Structural Studies on Foot–and–Mouth Disease Virus

Susan Mary Lea

Laboratory of Molecular Biophysics

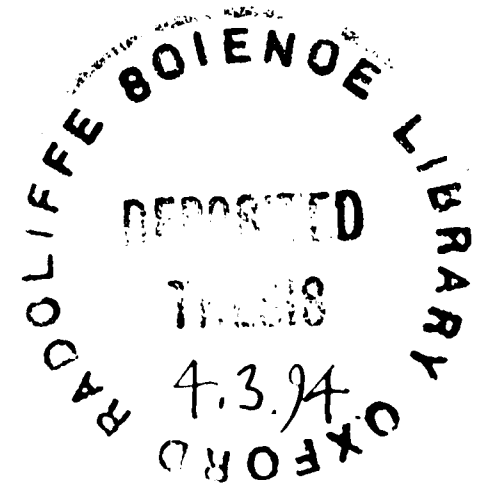
and New College, Oxford

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

at the University of Oxford

Michaelmas Term 1993



Structural Studies on Foot–and–Mouth Disease Virus

Susan Mary Lea

Laboratory of Molecular Biophysics and New College, Oxford

Submitted for the degree of Doctor of Philosophy

Michaelmas Term 1993

Abstract

Foot–and–mouth disease viruses (FMDVs) constitute the aphthovirus genus of the Picornaviridae. The structures of O_1 subtype viruses O_1K and G67 have been solved and comparisons reveal the structural basis of monoclonal antibody escape mutations in G67. Escape mutations are seen to occur at surface–exposed residues and to provoke structural changes limited to the altered side chains. Comparisons of the structures of O_1K and O_1BFS (Acharya *et al.*, Nature **337**, 709–716 (1989)) suggest that changes occurring 'in–the–field' in response to polyclonal antibody pressure may be subtly different from mutations produced by monoclonal antibody pressure *in vitro*. Field mutations are seen to alter less exposed residues and to have more far–reaching structural effects than the *in vitro*, monoclonal provoked mutations.

Crystals of G67 are seen to be 'intimately twinned', the data possessing extra symmetry due to a mis–packing of the crystals. A protocol, based on current real–space averaging procedures with a novel constraint imposed, has been used successfully to deconvolute these data. This method might be more generally applied to deconvolute the wavelength overlaps that occur when using the Laue method.

The structures of C-S8c1 and mutant SD6-6 have been solved at a resolution of 3.5Å. These structures enable comparisons between members of different FMDV serotypes to be made for the first time, namely: serotype O (O_1BFS) and serotype C (C-S8c1). Flexibility of the Arg–Gly–Asp containing G–H loop of VP1 is seen to be amongst the most conserved structural features. This loop is implicated in receptor binding and possible rôles for the observed flexibility are discussed. The C-S8c1 structure also reveals more detail in previously disordered regions of the capsid, namely: the N–terminal residues of VP2 and potential myristate density under the 5–fold axis of the virion.

Analysis of structures from the Protein Data Bank reveals different patterns of amino acid use in proteins involved in the two halves of the immune recognition event i.e. immunoglobulins and viruses. These patterns seem to be based not only on the characteristics of the most used amino acids but also on characteristics of the nucleotide codons used to code for them.

Acknowledgements

Thanks to Dave Stuart for taking on a half–made medic and turning me into a part–made crystallographer, letting me contribute to the 'FMDV story' in the process and to the Oxford members of the FMDV team, Liz Fry and Derek Logan for making me welcome.

Thanks to the virologists and 'crystal–providers', the members of the Picornavirus Group at the IAH Pirbright, Andrew King, Robin Abu–Ghazaleh, Wendy Blakemore, Stephen Curry, Terry Jackson and John Newman. Special thanks to Dave Goodrich, in his rôle as Disease Security Officer, for enduring and enlivening long nights at Daresbury with tales of FMDVs past and present.

To the members of T4, Ren, Richard, Robert, Jon, Liz (Duke) and Dale, as well as the semi–T4s in T2 (Mike, Ajit and John), thanks for the tea (and command files). Thanks also to my parents (and in–laws) for proof–reading at such speed.

Finally I would like to acknowledge the AFRC, IAH Pirbright and Rory for financial support.

For Rory.

Contents

List of Figures

List of Tables

Chapter 1

Introduction to FMDV

1.1 Picornavirus Structure

The picornavirus family consists of five genera of viruses responsible for a wide range of human and animal diseases. They are small, spherical viruses approximately 300Å in diameter and contain a single positive sense RNA genome some 7500–8500 bases in length, enclosed by a protein shell. The RNA is susceptible to proteolytic degradation by ribonucleases, one of the rôles of the protein coat is therefore to provide a physical barrier between the RNA and the external world. The protein coat also mediates interactions between the virus and the cell via receptor binding, assisting entry of the RNA into the cell. The protein coat must, necessarily, be coded for by the RNA it contains. This imposes strict constraints on the amount of protein coded for. Picornaviruses satisfy these constraints by using 60 copies of a protomer made up of only 4 proteins (VP1–4) to form a protein coat with T=1 symmetry (Caspar and Klug, 1962). The assembly of protomers into an icosahedral capsid occurs via a pentameric intermediate which is stable in solution and sediments at 12S (Svedberg units). Structural similarities exist between these animal viruses and the T=3 viruses such as tomato bushy stunt virus (Harrison *et al.*, 1978) with VP1–3 occupying positions analogous to the sequence identical A, B and C subunits. VP1–3 are exposed on the capsid surface and are approximately 30KD, whilst VP4 is smaller (7-8KD) and is completely internal.

X–ray crystallography has revealed a common fold for proteins VP1–3 amongst all the picornaviruses (and many other small RNA viruses) – an 8 stranded β–barrel. Figure 1.1 shows a schematic representation of a canonical β–barrel consisting of two

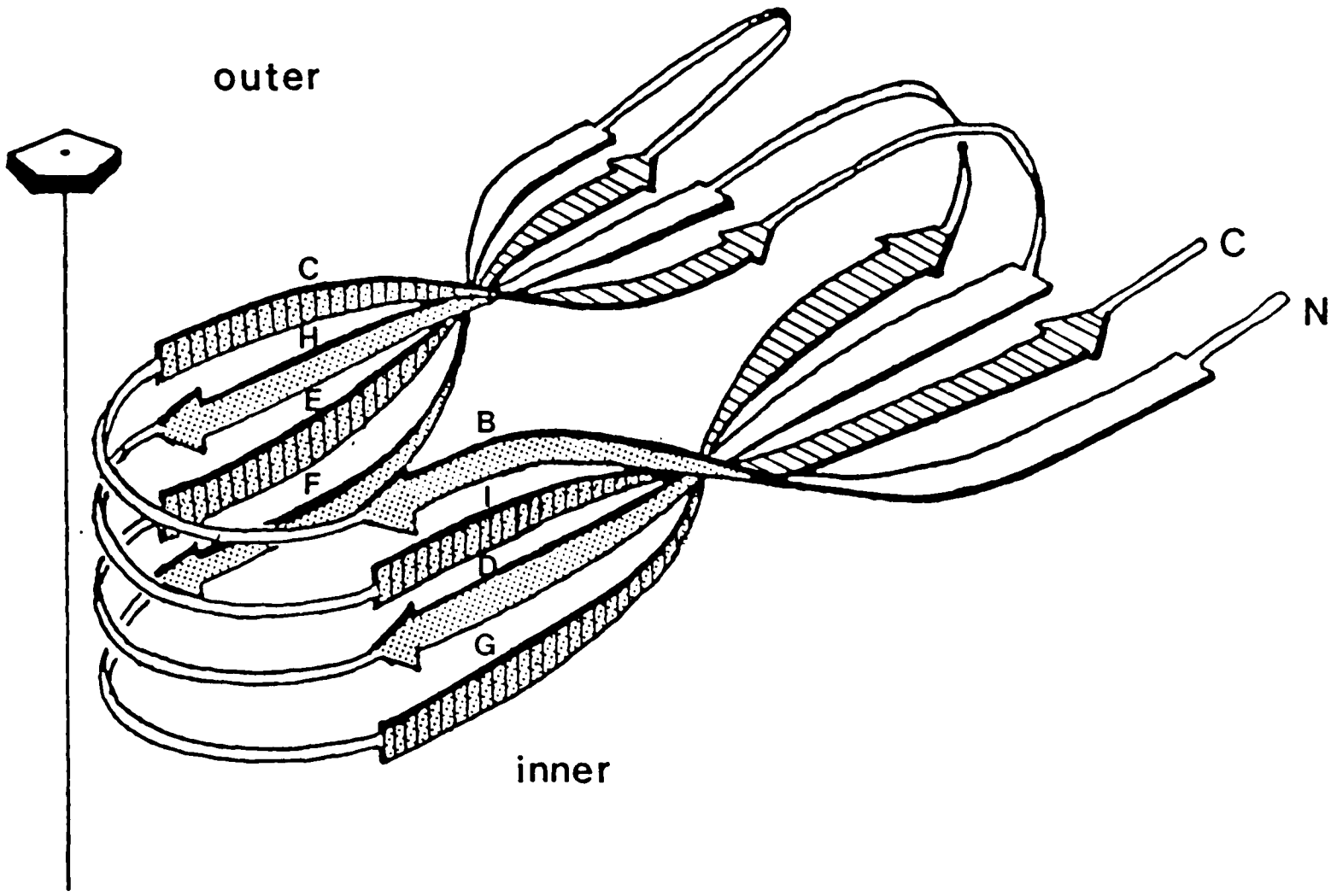


Figure 1.1: Canonical 8–stranded β-barrel

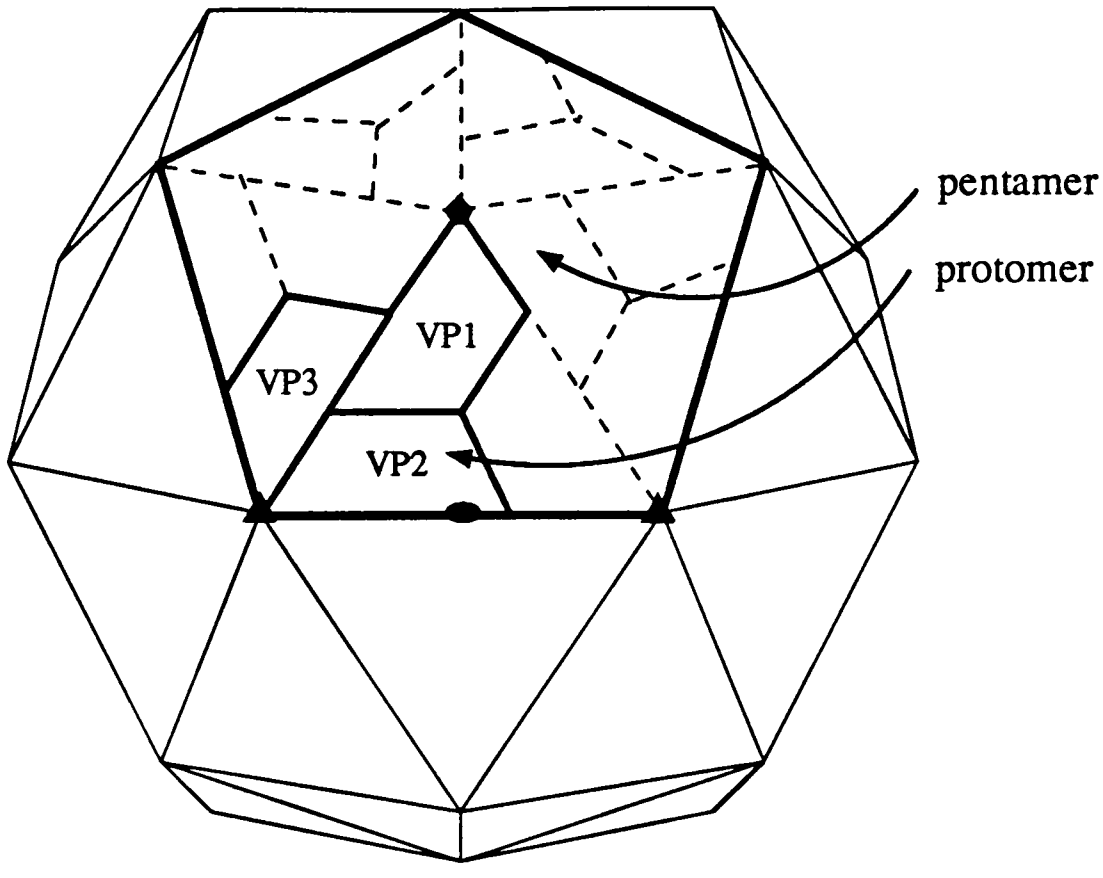


Figure 1.2: Arrangement of proteins within the icosahedral capsid.

β–sheets containing strands named B, I, D, G and C, H, E, F according to their order in the linear sequence. The barrel has a wedge–like shape which allows packing of the narrow end around the 3– and 5–fold axes, with the blunt end packing around the pseudo–3–fold axis (Figure 1.2). This core structure is conserved amongst all the picornaviruses but, variety in the length of the loops joining the strands, leads to very different surface properties.

High resolution structures now exist for representatives of four of the picornavirus genera (distinguished on the basis of physico–chemical properties, Table 1.1) namely: poliovirus types 1 and 3 (Hogle *et al.*, 1985; Filman *et al.*, 1989) for the enteroviruses, human rhinovirus serotypes 14, 1A and 16 (Rossmann *et al.*, 1985; Kim *et al.*, 1989; Oliveira *et al.*, 1993) for the rhinoviruses, mengo virus (Luo *et al.*, 1987) and two strains of Theiler's virus (Grant *et al.*, 1992; Luo *et al.*, 1992) for the cardioviruses and foot–and–mouth disease virus O_1BFS 1860 (Acharya *et al.*, 1989) for the aphthoviruses. Comparisons between sequences, physico–chemical properties and structures, show that the enteroviruses and rhinoviruses are most closely related, with the

Genus	Mol. Wt. of RNA $\times 10^{-6}$	Mol. Wt. of capsid proteins $\times 10^{-3}$	pH stability	buoyant density in CsCl (g/ml)	sedimentation coefficient (S)	diameter (Å)
enteroviruses	2.6	34, 30, 26, 7	stable 3–9	1.34	156	270–300
cardioviruses	2.7	32, 29, 25, 7	stable 3–9	1.34	156	"
rhinoviruses	2.5	32, 29, 26, 7	labile < 6	1.39–1.42	149	"
aphthoviruses	2.9	23, 25, 24, 8	labile < 7	1.43–1.45	142–146	"
hepatoviruses	2.6	30, 32, 41 (*)	stable 3–9	1.32–1.34	156	"

(*) No protein equivalent to VP4 identified

Table 1.1: Some physico–chemical properties of picornaviruses

cardioviruses located between these and the more distantly–related aphthoviruses. The hepatovirus genus contains echoviruses and coxsackie viruses which have been relatively little studied, only recently being reclassified as members of an independent genus.

1.2 FMDV

1.2.1 The Disease

Foot–and–mouth disease is amongst the most infectious of animal diseases affecting hoofed animals such as cattle, goats and sheep. The disease has been recognised for several centuries causing fever, anorexia and a drop in cattle milk production. These early signs are followed by the appearance of lesions in the mouth and feet leading to lameness and production of an excess of saliva. The disease is not necessarily fatal, but there is a loss of growth and productivity after the acute phase. In 1897 Loeffler

and Frosch (Loeffler and Frosch, 1897) noted that the infectious agent passed through the smallest available filters and that it required living cells to reproduce.

The economic effects of the disease are maximised by the ease of transmission; control has only been achieved by demanding that all infected animals are killed and incinerated to decrease the likelihood of air–borne infection over long distances. Countries free from FMD attempt to prevent outbreaks by stopping trade in potentially infected breeds with countries experiencing outbreaks. Where outbreaks are rare such slaughter can control the disease. However in countries where the disease is endemic (all continents except North America and Australia) chemically inactivated virus vaccines are used. Use of such vaccines is complicated by the need to maintain a 'cold–chain' – especially difficult in the warmer countries where FMD occurs at high levels. Protection requires high levels of circulating antibody and therefore requires frequent re–injection (2 to 3 times yearly). As for all chemically inactivated vaccines the risk of outbreaks caused by incomplete inactivation must always be borne in mind. Seven virus serotypes can be distinguished namely: O, A, C, SAT1–3 (South African Territories) and Asia-1. Infection by virus belonging to one serotype provides no protection against viruses of all other serotypes (first demonstrated by Vallée and Carré (Vallée and Carré, 1922)). However the rate at which the virus evolves in the field is so high that, even within a few months post–infection, no protection may remain against different subtypes within the same serotype.

1.2.2 Structural Characteristics

The structure of FMDV at 2.9Å (Acharya *et al.*, 1989) showed that the overall fold of proteins VP1-3 follows the same 8–stranded β–barrel pattern observed in the other picornaviruses. However the loops joining the strands are significantly truncated giving the virus a much 'smoother' surface. Truncation of VP1 is very marked by comparison with the other genera; as well as truncation of the loops, the β–strands are somewhat shortened. This leads to formation of a hole around the 5–fold axis

which allows permeation of Cs^+ ions and planar molecules, such as proflavin, into the capsid interior (Fry, 1991; Acharya *et al.*, 1990) and explains the unusually high buoyant density in CsCl (Table 1.1). As discussed below the 'canyon' implicated in receptor binding in the other genera (Rossmann *et al.*, 1985; Hogle *et al.*, 1985; Luo *et al.*, 1987; Kim *et al.*, 1989) is absent and the proposed receptor binding site is located on a surface exposed, disordered, loop (the G–H loop of VP1). The C–terminus of VP1 distinctively folds across the surface of the virus, in an extended conformation, to lie close to the G–H loop of a 5–fold related protomer, forming a complex immunogenic site (in type O_1 virus (McCahon *et al.*, 1989)). These two regions are also linked in determining receptor binding (Fox *et al.*, 1989).

VP2 is also truncated but the losses are confined to the surface loops, in particular the 'puff' following the βE strand is lost. The N–terminal residues ($\approx$ first 10) are absent from the electron density map as for the other picornaviruses. VP3 is more similar to the VP3s of other picornaviruses. VP4 is slightly larger in FMDV than in the other genera (Table 1.1). Biochemically it is known to be myristoylated at the N–terminus (Chow *et al.*, 1987; Belsham *et al.*, 1991) but no evidence concerning the location of this moiety is seen due to disorder of the first 14 residues. The myristate moiety has only been seen in one picornavirus structure namely poliovirus (Hogle *et al.*, 1985). Here it is seen to fill an hydrophobic cavity at the 5–fold axis. The nature of this cavity is somewhat altered in FMDV (Fry, 1991) and this may partially explain why no ordered myristate density is seen in FMDV. The visible portions of VP4 (15–39 and 64–85) possess little regular secondary structure except for a 7 residue α–helix beginning at residue 67.

1.2.3 The Genome

The FMDV genome (as for all picornaviruses) acts as mRNA and is translated via a single open reading frame to a single polyprotein which is cleaved into structural and

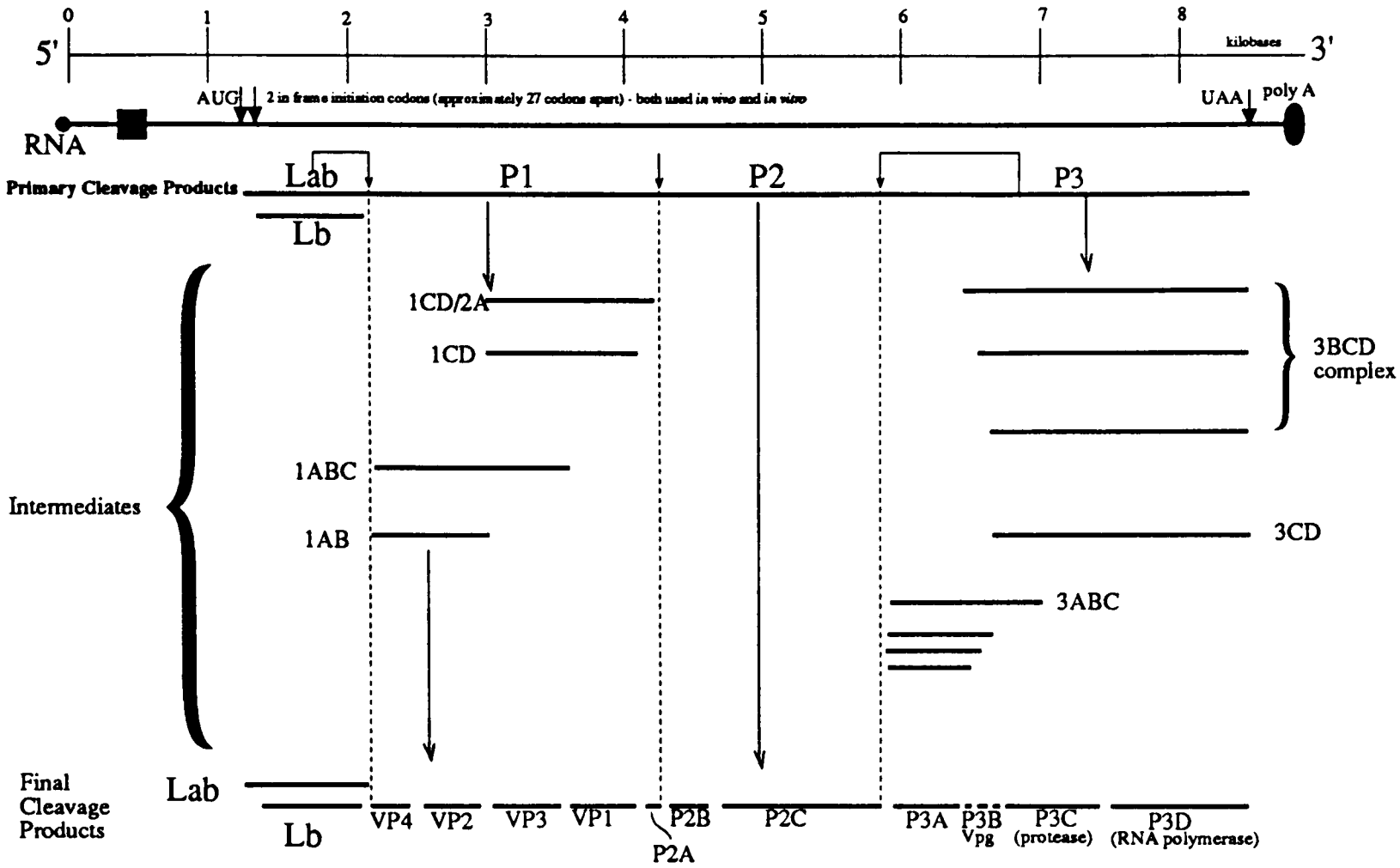
non–structural proteins. All picornaviruses have long uncapped 5' untranslated regions (UTRs) 600–1200 nucleotides in length, but the aphthoviruses have the longest 5' UTR which contains a poly–C tract (only present in one other genus - the cardioviruses). Despite these long untranslated regions the RNAs are translated efficiently. This is because initial binding of the ribosome to the RNA occurs not at the 5' end but at a site not far upstream from the correct AUG codon (there are multiple copies of the AUG). This binding of the ribosome is directed by a large piece of RNA, with a defined secondary structure, termed the internal ribosome entry site (IRES) (Pelletier and Sonenberg, 1988). The structure of this portion of the RNA is essential to its function, small insertions or deletions having drastic effects on virus growth e.g. a single transition, at position -376 (U to C) in C serotype FMDVs, causes an increase in activity of the IRES of 1.5 to 5–fold, compared to the parent, making the virus hypervirulent for BHK-21 cells (Martínez–Salas *et al.*, 1993).

In FMDV initiation occurs at two AUG codons leading to two different Leader (L) proteins, La and Lab, the latter being 28 amino acids longer (Clarke *et al.*, 1985). These are involved in host–cell shutoff by initiating cleavage of the p220 cap–binding protein complex.

The ability to translate in a cap–independent way, whilst shutting off cap–dependent translation allows reproduction of the viral RNA whilst host translation ceases. The 5' UTR contains many other regions predicted to have complex secondary structures (in addition to the IRES) - these elements are likely to be involved in initiation of RNA packaging and replication.

1.2.4 Polyprotein Processing

Processing of the single, translated polyprotein (Figure 1.3) begins during translation so that the full polyprotein is never observed. Processing intermediates are common to most picornaviruses but the three, distinct copies of VPg and the production of leader proteins are unique to the aphthoviruses and cardioviruses. The exact nature



Lab, Lb - Protease activity cleaving at Lb/P1 junction otherwise function unknown.
P1 - Structural protein.
P2A - Protease active site, P2A is larger in other picornaviruses and has been shown to be a protease involved in early cleavage.
P2B - Function unknown.
P2C - Function unknown but implicated in RNA replication.
P3A - Function unknown.
P3B - VPg - there are three different functional copies in FMDV and only one copy in the other picornaviruses.
P3C - Protease responsible for most of the maturation cleavages.
P3D - RNA polymerase - exact mechanism of replication still unclear.

Figure 1.3: Processing of the FMDV polypeptide.

of the proteolytic cascade leading to the mature proteins is not fully understood for FMDV but seems to begin with self–cleavage of the leader protein (L) and *cis* action of the 3C protease at the P2/P3 junction. The L protein inhibits cap–dependent mRNA translation (rather than 2A as in poliovirus) - demonstrated by Belsham and Brangwyn (1990). Cleavage then occurs at the C–terminal end of 2A to yield P1-2A and 2B (rather than P1 and 2AB as in enteroviruses and rhinoviruses) - this cleavage is mediated by 19 amino acids spanning the 2A region (Ryan 1991), which is only 16 residues long in FMDV.

FMDV 3C preferentially cleaves at Glu–Gly junctions (poliovirus 3C invariantly cleaves at Gln–Gly) although the larger 3C precursors, such as 3BCD, also seem to be involved in P1 processing. The cleavage of VP0 to give VP2 and VP4 - which occurs as the final maturation stage in the complete capsid, is thought to be associated with encapsidation of RNA. This association is not, however, clear–cut as empty capsids of FMDV type A_{22} - which contain no RNA, have the majority of the copies of VP0 cleaved (S. Curry, personal communication). The mechanism of this cleavage therefore remains unclear and must probably await the structure of a virus in which uncleaved VP0 is seen: interpretations of autocatalytic mechanisms based on the positions of either the N–terminus of VP2 (disordered in all picornavirus structures) or the C–terminus of VP4 do not allow for the vast structural rearrangements that must obviously accompany the cleavage.

With the exception of 2A the picornavirus proteases have sequence similarities with trypsin–like serine proteases, the active site Ser replaced by a Cys. Protease 3D is highly conserved amongst all the picornavirus genera, it acts as an RNA–dependent RNA polymerase. The 2C proteins contain some sequence motifs reminiscent of the helicase superfamily of proteins involved in promoting DNA and RNA unwinding. Sensitivity to the translational inhibitor guanidine hydrochloride has been mapped to the 2C gene but the functions of the 2B, 3A and 3B proteins remain obscure.

1.3 Effects of Immune Pressure

All RNA viruses share a great potential for variation due to the error–prone nature of RNA replication (Holland *et al.*, 1982). However the rate at which fixation of mutations occurs varies greatly between different viruses. Sobrino *et al.* (1986) have estimated that the fixation rate for FMDV is very high and varies between $< 4 \times 10^{-4}$ to 4.5×10^{-2} substitutions per site (s/s) per year (compared to 10^{-8} to 10^{-9} s/s per year for many cellular genes) during acute disease outbreaks. Villaverde *et al.* (1988) found that the rate at which mutations are fixed in the RNA coding for VP1 is 2.1 times higher than the rate for the 3D RNA. This is simply because mutations that destroy the function of 3D will result in non-viable virus and will therefore be selected against. Calculation of all these rates is complicated by the significant degree of heterogeneity observed amongst viruses isolated from the field at the same time.

Such error–produced random mutations in the capsid proteins may lead to a selective advantage for the virus if they enable escape from antibodies directed against the parental virus. The immune response generated when a virus infects an animal therefore acts to increase the rate at which mutations in antibody accessible residues are fixed, leading to rapid variation of antigenic phenotype during outbreaks. The heterogeneity of the virus population generated by this process, within an animal, makes it even more difficult for the immune system to clear the virus from the animal.

Within the aphthovirus genus the division into serotypes and subtypes is based simply on differences in the antigenic phenotypes of the viruses. It may be inferred that, at some distant time point, there was a single, ancestral FMDV from which all FMDVs today are descended. The differences between viruses in the field today simply representing the accumulation of mutations over time. Serotypes represent viruses that have diverged a long time ago and therefore no longer share any common antigenic features, whilst subtypes represent more recently separated evolutionary lines.

Esteban Domingo and colleagues have closely studied antigenic variation within C

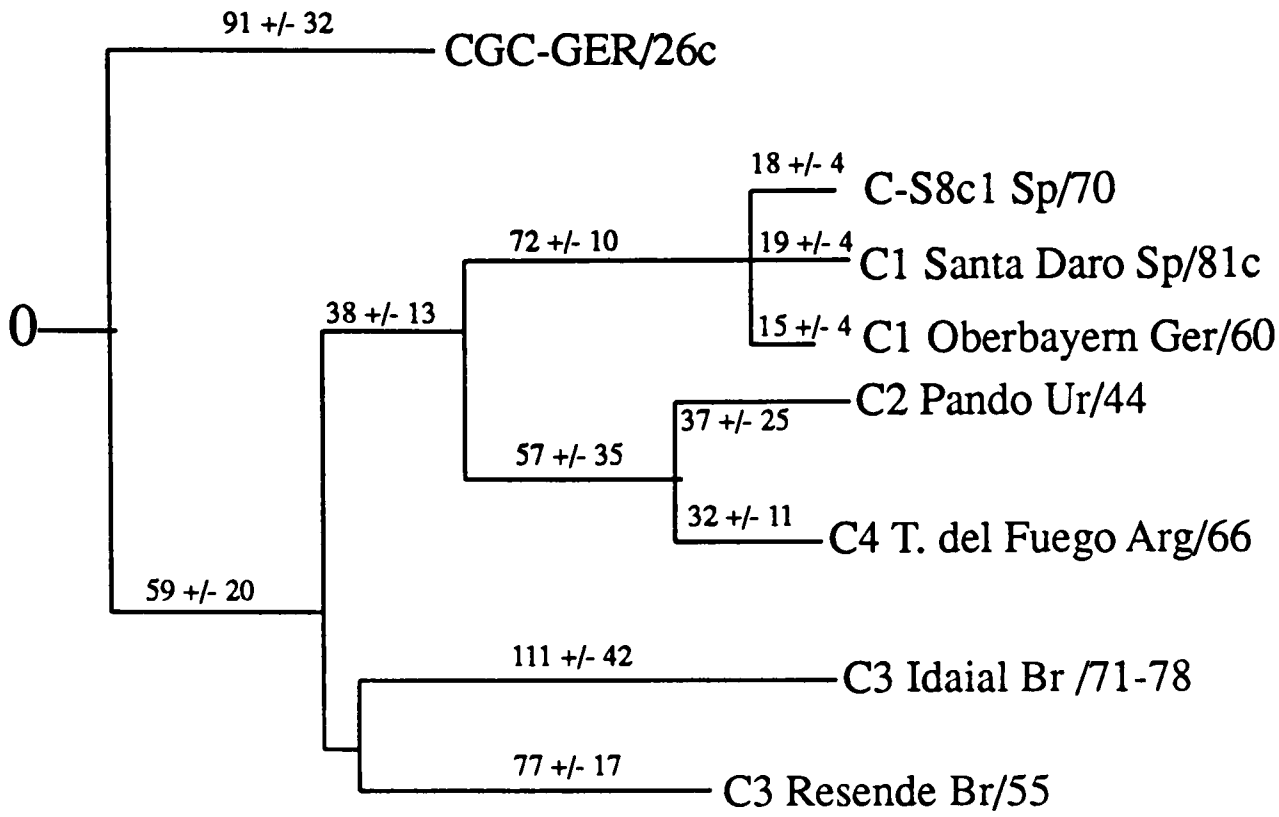


Figure 1.4: Phylogenetic tree derived from P1 RNA sequences of isolates of FMDV type C (taken from Martínez *et al.* 1992).

serotype FMDVs and believe they can place a virus sample isolated in 1926, CGC-Ger (Martínez *et al.* (1992) and Appendix A), at or near the root of an evolutionary tree for all C serotype viruses (Figure 1.4). This virus has not been seen in the field for over 50 years and seems therefore to be an example of extinction of a virus in the wild (Martínez *et al.*, 1992). The constantly changing nature of the antigenic phenotype makes control of FMDV by vaccination difficult and is therefore the focus of much work on the virus.

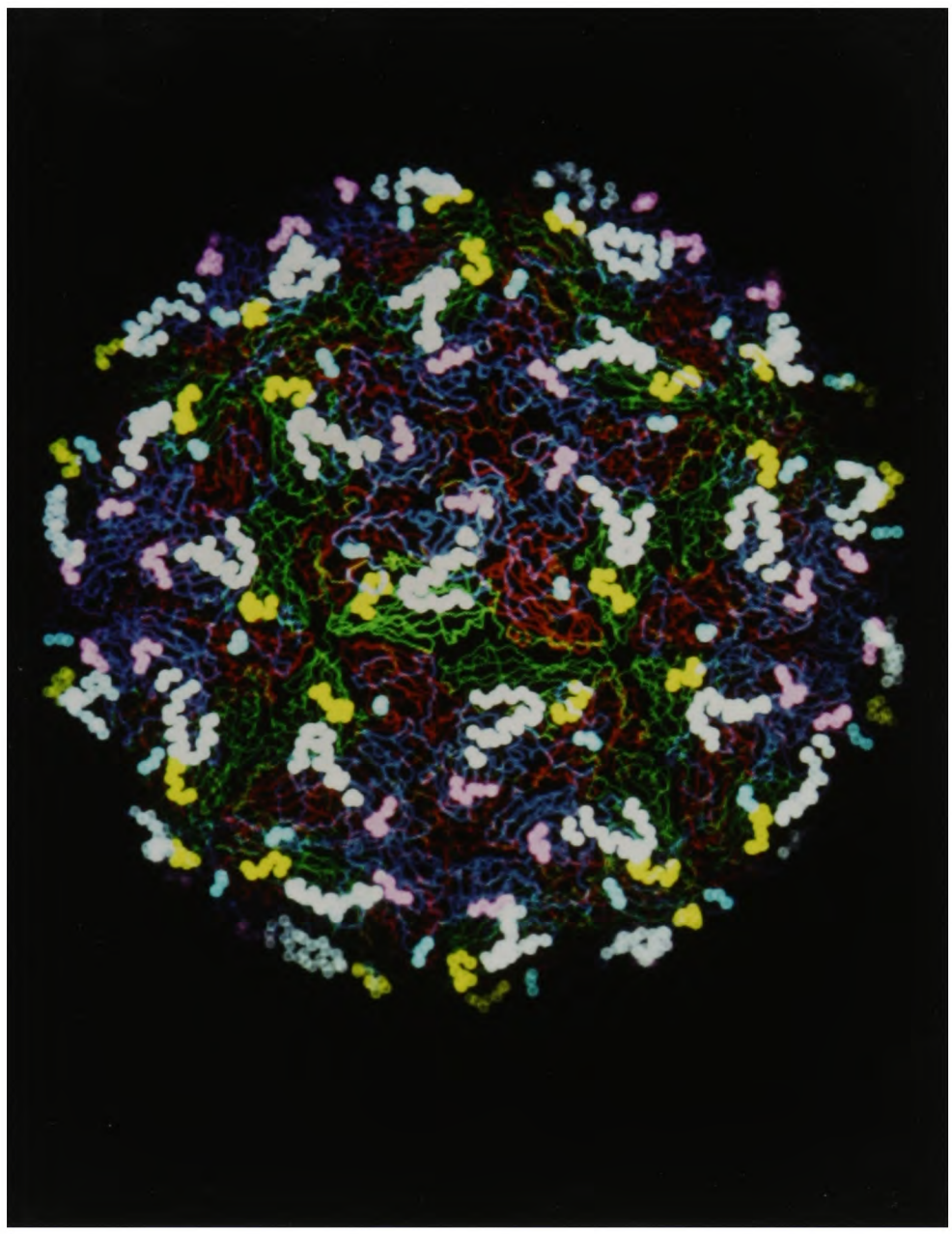
The method of choice for characterisation of picornavirus antigenic sites involved in virus neutralization has long been the isolation and sequencing of escape mutants. This approach has been used for members of all the picornavirus genera, notably to distinguish four independent antigenic sites on the surface of HRV14 (Sherry *et al.*, 1986) and also on poliovirus types 1 and 2 (Minor, 1990). Similar studies with different serotypes of FMDV have also divided the virion surface into different sites,

	VP1	VP2	VP3
O_1 (Xie *et al.*, 1987) (Kitson *et al.*, 1990) (Barnett *et al.*, 1989) (Stave *et al.*, 1988) (Pfaff *et al.*, 1989) (Crowther *et al.*, 1993)	144–159 (Site 1) 208–213 (Site1) 43–48 (Site 3) 149 (Site 5)	70–77 (Site 2)	56–58 (Site 4)
A_{10} (Thomas *et al.*, 1988)	140–160 (Group 1) 200–212 (Group 2)	80 (Group 3) 196 (Group 3)	58–61 (Group 3) 69–70 (Group 3) 136,139,195 (Group 3)
A_{12} (Baxt *et al.*, 1989)	145–168 (Site 1) 201, 209 (Site 1) 173 (Site 2)		175, 178 (Site 1)
A_{22} (Bolwell *et al.*, 1989)	137–144 (Group I) 149–154 (Group II)		
C (Mateu *et al.*, 1990) (Mateu *et al.*, 1989) (Mateu *et al.*, 1988)	140–156 (Site A) 200–213 (Site C) 193 (Site D)	70–80 (Site D)	58–61 (Site D)

Table 1.2: Antigenic sites identified in FMDV

although the precise definition of these sites varies somewhat between serotypes. The sites identified by different groups are summarised in Table 1.2 and the locations of the sites identified in O_1 FMDVs shown in Figure 1.5. As the table shows similar regions of the capsid are commonly identified as antigenic in the different serotypes but the linkage observed between the different capsid regions varies, hence the division into different sites by different groups. One site is common to all studies, the VP1 G–H loop. This region is often termed the major antigenic site of FMDV as the probability of generating a monoclonal antibody against this loop is much greater for this region than for any other region.

The determination of the structure of O_1BFS revealed no conformation for this loop as it was disordered and absent from the electron density map. However studies of escape mutants, bearing mutations in the B–C loop of VP1 (Parry *et al.*, 1990),



WHITE - site 1, the G–H loop of VP1; **YELLOW** - site 2; **MAGENTA** - site 3; **CYAN** - site 4

Figure 1.5: Antigenic sites on type O_1 BFS

gave some clues to the conformational variability of the loop. These mutant structures suggested that the loop could adopt either 'up' or 'down' conformations in response to mutations elsewhere in the capsid – mainly lying in the 'up' conformation in the original parental structure and in a 'down' conformation, lying over VP2, in the mutants. Further evidence for this conformational switching was gained when the structure of DTT (dithiothreitol) soaked crystals of O_1BFS was determined (Logan *et al.*, 1993). The reducing treatment lead to reduction of a disulphide bond linking VP1 (134) and VP2 (130) which allowed the G–H loop of VP1 to adopt an ordered conformation, lying over VP2, in a 'down' conformation. The loop has a well defined internal structure and, since, 'native' and 'reduced' virus cannot be easily distinguished using monoclonal antibodies directed at this site (personal communication, R. Abu–Ghazaleh) it seems likely that this internal conformation is maintained by the loop at all times, whatever it's orientation relative to the rest of the capsid. The disorder observed in the 'native' structure being simply due to pivoting of this region about the disulphide. Structural analysis of the virus at different times after release from the cells has shown that the reduced structure predominates in freshly released virions, the disulphide bonds forming over time such that, after 28 hours, the majority of VP1-VP2 bonds have formed (Logan *et al.*, 1993). This provides a time–dependent form of antigenic variation aiding immune escape, that can operate during the life–time of the virus, in addition to fixation of mutations that alter the antigenic properties of the virus and cause variation in phenotype between different generations of the virus i.e. acting over a longer time–scale.

1.4 The 'Canyon' Hypothesis

Receptors for poliovirus and rhinovirus have been identified and are members of the immunoglobulin superfamily. The poliovirus receptor (PVR) is a glycosylated 67KD membrane protein consisting of three immunoglobulin domains packed head–to–tail (Mendelsohn *et al.*, 1989). Virus binding properties have been localised to

domain 1. The major–group human rhinovirus receptor has been identified as the 95KD glycoprotein ICAM-1 (Staunton *et al.*, 1989; Greve *et al.*, 1989), one of the family of intercellular adhesion molecules. This molecule consists of 5 immunoglobulin domains packed head to tail, and appears as a 'bent-rod' in electron micrographs, binding to rhinovirus is largely dependent on the first domain (Staunton *et al.*, 1990).

The structures of HRV14 (Rossmann *et al.*, 1985), HRV1A (Luo *et al.*, 1989) and poliovirus (Hogle *et al.*, 1985) revealed a surface depression approximately 20Å across and 25Å deep encircling the 5–fold axes of the virion. Based on this structural feature Rossmann *et al.* (1985), (reviewed in (Chapman and Rossmann, 1993)) proposed that residues at the base of the depression would be involved in receptor binding. The shape of the depression will permit access of a slender receptor molecule (e.g. a single Ig domain) whilst preventing access of the bulkier antibody molecules (e.g. paired Ig domains). Residues around the canyon rim are hypervariable and aid immune escape, whilst residues at the base of the depression are protected from immune pressure and can remain constant, appropriate for receptor binding. In mengo virus the depression is partially filled by extensions of surface loops, but 5 discontinuous pits remain so a similar mechanism may operate.

This hypothesis is supported by much evidence including:-

1. Mutations of residues at the base of the canyon of HRV14 produce virus with altered affinity for the receptor (Colonno *et al.*, 1988).

2. Conformational changes in the canyon floor caused by binding of antiviral drugs, prevent attachment to the receptor (Smith *et al.*, 1986; Pevear *et al.*, 1989).

3. Modelling studies show that the immunoglobulin fold docks well with the HRV14 canyon structure, half of the model of ICAM-1 domain 1 fitting well into the canyon (Giranda *et al.*, 1990).

4. Cryo–electron microscopy (at 28Å resolution) of HRV16 with bound, soluble, ICAM-1 (domains 1 and 2) confirms the sight of binding as the canyon (Olson

et al., 1993; Oliveira *et al.*, 1993).

The canyon hypothesis therefore demonstrates a way in which the differing shapes of the attacking antibody molecules and the receptor can be exploited to prevent a constant receptor binding sequence becoming an antigenic liability.

1.5 The Receptor for FMDV ?

The receptor for FMDV has not been unambiguously identified but many lines of evidence suggest that it belongs to the integrin family. Much of this evidence is based on the apparent use of an RGD tripeptide as a major determinant of receptor binding. RGD is used as a recognition signal by many extracellular proteins involved in cell attachment, motility, adhesion and growth (Ruoshlati and Pierschbacher, 1987; Felding-Harbermann and Cheresh, 1993). These proteins include fibrinogen, vitronectin, osteopontin and fibronectin. If these proteins are used on affinity columns to extract other proteins from cells, a range of large, heterodimeric proteins called integrins are separated. Despite this common recognition signal many of the integrins are highly specific in binding to their ligand. This specificity often arises due to a secondary interaction between the integrin and binding protein in addition to the RGD tripeptide mediated interaction. In FMDV the RGD tripeptide is located at the tip of the VP1 G–H loop and is implicated in receptor binding in the following ways:-

1. All antibodies known to bind to the loop prevent cell attachment (Fox *et al.*, 1989).

2. Proteolytic cleavage of the loop, or the C–terminus of VP1 (which seems to be involved in the specificity–determining interaction) eliminates binding (Wild *et al.*, 1969; Cavanagh *et al.*, 1977).

3. The surface exposed residues are highly variable excepting the C–terminus of VP1 and RGD sequence within the G–H loop (Fox *et al.*, 1989), this strongly

implies these regions have important biological functions dependent on their sequence.

4. Short peptides including the tripeptide RGD, inhibit virus attachment (Fox *et al.*, 1989).

5. FMDV mutants bearing mutations in the RGD are non–viable viruses (Mason *et al.*, 1993) but can be made viable if an alternative entry route into the cells is provided using membrane bound antibody receptors and antibody–complexed virus.

6. Recognition of ligands by integrins is thought to be conformation–specific (Ruoshlati and Pierschbacher, 1987), conservative mutations within the RGD sequences having dramatic effects on binding (Pierschbacher and Ruoslahti, 1987). The conformation of the RGD tripeptide in FMDV, as revealed in the structure of reduced O_1BFS ((Logan *et al.*, 1993) and above), is very similar (RMS deviation in Cα positions 0.09Å) to the conformation observed in other RGD containing proteins known to bind integrins e.g γII–crystallin, and very different (RMS deviation in Cα positions 0.89Å) to the conformation of the tripeptide in proteins known not to be involved in integrin binding e.g. α–lytic protease.

7. All known integrins require divalent metal cations for attachment to their ligands – this may be mediated by an incomplete Ca^{2+} binding domain recognised in the α–integrin subunit. It has been suggested that this domain is similar to that found in calmodulin but lacks one of the Glu or Asp residues required to coordinate the Ca^{2+}. The Asp of the RGD tripeptide may complete the coordination. FMDV attachment to cells is Ca^{2+} dependent (Brown *et al.*, 1962).

Although the integrin hi–jacked by FMDV as a means of entry to the cell has not been identified, integrins are used by other picornaviruses as a means of cell entry. Echovirus types 1 and 8 (members of the hepatovirus genus) use a non-RGD dependent integrin VLA-2 to enter cells (Bergelson *et al.*, 1992; Bergelson *et al.*, 1993).

Coxsackievirus A9, a member of the enterovirus genus, may also use an integrin as its receptor (Chang *et al.*, 1989). Binding of this virus to the integrin is RGD dependent (the RGD sequence located near the C–terminus of VP1), although cleavage of the C–terminus exposes a second, RGD–independent receptor binding site (Roivainen *et al.*, 1991).

1.6 The 'Camouflage' Hypothesis

The structure of FMDV (Acharya *et al.*, 1989) revealed that, due to truncation of many surface loops, there was no canyon around the 5–fold axis. This implied that the mechanism employed by FMDV to hide it's receptor binding residues from antibodies would differ from that used by other picornaviruses e.g. human rhinoviruses and polioviruses. The finding of the RGD tripeptide, which is strongly implicated in receptor binding (see above), at the tip of the disordered FMDV loop suggested an alternative method to prevent the constant receptor binding region becoming an antigenic liability. The 'camouflage' hypothesis (Acharya *et al.*, 1989) suggests that, since the 'footprint' of an antibody on a protein surface is of the order of 700Å^2 (Alzari *et al.*, 1988), the incredible hypervariability of the residues surrounding the conserved RGD masks the constant region. If the footprint of the receptor is much smaller than that of the antibody it's binding determinant may be conserved with no risk of increasing the likelihood of antibody binding. The conformational variability of this loop (see above) also helps by altering the context in which the antibody will see the conserved residues, reducing the likelihood of binding. A similar camouflaging of a site by surrounding it with regions of great sequence variability is also seen in the influenza virus neuraminidase (Colman *et al.*, 1983) where the surface exposed residues involved in catalysing the release of sialic acid are surrounded by regions of great variability.

The location of the rhinovirus and poliovirus receptor sites at the base of the canyon may also be associated with the rôle of the receptor in inducing conformational

changes required for cell entry and uncoating. This is not required in FMDV where uncoating and cell entry seem to be initiated by acid–induced disruption of the virions within the acidified endosomal vesicle. Such differences in the way in which these viruses enter cells is also confirmed by recent studies (Mason *et al.*, 1993) which have shown that the Fc receptor of cells lacking the true virion receptor can act as a receptor for antibody–complexed FMDV, allowing infection of otherwise resistant cells, but, although binding to the FcR must still occur for virus–antibody complexes of poliovirus, infection does not result suggesting some additional receptor–mediated process is required.

Chapter 2

Crystallographic Techniques

2.1 Theory of X–Ray Crystallography

Protein crystallography involves the application of X–ray diffraction techniques to record information about protein molecules existing within an ordered crystal lattice. These X–rays interact with the electrons of atoms within the crystal to give coherent scattering. The electrons are stimulated by the incident X–rays and become secondary sources of scattered radiation. The diffraction pattern being the Fourier transform of the convolution of the aperiodic molecule with the periodic crystal lattice. The theory of diffraction is described in detail in Blundell and Johnson (Blundell and Johnson, 1976) and Cantor and Schimmel (Cantor and Schimmel, 1980) amongst other texts.

2.1.1 The Geometry of X–ray Scattering

Scattering by the crystal lattice may be considered simply as scattering by two point–electrons A and B (Figure 2.1). Interference between these waves will be determined by the relative phase of the diffracted vector s. The path length difference

$$\begin{aligned} AC - DB &= (\mathbf{r}\cdot\mathbf{s} - \mathbf{r}\cdot\mathbf{s_0}) \\ &= \mathbf{r}\cdot(\mathbf{s} - \mathbf{s_0}) \end{aligned}$$

If $S = \frac{\mathbf{s}-\mathbf{s_0}}{\lambda}$ then the phase difference is

$$\Delta\phi = 2\pi\mathbf{r}\cdot\mathbf{S}$$

For constructive interference to occur $\Delta\phi$ must be equal to $n2\pi$ (where n is integral) and $\mathbf{r}\cdot\mathbf{S}$ must therefore be integral. If A and B form part of the crystal lattice with

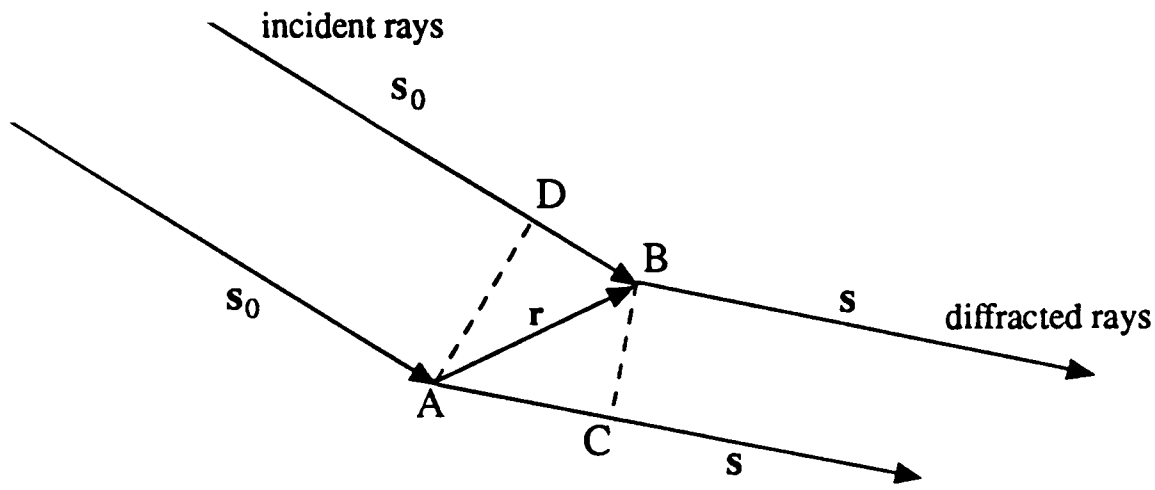


Figure 2.1: Scattering of X-rays

vectors **a**, **b** and **c** defining the unit cell then

$$\mathbf{r} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$$

where u, v and w are integers. Since $\mathbf{r} \cdot \mathbf{S}$ must be integral each component must be integral

$$\mathbf{a} \cdot \mathbf{S} = h$$

$$\mathbf{b} \cdot \mathbf{S} = k$$

$$\mathbf{c} \cdot \mathbf{S} = l$$

These equations are the Laue equations and h, k, l are the Miller indices defining a unique plane of reflection (i.e. a diffraction spot). Bragg's law provides a physical solution to these equations and predicts the angle of reflection of any diffracted ray from specific atomic planes such that

$$n\lambda = 2dsin\theta$$

The ends of the scattering vector **S** define a set of planes perpendicular to **a**, **b** and **c** respectively, the intersections of which form a 'reciprocal' lattice with primitive

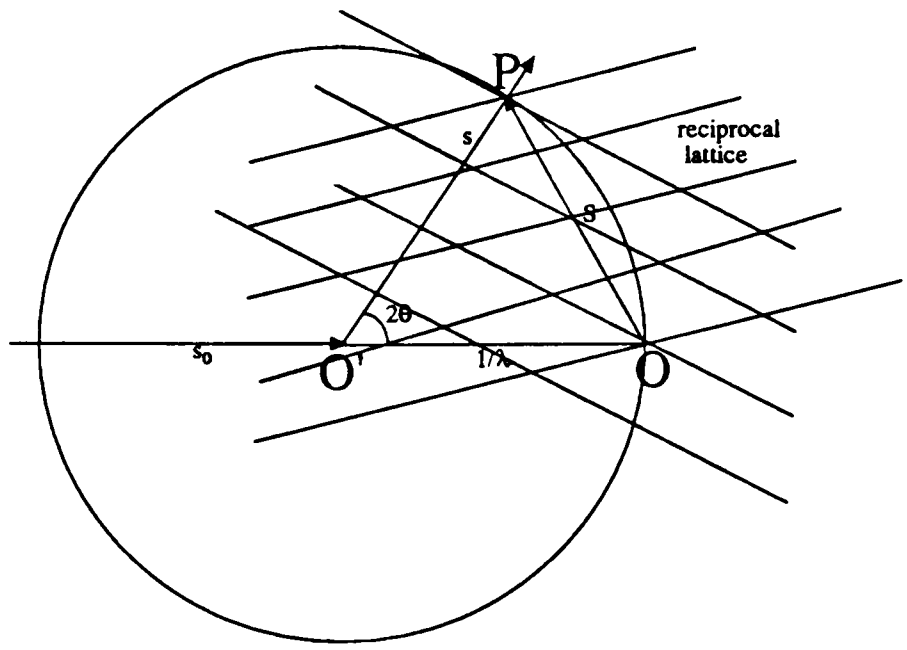


Figure 2.2: The Ewald Sphere

translations

$$\mathbf{a}* = \frac{1}{V}(\mathbf{b} \times \mathbf{c})$$

$$\mathbf{b}* = \frac{1}{V}(\mathbf{a} \times \mathbf{c})$$

$$\mathbf{c}* = \frac{1}{V}(\mathbf{a} \times \mathbf{b})$$

where V is the volume of the unit cell.

The Ewald sphere (Figure 2.2) is a geometrical construction which enables visualization of the Laue or Bragg conditions. As previously s and s_0 are represented by vectors of length $\frac{1}{\lambda}$. The reciprocal lattice origin is arbitrary and is placed on the surface of a sphere of radius $\frac{1}{\lambda}$, at the point of intersection of the incident beam. The crystal is located at O'. The angle $\angle OO'P = 2\theta$ and it can be seen that $\mathbf{r} \cdot \mathbf{S}$ can be integral only when a reciprocal lattice point (P) cuts the sphere. In such a position $OP = \frac{2sin\theta}{\lambda} = \frac{1}{d}$, and therefore $\lambda = 2dsin\theta$.

2.1.2 Scattering Of X–rays By The Molecular Crystal

Scattering by crystals consisting of real molecules with atoms in general positions may be considered by taking an infinitesimal electron density element $\rho(V)$ in a volume dV. It's scattering relative to an electron at the origin may be represented as $\rho(V)dVexp(-2\pi i\mathbf{r}_p \cdot \mathbf{S})$ where $\mathbf{r}_p$ is the position vector with respect to the origin of the unit cell. The scattering by the whole cell may therefore be represented by

$$F(S) = \int_v \rho(V)exp(2\pi i\mathbf{r}_p \cdot \mathbf{S})dV$$

since $\mathbf{r}_p = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$

$$F_{h,k,l} = \int_x \int_y \int_z \rho(x,y,z)exp(2\pi i(hx + ky + lz))dxdydz$$

The electron density may also be modelled as a summation over a number of spherical atoms so that

$$F_{h,k,l} = \sum_j f_j exp(2\pi i(hx_j + ky_j + lz_j)) \tag{2.1}$$

where f_j is the atomic scattering factor of atom j and is dependent on the number of electrons in the atom. This equation is for ideal point atoms, fixed in space and is therefore an oversimplification of the crystal. The assumption is usually made that the electron density of atoms is spherically symmetric and hence that the atomic scattering factor is also spherically symmetric. This expression is for stationary atoms. However, the thermal motion of atoms will affect the scattered intensities. Debye (1914) showed that the thermal motion of atoms causes a decrease in the intensity such that

$$\hat{f}_j = f_j exp(\frac{-Bsin^2\theta}{\lambda^2})$$

where $B = 8\pi^2\bar{u}^2/3$ and $\bar{u}^2$ is the mean square displacement of atoms along the normal to the reflecting plane. Atoms within a protein are not free to vibrate equally in all directions and this isotropic B factor should really be replaced by an anisotropic

temperature factor. Six parameters are required to model this vibration which commonly results (except at very high resolution e.g. 1.5Å) in too many parameters for the limited number of X–ray observations.

The Fourier inverse of Equation 2.1 gives the electron density equation

$$\rho(x,y,z) = \frac{1}{V}\sum_{h,k,l} \mid F_{h,k,l} \mid exp(i\alpha_{h,k,l})exp[-2\pi i(hx+ky+lz)]$$

This equation illustrates the 'phase problem' – $\mid F_{h,k,l} \mid^2$ is the measured quantity (i.e. intensity), $\alpha_{h,k,l}$ is unmeasured and must be separately determined. Phases may either be determined directly by isomorphous replacement or anomalous scattering or may be inferred by use of a known, related protein structure which, when correctly positioned and oriented, may be used to calculate initial, approximate phases for the unknown structure.

2.2 Data Collection

2.2.1 Use of Synchrotron Radiation

Virus crystals are almost universally small (i.e. <1mm in all dimensions), this combined with the large unit cell required to accommodate the virion results in relatively few unit cell repeats within the crystal to yield diffraction data. Use of a synchrotron radiation (SR) source enables the most to be gained from such crystals – although it is not always necessary, the structure of poliovirus was determined using only a laboratory source (Hogle *et al.*, 1985). The use of synchrotron radiation in protein crystallography is reviewed in detail in Helliwell (1992). SR has several characteristics which make it especially suitable for collection of data from such crystals.

1. Intense

2. Tunable to short wavelengths

3. Finely focussed

By matching the size of the beam to the crystal (using a fine collimator and bringing in horizontal and vertical slits), the amount of non–crystal, extraneous matter in the beam may be minimized leading to an enhanced signal to noise ratio. The combination of the small horizontal and vertical beam divergence with the low mosaicity of FMDV crystals also enhances the useable oscillation range on each image. With tight collimation adjacent spots can be resolved and the intensity of the non–spot, 'background' (i.e. the noise caused by diffuse scatter etc.) estimated from the areas between them.

Use of shorter wavelengths has also been shown to reduce radiation damage (Acharya *et al.*, 1989) to which virus crystals are particularly sensitive. The data presented herein have been collected using radiation of $\approx$ 1.488Å (Station 7.2 the SRS Daresbury) and $\approx$ 0.89Å (Stations 9.5 and 9.6). FMDV crystals rarely yield more than one useful image at the longer wavelength but up to 30 useful images (trial involving O_1BFS and the MAR Research image plate, Station 9.5) have been obtained from one crystal at the shorter wavelength. Absorption effects are also approximately proportional to λ^3 and so are minimized by use of shorter wavelengths.

2.2.2 Disease Security

Crystals are only allowed out of the IAH Pirbright for limited periods of time under MAFF disease security arrangements. These require that the crystals are mounted at Pirbright in wax–sealed quartz capillary tubes which must be dipped in a solution of citric acid and detergent before being transported to Daresbury in a locked metal box. The crystals must at all times be supervised by a disease security officer (Dave Goodridge) and returned to Pirbright within a specific time (24–60 hours). The extreme acid lability of FMDV means that safety at the synchrotron may be maintained by spraying the area surrounding the crystal containing capillaries with a mixture of citric acid and detergent. After exposure all capillaries are broken in a bath of citric acid and detergent.

2.2.3 Crystal Alignment

A maximum oscillation range of 0.5° is generally used with FMDV crystals. This has previously been determined by consideration of reflection overlap near the edge of the diffraction pattern (Fry, 1991). If the crystal is ideally oriented such that the beam lies along the $[\bar{1}0\bar{1}]$ axis a good distribution of symmetry related reflections will be obtained within an oscillation of 20° about the rotation axis. Crystal orientation is essentially random, however, the rhombic dodecahedra morphology often leads to alignment of the $[\bar{1}0\bar{1}]$ axis normal to the capillary wall.

The exposure is generally divided into multiple passes (5–15) over the oscillation range to smooth out long term fluctuations in the beam intensity. Crystals are centered at a given point using an optical goniostat then positioned accurately in the beam relative to the position of the direct beam recorded using radiation sensitive 'pink' or 'green' paper. Such accurate positioning is essential when using small crystals and a finely collimated beam.

2.2.4 The Detection System

Two X–ray detection systems have been employed during the course of this work, namely X–ray sensitive photographic film and imaging plate detectors. Detection systems have been compared in detail by Arndt (1986) but briefly, film is one of the few devices with the spatial resolution required to record virus diffraction data – when scanned using a $50\mu m$ raster a $12.5cm^2$ film yields 2250^2 useful pixels. The dynamic range is somewhat limited (0-255) but can be increased by using film packs containing 3 CEA Reflex films, the second and third separated by a blank film. The films are wrapped in one sheet of black paper and preceded by a double thickness of black paper to prevent blackening by stray light. The detection quantum efficiency is somewhat low, especially so at short wavelengths and 'chemical fogging' increases the 'noise' level. The 'difficulties' of handling and processing film means that the number of films used is minimized by use of a 'large' rotation range leading to build up of the

background. Imaging plate technology is based on EU^{2+}–doped BaFBR phosphor in which electrons are trapped in metastable states when excited by incident X–rays. When scanned by red He–Ne laser the stimulated luminescence is proportional to the number of X–ray photons absorbed. The plate can then be erased using visible light and re–used. The plate has a dynamic range of $\approx 10^5$ and has a much better quantum efficiency than film over the λ range we are interested in. Imaging plates however have two drawbacks:

1. The duty cycle i.e. the time taken to scan and then erase the plate ready for new data collection ranges between 90s and 12min for different systems compared to a few seconds to rotate a film carousel. This is compensated for (in part) by the increased detection efficiency which means that weak data can often be collected more quickly using such systems.

2. The pixel size (generally $\approx 150\mu m^2$) means that a large plate must be used to enable separation of spots. With FMDV data collection and a 90mm radius circular plate we generally limit the resolution of the data collected to 3.0Å (c.f. 2.6Å for film) – but this resolution is entirely sufficient to answer many biological questions.

Other factors relevant to collection of FMDV data have been discussed in detail in Fry (1991).

2.3 Data Processing

Films are digitised, prior to processing, using an Optronics P–1000 scanning microdensitometer and a raster of $50\mu m$ to produce an image of 2250 × 2250 pixels (described in detail in Logan (1991)). Data were also recorded using the MAR Research image plate on Station 9.5 (SRS Daresbury). This has a radius of 90mm and consists of 1200 150 × $150\mu m^2$ pixels.

2.3.1 Determination of Crystal Orientation

'Indexing' requires the determination of the crystal orientation in the beam with respect to an ideal relationship between the reciprocal lattice axes and the camera axes. If this ideal relationship is expressed as a matrix [**A**] then all orientations may be defined relative to this using three missetting angles ϕ_x, ϕ_y, ϕ_z around the the camera axes **X***, **Y***, **Z*** (defined in Figure 2.3). This may be expressed in the matrix [U]. [U][**A**] therefore defines the relationship between the axes. If this taken as the position at rotation angle $\phi = 0$ then the camera frame coordinates of any reciprocal lattice point is related to the indices *h,k,l* by:

$$\begin{bmatrix} X* \\ Y* \\ Z* \end{bmatrix} = [\Phi][\mathbf{U}][\mathbf{A}] \begin{bmatrix} h \\ k \\ l \end{bmatrix}$$

Where [Φ] corresponds to a rotation of ϕ about **Z***.

All the data described herein have been autoindexed using programs based on the algorithm of Wolfgang Kabsch (Kabsch, 1988). These programs derive indexing solutions from one (or more) small oscillation–range image by treating it as a still image taken at the mid–point of the ϕ range. Various implementations of this algorithm have been used in indexing – AUTO (D. Stuart, R. Acharya, E.Fry) for indexing of film images and REFIX / IDXREF (A. Leslie) for indexing of image plate images.

The first step requires generation of a list of spots with the x,y position on the film/image, a low intensity threshold is used to pick as many spots as possible and the spots selected plotted to a printer or the screen so that the intensity threshold may be monitored to prevent detection of 'false' spots in the background noise. A set of reciprocal space vectors is obtained from these coordinates and a list of low–resolution difference vectors $\mathbf{p}_i$ calculated and stored in a three–dimensional histogram. The most densely populated clusters of difference vectors are sorted on population. Current estimates of the unit cell and space group parameters are used to generate very low order difference vectors, which may be used to assign indices $\mathbf{h}_1$ and $\mathbf{h}_2$ to $\mathbf{v}_1$ and

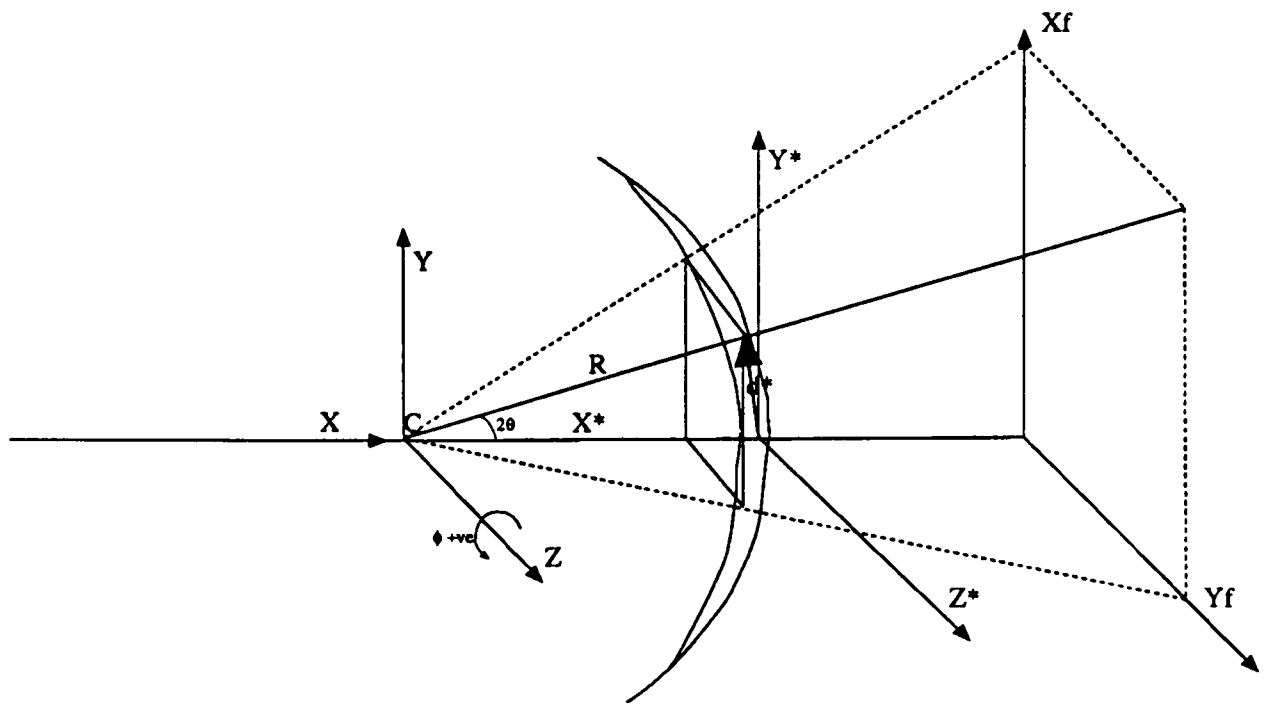


Diagram taken from Fry (1991). The Ewald construction is shown. Camera axes X*, Y* and Z* are defined in reciprocal space. Xf and Yf represent coordinates in the plane of the film. **R** is the vector from the centre of the sphere to some lattice point in a diffracting position. The magnitude of **R** will be one for a parallel monochromatic beam and a perfect crystal. **d*** is the vector from the origin of the reciprocal lattice to this point with magnitude d* given by $2sin\theta/\lambda$. The order H_i of the circle on which the i^{th} spot lies is given by

$$H_i = \frac{(1 - cos2\theta_i)}{d*}$$

where $d = \frac{\lambda}{d}$ (in reciprocal lattice units) and d is the length of the axis along the beam.

