Structural Studies on Negative Strand Virus Proteins

Jane Alexandra Potter

A Thesis Submitted for the Degree of Doctor of Philosophy

University of St Andrews

June 2002

Declaration

(i) I, Jane Potter, hereby certify that this thesis, which is approximately 23,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date 1/7/02 Signature of candidate

(ii) I was admitted as a research student in October 1998 and as a candidate for the degree of PhD in October 1999; the higher study for which this is a record was carried out in the University of St Andrews between 1998 and 2002.

Date 1/7/02 Signature of candidate

(iii) I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date 29/7/02 Signature of supervisor

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

Date 1/7/02 Signature of candidate

Abstract

Negative strand viruses (NSVs) include important human and animal pathogens. The filoviruses (Ebola and Marburg) are among the most pathogenic NSVs, yet knowledge about their genetics, natural history and pathogenicity is limited.

The most abundant protein of filoviruses is the nucleoprotein (NP), which polymerises to encapsidate the RNA genome. The first part of this project investigates the oligomeric state of recombinant Ebola NP *in vitro* and its suitability for structure elucidation by X-ray crystallography. Electron microscopy reveals that, in the absence of other viral proteins, recombinant NP oligomerises to form unusual ring-shaped structures that have not previously been observed either *in vitro* or in Ebola virions. Attempts were made to solubilise both truncated (residues 1-399) and full length NP for crystallisation purposes. Evidence that the truncated form corresponds to a separate sub-domain with the ability to self-assemble is discussed.

The second part of the project involves another family of NSVs – the Paramyxoviruses. An antibody (SV5-Pk), has previously been found to bind an epitope on the Simian Virus 5 Phosphoprotein with unusually high affinity. The aim of the work was to obtain the crystal structure of the Fab fragment of SV5-Pk in both the presence and absence of a 14 amino acid peptide containing the epitope sequence. The structures of the Fab fragment and the Fab-peptide complex have been obtained at 2.3 Å and 1.9 Å resolution respectively. The structure of the complex provides an insight into the unusually high antigen-binding affinity and has also facilitated the

design of peptides with slightly reduced affinity for use as potential tags for protein purification and immunodetection techniques.

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Garry Taylor for all his help and encouragement over the course of my PhD. Financial support for my work was provided by a BBSRC Case studentship and the Centre for Applied Microbiological Research (CAMR). I am grateful to Tony Fooks at CAMR for all his help with the Ebola work.

I would also like to thank the other members of the lab for providing assistance and useful suggestions, as well as for being good company both inside the lab and out.

Finally, I have to thank Matt for his support and good humour throughout and for generally just being Matt.

Table of Contents

Table of Figures

Chapter 3.

Chapter 4.

Abbreviations

Ab	Antibody
ACS	Antibody combining site
BSA	Bovine serum albumin
CDR	Complementarity determining region
CMC	Critical micellar concentration
CC	Correlation coefficient
Da	Dalton
DLS	Dynamic light scattering
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
F	Fusion protein
Fab	Antigen-binding antibody fragment
Fc	Readily crystallisable antibody fragment
FT	Fourier transform
GP	Glycoprotein
GST	Glutathione S-transferase
HN	Haemagglutinin Neuraminidase
Ig	Immunoglobulin
IPTG	Isopropyl-β -D-thiogalactoside
ITC	Isothermal titration calorimetry
LB	Luria broth
LDS	Lithium dodecyl sulfate

M	Matrix protein
mAb	Monoclonal antibody
MALDI-TOF	Matrix-Assisted Laser Desorption/İonisation Time-of-Flight
MHC	Major histocompatability complex
MOI	Multiplicity of infection
MV	Measles virus
NNS	Non-segmented negative strand (virus)
NP	Nucleoprotein
OD	Optical density
P	Phosphoprotein
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PMSF	phenymethylsulfonyl fluoride
RABV	Rabies virus
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RSV	Respiratory syncytial virus
SAH	S-adenosyl homocysteine
SC	Surface complementarity
SDR	Structurally-determining residue
SDS	Sodium dodecyl sulphate
SEN	Sendai virus

SMAA	Solid matrix antibody-antigen (complex)
SV5	Simian parainfluenza irus 5
TCEP	Tris (2-carboxyethyl) phosphine
TLS	Tubule-like structures
VP	Viral protein
VSV	Vesicular Stomatitis Virus

Amino Acids:

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

Chapter 1

Introduction

1. 1 Mononegavirales

Negative sense RNA viruses are classified within the order Mononegavirales and include the Filoviridae, Paramyxoviridae, Bornaviridae and Rhabdoviridae families. Examples of genera and species within each family are given in Table 1.

Family	Subfamily	Genus	Species
Paramyxoviridae	*Paramyxovirinae*	*Respirovirus*	human parainfluenza virus 1
		Morbillivirus	measles virus
		Rubulavirus	simian virus 5
	Pneumovirinae	*Pneumovirus*	human respiratory syncytial virus
Filoviridae		*Ebola-like viruses*	ebola virus
		Marburg-like viruses	marburg virus
Rhabdoviridae		*Vesiculovirus*	vesicular stomatitis virus
		Lyssavirus	rabies virus
		Ephemerovirus	bovine ephemeral fever virus
		Cytorhabdovirus	lettuce necrotic yellow virus
		Nucleorhabdovirus	potato yellow dwarf virus
Bornaviridae		*Bornavirus*	Borna disease virus

Table 1. Taxonomic Classification within the Mononegavirales

1.2 The Filoviruses – Ebola and Marburg

Viral haemorrhagic fever is caused by several RNA viruses including members of the Arenaviridae, Bunyaviridae, Flaviviridae and Filoviridae. The filoviruses are among the most pathogenic of human viruses and are classified as biosafety level 4 agents based on their high mortality rates, person-to-person transmission, aerosol infectivity and lack of a protective vaccine or antiviral drug(World Health Organization 1985). Despite extensive investigations, filoviruses are the most poorly understood of virus families - very little is known about the natural history and reservoirs of these viruses, and knowledge of their genetics and pathogenicity is limited.

The Filoviridae family includes four sub-types of Ebola (Zaire, Sudan, Ivory Coast and Reston) and the Marburg virus (ICTV 1996). Ebola virus first emerged in 1976 in two simultaneous outbreaks in Zaire and Sudan. Over 500 cases were reported, with mortality rates of 88% in Zaire and 53% in Sudan (Baron, McCormick et al. 1983). Ebola virus re-emerged in 1979 in south Sudan and again in 1995 in Kikwit, Zaire (1978; 1995). In 1992 the Ivory Coast subtype was isolated from a laboratory worker who autopsied a dead chimpanzee from the southwest Ivory Coast (Le Guenno, Formentry et al. 1995). Reston virus, serologically related to Ebola virus, was isolated in 1989 from cynomolgus monkeys which originated in the Philippines. This subtype is infectious to humans but does not appear to cause serious disease (1990).

1.2.1 Morphology

Filovirus particles are similar in morphology to rhabdoviruses but much longer, appearing under the electron microscope as filamentous forms, often U- or 6-shaped (ICTV 1991) (Fig. 1). Virions vary greatly in length but have a uniform diameter of approximately 80 nm. Particles are composed of a helical ribonucleocapsid complex (RNP) surrounded by a lipid envelope derived from the host cell plasma membrane (Fig. 2) (Peters, Muller et al. 1971; Murphy, van der Groen et al. 1978; Kiley, Bowen et al. 1982).

The nucleocapsid complex contains a central channel (20 nm in diameter) and an outer helical layer (50 nm in diameter) with cross striations at 5 nm intervals. The complex is composed of a single negative stranded RNA molecule associated with

four of the seven structural proteins: Nucleoprotein (NP), Virion structural protein 30 (VP30), VP35 and the large protein (L). The three remaining proteins (Glycoprotein (GP), VP24 and VP40) are membrane associated (Kiley, Regnery et al. 1980). GP is a transmembrane protein anchored with the C-terminus in the virion membrane and homotrimers of GP form spikes on the virion surface. VP24 and VP40 are probably located at the inner surface of the membrane. All Ebola viruses possess a non-structural protein, sGP.

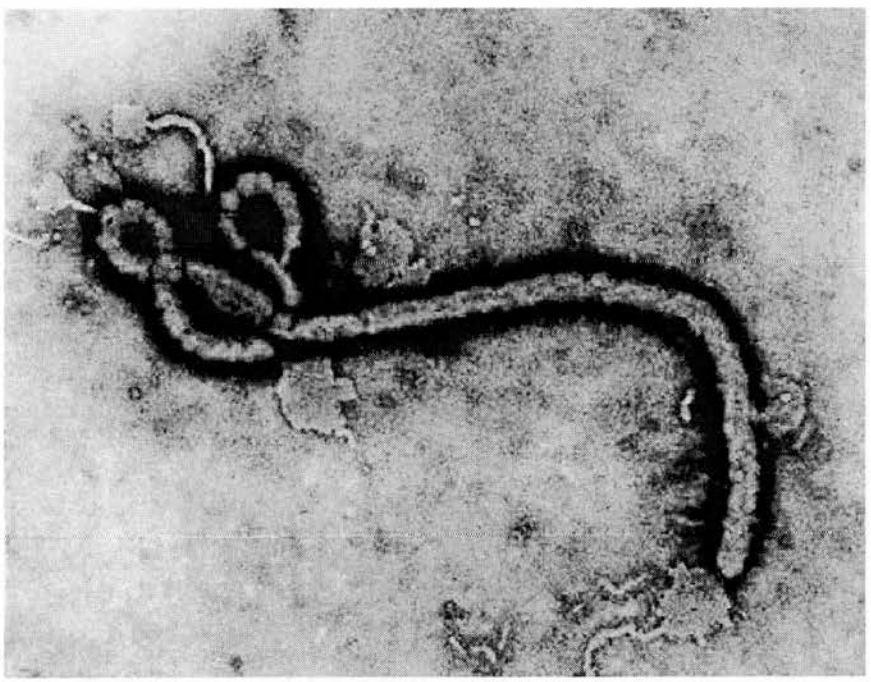


Fig 1. Electron Micrograph of an Ebola (Zaire)Virus virion. Specimen in cell culture at 160,000 X magnification. (Photograph taken by F. A. Murphy in 1976 at CDC).

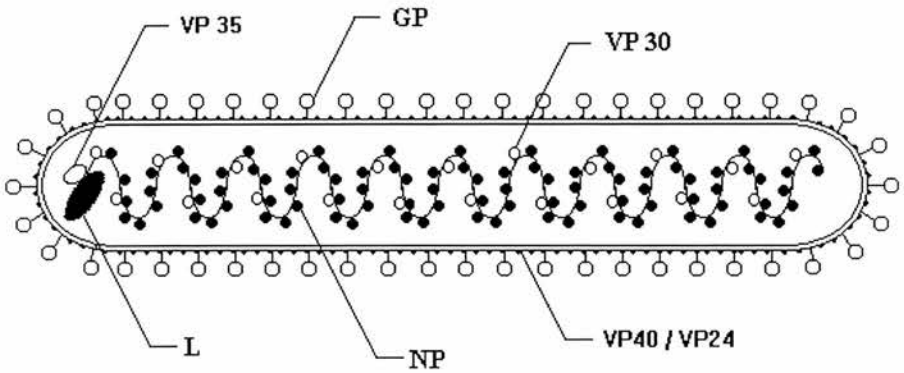


Fig.2 Diagrammatic representation of the Ebola Virus proteins

1.2.2 Filovirus Disease

The Ebola virus disease begins with a sudden onset of fever, severe headache, weakness, muscle aches followed by vomiting, abdominal pain, diarrhoea, pharyngitis and conjunctivitis. As the disease progresses, blood fails to clot and patients bleed into the gastrointestinal tract and internal organs. The incubation time for the disease is usually 5-10 days and the time from onset of illness until death or improvement is usually 7-10 days. Death is believed to result directly from damage to internal tissues (Peters and Khan 1999).

1.2.3 Treatment and Prevention

Transmission occurs primarily via close personal contact with hospital care workers or family members tending to the ill or dead. Unfortunately, no effective vaccine or treatments currently exist for filoviral infections. Ribavirin, an antiviral drug used to treat several other haemorrhagic fevers has no effect on filoviruses *in vitro* and fails to protect experimentally infected animals (Huggins 1989).

Eight patients in the 1995 Zaire outbreak were transfused with blood from convalescent patients (Mupapa, Massamba et al. 1999). Only one of the transfused patients (12.5%) died; this number is significantly lower than the overall fatality rate (80%) of the epidemic. However, because of the small number of patients and lack of control subjects, these data remain inconclusive.

Hyperimmune horse serum to Ebola virus has been shown to neutralise the virus both *in vitro* and in guinea pigs and mice (Jahrling, Geisbert et al. 1999). Onset of illness was delayed in infected Rhesus monkeys treated with the IgG but none survived.

GP is the only viral protein known to be present on the surface of virions and infected cells and may therefore be an important target for protective antibodies. A recent study has identified several protective monoclonal antibodies to Ebola virus GP (Huggins, Zhang et al. 1999). Three of the five unique epitopes bound by the mAbs are linear sequences on the molecule, the remaining two are conformational epitopes present in both GP and sGP. Ten of the fourteen mAbs protected mice when administered 24 hours before a lethal challenge with Ebola (Zaire). Some of these mAbs were effective even when administered two days after infection.

Targets for drug development include molecules that are required for viral replication, whose inhibition blocks replication either because the target is unique to the virus or the cellular analogue of the target is not affected at the concentrations required to inhibit viral replication. One such cell-encoded enzyme is *S*-adenosylhomocysteine (SAH) hydrolase. Huggins et al. demonstrated that an analogue inhibitor of the hydrolase, carbocyclic 3-deazaadenosine, was completely effective at protecting mice infected with Ebola virus and 90% effective when

administered two days after challenge (Huggins, Zhang et al. 1999). A second analogue, 3-deazaneplanocin A, has been shown to provide equal protection but without acute toxicity (Bray, Driscoll et al. 2000). These inhibitors are the first compounds demonstrated to cure animals from Ebola virus infection and, as the hydrolase is highly conserved among species, show promise as a potential therapy for filovirus infection in humans.

1.2.4 Genome

Filovirus genomes have a length of approximately 19 kb (Marburg 19.1 kb, Ebola 18.9 kb), significantly larger than other members of the order Mononegavirales (Regnery, Johnson et al. 1980). The following gene order is characteristic for filoviruses: 3` leader-NP-VP35-VP40-GP-VP30-VP24-L-5` trailer (Fig. 3) (Feldmann, Muhlberger et al. 1992; Volchkov, Blinov et al. 1993). Each gene is flanked by conserved transcriptional start (3`-CUNCNUNUAAUU-5`) and termination (3`-UAAUUCUUUUU-5`) signals. Both signals carry the pentamer 3`-UAAUU-5` which is a unique feature among filoviruses (Kiley, Wilusz et al. 1986; Sanchez, Kiley et al. 1992). The function of the pentamer is unclear, but it may serve as a recognition site for the polymerase complex. The genes possess unusually long non-coding sequences at their 3` and/or 5` ends. The 5` untranslated regions are predicted to form stable hairpin structures, which may have a role in transcript stability and ribosome binding (Sanchez, Kiley et al. 1992). The role of the gene overlaps that occur in Ebola virus and Marburg virus genomes is unknown.

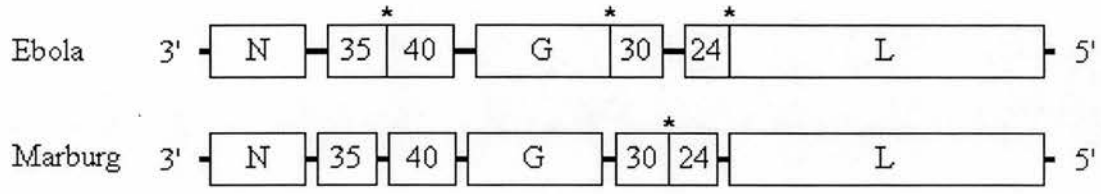


Fig. 3. Filovirus gene order. Asterisks indicate positions of gene overlap.

1.2.5 Replication

Marburg and Ebola viruses cause lytic infections in cells. Cell entry is presumably mediated by GP, the only surface protein of filovirus particles. Ultrastructural studies show that filoviruses are closely associated with coated pits along the plasma membrane suggesting that entry occurs by endocytosis of the virion (Geisbert and Jahrling 1995).

Transcription and translation occur in the cytoplasm of infected cells. GP and sGP are translated at membrane-bound ribosomes and are directed by an N-terminal hydrophobic domain into the endoplasmic reticulum where they undergo a complex series of processing events including signal peptide removal, N-glycosylation and trimerisation (Feldmann, Will et al. 1991; Geyer, Will et al. 1992; Will, Muhlberger et al. 1993). The glycoprotein is then transported via the Golgi body to the host cell membrane.

All other viral proteins are translated at free ribosomes in the cytosol. NP and VP40 are phosphorylated, possibly by the RNA-dependent RNA polymerase which is thought to possess kinase activity (Elliott, Kiley et al. 1985; Muhlberger, Sanchez et al. 1992).

The switch between transcription and replication is unknown but, as with other NNS RNA viruses, the synthesis of NP is thought to be a key factor. Replication occurs via a full length positive (+)-strand which serves as a template for the synthesis of the genomic negative (-)-strand RNA. Encapsidated RNA forms ribonucleoprotein (RNP) complexes that bud at the plasma membrane, incorporating GP, VP40 and VP24 (Schnittler, Mahner et al. 1993).

1.2.6 Filovirus Proteins

Gene 1 at the 3` end of the Ebola virus genome encodes the 739 amino acid nucleoprotein (NP)(Sanchez, Kiley et al. 1989; Bukreyev, Volchkov et al. 1995). The protein has a calculated molecular weight of 83 kDa and an observed weight of 104 kDa, as determined by SDS PAGE. Ebola virus and Marburg virus NP (78 kDa) possess unusually high molecular weights relative to other NNS RNA NPs which typically fall into the 42-68 kDa range.

Only the phosphorylated form of NP is found in virion particles, implying that phosphorylation is required to allow stable RNP complex formatio (Becker, Huppertz et al. 1994).

Sequence alignments of filoviral nucleoproteins reveal a region of very high homology between amino acids 130 and 392 including a 34 amino acid stretch of identical residues from position 296. This region also has a significant degree of identity with paramyxoviruses and to a lesser extent with those of rhabdoviruses (Sanchez, Kiley et al. 1992).

NP contains a large amount of negatively charged amino acids resulting in a highly acidic protein. The extreme net negative charge (-28) is the likely explanation

for the discrepancy between the calculated and observed Mr values, since binding of SDS to a peptide influences migration of a protein on SDS PAGE. The majority of the acidic and proline residues are found in the C-terminal half of the Marburg virus and Ebola virus NPs, and all three cysteine residues are found in the N-terminal third of the protein. Hydropathy plots of filovirus NPs show that the Ebola virus and Marburg virus proteins can be divided into a hydrophobic N-terminal half and a hydrophilic C-terminal half. A similar pattern is observed in other NNS RNA virus NPs such as Sendai NP and, to a lesser extent, Vesicular Stomatitis (VSV) NP. The similarity in the hydropathy profile is particularly prominent in the hydrophobic sequences from amino acids 130 to 320 (160 to 350 for Sendai). The N-terminal half may have a role in protein folding and/or RNA binding. It has been suggested that the less conserved C-terminus is exposed to the environment and may function in virion assembly by interacting with the matrix proteins (Kiley, Cox et al. 1988; Sanchez, Kiley et al. 1992).

Virion Structural protein 35 (VP35), encoded by gene 2, is also present within the RNP complex although its association is much weaker than that of NP and VP30 (Elliott, Kiley et al. 1985). Like NP, the protein may exist in both a phosphorylated and an unphosphorylated form. VP35 possesses an N-terminal hydrophilic region thought to be involved in template binding. The genome position of the corresponding gene and the fact that VP35 is associated with the RNP complex suggests that VP35 is functionally analogous to the phosphoproteins of paramyxoviruses and rhabdoviruses.

VP40 is not RNP associated but is known to bind non-specifically to nucleic acids (Elliott, Kiley et al. 1985). The protein is thought to have a role in regulation of

transcription/replication similar to that of the matrix protein of VSV. The protein's predominantly hydrophobic profile, its presence in large amounts in virions and the position of the corresponding gene indicate that VP40 is the matrix protein analogue of filoviruses (Kiley, Cox et al. 1988).

Gene 4 encodes the transmembrane glycoprotein. The protein is expressed as a precursor that is post-translationally cleaved into the disulphide-linked fragments GP1 and GP2 by a cellular furin (Volchkov, Feldmann et al. 1998). The two subunits are analogous to HA1 and HA2 from influenza, and the surface subunit and transmembrane subunit from retroviruses. In each case the first subunit binds to cell-surface receptors and the second mediates membrane fusion. Although filovirus GPs exhibit little homology with other NNS RNA virus envelope proteins, the glycoprotein contains a proteolytically resistant core that has remarkable structural similarity to a retrovirus envelope protein domain (Malashkevich, Schneider et al. 1999). The overall architecture of these domains, a trimer of helical hairpins, is a feature shared by many viral membrane-fusion proteins including those from influenza, HIV-1, SIV and SV5. GP is one of the main targets for the host immune response.

VP30, also known as the minor nucleoprotein, is associated with the RNP (Elliott, Kiley et al. 1985). Gene 6 encodes VP24, which is membrane associated and is thought to serve as a second matrix protein (Feldmann, Muhlberger et al. 1992).

The large protein (L) is regarded as an RNA-dependent RNA polymerase (Kamer and Argos 1984). Despite little overall sequence homology, the protein contains a 4 residue sequence (GDNQ) that is absolutely conserved among all known NNS RNA virus polymerases. The motif is flanked by hydrophobic residues and

appears to be required for polymerase activity (Sleat and Banerjee 1993; Schnell and Conzelmann 1995).

1.3 Paramyxoviridae

The Paramyxovirinae are divided into three genera: Respirovirus, Rubulavirus and Morbillivirus (ICTV 1996). Like filovirus genomes, the paramyxovirus genetic material consists of a non-segmented negative RNA strand. The distinguishing morphological feature of Paramyxoviridae is the size and shape of the nucleocapsids. Another feature of the Paramyxoviridae is the presence of the F protein that mediates cell-virus membrane fusion. Within the Paramyxovirinae, viruses are classified according to several criteria including the presence (Respirovirus and Rubulavirus) or absence (Morbillivirus) of neuraminidase activity and the coding capacity of the P gene.

1.3.1 Morphology

The Paramyxoviridae are enveloped viruses and generally spherical, with diameters ranging from 150 to 350 nm. The genetic material is encapsidated within the helical ribonucleoprotein (RNP) complex. The protein components of the complex (NP, P and L) are analogous to those present in the filoviral RNP. The RNP is surrounded by a lipid bilayer derived from the host cell membrane (Fraenkel-Conrat. 1975). The membrane contains two types of protein, hemagglutinin-neuraminidase (HN) and the fusion protein (F) (Fig. 4).

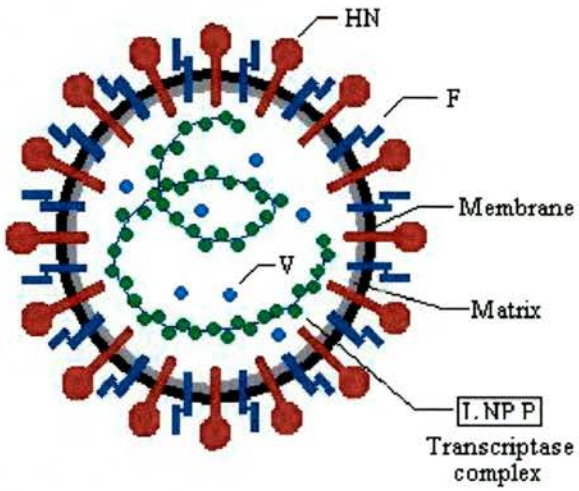


Fig. 4. Diagrammatic representation of Paramyxovirus proteins

1.3.2 Simian Parainfluenza Virus 5

Simian parainfluenza virus (SV5) is a member of the genus Rubula virus that also includes Sendai virus, Human parainfluenza virus types 1-4, Measles virus, Mumps virus and Newcastle disease virus. The 15.2 bp genome contains seven genes that encode eight proteins (Fig. 5). The genes are separated by repeated sequences: a polyadenylation signal at the end of the gene, an intergenic sequence, and a translation start signal at the beginning of the next gene (Fields, Knipe et al. 1996).

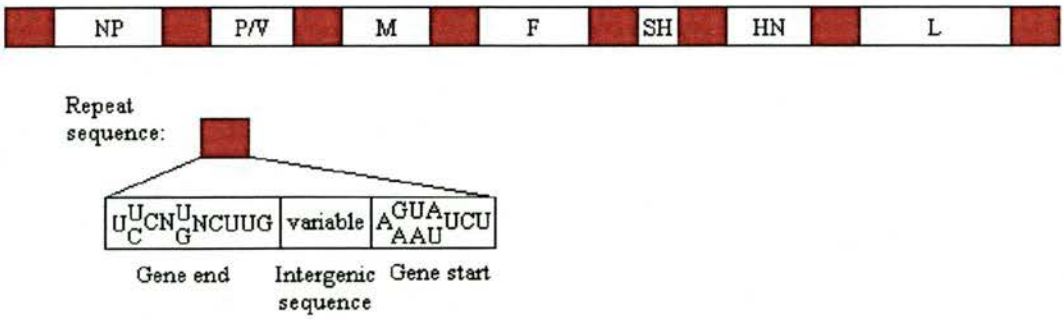


Fig. 5. Simian virus 5 gene order. The repeat sequence occurring between genes is shown in detail.

1.3.3 SV5 Proteins

HN is a multifunctional glycoprotein that recognises sialic acid on host cell receptors. The protein also cleaves sialic acid from progeny virions, preventing self-aggregation. In addition, HN of some paramyxoviruses assists virus entry by promoting the fusion activity of the F protein.

The most abundant of the Paramyxovirus proteins, the matrix protein (M), has a central role in virus assembly and budding (Blumberg, Rose et al. 1984). M has been shown to interact with proteins associated with both the lipid membrane (F and HN) and the ribonucleoprotein complex.

The rubulaviruses SV5 and Mumps virus and members of the Pneumovirinae possess a small integral membrane protein of unknown function called SH. Deletion of the SH gene from recombinant RSV has been shown to have little effect on viral replication both *in vitro* and *in vivo* (Bukreyev, Whitehead et al. 1997).

The coding capacity of the Paramyxovirinae genomes is greatly increased by RNA editing processes occurring during transcription of the P gene. The P gene of

SV5 encodes two proteins - V and the phosphoprotein (P) - which are identical in sequence from amino acids 1-164 but differ in their C-termini (Paterson, Harris et al. 1984). The V protein mRNA is a faithful copy of the P gene whereas the P mRNA contains two extra G residues (Thomas, Lamb et al. 1988). These two non-templated residues are thought to be added by the polymerase as it stutters at a sequence (3`-AAAAUUCU-5`) upstream of the insertion site. This sequence is similar to the polyadenylation signal at which the viral polymerase stutters, generating a stretch of Us that results in the long poly(A)$^{+}$ tail (Vidal, Curran et al. 1990).

Among the paramyxoviruses, the highly phosphorylated P proteins show little conservation of sequence. In contrast, the unique C-termini of the V proteins are highly similar and contain seven cysteine residues arranged in a putative Zinc finger. In SV5, V appears to play a role in circumvention of the host interferon (IFN) response by targeting the transcription factor STAT1 for degradation (Young, Didcock et al. 2000).

The phosphoprotein has a central role in both genome replication and transcription. The protein interacts with L to form the viral polymerase complex (Curran, Marq et al. 1992). P also aids the RNA encapsidation process by forming a complex with NP, keeping NP soluble prior to its binding to the nascent RNA chain (Curran, Boeck et al. 1995).

Sequence alignments indicate that Paramyxovirus phosphoproteins can be divided into several domains (Curran, Boeck et al. 1995). Sendai P contains an N binding domain (a.a. 33-41), an oligomerisation domain (a.a. 320-446) which also includes the L-binding domain (a.a. 412-445) and an N-RNA complex-binding domain (479-568) (Fig. 6). All these regions appear to be essential as deletions in any

of the domains renders P non-functional (Curran, Boeck et al. 1991). The X-ray structure of the Sendai P oligomerisation domain reveals that the protein forms a tetrameric coiled coil (Tarbouriech, Curran et al. 2000).

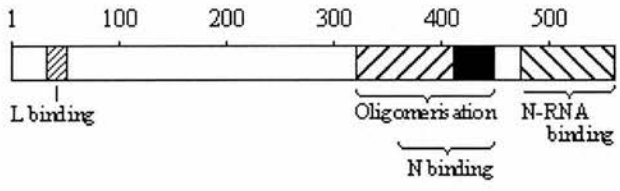


Fig. 6. Domains of the Sendai virus Phosphoprotein

1.4 The Ribonucleoprotein Complex

The interaction of nucleocapsid proteins is not only essential for assembly of the core structure but also has functions in viral transcription and replication.
To determine the composition of the ribonucleocapsid complex, Elliot *et al.* treated purified Ebola virus with detergent in increasing concentrations of NaCl followed by centrifugation in a sucrose cushion (Elliott, Kiley et al. 1985). Even in low salt conditions GP, VP40 and VP24 were dissociated from the RNP. VP35, loosely associated with the RNP, was removed with increasing salt concentrations. NP, L and VP30 remained in the pellet, even up to 1M salt, suggesting an intimate association with virion RNA.

Proteins analogous to each of the seven Ebola virus proteins have been shown to occur in Marburg virus (Becker, Huppertz et al. 1994). Upon salt dissociation of

isolated Marburg virions four proteins (NP, VP35, VP30 and L) remained attached to the core complex and were localised in inclusion bodies, presumed to be areas of nucleocapsid formation. The intracellular localisation of Marburg virus proteins has been investigated by immunofluorescent analysis of HeLa cells expressing recombinant nucleocapsid proteins (Becker, Rinne et al. 1998). When expressed in the absence of other viral proteins, VP35 and VP30 were distributed homogeneously throughout the cytoplasm instead of being concentrated in inclusion bodies, as in Marburg virus-infected cells. Large aggregates resembling these inclusion bodies were found only in cells expressing NP. Both VP30 and VP35, when co-expressed with NP, co-localised to the inclusion bodies suggesting NP-VP30 and NP-VP35 complex formations. No interactions were observed when NP was co-expressed with L, VP40 or GP. The presence of a further complex, between VP35 and L, was demonstrated by co-immunoprecipitation using both anti-VP35 and anti-L antibodies.

To further investigate complex formation between NP, VP35 and L, the proteins were co-expressed in the HeLa cells. The presence of VP35 was required to localise L to the NP-induced inclusion bodies. The same experiment in which VP35 was replaced with VP30 did not result in co-localisation of L. In summary, Marburg virus NP is able to form complexes with VP35 and VP30, and a trimeric complex occurs in which VP35 connects L and NP.

The complex between VP35 and L represents the active viral polymerase (Banerjee and Barik 1992). In the paramyxoviruses studied so far the binding site for P (the VP35 analogue of paramyxoviruses) is located in the N-terminal half of L proteins of Simian Virus 5, Sendai virus and Measles virus (Horikami, Smallwood et al. 1994; Parks 1994; Chandrika, Horikami et al. 1995). Similarly, the region

between amino acids 390 and 530 of Marburg virus L has been shown to be essential for binding VP35 (Becker, Huppertz et al. 1994). In contrast, the P binding site on the L proteins of two rhabdoviruses, VSV and Rabies virus, is located on the C-terminal half of the protein (Canter and Perrault 1996; Chenik, Schnell et al. 1998).

Little is known about nucleocapsid formation or the protein-protein interactions involved. The ability of NPs to self-assemble independently of other viral protein appears to be a general property among the order Mononegavirales. Kolesnikova and co-workers carried out electron microscopal studies to compare the behaviour of recombinant NP in HeLa cells. Recombinantly expressed NP assembled into tubule-like structures (TLS) very similar to those observed in Marburg virus-infected cells. The composition of the TLS was confirmed by immunogold labelling using a monoclonal antibody directed against Marburg virus NP (Kolesnikova, Muhlberger et al. 2000).

In Paramyxoviruses, nucleocapsid assembly has been most extensively studied in Sendai virus and to a lesser extent Measles virus. Measles NP expressed in animal and bacterial cells has been shown to form nucleocapsid-like structures in the absence of other viral proteins (Spehner, Kirn et al. 1991; Fooks, Stephenson et al. 1993). The capsids appeared to be less tightly coiled than those seen in measles-infected cells and the overall length of the structures was generally smaller. Further studies of measles NP showed that, after three days post-infection, NP recombinantly expressed in insect cells is cleaved into proteins of lower molecular weight. A similar process is observed in measles-infected mammalian cells (Fooks, Stephenson et al. 1993). Nucleocapsid-like structures partially purified from insect cells on CsCl gradients

were shown to exhibit classic herringbone morphology when viewed under the electron microscope. It is interesting that such structures were observed in both baculovirus-infected and measles-infected cells despite the fact that extensive cleavage of NP had occurred. In contrast to the studies on Vaccinia virus-infected animal cells, baculovirus-infected cells were found to contain a higher proportion of long nucleocapsids, suggesting that the length of the structures was not limited by the size of any associated RNA molecules. The buoyant density of the nucleocapsids, determined by CsCl gradient centrifugation, was lower (1.28 g/ml) than that reported for capsids from measles virus (1.31 g/ml). This finding is consistent with the hypothesis that the capsids from baculovirus-infected cells do not contain RNA. However, it is possible that a small amount of RNA is required to initiate capsid assembly.

Spehner *et al.* have studied the influence of P on the assembly and RNA-encapsidating properties of measles NP (Spehner, Kirn et al. 1991). Although recombinant measles NP expressed in the presence of P was shown by electron microscopy to form nucleocapsids, the process was much less efficient than in the absence of P. In addition, the buoyant density of nucleocapsids formed in the presence of P was slightly higher than the density of those formed when NP was expressed alone. The cause of this density increase is unclear. Analysis of the nucleic acid content of nucleocapsids showed that the RNA content did not differ significantly between the two types of capsid. It is more likely that the presence of P subtly modifies the conformation of the capsid resulting in a density change undetectable by electron microscopy.

Investigations of the effect of NP-P and P-L complexes on the replication of Sendai defective interfering virus *in vitro* was carried out using extracts of cells expressing the viral proteins (Horikami, Curran et al. 1992). Extracts that contained co-expressed NP and P were demonstrated to be effective in supporting replication of the defective virus genome whereas extracts in which NP and P were expressed separately, then mixed, were inactive. Co-expression of P and L was also required for activity. These findings suggest that the two complexes, NP-P and P-L are required for RNA replication and must form during or soon after the synthesis of the proteins. The P-L complex serves as the RNA polymerase and NP, donated from the NP-P complex, encapsidates the newly synthesised RNA. It appears that P acts as a chaperone for NP, maintaining it in a soluble state until nascent RNA genomes are available.

Buchholz *et al.* have analysed the domains of Sendai virus NP involved in nucleocapsid assembly using 29 deletion mutants of the protein (Buchholz, Spehner et al. 1993). Deletions between residues 1 and 82 and between 385 and 399 resulted in unstructured aggregates of NP. Deletions between residues 83 and 384 completely abolished all interactions. Only deletions in the C-terminal 124 amino acids had no effect on nucleocapsid assembly. Residues 1 to 399 therefore appear to contain all of the structural information necessary for assembly. This region of NP is the one best conserved among paramyxoviruses and is hydrophobic, with a high proportion of predicted β-sheet. In contrast, the C-terminal region is highly charged and much more variable. Analysis of further deletion mutants was carried out to indicate whether the C-terminal domain is involved in P protein binding (Buchholz, Retzler et al. 1994). Deletions between residues 400 and 439 did not effect NP-P complex

formation. A more likely region for P protein binding lies between amino acids 440 and 524 as deletions within these residues abolishes complex formation. This is supported by immunological studies in which a monoclonal antibody was reported to displace P from the nucleocapsid and also prevented exogenous P from binding. The epitope for the mAb was mapped between residues 405 and 524 of NP (Ryan, Portner et al. 1993). These data suggest that the NP can be divided into two domains that probably fold independently of each other and carry out completely different functions.

1.5 The Immune System

Immunity in vertebrates is mediated by lymphocytes, a type of white blood cell. The two major categories of immune response, humoral immunity and cellular immunity, involve different classes of lymphocyte. B cells, which develop in the adult bone marrow or foetal liver, produce antibodies and are involved in humoral immunity. T cells, produced in the thymus, are responsible for cell-mediated immunity and exist as either helper T cells or cytotoxic T cells. The helper T cells enhance the action of other lymphocytes by, for example, assisting antibody production in B cells. In contrast, cytotoxic T cells are directly involved in the destruction of infected cells.

