A STUDY OF THE 2A REGION OF APHTHOVIRUSES
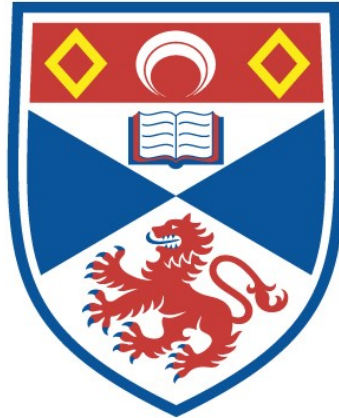
Michelle L. L. Donnelly

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

1997

A Study of the 2A Region of Aphthoviruses

by

Michelle L.L. Donnelly BSc.

Department of Chemistry

University of St. Andrews

A thesis submitted in partial fulfilment

of the requirements for the

degree of Doctor of Philosophy

June 1997

ProQuest Number: 10166285

ProQuest 10166285

TW
C362

DECLARATIONS

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

date 26/9/97   signature of candidate *MDonnelly*

I was admitted as a research student in October 1993 and as a candidate for the degree of Doctor of Philosophy in October 1993; the higher study for which this is a record was carried out in the University of St. Andrews between 1993 and 1996.

date 26/9/97   signature of candidate *MDonnelly*

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 this thesis in application for that degree.

date 26/9/97 signature of supervisor *David Cram*

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 a copy of the work may be made and supplied to any *bona fide* library or research worker.

date 26/9/97   signature of candidate *MDonnelly*

i

ACKNOWLEDGEMENTS

ABSTRACT

The proteins encoded by foot-and-mouth disease virus are expressed in the form of a polyprotein, which is processed by virus-encoded proteases to yield the mature viral proteins. The focus of this thesis is the 18 amino acid 2A region of foot-and-mouth disease virus (FMDV), which is (together with the first proline residue of 2B) capable of mediating a primary cleavage at its own carboxy-terminus, between the structural and replicative viral proteins. It was proposed that the 2A region performed this event *via* a novel proteolytic mechanism.

Vectors encoding the FMDV 2A region, in frame, between two foreign gene sequences were constructed to allow the 2A region to be investigated in isolation from the rest of the virus proteins using *in vitro* translation systems. Detailed quantitative densitometric analyses of the translation products were carried out, which suggested non-stoichiometric expression of the two cleavage products.

Single and multiple site-directed mutations were made to the 2A sequence and the activities of the resultant 2A regions determined to establish the identity of functional amino acid residues. The effect of the surrounding foreign protein sequence and the inclusion of progressively longer wild-type sequences prior to 2A was also examined.

The carboxy-termini of cardiovirus 2A regions show significant sequence similarity to the FMDV 2A region and are known to mediate a similar primary cleavage. The cleavage activities of the carboxy-termini of cardiovirus 2A regions were compared to that of the FMDV 2A region and found to equally capable of mediating cleavage.

The activity of FMDV 2A region was tested in prokaryotes and was found to be inactive.

These studies indicated that the 2A cleavage was not a proteolytic cleavage, but more likely a translational effect. A new model for 2A-mediated activity is presented.

CONTENTS

3. RESULTS 66

4. DISCUSSION 165

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ACTH	adrenocorticotropic hormone
ATP	adenosine 5'-triphosphate
BSA	bovine serum albumin
CAT	chloramphenicol acetyl transferase
C-	carboxy-
CLIP	corticotropin-like intermediate lobe peptide
DNA	deoxyribonucleic acid
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
dITP	2'-deoxyinosine 5'-triphosphate
dGTP	2'-deoxyguanine 5'-triphosphate
DMSO	dimethyl sulphoxide
dNTP	2'-deoxynucleoside 5'-triphosphate
ddNTP	dideoxynucleoside triphosphate
DTT	dithiothreitol
dTTP	2'-deoxythymidine 5'-triphosphate
E. coli	*Eschericia coli*
EDTA	ethylenediamine tetraacetic acid
eIF	eukaryotic initiation factor
ELH	egg-laying hormone
EMCV	encephalomyocarditis virus
ERV	equine rhinovirus
FMD	foot-and-mouth disease
FMDV	foot-and-mouth disease virus
GAG	group specific antigen
GUS	ß-glucuronidase
GTP	guanosine 5'-triphosphate

HAV	hepatitis A virus
HIV	human immunodeficiency virus
HRV	human rhinovirus
IPTG	isopropyl ß-D-thiogalactopyranoside
IRES	internal ribosome entry site
kDa	kilo Daltons
LB	Luria Broth
LPH	lipotropin
mRNA	messenger RNA
N-	amino-
NA	neuraminadase
NP40	Nonidet P-40
ns	non-structural
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
POMC	proopiomelanocortin
PSL	photo-stimulated luminescence
RNA	ribonucleic acid
rRNA	ribosomal RNA
S	Svedberg
SDS	sodium dodecyl sulphate
TAE	tris-actetate-EDTA buffer
TBE	tris-borate-EDTA buffer
TE	tris-EDTA buffer
TEMED	N,N,N',N'-tetramethylethylenediamine
TMEV	Theiler's murine encephalomyelitis virus

TnT	transcription and translation
tRNA	transfer RNA
TRIS	2-amino-2-(hydroxymethyj)propane-1,3-diol
UTR	untranslated region
UV	ultra violet
VPg	genome-linked viral protein
X-GAL	5-bromo-4-chloro-3-indolyl-ß-D-galactoside

Abbreviations for amino acids

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

1. INTRODUCTION

This thesis describes an investigation of the putative proteolytic activity of the short 18 amino acid 2A region (together with the adjacent proline residue of 2B) of the foot-and-mouth disease virus (FMDV). The introductory section brings together the various elements which have influenced both our understanding of proteolysis within FMDV replication and more specifically the creation of a model for 2A-mediated "cleavage" activity. This introduction describes FMDV itself, polyproteins and their adoption as a strategy in gene expression, proteases (cellular and viral), enzyme-independent proteolysis (*i.e.* deamidation and protein splicing) and, finally, picornavirus polyprotein processing in detail. The section is concluded with a presentation of some models in support of the hypothesis of a FMDV 2A region with proteolytic activity.

Section 1.1 Foot-and-mouth disease virus

1.1.1 Foot-and-mouth disease

Foot-and-mouth disease (FMD) is a highly infectious disease of even-toed ungulates (which include cattle, pigs, and sheep). FMD is rarely fatal (less than 5% in mature animals), except in very young animals. The disease causes a fever, followed by the formation of vesicles on the feet, in and around the mouth, and in other places where the skin is thin. It is characterised by lameness, salivation, reduced milk yield, and weight loss.

The virus is present in the vesicles and is released when they burst. The virus particles can survive for considerable lengths of time outside a host body and can be spread by the wind and objects that have been in contact with infected animals. Much of the spread of

1

the 1967 - 68 UK epidemic, which lasted over six months and involved 2397 outbreaks was attributed to the wind (West, 1992).

The disease is economically important not only through the failure of beef cattle to put on weight, the loss of milk in dairy cows, and the control measures required, but also at the international level, since an infected country would lose many of its export markets. Payments in compensation for the slaughter of over 211,000 cattle, 108,000 sheep, 113,000 pigs, and 50 goats during the 1967 - 68 UK epidemic were in excess of £27 million (West, 1992), a figure which would now equate to £billions. The use of chemically inactivated whole-virus vaccines in vaccination programmes, coupled with movement restrictions have been successful in the control of FMD. The elimination of FMD from Western Europe resulted in the cessation of routine vaccinations. Quarantine and import controls have since been used to exclude FMD, with a slaughter policy being implemented to prevent outbreaks becoming established. Such a policy involves the slaughter of all infected and exposed animals and the vaccination of all animals within a given radius. Compensation is awarded to owners of compulsorily slaughtered animals, but the cost of this is usually less than an annual vaccination programme.

1.1.2 Foot-and-mouth disease virus

1.1.2.1 Classification

The foot-and-mouth disease virus (FMDV) is the etiological agent of FMD. It belongs to the aphthovirus genus of the picornavirus family, which also includes four other genera: entero-, rhino-, cardio-, and hepatoviruses. The known picornaviruses are listed below in their current genera (Table 1.1). Aphthoviruses encompass both FMDV, of which there are seven distinct serotypes, O, A, C, SAT 1, SAT 2, SAT 3, and Asia 1. The recently sequenced equine rhinoviruses 1 and 2 (ERV1 and ERV2; Wutz *et al.*, 1996; Li *et al.*,

2

1996) show an extremely high degree of sequence similarity with aphthoviruses and will most probably be included within the aphthovirus genus at a later date.

Table 1.1 Classification of picornaviruses

Genera	Virus
• Enteroviruses	Poliovirus
	Coxsackie A and B viruses
	Swine vesicular disease virus
	Echovirus
	Human enteroviruses 68-71
	Bovine enteroviruses
	Simian enteroviruses
	Porcine enteroviruses
	Vilyuisk virus
• Rhinoviruses	Human rhinoviruses
	Bovine rhinoviruses
• Aphthoviruses	Foot and-mouth disease viruses
• Cardioviruses	Encephalomyocarditis viruses (EMCV)
	Theiler's murine encephalomyelitis viruses (TMEV)
	Mengovirus
• Hepatoviruses	Hepatitis virus A
• Unassigned	Equine rhinovirus 1
	Equine rhinovirus 2

1.1.2.2 Picornavirus genome structure

Picornaviruses are small RNA viruses in which one single stranded RNA molecule of positive polarity is encapsidated by an icosahedral protein shell of approximately 270 Å diameter. The genomic RNAs vary in length from 7209 (HRV-14) to 8450 bases (aphthovirus) and contain a single long open reading frame flanked by untranslated regions (UTRs). Additionally, the 3'-end of picornaviral RNA is polyadenylated, as is characteristic of most eukaryotic mRNAs. However, the 5'-ends of picornaviral RNA are not capped in the usual manner for eukaryotic mRNAs; cellular mRNAs contain a 5'-cap-structure which consists of 7-methylguanosine covalently linked by a 5'-5'-linkage to the terminal base of the mRNA. Instead of a 5'-cap-structure, picornaviruses have a small viral protein attached to their 5'-terminus, called 3B or VPg (genome-linked viral protein). Initiation of picornaviral protein synthesis occurs after an area of the 5'-UTR which is predicted to contain extensive secondary structure. This is termed the internal ribosome entry site (IRES), and is known to bind the 40S ribosomal subunit directly. The ribosomes then scan for a start codon as for normal cellular mRNAs. The aphthoviruses and cardioviruses have a poly(C) tract within their 5'-UTR, the function of which is unknown.

1.1.3 The strategy of picornaviral replication

The replication of picornaviruses has recently been reviewed by Reukert (1996), and is summarised here. The initial event in replication is the attachment of the virion to specific receptors on a host cell surface. This triggers an irreversible change in the conformation of the capsid, such that the capsid protein VP4 is lost, and the RNA genome is transferred across the cell membrane into the cytosol. Viral RNA itself is infectious when injected into the cell cytoplasm. The virus genome is then translated thereby supplying the viral proteins necessary for replication of its RNA. Virus proteins are

4

produced by translation of the single open reading frame into a long polyprotein which is cleaved *in statu nascendi* into the primary processing products (P1, P2 and P3, in the case of entero- and rhinoviruses; L, P1-2A, 2BC and P3, in the case of aphthoviruses; P1-2A, 2BC and P3 in hepatoviruses; and, L-P1-2A, 2BC and P3, in the case of cardioviruses (see Figure 1.1.1). The mature viral proteins are produced by further proteolytic processing. A strategy commonly used by positive strand RNA viruses is to encode proteins which are biochemically or temporally associated in discreet regions of the genome. Thus there is a segregation of the replicative and encapsidation functions within the polyprotein. In picornaviruses P1 encodes for the structural proteins, which compose the virion capsid structure, each capsid being made up of sixty copies of each protein. P2 and P3 encode the non-structural proteins.

The vRNA is then transcribed by a virally encoded RNA-dependent RNA-polymerase ($3D^{pol}$) to form complementary negative sense RNA, which in turn directs the synthesis of many new progeny positive sense RNA strands. These can either serve as templates for further protein or RNA synthesis, or be encapsidated to form new virions, depending on the stage in infection. Early in replication, newly synthesised positive sense RNA is recycled to form additional replication centres, but as the concentration of RNA increases a larger fraction is packaged into virions.

Before assembly begins P1, the precursor of the four mature capsid proteins, must be processed by the virally encoded protease, $3C^{pro}$ (or $3CD^{pro}$ in the entero- and rhinoviruses). This process is slow at early stages in infection, due to low concentrations of both P1 and 3C. Later, the increase in protease activity and the accumulation of more capsid proteins results in the assembly of capsids which can package VPg-RNA to form provirions. The maturation cleavage, in which capsid protein VP0 is cleaved into VP2 and VP4, is required to form the infectious virions that are released.

5

Entero-/Rhinoviruses

O—VpG | 1A | 1B | 1C | 1D | 2A^{pro} | 2B | 2C | 3A | 3C^{pro} | 3D^{pol} —A$_n$

3B

Capsid Proteins — Replicative Proteins

Cardioviruses

O—VpG | L | 1A | 1B | 1C | 1D | 2A^{pro} | 2B | 2C | 3A | 3C^{pro} | 3D^{pol} —A$_n$

3B

Capsid Proteins — Replicative Proteins

Hepatoviruses

O—VpG | 1A | 1B | 1C | 1D | 2A^{pro} | 2B | 2C | 3A | 3C^{pro} | 3D^{pol} —A$_n$

3B
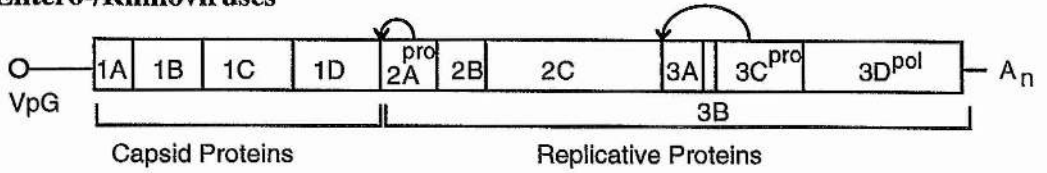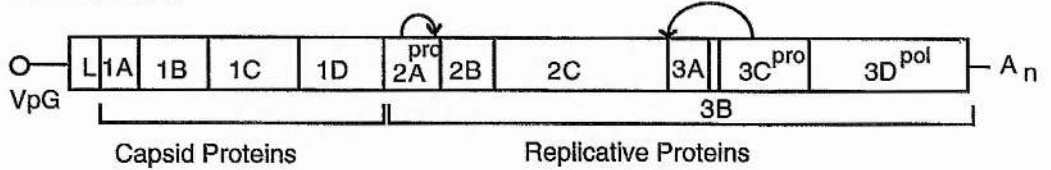
Capsid Proteins — Replicative Proteins

Aphthoviruses   *eg.* 1. FMDV

O—VpG | L^{pro} | 1A | 1B | 1C | 1D | 2B | 2C | 3A | 3C^{pro} | 3D^{pol} —A$_n$

2A   3B

Capsid Proteins — Replicative Proteins

eg. 2. ERV-1

O—VpG | L^{pro} | 1A | 1B | 1C | 1D | 2B | 2C | 3A | 3C^{pro} | 3D^{pol} —A$_n$

2 A   3B

Capsid Proteins — Replicative Proteins
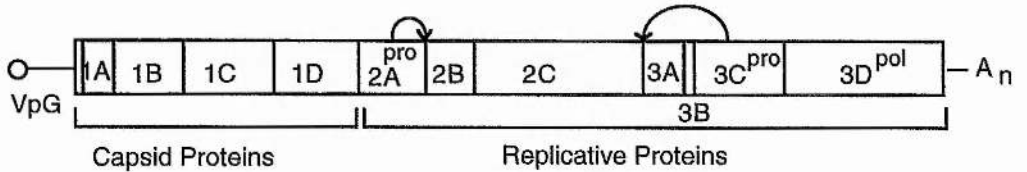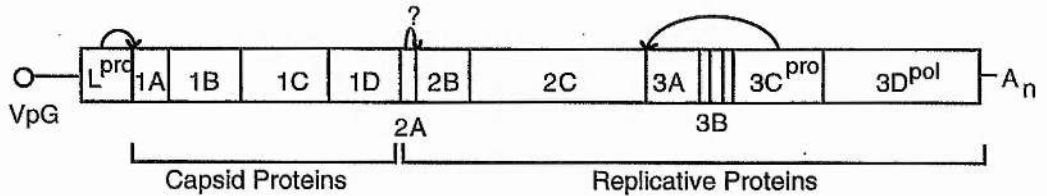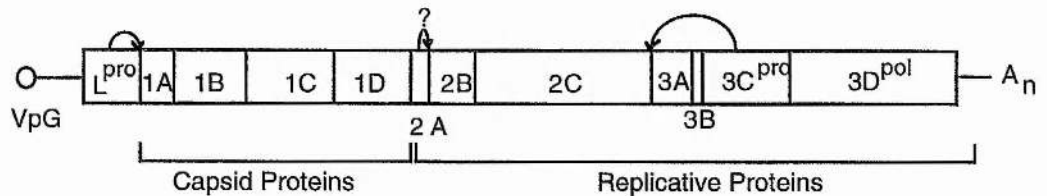
Figure 1.1.1 Picornavirus polyproteins and their primary processing events.
The polyprotein organisation within the single long open reading frame is shown.
Arrows indicate the sites of primary cleavages and the virus encoded proteases
responsible. The P1 region contains the mature proteins 1A to 1D which compose the
capsid. The mature protein products 2A to 2C compose P2 and 3A to 3D compose P3
together constituting the non-structural proteins involved in vRNA replication and protein
processing. The roles of those proteins involved in polyprotein processing, L, 2A and 3C,
will be discussed later in section 1.5. The functions of the proteins 2B, 2C and 3A have
still not been identified. Protein 3D is the RNA-dependent RNA polymerase, and 3B
encodes the virus protein VPg which blocks the 5'-terminus of the genome.

6

Section 1.2   Polyproteins as a strategy for gene expression

1.2.1 General overview

The concept of a one to one relationship between genes and proteins was proposed in 1941 (Beadle and Tatum). However it is now well established that there are many exceptions to this rule. Five different mechanisms are known to create multiple proteins from a single mRNA:

(1) In-phase overlapping genes with either read-through of a termination codon, or the use of in-phase internal initiation sites, relatively common in bacteriophage viruses, mitochondria and bacteria (Normark *et al.*, 1983).

(2) Ribosomal frame-shifting, a strategy employed by some viruses.

(3) Differential splicing of a pre-mRNA.

(4) Polyprotein processing, this has been well documented both for RNA viruses and for the expression of eukaryotic peptide hormones (see sections 1.2.2 and 1.2.4). More recently it has also been found to exist in prokaryotes (see section 1.2.3).

(5) Protein splicing (see section 1.4.1).

Polyprotein processing involves the transcription of a gene to a single mRNA which is translated into a single long polyprotein. This is then proteolytically processed either in *cis* (an intramolecular reaction) or in *trans* (an intermolecular reaction) to yield multiple protein products. The reactions in *cis* are often rapid, occuring co-translationally, and are referred to as autocatalytic. These follow zero order kinetics and are dilution insensitive. Reactions in *trans* are exemplified by a protein containing a proteolytic activity cleaving another protein or peptide at a substrate cleavage site and are of second order kinetics and dilution sensitive.

Polyproteins can be grouped into three classes with respect to the peptides that they generate: polyproteins composed of multiple copies of an identical sequence, those

7

containing related protein sequences often with related actions, and those generating peptides performing completely different biological functions. Examples of each of these will be discussed below.

There are various arguments for the evolutionary existence of polyproteins. It has been suggested that either polyproteins have originated as a consequence of the in-frame fusion of separate genes, or have evolved from the fragmentation of one entity with two or more distinct functions, by the invention of mechanisms for post-translational cleavage. There is evidence in support of both proposals. Polyproteins in which a protein or bio-active peptide is simply repeated are most easily explained by the theory of duplicated genes becoming fused (*e.g.* prepro-α-mating factor; Kurjan and Herskowitz, 1982). On the other hand the example of penicillin acylase, in which both α and β subunits are translated from the same mRNA (Bock *et al.*, 1983a), supports the theory of the separation of a single bifunctional protein into two enzymes. Neither subunit has an independent activity; they display activity only as a single functional unit. The evolution of a polyprotein from a single multifunctional protein requires that a protease cleavage site is developed and a protease recruited. This theory is also adhered to for penicillin acylase whose processing is either autocatalytic, or by a ubiquitous protease within bacteria (see later).

The strategy of polyprotein synthesis brings distinct advantages and disadvantages for gene expression. The synthesis and post-translational cleavage of a polyprotein constitutes the simplest means of guaranteeing a strict protein stoichiometry. It allows equal and efficient compartmentalisation of all proteins of a multicomponent system *e.g.*, translocation into the golgi apparatus. After post-translational cleavage the local concentration of the different components is increased rendering assembly of a multi-enzyme complex less dependent on diffusion. The polyprotein domains may fold interdependently removing the requirement of a chaperone. Additionally the polyprotein may perform a function distinct to those of its components. However, a severe drawback

8

to this strategy is that a mutation within the gene, leading to a stop codon, will lead to the absence of other protein products encoded beyond the mutation. For viruses the requirement for host cell proteases to process a polyprotein may limit the range of hosts.

In the following examples the use of polyproteins for the co-ordinated production of proteins is examined in a wide variety of systems.

1.2.2 Eukaryotic polyproteins

In eukaryotic cells the strategy of polyprotein synthesis and processing has been well documented for the production of neuroendocrine peptides, and more recently examples of larger cellular polyproteins have also been found.

Many eukaryotic genes contain coding sequences called exons which are interspersed with non-coding sequences called introns. The occurrence of split genes in eukaryotes has lead to the theory that introns separate functional domains in the gene-coding regions, thus allowing for the evolution of novel peptides. Indeed several polyprotein genes are organised such that exonic regions code for individual peptides or proteins.

The best known example of a eukaryotic polyprotein is the proopiomelanocortin (POMC) polyprotein found in the anterior pituitary gland of mammals. In 1977 cell-free protein synthesis from the mRNA of adrenocorticotropic hormone (ACTH), a steroidogenic hormone, demonstrated that the translation products contained β–lipotropin (β-LPH) as well as ACTH (Nakanishi *et al.*, 1977). This polyprotein, POMC, of 41 kDa is proteolytically cleaved to yield the two hormonal proteins ACTH and β-LPH. These products were known to be processed further themselves; ACTH to α-melanocyte-stimulating hormone and corticotropin-like intermediate lobe peptide (CLIP) (Scott *et al.*, 1973), and β-LPH to the opoid peptide, β-endorphin, and γ-lipotropin (Figure 1.2.1) (Li

9

and Chung, 1976; Ling *et al.*, 1976). The γ-melanocyte-stimulating hormone was also found to be encoded by the mRNA of POMC (Harris and Roos, 1956). The α- , β- and γ-melanocyte-stimulating hormones are encoded by three repetitive units of approximately 50 nucleotides in length, suggesting that this polypeptide has arisen by the duplication of an ancestral gene (Nakanishi *et al.*, 1980). However, POMC also provides an example of a polyprotein which consists of the unrelated sequences of hormonal proteins, opoid peptides and endorphins, which have separate biochemical actions. Depending on the tissue in which the precursor is synthesised, processing at pairs of basic amino acids, releases different combinations of polypeptides as mature products (Douglass *et al.*, 1984). The alternate pathways used in the processing of this polypeptide provides yet another mechanism to increase the number of proteins encoded by a single gene. There are many other such polypeptides encoding for unrelated hormones and neurotransmitters, including the two neurophysin-containing precursors which contain either arginine vasopressin or oxytocin and one of the two neurophysins, and prepro egg-laying hormone (ELH) which contains sequences of several different neurotransmitters in addition to (ELH) (see Douglass *et al.*, 1984; and Pierce and Parsons, 1981, for reviews).

Insulin is another well documented hormone whose biosynthesis involves polyprotein processing (Stryer, 1988). Insulin is composed of two peptide chains, A and B, joined by disulphide links. The precursor to insulin, preproinsulin, has a leader peptide followed by the sequences encoding the A and B chains, which are separated by a 30 amino acid connecting sequence. After formation of the disulphide bonds and removal of the leader sequence a proteolytic trypsin-like enzyme hydrolyses the polypeptide chain at dibasic sites to remove the connecting region and produce the mature insulin hormone (Figure 1.2.1). Polyprotein expression ensures the stoichiometry, and allows formation of the correct conformation and intramolecular disulphide bonds before proteolytic processing.

Prepro-α-mating factor found in yeast provides another example of multiple polypeptide products produced by a single polyprotein (Kurjan and Herskowitz, 1982). In this

10

example, however, it is multiple copies of a single 13 amino acid peptide (α-factor) which are produced by the polyprotein (Figure 1.2.1). The α-factor sequences are separated by 8 amino acid spacer regions which are similar in structure. Thus duplication of a unique (α-factor) / spacer region may have resulted in the present α-mating factor gene. Similar polyproteins producing duplications of peptides are the anti-freeze gene of the Arctic cod (Hsiao *et al.*, 1990), and both the preprocaerulin and thyrotropin-releasing hormone polyproteins of *Xenopus laevis* (Kuchler *et al.*, 1990; Richter *et al.*, 1986).

The first five steps of the biosynthesis of arginine from glutamate in the fungi *Neurospora crassa* occurs in the mitochondria. The second and third enzymes in this pathway are N-acetylglutamate kinase (AGK) and N-acetylglutamyl-γ-phosphate reductase (AGPR). These function as two separate enzymes, but are encoded by a single nuclear gene. These enzymes are synthesised in the cytosol as a polyprotein, before being transported to the mitochondria where processing occurs. Processing is mediated by the mitochondrial processing peptidase and processing enhancing enzyme (Gessert *et al.*, 1994). Existence as a polyprotein allows a more efficient transport of the enzymes into a mitochondrion. In *E. coli* a functionally equivalent biosynthesis takes place, with two genes 7 bases apart encoding the equivalent enzymes to AGK and AGPR although in the reverse order to *N. crassa*. It therefore seems likely that the gene encoding the polyprotein containing *N. crassa* AGK and AGPR enzymes was formed by fusion of the adjacent primitive genes with a rearrangement in the order. It does seem rather strange, however, that two genes should become fused only for their product to become processed into separate enzymes on import to the mitochondrion.

Other examples of fused genes are seen in the aromatic amino acid biosynthesis in *N. crassa* and *Saccharomyces cerevisiae* (Coggins *et al.*, 1987; Duncan *et al.*, 1987). These genes encode five different enzyme activities, however, they are not separated by polyprotein processing. It is thought that evolution favours the formation of covalently bound complexes to catalyse metabolic pathways when the intermediates are not required

Leader ▨□ ▤ ▨▨▨▨ ⟨⟨⟨⟨ **PRO-OPIOMELANO-**
CORTIN (POMC)

γ -Melanocyte-stimulating
hormone (γ -MSH) ▤

Adrenocorticotropic hormone
(ACTH) ▨▨

β-Lipotropic hormone (β-LPH)

▨▨▨▨ ⟨⟨⟨⟨

corticotropin-like
intermediate (CLIP) ▨

β-Melanocyte-stimulating
γ-Lipotropin hormone (β-MSH)

α -Melanocyte-stimulating
hormone (α-MSH) ▨

β-Endorphin
⟨⟨⟨⟨

Vertical lines separating regions of the polypeptide indicate dibasic cleavage sites.

(i)   PRO-OPIOMELANOCORTIN (POMC) - the single polyprotein produced from one gene is proteolytically cleaved to give copies of many different peptides.

PREPROINSULIN ▨▨▨▨▨
Leader

β-chain   α-chain
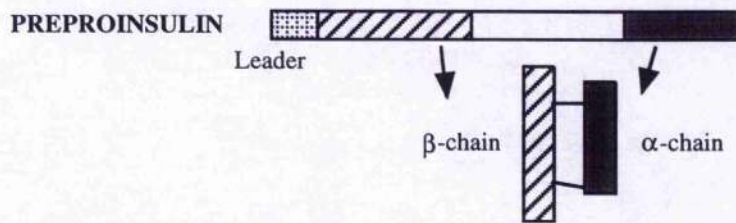
(ii)   PREPROINSULIN - The connecting sequence and leader peptide are removed by polyprotein processing.

PREPRO- α -MATING FACTOR ▨▨▨▨
Leader

■ ■ ■ ■

4 copies of the 13 amino acid peptide (α-factor)
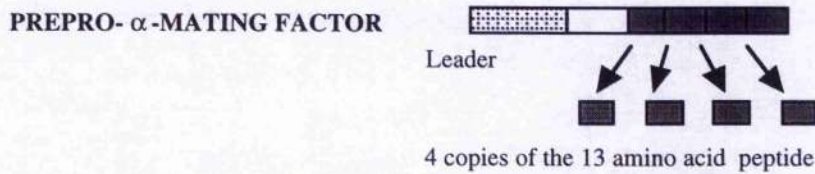
(iii) PREPRO- α-MATING FACTOR - multiple copies of a single peptide are produced via the single polyprotein encoded by one gene.

polyprotein precursor ▨▨▨
Leader

N-acetylglutamyl-γ-phosphate
reductase(AGPR) ▨▨

N-acetylglutamate kinase (AGK) ■
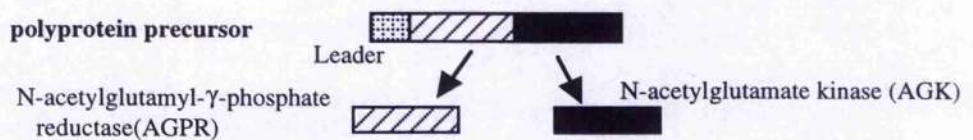
(iv)   The 871 amino acid polyprotein precursor of N-acetylglutamate kinase and N-acetylglutamyl-γ-phosphate is cleaved in the mitochondrion to the final mature products - two enzymes within a metabolic pathway.

Figure 1.2.1 Examples of eukaryotic polyproteins. The entire primary translation products are shown with vertical lines dividing regions which become final mature products.

for other systems (Srere, 1987). It has been suggested that the evolution of enzymes has progressed from the interaction of non-identical polypeptide chains to the single genetic unit-single polypeptide multi-enzyme system, performing more than one task (Bonner *et al.*, 1965).

1.2.3   Prokaryotic polyproteins

For many years proteolytic release of mature proteins from polyprotein precursors was thought to be a hallmark of eukaryotes. However, in 1983 the first example of a polyprotein in prokaryotes was identified (Böck *et al.*, 1983b). It was discovered that the *E.coli* Penicillin G acylase enzyme was composed of two dissimilar subunits of 23 and 69 kDa, both of which were encoded by the single open reading frame of the gene *pac* (Figure 1.2.2). The precursor polyprotein was found to be proteolytically processed to its mature products. Similar penicillin acylases or related enzymes have been found in other bacteria (Thöny-Meyer *et al.*, 1992). Such expression ensures the compartmentalisation of both components into the periplasma in equal amounts for the formation of hetero-dimers.

The *Bacillus subtlilis cotF* gene encodes a polyprotein precursor of 19 kDa which is proteolytically cleaved into two alkali soluble polypeptides of 5 and 8 kDa which are components of the spore coat (Figure 1.2.2; Cutting *et al.*, 1991). Again the initial synthesis of a polyprotein precursor ensures a 1:1 molar ratio of the mature proteins which could be necessary to establish a particular ordered structure.

In *Bacillus polymyxa* amylase is synthesised as a polyprotein precursor, 130 kDa, with α- and β-amylase activities. This was the first reported polyprotein precursor in prokaryotes to give rise to two enzymatic activities (Uozumi *et al.*, 1989). The polyprotein is proteolytically cleaved to produce multiform β-amylases, 70, 56 and 42 kDa, and α-amylase of 48 kDa (Figure 2). There is a linking sequence of 214 amino acids connecting

13

the α- and β-amylase regions, which contains two direct repeats of an 104 amino acid segment. The repeating connecting regions are susceptible to various proteases, suggesting that this linking sequence may be exposed on the surface of the molecule. It has been suggested that the direct repeating sequence may have somehow mediated a fusion between the two amylase genes. It is, however, still possible that this precursor could have been formed from one ancestral gene for α- and β-amylases. The application of a polyprotein would maintain an efficient means of secretion.

Other examples of prokaryotic polyproteins exist in the cytochrome bc_1 precursor of *Bradyrhizobium japonicum* (Thöny-Meyer *et al.*, 1991), and the *ptr* gene product of *Porphyromomas gingivalis* which encodes an arginine-specific thiol protease and multiple adhesins (Slakeski *et al.*, 1996).

E. coli Penicillin G acylase (*pac*)

Leader Peptide α-subunit Linker Peptide β-subunit

B. subtilis Spore Coat Proteins (*cotF*)

5 kDa 8 kDa

B. polymyxa Amylase

Leader Peptide β-amylase Linker Peptide α-amylase

↓ - Identified Processing Sites
▨ - Linker peptides
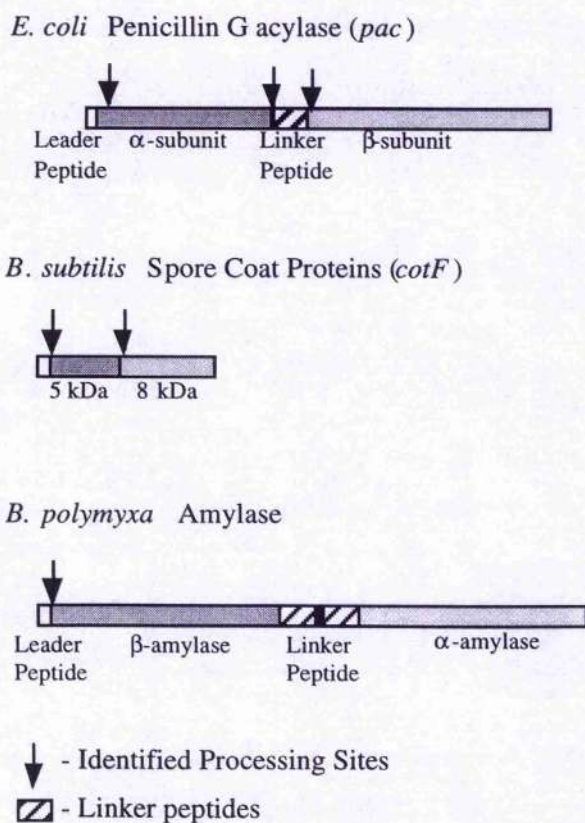
Figure 1.2.2 Examples of prokaryotic polyproteins. The entire primary translation products are shown. Vertical lines dividing those moieties with different structural or functional characteristics.

1.2.4   Viral polyproteins

Eukaryotic cells, unlike prokaryotic cells, usually initiate translation only from the 5'-end of an mRNA molecule and cannot recognise internal initiation sites. Hence they translate all mRNA molecules as if they are monocistronic messages containing only the information for a single protein. An RNA virus with a polycistronic genome must therefore have either:

(i)   a means of creating functional monocistronic mRNA for each viral protein in its genome, or

(ii) an ability to process the protein products of its genome to generate individual proteins.

One method of doing this is to have multipartite genomes with each mRNA strand packaged into a different virus particle. However this requires that all discrete parts of the virus genome must infect a single host cell to establish a productive infection. An alternative approach is that of segmented genomes which are then packaged into a single virus particle, though this requires accurate sorting of the individual segments. Yet another strategy is to produce subgenomic RNA. In the case of the positive strand RNA viruses this involves a first round of protein synthesis from the genomic RNA molecule to produce an mRNA-dependent RNA-polymerase, which synthesises negative sense RNA, and this in turn serves as a template from which subgenomic mRNA can be transcribed, to allow a second round of protein synthesis. A fourth solution is to encode a single polyprotein which is then proteolytically processed into many mature products. The major disadvantage of this approach is that a mutation within the genome leading to premature termination of translation can prevent the synthesis of multiple protein products, rather than the single final product within which the mutation exists. As mentioned earlier, in-phase overlapping genes and ribosomal frame-shifting, are also employed by some viruses, to aid in the production of the diverse proteins neccessary for

15

replication. Often a combination of these methods is employed for expression of a virus genome.

Many positive sense RNA viruses, retroviruses and some double stranded DNA viruses have employed the strategy of polyprotein synthesis and processing to produce multiple products from a single mRNA molecule. The first example of a viral polyprotein to be found was the P1 fragment of the poliovirus polyprotein encoding the capsid proteins (Jacobson and Baltimore, 1968).

It is clear that the processing of viral polyproteins by virus encoded proteases is not a simple enzyme substrate interaction. In a number of experimental systems, polyprotein precursors and protein products are generated and turned over in a temporal fashion, some short lived and others rapidly processed to final products (de Groot *et al.*, 1990). Differences in the rate of processing of a substrate by a protease under defined conditions have led to the proposal that differential proteolytic processing may be an integral part of genomic expression. Picornaviruses are unable to regulate gene expression at the level of transcription, however, the rates at which products are cleaved from a polyprotein vary considerably and hence certain proteins may accumulate faster than others.

Theoretically, alternative pathways of processing could be used to provide different subsets of biochemical functions from a single precursor. Models with such alternative processing pathways have been proposed for picornavirus non-structural proteins (Lawson and Semler, 1992; Flint, 1994). These are discussed in section 1.2.4.2.

1.2.4.1 Retrovirus polyproteins

In retroviruses, such as HIV, replication proceeds *via* reverse transcription of the viral RNA genome into a DNA intermediate, which integrates into the host cell chromosome

to generate a provirus. Gene expression begins by transcription of the entire proviral DNA sequence into a polycistronic mRNA. In the case of HIV three different strategies are used to generate the separate viral proteins. Differential splicing produces one final protein, whilst the entire genome molecule is used for translation of the GAG polyprotein and by a frameshift, the POL polyprotein. The infrequency of frame-shifting and the suppression of termination codons leads to an overproduction of structural proteins, which constitute the GAG polyprotein, compared to replicative enzymes. Each of the polyproteins is then cleaved into its mature protein components by a virus encoded proteinase (for a review see Dougherty and Semler, 1993).

1.2.4.2 Positive strand RNA virus polyproteins

The potyviruses, picornaviruses and togaviruses are all examples of viruses with positive strand RNA genomes. The togaviruses use two strategies, polyprotein processing and subgenomic mRNA to express their genome (Strauss and Strauss, 1986). They produce a polyprotein which is autocatalytically cleaved to produce an RNA-dependent-RNA-polymerase, amongst other products. This then makes an antigenome which acts as the template for a shorter positive-strand RNA molecule equivalent to the 3'-end of the genome. This is translated to give a polyprotein which is cleaved by a protease from the former polyprotein to give the structural gene products. The gene expression of potyviruses and picornaviruses are extreme examples of the use of polyprotein synthesis and processing, since this is their sole strategy for the production of mature viral proteins (Hellen *et al.*, 1989).

1.2.4.2.1 Picornavirus polyproteins

The entire genome of a picornavirus encodes a single open reading frame which is translated from a single polyprotein (Hellen *et al.*, 1989; Palmenberg, 1990). In entero-

and rhinoviruses there is a single initiation site. In cardio- and aphthoviruses there are an additional two alternative AUG codons at which initiation occurs, producing slightly shorter polyproteins. The polyprotein is cleaved *in statu nascendi*, into smaller polyproteins by rapid intramolecular cleavages, called primary cleavages, (see Figure 1.1.1 for the primary polyprotein cleavages of picornaviruses) such that isolation of the entire polyprotein is not possible from infected cells. However the full length polyprotein can be detected in infected cells in the presence of amino acid analogues (Jacobson and Baltimore, 1968) Further cleavages of the primary polyprotein processing products then occur in *cis* and *trans* to produce the mature viral protein products (see section 1.5). It has recently been reported that the cardiovirus, TMEV, uses an alternative reading frame to translate protein L*, from the genomic mRNA, which is involved in viral pathogenesis (Chen *et al.*, 1995). This is the only case found, thus far, of overlapping, but different, reading frames in picornaviruses.

All picornavirus polyprotein processing is exacted by virus encoded proteases; no cellular components involved in proteolytic processing have been isolated. Enzymic self-sufficiency must clearly be an advantage, and may partly explain why picornaviruses may efficiently infect a diverse range of host organisms.

The localisation of proteins, biochemically, or, temporally associated, into discreet regions of the genome, is a common strategy amongst positive strand RNA viruses. In picornaviruses the capsid proteins are encoded within P1, whilst P2 and P3 regions encode the precursors which are processed to the non-structural proteins involved in vRNA replication and protein processing. Thus upon cleavage of the initial polyprotein there is a separation of the structural and replicative proteins. The segregation of all capsid proteins on a single polyprotein theoretically ensures their equimolar expression at the same time and place, presumably aiding self-assembly.

18

Section 1.3 Proteases

Many enzymes catalyse a reaction either, according to the transition state theory suggested by Pauling in 1948, by binding a transition-state complex more tightly than the ground state and thereby stabilising the transition state, or by destabilising the reactants. Proteases are enzymes which catalyse the hydrolysis of peptide bonds *via* a nucleophilic attack on the carbonyl-carbon of the scissile bond, stabilising transition states, and providing a proton to the amine leaving group. The nature of the nucleophile is used to classify proteases into four major categories: serine, cysteine (or thiol), metallo, and aspartate (or acid) proteases. Recently a fifth category of protease has been identified, a threonine protease (Seemüller *et al.*, 1995). However, although only five types of nucleophile are used it is clear that nature has evolved many alternative mechanisms to form the required catalytic motifs.

Proteases are usually regarded as degradative enzymes, carrying out extensive hydrolysis of peptides and proteins to their constitutive amino acids and/or small oligopeptides. Nevertheless, proteolysis is involved in a productive manner in some important biosynthetic processes, including blood coagulation, and the production of peptide hormones (see section 1.2.2), and mature virions (see section 1.5.3). To provide such a role the extent of hydrolysis must be controlled and limited. The specificity of the reaction is imparted by the substrate binding pocket of the protease which interacts with substrate residues flanking the scissile bond. Usually proteolysis is limited by being directed toward the available polypeptide chain on the surface of a protein, and a number of proteases have been found which can recognise tertiary conformations rather than merely relying on the primary peptide sequence.

Proteases may be either endoproteases (also called proteinases *e.g.* trypsin), hydrolysing internal peptide bonds, or exoproteases, which cleave residues from the termini of proteins, often in a progressive manner, *e.g.* carboxypeptidase Y.

1.3.1 Cellular proteases

The four main groups of protease, metallo-, cysteine-, serine- and aspartate-proteases, are found in both prokaryotic and eukaryotic cells.

Sequence alignments and structural studies have revealed two major sub-types of serine protease. The large sub-group contains trypsin-like proteases, whereas the small sub-group contains enzymes similar to subtilisin. In both classes the catalytic residues, serine, histidine and aspartic acid, form a catalytic triad (Figure 1.3.1). The tertiary structure of the two classes differs such that the catalytic residues appear in a different order within the primary structure for each class (trypsin - His^{57}-Asp^{102}-Ser^{195}; subtilisin - Asp^{32}-His^{64}-Ser^{221}). In the absence of substrate the histidine residue is unprotonated, but poised to accept the serine hydroxyl proton, to which it is hydrogen-bonded. This makes the serine more nucleophilic than one would normally expect under physiological conditions. The hydroxyl group attacks the scissile carbonyl group to form an acyl-enzyme intermediate. The imidazole group abstracts the proton, the resulting positive charge being stabilised by the negatively charged aspartate group, and conveys the proton to the amine leaving group. Water catalysed hydrolysis of the acyl-enzyme intermediate leads to the formation of products (Phillips and Fletterick, 1992). The aspartic acid residue acts to orient the histidine and stabilise the charged histidine intermediate.

An important structural feature of serine protease catalytic sites is the oxyanion hole, which appears to be a component common to all serine proteases. It consists of two hydrogen bond donors, usually main-chain amides, which are essential to catalysis. Site directed mutational analysis has shown that they contribute 12-21 kJ/mol towards the transition state stabilisation (Bryan *et al.*, 1986). Furthermore, these hydrogen bonds accelerate the reaction by distorting the ground-state acyl-enzyme structure along the enzymatic pathway. The hydrogen bond between the imidazole and serine is thought to

be very weak and the serine is not intrinsically nucleophilic. Rather the serine reacts because it is positioned optimally to attack a tetrahedrally distorted carbonyl atom in the substrate (Kraut, 1977). Thus the interaction of the dipoles of the main-chain amides and the scissile peptide bond carbonyl-oxygen is thought to promote the formation of the transition state and increase the rate of deacylation of the intermediate (Whiting and Peticolas, 1994).

Recently novel serine proteases which use hydroxyl / amine catalytic dyads have been discovered (for a review see Paetzel and Dalbey, 1997). These proteases contain an essential lysine, but no essential histidines. Thus it has been proposed that an ε-amino group of this lysine residue acts as a general base to increase the nucleophilicity of the active site serine.
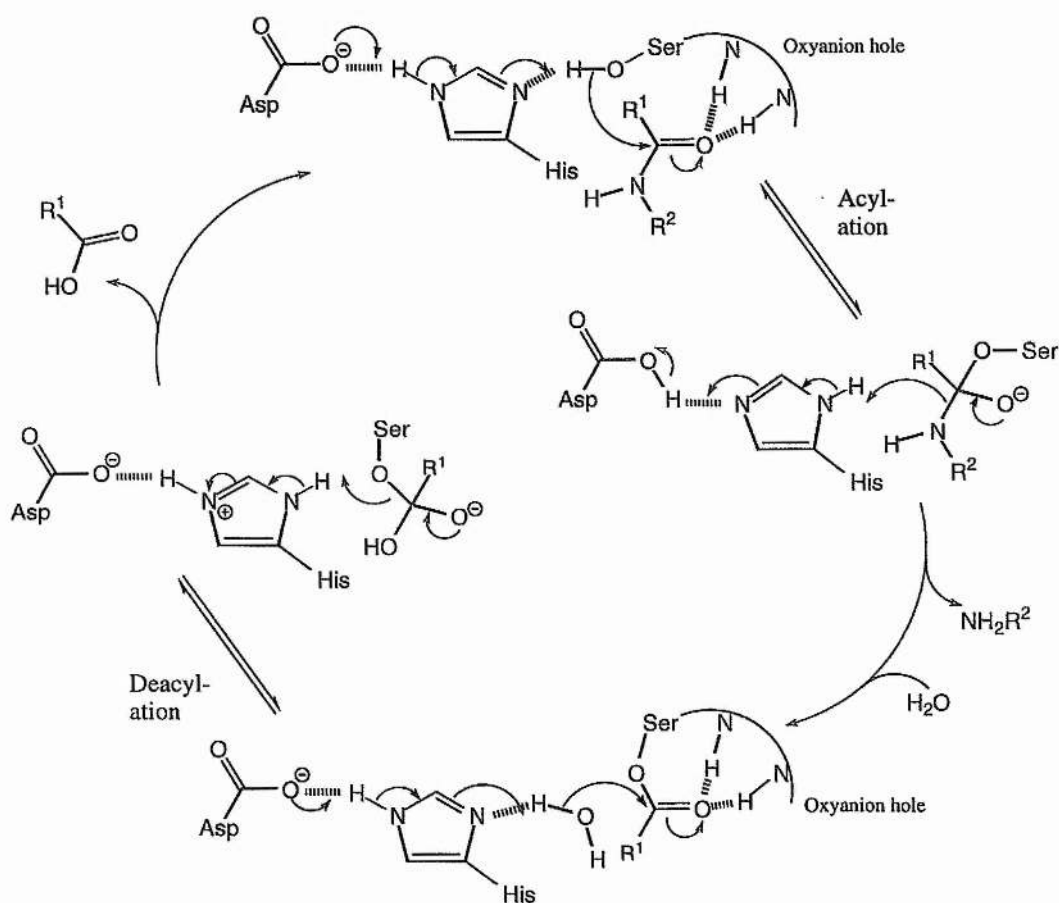
Figure 1.3.1 Mechanism of action of a serine protease.

21

Cysteine or thiol proteases, such as papain, contain an unusually nucleophilic active site cysteine residue. These proteases have a number of features in common with serine proteases, also proceeding *via* an acyl-enzyme intermediate, with an imidazole group assisting in the nucleophilic attack. Cysteine proteases also possess an oxyanion-hole, although its role in catalysis appears to be somewhat less than in serine proteases (Ménard *et al.*, 1991). Unlike serine proteases however, the thiol readily forms an ion pair with imidazole in the ground state (Figure 1.3.2) (Kollman, 1992; Phillips and Fletterick, 1992). A neutral asparagine stabilises the transition state.

Peptide bond hydrolysis by a metalloprotease proceeds *via* a general base hydrolysis mechanism (Figure 1.3.3). Metalloproteases, all of which contain a zinc atom in the catalytic centre, have been recently classified into distinct family groups depending upon their structural and sequence similarities (Hooper *et al.*, 1994). Carboxypeptidase A is the classic example of a metalloprotease. Proteolysis by all metalloproteases proceeds *via* a similar route to that shown in Figure 1.3.3 for carboxypeptidase A. A zinc atom is co-ordinated to one glutamate and two histidine residues, and a water molecule. Another glutamate acts as a general base to removes a proton from the zinc-bound water molecule, which attacks the scissile carbonyl group to form the tetrahedral oxyanion. This is stabilised by a positively charged arginine residue. In contrast to the serine and cysteine proteases, at no stage in catalysis is the substrate bound covalently to the metalloprotease.
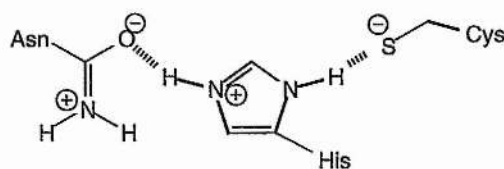
Figure 1.3.2 Catalytic residues of a cysteine protease.
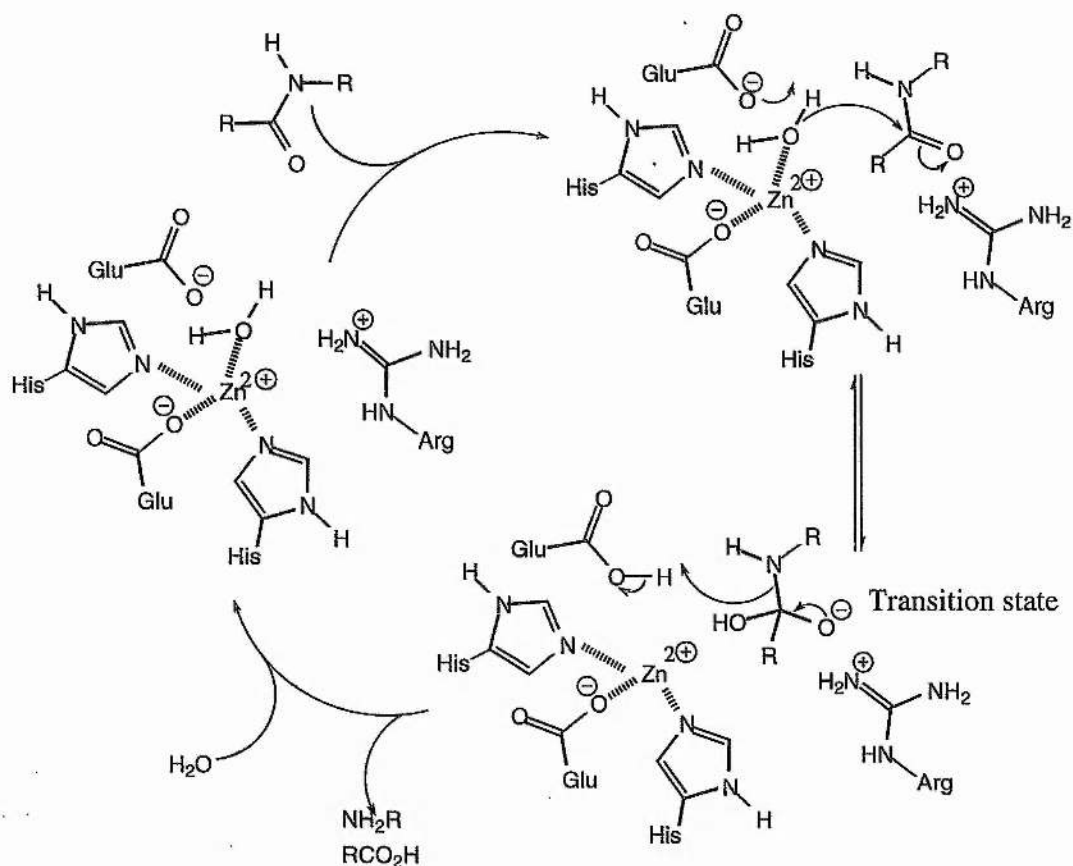
Figure 1.3.3 Catalytic mechanism of the metalloprotease, Carboxypeptidase A.