Figure 2.3: Coordinate system on the oscillation camera.

$\mathbf{v_2}$, the most populated difference vectors separated by 45°. The quantity

$$E = ([\mathbf{U}][\mathbf{A_0}]\mathbf{h_1} - \mathbf{v_1})^2 + ([\mathbf{U}][\mathbf{A_0}]\mathbf{h_2} - \mathbf{v_2})^2$$

may be minimized (Kabsch, 1976) to find an initial estimate for **U**. The top few solutions are kept. U is used to generate a list of short difference vectors $\mathbf{p}_{ij}$ and the indices $\mathbf{h}_{ij}$ determined. Indices $\mathbf{h}_i$ for the vectors $\mathbf{p}_i$ in the list already determined are found by minimising

$$H = \frac{1}{4}\sum_i \sum_j \omega_{ij}(\mathbf{h_i} - \mathbf{h_j} - \mathbf{h_{ij}})^2 + \frac{1}{2}(\sum_i \mathbf{h_i})^2$$

where $\omega_{\mathbf{ij}} = 1$ if i and j are close, 0 otherwise. Reflections may now be indexed locally with respect to the centroid of the nearest cluster. A better matrix **U** may then be obtained for global indexing by minimising

$$\sum_\mu \sum_j [[\mathbf{U}][\mathbf{A_0}]\mathbf{h_j^\mu} - (p_j^\mu - \mathbf{t}_\mu)]^2$$

where μ represents a connected set and $\mathbf{t}_\mu$ is the centroid of reciprocal lattice points p_j^μ. The relative offsets calculated as .

$$\mathbf{h}^\mu = ([\mathbf{U}][\mathbf{A_0}])^{-1}\mathbf{t}_\mu$$

may be added to each local index $\mathbf{h_j^\mu}$. Finally the offset common to all indices may be found and subtracted. Camera geometry and cell constants may now be refined starting at a low resolution and extending to a higher resolution by minimising the deviations between observed and calculated spot positions and spindle angles.

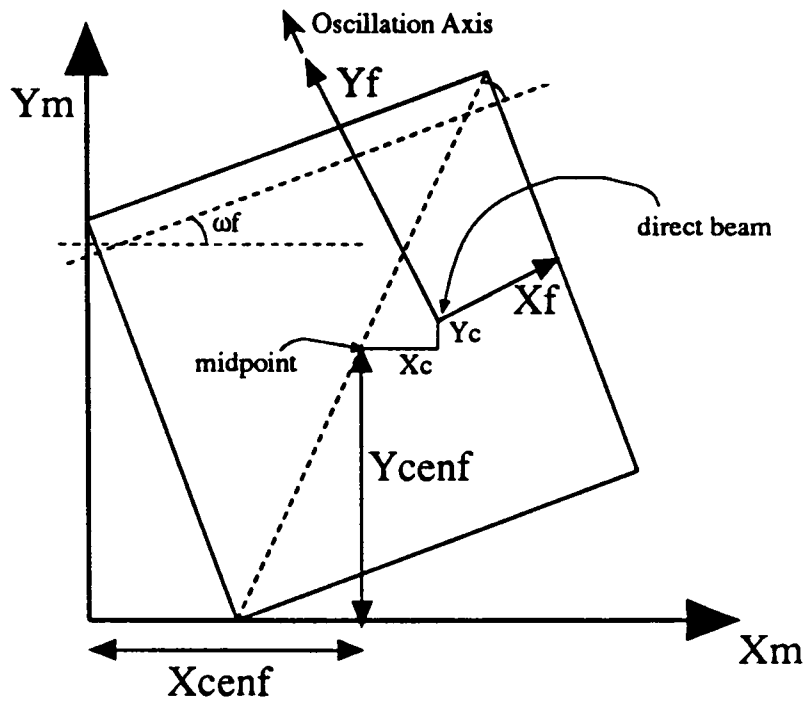


Figure 2.4: Camera geometry for MOSCO10

2.3.2 Spot Integration

The crystal orientation along with synchrotron parameters, estimates of the crystal mosaicity and radial cutoffs defining the useful region of the 'film' are used to predict spot positions in the 'film' coordinates. These positions (and the parameters used to produce them) are written to a file with the fraction recordedness and 'overlap' flags (OXGEN, OSCGEN). Initially camera parameters (defining misalignment of the camera to ideal) and scanner parameters (for image plate data, defining geometrical offsets to the scanner readings) are refined using low–resolution spots. Figure 2.4 illustrates these parameters.

Spots greater than $\frac{7}{8}$ recorded are profile fitted using a least squares fit to an average starting profile (calculated from a few strong spots covering the entire area

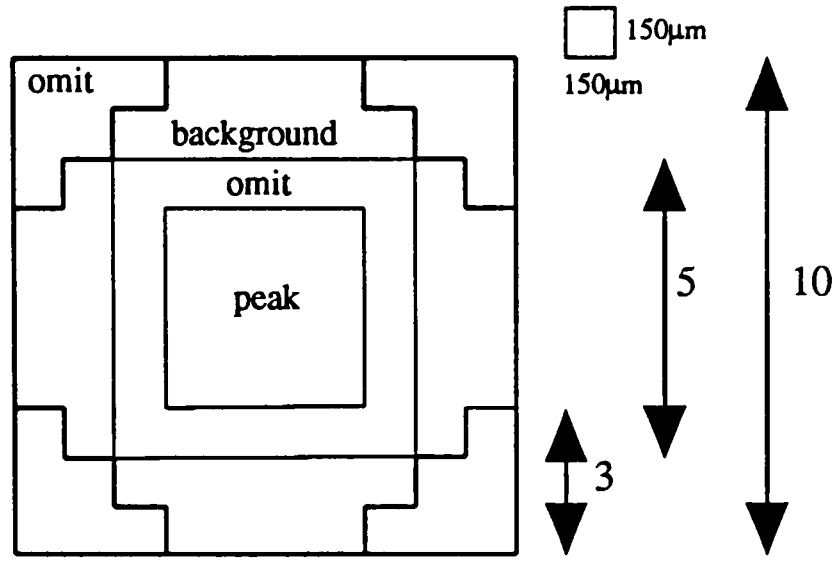


Figure 2.5: Spot box used for integration in MOSFLM

of the 'film'). The profile is continuously updated to allow for changes in spot shape over the 'film'. Background bit maps are set to designate whether pixels are in a spot, the background, spot and background or two spots simultaneously. Only those pixels not overlapping other spots are used for profile fitting, saturated pixels are also omitted. For film data an integration box of $\approx 350 \times 450 \mu m$ is used, for image plate data a box as shown in figure 2.5 is used.

Several programs have been used to integrate the data including one that diverged from the Cambridge film evaluation program MOSCO many years ago (MOSCO10, D. Stuart). Within this program the background is modelled by a plane Ax + By + C the areas for estimation being arranged around the spot with mm symmetry. Multiple passes are made with spots deviating badly from the model rejected in later passes and the model recalculated. Rejection criteria include:

1. The slope of the background plane is too great
2. Background intensity is too high
3. Background intensity is too low
4. Background is non–linear i.e. the diagonally opposite values of the plane are too different
5. Too many pixels are rejected from the background during the later passes
6. Too many saturated pixels within the spot

For image plate data MOSFLM (A. Leslie) has been used to integrate the data. Different programs give different statistics to evaluate the success of integration, however the distribution of I in $\frac{I}{\sigma(I)}$ bins gives an idea of the data quality.

2.3.3 'ABSCALE'

Film and image plate versions of ABSCALE are used to:

1. Calculate scale factors between A, B and C films from the same pack (film data only) according to the method of Fox and Holmes (1966).
2. Reduce the data to the unique set under the appropriate point group
3. Apply polarization and Lorentz corrections to the intensities

The polarization correction is based on modelling the X–ray beam incident on a crystal as a set of electrical vectors oscillating in all directions perpendicular to the directional vector. This may be modelled more simply by two components in perpendicular directions. After interaction with the electrons in the crystal the component parallel to the Bragg plane will be unaffected, but the component perpendicular to this plane will be attenuated by a factor dependent on $\cos 2\theta$. The intensities are

therefore corrected by a factor $\frac{1}{P}$. For an unpolarized incident beam (i.e. X–ray source without a monochromator) the polarisation may be expressed as:

$$P_0 = \frac{1}{2}(1 + cos^2 2\theta)$$

however synchrotron radiation is much more polarized in the horizontal plane so that the correction

$$P = P_0(1 + \tau'\alpha)$$

(Kahn *et al.*, 1982) must be applied where τ' is a measure of the polarization of the incident beam

$$\tau' = \frac{I'_0 - I'_\pi}{I'_0 + I'_\pi}$$

I'_0 and I'_π are the horizontal and vertical components of the polarized intensity,

$$\alpha = \frac{sin^2 2\theta cos 2\psi}{1 + cos^2 2\theta}$$

where ψ is the angle between the projection of the diffracted beam and the horizontal axis of the synchrotron.

The Lorentz correction allows for differences in intensity linked to the time the spot is in contact with the Ewald sphere. This is dependent on the velocity with which it travels through the sphere

$$L = (sin^2 2\theta - \zeta^2)^{-\frac{1}{2}}$$

where ζ is the projection of the reciprocal lattice vector on the rotation axis. L therefore varies from $\frac{1}{sin 2\theta}$ at the pole of the image (rotation axis horizontal) to ∞ at

λ	100mm	300mm	500mm	1000mm
1.488	0.90	0.73	0.59	0.35
0.89	0.98	0.93	0.89	0.80

Relative to 89% transmission of CuKα by 100mm air path (International Tables, Volume III)

Table 2.1: Absorption effects at different path lengths

the equator. Reflections lying too close to the rotation axis are discarded as L varies greatly even within a spot (Wonacott, 1977).

The recorded intensity is related to the 'true' intensity by

$$I = I_0 e^{-\mu t}$$

where μ is the absorption coefficient for the material the X–rays are passing through and t the thickness of the material. No correction is made for absorption by the crystal, the errors introduced by this approximation being small e.g. for a $0.3 \times 0.3 mm^2$ crystal cross–section in the beam, transmission of the (λ=0.9Å) incident beam will vary between 0.94 (at $\phi = 0$) to 0.91 (at $\phi = 45°$), causing an average error in the recorded intensity of 0.5% (Stuart and Walker, 1987). The effects of absorption by the 'air–path' between the irradiated crystal and the detector are shown in Table 2.1. This shows the fraction of the true intensity recorded for different crystal–to–detector distances and emphasises the importance of the short wavelength radiation provided by the synchrotron in allowing us to ignore these effects ($\mu \approx a\lambda^3$).

2.3.4 Determination of Indexing Hand

In space group I23 the auto–indexing solutions will divide into two hands since the crystal lattice possesses geometric symmetry that is not expressed in the intensities (i.e. there is no 4–fold axis). Autoindexing is independent of intensity information and can therefore produce equivalent solutions which are ambiguous with respect to permutation of the indices *h,k,l* to *k,h,l*. This may be thought of as an ambiguity

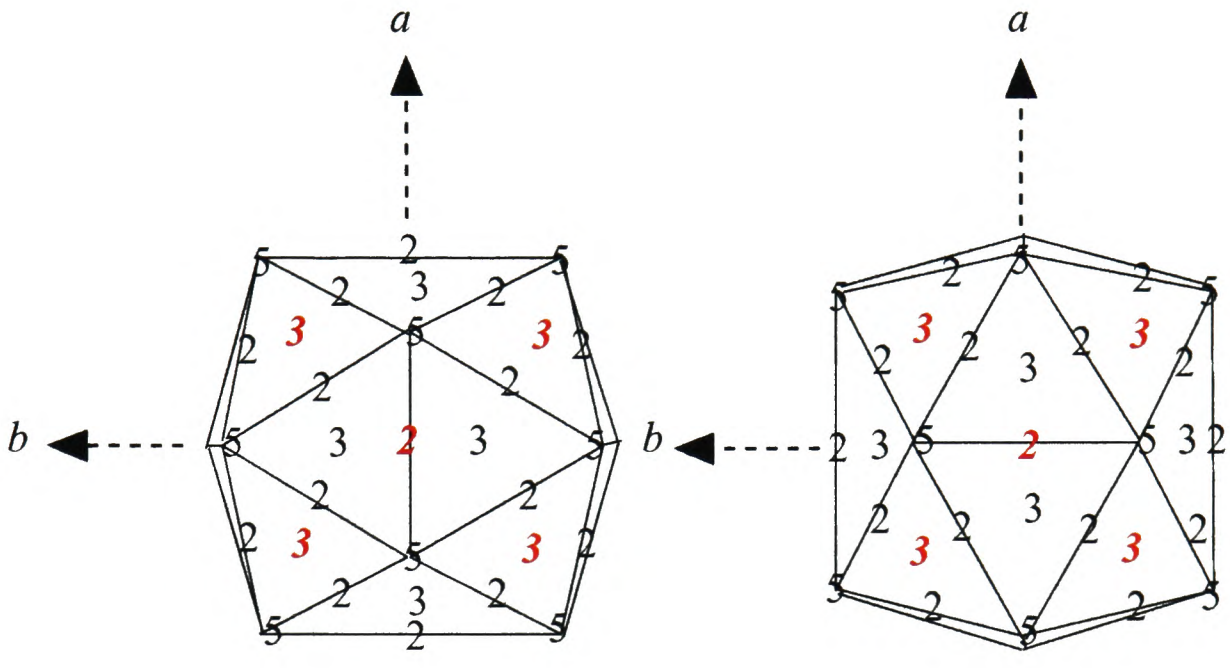


Figure 2.6: Two equivalent orientations of an icosahedron on a point of 23 symmetry

in the orientation of an icosahedron on a point of 23 symmetry (Figure 2.6). The icosahedron may be placed with its 5–fold axis in one of two directions, these being related by a 90° rotation about a particle 2–fold/cell axis. Postrefinement of the two groups of solutions together is unsuccessful (Fry, 1991) therefore a correlation coefficient is calculated (REINDLCF) between data from each crystal and a reference data set

$$C = \frac{\sum_h (< I_h > - I_h)(< I_{h,ref} > - I_{h,ref})}{[\sum_h (< I_h > - I_h)^2 \cdot \sum_h (< I_{h,ref} > - I_{h,ref})^2]^{\frac{1}{2}}}$$

to enable separation of the crystals into unswapped and swapped sets. C will be >75% where images are indexed similarly and <20% where the indices are permuted. Data are processed as two streams in parallel until, after postrefinement, the indices

of the 'swapped' set are permuted and the two streams merged.

2.3.5 Scaling and Postrefinement

Postrefinement involves refinement of the crystal missetting angles, and other geometric parameters, to optimise the classification of reflections as fully or partially recorded. Since the crystal has a finite reflection range and the beam a finite cross-fire, a single small oscillation image will contain many reflections ($\approx 25\%$ in the case of FMDV) that are only partially recorded. Since we often have no images from abutting oscillation ranges, these cannot be matched with the equivalent reflections on other images. By accurately estimating the degree of recordedness we can ensure that any partials[1] discarded are not mis–classified fullys[2], and we can also scale up some partials to fullys if the rocking curve (derivative of I with respect to rotation) is accurately modelled leading to accurate estimation of the fraction recordedness.

We use the methods of Schutt and Winkler (1977) as implemented in the program POSTREF (modified by P. Evans) incorporating refinement of anisotropic beam divergence and modelling of spectral dispersion and polarization (specific characteristics of synchrotron radiation). Partially recorded reflection intensities I'_{obs} are compared to the mean I of safe fully recorded equivalent reflections I'_{ref}. Thus the observed degree of partiality can be related to the calculated value P_{cal} derived from the crystal setting parameters. By minimising

$$P_{hi} = \sum_i \omega_i (I'_{obs} - P'_{cal} I'_{ref})^2$$

(w_i is the weight applied to reflection i) with respect to the crystal parameters optimum values of these parameters and the fraction recordedness P'_{cal} are obtained. ϕ_x may not be refined from a single image since the rotation has no component normal

[1] 'partials' are those reflections for which the reflecting range of the reciprocal lattice point lies only partially within the oscillation range of the image.

[2] 'fullys' are those reflections for which the reflecting range of the reciprocal lattice point lies entirely within the oscillation range of the image.

to the sphere of reflection, and is therefore fixed. An isotropic cosine function has proved to be the best rocking curve model for FMDV (Fry, 1991) and enables us to scale up reflections, more than 75% recorded, to fullys without degradation of the data quality.

Prior to postrefinement the data are scaled together using ROTAVATA to place them on a common scale. A single cycle of postrefinement is optimally used (Fry, 1991), a second cycle leading to only a small improvement in agreement whilst resulting in a loss of some unique data. After postrefinement the swapped and unswapped sets are internally merged using ROTAVATA and the unique sets given single batch numbers (e.g. 1 and 2). The h and k indices of the swapped set may then be permuted and the two sets combined in a final run of ROTAVATA.

2.3.6 Final Treatment of Observations

TRUNCATE is used to treat the final observed intensities. Due to counting statistics many of the small structure factors may have negative intensities which cannot, trivially, be converted to |F| ($\approx$ 15% for FMDV). The methods of French and Wilson (1978) are used to derive the positive distribution in the 'true' intensities $\mathbf{J}$, $P_J(\mathbf{J}/\mathbf{I})$ by Bayesian inference based on our prior beliefs about these and about the form of an ideal distribution.

$$P_J(\mathbf{J}/\mathbf{I}) \overset{\mathbf{J}}{\propto} P_I(\mathbf{I}/\mathbf{J})P_J\mathbf{J}$$

$P_J(\mathbf{J})$ represents our beliefs about the ideal distribution – that it obeys Wilson statistics (Wilson, 1949). Our belief in observation $\mathbf{I}$ given $\mathbf{J}$, $P_I(\mathbf{I}/\mathbf{J})$ assumes a normal distribution of $\mathbf{I}$ about $\mathbf{J}$. $P_J(\mathbf{J}/\mathbf{I})$ is the 'posterior' distribution which is normal and truncated at 0. All $\mathbf{I}$ may therefore be reduced to $|\mathbf{F}|$ without loss of the information contained in the weakest data.

ABSCALE has been shown (Fry, 1991) to incorrectly reduce reflections to the I23 asymmetric portion. FPAD is used to fill out the asymmetric unit with equivalent

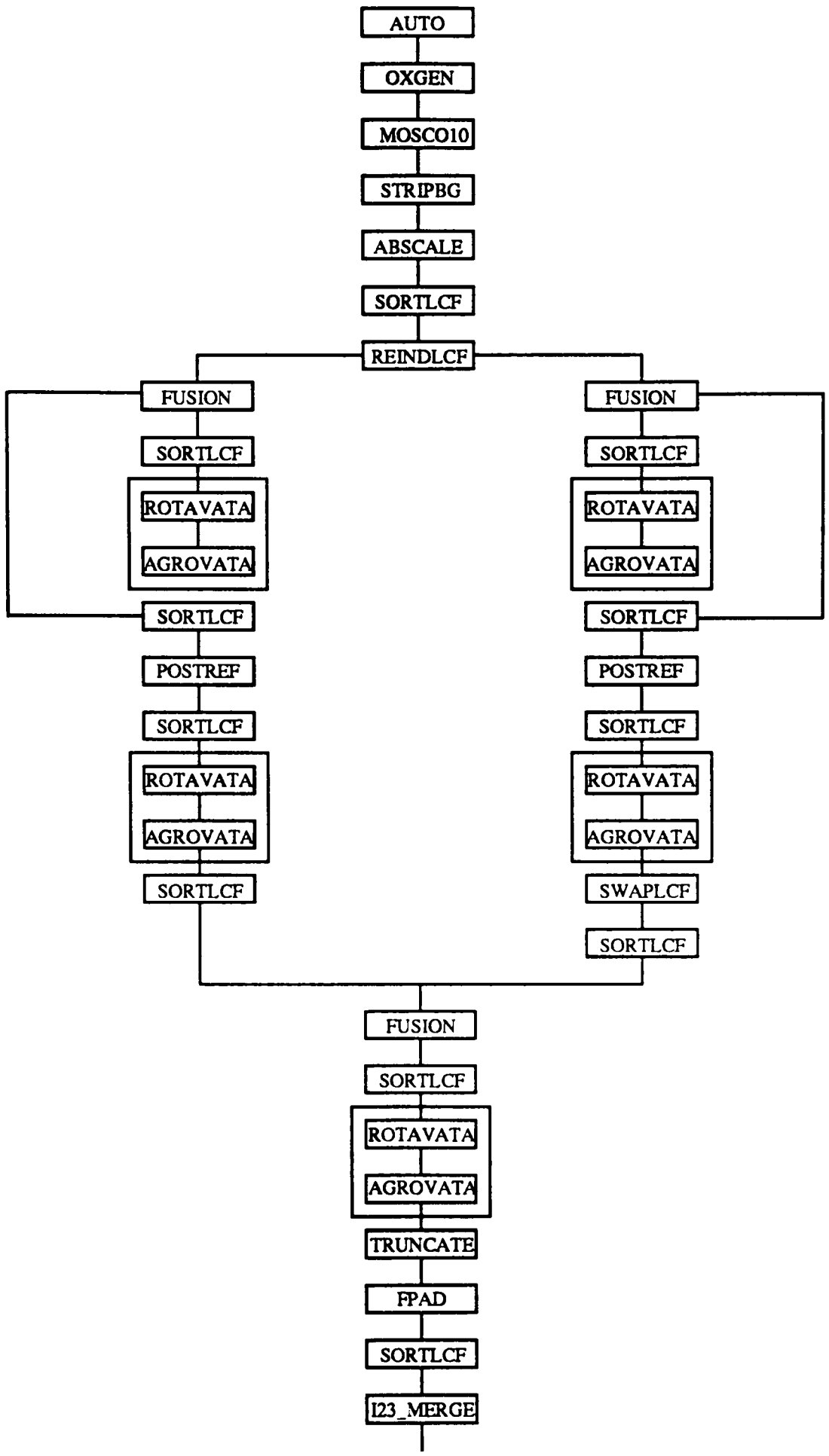


Figure 2.7: Summary of film processing.

reflections, and repeats are averaged in I23_MERGE – although very few reflections are involved in these operations (<1%).

The specific programs used in film and image processing are shown in flow diagrams in figures 2.7 and 2.8.

2.4 Model Building

Model building has mainly been performed using FRODO (Jones, 1985) on an ESV-10 RISC workstation. Certain problematical loops have been rebuilt using the structure databases within O (Jones *et al.*, 1991) and C–alpha (Esnouf, 1992).

2.5 Refinement Using X-PLOR

Refinement is the determination of shifts to the atomic parameters (coordinates and thermal parameters) to improve the agreement with the observed diffraction data. This is generally accomplished by least–squares minimisation of the differences between observed and calculated structure factors. The major problem which must be overcome is the lack of observations with which to restrain these shifts. This problem arises because of fall–off in protein diffraction at higher resolution due to structural mobility. For even a simple refinement of x,y,z and an isotropic B–factor for each atom at 3Å resolution (protein crystal with 50% solvent) there is only 3/4 of an observation per parameter being refined. With such a low ratio there is no single, unique solution to the observational equations. Two methods can be used to improve this ratio:

1. More 'observations' may be introduced in the form of restraints on the geometry of the molecule. These restraints are generally derived from knowledge of small molecule structures. This is the method used by most refinement programs (Jack and Levitt, 1978; Tronrud *et al.*, 1978; Hendrickson, 1985; Brünger, 1992).

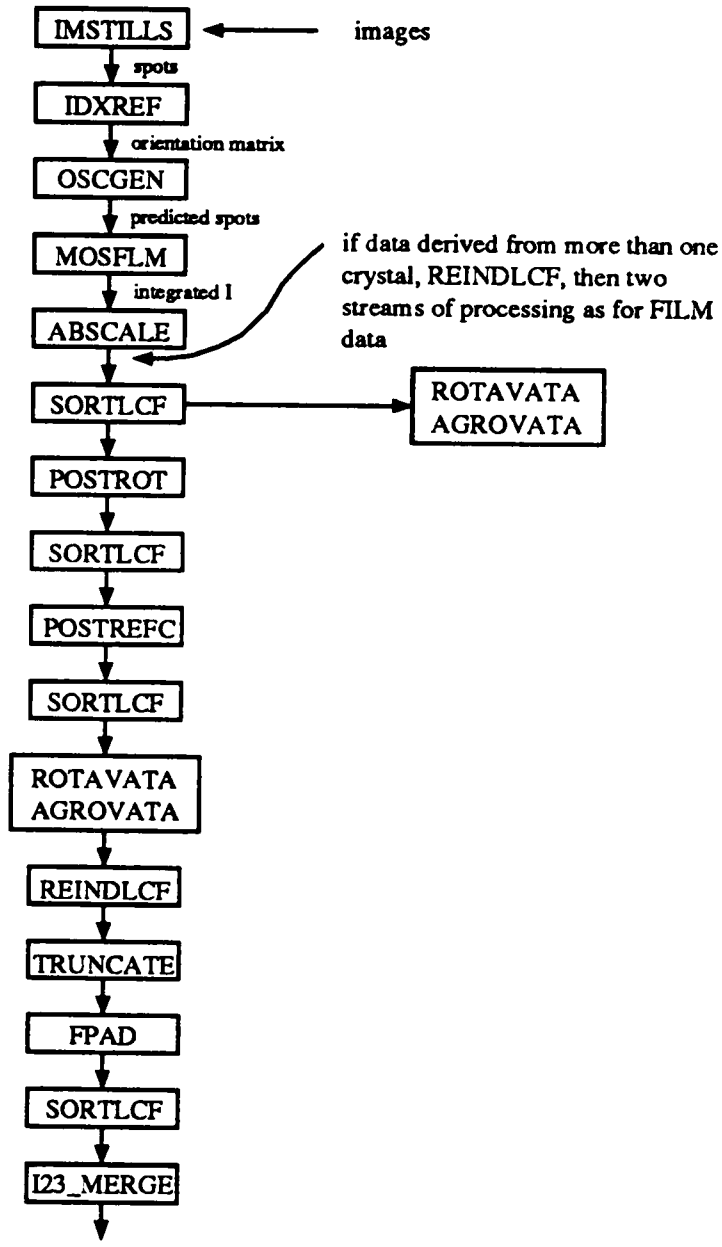


Figure 2.8: Summary of image plate processing.

2. The number of parameters to be refined is reduced by using rigid bodies / group refinement e.g. as implemented in CORELS (Sussmann *et al.*, 1977). Non–crystallographic symmetry provides a very effective constraint by refining only one of the copies in the asymmetric unit. If there are n copies in the a.u. the observation / parameter ratio will be increased n–fold. For FMDV C-S8c1 use of the 5–fold NCS to provide extra constraints gives a high ratio of 15 observations per non–hydrogen atom refined at 3.5Å.

The refinement described herein has been performed by use of simple energy minimisation by the method of conjugate gradients using X-PLOR. The parameters therefore follow the line of steepest descent of the energy surface to the nearest local minimum. Molecular dynamics protocols which would allow a greater exploration of conformational space have not been used.

2.5.1 The X-PLOR Energy Function

X-PLOR in versions 2.1, 3.0 and 3.1 (Brünger, 1992) has been used throughout this work. The energy function to be minimized consists of two classes of terms

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

$E_{EFFECTIVE}$ comprises the restraining energy terms that use experimental or other information. For simple refinement against crystallographic data with imposition of strict NCS restraints and with no phase restraints imposed

$$\begin{aligned} E_{EFFECTIVE} = E_{XREF} &= \frac{w_A}{N_A} \sum_{\mathbf{h}} w_{\mathbf{h}} [\mid F_{OBS}(\mathbf{h}) \mid - k \mid F_{CALC}(\mathbf{h}) \mid]^2 \\ N_A &= \sum_{\mathbf{h}} w_h \mid F_{OBS}(\mathbf{h}) \mid^2 \\ k &= \frac{\sum_{\mathbf{h}} w_{\mathbf{h}} \mid F_{OBS}(\mathbf{h}) \mid\mid F_{CALC}(\mathbf{h}) \mid}{\sum_{\mathbf{h}} w_{\mathbf{h}} \mid F_{CALC}(\mathbf{h}) \mid^2} \end{aligned}$$

$E_{EMPIRICAL}$ describes the energy of the molecule and may be broken down into several terms derived from bonds, angles, dihedral and improper angles and non–bonded contacts of van der Waals or electrostatic type. For bonds this takes an harmonic form

$$E_{BOND} = \sum_{BONDS} k_b(d_{model} - d_{ideal})^2$$

where k_b is the bond force constant. For angles the term has the same harmonic form. The dihedral angles energy term is

$$E_{DIHE} = \begin{cases} k_d(1 + cos(n\phi + \delta)) & \text{if } n > 0 \\ k_d(\phi - \delta)^2 & \text{if } n = 0 \end{cases}$$

where δ is the ideal value and n the periodicity of the angle. Improper angles are assessed by an equivalent term which uses very high weights (k_i) to maintain chirality of L–amino acids and planar aromatic groups. Non–covalent interactions are modelled by an electrostatic potential whilst hydrogen bonding is implicitly modelled in the partial charges and van der Waals parameters.

2.5.2 Use of a Bulk Solvent Mask

To enable use of the low resolution data (below 8–10Å) in refinement of atomic B–factors we have used the bulk solvent correction defined in X-PLOR. By Babinet's principle the solvent continuum will scatter out of phase with the protein component so that any model consisting of only protein atoms will have structure factors that are overestimated. The bulk solvent correction model sets all grid points not specified as belonging to protein as equal to a uniform electron density SOLDEN and stores the transform of this solvent mask in an array of partial structure factors FPART. The FPART are vectorially added to FCALC before calculation of R factors. A margin SOLRAD is left around the protein atoms and a scale (k_p) and B–factor (B_p) applied to the FPART array to minimise ripples caused by a sharp edge in the solvent mask.

$$FCALC = FCALC + k_p FPART e^{-B_p sin^2\theta/\lambda}$$

Optimal values of k_p, B_p and SOLRAD are determined using grid searches (Chapter 5).

2.5.3 Energy Minimisation

Simple positional and isotropic B–factor refinement have been performed using the Powell conjugate gradient minimizer implemented in X-PLOR. This is capable of finding the local energy minimum. The Powell minimizer is a variant of the steepest descent method where a displacement s_i is calculated along the negative gradient $-g_i$ at step i. Information on the previous step is included by taking a linear combination of the current gradient and previous search direction where b_k is a weighting factor dependent on the magnitudes of the two gradients.

$$\begin{aligned} s_k &= -g_k + b_k s_{k-1} \\ b_k &= \frac{|g_k|^2}{|g_{k-1}|^2} \end{aligned}$$

2.6 NCS Symmetry Averaging

All the FMDV electron density maps calculated as part of this work have been averaged over the 5 copies of the icosahedral subunit in the crystallographic asymmetric unit. Non–crystallographic symmetry (NCS) averaging provides a powerful method for both phase improvement at a given resolution and extension of phases to a higher resolution limit and, in combination with solvent flattening, has proved essential to most virus structure determinations e.g. TMV (Bloomer *et al.*, 1978), TBSV (Harrison *et al.*, 1978), SBMV (Abad–Zapatero *et al.*, 1981), STNV (Liljas *et al.*, 1982), HRV14 (Arnold *et al.*, 1987; Arnold and Rossmann, 1988), poliovirus (Hogle *et al.*, 1985; Filman *et al.*, 1989), FMDV (Acharya *et al.*, 1989; Fry *et al.*, 1993), mengovirus (Luo *et al.*, 1989) and HRV16 (Oliveira *et al.*, 1993). Initial phases may either be provided by single / multiple isomorphous replacement or molecular replacement. In theory, it should be possible to solve a high resolution virus structure starting from

a low resolution model based on simple geometric parameters, however, so far, such attempts have been relatively unsuccessful (e.g. CPV (Tsao *et al.*, 1991)).

The 'phasing power' available by use of NCS has been expressed by Arnold and Rossmann (1986)

$$P = \frac{\sqrt{nf}}{R(\frac{U}{V})}$$

where n = degree of NCS, f = fraction of data recorded, R = data quality and $\frac{U}{V}$ = ratio of protein volume to total volume of cell. However P is more likely to be proportional to n rather than $\sqrt{n}$.

In real space, if rotational ($[C_n]$) and translational (d_n) components of the n^{th} NCS operator are defined so that

$$x_n = [C_n]x_1 + d_n$$

then the electron density may be averaged over the n copies

$$\rho_{av}(x) = \frac{1}{N}\sum_{n=1}^{N}\rho(x_n)$$

The programs used in this work have all made use of such conceptually simple real–space formulations (Chapter 5) however Main and Rossmann (1986) have used this equation to derive a set of reciprocal–space equations linking structure factors

$$F_p = \frac{U}{NV}\sum_{\mathbf{h}} F_{\mathbf{h}} a_{\mathbf{hp}}$$

where the complex coefficients $a_{\mathbf{hp}}$ are derived only from the NCS elements. Significant values of $a_{\mathbf{hp}}$ occur in a thin shell surrounding h

Chapter 3

The structures of O_1K and G67

3.1 Introduction

Generation of monoclonal antibody escape mutants is the major method used to study antigenic variation of viruses *in vitro*. Many mutants have been produced and sequenced for FMDV (see Table 1.2) and for other picornaviruses (e.g. HRV (Sherry and Rueckert, 1985; Sherry *et al.*, 1986; Rossmann *et al.*, 1985) and poliovirus (Hogle and Filman, 1989; Page *et al.*, 1988)). These studies generally find that one, or only a few mutations are required to escape a single monoclonal antibody and that mutations provoked by different antibodies cluster into antigenic 'sites'. A mutation provoked by one MAb gives protection only against some of the other MAbs directed at the same site and gives no protection against MAbs directed at other sites. These clusters of mutations may occur in discontinuous regions of the primary sequence, but, are grouped into a single spatial location on the three–dimensional surface of the virion. Structural studies have shown that the surface area buried in the interface between an antigen and a binding antibody is $\approx$ 700–800Å^2 on each protein (Alzari *et al.*, 1988). This is a large area of contact, representing many interactions between different amino acids. However, the interaction is very specific and a single mutation within this huge area of complementarity is sufficient to abolish binding. To attempt to further understand the mechanism of escape we have solved the structures of two type O_1 FMDVs, very closely related to O_1BFS.

The first is a representative of the much studied O_1 Kaüfbeuren strain. O_1K was originally isolated from cattle during an outbreak in Southern Germany in 1966 and subsequently became the German vaccine strain. Whereas O_1BFS was isolated in

England in 1967 during a widespread outbreak of FMD that originated in infected meat imported from Argentina, it became the English vaccine strain. Structural differences between these two viruses should therefore represent differences that have arisen 'in–the–field' however the precise history of all viruses is complicated due to the high rate at which random mutations occur. This means that the differing 'passage histories'[1] of the two viruses will also have led to differences and it is impossible to state that a particular difference seen between *our* O_1BFS and O_1K structures represents the effects of field imposed pressures rather than random mutations - although both are considered to be relatively 'close–to–field' for laboratory strains (personal communication Andrew King). Both viruses are considered to be epidemiologically of a South American type and are clearly very closely related. The sequences of the two viruses are known, and they differ by only 5 or 6 amino acids (the O_1K population is heterogenous at one position (Kitson *et al.*, 1990)) in the 4 capsid proteins.

The second virus, G67, is derived from O_1K by sequential exposure to MAbs directed at the four antigenic sites defined in type O_1 FMDVs (Xie *et al.*, 1987; McCahon *et al.*, 1989; Kitson *et al.*, 1990), and sequencing shows it to contain escape mutations in each of these areas (Figure 3.1). Despite these mutations G67 is still recognised as a subtype O_1 virus by polyclonal sera (as are O_1K and O_1BFS). Structures of these viruses have been determined in both native and dithiothreitol (DTT) soaked forms to enable comparisons of the conformation(s) adopted by the VP1 G–H loop (Logan *et al.*, 1993) (Chapter 1).

3.2 Data Collection and Processing

Virus seed stocks were kindly provided by Dr. G. Belsham[2] and were grown and purified using the protocols established for O_1BFS (Fox *et al.*, 1987; Curry *et al.*, 1992), Crystals of both viruses were grown by R. Abu–Ghazaleh by microdialysis

[1]The passage history of a virus is the sequence of hosts in which the virus has replicated since isolation from the field.

[2]I.A.H. Pirbright

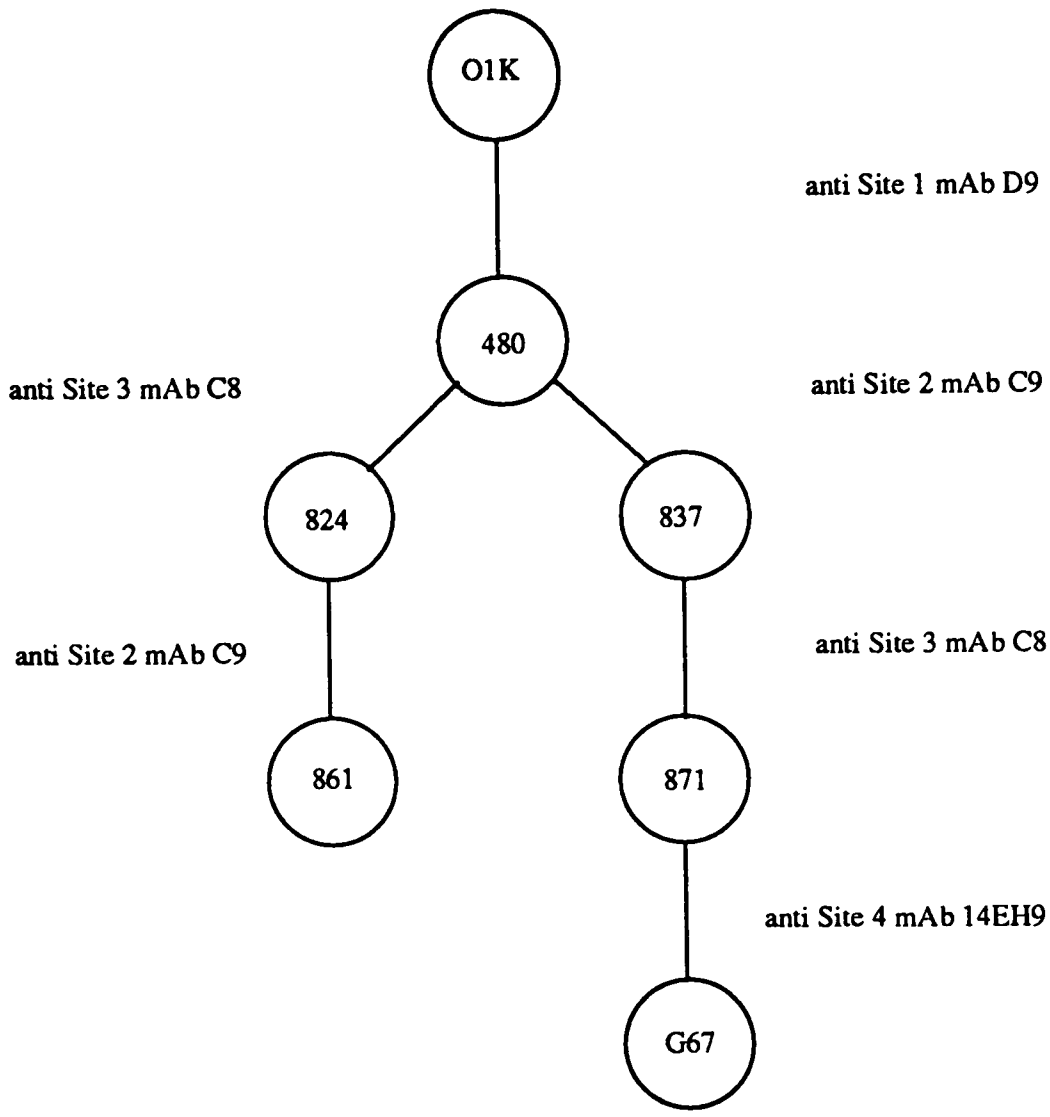


Figure 3.1: 'Family tree' showing MAbs used in production of G67.

using similar conditions to those used to crystallize O_1BFS (Fox *et al.*, 1987; Curry *et al.*, 1992), although a higher concentration of $(NH_4)_2SO_4$ was needed to produce crystals of G67 than that required to yield useful crystals of O_1K (21–23% c.f. 11–12%). In order to visualize the VP1 G–H loop both native and 'reduced' data sets were collected for the two viruses. The reduction procedure used was as described in Logan *et al.* 1993, i.e. the crystals were transferred, before mounting, to a well containing crystallization mother liquor and 10mM DTT. After soaking for two hours the crystals were quickly mounted in sealed quartz capillary tubes. Data were collected during many trips to the SRS Daresbury over a period of a year (Nov 1990–1991), the majority of the data being collected using Stations 7.2 and 9.6 with film as the detection media, although the 'final' (see below) reduced O_1K data were collected using the 90mm MAR Research imaging plate, installed on Station 9.5, in November 1991. The crystals varied in quality between different data collection trips – this was largely the cause of the extended time period needed to collect the data, several trips yielded only one or two useful film packs despite the large number of crystals taken (often more than 20 of any one virus).

The crystals of both viruses are rhombic dodecahedra and initial examination of the diffraction patterns suggested that both had crystallised isomorphously with O_1BFS i.e. space group I23, a = b = c = 345.0Å, $\alpha = \beta = \gamma = 90°$. Both O_1K data sets were successfully processed using these parameters and our standard procedures (Chapter 2 and below) however a mis–packing of the G67 crystal lattice led to an apparent increase in the symmetry of the data (discussed in detail in Chapter 4). Both G67 data sets were therefore processed in PG432 (see below).

3.2.1 Native O_1K

These data were collected during November / December 1990 on station 7.2 of the SRS Daresbury, using radiation with $\lambda = 1.488$Å (Table 3.6). The lifetime of FMDV crystals in the beam is significantly shorter at this longer wavelength (c.f. $\lambda = 0.89$Å), but

	Native O_1K	Native G67
λ	1.488Å	
horizontal divergence	0.200°	
vertical divergence	0.013°	
spectral dispersion	0.001°	
oscillation range	0.5°	
no. of oscillations per exposure	10–15	
time per oscillation	100–150s	
Crystal to film distance	88mm	100mm
resolution at edge of film	2.6Å	3.0Å
T (° C)	15–19	
no. of crystals	35	24
no. of useful film packs	27	18

Table 3.1: Data collection: Native O_1K and G67 - Station 7.2

the relatively large size (0.3mm in all dimensions) and good quality of these crystals enabled collection of 2 film packs from several crystals (although most crystals yielded only one useful film pack). Data were processed using our standard procedures (see Chapter 2) with REINDLCF separating the 27 useful packs into 17 unswapped and 10 swapped, by correlation with a standard, native, O_1BFS set. These streams were separately merged and postrefined (Table 3.2) until the hand of the swapped set was reversed and the two streams merged to give a data set 67% complete to 2.6Å.

3.2.2 Reduced O_1K

Twenty–five useful packs of data were collected from DTT soaked O_1K crystals in March 1991 (station 9.6, $\lambda = 0.89$Å), however, after processing the data, no evidence could be seen for an ordered structure for the G–H loop of VP1. Checks of the procedure used to reduce these crystals showed that a lower concentration of DTT had been used than that used previously to successfully reduce the VP1–VP2 disulphide, therefore, further data were collected as soon as new crystals of O_1K were produced and soaked. Three small crystals of O_1K were soaked and then mounted for a trip

Native O_1K	Unswapped	Swapped
AUTO SDXY SDPHI	 1.86±0.23 0.12±0.01	 1.68±0.24 0.11±0.00
MOSCO $I/\sigma(I)$ A film	3.6±1.3	3.1±1.1
ABSCALE R_{prof}(I) (%) R_{int}(I) (%)	 6.9±1.6 8.1±1.9	 7.6±1.8 9.3±2.2
$< I >$ before PR	210.87	478.04
R(I) (%) before PR	13.1	14.9
POSTREF $\Delta\phi y$ $\Delta\phi z$ $\Delta\lambda$ (Å $\times 10^{-5}$) ϵ F→P P→F total F	 -0.00±0.00 0.01±0.02 -99±83 0.043±0.062 4713 23489 63093	 0.00±0.01 0.00±0.01 -133±60 0.003±0.021 13135 42633 110983
Before Massagelcf R(I) (%) no. of Ind. Ref.	 11.0 83954	 11.7 106323
After Massagelcf R(I) (%) no. of Ind. Ref.	 12.6 53802	 13.2 67948
Combined Resolution Limit R(I) (%) Total No. Observed Ref. No. of Ind. Ref. % Complete	 2.6Å 11.3 237674 138727 67%	

SDXY - standard deviation in spot position (pixels); SDPHI - standard deviation in spindle angle (°); PR - postrefinement; ϵ - beam divergence≡mosaiscity of crystal; F - fully recorded reflections; P - partially recorded reflections; $R(I)=\frac{\sum_h \sum_i |(I_h - I_{hi})|}{\sum_h \sum_i I_{hi}} \times 100$

Table 3.2: Data processing: O_1K Native

	Reduced O_1K
λ	0.895Å
horizontal beam divergence	0.05°
vertical beam divergence	0.0005°
spectral dispersion	0.0015°
oscillation range	0.5°
no. of oscillations per exposure	7–10
time per oscillation	120s
current	202–253mA
ion chamber flux	1.45–1.88
Crystal to plate distance	325mm
resolution at edge of plate	3.3Å
T (°C)	17–21
no. of crystals	3
no. of useful images	15

Table 3.3: Image plate data collection: O_1K DTT soaked - Station 9.5

to the SRS Daresbury in November 1991. Station 9.5 was used where the increased sensitivity of the recently installed 90mm MAR Research imaging plate detector to short wavelengths (relative to film) enabled use of a total exposure time equivalent to that used on station 9.6 even though there is a 10–fold reduction in flux between the two stations. This enabled collection of 12 images from one of the crystals despite the small size (less than 0.2mm in all dimensions) of the crystals used. Data collection parameters are shown in Table 3.3.

Crystal 1 yielded no useful diffraction. Initial attempts to index images using IMSTILLS and REFIX (MOSFLM package version 4.22, A.Leslie) yielded a satisfactory solution for the 12 images from crystal 3, but no solution for the 3 images from crystal 2. Processing of the crystal 3 data therefore proceeded using MOSFLM (Chapter 2). Eventually an indexing solution was found for crystal 2 (by adjustments to the S/N ratio) and these 3 images were processed as a separate batch. Both crystals correlated better with the reference O_1BFS data in the swapped hand (REINDLCF), therefore,

following postrefinement (POSTREFC) and scaling up of all reflections greater than 75% recorded (MASSAGELCF), the hand of the two batches was swapped and the data merged to give a data set 58% complete to 3.3Å. Table 3.4 shows the quality of these data.

3.2.3 Native G67

As for the native O_1K data these data were collected on film using station 7.2 (Table 3.1), however, these crystals were of relatively poor quality and of a small size (0.15mm maximum in all dimensions), no crystal yielded more than one useful film pack. As discussed in Chapter 4, REINDLCF was unable to separate the G67 crystals into swapped and unswapped streams by comparison with O_1BFS or O_1K , the correlation in either hand being less than 0.5. This suggested extra 4–fold symmetry, and the data were successfully processed in PG432. After postrefinement (Table 3.5) and scaling up of reflections more than 75% recorded, the 18 films packs yielded a data set 76% complete to 3.0Å in PG432.

3.2.4 Reduced G67

The crystals used to collect these data were amongst the largest FMDV crystals ever produced (in excess of 0.5mm in all dimensions). Data were collected on Station 9.6 (Table 3.6) during June and July of 1991, using film as the detection media. Although 26 crystals were examined the useful film packs were produced by 12 of the largest crystals, several crystals producing 4 useful film packs each. As for the native data REINDLCF could not divide the data into swapped and unswapped streams and the data were processed in PG432 (Chapter 4) to give a data set 92% complete to 3.0Å in this point group (Table 3.7).

	Crystal 2	Crystal 3
REFIX		
SDXY	0.80	0.57
SDPHI	0.11	0.11
MOSFLM		
$I/\sigma(I)$	3.5±0.2	4.1±0.4
FULLYS		
ABSCALE		
$R_{symm}(I)$	25.5±4.3	20.8±3.9
REINDLCF		
UN	0.06	0.00
SW	**0.77**	**0.69**
$< I >$ before PR	9.69	12.77
R(I) (%)		
before PR	23.0	20.1
POSTREF		
$\Delta\phi y$	0.00±0.00	0.00±0.00
$\Delta\phi z$	0.00±0.00	0.00±0.00
$\Delta\lambda$ (Å $\times 10^5$)	-69	252
ϵ	0.029±0.010	0.028±0.023
F→P	67	349
P→F	2449	10128
total F	14421	57447
Before Massagelcf		
R(I) (%)	23.8	19.8
no. of Ind. Ref.	12486	46527
After Massagelcf		
R(I) (%)	24.3	19.9
no. of Ind. Ref.	14051	51452
Combined		
Resolution Limit	3.3Å	
R(I) (%)	21.0	
Total No. Observed Ref.	82112	
No. of Ind. Ref.	58738	
% Complete	58%	

SDXY - standard deviation in spot position (pixels); SDPHI - standard deviation in spindle position (°); PR - postrefinement; ϵ - beam divergence≡mosaiscity of crystal; F - fully recorded reflections; P - partially recorded reflections; $R(I)=\frac{\sum_h \sum_i |(I_h - I_{hi})|}{\sum_h \sum_i I_{hi}} \times 100$

Table 3.4: Data processing: O_1K DTT soaked

	Native G67
AUTO	
SDXY	1.63±0.23
SDPHI	0.11±0.00
MOSCO	
$I/\sigma(I)$	3.7±0.6
A film	
ABSCALE	
$R_{prof}(I)$ (%)	8.8±1.2
$R_{int}(I)$ (%)	11.0±1.7
R(I) (%)	
before PR	19.8
POSTREF	
$\Delta\phi y$	-0.00±0.01
$\Delta\phi z$	0.01±0.02
$\Delta\lambda$ (Å $\times 10^5$)	8±360
ϵ	0.135±0.091
F→P	15437
P→F	21920
total F	78123
Before Massagelcf	
R(I) (%)	18.4
no. of Ind. Ref.	40332
After Massagelcf	
Resolution Limit	3.0Å
R(I) (%)	19.2
Total No. Observed Ref.	101615
No. of Ind. Ref.	52219
% Complete	76%

SDXY - standard deviation in spot position (pixels); SDPHI - standard deviation in spindle angle (°); PR - postrefinement; ϵ - beam divergence≡mosaiscity of crystal; F - fully recorded reflections; P - partially recorded reflections; $R(I)=\frac{\sum_h \sum_i |(I_h - I_{hi})|}{\sum_h \sum_i I_{hi}} \times 100$

Table 3.5: Data processing: G67 Native

	Reduced G67
λ	0.89Å
horizontal divergence	0.200°
vertical divergence	0.013°
spectral dispersion	0.001°
oscillation range	0.5°
no. of oscillations per exposure	5–7
time per oscillation	100–120s
Crystal to film distance	190mm
resolution at edge of film	3.0Å
T (°C)	15–21
no. of crystals	26
no. of useful film packs	27

Table 3.6: Data collection: G67 DTT soaked - Station 9.6

3.3 Calculation of Difference Maps

Since O_1K crystallised isormorphously with O_1BFS we felt the easiest way to visualize the differences between the two viruses would be by calculation of difference maps. The difference Fourier is a very sensitive technique allowing shifts of atomic position of less than 1Å to be detected (Blundell and Johnson, 1976). Reflections are matched and scaled in resolution shells (SHELL_SCALE) to observed amplitudes for the other O_1 viruses, with R_{match}, C_{match} calculated (shown in Table 3.8 as R and C) as:

$$R_{match} = \frac{\sum_h \mid\mid F_{virus1} \mid_h - \mid F_{virus2} \mid_h \mid}{\sum_h \mid F_{virus1} \mid}$$

$$C_{match} = \frac{\sum_h (< F_{virus1} > - \mid F_{virus1} \mid_h)(< F_{virus2} > - \mid F_{virus2} \mid_h)}{[\sum_h (< F_{virus1} > - \mid F_{virus1} \mid)^2 \cdot \sum_h (< F_{virus2} > - \mid F_{virus2} \mid)^2]^{\frac{1}{2}}}$$

These scaled reflections are then matched with the 'best' O_1BFS, native, phases - produced after 719 cycles of NCS averaging (Acharya *et al.*, 1989; Fry, 1991; Fry *et al.*, 1993). Maps are calculated with coefficients:

$$\mid F_{virus1} \mid - \mid F_{virus2} \mid, \alpha_{O_1BFS}$$

	Reduced G67
AUTO SDXY SDPHI	 1.75±0.19 0.11±0.01
MOSCO $I/\sigma(I)$ A film	 6.1±2.7
ABSCALE R_{prof}(I) (%) R_{int}(I) (%)	 6.4±1.4 7.4±1.5
$< I >$ before PR	37.0
R(I) (%) before PR	 11.0
POSTREF $\Delta\phi y$ $\Delta\phi z$ $\Delta\lambda$ (Å $\times 10^5$) ϵ F→P P→F total F	 0.01±0.02 0.03±0.08 -85±256 0.069±0.185 8407 46573 128920
Before Massagelcf R(I) (%) no. of Ind. Ref.	 8.6 58072
After Massagelcf Resolution Limits R(I) (%) Total No. Observed Ref. No. of Ind. Ref. % Complete	 3.0Å 11.3 165285 64004 92%

SDXY - standard deviation in spot position (pixels); SDPHI - standard deviation in spindle angle (°); PR - postrefinement; ϵ - beam divergence≡mosaiscity of crystal; F - fully recorded reflections; P - partially recorded reflections; $R(I)=\frac{\sum_h \sum_i |(I_h - I_{hi})|}{\sum_h \sum_i I_{hi}} \times 100$

Table 3.7: Data processing: G67 DTT soaked

		O_1BFS		O_1K		G67◇	
		Native	Reduced	Native	Reduced	Native	Reduced
O_1BFS	Native	R(I)=14.1*	R=15.41 C=0.9274	R=13.97 C=0.9401	R=19.72 C=0.8823	R=15.00 C=0.8756	R=13.48 C=0.9029
	Reduced		R(I)=16.0†	R=17.18 C=0.9123	R=17.84 C=0.8995	R=15.80 C=0.8656	R=10.87 C=0.9363
O_1K	Native			R(I)=11.3	R=19.47 C=0.8832	R=13.00 C=0.9006	R=11.77 C=0.9248
	Reduced				R(I)=21.0	R=25.57 C=0.6809	R=13.09 C=0.7037
G67	Native					R(I)=19.2	R=14.64 C=0.8846
	Reduced						R(I)=11.3

◇ - All comparisons with G67 are made in PG432 between G67 data processed in this point group and I23 data sets folded into this point group (see text).
* - E. Fry, 1991.
† - D. Logan, 1991.

Table 3.8: Scaling statistics between O_1 virus data sets

at 3.5 and 4.0Å. These are then averaged over the 5 NCS related icosahedral subunits within the crystallographic asymmetric unit. Where comparisons were made between two relatively incomplete data sets (e.g. reduced O_1BFS and reduced O_1K) the incompleteness of the matched set of reflections led to noisier difference maps - although all the maps were easily interpretable in terms of the known sequence changes.

Calculation of difference maps using the G67 data was complicated by the mis–packing of the crystals. The data were processed in PG432 and could therefore not be simply matched to the other O_1 data sets and the available phases (which are all in I23). As discussed in Chapter 4 we found that a protocol involving 'folding' of an I23 data set into the unique portion for 432 and calculation of pseudo–432 intensities as the arithmetic mean of the I23 intensities for reflection pairs h,k,l and k,h,l allowed us to successfully mimic the G67 data. Comparisons between the other O_1 viruses and G67 were therefore made by producing 'folded' data sets for each virus which were scaled (as before with calculation of R_{match} and C_{match}) to the G67 data in 432 (Table 3.8). These scaled data are than 'unfolded' into I23 by duplication of reflection records and matched with the O_1BFS, native, phases to calculate difference maps (as for O_1K) in I23. After 5–fold averaging these maps were easily interpreted.

3.4 Results

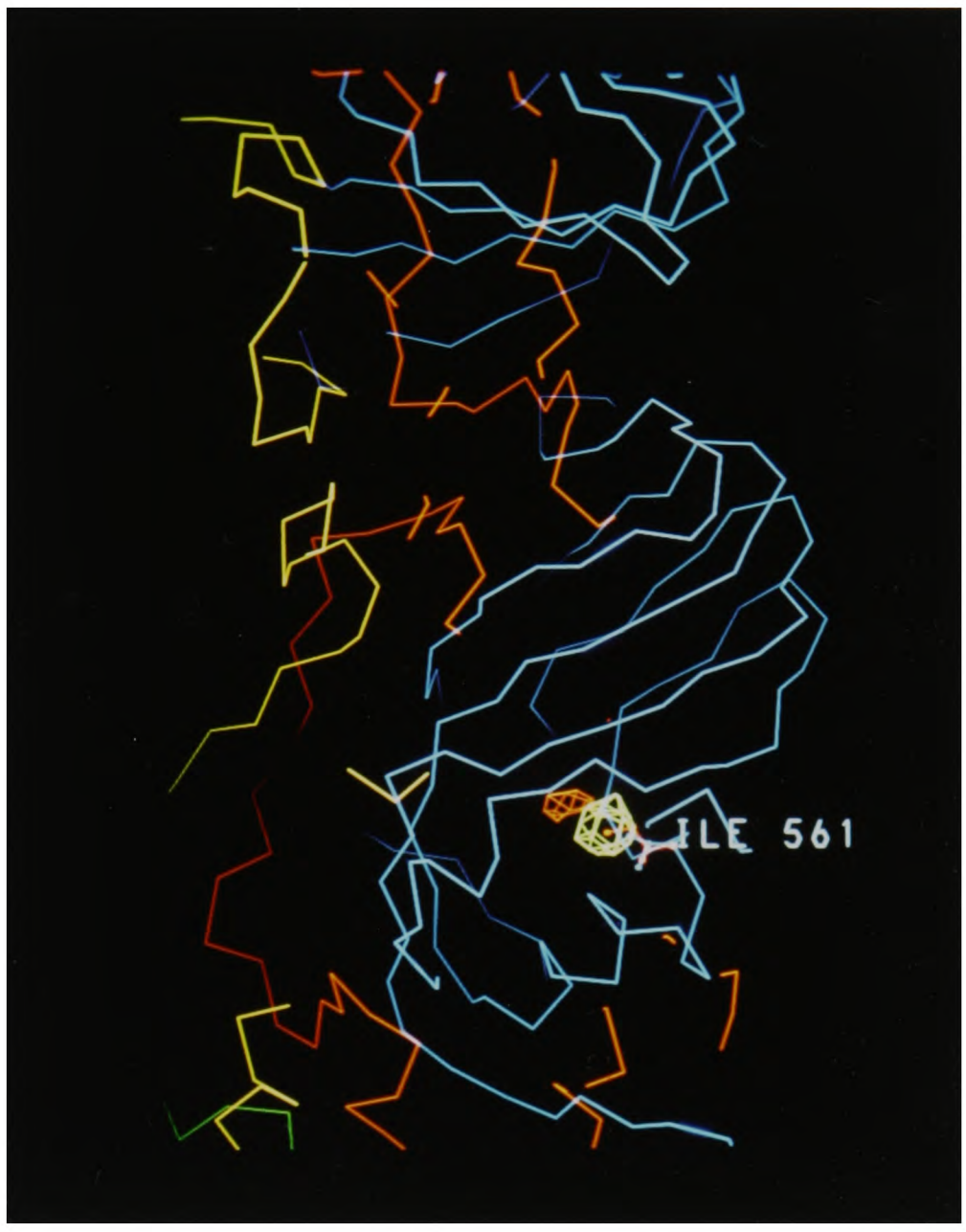
3.4.1 O_1K

The sequence of O_1K as reported by Kitson *et al.* (1990) differs from that of O_1BFS by 5 or 6 amino acid changes (Table 3.9). The ambiguity arises because the parental O_1K population was found to be mixed, some clones having Glu at position 133 of VP1 (as in O_1BFS) and some having a Gly at this position. Difference maps calculated between O_1K and O_1BFS showed clear areas of difference density corresponding to the four sequence changes outside the disordered 'FMDV' loop region (Figure 3.2). The density clearly showed that although the side chains were altered there was no

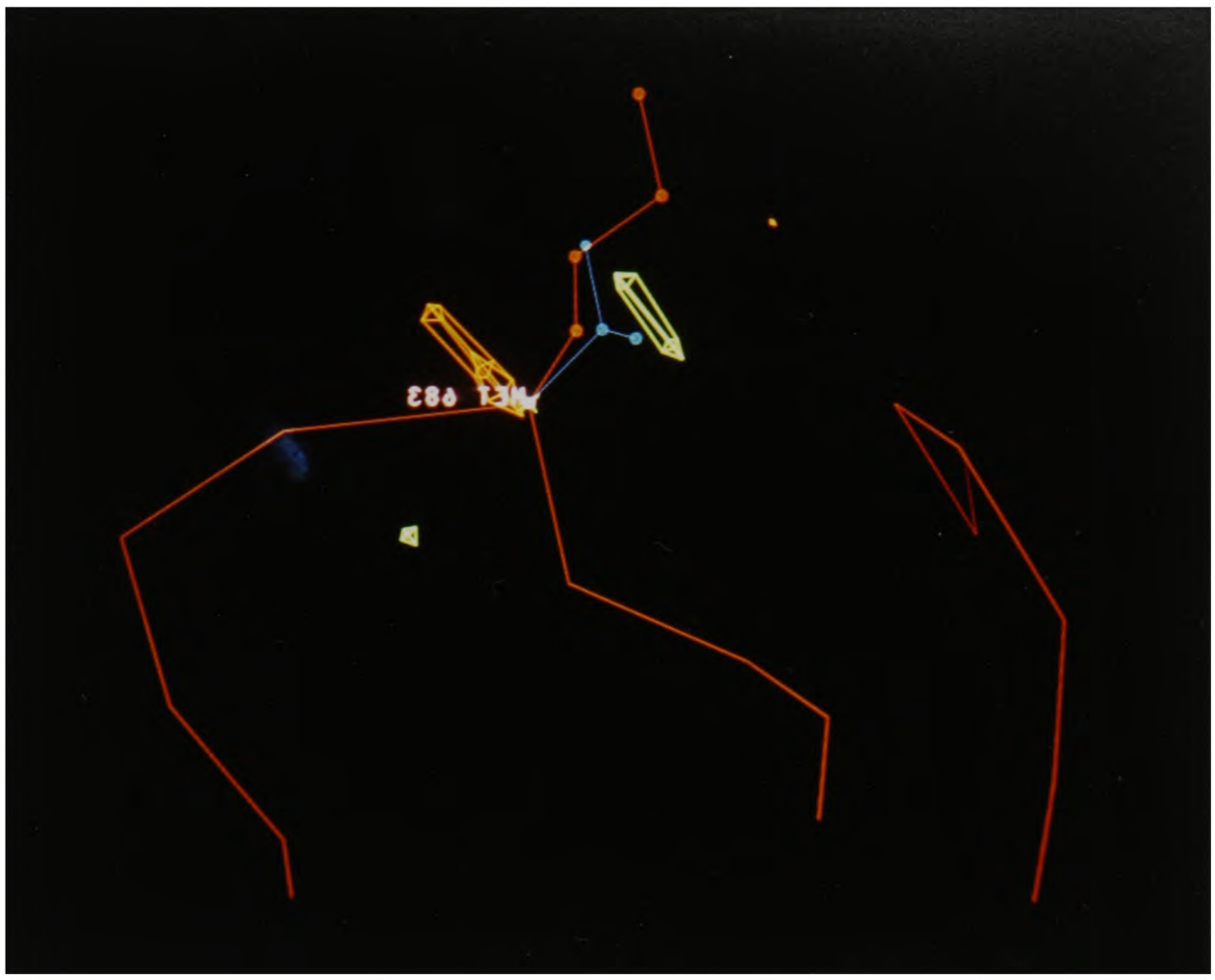
	O_1BFS	O_1K	H_2O Accessibility in O_1BFS	
			Å^2	%*see below*
VP1				
56	Val	Ile	1.0	0
133	Glu	Gly	49.1	27
137	Ser	Asn	77.1	62
VP2				
NONE				
VP3				
56	Arg	His	33.3	13
60	Gly	Asp	14.4	16
68	Thr	Met	89.9	60

$$\% = \frac{Accessibility\ in\ O_1BFS}{Accessibility\ in\ G-X-G\ extended\ tripeptide}$$

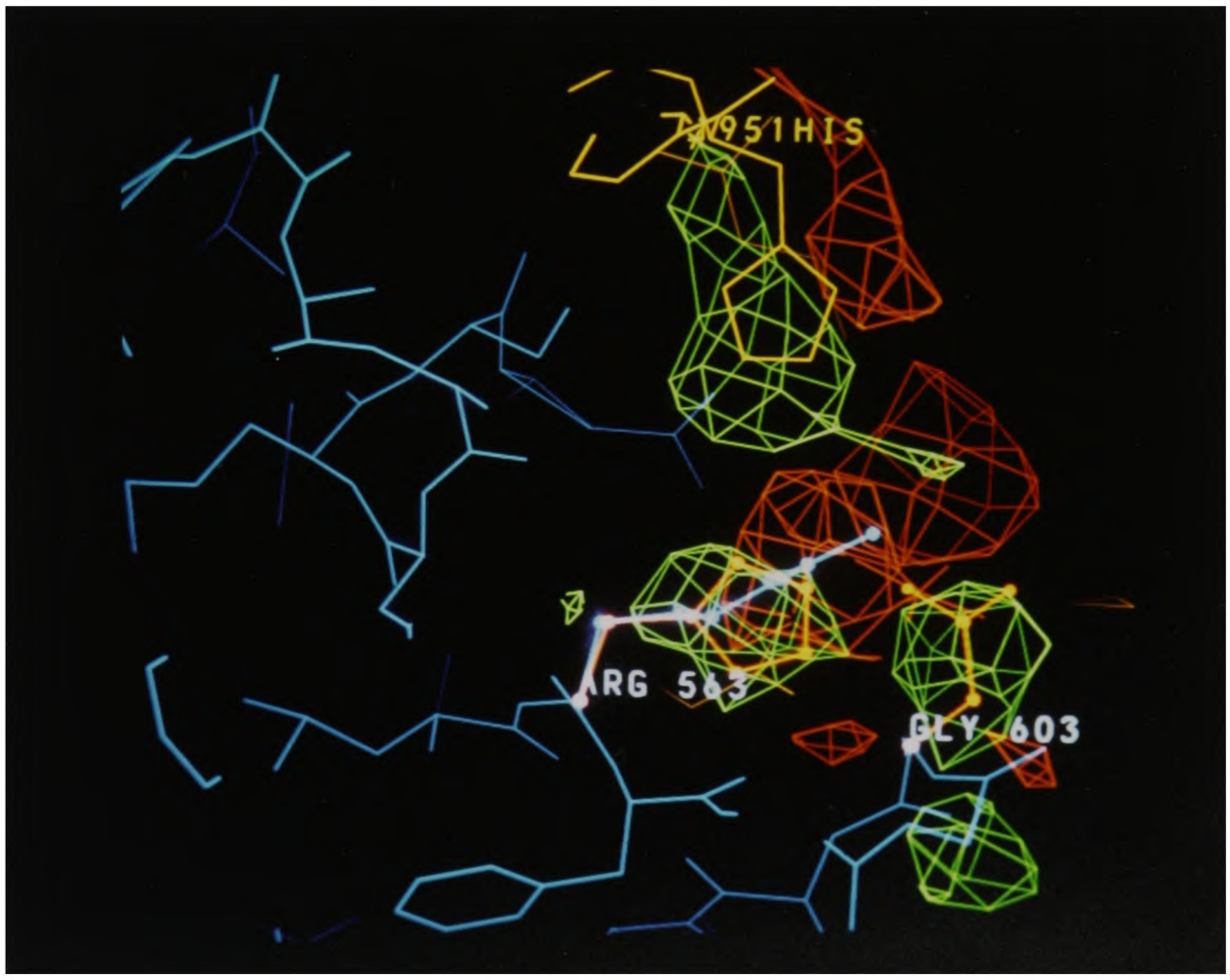
Table 3.9: Sequence differences seen between O_1K and O_1BFS



(A) Val→Ile, 56 of VP1



(B) Thr→Met, 68 of VP3



(C) Arg→His, 56 of VP3 and Gly→Asp, 60 of VP3

Contoured at an arbitrary level.
+ve electron density – green; -ve electron density – red.

