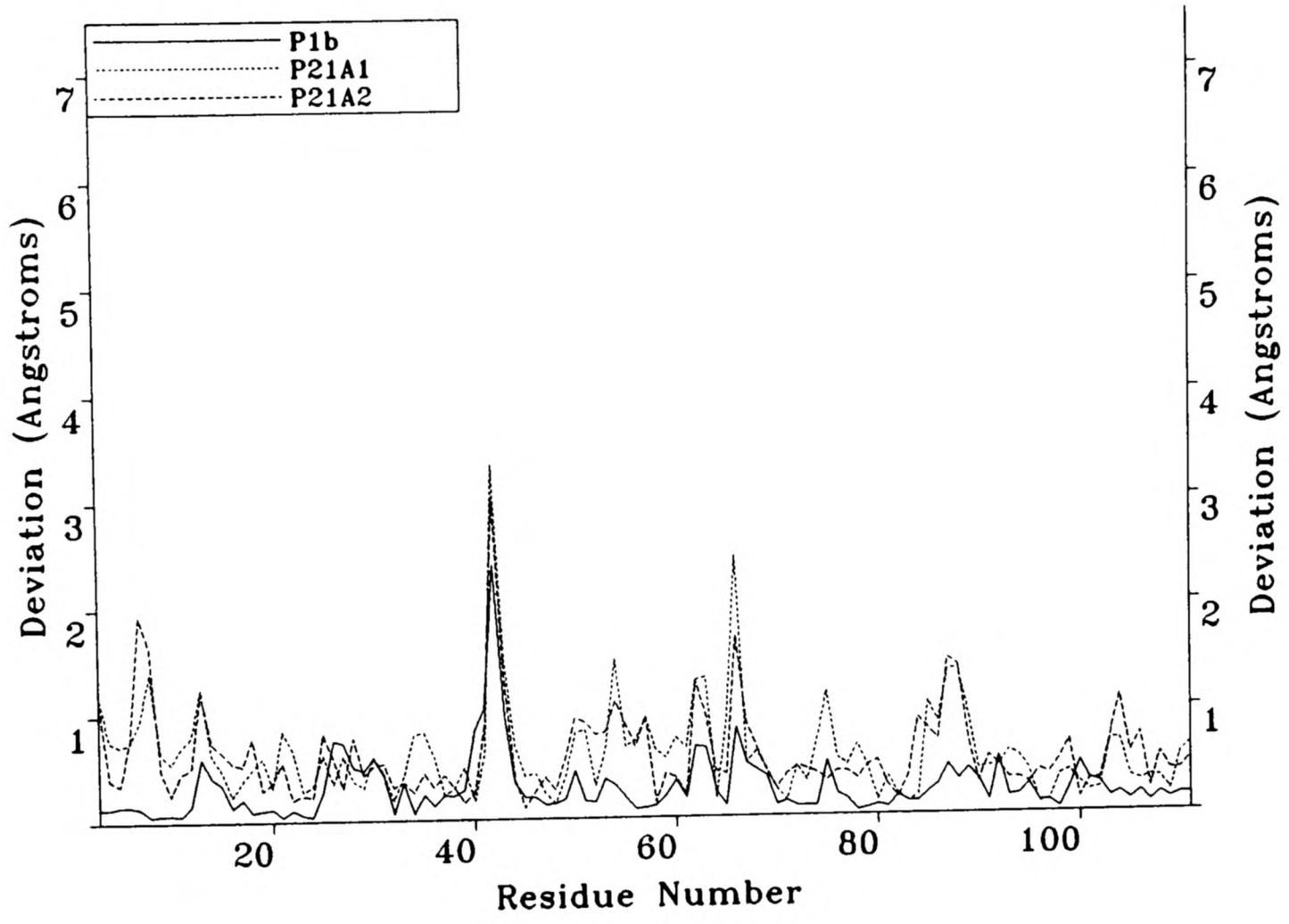
V_L: The P1b Fab is very similar to the P1b Fab with differences barely above the mean coordinate error ($\approx$0.35Å) for most residues. Fabs $P2_1A1$ and $P2_1A2$ differ principally in the region 40L–43L: a loop at the centre of the elbow bend region which may change conformation due to the elbow angle difference. The CDRs (24L–34L, 50L–56L and 89L–97L) do not show significant differences between the copies.

V_H: Again the P1b Fab is very similar to P1a except for the region 40H–43H, again a loop on the elbow region, which suggests a chain-tracing error in this region since the elbow angles of the two Fabs are very similar. The $P2_1A1$ and $P2_1A2$



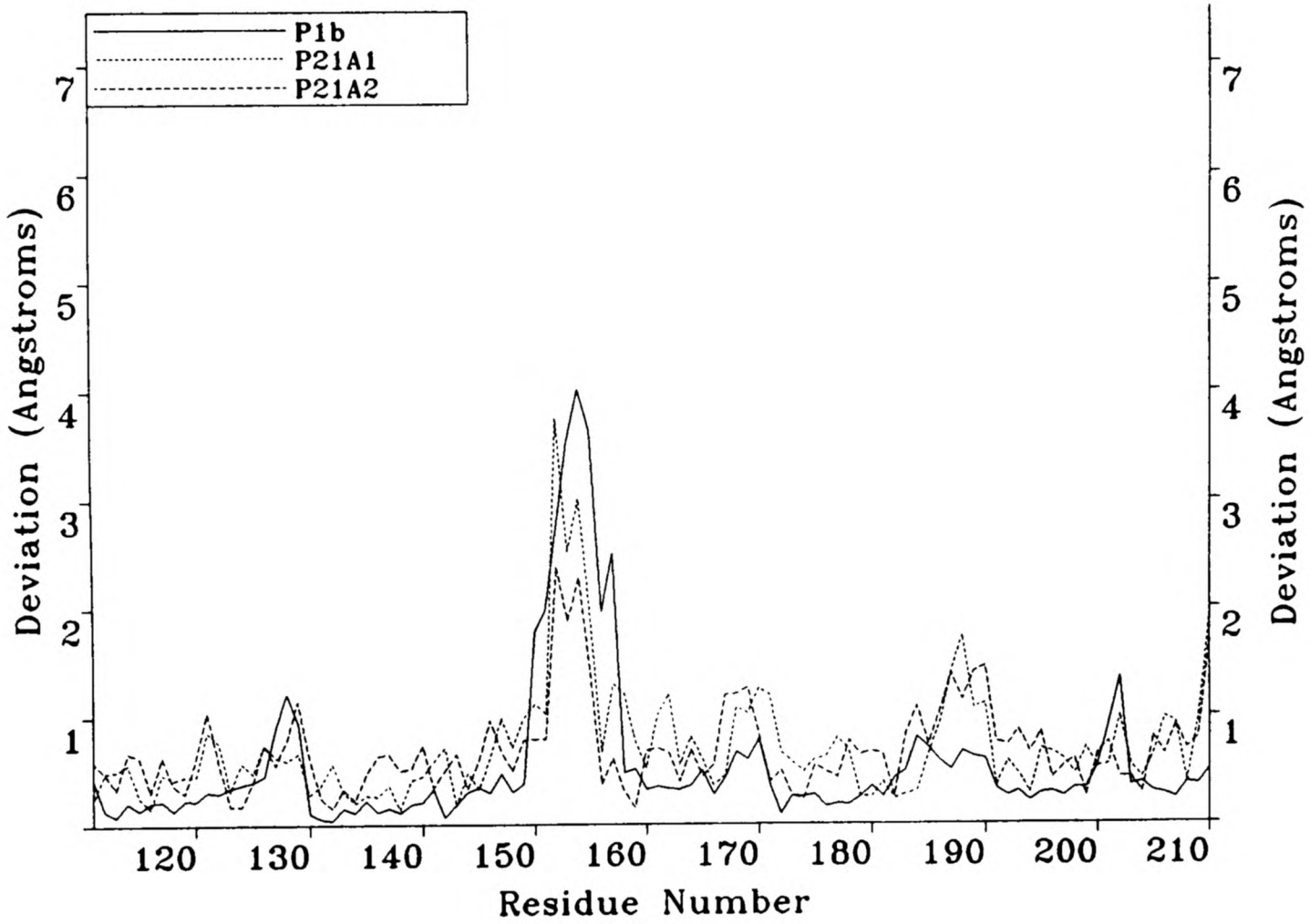
Distance between $C\alpha$ atoms at equivalent positions in the sequence plotted with respect to residue number.
Superimpositions performed on all $C\alpha$s from residues 1L–105L in the Gloop2 sequence (Appendix 1), including CDRs.
Reference coordinates: V_L from Fab P1a.

Figure 11: Deviations Between the Gloop2 V_L Domains



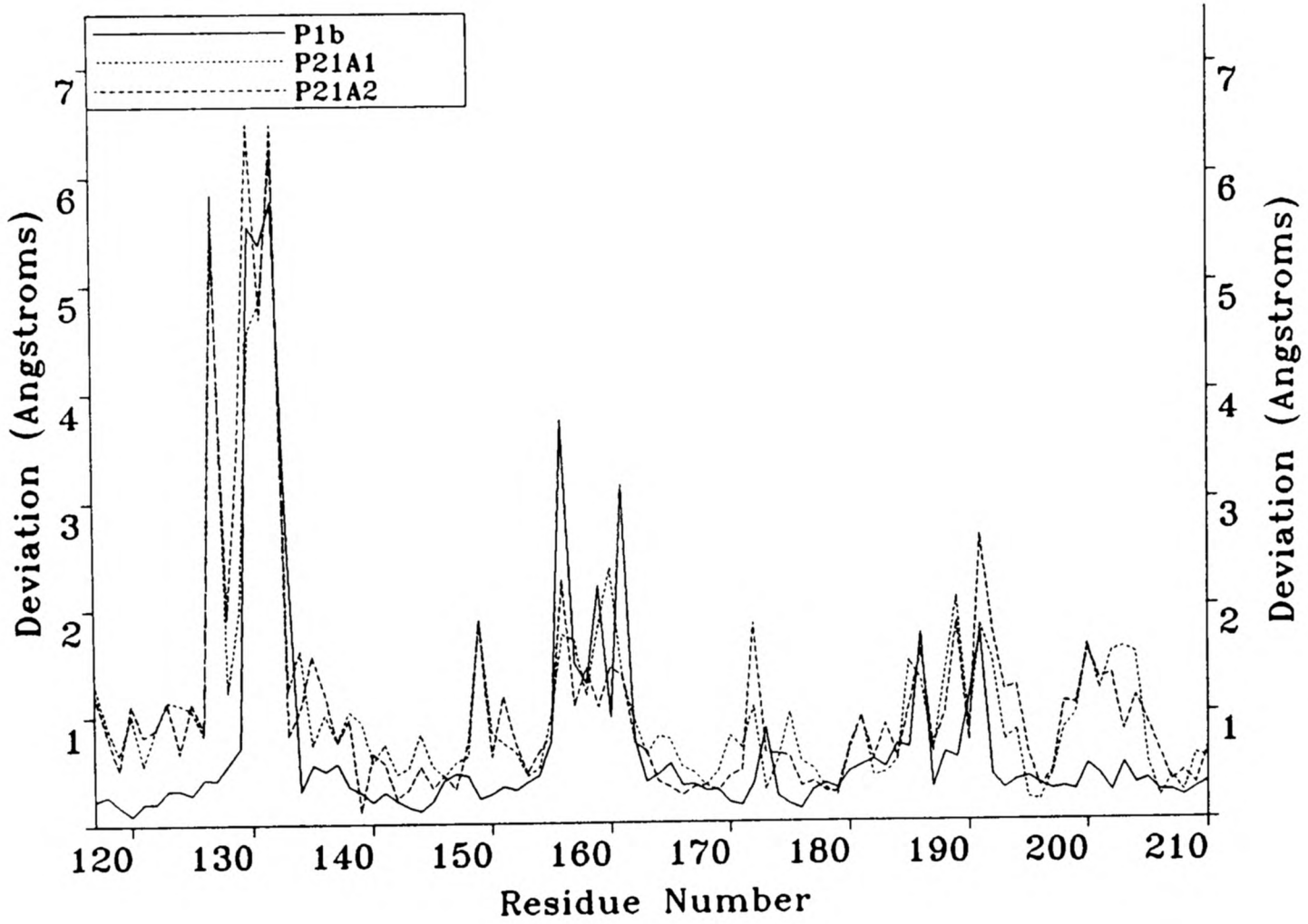
Distance between $C\alpha$ atoms at equivalent positions in the sequence plotted with respect to residue number.
Superimpositions performed on all $C\alpha$s from residues 3H–111H in the Gloop2 sequence (Appendix 1), including CDRs.
Reference coordinates: V_H from Fab P1a.

Figure 12: Deviations Between the Gloop2 V_H Domains



Distance between $C\alpha$ atoms at equivalent positions in the sequence plotted with respect to residue number. Superimpositions performed on all $C\alpha$s from residues 111L–210L in the Gloop2 sequence (Appendix 1). Reference coordinates: C_L from Fab P1a.

Figure 13: Deviations Between the Gloop2 C_L Domains



Distance between $C\alpha$ atoms at equivalent positions in the sequence plotted with respect to residue number. Superimpositions performed on all $C\alpha$s from residues 117H–210H in the Gloop2 sequence (Appendix 1). Reference coordinates: C_H1 from Fab P1a.

Figure 14: Deviations Between the Gloop2 C_H1 Domains

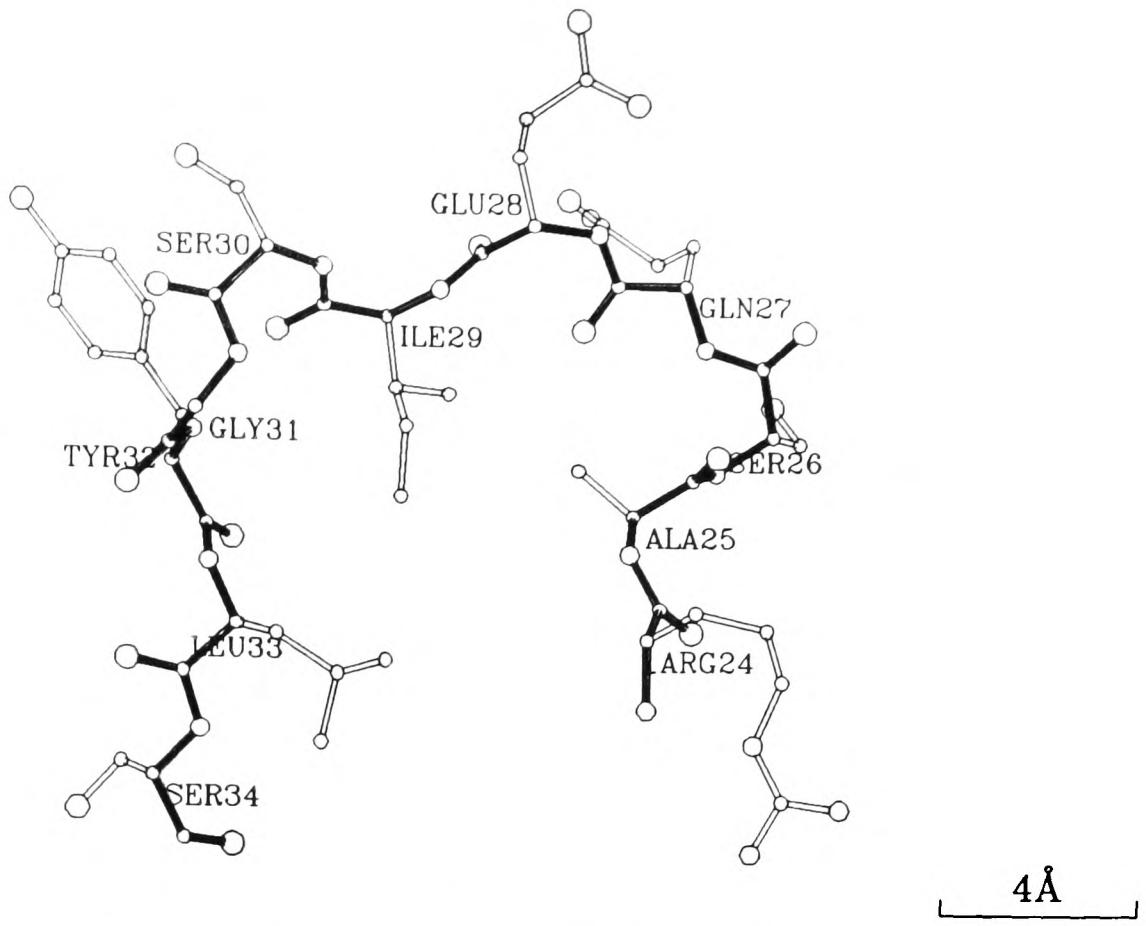
differ from the P1a Fab at the loop in the elbow bend region: 40H–43H, and the end of CDR H2: 60H–65H. CDR H2 has poor density in this region for the $P2_1$ 'A' form structures, and the chain trace may not be entirely reliable. The other regions within the CDRs (31H–35H, 50H–65H, 99H–102H) do not show significant deviation.

C_L: The major deviation of all three Fabs with respect to the P1a structure is the loop at 150L–158L. This loop is the most distant from the C_L:C_H1 interface and is affected by crystal packing forces in the crystal structures.

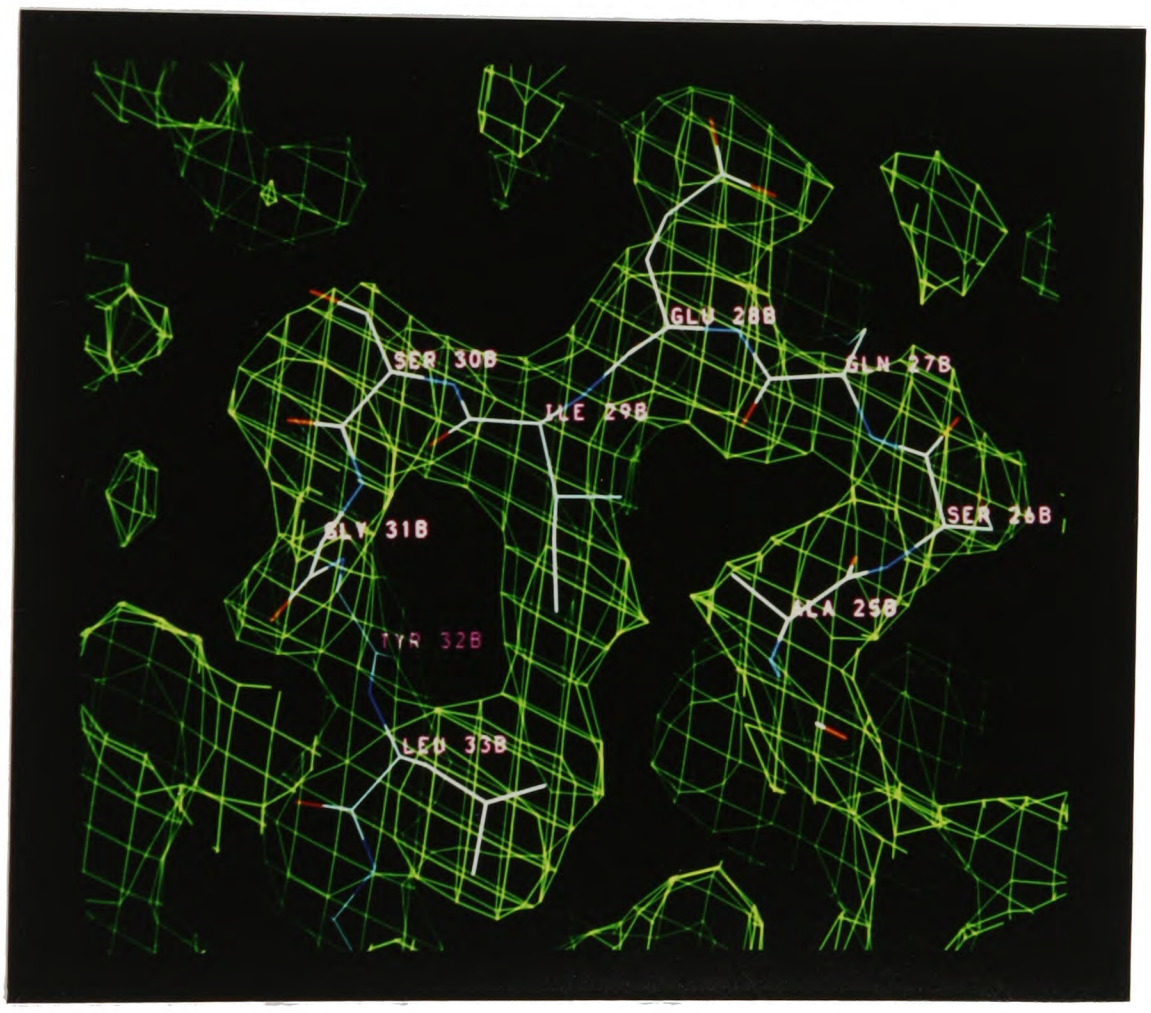
C_H1: The three main regions of deviation with respect to the P1a C_H1 correspond with the loops that have proven difficult to place with confidence in the $P1$ refinement. It would appear that these regions represent sources of general structural variability between the C_H1 domains, and the corresponding loops in C_H1 domains from other Fabs show significant variability, supporting this suggestion.

The deviations between the P1a and P1b Fabs are small in the regions where non-crystallographic symmetry restraints were applied (as expected). The CDRs, although loop regions, do not show any particular deviation between the crystal forms. The different crystal forms of the Gloop2 show the expected structural homology between domains when the error levels in the structures (Section 7.1) are taken into account and allowances are made for crystal packing interactions.

Figures 15–31 represent summaries of the conformations of the six CDRs in the four Fabs. The P1a CDR is shown alone as a "ball-and-stick" representation, and also with its associated $(2F_{obs}-F_{calc}),\alpha_{calc}$ electron density calculated with the CDR omitted, from approximately the same viewpoint as the individual CDR (subject to the constraints of displaying the map). The superimposition of all four CDRs is shown along with the superimposition of the most homologous CDR in length and



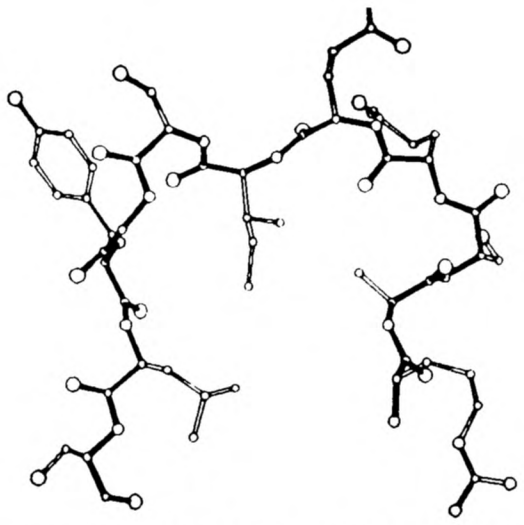
(a) CDR L1 from the P1a Gloop2 Fab



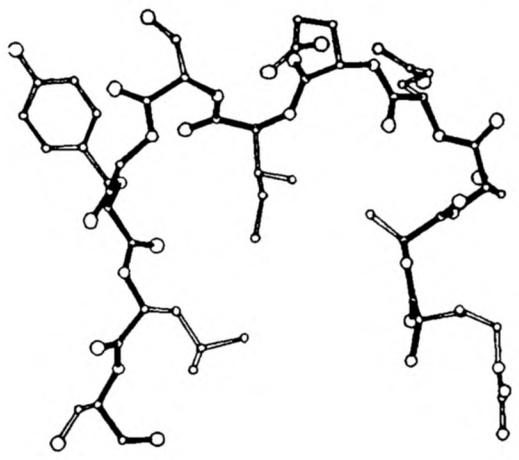
(b) $(2F_{obs} - F_{calc}), \alpha_{calc}$ omit map for some of the residues within CDR L1

'Omit' electron density map calculated from P1a coordinates from which the CDR L1 residues had been omitted. The $(2F_{obs} - F_{calc}), \alpha_{calc}$ map is contoured at 0.6xRMS of the map above zero, with $F(000)$ omitted.