The mechanism of action of the aspartate or acid family of proteases has been the subject of much discussion, but it is generally believed that they also act *via* general base catalysis in which a carboxylate activates a water molecule by hydrogen bonding (Figure 1.3.4) (Phillips and Fletterick, 1992; James and Sielencki, 1985; Davies, 1990; Pearl, 1987). The first step in many proposed mechanisms is the protonation of the scissile carbonyl oxygen by the proton shared between the two aspartic acid residues (James and Sielencki, 1985; Suguna *et al.*, 1987; Davies, 1990; Phillips and Fletterick, 1992). Crystallographic studies with inhibitors, however, also indicated that the carbonyl group is instead hydrogen bonded to by main-chain amide groups of the ß-hairpin "flaps" overlying the binding cleft (Pearl, 1987). It was suggested that the extended sub-site specificity of these enzymes produces strain energy which twists the peptide bond out of planarity, reducing the C-N double bond character and polarising the carbonyl bond

23

(Pearl, 1987; Suguna *et al.*, 1987). Nucleophilic attack of the carbonyl-carbon atom by water, with concomitant transfer of its proton to an aspartate follows. This would result in the formation of an uncharged tetrahedral intermediate if the carbonyl group was protonated by an aspartic acid residue. Otherwise, according to Pearl, the tetrahedral intermediate should be a tetrahedral oxyanion. The formation of products is completed by the collapse of the intermediate and protonation of the leaving group amine.
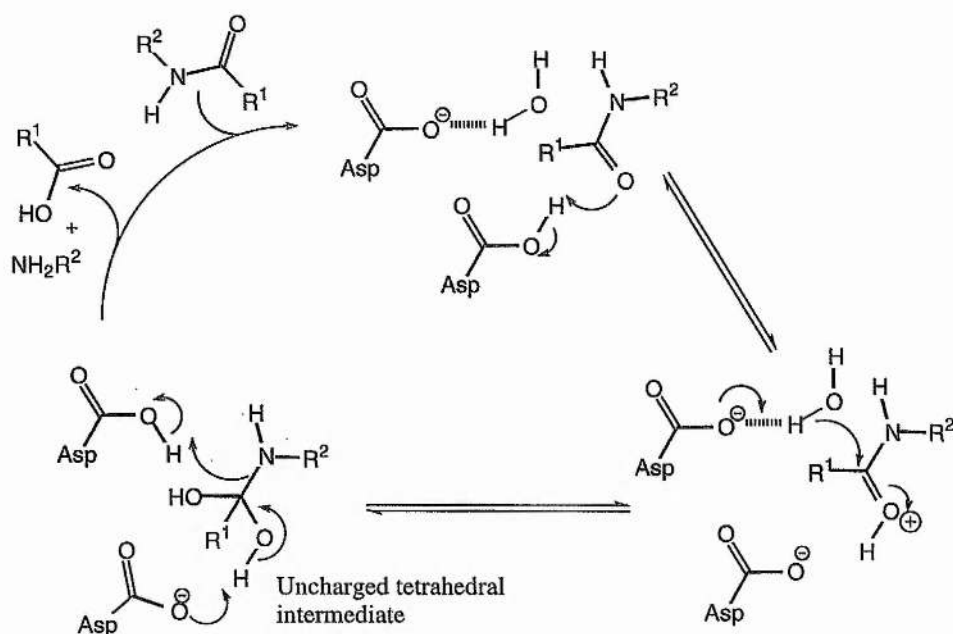
Figure 1.3.4 Catalytic mechanism of aspartate proteases. This shows the generally accepted mechanism of aspartate proteases in which an aspartic acid residue protonates the carbonyl group of the scissile bond.

Proteasomes, which have been found in eukaryotes, are protein complexes with many catalytic activities. In the case of the 20S proteasome, from the archaebacterium *Thermoplasma acidophilum*, the active sites were found to be of a single type, with the differing substrate-specificities imparted by the binding pockets. It was determined that the proteasome did not belong to the usual classes of protease, and site directed mutagenesis was then used to pinpoint the N-terminal threonine residue as the active-site nucleophile and an essential lysine group as being involved in the catalysis (Seemüller *et al.*, 1995). This is the only threonine protease characterised thus far.

24

1.3.2   Virus encoded proteases

One or more protease enzymes, of host or virus origin, are often involved in the proteolytic processing of a virus during its replicative cycle. Most of these are involved with gene expression, in the processing of a high molecular weight polyprotein into functional gene products. Some are also involved in the maturation process of structural proteins during virion assembly. The use of virus-encoded enzymes rather than host proteases allows less dependence on a host cell. Host cell proteases are in general not specific enough and are subject to competition by cellular protein substrates. Many viruses replicate in the cytoplasm, where there seem to exist very few specific, regulatory proteinases which could participate in viral proteolytic processing. The power and flexibility of this strategy is reflected very clearly in the overall success of the positive strand RNA viruses; in their numbers and the number of different hosts which they infect. Enigmatically, however, a total dependence on virally encoded proteases, which are not usually susceptible to regulation by the host cell inhibitors, and are unable to attack host cell proteins, could turn out to be the "Achilles heel" of many viruses. The design of a synthetic inhibitor to specifically block the action of a viral protease without interfering with host cell proteases is an extremely attractive and feasible target for anti-viral therapy, not only for single viruses but for whole families of viruses. Indeed, such a strategy has proved to be clinically useful in the treatment of AIDS (Fitzgerald, 1993); to date four HIV-1 protease inhibitors have been approved for clinical use by the Food and Drug Administration of the United States of America (McDonald and Kuritzkes, 1997).

The adoption of a polyprotein strategy reduces the genetic control of protein expression, eliminating the potential for temporal control of expression by transcription. However, subtle control of proteolysis through different pathways provides an alternative to this method of control. Spatial effects throughout the cell may also be employed; the gathering of associated proteins, into discreet regions of the genome such that they can be separately expressed in multipartite genomes or cleaved apart in polyproteins is commonly adopted by viruses. This enables proteins to be initially concentrated within

25

local areas of the cell encouraging their interaction early in replication. Very often a viral protease cleaves between specific amino acid pairs, but not all of the available pairs are actually cleaved, since other factors, including the nearby primary sequence and tertiary structure at scissile bonds, may also be important. The variation in these factors at cleavage sites may reflect differences in the efficiency with which they are cleaved, which could also be a method of regulating gene expression. Attention will be drawn to the use of such controls over proteolysis in the following examples of viral proteases.

Virally encoded proteases that have been characterised to date are thought to be related, either in a structural or mechanistic fashion, to their cellular counterparts. However, most are smaller than their counterparts, and usually display a lower turnover number and a higher degree of substrate specificity. Such proteases are encoded by many RNA viruses, all retroviruses, and many DNA viruses (Hellen and Wimmer, 1992). Picornaviral and retroviral proteases are currently the subjects of considerable interest. Both of these types are synthesised as integral parts of the polyprotein. In picornaviruses the proteases perform proteolytic cleavages co-translationally *in cis* and post-translationally both *in cis* and *in trans* modes. The retroviral protease, being active as a dimer (see later), is cleaved from its polyprotein *in trans*. Proteases have also been described for eukaryotic DNA viruses of the families *Herpesviridae* and *Adenoviridae* (Donaghy and Jupp, 1994). Most of these enzymes belong to either the serine, cysteine, or aspartate classes of protease enzyme. Additionally a serine-like cysteine protease is also found amongst virus encoded proteases. A viral metalloprotease activity has been suggested for hepatitis C virus (Lohmann *et al.*, 1996). For reviews see Dougherty and Semler (1993), Krausslich and Wimmer (1988) and Wellink and van Krammen (1988).

Viral aspartate proteases are functionally and structurally related to cellular aspartate proteases. However there are some notable differences. The cellular aspartate proteases are on average approximately 325 residues in length whereas the viral proteases range from 99 to 125 amino acids. The cellular aspartate proteinases are bi-lobal proteins containing two stereochemically equivalent domains, which are thought to have arisen by

26

gene duplication. A retroviral aspartic acid protease is required for retroviral polyprotein processing. Crystallographic studies, however, of the HIV-1 protease have shown the active form to be a dimer composed of two identical subunits, which constitute the two-domain-like structure (Navia *et al.*, 1989). Retroviral polyproteins are not cleaved immediately after synthesis but are instead transported to the plasma membrane and normally processed once budding of the immature particles occurs. The gag polyprotein, which encodes the structural precursor, does not require to be proteolytically processed in order to assemble immature virus particles (Kohl *et al.*, 1988). The N-terminal glycine residues of the gag and gag-pol polyproteins are myristylated, and this modification directs the polyproteins to concentrate on the membrane (di Marzo Veronese *et al.*, 1988). It is thought that this may promote the dimerisation and subsequent activation of the protease. Premature processing of the polyprotein is undesirable because the liberated polyprotein domains would not be myristylated and would hence be free to diffuse back into the cytoplasm. Once released, the protease is confined to the budding particle and can initiate a cascade of cleavage events which results in the maturation of the virus particle.

Virus-encoded serine proteases are found in the flavi- and alphaviruses and are involved in the processing of the single polyprotein produced by the genomic mRNA and the capsid polyprotein, respectively. Only one serine protease, the P1 protease of potyviruses, has been found within the picornavirus super-group (Ryan and Flint, 1997). These proteases are functionally similar to their previously described cellular counterparts.

A serine-like cysteine protease has been identified in picornaviruses (Gorbalenya *et al.*, 1989). In all picornaviruses the viral protein 3C is the major source of proteolytic activity (see section 1.5.3). It is responsible for the final primary cleavage event, between 2C and 3A (see Figure 1.1.1), and all secondary proteolysis (Palmenberg and Reuckert, 1982). It cleaves itself from its related precursors through a series of progressive auto-catalytic reactions. The 3C protease (3C^{pro}) is also active, and sometimes required, as a

larger precursor polypeptide 3CD^{pro} (Jore *et al.*, 1988). The 3C^{pro} region was found to be sufficient to cleave specific junctions in P2, but was required in the form 3CD^{pro} to process the structural precursor P1 (Ypma-Wong and Semler, 1987). This is an example of the use of alternative pathways in the processing of viruses.

On the basis of inhibitor studies, amino acid sequence comparisons and a site-directed mutagenesis study, the active site nucleophile was assumed to be a highly conserved cysteine (Pelham, 1978). Sequence alignments and structural molecular modelling, however, identified a significant homology to the cellular serine proteases (Gorbalenya *et al.*, 1989; Mathews *et al.*, 1994; Allaire *et al.*, 1994). The 3C protease is thus assigned as a trypsin-like protease in which a cysteine replaces serine in the catalytic triad along with histidine and aspartic acid. The picornavirus 2A protease is also thought to be a similar trypsin-like protease (Bazan and Fletterick, 1988). Similar proteases have also been found in poty-, como- and nepoviruses (Bazan and Fletterick, 1988).

Viral proteases functionally and structurally similar to their cellular cysteine homologues have only recently been identified in the poty- and alphaviruses (Oh and Carrington, 1989; Strauss *et al.*, 1992). The HC-Pro of potyviruses was demonstrated to be similar to a cellular cysteine protease. Following this, the non-structural protein nsP2 of sinbis virus was suggested to be a cysteine-like protease. The primary amino acid spacing between the histidine and cysteine appears to be shorter than in such cellular enzymes, but amino acid motifs are shared in both cellular and virus-encoded enzymes.

The L-proteinase of the aphthoviruses, which is responsible for its own cleavage from the polyprotein, between L and P1 (Strebel and Beck, 1986), was also proposed to be a cysteine protease (Gorbalenya *et al.*, 1991) and this was confirmed by subsequent site-directed mutagenesis (Roberts and Belsham, 1995).

Section 1.4 Enzyme-independent protein cleavages

A number of hydrolyses of peptide bonds cannot be explained in terms of proteolytic processing by a protease. These include the phenomena of protein splicing and deamidation with peptide bond cleavage. Also included in this category is the 2A region of the FMDV polyprotein which mediates a *cis*-acting hydrolysis of the peptide bond at its C-terminus, however, this will be discussed later in section 1.5.6.

1.4.1   Protein splicing

The linear expression of a gene can be interupted by unique segments of information encoded by the gene. The best studied of these special segments are the RNA introns. Recently protein introns called inteins have been discovered (For reviews see Cooper and Stevens, 1993 and 1995; and Shub and Goodrich-Blair, 1992). Inteins are protein sequences embedded in-frame within a protein precursor which are removed by protein splicing. This post-translational reaction involves the precise excision of an intervening protein sequence from a precursor protein, coupled to peptide bond formation between the flanking N- and C-terminal domains, to give a spliced protein product. Two separate products are produced from such a process, each of which performs a distinct purpose. The excised internal protein is encoded by a DNA insertional element. Once excised the protein can act as an endonuclease to mediate the movement of the element elsewhere in the genome. The host protein performs its original function as encoded by the gene before the insertional event.

Protein splicing has been found to occur both in eukaryotes, in *Saccharomyces cerevisae* and *Candida tropicalis* (Gu *et al.*, 1993; Davis *et al.*, 1991) and also in the prokaryotes, *Mycobacterium tuberculosis* and *Mycobacterium leprae* of the eubacteria (Cooper and

29

Stevens, 1995; Davis *et al.*, 1991), and *Thermococcus litoralis* and *Pyrococcus sp.* of the archaebacteria (Perler *et al.*, 1992; Xu *et al.*, 1993; Figure 1.4.1).

The first example to be reported was that of the yeast *S. cerevisae TFP1* gene which encodes the 69 kDa catalytic subunit of the vacuolar H$^+$-ATPase (Kane *et al.*, 1990). A continuous ORF predicted a 119 kDa protein, however, translation of the ORF produced a 454 amino acid, 50 kDa, intein and the spliced 69 kDa V-ATPase in equal ratios.

The *RecA* gene of *M. tuberculosis* and *M. leprae* were also found to contain inteins which were spliced from the precursor protein by a reaction similar to that of the *TFP1* gene (Davis *et al.*, 1991). The genes encoding DNA polymerases from *T. litoralis* and *Pyrococcus sp.* each contain two inteins (Perler *et al.*, 1992; Xu *et al.*, 1993).

Sequence comparisons of the known splicing events have revealed a thiol or hydroxyl at each extein border and the sequence His-Asn-Ser/Cys/Thr at the border between intein and C-terminal extein. All of these conserved residues are essential for the splicing event (Xu and Perler, 1996). Amino-terminal protein sequencing of an intermediate produced in the splicing reaction demonstrated that it was branched with amino-terminal residues of the N-extein and the intein free (Xu *et al.*, 1993). Various mechanisms have been proposed and are reviewed in Cooper and Stevens (1995). Recently site-directed mutagenesis of these conserved residues has clarified the situation and indicated that the mechanism shown in Figure 1.4.2 seems most plausible (Shao *et al.*, 1996; Xu and Perler, 1996). Shao *et al.* substituted a cysteine for the serine of the N-terminal extein (and an alanine for the cysteine residue of the C-terminal extein) and found that the linear intermediate formed was more labile in the presence of hydroxylamine. Since thioesters are known to be more rapidly hydrolysed by hydroxylamine than oxygen esters this supported the hypothesis that N-O acyl rearrangement occurs as the initial step. The serine residue was shown to be necessary for the N-terminal cleavage event and the formation of a branched intermediate. The serine residue at the C-terminal extein was

necessary for the formation of the branched intermediate. Mutation of the histidine residue caused the accumulation of the branched intermediate, showing that it was necessary for cleavage of this intermediate. Mutation of the N-terminal asparagine still permitted cleavage of the N-terminal extein, but no splicing of the branched intermediate or cleavage of the C-terminal extein was seen. This evidence supports the order of intermediates produced in Figure 1.4.2, and the residues involved. It has been suggested that the histidine residue assists in the asparagine cyclisation, probably acting as a proton donor or acceptor.
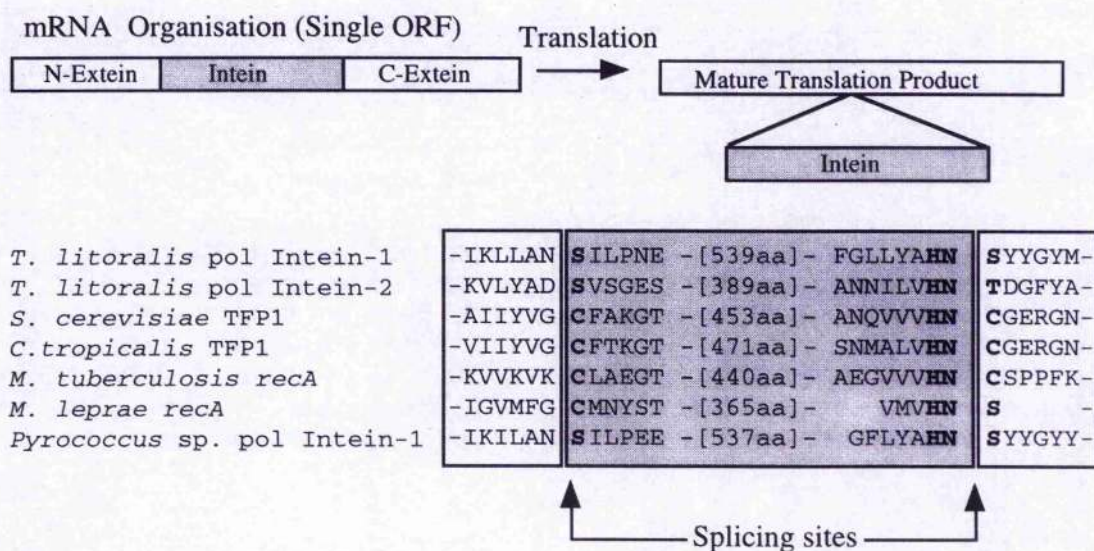
Figure 1.4.1 Known examples of protein splicing. The amino acid sequences around the splicing sites are shown, with those residues which are totally conserved shown in bold.
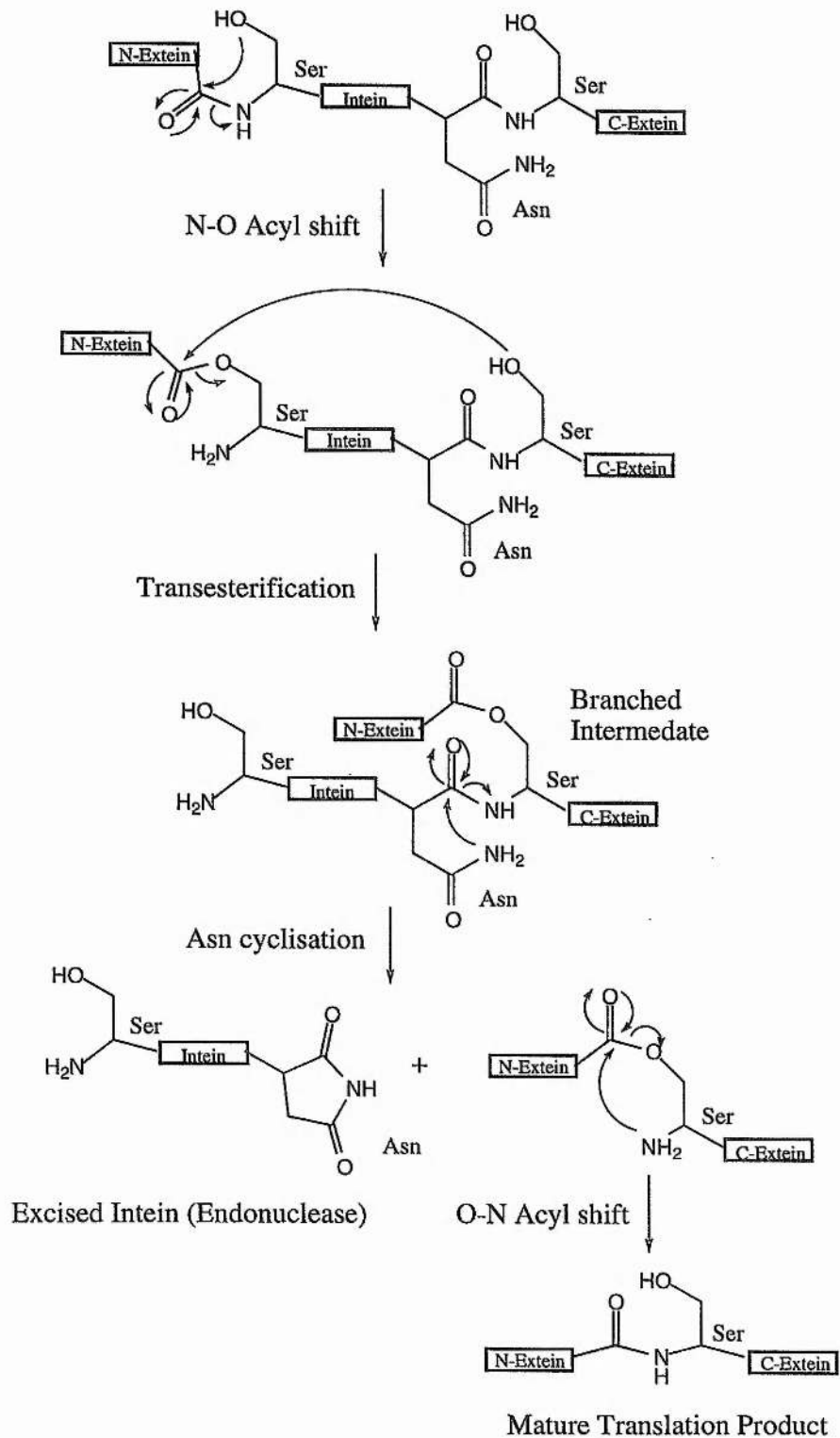
Figure 1.4.2 Mechanism of protein splicing (Xu and Perler, 1996). This mechanism is well supported by site directed mutagenesis data. The presence of an ester bond, branched intermediate, and succinimide intermediate as well as the involvement of the residues shown have all been demonstrated.

1.4.2   Deamidation-like peptide bond hydrolysis

Deamidation is a hydrolytic reaction requiring only water to convert an asparagine or glutamine residue into aspartate (or α-isoaspartate) or glutamate respectively. It has been identified in a wide range of proteins and peptides and is thought to be of physiological significance in the degradation of proteins. In many systems deamidation of both glutamine and asparagine contribute to protein "ageing" and accelerate protein turnover, especially for long lived proteins. The proteins of the eye lens, such as α-crystalin, are well known for their longevity, and have been thoroughly studied for deamidation (van Kleef et al., 1975; Voorter et al., 1988). A similar reaction has also been found to occur by which the main-chain peptide bond, rather than the side-chain amide bond, is hydrolysed.

Mechanisms for the deamidation of asparagine have implicated a cyclic succinimide intermediate which undergoes hydrolysis to products. A generally accepted mechanism for the deamidation of peptides with Asn-Gly, Asn-Ser and Asn-Ala sequences is the β-aspartyl shift mechanism (Wright, 1991(a); Klotz and Thomas, 1993; Bischoff and Kolbe, 1994). The main-chain amide nitrogen of the following residue functions' as a nucleophile attacking the side-chain amide (Figure 1.4.3 (a)). The resulting amino-succinimidyl residue is unstable in aqueous solution, and can be rapidly hydroysed on either side of the nitrogen to give two isomeric products, α-Asp or α-isoAsp, which differ in the main-chain connection of the peptide backbone (i.e. through the α- or β-aspartate carbonyl group). The deamidation of asparagine containing tetrapeptides has shown that direct solvent hydrolysis does not occur to an appreciable extent (Capasso et al, 1993). At low pH the rate of hydrolysis increases with hydroxide ion concentration indicating that the reaction is specific-base catalysed. The rate determining step is the cyclisation to form the cyclic intermediate. However, under neutral and basic conditons the cyclisation involves the formation of an anionic intermediate by the abstraction of an amide proton followed by the cyclisation to the imide. Under these conditions the collapse of the imide

was found to be the rate limiting step. As a general rule Asn deamidates more easily than Gln (Bischoff and Kolbe, 1994). This is consistent with the proposed mechanism which would proceed more slowly for the formation of the sterically less favoured 6-membered cyclic imide, compared to the five membered cyclic imide. The β-aspartyl shift mechanism requires that the residue that is immediately downstream of asparagine is small in order to permit cyclisation. Indeed, sequences surrounding known deaminating asparagine residues in proteins have been found show very high frequencies for glycine and serine C-terminal to the asparagine (Wright, 1991(b)). A relatively high occurrence of serine and threonine, both general acids within the specialised environments of proteins, either side of the asparagine might indicate that another mechanism may also mediate deamidation. Wright and Robinson have suggested the hydroxyl groups of these residues accelerate deamidation (1982). Kossiakoff studied the deamidating Asn residues of trypsin, using X-ray crystallography (1988), in which the conformation of the deamidated residues was clearly distinguished from other Asp residues by hydrogen bonding to the n+2 main-chain amide through the side-chain amide oxygen. He also suggested that a C-terminal serine hydroxyl group might hydrogen bond to the main-chain nitrogen making the amide proton more acidic.

A similar deamidation-like cleavage of peptide bonds has been shown to occur immediately after an asparagine residue (Voorter *et al.*, 1988; Blogett *et al.*, 1985; Klotz and Thomas, 1993; Geiger and Clarke, 1987). Whilst this is not a deamidation reaction of the side-chain asparagine its mechanism is similar to the β-aspartyl shift mechanism. The side-chain amido-NH_2 group attacks its own main-chain α-carbonyl C-atom with displacement of the downstream peptide chain from the α-carbonyl side of the labile asparagine to give an amido-succinimide as shown in Figure 1.4.3 (b).

Synthetic peptides containing a N^5-methyl asparagine, which imparts a stronger side-chain carbon-nitrogen bond and sterically hinders attack at the side-chain β-carbonyl C-atom, demonstrated a deamidation-like reaction to produce almost three times as much

main-chain hydrolysis reaction as side-chain deamidation (Klotz and Thomas, 1993). The activation energies for each pathway were similar, however entropic factors favoured the main-chain cleavage.

(a)

α-isoAsp-Xaa   α-Asp-Xaa

(b)
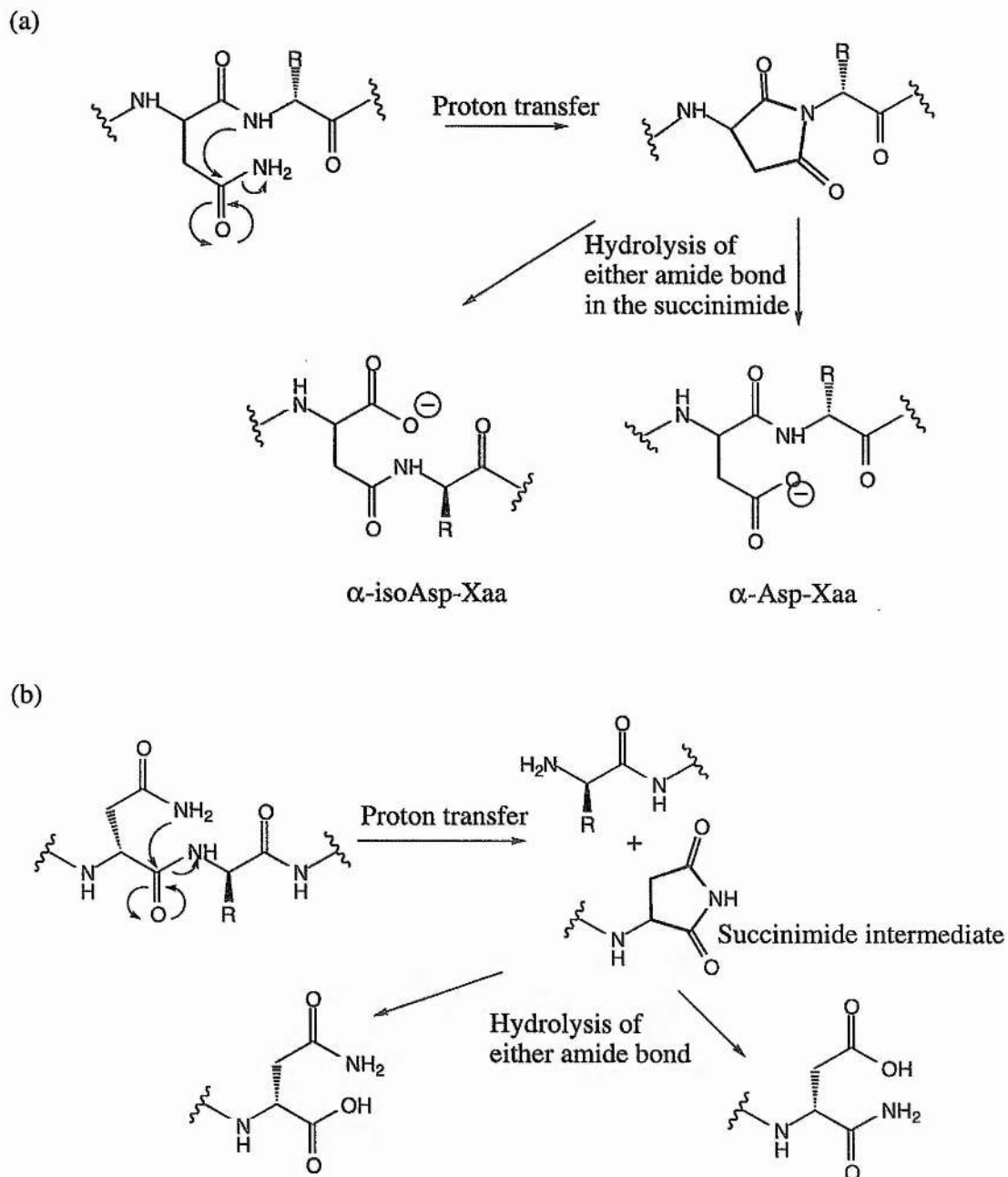
Figure 1.4.3 (a) The β-aspartyl shift mechanism of deamidation

(b) Deamidation- type reaction with peptide bond cleavage

Both reactions are very similar, and both occur for the same residue. The β-aspartyl shift deamidation reaction, however, is usually lower in energy, and predominates.

Section 1.5 Picornavirus polyprotein processing

The picornavirus polyprotein, as described in section 1.2.4.2, is processed entirely by virus encoded proteinases (for a review see Reuckert, 1996). The aphthovirus proteins are proteolytically cleaved by the L protease, 2A proteolytic element and the 3C protease. In rhino-, entero- and cardioviruses, the 2A and 3C proteases mediate proteolytic processing of the polyprotein. In hepatoviruses the 3C protease is the only proteolytic activity yet identified. The 2A region of hepatoviruses does not appear to have any proteolytic activity. An additional proteolytic cleavage, of a capsid protein precursor VP0 to VP4 and VP2, called the maturation cleavage, occurs by an unknown mechanism.

1.5.1 The L protease

A leader protein has been found in both cardio- and aphthoviruses, encoded at the amino-terminus of the polyprotein before P1. The leader protein of the aphthoviruses, also known as the L protease (L^{pro}), has been shown to cleave co-translationally at its own C-terminus, separating itself from the rest of the polyprotein (Strebel and Beck, 1986). The leader protein of cardioviruses has not been demonstrated to have a proteolytic activity, the 3C protease being responsible for the L-P1 cleavage event. In addition to this autocatalytic activity, L^{pro} can behave as a protease cleaving in an intermolecular fashion (in *trans*) (Devaney *et al.*, 1988). It is this activity which is responsible for the shut off of host cell protein synthesis. In the eukaryotic cell mRNA contains a cap structure which consists of 7-methyl guanosine covalently linked by a 5'-5'-linkage to the terminal base of the mRNA. Recognition of this cap is required for translation of eukaryotic cell proteins. A cap-binding complex (eIF-4F) is involved in the assembly of an initiation complex involving the 40S ribosomal subunit, which scans along mRNA for a start codon, upon which protein synthesis begins. Picornaviruses have no such cap-structure at the 5'-end of their RNA; they have a small protein (VPg) covalently attached to the 5'-end. They

36

must, therefore, recruit ribosomes in a cap-independent manner, and for this they have evolved an internal ribosome entry site (IRES). This enables them to selectively suppress initiation of translation of host protein synthesis, by cleaving the p220 (or eIF-4γ) component of the cap-binding complex. In aphthoviruses this is mediated by L^{pro}, whereas in entero- and rhinoviruses 2A^{pro} is responsible. The L^{pro} of cardioviruses is not involved this activity and no such rapid shutoff of host cell protein synthesis is seen.

Two forms of the L protein, Lab^{pro} and Lb^{pro}, are synthesised in FMDV-infected cells. These have resulted from the use of alternative translation start codons (Clarke *et al.*, 1985; Sangar *et al.*, 1987). The Lb form undergoes post-translational modification to a Lb' form. No functional differences have been found for the various forms of L (Medina *et al.*, 1993).

Amino acid sequence alignments suggested that L^{pro} was a cysteine protease related to papain (Gorbalenya *et al.*, 1991). Inhibitor studies and site directed mutagenesis of the active site cysteine demonstrated that L was indeed a cysteine protease (Kleina and Grubman, 1992; Roberts and Belsham, 1995).

1.5.2 The 2A protease of entero- and rhinoviruses

In the entero- and rhinoviruses a primary cleavage occurs at the N-terminus of the 2A region, generally between a tyrosine-glycine scissile pair, separating the capsid protein precursor, P1, from the precursor of the replicative proteins, P2-P3 (Sommergruber *et al.*, 1989; and Toyoda *et al.*, 1986, respectively). In both cases the 2A^{pro} was identified as mediating the cleavage. The 2A protease (2A^{pro}) was also found to mediate a second proteolysis between tyrosine and glycine, within the 3D protein of 3CD to yield 3C' and 3D' (Toyoda *et al.*, 1986). The significance of this reaction is not clear, and the products are not necessary for successful virus replication *in vitro* (Lee and Wimmer, 1989).

Sequence alignments of the entero- and rhinoviruses with the cellular subtilisin-like serine proteases suggested that the 2A^{pro} had a catalytic triad composed of His, Asp and Cys, where cysteine was the active site nucleophile rather than serine (Bazan and Fletterick, 1988). Inhibitor studies also supported the hypothesis of a cysteine serine-like protease (Konig and Rosenwirth, 1988). Recently the 2A^{pro} was found to contain a structural zinc ion which was essential for activity (Sommergruber *et al.*, 1994; Voss *et al.*, 1995).

The task of host cell protein synthesis shut-off, in entero- and rhinovirus infected cells, is effected by the 2A^{pro} in a similar fashion to L^{pro} in aphthovirus infected cells (as described in section 1.5.1; Bernstein *et al.*, 1985).

1.5.3 The 3C protease

All picornaviruses possess a 3C protease (3C^{pro}), which shows a high degree of similarity across the genera (Ryan and Flint, 1997). With the exception of the L^{pro}, 2A^{pro}, and maturation cleavage activities the 3C protease is responsible for all other picornavirus proteolytic processing activities. Like L^{pro} and 2A^{pro}, 3C^{pro} is responsible for a co-translational primary cleavage event, although in this case cleavage occurs at a site distal to 3C, at the 2C-3A junction (see Figure 1.1.1). Apart from the maturation cleavage, the 3C^{pro} is the only proteolytic activity to have been found for hepatoviruses. The 3C^{pro} cleaves itself from its related precursors through a series of progressive auto-catalytic reactions. Sequences extraneous to 3C within these precursors may act to modulate the specificity of the 3C^{pro}. Indeed, the 3C protease is also active, and sometimes required, as a larger precursor polypeptide 3CD^{pro} (Jore *et al.*, 1988); the 3C^{pro} was found to be sufficient to cleave specific junctions in P2, but was required in its precursor form 3CD^{pro} to process the structural precursor P1 (Ypma-Wong and Semler, 1987).

In 1978 it was shown that the activity of 3C^{pro} was inhibited by thiol protease inhibitors (Pelham, 1978). Sequences analyses identified the 3C^{pro} of picornaviruses to be a serine-like cysteine protease (Gorbalenya *et al.*, 1989). This was recently confirmed, when the X-ray crystal structure of the 3C^{pro} was solved for both HRV-14 (Mathews *et al.*, 1994), and Hepatitis A (Allaire *et al.*, 1994). The structures of both were found to be similar to the predicted chymotrypsin-like serine proteases. In the case of HAV 3C^{pro}, however, there appeared to be only a cysteine-histidine dyad in the catalytic site; an aspartate residue did not interact as expected with the catalytic histidine residue, but with the side-chain of a lysine residue and a main-chain amide group.

In poliovirus, 3C^{pro} mediates cleavages between 9 out of 13 glutamine-glycine pairs (with a preference for those with Ala, Thr, Val or Pro residues in the P4 position). However, it is less specific in aphthoviruses and mediates proteolysis between E-G, E-T, Q-L, and Q-I pairs (Robertson *et al.*, 1985). The preferences for cleavage at specific sites possibly allows some control over the order of proteolytic processing.

1.5.4 The maturation cleavage

The maturation cleavage is the final proteolytic event of the picornaviral replication cycle. The picornaviral capsid is composed of 60 copies of each of the four structural proteins. The P1 polyprotein, comprising a capsid protomer, probably folds itself into the tertiary structure of a protomeric unit, prior to proteolytic processing by the 3C protease between 1AB (VP0) and 1C (VP3), and between 1C and 1D (VP1). The processed protomers have been shown to self-assemble into pentamers *in vitro* (Palmenberg, 1982), twelve of which are required to form a capsid structure. After capsid assembly, and either concomitant with, or after association with viral RNA, most of the VP0 molecules (all except 1 or 2) are cleaved into 1A (VP4) and 1B (VP2) at an Asn-Ser scissile bond (Hellen and Wimmer, 1992). This cleavage is associated with the acquisition of

infectivity and greater virion stability. In empty capsids VP0 is uncleaved indicating that RNA is required for this cleavage. The cleavage site is located towards the interior of the side of the capsid, inaccessible to external proteases. Since picornaviruses are not known to package any of their proteases, it was postulated that the maturation cleavage was an autocatalytic event for which the presence of RNA was necessary. The mechanism for the maturation cleavage remains unknown.

A similar autocatalytic and assembly dependent cleavage occurs in insect nodaviruses. This cleavage is also associated with acquiring increased stability and infectivity. The mechanism of this event has been determined to require an assembly dependent strain, and a critical aspartic acid residue. Zlotnick *et al.* (1994) have suggested that the aspartic acid, lying within a hydrophobic area, should possess a side-chain displaying an elevated pKa and is, therefore, protonated. This carboxylic acid group could hydrogen bond to the oxygen atom of the carbonyl moiety of the scissile bond to increase its susceptiblity to attack by water.

1.5.5 Proteolysis in *cis* and *trans*

In picornaviruses and many other viruses proteolytic processing can proceed in both *cis* and *trans* modes. Early in infection the concentration of viral proteins is low, and therefore, it is kinetically more likely that processing in *cis* will dominate. Later in the infection the concentration of viral proteins will be greater and the amount of *trans* processing will increase. This could allow for a temporal switch in processing from *cis* to *trans* enabling different polyproteins to be predominant at different stages in replication.

1.5.6 The 2A region of aphtho- and cardioviruses

Picornaviral mRNA encodes a single polyprotein which is then proteolytically processed into many mature products. In entero- and rhinoviruses the capsid protein precursor P1

40

is separated from the replicative proteins by a rapid co-translational proteolytic cleavage mediated by the 2A protease itself as discussed earlier.

In cardio- and aphthoviruses a rapid co-translational cleavage is also mediated by 2A, but at the C-terminus of the 2A region (*c.f.* cleavage is N-terminal to 2A in entero- and rhinoviruses), releasing L-P1-2A and P1-2A, respectively, from 2BC and P3 such that precursors spanning the 2A / 2B cleavage site are not detected during native polyprotein processing (Sangar *et al.*, 1977; Ryan *et al.*, 1989; Palmenberg *et al.*, 1990).

Cleavages between 2A and 2B in aphthovirus and TMEV polyproteins were shown to require neither L nor 3C, the known proteolytic activities of these viruses (Vakharria *et al.*, 1987; Batson and Rundell, 1991; Clarke and Sangar, 1988; Roos *et al.*, 1989; Ryan *et al.*, 1989). In the case of EMCV, cleavage at this site has been shown to be mediated by a 2A protease of 143 amino acids in length (Jackson, 1986). It is interesting to note that the deletion of the N-terminal region of EMCV 2A reduces but does not destroy its activity (Palmenberg *et al.*, 1992). Analyses of recombinant FMDV polyproteins and artificial reporter gene polyproteins showed that a 20 amino acid oligopeptide sequence which contained the FMDV 2A region and the N-terminal proline of 2B was able to mediate a co-translational proteolytic cleavage (Ryan *et al.*, 1991; Ryan and Drew, 1994). An even shorter 16 amino acid region of FMDV 2A was also shown to mediate cleavage at its C-terminus when inserted in the foreign context between CAT and NA in a chimeric influenza virus (Percy *et al.*, 1994).

The 2A region is released from the capsid protein precursor by an intermolecular proteolytic cleavage mediated by the 3C^{pro} activity (Jackson, 1986; Clarke and Sangar, 1988). Although the 2A region is probably only 18 amino acids in length (see below) it is important for the virus that it is cleaved from 1D since its conserved nature, and position on the capsid's surface would make it an obvious immunogenic target (see Figure 1.5.1 for the location of the C-terminus of 1D). The C-terminus of capsid protein 1D was

41

determined by C-terminal degradation of radio labelled 1D protein (Bacharach *et al.*, 1973; Kurz *et al.*, 1981). N-terminal amino acid sequencing of the 2B protein then determined that the FMDV 2A region released must be only 16 amino acids in length (Robertson *et al.*, 1985). However, the data for the sequencing of the C-terminus of 1D seems inconclusive, and indeed the results of Kurz *et al.* for the amino acid composition of 1D appear to be equally consistent with a 2A region of 18 amino acids. A 2A region of 16 residues would mean that the 3C[pro] had cleaved an unprecedented Leu-Asn pair. The known scissile amino acid pairs recognised by FMDV 3C[pro] are Glu-Gly, Glu-Thr, Gln-Leu, and Gln-Ile (Robertson *et al.*, 1985). Moreover, a more typical amino acid pair for recognition by 3C[pro], Gln-Leu, is present which would result in an 18 amino acid 2A region. In the case of ERV, by comparison with FMDV, it has also been generally accepted that the 2A region is 16 amino acids in length although the 3C[pro] is then cleaving an unprecedented Thr-Asn pair (Wutz *et al.*, 1996; Li *et al.*, 1996). However, examination of the protein sequences of both ERV serotypes reveals that more typical 3C[pro] cleavage sites, Gln-Cys (ERV-1) and Glu-Ala (ERV-2), are present which would result in 2A regions of 18 and 19 amino acids respectively. Further, in support of FMDV 3C[pro] cleaving between the Gln-Leu pair, the x-ray crystal structure for the capsid of FMDV serotype C shows electron density only to the glutamine (Gln[211]) of the proposed scissile bond (see Figure 1.5.1; Lea *et al.*, 1994).

The aphtho- and cardiovirus 2A regions are obviously not equivalent to the 2A proteases of rhino- and enteroviruses. An additional function of 2A in entero- and rhinoviruses is the cleavage of p220, resulting in host cell translation shut off; in aphthoviruses the leader protein L[pro] cleaves p220 (Devaney *et al.*, 1988); in cardioviruses such host cell translation shut off is not seen. The 2A regions of aphthoviruses and cardioviruses show no homology with the 2A regions of entero- and rhinoviruses and display none of the known protease motifs.
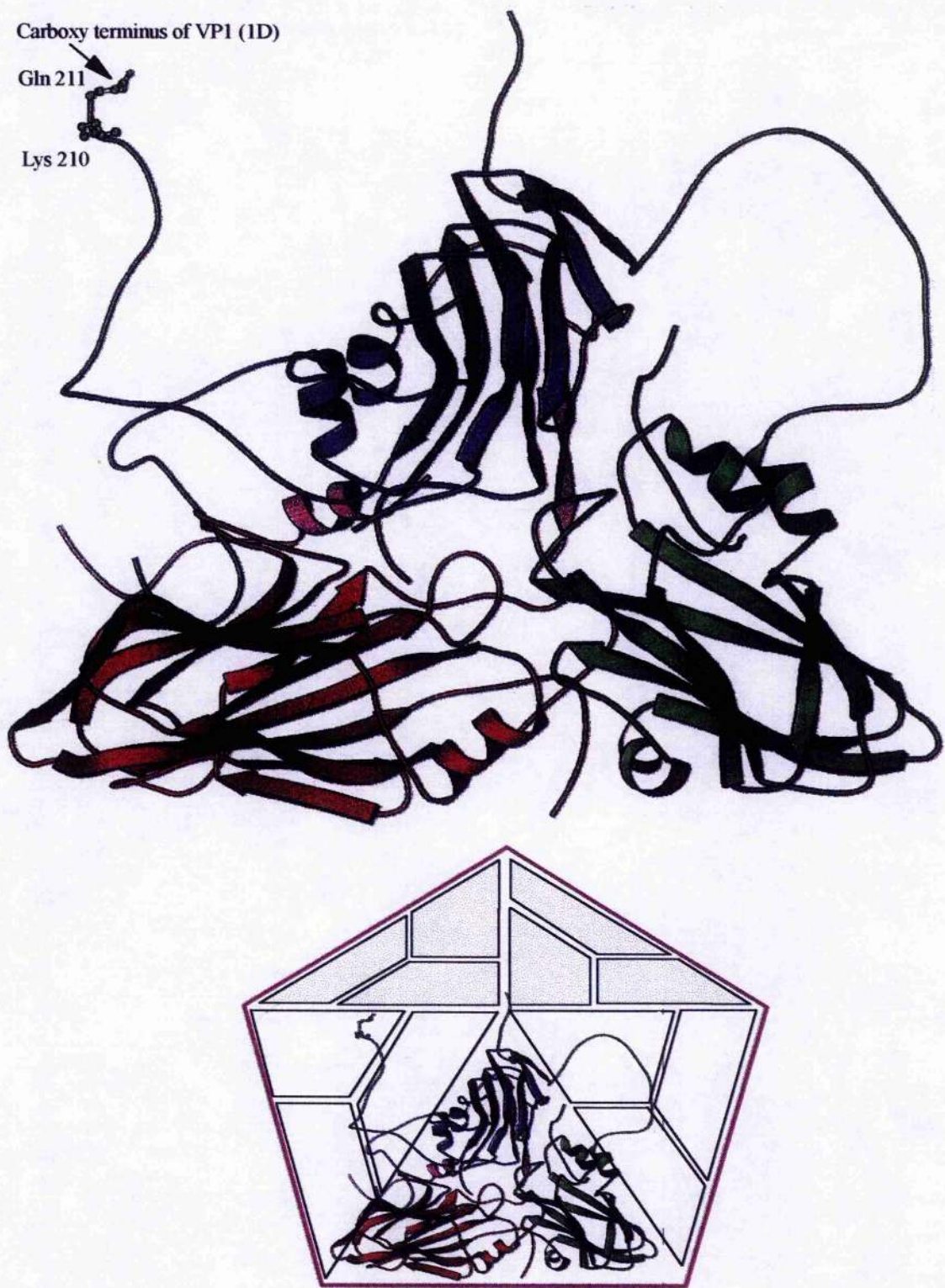
Figure 1.5.1. X-ray crystal structure of the capsid of FMDV serotype C (Lea *et al.*, 1994).
The 5S capsid protomer is shown, with its arrangement into capsid pentamer units illustrated
beneath. VP1 is shown in blue, VP2 in red, VP3 in green and VP4 in magenta. Note that
Glutamine-211 is the carboxy-terminal residue of VP1.

The 2A regions of the aphthoviruses are completely conserved and almost completely identical to those of the recently sequenced equine rhinovirus serotypes 1 and 2 (ERV-1 and ERV-2; Li *et al.*, 1996; Wutz *et al.*, 1996). The 2A proteins are highly conserved amongst Theiler's murine encephalitis (TME) viruses and amongst encephalomyocarditis (EMC) viruses, although only the C-terminal region is highly conserved across all cardioviruses. This C-terminal region also is also highly similar to the 2A regions of FMDV and ERV (see Figure 1.5.2). The proposed autocatalytic cleavage between 2A and 2B is necessary for the successful replication of the virus (Hahn and Palmenberg, 1996). One would thus expect all residues involved in the cleavage event to be well conserved. However, it is not possible to identify a motif of direct similarity to any known proteolytic element.

The C-terminal -NPGP- motif is extremely rare in nature with only a few examples recorded in published databases. It has been reported (although not directly published) that the synthetic peptide H_2N-NPGP-CO_2H is inherently unstable and spontaneously hydrolyses to H_2N-NP-CO_2H and H_2N-GP-CO_2H fragments (Palmenberg *et al.*, 1992). Although this may indicate the general fragility of -NPGP- containing sequences, this does not demonstrate cleavage of the same scissile bond as for 2A mediated cleavage, *i.e.* G-P, so it seems unlikely that this occurs by the same mechanism as 2A mediated cleavage. Attempts to repeat this within our laboratory were unsuccessful.

Site directed mutational analysis of the C-terminal well conserved element, -DIETNPGP-, of EMCV 2A has shown that all of the mutations introduced within this region are deleterious, or more commonly abolish 2A activity completely (Hahn and Palmenberg, 1996). The only substitutions to be found which were still able to cleave efficiently were Ile to Val, not unexpectedly since this merely mutates the EMCV sequence to one found naturally in mengovirus; Thr to Ala, a position where identity is not highly conserved in nature; and Glu to Asp, although a conservative mutation, it is in a naturally completely conserved position and indicates that this residue is not directly

44

involved in the auto catalytic mechanism. These mutations, which still allowed for a moderate 2A activity, however, appear to prevent the subsequent processing of P1-2A by 3C^{pro}. This demonstrates that 2A has a role in the formation of a suitable capsid precursor for cleavage by 3C^{pro}.

The 2A/2B cleavage occurs in a range of heterologous expression systems: rabbit reticulocyte lysates, wheat germ extracts, HTK-143 cells (Ryan *et al.*, 1994), insect cells (Roosien *et al.*, 1990), and tobacco plants (Claire Halpin, personal communication). However, it is yet to be proven that the 2A / 2B junction of FMDV is not merely being very efficiently recognised by a host protease which exists in a wide range of cells. Evidence indicating this may not be the case includes:

   i.   Dilution experiments do not effect the rate of cleavage (Ryan and Drew, 1994).

   ii.   Duration of reaction does not effect the quantity of cleavage products (Batson and Rundell, 1991; Ryan and Drew, 1994).

   iii.   Many known protease inhibitors do not inhibit reaction (Jackson *et al.*, 1986; Sangar *et al.*, 1977).