T and B cells occur at different levels of maturation - virgin cells, memory cells and activated cells. Upon first contact with antigen, virgin cells may become activated. Alternatively, the cells can be stimulated to mature into memory cells, which respond much more readily to antigen than do virgin cells. In the case of memory B cells this is brought about by increased antigen affinity. Memory T cells increase the efficiency of extracellular signal transduction by enhanced adhesion to other cells.

In contrast to B cells, which bind intact antigen, T cells recognise fragments of foreign proteins that have been partially degraded in the target cell and then displayed on the cell surface. The target cell proteins responsible for carrying antigenic fragments to the cell surface are the major histocompatability complex (MHC) molecules. Class I MHC molecules are displayed on the surface of nearly all nucleated cells and are recognised by cytotoxic T cells. In contrast, Class II molecules

are usually present in specialised cells, such as B cells and are recognised by helper T cells.

T cells also express co-receptors that function in both stabilisation of the T cell-target cell interaction and activation of the T cell. Two such co-receptors are CD4 (expressed on helper T cells) and CD8 (expressed on cytotxic T cells). Both proteins have a transmembrane domain, an extracellular immunglobulin-like domain and a cytoplasmic tail. The latter region possesses tyrosine kinase activity and is involved in T cell activation.

1.5.1 Antibody diversity

The immune system requires an enormous range of antibodies that must be able to cope with almost any potential antigen. In mammals, each type of Ig chain – heavy chains and κ- and λ-light chains – is encoded within a separate pool of gene segments. Antibody diversity is generated by somatic recombination of these gene pools.

In general, the pools contain a large number of gene segments encoding the variable regions of an Ig chain and a smaller number that encode the constant regions. Site-specific genetic recombination mechanisms combine the V and C coding sequences, a process that also activates transcription. The constant domains of antibodies are encoded by a single C gene segment. In contrast, gene sequences encoding the variable regions are combinations of more than one gene segment. In the case of light chain V regions, two segments (a long V segment and a short J segment) are combined. H chain V regions contain both V and J segments and an additional D (diversity) gene segment.

The antibody diversity generated by the large numbers of V, J and D gene sequences is greatly increased by combinatorial joining of the segments. Furthermore, the joining events themselves are inherently imprecise. This 'junctional diversification' is brought about by the random loss or insertion of nucleotides at the sites of recombination.

The generation of memory cells from antigen (or Th cell) stimulated B cells is usually associated with a progressive affinity between the antibody and its antigen. This process, affinity maturation, is a result of a dramatic increase in the rate of somatic mutation. The small proportion of B cells that produce higher affinity antibodies are preferentially stimulated by antigen to proliferate providing progressively better protection against the antigen.

1.5.2 Monoclonal Antibody Production

The development of hybridoma techniques in 1975 revolutionised the use of antibodies as tools in cell biology and facilitated structural studies of specific antibodies (Kohler and Milstein 1975). The technique, which allows production of monoclonal antibodies in large amounts, combines the specific antibody secreting properties of a clone of B lymphocytes with the 'immortality' of B lymphocyte tumour cells. This is achieved by fusing the required immunoglobulin-producing B lymphocytes with tumour cells and selecting for immortalized hybrid cells that produce a single antibody.

1.5.3 Antibody Structure

Antibodies are Y-shaped molecules with two identical antigen binding sites. The molecule has a flexible hinge region that allows the distance between the two antigen combining sites to vary, thereby increasing the efficiency of antigen binding. Antibodies consist of four polypeptide chains, two light (L) chains and two heavy (H) chains (Fig 7).

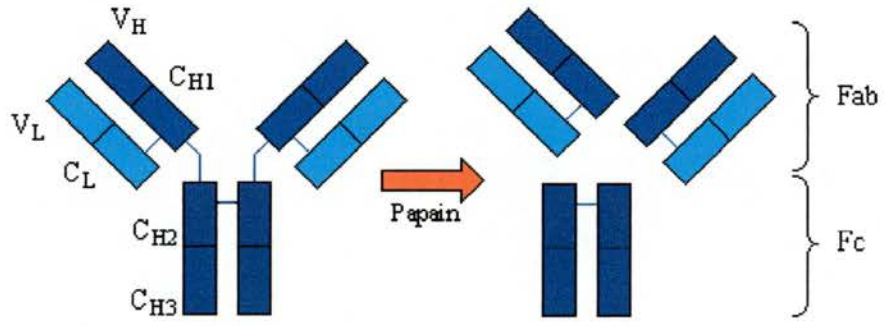


Fig. 7. Schematic diagram of an antibody illustrating the arrangement of immunoglobulin domains in an antibody (left) and the action of papain to produce Fab and Fc fragments (right).

Antibodies may be fragmented by the proteolytic enzymes papain and pepsin. Papain cleavage produces two identical antigen-binding (Fab) fragments and one readily crystallisable (Fc) fragment (Fig 6.). Pepsin cleavage produces one $F(ab')_2$ fragment which contains both antigen binding sites. The tail region of the molecule is broken down into smaller fragments.

In higher vertebrates there are 5 classes of antibody: IgA, IgD, IgE, IgG and IgM, each containing its own class of heavy chain: α, δ, ε, γ and μ respectively.

During the primary immune response, IgM is secreted into the blood stream as a pentamer containing another protein, called the J chain. Binding of IgM to the antigen activates the first component of the complement system, resulting in destruction of the invading organism.

IgG, the major class of antibody in the blood, is produced during secondary immune responses. Phagocytic cells possess receptors that recognise the Fc regions of IgG molecules and can ingest and destroy micro-organisms that have become coated in IgG antibodies produced in response to the infection.

IgE molecules tightly bind mast cells and basophils via their Fc regions. Subsequent binding of antigen to the IgE molecules stimulates the cells to release biologically active amines. These amines cause dilation and alter the permeability of blood vessels, and are a major factor in many allergic reactions.

In addition, higher vertebrates have two classes of light chain, κ and λ. An individual antibody always has identical heavy and identical light chains.

1.5.4 Fine Structure

The heavy and light chains of antibodies consist of repeating domains of about 110 amino acids, each containing one disulphide bond. Each domain (VL, CL, VH_1, CH_1, CH_2 and CH_3) forms an immunoglobulin fold consisting of two β-sheets lying anti-parallel to one another and joined by a disulphide bond.

The variable domains of heavy and light chains each contain three hypervariable loops, also known as complementarity-determining regions (CDRs), which are clustered together to form the antigen-binding site. The binding surfaces vary in size and shape depending on the antigen. Smaller antigens such as peptides

and haptens tend to bind to deeper pockets whereas antibodies that recognise larger molecules tend to have flatter antigen binding sites. The structure of the remainder of the variable domain is rigid and highly conserved. This structural organisation means that a great diversity of antibodies can be produced by variation of the residues in the hypervariable loops without affecting the overall structure of the antibody.

Chapter 2

Ebola Virus Nucleoprotein

2.1 Introduction

Prior to the start of the project, Ebola NP had been cloned into a baculovirus expression vector by A.R. Fooks at CAMR (Porton Down) for expression in insect cells. Expression trials at various MOI (multiplicities of infection) and TOI (times of infection) indicated that the protein yield was too low for crystallization experiments. A clone expressing high levels of hexahistidine-tagged Ebola NP (Gabon strain) was kindly provided by Dr M. Bouloy of Institut Pasteur, Paris.

At the time of starting the project, polymerisation of Ebola virus NP had not been observed in the absence of other viral proteins. However, as self-assembly of capsid proteins appears to be a common phenomenon among the Mononegavirales, the oligomeric state of Ebola NP needed to be investigated prior to crystallization attempts.

2.2 Recombinant Protein Expression

2.2.1 Glycerol stocks

BL21(DE3) cells transformed with plasmid pET-N.Ebola were streaked on agar and incubated overnight at 37°C. A single colony was picked and grown to an OD_{600} of 0.4 in 5 ml Luria broth (LB) containing 34 μg/ml chloramphenicol and 50 μg/ml ampicillin. 0.5 ml of culture was added to an equal volume 30% v/v glycerol and stored at –80 °C.

2.2.2 Expression of His-tagged Ebola NP

5 ml LB was inoculated with a scraping from a glycerol stock and shaken overnight at 37 °C. This culture was added to 500 ml LB media containing the

appropriate antibiotics, and induced with 0.4 mM isopropyl-β-D-thiogalactoside (IPTG) when the OD_{600} reached 0.6. After 3 hours the cells were spun down and stored at –20 °C.

2.3.1 Metal Chelate Affinity Chromatography

Cell pellets were defrosted on ice and resuspended in binding buffer (20 mM TrisHCl pH7.5, 0.5 M NaCl and 10 mM imidazole). Following sonication, cell debris was removed by centrifugation at 20 000g for 20 min. Cleared lysate was filtered through a 0.22 μm membrane prior to purification.

His-tagged proteins were purified on a POROS MC column (Perseptive Biosystems, Inc). The column was charged with 5 column volumes (CV) of 100 mM $NiSO_4$ and equilibrated with binding buffer. Protein was loaded onto the column and washed with 5 CV of 20mM TrisHCl buffer containing 500mM NaCl and 20mM imidazole. Bound protein was eluted by a 20mM to 800mM imidazole gradient. Peak-containing fractions were analysed by SDS PAGE.

2.4 Gel Electrophoresis

2.4.1 SDS PAGE

Denaturing electrophoresis was carried out on Bis-Tris-HCl buffered 4-12 % polyacrylamide NuPAGE gels (Invitrogen) according to the manufacturer's protocols. Briefly, samples were dissolved in NuPAGE LDS (lithium dodecyl sulfate) sample buffer and incubated at 70 °C for 10 min. Approximately 1 μg protein was loaded per well. After electrophoresis, the gel was stained in a solution containing 0.1 % Coomassie Brilliant Blue, 45 % methanol, 45 % glacial acetic acid.

2.4.2 Native PAGE

Proteins were run on 12% acrylamide gels in the absence of SDS or reducing agent to give an indication of solubility. The gels consisted of 7.5 ml of resolving gel (12 % acrylamide, 0.37 M TrisHCl pH8.8, 0.1 % SDS, 0.04 % ammonium persulphate) covered with 2.5 ml stacking gel (5 % acrylamide, 125 mM TrisHCl pH 6.8, 0.1 % SDS, 0.05 % ammonium persulphate). Samples were added to an equal volume of native loading buffer (0.125 M TrisHCl pH 6.8, 20 % glycerol, 0.004 % bromophenol blue). Electrophoresis was carried out in tank buffer (2.5 mM Tris, 0.2 M glycine) at a constant current of 25 mA.

2.5 Bradford Assay

Protein concentration was measured by the Bradford assay (Bradford 1976). 500 µl of protein solution, at a concentration between 10 and 100 µg/ml, was added to 500 µl Bradford reagent. After 5 min the optical density was measured at 595 nm and compared to a bovine serum albumin (BSA) standard curve.

2.6 Dynamic light scattering

250 µl protein solution at 1 mg/ml was filtered through a 0.1 µm Whatman filter. Light scattering measurements were performed using a DynaPro-801 Molecular Sizing Instrument (Protein Solutions Inc.).

2.7 Immunodetection

SDS gels were run as described above. Unstained gels were blotted onto nitrocellulose membranes using an Xcell II Blot Module (Invitrogen).

The membrane was washed twice for 10 minutes each with TBS buffer (10 mM TrisHCl, pH 7.5, 150 mM NaCl) then incubated for 1 hour at room temperature in blocking buffer (3% BSA in TBS). The membrane was washed twice in TBS-Tween/Triton buffer (20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 0.05 % Tween 20, 0.2 % Triton X-100) and once in TBS before incubating for 1 hour in anti-His antibody solution (1 in 1000 dilution in blocking buffer). After washing twice with TBS-Tween/Triton buffer and once with TBS, the membrane was incubated with secondary antibody solution for 1 hour. The secondary antibody (rabbit-anti-mouse/AP conjugate) was diluted 1 in 10,000 in blocking buffer. After four washes with TBS-Tween/Triton buffer, the membrane was stained with Western Blue (Promega), an alkaline phosphatase substrate.

2.8 N-terminal sequencing

Samples for N-terminal sequencing were run on SDS PAGE and blotted onto PVDF membrane. The bands, stained for 30 sec with amido black (0.1 % amido black, 40 % MeOH, 1% acetic acid), were excised and submitted to the University of St Andrews Protein Sequencing service. Sequencing was carried out on an Applied Biosystems Procise 491 using the Edman degradation method (Edman 1970).

2.9 Mass Spectrometry

Matrix-Assisted Laser Desorption/Ionisation Time-of-Flight (MALDI-TOF) Mass Spectrometry was carried out by the University of St Andrews MALDI-TOF MS service on a Tofspec 2E Micromass system. Briefly, the sample is mixed with a UV adsorbing matrix and ionised by a UV laser. A potential is applied to the sample

plate, accelerating the matrix and analyte ions. The time taken by the ions to reach the detector is converted to molecular weight.

2.10 Electron Microscopy

Electron microscopy was carried out by Dr P. A. Bullough at the University of Sheffield. NP samples were analysed at a concentration of 0.1 mg/ml in TrisHCl pH7.5. The protein was negatively stained with uranyl acetate.

2.11 Proteolytic digestion

2.11.1 Thrombin digest

125 µl of NP (0.14 mg/ml) was added to 1 unit of thrombin and incubated overnight at 37°C.

2.11.2 Trypsin digest

21µl of trypsin (140 µg/ml) was added to 500 µl of prewarmed NP (140 µg/ml). The solution was incubated at 37°C. 10 µl aliquots were removed at various time points and added to 1µl of 10mM phenymethylsulfonyl fluoride (PMSF) to quench the reaction. The samples were run on both SDS and native gels.

2.12 Construction of Clones

Restriction enzymes and DNA-modifying enzymes were obtained from Promega with the exception of AseI and NdeI which were purchased from New England Biolabs. Oligonucleotides were obtained from Oswel.

2.12.1 Plasmid DNA purification

Plasmid DNA was extracted from cells using QIAprep Miniprep or Maxiprep kits (Qiagen) following the manufacturer's instructions. For large numbers of samples, plasmid purification was achieved by alkaline lysis of cells followed by isopropanol precipitation: Cells from a 5 ml culture were spun down and resuspended in 0.3 ml resuspension buffer (50 mM TricHCl, pH 8; 10 mM EDTA; 100 μg/ml RnaseA). 0.3 ml of lysis buffer (200 mM NaOH, 1% SDS) was then added and cells were lysed at room temperature. After 4 min, the reaction was neutralized by the addition of 0.3 ml 3 M potassium acetate, pH 5.5. Samples were incubated on ice for 10 min and then centrifuged for 15 min at 13 000 g. The supernatant was removed to a fresh tube and respun. In another fresh tube, the second supernatant was added to 0.7 ml isopropanol. Precipitated DNA was pelleted by centrifugation and washed with 0.2 ml 70 % EtOH. The pellet was air dried for 5 min and resuspended in 10 μl H_2O.

DNA digestion was carried out to confirm the presence of insert DNA within the vector. A typical digestion reaction contained 7 μl DNA, 1 μl enzyme(s),1 μl Multicore buffer, and water up to 10 μl. Following incubation for 1 hr at 37 °C the DNA was analysed on agarose gel.

2.12.2 Agarose gel electrophoresis

For a 1% gel, 250 mg agarose was dissolved in 25 ml TAE buffer (40 mM Tris-acetate, 1 mM EDTA). When slightly cooled, 3 μl of 10 mg/ml ethidium bromide was added and the solution was poured into a mould. 10 μl DNA sample was

added to 2 µl of DNA sample buffer (0.05 % Bromophenol Blue, 40 % sucrose, 0.1 M EDTA, 0.5 % SDS) and loaded onto the gel. The gels were run at 80 V for 30 mins.

2.12.3 Polymerase chain reaction (PCR)

PCR amplification of five different sections of the NP gene was carried out using combinations of the oligonucleotides shown in Table 1. The primers were designed to incorporate 3` BamHI and 5` KpnI restriction sites into the ends of each amplified section.

The reaction components (1 µl 5` primer, 1 µl 3` primer, 1 µl plasmid DNA, 1 µl dNTPs, 5 µl 10x Thermopol buffer, 40 µl water, 1µl VENT polymerase) were mixed in a PCR tube and overlaid with 40 µl mineral oil. Amplification was achieved by carrying out 30 cycles of DNA denaturation (95 °C, 1.5 min), oligonucleotide annealing (60 °C, 1.5 min) and oligonucleotide elongation (72 °C, 2 min), followed by 7 min at 72 °C.

2.12.4 Digestion

Vector (pQE30, Qiagen) and insert DNA samples were digested separately with both KpnI and BamHI for 3 hr at 37 °C. The digested DNA was extracted from an agarose gel and purified using a QIAquick Gel Extraction Kit (Qiagen).

Residues 1-132	Forward	CAC AAA ATT GGA TCC ATG GAT TCT CGT
	Reverse	CCG GGA TGG TAC CAA GTT ATC TCT TAA TG
Residues 133-399	Forward	GGA AAA AAC ATT GGA TCC ACA CTT GCT G
	Reverse	CGT GAC CTT GGT ACC GCG TTA TTT CTT TAG
Residues 400-739	Forward	GCT ATG GTA ACT CTA GGA TCC GAG CGC CTG
	Reverse	ATC CCA TGG TAC CAT GCT CAT TCA CTG

Table 1. Primers used to amplify sections of Ebola NP.

2.12.5 Ligation

To ligate insert and plasmid DNA, 14 µl of insert was added to 1 µl vector. The DNA was heated at 45 °C for 5 min followed by 5 min on ice. 2 µl of T4 DNA ligase and 2 µl of ligase buffer were then added and the reaction mix was incubated at room temperature for 4-7 hours.

2.12.6 Preparation of competent cells

XL1-Blue cells, grown to an OD_{600} of 0.4, were pelleted and resuspended in 5ml 0.1M $MgCl_2$. The cells were repelleted and resuspended in 2ml 0.1M $CaCl_2$.

2.12.7 Transformation

20 µl of ligation reaction was added to 100 µl of competent cells. After 30 min incubation on ice, cells were heated at 42 °C for 90 sec. After a further 5 min on ice the cells were added to 100 µl of prewarmed LB media and shaken at 37 °C for 1 hr. 100 µl of cells were plated out on L-agar containing the appropriate antibiotic.

Following overnight incubation all colonies, up to maximum of 24, were picked for each construct and grown up in 5 ml volumes of LB. Plasmid purification of all the cultures was carried out. Identification of positive clones was achieved by incubation of 7 μl of each purified DNA sample with BamHI and KpnI followed by analysis on agarose gel.

2.13 Results

2.13.1 Characterisation of Ebola NP

Purified NP was obtained by metal chelate affinity chromatography. Dynamic light scattering (DLS) was carried out to investigate the oligomeric/aggregation state of the protein. DLS measurements gave a molecular weight estimate in the region of 17,000 kDa, clearly indicating either polymerisation or non-specific aggregation of NP. A test experiment with a BSA standard gave a molecular weight of 70 kDa, close to the expected value of 66 kDa.

As soluble protein was required for crystallization, various additives were tested for their ability to reduce aggregation. Addition of 0.5 M NaCl to the protein solution prior to DLS slightly reduced the observed molecular weight to 10,000 kDa. Addition of the reducing agent Tris (2-carboxyethyl) phosphine (TCEP) at concentrations of 0.1, 1 and 10 mM had a similar effect on the aggregation of NP. TCEP is a more stable, odourless alternative to DDT.

Detergents (Triton-X100, Tween20, n-Octyl-β-Glucopyranoside) were added at concentrations corresponding to their critical micellar concentrations (CMC). The lowest molecular weight was obtained with TritonX, however this value was still in the region of 6000 kDa, indicating that large aggregates were still present.

To investigate whether NP was self-assembling into nucleocapsid-like structures the aggregates were analysed by EM. The images revealed ring-shaped structures 40-50 nm in diameter (Fig. 1). No nucleocapsid-like structures were observed.

2.13.2 Proteolytic digestion

As attempts to solubilise the full-length protein failed, NP was subjected to proteolytic digestion so that a soluble portion of the protein might be obtained. In addition to intact NP (83.3 kDa), SDS PAGE analysis of the thrombin digestion products showed a band at 81 kDa and a faint band at 31 kDa. The likely cleavage site is after arginine 260. Cleavage here would result in two fragments of 29.4 kDa and 53.9 kDa. The latter would be highly negatively charged and would be expected to migrate more slowly on an SDS gel. The intact NP, which has an overall negative charge of -30, behaves in a similar manner. Unfortunately, neither peptide could be seen on a native gel, suggesting that they were not soluble.

Digestion with trypsin resulted in a 30 kDa peptide which could be observed on both SDS and native gels (Fig.2). The peptide was blotted onto PVDF membrane and the band was excised. N-terminal sequencing indicated that the N-terminus of the peptide corresponded to residue 133 of the full length NP. Initial MALDI-TOF Mass spectra of the digested NP did not give a clear indication of the exact molecular weight of the soluble peptide. Several peaks were present and it is likely that the peptide of interest was present in too low a concentration relative to other peptides.

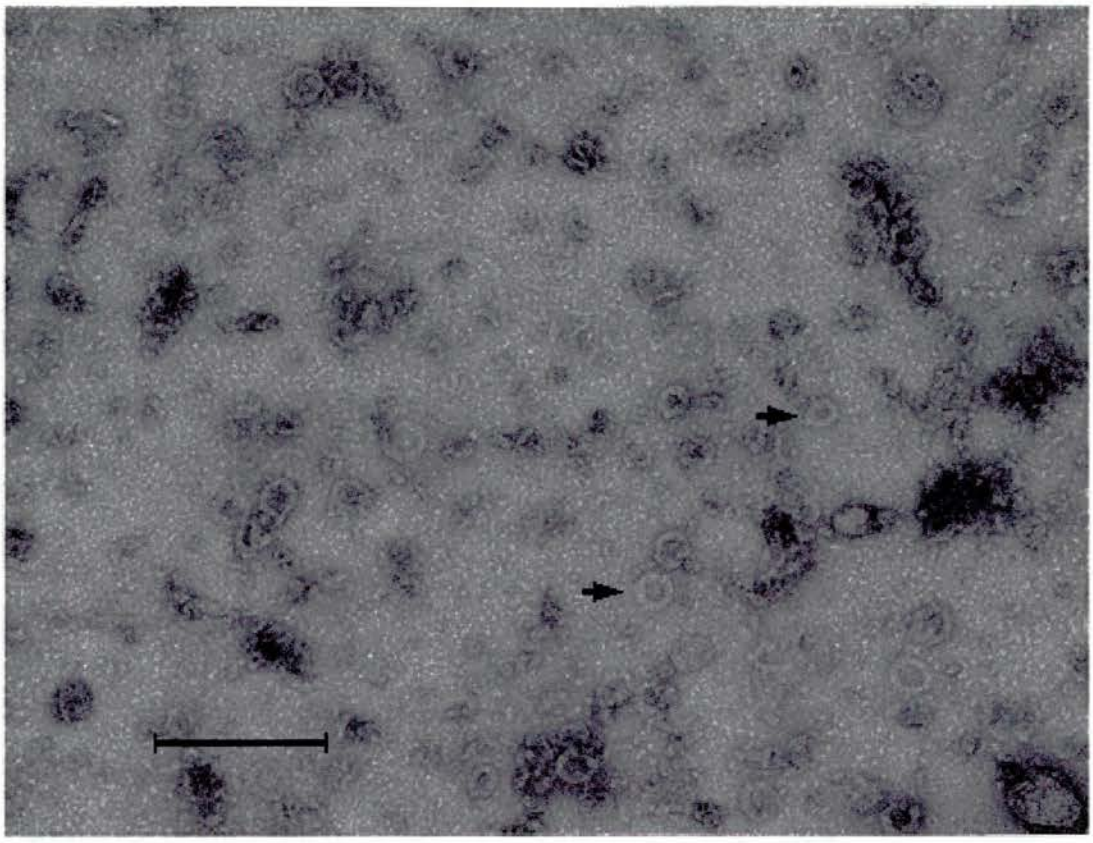


Fig 1. Scanning electron micrographs of purified NP. Arrows indicate ring structures. Bar: 200 nm

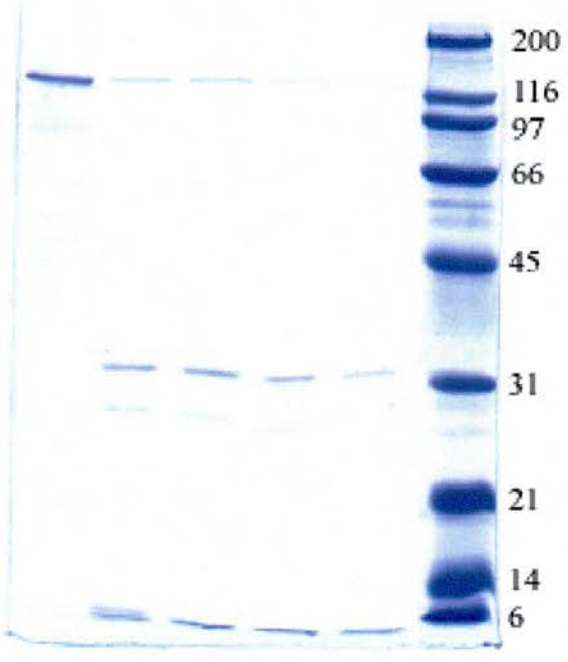


Fig. 2. SDS PAGE : trypsin digestion time course. 0 min, 2 min, 5 min, 10 min, 30 min, MW Markers.

The trypsin digested NP was loaded onto an anion exchange column to separate contaminating protein. SDS PAGE analysis showed that the 30 kDa peptide had been partially purified. After concentrating the peptide-containing fraction, mass spectrometry analysis was repeated (Fig. 3).

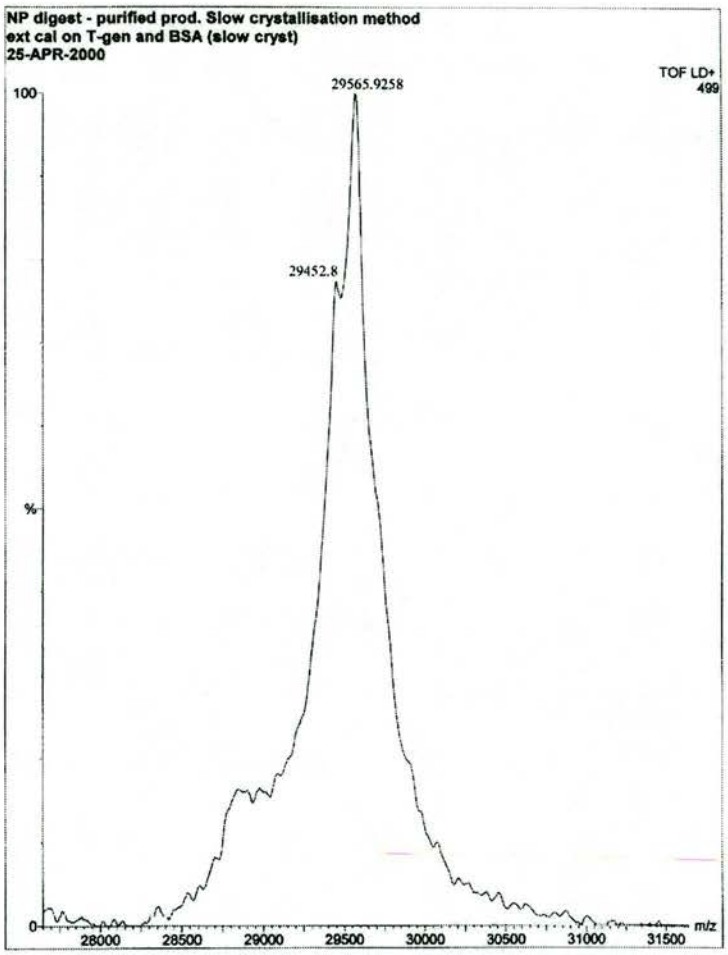


Fig. 3. MALDI-TOF MS of the partially purified 30 kDa trypsin digestion product.

A large peak at 29565.9 Da, with a small shoulder spike at 29452.8 Da, was observed. These values correspond to residues 133-399 (calculated weight = 29597.5) and 133-398 (calculated weight = 29469.4). Residues 398 (arginine) and 399 (lysine) are potential trypsin cleavage sites.

2.13.3 Construction of clones

Based on the above results, five segments of the nucleoprotein, including the central soluble portion were cloned for expression as N-terminal His-tagged proteins. These segments are represented in figure 4.

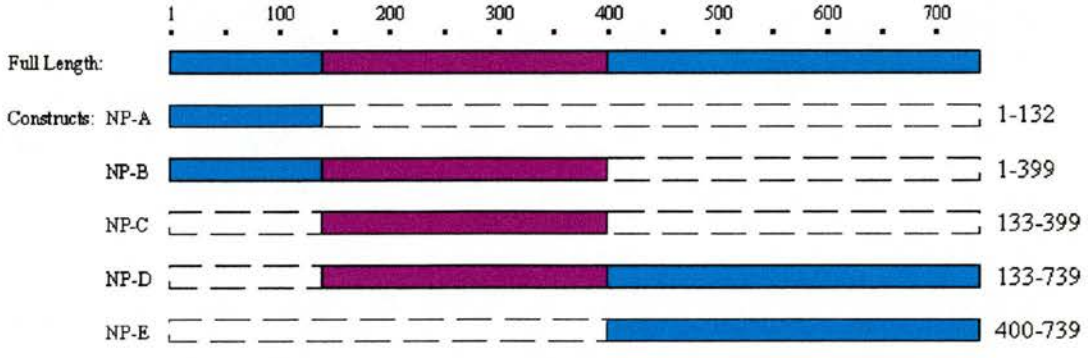


Fig 4. Ebola NP constructs A-E. The soluble segment obtained by trypsin digest is shown in magenta. Molecular weights: A (15.0 kDa), B (44.7 kDa), C (29.6 kDa), D (68.2 kDa), E (38.6 kDa).

Each PCR reaction amplified DNA of the correct size as shown in Figure 5. Each insert was successfully ligated into the vector pQE30 and incorporated into XL1-Blue cells. This was confirmed by BamHI/KpnI digestion of plasmid DNA purified from several XL1-Blue colonies.

Purified plasmid DNA from the positive colonies was used to transform the *E. coli* expression host M15. Constructs C, D and E were not expressed, even at a range of growth temperatures and IPTG levels. Constructs A and B were overexpressed as insoluble protein. No soluble recombinant protein could be obtained by altering growth and induction conditions. A second host, SG13009, was tested with each of the constructs. Unfortunately, peptides NP-A, NP-D and NP-E were not expressed and peptides NP-B and NP-C were present in insoluble inclusion bodies only.

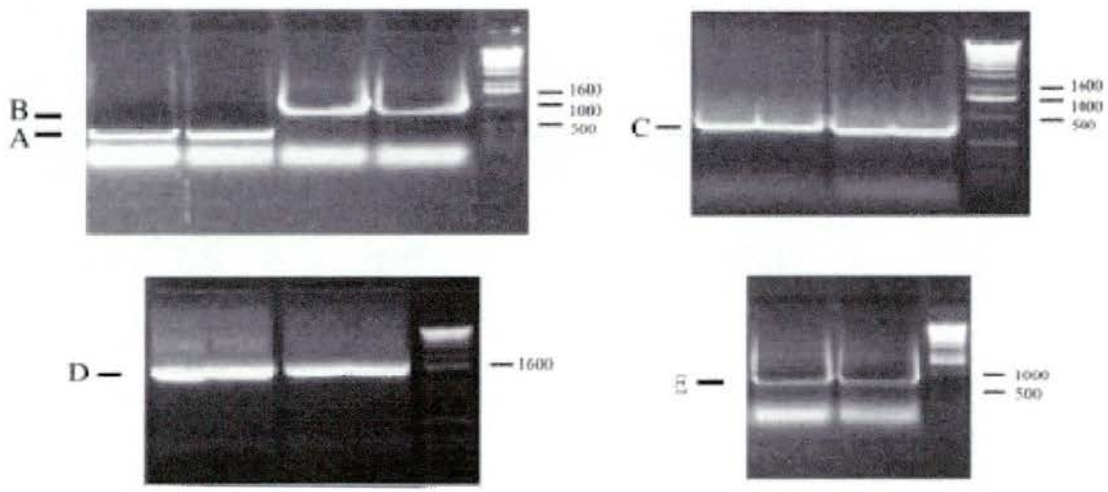


Fig 5. Agarose gels of PCR products. NP-A (396 bp), NP-B (1197 bp), NP-C (798 bp), NP-D (1818 bp), NP-E (340 bp).

Due to the difficulties encountered with this expression system, two of the constructs which had expressed (NP-B and NP-C) were subcloned into a pRSET vector for expression in (a) BL21*trxB*(DE3) and (b) BL21(DE3)pLysS cells. The former expression host possesses a thioredoxin reductase mutation (trxB) and may improve solubility and folding by facilitating disulphide bond formation. pLysS strains produce small amounts of T7 lysozyme, an inhibitor of T7 RNA polymerase. The T7 lysozyme provides additional stability to target genes by suppressing the basal

level of T7 RNA polymerase prior to induction. However, neither recombinant protein was expressed in these cells.

A small amount of soluble protein was eventually obtained from XL1-Blue cells transformed with pQE30-NP-B (Fig 6a, lane 5) and its identity was confirmed by immunodetection with anti-His antibody. The levels of soluble protein were improved by induction with 0.4 mM IPTG and overnight expression at 25 °C (Fig 6b). NP-B was purified by affinity chromatography on a Ni^{2+} column (Fig 6c).

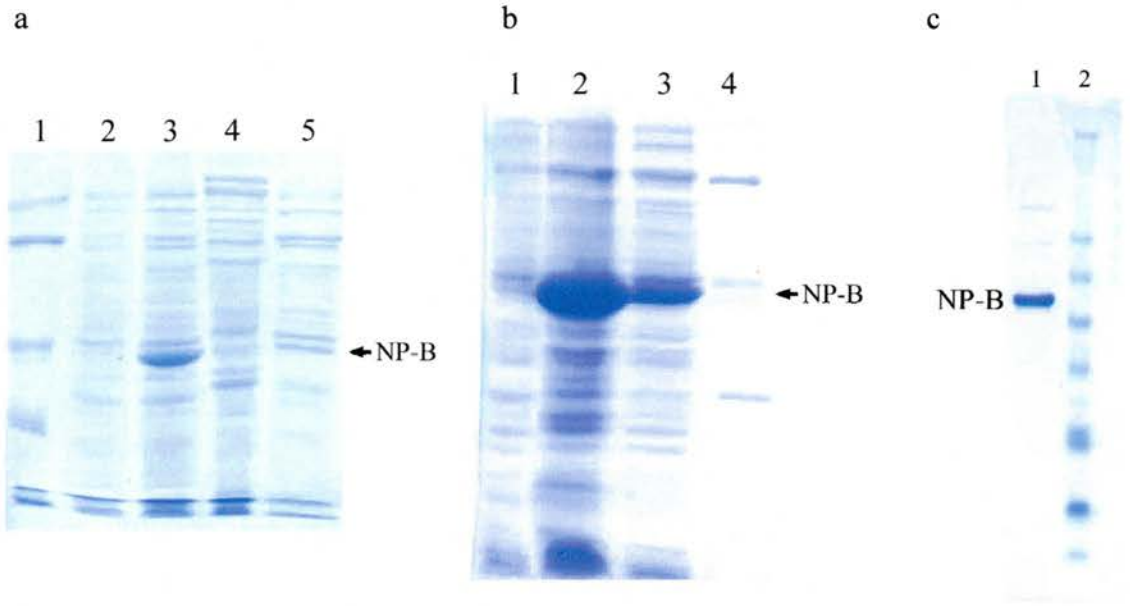


Fig 6. Expression of NP-B in XL1-Blue cells. (a) Initial expression, 1 mM IPTG. Lane 1, markers (see fig. 2); lane 2, non-induced control; lane 3, whole cell extract 1 hr post induction; lane 4, whole cell extract 2 hr post-induction; lane 5, soluble extract. (b) Optimised expression, overnight induction with 0.4 mM IPTG at 25°C. Lanel, non-induced control; lane 2, whole cell extract; lane 3, soluble extract; lane 4, markers. (c) Lane 1, Metal chelate affinity purified NP-B; lane 2, pre-stained markers (188 kDa, 62 kDa, 49 kDa, 38 kDa, 28 kDa, 18 kDa, 14 kDa, 6 kDa, 3 kDa).