Figure 15: The Conformation of CDR L1 in the P1a Fab and Associated Electron Density

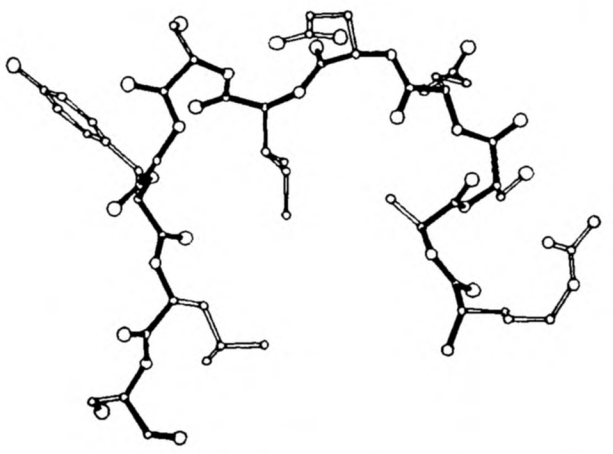


(a) CDR L1 from the P1a Fab

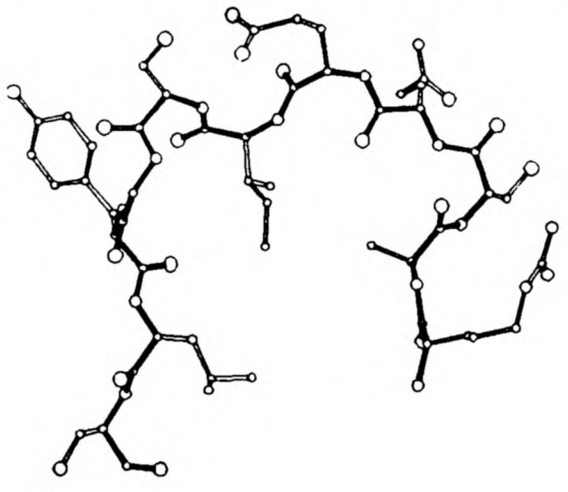


4Å

(b) CDR L1 from the P1b Fab



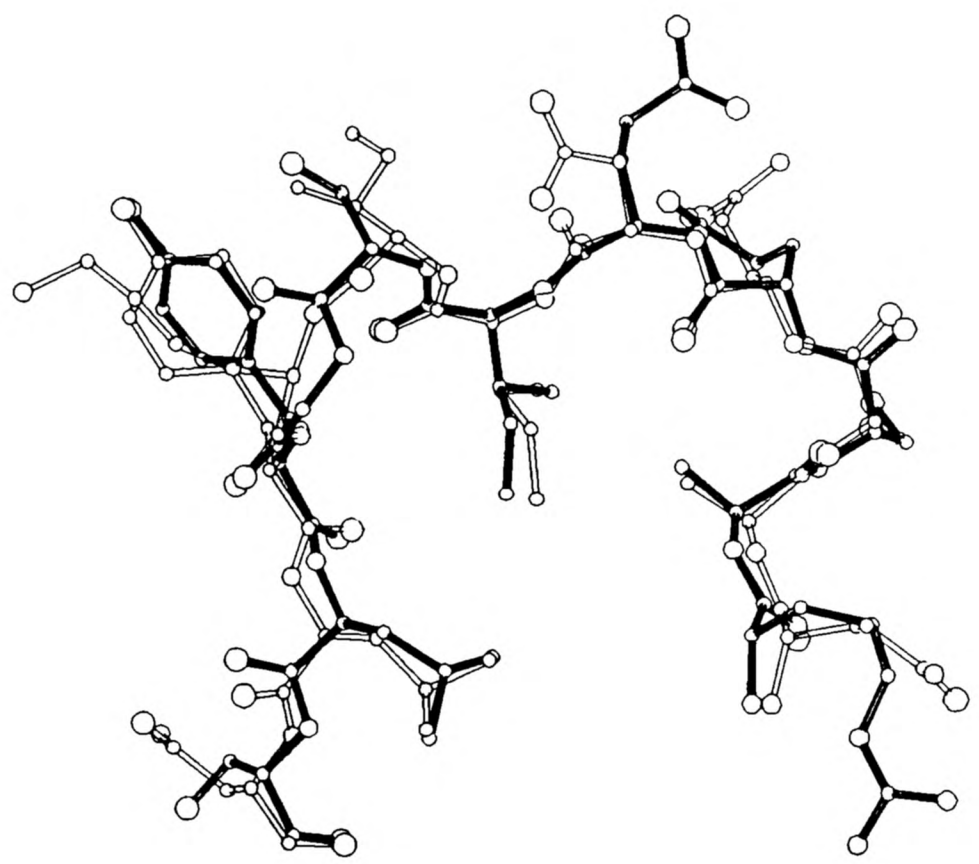
(c) CDR L1 from the $P2_1A1$ Fab



(d) CDR L1 from the $P2_1A2$ Fab

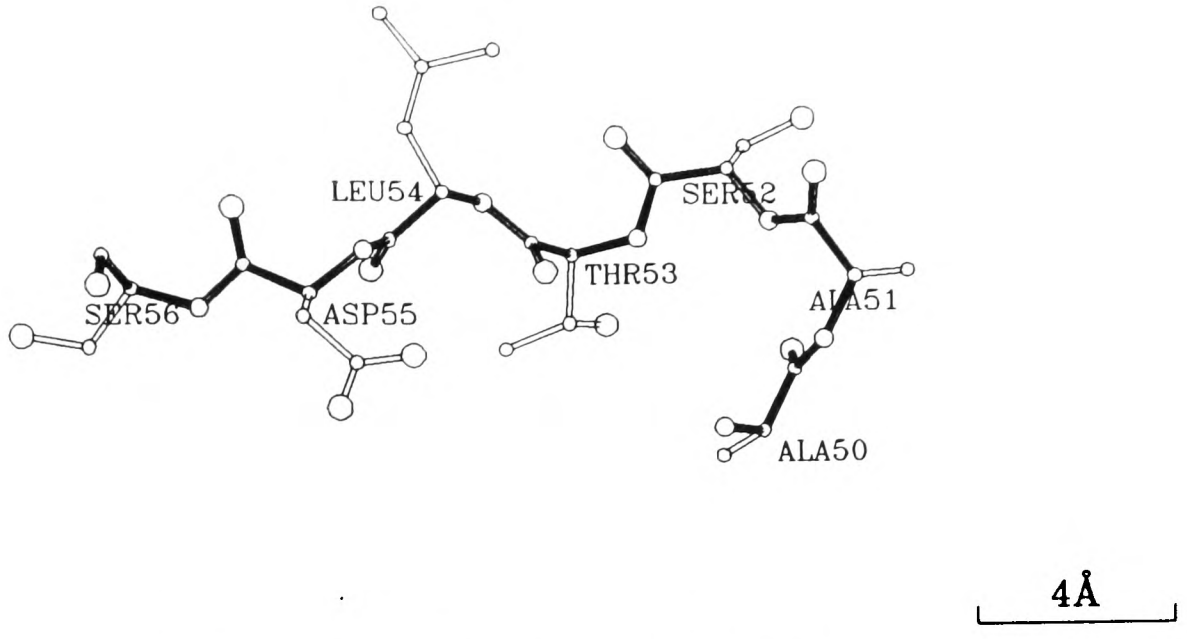
CDRs viewed from equivalent directions with respect to the whole V_L domain.

Figure 16: Comparison of the Conformations of CDR L1 in the Four Gloop2 Fabs

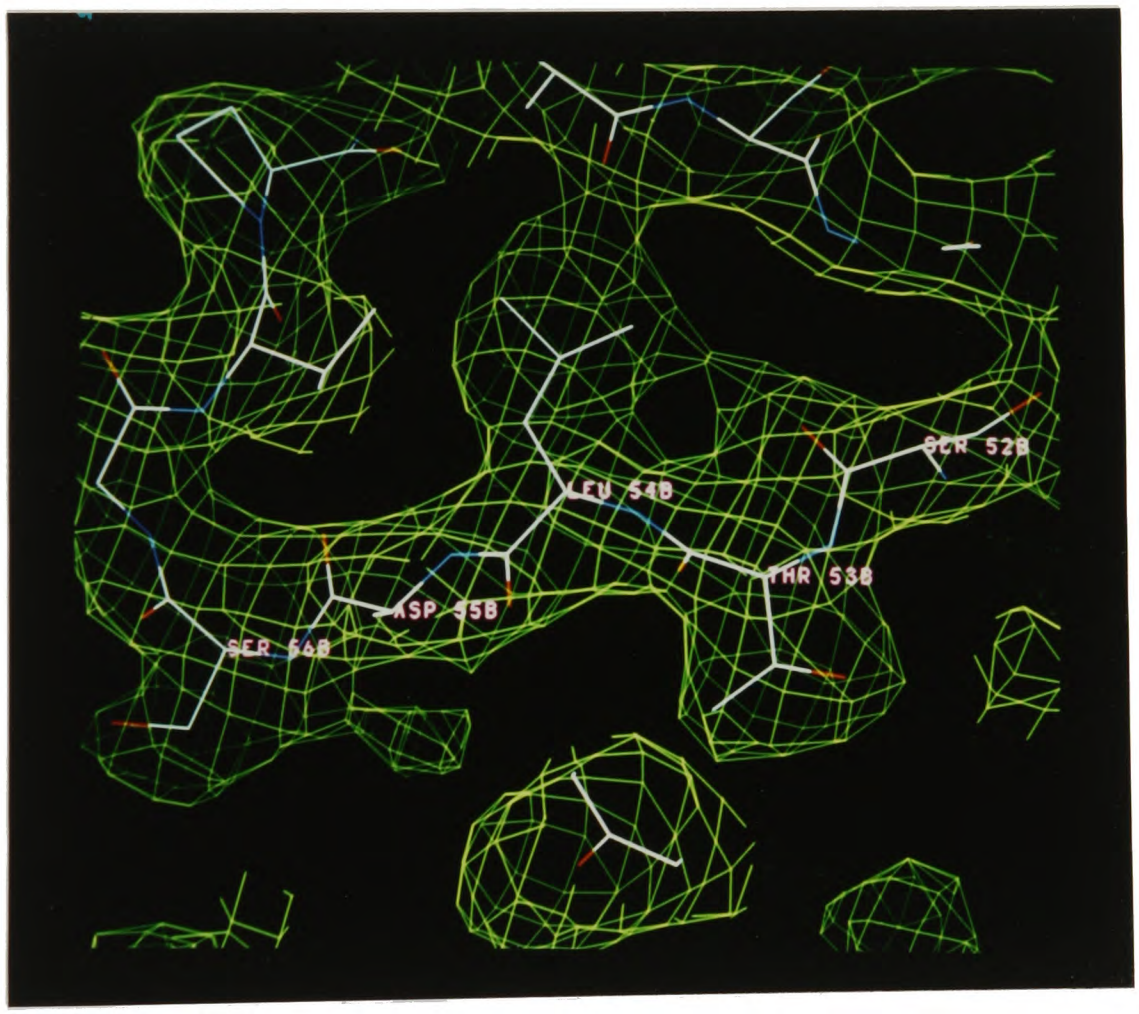


Superimposition of CDR L1 from P1a with CDR L1 from Rei
Superimposition of CDRs made on all $C\alpha$ atoms in the CDRs.

Figure 17: Comparison of CDR L1 from Gloop2 with that from Rei



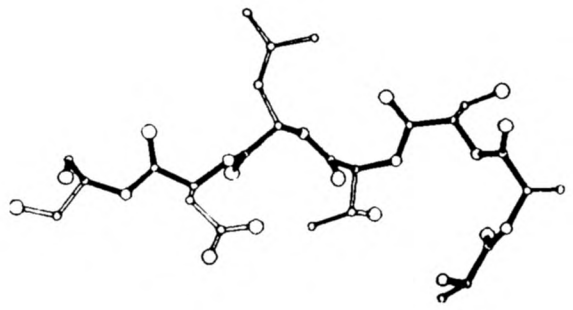
(a) CDR L2 from the P1a Gloop2 Fab



(b) $(2F_{obs} - F_{calc}), \alpha_{calc}$ omit map for some of the residues within CDR L2

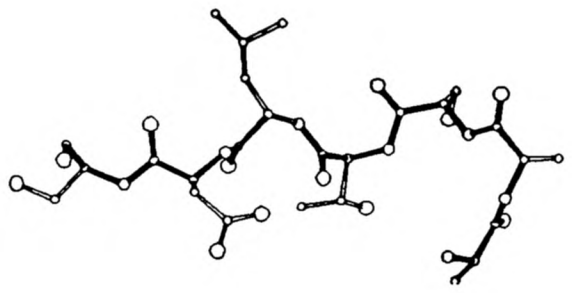
'Omit' electron density map calculated from P1a coordinates from which the CDR L2 residues had been omitted. The $(2F_{obs} - F_{calc}), \alpha_{calc}$ map is contoured at 0.6xRMS of the map above zero, with $F(000)$ omitted.

Figure 18: The Conformation of CDR L2 in the P1a Fab and Associated Electron Density

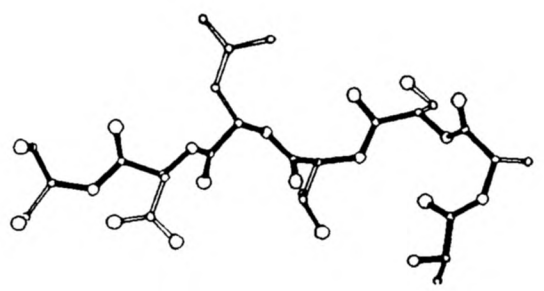


(a) CDR L2 from the P1a Fab

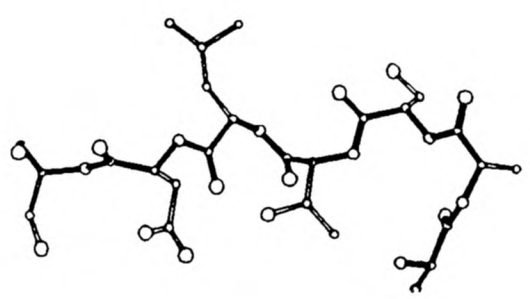
4Å



(b) CDR L2 from the P1b Fab



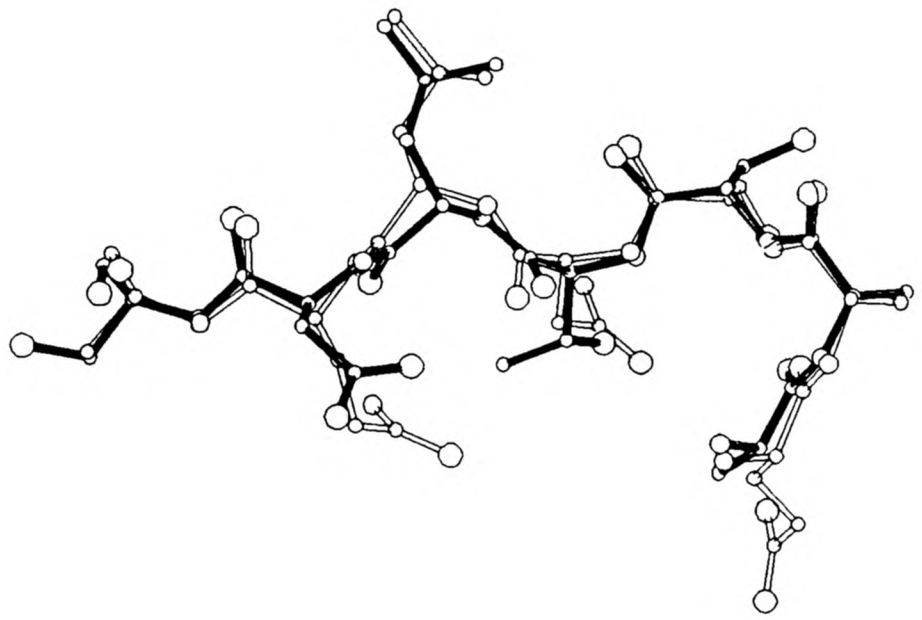
(c) CDR L2 from the $P2_1A1$ Fab



(d) CDR L2 from the $P2_1A2$ Fab

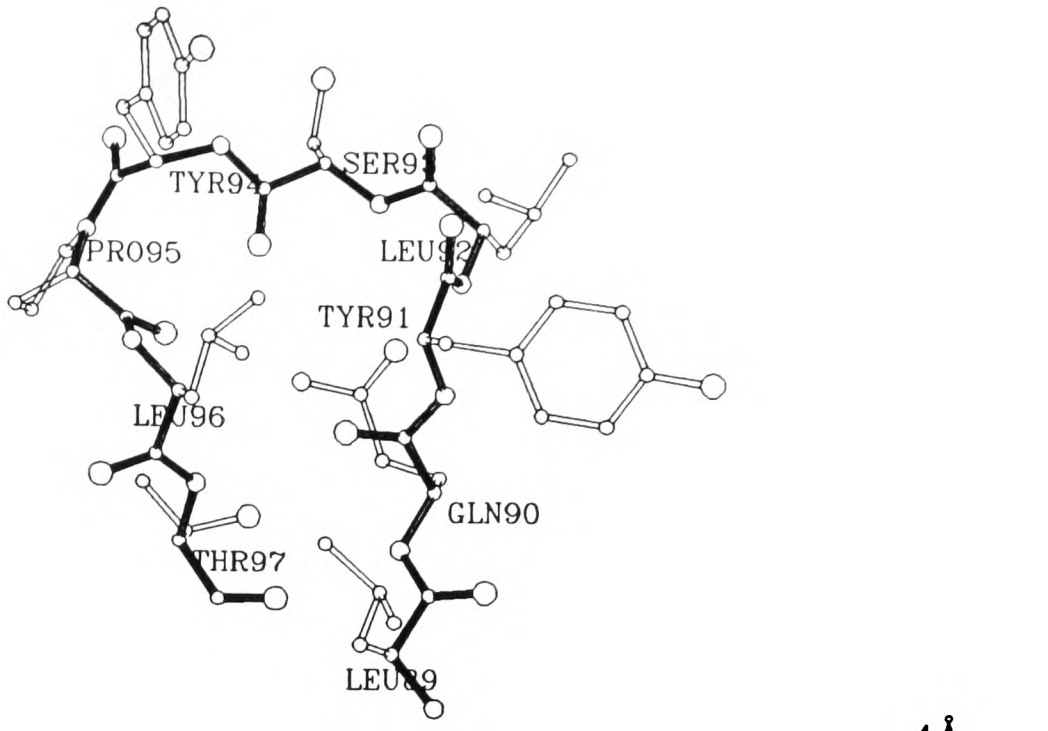
CDRs viewed from equivalent directions with respect to the whole V_L domain.

Figure 19: Comparison of the Conformations of CDR L2 in the Four Gloop2 Fabs

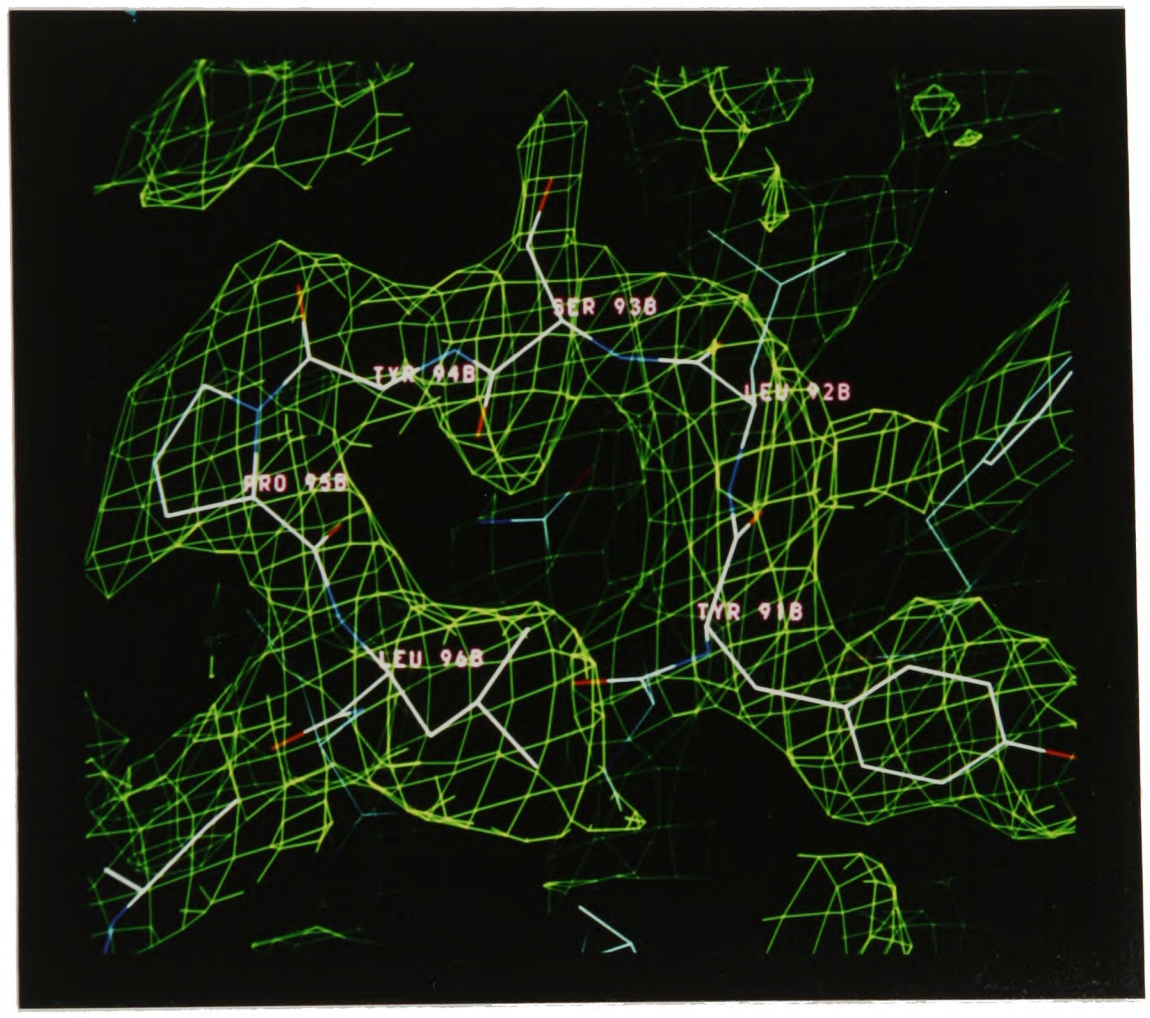


Superimposition of CDR L2 from P1a with CDR L2 from Rei
Superimposition of CDRs made on all $C\alpha$ atoms in the CDRs.

Figure 20: Comparison of CDR L2 from Gloop2 with that from Rei



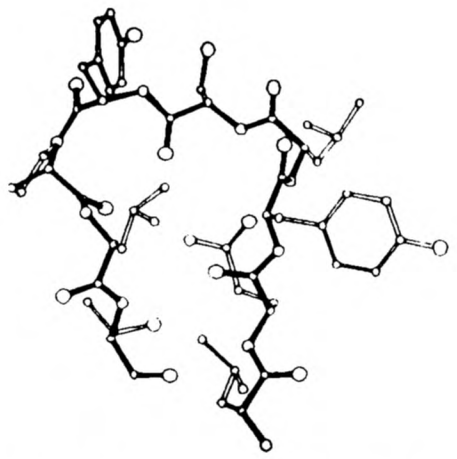
(a) CDR L3 from the P1a Gloop2 Fab



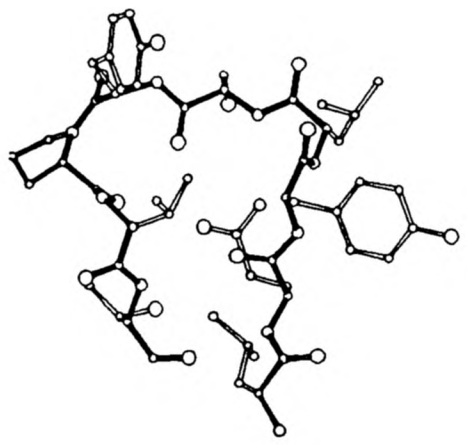
(b) $(2F_{obs} - F_{calc}), \alpha_{calc}$ omit map for some of the residues within CDR L3

'Omit' electron density map calculated from P1a coordinates from which the CDR L3 residues had been omitted. The $(2F_{obs} - F_{calc}), \alpha_{calc}$ map is contoured at 0.6xRMS of the map above zero, with $F(000)$ omitted.