Although this 19 amino acid sequence has been shown to be cleaved co-translationally in a biological system, it has not yet been possible in our laboratory to repeat this cleavage in a chemically synthesised peptide. One possible explanation, assuming that a host protease is not involved, is that the chemically synthesised peptides maybe folding in a different manner to those expressed in biological systems, perhaps because the physiological conditions have not been replicated correctly, or because it is chemically synthesised in the opposite direction (from carboxy- to amino-terminus) it is not possible to adopt the structure of the nascent peptide. Another explanation may be that in a biological system the proline-glycine peptide bond adjacent to the scissile bond may be formed in the *cis* orientation, whereas in the chemically synthesised peptide this bond most probably equilibrates to form predominantly the *trans* peptide bond.

The mechanism by which the 2A / 2B cleavage in FMDV occurs remains unclear. The self-cleavage hypothesis would invoke cleavage as an intrinsic property of 2A. The 18 amino acids of FMDV 2A might represent an active site / substrate couple which cleaves at its own C-terminus. FMDV 2A might represent an enzyme which has dispensed with those sequences necessary for both, (i) regulation of enzyme activity and, (ii) recognition and binding a substrate intermolecularly. Alternatively the 2A region may form a structure which is unstable under physiological conditions. Another interpretation of the data would be that the 2A sequence in some manner disrupts the normal peptide bond formation during translation.

In any case the 2A region does not fall into any of the known classes of proteases, and the 2A / 2B cleavage of FMDV appears to be unique. The aim of this project is to elucidate the mechanism of this cleavage event.

1.5.7 Potential biotechnological uses of the FMDV 2A region

As a potentially entirely novel self-cleaving peptide the 2A region of FMDV may be of considerable biotechnological use for co-ordinated gene expression.

The co-ordinated introduction of more than one gene into an organism for the stoichiometric expression of two or more interacting proteins, for example in a metabolic pathway, or in a protein complex, would provide a powerful tool. There are problems with the existing mechanisms for the co-ordinated introduction of two transgenes. It has been possible to use pairs of naturally occurring divergently transcribed genes to demonstrate co-ordinate expression of transgenes in tobacco plants and petunia (Gidoni *et al.* 1988). However, when foreign genes were introduced into these systems the degree of stoichiometry was reduced. Furthermore, expression of reporter genes driven by the naturally occurring pair of adjacent and divergent mas promoters has also been found to

be non-co-ordinated (Peach and Velten, 1991). Thus the reliability of co-ordination in such a system must be tested for each pair of transgenes. Tandem gene fusions which show a considerable similarity in their 5'-flanking regions show a good degree of co-ordination, however, again this is not always the case and stoichiometry must be tested for each pair of transgenes (Dean *et al.*, 1988). The FMDV 2A region may be used to stoichiometrically express two or more genes. When inserted in-frame between two genes it is capable of cleaving itself co-translationally from its C-terminal fusion protein (Ryan and Drew, 1994). The first protein produced, however, has a C-terminal extension of about 20 amino acids, nevertheless, this is so small that it is unlikely to interfere with the protein's activity.

The 2A region may also be of in use the permanent expression of foreign sequences by viruses. Short foreign gene sequences have been introduced directly into the ORF's of essential genes of influenza virus. However, this is limited because of size and sequence restrictions on inserted sequences. Since viral proteins are being altered large foreign gene sequences can severely perturb the proteins' structures and affect their biological activity. By the in-frame addition of 2A and a foreign gene sequence to an essential viral gene, then, so long as the short 2A sequence addition to either the viral or foreign protein can be tolerated, the foreign gene can be permanently expressed. This approach has been successfully adopted in the introduction of a permanently expressed gene to influenza virus (Percy *et al.*, 1994) and attenuated Sabin 3 poliovirus vectors (Mattion *et al.*, 1996). Thus this use of the FMDV 2A region has important implications for the use of viruses as expression vectors, particularly in the development of live-attenuated vaccines to protect against infection by other pathogens.

```
          1                                                           60
FMD-A10   ............................................................
FMD-A12   ............................................................
FMD-C1    ............................................................
FMD-O1K   ............................................................
FMD-SAT2  ............................................................
FMD-SAT3  ............................................................
ERV-1     ............................................................
ERV-2     ............................................................
EMC-R     SPNAL...DISRTYPTLHVLIQFNHRGLEVRLFRHGHFWAETRADVILRSKTKQVSFLSN
EMC-D     SPNAL...DISRTYPTLHILIQFNHGGLEIRLFRHVQFWAEAHADVILRSRTKQISFLNN
MENGO     SPNPL...DVSKTYPTLHILLQFNHRGLEARIFRHGQLWAETHAEVVLRSKTKQISFLSN
TME-BEAN  NPAALYRIDLFITFTDEFITFDYKVHGRPVLTFRIPGF.GLTPAGRMLVCMGEQ......
TME-GD7   NPASLYRIDLFITFTDELITFDYKVHGRPVLTFRIPGF.GLTPAGRMLVCMGAK......
TME-DA    NPAAFYRIDLFITFIDEFITFDYKVHGRPVLTFRIPGF.GLTPAGRMLVCMGEK......

          61                                                          120
FMD-A10   ............................................................
FMD-A12   ............................................................
FMD-C1    ............................................................
FMD-O1K   ............................................................
FMD-SAT2  ............................................................
FMD-SAT3  ............................................................
ERV-1     ............................................................
ERV-2     ............................................................
EMC-R     GNYPSMDSRAPWNPWKNTYQAVLRAEPCRVTMDIYYKRVRPFRLPLVQKEWPVREENVFG
EMC-D     GSFPSMDARAPWNPWKNTYHAVLRAEPYRVTMDVYHKRIRPFRLPLVQKEWNVREENVFG
MENGO     GSYPSMDATTPLNPWKSTYQAVLRAEPHRVTMDVYHKRIRPFRLPLVQKEWRTCEENVFG
TME-BEAN  .....PAHGPFTSSRSLYHVIFTATCSSFSFSIYKGRYRSWKKP.IHDELVDRGYTIFG
TME-GD7   .....PAHSPFTSSKSLYHVIFTSTCNSFSFTIYKGRYRSWKKP.IHDELVDRGYTTFR
TME-DA    .....PAHGPFTSSRSLYHVIFTATCSSFSFSIYKGRYRSWKKP.IHDELVDRGYTTFG

          121                             |-> 2B
FMD-A10   .........LLNFDLLKLAGDVESNPG PFFFADV.RSNFSKLVDTINQMQEDMSTKHGPD
FMD-A12   .........LLNFDLLKLAGDVESNPG PFFFADV.RSNFSKLVDTINQMQEDMSTKHGPD
FMD-C1    .........LLNFDLLKLAGDVESNPG PFFFSDV.RSNFSKLVETINQMQEDMSTKHGPD
FMD-O1K   .........TLNFDLLKLAGDVESNPG PFFFSDV.RSNFSKLVETINQMQEDMSTKHGPD
FMD-SAT2  .........LLNFDLLKLAGDVESNPG PFFFSDV.RSNFTKLVDTINQMQNDMSTKHGPD
FMD-SAT3  .........MCNFDLLKLAGDVESNPG P...........................
ERV-1     .........CTNYALLKLAGDVESNPG PTIFSKASADLNALSTSLGELTGMLKDLKAKAE
ERV-2     .........ATNFSLLKLAGDVELNPG PTIWSVMPGGELIEFATKVSQVYKTIKEKCTDP
EMC-R     .LYRIFNAHYAGYFADLLIHDIETNPG PFMFRPR.KQVFQTQGAAVSSMAQTLLPNDLAS
EMC-D     .LYSIFNAHYAGYFADLLIHDIETNPG PFMAKPK.KQVFQTQGAAVSSMAQTLLPNDLAS
MENGO     .LYHVFETHYAGYFSDLLIHDVETNPG PFTFKPRQRPVFQTQGAAVSSMAQTLLPNDLAS
TME-BEAN  EFFKAVRGYHADYYRQRLIHDVETNPG PV......QSVFQPQGAV...LTKSLAPQA...
TME-GD7   EFFKAVRGYHADYYKQRLIHDVEMNPG PV......QSVFQPQGAV...LTKSLAPQA...
TME-DA    EFFRAVRAYHADYYKQRLIHDVEMNPG PV......QSVFQPQGAV...LTKSLAPQA...
                              *  * * *** *
```

Figure 1.5.2   A sequence aligment of all known sequences of aphtho-, cardio-, and equine rhinoviruses.   There is complete homology between all aphthoviruses from L-6 to P-19.   A comparison between aphtho- and cardioviruses shows those residues marked * to be entirely conserved. It is also interesting to note that the area after the cleavage site is usually very hydrophobic.   Such obvious conservation of sequence indicates the importance of these residues for cleavage activity, and for the success of viral replication.

  Completely conserved throughout cardio and aphthoviruses
  Residues of high similarity or identity at this position

Section 1.6 Hypotheses and models for FMDV 2A-mediated cleavage.

The previously discussed results of Ryan and Drew (1994), Batson and Rundel (1991) and Jackson *et al.* (1986) implicated the 2A regions of cardio- and aphthoviruses in the cleavage event between 2A and 2B. Since it was not likely that this region was merely being recognised by a host cell protease, it seems entirely reasonable to suggest that the 2A region itself was capable of mediating a proteolytic reaction at its C-terminus. Molecular modelling studies with energy minimisation have provided various potential three-dimensional structures for the FMDV 2A region, and suggested possible structurally important and catalytic residues.

Molecular dynamics experiments can only provide models, however, which may be tested by site-directed mutagenesis, since many factors which would affect such calculations are unknown for the moment of cleavage. The most important of these factors are:

(i) The presence of 2B sequences. Modelling studies have thus far assumed only 4 residues of 2B exist when cleavage occurs.

(ii) The environment and hence dielectric constant of the 2A peptide.

Is it   (a) surrounded by the ribosome, in the ribosomal exit pore?

The formation of new peptide bonds occurs with the exclusion of water.

One might expect this environment to have a low dielectric constant,

i.e. approximately 4.

(b) on the interface between the ribosome and cytoplasm?

(c) surrounded by the cytoplasm?

One such optimised structure of the 2A region in an environment of dielectric constant of 4 is shown below (Figure 1.6.1). The aspartate side-chain oxygen atoms, in this proposed model, are separated from the scissile glycine-proline peptide bond by approximately 5 Å; enough to allow a water molecule, which has a spherical Van der Waals radius of about 3.2 Å to occupy a position between the aspartate residue and the scissile bond.

This supports a hypothesis of cleavage occurring *via* a general-base hydrolysis mechanism (Figure 1.6.2). This would require concerted removal of a proton from a water molecule by the conjugate base of an acidic residue, and attack of the carbonyl-carbon of the scissile amide bond by the oxygen of the same water molecule.

It is not known how much of the polypeptide C-terminal of the 2A sequence has been synthesised at the time of cleavage. However, if there are approximately 12 residues or more after the C-terminal proline, then molecular dynamics have shown that the conformation of the glycine-proline amide bond probably requires to be in the *cis* conformation for cleavage to occur. If it is in the *trans* conformation then the bulk of the continuing peptide may prevent the approach of the aspartate residue and hence prevent the cleavage event. Thus it was proposed that the scissile bond may be in the *cis* conformation.
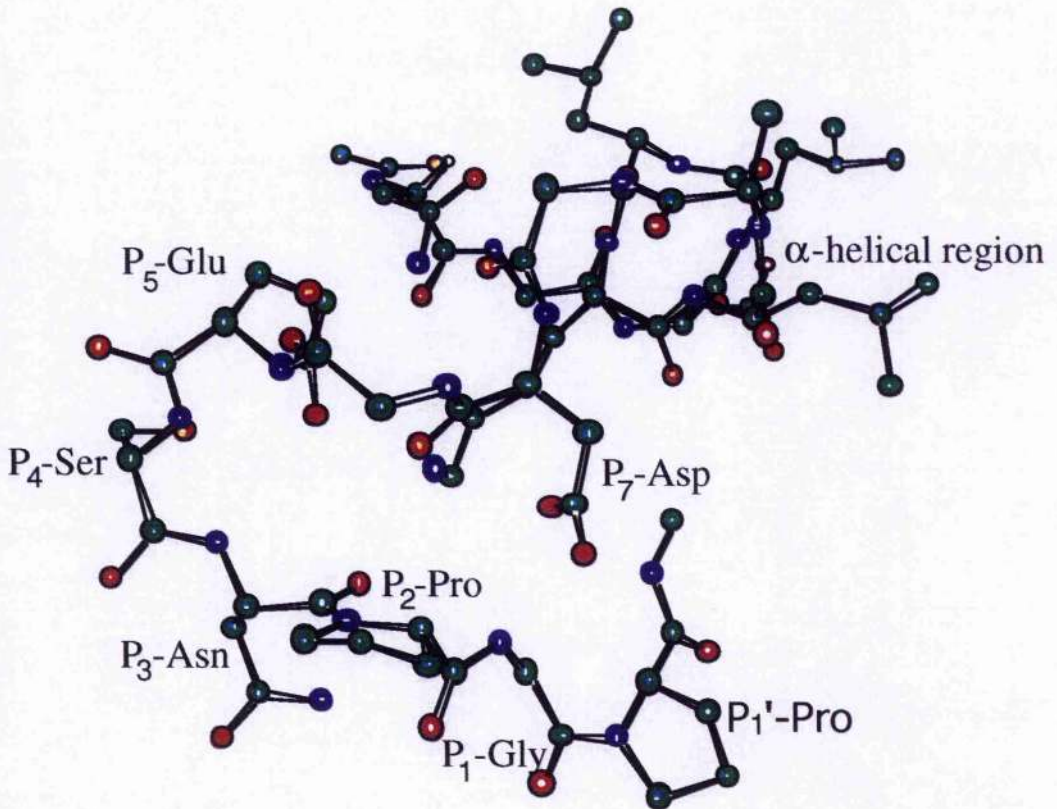
Figure 1.6.1 Optimised FMDV 2A structural model. The aspartate acid oxygens are positioned 4.9Å and 5.3Å from the scissile carbonyl-carbon.

A second similarly derived structural model also suggests a base catalysed mechanism (Figure 1.6.3). β-Sheet formation stabilises this conformation which has one very hydrophobic face and one hydrophilic face, suggesting that it might exist on the interface between the ribosome and the cytoplasm.

Rotation of the glutamate side-chain of this model produces a third model (Figure 1.6.4) which might operate *via* a mechanism similar to that of aspartate proteases (see section 1.3), with a glutamate residue substituted for one of the catalytic aspartate groups. This would assume that a protonated glutamate residue donates its proton to the scissile carbonyl-oxygen. Nucleophilic attack of the carbonyl followed by concomitant transfer of a proton to the aspartate or glutamate would follow resulting in an uncharged tetrahedral intermediate, which collapses with protonation of the amine leaving group.

Another possible mechanism for the 2A/2B cleavage is a deamidation type reaction. Deamidation occurs immediately after an asparagine residue *via* the cyclisation of the asparagine side-chain amido group with its own carbonyl-carbon of the peptide bond to form a succinimide (see section 1.4.2).

In the case of the 2A peptide an Asn-turn may be formed at the C-terminus of 2A, stabilised by an i, i+2 side-chain - main-chain interaction (Figure 1.6.5). The attack of the
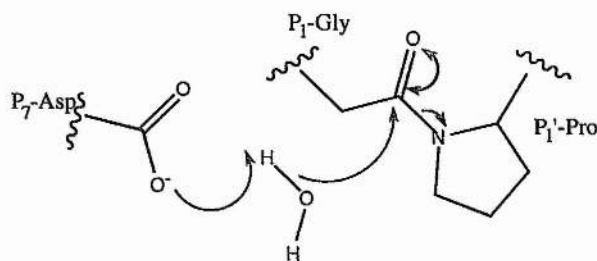
Figure 1.6.2 General-base hydrolysis mechanism for 2A mediated cleavage.

51

amido-NH_2 would occur at the P_1-glycine carbonyl group to produce an 11 membered ring, or at the P_2-proline carbonyl group to give a 5 membered ring.
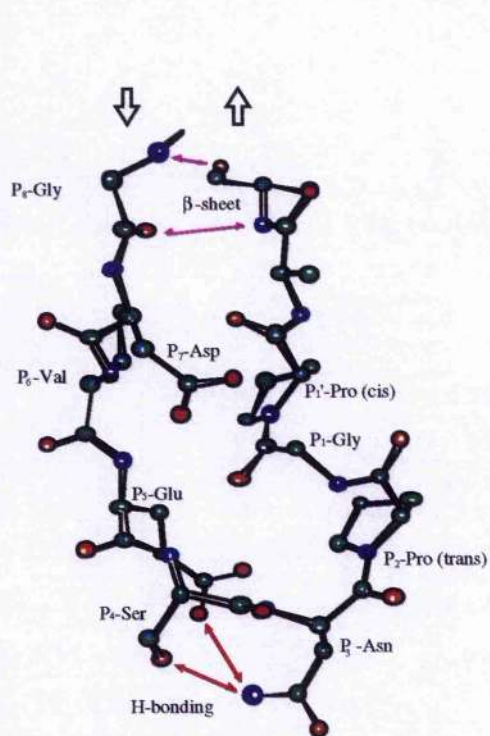
Figure 1.6.3 Optimised structural 2A model with proposed general-base catalysed mechanism
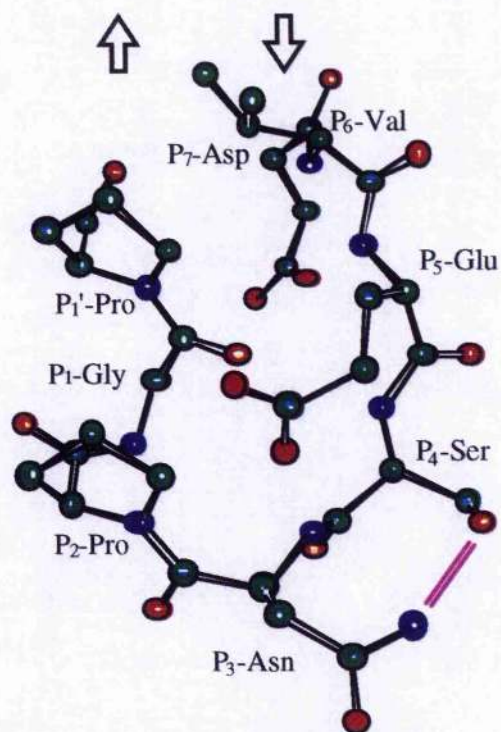
Figure 1.6.4 2A model with proposed mechanism similar to that of aspartate proteases
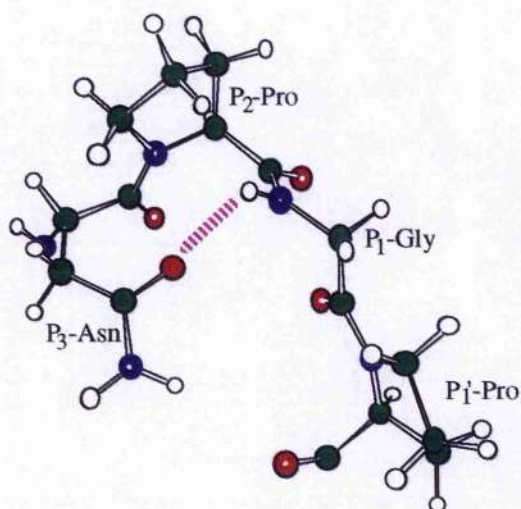
Figure 1.6.5 i, i+2 Side-chain - Main-chain interaction between Asp and Gly

Section 1.7 The aim of this project

The aim of the work undertaken for this thesis was to test the hypothesis that the FMDV 2A region was capable of mediating a proteolytic cleavage at its own C-terminus, and to provide site directed mutational data for the corroboration or elaboration of the models suggested by molecular modelling. It was hoped, therefore, to gain an insight into the cleavage mechanism of the FMDV 2A region. To aid in this a vector was constructed to encode the FMDV 2A region, in frame, between two foreign proteins. This allowed an examination of the activity of the isolated 2A region. Site-directed mutations were inserted within this sequence and the activity of mutated 2A regions compared to that of the wild-type. A comparison with the activity of truncated cardiovirus 2A regions and studies of the influence of the surrounding protein environment were also addressed in an attempt to clarify the mechanism of the FMDV 2A-mediated cleavage.

2. EXPERIMENTAL

Section 2.1 Materials

RNAse A, T4 DNA ligase, T4 DNA polymerase, S1 nuclease, mung bean nuclease, TAQ polymerase, 1kb DNA ladder, restriction enzymes (except for Afl II and Xmn I) and their buffer solutions were obtained from Promega. Afl II and Xmn I restriction enzymes and their buffer solutions were obtained from New England Biolabs. Protein size markers were obtained from Sigma. Oligonucleotides were synthesised using phophoramidite chemistry either on a Cruachem PS250 by W. Blakemore or on an Applied Biosystems 381A by I. Armit. Transformations, mini-preparations and maxi-preparations were carried out using the *E. coli* strains TG1 or XL1-Blue.

Solutions:

LB:
1 % [w/v] bacto-tryptone,
0.5 % [w/v] bacto-yeast extract,
1 % w/v NaCl, pH 7.0.

TSS:
1 % [w/v] bacto-tryptone,
0.5 % [w/v] bacto-yeast extract,
1 % [w/v] NaCl,
10 % [w/v] PEG 5000,
5 % DMSO,
20 mM $MgCl_2$, pH 6.5.

phenol/chloroform:
50 % phenol,
50 % chloroform,
overlaid with 0.001 M Tris.Cl.

TAE:
0.04 M tris-acetate,
0.001 M EDTA.

Agarose gel loading buffer:	5 % [v/v] glycerol, 0.005 % [w/v] bromophenol blue.
STET buffer:	0.1 M NaCl, 10 mM Tris.HCl (pH 8.0), 1 mM EDTA (pH 8.0), 5 % [v/v] Triton X-100.
TBE:	0.9 M tris-borate, 0.002 M EDTA.
fmol Sequencing buffer (5 x):	250 mM Tris.HCl (pH 9.0), 10 mM $MgCl_2$.
Destain:	20 % [v/v] methanol, 10 % [v/v] glacial acetic acid.
Tris glycine buffer:	0.1 % [w/v] SDS, 25 mM Tris, 250 mM glycine.
2 x SDS-PAGE loading buffer: (containing DTT)	100 mM Tris.HCl (pH 6.8), 4 % SDS, 0.2 % Bromophenol blue, 20 % glycerol, 200 mM DTT.
2 x SDS-PAGE loading buffer: (containing β-mercaptoethanol)	124 mM Tris.HCl (pH 6.8), 4 % SDS, 0.2 % Bromophenol blue, 20 % glycerol, 10 % β-mercaptoethanol.

Solutions for minipreparation of DNA by alkaline lysis:

Solution 1:	50 mM glucose, 25 mM Tris.HCl (pH 8.0), 10 mM EDTA.

Solution 2:	1 % SDS,
	0.2 M NaOH.

Solution 3:	3 M KAc,
	5 M acetic acid.

Solutions for immunoprecipitation of proteins:

NET/GEL:	0.15 M NaCl,
	0.05 M Tris.HCl (pH 7.4),
	0.005 M EDTA,
	0.02 % [w/v] sodium azide,
	0.05 % [v/v] NP40,
	0.25 % [w/v] gelatin.

NET/GEL/BSA:	NET/GEL and 0.2 % BSA.

NET/GEL/NaCl:	NET/GEL and 0.5 M NaCl.

NET/GEL/SDS:	NET/GEL and 0.1 % [w/v] SDS.

Section 2.2 Methods

2.2.1 Cloning techniques

Restriction enzyme digestions. All plasmid DNA restriction enzyme digestions were carried out under the conditions specified by the supplier. In general, 1 µg of DNA was digested with 1 unit of enzyme in a total volume of 20 µl, containing 2 µl of x10 restriction buffer at 37 °C unless otherwise specified.

Mung bean nuclease single-stranded DNA degradation. Mung bean nuclease was used to generate blunt ends from the single stranded 5'-overhangs of restricted DNA. A total volume of 100 µl, containing 20 µg of linearised double stranded DNA and 100 units (0.05 µl) of mung bean nuclease was incubated at 37 °C for 10 minutes in the following

buffer conditions: 30 mM sodium acetate (pH 4.6), 50 mM sodium chloride, 1 mM zinc chloride and 5 % (v/v) glycerol. The reaction was extracted twice with 100 μl of phenol/chloroform, and the DNA precipitated from the aqueous phase with 2 volumes of ethanol and 0.05 volumes of 2 M sodium acetate.

T4 DNA polymerase single stranded DNA degradation. T4 DNA polymerase was used to generate blunt ends from restricted DNA with 3'-overhangs. A total volume of 20 μl, containing restricted 25 μg of DNA, 0.5 μl (4 units) of T4 DNA polymerase, and each of the dNTPs composing the last base pair of the DNA duplex to a final concentration of 100 μM, was incubated at 37 °C for 10 minutes in the following conditions: 33 mM tris-acetate (pH 7.9), 66 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM DTT and 0.1 mg/ml BSA.

Agarose-gel electrophoresis. Flat bed agarose gels, of concentration 0.7 - 2 % w/v, were prepared with 1x TAE which contained ethidium bromide at a final concentration of 0.5 μg/ml. DNA samples were applied to the gel in agarose gel loading buffer. Electrophoresis was carried out at 100 - 150 V in 1 x TAE containing 0.5 μg/ ml ethidium bromide. Following electrophoresis the DNA bands were visualised by illumination from a UV transilluminator (UVP).

Agarose Gel Purification of DNA fragments. DNA was isolated from low melting point agarose using one of the two following methods:

(i) Phenol/acetate method. The DNA band was run into the low melting point agarose from were it was excised and placed in a 1.7 ml microcentrifuge tube. This was heated to 70 °C to melt the agarose, then filled with phenol/acetate (phenol equilibrated with 0.3 M sodium acetate), vortexed and incubated on ice for 10 minutes before being spun in a microfuge at 13000 rpm for 5 minutes and the top aqueous layer was removed to a fresh tube. The extraction was repeated and 400 μl of the top layer removed to a fresh tube. To precipitate the DNA, 1 ml of ethanol and 20 μl of 2 M sodium acetate were added

and mixed. The tube was kept at -20 °C overnight, spun at 14000 rpm at 4 °C in a microcentrifuge for 20 minutes and the supernatant discarded. The DNA was resuspended in sterile distilled water.

(ii) WIZARD method. DNA was excised from low melting point agarose as above and isolated using the WIZARD Prep DNA Purification system (Promega), according to the manufacturer's instructions.

Ligations. Ligation reactions were routinely carried out in a final volume of 20 μl. The ligation reaction mix generally consisted of 1 μl (4 Weiss units) of T4 DNA ligase , 0.5 μg of vector DNA, and insert concentrations at 2 fold, x fold or 10 fold molar ratios in T4 DNA ligase buffer (30 mM Tris.HCl (pH7.8), 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM ATP). The reaction mixes were incubated overnight at 16 °C .

Transformation of E. coli. Competent *E. coli* were prepared and transformed with plasmid DNA by one of the two following methods:

(i)TSS method. 30 ml of LB broth was inoculated with 0.6 ml from an overnight culture of LB broth inoculated with TG1, and incubated on a shaker until exponential phase (OD_{600} = 0.3) is reached. The cells were pelleted, the supernatant removed, and the pellet resuspended in ice cold TSS to 1/10 of its original volume. 250 μl of cells in TSS were mixed with plasmid DNA (0.25 μg) and stored on ice for 30 minutes. Transformed cells were then plated out on the appropriate bacterial plates.

(ii) Calcium chloride method. For transformation of ligation reactions, where a medium efficiency of transformation was required the calcium chloride method was used. 50 ml of LB were inoculated with 1 ml from an overnight culture of LB inoculated with *E. coli* , and incubated on a shaker until exponential phase (OD_{600} = 0.4) was reached, then the cells were cooled to 0 °C for 10 minutes in centrifuge tubes, pelleted at 4 °C, and the supernatant discarded. Each pellet was resuspended in 10 ml of ice cold 0.1 M calcium chloride and stored on ice for 15

minutes. The cells were pelleted, the supernatant discarded and then each pellet resuspended in 1 ml ice cold 0.1 M calcium chloride and stored on ice for 15 minutes (or until required). The transformation efficiency of cells prepared in this way is at an optimum between 12 and 24 hours after preparation. To 200 µl of cells 0.5 µg of DNA, or a ligation reaction, is added and sat on ice for 30 minutes. The cells were then heat shocked by heating to 42 °C for 90 seconds and rapidly transferring them to an ice bath for 2 minutes, before plating out on the appropriate bacterial plates.

DNA Preparations. Mini-, midi-, and maxi-preparations of plasmid DNA were used depending on the quantity and quality of DNA required. Midi- or maxi-preparations were required for transcription and translation.

(a) **Mini-preparation of plasmid DNA.** 10 µl of LB broth containing 100 µg/ml ampicillin was inoculated with a single colony and incubated in an orbital incubator overnight. The cells were pelleted, the supernatant removed, and the cells resuspended in 700 µl of STET buffer containing 5 mg/ml lysosyme and incubated at room temperature for 10 minutes. 30 µl of a 10 % SDS solution was added and mixed, followed by 70 µl of 4 M potassium acetate. The sample was mixed and kept at 0 °C for 30 minutes, then the solution was centrifuged and the supernatant transferred to a fresh eppendorf. 1 µl of RNAse A was added and left at room temperature for 15 minutes, the sample was then extracted by adding 700 µl of phenol/chloroform, vortexing, centrifuging and removing the upper aqueous layer to a fresh eppendorf. Two volumes of cold ethanol were then added, mixed, and the sample centrifuged for 10 minutes. The DNA precipitate was washed with 70 % (v/v) ethanol, dried, and then resuspended in 100 µl of distilled water .

(b) Maxi-preparation and Midi-preparation of plasmid DNA. Maxi-preparation of plasmid DNA was carried out *via* two different methods; the latter of these two methods was also used for midi-preparation of plasmid DNA:

(i) Caesium chloride gradient method. 200 ml of LB broth containing 100 μg/ml ampicillin was inoculated with a single colony and incubated in an orbital incubator overnight. The cells were pelleted and the supernatant decanted. The cells were then resuspended in 10 ml of solution 1. 20 ml of solution 2 was added and mixed. 15 ml of solution 3 was added and mixed. The mixture was then kept at 0 °C for 5 minutes, centrifuged for 15 minutes, and the supernatant poured through muslin into fresh Sorval tubes. Isopropanol (0.6 volumes) was added, left at room temperature for 30 minutes, then centrifuged for 5 minutes. The supernatant was pipetted off and the nucleic acid resuspended in 8.5 ml of TE. 10 g of Caesium chloride and 1 ml of 5 mg/ml ethidium bromide were added to the DNA solution, and the sample was centrifuged for 20 hours at 55K at 20 °C. The plasmid DNA bands were seen under UV light and the lower band, containing closed circular supercoiled DNA, was removed. The gradient fraction was extracted with 1 ml of caesium chloride saturated isopropanol, the sample vortexed and the upper phase removed, and this was repeated until all ethidium bromide was removed. 2 ml of distilled water, 300 μl of 2 M sodium acetate and 7.5 ml of ethanol were then added and the sample left at room temperature for 30 minutes to precipitate the DNA. This was then pelleted by centrifugation for 10 minutes, washed with 70 % ethanol and resuspended in 150 μl of TE.

(ii) Qiagen method. Maxi-preparations and midi-preparations of plasmid DNA were carried out according to the Qiagen Plasmid Maxi or Midi Kit protocol, as per manufacturer's instructions (Qiagen).

Polymerase chain reaction (PCR). PCR amplification of DNA was used to amplify genes for cloning purposes and for the production of gene fragments with site directed mutation within the FMDV 2A region. A typical PCR reaction was carried out using

20 ng of template DNA with 100 pmol of each primer, and 0.25 mmol of each dNTP in a total volume of 100 μl of TAQ DNA polymerase buffer (50 mM KCl, 10 mM Tris.HCl (pH 9.0), 0.1 % (v/v) Triton X-100, 1.5 mM MgCl$_2$). The reactions were overlaid with mineral oil to prevent evaporation of the reaction mixture during thermal cycling. Reactions were heated to 94 °C for 5 minutes then held at 85 °C whilst 2 units of TAQ DNA polymerase (Promega) were added to the aqueous phase of each reaction. Amplification was carried out on a thermal cycler using the following parameters: 94 °C for 2 minutes, to denature the DNA; 50 °C for 1 minute, to allow primers to anneal to the DNA; 72 °C for 1 minute for every thousand base pair required to be amplified, to allow TAQ DNA polymerase to extend from each primer. The amplification was carried out for 25 or 30 cycles, with the final 72 °C step being increased to 10 minutes, to ensure that the majority of final product was full length double stranded DNA.

Purification of PCR products. All PCR products were purified using the Wizard DNA direct rapid purification method (Promega) exactly as directed by the manufacturer.

2.2.2 Nucleotide dideoxy sequencing of recombinant DNA clones

Nucleotide dideoxy sequencing was routinely carried out *via* a method based on the fmol protocol (Promega), using cycle sequencing with a [^{32}P]-end-labelled primer. However it was also necessary to use the ^{35}S direct incorporation method, based on the Sequenase protocol (USB), occasionally to reduce compressions, produce less diffuse bands, and provide a better quality sequence.

(i) [^{32}P] end-labelling of primer with cycle sequencing. Cycle sequencing of plasmid DNA was carried out using the fmol protocol (Promega), except for the following changes. The majority of sequencing reactions were intended to determine the nucleotide sequence in the short 2A region close to the sequencing primer. Thus the concentration

of ddNTPs used in each reaction was increased by 5 % from the manufacturers instructions. Thus the stock solutions of d/ddNTPs were as follows in Table 2.1. It was found that an annealing temperature of 50 °C in the PCR program was preferable for all of my primers rather than that of 42 °C suggested in the fmol protocol.

Table 2.1 Stock solutions of d/ddNTPs.

d/ddNTP mixture	ddNTP	dGTP	dATP	dTTP	dCTP
ddGTP	31.5 μM	20 μM	20 μM	20 μM	20 μM
ddATP	367.5 μM	20 μM	20 μM	20 μM	20 μM
ddTTP	630 μM	20 μM	20 μM	20 μM	20 μM
ddCTP	210 μM	20 μM	20 μM	20 μM	20 μM

(ii) Direct incorporation of [^{35}S]-radiolabelled-ATP nucleotides by sequenase T7 DNA polymerase. The Sequenase Version 2.0 DNA Sequencing kit (USB) was used to sequence the construct pMD4 according to the manufacturer's instructions. Briefly the primer was extended by the sequenase T7 polymerase in the presence of [^{35}S]-radiolabelled-dATP, dITP, dCTP and dTTP, ddNTPs are then added and the reaction allowed to progress until all growing chains are terminated by a ddNTP. The reactions are terminated by the addition of EDTA and formamide, denatured and run on an electrophoresis gel.

2.2.3 Translation *in vitro*

Coupled transcription/translation (TnT) reactions. Proteins were expressed *in vitro* using coupled transcription/translation kits, for either wheat germ extract or rabbit reticulocyte lysate systems, according to the manufacturers instructions (Promega).

Proteins were generally radiolabelled with [^{35}S]-methionine unless otherwise stated. Reactions were incubated at 30 °C for 45 minutes, unless otherwise stated, then stopped by the addition of protein loading buffer. 3 µl aliquots of translation reactions were analysed by denaturing PAGE, or 10 µl aliquots were immunoprecipitated. When proteins were to be synthesised *in vitro* with both proline and methionine residues radiolabelled, then both [^{14}C]-proline and [^{35}S]-methionine were each added to a final concentration of 1 mM, and the amino acid mixture minus methionine from the kit was replaced with an amino acid mixture containing all amino acids except proline and methionine.

2.2.4 Protein expression in *E. coli*

LB with 100 µg/ml of ampicillin was inoculated with E. coli strain JM109 and grown overnight on an orbital incubator at 37 °C. 1 ml of this culture was added to 9 ml of LB with 100 µg/ml of ampicillin and grown , with shaking at 37 °C for 3 hours. The culture was divided into 2 x 5 ml, and to one of these was added IPTG to a final concentration of 1 mM, to induce expression of the fusion protein. Incubation was continued for a further 3 hours. The cells were then pelletted by spinning for 5 minutes at 3000 rpm, the supernatant discarded, and the cells resuspended in 250 µl of ice-cold PBS. The cells were then lysed using a probe sonicator for 3 x 10 seconds with 10 seconds on ice between sonicating. From each sample of whole sonicate, induced and uninduced, 100 µl was removed and spun in a microfuge. The supernatant was removed and the pellet resuspended in 100 µl. For both the induced and uninduced samples, 10 µl of supernatant and insoluble material were analysed by SDS-PAGE and western blotting.

2.5 Protein analysis

Denaturing Polyacrylamide gel electrophoresis (SDS-PAGE). The discontinuous buffer gel system based on that of Laemmli (1970) was used for denaturing polyacrylamide gel electrophoresis (SDS-PAGE). Unless otherwise noted all gels were constructed with a 4 % polyacrylamide stacking gel and a 10 % polyacrylamide resolving gel. Typically electrophoresis of each gel was carried out at a constant current of 15 - 20 mA throughout the stacking gel and 30 - 40 mA through the resolving gel.

Immunoprecipitation reactions. In vitro transcription/translation reactions were immunoprecipitated with either CAT or GUS antibodies as follows:

(a) Preparation of immunoprecipitin. Immunoprecipitin (Gibco-BRL), *i.e.* heat killed, formalin fixed *Staphalococcus aureus* cells, 1 g in 10 ml PBS, was centrifuged for 20 minutes at 3000 rpm and the supernatant discarded. The cell pellet was resuspended in PBS containing 10 % β-mercaptoethanol and 3 % SDS, and the sample incubated at 95 °C for 30 minutes. The cells were again centrifuged at 3000 rpm for 20 minutes, the supernatant discarded, and the cells resuspended in 900 μl of NET/BSA .

(b) Pre-clearing the sample. This removed proteins which would have bound non-specifically directly (*i.e.* not *via* the antibody) to the immunoprecipitin. 50 μl of NET/BSA was added to 10 μl of coupled TnT reactions in 0.5 μl microcentrifuge tube. 1.5 μl of immunoprecipitin solution was added and incubated at room temperature for 15 minutes, then spun in a microcentrifuge at 13000 rpm for 2 minutes and the pellet discarded.

(c) Immunoprecipitation reaction. 5 μl of the appropriate anti-serum was added to 50 μl of the supernatant of the precleared sample and incubated overnight at 4 °C. 6 μl of immunoprecipitin were added and incubated for 15 minutes at room temperature, spun in a microcentrifuge for 2 minutes at 13000 rpm, and the supernatant discarded. The pellet

was resuspended in 100 µl of NET/GEL/NaCl , spun in a microfuge for 2 minutes at 13000 rpm, and the supernatant discarded. The pellet was resuspended in 100 µl of NET/GEL/SDS , spun in a microfuge for 2 minutes at 13000 rpm, and the supernatant discarded. The pellet was resuspended in 100 µl of 10 mM tris.HCl (pH 7.5) and 0.1 % NP40, spun in a microfuge for 2 minutes at 13000 rpm, and the supernatant discarded. The pellet was resuspended in 1 x SDS-PAGE loading buffer containing 100 mM DTT, and the sample boiled for 2 minutes and spun in a microfuge for 2 minutes at 13000 rpm. The results were analysed by denaturing SDS-PAGE.

Western blotting. Proteins were separated by SDS-PAGE, before being transferred to PDVF membrane using a semi-dry transfer tank (Gibco-BRL).The membrane was left in blocking buffer (PBS, 20% (w/v) Marvel) either overnight at 4 °C or for 2 hours at 37°C with agitation. It was allowed to equilibrate to room temperature, and washed in washing buffer (PBS, 0.1 % (v/v) Tween 20, 1 % Marvel) for 10 minutes. The antiserum to be used was diluted in antibody diluent (PBS, 1 % (w/v) Marvel) to a dilution of 1:500 or 1:1000 (as stated in the figure legends), and incubated with the PDVF membrane for 30 minutes at room temperature, with agitation. The membrane was washed three times in washing buffer for 10 minutes each time, then incubated with a 1:1000 dilution of horseradish peroxidase labelled goat anti-rabbit IgG in antibody diluent for 45 minutes at room temperature. The PDVF membrane was then washed twice in 1 % Marvel in PBS for 5 minutes, then three times in PBS alone. The enhanced chemiluminescence (ECL) kit was used to develop the blotted membrane.

3. RESULTS

Section 3.1: Construction of a CAT2AGUS reporter system and control plasmids

3.1.1   Introduction

A transcription vector was required which would allow an analysis of the activity of the 2A region of FMDV *in vitro*. This section describes the construction of such plasmids, pMD1 and pMD2, and confirmation of their integrity by expression *in vitro*.

It was desirable to consider 2A activity in the absence of other FMDV sequences which might influence or contribute to activity. Thus it was chosen to insert the 2A region between two foreign gene sequences such that a single long open reading frame was produced. The reporter genes CAT and GUS were chosen for this purpose since a similar construct pCAT2AGUS(3), encoding [CAT2AGUS], and the control constructs pCAT14/21, pGUS12/23, and pCATGUS encoding CAT, GUS and [CATGUS] proteins, respectively, were already available within the laboratory. These constructs were based on the pGEM transcription vector system (Promega), with the gene encoding the protein of interest cloned downstream of a bacteriophage T7 RNA polymerase promotor. An additional consideration in the choice of reporter genes was that antibodies against both CAT and GUS proteins, required for confirmation of the identity of the translation products, were commercially available.

These constructs were designed with the intention that they would be used at a later stage to introduce mutations into the 2A nucleotide sequence using double stranded oligonucleotide adapter molecules. Restriction sites were not available within the existing plasmid pCAT2AGUS(3) for the introduction of adapter molecules encoding for mutations downstream of Asp-12. It was therefore necessary to adapt this creating a new plasmid,

66

pMD1, with unique Spe I and Bgl II restriction sites downstream of 2A (Figure 3.1.1). Construction of new control plasmids encoding CAT2A and GUS are also described.

```
      Q   L   L   N   F   D   L   L   K   L   A   G   D   V   E   S   N   P   G   P   M
     CAGCTGTTGAATTTTGACCTTCTTAAGCTTGCGGGAGACGTCGAGTCCAACCCTGGGCCCATG
     Pvu II                    Afl II          Aat II              Apa I
```

(a) Nucleic acid sequence of the 2A region in pCAT2AGUS(3).

```
      Q   L   L   N   F   D   L   L   K   L   A   G   D   V   E   S   N   P   G   P   F
     CAGCTGTTGAATTTTGACCTTCTTAAGCTTGCGGGAGACGTCGAGTCCAACCCTGGGCCCTTT
     Pvu II                    Afl II          Aat II              Apa I
```

```
      F   F   T   S   R   S   M
     TTTTTTACTAGTAGATCTATG
           Spe I   Bgl II
```

(b) Nucleic acid sequence of the 2A region in pMD1.

Figure 3.1.1 Nucleic acid sequences of the 2A regions in pCAT2AGUS(3) and pMD1. Restriction sites are underlined and amino acid sequences shown above the nucleic acid sequences.

3.1.2 Results

3.1.2.1 Construction of the control pCAT2A

Forward and reverse primers for CAT2A, ORM3 and ORM4 respectively, were designed, each with a BamH I restriction site such that the PCR product produced could be inserted into the plasmid pGem11zf(+) at the BamH I site in the multiple cloning site. A stop codon was placed immediately after the glycine codon in the reverse primer such that the translation product of pCAT2A would be exactly the same as the cleavage product CAT2A.

```
Protein sequence                M   G   K   K
ORM3   5'- GCGCGC GGA TCC ATG GAG AAA AAA -3'
                  BamH I   ↦CAT2A
```

67

```
Protein sequence            E    S    N    P    G
Region annealed to by oligo GAG  TCC  AAC  CCC  GGG
Oligo ORM4          3'-  CTC  AGG  TTG  GGG  CCC  ACT  CCT AGG  CGCGCG -5'
                                                  Stop  BamH I
```

The plasmid pGem11zf(+) was linearised with the restriction enzyme BamH I and treated
with calf intestinal alkaline phosphatase to prevent recircularisation of the vector in the
ligation reaction. The PCR product was also cut with BamH I, then ligated into
pGem11zf(+). The ligation reactions were then used to transform *E. coli* which were
screened by α-complementation, and DNA from the transformants was prepared using the
mini-preparation protocol. The nucleotide sequence was confirmed by dideoxynucleotide
sequencing.

3.1.2.2 Construction of pMD1

A 1792 base pair fragment including the sequence encoding CAT2A was excised from
pCAT2AGUS(3) by a double restriction digest using Xmn I and Apa I. pGUS22/17/88/89
was doubly restricted with Xmn I and Nde I and also partially with Apa I to obtain the
3746 base pair vector fragment. Both fragments were isolated by agarose gel
electrophoresis then ligated together. The resultant plasmid pMD1 contained Afl II, Aat II,
and Bgl II as unique restriction enzyme sites within a short region surrounding 2A (see
Figure 3.1.2 for an overview of the cloning strategy).

3.1.2.3 TnT reaction of pMD1 in wheat germ extract

Coupled transcription and translation (TnT) reactions in wheat germ extracts were
programmed with maxi-preparations of pMD1. $[^{35}S]$-methionine was used to radiolabel
translation products. Denaturing PAGE (11% polyacrylamide) was used to analyse the

GGGCCCTTTTTTTTTTACTAGTAGATCTGGATCCATG
ApaI   SpeI   BglII Bam HI→GUS
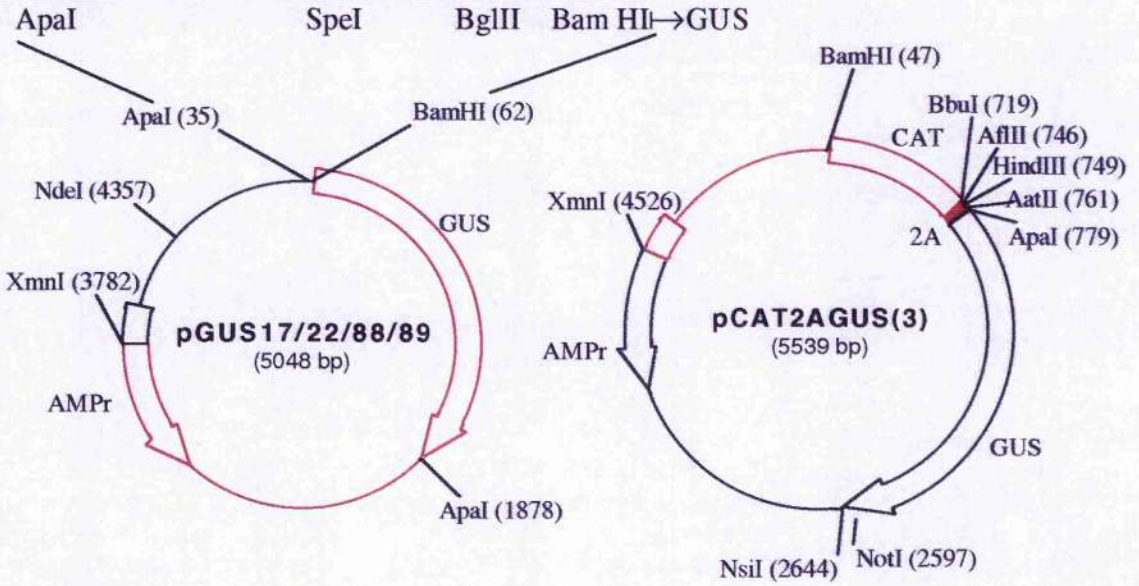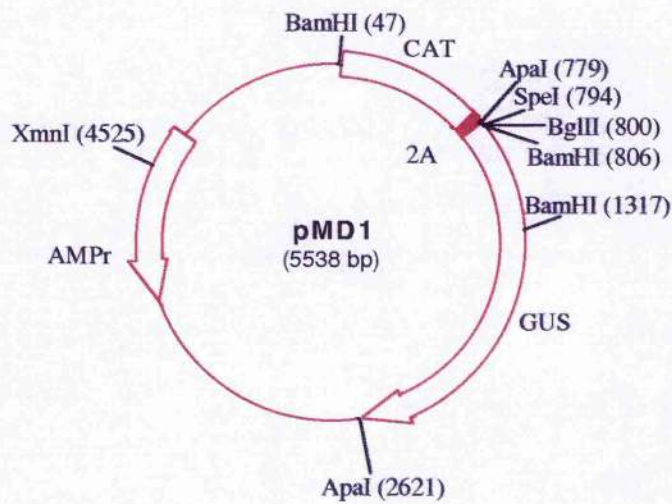
Figure 3.1.2 Construction of pMD1.

translation products for evidence of cleavage at the carboxy-terminal of FMDV 2A and was visualised by autoradiography. Five control reactions, each programmed with one of the plasmids pCAT14/21, pGUS17/22, pGUS17/22/88/89, pCATGUS and pCAT2AGUS(3), were also analysed alongside pMD1 to aid in the identification of bands (Figure 3.1.3).

3.1.2.4 Immunoprecipitation of the transcription and translation products of pMD1

Maxipreparations of the plasmids pCAT14/21, pGUS17/22/88/89, pCATGUS, pCAT2AGUS(3) and pMD1 were transcribed and translated in a coupled wheat germ extract system. The translation products were radiolabelled with [^{35}S]-methionine and immunoprecipitated with both anti-CAT and anti-GUS sera. The products were analysed by denaturing PAGE (11% polyacrylamide; Figure 3.1.4). However, these did not immunoprecipitate the expected GUS protein, and concomitant sequencing of pMD1 confirmed that there was a single base insertion within the GUS gene leading to an early stop codon.

3.1.2.5 Nucleotide sequencing of pMD1

Three primers, OR62, GUS64, and GUS65, were designed for nucleotide sequencing of pMD1 over the region encoding 2A and the 5' sequence of the GUS gene. These primers were forward sense annealing to regions approximately 200 bases apart.