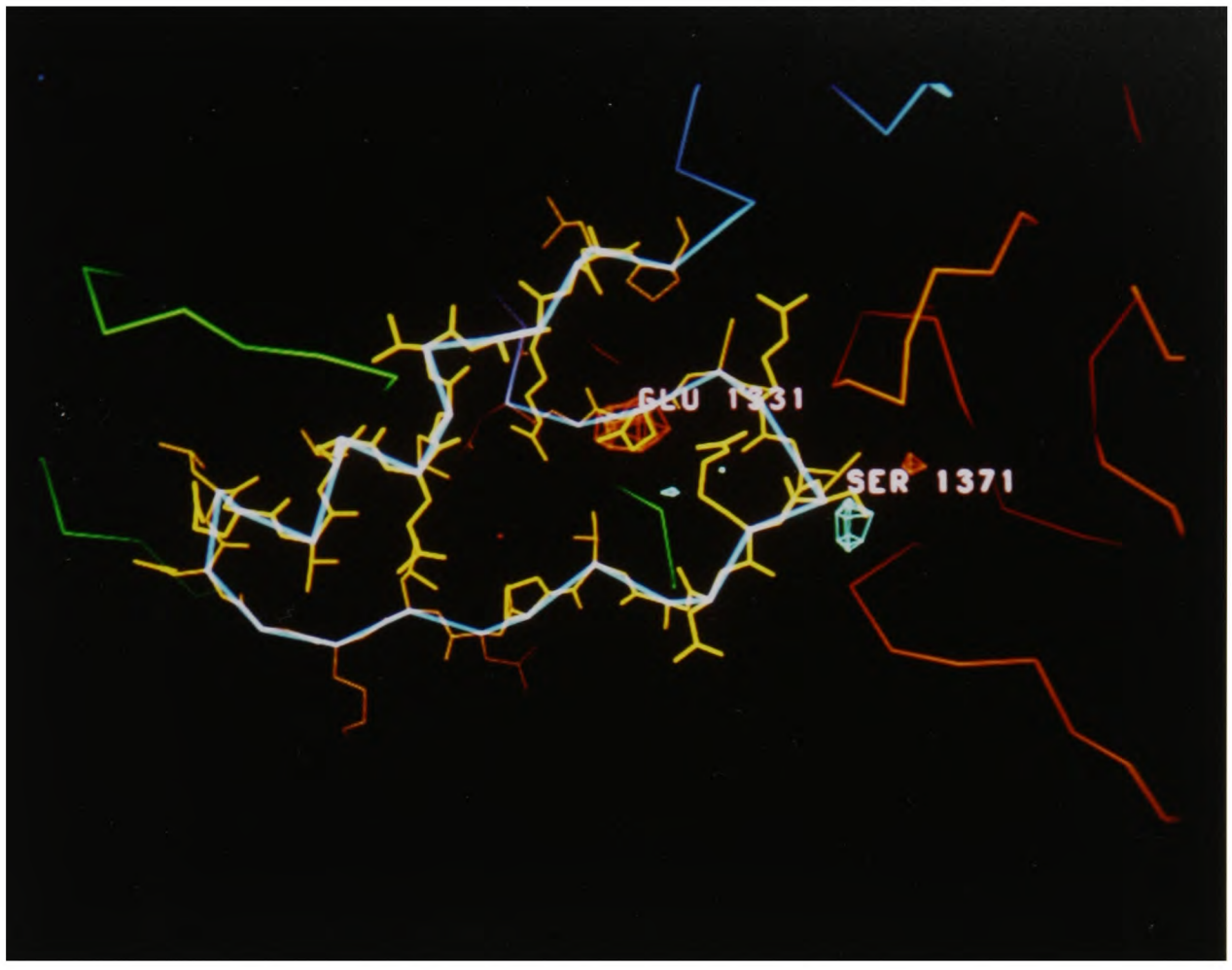
Figure 3.2: Sequence differences seen in $\mid F_{O_1K,native} \mid - \mid F_{O_1BFS,native} \mid$ difference maps.

movement of main chain atoms. The change at residue 56 of VP1 from a Val to an Ile is marked by a small peak of positive density associated with the increase in side chain length (Figure 3.2A). This change is unlikely to be important in determining antigenic differences between O_1K and O_1BFS since the residue is buried in an hydrophobic pocket in the capsid and is therefore inaccessible to antibody attack (H_2O accessibilities[3] for all the altered amino acid positions are shown in Table 3.9). A similar positive peak marks the change from Thr to Met at 68 of VP3 (Figure 3.2B) although this peak is perhaps smaller than might be expected for the gain of a sulphur; probably explained by the great mobility of this residue as shown by crystallographic B factors (The Thr in O_1BFS; B-factors for all side chain atoms are in excess of $50Å^2$).

These are relatively small features in the difference maps. The major feature is an area of difference density around the mutations at 56 and 60 of VP3 (Figure 3.2C). These changes lie close together on the viral surface. The substitution of an His for an Arg is associated with an area of negative density (shortening of the chain) and an area of positive density (accommodating the ring). The substitution at 60 of an arginine for a glycine is similarly clearly marked with a large area of positive density associated with the increase in side chain length. There is, in addition, other difference density associated with this region which implies loss of the sulphate ion bound to Arg 56 in O_1BFS (Fry *et al.*, 1990) and loss of two H_2O molecules from the area. The side chain of His 195 of the C-terminus of VP1 (which lies over this region) also moves in response to the mutations in VP3. Based on this large structural difference we predicted that a MAb directed at this region would be able to distinguish between the two viruses. At the time of the structure determination antigenic differences between the two crystal strains had not been studied. Assays using the anti–site 4 MAb 14EH9 showed that it was indeed able to distinguish between the viruses, binding to O_1K but not to O_1BFS.

The difference map calculated between reduced O_1K and reduced O_1BFS shows that the FMDV-loop has adopted the same conformation in O_1K as in O_1BFS on

[3]Calculated by the method of Lee and Richards (1971).



Map contoured at an arbitrary level.
+ve difference density – green; -ve difference density – red.

Figure 3.3: Sequence differences seen in 'FMDV–loop' region of $\mid F_{O_1K,reduced} \mid - \mid F_{O_1BFS,reduced} \mid$ difference map.

reduction of the disulphide linking 130 of VP1 and 134 of VP2. The two areas of difference density associated with the loop residues show that the virus from which the crystals had been grown contains the Glu to Gly mutation at 133 of VP1 in addition to the serine to asparagine mutation at residue 137 (Figure 3.3). The difference density for both these changes shows the differences to be clearly limited to the changes in the side chain with no movements of the main chain atoms compared to their positions in reduced O_1BFS.

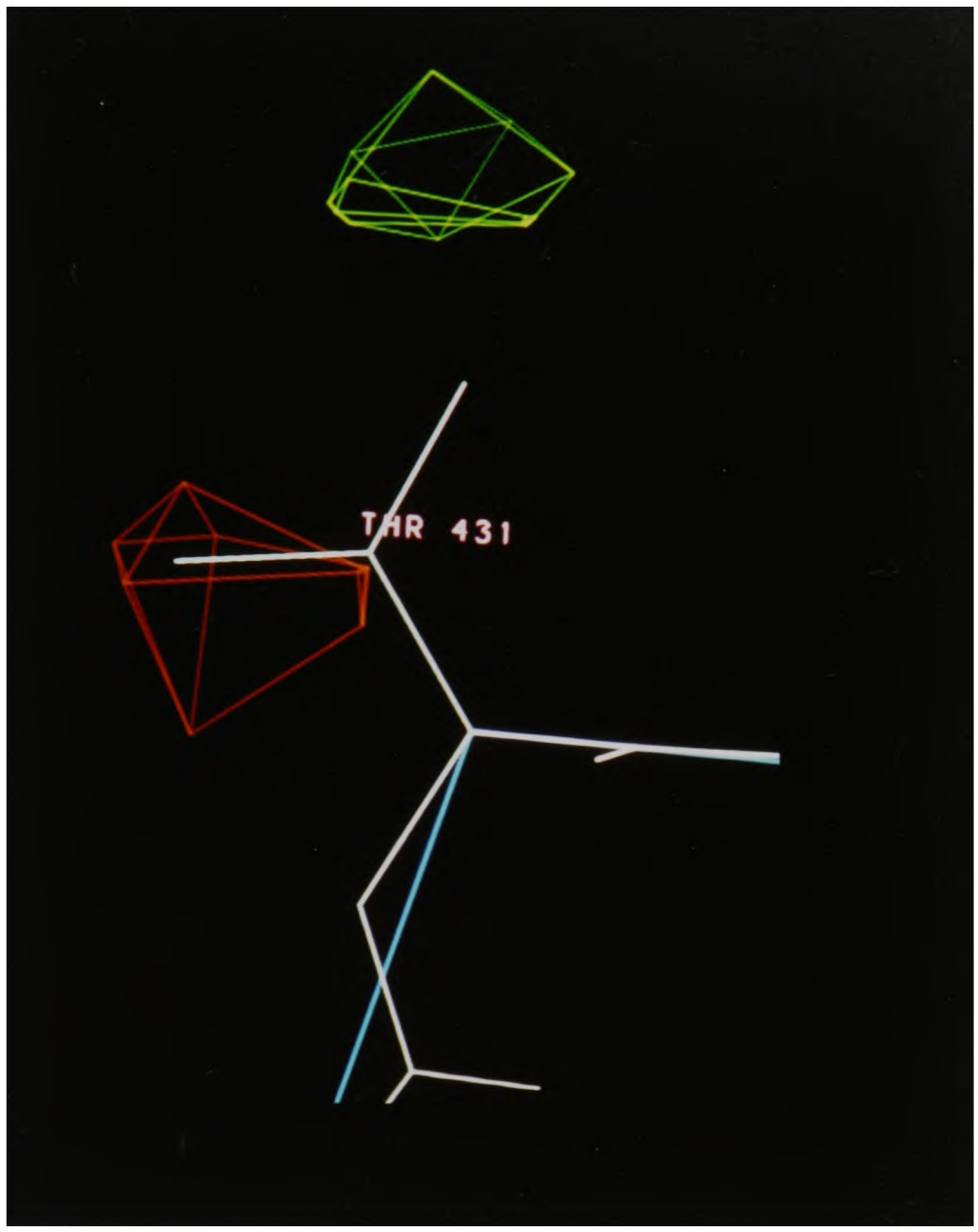
3.4.2 G67

G67 differs from the parental O_1K by a further six mutations (Table 3.10), three of

	O_1K	G67	H_2O Accessibility in O_1K	
			$Å^2$	% *see below*
VP1				
43	Thr	Lys	117.3	78
133	Gly	Glu	27.6	31
138	Arg	Lys	192.2	76
148	Leu	Arg	44.8	25
150	Val	Ala	30.1	19
VP2				
72	Ser	Asn	92.2	77
VP3				
58	Glu	Val	169.6	91

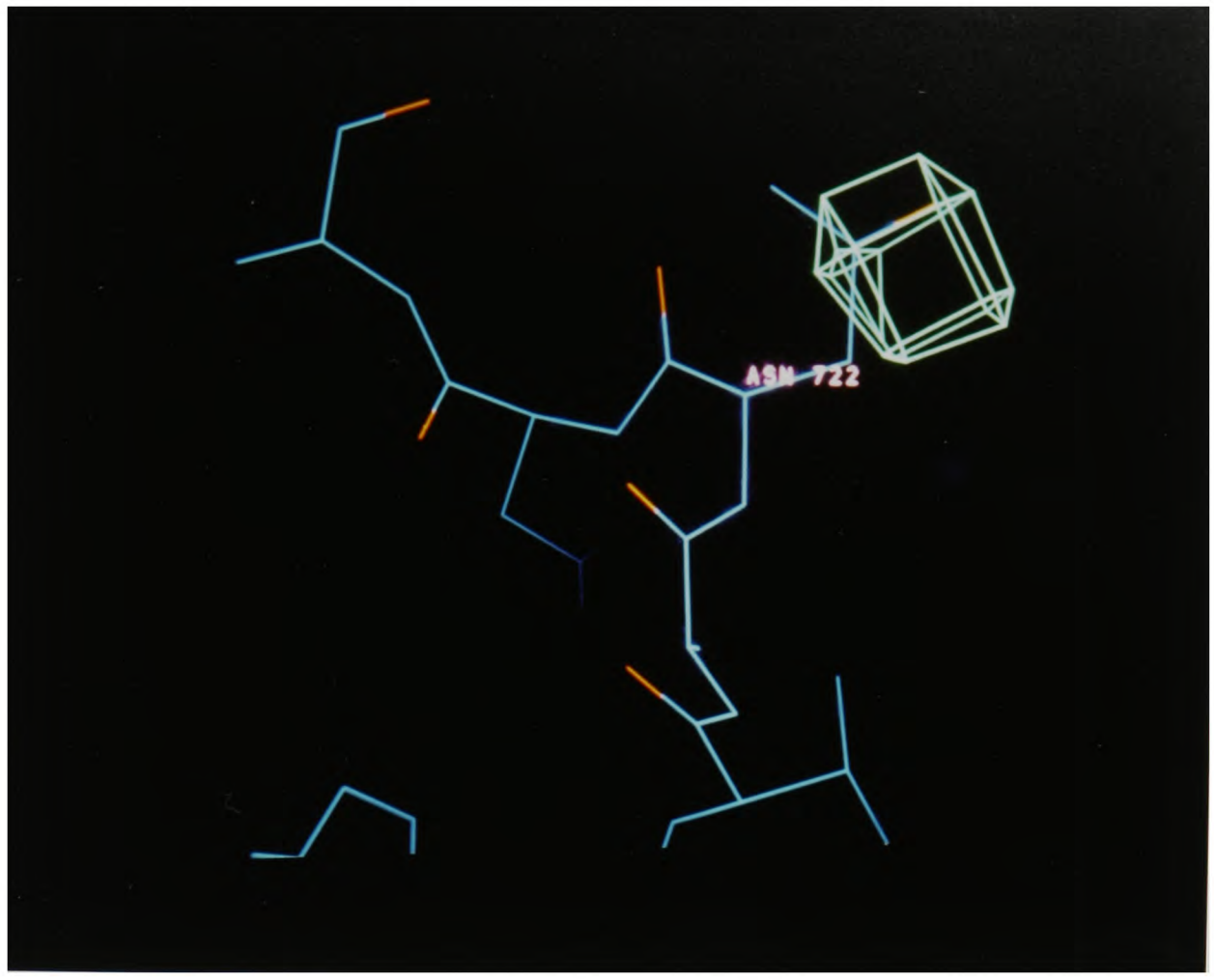
$$\% = \frac{Accessibility\ in\ O_1K}{Accessibility\ in\ G-X-G\ extended\ tripeptide}$$

Table 3.10: Sequence differences seen between O_1K and G67

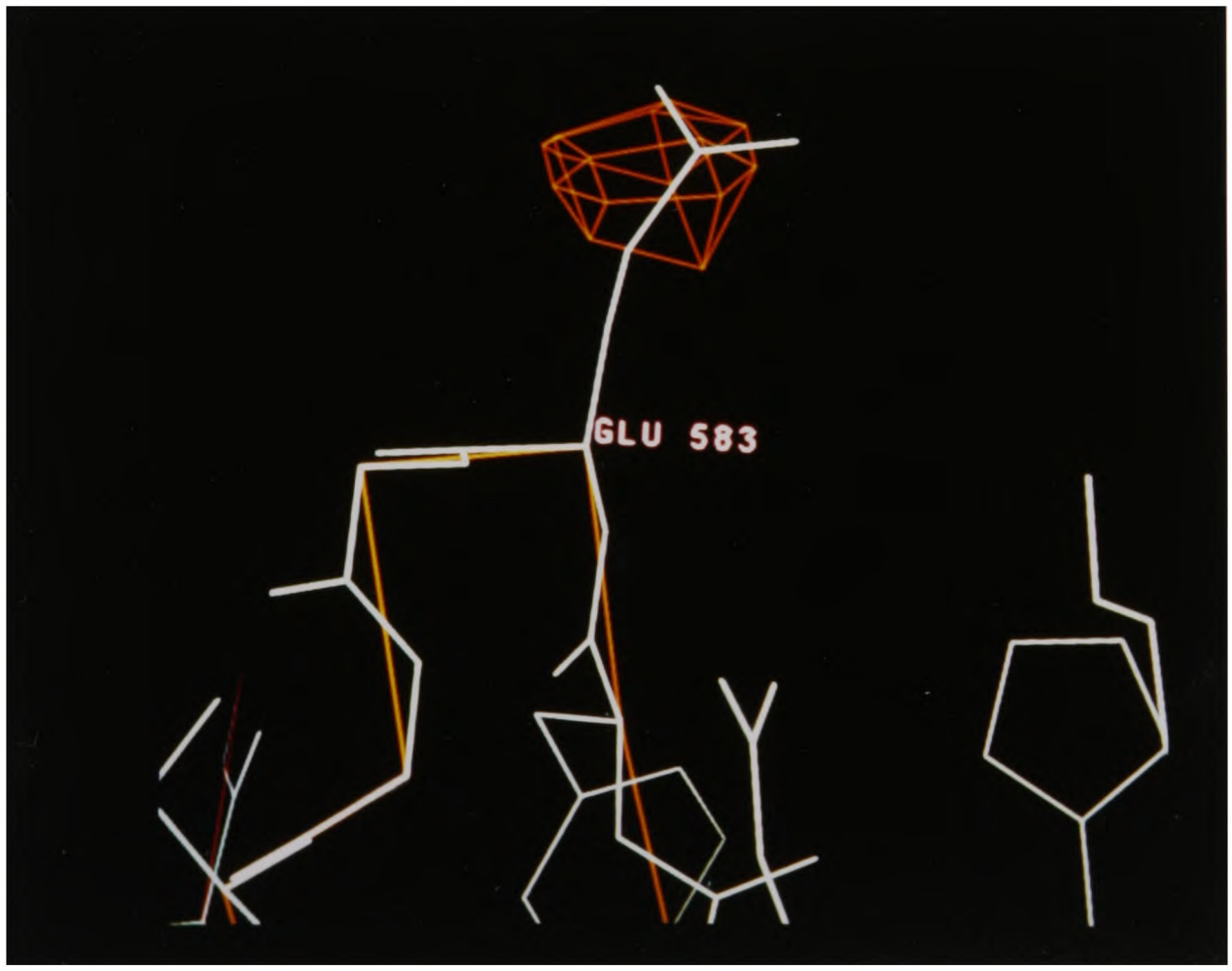


(A) Thr→Lys, 43 of VP1

these are in the loop region and can therefore only be seen in the maps calculated using data collected from the crystals treated with the reducing solution. The mutations can all be clearly seen (Figure 3.4) marked by appropriate difference density associated with the side chains of the mutated residues. Again there is no evidence for any movement of main chain atoms. The loop once more adopts the same main chain conformation in the reduced structure, despite radical substitutions at three positions in the loop (Figure 3.5). This is somewhat unexpected as the mutation (Leu to Arg) at position 148 of VP1 involves a residue revealed in the structure of the reduced



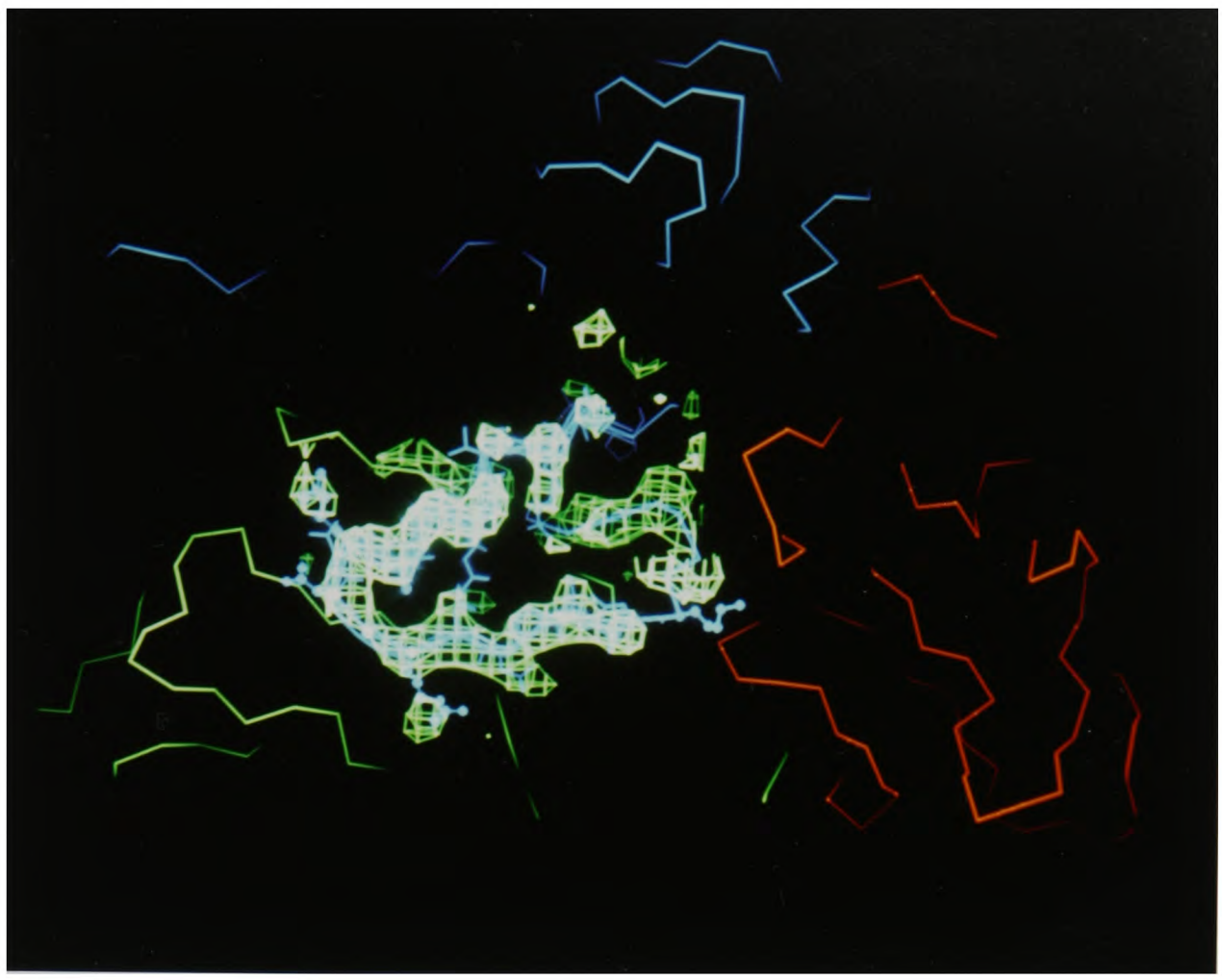
(B) Ser→Asn, 72 of VP2



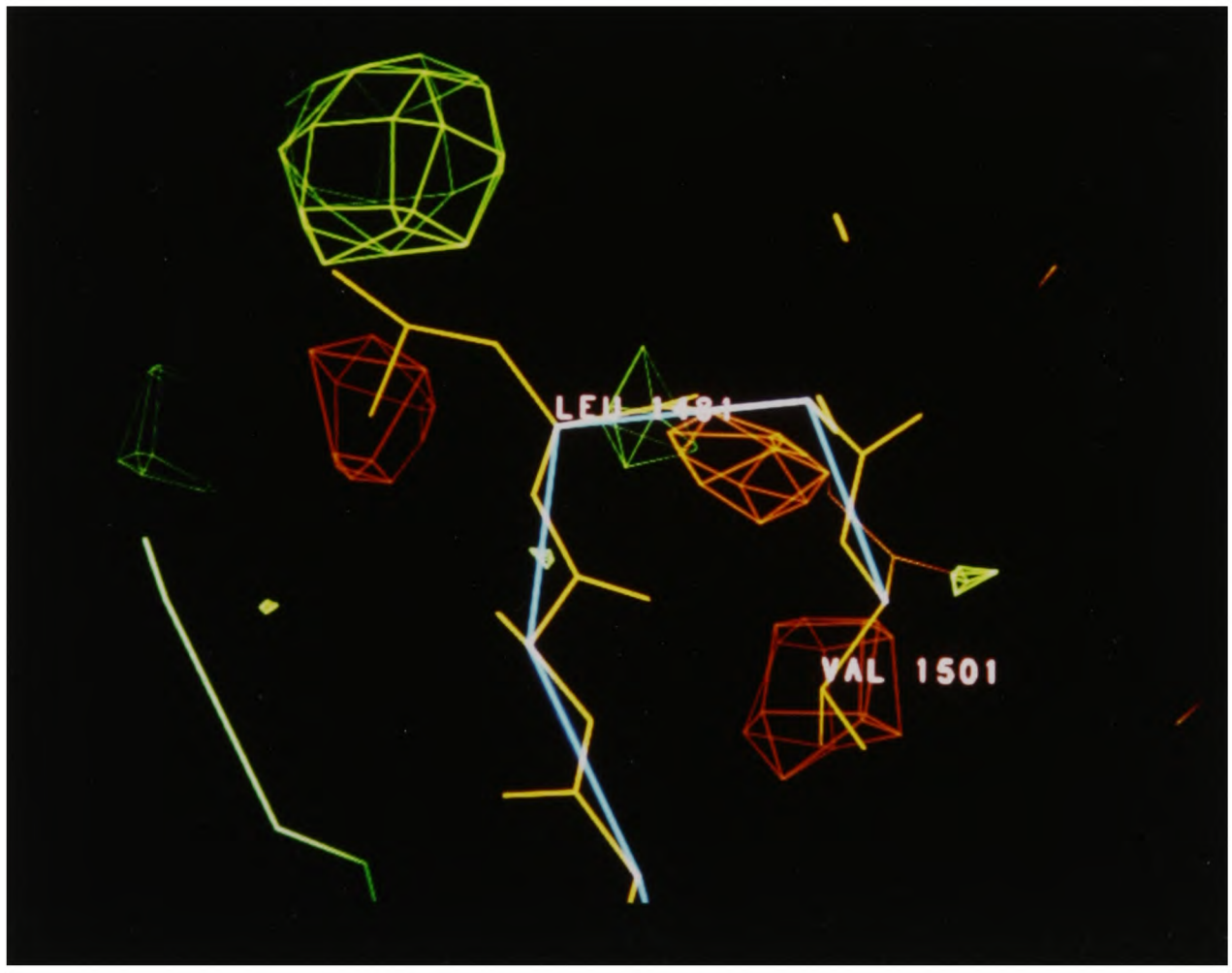
(C) Glu→Val, 58 of VP3.

Contoured at an arbitrary level.
+ve electron density – green; -ve electron density – red.

Figure 3.4: Sequence differences seen in $| F_{G67,native} | - | F_{O_1K,native} |$ difference map.



(A) $| F_{G67,reduced} | - | F_{O_1BFS,native} |$



(B) $| F_{G67,reduced} | - | F_{O_1K,reduced} |$

Contoured at an arbitrary level.
+ve electron density – green; -ve electron density – red.

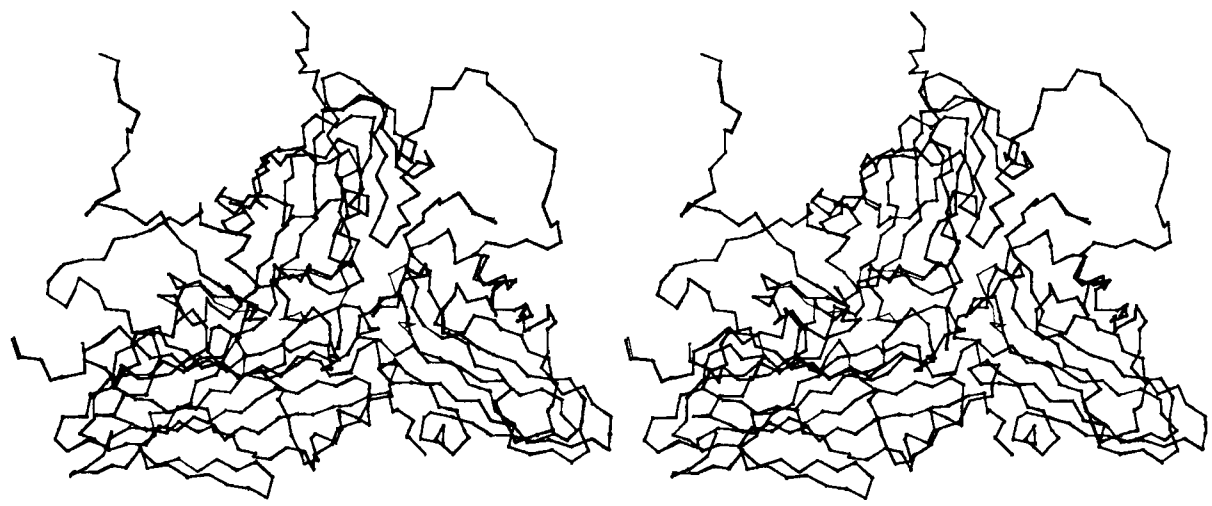
Figure 3.5: Sequence differences seen in $| F_{G67,reduced} | - | F_{O_1BFS} |$ or $| F_{O_1K} |$ difference maps.

virus to be directed inwards into a hydrophobic pocket in the surface of VP2. The change from a Leu to a polar Arg might therefore be expected to alter the way in which the loop packs against the surface of VP2, the difference density however, shows that this is clearly not the case, the only structural alterations seen occurring in the side chain, not the main chain atoms. The Arg side chain is easily accommodated as the atoms beyond Cδ are directed away from the hydrophobic pocket, back up towards the capsid surface, lying over VP2. Evidence for a further difference between O_1K and G67 is provided by a map calculated between reduced O_1BFS and reduced G67 as this shows no density associated with 133 of VP1. This shows our G67 to have a Glu at this position rather than the Gly of the O_1K. Either G67 has reverted to a Glu or the O_1K virus used to produce the reduced O_1K crystals was not the true parental clone of G67. This second explanation is the most likely cause of this difference - since selection of mutants involves repeated selection of individual viruses from plaques any heterogeneity in the parental population will be unrepresented in the cloned mutant. The parental virus used in the structure determination and G67 simply originate from different plaques.

3.5 Refinement of G–H Loop of VP1

Models have been built of both mutants using FRODO (Jones, 1985) by replacement of the mutated residue and movement of the side chain to satisfy the difference density. Refinement of these structures does not seem necessary to illustrate the differences between the viruses as there is no movement of main chain atoms however some refinement has been performed using the reduced O_1K data and model to further investigate the occupancy of this mobile loop. Reduction of the VP1–VP2 disulphide by DTT treatment causes rearrangement of three surface–oriented loops;

1. The VP1 G–H loop ('FMDV–loop'; 134–158) becomes ordered, adopting a stable conformation lying over VP2. Cys 134 (which formed half of the disulphide) moving by more than 12Å from its position in the native structure.



Produced using MOPLOT84 (D. Stuart, unpublished). Cα traces of O_1BFS Native (shown as thinner black lines) and O_1BFS Reduced (thicker black lines and small filled balls at Cα positions).

Figure 3.6: Stereo superposition of native and reduced O_1 BFS Structures

2. The VP2 E–F loop (the loop containing Cys 130, the other half of the disulphide) undergoes a rearrangement after release of VP1. The Cα of residue 130 moving by over 0.5Å (and the Cα of 131 by over 2.0Å).

3. The VP3 G–H loop (172–180) re–folds away from the surface of VP2 to accommodate the 'FMDV–loop'.

A stereo superposition of native and reduced O_1BFS structures is shown in Fig. 3.6. The occupancy of the 'FMDV–loop' has been investigated (Logan, 1991) by refinement with different loop occupancies – and the occupancy estimated as 80%. To further investigate these loops a model of reduced O_1K (based the reduced O_1BFS) was built in FRODO (Jones, 1985) and refined in XPLOR (Brünger, 1992) using a novel, weighted, bulk solvent mask.

The bulk solvent correction in XPLOR sets all grid points in the FFT calculation not specified as belonging to protein equal to a uniform electron density (SOLDEN; $1.0e^{-}Å^{-3}$) and stores the transform of this solvent mask in an array of partial structure factors, F_{part}. The corresponding F_{part} is vectorially added to each F_{calc} before calculation of R factors et.c.:

$$F_{calc} = F_{calc} + K_p F_{part} e^{-B_p sin^2\theta/\lambda^2}$$

k_p and B_p are determined by a grid search (Chapter 5), values of k_p = 0.32 and B_p = 110 were used.

The presence, or absence, of a 24 residue loop will significantly alter the bulk solvent correction. To generate a 'weighted' bulk solvent mask two sets of F_{part} are generated, one appropriate to the native model (i.e. no ordered G–H loop), the other appropriate to the reduced model (i.e. with a model for residues 134–158 of VP1). These F_{part} are then combined in different ratios to enable correct refinement of structures with different loop occupancies. The XPLOR script below generates F_{part} appropriate for a loop occupancy of 0.7

```
@xprepare_noloop.inp !!reads in ncs matrices, atomic scattering factors,
                     !!and FPART for model with NO loop
do(fcalc = fpart*0.3)
@xprepare_loop.inp     !!as above, but FPART correct for model WITH loop
do(fpart=fpart*0.7)
do complex (fpart = fcalc + fpart)
write reflection fobs fpart output=ro1k_occ07.fobs end
end
stop
```

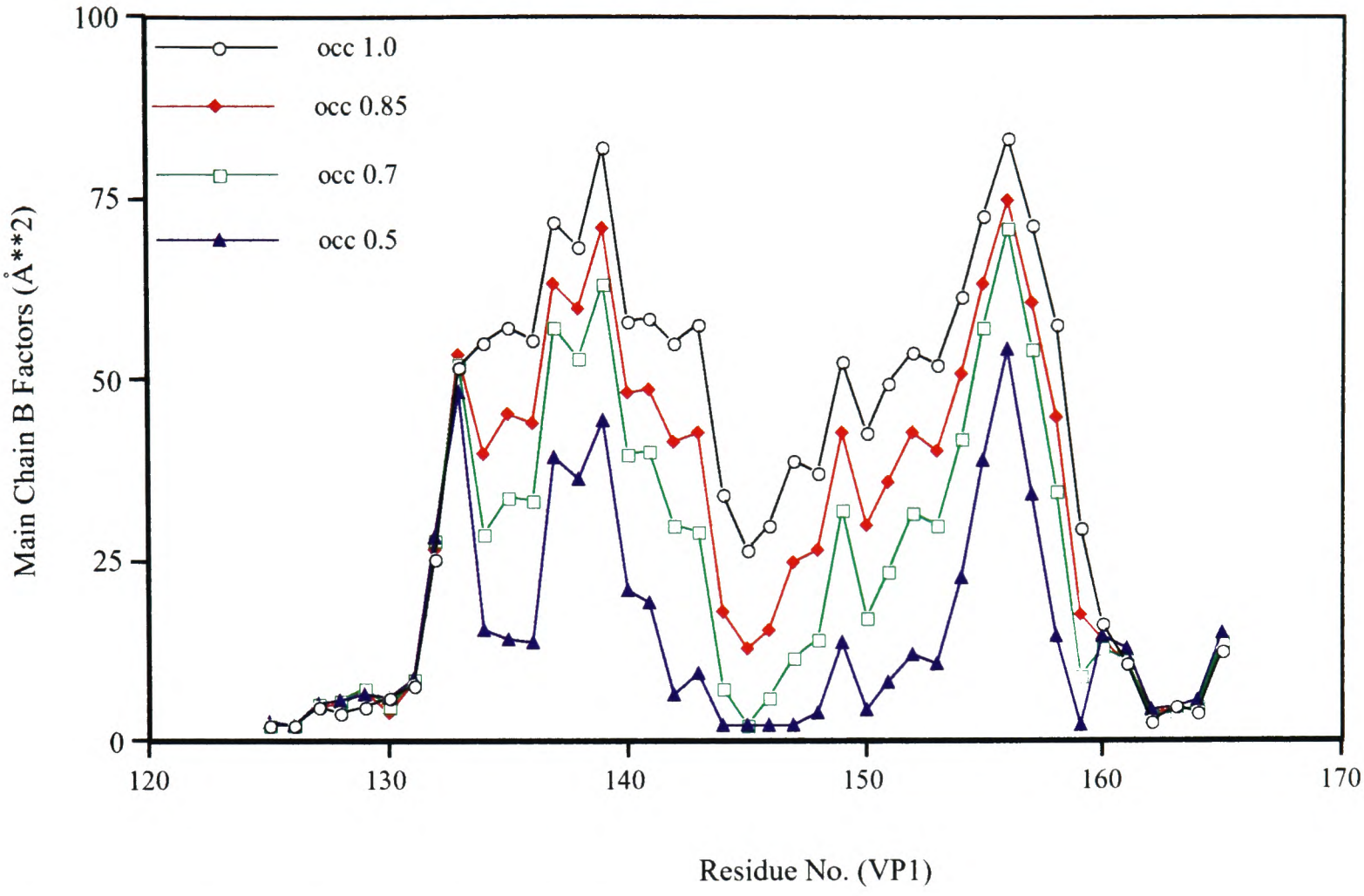


Figure 3.7: Main chain B–factors for loop residues after refinement with different occupancies.

Fifty cycles of positional refinement (tolerance[4] = 0.05Å) and thirty cycles of individual, isotropic, B–factor refinement were performed for five models with different loop occupancies, using the appropriate weighted bulk solvent mask. R factors proved insensitive monitors (R_C = 20% for all loop occupancies), however, individual B–factors for the loop residues provide a sensitive test of the appropriateness of the occupancy chosen. Refinements with loop occupancies of 0.5, 0.7, 0.85 and 1.0 gave different B–factors for the residues within the mobile loops (Figure 3.7). Occupancies lower than 0.85 always led to B–factors for residues within the mobile loops refining to the minimum allowed value of 2.0Å^2. Since these loops are mobile this indicates

[4]Tolerance specifies the distance (in Å) any atomic coordinate can deviate from the position when F_{CALC} was last computed during positional refinement. If tolerance is exceeded, F_{CALC} and all its derivatives with respect to atomic parameters are recomputed

the occupancy is being underestimated, the occupancy is therefore estimated as between 0.85 and 1.00. It is interesting to note that when refined with these occupancy values the amino acids of the RGD tripeptide, at the centre of the loop, have lower individual B factors than any other loop residues (Figure 3.7, and below).

3.6 Discussion

3.6.1 Receptor Binding

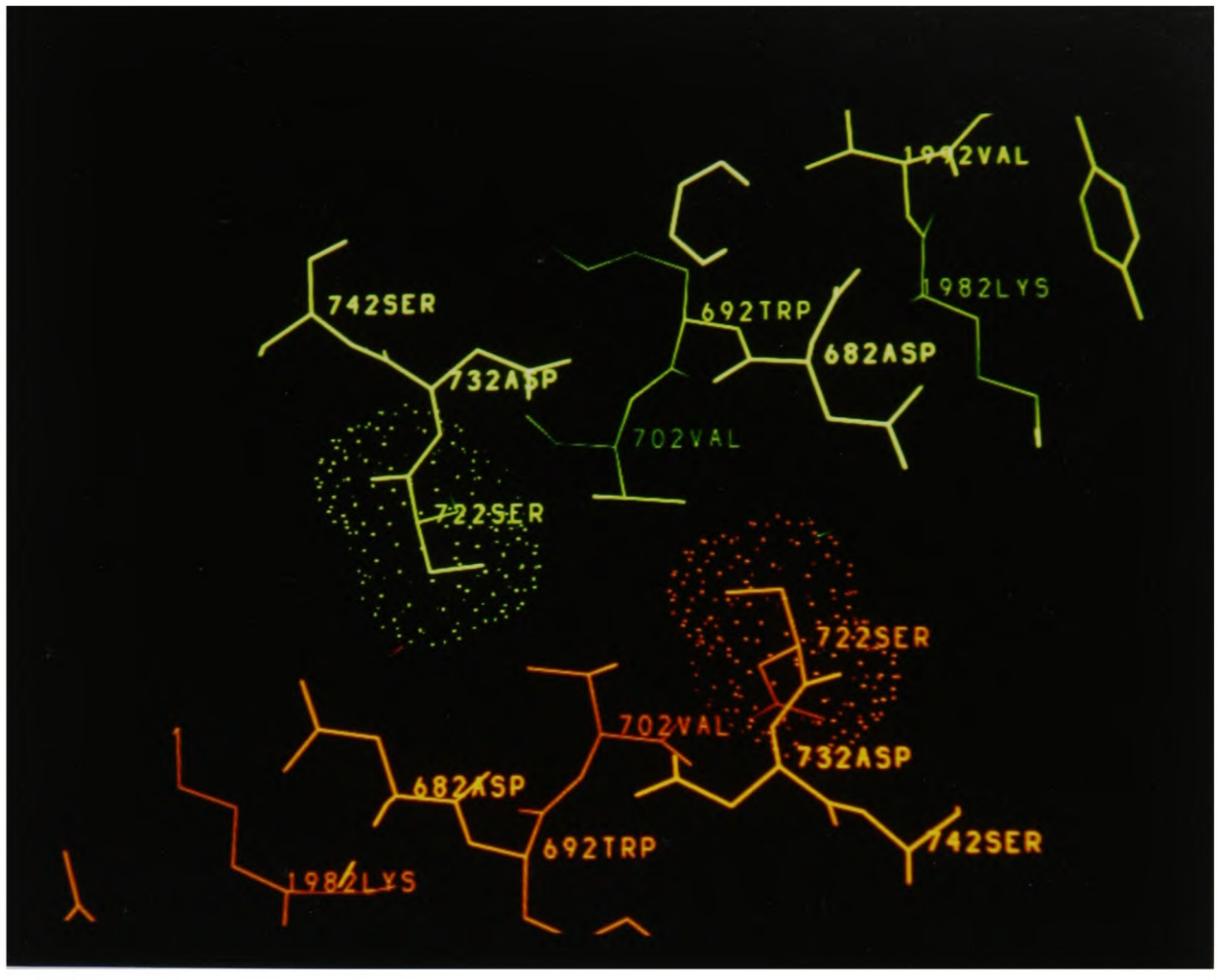
The conservation of the main chain structure of the FMDV-loop, especially the conformation and orientation of the RGD tripeptide, despite accumulation of 5 mutations (with respect to O_1BFS) within a 15 residue loop strengthens interpretations of virus activity based on this conformation. This confirms the idea that, since there is no difference in receptor binding (Logan *et al.*, 1993), or recognition of the virus by loop-directed antibodies (personal communication R. Abu–Ghazaleh), between oxidised and reduced forms of the virus the conformation of the loop as revealed in the reduced form of the virus is relevant to the oxidised form. This implies that the conformation of this loop in type O_1 viruses is constant, the orientation being the only variable, with the loop acting as an independent, hinged unit, pivoting about the disulphide–bond in the oxidised form of the virus. Refinement of the loop structure with the correct occupancy and weighted bulk solvent mask, show the RGD amino acids to be the least mobile of the loop residues – despite their location at the tip of this very mobile loop. This relative stability, taken with the conservation of the conformation of the tripeptide seen amongst these three O_1 viruses adds weight to previous comparisons (Logan *et al.*, 1993; Logan, 1991) between the conformation / orientation of this tripeptide in FMDV and other proteins. These analyses have shown that the conformation seen in FMDV is very similar to that seen in integrin–binding RGDs (e.g. γ–II crystallin (Wistow *et al.*, 1983)) but differs from the conformation seen in non–integrin–binding proteins (e.g. xylose–isomerase (Henrick *et al.*, 1987)

and α–lytic protease (Fujinaga *et al.*, 1985)). The structure revealed by the reducing treatment in these three viruses, therefore adds to the evidence implicating an integrin as the FMDV receptor.

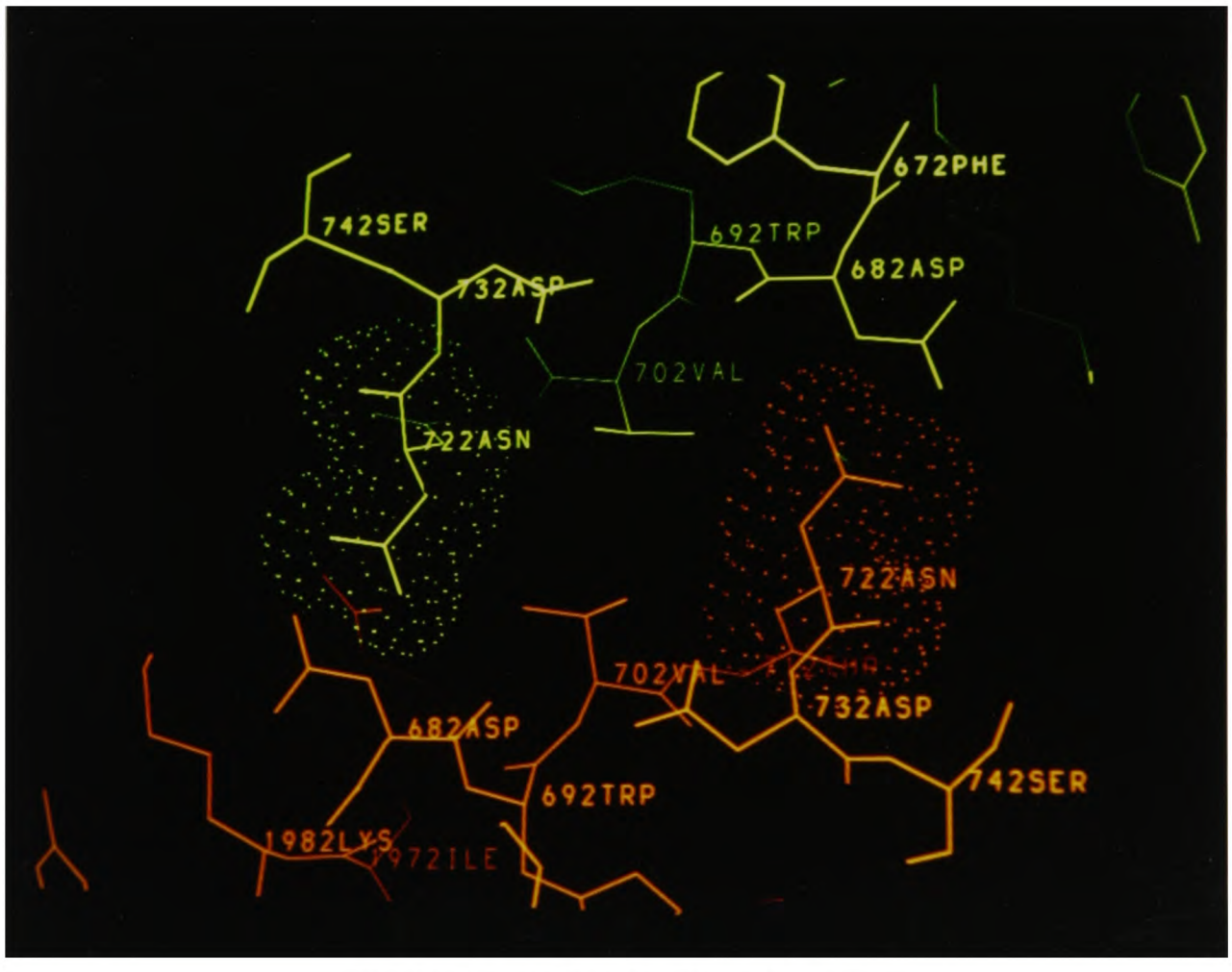
3.6.2 G67 Crystal Packing

The mutation at 72 of VP2 (Ser to Asn) is the likely cause of the altered packing of the G67 crystals. The I23 packing interaction consists of stacking of the Ser side chain over the side chains of Asp 68 and Gln 196 of VP2 of a symmetry related virion (Fig 3.8A). The Asp and Gln are hydrogen bonded to each other. A fairly tight junction is formed with a pair of equivalent interactions making up the crystal contact. By definition it is the information concerning packing between a virion and a 90° related virion that is lost it is therefore impossible to calculate even an unaveraged map to visualize the interactions occuring in th G67 crystals. In averaged maps even the I23 interaction is not clearly defined as the predominant conformation is seen i.e. the conformation for the protomers not involved in crystal contacts. Small rearrangements to improve the contacts will therefore be acceptable. By examining the two packing interactions using a graphics system it seems likely that it is less a case of the interaction between the virion and the 90° related virion becoming more favourable in the G67 crystals (there is no obvious partner for Asn to hydrogen bond to, Fig 3.8C) than of the I23 interaction becoming less favourable (the contact region becoming somewhat 'crowded' with the larger Asn side chain, Fig 3.8B). This implies that the forces holding the G67 crystals together will be weaker than those involved in the true I23 O_1 crystals - which seems to be confirmed by the need of higher salt concentrations to crystallize G67.

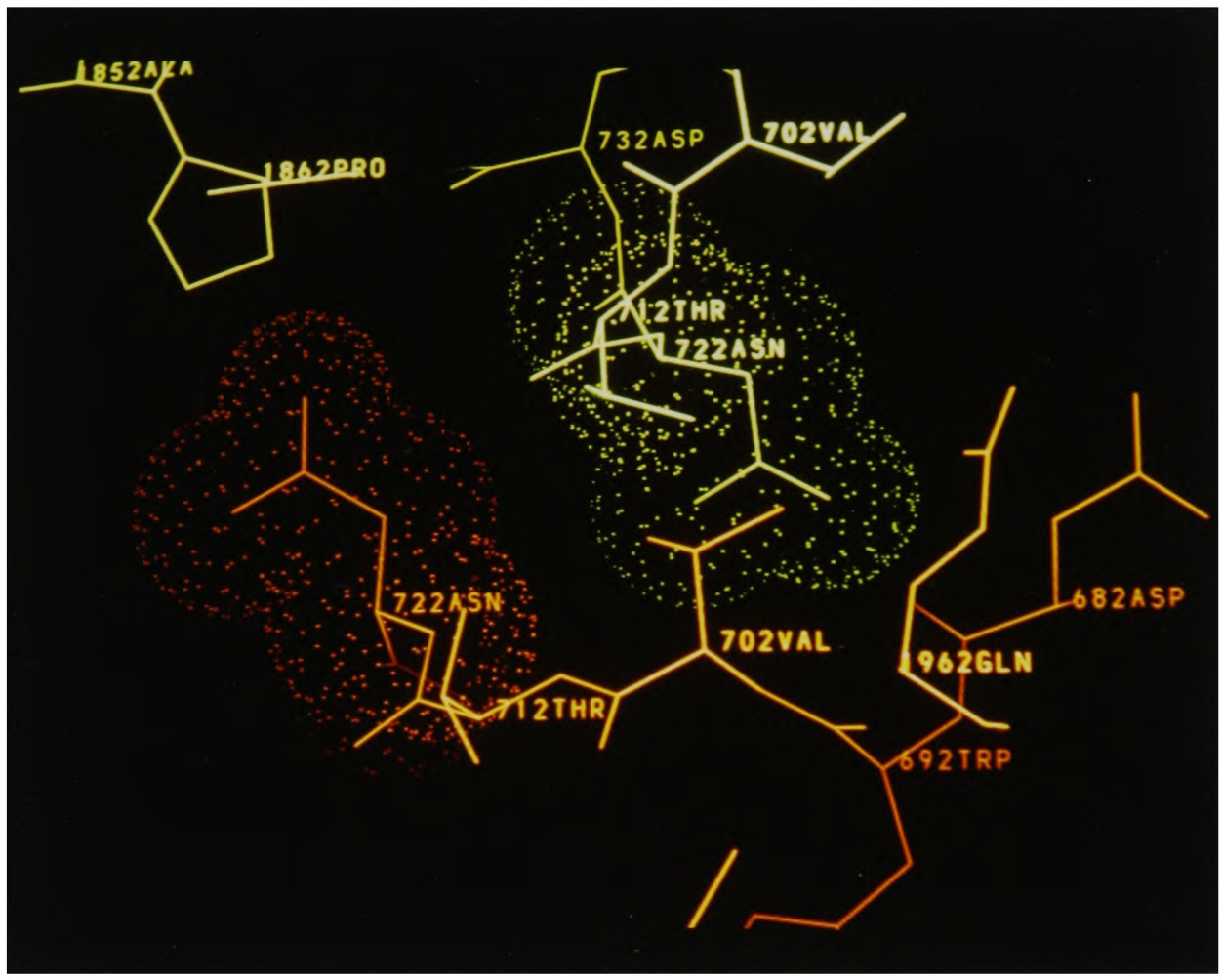
The packing might be further investigated by attempting to better model the interaction between the virion and the 90° related virion using simulated annealing. This would allow more long–range rearrangements of the loops containing 72 and 196 of VP2, whilst fixing the underlying structure. Such refinement might reveal if small



(A) 'True' I23 Packing in O_1K



(B) 'True' I23 Packing in G67



(C) Mis–packing between two 90° related virions in G67

Atoms from one virion are coloured red, from a symmetry related virion green. A van der Waals surface is shown for residue 72 of VP2.

Figure 3.8: Packing of G67 crystals

	O_1K	G67	H_2O Accessibility in O_1K		Size Change ?	Charge Change ?	MAb Eluded
			$Å^2$	% *see below*			
VP1							
43	Thr	Lys	117.3	78	Bigger	+1	C8
148	Leu	Arg	44.8	25	Bigger	+1	D9
VP2							
72	Ser	Asn	92.2	77	Bigger	–	C9
VP3							
58	Glu	Val	169.6	91	Smaller	+1	14EH9

$$\% = \frac{Accessibility\ in\ O_1K}{Accessibility\ in\ G-X-G\ extended\ tripeptide}$$

Table 3.11: G67 Mutations thought to be responsible for evading MAbs

movements of the loops would make the interaction more favourable.

3.6.3 Antigenic Structure

The structural changes occurring in G67 (relative to O_1K) as a result of MAb-escape driven point mutations are seen to be very limited having effect only by alterations in the nature of the side chain and not requiring movement of the main chain to produce antibody escape. The majority of the escape mutations involve surface–oriented side chains (Table 3.11) and since all of the mutations involve either substitution for a bulkier side chain or changes in the charge (or both), it is relatively simple to understand the process leading to escape in terms of direct interference with binding of the attacking antibody by the mutated side chain. The mutation at 148 of VP1

(Leu→Arg) appears at first sight to be more difficult to explain as an escape mutation in such simple terms since the side chain of this residue is directed inwards in O_1K and is not obviously exposed to antibody attack. Predictions based on the reduced O_1K structure might have suggested the effect of this mutation would be by alteration of the main–chain conformation of the G–H loop. The structure shows that this is not the mode of action since the larger Arg side chain is easily accommodated, lying over the surface of VP2 (directed towards the 5–fold) rather than penetrating the hydrophobic pocket. Since the Arg side chain is therefore more exposed than the Leu (H_2O accessibility 78.8Å^2, 31% of Arg in Gly-X-Gly tripeptide) it may interfere with binding of a MAb directed at this region and recognizing the loop as seen in the reduced structure. It must however be remembered that even in reduced virus a proportion of the G–H loops will be in an 'up' conformation and also that the time–course of experiments producing both MAbs and MAb–escape mutants is long enough that selection will be for MAbs that recognize the oxidised, 'native' virus. Loop directed MAbs selected for in this way may therefore recognize the loop in some orientation other than that seen in the reduced virus. Such MAbs should in theory be detected by their decreased affinity for reduced virus, however we must be careful in using the reduced virus structure to deduce much about interactions between such MAbs and the virus and about the way in which this interaction is disrupted in the MAb escape mutant.

The area of interaction between an antibody and an antigen is generally greater than 700Å^2 (Alzari *et al.*, 1988), monoclonal mutants such as G67 therefore demonstrate the exquisite sensitivity of this interaction in that small alterations (in terms of conformation and chemistry) such as the Ser to Asn change at 72 of VP2 are sufficient to provide escape from an antibody. These mutations are also seen to be very limited in terms of the structural changes produced, this has been noted in other studies of mutants bearing MAb-escape driven point mutations e.g. in studies of neuraminidase escape mutants (Tulip *et al.*, 1991).

The area of largest structural change between O_1K and O_1BFS is seen to be associated with the changes at 56 and 60 of VP3 (Figure 3.2C). The mutation at 56, Arg→His, being especially important in determining the structural changes due to the change in charge (-1) leading to loss of the bound sulphate ion and two H_2O molecules from the region, which leads in turn to the movement of side chain His 195 of VP1 which lies close to this region. A difference between the 56 and 60 mutations in O_1K and the MAb driven G67 mutations is seen if the accessibility of the residue replaced is considered. The G67 mutations (if the loop mutation is excluded) all occur at very exposed residues with solvent accessibilities greater than 75% of the accessibility of that residue type in an extended Gly-X-Gly tripeptide, the O_1K mutations at 56 and 60 of VP3 occur at less exposed residues with accessibilities less than 20% of the tripeptide value. By comparison with the MAb–induced G67 mutations, which result only in alteration of the effected side chain, these mutations, by effecting less exposed residues, are seen to be 'good–value' in terms of the structural changes produced. Residue 56 of VP3 is very variable amongst O_1 viruses, flipping between Arg and His both in the field and with passage *in vitro* (personal communication N. Knowles). As such a variable residue it has often been thought of as 'unimportant', however since this change produces such large structural changes within a defined antigenic site, it is likely to be very important. Based on the observed structural differences we predicted that MAbs directed at site 4 (this region of the capsid) would distinguish between O_1BFS (Arg at position 56) and O_1K (His at position 56). MAb 14EH9 was indeed able to do this, binding to O_1K but not to O_1BFS (personal communication J. Crowther) demonstrating that the structural alterations seen are antigenically relevant. By maintaining great variability at this position within the virus population in the field the likelihood of any one virus being recognised by an antibody directed at this site is decreased. Likelihood of recognition is also lowered if the residue frequently 'flips' during replication within an infected animal, providing a 'time–dependent' variation throughout the course of an infection aiding immune escape.

Chapter 4

Deconvolution of Data From Intimately Twinned Crystals of FMDV

This Chapter describes the novel procedures developed to make use of data collected from crystals of mutant G67 (see Chapter 3). As noted in the previous chapter this virus initially appeared to crystallize isomorphously with O_1BFS and also with its parent O_1K. However, when attempts were made to divide the data into swapped and unswapped streams the linear correlation coefficient in reciprocal space between the G67 data and either O_1BFS or O_1K data (for either indexing hand) was very low – for all the crystals less than 0.5 in either hand. This compares very poorly with the correlation seen between the O_1K and O_1BFS data where, in the correct hand, the correlation is of the order of 0.8. The inability of the correlation coefficient to distinguish between the two indexing hands suggests extra, 4–fold symmetry is present in these data. Such 4–fold symmetry is geometrically impossible for an icosahedral virus (which has no inherent 4–fold symmetry) in a cell that is too small to allow association of 4 viral particles around such a symmetry axis, however, statistically the data did appear to belong to point group 432. The data processed in this point group (Chapter 3) yielded excellent merging R factors – equivalent to the merging R factors for data collected from other O_1 virus crystals and processed in I23. The quality of these solutions suggests that no additional information has been lost by processing the data in this higher symmetry point group.

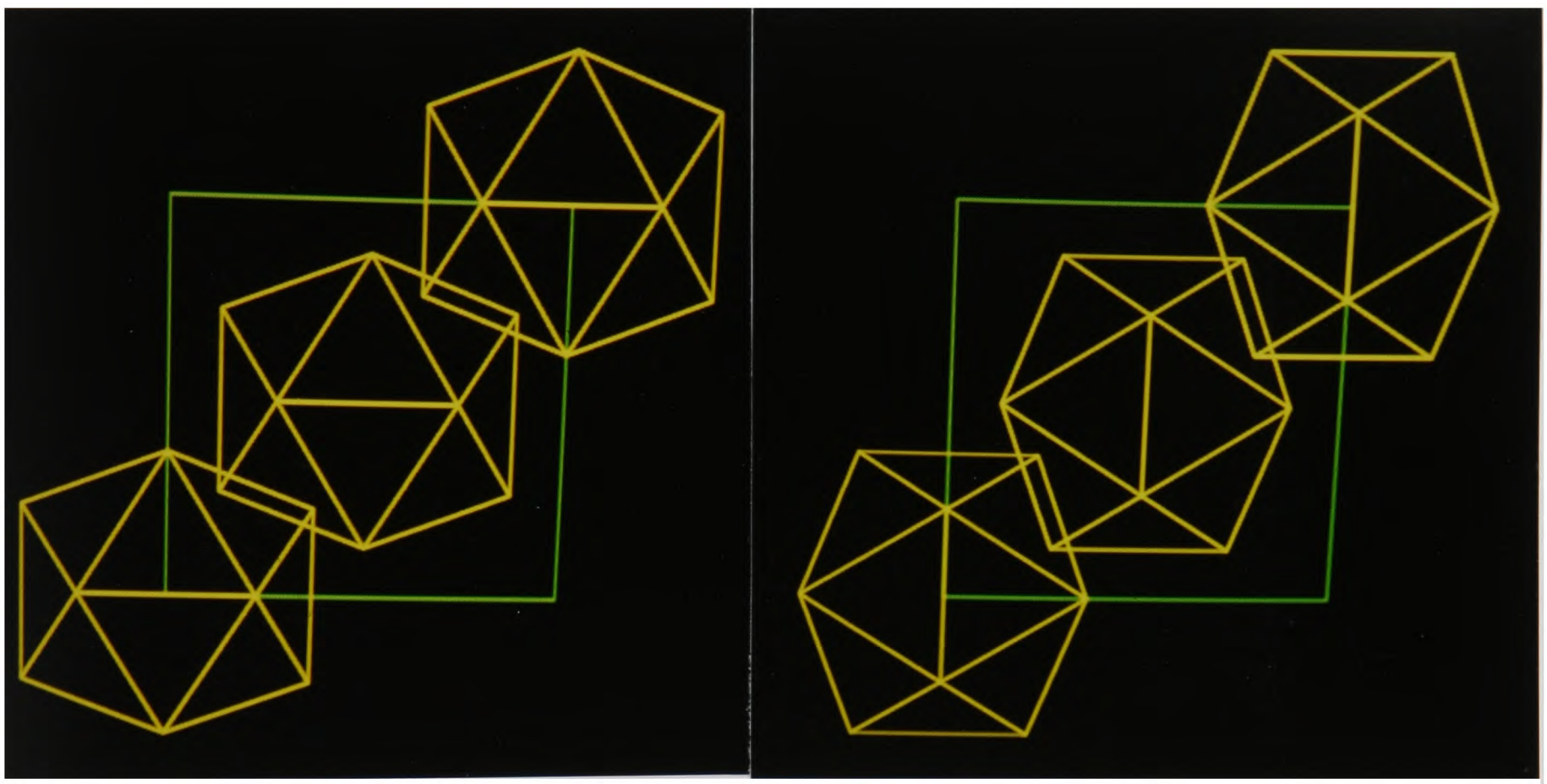


Figure 4.1: Packing ambiguity in I23.

4.1 Nature of the Disordering

This 'statistical symmetry' may be explained by assuming a disordering of the crystal packing. In space group I23 there is an ambiguity in the orientation of the virion particle in the cell (Fry *et al.*, 1993; Fry, 1991). As already noted (Chapter 2), this ambiguity corresponds to the two different ways in which an icosahedron can be placed on a specific threefold axis. The two orientations are related by a 90° rotation about a particle 2–fold (Figure 4.1). Within any crystal the particles are all in the same relative orientation, but the choice is random between crystals. We routinely divide data collected from such crystals into swapped and unswapped streams by comparison with a reference data set.