Figure 21: The Conformation of CDR L3 in the P1a Fab and Associated Electron Density

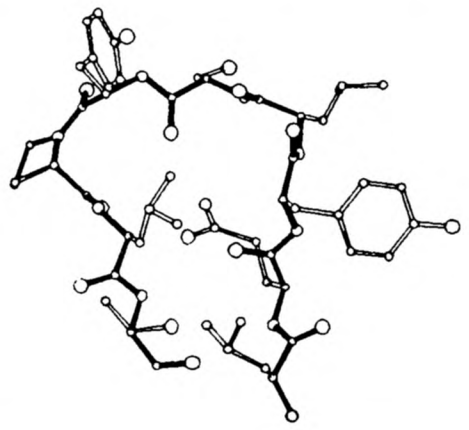


(a) CDR L3 from the P1a Fab

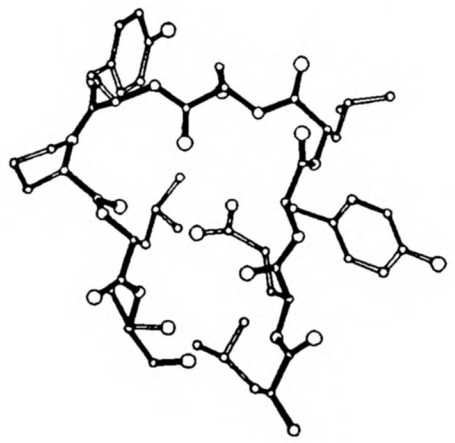


4Å

(b) CDR L3 from the P1b Fab



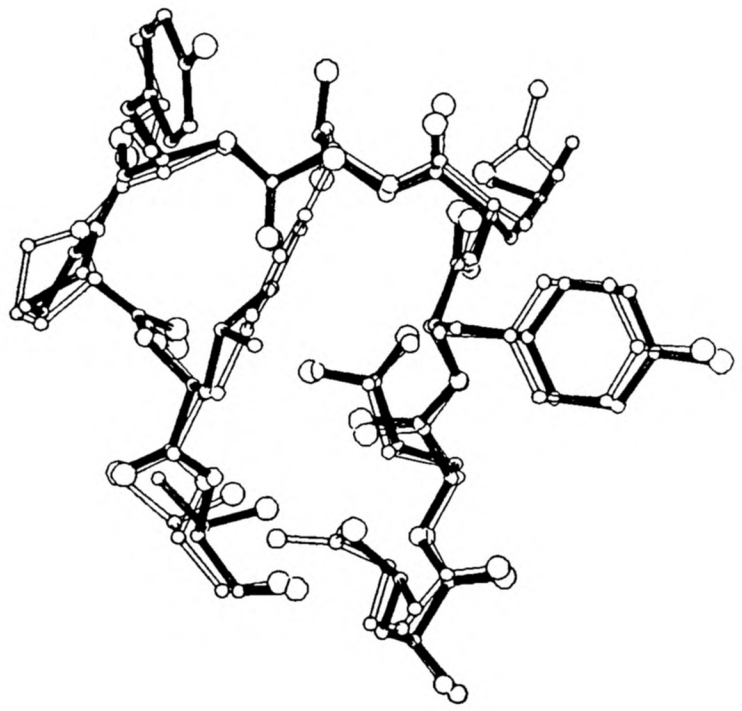
(c) CDR L3 from the $P2_1A1$ Fab



(d) CDR L3 from the $P2_1A2$ Fab

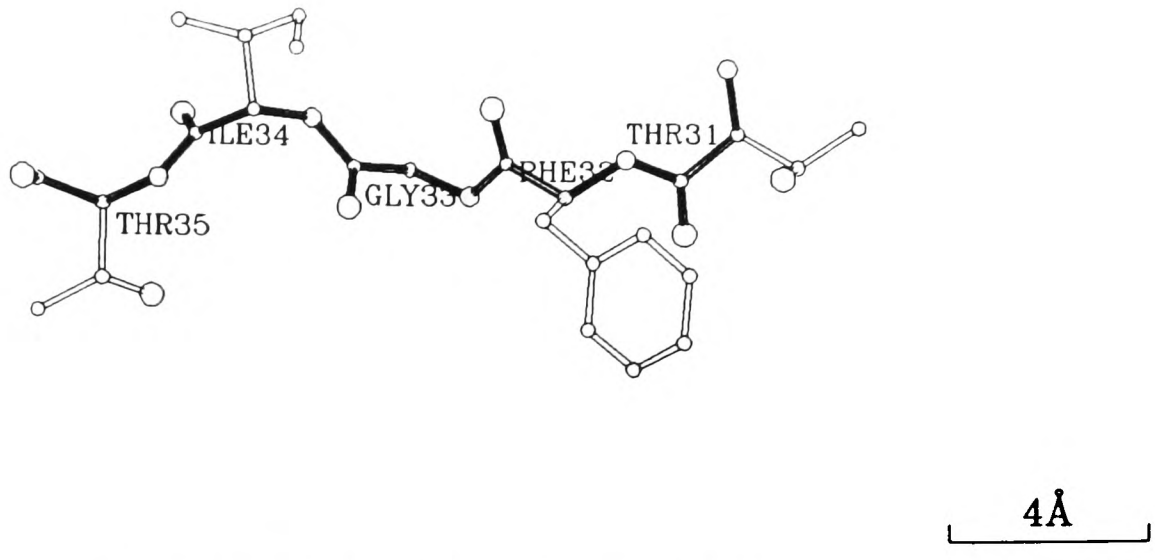
CDRs viewed from equivalent directions with respect to the whole V_L domain.

Figure 22: Comparison of the Conformations of CDR L3 in the Four Gloop2 Fabs

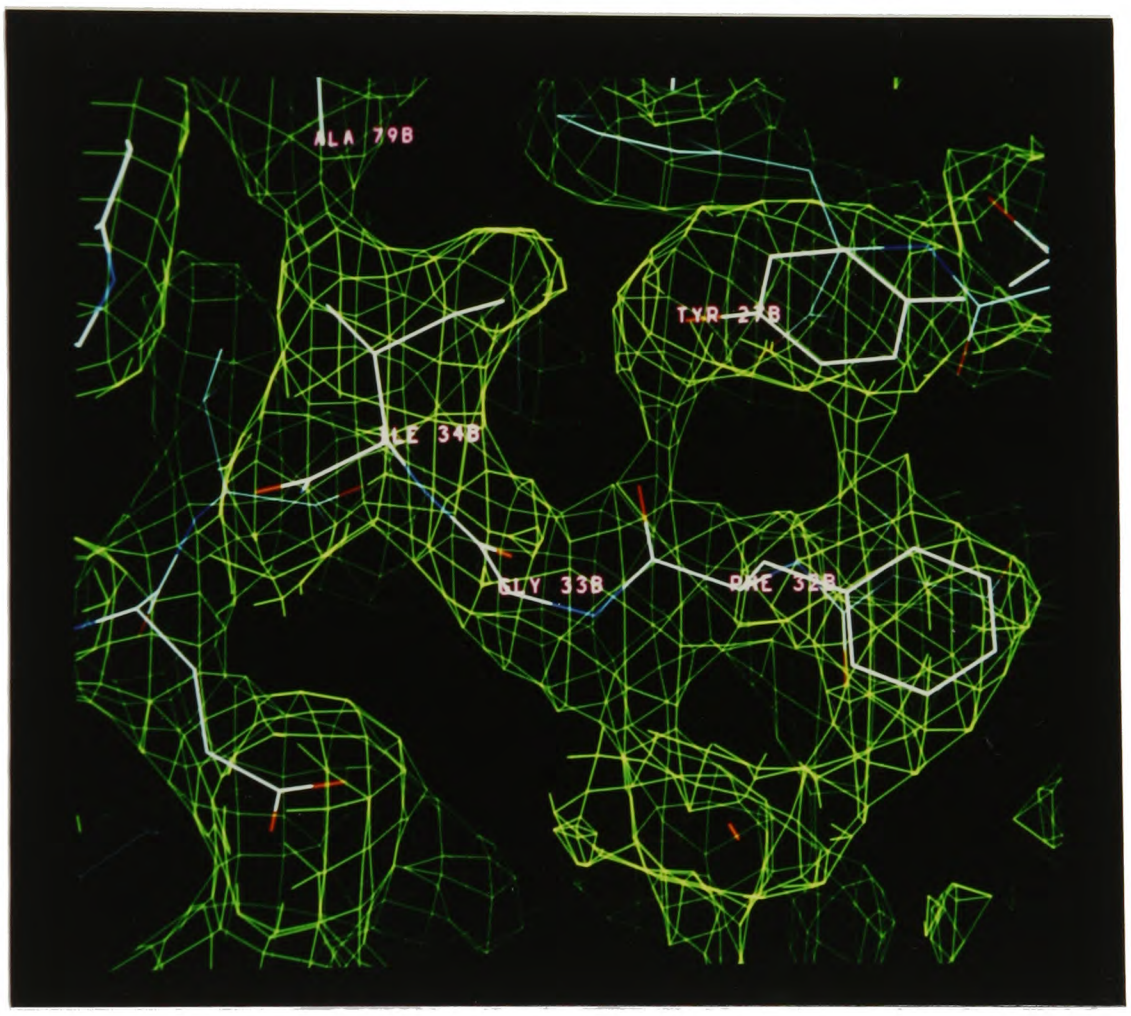


Superimposition of CDR L3 from P1a with CDR L3 from Rei
Superimposition of CDRs made on all $C\alpha$ atoms in the CDRs.

Figure 23: Comparison of CDR L3 from Gloop2 with that from Rei



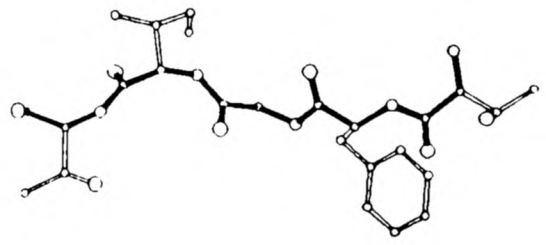
(a) CDR H1 from the P1a Gloop2 Fab



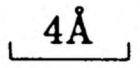
(b) $(2F_{obs} - F_{calc}), \alpha_{calc}$ omit map for some of the residues within CDR H1

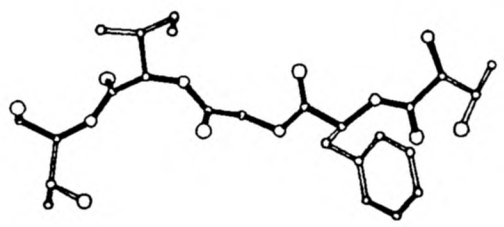
'Omit' electron density map calculated from P1a coordinates from which the CDR H1 residues had been omitted. The $(2F_{obs} - F_{calc}), \alpha_{calc}$ map is contoured at 0.6xRMS of the map above zero, with $F(000)$ omitted.

Figure 24: The Conformation of CDR H1 in the P1a Fab and Associated Electron Density

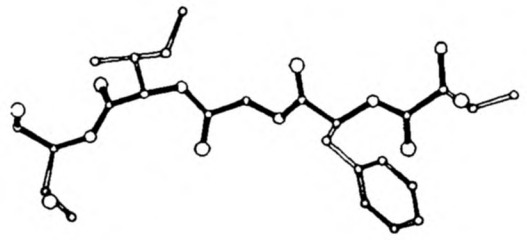


(a) CDR H1 from the P1a Fab

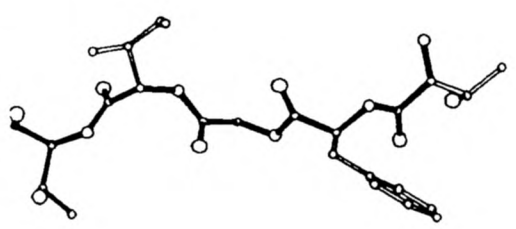




(b) CDR H1 from the P1b Fab



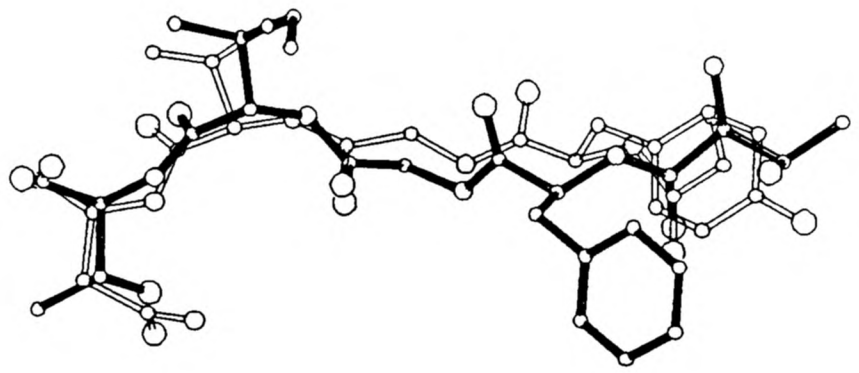
(c) CDR H1 from the $P2_1A1$ Fab



(d) CDR H1 from the $P2_1A2$ Fab

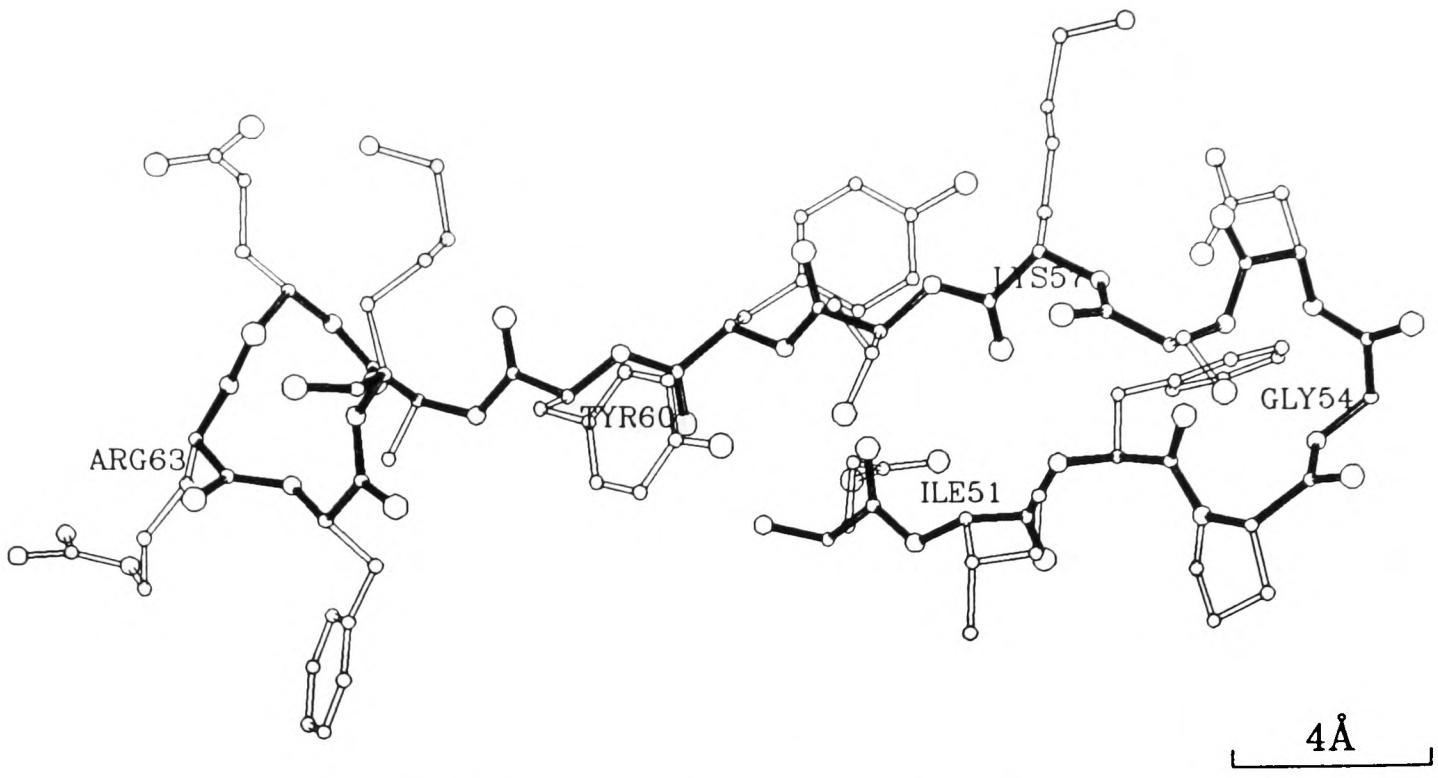
CDRs viewed from equivalent directions with respect to the whole V_H domain.

Figure 25: Comparison of the Conformations of CDR H1 in the Four Gloop2 Fabs

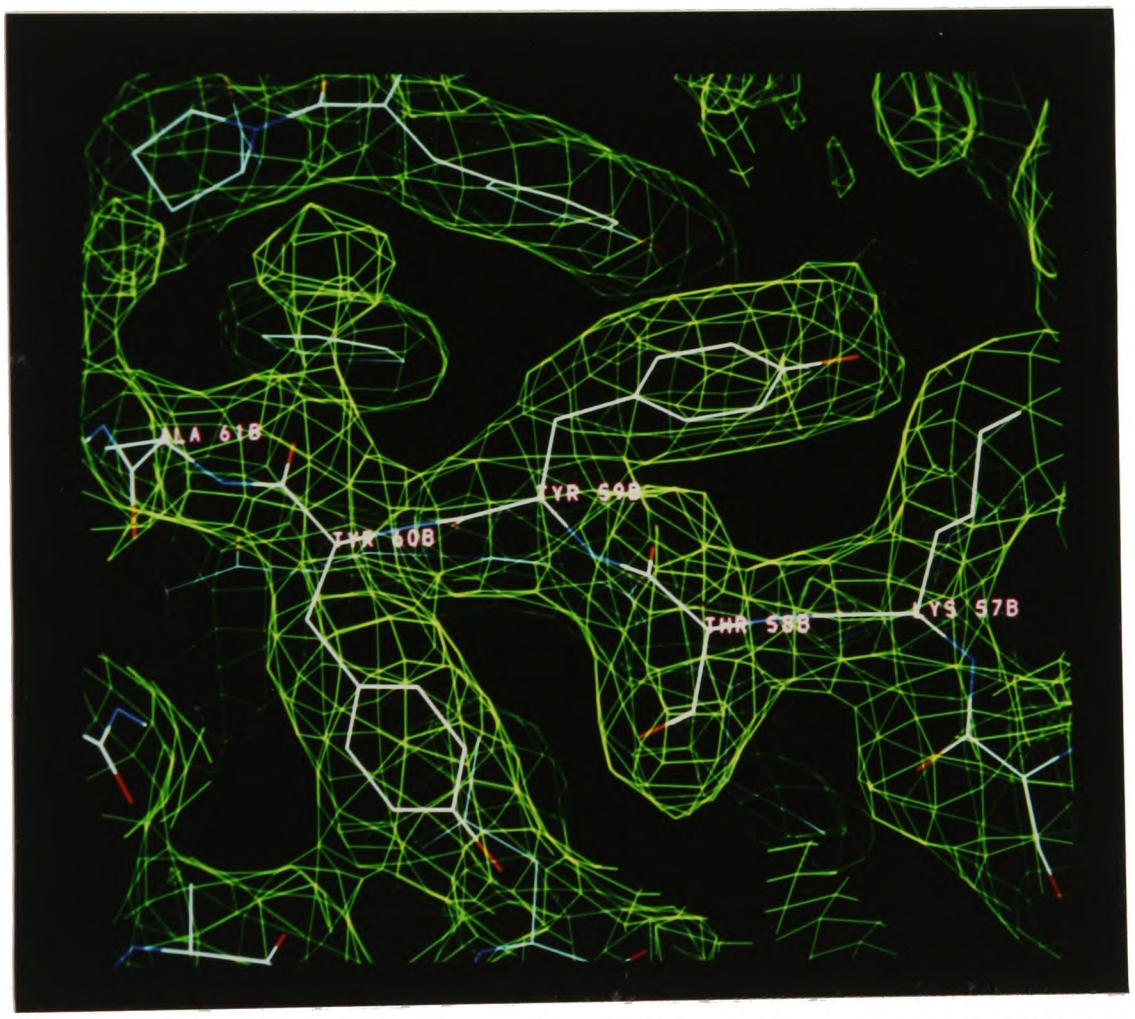


Superimposition of CDR H1 from P1a with CDR H1 from D1.3
Superimposition of CDRs made on all $C\alpha$ atoms in the CDRs.

Figure 26: Comparison of CDR H1 from Gloop2 with that from D1.3



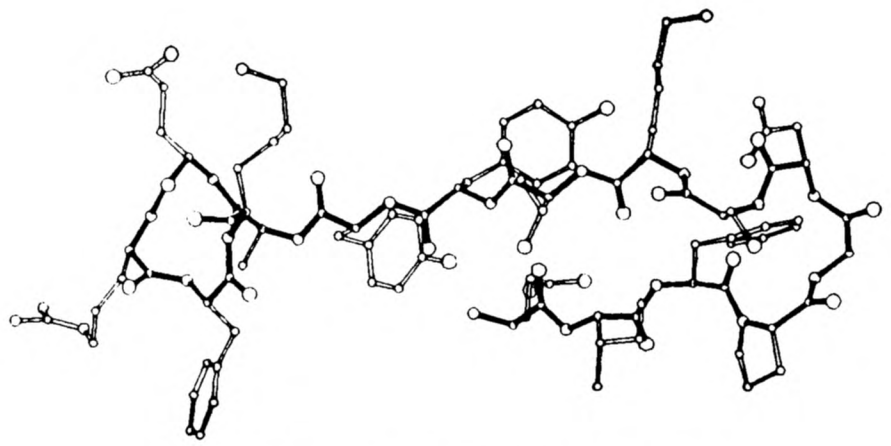
(a) CDR H2 from the P1a Gloop2 Fab



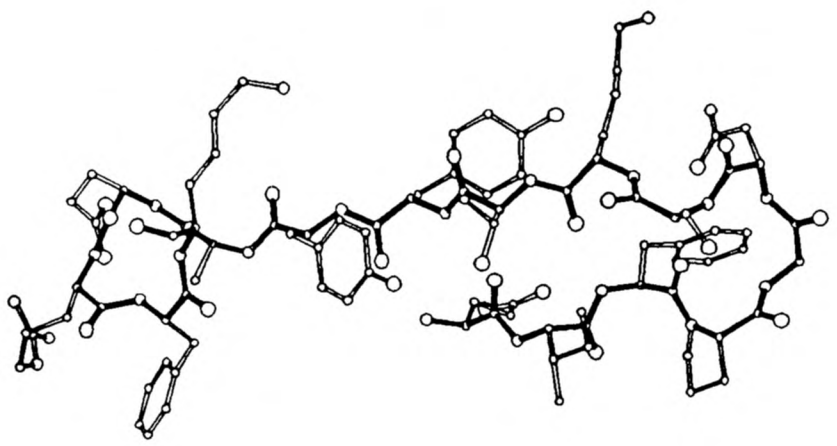
(b) $(2F_{obs} - F_{calc}), \alpha_{calc}$ omit map for some of the residues within CDR H2

'Omit' electron density map calculated from P1a coordinates from which the CDR H2 residues had been omitted. The $(2F_{obs} - F_{calc}), \alpha_{calc}$ map is contoured at 0.6xRMS of the map above zero, with $F(000)$ omitted.

