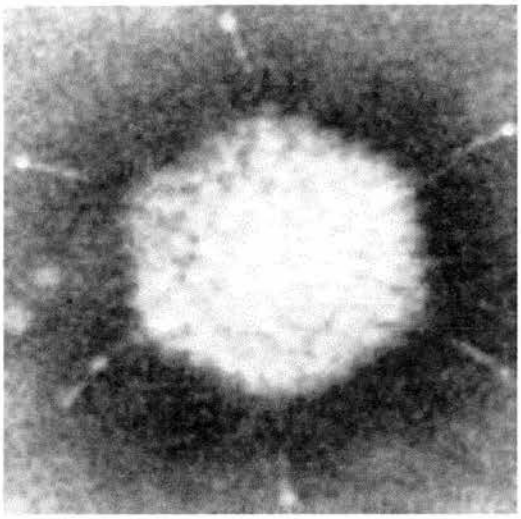
Characterisation of the Structure-Function Relationship of Adenovirus Protease



By Natalia Lynn Fowlis
University of St. Andrews
St. Andrews

A thesis submitted for the degree of Doctor of Philosophy at the University of St. Andrews, August 2004

This thesis is dedicated to my parents,
for their constant support and encouragement,
and for teaching me never to give up.

CONTENTS

DECLARATIONS

I Natalia Lynn Fowlis, hereby certify that this thesis, which is approximately 50,000 words in length, has been written by me, that it is a record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date 5/8/04 Signature of candidate..

I was admitted as a research student in October 2000 and as a candidate for the degree of Doctor of Philosophy in October 2001; the higher study for which this is a record was carried out in the University of St. Andrews between 2000 and 2004.

Date 5/8/04 Signature of candidate.

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St. Andrews and that the candidate is qualified to submit the thesis in application for that degree.

Date 5/8/04 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 5/8/04 Signature of candidate..

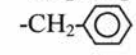
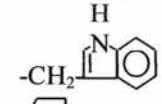
ABBREVIATIONS

Abs	Absorbance
ACQ. File	Acquisition File
Avp2	Human Adenovirus Serotype 2 Protease
Avp2-Mal	Avp2-Maltose Tag Fusion Protein
BFF.	Braided Format File
Ca_2Cl	Calcium Chloride
CAR	Coxsackie and Adenovirus Receptor
CE	Capillary Electrophoresis
CM-Sepharose	Carboxymethyl-Sepharose
CTL	$CD8^+$ cytotoxic cell
DBP	DNA-binding protein
DEAE-Sepharose	Diethylaminoethyl-Sepharose
D-MEM	Dulbecco's modified Eagle's medium
DNA	Deoxyribonucleic Acid
DNase	Deoxyribonuclease
dsDNA	Double-Stranded Deoxyribonucleic Acid
E. coli	*Escherichia coli*
EDS	Egg Drop Syndrome
EDSp	Egg Drop Syndrome Protease
EDTA	Ethylenediaminetetracetate
EM	Electron Microscope
ER	Endoplasmic Reticulum
GON Hexon	Group of Nine Hexon
GST	Glutathione-*S*-transferase
HC-Ad	High-Capacity Adenovirus
HCl	Hydrochloric acid
HEPES	N-2-hydroxyethylpiperazine-N'-2ethanesulfonic acid
HRP	Horseradish peroxidase
HS	Heparin-Sepharose
HSA	Human Serum Albumin
5-IAF	5-iodoacetamidofluorescein
ICNV	International Committee on Nomenclature of Viruses
ICTV	International Committee on Taxonomy of Viruses
IEC	Ion Exchange Chromatography
IPTG	Isopropyl-β-D-thiogalactopyranose
ITR	Inverted Terminal Repeat
LB	Luria Bertani

MCS	Multiple Cloning Site
MgCl	Magnesium Chloride
$MgSO_4$	Magnesium Sulphate
MHC	Major Histocompatibility Complex
MOWSE score	Molecular weight score
NaCl	Sodium Chloride
NEB	New England Biolabs
NES	Nuclear Export Signal
NF	Nuclear Factor
NLS	Nuclear Localisation Signal
NPC	Nuclear Pore Complex
OD	Optical Density
ori	Origin of replication
PBS	Phosphate Buffer Saline
pol	DNA polymerase
pTP	pre-terminal protein
PVDF	Polyvinyl Difluoride
pVIcT	C-terminal 11 Residues of pVI (Activating Peptide)
SBTI	Soyabean Trypsin Inhibitor
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
TE	Tris-ethylenediaminetetracetate
TEMED	N,N,N',N'-tetramethylethylenediamine
TFA	Trifluoroacetic Acid
TMV	Tobacco Mosaic Virus
TP	Terminal Protein
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer Ribonucleic Acid
ts1	Temperature sensitive 1
TTBS	Tween phosphate buffer saline
WT	Wild-type

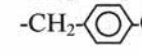
AMINO ACID R-CHAINS

HYDROPHOBIC AMINO ACIDS

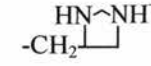
Glycine	Gly	G	-H
Alanine	Ala	A	$-CH_3$
Valine	Val	V	$-CH(CH_3)-CH_3$
Leucine	Leu	L	$-CH_2-CH(CH_3)-CH_3$
Isoleucine	Iso	I	$-CH(CH_3)-CH_2-CH_3$
Methionine	Met	M	$-CH_2-CH_2-S-CH_3$
Phenyalanine	Phe	F	
Tryptophan	Trp	W	
Proline	Pro	P	

POLAR AMINO ACIDS

Neutral

Serine	Ser	S	$-CH_2-OH$
Threonine	Thr	T	$-CH(OH)-CH_3$
Cysteine	Cys	C	$-CH_2-SH$
Asparagine	Asn	N	$-CH_2-CONH_2$
Glutamine	Gln	Q	$-CH_2-CH_2-CONH_2$
Tyrosine	Tyr	Y	

Positive Charge

Lysine	Lys	K	$-CH_2-CH_2-CH_2-CH_2-NH_3^+$
Arginine	Arg	R	$-CH_2-CH_2-CH_2-NH-C(NH_2)=NH_2^+$
Histidine	His	H	

Negative Charge

Aspartic Acid	Asp	D	$-CH_2-COO^-$
Glutamic Acid	Glu	E	$-CH_2-CH_2-COO^-$

GENETIC CODE

TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys
TTC	Phe	TCC	Ser	TAC	Tyr	TGC	Cys
TTA	Leu	TCA	Ser	TAA	Stop	TGA	Stop
TTG	Leu	TCG	Ser	TAG	Stop	TGG	Stop
CTT	Leu	CCT	Pro	CAT	His	CGT	Arg
CTC	Leu	CCC	Pro	CAC	His	CGC	Arg
CTA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CTG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser
ATC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
ATA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
ATG	Met*	ACG	Thr	AAG	Lys	AGG	Arg
GTT	Val	GCT	Ala	GTA	Asp	GGT	Gly
GTC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GTA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GTG	Val	GCG	Ala	GAG	Glu	GGG	Gly

*ATG is a start codon as well as being the codon for methionine.

ACKNOWLEDGEMENTS

I would firstly like to thank my supervisor, Dr. Graham Kemp for his guidance and endless patience during the practical and written elements of my PhD. I would also like to thank him for encouraging me throughout the Researchers in Residence Programme.

I wish to thank all of the students and staff of the BMS who helped me during my PhD.

Outwith St. Andrews I would like to thank all of my family and friends for their support during my PhD .

I am especially grateful to my parents, my brother Liam and my boyfriend Jamie for the immeasurable support, patience and love they have given me throughout my PhD.

This work was supported by the BBSRC, to whom I am indebted.

ABSTRACT

Human adenovirus serotype 2 cysteine protease (Avp2) is essential to the viral life cycle and is activated by an eleven residue peptide, pVIcT. In this work, I used bioinformatic and mutational analysis to investigate residues of the protease putatively involved in binding both the N-terminus of the activating peptide and Avp2 substrates. The results of this analysis suggested that Asp142 may be involved in binding the N-terminus of pVIcT, whilst residues Ser3, Glu5, Val53 and Trp55 were essential for substrate binding.

In addition to this work, the potential function of the C-terminus of Avp2 was investigated, as was a purification method for Egg Drop Syndrome Protease (EDSp). Subsequently, a number of tagged vector systems were investigated to purify Avp2 and EDSp.

Tagged vector systems were also used to purify the second adenoviral protein investigated in this work, the structural protein pVIII. Bioinformatic studies suggested that human Ad2 pVIII could be cleaved at potentially three distinct sites between residues 111-112, 131-132 and 157-158. Hence a pVIII-GFP fusion protein expressed in a mammalian *in vitro* transcription-translation system was incubated with Avp2 and pVIcT. The results of this suggested that the three sites could be cleaved by Avp2. Subsequently, purified virus was separated on an SDS-PAGE mini-gel and a band between 6-14K was N-terminal sequenced. Together, the results of the sequencing and the size of the band strongly suggested that mature VIII was formed by residues 132-227.

CHAPTER 1
INTRODUCTION AND AIMS

This Chapter describes background information to viruses, specifically Adenoviridae. It also discusses the current knowledge relating to the two adenoviral proteins investigated in this work; the protease and polypeptide VIII. Following this, the aims of the thesis are provided.

1.1 VIRUSES

A virus is an infectious agent which is capable of reproducing within a host cell. However outside of its host environment it is completely inert. A mature virus (virion) consists of a DNA or RNA core surrounded by a protein coat (capsid). Some viruses also have an envelope of lipids and proteins surrounding the capsid (enveloped virus). This envelope is acquired as the virus buds through host cell membrane. This sub-chapter provides a general background to the world of viruses, from their first discovery in the late 1800's as the causative agents of many diseases, to their taxonomic classification in the 21st century.

1.1.1 Discovery

The effects of viral transmission were first investigated by Adolf Mayer in 1879. He noted that when the leaves of a plant infected with tobacco mosaic virus (TMV) were ground up and the sap removed, the liquid extract could be used to infect healthy plants. Mayer made no conclusions as what the causative agent was. However, further experiments by him showed that the pathogen was neither of bacterial or fungal origin, as these could not be cultured from the sap (reviewed in Levine., 1996).

Three years later in 1892 Dimitri Ivanofsky repeated Mayers experiment, except after removing the sap he filtered it through a Chamberland Filter. This was a method of the time commonly used to remove bacteria. Therefore Ivanofskys results suggested that a pathogen distinct from bacteria caused TMV. Nevertheless even by 1903, when Ivanofsky published his thesis, he feared that the infection was caused by bacteria which he simply could not culture (reviewed in Levine., 1996).

Unaware of Ivanofskys work, Martinus Beijernick repeated the filtration experiment in 1898. Thus, it was Beijernick who made the suggestion that an agent other than bacteria caused tobacco mosaic disease. Following this conclusion, Beijernick went on to show that this pathogen could only replicate in its host cell and not in the cell free environment of the sap. This explained why the pathogen could not be cultured outside of its host (reviewed in Levine., 1996).

Together Mayer, Ivanofsky and Beijernicks insight led to the proposal of a new agent of infection. A filterable agent which was so small it could not be visualised using a light microscope, but which could cause devastating diseases by invading and replicating in host cells. The agent would later be aptly named virus, which is Latin for "poison".

Following the discovery of viruses, the early part of the 20th century was dedicated to viral purification. It was believed that viruses were proteins and therefore in 1929

Vinson and Petre purified TMV from the sap of infected plants. They showed that the virus sample could be separated in an electrical field, just like proteins, and this provided evidence that the virus was protein based (reviewed in Levine., 1996).

Once purification procedures were set in place, the first crystals of TMV were produced by Wendell Stanley in 1935 (reviewed in Levine., 1996). By 1936, Bawden and Pirie suggested that these crystals contained 0.5% Phosphorous and 5% RNA (reviewed in Levine., 1996). Therefore, these protein-RNA crystals were then used in x-ray crystallography to show that the virus was a rod-like shape (reviewed in Levine., 1996).

With the advent of electron microscopes, a TMV electron micrograph taken by Kausche, Ankuch and Ruska in 1939 confirmed the rod shape of TMV. Electron microscopy lead to elucidation of the structures of many other virus particles. All of the particles appeared to be spherical or helical in nature, and this distinct morphology paved the way for the first rational classification of viruses.

1.1.2 Taxonomy

As discussed in section 1.1.1, viruses were first classified as being agents which could be filtered through a Chamberland filter, could not be seen using a light microscope and yet had the ability to cause disease by replicating in host cells.

By the 1930's, electron microscopy provided evidence of the structure and composition of viruses and therefore these factors began to be used to classify viruses. However the large numbers of groups working in this area led to confusion. To overcome this problem the International Committee on Nomenclature of Viruses (ICNV) was established in 1966. The committees goal was to produce a single universal taxonomic scheme.

In 1973, the committee became the International Committee on Taxonomy of Viruses (ICTV). Today it is composed of 6 sub-committees, 45 study groups and over 400 participating virologists. The committees function is to analyse and weight certain characteristics of viruses and use these to divide viruses into Orders (suffix -virales, has only one member at present, Mononegevirales), Families (suffix -viridae), Subfamilies (suffix -virinae) and Genera (suffix -virus).

The committee however does not designate lower hierarchical levels, such as species and strain. This is the responsibility of international speciality groups. (ICTV web site - http://www.ncbi.nlm.nih.gov/ICTV/intro_to_univeral/intro_to_universal.html).

1.2 ADENOVIRIDAE

The term Adenovirus (Enders et al., 1956) describes a group of viruses first isolated in the early 1950's by two independent groups. These groups proposed that the virus could both infect human adenoidal tissue (Rowe et al., 1953) and cause acute respiratory illness in humans (Hillemann & Werner., 1954). This sub-chapter discusses Adenoviridae taxonomy, structure, life cycle, clinical effects in humans, ability to oncogenically transform cells and uses in gene therapy.

1.2.1 Taxonomy

The *Adenoviridae* family infects a number of species throughout the vertebrate kingdom, from fish to humans. The family itself was recently divided into four genera in the Seventh ICTV Report (http://www.ncbi.nlm.nih.gov/ICTVdb/index.htm). The genera were mastadenovirus, aviadenovirus, siadenovirus and atadenovirus.

As the names mastadenovirus and aviadenovirus infer, these genera infect mammalian and avian species respectively. However the siadenovirus and atadenovirus genera infect a wider host range. The siadenovirus genus has been found to infect birds and a frog, whilst atadenovirus infects mammals, birds, reptiles and even a marsupial (Possum).

Although only four genera have been accepted by the ICTV, a fifth genus has been proposed (Benko et al., 2002). To-date, this genus has only one member, a virus isolated from the fish, *Acipenser transmontanus* (White Sturgeon). Analysis of the DNA polymerase sequence from this virus suggested that the virus was distinct from the other four genera. The name proposed for this virus by Benkos' lab was ichtadenovirus but the name has not yet been approved by the ICTV (Benko & Harrach., 2003).

As expected, a vast amount of research has focused on human adenoviruses and the Seventh ICTV Report listed all human adenoviruses as being members of the mastadenovirus genus. Within this genus there are six recognised human subgroups, named human A-F and 51 serotypes, which are highlighted in Table 1.2.1.

Table 1.2.1: Human Mastadenovirus Subgroups and Serotypes
The Seventh ICTV Report listed six human species and 51 human serotypes.

Subgroup	Serotype
A	12, 18, 31
B	3, 7, 11, 14, 16, 21, 34, 35, 50
C	1, 2, 5, 6
D	8, 9, 10, 13, 15, 17, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 32, 33, 36, 37, 38, 39, 42, 43, 44, 45, 46, 47, 48, 49, 51
E	4
F	40, 41

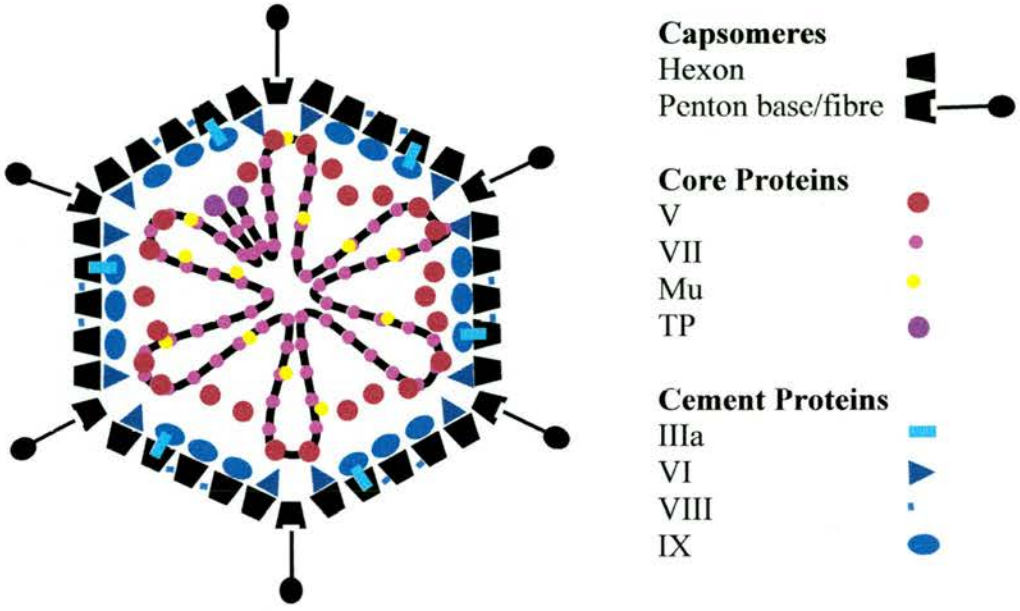
1.2.2 Structure

The first electron micrograph in 1959, by Horne et al., suggested that adenovirus serotype 5 was icosahedral in structure. It is now known that all serotypes of adenovirus are composed of a non-enveloped, icosahedral capsid, which surrounds a DNA-protein complex. The capsid is comprised of 12 pentameric-shaped structures and 240 hexameric-shaped structures, which are found at the 12 vertices and 20 facets, respectively (Valentine & Periera., 1965). The names designated to these are penton capsomeres (penton base and fibre) and hexon capsomeres (Ginsberg et al., 1966). The hexon capsomeres are subdivided into peripentonal- and Group of Nine (GON)- hexons. The capsomere subunits are linked together by an array of cement proteins (IIIa, VI, VIII, IX), whilst the viral DNA is bound by several core proteins (V, VII, TP and Mu). Together they form the mature virion particle (Figure 1.2.1 (a)).

The names given to each of the polypeptides were designated by Maizel et al, 1968. These workers separated whole virus (Roman numeral, I) on an SDS-PAGE mini-gel and designated the separated polypeptides, II-XII in order of their decreasing apparent molecular weight. The exceptions to this are X which is now named Mu and Terminal protein (TP). The molecular masses and the copy number of each of the structural proteins in the current human Ad2 virion model are given in Figure 1.2.1(b). However, XI and XII are absent from the table as they are believed to be degradation products formed by the cleavage of other viral proteins.

Figure 1.2.1: Adenovirus Particle
Adenovirus is composed of a non-enveloped, icosahedral capsid, which surrounds a DNA-protein complex. The capsid is comprised of hexon and penton capsomeres, which are linked together by cement proteins (IIIa, VI, VIII, IX). Within the capsid lies the viral DNA and several core proteins (V, VII, TP and Mu). A (a) schematic diagram of the adenovirus particle (adapted from Russell., 2000) and the molecular masses and copy number of each of the structural proteins in the current human Ad2 virion model are given in the (b) table (adapted from San Martin & Burnett., 2003).

(a) Schematic Diagram of Adenovirus



(b) Table of Components of Adenovirus
* Proteins which have not been imaged.

Polypeptide	Molecular mass (Da)	Copy number in current virion model
II (hexon)	109,077	240 trimers
III (penton base)	63,296	12 pentamers
IIIa (cement protein)	63,535	60 monomers
IV (penton fibre)	61,960	12 trimers
TP (core protein)	~55,000	2 *
V (core protein)	41,631	157 ± 1 *
VI (cement protein)	22,118	60 hexamers
VII (core protein)	19,412	833 ± 19 *
VIII (cement protein)	15,390	127 ± 3*
IX (cement protein)	14,339	80 trimers
Mu (core protein)	~4,000	~104*

1.2.2.1 Capsid Protein – Hexon

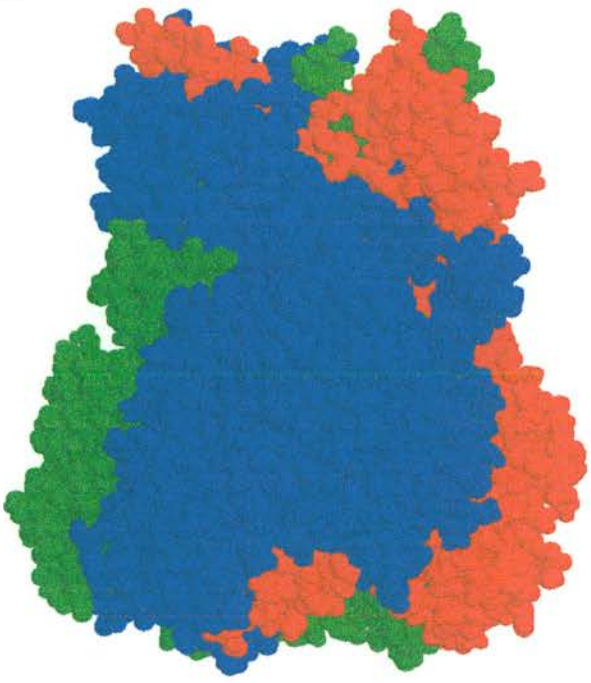
As was discussed in Section 1.2.2, the adenoviral capsid contains 252 capsomeres of which 240 are hexon capsomeres. These capsomeres are often referred to as hexons and are divided into two families depending on their placement in the virus; 5 peripentonal hexons surround each penton and 9 group of nine (GON) hexons form the facets of the virion (Figure 1.2.1; Figure 1.2.4).

The structure of hexons has been well established and they are known to be composed of three copies of protein II (Maizel et al., 1968; Horwitz et al., 1970). The structure of human Ad2 hexon was first crystallised by Pereira et al., 1968, and since this time the structure has been resolved to 6 Angstroms (Burnett et al., 1985), 2.9 Angstroms (Roberts et al., 1986; Anthappilly et al., 1994) and 2.2 Angstroms (Rux et al., 2003). Together, these structures have provided definitive proof that the three copies of protein II fit tightly together to form an N-terminal pseudo-hexagonal base and a C-terminal fibre, which projects away from the virion shell (Figure 1.2.2.).

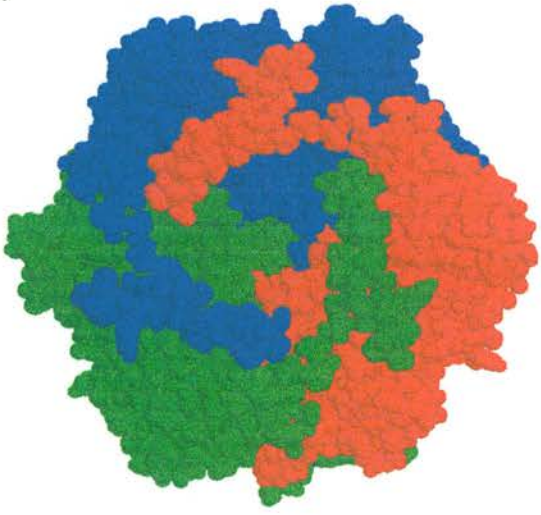
As both the well-packed structure and placement of hexons in the virion would suggest, their major function is to provide a barrier between the external environment and the inner core of the virus. However hexon may also play a role in viral entry - this will be discussed in section 1.2.3.

Figure 1.2.2: Crystal structure of Human Ad2 Hexon Trimer
The PDB file for the human Ad2 hexon trimer resolved to 2.2 Angstrom (Rux et al., 2003) was found at http://bioinfo.wistar.upenn.edu.pub/RKB_hexon_var.pdb. This figure was produced using the RASMOL version 2.6 computer package (Sayle & Milner-White, 1995). Hexon is composed of three copies of protein II, coloured blue, red and green. These proteins intertwine to form a (a) triangular shaped fibre which projects away from the virion shell. At the N-terminal foot of the fibre, lies a (b) pseudo-hexagonal base.

(a)



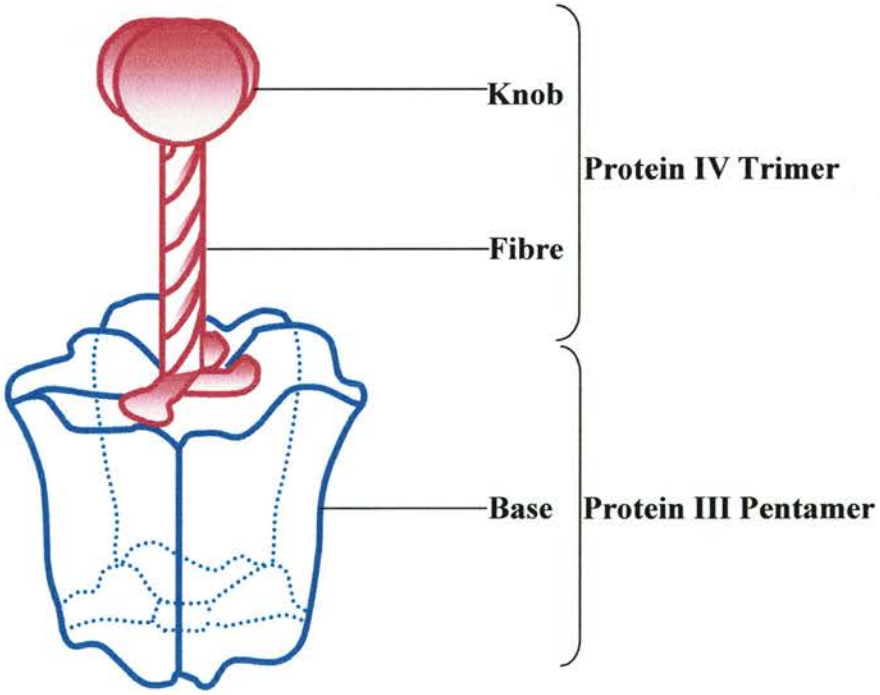
(b)



1.2.2.2 Capsid Protein – Penton Base and Fibre

Each adenovirus particle contains 12 penton capsomeres, found at the vertices of the icosahedral capsid. These capsomeres are commonly known as pentons and are composed of three domains, a penton base, fibre and knob (Figure 1.2.3). The base is comprised of five copies of protein III, whilst the fibre/knob consists of three copies of protein IV (van Oostrum & Burnett., 1985). As will be discussed in section 1.2.3, the penton knob and base play major roles in viral attachment and internalisation, respectively.

Figure 1.2.3: Schematic Diagram of Penton base, fibre and knob
Adnovirus pentons are composed of three domains, a penton base, fibre and knob. The base is comprised of five copies of protein III, whilst the fibre/knob consists of three copies of protein IV (adapted from Chroboczek et al., 2002).

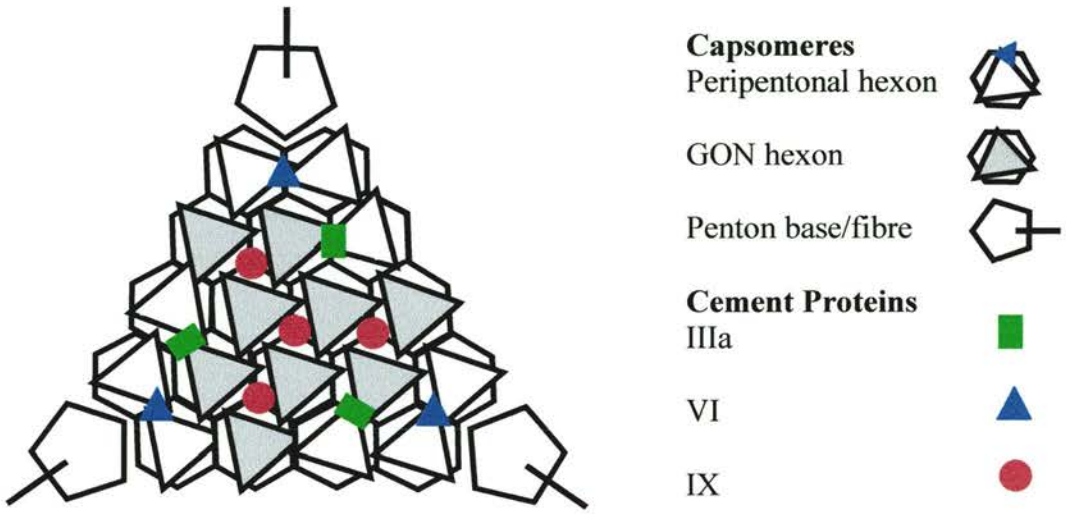


1.2.2.3 Cement Proteins – IIIa, VI, VIII, IX

The cement proteins are involved in viral stabilisation and the most recent information concerning their location in the viral capsid came from a combination of an Ad2 cryo-EM reconstruction and the crystal structure of hexon (Stewart et al., 1991; Stewart et al., 1993). The three cement proteins resolved by this method were IIIa, VI and IX (Figure 1.2.4). Protein IIIa found on the exterior of the virion is highlighted by a green rectangle in Figure 1.2.4 and is proposed to be involved in linking the facets of the virion. Wherease VI, highlighted by a blue triangles in Figure 1.2.4 is found on the interior and plays a dual role in connecting adjacent peripentonal hexons via their bases and also linking the capsid to the core (Russell & Precious., 1982). Likewise Stewart and colleagues proposed that protein VIII is found on the interior of the virion as had previously been suggested by Everitt et al., 1973. The fourth cement protein, IX has been well characterised and is found on the exterior of the virion (red circle in Figure 1.2.4).

Stoichiometric analysis of the virus first led to the proposal of a model for the localisation of IX in the viral capsid (van Oostrum & Burnett, 1985). This model suggested that the twelve copies of IX were found as trimers in the facet of the virion and that these trimers linked adjacent GON hexons on the same facet (Figure 1.2.4). This was later confirmed by a combination of images from a scanning-EM of the virus and the crystal structure of the hexon (Furcinitti et al., 1989).

Figure 1.2.4: Orientation of the Cement Proteins in the Adenoviral Capsid
The orientation of the cement proteins in the viral capsid came from a combination of an Ad2 cryo-EM reconstruction with the crystal structure of hexon. The model of the virus (adapted from Stewart et al., 1993) shows the hexon and penton capsomeres represented by hexamers with superimposed triangles and pentamers with superimposed black rods respectively. Protein IIIa is found on the exterior of the virion linking the facets together (green rectangles); protein VI is on the interior of the virion, and plays a dual role in connecting adjacent peripentonal hexons via their bases and also linking the capsid to the core (blue triangles); and protein IX is found on the exterior of the virion (red circle). The function of the twelve copies of protein IX are to form trimers (red circles) and link adjacent GON hexons on the same facet (hexamers with superimposed grey triangles).



Due to the location of IX it plays a major role in viral stability (Colby & Shenk., 1981). However, another function of IX is the activation of the adenovirus major late promoter as well as several other viral and cellular TATA-containing promoters (Lutz et al., 1997), thus the most recent work on IX has concentrated on the determination of the domains involved in transcriptional activation and in binding the capsid.

Analysis of IX by immunoelectron microscopy has shown that the C-terminal domain is on the surface of the capsid, whilst the N-terminus is found within the capsid (Akalu et al., 1999). Site-directed mutagenesis has also been employed to prove that the N-

terminal residues, 22-26 of Ad5 are essential for IX to be buried within the capsid, whilst the transcriptional activity is dependent upon the C-terminus (Rosa-Calatrava et al., 2001).

1.2.2.4 Core Proteins – V, VII, Mu, TP

The adenovirus core is composed of a single copy of double-stranded DNA and four core proteins, V, VII, Mu and terminal protein (TP). A copy of TP binds each end of the DNA strand, circularises it and functions as a primer for DNA replication (Rekosh et al., 1977). Proteins V and VII have also been shown to bind the viral DNA (Chatterjee et al., 1986). Protein VII is the major core protein and has been proven to act in a histone-like manner. Thus the DNA-VII structure has been designated the name adenosome (Mirza & Weber, 1982; Devaux et al., 1983; Vayda et al., 1983; Vayda & Flint, 1987). As well as binding DNA, protein V is also capable of attaching to the penton base and protein VI and therefore has been suggested to link the core to the capsid (Everitt et al., 1975; Chatterjee et al., 1986). However although Mu is known to bind DNA (Anderson et al., 1989), its function is yet to be elucidated.

1.2.2.5 Structural Similarity to Other Viruses

Adenovirus shares structural similarity with the dsDNA bacteriophage, PRD1 which infects the bacteria, *Escherichia coli* and *Salmonella typhimurium*. Analogous to adenovirus, PRD1 is icosahedral in structure and is composed mainly of the capsid protein, P3 which is found in a similar location to the adenovirus hexon, at 240 copies per virion (Butcher et al., 1995). Furthermore when the crystal structures of P3 (Benson et al., 1999) and hexon (Roberts et al., 1986; Anthappilly et al., 1994; Rux et al., 2003) were compared they were found to have similar folds (Benson et al., 1999). In addition to the requirement for a major capsid protien, PRD1, like adenovirus requires minor capsid proteins to

maintain viral stability (Rydman et al., 2001; San Martin et al., 2001, 2002). Accordingly, all of these structural similarities together with the fact that PRD1 and adenovirus share a common DNA replication system have led to the proposal that PRD1 and adenovirus share a mutual evolutionary ancestor (Belnap & Steven., 2000).

Likewise, the dsRNA virus, reovirus has been also been suggested to share a common evolutionary origin with adenovirus. This was proposed following the determination of the crystal structure of the reovirus attachment protein σ1 (Chapell et al., 2002). Like the penton of adenovirus , the σ1 protein is found at the twelve vertices in the capsid and is composed of a head and tail domain, both of which are structurally similar to the penton base and fibre of adenovirus, respectively.

Thus, the similarities between adenovirus and both PRD1 and reovirus have suggested an evolutionary link between dsDNA viruses which infect prokaryotes and eukaryotes and also a link between dsDNA and dsRNA-containing viruses

1.2.3 Life cycle

Host cell infection by adenovirus commences following viral attachment to the plasma membrane. The virion is subsequently internalised and transported to the nucleus, at which point the viral DNA moves into the nucleus. The virion then inhibits host defences and utilises host resources to begin self-replication prior to release from the cell. Each of these stages of the viral life cycle are described in detail in this section.

1.2.3.1 Attachment and Entry

Both hexons and pentons have been implicated in adenoviral attachment and subsequent entry to host cells. Hexons were proposed in a recent study by Balakireva et al., 2003 to be involved in entry to alveolar epithelial cells. These cells contain very low concentrations of the Coxsackie and Adenovirus Receptor (CAR), commonly bound by penton fibres in viral attachment. The paper proposed that hexons can bind Dipalmitoyl Phospatidylcholine (DPPC), a surfactant released by the alveolar cells. This surfactant re-enters the cell line during lung surfactant turnover, therefore the paper concluded that this may provide a pathway for viral entry into alveolar cells.

However for most cell lines the primary receptor involved in viral attachment is the CAR receptor, which as the name suggests was first discovered as a receptor for both coxsackievirus species B and adenovirus subgroup C (Lonberg-Holm et al., 1976; Bergelson et al., 1997; Tomko et al., 1997).

Following the findings that subgroup C could bind CAR, Roelvink et al., 1998 selected representative members of each of the human subgroups – subgroup A (Ad12, Ad31), subgroup B (Ad3), subgroup C (Ad2, Ad5), subgroup D (Ad9, Ad15, Ad19), subgroup E (Ad4) and subgroup F (Ad41) to analyse whether they all utilised CAR. The paper found that all subgroups except subgroup B could bind CAR. This had previously

been noted by Stevenson et al., 1995. Interestingly it was also discovered that although subgroup F contained two fibres, a short fibre and a long fibre, only the long fibre could bind CAR.

The binding of the knob to CAR was visualised with the publication of the 2.6 Angstrom crystal structure of Ad12 bound to CAR (Bewley et al., 1999). The structure proved definitively that each trimer of protein IV bound three receptors, and that a surface loop (AB loop) on the knob contributed to over 50% of the interaction between the penton and CAR and was therefore the major site of contact between the two.

However it has been suggested that CAR is not the sole primary receptor for subgroup C, as the MHC class I receptor has also been shown to be used by Ad5 when there is little or no cell surface expression of CAR (Hong et al., 1997; Davison et al., 1999). Likewise Ad37 (subgroup D) has been found to preferentially bind Sialic acid over CAR (Arnberg et al., 2000).

The length and flexiblity of penton fibres within different adenovirus serotypes has been suggested to be a key determinant in receptor binding. It is thought although Ad37 contains a CAR binding domain it does not bind the receptor efficiently due to its short, rigid fibre (Wu et al., 2003). The use of a variety of receptors is beneficial to adenovirus as it allows the virus family to attach to and invade a number of cell types.

Once cell attachment has occurred it has been shown that the penton base of subgroup A (Ad12), subgroup B (Ad3), subgroup C (Ad2 and Ad5) and subgroup E (Ad4) bind integrin receptors (Belin & Boulanger., 1993; Wickham et al., 1993; Mathias et al., 1994; Salone et al., 2003). The integrin family of receptors are heterodimeric, transmembrane receptors which are composed of an α-subunit and β-subunit. Several α- and β-subunits have been characterised and found to bind extracellular proteins such as

fibronectin and laminin. These proteins are recognised by their RGDS (arginine-glycine-aspartate-serine) motifs, however other motifs have been identified (reviewed in Wolfe., 1993). Likewise, each monomer of the adenovirus penton contains an RGD motif (Neumann et al., 1988) which is recognised by integrins. Once bound to the receptor, the virion enters the cell via receptor mediated endocytosis (Fitzgerald et al., 1983; Svennson & Perrson, 1984; Varga et al., 1991; Wang et al., 1998).

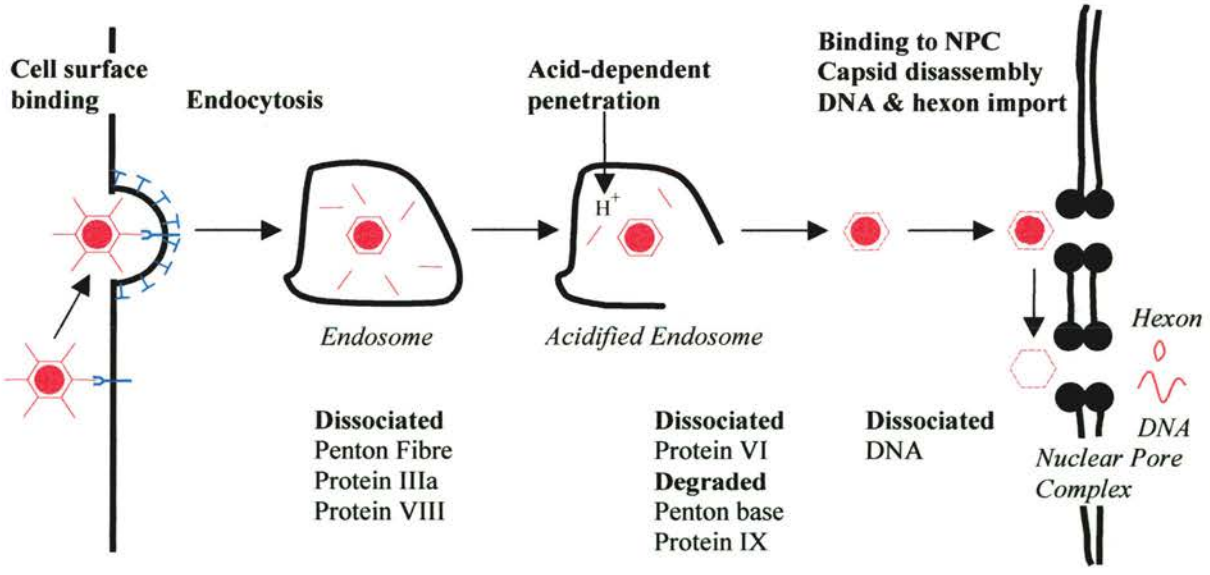
Following internalisation, the virion is released from the endosome into the cell cytoplasm and a number of factors have been implicated in this. Early studies investigating adenoviral infection suggested that a reduction in pH within the endosome was essential for penton binding to the membrane and subsequent viral release (Fitzgerald et al., 1983; Seth et al., 1984; Svennson & Perrson, 1984; Svennson, 1985; Seth, 1994). Penton base binding to the $\alpha_v\beta_5$ integrin was later implicated in this adenovirus-mediated permeabilisation of the membrane (Wang et al., 2000). It has also been suggested that although the penton fibre has begun to dissociate from the virion by this point in infection, it may play a role as a pH detector thereby regulating the timing and location of viral release into the cytoplasm (Miyazawa et al., 2001).

Another element which has been associated with virion release from the endosome is the formation of macropinosomes (formed by budding from membrane ruffles). Meier et al., 2002 proposed that macropinocytosis was stimulated by adenovirus binding the CAR receptor. The paper also suggested that there was correlation between macropinocytosis and endosomal leakage and that macropinosomes may play an as yet unidentified role in viral release from the endosome.

1.2.3.2 Nuclear Localisation and Import

Once released from the endosome, the virus is transported to the nucleus via hexon binding to microtubules (Dales & Chardonnet, 1973; Luftig et al., 1975; Suomalanien et al., 1999). From entry to the cell via endocytosis, to transport to the nucleus the virus is uncoated in a stepwise manner (Greber et al., 1993). The process of dismantling is shown in Figure 1.2.5.

Figure 1.2.5: Stepwise Dismantling of Adenovirus During Cell Entry
Adenovirus virions utilise receptor mediated endocytosis to enter host cells and are subsequently released from the endosome in a pH-dependent manner. During cell entry and transport to the nucleus the virion is sequentially dismantled by dissociation and degradation. Once it has reached the Nuclear Pore Complex (NPC), the viral DNA and hexon are transported across the Nuclear envelope into the nucleus (adapted from Greber et al., 1993).



Investigations into the dismantling process have suggested that degradation of protein VI is caused by the adenovirus protease (Avp) (Greber et al., 1996). It has been proposed that upon release from the host cell, the protease is inactivated by the oxidising environment of the extracellular matrix. It is then reactivated in the reducing environment

of the endosome. In this environment, Avp is capable of cleaving protein VI, who's function is to anchor DNA to hexon in the capsid. Thus the cleavage of VI frees the viral DNA from the capsid thereby allowing it to enter the nucleus (Greber et al., 1996). The viral DNA along with its covalently bound terminal protein (TP) are imported into the nucleus due to the Nuclear Localisation Signal (NLS) carried by the TP.

1.2.3.3 DNA Transcription and Replication

Following viral DNA entry to the nucleus, terminal protein (TP) forms a complex with the CAD (Carbamyl phosphate synthetase, aspartate carbamylase and dihydrorotase) enzyme of the nuclear matrix (Bodnar et al., 1989; Schaak et al., 1990; Fredman & Engler., 1993; Angeletti & Engler, 1998) and viral transcription ensues.

Adenovirus DNA transcription and replication occurs as a series of events. These events take place in the order, early gene transcription, DNA replication, intermediate gene transcription and finally late gene transcription. Each of the three transcription events produce a number of proteins which are listed in Table 1.2.2 and whose functions are discussed below.

Table 1.2.2: Early, Intermediate and Late Gene Transcription Products
Adenovirus DNA transcription occurs in the order, early (E) gene transcription, intermediate gene transcription and finally late (L) gene transcription. Each of the three transcription events produce a number of proteins which are listed in the table (adapted from San Martin & Burnett., 2003).
* Proteins are denoted by their molecular mass (K) or by their name.

Transcription Unit	**Proteins***
E1A	13, 27, 32
E1B	16, 21, 55
E2A	DNA-binding protein (DBP)
E2B	DNA polymerase, Terminal Protein (TP)
E3	4, 7, 8, 10,12, 13, 15, 19
E4	7, 13, 14, 15, 17
Intermediate	IX, IVa2
L1	p52/p55, IIIa
L2	pX, pVII, V, III
L3	Protease, pVI, II
L4	25, pVIII, 100K
L5	IV

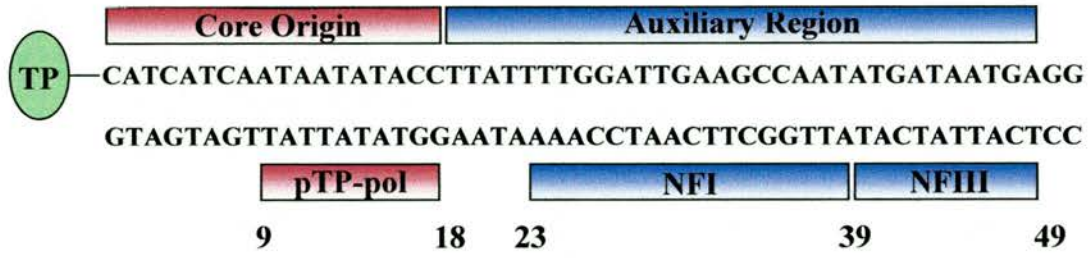
The four early gene transcription units are named E1 (E1A and E1B), E2 (E2A and E2B), E3 and E4. The proteins transcribed from these units are involved in maintaining the

cell to allow the production of viral progeny, combating host defences and allowing viral DNA replication to occur (reviewed in Shenk., 1996; Russell., 2000).

Following early gene transcription, DNA replication occurs at the origin of replication (*ori*). The sequences which define the *ori* are found within an Inverted Terminal Repeat (ITR), which is located at either end of the 36 Kbp viral genome. The *ori* is composed of a core origin and an auxiliary region in human Ad2 (Rawlins et al, 1984; Figure 1.2.6), which bind viral E2 gene transcripts and cellular nuclear factors (NF), respectively. The E2 gene cassette transcripts which are required are DNA polymerase (pol), pre-terminal protein (pTP) and DNA Binding Protein (DBP), whilst the cellular nuclear factors are NFI (Nagata et al., 1982), NFII (Nagata et al., 1983a) and NFIII (Pruijin et al., 1986).

As stated above, the viral and nuclear transcription factors bind independent sites. The pTP-pol heterodimer (Lichy et al., 1982) binds bases 9-18 of the core region, which is conserved throughout all human adenovirus serotypes (Stillman et al., 1982; Figure 1.2.6), whilst NFI and NFIII bind the recognition sites 23-38 (Nagata et al., 1982, 1983b; Jones et al., 1987; Adhya et al., 1986; Figure 1.2.6) and 39-49 (Pruijin et al., 1986; O'Neill & Kelly., 1988; Figure 1.2.6), respectively. However, although the binding of these cellular proteins to the viral auxiliary region have been found to be important in human Ad2 viral DNA replication it has been discovered that the first 18 bases alone are sufficient to initiate DNA replication (Challberg & Rawlins., 1984).

Figure 1.2.6: Core and Auxiliary Region of Human Ad2
DNA replication occurs at the origin of replication (*ori*). The sequences which define the *ori* are found within an Inverted Terminal Repeat (ITR), which is located at either end of the 36 Kbp viral genome. The *ori* is composed of a core origin and an auxiliary region in human Ad2. The pTP-pol heterodimer binds bases 9-18 of the core region, whilst NFI and NFIII bind the recognition sites 23-38 and 39-49, respectively (Adapted from Hay., 1996).



DNA replication begins with the coating of DNA by DBP (van der Vleit & Levine., 1973). The coating by DBP increases the affinity of NFI for its site (Cleat & Hay., 1989; Stuiver & van der Vliet., 1990), thereby rendering the NFI sites saturated. Following this, an interaction between NFIII, bound to its recognition site and pTP, recruits the pTP-pol complex to its' site (Coenjaerts et al., 1994; de Jong et al., 2002). The pre-initiation complex is then formed and is stabilised by NFI which binds to pol (Bosher et al., 1990; Chen et al., 1990; Mul et al., 1990; Mul & van der Vliet., 1992; Armentero., 1994). Furthermore, the genome-bound TP has also been found to stabilise the complex (Pronk & van der Vleit., 1993). Together, these three proteins are involved in orientating the pTP-pol heterodimer to its correct site.

Initiation is then proposed to occur by a jumping back mechanism (King & van der Vleit., 1994). Following the formation of the pre-initiation complex, the pTP-pol heterodimer is repositioned for initiation so that Ser580 of pTP is positioned opposite Guanine4 of the DNA (Figure 1.2.6). Addition of the three nucleotides at positions 4-6 then occurs to form a pTP-trinucleotide complex (pTP-CAT). The pTP-CAT complex then

jumps back three bases to GTA at positions 1-3 in the *ori*. pTP then begins to dissociate from pol to allow elongation of the DNA strand by pol (King et al., 1997).

Following DNA synthesis, the non-template strand is displaced and can form a duplex via its ITR's. This strand can be used in a further round of DNA replication (Stow et al., 1982; Hay et al., 1984). Alternatively, the displaced strands form dsDNA and are processed then packaged into the virion.

Subsequent to DNA replication the intermediate transcription proteins, IX and IVa2 are produced. This is followed by the Late transcription of the remaining proteins as shown in Table 1.2.2.

1.2.3.4 Assembly and Release

Subsequent to viral DNA transcription and replication, proteins are translated in the cytoplasm and then translocated to the nucleus where they are assembled to form new progeny. However a complete mechanism of adenoviral assembly is yet to be established.

The first structure assembled in the nucleus is the major capsid protein, hexon. Hexon monomers (polypeptide II) are folded to trimers in the cytoplasm by the L4 100K chaperone protein (Cepko & Sharp., 1982). Following this, as protein pVI has been shown to be capable of binding hexon (Matthews & Russell., 1994, 1995) it has been proposed that pVI transports hexon into the nucleus (Kauffman & Ginsberg., 1976; Wodrich et al., 2003). A model for this transport suggested by Wodrich and colleagues, 2003 proposed that pVI binds hexon and that the C-terminal Nuclear Localisation Signal (NLS) on pVI then interacts with the α/β importin to allow transport of the complex across the nuclear membrane. Once inside the nucleus the importin dissociates from pVI and then pVI detaches from the hexon. The hexon then remains in the nucleus whilst pVI Nuclear Export Signal (NES) binds an exportin and moves back to the cytoplasm, where the cycle

begins again. The paper further suggested that as the hexon concentration increased within the nucleus, the number of hexon binding sites for pVI increased therefore causing pVI to remain in the nucleus. Subsequently the large concentration of pVI would activate Avp from a basal level (Section 1.3.2.1) thus leading to N- and C-terminal cleavage of pVI by Avp. This would consequently result in the removal of the NLS and NES and also in the maturation of pVI to VI, changing the protein from a nuclear import protein to a structural protein.

Furthermore work analysing viral intermediates during assembly have discovered that capsids are produced in an order (D'Halluin et al., 1978; Evardsson et al., 1978). Firstly an empty capsid with a density of 1.31 g/mL is formed which contains hexon and penton capsomeres, pIIIa, pVI and pVIII. Subsequently a heavy intermediate with a density of 1.37 g/mL is created. This contains all of the components of the empty capsid, as well as viral DNA. However it lacks the core proteins V, VII and μ which are thought to be lost during the isolation procedure. Following viral assembly, packaging of DNA into the viral particle is believed to occur in a left-to-right manner and it is thought that this is due to an A-T rich region found at the left end of the viral genome (Daniell., 1976; Tibbetts., 1977; Grable & Hearing., 1990). Nevertheless it is still unknown whether the DNA enters the cell in the presence or absence of V, VII and μ. Once the viral DNA is packaged and the virion is formed, it is postulated that Avp is activated by the C-terminal 11-residue peptide from pVI (Section 1.3.2.1). Subsequently the viral precursors pIIIa, pVII, pVIII, pμ and pTP are processed by Avp to produce a mature virion. Avp has also been implicated in the cleavage of the cytokeratins K7 and K18. Cleavage would lead to the disruption of the host cell cytokeratin network, cell lysis and the eventual release of mature virions from the nucleus (Chen et al., 1993).

1.2.4 Infection of Humans

Adenoviruses infect a number of species and can be isolated from both healthy and diseased individuals. This section discusses four topics including the clinical conditions caused by the virus, the immune response to viral invasion, adenovirus as a cancer-causing agent and the therapeutic use of the virus in gene therapy.

1.2.4.1 Clinical Conditions

There are 51 serotypes of human adenoviruses and although these serotypes can infect a number of host cells they predominantly infect the respiratory and gastrointestinal tracts to cause a variety of diseases (reviewed in Shenk., 1996; Table 1.2.3).

Respiratory infections by adenovirus serotypes 1, 2, 3, 5 and 6 commonly occur in early childhood to cause 5-10% of mild paediatric respiratory infections, whilst serotype 7 causes acute respiratory infection with some mortality. Moreover, gastrointestinal infections involving serotypes 40 and 41 cause severe diarrhoea which can be fatal where water supplies are of a poor standard. However, the virus is generally asymptomatic.

Table 1.2.3: Clinical Conditions Caused by Adenovirus in Humans

Syndrome	Serotype
Upper respiratory illness	3, 7, 11, 14, 16, 21, 34, 35, 50
Lower respiratory illness	3, 4, 7, 21
Pertussis syndrome	5
Acute respiratory disease	4, 7, 21
Pharyconjunctival fever	3, 4, 7
Epidemic keratoconjuncitivitis	8, 19, 37
Acute haemorrhagic conjunctivitis	11
Acute haemorrhagic cystitis	7, 11, 21, 35
Infant gastroenteritis	2, 31, 40, 41
Penile/Labial ulcers and urethritis	2, 19, 37

1.2.4.2 Immune Response

The response to viral infection is divided into two categories, the innate- and the adaptive- immune responses. The adaptive immune response is further subdivided into the interferon-, apoptosis-, cellular-, and humoral- responses. The immune responses to adenovirus infection by all of these mechanisms are discussed below.

1.2.4.2.1 Innate

The innate immune response has a array of systems at its disposal to fight adenoviral infection. These include the antimicrobial peptides named defensins (reviewed in Ganz & Lehrer., 1998) which have been linked to protecting host cells from viral infection (Gropp et al., 1999). Cytokines are also released in response to invasion by adenovirus (Charles et al., 1999) which results in the recruitment of neutrophils to the site of infection (Muruve et al., 1999). Neutrophils then function to both phagocytose foreign material and to induce an immune response. Likewise NF-κB transcription factors are produced late in adenoviral infection to regulate the innate immune response (Clesham et al., 1998; Lieber et al., 1998).

1.2.4.2.2 Interferons

There are two types of interferons, Type 1 (α and β) and Type 2 (γ), both of which are produced following the appearance of adenovirus (Reich et al., 1988). Their function is to bind cellular receptors and activate the Jak/STAT pathway in the cell. This results in the production of STAT complexes within the cytoplasm of the cell. These complexes then translocate to the nucleus where they bind the interferon-response elements (ISRE's) of the cellular DNA. The function of the ISRE's is to regulate the transcription of gene products

involved in forming a defence against the virus. However as with many immune responses viruses can counteract them and it has been proven that adenoviral E1A gene products can down-regulate the coactivators of STAT thereby inhibiting their actions (Leonard & Sen., 1996, 1997; Look et al., 1998; McDonald & Reich., 1999; Paulson et al., 1999).

1.2.4.2.3 Apoptosis

Under certain conditions, for example viral infection, the immune system activates several pro-apoptotic proteins. With the most well-known being p53, which acts to modulate the transcription of genes required for apoptosis. However, it has been found that the adenovirus E1B 19K protein can repress this mechanism of apoptosis (Debbas & White., 1993).

A similar mechanism of apoptosis is by mediated by Tumour Necrosis Factor (TNF)-α. TNF-α is secreted by monocytes and leukocytes following activation by the innate immune system and has been shown to be important in clearing adenovirus (Elkon et al., 1997). It plays two roles in adenoviral elimination, the first of which is the activation of cytosolic phospholipase A2 (pLA2). pLA2 catalyses the release of arachidonic acid, which leads to apoptosis (Wolf & Laster., 1999). Normally this pathway is modulated by Bcl-2 (de Mossiac et al., 1999) and therefore it has been suggested that its viral analogue, E1B 19K may also be capable of this function. The second role of TNF-α in apoptosis is to mediate a caspase cascade (Kimura & Gelmann., 2000), but yet again the E3 gene products of adenovirus have been shown to subvert this mechanism of apoptosis (Lukashok et al., 2000).

Likewise, the apoptotic effect induced by the Fas-Fas ligand in liver cells infected with adenovirus vector (Chirmule et al., 1999) has also been shown to be inhibited by E3

gene products. These E3 proteins cause the ligand to be removed from its cellular receptor thereby inhibiting apoptosis (Tollefson et al., 1998).

It has also been discovered that the cellular protein FIP-3 can inhibit the activation of NF-κB. Therefore because NF-κB can inhibit apopotosis, it has been suggested that FIP-3 has a pro-apoptotic effect (Li et al., 1999). It has been further proposed that adenovirus E3 14.7K protein can bind FIP-3 to restore NF-κB activity and inhibit apoptosis (Li et al., 1999).

1.2.4.2.4 Cellular

The cellular immune response to adenovirus is mediated by the T cells, $CD8^+$ cytotoxic cells (CTL's) and $CD4^+$ helper cells. In terms of the CTL's, the response begins with a viral antigen being recognised by the class I Major Histocompatibility Complex (MHC) protein. The complex then moves to the plasma membrane of the cell where CTL's recognise the antigen. In response to this, perforin is produced and induces cell lysis. It has been shown that there is cross-reactivity of this particular immune response to different human subgroups thus suggesting that infection with one adenovirus may protect individuals in part from other adenovirus serotypes (Smith et al., 1998). However adenovirus can combat this immune response as E4 gene products play a role in resistance to lysis induced by T-cells (Kaplan et al., 1999) and E3 19K forms a complex with MHCI proteins (Kvist et al., 1978) causing it to be retained in the ER, therefore inhibiting its effect on apoptosis (Kaplan et al., 1999).

In a similar manner to CTL's, $CD4^+$ helper cells recognises an MHCII-antigen complex when it is presented on the plasma membrane of the infected cell. The antigens detected by $CD4^+$ cells have been suggested to be the adenoviral structural proteins, penton

fibre, polypeptide VI and IIIa (Souberbielle & Russell., 1995) and this detection is thought to provide most individuals with a long-lived $CD4^+$ response to all adenoviruses (Flomenberg et al., 1995). Recognition of the antigen subsequently leads to $CD4^+$cells stimulating the proliferation of B cells to produce antibodies for the humoral response (Section 1.2.4.2.5).

1.2.4.2.5 Humoral

As stated in section 1.2.4.2.4, the humoral response is dependent upon B cells and their proliferation is stimulated by $CD4^+$cells. The function of these B cells is to produce immunoglobulins which are expressed on the cell surface and are released into the blood system as well as other body fluids. In the case of adenovirus the antigen which these antibodies are directed against are the capsomeres (hexon and penton fibre/base) of the viral capsid (Willcox & Mautner., 1976; Gahery-Segard et al., 1997, 1998).

1.2.4.3 Oncogenesis

Adenoviruses mainly cause asymptomatic diseases in humans. However in 1962, Trentin and colleagues showed alarmingly that human Ad12 was capable of inducing tumours in newborn hamster cells. Thus human adenoviruses are now termed DNA tumour viruses (reviewed in Graham., 1984; Branton et al., 1985) and have been divided into three categories according to their oncogenecity. Those categorised into 'high' cause tumours at a high frequency within a few months of incubation, whilst 'moderate' tumours occur at low frequency after long incubation times and 'non-oncogenic' serotypes do not cause tumours (Table 1.2.4).

Thus, the role of early gene transcripts in the transformation of healthy cells and also the hit and run theory of oncongenesis and its implications in human cancers are discussed below.

Table 1.2.4: Classification of Oncogenic Human Adenoviruses
Human adenoviruses are divided into three categories according to their oncogenecity. Those categorised into 'high' cause tumours at a high frequency within a few months of incubation, whilst 'moderate' tumours occur at low frequency after long incubation times and 'non-oncogenic' serotypes do not cause tumours. The table below highlights the human adenovirus subgroups and the serotypes of each which have been examined for oncogenic potential (adapted from Endter & Dobner., 2004).

Subgroup	Serotype	Oncogenecity
A	12,18,31	High
D	9,10	High
B	3,7	Moderate
C,E,F	All	None

1.2.4.3.1 Function of Early Genes

Adenovirus genes transcribed early in infection play a major role in oncogenesis. In particular, the protein products of the E1A, E1B and E4 early transcription units have been proven to be essential in the transformation of host cells.

Early in infection E1A encodes two mRNA's, 12S and 13S which are translated to produce proteins of 289 and 244 amino acids, respectively. These proteins are capable of transforming cells (Haley et al., 1984; Moran et al., 1986). Nonetheless the transformation is incomplete when these proteins are expressed alone as the transformants neither grow to high density nor do they induce a high frequency of tumours (Gallimore et al., 1984).

However, complete transformation has been noted when E1A and E1B genes are expressed simultaneously (Ruley et al., 1983). The E1B gene products believed to be responsible for this are E1B 19K and E1B 55K. These proteins independently co-operate

with E1A gene transcripts to transform cells but transformation occurs with a greater efficiency when both E1B proteins are expressed (Bernards et al., 1986; White & Cipriani., 1990; McLorie et al.,1991). It is believed that E1B 19K and E1B 55K aid in cell transformation by suppressing the apoptotic effects of the E1A gene transcripts (Debbas & White., 1993; Lowe & Ruley., 1993; Teodoro et al., 1995).

Similarly, by the 1980's three lines of evidence indicated that E4 proteins were involved in transformation of cell lines by adenovirus. The evidence included, the detection of E4 proteins in transformed cell lines (Esche., 1982; Esche & Steigman., 1982), the finding that antisera from hamster tumours contained antibodies raised against E4 gene products (Sarnow et al., 1982; Downey et al., 1983), and the discovery that co-expression of E1 and E4 genes from human Ad12 could transform cells (Shiroki et al., 1984). More specifically, recent investigations have suggested that the E4orf3 and E4orf6 protein products of human subgroup C can co-operate with both E1A and E1B gene transcripts to transform rodent cell lines (Ohman et al., 1995; Nevels et al., 1997; Nevels et al., 1999).

1.2.4.3.2 Human Tumours

In respect of human cancers, regions of adenoviral genomes have been discovered in several malignant tumours including cervical tumours, astrocytomas, metastatic brain carcinoma, glioblastomas and melanomas to name but a few (Maitland et al., 1981; Ibelgaufts et al., 1982; Kuwano et al., 1997a, 1997b). However there is no reliable epidemiological, serological or genetic evidence to suggest that adenoviruses induce these tumours.

Nevertheless it has been shown that human Ad12 can induce tumours in cultured cells by a hit and run mechanism, whereby the viral oncogenes initiate transformation but are not required to maintain it, and continuous passage of the cells results in the loss of

adenoviral DNA (Paraskeva & Gallimore., 1980; Kuhlmann et al., 1982; Paraskeva et al., 1982; Pfeffer et al., 1999). To-date, the viral oncogenes suggested to provide this mutagenic function and therefore induce transformation by a hit and run mechanism are the E1A, E4orf3 and E4orf6 gene products (Shen et al., 1997; Nevels et al., 2001).

Thus, if adenoviruses induce tumours by a hit and run mechanism, then the viral genes would not persist in the transformed cells, therefore explaining why there is no reliable epidemiological, serological or genetic evidence which links adenoviral infection to cancer.

1.2.4.4 Gene Therapy

The use of adenovirus as a vector in gene therapy has three major advantages; they can be easily grown to high titres, they can infect a variety of host cells and newer vectors can maintain large quantities of foreign DNA. The interest in the use of adenovirus in gene therapy has led to a multitude of publications relating to the treatment of inherited and acquired diseases in animal models. However, it must be borne in mind that although the vectors appear to be safe in experimental models, they may be toxic to humans.

The uses of adenoviral vectors and the development of their cell specificity are discussed in this section.

1.2.4.4.1 E1- and E1-/E3-Deleted Vectors

The first adenovirus vectors produced for gene therapy had the E1 region deleted and were propagated in the murine 293 cell line (Graham et al., 1977), as these cell lines provide the E1 gene functions *in trans*. Removal of the E1 region provided two advantages, firstly, the viral replication potential was decreased and secondly, the gap in the viral genome allowed the insertion of a transgene. In addition to the removal of the E1

transcription unit, many first generation vectors also had the E3 region removed. This unit was removed as its main responsibility was to transcribe proteins essential for viral evasion of the host immune system and this was not believed to be necessary.

E1- and E1-/E3-deleted vectors were not expected to replicate and therefore were not anticipated to induce an immune response. However it soon became apparent that expression of other viral proteins led to an immune response and the clearance of transduced cells. This in turn meant that transgenes were only expressed for short periods of time. Although this initially appeared to be a disadvantage, it was in fact highly advantageous for the expression of non-persistent transgenes to be used in the treatment of cancer.

As in many cancers, the p53 gene is either mutated or lost. Therefore, first generation adenovirus vectors have been used to deliver the wild-type p53 gene to treat non-small cell lung cancer (Roth et al., 1998; Waugh et al., 1999) and head and neck squamous cell carcinoma (Clayman et al., 1998). Moreover the vectors have been used to express the genes for the anti-tumour agents, thymidine kinase and creatine deaminase, which have had encouraging results in clinical trials (Rogulski et al., 2000; Trask et al., 2000).

1.2.4.4.2 E2, E3 and E4 Mutant Vectors

As E1- and E1-/E3-deletion mutant vectors are not maintained long-term within transduced cell lines, investigations into the mutation of other early viral genes have been undertaken. Deletion of the E2 gene transcription unit has led to expression of transgenes for periods of up to 6 months, which is far in excess of the transient expression seen by E1-deletion vectors (Ding et al., 2001). The improvement is believed to be caused by a reduced immune response in the absence of the transcribed proteins, pTP and pol.

Similarly, E4- and E1-/E4-deletion mutants have also provided long-term expression of transgenes (Armentano et al., 1997,1999; Brough et al., 1997; Dedieu et al., 1997; Wang et al., 1997). Nevertheless, although the E3 gene is often deleted in adenoviral vectors (Section 1.2.4.4.1) recent studies have shown that when the E3 gene is maintained, its' anti-immune response preserves the transduced cell thus leading to persistent transgene expression (Lee et al., 1995; Bruder et al., 1997). Therefore, together, these results suggest that the key to designing an adenovirus vector which can express a transgene, long-term may be dependent upon removing some genes and retaining others.

In addition to changing the genome content of the virus to allow long-term transgene expression, it is important that the adenovirus vector is tailored to the correct cell line. This has been accomplished by replacing viral promoters with a tissue-specific promoter. Both the E1 promoter (Rodriguez et al., 1997; Hallenbeck et al., 1999; Hernandez-Alcocoeba et al., 2000; Kurihara et al., 2000; Adachi et al., 2001; Matsubara et al., 2001) and the E4 promoters (Doronin et al., 2000) have been replaced with a high degree of success in the treatment of cancer.

1.2.4.4.3 High-Capacity Adenovirus Vectors

High-Capacity Adenovirus (HC-Ad) vectors lack most of the adenoviral genome and were designed as previous vectors had a low capacity for foreign DNA and the expression of some viral proteins caused an immune response which led to clearance of the transduced cell (Mitani et al., 1995; Clemens et al., 1996; Fisher et al., 1996; Kochanek et al., 1996; Kumar-Singh & Chamberlain., 1996).

Although most of the viral genome is removed in HC-Ad vectors, two components are retained; the ITR's, which are essential for viral DNA replication (Section 1.2.3.3) and the sequence at the left terminus of the genome which is required for DNA encapsidation

(Section 1.2.3.4). As the vectors contain no other viral genes, they are propagated in 293 cells in the presence of E1-deleted helper viruses. They are then separated from the helper virions by caesium gradient purification.