OR62   5' - TGATGGCTTCCATGTCGGC - 3'

GUS64   5' - CCGGGCAATTGCTGTGCC - 3'

GUS65   5' - ACTATCCCGCCGGGAATG - 3'
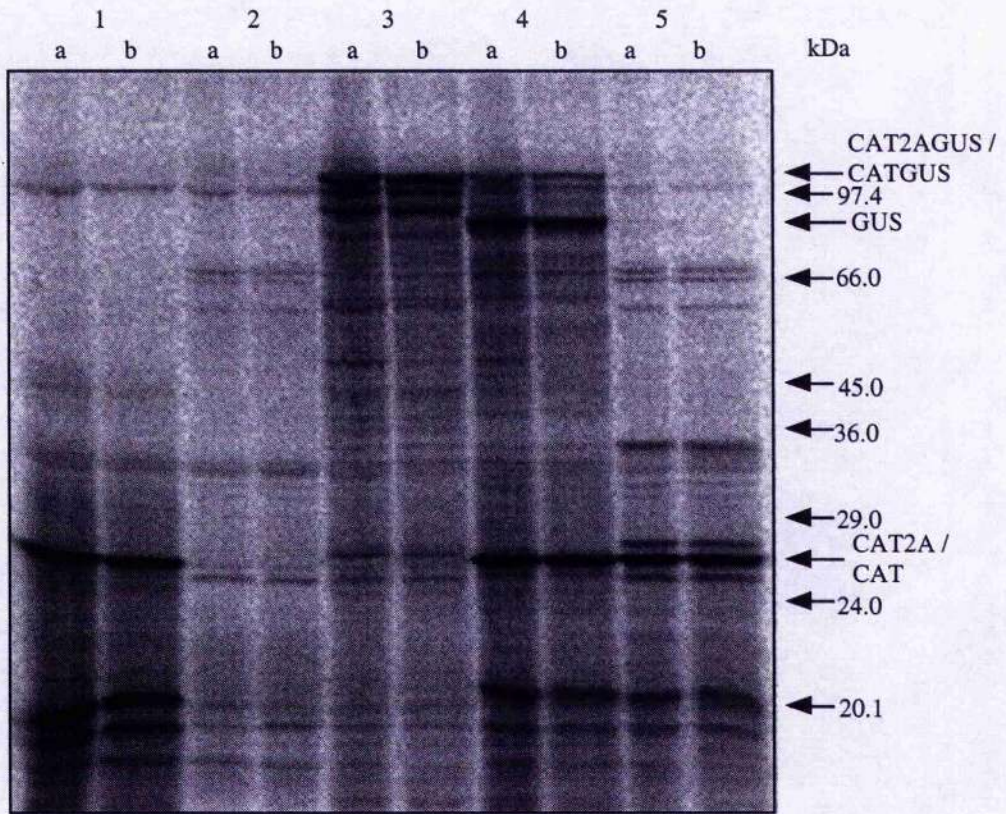
Figure 3.1.3 Translation products of pMD1. TnT wheat germ extract systems were programmed with the following plasmid DNA: lanes 1(a) and (b) - pCAT14/21, lanes 2(a) and (b) - pGUS17/22/88/89, lanes 3(a) and (b) - pCATGUS, lanes 4(a) and (b) - pCAT2AGUS(3), and lanes 5(a) and (b) - pMD1. Samples were loaded into lanes labelled (a) with SDS-PAGE loading buffer containing DTT, and those labelled (b) with SDS-PAGE loading buffer containing β-mercaptoethanol.

71
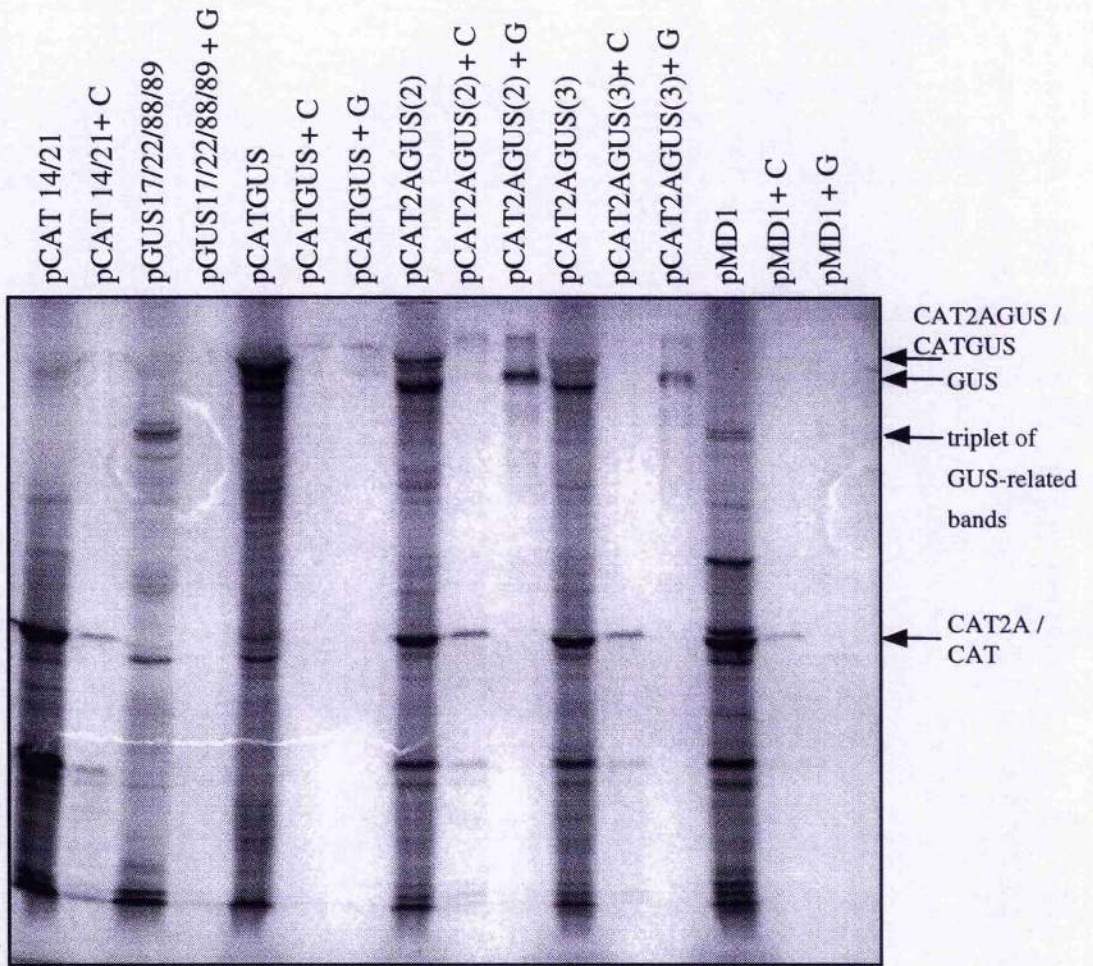
Figure 3.1.4 Immunoprecipitation of pMD1 translation products. TnT wheat germ extracts programmed with plasmid DNA and immunoprecipitated with anti-CAT or anti-GUS sera. Lanes containing samples immunoprecipitated by anti-CAT antibodies are indicated by "+ C", and those immunoprecipitated by anti-GUS antibodies are indicated by "+ G".

Each set of sequencing reactions was loaded onto the sequencing gel at three different times, spaced 1.5 hours apart, to provide a larger range of readable sequence data for each primer. The sequence of the 5' terminus of GUS in pMD1 is given in Figure 3.1.5, where the translation of the nucleic acid sequence is shown underneath the DNA sequence. Both the nucleic acid and amino acid sequences are numbered from the start codon of CAT and the numbering shown above the DNA sequence. The point mutations found are shown in bold red. The additional guanine base altered the reading frame such that a truncated protein of 20 amino acids was encoded after the 2A cleavage site. The protein sequence encoded by the frame shift is shown in red.

```
<----------------------2A-----------------------------------------------------/--GUS--->
678/226
CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC GAG TCC AAC CCT GGG CCC TTT
 L   L   N   F   D   L   L   K   L   A   G   D   V   E   S   N   P   G   P   F

738/246                                      ⇓
TTT TTT ACT AGT AGA TCT GGA TCC ATG TTG ACG TCC TGT AGA AAC CCC AAC CCG TGA AAT
 F   F   T   S   R   S   G   S   M   L   T   S   C   R   N   P   N   P   *   N

798/266
CAA AAA ACT CGA CGG CCT GTG GGC ATT CAG TCT GGA TCG CGA AAA CTG TGG AAT TGA TCA
 Q   K   T   R   R   P   V   G   I   Q   S   G   S   R   K   L   W   N   *   S

858/286
GCG TTG GTG GGA AAG CGC GTT ACA AGA AAG CCG GGC AAT TGC TGT GCC AGG CAG TTT TAA
 A   L   V   G   K   R   V   T   R   K   P   G   N   C   C   A   R   Q   F   *

918/306                                                          ⇓
CGA TCA GTT CGC CGA TGC AGA TAT TCG TAA TTA TGC GGG CAA CGT CTG ATA TCA GCG CGA
 R   S   V   R   R   C   R   Y   S   *   L   C   G   Q   R   L   L   S   A   R
```

Figure 3.1.5 Sequencing of pMD1.

3.1.2.6 Synthesis of a new GUS gene

A new GUS gene containing fragment, to correct the point mutation within GUS of pMD1, was amplified *via* PCR using the template pGUS12/23, which had previously been shown to transcribe and translate to give a protein of the correct size.

73

Two primers, GUS66 and GUS67, were designed to prime the forward and reverse strands respectively. GUS66 was designed such that a Bgl II site, needed for cloning the gene back into the vector, was encoded at the 5' end of the primer. The primer required for the 'reverse strand' encoded a stop codon, followed by a Not I site, also for cloning back into the vector. However, it was noted that restriction at the Not I site within the GUS gene of GUS12/23 would produce a stop codon closer to the end of the natural GUS sequence once inserted into either pMD1 or GUS17/22/88/89.

```
Protein sequence                          M   L   R   L   V   E
GUS66         5'- GCG CGC AGA TCT ATG TTA CGT CCT GTA GAA -3'
                          Xba I      ⊢→GUS

Protein sequence         V   G   C   I   A
Region of annealing     GTT GGA TGC ATA GCT
GUS67        3'- CAA CCT ACG TAT CGA ACT CGC CGG CGT ATATA-5'
                                     Stop      Not I
```

3.1.2.7   Construction of the control plasmid pGUS66/67

The plasmid pGUS66/67 was constructed, containing the new corrected GUS gene, to provide a control for the GUS protein in TnT reactions. The vector pGUS17/22/88/89 was completely digested with the restriction enzymes, Bgl II and Not I, and the fragment of 3212 bp was purified by agarose gel isolation. The similarly restricted GUS PCR product, 1819 bp, was also agarose gel purified before being ligated into the restricted pGUS17/22/88/89 vector producing the new construct pGUS66/67 of 5031 bp.

3.1.2.8   Construction of plasmid pMD2

Plasmid pMD2 (encoding [CAT2AGUS]) was constructed to correct the GUS gene within pMD1. The vector pMD1 was completely digested with the restriction enzymes Bgl II and

74

Not I, and the fragment of 3703 bp was purified by agarose gel isolation. The similarly restricted GUS PCR product, 1819 bp, was also purified by agarose gel isolation and then ligated into the restricted vector to produce pMD2 of 5522 bp (Figure 3.1.6).

3.1.2.9 TnT reactions of pMD2 and pGUS66/67 in a wheat germ extract system

DNA maxi-preparations of plasmid pGUS66/67 and plasmid pMD2 were transcribed and translated in a coupled wheat germ extract system. Six control reactions, each programmed with one of pCAT14/21, pGUS17/22/88/89, pGUS12/23, pCATGUS, pCAT2AGUS(2), and pMD1 were analysed alongside two samples of pMD2 and four of pGUS66/67. The translation products were radiolabelled with [^{35}S]-methionine. The products were analysed by denaturing PAGE (11% polyacrylamide) and the results were visualised using phosphor-imaging techniques (Figure 3.1.7).

3.1.2.10 TnT reaction of pMD2 in rabbit reticulocyte lysates and immunoprecipitation of the translation products

A DNA maxi-preparation of plasmid pMD2 was transcribed and translated in a coupled rabbit reticulocyte lysate system. The translation product was radiolabelled with [^{35}S]-methionine and immunoprecipitated with both anti-CAT and anti-GUS antisera. The products were analysed by denaturing PAGE. The results were visualised by auto-radiography (Figure 3.1.8).

GUS gene sequences were amplified by PCR using GUS66 and GUS67 oligonucleotide primers with the template pGUS12/23.
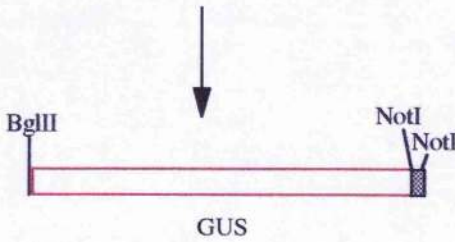
GUS66/67 PCR product
(1892 bp)

Restrict with BglII and NotI.
Gel Isolate1819 bp fragment.
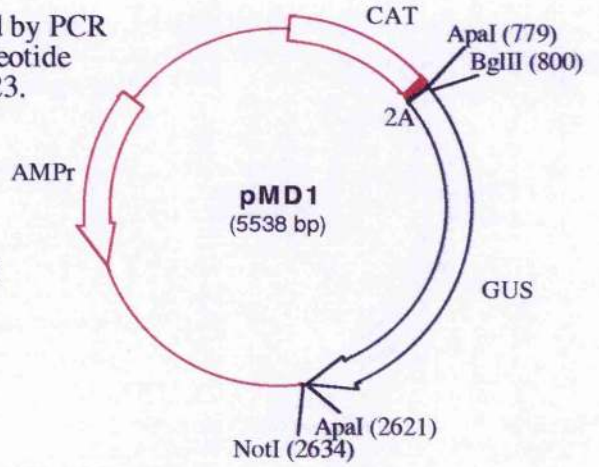
Restrict with BglII and NotI.
Gel Isolate 3703 bp fragment.

Ligate
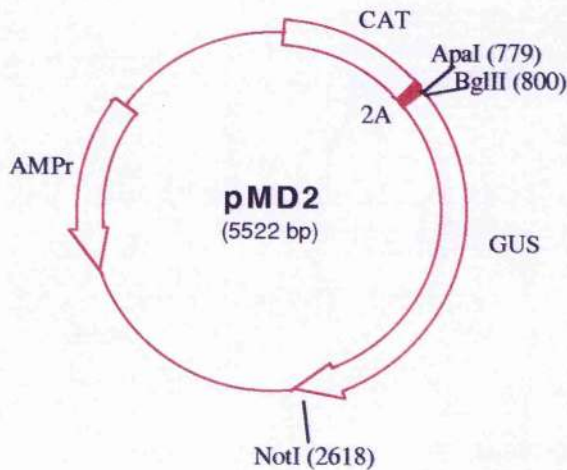
Figure 3.1.6   Construction of plasmid pMD2.

76
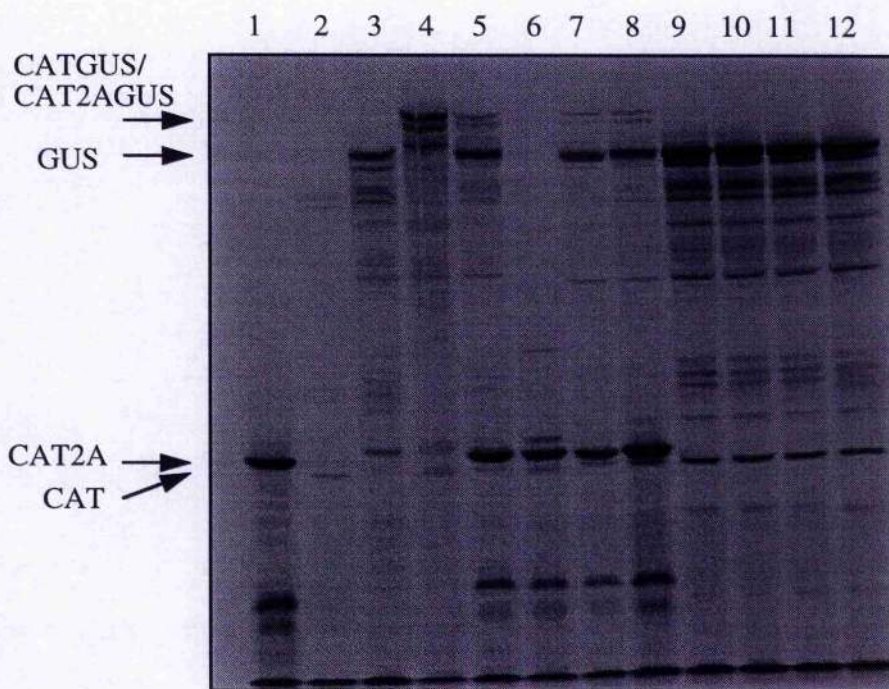
Figure 3.1.7 TnT reactions of pMD2 and pGUS66/67.
TnT reactions in the wheat germ extract system were programmed with the plasmid DNA indicated.

1. pCAT14/21	5. pCAT2AGUS(2)	9. pGUS66/67 (A)
2. pGUS17/22/88/89	6. pMD1	10. pGUS66/67 (B)
3. pGUS12/23	7. pMD2 (A)	11. pGUS66/67 (C)
4. pCATGUS	8. pMD2 (B)	12. pGUS66/67 (D)
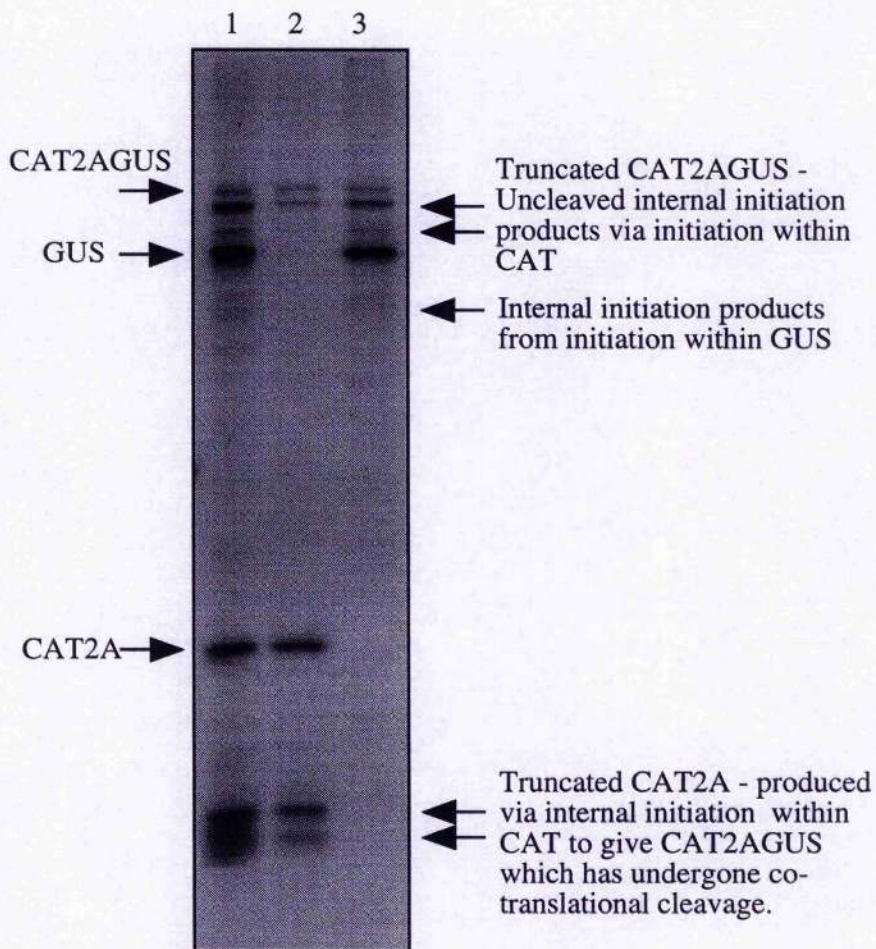
Figure 3.1.8 Immunoprecipitation of pMD2 translation products.
Immunoprecipitation with anti-CAT and anti-GUS sera of the translation
products of pMD2 in a rabbit reticulocyte lysate TnT reaction.

Lane   1. pMD2
         2. pMD2 immunoprecipitated with anti-CAT
         3. pMD2 immunoprecipitated with anti-GUS

3.1.3   Discussion

3.1.3.1   The construction pCAT2A

Mini-preparations of the plasmid pCAT2A were sequenced using the oligo OR62 which anneals to the 3' end of the CAT gene. The DNA sequencing of the region of the BamH I site following 2A allowed the identification of those plasmids in which the insert had ligated in the correct orientation. The first of those plasmids identified with the correct orientation had a mutation in the BamH I site, however, this can be tolerated since it does not affect the translation product.

3.1.3.2   Analysis of the phenotype of pMD1

Analysis of the TnT reactions in rabbit reticulocyte lysate and wheat germ extract systems showed a high degree of internal initiation of translation. The translation products of all constructs containing a GUS gene contained a triplet of bands of high molecular weight, presumed to be GUS itself and two internal initiation products. These, and the presence of other internal initiation products caused some difficulty in the initial elucidation of the results. A band of appropriate size for CAT2A was seen for pMD1. However no band corresponding to the expected position for uncleaved [CAT2AGUS] polyprotein was seen. For the control construct, pCAT2AGUS(3), some uncleaved [CAT2AGUS] polyprotein was seen in addition to cleaved CAT2A and GUS. This latter plasmid did not contain the hydrophobic residues, Phe-Phe-Phe, which are present both at the start of the native FMDV 2B protein and in pMD1, and since residues closely flanking cleavage sites might be expected to influence cleavage, pMD1 might have been expected to have a higher efficiency of cleavage. The triplet of bands which seemed to indicate the presence of GUS migrated further than the GUS product achieved by the control pCAT2AGUS(3), although they migrated the same distance as the controls pGUS17/22 and pGUS17/22/88/89. To

79

ensure that the triplet was not merely a gel artefact (*i.e.* due to different conformations of GUS caused by not fully reducing the disulphide bonds), different loading buffer systems with dithiothreitol (DTT) or β-mercaptoethanol were tried. There was no change in the band pattern (Figure 3.1.3). Comparison with protein size markers demonstrated that the triplet of bands indeed represented proteins too small to be GUS *i.e.* 50-60 kDa rather than the 70 kDa of the authentic GUS gene, and that these bands, in all probability, were due to internal initiation of translation within the GUS gene.

Immunoprecipitation reactions (Figure 3.1.4) further clarified the interpretation. The product CAT2A was immunoprecipitated as expected from the TnT reactions programmed with pMD1 and pCAT2AGUS(3), with anti-CAT serum. Also as expected CAT and [CATGUS] proteins were similarly immunoprecipitated from the reactions programmed with pCAT14/21 and pCATGUS, respectively. However, the results of the immuno-precipitations with anti-GUS serum were not as predictable. Protein products [CATGUS], and both [CAT2AGUS] and GUS were precipitated from reactions programmed with pCATGUS and pCAT2AGUS(2) and (3), respectively. However, no high molecular weight translation product was immunoprecipitated from the reactions containing either pGUS17/22/88/89 or pMD1.

The only reasonable explanations for the results of the immunoprecipitation reactions and the sizes of the translation products are that either there was a point mutation, or a deletion or insertion resulting in a frame shift leading to the early occurrence of a stop codon within the GUS gene. Such a codon occurring close to the end of 2A would result in no noticeable change in the size of the CAT2A translation product whether or not cleavage had occurred. If there was cleavage of the [CAT2AGUS] polyprotein the GUS translation product was obviously very small since the immunoprecipitations with anti-GUS antibodies did not indicate that GUS was present on the gel. Thus, it was assumed that a mutation had probably occurred towards the 5' end of GUS and hence the GUS gene was sequenced from the 5' end to look for the mutation. Indeed, an inserted deoxyguanosine

had caused a frame shift and hence the premature occurrence of a stop codon within 21 codons of the 2A/2B cleavage site. The GUS peptide fragment which would have resulted from cleavage at the 2A/2B junction would have been only 2245 Da in molecular weight, and as such would have run off of the bottom of the protein gel. If cleavage had not occurred the size difference in CAT2A would not have been easily detectable.

3.1.3.3   Analysis of the phenotype of pMD2

The controls pGUS12/23 and pCAT2AGUS(2), both demonstrated a band at 70 kDa for GUS which is absent for pGUS17/22/88/89 and pMD1. A protein of the same approximate size was also found for all samples of pGUS66/67 and pMD2. pMD2 also showed, as expected, a protein corresponding to CAT2A at 28 kDa and to a much lesser extent a protein corresponding to the uncleaved polyprotein [CAT2AGUS] at 98 kDa . Fairly high levels of internal initiation were found throughout the reactions and are responsible for all other bands on the gel.

The assignment of the protein bands was confirmed by the immunoprecipitation of the translation products of pMD2 in rabbit reticulocyte lysates. Anti-CAT serum immunoprecipitated polyprotein [CAT2AGUS] at 98 kDa, CAT2A at 28 kDa, and truncated versions of CAT2A and [CAT2AGUS] produced *via* internal initiation within CAT. Anti-GUS serum also immunoprecipitated polyprotein [CAT2AGUS] and its truncated forms from internal initiation, as well as the GUS protein at 70 kDa. Since all of the major bands were immunoprecipitated by either CAT or GUS antiserum it was determined that all of the major small proteins were produced by internal initiation within the correct reading frame.

3.1.4   Conclusions

- A vector system encoding the polyprotein [CAT2AGUS] with suitable sites for the later insertion of double stranded oligonucleotide adapter molecules to produce site directed mutations has been created.

- This polyprotein undergoes an apparent auto-proteolytic cleavage in both rabbit reticuloctye lysate and wheat germ extract producing CAT2A and GUS proteins.

- Control constructs pCAT2A and pGUS66/67 encoding CAT2A and GUS have been created.

Section 3.2: Densitometric Analyses of the translation products of pMD2

3.2.1   Introduction

The protein PAGE analysis of pMD2 was complicated by additional unassigned protein products thought to be due to internal initiation within the CAT gene. Thus, for a clearer interpretation of the PAGE gels it was necessary to assign some of the internal initiation products and assess the processivity of 2A within these products.

A ribosome scans along mRNA until the first start codon, AUG, is found upon which protein synthesis begins. However, the sequences flanking the AUG start codon influence its recognition by eukaryotic ribosomes. A consensus sequence, G C C G C C **A/G** C C **A U G** G, called the Kozak consensus sequence is required for optimal initiation (Kozak, 1986). Those nucleotides shown in bold are most important for recognition. Some ribosomes may, therefore, scan through the first AUG and initiate at another downstream AUG producing internal initiation products. An alignment of the sequences surrounding the methionines of the CAT gene with the Kozak consensus sequence is seen below (Table 3.2.1).

Kozak consensus sequence	G	C	C	G	C	C	**A/G**	C	C	**A**	**U**	**G**	G
Met-1	G	A	G	G	G	A	U	C	C	**A**	**U**	**G**	G
Met-67	G	C	C	C	G	C	C	U	G	**A**	**U**	**G**	A
Met-75	G	A	A	U	U	C	C	G	U	**A**	**U**	**G**	G
Met-77	C	G	U	A	U	G	**G**	C	A	**A**	**U**	**G**	A
Met-142	A	U	U	G	A	G	**A**	A	U	**A**	**U**	**G**	U
Met-163	G	U	G	G	C	C	**A**	A	U	**A**	**U**	**G**	G
Met-173	G	U	U	U	U	C	**A**	C	C	**A**	**U**	**G**	G
Met-185	A	A	G	G	U	G	C	U	G	**A**	**U**	**G**	C
Met-205	G	U	C	G	G	C	**A**	G	A	**A**	**U**	**G**	C

Table 3.2.1   Alignment of Kozak consensus sequence with the sequences encoding methionines within the CAT gene.

83

By comparing this alignment with the migration of the internal initiation products by PAGE it was estimated that these may have been products from initiation at Met[77] and Met[163], in addition to Met[1]. Thus, it was planned to truncate the CAT protein by deleting the upstream context from Met[77] and Met[163] of polyprotein [CAT2AGUS], as encoded by the construct pMD2, to produce pMD12/4 and pMD13/4 respectively.

This section describes the molecular cloning required for the assignment of [ΔCAT2AGUS] translation products caused by internal initiation within CAT, a densitometric analysis of all cleavage products and thus a determination of the relative quantities of CAT2A and GUS proteins.

3.2.2   Results

3.2.2.1   Construction of plasmids encoding truncated versions of [CAT2AGUS]

Two forward primers, ORM12 and ORM13, were designed to encode a BamH I site followed by six codons from within CAT starting at Met[77] and Met[163] respectively. A third primer ORM4 with a 3' sequence complementary to the 5' sequence of 2A, *i.e.* complementary to the sequence encoding ESNPG was used as the reverse primer. Truncated versions of CAT2A initiating at Met-77 and Met-163 were amplified by PCR from pMD2 using the oligonucleotide primer pairs ORM12 and ORM4, and ORM13 and ORM4 respectively.

```
Protein Sequence:              M⁷⁷  K    D    G    E    L
ORM12:  5'- GCGCGC GGA TCC ATG AAA GAC GGT GAG CTG -3'
                   BamH I   ↦ΔCAT2A
```

```
Protein Sequence:              M¹⁶³ D    N    F    F    A
ORM13:  5'- GCGCGC GGA TCC ATG GAC AAC TTC TTC GCC -3'
                   BamH I   ↦ΔCAT2A
```

ORM4:   5'- GCGCGC <u>GGA TCC</u> TCA CCC GGG GTT GGA CTC -3'
                    BamH I

The PCR products were then restricted with BamH I and Aat II, and after agarose gel
purification ligated into the similarly restricted and purified pMD2 fragment of 4804 base
pairs.   The 5294 and 5036 base pair products, entitled pMD12/4 and pMD13/4
respectively, were then transformed into *E. coli*, mini-prepared, sequenced and maxi-
prepared using the Qiagen method.

3.2.2.2   Construction of a new plasmid, pMD5, encoding [CAT2AGUS] and removing methionine from the start of GUS

A new GUS containing fragment, with the natural methionine start codon removed, was
amplified by the PCR using the template pMD2.   The primer GUS30 was designed to
prime the forward strand, whilst the SP6 promoter primer would prime the reverse strand.

```
                 T    S    R    S    L    R    P
GUS30    5'- ACT AGT AGA TCT TTA CGT CCT -3'
                Spe I       Bgl II   ->GUS

SP6      5'- GATTTAGGTGACACTATAG -3'
```

The 1869 base pair PCR product was cut with Bgl II and Not I, to produce a 1816 base
pair fragment which was purified by agarose gel isolation.   The 3703 base pair fragment of
the similarly digested pMD2 was also purified by agarose gel isolation then ligated to the
digested PCR product producing the plasmid pMD5. This was transformed into *E. coli*,
mini-prepared, sequenced and maxi-prepared using the Qiagen method.

3.2.2.3 TnT reactions of pMD12/4, pMD13/4 and pMD5 in rabbit reticulocyte lysate and wheat germ extract systems

Maxipreparations of pMD2, pMD12/4, pMD13/4 and pMD5 were transcribed and translated (alongside the controls pGUS 66/67, pCATGUS, pMD2 x Bgl II, and pCAT2A) in coupled rabbit reticulocyte lysate and wheat germ extract systems. The translation products were separated by PAGE, and visualised by phosphoimaging (Figure 3.2.1), then subjected to densitometric analysis.

3.2.2.4 Study of the stability of translation products

The stability of the translation products from both rabbit reticulocyte lysates and wheat-germ extract reactions was investigated. A 25 µl coupled reaction for each system was incubated at 30 °C for 45 minutes. Protein synthesis was then stopped by the addition of cycloheximide to a final concentration of 0.9 mg/ml and the mRNA degraded by adding RNase A to a final concentration of 60 µg/ml. The incubation was continued at 30 °C and at time intervals, from 0 to 20 hours, a 3.5 µl sample was removed, and added to 3.5 µl of loading buffer to stop any further processing and degradation. The reactions were frozen until being subjected to PAGE (Figures 3.2.2 and 3.2.3), and analysed densitometrically.

3.2.2.5 Error analysis for the densitometry of phosphoimages

DNA from a maxi-preparation of plasmid pMD2 was used in a coupled rabbit reticulocyte lysate reaction and the sample separated on four different polyacrylamide gels by PAGE, then visualised by phosphoimaging. Densitometric analysis was carried out using Mac Bas Version 2 software. Briefly, the major bands and a background area for each lane were encircled by freehand and by drawing boxes around them, then the (PSL-background)
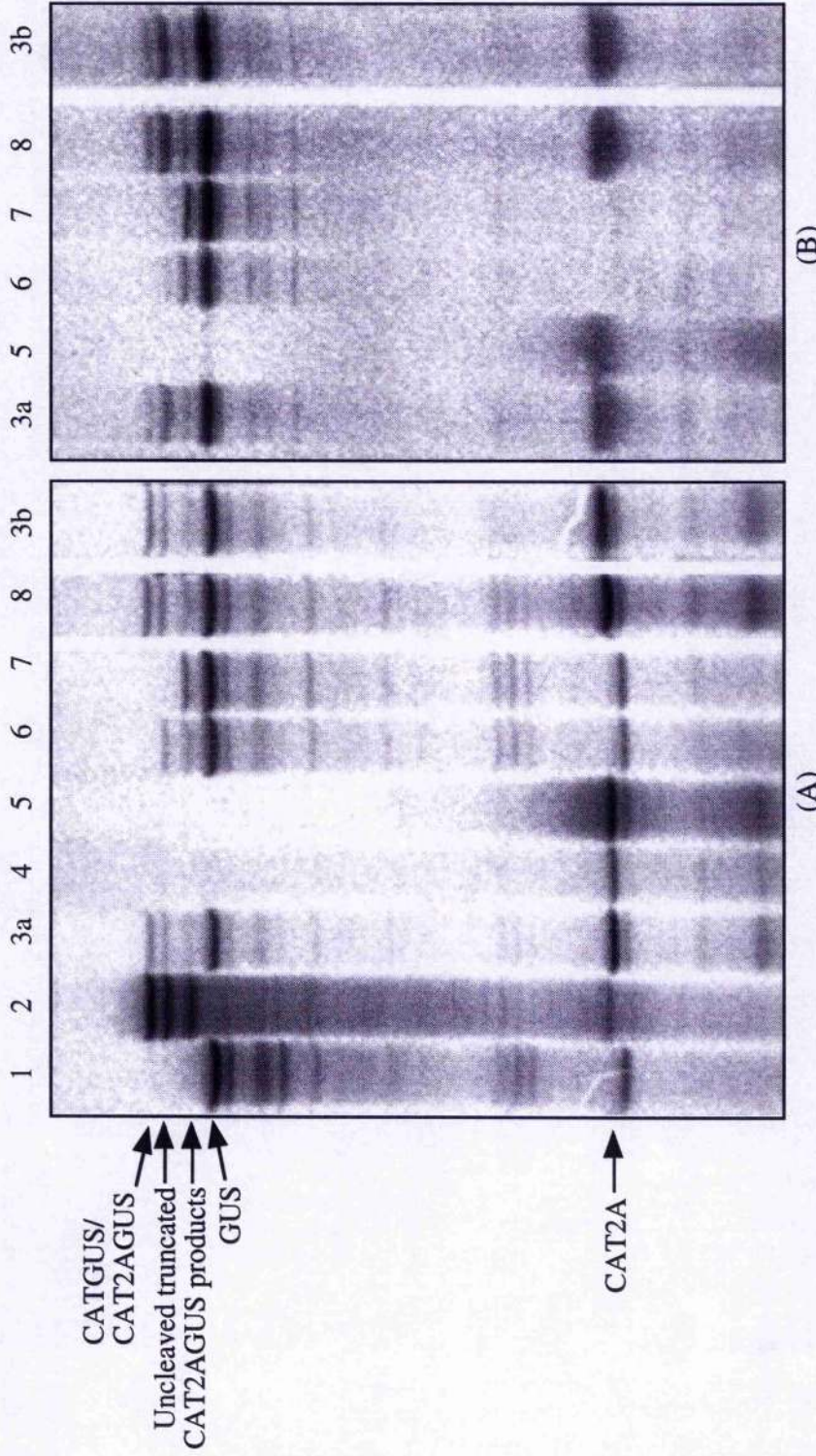
Figure 3.2.1 TnT reactions to assign internal initiation products. The reaction products of TnT reactions in (A) wheat germ extract and (B) rabbit reticulocytes were programmed with the following DNA:

1. pGUS66/67
2. pCATGUS
3a, 3b. pMD2
4. pMD2xBgl II
5. pCAT2A
6. pMD12/4
7. pMD13/4
8. pMD5(4)

87
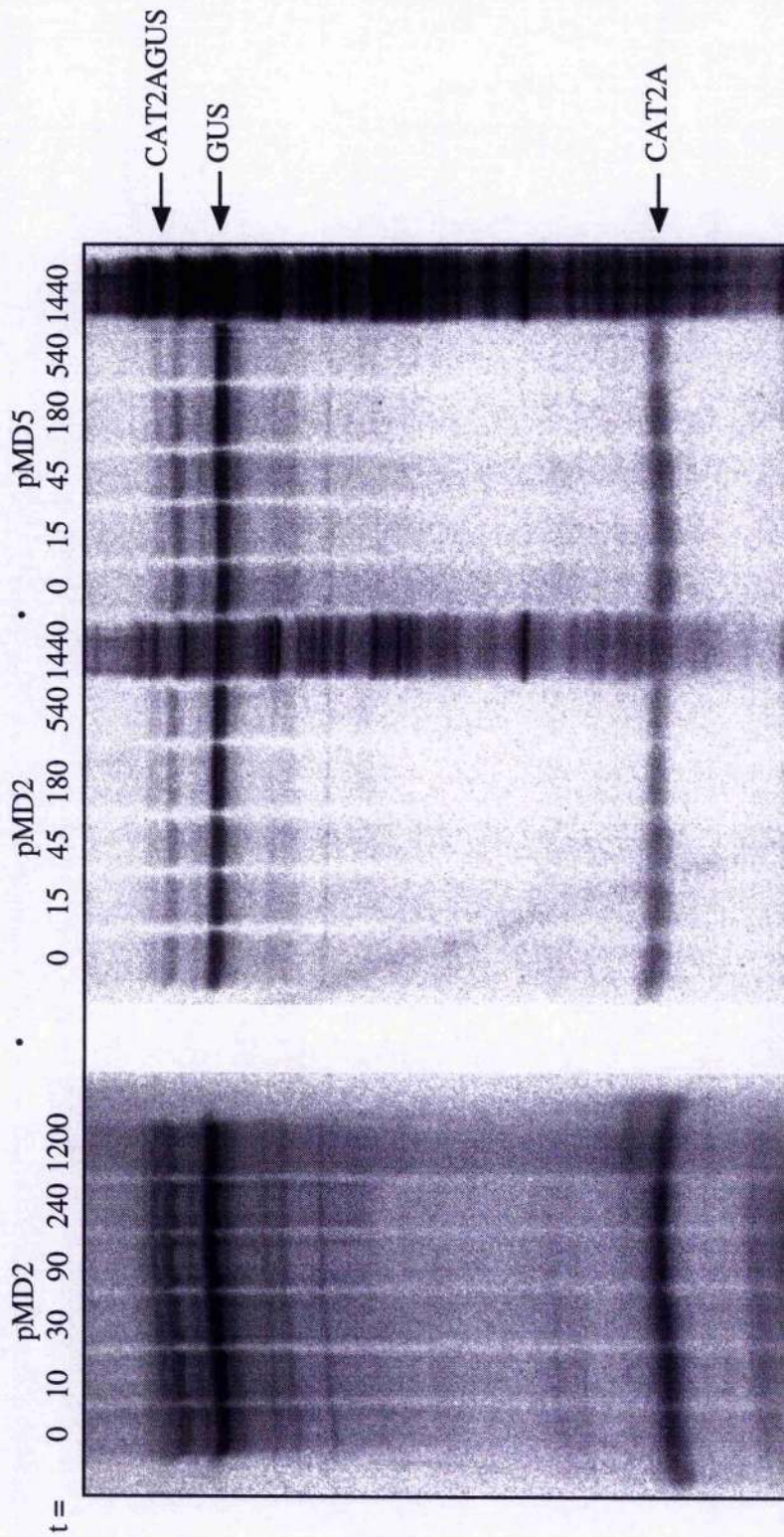
Figure 3.2.2 SDS-PAGE of the rabbit reticulocyte lysate translation products of pMD2 and pMD5 after incubation at 30 °C. After the translation reaction had been stopped the translation products of pMD2 and pMD5 were incubated at 30 °C for the time, t (minutes), indicated. Each lane (except for those with the incubation time of 1440 minutes, where considerable degradation has taken place) was analysed densitometrically to investigate the degree of degradation.
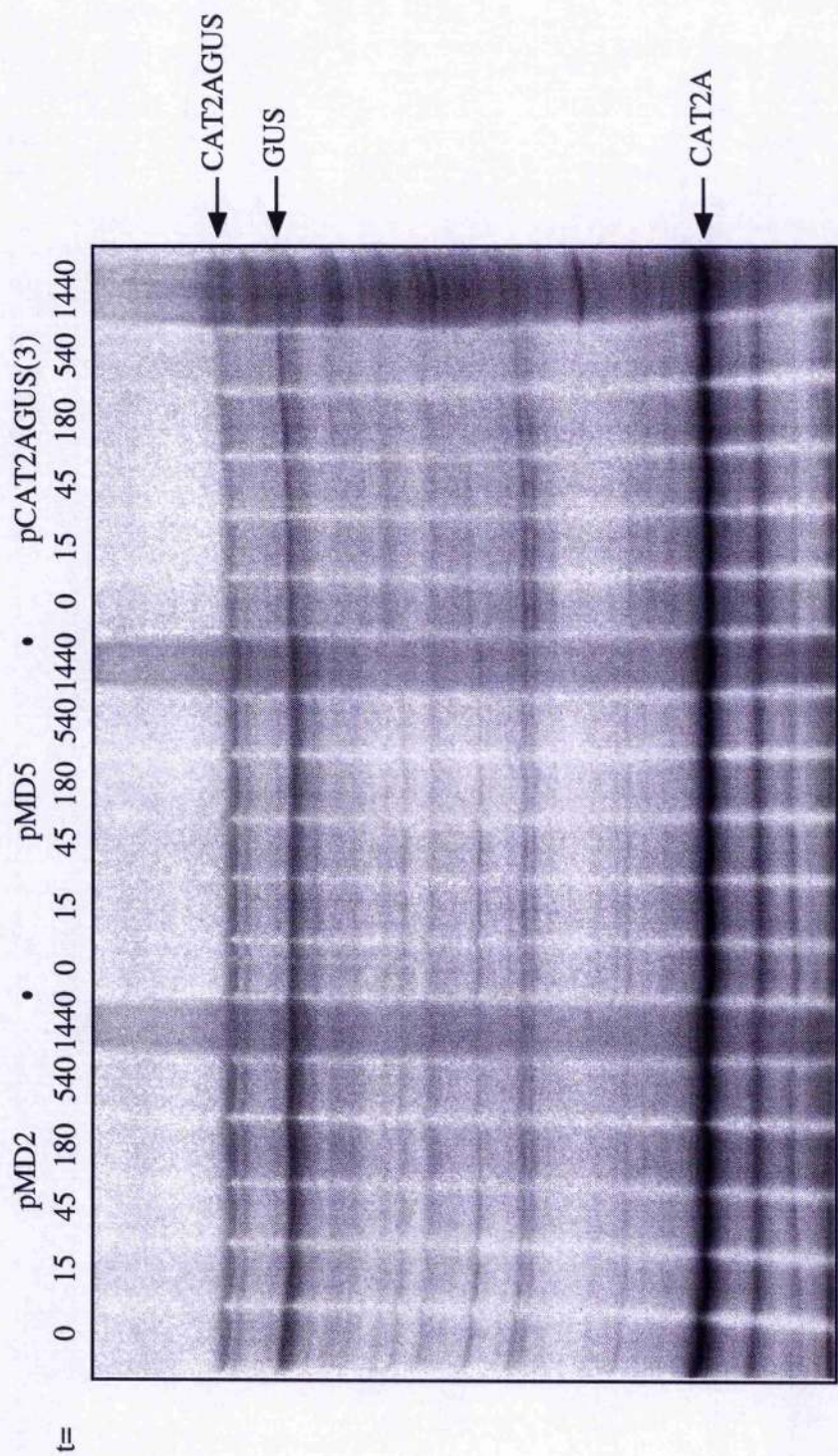
Figure 3.2.3 SDS-PAGE of the wheat germ extract translation products of pMD2, pMD5, and pCAT2AGUS(3) after incubation at 30 °C. After the translation reaction had been stopped the translation products of pMD2 and pMD5 were incubated at 30 °C for the time, t (minutes), indicated. Each lane was analysed densitometrically to investigate the degree of degradation.

89

count was read. A third method was also used, whereby a profile of the lane was obtained and the peaks corresponding to the major bands of interest were integrated to find values for (PSL-background) (Figure 3.2.4). The results obtained are displayed in Table 3.2.2.

3.2.3 Discussion

3.2.3.1 Assignment of internal initiation products

The highest molecular weight product of the lanes containing the translation products of pMD12/4 corresponded to the truncated polyprotein [Δ^{77Met}CAT2AGUS], and appeared to be slightly smaller than the first internal initiation product of pMD2 reactions. Thus it was concluded that the first internal initiation product of pMD2 was not [Δ^{77Met}CAT2AGUS] but the slightly larger protein derived by initiation at Met^{75} - [Δ^{75Met}CAT2AGUS].

The highest molecular weight product of the TnT reaction of pMD13/4, [Δ^{163Met}CAT2AGUS], co-migrated with the second internal initiation product of pMD2, confirming its identity.

3.2.3.2 Densitometric analysis of internal initiation products

With the identity of each major internal initiation product now assigned, and hence the number of methionines within each known, the percentage of initiation at each methionine could be found for both the rabbit reticulocyte lysate and wheat germ extract systems *via* densitometric analysis. However, for the polyprotein [CAT2AGUS] and its truncated versions the possibility of a variation in the processivity of each version had to be considered. Therefore, the percentage of initiation at each methionine for the control

90
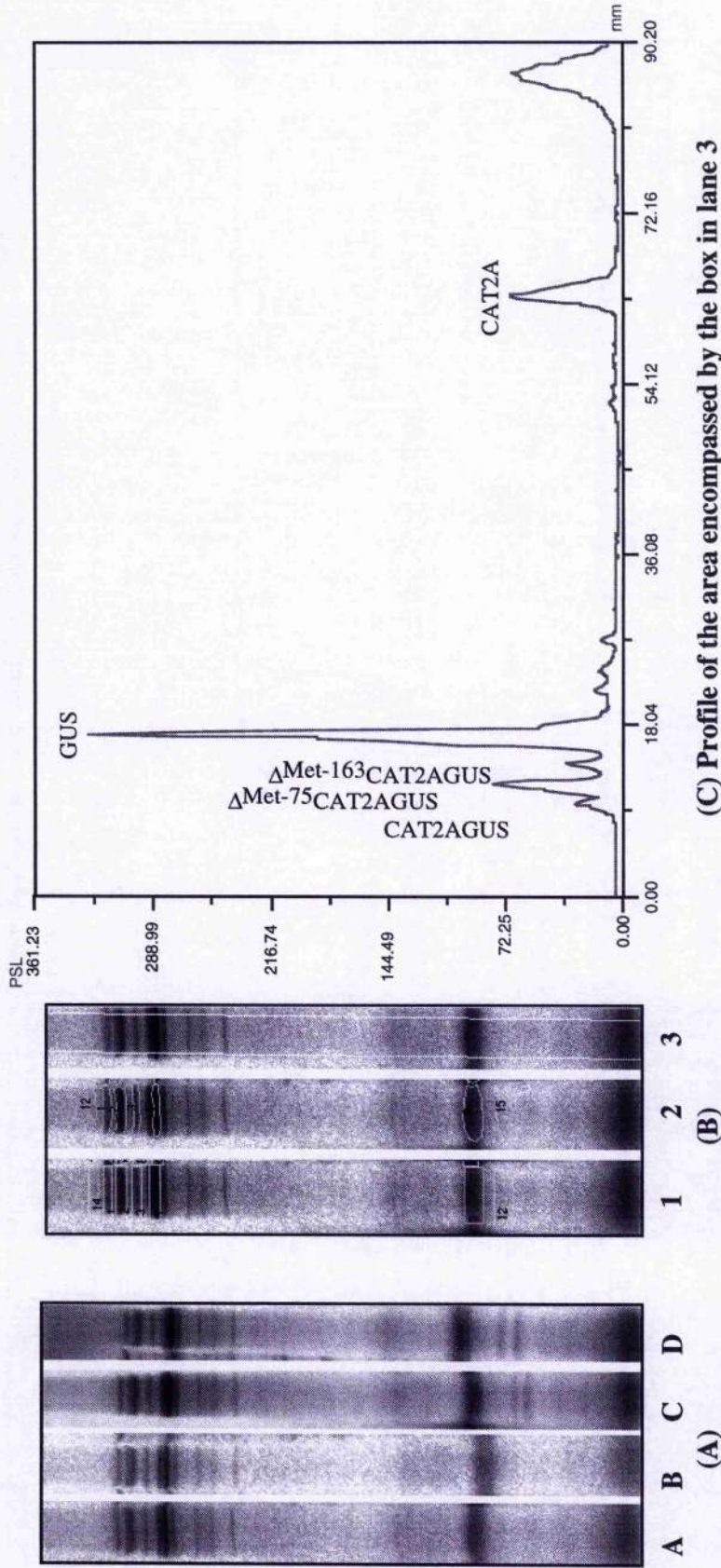
Figure 3.2.4 Densitometric measurement of pMD2 translation products.

(A) The four different lanes containing the translation products of a single TnT reaction of pMD2.

(B) Lane A containing the translation products from the coupled TnT reaction of pMD2 in rabbit reticulocyte lysate was analysed densitometrically by either drawing the boxes or freehand shapes around the bands, or by integrating the named peaks of the profile of the entire lane.

(C) The profile of the area in lane 3.

91

lane	measurement	CAT2AGUS		75MetCAT2AGUS		163MetCAT2AGUS		GUS		CAT2A	
		PSL	% lane	PSL	% lane	PSL	% lane	PSL	% lane	PSL	% lane
A	Int.	489.	3.09	1813	11.47	562.3	3.56	8868	52.15	4070	25.76
	Freehand	387.7	3.04	1503	11.79	346.3	2.71	7545	59.20	2964	23.25
	Boxed	419.3	2.81	1708	11.43	487.3	3.26	7789	56.12	4533	30.34
B	Int.	512.	3.19	2103	13.11	518	3.86	9349	58.28	3459	21.56
	Freehand	489.	3.61	2320	17.14	502	3.70	7684	56.75	2544	18.79
	Boxed	636.	3.79	2028	12.10	837.8	5.00	9432	56.27	3826	22.83
C	Int.	88.2	4.94	321.0	17.96	97.09	5.43	888.0	49.69	392.6	21.97
	Freehand	89.1	4.84	273.7	14.84	96.36	5.23	944.8	51.23	440.0	23.86
	Boxed	107.	4.79	349.2	15.60	126.2	5.64	1117	49.96	536.6	23.99
D	Int.	144.8	4.51	458.2	14.28	155.5	4.85	1719	55.81	659	20.55
	Freehand	128.9	4.31	468.4	15.68	133.2	4.46	1520	50.90	736	24.64
	Boxed	124.2	3.58	481.41	13.88	164.8	4.75	1930	55.81	769	22.16
Mean % of lane		3.88±0.74		14.11±2.12		4.37±0.89		54.33±3.20		23.31±2.7	

Table 3.2.2 Densiometric measurement of the PSL distribution for the rabbit reticulocyte TnT products of pMD2. The number of (PSL-background) for each of the major bands corresponding to the protein products of interest, CAT2AGUS, 75MetCAT2AGUS, 163MetCAT2AGUS, GUS, and CAT2A was measured three times for each of the four lanes, named A, B, C andD, and is displayed under the column heading of PSL for each product. A direct comparison between different measurements for each product was permitted by summing each of the independent sets of measurements, integration of profile peaks (termed Int.), boxed and freehand encircling, then each PSL measurement was expressed as a percentage of the total measured PSL for that lane (shown in the column labelled % lane). The mean value for the percentage of total PSL measured for each protein product ± the standard deviation is shown as 'mean % of lane'.

92

polyprotein [CATGUS] was also calculated for comparison with [CAT2AGUS] (Tables 3.2.3 and 3.2.4).