Figure 27: The Conformation of CDR H2 in the P1a Fab and Associated Electron Density

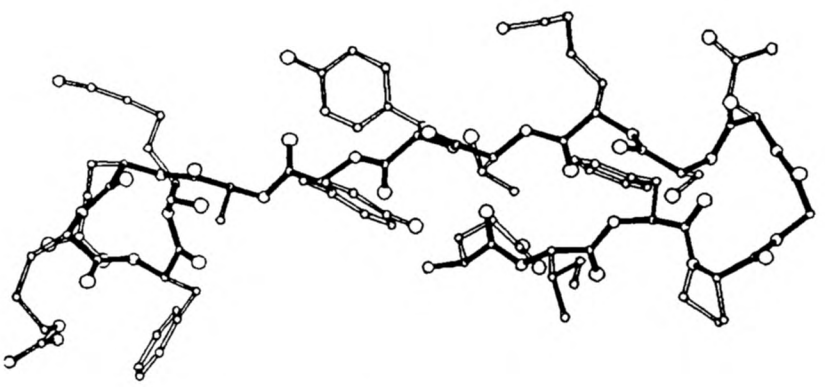


(a) CDR H2 from the P1a Fab

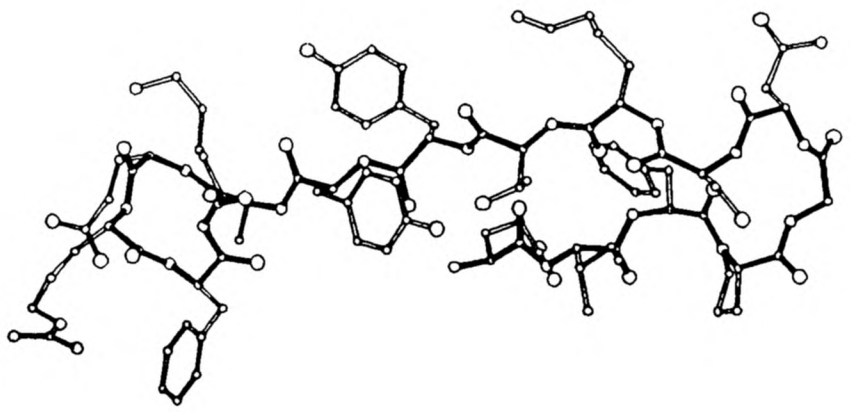


4Å

(b) CDR H2 from the P1b Fab



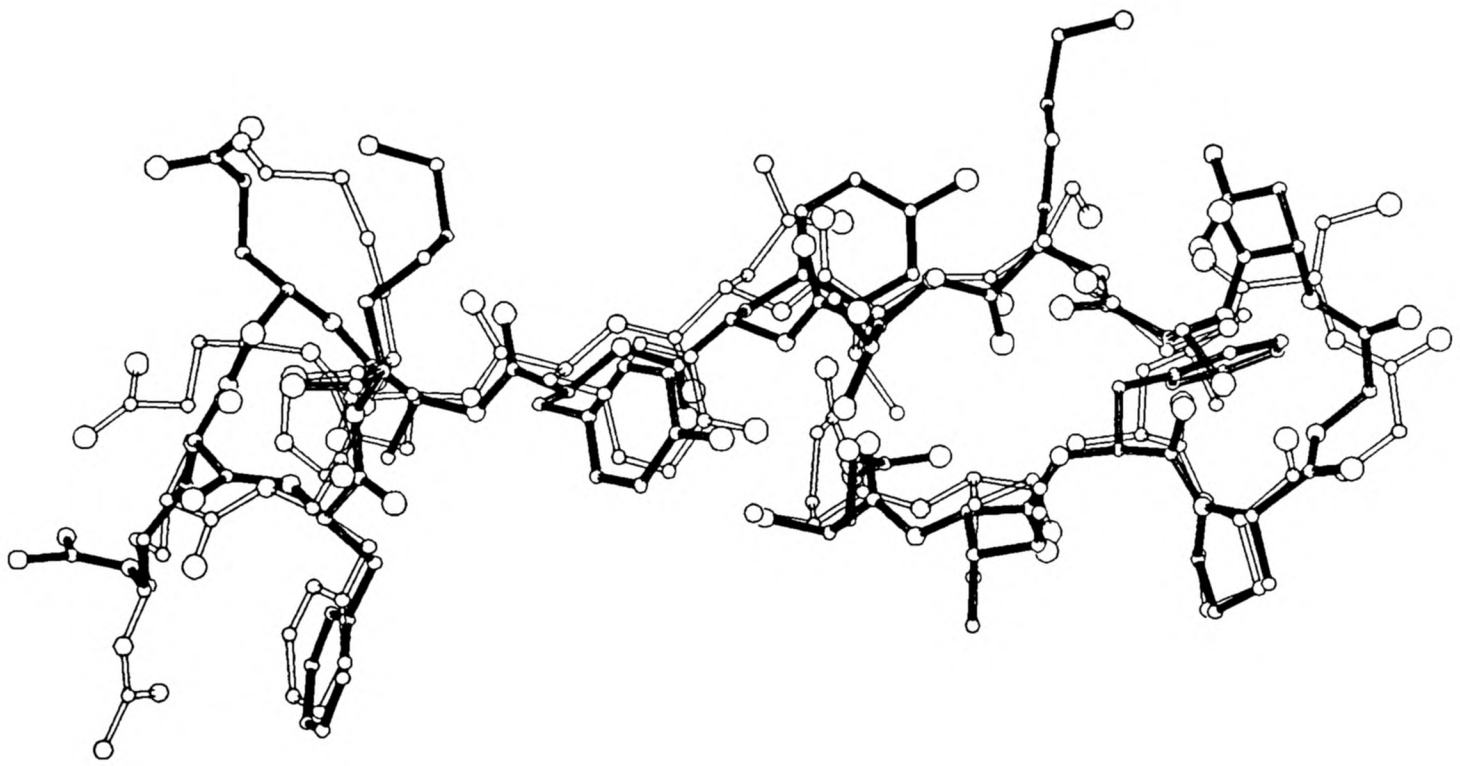
(c) CDR H2 from the $P2_1A1$ Fab



(d) CDR H2 from the $P2_1A2$ Fab

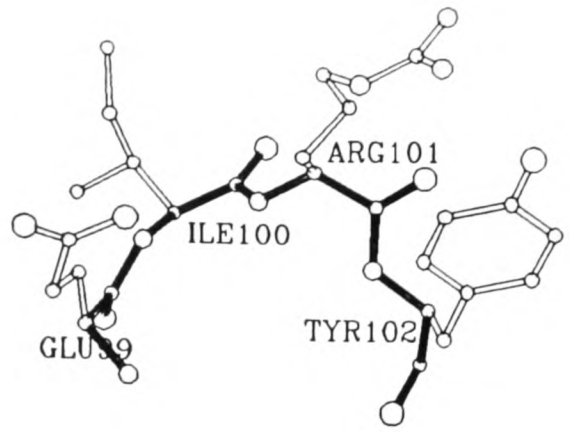
CDRs viewed from equivalent directions with respect to the whole V_H domain.

Figure 28: Comparison of the Conformations of CDR H2 in the Four Gloop2 Fabs



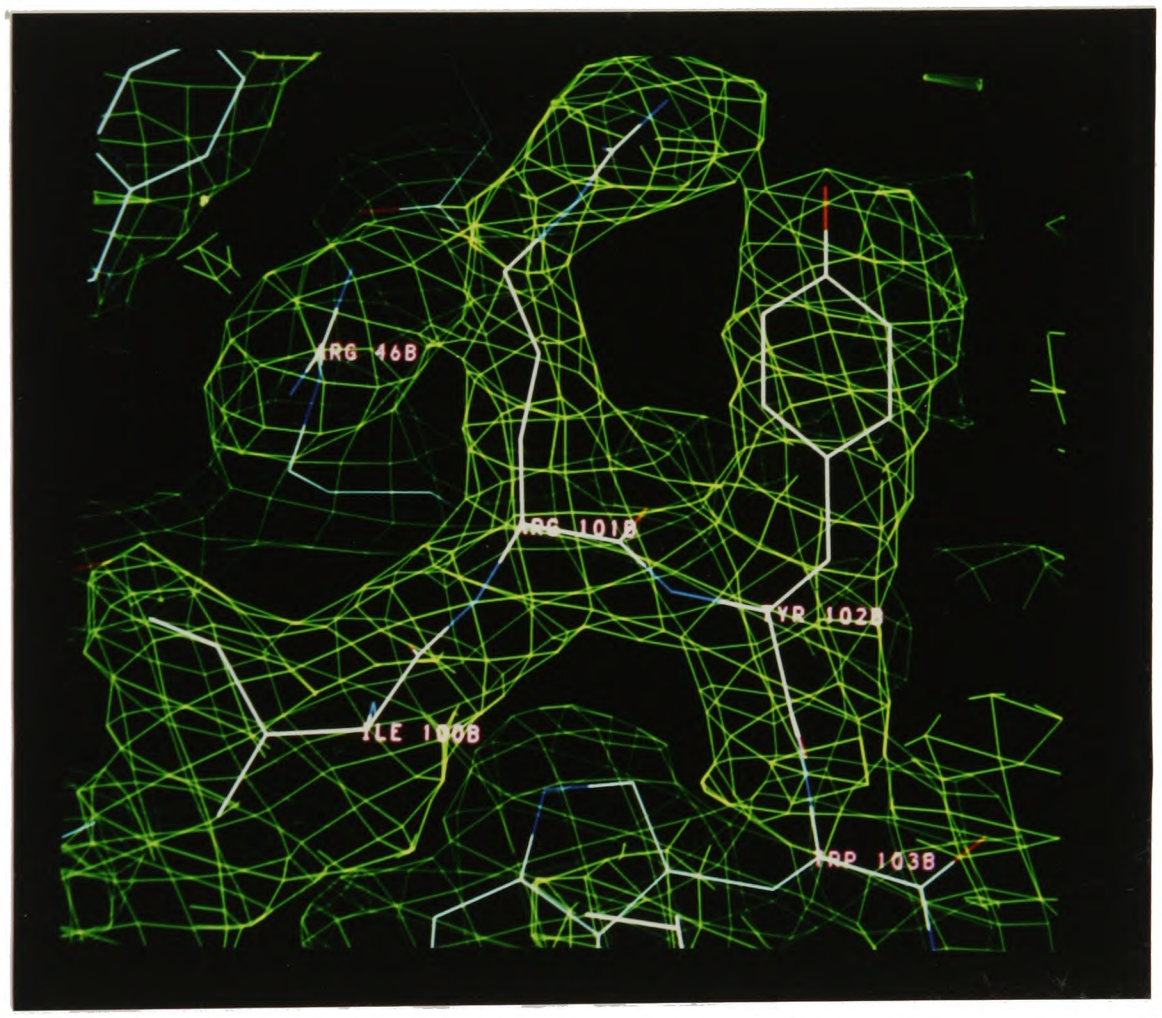
Superimposition of CDR H2 from P1a with CDR H2 from HyHEL-5
Superimposition of CDRs made on all $C\alpha$ atoms in the CDRs.

Figure 29: Comparison of CDR H2 from Gloop2 with that from HYHEL-5



4Å

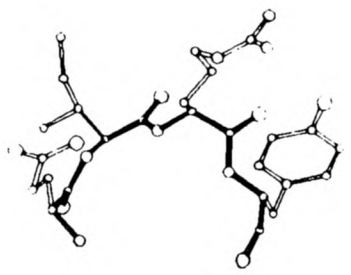
(a) CDR H3 from the P1a Gloop2 Fab



(b) $(2F_{obs} - F_{calc}), \alpha_{calc}$ omit map for some of the residues within CDR H3

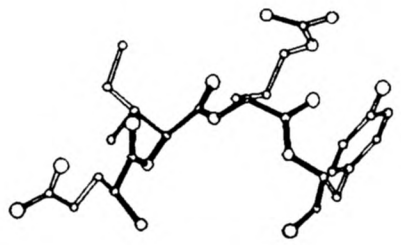
'Omit' electron density map calculated from P1a coordinates from which the CDR H3 residues had been omitted. The $(2F_{obs} - F_{calc}), \alpha_{calc}$ map is contoured at 0.6xRMS of the map above zero, with $F(000)$ omitted.

Figure 30: The Conformation of CDR H3 in the P1a Fab and Associated Electron Density

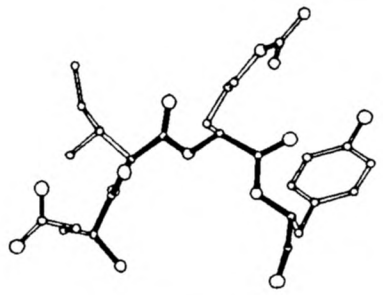


(a) CDR H3 from the P1a Fab

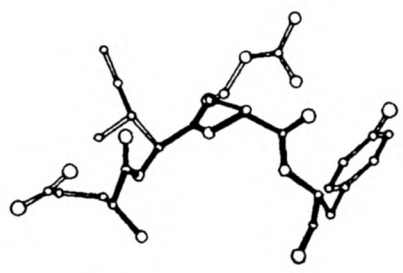
4Å



(b) CDR H3 from the P1b Fab



(c) CDR H3 from the $P2_1A1$ Fab



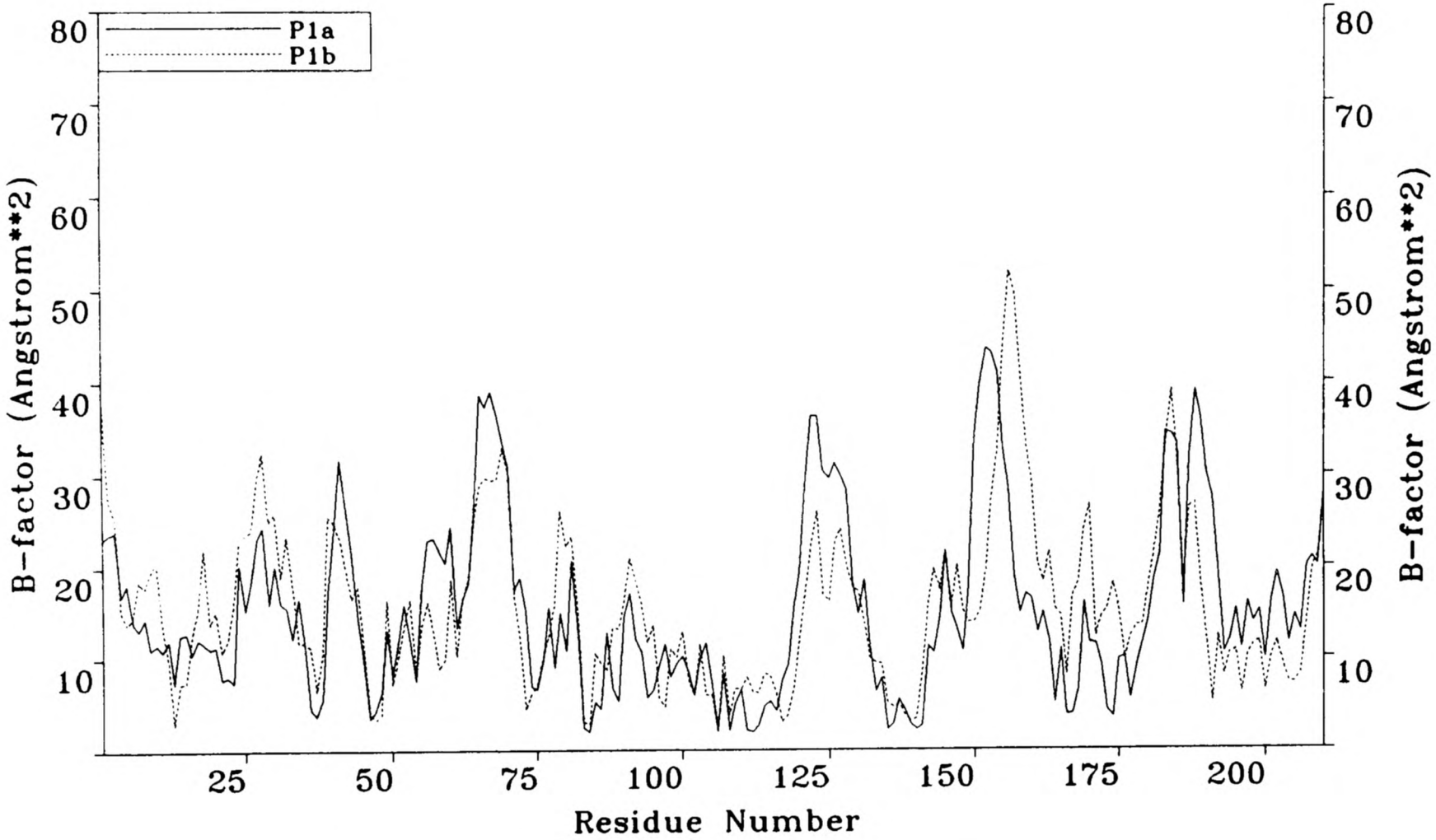
(d) CDR H3 from the $P2_1A2$ Fab

CDRs viewed from equivalent directions with respect to the whole V_H domain.

Figure 31: Comparison of the Conformations of CDR H3 in the Four Gloop2 Fabs

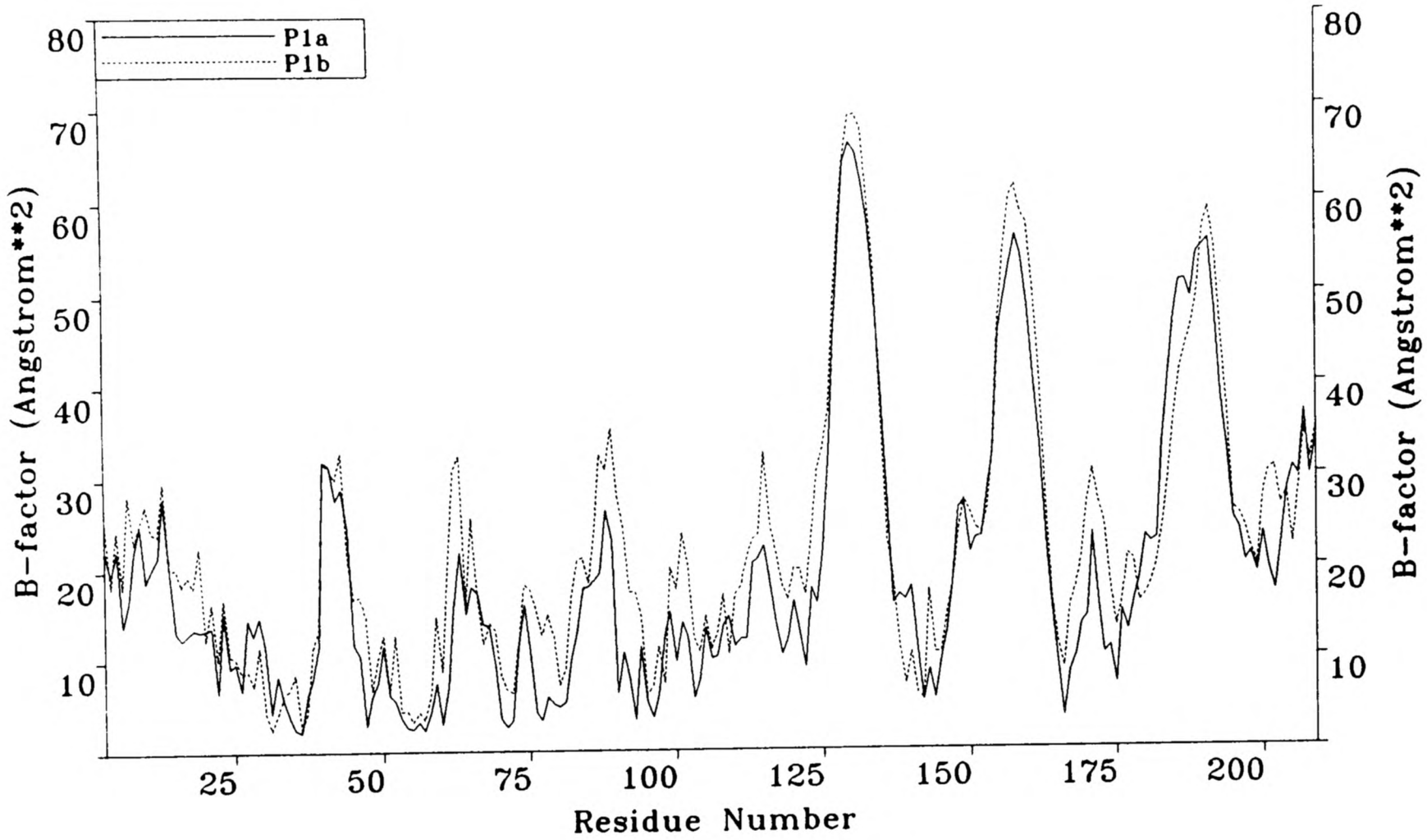
sequence from another antibody structure (discussed below). Within the limits of the resolutions of the studies the CDR conformations are very similar in the four Gloop2 Fabs. The most significant difference is the rotation of Phe52H and Tyr59H from CDR H2 in the $P2_1A1$ and $P2_1A2$ Fabs with respect to the orientations in the P1a and P1b Fabs. This can be explained in terms of crystal packing: in the $P2_1$ 'A' cells the C_L domain from a symmetry-related Fab packs on top of the V_H domain. These two changes are the only significant changes throughout the whole combining site between the Gloop2 Fabs, indicating that the CDRs are not a site of particular flexibility. The mean B factors for the P1a CDRs are 18.1, 14.0, 10.6, 5.0, 8.7 and 12.8Å^2 for CDRs L1, L2, L3, H1, H2 and H3 respectively (mean value taken over all atoms in each CDR). These compare favourably with the mean values of 15.5Å^2 and 19.5Å^2 for the V_L and V_H domains respectively.

Since individual B factors have only been determined for the two Fabs in the $P1$ crystal form, it is only possible to correlate the B factors between the P1a and P1b Fabs. As the refinement stood at the end of cycle 25 (Chapter 6) the B factors had been refined with non-crystallographic symmetry restraints imposed ($\sigma_B = 2.0$). In order to obtain less biased values for the B factors the current P1a and P1b models had their mean B factors set to 15.0Å^2 for all atoms (except hydrogens) and individual B factors were then refined for each chain in the absence of non-crystallographic symmetry restraints. The B factors were restrained along the chain by $\sigma_{12} = 1.5$ and $\sigma_{13} = 2.0$ (see Chapters 2 and 6 for definitions of these parameters). This type of B factor refinement was not entirely unbiased since these B factors were determined for a structure which was refined with B factors in place which *were* restricted by non-crystallographic symmetry, but at least it gave the opportunity for the B factors to deviate between the chains. The results of the B factor refinement is given in Figures 32 and 33 as the mean B factor for



Mean *B* factor for each residue in the light chain (L) plotted against residue number.
Mean *B* factors for whole chain — P1a = 15.5Å²; P1b = 15.8Å².
Correlation Coefficient = 0.6540
(Form of the correlation coefficient is defined in the text)

Figure 32: Correlation Between Thermal Factors in Fabs P1a and P1b — Light Chain



Mean *B* factor for each residue in the heavy chain (H) plotted against residue number.
Mean *B* factors for whole chain — P1a = 19.5Å^2; P1b = 22.8Å^2.
Correlation Coefficient = 0.9343
(Form of the correlation coefficient is defined in the text)

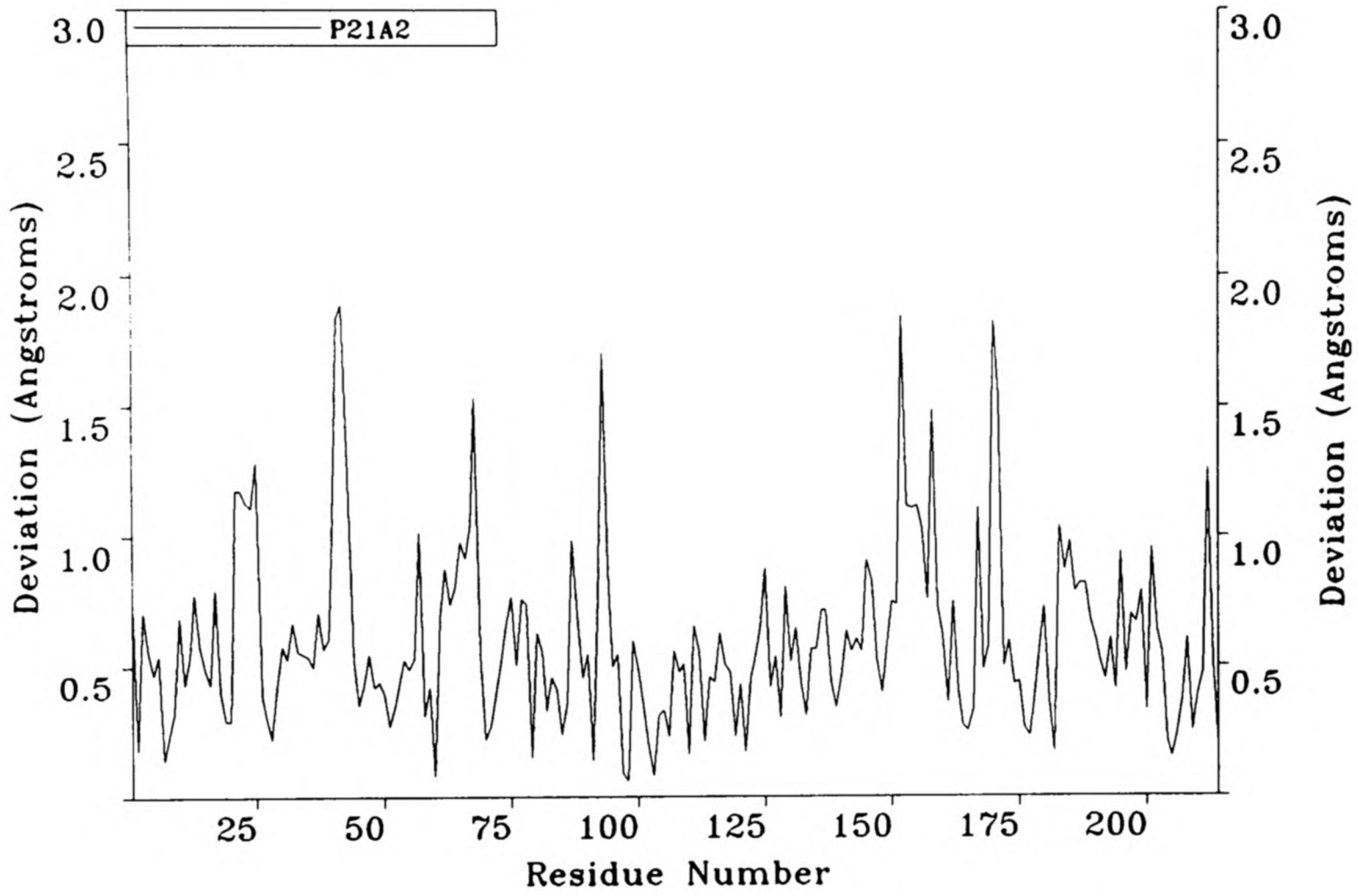
Figure 33: Correlation Between Thermal Factors in Fabs P1a and P1b — Heavy Chain

each residue for the heavy and light chains plotted against residue number. The agreement between the B factors on the two Fabs was obvious even visually, and is confirmed by the correlation coefficients (65.4% for the light chain, 93.4% for the heavy chain). The form of the correlation coefficient used was:

$$CC = \frac{\sum(B_1 - \langle B_1 \rangle) * (B_2 - \langle B_2 \rangle)}{\sqrt{\sum(B_1 - \langle B_1 \rangle)^2 * \sum(B_2 - \langle B_2 \rangle)^2}}$$