The only restriction of this vector is that the maximum amount of DNA which can be inserted is 37Kbp. The lowest amount is also important as vectors containing less than 27Kbp undergo DNA rearrangements, have inefficient packaging and show low level, transient expression of the transgene (Lieber et al., 1997; Parks & Graham., 1997). However a lack of DNA can be overcome by the addition of extra DNA termed "stuffer" DNA.

Therefore this vector is very promising for targeted, long-term expression of genes in host cell lines.

1.3 ADENOVIRUS PROTEASE

Investigators have been studying adenovirus protease for over twenty years, which has led to the protease acquiring several names, thus for the sake of simplicity it will either be referred to as 'Avp' or 'the protease' throughout this work. The protease is essential for the viral life cycle and until recently was the only member of the CE clan of cysteine proteases. As characterising the residues of the protease involved in the S2 and S4 substrate specificity pockets was the major focus of this thesis, this sub-chapter provides background information about the cysteine protease, its activation, structure and substrate specificity.

1.3.1 Cysteine Protease

The adenovirus protease was found to be a 23K polypeptide after mapping of the Leucine-Proline (P137L) temperature sensitive (ts1) mutation to the L3 region of the genome (Yeh-kai et al., 1983). The P137L mutation leads to the protease not being packaged into the complete virion at the non-permissive temperature of 39°C. Therefore non-infectious virions are not produced as the structural proteins are not cleaved (Rancourt et al., 1995).

Since this discovery, the 23K protein has been purified to homogeneity from human adenovirus serotype 2 and its proteolytic activity has been tested using a synthetic substrate. This proved definitively that the 23K protein was not only required for proteolytic activity but was in fact the protease (Webster & Kemp, 1993).

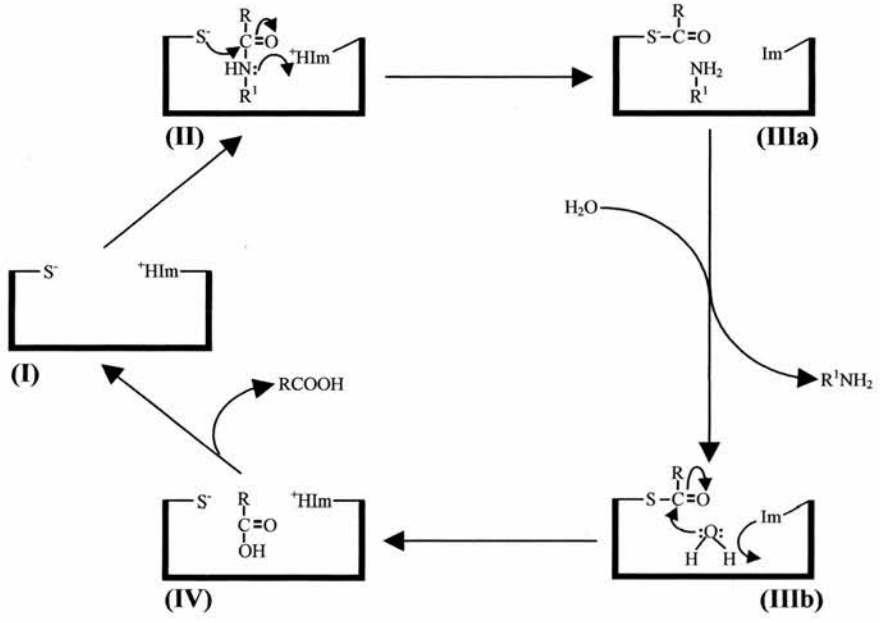
Following the identification of the protease as the L3 23K polypeptide, of which there are around 10 (Anderson, 1990) to 50 (Brown et al., 1996) copies per virion, it became important to classify the protease as a Serine-, Cysteine-, Metallo- or Aspartic-protease. The protease was initially considered to be a Serine protease due to inhibitor studies (Bhatti & Weber, 1979a; Tremblay et al., 1983) and sequence alignments (Houde & Weber, 1990a). Thereafter, inhibitor experiments contradicted these findings and instead suggested that the protease was a Cysteine protease (Tihanyi et al., 1993; Webster et al., 1989a), belonging to a unique Cysteine protease class (Webster & Kemp, 1993; Mangel et al., 1997). Classification as a cysteine protease was also proposed following a sequence alignment of Avp2 with the Poliovirus 3C proteinase which indicated that His54 and Cys122 were members of the catalytic triad (Webster et al., 1989a). These residues were then confirmed to form the catalytic site by mutational analysis (Grierson et al., 1994, Jones et al., 1996).

The third member of the catalytic triad was determined by two groups, Rancourt et al., 1996 and Ding et al., 1996. Whilst Rancourt et al. utilised mutational studies, Ding et al. determined the crystal structure of Avp2 to a 2.6 Angstrom resolution. Together, the papers made a convincing proposal that the third member of the catalytic triad for human adenovirus serotype 2 protease was Glu71. Consequently the three members of the triad are His54, Glu71 (Asp in all serotypes except human serotype 2 and 5) and Cys122 (numbering for Avp2).

The crystal structure was also used to superimpose Avp2 and Papain (Ding et al., 1996). From this, the group proposed that the catalytic triad of Papain (Cys25, His159, Asn175) lay in an identical position to that of Avp2 (Cys122, His54, Glu71). The group therefore suggested that Avp2 may share a common catalytic mechanism with Papain.

Papain is the archetype of cysteine proteases and has provided a general catalytic mechanism (reviewed in Storer & Menard, 1994; Figure 1.3.1). The first step in the hydrolysis of amino acids by cysteine proteases involves (**I**) the free enzyme (**II**) binding to the substrate. This is followed by (**IIIa**) acylation of the enzyme and the formation of the first product. (**IIIb**) Water then reacts with the acyl-enzyme in a deacylation step to form (**IV**) the second product and return the enzyme to a (**I**) free state (Figure 1.3.1). It has been proposed that the function of the third member the catalytic triad, Asn175 for Papain and Glu71 for Avp2 is to ensure that the imidazole ring of the histidine is in an optimal position throughout the different reactions.

Figure 1.3.1: General Catalytic Mechanism for Cysteine Proteases
It has been suggested that Avp2 may share a common catalytic mechanism with Papain. Papain is the archetype of cysteine proteases and has provided a general catalytic mechanism for cysteine proteases. The diagram below denotes S^- as the sulphur group of the catalytic cysteine and ^+HIm as the H^+ of the imidazole of the catalytic histidine. The first step in the hydrolysis of amino acids involves (**I**) the free enzyme (**II**) binding to the substrate. This is followed by (**IIIa**) acylation of the enzyme and the formation of the first product. (**IIIb**)Water then reacts with the acyl-enzyme in a deacylation step to form (**IV**) the second product and return the enzyme to a (**I**) free state. (adapted from Storer & Menard., 1994).



As discussed, Papain and Avp have two identical characteristics, firstly they have been proposed to share a common catalytic mechanism and secondly they are both cysteine-type proteases. However, cysteine-type proteases are divided into clans, and whilst Papain is member of the CA clan, Avp is a member of the CE clan (http://merops.sanger.ac.uk).

Clans are further sub-divided into families according to their tertiary structure, although the order of the catalytic triad in the primary amino acid sequence can also be useful. When the structure of Avp was first discovered (Ding et al., 1996) it was found to

have a unique fold and this led to the establishment of a new cysteine protease clan, the CE clan. Until recently, Avp remained the only family of the clan, the C5 family. However the discovery of the Ubiquitin-like protease (Ulp1) in the yeast, *Saccharomyces cerevisiae* by Li & Hoschstrasser, 1999 constituted the addition of the family C48 to the CE clan. More recently, the clan has been extended to include the families C55, C57 and C63, which are represented by the members, YopJ protease (*Yersinia pseudotuberculosis*), Vaccinia virus I7L processing peptidase (Vaccinia virus) and African Swine Fever Virus processing peptidase (African Swine Fever Virus).

Of these families, structures have been published for Avp2 in complex with its activating peptide (Ding et al., 1996; McGrath et al., 2003) and Ulp1 in complex with its substrate, Smt3 (Mossessova & Lima, 2000). These structures were used as a basis for much of the bioinformatic analysis employed in this work to investigate substrate binding of Avp (Sub-chapter 3.1).

1.3.2 Activation

Early in the characterisation of the protease, groups began to study its control. It was first proposed that regulation occurred via autocatalysis of the 23K polypeptide to a mature 19K polypeptide (Chatterjee & Flint, 1987; Houde & Weber, 1990a), and this cleavage was suggested to occur between the conserved residues Ala46 and Gly47 (Avp2 numbering) (Houde & Weber., 1990a). However, as Anderson (1990) noted, the residues surrounding this dipeptide fitted neither of the consensus cleavage motifs for Avp (Section 1.3.4). Due to this and the fact that recombinant protease showed no activity, autocatalysis was disregarded as a means of activation of the protease. Thus implying that a cofactor may be required to activate the protease. Two viral cofactors and one cellular cofactor have now been proposed.

1.3.2.1 pVIcT

In 1993, the eleven residue peptide, GVQSLKRRRCF from the C-terminus of human Ad2 pVI was found to activate recombinant Avp2 expressed in *E. coli* (Mangel at al., 1993; Webster et al., 1993). Controversy followed the assignment of this cofactor as Weber and co-workers published work which indicated that recombinant Avp2 expressed in *E.coli* and insect cells showed proteolytic activity in the absence of pVIcT (Keyvani-Amineh et al., 1995; Tihanyi et al, 1993). The group later concluded that the protease has a great enough redox state to be active in the absence of pVIcT but has maximal activity in the presence of the peptide (Keyvani-Amineh et al., 1996), thus at any given time a number of protease molecules must have the correct, active conformation even in the absence of the activating peptide.

Following the determination of the crystal structure of Avp2 in the presence of pVIcT (Ding et al., 1996; McGrath et al., 2003) it was clear that the protease is divided into

two distinct domains, which are connected by pVIcT (Figure 1.3.3). It is believed that pVIcT functions as a 'strap' to bring these two domains together, therefore stabilising the catalytic site (Mangel et al., 1996).

As indicated by the crystal structure, the activating peptide forms a number of bonds with the protease. Those residues of particular significance are the highly conserved residues, Gly1', Val2', Cys10' and Phe11'. Mutational analysis of Gly1' and Val2' first suggested the importance of these residues in binding and activating the protease (Cabrita et al., 1997; Ruzindana-Umunyana et al., 2000). Further analysis, proved that in fact Gly1' and Phe11' were major components in binding whilst Val2' and Phe11' were important in stimulating enzyme activity (Baniecki et al., 2001).

Likewise, Cys10' of pVIcT has been shown to form a disulphide bond with Cys104 of the protease both *in vitro* (Grierson et al., 1994; Jones et al., 1996) and *in vivo* (McGrath et al., 2002).

1.3.2.2 DNA

Viral DNA was suggested to be a viral cofactor after it was discovered that Avp2 activity was lost when virus was incubated with DNase but reactivated by the addition of DNA (Mangel et al., 1993). The DNA does not have a specific sequence, as is the case for pVIcT and it is beleived that a polyanion, with a highly negative charge is all that is required for activation (Mangel et al., 1993). The proposed stoichiometry of binding is proportional to the length of the DNA, 3 and 6 molecules of Avp-pVIcT complex bind an 18-mer and a 36-mer of DNA, respectively (McGrath et al., 2001).

DNA as a viral cofactor was disputed by Webster et al., 1994, when this group found that DNA had no effect on the activity of the protease. There has been no resolution to the matter to-date.

1.3.2.3 Actin

Actin is a major cellular component and it has been shown *in vitro* to bind and activate Avp2, thereby allowing it to cleave cytokeratin K18 (Brown et al., 2002). When the C-terminal sequence of α-actin, β-actin, γ-actin and pVIcT were compared, actin was found to have homology to pVIcT (Figure 1.3.2). However two of the four residues which have been identified as being important in binding and activation of the protease are absent from the actin sequence - Gly1' and Val2' (Cabrita et al., 1997; Ruzindana-Umunyana et al., 2000; Baniecki et al, 2001). Although these residues have been cited as being important, they are not necessarily essential, as other residues also form bonds with the protease.

Figure 1.3.2: Comparison of the C-terminus of α-actin, β-actin, γ-actin and pVIcT
Actin has been shown *in vitro* to bind and activate Avp2, thereby allowing it to cleave cytokeratin K18. A primary sequence alignment of the of C-terminus of α-actin, β-actin, γ-actin and human Ad2 pVIcT was carried out and actin was found to show homology to pVIcT. Amino acid residues are coloured blue for identity and red for homology. (taken from Brown et al., 2002).

Human Ad2 pVIcT	G	V	Q	S	L	K	R	R	R	C	F
α-actin C-terminus	A	G	P	S	I	V	H	R	K	C	F
β-actin C-terminus	S	G	P	S	I	V	H	R	K	C	F
γ-actin C-terminus	S	G	P	S	I	V	H	R	K	C	F

Although no *in vivo* work has been published to prove that actin is a cellular cofactor, it has been shown that cytokeratin K18 is cleaved by the protease *in vivo* (Chen et al., 1993). Actin therefore remains a strong candidate as a cellular cofactor for Avp (Brown et al., 2002).

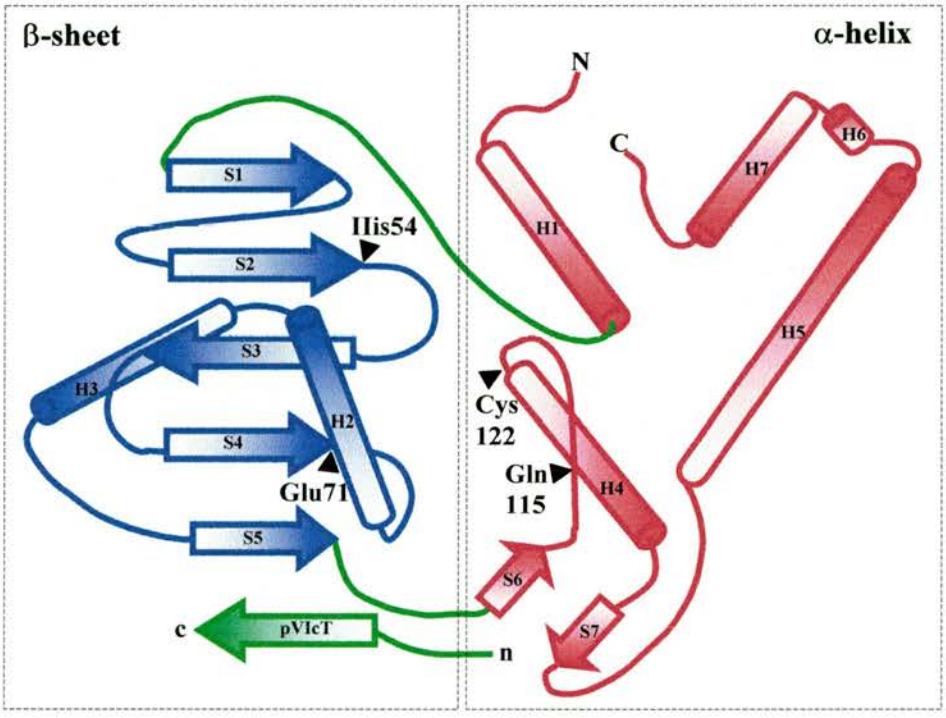
1.3.3 Structure

The crystal structure of Avp2 in the presence of its 11 amino acid activating peptide (pVIcT) has been resolved to both a 2.6 and a 1.6 Angstrom resolution (Ding et al., 1996; McGrath et al., 2003). These structures show clearly that the overall shape of the protease is ovoid. The structure is divided into two distinct domains, which are connected by two loops from Avp2 and a β-strand formed by pVIcT (Figure 1.3.3).

The domains have been assigned the names β-sheet domain and the α-helix domain (McGrath et al., 2003). The names indicate the major constituents of each domain. The β-sheet domain contains a β-sheet composed of the S1 to S5 Avp2 β-strands and the pVIcT β-strand, it also contains the central α-helix H2 and the peripheral α-helix, H3. The α-helix domain is composed of the N-terminal α-helix, H1, a central α-helix, H4, the C-terminal α-helices, H5, H6 and H7, and the two central β-strands, S6 and S7. Situated within the β-sheet domain are the active site residues, His54 and Glu71, whilst Gln115 and Cys122 are found within the α-helix domain. Therefore it has been proposed that pVIcT locks Avp into its correct conformation, thus bringing the active site residues into their correct orientation (Mangel et al., 1996).

The crystal structure of the protease in the presence of its activating peptide has given much insight into the structure of the active Avp. However a comparison with the newly crystallised protease in the absence of pVIcT (Baniecki et al., 2002) promises to shed light on the activation mechanism of the protease. Likewise, a crystal structure of Avp2 with a substrate trapped in the putative substrate cleft would provide us with the knowledge of how and where substrates bind.

Figure 1.3.3: Topological Diagram of the Secondary Structure of Avp2-pVIcT
The β-strands (labelled S1 to S7, N- to C-terminus) and α-helices (labelled H1 to H7, N- to C-terminus) are shown by arrows and rods, respectively. The length of the arrow or rod is indicative of the number of residues found in the structure. The structure is divided into two distinct domains, which are connected by two loops from Avp2 and pVIcT, both of which are coloured green. The β-sheet domain (highlighted in blue) contains a β-sheet composed of the Avp2 β-strands, S1 through S5 and the β-strand from pVIcT, it also contains the central α-helix H2 and the peripheral α-helix, H3. The α-helix domain (highlighted in red) is composed of the N-terminal α-helix, H1, a central α-helix, H4, the C-terminal α-helices, H5, H6 and H7, and the two central β-strands, S6 and S7. Situated between the two domains are the active site residues, His54, Glu71, Gln115 and Cys122. (adapted from McGrath et al., 2003).



1.3.4 Substrate Specificity and Drug Design

The elucidation of the substrate specificity motifs for Avp began in the early 1980's, after a number of virally-encoded substrates were investigated. The studies determined that cleavage occurred at either an Ala-Gly bond (Tremblay et al., 1983) or a Gly-Ala bond (Akusjarvi & Perrson, 1981; Hannan et al, 1983; Sung et al., 1983; Tremblay et al, 1983). Moreover, it was suggested that substrate specificity may not only be dependent upon the peptide bond but also on the secondary structure surrounding the bond (Tremblay et al., 1983). The motifs were later extended to (M,L,I)XGX-G and (M,L,I)XGG-X (Webster et al., 1989b; Anderson, 1990). Following this, cleavage of both synthetic substrates and pVI by purified recombinant Avp2 showed that the GX-G motif was hydrolysed more quickly than the GG-X motif (Diouri et al., 1996).

As the cleavage motifs until this point were determined using the human Group C adenoviruses (Ad2 & Ad5), it was unknown whether the same motifs would be cleaved by all adenoviruses. Therefore, cleavage of human Ad2 pVII by human species B (Ad3 & Ad7), C (Ad2 & Ad5), D (Ad9) and E (Ad4) were analysed (Houde & Weber, 1990b). pVII was cleaved by all of the species, thus it was proposed that the protease and the substrate cleavage motifs are highly conserved.

In the past few years Ad4 pTP has been shown to be cleaved at QRGF-G by both human Ad2 and Ad4 protease (Webster et al., 1997), and ovine Ad7 pVII was shown to be cleaved by ovine Ad7 protease at both an MRAT-G and an NTGW-G site (Vrati et al., 1996). Together, these reports extended the general cleavage motif for Avp to (M,L,I,N,Q)X(A,G)X-G and (M,L,I)XGG-X. More recently in a review covering the genetic content and evolution of adenoviruses, an alignment of pX proteins from the four established genera of adenovirus was carried out (Davison et al., 2003). The alignment highlighted both confirmed and putative cleavage sites and from this, proposed that the

motifs be changed to (M,L,I ,N,Q,V)X(A,G)X-G and (M,L,I,V,F)XGG-X. However those residues which are underlined are merely putative P4 residues, thus until these sites have been proven to be cleaved by Avp the accepted general motifs are (M,L,I,N,Q)X(A,G)X-G and (M,L,I)XGG-X. Nevertheless as this thesis focuses on the protease of human Ad2, the only proven motifs for this particular protease are (M,L,I,Q)XGX-G and (M,L,I)XGG-X.

As Avp has been suggested to have a similar substrate specificity throughout the serotypes (Houde & Weber, 1990b) investigations into irreversible inhibitors which bind at the putative substrate cleft (Ding et al., 1996; McGrath et al., 2003) have been investigated. Cornish et al., 1995, designed tetrapeptides composed of LAGG, with a C-terminal ester, amide, nitrile and dimethylacetal, with the C-terminal ester compounds showing greatest potential. More recently, an irreversible inhibitor for Avp2 was discovered, 2,4,5,7-tetranitro-9-fluorenone (TNFN) and has been proposed to be an ideal lead compound (Pang et al., 2001).

Likewise, the sites at which the two cofactors, pVIcT and DNA bind to the protease have been suggested to be excellent drug targets (Mangel et al., 2001). Anti-adenoviral drugs raised against these sites would provide one of two effects - either they would prematurely activate the protease, causing early cleavage of viral structural proteins, therefore resulting in the loss of viral assembly, or alternatively they would inactivate the protease so that neither viral nor cellular proteins were cleaved, therefore disrupting the viral life cycle.

Drugs raised against any of the three sites described above would require an examination of the Avp2 residues involved in their binding. An initial analysis of residues putatively involved in binding the substrate and pVIcT were investigated in this work.

1.4 Adenovirus Polypeptide VIII

The precursor protein, pVIII is cleaved by adenovirus protease to form mature VIII. Once produced, polypeptide VIII maintains the stability of the adenoviral capsid along with another three cement proteins. However the cleavage site(s) required to produce mature VIII have not been determined and therefore this work aimed to identify potential sites. This sub-chapter discusses the little amount of knowledge which has been acquired about polypeptide VIII over the past three decades.

1.4.1 Maturation of pVIII to VIII

The capsid of mature adenovirus is composed of two capsomeres hexon and penton capsomeres and four cement proteins (IIIa, VI, VIII and IX). Although much work has been published describing the capsomeres of the virus, the cement proteins have not been well studied. This is surprising as the positions of the proteins in the capsid raise the possibility of drugs targeted against these sites which would under- or over- stabilise the virus, ultimately leading to viral destruction.

Of these proteins, polypeptide VIII is particularly interesting as it has been found to be essential for viral thermostability (Lui et al., 1985). Studies of the polypeptide have suggested that it lies on the interior of the virus (Everitt et al., 1973; Stewart et al., 1993), bound to the base of the hexon capsomere (Everitt et al., 1973). However, the sequence of the mature polypeptide has never been identified although it is known to be cleaved from the precursor protein, pVIII. Therefore, a major aim of this study was to investigate which site(s) were required to be cleaved by Avp to produce mature VIII.

1.5 AIMS

The proteins investigated in this work were the adenovirus protease and the structural polypeptide VIII. This sub-chapter discusses the aims of the thesis in terms of these proteins.

1.5.1 Aims

The major aim of this work was to further our knowledge of Avp2 substrate specificity by investigating both the residues involved in the S2 and S4 substrate specificity pockets of Avp2 and also by examining the cleavage of pVIII to VIII.

The residues involved in both pockets were to be investigated using several bioinformatic techniques, including a primary structure alignment of Avp proteins, a primary structure alignment of the families which form the CE clan of cysteine proteases and a tertiary structure superimposition of Avp2 and Ulp1, with its Smt3 substrate. Avp2 and Ulp1 are members of the CE clan of cysteine proteases and they were to be superimposed to produce an Avp2-Smt3 substrate model. Following assignment of the putative residues involved in forming the S2 and S4 substrate specificity pockets an alanine screen was to be carried out.

In addition to this work, the binding of pVIcT and the requirement for the C-terminal tail of the protease were to be analysed using further bioinformatic and mutational analysis. Several tagged vector systems were also to be used to purify large quantities of Avp2 for use in kinetic studies and crystallisation trials.

Likewise human Ad2 pVIII was to be cloned into a number of tagged vector systems in the hope of purifying the polypeptide. Furthermore, primary sequence alignments, an in vitro trascription-translation system and analysis of purified Ad2 were to be employed to elucidate the pVIII site(s) cleaved by Avp2 to produce mature VIII.

CHAPTER 2
MATERIALS AND METHODS

This Chapter describes all of the techniques used in this work.

2.1 BIOINFORMATICS

2.1.1 Alignment of Primary Amino Acid Sequences

The sequence of a single protein was searched for using NCBI (http://www.ncbi.nlm.nih.gov/). The sequence was then pasted into the WU-BLAST2 programme at EBI, Hinxton (http://www.ebi.ac.uk/blast2/) where a similarity search was carried out. Each of the sequences from this search were pasted into the CLUSTALW alignment programme at EBI, Hinxton (Thompson et al., 1994;http://www.ebi.ac.uk/clustalw/) and an alignment was performed.

2.1.2 Tertiary Structure Analysis

Three methods described below were used to analyse proteins at the tertiary level.

2.1.2.1 Tertiary Structure Superimposition

Tertiary structures were superimposed using the Swiss-PDB Viewer Computer Programme (Guex & Peitsch., 1997; http://www.expasy.org/spdbv/). A similar residue from each protein structure was selected (specified in Chapter 3) in the Control Panel. 'Fit molecules (from selection)' was then chosen from the pull-down menu of 'Fit' in the Toolbar, and then 'all atoms' was selected. The computer package automatically superimposed all of the atoms of the two selected residues.

To improve the superimposition, 'Improve fit' was selected from 'Fit' pull-down menu in the Toolbar. As the fit improved, the residues which superimposed changed. Therefore 'Improve fit' was selected until all residues remained constant.

The superimposition was then viewed in the RASMOL Version 2.6 computer package (Sayle & Milner-White., 1995).

2.1.2.2 Using the Radius Function of Swiss-PDB Viewer

The Radius function of the Swiss-PDB viewer Computer Programme (Guex & Peitsch, 1997; http://www.expasy.org/spdbv/) was used to analyse which residues of a molecule were within a set radius of another residues' α-carbon. The 'Radius symbol' was chosen from the Toolbar and then the α-carbon of the selected residue was clicked on using the pointer. Using the pop-up menu, 'Display only groups that are within' was selected and the radius was specified in Angstroms. Those residues which were within the chosen radius were shown in the Control Panel

2.1.2.3 Modelling Tertiary Structures

The protein sequence was searched for using NCBI (http://www.ncbi.nlm.nih.gov/). The sequence was then pasted into the First Approach Mode at SwissModel (http://swissmodel.expasy.org) and sent to be processed.

SwissModel compared this primary sequence to all crystallised proteins in its databank. If a similar sequence was found then a tertiary structure model was built of the query sequence.

2.2 Cloning of Recombinant Proteins

2.2.1 Polymerase Chain Reaction (PCR)

Three methods of PCR were utilised and are described below. The method used in each case was dependent upon where the mutation was to be introduced into the sequence.

2.2.1.1 Method 1 – Forward Primer and Reverse Primer

This method of PCR allowed a gene to be cloned from a template either without mutations, or with mutations close to the N- or C-terminus. The PCR mixture contained 1µl mini-prep DNA (Method 2.2.7), 10µl x10 VENT DNA polymerase buffer (NEB), 250µM dNTP's (Promega), 0.5µM forward and reverse primers, 0.5µl VENT DNA polymerase (NEB) made up to 100µl water. Temperatures and length of time for Denaturation, Annealing and Extension varied and are specified in Chapter 3.

2.2.1.2 Method 2 – Forward Mega-primer and Reverse Primer

This method was used for mutagenesis and required a two step PCR. Firstly a PCR product was produced using a forward primer and an internal reverse primer (containing base substitutions) using Method 2.2.1.1. This large forward primer (Mega-primer) was then Gel Extracted (Method 2.2.3) and used in a further PCR. The second PCR mixture was the same as the first, except 5µl of the Mega-primer was used as the forward primer, along with a reverse primer. Temperatures and length of time for Denaturation, Annealing and Extension varied and are specified in Chapter 3.

2.2.1.3 Method 3 – Forward Primer and Reverse Mega-primer

This method was used for mutagenesis and required a two step PCR. Firstly a PCR product was produced using an internal forward primer (containing base

substitutions) and a reverse primer, using Method 2.2.1.1. This reverse Mega-primer was then Gel Extracted (Method 2.2.3) and used in a further PCR. The second PCR mixture was the same as the first, except 5µl of the Mega-primer was used as the reverse primer, along with a forward primer. Temperatures and length of time for Denaturation, Annealing and Extension varied and are specified in Chapter 3.

2.2.2 Agarose Gels

100ml of TAE electrophoresis buffer (40mM Tris, 20mM Acetic acid, 1mM EDTA, pH8) was prepared. 50ml of this was retained as buffer whilst the other 50ml was used to make a 0.9% (w/v) Agarose solution. The Agarose in solution was melted, cooled to 60°C and then Ethidium Bromide was added to a final concentration of 0.5μg/ml. Agarose solution was poured into a gel rig (Bioscience Services) to a depth of 3-5mm, a comb was set in place and the gel was left to solidify for 30-45min at 25°C. The comb was then removed and the gel was overlayed with TAE buffer to a depth of 1mm.

The DNA sample was mixed with loading buffer (0.04% (w/v) Bromophenol blue, 7% (w/v) sucrose). The mixture was loaded onto the gel and a 1Kbp DNA ladder (Promega) was separated alongside it, at a constant voltage of 70V.

2.2.3 DNA Gel Extraction

The procedure for DNA gel extraction was based on the method described in QIAGEN QIAquick Spin Handbook, July 2002. The QIAquick Gel Extraction Kit Protocol used the purification of DNA from 70bp to 10Kbp.

Firstly, a DNA band was excised from an Agarose Gel and then 3 volumes of Buffer QG (QIAGEN) was added to 1 volume of gel. The gel slice was incubated at 50°C for 10min or until the gel had dissolved (the tube was vortexed every 2-3min to help the gel dissolve).

Once dissolved, 1 gel volume of Isopropanol was added to the sample and the sample was mixed. The DNA was bound to the QIAquick spin column (QIAGEN) by centrifuging in a MSE Microcentaur centrifuge at 13,000rpm for 1min. The flow-through was discarded and a further 500µl of Buffer QG (QIAGEN) was passed through the column at 13,000rpm for 1min. The column was then washed by centrifuging with 750µl of Buffer PE (QIAGEN) at 13,000rpm for 1min, and the flow-through was discarded. The column was then centrifuged for a further 1min at 13,000rpm to remove all traces of Ethanol.

The DNA was eluted by applying 50µl of water to the column. The column was left to stand for 1min and then centrifuged for 1min at 13,000rpm. The DNA was aliquoted and stored at –20°C.

2.2.4 Restriction Digest of Insert and Plasmid

The inserts and plasmids used for the protocol were produced by PCR (Method 2.2.1) and mini-prep (Method 2.2.7), respectively. The inserts and plasmids were digested by using one of two methods – either, they were digested by using both enzymes together i.e. a double digestion, or alternatively, they were digested using two separate digestions with a Gel Extraction (Method 2.2.3) in between.

The digestions involved mixing 50µl of insert or plasmid DNA with 5µl of either one or two restriction enzymes and 7.5µl of the appropriate buffer, this was then made up to 75µl with water. The solution was then incubated for 3hr at 37°C (in the case of NdeI the reaction was left for 12-16hr as it cleaves its restriction site slowly). After incubation, the DNA was Gel Extracted.

2.2.5 Ligation of Insert

In some cases, prior to ligation, the vector was treated with Calf Intestinal Alkaline Phosphatase (CIAP) (Promega). The function of the CIAP treatment was to dephosphorylate the restricted cloning sites of the vector and prevent religation, thereby reducing background. The dephosphorylation was setup with 35µl of digested mini-prep vector DNA, buffer (200mM Tris-HCl, 4mM $MgCl_2$, 0.4mM $ZnCl_2$, 4mM Spermidine, pH9.3) (Promega) and 1µl of CIAP (Promega). This was then incubated for 30min at 37°C. The reaction was stopped by adding 25mM EDTA and then incubating for 20min at 65°C. The vector could then be used in the ligation reaction.

The ligation was set up with 10µl restricted insert, 1µl restricted plasmid, 6µl water, T4 DNA ligase buffer (1mM $MgCl_2$, 1mM Ditiothreitol, 0.1mM ATP, 0.25%(w/v) BSA, 5mM Tris-HCl, pH 7.5) (NEB) and 1µl T4 DNA ligase (NEB). As a control for religation of the plasmid, the same reaction was set up without the restricted insert and the solution was made up to 20µl with water. The reactions were left to incubate at 4°C for 12-16hr.

2.2.6 Transformation of Cells

A colony of JM109 or DH5α cells were selected from a stock plate (Method 2.2.10.1) and this was used to inoculate 5ml of LB/antibiotic (LB medium contained 1% bacto-tryptone, 0.5% bacto-yeast, 1% NaCl, pH7.0), which was grown for 12-16hr at 37°C with shaking. The 5ml solution was used to inoculate 50ml of LB/antibiotic, which again was grown at 37°C with shaking until an exponential stage was reached (OD_{600}=0.4-0.6). The bacterial culture was split between two tubes and centrifuged at 4°C, 3000rpm for 10min in a Howe Sigma Laborzentrifugen 3K10 centrifuge. Subsequently each pellet was re-suspended in 5ml of filter-sterilised, ice cold 0.1M $MgCl_2$, centrifuged as before, and again re-suspended in 2ml of filter-sterilised, ice-cold 0.1M $CaCl_2$. The samples were combined and stored at 4°C, where they remained competent for 4-5 days but were at their optimum activity within the first 24hr.

10-200ng of plasmid DNA (Methods 2.2.7 & 2.2.8) were added to 100µl of competent cells, mixed well and kept on ice for 30min. The cells were heat-shocked for 90sec at 42°C, and then left on ice for 5min. 900µl of LB was added to the cells and they were left at 37°C for 1hr to allow for cell recovery. 100µl of the cells were then smeared on to a LB-agar/antibiotic plate (Method 2.2.10.1), the remaining 900µl were centrifuged at 13,000rpm for 10min in a MSE Microcentaur centrifuge, 800µl of the media was removed and the cells were re-suspended in the remaining 100µl, which was also spread onto a LB-agar/antibiotic plate (Method 2.2.10.1). Both plates were incubated for 12-16hr at 37°C.

2.2.7 Mini-prep

The protocol was based on the method described in the QIAGEN QIAprep Miniprep Kit Handbook, March 2002. The QIAquick Miniprep Kit used a microcentrifuge to purify up to 20μg of high-copy plasmid DNA from a 1-5ml overnight culture of *E. coli* grown in LB/antibiotic.

Transformed JM109 or DH5α cells were plated onto LB-agar/antibiotic plates (Method 2.2.10.1) and left to grow for 12-16hr at 37°C. Single colonies were picked from the plate and used to inoculate 5ml of LB/antibiotic which were grown for 12-16hr at 37°C. 2.5ml of this, was centrifuged at 5,000rpm for 10min in a Howe Sigma Laborzentrifugen 3K10 centrifuge. The supernatant was discarded and the cell pellet was re-suspended in 250μl Buffer P1 (QIAGEN). 250μl of buffer P2 (QIAGEN) was added, the solution was inverted 4-6 times and then left to incubate at 25°C for 5mins. 350μl Buffer N3 was added, the solution was inverted 4-6 times and then centrifuged at 13,000rpm for 10min in an MSE Microcentaur centrifuge. The supernatant was applied to a QIAprep column and centrifuged at 13,000rpm for 1min and the flow-through discarded. The column was washed with 500μl of Buffer PB by centrifuging at 13,000rpm for 1min and the flow-through was discarded. The column was washed again with 750μl of Buffer PE and centrifuged at 13,000rpm for 1min, the flow-through was discarded and the column was centrifuged for a further 1min to remove residual buffer. The QIAprep column was placed in a clean 1.5ml microcentrifuge tube, 50μl water was applied to the column and the column was left to stand for 1min before being centrifuged for 1min at 13,000rpm. The sample was aliquoted and stored at –20°C.

2.2.8 Determination of DNA Concentration and Purity

To calculate double-stranded DNA (dsDNA) concentration, the DNA sample was diluted 1000-fold and the Absorbance was measured at 260nm. A DNA concentration of 50μg/ml gave an Abs_{260}=1 and therefore from this, the concentration in μg/ml could be calculated.

Similarly, the purity of the dsDNA sample was determined by dividing Abs_{260} by Abs_{280}. A relatively pure sample gave a value of around 1.8, whilst a sample contaminated with either phenol or protein produced a value significantly less than this.

2.2.9 DNA sequencing

The Automated DNA sequencing unit based at the University of St. Andrews was employed to confirm that the DNA fragment had been ligated in the appropriate conformation into the correct vector.

The DNA sequencing unit utilised Dye Terminator sequencing technology on an ABI PRISM^{TM} 377 DNA Sequencer. This form of sequencing required 500ng of purified DNA (Method 2.2.7 & 2.2.8) to act as a template for the PCR, 5pmol of an appropriate primer and the ABI PRISM^{R} BigDye^{TM} Terminator v3.1 Ready Reaction Cycle Sequencing Kit (Applied Biosystems).

The Ready Reaction Cycle Sequencing Kit contained buffer, the Ampli Taq FS DNA polymerase, deoxynucleotides and fuorescently labelled dideoxynucleotides. This mixture was added to the DNA template and Primer and a PCR reaction was carried out – hot start (96°C for 5min), 25 cycles of Denaturation (96°C for 10sec), Annealing (50°C for 5sec) and Extension (60°C for 4min).

2.2.10 Maintaining Plasmid Stocks

All plasmid stocks were retained in three different forms to reduce the risk of loss or contamination.

2.2.10.1 Method 1 – Agar Stocks

Luria Bertani-agar (15g of bacto-agar to 1L of LB) was melted, and equilibrated to 55°C. The appropriate antibiotic was added (Table 2.2.1), and the LB-agar was plated out. Once solidified plates were stored at 4°C for a maximum of two weeks prior to usage.

Plate stocks were produced by spreading colonies from a glycerol stock onto an LB-agar/antibiotic plate. The colonies on the plates were grown for 12-16hr at 37°C and then stored at 4°C for no longer than a month.

Table 2.2.1: Working Concentrations of Antibiotics

Antibiotic	Working Concentration (μg/ml)
Carbenecillin	50
Chloramphenicol	25
Spectinomycin	20
Kanamycin	50

2.2.10.2 Method 2 – Glycerol Stocks

A colony from a stock plate (Method 2.2.10.1) was used to inoculate 2ml of LB/antibiotic (Method 2.2.6) and was grown for 12-16hr at 37°C. The inoculated solution was mixed with glycerol to a final concentration of 15% (v/v). This was aliquoted in to four storage vials and stored at -70°C. Glycerol stocks were used for long term storage.

2.2.10.3 Method 3 – DNA Stocks

A Deoxyribonucleic acid stock, produced by Mini-prep (Method 2.2.7) was stored at –20°C due to the possibility of agar and glycerol stocks becoming contaminated.

2.3 Peptide Synthesis and Purification

2.3.1 Peptide Synthesis and Purification

Dr. Graham Kemp at the University of St. Andrews, St. Andrews synthesised all peptides used in this work, using the fluorenylmethoxycarbonyl-polyamide procedure. The peptides were then purified using reverse-phase High-Performance Liquid Chromatography (HPLC). Following this, the peptide was analysed using Mass Spectrometry (Method 2.4.7.1).

2.4 Analysing Proteins and Peptides

2.4.1 Mini-gels

Two mini-gel techniques were used in this work to analyse proteins and peptides. The protocols for each are given below.

2.4.1.1 Method 1 - SDS-PAGE Mini-gels

SDS-PAGE mini-gels consist of two layers, the separating gel and stacking gel. The SDS-PAGE mini-gels used either a 15% separating gel (15% (w/v) Acrylamide, 0.09% (w/v) Bisacrylamide, 0.1% (w/v) SDS, 0.08% (v/v) TEMED, 0.03% (w/v) Ammonium Persulphate, 0.4M Tris-HCl, pH8.7) or a 20% separating gel (20% (w/v) Acrylamide, 0.2% (w/v) Bisacrylamide, 0.1% (w/v) SDS, 0.07% (v/v) TEMED, 0.03% (w/v) Ammonium Persulphate, 0.4M Tris-HCl, pH8.7), as specified in Chapter 3. When prepared, the separating gel solution was poured into the gel rig, overlayed with water saturated butanol and left to solidify for 30-45min at 25°C.

Once set, the butanol was removed and replaced with stacking gel solution (5% (w/v) Acrylamide, 0.14% (w/v) Bisacrylamide, 0.1% (w/v) SDS, 0.2% (v/v) TEMED, 0.05% (w/v) Ammonium Persulphate, 0.12M Tris-HCl, pH6.9). Plastic combs were inserted into the stacking gel solution to form wells and the solution was left to solidify for 10-15min at 25°C.

Once set, the combs were removed and the rig was placed in Chamber buffer (0.3% (w/v) Tris, 1.4% (w/v) Glycine, 0.1% (w/v) SDS). 20µl protein samples were mixed with 5µl Reducing solution (10% (w/v) SDS, 25% (v/v) Glycerol, 0.05% (w/v) Bromophenol Blue, 25% (v/v) β-Mercaptoethanol, 0.25M Tris-HCl, pH 6.9). 10µl of the mixture was loaded onto the wells and a constant current of 40mA was passed through the gels for 50min.

2.4.1.2 Method 2 – Tris-Tricine Mini-gels

Tris-tricine mini-gels consist of two layers, the separating gel (10% (w/v) Acrylamide, 0.4% (w/v) Bisacrylamide, 0.1% (w/v) SDS, 10% (v/v) Glycerol, 0.02% (w/v) Ammonium Persulphate, 0.05% (v/v) TEMED, 1M Tris-HCl, pH8.45) and the stacking gel (3% (w/v) Acrylamide, 0.09% (w/v) Bisacrylamide, 0.07% (w/v) SDS, 0.04% (w/v) Ammonium Persulphate, 0.2% (v/v) TEMED, 0.7M Tris-HCl, pH8.45). The separating gel solution was poured into the gel rig, overlayed with water saturated butanol and left to solidify for 30-45min at 25°C.

Once set, the butanol was removed and replaced with stacking gel. Plastic combs were inserted into the stacking gel solution to form wells and the solution was left to solidify for 10-15min at 25°C.

Once set, the combs were removed and the rig was placed in a chamber. The Cathode buffer used was 0.1M Tris, 0.1M Tricine, 0.1% SDS, whilst the Anode buffer used was 0.2M Tris-HCl, pH 8.9. 20µl protein samples were mixed with 5µl Reducing solution (10% (w/v) SDS, 25% (v/v) Glycerol, 0.05% (w/v) Bromophenol Blue, 25% (v/v) β-Mercaptoethanol, 0.25M Tris-HCl, pH 6.9). 10µl of the mixture was loaded onto the wells and a constant current of 40mA was passed through the gels for 50min.

2.4.2 Staining Mini-gels

The methods describing two different techniques for staining mini-gels are given below.

2.4.2.1 Method 1 – Coomassie Blue Staining

A mini-gel (Method 2.4.1) was stained in ~ 20ml of Coomassie Brilliant Blue R250 stain (0.25% (w/v) Coomassie Brilliant Blue R250, 45% (v/v) Methanol, 10% (v/v) Acetic acid) for 10-15 min, and were then washed several times in destain (25% (v/v) methanol, 7.5% (v/v) Acetic acid).

2.4.2.2 Method 2 – Silver Staining

A mini-gel fixed by Coomassie Blue staining (Method 2.4.2.1) was washed with 30% (v/v) Ethanol (3 x 5min) and then distilled Water (2x 5min). The gel was soaked for a further 1min in 1.4mM Sodium dithionite and then washed in distilled Water (2 x 1min). It was then soaked for 5min in 12mM Silver nitrate, 1mM formaldehyde. Following this, it was rinsed in distilled water (3 x 20secs) and then developed in 0.57M Na_2CO_3, 6mM Formaldehyde, 20μM Sodium Dithionite. 3.5% (v/v) Acetic acid was added and the gel was shaken for 20min to inhibit further development. The gel was then rinsed in distilled Water (4 x 5min).

2.4.3 Western Blotting

Protein samples were separated on an SDS-PAGE mini-gel (Method 2.4.1.1) and the proteins were then transferred onto the solid support, Polyvinyl difluoride (PVDF) membrane in the presence of the transfer buffer (48mM Tris, 39mM Glycine, 20% (v/v) Methanol, 0.0375% (w/v) SDS) for 2hrs, at 300mA, 100V. Once transferred, the membrane was blocked with blocking buffer (250ml PBS (140mM NaCl, 2.7mM KCl, 8mM Na_2HPO_4, 1.5mM KH_2PO_4, pH 7.4) , 5% (w/v) Milk powder (TESCO Superstores Plc.), 0.1% (v/v) Polyoxyethylene Sorbitan Monolaurate) for 10min.

Following blocking, the PVDF was incubated for 30min in 20ml blocking buffer, 5µl R11 primary antibody (1 in 4000 dilution), which was raised against the N-terminal 14 residues of the protease (Webster & Kemp., 1993). The membrane was washed for a further 2 x 10min in blocking buffer before being incubated for 30min with 20ml blocking buffer, 4µl Horseradish peroxidase (HRP) labelled anti-rabbit IgG secondary antibody (1 in 5000 dilution).

After incubation, the membrane was washed 2 x 10min with blocking buffer and then rinsed for 5min with TBBS (blocking buffer with no milk powder). The blot was then soaked in 2ml Buffer 1 (2.5mM Luminol (Fluka), 0.369mM p-coumaric acid, 0.1M Tris-HCl, pH8) and 2ml Buffer 2 (0.0192% (v/v) Hydrogen peroxide and 0.1M Tris-HCl, pH8.5) together. The membranes were drained on filter paper, placed between two acetate sheets, exposed to X-ray film (Fuji Medical X-Ray Film RX) and the film was developed using a Kodak M35 X-OMAT Processor.

To compare the bands on the blot to the bands separated on the gel, the PVDF membrane was stained with Amido Black solution (0.1% (w/v) Amido Black, 40% (v/v) Methanol, 1% (v/v) Acetic acid) for 1min and then rinsed in distilled water for 30min before being air dried.

2.4.4 Evaluating Overall and Soluble Expression

Two methods for analysing expression of proteins in *E. coli* were used and these are described below.

2.4.4.1 Method 1 – SDS-PAGE Mini-gels

A colony from a stock plate (Method 2.2.10.1) was used to inoculate 5ml of LB/antibiotic and this was grown for 12-16hr at 25°C. The 5ml culture was used to inoculate 20ml of LB/antibiotic and this was grown until OD_{600} = 0.4-0.6 The culture was split into 7 x 3ml fractions and induced with 0mM, 0.02mM, 0.04mM, 0.08mM, 0.1mM, 0.2mM, 0.4mM IPTG for 6hr at 25°C.

To analyse overall expression a 1ml sample was removed from each induction. These were then centrifuged at 13,000rpm for 10min in an MSE Microcentaur centrifuge and the pellets were retained. The pellets were mixed with 50µl Reducing solution (10% (w/v) SDS, 25% (v/v) Glycerol, 0.05% (w/v) Bromophenol Blue, 25 (v/v) β-Mercaptoethanol, 0.25M Tris-HCl, pH 6.9). 5µl of each sample was separated on a 15% SDS-PAGE mini-gel (Method 2.4.1.1) and the gel was stained with Coomassie Blue (Method 2.4.2.1).

Similarly soluble/insoluble expression was analysed by removing a 1ml fraction from each induced sample. These were centrifuged at 13,000rpm for 10min and the pellets were re-suspended in 100µl Suspension buffer (refer to Chapter 3 for buffers used). The samples were then sonicated for 6 x 10sec with 1min on ice between sonications. The samples were further centrifuged at 13,000rpm for 10min, the supernatant (soluble fraction) was mixed with 50µl Reducing solution and the pellet (insoluble fraction) was re-suspended in 100µl Reducing solution. 5µl of each sample

were then separated on a 15% SDS-PAGE mini-gel and the gel was stained with Coomassie Blue.

2.4.4.2 Method 2 – Western Blotting

In some cases analysing a single protein band on an SDS-PAGE mini-gel (Method 2.4.1.1) is impractical as the protein band is faint. Therefore a Western blot technique was developed. This method produced a single band which could be analysed using the Gel-doc programme.

The saturation point of the X-ray film was first investigated. 5µl of purified Avp2ET11c (Method 2.5.4) was separated on four different 15% SDS-PAGE mini-gels and then Western blots were carried out using the R11 primary antibody (Method 2.4.3). The exposure time to the X-Ray film was 0min, 0.25min, 0.5min, 0.75min, 1min, 1.25min, 1.5min, 1.75min, 2min and the OD of the band minus the OD of the background was studied under white light using the Gel-doc programme.

The mean OD of the band $^{+}$/- standard deviation at each time point was plotted against time. From this, the saturation point of the X-ray film was determined and an OD below the saturation point was selected.

Using this method a number of variables which effect protein expression were investigated, and these are specified in Chapter 3.

2.4.5 Determining Protein Concentration

Two methods described below were used to determine protein concentrations using a Soya Bean Trypsin Inhibitor (SBTI) standard curve.

2.4.5.1 Method 1 – SDS PAGE Mini-gels

5µl, 10µl, 15µl of 0.1mg/ml SBTI, 5µl of 1mg/ml SBTI and 10µl of the unknown protein sample (diluted appropriately) were loaded onto a 15% SDS-PAGE mini-gel (Method 2.4.1.1). The gel was stained with Coomassie Blue (Method 2.4.2.1) and the OD of each band minus the background OD was calculated using the Gel-doc system. From this a standard curve was produced and the unknown protein concentration was determined.

2.4.5.2 Method 2 – Bradford Assay

0mg/ml, 0.25mg/ml, 0.5mg/ml, 0.75mg/ml and 1mg/ml SBTI standards were made and the unknown protein sample was diluted appropriately. 30µl of each sample was added to 970µl of Bradford reagent (Sigma) and the OD was measured at 595nm. From this a standard curve was produced and the unknown protein concentration was determined (Bradford., 1976).

2.4.6 Protein and Peptide Sequencing

The automated protein sequencing unit based at the University of St. Andrews used an Applied Biosystems Procise 491 Protein Sequencer. Samples were sent to protein sequencing either in solution, blotted onto Polyvinyl difluoride (PVDF) membrane or following an in-gel extraction. The methods are described below.

2.4.6.1 Method 1 – Samples on PVDF Membrane

Protein sample separated on an SDS-PAGE mini-gel (Method 2.4.1.1) was blotted onto Hyperbond PVDF (Porton Instruments, Incorporated) (Method 2.4.3). The blotting procedure was as is stated in Method 2.4.3, except the buffer used was CAPS buffer (10% (v/v) Methanol, 10mM CAPS-NaOH, pH 11) . The buffer was changed because the buffer in Method 2.4.3 contained Glycine.

After blotting, the PVDF membrane was stained with Amido Black solution (0.1% (w/v) Amido Black, 40% (v/v) Methanol, 1% (v/v) Acetic acid) for 1min and then rinsed in distilled water for 30min before being air dried. This allowed the bands to be visualised and then excised prior to protein sequencing.

2.4.6.2 Method 2 – Samples Produced by In-gel Extraction

Peptide sample was separated either on an SDS-PAGE mini-gel (Method 2.4.1.1) or a Tris-Tricine mini-gel (Method 2.4.1.2). Once separated on the gel, the band was visualised by staining with Coomassie Blue (Method 2.4.2.1), or by placing under a UV light. The band of interest was excised from the gel and the peptide was extracted from the gel twice with 50µl of 60% (v/v) Acetonitrile, 0.1% (v/v) TFA for 20min. The supernatants were combined and then dried using a Savant Speedvac SPD111V. The sample was then sent to be sequenced.

2.4.7 Mass Spectrometry

The Mass Spectrometry unit based at the University of St. Andrews used a Micromass TofSpec-2E Mass Spectrometer to analyse protein samples either in solution or extracted from an SDS-PAGE mini-gel . The machine is a Time of Flight Mass Spectrometer which utilises Matrix Assisted Laser Desorption Ionisation and is therefore called a MALDI-TOF spectrometer.

Samples were sent to Mass Spectrometry either in solution or as a gel slice. The procedures required for the preparation of each of these types of sample are discussed below.

2.4.7.1 Method 1 – Samples in Solution

Peptides in solution which were analysed by Mass Spectrometry were those synthesised and purified as in Method 2.3. MALDI-TOF was used to identify that the peptide was of the expected mass and also that it was purified after Reverse Phase Chromatography Purification.

2.4.7.2 Method 2 – Samples from an SDS-PAGE Mini-gel

Protein sample was separated on an SDS-PAGE mini-gel (Method 2.4.1.1), which was stained with Coomassie Blue (Method 2.4.2.1). The protein band was then excised and submitted to the Mass Spectrometry Unit.

2.5 Recombinant Protein Expression and Purification

2.5.1 Binding Tagged Proteins to Nickel Beads or Amylose Resin

A colony of cells was picked from an LB-agar/antibiotic plate (Method 2.2.10.1) and used to inoculate 5ml of culture. The culture was grown at 25°C for 12-16hr and then induced with 0.4mM IPTG at 25°C for 6hr. A 1ml sample was taken pre- and post-induction, centrifuged at 13,000rpm for 10min in an MSE Microcentaur centrifuge and the cell pellet was then re-suspended in 50µl Reducing solution (10% (w/v) SDS, 25% (v/v) Glycerol, 0.05% (w/v) Bromophenol Blue, 25% (v/v) β-Mercaptoethanol, 0.25M Tris-HCl, pH 6.9).

A further 3ml sample was removed post-induction, centrifuged at 13,000rpm for 10min and re-suspended in 500µl buffer (Chapter 3). The sample was then sonicated for 6 x 10sec with 1min on ice between sonications. The sample was centrifuged and the cell pellet (insoluble fraction) was re-suspended in 50µl Reducing solution. 50µl of the supernatant (soluble fraction) was mixed with 50µl Reducing solution. The remaining 450µl of supernatant was incubated at 25°C for 30min with either 50µl of Nickel beads or Amylose resin (Chapter 3).

After 30min, the Protein-bead/resin complex was washed with 3 x 500µl buffer (Chapter 3) and the pellet was re-suspended in 50µl Reducing solution. 5µl of each sample was separated on a 15% SDS-PAGE mini-gel (Method 2.4.1.1), which was stained with Coomassie Blue (Method 2.4.2.1) and then Western blotted if appropriate (Method 2.4.3).

2.5.2 Bacterial Growth and Induction of Protein Expression

A colony of host cells was picked from an LB-agar/antibiotic plate (Method 2.2.10.1) and was used to inoculate 10ml LB/antibiotic which was grown for 12-16hr at 37°C.

Subsequently the cells were centrifuged at 5,000rpm for 10min in a Howe Sigma Laborzentrifugen 3K10 centrifuge. The pellet was re-suspended in 20ml of LB/antibiotic and this was added to 1l of LB/antibiotic.

The cells were grown with shaking at 37°C until OD_{600} = 0.4-0.6. They were then induced by the addition of IPTG (Chapter 3). After induction, the cells were harvested by centrifugation at 10,000rpm for 10min at 4°C in a JA14 rotor using a J2-21 Beckman Centrifuge. The pellet was washed, re-suspended in an appropriate buffer specified in Chapter 3 and then stored at –20°C prior to extraction.

2.5.3. Extraction of Protein from *E. coli*

Recombinant protein expressed in *E. coli* was extracted from the bacteria by two methods which are described below.

2.5.3.1 Method 1 – Freeze-Thaw Cycle

The pellet from Method 2.5.2 was thawed at 37°C then incubated with 0.125mg/ml Lysozyme for 15min at 37°C, freeze-thawed three times using liquid Nitrogen-37°C and then incubated in the presence of 0.4mg/ml DNase I and 12.5mM $MgSO_4$ at 37°C for 30min or until fluid. The solution was centrifuged at 13,000rpm for 10mins in a MSE microcentaur centrifuge and the supernatant was retained for purification.

2.5.3.2 Method 2 – Sonication

The pellet from Method 2.5.2 was thawed in an ice-water bath. The suspension was then sonicated for 8 x 15sec with 1min on ice between each sonication. The sample was centrifuged at 5,000rpm for 10min in a Howe Sigma Laborzentrifugen 3K10 Centrifuge and the supernatant was retained for purification.

2.5.4 Purification of Avp2ET11c

A method for native Avp2 purification by Ion Exchange Chromatography (IEC) was developed to purify the protease by Pollard, 2001. The purification method is summarised below and it has acted as a basis upon which WT and all of the mutant proteases discussed in this work were purified.

The 1 x 30cm Diethylaminoethyl (DEAE) -Sepharose, 1 x 10cm Heparin-Sepharose (HS) and 1 x 3cm Carboxymethyl (CM) -Sepharose columns were equilibrated with 50mM Tris-HCl, pH8 for 1hr prior at a flow rate of 1ml/min. The supernatant from Method 2.5.3.1 was then applied to the DEAE-Sepharose column for 18-20min at flow rate 1ml/min and protein elution was monitored at 280nm by a chart recorder. Following this, the HS and CM-Sepharose columns were connected, in the order DEAE-Sepharose, HS, CM-Sepharose, and the buffer was passed through the columns for a further 18-20min. The DEAE-Sepharose and HS columns were then disconnected.

Once base line was reached, Avp2 was eluted from the CM-Sepharose column using a gradient of 50mM Tris-HCl, pH8 to 50mM Tris-HCl, 200mM NaCl, pH8 over a 30min period.

2.5.5 Purification of Smt3ET28b

Following extraction (Method 2.5.3.2), the Smt3ET28b supernatant was applied to a 1 x 5cm Nickel bead Column. The column was washed with 15ml of Water at 1ml/min and then further washed with 30ml of 20mM Tris-HCl, 10mM Imidazole, pH8. The supernatant from Method 2.5.3.2 was applied to the Nickel column and the column was washed with 15ml of 20mM Tris-HCl, 20mM Imidazole, pH8. Following this, the Smt3ET28b was eluted into 2ml fractions at 1ml/min with 15ml of 20mM Tris-HCl, 250mM Imidazole, pH8.

2.5.6 Removal of Tags

To analyse the cleavage of a tag, a 100µl sample of protein was incubated with the appropriate amount of enzyme (Chapter 3) at 25°C with gentle shaking for 24hr. 20µl aliquots were taken at 2hr, 4hr, 8hr and 24hr and a control was set up for 24hr in the absence of the enzyme.

2.5.7 Regeneration of Beads/Resin

The two columns used in this work which could be regenerated and reused were Nickel columns and Amylose resin columns. The methods for the regeneration and subsequent storage of each are described below.

2.5.7.1 Method 1 – Regenerating Nickel Columns

To regenerate a Nickel column after use the following solutions were passed through the column, in the order - 3 column volumes of water, 3 column volumes of 6M GuCl, 0.2M Acetic acid, 3 column volumes of water, 3 column volumes of 100mM $NiSO_4$, 3 column volumes of 6M GuCl, 0.2M Acetic acid, 3 column volumes of 0.02% Sodium Azide.

2.5.7.2 Method 2 – Regeneration of Amylose Columns

To regenerate Amylose columns after use, the following solutions were passed through the column, in the order - 3 column volumes of water, 3 column volumes of 0.1% SDS, 1 column volume of water, 3 column volumes of column buffer (20mM Tris-HCl, pH7.4, 200mM NaCl, 1mM EDTA) and 3 column volumes of 0.02% Sodium Azide, Column buffer.

2.5.8 Concentrating Protein Samples

The protocol used was based on the method described in the Millipore Amicon Bioseparations Centricon^{R} Centrifugal Device for a YM-10 2ml filter handbook. 2ml of the sample was loaded onto the sample reservoir and then centrifuged using a Howe Sigma Laborzentrifugen 3K10 centrifuge at 5,000rpm until there was a retentate of ~200µl. The retentate was then collected by turning the sample reservoir up-side down and centrifuging at 1,000rpm for 2min.

2.6 Determination of Specific Activities

2.6.1 Assay

Initial rates of cleavage were calculated when less than 20% of the substrate was cleaved. Therefore because every protease investigated had a different level of activity, both the concentration of protease added to the assay mix and the length of time taken to digest 20% of the substrate had to be determined experimentally (Chapter 3).

The assay solution contained Protease, 100 Molar excess of Activating Peptide to Protease, 229μM AcLRGAGRSR made up to 50μl with 50mM Tris, 10mM EDTA, 2μM Pepstatin, 2mM AEBSF, pH8. At each time point, 10μl of the assay mix was removed and added to 10μl of 1% TFA which inhibited any further cleavage. The aliquot was then frozen for 12-16hr before being analysed by Capillary Electrophoresis (CE).

2.6.2 Capillary Electrophoresis

Prior to using Capillary Electophoresis (CE) to detect the cleavage of AcLRGAGRSR by Avp2, 80µl of water was added to the samples from Method 2.6.1. These samples were then separated on a BioFocus Capillary Electrophoresis System (Bio-Rad), in 0.1M H_3PO_4-NaOH, pH2.5 at 20°C.

The system worked by injecting a small amount of the sample from Method 2.6.1 onto a capillary. The three components of the sample i.e. substrate and two products, then travelled down the capillary by electrophoresis. This allowed the separation of the components according to their electrophoretic mobility (dependent upon charge:size ratio), in the order, GRSR, AcLRGAGRSR, AcLRGA.

The capillary itself, had an internal diameter of 25-50 microns, which is a compromise between heat dissipation and the need for a light path for detection. In the Biofocus system, as each analyte electrophoretically migrated past the UV detector (detection at 200nm), its UV absorption generated a signal proportional to its concentration. This signal produced an electropherogram which was collected by the BioFocus V3.10 Computer Package (Bio-Rad). These files were then stored as .BFF (Braided Format File).

2.6.3 Determination of Specific Activities

The electropherograms from Method 2.6.2 were analysed using the Bio-Rad software package Integrator V3.01. This package, converted the .BFF files to .ACQ (Acquisition) files. This then allowed manual integration of the peaks. The percentage area under each of the peaks, allowed the calculation of percentage cleavage over a time period. A graph was then drawn of time versus percentage cleavage of substrate, and the gradient of the tangent of the progress curve was taken as the initial rate. Using this, the Specific Activity was determined (Cabrita., 1997).

2.7 Protein Cleavage by Avp2

2.7.1 Smt3 and SUMO1 Cleavage

The cleavage assay was set up with 1mg/ml Smt3 or SUMO1, 1μM Avp2, 100μM pVIcT, 12.5mM Tris-HCl, 2.5mM EDTA, 0.5μM Pepstatin, 0.5mM AEBSF, 0.5mM β-Mercaptoethanol, pH8. There were also three controls set-up, +Avp2/-pVIcT; -Avp2/+pVIcT; and -Avp2/-pVIcT. The assay was left to incubate at 37°C for 24hr.

2.7.2 pVIII Cleavage

The protocol was based on the method described in the Promega TNT Quick Coupled Transcription/Translation System Handbook. 2µl of [^{35}S]methionine, 2µl of 0.5µg/µl circular plasmid, 6µl sterile water were added to 40µl of the Quick Master Mix (rabbit reticulocyte lysate, reaction buffer, RNA polymerase, amino acid mixture and ribonuclease inhibitor) (Promega). A control was also set up containing 10µl Quick Master Mix and 2.5µl sterile water. The reactions were then incubated for 90min at 30°C to produce the GFP-pVIII protein.

Following the TNT reaction, the cleavage assays were set up with 12.5µl TNT reaction mix, 0.2µM Avp2, 20µM pVIcT, made up to 40µl with Reaction Buffer (50mM Tris-HCl, 10mM EDTA, 2µM Pepstatin, 2mM AEBSF, pH8) ; *(b)*12.5µl TNT reaction, 0.2µM Avp2, made up to 40µl with Reaction buffer; *(c)*12.5µl TNT reaction, 20µM pVIcT, made up to 40µl Reaction buffer; *(d)*12.5µl TNT reaction, made up to 40µl with Reaction buffer. A control to measure back ground was also set up with *(e)*12.5µl Quick Master Mix/water incubated control, made up to 40µl with Reaction buffer. The samples were then incubated at 37°C and 20µl samples were removed at 30min and 1hr.

Subsequently, the samples were separated on a 15% SDS-PAGE (Method 2.4.1.1) which was then dried on Biorad Model 583 gel dryer. The gel was then exposed to BioMax MR film (Kodak) for 12-16hr, the film was developed with a Kodak M35 X-OMAT processor. To study band intensity, a box was drawn around the first lane and then copied and pasted to all other lanes. Bands within each lane were selected and the band intensity minus background was calculated. Each band was then divided by the number of methionines it contained and the percentage of each band was calculated.

CHAPTER 3
RESULTS AND DISCUSSION

This Chapter describes and discuses the results of work carried out on two proteins – Avp and pVIII of Adenovirus.

3.1 Characterisation of AVP

Introduction

The tertiary structure of Avp2 in complex with pVIcT has been published to both a 2.6 (Ding et al., 1996) and 1.6 Angstrom resolution (McGrath et al., 2003). There has also been work published discussing which N-terminal residues of pVIcT are involved in binding the protease (Cabrita et al., 1997; Ruzindana-Umunyana et al., 2000; Baniecki et al., 2001). However there is no mutational evidence to suggest which residues of Avp2 bind the substrate or the N-terminus of pVIcT.

Therefore, bioinformatic analysis of Avp2 was carried out at the primary and tertiary level to hypothesise which residues may be involved in both the binding of the substrate and the N-terminus of pVIcT. Following this analysis, an Alanine screen was prepared for these residues. In addition to this work, the potential function of the C-terminal tail of Avp2 was investigated using bioinformatics and mutational analysis.

Prior to cloning mutant Avp2 proteins, improvements were made to the expression and purification systems used to produce recombinant, native WT Avp2. It was subsequently hoped that a similar purification method could be used to purify the Egg Drop Syndrome protease (EDSp) as EDS and human Ad2 are members of different genera within the *Adenoviridae* family.

Results and Discussion

3.1.1 Improving Soluble Expression and Purification of Native Avp2

There was a requirement to improve the current expression and purification methods used for native Avp2 expressed in the pET11c vector (Novagen). The improvements were made to native rather than a tagged Avp2 as it was thought that it would not be beneficial to have a tag at the N- or C- terminus where mutational analysis was to be carried out later in this work.

Avp2 was originally cloned into the pET11c vector, expressed in the BL21(DE3) cell line, and induced at 37°C for 4hr, using 0.6mM Isopropyl-β-D-thiogalactopyranose (IPTG). In order to improve the current expression levels, a rare codon analysis was carried out to investigate the codon usage of Avp2. Following this, three variables which are known to affect expression levels of recombinant proteins were analysed. These factors were, host cell line, temperature of induction and IPTG concentration.

Likewise, improvement of purification was attempted by introducing a salt gradient at the end of the purification procedure.

3.1.1.1 Avp2 Rare Codon Analysis

As with many species, *Escherichia coli* does not use all 61 amino acid codons with the same frequency, and those which are used at a frequency of <1% are called rare codons.

This codon usage has an impact on the expression of recombinant proteins in *E. coli*, as was discussed in a recent review by Kane, 1995. The review suggested that there was a reduction in the quantity and quality of recombinant protein, if the protein contained any of the codons – AGG, AGA, CGA, CTA, ATA, or CCC. Similarly, Novagen has produced a cell line (Rosetta(DE3)) which expresses the tRNAs for AGG, AGA, CTA,

ATA, CCC and GGA, contained on a chloramphenicol resistant plasmid. This cell line has been shown to reduce premature termination of transcription and/or translation, thereby decreasing the amount of truncated protein. This consequently improves the quantity of full length recombinant protein. Prior to transforming the Avp2pET11c vector into the Rosetta(DE3) cell line and comparing its expression levels to that of the BL21(DE3) cell line, a rare codon analysis was carried out on Avp2. The total number of each rare codon is given in Table 3.1.1 and those highlighted in red were the rare codons coded for by the Rosetta(DE3) cell line. As can be seen from the Table 3.1.1, 15 of the 45 rare codons were coded for by the Rosetta(DE3) cell line.