DLS measurements gave an estimated molecular weight value in the region of 20,000 kDa, indicating that the protein was not monomeric. As in the case of the full length protein, NP-B was not solubilized by the addition of salt or detergents, or by raising or lowering the pH.

2.14 Summary

Although full length NP could be purified almost to homogeneity, crystallization trials were not attempted. The crystallizability of proteins is strongly correlated to the monodispersity of the solution (Ferre-D'Amare and Burley 1994). However, DLS showed that NP was polydisperse and unlikely to produce crystals. In addition, molecular weight estimates from DLS were 200-fold higher than that expected for monomeric NP, suggesting that the protein was severely aggregated. To determine whether NP was polymerising into the coiled rods seen in Ebola virus-infected cells or aggregating non-specifically, a sample of protein was analysed under by EM. Surprisingly, the protein appeared to be forming numerous ring-shaped structures. Although such structures have not previously been reported for a filoviral NP, recombinant Rabies nucleoprotein has been shown to form similar rings, in addition to longer, nucleocapsid-like structures (Iseni, Barge et al. 1998). The rings, which are similar in diameter to the full length nucleocapsid, may be intermediate structures in the nucleocapsid assembly process.

During the project, EM images of HeLa cells expressing recombinant MBGV NP were published by Kolesnikova *et al.* (Kolesnikova, Muhlberger et al. 2000). The NP, immunogold-labelled using a monoclonal antibody against MBGV NP, formed tubule-like structures (TLS), appearing in cross section as rings similar in size to those formed by the Ebola protein. It is possible that recombinant Ebola NP forms both rings and longer, nucleocapsid-like structures but that the latter were lost during purification.

For the purpose of finding a soluble segment for crystallization, NP was subjected to treatment with thrombin and trypsin. 1 hour trypsin digestion produced a soluble 29 kDa peptide. N-terminal sequencing and mass spectrometry confirmed that this peptide corresponded to amino acids 133-399.

Of the five constructs, only NP-B could be expressed as a soluble protein. This peptide corresponds to the first 399 amino acids of NP. Initial crystallization trials using an NP-B concentration of 4 mg/ml produced heavy precipitation in over 80 % of wells. DLS indicated that, like the full length protein, NP-B was not monomeric.

The peptide NP-B corresponds to a region of the full length NP that is relatively conserved among filoviruses. This region shows a high degree of homology to the N-terminal 400 amino acids of MBGV NP (54% identity, compared to 20% identity for the remainder of the protein). A small section of this N-terminal domain running from residues 150 to 162 also contains significant homology to nucleoproteins of paramyxoviruses (Barr, Chambers et al. 1991). In the nucleoproteins of Respiratory syncytial virus (RS), Pneumonia virus of mice (PVM), Sendai virus (SEN), Measles virus (MV), Rabies virus (RAB), Vesicular Stomatitis virus (VSV), MBGV and Ebola virus, this short sequence is composed of hydrophobic regions interrupted by either a lysine or an arginine residue.

Inspection of the predicted secondary structure (Rost 1996) of NP reveals that the vast majority of predicted helix and sheet lies within the N-terminal 423 amino acids (Fig. 7). This N-terminal domain is hydrophobic and slightly positively charged, consistent with putative roles in RNA binding and NP-NP interactions. Conversely, the remainder of the protein is hydrophilic (Fig. 8) with a strong net negative charge

of -41. The lack of predicted secondary structure in this region may indicate that folding of the C-terminal domain requires interactions with other proteins such as VP40.

Fig. 7. Ebola NP sequence highlighting predicted secondary structure. Strand (cyan), Helix (blue).

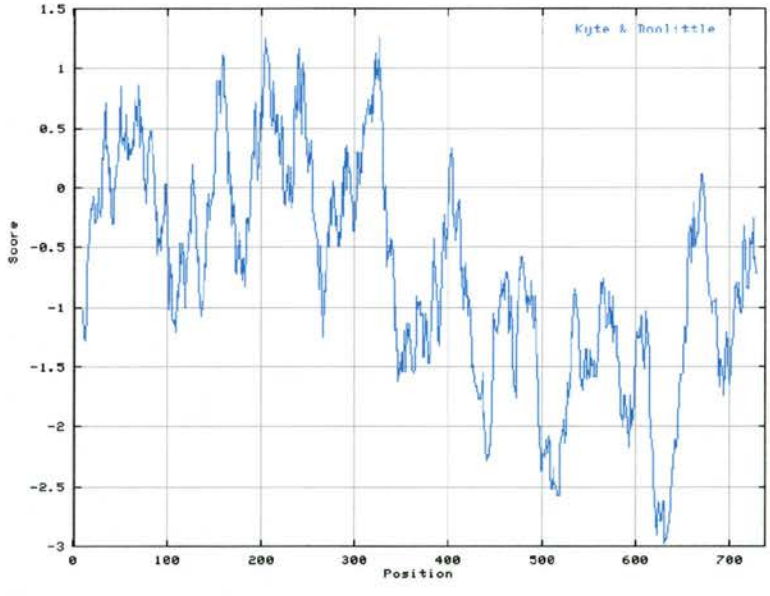


Fig. 8. Hydropathy plot of Ebola NP using the method of Kyte & Doolittle (Kyte and Doolittle 1982).

These observations suggest that Ebola NP has an N-terminal domain with regions of conserved structure that are involved in functions common to the NNS

RNA virus N proteins. These functions appear to include the protein-protein interactions required for nucleocapsid assembly. Buchholz et al (Buchholz, Spehner et al. 1993) tested 29 mutants of Sendai virus NP for their ability to self-assemble and showed that deletions within the N-terminal 399 amino acids abolished nucleocapsid formation. C-terminal deletions did not affect assembly. Similarly, a truncated form of SV5 NP in which the C-terminal 17 kDa had been cleaved was still able to form nucleocapsid-like structures (Mountcastle, Compans et al. 1974). Although it could not be confirmed by EM, it is likely that NP-B also retains the ability to self-assemble, making it a poor candidate for crystallisation.

Chapter 3

Crystallisation and Structure Determination of SV5-Pk Fab

3.1 SV5-Pk Antibody

The monoclonal antibody SV5-Pk was obtained by Randall *et. al.* from mice immunised with Simian Virus 5 (SV5) (Randall, Young et al. 1987). The antibody has been shown to be specific for an epitope (GKPIPNPLLGLDST) in the P and V proteins of SV5 (Southern, Young et al. 1991). A 14 amino acid peptide (V5) derived from this epitope retains its antigenicity when fused to the C-terminus of larger proteins and binds SV5-Pk very strongly. The picomolar affinity (K_d = 23.6 pmol) of the SV5-Pk/peptide system has been exploited in a wide variety of applications including immunodetection, purification and construction of solid matrix antibody-antigen (SMAA) complexes (Randall and Young 1991). In the latter technique, the mAb SV5-Pk is linked to a solid matrix to form a complex which can then be coupled to a recombinant protein tagged with the V5 peptide for vaccination purposes.

The aim of this part of the project was to obtain the three-dimensional structure of the Fab portion of SV5-Pk.

3.2 Fab Fragment Production and Purification

3.2.1 SV5-Pk antibody purification

Monoclonal SV5-Pk antibody was kindly provided by R. Randall of the University of St Andrews. The protein had previously been isolated from hydridoma cells and pelleted with $(NH_4)_2SO_4$. The pellet was resuspended in binding buffer (1M glycine pH 8.6, 150 mM NaCl) and loaded onto a Protein A affinity column. The column was washed with binding buffer until the flow through no longer absorbed at 280 nm. The antibody was eluted with 0.1 M citrate, pH 3.0. UV-absorbing fractions

were analysed on SDS PAGE (Fig. 1) then pooled and dialysed overnight into phosphate buffered saline (PBS).

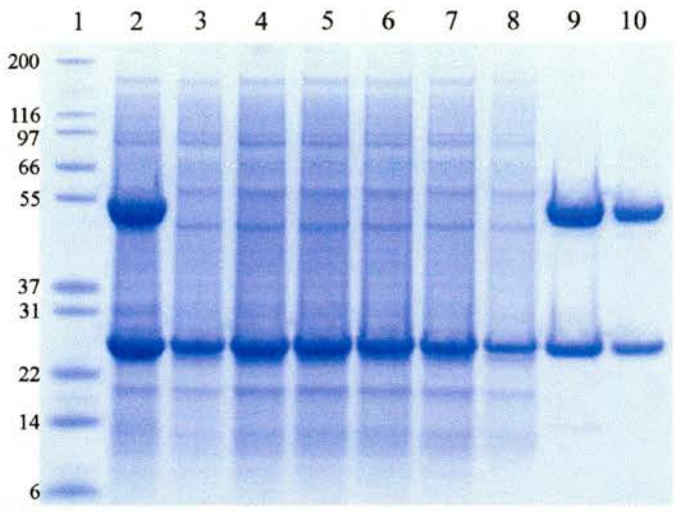


Fig 1. SDS PAGE: Antibody purification. Lane 1, Pre-stained markers; lane 2, crude antibody; lanes 3-8, flow through fractions; lanes 9 and 10, elution.

3.2.2 SV5-Pk Fab fragment production

Fab and Fc fragments of the antibody were obtained by papain cleavage. Antibody was digested with varying papain concentrations and incubation times to ascertain optimum conditions: 1 mg/ml SV5-Pk was incubated with 20 mM EDTA, 100 mM L-cysteine and 10 μg/ml papain for 2 hours at 37 °C. The digestion was quenched by adding 500 μl of 200 mM iodoaminoacetate (IAA) to 500 μl reaction, and incubating at room temperature for 30 min.

The digested antibody was dialysed into binding buffer and the fragments were purified on a Protein A column (Fig. 2). The flow through containing the Fab fragments was collected and dialysed into antibody buffer (50 mM NaCl, 20 mM Tris pH 7.6, 0.1 mM KI). The Fab fragments (Fig. 2, lanes 3-9) appear as two bands of

close molecular weight corresponding to the light chain (23.6 kDa) and the Fab portion of the heavy chain (23.1 kDa). The dialysed Fab was then concentrated to 10 mg/ml in a Centricon 10 kDa molecular weight cut-off concentrator (Amicon).

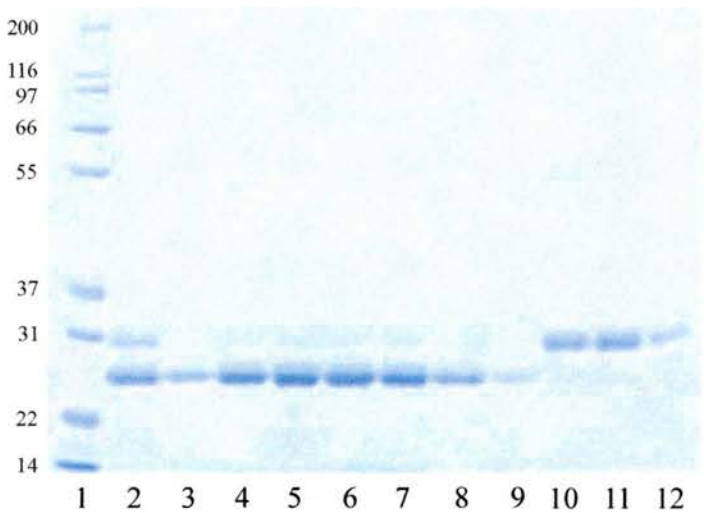


Fig 2. Fab fragment purification. SDS PAGE: Lane 1, MW markers; lane 2, pre-column; lanes 3-9, flow through (Fab fragment); lanes 10-12, elution (Fc fragment).

3.3 Crystallisation of SV5-Pk Fab

Initial crystallisation conditions were determined using the hanging-drop vapour diffusion method. Trials were set up on LINBRO crystallization plates at room temperature using Hampton Crystal Screen Kits (Hampton Research). 1 µl each of 10 mg/ml Fab and reservoir solution were mixed on siliconized glass coverslips and allowed to equilibrate over wells containing 0.5 ml crystallant.

After 24 hours equilibration, plate-like crystals had appeared in four of the 50 Hampton Screen I conditions, the largest of which grew in 0.2 M Calcium Acetate, 0.1

M Sodium Cacodylate pH 6.5, 18% PEG 8K (Fig. 3). These crystals were used for in-house data collection.

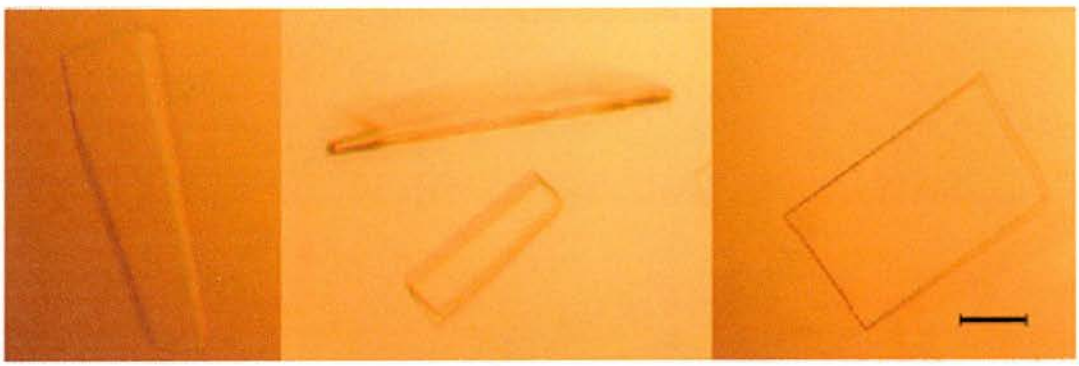


Fig 3. Crystals of SV5-Pk Fab grown in 0.2 M Calcium Acetate, 0.1 M Sodium Cacodylate pH 6.5, 18% PEG 8K. Bar: 50 μm

3.4 Data Collection

The crystals were transferred to a drop containing reservoir solution plus 20% glycerol prior to flash freezing in a nitrogen stream. X-ray diffraction data were collected at 100K from $Cu_{K\alpha}$ radiation generated by a Rigaku rotating anode generator using an R-axis IV imaging plate. 180 frames of data were collected in 1° oscillations at a crystal-to-detector distance of 180 mm. Diffraction data sets were processed and scaled using DENZO and SCALEPACK software (Otwinowski and Minor 1997). Data reduction was carried out using programs from the CCP4 suite (Collaborative Computational Project Number 4 1994). Table 1 summarises the data collection statistics.

	All Data	Highest resolution Shell (2.38-2.30 Å)
Crystal Parameters		
Space Group	C2	-
Unit cell dimensions	a=69.0 b=53.7 c=127.9 (Å)	-
	$\alpha=\gamma=90.0$ $\beta=98.3$ (°)	-
Resolution Limit (Å)	2.3	-
Mosaicity	0.48	-
Data Processing		
Total Nº of observations	126018	-
Nº of unique reflections	20029	1973
R_{merge} [a] (%)	5.5	14.1
$I/(\sigma I)$	20.5	8.8
Completeness (%)	97.0	94.2

Table 1. X-ray diffraction data for crystals of SV5Pk Fab fragment. [a] $R_{merge} = \Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle$, where $\langle I\rangle$ represents the average intensity of symmetry-equivalent reflections.

3.5 Structure Determination

The Fab crystal structure was solved by the Molecular Replacement method using CNS(Brunger, Adams et al. 1998). A high resolution Fab structure (Brookhaven Protein Data Bank entry 1SBS: mouse anti-Hcg Fab) with high variable domain homology to SV5-Pk Fab was selected as the phasing model. The sequence identities of the H and L variable chains of the model and the SV5-Pk Fab fragments are 67 %

and 68 % respectively. The solution was found by searching separately for the constant domain (C_HC_L) and the variable domain (V_HV_L) to allow for any variability in the elbow angles. The solution for the V_HV_L domain was found first. The constant domain was then subjected to a cross rotation search. The rotated, translated variable domain coordinates and the rotated constant domain coordinates were combined in a PDB file. These coordinates were subjected to a further translation search in which the variable domain remained fixed. Sensible packing of the molecule in the crystal was confirmed visually using O (Jones, Zou et al. 1991).

The solvent content was determined using the equation below (Matthews 1968). The calculated solvent content is 51% with one molecule in the asymmetric unit.

$$1 - {}^{1.23}/_{Vm} \quad \text{where} \quad Vm = \text{Volume of asymmetric unit } (\text{Å}^3)/_{Mr\ (Da)}$$

3.6 Refinement

Rigid body refinement of the correct solution gave an R-factor of 42.8 % at 4 Å resolution. Rounds of model building and refinement were carried out using a combination of 2Fo-Fc, Fo-Fc and omit maps in O followed by refinement using CNS. The procedures included positional refinement, simulated annealing and individual isotropic B-factor refinement. 211 water molecules were added automatically by CNS in peaks of electron density $\geq 3\sigma$ in the Fo-Fc difference map and checked manually for suitable hydrogen bonding geometry. Variable domain residues that differed from those of the model were mutated to the correct amino

acid. The CDRs of the model were omitted from 2Fo-Fc map calculations. Constant domain residues, for which the exact sequence was unknown, were mutated to residues that fitted the electron density. The sequence of the constant domain was later confirmed by a 1.9 Å resolution structure (see Chapter 4) and corresponded to the λ-subclass.

The structure was refined at 2.3 Å resolution to a final working R-factor of 21.2 % and an R_{free} (calculated with 10% of data omitted from refinement) of 25.7 % (Fig. 4).

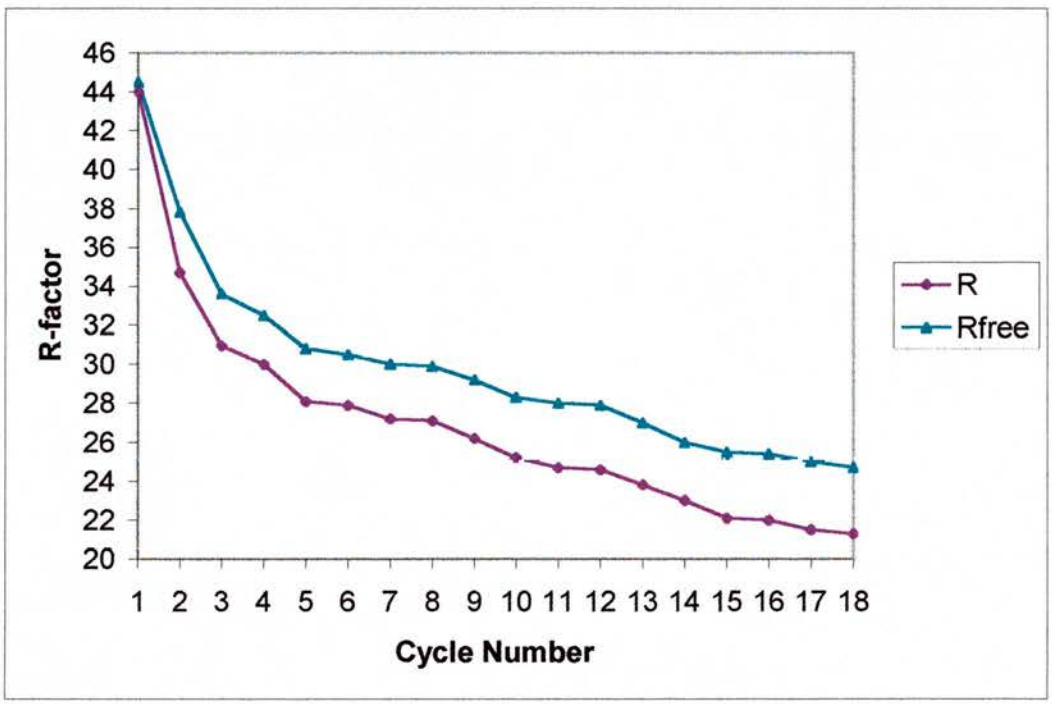


Fig. 4. R-factor and R-free values during refinement.

3.7 Structure Description

3.7.1 Overall Structure

The structure of SV5-Pk Fab fragment, shown in Fig. 5, is that of a typical Fab. The four immunoglobulin domains each contain a β-barrel formed by two β-

sheets connected by an internal disulphide bridge. The V_HV_L and C_HC_L domains each comprise two immunoglobulin folds that are related by a pseudo two-fold axis of rotation. The angle between the V_HV_L pseudo-dyad and the C_HC_L pseudo-dyad (the elbow angle) is 148°. To calculate the angle, the O program LSQ was used to find the transformations that map V_H onto V_L and C_H onto C_L. These matrices were converted to vectors using NCS STRICT in X-PLOR (Brunger 1992). The dot product of the vectors is equal to the inverse cosine of the elbow angle.

Solvent accessible surface area was calculated using NACCESS with a probe radius of 1.4 Å. The program uses the method of Lee & Richards (Lee and Richards 1971). The surface areas of the constant and variable domains are 10438 $Å^2$ and 10405 $Å^2$ respectively and the interface between them is 1477 $Å^2$. The buried surface area between the V_H and V_L domains is 1588 $Å^2$.

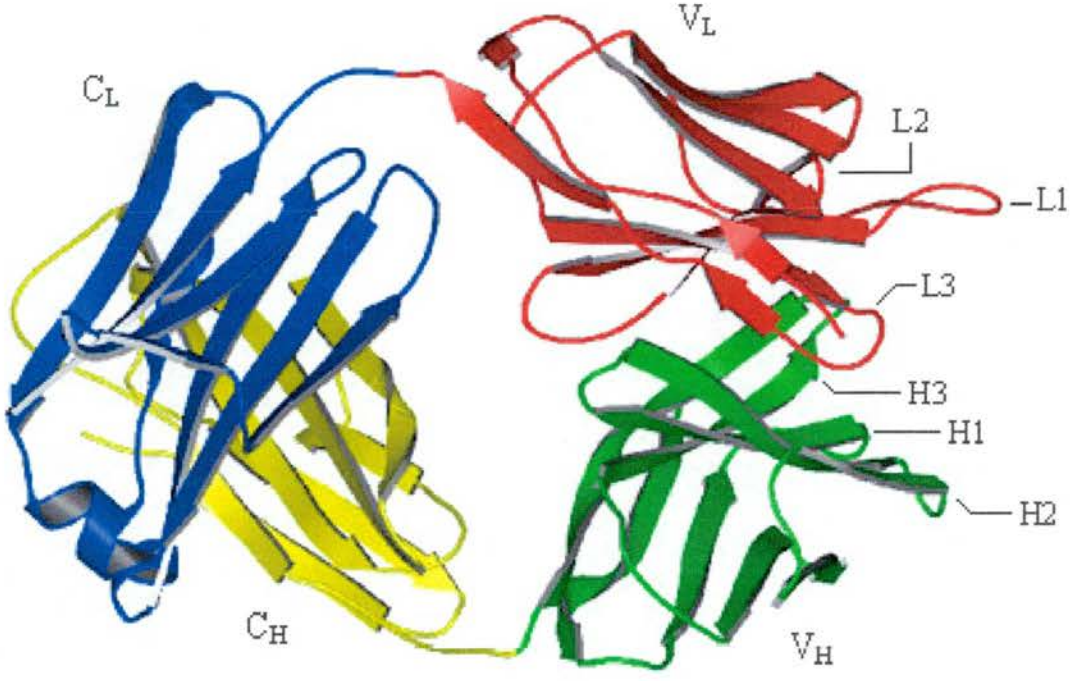


Fig 5. Ribbon representation of the Fab structure showing the positions of the CDRs. The figure was produced using Molscript (Kraulis 1991).

3.7.2 Complementarity Determining Regions (CDRs)

The structures of five of the six CDRs of antibodies appear to conform to a small number of canonical forms (Chothia and Lesk 1987). The canonical class of CDRs L1, L2, L3, H1 and H2 depends on both the length of the loops and the presence of structurally determining residues (SDRs). The CDR loops of SV5-Pk Fab were each assigned a canonical class using the automated method of Martin *et.al.* (Martin and Thornton 1996) and are summarised in Table 2. CDRs L2 and H1 match class 1. L1 is similar to class 4 but has a disallowed residue at L2. However, superimposition of SV5-Pk CDR L1 onto a representative class 4 CDR (1tet) yields an rmsd of 0.45 Å, suggesting that the loop does indeed belong to this class. Similarly, L3 and H2 can be superimposed onto representatives of classes 1 and 3 respectively. The superpositions are shown in figure 6.

CDR	Class	Mismatches
L1	Similar to class 4	L2 = I (allows V), L51 = M (allows V), L27 = K (allows Q), L92 = L (allows TS), L34 = Y (allows HEN), L93 = E (allows H),
L2	Class 1	-
L3	Similar to class 1	L91 = I (allows NFGSRDHTYV) L92 = L (allows NYWTSRQHAD)
H1	Class 1	-
H2	Similar to class 3	H52 = N (allows SFWH)

Table 2. Canonical classes of CDRs L1, L2, L3, H1 and H2 assigned by the method of Martin *et. al.* [Martin, 1996 #101]. The third column lists CDR residues that differ from the canonical definitions.

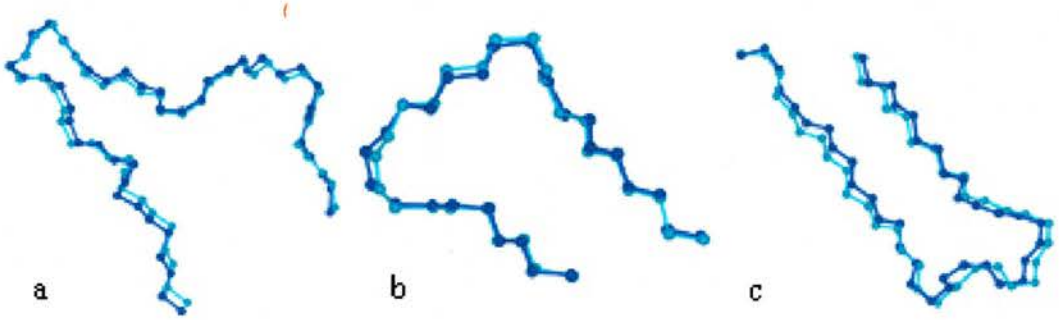


Fig 6. The backbone atoms of SV5-Pk CDRs (cyan) that could not confidently be assigned a canonical class were superimposed onto loops of known class (blue).
(a) CDR: L1, Class = 4, PDB code of representative Fab = 1rmf, rmsd = 0.45 Å;
(b) CDR: L3, Class = 1, PDB code = 1tet, rmsd = 0.3 Å;
(c) CDR: H2, Class = 3, PDB code = 1igc, rmsd = 0.32 Å.

The third hypervariable region of V_H varies greatly among antibodies and is therefore more difficult to divide into structural classes. However, analysis of 77 different H3 sequences indicates that some H3 loop conformations can be assigned canonical forms (Oliva, Bates et al. 1998). The loop is divided into three sections: the framework, take-off and apex regions. The framework is usually conserved in structure and the apex region forms a β-hairpin. The take-off region adopts either an extended or a kinked form. The kinked conformation is characterized by a conserved Trp103H to carbonyl hydrogen bond. In the extended form the hydrogen bond occurs between Trp103H and the side chain of a conserved Asp (101H). As shown in Fig. 7, the take-off region of SV5-Pk Fab H3 adopts a kinked conformation.

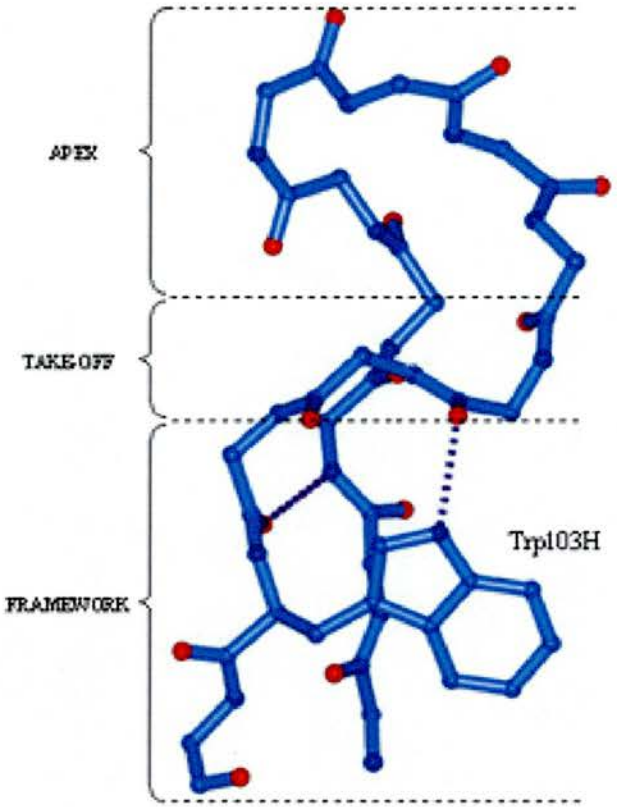


Fig.7 Kinked conformation of the SV5-Pk Fab CDR H3 showing the hydrogen bond formed between Trp103H and a main chain carbonyl.

Table 3 shows Kabat and Contact definitions of the CDRs. The Kabat definition (Wu and Kabat 1970) is based on sequence variability whereas the contact definition is based on analysis of antibody-antigen complex structures (Martin and Thornton 1996). The positions of the CDRs are shown in Fig 5.

CDR	Residues																		
L1	Residue:	24	25	26	27	27a	27b	27c	27d	27e	28	29	30	31	32	33	34	35	36
	Kabat:	R	S	S	K	S	L	V	H	S	N	G	I	T	Y	L	Y		
	Contact:												I	T	Y	L	Y		
L2	Residue:	46	47	48	49	50	51	52	53	54	55	56							
	Kabat:					Q	M	S	S	L	A	S							
	Contact:	L	L	I	Y	Q	M	S	S	L	A								
L3	Residue:	89	90	91	92	93	94	95	96	97									
	Kabat:	G	Q	I	L	E	L	P	F	T									
	Contact:	G	Q	I	L	E	L	P											
H1	Residue:	30	31	32	33	34	35												
	Kabat:		S	F	G	M	H												
	Contact:	S	S	F	G	M	H												
H2	Residue:	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65
	Kabat:				Y	I	N	T	D	S	T	T	I	Y	Y	G	T	V	K	G
	Contact:	W	V	A	Y	I	N	T	D	S	T	T	I	Y						
H3	Residue:	93	94	95	96	97	98	99	100	100	101	102							
	Kabat:			A	G	P	Y	Y	G	F	D	Y							
	Contact:	A	S	A	G	P	Y	Y	G	F	D								

Table 3. Amino acid composition of SV5-Pk Fab CDRs determined using the Kabat and Contact definitions. Residues are numbered according to the Kabat system.

3.8 Quality of the Structure

The quality of the model was verified by PROCHECK (Laskowski, MacArthur et al. 1993), the outputs of which are shown in Fig 10. 87.6 % of residues are in favoured regions of the Ramachandran Plot with a further 11.6 % in additional allowed regions. Three residues lie in disallowed regions (L51, H138, H139). L51, clearly defined in the electron density, is in the (i+1) position of a γ-turn and corresponding residues in other Fab structures are frequently observed to possess unusual geometry (Al-Lazikani, Lesk et al. 1997). The remaining two disallowed residues form part of a loop with poorly defined density connecting strands A and B of the heavy chain constant domain. This loop, running from H134 to H144, is often disordered in Fab structures (for examples see (Cheetham, Hale et al. 1998) (Li, Li et

al. 2000) (Augustine, de La Calle et al. 2001)). In addition, eight out of the remaining ten main-chain and side-chain parameters are described as 'better' than the PROCHECK comparison values.

In general, the electron density of the structure is of good quality as shown in Fig 8. However, a small number of residues (H64, L74, L77, L107) have no observable side chain electron density. A plot of the average B-factors for each Fab residue is shown in Fig. 9. Unsurprisingly, the region of highest B-factor occurs in the disordered C_H loop mentioned above. For the L chain, the highest B-factor values are present between L157 and L165. This loop has well-defined electron density suggesting that the high B-factors are due to flexibility rather than disorder.

Rmsd values for bonds and angles indicate good geometry. Table 4 summarises the refinement statistics for SV5-Pk Fab.

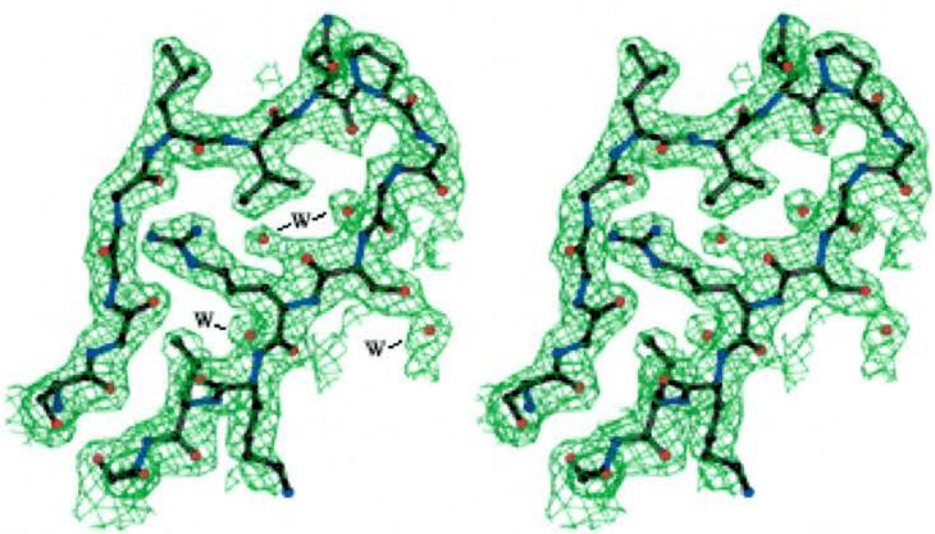


Fig 8. Stereo diagram of a region of the final 2.3 Å electron density map showing bound water molecules (W). The 2fo-fc map is contoured at 1σ.

Refinement Parameters	
Resolution (Å)	500 – 2.3
Nº of protein atoms	3252
Nº of water molecules	211
R Factors	
R[a] (%)	21.2
Rfree[b] (%)	25.7
Atomic Temperature Factors	
All atoms (Å^2)	32.28
Main-chain atoms (Å^2)	31.73
Side-chain atoms (Å^2)	32.88
Water (Å^2)	34.78
RMS Deviations from Ideal Geometry	
Bonds (Å)	0.006
Angles (°)	1.4
Ramachandran Plot	
Favourable (%)	87.6
Additional (%)	11.6
Generous (%)	0.0
Disallowed (%)	0.8

Table 4. Refinement Statistics for the PK Fab structure. [a] R-factor = $\Sigma \mid F_{obs}-F_{calc} \mid / \Sigma F_{obs}$. [b]R-free was calculated with 10% of the data omitted from refinement.