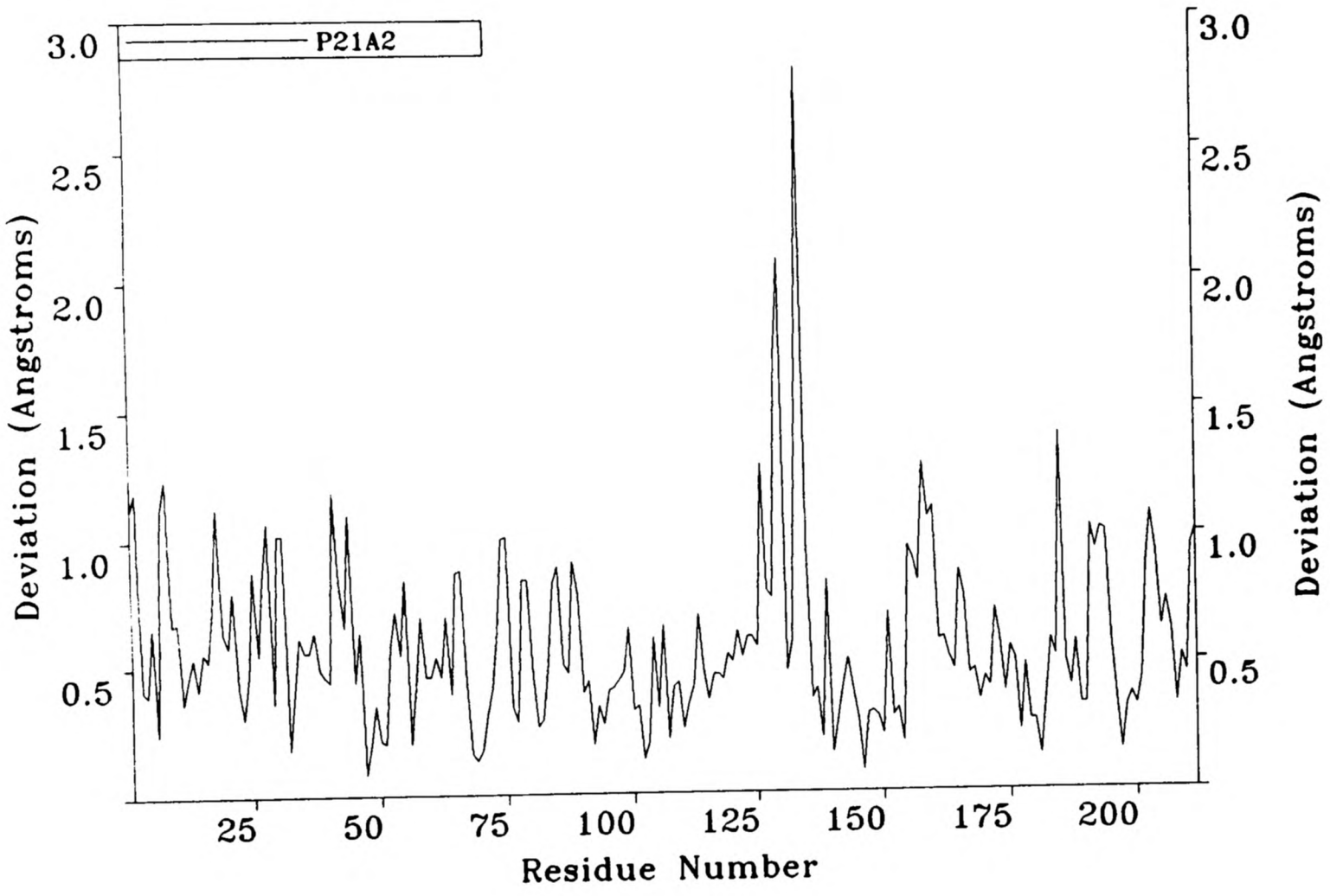
It would appear that the B factors are not significantly affected by differences in the crystal environments of the two Fabs. The B factor refinements of Foot and Mouth Disease Virus and Tumour Necrosis Factor using X-PLOR also show good correlation between chains related by non-crystallographic symmetry (Taylor, 1989b). In contrast Sheriff *et al.* (1985) found that lattice and oligomeric contacts decreased the B factors in the myohemerythrin monomer and hemerythrin octamer. The three major peaks in the heavy chain corresponded to the loops in C_H1 referred to above.

The sources of the non-isomorphism between the two $P2_1$ 'A' form cells were both rigid-body rotations and small changes in the polypeptide chain conformation. The V_L and V_H domains moved by 1.2° and the C_L and C_H1 domains moved by 1.3–1.4°, between the two Fabs ($P2_1A1$ and $P2_1A2$). Although the rotations for each domain were similar in magnitude, the direction of the rotation axes was different: the V_L:V_H domain pairing changed by 0.85° and the C_L:C_H1 pairing by 2.37° between the two Fabs (determined by superimpositions on all $C\alpha$s *including* CDRs, unlike the values in Tables 5 and 6). However the relative orientations of the intermolecular two-folds were only slightly altered (elbow bend in $P2_1A1$ is 151.9° and in $P2_1A2$ is 152.0°). In addition to these rigid body movements there was some movement in the loop regions of the domains. These changes were small translational displacements of parts of the loops which otherwise maintained their



Distance between $C\alpha$ atoms at equivalent positions in the sequence plotted with respect to residue number.
Superimpositions performed on all $C\alpha$s from residues 1L–214L in the Gloop2 sequence (Appendix 1), including CDRs.

Figure 34: Deviations Between the Gloop2 $P2_1A1$ and $P2_1A2$ Fabs



Distance between $C\alpha$ atoms at equivalent positions in the sequence plotted with respect to residue number.
Superimpositions performed on all $C\alpha$s from residues 1H–212H in the Gloop2 sequence (Appendix 1), including CDRs.

Figure 35: Deviations Between the Gloop2 $P2_1A1$ and $P2_1A2$ Fabs — Heavy Chain

		Superimpositions		
Domain	Residues	$C\alpha$ (Å)	B/B (Å)	ALL (Å)
V_L	1L–105L	0.67	0.76	0.97
C_L	111L–214L	0.63	0.74	0.97
V_H	1H-111H	0.61	0.75	0.97
C_H1	117H-212H	0.71	0.85	1.01

B/B = C, O, N, $C\alpha$.
ALL = all atoms except hydrogens.

Table 8: Comparison of Fabs $P2_1A1$ and $P2_1A2$

local conformations in the two Fabs. Figures 34 and 35 show the distances between equivalent $C\alpha$ atoms in the two Fabs, on the superimposition of whole domains. The RMS fits on superimposing each domain in the $P2_1A1$ Fab onto the corresponding $P2_1A2$ domain are given in Table 8.

7.2.5 Comparisons With Other Fabs

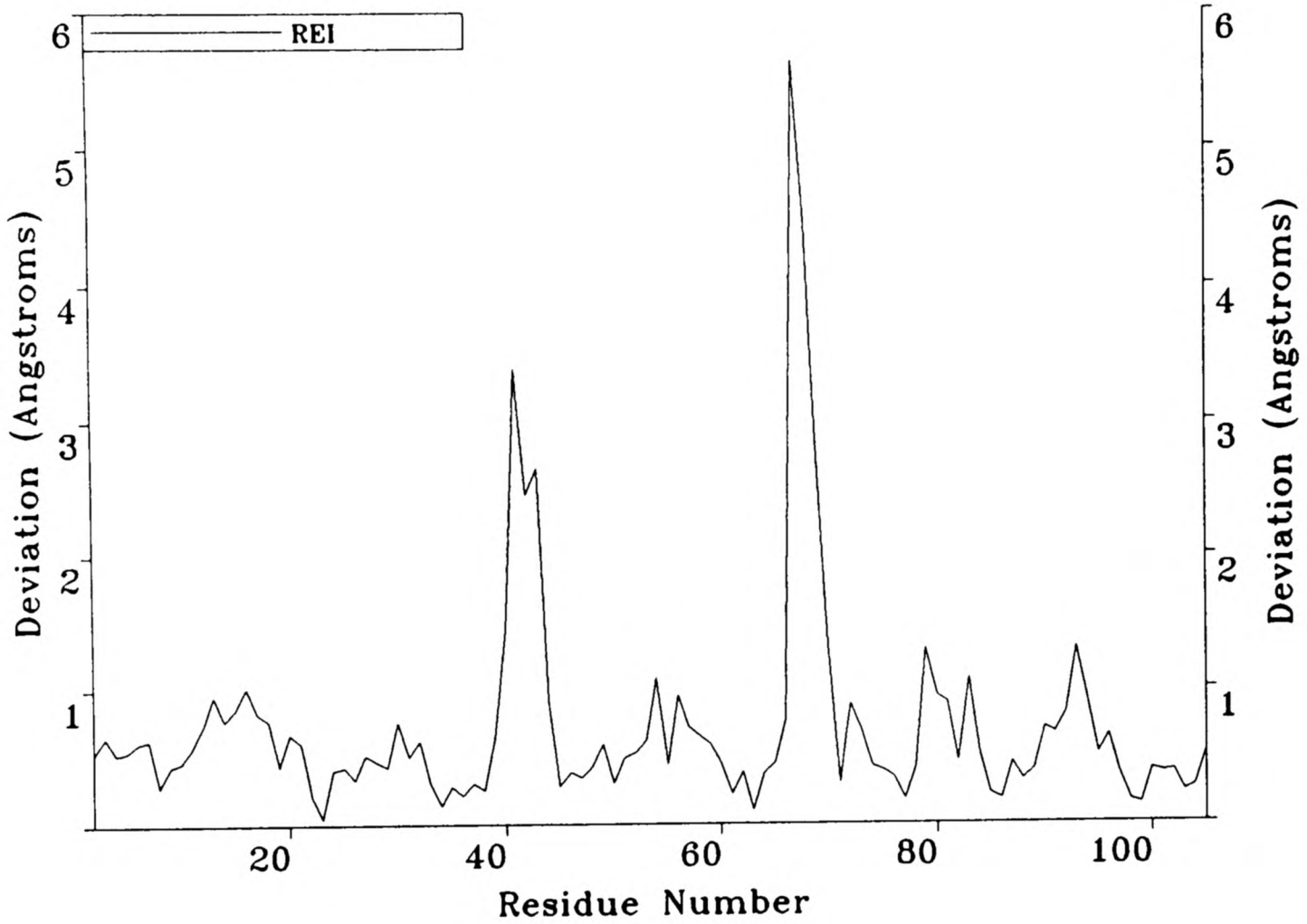
The molecular replacement solution of the Gloop2 crystal structures (Chapter 4) established the similarity in tertiary and quaternary structure between the Gloop2 Fabs and existing Fab structures. The quaternary structure of the Gloop2 Fabs has been compared to that of other Fabs in previous sections. In this section the domain and CDR conformation is compared to a representative antibody. The antibodies used in the comparisons were chosen on the basis of maximum sequence identity with Gloop2. For the purposes of comparing V_L and the light chain CDRs the 'A' chain from the Rei V_L dimer (Epp *et al.*, 1975) was chosen. The Rei V_L also has the convenient property of being the same length as the Gloop2 V_L. The V_H, C_L and C_H1 domains were compared to the corresponding segment of the HyHEL-5 chain (Sheriff *et al.*, 1987). CDR H1 was compared to the H1 of D1.3 (Amit *et al.*, 1986; coordinates supplied by Dr. Roberto Poljak) and CDR H2 was compared to the HyHEL-5 H2. CDR H3 in Gloop2 has no structural homologue in other crystal structures. The Gloop2 P1a Fab was used for the comparison in each case.

The comparison of the P1a V_L with the Rei V_L was particularly gratifying. Inspection of the least-squares superimposition on all $C\alpha$ atoms in the domain including CDRs (RMS = 1.12Å) showed that the orientation of the peptide planes was essentially identical between the Rei (2.0Å resolution) structure and the Gloop2 P1a (2.8Å resolution) structure. The deviation in $C\alpha$ positions between structurally equivalent residues is shown in Figure 36. The agreement is very close except for two

regions: 40L–43L and 65L–70L (Gloop2 numbering). The former of these regions is the elbow bend loop which also deviates between the Gloop2 structures. This loop may be affected by the presence of the C_L:C_H1 domains in Gloop2 (they are absent in Rei) or this may reflect an error in chain tracing. The other deviation is considerably more interesting. This loop (65L–70L) is a non-CDR loop that packs against CDR L1 at the antigen-binding end of V_L. The sequences for this region for the solved antibody structures fall into two classes: Gly at 65L and Tyr at 71L for κ chains (HyHEL-5 (Sheriff *et al.*, 1987); McPC603 (Satow *et al.*, 1986); J539 (Suh *et al.*, 1986)) and Lys at 65L and Ala at 71L for λ chains (New (Saul *et al.*, 1978); Kol (Marquart *et al.*, 1980)), numbered according to the Gloop2 sequence. The sequence of Gloop2 (κ chain, Appendix 1) is different from both, with Arg at 65L and Tyr at 71L. The conformation of the Kol loop, Rei loop and the Gloop2 loop are shown in Figure 37. In Gloop2 Arg65L packs between the 65L–71L loop and CDR L1 in a similar manner to the lysine of Kol. However the presence of Tyr71 in a similar position to that observed for the Rei V_L causes the loop to deviate from both the previously observed conformations in V_κ and V_λ domains. The *cis*-prolines at positions 8L (framework) and 95L (CDR L3) are also found to be *cis* in the Rei V_L.

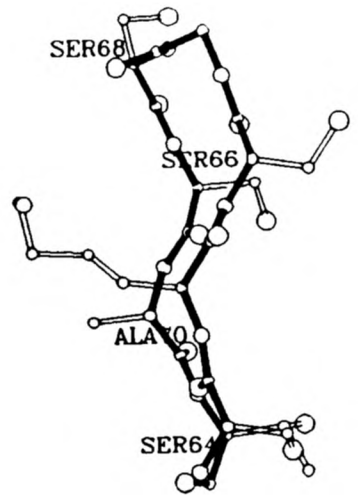
Superimpositions of the Rei CDRs onto the Gloop2 P1a CDRs (Figures 17,20,23) using the $C\alpha$ atoms shows marked similarity in all three cases. The major deviation in CDR L1 is the main-chain where replacement of Gly (in Gloop2) for Lys (in Rei) at residue 27L (Gloop2 numbering) allows adoption of less strained $\phi\psi$ angles for the preceding residue in Gloop2 (Ser26L has $(\phi,\psi) = (40°, 45°)$ in the P1a Fab, Ile26L has $(79°, -78°)$ in Rei). The other two CDRs are very similar to their Rei counterparts, even sharing similar side-chain conformations.

The deviations in $C\alpha$ positions between the of the V_H of HyHEL-5 onto the P1a



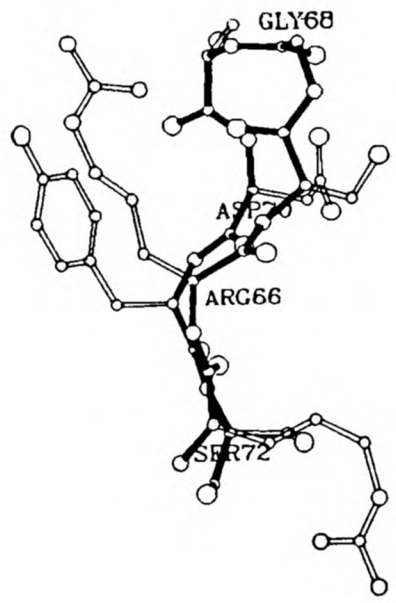
Distance between $C\alpha$ atoms at equivalent positions in the sequence plotted with respect to the Gloop2 residue number. Superimpositions performed on all $C\alpha$s from the V_L domain of Fab P1a and the 'A' chain of Rei (Epp *et al*, 1975).

Figure 36: Deviations Between Gloop2 (P1a Fab) and Rei V_L Domain

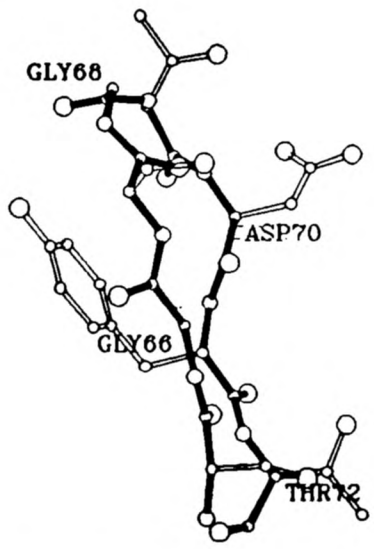


NEW

4Å



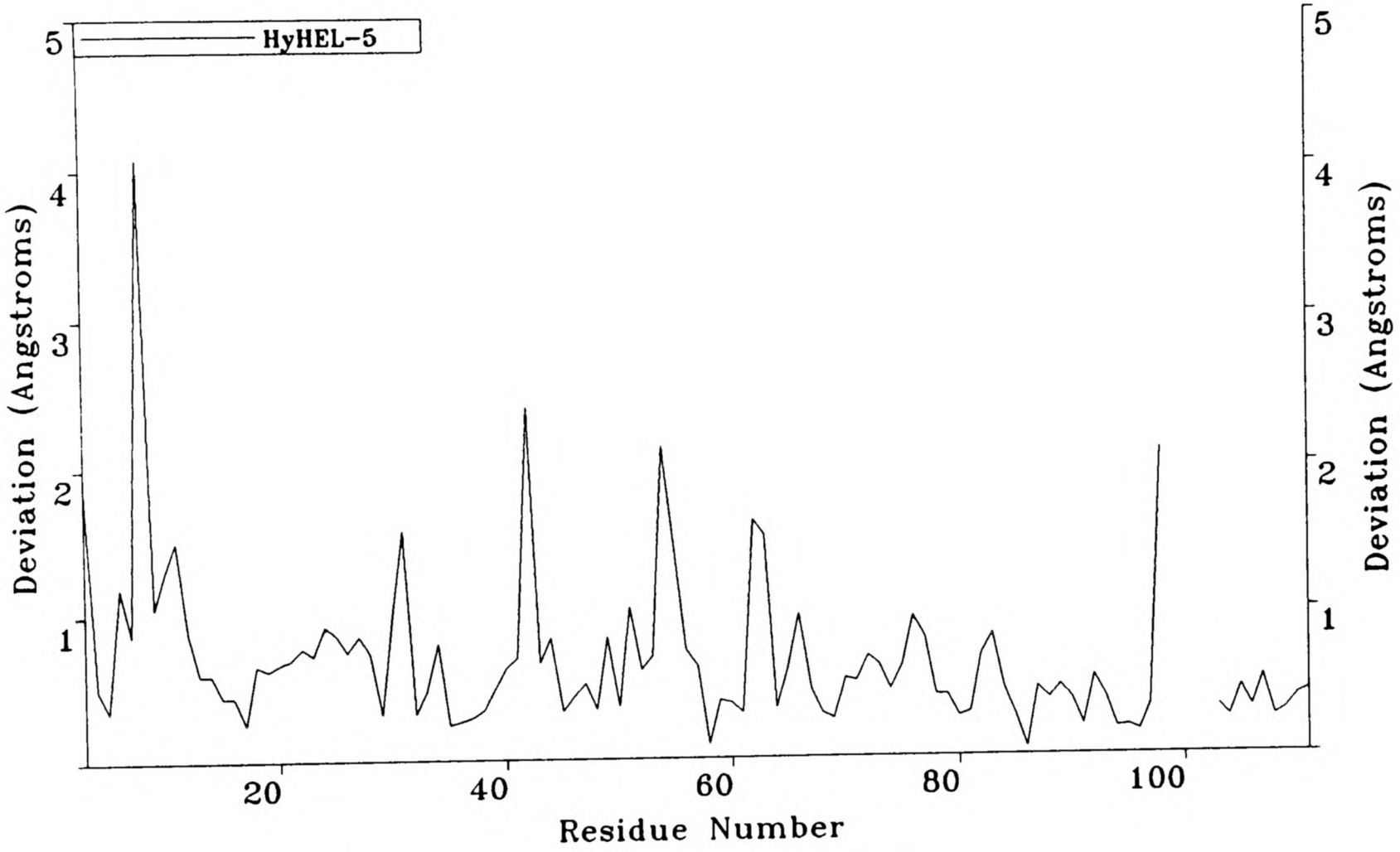
GLOOP2 P1a



REI

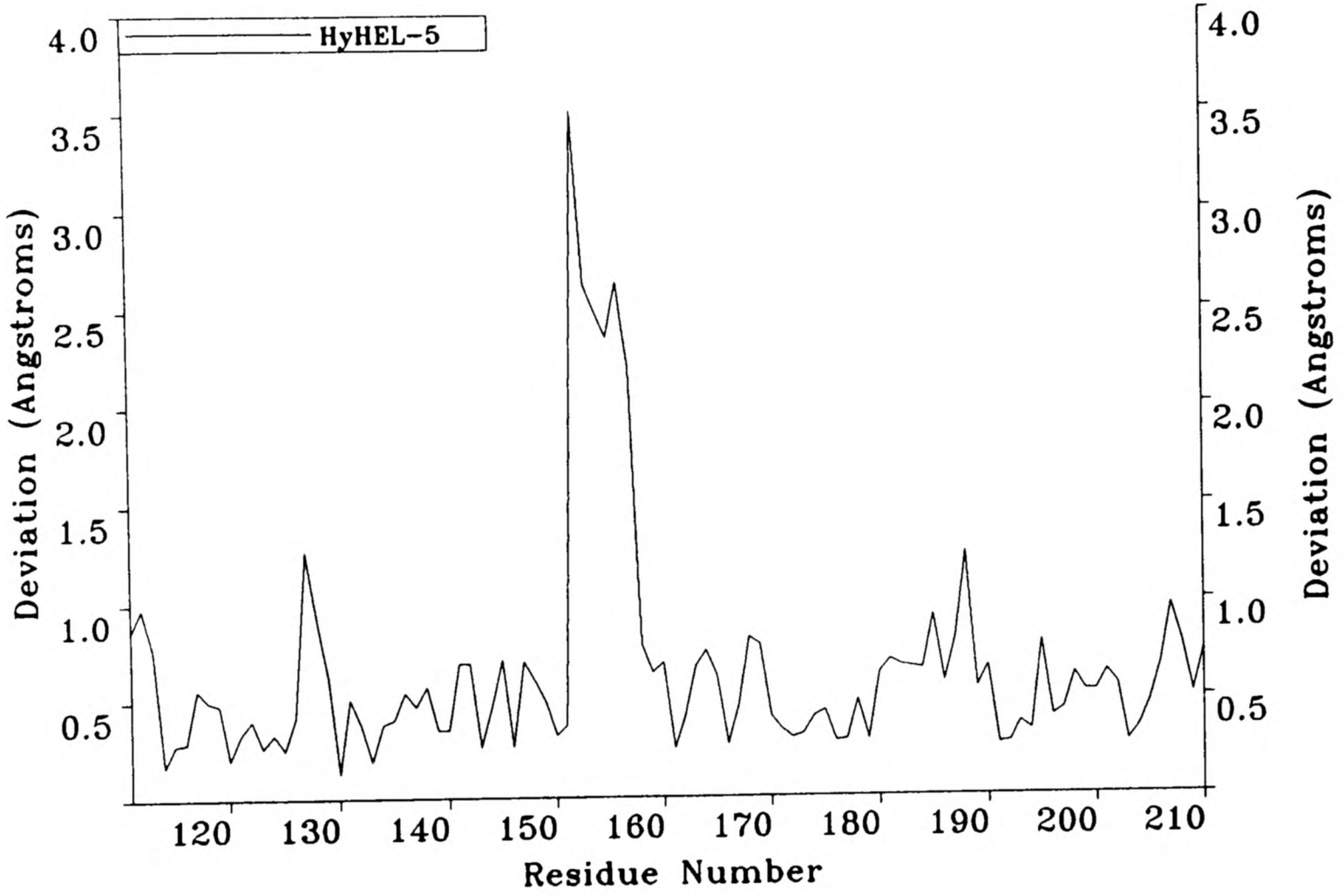
All images from equivalent viewpoints with respect to V_L
CDR L1 is to the left of the loops as viewed here