Table 3.1.1: Rare Codon Analysis of Avp2
A rare codon analysis was carried according to the rare codons suggested by Kane, 1995. The total number of each rare codon is given below and those highlighted in red were the rare codons coded for by the Rosetta(DE3) cell line.

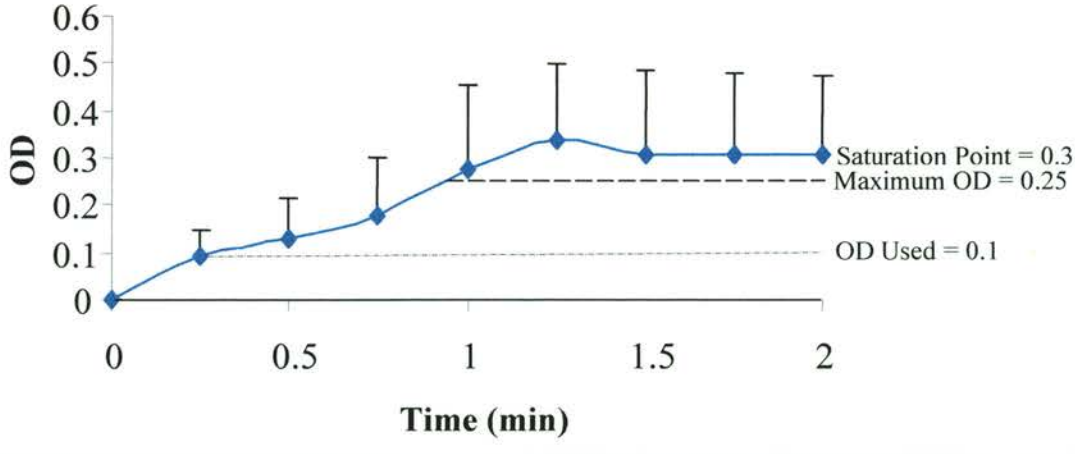
Rare Codon	**Total No. of Rare Codons (No. coded for by Rosetta(DE3))**
AGG	1
AGA	0
CGA	1
CTA	1
ATA	2
CCC	10
CGG	0
TGT	3
TGC	4
ACA	1
CCT	0
TCA	2
GGA	1
AGU	3
TCG	2
CCA	3
TCC	5
GGG	4
CTC	2
Total Number	**45 (15)**

3.1.1.2 Analysing the Saturation Point of a Western Blot

Three different variables were analysed to improve the expression levels of Avp2, these were different cell lines, temperatures of induction, and IPTG concentrations. To analyse the effects of each variable, a small scale growth was made, and a crude extract removed. However, the extract could not be analysed using SDS-PAGE mini-gels because the Avp2 was expressed at such low concentrations and there were a large number of contaminating bands. Therefore, the expression was examined by Western blot, and the OD of the band was calculated using the Gel-doc system.

Before analysing each variable, the saturation point of the Western blot was investigated (Method 2.4.4.2). As Figure 3.1.1 shows the saturation point of the X-ray film was at an OD of 0.3 and therefore a maximum OD of 0.25 was selected.

Figure 3.1.1: Saturation point of X-ray Film
Purified Avp2 was separated on a 15% SDS-PAGE mini-gel and then Western blotted using the primary and secondary antibodies, R11 and anti-Rabbit IgG, respectively. The mean OD (+/- SD) of the protein band minus the mean background OD was calculated in quadruplicate and then plotted against time. From this, the saturation point of the X-ray film and therefore the OD to be used for the superseding experiments were discerned.



3.1.1.3 Improving Expression of Native Avp2

Three variables were investigated to improve expression levels of native Avp2. These were different cell lines, induction temperatures and IPTG concentrations. To study each variable, four single colonies were grown to OD_{600} = 0.4-0.6, induced and then analysed by Western blot.

Firstly two pET vector host strains were investigated for native WT protease expression in the pET11c vector. These strains were BL21(DE3) (Promega) and Rosetta(DE3) (Novagen). The Rosetta(DE3) cell line were Tuner derivatives which were themselves mutants of the BL21(DE3) cell line. This mutation allowed constitutive entry of ITPG into all cells in the culture and therefore expression could be tightly regulated.

The Rosetta(DE3) cell line was selected to compare to the previously used BL21(DE3) cell line because it had been designed to enhance expression of proteins which contained codons rarely used by *E. coli*. The Novagen Rosetta(DE3) strain had a Chloramphenicol resistant plasmid which supplied tRNA's for AGG, AGA, AUA, CUA, CCC, GGA.

Four colonies of similar size were selected from a stock plate (Method 2.2.10.1) of Avp2pET11c expressed in BL21(DE3) and Rosetta(DE3) cell lines, which were used to inoculate 2ml of LB/Carbenicillin and 2ml of LB/Carbenicillin/Chloramphenicol, respectively. These were then grown for 12-16hr at 25°C. The 2ml cultures were then used to inoculate 18ml of LB/antibiotic and were grown at 37°C until OD_{600} = 0.4-0.6. Each culture was then induced with 0.6mM IPTG for 4hr at 37°C.

Post-induction, 1ml was removed, centrifuged at 13,000rpm for 10min in an MSE Microcentaur centrifuge and then re-suspended in 100µl Suspension Buffer (50mM Tris, 5mM EDTA, 4% Glycerol). Each sample was sonicated 6 x 10sec with 1min on ice between sonications. These samples were centrifuged at 13,000rpm for 10min, the

supernatant (soluble fraction) was retained and mixed with 50µl Reducing solution (10% (w/v) SDS, 25% (v/v) Gycerol, 0.05% (w/v) Bromophenol Blue, 25% (v/v) β-Mercaptoethanol, 0.25M Tris-HCl, pH6.9). 5µl of each sample was separated on a 15% SDS-PAGE mini-gel (Method 2.4.1.1). The four cultures from each cell line were separated in the same mini-gel to reduce variation between gels. Once Western blotted, the mean OD $^{+}$/- standard deviation was acquired, according to Method 2.4.4.2. As can be seen from Figure 3.1.2(a), expression of Avp2ET11c was far greater in the Rosetta(DE3) cell line than the BL21(DE3) cell line. Therefore, the next step was to analyse the effect of temperature on expression levels.

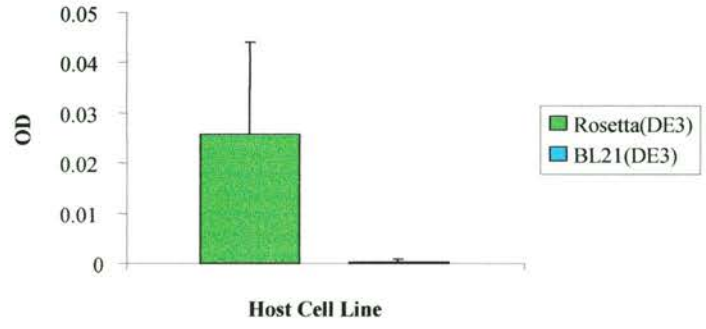
Reducing the temperature of induction is a common method used to improve soluble expression, therefore the temperatures of induction chosen were 25°C and 37°C. The same protocol was used as described above for growth and induction, except four similar sized colonies of Rosetta(DE3) cells were induced with either 0.6mM IPTG for 6hr at 25°C or 0.6mM IPTG for 4hr at 37°C. Reducing the temperature did not appear to improve soluble expression of the protease (Figure 3.1.2(b)) and so induction at 37°C was retained.

Likewise the results from the induction of Rosetta(DE3) cells with 0mM, 0.2mM, 0.4mM, 0.6mM, 0.8mM and 1mM IPTG at 37°C for 4hr also did not appear to improve expression (Figure 3.1.2(c)), because even in the absence of IPTG some basal expression was noted.

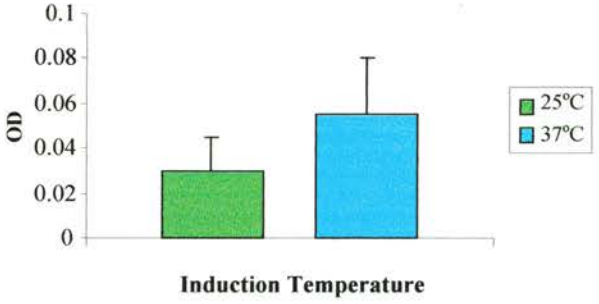
Therefore, for future experiments, both Avp2 WT and Avp2 mutants cloned into the pET11c vector were expressed in the Rosetta(DE3) cell line and induced with 0.6mM IPTG for 4hr at 37°C.

Figure 3.1.2: Improving Soluble Expression of Native Avp2
Results of the Soluble Expression tests for WT Avp2 grown in (a) Rosetta(DE3) and BL21(DE3) host cells, induced with 0.6mM IPTG at 37°C, (b) Rosetta(DE3) cells induced with 0.6mM IPTG at 25°C and 37°C for 6hr and 4hr respectively, (c) Rosetta(DE3) cells induced with 0mM, 0.2mM, 0.4mM, 0.6mM, 0.8mM and 1mM IPTG at 37°C for 4hr.

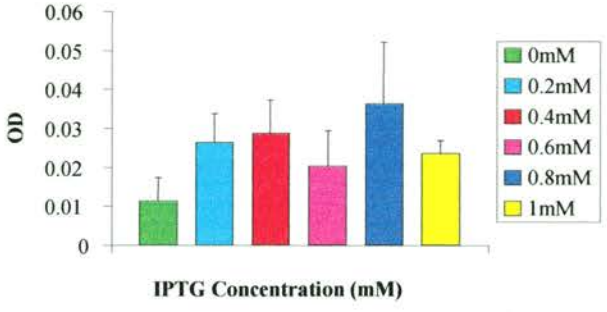
(a)



(b)



(c)



3.1.1.4 Improving Purification Method for Native Avp2

Avp2ET11c expressed in the Rosetta(DE3) cell line was grown in LB/Carbenecillin/Chloramphenicol according to Method 2.5.2. The cells were induced with 0.6mM IPTG for 4hr at 37°C. Following this, they were re-suspended in 8ml of Suspension Buffer (50mM Tris, 5mM EDTA, 4% Glycerol) and then stored at –20°C.

Previously, Avp2ET11c was purified by a three step process (Pollard, 2001) – firstly it was loaded onto the Diethylaminoethyl (DEAE) -Sepharose column and then onto the Heparin Sepharose (HS) column and finally the Carboxymethyl (CM) -Sepharose column. It was then eluted from the CM-Sepharose column in the presence of 200mM NaCl (Figure 3.1.3(a)). This produced a relatively pure sample of Avp2, which was the strongest band on a 15% SDS-PAGE mini-gel (Method 2.4.1.1), when stained with Coomassie Blue (Method 2.4.2.1) (Figure 3.1.3(b)). However, when the same gel was Silver stained (Method 2.4.2.2) it became obvious that there were a number of other contaminating bands (Figure 3.1.3(c)). Western blotting (Method 2.4.3) (Figure 3.1.3(d)) and Mass Spectrometry (Method 2.4.7.2) were used to confirm the Avp2 band.

In order to improve this purification, the same method was followed except elution from the CM-Sepharose column was carried out using a gradient of 0-200mM NaCl over a 30min period (Method 2.5.4). The eluted fractions (Figure 3.1.4(a)) were separated on a 15% SDS-PAGE mini-gel, which was stained with Coomassie Blue (Figure 3.1.4(b)) and then Silver stain (Figure 3.1.4(c)). The gel was also Western blotted (Figure 3.1.4(d)), which together with Mass Spectrometry confirmed the Avp2 band.

When Figure 3.1.3(c) and 3.1.4(c) were compared it was clear that contamination was greatly reduced when a salt gradient was used to elute the protein from the CM-Sepharose column. Therefore elution over a larger number of samples allowed those samples which were of highest purity to be pooled and concentrated (Method 2.5.8).

Figure 3.1.3: Purification of Native WT Avp2
Purification of native WT Avp2 was carried out using Ion Exchange Chromatography. The bacterial cell extract expressing Avp2 was passed through three columns at 1ml/min in the order, DEAE-Sepharose, HS and CM-Sepharose. 1ml fractions were collected and (a) the elution of proteins from the column was recorded at 280nm. Once passed onto the CM-Sepharose column the Avp2 was eluted with 200mM NaCl between fractions 50-55. The fractions were separated on a 15% SDS-page gel and stained with (b) Coomassie Blue then (c) Silver stain. (d) Western blotting, using the R11 primary antibody and Mass Spectrometry confirmed the Avp2 band.

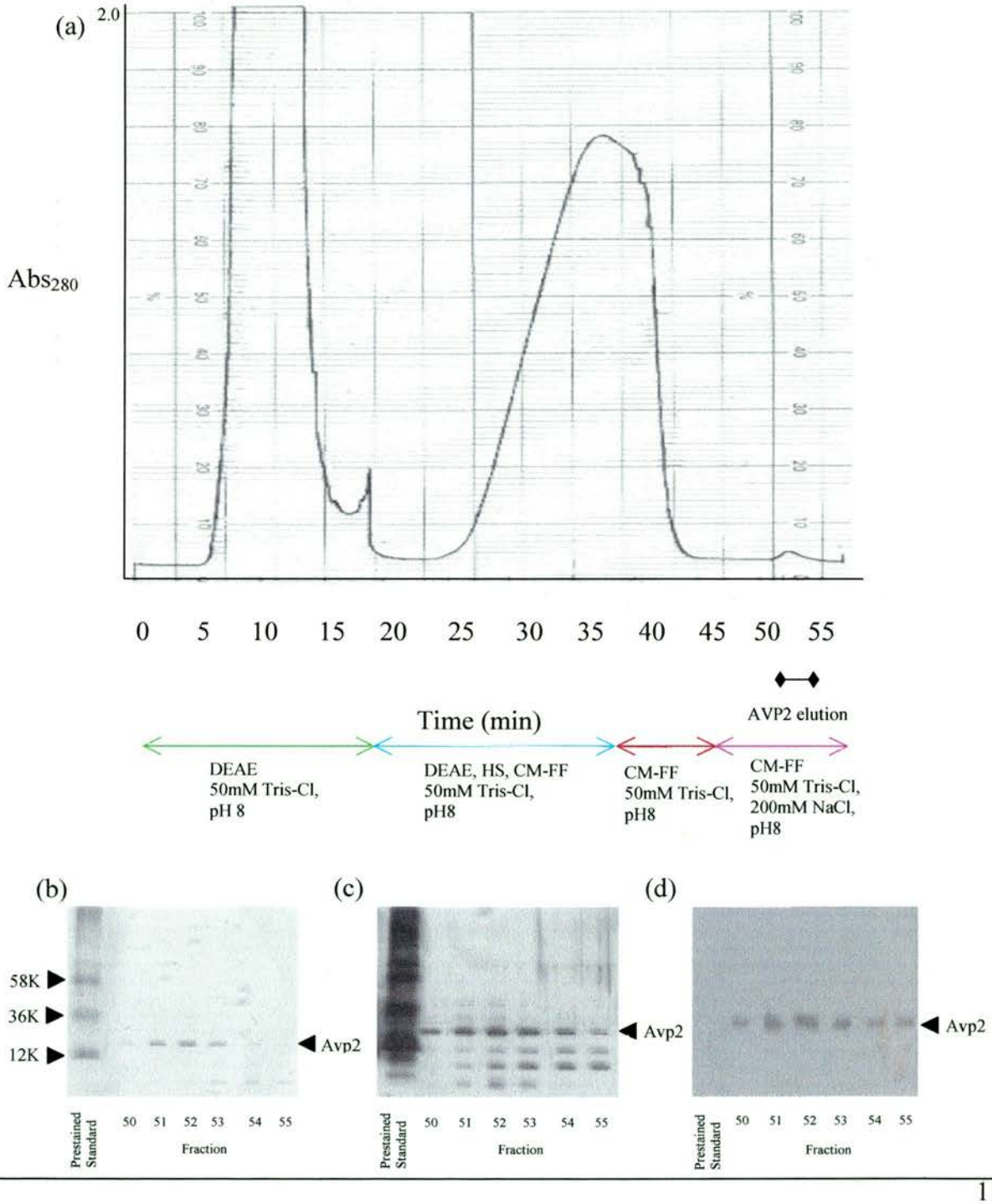
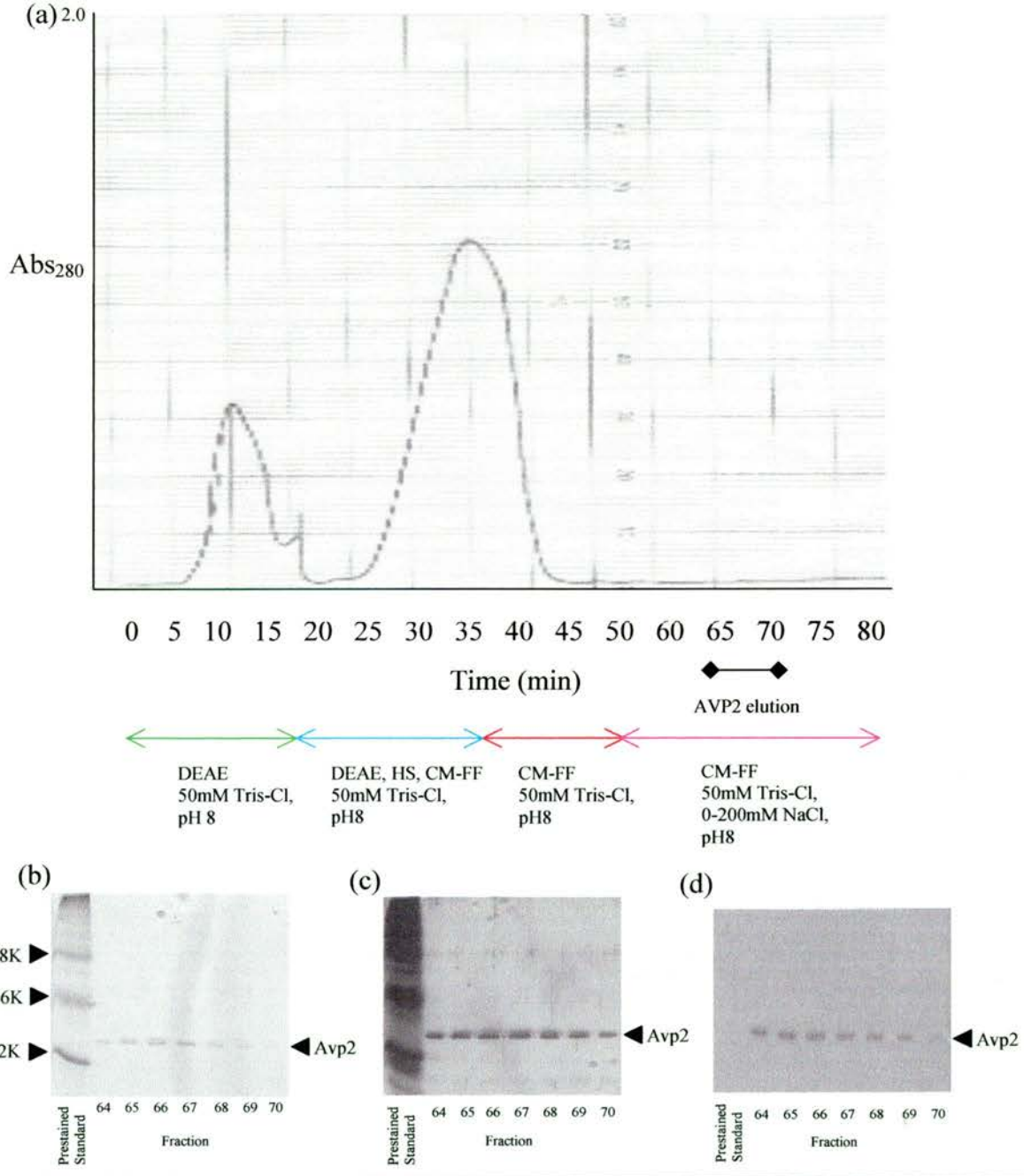


Figure 3.1.4: Improving the Purification of Native WT Avp2
An example purification of WT Avp2 is given below, a similar profile was found for all native mutants. Purification of native WT Avp2 was carried out using Ion Exchange Chromatography. The bacterial cell extract expressing Avp2 was passed through three columns at 1ml/min in the order, DEAE-Sepharose, HS and CM-Sephaorse. 1ml fractions were collected and (a) the elution of proteins from the column was recorded at 280nm. Once passed onto the CM-Sepharose column the Avp2 was eluted using a 0mM-200mM $NaCl_2$ gradient over a 30min period. The protease was eluted between fractions 64-70. The fractions were separated on a 15% SDS-page gel and stained with (b) Coomassie Blue then (c) Silver stain. (d) Western blotting, using the R11 primary antibody and Mass Spectrometry confirmed the Avp2 band. This method was followed for all of the mutants.



3.1.1.5 Discussion

To improve soluble expression of native Avp2ET11c, a number of variables which have previously been shown to improve soluble expression were examined. These included different cell lines, temperatures of induction and IPTG concentrations. The expression levels for each variable were measured in quadruplicate using a Western blot method, as protein expression was low. This method produced large standard deviations from the mean which may have been due to intrinsic differences between the selected colonies.

As the results of the Western blot method proved, the only variable which caused an increase in expression levels was the use of the Rosetta(DE3) cell line. However, this improvement was not dramatic and it may be have been beneficial to replace the rare codons in Avp2 with codons which are less rarely used by *E. coli*. This work would have been lengthy, therefore, instead I began to investigate other tagged vector systems to improve soluble expression levels (Sub-chapter 3.2).

Although expression levels were low, the expressed protein was found in high enough concentrations to be purified and used in an assay designed to measure specific activity (Webster et al., 1989a). Therefore the purification method was improved to reduce contaminants, including those capable of cleaving the synthetic AcLRGAGRSR substrate. Improvement of the purification method was successful as it reduced contaminantion dramatically.

The methods developed in this section for WT Avp2 were subsequently used to express and purify all of the proteases discussed in Sections 3.1.2 to 3.1.5.

3.1.2 Substrate Specificity Pockets

Although the crystal structure of the Avp2-pVIcT complex has been published to 2.6 Angstroms (Ding et al, 1996) and 1.6 Angstroms (McGrath et al, 2003), no Avp2-substrate complex has been published to-date. Therefore, it is not known which residues of Avp2 are involved in binding the substrate.

It is however accepted that the substrate motifs cleaved by Avp2 are, (M,L,I,Q)XGX-G and (M,L,I,)XGG-X (Webster et al, 1989b; Anderson et al, 1990; Webster et al., 1997). Thus, as the major established sites of recognition for the substrate are at the P2 and P4 sites (Webster et al, 1989b), the bioinformatic and mutational analysis used in this work analysed the S2 and S4 substrate specificity pockets of Avp2.

3.1.2.1 Bioinformatic Analysis

Bioinformatics were utilised to investigate which residues of Avp2 may be involved in substrate binding. The bioinformatic procedures used were - a tertiary structure superimposition of Avp2 and Ulp1, with its Smt3 substrate, a primary structure alignment of Avp proteins and a primary structure alignment of the families which form the CE clan of Cysteine proteases.

3.1.2.1.1 Avp2/Ulp1 Superimposition

Avp2 and Ulp1 belong to the C5 and C48 Cysteine protease family, respectively. These families are both members of the CE clan of Cysteine proteases and have similar substrate specificity in that both cleave at a GG-X bond. In addition it has also been proposed that they share structural similarities (Li & Hoschstrasser., 1999).

In this thesis, a preliminary investigation of the S2 and S4 substrate specificity pockets was carried out by superimposing the 2.6 Angstrom structure of Avp2 (this was the

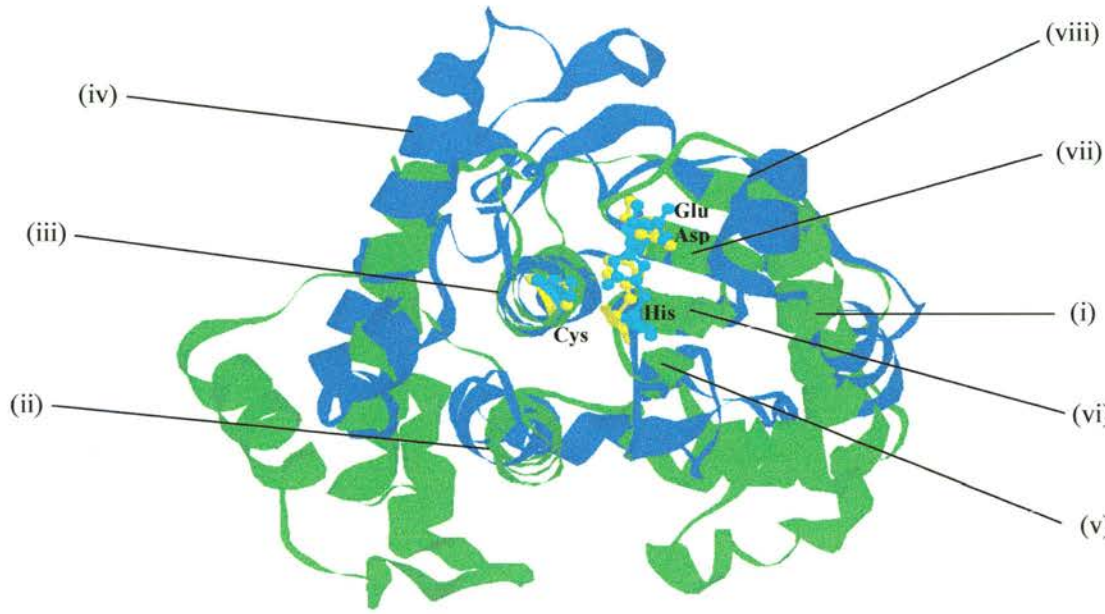
only structure available whilst this research was being carried out) and the 1.8 Angstrom structure of the Ulp1-Smt3 complex (Mossessova & Lima, 2000) using the Swiss-PDB Viewer Programme (Guex & Peitsch., 1997; http://www.expasy.org/spdbv/). The protocol used to superimpose the two proteins is described in Method 2.1.2.1, and the residue selected for superimposition was the catalytic cysteine (Cys122 for Avp2 and Cys580 for Ulp1).

The superimposition was viewed in the RASMOL Version 2.6 computer package (Sayle & Milner-White., 1995). Using this package, the Ulp1 substrate, Smt3 was removed to leave only Avp2 and Ulp1 (Figure 3.1.5). Analysis of the superimposition showed that as was expected, Avp2 and Ulp1 shared similar structural characteristics. The superimposition is shown in Figure 3.1.5(a) with Avp2 and Ulp1 highlighted in blue and green, respectively. There were a number of secondary structural characteristics which were maintained in both proteases. These characteristics were four α-helices and four β-strands. In particular, α-helix(a)(iii), β-strand(a)(vi) and β-strand (a)(vii) gave credence to using this as a model because the catalytic triad, cysteine, histidine, glutamate/aspartate lie on each of these secondary structures, respectively. Moreover as Figure 3.1.5(a) and 3.1.5(b) shows the catalytic triads were superimposed, as were a number of other residues.

Following this, Ulp1 was removed from the superimposition to leave only Avp2 and Smt3, and as is shown in Figure 3.1.6 Smt3 fits into the putative substrate specificity cleft of Avp2.

Figure 3.1.5: Superimposition of Avp2 and Ulp1
Avp2 and Ulp1 were superimposed using the Swiss-PDB Viewer Programme (Guex & Peitsch, 1997; http://www.expasy.org/spdbv/) and the (a) superimposition was visualised in the RASMOL Version 2.6 computer package (Sayle & Milner-White., 1995). (a) Avp2 and Ulp1 coloured blue and green, respectively, have (i)-(iv) four α-helices and (v)-(viii) four β-sheets, which superimpose. The catalytic triad of Avp2 and Ulp1are coloured cyan and yellow, respectively in the (a) superimposition. The (b) table shows a number of amino acid were superimposed, including the catalytic triad of Avp2 and Ulp1, which are highlighted in blue and green respectively.

(a) Superimposition

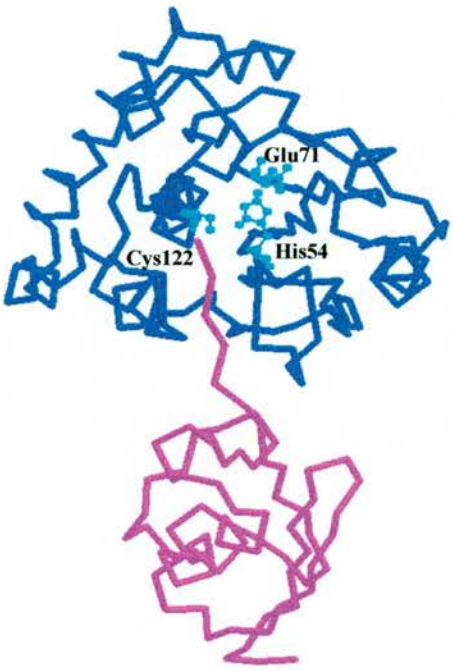


(b) Table of Superimposed Residues

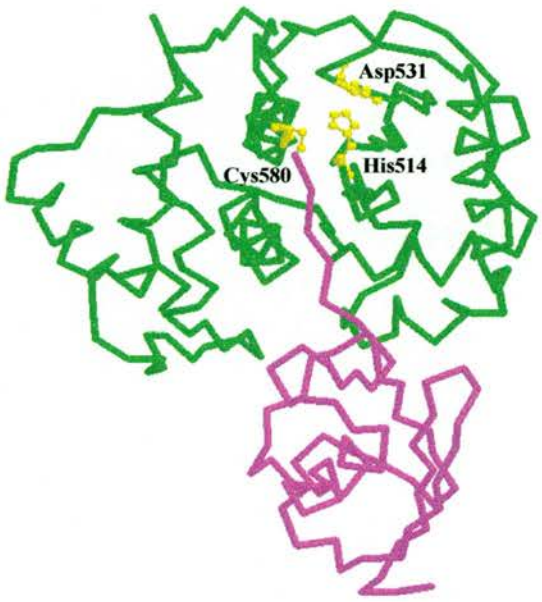
Avp2	Ulp1
3 SSEQE 7	448 WLNDT 452
8 LKA 10	454 IEF 456
17 CG 18	461 IE 462
20 YFLGT 24	469 VAFNS 473
38 LACAIVN 44	503 KIFTPIN 509
52 GVHWMAFAWNPRSKTCYLFEPFG 74	512 QSHWALGIIDLKKKTIGYVDSLS 534
76 S 76	536 G 536
103 RCITLEK 109	562 EDFDLIH 568
114 VQ 115	573 QQ 574
119 SAACGLFCCMFLHAFANW 136	577 GYDCGIYVCMNTLYGSAD 594
171 QEQLYS 176	601 YKDAIR 606
183 P 183	613 H 613

Figure 3.1.6: Smt3 as a Substrate for Avp2 and Ulp1
Avp2 and the Ulp1-Smt3 complex were superimposed using the Swiss-PDB Viewer Programme (Guex & Peitsch, 1997; http://www.expasy.org/spdbv/). The (a) Avp2-Smt3 model and (b) Ulp1-Smt3 structure were then visualised in the RASMOL Version 2.6 computer package (Sayle & Milner-White, 1995). Smt3 sits in the putative catalytic cleft for Avp2. This cleft has a similar secondary structure to Ulp1 and the catalytic triad sits in exactly the same orientation (Figure 3.1.5). The catalytic cysteine sits at the end of the central α-helix, whereas, catalytic glutamate/aspartate and histidine lie on two β-strands.

(a) Avp2-Smt3 Model



(b) Ulp1-Smt3 Structure



3.1.2.1.2 Avp2-Smt3 Molecular Model

As Ulp1 superimposed well onto Avp2 and also because Smt3 fitted into the putative substrate specificity cleft of Avp2 (Section 3.1.2.1.1), it was decided to use the Avp2-Smt3 complex as a model. The purpose of the model was to identify potential residues of Avp2 involved in the S2 and S4 substrate specificity pockets. To detect these residues, the 'Radius Function' of the Swiss-PDB Viewer Programme (Guex & Peitsch, 1997; http://www.expasy.org/spdbv/) was utilised (Method 2.1.2.2). This allowed the α-Carbon of either the P2 residue (Gly97) or the P4 residue (Gln95) of Smt3 to be chosen. The reason for selecting the α-Carbon was that this was independent of the side chain. Once the α-Carbon was selected, residues of Avp2 within a radius of 1 Angstrom to 10 Angstroms of it could be found. The results for the P2 site and P4 site are shown in Tables 3.1.2 and 3.1.3, respectively. Both tables begin at a 4 Angstrom radius as no Avp residues were found at 1, 2 or 3 Angstroms.

A 6 Angstrom radius 'cut off' was selected as this produced a list of 14 residues which were putatively involved in the substrate specificity pockets. The residues to be investigated further were highlighted in blue in Table 3.1.2 and 3.1.3. Whilst those residues not highlighted in column 6 of Table 3.1.2 are members of the catalytic triad and have already been examined by mutational analysis (Grierson et al., 1994; Jones et al., 1996).

Table 3.1.2: Avp2 Residues Which Lie in a 1-10 Angstrom Radius of the Smt3 P2 Site
Avp2 and the Ulp1-Smt3 complex were superimposed using the Swiss-PDB Viewer Programme (Guex & Peitsch, 1997; http://www.expasy.org/spdbv/). The 'Radius Function' of the programme allowed the α-Carbon of the P2 residue (Gly97) of Smt3 to be chosen. Once the α-Carbon was selected, residues of Avp2 within a radius of 1 Angstrom to 10 Angtroms of it were found. A 6 Angstrom radius 'cut off' was selected as this produced a list of 7 residues which were putatively involved in the S2 substrate specificity pocket. The residues to be investigated further were highlighted in blue in column 6, whilst those residues not highlighted are members of the catalytic triad.

Residues within a radius of:						
4 Angstroms	**5 Angstroms**	**6 Angstroms**	**7 Angstroms**	**8 Angstroms**	**9 Angstroms**	**10 Angstroms**
Ser3	Ser3	Ser3	Met1	Met1	Met1	Met1
Val53	Ser4	Ser4	Ser3	Gly2	Gly2	Gly2
Trp55	Asn44	Asn44	Ser4	Ser3	Ser3	Ser3
	Gly52	Gly52	Glu5	Ser4	Ser4	Ser4
	Val53	Val53	Leu8	Glu5	Glu5	Glu5
	His54	His54	Asn44	Gln6	Gln6	Gln6
	Trp55	Trp55	Gly51	Leu8	Leu8	Glu7
	Cys122	Cys122	Gly52	Asn44	Ile42	Leu8
		Gly123	His54	Thr45	Val43	Tyr25
			Trp55	Gly51	Asn44	Ile42
			Ala120	Gly52	Thr45	Val43
			Cys122	Val53	Ala46	Asn44
			Gly123	His54	Gly51	Thr45
				Trp55	Gly52	Ala46
				Met56	Val53	Thr50
				Ala120	His54	Gly51
				Ala121	Trp55	Gly52
				Cys122	Met56	Val53
				Gly123	Gln115	His54
					Ser119	Trp55
					Ala120	Met56
					Ala121	Pro72
					Cys122	Tyr84
					Gly123	Gln115
					Leu124	Ser119
					Ala195	Ala120
					Thr196	Ala121
					Leu201	Cys122
						Gly123
						Leu124
						Cys126
						Ala195
						Thr196
						His 200
						Leu201

Table 3.1.3: Avp2 Residues Which Lie in a 1-10 Angstrom Radius of the Smt3 P4 Site
Avp2 and the Ulp1-Smt3 complex were superimposed using the Swiss-PDB Viewer Programme (Guex & Peitsch, 1997; http://www.expasy.org/spdbv/). The 'Radius Function' of the programme allowed the α-Carbon of the P4 residue (Gln95) of Smt3 to be chosen. Once the α-Carbon was selected, residues of Avp2 within a radius of 1 Angstrom to 10 Angtroms of it were found. A 6 Angstrom radius 'cut off' was selected as this produced a list of 8 residues which were putatively involved in the S4 substrate specificity pocket. The residues to be investigated further were highlighted in blue in column 6.

		Residues within a radius of:				
4 Angstroms	**5 Angstroms**	**6 Angstroms**	**7 Angstroms**	**8 Angstroms**	**9 Angstroms**	**10 Angstroms**
Gln6	Gln6	Met1	Met1	Met1	Met1	Met1
Gly51	Gly51	Glu5	Ser4	Ser4	Ser3	Gly2
	Trp55	Gln6	Glu5	Glu5	Ser4	Ser3
		Asn44	Gln6	Gln6	Glu5	Ser4
		Arg48	Asn44	Glu7	Gln6	Glu5
		Gly51	Arg48	Leu8	Glu7	Gln6
		Gly52	Gly51	Thr24	Leu8	Glu7
		Trp55	Gly52	Tyr25	Thr24	Leu8
			Trp55	Asp26	Tyr25	Lys9
				Asn44	Asp26	Thr24
				Gly47	Asn44	Tyr25
				Arg48	Ala46	Asp26
				Thr50	Gly47	Lys27
				Gly51	Arg48	Ile42
				Val53	Glu49	Val43
				Trp55	Thr50	Asn44
					Gly51	Thr45
					Gly52	Ala46
					Val53	Gly47
					Trp55	Arg48
						Glu49
						Thr50
						Gly51
						Gly52
						Val53
						His54
						Trp55

3.1.2.1.3 Primary Sequence Alignments

Two primary sequence alignments involving Avp2 were carried out. These were an alignment of Avp proteins and also an alignment of representative members of the families which form the CE clan of Cysteine proteases. The alignments were carried out to analyse if any of the residues which lay within a 6 Angstrom radius of the P2 or P4 α-carbon of Smt3 (Section 3.1.2.1.2) were conserved.

As is shown in Figure 3.1.7, 28 Avp sequences representing the four genera of the *Adenoviridae* family were aligned (Method 2.1.1). The residues of Avp2 which were within a 6 Angstrom radius of the P2 or P4 α-carbon atom of Smt3 in the Avp2-Smt3 model (Section 3.1.2.1.2) were highlighted in blue and their position was indicated by a 2 or a 4, respectively. The catalytic residues were indicated by a blue triangle.

As Figure 3.1.7 shows, of the 14 residues putatively involved in the S2 and S4 substrate specificity pockets, 7 were conserved in the Avp alignment (Asn44, Gly51, Gly52, His54, Trp55, Cys122 and Gly123). This alignment gave further credence to the importance of these residues in the S2 and S4 substrate specificity pockets.

Following this, an alignment of representative members from each of the five families of the CE clan (Family C5 was represented by Avp2 (human Ad2 protease); family C48 was represented by Ulp1 (*Saccharomyces cerevisiae* protease); family C63 was represented by ASFV (African swine fever virus, strain BA71V processing peptidase); family C57 was represented by Vaccinia virus I7L (Vaccinia virus, strain Copenhagen I7L processing peptidase); family C55 was represented by YopJ (*Yersinia pseudotuberculosis* protease)) was carried out. Each sequence in the alignment was found using NCBI (http://www.ncbi.nlm.nih.gov/). The sequences were then pasted into the CLUSTALW

alignment programme at EBI, Hinxton (http://www.ebi.ac.uk/clustalw/) and an alignment was performed (Figure 3.1.8(a)).

As previously discussed, the aim of the CE clan alignment was to study whether any of the residues which were within a 6 Angstrom radius of either the P2 or P4 α-carbon of Smt3 in the Avp2-Smt3 model were maintained. Figure 3.1.8(a) highlights the known members of the catalytic triads of Avp2 (His54, Glu71, Cys122) (Grierson et al, 1994; Ding et al., 1996; Jones et al., 1996; Rancourt et al., 1996), Ulp1 (His514, Cys580) (Li and Hoschstrasser, 1999), ASFV (His168, Cys232) (Andres et al., 2001), Vaccinia virus (His241, Asp248, Cys328) (Byrd et al., 2003), YopJ (His109, Asp248, Cys172) (Orth et al., 2000) in blue and putative members in red. Nevertheless, preliminary analysis suggested that the members of the CE clan shared little sequence similarity and none of the residues potentially involved in the S2 and S4 substrate specificity pockets aligned (Figure 3.1.8(a)).

However as noted by Li and Hoschstrasser, 1999, members of the CE clan – Avp2, Ulp1, ASFV and Vaccinia virus I7L proteases contained a 'core domain' which could be aligned. Following this, Orth et al., 2000, aligned Avp2 and Ulp1 with the bacterial protease, YopJ using the same principal. Therefore when the representative members from each of the five families of the CE clan were aligned using the core domain (Figure 3.1.8(b)), five conserved residues were found – the catalytic histidine with adjacent tryptophan/phenyalanine (conserved substitution), the catalytic glutamate/aspartate/aspargine (conserved substitution) and the catalytic cysteine with an adjacent glycine.

This provided further evidence of the importance of His54, Trp55, Glu71, Cys122 and Cys123 in Avp2 substrate binding.

Figure 3.1.7: Primary Sequence Alignment of *Adenoviridae* Proteases
28 Avp sequences representing the four genera of the *Adenoviridae* family were aligned. The residues of Avp2 which were within a 6 Angstrom radius of either the P2 or P4 residues of Smt3 in the Avp2-Smt3 model were highlighted in blue and their position was indicated by a 2 or a 4, respectively. Catalytic residues were indicated by a blue triangle.
* Sequence for Canine 2 (strain Toronto A26-61).
** Sequence for Canine 1 (strains Utrecht, CLL and RI261).

```
                  22                                                  2       2222
                 4       44                                          4   4   4 ▼
Human 2       --MGSS----EQELKAIVKDLGCGPYFLGTYDKRFPGFVSPHKLACAIVNTAGRETGGVH 54
Human 5       --MGSS----EQELKAIVKDLGCGPYFLGTYDKRFPGFVSPHKLACAIVNTAGRETGGVH 54
Human 40      --MGSS----EQELVAIVRELGCGPYFLGTFDKRFPGFMAPHKLACAIVNTAGRETGGVH 54
Human 41      --MGSS----EQELVAIARDLGCGSYFLGTFDKRFPGFMAPNKLACAIVNTAGRETGGVH 54
Human 3       MTCGSGNGSSEQELKAIVRDLGCGPYFLGTFDKRFPGFMAPDKLACAIVNTAGRETGGEH 60
Human 11      MACGSQNGSSEQELRAIVQDLGCGPYFLGTYDKRFPGFMAPDKLACAIVNTAGRETGGEH 60
Human 12      --MGSS----EQELTAIVRDLGCGPYFLGTFDKRFPGFVSRDRLSCAIVNTAGRETGGVH 54
Human 17      -MSGSS----ERELASIVRDLGCGPTFWAPTTQRFPGFLAGDKLACAIVNTAGRETGGVH 55
Canine 2*     MAEGGS---SEEELRAIVHNLGVSPFFLGTFDKRFPGFISSQRMACAIVNTAGRETGGVH 57
Canine 1**    MAEGGS---SEEELRAIVRDLAVTPFFLGTFDKRFPGFISSQRITCAVVNTAGRETGGVH 57
Bovine 3      --MGSR----EEELRFILHDLGVGPYFLGTFDKHFPGFISKDRMSCAIVNTAGRETGGVH 54
Human 4       MAAGSG----EQELRAIIRDLGCGPYFLGTFDKRFPGFMAPHKVACAIVNTAGRETGGEH 56
Bovine 2      --MGSR----EEELRAIVRDLGISPYFLGTFDKRFPGFLHKDKLSCAIVNTAARETGGAH 54
Porcine 5     --MGST----EELRAIVRDLGVGAYFLGTFDNRFPGFLNKDKMSCAIVNTDARETGGRH 54
Bovine 10     --MGTS----EELKHIVIDLGCGPFFLGINDKHFPGFLNKQSNACAIVNTASRETGGVH 54
Mouse 1       --MGSS----ETELRQLVADLGI-GSFLGIFDKHFPGFISVNKPACAIVNTASRETGGVH 53
Porcine 3     --MGST----EDELRAMARDLQL-PRFLGTFDKSFPGFLQESQRCCAIVNTAARHTGGRH 53
Equine 2      --MGST----ETELKRILSDLNVTPLLLGTYDKRFPGFVSKAK-PCPIVNTAFGETGGEH 53
Fowl 1        -MSGTT----ETQLRDLLSSMHLRHRFLGVFDKSFPGFLDPHVPASAIVNTGSRASGGMH 55
Turkey 3      -MAGTS----SSELITLVRSLGLGSYFLGVYDKHFPGFLNDRRLAYAIVNTGDYMSGGLH 55
Fowl 8        -MSGTT----ESQLNQLVGAMHLRHRFLGVFDKTFPGFLDPNRPASAIVNTGSRATGGMH 55
Snake 1       -MSGSS----EQELKMMVRSLNLESGFLGTFDCRFPGFISKDRRQTAIVNTGPREQGGVH 55
Frog 1        --MGTS----GADLENIVLSLGLHSGFLGIFDKHFPGFLNVNKPSFAIVNTGDIIQGGLH 55
Possum 1      -MSGTS----EHELKSLIASLGVSSGFLGTFDCRFPGFIQKEKFQTAIVNTGPREKGGIH 55
Bovine 4      -MSGTS----ESELKHLLSSLHLTYGFLGTFDCRFPGFLQKNKVQTAIVNTGPREKGGVH 55
Bovine 7      -MSGLS----EKEVFLLLSSLQCTHGFLGTFDCRFPGFINKVKVQTAIINTGPREQGGIH 55
Duck 1        -MSGTS----ESELKALMKSLGIAGNFLGTFDCTFPGFINKHKRQTAIINTGSRASGGLH 55
Ovine 7       -MSGTS----ESELKNLISSLHLNNGFLGIFDCRFPGFLQKSKIQTAIINTGPREQGGIH 55
                 *        ::  :   :     : .     ****:       .::**     ** *

              2
              4
              ▼
Human 2       WMAFAWNPRSKTCYLFEPFGFSDQRLKQVYQFEYESLLRRSAIAS-SPDRCITLEKSTQS 114
Human 5       WMAFAWNPHSKTCYLFEPFGFSDQRLKQVYQFEYESLLRRSAIAS-SPDRCITLEKSTQS 114
Human 40      WLALAWNPKNRTCYLFDPFGFSDERLKQIYQFEYEGLLKRSALAS-TPDHCITLIKSTQT 114
Human 41      WLALAWNPKSHTCYLFDPFGFSDERLKQIYQFEYEGLLKRSALAS-TPDHCITLVKSTQT 114
Human 3       WLAFGWNPRYNTCYLFDPFGFSDERLKQIYQFEYEGLLRRSALA--TKDRCITLEKSTQS 119
Human 11      WLAFGWNPRSNTCYLFDPFGFSDDRLKQIYQFEYEGLLRRSALA--TKDRCITLEKSTQT 119
Human 12      WLAFGWNPKSHTCYLFDPFGFSDQRLKQIYQFEYESLLRRSALAA-TKDRCVTLEKSTQT 114
Human 17      WLAFGWNPRSRTCYMFDPFGFSDRRLKQIYSFEYEAMLRRSAVAS-SPDRCLSLEQSTQT 115
Canine 2*     WLAMAWNPRSKTFYMFDPFGFSDSKLKQVYSFEYEGLLRRSAIAS-SPDRCVTLAKSNET 117
Canine 1**    WLAMAWNPRSKTFYMFDPFGFSDSKLKQVYSFEYEGLLRRSAIAS-TPDRCVTLAKSNET 117
Bovine 3      WLAMAWHPASQTFYMFDPFGFSDQKLKQIYNFEYQGLLKRSALTS-TADRCLTLIQSTQS 114
Human 4       WLAFAWNPRSNTCYLFDPFGFSDQRLKQIYQFEYEGLLRRSALA--TKDRCVT-WKSHQT 114
Bovine 2      WLALAWFPNAKNFYFFDPFGFSDHKLKQIYQFEYEGLLRRSALA---GDGCVNLVKSTET 112
Porcine 5     WLALAWYPPKRRVYFFDPFGFSDAKLKQIYHFEYEGLLKRSALN---GDRCVELEKSTDT 112
Bovine 10     WIAMGWHPPSN-FYLFDPFGFSDKKLLQIYQFEYNALLKRSAITS-SPDRCVQLFQNNES 113
Mouse 1       WLAMAWYPTSSTFYLFDPFGFSDRKLQQVYKFEYERLLKRSAVSS-SSSKCVTLVKSHQT 113
Porcine 3     WLAVAWEPASRTFYFFDPFGFSDRELAQVYDFEYQRLLRKSAIQS-TPDRCLTLVKSTQS 113
Equine 2      WIAMAWYPPNNSFYMFDPFGFSDQKLKQIYDFEYQGLLRRSALTS-SKDRCVQLIRSTDT 113
Fowl 1        WIGFAFDPAAGRCYMFDPFGWSDQKLWELYRVKYNAFMRRTGLRQP--DRCFTLVRSTEA 114
Turkey 3      WIAFAYDPNGRKFYIFDPFGWSKKELWKFYKFQYDRIVRRTALQN---GRCIKLVRSVDT 113
Fowl 8        WIAFAFDPIARKCYMFDPFGWSDRELWNLYKVKYDAFLRRTGLRQP--DKCFELVRSVEA 114
Snake 1       WIALAWDPTSRKMFLFDPLGWTNAELKKYYGFSYQNMVQRSALSS--ADRCVTLERNTQA 114
Frog 1        WIAFAFDNVTSTFFMFDPFGWSDMELYRKYEFQYHRILKSTALTKP--SRCIKLVKSKEA 113
Possum 1      WIALAWNPLNFKMYLFDPLGWKEKQLAKYYGFSYRNMIARSALNT--PDRCVTLVRNEEA 114
Bovine 4      WVAMAWDPIYYKMYIFDPLGWKESQLQSLYNYSYQSMLKRSALTE--SERCITVEKNTQS 114
Bovine 7      WIALAWDPKSYQMFIFDPLGWKNDQLMKYYKFSYSNLIKRSALSS--PDKCVKVIKNSQS 114
Duck 1        WLAFAWDPLRYTIYMFDPLGWKEKDLFKLYGFSYKTMIKRSALQS--DNRCVKLVKNTEA 114
Ovine 7       WITLALEPISYKLFIFDPLGWKDTQLIKFYNFSLNSLIKRSALNN--SDRCITVERNTQS 114
              *: ..        ::*:*:*:..  *   *  .   ::  :.:       *.   :. ::
```

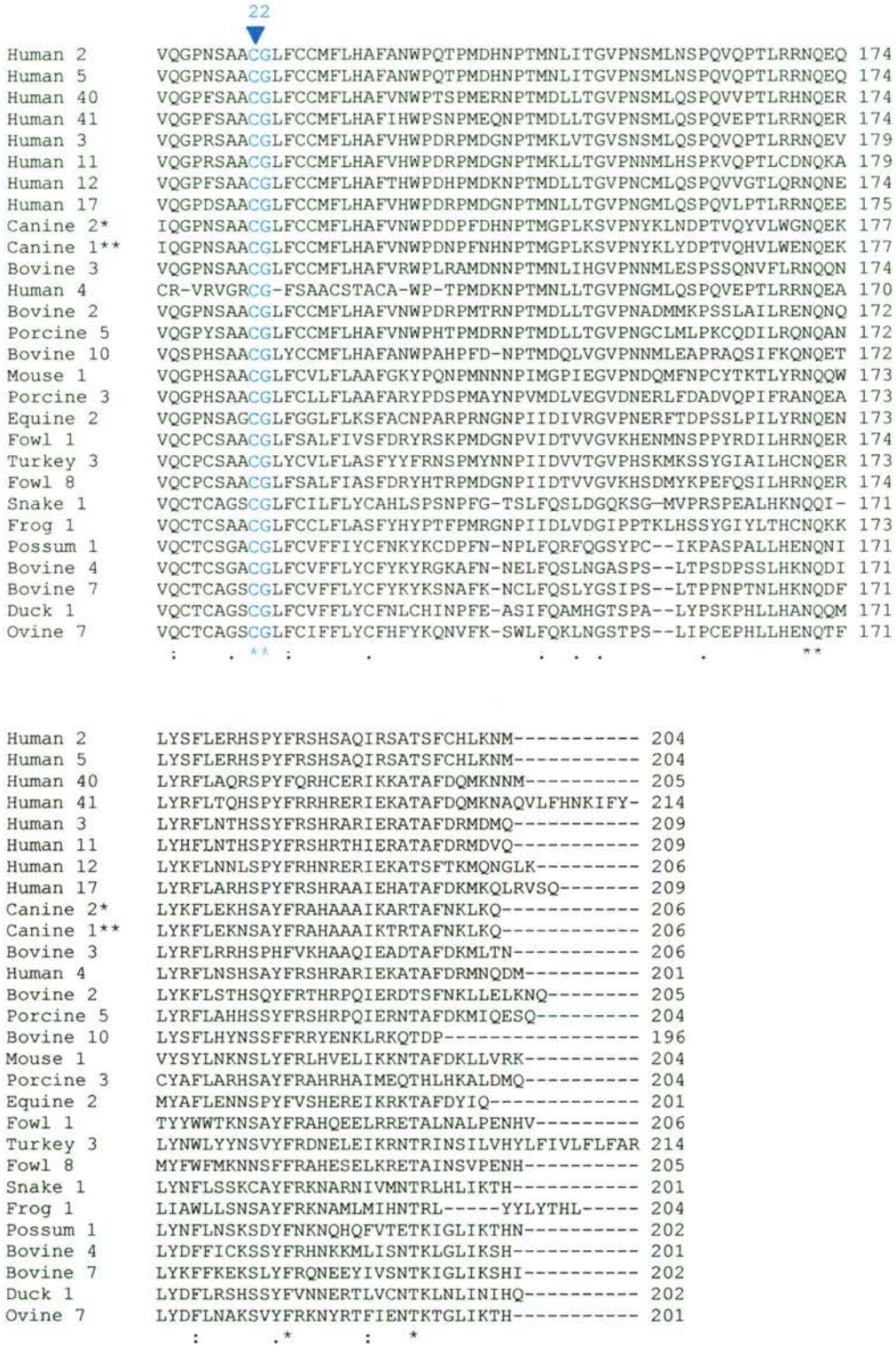

22
Human 2 VQGPNSAACGLFCCMFLHAFANWPQTPMDHNPTMNLITGVPNSMLNSPQVQPTLRRNQEQ 174
Human 5 VQGPNSAACGLFCCMFLHAFANWPQTPMDHNPTMNLITGVPNSMLNSPQVQPTLRRNQEQ 174
Human 40 VQGPFSAACGLFCCMFLHAFVNWPTSPMERNPTMDLLTGVPNSMLQSPQVVPTLRHNQER 174
Human 41 VQGPFSAACGLFCCMFLHAFIHWPSNPMEQNPTMDLLTGVPNSMLQSPQVEPTLRRNQER 174
Human 3 VQGPRSAACGLFCCMFLHAFVHWPDRPMDGNPTMKLVTGVSNSMLQSPQVQPTLRRNQEV 179
Human 11 VQGPRSAACGLFCCMFLHAFVHWPDRPMDGNPTMKLLTGVPNNMLHSPKVQPTLCDNQKA 179
Human 12 VQGPFSAACGLFCCMFLHAFTHWPDHPMDKNPTMDLLTGVPNCMLQSPQVVGTLQRNQNE 174
Human 17 VQGPDSAACGLFCCMFLHAFVHWPDRPMDGNPTMNLLTGVPNGMLQSPQVLPTLRRNQEE 175
Canine 2* IQGPNSAACGLFCCMFLHAFVNWPDDPFDHNPTMGPLKSVPNYKLNDPTVQYVLWGNQEK 177
Canine 1** IQGPNSAACGLFCCMFLHAFVNWPDNPFNHNPTMGPLKSVPNYKLYDPTVQHVLWENQEK 177
Bovine 3 VQGPNSAACGLFCCMFLHAFVRWPLRAMDNNPTMNLIHGVPNNMLESPSSQNVFLRNQQN 174
Human 4 CR-VRVGRCG-FSAACSTACA-WP-TPMDKNPTMNLLTGVPNGMLQSPQVEPTLRRNQEA 170
Bovine 2 VQGPNSAACGLFCCMFLHAFVNWPDRPMTRNPTMDLLTGVPNADMMKPSSLAILRENQNQ 172
Porcine 5 VQGPYSAACGLFCCMFLHAFVNWPHTPMDRNPTMDLLTGVPNGCLMLPKCQDILRQNQAN 172
Bovine 10 VQSPHSAACGLYCCMFLHAFANWPAHPFD-NPTMDQLVGVPNNMLEAPRAQSIFKQNQET 172
Mouse 1 VQGPHSAACGLFCVLFLAAFGKYPQNPMNNNPIMGPIEGVPNDQMFNPCYTKTLYRNQQW 173
Porcine 3 VQGPHSAACGLFCLLFLAAFARYPDSPMAYNPVMDLVEGVDNERLFDADVQPIFRANQEA 173
Equine 2 VQGPNSAGCGLFGGLFLKSFACNPARPRNGNPIIDIVRGVPNERFTDPSSLPILYRNQEN 173
Fowl 1 VQCPCSAACGLFSALFIVSFDRYRSKPMDGNPVIDTVVGVKHENMNSPPYRDILHRNQER 174
Turkey 3 VQCPCSAACGLYCVLFLASFYYFRNSPMYNNPIIDVVTGVPHSKMKSSYGIAILHCNQER 173
Fowl 8 VQCPCSAACGLFSALFIASFDRYHTRPMDGNPIIDTVVGVKHSDMYKPEFQSILHRNQER 174
Snake 1 VQCTCAGSCGLFCILFLYCAHLSPSNPFG-TSLFQSLDGQKSG-MVPRSPEALHKNQQI- 171
Frog 1 VQCTCSAACGLFCCLFLASFYHYPTFPMRGNPIIDLVDGIPPTKLHSSYGIYLTHCNQKK 173
Possum 1 VQCTCSGACGLFCVFFIYCFNKYKCDPFN-NPLFQRFQGSYPC--IKPASPALLHENQNI 171
Bovine 4 VQCTCSGACGLFCVFFLYCFYKYRGKAFN-NELFQSLNGASPS--LTPSDPSSLHKNQDI 171
Bovine 7 VQCTCAGSCGLFCVFFLYCFYKYKSNAFK-NCLFQSLYGSIPS--LTPPNPTNLHKNQDF 171
Duck 1 VQCTCAGSCGLFCVFFLYCFNLCHINPFE-ASIFQAMHGTSPA--LYPSKPHLLHANQQM 171
Ovine 7 VQCTCAGSCGLFCIFFLYCFHFYKQNVFK-SWLFQKLNGSTPS--LIPCEPHLLHENQTF 171
Human 2 LYSFLERHSPYFRSHSAQIRSATSFCHLKNM---------- 204
Human 5 LYSFLERHSPYFRSHSAQIRSATSFCHLKNM---------- 204
Human 40 LYRFLAQRSPYFQRHCERIKKATAFDQMKNNM---------- 205
Human 41 LYRFLTQHSPYFRRHRERIEKATAFDQMKNAQVLFHNKIFY- 214
Human 3 LYRFLNTHSSYFRSHRARIERATAFDRMDMQ---------- 209
Human 11 LYHFLNTHSPYFRSHRTHIERATAFDRMDVQ---------- 209
Human 12 LYKFLNNLSPYFRHNRERIEKATSFTKMQNGLK--------- 206
Human 17 LYRFLARHSPYFRSHRAAIEHATAFDKMKQLRVSQ------- 209
Canine 2* LYKFLEKHSAYFRAHAAAIKARTAFNKLKQ---------- 206
Canine 1** LYKFLEKNSAYFRAHAAAIKTRTAFNKLKQ---------- 206
Bovine 3 LYRFLRRHSPHFVKHAAQIEADTAFDKMLTN---------- 206
Human 4 LYRFLNSHSAYFRSHRARIEKATAFDRMNQDM---------- 201
Bovine 2 LYKFLSTHSQYFRTHRPQIERDTSFNKLLELKNQ-------- 205
Porcine 5 LYRFLAHHSSYFRSHRPQIERNTAFDKMIQESQ--------- 204
Bovine 10 LYSFLHYNSSFFRRYENKLRKQTDP----------------- 196
Mouse 1 VYSYLNKNSLYFRLHVELIKKNTAFDKLLVRK---------- 204
Porcine 3 CYAFLARHSAYFRAHRHAIMEQTHLHKALDMQ---------- 204
Equine 2 MYAFLENNSPYFVSHEREIKRKTAFDYIQ------------- 201
Fowl 1 TYYWWTKNSAYFRAHQEELRRETALNALPENHV--------- 206
Turkey 3 LYNWLYYNSVYFRDNELEIKRNTRINSILVHYLFIVLFLFAR 214
Fowl 8 MYFWFMKNNSFFRAHESELKRETAINSVPENH---------- 205
Snake 1 LYNFLSSKCAYFRKNARNIVMNTRLHLIKTH---------- 201
Frog 1 LIAWLLSNSAYFRKNAMLMIHNTRL-----YYLYTHL----- 204
Possum 1 LYNFLNSKSDYFNKNQHQFVTETKIGLIKTHN---------- 202
Bovine 4 LYDFFICKSSYFRHNKKMLISNTKLGLIKSH---------- 201
Bovine 7 LYKFFKEKSLYFRQNEEYIVSNTKIGLIKSHI--------- 202
Duck 1 LYDFLRSHSSYFVNNERTLVCNTKLNLINIHQ---------- 202
Ovine 7 LYDFLNAKSVYFRKNYRTFIENTKTGLIKTH---------- 201

Figure 3.1.8: Primary Structure Alignment of the CE Clan of Cysteine Proteases
The CE clan of cysteine proteases is composed of five families. (a) A representative member from each family was selected and the primary sequences were aligned using the CLUSTALW programme at EBI (http://www.ebi.ac.uk/clustalw). Family C5 was represented by Avp2 (human Ad2 protease); family C48 was represented by Ulp1 (*Saccharomyces cerevisiae* protease); family C63 was represented by ASFV (African swine fever virus, strain BA71V processing peptidase); family C57 was represented by Vaccinia virus 17L (Vaccinia virus, strain Copenhagen 17L processing peptidase); family C55 was represented by YopJ (*Yersinia pseudotuberculosis* protease). The known members of the catalytic triads for Avp2, Ulp1, ASFV, Vaccinia virus and, YopJ were highlighted in blue and putative members in red. The (b) core domain of each of the proteases was then aligned to show five conserved residues (highlighted in magenta) the catalytic histidine and adjacent tryptophan/phenyalanine, the catalytic glutamate/aspartate/aspargine and the catalytic cysteine with adjacent glycine.

(a) Alignment of Representative Members from each of the Five CE Clan Families

```
Avp2            ---------------MGSSEQELKAIVKDLGCGP-------------YFLGTYDKRFPGF 32
Ulp1            MSVEVDKHRNTLQYHKKNPYSPLFSPISTYRCYPRVLNNPSESRRSASFSGIYKKRTNTS 60
ASFV            ----------------------MSILEKITSSP-------------SECAEHLTNKDSC 24
Vaccinia        ------MERYTDLVISKIPELGFTNLLCHIYSLAGLCSNIDVSKFLTNCNGYVVEKYDKS 54
YopJ            ----------------------MIGPISQINIS----------------GGLSEKETSSL 22

Avp2            VSPHKLACAIVNTAGRETGGV----------------------HWMAFAWNPRSKTCYL 69
Ulp1            RFNYLNDRRVLSMEESMKDGSDRASKAGFIGGIRETLWNSGKYLWHTFVKNEPRNFDGSE 120
ASFV            LSKKIQKELTSFLEKKETLGCD----------------------SESCVITHPAVKAYAQ 62
Vaccinia        TTAGKVSCIPIGMMLELVESG-----------------HLSRPNSSDELDQKKELTDELK 97
YopJ            ISNEELKNIITQLETDISDGS--------------------WFHKNYSRMDVEVMPALV 61

Avp2            FEPFGFS-----------DQRLKQVYQFEYESLLRRSAIASS------------------ 100
Ulp1            VEASGNSDVESR-SSGSRSSDVPYGLRENYSSDTRKHKFDTSTWALPNKRRRIESEGVGT 179
ASFV            QKGLDLSKELET-RFKAPGPRNNTGLLTNFNIDETLQRWAIKYTKFFN------------ 109
Vaccinia        TRYHSIYDVFELPTSIPLAYFFKPRLREKVSKAIDFSQMDLKIDDLSRK------------ 146
YopJ            IQANNKYPEMNL-NLVTSPLDLSIEIKNVIENGVRSSRFIIN------------------ 102

Avp2            -----------------PDRCITLEKSTQSVQGPNSAACG------------LFCCMFLH 131
Ulp1            PSTSPISSLASQKSNCDSDNSITFSRDPFGWNKWKTSAIGSNSENNTSDQKNSYDRRQYG 239
ASFV            ----------------CPFSMMDFERVHYKFNQVDMVKVYKGEELQYVEGKVVKRPCNTF 153
Vaccinia        -------GIHTGENPKVVKMKIEPERGAWMSNRSIKNLVSQFAYGSEVDYIGQFDMRFLN 199
YopJ            ----------------MGEGGIHFSVIDYKHINGKTSLILFEPAN--------FNSMGPA 138

Avp2            AFANWPQTPMDHN------------------------------------------------ 144
Ulp1            TAFIRKKKVAKQNINNTKLVSRAQSEEVTYLRQIFNGEYKVPKILKEERERQLKLMDMDK 299
ASFV            GCVLNTDFSTGTG-------------------------------------KHWVAIFVDM 176
Vaccinia        SLAIHEKFDAFMNKHILSYILKDKIKSSTSRFVMFGFCYLSHWKCVIYDKKQCLVSFYDS 259
YopJ            MLAIRTKTAIER--------------------------YQLP-------DCHFSMVEMDI 165

Avp2            ------------------------------------------------------------
Ulp1            EKDTGLKKSIIDLTEKIKTILIENNKNRLQTRNENDDDLVFVKEKKISSLERKHKDYLNQ 359
ASFV            RG---------------------------------------------------------- 178
Vaccinia        GGNIP------------------------------------------------------TEF 267
YopJ            QRSS-------------------------------------------------------- 169

Avp2            ------PTMNLITGVPNS------------------------------------------ 156
Ulp1            KLKFDRSILEFEKDFKRYNEILNERKKIQEDLKKKKEQLAKKKLVPELNEKDDDQVQKAL 419
ASFV            ----DCWSIEYFNSAGNS------------------------------------------ 192
Vaccinia        HHYNNFYFYSFSDGFNTN------------------------------------------ 285
YopJ            ---SECGIFSFALAKKLY------------------------------------------ 184
```

```
Avp2        -------MLNSPQVQPTLRRNQ-------------EQLYSFLERHSP-------YFRSHS 189
Ulp1        ASRENTQLMNRDNIEITVRDFKTLAPRRWLNDTIIEFFMKYIEKSTPNTVAFNSFFYTNL 479
ASFV        --------PPGPVIRWMERVKQQLLKIHHTVKTLAVTNIRHQRSQTECGPYSLFYIRARL 244
Vaccinia    -HRHSVLDNTNCDIDVLFRFFECTFGAKIG--CINVEVNQLLESECGMFISLFMILCTRT 342
YopJ        IERDSLLKIHEDNIKGILSDGENPLPHDKLDPYLPVTFYKHTQGKKR----LNEYLNTNP 240

Avp2        AQ---------------------------------------IRSATSFCHLKNM---- 204
Ulp1        SERGYQGVRRWMKRKKTQIDKLDKIFTPINLNQSHWALGIIDLKKKTIGYVDSLSNGPNA 539
ASFV        DN---------------------------------------VSYAHFISARITDEDMY 263
Vaccinia    PP-----------------------------------KSFKSLKKVYTFFKFLADKKMT 366
YopJ        QE-----------------------------------LVLLLTKKMKPSLIDLITINPL- 264

Avp2        ------------------------------------------------------------
Ulp1        MSFAILTDLQKYVMEESKHTIGEDFDLIHLDCPQQPNGYDCGIYVCMNTLYGSADAPLDF 599
ASFV        KFRTHLFRIA---------------------------------------------------- 273
Vaccinia    LFKSILFNLHDLSLDITETDNAG-----LKEYKRMEKWTKKSINVICDKLTTKLNRIVDD 421
YopJ        ------------------------------------------------------------

Avp2        ----------------------
Ulp1        DYKDAIRMRRFIAHLILTDALK 621
ASFV        ----------------------
Vaccinia    DE--------------------- 423
YopJ        ----------------------
```

(b) Alignment of the Core Domain of Representative Members from each of the Five CE Clan Families

```
Avp2        HWMAFAWNPRSKTC--YLFEPFGFSDQRLKQVYQFEYESLLR------RSAIASSPDRCITLEKSTQS------VQGPNSAA-------CG
Ulp1        HWALGIIDLK-KKT-IGYVDSLSNGP------NAMSFAILTD-----LQKYVMEESKHTIGEDFDLIH------LDCPQQPN----GYDCG
ASFV        HWVAIFVDMRGDCWSIEYFNSAGNSP------PGPVIRWMER-----VKQQLLK-IHHTVKT-LAVTN------IRHQRSQT------ECG
Vaccinia    HWKCVIYDKK-QCL-VSFYDSGGNIPTEFHHYNNFYFYSFSDGFNTNHRHSVLDNTNCDIDVLFRFFECTFGAKIGCINVEVNQLLESECG
YopJ        HFSVIDYKHINGKTSLILFEPANFNS------MGPAMLAIRT------KTAIER--YQLPDCHFSMVE------MDIQRSSS------ECG
            *:     .          :. .              :       : :                    :   .         **
```

3.1.2.1.4 Residues Selected for Mutational Analysis

As the results have indicated, a number of residues were putative members of the S2 substrate specificity pocket, Ser3, Ser4, Asn44, Gly52, Val53, His54, Trp55, Cys122 and Gly123 (Figure 3.1.9(a)(i)). Likewise the residues, Met1, Glu5, Gln6, Asn44, Arg48, Gly51, Gly52 and Trp55 were potential members of the S4 substrate specificity pocket (Figure 3.1.9(b)(i)).