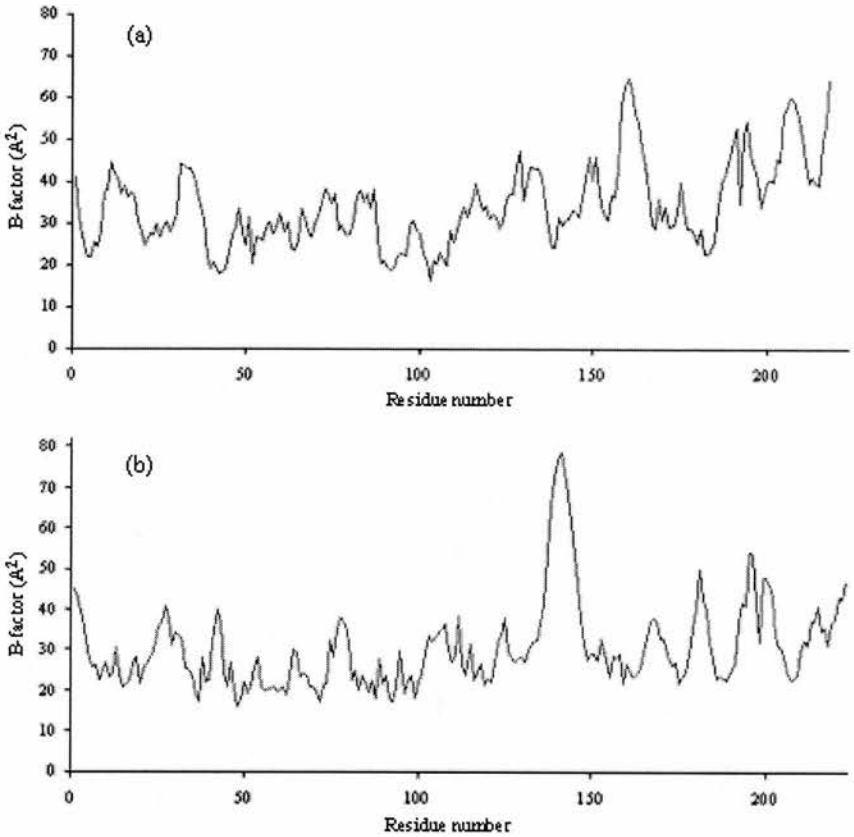
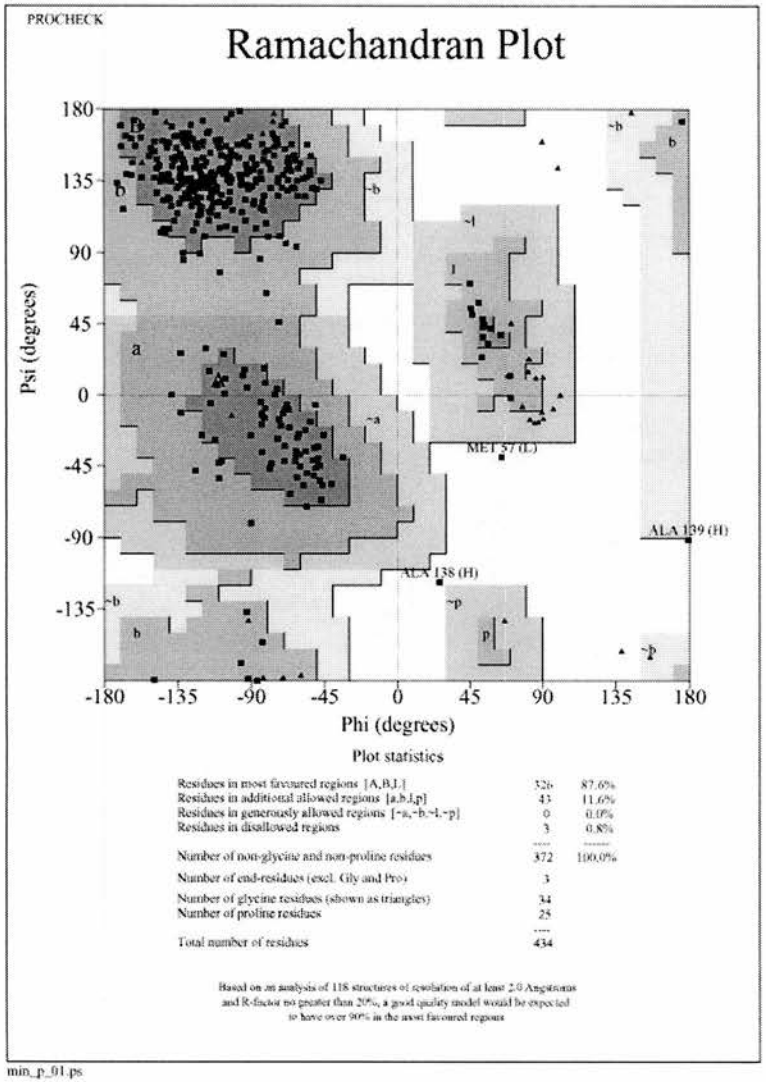


Fig 9. B-factor plots for (a) Light chain, (b) Heavy chain

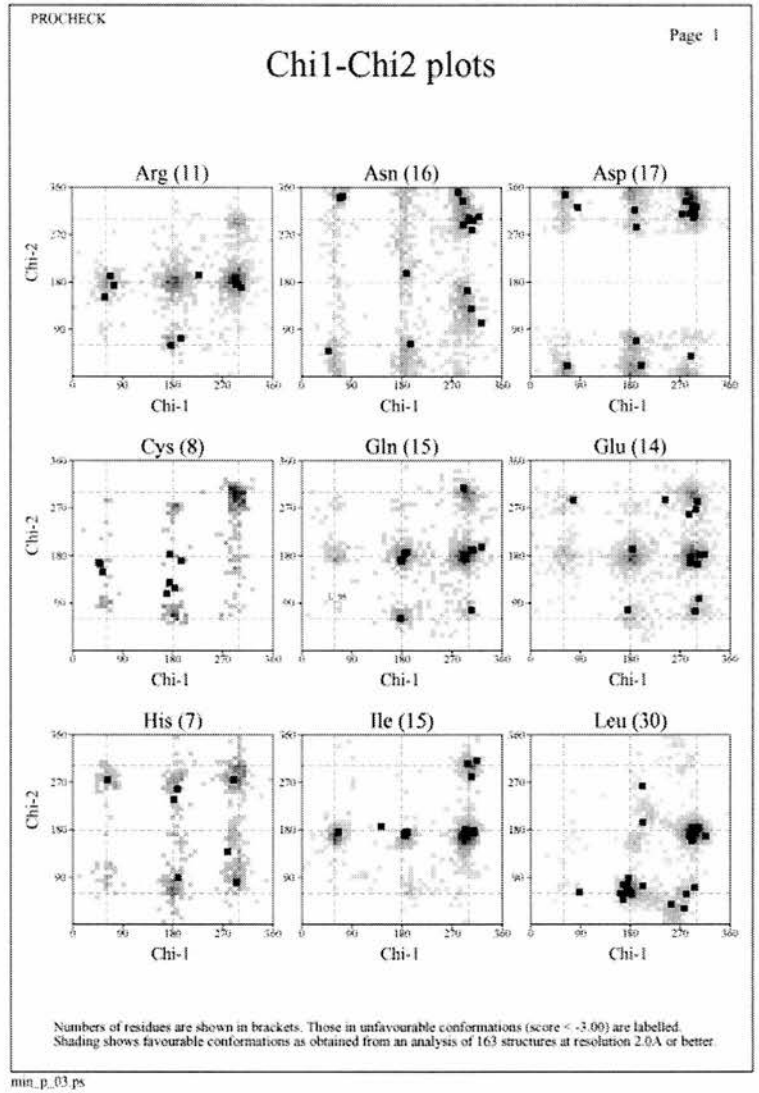
3.9 Summary

The structure of SV5-Pk Fab was solved by Molecular Replacement at 2.3 Å resolution to an R-factor of 21.2 % and an R-free of 25.7 %. The molecule has an elbow angle of 148°, which falls within the range of angles observed for Fab structures (127° to 194° (Wilson and Stanfield 1994)). The whole Fab is present in the structure with the exception of one lysine (H64) and three arginine (L74, L77, L107) side-chains.

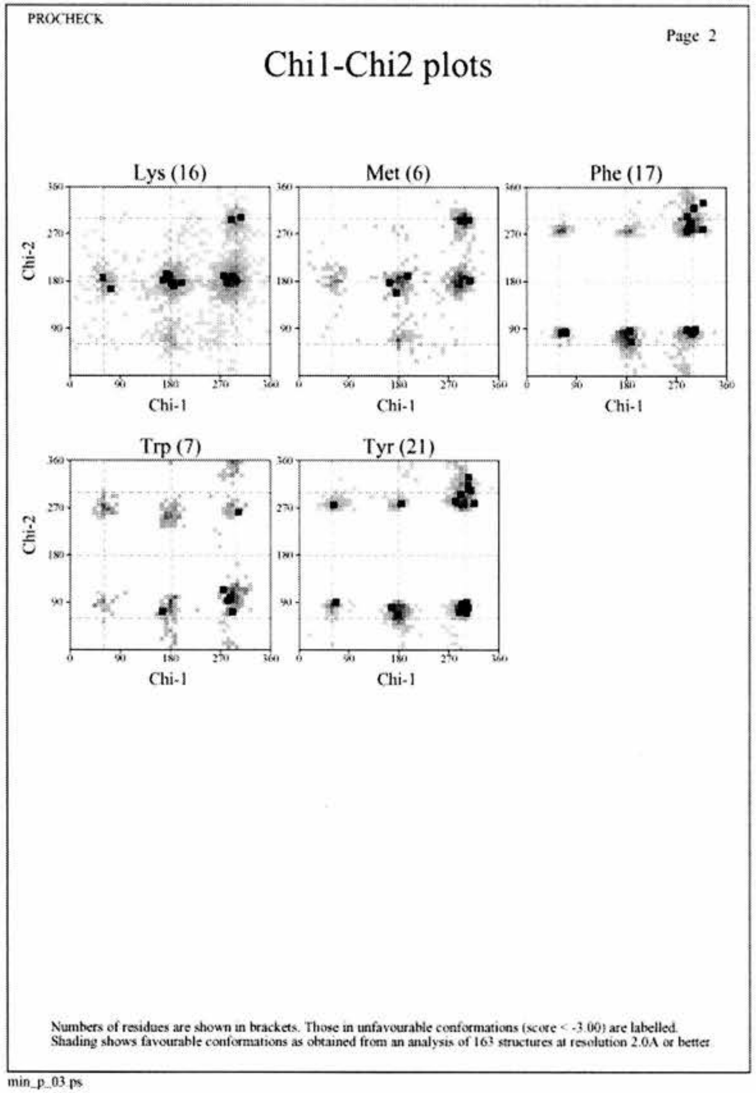
Five of the six CDRs of antibodies can be described by their canonical class, each CDR having only a small number of potential conformations. Of the 300 possible combinations of these classes, only 10 % occurred in an analysis of 381 antibody sequences carried out by Vargas-Madrazo et al (Vargas-Madrazo, Lara-Ochoa et al. 1995). The combination of SV5-Pk CDR loops, (1 3 4 1 1) for H1, H2, L1, L2 and L3 respectively, is one of three combinations that occur in peptide-binding antibodies. The (1 3 4 1 1) class is defined as multi-specific as it is also observed in antibodies that bind proteins, polysaccharides and nucleic acids. The categorisation of the combining site as multi-specific is not surprising as the antibody binds tightly to the V5 peptide whether on its own or as part of a larger recombinant protein.

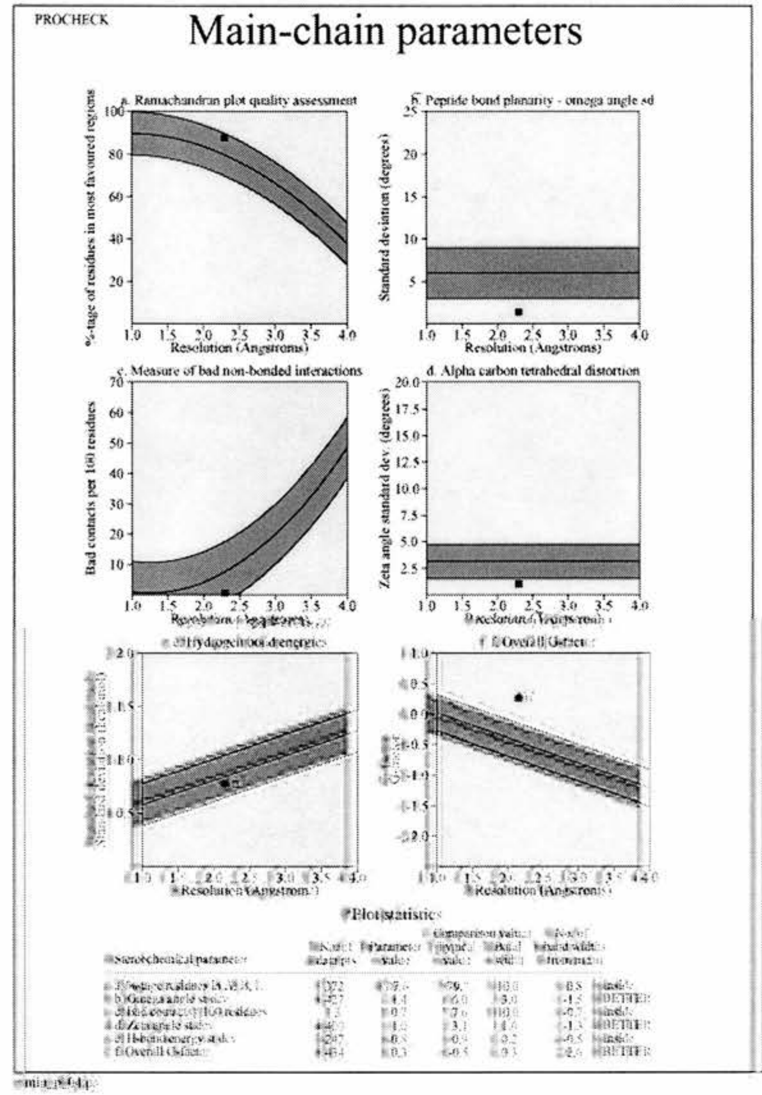


min_p_01.ps



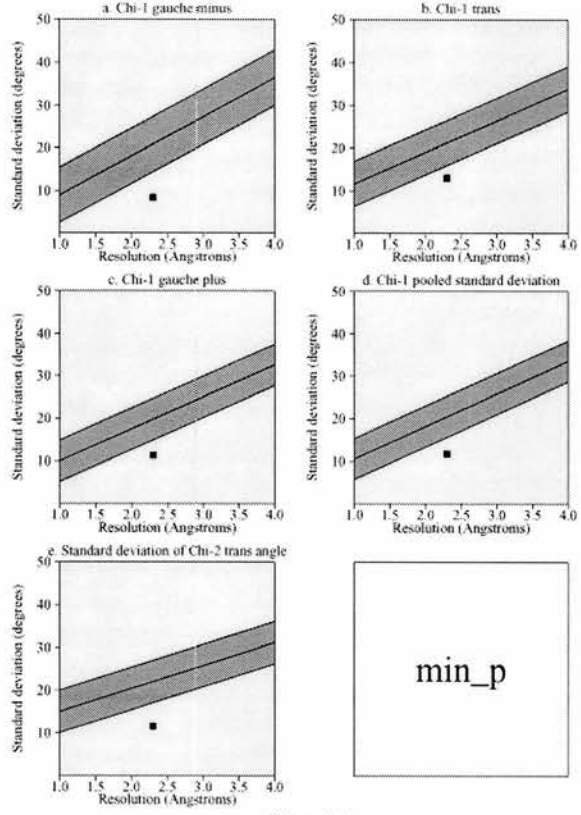
min_p_03.ps


PROCHECK
Page 2
Chi1-Chi2 plots
Lys (16)
Met (6)
Phe (17)
Trp (7)
Tyr (21)
Chi-1
Chi-2
Numbers of residues are shown in brackets. Those in unfavourable conformations (score < -3.00) are labelled.
Shading shows favourable conformations as obtained from an analysis of 163 structures at resolution 2.0A or better
min_p_03.ps


PROCHECK
Main-chain parameters
a. Ramachandran plot quality assessment
%-tage of residues in most favoured regions
Resolution (Angstroms)
b. Peptide bond planarity - omega angle sd
Standard deviation (degrees)
c. Measure of bad non-bonded interactions
Bad contacts per 100 residues
d. Alpha carbon tetrahedral distortion
Zeta angle standard dev. (degrees)
Plot statistics

PROCHECK

Side-chain parameters



Plot statistics

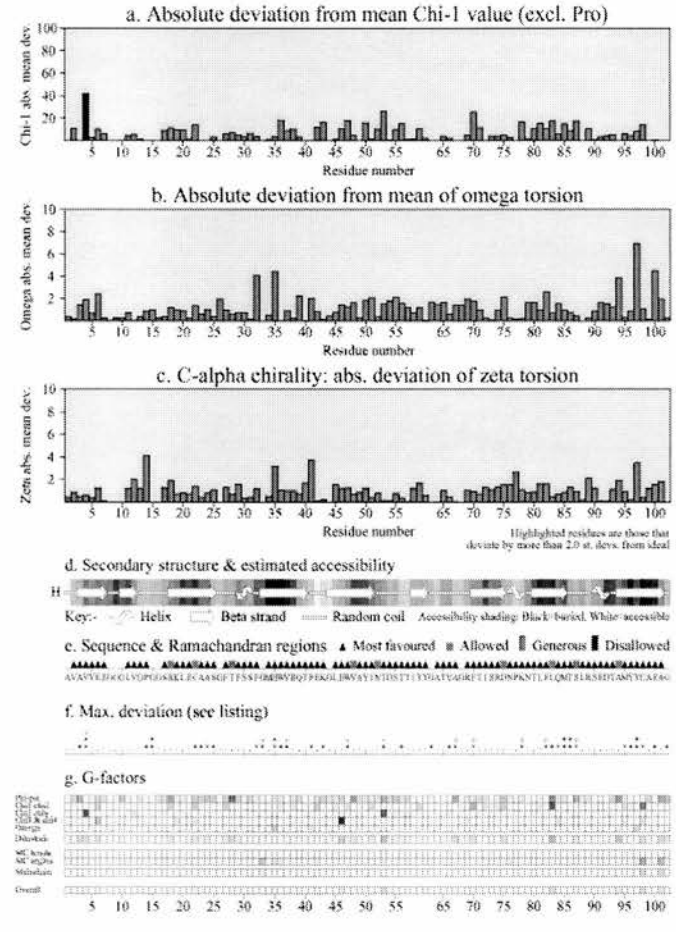
Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. Chi-1 gauche minus st dev	67	8.4	20.9	6.5	-1.9	BETTER
b. Chi-1 trans st dev	100	13.0	21.2	5.3	-1.5	BETTER
c. Chi-1 gauche plus st dev	168	11.3	19.8	4.9	-1.7	BETTER
d. Chi-1 pooled st dev	335	11.7	20.5	4.8	-1.8	BETTER
e. Chi-2 trans st dev	73	11.5	22.0	5.0	-2.1	BETTER

min_p_05.ps

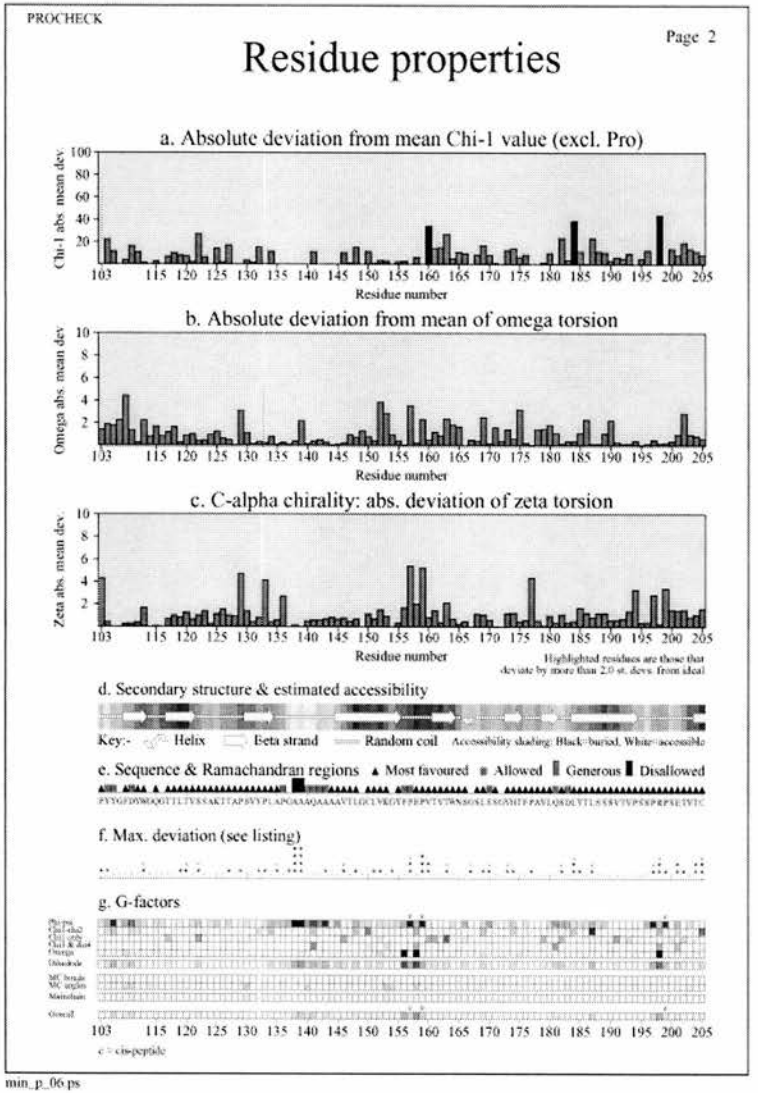
PROCHECK

Page 1

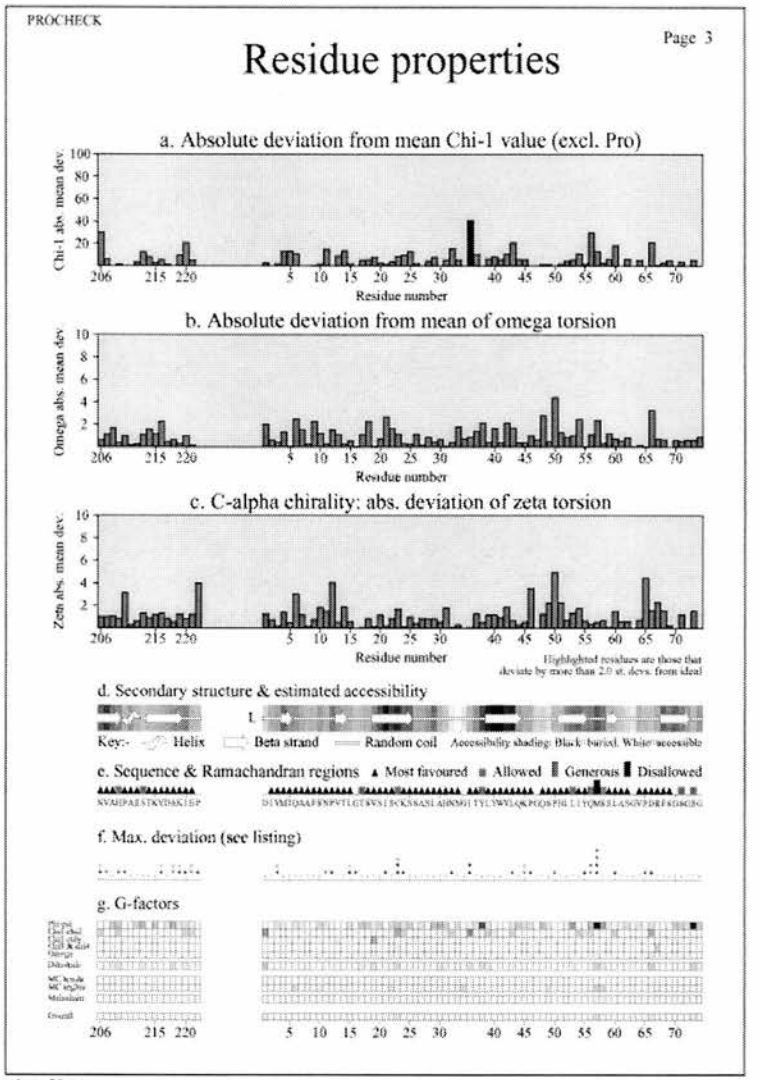
Residue properties



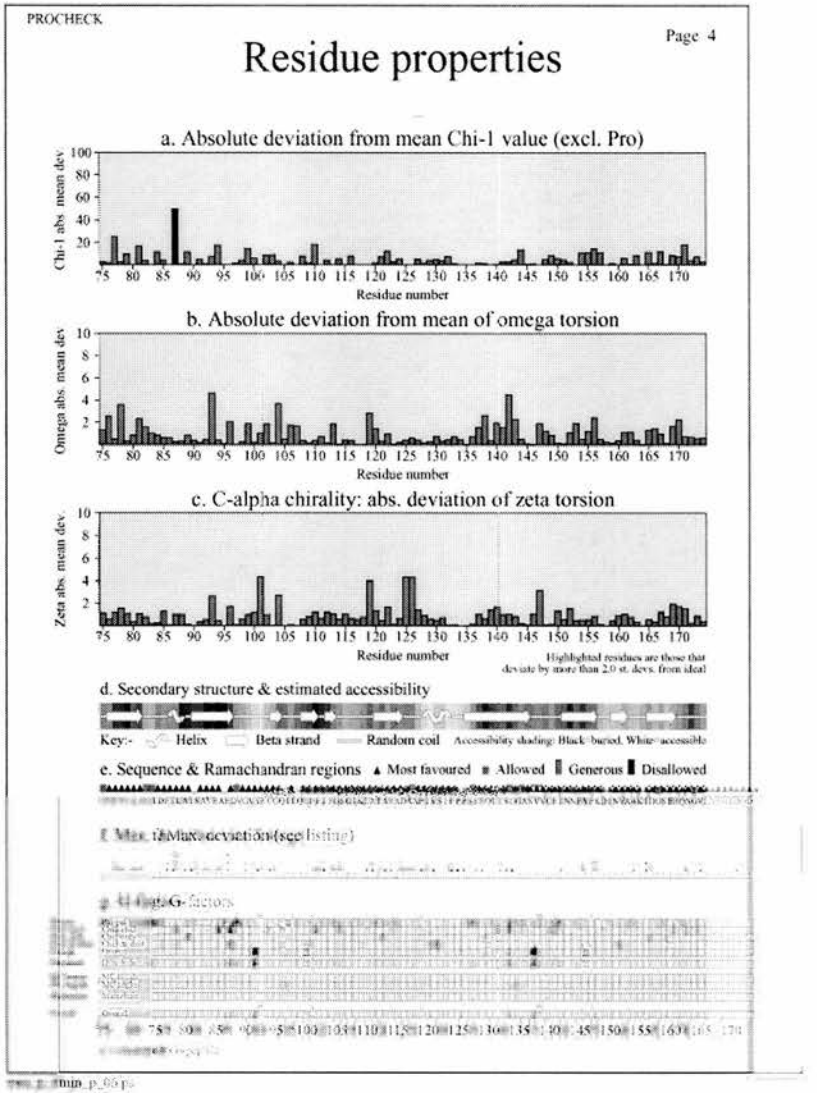
min_p_06.ps



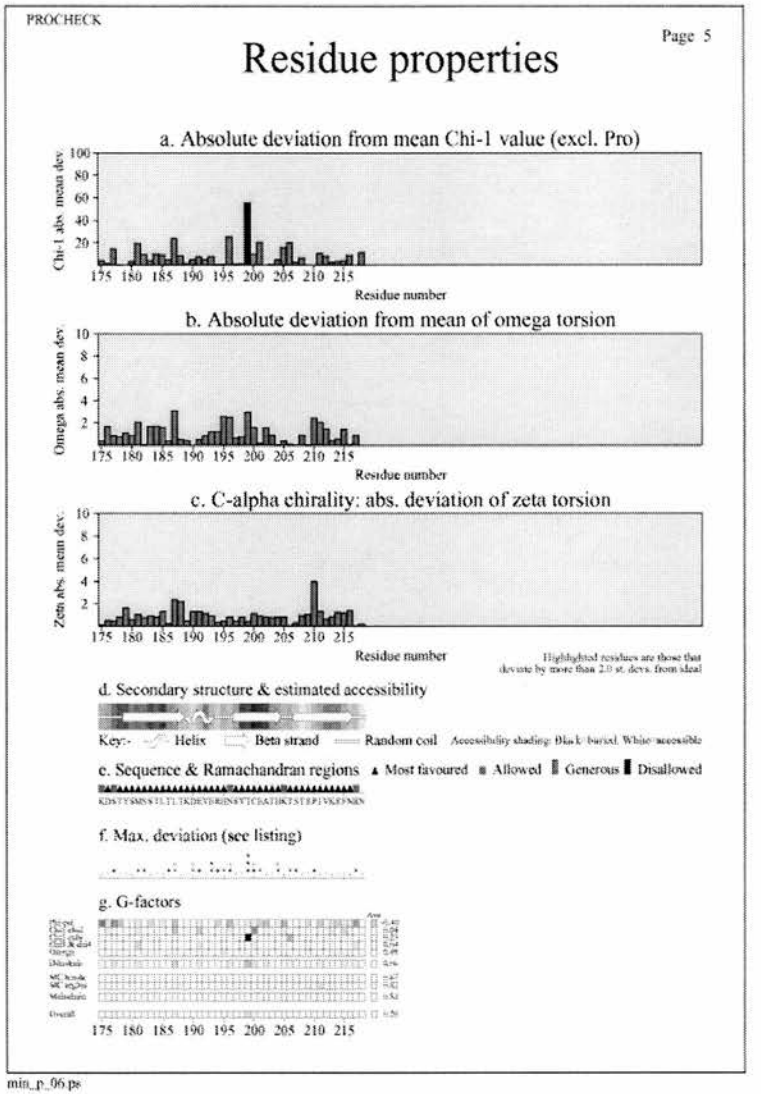
min_p_06.ps



min_p_06.ps



min_p_06.ps

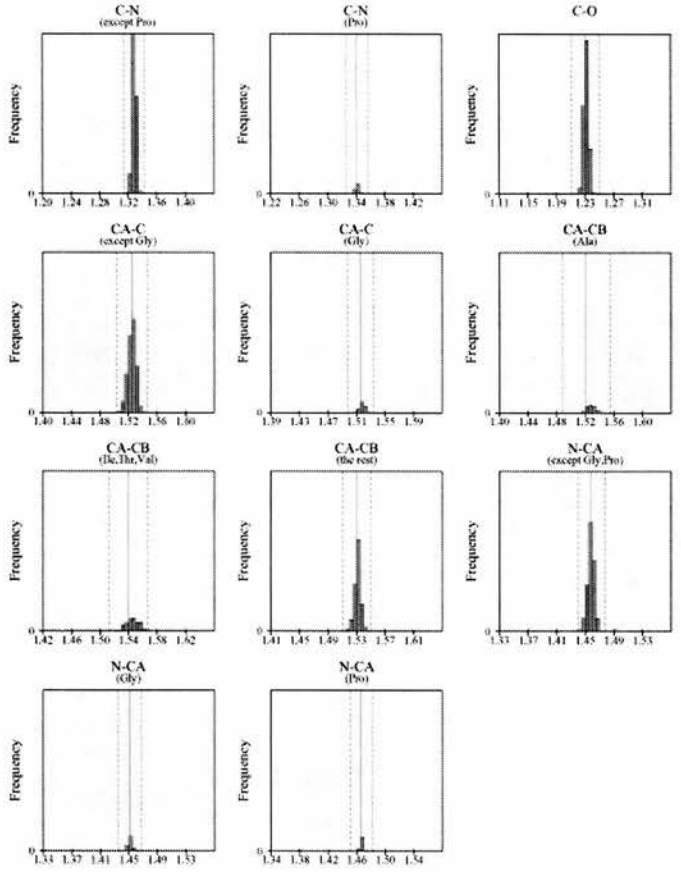


min_p_06.ps

PROCHECK

Page 1

Main-chain bond lengths



Black bars > 2.0 st. devs. from mean.

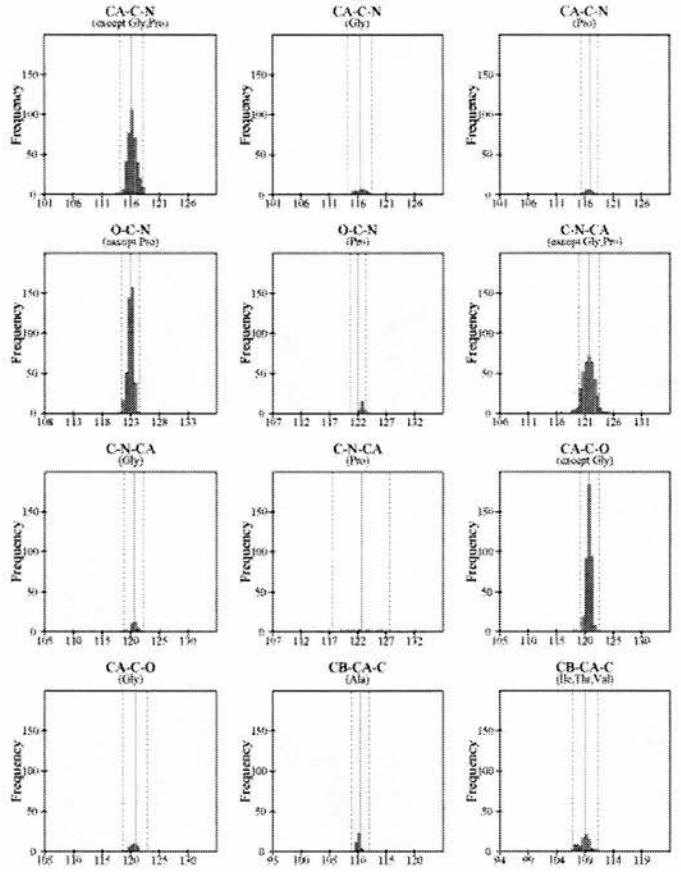
Solid and dashed lines represent the mean and standard deviation values as per Engh & Huber small-molecule data.

min_p_07.ps

PROCHECK

Page 1

Main-chain bond angles



Black bars > 2.0 st. devs. from mean.

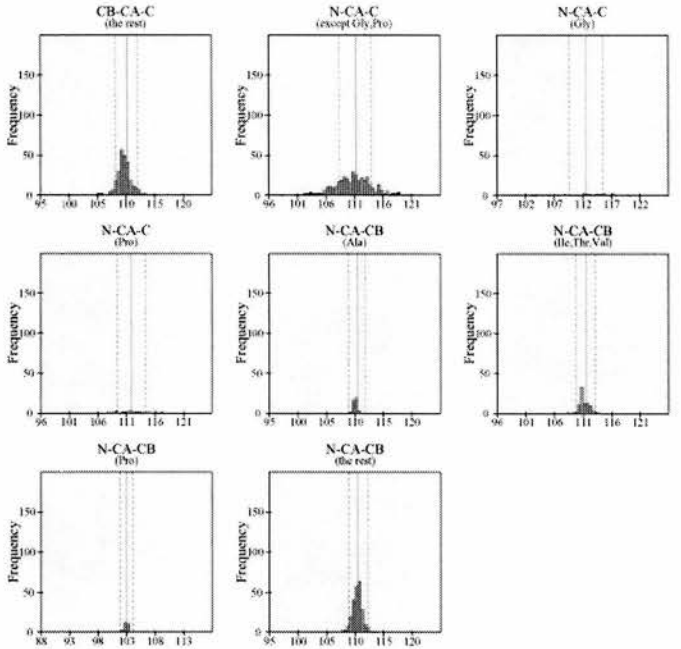
Solid and dashed lines represent the mean and standard deviation values as per Engh & Huber small-molecule data.

min_p_08.ps

PROCHECK

Page 2

Main-chain bond angles



Black bars > 2.0 st. devs. from mean.

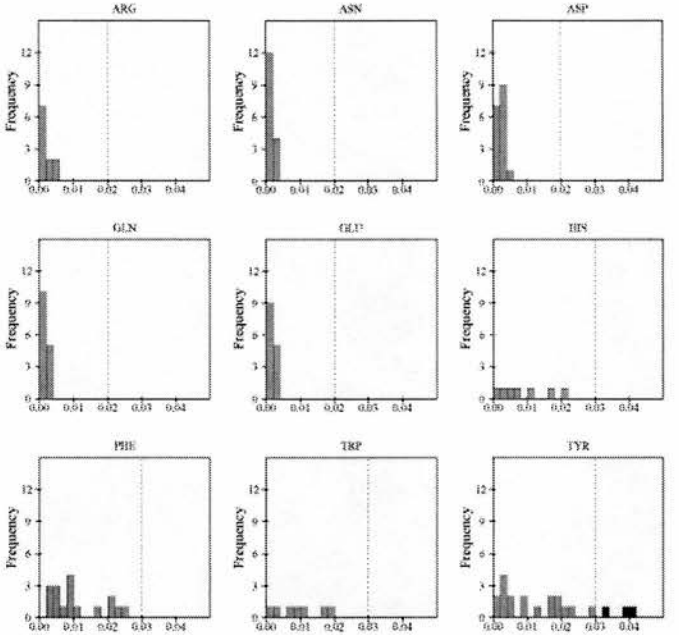
Solid and dashed lines represent the mean and standard deviation values as per Engh & Huber small-molecule data.

min_p_08.ps

PROCHECK

Page 1

RMS distances from planarity



Histograms showing RMS distances of planar atoms from best-fit plane.
Black bars indicate large deviations from planarity: RMS dist > 0.03 for rings, and > 0.02 otherwise.

min_p_09.ps

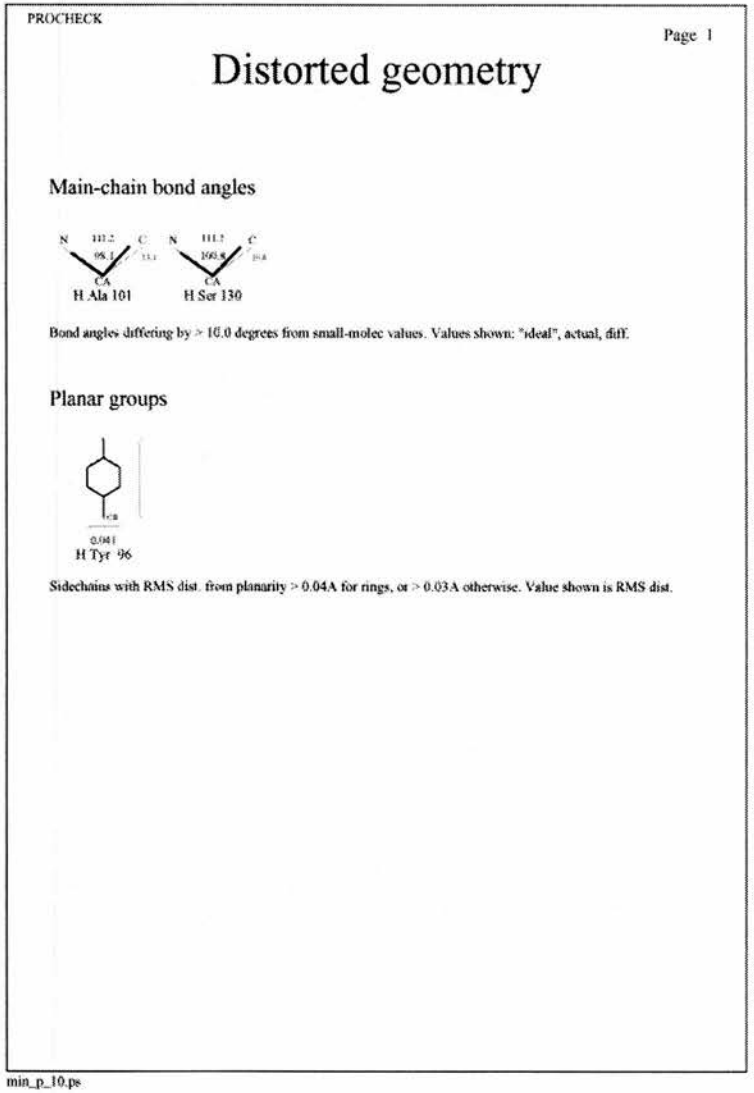


Fig. 10. PROCHECK results detailing the geometry of the SV5-Pk Fab structure.