Within the crystals of G67 the extra (statistical) symmetry may be thought of

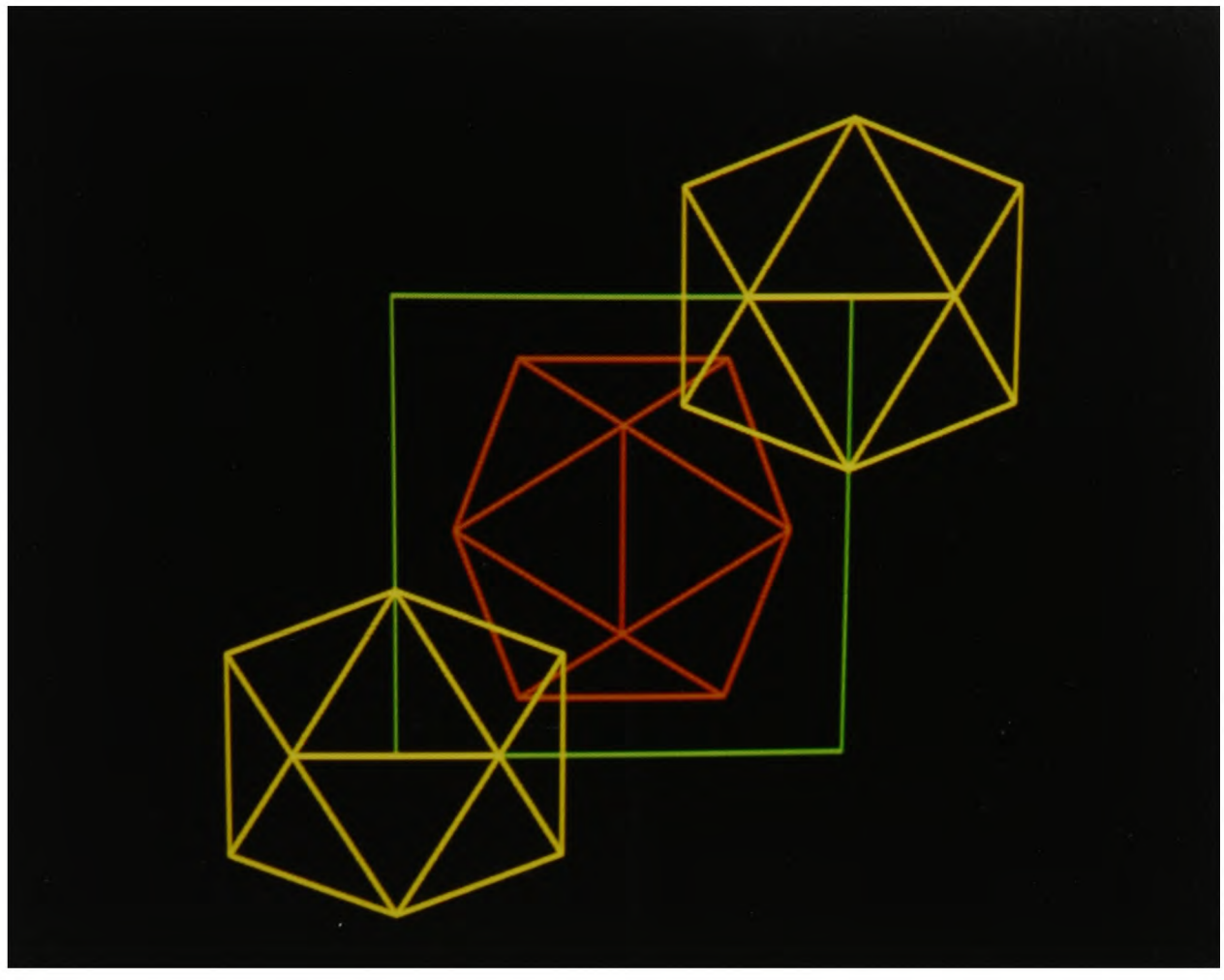


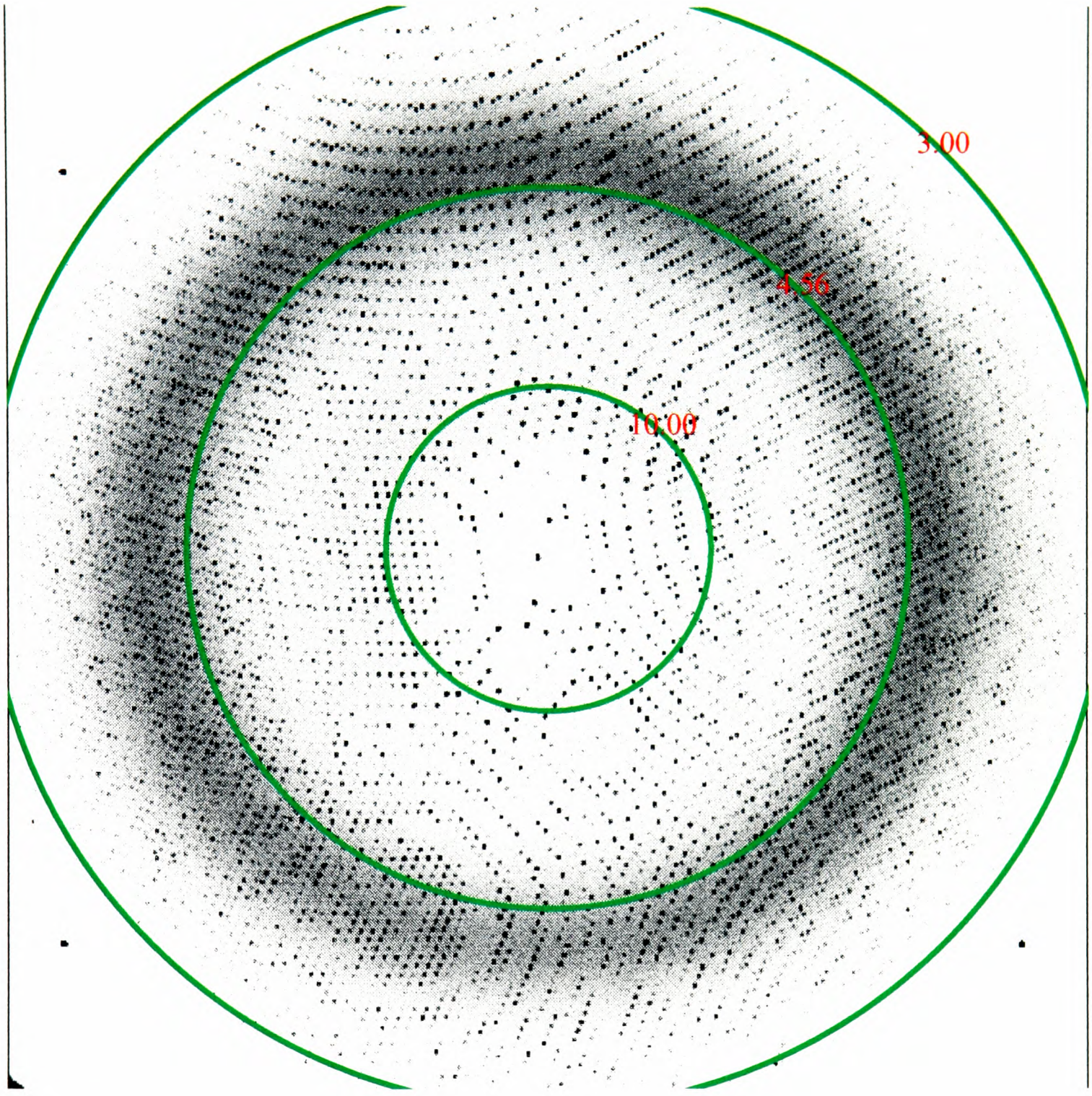
Figure 4.2: Packing of G67 crystals.

as arising from a mis–packing of the virions such that both orientations are present within the same crystal (Figure 4.2). Since the two lattices are exactly superimposed it is impossible to identify this 'twinning' phenomenon simply by examination of the diffraction pattern.

If the ratio of the two orientations is 50:50 then the data will have apparent 4–fold symmetry. The success with which the data could be processed in PG432 implies that the ratio of the two orientations must be very close to exactly 50:50.

To further understand the disordering we have attempted to use I23 data from the closely related parental virus (O_1K) to produce an accurate model of the G67 data.

In point group 23 reflection h,k,l is independent of reflection k,h,l i.e. $F_{h,k,l} \neq F_{k,h,l}$ however, in point group 432, these reflections are equivalent; $F_{h,k,l}=F_{k,h,l}$ (Table 4.1). This means that to reduce the I23 data to 432 a model for these reflections must be



Scanned Film Data

Oscillation 0.00° to 0.50°, Wavelength 0.890A, Distance 190.0mm

Figure 4.3: G67 Film Image

	I23	I432
Laue Symmetry	m3	m3m
Reflection conditions	$h + k + l = 2n$	
Equivalences	$[F_{h,k,l}] = [F_{\bar{h},\bar{k},\bar{l}}] = [F_{\bar{h},k,l}] = [F_{h,k,\bar{l}}]$ $\alpha_{h,k,l} = -\alpha_{\bar{h},\bar{k},\bar{l}} = -\alpha_{\bar{h},k,l} = -\alpha_{h,\bar{k},l} = -\alpha_{h,k,\bar{l}}$	
Asymmetric Unit	$0 \leq x \leq 1$ $0 \leq y \leq \frac{1}{2}$ $0 \leq z \leq \frac{1}{2}$ $y \leq min(x, 1-x)$ $z \leq y$	$0 \leq x \leq \frac{1}{2}$ $0 \leq y \leq \frac{1}{2}$ $0 \leq z \leq \frac{1}{4}$ $z \leq min(x, \frac{1}{2}-x, y, \frac{1}{2}-y)$

Table 4.1: Crystallographic parameters for I23 and I432.

	R–factor	Correlation Coefficient
VECTOR MEAN	0.30	0.70
ARITHMETIC MEAN	0.14	0.89

Table 4.2: Scaling of model data based on O_1K I23 data to 432 G67 data.

constructed. Two extreme models for the disordering differing in their treatment of these equivalent reflections are possible:

1. If all virion 2–folds are randomly arranged throughout the crystals then interference effects must be considered and a vector mean of the phased I23 structure factors will model the 432 data i.e.:

$$F^{432}_{h,k,l} = (F^{23}_{h,k,l} + F^{23}_{k,h,l})/2$$

2. Mosaic blocks of the crystal contain similarly oriented particles, but the arrangement of 2–folds between different mosaic blocks is random. In this case there will be no interference effects and the arithmetic mean of the intensities should model the 432 data i.e:

$$I^{432}_{h,k,l} = (I^{23}_{h,k,l} + I^{23}_{k,h,l})/2$$

'Folded' data sets based on the O_1K I23 data were constructed using both these models. These folded data were scaled in resolution shells to the observed G67 data and R factors and correlation coefficients calculated to monitor the agreement between the data sets. The results are shown in Table 4.2. This clearly shows that the arithmetic mean of the intensities provides an excellent model for the disordering, implying the disorder is occurring between mosaic blocks within the crystals.

4.2 Calculation of Difference Maps

The ability to effectively model the disorder enabled development of a protocol to calculate difference maps between G67 and other O_1 viruses which had crystallised

in I23. Firstly the I23 data is 'folded' into 432 using the arithmetic mean of the I23 intensities to model the 432 intensities. This folded data is scaled to the observed G67 data in resolution shells. The scaled data may then be 'unfolded' by duplicating the reflection records so that two copies of h,k,l are written as h,k,l and k,h,l with the same amplitude. By associating known, I23, phases with these amplitudes difference maps may be calculated in I23. These difference maps are somewhat noisy (as might be expected) but can be improved, by one cycle of 5–fold averaging over the NCS related subunits within the crystallographic asymmetric unit. By averaging about the 5–fold of one orientation, that orientation is reinforced whilst information corresponding to the other orientation is destroyed. A single cycle produces sufficient improvement in the maps to reveal the structural differences between the viruses (Chapter 3).

4.3 Deconvolution Strategy

The nature of this disordering suggested a possible route for further deconvolution of the data.

Within these crystals two non–equivalent lattice points in I23 are superimposed, leading to a loss of information. Such disorder and overlap of data in reciprocal space is not uncommon but we are presented here with a system in which the data inherently oversamples the molecular transform. This oversampling arises from two sources:

1. The disordered solvent

2. The 5–fold non–crystallographic symmetry present in this system

We can therefore use solvent flattening and NCS averaging to produce improved estimates of the true I23 structure factors. Our observed data provides a series of constraints in reciprocal space, one for each pair of reflections:

$$|F_{h,k,l}^{OBS-432}| = \sqrt{\frac{|F_{h,k,l}^{23}|^2+|F_{k,h,l}^{23}|^2}{2}}$$

These constraints maybe applied to further improve the estimates of the unknown intensities in I23. Cycling through map calculation and constraint imposition therefore seems likely to provide a practical way to regain some of the lost information.

4.3.1 Test Data

To develop a deconvolution protocol 'test' data were generated from a model of O_1BFS DTT soaked (Logan *et al.*, 1993). I23 $|F_{calc}|$ were generated in the range 30 to 3.5Å for two models:

1. To model the parental data I23 $|F_{calc}|$ and phases were generated from a model with residues 130 to 160 of VP1 (corresponding to the surface exposed G–H loop) deleted.

2. To model the G67, mutant, data the complete O_1BFS DTT soaked model was used to generate $|F_{calc}|$ in I23. These were 'folded' into 432 using the process described above and the I23 $|F_{calc}|$ were 'thrown away' for all further calculations, except where they could be used as a check of the process.

The success of the deconvolution process may therefore be monitored in two ways. A subjective assessment may be made based on the quality of the density for the truncated loop region - reconstructing this density is taken as the primary aim of the deconvolution. Secondly, as this is synthetic data the agreement between the true I23 amplitudes and the estimates of these amplitudes obtained from the deconvolution process provides an objective measure of the success of the deconvolution.

The programs used to perform the envelope determination, NCS symmetry averaging and solvent flattening are described in Chapter 5. I will only describe those procedures unique to the deconvolution process in detail here. The process is summarised in Figure 4.4. Envelopes were generated using the parental I23 data – the

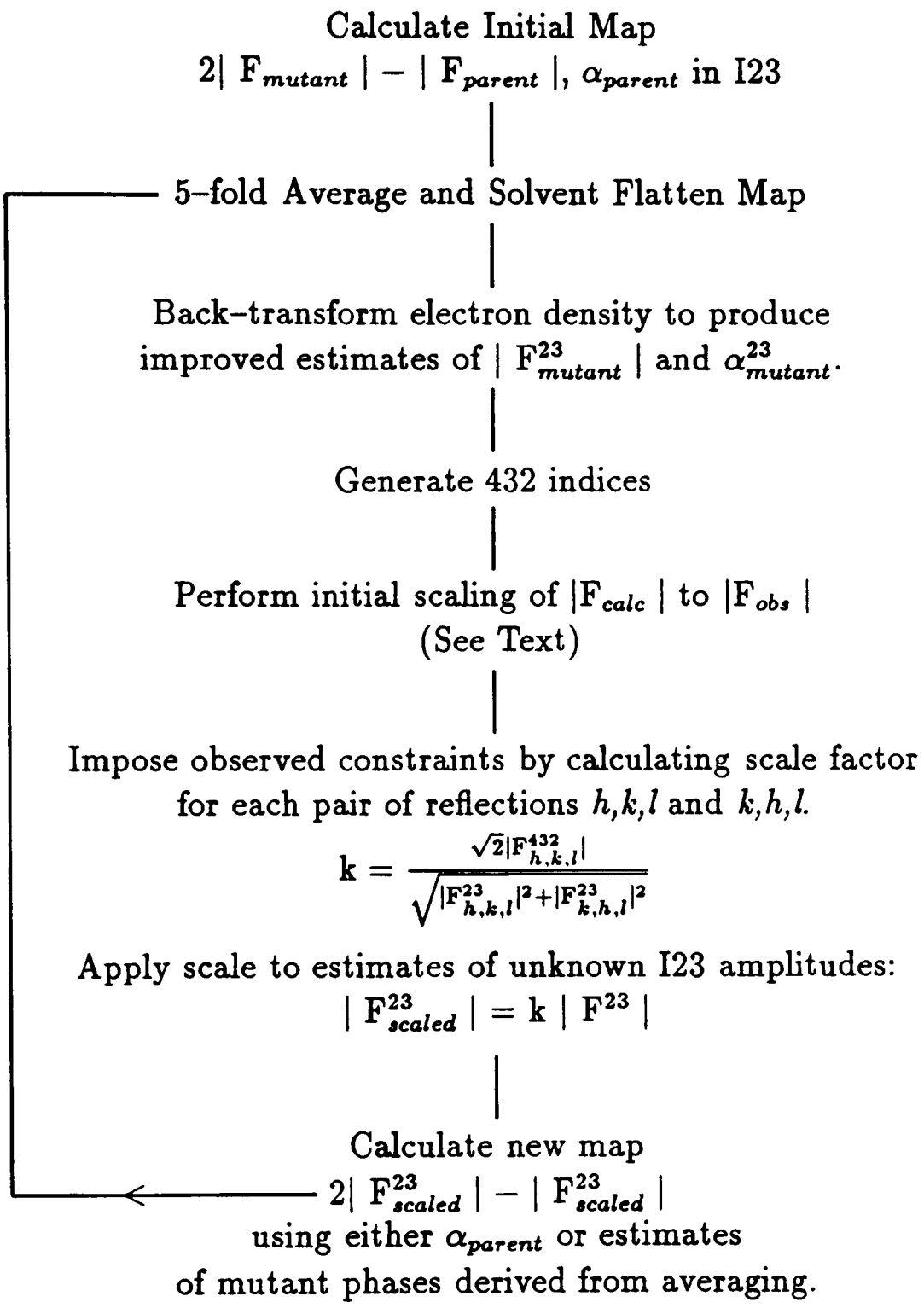


Figure 4.4: Process used to deconvolute data

parent and mutant structures are so similar these should prove entirely appropriate, and of better quality than envelopes calculated using the noisy 'unfolded' data.

Back–transformation of the averaged and flattened map generates new estimates of the amplitudes and phases, in I23, for the mutant virus. Initially, any phase information obtained from this back–transformation was discarded. This seems appropriate since we have a very accurate set of phases for the parent virus obtained by 719 cycles of NCS symmetry averaging (Fry *et al.*, 1993; Fry, 1991) – the structural differences between the parent and the mutant are small, therefore, these accurate parental phases are likely to be closer to the true I23 mutant phases than those phase estimates produced by the deconvolution process.

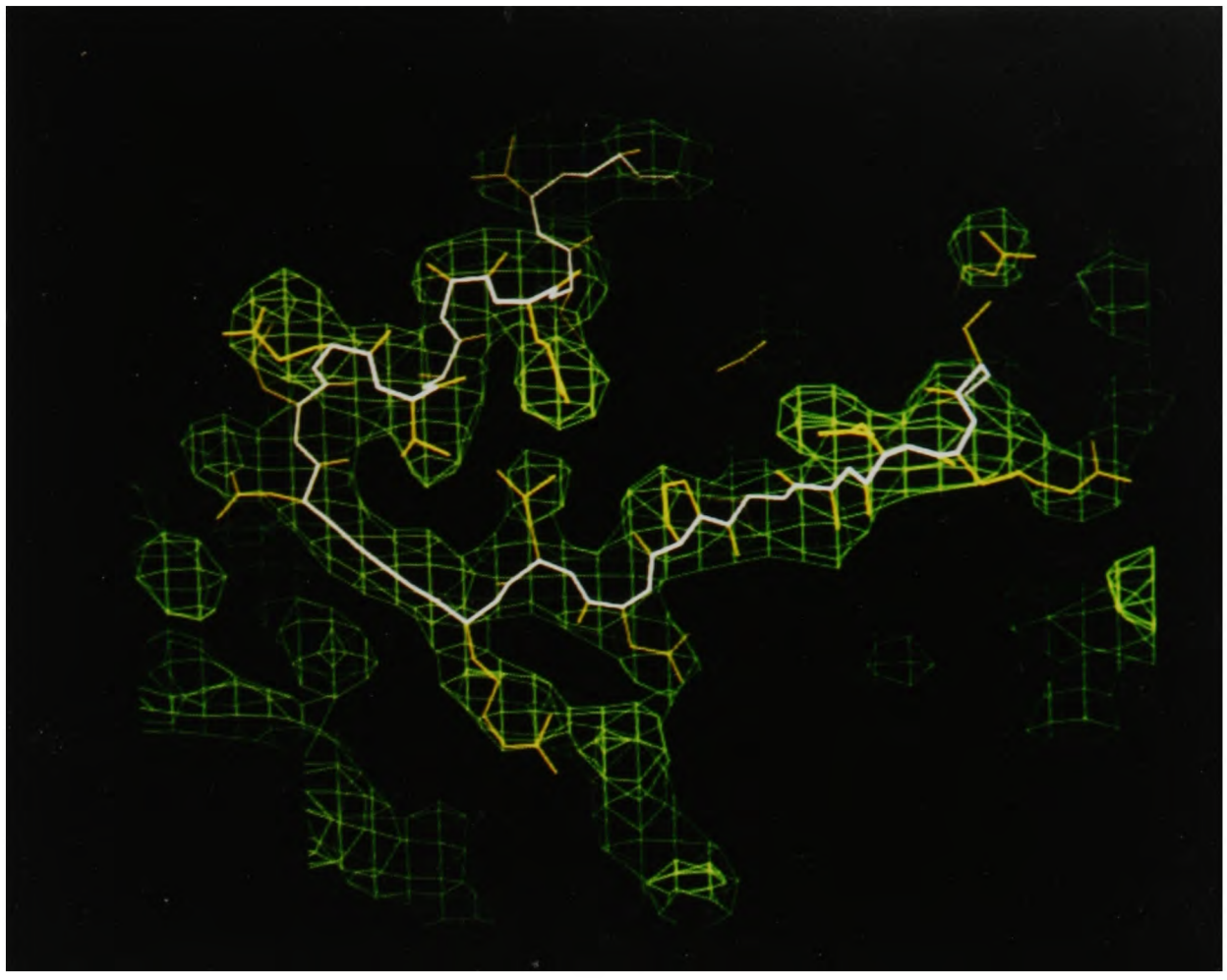
The new, mutant, I23 amplitudes will not be on the same absolute scale as the observed data, therefore, before applying the scaling constraints arising from the reflection pairs an overall scaling of the calculated to the observed data must be performed. The overall scale is calculated by generating the 432 indices and keeping large reflection records for each unique I23 reflection containing I23 indices, 432 indices, observed 432 amplitudes and the new estimates of the I23 amplitudes. A first scaling pass is made through the data where scales are calculated in resolution shells based on the 'singlet' reflections within each shell. These reflections are those such as the $h = k$ reflections for which there is only one copy in the I23, and in the folded, data sets. These singlet–scales may then be applied to all the reflections within each resolution shell. Once this global scaling has been performed a second scaling pass is made applying the constraints derived from the observed data, to calculate and apply independent scales for each reflection pair.

Initial attempts at deconvolution were made with $|\ \Delta\ F|$ calculated as $|F_{mutant}| - |F_{parent}|$, these differences were scaled and then the next map calculated as $|\Delta F|, \alpha_{parent}$. This process proved to be very unstable, the map quality degrading rapidly and tests made calculating $|F_{parent}|, \alpha_{parent}$ maps showed that the FFT did not seem to handle single amplitude map calculations correctly (i.e. the map was noisier than a map calculated with coefficients $2|F_{parent}| - |F_{parent}|, \alpha_{parent}$). This

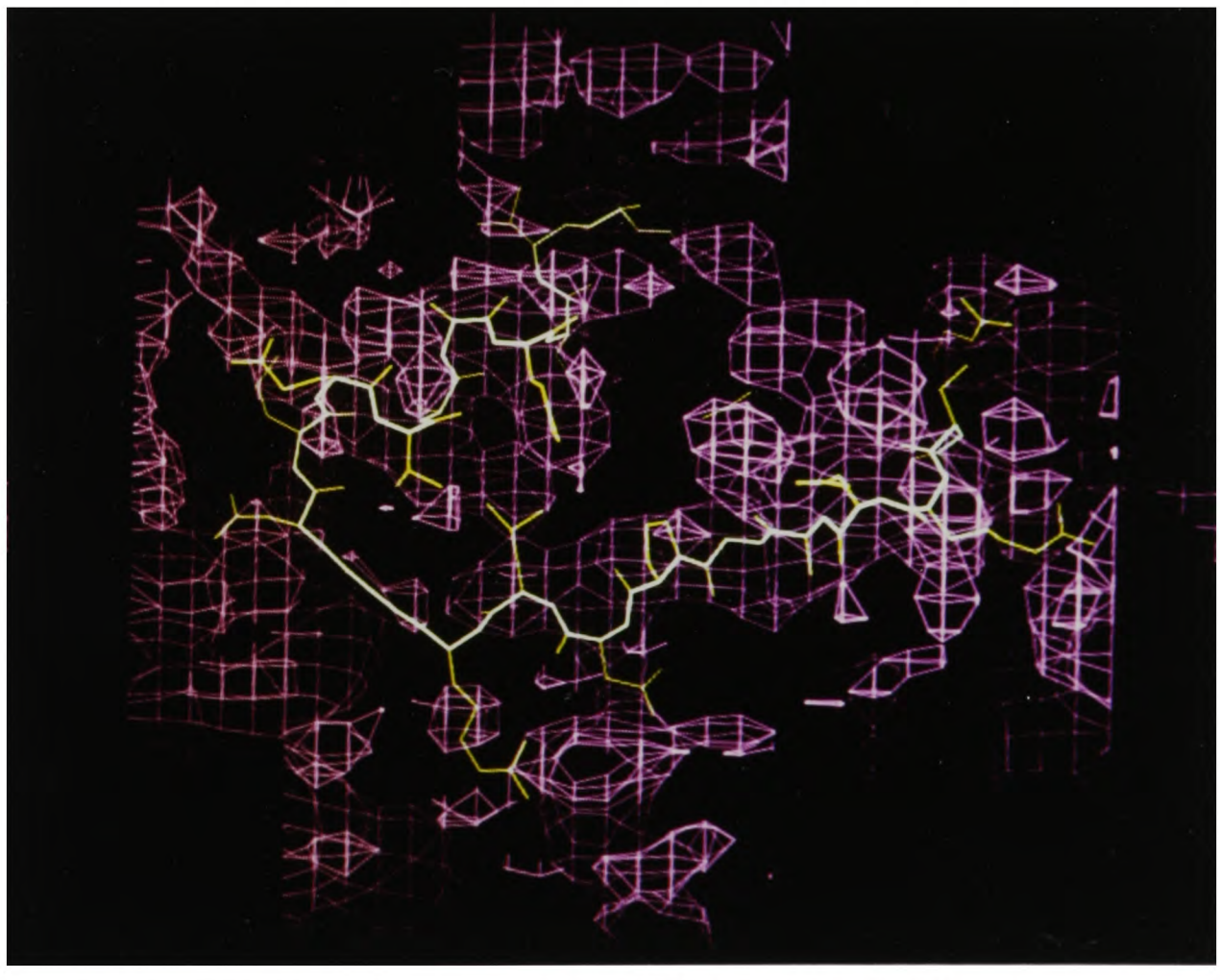
resulted in our protocol of calculating '2 – 1' maps so that after applying the constraints imposed by the observed data the new map for re–averaging is calculated with coefficients $2|F_{scaled}| - |F_{scaled}|$, α_{parent}.

The density produced by the deconvolution process is shown in Figure 4.5. The first cycle of 5–fold averaging and flattening is seen to have the greatest effect on the quality of the map. Figure 4.5A shows the density calculated in I23 using 'unfolded' amplitudes for both the parent and mutant data, the 'unfolded' mutant amplitudes scale to the 'true' amplitudes with R_C=29.9%, C=0.70. Figure 4.5B shows the map after one cycle of averaging and flattening - the estimates of the I23 amplitudes now scale to the 'true' amplitudes with R_C=15.1%, C=0.93. Applying the constraints imposed by the observed data to these estimates of the I23 amplitudes increased their agreement to the observed (432) data - from R_C=25.0% to 24.0% (C=0.73 to 0.77) - and also increased the agreement with the 'true' amplitudes - from R_C=15.1% to 11.3% (C=0.95 to 0.96). The next two cycles of deconvolution further improve the agreement between the estimates and the 'true' amplitudes to R=8.6%, C=0.97 (Figure 4.6). The process has converged by the third cycle and no further improvement is seen, although seven cycles are completed (Fig 4.5C shows the map after seven cycles). By the time convergence has occurred the I23 estimates have diverged from the observed data (Table 4.3).

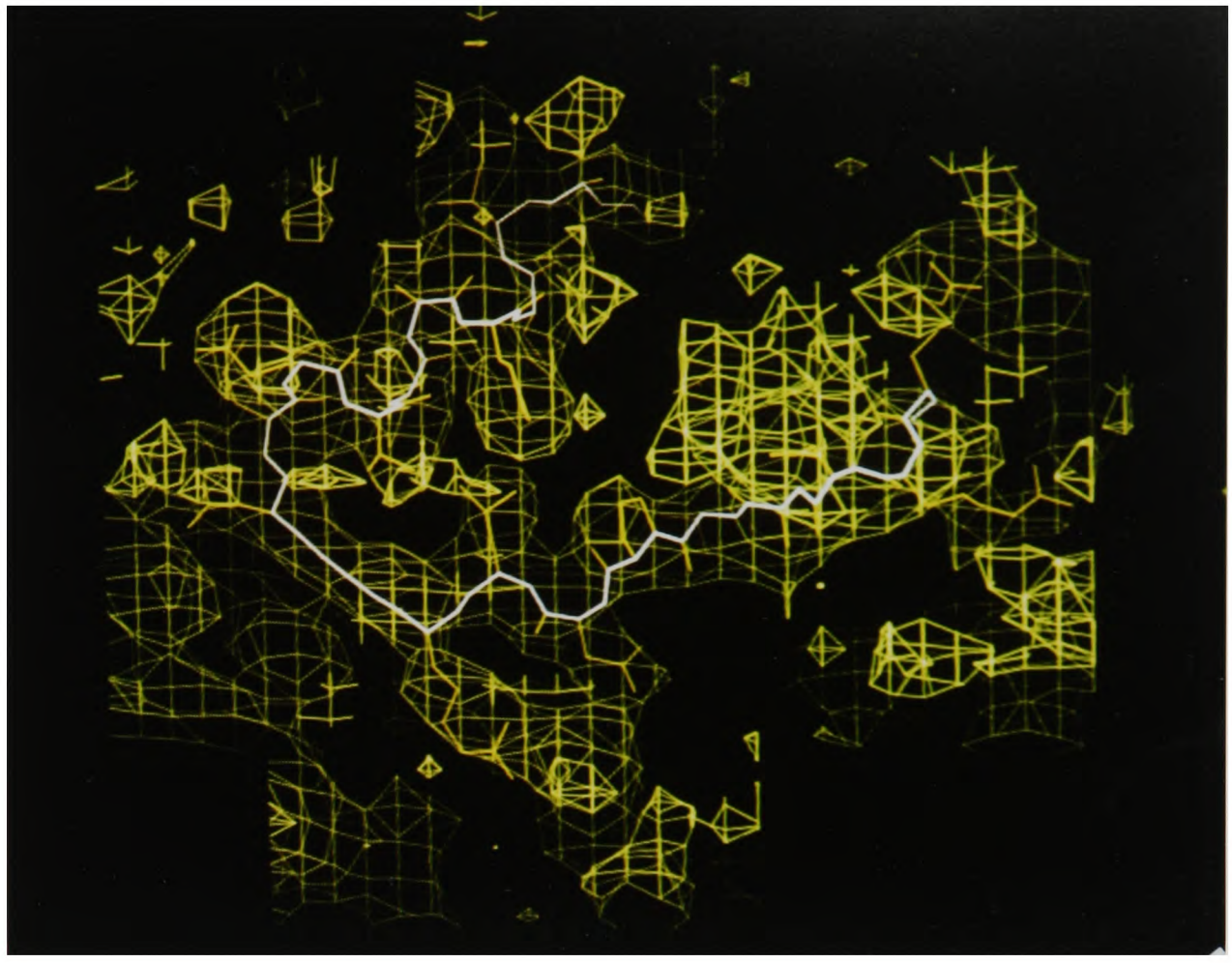
Three cycles appears to produce the maximum improvement in the amplitude estimates while still reverting to the parental phases for each map calculation. The estimates of the mutant phases produced by back–transformation of the averaged and flattened map at cycle 3 were therefore taken as the starting point for three further cycles of deconvolution using the new mutant phase estimates to calculate the maps for re–averaging. This process converged after three cycles - the mean phase change between the 5^{th} and 6^{th} cycles being less than 2°. Although this process managed to move the phases away from the parental phases (mean phase difference 15.8° at cycle 6) the phases did not become any closer to the 'true' mutant phases (mean phase difference 14.6° at cycle 3, 16.0° at cycle 6) and the agreement between the estimates



(A) 'True' Density - $2|F^{23}_{mutant}| - |F^{23}_{parent}|, \alpha_{parent}$



(B) Unaveraged Density - $2|F^{unfolded}_{mutant}| - |F^{unfolded}_{parent}|, \alpha_{parent}$



(C) Cycle 1

(D) Cycle 7

Figure 4.5: Density produced at different stages of deconvolution of Test data.

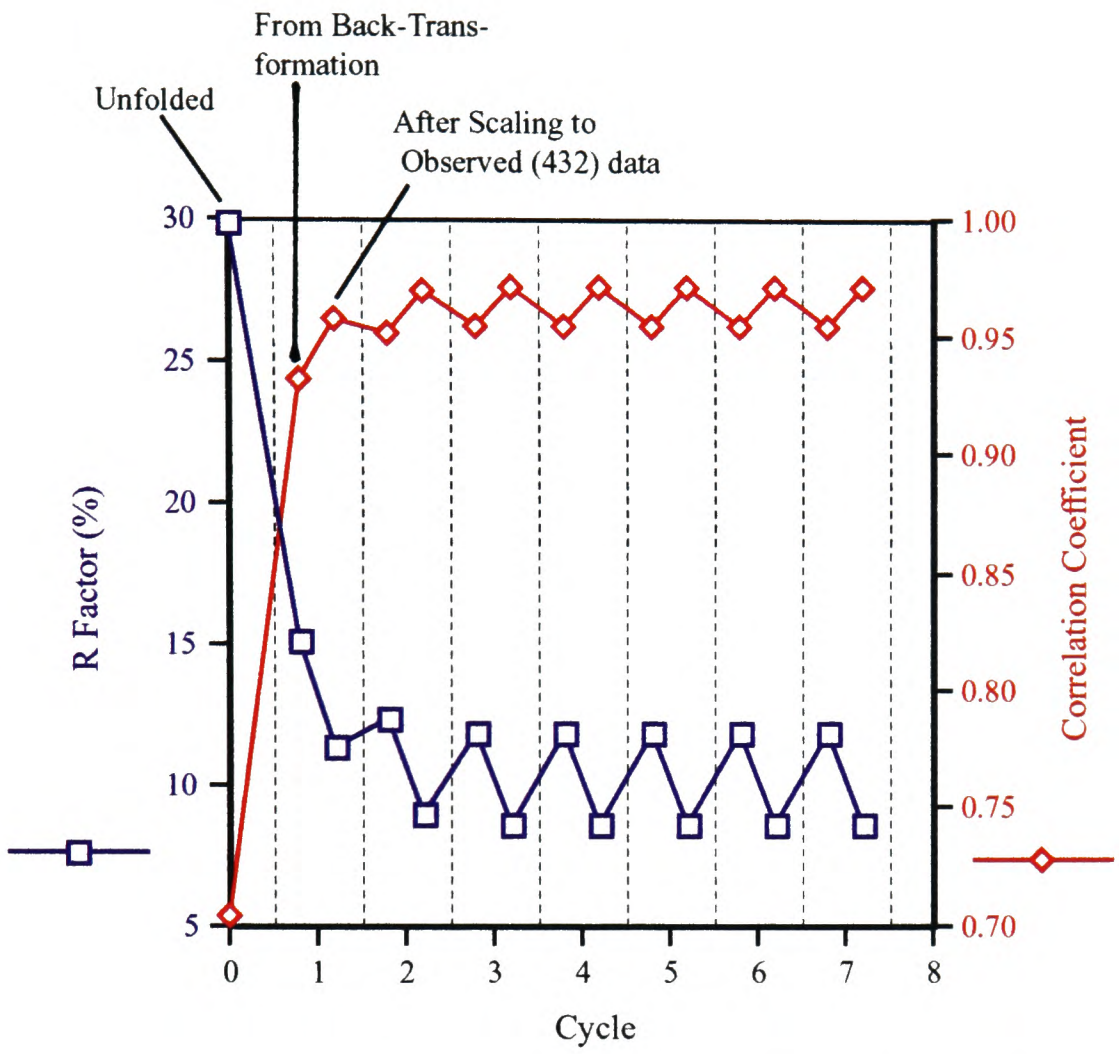


Figure 4.6: Agreement between 'true' and estimated $|F^{23}_{mutant}|$ during the deconvolution.

Cycle	R_C	C
'unfolded'	0.0	1.00
1		
After averaging	26.0	0.73
After Scaling to 432 data	24.0	0.77
2		
After averaging	28.7	0.70
After Scaling to 432 data	27.3	0.73
3		
After averaging	29.7	0.70
After Scaling to 432 data	28.4	0.72
– convergence achieved –		

Table 4.3: Agreement between $|F^{432}_{mutant}|$ and estimates of $|F^{23}_{mutant}|$ at different stages of the deconvolution process.

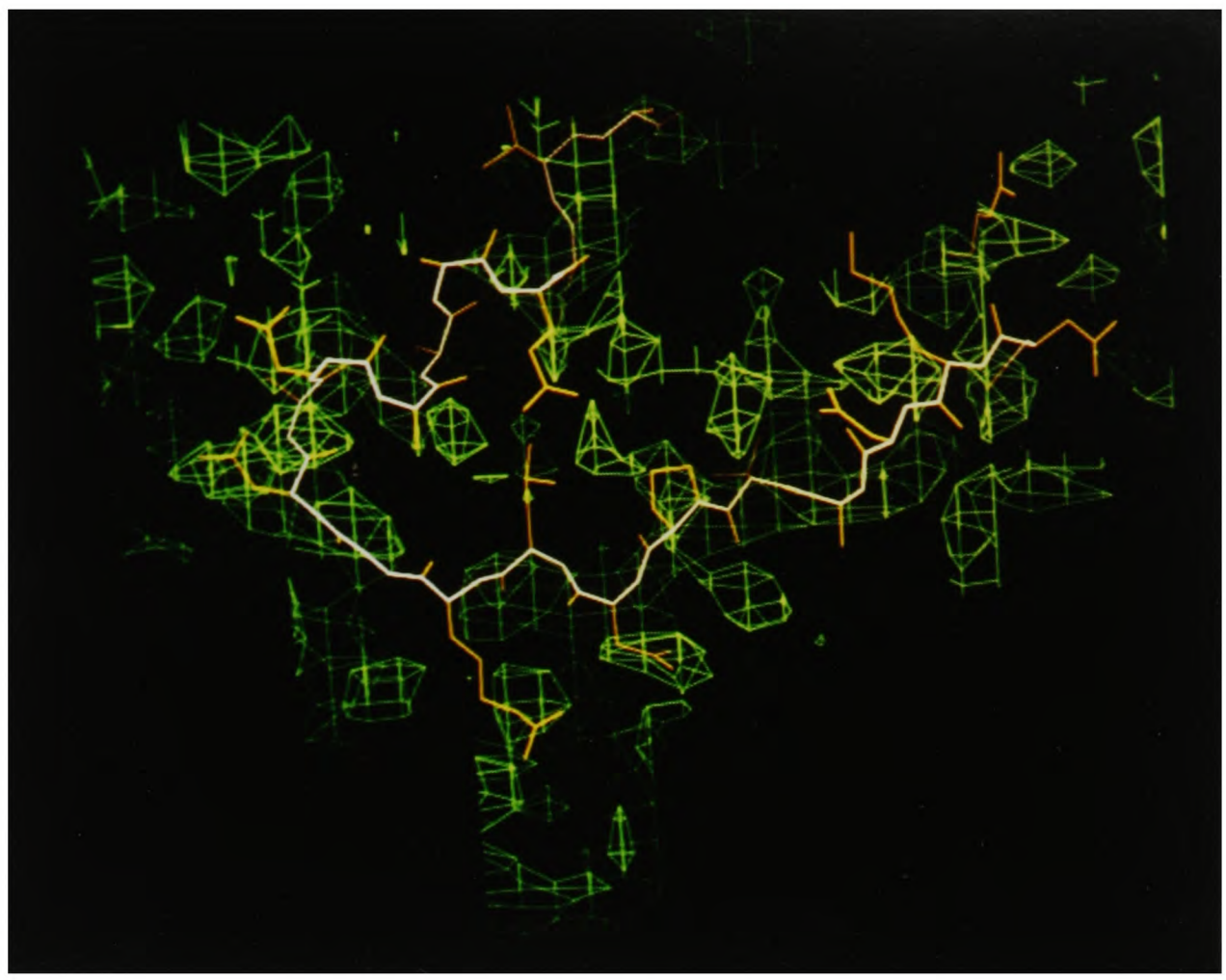
of the mutant I23 amplitudes and the true amplitudes was slightly worse then is had been before release of the phases with R_C=12.0% compared to 11.9% (C unchanged).

This test demonstrates that the deconvolution process can successfully reconstruct amplitudes, but that the phasing power (in this case) is not sufficient to also reconstruct an accurate set of phases. It is possible however, that optimization of the RNA and solvent envelope parameters to maximise the phasing power obtained by 'solvent' flattening might allow successful refinement of the phases in addition to the amplitudes.

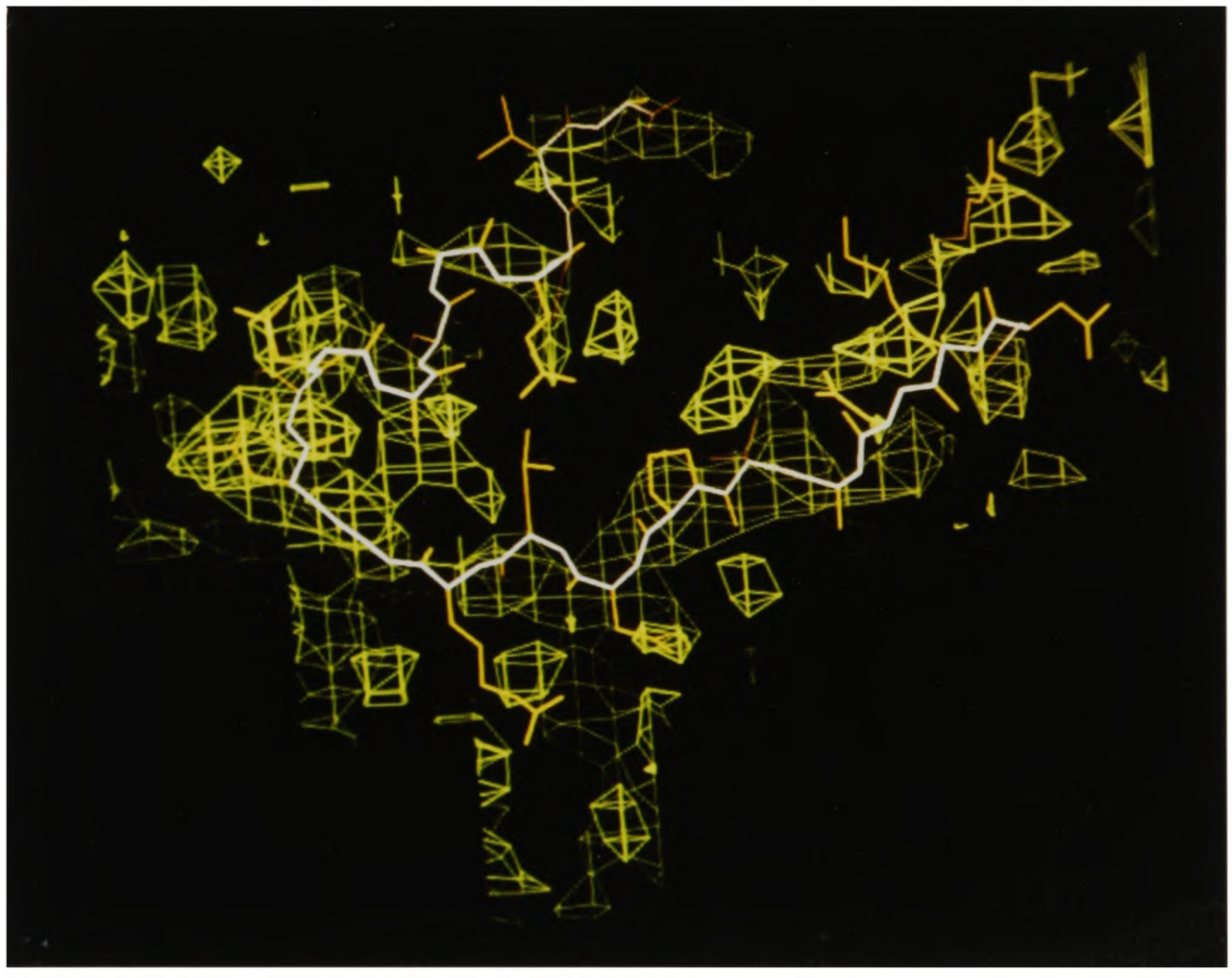
4.3.2 G67 Data

This deconvolution process was then applied to the actual, observed, G67 data. Attempts were only made using the data set derived from DTT soaked crystals since the quality of the these data is significantly better than the native data (R(I)=11% c.f. 19%) and the data set much more complete (0.90 to 3.0Å c.f. 0.76). The success of cyclic averaging is, to a large extent, determined by the completeness of the data[1], handicapping ourselves by using the less complete set seemed foolish. Three cycles of deconvolution using the observed G67-DTT data were performed. Judging the success or failure of this process is hindered by the lack of 'true' G67-DTT amplitudes to compare with the estimates obtained from the deconvolution. The maps produced are shown in Figure 4.7; (A) shows the map calculated using 'unfolded' O_1BFS and G67-DTT amplitudes, before averaging; (B) the map after one cycle of averaging; (C) the map after three cycles of deconvolution - at this point the process has converged and cycling was stopped. As for the test data the map quality is significantly improved by one averaging cycle, and then undergoes some further, slight, improvement by deconvolution. Since comparisons with 'true' G67-DTT amplitudes cannot be made R_C and C were calculated between the G67-DTT estimates and the observed O_1BFS amplitudes. After three cycles R_C=22.0% C=0.85 whereas, for the

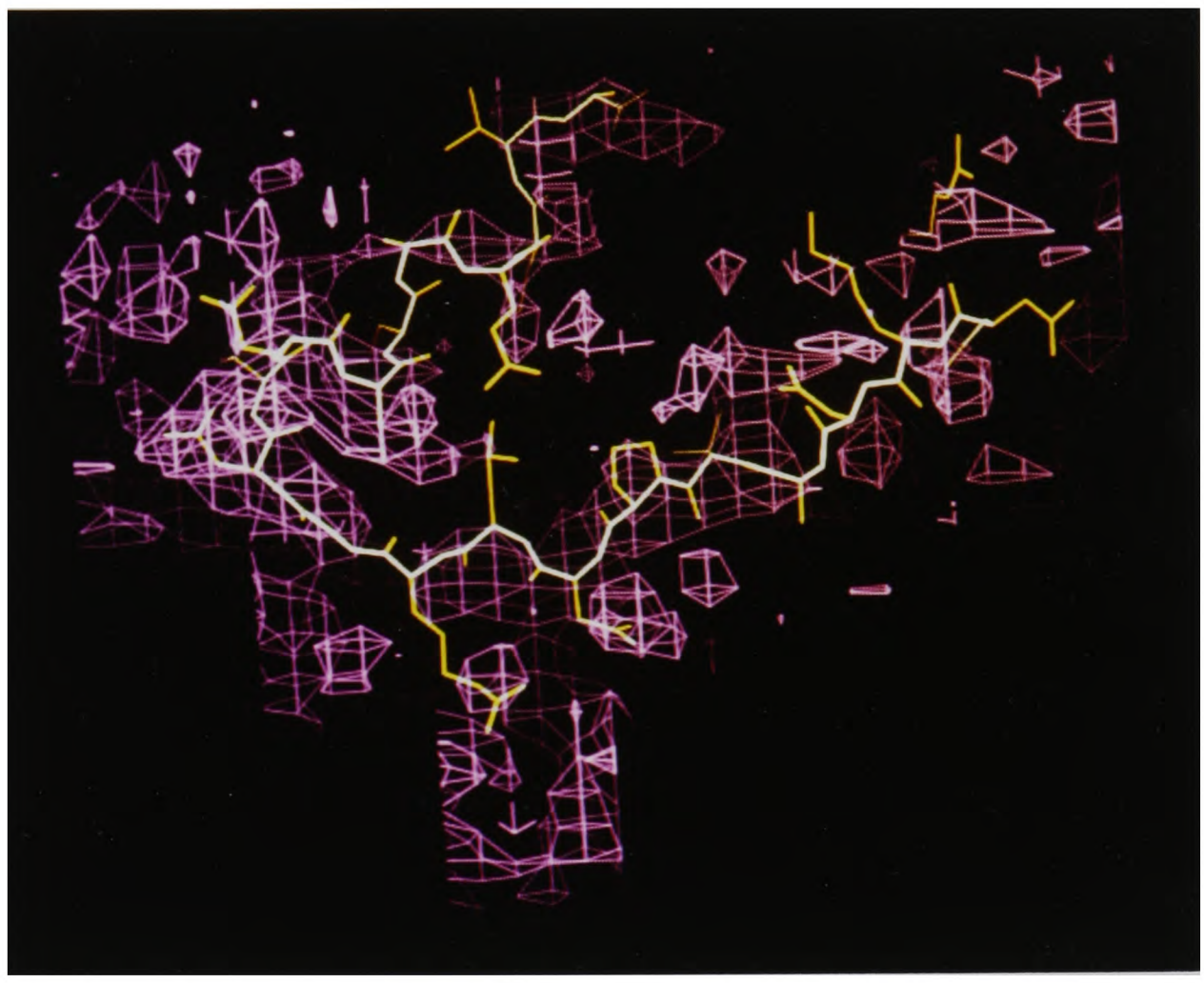
[1]Although at all stages in the deconvolution calculated data are used to fill in unobserved reflections.



(A) Unaveraged Density - $2|F^{unfolded}_{G67-DTT}| - |F^{unfolded}_{parent}|, \alpha_{parent}$



(B) Cycle 1



(C) Cycle 3

Figure 4.7: Density produced at different stages of deconvolution of G67-DTT data.

test data, the agreement between the mutant amplitude estimates and the parental amplitudes gave R_C=12.9% C=0.96. This difference reflects the larger structural differences between G67 and O_1BFS than those modelled in the test data and must also reflect the noise levels of the two observed data sets.

The deconvolution process is therefore seen to produce a rapid improvement in map quality - although the bulk of this occurs during the first round of averaging and flattening - further improvement is seen for a few cycles of deconvolution. Again, as for the test data, optimization of the envelopes might have further improved the result, as would collection of a more complete data set.

4.4 Conclusions

The method presented here works well but is only one of a variety of approaches that might have been taken to use this data. Other approaches would include use of difference coefficients combined with envelopes delimiting the areas (in real space) within which structural alterations are presumed to occur. By then flattening the vast majority of the map more stringent constraints will be imposed, however this is a less general method and requires a prior estimate of the extent of the structural changes.

The approaches of G. Bricogne (Bricogne, 1993; Bricogne, 1991) using maximisation of entropy and likelihood would allow a rigorous treatment of such effects, indeed they are already proving effective in the case of powder diffraction data where the ambiguity is significantly increased compared to our case. However the method we have developed is rather simple and fits in easily with the phase refinement procedures currently in use in many laboratories.

The power of this technique would be significantly increased in systems with a greater amount of NCS symmetry. The 5–fold averaging with which FMDVs crystallising in this form present us, is in effect, reduced to $2\frac{1}{2}$–fold by the 2–fold ambiguity introduced by the disordering. Many virus systems crystallize with much

larger amounts of NCS e.g. Mengo (Luo *et al.*, 1989) and B.E.V. (Smyth *et al.*, 1993) with 60–fold NCS. In such cases, with greater phasing power to be gained from the averaging, there is hope that such a deconvolution process could successfully deconvolute entirely overlapped reflections even in the absence of an accurate set of starting phases.

The method, as described, is not specific to the deconvolution of completely overlapped reflections produced by crystal mis–packing and provides a simple and rather general method that may be applied in other cases e.g. to the deconvolution of the wavelength overlaps that occur using the Laue method.

Chapter 5

Structure of virus C-S8cl

5.1 Introduction

Different serotypes of FMDV are antigenically very distinct - infection of an animal with a member of one serotype results in production of polyclonal sera able to recognize other members of the same serotype but showing no reactivity with FMDVs belonging to different serotypes. Despite such radically altered antigenic characters members of different serotypes of FMDV share 75-85% sequence identity in the four capsid proteins and must therefore be structurally very similar. To investigate the conformational differences underlying antigenic variation between different serotypes of FMDV we have solved the structure of a representative type C virus.

C-S8cl and O_1BFS show 82% sequence identity over the four capsid proteins with the identity in VP1 the lowest at only 70%. C serotype viruses are responsible for outbreaks of FMD in both Europe and South America, the virus we have initially studied is a member of the C_1 subtype - most commonly found in Europe (as is O_1BFS). Seed stocks of the plaque–purified virus (Mateu *et al.*, 1989) were provided by Esteban Domingo (Centro de Biología Molecular, Madrid, Spain). The parental (C_1 Santa Pau Sp/70) virus from which this clone is derived originates in an outbreak of the disease in Spain in 1970 (Sobrino *et al.*, 1989; Sobrino *et al.*, 1986). This original field isolate was then plaque–purified by dilution and plating (Sobrino *et al.*, 1983) on BHK-21 cell monolayers to produce the virus used in the structure determination.

In addition to illuminating the nature of inter-serotypic structural differences, the structure of this virus may be used to interpret specific immunological information relating to MAb–escape mutants of this virus (Hernández *et al.*, 1992; Mateu *et*

al., 1990; Mateu *et al.*, 1992) as well as a wealth of information accumulated by Esteban Domingo and colleagues, relating to antigenic variation of C viruses in the field (Domingo *et al.*, 1990; Mateu *et al.*, 1988) and evolutionary relationships between C virus sequences (Martínez *et al.*, 1992). Of special interest is the structure of a closely related virus (R100; (Diéz *et al.*, 1990)), isolated after 100 passages of a BHK cell line persistently infected with C-S8cl. persistent viral infection of the upper respiratory tract of animals which have apparently recovered from FMD is an important problem in management of the disease (Van Bekkum *et al.*, 1959). R100 contains multiple mutations throughout the genome and is antigenically distinct from C-S8cl even though it has never been subjected to any form of immune pressure (Diéz *et al.*, 1990). Several of the mutations seen are unique to R100 amongst all C viruses studied to date, one of these mutations being the substitution at position 7 of VP3 Cys→Val. In O_1BFS (Acharya *et al.*, 1989) this Cys is seen to play an important rôle in determining capsid stability by forming a network of disulphide bonds around a pore at the 5–fold axis of the virion, with symmetry related residues. This hole (which is unique to FMDV among the picornaviruses; Chapter 1, (Acharya *et al.*, 1989)) allows permeation of Cs^+ ions and planar molecules such as proflavin, giving FMDV an unusually high (for picornaviruses) buoyant density in CsCl. The disulphide bonds formed between symmetry related copies of VP3 effectively limit the diameter of the hole and determine the virus permeability. R100, despite lacking this Cys, is unexpectedly (in light of the rôle of this residue in linking trimers around the 5–fold) still a viable virus, although it is somewhat less stable than the parental C-S8cl. R100 also has a higher than usual CsCl buoyant density confirming an alteration of the structure around the 5-fold due to loss of the ability to form these disulphides. Knowledge of the structure of the parental virus should aid interpretation of the biochemical and eventually structural data pertaining to this mutant.

Crystals of R100 have been grown, but none have yielded useful diffraction - the majority of the crystals being less than 0.15mm in all dimensions. Trials varying the crystallisation conditions are underway to attempt to improve the crystal quality.

C-S8cl	
Date	Nov 1991
Station	9.6 Daresbury
Media	CEA Reflex 25 (film)
λ	0.89Å
Oscillation Range	0.5°
Crystal to film	185.6 mm
Exposure	12.5–15 min
No. of Crystals Examined	52
No. of Useful Film Packs	35
Precipitant used to grow crystals	9.5–11.5% $(NH_4)_2SO_4$
Crystal Morphology	Rhombic Dodecahedra
Crystal Size	$\approx$ 0.3 mm in all dimensions
Space Group	I23
Cell	a=b=c=347.6Å $\alpha = \beta = \gamma$=90.0°

Table 5.1: C-S8cl - Data collection and crystal characteristics.

C-S8cl lacks the Cys residues involved in the disulphide bond (between residues 130 of VP1 and 134 of VP2) known to destabilize the VP1 G–H loop (Acharya *et al.*, 1989; Logan *et al.*, 1993) (Chapters 1 and 3). We therefore also hoped that the structure of this virus might reveal a conformation for this elusive loop.

5.2 Data Collection.

C-S8cl was grown and crystallised by Robin Abu–Ghazaleh at the IAH Pirbright by variations of our standard protocols (Curry *et al.*, 1992). The virus was crystallised by vapour diffusion from 9.5–11.5% saturated $(NH_4)_2SO_4$ in the presence of 10mM dithiothreitol. The crystals were of the same morphology (rhombic dodecahedra) and similar dimensions ($0.3 \times 0.3 \times 0.3$ mm^3) to those used in previous structure determinations of O_1 viruses (Acharya *et al.*, 1989) (see Chapter 3). The crystals were mounted at Pirbright in quartz capillary tubes and transported to the SERC

Synchrotron Radiation Source, Daresbury, in accordance with agreed MAFF disease security protocols (see Chapter 2).

Crystals of C-S8cl were first taken to the synchrotron in October 1991. Several of these crystals were larger than those previously obtained for O_1 viruses (in excess of 0.4 mm in the longest dimension) however no diffraction was seen. It was noted that the crystals had degraded between crystal mounting and examination at the synchrotron (becoming opaque), an age dependent phenomenon which was later found to be characteristic of these crystals, occurring when the crystals exceed $\approx$ 10 days old. Further crystals were produced in November and taken to the synchrotron within 7–10 days of the crystallization experiments being set up.

The data collection from these younger crystals was much more successful. Data were collected in November 1991 on station 9.6 using radiation with λ=0.89Å. At this time data collection used photographic film (as for the O_1BFS structure determination; (Acharya *et al.*, 1989)) - with each film pack containing 3 CEA Reflex 25 films labelled A, B and C, B and C being separated by an extra, previously exposed film to increase the dynamic range of the pack. A summary of the data collection can be seen in Table 5.1. The crystals again proved to be unstable, several having degraded to the opaque form during transit (opaque crystals never diffracted) and with those giving useful diffraction having a life–time in the beam which only rarely allowed two exposures from the same crystal. The resolution to which the crystals diffract is also limited being 3.5Å (compared to better than 2.8Å for many of the O_1 crystals). 3.5Å was therefore taken as the resolution limit for the analysis.

5.3 Data Processing

By–eye examination of the films shows C-S8cl to have crystallised in the same space group, I23, as crystals of O_1BFS. Initial attempts at auto–indexing indicated that the cell dimensions are slightly expanded, estimated as a=348Å compared to a=345Å for the O_1 viruses. Data processing followed our standard route (Acharya *et al.*,

Crystal	AUTO		MOSCO	ABSCALE		REINDLCF	
	SDXY	SDPHI	I/σI on A film	$R_{prof}(I)$	$R_{int}(I)$	UN	SW
4001	1.17	0.10	0.9	6.7	10.3	**0.50**	0.16
4003	1.45	0.22	1.4	6.3	8.2	0.07	**0.49**
4004	"	"	1.4	5.9	8.2	0.05	**0.56**
4012	1.39	0.10	1.1	6.0	8.5	**0.44**	0.06
4013	"	"	1.4	6.8	9.5	**0.60**	0.25
4016	1.18	0.12	1.4	6.1	8.1	0.12	**0.65**
4018	1.86	0.11	3.5	6.4	7.2	0.09	**0.54**
4021	1.28	0.14	8.0	5.6	6.2	0.05	**0.67**
4024	1.52	0.10	3.2	6.2	7.0	0.09	**0.57**
4025	"	"	3.0	6.1	7.2	0.09	**0.70**
4028	1.28	0.10	1.3	6.6	9.7	**0.55**	0.08
4029	1.36	0.10	3.0	6.5	8.0	**0.42**	0.06
4030	1.59	0.11	2.2	7.2	8.7	**0.73**	0.12
4032[a]	1.13	0.10	1.8	-	-	**0.43**	0.07
4035	1.43	0.10	1.6	6.3	8.7	**0.58**	0.16
4036	1.53	0.11	1.6	6.8	8.4	-0.05	**0.40**
4037	1.42	0.10	1.7	6.1	7.8	0.10	**0.69**
4038	1.27	0.10	1.4	9.5	10.3	-0.07	**0.44**
4039[a]	1.19	0.11	2.1	-	-	**0.50**	0.07
4040	1.57	0.10	1.1	7.0	9.4	**0.53**	0.13
4042	1.81	0.11	2.6	7.6	9.4	**0.60**	0.00
4044	1.91	0.10	1.8	6.9	8.8	**0.38**	0.02
4045	0.93	0.12	3.2	6.2	8.4	**0.83**	-0.05
4046	1.58	0.11	0.9	7.0	9.9	**0.59**	0.24
4047	1.83	0.10	1.5	6.8	9.1	0.04	**0.60**
4048	1.45	0.10	1.0	6.8	9.9	0.01	**0.51**
4050	1.80	0.11	1.7	6.8	8.5	0.01	**0.57**
4051	1.25	0.11	2.0	7.5	11.0	**0.65**	-0.67
4055	1.78	0.14	8.9	4.7	5.4	0.13	**0.59**
4056	0.95	0.14	2.2	7.5	10.1	0.10	**0.48**
4057	1.88	0.12	2.1	7.8	9.0	0.10	**0.69**
4058	1.65	0.10	1.4	7.2	9.1	**0.42**	0.08
4059[a]	1.95	0.12	3.4	-	-	**0.35**	0.27
3025	1.36	0.11	1.3	8.8	14.6	**0.69**	0.03
3066	1.67	0.14	1.3	8.6	11.7	-0.02	**0.47**

a - B and C films not used from these packs.

Table 5.2: Indexing and integration statistics for C-S8cl

1989) (Chapter 2). Thirty-two auto-indexing solutions were required for the 35 useful film packs (Table 5.2) - three crystals yielded two useful exposures each, hence only one indexing solution was required for each pair of images. The indexing hand was determined by comparison with the standard O_1BFS reference set (see Chapter 2). This gave a lower correlation in the correct hand than is usually seen (due to the non-isomorphism and many differences between the viruses) but still clearly indicated the distinct 'handedness' of the data - characteristic of I23. Data processing statistics for all packs are shown in Table 5.2. The merging R(I)[1] for films from the same film pack is typically 7% whilst initial scaling of the packs together gave R(I) for the unswapped data of 18.6% and for the swapped data of 22.2%.

Postrefinement (Table 5.3), which decreased the mosaicity from an initial overestimate of 0.2° to a mean of 0.09°, improved the agreement between the data. This was followed by reclassification of all partials greater than 75% recorded and with $I > -2\sigma$ as fully recorded reflections using Massagelcf. After swapping the hand of the swapped data set R(I) is 18.0% for all data, 64849 independent reflections, 74% complete to 3.5Å (50% complete for data with $I > 3\sigma I$) (Figure 5.1).

5.4 Accurate determination of the cell edge.

When collecting data using synchrotron radiation it is often difficult to be certain of the wavelength of the radiation used. For this reason we decided to refine our estimate of the unit cell dimensions of C-S8cl with reference to the cell dimensions of O_1BFS - taken as exactly 345.0Å . Using XPLOR (Brünger, 1992) the coordinates for O_1BFS were placed in cells with the edges varying from 347.0Å to 349.0Å and amplitudes calculated in each cell compared to the observed C-S8cl data. Since the cell is cubic this greatly simplified the search procedure - a simple one–dimensional

[1] $R(I)=\frac{\sum_h \sum_i |(I_h - I_{hi})|}{\sum_h \sum_i I_{hi}} \times 100$

Pack	< I >	R_{merge}(I)		Post–Refinement						
	before PR	before PR	after PR	$\Delta\phi$y	$\Delta\phi$z	$\Delta\lambda$ (Å $\times 10^5$)	ϵ	F→P	P→F	total F
Unswapped	110.56	18.6	18.9					512	14631	49895
3025	133.18	21.9	23.3	0.01	-0.02	545	0.147	106	19	512
4001	114.99	33.2	33.7	0.01	0.01	136	0.057	4	927	2872
4012	111.54	15.1	16.3	0.00	-0.01	-99	0.021	0	1554	4742
4013	129.39	25.9	26.4	-0.02	0.01	-110	0.078	95	588	2025
4028	104.93	21.3	22.7	-0.01	-0.01	-22	0.070	0	650	2738
4029	97.99	4.8	5.0	0.00	-0.01	-94	0.047	0	1074	3993
4030	134.50	15.5	15.9	0.01	0.01	-25	0.18	233	12	808
4032	128.62	24.3	26.0	0.01	0.02	-288	0.085	0	481	2024
4035	92.19	10.1	10.6	0.00	0.00	12	0.017	0	2238	5824
4039	134.18	26.4	25.9	0.00	0.04	99	0.176	52	7	629
4040	98.69	24.2	25.2	0.00	0.00	-288	0.027	0	1569	4345
4042	126.28	12.7	12.7	0.00	0.00	-136	0.048	0	434	1463
4044	96.67	8.5	9.6	-0.01	-0.03	-16	0.024	0	1614	5122
4045	131.46	9.8	10.3	-0.01	-0.01	-359	0.085	3	166	726
4046	93.29	32.9	31.4	-0.01	-0.01	48	0.051	0	1096	3829
4051	166.09	17.3	16.3	0.00	-0.01	-266	0.112	19	30	382
4058	100.58	15.4	16.5	-0.01	0.02	-148	0.057	0	1207	4022
4059	116.61	14.4	15.0	0.01	-0.01	45	0.054	0	964	3794
Swapped	185.14	22.2	18.8					6364	12625	49047
3066	169.89	15.8	15.5	0.00	-0.01	-425	0.048	46	1413	4787
4003	159.65	29.9	29.0	0.01	0.04	-272	0.206	648	126	2257
4004	184.23	30.4	25.7	0.01	0.05	-387	0.212	627	235	1648
4016	190.24	29.2	29.2	-0.02	0.04	-352	0.195	492	104	1925
4018	175.10	4.6	3.9	0.01	-0.01	-121	0.008	0	2407	6201
4021	305.34	18.1	16.4	0.01	-0.01	-134	0.265	248	0	130
4024	189.34	5.1	4.4	0.01	0.00	20	-0.019	0	2991	6821
4025	179.17	5.4	4.9	-0.01	-0.01	-20	0.009	0	2303	6118
4036	191.73	30.2	27.7	0.00	-0.01	-85	0.107	67	262	2251
4037	208.63	20.6	20.1	0.00	0.00	159	0.081	24	586	2682
4038	173.83	32.5	33.1	-0.01	-0.02	-299	0.167	700	0	2149
4047	181.30	25.6	21.4	0.00	-0.01	-133	0.007	0	1410	3755
4048	184.27	46.3	39.1	0.02	-0.04	-233	0.095	172	181	2120
4050	177.84	16.1	16.5	0.01	0.02	-225	0.115	132	154	3172
4055	243.01	7.0	3.9	-0.01	0.00	-53	0.128	76	93	651
4056	177.48	18.3	16.0	0.00	0.00	-45	0.296	1282	0	693
4057	194.83	33.9	41.7	0.01	-0.08	-17	0.181	1832	0	1637

Input Parameters: mosaicity 0.9° ; $\lambda = 0.89$Å ; beam divergence - 0.1° horizontal / 0.05° vertical; spectral dispersion 0.0012 / correlated component 0.00025. **Key:** F - fully recorded; P - partially recorded; ϵ - beam divergence $\equiv$ mosaicity.