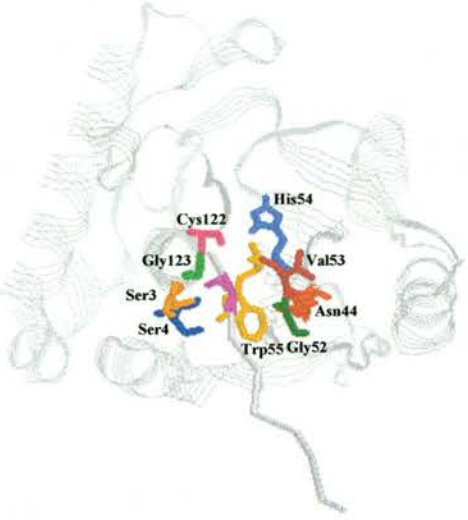
Whilst the mutational analysis in this study was being carried out, Balakirev et al., 2002 published a superimposition of Avp2 and YUH1 in complex with Ubal. As the substrate, Ubal sat in the putative substrate cleft for Avp2, this allowed the group to determine residues potentially involved the S2 and S4 substrate specificity pockets. The residues proposed were Ser3, Ser4, Val53 and Trp55 for the S2 pocket and for the S4 pocket, Glu5, Gln6, Arg48, Gly51 and Gly52, all of which corroborated the findings of Section 3.1.2.1.

Therefore the residues selected for mutational analysis in this thesis were Ser3, Ser4, Gly52, Val53 and Trp55 for the S2 pocket (Figure 3.1.9(a)(ii)) and for the S4 pocket Glu5, Gln6, Arg48, Gly51, Gly52 and Trp55 (Figure 3.1.9(b)(ii)).

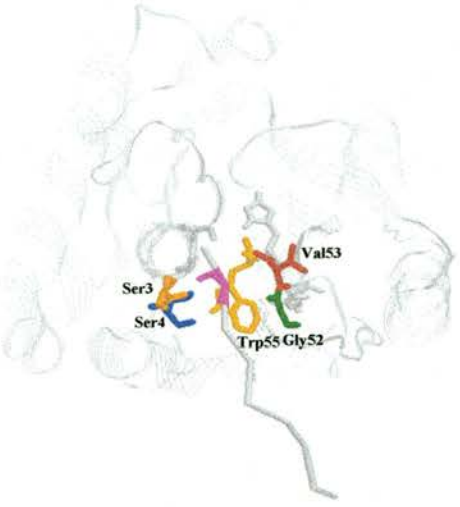
Figure 3.1.9: Mutational Analysis of the S2 and S4 Substrate Specificity Pockets
The Avp2 (grey/strands) – Smt3 (grey/stick) model was viewed in the RASMOL Version 2.6 computer package (Sayle & Milner-White., 1995) and the P2 (Gly97) and P4 (Gln95) residues of Smt3 were highlighted in magenta. The putative members of the (a)(i) S2 pocket were Ser3 (orange), Ser4 (dark blue), Asn44 (red), Gly52 (dark green), Val53 (dark brown), His54 (petrol), Trp55 (sulphur yellow), Cys122 (pink) and Gly123 (green), whilst those of the (b)(i) S4 pocket were Met1 (yellow), Glu5 (purple), Gln6 (cyan), Asn44 (red), Arg48 (blue), Gly51 (light brown), Gly52 (dark green) and Trp55 (sulphur yellow). The residues mutated were (a)(ii) Ser3, Ser4, Gly52, Val53 and Trp55 for the S2 pocket and (b)(ii) Glu5, Gln6, Arg48, Gly51, Gly52 and Trp55 for the S4 pocket.

(a) S2 Substrate Specificity Pocket

(i) Putative Members

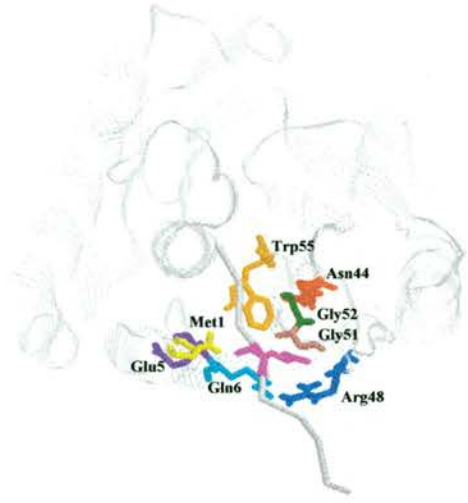


(ii) Residues Selected for Alanine Screen

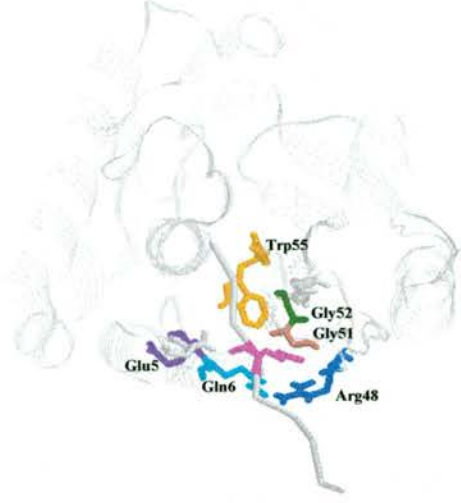


(a) S4 Substrate Specificity Pocket

(i) Putative Members



(ii) Residues Selected for Alanine Screen



3.1.2.2 Smt3 and SUMO1 Cleavage by Avp2

Prior to carrying out an Alanine screen of the residues thought to be involved in Avp2 substrate specificity, according to the Avp2/Ulp1-Smt3 superimposition (Section 3.1.2.1), it was important to analyse whether or not Smt3 was a substrate for Avp2. Therefore, a sample of Smt3 was purified and subjected to cleavage by Avp2.

Following this, the cleavage of SUMO1 (human homologue of Smt3) by Avp2 was also investigated.

3.1.2.2.1 Smt3ET28b Purification

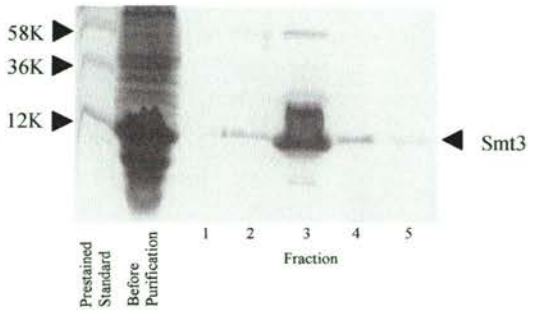
The DNA coding for the C-terminal His tagged Smt3 was ligated into the pET28b vector (Novagen) between the restriction sites NcoI and XhoI (a gift from Prof. C.D. Lima). This vector was then expressed in the BL21(DE3) cell line (Novagen). The protocol for the expression and purification of Smt3ET28b was modified from the paper by Mossessova & Lima., 2000.

Smt3pET28b expressed in the BL21(DE3) cell line was grown in LB/Kanamycin according to Method 2.5.2. The cells were then induced with 0.4mM IPTG for 6hrs at 25°C. Following this they were re-suspended in 8ml of Suspension buffer (20mM Tris-HCl, 10mM Imidazole, pH8).

The protein was then purified according to Method 2.5.5 and separated on a 20% SDS-PAGE mini-gel (Method 2.4.1.1) which was stained with Coomassie Blue (Method 2.4.2.1). As Figure 3.1.10 shows, the protein was eluted between fractions 2-5. These bands were confirmed to be Smt3 by Mass Spectrometry (Method 2.4.7.2).

Following elution, Fraction 3 was dialysed against 20mM Tris-HCl, pH8 for 4hr with two changes of buffer, to remove Imidazole.

Figure 3.1.10: Nickel Bead Purification of Smt3ET28b
Smt3pET28b was expressed in the BL21(DE3) cell line. Following growth, induction and extraction, Smt3ET28b was passed through a 1 x 5cm Nickel Column in the presence of 20mM Tris-HCl, 20mM Imidazole, pH8. The protein was eluted from the column in the presence of 20mM Tris-HCl, 250mM Imidazole, pH8 and 2ml fractions were collected at 1ml/min. The fractions were separated on a 20% SDS-PAGE mini-gel which was stained with Coomassie Blue.



3.1.2.2.2 Smt3 Cleavage by Avp2

The primary amino acid sequence of the Smt3 substrate used in the cleavage analysis is given in Figure 3.1.11. The vector was constructed by Prof. Lima and colleagues and during cloning the first N-terminal residues were omitted (highlighted in green). The C-terminus was also modified by a polylinker region (highlighted in red) followed by a His tag (highlighted in blue).

Figure 3.1.11: Smt3 Sequence
The primary amino acid sequence of Smt3 used in the cleavage analysis is given below. During cloning, the first 12 N-terminal residues were omitted (highlighted in green). The C-terminus was also modified to contain a polylinker region (highlighted in red), and a His tag (highlighted in blue). Preceding the His tag is the putative cleavage site for Avp, highlighted by a purple triangle.

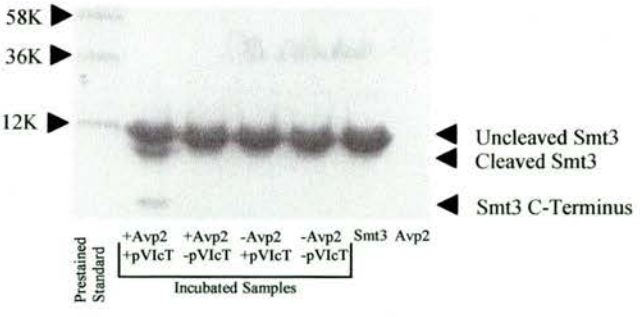
----------EVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKR
N-Terminus Removed

QGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGATYLEHHHHHH
Polylinker Region
His tag

The cleavage assay was set up with +Avp2/+pVIcT, +Avp2/-pVIcT, -Avp2/+pVIcT; and -Avp2/-pVIcT (Method 2.7.1). The assay was incubated for 24hr before being separated on a 20% SDS-PAGE mini-gel (Method 2.4.1.1), which was then stained with Coomassie Blue (Method 2.4.2.1). The results of this gel, showed that Smt3 was only cleaved in the presence of both Avp2 and pVIcT, as it was only in this lane that two extra bands appeared (Figure 3.1.12).

The top two bands of the +Avp2/+pVIcT lane were confirmed to be Smt3 by Mass Spectrometry (Method 2.4.7.2), whilst Protein Sequencing (Method 2.4.6.2) identified the lower band as the C-terminus of Smt3, ATYLEH. Therefore cleavage occurred at the proposed cleavage site, QIGG-A (highlighted by a purple triangle in Figure 3.1.11).

Figure 3.1.12: Cleavage Analysis of Smt3 by Avp2
The cleavage assay was set up with 1mg/ml Smt3, 1μM Avp2, 100μM pVIcT, 12.5mM Tris-HCl, 2.5mM EDTA, 0.5μM Pepstatin, 0.5mM AEBSF, 0.5mM β-Mercaptoethanol, pH8. There were also three controls set-up, +Avp2/-pVIcT, -Avp2/+pVIcT and -Avp2/-pVIcT. The assay was left to incubate for 24hr before being separated on a 20% SDS-PAGE mini-gel, which stained with Coomassie Blue.



3.1.2.2.3 SUMO1 Cleavage by Avp2

The purified human SUMO1 protein was a gift from Dr. L. Shen. The DNA was cloned into the pGEX2T vector (Amersham Biosciences) between BamHI and EcoRI, and expressed in the BL21(DE3) cell line. The protein was then purified on a Glutathione

column. Following this, the N-terminal GST tag was removed to leave a glycine and a serine at the N-terminus of SUMO1 (highlighted in green in Figure 3.1.13(a)). The SUMO1 protein was then further purified on a Glutathione column.

SUMO1 not only had additional amino acids at the N-terminus, but also at the C-terminus, where a cysteine was cloned in (highlighted in red in Figure 3.1.13(a)). The reason for both cloning a cysteine at the C-terminus and mutating Cys54 to an alanine (highlighted in red in Figure 3.1.13(a)) was to allow the fluorescent tag, 5-iodoacetamidofluorescein (5-IAF) (Invitrogen – Molecular Probes) to be linked to the C-terminus of the protein (Figure 3.1.13(b)).

The tag gave the advantage of allowing the visualisation of the SUMO1 and its product were it to be cleaved by Avp2 at its putative cleavage site (Figure 3.1.13(a)).

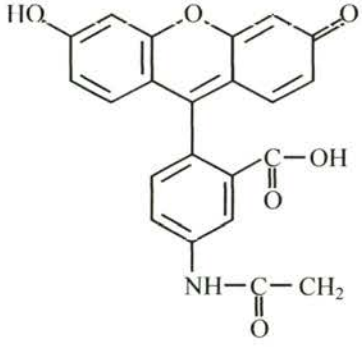
Figure 3.1.13: SUMO1 Sequence
Human SUMO1 protein was cloned into the pGEX2T vector and then purified using a Glutathione column. As the (a) cloned protein sequence shows there were two polylinker residues at the N-terminus (green), Cys54 was mutated to an alanine (red) and a cysteine residue was cloned in at the C-terminus. The reason for the alanine substitution and C-terminal cysteine addition was to allow the fluorescent tag, (b) 5-iodoacetamidofluorescein (5-IAF) (Invitrogen – Molecular Probes) to be linked at the C-terminus of the protein. This had the advantage of allowing the visualisation of the SUMO1 and its product were it to be cleaved by Avp2 at the (a) putative cleavage site for Avp2, marked by a purple triangle.

(a) SUMO1 Sequence

GSMSDQEAKPSTEDLGDKKEGEYIKLKVIGQDSSEIHFKVKMTTHLKKLKESYAQRQGVP 60
Polylinker Region

MNSLRFLFEGQRIADNHTPKELGMEEEDVIEVYQEQTGGHSTVC 104

(b) Structure of 5-IAF



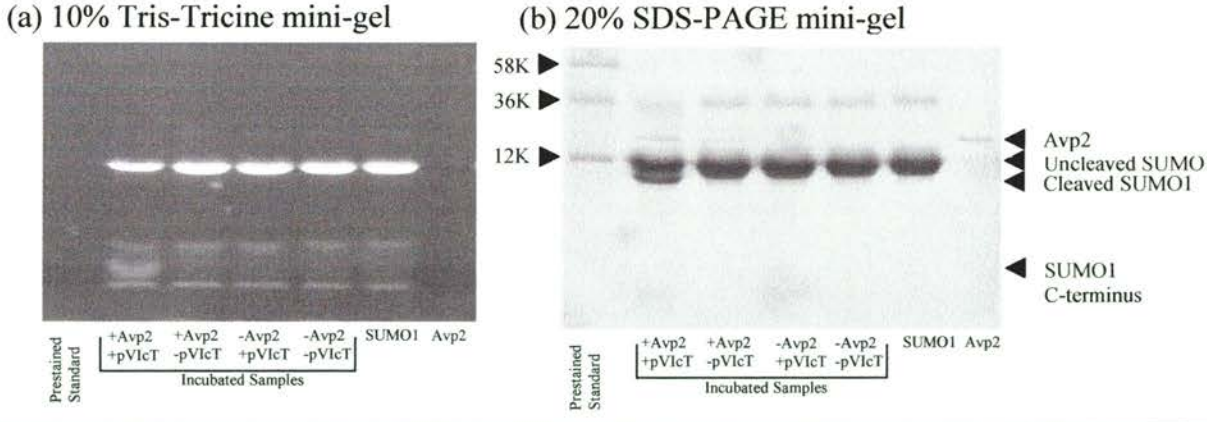
Molecular Mass = 515.26Da

The cleavage assay was set up with +Avp2/+pVIcT, +Avp2/-pVIcT, -Avp2/+pVIcT; and -Avp2/-pVIcT (Method 2.7.1). The assay was incubated for 24hr before being separated on a 10% Tris-Tricine mini-gel (Method 2.4.1.2) and 20% SDS-PAGE mini-gel (Method 2.4.1.1). The Tris-Tricine mini-gel was placed under a UV light to visualise any bands containing the 5-IAF fluorescent tag and the SDS-PAGE mini-gel was stained with Coomassie Blue (Method 2.4.2.1). These results are shown in Figure 3.1.14.

The results showed that SUMO1 was only cleaved in the presence of both Avp2 and pVIcT, because in the Tris-Tricine mini-gel the SUMO1 protein band reduced in intensity and an extra band was seen (Figure 3.1.14(a)). Likewise, an extra band was seen just below the uncleaved SUMO1 protein in the SDS-PAGE mini-gel (Figure 3.1.14(b)).

The top two bands on the SDS-PAGE mini-gel (Figure 3.1.14(b)) were confirmed to be SUMO1 by Mass Spectrometry (Method 2.4.7.2), whilst Protein Sequencing (Method 2.4.6.2) identified the lower band on the Tris-Tricine mini-gel (Figure 3.1.14(a)) as the C-terminus of SUMO1, HSTV. Therefore cleavage occurred at the proposed cleavage site, QTGG-H (highlighted by a purple triangle in Figure 3.1.13).

Figure 3.1.14: Cleavage Analysis of SUMO1 by Avp2
The cleavage assay was set up with 1mg/ml SUMO1, 1μM Avp2, 100μM pVIcT, 12.5mM Tris-HCl, 2.5mM EDTA, 0.5μM Pepstatin, 0.5mM AEBSF, 0.5mM β-Mercaptoethanol, pH8. There were also three controls set-up, +Avp2/-pVIcT, -Avp2/+pVIcT and -Avp2/-pVIcT. The assay was left to incubate for 24hr before being separated on (a) 10% Tris-Tricine min-gel and a (b) 20% SDS-PAGE mini-gel which were visualised by placing the gels under a UV light and by staining with Coomassie Blue, respectively.



3.1.2.3 Cloning of Mutants into pET11c for Alanine Screen

After determining which residues were potentially involved in binding the substrate (Section 3.1.2.1), an Alanine screen was carried out. W55A was cloned by Dr. S. Jones, whilst the other eight mutants were produced using one of three PCR protocols.

The first method used forward and reverse primers; the second used a forward primer, an internal reverse primer and a reverse primer; and the third method used a forward primer, an internal forward primer and a reverse primer. These inserts were then cloned into the pET11c vector (Novagen.) and expressed in the Rosetta(DE3) host cell line (Novagen).

3.1.2.3.1 Cloning of S3A, S4A, E5A and Q6A

The PCR reaction for S3A, S4A, E5A and Q6A used a mutant forward primer and a WT reverse primer. As these mutants were at the N-terminus, there was no requirement to use an internal primer. A mini-prep (Method 2.2.7) was made of WT Avp2pET11c vector and this was used as a template for the PCR reaction (Method 2.2.1.1). The forward and reverse primers in Table 3.1.4 were used to PCR the Avp2 fragment for S3A, S4A, E5A and Q6A. The PCR reaction was carried out at 94°C for 5min (hot start), 94°C for 30 sec, 50°C for 30sec, 72°C for 45sec, 72°C for 7min over 25 cycles.

The vector and insert were then digested with NdeI and BamHI (Method 2.2.4), ligated (Method 2.2.5), used to transform the JM109 cell line (Promega) (Method 2.2.6) and left to grow on an LB-agar/Carbenicillin plate (Method 2.2.10.1) at 37°C for 12-16hr . Six colonies were selected from the transformants and a mini-prep was carried out on each. 10µl of each mini-prep was double digested by XbaI (restriction site of the vector) and BamHI and then separated by 0.9% (w/v) Agarose gel electrophoresis (Method 2.2.2).

Finally one of the colonies which appeared to be positive according to the double digestion was identified by DNA sequencing (Method 2.2.9).

After confirmation by DNA sequencing, a glycerol stock (Method 2.2.10.2) was made of the JM109 cell line. The vector was then transformed into the Rosetta(DE3) cell line (Novagen) and a glycerol stock was made of these.

Table 3.1.4: Primers Used to Clone S3A, S4A, E5A and Q6A into the pET11c Vector
The PCR reaction for S3A, S4A, E5A and Q6A used a mutant forward primer and a WT reverse primer. The mutations in the WT sequence used to produce the mutants are highlighted in red.

Primer	Oligonucleotide primer sequence (5'-3')
S3A forward	GTAATTACATATGGGCGCGAGTGAGCAGGAACTG
S4A forward	GTAATTACATATGGGCTCCGCGGAGCAGGAACTG
E5A forward	GTAATTACATATGGGCTCCAGTGCGCAGGAACTG
Q6A forward	GTAATTACATATGGGCTCCAGTGAGGCGGAACTG
Reverse	GCGGGATCCTTACATGTTTTTCAAGTGACAAAAAGA

3.1.2.3.2 Cloning of R48A, G51A, G52A and V53A

The amplification of the internal mutants – R48A, G51A, G52A and V53A was complex, requiring two PCR steps and three primers. A mini-prep (Method 2.2.7) was made of WT Avp2pET11c vector and this was used as a template for both PCR reactions (Method 2.2.1.2).

The first PCR required the WT forward primer and the mutant reverse internal primers in Table 3.1.5 to PCR a forward mega-primer for R48A, G51A, G52A and V53A. The PCR reaction was carried out at 94°C for 5min (hot start), 94°C for 30 sec, 50°C for 30sec, 72°C for 45sec, 72°C for 7min over 25 cycles.

The mega-primer for each mutant was separated by 0.9% (w/v) Agarose gel electrophoresis (Method 2.2.2) and then gel extracted (Method 2.2.3). The mega-primer

was used with the WT reverse primer in Table 3.1.5 in a second PCR reaction (Method 2.2.1.2). The PCR reaction was carried out at 94°C for 5min (hot start), 94°C for 30 sec, 55°C for 60sec, 72°C for 60sec, 72°C for 7min over 25 cycles.

Unfortunately, G51A mega-primer created no secondary PCR product. Therefore an increase in Magnesium concentration, a decrease in annealing temperature, an increase in annealing time, and an increase/decrease in extension time were all investigated (data not shown) but none of these produced a PCR product. Thus another PCR method was attempted (Section 3.1.2.3.3).

For the other three mutants, the vector and insert were digested with NdeI and BamHI (Method 2.2.4), ligated into pET11c (Method 2.2.5), used to transform the JM109 cell line (Promega) (Method 2.2.6) and left to grow on an LB-agar/Carbenicillin plate at 37°C for 12-16hr (Method 2.2.10.1). Six colonies were selected from the transformants and a mini-prep was carried out on each. 10µl of each mini-prep was double digested by XbaI (restriction site of the vector) and BamHI and then separated by 0.9%(w/v) Agarose gel electrophoresis (Method 2.2.2). Finally one of the colonies which appeared to be positive according to the double digestion was identified by DNA sequencing (Method 2.2.9).

After confirmation by DNA sequencing, a glycerol stock (Method 2.2.10.2) was made of the JM109 cell line. The vector was then transformed into the Rosetta(DE3) cell line (Novagen) and a glycerol stock was made of these.

Table 3.1.5: Primers Used to Clone R48A, G51A, G52A and V53A into the pET11c Vector

The PCR reaction for R48A, G51A, G52A and V53A used a WT forward primer, a mutant reverse internal primer and a WT reverse primer. The mutations in the WT sequence used to produce the mutants are highlighted in red.

Primer	Oligonucleotide primer sequence (5'-3')
Forward	GTAATTACATATGGGCTCCAGTGAGCAGGAACTG
R48A reverse internal	CCCAGTCTCCGCACCGGCCGTATTGAC
G51A reverse internal	CATCCAGTGTACGCCCGCAGTCTCGCGACC
G52A reverse internal	CATCCAGTGTACCGCCCCAGTCTCGCGACC
V53A reverse internal	GGCAAAGGCCATCCAGTGCGCGCCCCCAGTCTC
Reverse	GCGGGATCCTTACATGTTTTTCAAGTGACAAAAAGA

3.1.2.3.3 Cloning of G51A

The PCR of the internal mutant, G51A required two PCR steps and three primers. A mini-prep (Method 2.2.7) was made of WT Avp2pET11c vector and this was used as a template for both PCR reactions (Method 2.2.1.3).

The first PCR required the mutant internal forward primer and the WT reverse primer in Table 3.1.6 to PCR a reverse mega-primer fragment for G51A. The PCR reaction was carried out at 94°C for 5min (hot start), 94°C for 30 sec, 55°C for 60sec, 72°C for 45sec, 72°C for 7min over 25 cycles.

The reverse mega-primer was separated on a 0.9% (w/v) Agarose gel (Method 2.2.2) and then gel extracted (Method 2.2.3). The mega-primer was used with the WT forward primer in Table 3.1.6 in a second PCR reaction (Method 2.2.1.3). The PCR reaction was carried out at 94°C for 5min (hot start), 94°C for 30 sec, 55°C for 60sec, 72°C for 60sec, 72°C for 7min over 25 cycles.

The vector and insert were then digested with NdeI and BamHI (Method 2.2.4), ligated into pET11c (Method 2.2.5), used to transform the JM109 cell line (Promega) (Method 2.2.6) and left to grow on an LB-agar/Carbenicillin plate at 37°C for 12-16hr . Six

colonies were selected from the transformants and a mini-prep was carried out on each. 10µl of each mini-prep was double digested by XbaI (restriction site of the vector) and BamHI and then separated by Agarose gel electrophoresis (Method 2.2.2). Finally one of the colonies which appeared to be positive according to the double digestion was identified by DNA sequencing (Method 2.2.9).

After confirmation by DNA sequencing, a glycerol stock (Method 2.2.10.2) was made of the JM109 cell line. The vector was then transformed into the Rosetta(DE3) cell line (Novagen) and a glycerol stock was made of these.

Table 3.1.6: Primers Used to Clone G51A into the pET11c Vector
The PCR reaction for G51A used a mutant forward internal primer, a WT forward primer and a WT reverse primer. The mutation in the WT sequence to produce the mutant is highlighted in red.

Primer	**Oligonucleotide primer sequence (5'-3')**
Forward	GTAATTACATATGGGCTCCAGTGAGCAGGAACTG
G51A forward internal	CGCGAGACTGCGGGCGTACAC
Reverse	GCGGGATCCTTACATGTTTTTCAAGTGACAAAAAGA

3.1.2.4 Determination of Specific Activities

The Alanine screen mutants expressed in the Rosetta(DE3) cell line (Section 3.1.2.3) were grown in LB/Carbenicillin/Chloramphenicol according to Method 2.5.2. They were induced with 0.6mM IPTG for 4hr at 37°C (Method 2.5.2). Following this, they were extracted (Method 2.5.3.1), re-suspended in 8ml of Suspension buffer (50mM Tris, 5mM EDTA, 4% (v/v) Glycerol) and purified (Method 2.5.4). However, G52A was expressed in too low concentrations to be purified (data not shown).

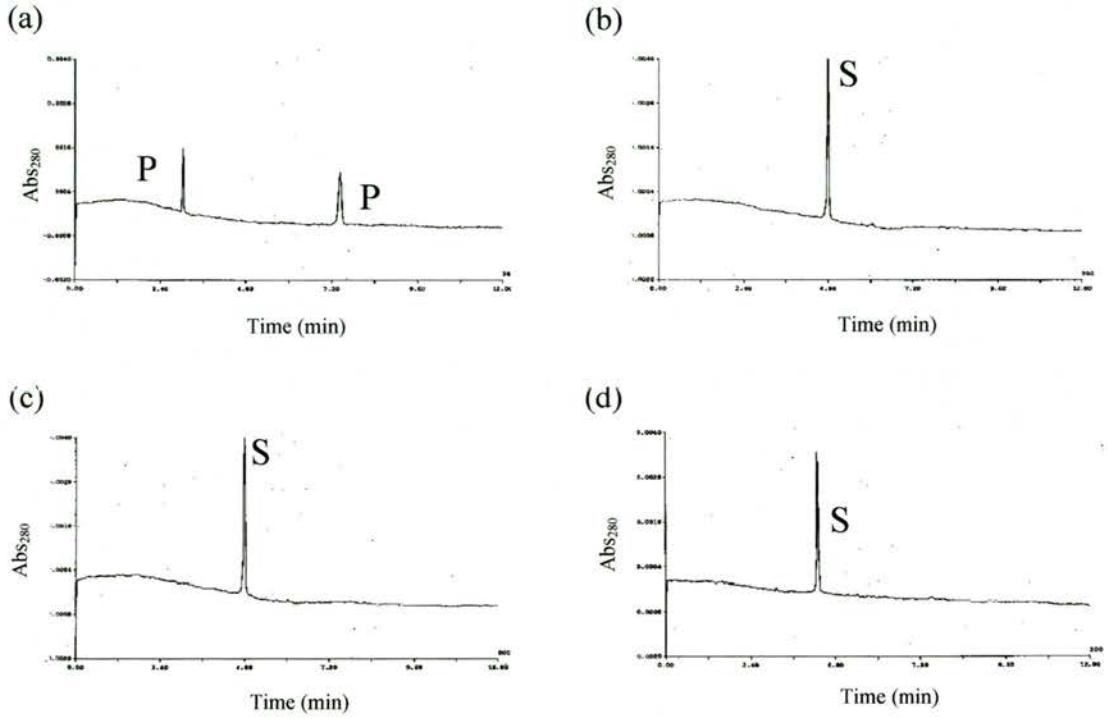
The Specific activities of the mutants were then determined by CE analysis and compared to that of WT Avp2 (Method 2.6).

3.1.2.4.1 Analysing Cleavage of AcLRGAGRSR over a 24hr period

Prior to the determination of the Specific activity for WT or any of the mutant Avp2 proteins, an assay was set up. The assay was incubated at 37°C for 24hr with 229μM AcLRGAGRSR, 0.02μM Avp2, 2μM pVIcT made up to 50μl with 50mM Tris, 10mM EDTA, 2μM pepstatin, 2mM AEBSF, pH8. Three controls were also set up with +Avp2/-pVIcT, -Avp2/+pVIcT, and -Avp2/-pVIcT. Following incubation, the assay mix was investigated by CE analysis, according to Method 2.6.2.

The results from this investigation indicated that for WT and for all of the mutants there was only cleavage in the presence of both Avp2 and pVIcT (Figure 3.1.15). This was confirmed by Mass spectrometry (Method 2.4.7.1) which showed that cleavage occurred at the consensus site. Therefore the substrate, AcLRGAGRSR was cleaved to the two products, AcLRGA and GRSR.

Figure 3.1.15: Example Cleavage of AcLRGAGRSR Over a 24hr Period
Prior to the determination of the Specific activity for WT or any of the mutant Avp2 proteins, an assay was set up to follow the cleavage of AcLRGAGRSR (**S**) to AcLRGA (**P**) and GRSR (**P**). The assay was incubated at 37°C for 24hr with 229μM AcLRGAGRSR (**S**), 0.02μM Avp2, 2μM pVIcT made up to 50μl with 50mM Tris, 10mM EDTA, 2μM pepstatin, 2mM AEBSF, pH8. Three controls were also set up with (b) +Avp2/-pVIcT, (c) -Avp2/+pVIcT, and (d) -Avp2/-pVIcT. Following incubation, the assay mix was investigated by CE analysis.



3.1.2.4.2 Protease Concentrations and Times used for Assays

Initial rates of cleavage were calculated when less than 20% of the substrate had been cleaved (Cabrita., 1997). Therefore because each mutant protease investigated had a different level of activity, both the concentration of protease added to the assay mix and the length of time taken to digest less than 20% of the substrate had to be determined experimentally preceding the calculation of the Specific Activity. The concentrations of protease and the time points taken are given in Table 3.1.7.

Table 3.1.7: Concentrations of Protease and Time Points Taken
Initial rates of cleavage were calculated when less than 20% of the substrate had been cleaved. Therefore because each mutant protease investigated had a different level of activity, both the concentration of protease added to the assay mix and the length of time taken to digest less than 20% of the substrate had to be determined experimentally preceding the calculation of the Specific activity.

Protease	Final Protease Concentration (μM)	Time Points Taken (min)
WT	0.109	0, 5, 10, 15, 30
S3A	0.069	0, 60, 120, 180, 240
S4A	0.127	0, 5, 10, 15, 30
E5A	0.669	0, 15, 30, 45, 60
Q6A	0.102	0, 5, 10, 15, 30
R48A	0.149	0, 5, 10, 15, 30
G51A	0.086	0, 5, 10 ,15 ,30
G52A	Not done	Not done
V53A	0.486	0, 5, 10, 15, 30
W55A	0.017	0, 150, 345, 450, 600

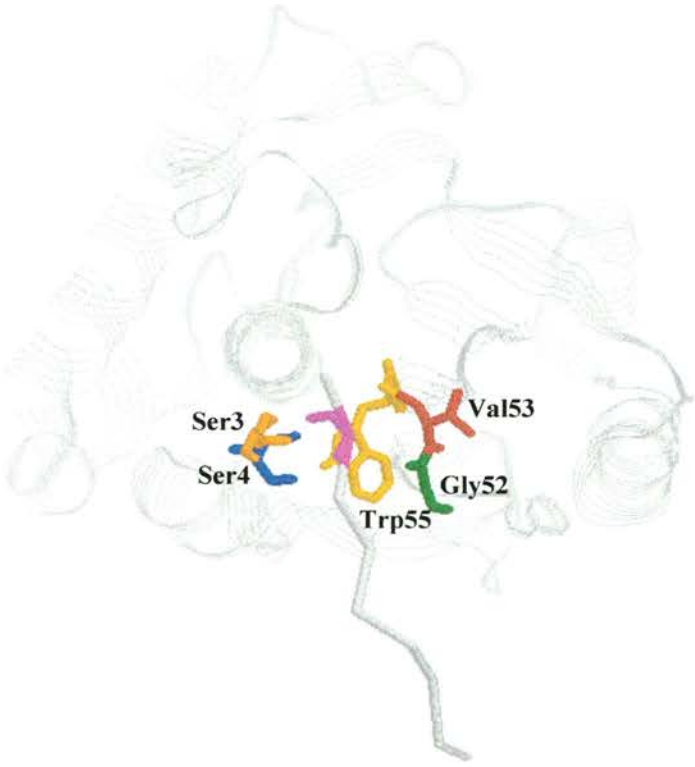
3.1.2.4.3 Specific Activities Determined by CE Analysis

The Specific activities were determined in quadruplicate and the results are shown in Figure 3.1.16 and 3.1.17. As discussed in Section 3.1.2.1.4, the residues putatively involved in the S2 substrate specificity pocket which were mutated in the Alanine screen were, Ser3, Ser4, Gly52, Val53 and Trp55. Whilst those of the S4 pocket were, Glu5, Gln6, Arg48, Gly51, Gly52 and Trp55.

From the Specific activity analysis of these mutations, it appeared that the S2 substrate specificity pocket mutations, S3A, V53A and W55A all caused a large reduction in activity, whilst S4A caused no reduction in activity (Figure 3.1.16(b)). Similarly for the S4 site, whilst E5A and W55A caused a large reduction in activity, R48A and G51A caused a small reduction in activity and Q6A caused a small increase in activity (Figure 3.1.17(b)).

Figure 3.1.16: Specific Activities for the S2 Substrate Specificity Pocket Mutants
The (a) Avp2 (grey/strands) – Smt3 (grey/stick) model was viewed in the RASMOL Version 2.6 computer package (Sayle & Milner-White., 1995) and the P2 (Gly97) residue of Smt3 was highlighted in magenta. The residues putatively involved in the S2 substrate specificity pocket which were mutated in the Alanine screen were Ser3 (orange), Ser4 (dark blue), Gly52 (dark green), Val53 (dark brown) and Trp55 (sulphur yellow). The Specific activities determined in quadruplicate for the putative S2 substrate specificity pocket mutants are given in the (b) table.

(a) Putative Residues Involved in the S2 Substrate Specificity Pocket.

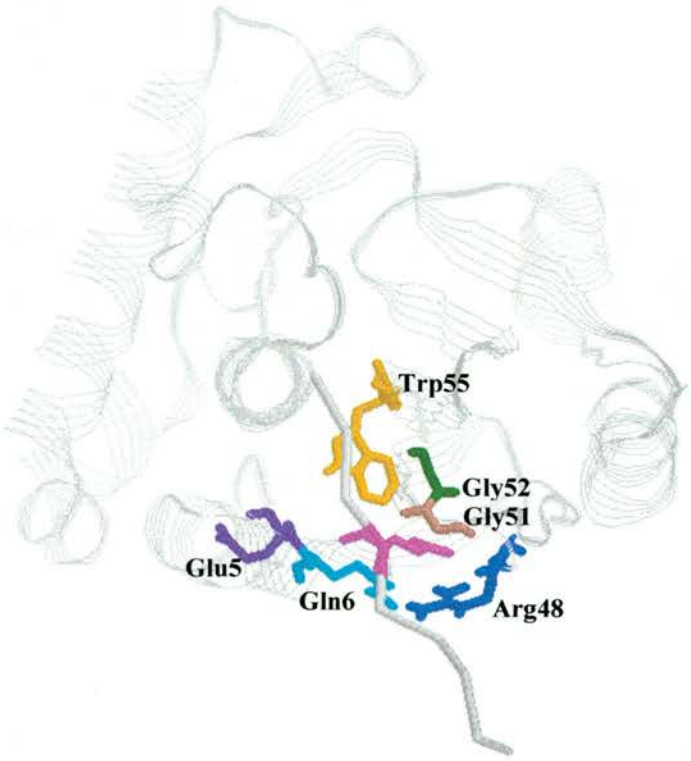


(b) Table of Specific Activities for S2 Substrate Specificity Pocket Mutants

Mutant	Mean Specific Activity +/- S.D (mol substrate transformed/min/mol enzyme)	Specific Activity in comparison to WT (%)
WT	11.22 +/- 0.62	100
S3A	2.49 +/- 0.19	22
S4A	10.94 +/- 0.54	97
G52A	Not done	Not done
V53A	2.95 +/- 0.33	26
W55A	1.76 +/- 0.76	16

Figure 3.1.17: Specific Activities for the S4 Substrate Specificity Pocket Mutants
The (a) Avp2 (grey/strands) – Smt3 (grey/stick) model was viewed in the RASMOL Version 2.6 computer package (Sayle & Milner-White, 1995) and the P4 (Gln95) residue of Smt3 was highlighted in magenta. The residues putatively involved in the S4 substrate specificity pocket which were mutated in the Alanine screen were Glu5 (purple), Gln6 (cyan), Arg48 (blue), Gly51 (light brown), Gly52 (dark green) and Trp55 (sulphur yellow). The Specific activities determined in quadruplicate for the putative S4 substrate specificity pocket mutants are given in the (b) table.

(a) Putative Residues Involved in the S4 Substrate Specificity Pocket



(b) Table of Specific Activities for S4 Substrate Specificity Pocket Mutants

Mutant	Mean Specific Activity +/- S.D (mol substrate transformed/min/mol enzyme)	Specific Activity in comparison to WT (%)
WT	11.22 +/- 0.62	100
E5A	0.45 +/- 0.03	4
Q6A	15.01 +/- 2.58	134
R48A	9.30 +/- 0.54	83
G51A	9.82 +/- 1.01	87
G52A	Not done	Not done
W55A	1.76 +/- 0.76	16

3.1.2.5 Comparison of the 2.6 and 1.6 Angstrom Resolution Structure of Avp2

The 2.6 Angstrom (Ding et al., 1996) and 1.6 Angstrom (McGrath et al., 2003) structures of Avp2 are identical, with the exception of the seven N-terminal residues. The reason for this was that Met1 could not be found in the 1.6 Angstrom structure which led to the remodelling of the six residues preceding this residue.

Within this region are four residues which have previously been selected as putative members of the S2 and S4 substrate specificity pockets, Ser3/Ser4 and Glu5/Gln6, respectively (Section 3.1.2.1). The placement of the residues in both structures was compared using the RASMOL Version 2.6 computer package (Sayle & Milner-White, 1995) (Figure 3.1.18 and Figure 3.1.19)). As can be seen from the Avp2-Smt3 model of the S2 pocket (Figure 3.1.18(a)), the R-chains of Ser3 and Ser4 residues which faced into the S2 pocket in the 2.6 Angstrom structure, no longer do so in the 1.6 Angstrom structure. Likewise, the residues, Glu5 and Gln6 of the P4 site have virtually swapped places (Figure 3.1.19(a)).

Following this work, the Radius function of the Swiss-PDB viewer Computer Programme (Guex & Peitsch., 1997; http://www.expasy.org/spdbv/) (Method 2.1.2.2) was used to measure the distance between the Ser3, Ser4, Glu5 and Gln6 α-Carbons of Avp2 and the respective Smt3, P2 or P4 residue α-Carbons. This was carried out with the Avp2-Smt3 model, using either the 2.6 or 1.6 Angstrom structure of Avp2. These results are shown in Figure 3.1.18(b) and 3.1.19(b).

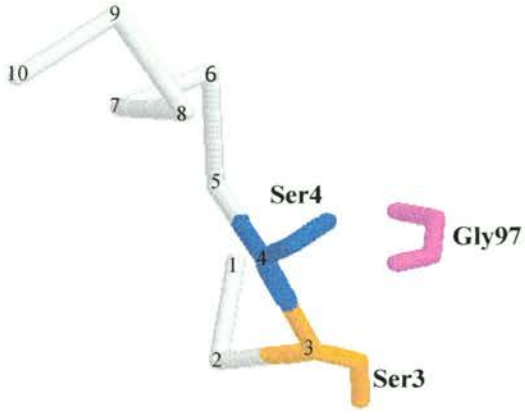
The results of Figure 3.1.18 show clearly that the R-chain of Ser3 and Ser4 not only face away from the S2 pocket but in fact are further away from the Smt3 P2 residue in the 1.6 Angstrom Avp2-Smt3 model. Similarly, Glu5 sits closer to the P4 residue in the 1.6 Angstrom Avp2-Smt3 model than Gln6 (Figure 3.1.19).

Figure 3.1.18: Comparison of the P2 Site for the 2.6 and 1.6 Angstrom Resolution Structure of Avp2

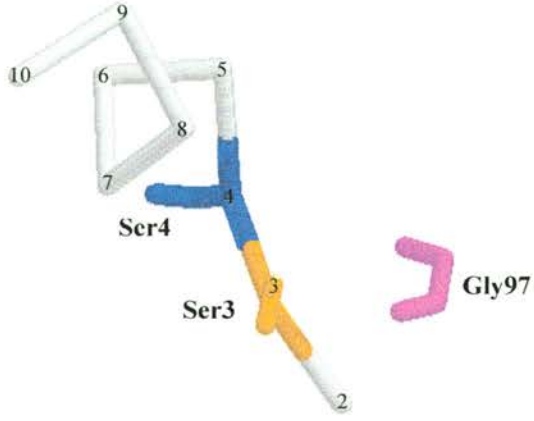
The placement of Ser3 (orange) and Ser4 (dark blue) in the S2 pocket of both the (a)(i) 2.6 and (a)(ii) 1.6 Angstrom structure of Avp2 were compared using the RASMOL Version 2.6 computer package (Sayle & Milner-White., 1995). Further analysis using the (b) Radius function of the Swiss-PDB viewer Computer Programme (Guex and Peitsch., 1997; http://www.expasy.org/spdbv/) was used to measure the distance between the α-Carbon of Ser3 and Ser4 and the α-Carbon of the P2 (magenta) residue.

(a) Avp2-Smt3 Model

(i) 2.6 Angstrom Structure



(ii) 1.6 Angstrom Structure



(b) Table of Distances Between the α-Carbons of Ser3 and Ser4 and the P2 α-Carbon (given to the nearest Angstrom)

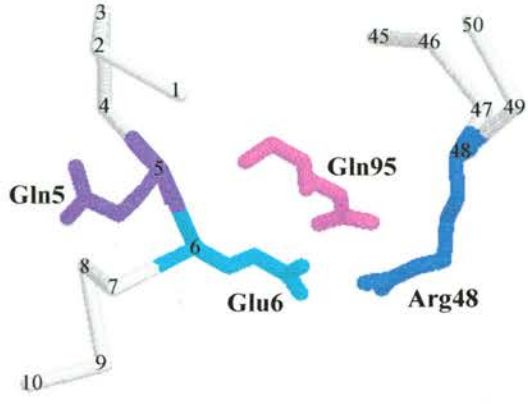
Avp2 Residue	Distance from P2 in 2.6 Angstrom Structure (Angstroms)	Distance from P2 in 1.6 Angstrom Structure (Angstroms)
Ser3	3	5
Ser4	4	7

Figure 3.1.19: Comparison of the P4 Site for the 2.6 and 1.6 Angstrom Resolution Structure of Avp2

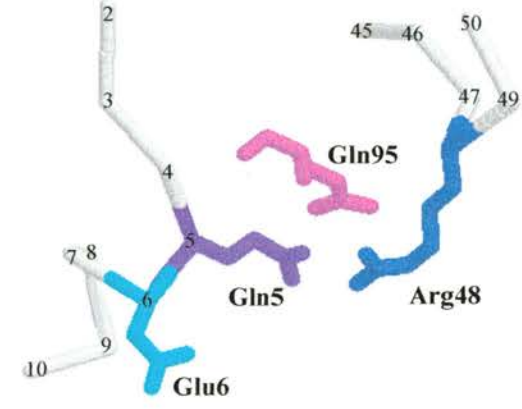
The placement of Glu5 (purple) and Gln6 (cyan) in the S4 pocket of both the (a)(i) 2.6 and (a)(ii) 1.6 Angstrom structure of Avp2 were compared using the RASMOL Version 2.6 computer package (Sayle & Milner-White., 1995). Further analysis using the (b) Radius function of the Swiss-PDB viewer Computer Programme (Guex & Peitsch., 1997; http://www.expasy.org/spdbv/) was used to measure the distance between the α-Carbon of Glu5 and Gln6 and the α-Carbon of the P4 (magenta) residue.

(a) Avp2-Smt3 Model

(i) 2.6 Angstrom Structure



(ii) 1.6 Angstrom Structure



(b) Table of Distances Between the α-Carbons of Ser3 and Ser4 and the P4 α-Carbon (given to the nearest Angstrom)

Avp2 Residue	Distance from P2 in 2.6 Angstrom Structure (Angstroms)	Distance from P2 in 1.6 Angstrom Structure (Angstroms)
Glu5	6	4
Gln6	4	8

3.1.2.6 Discussion

Cleavage of Smt3 and SUMO1 by Avp2

The Avp2-Smt3 model was used together with other bioinformatic techniques to identify putative residues of Avp2 involved in the S2 and S4 substrate specificity pockets. As discussed previously the P2 and P4 residues of Smt3 were a glycine and a glutamine, respectively and as such did not follow the motif, (M,L,I)XGG-X. To analyse this the putative Smt3 substrate was subjected to cleavage by Avp2 *in vitro*. Following this the human homologue of Smt3, SUMO1 was also cleaved by Avp2.

Smt3 and SUMO1 were found to be cleaved at QIGG-A and QTGG-H, respectively by protein sequencing and therefore a new motif was proposed (M,L,I,Q)XGG-X. The cleavage of Smt3 by Avp2 has also been shown by Balakirev et al., 2002. However unlike our work where there were only eleven amino acids following the cleavage site, Balakirev and colleagues used an Smt3-GFP substrate. Nevertheless, in both cases Smt3 was shown to be cleaved very slowly, as was SUMO1 in our analysis.

This suggested two potential theories. Firstly that SUMO proteins are cleaved by Avp2 slowly as a temporal mechanism within host cells infected with adenovirus. If this was found to be the case *in vivo*, and Avp2 could cleave SUMO proteins from their target proteins then this carries with it implications for host cell invasion by adenovirus, as sumoylation of proteins has been found to have several effects at the cellular level, including repression/activation of transcription, modifications to DNA structure, alterations to the sub-nuclear structure, increased oncogenic potential and changes to membrane transport (reviewed in Tatham & Hay, 2003).

The second hypothesis was that these proteins were not suitable substrates for the protease, and it must be remembered that previous studies which investigated glutamine at the P4 site both in synthetic substrates (Webster et al., 1989b) and in human Ad4 pTP

(Webster et al.,1997) showed that although glutamine could be accepted at the P4 site of the motif (M,L,I,Q)XGX-G, cleavage was slow in its presence. To analyse whether cleavage of Smt3 and SUMO1 is slow due to the presence of the glutamine in the P4 site or due to some other structural feature it may be beneficial to mutate the glutamine to either a methionine, leucine or isoleucine residue as a control, as these amino acids are well established P4 site residues.

Nevertheless it must be borne in mind that regardless of whether or not the glutamine is a suitable candidate for the P4 site, the work carried out in this study used the α-Carbon of Smt3 in the Avp2-Smt3 model to analyse which residues of Avp2 sat within a 6 Angstrom radius of the P2 and P4 site and therefore this analysis was unaffected by the R-chain of the P2 and P4 residues. It must also be remembered that this investigation along with other bioinformatic studies were simply used to indicate residues potentially involved in forming the S2 and S4 Avp2 substrate specificity pockets and were by no means a conclusive picture of those residues involved in substrate binding.

Bioinformatic and Mutational Analysis of the S2 and S4 Pockets of Avp2

The complex between Ulp1 and Smt3 in its thiohemiacetal transition state proved that this was the active conformation of Ulp1. Therefore, when Avp2 and Ulp1 were superimposed it became clear that Avp2 had also been crystallised in its active conformation as the catalytic core domain of the two proteases superimposed and Smt3 fitted into the proposed catalytic cleft of Avp2. Further analysis of the Avp2-Smt3 model determined the orientation of the P1, P2, P3 and P4 Smt3 residues in the catalytic cleft of Avp2; P1, P2 and P4 pointed into the protease, whilst P3 pointed away from it (data not shown). This was also proposed by Ding et al., 1996 who had previously superimposed Avp2 and Papain in complex with the synthetic substrate, QAAAA-chloromethyl ketone.

As the P2 and P4 residues were predicted to point towards Avp2 and also because the two major determinants of substrate specificity were the P2 and P4 residues according to the cleavage motifs (Webster *et al*, 1989b), the residues involved in the S2 and S4 substrate specific pockets were investigated using bioinformatics and mutational analysis.

Primarily bioinformatics were employed to indicate residues putatively involved in the S2 and S4 pockets. Those residues predicted to be members of the S2 substrate specificity pocket were Ser3, Ser4, Asn44, Gly52, Val53, His54, Trp55, Cys122 and Gly123, whilst those of the S4 substrate specificity pocket were Met1, Glu5, Gln6, Asn44, Arg48, Gly51, Gly52 and Trp55.

The residues selected for the Alanine screen were Ser3, Ser4, Gly52,Val53 and Trp55 for the S2 pocket and Glu5, Gln6, Arg48, Gly51, Gly52 and Trp55 for the S4 pocket. Unfortunately G52A could not be purified as it was produced in too low concentrations.

Although G52A could not be purified, all other mutants were purified and their specific activities determined. Whilst S4A caused no reduction in activity for the S2 pocket, S3A, V53A and W55A all caused around an 80% reduction in activity, suggesting their importance in this pocket.

The residues implicated in the S4 pocket had varying effects on activity when mutated to an alanine. Q6A produced a small increase in activity, whilst R48A and G51A produced a 13% and 17% reduction in activity, suggesting that these three residues are all relatively unimportant in substrate binding. However, the substitution used for Gly51 of glycine to alanine may not have been a particularly effective as the side chains are very similar, -H and -CH_3, respectively. Unlike these residues, the mutants, E5A and W55A had a dramatic effect on specific activity. Whilst E5A virtually caused a complete loss of activity (4% of the activity of WT), W55A showed a mere 16% of the activity of WT.

Mutation of the equivalent tryptophan residue in other members of the CE clan, Ulp1 and Vaccinia virus protease also had serious consequences. The ULP1 gene in yeast is essential therefore Mossessova & Lima, 2000 proposed that mutations which inhibit the activity of the protease would produce a lethal growth phenotype. To analyse this they produced a *ulp1Δ* yeast strain. They transformed the strain with plasmids containing either WT or mutant alleles of Ulp1 and found that the strain could be rescued by WT Ulp1 but not with a W515A mutant (equivalent to Trp55). Likewise, Byrd et al., 2003 carried out an alignment of the catalytic core of Vaccinia virus I7L protease, ASFV protease, Avp2 and Ulp1 to find that there were seven conserved residues. Each of the seven residues were mutated in Vaccina virus 17L protease to an alanine and their activities were analysed. Of these residues, six showed a complete loss of activity, these were the three members of the catalytic triad, the glutamine involved in forming the oxyanion hole, the glycine C-terminal to the catalytic cysteine and the tryptophan C-terminal to the catalytic histidine (equivalent to Trp55). Together, the [illegible] further support for the importance of Trp55 in substrate specificity and enzyme activity.

Similarly, the deletion mutant, Δ5-13 for Avp2 showed no activity (Anderson, 1990) indicating that this region is important for the activity of the protease. In this ar mutated Glu5 and Gln6 and as the results indicated Glu5 is particularly important as mutation of this residue to an alanine caused an almost complete loss of activity. Nevertheless, this was not consolidated by the orientation of Glu5 in the 2.6 Angstrom structure of Avp2 (Ding et al., 1996) as Glu5 points away from the proposed S4 site. structure instead indicated that Gln6 may play a more important role because together Arg48 it closes the S4 site (Figure 3.1.19(a)).

However, when the 1.6 Angstrom resolution structure of Avp2 (McGrath et al., 2003) was published, the positions of Glu5 and Gln6 had changed dramatically and Glu5 was now proposed to close the S4 site with Arg48 (Figure 3.1.19(b)). The 1.6 Angstrom crystal paper stated that Met1 could not be found in the crystal structure resulting in the re-modelling of the first seven residues. This therefore left the unanswered question of whether the methionine just simply could not be identified in the crystal or whether it was there at all. If the methionine could not be identified then this would indicate that the function of Glu5 is to form an ionic bond with Arg48 to close the groove of the S4 site. This would explain the reduction in activity when the glutamate was mutated to an alanine as this would inhibit the formation of the bond. However, if instead the methionine was lost during expression and/or purification then it must be assumed that the 2.6 Angstrom structure is the more accurate of the two structures and therefore it is the function of Gln6 and Arg48 to close the S4 pocket.

The second theory is more in keeping with the results as Gln6 and Arg48 both caused only a small change in activity which would be consistent with similar functions, whereas Glu5 caused an almost complete loss of activity.

It is therefore proposed that the 2.6 Angstrom resolution structure of Avp2 is more accurate at the N-terminus than the 1.6 Angstrom resolution structure and that Gln6 and Arg48 form a bond to close the groove of the S4 site. It is further proposed that Glu5 plays an as yet unidentified, essential role in the binding of the substrate, which is probably based on charge. To examine this further it would be interesting to mutate the residue to a glutamine which is the uncharged derivative of glutamate, thereby effecting charge but not size.

As discussed, a number of residues were identified as being important in substrate specificity by bioinformatic analysis and most of these residues were investigated by

mutational analysis. However three of the proposed residues were not mutated, Met1, Asn44 and Gly123. In hindsight these residues should also have been part of the Alanine screen as Asn44 and Gly123 were particularly interesting. Asn44 was found within a 5 and a 6 Angstrom radius of the P2 and P4 residues, respectively in the Avp2-Smt3 model, was conserved in both proteases in the Avp2/Ulp1 superimposition and was completely conserved in the Avp alignment. Likewise, Gly123 was found within a 6 Angstrom radius of the P2 residue in the Avp2-Smt3 model, was conserved in both proteases in the Avp2/Ulp1 superimposition, sits next to the catalytic cysteine and was conserved in the alignment of the core domain of representative members of each of the five CE clan families. Further evidence for the importance of Gly123 was produced by Byrd et al., 2003 who analysed the equivalent residue in the CE clan protease of Vacinnia virus I7L, where mutation of this residue to an alanine caused a loss of enzyme activity.

In summary those residues believed to be involved in substrate specificity are Ser3, Gln5, Val53, Asn44, Trp55 and Gly123, of which Asn44 and Gly123 still remain to be examined by mutational analysis.

3.1.3 pVIcT Binding

Avp2 is activated by an 11-residue peptide, GVQSLKRRRCF, called pVIcT, which has been proposed to bind the protease and act as a strap to stabilise the active site (Mangel et al., 1996; Figure 1.3.3; Figure 3.1.20(a)(i)). The aims of this section were to analyse the residues of Avp2 involved in binding the N-terminus of pVIcT and to examine whether Ulp1 requires a pVIcT-like β-strand for activation.

3.1.3.1 Avp2 Residues Involved in Binding the N-terminus of pVIcT

The Gly1' and Val2' N-terminal residues of pVIcT have been suggested to be important in binding and activating the protease (Cabrita et al., 1997; Ruzindana-Umunyana et al., 2000; Baniecki et al., 2001). They are also highly conserved and sit in a small hydrophobic pocket. However the residues of the Avp2 pocket have not been investigated to analyse which of these may be involved in binding Gly1' and Val2', thus bioinformatic and mutational analysis were used in this section to examine these residues.

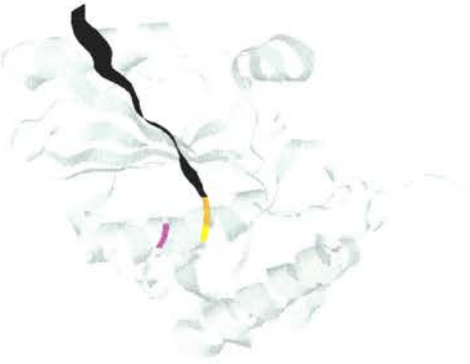
3.1.3.1.1 Bioinformatic Analysis

To investigate which residues of Avp2 may be important in binding the N-terminus of pVIcT, the 2.6 Angstrom resolution Avp2-pVIcT structure was viewed in the RASMOL Version 2.6 computer package (Sayle & Milner-White., 1995). It was apparent from the structure that the pocket contained only one charged residue, Asp142 (Figure 3.1.20(a)). Following this initial analysis, Asp142 was investigated using an alignment of 28 Avp sequences representing the four genera of the *Adenoviridae* family (Method 2.1.1). It was found to be highly conserved, further supporting its importance (Figure 3.1.20(b)).

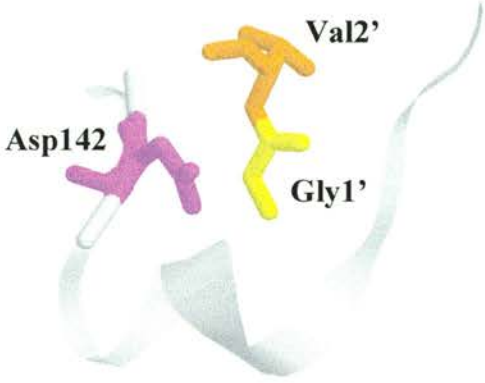
Figure 3.1.20: Bioinformatic Analysis of Avp2 Binding to the N-terminus of pVIcT
(a) Avp2 (grey) in complex with pVIcT (black) was viewed in the RASMOL Version 2.6 computer package (Sayle and Milner-White, 1995). The region of Avp2 (Asp142-magenta) proposed to be involved in binding the N-terminus of pVIcT (Gly1'-yellow; Val2'-orange) was then investigated using a (b) primary sequence alignment of 28 Avp sequences representing the four genera *Adenoviridae*.

(a) N-terminus pVIcT Sitting in the Hydrophobic pocket of Avp

(i) pVIcT Acting as a Strap to Bring the Two Domains of Avp Together



(ii) Expanded view of Hydrophobic Pocket



(b) Avp Primary Sequence Alignment

* Sequence for Canine 2 (strain Toronto A26-61).
** Sequence for Canine 1 (strains Utrecht, CLL and RI261).

```
             132       142       152
             |         |         |
Human 2      AFANWPQTPMDHNPTMNLITG
Human 5      AFANWPQTPMDHNPTMNLITG
Human 40     AFVNWPTSPMERNPTMDLLTG
Human 41     AFIHWPSNPMEQNPTMDLLTG
Human 3      AFVHWPDRPMDGNPTMKLVTG
Human 11     AFVHWPDRPMDGNPTMKLLTG
Human 12     AFTHWPDHPMDKNPTMDLLTG
Human 17     AFVHWPDRPMDGNPTMNLLTG
Canine 2*    AFVNWPDDPFDHNPTMGPLKS
Canine 1**   AFVNWPDNPFNHNPTMGPLKS
Bovine 3     AFVRWPLRAMDNNPTMNLIHG
Human 4      ACA-WP-TPMDKNPTMNLLTG
Bovine 2     AFVNWPDRPMTRNPTMDLLTG
Porcine 5    AFVNWPHTPMDRNPTMDLLTG
Bovine 10    AFANWPAHPFD-NPTMDQLVG
Mouse 1      AFGKYPQNPMNNNPIMGPIEG
Porcine 3    AFARYPDSPMAYNPVMDLVEG
Equine 2     SFACNPARPRNGNPIIDIVRG
Fowl 1       SFDRYRSKPMDGNPVIDTVVG
Turkey 3     SFYYFRNSPMYNNPIIDVVTG
Fowl 8       SFDRYHTRPMDGNPIIDTVVG
Snake 1      CAHLSPSNPFG-TSLFQSLDG
Frog 1       SFYHYPTFPMRGNPIIDLVDG
Possum 1     CFNKYKCDPFN-NPLFQRFQG
Bovine 4     CFYKYRGKAFN-NELFQSLNG
Bovine 7     CFYKYKSNAFK-NCLFQSLYG
Duck 1       CFNLCHINPFE-ASIFQAMHG
Ovine 7      CFHFYKQNVFK-SWLFQKLNG
             .             :  . .
```

3.1.3.1.2 Cloning of D142AET11c

As Asp142 of Avp2 was believed to be important in binding the N-terminus of pVIcT, an Alanine mutant was made at this site and the mutant was cloned into the pET11c vector. The PCR for the internal mutant, D142A required a two step PCR and three primers. A mini-prep (Method 2.2.7) was made of WT Avp2pET11c vector and this was used as a template for both PCR reactions (Method 2.2.1.3).

The first PCR required the mutant internal forward primer and the WT reverse primer in Table 3.1.8 to PCR a reverse mega-primer fragment for D142A. The PCR reaction was carried out at 94°C for 5min (hot start), 94°C for 30 sec, 55°C for 60sec, 72°C for 45sec, 72°C for 7min over 25 cycles.

The reverse mega-primer was separated on a 0.9% (w/v) Agarose gel (Method 2.2.2) and then gel extracted (Method 2.2.3). The mega-primer was used with the WT forward primer in Table 3.1.8 in a second PCR reaction. The PCR reaction was carried out at 94°C for 5min (hot start), 94°C for 30 sec, 55°C for 60sec, 72°C for 60sec, 72°C for 7min over 25 cycles.

The vector and insert were then digested with NdeI and BamHI (Method 2.2.4), ligated (Method 2.2.5), used to transform the JM109 cell line (Promega) (Method 2.2.6) and left to grow on an LB-agar/Carbenicillin plate at 37°C for 12-16hr (Method 2.2.10.1). Six colonies were selected from the transformants and a mini-prep was carried out on each. 10µl of each mini-prep was double digested by XbaI (restriction site of the vector) and BamHI and then separated by Agarose gel electrophoresis (Method 2.2.2). Finally one of the colonies which appeared to be positive according to the double digestion was identified by DNA sequencing (Method 2.2.9).

After confirmation by DNA sequencing, a glycerol stock (Method 2.2.10.2) was made of the JM109 cell line. The vector was then transformed into the Rosetta(DE3) cell line (Novagen) and a glycerol stock was made of these.

Table 3.1.8: Primers Used to Clone D142A into the pET11c Vector
The PCR reaction for D142A used a mutant forward internal primer, a WT forward primer and a WT reverse primer. The mutation in the WT sequence to produce the mutant is highlighted in red.

Primer	Oligonucleotide primer sequence (5'-3')
Forward	GTAATTACATATGGGCTCCAGTGAGCAGGAACTG
D142A forward internal	CAAACTCCCATGGCTCACAACCCCACC
Reverse	GCGGGATCCTTACATGTTTTTCAAGTGACAAAAAGA

3.1.3.1.3 Determination of Specific Activity

The Specific activity of D142A was calculated in the same manner as is described in Section 3.1.2.4 and the Avp concentration used was 0.1902μM over a time course of 5min, 10min, 15min, 30min. The mean Specific activity +/- S.D was found to be 6.4 +/- 0.32mol substrate transformed/min/mol enzyme, which was 57% of the activity of WT protease.

3.1.3.2 Ulp1 Requirement for pVIcT-like β-strand

The 2.6 Angstrom structure of the Avp2-pVIcT complex was superimposed onto the 1.8 Angstrom structure of the Ulp1-Smt3 complex (Mossessova & Lima, 2000) as described in Section 3.1.2.1.1.

The superimposition viewed in the RASMOL Version 2.6 computer package (Sayle & Milner-White., 1995) revealed that a portion of the Avp2 β-strand (residues 103-109)

involved in binding pVIcT superimposed onto a Ulp1 β-strand (residues 562-568), this is highlighted by the table in Figure 3.1.5.

3.1.3.3 Discussion

In this section, an initial investigation of an Avp2 residue involved in binding the N-terminus of pVIcT was made. The N-terminal residues, Gly1' and Val2' of pVIcT sit in a hydrophobic pocket of Avp2, which contains only one polar residue, Asp142, therefore Asp142 was mutated to an alanine and its specific activity was determined to be 57% of WT. This is a substantial reduction in activity caused by a single point mutation, thereby suggesting the importance of this residue in binding the activating peptide. However it must be remembered that the protease interacts with activating peptide through at least two other residues, Cys10' (Grierson et al., 1994; Jones et al., 1996; McGrath et al., 2002) and Phe11' (Baniecki et al, 2001) and so mutation at one site in the N-terminal hydrophobic pocket would not be expected to cause a large reduction in activity.