Chapter 4

Crystallisation and Structure Determination of the SV5-Pk Fab-Peptide Complex

4.1 Introduction

The 14 amino acid V5 peptide (GLPIPNPLLGLDST), derived from a linear epitope on the SV5 phosphoprotein (P), is widely used to tag recombinant proteins which may then be detected with the SV5-Pk antibody. The peptide binds the antibody with very high affinity ($K_d = 2.36 \times 10^{-11}$ M) and requires harsh conditions for elution, making the system unsuitable for purification of tagged recombinant proteins.

The three-dimensional structure of the SV5-Pk Fab/V5 peptide complex will provide insights into the interactions involved. This information may help explain the unusually tight binding and allow the design of a peptide with slightly reduced affinity for use as an aid to purification of proteins tagged with the V5 peptide.

4.2 Peptide soaks

Attempts were made to soak the peptide into the Fab crystals obtained previously (see Chapter 3). However, at relatively high concentrations of peptide (10 μg/ml), the crystals became "furry" in appearance and eventually cracked. At lower concentrations (1μg/ml) the crystals appeared normal but subsequent data collection and analysis of the structure gave no indication that the peptide had bound.

4.3 Co-crystallisation

Co-crystallisation trials were set up using a solution containing 5 mg/ml Fab and 0.2 mg/ml peptide. Small needle-shaped crystals were obtained in 8 of the Hampton Screen I conditions and plate-like crystals were observed in one condition: 20% isopropanol, 0.1 M Na Citrate pH 5.6, 20 % PEG 4K. The latter crystals were

used for in-house data collection. Data processing revealed high mosaicity (1.8°). The 3 Å resolution structure was solved by molecular replacement using the native structure as the phasing model. Inspection of the 2Fo-Fc electron density map indicated that the peptide was present in the combining site. However, the peptide was very poorly defined and the structure was not refined further.

Further crystal trials of the complex were carried out in an attempt to obtain more ordered crystals and higher resolution data. Improved crystals were obtained in one Hampton Cryoscreen condition (0.08 M Na Acetate pH 4.6, 0.16 M Ammonium Sulphate, 20 % PEG 4K, 20 % glycerol).

4.4 Data Collection

The crystals, which already contained cryoprotectant, were flash-frozen in a nitrogen stream. Diffraction data were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) ID14.4 beamline. 180 frames of data were collected in 1° oscillations at a crystal-to-detector distance of 175 mm (a diffraction image is shown in Fig. 1). The crystals diffracted to 1.9 Å. Data processing and scaling was carried out using DENZO and SCALEPACK software (Otwinowski and Minor 1997). Table 1 summarises the data collection statistics.

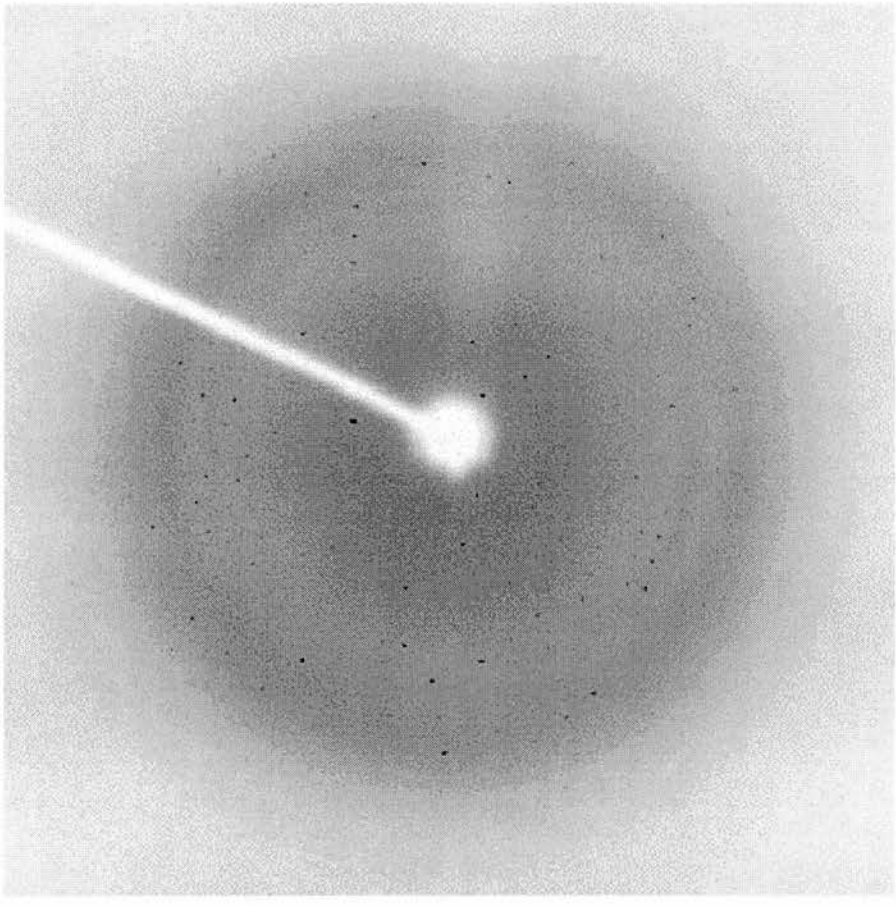


Fig 1. Diffraction pattern of an SV5-Pk Fab/peptide crystal. The resolution at the edge of the image is 1.9Å.

4.5 Structure Determination and Refinement

The structure was solved by Molecular Replacement with AMoRe (Navaza 1994) using the non-liganded SV5-Pk Fab structure as the phasing model. The initial R-factor and R-free values were 44 % and 44.5 % respectively. The peptide was modelled into the electron density within the antigen-binding site.

Repositioning of some combining site residues was required, particularly in the third CDR of the heavy chain. Subsequent rounds of model building, water-picking and minimisation reduced the R-factor to 21.4 % and the R_{free} to 24.7 % (Table 2).

	All Data	Highest resolution Shell (1.97-1.9 Å)
Crystal Parameters		
Space Group	C2	-
Unit cell dimensions	a=176.9 b=40.3 c=64.4 (Å)	-
	$\alpha = \gamma = 90.0$ $\beta = 98.7$ (^{o})	-
Resolution Limit (Å)	1.9	-
Mosaicity (^{o})	0.57	-
Data Processing		
Total N^{o} of observations	215808	-
N^{o} of unique reflections	35937	3428
R_{merge} [a] (%)	7.0	15.3
$I/(\sigma I)$	22.2	17.2
Completeness (%)	98.9	97.8

Table 1. X-ray diffraction data for crystals of the SV5-Pk Fab/peptide complex. [a] $R_{merge} = \Sigma|I - \langle I \rangle|/\Sigma\langle I \rangle$, where $\langle I \rangle$ represents the average intensity of symmetry-equivalent reflections.

Refinement Parameters	
Resolution (Å)	500 – 1.9
Nº of protein atoms	3387
Nº of water molecules	372
R Factors	
R[a] (%)	21.4
Rfree[b] (%)	24.7
Average B factors	
All atoms (Å^2)	24.0
Main-chain atoms (Å^2)	22.8
Side-chain atoms (Å^2)	25.1
Water (Å^2)	34.4
Peptide (all atoms)	30.4
RMS Deviations from Ideal Geometry	
Bonds (Å)	0.0065
Angles (°)	1.43
Ramachandran Plot	
Favourable (%)	89.2
Additional (%)	10.3
Generous (%)	0.0
Disallowed (%)	0.5

Table 2. Refinement Statistics for the SV5-Pk/V5 peptide complex structure. [a] R-factor = $\Sigma \mid F_{obs} - F_{calc} \mid / \Sigma F_{obs}$. [b]R-free was calculated with 10% of the data omitted from refinement.

4.6 Quality of the Structure

The final structure contained 372 water molecules. Peptide residues P1 to P11 had clearly defined electron density. Residues P12 to P13 were poorly defined and P14 could not be modelled into the structure. Figure 2 shows residues P1 to P11 and the corresponding 2Fo-Fc electron density map.

In the 1.9 Å resolution structure, the H134-H144 loop that was poorly defined in the previous 2.3 Å structure could be assigned more accurately. In contrast to the lower resolution structure, these residues possess allowed Phi-Psi angles. All residues lay in allowed areas of the Ramachandran plot with the exception of L51 (see Chapter 3). Rms deviation from ideal bond distances is 0.0065 Å and from ideal angles is 1.4°.

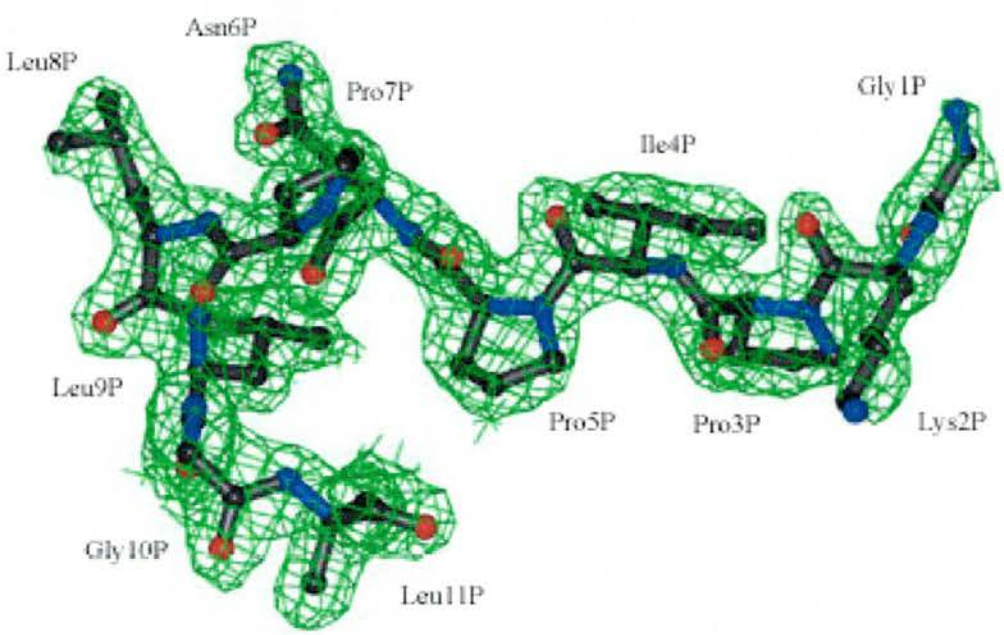


Fig 2. V5 peptide (residues P1 - P11) 2Fo-Fc electron density map contoured at 1σ.

4.7 Fab-peptide interactions

Residues Gly1P, Asp12P and Ser13P, at the ends of the peptide extend out into the solvent. The central portion of the peptide binds a shallow U-shaped groove formed by the CDRs (Fig. 3) and includes a section of 3_{10} helix formed by residues Asn6P to Gly10P. The side-chains of residues Asn6P, Leu9P and Leu11P extend into the combining site.

48 % of the peptide surface (691 $Å^2$) is buried upon binding. The buried surface area of the Fab at the binding site is 546 $Å^2$. The heavy and light chain CDRs contribute 54 % and 46 % of the Fab buried surface area respectively. Buried surface areas were measured using the method of Lee and Richards (Lee and Richards 1971). Table 3 lists the percentage of each peptide residue that is buried upon Fab binding.

Tables 4 and 5 list the contacts between the ligand and water molecules and between ligand and Fab atoms. The hydrogen bonds are also illustrated in Fig. 4. 11 SV5-Pk Fab residues from CDRs H2, H3, L1, L2 and L3 interact with the peptide directly. CDR H1 does not contribute to antigen binding. The antibody-antigen interactions include seven hydrogen bonds, about average for antibody-peptide interactions. Tyr50H, Asn52H and Tyr59H from CDR H2 are involved in three of the hydrogen bonds, with the remaining hydrogen bonds being formed by residues Tyr34L, Tyr49L, Ile91L, Leu92L from the three light chain CDRs. Interestingly, only one side chain atom of the peptide Asn6P($N^{\delta 2}$) is involved in hydrogen bonding with the Fab, the remainder occurring via main chain atoms. Asn6P is buried in the combining site and makes van der Waals contacts with CDR L1.

CDR H3, in particular Tyr99H, makes extensive van der Waals contacts with residues P3, P5 and P11 of the peptide.

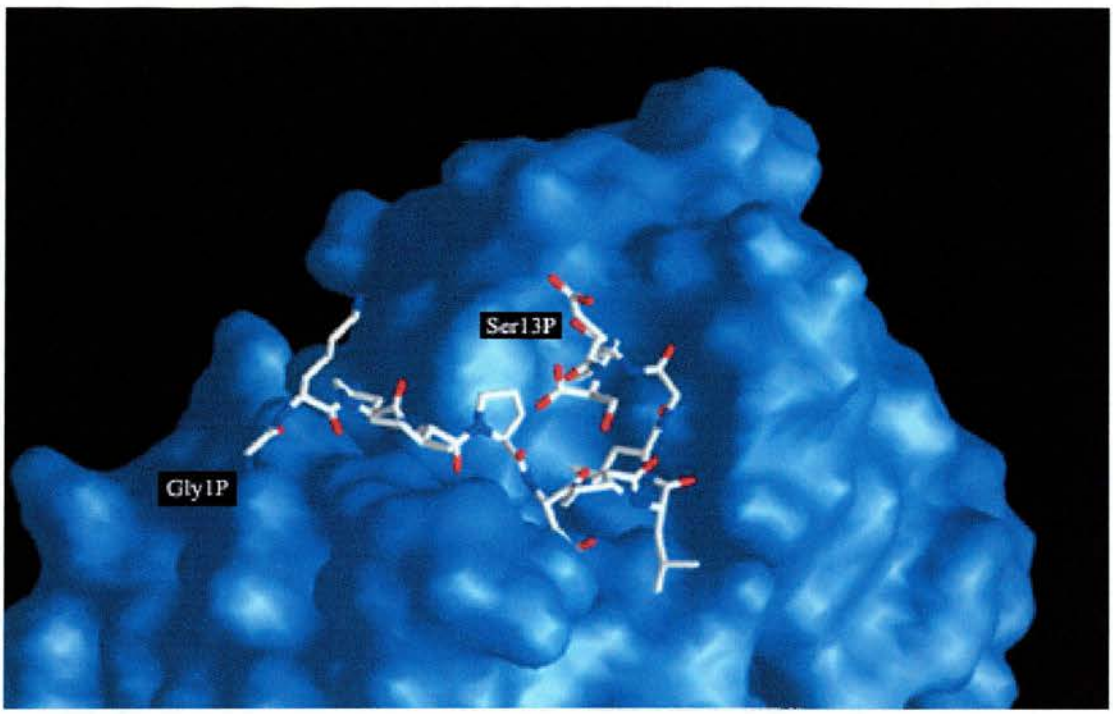


Fig 3. Position of the peptide in the antibody combining site. The first and last peptide residues are labelled. The figure was produced using GRASP (Nicholls and Honig 1991).

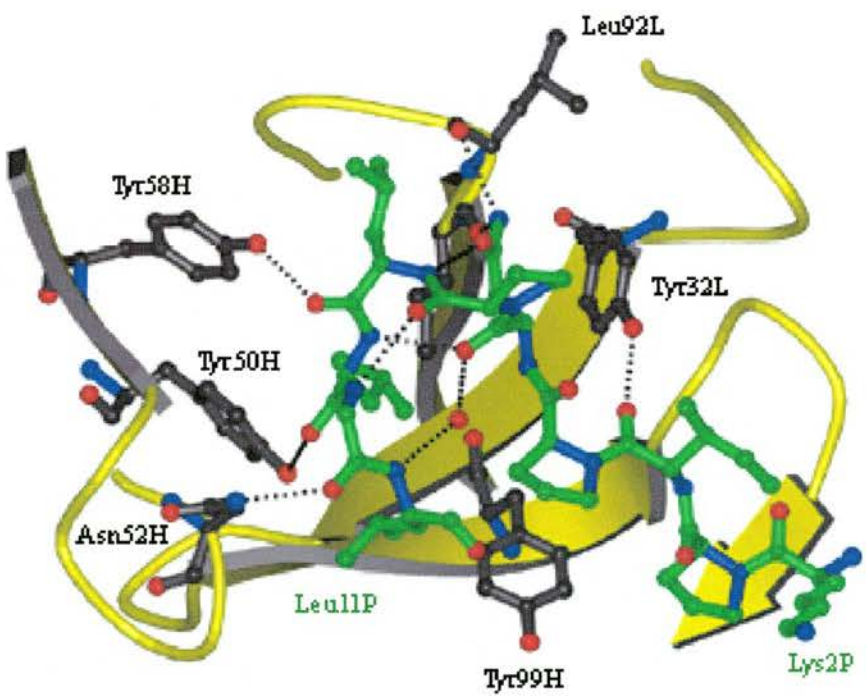


Fig 4. Hydrogen bonds at the Fab-peptide interface. Some CDR side chains (black) and residues P2 to P11 of the peptide are shown (green).

Peptide Residue	1	2	3	4	5	6	7	8	9	10	11	12
% Buried in Complex	61.4	15.6	83.1	81.3	92.9	100	79.7	98.2	100	71.9	88.0	8.2

Table 3. Percentage surface area of each peptide residue buried upon SV5-Pk Fab binding

Fab atom		Peptide atom		**Distance (Å)**
L1 Tyr34L	(OH)	Ile4P	(O)	2.66
L2 Tyr49L	(OH)	Gly1P	(O)	2.73
L3 Ile91L	(O)	Asn6P	($N^{\delta 2}$)	2.78
Leu92L	(O)	Asn6P	($N^{\delta 2}$)	3.08
H2 Tyr50H	(OH)	Leu9P	(O)	2.59
Asn52H	($N^{\delta 2}$)	Gly10P	(O)	3.01
Tyr59H	(OH)	Leu8P	(O)	2.67
Peptide atom 1		**Peptide atom 2**		
Gly10P	(N)	Pro7P	(O)	3.22
Leu9P	(N)	Asn6P	(O)	2.96
Leu8P	(N)	Asn6P	($O^{\delta 1}$)	2.77
Water		**Peptide atom**		
S99		Asn6P	(N)	3.17
S154		Gly1P	(O)	2.73
S182		Pro3P	(O)	2.73
S259		Pro3P	(O)	2.78
S262		Pro7P	(O)	3.02
S307		Asn6P	($O^{\delta 1}$)	2.84
S341		Asn6P	(O)	2.77
S344		Pro5P	(O)	2.53
S344		Asp12P	(O)	3.05

Table 4. Hydrogen bonds in the Fab-peptide complex including water-peptide and peptide-peptide interactions. Contacts and distances were calculated using the CNS program CONTACT.

17 water molecules are present in the combining site. Nine water molecules contact the peptide directly, four of which (S99, S307, S341 and S344) are completely buried. S99 fills a cavity between Asn6P of the ligand and the Fab residues Tyr108H and Tyr34L. S341 links two peptide residues (Asn6P and Leu11P). S307 and S344 occur in two separate hydrogen bond networks (fig. 5).

The two surfaces have a shape correlation statistic of 0.76, as determined by the CCP4 program SC. This value is relatively high for antibody-protein interactions, indicating good shape complementarity between SV5-Pk Fab and the peptide. The shape complementarity is enhanced by the presence of several buried water molecules (S23, S99, S307, S341 and S344).

Fab atom		**Peptide atom**		**Distance (Å)**
L1 Tyr34L	($C^{\varepsilon 2}$)	Asn6P	(C^{α})	3.80
	(C^{ζ})	Asn6P	(C^{α})	3.57
	($C^{\varepsilon 2}$)	Asn6P	(C^{β})	3.84
	(C^{ζ})	Asn6P	(C^{β})	3.87
L2 Tyr49L	($C^{\varepsilon 2}$)	Pro3P	(C^{γ})	3.50
	(C^{ζ})	Pro3P	(C^{γ})	3.89
Thr97L	(C^{ζ})	Leu9P	($C^{\delta 2}$)	3.78
H3 Tyr98H	(C^{α})	Pro3P	(C^{β})	3.88
Tyr99H	($C^{\delta}1$)	Pro3P	(C^{β})	3.70
	(C^{β})	Pro5P	(C^{α})	3.75
	(C^{γ})	Pro5P	(C^{α})	3.87
	($C^{\delta 1}$)	Pro5P	(C^{β})	3.48
	($C^{\delta 2}$)	Pro5P	(C^{β})	3.45
	(C^{γ})	Pro5P	(C^{β})	3.55
	($C^{\delta 1}$)	Pro5P	(C^{δ})	3.61
	($C^{\varepsilon 1}$)	Pro5P	(C^{δ})	2.53
	($C^{\delta 2}$)	Leu11P	($C^{\delta 1}$)	3.57
	($C^{\varepsilon 2}$)	Leu11P	($C^{\delta 1}$)	3.59

Table 5. Non-specific van der Waals contacts between Fab and the peptide.

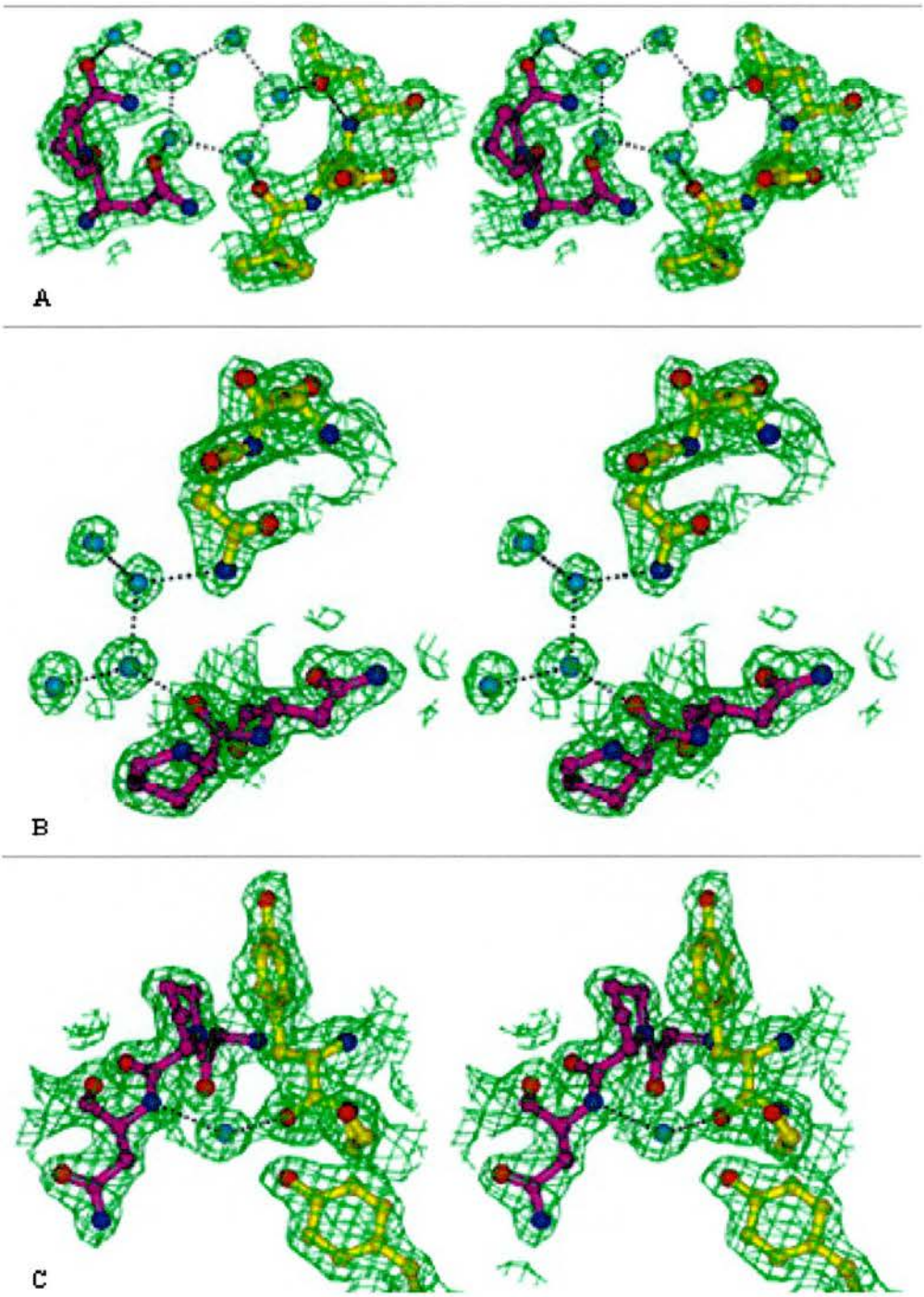


Fig 5. Stereo views of water molecules involved in peptide (magenta) and Fab (yellow) binding. A. CDR L3 connecting Asn6P and Pro7P via a network of waters; B. waters bridging CDR L1 and Pro5P; C. A buried water connecting Asn6P and CDR H3.

4.8 Comparison of Native and Peptide-bound Fab

The unliganded and peptide-bound Fab structures are very similar with rmsd values for the C_H, C_L, V_H and V_L sub-domains of 0.52, 0.26, 0.68 and 0.40 respectively.

The greatest deviation, in V_H, is mainly due to rearrangements in CDR H3. Upon complex formation, the C-α backbone of H3 moves by up to 2.2 Å. This shift is accompanied by large movements in the side chains of two tyrosine residues (fig. 6a). Tyr98H, which has very little contact with the peptide, moves by more than 6 Å at its OH position. Tyr99H also shifts significantly (5.2 Å in the OH group) to prevent a steric clash with Pro5P of the peptide. CDR H1, which is not directly involved in antigen binding, undergoes a rotamer change in Phe32H to accommodate the movement in CDR H3. Fig 6b shows the superposition of the liganded and unliganded antibody combining sites.

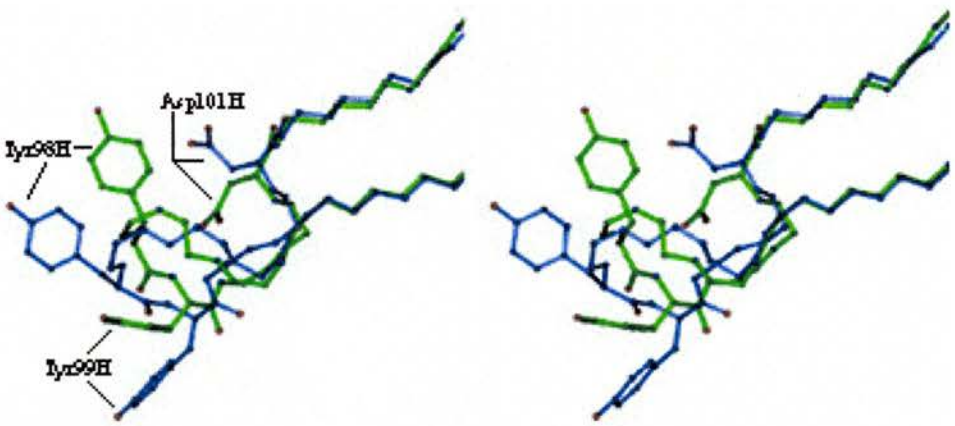


Fig 6a. Superposition of CDR H3 from unliganded (blue) and peptide-bound (green) SV5-Pk Fab showing the side chains of tyrosines H98 and H99.

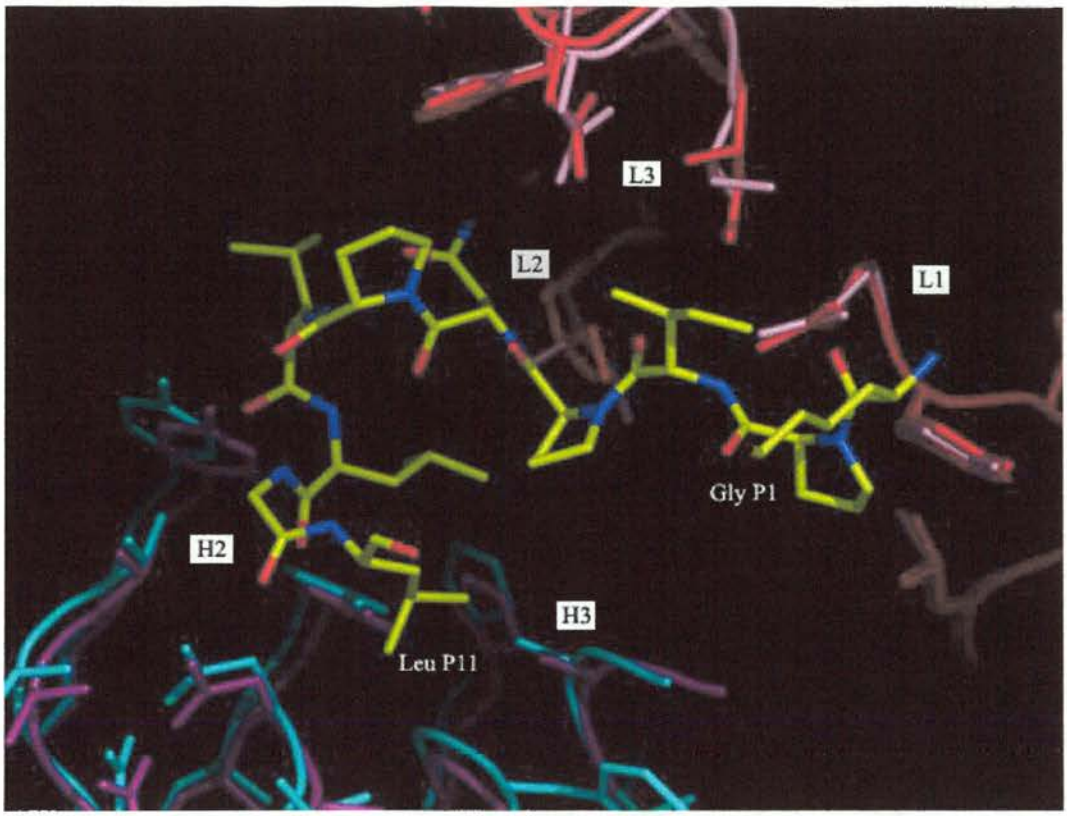


Fig 6b. Superposition of the unliganded (purple and pink) and peptide-bound (blue and red) SV5-Pk antibody conmbining sites. The peptide is shown in yellow.

There is a small change in the elbow angle from 148° in the native Fab to 143° in the liganded structure. Similar differences in the elbow angle have been seen between different crystal forms of the same Fab so it unclear whether the shift is due to peptide binding or the effects of crystal packing.

The surface area buried at the V_H/V_L interface area falls from 1588 $Å^2$ in the native Fab to 1462 $Å^2$ in the complex, a reduction of approximately 9%. This change is accompanied by a 10° rotation of V_H relative to V_L (Fig. 7).

The B-factors of the Fab atoms are higher overall in the unliganded molecule (the average B-factor over all atoms is 24 $Å^2$ in the complex and 32.3 $Å^2$ in the native). The increases are mainly confined to the constant regions of L and H, as shown in Fig. 8.

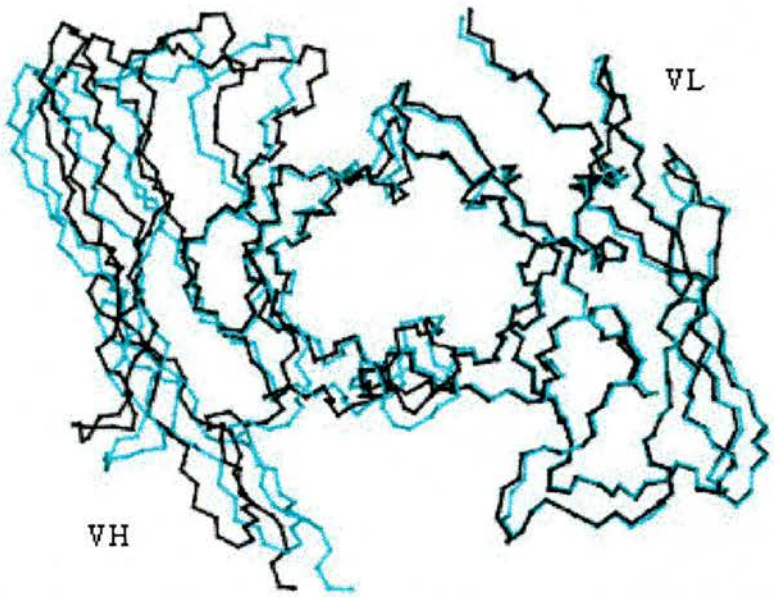


Fig 7. Unliganded (cyan) and liganded (black) variable regions shown superimposed at the V_L domains.

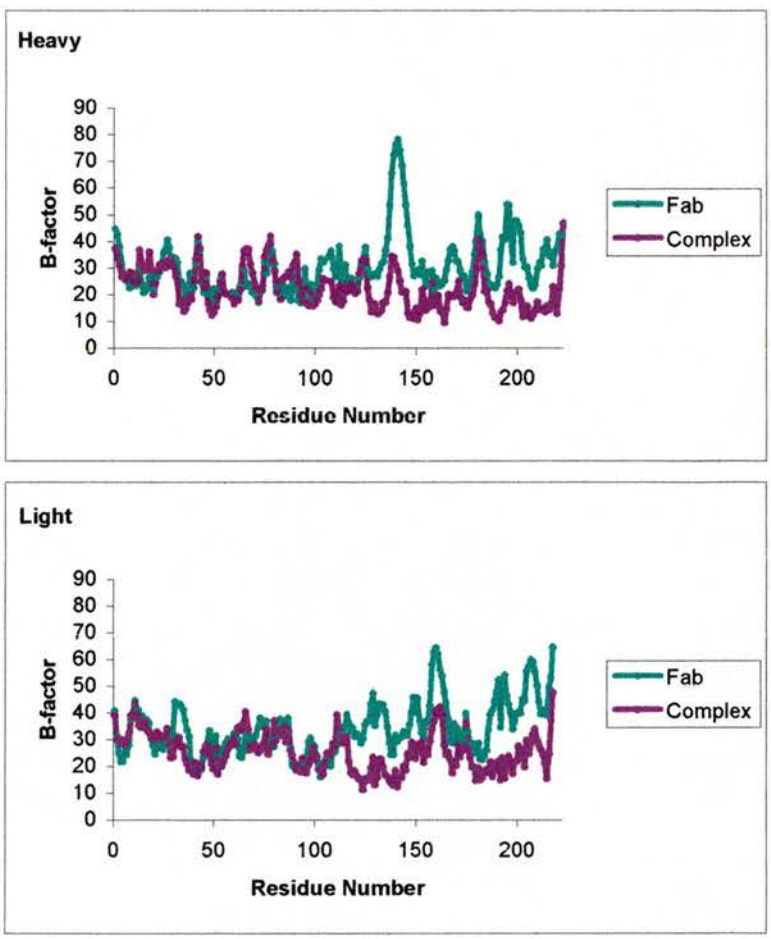


Fig. 8. B-factor plots of unliganded (green) and peptide-bound (magenta) Fab molecules.