Table 3.2.3 referring to reactions in wheat germ extract, shows that the average percentage of each [CAT2AGUS] version is comparable to the complementary [CATGUS] version;

[CAT2AGUS] 35.5 % [Δ75MetCAT2AGUS] 32.5 % [Δ163MetCAT2AGUS] 32 %,

c.f. [CATGUS] 37 %   [Δ75MetCATGUS] 32 %   [Δ163MetCATGUS] 31 %.

It could hence be concluded that the cleavage activity of each truncated version was the same as that of the full length translation product.

This was again exhibited in rabbit reticulocyte lysates (Table 3.2.4), although the percentages of initiation at each position were different;

[CAT2AGUS] 17 %   [Δ75MetCAT2AGUS] 55 %   [Δ163MetCAT2AGUS] 29 %,

c.f [CATGUS] 16 %   [Δ75MetCATGUS] 58 %   [Δ163MetCATGUS] 26 % .

Slight variation is seen between and within batches of rabbit reticulocyte lysate or wheat germ extract.

3.2.3.3   Densitometric analysis of the cleavage products of pMD2

With the percentage of initiation at the first methionine calculated it was then possible to calculate the relative levels of CAT2A and GUS (Table 3.2.5). The values for CAT2A were corrected to a value which would have resulted from all initiation of protein synthesis occurring at the first methionine of CAT. The values for the ratios of CAT2A to GUS and the percentage cleavage with respect to CAT2A were calculated using this corrected value for CAT2A. Since ratios of CAT2A : GUS were not found to be 1 : 1, the percentage of cleavage was calculated based on (i) CAT *i.e.* $\dfrac{[CAT2A]}{[CAT2A]+[CAT2AGUS]}$

and   (ii) GUS *i.e.* $\dfrac{[GUS]}{[GUS]+[CAT2AGUS]+ \text{all } [\Delta CAT2AGUS]}$.

Figure/lane	Polyprotein	PSL-Background (B) for each product produced by initiation at the Met named						Total	% Initiation at each Met.		
		Met-1	Met-1/22	Met-75	Met-75/20	Met-163	Met-163/17		Met-1	Met-75	Met-163
Appendix 1/3	CATGUS	301.0	13.7	996	49.8	373	21.9	85.4	16	58	26
Appendix 1/3	CAT2AGUS	35.3	1.6	113	5.7	51.0	3.0	10.7	16	55	29
3.1.7/1	CAT2AGUS	39.8	1.81	85.7	4.29	44.6	2.62	8.72	21	49	30
3.2.1B/3a	CAT2AGUS	1.88	0.09	5.83	0.29	3.23	0.19	0.57	15	51	34
3.2.1B/3b	CAT2AGUS	4.6	0.21	13.5	0.68	6.79	0.4	1.29	16	53	31
Mean value	CATGUS	--------							16	62	23
Mean value	All CAT2AGUS	--------							17	54	29
Mean value	CATGUS and CAT2AGUS	--------							17	55	29

Table 3.2.4 Densitometric Analysis to find the pecentage of initiation at Met-1, Met-75, and Met-163 of the CAT protein in rabbit reticulocyte lysate.

Figure/lane	Polyprotein	PSL–Background (B) for each product produced by initiation at the Met named						Total	% Initiation at each Met		
		Met-1	Met-1/22	Met-75	Met-75/20	Met-163	Met-163/17		Met-1	Met-75	Met-163
Appendix 1/3	CATGUS	301.0	13.7	996	49.8	373	21.9	85.4	16	58	26
Appendix 1/3	CAT2AGUS	35.3	1.6	113	5.7	51.0	3.0	10.7	16	55	29
3.1.7/1	CAT2AGUS	39.8	1.81	85.7	4.29	44.6	2.62	8.72	21	49	30
3.2.1B/3a	CAT2AGUS	1.88	0.09	5.83	0.29	3.23	0.19	0.57	15	51	34
3.2.1B/3b	CAT2AGUS	4.6	0.21	13.5	0.68	6.79	0.4	1.29	16	53	31
Mean value	CATGUS	---						---	16	62	23
Mean value	All CAT2AGUS	---						---	17	54	29
Mean value	CATGUS and CAT2AGUS	---							17	55	29

Table 3.2.4 Densitometric Analysis to find the pecentage of initiation at Met-1, Met-75, and Met-163 of the CAT protein in rabbit reticulocyte lysate.

| in vitro system | Figure/Lane | PSL | | % Full length i.e. init. at Met-1 | Corrected CAT2A | CAT2A : GUS | % Cleaved | |
		CAT2A-B 9	GUS-B 13				Based on CAT2A	Based on GUS
rabbit reticulocyte lysate	3.1.7/1	14.3	12.7	21	68	5.3 : 1	88	59
	Appendix 1/3	16.0	34.7	16	100	2.9 : 1	91	77
	3.2.1B/3a	1.3	2.27	15	8.5	3.1 : 1	94	83
	3.2.1B/3b	1.6	4.26	16	10.1	2.4 : 1	88	77
wheatgerm extract	3.2.1A/3a	6.9	3.16	35	19.8	6.3 : 1	96	80
	3.2.1.A/3b	9.5	3.93	36	26.5	6.7 : 1	97	82

Table 3.2.5 Ratio of CAT2A : GUS and percentage of cleavage for polyprotein [CAT2AGUS].

The ratio of CAT2A : GUS formed, and the percentage of cleavage occuring was calculated for the polyprotein [CAT2AGUS] both in rabbit reticulocyte lysate and wheat germ extract. Measured values of PSL for CAT2A and GUS (Table 3.2.2) were divided by the number of methionines in each protein product to provide relative molar quantities of CAT2A and GUS. The relative molar quantities of CAT2A were adjusted to take initiation at internal methionines into account. Values for the percentage of [CAT2AGUS] which was cleaved was calculated with respect to the quantities of both cleavage products - corrected CAT2A and GUS.

and on GUS. In rabbit reticulocytes the percentage of cleavage was on average 90 % based on CAT2A, and 74 % based on GUS. In the wheat germ extract system these figures were a little higher at 97 % and 81 %, respectively. Although there were large variations between reactions there was consistently 2 to 5 times more CAT2A than GUS in rabbit reticulocytes, and approximately 6 times more CAT2A than GUS in the wheat germ extract system.

3.2.3.4 A comparison of the densitometric analysis of the translation products of pMD5 with those of pMD2

A methionine was encoded at the usual position of the start codon for GUS in pMD2. It was necessary to ensure that the protein band representing GUS was caused by the cleavage of [CAT2AGUS] and not merely internal initiation at this start codon for GUS. Therefore a new plasmid pMD5 was constructed in which this methionine was removed. The densitometric analysis of the translation products of pMD5 was carried out in a similar fashion to that for pMD2 and the results compared to those for pMD2 (Table 3.2.6). The values were extremely close and within experimental error demonstrating that there was no measurable internal initiation at the methionine at the start of the GUS gene. Thus, all of the 70 kDa GUS protein product must have been produced by the cleavage event.

In vitro system	Translation reaction of the construct named	% Initiation at each Methionine named			% Cleavage based on GUS or corrected CAT2A	
		Met-1	Met-75	Met-163	GUS	CAT2A
Wheat germ Extract	pMD2 (Mean value)	37	32	31	81	97
	pMD5 (Fig. 3.2.1A)	37	33	32	77	99
Rabbit reticul-ocyte lysate	pMD2 (Mean value)	17	55	29	74	90
	pMD5 (Fig. 3.2.1B)	16	53	31	77	94

Table 3.2.6 Comparison of translation profiles of pMD2 and pMD5

3.2.3.5 Analysis of the stability of CAT2A and GUS protein products

Samples of the translation products of pMD2, pCAT2AGUS(3) and pMD5 from both rabbit reticulocyte lysates and wheat germ extracts were incubated at 30 °C and aliquots removed at intervals. These were analysed by PAGE to see whether either of the cleavage products, CAT2A and GUS, had undergone significant degradation. The polyacrylamide gels are shown in Figures 3.2.2 and 3.2.3 and the quantities of CAT2A and GUS, and the ratio of CAT2A : GUS, as measured by their measured PSL values, after the various incubation periods were depicted in graphical format (Figures 3.2.5 and 3.2.6). Examination of these graphs shows that there was some degradation of both products over the first 12 hours of incubation. The ratio of CAT2A : GUS decreases over the incubation period demonstrating that CAT2A must have been degraded faster than GUS.

3.2.3.6 Error analysis for the densitometry of phosphoimages

Quantitative analysis of the products of *in vitro* TnT reactions was carried out by densitometric analyses of products separated by PAGE and visualised by phosphoimaging. The measurement of total (PSL-background) values was obviously dependent on both the exposure time and the quantity of reaction loaded in a lane, so no correlation could be drawn between the (PSL-background) counts for a particular band from different exposures or gel-loadings. However, the ratios of (PSL-background) measurements for any two or more bands within a lane should be consistent (within experimental error) over different exposures and gel-loadings for any single sample.

There were, however, errors involved in making these measurements; when measuring *via* the integration of peaks, difficulty was encountered if the peaks were so close together that they overlapped to some extent, so raising the baseline (such as those peaks for [CAT2AGUS], [Δ^{75Met}CAT2AGUS] and [Δ^{163Met}CAT2AGUS]); and when measuring

98

Figure 3.2.5 Graphical depiction of the degradation of CAT2A and GUS
in rabbit reticulocyte lysate.   Presented here is the data from densitometric
analysis of the polyacrylamide gel in Figure 3.2.2.

Degradation of CAT2A and GUS proteins in wheat germ extract
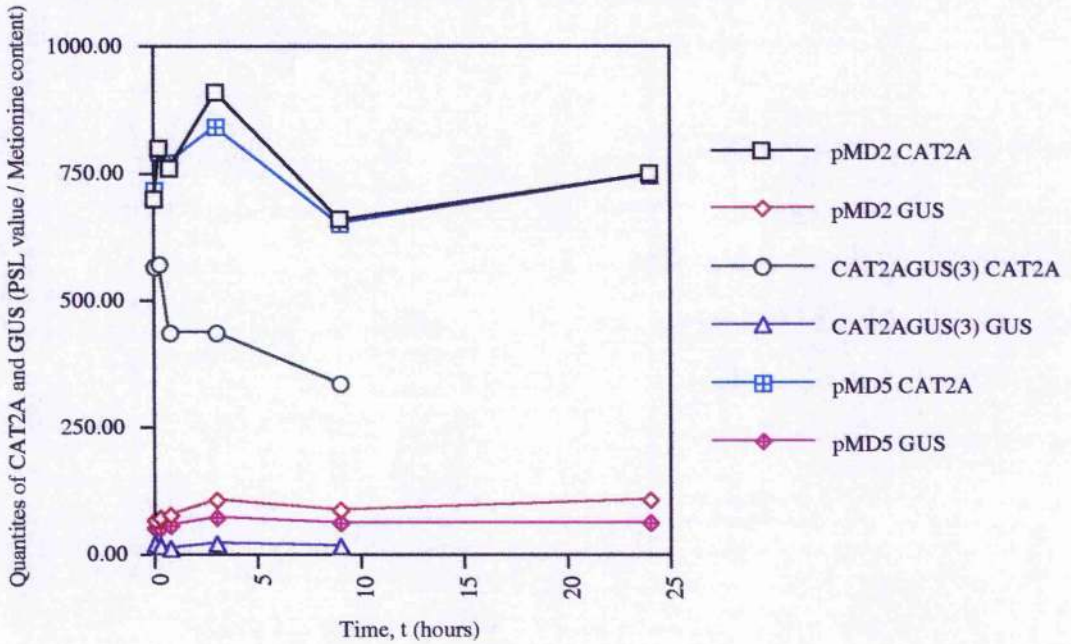
Variation of the ratio of CAT2A : GUS with incubation time in wheat germ extract
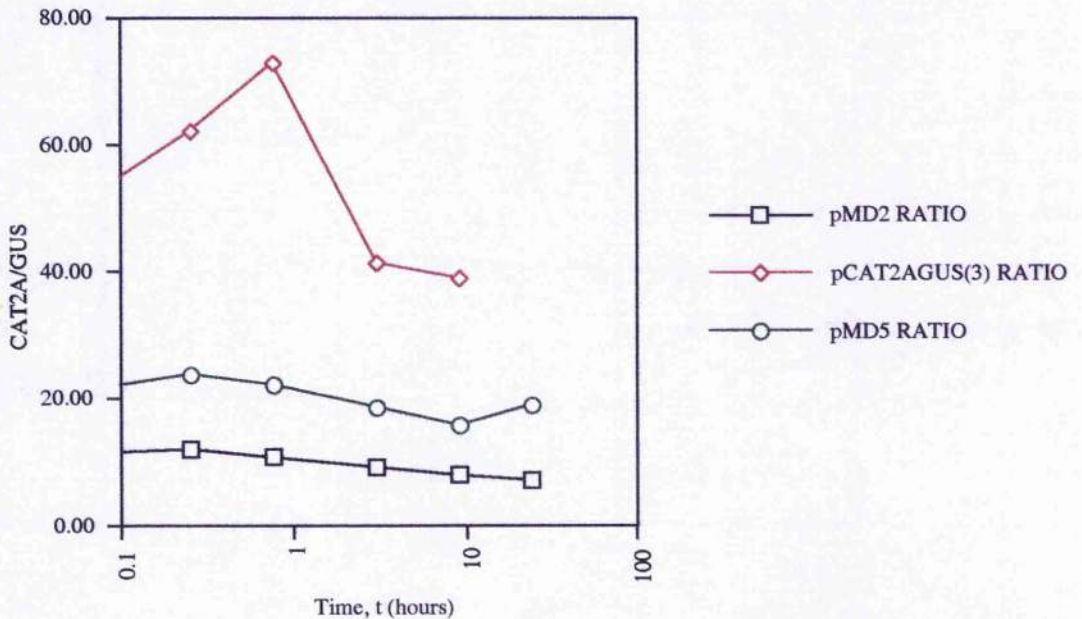
Figure 3.2.6 Graphical depiction of the degradation of CAT2A and GUS in wheat germ extract. Presented here is the data from densitometric analysis of the polyacrylamide gel in Figure 3.2.3

100

either by drawing boxes or drawing freehand around the bands then error was involved in choosing a representative background area, and in the accuracy of encircling solely and entirely the band of interest, especially if the band was weak in intensity, diffuse, or overlapping another.

The error in phosphoimaging densitometry was determined experimentally. A single rabbit reticulocyte lysate TnT reaction programmed with plasmid pMD2 was analysed by PAGE on four separate polyacrylamide gels, and each of these visualised using phosphoimaging techniques. Three separate measurements were made for the (PSL-background) for all of the major bands corresponding to the protein products of interest: [CAT2AGUS], [Δ^{75Met}CAT2AGUS], [Δ^{163Met}CAT2AGUS], GUS, and CAT2A. Since only ratios rather than individual measurements can be compared between gels and exposures, the ratio of (PSL-background) for each band to the total (PSL-background) of all bands measured was found and was expressed as a percentage of the total measured PSL for that lane. These expressions are directly comparable between different lanes and exposures using the same sample. Thus, a mean value and the standard deviation in the densitometric determination of each individual band measurement could be obtained. This data is presented in Table 3.2.2.

The information required from each gel lane was usually the percentage of initiation at Met^1, Met^{75} and Met^{163}, the ratio of CAT2A to GUS and the percentages of cleavage based on the quantities of CAT2A and GUS measured. This information was determined three times for each of the four gels using the three different ways of measuring discussed. The mean value and the standard deviation in determining these quantities is stated in Table 3.2.7.

3.2.4   Conclusions

- There is more CAT2A protein measured than GUS.

- This excess is not due to the degradation of GUS, since it has been shown that CAT2A was degraded faster than GUS.

- The only reasonable suggestion is, therefore, that there is more CAT2A *synthesised* than GUS. This is not consistent with the model of a co-translational auto-proteolytic cleavage occurring.

Product	Nº. Met	CAT2AGUS		75MetCAT2AGUS		163MetCAT2AGUS		GUS		corrected CAT2A		CAT2A GUS
		PSL/22	%	PSL/20	%	PSL/17	%	PSL/13	% cleaved	PSL/9	% cleaved	GUS
A Int.		23.27	14.21	105.15	63.80	36.39	22.08	719.15	81.36	2721.88	94.29	3.78
A Freehand		22.23	13.25	116.00	69.15	29.53	17.60	591.08	77.90	2133.36	92.70	3.60
A Boxed		28.91	16.10	101.40	56.49	49.28	27.44	725.53	80.16	2640.43	93.63	3.64
B Int.		6.58	17.03	22.91	59.29	9.51	23.68	137.77	78.10	430.16	91.76	3.12
B Freehand		5.86	15.89	23.42	63.09	7.84	21.12	116.92	75.90	514.79	93.27	4.40
B Boxed		5.65	14.34	24.07	61.08	9.69	24.59	148.46	79.02	595.96	93.79	4.00
C Int.		22.23	15.23	90.65	62.11	33.08	22.66	682.15	82.37	2969.27	95.31	4.35
C Freehand		17.62	15.57	75.15	66.42	20.37	18.00	580.38	83.69	3517.25	94.92	3.64
C Boxed		19.06	14.32	85.40	64.15	28.66	21.53	599.15	81.82	2115.16	96.35	5.80
D Int.		4.01	15.56	16.05	62.28	5.71	22.16	68.31	72.61	280.33	91.58	4.10
D Freehand		4.05	17.30	13.69	58.48	5.67	24.22	72.68	75.64	364.87	92.46	3.89
D Boxed		4.86	16.34	17.46	58.71	7.42	24.95	85.92	74.29	282.60	92.35	4.25
mean values		------	15.42±1.17	------	62.09±3.43	------	22.50±2.69	------	78.57±3.32	------	93.53±1.41	4.05±1 0.63

Table 3.2.7 Densiometric analysis of the PSL distribution for the rabbit reticulocyte TnT products of pMD2. The (PSL-background) measurements (from table 3.2.2) were divided by the number of methionines in each protein product to give a number proportional to the number of moles of protein within each protein band. The percentage of initiation at each of Met-1, Met-75, and Met-163 corresponds directly to the molar ratios of full length and truncated versions of CAT2AGUS. The values of PSL for CAT2A were corrected to a value which would have resulted from 100% initiation occuring at the first methionine by multiplying by a factor of 100 divided by the % Initiation at Met-1. Values for the ratio of CAT2A to GUS were calculated using the corrected value for CAT2A. The percentage cleavage of CAT2AGUS was calculated with respect to both the measured PSL for GUS and the corrected CAT2A value.

Section 3.3: Analysis of the activity of truncated cardiovirus 2A regions

3.3.1   Introduction

The full length 2A region of the TME virus shows 100 % cleavage activity when inserted into a foreign sequence. It has been demonstrated that the deletion of the amino terminal two thirds of EMCV 2A region to give a truncated 2A protein of 46 amino acids in length does not decrease cleavage activity (Palmenberg *et al.*, 1990).

It was proposed to examine the cleavage activity of the 2A regions of both EMCV-R and TMEV GDVII when all but the 19 C-terminal residues, analogous to the 2A region together with the last residue of 1D of FMDV, are deleted.

This section describes the construction of plasmids encoding [CAT-Δ2A-GUS] artificial polyprotein systems for both EMCV and TMEV, and an analysis of their translation products.

3.3.2   Results

3.3.2.1   Construction of pTME2AGUS

Plasmid pTMEGDVII was a kind gift of Dr. T.D.K. Brown (Law and Brown, 1990). The plasmid pTME2AGUS was constructed by Susan Monaghan in this laboratory. pCATGUS was doubly restricted with BamH I and Apa I and the large fragment isolated by agarose gel electrophoresis. The 2A region of TME strain GDVII was amplified from plasmid pTMEGDVII by PCR using the primers OR98a and OR99, doubly restricted with

Bgl II and Apa I. The restricted PCR product was then isolated by agarose gel electrophoresis and ligated into the restricted fragment of pCATGUS.

```
Protein Sequence:          S    R    M    P    A    S    L    Y    R
Oligo OR98a: 5' -  TTTTTT  AGA  TCT  ATG  CCC  GCA  TCT  CTC  TAC  CGC  -3'
                           Xba I   ⊢→TME 2A
```

```
Protein sequence:D   V    E    M    N    P    G    P    S    R    S    K
Sequence annealed
to by primer: 5'-GAT  GTA  GAA  ATG  AAC  CCC  GGG  CCC  TCC  AGG  TCC  AAA -3'
Oligo OR99: 3'-CTA  CAT  CTT  TAC  TTG  GGG  CCC  GGG  AGG  CCT  AGG  TTT -5'
                                   Sma I, Apa I
```

3.3.2.2   Construction of pCAT-ΔEMCV2A-GUS

The plasmid pMD5, encoding [CAT2AGUS], was doubly restricted with Xba I and Apa I, excising the 2A region of FMDV, and the 5443 base pair fragment purified by agarose gel electrophoresis. The oligonucleotides EMC-F and EMC-R were annealed forming a double stranded oligonucleotide adapter encoding for 19 C-terminal residues of the 2A region of EMCV between Xba I and Apa I compatible overhangs (see below).

```
Protein Sequence:    R    H    Y    A    G    Y    F    A    D    L    L    I    H    D    I
Oligo.EMC-F:5'-CTAGACACTACGCTGGTTACTTTGCGGACCTACTGATTCATGACATT
Oligo.EMC-R:3'-____TGTGATGCGACCAATGAAACGCCTGGATGACTAAGTACTGTAA
               Xba I
```

```
Protein sequence:    E    T    N    P    G
Oligo. EMC-F:      GAGACAAATCCAGGGCC    -3'
Oligo. EMC-R:      CTCTGTTTAGGTC____    -5'
                              Apa I
```

The adapter was ligated into the gel purified fragment of pMD5 and *E. coli* were transformed with the recombinant plasmid DNA, and selected from ampicillin plates for mini-preparations. The cDNA constructs were screened for the absence of an Aat II restriction site. Those which were not restricted were sequenced to check the insert was correct and then maxi-prepared using the Qiagen method.

105

3.3.2.3   Construction of pCAT-ΔTMEV2A-GUS

Forward and reverse primers, ORM1 and ORM2 respectively, were designed for the PCR to amplify the nucleic acid sequence encoding the 18 C-terminal residues of TME 2A, using pTME2AGUS as the template. ORM1 encoded an Xba I site prior to the 2A region, whilst ORM2 encoded the Apa I site at the 2A cleavage site, for cloning the DNA fragment into pMD2. In the TME 2A region either methionine or threonine can occupy the P4 position, depending upon the strain of TMEV. To permit this conservative mutation within the PCR product, 50 % of each of A and G was used for the second nucleotide in the codon at this P4 position. Thus, both versions of the TME 2A region can be obtained with either A$\underline{T}$G (encoding methionine), or (A$\underline{C}$G encoding threonine) at this position. The PCR product and the plasmid pMD2 were digested with Xba I and Apa I. The PCR product and the 4849 base pair fragment of pMD2 were then purified by agarose gel isolation, and were subsequently ligated together. *E. coli* were transformed with the recombinant plasmid DNA and then selected from ampicillin plates for mini-preparations. The cDNA constructs were screened for the absence of an Afl II restriction site. Those which were not restricted were sequenced to check the insert was correct. The DNA samples for both desired versions of TME 2A, together with two other samples containing mutations within the 2A region, were maxi-prepared using the Qiagen method.

```
Protein sequence:              S    R    Y    H    A    D    Y    Y
Oligo ORM1:  5'- GCGCGC  TCT  AGA  TAC  CAT  GCT  GAC  TAC  TAC -3'
                          Xba I


Protein sequence:              V    E    M    N    P    G    P
Region oligo anneals to :  5'- GAT  GAA   ATG   AAC  CCC  GGG  CCC -3'
Oligo ORM2:      ⁄         3'- CTA  CTT  TA/GC  TTG  GGG  CCC  GGG  CGCGCG -5'
                                                           Apa I
```

3.3.2.4 Transcription and translation *in vitro* of constructs containing truncated cardiovirus 2A regions

Maxi-preparations of pCAT2A, pGUS66/67, pMD2, pTME2AGUS, pCAT-ΔEMCV2A-GUS, pCAT-ΔTMEV2A-GUS, containing the wild type methionine, and pMD1/2(2) encoding [CAT-ΔTME2A-GUS], with threonine in the P4 position of 2A, were transcribed and translated in both the coupled rabbit reticulocyte lysate and wheat germ extract systems. The products were analysed by denaturing PAGE and visualised using phosphoimaging techniques (Figure 3.3.1A). The two plasmids pMD1/2(4) and pMD1/2(6) containing unintended mutations were also transcribed and translated in both coupled *in vitro* systems (Figure 3.3.1B).

3.3.3 Discussion

The sequences encoding the C-terminal 18 amino acids of EMCV-R, TME-GDVII wild-type and with the mutation of methionine to threonine as seen in other cardioviruses, were confirmed by nucleotide sequencing to have been successfully inserted between the CAT and GUS genes. In addition, two other unintended constructs were discovered during sequencing, in which mutated, truncated TME 2A sequences were inserted between CAT and GUS. These mutated sequences originated from the PCR extension of incorrect primers. Although unintentional, these constructs were of considerable utility in giving clues to the involvement of the mutated residues in the mechanism of 2A mediated cleavage. An alignment of the 2A sequences of all the constructs obtained is shown in Table 3.3.1.

107
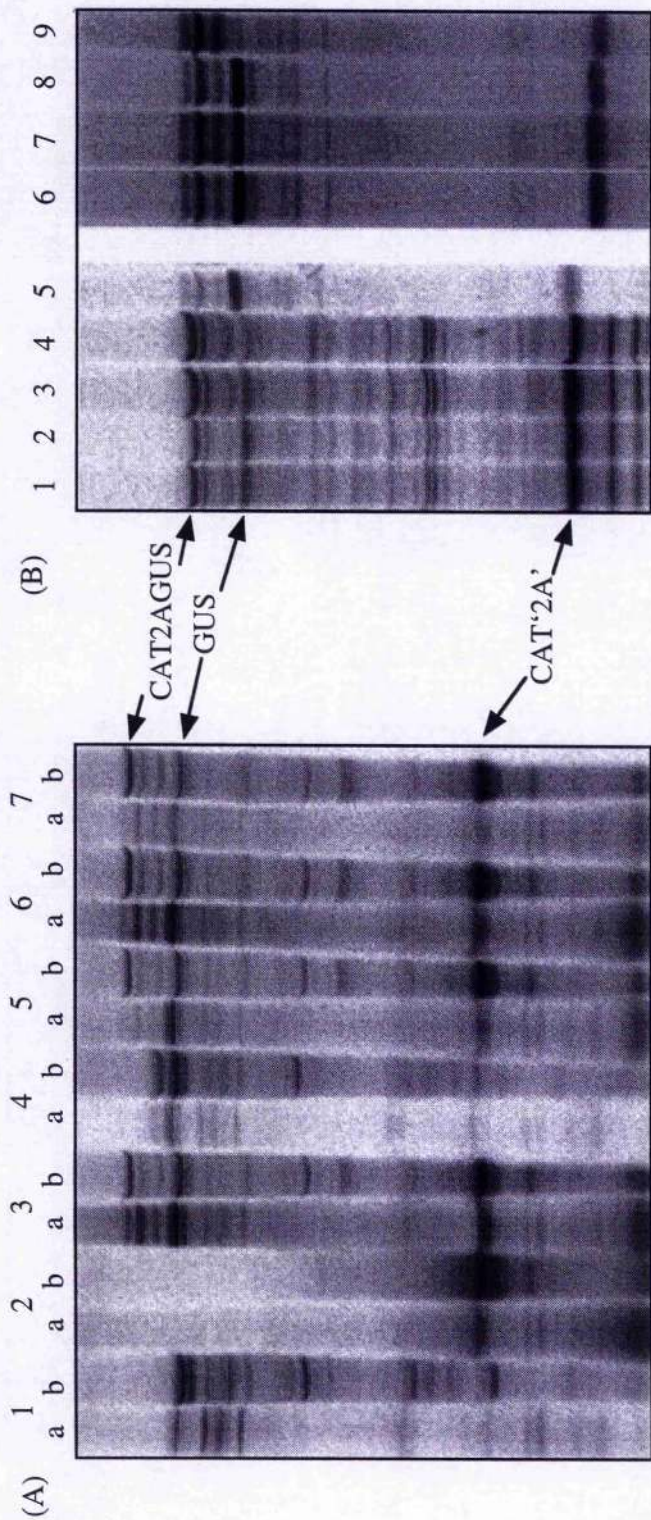
Figure 3.3.1 SDS-PAGE of the translation products of constructs encoding CAT- Cardiovirus 2A-GUS polyproteins.

(A)   Lanes entitled 'a' are from rabbit reticulocyte lysate coupled reactions, whilst those lanes entitled 'b' are from
wheatgerm extract coupled reactions programmed with the following DNA:

1. pGUS66/67   3. pMD2   5. pCATΔEMCV2AGUS   7. pCATΔTMEV2AGUS
2. pCAT2A   4. pTME2AGUS   6. pMD1/2(2)

(B)   Coupled reactions (1-5 in wheatgerm extract, 6-9 in rabbit reticulocyte lysate) of constructs encoding mutated
TMEV2A regions in CAT- Cardiovirus 2A-GUS polyproteins

1. pCATΔTMEV2AGUS   3. pMD1/2(4)   5. pMD2   7. pMD1/2(4)   9. pMD1/2(6)
2. pMD1/2(2)   4. pMD1/2(6)   6. pMD1/2(2)   8. pMD1/2(2)

108

pCATΔEMC2AGUS

S	R	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P

TCT AGA CAC TAC GCT GGT TAC TTT GCG GAC CTA CTG ATT CAT GAC ATT GAG ACA AAT CCA GGG CCC

pCATΔTME2AGUS

S	R	Y	H	A	D	Y	Y	K	Q	R	L	I	H	D	V	E	M	N	P	G	P

TCT AGA TAC CAT GCT GAC TAC TAC AAA CAG AGA CTC ATA CAC GAT GTA GAA ATG AAC CCC GGG CCC

pMD1/2(2)

S	R	Y	H	A	D	Y	Y	K	Q	R	L	I	H	D	V	E	T	N	P	G	P

TCT AGA TAC CAT GCT GAC TAC TAC AAA CAG AGA CTC ATA CAC GAT GTA GAA ACG AAC CCC GGG CCC

pMD1/2(4)

S	R	Y	H	A	D	Y	S	K	Q	R	L	A	H	D	V	E	M	N	P	G	P

TCT AGA TAC CAT GCT GAC TAC TCC AAA CAG AGA CTC ATA CAC GAT GTA GAA ATG AAC CCC GGG CCC

pMD1/2(6)

S	R	Y	H	A	D	Y	Y	K	Q	R	L	I	H	N	V	E	I	N	P	G	P

TCT AGA TAC CAT GCT GAC TAC TAC AAA CAG AGA CTC ATA CAC AAT GTA GAA ATC AAC CCC GGG CCC

Table 3.3.1   Alignment of cloned cardiovirus 2A regions. The cloned sequences of cardiovirus 2A regions, from the Xba I to Apa I restriction sites (underlined), which have been inserted within CAT/GUS reporter system are shown. The protein sequences are shown above the nucleotide sequences. Mutations from the wild type sequences are shown in bold.

Figure 3.3.1A shows the PAGE separation of the rabbit reticulocyte lysate and wheat germ extract translation products of pGUS 66/67 and pCAT2A as controls, alongside pMD2, pTME2AGUS, pCATΔEMCV2AGUS, pCATΔTMEV2AGUS, and pMD1/2(2). The lanes for pGUS66/67 contained a band for GUS at 70 kDa as well as numerous smaller internal initiation products. Both lanes for pCAT2A demonstrated a strong band for CAT2A. In the rabbit reticulocyte lysate lane, in particular, there was also a strong band very close to the bottom of the gel, which corresponds to a truncated CAT2A product caused by internal initiation within the CAT gene, followed by cleavage by 2A. This band was evident in all the translation profiles of constructs containing the CAT gene. The lanes representing the translation of pMD2, encoding [CAT-FMDV2A-GUS], both demonstrated strong bands for CAT2A and GUS proteins as well as lesser bands for full length and truncated forms (from internal initiation within CAT) of uncleaved [CAT-FMDV2A-GUS] as previously described in Section 3.2. The translation profiles for pTME2AGUS

109

encoding the full length TME 2A protein and GUS in a single ORF demonstrated in the case of rabbit reticulocyte lysates, a faint band corresponding to uncleaved [TME2AGUS] polyprotein and a band at 70 kDa corresponding to GUS; in the case of the wheat germ extract reaction, although it was evident that there was a greater amount of protein in total, the percentage of uncleaved [TME2AGUS] polyprotein appeared to be greater. However, there was still a large quantity of protein corresponding to GUS. The TME 2A protein is only 133 amino acids in length (15.5 kDa), and as such was to small to be visualised on this gel. In the next six lanes the translation products for constructs encoding [CAT-ΔCardiovirus2A-GUS] polyproteins were very similar to those for pMD2; there were bands corresponding to uncleaved polyprotein, and the cleaved products GUS and CATΔ2A. The lanes representing the translation products for pCAT-ΔEMCV2A-GUS especially when translated in rabbit reticulocyte lysates appeared by eye to have cleaved as well as, if not better than [CAT-FMDV2A-GUS].

A densitometric analysis of the gels in Figure 3.3.1, and others where available, was used to determine ratios of CAT2A to GUS and percentages of cleavage, based on either CAT2A or GUS, in both *in vitro* systems, for all of the [CAT-ΔCardiovirus2A-GUS] polyproteins (Table 3.3.2). This analysis demonstrated that [CAT-ΔEMCV2A-GUS] was indeed processed as effectively as the [CAT-FMDV2A-GUS] polyprotein. Unexpectedly the [CAT-ΔTMEV2A-GUS] polyprotein with the mutation to threonine at the P4 position, cleaves more efficiently than that with the wild-type sequence, and at a level equivalent to that containing FMDV 2A. The mutation of tyrosine to serine in the TME 2A sequence was slightly deleterious to the cleavage activity, reducing the percentage cleavage from about 92 % to 85 % with respect to CAT2A in rabbit reticulocyte lysates. This construct and that with the wild-type TME 2A produced less GUS protein than the other constructs. This resulted in much greater ratios for CAT2A : GUS. The general trends in the values calculated for reactions in rabbit reticulocyte lysates were mirrored and often accentuated in the values obtained for wheat germ extract reactions. Individual calculated values for cleavage varied little, but the individual ratios calculated for CAT2A : GUS did vary,

Construct	Rabbit reticulocyte lysate reactions			Wheatgerm extract reactions		
	Ratio $\frac{CAT2A}{GUS}$	% Cleavage based on CAT2A or GUS		Ratio $\frac{CAT2A}{GUS}$	% Cleavage based on CAT2A or GUS	
		CAT2A	GUS		CAT2A	GUS
pMD2	3.5±0.4	94±2	80±1	14±3	98±1	77±5
pCAT- EMCV2A-GUS †	2.1±1.3	98±1	97±1	17±4	97±1	65±4
pCAT- TMEV2A-GUS	8.0±1.6	92±2	61±3	23±2	95±1	42±3
pMD1/2(2)	4.5±1.7	93±3	78±4	16±2	96±0.4	56±2
pMD1/2(4)	7.7±0.6	85±1	42±1	21±2	91±1	31±2

Table 3.3.2 Results of the densitometric analysis of the translation products seen in Figure 3.3.1. The densitometric analysis was carried out exactly as described in Results section 2. The densitometric analysis of the gels in Figure 3.3.1 was compared to those of other gels (not shown) to obtain the mean values shown in the table and values for the standard deviation of measurements.

† It was not possible to measure directly the percentage of initiation at the first methionine of this translation since there is so little uncleaved material, however, by comparison with the translation products from other constructs translated at the same time it was possible to estimate this percentage to be 17 %. Using this estimation the above values were calculated for pCAT- EMCV2A-GUS.

depending on the batch of rabbit reticulocyte lysate or wheat germ extract. There was, however, consistently a greater ratio of CAT2A:GUS in wheat germ extract reactions than in rabbit reticulocyte lysate reactions. The double mutation of aspartate and methionine to asparagine and isoleucine, respectively, inactivated the 2A region, such that three major bands for [CAT2AGUS] and its truncated versions were seen. An additional band corresponding to a protein just smaller than CAT2A was also seen but could not be identified. This result indicated that either the aspartate or the methionine or both were required for activity. It is most likely that the aspartate residue is essential for activity, since mutation of the methionine alone, although conservative, was not at all detrimental to activity.

3.3.4   Conclusions

- The 18 C-terminal amino acids of TMEV 2A and EMCV 2A are capable of mediating cleavage in rabbit reticulocyte lysate and wheat germ extract to approximately the same extent as the FMDV 2A region.

- The mutation of methionine to threonine slightly increased the cleavage activity of TMEV 2A. The mutation of tyrosine to serine in TMEV 2A decreased cleavage activity by about 8 %. The double mutation of aspartate and methionine to asparagine and isoleucine, respectively, resulted in an inactive 2A region (see Figure 3.3.2). Thus, the aspartate residue is most probably required for 2A mediated activity.

Figure 3.3.2   The TME 2A C-terminal sequence and mutations inserted within.

112

Section 3.4 Mutagenesis of the FMDV 2A sequence

3.4.1 Introduction

It was proposed to investigate the mechanism of the FMDV 2A cleavage event *via* site directed mutagenesis. The codons for individual amino acids, within the region of pMD2 encoding 2A, were mutated to test the roles of these amino acids. The accepted system for reference to individual residues of a protease substrate is that of Scheuchter and Berger (1967) by which P1, P2, *etc.* refer to residues progressively N-terminal of the cleavage site, and P1', P2', *etc.* are residues progressively C-terminal to the hydrolysed peptide bond. Based upon this system, the numbering system for the amino acid residues of 2A, shown below, will be used throughout this section.

							2A										⇓ 2B
P18	P17	P16	P15	P14	P13	P12	P11	P10	P9	P8	P7	P6	P5	P4	P3	P2 P1	P1'
L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P G	P

One would expect the residues involved in the cleavage event to be well conserved. Indeed there is complete homology between all aphthoviruses from P13-leucine to the P1'-proline. The truncated 2A sequence -LKLAGDVESNPGP- is the shortest sequence shown to exhibit activity (Ryan *et al.*, 1991). Thus it was residues from this region, especially those shown to be highly conserved in sequence alignments of both the aphtho- and cardioviruses, which were of greatest interest in site-directed mutagenesis studies. Dynamic molecular modelling studies within our laboratory, and the alignment of the aphtho- and cardiovirus 2A regions (Figure 1.5.2) were used to determine which mutations to make.

This chapter describes the construction of a series of plasmids with site directed mutations within the 2A coding sequence, and their transcription and translation in *in vitro* systems.

3.4.1.1 Residues chosen for site directed mutagenesis of FMDV 2A

The residues glutamic acid, asparagine and serine had already been mutated within the construct pCAT2AGUS(3) in our laboratory; Glutamic acid to glutamine and aspartic acid; asparagine to histidine, glutamic acid and glutamine; and serine to phenylalanine and isoleucine. Only confirmation of their phenotypes was required.

It was firstly chosen to mutate P11-lysine, since the truncated sequence, which has up to and including this lysine deleted, -LAGDVESNPGP-, was shown to be inactive (Ryan *et al.*, 1990). Mutation to arginine would preserve the basic nitrogen, which is positively charged at physiological pH. Conversion to glutamine would remove the basic nitrogen functionality, and the residue would be neutral.

Our model suggested that the P7-aspartic acid residue's role may have been to activate a nucleophile to attack the scissile bond. This was supported by the low energy molecular modelling structure for 2A, shown in Figure 1.6.1. Hence it was decided to mutate this aspartic acid to glutamic acid, asparagine and glutamine. The mutant peptide containing glutamic acid might be expected to retain some degree of activity. Those containing neutral, non-nucleophilic residues asparagine and glutamine, although maintaining the approximate size of aspartic acid and glutamic acid respectively, would be expected to be inactive.

The small size of the P1-glycine adjacent to the scissile bond was thought to be important for activity since it could allow access of a nucleophile. Mutation to the more bulky groups, alanine and valine, might then cause a progressive reduction in activity.

Proline residues are known to have strong influences on protein and peptide structure, and hence it was thought that both of the proline residues in 2A were probably important in the formation of an active peptide. It was of particular interest to mutate the P1'-proline since

114

dynamic molecular modelling studies suggested that a requirement for 2A activity might be a *cis* glycine-proline amide bond. Mutation of this proline residue to any other amino acid would eliminate the possibility of a *cis*-amide bond. The mutation to serine was of particular interest since serine is known to be the closest natural amino acid analogue to proline; H-bonding between the hydroxyl-hydrogen and nitrogen means that serine under some circumstances may be structurally similar to proline, although it exists only in a *cis*-amide bond configuration.

It was proposed to change P5-glutamic acid and P3-asparagine as a double mutation, since the molecular modelling studies suggested they might be involved in the formation of a hair-pin loop motif during folding, which was thought to place the scissile bond in an accessible position for hydrolysis.

The -KLAG- motif is found throughout the aphthoviruses, but not in the cardioviruses. In the cardioviruses it was thought that this might be represented by a conserved -IH- motif, so it was decided to replace the residues KLAG with either IH or HI.

The final site directed mutations that were planned were intended to disrupt an area predicted to be α-helical by molecular modelling studies. It was planned to make insertions of proline as a helix breaker, or alanine as a spacer residue, both between the P10-leucine and P9-alanine, and the P13- and P14-leucines. The alanine spacer residues would break up the proposed amphipathic nature of the helix as shown in Figure 1.6.1.

If the 2A-mediated cleavage is a proteolytic reaction then under conditions where there is no significant degradation, as has been shown for the *in vitro* studies of [CAT2AGUS] activity, one would expect to see 1 : 1 ratios of the cleavage products. However an exces of the protein preceding the cleavage site is always seen. The lack of stoichiometry in the 'cleavage' products of [CAT2AGUS] led to the hypothesis that the 2A region might be mediating cleavage by somehow interfering with translation. To investigate whether

115

'cleavage' occurs at the level of translation of mRNA to protein, or post-translationally as a peptide cleavage it was proposed to alter the reading frame of 2A. Hence the mRNA sequence would remain the same throughout 2A but the protein encoded would be completely different.

3.4.2   Results

3.4.2.1   Design of oligonucleotide adapter molecules

A series of oligonucleotide double stranded adapters were designed to introduce some of the required site directed mutations by their insertion between the unique restriction sites created within 2A. The sequences of these oligonucleotides are listed in Table 3.4.1. The amino acid sequences encoded by the oligonucleotides are shown above the second nucleotide of each codon, and the restriction sites contained within each is shown below each pair.

3.4.2.2   Construction of mutant 2A sequences using double stranded adapter molecules

Prior to the discovery of the mutation within the GUS gene of plasmid pMD1, pMD1 was restricted and modified to create suitable vector fragments for the insertion of double stranded oligonucleotide pairs as follows: pMD1.1- pMD1 was doubly restricted with Aat II and Afl II; pMD1.2, pMD1.3, pMD1.4 and pMD1.6- pMD1 was doubly restricted with Bgl II and Aat II; pMD1.5- pMD1 was restricted with Afl II then treated with Mung Bean nuclease to remove the overhang, before digesting with Aat II; pMD1.7- pMD1 was digested with Aat II, then treated with T4 DNA polymerase to remove the overhang before digesting with Afl II. The large restricted vector fragments were then purified by agarose

116

Mutant series/ Oligonucleotides	Sequences
pMD2.1 Oligo. OMD1 Oligo. OMD2	``` L K L A G N V TTAAGCTTGCGGGAAACGT CGAACGCCCTT ```
pMD2.2 Oligo. OMD3 Oligo. OMD4	``` A E S N P V P F F F T S CGAGTCCAACCCTGYGCCCTTTTTTTTTTACTAGTA TGCAGCTCAGGTTGGGACRCGGGAAAAAAAAAATGATCATCTAG Spe I ```
pMD2.3 Oligo. OMD5 Oligo. OMD6	``` E S N P G Z F F F T S CGAGTCCAACCCTGGGNNNTTTTTTTTTTACTAGTA TGCAGCTCAGGTTGGGACCCNNNAAAAAAAAAATGATCATCTAG Spe I ```
pMD2.4 Oligo. OMD7 Oligo. OMD8	``` B S J P G P F F F T S CXASTCCSASCCTGGGCCCTTTTTTTTTTACTAGTA TGCAGZTSAGGSTSGGACCCGGGAAAAAAAAAATGATCATCTAG Apa I ```
pMD2.5 Oligo. OMD15 Oligo. OMD16	``` Q L R L A G D V TCCRACTTGCGGGAGACGT AGGYTGAACGCCCTC ```
pMD2.6 Oligo. OMD11 Oligo. OMD12	``` E S N Z G P F F F T S CGAGTCCAACNNNGGGCCCTTTTTTTTTTACTAGTA TGCAGCTCAGGTTGNNNCCCGGGAAAAAAAAAATGATCATCTAG Apa I Spe I ```
pMD2.7 Oligo. OMD13 Oligo. OMD14	``` E L K L A G Q V TTAAGCTTGCGGGASAGGT CGAACGCCCTSTCCA ```

Table 3.4.1 Oligonucleotides used as double stranded adapters for SDM

The series of mutant constructs which each oligonucleotide pair is used to make is named in the left column above the oligonucleotide names. The amino acid sequences encoded by the oligonucleotides are shown above the second nucleotide of each codon, and the restriction sites contained within each is shown below each pair.

Key: Within nucleotide sequences, R = A or G   W = C, G or T

S = C or G   X = A, C or G

N = A, C, G or T   Y = C or T

Within peptide sequences,   Z = Any residue

B = N, K, H, Q, D, or E

J = H, Q, D, or E

Mutated residues and codons shown in red.

117

gel electrophoresis. Oligonucleotide pairs were annealed and then ligated into the appropriately restricted and purified vector fragment. *E. coli* were transformed with the recombinant plasmid DNA, and selected from ampicillin plates for mini-preparations. The plasmids isolated were in all possible cases primarily screened by restriction digests to identify those containing mutations (see Table 3.4.2). All of the plasmid constructs were then sequenced to determine or confirm their identities.

Upon the discovery of the stop codon within the GUS gene of pMD1 (see Section 3.2), the mutated sequences were transferred into the plasmid pMD2. pMD2 was linearised by a restriction digest with Xmn I, cutting within the ampicillin resistance gene, then restricted with Bgl II to produce two fragments of 3173 and 1809 base pair. The 3173 base pair fragment, containing the corrected GUS gene was purified by agarose gel isolation. The panel of pMD1 based vectors containing site directed mutations within 2A were similarly restricted producing two fragments of 3729 and 1809 base pairs. The 1809 base pair fragments of each mutant pMD1 vector, containing CAT and the mutated 2A region, were also gel purified, then ligated to the 3173 base pair fragment of pMD2. Each ligation mixture was transformed into *E. coli*, and the DNA mini-prepared, sequenced and maxi-prepared.

Table 3.4.2. **Change of restriction mapping in creation of mutated pMD1 sequences**

Mutant plasmid series	Residue mutated	Change in restriction mapping
pMD1.1	P7-D	Aat II destroyed
pMD1.2	P1-G	Apa I destroyed
pMD1.3	P1'-P	Apa I destroyed
pMD1.4	P5-E / P3-N	None
pMD1.5	P11-K	Afl II destroyed
pMD1.6	P2-P	None
pMD1.7	P7-D	Aat II destroyed

3.4.2.3 Construction of plasmids replacing the KLAG motif with IH or HI

Plasmid constructs, in which the codons encoding the -KLAG- motif were replaced by those for -IH-, -HI- and -RI-, were made by using the PCR to amplify a mutated 2A region to replace the wild-type 2A region of pMD2. Three primers were designed to prime PCR fragments encoding CAT and the part of 2A containing the required mutation from the template pMD2. The oligonucleotide ORM31 was used as the forward primer for the reactions, introducing a BamH I site before the beginning of the CAT gene. ORM5 and ORM6 were designed to encode the replacement motifs for the residues KLAG, IH and HI respectively. A degeneracy within the codon introduced to encode the histidine residue, *i.e.* CAC (for His) or CGC (for Arg), would also allow arginine be produced in this position, which would also be of interest. Thus the primers were designed such that the second nucleotide of the codon for histidine was 50 % A and 50 % G (or since this is a reverse primer 50 % T and 50 % C). The PCR products and pMD2 were restricted with BamH I and Aat II. The 4823 base pair fragment from pMD2 and the restricted PCR products were purified by agarose gel electrophoresis, and the PCR products each ligated into the cut pMD2 vector. The ligation reactions were used to transform *E. coli* which were picked from ampicillin plates, for DNA mini-preparation. The resultant plasmid DNA was screened by restriction digest with Afl II. Those cDNA constructs lacking an Afl II site were sequenced and the mutations encoding -HI-, -RI-, and -IH- were isolated in plasmids entitled pMD31/6(c), pMD31/6(1),and pMD31/5 respectively.

```
Protein sequence:                   M    E    K    K
ORM31:    5'- GCGCGC GGA TCC GATG GAG AAA AAA -3'
                       BamH I   ⊢→CAT
```

```
Protein Sequence encoded:  L   N   F   D   L   L   I   R/H   D   V
Coding sequence:     5'- TTGAATTTTGACCTTCTTATCCG / ACGACGTCGCGCGC -3'
ORM5:               3'- AACTTAAAACTGGAAGAATAGGC / TGCTGCAGCGCGCG -5'
                                                      Aat II
```

```
Protein Sequence encoded:  N   F   D   L   L   R/H   I   D   V
Coding sequence::    5'- AATTTTCAGCTTCTTCG / ACATCCAGGTCGCGCGC-3 '
ORM6:               3'- TTAAAAGTCGAAGAAGC / TGTAGGTCCAGCGCGCG -5'
                                                   Aat II
```

3.4.2.4 Construction of plasmids encoding for proline or alanine insertions between P13 and P12 positions, and P10 and P9 positions.

Two oligonucleotides, ORM10 and ORM11, were designed as reverse primers for PCR reactions with the forward primer ORM31 and template pMD2, to create PCR products encoding CAT and some of 2A, with a proline or alanine inserted between P10 and P9, and P13 and P12, respectively, *i.e.* -KL<u>A/P</u>AG- and -DL<u>A/P</u>LK-. Both primers had 50 % C and 50 % G incorporated at the position complementary to the first nucleotide of the alanine (<u>G</u>CC) or proline (<u>C</u>CC) codon, thereby allowing two products to be produced from the single PCR reaction. ORM10 encoded a new unique Sac I site which was useful in screening later.

```
Protein sequence encoded:   Q   L   L   K   L   A/P   A   G   D   V
Complementary sequence
 to primer:            5'- CAGCTTCTTAAGCTTG / CCCGCGGGAGACGTCGCGCGC-3'
ORM10:                 3'- GTCGAAGAATTCGAAC / GGGCGCCCTCTGCAGCGCGCG-5'
                                               Sac I      Aat II

Protein sequence:               L   N   F   D   L   A/P   L   K
Complementary sequence:   5'- TTGAATTTTGACCCTG / CCCCTTAAGGCGCGC -3'
ORM11:                    3'- AACTTAAAACTGGGAC / GGGGAATTCCGCGCG -5'
```

pMD2 was digested with BamH I and Aat II producing 718 and 4804 base pair fragments, and with BamH I and Afl II producing 699 and 4823 base pair fragments. The large fragments were purified by agarose gel electrophoresis. The PCR product produced using the oligonucleotides ORM10 and ORM31 was digested with BamH I and Aat II, whilst the PCR product produced using the oligonucleotides ORM11 and ORM31 was digested with BamH I and Afl II. Both products were then purified by agarose gel electrophoresis, then ligated into the similarly restricted and purified pMD2 vector fragment. The ligation mixtures were used to transform *E. coli* which were picked from ampicillin plates for miniprepations of plasmid DNA. Those constructs originating from the ligation of the PCR product produced using ORM10 were screened with the restriction enzyme Sac I. The DNA constructs which were linearised with Sac I were sequenced and both mutated

120

sequences encoding -KL<u>A</u>AG- and -KL<u>P</u>AG- were isolated in plasmids entitled pMD31/10 (5), and pMD31/10(3), respectively. The mini-prepared DNA originating from the ligation of the PCR using ORM11 could not be screened by restriction mapping and was sequenced directly. Only the mutation -DL<u>P</u>LK- and the unexpected mutation -D<u>PP</u>LK- (caused by amplification from an incorrect primer) were obtained, in plasmids entitled pMD31/11(a) and pMD31/11(2), respectively.