Table 5.3: Summary of postrefinement of C-S8c1

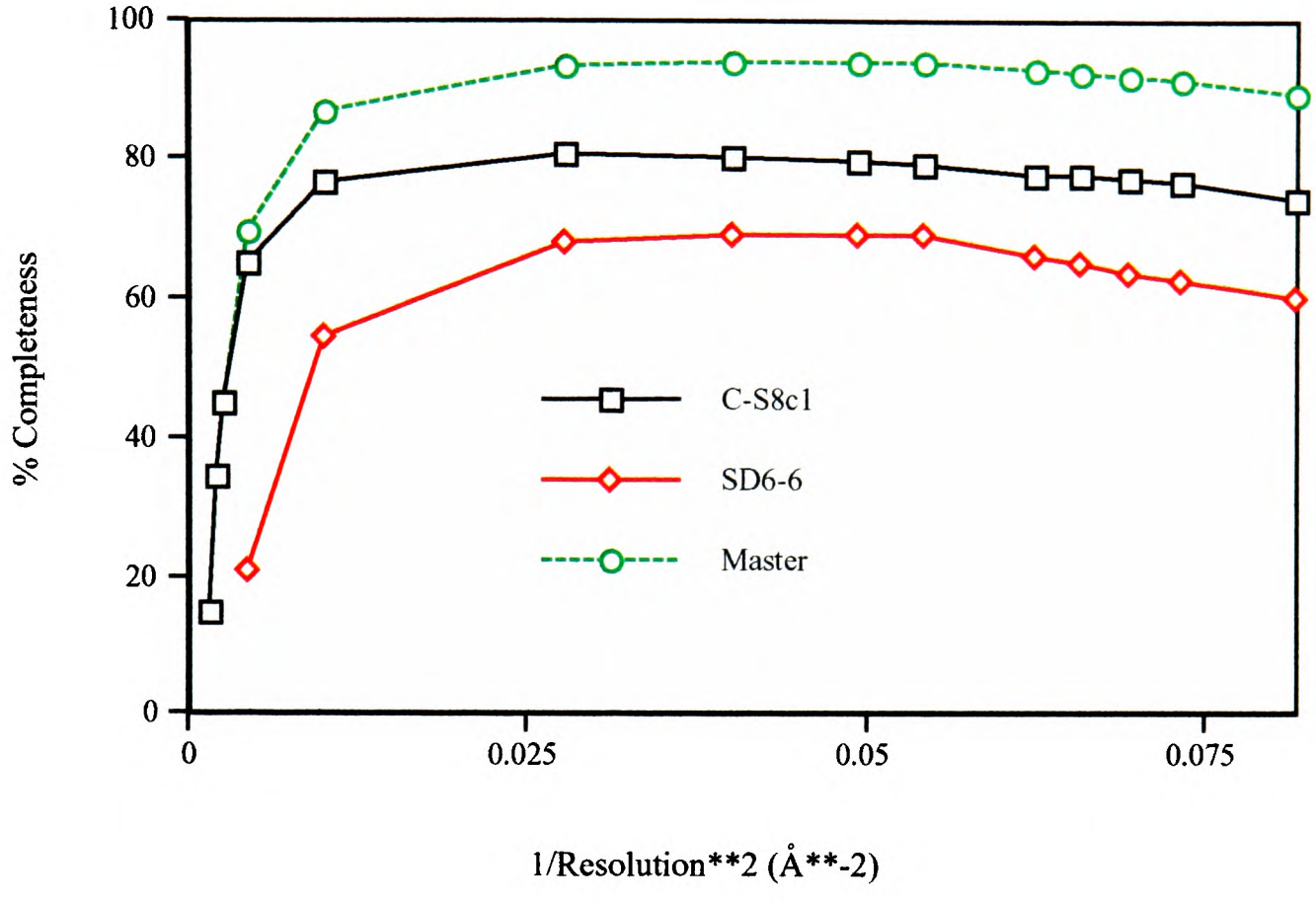


Figure 5.1: Completeness of C-S8c1 data Sets.

	Before 'Massagelcf'		After 'Massagelcf'	
	R(I) (%)	No. of Ind. Refl.	R(I) (%)	No. of Ind. Refl.
Unswapped	18.9	32895	19.3	40935
Swapped	18.8	32014	20.5	44729
Combined	–	–	18.0	64849

Table 5.4: Final data processing statistics for C-S8c1

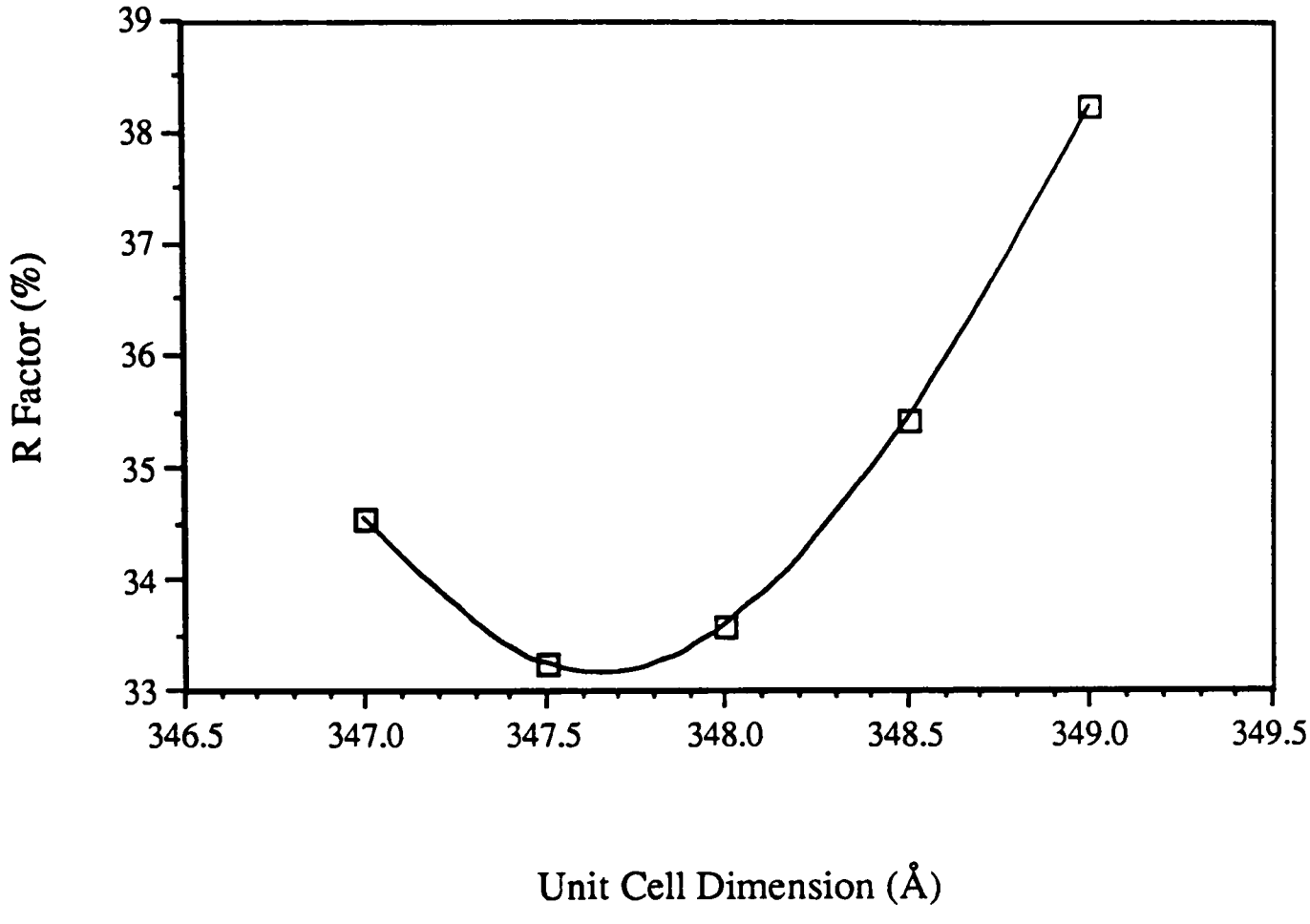


Figure 5.2: R-factor for O_1BFS model against C-S8c1 data in different unit cells.

search for the cell length is all that is required. R_C were calculated[2] for data in the range 8.0–3.5Å , the low resolution cutoff of 8.0Å used so that the contribution to the scattering made by the disordered solvent need not be considered. Plotting the results of a simple grid-search revealed a clear minimum value of R_C at a=347.6Å (Figure 5.2). This value is therefore taken as the true cell dimension for C-S8cl (n.b. it must always be remembered that this is not an absolute value, it is simply a value relative to the cell for O_1BFS; a=345.0Å).

[2] $R_C = \frac{\sum_h |(|F_{h,obs}| - |F_{h,calc}|)|}{\sum_h |F_{h,obs}|} \times 100$

5.5 Calculation of initial phases.

Using the O_1BFS model placed in the C-S8cl cell enabled calculation of initial phases for C-S8cl. To enable inclusion of the low resolution data a bulk solvent correction was applied using parameters determined by Derek Logan (Logan, 1991) during refinement of O_1BFS. Before applying this correction the O_1BFS model in the C-S8cl cell gave R_C=0.34 over the resolution range 50.0–3.5Å . A bulk solvent correction with uniform electron density 1.0 e^- $Å^{-3}$ (SOLDEN) and no margin around the protein atoms (i.e. SOLRAD=0.0) was applied. F_{calc} s calculated with the F_{part} bulk solvent contribution weighted by k_p=0.41 and B_p=110 where:

$$F_{calc} = F_{calc} + k_p F_{part} e^{-B_p sin^2\theta/\lambda^2}$$

scaled to the F_{obs} (using Shell–Scale) with R_C=0.31 and C=0.73[3] over all the data.

5.6 Refinement of phases.

Phases calculated in this way are obviously based on the known structure - we therefore decided to use a procedure based on NCS symmetry averaging (Chapter 2) and solvent flattening to gain improved estimates of the true C-S8cl phases. The procedure is based on that previously described (Fry *et al.*, 1993) for phase extension and refinement during the determination of the structure of O_1BFS. As used here no phase extension is required as phases to the limit of the observed data can be estimated from the O_1BFS model[4], these simply require refinement towards the true C-S8cl phases. The programs used are based on those of G. Bricogne (Bricogne, 1976) which are largely designed around the limited amount of real computer memory avaliable in the 1970s. This meant that an entire map could not be held in real memory at one time therefore random access to the map pixels was impossible. To overcome this

[3] $C=\frac{\sum_h(\bar{F}_{obs}-|F_{h,obs}|)\cdot(\bar{F}_{calc}-|F_{h,calc}|)}{[\sum_h(\bar{F}_{obs}-|F_{h,obs}|)^2\cdot\sum_h(\bar{F}_{calc}-|F_{h,calc}|)^2]^{\frac{1}{2}}}$

[4] R_C for shell 3.55–3.50Å is 40%

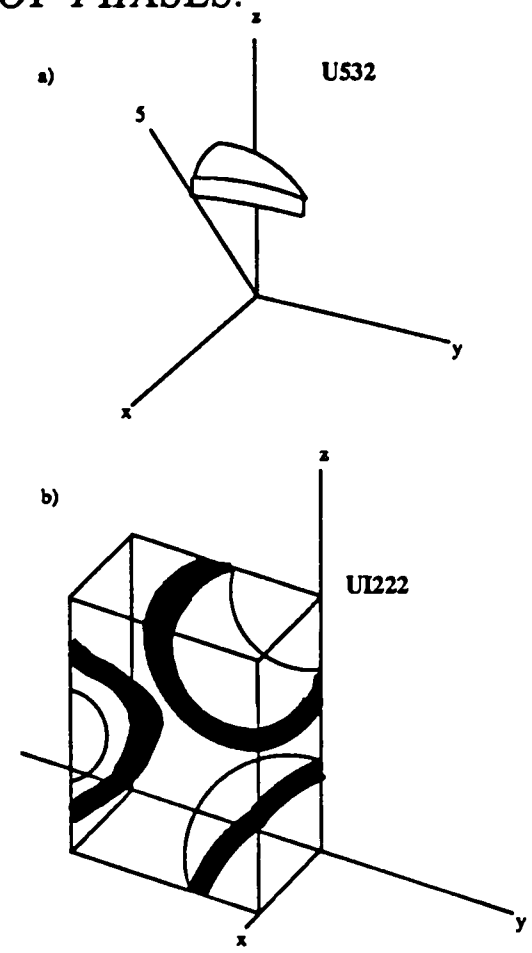


Figure 5.3: Different portions of map calculated during averaging.

limitation a 'double–sorting' technique was devised. Records are generated (GENERATE) for each integral grid point in the domain containing the averaged map which point to pixels in a more finely sampled map which will be used to interpolate and average the symmetry related densities. To allow sequential access to the fine map the records are sorted according to its sectioning access (BSORT). After interpolation the averaged density records are sorted on the sectioning axis of the new map. The FFT algorithm, as programmed, cannot handle the diagonal three–fold axis of I23, the FFTs are therefore performed in space group I222 (the different map portions calculated are shown in Figure 5.3). The application of these programs to 5–fold averaging of FMDV has been discussed in detail elsewhere (Fry *et al.*, 1993; Fry, 1991). I will therefore give only a brief description of the protocol used here.

The procedure used in the envelope determination is summarised in Figure 5.4

A $2|F_{obs}| - |F_{calc}|, \alpha_{calc}$ map calculated between 30 and 3.48Å is used as the starting point for the envelope determination and averaging. This map is calculated on two FFT grids (see above), the coarse grid (256) sampling the transform at $\approx 2/5$

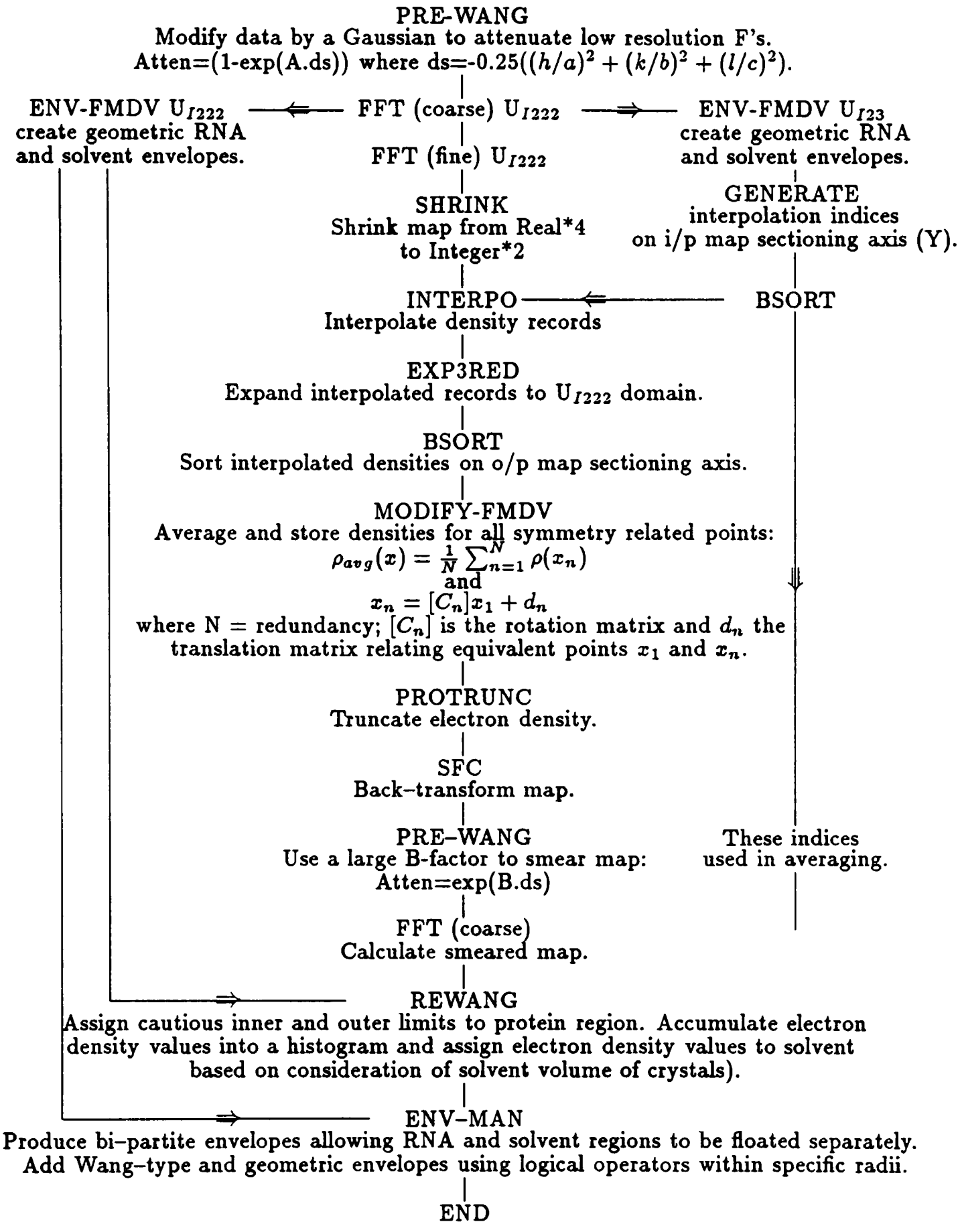


Figure 5.4: Procedure used to determine detailed RNA and solvent envelopes.

of the resolution limit and the fine grid (512) at $\approx 1/5$ of the resolution limit. The data are initially modified by an attenuating factor to help reduce the effects of series termination errors and to eliminate distortion of the envelope by any low resolution structure within the RNA region. A B factor of 350Å^2 (constant A in Figure 5.4) has been previously determined as appropriate (Fry, 1991). The map is then averaged over the 5 NCS related subunits and the amplitudes produced by back–transformation of this map further attenuated by a B Factor of 1000Å^2 (constant B in Figure 5.4). These attenuated data are used to calculate a 'smeared' map which is used to produce the envelope.

The electron density values within the smeared map are accumulated into a histogram and, based on an estimate of the solvent content, those grid points with electron density values below a certain value are marked as solvent. The envelope produced is defined in a logical format and a bi–partite envelope may be constructed (using logical operators) between this 'Wang–type' (Wang, 1985) envelope and geometric envelopes defining the RNA and solvent regions. This allows these regions to be separately flattened to different values (the RNA envelope will have a higher mean density) during the averaging. Envelopes are periodically re–determined during the course of the averaging as phase accuracy and hence map quality improves.

The procedure used for the cyclic averaging is summarised in Fig 5.5. In each cycle, after averaging over the 5 NCS related subunits, the bi–partite envelope is used to separately flatten the RNA and solvent regions to their mean value. The averaged electron density is back–transformed and the resultant calculated phases and amplitudes used to calculate a new $2 \mid F_{obs} \mid - \mid F_{averaged} \mid , \alpha_{averaged}$ map which may then be re–averaged. Cycling of averaging and flattening continues until convergence of the phases occurs. During each cycle calculated structure factors are included to fill in gaps in the sampling of the transform to prevent propagation of ripples in the FFT. The weighting scheme used throughout varies involving initial use of Rayment weights (Rayment, 1983) where reflections are weighted according to the error in the phases based on the fit of the F_{calc} to F_{obs};

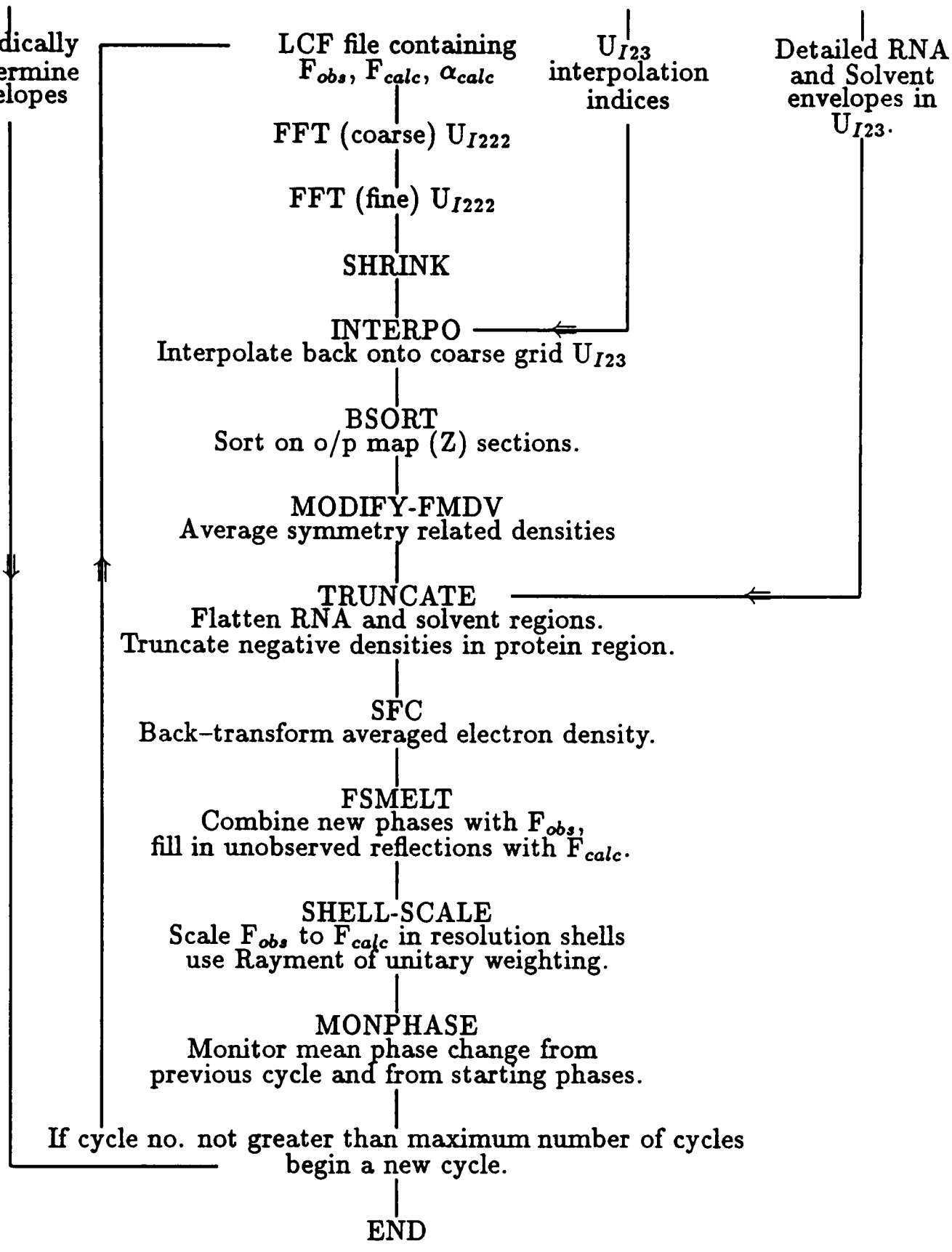


Figure 5.5: Procedure used for cyclic averaging of C-S8c1

$$W = exp(\frac{-||F_{calc}|-|F_{obs}||}{|F_{obs}|})$$

After a number of cycles using this weighting scheme, unitary weighting is used for a few cycles to allow the phases to settle (Rayment, 1983). $F_{averaged}$ were scaled to F_{obs} in resolution shells with several rejection criteria[5] set to reject outliers so preventing bias from spurious measurements creeping in (the total number of reflections rejected was less than 0.3% in each cycle) and the agreement between $|F_{obs}|$ and $|F_{calc}|$ monitored by calculation of R_C and C. The RNA and solvent envelopes were periodically recalculated as improved phase estimates were produced.

The O_1BFS model in the C-S8c1 cell was used to give starting phase estimates. Solvent and RNA Wang–type masks were generated with an inner radial boundary of 72 pixels (97Å) and an outer radial boundary of 114 pixels (154Å). These were taken as the inner and outer limits within which the protein might be expected to occur. One cycle of 5–fold averaging within the protein area defined by these masks resulted in an improvement of the agreement between the amplitudes produced by back–transforming the averaged electron density and the observed data from R_C = 31% to R_C = 23% (C=0.73 to C=0.86), with a mean phase change over the resolution range 30–3.48Å of 28°. Calculated structure factors were used to fill in the unobserved data and new RNA and solvent masks calculated using this complete data and new phase estimates. These envelopes used a tighter inner boundary (78 pixels/ 105Å) and loosened the external boundary to prevent accidental truncation of density for surface features such as the VP1 G–H loop (118 pixels/159Å). These boundaries correspond to $\approx$ 35% of the cell volume occupied by disordered solvent, 26% by RNA. These envelopes were much smoother than the original masks. Cyclic averaging using these envelopes proceeded for 20 cycles – the first ten utilising Rayment weights (until the mean phase change per cycle dropped below 2°), the next 10 cycles using a unitary weighting scheme. After these 20 cycles the averaging appeared to have converged with the mean phase change per cycle only 0.77°, and gave a reasonable solution; R_C

[5](i) reflections where $|F_{obs}| - |F_{calc}| > 6\sigma |F_{obs}|$ rejected. (ii) reflections where $|F_{obs}| - |F_{calc}| > 0.8\bar{F}_{obs}$ where $\bar{F}_{obs}$ is the mean value of F_{obs} for that resolution shell.

= 15.32%, C = 0.9266. This compares well with the agreement between calculated and observed data seen during phase extension by NCS averaging of FMDV O_1BFS (Fry *et al.*, 1993; Fry, 1991) which gave R_C = 16.28% C = 0.92 at 3.5Å . The mean phase change from the starting phases was 31°. New envelopes were determined and the averaging allowed to continue for a further 20 cycles (with a mix of Rayment and unitary weights), but no further improvement in phase accuracy was seen. The mean phase change per cycle gradually increased to 1.8° as the averaging diverged and the agreement between $|F_{obs}|$ and $|F_{calc}|$ decreased to R_C = 16.3% C = 0.9151.

At this stage we noted that several classes of reflections appeared to be absent from the LCF file output by CCP4 program SFC for instance the *0 0 l*. Inspection of the code showed that the reflections were incorrectly reduced to the I_{23} unique portion, resulting in some required reflection classes being rejected. Correction of this fault resulted in an increase of 1946 ($\approx$ 2%) in the number of reflections written out by SFC when back–transforming the averaged electron density in the resolution range 30 to 3.48Å . This error meant that, for the previous averaging cycles, several low resolution reflections were absent – possibly leading to ripples in the FFT. The phases produced by cycle 20 of the averaging were therefore taken as the start point for further averaging. Two parallel attempts were made:

1. The same geometric envelope parameters were used as in the previous cycles i.e. inner radial boundary 78 pixels (105Å), outer radial boundary 118 pixels (159Å).

2. A looser envelope was defined – there was evidence of ordered density on the internal surface of the virus within the area defined as disordered RNA by the previous mask. The new masks used an inner radius of 70 pixels (95Å) and an outer radius 120 pixels (162Å).

10 cycles of averaging using each of these envelopes (5 cycles Rayment weighting / 5 cycles unitary weighting) and the 'new' SFC gave a clear result. Table 5.5 shows the improved solution produced by averaging within the looser envelope. The quality

	R factor	C	Mean Phase change	
			previous cycle	starting phases
Tighter Envelope	16.37%	0.92	1.4°	-
Looser Envelope	13.33%	0.94	1.9°	39.0°

Table 5.5: C-S8cl: Results of final NCS averaging.

of this solution compares well with the result obtained after 719 cycles of averaging for O_1BFS at 2.8Å where the agreement between F_{obs} and F_{calc} gave a final R_C of 14.89% and C of 0.94.

5.7 Model Building

An initial model was built into the $2|F_{obs}|-|F_{av}|$, α_{av} map produced using the looser envelope and new SFC at 3.5Å resolution. This map was bricked and displayed using FRODO (Jones, 1985). The O_1BFS coordinates were used as a starting model and an initial build performed where sequence differences were modelled and any residues for which the density was ambiguous, deleted. The sequence differences could be clearly seen in all the coat proteins. The majority of the sequence changes were associated with little conformational change – those occurring in the β–sheets generally resulted in no main chain movement – however, some of the groups of changes in the external loops joining the sheets produced changes in the main chain. The G–H loop of VP3 showed some of the greatest main chain movements between the two viruses. The C–terminus of VP3 has also 'flipped' into a new conformation on the virus surface.

Disappointingly, it soon became apparent that as for the 'native' O_1BFS, no clear density could be seen for the VP1 G–H loop. The envelopes used in the averaging were carefully checked to ensure the density was not simply missing because it had been truncated, but no model could be built for residues 132 to 156 of VP1 (inclusive). Also missing from the electron density map were the two C-terminal residues of VP1 (213 and 212) and portions of VP4 – although all the residues present in the O_1BFS

model could be located, the portions disordered in that virus were no more ordered in C-S8c1.

Two areas of the map showed ordered density in regions not accounted for by a model based on O_1BFS;

1. The N–terminus of VP2, which was truncated in O_1BFS to residue 8, appeared better ordered in C-S8c1. Three symmetry related copies of VP2 associate around the 3–fold icosahedral axes and density could be seen running down the 3–fold axis and then across the internal surface of the capsid. At the resolution of this map however, interpretation of the density was problematical – it tended to become 'blurred' into one tube of density on the axis, rather than 3 symmetry related pieces. No model was built for residues 1-8 of VP2 at this time.

2. Under the icosahedral 5–fold axis on the interior of the capsid (at a radius of 103Å) was a large piece of ordered density associated with VP4. The density showed obvious 5–fold symmetry, and by analogy to poliovirus (Filman *et al.*, 1989) it seemed likely that the density would correspond to the myristate moieties known to be attached to the N–terminal Gly of VP4 (see Chapter 6 for further discussion). However, a model could not be built for this region.

This initial model was then refined using XPLOR (Brünger, 1992).

5.8 Refinement of the Model

This model gave R_C = 33% for all data between 30 and 3.5Å without solvent correction (31% after bulk solvent correction and scaling up of atomic B factors, see below). The C-S8c1 crystals were significantly less ordered than the O_1BFS crystals suggesting the atoms were likely to be more mobile in C-S8c1. At all temperatures atoms have a finite amplitude of oscillation such that the atomic scattering factor f_0 calculated by assuming spherical symmetry must be modified by a scale;

$$f = f_0 T$$

$$T = exp(-Bsin^2\theta/\lambda^2)$$

Where T is the temperature scale at any given angle of diffraction, B a constant and θ the Bragg angle. We can use this equation to give an estimate of the C-S8c1 atomic B factors (Wilson, 1942) by scaling the C-S8c1 observations to O_1BFS data. This gives us resolution dependent scale factors between the two sets of observations such that at any resolution;

$$|\mathrm{F}_{C-S8c1}| = |\mathrm{F}_{O_1BFS}| \; K$$

where K consists of a scale between the two data sets (k) which is independent of resolution and the temperature factor T which is resolution dependent.

$$K = kT$$

and at any resolution X

$$|\mathrm{F}_{C-S8c1}|^X = |\mathrm{F}_{O_1BFS}|^X \; kT^X$$

We can therefore use the scales K, between $|\mathrm{F}_{C-S8c1}|$ and $|\mathrm{F}_{O_1BFS}|$ calculated at two different resolution limits to give a pair of simultaneous equations which may be solved to give a value for B. This B represents the difference in the overall B factors for the two viruses. K was equal to 1.05 at 5.0Å and 0.74 at 3.5Å, giving a value for B of 32.33Å^2. Individual atomic B–factors were therefore assigned in the model by adding 30Å^2 to the B–factors for the corresponding atoms in O_1BFS. Cycles of simple positional and B–factor refinement, carried out using a Convex C210, the weight on the X–ray terms set to $w_a = 1 \times 10^7$ with positional tolerances (the distance atoms are allowed to move before recalculation of an FFT) between 0.15 and 0.05. Gaussian distributions of half–width 2.0 and 4.0Å^2 were applied to main and side chain B-factors before each B–refinement step, this allowed use of only a small rweight[6] for B–factor refinement (0.9 or 1.8; Table 5.6) allowing essentially free refinement of B–factors. During all the refinement strict icosahedral, 5–fold non–crystallographic symmetry was imposed to maximise the observations to parameters

[6]rweight is the weight given to B factor restraints during refinement. If rweight is zero no restraints are computed

ratio for the refinement (number of non–hydrogen atoms = 5197, number of X–ray observations = 64671). Cycles of refinement were alternated with manual rebuilds to correct features highlighted in $|F_O| - |F_C|$ maps. The progress of the refinement can be seen in Table 5.6. After the first round of refinement a $2|F_{obs}|-|F_{calc}|$, α_{calc} map was used to rebuild the less good regions of the model and to build an initial model for residues 1 to 8 (inclusive) of VP2 (see Chapter 6 for a description of this structure). These residues were included in all further refinements. Prior to the final refinement round (round 3) the parameters defining the bulk solvent correction were redetermined using several, independent, grid searches. Until this point the parameters as determined by D. Logan (Logan, 1991) in refining O_1BFS had been used:

$$\text{SOLRAD} = 0.0$$
$$\text{k} = 0.41$$
$$\text{B} = 110$$

The grid searches revealed clear minimum values for R_C (Figure 5.6) with the parameters:

$$\text{SOLRAD} = 0.3$$
$$\text{k} = 0.32$$
$$\text{B} = 110$$

These parameters were therefore used for all further refinement.

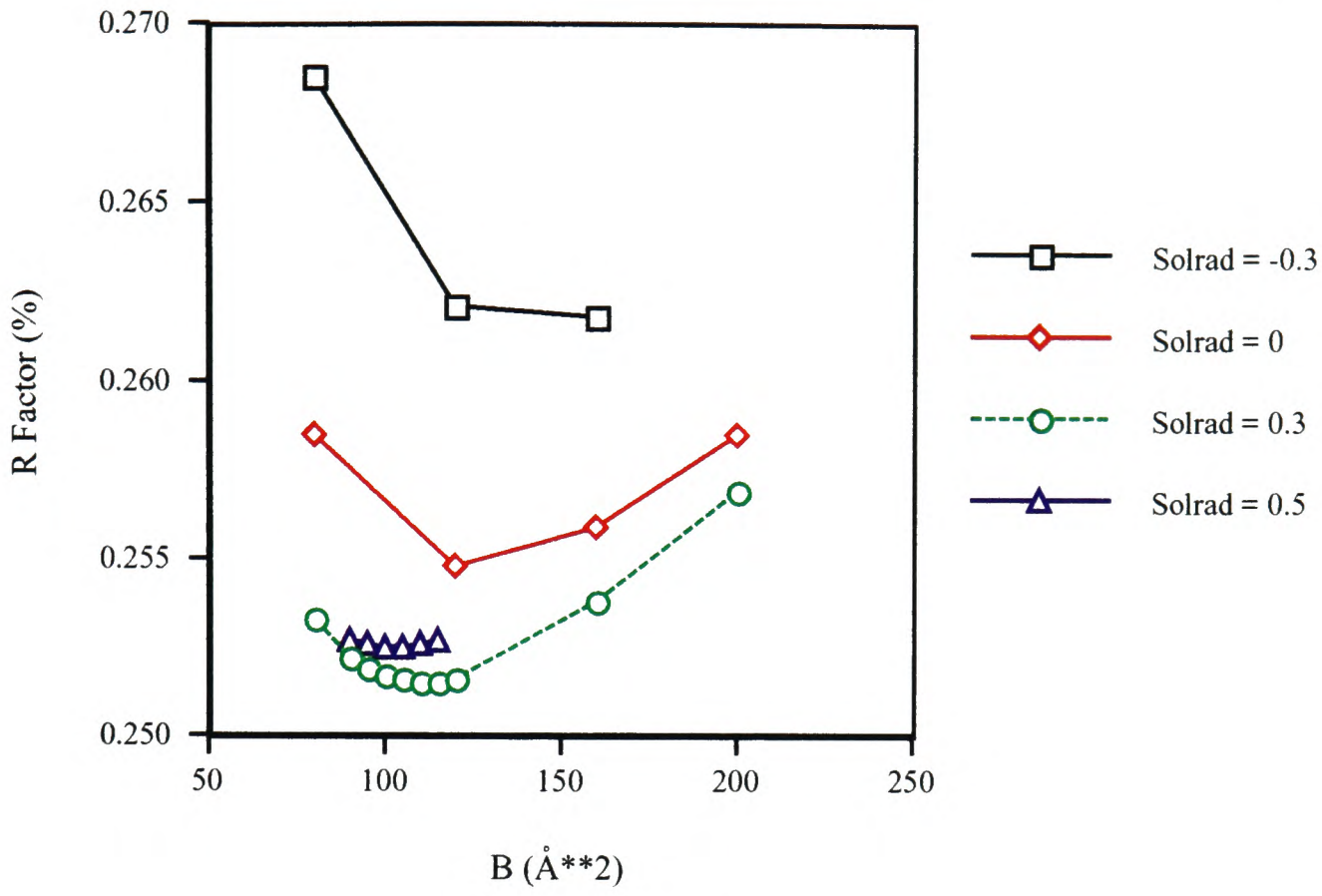
Final scaling of $|F_{calc}|$ after the third refinement round (with bulk solvent correction) to the observed data using SHELL_SCALE gave $R_C = 20.7\%$ (C = 0.88) for all data between 25 and 3.5Å (see Table 5.6.

Cycle No.	Type of Refinement	R_C	Comments
1		31%	
	Positional		tol=0.15
50		24%	
	Positional		tol=0.05
100		23%	
	B–factor		rweight=0.9
110		21%	
	B–factor		rweight=1.8
120		20%	
Geometry at this stage: RMSD bonds 0.030 Å RMSD angles 5.00°			
Manual Rebuild / N–terminal residues of VP2 added			
		24%	
	Positional		tol=0.15
170		22%	
	Positional		tol=0.05
220		22%	
	B–factor		rweight=0.9
235		21%	
	B–factor		rweight=1.8
250		21%	
Geometry at this stage: RMSD bonds 0.026 Å RMSD angles 4.75°			
Manual Rebuild / Redetermine Bulk Solvent Parameters			
		25%	
	Positional		tol=0.15
300		22%	
	Positional		tol=0.05
400		22%	
	B–factor		rweight=0.9
440		21%	
	B–factor		rweight=1.8
480		21%	
Geometry at this stage: RMSD bonds 0.016 Å RMSD angles 3.51° Δ B Bonds 6.98Å Δ B Angles 10.38Å			

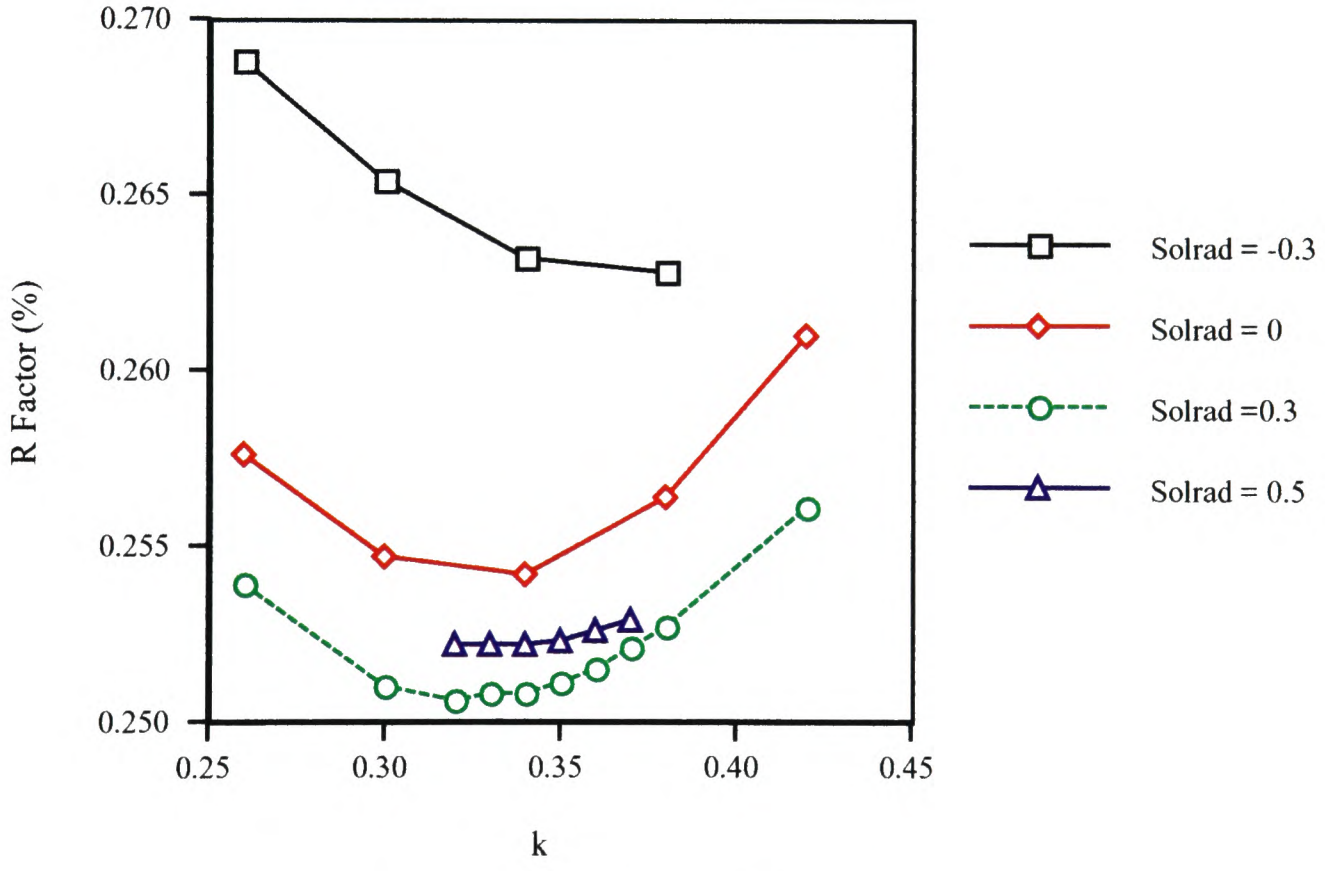
RMSD bonds - The root mean square deviation from ideal covalent bond lengths;**RMSD angles** - The root mean square deviation from ideal covalent bond angles; **ΔB Bonds** - The root mean square difference in isotropic B–value between covalently bonded atoms; **ΔB Angles** - The corresponding difference between covalently bonded next nearest neighbours.

All data 25-3.5Å (including weak data) were used throughout the refinement (64671 reflections).

Table 5.6: Refinement of C-S8c1



Search for B, k=0.32.



Search for k, B=110Å^2.

Figure 5.6: Grid searches for bulk solvent correction parameters.

5.9 Mutant SD6-6

5.9.1 The Virus

SD6-6 is a monoclonal–antibody (MAb SD6) resistant escape mutant of FMDV C-S8c1 (Mateu *et al.*, 1990) bearing a single amino acid substitution in VP1, Ser(139)→Ile. This residue lies within the 'FMDV loop' (the G–H loop of VP1) a region which is not clearly seen in the C-S8c1 structure (see above and Chapter 6). Since this mutation produces a conformational change large enough to prevent binding of an antibody, structural differences between the parent and mutant viruses may help interpretation of the scarce evidence concerning the location of the VP1 G–H loop in either the parent, the mutant or both viruses.

5.9.2 Data Collection

Crystals of SD6-6 grew under the same conditions as those of the parent virus and had the same morphology (Tab 5.7). Data were collected on station 9.6 of the SRS Daresbury using radiation of wavelength 0.89Å; however, since collection of the parental data, a RAXIS imaging plate (Molecular Structures Corporation) had been installed and this was used in place of film. Data collection is summarised in Table 5.7. The increased sensitivity of the image plate resulted in much shorter exposure times and enabled collection of several images from each crystal (allowing collection of a reasonably complete data set despite the small number of crystals available). The increased 'granularity' of the detector (pixels 101.7μm by 105μm rather than the 50μm^2 for film) meant the data were recorded at a much larger crystal–to–detector distance to ensure adequate spot separation (the plate measuring 2000^2 pixels). As for the C-S8c1 crystals the SD6-6 crystals gave relatively weak diffraction with little diffraction seen beyond 3.5Å . Since this is also the resolution limit of the parental data 3.5Å is taken as the resolution limit for the analysis.

SD6-6	
Date	November 1992
Station	9.6 Daresbury
Media	RAXIS imaging plate (fine scan)
λ	0.89Å
Oscillation Range	0.5°
Crystal to Plate	325.5 mm
Exposure	60s
No. of Crystals Examined	5
No. of Useful Images	24
Precipitant used to grow crystals	9.5–11.5% $(NH_4)_2SO_4$
Crystal Morphology	Rhombic Dodecahedra
Crystal Size	$\approx$ 0.3mm in all dimensions
Space Group	I23
Cell	a=b=c=347.6Å $\alpha = \beta = \gamma$=90.0°

Table 5.7: SD6-6 - Data collection and crystal characteristics.

5.9.3 Data Processing

Data were processed using the MOSFLM package (Andrew Leslie) and CCP4 data reduction programs. Indexing solutions were easily produced using all the images for each crystal. Integration with MOSFLM required careful optimization of the scanner parameters – largely by trial and error. The parameters producing the best integration (as judged by R(I), measurements of I/σI et.c.) are shown in Table 5.9. Crystal 3 gave only limited diffraction and was processed to a Bragg spacing of 4.2Å the other crystals were processed to 3.5Å.

Symmetry related reflections on the same images were scaled using ABSCALE and the data divided into unswapped and swapped streams by comparison with the C-S8c1 data. Attempts to merge reflections between different images within each data stream (ROTAVATA) revealed a non–Gaussian distribution in the estimates of the errors of the intensities. This arises from the fact that the original assumption,

Crystal	REFIX		MOSFLM I/σI		ABSCALE	REINDLCF	
	SDXY	SDPHI	Overall	3.5Å	R_{symm}(I)	UN	SW
Crystal 1	0.47	0.09	†			0.14	**0.82**
001					14.7		
002					19.3		
003					20.6		
004					24.0		
Crystal 2	0.52	0.10				**0.84**	0.15
001			10.5	4.5	22.0		
002			10.6	4.3	12.9		
003			10.1	4.5	24.8		
004			9.9	4.1	23.9		
005			9.3	3.4	17.8		
006			8.1	2.7	13.2		
Crystal 3	0.80	0.10				**0.88**	0.14
001			7.3	4.7	21.5		
002			7.7	5.2	16.0		
003			7.7	4.7	17.2		
004			6.7	3.8	24.0		
005			7.7	4.9	21.4		
006			7.2	4.2	17.7		
007			6.9	3.5	19.2		
008			6.5	3.8	10.5		
Crystal 5	0.42	0.09				0.14	**0.78**
001			9.8	3.8	23.1		
002			10.1	4.3	24.8		
003			10.1	3.8	23.6		
004			9.8	3.7	24.0		
005			9.1	3.6	24.3		
006			8.9	3.3	27.0		

†- o/p lost

Table 5.8: Indexing and integration statistics for SD6-6

CCX	-0.25
CCY	0.40
CCOM	0.02
TILT	-5
TWIST	-25
GAIN	10
Scanner Error Factor	0.11

Table 5.9: Refined MOSFLM scanner parameters for RAXIS

that $\sigma(I)\alpha\sqrt{I}$ (i.e. errors arise from counting statistics), was inadequate. Some attempts were made to model systematic errors by adjusting the scanner parameters however the errors on the strong data were still (for this data) being underestimated. The distribution of $\sigma(I)$ may be normalised using parameter SDADD in ROTAVATA where:

$$\sigma_{new}(I) = \sqrt{\sigma(I)^2 + (SDADD.I)^2}$$

It has been shown that when the reflections are normally distributed $< \Delta I >$[7] should be 0 and $\sigma(\Delta I)$ should be 1.0 (French and Wilson, 1978). A coarse grid search for SDADD (SDADD = 0.02, 0.05, 0.07, 0.10, 0.15, 0.20, 0.25) lead to the selection of a value of 0.25 for this data. Applying this correction the two data streams scaled reasonably well with R(I) 19.2% and 22.6% for the unswapped and swapped streams respectively. Postrefinement (Table 5.10) slightly increased the estimates of the crystal mosaicity, and also refined the oscillation ranges such that a total oscillation range of 0.5° was reduced by an average of 0.05°. This suggests an error in the spindle motor, or in the timing of the shutter opening and closing. This ill-determination of the oscillation range probably accounts, in part, for the mediocre quality of the data (as judged by R(I)). The two data streams were merged after swapping the indices of the swapped stream and reflections with I > -2σ and more than 75% recorded appropriately scaled and marked as fully recorded to give a 60%

[7] where ΔI_h is $\sum_j || I_{hj} | - < I_h > |$

Pack	< I >	R_{merge}(I)		Post–Refinement						
	before PR	before PR	after PR	$\Delta\phi$y	$\Delta\phi$z	$\Delta\lambda$ (Å $\times 10^5$)	ϵ	F→P	P→F	total F
Unswapped	262.83	19.2	16.5					8792	167	36916
202	246.13	16.7	12.0	0.00	-0.05	150	0.088	1143	0	3119
203	244.48	15.0	10.6	0.00	-0.05	88	0.060	690	5	3612
204	238.83	16.7	12.7	0.00	-0.04	76	0.066	751	6	3628
205	244.43	17.6	14.1	0.00	-0.06	93	0.039	478	57	3829
206	250.74	17.4	14.5	0.00	-0.05	7	0.060	624	5	3626
207	243.63	17.5	14.0	0.00	-0.04	119	0.073	792	11	3511
302	259.99	22.2	19.5	0.01	-0.06	167	0.080	624	1	1879
303	282.45	19.7	19.7	0.00	-0.03	180	0.092	564	0	1916
304	281.23	19.7	17.8	0.01	-0.02	255	0.097	537	7	1940
305	281.04	22.1	21.5	0.00	-0.02	295	0.091	536	7	1954
306	282.40	19.4	17.6	0.00	-0.03	147	0.092	462	12	2041
307	293.94	20.8	18.5	0.00	-0.02	199	0.097	523	10	1951
308	281.13	22.8	21.4	0.01	-0.04	227	0.067	407	52	2104
309	294.88	24.3	23.7	0.00	0.00	198	0.130	661	0	1806
Swapped	265.86	22.6	18.1					4599	451	27863
102	244.68	18.4	†	0.00	-0.07	216	0.034	546	75	3825
103	256.71	19.8		0.00	-0.05	260	0.026	339	181	4154
104	229.88	19.8		0.00	-0.06	67	0.052	690	3	3676
105	243.88	21.0		0.00	-0.06	64	0.069	895	0	3391
502	295.32	24.3		-0.01	-0.07	165	0.015	136	71	1842
503	286.74	24.2		0.00	-0.05	52	0.048	316	0	1952
504	270.95	23.7		-0.01	-0.07	225	0.025	245	39	2268
505	296.49	22.2		-0.01	-0.07	168	0.009	188	79	2502
506	281.83	26.5		0.00	-0.05	-40	0.089	673	0	2039
507	279.27	26.3		0.01	-0.08	211	0.049	571	3	2214

†- o/p lost

Input Parameters: mosaicity 0.05° ; λ 0.89Å ; beam divergence - 0.030° horizontal / 0.015° vertical; spectral dispersion 0.001 / correlated component 0.0000. **Key:** F - fully recorded; P - partially recorded; ϵ - beam divergence $\equiv$ mosaicity

Table 5.10: Summary of postrefinement of SD6-6

	Before 'Massagelcf'		After 'Massagelcf'	
	R(I) (%)	No. of Ind. Refl.	R(I) (%)	No. of Ind. Refl.
Unswapped	16.5	29905	20.3	27301
Swapped	18.1	23596	18.2	36416
Combined	-	-	17.9	52677

Table 5.11: Final data processing statistics for SD6-6

complete data set (R(I)=18.2%) to 3.5Å (Table 5.11, Figure 5.1).

5.9.4 Calculation of Difference Maps

The SD6-6 data scaled to the C-S8c1 data in Shell_Scale with R_C=17.8%, C=0.89 for all data 18–3.5Å (R_C=23.2, C=0.69 for the shell 3.52–3.50Å). Difference maps were calculated at 4Å and 3.5Å with amplitudes $|F^{OBS}_{CS8-c1}|$ - $|F^{OBS}_{SD6-6}|$ using the C-S8c1 phases derived from the NCS averaging, these were then 5-fold averaged. Such maps showed no peaks above the noise in any region of the map.

5.9.5 Implications of the structure of SD6-6

The single amino acid substitution in SD6-6 lies within the VP1 G–H loop. The parental structure reveals only hints concerning the location of the loop residues in the form of discontinuous density lying over the surface of VP2. The lack of any detectable difference implies the location and occupancies of the residues giving rise to this lower level density are not significantly altered in the mutant with respect to the parent. This density, most probably, corresponds to the main chain atoms of the loop (these are likely to be significantly more ordered than the side chain atoms), the lack of difference density therefore implies that the mutant achieves antigenic escape simply by alteration in the chemical nature/conformation of the mutated side chain without requiring re-arrangement of the underlying main–chain atoms. This result is highly analogous to the level of structural changes seen in O_1 mutants (Chapter 3)

and emphasises that relatively small structural changes can have dramatic antigenic effects.

5.10 Creation of 'Master' Data Set

The lack of any detectable structural differences between SD6-6 and C-S8c1 suggested combination of the two data sets to produce a more complete 'master' data set would be valid. Merging the data sets using Rotavata gave a 90% complete data set to 3.5Å with R(I) = 17.8% (78486 independent reflections, Fig. 5.1).

5.11 GAP NCS Averaging

The completeness of a data set has a great effect on the quality of the phases given by NCS averaging – using an incomplete data set means that there are gaps in the sampling of the transform which will result in ripples in the FFT (hence the use of calculated structure factors to fill in gaps in observed data during averaging). The creation of a more complete 'master' data set therefore suggested improved phase estimates could be gained by further cycles of NCS averaging using these amplitudes. Since many of the 'more interesting' regions of the structure (e.g. density around the interior of the 5–fold axis, N–terminus of VP2 et.c.) have relatively weak density, any improvement in phase accuracy should aid interpretation. The 'best' phases from the previous NCS averaging were taken as a start point and the more complete 'master' data set used. A new protocol using program GAP (J. Grimes and D. Stuart unpublished) for envelope determination and averaging was used.

GAP uses the larger memory capacity of modern computers to store all maps and envelopes as large arrays in 'real' memory. This allows random access to all pixels and removes the need for tedious sorting of the maps. Each map and envelope is read in only once in every cycle so less time is spent performing I/O. GAP also uses 11–point nonlinear interpolation as proposed by Nordman (Nordman, 1980) (rather than the

linear interpolation used in the Bricogne–type averaging) which requires the map to be sampled at only 1/3 of the maximum resolution. A 256 grid is therefore adequate for averaging of C-S8c1 at 3.5Å reducing the data storage problems associated with the large (fine) maps. The program internally folds and expands between the U_{I23} and U_{I222} asymmetric units as appropriate.

Envelopes are produced using similar principles to those already described, but the parameters used to define them are somewhat altered. A geometric envelope constructed from the logical addition of 2 spheres (inner radius 95Å, outer radius 154Å) and a plane cut envelope (using planes perpendicular to the crystallographic three–fold axes of symmetry, at a distance of 149Å as before) are used to flatten an input $2|F_{obs}| - |F_{calc}|, \alpha_{calc}$ map calculated on a 256 grid, using data between 30 and 3.5Å. The 'Wang–type'(Wang, 1985) envelope is generated in two steps:

1. A sigma cut is applied to the density:

$$\text{abs } \rho < 0.5\sigma \text{ then } \rho \text{ set to } 0$$
$$\text{abs } \rho > 5.0\sigma \text{ then } \rho \text{ set to } 5.0\sigma$$

2. The density is 'flipped' to its absolute value.

3. The density is convoluted using a Gaussian of width equivalent to a B factor of $40Å^2$.

4. Density histogramming reclassifies pixels based on the estimated solvent content. A value of 55% was used – although this underestimates the actual solvent content of the crystals which is nearer 80% – as the envelopes produced using these parameters were seen to fit the C-S8c1 model well when displayed in FRODO (Jones, 1985).

5. TIDY then reclassifies pixels as solvent or protein based on their neighbouring pixels (less than 20 in 2×10^6 pixels were reclassified).

Envelopes were calculated at cycles 1 and 14 of the GAP averaging and appeared to be of a very high quality as judged by examination in FRODO or by looking at plotted sections (Figure 5.7). A Rayment weighting scheme (with rejection criteria as before) was used for cycles 1 to 16 and a unitary scheme thereafter.

The GAP averaging converged after 30 cycles (the mean phase change between cycles 29 and 30 was 0.42°). The phases from this averaging cycle showed a mean phase difference from the previous 'best' phases of 30.4° and from the phases used as a start point of the original averaging i.e. those generated using the O_1BFS model in the C-S8c1 cell, 37°. The solution given by the averaging is seen to be of a very high quality as shown by the agreement between the observed amplitudes and the amplitudes produced by back–transforming the final map (R_C = 11.86%, C = 95.44%; Fig. 5.8). This map confirmed the previous maps but gave improved clarity in the N–terminal region of VP2 and the density under the 5–fold axis.

5.12 Final Refinement and Model Status

Careful checking of the model against the new map enabled re–orientation of some of the side chains for residues within the region 1–12 of VP2. However even in this map, location of all the side chains in this region is problematical and Glu 6 is modelled as a Gly. The main chain conformation of this region is however confirmed by the new map. The only other region of the model which required alteration was the B–C loop of VP3. Although continuous density is seen in the region that must be occupied by this loop, building a single conformation, that is able to account for all the density, and which has reasonable stereochemistry is impossible. Although the loop almost certainly spends part of its time in a conformation similar to that seen for this region in O_1BFS repeated attempts to refine this conformation against the C-S8c1 data proved unsatisfactory. The loop has therefore been rebuilt into an altered conformation which provides a reasonable explanation of the density seen. This conformation is unlikely to to be a complete representation of the loop structure,

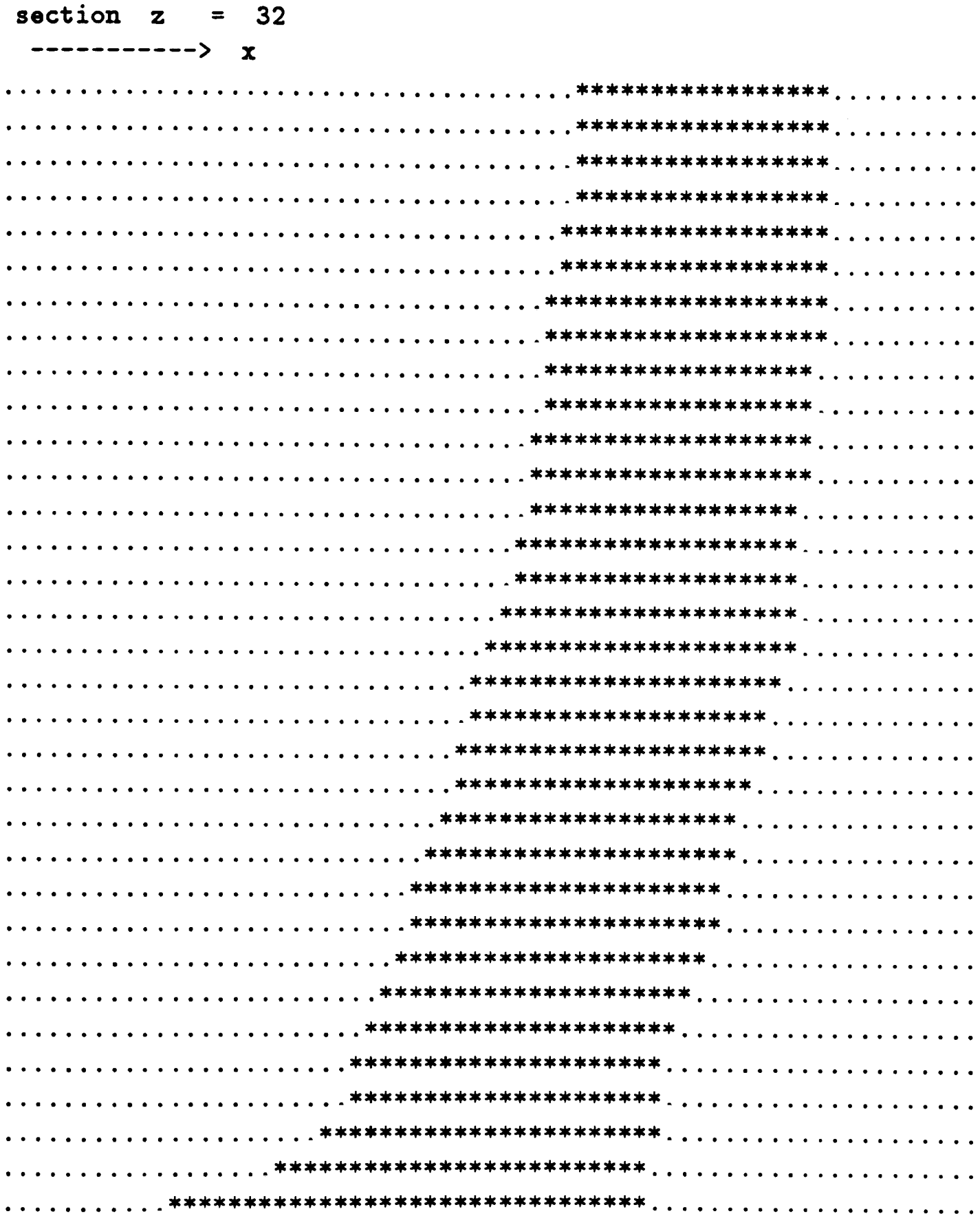


Figure 5.7: Section of GAP 'Wang-type' envelope.

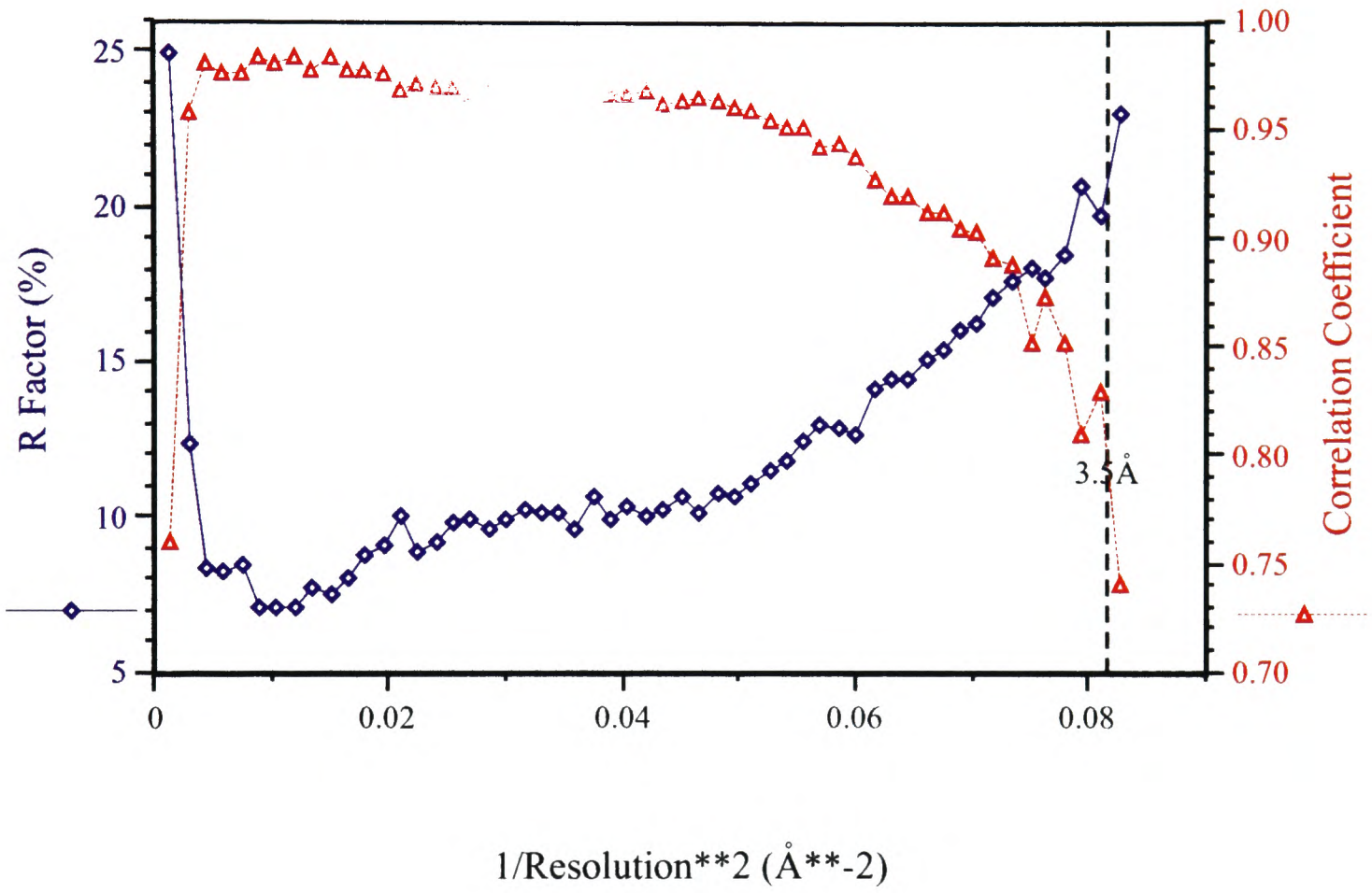
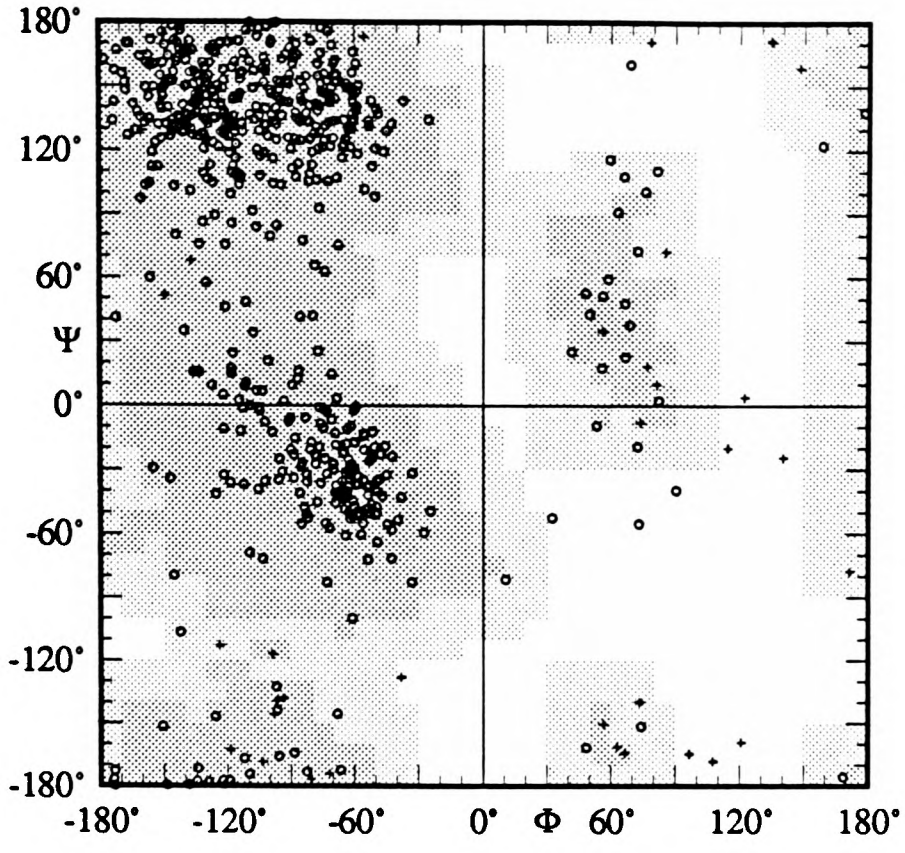


Figure 5.8: Quality of GAP averaging solution.