The unit cell lengths of the liganded and unliganded forms of Fab crystals are noticeably different (179.6, 40.3, 64.4 and 69, 53.7, 127.9 respectively). Fig 9. shows the different packing arrangements of the two forms.

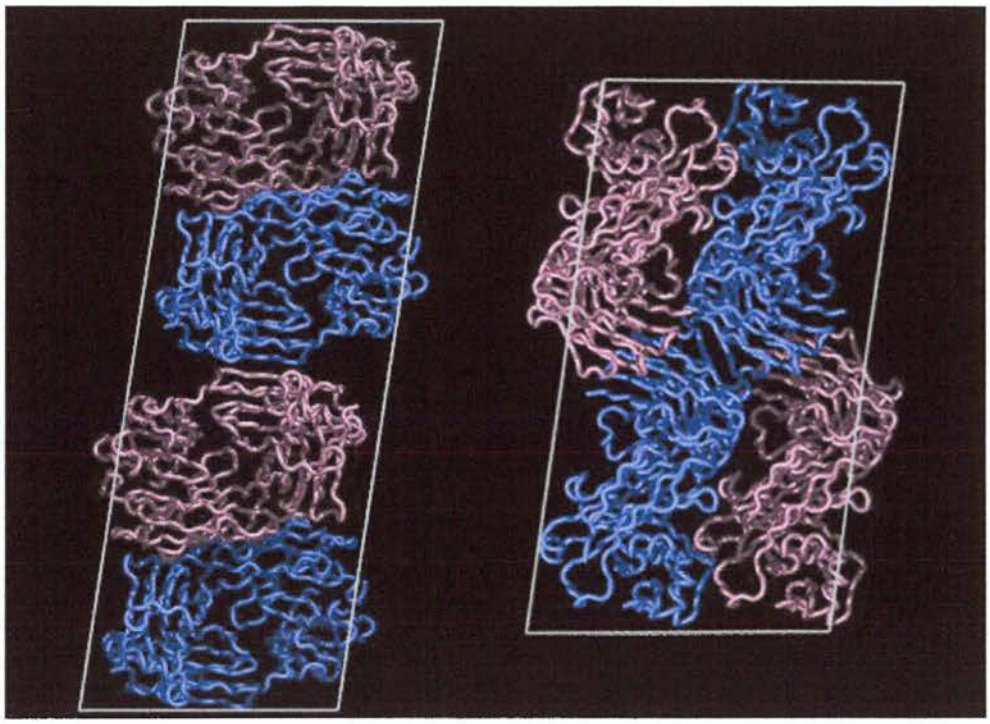


Fig. 9. Arrangement of Fab molecules in the the peptide-bound (left) and unliganded (right) crystals.

4.9 Comparison of SV5Pk-V5 with other Ab-peptide complexes.

The Protein Data Bank currently (Sep 2001) contains the co-ordinates of 27 Fab/Fv-peptide complexes and includes 20 different Fab molecules. For eight of these Fabs, the structures of the free antibody fragments are also known. Table 5 summarises some properties of known Fab-peptide complexes.

PDB code	Antibody	Res. (Å)	Peptide Length[a]	SC[b]	Peptide 2° Structure[c]	Fab-peptide interface[d]			Buried Surface Area		
						N° of Hydrogen Bonds	% Non-polar (Peptide)	% Non polar (Fab)	Peptide	Fab	% of peptide buried
SV5-Pk	SV5-Pk	1.9	12	0.76	3/10-helix	7	73.5	55	691	547	48
1a3r	8F5	2.1	15	0.76	3/10-helix	19	62.5	53	1094	861	59
1acy	59.1	3.0	11	0.8	β-turn	5	72.0	51	605	439	47
1ail	59.1	2.8	11	0.76	β-turn	4	72.0	52	569	395	46
1bog	Cb41	2.6	11	0.66	β-turn	4	52.3	57	812	639	57
1ce1	Campath-1H	1.9	8	0.80	β-turn	13	64.4	48	518	350	62
1cfn	Cb41	2.7	10	0.56	β-turn	4	56.7	54	550	472	51
1cfs	Cb41	2.8	11	0.73	β-turn	5	69.7	60	762	552	52
1cft	Cb41	2.8	5	0.61	β-turn	2	78.0	59	435	356	55
1cu4	3F4	2.9	10	0.75	β-turn	8	66.1	50	843	620	59
1e4w	Tab2	2.0	7	0.86	β-turn	7	56.0	60	522	402	54
1ejo	4C4	2.3	13	0.75	β-turn	11	61	52	813	601	54
1f58	58.2	2.0	13	0.79	β-turn	10	63.9	60	799	633	57
1f90	LKNB-2	2.6	9	0.69	α-helix	4	64.3	54	562	432	50
1fpt	C3	3.0	11	0.74	β-turn	7	52.8	56	612	440	43
1frg	Fab26/9	2.8	8	0.81	β-turn	10	69.7	53	646	472	57
1ggi	50.1	2.8	11	0.79	β-turn	5	73.8	48	665	454	50
1hh6	Cb41	2.6	11	0.67	β-turn	4	60.6	57	735	592	54
1hh9	Cb41	2.7	11	0.69	β-turn	3	62.3	54	760	608	57
1hi6	Cb41	2.6	11	0.68	β-turn	7	65.1	56	846	571	59
1ifh	17/9	2.8	7	0.80	β-turn	8	59.2	48	578	430	57
1mpa	Mn12H2	2.6	9	0.66	β-turn	11	43.1	57	605	435	67
1qkz	Mn14C11.6	2.0	10	0.67	β-turn	5	54.2	57	609	472	55
1sm3	Sm3	2.0	9	0.76	extended	4	59.2	57	594	482	50
1tet	Te33	2.3	12	0.79	β-turn	8	55.2	54	742	559	57
2igf	B13I2	2.8	7	0.75	β-turn	12	62.8	74	658	490	60
2f58	58.2	2.8	12	0.74	β-turn	7	61.3	53	742	559	57
2h1p	2H1	2.4	11	0.63	β-turn	7	68.0	63	672	581	49

Table 5. Comparison of some properties of antigen binding among known Fab/peptide complex structures. [a]Length of peptide visible in the structure is given. [b]Surface complementarity statistic calculated by the program SC (Lawrence and Colman 1993). [c]Peptide secondary structure was calculated using PROMOTIF (Hutchinson and Thornton 1996). [d]Fab-peptide interactions were analysed using the Protein-Protein Interaction Server (Jones and Thornton 1996).

4.10 Discussion

The crystallographic studies of the SV5Pk-Peptide complex provides structural explanations for some of the mutagenesis data previously obtained. Dunn *et al* used truncation mutants of the V5 peptide to find the minimal binding epitope which extends from Ile4P to Leu11P (IPNPLLGL) (Dunn, O'Dowd et al. 1999). This is in good agreement with the Fab-peptide structure in which residues Pro3P to Leu11P have well defined electron density and low B-factors (< 30 $Å^2$) relative to the ends of the peptide which are disordered (Fig. 9). Furthermore, all non-specific van der Waals interactions and all but one of the Fab to peptide hydrogen bonds occur between Ile4P and Gly10P. All of the intra-peptide bonds are confined to the helical portion of the peptide running from Asn6P to Gly10P.

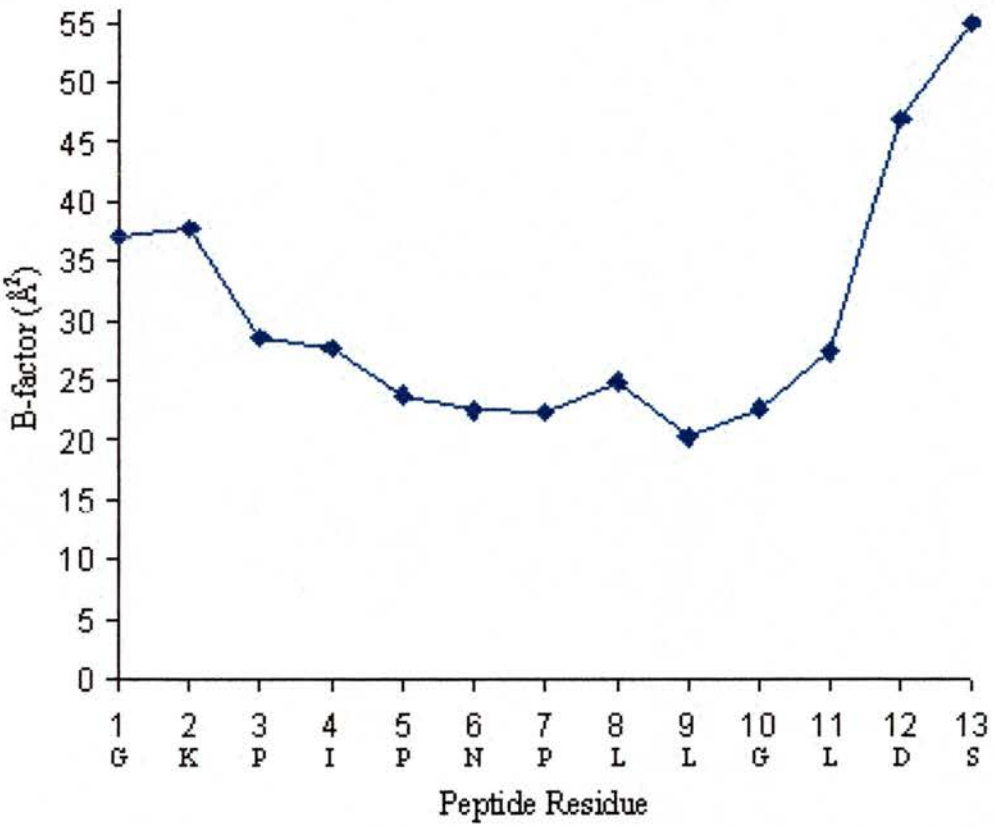


Fig. 9. Average B-factors for each peptide residue

Dunn *et al* also investigated the effect of cysteine substitutions and insertions on SV5-Pk binding. Binding was completely prevented by mutation of Leu9P to cysteine, and mutation of Pro5P or Asn6P to cysteine resulted in reduced affinity of the peptide for the antibody. Mutation of the proline residue may affect binding by distorting the peptide and reducing shape complementarity in a deeply buried region. Indeed, insertion of a cysteine residue between Pro5P and Asn6P totally abolishes binding. This effect is not produced by insertion in any other region of the peptide. Asn6P is the only residue that hydrogen bonds to the antibody via side chain atoms. These bonds would not exist in a cysteine substituted peptide. In addition, the asparagine side chain is completely buried in the antibody-combining site so a mutation at this residue may be expected to have a substantial effect on antibody binding. It is unclear why the Leu9Cys mutation prevents any interaction with the antibody as it makes only one hydrogen bond via its main chain oxygen which should be unaffected by the mutation. However, this residue forms part of the 3_{10} helix which may be disrupted by the mutation.

The amino acid sequences of the SV5-Pk V_H and V_L domains have been reported and are in agreement with the 1.9 Å structure (Hanke, Botting et al. 1994). The full Fab sequence is given in Fig. 10.

All CDRs except H1 contribute to the Fab-peptide interaction. CDR H3, almost always most prominent in antigen binding, forms no hydrogen-bonds but makes extensive van der Waals contacts with the peptide (Table 4). Seven hydrogen bonds are formed between the peptide and the Fab, excluding those made by bridging water molecules.

Heavy Chain

DVQVVESGGG LVQPGGSRKL SCAASGFTFS SFGMHWVRQT PEKGLEWVAY
INTDSTTIYY GDTVKGRFTI SRDNPKNTLF LQMTSLRSED TAMYYCASAG
PYYGFDYWGQ GTTLTVSSAK TTAPSVYPLA PVAGDTTGSS VTLGCLVKGY
FPEPVTLTWN SGSLSSGVHT FPAVLQSDLY TLSSSVTVTS STWPSETITC
NVAHPASSTK VDRKIEPR

Light Chain

DIVMTQAAFS NPVTLGTSVS ISCRSSKSLV HSNGITYLYW YLQKPGQSPH
LLIYQMSSLA SGVPDRFSGS GSGTDFTLRI SRVEAEDVGV YYCGQILELP
FTFGSGTKLE IRRADAAPTV SIFPPSSEQL TSGGASVVCF LNNFYPKDIN
VKWKIDGSER QNGVLNSWTD QDSKDTYSMS STLTLTKDEY ERHNSYTCEA
THKTSTSPIV KSFNRN

Fig 10. Amino acid sequence of SV5-Pk Fab. The C_H and C_L residues are underlined. The variable domain sequences were obtained from Hanke *et al* (1995) and the constant domain sequences were elucidated from the 1.9 Å Fab-peptide structure.

The helical conformation of the peptide within the combining site is unusual among known Fab-peptide complexes, β-turns being by far the most common conformation for Fab-bound peptides. Only two other complexes have helical peptides, 1a3r (a 3_{10} helix) and 1f90 (an α-helix) (Table 5).

It is not immediately obvious from the structure why the Fab-peptide interaction is so strong ($K_d = 2.36 \times 10^{-11}$ M). The shape complementarity statistic, number of hydrogen bonds and buried surface areas on both SV5-Pk Fab and ligand appear fairly typical for Fab/peptide complexes (Table 5). In common with other antibody-peptide complexes, the two binding surfaces are highly complementary. The buried surface areas on bound peptides are higher than those reported in the literature. The values given in Table 5 were calculated using the surface accessibility

algorithm of Lee and Richards (Lee and Richards 1971) which calculates the area covered by the centre of a 1.4 Å probe rolled over the molecular surface. Surface areas calculated with this method are therefore artificially low for concave surfaces (i.e. the antibody combining site) and artificially high for convex surfaces (i.e. the peptide). However, the buried surface areas are consistently higher than those calculated using the Connolly's molecular surface calculation (Connolly 1993) and would still be expected to highlight any unusual features.

The SV5-Pk/V5 complex is the highest resolution (1.9 Å) Fab/peptide structure obtained to date and provides information about the interaction that can be used in the design of peptides with altered binding properties. The design, synthesis and binding properties of mutated peptides are discussed in Chapter 5.

Chapter 5

Peptide mutants: Crystallisation and Binding Studies

5.1 Introduction

Four peptide residues (Pro5P, Asn6P, Leu9P and Gly10P) were mutated separately in an attempt to produce a peptide with slightly reduced affinity to the mAb SV5Pk. The Fab-peptide interactions are summarised in Fig. 1.

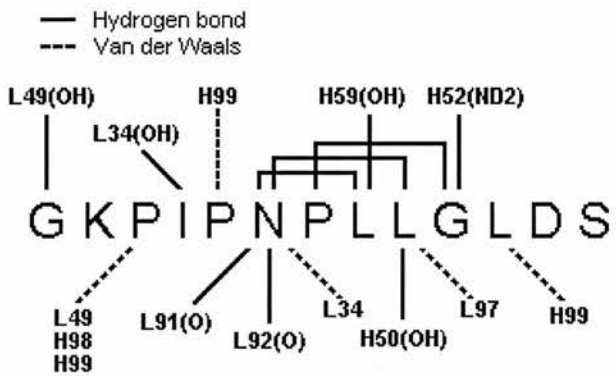


Fig. 1 Summary of the H-bonding and van der Waals contacts between the V5 peptide and SV5Pk

As discussed Chapter 4, Pro5P packs closely against a tyrosine residue from CDR H3 and has previously been highlighted as an important residue in SV5-Pk binding. It was decided to substitute this residue with a glycine to disrupt its van der Waals interactions with the Fab molecule. Asn6P is also an important residue, forming four of the ten Fab-peptide and intrapeptide hydrogen bonds. This amino acid was mutated to alanine in an attempt to remove the two peptide-Fab bonds made by the asparagine side chain. Leu9P, present in the 3_{10} helix portion of the peptide was mutated to an alanine. Gly10P was substituted by a proline residue. Such a mutation would be expected to disrupt two hydrogen bonds in addition to distorting the helix.

5.2 Site directed mutagenesis

The four mutated peptides (Pro5Gly, Asn6Ala, Leu9Ala and Gly10Pro) were produced by site-directed mutagenesis of the V5 peptide. The vector, pGEX-2T(Fig. 2), containing the V5 sequence (kindly provided by A. Diassiti of the University of St Andrews) encodes a thrombin cleavage site between a *Schistosoma japonica* glutathione S-transferase (GST) domain and the cloning site. GST-fused proteins may be affinity purified using glutathione beads. A thrombin cleavage step is then carried out followed by a further purification step to separate the target protein from GST.

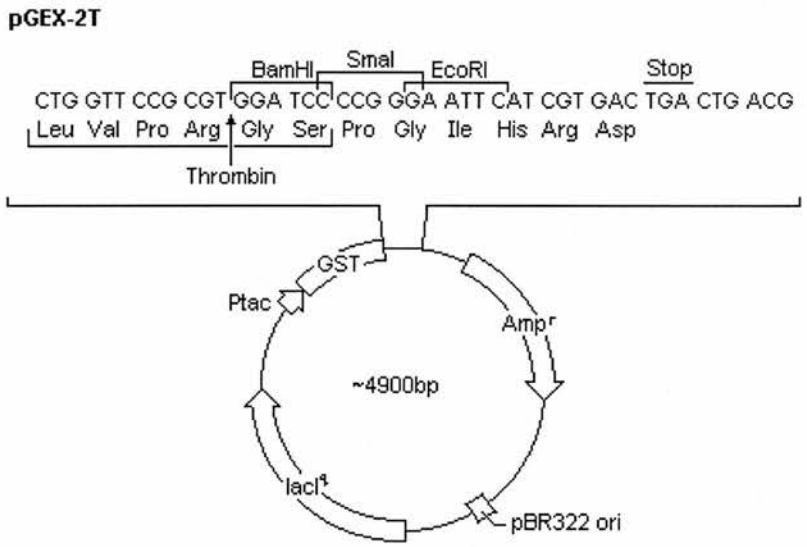


Fig. 2 The pGEX-2T vector with the cloning site shown in detail.

The vector, transformed into B834 *E. coli* cells was purified from a 500 ml culture as described previously. For each mutant, a complementary pair of primers was designed comprising the mutation plus 12 to 18 bases either side. The sequences of the primers are listed in Table 1.

Original Pk Sequence	
5`-GGA AAG CCG ATC CCA AAC CCT TTG CTG GGA TTG GAC TCC ACC-3`	
Pro5Gly	
Forward	5`-GGA AAG CCG ATC GGA AAC CCT TTG CTG GGA TTG-3`
Reverse	5`-CAA TCC CAG CAA AGG GTT TCC GAT CGG CTT TCC-3`
Asn6Ala	
Forward	5`-GGA AAG CCG ATC CCA GCC CCT TTG CTG GGA TTG-3`
Reverse	5`-CAA TCC CAG CAA AGG GGC TGG GAT CGG CTT TCC-3`
Leu9Ala	
Forward	5`-ATC CCA AAC CCT TTG GCG GGA TTG GAC TCC ACC-3`
Reverse	5`-GGT GGA GTC CAA TCC CGC CAA AGG GTT TGG GAT-3`
Gly10Pro	
Forward	5`-ATC CCA AAC CCT TTG CTG CCA TTG GAC TCC ACC-3`
Reverse	5`-GGT GGA GTC CAA TGG CAG CAA AGG GTT TGG GAT-3`

Table 1. Primers used in the construction of mutant peptides. Mutated residues are underlined.

A PCR reaction was carried out to denature the plasmid DNA, anneal the primers, elongate the DNA and amplify the mutated vector. 1 µl of each primer was added to 1 µl purified plasmid DNA, 1 µl dNTPs, 1 µl PFU polymerase (Promega) and 5 µl PFU polymerase buffer. The reaction was made up to 50 µl with dH_2O and the following PCR mutagenesis method was used:

- 5 min at 94 °C
- 16 cycles: 30 sec at 94 °C, 30 sec at 50 °C, 12 min at 68 °C
- 7 min at 72 °C

The 50 µl reaction mix was incubated with DpnI for 1 hr at 37 °C. Treatment with DpnI removes template DNA by preferential degradation of methylated nucleic acid.

Competent XL1-Blue cells were transformed with 3 µl of the vector and plated out on L-agar. A single colony was picked for each mutant and grown up in LB

media for plasmid DNA purification. Sequencing of the plasmid DNA extractions confirmed that each construct had incorporated the appropriate mutation.

5.3 Transformation and Expression

The peptides were expressed in an *E.coli* strain, B834. 100 µl of competent B834 cells were transformed with 5 µl vector. Colonies of transformed cells, selected for ampicillin resistance, were picked and grown in LB media for peptide expression. Cells were grown to an OD_{600} of 0.6 and induced with 50 µM IPTG for 3 hr. All peptide-GST fusion proteins, each visible as a 28 kDa band on SDS PAGE, were over-expressed as soluble protein.

5.4 GST-peptide purification

The GST fusion proteins were purified by a batch method using glutathione agarose (Sigma), according to the manufacturer's instructions. Briefly, 140 mg of glutathione-agarose beads were swollen in water and washed thoroughly. Filtered cell extract from a 500 ml culture was mixed gently with 4 ml (50% v/v) of glutathione agarose in PBS for 30 min at 4°C. After four washes with PBS, the bound protein was eluted four times with 10 mM reduced glutathione in 50 mM TrisHcl, pH 8.0. Elute fractions contained high concentrations of GST fusion protein (Fig. 3). A small quantity of higher molecular weight contaminants was present in each fusion protein solution.

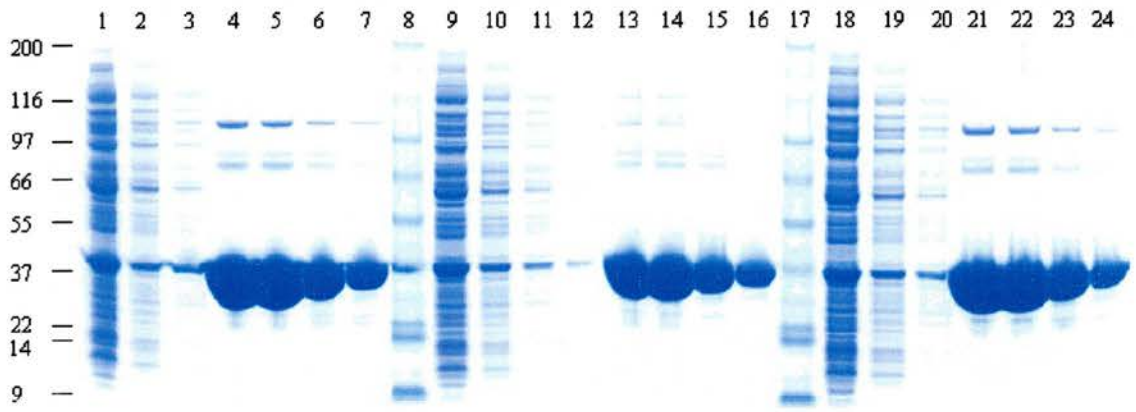


Fig. 3. GST-peptide purification. Lanes 1-3 Pro5Gly-GST wash; lanes 4-7 Pro5Gly-GST elution; lanes 8 & 17 markers; lanes 9-12 Leu9Ala-GST wash; lanes 13-16 Leu9Ala-GST elution; lanes 18-20 Gly10Pro-GST wash; lanes 21-24 Gly10Pro-GST elution.

5.5 Thrombin digest

The fusion proteins contain a thrombin cleavage site to facilitate separation of GST and peptide. 1 ml of GST-peptide containing approximately 1 mg of protein in PBS was incubated overnight at room temperature with 1 unit of thrombin. Successful digestion was confirmed on SDS PAGE (Fig. 4).

5.6 Peptide purification

The digestion products were separated using a microconcentrator with a 10 kDa molecular weight cut off. The filtrate was analysed by MALDI-TOF MS to confirm the presence of peptide (Fig. 5). Table 2 shows the mass spectroscopy results for each of the peptides. In each case the observed mass was within 0.5 Daltons of the calculated value.

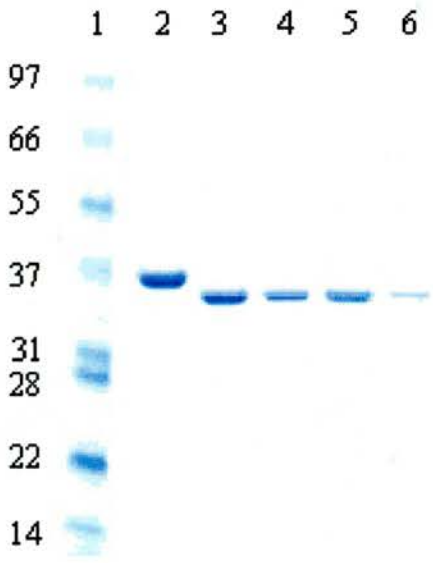


Fig 4. Thrombin digestion of GST-fused peptides. Lane 1, molecular weight markers; lane 2, non-digested control; lane 3, Pro5Gly; lane 4, Asn6Ala; lane 5, Leu9Ala; lane 6, Gly10Pro.

	Pro5Gly	Asn6Ala	Leu9Ala	Gly10Pro
Calc. MW (Da)	1890.17	1887.21	1888.15	1970.3
Obs. MW (Da)	1890.07	1887.02	1887.96	1969.80

Table 2. Calculated and observed molecular weights of the four peptides.

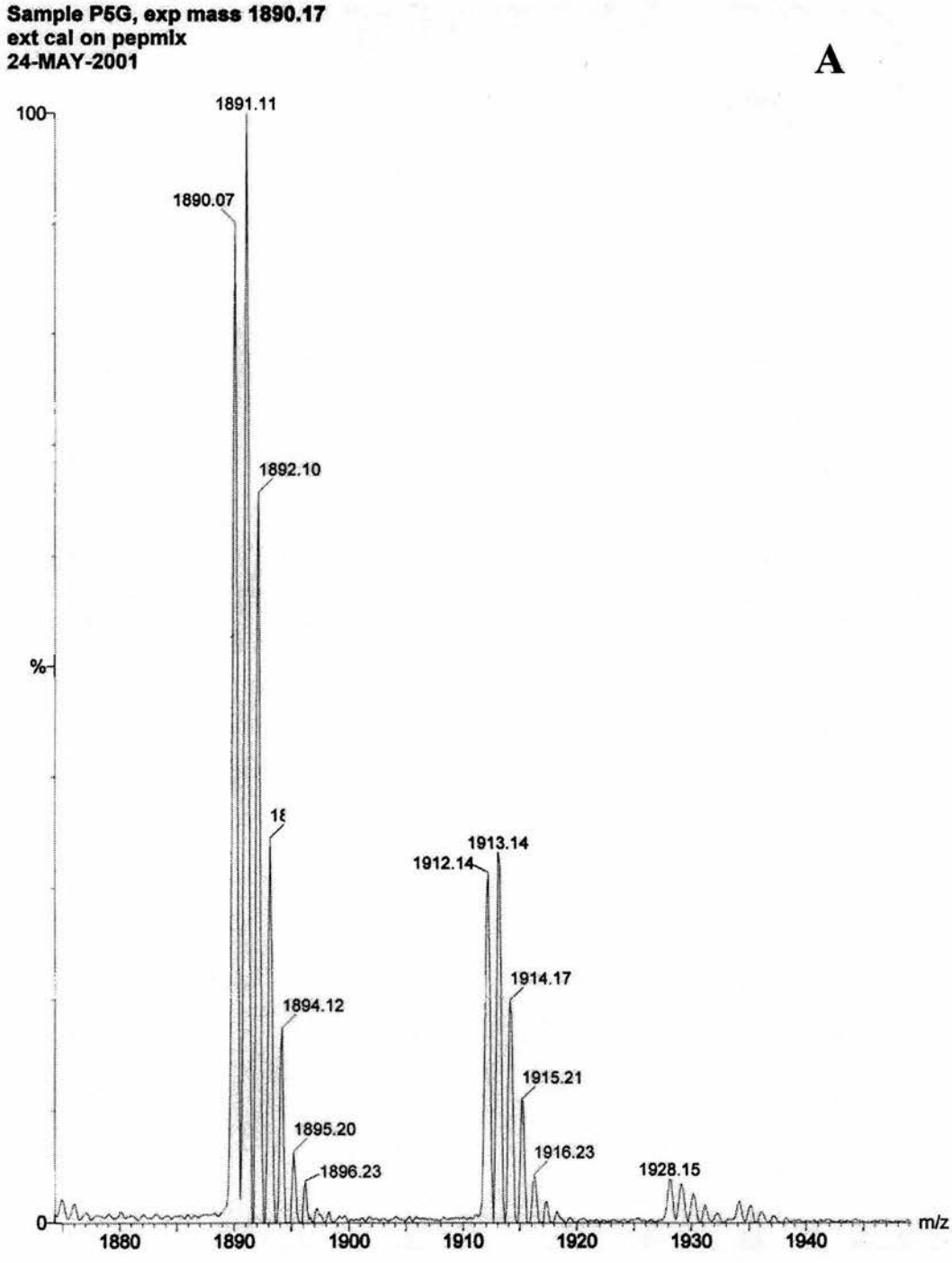
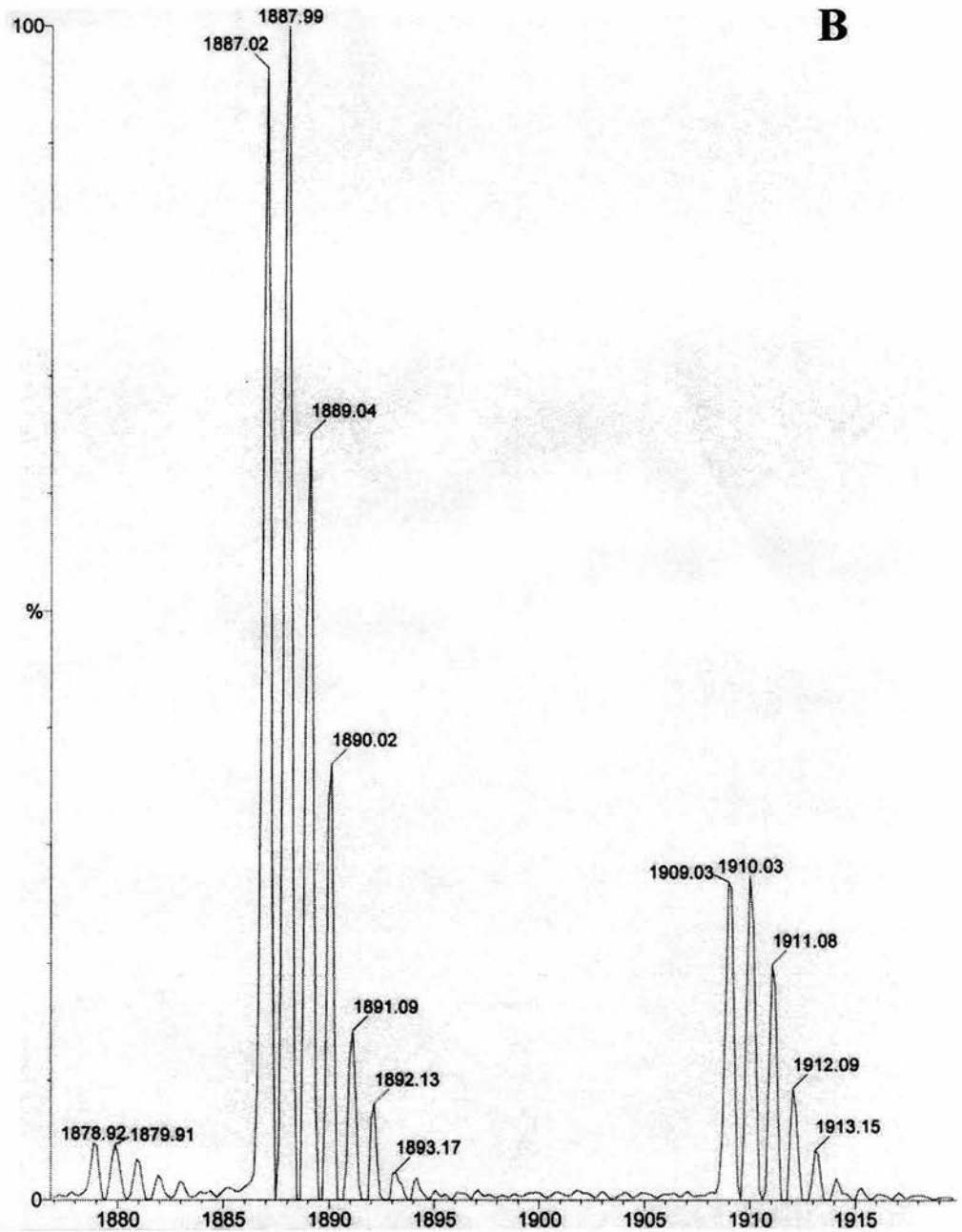
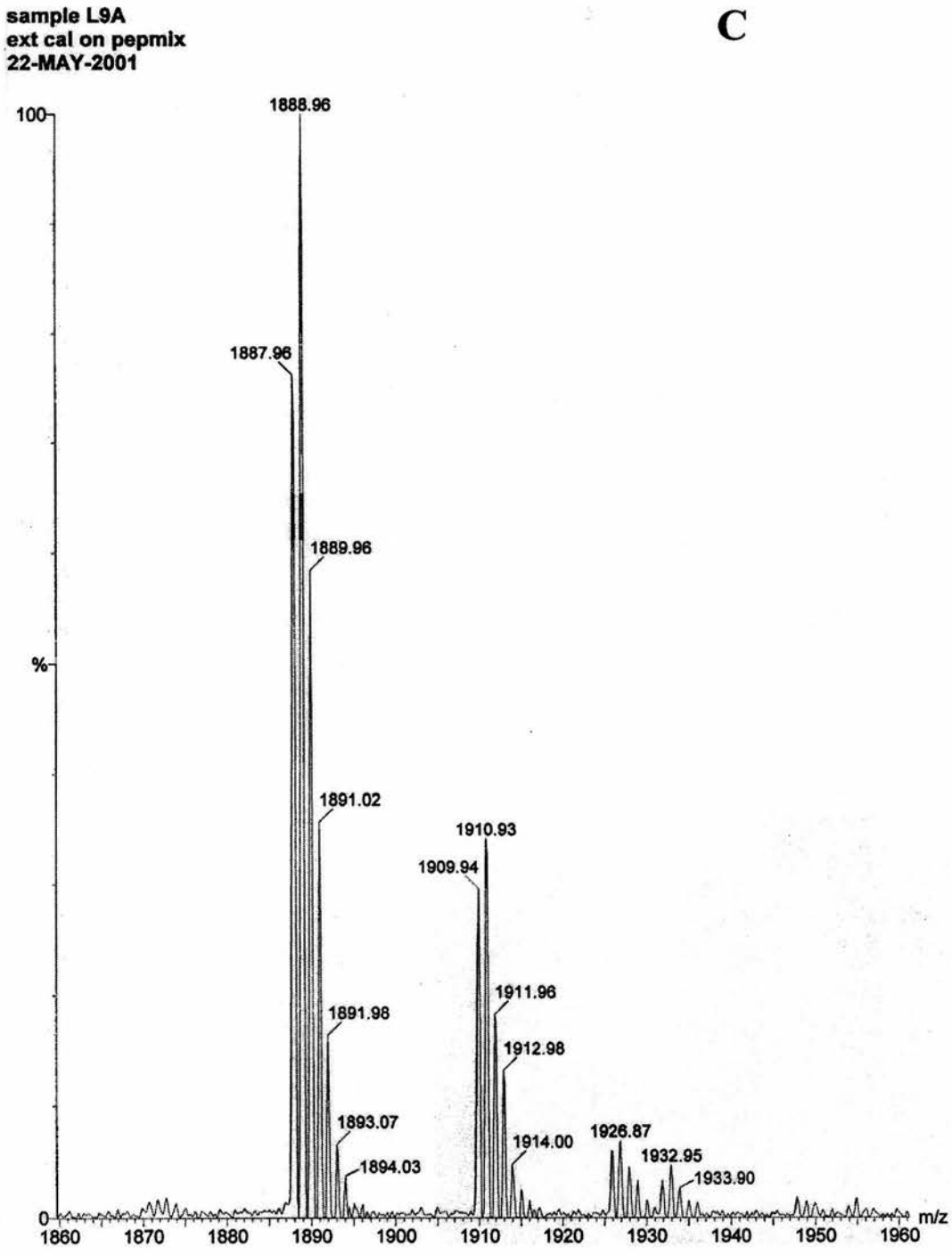
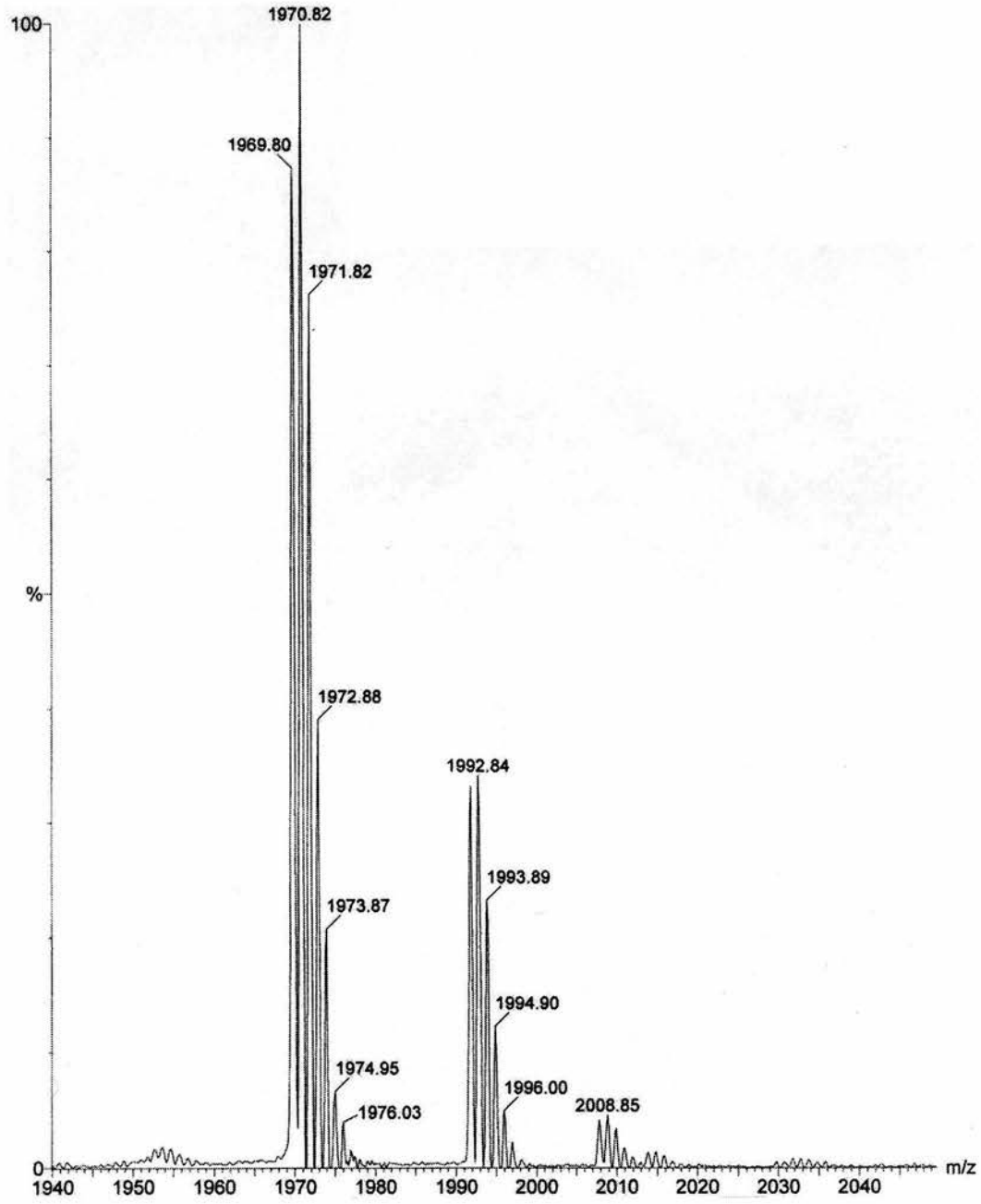


Fig. 5. Mass spectroscopy results for the purified peptides. A, Pro5Gly; B, Asn6Ala; C, Leu9Ala; D, Gly10Pro.