3.4.2.5 Construction of plasmids pCAT2AGUS 31/10 (5) and pMD6.7

Plasmid constructs pMD31/10 (5) and pCAT2AGUS 06 (7) were doubly digested with Xmn I (within the β-lactamase gene) and Apa I (at the 3' end of 2A). Both fragments from each plasmid were purified by agarose gel electrophoresis. The large fragment from pMD31/10 (5) was ligated to the small fragment of pCAT2AGUS 06 (7). The resultant plasmid, named pCAT2AGUS 31/10 (5), was obtained by transformation into *E. coli* and screening of DNA mini-preparations. The large fragment from pCAT2AGUS 06 (7) was ligated to the small fragment of pMD31/10 (5). The resultant plasmid, named pMD6.7, was also obtained by transformation into *E. coli* and screening of DNA mini-preparations.

3.4.2.6 Molecular cloning to investigate the 'interruption in translation' hypothesis

It was proposed to alter the reading frame of 2A to investigate whether 'cleavage' occurs at the level of translation of mRNA to protein, or post-translationally as a protein cleavage. The mRNA encoding 2A would thus remain (almost) unchanged but the protein encoded would be completely different. The protein sequences obtained from all three reading frames of the DNA sequence encompassing 2A are shown over leaf.

Frame		BbuI										AflII			
	2A	GGA	<u>GCA TGC</u>	CAG	CTG	TTG	AAT	TTT	GAC	CTT	<u>CTT AAG</u>	CTT	GCG	GGA
0		G	A C	Q	L	L	N	F	D	L	L K	L	A	G
+1		E	H A	S	C	*	I	L	T	F	L S	L	R	E
-1		S	M P	A	V	E	F	*	P	S	*	A	C	G

													Bgl II	
	GAC	GTC	GAG	TCC	AAC	CCT	GGG	CCC	TTT	TTT	TTT	ACT	AGT	<u>AGA TCT</u>
0	D	V	E	S	N	P	G	P	F	F	F	T	S	R S
+1	T	S	S	P	T	L	G	P	F	F	L	L	V	D –
-1	R	R	R	V	Q	P	W	A	L	F	F	Y	*	* I

It was chosen to remove a nucleotide before 2A thus converting it to the +1 reading frame, since this created only one stop codon (highlighted in blue) which was further from the cleavage site than the two created in the -1 frame. The stop codon thus created can be destroyed by changing the codon TTG for leucine in 2A to CTT which would also have encoded leucine in the correct reading frame for 2A. An overview of the cloning strategy is shown in Figure 3.4.1.

Two oligonucleotides, ORM7 and ORM8, were designed such that they would form a double-stranded oligonucleotide adapter molecule, to fit between the Bbu I site (position 723) and the unique Afl II site at 746 base pairs in pMD2. These encoded a +1 frameshift at the start of 2A.

```
Protein sequence encoded:           C  S  C  L  I  L  T  F
ORM7                      5'-      CAGCTGCTTAATTTTGACCTTC       -3'
ORM8                      3'-  GTACGTCGACGAATTAAAACTGGAAGAATT  -5'
                                  ↑ C omitted to change to +1 frame
```

pMD2 was partially digested with Bbu I and the linearised 5522 base pair band, which contains plasmid linearised at position 2628 as well as the required 723, was purified by agarose gel isolation. This was then digested completely with Afl II to produce the four bands at 23, 1882, 3640, and 5499 base pairs. The latter band was purified by agarose gel electrophoresis. The oligomers ORM7 and ORM8 were annealed and ligated into the cut pMD2 vector. However, transformation of E. coli with the ligation mixes and a control containing ligated restricted pMD2 vector showed a high background of linearised pMD2.

122
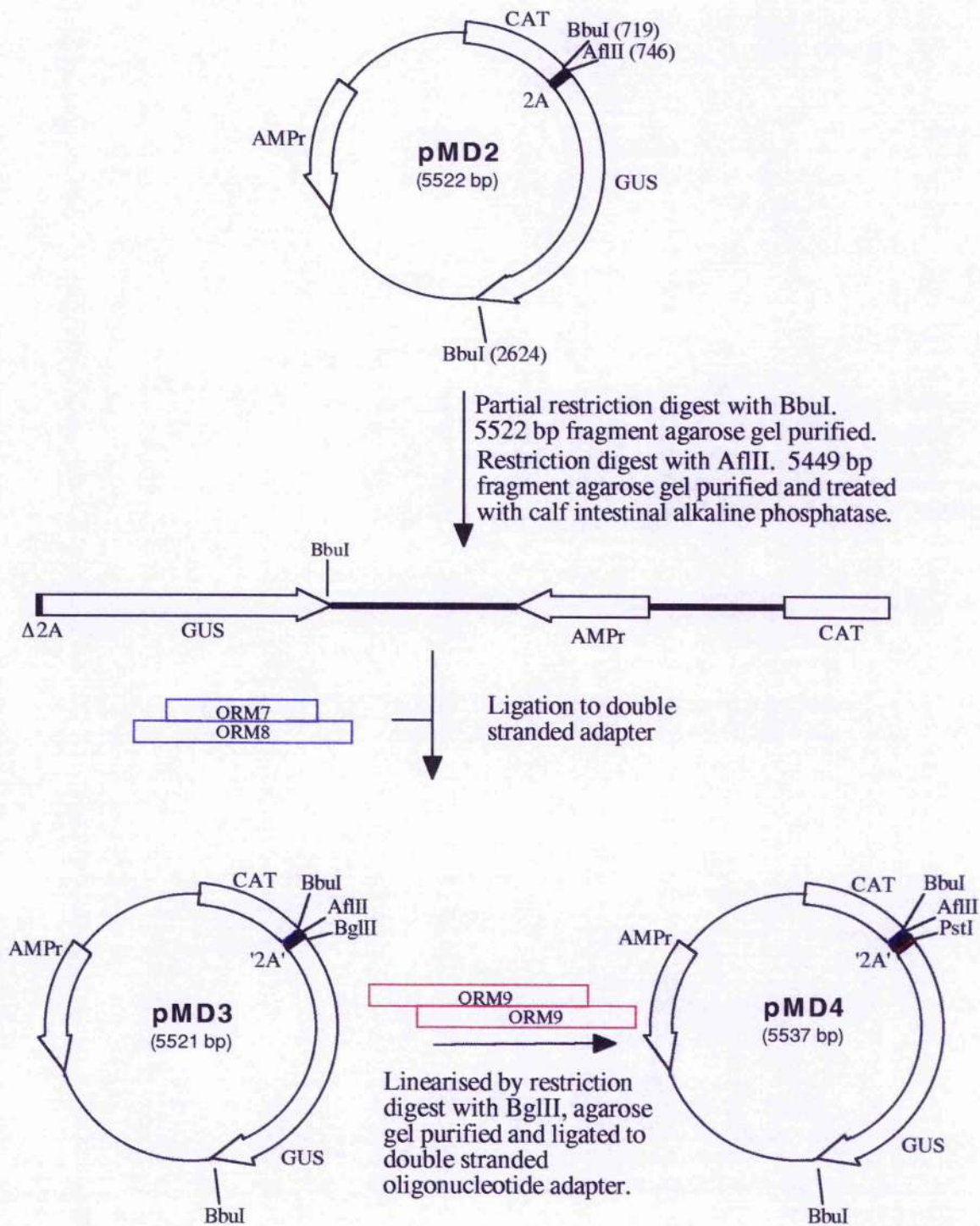
Figure 3.4.1 Construction of plasmid pMD4.

N.B. Site positions of restriction enzyme sites are given as the first nucleotide of the recognition sequence

123

This was not surprising since the isolated fragment at 5499 base pairs would not be separated from any remaining 5522 base pair linearised DNA *via* agarose gel purification. Therefore, the restricted vector was also treated with calf intestinal alkaline phosphatase to prevent recircularisation of the contaminating linearised vector. The double stranded oligonucleotide adapter molecules were treated with T4 DNA polynucleotide kinase and then ligated to the restricted vector fragment to form the new construct pMD3. *E. coli* were transformed with the ligation mixtures and five colonies were used for mini-preparations of DNA. The mini-prepared DNA was linearised with Bgl II and two samples appeared to be of the correct size.

In order to return the reading frame back to normal for the GUS gene a second double stranded adapter molecule was designed. The oligomer ORM9, of 16 nucleotides ($3n+1$ nucleotides, where n is a whole number, are required to decrease the reading frame by 1) was designed such that it could be inserted after 2A in the Bgl II site. ORM9 was palindromic, thus the direction of insertion of the adapter was not a factor requiring consideration and only one oligomer was required which could be annealed to itself to form a double stranded adapter. The oligonucleotide encoded a Pst I site, which would be unique in the new plasmid and useful for screening.

```
ORM9   5'-  GATC GGG CTG CAG CCC  -3'
ORM9          3' CCC GAC CTG GGG CTAG  -5'
                     Pst I
```
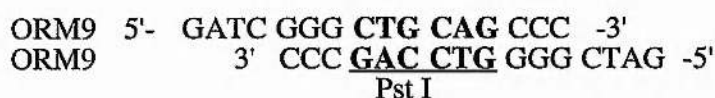
The oligonucleotides were annealed to form the adapter molecule, which was then treated with T4 DNA polynucleotide kinase. pMD3 which had been linearised with Bgl II was treated with calf intestinal alkaline phosphatase to prevent recircularisation of the vector and ligated to the adapter molecule. The ligation mixtures were transformed into *E. coli* and the mini-preparations of DNA screened for the Pst I site. Mini-preparations containing the site were then maxi-prepared and sequenced using ^{32}P end-labelling dideoxy cycle sequencing techniques. However although the sequence confirmed that the double stranded oligonucleotide adapter molecule ORM7&8 had been inserted correctly before the

2A region, a compression of the sequence within and after the adapter ORM9 suggested that these cDNA products may have been incorrect. Sequencing of the mini-preparations was repeated using a [^{35}S] direct incorporation sequencing technique, with ITP rather than GTP to reduce the occurrence of compressions. This revealed that the sequence over the area in question was, indeed, correct, although there was some 'four-tracking' to the 5' of the area. Thus, by combination of the information gleaned by both techniques the sequence of pMD4 was confirmed as correct.

3.4.2.7 TnT reactions of plasmids containing mutated 2A sequences in rabbit reticulocyte lysate and wheat germ extract systems.

The plasmids encoding mutations within the 2A region of pMD2 and pCAT2AGUS(3) were transcribed and translated in both coupled wheat germ extract and rabbit reticulocyte lysate systems. The translation products were analysed by PAGE and the results were visualised by phospho-imaging (Figure 3.4.2). Plasmids pMD2 and pCAT2AGUS(3) were also transcribed and translated as controls.

3.4.2.8 TnT reaction of pMD4 in rabbit reticulocyte lysates and immunoprecipitation of the translation products

Maxipreparations of the plasmids pMD4, which encoded a frameshift across the 2A region, and pMD2 were transcribed and translated in both coupled rabbit reticulocyte lysate and wheat germ extract systems. The translation products from the rabbit reticulocyte lysate reaction of pMD4 were immunoprecipitated with both anti-CAT and anti-GUS sera. The products were analysed by 10 % PAGE (Figure 3.4.3).

125
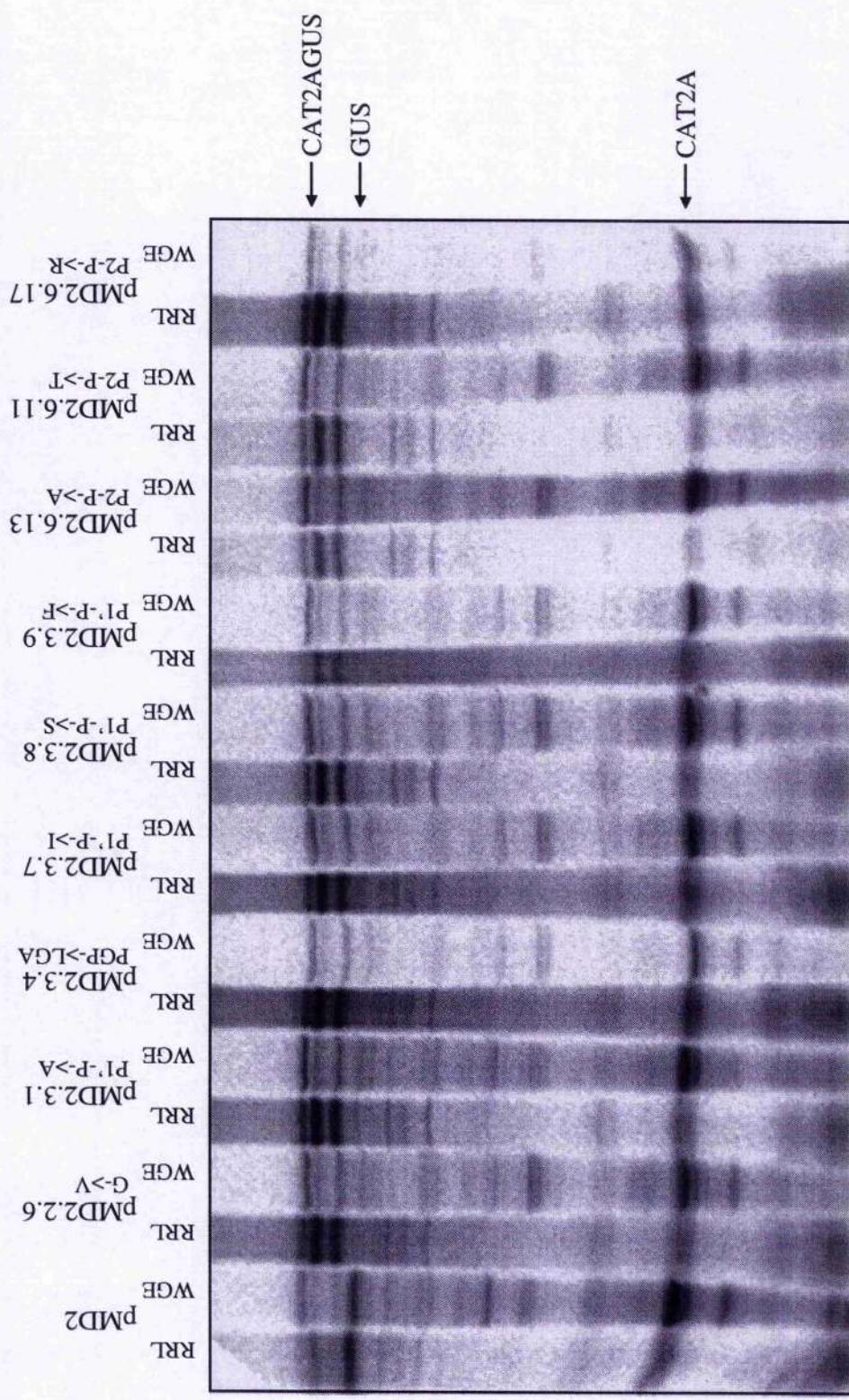
Figure 3.4.2 TnT reactions of constructs containing mutated 2A regions. Reactions in rabbit reticulocyte lysate (RRL) and wheatgerm extract (WGE) for each construct were separated by 10% PAGE.
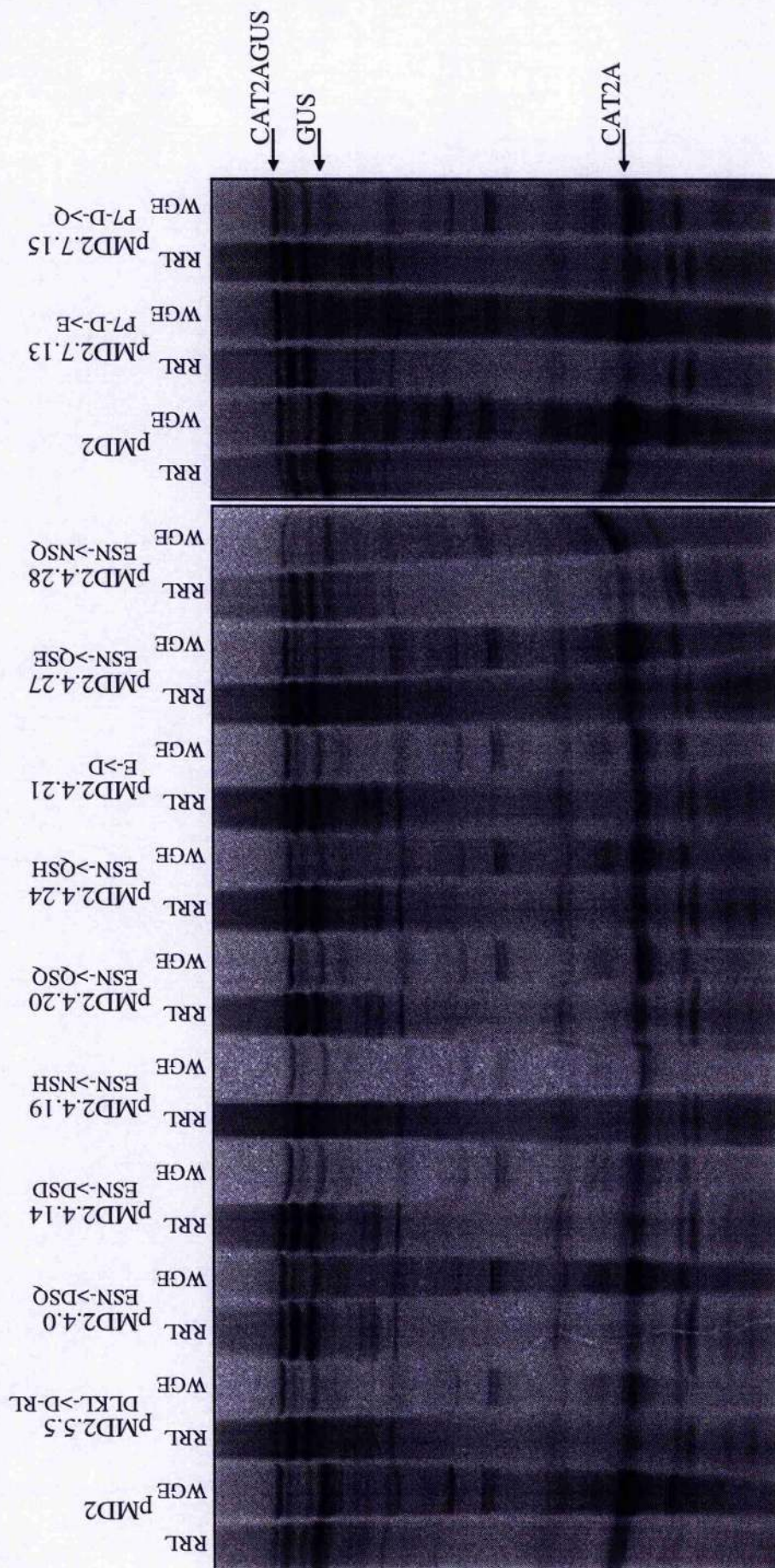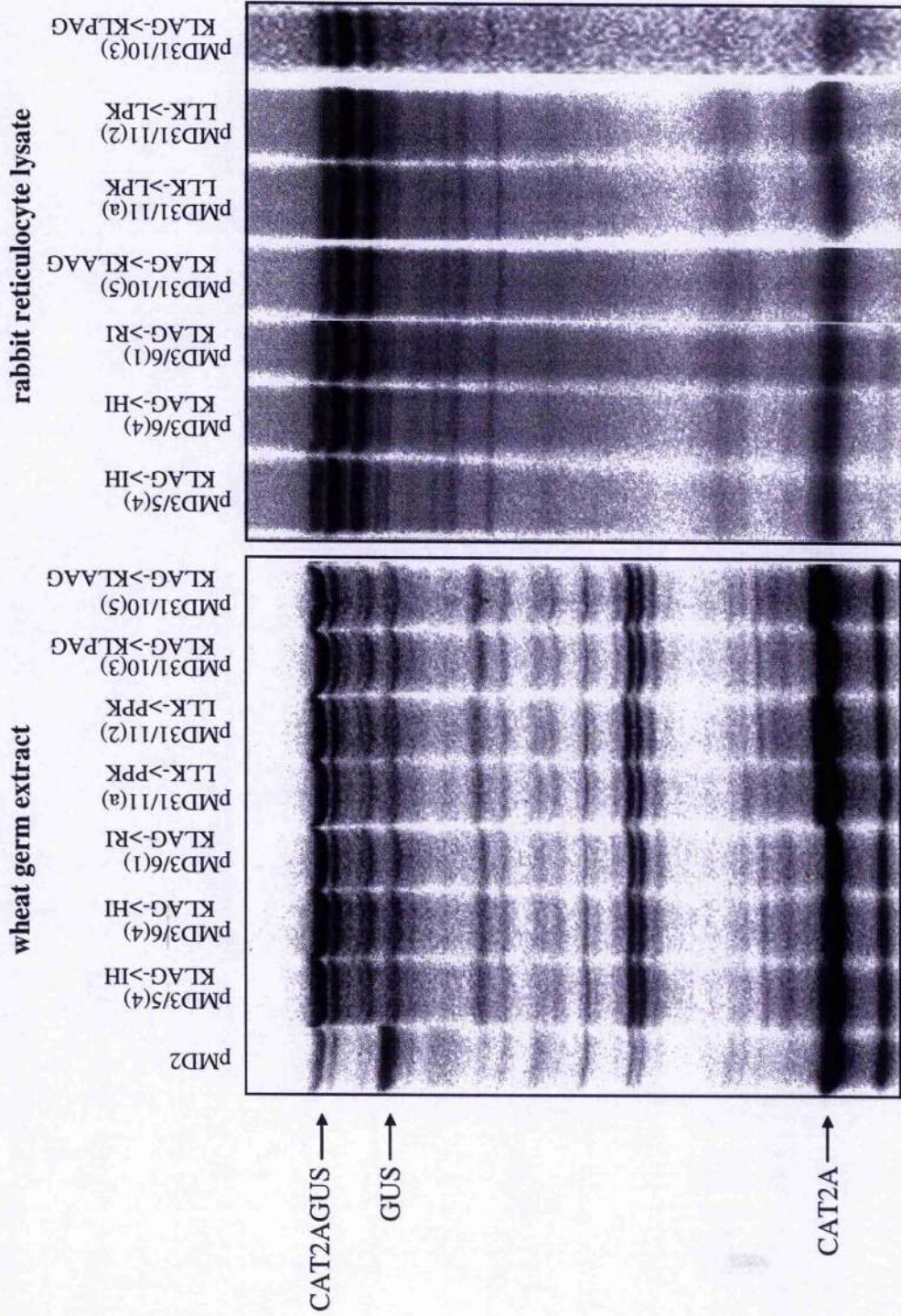
Figure 3.4.2 TnT reactions of constructs containing mutated 2A regions (cont.).

Figure 3.4.2 TnT reactions of constructs containing mutated 2A regions (cont.).

128
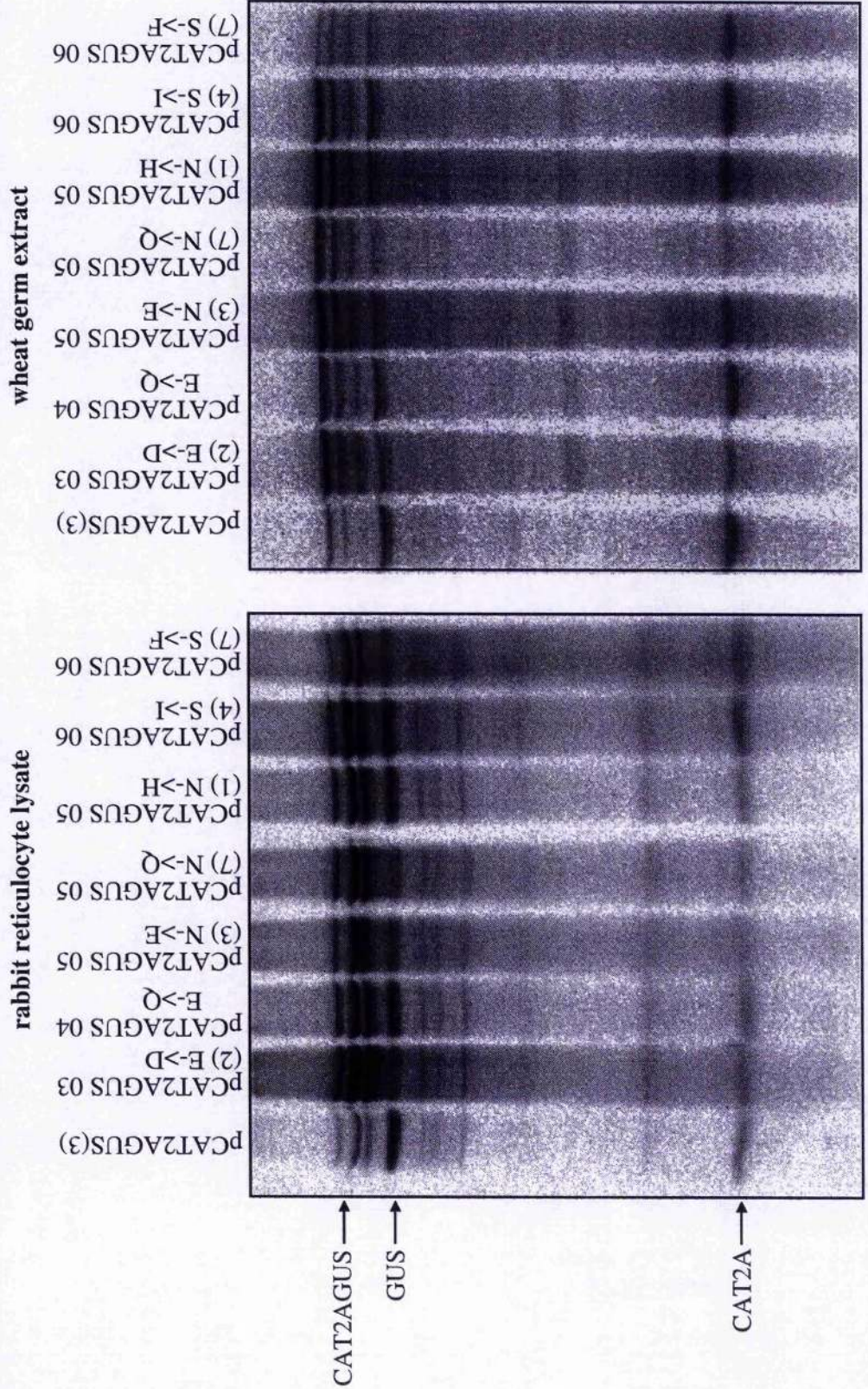
Figure 3.4.2 TnT reactions of constructs containing mutated 2A regions (cont.).
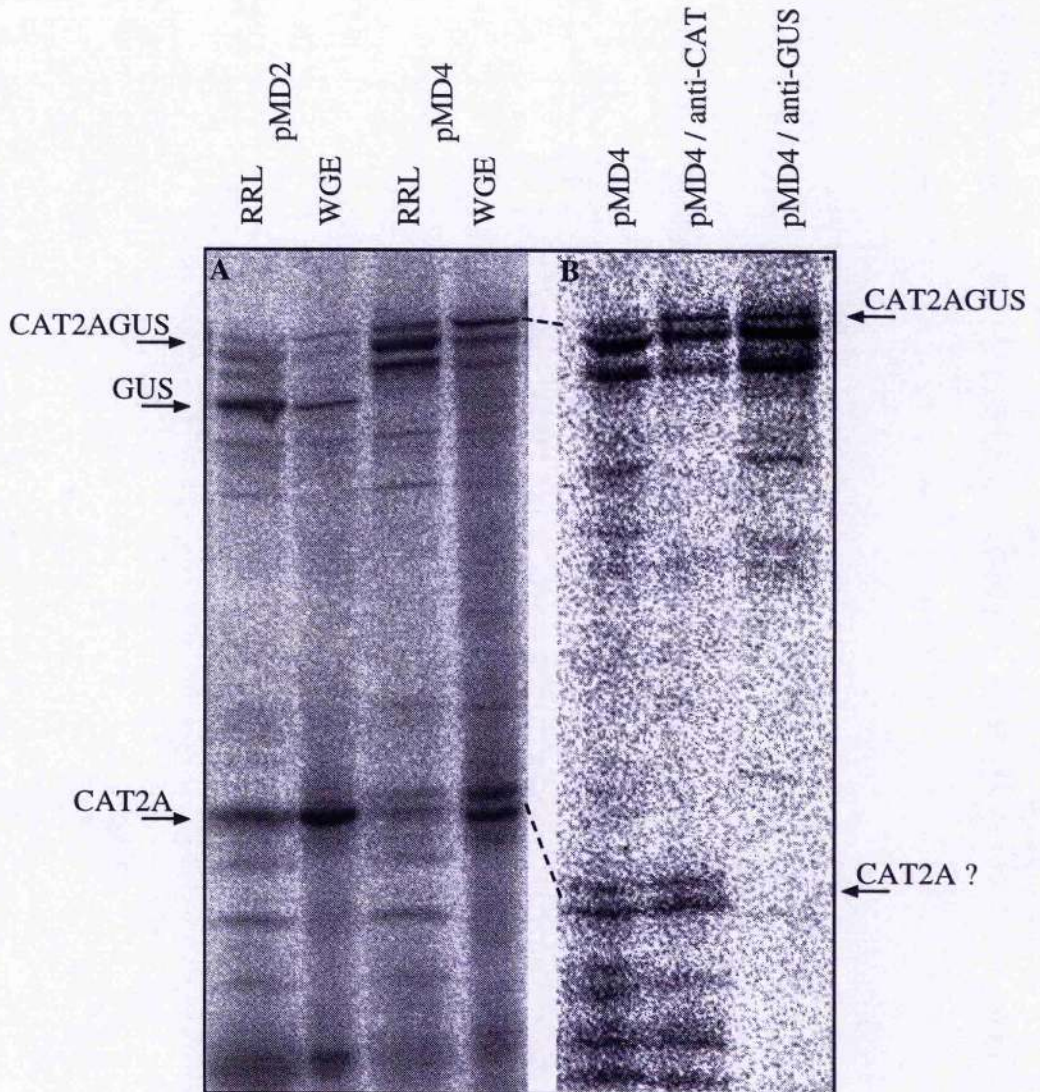
Figure 3.4.3 TnT and immunoprecipitation reactions of pMD4.

A. Plasmids pMD2 and pMD4 were transcribed and translated in both wheat germ extract and rabbit reticulocyte lysate.

B. The rabbit reticulocyte reaction containing pMD4 was immunoprecipitated with anti-CAT and anti-GUS antibodies.

The products of reactions were separated by 10 % PAGE. Dashed lines indicate equivalent points on the two panels.

3.4.2.9 Immunoprecipitation of the translation products of pMD2.3.8 in rabbit reticulocyte lysates

A maxi-preparation of the plasmid pMD2.3.8, which encoded the mutation of P1-proline to serine was transcribed and translated in a coupled rabbit reticulocyte lysate system. The translation products from the reaction were immunoprecipitated with both anti-CAT and anti-GUS sera. The products were analysed by 10 % PAGE (Figure 3.4.4).

3.4.3   Discussion

3.4.3.1   Construction of mutated sequences

Most of the planned mutations were carried out by inserting double stranded oligo-nucleotide adapter molecules between the restriction sites which had been engineered into the 2A region. It was unfortunate that the construction of plasmids *via* this method was begun before the translation of pMD1, which resulted in the discovery of the stop codon within the GUS gene. However, all of these mutated sequences were easily transferred into the new plasmid pMD2 which correctly encoded [CAT2AGUS]. The construction of all of the planned mutations was not achieved, since the coupled transcription and translation reactions of those first synthesised indicated that the completion of all planned syntheses would not be a fruitful course of action for obtaining more information on the mechanism of 2A (see later). A list of all the mutations achieved is seen in Table 3.4.3.

In particular the mutations of aspartic acid to asparagine encoded by pMD1.1 and lysine to either glutamine or arginine have not been achieved. Transformations carried out with the ligation mixture for pMD1.1 appeared to have usual efficiencies, however very few of the plasmids were not linearised in the restriction digest with Aat II, *i.e.* most constructs

Figure 3.4.5 Comparison of mutations in pCAT2AGUS and pMD2. Plasmids pCAT2AGUS 06(7) and pMD6.7 both encoding the mutation S->F, and pCAT2AGUS 31/10(5) and pMD31/10(5) both encoding KLAG->KLAAG were transcribed and translated in wheat germ extract (WGE) and rabbit reticulocyte lysate (RRL).

Figure 3.4.4 Immunoprecipitation of the translation products of pMD2.3.8. Plasmid pMD2.3.8. was transcribed and translated in rabbit reticulocyte lysate and the products immunoprecipitated with anti-CAT and anti-GUS antibodies as indicated.

Table 3.4.3 FMDV Mutant 2A Constructs.

Mutant sequences are listed below. pCAT2AGUS and pMD2 are also listed for comparison. Restriction sites are underlined and mutations from the wild-type 2A sequences are shown in red.

pCAT2AGUS;

Q	L	L	N	F	D	L	K	L	A	G	D	V	E	S	N	P	G	P	M	L	R	P	V	E	T	P	T
CAG	CTG	TTG	AAT	TTT	GAC	CTT	AAG	CTT	GCG	GGA	GAC	GTC	GAG	TCC	AAC	CCT	GGG	CCC	ATG	TTA	CGT	CCT	GTA	GAA	ACC	CCA	ACC
	PvuII						AflII-HindIII				AatII						ApaI										

pMD2;

Q	L	L	N	F	D	L	K	L	A	G	D	V	E	S	N	P	G	P	F	F	T	S	R	S	M	L
CAG	CTG	TTG	AAT	TTT	GAC	CTT	AAG	CTT	GCG	GGA	GAC	GTC	GAG	TCC	AAC	CCT	GGG	CCC	TTT	TTT	ACT	AGT	AGA	TCT	ATG	TTA
	PvuII						AflII-HindIII				AatII						ApaI				SpeI		BglII			

P1'-Proline Mutants;

pMD2.3.1;

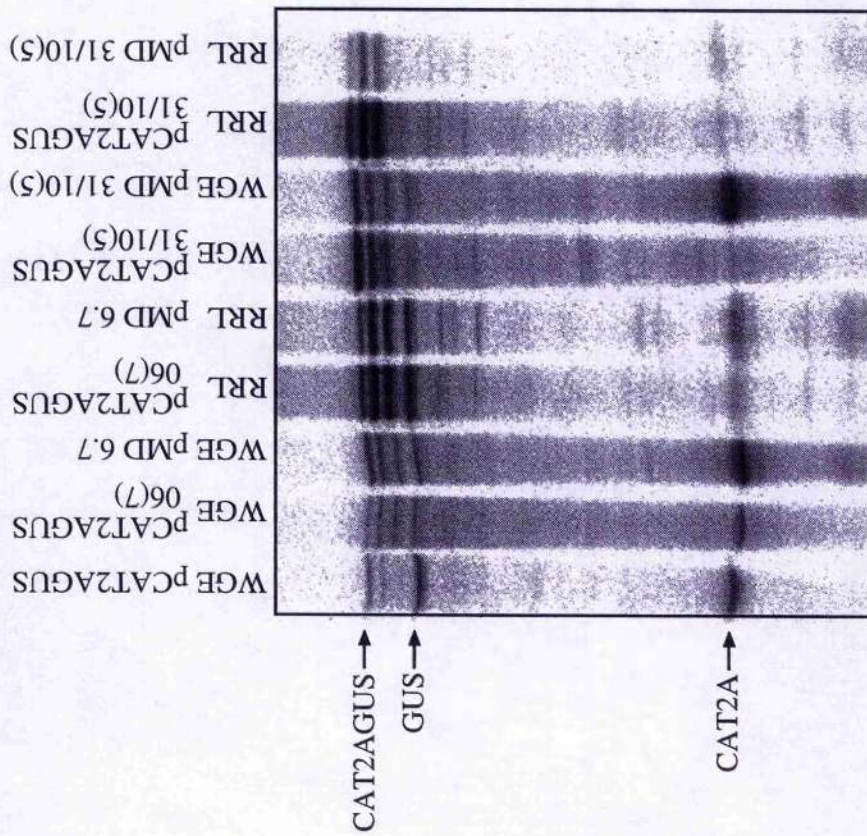| Q	L	L	N	F	D	L	K	L	A	G	D	V	E	S	N	P	G	A	F	F	T	S	R	S	M	L
CAG	CTG	TTG	AAT	TTT	GAC	CTT	AAG	CTT	GCG	GGA	GAC	GTC	GAG	TCC	AAC	CCT	GGG	GCT	TTT	TTT	ACT	AGT	AGA	TCT	ATG	TTA

pMD2.3.7;

Q	L	L	N	F	D	L	K	L	A	G	D	V	E	S	N	P	G	I	F	F	T	S	R	S	M	L
CAG	CTG	TTG	AAT	TTT	GAC	CTT	AAG	CTT	GCG	GGA	GAC	GTC	GAG	TCC	AAC	CCT	GGG	ATT	TTT	TTT	ACT	AGT	AGA	TCT	ATG	TTA

pMD2.3.8;

Q	L	L	N	F	D	L	K	L	A	G	D	V	E	S	N	P	G	S	F	F	T	S	R	S	M	L
CAG	CTG	TTG	AAT	TTT	GAC	CTT	AAG	CTT	GCG	GGA	GAC	GTC	GAG	TCC	AAC	CCT	GGG	AGT	TTT	TTT	ACT	AGT	AGA	TCT	ATG	TTA

pMD2.3.9;

Q	L	L	N	F	D	L	K	L	A	G	D	V	E	S	N	P	G	F	F	F	T	S	R	S	M	L
CAG	CTG	TTG	AAT	TTT	GAC	CTT	AAG	CTT	GCG	GGA	GAC	GTC	GAG	TCC	AAC	CCT	GGG	TTT	TTT	TTT	ACT	AGT	AGA	TCT	ATG	TTA

pMD2.3.12;

Q	L	L	N	F	D	L	K	L	A	G	D	V	E	S	N	P	G	I	F	F	T	S	R	S	M	L
CAG	CTG	TTG	AAT	TTT	GAC	CTT	AAG	CTT	GCG	GGA	GAC	GTC	GAG	TCC	AAC	CCT	GGG	ATT	TTT	TTT	ACT	AGT	AGA	TCT	ATG	TTA

P1-Glycine Mutant;

pMD2.2.6;

Q	L	L	N	F	D	L	K	L	A	G	D	V	E	S	N	P	V	P	F	F	T	S	R	S	M	L
CAG	CTG	TTG	AAT	TTT	GAC	CTT	AAG	CTT	GCG	GGA	GAC	GTC	GAG	TCC	AAC	CCT	GTG	CCC	TTT	TTT	ACT	AGT	AGA	TCT	ATG	TTA

P1', P2-Proline Mutant;

pMD2.3.4;

Q	L	L	N	F	D	L	K	L	A	G	D	V	E	S	N	L	G	A	F	F	T	S	R	S	M	L
CAG	CTG	TTG	AAT	TTT	GAC	CTT	AAG	CTT	GCG	GGA	GAC	GTC	GAG	TCC	AAC	CTT	GGG	GCT	TTT	TTT	ACT	AGT	AGA	TCT	ATG	TTA

133

P2-Proline Mutants;

pMD 2.6.11;

```
Q   L   N   F   D   L   L   K   L   A   G   D   V   E   S   N   A   G   P   F   F   F   T   S   R   S   M   L
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC GAG TCC AAC GCT GGG CCC TTT TTT ACT AGT AGA TCT ATG TTA
```

pMD 2.6.13;

```
Q   L   N   F   D   L   L   K   L   A   G   D   V   E   S   N   T   G   P   F   F   F   T   S   R   S   M   L
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC GAG TCC AAC ACG GGG CCC TTT TTT ACT AGT AGA TCT ATG TTA
```

pMD 2.6.17;

```
Q   L   N   F   D   L   L   K   L   A   G   D   V   E   S   N   R   G   P   F   F   F   T   S   R   S   M   L
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC GAG TCC AAC AGA GGG CCC TTT TTT ACT AGT AGA TCT ATG TTA
```

P3-Aspartate Mutants;

pMD 2.7.13;

```
Q   L   N   F   D   L   L   K   L   A   G   E   V   E   S   N   P   G   P   F   F   F   T   S   R   S   M   L
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAG GTC GAG TCC AAC CCT GGG CCC TTT TTT ACT AGT AGA TCT ATG TTA
```

pMD 2.7.15;

```
Q   L   N   F   D   L   L   K   L   A   G   Q   V   E   S   N   P   G   P   F   F   F   T   S   R   S   M   L
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA CAG GTC GAG TCC AAC CCT GGG CCC TTT TTT ACT AGT AGA TCT ATG TTA
```

P4-Serine Mutants;

pCAT2AGUS / 06; (4)

```
Q   L   L   N   F   D   L   L   K   L   A   G   D   V   E   I   N   P   G   M   L   R   P   V   E   T   P   T
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC GAG ATT AAC CCT GGG ATG TTA CGT CCT GTA GAA ACC CCA ACC
```

pCAT2AGUS / 06; (7)

```
Q   L   L   N   F   D   L   L   K   L   A   G   D   V   E   F   N   P   G   M   L   R   P   V   E   T   P   T
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC GAG TTT AAC CCC GGG ATG TTA CGT CCT GTA GAA ACC CCA ACC
```

pMD 6.7 ;

```
Q   L   L   N   F   D   L   L   K   L   A   G   D   V   E   F   N   P   G   P   F   F   F   T   S   R   S   M   L
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC GAG TTT AAC CCC GGG CCC TTT TTT ACT AGT AGA TCT ATG TTA
```

P5-Glutamate Mutants;

pMD 2.4.21;

```
Q   L   L   N   F   D   L   L   K   L   A   G   D   V   D   S   N   P   G   P   F   F   F   T   S   R   S   M   L
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC GAC TCC AAC CCT GGG CCC TTT TTT ACT AGT AGA TCT ATG TTA
```

pCAT2AGUS /03; (2)

```
Q   L   L   N   F   D   L   L   K   L   A   G   D   V   D   S   N   P   G   P   M   L   R   P   V   E   T   P   S
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC GAC TCC AAC CCC GGG CCC ATG TTA CGT CCT GTA GAA ACC CCA ACC
```

134

P5-Glutamate Mutants (cont.);

pCAT2AGUS / 04;
```
Q   L   N   F   D   L   L   K   L   A   G   D   V   Q   S   N   P   G   P   M   L   R   P   V   E   T   P   S
CAG CTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC CAG TCC AAC CCC GGG CCC ATG TTA CGT CCT GTA GAA ACC CCA ACC
```

P5-Glutamate / P3-Asparagine Double Mutants;

pMD 2.4.0;
```
Q   L   L   N   F   D   L   L   K   L   A   G   D   V   D   S   Q   P   G   P   F   F   T   S   R   M   L
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC GAC TCC CAG CCT GGG CCC TTT TTT ACT AGT AGA ATG TTA
```

pMD 2.4.14;
```
Q   L   L   N   F   D   L   L   K   L   A   G   D   V   D   S   D   P   G   P   F   F   T   S   R   M   L
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC GAC TCC GAC CCT GGG CCC TTT TTT ACT AGT AGA ATG TTA
```

pMD 2.4.20;
```
Q   L   L   N   F   D   L   L   K   L   A   G   D   V   D   S   Q   P   G   P   F   F   T   S   R   M   L
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC GAC TCC CAG CCT GGG CCC TTT TTT ACT AGT AGA ATG TTA
```

pMD 2.4.24;
```
Q   L   L   N   F   D   L   L   K   L   A   G   D   V   Q   S   Q   P   G   P   F   F   T   S   R   M   L
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC CAG TCC CAG CCT GGG CCC TTT TTT ACT AGT AGA ATG TTA
```

pMD 2.4.27;
```
Q   L   L   N   F   D   L   L   K   L   A   G   D   V   Q   S   E   P   G   P   F   F   T   S   R   M   F
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC CAG TCC GAG CCT GGG CCC TTT TTT ACT AGT AGA ATG TTA
```

pMD 2.4.19;
```
Q   L   L   N   F   D   L   L   K   L   A   G   D   V   N   S   H   P   G   P   F   F   T   S   R   M   L
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC AAC TCC CAC CCT GGG CCC TTT TTT ACT AGT AGA ATG TTA
```

pMD 2.4.28;
```
Q   L   L   N   F   D   L   L   K   L   A   G   D   V   N   S   Q   P   G   P   F   F   T   S   R   M   L
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC AAC TCC CAG CCT GGG CCC TTT TTT ACT AGT AGA ATG TTA
```

P3-Asparagine Mutants;

pCAT2AGUS/05; (1)
```
Q   L   L   N   F   D   L   L   K   L   A   G   D   V   E   S   H   P   G   P   M   L   R   P   V   E   T   P   T
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC GAG TCC CAC CCC GGG CCC ATG TTA CGT CCT GTA GAA ACC CCA ACC
```

pCAT2AGUS/05; (3)
```
Q   L   L   N   F   D   L   L   K   L   A   G   D   V   E   S   E   P   G   P   M   L   R   P   V   E   T   P   T
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC GAG TCC GAG CCC GGG CCC ATG TTA CGT CCT GTA GAA ACC CCA ACC
```

pCAT2AGUS/05; (7)
```
Q   L   L   N   F   D   L   L   K   L   A   G   D   V   E   S   Q   P   G   P   M   L   R   P   V   E   T   P   T
CAG CTG TTG AAT TTT GAC CTT CTT AAG CTT GCG GGA GAC GTC GAG TCC CAG CCC GGG CCC ATG TTA CGT CCT GTA GAA ACC CCA ACC
```

Δ Leucine / P11-Lysine Mutant;

pMD 2.5.5;

```
Q    L    L    N    F    D    L    -    R    L    A    G    D    V    E    S    N    P    G    P    F    F    T    S    R    S    M    L
CAG  CTG  TTG  AAT  TTT  GAC  CTC  ---  CGA  CTT  GCG  GGA  GAC  GTC  GAG  TCC  AAC  CCT  GGG  CCC  TTT  TTT  ACT  AGT  AGA  TCT  ATG  TTA
```

Δ(-KLAG-), Insertion (-IH- etc.);

pMD 3 / 5;

```
Q    L    L    N    F    D    L    -    I    H    -    D    V    E    S    N    P    G    P    F    F    T    S    R    S    M    L
CAG  CTG  TTG  AAT  TTT  GAC  CTT  ---  ATC  CAC  ---  GAC  GTC  GAG  TCC  AAC  CCT  GGG  CCC  TTT  TTT  ACT  AGT  AGA  TCT  ATG  TTA
```

pMD 3 / 6; (c)

```
Q    L    L    N    F    D    L    -    H    I    -    D    V    E    S    N    P    G    P    F    F    T    S    R    S    M    L
CAG  CTG  TTG  AAT  TTT  GAC  CTT  ---  CAC  ATC  ---  GAC  GTC  GAG  TCC  AAC  CCT  GGG  CCC  TTT  TTT  ACT  AGT  AGA  TCT  ATG  TTA
```

pMD 3 / 6; (i)

```
Q    L    L    N    F    D    L    -    R    I    -    D    V    E    S    N    P    G    P    F    F    T    S    R    S    M    L
CAG  CTG  TTG  AAT  TTT  GAC  CTT  ---  CGC  ATC  ---  GAC  GTC  GAG  TCC  AAC  CCT  GGG  CCC  TTT  TTT  ACT  AGT  AGA  TCT  ATG  TTA
```

Insertion Mutants;

pMD 31/10(3);

```
Q    L    L    N    F    D    L    L    K    L    P    A    G    D    V    E    S    N    P    G    P    F    F    T    S    R    S    M    L
CAG  CTG  TTG  AAT  TTT  GAC  CTT  CTT  AAG  CTT  CCC  GCG  GGA  GAC  GTC  GAG  TCC  AAC  CCT  GGG  CCC  TTT  TTT  ACT  AGT  AGA  TCT  ATG  TTA
```

pMD 31/10(5);

```
Q    L    L    N    F    D    L    L    K    L    A    A    G    D    V    E    S    N    P    G    P    F    F    T    S    R    S    M    L
CAG  CTG  TTG  AAT  TTT  GAC  CTT  CTT  AAG  CTT  GCC  GCG  GGA  GAC  GTC  GAG  TCC  AAC  CCT  GGG  CCC  TTT  TTT  ACT  AGT  AGA  TCT  ATG  TTA
```

pCAT2AGUS; 31/10(5)

```
Q    L    L    N    F    D    L    P    L    K    L    A    A    G    D    V    E    S    N    P    G    M    R    P    V    E    T    P    T
CAG  CTG  TTG  AAT  TTT  GAC  CTT  CCC  CTT  AAG  CTT  GCC  GCG  GGA  GAC  GTC  GAG  TCC  AAC  CCT  GGG  ATG  CGT  CCT  GTA  GAA  ACC  CCA  ACC
```

pMD 3/11(a);

```
Q    L    L    N    F    D    L    P    L    K    L    A    G    D    V    E    S    N    P    G    P    F    F    T    S    R    S    M    L
CAG  CTG  TTG  AAT  TTT  GAC  CTT  CCC  CTT  AAG  CTT  GCG  GGA  GAC  GTC  GAG  TCC  AAC  CCT  GGG  CCC  TTT  TTT  ACT  AGT  AGA  TCT  ATG  TTA
```

pMD 31/11(2);

```
Q    L    L    N    F    D    P    P    L    K    L    A    G    D    V    E    S    N    P    G    P    F    F    T    S    R    S    M    L
CAG  CTG  TTG  AAT  TTT  GAC  CCT  CCC  CTT  AAG  CTT  GCG  GGA  GAC  GTC  GAG  TCC  AAC  CCT  GGG  CCC  TTT  TTT  ACT  AGT  AGA  TCT  ATG  TTA
```

136

isolated were merely pMD1. Those constructs appearing to have no Aat II site by restriction digest were sequenced, but all of them had the natural 2A sequence. This highlighted a problem with my strategy of inserting oligonucleotide pairs into a sequence which already contained multiple restriction sites. The maximum number of restriction sites possible within the natural amino acid sequence of 2A was already present. Thus it was only possible to destroy, not create, restriction sites for screening within the mutant plasmids, pMD1.1-7, without altering the amino acid sequence further. The results of such experiments were not conclusive. The inability of an enzyme to linearise the DNA might have been due to too much salt in the DNA sample, inactive or insufficient enzyme or insufficient incubation time as well as the absence of a restriction site.