Figure 37: Conformation of the 65–70L Loop in Rei, New and Gloop2



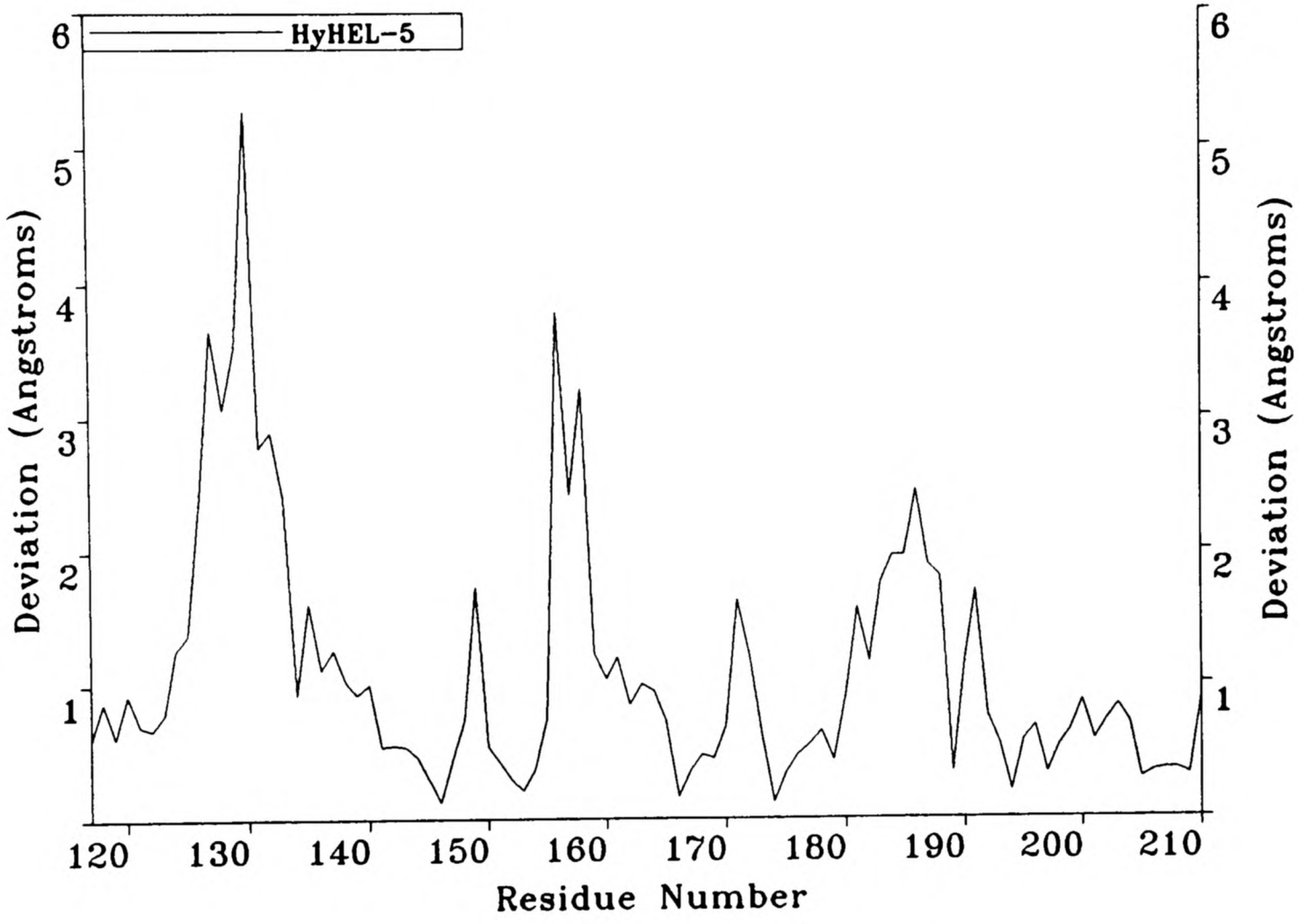
Distance between $C\alpha$ atoms at equivalenced positions in the sequence plotted with respect to the Gloop2 residue number. Superimpositions performed on equivalenced $C\alpha$s from the V_H domain of Fab P1a and the HyHEL-5 V_H (Sheriff *et al.*, 1987).

Figure 38: Deviations Between Gloop2 (P1a Fab) and HyHEL-5 V_H Domain



Distance between $C\alpha$ atoms at equivalenced positions in the sequence plotted with respect to the Gloop2 residue number. Superimpositions performed on equivalenced $C\alpha$s from the C_L domain of Fab P1a and the HyHEL-5 V_H (Sheriff *et al.*, 1987).

Figure 39: Deviations Between Gloop2 (P1a Fab) and HyHEL-5 C_L Domain



Distance between $C\alpha$ atoms at equivalenced positions in the sequence plotted with respect to the Gloop2 residue number. Superimpositions performed on equivalenced $C\alpha$s from the C_H1 domain of Fab P1a and the HyHEL-5 V_H (Sheriff *et al.*, 1987).

Figure 40: Deviations Between Gloop2 (P1a Fab) and HyHEL-5 C_H1 Domain

V_H are shown in Figure 38. The RMS fit on all $C\alpha$ atoms is 0.87Å (excluding CDR H3). Deviations occur in the loop regions around residues 8H and 42H (the latter in a position of structural diversity between Gloop2 Fabs, see Figure 12) and within the three CDRs (31H–35H, 50H–65H and 99H-102H). Deviations are expected for CDR H3 as the HyHEL-5 H3 is longer than Gloop2 and does not have a similar conformation. Comparison of the conformation of the Gloop2 and D1.3 CDR H1 (Figure 26) shows close similarity: the major deviation is between the ring positions of the phenylalanine and tyrosine. The comparison between the CDR H2 of Gloop2 and HyHEL-5 is given in Figure 29. There are obvious differences at 54H and 55H and similarly at 61H–65H (although this is difficult to interpret from the figure). From the comparison with HyHEL-5 it would appear that Pro53H in Gloop2 has been erroneously built in the *cis*-configuration. In both J539 (Suh*et al.*, 1986) and HyHEL-5 (Sheriff *et al.*, 1987) this proline is trans. This error may be sufficient to account for the deviations in residues 54H and 55H.

The superimpositions of the HyHEL-5 C_L and C_H1 domains (Figures 39 and 40) have a strong similarity to the equivalent plots of structural deviations between the Gloop2 C_L and C_H1 domains. The RMS deviations on all $C\alpha$ atoms is 0.86Å for C_L and 1.45Å for C_H1. The latter figure is larger due to the three regions of significant deviation. The homology in the framework regions in C_H1 is comparable to that between the C_L domains, and the major deviation between the HyHEL-5 and Gloop2 C_L domains is the outer loop region 152L-157L. The deviations between the C_H1 domains are in the regions around 130H, 157H and 185H. The *cis*-prolines at 141L in C_L, and at 147H in C_H1 are also *cis* in the HyHEL-5 structure.

7.3 Antigen Binding

In the absence of the Gloop2:HEL and Gloop2:loop complex structures it is difficult to make any definite conclusions about the nature of the interaction of Gloop2 with the loop peptide (whose structure in solution is not known) or with HEL.

Experimental evidence for the mode of antigen binding comes from three sources: NMR data on the binding of the loop peptide to the Gloop1 Fab (J. C. Cheetham *et al.*, manuscript in preparation); the serological mapping of the Gloop2 epitope using a panel of lysozymes (Darsley and Rees, 1985a); and mutations in the CDRs of the Gloop2 Fab (Roberts *et al.*, 1987; Roberts and Rees, 1986; Sally Roberts, unpublished results; Kate Hilyard, unpublished results).

Serological mapping of the Gloop2 epitope on HEL involved the determination of the binding affinities of Gloop2 for nine lysozyme species (Darsley and Rees, 1985a). Species which had no mutations with respect to HEL in the loop region, but mutations elsewhere on the protein, bound to Gloop2 with the same affinity as HEL. Turkey Egg Lysozyme (mutation Arg73→Lys) bound Gloop2 with the same affinity as HEL, but Ring-Necked Pheasant Egg Lysozyme (Arg73→Lys, Asn77→His) showed decreased affinity. Bob-white Quail Egg Lysozyme (Arg68→Lys) also had decreased affinity for Gloop2. The epitope suggested for Gloop2 binding to HEL involved Arg68 and Asn77 and the region 69-73, inferred from the lack of affinity of a proteolytic fragment of the pep1 peptide, which did not have these residues, for Gloop2.

The structures of Fabs and Fab:antigen complexes in solution are inaccessible to the technique of 2-dimensional NMR at present. Molecules of that size tumble too slowly in solution and the resulting resonances are too broad to be useful. The

Mutant	Pep1	HEL
CDR L1		
Q27L	NC	↑
E28S†	NC	↑
E28I	NC	↑
E28R	NC	NC
S30A	NC	↑
E28SK56Q†	↑	↑
Q27LE28IS30A	NC	↑
Y32F‡	NC	NC
Y32S‡	↓	ND
Y32D‡	↓	ND
Y32DL92D‡	↓	ND
CDR L3		
L92A	↓	NC
L92D	↓	ND
L92SS93Q	↓	NC
L92NS93Q	↓	NC
L92QS93Q	↓	NC
CDR H1		
E50Q	NC	ND
E50R	NB	ND
E50S†	NB	ND
E50RR101E	NB	ND
K56Q	NC	NC
CDR H3		
E99R	NB	ND
E99Q	↓	ND
E99D‡	NB	ND
R101E	NC	NC
R101Q	NC	NC

NC — no change in binding affinity.
NB — no detectable binding.
ND — not determinined.
↑ — increase in binding affinity.
↓ — decrease in binding affinity.
(Sally Roberts, unpublished results)
† (Roberts *et al.*, 1987; Roberts and Rees, 1986)
‡ (Kate Hilyard, unpublished results)

Table 9: Site-Directed Mutagenesis of Residues in the Gloop2 ACS

broadening of these resonances has been used in order to map the epitope of the loop peptide as recognised by the Gloop1 Fab. The resonances from each residue of the loop peptide in solution are observable since the peptide tumbles rapidly. When the Gloop1 Fab binds to a residue in the peptide the resonances from that residue broaden as the residues tumbles at the same rate as the Fab. In principle, any residues not part of the loop/Gloop1 epitope should remain conformationally labile and these resonances will not disappear from the loop peptide spectrum. Whether this will be the case depends on the type of residue and the conformational restraints imposed upon it by neighbouring residues that might be bound to the Fab. For instance if residues either side of a glycine are bound to the antigen, but not the glycine itself, this may be impossible to detect. If, however, the residue was a lysine the flexibility of the long side-chain would give clear indication that that lysine (or at least the lysine side-chain) was not involved in antibody binding. This method is especially valuable for peptide antigens, but is less applicable to protein antigens where the lower degree of conformational flexibility and increased size limits its usefulness.

The Gloop1 and Gloop2 epitopes are indistinguishable by serological mapping techniques, and the two Fabs are very homologous in the variable regions (V_L domains have 99% identity, V_H domains 91% identity; sequences in Appendix 1). The binding of loop peptide should therefore be similar between the Gloop1 and Gloop2 antibodies. These epitope mapping experiments partially contradict the serological data. The region 71–73 on the loop peptide is seen as mobile in the complex and inferred as not participating in antibody:antigen interactions. The N- and C-termini (57-59 and 80-84) are mobile also but the rest of the residues are apparently involved in antigen binding. The two HEL active site tryptophans, 62 and 63, appear to be implicated by this method, despite the fact that Gloop2 does not inhibit HEL

activity (Darsley and Rees, 1985a).

Table 9 provides a list of the binding affinities of the mutant Gloop2 Fab for the pep1 peptide (loop peptide plus 31–35 of the HEL sequence) and HEL. The loop peptide and the pep1 peptide bind to Gloop2 with comparable affinity (Darsley and Rees, 1985a). The methods used to produce the mutants and to determine the binding affinities have been previously reported (Roberts *et al.*, 1988). Mutations of Tyr32L in CDR L1 indicate that the phenolic hydroxyl is not important in antigen binding (Y32→F mutant) but that the presence of the aromatic ring is important in pep1 recognition. The remainder of CDR L1 appears not to participate in the binding of pep1, and mutations in this area which reduce its hydrophilicity increase the affinity for HEL antigen. Mutation of Leu92L (the solvent-exposed leucine in CDR L3) appears to affect pep1 binding only. These two sets of mutations illustrate that pep1 and HEL bind to Gloop2 in different ways: whilst the CDR L1 results could simply be attributed to the absence of the appropriate part of polypeptide chain in the pep1 antigen, the differing importance of Leu92L is a clear indication of dissimilar recognition. The results for CDR L1 also tie in with the origins of Gloop2: since Gloop2 was an antibody produced in response to the loop antigen (Darsley and Rees, 1985a), CDR L1 does not optimally match the conformation/charge distribution of the regions of the intact protein not present in the loop peptide. The suggestion that CDR L1 does not interact with antigen is not without precedent: in the NC41 and NC10 Fab:neuraminidase complexes CDR L1 made no contact with the antigen (Tulip *et al.*, 1989).

Mutation results for CDRs H1 and H3 indicate the importance of Glu50H and Glu99H in pep1 binding. The presence of two carboxylate groups at the centre of the ACS suggests that salt bridges may play a part in antigen recognition. The structure of the HyHEL-5 complex, whose epitope partially overlaps the proposed Gloop2

epitope (Chapter 1), contains two salt bridges between Arg68 and Arg45 in HEL and Glu35H (CDR H1) and Glu50H (CDR H2) in HyHEL-5 (Sheriff *et al.*, 1987). Arg68 was implicated in the Gloop2 epitope by serological mapping (Darsley and Rees, 1985a) and the mutation of Arg68→Lys reduces antigen binding by a factor of 30 in Gloop2 (Roberts, 1987) and 1000 in HyHEL-5 (Smith-Gill *et al.*, 1982). The superimposition of the Gloop2 V_H onto that of HyHEL-5 indicates that Glu50H in both structures appears at the same position with respect the the V_H domain (the equivalent residue to Glu35H in HyHEL-5 is Thr35H in Gloop2). However Glu35H from HyHEL-5 and Glu99H from Gloop2 are not in similar positions: the $C\delta$–$C\delta$ distance between Glu50H and Glu35H in HyHEL-5 is 4.27Å, but the equivalent distance between Glu50H and Glu99H in Gloop2 is 7.83Å for the P1a Fab and 5.44Å in the P1b Fab. The difference in the separation of the glutamates in Fabs P1a and P1b has no obvious reason from inspection of omit maps. However the position of the solvent molecule bound to Glu50H in the P1a Fab is occupied by the NH_3^+ group of Lys169L of a symmetry-related P1a Fab and this compensation of the charge of Glu50H by Lys169L may allow closer approach of Glu99H in the P1b Fab than in the uncompensated P1a Fab. By the same token the salt-bridges between the glutamates in HyHEL-5 and Arg45 and Arg68 from HEL may explain why these two glutamates can approach closer still.

Despite giving valuable evidence as to some aspect of the interactions between Gloop2 and the loop peptide and HEL antigens there is still no indication of the conformation of HEL or peptide in the complex. The presence of the two glutamates and the effect of their mutation on binding suggests some salt-bridge interaction in the complex, although not necessarily similar to that of HyHEL-5. Unlike HyHEL-5, Gloop2 does not inhibit the lysis of *M.lysodeikticus* bacteria *in vitro* (Smith-Gill *et al.*, 1982; Darsley and Rees, 1985a). The mutagenesis data indicates that

the peptide binding involves mainly those residues in the groove, whilst residues interacting with HEL may extend outside the groove.

Any suggestions on the Gloop2:antigen will remain a matter for speculation until the crystal structures become available. There has been no success with the Gloop2:HEL crystallisations, but small Gloop2:peptide crystals have been grown (R. E. Griest, unpublished results). The sensitivity of these crystals to radiation damage will require flash freezing techniques for collection of a native dataset, but the Gloop2:peptide complex structure may yet become available.

APPENDIX 1

SEQUENCES OF THE GLOOP ANTIBODIES AND SEQUENCE ALIGNMENTS

A. Sequences of the Gloop Variable Regions

Variable region sequences include elbow region sequence. Variable region sequences for Gloop1–5 taken from (Darsley and Rees, 1985b). Gloop2 V_L and V_H sequences taken from Roberts (1987). Numbering corresponds to Gloop2 sequential numbering. Blanks spaces are unsequenced regions, dashes are insertions to optimise manual alignment.

V_L

	1																							
Gloop1	D	I	Q	M	T	Q	S	P	S	S	L	S	A	S	L	G	E	R	V	S	L	T	C	R	A
Gloop2	D	I	Q	M	T	Q	S	P	S	S	L	S	A	S	L	G	E	R	V	S	L	T	C	R	A
Gloop4						Q	I	P	L	S	L	P	V	S	L	G	D	Q	A	S	I	S	C	R	S
Gloop5	G	I	V	M	S	Q	S	P	S	S	L	A	V	S	V	G	E	K	V	T	M	S	C	K	S

	26																							
Gloop1	S	Q	E	I	S	G	–	–	–	–	–	–	Y	L	S	W	L	Q	Q	K	P	D	G	T	I
Gloop2	S	Q	E	I	S	G	–	–	–	–	–	–	Y	L	S	W	L	Q	Q	K	P	D	G	T	I
Gloop4	S	Q	S	I	V	H	–	S	N	G	N	T	Y	L	E	W	Y	Q	Q	K	P	G	Q	T	P
Gloop5	S	Q	S	L	F	Y	S	S	N	Q	K	N	S	L	A	W	Y	Q	Q	R	P	G	Q	S	P

	45																							
Gloop1	K	R	L	I	Y	A	A	S	T	L	D	S	G	V	P	K	R	F	S	G	S	R	S	G	S
Gloop2	K	R	L	I	Y	A	A	S	T	L	D	S	G	V	P	K	R	F	S	G	R	R	S	G	S
Gloop4	K	V	L	I	Y	K	V	S	N	R	F	S	G	V	P	D	R	F	S	G	S	G	S	G	T
Gloop5	K	L	L	I	Y	W	A	S	T	R	E	S	G	V	P	D	R	F	T	G	S	G	S	G	T

	70																							
Gloop1	D	Y	S	L	T	I	S	S	L	E	S	E	D	F	A	D	Y	Y	C	L	Q	Y	L	S	Y
Gloop2	D	Y	S	L	T	I	S	S	L	E	S	E	D	F	A	D	Y	Y	C	L	Q	Y	L	S	Y
Gloop4	D	F	T	L	K	I	S	R	V	E	A	E	D	L	G	V	Y	Y	C	F	Q	G	S	H	V
Gloop5	D	F	T	L	T	I	S	S	L	K	A	E	D	L	G	V	Y	Y	C	Q	Q	Y	Y	S	Y

	95															
Gloop1	P	L	T	F	G	A	G	T	K	L	E	L	K	R	A	D
Gloop2	P	L	T	F	G	A	G	T	K	L	E	L	K	R	A	D
Gloop4	P	W	T	F	G	G	G	T	K	L	E	I	K	R	A	D
Gloop5	P	W	T	F	G	G	G	T	K	L	E	I	K	R	A	D

$\underline{V_H}$

	1																			
Gloop1													G	A	S	V	R	L	S	C	K	A	S
Gloop2	Q	V	Q	L	Q	Q	S	G	T	E	L	A	R	P	G	A	S	V	R	L	S	C	K	A	S
Gloop4							S	G	P	E	L	V	E	P	G	A	S	V	R	I	S	C	T	A	S
Gloop5	Q	V	Q	L	Q	Q	P	G	S	V	L	V	G	P	G	D	S	V	K	L	S	C	K	A	S

	26																			
Gloop1	G	Y	S	F	T	T	F	G	I	T	W	V	K	Q	R	P	G	Q	G	L	E	W	I	G	E
Gloop2	G	Y	T	F	T	T	F	G	I	T	W	V	K	Q	R	T	G	Q	G	L	E	W	I	G	E
Gloop4	G	Y	T	F	T	N	Y	Y	I	H	W	L	K	Q	R	P	G	Q	G	L	E	W	I	G	W
Gloop5	G	Y	T	F	T	S	Y	W	M	R	W	M	K	Q	R	P	G	Q	G	L	E	W	I	G	E

	51																								
Gloop1	I	F	P	G	N	N	N	I	Y	Y	N	G	K	F	K	G	K	A	T	L	T	A	D	K	S
Gloop2	I	F	P	G	N	S	K	T	Y	Y	A	E	R	F	K	G	K	A	T	L	T	A	D	K	S
Gloop4	I	Y	P	G	N	G	N	T	K	Y	N	E	N	F	K	G	K	A	T	L	T	A	D	K	S
Gloop5	I	Y	P	N	S	G	R	T	N	Y	N	E	K	F	R	D	K	A	S	L	T	V	D	T	S

	76																								
Gloop1	S	S	S	A	Y	M	G	L	S	S	L	T	S	E	D	S	A	V	Y	F	C	A	R	E	I
Gloop2	S	T	T	A	Y	M	G	L	S	S	L	T	S	E	D	S	A	V	Y	F	C	A	R	E	I
Gloop4	S	S	T	A	F	M	G	I	S	S	L	T	S	E	D	S	A	V	Y	F	C	A	R	Y	T
Gloop5	S	S	T	A	Y	V	D	L	S	S	L	T	S	E	D	S	A	V	Y	Y	C	A	R	G	T

	101														
Gloop1	–	R	Y	W	G	Q	G	T	T	L	T	V	S	S	A	K	T
Gloop2	–	R	Y	W	G	Q	G	T	T	L	T	V	S	S	A	K	T
Gloop4	–	H	Y	W	G	Q	G	T	T	L	T	V	S	S	A	K	T
Gloop5	V	D	Y	W	G	Q	G	T	T	L	T	V	S	S	A	K	T

B. Sequences of the Gloop2 Constant Regions

Gloop2 C_L and C_H1 sequences taken from Roberts (1987). Numbering corresponds to Gloop2 sequential numbering.

C_L

```
                111
Gloop2          A A P T V S I F P P S S E Q L T S G G A S V V C F

                136
Gloop2          L N N F Y P K D I N V K W K I D G S E R Q N G V L

                161
Gloop2          N S W T D Q D S K D S T Y S M S S T L T L T K D E

                186
Gloop2          Y E R H N S Y T C E A T H K T S T S P I V K S F N

                211
Gloop2          R N E C
```

C_H1

```
                117
Gloop2          T P P S V Y P L A P G C G A T T G S S V T L G C L

                142
Gloop2          V K G Y F P E S V T V T W N S G S L S S S V H T F

                167
Gloop2          P A L L Q S G L Y T M S S S V T V P S S T W P S Q

                182
Gloop2          T V T C S V A H P A S S T T V D K K L G P
```

C. Alignment of Gloop2 Sequences with HyHEL-5 and McPC603

Gloop2 sequence as above. HyHEL-5, McPC603 and Rei sequences taken from the Protein Databank Files 2HFL, 1MCP and 1REI respectively.