Sample N6A, exp mass 1887.21
ext cal on pepmix
24-MAY-2001
B
100
%
0
1887.99
1887.02
1889.04
1890.02
1891.09
1892.13
1893.17
1878.92
1879.91
1909.03
1910.03
1911.08
1912.09
1913.15
1880
1885
1890
1895
1900
1905
1910
1915


sample L9A
ext cal on pepmix
22-MAY-2001
C
100
%
0
1888.96
1887.96
1889.96
1891.02
1891.98
1893.07
1894.03
1910.93
1909.94
1911.96
1912.98
1914.00
1926.87
1932.95
1933.90
1860
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
m/z

D


Sample G10P, exp mass 1970.30
ext cal on pepmix
24-MAY-2001
100
%
0
1970.82
1969.80
1971.82
1972.88
1992.84
1993.89
1973.87
1994.90
1974.95
1996.00
2008.85
1976.03
1940
1950
1960
1970
1980
1990
2000
2010
2020
2030
2040
m/z

5.7 Fab-peptide Co-crystallisation

Attempts were made to crystallise SV5-Pk Fab in the presence of mutated peptide. For each peptide, hanging drops were set up as described in Chapter 4 using the same crystallisation conditions that provided crystals of the original complex: 0.08 M Na Acetate pH 4.6, 0.16 M Ammonium Sulphate, 20 % PEG 4K, 20 % glycerol.

Crystals were obtained for each peptide. However, these were very small and diffracted poorly. No data was collected.

5.8 Immunodetection: Dot Blot

A Western blot was carried out to give an indication of the affinity of the peptides for SV5-Pk. Purified GST-fused peptides were adjusted to a concentration of 0.1 mg/ml, 1 µl of which was applied directly onto nitrocellulose membrane and allowed to dry. The membrane was washed twice in TBS buffer (10 mM TrisHCl, pH 7.5, 150 mM NaCl) and then incubated for 1 hour at room temperature in blocking buffer (3% BSA in TBS). The membrane was washed twice in TBS-Tween/Triton buffer (20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 0.05 % Tween 20, 0.2 % Triton X-100) and once in TBS before incubating for 1 hour in a solution of SV5-Pk (0.1 µg/ml antibody in 3% BSA). After washing twice with TBS-Tween/Triton the membrane was incubated with the secondary antibody, an anti-mouse/alkaline phosphatase conjugate diluted 1 in 10,000 in 3% BSA. The membrane was stained with Western Blue (Promega) to determine which GST-fused peptides were recognised by the SV5-

Pk antibody. The results showed that Pro5Gly-GST, Asn6Ala-GST and a positive control (V5-GST) bound the antibody. A very faint dot was seen for Leu9Ala-GST and no evidence of binding was observed for the fourth peptide, Gly10Pro. A negative control, GST on its own, was not recognised by SV5-Pk.

5.9 Isothermal Titration Calorimetry

The binding affinity of the peptides for SV5-Pk was investigated by isothermal titration calorimetry (ITC). Measurements were taken using an MCS Isothermal Titration Calorimeter (MicroCal) at the University of Glasgow. The equipment can be used to measure a range of physical characteristics including affinity (K_a), heat of antigen binding (ΔH) and the number of antigen-binding sites. In ITC, successive amounts of ligand are injected into a cell containing a solution of the antibody. The heat released upon binding (ΔH) is measured. The titration is continued until the binding sites reach saturation and only heats of dilution are observed.

The antibody SV5-Pk and GST-fused peptides were dialysed extensively against PBS. Protein concentrations were determined spectrophotometrically using the Beer-Lambert law, $A=\varepsilon cl$, where A is the absorbance of the protein solution at 280 nm, ε equals the extinction coefficient, c is the concentration and l refers to the path length.

Each experiment required 2 ml of SV5-Pk antibody at a concentration of 5 μM and 500 μl of GST-fused peptide at a concentration of 200 μM. Prior to each experiment, the calorimeter was equilibrated with PBS. The antibody was added to

the sample cell and pipetted up to down to remove air bubbles. The syringe was filled with ligand solution and placed into the cell. Following equilibration, the ligand solution was titrated into the cell in approximately 25 injections of 10 μl. Data analysis was performed using the program Origin. The last three values (after saturation) were averaged and subtracted from the raw data to correct for any heat of dilution. Figure 6 shows the results of the titration of SV5-Pk antibody with the four different peptides.

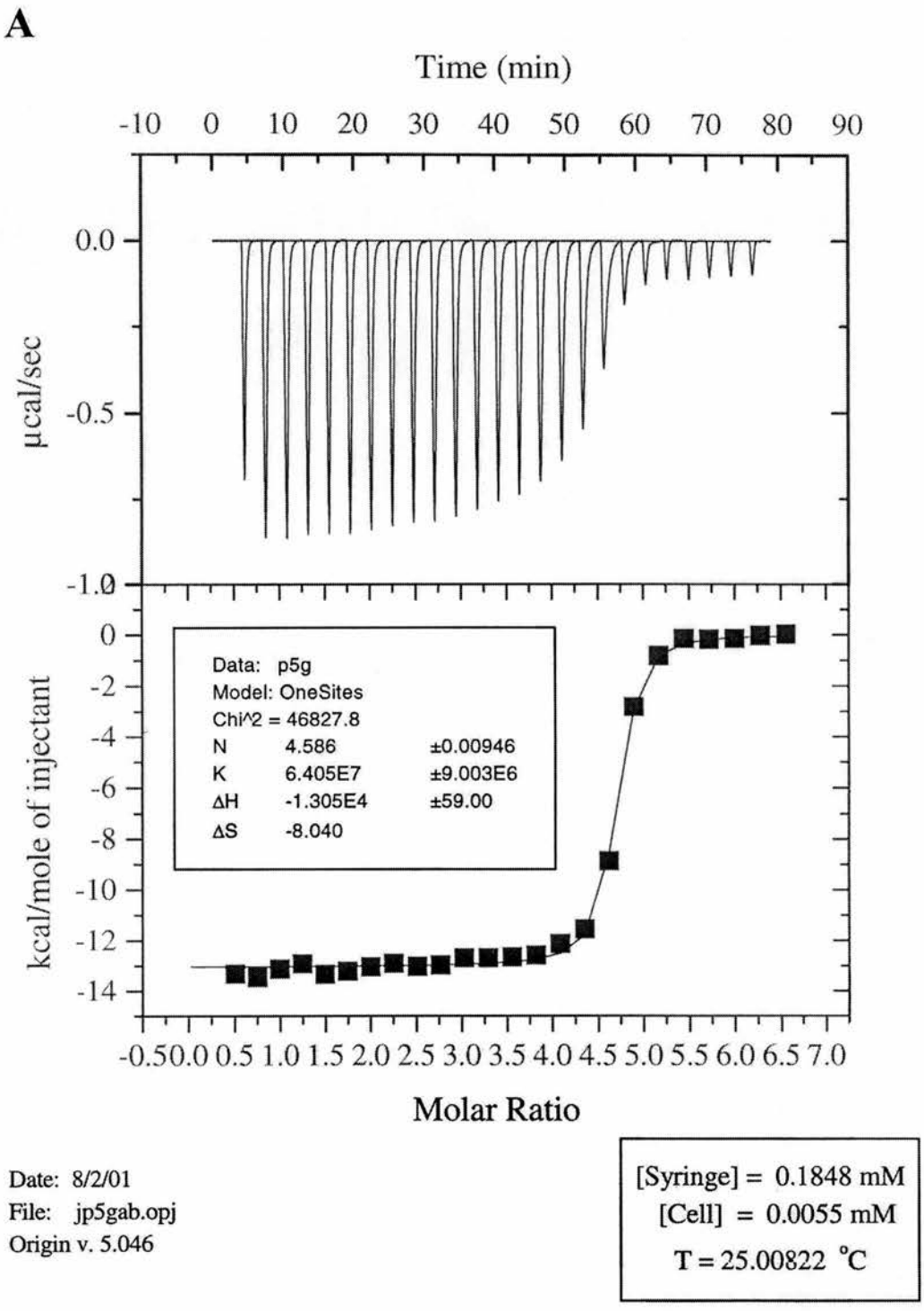
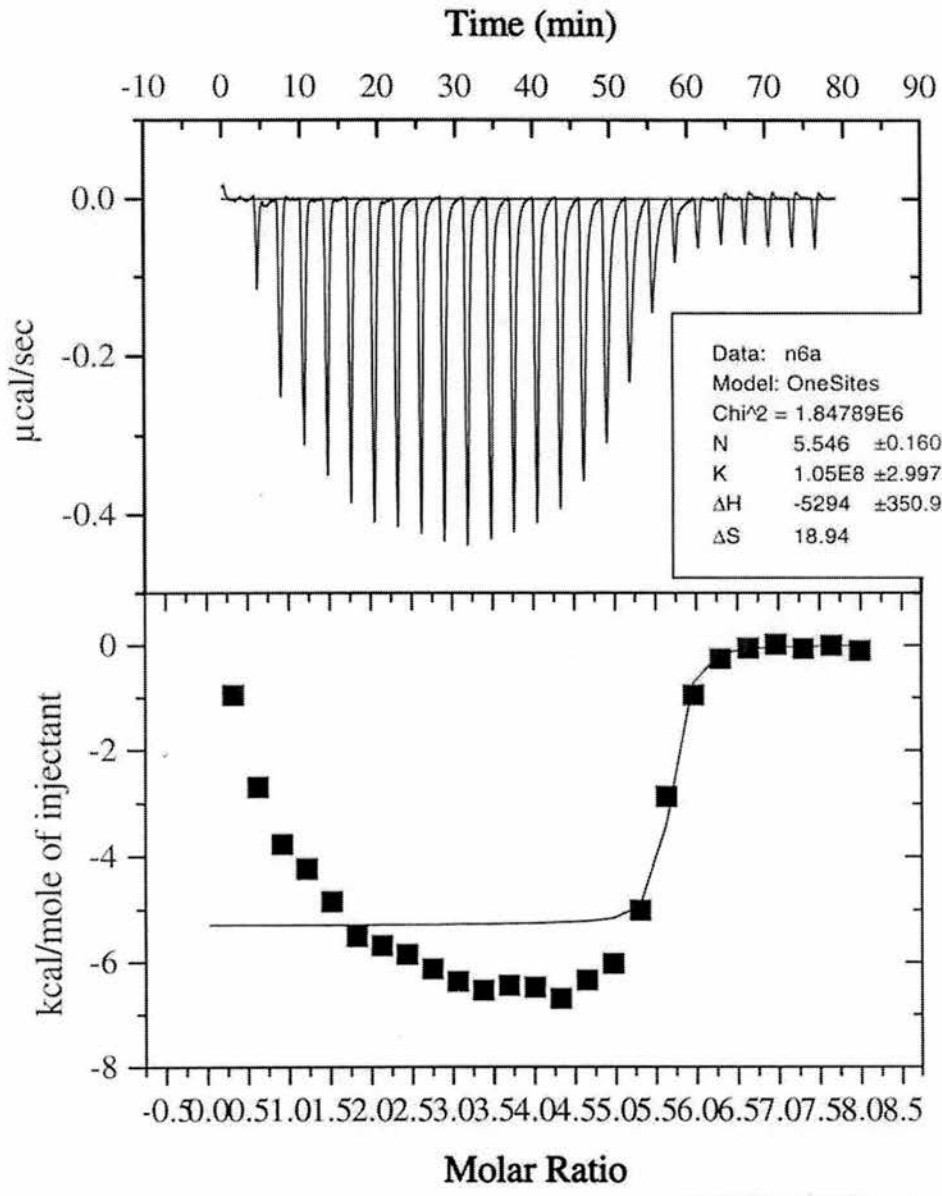


Fig.6 Isothermal titration calorimetry of the interaction between SV5-Pk antibody and (a) Pro5Gly, (b) Asn6Ala, (c) Leu9Ala, (d) Gly10Pro.

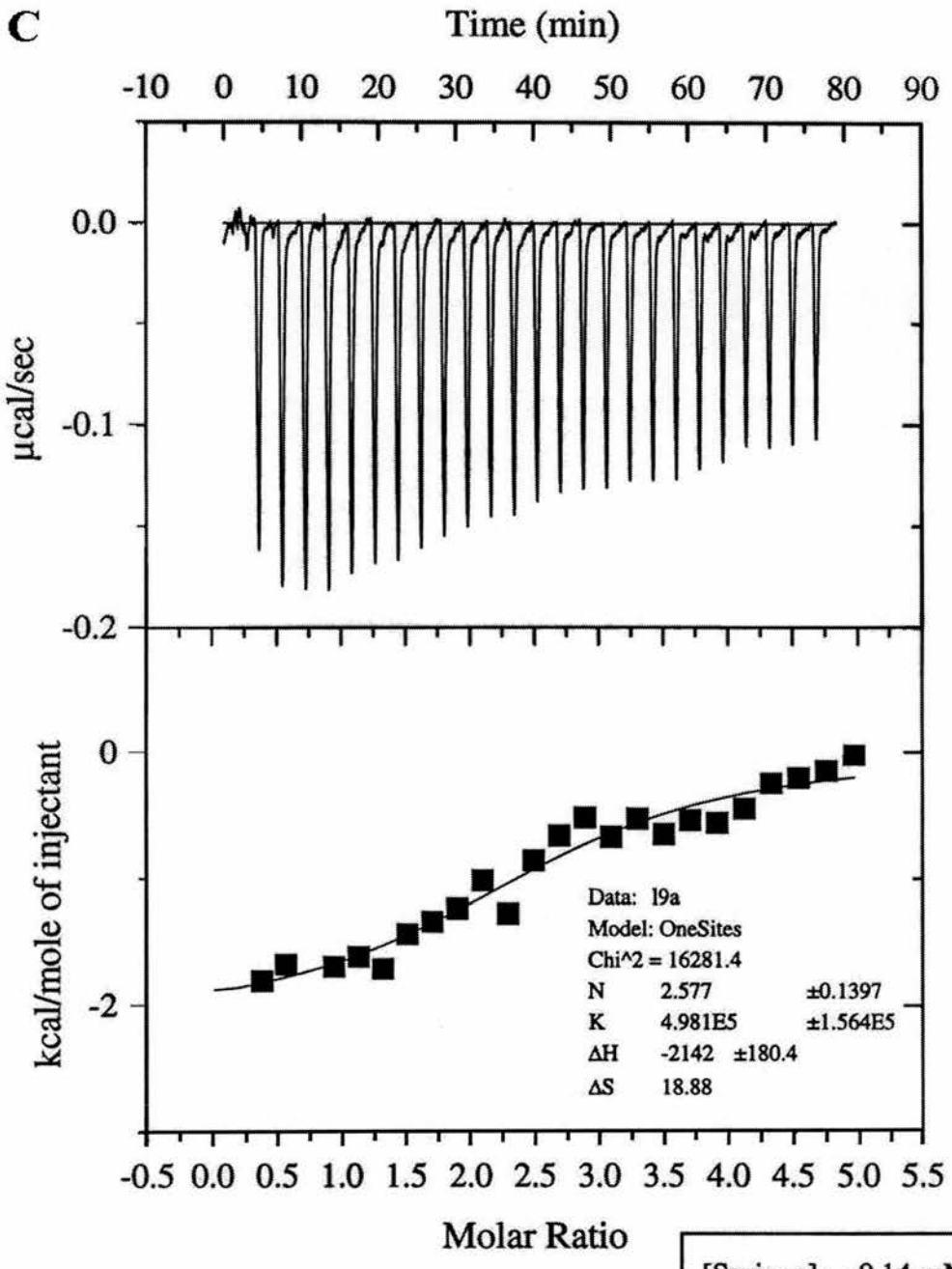
B



Date: 8/2/01
File: jn6aab.opj
Origin v. 5.046

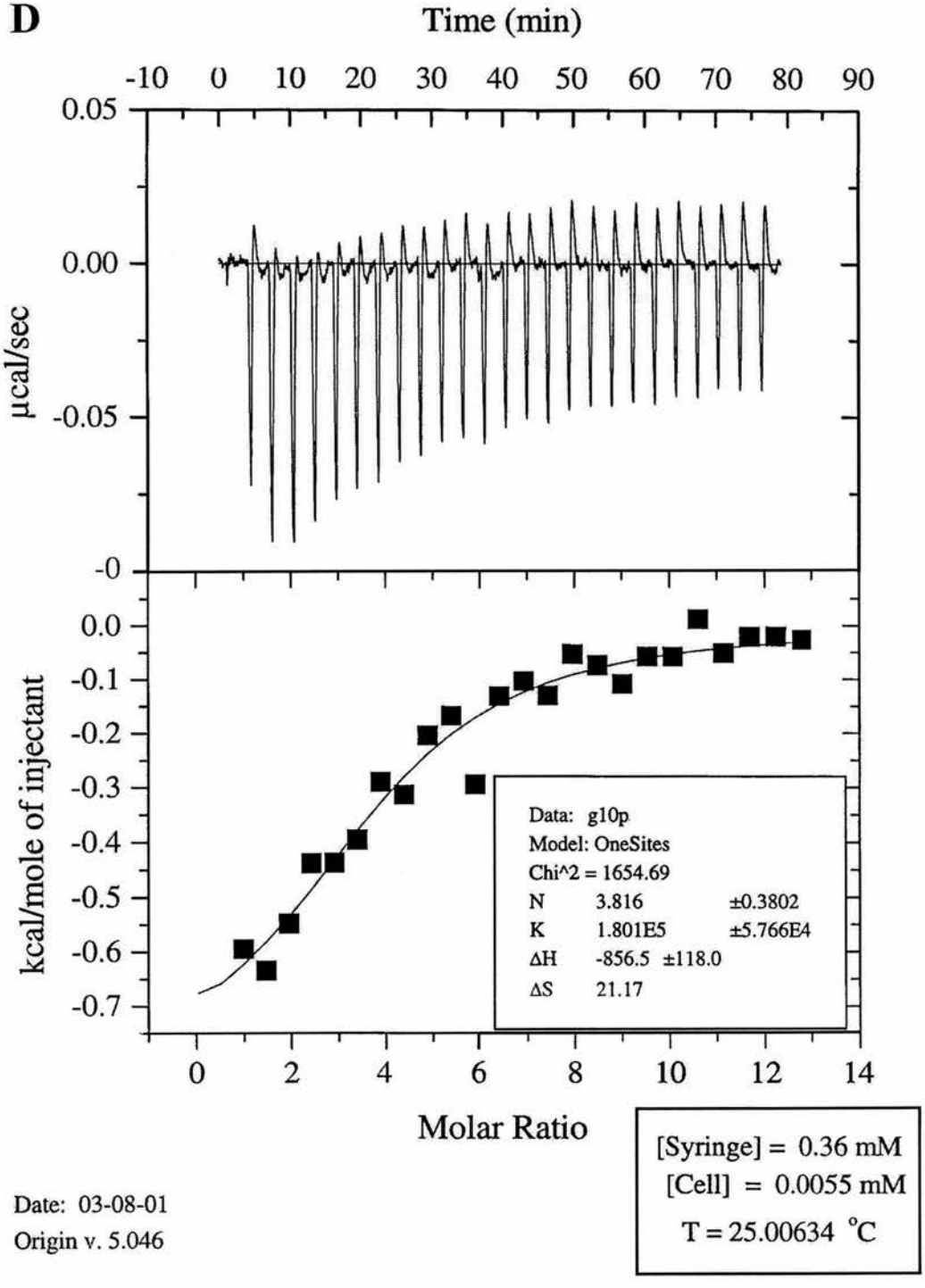
[Syringe] = 0.225 mM
[Cell] = 0.0055 mM
T = 25.00693 °C

C



[Syringe] = 0.14 mM
[Cell] = 0.0055 mM
T = 25.01245 °C

Date: 03-08-01
Origin v. 5.046


D
Time (min)
-10 0 10 20 30 40 50 60 70 80 90
0.05
0.00
-0.05
-0
µcal/sec
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
kcal/mole of injectant
Data: g10p
Model: OneSites
Chi^2 = 1654.69
N 3.816 ±0.3802
K 1.801E5 ±5.766E4
ΔH -856.5 ±118.0
ΔS 21.17
0 2 4 6 8 10 12 14
Molar Ratio
[Syringe] = 0.36 mM
[Cell] = 0.0055 mM
T = 25.00634 °C
Date: 03-08-01
Origin v. 5.046

5.10 Summary

Three of the peptides (Pro5Gly, Asn6Ala and Leu9Ala) showed SV5-Pk affinity high enough to be detected by ITC. The results for the titration with Pro5Gly-GST (Fig. 6a) show a typical pattern of heat release. Unfortunately, the experimental molar ratio (N) of ligand to antibody appears to be in the range of four to five rather than the expected two. It is possible that GST may have aggregated, complicating the binding interactions. However, although the results are not quantitative, they indicate that the mutant Pro5Gly is recognised by the antibody. Titration of SV5-Pk with Asn6Ala-GST produced a more unusual pattern of heat release which could again be a consequence of GST aggregation. As indicated in the lower half of Fig. 6b, a binding model could not be produced to fit this data. This peptide-GST fusion protein appears to bind the antibody with lower affinity than Pro5Gly-GST as suggested by the lower ΔH. Binding of the third ligand (Leu9Ala-GST) to SV5-Pk was only just detectable by ITC (Fig. 6c) and the affinity of Gly10Pro-GST for the antibody is lower again. The results are consistent with the dot blot analysis described above.

The decision to use GST-fused ligands was taken after difficulties were encountered in measuring the peptide concentrations. The peptides are too small to be measured by the standard assays and contain no aromatic residues that would allow them to be measured spectrophotoscopically. For this type of experiment it may be preferable to try to produce a large amount of peptide that could be freeze-dried and weighed directly.

The dramatic effects of the mutations at peptide residues Leu9P and Gly10P on SV5-Pk binding are consistent with the crystal structure and previous mutagenesis results. The two amino acids reside within the 3_{10} helix which may be important for

antibody recognition. In contrast, the mutation of Pro5P to glycine may exert a smaller effect by increasing flexibility in the chain and/or by removing van der Waals contacts. Asn6P makes several connections both to the antibody and within the peptide. These include side chain interactions which would not exist in the Asn6Ala mutation.

The precise effects of these mutations on Fab binding could not be confirmed crystallographically as the crystals of the complexes that were grown were of insufficient quality for data collection. The mutations may have had a large enough effect to alter the crystal growth, or the problem may lie in the difficulties in measuring exact concentrations of peptide.

The preliminary ITC experiments could not be developed further due to time constraints. However, the experiments indicate that the Pro5Gly mutant exhibits slightly reduced affinity for SV5-Pk and therefore has potential for use in tagging recombinant proteins for purification.

Appendix I - SV5 Phosphoprotein

1. Introduction

As discussed in Chapter 1, the phosphoproteins of Paramyxoviruses act as both a cofactor for the L protein and a chaperone for unassembled N protein. The phosphoprotein (P) from SV5 is 42 kDa and is thought to be oligomeric. The sequence of the phosphoprotein is shown below with the SV5Pk epitope underlined.

```
MDPTDLSFSP DEINKLIETG LNTVEYFSTQ QVTGTSSLGK NTIPPGVTGL 50
LTNAAEAKIQ ESTNHQKGSV GGGAKPKKPR PKIAIVPADD KTVPGKPIPN 100
PLLGLDSTPS TQTVLDLSGK TLPSGSYKGV KLAKFGKENL MTRFIEEPRE 150
NPIATSSPID FKRGAGIPAG SIEGSTQSDG WEMKSRSLSG AIHPVLQSPL 200
QQGDLNALVY SVQSLALNVN EILNTVRNLD SRMNQLETKV DRILSSQSLI 250
QTIKNDIVGL KAGMATLEGM ITTVKIMDPG VPSNVTVEDV RKTLSNHAVV 300
VPESFNDSFL TQSEDVISLD ELARPTATSV KKIVRKVPPQ KDLTGLKITL 350
EQLAKDCISK PKMREEYLLK INQASSEAQL IDLKKAIIRS AI          392
```

The following describes preliminary cloning and purification of SV5 P for X-ray crystallographic study.

2. Expression and Purification

The SV5 phosphoprotein construct was received in a pGEM3 vector transformed into the expression host, BL21-DE3. The protein was expressed by inducing the cells with 1 mM IPTG when the OD_{600} reached 0.4 and incubating for three hours. The cells were pelleted and resuspended in 20mM TrisHCl pH 9.2 in preparation for anion exchange. This buffer was used because SV5 P, which has a

calculated pI of 8.6, would not bind an anion exchange column at a pH lower than 9. The protein eluted from the anion exchange column at approximately 400 mM NaCl along with several, mainly lower molecular weight, contaminants.

The identity of the protein was confirmed by immunodetection with the SV5-Pk mAb. The Western blot was carried out as described in Chapter 2. SV5-Pk was used as the primary antibody, followed by the secondary antibody (rabbit-anti-mouse/alkaline phosphatase conjugate) and Western Blue stain (Promega).

As P is thought to be oligomeric, the protein was run on a gel filtration column. Unfortunately, many impurities remained, even after a further hydrophobic column step.

3. Cloning

3.1 Site-directed mutagenesis

As the SV5 phosphoprotein was not being expressed in large amounts and was proving difficult to purify to homogeneity, the P gene was sub-cloned into a vector for over-expression as a dodeca-His tagged protein. The chosen vector (C12HIS) required an insert with a 3` NdeI restriction site and a 5` BamHI restriction site. A restriction map of the entire P gene revealed BamHI cleavage sites at positions 3 and 373. The latter site was removed by site-directed mutagenesis to introduce a silent mutation such that BamHI no longer cleaved and the encoded amino acid was not altered.

The vector, pGEM3 containing the P gene was purified as described previously. A double digest with SalI and KpnI was carried out to confirm the presence of the gene, visible as a 1200 bp band on agarose gel. For the site-directed mutagenesis, a PCR reaction was carried out using a pair of complementary primers

covering the silent mutation plus 17 bases either side (5`-AAA ACA TTA CCA TCA GGG TCC TAT AAG GGG GTT AA-3` and 3`-TTA AAC CCC TTA TAG GAC CCT GAT GGT AAT GTT TT-5`).

Site directed mutagenesis PCR was carried out as follows:

- 5 min at 94 °C
- 16 cycles of: 30 sec at 94 °C, 30 sec at 50 °C, 2 min per 1 Kb plus 2 min at 68 °C
- 7 min at 72 °C

The PCR reaction products were treated with DpnI to cleave methylated (template) DNA. 1µl of DpnI was added to 50 µl of PCR and incubated for 1hr at 37 °C.

5 µl of the digest was used to transform competent XL1-Blue cells, which were then plated out on L-agar containing 50 µg/ml ampicillin. Eight colonies were picked and grown overnight in 5 ml LB for plasmid purification. Incubation of two of the plasmid preparations with BamHI, SalI and KpnI each resulted in a clear band of 1200 bp (Fig. 1, lanes 6 and 8), confirming mutation of the BamHI restriction site. A control incubation (Fig. 1, lane 9) containing non-mutated DNA resulted in two bands of 300 and 900 bp indicating that the DNA had been cleaved at the BamHI site.

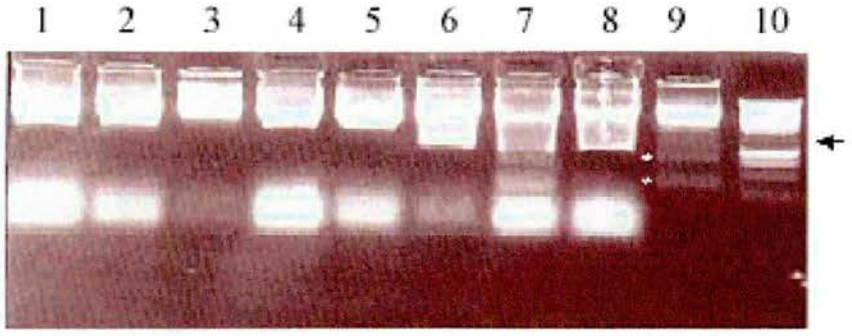


Fig. 1. Digestion of purified plasmid DNA with BamHI, SalI and KpnI. Lanes 1 to 8, digestion of plasmid DNA purified from eight colonies grown from cells transformed with mutated vector. Lane 9, control digestion of non-mutated DNA resulting in two bands of 300 bp and 900 bp (indicated by white asterisks). Lane 10, 1Kb DNA ladder. Arrow indicates 1200 bp bands present in lanes 6 and 8.

3.2 Cloning into the C12HIS vector

The following primers were used to amplify the P gene and incorporate the 3` AseI and 5` XbaI restriction sites (Forward: 5`-GCG CTC ATT AAT ATG GAC CCC ACT-3`, Reverse: 3`-GGT CGA CGC GGA TCC TCA AAT TGC-5`). In addition, the forward primer introduced a silent mutation that removed the BamHI site at position 3. AseI was chosen for the 3` site as digestion with this enzyme results in the same overhang sequence as NdeI digestion.

A PCR was carried out and the reaction products were gel purified as described previously and digested with 2 µl AseI, 2 µl BamHI and 1.5 µl buffer 2 (New England Biolabs). 1 µl of the C12HIS vector was digested with 1 µl BamHI, 1 µl NdeI, 1 µl buffer and 6 µl H_20 for 3 hours at 37°C.

The digestion products were gel purified and ligated with digested C12HIS as described previously. Competent XL1Blue cells were transformed with the ligated DNA ('C12HIS-SV5P') and plated out on L-agar containing 50 μg/ml ampicillin.

The four colonies that grew overnight were picked and grown in LB media for plasmid purification. Subsequent digestion with XbaI and BamHI indicated that three of the colonies contained the correct construct (fig. 2).

4. Expression of SV5P-12HIS

1 μl of DNA from colony number 1 (fig. 2) was used to transform the expression host, BL21-Star (DE3) (Invitrogen). Overnight cultures of six of the colonies were used to inoculate 25 ml LB media containing ampicillin. When the OD_{600} reached 0.4, cultures were induced with 0.5 mM IPTG and samples were removed at 0, 1, 2 and 3 hours post-induction. SDS PAGE analysis indicated the presence of a highly expressed protein of the expected weight (45 kDa).

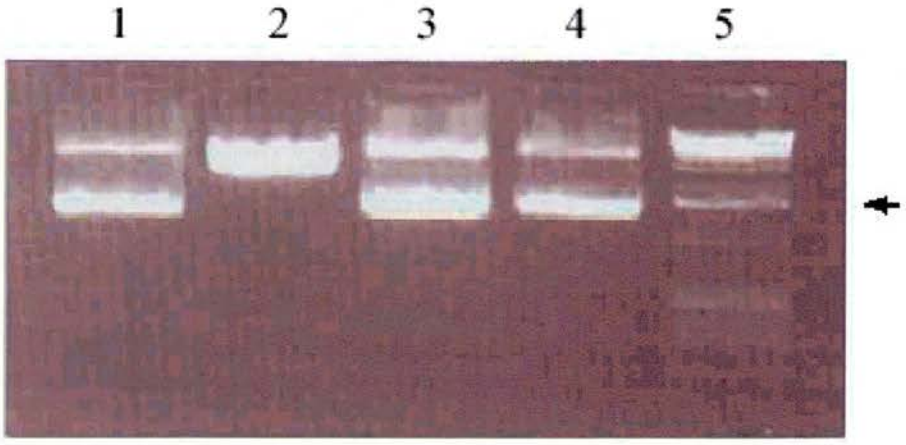


Fig. 2. Lanes 1-4, XbaI/BamHI digestion of plasmid DNA purified from four colonies of XL1-Blue cells transformed with C12HIS-SV5P. Lane 5, 1 Kb DNA ladder. Arrow indicates 1200 bp bands in lanes 1, 3 and 4

5. Purification

Cleared lysate from the soluble cell extract was loaded onto a POROS MC column as described in Chapter 2. Unfortunately, very little of the P protein remained bound to the column after washing. To investigate the reasons for this, protein minipreps were carried out using Qiagen nickel spin columns in varying concentrations of urea. The spin columns were equilibrated with buffer (20 mM Tris pH 7.5, 300 mM NaCl). After addition of the cleared lysate the column was washed twice with buffer containing 10 mM imidazole. Bound protein was then eluted with 250 mM imidazole.

As fig. 3 shows, under native conditions and in the presence of 1M urea almost all the P protein is removed from the column after the first wash. However, at higher urea concentrations (5M and 8M) P is able to remain bound to the column until eluted with 250 mM imidazole.