Furthermore, it may be interesting to investigate the other residues of this pocket, residues 141-152. From the 1.6 Angstrom crystal structure of Avp2 in complex with pVIcT, a number of the hydrophobic pocket residues were suggested to form hydrogen bonds with the Gly1' and Val2' of pVIcT, including 141, 142, 147, 150 and 152 and are therefore worthy of investigation.

Likewise, it may also be interesting to investigate whether Ulp1 requires a pVIcT-like activator because bioinformatic studies showed that a portion of the Avp2 β-strand involved in binding pVIcT superimposed onto a portion of Ulp1 β-strand.

3.1.4 C-terminal Tail

The aim of this section was to investigate whether the C-terminal tail, residues 161-204 of Avp2 had a function in proteolytic activity. This was investigated using bioinformatic and mutational analysis.

3.1.4.1 Bioinformatic Analysis

Analysis of twenty-eight Avp sequences showed that the C-terminal tail contained eight conserved residues, four of which were completely conserved (Figure 3.1.21(a)) When the 2.6 Angstrom crystal structure of Avp2 (Ding et al., 1996) was analysed, it appeared that the C-terminal tail of Avp2 did not have a function in either binding pVIcT or substrate binding (data not shown).

The tail was viewed in the RASMOL Version 2.6 computer package (Sayle & Milner-White., 1995) and the polar residues (Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Lys, Ser, Thr Tyr) and hydrophobic residues (Ala, Leu, Ile, Pro, Phe, Met, Trp, Val) determined by the package were highlighted in red and blue, respectively (Figure 3.1.21(b)). As expected, the hydrophobic residues faced the interior of the protease, whereas the polar residues pointed away from the structure. Three hydrophobic and five polar residues were conserved (Figure 3.1.21(a)).

Figure 3.1.21: C-terminus of Avp2
(a) 28 Avp sequences were aligned and the polar and hydrophobic residues were highlighted in red and blue, respectively. The (a) front and (b) side view of the C-terminus of Avp2 were then visualised in the RASMOL Version 2.6 computer package (Sayle & Milner-White., 1995).

(a) Primary Sequence Alignment of the C-terminus of 28 Avp Sequences
* Sequence for Canine 2 (strain Toronto A26-61).
** Sequence for Canine 1 (strains Utrecht, CLL and RI261).

```
                161                                             204
                |                                               |
Human 2         SPQVQPTLRRNQEQLYSFLERHSPYFRSHSAQIRSATSFCHLKNM-----------
Human 5         SPQVQPTLRRNQEQLYSFLERHSPYFRSHSAQIRSATSFCHLKNM-----------
Human 40        SPQVVPTLRHNQERLYRFLAQRSPYFQRHCERIKKATAFDQMKNNM----------
Human 41        SPQVEPTLRRNQERLYRFLTQHSPYFRRHRERIEKATAFDQMKNAQVLFHNKIFY-
Human 3         SPQVQPTLRRNQEVLYRFLNTHSSYFRSHRARIERATAFDRMDMQ-----------
Human 11        SPKVQPTLCDNQKALYHFLNTHSPYFRSHRTHIERATAFDRMDVQ-----------
Human 12        SPQVVGTLQRNQNELYKFLNNLSPYFRHNRERIEKATSFTKMQNGLK---------
Human 17        SPQVLPTLRRNQEELYRFLARHSPYFRSHRAAIEHATAFDKMKQLRVSQ-------
Canine 2*       DPTVQYVLWGNQEKLYKFLEKHSAYFRAHAAAIKARTAFNKLKQ------------
Canine 1**      DPTVQHVLWENQEKLYKFLEKNSAYFRAHAAAIKTRTAFNKLKQ------------
Bovine 3        SPSSQNVFLRNQQNLYRFLRRHSPHFVKHAAQIEADTAFDKMLTN-----------
Human 4         SPQVEPTLRRNQEALYRFLNSHSAYFRSHRARIEKATAFDRMNQDM----------
Bovine 2        KPSSLAILRENQNQLYKFLSTHSQYFRTHRPQIERDTSFNKLLELKNQ--------
Porcine 5       LPKCQDILRQNQANLYRFLAHHSSYFRSHRPQIERNTAFDKMIQESQ---------
Bovine 10       APRAQSIFKQNQETLYSFLHYNSSFFRRYENKLRKQTDP-----------------
Mouse 1         NPCYTKTLYRNQQWVYSYLNKNSLYFRLHVELIKKNTAFDKLLVRK----------
Porcine 3       DADVQPIFRANQEACYAFLARHSAYFRAHRHAIMEQTHLHKALDMQ----------
Equine 2        DPSSLPILYRNQENMYAFLENNSPYFVSHEREIKRKTAFDYIQ------------
Fowl 1          SPPYRDILHRNQERTYYWWTKNSAYFRAHQEELRRETALNALPENHV---------
Turkey 3        SSYGIAILHCNQERLYNWLYYNSVYFRDNELEIKRNTRINSILVHYLFIVLFLFAR
Fowl 8          KPEFQSILHRNQERMYFWFMKNNSFFRAHESELKRETAINSVPENH----------
Snake 1         PRSPEALHKNQQI-LYNFLSSKCAYFRKNARNIVMNTRLHLIKTH-----------
Frog 1          SSYGIYLTHCNQKKLIAWLLSNSAYFRKNAMLMIHNTRL-----YYLYTHL-----
Possum 1        KPASPALLHENQNILYNFLNSKSDYFNKNQHQFVTETKIGLIKTHN----------
Bovine 4        TPSDPSSLHKNQDILYDFFICKSSYFRHNKKMLISNTKLGLIKSH-----------
Bovine 7        TPPNPTNLHKNQDFLYKFFKEKSLYFRQNEEYIVSNTKIGLIKSHI----------
Duck 1          YPSKPHLLHANQQMLYDFLRSHSSYFVNNERTLVCNTKLNLINIHQ----------
Ovine 7         IPCEPHLLHENQTFLYDFLNAKSVYFRKNYRTFIENTKTGLIKTH-----------
                 .       **     :      .*      :   *
```

(b) C-terminus of Avp2

(i) Front View

(ii) Side View

3.1.4.2 Expression

There are three secondary structures in the C-terminus, both 161-181 and 182-195 form an α-helix, whilst 196-204 forms a piece of random coil. To analyse the function of each of these regions three deletion mutants were made, Δ161-204, Δ182-204, Δ196-204 by Dr S. Jones (Figure 3.1.22).

The mutants were cloned into the pET11c vector (Novagen) at the XbaI and BamHI restriction sites and were then transformed into the Rosetta(DE3) cell line. The cell line was then grown in LB/Carbenicillin/Chloramphenicol according to Method 2.5.2. They were induced with 0.6mM IPTG for 4hr at 37°C (Method 2.5.2). Following this, they were re-suspended in 8ml of Suspension buffer (50mM Tris, 5mM EDTA, 4% (v/v) Glycerol). However they could not be purified according to Method 2.5.4 due to low yields.

Figure 3.1.22: C-terminal Deletion Mutants for Avp2
The C-terminus of Avp2 was visualised in the RASMOL Version 2.6 computer package (Sayle & Milner-White., 1995). The C-terminus was composed of two α-helices and a section of random coil. The first and second α-helices were formed by residues 161-181 (highlighted magenta) and 182-195 (highlighted purple), respectively, whilst the random coil was formed by residues 196-204 (highlighted blue). The deletion mutants represented the removal of each of these secondary structures - Δ161-204, Δ182-204, Δ196-204.

3.1.4.3 Discussion

The C-terminal tail of Avp2 does not appear to have a function in respect of binding the activating peptide or the substrate. However, eight residues were found to be conserved within this region, four of which were completely conserved therefore suggesting that the C-terminal tail has an as yet unidentified function. As expected, the tail was amphipathic, with hydrophobic residues facing towards the protease and polar residues facing away from it.

The C-terminus is composed of three secondary structures, two α-helices formed by 161-181 and 182-195 and a random coil formed by 196-204. Therefore to analyse each of these regions function in proteolytic activity three deletion mutants were made, Δ161-204, Δ182-204, Δ196-204. The theoretical pI for the Δ161-204, Δ182-204, Δ196-204 mutants were 8.18, 8.49 and 8.88, respectively, compared to the theoretical pI of WT Avp2 which is 8.95. Consequently, as the purification method for WT Avp2 is dependant upon Ion Exchange Chromatography, purification of the C-terminal mutants meant the development of a new purification method for each protease. A number of attempts were made to purify the mutants but their expression levels were particularly low, thus making purification difficult.

Hence, the question of whether the C-terminus of Avp2 is required for activity remains unanswered. Therefore to analyse this, it may be advantageous in the future to mutate each of the eight residues conserved within the C-terminus and analyse their effect on activity. Alternatively, it was hoped that if a tagged system was developed in Sub-chapter 3.2 to purify WT Avp2 then this system could be used to purify the C-terminal deletion mutants.

3.1.5 EDS Protease

Egg Drop Syndrome (EDS) virus is a member of the Atadenovirus genera of the *Adenoviridae* family, which infects hens, causing the production of soft-shelled and shell-less eggs. As EDS and human Ad2 are members of the Atadenovirus and Mastadenovirus genera, respectively it would be interesting to compare the structures of their proteases.

To compare the differences/similarities between Egg Drop Syndrome Protease (EDSp) and Avp2, the EDSp would have to be produced in large enough quantities for crystal trials. Therefore, the aim of this section was to clone EDSp into the pET11c vector (Novagen) to produce purified, recombinant EDSp. As there was already a purification method developed for Avp2 it was decided to use this as a basis for the purification of EDSp.

3.1.5.1 Cloning of EDSET11c

A genomic fragment of EDS DNA coding for the protease cloned into the Bluescript vector (a gift from Dr. D. Todd) was used as a template for the PCR reaction (Method 2.2.1.1). The forward and reverse primers in Table 3.1.9 were used to PCR the EDSp fragment. The PCR reaction was carried out at 94°C for 5min (hot start), 94°C for 30 sec, 50°C for 30sec, 72°C for 45sec, 72°C for 7min over 25 cycles.

The vector and insert were then digested with NdeI and BamHI (Method 2.2.4), ligated into pET11c (Method 2.2.5), used to transform the JM109 cell line (Promega) (Method 2.2.6) and left to grow on an LB-agar/Carbenicillin plate (Method 2.2.10.1) at 37°C for 12-16hr . Six colonies were selected from the transformants and a mini-prep was carried out on each. 10μl of each mini-prep was double digested by XbaI (restriction site of the vector) and BamHI and then separated by 0.9% (w/v) Agarose gel electrophoresis

(Method 2.2.2). Finally one of the colonies which appeared to be positive according to the double digestion was identified by DNA sequencing (Method 2.2.9).

After confirmation by DNA sequencing, a glycerol stock (Method 2.2.10.2) was made of the JM109 cell line. The vector was then transformed into the Rosetta(DE3) cell line (Novagen) and a glycerol stock was made of these.

Table 3.1.9: Primers Used to Clone EDS into the pET11c Vector
The PCR reaction for EDSp used a forward and reverse primer.

Primer	Oligonucleotide primer sequence (5'-3')
Forward	GTAATTACATATGAGCGGCACATCCGAATCA
Reverse	GCGGGATCCTTATTGATGAATGTTTATTAA

3.1.5.2 Purification

The purification of native Avp2ET11c is dependent upon three Ion Exchange columns, Diethylaminoethyl (DEAE)-Sepharose, Heparin-Sepharose (HS) and Carboxymethyl (CM)-Sepharose. The theoretical pI of Avp2 is 8.95 thus at pH8 it is positively charged and passes through the DEAE-sepharose anion exchange column, it also then passes through the cation exchange column, HS and then finally binds the CM-Sepharose cation exchange column only to be eluted in the presence of NaCl. Therefore, as the theoretical isoelectric point of EDSp is 9.06 this method was used to purify it.

EDSET11c expressed in the Rosetta(DE3) cell line was grown according to Method 2.5.2. The cells were induced with 0.6mM IPTG for 4hr at 37°C (Method 2.5.2). Following this, they were re-suspended in 8ml of Suspension buffer (50mM Tris, 5mM EDTA, 4% (v/v) Glycerol) and purified (Method 2.5.4). However following purification the protease could not be identified in the eluted fractions.

3.1.5.3 Discussion

To analyse differences/similarities between the structures and the mechanisms of activation of EDSp and Avp2, the EDSp would have to be produced in large enough quantities for crystal trials, thus, EDSp was cloned into the pET11c vector and the purification method for AvpET11c was used to attempt to purify the protease. However EDSp was expressed in low yields and could not be visualised in the cell extract using an SDS-PAGE mini-gel. Following purification, the protease again could not be identified using an SDS-PAGE mini-gel. In the absence of an appropriate antibody, the lack of purified protease could not be verified as being caused by low yields of the protease or the purification method being unsuitable for EDSp. However it may have been useful to analyse the activity of the protease using the crude lysate.

The problem of low yields had also been found for several mutant Avp2 proteins, therefore tagged vectors systems were employed to improve soluble expression of Avp2 in the next Sub-chapter. It was hoped that if a purification method could be developed for tagged Avp2 then this could also be used to purify EDSp.

3.2 Cloning, Expression and Purification of Tagged Avp

Introduction

In the previous Sub-chapter, mutant Avp proteins were purified using native expression in the pET11c vector. However several of these proteins could not be purified due to low protein expression, therefore their specific activities could not be calculated. This was also the case for the duck adenovirus protease, EDSp. In addition although many of the Avp2 proteins could be expressed and purified, further kinetic work could not be carried out due to low yields.

The aim of this Sub-chapter was to express and purify large quantities of Avp2 to carry out kinetic studies on the mutant Avp2 proteins and to produce crystals of WT Avp2 in the absence of pVIcT. Moreover it was hoped that if a purification method could be set in place for Avp2 then this could be used as a basis to purify EDSp, so that its specific activity could be determined and crystal trials could be carried out.

It was therefore decided to express WT Avp2 in tagged vector systems. Several systems were investigated and these are discussed in the following results section.

Results and Discussion

3.2.1 Cloning, Expression and Purification of Tagged Avp2

3.2.1.1 Avp2ET28a

Avp2 was first cloned into the pET28a His tag vector (Novagen) because the His tag system would allow a simple purification. Avp2 was cloned into the vector and soluble expression was analysed using SDS-PAGE mini-gels. Following this analysis, the Avp2pET28a vector was co-expressed with the GroELGroESpET29c vector which expressed the chaperone proteins GroEL and GroES.

3.2.1.1.1 Cloning and Expression

A mini-prep (Method 2.2.7) was made of WT Avp2pET11c vector and this was used as a template in the PCR reaction (Method 2.2.1.1). The forward and reverse primers in Table 3.2.1 were used in the PCR. The PCR reaction was carried out at 94°C for 5min (hot start), 94°C for 30sec, 50°C for 30sec, 72°C for 45sec, 72°C for 7min over 25 cycles.

Table 3.2.1 : Primers Used to Clone Avp2 into the pET28a Vector
The PCR reaction for cloning Avp2 into the pET28a vector used a forward and reverse primer.

Primer	Oligonucleotide primer sequence (5'-3')
Forward	GTAATTACATATGGGCTCCAGTGAGCAGGAACTG
Reverse	GCGGGATCCTTACATGTTTTTCAAGTGACAAAAAGA

The vector and insert were then digested with NdeI and BamHI (Method 2.2.4), ligated (Method 2.2.5), used to transform the JM109 cell line (Promega) (Method 2.2.6) and left to grow on an LB-agar/Kanamycin plate at 37°C for 12-16hr (Method 2.2.10.1). Six colonies were selected from the transformants and a mini-prep was carried out on each.

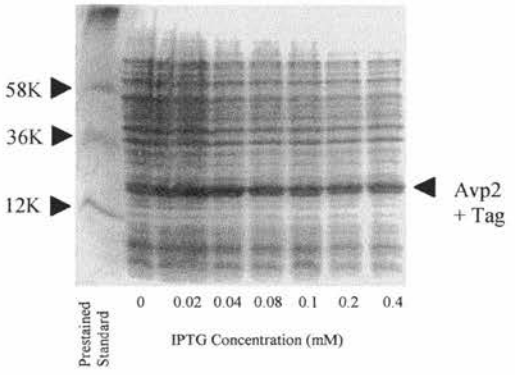
10µl of each mini-prep was double digested by XbaI (restriction site of the vector) and BamHI and then separated by Agarose gel electrophoresis (Method 2.2.2). Finally one of the colonies which appeared to be positive according to the double digestion was identified by DNA sequencing (Method 2.2.9).

After confirmation by DNA sequencing, a glycerol stock (Method 2.2.10.2) was made of the JM109 cell line. The vector was then transformed into the Rosetta(DE3) cell line (Novagen) and a glycerol stock was made of these.

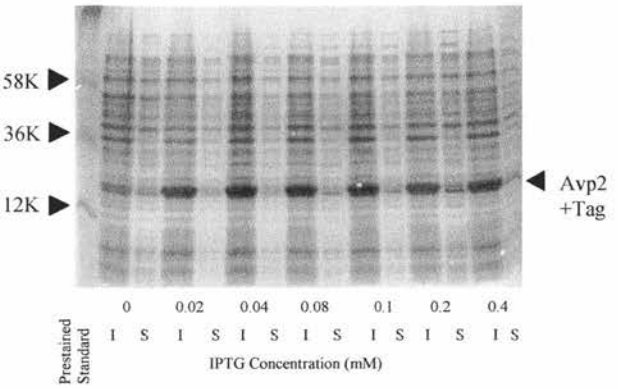
An LB-agar/Kanamycin/Chloramphenicol plate was made from the glycerol stock and a single colony was selected to analyse the Overall/Soluble expression of Avp2 (Method 2.4.4.1). The cells were grown in LB/Kanamycin/Chloramphicol and then re-suspended in 20mM Tris-HCl, 10mM Imidazole, pH8 according to Method 2.4.4.1. Unfortunately, the results of the expression and solubility test (Figure 3.2.1) showed that although the Avp2 was expressed very well in this vector, it was mainly found in the insoluble fraction.

Figure 3.2.1 : Expression of Avp2ET28a
WT Avp2 was cloned into the pET28a His tag vector and expressed in the Rosetta(DE3) cell line. 3ml of cells grown to log phase were induced with 0, 0.02, 0.04, 0.08, 0.1, 0.2 and 0.4mM IPTG for 6hrs at 25°C. Two 1ml fractions of the cells were removed and centrifuged to form a cell pellet. One pellet was loaded directly onto a 15% SDS-PAGE mini-gel, which was stained with Coomassie Blue, whilst the other pellet was sonicated before loading. These fractions were then used to analyse the (a) Overall and (b) Insoluble(I)/Soluble(S) expression levels respectively.

(a) Overall Expression



(b) Insoluble/ Soluble Expression



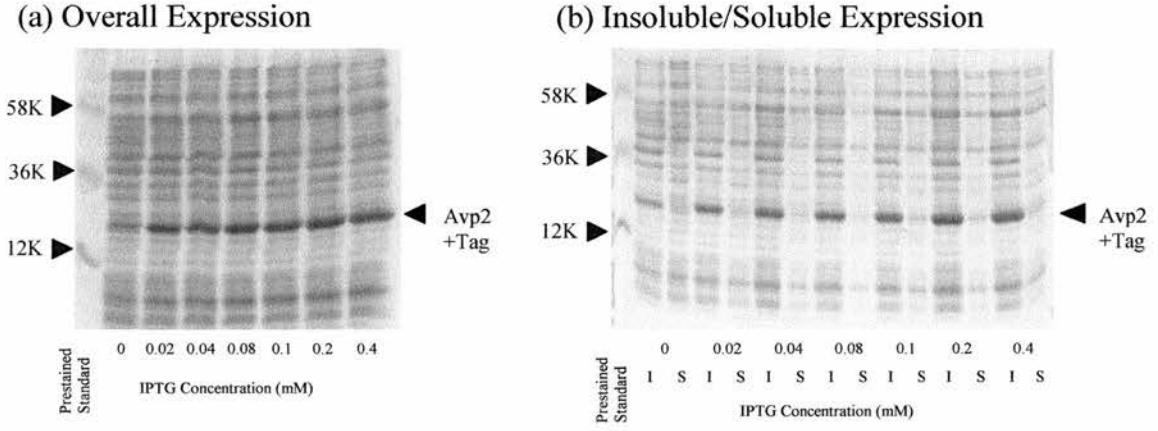
3.2.1.1.2 Co-expressing Avp2ET28a and GroELGroESET29c

As expression of Avp2ET28a produced mainly insoluble protein (Section 3.2.1.1.1), the Avp2pET28a vector was co-expressed with the GroELGropESET29c vector. It was hoped that this co-expression in the Rosetta(DE3) cell line would cause the over-production of the GroEL and GroES chaperone proteins which would in turn allow the refolding of the insoluble His tagged Avp2.

The procedure required the Rosetta(DE3) (Novagen) cell line already expressing Avp2pET28a to be first made competent and then re-transformed (Method 2.2.6) with the GroELGroESpET29c vector. Transformed cells were grown on the selective media LB-agar/Kanamycin/Chloramphenicol/Spectinomycin plate (Method 2.2.10.1) at 37°C for 12-16hr. From this plate a single colony was selected to analyse the Overall/Soluble expression of Avp2 (Method 2.4.4.1). The cells were grown in

LB/Kanamycin/Chloramphicol/Spectinomycin and then re-suspended in 20mM Tris-HCl, 10mM Imidazole, pH8 according to Method 2.4.4.1. Unfortunately, the results of the expression and solubility test (Figure 3.2.2) showed that although the Avp2 was expressed very well in this system, the production of the chaperone proteins had not led to Avp2 refolding as it was still found mainly in the insoluble fraction.

Figure 3.2.2: Co-expression of Avp2ET28a and GroELGroESET29c
Avp2ET28a was co-expressed in the Rosetta(DE3) cell line along with the GroELGroESET29c. 3ml of cells grown to log phase were induced with 0, 0.02, 0.04, 0.08, 0.1, 0.2 and 0.4mM IPTG for 6hrs at 25°C. Two 1ml fractions of the cells were removed and centrifuged to form a cell pellet. One pellet was loaded directly onto a 15% SDS-PAGE mini-gel which was stained with Coomassie Blue, whilst the other pellet was sonicated before loading. These fractions were then used to analyse the (a) overall and (b) insoluble(I)/soluble(S) expression levels respectively.



3.2.1.2 Avp2EHISTEV

Avp2 was cloned into the His tag vector pEHISTEV (a gift from Dr. H. Lui; Appendix I.A). The vector was a modified pET30c (Novagen) vector which had had its Multiple Cloning Site (MCS) and Enterokinase cleavage site removed and replaced with a new MCS and the Tev cleavage site, respectively. The reason for trying another His tag vector was that this vector had previously been used with the pTP protein of human Ad2 and the protein was found to be expressed in a soluble form.

3.2.1.2.1 Cloning and Expression

A mini-prep (Method 2.2.7) was made of WT Avp2pET11c vector and this was used as a template for the PCR reaction (Method 2.2.1.1). The forward and reverse primers in Table 3.2.2 were used in the PCR reaction. The PCR reaction was carried out at 94°C for 5min (hot start), 94°C for 30sec, 50°C for 30sec, 72°C for 45sec, 72°C for 7min over 25 cycles.

Table 3.2.2: Primers Used to Clone Avp2 into the pEHISTEV, pMALc2E, pLou2 and pLou3 Vectors

The PCR reaction for cloning Avp2 into the pEHISTEV, pMALc2E, pLou2 and pLou3 vectors used a forward and reverse primer.

Primer	Oligonucleotide primer sequence (5'-3')
Forward	ATAGGATCCATGGGCTCCAGTGAGCAGGAACTG
Reverse	GCGGCGAAGCTTTTATTACATGTTTTTCAAGTGACA

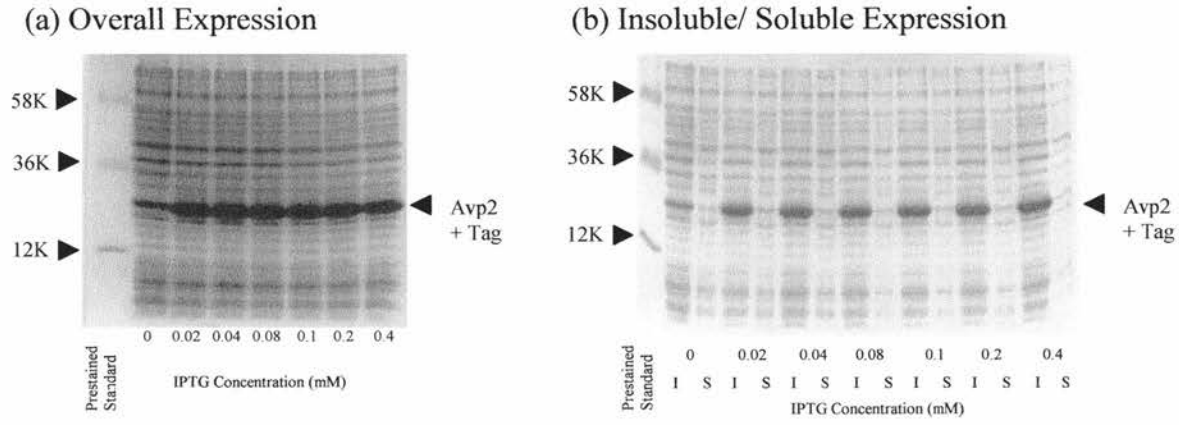
The vector and insert were then double digested by BamHI and HindIII (Method 2.2.4), ligated (Method 2.2.5), used to transform the JM109 cell line (Promega) (Method 2.2.6) and left to grow on an LB-agar/Kanamycin plate (Method 2.2.10.1) at 37°C for 12-16hr . Six colonies were selected from the transformants and a mini-prep was carried out

on each. 10µl of each mini-prep was double digested by BamHI and HindIII and then separated by Agarose gel electrophoresis (Method 2.2.2). Finally one of the colonies which appeared to be positive according to the double digestion was identified by DNA sequencing (Method 2.2.9).

After confirmation by DNA sequencing, a glycerol stock (Method 2.2.10.2) was made of the JM109 cell line. The vector was then transformed into the Rosetta(DE3) cell line (Novagen) and a glycerol stock was made of these.

An LB-agar/Kanamycin/Chloramphenicol plate was made from the Rosetta(DE3) glycerol stock and a single colony was selected to analyse the Overall/Soluble expression of Avp2 (Method 2.4.4.1). The cells were grown in LB/Kanamycin/Chloramphenicol and then re-suspended in 20mM Tris-HCl, 10mM Imidazole, pH8 according to method 2.4.4.1. Unfortunately, the results of the expression and solubility test (Figure 3.2.3) showed that although the Avp2 was expressed very well in this vector, it was mainly found in the insoluble fraction.

Figure 3.2.3: Expression of Avp2EHISTEV
WT Avp2 was cloned into the pEHISTEV vector and expressed in the Rosetta(DE3) cell line. 3ml of cells grown to log phase were induced with 0, 0.02, 0.04, 0.08, 0.1, 0.2 and 0.4mM IPTG for 6hrs at 25°C. Two 1ml fractions of the cells were removed and centrifuged to form a cell pellet. One pellet was loaded directly onto a 15% SDS-PAGE mini-gel, which was stained with Coomassie Blue whilst the other pellet was sonicated before loading. These pellets were then used to analyse the (a) overall and (b) insoluble(I)/soluble(S) expression levels respectively.



3.2.1.2.2 Analysing Binding to Nickel Beads

WT Avp2pEHISTEV transformed into Rosetta(DE3) cells produced mainly insoluble Avp2 (Section 3.2.1.2.1). However, it has been found that proteins expressed with His tags may produce enough soluble material to be purified using a Nickel Column, as Nickel beads can facilitate the purification of small amounts of protein.

Therefore prior to large scale purification, a small scale experiment was carried out to investigate whether a proportion of the expressed protein was soluble and could bind Nickel beads (Method 2.5.1).

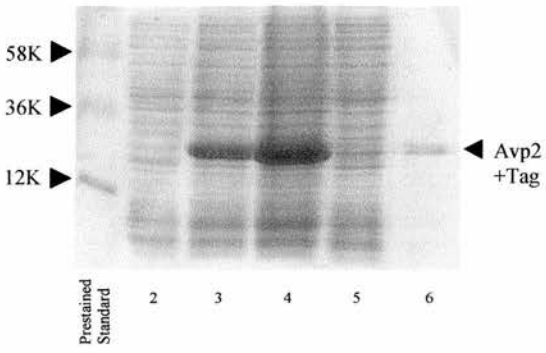
5ml of culture was grown at 25°C for 12-16hr in LB/Kanamycin/Chloramphenicol, then induced with 0.4mM IPTG at 25°C for 6hr and the buffer used for re-suspension was 20mM Tris-HCl, 10mM Imidazole, pH8. Samples taken were pre-induction cell pellet; post-induction cell pellet; post-sonication cell pellet (insoluble fraction); post-sonication

supernatant (soluble fraction) and the protein bound to Nickel beads (washed with Wash buffer (20mM Tris-HCl, 30mM Imidazole, pH8)). These samples were separated on a 15% SDS-PAGE mini-gel (Method 2.4.1.1) which was stained with Coomassie Blue (Method 2.4.2.1) and the results are shown in Figure 3.2.4(a).

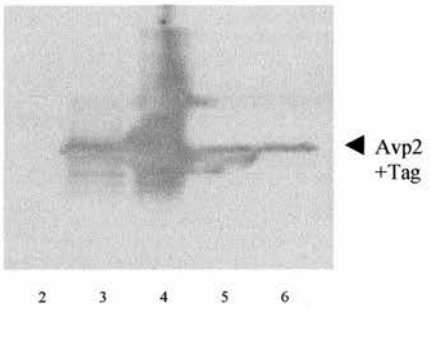
The results in Figure 3.2.4 showed that the expressed protein was found mainly in the cell pellet after sonication (Lane 4) but there was some protein capable of binding the Nickel beads (Lane 6). The band was confirmed to be Avp2 by Western blotting with the R11 primary antibody (Method 2.4.3) (Figure 3.2.4(b)).

Figure 3.2.4: Binding of Avp2EHISTEV to Nickel Beads
Avp2 cloned into the pEHISTEV vector and expressed in the Rosetta(DE3) cell line produced mainly insoluble protein. Therefore to analyse its ability to bind Nickel beads, 5ml of culture was grown at 25°C for 12-16hr and then induced with 0.4mM IPTG at 25°C for 6hr. Samples taken were, pre-induction cell pellet **(Lane 2)**; post-induction cell pellet **(Lane 3)**; post-sonication cell pellet (insoluble fraction) **(Lane 4)**; post-sonication supernatent (soluble fraction) **(Lane 5)** and the protein bound to the Nickel beads **(Lane 6)**. These samples were separated on a 15% SDS-PAGE mini-gel which was (a) stained with Coomassie Blue and (b) Western blotted using the R11 primary antibody.

(a) 15% SDS-PAGE mini-gel



(b) Western blot



However, it was decided not to continue with this method of purification as it produced a low yield of protease.

3.2.1.3 Avp2MALc2E

Avp2 was cloned into the Maltose tag vector pMALc2E (NEB). This tag was chosen as the use of His tags had lead to the expression of mainly insoluble protein (Section 3.2.1.1 and 3.2.1.2) and it was thought that the large N-terminal Maltose tag may improve soluble expression.

3.2.1.3.1 Cloning and Expression

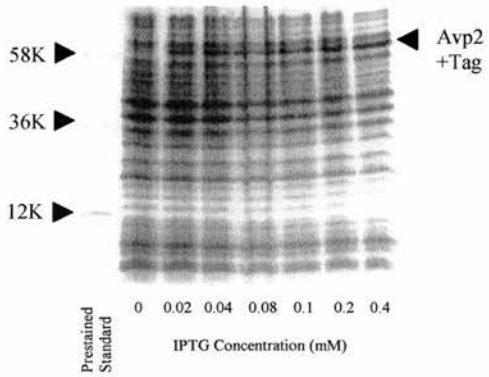
Avp2 was cloned in to the pMALc2E vector according to the method used for Avp2pEHISTEV (Section 3.2.1.2.1), except the PCR reaction was carried out at 94°C for 5min (hot start), 94°C for 30sec, 55°C for 1min, 72°C for 45sec, 72°C for 7min over 25 cycles and the antibiotic used was Carbenicillin.

After confirmation by DNA sequencing, a glycerol stock (Method 2.2.10.2) was made of the JM109 cell line. The vector was then transformed into the Rosetta(DE3) cell line (Novagen) and a glycerol stock was made of these.

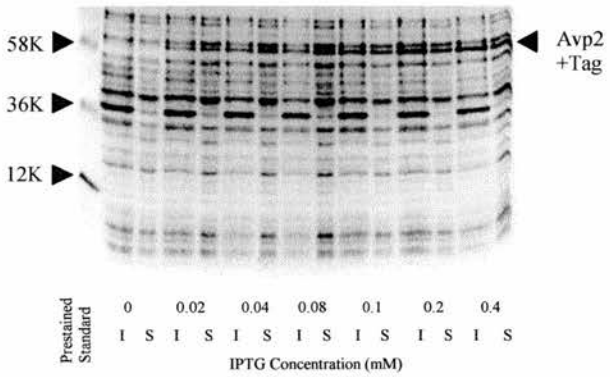
An LB-agar/Carbenicillin/Chloramphenicol plate (Method 2.2.10.1) was made from the glycerol stock and a single colony was selected to analyse the Overall/Soluble expression of Avp2 (Method 2.4.4.1). The cells were grown in LB/Carbenicillin/Chloramphenicol, 0.2% (v/v) Glucose and then re-suspended in Column buffer (20mM Tris-HCl, 200mM NaCl, 1mM EDTA, pH7.4) according to Method 2.4.4.1. The results of the expression and solubility test (Figure 3.2.5) showed that Avp2 was expressed in this vector and was found in both the soluble and insoluble fractions.

Figure 3.2.5: Expression of Avp2MALc2E
WT Avp2 was cloned into the pMALc2E Maltose binding vector and was expressed in the Rosetta(DE3) cell line. 3ml of cells grown to log phase were induced with 0, 0.02, 0.04, 0.08, 0.1, 0.2 and 0.4mM IPTG for 6hrs at 25°C. Two 1ml fractions of the cells were removed and centrifuged to form a cell pellet. One pellet was loaded directly onto a 15% SDS-PAGE mini-gel, whilst the other pellet was sonicated before loading. These pellets were then used to analyse the (a) overall and (b) insoluble(I)/soluble(S) expression levels respectively.

(a) Overall Expression



(b) Insoluble/Soluble Expression



3.2.1.3.2 Analysing Binding to Amylose Resin

WT Avp2MALc2E expressed in Rosetta(DE3) cells produced soluble Avp2 (Section 3.2.1.3.1). Therefore, preceding large scale purification, a small scale experiment was carried out to investigate whether the soluble protein could bind to the Amylose resin (Method 2.5.1) and only after this, was a large scale purification method investigated.

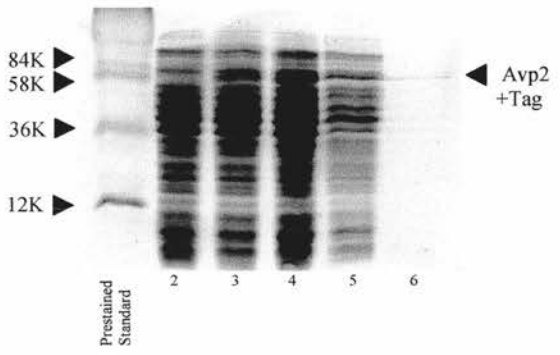
5ml of culture was grown at 25°C for 12-16hr in LB/Carbenicillin/Chloramphenicol, 0.2% (v/v) Glucose then induced with 0.4mM IPTG at 25°C for 6hr and the buffer used for re-suspension was Column buffer (20mM Tris-HCl, 200mM NaCl, 1mM EDTA, pH7.4). Samples taken were pre-induction cell pellet; post-induction cell pellet; post-sonication cell pellet (insoluble fraction); post-sonication supernatent (soluble fraction) and the protein bound to the Amylose resin (washed with Column buffer) (Method 2.5.1). These samples were separated on a 15% SDS-PAGE

mini-gel (Method 2.4.1.1), which was stained with Coomassie Blue (Method 2.4.2.1) and the results are shown in Figure 3.2.6.

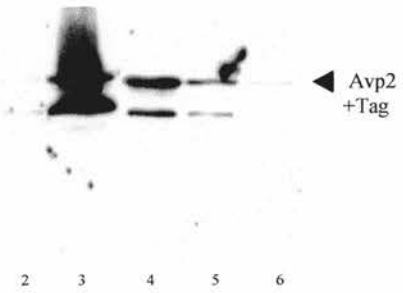
The results showed that the expressed protein was found mainly in the supernatant after sonication (Lane 5) and that the protein was capable of binding Amylose Resin (Lane 6). The band was confirmed to be Avp2 by Western blotting with R11 primary antibody (Method 2.4.3).

Figure 3.2.6: Binding of Avp2MALc2E to Amylose Resin
Avp2 cloned into the pMALc2E vector and expressed in the Rosetta(DE3) cell line produced soluble Avp2-Mal protein. To analyse its ability to bind Amylose resin, 5ml of culture was grown at 25°C for 12-16hr and then induced with 0.4mM IPTG at 25°C for 6hr. Samples taken were, pre-induction cell pellet **(Lane 2)**; post-induction cell pellet **(Lane 3)**; post-sonication cell pellet (insoluble fraction) **(Lane 4)**; post-sonication supernatant (soluble fraction) **(Lane 5)** and the protein bound to the Amylose resin **(Lane 6)**. These samples were separated on a 15% SDS-PAGE mini-gel, which was (a) stained with Coomassie Blue and (b) Western blotted using the R11 primary antibody.

(a) 15% SDS-PAGE mini-gel



(b) Western blot



3.2.1.3.3 Purification

Avp2MALc2E expressed in the Rosetta(DE3) cell line was grown in LB/Carbenicillin/Chloramphicol, 0.2% (v/v) Glucose according to Method 2.5.2. They were then induced with 0.4mM IPTG for 6hr at 25°C (Method 2.5.2). Following this, they

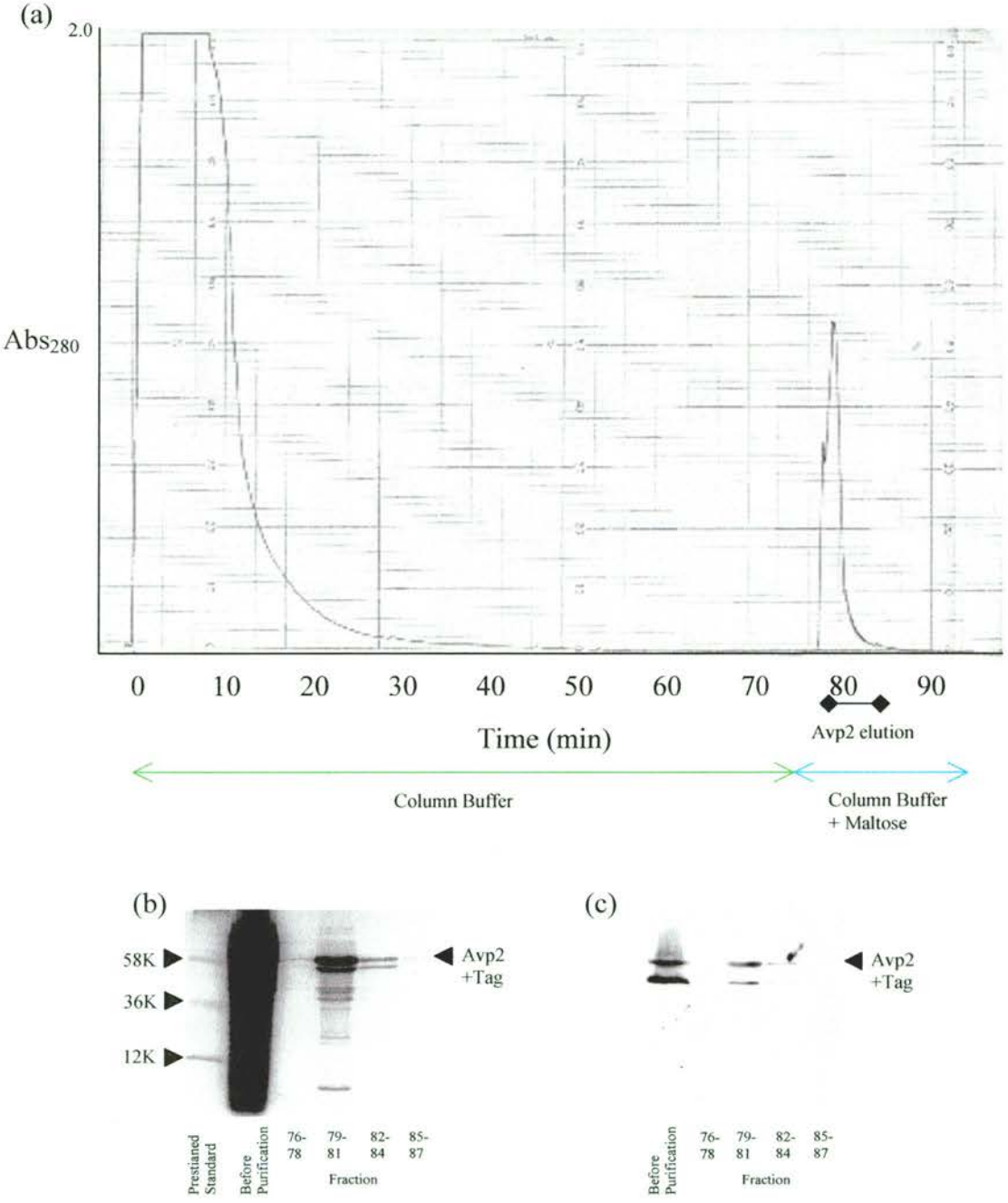
were extracted (Method 2.5.3.2) and then re-suspended in 10ml Column buffer (20mM Tris-HCl, 200mM NaCl, 1mM EDTA, pH7.4).

A 1 x 5cm Amylose resin (NEB) column was equilibrated with Column buffer (20mM Tris-HCl, 200mM NaCl, 1mM EDTA, pH7.4) for 60min at a constant flow rate of 1ml/min. The cell extract from Method 2.5.3.2 was then applied to the column in the presence of Column Buffer at a flow rate of 1ml/min for a further 75min with protein elution being monitored at 280nm. Following this, the protein was eluted into 3ml samples in the presence of Elution Buffer (20mM Tris-HCl, 200mM NaCl, 1mM EDTA, 10mM Maltose, pH7.4) for 20min, at a flow rate of 1ml/min.

Figure 3.2.7 shows the elution profile of expressed Avp2 with the Maltose tag. The first peak was the elution of a number of contaminating proteins which cannot bind the column. Whilst the second peak was the elution of Avp2 and other contaminating proteins which are capable of binding the Amylose column and are only eluted in the presence of Maltose. The elution of these proteins was followed using a 15% SDS-PAGE mini-gel (Method 2.4.1.1), which was stained with Coomassie Blue (Method 2.4.2.1) and the Avp2 band was confirmed by Western blotting using the R11 primary antibody (Method 2.4.3).

Immediately below the band considered to be the Avp2-fusion protein was another band. This band was a degradation product, which was thought to be caused by the degradation of the Maltose tag as degradation of Avp2 had not been previously been noted either in the absence of a tag or in the presence of a His tag.

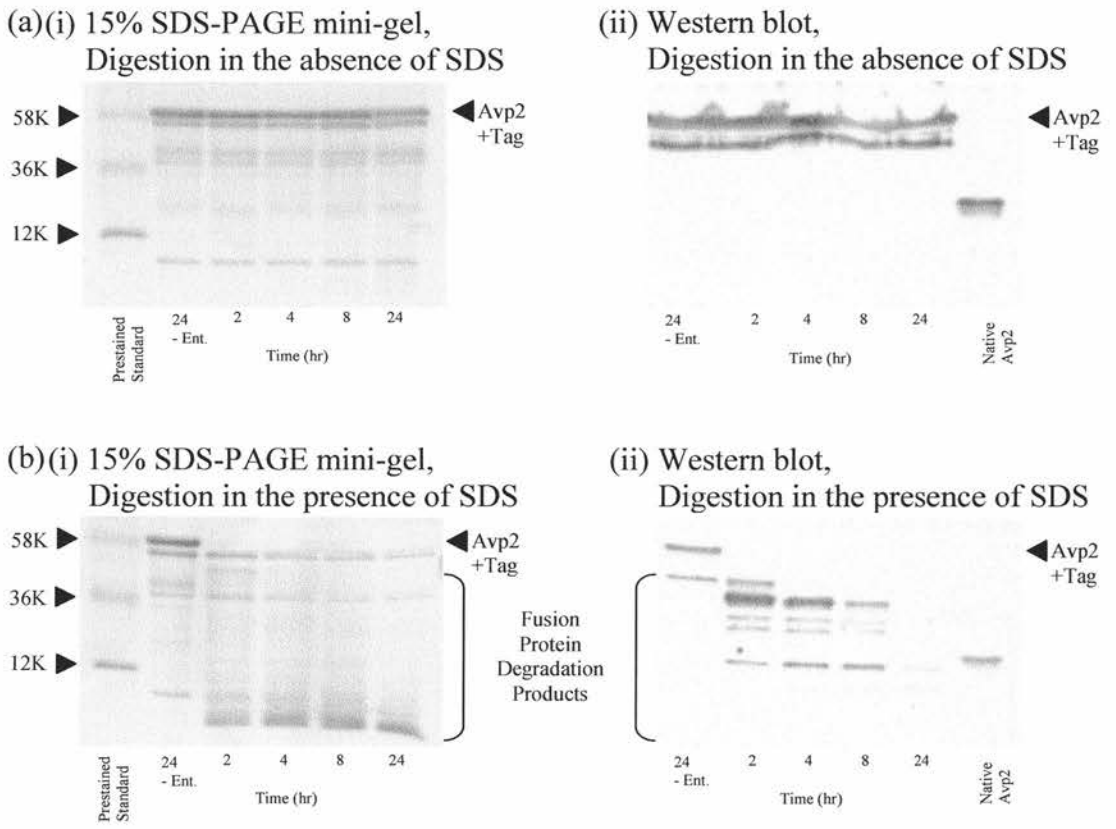
Figure 3.2.7: Amylose Purification of Avp2MALc2E
Avp2 with the Maltose tag was purified on a (a) 1 x 5cm Amylose column and the elution of proteins from this column was recorded at 280nm. Once the cell extract was applied to the column, contaminating proteins were eluted over the first 40min in the presence of Column Buffer at a flow rate of 1ml/min. Column Buffer continued to be passed through the column for a total of 75min to remove all traces of these contaminating proteins. Following this, Elution Buffer (Column Buffer + 10mM Maltose) was passed through the column to elute the Avp2-Maltose protein at a flow rate of 1ml/min and 3ml fractions were collected. Avp2 and some contaminating proteins were eluted from the column between 76-87 min. This elution was confirmed by (b) SDS-PAGE analysis and (c) Western blotting using the R11 antibody.



Following purification, fraction 79-81 was used in an initial attempt to cleave the tag from the Avp2 using Enterokinase (Method 2.5.6). The solution was incubated at 25°C with 1ng of Enterokinase and samples were taken at 2hr, 4hr, 8hr and 24hr. A control was also set up for 24hr in the absence of Enterokinase. These samples were then separated on a 15% SDS-PAGE mini-gel which was stained with Coomassie Blue (Figure 3.2.8(a)(i)) and then Western blotted using the R11 primary antibody (Figure 3.2.8(a)(ii)). However, the results of the analysis suggested that there was no cleavage. It was believed that this was due to the Avp2 protein folding over the recognition site for Enterokinase.

Therefore the experiment was repeated in the presence of 0.05% SDS. It was hoped that the SDS would bind the protein and cause a conformational change which would remove it from the cleavage site. The results of the SDS-PAGE mini-gel (Figure 3.2.8(b)(i)) and Western blot (Figure 3.2.8(b)(ii)) suggested that the protein was degraded when both Enterokinase and SDS were present. The reason for this may have been that the SDS unfolded the protein, thus allowing the Enterokinase to cleave it.

Figure 3.2.8: Cleavage of the Maltose Tag from Avp2MALc2E
Cleavage of the tag from the purified Avp2 Maltose tag fusion protein was carried out over a 24hr period at 25°C in the presence of Enterokinase +/- SDS. Aliquots (including a control, which was left for 24hr and had no Enterkinase added) were taken at 2hr, 4hr, 8hr and 24hr. 15% SDS-PAGE mini-gels and Western blots using the R11 primary antibody were used to analyse these samples which had been incubated with Enterokinase both (a) in the absence and (b) presence of SDS.



3.2.1.4 Avp2Lou2

The reason for investigating another Maltose tag vector was that the previously studied pMALc2E vector allowed the production of large quantities of recombinant Avp2 but the Maltose tag could not be cleaved at the Enterokinase site (Section 3.2.1.3.3).

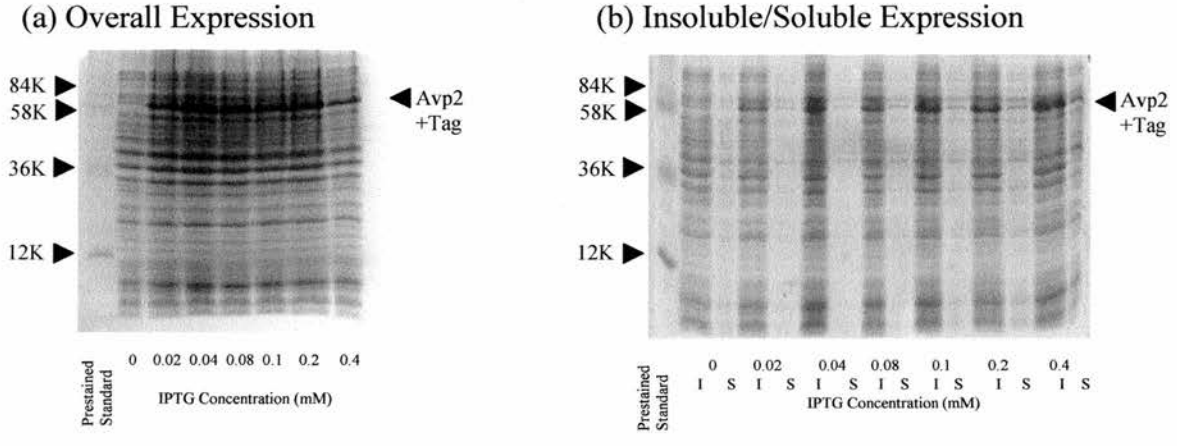
Therefore, Avp2 was cloned into the Maltose tag vector pLou2 (a gift from Dr. L. Major; Appendix I.C), which was a modified pMALc2X vector (NEB). The vector had its

Enterokinase cleavage site removed and replaced with a Tev cleavage site. It was thought that this construct would allow the Maltose tag to be removed from Avp2.

3.2.1.4.1 Cloning and Expression

Avp2 was cloned and expressed as discussed for Avp2pMALc2E (Section 3.2.1.3.1). The results of the expression and solubility test (Figure 3.2.9) showed that Avp2 was expressed in this vector and was found in both the soluble and insoluble fractions.

Figure 3.2.9: Expression of Avp2Lou2
WT Avp2 was cloned into the pLou2 Maltose binding vector and was expressed in the Rosetta(DE3) cell line. 3ml of cells grown to log phase were induced with 0, 0.02, 0.04, 0.08, 0.1, 0.2 and 0.4mM IPTG for 6hrs at 25°C. Two 1ml fractions of the cells were removed and centrifuged to form a cell pellet. One pellet was loaded directly onto a 15% SDS-PAGE mini-gel, whilst the other pelllet was sonicated before loading. These pellets were then used to analyse the (a) overall and (b) insoluble(I)/soluble(S) expression levels respectively.



3.2.1.4.2 Analysing Binding to Amylose

WT Avp2pLou2 transformed into Rosetta(DE3) cells produced soluble Avp2 (Section 3.2.1.4.1). Preceding large scale purification, a small scale experiment was carried

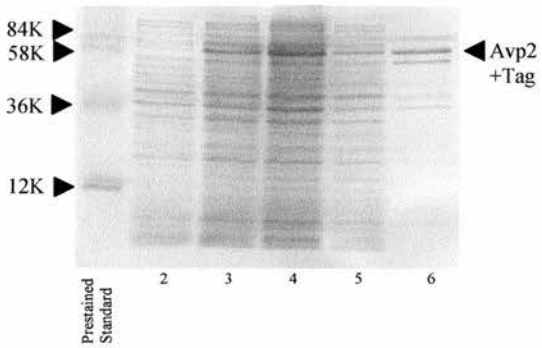
out to investigate whether the soluble protein could bind the Amylose resin (Method 2.5.1) and only after this, was a large scale purification method investigated.

5ml of culture was grown at 25°C for 12-16hr in LB/Carbenicillin/Chloramphenicol, 0.2% (v/v) Glucose, then induced with 0.4mM IPTG at 25°C for 6hr and the buffer used for re-suspension was Column buffer 1 (20mM Tris-HCl, 200mM NaCl, 1mM EDTA, pH7.4). Samples taken were pre-induction cell pellet; post-induction cell pellet; post-sonication cell pellet (insoluble fraction); post-sonication supernatent (soluble fraction) and the protein bound to the Amylose resin (washed with Column buffer 1) (Method 2.5.1). These samples were separated on a 15% SDS-PAGE mini-gel (Method 2.4.1.1), which was stained with Coomassie Blue (Method 2.4.2.1) and the results are shown in Figure 3.2.10(a).

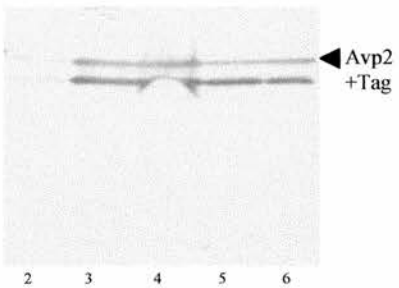
The results showed that the expressed protein was found in the supernatant after sonication (Lane 5) and this protein was capable of binding Amylose Resin (Lane 6) (Figure 3.2.10). The band was confirmed to be the Maltose tagged Avp2 by Western blotting with the R11 primary antibody (Method 2.4.3) (Figure 3.2.10(b)).

Figure 3.2.10: Binding of Avp2Lou2 to Amylose Resin
Avp2 cloned into the pLou2 vector and expressed in the Rosetta(DE3) cell line produced soluble Avp2-Mal protein. To analyse its ability to bind Amylose resin, 5ml of culture was grown at 25°C for 12-16hr and then induced with 0.4mM IPTG at 25°C for 6hr. Samples taken were, pre-induction cell pellet **(Lane 2)**; post-induction cell pellet **(Lane 3)**; post-sonication cell pellet (insoluble fraction) **(Lane 4)**; post-sonication supernatant (soluble fraction) **(Lane 5)** and the protein bound to the Amylose resin **(Lane 6)**. These samples were separated on a 15% SDS-PAGE mini-gel, which was (a) stained with Coomassie Blue and (b) Western blotted using the R11 primary antibody.

(a) 15% SDS-PAGE mini-gel



(b) Western blot



3.2.1.4.3 Purification

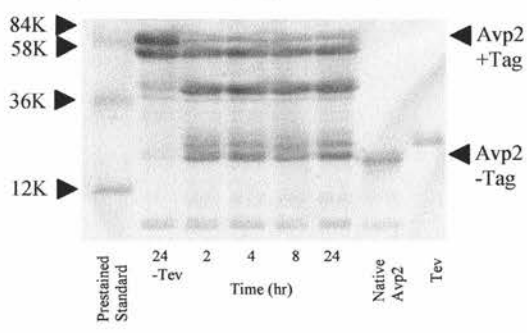
Avp2Lou2 expressed in the Rosetta(DE3) cell line was grown in LB/Carbenicillin/Chloramphenicol, 0.2% (v/v) Glucose according to Method 2.5.2. They were then induced with 0.4mM IPTG for 6hr at 25°C (Method 2.5.2). Following this, they were extracted (Method 2.5.3.2) and then re-suspended in 10ml Column buffer 1(20mM Tris-HCl, 200mM NaCl, 1mM EDTA, pH7.4).

A 1 x 5cm Amylose resin (NEB) column was equilibrated with Column buffer 1. The cell extract (Method 2.5.3.2) was then applied to the column in the presence of Column Buffer 1 at a flow rate of 1ml/min for 75min. Following this, the protein was eluted into 3ml samples in the presence of Elution Buffer (20mM Tris-HCl, 200mM NaCl, 1mM EDTA, 10mM Maltose, pH7.4) for 20min, at a flow rate of 1ml/min. The protein was mainly eluted in the first 6ml, as had been found for Avp2MALc2E (Section 3.2.1.3.3).

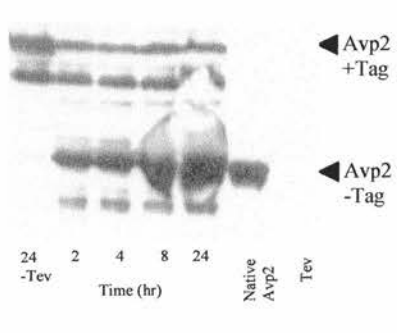
Once eluted from the column, the protein was dialysed against Column buffer 2 (20mM Tris-HCl, 200mM NaCl, pH7.4) for 4hr with two changes of buffer. This removed the EDTA from the solution (EDTA was removed because the protein solution was to be passed through a Nickel column later in the protocol to remove the His-Tev protease). However, this did not remove the Maltose from the solution as this is difficult to remove by dialysis. Following dialysis the protein solution was diluted to 5mg/ml and the Maltose tag was cleaved from Avp2 over a time period. 1μg of Tev protease was added to the solution and the solution was left to digest at 25°C for 2hr, 4hr, 8hr and 24hr. A control incubation without Tev was carried out (Method 2.5.6). As Figure 3.2.11 shows, these samples were separated on a 15% SDS-PAGE mini-gel (Method 2.4.1.1) which was stained with Coomassie Blue (Method 2.4.2.1). Avp2 with and without the fusion tag were confirmed by Western blotting using the R11 primary antibody (Method 2.4.3). The Figure shows that cleavage occurred after 2hr. Therefore this Method was scaled up for the cleavage of the 6ml of sample.

Figure 3.2.11: Cleavage of the Maltose Tag from Avp2Lou2
Cleavage of the tag from the purified Avp2-Maltose tag fusion protein was carried out over a 24hr period at 25°C in the presence of 1μg of Tev Protease in 100μl of sample. 20μl aliquots (including a control) were taken at 2hr, 4hr, 8hr and 24hr. The samples were separated on a 15% SDS-PAGE mini-gel which was (a) stained with Coomassie Blue and (b) Western blotted using R11 primary antibody.

(a) 15% SDS-page mini-gel



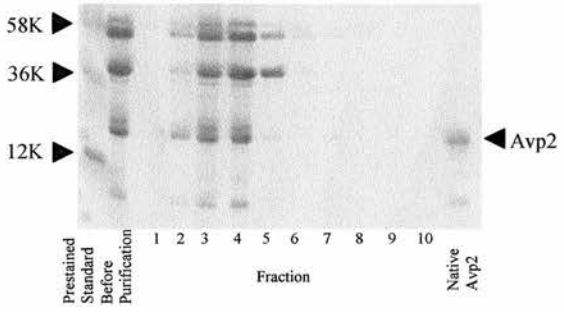
(b) Western blot



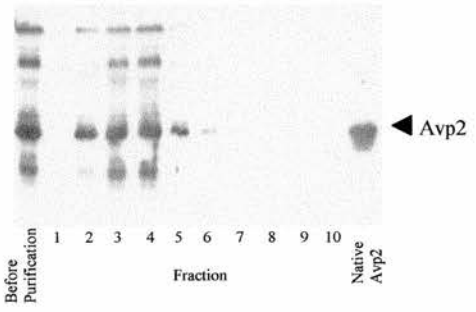
After cleavage of the Maltose tag from Avp2, the sample was applied to a Nickel column to remove the contaminating His tagged Tev Protease. The 1 x 5cm column was washed for 15min with water at 1ml/min and then further washed for 15min with Column buffer 2. The sample was then applied to the column in the presence of Column buffer 2 at a flow rate of 1ml/min, and 2ml samples were collected for 20min. These samples were separated on a 15% SDS-PAGE mini-gel which was stained with Coomassie Blue (Figure 3.2.12(a)) and Western blotted with the R11 primary antibody (Figure 3.2.12(b)). As Figure 3.2.12 shows, the protein was eluted between fractions 2-6, with most being eluted in fractions 3 and 4.

Figure 3.2.12: Nickel Bead Purification of Avp2Lou2
After the Maltose tag was removed from Avp2, the sample was passed applied to a Nickel column to remove the contaminating His tagged Tev Protease. The sample was applied to the column in the presence of 20mM Tris-HCl, 200mM NaCl, pH7.4 at a flow rate of 1ml/min, and 2ml samples were collected for 20min. Samples were separated on a 15% SDS-PAGE mini-gel which was (a) stained with Coomassie Blue and (b) Western blotted with R11 primary antibody.

(a) 15% SDS-PAGE mini-gel



(b) Western blot

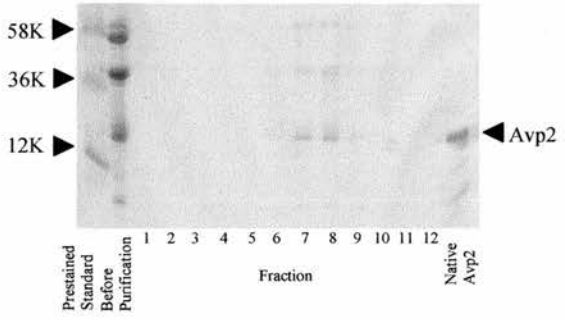


After removing the Tev protease, fractions 2-6 were pooled and applied to a DEAE ion exchange column. The function of this column was to remove Maltose from the solution, as Maltose would interfere with the last step of purification – the Amylose resin purification.

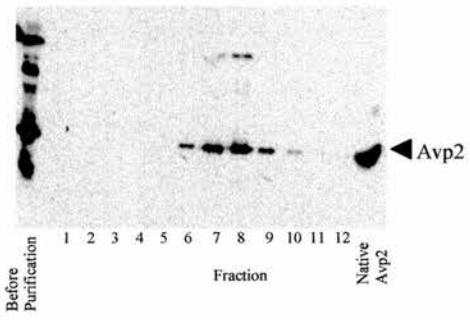
Fractions 2-6 were pooled and applied to a 1 x 30cm DEAE ion exchange column in the presence of Column buffer 2 at a rate of 1ml/min, and 2ml samples were collected for 30min. These samples were separated on a 15% SDS-PAGE mini-gel which was stained with Coomassie Blue (Figure 3.2.13(a)) and Western blotted using R11 primary antibody (Figure 3.2.13(b)). As Figure 3.2.13 shows, the protein was eluted between fractions 6-10, with most being eluted in fractions 7 and 8.

Figure 3.2.13: DEAE Purification of Avp2Lou2
After the His tagged Tev Protease was removed, the sample was applied to a 1 x 30cm DEAE column in the presence of 20mM Tris-HCl, 200mM NaCl, pH7.4 at a flow rate of 1ml/min, and 2ml samples were collected for 30min. Samples were separated on a 15% SDS-PAGE mini-gel which was (a) stained with Coomassie Blue and (b) Western blotted using R11 primary antibody.

(a) 15% SDS-page mini-gel



(b) Western blot

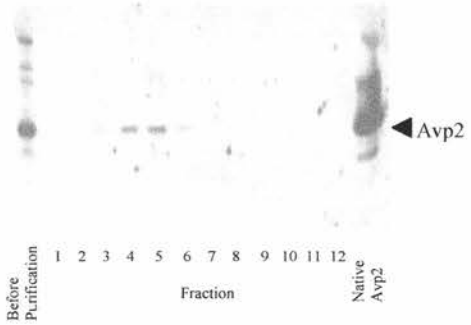


Following DEAE purification, fractions 6-10 were pooled and applied to an equilibrated 1 x 5cm Amylose resin column to remove contaminants. The sample was applied to the column in the presence of Column buffer 2 at a flow rate of 1ml/min, and 2ml samples were collected for 30min. The fractions were separated on a 15% SDS-PAGE mini-gel which was stained with Comassie blue and then Silver stained (Method 2.4.2.2). However, the protein was in such low yields that it could only be visualised when Western

blotted. As Figure 3.2.14 shows, the protein was eluted between fractions 3-6, with most being eluted between fractions 4 and 5.

The sample was concentrated to 200µl (Method 2.5.8) and found to be 0.03mg/ml (Method 2.4.5.2).

Figure 3.2.14: Amylose Resin Purification of Cleaved Avp2Lou2
After the Maltose was removed, the sample was applied to 1 x 5cm Amylose resin column to remove undigested Avp2 and other contaminating proteins. The sample was applied to the column in the presence of 20mM Tris-HCl, 200mM NaCl, pH7.4 at a flow rate of 1ml/min, and 2ml samples were collected for 30min. Samples were separated on a 15% SDS-PAGE mini-gel which was Western blotted using the R11 primary antibody.



3.2.1.5 Avp2Lou3

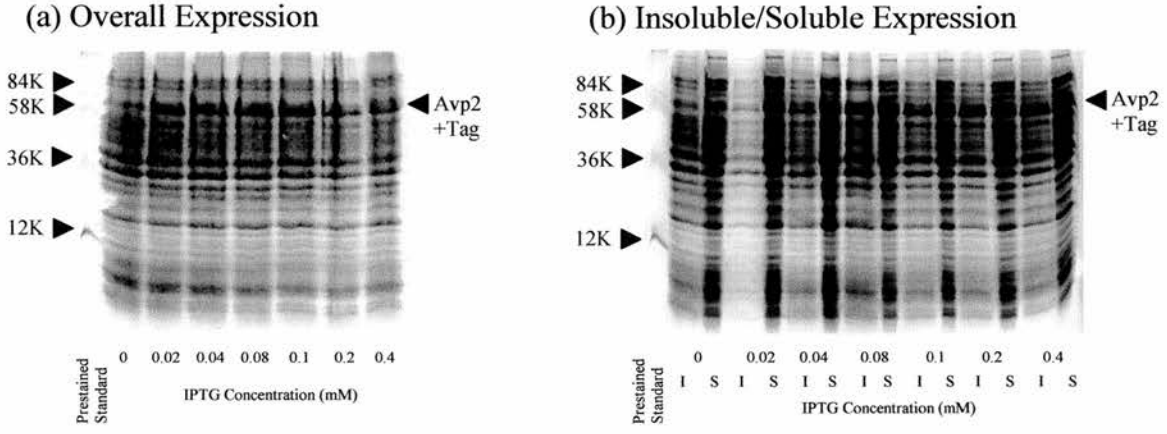
Avp2 cloned into the pLou2 vector produced low yields of purified Avp2 (Section 3.2.1.4.3). The production of such low concentrations of Avp2 was believed to be caused by the number of purification steps. Therefore another variation on this vector was analysed, pLou3 (Appendix I.C).

This vector had a tag placed in the order, His tag - Maltose tag - Tev cleavage site - MCS. Therefore, this was chosen for Avp2, for two reasons, firstly the Maltose tag should aid in improving the solubility of Avp2 and secondly, the purification required only two purification steps using a Nickel bead column.

3.2.1.5.1 Cloning and Expression

Avp2 was cloned and expressed as discussed for Avp2pMALc2E (Section 3.2.1.3.1), except the Column buffer used was 20mM Tris-HCl, 500mM NaCl, 10mM Imidazole, pH8. The results of the expression and solubility test (Figure 3.2.15) showed that Avp2 was expressed in this vector and was found in both the soluble and insoluble fractions.