The residues for mutation in both pMD1.7 and pMD1.5 mutant plasmids, were encoded by codons within a restriction site. The protruding termini produced by restriction digests at these sites had to be removed for the ligation of the oligonucleotides. The 3' to 5' exonuclease activity of T4 DNA polymerase was used successfully for the removal of nucleotides from the 3' overhang of the Aat II site in making pMD1.7. However, to make the desired mutations encoded by pMD1.5, a 5' overhang required removal. The only available enzymes with 5' exonuclease activity are Mung bean nuclease and S1 nuclease. Incubation of the DNA with S1 nuclease for 10 minutes completely shredded the DNA. Incubation with Mung bean nuclease, which was known to be milder, for periods as short as ten minutes still resulted in degradation of the double stranded DNA. Hence only mutations about lysine which also encode a deletion or frame shift have been found.

3.2.3.2 The phenotypes of constructs containing mutated 2A regions

The translation products of all of the 2A mutation-containing constructs derived from pMD2 appeared to be similar (Figure 3.4.2). The lanes resulting from reactions in rabbit reticulocyte lysate all displayed a large quantity of uncleaved [CAT2AGUS] and a diffuse

band co-migrating with CAT2A. Very faint bands co-migrating with GUS were seen in some cases. In the majority of lanes this band migrated slightly further than the genuine GUS cleavage product and was attributed to internal initiation at the methionine start codon for GUS. Internal initiation was more pronounced in the wheat germ system. In wheat germ extract there was also some uncleaved material for each of the mutant polyproteins. However, there appeared to be more of the CAT2A cleavage products in the wheat germ system. The band co-migrating with GUS was again much fainter than that for CAT2A. In all of these cases "2A-mediated proteolytic cleavage" as such does not appear to be occurring. Rather an interruption in translation after the 2A region seemed to have taken place. Plasmid pMD2.4.28 which encodes the mutations E->N and N->Q shows notably more cleavage products than the other mutants. Plasmids pMD2.3.8 and pMD2.3.7 which encode a P1' mutation of P->S and I, respectively, also appeared to be cleaved to CAT2A and GUS to a small extent. Immunoprecipitation of pMD2.3.8 with anti-CAT and anti-GUS polyclonal antibodies confirmed the product migrating with CAT2A to be derived from the CAT protein and that migrating with GUS to be GUS specific (Figure 3.3.4).

The translation products of all the 2A mutant constructs derived from pCAT2AGUS are more consistent with what one would expect from a proteolysis reaction. The efficiency of 2A cleavage is reduced in all of the mutated forms, however all mutants produce both CAT2A and GUS to some extent. The most efficient cleaving mutant 2A regions are those with E->Q, N->H , S->I and S->F mutations. The polyproteins containing the mutations E->D and N>E, and to a lesser extent the mutation of N->Q, produced small quantities of CAT2A and GUS proteins. The distinctive feature of these plasmids compared to those derived from pMD2 is that the GUS gene commences immediately after the first proline residue of FMDV 2B (see Table 3.4.3). This might suggest that the sequence immediately following 2A is important for translation of the second cleavage product.

The mutations encoding S->F and KLAG->KLAAG were placed in the context of both pMD2 and pCAT2AGUS(3) to allow a direct comparison of their activities. It was hoped

that this might clarify the differences in the activities of 2A in the two systems. The translation products of these plasmids from both rabbit reticulocyte lysate and wheat germ extract reactions are seen in Figure 3.4.5. In both systems the 2A region containing mutation of serine to phenylalanine cleaves well. However, there is a marked increase in the ratio of CAT2A : GUS for pMD6.7 especially in the wheat germ extract system. The insertion of an alanine residue to give the motif KLAAG inactivates the proteolytic activity of the 2A region. Once again however there was some CAT2A protein produced in the reactions programmed with pMD31/10(5). This substantiated the hypothesis that the sequence following 2A was important for the translation of the second cleavage product.

The products of transcription and translation of pMD4, encoding [CAT-'2A'-GUS] with a +1 frameshift over the 2A region, were somewhat unanticipated (Figure 3.4.3); if 2A were mediating a proteolytic cleavage dependent upon its primary sequence then the product of this reaction would surely be an uncleaved [CAT-'2A'-GUS] polyprotein. In both rabbit reticulocyte and wheat germ systems there were bands, as expected, co-migrating with uncleaved [CAT2AGUS] polyprotein, and surprisingly a doublet of bands migrating similarly to CAT2A. Immunoprecipitation reactions using anti-CAT and anti-GUS polyclonal antibodies identified the doublet of bands as CAT specific. Thus, it appeared that no 2A-mediated cleavage had occurred. Instead an interruption in translation seemed to occur in a small percentage of cases, although it produced two products of the approximate size of CAT2A rather than one. One could postulate, therefore, that it was the RNA sequence of 2A which was important for the arrest of translation.

3.4.4   Conclusions

• All of the [CAT2AGUS] polyproteins synthesised with mutated 2A regions displayed similar profiles. All had large amounts of uncleaved [CAT2AGUS] polyprotein, relatively large quantities for CAT2A and low levels of GUS. Thus the CAT2A : GUS ratio was very large, and much greater than for the wild type sequence. This is in contrast with a proteolytic cleavage hypothesis.

• Translation of CAT'2A'GUS, in which the 2A region is in the +1 reading frame, appeared to result in two CAT specific products which were of the approximate size of CAT2A. This indicated that only the RNA sequence of 2A and / or the short stretch of 2B linker sequence encoding PFFF was important for the interruption in translation.

Section 3.5 The effect of N-terminally extending the FMDV 2A region

3.5.1 Introduction

Previous analyses of recombinant FMDV polyproteins demonstrated that deletion of the region 3' to 2A does not affect 2A cleavage activity (Ryan *et al.*, 1991). However the presence of the upstream context of 2A was been found to increase the efficiency of 2A activity, although it was not necessary for cleavage (Ryan *et al.*, 1991). It was planned to extend the FMDV sequence, within the constructs encoding [CAT2AGUS] polyproteins, to include parts of the FMDV capsid protein 1D region as well as 2A itself. It was thus hoped to gain an insight into the degree of upstream context required to return 2A cleavage activity towards 100 %. This section describes the syntheses and translation of three cDNA constructs; pTG393, pTG394 and pTG395 each encoding CATΔ1D2AGUS polyproteins in which the truncated C-terminal 1D regions were of 180, 39 and 5 amino acids respectively.

3.5.2 Results

3.5.2.1 Construction of pTG393

Sequences encoding the 180 C-terminal residues of FMDV 1D were amplified from plasmid pMR31 (encoding FMDV P1P2ΔP3) using the oligonucleotides OR393 and OR48 (Ryan *et al.*, 1989), as forward and reverse primers, respectively. OR393 encoded an Xba I site followed by some sequence of 1D beginning at its 33rd codon. OR48 was designed to anneal to the sequence encoding the 2B region. The PCR product, encoding Δ1D2AΔ2B, was doubly restricted with Xba I and Apa I, and isolated by agarose gel electrophoresis. This was ligated into the large fragment of the similarly restricted plasmid pMD2 and transformed into *E. coli* for mini-preparations. The resultant cDNA constructs

were screened for the unique Sma I site, present within the 1D region, then pTG393 was sequenced and maxi-prepared *via* the Qiagen method.

```
Protein sequence encoded:        S   R   S   F   I   M   D   R
OR393:              5'- TTTTTTTCTAGATCGTTCATCATGGACAGA -3'
                             Xba I
```

```
Protein sequence encoded      A   I   I   W   P   H   A   D
complementary sequence:   GCCATCATCTGGCCGCACGCCGATAC
OR48:                  3'- CGGTAGTAGACCGGCGTGCGGCTATAG -5'
```
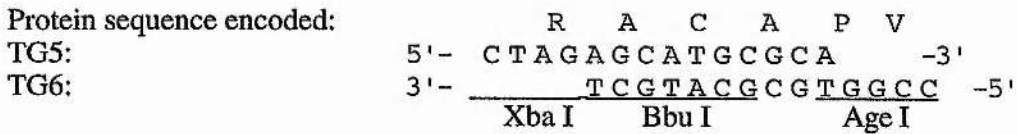
3.5.2.2   Construction of pTG394

Sequences encoding the 39 C-terminal residues of FMDV 1D were amplified from plasmid pMR31 (see above) using oligonucleotides OR394 and OR48 as forward and reverse primers, respectively. OR394 encoded an Xba I site followed by the sequence of 1D from the 173rd codon. The PCR product was doubly restricted with Xba I and Apa I and isolated by agarose gel electrophoresis. It was then ligated into the large fragment of the similarly restricted plasmid pMD2 and transformed into *E. coli* for mini-preparations. The resultant cDNA constructs were screened for the presence of an Age I restriction site, maxi-prepared *via* the Qiagen method and sequenced.

```
Protein sequence encoded:        S   R   V   T   E   L   L   Y
OR394:              5 '- TTTTTTTCTAGAGTCACCGAGTTGCTTTAC -3'
```

3.5.2.3   Construction of pTG395

Plasmid pTG394 was doubly digested with Xba I and Age I, and the 5517 base pair fragment, from which most of the 1D encoding sequence had been excised, was purified by agarose gel isolation. A double stranded oligonucleotide adapter prepared by annealing the two oligonucleotides TG5 and TG6 was then ligated into this fragment and the resultant cDNA constructs transformed into *E. coli* for mini-preparations. The mini-preparations

142

were screened for the additional Bbu I restriction site (there was already one present at the terminus of the GUS gene), then maxipreped using the Qiagen method before sequencing.

```
Protein sequence encoded:        R   A   C   A   P   V
TG5:                    5'- C T A G A G C A T G C G C A      -3'
TG6:                    3'- _____T C G T A C G C G T G G C C  -5'
                            Xba I      Bbu I        Age I
```

3.5.2.4   TnT reactions of plasmids encoding CATΔ1D2AGUS polyproteins

Maxi-preparations of plasmid DNA, pMD2, pTG393, pTG394 and pTG395 were transcribed and translated in coupled rabbit reticuloctye lysate and wheat germ extract systems. The products were analysed by PAGE (Figure 3.5.1).

3.5.3   Discussion

The plasmids synthesised for use in this study, pTG393, pTG394, and pTG395 are shown in Figure 3.5.2. It should be noted that the sequence encoding Δ1D2A regions of FMDV were the wild type sequences. pMD2 DNA was mutated from the wild type FMDV sequence for 2A in order to introduce restriction sites. The translation profile derived from pMD2 was compared to those from pTG393, pTG394, and pTG395, which have progressively shorter 1D sections encoded before the 2A region. The percentage of cleavage for each polyprotein in both rabbit reticulocyte lysate and wheat germ extract was calculated exactly as described in the Section 3.2. Since unequal amounts of CAT2A and GUS have been formed by the cleavage of such polypeptides thus far, the percentage of cleavage stated here refers to the quantity of CAT2A found, compared to that of the uncleaved [CAT2AGUS].
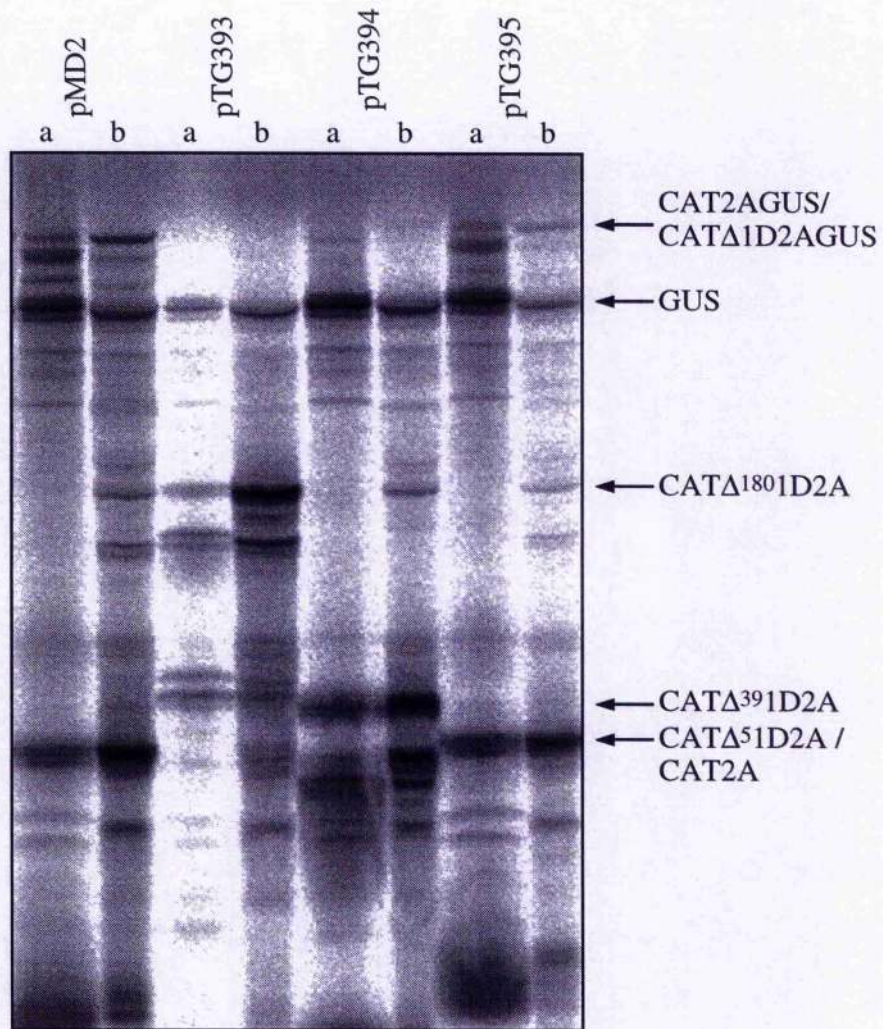
Figure 3.5.1 Translation *in vitro* of constructs encoding CATΔ1D2AGUS.
Coupled transcription/translation reactions in both rabbit reticulocyte lysate, lanes labelled 'a', and wheat germ extract, lanes labelled 'b', were programmed with the plasmid DNA as indicated. Translation products were analysed by PAGE. The translation products are indicated, with the number of residues of 1D present shown by the superscript number in where appropriate.
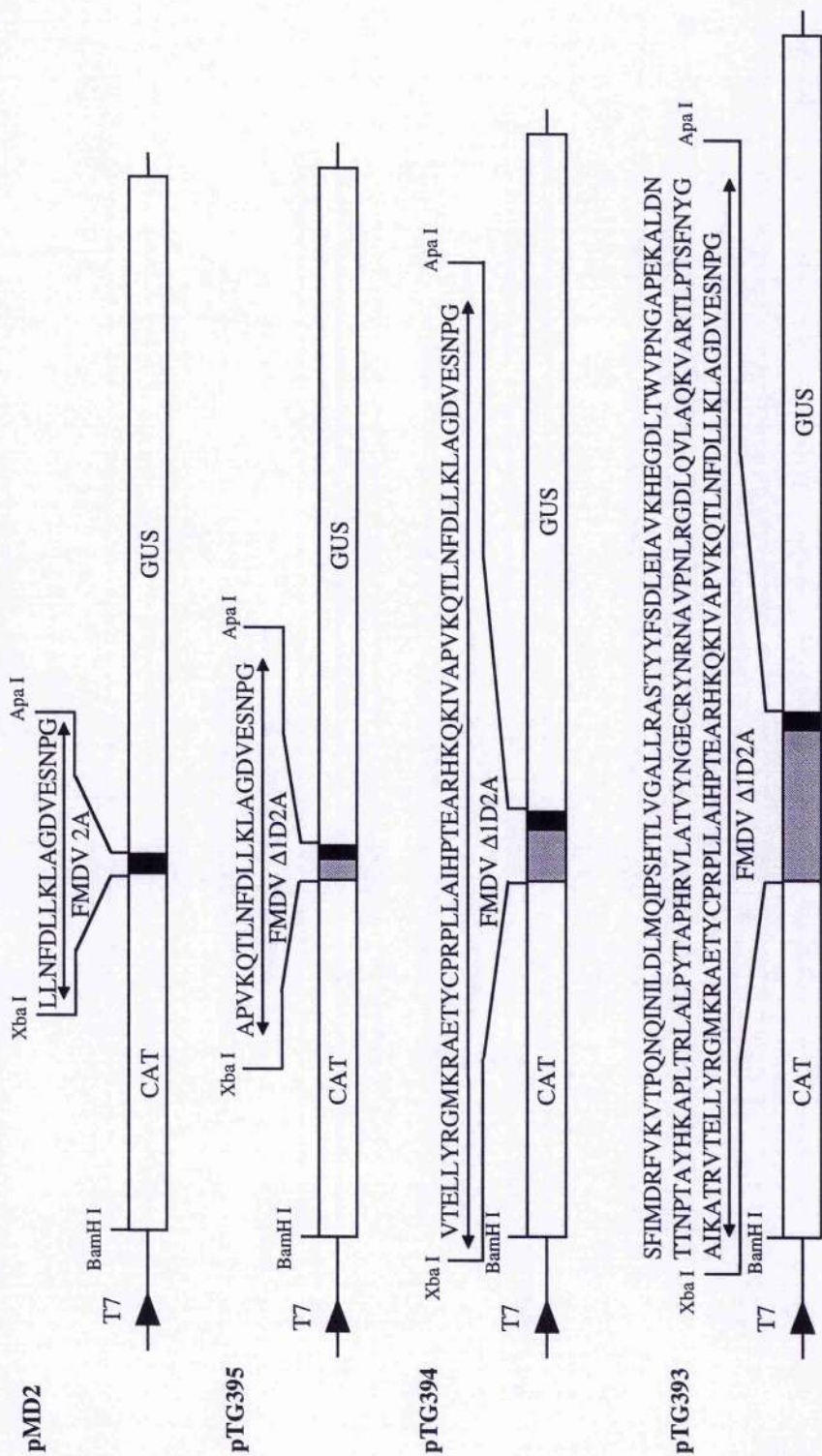
Figure 3.5.2 Plasmid constructs. The boxed areas indicate the single open reading frames encoding the artificial polyproteins. Solid black areas represent the 2A region of FMDV. ▨ Represents the truncated 1D regions. The protein sequence encoded between the Xba I and Apa I sites is listed above the area. N.B. The GUS gene sequence is preceeded by seqences encoding PFFF of the FMDV 2B region and RS from a Bgl II site.

145

For all translations large quantities of internal initiation products were produced. In the majority of cases these products resulted from internal initiation within CAT2A as explained in the Section 3.2. Only products resulting from full length translation products are indicated in Figure 3.5.1.

The results of densitometric analysis of the cleavage products are shown in Table 3.5.1. In rabbit reticulocyte lysates the control pMD2 encoding [CAT2AGUS] was translated to give a product which cleaved in 96 % of cases to give CAT2A, with a CAT2A : GUS ratio of 4.6 : 1. In the wheat germ system 98 % of cleavage is seen with a CAT2A : GUS ratio of 14.1 : 1. The translation profiles, for both *in vitro* systems, derived from pTG393, encoding the CAT gene, almost the entire FMDV capsid protein 1D, FMDV 2A, and GUS showed complete cleavage, with no detectable full length polyprotein. pTG394, encoding CAT the 39 C-terminal residues of 1D, 2A, and GUS, cleaved completely such that no uncleaved polyprotein was visible in the wheat germ system. In the rabbit reticulocyte lysate a small amount of uncleaved material was just visible, but the level of cleavage was still found to be greater than 99 %. The translations of pTG395, which encodes CAT, 5 residues of 1D, 2A, and GUS, have profiles comparable to those for pMD2. Although a comparison of the lanes of pMD2 with those of pTG395, by eye, finds less uncleaved polyprotein in the case of pTG395, the densitometric analysis of the products of pTG395 showed that the level of cleavage was not significantly different from that of the products of pMD2. In the rabbit reticulocyte lysate system the full length and truncated forms (from internal initiation) of this [CATΔ1D2AGUS] polyprotein were detected, and the percentage of cleavage was measured to be 96%. In the wheat germ extract system only the full length polyprotein was detectable. By comparison with the profile for pMD2 this was estimated to result from 73 % initiation at the first methionine within this batch of wheat germ extract. Extrapolation of this data enabled an estimation of 98 % to be made for the efficiency of cleavage. The ratio of CAT2A to GUS in the rabbit reticulocyte system was found to be fairly constant, within experimental error, for pMD2 and the three constructs encoding N-terminal extensions to 2A, with a mean value of 4.6 :1. The ratios found for

Construct	Rabbit reticulocyte lysate reactions			Wheatgerm extract reactions		
	Ratio CAT2A/GUS	% Cleavage based on CAT2A or GUS		Ratio CAT2A/GUS	% Cleavage based on CAT2A or GUS	
		CAT2A			CAT2A	
pMD2	4.6	96		14.1	98	
pTG393	4.8	100		20.5	100	
pTG394	5.1	>99		12.1	100	
pTG395	4.7	96		20.8	99	

Table 3.5.1 Results of the densitometric analysis of the translation products seen in Figure 3.5.1. The densitometric analysis was carried out exactly as described in Results part 2. The percentage of internal initiation could not be measured directly for pTG393 and pTG394 in either system, nor for pTG395 in the wheatgerm extract system. For these constructs the percentage of initiation at the first methionine was assumed to be the same as that found for pMD2.

the wheat germ extract system varied considerably from 12 : 1 for pTG394 to 21 : 1 for both pTG393 and pTG395.

3.5.4   Conclusions

- The ability of the FMDV 2A region to mediate a cleavage at its C-terminus was significantly enhanced by the addition of 39 amino acids from the C-terminus of FMDV 1D region.

- The addition of 180 amino acids from the C-terminus of FMDV 1D region resulted in complete cleavage occurring, such that an uncleaved polyprotein was not detectable.

- The addition of 5 amino acids from the C-terminus of FMDV 1D region did not significantly increase the cleavage ability of FMDV2A.

- N-terminal additions to the FMDV 2A region did not affect the ratio of CAT2A : GUS produced.

Section 3.6   A study of FMDV 2A activity in prokaryotes

3.6.1   Introduction

cDNA constructs encoding [CAT2AGUS] have previously been expressed in *E. coli* and anti-GUS antibodies used to probe for the presence of GUS as an indication of 2A cleavage activity (see Figure 3.6.1; Donnelly *et al.*, 1997). However, no GUS protein was detected. Anti-GUS antibodies indicated only the presence of the uncleaved [CAT2AGUS] polyprotein. No attempt was made at this time to probe for CAT2A, since it was assumed that it would only be found in the presence of GUS. In the light of the results of the site directed mutagenesis experiments of 2A, in which the CAT2A protein was found in the absence of GUS, it was necessary to repeat this and probe also for the CAT2A cleavage product.
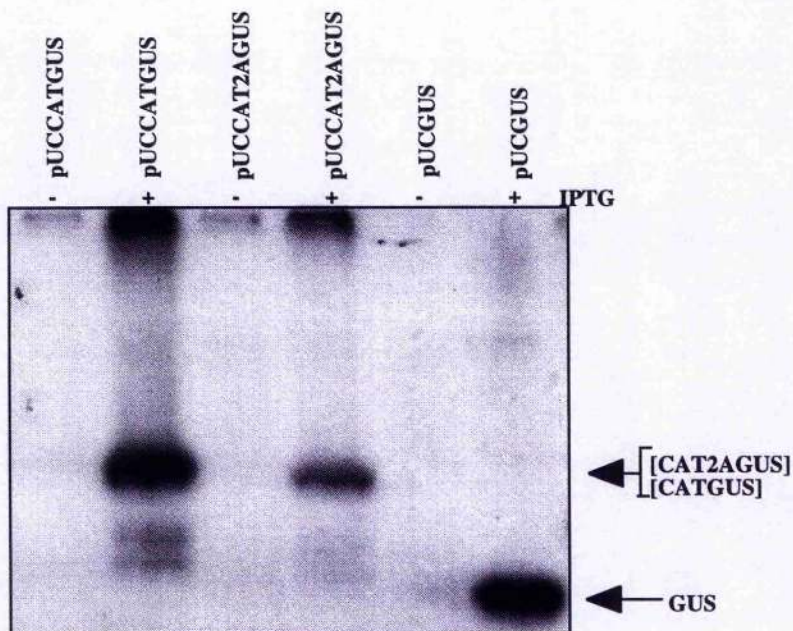

Figure 3.6.1. Prokaryotic expression. Artificial polyproteins were expressed in *E. coli* and the cleavage activities assessed by probing a western blot of total cellular proteins with anti-GUS antibodies.

149

It was proposed to express the polyproteins [CAT2AGUS] and [GFP2AGUS] in *E. coli* and analyse the products by 10 % PAGE and western blotting with anti-CAT and anti-GFP antibodies as well as anti-GUS antibodies.

It was chosen to base the constructs, required to express these polyproteins, on the pUC plasmid series. These plasmids contain the *lacZ'* fragment, encoding the α-peptide of ß-galactosidase, with a multiple cloning site (MCS) downstream of its translation-initiation codon. This part of the *lacZ* gene is sufficient to complement the M15 deletion of the *lacZ* gene in the host genome. Insertion of DNA fragments into the multiple cloning site generally leads to a non functional *lacZ'* gene providing blue/white screening. Genes inserted into the MCS are expressed as fusion products under control of the *lac* promotor which can be derepressed by IPTG. Proteins expressed therefore contain the N-terminal 10 amino acids of the α-peptide of ß-galactosidase.

3.6.2   Results

3.6.2.1   Construction of pUC:CAT2AGUS and the controls pUC:CAT, pUC:GUS and pUC:CATGUS

The cDNA constructs pUC:CAT2AGUS, pUC:CAT, pUC:GUS and pUC:CATGUS were constructed by Susan Monaghan in this laboratory. Briefly, the sequences encoding [CAT2AGUS], CAT, GUS and [CATGUS] were excised from pCAT2AGUS, pCAT14/20, pGUS17/22 and pCATGUS, respectively, by restriction digests with BamH I and Nde I, and purified by agarose gel electrophoresis. These fragments were ligated into the similarly restricted and purified vector pUC8 and transformed into *E. coli*. Transformants were selected by blue/white screening and the DNA maxi-prepared.

150

3.6.2.2 Construction of pUC:GFP2AGUS and the control pUC:GFPGUS

The cDNA constructs pUC:CAT2AGUS and pUC:CATGUS were doubly digested with the restriction enzymes BamH I and Xba I. The large fragments, from which the CAT gene had been removed, were purified by agarose gel electrophoresis. A DNA fragment, encoding green fluorescent protein, was excised from the commercially available vector GFP-N$_2$ via a BamH I and Xba I double restriction digest. This was purified by agarose gel electrophoresis, before being ligated into the similarly restricted vector fragments. *E. coli* were transformed with the ligation reactions and colonies picked for maxi-preparation of the recombinant plasmids pUC:GFP2AGUS and pUC:GFPGUS.

3.6.2.3 Prokaryotic expression and analysis of recombinant polyproteins

Plasmid DNA was transformed into *E. coli* strain JM109 (recA1 supE44 endA1 hsdR17 gyrA96 relA1 thi Δ(lac-proAB), F'[traD36 proAB$^+$ lacI^q lacZΔM15]), and expression of the fusion protein induced by IPTG as described in section 2.2. The cells were harvested three hours after induction and disrupted by sonication and the insoluble material separated by centrifugation. The proteins from samples of uninduced whole cells, induced whole cells, induced soluble cell fractions and the induced insoluble cell fractions were separated by PAGE and the proteins stained with Coomassie brilliant blue. This revealed good expression for CAT and GUS recombinant proteins, and a lesser expression of [CATGUS], [CAT2AGUS], GFPGUS and GFP2AGUS recombinant polyproteins. CAT and GUS proteins were strongly indicated in the soluble cell fractions, whereas the others were most strongly indicated in the insoluble fractions. It was therefore decided to use the total cellular protein fraction to probe for GFP, GUS and CAT containing proteins in a western blot. A sample of the total cellular proteins from each induced culture was separated by 10 % PAGE and the proteins transferred to a PDVF membrane and analysed by western blotting. The distribution of proteins was determined by probing the membrane

151

with anti-GUS, anti-CAT and anti-GFP antibodies (kind gift of Simon Santa-Cruz). The bound antibodies were visualised by enhanced chemiluminescence (ECL) as per the manufacturers instructions. Figure 3.6.2 shows the results of western blotting of the protein products from the expression of plasmids pUC:CAT2AGUS and pUC:GFP2AGUS as well as the controls pUC:GUS, pUC:CAT, pUC:CATGUS and pUC:GFPGUS. The anti-GUS antibodies bound only to [CATGUS], [CAT2AGUS], GFPGUS and GFP2AGUS polyproteins in samples derived from *E. coli* transformed with pUC:CATGUS, pUC:CAT2AGUS, pUC:GFPGUS and pUC:GFP2AGUS respectively, and to GUS protein expressed from plasmid pUC:GUS. The anti-GFP antibodies bound only to GFPGUS and GFP2AGUS polyproteins, from *E. coli* transformed with pUC:GFPGUS and pUC:GFP2AGUS respectively. The anti-CAT antibodies bound to [CATGUS] and [CAT2AGUS] polyproteins from *E. coli* transformed with pUC:CATGUS and pUC:CAT2AGUS respectively, and to CAT protein expressed from plasmid pUC:CAT. Neither GFP2A nor GUS proteins were detected in samples from *E. coli* transformed with pUC:GFP2AGUS. Similarly neither CAT2A nor GUS protein cleavage products were detected in samples from *E. coli* transformed with pUC:CAT2AGUS.

3.6.3   Discussion

The products from 2A mediated cleavage of [CAT2AGUS] and [GFP2AGUS] were not observed. Only the full length uncleaved polyproteins were indicated by western blotting. The FMDV 2A region therefore appears to be completely inactive in prokaryotes, neither pausing translation at its C-terminus nor mediating proteolytic cleavage.

3.6.4   Conclusion

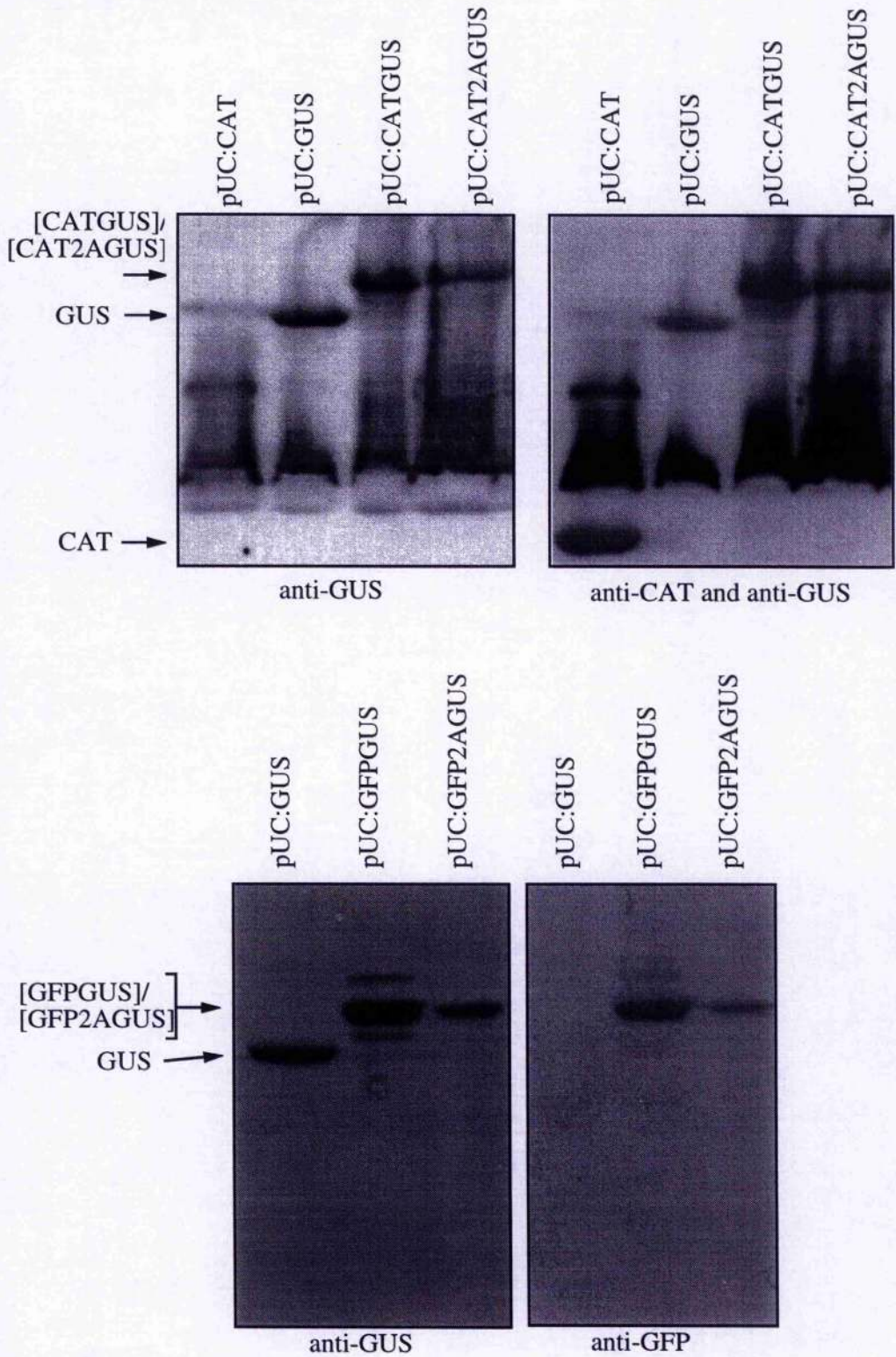- FMDV 2A region does not display any cleavage activity in *E. coli*.

Figure 3.6.2   Prokaryotic expression of GFP, CAT and GUS containing polyproteins. Artificial polyproteins were expressed in *E. coli* and the cleavage activities assessed by probing an western blot of total cellular proteins with anti-CAT, anti-GFP and anti-GUS antibodies.

Section 3.7 The effect of changing the protein environment of the FMDV 2A region.

3.7.1 Introduction

The wild-type upstream context of the FMDV 2A region was shown to increase the efficiency of 2A mediated cleavage (Ryan and Drew, 1994; Section 3.5 this thesis). It was shown by Ryan *et al.* (1991) that the downstream context of FMDV 2A did not effect the efficiency of 2A mediated cleavage.

We wished to ascertain whether the foreign gene sequences, CAT and GUS, surrounding the 2A region were influencing 2A activity, either in efficiency of cleavage or in the ratio of CAT2A and GUS products formed. In order to do this, it was planned to compare the products of TnT reactions programmed with plasmid constructs which had 2A inserted between a range of foreign genes.

This section describes the construction of plasmids pGFP2ACAT2AGUS and pGFP2AGUS. The translation products of these plasmids as well as pCAT2AGUS, and pCAT2ANEO and pMR101:GFP (already available within our laboratory, encoding [CAT2ANEO] and [GFP2ANEO]) should allow a comparison of 2A activities of when either GFP or CAT precede the 2A region and CAT, NEO or GUS follow it.

3.7.2 Results

3.7.2.1 Construction of pGFP2AGUS

pCAT2AGUS was digested with BamH I and Xba I to excise the CAT gene and the large remaining fragment purified by agarose gel electrophoresis. pGFP-N2 (CLONTECH
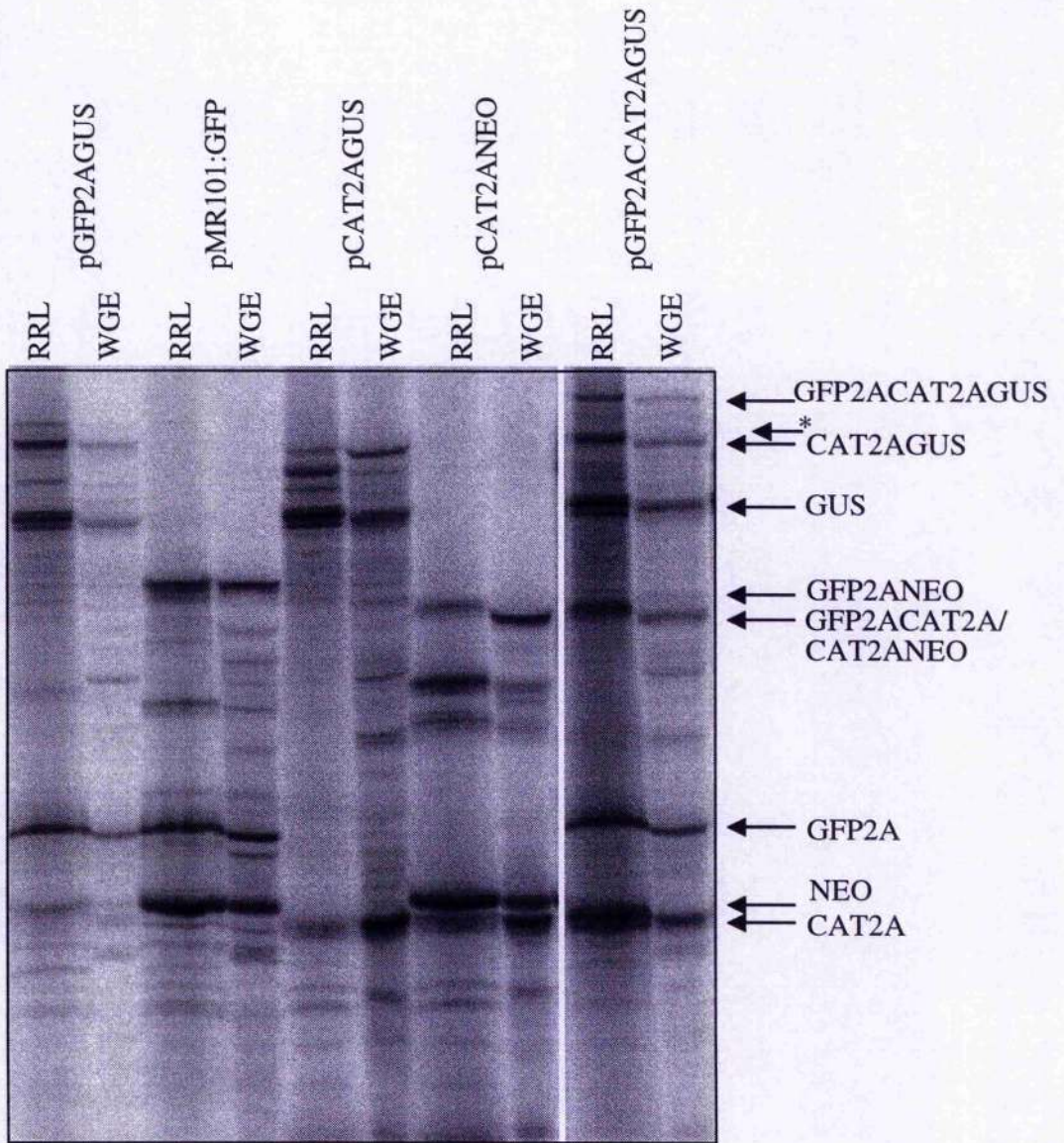
Laboratories Inc.) was also digested with BamH I and Xba I and the fragment containing the GFP gene purified by agarose gel electrophoresis. The purified fragments were ligated to form pGFP2AGUS and transformed into *E. coli*. pGFP2AGUS was then maxi-prepared from both XL1 Blue *E. coli*, for use in TnT reactions, and a Dam (-) strain of *E. coli* which would allow use of the Xba I site as necessary for further cloning.

3.7.2.2   Construction of pGFP2ACAT2AGUS

The plasmid pGFP2AGUS was linearised by the restriction enzyme Apa I at the 2A cleavage site, purified by agarose gel electrophoresis, and treated with calf intestinal alkaline phosphatase to prevent recircularisation. A PCR product containing the sequences encoding CAT2AΔGUS was amplified from pMD2 (encoding [CAT2AGUS]) using the forward primer CAT14 and reverse primer OR63. The PCR product was purified, digested with Apa I and the large fragment encoding CAT2A was ligated into the linearised pGFP2AGUS. The ligation reactions were transformed into *E. coli* and transformants screened by restriction mapping.

```
                          M  E  K  K  I  T
CAT14:   5'    CGCGCGGGGCCCATGGAGAAAAAAATGACT    3'
                    ApaI
```

```
OR63:  5' ACTGAATGCCCACAGGCCGTC 3'
```

3.7.2.3 TnT reactions of plasmids encoding 2A in reporter gene cassettes

Maxipreparations of the plasmids pGFP2AGUS, pCAT2ANEO, pCAT2AGUS, pMR101:GFP (encoding [GFP2ANEO]) and pGFP2ACAT2AGUS were transcribed and translated in coupled rabbit reticulocyte lysate and wheat germ extract systems. The products were analysed by denaturing PAGE (Figure 3.7.1).

155

Figure 3.7.1 Translation *in vitro* **of constructs encoding 2A between two reporter genes.** Coupled transcription/translation reactions in both rabbit reticulocyte lysate (RRL) and wheat germ extract (WGE) were programmed with plasmid DNA as labelled. The translation products are indicated. The symbol * refers to the product GFP2AGUS. Those products not identified are caused by internal initiation.

3.7.3   Discussion

The plasmids constructs used in this section, pCAT2AGUS, pCAT2ANEO, pGFP2AGUS, pMR101:GFP (encoding [GFP2ANEO]) and pGFP2ACAT2AGUS are depicted in Figure 3.7.2. The linker sequence immediately before 2A is the same in all constructs. Immediately after the 2A region in all of these constructs is the codon for the proline residue of the scissile bond, followed by the sequence of the next gene.

These constructs were transcribed and translated in rabbit reticulocyte lysate and wheat germ extract systems (Figure 3.7.1). In each case the expected 2A mediated cleavage products were obtained together with uncleaved polyprotein. Densitometric analyses of the translation products, the results of which are shown in Tables 3.7.1 - 3.7.3, were performed as in Section 3.2 to determine relative quantities of the two cleavage products and the percentage of cleavage exhibited for each polyprotein.

Values for the percentage of cleavage were calculated with respect to the quantity of the first cleavage product (*i.e.* CAT2A or GFP2A) and were found to be between 77 - 85% in rabbit reticulocyte lysate and 84 - 96 % in wheat germ extract, which within the expected error of these experiments is not significantly different from one another. Lower values were obtained for the polyprotein [GFP2ANEO], but these were due to the trivial reasons explained below.

The similarity in the percentages of cleavage was not particularly surprising since it was previously suggested that only the protein prior to 2A was important for the efficacy of cleavage and the linker sequence of 6 amino acids, -SRGACQ-, immediately preceding 2A was the same in all polyproteins. One can however assume that the CAT or GFP sequence before the 2A region was not having a marked affect on the cleavage reaction. The linker sequence was shown in Section 3.5 to result in a reduced cleavage efficacy as opposed to the wild-type sequence.
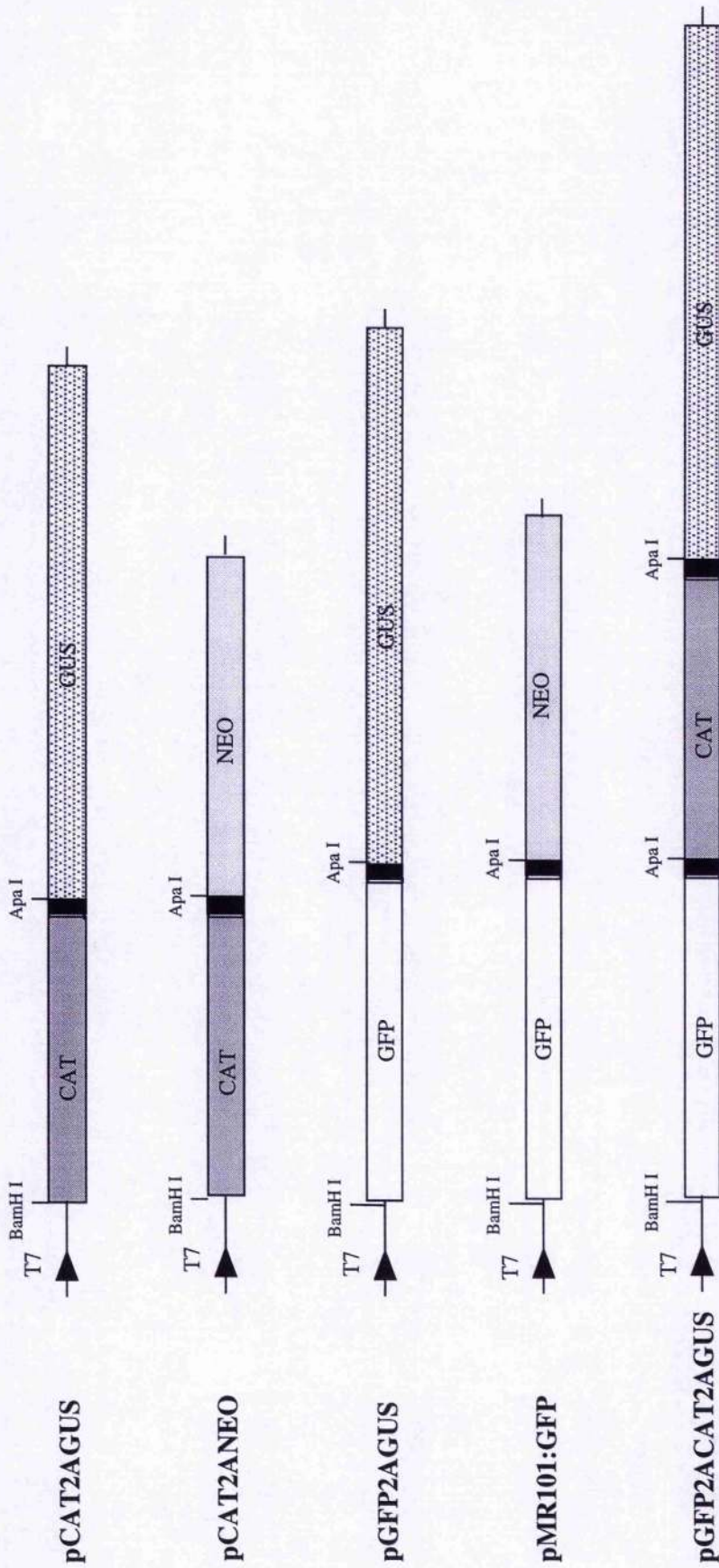
Figure 3.7.2 Plasmid constructs. The boxed areas indicate the single open reading frames encoding the artificial polyproteins. Solid black areas represent the 2A region of FMDV. In all cases the linker sequence immediately preceding 2A is the same; the sequence contains Xba I and Bbu I restriction sites and encodes the amino acid sequence SRGACQ. Prior to the 1st methionine of each gene following 2A is the codon for proline from the FMDV 2B region.

Construct	Rabbit reticulocyte lysate reactions		Wheatgerm extract reactions	
	Ratio $\frac{1st\ protein}{2nd}$	% Cleavage based on 1st protein	Ratio $\frac{1st\ protein}{2nd}$	% Cleavage based on 1st protein
pCAT2AGUS	1.2	82	8.4	96
pCAT2ANEO	1.2	85	1.8	84
pGFP2AGUS	1.4	77	5.3	96
pGFP2ANEO *	0.5	72	0.9	79

Table 3.7.1 Results of the densitometric analysis of the translation products seen in Figure 3.7.1. The densitometric analysis was carried out exactly as described in the Results Section 2. In all cases there was some degree of internal initiation of protein synthesis. The percentage of initiation at the first methionine was calculated and the quantity of the first protein corrected to account for 100 % of initiation. This corrected figure was used to calculate the ratios of the first and second cleavage products.

* The ratio of cleavage products GFP2A : NEO for GFP2ANEO was unexpectedly low in both systems. This can be attributed to the comigration of a GFP specific internal initiation product with NEO.

159

The ratio GFP2A : NEO was found to be <1 : 1 and the percentage of cleavage measured was lower than for the other construct examined here. However, closer examination of both [GFP2AGUS] and [GFP2ANEO] translation profiles (Figure 3.7.3) revealed that a GFP specific product, probably due to internal initiation within GFP and subsequent cleavage by 2A, was co-migrating with the NEO protein. Thus the measurement of PSL attributed to NEO was elevated from its true value. This product is most evident in the rabbit reticulocyte reactions hence the larger deviation from expected values in rabbit reticulocyte lysate reaction programmed with DNA encoding GFP2ANEO.

The ratio of cleavage products from [CAT2ANEO], [CAT2AGUS] and [GFP2AGUS] demonstrated an excess of the first cleavage product. In the rabbit reticulocyte systems ratios ranging from 1.2 :1 and 1.4 : 1 are fairly similar. However in the wheat germ extract system the CAT2A :GUS and GFP2A : GUS ratios of 8.4 : 1 and 5.3 : 1 demonstrate a very obvious excess of the first cleavage product. This result might suggest that when a GUS gene follows the 2A region it leads to a decrease in the continuation of translation.
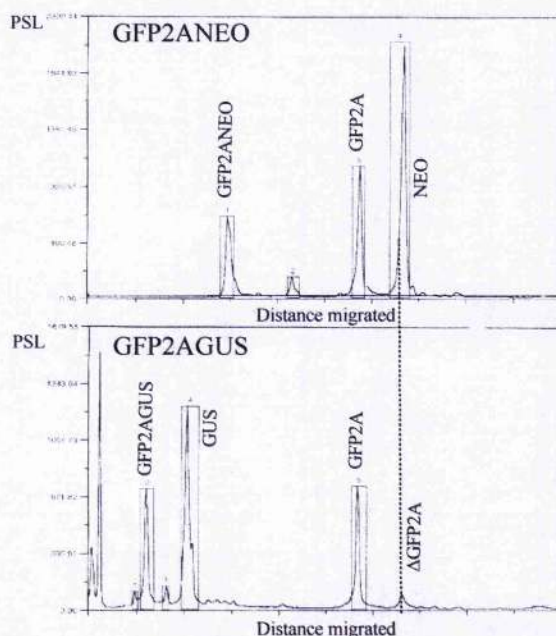
Figure 3.7.3   Profiles of [GFP2AGUS] and [GFP2ANEO]. Profiles of the lanes containing the rabbit reticulocyte reactions programmed with pGFP2AGUS and pGFP2ANEO from Figure 3.7.1.

The results of densitometric analysis of the translation products of the polyprotein [GFP2ACAT2AGUS] are shown in Table 3.7.2. This polyprotein was able to be processed at either or both of the two FMDV 2A regions; cleavage only after the first 2A region resulted in GFP2A and [CAT2AGUS] polyproteins; cleavage only after the second 2A region resulted in [GFP2ACAT2A] and GUS polyproteins; cleavage at both 2A regions produced GFP2A, CAT2A and GUS proteins.

Relative molar quantities of each product were derived from the PSL count. In the rabbit reticulocyte system the relative corrected molar quantity of polyprotein which commenced translation at the start of GFP, equal to the sum of all GFP containing products, was 2745. The relative molar quantity of GFP2A released by 2A mediated cleavage at the first 2A region was 2052 which is equivalent to a 75 % efficiency of cleavage. Conversely, the sum of molar quantities of polyproteins for which this 2A junction was uncleaved was found to be 25 %. The relative molar quantity of both [CAT2AGUS] and CAT2A was 2100, thus, within the experimental error, there are equal numbers of cleavage products from the first 2A mediated cleavage. With a ratio of cleavage products for the first 2A region of 1:1 (*i.e.* there has been no termination of translation) there was still a relative molar quantity of 2745 translating polyproteins available to be cleaved by 2A at the second 2A cleavage site. A relative molar quantity of 254 polyproteins (9 %) experienced no cleavage at this site and continued translating to produce either [GFP2ACAT2AGUS] or [CAT2AGUS] polyprotein products. Cleavage at this second 2A cleavage site produced the CAT2A containing products [GFP2ACAT2A] and CAT2A itself in the relative molar quantity of 2491 or 91 %. However the corresponding relative molar quantity of GUS protein produced from this cleavage reaction was only 1563, a value equivalent to 57 % of cleavage at this site. The missing 34 % of GUS protein therefore must be due some sort of interruption of translation after this site.