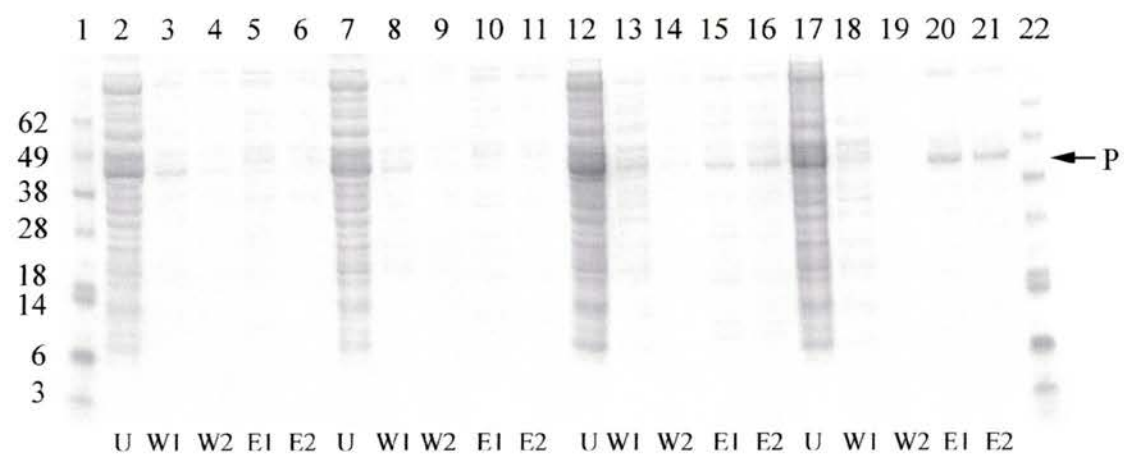


Fig. 3. Purification of SV5P on protein miniprep nickel columns at different concentrations of urea. Lanes 1 and 22, molecular weight markers. Lanes 2-6, native conditions. Lanes 7-11, 1M urea. Lanes 12-16, 5M urea. Lanes 17-21, 8M urea. U = unbound, W1 = wash 1, W2 = wash2, E1 = Elute 1, E2 = Elute 2

6. Multiple sequence alignment and secondary structure prediction

Fig. 4 shows the sequence of SV5 P aligned to five other paramyxovirus phophoproteins. The areas of predicted secondary structure, including the putative coiled coil region are also indicated.

7. Discussion

Initially, it was hoped that the interaction between P and the antibody SV5-Pk could be exploited for the purification of P. The phosphoprotein-antibody complex, isolated using a protein A column, could then be subjected to papain digestion to remove the Fc region of the antibody in preparation for crystallisation. However, preliminary incubations of P with papain indicated that the protein was susceptible to proteolytic digestion by the enzyme indicating that this approach was unsuitable.

As the original SV5 P construct could not easily be purified by anion exchange, gel filtration and hydrophobic columns, the P gene was cloned into a vector for expression as a 12-his tagged protein for metal chelate affinity purification. Unfortunately, it appeared that the tag was too inaccessible to bind to the nickel column. This was indicated by the behaviour of the protein in urea. As the concentration of denaturant was increased, more of the histidine tag became exposed and consequently more phosphoprotein was able to bind. Due to time constraints, scale-up of purification in urea and protein refolding could not be carried out.

Multiple sequence alignments of some paramyxovirus phosphoproteins (excluding SV5 P) have been carried out by Tarbouriech et al. (Tarbouriech, 2000a).

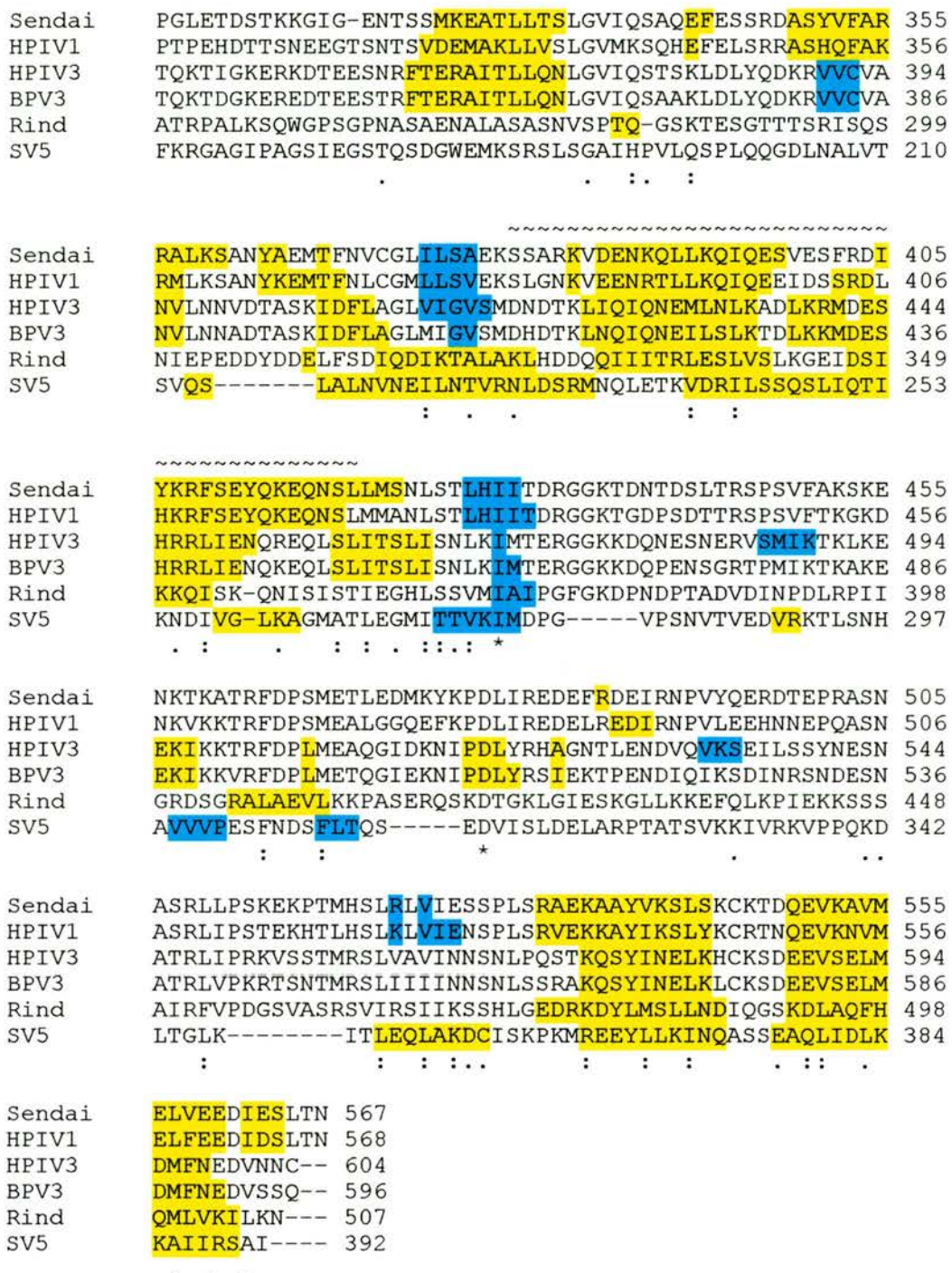


Fig. 4. Sequence alignments and secondary sequence predictions for the C-terminal regions of Sendai virus, Human Parainfluenza virus 1, Human Parainfluenza virus 3, Bovine Parainfluenza virus 3, Rinderpest and SV5 Phosphoproteins. The multiple alignment was carried out with ClustalW 1.6 and the secondary structure prediction with PHD {Rost, 1996 #124}. Regions of predicted α-helix and β-sheet are highlighted in yellow and blue respectively. The predicted coiled coil region is indicated by ~.

To ascertain the likely positions of functional domains in SV5 P, its sequence has been aligned to five other paramyxovirus phosphoprotein sequences (fig. 4). Although there is little sequence homology between the proteins, there are similarities in predicted secondary structure in two regions. The first region is the predicted multimerisation domain. The region of Sendai P corresponding to this domain has been shown by X-ray crystallography to form a tetrameric coiled coil (Tarbouriech, 2000b). As fig. 4 shows, SV5 P is also predicted to form a coiled coil in this domain. The second part of the phosphoprotein with similar predicted secondary structure is located towards the C-terminus. This domain is thought to bind to the N protein of the N:RNA template (Ryan, 1991). The phosphoproteins in fig.4, including SV5 P, are predicted to be α-helical in this region.

Appendix II - Crystallography theory

1. Crystallisation

In the most common method of growing protein crystals, the purified polypeptide is added to a solution containing a precipitant at a concentration just below that necessary to precipitate the protein. Water is removed gradually by controlled evaporation. This is often achieved by vapour diffusion using, for example, the hanging drop method (Fig. 1). The concentration of precipitant in the drop is half that of the reservoir buffer and the two solutions are allowed to equilibrate by vapour diffusion. As the reservoir is much larger than the drop, the precipitant concentrations of the two solutions at equilibrium are almost identical. The concentration of precipitant in the drop is therefore maintained at that optimum for precipitation.

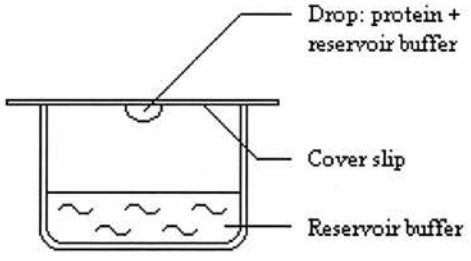


Fig. 1. Vapour diffusion: The hanging drop method.

Once crystals are obtained, the growth conditions may need to be optimised to produce better quality crystals. The many variables for growth include temperature, pH, precipitant concentration, protein concentration, drop size and protein purity.

2. Crystal mounting

Data collection is commonly carried out at low temperatures (-100°C) to help improve molecular order and decrease radiation damage to the crystal. Before mounting, the crystal is added to a cryoprotectant such as glycerol, typically for 5-30 seconds. The crystal may then be simultaneously mounted onto the goniometer and flash frozen in a stream of nitrogen gas in preparation for data collection.

3. Diffraction

The dimensions of the unit cell of a crystal is described by the lengths of the cell edges (a, b, c) and three unique angles (α, β, γ). The different types of cell are classified as shown in table 1.

Cell type	Cell dimensions	
monoclinic	$a \neq b \neq c$	$\alpha = \gamma = 90^\circ, \beta > 90^\circ$
triclinic	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma$
hexagonal	$a = b$	$\alpha = \beta = 90^\circ, \gamma = 120^\circ$
cubic	$a = b = c$	$\alpha = \beta = \gamma = 90^\circ$
tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90^\circ$
orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90^\circ$

Table 1. Dimensions of the six cell types

For the purposes of structure calculation, diffraction of X-ray beams by atoms in a crystal can be treated as reflections from sets of parallel planes of atoms. The sets of planes can be described by lattice indices (*hkl*). The *h* index refers to the number of planes per unit cell in the x direction. Similarly, *k* and *l* indicate the number of planes per unit cell in the y and z directions respectively.

Diffraction of an X-ray beam is described by several factors: the sets of planes (*hkl*), the interplanar distance (d), the wavelength of the x-ray beam (λ) and the angle

of the beam (θ). A reflection will only be produced if the value of θ satisfies Bragg's law:

$$2d_{hkl}\sin\theta = n\lambda, \text{ (where n is an integer)}$$

As illustrated in Fig. 2, the extra distance travelled by ray 2 compared to ray 1 is equal to $2d_{hkl}\sin\theta$. If this distance is equal to the wavelength multiplied by an integer, the beams will emerge from the crystal in phase and produce a strong reflection.

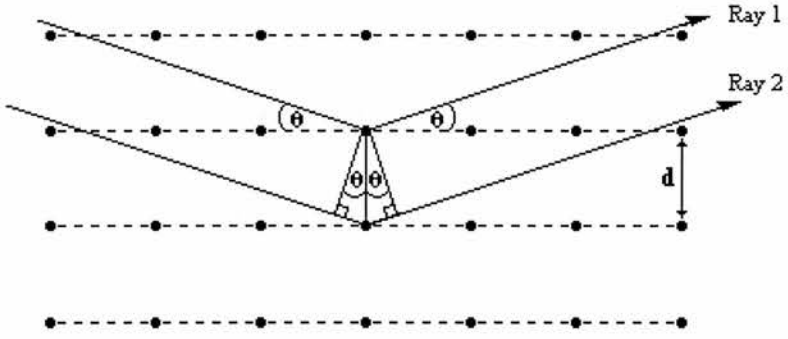


Fig. 2. Diagram demonstrating the geometry involved in Bragg's law

4. Data collection

The aim of a data collection experiment is to obtain a dataset that covers most of the theoretically possible reflections. Only the contents of the reciprocal space asymmetric unit cell are needed to reconstruct the electron density. As the reciprocal lattice spacings are the inverse of the real space lattice, a data collection strategy requires knowledge of the unit-cell dimensions.

Fig.3 describes Bragg's law in reciprocal space. The crystal (C) is at the centre of a sphere of radius $1/\lambda$. O is the origin of the reciprocal lattice. Diffraction maxima

only occur when a reciprocal lattice point (P) lies on the edge of the sphere. This condition occurs when Bragg's law is satisfied.

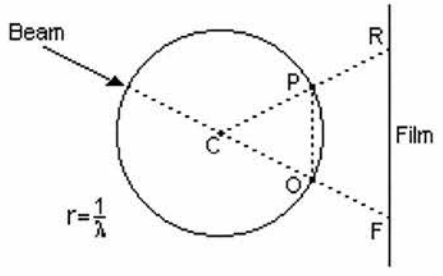


Fig 3. Unit cell dimensions are inversely proportional to reflection spacings on the detector

Fig. 3 also shows the detector film where F is the origin on the film and R is the reflection recorded on the film. It can be seen from the diagram that:

$$RF/CF = PO/CO = PO/(1/\lambda)$$

Therefore: $PO = RF/CF\lambda$

PO is equal to the reciprocal of unit cell edge a (d_{100}) giving:

$$d_{100} = CF\lambda/RF$$

For this simple example of an orthorhombic crystal, the length of the cell edge can be calculated by dividing the crystal to film distance by the distance between the reflection and the film origin and multiplying by the wavelength of the beam. Similarly, the length of the other unit cell edges can be obtained from the spacing of reflections along 0k0 and 00l.

The total angle of data collection can be further reduced by the use of symmetry information. The 230 possible space groups condense to 11 Laue groups when an additional centre of symmetry is added (a reciprocal lattice possesses the

same symmetry elements as the corresponding real space lattice, plus an extra centre of symmetry). The fraction of unique reflections expected for each Laue group can be calculated and used to reduce the amount of data collected. The symmetry is determined by certain systematic absences in the diffraction pattern that are characteristic of different space groups.

5. Scaling and processing

The collected data must be scaled to allow for differences in, for example, the diffracting power of different crystals, beam intensities and film sensitivities. This is achieved by comparing equivalent reflections from e.g. different films and rescaling those blocks of data so the equivalent reflections are of equal intensity.

6. Structure Factors and The Phase Problem

During a diffraction experiment, the intensities and positions of reflections are measured. The structure factor equation (1) specifies the amplitude $|F|$, frequency and phase (α_{hkl}) of each reflection (h,k,l).

$$F_{(h,k,l)} = \Sigma_{j=1} f_{(j)} \exp[2\pi \cdot i(hx_{(j)} + ky_{(j)} + lz_{(j)})] \qquad (1)$$

This equation is a summation over all atoms (j), the fractional co-ordinates of which are represented by x, y and z. $F_{(j)}$, the scattering factor of atom j depends on the diffraction angle of the corresponding reflection (h,k,l) and the atom type. Equation (1) allows the calculation of structure factors from a known structure. In order to carry out the reverse, i.e. determine a structure from known structure factors, a Fourier Transformation (FT) is applied to the structure factor equation to obtain the electron

density equation (2). In this equation, $F_{(h,k,l)}$ is in units of electrons and the summation is divided by the cell volume, V. The dimension of $\rho(x,y,z)$ is therefore electron density. Thus, if the structure factors are known, the actual structure (the density of electrons in real space) can be calculated.

$$\rho\,(x,y,z) = {}^{1}/_{V} \Sigma_{hkl} F_{hkl} \exp[-2\pi(hx+ky+lz)] \quad (2)$$

As mentioned above, the structure factors specify the amplitude $|F|$, frequency and phase (α_{hkl}) of each reflection. The amplitudes, proportional to $\sqrt{(I_{hkl})}$, are obtained from the diffraction data and the frequency is that of the X-ray source. However, the data do not provide phase information. Common methods of determining phases include isomorphous replacement, anomalous dispersion and molecular replacement. If the target is similar to a protein of known structure, as in the case of Fab molecules, molecular replacement is the method of choice for estimating phases.

The aim of the method is to superimpose the known structure, or phasing model, onto the target molecule by first orientating and then translating the model into the correct position. The first step, the rotation search, is achieved using the Patterson function, $P_{(u)}$, which does not require phases and uses only the intensities as Fourier coefficients:

$$P_{(u)} = {}^{1}/_{a}\Sigma |F_h| \cos(2\pi hu)$$

The Patterson map of a molecule contains peaks corresponding to inter-atomic vectors between atoms (self-rotation function) or intra-atomic vectors between equivalent molecules (cross-rotation function). During the rotation search, Patterson

maps of the phasing model in various orientations are calculated and compared with the measured map. The search is monitored by the correlation coefficient (CC). The correct solution should ideally have a high CC with a clear gap between the top CC and the lower values.

In CNS, the top ten rotation solutions are then subjected to a translation search, which is monitored by comparing the observed structure factor amplitudes (F_{obs}) with amplitudes calculated from the translated model (F_{calc}). The agreement between the two sets of amplitudes is indicated by the R-factor, defined below:

$$R=\frac{\Sigma||F_{obs} - F_{calc}||}{\Sigma|F_{obs}|}$$

7. Refinement

The first phase estimates will often be incomplete and will need to be refined before an interpretable electron density map can be produced. Refinement minimises the differences between the observed and calculated structure factors, bringing the model towards its global minimum. The starting model must be within the radius of convergence, i.e. it must be near enough to the correct structure that refinement will cause the model to converge upon the global minimum. The radius of convergence may be increased by several means, for example by constraining the occupancy, or by restraining bond angles and lengths to within a certain range of values. The radius of convergence may also be increased by energy refinement. In simulated annealing, the model is allowed to move as if at high temperature and is then slowly cooled. This may also serve to remove the model from incorrect local energy minima.

Later stages of refinement involve alternate rounds of model building and energy minimisation. In the final stages, atomic temperature factors are refined independently and water molecules are added.

8. Validation

The progress of refinement is commonly monitored using the R-factor, described above. To reduce the risk of introducing bias into the model, a second criterion, the free R-factor (R_{free}) is used. The R_{free} is calculated using a subset of the data, typically around 10% of the total, which has been excluded from the refinement. Other methods of validation evaluate the various structural parameters (e.g. bond lengths, bond angles, phi/psi angles) by comparing their values with those of known high resolution structures.

References

(1978). "Ebola haemorrhagic fever in Sudan, 1976. Report of a WHO/International Study Team." Bull World Health Organ **56**(2): 247-70.

(1990). "Update: filovirus infection in animal handlers." MMWR Morb Mortal Wkly Rep **39**(13): 221.

(1995). "From the Centers for Disease Control and Prevention. Outbreak of Ebola viral hemorrhagic fever--Zaire, 1995." Jama **273**(22): 1747-8.

Al-Lazikani, B., A. M. Lesk, et al. (1997). "Standard conformations for the canonical structures of immunoglobulins." J Mol Biol **273**(4): 927-48.

Augustine, J. G., A. de La Calle, et al. (2001). "The crystal structure of the fab fragment of the monoclonal antibody MAK33. Implications for folding and interaction with the chaperone bip." J Biol Chem **276**(5): 3287-94.

Banerjee, A. K. and S. Barik (1992). "Gene expression of vesicular stomatitis virus genome RNA." Virology **188**(2): 417-28.

Baron, R. C., J. B. McCormick, et al. (1983). "Ebola Haemorrhagic fever in southern Sudan: hospital dissemination and intrafamilial spread." Bulletin World Health Organ **6**: 997-1003.

Barr, J., P. Chambers, et al. (1991). "Sequence of the major nucleocapsid protein gene of pneumonia virus of mice: sequence comparisons suggest structural homology between nucleocapsid proteins of pneumoviruses, paramyxoviruses, rhabdoviruses and filoviruses." J Gen Virol **72**(Pt 3): 677-85.

Becker, S., S. Huppertz, et al. (1994). "The nucleoprotein of Marburg virus is phosphorylated." J Gen Virol **75**(Pt 4): 809-18.

Becker, S., C. Rinne, et al. (1998). "Interactions of Marburg virus nucleocapsid proteins." Virology **249**(2): 406-17.

Blumberg, B. M., K. Rose, et al. (1984). "Analysis of the Sendai virus M gene and protein." J Virol **52**(2): 656-63.

Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." Anal Biochem **72**: 248-54.

Bray, M., J. Driscoll, et al. (2000). "Treatment of lethal Ebola virus infection in mice with a single dose of an S-adenosyl-L-homocysteine hydrolase inhibitor." Antiviral Res **45**(2): 135-47.

Brunger, A. T. (1992). X-PLOR, Version 3.1. A system for X-ray crystallography and NMR. New Haven, Yale University Press.

Brunger, A. T., P. D. Adams, et al. (1998). "Crystallography and NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54**(Pt 5): 905-21.

Buchholz, C. J., C. Retzler, et al. (1994). "The carboxy-terminal domain of Sendai virus nucleocapsid protein is involved in complex formation between phosphoprotein and nucleocapsid- like particles." Virology **204**(2): 770-6.

Buchholz, C. J., D. Spehner, et al. (1993). "The conserved N-terminal region of Sendai virus nucleocapsid protein NP is required for nucleocapsid assembly." J Virol **67**(10): 5803-12.

Bukreyev, A., S. S. Whitehead, et al. (1997). "Recombinant respiratory syncytial virus from which the entire SH gene has been deleted grows efficiently in cell culture and exhibits site-specific attenuation in the respiratory tract of the mouse." J Virol **71**(12): 8973-82.

Bukreyev, A. A., V. E. Volchkov, et al. (1995). "The complete nucleotide sequence of the Popp (1967) strain of Marburg virus: a comparison with the Musoke (1980) strain." Arch Virol **140**(9): 1589-600.

Canter, D. M. and J. Perrault (1996). "Stabilization of vesicular stomatitis virus L polymerase protein by P protein binding: a small deletion in the C-terminal domain of L abrogates binding." Virology **219**(2): 376-86.

Chandrika, R., S. M. Horikami, et al. (1995). "Mutations in conserved domain I of the Sendai virus L polymerase protein uncouple transcription and replication." Virology **213**(2): 352-63.

Cheetham, G. M., G. Hale, et al. (1998). "Crystal structures of a rat anti-CD52 (CAMPATH-1) therapeutic antibody Fab fragment and its humanized counterpart." J Mol Biol **284**(1): 85-99.

Chenik, M., M. Schnell, et al. (1998). "Mapping the interacting domains between the rabies virus polymerase and phosphoprotein." J Virol **72**(3): 1925-30.

Chothia, C. and A. M. Lesk (1987). "Canonical structures for the hypervariable regions of immunoglobulins." J Mol Biol **196**(4): 901-17.

Collaborative Computational Project Number 4 (1994). "The CCP4 suite: Programs for Protein Crystallography." Acta Cryst. **D50**: 760-763.

Connolly, M. L. (1993). "The molecular surface package." J Mol Graph **11**(2): 139-41.

Curran, J., R. Boeck, et al. (1991). "The Sendai virus P gene expresses both an essential protein and an inhibitor of RNA synthesis by shuffling modules via mRNA editing." Embo J **10**(10): 3079-85.

Curran, J., R. Boeck, et al. (1995). "Paramyxovirus phosphoproteins form homotrimers as determined by an epitope dilution assay, via predicted coiled coils." Virology **214**(1): 139-49.

Curran, J., J. B. Marq, et al. (1992). "The Sendai virus nonstructural C proteins specifically inhibit viral mRNA synthesis." Virology **189**(2): 647-56.

Dunn, C., A. O'Dowd, et al. (1999). "Fine mapping of the binding sites of monoclonal antibodies raised against the Pk tag." J Immunol Methods **224**(1-2): 141-50.

Edman, P. (1970). "Sequence determination." Mol Biol Biochem Biophys **8**: 211-55.

Elliott, L. H., M. P. Kiley, et al. (1985). "Descriptive analysis of Ebola virus proteins." Virology **147**(1): 169-76.

Feldmann, H., E. Muhlberger, et al. (1992). "Marburg virus, a filovirus: messenger RNAs, gene order, and regulatory elements of the replication cycle." Virus Res **24**(1): 1-19.

Feldmann, H., C. Will, et al. (1991). "Glycosylation and oligomerization of the spike protein of Marburg virus." Virology **182**(1): 353-6.

Ferre-D'Amare, A. R. and S. K. Burley (1994). "Use of dynamic light scattering to assess crystallizability of macromolecules and macromolecular assemblies." Structure **2**(5): 357-9.

Fields, B. N., D. M. Knipe, et al., Eds. (1996). Fundamental Virology. Philadelphia, Raven Publishers.

Fooks, A. R., J. R. Stephenson, et al. (1993). "Measles virus nucleocapsid protein expressed in insect cells assembles into nucleocapsid-like structures." J Gen Virol **74**(Pt 7): 1439-44.

Fraenkel-Conrat, H. and W. R.R., Eds. (1975). Comprehensive Virology. New York.

Geisbert, T. W. and P. B. Jahrling (1995). "Differentiation of filoviruses by electron microscopy." Virus Res **39**(2-3): 129-50.

Geyer, H., C. Will, et al. (1992). "Carbohydrate structure of Marburg virus glycoprotein." Glycobiology **2**(4): 299-312.

Hanke, T., C. Botting, et al. (1994). "Expression and purification of nonglycosylated SIV proteins, and their use in induction and detection of SIV-specific immune responses." AIDS Res Hum Retroviruses **10**(6): 665-74.

Horikami, S. M., J. Curran, et al. (1992). "Complexes of Sendai virus NP-P and P-L proteins are required for defective interfering particle genome replication in vitro." J Virol **66**(8): 4901-8.

Horikami, S. M., S. Smallwood, et al. (1994). "An amino-proximal domain of the L protein binds to the P protein in the measles virus RNA polymerase complex." Virology **205**(2): 540-5.

Huggins, J., Z. X. Zhang, et al. (1999). "Antiviral drug therapy of filovirus infections: S-adenosylhomocysteine hydrolase inhibitors inhibit Ebola virus in vitro and in a lethal mouse model." J Infect Dis **179 Suppl 1**: S240-7.

Huggins, J. W. (1989). "Prospects for treatment of viral hemorrhagic fevers with ribavirin, a broad-spectrum antiviral drug." Rev Infect Dis **11 Suppl 4**: S750-61.

Hutchinson, E. G. and J. M. Thornton (1996). "PROMOTIF--a program to identify and analyze structural motifs in proteins." Protein Sci **5**(2): 212-20.

ICTV (1991). "The order Mononegavirales. Paramyxovirus Study Group of the Vertebrate Subcommittee." Arch Virol **117**: 137-140.

ICTV (1996). International Committee on Taxonomy of Viruses. Xth International Congress of Virology, Jerusalem.

Iseni, F., A. Barge, et al. (1998). "Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures." J Gen Virol **79**(Pt 12): 2909-19.

Jahrling, P. B., T. W. Geisbert, et al. (1999). "Evaluation of immune globulin and recombinant interferon-alpha2b for treatment of experimental Ebola virus infections." J Infect Dis **179 Suppl 1**: S224-34.

Jones, S. and J. M. Thornton (1996). "Principles of protein-protein interactions." Proc Natl Acad Sci U S A **93**(1): 13-20.

Jones, T. A., J. Y. Zou, et al. (1991). "Improved methods for binding protein models in electron density maps and the location of errors in these models." Acta Crystallogr A **47**(Pt 2): 110-9.

Kamer, G. and P. Argos (1984). "Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses." Nucleic Acids Res **12**(18): 7269-82.

Kiley, M. P., E. T. Bowen, et al. (1982). "Filoviridae: a taxonomic home for Marburg and Ebola viruses?" Intervirology **18**(1-2): 24-32.

Kiley, M. P., N. J. Cox, et al. (1988). "Physicochemical properties of Marburg virus: evidence for three distinct virus strains and their relationship to Ebola virus." J Gen Virol **69**(Pt 8): 1957-67.

Kiley, M. P., R. L. Regnery, et al. (1980). "Ebola virus: identification of virion structural proteins." J Gen Virol **49**(2): 333-41.

Kiley, M. P., J. Wilusz, et al. (1986). "Conservation of the 3' terminal nucleotide sequences of Ebola and Marburg virus." Virology **149**(2): 251-4.

Kohler, G. and C. Milstein (1975). "Continuous cultures of fused cells secreting antibody of predefined specificity." Nature **256**(5517): 495-7.

Kolesnikova, L., E. Muhlberger, et al. (2000). "Ultrastructural organization of recombinant Marburg virus nucleoprotein: comparison with Marburg virus inclusions." J Virol **74**(8): 3899-904.

Kraulis, P. J. (1991). "MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures." Journal of Applied Crystallography **24**: 946-950.

Kyte, J. and R. F. Doolittle (1982). "A simple method for displaying the hydropathic character of a protein." J Mol Biol **157**(1): 105-32.

Laskowski, R. A., M. W. MacArthur, et al. (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." Journal of Applied Crystallography **26**: 283-291.

Lawrence, M. C. and P. M. Colman (1993). "Shape complementarity at protein/protein interfaces." J Mol Biol **234**(4): 946-50.

Le Guenno, B., P. Formentry, et al. (1995). "Isolation and partial characterisation of a new strain of Ebola virus [see comments]." Lancet **345**(8960): 1271-4.

Lee, B. and F. M. Richards (1971). "The interpretation of protein structures: estimation of static accessibility." J Mol Biol **55**(3): 379-400.

Li, Y., H. Li, et al. (2000). "Three-dimensional structures of the free and antigen-bound Fab from monoclonal antilysozyme antibody HyHEL-63(,)." Biochemistry **39**(21): 6296-309.

Malashkevich, V. N., B. J. Schneider, et al. (1999). "Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9-A resolution." Proc Natl Acad Sci U S A **96**(6): 2662-7.

Martin, A. C. and J. M. Thornton (1996). "Structural families in loops of homologous proteins: automatic classification, modelling and application to antibodies." J Mol Biol **263**(5): 800-15.

Matthews, B. W. (1968). "Solvent content of protein crystals." J Mol Biol **33**(2): 491-7.

Mountcastle, W. E., R. W. Compans, et al. (1974). "Proteolytic cleavage of subunits of the nucleocapsid of the paramyxovirus simian virus 5." J Virol **14**(5): 1253-61.

Muhlberger, E., A. Sanchez, et al. (1992). "The nucleotide sequence of the L gene of Marburg virus, a filovirus: homologies with paramyxoviruses and rhabdoviruses." Virology **187**(2): 534-47.

Mupapa, K., M. Massamba, et al. (1999). "Treatment of Ebola hemorrhagic fever with blood transfusions from convalescent patients. International Scientific and Technical Committee." J Infect Dis **179 Suppl 1**: S18-23.

Murphy, F. A., G. van der Groen, et al. (1978). Ebola virus Hemorrhagic Fever. Amsterdam, Elsevier.

Navaza, J. (1994). "AMoRe: an automated package for molecular replacement." Acta Crystallographica Section A **A50**: 157-163.

Nicholls, A. and B. Honig (1991). "GRASP: a program for the graphical representation and analysis of surface properties." J. Comp. Chem **12**: 435-445.

Oliva, B., P. A. Bates, et al. (1998). "Automated classification of antibody complementarity determining region 3 of the heavy chain (H3) loops into canonical forms and its application to protein structure prediction." J Mol Biol **279**(5): 1193-210.

Otwinowski, Z. and W. Minor (1997). "Processing of X-ray diffraction data collected in oscillation mode." Methods in Enzymology **276**: 307-326.

Parks, G. D. (1994). "Mapping of a region of the paramyxovirus L protein required for the formation of a stable complex with the viral phosphoprotein P." J Virol **68**(8): 4862-72.

Paterson, R. G., T. J. Harris, et al. (1984). "Analysis and gene assignment of mRNAs of a paramyxovirus, simian virus 5." Virology **138**(2): 310-23.

Peters, C. J. and A. S. Khan (1999). "Filovirus diseases." Curr Top Microbiol Immunol **235**: 85-95.

Peters, D., G. Muller, et al. (1971). Marburg Virus Disease. New York, Springer-verlag.

Randall, R. E. and D. F. Young (1991). "Solid matrix-antibody-antigen complexes induce antigen-specific CD8+ cells that clear a persistent paramyxovirus infection." J Virol **65**(2): 719-26.

Randall, R. E., D. F. Young, et al. (1987). "Isolation and characterization of monoclonal antibodies to simian virus 5 and their use in revealing antigenic differences between human, canine and simian isolates." J Gen Virol **68 (Pt 11)**: 2769-80.

Regnery, R. L., K. M. Johnson, et al. (1980). "Virion nucleic acid of Ebola virus." J Virol **36**(2): 465-9.

Rost, B. (1996). "PHD: predicting one-dimensional protein structure by profile-based neural networks." Methods Enzymol **266**: 525-39.

Ryan, K. W., A. Portner, et al. (1991). c Two noncontiguous regions of sendai virus P protein combine to form a single nucleocapsid binding domain. " Virology **180**: 126-134.

Ryan, K. W., A. Portner, et al. (1993). "Antibodies to paramyxovirus nucleoproteins define regions important for immunogenicity and nucleocapsid assembly." Virology **193**(1): 376-84.

Sanchez, A., M. P. Kiley, et al. (1989). "The nucleoprotein gene of Ebola virus: cloning, sequencing, and in vitro expression." Virology **170**(1): 81-91.

Sanchez, A., M. P. Kiley, et al. (1992). "Sequence analysis of the Marburg virus nucleoprotein gene: comparison to Ebola virus and other non-segmented negative-strand RNA viruses." J Gen Virol **73**(Pt 2): 347-57.

Schnell, M. J. and K. K. Conzelmann (1995). "Polymerase activity of in vitro mutated rabies virus L protein." Virology **214**(2): 522-30.

Schnittler, H. J., F. Mahner, et al. (1993). "Replication of Marburg virus in human endothelial cells. A possible mechanism for the development of viral hemorrhagic disease." J Clin Invest **91**(4): 1301-9.

Sleat, D. E. and A. K. Banerjee (1993). "Transcriptional activity and mutational analysis of recombinant vesicular stomatitis virus RNA polymerase." J Virol **67**(3): 1334-9.

Southern, J. A., D. F. Young, et al. (1991). "Identification of an epitope on the P and V proteins of simian virus 5 that distinguishes between two isolates with different biological characteristics." J Gen Virol **72**(Pt 7)(2): 1551-7.

Spehner, D., A. Kirn, et al. (1991). "Assembly of nucleocapsidlike structures in animal cells infected with a vaccinia virus recombinant encoding the measles virus nucleoprotein." J Virol **65**(11): 6296-300.

Tarbouriech, N., J. Curran, et al. (2000a). "Tetrameric coiled coil domain of Sendai virus phosphoprotein." Nat Struct Biol **7**(9): 777-81.

Tarbouriech, N., J. Curran, et al. (2000b). " On the domain structure and the polymerization state of the sendai virus P protein." Virology **266**(1): 99-109

Thomas, S. M., R. A. Lamb, et al. (1988). "Two mRNAs that differ by two nontemplated nucleotides encode the amino coterminal proteins P and V of the paramyxovirus SV5." Cell **54**(6): 891-902.

Vargas-Madrazo, E., F. Lara-Ochoa, et al. (1995). "Canonical structure repertoire of the antigen-binding site of immunoglobulins suggests strong geometrical restrictions associated to the mechanism of immune recognition." J Mol Biol **254**(3): 497-504.

Vidal, S., J. Curran, et al. (1990). "Editing of the Sendai virus P/C mRNA by G insertion occurs during mRNA synthesis via a virus-encoded activity." J Virol **64**(1): 239-46.

Volchkov, V. E., V. M. Blinov, et al. (1993). The full length nucleotide sequence of the Ebola virus. IXth International Congress of Virology, Glasgow, Scotland.

Volchkov, V. E., H. Feldmann, et al. (1998). "Processing of the Ebola virus glycoprotein by the proprotein convertase furin." Proc Natl Acad Sci U S A **95**(10): 5762-7.

Will, C., E. Muhlberger, et al. (1993). "Marburg virus gene 4 encodes the virion membrane protein, a type I transmembrane glycoprotein." J Virol **67**(3): 1203-10.

Wilson, I. A. and R. L. Stanfield (1994). "Antibody-antigen interactions: new structures and new conformational changes." Curr Opin Struct Biol **4**(6): 857-67.

World Health Organization (1985). "Arthropod-Borne and Rodent-Borne Viral Diseases." WHO Technical Report Series 719.

Wu, T. T. and E. A. Kabat (1970). "An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity." J Exp Med **132**(2): 211-50.

Young, D. F., L. Didcock, et al. (2000). "Paramyxoviridae use distinct virus-specific mechanisms to circumvent the interferon response." Virology **269**(2): 383-90.