Gly marked as +, all others as o
Non–Gly residues in 'BAD' regions are Glu 5 and Thr 8 of the N–terminus of VP2, Ser 87 and Ala 196 of VP3.

Figure 5.9: Ramachandran plot for final model C-S8cl.

with the loop likely to be very flexible (B–factors for the Cα atoms as built are high with a mean value of 89Å^2; Figure 6.2) adopting multiple conformations over time and between different copies of the biological protomer in the virus.

Final positional and B–factor refinement in XPLOR against the 'master' data set has resulted in a model with R_C = 21% and C = 0.87 for all data between 25 and 3.5Å . This model contains 670 residues and has good stereochemistry (Tab. 5.12, fig 5.9).

Refinement	
No. of Independent Reflections	79016
No. of Non–Hydrogen Atoms in Refinement	5197
Stereochemistry of Model	
RMS Δ Bonds	0.012Å
RMS Δ Angles	2.0°
ΔB Bonds	4.86Å^2
ΔB Angles	8.02Å^2
Residues Included in Model[a]	
VP1	1–132, 157–211
VP2	1–218[b]
VP3	1–220[c]
VP4	15–39, 65–85

RMS Δ bonds refers to the root mean square deviation from ideal covalent bond lengths. **RMS Δ angles** refers to the mean square deviation from ideal covalent bond angles. **RMS ΔB Bonds** refers to the root mean square difference in isotropic B–value between covalently bonded atoms. **RMS ΔB angles** refers to the corresponding difference between covalently bonded next nearest neighbours.

a – Numbering of residues is as for O_1BFS, therefore there are no residues 135–138 of VP1, or residue 60 of VP3.

b – Glu 6, no evidence for position of side chain. Included as Gly in the Final Model.

c – Numbering relative to O_1BFS, there is therefore no residue 59 in this model.

Table 5.12: Stereochemistry, and residues included in final model of C-S8c1

Chapter 6

Analysis of the Structure of Virus C-S8c1

C-S8c1 and O_1BFS share 82% sequence identity in the 4 capsid proteins. As expected the C-S8c1 structure is basically very similar to that previously described for O_1BFS (Acharya *et al.*, 1989). Figure 6.1 shows individual stereo plots[1] for the four capsid proteins with superimpositions of the main chain atoms of C-S8c1 (thicker lines and black circles) and O_1BFS (thinner lines). The β–barrel cores of the 3 main proteins are conserved with an RMS[2] deviation of Cα positions between C-S8c1 and O_1BFS of 0.3Å^2, the limit of accuracy of the data (Figure 6.2). This is comparable to the RMS deviation in Cα positions of the β–barrel core residues seen in the comparison of poliovirus serotypes 1 and 3 (Filman *et al.*, 1989). Significant differences in Cα positions are observed only in the loops connecting the β–strands of these core regions and in the termini of the proteins. Figure 6.2 shows that there is a marked correlation between residue mobility (as judged from crystallographic B–factors) and conformational differences between O_1BFS and C-S8c1. A correlation coefficient may be calculated to invetsigate the link between mobility and positional differences between two proteins (Acharya *et al.*, 1991). This takes the form:

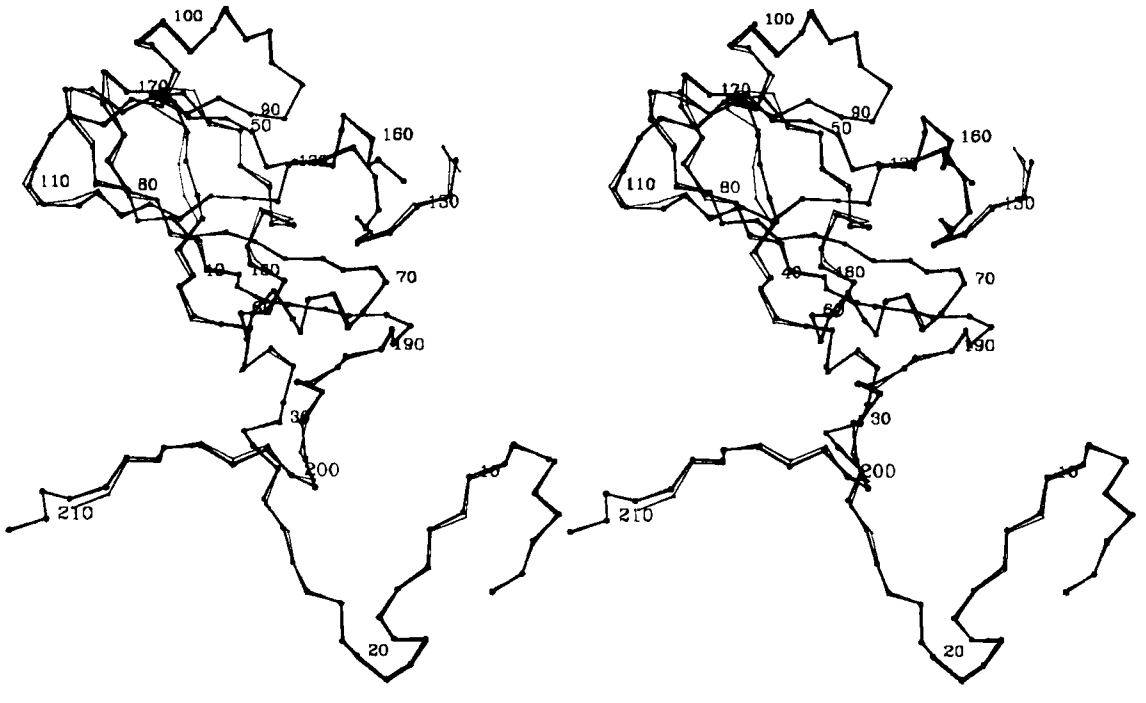
$$C_{BD} = \frac{\sum_j((B_{j,1} - < B_1 >) + (B_{j,2} - < B_2 >)) \times (D_j - < D >)}{[\sum_j((B_{j,1} - < B_1 >) + (B_{j,2} - < B_2 >))^2 \times \sum_j (D_j - < D >)^2]^{\frac{1}{2}}}$$

where $D_j = X_{1,j} - X_{2,j}$; X_j is the position of atom j (after superimpoisition of structures 1 and 2); B values are isotropic temperature factors and the sum $\sum_j$, represents the sum over atoms C, O, N, Cα and Cβ.

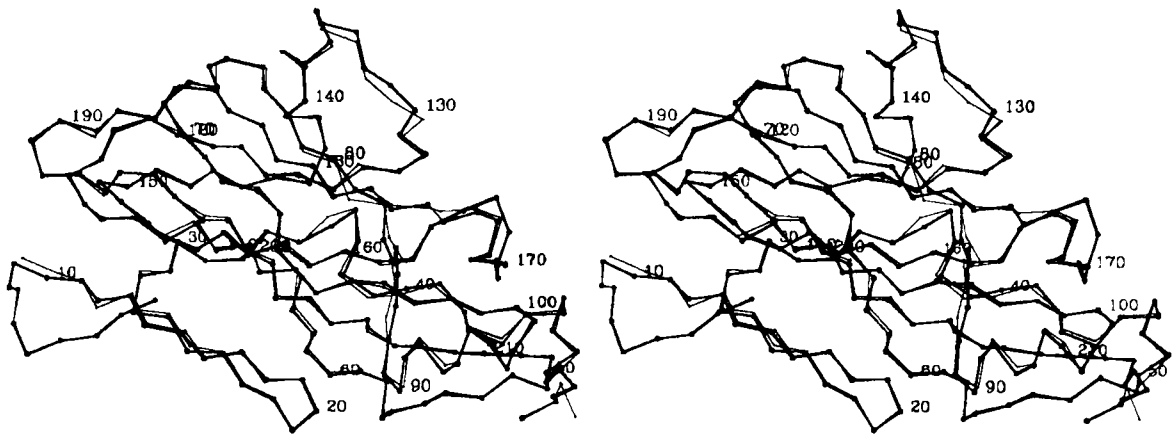
Acharya *et al.* (1991) have compared the structures of human and baboon α–lactalbumin using this measure. These proteins share 92% sequence identity and

[1]Produced using MOPLOT84, D. Stuart.

[2]RMS=Root mean square


100
120
80
50
160
110
80
150
70
190
30
200
210
10
20
100
120
80
50
160
110
80
150
70
190
30
200
210
10
20

VP1


140
130
190
180
150
30
60
170
10
40
100
90
20
140
130
190
180
150
30
60
170
10
40
100
90
20

VP2

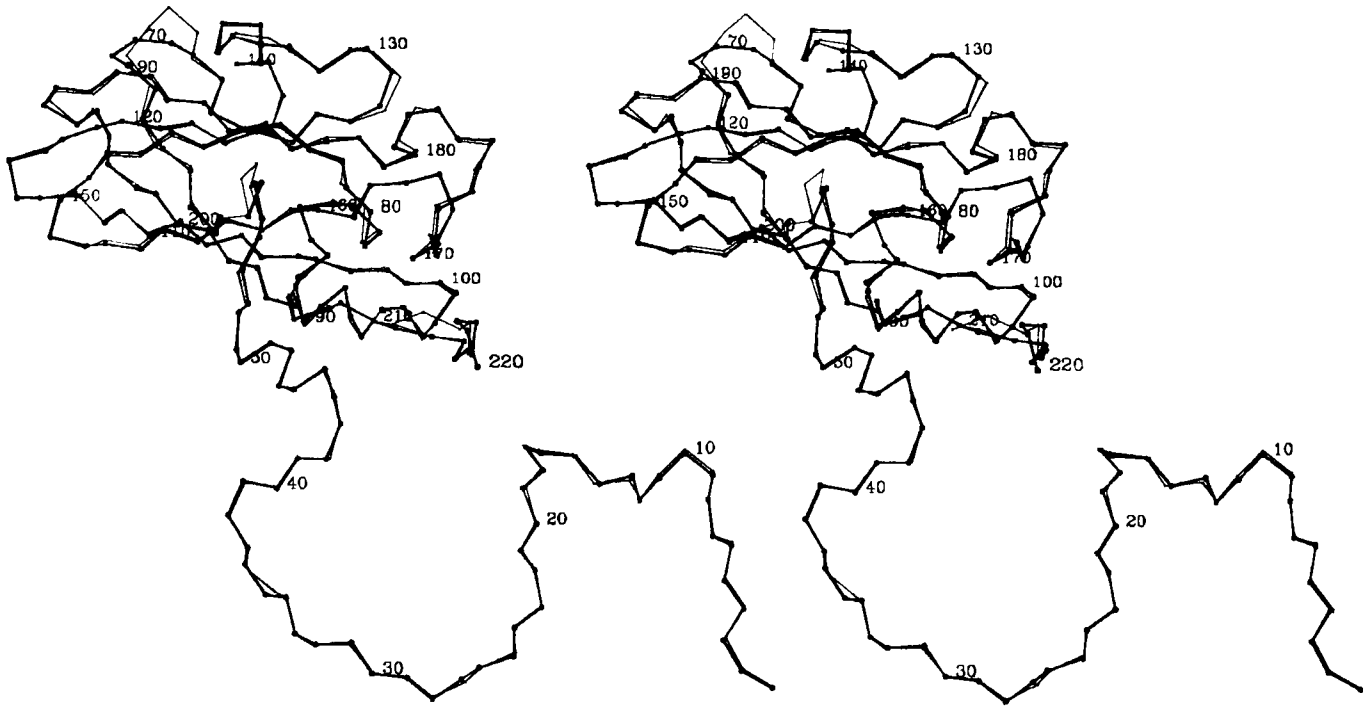
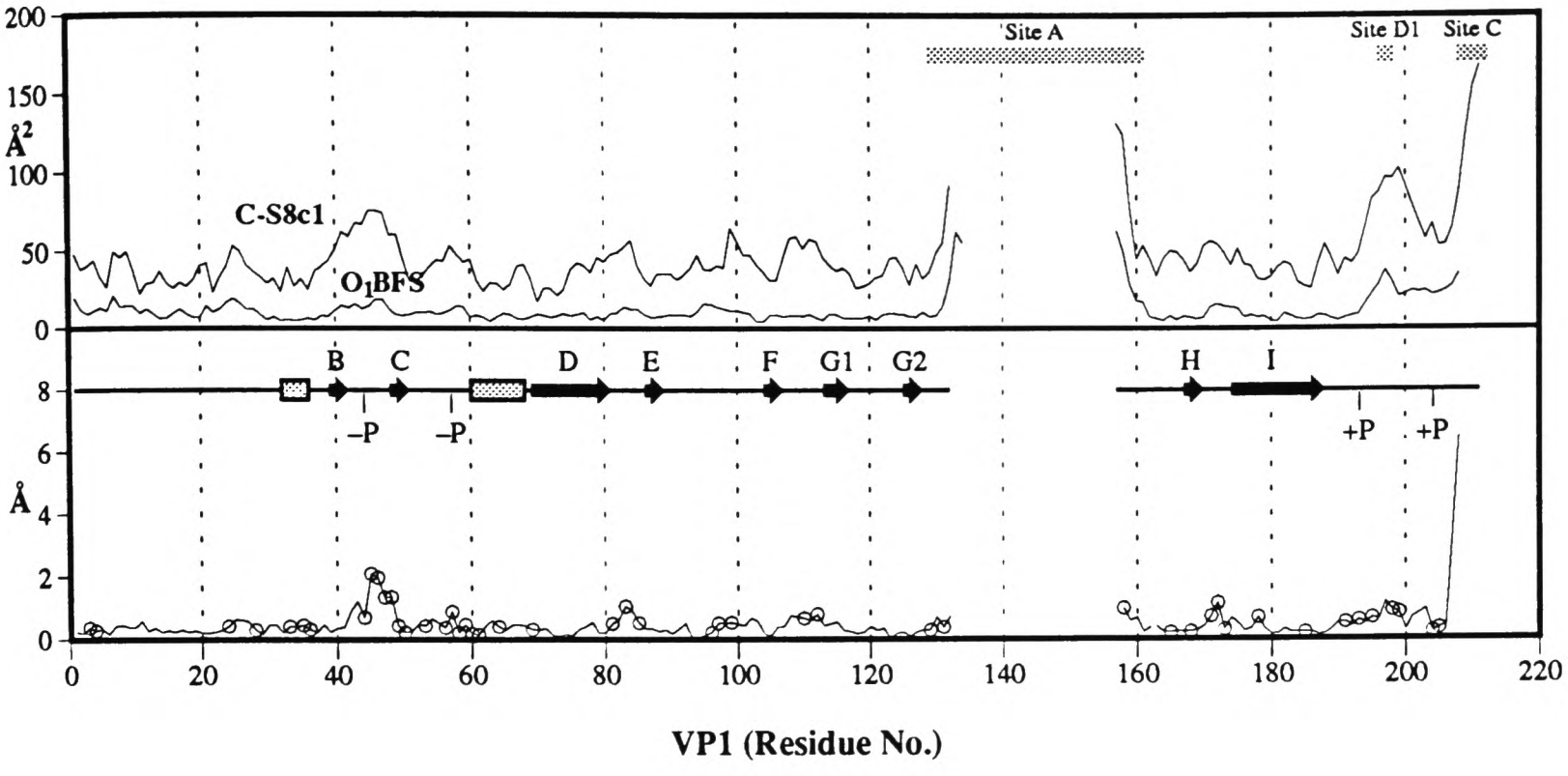
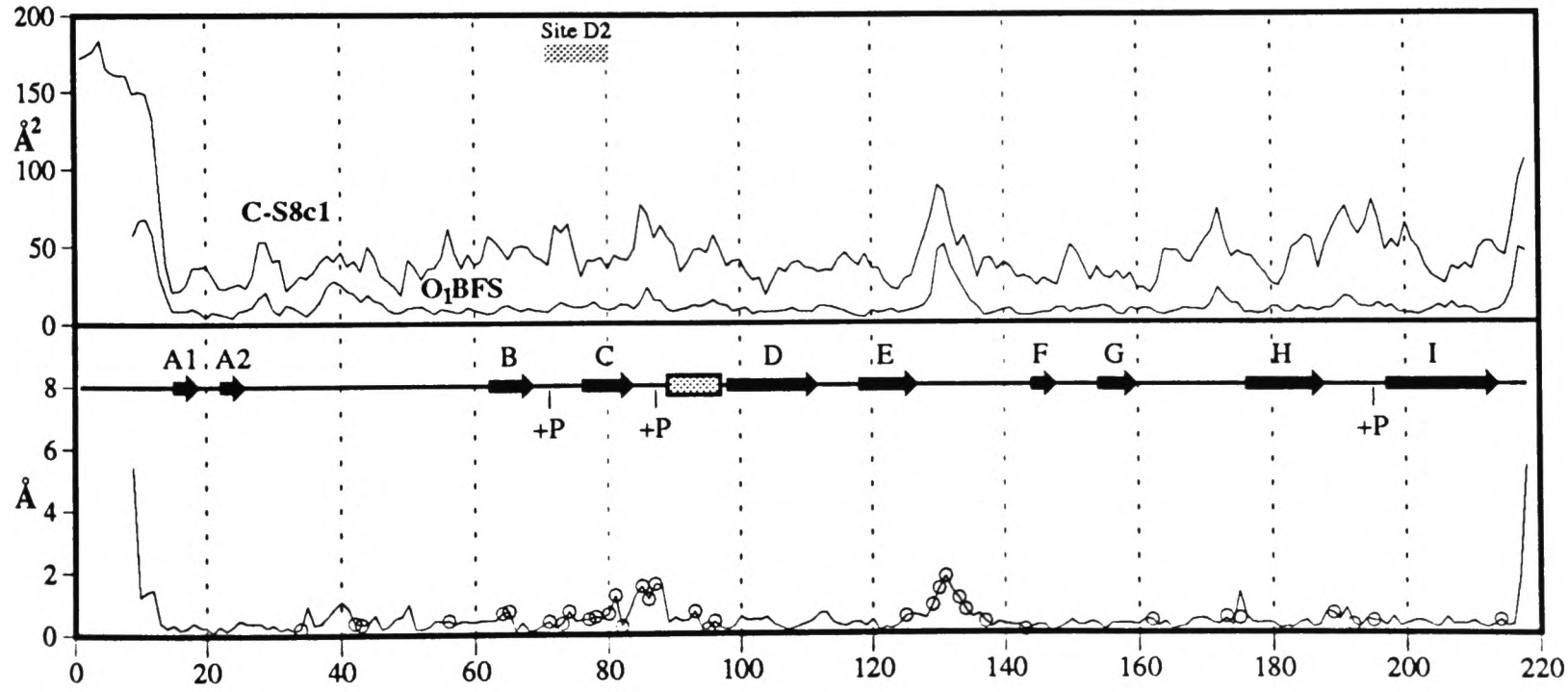
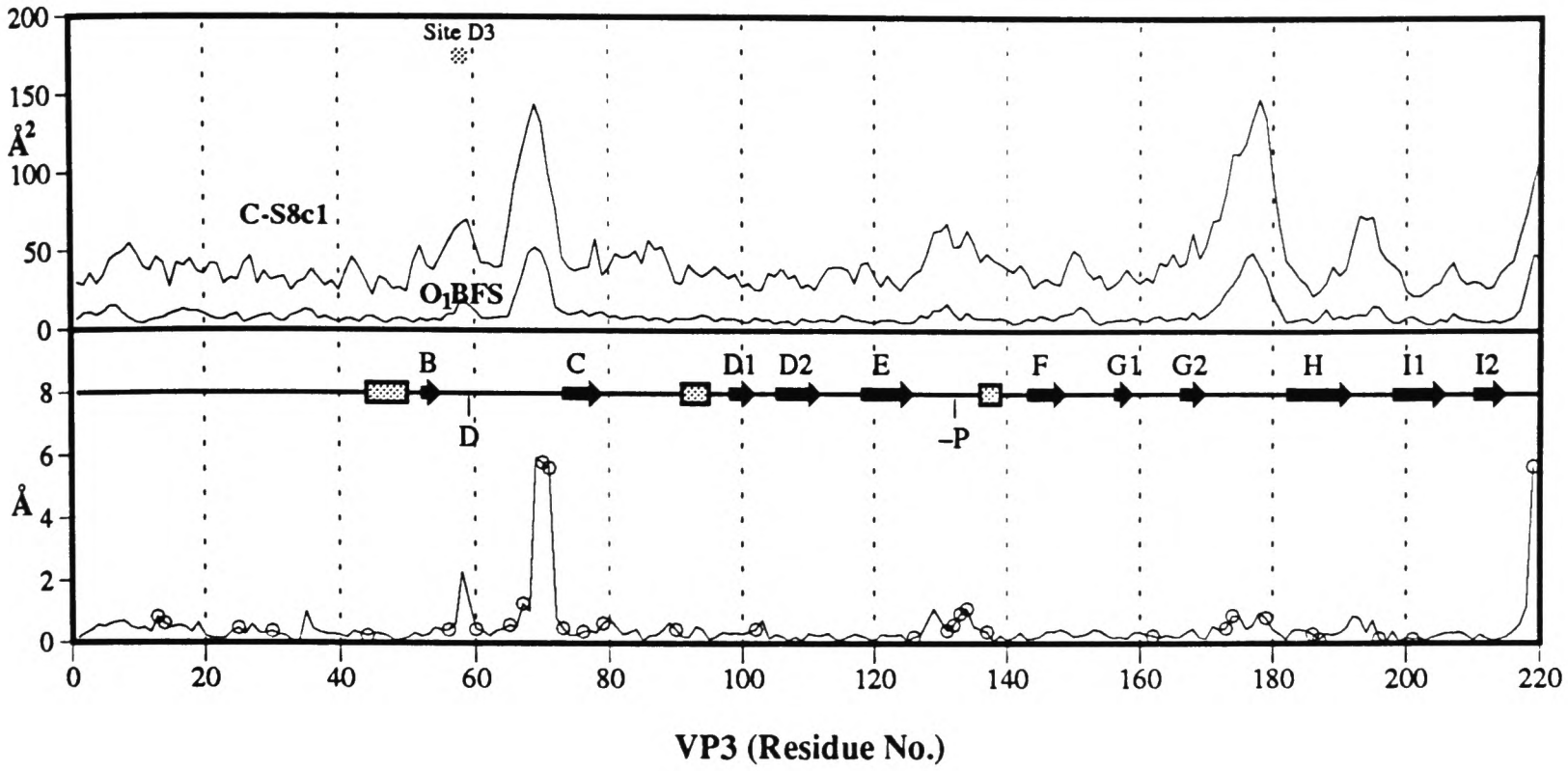
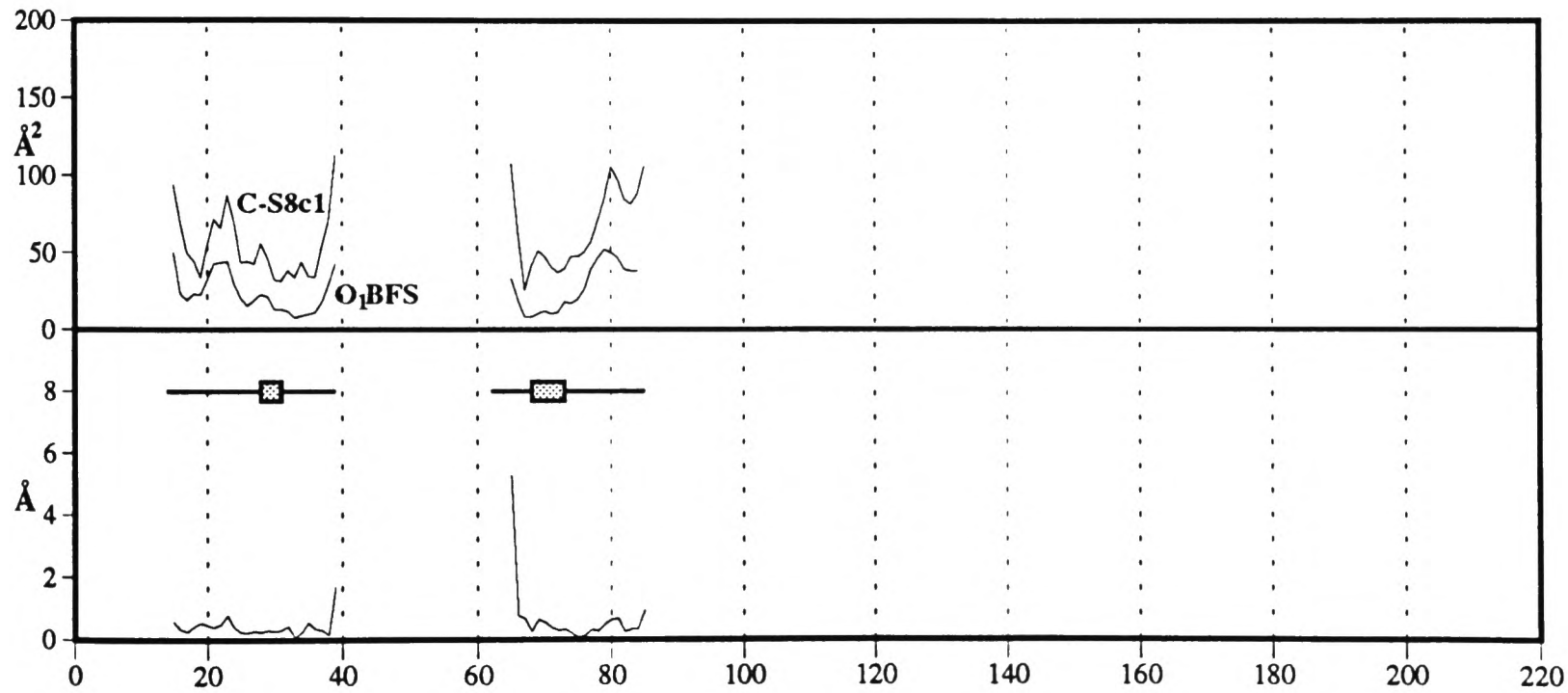


Figure 6.1: Stereo views of superimposition of C-S8c1 and O_1BFS.


Site A
Site D1
Site C
C-S8c1
O1BFS
B
C
D
E
F
G1
G2
H
I
–P
–P
+P
+P
Å2
Å
VP1 (Residue No.)


Site D2
C-S8c1
O1BFS
A1
A2
B
C
D
E
F
G
H
I
+P
+P
+P
Å2
Å
VP2 (Residue No.)





The upper trace shows the main chain B Factors for O_1BFS and C-S8c1, also marked are positions of antigenic sites (labelled, shaded bars) defined by MAb-escape mutations in type C FMDVs. The lower trace shows the RMS deviation in Cα positions between O_1BFS and C-S8c1 with the positions of sequence differences marked as open circles. The upper portion of this box contains a representation of the secondary structural elements (β-sheet - filled arrow; α-helix - shaded box) as defined by program DSSP (Kabsch and Sander, 1983). +P/-P - Substitution by or for a Proline in C-S8c1.

Figure 6.2: Analysis of C-S8c1.

both structures are refined against high resolution data (both at 1.7Å) - the atomic B–factors and positions are therefore well defined. C_{BD} is 0.82 for the α–lactalbumins. When calculated using the coordinates of O_1BFS and C-S8c1, C_{BD} is 0.73. This compares very favourably to the α–lactalbumin value, considering the lower degree of sequence identity between the viruses, and as such emphasizes the linkage between residue mobility and conformational change.

Similarly a correlation between atomic B–factors for equivalent atoms in the two structures may be calculated:

$$C_{BB} = \frac{\sum_j (B_{j,1} - < B_1 >)(B_{j,2} - < B_2 >)}{[\sum_j (B_{j,1} - < B_1 >)^2 \cdot \sum_j (B_{j,2} - < B_2 >)^2]^{\frac{1}{2}}}$$

The value of C_{BB} for O_1BFS and C-S8c1 is 0.79. This compares well with the value obtained when comparing the α–lactalbumins (Acharya *et al.*, 1991), C_{BB}=0.84. This strong correlation between the B–factors for the O_1BFS and C-S8c1 structures indicates that even with X–ray data of limited resolution we may interpret B–factors with confidence.

6.1 The G–H Loop of VP1

Unexpectedly, despite lacking the disulphide bond known to destabilize the VP1 G–H loop in O_1BFS (Logan *et al.*, 1993), this important feature is again found to be significantly less well ordered than other surface exposed loops in C-S8c1. No model has been constructed for residues 133 to 156 [3]. The missing region corresponds closely to the disordered portion in O_1BFS and analysis of re–dissolved crystals confirms that there has been no proteolysis of the C-S8c1 virus. Although the loop appears relatively disordered, some low level density can be seen. The electron density is not of sufficient quality to allow a *definitive model* to be built, however a crude model based on the reduced O_1BFS structure has been constructed and explains most of the observed

[3]This numbering is relative to O_1BFS. Since there is a deletion of 4 residues in C-S8c1 relative to O_1BFS after residue 134 of VP1, this is equivalent to residues 133–152 using a C-S8c1 numbering scheme.

electron density. The loop conformation adopted in CS8–c1 appears different to that seen in the reduced O_1BFS structure (not surprisingly, since there are a number of sequence differences and a 4 residue deletion) but the chain occupies the same region of the capsid surface (lying predominantly over VP2) in both cases. This is in contrast to the structure of native O_1BFS where the loop appears to lie nearer the B–C loop of VP1, in an 'up' conformation (Parry *et al.*, 1990). We would emphasize that this rough model may be only one of several very different conformations that can be adopted by the G–H loop.

Conservation of this disorder between FMDVs of type O and C suggests flexibility of the VP1 G–H loop region is of advantage to the virus.

The importance of the VP1 G–H loop in FMDV is well established. Numerous studies suggest that it is perhaps the most important antigenic feature on the virion surface (Strohmaier *et al.*, 1982; Bittle *et al.*, 1982; Pfaff *et al.*, 1982). For example, a synthetic peptide corresponding to the G–H loop capable of protecting cattle against challenge with live FMDV (Bittle *et al.*, 1982). Moreover, sequence analyses show that the G–H loop is one of the most variable in the virus (Palmenberg, 1989; Martínez *et al.*, 1991). Clearly a very flexible loop places fewer structural constraints on the types of amino acid found within it. Preservation of the flexibility of the G–H loop of FMDV therefore allows considerable sequence, and thus antigenic, diversity.

Intriguingly, at the centre of this sequence hypervariability is one of the most conserved portions of the virion - the RGD tripeptide a critical component of receptor binding (Fox *et al.*, 1989). The identity of the FMDV receptor is unknown but an integrin is strongly implicated (Fox *et al.*, 1989; Geysen *et al.*, 1985; Surovoi *et al.*, 1988; Acharya *et al.*, 1989). The structures of a number of integrin binding proteins, all of which contain the RGD motif, have been solved recently (Wistow *et al.*, 1983; Leahy *et al.*, 1992; Saudek *et al.*, 1991; Adler *et al.*, 1991). In all cases the RGD tripeptide is found at the tip of a flexible and highly exposed loop. The similarity with FMDV is striking and underscores the likelihood that FMDV binds to an integrin as the first step in cell entry. It may be that conservation of the RGD

motif within a flexible loop is necessary for good receptor interactions (e.g. allowing induced fit) - or it may just enable better antigenic camoflague of the conserved portion. The observed flexibility of the G–H loop of VP1 may thus facilitate *two* purposes: receptor binding and antigenic escape.

The limited structural evidence for C-S8c1 concerning the location of this loop does however indicate that the loop residues spend at least part of the time overlying VP2, near the viral 2–fold and away from the C–terminus of VP1 (Site *C*). This appears to differ from the native type O_1 structure where indirect evidence suggests that the loop residues are mainly located over VP1, near the viral 5–fold, close to the C–terminus of VP1 (Parry *et al.*, 1990). A different location of antigenic site *A* (the VP1 G–H loop) in FMDV of serotypes O and C provides a rational explanation for the finding of a single discontinuous site – involving both the VP1 G–H loop and the VP1 C–terminus – in type O (Kitson *et al.*, 1990) and of two independent, continuous antigenic sites in type C (manuscript in preparation and Chapter 3).

The absence of any detectable structural differences between SD6-6 and C-S8c1 suggests that mutations in the VP1 G–H loop, sufficient to cause a dramatic alteration in the antigenicity of the virus, do not require alteration of the main chain conformation to achieve antibody escape. This confirms our earlier work with type O_1 variants (manuscript in preparation and chapter 5) and work by others (Tulip *et al.*, 1991) with neuraminidase MAR mutants, which has shown that although the footprint of an antibody involves a large surface area structural differences in MAR mutants are often confined to changes in the substituted side chain which are however sufficient to abolish MAb binding.

6.2 New Structural Features

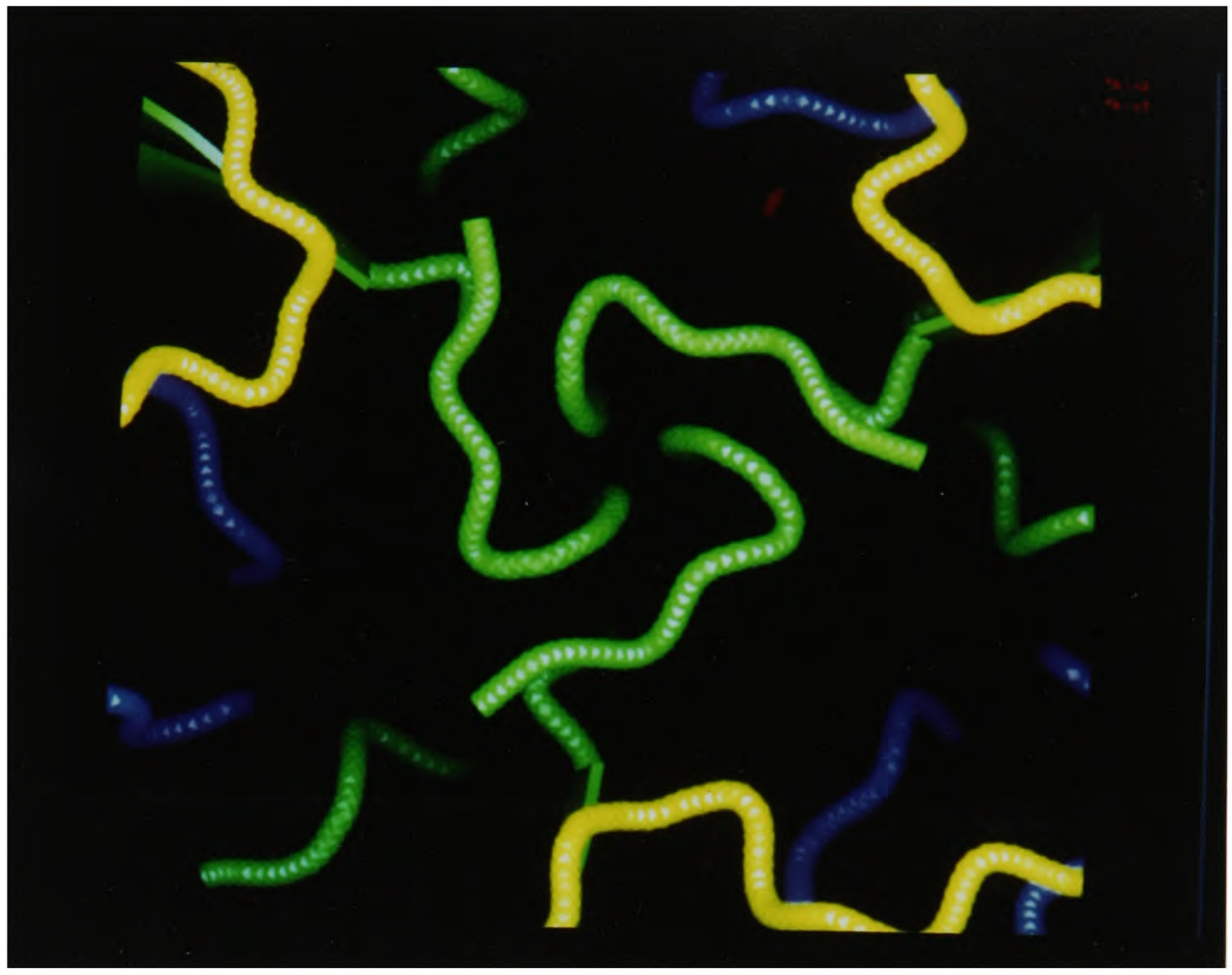
Despite the increased mobility of the atoms (as previously noted Figure 6.2) and limited resolution of the data (as compared to O_1BFS) several additional features have been identified in the electron density map which account for some regions of

the capsid structure that remained obscure in the O_1BFS analysis.

6.2.1 The N–terminus of VP2

Description of the Structure at the 3–Fold Axis

Extra density (not accounted for by a model based on O_1BFS) is seen around the icosahedral 3–fold axes on the inside of the capsid, extending from residue 9 of VP2 (the last residue included in the O_1BFS model) to the 3–fold axis and running back away from the 3–fold over the internal surface of VP2. A structure for these N–terminal residues has not been previously described for any picornavirus. Initially a model has been built for residues 9 to 1 (inclusive). The first 6 residues of VP2 run across the internal surface of the capsid towards the icosahedral 3–fold axis. Residues 6–9 then follow a path radially outwards (to a higher radius), parallel to, and close to the 3–fold axis, forming a tight core (Figure 6.3). Residues 10–12 loop back towards the interior of the capsid placing residue 12 between residues 3 and 27 from the same copy of VP2. Residue Asp 12 of VP2 (which is almost completely conserved amongst picornaviruses sequenced to date (Palmenberg, 1989)) seems important in stabilizing the conformation of this loop - making hydrogen–bonds via side and main chain atoms to residues 12 and 3 of the same copy of VP2 as well as to residue 5 of a symmetry related loop. This structure behaves reasonably well in refinement, although the chemistry is less than ideal and B–Factors for many atoms are still high (≈ 150Å^2). To attempt to highlight any errors in the model an omit map was calculated with residues 1–12 of VP2 omitted from the phasing model. This map seems to confirm the positioning of the main chain atoms as currently built although the density is somewhat diffuse (due largely to the obvious mobility of the protein in this region), and positioning of the side chains in the 'core' region running down the 3–fold must mainly be by consideration of steric clashes from symmetry related residues. The connectivity at residue Glu 6 is not certain (and this residue is currently modelled as a Gly due to lack of clear side chain density), however it seems to be confirmed by



View looking down the icosahedral 3–fold axis from the centre of the capsid, towards the internal surface.

View cut through the icosahedral 3–fold axis.

Figure 6.3: The N-terminus of VP2.

comparisons with the structures of mengovirus (Krishnaswamy and Rossmann, 1990) and poliovirus (Hogle *et al.*, 1985). In the structure of C-S8c1 residues 1–4 of VP2 run towards the 3–fold axis and in so doing fill a shallow depression on the internal surface of the capsid of the same protomer. In doing this the N–terminal residues of VP2 occupy a similar location to the N–terminal extension of VP1 in poliovirus (Hogle *et al.*, 1985) resulting in a 'smoothing' of the internal surface.

The model, at present, is not ideal – the packing of the three symmetry related strands is too tight to allow sensible stabilizing interactions between the protomers. However, we are confident about the connectivity, especially as data collected recently from a related C virus (CGC-Ger; see Appendix) to a higher resolution seems to confirm the overall structure presented here. This structure is of importance in considering the biology of FMDV and other picornaviruses as it allows, for the first time, visualization of both the N–terminus of VP2 and C–terminus of VP4 (Figure 6.4).

Cleavage of VP0 to form VP2 and VP4, releasing the N–terminus of VP2 and the C–terminus of VP4, is the final maturation stage for all picornaviruses and is associated with encapsidation of RNA[4]. In C-S8c1 the N–terminus of VP2 lies closest (14Å c.f. 31Å) to the C–terminus of a VP4 molecule which is associated with a symmetry related protomer i.e. with a different pentamer. In poliovirus empty particles where VP0 is uncleaved the VP2 N–terminus is covalently linked to the C–terminus of a VP4 molecule associated with the same protomer (personal communication J. Hogle). This would imply that in C-S8c1 the VP4 molecule to which the VP2 N–terminus was originally linked is the one which is seen to lie 31Å away in the mature capsid structure. The relative instability of the VP2 N–terminal structure compared to the C–terminus of VP4 suggests that it is the VP2 residues that are rearranged following cleavage – and that the structure described here for the N–terminal residues may not form unless cleavage has occurred and RNA encapsidation has filled the interior of the virion leaving no room for these residues except in the depression around

[4]Although 'empty' capsids lacking RNA in which the cleavage has occurred are frequently found in preparations of type A_{22} FMDV (personal communication S. Curry).

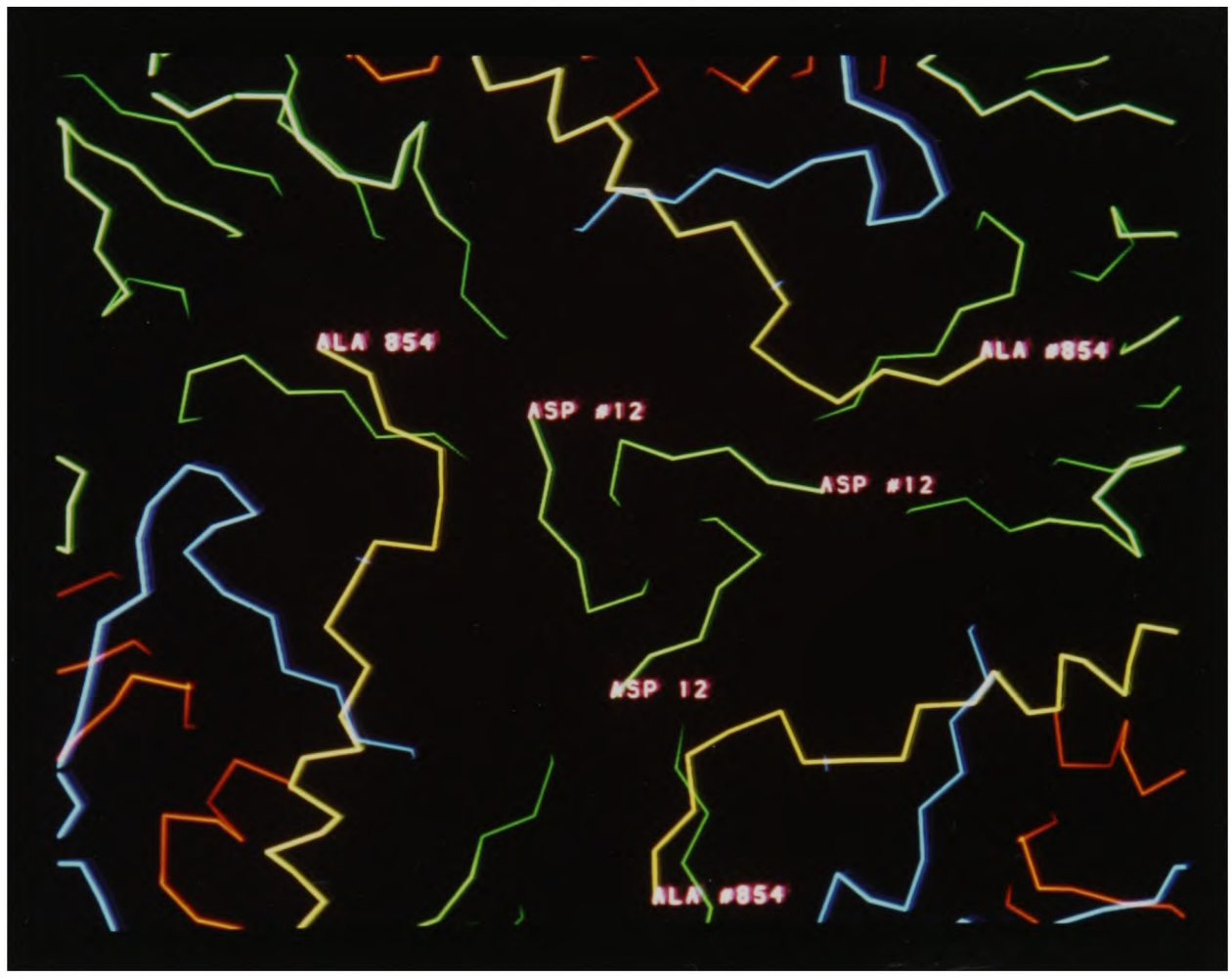


Figure 6.4: Positioning of N–termini of VP2 and C–termini of VP4.

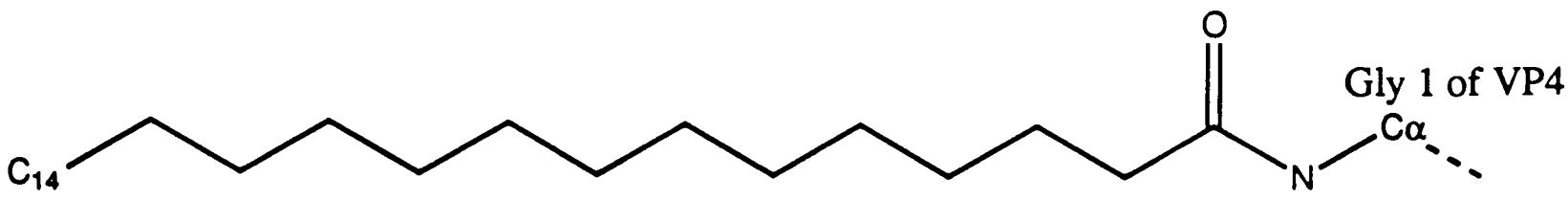


Figure 6.5: Attachment of myristate to VP4

the 3–fold. This may explain the relative absence of stabilizing interactions within this structure.

6.2.2 Extra Density Around the Interior of the 5–Fold Axis

Myristoylation in Picornaviruses

Post–translational modification of proteins is often of great importance in determining correct localization, stability and functioning. Many viral proteins are found to be acylated with fatty acids and this is commonly important for localization on the plasma membrane of cells. Picornaviruses are non–enveloped but must still interact specifically with the cell membrane at different stages during their life–cycle i.e. when entering the cell and during assembly of virus particles. Myristoylation of the N–terminus of VP4 has been demonstrated in all picornavirus genera (Chow *et al.*, 1987) – the myristic acid (*n*–tetradecanoic acid, shown in Figure 6.5) being covalently linked via an amide bond to the free α–amino group of the N–terminal Gly of VP4 (the immediate precursors of VP4, namely VP0 and P1, are also found to be myristoylated). Formation of this amide bond requires processing of the polypeptide either by removal of the initiating Met (rhinoviruses and enteroviruses) or by cleavage of a leader protein (cardioviruses and aphthoviruses). Myristoylation is commonly associated with such processing (Chow *et al.*, 1987) and may be directly coupled to the cleavage although detailed mechanisms are unknown. Myristoylation of FMDV will occur in the absence of all non–structural viral proteins and will even occur in insect cells (Belsham *et al.*, 1991) suggesting the mechanism for this modification is highly conserved. Myristoylation is associated with a consensus sequence revealed after the

protein processing, to be Gly–X–X–X–Ser/Thr (Chow *et al.*, 1987; Belsham *et al.*, 1991; Carr *et al.*, 1982; Aitken *et al.*, 1982) – mutagenesis of this sequence is lethal (Krausslich *et al.*, 1990; Marc *et al.*, 1989) – suggesting this modification is essential at some stage of the virus life–cycle. RNA transcripts of myristoylation–deficient mutants of poliovirus will replicate and produce processed products within cells but do not assemble efficiently (Marc *et al.*, 1990) suggesting a rôle in aiding viral assembly, possibly by localization of pentameric intermediates on the membrane of the RER.

The structure of this aliphatic moiety has been seen in only one picornavirus to date – poliovirus (Chow *et al.*, 1987; Hogle *et al.*, 1985), this structure is described below but one feature is the importance of Thr 28 of VP4 in stabilizing the conformation firstly by a side chain hydroxyl mediated hydrogen bond to the myristate carbonyl and secondly by hydrophobic interactions of the methyl group. Moscufo and Chow (1992) have probed the importance of these interactions by site–directed mutagenesis and have found that replacement of the Thr by either Lys or Gly is lethal resulting in production of non–viable virus. However, replacement with Ser (conserving the hydrogen bond) or Val (conserving the hydrophobic interactions) results in production of a viable virus – although these viruses are 'poorly' being less infectious (virus yield reduced 5- to 10–fold) and more sensitive to thermal inactivation, they also show defects in assembly leading to a build–up of assembly intermediates. These mutants confirm the importance of the myristate moiety in determining correct assembly of viable virions. Infection of cells with cDNA coding for the Thr→Gly (non–viable) mutant reveals that this has an assembly phenotype similar to the Thr→Ser and Thr→Val mutants i.e. some correctly formed virus particles are made (Moscufo *et al.*, 1993). These viral particles will also bind to the poliovirus receptor and undergo the conformational transition from 150S particles to 135S (i.e. $VP4^-$) particles – however no infectious particles are produced (Moscufo *et al.*, 1993). This suggests that in poliovirus the myristoylation functions to enable virus uptake into the cell at a stage after receptor binding and expulsion of VP4 from the particle has occurred.

However, this does not necessarily imply the same rôle for myristoylation in the uptake of FMDV into the cell since the route by which this occurs is quite different in these two viruses, for instance the acid lability of FMDV means that acidification of the virus–containing endosome is sufficient to cause release of the viral RNA.

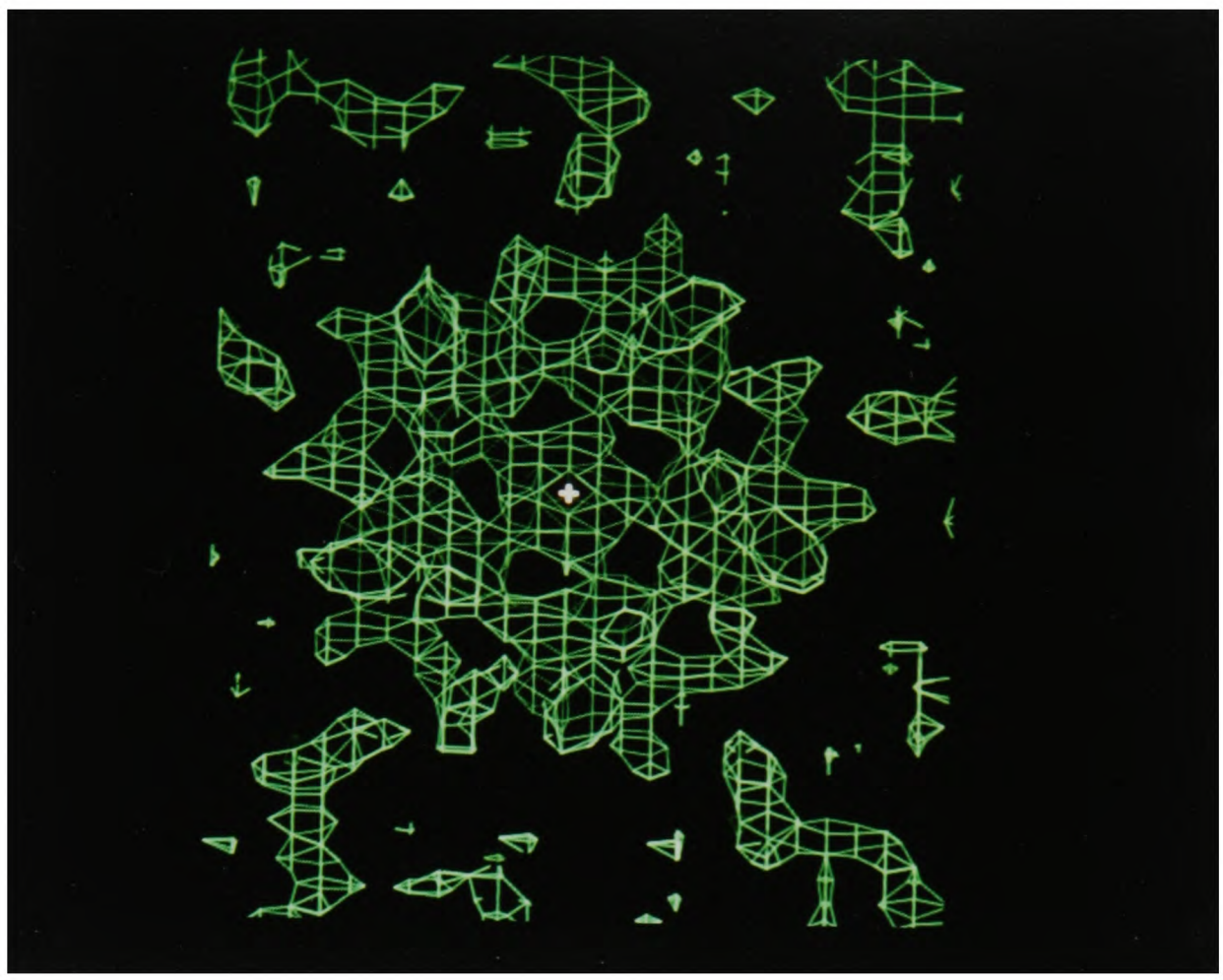
The Structure of the Myristate Moieties in Poliovirus.

The structures of Mahoney strain type 1 (Hogle *et al.*, 1985), and Sabin strain type 3 (Filman *et al.*, 1989) polioviruses have revealed a common structure for these aliphatic moieties. The N–terminus of VP4 is located on the inside of the capsid close to the icosahedral 5–fold axis. At the 5–fold the symmetry related copies of VP3 intertwine to form a twisted tube of parallel β–structure which is flanked on the inner capsid surface by 5 short, symmetry related pieces of two–stranded, anti–parallel, β–sheet formed by residues near to the N–terminus of VP4. The 5 myristyl groups extend from the N–terminus of VP4 initially forming a tight cluster near the 5–fold, then, splaying apart and running to a higher radius, to enclose the twisted tube formed by the N–termini of VP3. Each myristate group interacts with several residues in VP3 and VP4. As noted above an hydrogen–bond between Thr 28 of VP4 and the myristate carbonyl (of a symmetry related moiety) is of great importance in stabilizing this structure.

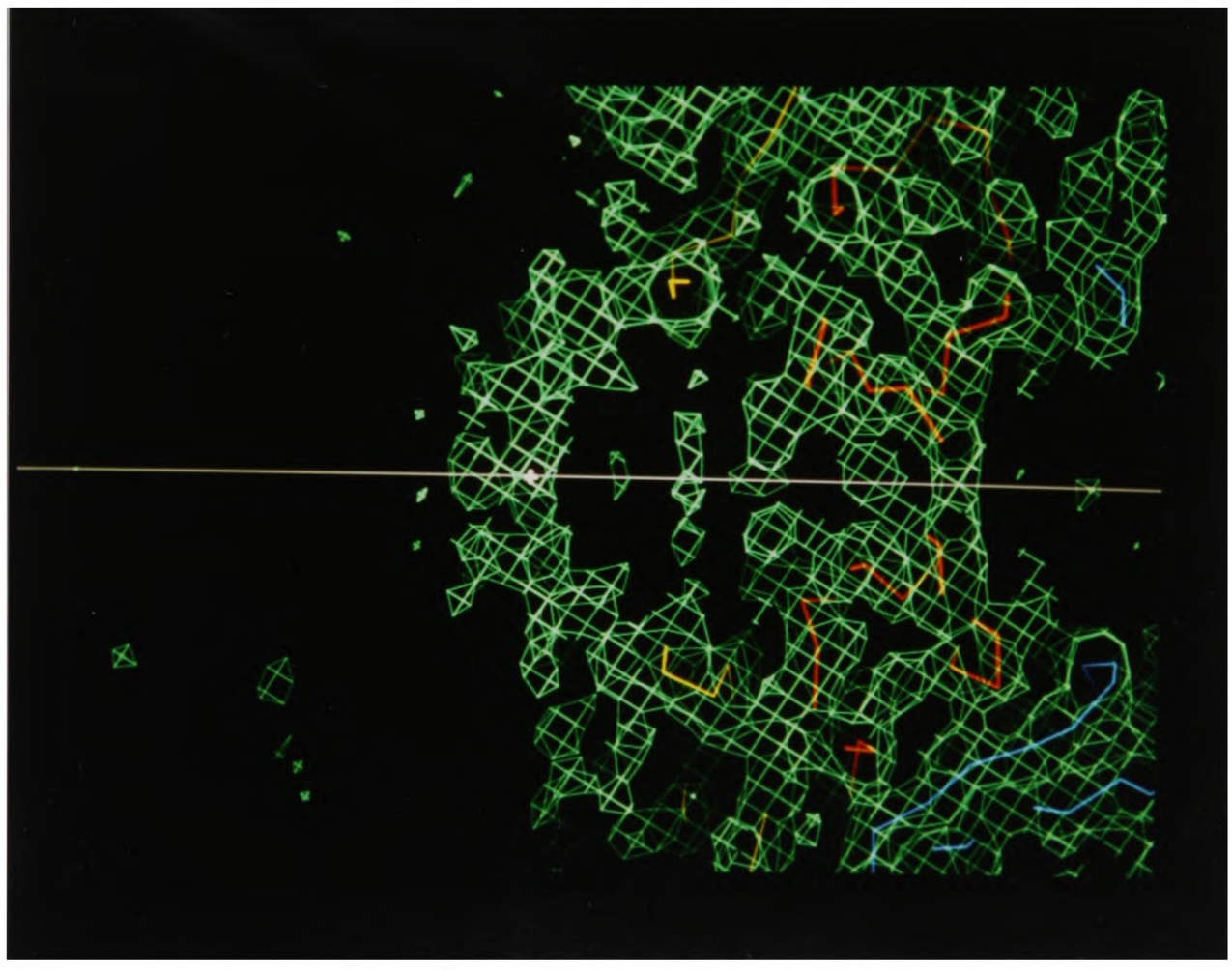
Density Seen in C-S8c1.

In C-S8c1 extra density (not accounted for by the model) is seen around the interior surface of the 5–fold axis. This density has obvious 5–fold symmetry – two views related by a 90° rotation are shown in Figure 6.6.

As noted by Fry *et al.* (1990) the 5–fold environment of FMDV is somewhat different from that seen in poliovirus due to a canting of the VP1 β–barrel, the strands of the barrel being shifted approximately 3Å (outwards) compared to poliovirus. The extra density observed (in the C-S8c1 map) is at the same radius as residue 1 of VP4 in the poliovirus model – the start of the myristyl moiety. Since in FMDV the other



View looking down the icosahedral 5–fold axis from the centre of the capsid, towards the internal surface.



View cut through the icosahedral 5–fold axis.

Figure 6.6: Density at interior of 5–fold axis.

proteins are then shifted outwards with respect to this position, there is more room to accommodate the hydrocarbon chains. At present we cannot distinguish between two possible models which could both account for the extra density. In the first explanation the myristate group is assumed to occupy the depression on the inner capsid surface around the 5–fold axis, as in poliovirus (the absence of density for this group being explained by the increased space available due to the canting of the β–barrels). The myristyl carbonyl may then be placed at a position equivalent to that seen in poliovirus where it can H–bond to the hydroxyl group of Asn 23 of VP4 in an interaction analogous to the interaction with Thr 28 of VP4 seen in the poliovirus model. The density observed may therefore be interpreted as residues 6-1 of VP4 (with residue 6 the most internal). However as no density can be seen for side–chains we cannot unequivocally identify this feature as protein. The shape of the density favours an alternative explanation whereby the density corresponds to the myristate chains forming a twisted barrel with the 5 copies related by the icosahedral 5–fold axis running inwards towards the virion centre i.e. in the *opposite* direction to that observed in poliovirus. The myristate may then be stabilized by non–specific interactions with the RNA and the carbonyl oxygen can still H–bond to residue 23 as described above. In both models the lack of continuous density for VP4 is explained by flexibility of the appropriate portion.

6.2.3 Conclusions

Both of these new structural features raise questions which it is difficult to answer at the current resolution limit of the data. Studies with a closely related C virus (CGC–Ger; see Appendix) which yields better diffracting crystals will hopefully shed more light on, and allow a fuller analysis of these regions.

It is not clear why these features are ordered in C serotype FMDVs when they are absent from O (Acharya *et al.*, 1989) and A (personal communications E. Fry and S. Curry) serotype maps – especially as the residues involved around the 3–fold and

5–fold are very highly conserved (the sequence of VP4 is completely conserved and there are no changes within the first 30 residues of VP2 between O_1BFS and C-S8c1). It is noteworthy that the only other picornavirus structure determination which has revealed details of internal features (i.e. poliovirus (Hogle *et al.*, 1985; Filman *et al.*, 1989)) has involved cryo–cooling which may help to increase order in these mobile regions. As such, cryo–cooling might provide a method to obtain more detail in these features in C-S8c1 – although construction of a protocol for cryo–cooling, that maintains disease security for FMDV, would make collection of such data complex and initial studies with O_1BFS were unpromising (D. Stuart, unpublished).

6.3 Structural Predictions of Antigenicity.

Antigenic sites on viruses are generally defined using monoclonal antibodies (MAbs) to apply pressure to virus *in vitro* to produce escape mutants. Mapping of these mutations to both the primary sequence and to the three–dimensional structure generally results in spatial clustering of mutations. Competition between different MAbs leads to definition of these clusters as separate sites with there being little or no competition between MAbs directed at separate sites. These separate sites are seen to correspond to spatially distinct clusters of mutations on the virion surface the distance between the different sites explaining why an antibody binding to one site has no effect on antibodies binding to the other sites. Production of more and more escape mutants eventually builds up a picture of the overall immunogenicity of the virus. As there is a strong element of chance in the generation of these mutants we have wondered whether the three–dimensional structure of a virus might be used to directly predict which residues would be involved in antibody binding.

6.3.1 Predictions Based on Accessibility.

The surface of FMDV is relatively smooth and this means that almost all solvent accessible residues are potentially accessible to attacking antibodies. This differs

greatly from the structure of viruses such as HRV (Rossmann *et al.*, 1985) which contain canyons large enough to permit access of solvent molecules and other small objects but too narrow to permit access of large antibody molecules. We therefore wondered if classifying residues on the basis of accessibility would prove an accurate method by which to predict the locations of antigenic sites in FMDV.

Method.

The accessible surface area of C-S8c1 was calculated using the method of Lee and Richards (1971). The accessible area is defined as the continuous path of the centre of a spherical probe as it is rolled over the van der Waals surface of the protein. For multimeric proteins the calculation of accessible area may be made for the unique portion, but must be made in the context of the whole assembly. For FMDV this may be achieved by making the calculation for an icosahedral subunit in the context of a fringe of residues (fringe radius > 2× probe radius) from symmetry related units. To ensure that the internal surface, which must necessarily be inaccessible to antibodies in the intact virion, is not considered, residues are only considered potentially external if they are at a capsid radius greater than 125Å. Residues are considered probe accessible if their accessibility in the C-S8c1 structure exceeds 10% of the amino acid accessibility in a Gly–X–Gly extended tripeptide. Appropriate tripeptides were built for the 20 amino acids using the β–extend option in FRODO (Jones, 1985). Novotńy *et al.* (1987) have previously used probes of radius 10Å to mimic interactions between immunoglobulins and protein antigens, however, for this analysis we are not only interested in those residues which are accessible to an immunoglobulin in C-S8c1, but also in those which, when mutated, might become accessible and make contact with an antibody. Therefore a probe size of 2Å is used.

Results.

Those residues which are considered accessible based on the parameters defined above are marked on Figure 6.7 by an asterisk beneath the C-S8c1 sequence. The residues

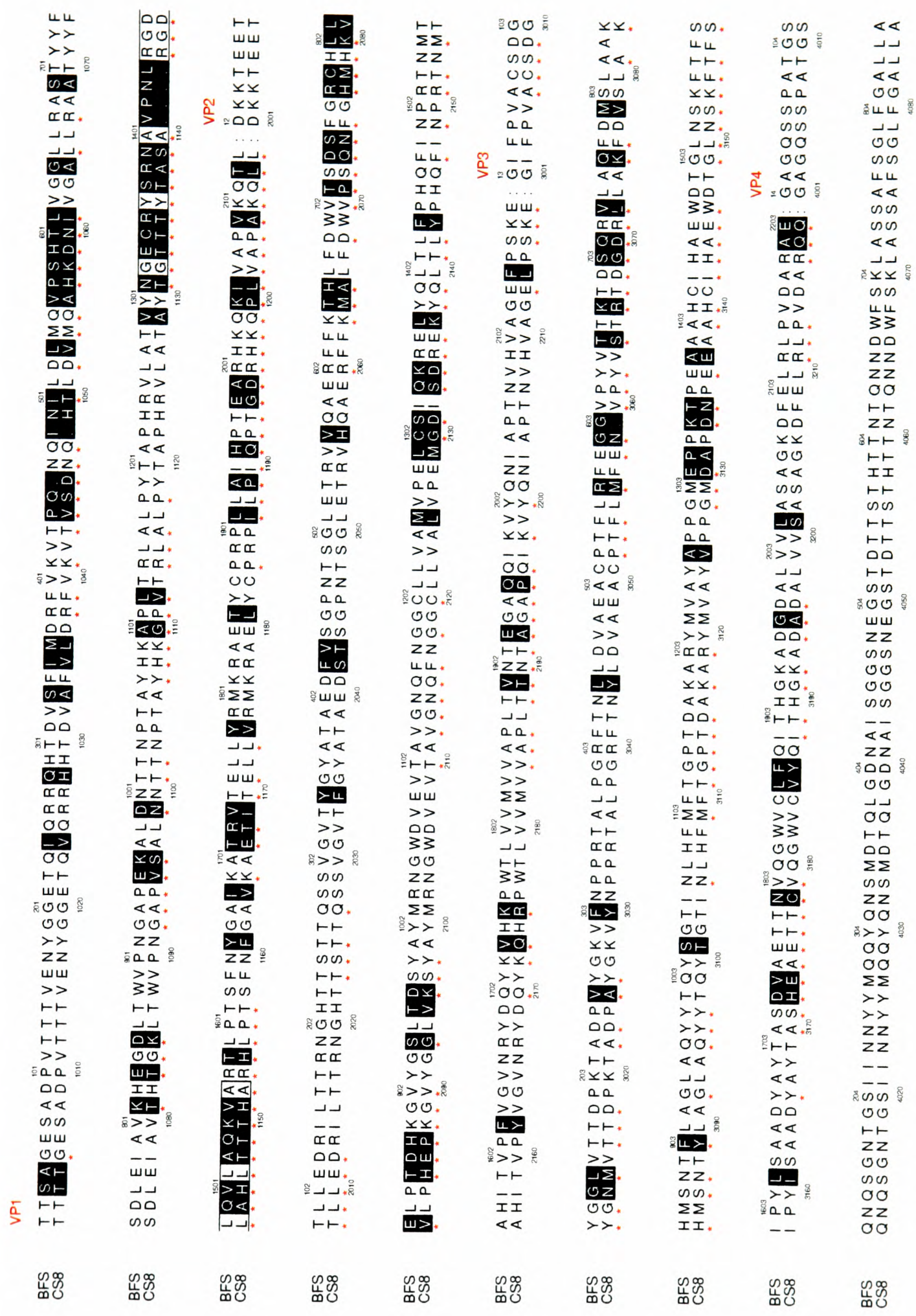


Figure 6.7: Sequence alignment of C-S8c1 and O_1BFS.

located within the disordered VP1 G–H loop are also marked as accessible since all the available structural evidence (Acharya *et al.*, 1989; Parry *et al.*, 1990; Logan *et al.*, 1993) suggests these residues will be surface exposed.