Figure 3.2.15: Expression of Avp2Lou3
WT Avp2 cloned into the pLou3 vector was expressed in the Rosetta(DE3) cell line. 3ml of cells grown to log phase were induced with 0, 0.02, 0.04, 0.08, 0.1, 0.2 and 0.4mM IPTG for 6hrs at 25°C. Two 1ml fractions of the cells were removed and centrifuged to form a cell pellet. One pellet was loaded directly onto a 15% SDS-PAGE mini-gel, whilst the other pellet was sonicated before loading. These pellets were then used to analyse the (a) overall and (b) insoluble(I)/soluble(S) expression levels respectively.



3.2.1.5.2 Analysing Binding to Nickel Beads

WT Avp2pLou3 transformed into Rosetta(DE3) cells produced soluble Avp2 (Section 3.2.1.5.1). Preceding large scale purification, a small scale experiment was carried out to investigate whether the soluble protein could bind Nickel beads (Method 2.5.1) and only after this, was a large scale purification method investigated.

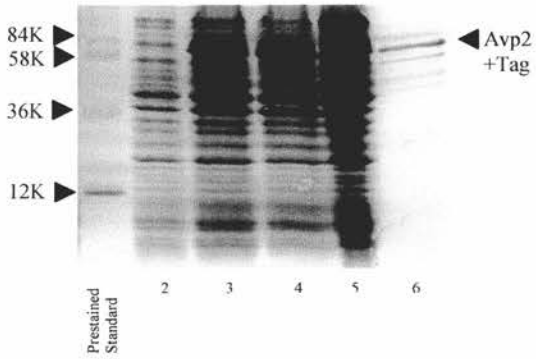
5ml of culture was grown at 25°C for 12-16hr in LB/Carbencillin/Chloramphenicol, 0.2% (v/v) Glucose, then induced with 0.4mM IPTG at 25°C for 6hr and the buffer used for re-suspension was Column buffer (20mM Tris-HCl, 500mM NaCl, 10mM Imidazole, pH8). Samples taken were pre-induction cell pellet; post-induction cell pellet; post-sonication cell pellet (insoluble fraction); post-sonication supernatant (soluble fraction) and the protein bound to Nickel beads (washed with Wash buffer (20mM Tris-HCl, 500mM NaCl, 30mM Imidazole, pH8). These samples were separated on a 15% SDS-PAGE mini-gel (Method 2.4.1.1.) which was stained with Coomassie Blue (Method 2.4.2.1) and the results are shown in Figure 3.2.16(a).

The results showed that the expressed protein was found in the supernatant after sonication (Lane 5) and this protein was capable of binding Nickel beads (Lane 6) (Figure 3.2.16). The band was confirmed to be the Avp2 fusion protein by Western blotting with the R11 primary antibody (Method 2.4.3) (Figure 3.2.16(b)).

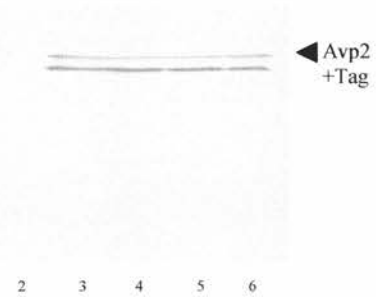
As has previously been shown for the pMALc2E (Figure 3.2.6) and pLou2 (Figure 3.2.10) vectors, there was a band immediately below the Avp2-fusion protein band. This band was a degradation product, which was thought to be caused by the degradation of the Maltose tag as degradation of Avp2 had not been previously been noted in the presence of a His tag.

Figure 3.2.16: Binding of Avp2Lou3 to Nickel beads
Avp2Lou3 expressed in the Rosetta(DE3) cell line produced soluble Avp2-His/Mal. To analyse its ability to bind Nickel beads, 5ml of culture was grown at 25°C for 12-16hr and then induced with 0.4mM IPTG at 25°C for 6hr. Samples taken were, pre-induction cell pellet **(Lane 2)**; post-induction cell pellet **(Lane 3)**; post-sonication cell pellet (insoluble fraction) **(Lane 4)**; post-sonication supernatant (soluble fraction) **(Lane 5)** and the protein bound to Nickel beads **(Lane 6)**. These samples were separated on a 15% SDS-PAGE mini-gel, which was (a) stained with Coomassie Blue and (b) Western blotted using R11 primary antibody.

(a) 15% SDS-PAGE mini-gel



(b) Western blot



3.2.1.5.3 Purification

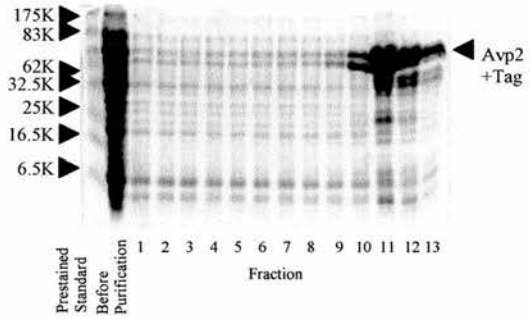
Avp2Lou3 expressed in the Rosetta(DE3) cell line was grown in LB/Carbenicillin/Chloramphenicol, 0.2% (v/v) Glucose according to Method 2.5.2. Following this, they were extracted (Method 2.5.3.2) and then re-suspended in 50ml Column buffer (20mM Tris-HCl, 500mM NaCl, 10mM Imidazole, pH8).

A 1.5 x 10cm Nickel column was washed for 45min with water at a flow rate of 1ml/min and then further washed for 45min with Column buffer. Following this, the cell extract was applied to the column for 45min in the presence of Wash buffer ((20mM Tris-HCl, 500mM NaCl, 30mM Imidazole, pH8)). Avp2Lou3 was then eluted into 2ml fractions for 45min with Elution buffer (20mM Tris-HCl, 500mM NaCl, 150mM Imidazole, pH8).

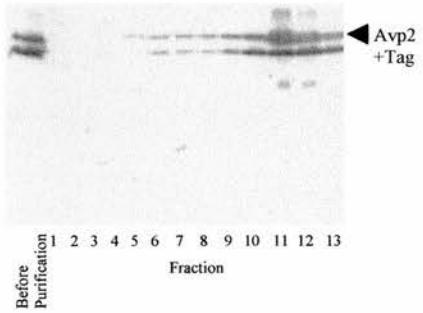
The eluted fractions were separated on a 15% SDS-PAGE mini-gel (Method 2.4.1.1) which was stained with Coomassie Blue (Method 2.4.2.1) and then Western blotted using R11 primary antibody (Method 2.4.3). The results are shown in Figure 3.2.17. Fractions 10-13 were pooled and dialysed against 20mM Tris-HCl, pH8 for 4hr with two changes of buffer.

Figure 3.2.17: First Nickel Bead Purification of Avp2Lou3
Avp2Lou3 was purified on a 1.5 x 10cm Nickel column. The extract was applied to the column and the column was washed for 45min at 1ml/min in the presence of 20mM Tris-HCl, 500mM NaCl, 30mM Imidazole, pH8. Following this, the tagged Avp2 was eluted into 2ml fractions at 1ml/min for 45min with 20mM Tris-HCl, 500mM NaCl, 150mM Imidazole, pH8. The fractions were separated on a 15% SDS-PAGE mini-gel which was (a) stained with Coomassie Blue and (b) Western blotted using R11 primary antibody.

(a) 15% SDS-page mini-gel



(b) Western blot



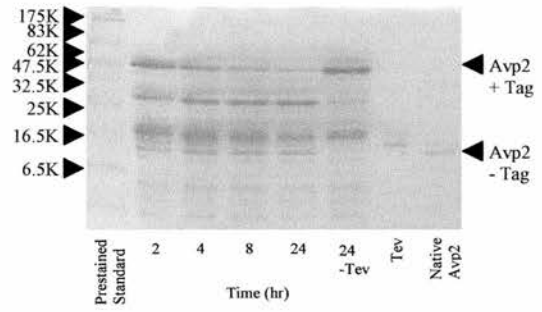
Following dialysis, the protein solution was diluted to 5mg/ml and the tag was cleaved from Avp2 over a time period. 1μg of Tev protease was added to solution and the solution was left to digest with gentle agitation at 25°C for 2hr, 4hr, 8hr and 24hr. A control incubation without Tev was also carried out (Method 2.5.6).

The cleavage products were separated on a 15% SDS-PAGE mini-gel which was stained with Coomassie Blue (Figure 3.2.18(a)) and then Western blotted using R11 primary antibody (Figure 3.2.18(b)). This cleavage was scaled up for large scale cleavage

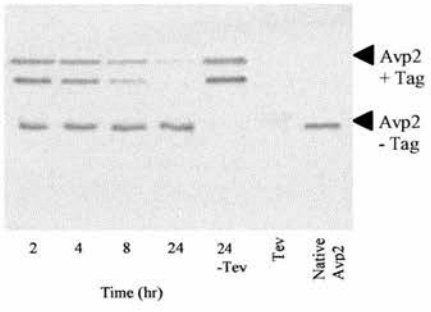
and 600μg Tev was added to the solution. The solution was left to digest with gentle agitation for 12-16hr.

Figure 3.2.18: Cleavage of the His/Maltose Tag from Avp2Lou3
Cleavage of the tag from the His/Maltose tagged Avp2Lou3 was carried out over a time course of 24hr at 25°C in the presence of 1μg of Tev protease added to 100μl sample. 20μl aliquots (including a control) were taken at 2hr, 4hr, 8hr and 24hr. The samples were separated on a 15% SDS-PAGE mini-gel which was (a) stained with Coomassie Blue and (b) Western blotted using R11 primary antibody.

(a) SDS-page mini-gel



(b) Western blot



Following cleavage, some protein had precipitated (Figure 3.2.19(a)). Therefore, the protein solution was centrifuged at 5,000rpm for 10min and the supernatant was removed. The supernatant was applied to the 1.5 x 10cm Nickel bead column and washed for 45min at a constant flow rate of 1ml/min with 20mM Tris-HCl, pH8 to elute the cleaved protein into 2ml fractions.

The results of the purification are shown in Figure 3.2.19. As the 15% SDS-PAGE mini-gel stained with Coomassie Blue shows (Figure 3.2.19(a)), the fractions were contaminated by other proteins and were in such low yields that the Avp2 could only be identified by Western blotting using R11 primary antibody (Figure 3.2.19(b)).

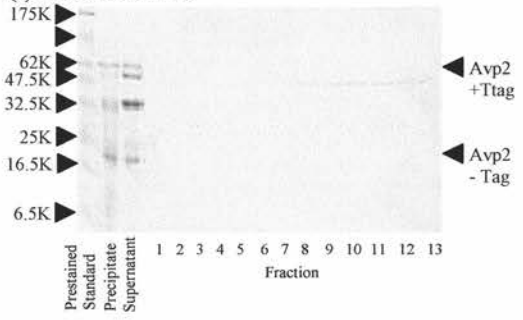
Therefore, in a bid to remove the contaminants, fractions 8-21 were pooled, concentrated to 8ml and then applied to a 1 x 30cm Diethylaminoethyl (DEAE) sepharose

column at 1ml/min in the presence of PBS, pH8. Analysis of the eluted fractions by Western blotting using R11 primary antibody suggested that the protein was either too dilute to be visualised, or had bound the column.

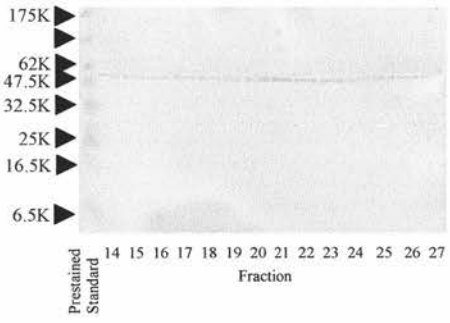
Figure 3.2.19: Second Nickel Bead Purification of Avp2pLou3
The protein solution was centrifuged and the supernatant was loaded onto the 1.5 x 10cm Nickel bead column. Fractions containing cleaved Avp2 were visualised using (a(i)(ii)) a 15% SDS-PAGE mini-gel, stained with Coomassie Blue and (b(i)(ii)) Western blotting using R11 primary antibody.

(a) 15% SDS-page mini-gel

(i) Fractions 1-13

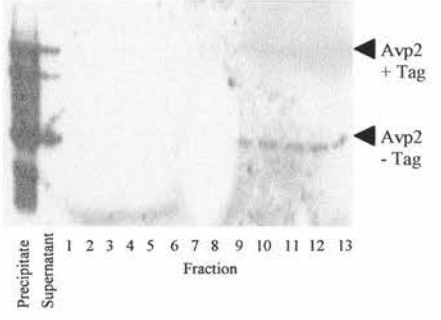


(ii) Fractions 14-27

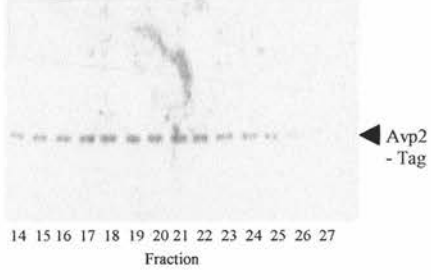


(b) Western blot

(i) Fractions 1-13



(ii) Fractions 14-27



3.2.1.6 Discussion

Avp2 was cloned into a number of tagged vector systems each of which had its own intrinsic problems. Factors used in the induction steps to try to increase soluble expression included, inducing at a low temperature of induction and using a low IPTG concentration. Together, it was hoped that these factors would reduce the rate of protein synthesis thereby allowing the proteins to fold correctly and thus be synthesised in soluble form. Likewise the Rosetta(DE3) cell line was used because it had been designed to enhance expression of proteins which contained codons rarely used by *E. coli*.

However, none of these factors appeared to aid soluble expression of His tagged Avp2 as the two His tag vectors investigated both produced mainly insoluble protein, even in the presence of chaperone proteins. Nevertheless, the soluble protein although produced in low concentrations was capable of binding Nickel beads and it may have been beneficial to continue using the N-terminal His-tag to purify under native conditions as the His tag vector, pET28b had previously been used by Mossessova & Lima, 2002 to purify the CE clan protease, Ulp1. Likewise, it may have been advantageous to use the partially purified His-tagged Avp2 protein in the previously developed IEC column purification method (Section 3.1.1). However it would be expected that the greatest amount of protein could have been purified under denaturing conditions, although it cannot be guaranteed that the protein would have been refolded correctly.

Therefore, three vectors which utilised a Maltose tag were then investigated. Primarily the MALc2E tag with an enterokinase cleavage site was used but although this produced soluble protein, the tag could not be removed. This led to the expression of Avp2 in the pLou2 vector, where the enterokinase cleavage site had been replaced with a Tev cleavage site. In this case, the tag was removed but the large number of purification steps led to the most of the protein being lost. Therefore, the last vector examined, pLou3

contained a Maltose tag which would help with solubility, a Tev cleavage site which had already been proven to be cleaved and an N-terminal His tag which meant only two purification steps were required. However, following purification the protein was found to be produced in low yields and was contaminated.

In summary the major problem was low expression of Avp2 in *E. coli* and therefore in future other expression systems should be investigated.

3.3 Maturation of pVIII

Introduction

As discussed in Sub-chapter 1.4, the sequence of mature VIII has not been published. However polypeptide VIII is thought to be cleaved from the precursor protein pVIII by Avp2 at one of three putative cleavage sites which conform to the cleavage motifs, (M,L,I)XGX-G and (M,L,I)XGG-X (Webster et al., 1989b; Anderson., 1990). This knowledge was the basis of work carried out in our laboratory by Dr. W.S. Annan (1999) to investigate the maturation of pVIII.

Dr. Annan made considerable progress towards the elucidation of these cleavage sites by utilising a recombinant His tagged pVIII protein The protein was purified under denaturing conditions, re-folded and then cleaved by Avp2. The cleavage products were subsequently studied using N-terminal sequencing and molecular mass size markers on an SDS-PAGE mini-gel. The results of this investigation proposed that pVIII could be cleaved at the three sites, 111-112, 131-132 and 157-158.

Although this gave an indication as to which residues of pVIII were cleaved by Avp2 it is not known whether the protein was folded correctly and therefore if all three sites are accessible by Avp2. Likewise no indication as to the order of cleavage was produced. Therefore the aims of this Sub-chapter were to analyse the maturation of pVIII to VIII and to investigate the putative structure of pVIII. Initially, a primary structure alignment of pVIII was utilised to examine whether any of the potential cleavage sites of pVIII were conserved. Following this, bioinformatics were employed to analyse both the putative secondary and tertiary structures of pVIII.

Subsequently, to further characterise the cleavage sites of pVIII an investigation into purifying recombinant pVIII using non-denaturing conditions was investigated, as was a mammalian *in vitro* transcription-translation expression system.

Results and Discussion

3.3.1 Bioinformatic Analysis

According to the cleavage motifs (M,L,I)XGX-G and (M,L,I)XGG-X for Avp2, the human Ad2 pVIII sequence contains three putative cleavage sites between residues, 108-112, 128-132, 154-158 (Figure 3.3.1). To investigate whether any of these sites were conserved, thirty-six pVIII proteins which represented the four genera of the *Adenoviridae* family were aligned. Following this, predictions of the secondary and tertiary structure of human Ad2 pVIII were made.

Figure 3.3.1: pVIII Sites Putatively Cleaved by Avp2
The consensus cleavage motifs for Avp2 were (M,L,I)XGX-G and (M,L,I,)XGG-X, therefore, pVIII contained three potential cleavage sites for Avp2. These cleavage sites were between residues 111-112, 131-132 and 157-158, and are highlighted in red, blue and purple boxes, respectively.

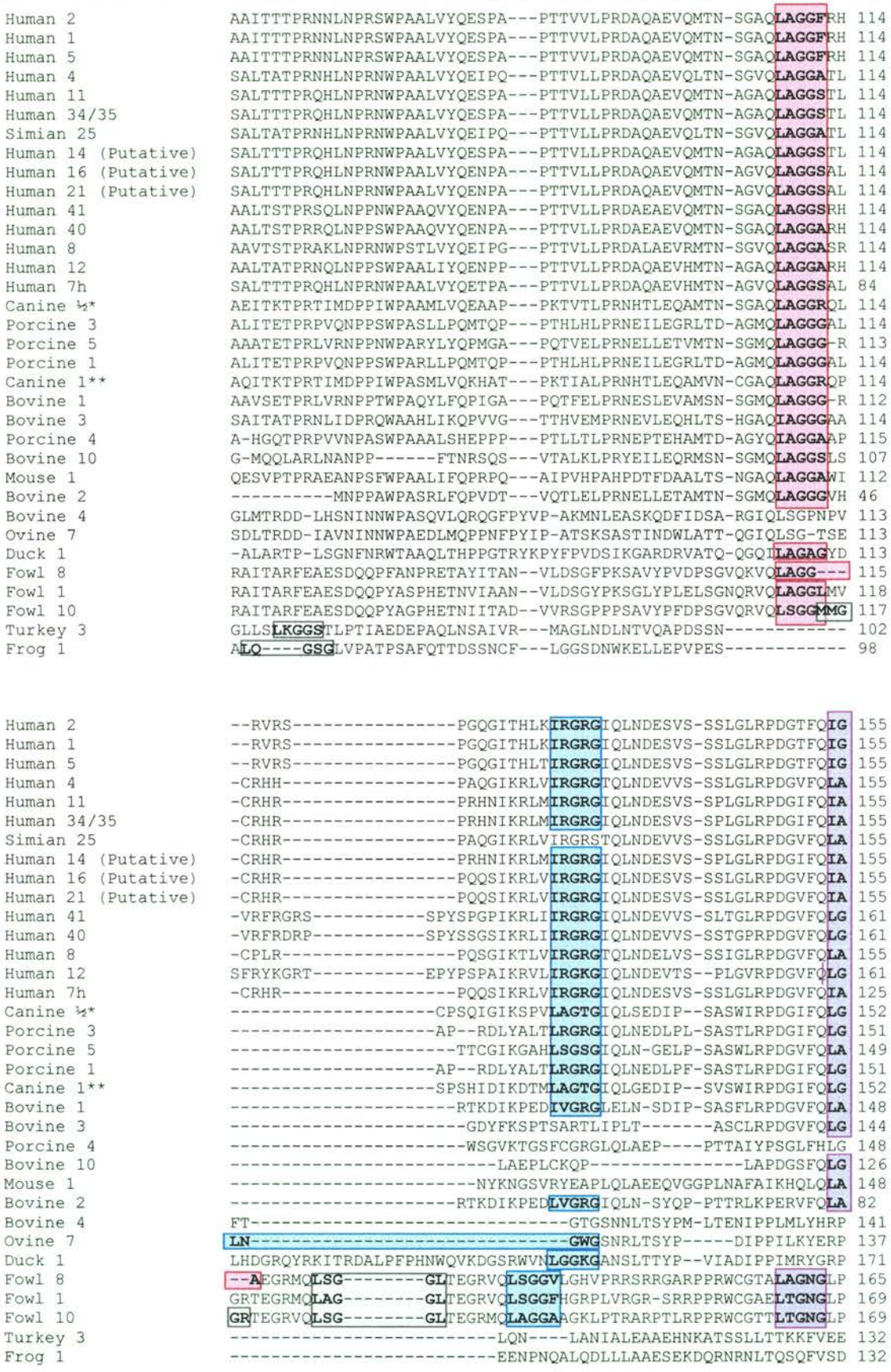
```
MSKEIPTPYMWSYQPQMGLAAGAAQDYSTRINYMSAGPHMISRVNGIRAHRNRILLEQAA  60

ITTTPRNNLNPRSWPAALVYQESPAPTTVVLPRDAQAEVQMTNSGAQLAGGFRHRVRSPG  120

QGITHLKIRGRGIQLNDESVSSSLGLRPDGTFQIGGAGRSSFTPRQAILTLQTSSSEPRS  180

GGIGTLQFIEEFVPSVYFNPFSGPPGHYPDQFIPNFDAVKDSADGYD               227
```

3.3.1.1 Primary Sequence Alignment of pVIII Proteins

Thirty-six *Adenoviridae* pVIII protein sequences were aligned according to Method 2.1.1. Subsequently, the sequences of the alignment were examined for the Avp2 cleavage motifs, (M,L,I)XGX-G and (M,L,I)XGG-X. Those sites which aligned with the putative human Ad2 pVIII cleavage sites at 108-112, 128-132, 154-158 (numbering for human Ad2 pVIII), were highlighted in red, blue and purple, respectively, whilst those highlighted with grey boxes, were cleavage motifs which lay outside of these sites (Figure 3.3.2).

Figure 3.3.2: Primary Sequence Alignment of *Adenoviridae* pVIII Proteins
The human pVIII sequence, found at NCBI (http://www.ncbi.nlm.nih.gov/) was pasted into WU-BLAST at EBI (http://www.ebi.ac.uk/blast2) where a similarity search was carried out. Each of the sequences from this search were then pasted into the CLUSTALW programme at EBI (http://www.ebi.ac.uk/clustalw/) and an alignment was performed. The three putative Avp cleavage sites of human Ad2 pVIII which were conserved in the alignment were highlighted. The cleavage site motifs selected were 108-112, 128-132, 154-158 (numbering for human Ad2 pVIII), and were highlighted with red, blue and purple boxes, respectively. Whilst those highlighted with grey boxes, were cleavage motifs which lay outside of these sites.
*Sequence for Canine 1 (strain Utrecht) and Canine 2 (strain Toronto A 26-61).
**Sequence for Canine 1 (strain CLL and strain RI261).

```
Human 2                --MSKEIPTPYMWSYQPQMGLAAGAAQDYSTRINYMSAGPHMISRVNGIRAHRNRILLEQ 58
Human 1                --MSKEIPTPYMWSYQPQMGLAAGAAQDYSTRINYMSAGPHMISRVNGIRAHRNRILLEQ 58
Human 5                --MSKEIPTPYMWSYQPQMGLAAGAAQDYSTRINYMSAGPHMISRVNGIRAHRNRILLEQ 58
Human 4                --MSKEIPTPYMWSYQPQMGLAAGAAQDYSTRMNWLSAGPGMISRVNDIRAHRNQILLEQ 58
Human 11               --MSKEIPTPYMWSYQPQMGLAAGASQDYSTRMNWLSAGPSMISRVNDIRAYRNQILLEQ 58
Human 34/35            --MSKEIPTPYMWSYQPQMGLAAGASQDYSTRMNWLSAGPSMISRVNDIRAYRNQILLEQ 58
Simian 25              --MSKEIPTPYMWSYQPQMGLAAGAAQDYSTRMNWLSAGPAMISRVNDIRAHRNQILLEQ 58
Human 14 (Putative)    --MSKEIPTPYMWSYQPQMGLAAGASQDYSTRMNWLSAGPSMISRVNDIRAYRNQILLEQ 58
Human 16 (Putative)    --MSKEIPTPYMWSYQPQMGLAAGASQDYSTRMNWLSAGPSMISRVNDIRAYRNQLLLEQ 58
Human 21 (Putative)    --MSKDIPTPYMWSYQPQMGLAAGASQDYSTRMNWLSAGPSMISRVNDIRAYRNQLLLEQ 58
Human 41               --MSKEIPTPYMWSYQPQMGLAAGASQDYSSRMNWLSAGPHMIGRVNGIRATRNQILLEQ 58
Human 40               --MSKDIPTPYMWSYQPQMGLAAGASQDYSSRMNWLSAGPHMIGRVNGIRATRNQILLEQ 58
Human 8                --MSKEIPTPYMWSYQPQMGLAAGASQDYSTRMNWLSAGPSMISQVNGVRNHRNQILLEQ 58
Human 12               --MSKDIPTPYMWSFQPQMGLAAGAAQDYSSKMNWLSAGPHMISRVNGVRARRNQILLEQ 58
Human 7h               ------------------------------MNWLSAGPSMISRVNDIRAYRNQLLLEQ 28
Canine ½*              --MSKEIPTPYMWSYQPQTGHAAGASQDYSTQMNWFSAGPSMISQVYGIRDLRNKVLITQ 58
Porcine 3              --MSKQIPTPYMWSYQPQSGRAAGASVDYSTRMNWLSAGPSMIGQVNDIRHTRNQILIRQ 58
Porcine 5              --MSKDIPTPYVWTYQPQLGQAAGASQDYSTRMNWLSAGPSMIDQVNQIRVERNNILLRQ 58
Porcine 1              --MSKQIPTPYMWSYQPQSGRAAGASVDYSTRMNWLSAGPSMIGQVNDIRHTRNQILIRQ 58
Canine 1**             --MSKEIPTPYIWSYQPQTGHAAGASQDYSTQMNWFSAGPSMISHVYGIRDLRNKVLMTQ 58
Bovine 1               --MSKDIPTPYVWTFQPQLGGCG-ASQDYSTRMNWLSAGPSMINQVNSVRADRNRILLRQ 57
Bovine 3               --MSKEIPTPYVWTFQPQMGAAAGASQDYSTRMNWFSAGPDMIHDVNNIRDAQNRILMTQ 58
Porcine 4              MPLAKEIPTPYVWSYQPQMGVPAGASQDYSTKINCLSAGPRMAQTVFALRDQRNRVLTAE 60
Bovine 10              --MSKEIPTPYVWNYQPQAGTAAGASQDYSTRLHWLSAGKSMISKVNTICENRNSILLSK 58
Mouse 1                --MS--APSPYVWTFQPQRGTAAGASQDYSTRINWLSAGPELRGKVVQLNEARNAILMKE 56
Bovine 2               ------------------------------------------------------------
Bovine 4               ----MQPVTPYIWKYQPETGHTAGAHQDYGSVINWLQSNPQMFNRIRDINLQRNNIDQTQ 56
Ovine 7                ---MAQPVTPYVWKYQPETGYTAGAHQNYNTVINWLHANPQMFARIQHINTARNVMDKFR 57
Duck 1                 ----MQPVTPYLWRYQPETGTAAGARQDYGAVINWFNSGPDLYRRIRDVNITRNNVEQTR 56
Fowl 8                 MNLLDATPTEYVWKYNPLSGIPAGAQQNYGATINWVVPGGNSFAYAADEIRRNTLSPAVT 60
Fowl 1                 MNLMNATPTEYVWKYNPVSGIPAGAQQNYGATIDWVLPGGTGFAIATNDIRRQTLNPAVT 60
Fowl 10                MNLLNAAPTPYVWKYNPVTGKCAGANRTT-AHYRLVLPGGNSFAYAADEIRRRFPEPAVT 59
Turkey 3               ---MDPVPLEYIWQYNPVTGRVGGANQNYGQRINVLHTNRYLYNRMQNVQKKSNERATER 57
Frog 1                 ---MEAEPSEYIWQYNPVTGRVVGASQNFGARINTLHASPQLWARMQQVQRRRNETVTLS 57
```


Human 2 AAITTTPRNNLNPRSWPAALVYQESPA---PTTVVLPRDAQAEVQMTN-SGAQLAGGFRH 114
Human 1 AAITTTPRNNLNPRSWPAALVYQESPA---PTTVVLPRDAQAEVQMTN-SGAQLAGGFRH 114
Human 5 AAITTTPRNNLNPRSWPAALVYQESPA---PTTVVLPRDAQAEVQMTN-SGAQLAGGFRH 114
Human 4 SALTATPRNHLNPRNWPAALVYQEIPQ---PTTVLLPRDAQAEVQLTN-SGVQLAGGATL 114
Human 11 SALTTTPRQHLNPRNWPAALVYQESPA---PTTVLLPRDAQAEVQMTN-AGAQLAGGSTL 114
Human 34/35 SALTTTPRQHLNPRNWPAALVYQESPA---PTTVLLPRDAQAEVQMTN-AGAQLAGGSTL 114
Simian 25 SALTATPRNHLNPRNWPAALVYQEIPQ---PTTVLLPRDAQAEVQLTN-SGVQLAGGATL 114
Human 14 (Putative) SALTTTPRQHLNPRNWPAALVYQESPA---PTTVLLPRDAQAEVQMTN-AGAQLAGGSTL 114
Human 16 (Putative) SALTTTPRQHLNPRNWPAALVYQENPA---PTTVLLPRDAQAEVQMTN-AGVQLAGGSAL 114
Human 21 (Putative) SALTTTPRQHLNPRNWPAALVYQENPA---PTTVLLPRDAQAEVQMTN-AGVQLAGGSAL 114
Human 41 AALTSTPRSQLNPPNWPAAQVYQENPA---PTTVLLPRDAEAEVQMTN-SGAQLAGGSRH 114
Human 40 AALTSTPRRQLNPPSWPAAQVYQENPA---PTTVLLPRDAEAEVQMTN-SGAQLAGGARH 114
Human 8 AAVTSTPRAKLNPRNWPSTLVYQEIPG---PTTVLLPRDALAEVRMTN-SGVQLAGGASR 114
Human 12 AALTATPRNQLNPPSWPAALIYQENPP---PTTVLLPRDAQAEVHMTN-AGAQLAGGARH 114
Human 7h SALTTTPRQHLNPRNWPAALVYQETPA---PTTVLLPRDAQAEVHMTN-AGVQLAGGSAL 84
Canine ½* AEITKTPRTIMDPPIWPAAMLVQEAAP---PKTVTLPRNHTLEQAMTN-SGAQLAGGRQL 114
Porcine 3 ALITETPRPVQNPPSWPASLLPQMTQP---PTHLHLPRNEILEGRLTD-AGMQLAGGGAL 114
Porcine 5 AAATETPRLVRNPPNWPARYLYQPMGA---PQTVELPRNELLETVMTN-SGMQLAGGG-R 113
Porcine 1 ALITETPRPVQNPPSWPARLLPQMTQP---PTHLHLPRNEILEGRLTD-AGMQLAGGGAL 114
Canine 1** AQITKTPRTIMDPPIWPASMLVQKHAT---PKTIALPRNHTLEQAMVN-CGAQLAGGRQP 114
Bovine 1 AAVSETPRLVRNPPTWPAQYLFQPIGA---PQTFELPRNESLEVAMSN-SGMQLAGGG-R 112
Bovine 3 SAITATPRNLIDPRQWAAHLIKQPVVG---TTHVEMPRNEVLEQHLTS-HGAQIAGGGAA 114
Porcine 4 A-HGQTPRPVVNPASWPAAALSHEPPP---PTLLTLPRNEPTEHAMTD-AGYQIAGGAAP 115
Bovine 10 G-MQQLARLNANPP------FTNRSQS---VTALKLPRYEILEQRMSN-SGMQLAGGSLS 107
Mouse 1 QESVPTPRAEANPSFWPAALIFQPRPQ---AIPVHPAHPDTFDAALTS-NGAQLAGGAWI 112
Bovine 2 ----------MNPPAWPASRLFQPVDT---VQTLELPRNELLETAMTN-SGMQLAGGGVH 46
Bovine 4 GLMTRDD-LHSNINNWPASQVLQRQGFPYVP-AKMNLEASKQDFIDSA-RGIQLSGPNPV 113
Ovine 7 SDLTRDD-IAVNINNWPAEDLMQPPNFPYIP-ATSKSASTINDWLATT-QGIQLSG-TSE 113
Duck 1 -ALARTP-LSGNFNRWTAAQLTHPPGTRYKPYFPVDSIKGARDRVATQ-QGQILAGAGYD 113
Fowl 8 RAITARFEAESDQQPFANPRETAYITAN--VLDSGFPKSAVYPVDPSGVQKVQLAGG--- 115
Fowl 1 RAITARFEAESDQQPYASPHETNVIAAN--VLDSGYPKSGLYPLELSGNQRVQLAGGLMV 118
Fowl 10 RAITARFEAESDQQPYAGPHETNIITAD--VVRSGPPPSAVYPFDPSGVQRVQLSGGMMG 117
Turkey 3 GLLSLKGGSTLPTIAEDEPAQLNSAIVR---MAGLNDLNTVQAPDSSN---------- 102
Frog 1 ALQ----GSGLVPATPSAFQTTDSSNCF---LGGSDNWKELLEPVPES---------- 98
Human 2 --RVRS----------PGQGITHLKIRGRGIQLNDESVS-SSLGLRPDGTFQIG 155
Human 1 --RVRS----------PGQGITHLKIRGRGIQLNDESVS-SSLGLRPDGTFQIG 155
Human 5 --RVRS----------PGQGITHLTIRGRGIQLNDESVS-SSLGLRPDGTFQIG 155
Human 4 -CRHH----------PAQGIKRLVIRGRGTQLNDEVVS-SSLGLRPDGVFQLA 155
Human 11 -CRHR----------PRHNIKRLMIRGRGIQLNDESVS-SPLGLRPDGIFQIA 155
Human 34/35 -CRHR----------PRHNIKRLMIRGRGIQLNDESVS-SPLGLRPDGIFQIA 155
Simian 25 -CRHR----------PAQGIKRLVIRGRSTQLNDEVVS-SSLGLRPDGVFQLA 155
Human 14 (Putative) -CRHR----------PRHNIKRLMIRGRGIQLNDESVS-SPLGLRPDGIFQIA 155
Human 16 (Putative) -CRHR----------PQQSIKRLVIRGRGIQLNDESVS-SSLGLRPDGVFQIA 155
Human 21 (Putative) -CRHR----------PQQSIKRLVIRGRGIQLNDESVS-SSLGLRPDGVFQIA 155
Human 41 -VRFRGRS----------SPYSPGPIKRLIIRGRGIQLNDEVVS-SLTGLRPDGVFQLG 161
Human 40 -VRFRDRP----------SPYSSGSIKRLIIRGRGIQLNDEVVS-SSTGPRPDGVFQLG 161
Human 8 -CPLR----------PQSGIKTLVIRGRGTQLNDELVS-SSIGLRPDGVFQLA 155
Human 12 SFRYKGRT----------EPYPSPAIKRVLIRGKGIQLNDEVTS--PLGVRPDGVFQLG 161
Human 7h -CRHR----------PQQSIKRLVIRGRGIQLNDESVS-SSLGLRPDGVFQIA 125
Canine ½* ----------CPSQIGIKSPVLAGTGIQLSEDIP--SASWIRPDGIFQLG 152
Porcine 3 ----------AP--RDLYALTLRGRGIQLNEDLPL-SASTLRPDGIFQLG 151
Porcine 5 ----------TTCGIKGAHLSGSGIQLN-GELP-SASWLRPDGVFQLA 149
Porcine 1 ----------AP--RDLYALTLRGRGIQLNEDLPF-SASTLRPDGIFQLG 151
Canine 1** ----------SPSHIDIKDTMLAGTGIQLGEDIP--SVSWIRPDGIFQLG 152
Bovine 1 ----------RTKDIKPEDIVGRGLELN-SDIP-SASFLRPDGVFQLA 148
Bovine 3 ----------GDYFKSPTSARTLIPLT-------ASCLRPDGVFQLG 144
Porcine 4 ----------WSGVKTGSFCGRGLQLAEP----PTTAIYPSGLFHLG 148
Bovine 10 ----------LAEPLCKQP--------------LAPDGSFQLG 126
Mouse 1 ----------NYKNGSVRYEAPLQLAEEQVGGPLNAFAIKHQLQLA 148
Bovine 2 ----------RTKDIKPEDLVGRGIQLN-SYQP-PTTRLKPERVFQLA 82
Bovine 4 FT----------GTGSNNLTSYPM-LTENIPPLMLYHRP 141
Ovine 7 LN----------GWGSNRLTSYP-----DIPPILKYERP 137
Duck 1 LHDGRQYRKITRDALPFFPHNWQVKDGSRWVNLGGKGANSLTTYP--VIADIPPIMRYGRP 171
Fowl 8 --AEGRMQLSG--------GLTEGRVQLSGGVLGHVPRRSRRGARPPRWCGTALAGNGLP 165
Fowl 1 GRTEGRMQLAG--------GLTEGRVQLSGGFHGRPLVRGR-SRRPPRWCGAELTGNGLP 169
Fowl 10 GRTEGRVQLSG--------GLTEGRMQLAGGAAGKLPTRARPTLRPPRWCGTTLTGNGLP 169
Turkey 3 ----------LQN----LANIALEAAEHNKATSSLLTTKKFVEE 132
Frog 1 ----------EENPNQALQDLLLAAESEKDQRNRNLTQSQFVSD 132

```
Human 2                GAGRSSFTP-RQAILTLQTS-SSEPRSGGIGT--LQFIEEFVPSVYFNPFSGP-PGHYPD 210
Human 1                GAGRSSFTP-RQAILTLQTS-SSEPRSGGIGT--LQFIEEFVPSVYFNPFSGP-PGHYPD 210
Human 5                GAGRPSFTP-RQAILTLQTS-SSEPRSGGIGT--LQFIEEFVPSVYFNPFSGP-PGHYPD 210
Human 4                GSGRSSFTP-RQAVLTLESS-SSQPRSGGIGT--LQFVEEFTPSVYFNPFSGS-PGHYPD 210
Human 11               GCGRSSFTP-RQAVLTLESS-SSQPRSGGIGT--VQFVEEFTPSVYFNPFSGS-PGHYPD 210
Human 34/35            GCGRSSFTP-RQAVLTLESS-SSQPRSGGIGT--VQFVEEFTPSVYFNPFSGS-PGHYPD 210
Simian 25              GSGRSSFTP-RQAVLTLESS-SSQPRSGGIGT--LQFVEEFTPSVYFNPFSGS-PGHYPD 210
Human 14 (Putative)    GCGRSSFTP-RQTVLTLESS-SSQPRSGGIGT--VQFVEEFTPSVYFNPFSGS-PGHYPD 210
Human 16 (Putative)    GCGRSSFTP-RQAVLTLESS-SSQPRSGGIGT--LQFVEEFTPSVYFNPFSGS-PGQYPD 210
Human 21 (Putative)    GCGRSSFTP-RQAVLTLESS-SSQPRSGGIGT--LQFVEEFTPSVYFNPFSGS-PGQYPD 210
Human 41               GAGRSSFTP-RQAYLTLQSS-SSQPRSGGIGT--LQFVEEFVPSVYFNPFSGA-PGLYPD 216
Human 40               GAGRSSFTP-RQAYLTLQSS-SSQPRSGGIGT--LQFVEEFVPSVYFNPFSGA-PGLYPD 216
Human 8                GAGRSSFTP-HQAYLTLQSS-SSEPRSGGIGT--LQFVEEFVPSVYFNPFSGS-PGLYPD 210
Human 12               GSGRSSFTA-RQAYLTLQSS-SSAPRSGGIGT--LQFVEEFTPSVYFNPFSGS-PGHYPD 216
Human 7h               GCGRSSFTP-RQAVLTLESS-SSQPRSGGIGT--LQFVEEFTPSVYFNPFSGS-PGQYPD 180
Canine ½*              GGSRSSFSP-TQAFLTLQQA-SSTPRAGGVGT--YQFVREFVPEVYLNPFSGP-PDTFPD 207
Porcine 3              GGGRSSFNP-TDAYLTLQNS-SSLPRSGGIGS--EQFVREFVPTVYINPFSGP-PGTYPD 206
Porcine 5              GGSRSSFSPGVSTLLRLEPS-SSLPRSGGIGS--TQFVQEFVPAVYFQPFSGP-PGTYPD 205
Porcine 1              GGGRSSFNP-TDAYLTLQNS-SSLPRSGGIGS--DQFVREFVPTVYINPFSGP-PGTYPD 206
Canine 1**             GGSRSSFSP-TQAFLTLQQA-SSTPRTGGVGS--YQFVREFVPEVYLNPFSGP-PDTFPD 207
Bovine 1               GGSRSSFNPGLSTLLTVQPA-SSLPRSGGIGE--VQFVHEFVPSVYFQPFSGP-PGTYPD 204
Bovine 3               GGSRSSFNP-LQTDFAFHAL-PSRPRHGGIGS--RQFVEEFVPAVYLNPYSGP-PDSYPD 199
Porcine 4              RGQRLVLDHQPARTLLLESA-PSVPRYGGIGA--RQFLKEFTPAVYPQPYSGP-PNTFPD 204
Bovine 10              GGAK-VFSRGATNYAIRKTI-LLLEYSGGMGE--KQFVKEFVPAVYFNPYSGL-PASYPD 181
Mouse 1                GGALSASMS------EMSGA-PRIPRSGGIGS--WQFSREFPPTVYLNPFSGS-PDTFPH 198
Bovine 2               GGSRSSFNPSINTLLTLQPA-ASVPRSGGIGE--VQFVHEFVPSVYFQPFSGP-PGSYPD 138
Bovine 4               GQQLQGMGTISKENFHLLEEVSRVPRSGGLTP--TEFLTQFPPIVYRHPFSSS-LIYFPK 198
Ovine 7                GQQLQGQGLFKQENIHLFYESPRLPRSGGLTP--QQFVKEFPPVVYNNPFSES-MSVFPK 194
Duck 1                 GQQLQGSG-FHRPSPALLTEEARVPRSHGMTV--RQFVHEFPPVVFNHPFSED-ITYFPK 227
Fowl 8                 ED-AEMVSDTYKYFLRTQGPSQVVSEPGVYSQ--RQFMTTFLPAVVPHPFDSPNPNDFPA 222
Fowl 1                 EQ-AEVTSDTYKYFLRTQGPSQVVEEPGVFSQ--RQFMTTFLPSVVPHPFDSTNPGDFPA 226
Fowl 10                ADYPEMTPDAFKYYLRVQGPSQEVDEPGVMSQ--RQFMTTFLPAMVPHPFDSESPDAFPA 227
Turkey 3               FPPVVYENPFSGSNFAYEFNPLYSPSGNEFSNPPFKFTGGAISLTGQHPVLSG-----GA 187
Frog 1                 FPPVVYERPFTGSDFPLEFSPLYSPDGNQFTN-----TG-WISQS-FPPNQSF-----PE 180
                                                                    .     *

Human 2                QFIPNFDAVKDSADGYD------- 227
Human 1                QFIPNFDAVKDSADGYD------- 227
Human 5                QFIPNFDAVKDSADGYD------- 227
Human 4                EFIPNFDAISESVDGYD------- 227
Human 11               EFIPNFDAISESVDGYD------- 227
Human 34/35            EFIPNFDAISESVDGYD------- 227
Simian 25              EFIPNFDAISESVDGYD------- 227
Human 14 (Putative)    EFIPNFDAISESVDGYD------- 227
Human 16 (Putative)    EFIPNFDAISESVDGYD------- 227
Human 21 (Putative)    EFIPNFDAISESVDGYD------- 227
Human 41               DFIPNYDAVSESVDGYD------- 233
Human 40               DFIPNYDAVSESVDGYD------- 233
Human 8                EFIPNFDAVREAVDGYD------- 227
Human 12               AFIPNFDAVSESVDGYD------- 233
Human 7h               EFIPNFDAISESVDGYD------- 197
Canine ½*              QFIPNYDIVTNSVDGYD------- 224
Porcine 3              QFIANYNILTDSVAGYD------- 223
Porcine 5              EFIYNYDIVSDSVDGYD------- 222
Porcine 1              QFIANYNILTDSVAGYD------- 223
Canine 1**             QFIPNYDIVTNSVDGYD------- 224
Bovine 1               EFIYNYDIVSDSVDGYD------- 221
Bovine 3               QFIRHYNVYSNSVSGYS------- 216
Porcine 4              YFCFNYDSVSNSVDGYS------- 221
Bovine 10              QFISNYDAVTDSVDGYA------- 198
Mouse 1                QFLSNYDSFSHTVDGYD------- 215
Bovine 2               EFIYNFDVATDSIDGYA------- 155
Bovine 4               EFDPLFSPHNDPRLTSSKTLQYVP 222
Ovine 7                EFSPLFNPSESLKKTSSQTLQYK- 217
Duck 1                 EFNPLFNPSEDYRVSSDRTLQYV- 250
Fowl 8                 QYSAIYKGTNAYEDVFWDW----- 241
Fowl 1                 QYSAIYKGRTAFEDTFWDW----- 245
Fowl 10                YFSSVYKGTNGFEPVFWQG----- 246
Turkey 3               VILSGQNPVLDKA----------- 200
Frog 1                 IINHPTQHSSTTI----------- 193
                                .
```

Following the alignment, a table was constructed of the potential cleavage sites for each pVIII protein according to the cleavage motifs, (M,L,I)XGX-G and (M,L,I)XGG-X (Table 3.3.1). As the table clearly shows, the cleavage site 111-112 was maintained throughout the Mastadenovirus genus with the exception of bovine Ad4. It was also found in both the Atadenovirus and Aviadenovirus genera.

Likewise, examples of site 131-132 were found in the Mastadenovirus, Atadenovirus and Aviadenovirus genera, although it was not as well conserved in the Mastadenovirus genus as had been found for site 111-112. Site 157-158 was also highly conserved within the Mastadenovirus genus and this is the only putative cleavage site that the bovine Ad4 sequence contained. Site 157-158 was however only found in the Mastadenovirus and Aviadenovirus genera.

Surprisingly, the sequences of the two Siadenovirus species did not contain any of the proposed cleavage sites and frog Ad1 pVIII contained only one putative cleavage site.

Therefore the alignment did not shed light upon whether one putative human Ad2 cleavage site was more important than any other.

Table 3.3.1: Avp2 Cleavage Sites of *Adenoviridae* pVIII Proteins

The table was constructed from the primary sequence alignment of 36 pVIII proteins. The cleavage sites are numbered according to the Ad2 pVIII sequence. The sites selected were 111-112, 131-132, 157-158 and were highlighted with red, blue and purple boxes, respectively. Whilst those sites highlighted with grey boxes were cleavage motifs which lay outside of these sites.

*Sequence for Canine 1 (strain Utrecht) and Canine 2 (strain Toronto A 26-61).

**Sequence for Canine 1 (strain CLL and strain RI261).

		Cleavage Sites (Numbering for Human Ad2 pVIII)					
Genus	**Species**	**From 1-110**	**At 111-112**	**From 113-130**	**At 131-132**	**At 157-158**	**From 159-End**
Mastadenovirus	*Human 2*						
Mastadenovirus	*Human 1*						
Mastadenovirus	*Human 5*						
Mastadenovirus	*Human 4*						
Mastadenovirus	*Human 11*						
Mastadenovirus	*Human 34/35*						
Mastadenovirus	*Simian 25*						
Mastadenovirus	*Human 14 (Put.)*						
Mastadenovirus	*Human 16 (Put.)*						
Mastadenovirus	*Human 21 (Put.)*						
Mastadenovirus	*Human 41*						
Mastadenovirus	*Human 40*						
Mastadenovirus	*Human 8*						
Mastadenovirus	*Human 12*						
Mastadenovirus	*Human 7h*						
Mastadenovirus	*Canine ½**						
Mastadenovirus	*Porcine 3*						
Mastadenovirus	*Porcine 5*						
Mastadenovirus	*Porcine 1*						
Mastadenovirus	*Canine 1***						
Mastadenovirus	*Bovine 1*						
Mastadenovirus	*Bovine 3*						
Mastadenovirus	*Porcine 4*						
Mastadenovirus	*Bovine 10*						
Mastadenovirus	*Mouse 1*						
Mastadenovirus	*Bovine 2*						
Mastadenovirus	*Bovine 4*						
Atadenovirus	*Ovine 7*						
Atadenovirus	*Duck 1*						
Aviadenovirus	*Fowl 8*						
Aviadenovirus	*Fowl 1*						
Aviadenovirus	*Fowl 10*						
Siadenovirus	*Turkey 3*						
Siadenovirus	*Frog 1*						
TOTAL	***36***	***5***	***32***	***3***	***28***	***33***	***1***

3.3.1.2 Secondary and Tertiary Structure Modelling

Following analysis of human Ad2 pVIII at the primary level, the putative secondary structure of the protein was investigated using the HNN Secondary Structure Prediction Method at Pole Bioinformatique, Lyonnais (Guermeur., 1997; http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa.nn.html). The sequence was predicted to be composed of 66 % random coil, 14% α-helix and 20% β-strand (Figure 3.3.3).

Subsequently the First Approach Mode at SwissModel (Method 2.1.2.3; http://swissmodel.expasy.org) was used to compare the primary sequence of human Ad2 pVIII polypeptide to all crystallised proteins in its databank. However no protein similar to pVIII had been crystallised and therefore a tertiary structure model could not be built.

Figure 3.3.3: Secondary Structure Prediction for Human Ad2 pVIII
The putative secondary structure of the protein was investigated using the HNN Secondary Structure Prediction Method at Pole Bioinformatique, Lyonnais (Guermeur., 1997; http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa.nn.html). The sequence was predicted to be composed of 66 % random coil (highlighted by a green 'c'), 14% α-helix (highlighted by a blue 'h') and 20% β-strand (highlighted by a red 'e').

```
MSKEIPTPYMWSYQPQMGLAAGAAQDYSTRINYMSAGPHMISRVNGIRAHRNRILLEQAA  60
ccccccccccccccccccccccchhchceeeeeeccccchheecchhhhhhhehhhhe

ITTTPRNNLNPRSWPAALVYQESPAPTTVVLPRDAQAEVQMTNSGAQLAGGFRHRVRSPG  120
ecccccccccccccccceeeeccccccceeeeccccchhhhhhccchhccccceeeeccc

QGITHLKIRGRGIQLNDESVSSSLGLRPDGTFQIGGAGRSSFTPRQAILTLQTSSSEPRS  180
ccceeeeeccccceeeccccccccccccccceeecccccccccccceeeeeeccccccc

GGIGTLQFIEEFVPSVYFNPFSGPPGHYPDQFIPNFDAVKDSADGYD              227
ccccceehhhhhcccceeeccccccccccccccccccccchccccccc
```

3.3.1.3 Discussion

The human Ad2 precursor protein, pVIII is thought to be cleaved by Avp2 to produce polypeptide, VIII. The precursor protein contains three putative Avp2 cleavage sites between residues 111-112, 131-132 and 157-158. However it is unknown whether one, two or three of these sites are required to be cleaved to form mature VIII. Therefore, in this section, 36 pVIII proteins were aligned to analyse the conservation of these sites.

The results showed that all three putative cleavage sites were well conserved throughout the Mastadenovirus genus. However none of them were completely conserved and a number of species contained putative cleavage sites out-with the main three (Table 3.3.1). Therefore defining one cleavage site as being more important than another was impossible and the results suggested that all three cleavage sites may be required in the Mastadenovirus genus.

The proposal that more than one cleavage site is required was further strengthened by an examination of the Atadenovirus and Aviadenovirus genera sequences. Atadenovirus although represented by only two species contained examples of the three cleavage sites. Likewise the three members of the Aviadenovirus genus contained all three cleavage sites as well as two other peripheral cleavage sites.

In contradiction to the theory that all three sites are essential, the Mastadenovirus and Siadenovirus pVIII proteins from bovine Ad4 and frog Ad1, respectively both contained only one cleavage site. However it must be borne in mind that this work did not utilise the full cleavage motifs for Avp but instead used the slightly more restrictive motifs of Avp2. This may have resulted in some of the cleavage sites of the bovine Ad4 and frog Ad1 proteins being omitted. In addition as the general cleavage motifs for Avp develop it may be found that the bovine Ad4 and frog Ad1 proteins contain other as yet unidentified cleavage sites.

The alignment therefore provided no definitive conclusions as to which cleavage sites were essential for the production of mature pVIII. Moreover, there was conflicting evidence over the number and location of cleavage sites required, because Siadenovirus genus contained only two peripheral cleavage sites which were not in a comparable position to any of the three major human Ad2 sites. This suggested that the Siadenovirus genus may not share the same cleavage requirement as Mastadenovirus genus.

Therefore although we began with one simple question, we are now left with three. Which site(s) are essential for the production of mature VIII? Are the other sites redundant and if not what is their function(s)? Do each of the genera have their own requirement for cleavage?

An investigation into the secondary and tertiary structures of human Ad2 pVIII was equally as elusive, indicating that the protein was composed mainly of random coil and shared no similarity to any other crystallised protein in the SWISS-PROT databank.

As this was the case, the next section of this work aimed to express and purify recombinant pVIII protein, firstly to analyse its cleavage by Avp2 *in vitro* and secondly to elucidate its structure using crystallography.

3.3.2 Cloning, Expression and Purification of Tagged pVIII

As discussed in the previous section, there were two reasons for purifying recombinant pVIII. Firstly, to produce enough soluble pVIII to be used in crystal trials, and secondly to analyse the cleavage of pVIII by Avp2.

Previous experiments in our laboratory aimed at purifying pVIII had used the vectors pRSETA (Invitrogen), pMALp2E (NEB) and pMALc2E (NEB) (Annan., 1999). Unfortunately the pRSETA vector produced insoluble His tagged protein, the pMALp2E vector produced very small amounts of Maltose tagged protein and although the pMALc2E vector produced soluble pVIII, the protein did not bind the amylose column. However, with new techniques it was hoped that well expressed, soluble protein could be produced and a purification method could be set in place. Several systems were investigated and these are discussed in the following results section.

3.3.2.1 pVIIIGEX4Tev

pVIII was first cloned into the GST tag vector pGEX4Tev (a gift from Mr. B. Precious; Appendix I.B), which is a modified pGEX4-T-3 vector. The vector had its Thrombin cleavage site removed and replaced with a Tev cleavage site. In addition the MCS was also modified. This vector was chosen because GST tags are well established for improving solubility of proteins and the Tev protease is readily available and inexpensive.

3.3.2.1.1 Cloning and Expression

A mini-prep (Method 2.2.7) was made of pVIIIpRSETA vector and this was used as a template for the PCR reaction (Method 2.2.1.1). The forward and reverse primers in Table 3.3.2 were used in the PCR. The PCR reaction was carried out at 94°C for 5min (hot start), 94°C for 30 sec, 55°C for 60sec, 72°C for 60sec, 72°C for 7min over 25 cycles.

Table 3.3.2: Primers Used to Clone pVIII in to the pGEX4Tev Vector
The PCR reaction for cloning pVIII into the pGEX4Tev vector used a forward and reverse primer.

Primer	Oligonucleotide primer sequence (5'-3')
Forward	AGGACCATGGCCATGAGCAAGGAAATTCCCACG
Reverse	ATACTCGAGTTATCAGTCGTAGCCGTCCGCCGA

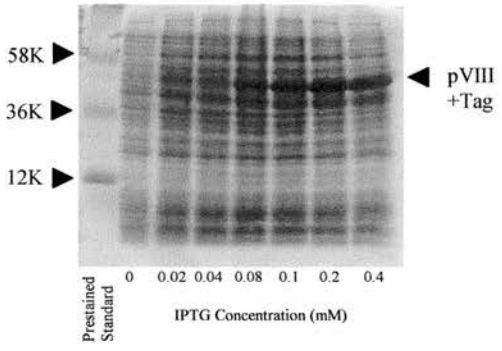
The vector and insert were then double digested by NcoI and XhoI (Method 2.2.4), ligated (Method 2.2.5), used to transform the JM109 cell line (Promega) (Method 2.2.6) and left to grow on an LB-agar/Carbenicillin plate at 37°C for 12-16hr (Method 2.2.10.1). Six colonies were selected from the transformants and a mini-prep was carried out on each. 10µl of each mini-prep was double digested by NcoI and XhoI and then separated by Agarose gel electrophoresis (Method 2.2.2). Finally one of the colonies which appeared to be positive according to the double digestion was identified by DNA sequencing (Method 2.2.9).

After confirmation by DNA sequencing, a glycerol stock (Method 2.2.10.2) was made of the JM109 cell line. The vector was then transformed into the BL21(DE3) cell line (Novagen) and a glycerol stock was made of these.

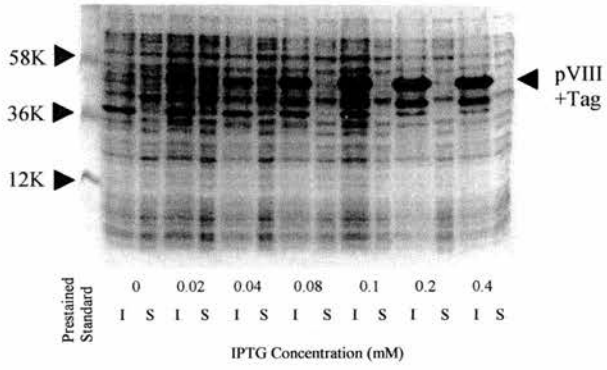
An LB-agar/Carbenicillin plate was made from the glycerol stock and a single colony was selected to analyse the Overall/Soluble expression of pVIII (Method 2.4.4.1). The cells were grown in LB/Carbenicillin and then re-suspended in 20mM Tris-HCl, pH8. Unfortunately, the results of the expression and solubility test (Figure 3.3.4) showed that although the pVIII was expressed very well in this vector, it was mainly found in the insoluble fraction.

Figure 3.3.4: Expression of pVIIIGEX4Tev
pVIII was cloned into the pGEX4Tev vector and expressed in the BL21(DE3) cell line. 3ml of cells grown to log phase were induced with 0, 0.02, 0.04, 0.08, 0.1, 0.2 and 0.4mM IPTG for 6hrs at 25°C. Two 1ml fractions of the cells were removed and centrifuged to form a cell pellet. One pellet was loaded directly onto a 15% SDS-PAGE mini-gel, which was stained with Coomassie Blue, whilst the other pellet was sonicated before loading. These samples were then used to analyse the (a) overall and (b) insoluble(I)/soluble(S) expression levels respectively.

(a) Overall Expression



(b) Insoluble/Soluble Expression



3.3.2.2 pVIIIRSETA

As the GST tag had produced insoluble protein (Section 3.3.2.1.1), the His tag vector, pRSETA (Invitrogen) was investigated. pVIII was previously cloned into the vector the BamHI site by Dr. W.S. Annan (1999).

Prior to large scale purification, a small scale experiment was carried out to analyse the soluble expression levels of pVIII, and whether or not the pVIII-His could bind Nickel beads. Following this three purification methods were investigated.

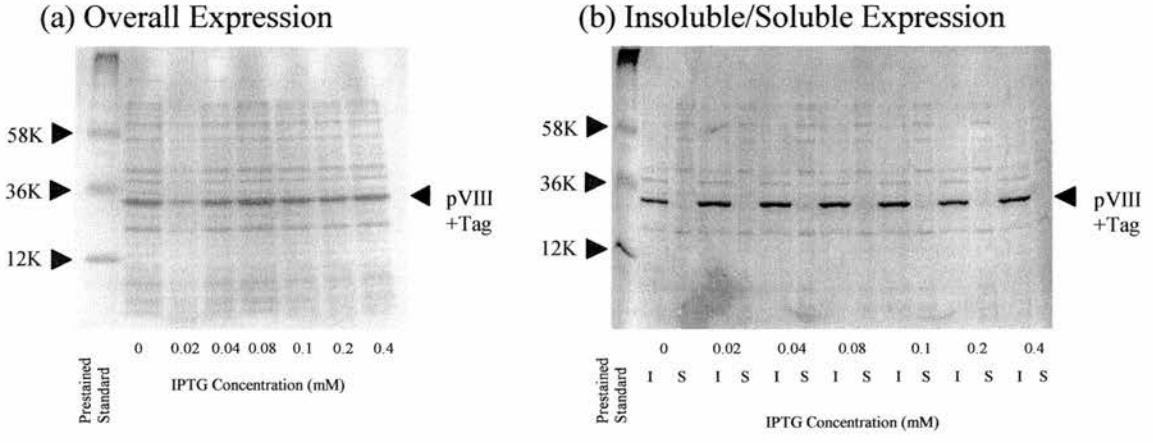
3.3.2.2.1 Expression

pVIIIpRSETA was previously grown to an OD_{600} = 0.4-0.6 and then induced for 2h at 37°C with 0.4mM IPTG. However, the expressed protein was found to be mainly

insoluble. Since this time, it has became common practice to reduce induction temperatures and IPTG concentrations to improve solubility. Therefore, the culture was induced with 0, 0.02, 0.04, 0.08, 0.1, 0.2 and 0.4mM IPTG for 6hrs at 25°C according to Method 2.4.4.1.

An LB-agar/Carbenicillin plate (Method 2.2.10.1) was made from the glycerol stock of pVIIIpRSETA expressed in the BL21(DE3) cell line and a single colony was selected to analyse the overall/soluble expression of pVIII (Method 2.4.4.1). The cells were grown in LB/Carbenicillin, and then re-suspended in Column buffer (20mM Tris-HCl, 10mM Imidazole, pH8) according to Method 2.4.4.1. The protein was found to express well but was found mainly in the insoluble fraction (Figure 3.3.5).

Figure 3.3.5: Expression of pVIIIRSETA
pVIIIRSETA was expressed in the BL21(DE3) cell line. 3ml of cells grown to log phase were induced with 0, 0.02, 0.04, 0.08, 0.1, 0.2 and 0.4mM IPTG for 6hrs at 25°C. Two 1ml fractions of the cells were removed and centrifuged to form a cell pellet. One pellet was loaded directly onto a 15% SDS-PAGE mini-gel, whilst the other pellet was sonicated before loading. These samples were then used to analyse the (a) overall and (b) insoluble(I)/soluble(S) expression levels respectively.



3.3.2.2.2 Analysing Binding to Nickel Beads

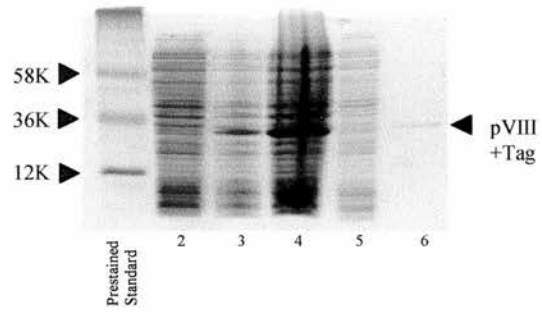
pVIII expressed in the pRSETA vector produced mainly insoluble protein (Section 3.3.2.2.1) . However, it is often found that proteins expressed with His tags produce enough soluble material to be purified using a Nickel Column, as Nickel beads can facilitate the purification of small amounts of protein.

Therefore prior to large scale purification, a small scale experiment was carried out to investigate whether a proportion of the expressed protein was soluble and could bind Nickel beads (Method 2.5.1).

5ml of culture was grown at 25°C for 12-16hr LB/Carbenicillin and then induced with 0.4mM IPTG at 25°C for 6hr and the buffer used for re-suspension was Column buffer (20mM Tris-HCl, 10mM Imidazole, pH8). Samples taken were pre-induction cell pellet; post-induction cell pellet; post-sonication cell pellet (insoluble fraction); post-sonication supernatant (soluble fraction) and the protein bound to Nickel beads (washed with Wash buffer (20mM Tris-HCl, 30mM Imidazole, pH8)). These samples were separated on a 15% SDS-PAGE mini-gel (Method 2.4.1.1) which was stained with Coomassie Blue (Method 2.4.2.1) and the results are shown in Figure 3.3.6.

The results showed that the expressed protein was found mainly in the cell pellet after sonication (Lane 4) but there was some protein capable of binding the Nickel beads (Lane 6). The band excised from the gel was identified as human Ad2 pVIII by Mass Spectrometry (Method 2.4.7.2).

Figure 3.3.6: Binding of pVIIIRSETA to Nickel Beads
pVIIIRSETA expressed in the BL21(DE3) cell line produced mainly insoluble pVIII protein. Therefore to analyse its ability to bind Nickel beads, 5ml of culture was grown at 25°C for 12-16hr and then induced with 0.4mM IPTG at 25°C for 6hr. Samples taken were, pre-induction cell pellet **(Lane 2)**; post-induction cell pellet **(Lane 3)**; post-sonication cell pellet (insoluble fraction) **(Lane 4)**; post-sonication supernatent (soluble fraction) **(Lane 5)** and the protein bound to the Nickel beads **(Lane 6)**. These samples were separated on a 15% SDS-PAGE mini-gel which was stained with Coomassie Blue.



3.3.2.2.3 Purification Method 1

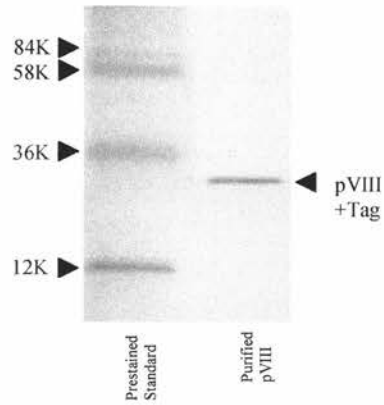
As the His tagged pVIIIRSETA protein could bind the Nickel Beads, the first purification method to be investigated involved leaving the protein attached to the Nickel beads. Although this type of purification was inappropriate for crystal trials, it was hoped that it may be useful for analysing cleavage of pVIII by Avp2.

pVIIIpRSETA expressed in the BL21(DE3) cell line was grown in LB/Carbenicillin according to Method 2.5.2. The cells were then induced with 0.4mM IPTG for 6hr at 25°C (Method 2.5.2). Following this they were extracted (Method 2.5.3.2) and then re-suspended in 8ml of Column buffer (20mM Tris-HCl, 10mM Imidazole, pH8).

5ml of Nickel beads were washed with 3 x 10ml water and then 3 x 10ml Column buffer. The cell extract was then applied to the beads and the beads/cell extract slurry was gently agitated at 25°C for 30min. The beads were then washed with 3 x 10ml Wash buffer (20mM Tris-HCl, 30mM Imidazole, pH8). Subsequently the beads were re-suspended in

500µl Wash buffer and the sample was aliquoted into 20µl fractions and stored at –70°C. Figure 3.3.7 shows the pVIII purified by this method, separated on a 15% SDS-PAGE mini-gel (Method 2.4.1.1) which was stained with Coomassie Blue (Method 2.4.2.1). The band was confirmed to be pVIII by Mass Spectrometry (Method 2.4.7.2). However when the gel was silver stained (Method 2.4.2.2) a number of contaminating bands appeared. These contaminants made cleavage analysis impossible.

Figure 3.3.7: pVIIIRSETA Purified by Attaching to Nickel Beads
pVIIIRSETA expressed in the BL21(DE3) cell line was purified by leaving the protein attached to Nickel Beads.



3.3.2.2.4 Purification Method 2

Impure pVIII was produced by leaving the protein attached to Nickel beads (Section 3.3.2.2.3). This method of purification was therefore unacceptable for analysing the cleavage of pVIII, thus Purification Method 2 aimed to purify pVIII on a Nickel column, remove the His tag and then further purify by re-loading onto the Nickel column.