V_L

```
Gloop2     D I Q M T Q S P S S L S A S L G E R V S L T C R A
Rei        D I Q M T Q S P S S L S A S V G D R V T I T C Q A
HyHEL-5    D I V L T Q S P A I M S A S P G E K V T M T C S A
McPC603    D I V M T Q S P S S L S V S A G E R V T M S C K S

Gloop2     S Q E I S G - - - - - - - Y L S W L Q Q K P D G T I
Rei        S Q D I I K - - - - - - - Y L N W Y Q Q T P G K A P
HyHEL-5    S S S V N - - - - - - - - Y M Y W Y Q Q K S G T S P
McPC603    S Q S L L N S G N Q K N F L A W Y Q Q K P G Q P P

Gloop2     K R L I Y A A S T L D S G V P K R F S G R R S G S
Rei        K L L I Y E A S N L Q A G V P S R F S G S G S G T
HyHEL-5    K R W I Y D T S K L A S G V P V R F S G S G S G T
McPC603    K L L I Y G A S T R E S G V P D R F T G S G S G T

Gloop2     D Y S L T I S S L E S E D F A D Y Y C L Q Y L S Y
Rei        D Y T F T I S S L Q P E D I A T Y Y C Q Q Y Q S L
HyHEL-5    S Y S L T I S S M E T E D A A E Y Y C Q Q W G R N
McPC603    D F T L T I S S V Q A E D L A V Y Y C Q N D H S Y

Gloop2     P L T F G A G T K L E
Rei        P Y T F G Q G T K L Q
HyHEL-5    P - T F G G G T K L E
McPC603    P L T F G A G T K L E
```

V_H

```
Gloop2     Q V Q L Q Q S G T E L A R P G A S V R L S C K A S
HyHEL-5    Q V Q L Q Q S G A E L M K P G A S V K I S C K A S
McPC603    E V K L V E S G G G L V Q P G G S L R L S C A T S
D1.3       Q V Q L K E S G P G L V A P S Q S L S I T C T V S

Gloop2     G Y T F T T F G I T W V K Q R T G Q G L E W I G E
HyHEL-5    G Y T F S D Y W I E W V K Q R P G H G L E W I G E
McPC603    G F T F S D F Y M E W V R Q P P G K R L E W I A A
D1.3       G F S L T G Y G V N W V R Q P P G K G L E W L G -
```

Gloop2	– – I F P G N S K T Y Y A E R F K G K A T L T A D
HyHEL-5	– – I L P G S G S T N Y H E R F K G K A T F T A D
McPC603	S R N K G N K Y T T E Y S A S V K G R F I V S R D
D1.3	– – M I W G D G N T D Y N S A L K S R L S I S K D

Gloop2	K S S T T A Y M G L S S L T S E D S A V Y F C A R E
HyHEL-5	T S S S T A Y M Q L N S L T S E D S G V Y Y C L H G
McPC603	T S Q S I L Y L Q M N A L R A E D T A I Y Y C A R N
D1.3	N S K S Q V F L K M N S L H T D D T A R Y Y C A R E

Gloop2	– – – – – – – – I R Y W G Q G T T L T V
HyHEL-5	N Y D – – – – – F D G W G Q G T T L T V
McPC603	Y Y G S T W Y F D V W G A G T T V T V
D1.3	R D Y R – – – L D Y W G Q G T T L T V

$\underline{C_L}$

Gloop2	A A P T V S I F P P S S E Q L T S G G A S V V C F
HyHEL-5	A A P T V S I F P P S S E Q L T S G G A S V V C F

Gloop2	L N N F Y P K D I N V K W K I D G S E R Q N G V L
HyHEl-5	L N N F Y P K D I N V K W K I D G S E R Q N G V L

Gloop2	N S W T D Q D S K D S T Y S M S S T L T L T K D E
HyHEL-5	N S W T D Q D S K D S T Y S M S S T L T L T K D E

Gloop2	Y E R H N S Y T C E A T H K T S T S P I V K S F N
HyHEL-5	Y E R H N S Y T C E A T H K T S T S P I V K S F N

Gloop2	R N E C
HyHEL-5	R N E C

$\underline{C_H1}$

Gloop2	T P P S V Y P L A P G C G A T T G S S V T L G C L
HyHEL-5	T P P S V Y P L A P G S A A Q T N S M V T L G C L

Gloop2	V K G Y F P E S V T V T W N S G S L S S S V H T F
HyHEL-5	V K G Y F P E P V T V T W N S G S L S S G V H T F

Gloop2	P A L L Q S G L Y T M S S S V T V P S S T W P S Q
HyHEL-5	P A V L Q S D L Y T L S S S V T V P S S P R P S E

Gloop2	T V T C S V A H P A S S T T V D K K L G P
HyHEL-5	T V T C N V A H P A S S T K V D K K I

APPENDIX 2

PROGRAMS USED

A. Programs for Data Collection and Processing

XENGEN	A. J. Howard, (Howard *et al.*, 1987). Supplied as part of the commercial Nicolet-Xentronics detector package. Contains all programs necessary for collection and processing of data. Described in detail in Chapter 4.
VIMVS	G. L. Taylor. Allows inspection of frames (and the sum or difference of frames) on a microVAX 3000 series workstation screen. Pixel counts represented on a grey scale. Ramtek version allows inspection of frames at high magnification.
LATTICE	G. L. Taylor (Taylor, 1989a). Representation of reciprocal lattice weighted by the structure factor amplitude. Viewed in conjunction with FRODO (Jones, 1985).
DIFFER	P. Shaw *et al.* Scaling between datasets by anisotropic temperature factor and calculation of mean fractional difference.

B. Programs for Molecular Replacement Calculations

BRUTE	M. Fujinaga and R. J. Read, (1987). Version supplied by S. Sheriff with modification to output maps in a format recognised by MERLOT. Implemented on the Convex C210 in Oxford with insertion of a machine-specific subroutine call for matrix multiplication. Provision for the MERLOT convention for c-unique space groups added.
MERLOT	P. M. D. Fitzgerald, (1988). Version 2.3 supplied by P. M. D. Fitzgerald. Additional code in subroutine MAPOUT to handle BRUTE maps (modification suggested by Dr. S. Sheriff) and changes to OPEN statements to eliminate file overwriting on VAX's.

C. Programs for Map Calculations

CADLCF	E. J. Dodson, CCP4 program suite. Sort observed data into appropriate order for GENSFC.
GENSFC	E. J. Dodson, CCP4 program suite. Calculation of electron density map from model coordinates and transformation to F_{calc}.
SHELL_SCALE	D. I. Stuart and E. Fry. Scaling of F_{obs} to F_{calc} in shells of $\sin^2\theta/\lambda^2$ and application of Rayment weighting (Rayment, 1983). Calculation of residual.

FFT	L. Ten Eyck, CCP4 program suite. Fast Fourier Transform electron density/Patterson map calculation.
EXTEND	P. R. Evans, CCP4 program suite. Expansion of map from asymmetric unit to desired fraction of cell.
MAPBRICK	P. R. Evans, CCP4 program suite. Contour electron density maps for use with FRODO (Jones, 1985).

D. Programs for Structural Analysis

ASH	D. I. Stuart. Rigid body superimposition (Hendrickson, 1979) of equivalently numbered molecules and analysis of structural homology and B factors.
SHP	D. I. Stuart. Superimposition of structurally related molecules by the method of Rossmann and Argos (1976).

E. Other Programs

FRODO	T. A. Jones. Modified by J. Pflugrath, M. Saper, R. Hubbard and P. R. Evans. Coordinate and electron density map manipulation.

APPENDIX 3

CALCULATION OF MAPS

The electron density maps used in this study were $|2F_{obs} - F_{calc}|\exp(i\alpha_{calc})$ and $|F_{obs} - F_{calc}|\exp(i\alpha_{calc})$ maps. All data with $F > 0$ in the range 8.0–2.8Å ($P1$ cells) or 8.0–3.3Å ($P2_1$ cells) were used in the calculations, and the $F(000)$ term was omitted. The maps were calculated according to the following protocol:

1. CADLCF	Sorting of observed data into an order acceptable for input to GENSFC.
2. GENSFC	Calculation of F_{calc} from the model. F_{calc} added to the native file where there was an equivalent native reflection.
3. SHELL_SCALE	Scaling of F_{calc} to F_{obs} in shells of $\sin^2\theta/\lambda^2$. Rayment weighting (factor of $\exp\left(-\lvert F_{obs}-F_{calc}\rvert/F_{obs}\right)$) applied to the structure factors. Report residual $R = \sum \lvert F_{obs} - F_{calc}\rvert / \sum F_{obs}$.
4. FFT	Calculation of $F_{obs} - F_{calc}$ and $2F_{obs} - F_{calc}$ electron density maps. Unit cell sampling set to at least $\approx 1/3\ d_{min}$.
5. EXTEND	Expansion of map from the asymmetric portion of the unit cell to the portion required to cover the coordinates by application of symmetry operators.
6. MAPBRICK	Contouring of the electron density map and conversion into compressed format readable by FRODO (Jones, 1985).

REFERENCES

Abola, E. E., Ely, K. R. and Edmundson, A. B. (1980) Biochemistry **19**, pp.432–439

Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D. and Brown, F. (1989) Nature **327**, pp.709–716

Al Moudallal, Z., Briand, J. B. and Van Regenmortel, M. H. V. (1985) EMBO J. **4**, pp.1231–1235

Al Moudallal, Z., Briand, J. B. and Van Regenmortel, M. H. V. (1982) EMBO J. **1**, pp.1005–1008

Alzari, P. M., Lascombe, M.-B. and Poljak, R. J. (1988) Ann. Rev. Immunol. **6**, pp.555–580

Alzari, P. M., Mariuzza, R. A., Saludjan, P. and Poljak, R. J. (1987) Acta Cryst. A**43** Supplement, C-19

Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. and Poljak, R. J. (1986) Science **233**, pp.747–753

Amzel, L. M. and Poljak, R. J. (1979) Ann. Rev. Biochem. **48**, pp.961–997

Amzel, L. M., Poljak, R. J., Saul, F., Varga, J. M. and Richards, F. F. (1974) Proc. Natl. Acad. Sci. USA **71**, pp.1427–1430

Argos, P., Ford, G. C. and Rossmann, M. G. (1975) Acta Cryst. A**31**, pp.499–506

Arnold, E. and Rossmann, M. G. (1988) Acta Cryst. A**44**, pp.270–282

Arnold, E. and Rossmann, M. G. (1986) Proc. Natl. Acad. Sci. USA **83**, pp.5489–5493

Arnon, R. and Maron, E. (1971) J. Mol. Biol. **61**, pp.225–235

Arnon, R. and Sela, M. (1969) Proc. Natl. Acad. Sci. USA **62**, pp.163–170

Artymiuk, P. J., Blake, C. C. F., Rice, D. W. and Wilson, K. S. (1982) Acta Cryst. B**38**, p.778

Artymiuk, P. J., Blake, C. C. F., Grace, D. E. P, Oatley, S. J., Phillips, D. C. and Sternberg, M. J. E. (1979) Nature **280**, pp.563–568

Atassi, M. Z. (1984) Eur. J. Biochem. **145**, pp.1–20

Atassi, M. Z. (1978) Immunochem. **15**, pp.909–933

Audibert, F., Jolivetr, M., Chedid, L., Alouf, J. E., Boquet, P., Rivaille, P. and Siffert, O. (1981) Nature **289**, pp.593–594

Avey, H. P., Poljak, R. J., Rossi, G. and Nisonoff, A. (1968) Nature **220**, pp.1248–1249

Baker, A. T., Varghese, J. N., Laver, W. G., Air, G. M. and Colman, P. M. (1987) Proteins **2**, p.111

Baker, A. T. and Hubbard, R. E. (1984) Progr. Biophys. Molec. Biol. **44**, pp.97–179

Barford, D., and Johnson, L. N. (1989) Nature **340**, pp.609–616

Barlow, D. J., Edwards, M. S. and Thornton, J. M. (1986) Nature **322**, pp.747–748

Beale, D. and Feinstein, A. (1976) Quart. Rev. Biophys. **9**, pp.135–180

Becker, J. W. and Reeke G. (1985) Proc. Natl. Acad. Sci. USA **82**, pp.4225–4229

Benjamin, D. C., Berzofsky, J. A., Easy, I. J., Gurd, F. R. N., Hannum, C., Leach, S. J., Margoliash, E., Michael, J. G., Miller, A., Prager, E. M., Reichlin, M., Sercarz, E. E., Smith-Gill, S. J., Todd, P. E. and Wilson, A. C. (1984) Ann. Rev. Immunol. **2**, pp.76–101

Benkovic, S. J., Napper, A. D. and Lerner, R. A. (1988) Proc. Natl. Acad. Sci. USA **85**, pp.5355–5358

Bentley, G. A., Bhat, T. N., Boulot, G., Fischmann, T., Navaza, J., Poljak, R. J., Riottot, M.-M., and Tello, D. (1989) Cold Spring Harbor Symposium on Immune Recognition, In press.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., Tasumi, M., (1977) J. Mol. Biol. **112**, pp.535–542

Berzofsky, R. A. (1985) Science **229**, pp.932–940

Bird, R. E., Hardman, K. D., Jacobson, J. W., Johnson, S., Kaufman, B. M, Lee, S.-M., Lee, T., Pope, S. H., Riordan, G. S. and Whitlow, M. (1988) Science **242**, pp.423–426

Bittle, J. L., Houghten, R. A., Alexander, H., Shinnick, T. M., Sutcliffe, J. G., Lerner, R. A., Rowlands, D. J. and Brown F. (1982) Nature **298**, pp.30–33

Bjorkman, B. J., Saper. M. A., Samraoui, B., Bennet, W. S., Strominger, J. L. and Wiley, D. C. (1987) Nature **329**, pp.506–512

Bjorkman, B. J., Saper. M. A., Samraoui, B., Bennet, W. S., Strominger, J. L. and Wiley, D. C. (1987) Nature **329**, pp.512–518

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C. and Sarma, V. R. (1965) Nature **206**, p.757

Blow, D. M. (1985) "Molecular Replacement" — Proceedings of the Daresbury Study Weekend, 15–16 February 1985 (ed. P. A. Machin), pp.2–7. SERC, Daresbury.

Blum, M., Metcalf, P., Harrison, S. C. and Wiley, D. C. (1987) J. Appl. Cryst. **20**, pp.235–242

Blundell, T. L. and Johnson, L. N. (1976) "Protein Crystallography". Academic Press, London.

Boisset, N., Frank, J., Taveau, J. C., Billiald, P., Motta, G., Sizaret, P. Y. and Lamy, J. (1988) Proteins **3**, pp.161–183

Boquet, P., Alouf, J. E., Duflot, E., Siffert, O. and Rivaille, P. (1982) Mol. Immunol. **19**, pp.1541–1549

Bott, R. R., Navia, M. A. and Smith, J. L. (1982) J. Biol. Chem. **257**, pp.9883–9886

Boulianne, G. L., Hozumi, N. and Shulman, M. J. (1984) Nature **312**, pp.643–646

Bricogne, G. (1974) Acta Cryst. A**30**, pp.395–405

Brown, J. H., Jardetzky, T., Saper, M. A., Samraoui, B., Bjorkman, P. J. and Wiley, D. C. (1988) Nature **332**, pp.845–850

Brown, S. E., Howard, C. R., Zuckerman, A. J. and Steward, M. W. (1984) Lancet ii, p.184

Bruccoleri, R. E., Haber, E. and Novotný, J. (1989) Nature **335**, pp.564–568

Bruggermann, M., Williams, G. T., Bindon, C. I., Clark, M. R., Walker, M. R., Jefferis, R., Waldmann, H. and Neuberger, M. S. (1987) J. Exp. Med. **166**, pp.1351–1361

Brünger, A. T. (1989) Acta Cryst. A**45**, pp.42–50

Brünger, A. T. (1988a) J. Mol. Biol. **203**, pp.803–816

Brünger, A. T. (1988b) X-PLOR Manual for v1.3

Brünger, A. T., Karplus, M. and Petzko, G. (1989) Acta Cryst. A**45**, pp.50–61

Brünger, A. T., Kuriyan, J. and Karplus, M. (1987) Science **235**, pp.458–460

Bryant, S. H., Amzel, L. M., Phizackerley, R. P. and Poljak, R. J. (1985) Acta Cryst. B**41**, pp.362–368

Burnens, A., Demotz, S., Corradin, G., Binz, H. and Bosshard, H. R. (1987) Science **235**, pp.781–783

Burton, D. R. (1985) Mol. Immunol. **22**, pp.161–206

Cann, G., Zaritsky, A. and Koshland, M. E. (1982) Proc. Natl. Acad. Sci. USA **79**, pp.6656–6660

Capon, D. J., Chamow, S. M., Mordenti, J., Marsters, S. A., Gregory, T., Mitsuya, H., Byrn, R. A., Lucas, C., Wurm, F. M., Groopman, J. E., Broder, S. and Smith, D. H. (1989) Nature **337**, pp.525–531

Cease, K. B., Margalit, H., Cornette, J. L., Putney, S. D., Robey, W. G., Ouyang, C., Streicher, H. Z., Fischinger, P. J., Gallo, R. C., DeLisi, C. and Berzofsky, J. A. (1987) Proc. Natl. Acad. Sci. USA **84**, pp.4249–4253

Chang, C.-W., Carperos, W. E., Ainsworth, C. F., Olsen, K. W. and Schiffer, M. (1987) Acta Cryst. A**43** Supplement, C-19

Chang, C.-W., Short, M. T., Westholm, F. A., Stevens, F. J., Wang, B.-C., Furey, W., Solomon, A. and Schiffer, M. (1985) Biochemistry **24** pp.4890–4897

Chaudray, V. K., Queen, C., Junghans, R. P., Waldmann, T. A., FitzGerald, D. J. and Pastan, I. (1989) Nature **339**, pp.394–397

Chothia, C., Boswell, D. R. and Lesk, A. M. (1988) EMBO J. **7**, pp.3745–3755

Chothia, C. and Lesk, A. M. (1987) J. Mol. Biol. **196**, pp.901–917

Chothia, C., Novotný, J., Bruccoleri, R. and Karplus, M. (1985) J. Mol. Biol. **186**, pp.651–663

Chothia, C. and Janin, J. (1982) Biochemistry **21**, pp.3955–3965

Chothia, C. and Janin, J. (1981) Proc. Natl. Acad. Sci. USA **78**, pp.4146–4150

Colman, P. M. (1988) Adv. Immunol. **43**, pp.99–132

Colman, P. M., Tulip, W. R., Varghese, J. N., Tulloch, P. A., Baker, A. T., Air, G. M. and Webster, R. G. (1989) Phil. Trans. R. Soc. Lond. B**323**, pp.511–518

Colman, P. M., Laver, W. G., Varghese, J. N., Baker, A. T., Tulloch, P. A., Air, G. M. and Webster, R. G. (1987) Nature **326**, pp.358–363

Colman, P. M. and Webster, R. G. (1987) in "Biological Organization: Macromolecular Interactions at High Resolution" (ed. R. M. Burnett, H. J. Vogel), pp.193–214. Academic Press, New York.

Colman, P. M., Schramm, H. J. and Guss, J. M. (1977) J. Mol. Biol. **116**, pp.73–79

Colman, P. M., Deisenhofer, J., Huber, R. and Palm, W. (1976) J. Mol. Biol. **100**, pp.257–282

Cooper, H. M., Jemmerson, R., Hunt, D. F., Griffin, P. R., Yates, J. R., Shabanowitz, J., Zhu, N.-Z. and Paterson, Y. (1987) J. Biol. Chem. **262**, pp.11591–11597

Crowther, R. A. and Blow, D. M. (1967) Acta Cryst. **23**, pp.544–548

Crowther, R. A. (1972) in "The Molecular Replacement Method" (ed. M. G. Rossmann), pp.174–178. New York: Gordon and Breach.

Cygler, M. and Anderson, W. F. (1988a) Acta Cryst. A**44**, pp.38–45

Cygler, M. and Anderson, W. F. (1988b) Acta Cryst. A**44**, pp.300–308

Cygler, M., Boodhoo, A., Lee, J. S. and Anderson, W. F. (1987)
J. Biol. Chem. **262**, pp.643–648

Darsley, M. J. and Rees, A. R. (1985a) EMBO J. **4**, pp.383–392

Darsley, M. J. and Rees, A. R. (1985b) EMBO J. **4**, pp.393–398

Davies, D. R., Sheriff, S. and Padlan, E. A. (1988) J. Biol. Chem. **263**,
pp.10541–10544

Davies, D. R. and Metzger, H. (1983) Ann. Rev. Immunol. 1, pp.87–117

Davies, D. R., Padlan, E. A. and Segal, D. M. (1975) Ann. Rev. Biochem. **44**,
pp.961–997

de la Paz, P., Sutton, B. J., Darsley, M. J. and Rees, A. R. (1986) EMBO J. **5**,
pp.415–425

Deisenhofer, J. (1981) Biochemistry **20**, pp.2362–2370

Derewenda, Z. S. (1989) Acta Cryst. A**45**, pp.227–234

Derewenda, Z. S. and Helliwell, J. R. (1989) J. Appl. Cryst. **22**, pp.123–137

Derewenda, Z. S., Dodson, E. J., Dodson, G. G. and Brzozowski, A.M. (1981)
Acta Cryst. A**37**, pp.407–413

Diamond, R. (1971) Acta Cryst. A**27**, pp.436–452

Dijkstra, B. W., Renetseder, R., Kalk, K. H., Hol, W. G. J., and Drenth, J. (1982)
J. Mol. Biol. **168**, p.163

Dodson, E. J. (1985) "Molecular Replacement" — Proceedings of the Daresbury
Study Weekend, 15–16 February 1985 (ed. P. A. Machin), pp.33–45.
SERC, Daresbury.

Driessen, H. P. C. and White, H. (1985) "Molecular Replacement" — Proceedings of
the Daresbury Study Weekend, 15–16 February 1985 (ed. P. A. Machin)
pp.27–32. SERC, Daresbury.

Duncan, A. R. and Winter, G. (1988) Nature **332**, pp.738–740

Duncan, A. R., Woof, J. M., Partridge, L. P., Burton, D. R. and Winter, G. (1988) Nature **332**, pp.563–564

Durbin, R. M., Burns, R., Moulai, J., Metcalf, P., Freymann, D., Blum, M., Anderson, J. E., Harrison, S. C. and Wiley, D. C. (1986) Science **232**, pp.1127–1132

Dwek, R. A., Wain-Hobson, S., Dower, S., Gettins, P., Sutton, B., Perkins, S. J. and Givol, D. (1977) Nature **266**, pp.31–37

Edelman, G. M. (1970) Biochemistry **9**, p.3197

Edelman, G. M. and Gall, W. E. (1969) Ann. Rev. Biochem. **38**, p.415

Ely, K. R., Colman, P. M., Abola, E. E., Hess, A. C., Peabody, D. S., Parr, D. M., Connell, G. E., Laschinger, C. A. and Edmundson, A. B. (1978) Biochemistry **17**, pp.820–823

Edmundson, A. B., Ely, K. R., Herron, J. R. and Cheson, B. D. (1987) Mol. Immunol. **24**, pp.915–935

Edmundson, A. B. and Ely, K. R. (1985) Mol. Immunol. **22**, pp.463–475

Edmundson, A. B., Ely, K. R., Herron, J. N. and Gibson, A. L. (1984) "Investigation and Exploitation of Antibody Combining Sites", (Eds E. Reid, G. M. W. Cook, D. J. Morre), pp.33–50. Plenum Press, New York.

Edmundson, A. B., Ely, K. R., Abola, E. E., Schiffer, M. and Panagiotopoulos, N. (1975) Biochemistry **14**, pp.3953–3961

Emini, E., Jameson, B. A. and Wimmer, E. (1983) Nature **304**, p.699

Epp, O., Lattman, E. E., Schiffer, M., Huber, R. and Palm, W. (1975) Biochemistry **14**, pp.4943–4952

Evans, P. R., Farrants, G. W., Lawrence, M. C. and Shirakihara, Y. (1985) "Molecular Replacement" — Proceedings of the Daresbury Study Weekend, 15–16 February 1985 (ed. P. A. Machin) pp.53–55. SERC, Daresbury.

Evin, G., Galen, F.-X., Carlson, W. D., Handschumacher, M., Novotný, J., Bouhnik, J., Menard, J., Covol, P. and Haber, E. (1988) Biochemistry **27**, pp.156–164

Ewald, P. P. (1921) Z. Kristallagr. Miner. **56**, p.129

Feinstein, A. and Beale, D. (1977) in Structure and Function of Antibodies (eds. L. E. Glynn and M. W. Steward), pp.263–306. Wiley, New York.