Unsurprisingly all the antigenic sites (as defined by MAb escape mutant mapping studies (Mateu *et al.*, 1988; Mateu *et al.*, 1989; Martínez *et al.*, 1991; Martínez *et al.*, 1992)) contain residues identified as accessible by these criteria. The extent of these sites is increased by considering accessibility. This ties in with the observation that as more and more escape mutants are mapped the areas of antigenic sites tend to become extended and the distinctions between different sites blurred, the limits placed on a site at any time simply representing the incompleteness of the Mab data.

Other regions defined as accessible fall outside of the antigenic sites defined in type C FMDVs but within regions defined as antigenic in types O and/or A viruses. The finding of accessible residues within these regions in type C virus suggests that these regions are antigenic in the C viruses and are simply not identified as such due to the incompleteness of the MAb data. One such region which contains accessible residues but is not currently identified as an antigenic site in type C is the B–C loop of VP1. This region is defined as Site 3 in type O (Barnett *et al.*, 1989; Kitson *et al.*, 1990; Parry *et al.*, 1989; Pfaff *et al.*, 1989; Stave *et al.*, 1988; Xie *et al.*, 1987) and type A (Bolwell *et al.*, 1989; Baxt *et al.*, 1989; Saiz *et al.*, 1991; Thomas *et al.*, 1988) FMDVs and is also antigenic in HRV (Kim *et al.*, 1989) and poliovirus (Page *et al.*, 1988; Hogle *et al.*, 1985). Table 6.1 shows an alignment of 15 sequences of C serotype field variants[5] in this region. The residues are marked:

- To show the secondary structural element they are part of[6].
- With a (–) is they meet the accessibility criteria.
- With a ($\triangle$) if escape mutations have been identified at that position in type O or type A viruses.

[5]Sequences kindly provided by Dr. E. Domingo, Centro de Biología Molecular, Madrid, Spain.
[6]Defined using program DSSP (Kabsch and Sander, 1983)

Virus			βB						B-C loop							βC				
CGC	V	L	D	R	F	V	K	V	Q	V	S	G	N	Q	.	H	T	L	D	I
CGCC1	V	L	D	R	F	V	K	V	Q	V	S	G	N	Q	.	H	T	L	D	I
CS8C1JH	V	L	D	R	F	V	K	V	T	V	S	D	N	Q	.	H	T	L	D	V
CS8CE	V	L	D	R	F	V	K	V	T	V	S	D	N	Q	.	H	T	L	D	V
C1OBB	V	L	D	R	F	V	K	V	T	V	S	G	N	Q	.	H	T	L	D	V
C1HLOIRE	V	L	D	R	F	V	K	V	T	V	S	G	N	Q	.	H	T	L	D	V
C1SDARO	I	L	D	R	F	V	K	V	P	V	S	G	N	Q	.	H	T	L	D	V
C384	V	L	D	R	F	V	K	V	P	V	L	D	G	Q	Q	H	T	L	D	V
C3RES55a	V	L	D	R	F	V	K	V	P	V	S	D	R	Q	Q	H	T	L	D	V
C5ARG69	V	L	D	R	F	V	K	V	P	V	S	G	G	Q	Q	H	T	L	D	V
C3RES55b	V	L	D	R	F	V	K	V	P	V	S	D	R	Q	Q	H	T	L	D	V
C3IND78	V	L	D	R	F	V	K	V	H	V	S	G	N	Q	.	H	T	L	D	V
C385	V	L	D	R	F	V	K	V	Q	V	S	G	N	Q	.	H	T	L	D	V
C2PANDO	V	L	D	R	F	V	K	V	N	V	S	G	N	Q	.	H	T	L	D	M
C4TDF66	V	L	D	R	F	V	K	V	K	A	S	G	N	Q	.	H	T	L	D	V
Accessible Residues				-	-		-		-	-	-	-	-				-		-	
MAb Escape Mutation Positions							Δ	Δ		Δ				Δ						

Table 6.1: Variation in the field within the B–C loop of VP1 in type C FMDVs.

From the alignment it can be seen that C sequences vary significantly at several positions in this region it is therefore very probable that these accessible residues which form an antigenic site in FMDVs of types O and A are also part of such a site in type C viruses. Further extension of Mab–escape mutant studies is likely to eventually identify this region as antigenically important.

There remain about 30 residues identified by these criteria as potentially accessible but falling outside presently identified FMDV antigenic sites. Two such regions include the D–E loop of VP2 and the C–terminus of VP3. The primary sequence of the D–E loop of VP2 is completely conserved amongst the 15 field variants of type C FMDV suggesting that mutations in these residues do not form a mechanism of antigenic escape in the field. This may be because the residues fulfil some structural rôle, although this region does not seen more obviously structurally constrained than other surface exposed regions but a more likely explanation involves the G–H loop of VP1. This loop is not included in the model used to calculate the accessibilities, however, the limited structural evidence we have concerning its location, suggests

that for a least part of the time it lies over VP2 (including over the D–E loop). Taken with the lack of sequence variation (more typical of internal regions) this suggests that access of an antibody to these residues is hindered by the mobile G–H loop of VP1 and they therefore do not form part of an antigenic site. Their classification as accessible simply reflects a deficiency in the model.

The C–terminus of VP3 shows great sequence variation (especially at position 219), and this, combined with its exposed location suggests that it may be a target for antibody binding. The flexibility of the terminus may however decrease the frequency with which antibodies bind so reducing the likelihood of isolating MAb–escape mutants with mutations within this region. Accessibility is therefore seen as a fairly good criterion on which to base predictions of residues likely to be involved in immunoglobulin binding, although it is more difficult to predict which residues will have been identified by MAb–escape mutant mapping.

6.3.2 Predictions Based on Inter–Serotypic Structural Differences.

For viruses like FMDV, against which the humoral response constitutes the main immune defence, it is important that areas of the capsid liable to attack by antibodies should be capable of antigenic change. In principle this can be achieved either by substitutions of antibody–contact residues or by changes in protein conformation. Since such changes are usually associated with each other, we might expect antigenic sites to be correlated with regions of greatest conformational divergence between two serotypes.

Figure 6.2 confirms this expectation, in that all the sites previously identified in serotype C viruses (Mateu *et al.*, 1989; Marc *et al.*, 1990; Martínez *et al.*, 1992), except D_2 (see below), coincide with regions of greater than average RMS deviation in Cα position, although in the case of Site A (the G–H loop of VP1), the difference is not

quantifiable owing to the loop's disorder in C-S8c1. In addition to these regions the B–C loops of VP1 and VP3 and the E–F loop of VP2, which are not recognised as parts of antigenic sites in C serotype FMDVs, exhibit very large conformational differences. Interestingly, in serotype O and A strains, antigenic residues have been identified in all of these loops (Kitson *et al.*, 1990; Thomas *et al.*, 1988). As noted above regions likely to be antigenically important may also be identified by examining sequence variation occurring in the virus in the field. Some of this variation will be connected with non–immune pressures and with random, allowed, mutations that have no deleterious effect on the virus, however the areas of greatest variability are likely to be those involved in antigenic sites which are mutating in response to pressures applied by the host immune system. The B–C loops of VP1 and VP3 and the E–F loop of VP2 are all found to be variable in such sequence alignments, thus, if all the avaliable information on antigenic sites in FMDV is taken into account, variability in conformation of surface regions is seen to be an excellent predictor of potential antigenic change. Whether the serotypic differences in antigenicity as defined by escape mutants reflect actual immunological preferences, or simply incompleteness of the MAR mutant data is unclear.

Links between protein mobility (as measured by crystallographic B–factors) and antigenicity have previously been made (Geysen *et al.*, 1987; Getzoff *et al.*, 1988; Westhof *et al.*, 1984). These are complicated by the fact that antigenic residues will be surface oriented and so less tightly constrained by protein–protein interactions. They will therefore be more mobile than the core protein regions. The strong correlation between the main chain B–factors for C-S8c1 and O_1BFS (see above and shown in Figure 6.2) indicates that even with X–ray data of limited resolution we may interpret B–factors (as indicative of relative structural mobility), with confidence. As for conformational change, there is some correlation between areas with greater than average B–factors and positioning of antigenic sites. As discussed and quantified above, mobility and conformational change are linked by the fact that these will be characteristic of less conformationally constrained regions of the protein e.g. surface exposed loops. This lack of constraint gives greater tolerance to different amino acid

substitutions facilitating antibody escape.

Site D_2

There remains the sole exception of site D_2, located in the B–C loop of VP2, which, while antigenic in both C and type O, does not exhibit an RMS deviation in Cα positions significantly above the level of experimental accuracy of the coordinates. Although the main chain protein conformation in this region, is conserved between the two viruses the sequences are very different and the site will therefore appear different (in terms of surface shape and chemical nature) to an antibody attempting to recognize this region. A careful analysis of the hydrogen–bonding pattern in this site demonstrates that antigenic change is accomplished by mutations occurring at positions with surface oriented side chains (Table 6.2) not involved in interactions with the surrounding protein (Figure 6.8). In this site there is only one position where the side chain forms an H–bond (His 79) at which a substitution is seen. The only replacement seen at this positions is His→Arg and modelling of this change suggests the H-bond can be conserved. This analysis may provide an explanation for the observation that when MAR mutants of viruses are selected the same amino acid replacement is repeatedly found in mutants obtained from independent clonal pools of a virus (Mateu *et al.*, 1989). Previous explanations of the phenomenon have suggested that those residues repeatedly substituted for are "key" residues (Air *et al.*, 1990) in the interaction with the antibody and/or the need to produce the MAR phenotype as a result of a single base change (point mutation). The structure of C-S8c1 site D_2 suggests that an alternative explanation may be limited tolerance to amino acid change due to structural constraints within the capsid operating to filter those substitutions allowed in response to the selective pressure.

Residue Number (VP2)[a]	Residue Type	H_2O Accessibility[b]
70	Val	39.0
71	Pro	33.6
72	Ser	92.8
73	Gln	25.5
74	Asn	89.3
75	Phe	31.3
76	Gly	5.8
77	His	79.0[c]
78	Met	35.2
79	His	52.0
80	Lys	76.1[d]

a - Residues 72, 74 and 79 are found mutated in MAR mutants.
b - Accessibilities calculated by the method of Lee and Richards (Lee and Richards, 1971). A probe of radius 1.4Å is rolled over the surface of the protein - the accessible area being defined as the continuous path of the centre of this probe.
c - There seems no structural reason for this residue to remain unaltered when antigenic pressure is applied.
d - Although the accessible area for this residue is large the residue is not surface directed; it lies across the capsid surface (a fully exposed Lys would have an H_2O accessible area of 218.5Å ^{2}. Access to this residue by a large molecule such as an antibody may also be restricted by residue 131 (VP2) from a neighbouring loop, although this residue does not lie sufficiently close to restrict access of the H_2O probe.

Table 6.2: H_2O Accessibilities of residues in B–C loop of VP2 (Site D_2).

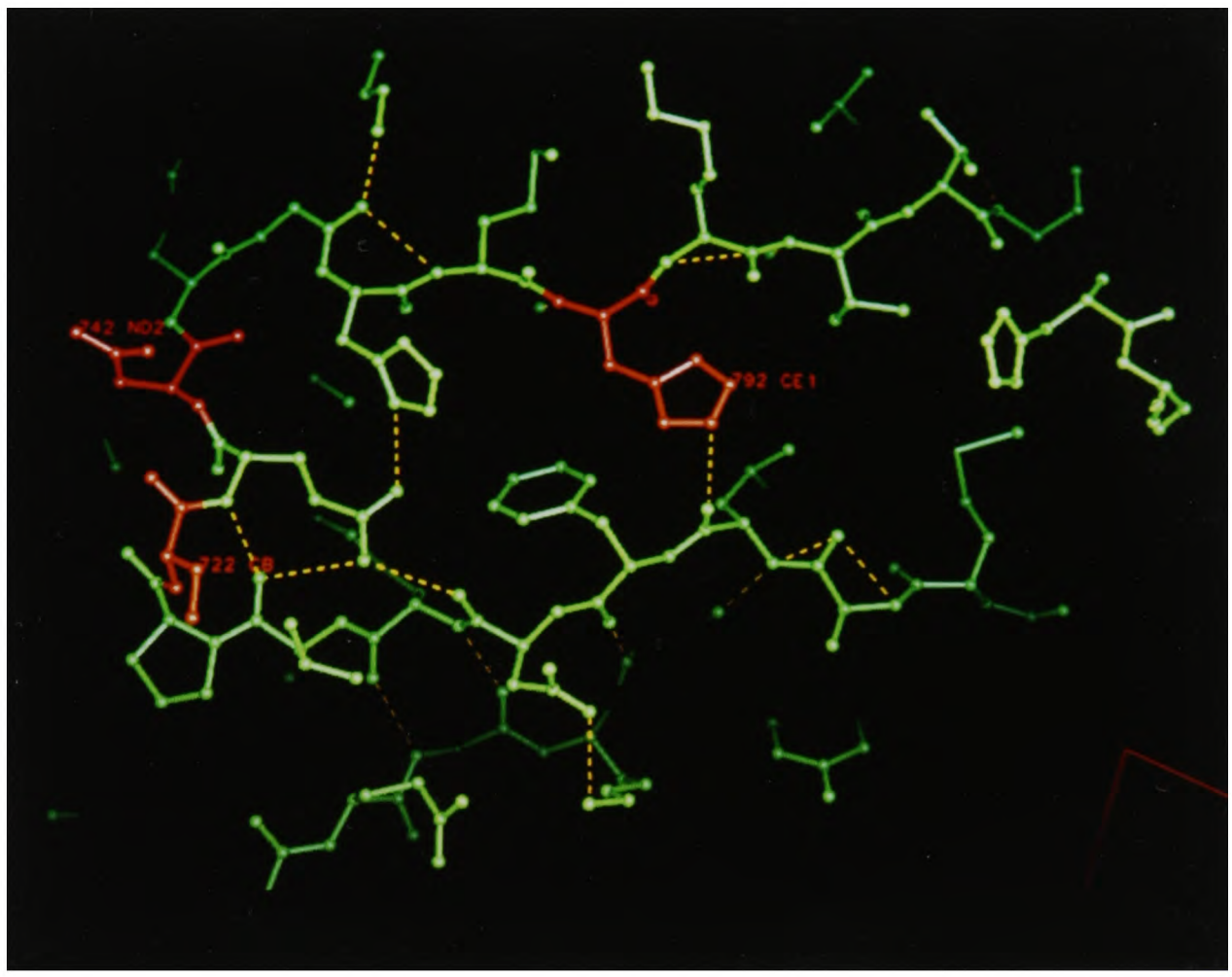


Figure 6.8: Hydrogen bonding pattern in site D_2.

6.3.3 Conclusions

To conclude, structural information in the form of accessibility analyses or conformational change between serotypes can be used to predict antigenic regions, and even to pinpoint specific residues capable of undergoing antigenic change. The accuracy and usefulness of such predictions can be dramatically improved by considering many different sources of information including these structural predictions and also sequence variation. These methods will not necessarily predict which sites have been identified by the *in vitro* techniques (we must recognize the limitations of the existing MAR mutant data) but may provide a more biological description of the antigenic nature of the virus. The smooth capsid of FMDV, with no canyons of pits separating antigenic protrusions, may render nearly the entire surface accessible to antibodies. Extension of the Mab mapping studies may therefore eventually blur the distinctions between the different sites and favour the view that the entire FMDV surface accessible to antibodies may prove antigenic to some extent. This concept is supported by previous studies with lysozyme (Davies and Padlan, 1990) and the human growth hormone (Jin *et al.*, 1992).

6.4 Future Methods of Comparison.

Making meaningful, quantitative comparisons of complex 3–dimensional, charged objects such as proteins is difficult to achieve. The basic quantitative comparison presented here is the calculation of the RMS deviation between the Cα positions in the two viruses. This is a crude measure of conformational change – changes at this level are fairly 'gross' changes in terms of altering how the virus appears to other proteins, it is also dependent on a one–to–one correspondence between residues in the two structures being compared. This is fairly true between two such closely related viruses as O_1BFS and C-S8c1, but will tend to exaggerate the structural differences between viruses where different portions of the linear sequence fulfil similar structural rôles. In designing a better method several concepts must be considered:

- The overall 'shape' – i.e. the volume occupied by the viral atoms should be considered rather than comparisons based on the linear sequence, this would allow better comparisons of regions in which complementary mutations have occurred which would radically alter a residue–by–residue comparison but have little net effect on the overall shape of the virus.

- Charge distribution – we now have structures of representatives of 3 FMDV serotypes, O (O_1BFS), C (C-S8c1) and A (A_{22}). These all show great similarity in the conformations of even surface exposed loops. Comparisons based on charge distribution over the surface of these viruses (e.g. that calculated by DELPHI (Gilson *et al.*, 1987)) might reveal either common patterns of usage giving clues to receptor binding and other conserved functions, or differing patterns of use, giving further clues to the nature of antigenic escape and serotypic characteristics.

An ideal method of comparison would provide a three–dimensional mapping of both conformational/volume changes and charge changes allowing comparisons to be made in a way that is entirely dependent on the structure and independent of the linear sequence.

Chapter 7

Analysis of Antigenic Surfaces of Proteins

7.1 Introduction

The complex network of reactions occurring when a foreign substance / organism exposes itself to the immune system of a host may be considered, at a most basic level, as consisting of a series of protein–protein interactions with wide–ranging effects. These interactions will be between foreign proteins and host proteins and also host–host and foreign–foreign interactions. The host's immune system will attempt to clear the body of the invading organism by specific destructive interactions mediated by B cells (i.e. mechanisms involving immunoglobulins) and T cells (e.g. $CD8^+$ T cells – 'killer' T cells). Such specific interactions must, in the end, be based on recognition of one foreign protein (or peptide fragment in the case of T cell mediated interactions) by another, host protein. One part of this immune recognition involves interactions between native, foreign proteins and host immunoglobulins. The interactions involve surface exposed residues of both halves of the recognition event and therefore must impose differing evolutionary pressures on the proteins' surfaces. The foreign effect will be as if the protein were attempting to maintain a surface that minimises the likelihood of recognition by the immunoglobulin whereas the host attempts to produce immunoglobulins capable of specifically recognizing many different foreign proteins (i.e. it will attempt to maximise the likelihood of the recognition event occurring). Since protein–protein interactions must be determined by the amino acid composition of the proteins' surfaces (even if there is some sort of 'induced fit' occurring, the initial

interaction must be between the two surfaces seen in the uncomplexed proteins) we wondered whether the pattern of amino acid use on the surfaces of proteins that need to evade the immune system to survive would differ from the patterns of use found on the surfaces of proteins involved in immune recognition.

7.2 Method

Amino acids are of different sizes, therefore simply comparing the numbers of different amino acids expressed on the surface of a protein gives no real clue as to the nature of that surface. The 'accessible surface' area of a protein may be calculated by the method of Lee and Richards (1971) – a spherical probe is rolled over the protein surface (generated by applying suitable van der Waals radii to a set of atomic coordinates), the accessible area is defined as the continuous path of the centre of this probe[1]. Previous studies of antigenicity using spherical probes (Novotńy *et al.*, 1986; Novotńy *et al.*, 1987) have used the algorithm of Lee and Richards to calculate a 'contact area' – a set of discontinuous patches on the antigen representing those portions of the atomic surfaces that are in contact with the surface of the probe as it is rolled over the antigen. Although accessible areas and contact areas are directly related the latter seems less biologically relevant, leading to the perception of the protein surface as a series of either completely accessible or inaccessible regions. The surface defined by the accessible area (providing an appropriate probe size is chosen) seems to better mimic the antigenic surface as it would be 'seen' by another protein. By comparing the percentage of the total accessible area occupied by one amino acid type relative to other types meaningful comparisons of the surface nature may be made.

[1]From this definition it is apparent that the radius of the probe used will directly alter the absolute value of the accessible surface area (see below).

7.2.1 Selection of Protein Classes

The surface of a protein will not be made up of equal proportions of all 20 amino acids – even in the absence of immune pressure – as most proteins exist in an aqueous environment. This provides a very strong evolutionary pressure and gives a selective advantage to those proteins displaying hydrophilic residues (able to make energetically favourable interactions with the surrounding solvent), compared to those displaying hydrophobic residues. Charged residues such as Lys and Glu will therefore cover a disproportionate fraction of the surface. The effect of this aqueous environment in determining the use of amino acids must be 'subtracted' when looking for patterns imposed by other evolutionary pressures, three classes of proteins have therefore been studied:

- **Proteins subject to immune pressure (Group I)** – picornaviruses, as animal viruses, need to evade the host's immune system to survive. Clearance of the virus from the host is mainly mediated by specific recognition of the virus by immunoglobulins. Structures of several picornaviruses have been determined at high resolution and reveal in atomic detail the protein surface involved in this recognition event. Residues from all three of the larger coat proteins (VP1, VP2 and VP3) are seen to be surface exposed and therefore potentially involved in such an interaction. Mutations provoked by immunoglobulin pressure are commonly mapped to such exposed residues (Kitson *et al.*, 1990; Rossmann *et al.*, 1985; Page *et al.*, 1988; Hogle and Filman, 1989; Acharya *et al.*, 1989) confirming surface exposure as a criteria for involvement in immune interactions. The virus residues are divided into two classes on the basis of an average capsid radius, those at a higher radius are considered external (Ia) and subject to immune pressure, those at a lower radius internal (Ib) and not subject to immune pressure.

- **Proteins involved in immune system recognition (Group II)** – immunoglobulins are responsible for one type of immune recognition. Many high

resolution structures are known both of isolated immunoglobulin fragments and also of complexes of such fragments and antigen. However, in such complexes, only a limited subset of the surface exposed residues are seen to be directly involved in the interaction with the antigen (Davies *et al.*, 1989; Satow *et al.*, 1986; Amit *et al.*, 1986; Tulip *et al.*, 1991) (these residues are also those which are seen to vary most between different immunoglobulins), this subset of residues lie within a single spatial location although different parts of the linear sequence are involved. Therefore for this analysis, only these residues from the complementary determining regions (CDRs) which are directly involved in immune recognition are considered. The 7 immunoglobulin structures used were aligned using SHP (D. Stuart, unpublished) so that structurally equivalent residues could be selected from each immunoglobulin, with the results of the alignment checked visually using a graphics system.

- **Control (Group III)** – proteins not subject to immune pressure are taken as a control to determine the effect of the aqueous environment. Those proteins not usually exposed to a foreign immune system, or involved in immune recognition are used.

For all these groups only those proteins for which well refined, high resolution structures were available (either from the Brookhaven Protein Data Bank or by personal communication) were studied. For proteins which normally exist as oligomers the calculations were made for the oligomeric complexes, or for the viruses for the unique portion of the complex. Table 7.1 summarizes the details of the protein structures used.

Two plant viruses were also analysed (Table 7.1), the lack of any known plant immune system might cause them to more resemble the control group of proteins in their amino acid use, or reveal some different pattern. Unfortunately there was much less consistency between these two viruses than between any other two members of the same group, therefore no general pattern for this group may be discerned.

GROUP	PDB FILE NAME	DESCRIPTION
GROUP I CONTROLS 3870 amino acids	PDB7API.ENT (Engh *et al.*, 1989)	α antitrypsin
	PDB1CA2.ENT (Eriksson *et al.*, 1988)	carbonic anhydrase II
	PDB3CLN.ENT (Babu *et al.*, 1988)	calmodulin
	PDB8DFR.ENT[a]	dihydrofolate reductase
	PDB1MBO.ENT (Phillips, 1980)	myoglobin
	PDB9PAP.ENT (Kamphuis *et al.*, 1984)	papain
	–	PGK[b]
	PDB5TNC.ENT (Herzberg and James, 1988)	troponin C
	–	phosphorylase R state dimer[c]
	PDB1GD1.ENT (Skarżyński *et al.*, 1987)	glyceraldehyde 3 phosphate dehydrogenase
GROUP II PICORNAVIRUSES 3154 amino acids	PDB4RHV.ENT (Arnold and Rossmann, 1988)	HRV 14
	–	reduced FMDV (Logan *et al.*, 1993)
	PDB2MEV.ENT (Luo *et al.*, 1987)	MENGO
	PDB2PLV (Filman *et al.*, 1989)	POLIO SABIN 2
GROUP III IMMUNOGLOBULINS 582 amino acids	PDB1MCP.ENT (Satow *et al.*, 1986)	mouse IgA V_κ
	PDB2IG2.ENT (Marquart *et al.*, 1980)	human myeloma IgG1 V_λ
	"GLOOP 1"[d]	mouse IgG2b V_κ
	"GLOOP 2"[d]	mouse IgG2a V_κ
	PDB1F19.ENT (Lascombe *et al.*, 1989)	mouse IgG2b V_κ
	PDB2FBJ.ENT[e]	mouse IgA V_κ
	PDB3FAB.ENT[f]	human IgG1 V_λ
OTHER VIRUSES	PDB4SBV.ENT (Silva and Rossmann, 1985)	SBMV
	–	CMV[g]

[a] Davies, J.F. *et al.*; [b] personal communication K. Harlos / C.C.F. Blake, Oxford; [c] personal communication K. Woods, Oxford; [d] personal communication G. Taylor, Bath; [e] Bhat, T.N. *et al.*; [f] Poljak, R.J. *et al.*; [g] personal communication E. Fry, Oxford.

Table 7.1: Protein structures studied

7.2.2 Selection of Probe Size

The accessible surface area of a protein as determined by the method of Lee and Richards (1971) will vary according to the probe size used (see definition above), as will the atoms considered to be contributing to the surface. A probe with radius 1.4 Å may be used to generate the solvent accessible area (such a probe size is comparable to that of a molecule of H_2O), however, as we are interested in an interaction between two large protein molecules rather than a small water molecule and a protein, use of such a small probe will implicate residues as part of the surface that will not participate in the protein–protein interaction. Novotńy *et al.* (Novotńy *et al.*, 1986; Novotńy *et al.*, 1987) have previously used a probe with radius 10 Å (see above) to mimic the interaction between an immunoglobulin and an antigen, however, our results with monoclonal–antibody escape mutants of FMDV (Chapters 3 and 5) and the results of Tulip *et al.* (1991) with neuraminidase escape mutants have shown that single, conservative side chain substitutions where very little conformational change occurs can result in an entire antigenic site becoming unrecognizable. This suggests a probe with a smaller radius may be appropriate. Trials with probes varying between 1.4 and 10 Å radius were made. In each case the accessible surface area of DTT soaked FMDV O_1BFS (Logan *et al.*, 1993) was calculated and the percentage of the total accessible area (not the absolute area) occupied by different amino acids compared. Figure 7.1shows that the percentage of the total accessible area occupied by 5 typical amino acids is remarkably constant for probes of different radii. Since the mutant virus structures warn us that the conformational changes needed to escape immune recognition (and therefore by inference to result in immune recognition) are small a probe of radius 5 Å is used for this analysis.

7.2.3 Selection of Cut–Off

To further clarify the patterns of use an accessibility cut-off is imposed – this attempts to exclude residues with only one or two atoms partially exposed to the probe from the

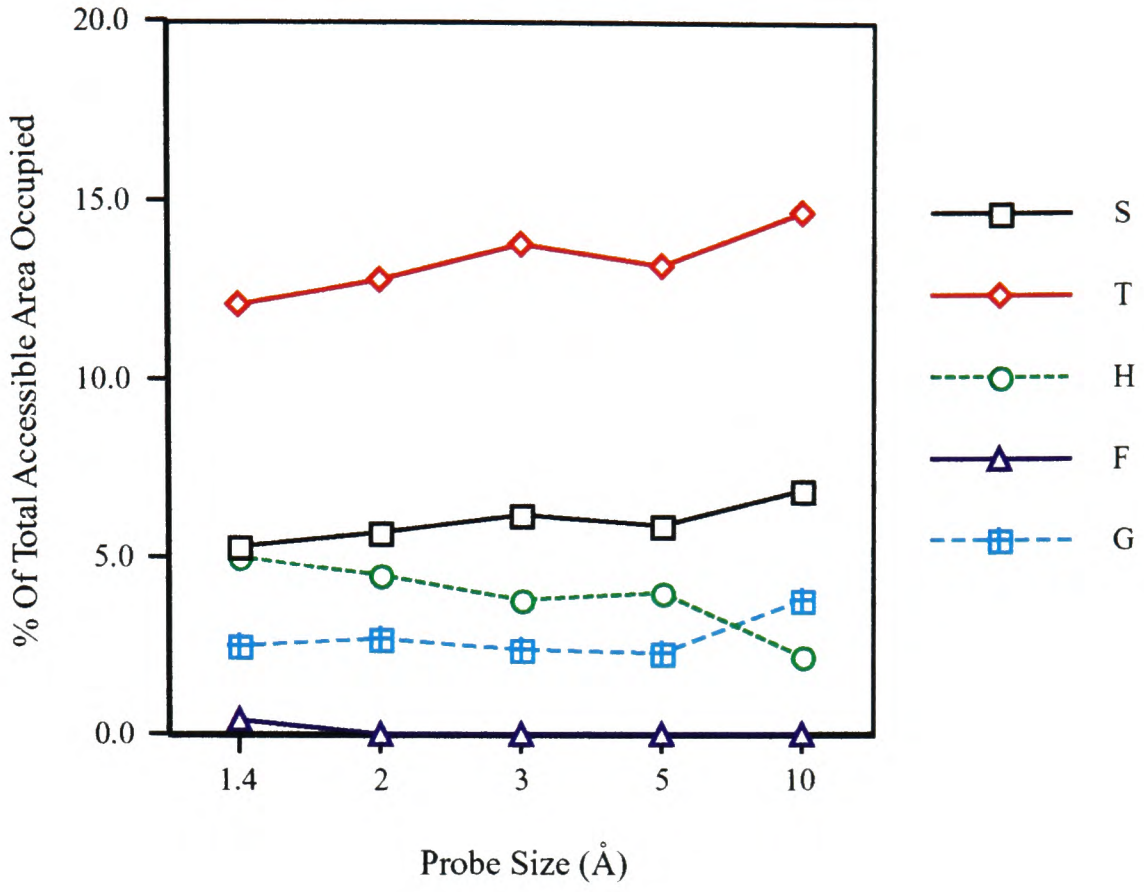


Figure 7.1: Effect of probe size on percentage of total accessible area occupied by specific amino acid types.

analysis. Such residues will have only a very limited effect on the interaction and will therefore be likely to evolve in response to pressures other than those imposed by the immune system. For the 5 Å probe analysis of DTT soaked FMDV O_1BFS (Logan *et al.*, 1993) different cutoffs varying between 1.4 and 120Å^2 were used . Cutoff values in the range 1.4–20Å^2 are seen to have little effect on the percentage of the total surface area occupied by each residue type (Figure 7.2), the larger cutoffs simply reducing the number of residues considered to contribute to the surface structure (Figure 7.3).

A combination of a 10 Å^2 cutoff with a 5 Å radius probe accessibility seems to classify residues as accessible/inaccessible in a way that is consistent with visual examination using a graphics system. This is therefore the criteria of accessibility used throughout the analysis.

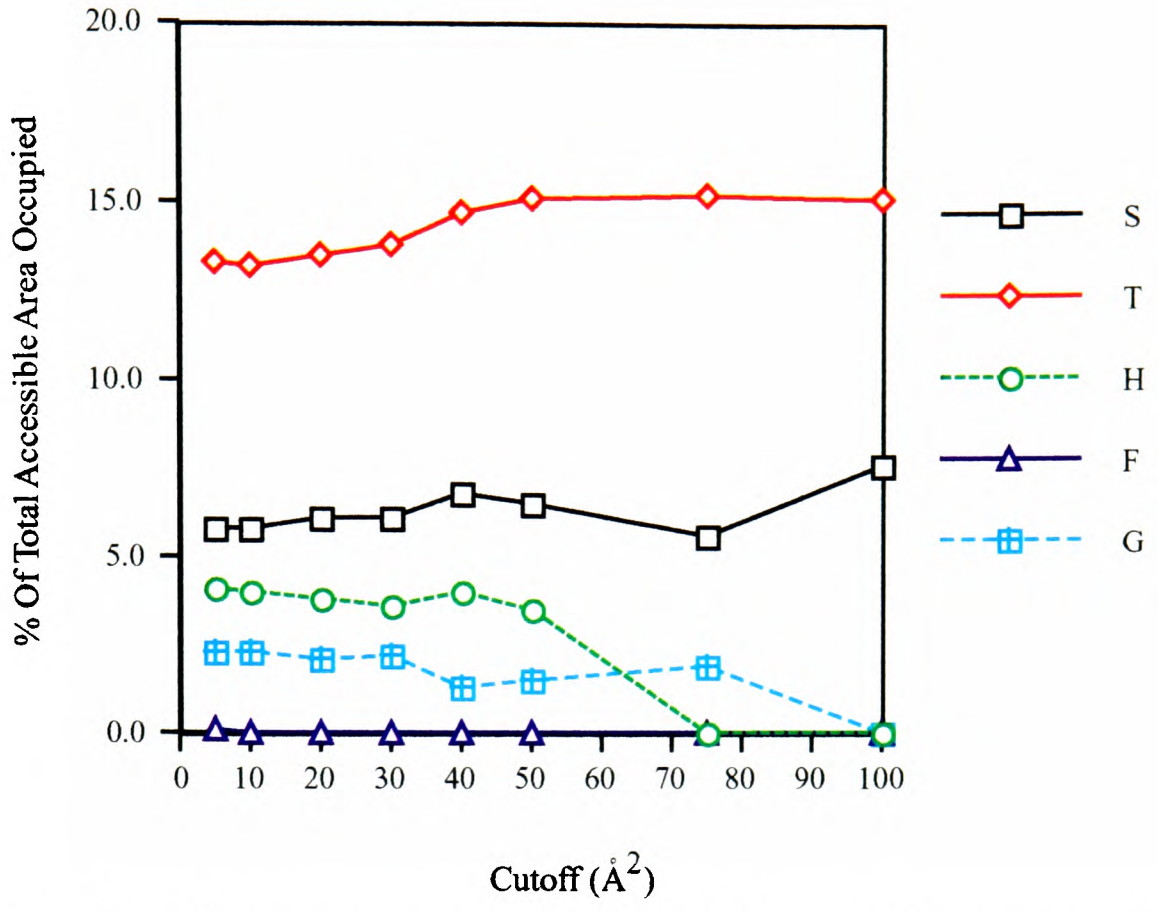


Figure 7.2: Effect of cutoff value on percentage of total accessible area occupied by specific amino acid types.

7.3 Results

The results are shown in two ways. In Figure 7.4 the mean percentage of the total accessible area occupied by each residue type is plotted as a bar chart (the error bars show one S.D. of the data), with amino acids referred to throughout by their one–letter code separately for each group of proteins. The column marked by a * shows the mean percentage of the total accessible area occupied by the aromatic residues (His, Tyr, Trp, Phe) for the group. The groups are as described above. In Figure 7.5 the results for the different groups are plotted against each other with error bars showing one SD of the data. The positions corresponding to Ser and Thr are highlighted (see below). The control proteins (group III) show the pattern expected of proteins which exist in an aqueous environment i.e. great use of hydrophilic residues capable of making energetically favourable interactions with H_2O, and relative suppression of hydrophobic residues. The large, charged side chains such as Lys and Glu form a

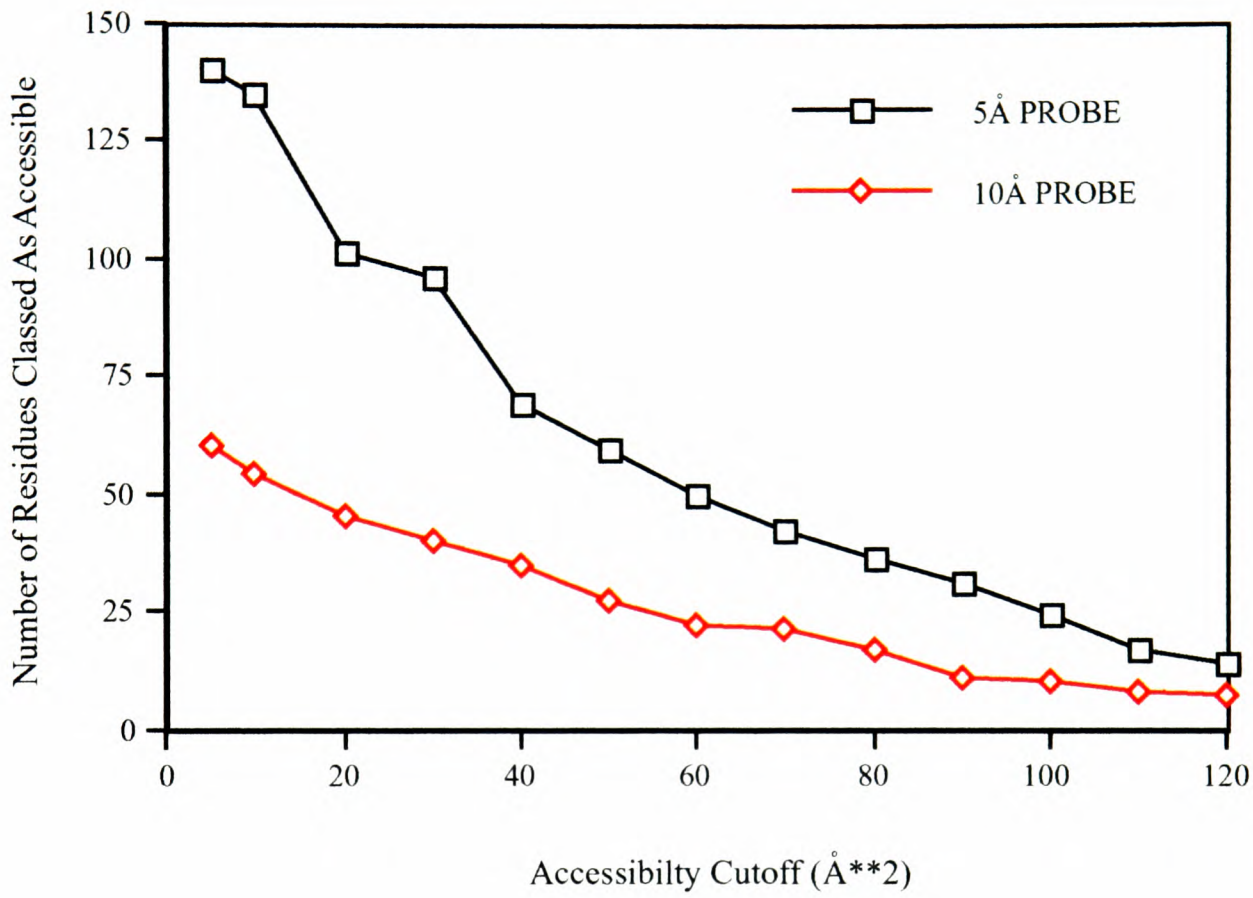
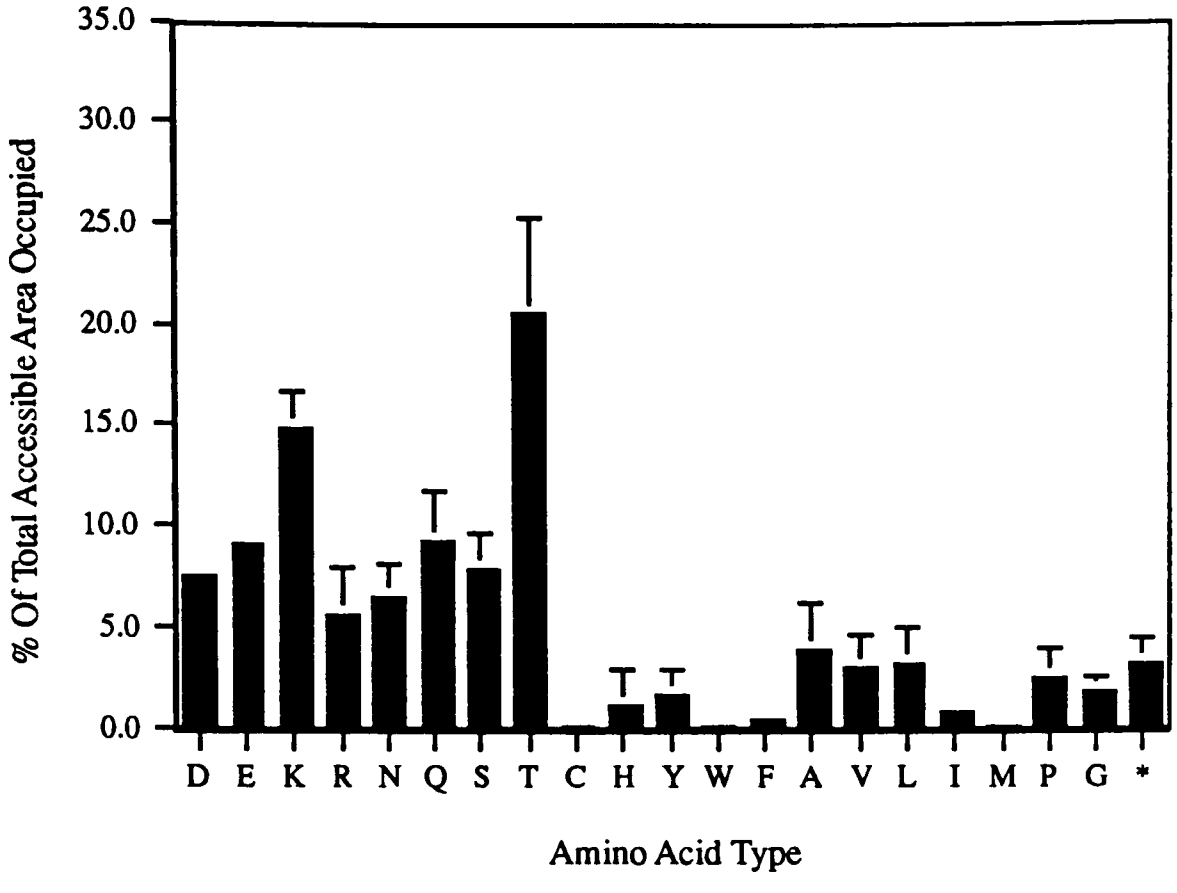
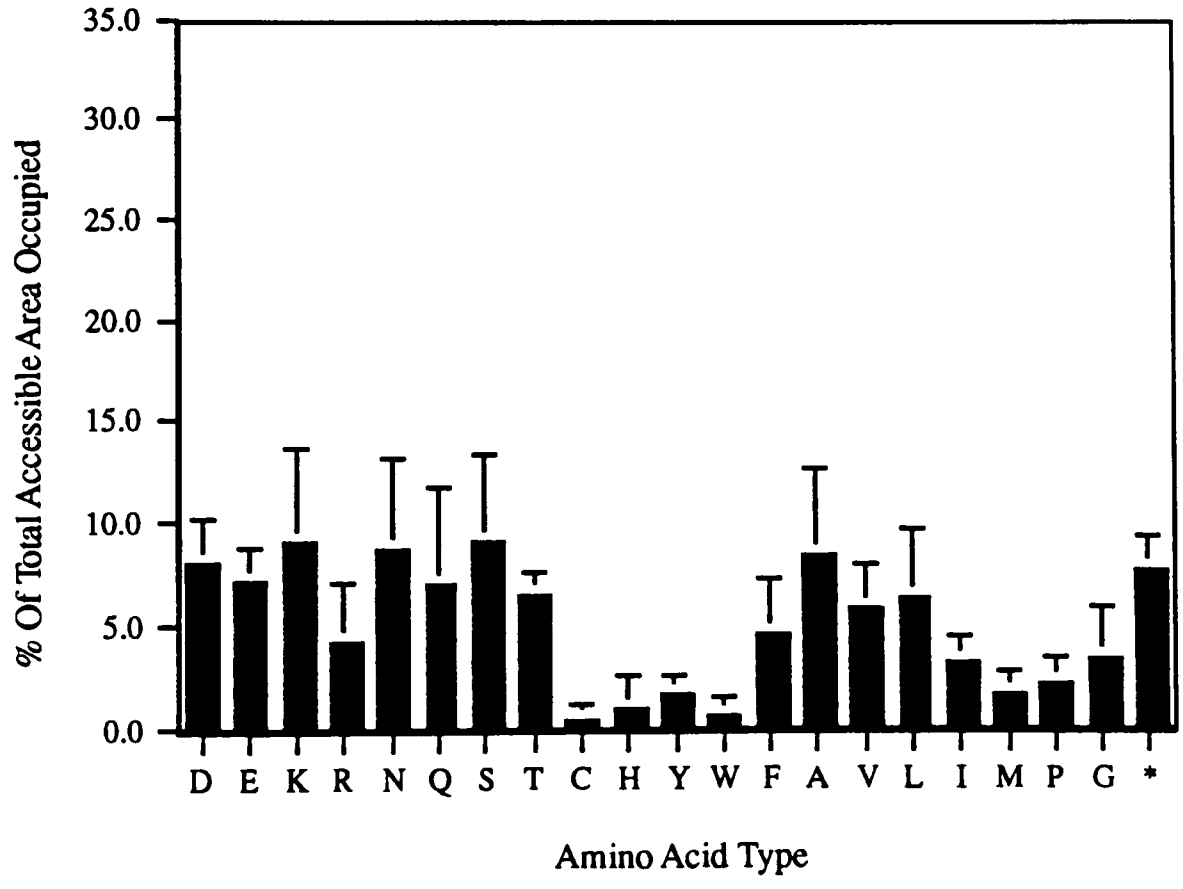


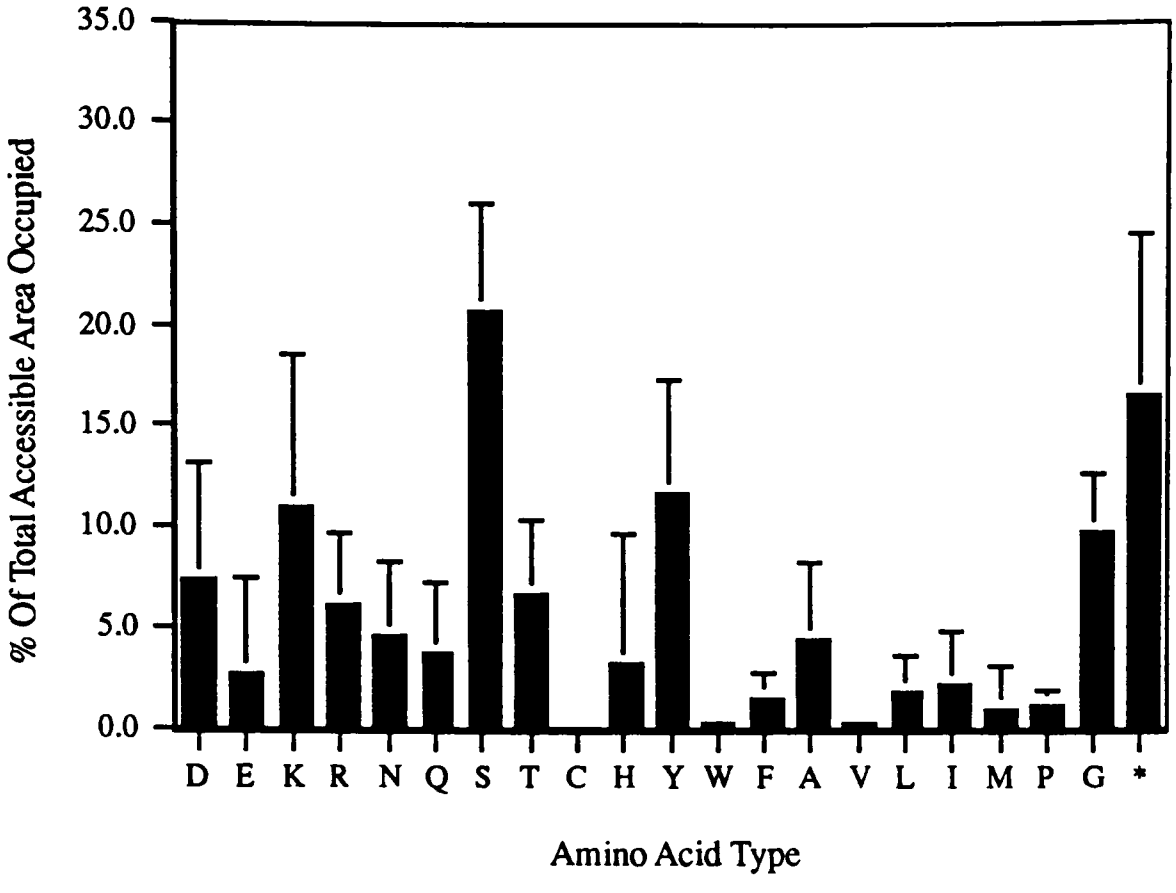
Figure 7.3: Effect of accessibility cut–off on the number of residues classed as accessible.



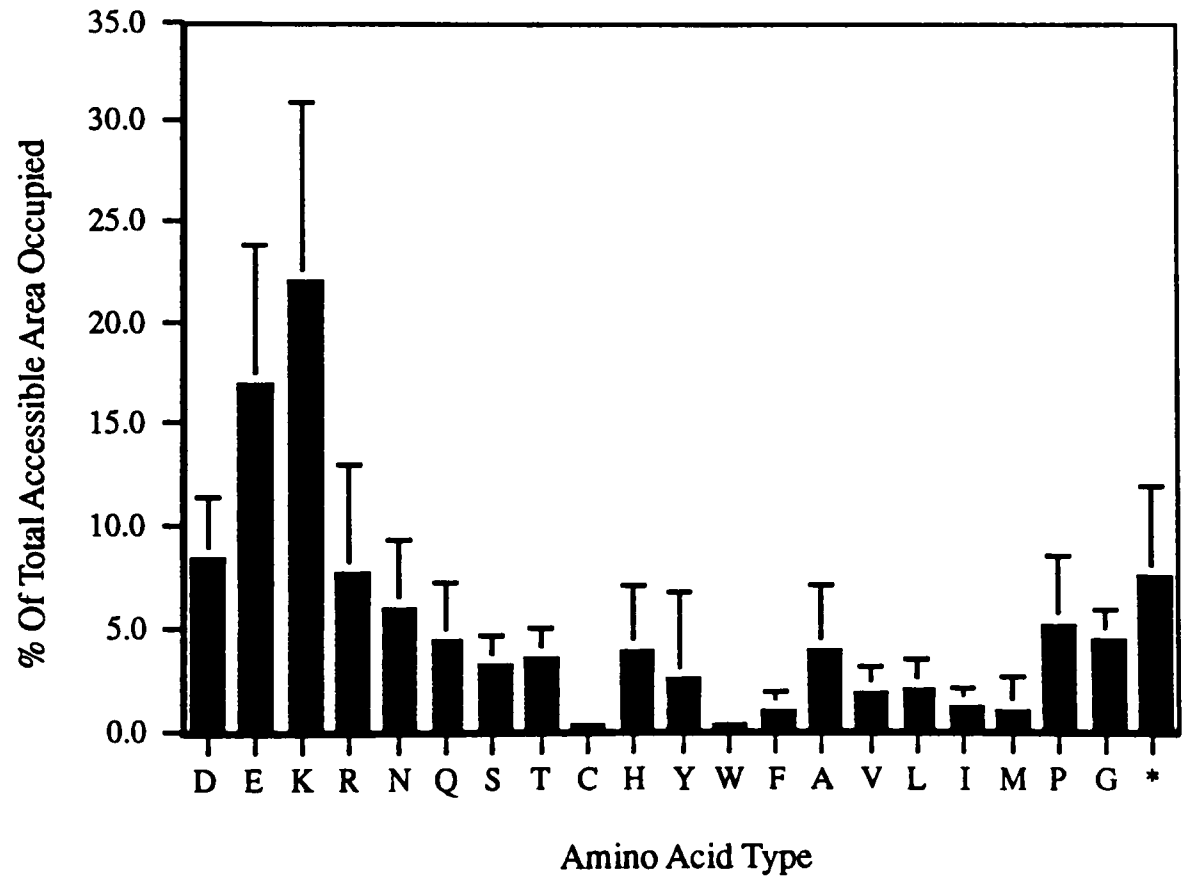
A. Group Ia.



B. Group Ib.

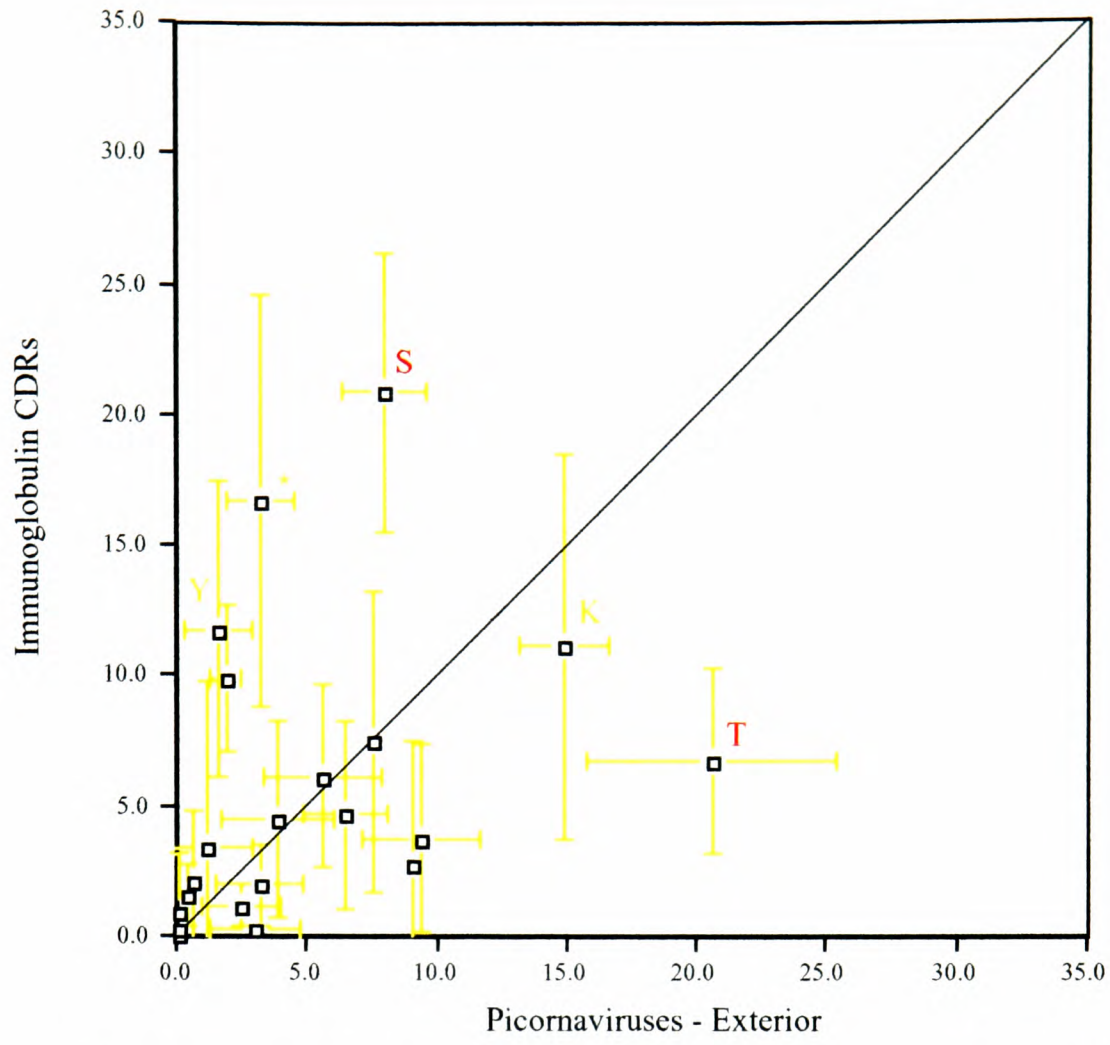


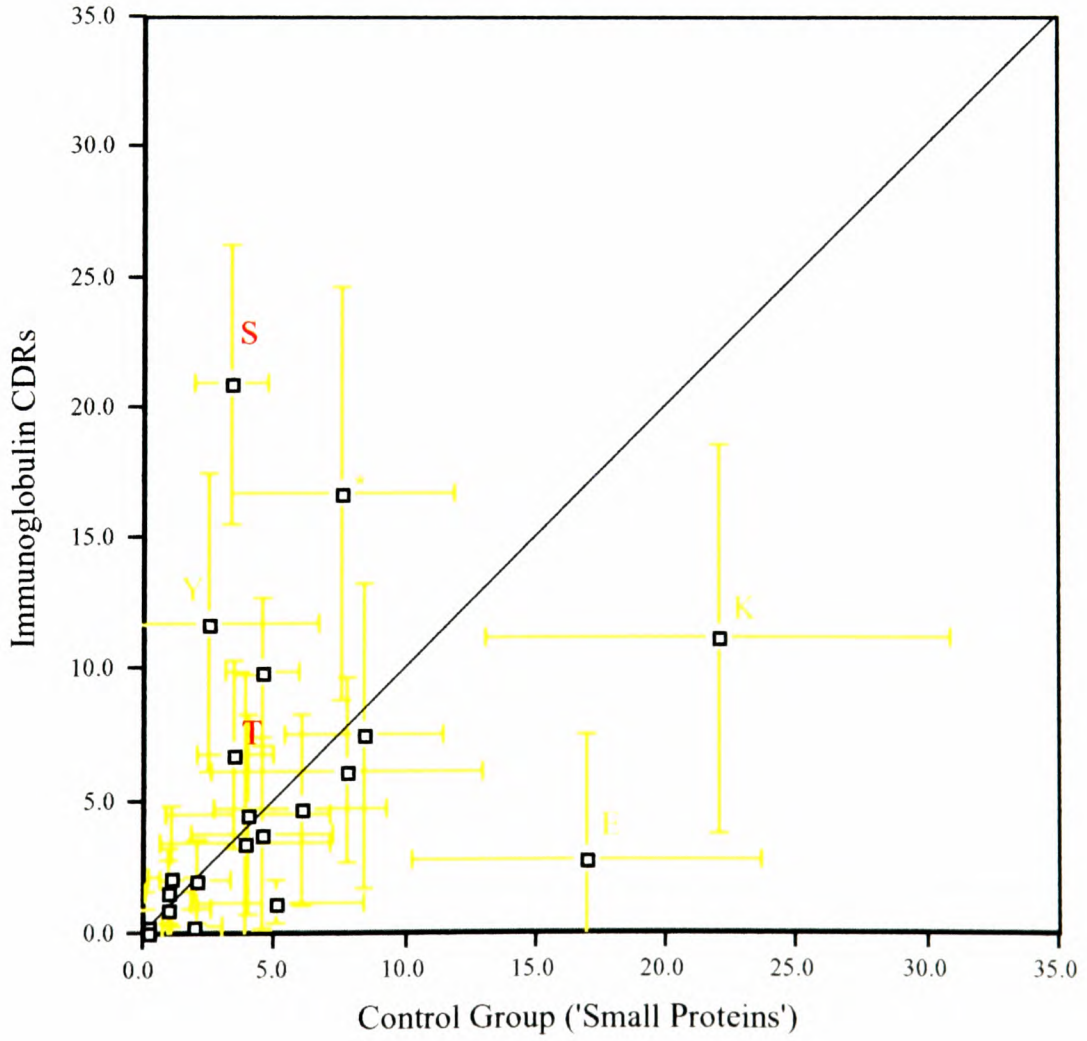
C. Group II.

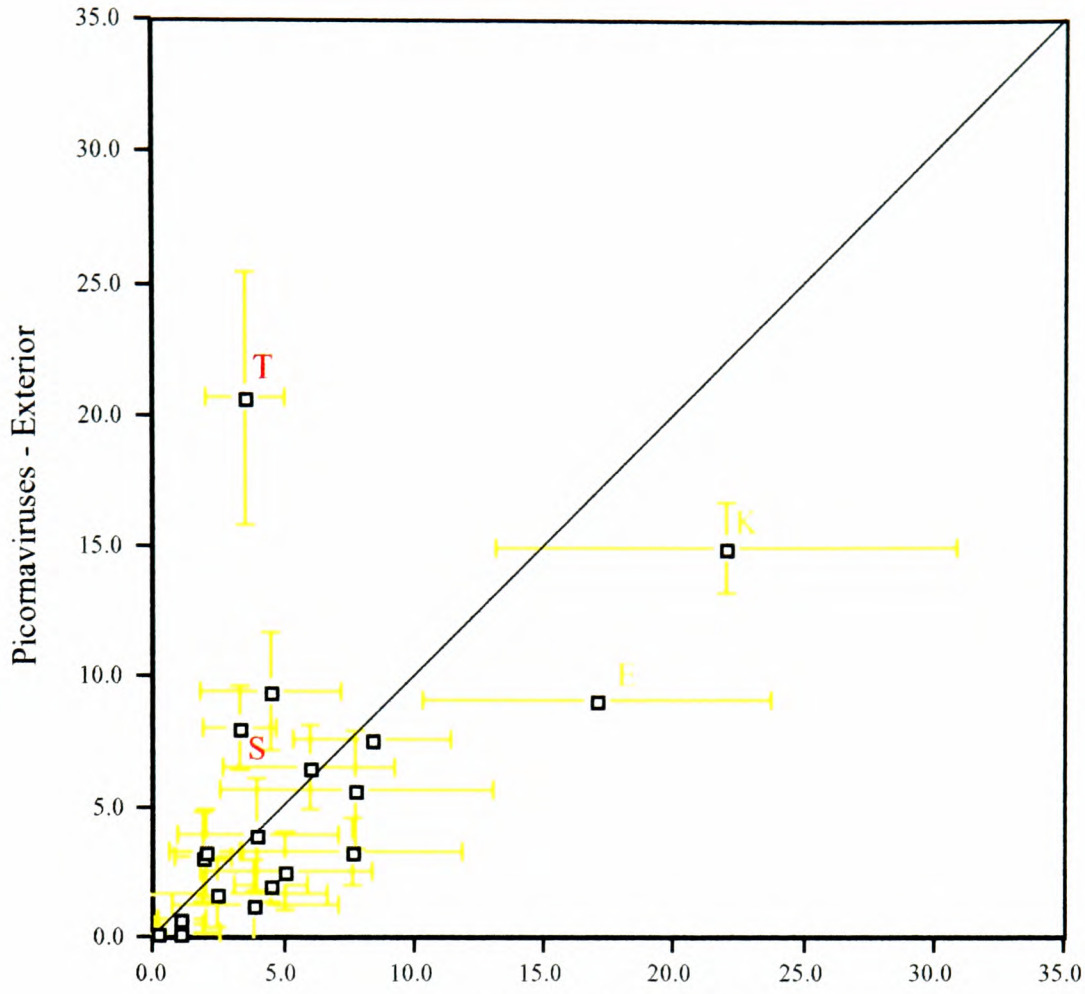


D. Group III.

Figure 7.4: The mean percentage of the total accessible area occupied by residue type for the four categories of proteins.


Immunoglobulin CDRs
Picornaviruses - Exterior
0.0
5.0
10.0
15.0
20.0
25.0
30.0
35.0
S
K
T
Y


Immunoglobulin CDRs
Control Group ('Small Proteins')
0.0
5.0
10.0
15.0
20.0
25.0
30.0
35.0
S
K
T
Y
E



Where the size of the error bar does not exceed the size of the plotted character, the error bar is not shown.

Figure 7.5: The mean percentage of the total accessible area occupied by residue type for the different groups.

large part of the surface enabling many stabilizing interactions with the surrounding solvent.

The Group I picornaviruses show very different patterns of use on the internal and external surfaces. The external surface (Ia) shows a similar pattern of use to that seen in the control group although there is a significant increase in the use of Thr and relative suppression of the use of Lys and Glu. The internal surface (Ib) shows a relative suppression of all hydrophilic residues (with respect to the control) so that the pattern of use is much flatter across all amino acid types. There is a significant increase in the use of Phe although overall use of aromatic residues in not increased compared to the controls. The differences between the internal and external surfaces must reflect, in part, the rôle of the internal surface in interacting with the viral RNA – the increased use of Phe may reflect the favourable interactions possible between the aromatic ring and the nucleotide bases (although it is noteworthy that the overall use of aromatic residues is not increased).

The immunoglobulin CDRs show a significant overuse of Ser and a suppression of Thr by comparison to the control group.

Overington *et al.* (1992) have shown that a Thr in an accessible site is most likely to be substituted for by a Ser and a Ser by a Thr, this leads to the prediction that approximately equal numbers of the two residues should be found on exposed protein surfaces. Thr and Ser are of similar sizes and therefore should occupy, in this analysis, roughly equal percentages of the total surface area. This is the pattern of use seen in the control group of proteins however, this pattern is significantly disturbed on the exterior surfaces of the picornaviruses and within immunoglobulin CDRs: the picornaviruses overuse Thr and the Ig CDRs overuse Ser.

7.3.1 Test to Distinguish Between Conservation and Re–Selection

At any evolutionary time point (as represented by a specific crystal structure) the amino acid composition of the protein's surface will reflect a mixture of invariant (conserved) residues which will remain constant under evolutionary pressure, and other more variable residues which will alter under evolutionary pressure. By looking at sequences of closely related proteins separated by a few mutations (i.e. a short evolutionary time) it is possible to distinguish between those amino acids which are fixed in the population and those which are allowed to vary. The overuse of Thr or Ser seen in groups I and II may occur by two different routes:

- The residue type is playing an important structural rôle and is fixed in the sequence.
- The residue is repeatedly re–selected in response to rapidly changing evolutionary pressures.

By examining whether specific residues are conserved (or not) across many sequence alignments it is possible to distinguish between these two mechanisms.

Known structures were used to predict which residues in multiple sequence alignments would be accessible (according to the criteria used above). An index of conservation was then calculated for each residue type:

$$\text{aa\_index} = \frac{\sum_{n=1}^{\text{NRES}} \left(\frac{(N_{OCC})^2}{\text{NPROT}} \right)}{T_{OCC}}$$

NRES = Number of residue positions which are predicted to be accessible.

N_{OCC} = The number of occurrences of that amino acid type at one *accessible position* in the sequence, for all the sequences in the alignment.

NPROT = The number of proteins in the alignment.

T_{OCC} = The total number of occurrences of that amino acid type at all accessible positions in the alignment.

This index will give a value of 1.0 if the residue is completely conserved, hence a lower value indicates variability. Since the index is related to the total number of occurrences it is plotted against T_{OCC} (Figure 7.6). A 'best fit' line is shown indicating the mean variability.

Alignment of immunoglobulin CDR sequences can be problematical due to the varying lengths of the loops forming these regions. We therefore chose to look at the light chain CDR 2 which is of a constant length in 68 of the light chain sequences available from the SWISSPROT database. These sequences were aligned using the GCG program suite (Genetics Computer Group, 1991) and the PDB2FBJ.ENT 5Å probe / 10Å ^{2} cutoff accessibility used to code residues (in CDR 2) for inclusion in the conservation calculations. Figure 7.6, (A) shows that both Ser and Thr (Positions indicated by one letter codes on graph) are more variable than the majority of residue types.