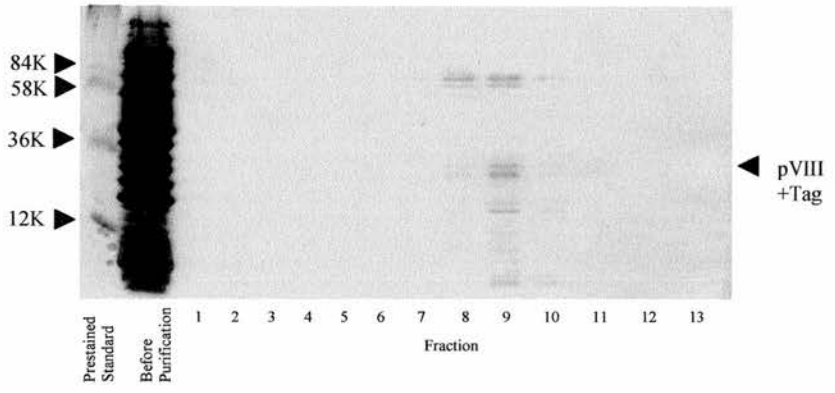
pVIIIpRSETA expressed in the BL21(DE3) cell line was grown in LB/Carbenicillin according to Method 2.5.2. The cells were then induced with 0.4mM IPTG for 6hr at 25°C

(Method 2.5.2). Following this they were extracted (Method 2.5.3.2) and then re-suspended in 8ml of Column buffer (20mM Tris-HCl, 10mM Imidazole, pH8).

A 1 x 5cm Nickel column was washed for 15min with water at a flow rate of 1ml/min and then further washed for 15min with Column buffer. Following this, the cell extract was applied to the column and then further washed for 15min in the presence of Wash buffer ((20mM Tris-HCl, 30mM Imidazole, pH8)). pVIIIRSETA was then eluted into 1ml fractions for 15min at 1ml/min with Elution buffer (20mM Tris-HCl, 250mM Imidazole, pH8).

The fractions were separated on a 15% SDS-PAGE mini-gel, which was stained with Coomassie Blue and as Figure 3.3.8 shows, the protein was eluted between fractions 8-11. The pVIII fusion protein band was identified by Mass Spectrometry.

Figure 3.3.8: Nickel Bead Purification of pVIIIRSETA
The pVIIIRSETA protein solution was applied to a 1 x 5cm Nickel column and was eluted in the presence of 250mM Imidazole. The protein was eluted into 1ml fractions and these fractions were separated on a 15% SDS-PAGE mini-gel, which was stained with Coomassie Blue. The pVIII band was confirmed by Mass Spectrometry.

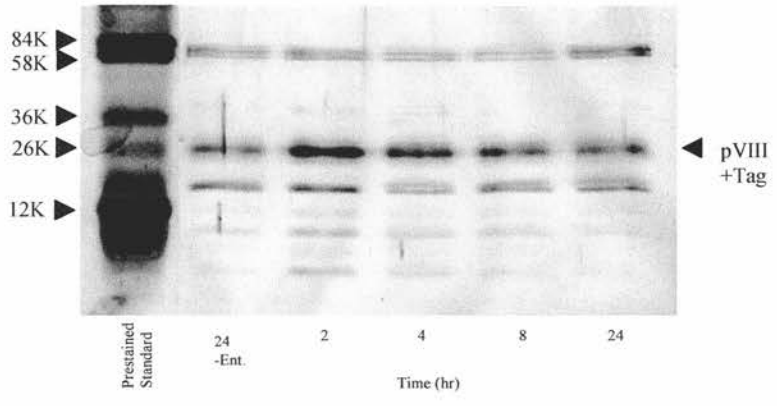


Fractions 8-11 were pooled and dialysed against 1L of 20mM Tris-HCl, pH8 for 4hr with two changes of buffer. Following dialysis, the protein solution was diluted to 5mg/ml

and the tag was cleaved from pVIIIRSETA over a time period. 1ng of Enterokinase was added to the solution and the solution was left to digest with gentle agitation at 25°C for 2hr, 4hr, 8hr and 24hr. A control incubation in the absence of Enterokinase was also carried out (Method 2.5.6).

The incubated samples were analysed using a 15% SDS-PAGE mini-gel, which was Silver stained (Method 2.4.2.2) and the results are shown in Figure 3.3.9. There was no cleavage of the pVIII band even after 24hr and so this method only produced partially purified protein which could not have its tag removed.

Figure 3.3.9: Cleavage of the His Tag from pVIIIRSETA
Cleavage of the His tag from the purified pVIIIRSETA fusion protein was carried out over a 24hr period at 25°C in the presence of 1ng of Enterokinase added to a 100µl sample. 20µl aliquots (including a control, which was left for 24hr and had no Enterkinase added) were taken at 2hr, 4hr, 8hr and 24hr. The samples were then separated on a 15% SDS-PAGE mini-gel, which was Silver stained.



3.3.2.2.5 Purification Method 3

As the His tag could not be cleaved from pVIII (Section 3.3.2.2.4), this meant that pVIII could only be partially purified and therefore a further purification step was required.

Thus, the fractions eluted from the Nickel column (Section 3.3.2.2.4), containing pVIIIRSETA were pooled and dialysed against 1L of Column buffer (20mM Tris-HCl, pH8) for 4hr with two changes of buffer to remove Imidazole. The sample was then applied to a 1 x 30cm DEAE column at 1ml/min for 1hr in the presence of Column buffer and 3ml fractions were collected. Following this, a solution of Elution buffer (20mM Tris-HCl, 1M NaCl, pH8) was passed through the column for 1hr at 1ml/min, to remove any residual proteins and these fractions were also collected.

All of the fractions were separated on a 15% SDS-PAGE mini-gel (Method 2.4.1.1) and the gel was Silver stained (Method 2.4.2.2). However, a band corresponding to pVIII could not be identified and it was assumed to be too low in yield to be visualised.

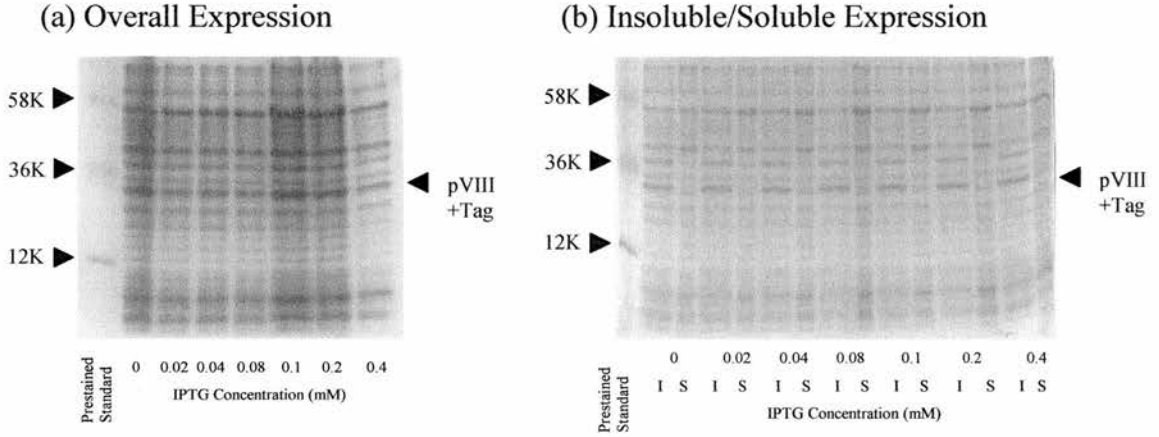
3.3.2.2.6 Co-expressing pVIIIRSETA and GroELGroESET29c

As all attempts at purifying large quantities of His tagged pVIIIRSETA had been unsuccessful, pVIIIpRSETA was co-expressed with the GroELGroESpET29c vector in the BL21(DE3) cell line. It was hoped that the over-production of the GroEL and GroES Chaperone proteins would allow the refolding of the insoluble His tagged pVIII.

The procedure required the BL21(DE3) (Novagen) cell line already expressing pVIIIpRSETA to be transformed (Method 2.2.6) with the GroELGroESpET29c vector. Transformed cells were grown on the selective media LB-agar/Carbenicillin/Spectinomycin plate (Method 2.2.10.1) at 37°C for 12-16hr, as the pRSETA and GroELGroESpET29c vectors were Carbenicillin and Spectinomycin resistant, respectively. From this plate a single colony was selected to analyse the overall/soluble expression of pVIII (Method 2.4.4.1). The cells were grown in LB/Carbenicillin/Spectinomycin, and then re-suspended in Column buffer (20mM Tris-HCl, 10mM Imidazole, pH8) according to Method 2.4.4.1.

The results of the expression and solubility test (Figure 3.3.10) showed that although pVIII was expressed very well in this system, it was not refolded and still remained in the insoluble fraction.

Figure 3.3.10: Co-expression of pVIIIRSETA and GroELGroESET29c
pVIIIpRSETA was co-expressed in the BL21(DE3) cell line along with the GroELGroES-pET29c heat shock vector, which produced the Chaperone proteins, GroEL and GroES. 3ml of cells grown to log phase were induced with 0, 0.02, 0.04, 0.08, 0.1, 0.2 and 0.4mM IPTG for 6hrs at 25°C. Two 1ml fractions of the cells were removed and centrifuged to form a cell pellet. One pellet was loaded directly onto a 15% SDS-PAGE mini-gel, which was stained with Coomassie Blue, whilst the other pellet was sonicated before loading. These samples were then used to analyse the (a) overall and (b) insoluble(I)/soluble(S) expression levels respectively.



3.3.2.3 pVIIILou3

As the GST and His tags had not produced purified pVIII, the next tag to be investigated was a Maltose tag. Therefore, pVIII was cloned into the His/Maltose tag vector pLou3 (a gift from Dr. L. Major; Appendix I.C), which was a modified pMALc2X vector (NEB). The vector had its MCS replaced, the recognition site for the Tev protease was placed after the Maltose tag to allow cleavage of the tag, and a His tag was placed before the Maltose tag to allow purification. The two major advantages of using the vector were that, firstly pVIII had previously been shown to be soluble when tagged to a Maltose

tag (Annan., 1999) and secondly the His tag/Tev protease system allowed the use of a single column in the purification process.

3.3.2.3.1 Cloning and Expression

A mini-prep (Method 2.2.7) was made of pVIIIpRSETA vector and this was used as a template for the PCR reaction (Method 2.2.1.1). The forward and reverse primers in Table 3.3.3 were used in the PCR. Although the pVIII sequence contained its own stop codon at the C-terminus, this isn't a very reliable stop codon and so another stop codon (TAA) was placed after it in the reverse primer.

The PCR reaction was carried out at 94°C for 5min (hot start), 94°C for 30 sec, 55°C for 60sec, 72°C for 45sec, 72°C for 7min over 25 cycles.

Table 3.3.3: Primers Used to Clone pVIII in to the pLou3 Vector
The PCR reaction for cloning pVIII into the pLou3 vector used a forward and reverse primer.

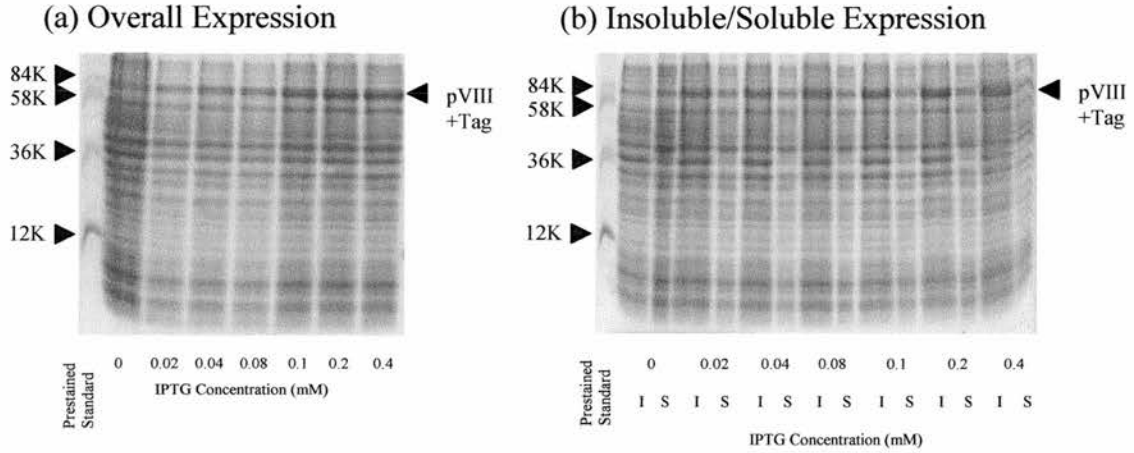
Primer	Oligonucleotide primer sequence (5'-3')
Forward	ATAGCCATGGATATGAGCAAGGAAATTCCCACG
Reverse	ATAGGATCCTTATCAGTCGTAGCCGTCCGCCG

The vector and insert were then double digested by NcoI and BamHI (Method 2.2.4), ligated (Method 2.2.5), used to transform the JM109 cell line (Promega) (Method 2.2.6) and left to grow on an LB-agar/Carbenicillin plate (Method 2.2.10.1) at 37°C for 12-16hr. Six colonies were selected from the transformants and a mini-prep was carried out on each. 10µl of each mini-prep was double digested by NcoI and BamHI and then separated by Agarose gel electrophoresis (Method 2.2.2). Finally one of the colonies which appeared to be positive according to the double digestion was identified by DNA sequencing (Method 2.2.9).

After confirmation by DNA sequencing, a glycerol stock (Method 2.2.10.2) was made of the JM109 cell line. The vector was then transformed into the BL21(DE3) cell line (Novagen) and a glycerol stock was made of these.

An LB-agar/Carbenicillin plate was made from the glycerol stock and a single colony was selected to analyse the overall/soluble expression of pVIII (Method 2.4.4.1). The cells were grown in LB/Carbenicillin and then re-suspended in Column buffer (20mM Tris-HCl, 500mM NaCl, 10mM Imidazole, pH8). The results of the expression and solubility test (Figure 3.3.11) showed that pVIII was found in both the soluble and insoluble fractions.

Figure 3.3.11: Expression of pVIIILou3
pVIIIpLou3 was expressed in the BL21(DE3) cell line. 3ml of cells grown to log phase were induced with 0, 0.02, 0.04, 0.08, 0.1, 0.2 and 0.4mM IPTG for 6hrs at 25°C. Two 1ml fractions of the cells were removed and centrifuged to form a cell pellet. One pellet was loaded directly onto a 15% SDS-PAGE mini-gel, which was stained with Coomassie Blue, whilst the other pellet was sonicated before loading. These samples were then used to analyse the (a) overall and (b) insoluble(I)/soluble(S) expression levels respectively.

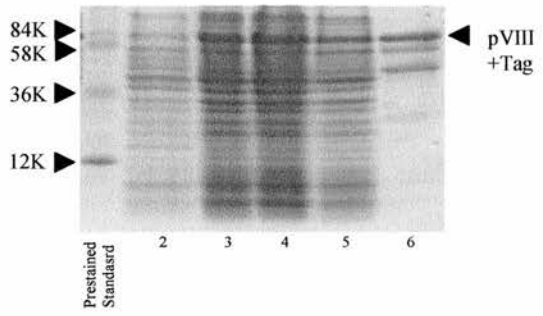


3.3.2.3.2 Analysing Binding to Nickel Beads

The pVIIIpLou3 vector expressed soluble His/Mal tagged pVIIILou3 (Section 3.3.2.3.1). Therefore prior to large scale purification, a small scale experiment was carried out to investigate whether the protein could bind Nickel beads (Method 2.5.1).

5ml of culture was grown at 25°C for 12-16hr in LB/Carbenicillin and then induced with 0.4mM IPTG at 25°C for 6hr and the buffer used for re-suspension was Column buffer (20mM Tris-HCl, 500mM NaCl, 10mM Imidazole, pH8). Samples taken were pre-induction cell pellet; post-induction cell pellet; post-sonication cell pellet (insoluble fraction); post-sonication supernatant (soluble fraction) and the protein bound to Nickel beads (washed with Wash buffer (20mM Tris-HCl, 500mM NaCl, 30mM Imidazole, pH8)) (Method 2.5.1). These samples were separated on a 15% SDS-PAGE mini-gel (Method 2.4.1.1), which was stained with Coomassie Blue (Method 2.4.2.1) and the results are shown in Figure 3.3.12. The results showed that the expressed protein was found in the supernatant after sonication (Lane 5) and the protein was capable of binding Nickel beads (Lane 6). The band was confirmed to be pVIII by Mass Spectrometry (Method 2.4.7.2).

Figure 3.3.12: Binding of pVIIILou3 to Nickel Beads
5ml of culture was grown at 25°C for 12-16hr and then induced with 0.4mM IPTG at 25°C for 6hr. Samples taken were, pre-induction cell pellet **(Lane 2)**; post-induction cell pellet **(Lane 3)**; post-sonication cell pellet (insoluble fraction) **(Lane 4)**; post-sonication supernatant (soluble fraction) **(Lane 5)** and the protein bound to the Nickel beads **(Lane 6)**. These samples were separated on a 15% SDS-PAGE mini-gel, which was stained with Coomassie Blue.

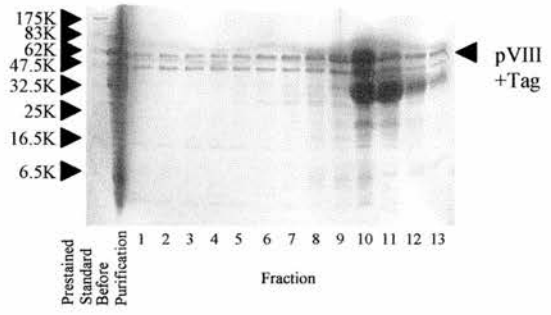


3.3.2.3.3 Purification Method 1

pVIIILou3 expressed in the BL21(DE3) cell line was grown in LB/Carbenicillin according to Method 2.5.2. The protein was then extracted (Method 2.5.3.2) and the cell extract was re-suspended in 50ml Column buffer (20mM Tris-HCl, 500mM NaCl, 10mM Imidazole, pH8). Following this, the protein was purified as has previously been described for Avp2Lou3 (Section 3.2.1.5.3).

The eluted fractions were separated on a 15% SDS-PAGE mini-gel (Method 2.4.1.1) which was stained with Coomassie Blue (Method 2.4.2.1) and the results are shown in Figure 3.3.13. The band highlighted in this figure was confirmed by Mass Spectrometry (Method 2.4.7.2) to be the His/Maltose tagged pVIII protein. Fractions 10-11 were pooled and dialysed against 20mM Tris-HCl, pH8 for 4hr with two changes of buffer.

Figure 3.3.13: Nickel Bead Purification of pVIIILou3
pVIIILou3 was purified on a 1.5 x 10cm Nickel column. The extract was applied to the column and the column was washed for 45min at 1ml/min in the presence of 20mM Tris-HCl, 500mM NaCl, 30mM Imidazole, pH8. Following this, the tagged Avp2 was eluted into 2ml fractions at 1ml/min for 45min with 20mM Tris-HCl, 500mM NaCl, 150mM Imidazole, pH8. The fractions were separated on a 15% SDS-PAGE mini-gel which was stained with Coomassie Blue.

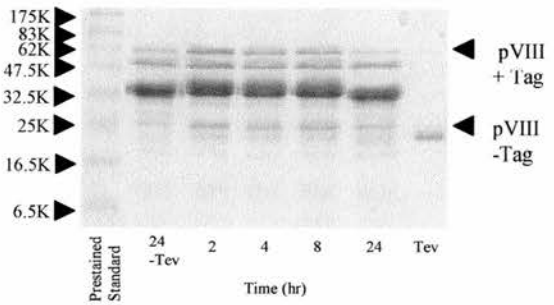


Following dialysis the concentration of the protein solution was calculated to be 30mg/ml, by the Bradford Assay (Method 2.4.5.2). Therefore prior to cleavage of the His/Mal tag by the Tev protease, the solution was diluted to 5mg/ml and the tag was

cleaved from pVIII over a time period. 1μg of Tev protease was added to solution and the solution was left to digest with gentle agitation at 25°C for 2hr, 4hr, 8hr and 24hr. A control incubation without Tev was also carried out (Method 2.5.6).

The cleavage products were separated on a 15% SDS-PAGE mini-gel which was stained with Coomassie Blue (Figure 3.3.14) and the cleaved pVIII band was confirmed by Mass Spectrometry. This system was scaled up for large scale cleavage.

Figure 3.3.14: Cleavage of the His/Maltose Tag from pVIIIpLou3
Cleavage of the tag from the His/Maltose tagged pVIIILou3 was carried out over a 24hr period at 25°C in the presence of 1μg of Tev Protease in 100μl of sample. 20μl aliquots (including a control, which was left for 24hr in the absence of Tev) were taken at 2hr, 4hr, 8hr and 24hr. A 15% SDS-PAGE mini-gel stained with Coomassie Blue was used to analyse these samples.



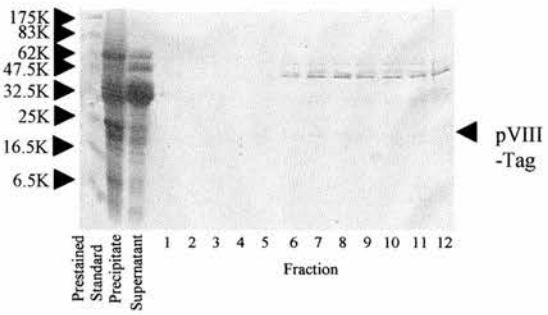
The large scale cleavage was left for 12hr at 25°C, with gentle agitation. However, after incubation, it was noted that a large quantity of the sample had precipitated (Figure 3.3.15), even though two measures had been set in place to reduce this occurring. These measures involved diluting the concentration of the protein sample from 30mg/ml to 5mg/ml, and decreasing the temperature of cleavage from 37°C to 25°C. It was thought that reducing these factors would allow slower cleavage, therefore giving the cleaved protein a greater amount of time to refold, thus diminishing the amount of precipitation.

Reducing these factors may have had some effect, because a portion of protein which had remained soluble was cleaved by the Tev protease (Figure 3.3.15). Therefore, after cleavage the sample was applied to a 1.5cm x 10cm to Nickel column to remove the contaminating tagged pVIII protein, the His tagged Tev Protease and any other proteins which had previously bound the Nickel column.

Prior to loading the sample onto the column, the sample was centrifuged at 5,000rpm for 10min in a Howe Sigma Laborzentrifugen 3K10 centrifuge, to remove the precipitant. The sample was then applied to the column in the presence of 20mM Tris-HCl, pH8 at a rate of 1ml/min, and 2ml samples were collected for 60min.

These samples were then separated on a 15% SDS-PAGE mini-gel which was stained with Coomassie Blue and as Figure 3.3.15 shows, the protein was eluted between fractions 7 to 11. The cleaved pVIII band was again identified by Mass Spectrometry. Nevertheless, the method only produced a small amount of pVIII protein which was contaminated by a number of other proteins.

Figure 3.3.15: Secondary Nickel Bead Purification of pVIIIpLou3
After cleavage, the precipitated protein was removed by centrifugation, and the protein sample was applied to a 1.5 x 10cm Nickel column in the presence of 20mM Tris-HCl, pH8 at a rate of 1ml/min, and 2ml samples were collected for 60min. The samples were then separated on a 15% SDS-PAGE mini-gel which was stained with Coomassie Blue, and the protein was eluted between fractions 7 to 11. The cleaved pVIII band was identified by Mass Spectrometry. Nevertheless, the method only produced a small amount of pVIII protein which was contaminated by a number of other proteins.

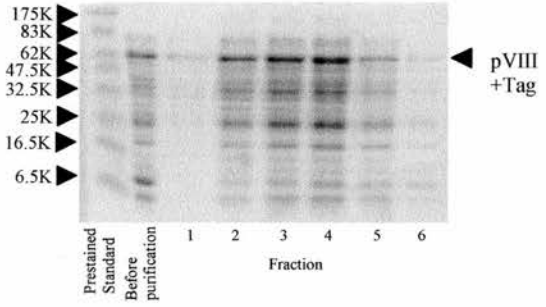


3.3.2.3.4 Purification Method 2

As pVIII mainly precipitated when the His/Maltose tag was removed (Section 3.3.2.3.3) it was decided to retain the tag and to further purify the protein on a Carboxymethyl(CM)-sepharose column. This column was chosen as the theoretical pI of the protein construct was 5.83. Therefore the protein was negatively charged at pH8 and was expected to pass through the column.

pVIIILou3 was grown, extracted and purified as discussed in Section 3.3.2.3.3 except the tag was not removed and the second Nickel bead purification was not carried out. Following purification the samples were applied to a 1 x 3cm CM-Sepharose column for 1hr at 1ml/min in the presence of 20mM Tris-HCl, pH8, with 2ml fractions being collected. All of the fractions were then separated on a 15% SDS-PAGE mini-gel (Method 2.4.1.1) which was stained with Coomassie Blue (Method 2.4.2.1), and as Figure 3.3.16 shows the pVIII protein was eluted between fractions 1-6. However the samples still remained contaminated and were therefore useless for cleavage studies by Avp2.

Figure 3.3.16: CM Column Purification of pVIIILou3
The tagged pVIII purified using a Nickel column the fractions were pooled and dialysed against 20mM Tris-Hcl, pH8. The samples were then applied to a 1 x 3cm CM-sepharose column for 1hr at 1ml/min in the presence of 20mM Tris-HCl, pH8 and the fractions were then separated on a 15% SDS-PAGE mini-gel.



3.3.2.4 Discussion

pVIII of the *Adenoviridae* family has never been purified under native conditions and therefore that was the aim of this Section. As has previously been discussed for tagged Avp2, the tagged pVIII proteins were expressed in *E. coli* and were induced using both a low concentration of IPTG and a low temperature of induction, which were hoped to reduce protein synthesis thus allowing the protein to be folded correctly.

Three vectors were investigated in this work, the first of which was a GST vector which surprisingly produced insoluble protein. Thus the His tag vector, pRSETA was investigated but the protein was again found to be insoluble even when co-expressed with chaperone proteins. However a small amount of soluble protein was capable of binding Nickel beads therefore leading to three purification methods. The first method involved leaving the protein attached to the beads in the hope that this could be used in a cleavage analysis, but the protein sample was contaminated therefore making the analysis impossible. The second method involved purification using a Nickel column but subsequently the His tag could not be removed, leading to the third and final purification method. This method again used a Nickel column followed by a DEAE column but the protein was purified in low concentrations.

The third vector investigated, pLou3 (His/Maltose tag followed by a Tev cleavage site), produced soluble protein which was capable of binding Nickel beads. However, removal of the His/Maltose caused the protein to precipitate. Therefore a second purification method using a Nickel column followed by a Carboxymethyl-Sepharose column which allowed the tag to be retained was investigate, but following purification the protein sample was found to be contaminated with several other proteins. Nevertheless, this was the first work to have purified non-denatured pVIII to any extent and it may lead to a complete purification method being developed whereby the His/Maltose tag is retained.

Regardless of the purification method being refined the partially purified His/Maltose tagged pVIII protein may serve as a substrate for Avp2 in future crystal trials. A purified fraction of pVIII containing a single cleavage site for Avp2 would be added to a reducing solution containing Avp2 and pVIcT. pVIII may then form a thiohemiacetal intermediate complex with Avp2 as has previously been shown for Ulp1-Smt3 (Mossessova & Lima., 2000). The complex could then be applied to a Nickel column to be purified prior to crystal trials.

One of the reasons for purifying non-denatured, recombinant pVIII was to examine the cleavage of pVIII by Avp2 *in vitro*. However because the purification systems were unsuccessful in this respect, the next Section discusses the use of a mammalian *in vitro* transcription-translation system and also the analysis of an adenovirus preparation to investigate this.

3.3.3 Maturation of pVIII

As human Ad2 pVIII had proven difficult to purify using an *E. coli* expression system it was decided to investigate cleavage of pVIII using a mammalian *in vitro* transcription-translation expression system and also to analyse an adenoviral preparation. The results of these investigations are discussed below.

3.3.3.1 pVIII in the TNT System

The two advantages of the mammalian *in vitro* transcription-translation expression system were that the protein should be correctly folded and the protein did not require purification.

The DNA for pVIII was first cloned into the pEGFP-N1 vector (Clontech). The pVIIIGFP DNA sequence was then sub-cloned into the pcDNA3.1(-) vector (Invitrogen; Appendix I.D) and this vector was used in the TNT system. The reason for tagging the protein with GFP at the C-terminus was that the proteins in this system are detected by radio-labelled Methionines and the C-terminal portion of pVIII did not contain any Methionines, whereas GFP did contain Methionine. Following cloning the pVIII-GFP protein was expressed in the TNT system and digestion by Avp2 was analysed using autoradiography.

3.3.3.1.1 Cloning pVIIIGFPpcDNA3.1(-)

A mini-prep (Method 2.2.7) was made of the pVIIIpRSETA vector and this was used as a template for the PCR reaction (Method 2.2.1.1). The forward and reverse primers in Table 3.3.4 were used in the PCR reaction. The reverse primer did not contain the C-terminal pVIII stop codon as this would not allow the polymerase to run through the

sequence to include the GFP tag. The PCR reaction was carried out at 94°C for 5min (hot start), 94°C for 30 sec, 55°C for 60sec, 72°C for 45sec, 72°C for 7min over 25 cycles.

Table 3.3.4: Primers Used to Clone Human Ad2 pVIII in to the pEGFP Vector.
The PCR reaction for cloning human Ad2 pVIII into the pEGFP vector used a forward and reverse primer.

Primer	Oligonucleotide primer sequence (5'-3')
Forward	CCCAAGCTTATGAGCAAGGAAATTCCC
Reverse	CGCGGATCCCGGTCGTAGCCGTCCGCCGAGTC

The vector and insert were then double digested by HindIII and BamHI (Method 2.2.4). Prior to ligation, the vector was treated with Calf Intestinal Alakaline Phosphatase (CIAP) (Method 2.2.5). The vector and insert were then ligated (Method 2.2.5), used to transform the JM109 cell line (Promega) (Method 2.2.6) and left to grow on an LB-agar/Kanamycin plate at 37°C for 12-16hr (Method 2.2.10.1). Six colonies were selected from the transformants and a mini-prep was carried out on each. 10µl of each mini-prep was double digested by HindIII and BamHI and then separated by Agarose gel electrophoresis (Method 2.2.2). Finally one of the colonies which appeared to be positive according to the double digestion was identified by DNA sequencing (Method 2.2.9). After confirmation by DNA sequencing, a glycerol stock (Method 2.2.10.2) was made of the JM109 cell line containing pVIIIpEGFP vector.

A mini-prep was then made of both the pVIIIpEGFP vector and the pCDNA3.1(-) vector. The pVIIIpEGFP vector was then double digested by XhoI and NotI. The pCDNA3.1(-) vector was digested with XhoI, for 1.5hr before NotI was added to the reaction mix. This was because XhoI and NotI sites are adjacent to each other in the pCNDA3.1(-) vector. Therefore XhoI was used first so that the cleavage site of NotI was

left to be digested. If digestion was carried out in the opposite order, then the NotI digestion would remove the XhoI restriction site. Following digestion, the vector was treated with CIAP.

The pVIIIGFP insert was ligated into the pcDNA3.1(-) vector. The ligated vector was then used to transform the JM109 cell line and left to grow on an LB-agar/Carbenicillin plate at 37°C for 12-16hr . Six colonies were selected from the transformants and a mini-prep was carried out on each. 10µl of each mini-prep was double digested by XhoI and NotI and then separated by Agarose gel electrophoresis. Finally one of the colonies which appeared to be positive according to the double digestion was identified by DNA sequencing. After confirmation by DNA sequencing, a glycerol stock was made of the JM109 cell line.

3.3.3.1.2 pVIII Cleavage by Avp2

The pVIIIGFPpcDNA3.1(-) vector was then expressed in a TNT T7 Rabbit Reticulocyte Quick Coupled Transcription/Translation Reaction (Promega), using Plasmid DNA and [^{35}S] Methionine. In this system, as the DNA was translated, the radiolabelled Methionine was incorporated into the sequence.

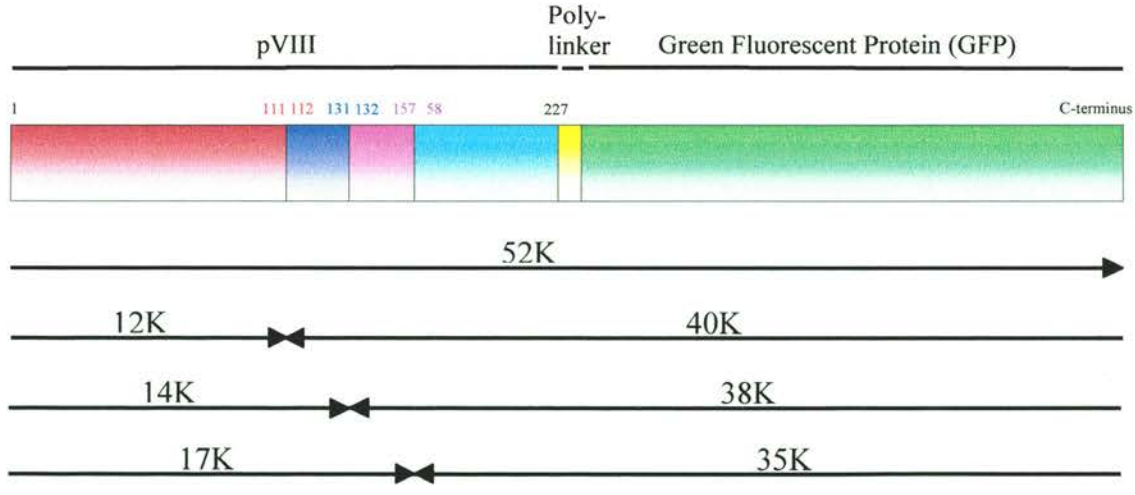
Following translation, the protein was cleaved by Avp2, separated on an SDS-PAGE gel, and the radiolabelled whole protein/cleaved products were visualised by autoradiography. As the pVIII protein and its cleavage products were visualised using this system, it was disadvantageous to clone pVIII alone into the pcDNA3.1 vector because if cleavage occurred at the expected sites, then only fragments 1-111, 1-131, 1-157 would be visualised on the gel. This was because Methionines were only found at residues 1, 10, 17, 34, 40 and 101 of pVIII (Figure 3.3.17). Therefore, pVIII was tagged with GFP at the C-terminus, prior to ligation into the pcDNA3.1- vector.

Figure 3.3.17: Availability of pVIII Methionines for Radiolabelling
The Methionines which may be radiolabelled in the TNT system are highlighted in bold at residues 1, 10, 17, 34, 40 and 101, whilst the potential cleavage sites for Avp2, are highlighted with red, blue and purple boxes.

```
MSKEIPTPYMWSYQPQMGLAAGAAQDYSTRINYMSAGPHMISRVNGIRAHRNRILLEQAA   60

ITTTPRNNLNPRSWPAALVYQESPAPTTVVLPRDAQAEVQMTNSGAQLAGGFRHRVRSPG  120

QGITHLKIRGRGIQLNDESVSSSLGLRPDGTFQIGGAGRSSFTPRQAILTLQTSSSEPRS  180

GGIGTLQFIEEFVPSVYFNPFSGPPGHYPDQFIPNFDAVKDSADGYD               227
```

The use of GFP at the C-terminus allowed the visualisation of the bands from 112-C-terminus, 132-C-terminus, and 158-C-terminus (Figure 3.3.18) using autoradiography. However there were still three small non-methionine containing internal regions from 112-131, 112-157 and 132-157, which could not be visualised.

Figure 3.3.18: Expected Cleavage Sizes for pVIIIGFP
A pVIIIGFP construct was inserted in to the pcDNA3.1- vector and this was expressed in a mammalian TNT system. Following expression, the protein was digested by Avp2. It was expected that the construct would be cleaved at three sites, between residues, 111-112, 131-132 and 157-158. If cleavage occurred at the consensus sites, then six fragments of sizes - 12K, 14K, 17K, 35K, 38K and 40K would appear. The derivation of these sizes is shown in the figure below.



The TNT reaction was set up in triplicate, with TNT Quick Master Mix, [^{35}S]Methionine and +/-plasmid DNA (Method 2.7.2). The samples were then incubated for 90min at 30°C (Method 2.7.2). Following incubation, five reactions were set up containing TNT reaction mix, buffer and either +Avp2/+pVIcT/+DNA, +Avp2/-pVIcT/+DNA, -Avp2/+pVIcT/+DNA, -Avp2/-pVIcT/+DNA, or -Avp2/-pVIcT/-DNA. These samples were then incubated for a further 30min at 37°C (Method 2.7.2).

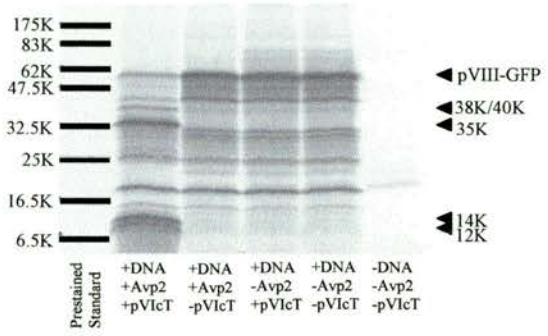
The samples were then separated on two 15% SDS-PAGE mini-gels (Method 2.4.1.1). The first gel was dried and then exposed to BioMax MR Film (Kodak). The film was then developed with a Kodak M35 X-OMAT Processor, this allowed the visualisation of the cleavage products from 1-111, 112-C-terminus, 1-131, 132-C-terminus, 1-157 and 158-C-terminus. The second gel was Western blotted using a GFP primary antibody. The Western blot was based on Method 2.4.3 except the GFP primary antibody used was a 1 in 1000 dilution and the anti-mouse IgG secondary antibody was used at a 1 in 4000 dilution. The Western blot allowed the visualisation of the potential cleavage products, from 112-C-terminus, 132-C-terminus and 158-C-terminus.

The developed autoradiography film, shown in Figure 3.3.19(a) indicated that there was cleavage of pVIII only in the presence of both Avp2 and pVIcT, as only in this lane were four additional bands found. The bands are highlighted in Figure 3.3.19(a) by arrows. Their mobilities are consistent with bands corresponding to fragments 1-111 (12K band), 1-131 (14K band), 158-C-terminus (35K band), and 132-C-terminus (38Kband) or 112-C-terminus (40K band) (Figure 3.3.18).

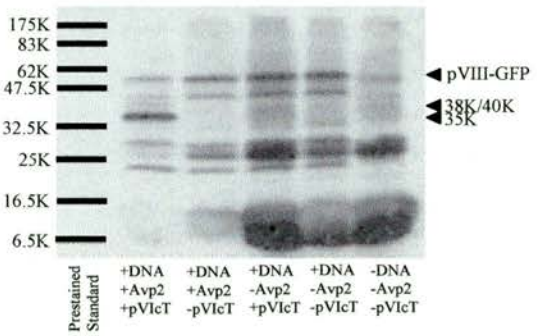
Likewise, the Western blot in Figure 3.3.19(b) indicated that cleavage had occurred between residues 158-C-terminus (35K band) and 132-C-terminus (38K band) or 112-C-terminus (40K band).

Figure 3.3.19: Avp2 Cleavage of TNT-produced pVIII
The TNT reaction was set up with TNT Quick Master Mix, [^{35}S]Methionine and +/-plasmid DNA. The samples were incubated for 90min at 30°C. Following incubation, five reactions were set up containing TNT reaction mix, buffer and either +Avp2/+pVIcT/+DNA; +Avp2/-pVIcT/+DNA; -Avp2/+pVIcT/+DNA; -Avp2/-pVIcT/+DNA; or -Avp2/-pVIcT/-DNA. These samples were then incubated for a further 30min at 37°C. The samples were separated on two 15% SDS-PAGE mini-gel. One gel was (a) exposed to BioMax MR Film (Kodak) and the Film was developed with a Kodak M35 X-OMAT Processor, whilst the other gel was (b) Western blotted with a GFP primary antibody. The (c) percentage of protein for each band was then calculated

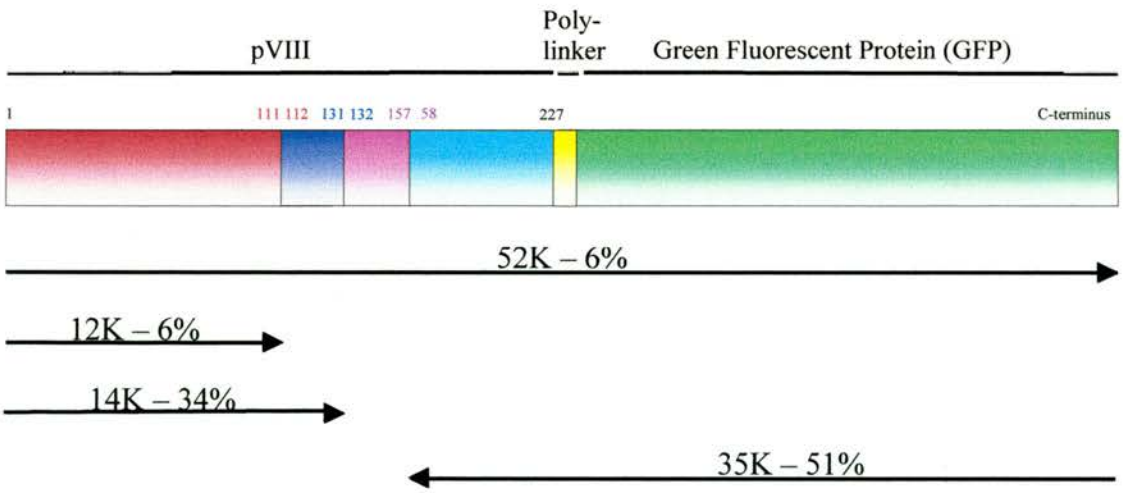
(a) Autoradiograph



(b) Western blot



(c) Percentage of pVIII Cleavage Products



The results suggested that all of the pVIII cleavage sites were cleaved by Avp, therefore the percentage of protein for each band was calculated (Method 2.7.2; Figure 3.3.19(c)). The results showed that most of the substrate was cleaved, with only 6%

remaining, and that the minor cleavage products were 132-C-terminus or 112-C-terminus (38Kor 40K band) and 1-111 (12K band), with only 3% and 6%, respectively, of the total protein content. Whilst the major cleavage products were 158-C-terminus (35K band) and 1-131 (14K band), as their percentage of total protein were 51% and 34%, respectively (Figure 3.3.19(c)).

Figure 3.3.19(a) and (b) also showed that there were a number of bands found in all four samples which had had DNA added to them during the TNT reaction. These additional bands were attributed to a number of factors - DNA degradation, premature termination during transcription, protein degradation and internal initiation. Of these four factors, only the effect of internal initiation could be analysed, by calculating the length of the expressed construct beginning at different internal initiation sites.

Initiation may have occurred at Methionines 10, 17, 34, 40 and 101 of pVIII, to produce bands at 51K, 50K, 48K, 47K and 40K. All other bands were attributed to DNA degradation, premature termination and protein degradation.

3.3.3.2 Mature VIII in Viral Preparations

Following the analysis of cleavage of pVIII in the 'TNT' system by Avp2, a purified human Ad2 viral preparation (prepared by Dr. A. Webster) was separated on a 15% SDS-PAGE mini-gel (Method 2.4.1.1) and was then transferred to PVDF according to Method 2.4.6.1. A band which lay between 6-14kDa (according to the protein standard) and corresponded to polypeptide VIII was then excised from the membrane and analysed by N-terminal protein sequencing. The sequence was identified as GIQLNDESV.

Subsequently the sequence was analysed using the PATTINPROT computer package at Pole Bioinformatique, Lyonnais (http://npsa-pbil.ibcp.fr/cgi-bin/pattern_pattinprot.pl). The proteins which were found to share 100% similarity to this

sequence were pVIII proteins of the Mastadenovirus genus. More specifically the sequence was found to correspond to cleavage having occurred at the 131-132 site of human Ad2 pVIII.

3.3.3.3 Discussion

The bioinformatic work utilised to investigate the cleavage sites of human Ad2 pVIII suggested that all three cleavage sites between residues 111-112, 131-132 and 157-158 were equally important (Section 3.3.1). Likewise, the cleavage results of pVIII produced in the TNT system indicated that pVIII could be cleaved at the sites.

This consolidated the work previously carried out in our laboratory by Dr. W.S. Annan who showed that His tagged pVIII protein purified under denaturing conditions could be cleaved. N-terminal sequencing and size markers on an SDS-PAGE mini-gel were used to identify cleavage products 1-111, 1-131, 132-227, 1-157 and 158-227.

Although the purification of pVIII using denaturation and subsequent refolding may not be ideal for this experiment, at the very least if the protein was not correctly folded then this method proved that the primary amino acid sequence could be cleaved at these sites.

In this work the recombinant protein produced in the TNT system would be expected to be folded correctly as it was synthesised in a eukaryotic system. Therefore, following cleavage, autoradiography allowed the visualisation of three cleavage products who's sizes were concurrent with fragments, 1-111 (12K), 1-131 (14K), 158-C-terminus (35K), and a fourth band which could either have been 132-C-terminus (38K) or the 112-C-terminus (40K). Likewise, the C-terminal antibody for GFP was also employed to analyse cleavage products and as expected the bands detected were 35K and 38K/40K.

Subsequently autoradiograph was used to analyse which bands constituted the major cleavage products, and these were found to be 158-C-terminus and 1-131. However

it must be borne in mind that this was over a distinct time course and over a longer time course I would expect to find that the major cleavage products were 1-111 and 158-C-terminus as the results indicated that 1-131 was being cleaved to 1-111. Thus a time course experiment would be essential to examine the order of cleavage *in vitro*.

Nevertheless the results of the TNT system have shown that folded pVIII can be cleaved at all three sites. Subsequently, an investigation of the human Ad2 virus using N-terminal sequencing discovered the sequence GIQLNDESV which is consistent with residues 132-140. This, together with the size of the protein, 6-14K (Table 3.3.5) suggested that the mature VIII protein found in the viral capsid of human Ad2 was composed of residues 132-227.

Table 3.3.5: Molecular Weights of the Potential Cleavage Products of pVIII

Sequence	Molecular Mass (K)
1-227	24.7
1-157	17.1
1-132	14.5
112-227	12.6
1-111	12.1
132-227	10.4
158-227	7.6
112-157	5.0
132-157	2.6
112-131	2.3

In summary the results showed that human Ad2 pVIII can be cleaved at all three sites and that mature VIII within the virus capsid is composed of residues 132-227. Therefore future analysis in this area should focus on the function(s) of the other cleavage sites.

CHAPTER 4
CONCLUSIONS

This Chapter summarises the results of Chapter 3 and also discusses future work.

4.1 Characterisation of the Structure-Function Relationship of Avp2

This sub-chapter examines the problems encountered during the purification of Avp2, EDSp and pVIII and describes potential solutions to them. It also makes conclusions over the characterisation of the residues involved in the S2 and S4 substrate specificity pockets, as well as those involved in binding the N-terminus of pVIcT. In addition, a putative function of the Avp2 C-terminus is proposed and the Avp2 cleavage sites of pVIII are discussed.

4.1.1 Purifying Adenovirus Proteins - Problems and Possible Solutions

As was noted in Section 3.1.1, native Avp2 cloned into the pET11c vector was not expressed in large quantities, therefore a number of factors which are known to improve expression were investigated. These included different cell lines, temperatures of induction and IPTG concentrations. Although none of these factors increased expression dramatically the yields were high enough to determine the specific activities for WT and most of the mutant Avp2 proteins (Sections 3.1.2 - 3.1.4).

However some of the mutant Avp2 proteins could not be purified due to low yields. Likewise the duck adenovirus protease, EDSp could not be purified using the system developed for native, recombinant Avp2 (Section 3.1.5) and therefore tagged vector systems were investigated (Sub-chapter 3.2).

These systems carried with them their own intrinsic problems, for example Avp2 expressed in two separate His tag vectors in the presence and absence of recombinant chaperone proteins expressed large quantities of Avp2, but the protein was found to be mainly insoluble. Nevertheless some of the protein was soluble and could bind Nickel beads. However, this route of purification was not followed up as it still produced low levels of purified protein, but be may worthy of future examination as the CE clan protease, Ulp1 was purified using a His tag vector (Mossesova & Lima., 2000).

As well as the small His tags, the larger Maltose fusion protein was investigated. These vectors did not cause large amounts of expression but the expressed protein was mainly soluble. The problem with the Maltose vectors was therefore not expression but purification, as the removal of the Maltose fusion protein resulted in precipitation.

Similarily this was the case for pVIII, as neither His tags nor a GST tag increased solubility, and the removal of the Maltose tag during protein purification also led to precipitation (Section 3.3.2).

One possible solution to purifying these adenovirus proteins would be leave the Maltose tag attached to the N-terminus, but in this system the proteins would not be completely purified. Alternatively, the DNA sequence could be analysed for rare codons as was carried out for Avp2 (Section 3.1.1). These rare codons could then be replaced with codons more commonly used by *E. coli*. However it may be worthwhile refraining from using *E. coli* as an expression system and instead investigating other systems such as insect cells or yeast in future work.

In the next Section the mutant Avp2 proteins for which a tagged vector purification system could be used are described. These include mutations in sites believed to be involved in substrate specificity, N-terminal pVIcT binding and the C-terminal tail of the protease. However as discussed in the introduction to Section 3.1.1 it was beneficial to use native Avp2 in the mutational analysis of substrate specificity as many of the mutations were at the N-terminus.

4.1.2 Characterisation of Avp2 Structure-Function

One of the major aims of this thesis was to investigate the structure-function relationship of three structural characteristics of Avp2. These were the S2 and S4 substrate specificity pockets, the pocket involved in binding in the N-terminus of pVIcT, and the C-terminal tail (Sections 3.1.2 – 3.1.4). The conclusions made from this work are discussed below.

4.1.2.1 Residues Involved in the S2 and S4 Substrate Specificity Pockets

The initial conclusions made from the Avp2/Ulp1-Smt3 superimposition (Section 3.1.2) was that Avp2 had been crystallised in its active conformation and that in the absence of a crystal structure of an Avp2-substrate complex, this was an ideal model. Furthermore, analysis of the Avp2-Smt3 model determined the orientation of the P1, P2, P3 and P4 Smt3 residues in the putative catalytic cleft of Avp2; P1, P2 and P4 pointed into the protease, whilst P3 pointed away from it. These results confirmed a similar proposal by Ding et al., 1996 who had superimposed Avp2 and Papain in complex with the synthetic substrate, QAAAA-chloromethyl ketone.

In addition to analysis of the Avp2/Ulp1-Smt3 superimposition, primary sequence alignments were employed to indicate residues putatively involved in the S2 and S4 substrate specificity pockets. The residues were then investigated using an Alanine screen and several of the mutants were found to cause large reductions in activity. These included S3A and V53A for the S2 pocket, E5A for the S4 pocket and W55A for both the S2 and S4 pockets (Section 3.1.2).

Of particular interest was Trp55, because it is believed to form part of both the S2 and S4 pockets, and when mutated, it caused an 84% reduction in activity compared to that of WT. In addition previous reports have suggested the importance of the equivalent

tryptophan in the CE clan members, Ulp1 (Mossessova & Lima., 2000) and Vaccinia virus protease (Byrd et al., 2003). The results of this thesis together with these reports provide evidence that the conserved tryptophan residue which lies adjacent to the catalytic cysteine, plays an essential role across the CE clan of cysteine proteases.

Although most of the Avp2 residues identified as being important in substrate specificity by bioinformatic analysis were involved in the Alanine screen, three were not. In hindsight residues, Asn44 and Gly123 should also have been part of the screen for numerous reasons (Section 3.1.2). Asn44 was found within a 5 and a 6 Angstrom radius of the P2 and P4 residues, respectively in the Avp2-Smt3 model, was conserved in both proteases in the Avp2/Ulp1 superimposition and was also conserved in the Avp alignment. Likewise, Gly123 was found within a 6 Angstrom radius of the P2 residue in the Avp2-Smt3 model, was conserved in both proteases in the Avp2/Ulp1 superimposition, sits next to the catalytic cysteine and was conserved in the alignment of the core domain of representative members of each of the five CE clan families. Further evidence for the importance of Gly123 was produced by Byrd et al., 2003 who analysed the equivalent residue in the CE clan protease of Vacinnia virus I7L, where mutation of this residue to an alanine caused a loss of enzyme activity.

In summary those residues believed to be involved in substrate specificity are Ser3, Gln5, Asn44, Val53, Trp55 and Gly123, of which Asn44 and Gly123 still remain to be examined by mutational analysis. However, although the Avp2/Ulp1-Smt3 superimposition was an excellent model for identifying residues putatively involved in the S2 and S4 substrate specificity pockets, in the absence of a crystal structure of Avp2 in complex with a substrate no functions were assigned to these residues.

As well as allowing the determination of residues putatively involved in the S2 and S4 substrate specificity pockets, the Avp2/Ulp1-Smt3 superimposition raised the question

of whether SUMO proteins were substrates for Avp2. As described in Section 3.1.2, both Smt3 and the human homologue, SUMO1 were cleaved by Avp2, albeit very slowly. This work was consolidated by cleavage analysis of a synthetic substrate, QSGGAFSW (Webster et al., 1989b) and the human Ad4 pTP protein (Webster et al.,1997), both of which contained a glutamine at the P4 site. These findings together with the results of this thesis suggested that the although glutamine could be accepted at the P4 site, cleavage was slow in its presence and therefore was not a true substrate. It is proposed that this is because glutamine is a polar, hydrophilic residue, whilst methionine, leucine and isoleucine are hydrophobic in nature.

However, a second theory of why SUMO proteins are cleaved slowly is that this acts as a temporal mechanism within host cells infected with adenovirus. If this was found to be the case *in vivo* and Avp2 functioned to de-sumoylate target proteins then this has ramifications for invasion by adenovirus, as sumoylation of proteins has been found to have several effects at the cellular level, including repression/activation of transcription, modifications to DNA structure, alterations to the sub-nuclear structure, increased oncogenic potential and changes to membrane transport (Reviewed in Tatham & Hay., 2003).

In the future it would be interesting to analyse whether cleavage of Smt3 and SUMO1 is slow due to the presence of the glutamine in the P4 site or due to some other structural feature, therefore it may be beneficial to mutate the glutamine to either a methionine, leucine or isoleucine residue as a control, as these amino acids are well established P4 site residues for Avp2. In addition current work in our laboratory utilising 2D-gel electrophoresis to compare Ad2 non-infected and infected cell lines will perhaps find that proteins which are sumoylated in non-infected cells are de-sumoylated in infected cells, thus providing further evidence for the temporal cleavage theory.

4.1.2.2 Residues Involved in Binding the N-terminus of pVIcT

Analysis of the 2.6 Angstrom Avp2-pVIcT structure suggested that the N-terminus of pVIcT sits in a pocket of Avp2 (Section 3.1.3). In addition, further bioinformatic and mutational analysis was used to identify the only polar residue of the pocket, Asp142 as a major contributor to the binding of pVIcT. Moreover it may be interesting to investigate residues 141, 142, 147, 150 and 152 of this pocket as they have all been suggested to form hydrogen bonds with either the N-terminal Gly1' or Val2' of pVIcT (McGrath et al., 2003).

To-date, Avp remains unique within the CE clan in its requirement for an activating peptide, therefore the Avp2-pVIcT/Ulp superimposition was analysed to examine whether Ulp1 requires a pVIcT-like β-strand. The analysis revealed that Ulp contains a β-strand which partially aligns with the Avp2 β-strand involved in binding the pVIcT. However, in the absence of evidence for an activating peptide and also in the absence of the complete structure of Ulp1 (the Ulp1 catalytic domain may be activated by the remaining domain which is not crystallised) then this question will remain unanswered and it must be presumed that Ulp1 does not require a co-activator.

4.1.2.3 Putative Function of the C-terminus

Examination of the C-terminal tail in the 2.6 Angstrom structure of Avp2 suggested that the tail was too distant from either the proposed substrate specificity cleft or the known pVIcT binding sites to be involved in any either of these functions (Section 3.1.4). However when a primary sequence alignment was carried out eight residues were found to be conserved within this region, four of which were completely conserved therefore suggesting that the C-terminal tail had an as yet unidentified function.

Unfortunately, the deletion mutants of the three secondary structures of the C-terminus (two α-helices formed by 161-181 and 182-195 and a random coil formed by 196-204) could not be purified due to low yields and it was hoped that the in future work these could be purified using another system (Section 4.1.1).

Hence, the question of whether the C-terminus of Avp2 is required for activity remains unanswered. However, the results showed that the tail was amphipathic, with hydrophobic residues facing towards the protease and polar residues facing away from it, therefore it is speculated that the tail plays a role in attaching the protease to a cellular or viral target. This may explain why the protease is mainly insoluble except when tagged to a large fusion protein such as maltose (Sub-chapter 3.2).

To investigate this the technique of West-western blotting could be used to analyse the ability of the protease to bind cellular or viral extracts which have been separated on an SDS-PAGE mini-gel and then transferred to PVDF. In addition it may be also be advantageous in the future to mutate each of the eight residues conserved within the C-terminus and analyse their effect on activity because perhaps the C-terminus does play a role in proteolytic activity.

As the conclusions made from the characterisation of the three structural components of Avp2 using bioinformatic and mutational analysis have now been discussed, the next Section examines the function of Avp2 in terms of cleavage of human Ad2 pVIII.

4.1.3 Characterisation of Avp2 Cleavage Sites in pVIII

Initial investigations of human Ad2 pVIII using a primary sequence alignment suggested that the three Avp2 cleavage sites between residues 111-112, 131-132 and 157-158 were all of equal importance (Section 3.3.1). These three sites were confirmed by both Avp2 cleavage of pVIII produced in a mammalian *in vitro* transcription-translation system in this thesis (Section 3.3.3) and also by Avp2 cleavage of a denatured, refolded, recombinant pVIII protein (Annan., 1999).

Furthermore N-terminal sequencing and a molecular mass size marker separated on an SDS-PAGE mini-gel were used to determine the cleavage site of mature polypeptide VIII. The protein was 6-14K in size and began with the sequence GIQLNDESV which is consistent with residues 132-227 (Section 3.3.3).

In summary the results showed that human Ad2 pVIII can be cleaved at all three sites and that mature VIII within the virus capsid is composed of residues 132-227. Therefore future analysis in this area should focus on the function(s) of the other cleavage sites. Additionally, it may be beneficial to use the mammalian *in vitro* transcription-translation system to carry out a time course experiment to examine the order of pVIII cleavage by Avp2 *in vitro*.

4.1.4 Future Work

Most of the work discussed in this thesis merits future investigation. This includes the further characterisation of Avp2 residues involved in binding the substrate and the N-terminus of the activating peptide. In addition to this work it may be beneficial to investigate the putative function of the Avp2 C-terminal tail, which remains undetermined.

As well as investigations into the characterisation of the structure-function relationship, further examination of purification methods used to produce recombinant Avp2 and pVIII would be invaluable. Purification of these enzymes would allow three crystal structures to be produced, Avp2 in the absence of pVIcT would provide insight into its inactive state, pVIII and fragments of pVIII would allow the analyse of similarities/differences in structure between pVIII and its cleavage products, and Avp2-pVIcT in complex with pVIII would permit residues involved in the substrate binding pockets to be further studied.

The purification of recombinant pVIII would also be advantageous as the protein could be used in a time point assay which would provide *in vitro* evidence for the order of human Ad2 pVIII cleavage.

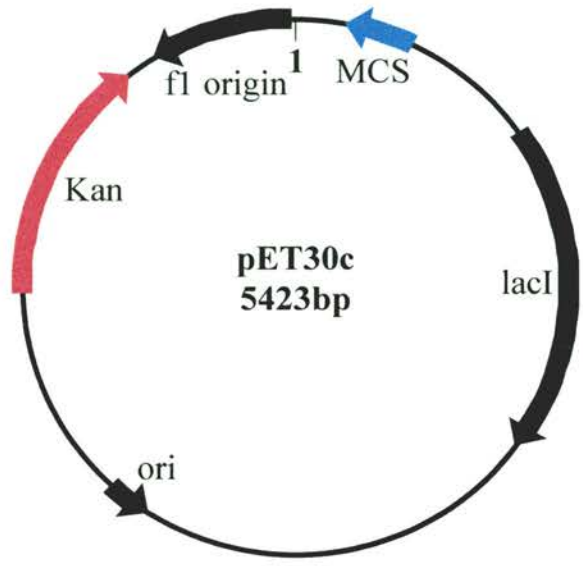
In conclusion, the major aim of the thesis was 'to further our knowledge of Avp2 substrate specificity by investigating both residues involved in the S2 and S4 substrate specificity pockets of Avp2 and also by examining the cleavage of pVIII to VIII' (Sub-chapter 1.5). These aims have been fulfilled. However further characterisation is required and many unanswered questions remain.

APPENDICES

APPENDIX I VECTOR MAPS

Appendix I.A pEHISTEV

Original Vector upon which pEHISTEV was based



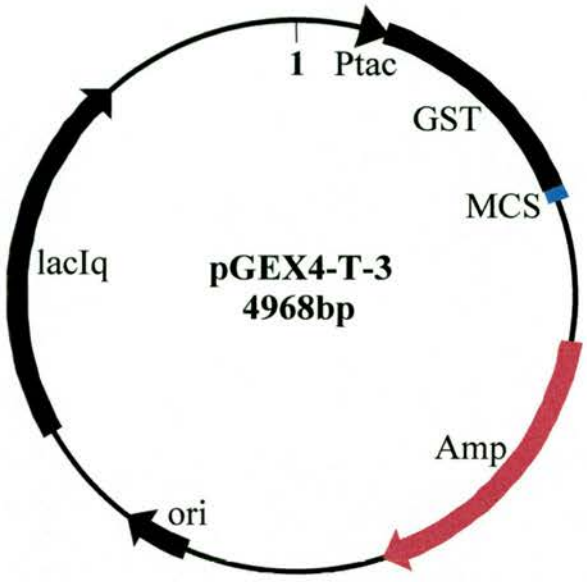
pEHISTEV Tev Protease Cleavage Site and Multiple Cloning Sites

- The MCS of pET30c was replaced with the MCS and Tev protease site given below.

```
  NdeI                          6 x His
CAT ATG TCG TAC TAC CAT CAC CAT CAC CAT CAC GAT TAC GAT ATC CCA
 M   S   Y   Y   Y   H   H   H   H   H   H   D   Y   D   I   P

                Tev Protease Site ▼       NcoI                BamHI
ACG ACC GAA AAC CTG TAT TTT CAG GGC GCC ATG GCT GAT ATC GGA TCC
 T   T   E   N   L   Y   F   Q   G   A   M   D   D   I   G   S

 EcoRI   SacI       SalI  HindIII    NotI       XhoI
GAA TTC GAG CTC CGT CGA CAA GCT TGC GGC CGC ACT CGA GCA CCA
 E   F   E   L   R   R   E   A   C   G   R   T   R   A   P
```

Appendix I.B pGEX4TEV

Original Vector upon which pGEX4TEV was based



pEHISTEV Tev Protease Cleavage Site and Multiple Cloning Sites

- The MCS of pGEX4-T-3 was replaced with MCS and Tev protease site given below.
- The V-protein was digested out of the vector and replaced with pVIII.

Tev Protease Site ▼ *NcoI* *V-Protein*

GGA ATT GAA AAC CTG TAT TTT CAG GGC GCC ATG ▬▬▬▬▬▬

P N E N L Y F Q G A M

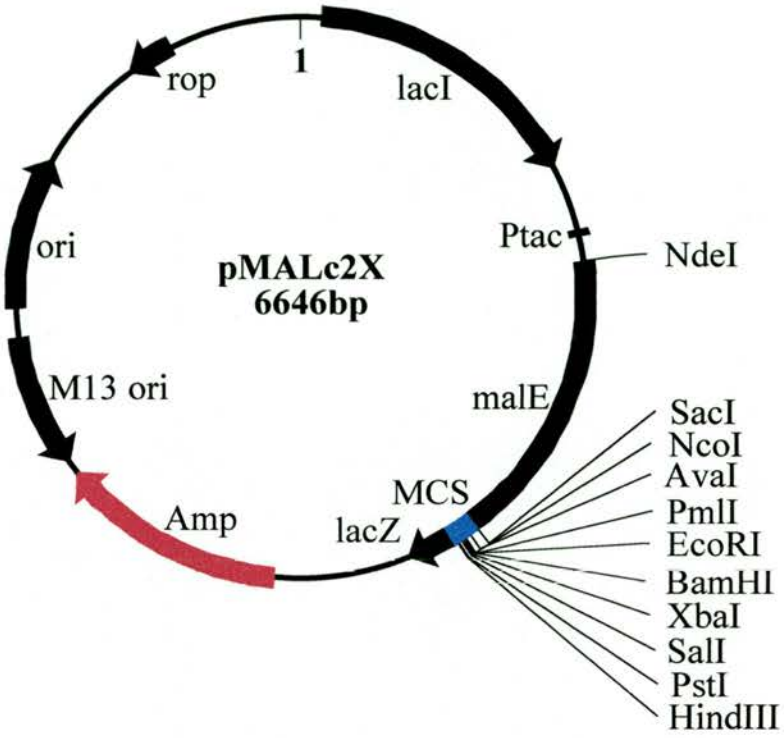
XhoI *NotI*

GAC TCG AGC GGC CGC ATC GTG ACT GAC TGA

D S S G R I V T D STOP

Appendix I.C pLou2 and pLou3

Original Vector upon which pLou2 and pLou3 were based

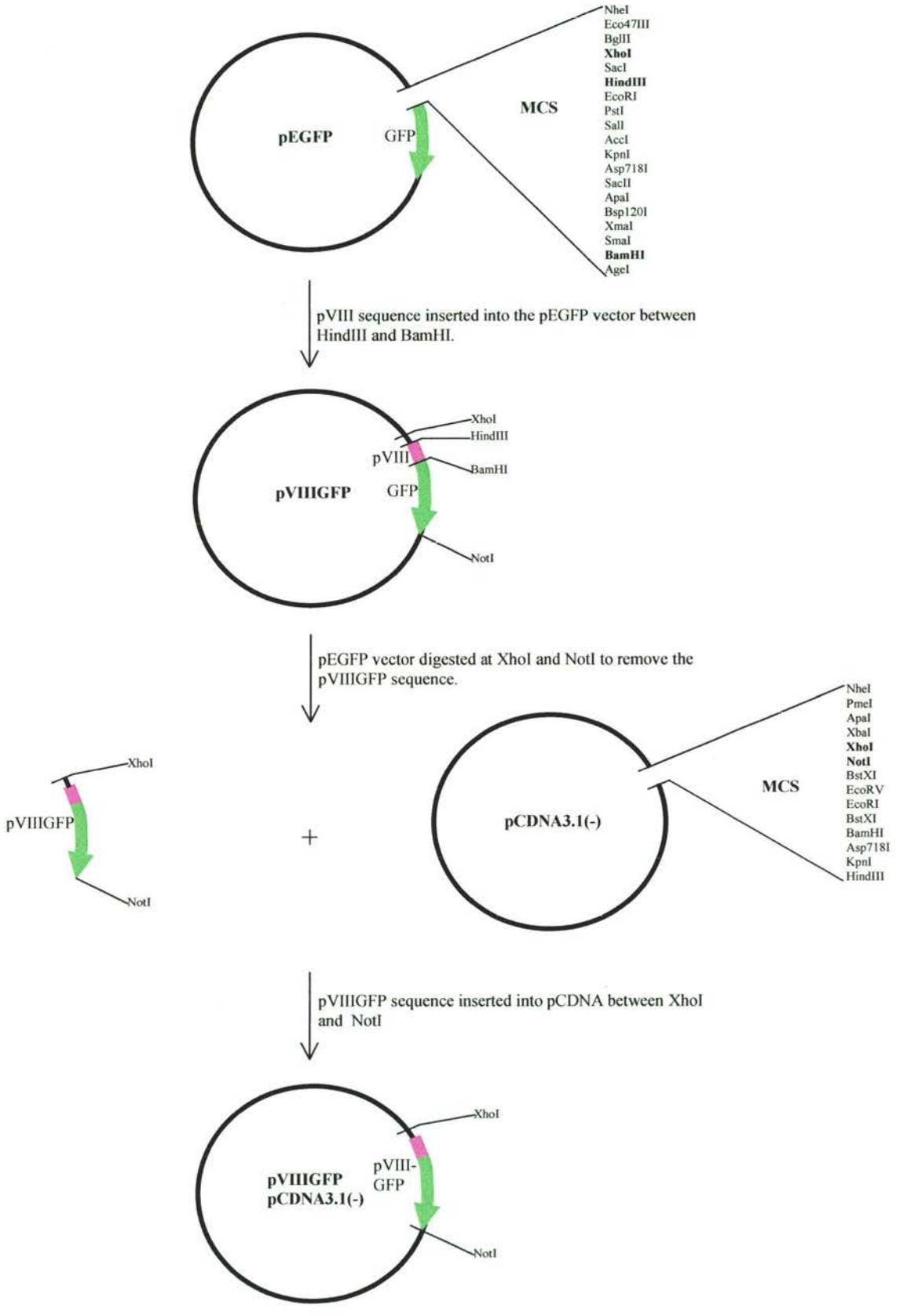


pLou2 and pLou3 Tev Protease Cleavage Site and Multiple Cloning Sites

- The MCS of pMALc2E was replaced with MCS and Tev protease site given below.
- Both vectors shared the same MCS but pLou3 had a 6 x His tag inserted at NdeI (1526).

```
  AvaI        Tev Protease Site ▼        NcoI          PmlI
CTC GGG GAA AAC CTG TAT TTT CAG GGC GCC ATG GAT CAC GTG
 L   G   E   N   L   Y   F   Q   G   A   M   D   H   V

 EcoRI    BamHI    XbaI     SalI    PstI       HindIII
GAA TTC  GGA TCC TCT AGA GTC GAC CTG CAG GCA AGC TTG GCA
 E   F    G   S   S   R   V   D   L   Q   A   S   L   A
```

Appendix I.D pVIIIGFPpcDNA3.1(-)



REFERENCES

ADACHI, Y., REYNOLDS, P.N., YAMAMOTO, M., WANG, M., TAKAYAMA, K., MATSUBARA, S., MURAMATSU, T. & CURIEL, D.T. (2001). A midkine promoter-based conditionaly replicative adenovirus for treatment of paediatric solid tumors and bone marrow tumor purging. *Cancer Research* **61**, 7882-7888.