Feldmann, R. J., Potter, M. and Glaudemans, C. P. J. (1981) Mol. Immunol. **18**, pp.683–698

Fieser, T. M., Tainer, J. A., Geysen, H. M., Houghten, R. A. and Lerner, R. A. (1987) Proc. Natl. Acad. Sci. USA **84**, pp.8568–8572

Fine, R. M., Wang, H., Shenka, P. S., Yarmush, D. L. and Levithal, C. (1986) Proteins **1**, pp.342–362

Firca, J. R., Ely, K .R., Kremser, P., Westholm, F. A., Dorrington, K. J. and Edmundson, A. B. (1978) Biochemistry **17**, pp.148–158

Fitzgerald, P. M. D. (1988a) J. Appl. Cryst. **21**, pp.273–278

Fitzgerald, P. M. D. (1988b) Documentation for MERLOT v1.2

Fitzgerald, P. M. D. and Love, W. E. (1979) J. Mol. Biol. **132**, pp.603–619

Fujinaga, M., Gros, P. and van Gunsteren, W. F. (1989) J. Appl. Cryst. **22**, pp.1–8

Fujinaga, M. and Read, R. (1987) J. Appl. Cryst. **20**, pp.517–521

Furey, W., Wang, B. C., Yoo, C. S. and Sax, M. (1979) Acta Cryst. A**35**, pp.810–817

Gaykema, W. P. J., Hol, W. G. J., Vereijken, J. M., Soeter, N. M., Bak, H. J. and Beintema, J. J. (1984) Nature **309**, pp.23–29

Getzoff, E. D., Tainer, J. A., Lerner, R. A. and Geysen, H. M. (1988) Adv. Immunol **43**, pp.1–98

Getzoff, E. D., Geysen, H. M., Rodda, S. J., Alexander, H., Tainer, J. A. and Lerner, R. A. (1987) Science **235**, pp.1191–1196

Geysen, H. M., Rodda, S. J., Mason, T. J., Tainer, R. A., Alexander, H., Getzoff, E. D. and Lerner, R. A. (1987a) Science **238**, pp.1584–1586

Geysen, H. M., Tainer, J. A., Rodda, S. J., Mason, T. J., Alexander, H., Getzoff, E. D. and Lerner, R. A. (1987b) Science **235**, pp.1184–1190

Giusti, A. M., Chien, N. C., Zack, D. J., Shin, S.-U. and Scharff, M. D. (1987) Proc. Natl. Acad. Sci. USA **84**, pp.2926–2930

Gras-Masse, H. S., Jolivert, M. E., Audibert, F. M., Beachey, E. H., Chedid, L. A. and Tartar, A. L. (1986) Mol. Immunol. **23**, pp.1391–1395

Green, N., Alexander, H., Olson, A., Alexander, S., Shinnick, T. M., Sutcliffe, J. G. and Lerner, R. A. (1982) Cell **28**, pp.477–487

Gregory, L., Davis, K. G., Sheth, B., Boyd, J., Jefferis, R., Nave, C. and Burton, D. R. (1987) Mol. Immunol. **24**, pp.821–829

Hamlin, R. (1985) Methods In Enzymology **114** (Eds. H. W. Wycoff, C. H. W. Hirs, S. N. Timasheff), pp.416–452

Hanson, D. C., Yguerabide, J. and Schumaker, V. N. (1985) Mol. Immunol. **22**, pp.237–244

Harada, Y., Lifchitz, A., Berthou, J. and Jolles, P. (1981) Acta Cryst. A**37**, pp.398–406

Harrison, E. T. and Mage, M. G. (1967) Biochem. Biophys. Acta **147**, pp.52–59

Hendrickson, W. A. (1985) Methods In Enzymology **115** (Eds. H. W. Wycoff, C. H. W. Hirs, S. N. Timasheff), pp.252–270

Hendrickson, W. A. and Konnert, J. H. (1980) Computing in Crystallography (Eds. R. Diamond, S. Ramaseshan, K. Venkatesan), pp.13.01–13.26. Indian Academy of Sciences.

Hendrickson, W. A. and Ward, K. B. (1976) Acta Cryst. A**32**, pp.778–780

Herrington, D. A., Clyde, D. F., Losonsky, G., Cortesia, M., Murphy, J. R., Davis, J., Baqar, S., Felix, A. M., Heimer, E. P., Gillessen, D., Nardin, E., Nussenzweig, R. S., Nussenzweig, V., Hollingdale, M. R. and Levine, M. M. (1987) Nature **328**, pp.257–259

Herron, J. N., He, X.-M., Gibson, A. L., Voss, E. W. and Edmundson, A. B. (1987) Fed. Proc. **46**, p.2204

Herron, J. N., He, X.-M., Mason, M. L., Voss, E. W. and Edmundson, A. B. (1989) Proteins **5**, pp.271–280

Hilyard, K. L., Roberts, R., Cheetham, J. C. and A. R. Rees (1989) J. Cell. Biochem. Supplement 13A, p.86

Honjo, T. (1983) Ann. Rev. Immunol. **1**, pp.499–528

Hopp, T. P. and Woods, K. R. (1981) Proc. Natl. Acad. Sci. USA **78**, pp.3824–3828

Howard, A. J. (1988) User Guide to XENGEN v1.3

Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulos, T. L., Ohlendorf, D. H. and Salemme, F. R. (1987) J. Appl. Cryst. **20**, pp.383–387

Huber, R. (1985) "Molecular Replacement" — Proceedings of the Daresbury Study Weekend, 15–16 February 1985 (ed. P. A. Machin), pp.58–61. SERC, Daresbury.

Huber, R. (1965) Acta Cryst. **19**, pp.353–356

Hunkapiller, T. and Hood, L. (1988) Adv. Immunol. **44**, pp.1–64

Huston, J. S., Levinson, D., Mudgett-Hunter, M., Tai, M.-S., Novotný, J., Margolies, M. N., Ridge, R. J., Bruccoleri, R. E., Haber, E., Crea, R. and Oppermann, H. (1988) Proc. Natl. Acad. Sci. USA **85**, p.5879

Iverson, B. L. and Lerner, R. A. (1989) Science **243**, pp.1184–1188

Jack, A. and Levitt, M. (1978) Acta Cryst. A**34**, pp.931–935

James, M. N. G., and Sielecki, A. R. (1986) Nature **319**, pp.33–38

Jemmerson, R. and Paterson, Y. (1986) Science **232**, pp.1001–1004

Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. and Winter, G. (1986) Nature **321**, pp.522–525

Jones, T. A. (1985) Methods In Enzymology **115** (Eds. H. W. Wycoff, C. H. W. Hirs, S. N. Timasheff), pp.157–171

Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. and Perry, H. (1987) Sequences of Proteins of Immunological Interest vol. 1, (U.S. Dept. of Health and Public Services, Washington, D.C.

Kabat, E. A., Wu, T. T. and Bilofsky, H. (1977) J. Biol. Chem. **252**, pp.6609–6616

Kabat, E. A. and Wu, T. T. (1971) Ann. N. Y. Acad. Sci. **190**, p.382

Knossow, M., Daniels, R. S., Douglas, A. R., Skehel, J. J. and Wiley, D. C. (1984) Nature **311**, pp.678–680

Kohen, F., Kim, J. B., Lindner, H. R., Eshar, Z. and Green, B. (1980) FEBS Lett. **111**, pp.427–431

Köhler, G. and Milstein, C. (1975) Nature **256**, pp.495–497

Konnert, J. H. (1976) Acta Cryst. A**32**, pp.614–617

Krystek, S. R., Dias, J. A., Riechert, L. E. and Andersen, T. T. (1985) Endocrinology **117**, pp.1125–1131

Kuriyan, J., Brünger, A. T., Karplus, M. and Hendrickson, W. A. (1989) Acta Cryst. A**45**, pp.396–409

Langs, D. A. (1985) Acta Cryst. A**41**, pp.305–308

Lascombe, M.-B., Alzari, P. M., Boulot, G., Saludjian, P., Tougard, P., Berek, C., Haba, S., Rosen, E. M., Nisonoff, A. and Poljak, R. J. (1989) Proc. Natl. Acad. Sci. USA **86**, pp.607–611

Lattman, E. E. (1985) Methods In Enzymology **115** (Eds. H. W. Wycoff, C. H. W. Hirs, S. N. Timasheff), pp.55–77

Lattman, E. E. and Love, W. E. (1970) Acta Cryst. B**26**, pp.1854–1857

Lerner, R.A. (1984) Adv. Immunol. **36**, pp.1–44

Lerner, R.A. and Benkovic, S. J. (1988) BioEssays **9**, pp.107–112

Lerner, R.A. and Tramontano, A. (1988) Scientific American, March 1988, pp.42–50

Lerner, R.A. and Tramontano, A. (1987) Trend Biochem. Sci. **12**, pp.427–430

Lesk, A. M. and Chothia C. (1988) Nature **335**, pp.188–190

Lesk, A. M. and Chothia C. (1982) J. Mol. Biol. **160**, pp.325–342

Lewis, M. and Rees, D. C. (1985) Science **229**, pp.1163–1165

LoBuglio, A. F., Wheeler, R. H., Trang, J., Haynes, A., Rogers, K., Harvey, E. B., Sun, L., Ghrayeb, J. and Khazaeli, M. B. (1989) Proc. Natl. Acad. Sci. USA **86**, pp.4220–4224

Luo, M., Vriend, G., Kamer, G., Minor, E., Arnold, E., Rossmann, M. G., Boege, U., Scraba, D. G., Duke, D. M. and Palmenberg, A. C. (1987) Science **235**, p.182–191

Luzzati, V. (1952) Acta Cryst. **5**, pp.802–810

Mainhart, C. R., Potter, M. and Feldmann, R. J. (1984) Mol. Immunol. **21**, pp.469–478

Mariuzza, R. A., Phillips, S. E. V. and Poljak, R. J. (1987) Ann. Rev. Biophys. Biophys. Chem. **16**, pp.139–159

Mariuzza, R. A., Poljak, R. J., Mihaesco, C. and Mihaesco, E. (1983) J. Mol. Biol. **165**, pp.559–561

Maron, E., Shiozawa, C., Arnon, R. and Sela, M. (1971) Biochemistry **10**, pp.763–771

Marquart, M., Deisenhofer, J., Huber, R. and Palm, W. (1980) J. Mol. Biol. **141**, pp.369–391

Martin, A. C. R., Cheetham, J. C. and Rees, A. R. (1989) Proc. Natl. Acad. Sci. USA , in press.

Matsushima, M., Marquart, M., Jones, T. A., Colman, P. M., Bartels, K., Huber, R. and Palm, W. (1978) J. Mol. Biol. **121**, pp.441–459

Matthews, B. W. (1968) J. Mol. Biol. **33**, pp.497–497

Messerschmidt, A., Rossi, A., Ladenstein, R., Huber, R., Bolognesi, M., Gatti, G., Marchesini, A., Pertruzzelli, R. and Finazzi-Agro, A. (1989) J. Mol. Biol. **206**, pp.513–529

Morrison, S. L. and Oi, V. T. (1988) Adv. Immunol. **44**, pp.65–92

Morrison, S. L., Johnson, M. J., Herzenberg, L. A. and Oi, V. T. (1984) Proc. Natl. Acad. Sci. USA **81**, pp.6851–6855

Moult, J., Yonath, A., Traub, W., Smilansky, A., Podjarny, A., Rabinovich, D. and Saya, A. (1976) J. Mol. Biol. **100**, p.179

Muller, S., Plane, S., Couppez, M. and Van Regenmortel, M. H. V. (1986) Mol. Immunol. **23**, pp.593–601

Navia, M. A., Segal, D. M., Padlan, E. A., Davies, D. R., Rao, N., Rudikoff, S. and Potter, M. (1979) Proc. Natl. Acad. Sci. USA **76**, pp.4071–4074

Neuberger, M. S., Williams, G. T., Mitchell, E. B., Jouhal, S. S., Flanagan, J. G. and Rabbitts, T. H. (1985) Nature **314**, pp.268–270

Neuberger, M. S., Williams, G. T., and Fox, R. O. (1984) Nature **312**, pp.604–608

Nixon, P. E. and North, A. C. T. (1976) Acta Cryst. A**32**, pp.320–325

Novotný, J., Bruccoleri, R. E. and Saul F. A. (1989) Biochemistry **28**, pp.4735–4749

Novotný, J., Handschumacher, M., and Bruccoleri, R. E. (1987a) Immunol. Today **8**, pp.26–31

Novotný, J., Bruccoleri, R. E., Carlson, W. D., Handschumacher, M. and Haber, E. (1987b) Science **238**, p.1584

Novotný, J. and Haber, E. (1986) Biochemistry **25** pp.6748–6754

Novotný, J., Handschumacher, M. and Haber, E. (1986a) J. Mol. Biol. **189**, pp.715–721

Novotný, J., Handschumacher, M., Haber, E., Bruccoleri, R. E., Carlson, W. B., Fanning, D. W., Smith, J. A. and Rose, G. D. (1986b) Proc. Natl. Acad. Sci. USA **83**, pp.226–230

Novotný, J., Tonegawa, S., Saito, H., Kranz, D. M. and Eisen, H. N. (1986c) Proc. Natl. Acad. Sci. USA **83**, pp.742–746

Novotný, J. and Haber, E. (1985) Proc. Natl. Acad. Sci. USA **82**, pp.4592–4596

Novotný, J., Bruccoleri, R., Newell, J., Murphy, D., Haber, E. and Karplus, M. (1983) J. Biol. Chem. **258**, pp.14433–14437

Padlan, E. A. and Kabat, E. A. (1988) Proc. Natl. Acad. Sci. USA **85**, pp.6885–6889

Padlan, E. A., Cohen, G. H. and Davies, D. R. (1987) in "Biological Organization:" Macromolecular Interactions at High Resolution" (ed. R. M. Burnett, H. J. Vogel), pp.193–214. New York, Academic Press.

Padlan, E. A., Davies, D. R., Pecht, I., Givol, D. and Wright, C. (1976) Cold Spring Harbor Symp. Quant. Biol. **41**, pp.627–637

Padlan, E. A., Segal, D. M., Spande, T. F., Davies, D. R., Rudikoff, S. and Potter, M. (1973) Nature **245**, pp.165–167

Parker, J. M. R. , Guo, D. and Hodges, R. S. (1986) Biochemistry **25**, pp.5425–5432

Paul, S., Volle, D. J., Beach, C. M., Johnson, D. R., Powell, M. J. and Massey, R. J. (1989) Science **244**, pp.1158–1162

Patarroyo, M. E., Amador, R., Clavijo, P., Moreno, A., Guzman, F., Romero, P., Tascon, R., Franco, A., Murillo, L. A., Ponton, G. and Trujillo, G. (1988) Nature **332**, pp.158–161

Patarroyo, M. E., Romero, P., Torres, M. L., Clavijo, P., Moreno, A., Martinez, A., Rodriguez, R. Guzman, F. and Cabezas, E. (1987) Nature **328**, pp.629–632

Patterson, A. L. (1934) Phys. Rev. **46**, pp.372–376

Pawlita, M., Mushinski, E., Feldmann, R. J. and Potter, M. (1981) J. Exp. Med. **154**, pp.1946–1956

Phillips, S. E. V. (1980) J. Mol. Biol. **142**, pp.531–554

Pollack, S. J., Nakayama, G. R. and Schultz, P. G. (1988) Science **242**, pp.1038–1040

Pollack, S. J., Jacobs, J. W. and Schultz, P. G. (1986) Science **234**, p.1570

Porter, R. R. (1959) Biochem. J. **73**, pp.119–127

Prasad, L., Vandonselaar, M., Lee, J. S. and Delbaere, L. T. J. (1988) J. Biol. Chem. **263**, pp.2571–2574

Pumphrey, R. S. H. (1986a) Immunol. Today **7**, pp.206–210

Pumphrey, R. S. H. (1986b) Immunol. Today **7**, pp.174–178

Rajan, S. S., Ely, K. R., Abola, E. E., Wood, M. K., Colman, P. M., Athay, R. J. and Edmundson, A. B. (1983) Mol. Immunol. **20**, pp.797–799

Rao, S. T., Hogle, J., Sundralingham, M. (1983) Acta Cryst. C**39**, p.237

Rayment, I. (1983) Acta Cryst. A**39**, pp.102–116

Rees, A. R. and de la Paz, P. (1986) Trends Biochem. Sci. **11**, pp.144–148

Renneboog-Squilbin, C. (1972) J. Mol. Biol. **64**, pp.221–236

Reynolds, R. A., Remington, S. J., Weaver, L. H., Fisher, R. G., Anderson, W. F., Ammon, H. L., and Matthews, B. W. (1985) Acta Cryst. B**41**, pp.139–147

Richardson, J. S. (1981) Adv. Protein Chem. **34**, pp.168–339

Riechmann, L., Clark, M., Waldmann, H. and Winter G. (1988) Nature **332**, pp.323–327

Roberts, S. (1987) D.Phil. Thesis, University of Oxford.

Roberts, S., Cheetham, J. C. and Rees, A. R. (1987) Nature **328**, pp.731–734

Roberts, S. and Rees, A. R. (1986) Protein Engineering **1**, pp.59–65

Rossi, G., Choi, T. K. and Nisonoff, A. (1969) Nature **223**, pp.837–838

Rossmann, M. G. and Argos, P. (1976) J. Mol. Biol. **105**, pp.75–95

Rossmann, M. G. and Blow, D. M. (1963) Acta Cryst. **16**, pp.39–45

Rossmann, M. G. and Blow, D. M. (1962) Acta Cryst. **15**, pp.24–31

Roux, K. H., Monafo, W. J., Davie, J. M. and Greenspan, N. S. (1987) Proc. Natl. Acad. Sci. USA **84**, pp.4984–4988

Ru-Chang, Bi, Cutfield, S.M., Dodson, E. J., Dodson, G. G., Giordano, F., Reynolds, C. D. and Tolley, S. P. (1983) Acta Cryst. B**39**, pp.90–98

Rudikoff, S. and Pumphrey, J. G. (1986) Proc. Natl. Acad. Sci. USA **83**, pp.7875–7878

Rudikoff, S., Potter, M., Segal, D. M., Padlan, E. A., and Davies, D. R. (1974) Proc. Natl. Acad. Sci. USA **71**, pp.4298–4302

Sarma, V. R., Silverton, E. W., Davies, D. R. and Terry, W. D. (1971) J. Biol. Chem. **246**, pp.3753–3759

Satow, Y., Cohen, G. H., Padlan, E. A. and Davies, D. R. (1986) J. Mol. Biol. **190**, pp.593–604

Saul, F. A., Amzel, L. M. and Poljak, R. J. (1978) J. Biol. Chem. **253**, pp.585–597

Sawutz, D. G., Kuory, R. and Homcy, C. J. (1987) Biochemistry **26**, pp.5275–5282

Schierbeek, A. J., Renestseder, R., Dijkstra, B. W. and Hol, W. G. J., (1985) "Molecular Replacement" — Proceedings of the Daresbury Study Weekend, 15–16 February 1985 (ed. P. A. Machin), pp.16–21. SERC, Daresbury.

Schiffer, M., Ainsworth, C., Xu, Z.-B., Carperos, W., Olsen, K., Solomon, A., Stevens, F. J. and Chang, C.-H. (1989) Biochemistry **28**, pp.4066–4072

Schiffer, M., Stevens, F. J., Westholm, F. A., Kim, S. S. and Carlson, R. D. (1982) Biochemistry **21**, pp.2874–2878

Schultz, P.G. (1988) Science **240**, pp.426–433

Segal, D. M., Padlan, E. A, Cohen, G. H., Rudikoff, S., Potter, M. and Davies, D. R. (1974) Proc. Natl. Acad. Sci. USA **71**, pp.4298–4302

Sheriff, S., Silverton, E. W., Padlan, E. A., Cohen, G. H., Smith-Gill, S. J., Finzel, B. C. and Davies, D.R. (1987) Proc. Natl. Acad. Sci. USA **84**, pp.8075–8079

Sheriff, S., Hendrickson, W. A., Stenkamp, R. E., Sieker, L. C., and Jensen, L. H. (1985) Proc. Natl. Acad. Sci. USA **82**, pp.1104–1107

Shih, H. H.-L., brady, J. and karplus, M. (1985) Proc. Natl. Acad. Sci. USA **82**, pp.1697–1700

Shokat, K. M., Leumann, C. J., Sugasawara, R. and Schultz, P. G. (1989) Nature **338**, pp.269–271

Smith-Gill, S. J., Mainhart, C. R., Lavoie, T. B., Feldmann, R. J., Drohan, W. and Brooks, B. R. (1987) J. Mol. Biol. **194**, pp.713–724

Smith-Gill, S. J., Wilson, A. C., Potter, M., Prager, E. M., Feldmann, R. J. and Mainhart, C. R. (1982) J. Immunol. **128**, pp.314–322

Snow, M. E. and Amzel, L. M. (1986) Proteins **1**, pp.267–279

Steward, M. W. and Howard, C. R. (1987) Immunol. Today **8**, pp.51–58

Stevens, F. J., Chang, C.-H., Schiffer, M. (1988) Proc. Natl. Acad. Sci. USA **85**, pp.6895–6899

Stevens, V. (1986) in "Synthetic Peptides and Antigens" (Ciba Symposium **119**) (eds. R. Porter and J. Whelan), p.200. Wiley, Chichester.

Stuart, D. (1979) Ph.D. Thesis, University of Bristol

Suh, S. W., Bhat, T. N., Navia, M. A., Cohen, G. H., Rao, D. N., Rudikoff, S. and Davies, D. R. (1986) Proteins **1**, pp.74–80

Sun, L. K., Curtis, P., Rakowicz-Szulczynska, E., Ghrayeb, J., Chang, N., Morrison, S. L. and Koprowski, H. (1987) Proc. Natl. Acad. Sci. USA **84**, pp.214–218

Sussman, J. L., Holbrook, S. R., Church, G. M. and Kim, S-H. (1977) Acta Cryst. A**33**, pp.800-804

Sutton, B. J. and Phillips, D. C. (1983) Biochem. Soc. Trans. **11**, pp.130–132

Tainer, J. A., Getzoff, E. D., Alexander, H., Houghten, R. A., Olson, A. J., Lerner, R. A. and Hendrickson, W. A. (1984) Nature **312**, pp.127–134

Tanaka, N. (1977) Acta Cryst. A**33**, pp.191–193

Taylor, G. L. (1989a) Newsletter on Protein Crystallography **23**, pp.13–14

Taylor, G. L. (1989b) "Molecular Simulation and Protein Crystallography" — Proceedings of the Daresbury Study Weekend, 27–28 January 1989. In press.

Thornton, J. M., Edwards, M. S., Taylor, W. R. and Barlow, D. J. (1986) EMBO J. **5**, pp.409–413

Tickle, I. J. (1985) "Molecular Replacement" — Proceedings of the Daresbury Study Weekend, 15–16 February 1985 (ed. P. A. Machin), pp.16–21. SERC, Daresbury.

Tollin, P. (1966) Acta Cryst. **21**, pp.613–614

Tollin, P. and Rossmann, M. G. (1966) Acta Cryst. **21**, pp.872–876

Tonegawa, S. (1983) Nature **302**, pp.575–581

Tramontano, A., Ammann, A. A. and Lerner, R. A. (1988) J. Am. Chem. Soc. **110**, pp.2282–2286

Tramontano, A. and Lerner, R. A. (1986) Science **234**, pp.1566–1570

Tulip, W. R., Varghese, J. N., Webster, R. G., Air, G. M., Laver, W. G. and Colman, P. M. (1989) Cold Spring Harbor Symposium on Immune Recognition. In press.

Tulloch, P. A., Laver, W. G., Davis, P. C., Webster, R. G., Varghese, J. N. and Colman, P. M. (1985) "Immune Recognition of Protein Antigens" — Current Communications in Molecular Biology, pp.150–155. Cold Spring Harbor Laboratory.

Valentine, R. C. and Green, N. M. (1967) J. Mol. Biol. **248**, p.2818

Van Regenmortel (1989) Phil. Trans. R. Soc. Lond. B**323**, pp.417–466

Varghese, J. N., Webster, R. G., Laver, W. G. and Colman, P. M. (1988) J. Mol. Biol. **200**, pp.201–203

Vitetta, E. S. and Uhr, J. W. (1985) Ann. Rev. Immunol. **3**, pp.197–212

Wang, B.-C., Yoo, C. S. and Sax, M. (1979) J. Mol. Biol. **129**, pp.657–674

Wang, H., Wu, J. and Tang, P. (1989) Nature **337**, p.514

Waser, J. (1963) Acta Cryst. **16**, pp.1091–1094

Watenpaugh, K. D., Siecker, L. C., Herriott, J. R. and Jensen, L. H. (1973) Acta Cryst. B**29**, pp.943–956

Welling, G. W., Weijer, W. J., van der Zee, R. and Welling-Webster, S. (1988) FEBS Lett. **188**, pp.215–218

Westhof, E., Altschuh, D., Moras, D., Bloomer, A. C., Mondragon, A., Klug, A. and Van Regenmortel, M. H. V. (1984) Nature **311**, pp.123–126

Williams, A. F. and Barclay, A. N. (1988) Ann. Rev. Immunol. **6**, pp.381–405

Wilson, I. A., Niman, H. L., Houghten, R. A., Cherenson, A. R., Connolly, M. L. and Lerner, R. A. (1984) Cell **37**, pp.767–778

Wu, T. T. and Kabat, E. A. (1970) J. Exp. Med. **132**, p.211

Yeates, T. O. (1989) Acta Cryst. A**45**, pp.309–314

Zikan, J., Novotný, J., Trapare, T. L., Koshland, M. E., Urry, D. W., Bennet, J. C. and Mestecky, J. (1985) Proc. Natl. Acad. Sci. USA **82**, pp.5905–5909