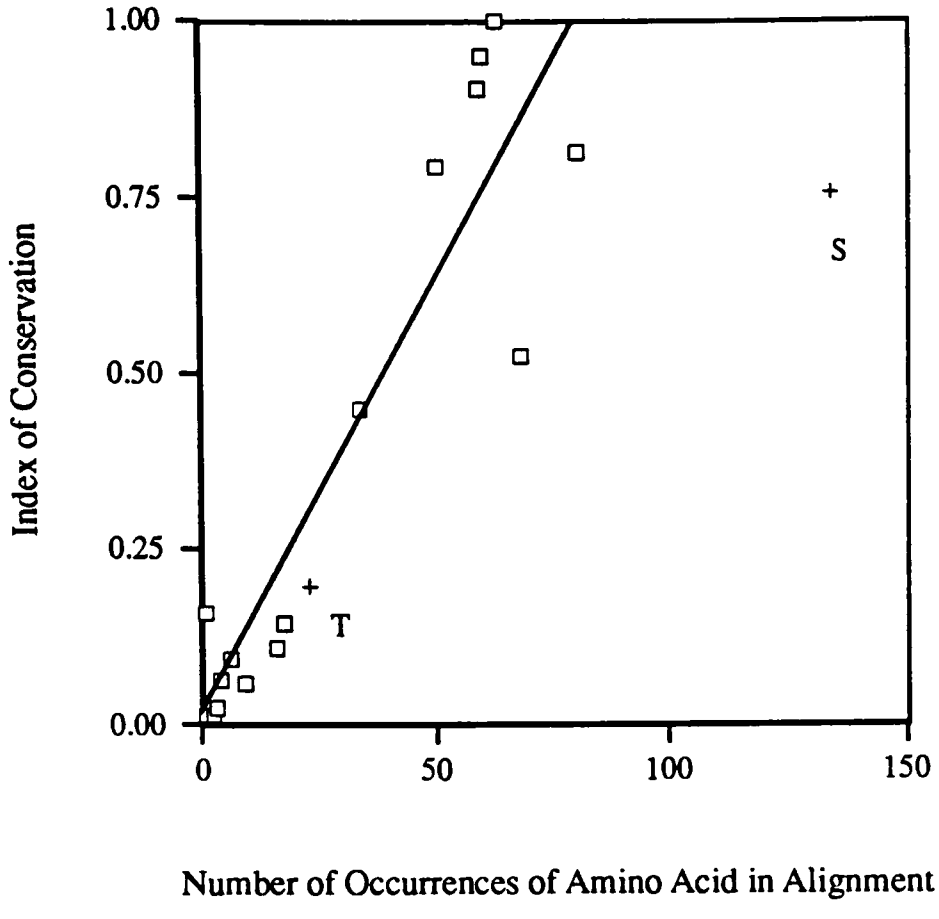
18 FMDV sequences (provided by Nick Knowles[2] and Esteban Domingo[3]) were aligned using the GCG suite, and coded using the 5Å probe/10Å^2 cutoff accessibility of reduced O_1BFS (Logan *et al.*, 1993). The positions of Ser and Thr are again indicated using one letter code showing, once more, greater than average variability.

7.4 Conclusions

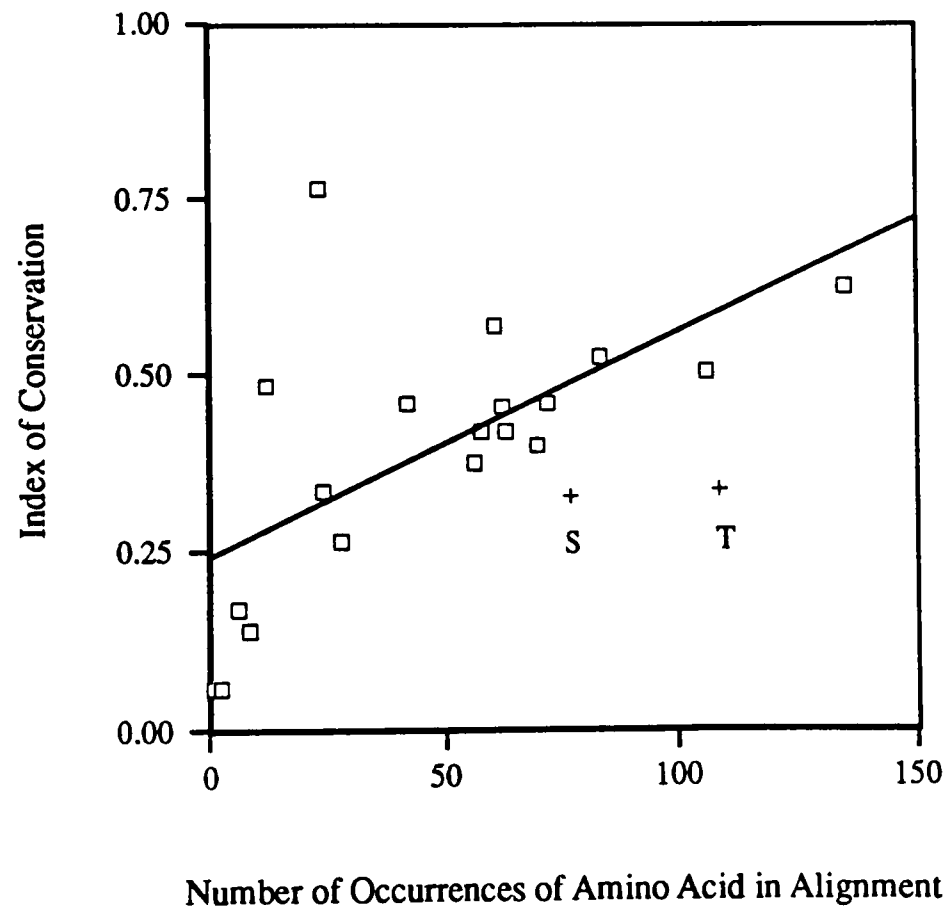
Our analysis showing that the most frequently used residue types are the least conserved suggests that repeated reselection of the amino acid is occurring. This reselection may occur for at least two reasons.

Most obviously reselection may occur because the residue is especially suited to

[2]AFRC IAH, Pirbright, UK.
[3]Centro de Biología Molecular, Madrid, Spain.



A. Immunoglobulin CDR 2.



B. FMDV VP1.

Figure 7.6: The index of conservation plotted against the number of occurrences of the amino acid (in the entire alignment) for all amino acid types.

the biological task. Previous theories concerning amino acid use in immunoglobulins have concentrated on the properties of a particular amino acid that may make it well suited to the task of promoting specific interactions with another protein. Greater than expected use of aromatic residues, and especially of Tyr has been noted in several immunoglobulin structures (Satow *et al.*, 1986; Amit *et al.*, 1986; Davies *et al.*, 1989) and it has been proposed (Davies *et al.*, 1989) that such overuse may be due to specific properties of Tyr that make it well suited to complex formation. These properties include:

1. It is a neutral residue capable of hydrogen bonding through the terminal hydroxyl group and is also able to accept hydrogen–bonds directed at the centre of the aromatic ring.

2. The aromatic component is capable of large packing adjustments by a small rotation about a single bond.

3. Tyr has a low conformational entropy – this minimises the decrease in entropy caused by complex formation.

Although we see an increase in the use of Tyr by the immunoglobulin CDRs (with respect to the controls) this is barely statistically significant in our analysis. Overall, the immunoglobulin CDRs, show an increase in the use of aromatic residues as a whole – although the large SD of this class indicates great variability in the use of aromatic residues between different immunoglobulins.

Ser is more frequently found in enzyme active sites than other hydrophillic residues (personal communication T. Blundell). Whilst some of this overuse will obviously be due to the use of Ser by some enzymes in their catalytic mechanism, it has been suggested that the great flexibility and short length of the Ser side chain make it well suited to the task of recognition of the enzyme substrate. Such properties obviously suit Ser to recognition of an antigen and therefore may explain the overuse of Ser by the immunoglobulins. Conversely the relative rigidity of the Thr side chain may

make it ill suited to the task of recognition and so explain the overuse of this residue type by viruses.

An alternative explanation for this overuse might lie within the nucleotides coding for the residues. The genetic code for Thr is AC* and for Ser UC* or AGU/AGC. Nucleotide point mutations are frequent occurrences but tend to be biased towards transitions (i.e. A to G, C to T or *vice versa*). This means that the most likely outcome of a point mutation in a codon coding for Thr (other than the mutation being silent and still producing Thr) is production of Ala (1/2), Ile (1/2) or Met (1/4) at that position. Thr, Ala and Ile are chemically and structurally fairly 'unexciting' containing no charged groups or long side chains, they may therefore be less likely to provoke strong interactions with residues within an immunoglobulin CDR. This may render the virus less visible to the host immune system. The virus therefore over-uses Thr as it lies at a 'crossroads' in the genetic code where single base mutations are unlikely to result in an amino acid being expressed that increases the likelihood of recognition by the immune system.

Transitions in the nucleotides coding for Ser will lead to expression of Phe (1/6), Leu (1/6), Pro (1/3), Gly (1/3), Lys (1/6) and Asn (1/6). These residues are all chemically dissimilar and many consist of relatively large side chains bearing charged groups. A mutation from Ser to any of these residue types will significantly alter the characteristics of the immunoglobulin CDR and so alter the immunoglobulin specificity.

This analysis suggests that the genetic code for a residue may be as important in determining use of that residue as the more obviously important characteristics of side chain structure and chemical nature. The abundance of a certain residue may therefore be not only connected with the characteristics of that amino acid, but may also be intimately connected with the characteristics of amino acids with base codes single base mutations away from its code.

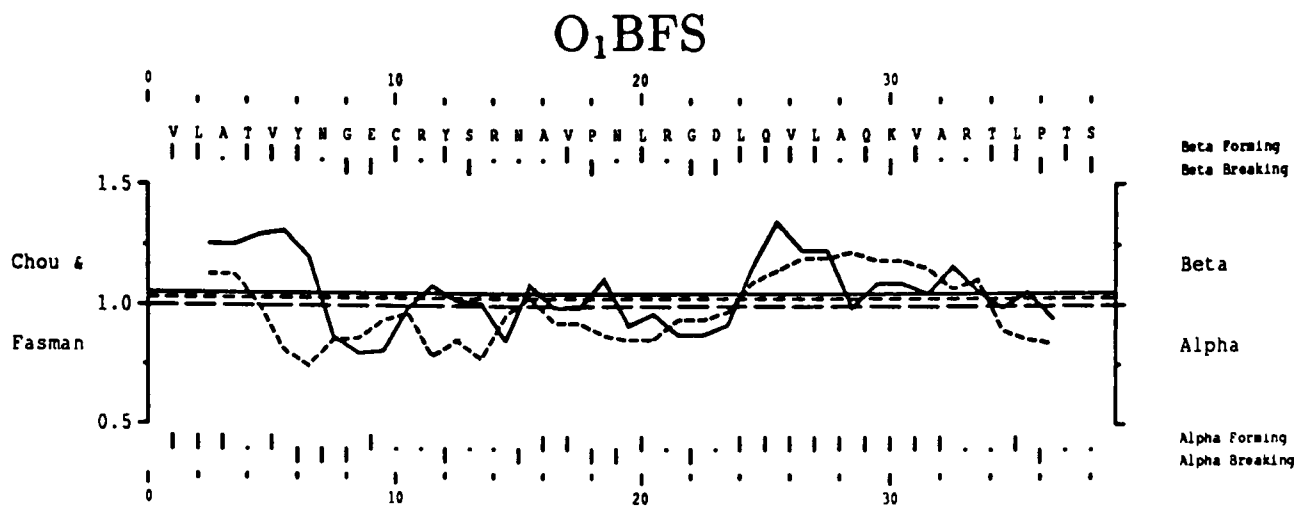
Chapter 8

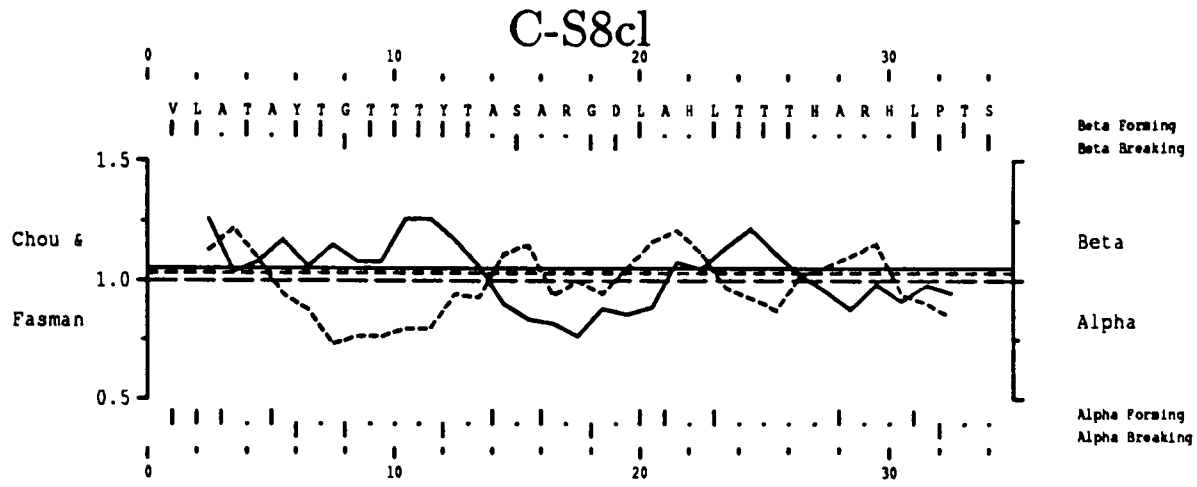
Conclusions

8.1 Receptor Binding

We now have structures for members of 3 FMDV serotypes; O_1BFS (Acharya *et al.*, 1989), C-S8cl (Chapters 5 and 6), A_{22} (Curry *et al.*, unpublished), A_{10} (Fry *et al.*, unpublished). In all of these structures flexibility/disorder of the loop containing the primary receptor–binding determinant (the RGD tripeptide within the G–H loop of VP1) is amongst the most conserved structural feature. This is despite differences in the length and sequence of this loop between all the viruses. These differences in the length can be extreme by comparision to differences seen in other regions of the capsid; the shortest and the longest sequenced loops differing by 7 residues in length. The family of O_1 structures in 'native' and 'reduced' forms[1] have revealed that in this subtype the loop has an inherently stable internal structure and flexibility is maintained by formation of a 'stressed' disulphide–bond between a loop residue and VP2. This disulphide prevents packing of the ordered loop against VP2, and results in the observed flexibility/disorder - the loop acting as a separate structural unit hinged to the virion via the disulphide. The structures of viruses lacking the disulphide still show the flexibility/disorder without the presence of a destabilizing disulphide. The question therefore arises whether in these viruses the loop has some stable internal structure or not. Figure 8.1 shows secondary structure predictions (produced using PEPPLOT (Genetics Computer Group, 1991)) for VP1 G–H loop sequences for 4 viruses for which structural information is known (O_1BFS, C-S8cl, A_{10} and

[1] O_1BFS (Acharya *et al.*, 1989; Logan *et al.*, 1993), O_1K, G67, O_1K B64 (Chapter 3 and Appendix A)





In all plots secondary structure predictions for the residues equivalent to the 125–162 regions of VP1 in FMDV O_1BFS are shown. This region covers the G–H loop and parts of the G and H β–strands. The plots show the Chou and Fasman index (Chou and Fasman, 1974) for that region of the sequence, a value in excess of 1 shows an increasing tendency to form the structural element, below 1 an increasing tendency to break the structural element. The tendency to form β–sheet is indicated by the solid black line, and to form α–helix by the broken line. β / α forming /breaking residues are shown as 'tick' marks above and below the plots.

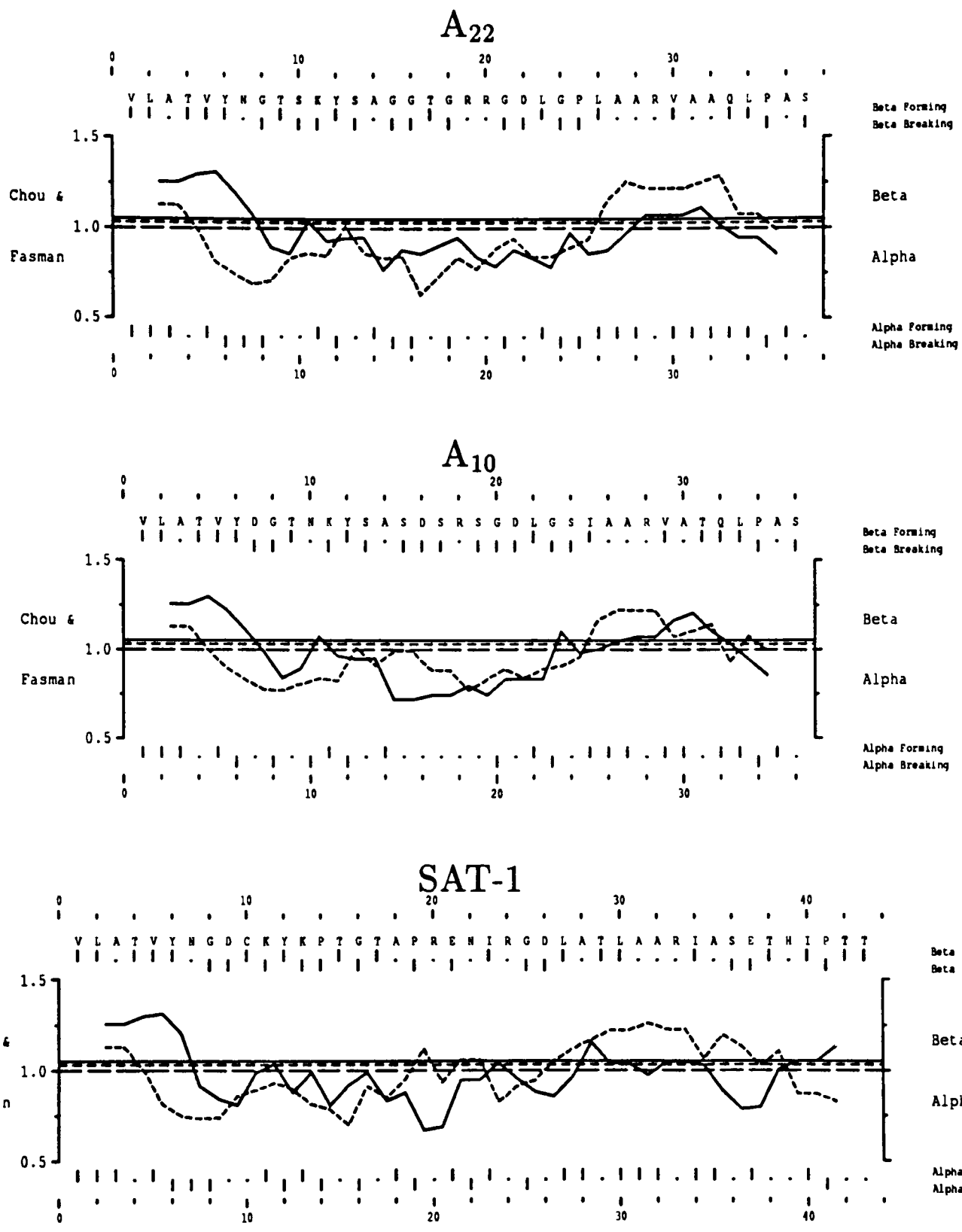


Figure 8.1: Secondary structure predictions for the VP1 G–H loop

A_{22}) and for one for which no structure currently exists (SAT-1). Secondary structure predictions are difficult to make accurately, the simple sequence based method used here does not attempt to use any information about the likely environment of the sequence to increase the accuracy of the prediction. Such predictions must not be over–interpreted but the O_1BFS sequence contains a large group of consecutive residues likely to form a secondary structural element (the 9 residues predicted to be α forming following the RGD). This is indeed the region found to form a 3_{10} helix in the reduced O_1 structures lending credence to the structure prediction.

The other sequences all seem less likely to form ordered secondary structures. The C-S8cl sequence, in particular, has little tendency to adopt a single, defined, structure - this correlates well with the structural information since fewer residues are visible at the N- and C-terminal ends of the disordered loop in C-S8cl than in the other FMDV structures. The A_{22}, A_{10} and SAT-1 sequences show some tendency to adopt helical conformations in regions equivalent to the helical region in O_1, but in all cases the prediction based evidence is not as compelling as for the O_1 viruses, with short, interrupted patches of residues likely to be helix forming rather than a single, longer, consecutive region of such residues. Considering the structural results and predictions together I would be surprised if a SAT-1 structure revealed an ordered conformation for this loop.

Flexibility of loops bearing RGD tripeptides involved in receptor binding is not only conserved amongst FMDVs but also amongst other integrin binding proteins e.g.: γ-II crystallin (Wistow *et al.*, 1983). This implies that flexibility may be important in allowing efficient receptor binding - the O_1BFS sequence with its predisposition to adopt an ordered, stable conformation may therefore only be tolerated in the context of VP1–VP2 disulphide which maintains flexibility by preventing packing of an ordered G–H loop against VP2. In the FMDVs lacking such a disulphide flexibility must be maintained in other ways, this may therefore lead to selective pressures against mutations which increase the propensity of the loop to form ordered secondary structure.

8.2 Antigenic Escape

Although antigenically very different (sharing no common epitopes) FMDVs of different serotypes are seen to be structurally very similar with changes limited mainly to alterations of side chains with little associated main chain movement. Where loops are seen to alter in their underlying conformation the changes are relatively small. This confirms results obtained with different viruses e.g. HRVs (Oliveira *et al.*, 1993; Kim *et al.*, 1989) and polioviruses (Filman *et al.*, 1989), which have shown that antigenically dissimilar viruses are structurally very similar. Still unanswered is the question of why mutations to evade immune pressure result in viruses falling into distinct serotypes, with similar antigenic phenotypes, rather than forming a broad spectrum of all possible intermediate phenotypes. The structure of C-S8c1 site D_2 may perhaps give some clues. Amongst MAb–mutant viruses differences in amino acids are seen only in those positions which result in conservation of the underlying main chain of the loop. A mutation which results in alteration of the main chain structure may cause such a radical alteration of the local conformation that no common antibody epitope exists between the parental and mutant viruses i.e. the mutant virus belong to a new serotype (at least in terms of that antigenic site). Perhaps our perception of viruses as belonging to distinct serotypes is not really due to startling differences in the viruses but is instead the result of the exquisite specificity of antibodies - simply another reflection of how small a conformational change is required to render a site unrecognizable, as demonstrated by the structure of the quadruple MAb escape mutant G67.

The O_1 structures presented herein seem also to caution us that we cannot necessarily assume that information gained using MAbs *in vitro* gives direct insight into the way in which viruses evolve in response to the complex polyclonal pressures confronted in animals *in vivo*. It may be that the time–dependent structural variations are important in determining field escape i.e. the relatively short time–span variations produced by formation of the O_1 disulphide over a few hours and the variation

produced by 'flipping' of residues at key structural positions between two amino acids over different generations.

The analysis of protein structures presented in Chapter 7 also warns that we must always remember that that proteins are initially coded for by nucleotides and that evolutionary pressures will act at this level as well as at the amino acid level. The amino acid sequence may therefore sometimes be determined by factors other than the properties of the amino acids of which it consists at any time–point.

8.3 Future Directions

The future holds many exciting ways in which structures may be further used to aid our understanding of viruses. The structures already determined could yield more information with the development of better methods to compare several related structures. Currently we are limited to simplistic comparisons based on crude measures of a structure such as comparisons between Cα positions, residue by residue accessibility or charge, or subjective 'by–eye' comparisons of different structures. Ways need to be found to combine all the available information in a way that is independent of the primary sequence so that compensatory mutations in different regions of the primary sequence brought together by folding are considered as regions of 'spatial conservation' rather than separate instances of sequence related change.

In terms of new structures the most exciting possibilities lie in direct visualization of antibody–virus complexes, either as FAb–virus (Smith *et al.*, 1993) or Fv–virus (T. Jackson personal communication) and, once the FMDV receptor is identified, in receptor–virus complexes. The structure of a virus in which VP0 is uncleaved would also be very exciting and is likely to be possible using a Vaccinia expressed clone of A_{10} crystallised at 4°C (S. Curry, personal communication). Cryo–temperature protocols might be developed in an attempt to improve order in the currently partially ordered regions e.g. the N-terminus of VP2 and 'myristate' density under the 5–fold axis.

Appendix A

Other Data Sets

A.1 O_1K B64

O_1K B64 is derived from the same field isolate as O_1K (Chapter 3) but differs in its passage history (it has been passaged 64 times in BHK[1] cells). This is therefore a tissue–culture adapted virus showing markedly altered reactivities with MAbs with minimal sequence changes (Abu–Ghazaleh, unpublished observations). The virus crystallises in the standard O_1 morphology (Table A.1). Data for both 'native' and 'reduced' crystals of O_1K B64 have been collected and processed (Table A.1). Difference maps calculated between B64 and the other O_1 viruses show clear areas of structural change. Final analysis awaits re–sequencing of the region around 130 of VP2 and the FMDV loop as the structural results suggest the 'crystal' virus does not share the published sequence - although the structural changes are large enough that interpretation of the maps in terms of an unambiguous sequence is not possible.

A.2 A22 Variant

This virus is a MAb–escape mutant of virus A_{22} showing alterations in its interaction with loop directed antibodies and also in its host–range (Bolwell *et al.*, 1989). When sequenced no mutations were found within the loop, the only surface exposed mutations occurring in VP2, in the region lying under the loop as revealed in reduced O_1BFS. This virus crystallised isomorphously with the parental A_{22} (Curry *et al.*, 1992) in space group I222. Data Processing is summarised in Table A.1 and shows

[1]BHK = Baby Hamster Kidney.

Virus	O_1K B64		A22 Variant	CGC-Ger
	Native	Reduced		
Image Plate Used	18cm MAR	30cm MAR	RAXIS (fine)	18cm MAR
Station	9.5	9.6	9.6	9.5
No. of Crystals	2	3	1	10
No. of images	19	19	16	30
Space Group	I23	I23	I222	I23
Cell Parameters	a=345.0Å α=90°	a=345.0Å α=90°	a=348.2Å b=341.8Å c=363.75Å α=90°	a=347.6Å α=90°
R(I) %	14.3	15.2	9.6	13.5
Resolution (Å)	3.2Å	2.8Å	2.4Å	3.2Å
No. of Ind. Refl.	78443	116150	110781	90564
% complete	73	70	19	79

Table A.1: Brief summary of data processing for various FMDVs

the unusually good merging R–factor - all the data were collected from one crystal. The structure has been solved by calculation of difference maps using the parental phases (Stephen Curry, unpublished data) and suggests that only one of the two mutations in VP2 implicated in the altered antigenic phenotype appears to be present in our 'crystal' virus. Re-sequencing and checking of the 'crystal' virus phenotype is being undertaken, although clear differences in the occupancy of residues near the disordered termini of the G–H loop suggest the crystal virus will have the altered phenotype. This structure again shows that mutations in one capsid protein can have drastic effects on the conformation of an antigenic loop located in another protein in a similar way to the O_1 mutants described in Parry *et al.* (1990).

A.3 CGC-Ger

This virus appears (by sequence analysis and date of isolation in the field) to be a representative of an ancestral C subtype now extinct in the wild (Martínez *et al.*, 1992). As such it is of interest in allowing us to examine a different time point in the evolution of the C serotype FMDVs. CGC-Ger differs from C-S8c1 by only 44

PROTEIN	NO. OF DIFFERENT AMINO ACIDS	% IDENTITY
VP1	17	92
VP2	16	93
VP3	6	97
VP4	5	95
TOTAL	44	94

Table A.2: Comparison of C-S8c1 and CGC GER/26.

mutations in the 4 capsid proteins (Table A.2) and crystallises isomorphously with C-S8c1, but the crystals diffract to higher resolution (at least 2.8Å). An initial data set was collected using the 18cm MAR plate (Table A.1) to 3.2Å. The data scaled to the C-Master data with R=20.0%, C=0.89. The C-S8c1 phases were used as the start point for cyclic averaging and phase extension (GAP) which produced phases differing from the starting phases by a mean phase difference of 30.1°(to 3.5Å) after 90 cycles, with the agreement between calculated and observed amplitudes being R=17.3%, C=0.90 at 3.2Å. A $2|F_{OBS}| - |F_{CALC}|$ map calculated at 3.2Å, with these phases shows all the sequence changes clearly. The map also reveals further detail in the density around the N–terminus of VP2 and the 'myristate' density under the 5–fold axis. A model based on the C-S8c1 structure and incorporating the correct sequence has been built and refined (using XPLOR) to an R–factor of 20.7%[2]. Further data to 2.8Å have been collected using the 30mm MAR plate installed on station 9.6 but not yet analysed.

[2]RMS deviation from ideality of bond lengths being 0.012Å, and of angles 1.9°.

Appendix B

Glossary

BBV Black beetle virus

BEV Bovine enterovirus

BHK Baby Hamster Kidney (cells)

$$\mathbf{C} = \frac{\sum_h (<F_{obs}> - |F_{obs}|_h)(<F_{calc}> - |F_{calc}|_h)}{[\sum_h (<F_{obs}> - |F_{obs}|_h)^2 \cdot \sum_h (<F_{calc}> - |F_{calc}|_h)^2]^{\frac{1}{2}}}$$

$$\mathbf{C}_{match} = \frac{\sum_h (<F_{virus1}> - |F_{virus1}|_h)(<F_{virus2}> - |F_{virus2}|_h)}{[\sum_h (<F_{virus1}> - |F_{virus1}|_h)^2 \cdot \sum_h (<F_{virus2}> - |F_{virus2}|_h)^2]^{\frac{1}{2}}}$$

CAV Coxsackie A virus

CAV9 Coxsackievirus A9

CDR Complementarity determining region

CMV Carnation mottle virus

CPV Canine parvovirus

ϵ used in postrefinement to signifiy beam divergence, equivalent in this context to the mosaicity of the crystal

Fullys Those reflections for which the reflecting range of the reciprocal lattice point lies entirely within the oscillation range of the image

FAb Antigen binding antibody fragment

FMD Foot–and–mouth disease

FMDV Foot–and–mouth disease virus

Fv Variable region antibody fragment

HRV Human rhinovirus

Ig Immunoglobulin

MAb Monoclonal antibody

Partials Those reflections for which the reflecting range of the reciprocal lattice point lies only partially within the oscillation range of the image

PR Postrefinement

$$\mathbf{R}_C = \frac{\sum_h |(|F_{obs}|_h - |F_{calc}|_h)|}{\sum_h |F_{obs}|_h} \times 100$$

$$\mathbf{R(I)} = \frac{\sum_h \sum_i |(I_h - I_{hi})|}{\sum_h \sum_i |I_{hi}|} \times 100$$

$$\mathbf{R}_{match} = \frac{\sum_h ||F_{virus1}|_h - |F_{virus2}|_h|}{\sum_h |F_{virus1}|_h}$$

RMS Root Mean Square

RMSD Root Mean Square Deviation

SBMV Southern bean mosaic virus

SD Standard deviation

SDPHI Used to denote the quality of an auto–indexing solution, the standard deviation in the spindle angle position of the spots, units are °.

SDXY Used to denote the quality of an auto–indexing solution, the standard deviation in the xy position of the spots, usual units are pixels.

TBSV Tomato bushy stunt virus

TMV Tobacco mosaic virus

References

Abad–Zapatero, C., Abdel–Meguid, S., Johnson, J., Leslie, A., Rayment, I., Rossmann, M., Suck, D. and Tsukihara, T. (1981). A description of techniques used in the structure determination of southern bean mosaic virus at 2.8Å resolution. *Acta Cryst.*, **B37**, 2002–2018.

Acharya, K., Ren, J., Stuart, D., Phillips, D. and Fenna, R. (1991). Crystal structure of human α–lactalbumin at 1.7Å resolution. *J. Mol. Biol.*, **221**, 571–581.

Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D. and Brown, F. (1989). The three–dimensional structure of foot–and–mouth disease virus at 2.9Å resolution. *Nature (London)*, **337**, 709–716.

Acharya, R., Fry, E., Logan, D., Stuart, D., Brown, F. and Rowlands, D. *New apsects of positive strand RNA viruses*, chapter Some observations on the three–dimensional structure of foot–and–mouth disease virus. Brinton and Heinz, (1990).

Adler, M., Lazarus, R., Dennis, M. and Wagner, G. (1991). Solution structure of kistrin, a potent platelet aggregation inhibitor and GP IIb-IIIa antagonist. *Science*, **253**, 445–448.

Air, G., Laver, W. and Webster, R. (1990). Mechanism of antigenic variation in an individual epitope on influenza N9 neuraminidase. *J. Virol.*, **64**, 5797–5803.

Aitken, A., Cohen, P., Santikarn, S., Williams, D., Calder, A., Smith, A. and Klee, C. (1982). Identification of the NH_2–terminal blocking group of calcineurin B as myristic acid. *FEBS letter*, **150**, 314–318.

Alzari, P., Lascombe, M. and Poljak, R. (1988). Three–dimensional structure of antibodies. *Ann. Rev. Immunol.*, **6**, 555–580.

Amit, A., Mariuzza, R., Phillips, S. and Poljak, R. (1986). Three–dimensional structure of an antigen–antibody complex at 2.8Å resolution. *Science*, **233**, 747–753.

Arndt, U. (1986). X–ray position–sensitive detectors. *J. Appl. Cryst.*, **19**, 145–163.

Arnold, E. and Rossmann, M. (1986). Effect of errors, redundancy and solvent content on the molecular replacement procedure for the structure determination of biological macromolecules. *Proc. Natl. Acad. Sci. (USA)*, **83**, 5489–5493.

Arnold, E., Vriend, G., Luo, M., Griffith, J., Kamer, G., Erickson, J., Frankenberger, E., Johnson, J. and Rossmann, M. (1987). Structure determination of a common cold virus, human rhinovirus 14. *Acta Cryst.*, **A43**, 346–361.

Arnold, E. and Rossmann, M. (1988). The use of molecular replacement phases for the refinement of the human rhinovirus 14 structure. *Acta. Cryst.*, **A44**, 270–282.

Babu, Y., Bugg, C. and Cook, W. (1988). Structure of calmodulin refined at 2.2Å resolution. *J. Mol. Biol.*, **204**, 191–204.

Barnett, P., Ouldridge, E., Rowlands, D., Brown, F. and Parry, N. (1989). Neutralising epitopes on type O foot–and–mouth disease virus. I. Identification and characterisation of three functionally independent, conformational sites. *J. Gen. Virol.*, **70**, 1483–1491.

Baxt, B., Vakharia, V., Moore, D., Franke, A. and Morgan, D. (1989). Analysis of neutralising antigenic sites on the surface of type A_{12} foot–and–mouth disease virus. *J. Virol.*, **63**, 2143–2151.

Belsham, G. and Brangwyn, J. (1990). A region of the 5' non–coding region of foot–and–mouth disease virus RNA directs efficient internal initiation of protein synthesis within cells: Involvement with the rôle of L protease in translational control. *J. Virol.*, **64**, 5389–5395.

Belsham, G., Abrams, C., King, A., Roosien, J. and Vlak, J. (1991). Myristoylation of foot–and–mouth disease virus capsid precursors is independent of other viral proteins and occurs in both mammalian and insect cells. *J. Gen. Virol.*, **72**, 747–751.

Bergelson, J., Shepley, M., Chan, B., Hemler, M. and Finberg, R. (1992). Identification of the integrin VLA-2 as a receptor for echovirus 1. *Science*, **255**, 1718–1720.

Bergelson, J., St. John, N., Kawaguchi, S., Chan, M., Stubdal H. Modlin, J. and Finberg, R. (1993). Infection by echovirus 1 and 8 depends on the $\alpha 2$ subunit of human VLA-2. *J. Virol.*, **67**, 6847–6852.

Bittle, J., Houghten, R., Alexander, H., Shinnick, T., Sutcliffe, J., Lerner, R., Rowlands, D. and Brown, F. (1982). Protection against foot–and–mouth disease by immunisation with a chemically synthesised peptide predicted from the viral nucleotide sequence. *Nature (London)*, **298**, 30–33.

Bloomer, A., Champness, J., Bricogne, G., Staden, R. and Klug, A. (1978). Protein disk of tobacco mosaic virus at 2.8Å resolution showing the interactions within and between subunits. *Nature (London)*, **276**, 362–368.

Blundell, T. and Johnson, L. *Protein crystallography.* Academic Press, New York, (1976).

Bolwell, C., Clarke, B., Parry, N., Ouldridge, E., Brown, F. and Rowlands, D. (1989). Epitope mapping of foot–and–mouth disease virus with neutralising monoclonal antibodies. *J. Gen. Virol.*, **70**, 59–68.

Bricogne, G. (1976). Methods and programs for direct–space exploitation of geometric redundancies. *Acta Cryst.*, **A32**, 832–847.

Bricogne, G. (1991). A multisolution method of phase determination by combined maximisation of entropy and likelihood. III. Extension to powder diffraction data. *Acta Cryst.*, **A47**, 803–829.

Bricogne, G. (1993). Direct phase determination by entropy maximisation and likelihood ranking: status report and perspectives. *Acta Cryst.*, **D49**, 37–60.

Brown, F., Cartwright, B. and Stewart, D. (1962). Further studies on the infection of pig–kidney cells by foot–and–mouth disease virus. *Biochim. Biophys. Acta.*, **55**, 768–774.

Brünger, A. *XPLOR Version 3.0.* New Haven, C.T., (1992).

Cantor, C. and Schimmel, P. *Biophysical chemistry Part II: Techniques for the study of biological structure and function.* W.H. Freeman and Co., San Francisco, (1980).

Carr, S., Biemann, A., Shoji, S., Parmelee, S. and Titani, K. (1982). *n*–Tetradecanoyl is the NH_2–terminal blocking group of the catalytic subunit of cyclic AMP–dependent protein kinase from bovine cardiac muscle. *Proc. Natl. Acad. Sci. USA*, **79**, 6128–6131.

Caspar, D. and Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold. Spring Harbour Symp. Quant. Biol.*, **27**, 12219–12223.

Cavanagh, D., Sangar, D., Rowlands, D. and Brown, F. (1977). Immunogenic and cell attachment site of FMDV: further evidence for their location in a single capsid polypeptide. *J. Gen. Virol.*, **35**, 149–158.

Chang, K., Auvinen, P., Hyppiä, T. and Stanway, G. (1989). The nucleotide sequence of coxsackie virus A9; implications for receptor binding and enterovirus classification. *J. Gen. Virol.*, **70**, 3269–3280.

Chapman, M. and Rossmann, M. (1993). Comparison of surface properties of picornaviruses: strategies for hiding the receptor site from immune surveillance. *Virol.*, **185**, 745–756.

Chou, P. and Fasman, G. (1974). Predictions of protein conformation. *Biochemistry*, **13**, 222–244.

Chow, M., Newman, J., Filman, D., Hogle, J., Rowlands, D. and Brown, F. (1987). Myristylation of picornavirus capsid protein VP4 and its structural significance. *Nature (London)*, **327**, 482–486.

Clarke, B., Sangar, D., Burroughs, J., Newton, S., Carroll, A. and Rowlands, D. (1985). Two initiation sites for foot–and–mouth disease virus polyprotein *in vivo*. *J. Gen. Virol.*, **66**, 2615–2626.

Colman, P., Varghese, J. and Laver, W. (1983). Structure of the catalytic and antigenic sites in influenza virus neuraminidase. *Nature*, **303**, 41–44.

Colonno, R., Candra, J., Mizutani, S., Callahan, P., Davies, M. and Murko, M. (1988). Evidence for the direct involvement of the rhinovirus canyon in receptor binding. *Proc. Natl. Acad. Sci. USA*, **85**, 5449–5453.

Crowther, J., Farias, S., Carpenter, W. and Samuel, A. (1993). Identification of a fifth neutralizable site on type O foot–and–mouth disease virus following characterisation of single and quintuple monoclonal antibody escape mutants. *J. Gen. Virol.*, **74**, 1547–1553.

Curry, S., Abu–Ghazaleh, R., Blakemore, W., Fry, E., Jackson, T., King, A., Lea, S., Logan, D., Newman, J. and Stuart, D. (1992). Crystallisation and

preliminary X–ray analysis of three serotypes of foot–and–mouth disease virus. *J. Mol Biol.*, **228**, 1263–1268.

Davies, D., Sheriff, S. and Padlan, E. *Cold Spring Harbour Symposia on Quantitative Biology.*, chapter Comparative study of two Fab–lysozyme crystal structures., pages 233–238. Volume LIV, Cold Spring Harbour Laboratory Press, New York, (1989).

Davies, D. and Padlan, E. (1990). Antibody–antigen complexes. *Annu. Rev. Biochem.*, **59**, 439–473.

Debye, P. (1914). Interferenz von röntgenstrahlen und wärmebewegung. *Annl. Phys.*, **43**, 49–95.

Diéz, J., Dávila, M., Escarmís, C., Mateu, M., Dominguez, J., Pérez, J., Giralt, E., Melero, J. and Domingo, E. (1990). Unique amino acid substitutions in the capsid proteins of foot–and–mouth disease virus from a persistent infection in cell culture. *J. Virol.*, **64**, 5519–5528.

Domingo, E., Mateu, M., Martínez, M., Dopazo, J., Moya, A. and Sobrino, F. *Applied Virology Research*, chapter Genetic variability and antigenic diversity in foot–and–mouth disease virus, pages 233–266. Volume 2, Plenum Publishing, New York, (1990).

Engh, R., Lobermann, H., Schneider, M., Wiegand, G., Huber, R. and Laurell, C. (1989). The S variant of human alpha 1–antitrypsin, structure and implications for function and metabolism. *Prot. Eng.*, **2**, 407–415.

Eriksson, A., Jones, T. and Liljas, A. (1988). Refined structure of carbonic anhydrase II at 2.0Å resolution. *Prot. Struct., Func.*, **4**, 274–282.

Esnouf, R. *Protein structure from simulated NMR databases.* D. Phil. thesis, Oxford University, (1992).

Felding-Harbermann, B. and Cheresh, D. (1993). Vitronectin and its receptors. *Current Opinion in Cell Biology*, **5**, 864–868.

Filman, D., Syed, R., Chow, M., Macadam, A., Minor, P. and Hogle, J. (1989). Structural factors that control conformational transitions and serotype specificity in type 3 poliovirus. *EMBO J.*, **8**, 1567–1579.

Fox, G. and Holmes, K. (1966). An alternative method of solving the layer scaling equations of Hamilton, Rollett and Sparks. *Acta Cryst.*, **20**, 886–891.

Fox, G., Stuart, D., Acharya, K., Fry, E., Rowlands, D. and Brown, F. (1987). Crystallisation and preliminary X–ray diffraction analysis of foot–and–mouth disease virus. *J. Mol. Biol.*, **196**, 591–597.

Fox, G., Parry, N., Barnett, P., McGinn, B., Rowlands, D. and Brown, F. (1989). Cell attachment site on foot–and–mouth disease virus includes the amino acid sequence RGD (arginine–glycine–aspartic acid). *J. Gen. Virol*, **70**, 625–637.

French, S. and Wilson, K. (1978). On the treatment of negative intensity observations. *Acta Cryst.*, **A34**, 517–525.

Fry, E., Logan, D., Acharya, R., Fox, G., Rowlands, D., Brown, F. and Stuart, D. (1990). Architecture and topography of an aphthovirus. *Sem. in Virol.*, **1**, 439–451.

Fry, E. *The three–dimensional structure of foot–and–mouth disease virus.* D. Phil. thesis, Oxford University, (1991).

Fry, E., Acharya, R. and Stuart, D. (1993). Methods used in the structure determination of foot–and–mouth disease virus. *Acta Cryst.*, **A49**, 45–55.

Fujinaga, M., Delbaere, L., Brayer, G. and James, M. (1985). Refined structure of α–lytic protease at 1.7Å resolution. Analysis of hydrogen bonding and solvent structure. *J. Mol. Biol.*, **184**, 479–502.

Genetics Computer Group. *Program Manual for the GCG Package, Version 7.* Wisconsin, USA., (1991).

Getzoff, E., Tainer, J., Lerner, R. and Geysen, H. (1988). The chemistry and mechanism of antibody binding to protein antigens. *Adv. In Immunol.*, **43**, 1–98.

Geysen, H., Barteling, S. and Meloen, R. (1985). Small peptides induce antibodies with a sequence and structural requirement for binding antigen comparable to antibodies raised against the native protein. *Proc. Natl. Acad. Sci. (USA)*, **82**, 178–182.

Geysen, H., Tainer, J., Rodda, S., Mason, T., Alexander, H., Getzoff, E. and Lerner, R. (1987). Chemistry of antibody binding to a protein. *Science*, **235**, 1184–1190.

Gilson, M., Sharp, K. and Honig, B. (1987). Calculating the electrostatic potential of molecules in solution: Method and error assessment. *J. Comp. Chem.*, **9**, 327–335.

Giranda, V., Chapman, M. and Rossmann, M. (1990). Modelling of the human intracellular adhesion molecule-1, the human rhinovirus major group receptor. *Proteins*, **7**, 227–233.

Grant, R., Filman, D., Fujinami, R., Icenogle, J. and Hogle, J. (1992). Three–dimensional structure of theiler virus. *Proc. Natl. Acad. Sci. USA.*, **89**, 2061–2065.

Greve, J., Davis, G., Meyer, A., Forte, C., Connolly–Yost, A., Marlor, C., Kamarck, M. and McClelland, A. (1989). The major human rhinovirus receptor is ICAM-1. *Cell*, **56**, 839–847.

Harrison, S., Olson, A., Schutt, C., Winkler, F. and Bricogne, G. (1978). Tomato bushy stunt virus at 2.9Å . *Nature (London)*, **276**, 368–373.

Helliwell, J. *Macromolecular crystallography with synchrotron radiation.* Cambridge University Press, (1992).

Hendrickson, W. (1985). Stereochemically restrained refinement of macromolecular structures. *Meth. Enzym.*, **115**, 252–303.

Henrick, K., Blow, D., Carrell, H. and Glusker, J. (1987). Comparison of backbone structures of glucose isomerase from *Streptomyces* and *Arthobacter. Protein Eng.*, **1**, 467–469.

Hernández, J., Martínez, M., Rocha, E., Domingo, E. and Mateu, M. (1992). Generation of a subtype–specific neutralisation epitope in foot–and–mouth disease virus of a different subtype. *J. Gen. Virol.*, **73**, 213–216.

Herzberg, O. and James, M. (1988). Refined crystal structure of tropnin C from turkey skeletal muscle at 2.0Å resolution. *J. Mol. Biol.*, **203**, 761–769.

Hogle, J., Chow, M. and Filman, D. (1985). Three–dimensional structure of poliovirus at 2.9Å resolution. *Science*, **229**, 1358–1365.

Hogle, J. and Filman, D. (1989). The antigenic structure of poliovirus. *Phil. Trans. R. Soc. Lond.*, **B 323**, 467–478.

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. and Van de Pol, S. (1982). Rapid evolution of RNA genomes. *Science*, **215**, 1577–1585.

Jack, A. and Levitt, M. (1978). Refinement of large structures by simultaneous minimisation of energy and R factors. *Acta Cryst.*, **A34**, 931–935.

Jin, L., Fendley, B. and Wells, J. (1992). High resolution functional analysis of antibody–antigen interactions. *J. Mol. Biol.*, **226**, 851–865.

Jones, T. (1985). Interactive computer graphics: FRODO. *Meth. Enzym.*, **115**, 157–171.

Jones, T., Zou, Y., Cowan, S. and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.*, **A47**, 110–119.

Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Cryst.*, **A32**, 922–923.

Kabsch, W. and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

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

Kahn, R., Fourme, R., Gadet, A., Janin, J. and André, D. (1982). Macromolecular crystallography with synchrotron radiation: Photographic data collection and polarisation correction. *J. Appl. Cryst.*, **15**, 330–337.

Kamphuis, I., Kalk, K., Swarte, M. and Drenth, J. (1984). Structure of papain refined at 1.65Å resolution. *J. Mol. Biol.*, **179**, 233–256.

Kim, S., Smith, T., Chapman, M., Rossmann, M., Pevear, D., Dutko, F., Felock, P., Diana, G. and McKinlay, M. (1989). Crystal structure of human rhinovirus serotype 1A (HRV1A). *J. Mol. Biol.*, **210**, 91–111.

Kitson, J., McCahon, D. and Belsham, G. (1990). Sequence analysis of monoclonal antibody resistant mutants of type O foot–and–mouth disease virus: evidence for involvement of the three surface exposed capsid proteins in four antigenic sites. *Virology*, **179**, 26–34.

Krausslich, H., Holscher, C., Reuer, Q., Harber, J. and Wimmer, E. (1990). Myristylation of the poliovirus polyprotein is required for proteolytic processing of the capsid and for viral infectivity. *J. Virol.*, **64**, 2433–2436.

Krishnaswamy, S. and Rossmann, M. (1990). Structural refinement and analysis of mengo virus. *J. Mol. Biol.*, **211**, 803–844.

Lascombe, M., Alzari, P., Boulot, G., Saludjian, P., Tougard, P., Berek, C., Haba, S., Rosen, E., Nisonoff, A. and Poljak, R. (1989). Three-dimensional structure of Fab R19.9, a monoclonal murine antibody specific for the *p*–azobenzenearsonate group. *Proc. Natl. Acad. Sci. (USA)*, **86**, 607–611.

Leahy, D., Hendrickson, W., Aukhil, I. and Erickson, H. (1992). Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science*, **258**, 987–991.

Lee, B. and Richards, F. (1971). The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.*, **55**, 379–400.

Liljas, L., Unge, T., Jones, T., Fridborg, K., Lövgren, S., Skoglund, U. and Strandberg, B. (1982). Structure of satellite tobacco necrosis virus at 3.0Å res-
olution. *J. Mol. Biol.*, **159**, 93–108.

Loeffler, F. and Frosch, P. (1897). Summarischer bericht über der ergebnisse der untersuchungen zur erforschung der maul- und klauenseuche. *Zentralblatt Bakt. Abt. Orig.*, **22**, 257–259.

Logan, D. *Investigation of the three–dimensional structure of viruses.* D. Phil. thesis, Oxford University, (1991).

Logan, D., Abu–Ghazaleh, R., Blakemore, W., Curry, S., Jackson, T., King, A., Lea, S., Lewis, R., Newman, J., Parry, N., Rowlands, D., Stuart, D. and Fry, E. (1993). Structure of a major immunogenic site on foot–and–mouth disease virus. *Nature (London)*, **362**, 566–568.

Luo, M., Vriend, G., Kamer, G., Minor, I., Arnold, E., Rossmann, M., Boegé, U., Scraba, D., Duke, G. and Palmenberg, A. (1987). The atomic structure of mengo virus at 3.0Å resolution. *Science*, **235**, 182–191.

Luo, M., Vriend, G., Kamer, G. and Rossmann, M. (1989). Structure determination of mengo virus. *Acta Cryst.*, **B45**, 85–92.

Luo, M., He, C., Toth, K., Zhang, C. and Lipton, H. (1992). Three–dimensional structure of theiler murine encephalomyelitis virus (BeAn strain). *Proc. Natl. Acad. Sci. USA.*, **89**, 2049–2413.

Main, P. and Rossmann, M. (1966). Relationships among structure factors due to identical molecules in different crystallographic enviroments. *Acta Cryst.*, **21**, 67–72.

Marc, D., Drugeon, G., Haenni, A., Girard, M. and van der Werf, S. (1989). Rôle of myristoylation of poliovirus capsid protein VP4 as determined by site–directed mutagenesis of its N–terminal sequence. *EMBO J.*, **8**, 2661–2668.

Marc, D., Masson, G., Girard, M. and van der Werf, S. (1990). Lack of myristoylation of poliovirus polypeptide VP0 prevents the formation of virions or results in assembly of non–infectious viral particles. *J. Virol.*, **64**, 4099–4107.

Marquart, M., Deisenhofer, J. and Huber, R. (1980). Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen–binding fragment at 3.0Å and 1.9Å. *J. Mol. Biol.*, **141**, 369–391.

Martínez, M., Hernández, J., Piccone, M., Palma, E., Domingo, E. and Knowles, N. (1991). Two mechanisms of antigenic diversification of foot–and–mouth disease virus. *Virology*, **167**, 695–706.

Martínez, M., Dopazo, J., Hernández, J., Mateu, M., Sobrino, F., Domingo, E. and Knowles, N. (1992). Evolution of the capsid protein genes of foot–and–mouth disease virus: antigenic variation without accumulation of amino acid substitutions over six decades. *J. Virol*, **66**, 3557–3565.

Martínez–Salas, E., Sáiz, J., Dávila, M., Belsham, G. and Domingo, E. (1993). A single nucleotide substitution in the internal ribosome entry site of foot–and–mouth disease virus leads to enhanced cap–independent translation *in vivo*. *J. Virol.*, **67**, 3748–3755.

Mason, P., Baxt, B., Brown, F., Harber, J., Murdis, A. and Wimmer, E. (1993). Antibody-complexed foot–and–mouth disease virus, but not poliovirus, can infect normally insusceptible cells via the Fc receptor. *Virol.*, **192**, 568–577.

Mateu, M., Da Silva, J., Rocha, E., de Brum, D., Alonso, A., Enjuanes, L., Domingo, E. and Barahoma, H. (1988). Extensive antigenic heterogeneity of foot–and–mouth disease virus of serotype C. *Virology*, **167**, 113–124.

Mateu, M., Martínez, M., Rocha, E., Andreu, D., Parejo, J., Giralt, E., Sobrino, F. and Domingo, E. (1989). Implications of a quasispecies genome structure: effect on frequent, naturally occurring amino acid substitutions on the antigenicity of foot–and–mouth disease virus. *Proc. Natl. Acad. Sci. USA.*, **86**, 5883–5887.

Mateu, M., Martínez, M., Capucci, L., Andreu, D., Giralt, E., Sobrino, F., Brocchi, E. and Domingo, E. (1990). A single amino acid substitution affects multiple overlapping epeitopes in the major antigenic site of foot–and–mouth disease virus of serotype C. *J. Gen. Virol.*, **71**, 629–637.

Mateu, M., Andreu, D., Carreño, C., Roig, X., Cairó, J., Camarero, J., Giralt, E. and Domingo, E. (1992). Non–additive effects of multiple amino acid substitutions on antigen–antibody recognition. *Eur. J. Immunol.*, **22**, 1385–1389.

McCahon, D., Crowther, J., Belsham, G., Kitson, J., Duchesne, M., Have, P., Meloen, R., Morgan, D. and De Simone, F. (1989). Evidence for at least four antigenic sites on type O foot–and–mouth disease virus involved in neutralisation; identification by single and multiple monoclonal antibody–resistant mutants. *J. Gen. Virol.*, **70**, 639–645.

Mendelsohn, C., Wimmer, E. and Racaniello, V. (1989). Cellular receptor for poliovirus: molecular cloning, nucleotide sequence and expression of a new member of the immunoglobulin superfamily. *Cell*, **56**, 855–865.

Minor, P. *Curr. Top. In Microbiol. and Immunol.*, chapter Antigenic structure of picornaviruses., pages 121–154. Volume 161, Springer–Verlag, Berlin, (1990).

Moscufo, N. and Chow, M. (1992). Myristate–protein interactions in poliovirus: interactions of VP4 threonine 28 contribute to the structural conformation of assembly intermediates and the stability of assembled virions. *J. Virol*, **66**, 6849–6857.

Moscufo, N., Yafal, A., Rogove, A., Hogle, J. and Chow, M. (1993). A mutation in VP4 defines a new step in the late stages of cell entry by poliovirus. *J. Virol*, **67**, 5075–5078.

Nordman, C. (1980). Procedures for detection and idealisation of non–crystallographic symmetry with application to phase refinement of the satellite tobacco necrosis virus structure. *Acta Cryst.*, , 747–754.

Novotńy, J., Handschumacher, M., Haber, E., Brucolleri, R., Carlson, W., Fanning, D., Smith, J. and Rose, G. (1986). Antigenic determinants in proteins coincide with surface accessible to large probes (antibody domains). *Proc. Natl. Acad. Sci. USA*, **83**, 226–230.

Novotńy, J., Handschumacher, M. and Bruccoleri, R. (1987). Protein antigenicity: a static surface property. *Immunology Today*, **8**, 26–31.

Oliveira, M., Zhao, R., Lee, W., Kremer, M., Minor, I., Rueckert, R., Diana, G., Pevear, D., Dutko, F., McKinlay, M. and Rossmann, M. (1993). The structure of human rhinovirus 16. *Structure*, **1**, 51–68.

Olson, N., Kolatkar, P., Oliveira, M., Cheng, R., Greve, J., McClelland, A., Baker, T. and Rossmann, M. (1993). Structure of a human rhinovirus complexed with its receptor molecule. *Proc. Natl. Acad. Sci. USA*, **90**, 507–511.

Overington, J., Donnelly, D., Johnson, M., Sali, A. and Blundell, T. (1992). Enviroment–specific amino acid substitution tables: tertiary templates and prediction of protein folds. *Prot. Sci.*, **1**, 216–226.

Page, G., Mosser, A., Holge, J., Filman, D., Rueckert, R. and Chow, M. (1988). Three–dimensional structure of poliovirus serotype 1 neutralising determinants. *J. Virol.*, **62**, 1781–1794.

Palmenberg, A. *Molecular Aspects of Picornavirus Infection and Detection.* American Society for Microbiology, Washington D.C., (1989).

Parry, N., Ouldridge, E., Barnett, P. and Brown, F. (1989). Neutralising epitopes of type O foot–and–mouth disease virus. II. Mapping three conformational sites with synthetic peptide reagents. *J. Gen. Virol.*, **70**, 1493–1503.

Parry, N., Fox, G., Rowlands, D., Brown, F., Fry, E., Acharya, R. and Stuart, D. (1990). Structural and serological evidence for a novel mechanism of immune evasion in foot–and–mouth disease virus. *Nature*, **347**, 569–572.

Pelletier, J. and Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature*, **334**, 320–325.

Pevear, D., Fancher, M., Felock, P., Rossmann, M., Miller, M., Diana, G., Treasurywala, A., McKinlay, M. and Dutko, F. (1989). Conformational change in the floor of the human rhinovirus canyon blocks absorption to HeLa cell receptors. *J. Virol.*, **63**, 2002–2007.

Pfaff, E., Mussgay, M., Bohm, H., Schulz, G. and Schaller (1982). Antibodies against a preselected peptide recognise and neutralise foot–and–mouth disease virus. *EMBO J.*, 1, 869–874.

Pfaff, E., Müller, H., Haas, B. and Thiel, H. *Vaccines '89: Modern approaches to new vaccines including prevention of aids*, chapter Molecular and immunological analysis of foot–and–mouth disease virus mutants., pages 449–455. Cold Spring Harbour Laboratory, New York, (1989).

Phillips, S. (1980). Structure and refinement of oxymyoglobin at 1.6Å resolution. *J. Mol. Biol.*, **142**, 531–554.

Pierschbacher, M. and Ruoslahti, E. (1987). Influence of streochemistry of the sequence Arg–Gly–Asp–Xaa on binding specificity in cell adhesion. *J. Biol. Chem.*, **262**, 17294–17298.

Rayment, I. (1983). Molecular replacement method at low resolution: optimum strategy and intrinsic limitations as determined by calculations on icosahedral virus models. *Acta Cryst.*, **A39**, 102–116.

Roivainen, M., Hyypiä, T., Pirainen, L., Kalkkinen, N., Stanway, G. and Hovi, T. (1991). RGD–dependent entry of coxsackievirus A9 into host cells and its bypass after cleavage of VP1 protein by intestinal proteases. *J. Virol.*, **65**, 4735–4740.

Rossmann, M., Arnold, E., Erickson, J., Frankenburger, E., Griffith, J., Hecht, H., Johnson, J., Kamer, G., Luo, M., Mosser, A., Rueckert, R., Sherry, B. and Vriend, G. (1985). Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature (London)*, **317**, 145–153.

Ruoshlati, E. and Pierschbacher, M. (1987). New perspectives in cell adhesion: RGD and integrins. *Science*, **238**, 491–497.

Saiz, J., González, M., Borca, M., Sobrino, F. and Moore, D. (1991). Identification of neutralising sites on VP1 and VP2 of type A5 foot–and–mouth disease virus, defined by neutralisation resistant variants. *J. Virol.*, **65**, 2518–2524.

Satow, Y., Cohen, G., Padlan, E. and Davies, D. (1986). Phosphocholine binding immunoglobulin Fab McPC603 an X–ray diffraction study at 2.7Å. *J. Mol. Biol.*, **190**, 593–604.

Saudek, V., Atkinson, R. and Pelton, J. (1991). Three–dimensional structure of echistatin, the smallest active RGD protein. *Biochemistry*, **30**, 7369–7372.

Schutt, C. and Winkler, F. *The rotation method in crystallography*, chapter The oscillation method for very large unit cells., pages 173–186. North–Holland, Amsterdam, (1977).

Sherry, B. and Rueckert, R. (1985). Evidence for at least two dominant neutralisation antigens on human rhinovirus 14. *J. Virol.*, **53**, 137–143.

Sherry, B., Mosser, A., Colonno, R. and Rueckert, R. (1986). Use of monoclonal antibodies to identify four neutralisation epiptopes on a common cold picornavirus, human rhinovirus 14. *J. Virol.*, **57**, 246–257.

Silva, A. and Rossmann, M. (1985). The refinement of southern bean mosaic virus in reciprocal space. *Acta Cryst.*, **B41**, 147–157.

Skarżyński, T., Moody, P. and Wonacott, A. (1987). Structure of holo–glyceraldehyde–3–phosphate dehydrogenase from *Bacillus stearothermophilus* at 1.8Å resolution. *J. Mol. Biol.*, **193**, 171–187.

Smith, T., Kramer, M., Luo, M., Vriend, G., Arnold, E., Kamer, G., Rossmann, M., McKinlay, M., Diana, G. and Otto, M. (1986). The site of attachment in human rhinovirus 14 for antiviral agents that inhibit uncoating. *Science*, **233**, 1286–1293.

Smith, T., Olson, N., Cheng, R., Liu, H., Chase, E., Lee, W., Leippe, D., Mosser, A., Rueckert, R. and Baker, T. (1993). Structure of human rhinovirus complexed with FAb fragments from a neutralizing antibody. *J. Virol.*, **67**, 1148–1158.

Smyth, M., Fry, E., Stuart, D., Lyons, C., Hoey, E. and Martin, S. (1993). Preliminary crystallographic analysis of bovine enterovirus. *J. Mol. Biol.*, **231**, 930–032.

Sobrino, F., Dávila, M., Ortín, J. and Domingo, E. (1983). Multiple genetic variants arise in the course of replication of foot–and–mouth disease virus in cell cultures. *Virology*, **128**, 310–318.

Sobrino, F., Palma, E., Beck, E., Dávila, M., de la Torre, J., Negro, P., Villanueva, N., Ortín, J. and Domingo, E. (1986). Fixation of mutations in the viral genome during an outbreak of foot–and–mouth disease: heterogeneity and rate mutations. *Gene*, **50**, 149–159.

Sobrino, F., Martínez, M., Carrillo, C. and Beck, E. (1989). Antigenic variation of foot–and–mouth disease virus of serotype C during propagation in the field is mainly restricted to only one structural protein (VP1). *Virus. Res.*, **14**, 273–280.

Staunton, D., Merluzzi, V., Rothlein, R., Barton, R., Marlin, S. and Springer, T. (1989). A cell adhesion molecule, ICAM-1, is the major surface receptor for rhinoviruses. *Cell*, **56**, 849–853.

Staunton, D., Dustin, M., Erickson, H. and Springer, T. (1990). The arrangement of the immunoglobulin–like domain of ICAM-1 and the binding sites for LFA-1 and rhinovirus. *Cell*, **61**, 243–354.

Stave, J., Card, J., Morgan, D. and Vakharia, V. (1988). Neutralisation sites of type O_1 foot–and–mouth disease virus defined by monoclonal antibodies and neutralisation–escape mutants. *Virol.*, **162**, 21–29.

Strohmaier, K., Franze, R. and Adam, K. (1982). Location and characterisation of the antigenic portion of the FMDV immunising protein. *J. Gen. Virol.*, **59**, 295–306.

Stuart, D. and Walker, N. Absorption corrections. In *Computational aspects of protein crystal data analysis.*, pages 25–37, (1987).

Surovoi, A., Ivanov, V., Chepurkin, A., Ivanyuschenkov, V. and Dryagalin, N. (1988). Is the Arg–Gly–Asp sequence the site for foot–and–mouth disease virus binding with cell receptor? *Soviet J. Bioorg. Chem.*, **14**, 965–968.

Sussmann, J., Holbrook, S., Church, G. and Kim, S. (1977). A structure factor least–squares minimisation procedure for macromolecular structures using constrained *and* restrained parameters. *Acta Cryst.*, **A33**, 800–804.

Thomas, A., Woortmeijer, R., Puijk, W. and Barteling, S. (1988). Antigenic sites on foot–and–mouth disease virus type A_{10}. *J. Virol.*, **62**, 2782–2789.

Tronrud, D., Ten Eyck, L. and Matthews, B. (1978). An efficient general purpose least–squares refinement program for macromolecular structures. *Acta Cryst.*, **A34**, 931–935.

Tsao, J., Chapman, M., Agbandje, M., Keller, W., Smith, K., Wu, H., Luo, M., Smith, T., Rossmann, M., Compans, R. and Parrish, C. (1991). The three–dimensional structure of canine parvovirus and its functional implications. *Science*, **251**, 1456–1464.

Tulip, W., Varghese, J., Baker, A., van Dankelaar, A., Laver, G., Webster, R. and Colman, P. (1991). Refined atomic structures of N9 subtype influenza virus neuraminidase and escape mutants. *J. Mol. Biol.*, **221**, 487–497.

Vallée, H. and Carré, H. (1922). Sur la pluralité du virus aphteux. *Compt. Rend. Hebd. Seanc. Acad. Sci. Paris*, **174**, 1498–1500.

Van Bekkum, J., Frenkel, H., Frederiks, H. and Frenkel, S. (1959). Observations on the carrier state of cattle exposed to foot–and–mouth disease virus. *Tijschr. Diergeneeskd.*, **84**, 1159–1164.

Villaverde, A., Martínez–Salas, E. and Domingo, E. (1988). 3D gene of foot–and–mouth disease virus. Conservation by convergence of average sequences. *J. Mol. Biol.*, **204**, 771–776.

Wang, B. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Meth. Enzym.*, **115**, 90–117.

Westhof, E., Altschuh, D., Moras, D., Bloomer, A., Mondragon, A., Klug, A. and van Regenmortel, M. (1984). Correlation between segmental mobility and the location of antigenic determinants in proteins. *Nature (London)*, **311**, 123–126.

Wild, T., Burroughs, N. and Brown, F. (1969). Surface structure of foot–and–mouth disease virus. *J. Gen. Virol.*, **4**, 313–320.

Wilson, A. (1942). Determination of absolute from relative X–ray intensity data. *Nature (London)*, **150**, 151–152.

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

Wistow, G., Turnell, B., Summers, L., Slingsby, C., Moss, D., Miller, L., Lindley, P. and Blundell, T. (1983). X–ray analysis of the eye lens protein γ–II crystallin at 1.9Å resolution. *J. Mol. Biol.*, **170**, 175–202.

Wonacott, A. *The rotation method in crystallography*, chapter Geometry of the rotation method, pages 75–103. North–Holland, Amsterdam, (1977).

Xie, Q., McCahon, D., Crowther, J., Belsham, G. and McCullough, K. (1987). Neutralisation of foot–and–mouth disease virus can be mediated through any of a least three separate antigenic sites. *J. Gen. Virol.*, **68**, 1637–1647.