ADHYA, S., SHNEIDMAN, P.S. & HURWITZ, J. (1986). Reconstruction of adenovirus replication origins with a human nuclear factor I binding site. *Journal of Biological Chemistry* **261**, 3339-3346.

AKALU, A., LIEBERMANN, H., BAUER, U., GRANZOW, H. & SEIDEL, W. (1999). The subgenus specific C-terminal region of protein IX is located on the surface of the adenovirus capsid. *Journal of Virology* **73**, 6182-6187.

AKUSJARVI,G & PERRSON, H. (1981). Gene and mRNA for precursor polypeptide VI from adenovirus type 2. *Journal of Virology* **38**, 469-482.

ANDERSON, C.W. (1990). The proteinase polypeptide of adenovirus serotype 2 virions. *Virology* **177**, 259-272.

ANDERSON, C.W., YOUNG, M.E. & FLINT, S.J. (1989). Characterization of the adenovirus 2 virion protein, Mu. *Virology* **172**, 506-512.

ANDRES, G., ALEJO, A., SIMON-MATEO, C & SALAS, M.L. (2001). African swine fever virus protease, a new viral member of the SUMO-1-specific protease family. *Journal of Biological Chemistry* **276**, 780-787.

ANGELETTI, P.C. & ENGLER, J.A. (1998). Adenovirus preterminal protein binds to the CAD enzyme at active sites of viral DNA replication on the nuclear membrane. *Journal of Virology* **72**, 2896-2904.

ANNAN, W.S. (1999). Cloning and purification of the adenovirus protein pVIII and its cleavage products. *Thesis*.

ANTHAPPILLY, F.K., MURALI, R., RUX, J.J., CAI, Z. & BURNETT, R.M. (1994). The refined crystal structure of hexon, the major coat protein of adenovirus type 2, at 2.9A resolution. *Journal of Molecular Biology* **242**, 430-455.

ARMENTANO, D., SMITH, M.P., SOOKDEO, C.C., ZABNER, J., PERRICONE, M.A., ST. GEORGE, J.A., WADSWORTH, S.C. & GREGORY, R.J. (1999). E4ORF3 requirement for achieving long-term transgene expression from the cytomegalovirus promoter in adenovirus vectors. *Journal of Virology* **73**, 7031-7044.

ARMENTANO, D., ZABNER, J., SACKS, C., SOOKDEO, C.C., SMITH, M.P., ST. GEORGE, J.A., WADSWORTH, S.C., SMITH, A.E. & GREGORY, R.J. (1997). Effect of the E4 region on the persistence of transgene expression from adenovirus vectors. *Journal of Virology* **71**, 2408-2416.

ARMENTERO, M., HORWITZ, M. & MERMOD, N. (1994). Targeting of the DNA polymerase to the adenovirus origin of DNA replication by interaction with nuclear factor I. *Proceedings of the National Academy of Science* **91**, 11537-11541.

ARNBERG, N., EDLUND, K., KIDD, A.H. & WADELL, G. (2000). Adenovirus type 37 uses Siallic acid as a cellular receptor. *Journal of Virology* **74**, 42-48.

BALAKIREV, M.Y., JAQUINOD, M., HAAS, A.L. & CHROBOCZEK, J. (2002). Deubiquitinating function of adenovirus proteinase. *Journal of Virology* **76**, 6323-6331.

BALAKIREVA, L., SCHOEN, G., THOUVENIN, E. & CHOROBOCZEK, J. (2003). Binding of Adenovirus capsid to Dipalmitoyl Phosphatidylcholine provides a novel pathway for virus entry. *Journal of Virology* **77**, 4858-4866.

BANIECKI, M.L., MCGRATH, W.J., DAUTER, Z. & MANGEL, W.F. (2002). Adenovirus proteinase: crystallisation and preliminary X-ray diffraction studies to atomic resolution. *Acta Crystallographica* **D58**, 1462-1464.

BANIECKI, M.L., MCGRATH, W.J., MCWHIRTER, S.M., LI, C., TOLEDO, D.L., PELLICENA, P., BARNARD, D.L., THORN, K.S. & MANGEL, W.F. (2001). Interaction of the human adenovirus proteinase with its 11-amino acid cofactor pVIc. *Biochemistry* **40**, 12349-12356.

BELIN, M. & BOULANGER, P. (1993). Involvement of cellular adhesion sequences in the attachment of adenovirus to the HeLa cell surface. *Journal of General Virology* **74**, 1485-1497.

BELNAP, D.M. & STEVEN, A.C. (2000). 'Déjà vu all over again': the similar structures of bacteriophage PRD1 and adenovirus. *Trends in Microbiology* **8**, 91-93.

BENKO, M., ELO, P., URSU, K., AHNE, W., LAPATRA, S., THOMSON, D. & HARRACH, B. (2002). First molecular evidence for the existence of distinct fish and snake adenoviruses. *Journal of Virology* **76**, 10056-10059.

BENKO, M. & HARRACH, B. (2003). Molecular evolution of adenoviruses. In: Adenoviruses: Model and vectors in virus-host interaction. *Current Topics in Microbiology and Immunology*. **272**, 1-35. Ed. by: Doerfler, W & Bohm, P. Springer-Verlag, Berlin, Heidelberg, New York.

BENSON, S.D., BAMFORD, J.K.H., BAMFORD, D.H. & BURNETT, R.M. (1999). Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* **98**, 825-833.

BERGELSON, J.M., CUNNINGHAM, J.A., DROGUETT, G., KURTJONES, E.A., KRITHIVAS, A., HONG, J.S., HORWITZ, M.S., CROWELL, R.L. & FINBERG, R.W. (1997). Isolation of a common receptor for coxsackie B viruses and adenovirus 2 and 5. *Science* **275**, 1320-1323.

BERNARDS, R., DE LEEUW, M.G.W., HOUWELING, A. & VAN DER EB, A.J. (1986). Role of the adenovirus early region 1B tumor antigens in transformation and lytic infection. *Virology* **150**, 126-139.

BEWLEY, M.C., SPRINGER, K., ZHANG, Y., FEIMUTH, P. & FLANAGAN, J.M. (1999). Structural analysis of the mechanism of adenovirus binding to its human cellular receptor, CAR. *Science* **286**, 1579-1583.

BHATTI, A.R. & WEBER, J. (1979). Protease of adenovirus type 2: partial characterisation. *Virology* **96**, 478-485.

BODNAR, J.W., HANSON, P.I., POLVINO-BODNAR, M., ZEMPSKY, W & WARD, D.C. (1989). The terminal regions of adenovirus and minute virus of mice DNAs are preferentially associated with the nuclear matrix in infected cells. *Journal of Virology* **63**, 4344-4353.

BOSHER, J., ROBINSON, E.C. & HAY, R.T. (1990). Interactions between the adenovirus type 2 DNA polymerase and the DNA binding domain of nuclear factor I. *New Biology* **2**, 1083-1090.

BRADFORD, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilising the principle of protein-dye binding. *Analytical Biochemistry* **72**, 248-254.

BRANTON, P.E., BAYLEY, S.T. & GRAHAM, F.L. (1985). Transformation by human adenoviruses. *Biochimica et Biophysica Acta* **780**, 67-94.

BROUGH, D.E., HSU, C., KULESA, V.A., LEE, G.M., CANTOLUPO, L.J., LIZONOVA, A. & KOVESDI, I. (1997). Activation of transgene expression by early 4 region is responsible for a high level of persistent transgene expression from adenovirus vectors in vivo. *Journal of Virology* **71**, 9206-9213.

BROWN, M.T., MCBRIDE, K.M., BANIECKI, M.L., REICH, N.C., MARRIOTT, G. & MANGEL, W.F. (2002). Actin can act as a cofactor for a viral proteinase in the cleavage of the cytoskeleton. *Journal of Biological Chemistry* **277**, 46298-46303.

BROWN, M.T., MCGRATH, W.J., TOLEDO, D.L. & MANGEL, W.F. (1996). Different modes of inhibition of human adenovirus proteinase, probably a cysteine proteinase, by bovine pancreatic trypsin inhbitor. *FEBS Letters* **388**, 233-237.

BRUDER, J.T., JIE, T., MCVEY, D.L. & KOVESDI, I. (1997). Expression of gp19K increases the persistence of transgene expression from an adenovirus vector in the mouse lung and liver. *Journal of Virology* **71**, 7623-7628.

BURNETT, R.M., GRUTTER, M.G. & WHITE, G.L. (1985). The structure of the adenovirus capsid. II. The packing symmetry of hexon and its implications for viral architecture. *Journal of Molecular Biology* **185**, 125-143.

BUTCHER, S.J., BAMFORD, D.H & FULLER, S.D. (1995). DNA packaging orders the membrane of bacteriophage PRD1. *EMBO Journal* **14**, 6078-6086.

BYRD, C.M., BOLKEN, T.C. & HRUBY, D.E. (2002). The vaccinia virus I7L gene product is the core protein proteinase. *Journal of Virology* **76**, 8937-8976.

CABRITA, G., IQBAL, M., REDDY, H. & KEMP, G. (1997). Activation of the Adenovirus Protease Requires Sequence Elements from Both Ends of the Activating Peptide. *The Journal of Biological Chemistry* **272**, 5635-5639.

CEPKO, C.L. & SHARP, P.A. (1982). Assembly of adenovirus major capsid protein is mediated by a nonvirion protein. *Cell* **31**, 407-415.

CHALLBERG, M.D. & RAWLINS, D.R. (1984). Template requirements for the initiation of adenovirus replication. *Proceedings of the National Academy of Science* **81**, 100-104.

CHAPPELL, J.D., PROTA, A.E., DERMODY, T.S. & STEHLE, T. (2002). Crystal structure of reovirus attachment protein σ1 reveals evolutionary relationship to adenovirus fiber. *EMBO Journal* **21**, 1-11.

CHARLES, P.C., CHEN, X., HORWITZ, M.S. & BROSNAN, C.F. (1999). Differential chemokine induction by the mouse adenovirus type-1 in the central nervous system of susceptible and resistant strains of mice. *Journal of Neurovirology* **5**, 55-64.

CHATTERJEE, P.K. & FLINT, S.J. (1987). Adenovirus type 2 endopeptidase: An unusual phosphoprotein enzyme matured by autocatalysis. *Proceedings of the National Academy of Science, USA* **84**, 714-718.

CHATTERJEE, P.K., VAYDA, M.E. & FLINT, S.J. (1986). Identification of proteins and protein domains that contact DNA within adenovirus nucleoprotein cores by ultraviolet light crosslinking of oligonucleotides ^{32}P-labelled *in vivo*. *Journal of Molecular Biology* **188**, 23-37.

CHEN, M., MERMOD, N. & HORWITZ, M.S. (1990). Protein-protein interactions betwenn DAN polymerase and nuclear factor I mediate formation of the DNA replication preinitiation complex. *Journal of Biological Chemistry* **265**, 18634-18642.

CHEN, P.H., ORNELLES, D.A. & SHENK, T. (1993). The adenovirus L3 23-Kilodalton proteinase cleaves the amino-terminal head domain from cytokeratin 18 and disrupts the cytokeratin network of HeLa cells. *Journal of Virology* **67**, 3507-3514.

CHIRMULE, N., MOSCIONI, A.D., QIAN, Y., QIAN, R., CHEN, Y. & WILSON, J.M. (1999). Fas-Fas ligand interactions play a major role in effector functions of cytotoxic T lymphocytes after adenovirus vector-mediated gene transfer. *Human Gene Therapy* **10**, 259-269.

CHROBOCZEK, E., GOUT, E, FAVIER, A.L. & GALINIER, R. (2003). Novel partner proteins of adenovirus penton. In: Adenoviruses: Model and vectors in virus-host interaction. *Current Topics in Microbiology and Immunology*. **272**, 37-55. Ed. by: Doerfler, W & Bohm, P. Springer-Verlag, Berlin, Heidelberg, New York.

CLAYMAN, G.L., EL-NAGGAR, A.K., LIPPMAN, S.M., HENDERSON, Y.C., FREDERICK, M., MERRITT, J.A., ZUMSTEIN, L.A., TIMMONS, T.M., LIU, T.J., GINSBERG, L., ROTH, J.A., HONG, W.K., BRUSO, P. & GOEPFERT, H. (1998). Adenovirus-mediated p53 gene transfer in patients with advanced recurrent head and neck squamous cell carcinoma. *Journal of Clinical Oncology* **16**, 2221-2232.

CLEAT, P.H. & HAY, R.T. (1989). Co-operative interactions between NFI and the adenovirus DNA binding protein at the adenovirus origin of replication. *EMBO Journal* **8**, 1841-1848.

CLEMENS, P.R., KOCHANEK, S., SUNADA, Y., CHAN, S., CHEN, H-H., CAMPBELL, K.P. & CASKEY, C.T. (1996). In vivo muscle gene transfer of full-length dystorphin with an adenoviral vector that lacks all viral genes. *Gene therapy* **3**, 965-972.

CLESHAM, G.J., ADAM, P.J., PROUDFOOT, D., FLYNN, P.D., EFSTATHIOU, S. & WEISSBERG, P.L. (1998). High adenoviral loads stimulate NF kappaB-dependant gene expression in human vascular smooth muscle cells. *Gene Therapy* **5**, 174-180.

COENJAERTS, S.J., VAN OOSTERHAUT, J.A.W.M. & VAN DER VLIET, P.C. (1994). The Oct-1 POU domain stimulates adenovirus DNA-replication between the viral precursor terminal protein-DNA polymerase complex and the POU homeodomain. *EMBO Journal* **13**, 5401-5409.

COLBY, W.W. & SHENK. (1981). Adenovirus type 5 virions can be assembled *in vivo* in the absence of detectable polypeptide IX. *Journal of Virology*, **39**, 977-980

CORNISH, J.A., MURRAY, H., KEMP, G.D. & GANI, D. (1995). Inhibitors of the adenovirus type 2 proteinase based on substrate-like tetrapeptide nitriles. *Bioinorganic & Medicinal Chemistry Letters*, **5**, 25-30.

DALES, S & CHARDONNET, Y. (1973). Early events in the interaction of adenoviruses with HeLa cells IV. Association with microtubules and the nuclear pore complex during vectorial movement of the inoculum. *Virology* **56**, 465-483.

DANIELL, E. (1976). Genome structure of incomplete particles of adenovirus. *Journal of Virology* **19**, 685-708.

DAVISON, E., KIRBY, I., ELLIOTT, T. & SANTIS, G. (1999). The human HLA-A*0201 allele, expressed in hamster cells, is not a high-affinity receptor for adenovirus Type 5 fiber. *Journal of Virology* **73**, 4513-4517.

DAVISON, A.J. BENKO, M. & HARRACH, B. (2003). Genetic content and evolution of adenoviruses. *Journal of General Virology* **84**, 2895-2908.

DEBBAS, M. & WHITE, E. (1993). Wild-type p53 mediates apoptosis by E1A, which is inhibited by E1B. *Genes and Development* **7**, 546-554.

DEDIEU, J.F., VIGNE, E., TORRENT, C., JULLIEN, C., MAHFOUZ, I., CAILLUND, J.M. AUBAILLY, N., ORSINI, C., GUILLAUME, J.M., OPOLON, P., DELAERE, P., PERRICAUDET, M.

& YEH, P. (1997). Long-term gene delivery into the livers of immunocompetent mice with E1/E4-defective adenoviruses. *Journal of Virology* **71**, 4626-4637.

DE JONG, R.N., MYSIAK, M.E., MEIJER, L.A.T., VAN DER LINDEN, M. & VAN DER VLIET, P.C. (2002). Recruitment of the priming protein pTP and DNA binding occur by overlapping Oct-1 POU homeodomian surfaces. *EMBO Journal* **21**, 725-735.

DE MOSSAIC, D., ZHENG, H. & KIRSHENBAUM, L.A. (1999). Linkage of the BH4 domain of Bcl-2 and the nuclear factor κB signalling pathway for suppression of apoptosis. *Journal of Biological Chemistry* **274**, 29505-29509.

DEVAUX, C, TIMMINS, PA. & BERTHET-COLOMINAS, C. (1983). Structural studies of adenovirus type 2 by neutron and X-ray scattering. *Journal of Molecular Biology* **167**, 119-132.

D'HALLUIN, J-C,, MILLEVILLE, M., BOULANGER, P.A. & MARTIN, G.R. (1978). Temperature-sensitive mutatn of adenovirus type 2 blocked in virion assembly: Accumulation of light intermediate particles. *Journal of Virology* **26**, 344-356.

DING, E.Y., HODGES, B.L., HU, H., MCVIE-WYLIE, A.J., SERRA, D., MIGONE, F.K., PRESSLEY, D., CHEN, Y.T. & AMALFITANO, A. (2001). Long-term efficacy after E1-, polymerase- adenovirus-mediated transferof human acid-alpha-glucosidase gene into glycogen storage disease type II knockout mice. *Human Gene Therapy* **12**, 955-965.

DING, J., M^{C}GRATH , W.J., SWEET, R.M & MANGEL, W.F. (1996). Crystal structure of the human adenovirus proteinase with its 11 amino acid cofactor. *EMBO Journal* **15**, 1778-1783.

DIOURI, M., KEYVANI-AMINEH, H., GEOGHEGAN, K.F. & WEBER, J.M. (1996). Cleavage efficiency by adenovirus protease is site-dependant. *Journal of Biological Chemistry* **271**, 32511-32514.

DORONIN, K., TOTH, K., KUPPUSWAMY, M., WARD, P., TOLLEFSON, A.E. & WOLD, W.S.M. (2000). Tumor-specific, replication-competent adenovirus vectors overexpressing the adenovirus death protein. *Journal of Virology* **74**, 6147-6155.

DOWNEY, J.F., ROWE, D.T., BACCHETTI, S., GRAHAM, F.L. & BAYLEY, S.T. (1983). Mapping of a 14,000-dalton antigen to early region 4 of the human adenovirus 5 genome. *Journal of Virology* **45**, 514-523.

EDVARDSSON, B., USTACELEBI, S., WILLIAMS, J. & PHILIPSON, L. (1978). Assembly intermediates among adenovirus type 5 temperature-sensitive mutants. *Journal of Virology* **25**, 641-651.

ELKON, K.B., LIU, C-C., GALL, J.G., TREVEJO, J., MARINO, M.W., ABRAHAMSEN, K.A., SONG, X., ZHOU, J-L., OLD, L.J., CRYSTAL, R.G. & FALCK-PEDERSEN, E. (1997). Tumor necrosis factor α plays a central role in immune-mediated clearance of adenoviral vectors. *Proceedings of the National Academy of Science* **94**, 9814-9819.

ENDERS, J.F., BELL, J.A., DINGLE, J.H., FRANCIS, T.(JR)., HILLEMANN, M.R., HEUBNER, R.J. & PAYNE, A.M.M. (1956). "Adenoviruses": Group name proposed for new respiratory-tract viruses. *Science* **124**, 119-120.

ENDTER, C. & DOBNER, T. (2004). Cell transformation by human adenoviruses. *Current Topics in Microbiology and Immunology* **273**, 163-214. Ed. by: Doerfler, W & Bohm, P. Springer-Verlag, Berlin, Heidelberg, New York.

ESCHE, H. (1982). Viral gene products in adenovirus type 2-transformed hamster cells. *Journal of Virology* **41**, 1076-1082.

ESCHE, H & STEIGMAN, B. (1982). Expression of early gene products in adenovirus type 12-infected and –transformed cells. *Journal of General Virology* **60**, 99-113.

EVERITT, E., SUNDQUIST, B., PETTERSSON, U. & PHILIPSON, L. (1973). Structural proteins of adenoviruses. X. Isolation and topography of low molecular weight antigens from the virion of adenovirus type 2. *Virology* **52**, 130-147.

EVERITT, E., LUTTER, L. & PHILIPSON, L. (1975). Structural proteins of adenoviruses. XII. Location and neighbour relationship among proteins of adenovirion type 2 as revealed by enzymatic iodination, immunoprecipitation and chemical cross-linking. *Virology* **67**, 197-208.

FISHER, K.J., CHOI, H., BURDA, J., CHEN, S-J. & WILSON, J.M. (1996). Recombinant adenovirus deleted of all viral genes for gene therapy of cystic fibrosis. *Virology* **217**, 11-22.

FITZGERALD, D.J.P., PADMANABHAN, R., PASTAN, I. & WILLINGHAM, M.C. (1983). Adenovirus-induced release of epidermal growth factor and pseudomonas toxin into the cytosol of KB cells during receptor-mediated endocytosis. *Cell* **32**, 607-617.

FLOMENBERG, P., PIASKOWSKI, V., TRUITT, R.L. & CASPER, J.T. (1995). Characterization of human proliferative T cell responses to adenoviruses. *Journal of Infectious Diseases* **171**, 1090-1096.

FREDMAN, J.N. & ENGLER, J.A. (1993). Adenovirus precursor to terminal protein interacts with the nuclear matrix in vivo and in vitro. *Journal of Virology* **67**, 3384-3395.

FURCINITTI, P.S., VAN OOSTRUM, J. & BURNETT, R.M. (1989). Adenovirus polypeptide IX revealed as capsid cement by difference images from electron microscopy and crystallography. *EMBO Journal* **8**, 3563-3570.

GAHERY-SEGARD, H., FARACE, F., GODFRIN, D., GASTON, J., LENGAGNE, R., TURSZ, T., BOULANGER, P. & GUILLET, J.G. (1998). Immune response to recombinant capsid proteins of adenovirus in humans: antifiber and anti-penton base antibodies have a synergistic effect on neutralizing activity. *Journal of Virology* **72**, 2388-2397.

GAHERY-SEGARD, H., JUILLARD, V., GASTON, J., LENGAGNE, R., PAVIRANI, A., BOULANGER, P. & GUILLET, J.G. (1997). Humoral immune response to the capsid

components of recombinant adenoviruses: routes of immunization modulate virus-induced Ig subclass shifts. *European Journal of Immunology* **27**, 653-659.

GALLIMORE, P.H., BYRD, P., GRAND, R.J., WHITAKER, J.L., BREIDING, D., WILLIAMS, J. (1984). An examination of the transforming and tumor-inducing capacity of a number of adenovirus type 12 early region 1, hostrange mutants and cells transformed by subgenomic fragments of Ad12 E1 region. *Cancer Cells* **2**, 519-526.

GANZ, T. & LEHRER. (1998). Antimicrobial peptides of vertebrates. *Current Opinion in Immunology* **10**, 41-44.

GINSBERG, H.S., PERIERA, H.G., VALENTINE, R.C. & WILCOX, W.C. (1966). A proposed terminology for the adenovirus antigens and virion morphological subunits. *Virology* **28**, 782-783.

GRABLE, M. & HEARING, P. (1990). Adenovirus type 5 packaging domain is composed of a repeated element that is functionally redundant. *Journal of Virology* **64**, 2047-2056.

GRAHAM, F.L. (1984). Transformation by and oncogenecity of human adenoviruses. In: The Adenoviruses, 339-398. Ed. by: Ginsberg, H.S. Plenum Press, New York.

GRAHAM, F.L., SMILEY, J., RUSSELL, W.C. & NAIRN, R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *Journal of General Virology* **36**, 59-74.

GREBER, U.F., WILLETTS, M., WEBSTER, P. & HELENIUS, A. (1993). Stepwise dismantling of adenovirus 2 during entry into cells. *Cell* **75**, 477-486.

GREBER, U.F., WEBSTER, P., WEBER, J. & HELENIUS, A. (1996). The role of adenovirus protease in virus entry into cells. *EMBO Journal* **15**, 1766-1777.

GRIERSON, A.W., NICHOLSON, R., TALBOT, P., WEBSTER, A. & KEMP , G. (1994). The protease of adenovirus serotype 2 requires cysteine residues for both activation and catalysis. *Journal of General Virology* **75**, 2761-2764.

GROPP, R., FRYE, M., WAGNER, T.O. & BARGON, J. (1999). Epithelial defensins impair adenoviral infection. *Human Gene Therapy* **10**, 957-964.

GUERMEUR, Y. (1997). Combinasion de classifieurs statistiques application a la prediction de structure secondaire des proteines. *Thesis*.

GUEX, N. & PEITSCH, M.C. (1997). SWISS-MODEL and the Swiss-Pdb viewer: An environment for comparitive protein modelling. *Electrophoresis* **18**, 2714-2723.

HALEY, K.P., OVERHAUSER, J., BABISS, L.E., GINSBERG, H.S. & JONES, N.C. (1984). Transformation properties of type 5 adenovirus mutants that differentially express the E1A gene products. *Proceedings of the National Academy of Science* **81**, 5734-5738.

HALLENBECK, P.L., CHANG, Y.N., HAY, C., GOLIGHTLY, D., STEWART, D., LIN, J., PHIPPS, S., CHIANG, Y.L. (1999). A novel tumor-specific replication-restircted adenoviral vector for gene therapy of hepatocellular carcinoma. *Human Gene Therapy* **10**, 1721-1733.

HANNAN, C., RAPTIS, L.H., DERY, C.D. & WEBER, J. (1983). Biological and structural studies with an adenovirus type 2 temperature-sensitive mutant defective for uncoating. *Intervirology* **19**, 213-223.

HAY, R.T. (1996). Adenovirus DNA replication. In: DNA replication in eukaryotic cells, 699-719. Ed. by: DePamphilis, M.L. Cold Spring Harbour Laboratory Press, New York.

HAY, R.T., STOW, N. & MCDOUGALL, I.M. (1984). Replication of adenovirus mini-chromosomes. *Journal of Molecular Biology* **175**, 493-510.

HERNANDEZ-ALCOCEBA, R., PIHALJA, M., WICHA, M.S. & CLARK, M.F. (2000). A novel, conditionally replicative adenovirus for the treatment of breast cancer that allows controlled replication of E1a-deleted adenoviral vectors. *Human Gene Therapy* **11**, 2009-2024.

HILLEMAN, M.R. & WERNER, J.H. (1954). Recovery of new agents from patients with acute respiratory illness. *Proceedings of the Society for Experimental Biology and Medicine* **85**, 183-188.

HONG, S.S., KARAYAN, L., TOURNIER, J., CURIEL, D.T. & BOULANGER, P.A. (1997). Adenovirus type 5 fiber knob binds to MHC class I α2 domain at the surface of human epithelial and B lymphoblastoid cells. *EMBO Journal* **16**, 2294-2306.

HORNE, R.W., BRENNER, S., WATERSON, A.P. & WILDY, P. (1959). The icosahedral form of adenovirus. *Journal of Molecular Biology* **1**, 84-86.

HORWITZ, M.S., MAIZEL J.V. & SCHARFF, M.D. (1970). Molecular weight of adenovirus type 2 hexon polypeptide. *Journal of Virology* **6**, 569-571.

HOUDE, A. & WEBER, J.M. (1990a). Adenovirus proteinases: comparison of amino acid sequence and expression of the cloned cDNA in *Escherichia coli*. *Gene* **88**, 269-273.

HOUDE, A & WEBER, J.M. (1990b). Adenovirus type 2 precursor proteins are cleaved by proteinases of other adenoviruses. *Virology* **179**, 485-486.

IBELGAUFTS, H., JONES, K.W., MAITLAND, N. & SHAW, J.F. (1982). Adenovirus-related RNA sequences in human neurogenic tumours. *Acta Neuropathology* **56**, 113-117.

JONES, S.J., IQBAL, M., GRIERSON, A.W. & KEMP, G. (1996). Activation of the protease from human adenovirus type 2 is accompanied by a conformational change that is dependant on cysteine-104. *Journal of General Virology* **77**, 1821-1824.

JONES, K.A., KADONAGA, J.T., ROSENFIELD, P.J., KELLY, T.J. & TIJAN, R. (1987). A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. *Cell* **48**, 79-89.

KAPLAN, J.M., ARMENTANO, D., SCARIA, A., WOODWORTH, L.A., PENNINGTON, S.E., WADSWORTH, S.C., SMITH, A.E. & GREGORY, R.J. (1999). Novel role for E4 region genes in protection of adenovirus vectors by cytotoxic T lymphocytes. *Journal of Virology* **73**, 4489-4492.

KANE, J.F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current Opinion in Biotechnology* **6**, 494-500.

KAUFFMAN, R.S. & GINSBERG, H.S. (1976). Characterization of a temperature-sensitive, hexon transport mutant of type 5 adenovirus. *Journal of Virology* **19**, 643-658.

KEYVANI-AMINEH, H., DIOURI, M., TIHANYI, K. & WEBER, J.M. (1996). Adenovirus type 2 endoprotease: isoforms and redox effects. *Journal of General Virology* **77**, 2201-2207.

KEYVANI-AMINEH, H., LABRECQUE, P., CAI, F., CARSTENS, E.B. & WEBER, J.M. (1995). Adenovirus protease expressed in insect cells cleaves adenovirus proteins, ovalbumin and baculovirus protease in the absence of activating peptide. *Virus Research* **37**, 87-97.

KIMURA, K. & GELMANN, E.P. (2000). Tumor necrosis factor-α and Fac activate complimentary Fas-associated death domain-dependant pathways that enhance apoptosis induced by γ-irradiation. *Journal of Biological Chemistry* **275**, 8610-8617.

KING, A.J. & VAN DER VLIET, P.C. (1994). A precursor terminal protein-trinucleotide intermediate during initiation of adenovirus DNA replication: regeneration of molecular ends in vitro by a jumping back mechanism. EMBO *Journal* **13**, 5786-5792.

KING, A.J., WIEKE, R.T. & VAN DER VLIET, P.C. (1997). Dissociation of the protein primer and DNA polymerase after initiation of adenovirus DNA replication. *Journal of Biological Chemistry* **272**, 24617-24623.

KOCHANEK., S., CLEMENS, P.R., MITANI, K., CHEN, H-H., CHAN, S. & CASKEY, C.T. (1996). A new adenoviral vector: Replacement of all viral coding sequences with 28kb of DNA independently expressing both full-length dystorphin and β-galacotsidase. *Proceedings of the National Academy of Science* **93**, 5731-5736.

KUHLMANN, I., ACHTEN, S., RUDOLPH, R. & DOERFLER, W. (1982). Tumour induction by human adenovirus type 12 in hamsters: loss of the viral genome from adenovirus type 12-induced tumor cells is compatible with tumor formation. *EMBO Journal* **1**, 79-86.

KUMAR-SINGH, R. & CHAMBERLAIN, J.S. (1996). Encapsidated adenovirus minichromosomes allow delivery of a 14kb dystorphin cDNA. *Human Molecular Genetics* **5**, 913-921.

KURIHARA, T., BROUGH, D.E., KOVESDI, I. & KUFE, D.W. (2000). Selectivity of a replication-competent adenovirus for human breast carcinoma cells expressing the MUC1 antigen. *Journal of Clinical Investigation* **106**, 763-771.

KUWANO, K., KAWASAKI, M., KUNITAKE, R., HAGIMOTO., N,. NOMOTO, Y., MATSUBA, T., NAKANISHI, Y. & HARA, N. (1997a). Detection of group C adenovirus DNA in small-cell lung cancer with the nested polymerase chain reaction. *Journal of Cancer Research and Clinical Oncology* **123**, 377-382.

KUWANO, K., NOMOTO, Y., KUNITAKE, R., HAGIMOTO., N., MATSUBA, T., NAKANISHI, Y. & HARA, N. (1997b). Detection of adenovirus E1A DNA in pulmonary fibrosis using nested polymerase chain reaction. *European Respiratory Journal* **10**, 1445-1449.

KVIST, S., OSTBERG, L., PERRSON, H., PHILIPSON, L. & PETERSON, P.A. (1978). Molecular association between transplantation antigens and cell surface antigen in adenovirus-transformed cell line. *Proceedings of the National Academy of Science* **75**, 5674-5678.

LEE, M.G., ABINA, M.A., HADDADA, H. & PERRICAUDET, M. (1995). The constitutive expression of the immunomodulatory gp19K protein in E1-, E3- adenoviral vectors strongly reduces the host cytotoxic T cell response against the vector. *Gene Therapy* **2**, 256-262.

LEONARD, G.T. & SEN, G.C. (1996). Effects of Adenovirus E1A protein on interferon-signalling. *Virology* **224**, 25-33.

LEONARD, G.T. & SEN, G.C. (1997). Restoration of interferon responses of adenovirus E1A-expressing HT1080 cell lines by overexpression of p48 protein. *Journal of Virology* **71**, 5095-5101.

LEVINE, A.J. (1996, Third Edition). The origins of virology. In: Fundamental virology, 1-14. Ed. by: Fields, B.N., Knipe, D.M. & Howley, P.M. Lippincott-Raven Publishers, Philadelphia, New York.

LI, S.-J. & HOCHSTRASSER, M. (1999). A new protease required for cell-cycle progression in yeast. *Nature* **398**, 246-251.

LI, Y., KANG, J., FRIEDMAN, J., TARASSISHIN, L.,YE, J., KOVALENKO, A., WALLACH, D. & HORWITZ, S. (1999). Identification of a cell protein (FIP-3) as a modulator of NF-κB activity and as a target of an adenovirus inhibitor of tumor necrosis factor α-induced apoptosis. *Immunology* **96**, 1042-1047.

LICHY, J.H., FIELD, J., HORWITZ, M. & HURWITZ, J. (1982). Separation of the adenovirus terminal protein precursor from its associated DNA polymerase: Role of both proteins in the initiation of adenovirus replication. *Proceedings of the National Academy of Science* **79**, 5225-5229.

LIEBER, A., HE., C-Y. & KAY, M.A. (1997). Adenoviral preterminal protein stabilizes min-adenoviral genomes in vitro and in vivo. *Nature Biotechnology* **15**, 1383-1387.

LIEBER, A., HE, C.Y., MEUSE, L., HIMEDA, C., WILSON, C. & KAY, M.A. (1998). Inhibition of NF-κB activation in combination with bcl-2 expression allows for persistance of first-generation adenovirus vectors in the mouse liver. *Journal of Virology* **72**, 9267-9277.

LIU, G.-Q., BABIS, L.E., VOLKERT, F.C., YOUNG, C.S.H. & GINSBERG, H.S. (1985). A thermolabile mutant of adenovirus 5 resulting from a substitution mutation in the protein VIII gene. *Journal of Virology* **53**, 920-925.

LONBERG-HOLM, K., CROWELL, R.L. & PHILIPSON, L. (1976). Unrelated animal viruses share receptors. *Nature* **259**, 679-681.

LOOK, D.C., ROSWIT, W.T., FRICK, A.G., GRIS-ALEVY, Y., DICKHAUS, D.M., WALTER, M.J. & HOLTZMAN, M.J. (1998). Direct suppression of Stat1 function during adenoviral infection. *Immunity* **9**, 871-880.

LOWE, S.W. & RULEY, H.E. (1993). Stabilization of the p53 tumor suppressor is induced by adnovirus 5 E1A and accompanies apoptosis. *Genes and Development* **7**, 535-545.

LUFTIG, R.B. & WEIHING, R.R. (1975). Adenovirus binds to rat brain microtubules in vitro. *Journal of Virology* **16**, 696-706.

LUKASHOK, S.A., TARASSISHIN, L., LI, Y. & HORWITZ, M.W. (2000). An adenovirus inhibitor of tumor necrosis factor alpha-induced apoptosis complexes with dynein and a small GTPase. *Journal of Virology* **74**, 4705-4709.

LUTZ, P., ROSA-CALATRAVA, M. & KEDINGER, C. (1997). The product of the adenovirus intermediate gene IX is a transcriptional activator. *Journal of Virology* **71**, 5102-5109.

MAITLAND, N.J., KINROSS, J.H., BUSUTILL, A., LUDGATE, S.M., SMART, G.E. & JONES, K.W. (1981). The detection of DNA tumour virus-specific RNA sequences in abnormal human cervical biopses by in situ hybridization. *Journal of General Virology* **55**, 123-137.

MAIZEL, J.V., WHITE, D.O. & SCHARFF, M.D. (1968). The Polypeptides of adenovirus II. soluble proteins, cores, top components and the structure of the virion. *Virology* **36**, 126-136.

MANGEL, W.F., M^{C}GRATH, J., TOLEDO, D.L. & ANDERSON, C. (1993). Viral DNA and a viral peptide can act as cofactors of adenovirus virion proteinase activity. *Nature* **361**, 274-275.

MANGEL, W.F., M^{C}GRATH, W.J., BROWN, M.T, BANIECKI, M.L., BARNARD, D.L. & PANG, Y.-P. (2001). A new form of antiviral combination therapy predicted to prevent resistance from arising, and a model system to test it. *Current Medicinal Chemistry* **8**, 933-939.

MANGEL, W.F., TOLEDO, D.L., BROWN, M.T., MARTIN, J.H & MCGRATH, W.J. (1996). Characterization of three components of human adenovirus proteinase activity in vitro. *Journal of Biological Chemistry* **271**, 536-543.

MANGEL, W.F., TOLEDO, D.L., DING, J., SWEET, R.M.& M^{C}GRATH, W.J. (1997). Temporal and spatial control of the adenovirus proteinase by both a peptide and viral DNA. *TIBS* **22**, 393-398.

MATHIAS, P., WICKHAM, T., MOORE, M. & NEMEROW, G. (1994). Multiple adenovirus serotypes use αv integrins for infection. *Journal of Virology* **68**, 6811-6814.

MATSUBARA, S., WADA, Y., GARDNER, T.A., EGAWA, M., PARK, M-S., HSEIH, C-L., ZHAU, H.E., KAO, C., KAMIDONE, S., GILLENWATER, J.Y. & CHUNG, L.W.K. (2001). A conditional replication-competent adenoviral vector, ad-OC-E1a, to cotarget prostate cancer and bone stroma in an experimental model of androgen-independent prostate cancer bone metastasis. *Cancer Research* **61**, 6012-6019.

MATTHEWS, D.A. & RUSSELL, W.C. (1994). Adenovirus protein-protein interactions: hexon and protein VI. *Journal of General Virology*. **75**, 3365-3374.

MATTHEWS, D.A. & RUSSELL, W.C. (1995). Adenovirus protein-protein interactions: molecular parameters governing the binding of protein VI to hexon and the activation of the adenovirus 23K protease. *Journal of General Virology*. **76**, 1959-1969.

MCDONALD, C. & REICH, N.C. (1999). Cooperation of the coactivators CBP and p300 with Stat6. *Journal of Interferon and Cytokine Research* **19**, 711-722.

MCGRATH, W.J., AHERNE, K.A. & MANGEL, W.F. (2002). In the virion, the 11-amino acid peptide cofactor pVIc is covalently linked tot he adenovirus proteinase. *Virology* **296**, 234-240.

MCGRATH, W.J., BANIECKI, M.L., LI, C., MCWHIRTER, S.M., BROWN, M.T., TOLEDO, D.L. & MANGEL, W.F. (2001). Human adenovirus proteinase: DNA binding and stimulation of proteinase activity by DNA. *Biochemistry* **40**, 13237-13245.

MCGRATH, W.J., DING, J., DIDWANIA, A., SWEET, R.M. AND MANGEL, W. F. (2003). Crystallographic structure at 1.6A resolution of the human adenovirus proteinase in a covalent complex with its 11-amino-acid peptide cofactor: insights on a new fold. *Biochimica et Biophysica Acta – Proteins and Proteomics* **1648**, 1-11.

MCLORIE, W., MCGLADE, C.J., TAKAYESU, D. & BRANTON, P.E. (1991). Individual E1B proteins induce transformation independently but by additive pathways. *Journal of General Virology* **72**, 1467-1471.

MEIER, O., BOUCKE, K., VIG HAMMER, S., KELLER, S., STIDWILL, R.P., HEMMI, S. & GREBER, U.F. (2002). Adenovirus triggers macropincytosis and endosomal leakage together with its clathrin-mediated uptake. *Journal of Cell Biology* **158**, 1119-1131.

MIRZA, M.A. & WEBER, J. (1982). Structure of adenovirus chromatin. *Biochimica et Biophysica Acta* **696**, 76-86.

MITANI, K., GRAHAM, F.L., CASKEY, C.T. & KOCHANEK, S. (1995). Rescue, propagation , and partial purification of a helper virus-dependent adenovirus vector. *Proceedings of the National Academy of Science* **92**, 3845-3858.

MIYAZAWA, N., CRYSTAL, R.G. & LEOPOLD, P.L. (2001). Adenovirus serotype 7 retention in a late endosomal compartment prior to cytosol escape is modulated by fiber protein. *Journal of Virology* **75**, 1387-1400.

MORAN, E., GRODZICKER, T., ROBERTS, R.J., MATTHEWS, M.B. & ZERLER, B. (1986). Lytic and transforming functions of individual products of the adenovirus E1A gene. *Journal of Virology* **57**, 765-775.

MOSSESSOVA, E. & LIMA, C.D. (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Molecular Cell* **5**, 865-876.

MUL, Y.M. & VAN DER VLIET, P.C. (1992). Nuclear factor-I enhances adenovirus DNA-replication by increasing the stability of a preinitiation complex. *EMBO Journal* **11**, 751-760.

MUL, Y.M., VERRIJZER, P. & VAN DER VLIET, P.C. (1990). Transcritpion factors NFI and NFIII/oct-1 function independantly, employing different mechanisms to enhance adenovirus DNA replication. *Journal of Virology* **64**, 5510-5518.

MURUVE, D.A., BARNES, M.J., STILLMAN, I.E. & LIBERMANN, T.A. (1999). Adenoviral gene therapy leads to rapid induction of multiple chemokines and acute neutrophil-dependant hepatic injury in vivo. *Human Gene Therapy* **10**, 965-976.

NAGATA, K., GUGGENHEIMER, R.A., ENOMOTO, T., LICHY, J.H. & HURWITZ, J. (1982). Adenovirus DNA replication in vitro: Identification of a host factor that stimulates synthesis of the preterminal protein-dCMP complex. *Proceedings of the National Academy of Science* **79**, 6438-6442.

NAGATA, K., GUGGENHEIMER, R.A. & HURWITZ, J. (1983a). Adenovirus DNA replication in vitro: Synthesis of full-length DNA with purified proteins. *Proceedings of the National Academy of Science* **80**, 4266-4270.

NAGATA, K., GUGGENHEIMER, R.A. & HURWITZ, J. (1983b). Specific binding of a cellular DNA replication protein to the origin of replication of adenovirus DNA. *Proceedings of the National Academy of Science* **80**, 6177-6181.

NEUMANN, R., CHROBOCZECK & BERNARD JACROT. (1988). Determination of the nucleotide sequence for the penton-base gene of human adenovirus type 5. *Gene* **69**, 153-157.

NEVELS, M., RUBENWOLF, S., SPRUSS, T., WOLF, H. & DOBNER, T. (1997). The adenovirus E4orf6 protein can promote E1A/E1B-induced focus formation by interfering with p53 tumor suppressor function. *Proceedings of the National Academy* **94**, 1206-1211.

NEVELS, M., TAUBER, B., KREMMER, E., SPRUSS, T, WOLF, H. & DOBNER, T. (1999). Transforming potential of the adenovirus type5 E4orf3 protein. *Journal of Virology* **73**, 1591-1600.

NEVELS, M., TAUBER, B., SPRUSS, T, WOLF, H. & DOBNER, T. (2001). "Hit-and-Run" transformation by adenovirus oncogenes. *Journal of Virology* **75**, 3089-3094s.

OHMAN, K., NORDQUIST, K., LINDER, S. & AKUSJARVI, G. (1995). Effects of adenovirus-2 early region 4 products on E1 transformation. *International Journal of Oncology* **6**, 663-668.

O'NEILL, E.A. & KELLY, T.J. (1988). Purification and characterization of nuclear factor III (origin recognition protein C), a sequence specific DNA binding protein required for efficient initiation of adenovirus DNA replication. *Journal of Biological Chemistry* **263**, 931-937.

ORTH, K., XU, Z., MUDGETT, M.B., QIN BAO, Z., PALMER, L.E., BLISKA, J.B., MANGEL, W., STASKAWICZ, B. & DIXON, J.E. (2000). Disrtuption of signalling by *Yersinia* effector YopJ, a ubiquitin-like protein protease. *Science* **290**, 1594-1597.

PANG, Y.-P., XU, K., KOLLMEYER, T.M., PEROLA, E., MCGRATH, W.J., GREEN, D.T. & MANGEL, W.F. (2001). Discovery of a new inhibitor lead of adenovirus proteinase: steps towards selective, irreversible inhibitors of cysteine proteinases. *FEBS Letters* **502**, 93-97.

PARASKEVA, C., BROWN, K.W., DUNN, A.R. & GALLIMORE, P.H. (1982). Adenovirus type 12-transformed rat embryo brain and rat liver epithelial cell lines: adenoviruses type 12 genome content and viral protein expression. *Journal of Virology* **44**, 759-764.

PARASKEVA, C. & GALLIMORE, P.H. (1980). Tumorigenicity and in vitro characteristics of rat epithelial cells and their adenovirus-transformed derivatives. *International Journal of Cancer* **25**, 631-639.

PARKS, R.J. & GRAHAM, F.L. (1997). A helper-dependent system for adenovirus vector production helps define a lower limit for efficient DNA packaging. *Journal of Virology* **71**, 3293-3298.

PAULSON, M., PISHARODY, S., PAN, L., GAUDAGNO, S., MUI, A.L. & LEVY, D.E. (1999). Stat protein transactivation domains recruit p300/CBP through widely divergent sequences. *Journal of Biological Chemistry* **274**, 25343-25349.

PERIERA, H.G., VALENTINE, R.C. & RUSSELL, W.C. (1968). Crystallisation of an adenovirus protein (the hexon). *Nature* **219**, 946-947.

PFEFFER, A., SCHUBBERT, R., OREND, G., HILGER-EVERSHEIM, K. & DOERFLER, W. (1999). Integrated viral genomes can be lost from adenovirustype 12-induced hamster tumor cells in a clone-specific, multistep process with retention of the oncogenic phenotype. *Virus Research* **59**, 113-127.

POLLARD, B.D. (2001). The activation of Adenovirus type 2 protease and the function of the activating peptide. *Thesis*.

PRONK, R. & VAN DER VLIET, P.C. (1993). The adenovirus terminal protein influences the binding of replication proteins and changes the origin of replication. *Nucleic Acids Research* **21**, 2293-2300.

PRUIJIN, G. J.M., VAN DRIEL, W., VAN DER VLEIT, P.C. (1986). Nuclear factor III, a novel sequence-specific DNA-binding protein from HeLa cells stimulating DNA replication . *Nature* **322**, 656-659.

RANCOURT, C., KEYVANI-AMINEH, H., SIRCAR, S., LABRECQUE, P. & WEBER, J.M. (1995). Proline 137 is critical for adenovirus protease encapsidation and activation but not enzyme activity. *Virology* **209**, 167-173.

RANCOURT, C., KEYVANI-AMINEH, H., DIOURI, M. & WEBER, J.M. (1996). Mutagenesis of conserved residues of the adenovirus protease. *Virology* **224**, 561-563.

RAWLINS, D.R., ROSENFIELD, P.J., WIDES, R.J., CHALLBERG, M.D. & KELLY, T.J.K. (JR). (1984). Structure and function of the adenovirus origin of replication. *Cell* **37**, 309-319.

REICH, N., PINE, R., LEVY, D. & DARNELL, J.E.(JR). (1988). Transcription of interferon-stimulated genes is induced by adenovirus particles but is suppressed by E1A gene products. *Journal of Virology* **62**, 114-119.

REKOSH, D.M.K., RUSSELL, W.C. & BELLET, A.J.D.(1977). Identification of a protein linked to the ands of adenovirus DNA. *Cell* **11**, 283-295.

ROBERTS, M.M., WHITE, J.L., GRUTTER, M.G. & BURNETT, R.M. (1986). Three-dimensional structure of the adenovirus major coat protein hexon. *Science* **232**, 1148-1151.

RODRIGUEZ, R., SCHUUR, E.R., LIM, H.Y., HENDERSON, G.A., SIMONS, J.W. & HENDERSON, D.R. (1997). Prostate attenuated replication competent adenovirus (ARCA) CN706: a selective cytotoxic for prostate-specific antigen-positive prostate cancer cells. *Cancer Research* **57**, 2559-2563.

ROELVINK, P.W., LIZONOVA, A., LEE, J.G.M, LI, Y., BERGELSON, J.M., FINBERG, R.W., BROUGH, D.E., KOVESDI, I. & WICKHAM, T.J. (1998). The Coxsackie-Adenovirus receptor protein can function as a cellular attachment protein for adenovirus serotypes from subgroups A, C, D, E and F. *Journal of Virology* **72**, 7909-7915.

ROGULSKI, K.R., WING, M.S., PAIELLI, D.L., GILBERT, J.D., KIM, J.H. & FREYTAG, S.O. (2000). Double suicide gene therapy augments the antitumour activity of a replication-competent lytic adenovirus. *Journal of Virology* **57**, 765-775.

ROSA-CALATRAVA, M, GRAVE, L., PUVION-DUTILLEUL,F., CHATTON, B & KEDINGER, C. (2001). Functional analysis of adenovirus protein IX identifies domains involved in capsid stability, transcriptional activation, and nuclear reorganization. *Journal of Virology* **75**, 7131-7141.

ROTH, J.A., SWISHER, S.G., MERRITT, J.A., LAWRENCE, D.D., KEMP, B.L., CARRASCO, C.H., EL-NAGGAR, A.K., FOSSELLA, F.V., GLISSON, B.S., HONG, W.K., KHURL, F.R., KURIE, J.M.., NESBITT, J.C., PISTERS,K., PUTNAM, J.B., SCHRUMP, J.S., SHIN, D.M. & WALSH, G.L. (1998). Gene therapy for non-small cell lung cancer: a preliminary report of a phase I trial of adnoviral p53 gene replacement. *Seminars in Oncology* **25**, 33-37.

ROWE, W.P., HUEBNER, R.J., GILMORE, L.K., PARROTT, R.H. & WARD, T.G. (1953). Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. *Proceedings of the Society for Experimental Biology and Medicine* **84**, 570-573.

RULEY, H.E. (1983). Adenovirus early region E1A enables viral and cellular transforming genes to transform primary cells in culture. *Nature* **304**, 602-606.

RUSSELL, W. (2000). Update on adenovirus and its vectors. *Journal of General Virology* **81**, 2573-2604.

RUSSELL, W & PRECIOUS, B. (1982). Nucleic acid-binding properties of adenovirus structural proteins. *Journal of General Virology* **63**, 69-79.

RUX, J.J., KUSER, P.R. & BURNETT, R.M. (2003). Structural and phylogenetic analysis of adenovirus hexons by use of high resolution x-Ray crystallographic, molecular modelling, and sequence based methods. *Journal of Virology* **77**, 9553-9566.

RUZINDANA-UMUNYANA, A., SIRCAR, S., WEBER, J.M. (2000). The Effect of Mutant Peptide Cofactors on Adenovirus Protease Activity and Virus Infection. *Virology* **270**, 173-179.

RYDMAN, P.S. BAMFORD, J.K.H. & BAMFORD, D.H.. (2001). A minor capsid protein P30 is essential for bacteriophage PRD1 capsid assembly. *Journal of Molecular Biology* **313**, 785-795.

SALONE, B., MARTINA, Y., PIERSANTI, S., CUNDARI, E., CHERUBINI, G., FRANQUEVILLE, L., FAILLA, C.M., BOULANGER, P. & SAGGIO, I. (2003). Integrin $\alpha 3\beta 1$ is an alternative cellular receptor for adenovirus serotype 5. *Journal of Virology* **77**, 13448-13454.

SAN MARTIN, C. & BURNETT, R.M. (2003). Stuctural studies on Adenoviruses. In: Adenoviruses: Model and vectors in virus-host interaction. *Current Topics in Microbiology and Immunology* **272**, 58-94. Ed. by: Doerfler, W & Bohm, P. Springer-Verlag, Berlin, Heidelberg, New York.

SAN MARTIN, C., BURNETT, R.M., DE HASS, F., HEINKEL, R., RUTTEN, T., FULLER, S.D., BUTCHER, S.J. & BAMFORD, D.H. (2001). Combined EM/X-ray imaging yields a quasi-atomic model of the adenovirus-related bacteriophage PRD1 and shows key capsid and membrane interaction. *Structure* **9**, 917-930.

SAN MARTIN, C., HUISKONEN, J.T., BAMFORD, J.K.H., BUTCHER, S.J., FULLER, S.D., BAMFORD, D.H. & BURNETT, R.M. (2002). Minor proteins, mobile arms and membrane-

capsid interactions in the bacteriophage PRD1 capsid. *Nature structural biology* **9**, 756-763.

SARNOW, P., HEARING, P., ANDERSEN, C.W., REICH, N. & LEVINE, A.J. (1982). Identification and characterization of an immunologically conserved adenovirus early region 11,000 M_r protein and its association with nuclear matrix. *Journal of Molecualr Biology* **162**, 565-583.

SAYLE, R. & MILNER-WHITE, E.J. (1995). Biomolecular graphics for all. *Trends in Biochemical Sciences* **20**, 374.

SCHAAK, J., HO, W.Y., FREIMUTH, P. & SHENK, T. (1990). Adenovirus terminal protein mediates both nuclear matrix association and efficient transcription of adenovirus DNA. *Genes and Development* **4**, 1197-1208.

SETH, P. (1994). Adenovirus-dependant release of choline from plasma membrane vesicles at an acidic pH is mediated by the penton base protein. *Journal of Virology* **68**, 1204-1206.

SETH, P., FITZGERALD, D.J.P., WILLINGHAM, M.C. & PASTAN, I. (1984). Role of a low-pH environment in adenovirus enhancement of the toxicity of a *Pseudomonas* exotoxin-epidermal growth factor conjugate. *Journal of Virology* **51**, 650-655.

SHEN, Y. ZHU, H & SHENK, T. (1997). Human cytomegalovirus IE1 and IE2 proteins are mutagenic and mediate "hit-and-run" oncogenic transformation in cooperation with adenovirus E1A proteins. *Proceedings of the National Academy of Science* **94**, 3341-3345.

SHENK, T. (1996, Third Edition). Adenoviridae: The viruses and their replication. In: Fundamental virology, 979-1016. Ed. by: Fields, B.N., Knipe, D.M. & Howley, P.M. Lippincott-Raven Publishers, Philadelphia, New York.

SHIROKI, K., HASHIMOTO, S., SAITO, I., FUKUI, Y., FUKUI, Y., KATO, H., & SHIMOJO, H. (1984). Expression of the E4 gene is required for establishment of soft agar colony-forming rat cell lines transformed by the adenovirus 12 E1 gene. *Journal of Virology* **50**, 854-863.

SMITH, C.A., WOODRUFF, L.S., ROONEY, C. KITCHINGHAM, G.R. (1998). Extensive cross-reactivity of adenovirus-specific cytotoxic T cells. *Human Gene Therapy* **9**, 1419-1427.

SOUBERBIELLE, B.E. & RUSSELL, W.C. (1995). Human T cell proliferative response to polypeptides from adenovirus type 2. *Journal of Infectious Diseases* **172**, 1421-1422.

STEVENSON, S.C., ROLLENCE, M., WHITE, B., WEAVER, L. & MCLELLAND, A. (1995). Human adenovirus serotypes 3 and 5 bind to two different cellular receptors via the fiber head domain. *Journal of Virology* **69**, 2850-2857.

STEWART, P.L., BURNETT, R.M., CYRKLAFF, M. & FULLER, S.D. (1991). Image reconstruction reveals the complex molecular organisation of adenovirus. *Cell* **67**, 145-154.

STEWART, P.L., FULLER, S.D. & BURNETT, R.M. (1993). Difference imaging of adenovirus: bridging the resolution gap between X-ray crystallography and electron microscopy. *EMBO Journal* **12**, 2589-2599.

STILLMAN, B.W., TOPP, W.C. & ENGLER, J.A. (1982). Conserved sequences at the origin of adenovirus DNA replication. *Journal of Virology* **44**, 530-537.

STORER, A.C. & MENARD, R. (1994). Catalytic mechanism in papain family of cysteine peptidases. *Methods in Enzymology* **244**, 486-500.

STOW, N.D. (1982). The infectivity of adenovirus genomes lacking DNA sequences from their left-hand termini. *Nucleic Acids Research* **10**, 5105-5119.

STUIVER, M.H & VAN DER VLIET, P.C. (1990). Adenovirus DNA-binding protein forms a multimeric protein complex with double-stranded DNA and enhances binding of nuclear factor I *Journal of Virology* **65**, 379-386.

SUOMALANIEN, M., NAKANO, M.Y., KELLER, S., BOUCKE, K., STIDWILL, R.P. & GREBER, U.F. (1999). Microtubule-dependant plus- and minus end-directed motilities are Competing processes for nuclear targeting of adenovirus. *The Journal of Cell Biology* **144**, 657-672.

SVENNSON, U. (1985). Role of vesicles during adenovirus 2 internalization into HeLa cells. *Journal of Virology* **55**, 442-449.

SVENNSON, U & PERRSON, R. (1984). Entry of adenovirus 2 into HeLa cells, *Journal of Virology* **51**, 687-694.

SUNG, M.T., CAO, T.M., LISCHWES, M.A. & COLEMAN, R.T. (1983). Molecular processing of adenovirus proteins. *Journal of Biological Chemistry* **258**, 8266-8272.

TEODORO, J.G., SHORE, G.C. & BRANTON, P.E. (1995). Adenovirus E1A proteins induce apoptosis by both p53-dependent and p53-independent mechanisms. *Oncogene* **11**, 467-474.

THOMPSON, J.D., HIGGINS, D.G. & GIBSON, T.J. (1994). CLUSTAL-W improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-4680.

TIBBETTS, C. (1977). Viral DNA sequences from incomplete particles of human adenovirus type 7. *Cell* **12**, 243-249.

TIHANYI, K., BOURBONNIERE, M., HOUDE, A., RANCOURT, C & WEBER, J.M. (1993). Isolation and properties of adenovirus type 2 proteinase. *Journal of Biological Chemistry* **268**, 1780-1785.

TOLLEFSON, A.E., HERMISTON, T.W., LICHTENSTEIN, D.L., COLLE, C.F., TRIPP, R.A., DIMITROV, T., TOTH, K, WELLS, C.E., DOHERTY, P.C. & WOLD, W.S.M. (1998). Forced degradation of Fas inhibits apoptosis in adenovirus-infected cells. *Nature* **392**, 726-730.

TOMKO, R.P., XU, R. & PHILIPSON, L. (1997). HCAR and MCAR: The human and mouse cellular receptors for subgroup C adenoviruses and group B coxsakieviruses. *Proceedings of the National Academy of Science, USA* **4**, 3352-3356.

TRASK., T.W., TRASK, R.P., AGUILAR-CORDOVA, E., SHINE, H.D., WYDE, P.R., GOODMAN, J.C., HAMILTON, W.J., ROJAS-MARTINEZ, A., CHEN, S.H., WOO, S.L. & GROSSMAN, R.G. (2000). Phase I study of adenoviral delivery of the HSv-tk gene and ganciclovir administration in patients with current malignant brain tumours. *Molecular Therapy* , 195-203.

TREMBLAY, M.L., DERY, C.V., TALBOT, B.G. & WEBER, J. (1983). In vitro cleavage specificity of the adenovirus type 2 proteinase. *Biochimica et Biophysica Acta* **743**, 239-245.

TRENTIN, J.J., YABE, Y. & TAYLOR, G. (1962). The quest for human cancer viruses: A new approach to an old problem reveals cancer induction in hamsters by human adenovirus. *Science* **137**, 835-849.

VALENTINE, R.C. & PERIERA, H.G. (1965). Antigens and structure of the adenovirus. *Journal of Molecular Biology* **13**, 13-20.

VAN DER VLIET, P.C. & LEVINE, A.J. (1973). DNA binding proteins specific for cells infected by adenovirus. *Nature New Biology* **246**, 170-174.

VAN OOSTRUM, J & BURNETT, R. (1985). Molecular composition of the adenovirus type 2 virion. *Journal of Virology* **56**, 439-448.

VARGA, M.J., WEIBULL, C. & EVERITT, E. (1991). Infectious entry pathway of adenvoirus type 2. *Journal of Virology* **65**, 6061-6070.

VAYDA, M.E. & FLINT, S.J. (1987). Isolation and characterization of adenovirus core nucleoprotein subunits. *Journal of Virology* **61**, 3335-3339.

VAYDA, M.E., ROGERS, A.E. & FLINT, S.J. (1983). The structure of nucleoprotein cores released from adenovirions. *Nucleic Acids Research* **11**, 441-460.

VRATI, S., BROOKES, D.E., STRIKE, P., KHATRI, A., BOYLE, D.B. & BOTH, G.W. (1996). Unique genome arrangement of an ovine adenovirus: Identification of a new proteins and proteinase cleavage sites. *Virology* **220**, 186-199.

WANG, K., GUAN, T., CHERESH, D.A. & NEMEROW, G.R. (2000). Regulation of adenovirus membrane penetration by the cytoplasmic tail of integrin β5. *Journal of Virology* **74**, 2731-2739.

WANG, Q., GREENBERG, G., BUNCH, D., FARSON, D. & FINER, M.H. (1997). Persistent transgene expression in mouse liver following in vivo gene transfer with a delta E1/delta E4 adenovirus vector. *Gene Therapy* **780**, 67-94.

WANG, K., HUANG, S., KAPOOR-MUNSHI, A & NEMEROW, G.R. (1998). Adenovirus internalization and infection require dynamin. *Journal of Virology* **72**, 3455-3458.

WAUGH, M.K., SWISHER, S.G., ROTH, J.A., NEMUNAITIS, J., LAWRENCE, D.D., KEMP, B.L., CARRASCO, C.H., CONNORS, D.G., EL-NAGGAR, A.K., FOSSELLA, F., GLISSON, B.S., HONG, W.K., KHURI, F.R., KURIE, J.M., LEE, J.J., LEE, J.S., MACK, M., MERRITT, J.A., NGUYEN, D.M., NESBITT, J.C., PEREZ-SOLER, R., PISTERS, K.M., PUTNAM, J.B.JR., RICHLIE, WR. & SAVIN, M. (1999). Adenovirus-mediated p53 gene transfer in advanced non-small-cell lung cancer. *Journal of the National Cancer Intsitute* **91**, 763-771.

WEBSTER, A., HAY, R.T. & KEMP, G. (1993). The adenovirus protease is activated by a virus-coded disulphide-linked peptide. *Cell* **72**, 97-104.

WEBSTER, A. & KEMP, G. (1993). The active adenovirus protease is the intact L2 23K protein. *Journal of General Virology* **74**, 1415-1420.

WEBSTER, A., LEITH, I.R. & HAY, R.T. (1994). Activation of Adenovirus-Coded Protease and Processing of the Preterminal Protein. *Journal of Virology* **68**, 7292-7300.

WEBSTER, A., LEITH, I.R., NICHOLSON, J., HOUNSELL, J. & HAY, R.T. (1997). Role of preterminal protein processing in adenovirus replication. *Journal of Virology* **71**, 6381-6389.

WEBSTER, A., RUSSELL, W.C. & KEMP, G. (1989a). Characterization of the adenovirus proteinase: Development and use of a specific peptide assay. *Journal of General Virology* **70**, 3215-3223.

WEBSTER, A., RUSSELL, S., TALBOT, P., RUSSELL, W.C. & KEMP, G.D. (1989b). Characaterization of the adenovirus proteinase: Substrate specificity. *Journal of General Virology* **70**, 3225-3234.

WICKHAM, T.J., MATHIAS, P., CHERESH, D.A. & NEMEROW, G.R. (1993). Intergrins $\alpha_v\beta_3$ and $\alpha_v\beta_5$ promote adenovirus internalization but not virus attachment. *Cell* **73**, 309-319.

WILLCOX, N. & MAUTNER, V. (1976). Antigenic determinants of adenovirus capsids. I. Measurement of antibody cross-reactivity. *Journal of Immunology* **116**, 19-24.

WHITE, E. & CIPRIANI, R. (1990). Role of adenovirus E1B proteins in transformation: altered organization of intermediate filaments in transformed cells that express the 19-kilodalton protein. *Molecular and Cellular Biology* **10**, 120-130.

WODRICH, H., GUAN, T., CINGOLANI, G., VON SEGGERN, D., NEMEROW, G & GERACE, L. (2003). Switch from capsid protein import to adenovirus assembly by cleavage of nuclear transport signals. *EMBO Journal* **22**, 6245-6255.

WOLF, L.A. & LASTER, S.M. (1999). Characterization of arachidonic acid-induced apoptosis. *Cell Biochemistry and Biophysics* **30**, 353-368.

WOLFE, S.L. (1993). Molecular and Cellular Biology Ed. by: Wadsworth Publishing Company, Belmont, California.

WU, E., PACHE, L., VON SEGGERN, D.J., MULLEN, T., MIKYAS, Y., STEWART, P. & NEMEROW, G. (2003). Flexibility of the adenovirus fiber is required for efficient receptor interaction. *Journal of Virology* **77**, 7225-7235.

YEH-KAI, L., AKUSJARVI, G., ALESTROM, P., PETTERSSON, U., TREMBLAY, M. & WEBER, J. (1983). Genetic Identification of an endoproteinase encoded by the adenovirus genome. *Journal of Molecular Biology* **167**, 217-222.

All bacterial cells/cell lines, including BL21(DE3), Rosetta(DE3) and JM109 discussed in this work are *Escherichia coli* cell lines.