In the wheat germ extract system the relative corrected molar quantity of polyprotein which commenced translation at the start of GFP was 471. The relative molar quantity of GFP2A

Protein product	Rabbit reticulocyte lysate			Wheat germ extract		
	PSL reading	Relative molar quantities	Adjusted for GFP internal initation	PSL reading	Relative molar quantities	Adjusted for GFP internal initation
GFP2ACAT2AGUS	1053	38	54	96	3.4	5
GFP2ACAT2A	6809	454	639	876	58	85
CAT2AGUS	4398	200	-	430	20	-
GFP2A	8739	1457	2052	1567	261	381
CAT2A	17100	1900	-	2700	300	-
GUS	20320	1563	-	1521	117	-

Table 3.7.2 Results of densitometry of the translation profiles of GFP2ACAT2AGUS. The densitometric analysis was carried out exactly as described in Results part 2. Some internal initiation within GFP at Met-153 occured. The percentages of full length polyprotein was densitometrically measured in rabbit reticulocytes to be 71 % and in wheat germ extract to be 69 %. The relative molar quantities for proteins containing GFP were therefore adjusted to account for this.

recombinant protein	Rabbit reticulocyte lysate		Wheat germ extract	
	Summation of relative quantity of translation products containing the recombinant protein	Normalised to GFP2A	Summation of translation products containing the recombinant protein	Normalised to GFP2A
GFP2A	54+639+2052=2745	1.0	5+85+381=471	1.0
CAT2A	54+639+200+1900=2793	1.0	5+85+20+300=410	0.9
GUS	54+200+1563=1817	0.7	5+20+117=142	0.3

Table 3.7.3 Summation of the total products of GFP2ACAT2AGUS. Relative quantities of translation products containing GFP2A, CAT2A or GUS were calculated to allow a calculation of the ratios of genes translated before and after each 2A region.

released by 2A mediated cleavage at the first 2A region was 381 which is equivalent to cleavage occurring in 81 % of polyproteins. Conversely, the sum of molar quantities of polyproteins for which this 2A junction was uncleaved was found to be 19 %. The relative molar quantity of both [CAT2AGUS] and CAT2A was 320, a value equivalent to 68 % of cleavage at this site, thus an interruption in translation must have occurred in the 13 % of cases unaccounted for. The ratio of cleavage products for the first 2A region is thus 1 : 0.84. The molar quantity of 410 translating polyproteins was hence available to be cleaved by 2A at the second 2A cleavage site to produce [GFP2ACAT2A], CAT2A and GUS. A relative molar quantity of 25 polyproteins (6 %) experienced no cleavage at this site and continued to translate producing either [GFP2ACAT2AGUS] or [CAT2AGUS] polyprotein products. This cleavage reaction produced the CAT2A containing products, [GFP2ACAT2A] and CAT2A itself, in a relative molar quantity of 385 or 94 %. However, the corresponding relative molar quantity of GUS protein produced from this cleavage reaction was only 117, a value equivalent to only 29 % of cleavage at this site. Thus, an interruption of translation after this site in 65 % cases must have occurred.

The analysis of the ratios of cleavage products before and after 2A is complicated since 2A mediated cleavage of this polyprotein is not complete at either 2A region. A final determination of the ratios was found by the summation and comparison of all cleavage products containing either GFP2A, CAT2A, or GUS, as seen in Table 3.7.3. In rabbit reticulocyte lysate and wheat germ extract the ratio of GFP2A : CAT2A : GUS of approximately 1 : 1 : 0.7 and 1 : 0.9 : 0.3, respectively. Both of these results illustrate a interruption of translation following the second 2A region and following the first 2A region in wheat germ extract. As observed previously, the effect is greater in the wheat germ system. Cleavage at the first 2A region, which is followed by CAT, in different batches of rabbit reticulocyte lysate has consistently been found to produce its components in a 1:1 ratio.

3.7.4   Conclusions

• There was no significant difference in the cleavage efficiency of the various polyproteins examined here.   However, although the sequence following the 2A regions varied considerably, the sequence immediately preceeding 2A was the same for all polyproteins. This similarity is perhaps being reflected in the efficiency of 2A cleavage. No correlation could be drawn between the protein preceeding 2A and the ratio of cleavage products.

• Only when the CAT protein followed the 2A region was there a 1 : 1 ratio of cleavage products. When any other protein followed the 2A region there resulted an excess of the first cleavage product (see Figure 3.7.4).

• It seems that the context prior to the 2A region may be important only for the efficiency of cleavage, and the context following the 2A region is important in determining the degree to which translation continues to produce both cleavage products in equal amounts. The imbalance of cleavage products seen here is again directly at variance with a proteolytic cleavage hypothesis.

Comparison of the relative molar quantities of cleavage products
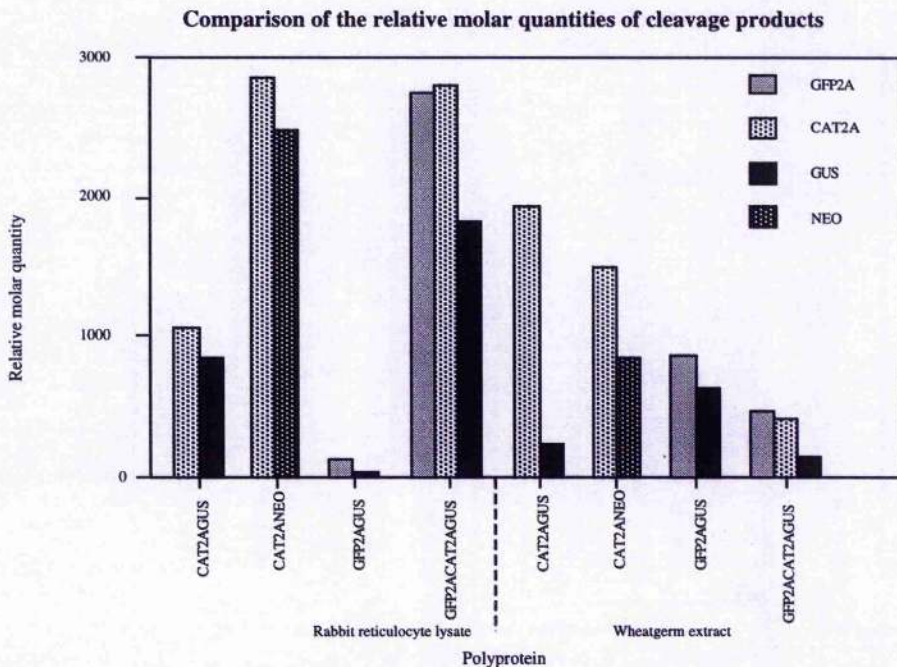
Figure 3.7.4   Comparison of the measured relative molar quantities of cleavage products.

4. DISCUSSION

On commencing this project we wished to test the hypothesis that the FMDV 2A region possessed proteolytic activity and then, if possible, determine its mechanism of action. This discussion begins with the results of studies of synthetic 2A containing peptides which came to light at the start of this project. These results suggested that the nature of the bond to the P1'-proline might have been important, and hence the possibe effect of prolyl *cis / trans* isomerism on 2A activity is considered. The results of experiments to examine the activity of the 2A regions of FMDV (both wild-type and mutated) and cardioviruses using *in vitro* translation systems and in prokaryotes are then discussed. These studies showed that the 2A "cleavage" event was in all probablity not caused by proteolysis, but was more likely to be a translational effect. A brief introduction to translation is, therefore, given before the final presentation of our model for 2A-mediated activity.

Section 4.1 The activity of FMDV 2A synthetic peptides

4.1.1 Synthetic peptides exhibit no cleavage activity

At the outset of this project it was proposed that the 2A "proteolytic" element itself, being capable of mediating a cleavage event in a completely foreign context (Ryan and Drew, 1994), might also be capable of mediating cleavage when chemically synthesised. The intention was therefore to chemically synthesise the peptide of the FMDV 2A region, and allow, or induce it to cleave, then study the kinetics of this cleavage. The syntheses of many FMDV 2A peptides with mutations at specific positions would be quick and facile, and a study of their activity was intended. An extensive range of conditions were tested, by Rhonda Hussain within our laboratory, to induce hydrolysis of the peptide bond at the C-terminus of 2A, with a peptide of 29 residues in length spanning the complete 2A

region. However, all attempts to induce cleavage to occur were in vain and, in fact, no cleavage of a synthetic 2A peptide has been reported to date. It was however reported as a personal communication that the tetra-peptide NPGP corresponding to the three C-terminal residues of 2A and proline of 2B, cleaved to give two di-peptides, NP and GP, by the terminal-amino group (Palmenberg, 1990). No evidence to support this claim has been published to date. Although this report might represent the inherent instability of the NPGP sequence, it demonstrates hydrolysis of a different peptide bond to that in 2A mediated cleavage.

One possible explanation for the lack of cleavage in chemically synthesised peptides might be that in a biological system the proline-glycine peptide bond may be formed in the *cis* orientation, whereas in the chemically synthesised peptide this bond most probably equilibrates to form predominantly the *trans* peptide bond.

4.1.2 Proline and *cis / trans* isomerism

Proline is distinct amongst the components of proteins in that it possesses a secondary amino group which results in N-dialkyl or tertiary amide bonds when incorporated into a peptide. Lacking an amide proton when found in proteins or peptides, hydrogen bonding between proline and hydrogen bond acceptors is precluded. Proline also perturbs the structure of proteins due to steric interactions, thus, having a strong influence on the structure of proteins. It disrupts regular α-helices both sterically and by preventing hydrogen bond formation with carbonyls in the previous turn of the helix, such that it is hardly ever found at the C-terminus and rarely in the centre of an α-helix (Yaron and Naider, 1993). Proline residues are found where the backbone changes direction, especially in β-turns. The tertiary amide character of the X-Pro bond confers an increased basicity to the attached carbonyl oxygen atom and hence a greater propensity for the involvement of its carbonyl group in such specific hydrogen-bonded structures.

It was shown by Linus Pauling 45 years ago that for most peptide bonds amino acid side chains are held as far apart as possible, in the *trans* configuration (Figure 4.1.1; Pauling and Corey, 1953). The hydrogen atom attached to nitrogen is small, so there is significantly less steric congestion in the *trans* conformation, and the *trans* isomer is up to 40 kJ mol^{-1} more stable than the *cis* isomer (Juvvadi *et al.*, 1992). Amide bonds can also be represented as their oxime tautomers; so called amide resonance. This partial double bond character of the carbon-nitrogen bond (bond length 1.33 Å) would retard rotation. Thus the energy of the transition state between *trans* and *cis* amide bonds is so large, at approximately 85 kJ mol^{-1}, that any interconversion from *trans* to *cis* is extremely slow.

However, there are many known examples of *cis*-X-proline amide bonds in nature (where X is another amino acid); the observed *cis*-X-proline abundance in small peptides is 10 to 30 %, compared to less than 0.1 % for other amino acids (MacArthur and Thornton, 1991). The energies of *cis* and *trans* amide bonds to proline residues are comparable since neither is particularly more sterically hindered, and which of these conformations has the lowest energy is determined by the surrounding peptide environment. The partial double bond character is slightly reduced in an amide bond involving the nitrogen of proline (bond length 1.36 Å), since it has no hydrogen attached and the polarisation of charge at nitrogen is, therefore, slightly decreased.
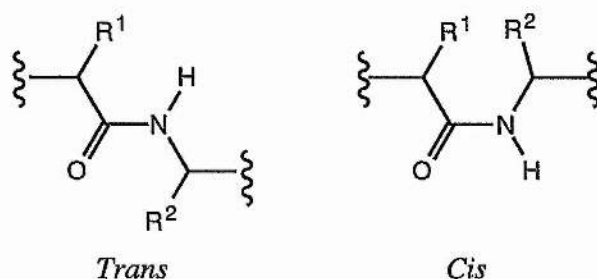

Trans   *Cis*

Figure 4.1.1. *Trans* **and** *cis* **conformations of amide bonds.**

The energy barrier to interconversion of the *cis* and *trans* conformers is, hence, reduced to about 55 kJ mol[-1] (MacArthur and Thornton, 1991), such that rotation about this amide bond is possible although slow (see Figure 4.1.2). The *trans*-X-proline conformation is more stable by 4-10 kJ mol[-1] in water (Jorgensen and Gao, 1987; Eberhardt *et al.*, 1993; Wuthrich and Grathwohl, 1974; Vasquez, 1983), and occurs more often in folded proteins, although the *cis* isomer easily occurs in unfolded proteins. Indeed the isomerisation of proline is thought to be a rate limiting step in the folding of proteins (Brandts *et al.*, 1975).

The proposed requirement of a *cis* peptide bond between glycine and proline is further supported by the reported observation of large *cis / trans* ratios in other peptides containing the sequence glycine-proline-phenylalanine. This sequence has been found in (Gly[6]) bradykinin, for which the *cis* conformation accounts for 38 % of the population (Anteunis *et al.*, 1981; London *et al.*, 1979).
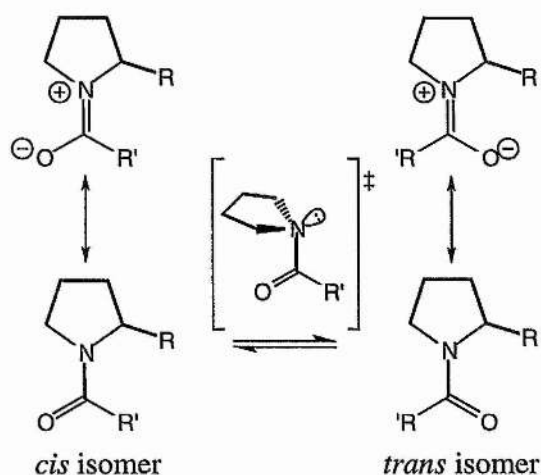

cis isomer   trans isomer

Figure 4.1.2 Prolyl *cis-trans* isomerisation.

168

The presence of a proline residue restricts a proteins susceptibility to most peptidases, particularly for the X-pro bond, but also for bonds not adjacent to proline (Freedman, 1979). Thus, it is not surprising that only a few, specific, enzymes participate in the cleavage of bonds involving proline. Those enzymes which are known to perform a cleavage of X-pro are prolidases, aminopeptidase P and HIV 1-protease. Aminopeptidase P is a metallo-protease, which requires a *trans* X-pro bond and for proline to be the penultimate residue. Closely related are prolidases which hydrolyse dipeptides in the *trans* conformation, also at the X-pro bond. The HIV 1-protease is highly specific for X-pro sequences in the middle of a polypeptide (*i.e.* it is an endopeptidase). It belongs to the family of aspartate proteases, several members of which have also been found to exhibit this activity. This resistance of X-pro bonds to proteases substantiates the theory that the FMDV 2A sequence is not merely being recognised by a host protease closely coupled with the ribosome. Previous experiments have demonstrated that any putative protease cleaving the X-pro bond would need to be very tightly coupled to the ribosomes (Ryan and Drew, 1994; Section 3.2 this thesis), however, no such protease has been descibed in the literature.

Section 4.2 Activity of FMDV 2A in eukaryotic *in vitro* translation systems

4.2.1 The activity of the FMDV 2A region in the polyprotein CAT2AGUS
- A surprising result!

We wished to examine the activity of the 2A region of FMDV in the absence of other FMDV sequences which might influence 2A-mediated activity. A further advantage of removing the FMDV 2A region from the context of the FMDV polyprotein itself, was that the FMDV 3C^{pro} polyprotein processing would have otherwise considerably complicated the interpretation of the data. It was, therefore, chosen to study the 2A region of FMDV in an artificial reporter gene polyprotein system, within which the sequence encoding the FMDV 2A region was placed between two reporter genes. The reporter genes used initially were CAT and GUS in the plasmid pMD2.

In vitro coupled transcription and translation of the plasmid pMD2, encoding [CAT2AGUS], produced the polyprotein [CAT2AGUS], and the cleavage products CAT2A and GUS (as identified by immunoprecipitation with anti-CAT and anti-GUS serum) in agreement with previous reports (Ryan and Drew, 1994). There were considerable amounts of internal initiation products from the coupled transcription / translation reactions, which could have been reduced by improving the Kozak consensus sequence of the translation-initiating codon of CAT, although this is a typical problem with these *in vitro* systems. However, an assignment of the major internal initiation products, and a densitometric analysis of their relative abundance, provided sufficient information to allow the continued use of plasmid pMD2 and other plasmids based upon it. A comparison of the translation products of pCATGUS and pMD2 demonstrated that the truncated polyproteins, [Δ^{75Met}CAT2AGUS] and [Δ^{163Met}CAT2AGUS] produced by internal initiation at Met-75 and Met-163, respectively, within the CAT gene, and the full length [CAT2AGUS] polyprotein were processed equally efficiently by the FMDV 2A region.

170

Unexpectedly, densitometric analyses of the cleavage products, taking the internal initiation into account, demonstrated CAT2A and GUS *were not present in equal quantities*. The ratio of CAT2A : GUS was calculated to be between 2.4 : 1 and 5.6 : 1 in the rabbit reticulocyte system. In the wheat germ extract system the ratio of CAT2A : GUS was even greater, although it was more variable, at between 6.2 : 1 and 14 : 1.

Initially these differences in stoichiometry were thought to be due to trivial reasons such as degradation, or incomplete synthesis of the GUS protein. However degradation experiments in which the transcription / translation reaction of pMD2 was arrested and the reaction mixture then incubated at 30 °C (the temperature at which transcription / translation reactions are carried out) for up to 24 hours demonstrated that although both CAT2A and GUS were degraded to some extent, the GUS protein was degraded more slowly than the CAT2A protein, and not to the extent which would be required to produce such differences in stoichiometry. One would thus expect that if equal quantities of CAT2A and GUS were made then, should degradation have been a factor requiring consideration, there would have been more GUS than CAT2A detected. It has been reported that the synthesis of protein from larger genes in rabbit reticulocyte lysate is less efficient than the synthesis of smaller proteins (Dasso and Jackson, 1989). If the synthesis of the GUS gene was not being completed for this reason then one might have expected to see a "smear" of protein representing a range of C-terminally truncated GUS proteins. However this was seen neither in the polyacrylamide gel analyses of the transcription / translation reactions nor in the immunoprecipitation of their products. One might also have expected that there would be a greater "smear" for the wheat germ extract reactions in which the ratio of CAT2A to GUS is greatest. This, however, was not the case either. Thus, it was concluded that the synthesis of GUS was not being hampered by its size.

The only explanation for an excess of CAT2A was that there must have been more CAT2A *translated* than GUS. This was most likely to be controlled at the level of

translation than transcription since translation reactions programmed with RNA from an *in vitro* transcription reaction also displayed this effect. The ratios of CAT2A : GUS can be explained by an 'interruption' in translation, which in some cases results in termination producing CAT2A only, in others it releases CAT2A then continues to produce GUS (termed cleavage), and in others the ribosome continues as normal to produce CAT2AGUS (termed non-cleavage). Such a concept is not compatible with the theory of a 2A-mediated proteolytic cleavage.

4.2.2 The importance of the 2A RNA sequence in 2A activity

If an interruption in translation was occurring, then it was thought that perhaps the 2A RNA sequence or secondary structure was important for FMDV 2A activity. The nucleotide sequence encoding 2A in pMD2 had been altered from the wild-type sequence in order to introduce new restriction enzyme sites. It was feared that this might have resulted in a change from 2A wild-type activity. However, plasmid pTG395 which encoded CAT-(5 C-terminal residues of 1D)-2A-GUS, in which the 1D/2A sequences were completely wild-type displayed similar cleavage activity and product ratios (Donnelly *et al.*, 1997). It can therefore be concluded that the activity of the 2A region encoded by pMD2 was not affected by the alteration of the nucleotide sequence to produce new restriction enzyme sites.

The introduction of a +1 frameshift throughout the FMDV 2A region was employed to study any effects mediated by the RNA sequence encoding 2A, rather than the 2A protein itself. As expected the majority of translation products from the construct encoding [CAT(+1)2AGUS] was the uncleaved polyprotein and its truncated forms (due to internal initiation). Unexpectedly, however, this was accompanied by a small quantity of a product of the same size as CAT2A which was immunoprecipitated by anti-CAT

antibodies. This result cannot be explained but may indicate that the RNA sequence of FMDV 2A does play a role in termination of translation after translation of the 2A region. The 2A RNA sequence was examined for possible secondary structure elements using the programme RNAdraw (Hofacker *et al.*, available at htp.itc.univie.ac.at.), however, no consistent features were found within the predicted 2A RNA secondary structures of the aphtho- and cardioviruses. An examination of the primary RNA sequence for the C-terminus of 2A revealed no consistency in codon usage between aphtho- and cardioviruses, although within the aphthoviruses themselves there was reasonably high identity.

4.2.3 The influence of the protein context of 2A on its cleavage ability

The 2A-mediated cleavage occurs co-translationally. It is, therefore, reasonable to assume that the wild-type FMDV sequence following 2A is not necessary for cleavage and indeed this was shown by Ryan *et al.* (1991). In the FMDV virus 2A acts as the species P1-2A, so it is not surprising that P1 is influential in mediating 2A activity (Ryan *et al.*, 1991). This influence may be asserted either by residues close in the primary sequence or those which are spatially close in the tertiary structure of P1-2A. The x-ray crystallographic data of the capsid protomer (see Figure 1.5.1; Lea et al, 1994) demonstrated that the C-terminus of 1D was not associated with any other part of the protomer but with the adjacent protomer in the pentamer. Since 2A-mediated cleavage is a co-translational process, which may be occurring in the ribosomal exit pore from which residues distant in the primary sequence would already have been extruded, it is unlikely that residues very distant in the primary sequence are involved. Hence, it was assumed most likely that residues from within 1D were required to enhance cleavage activity to 100 %. Indeed, the inclusion of C-terminal 1D sequences of 180 and 39 amino acids in length did increase the efficiency of cleavage (this thesis, Section 3.3). However, they *did not improve the stoichiometry of the cleavage reaction.*

173

Proteins GFP and CAT were both used to precede the 2A region in artificial reporter gene cassettes (see Section 3.7). No correlation with the preceding reporter gene protein could be found for either the efficiency of cleavage or the ratio of cleavage products. Proteins CAT, NEO and GUS were each encoded after 2A regions. Only when the CAT protein was C-terminal to 2A was the ratio of cleavage products found to be 1 : 1. As predicted no correlation was found for the protein following 2A and the extent of cleavage.

4.2.4 Comparing rabbit reticulocyte lysate and wheat germ extracts

The ratios of CAT2AGUS cleavage products in both rabbit reticulocyte lysate and wheat germ extract varied from batch to batch, but an excess of CAT2A always existed and was always greater in the wheat germ extract system. The percentage of cleavage based either on the quantity of uncleaved material or CAT2A was more consistent in rabbit reticulocyte lysate with average values of 91 % in rabbit reticulocyte lysate and 96 % in wheat germ extract. Qualitatively, in both translation systems the activity of 2A (and mutated 2A regions) was similar.

The relatively similar percentages of cleavage reaction occurring between both batches and translation systems may be read as an indication that the ability to mediate 2A cleavage resides with the 2A region alone. However, the occurrence of large quantitative differences between the two systems for the ratios of CAT2A : GUS (and smaller batch variations), may infer that the translational machinery of these *in vitro* systems is playing a role in the determining the ratio of products (*i.e.* determining the frequency of termination of translation after 2A). If 2A-mediated cleavage was merely a proteolytic event then its dependence on ribosomes would be expected to be less.

Section 4.3   The activity of mutated FMDV 2A regions

The site-directed mutagenesis of the FMDV 2A region was designed to test which residues might be important for 2A-mediated proteolysis. The activities of mutated 2A regions however were not consistent with a proteolytic activity and the results of these experiments posed more questions than answers.

The site-directed mutations were inserted in two distinct contexts. Firstly, most mutations were inserted into the 2A sequence of pMD2, which encoded the wild-type sequence (-PFFF-) immediately after the 2A cleavage site. Upon translation all of these mutations resulted largely in unprocessed [CAT2AGUS] polyprotein, but also in the cleavage product CAT2A. Only a few showed evidence of the accompanying GUS cleavage product. A similar effect was also seen in some of the EMCV 2A mutation experiments of Hahn and Palmenberg (1996). In both cases at least four wild-type residues followed the cleavage site. These data were, therefore, not consistent with a co-translational proteolytic cleavage hypothesis, for which both cleavage products would have been produced in equal amounts. Secondly, some mutations were made within the 2A sequence of plasmid pCAT2AGUS(3), which encoded the methionine of GUS immediately following the P1'-proline of the 2A cleavage site. The polyproteins, thus encoded, yielded reduced levels of both cleavage products compared to the wild-type sequence, and as with the wild-type sequence the CAT2A product was produced in excess.

The differences between the two plasmid systems may again indicate that the context following the 2A region is important for the "cleavage" reaction in determining the ratio of products. Since the sequence following the 2A region may not have been translated at the moment of "cleavage", either the following protein and / or RNA sequence may be implicated in this effect. Three mutated 2A sequences were made in both plasmid systems: the mutations of P5-glutamate to aspartate, and P4-serine to phenylalanine, and

175

the insertion of an alanine within the -KLAG- motif to give -KLAAG. The mutation of P5-glutamate to aspartate and the insertion of alanine in the 2A sequence of pMD2 both resulted in a 2A region which was largely unprocessed but also made a small quantity of CAT2A. When made in the context of pCAT2AGUS(3) the alanine insertional mutant was completely inactive resulting in uncleaved polyprotein alone, whereas the glutamate to aspartate mutant "cleaved" very poorly, also yielding an excess of CAT2A. The observation that the CAT2A "cleavage" product of the insertional mutation in pMD2 was abolished on being placed in the context of pCAT2AGUS(3) strongly supports the theory that the sequence C-terminal to 2A is important for termination or pausing of translation at the C-terminus of 2A. The mutation of P4-serine to phenylalanine showed moderate activity in both systems, although the excess of CAT2A was greater in the translation of the pMD2 derived plasmid.

The FMDV 2A region appears either to be capable of two separate activities - termination of translation and protein cleavage - or to have a single activity which has two alternative consequences - protein termination or protein cleavage. The small size of the 2A region perhaps makes it more likely that it possesses a single activity which can result in either termination of translation or an apparent proteolytic cleavage.

The mutated sequences which were obtained are shown in Figure 4.3.1. Their "cleavage" activities are indicated and compared with the mutations of the EMCV 2A sequence made by Hahn and Palmenberg (1996). To consider the activity of each 2A region only the production of *both* CAT2A and GUS was considered as cleavage activity.

Significant cleavage activity was seen for the mutations of P4-serine to isoleucine encoded by pCAT2AGUS(3) and P4-serine to phenylalanine encoded by both pCAT2AGUS(3) and pMD2. Hence, it can be concluded that P4-serine, (corresponding to methionine or threonine in the cardioviruses) is neither a catalytic residue nor important for formation of the active structure.

176

```
pCAT2AGUS        ..S R G A C Q L L N F D L L K L A G D V E S N P G P M L R P
-single mutations                              A*        Q F H
                                                         D I E
                                                           Q


pMD2             ..S R G A C Q L L N F D L L K L A G D V E S N P G P F F F M
-single mutations              P*              P*        E   D F A V A
                                               A*        Q     R S
                                                                 T F
                                                                   I

-double mutations                   --IH†--              N   Q
                                    --HI†--              D   Q
                                    --RI†--              N   H
                                                         Q   H
                                                             E
                                                         D   D
                                                             Q
                                                         Q   Q


Hahn and Palmenberg  ..Y F A D L L I H D I E T N P G P F M F R
-single mutations                    H V D A L A L
                                     N F     R E R
                                               Q V
                                                 W
```

Figure 4.3.1 An alignment of site-directed mutations created within 2A regions. The identification of significant cleavage activity is indicated in red, a small amount of cleavage activity in magenta and no cleavage activity in blue. The termination of translation after 2A was not considered, since this appears to be determined in part at least by the plasmid construct, rather than the mutation. Only the cleavage activity to give both CAT2A and GUS is considered here. † Indicates the replacement of the KLAG motif. * Indicates an insertional mutation.

177

Mutations of the P3-asparagine residue to histidine and glutamate both resulted in 2A regions which retained a significant amount of activity. This was interesting since this residue is entirely conserved throughout cardio- and aphthoviruses. The proposed structural function of the asparagine in its interaction with the P5-glutamate during folding (see Section 1.6) is not supported by this data, since mutation to glutamate would surely abolish this interaction. However, this data does not disprove such a theory, it merely shows that if an interaction occurs it is not a requirement for 2A activity.

Mutations of the P5-glutamate residue were also fairly active when encoded in the plasmid pCAT2AGUS(3). Surprisingly the mutation to aspartate was less efficient than the mutation to glutamine, suggesting that the side-chain length was important for the function of this group.

The P7-aspartate residue appeared to be essential for activity with mutations to both glutamate and glutamine producing inactive 2A regions, suggesting that this residue plays a critical role, either structurally or directly in catalysis. In the mutational studies of Hahn and Palmenberg this residue was also essential for activity with neither mutation to histidine nor asparagine being active (1996).

The insertion of residues towards the N-terminus of 2A all resulted in inactive 2A regions. The insertion of proline residues, within the sequence -LLKLAG-, would have disrupted the proposed α-helical nature of the 2A region and both proline and alanine residues could have disrupted possible stabilising side-chain / side-chain interactions of P14-aspartate and lysine or lysine and P7-aspartate. Inactivation of the 2A region by replacement of the -KLAG- motif with the cardiovirus motif, -IH-, and the mutants, -HI- and -IR-, may also have prevented the side-chain interactions stabilising the proposed α-helix. The side-chain interactions which might be possible within the FMDV 2A sequence and these mutated sequences are shown in Figure 4.3.2.

178

The P1-glycine and P2-proline residues appear to be essential for activity. Mutation of these residues, within either the FMDV 2A sequence or EMCV 2A sequence (Hahn and Palmenberg, 1996), resulted in 2A regions which demonstrated no cleavage activity. One can therefore assume that these residues are essential for the formation of the correct conformation and / or environment at the scissile bond. Very slight activity was observed for two of the mutations of P1'-proline (to serine and isoleucine). Similar activities were reported by Hahn and Palmenberg for mutations of this proline within the EMCV sequence.

FMDV 2A   Q L L N F D L L K L A G D V E S N P G

Alanine
insertion   Q L L N F D L L K L A A G D V E S N P G

KLAG
replaced   Q L L N F D L L I H D V E S N P G
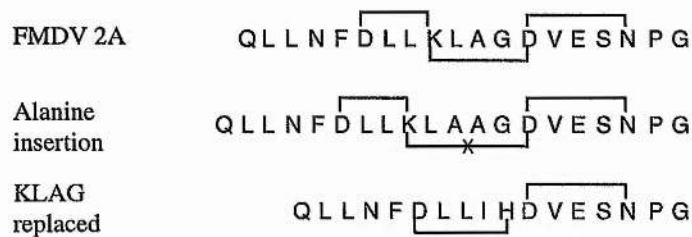
Figure 4.3.2 Side-chain interactions of an α-helical 2A region. The possible side chain interactions of the FMDV 2A region and the mutated 2A regions in the proposed α-helical conformation are indicated. Both of these mutations shown would reduce the side-chain interactions of an α-helical FMDV 2A region.

179

Section 4.4 Comparison of the activity of Aphthovirus and Cardiovirus 2A regions

An examination of the alignment of cardiovirus and aphthovirus 2A proteins shows a high degree of identity within TME viruses and within EMC viruses, but not between EMC and TME viruses - with the exception of the C-terminal region, which is also very similar to the FMDV and ERV 2A regions (Figure 1.5.2). In Section 3.3 it was shown that truncated cardiovirus 2A regions of equivalent size to the FMDV 2A region possess the same degree of cleavage activity. Although the extent to which cleavage occurred was surprisingly high for the truncated cardiovirus 2A regions, it had been expected that some degree of activity would be shown by the C-terminal portion since it appeared to contain equivalent functional groups to the FMDV 2A region.

Using dynamic molecular modelling studies the FMDV 2A region was predicted to form an amphipathic α-helix which was stabilised by a number of side-chain / side-chain interactions followed by a turn at the C-terminus (see Figure 1.6.1). The corresponding regions of TMEV, EMCV and ERV-1 and 2 contain residues with the propensity to form a similar amphipathic helical structure (Figure 4.4.1). The extent to which the putative polar and charged side-chains can interact is most limited in the EMCV C-terminus of the 2A region, and most extensive in the TMEV 2A region. Truncation of the either EMCV or TMEV 2A protein would not effect these interactions. However, in the case of FMDV and ERV-1 an interaction between P16-asparagine and the lysine residue at the end of the 1D chain is destroyed when the 2A region is inserted into a foreign context. Notably, restoration of this putative interaction in the translation product of pTG395 (Section 3.5) did not restore 2A activity to 100 %.

The residues predicted to be involved in the formation of a turn -ENPGP- are completely conserved throughout cardio and aphthoviruses. Therefore, it can be concluded that the

same structural motifs which are present in the FMDV 2A region are very likely also present in the equivalent C-terminus of the cardiovirus 2A region.

The putative catalytic P7-aspartate residue, which was found to be essential for catalytic activity in the site directed mutagenesis studies of FMDV 2A, is entirely conserved throughout cardio- and aphthoviruses (see Section 4.3). The double mutation, within TMEV 2A, of P7-aspartate and P4-methionine to asparagine and isoleucine, respectively, was inactive. However mutation of the P4-methionine residue alone to threonine appeared to slightly increase the cleavage activity. Within the EMCV sequence Hahn and Palmenberg have mutated the corresponding threonine to alanine and found the resultant 2A region to retain a high degree of activity. The corresponding serine residue in FMDV was substituted by phenylalanine and isoleucine and in both cases 2A was found to be active (Section 4.4); 2A activity appears to be remarkably resistant to change at this position. Thus the fortuitous discovery of the double mutation indicates that an aspartate at this position is also necessary for cardiovirus 2A-mediated cleavage. Hahn and Palmenberg reported that within the EMCV sequence 10-30 % of 2A activity still remained after mutation of this aspartate to asparagine, but very little evidence for this is seen.

FMDV 2A	Q L L N F D L L K L A G D V E S N P G
ERV-1 2A	Q C T N Y S L L K L A G D V E S N P G
EMCV 2A	A Y A G Y F A D L L I H D I E T N P G
TMEV 2A	H A D Y Y K Q R L L I H D I E T N P G
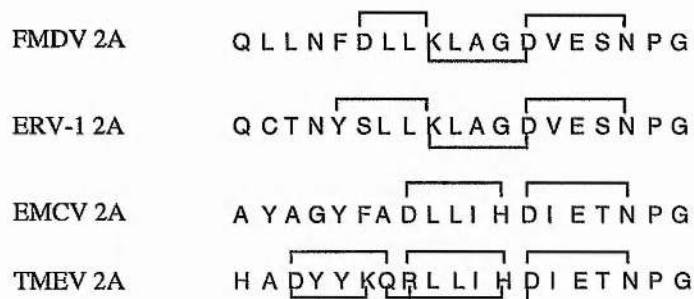
Figure 4.4.1 Side-chain interactions of aphtho cardio- and equine rhinovirus 2A regions. The C-terminal regions of cardiovirus' 2A proteins and the 2A regions of aphthoviruses are shown with possible side-chain interactions indicated.

181

The mutation of P14-tyrosine to a serine resulted in a small decrease in cleavage efficiency. However a more notable defficiency of this construct was its ability to then synthesise the GUS gene. This is consistent with a general trend noticed for mutants of the FMDV 2A region in which the poorer cleaving mutants are also less capable of producing the cleaved GUS protein, and have a higher CAT2A : GUS ratio.

The high levels of cleavage activity observed in the [CAT-ΔTME2A-GUS] and [CAT-ΔEMCV2A-GUS] polyproteins raises the question what is the function of the remaining approximately 85 % of the cardiovirus 2A protein? It seems unlikely that its function is merely to enhance the C-terminal regions activity to 100 % when FMDV can do this with a total length of between 24 and 55 amino acids. It is equally unlikely that the 140 amino acid cardiovirus regions have, under evolutionary pressure, been preserved with the sole purpose of being able to cleave themselves from the rest of the P2 region and then to provide a site to be cleaved by 3C from the P1 region. It seems more probable that it is responsible for another function in addition to the 2A / 2B cleavage. Since the aphtho- and cardioviruses cleave C-terminally in contrast to the rhino- and enteroviruses any function of the cardiovirus 2A region other than the primary cleavage event could perhaps be *cis*-acting in capsid protein precursor processing or encapsidation, or perhaps less likely, *trans*-acting in replication.

Section 4.5 The activity of the FMDV 2A region in prokaryotes

The reason for the inability of FMDV 2A to mediate a cleavage in a prokaryotic system is not clear. The rapidity of 2A-mediated cleavage and the stability of the small proportion of uncleaved material in eukaryotic systems is evidence of a co-translational cleavage. Since this is achieved in a wide range of eukaryotic cells but not prokaryotic cells it appears to indicate a requirement for 80S ribosomes or another eukaryotic factor probably closely coupled to translation. This claim is supported by the inability of chemically synthesised peptides to cleave themselves.

The consistent observation of an excess of the first cleavage product over the second indicates that this is not merely an unusual proteolytic cleavage which is occurring. It is more accurately described as a novel event which results in the production of two products from a single mRNA, and which appears to involve some sort of interruption in translation such that more of the translation product N-terminal of 2A is formed.

For such an event to occur it is conceivable that the translating ribosomes play a role in creating an environment for efficient "cleavage", if not a more active part. Since FMDV infects mammals only then it would not be surprising that this 2A-mediated processing event has been optimised to occur on 80 S eukaryotic ribosomes under the conditions of protein synthesis imposed by eukaryotic translation factors. Although in outline the protein synthesis of both prokaryotes and eukaryotes is similar there are some major differences. The overall roles of ribosomal particles, initiator tRNAs, elongation and release factors are analogous, however, their structures are quite different. Prokaryotic 70S ribosomes are composed of two subunits: a small 30S subunit containing 16S RNA and 21 proteins and a large 50S subunit containing two rRNA molecules of 5S and 23S and 31 proteins. Eukaryotic 80S ribosomes are also composed of two subunits: a small 40S subunit, containing 18S RNA and 33 proteins, and a large 60S subunit, containing

three rRNA molecules of 5 S, 5.8 S and 25-28 S as well as 50 known proteins. Initiation of translation is the most dissimilar process between the two systems with at least 10 eukaryotic initiation factors required compared to only 3 for prokaryotes, and the interactions between the small ribosomal subunit, initiator tRNAs, and mRNA being completely different. Another interesting difference is in the eukaryotic elongation factor 2 (eEF-2) which, with a GTP cofactor, promotes translocation. This factor was found to possess the ability to bind to RNA which was not observed for its prokaryotic counterpart EF-G (Domogatsky *et al.*, 1978). Indeed the inhibition of the synthesis of proteins following the 2A region of EMCV, in a fractionated cell-free system from Krebs-2 cells, was abolished by the addition of the eukaryotic elongation factor 2 (eEF-2) (Svitkin and Agol, 1983). It would not be surprising if such differences between the two translation systems might not permit a 2A-mediated "cleavage" to occur in prokaryotes.

Section 4.6   A model for the mechanism of FMDV 2A mediated activity

In order to provide a possible model for aphtho- and cardiovirus 2A-mediated "cleavage" the salient points gleaned from the literature and this research are summarised below. Since the cleavage reaction is closely coupled to translation a discussion of eukaryotic translation is given before the presentation of a model for 2A activity which appears to be consistent with all of the known data.

4.6.1 Summary of FMDV 2A data

- Chemically synthesised peptides encompassing the FMDV 2A region and the first residue of 2B did not display any proteolytic cleavage ability in a variety of solvents.

- Both the 2A region of FMDV, and the equivalent sequence of the C-terminal part of cardiovirus 2A regions, together with the first proline residue of the 2B sequences are capable of mediating a co-translational "cleavage" at their C-termini.

- Post-translational processing of the small quantity of uncleaved material produced in addition to the cleavage products does not occur.

- An excess of the first cleavage product is usually produced.

- Only the first cleavage product was seen in translation reactions programmed with constructs containing certain mutations.

- Only the translation product encoded prior to the 2A cleavage site (*i.e.* L-P1-2A) was observed upon translation of EMCV RNA in a fractionated Krebs-2 cell free extract. Translation of 2BC was restored by the addition of eEF-2 (Svitkin and Agol, 1983).

- The protein and / or RNA sequence immediately following the 2A region is important in determining whether translation of the second product occurs and to what extent.

- No 2A-mediated "cleavage" was observed in prokaryotes.

- The residues P5-Gln, P4-Ser, P3-Asn and P1'-proline are not essential for 2A-mediated "cleavage" activity.

- The Asp residue appears to be essential for cleavage activity and as such could be a possible nucleophile in the "cleavage" reaction.

- The FMDV 2A region has been predicted by molecular modelling studies to be largely α-helical with a turn towards the C-terminus.

4.6.2 2A -Action within the context of translation

The translation of mRNA into protein in all species occurs in three distinct phases; initiation, elongation and termination. It is after elongation of the 2A peptide by eukaryotic ribosomes that 2A-mediated cleavage occurs and this seems in some cases to interfere with (or perhaps terminate) the continued elongation though the following protein. The excess of the first cleavage product cannot otherwise be accounted for. This outline of translation will therefore concentrate on the elongation and termination steps of eukaryotic translation.

A number of initiation factors (eIFs) are required to initiate translation. A cap-binding complex is involved in the recognition of cellular capped mRNA and the subsequent formation of an initiation complex which includes the 40S small subunit and the mRNA. The ribosome then scans along the mRNA until a start codon is found upon which protein synthesis begins. The initiating methionyl-tRNA occupies a position termed the peptidyl-site of the 40S subunit with its anti-codon base paired to the AUG codon of the mRNA and the 60S subunit joins the complex to form the final 80S initiation complex.

mRNA is selected by the small ribosomal subunit and during translation interacts with both the small subunit and the tRNA anti-codons. Ribosomes have been found to contain three sites which the tRNA can occupy during elongation: the acceptor (A)-site, at which translation of all internal codons occurs, the peptidyl (P)-site, to which a tRNA is transferred after the peptidyl transferase reaction, and the exit (E)-site, which is occupied

186

by a newly deacylated tRNA before its release from the ribosome (see Figure 4.6.1). In the A-site the tRNA interacts with 16S rRNA as well as mRNA; the interaction with rRNA is weak, presumably so as not to distract from the anticodon-codon interaction (Kaziro *et al*, 1978). There is no evidence for aminoacyl-tRNA binding to the large subunit within the A-site (Porse et al, 1995). Within the P-site tRNA interacts with both subunits. An interaction with 23S rRNA is probably required to position the acceptor end of t-RNA in the peptidyl-transferase site. Deacylated tRNA is then transferred to the E-site on the large subunit from which it can spontaneously dissociate. However, although there are only three sites for the tRNA molecules to occupy, there are more than 3 states within which an elongating ribosome can exist, since the tRNA can be bound in hybrid A/P and P/E sites between both subunits (see Figure 4.6.2)
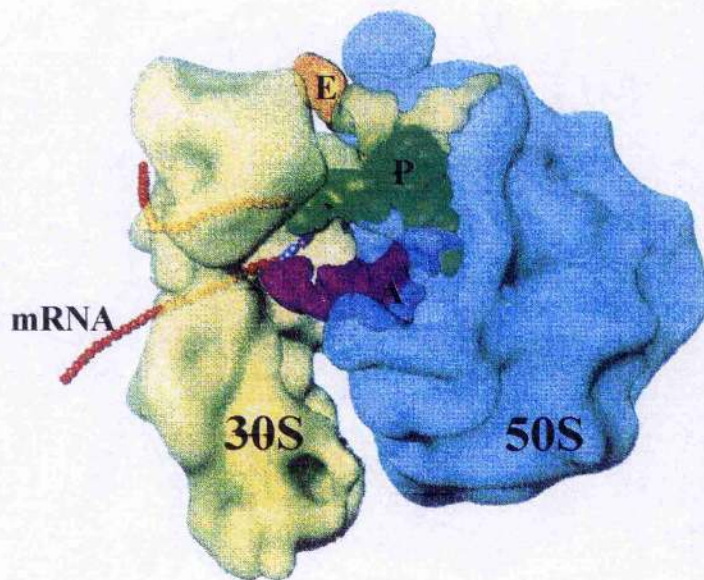
Figure 4.6.1 A model of a eukaryotic ribosome (Frank, 1997). The A-, P- and E-sites which tRNA can occupy are indicted.

Once the elongation complex has been formed a series of cyclic events resulting in peptide chain elongation follow until a termination codon is reached (see Figure 4.6.2). There are three steps required for elongation; (i) eEF-1α binds GTP and aminoacyl-tRNA leading to codon-dependent placement of this aminoacyl-tRNA at the A-site of the ribosome; (ii) peptidyl transferase catalyses peptide bond formation; (iii) after peptide bond formation the binding of eEF-2.GTP promotes translocation of the mRNA by one codon to allow a new aminoacyl-tRNA to occupy the A-site.

The prokaryotic equivalent of eEF-1α, EF-Tu, recognises the overall spatial structure of the stacked CCA 3'-end of the tRNA as well as the 3'-adenosine and NH_3^+ of the aminoacyl group. Presumably eEF-1α may also recognise these features, and it is known to form a ternary complex with GTP and aminoacyl-tRNA. This complex is required for efficient entry of aminoacyl-tRNA to the A-site of the ribosome, the rate limiting step in elongation of the peptide chain. The CCA-aminoacyl moiety is buried in the ternary complex and only available for binding within the large ribosomal subunit after release from the complex by GTP hydrolysis. Peptide bond formation cannot occur before this hydrolysis and the resulting delay caused enhances the fidelity of translation by allowing non-cognate codon / anticodon interactions to be broken. The eEF-1α.GDP complex is released from the ribosome and the elongation factors eEF-1$\beta\gamma$ facilitate the exchange of bound GDP for GTP.

On release from the ternary complex the 3'-end of the aminoacyl-tRNA may move relative to the ribosome towards the P-site of the large ribosomal subunit. A conformational change is induced leading to an opening of the catalytic centre. An interaction between the CCA tRNA 3'-terminus and 23S rRNA occurs (demonstrated in prokaryotes), which probably serves to position the acceptor end of the tRNA in the peptidyl transferase site (Samaha *et al.*, 1995). On access to the catalytic site the peptidyl transferase reaction occurs spontaneously and rapidly.
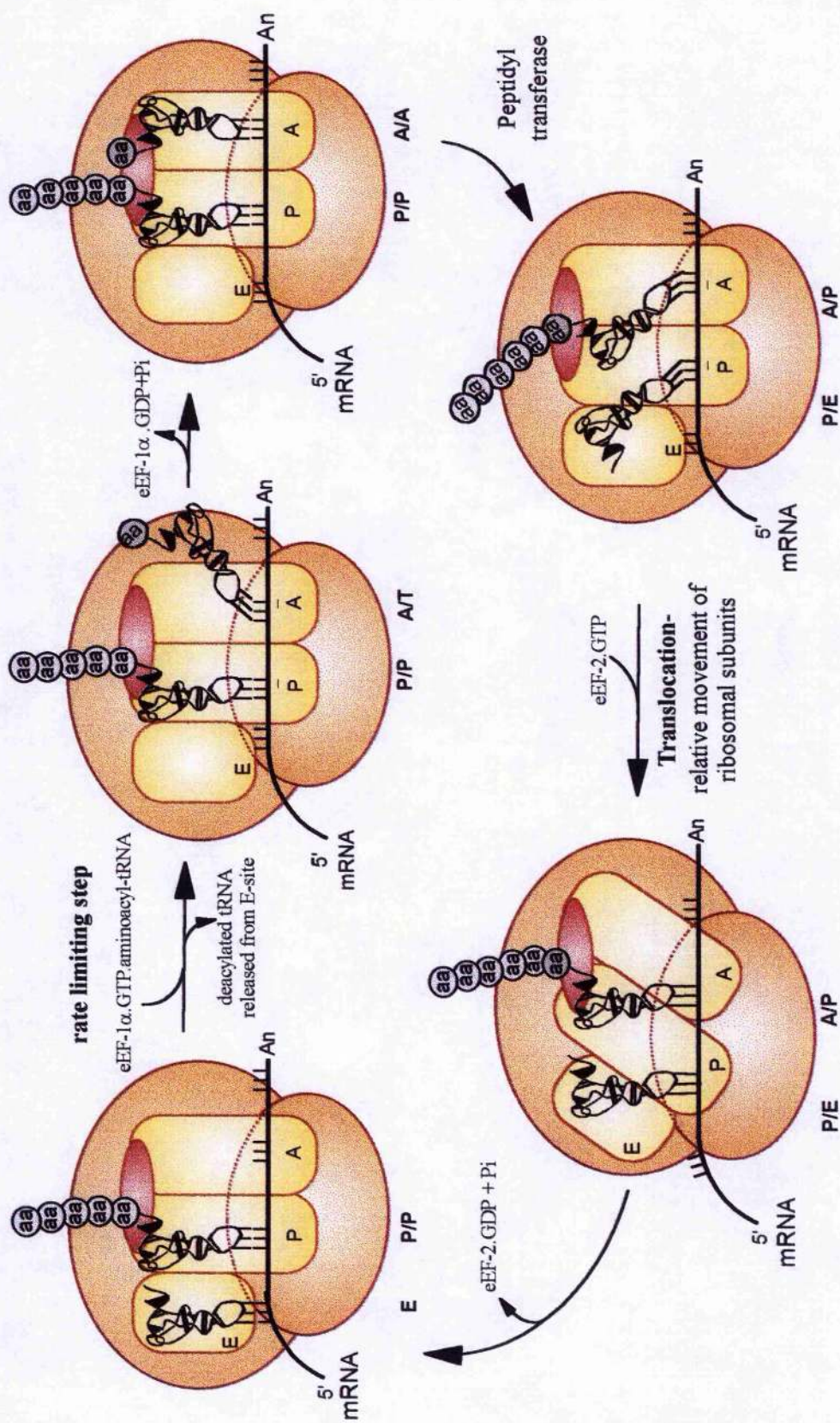
Figure 4.6.2 Schematic representation of the elongation cycle of translation. The A, P and E sites on the ribosomal subunits are indicated. Shown below each tRNA molecule is the state within which the tRNA is said to be. The peptidyl transferase centre is represented by the pink area.

189

The peptidyl transferase activity resides in the large ribosomal subunit. It was shown that a 50 S subunit depleted of over 90 % of its protein component could still catalyse the peptidyl transferase reaction (Noller *et al.*, 1992). This activity was destroyed by ribonuclease treatment, thus demonstrating the importance of rRNA in this reaction. The standard free energy of hydrolysis of the ester between the aminoacyl-residue and the tRNA is about -30 kJ mol^{-1} compared with -2 kJ mol^{-1} for an amide. Consequently, although the loss in entropy is approximately 10.5 kJ mol^{-1} for each peptide bond formed, there is still an overall release of energy associated with the process of peptide bond formation (Spirin, 1985). The elongation factors and GTP could well, therefore, only provide a kinetic contribution for the reaction. At this pre-translocation stage of translation the new peptidyl-tRNA and deacylated tRNA enter the "hybrid" states. The deacylated tRNA is bound to the P-site of the small ribosomal subunit in a P/E hybrid state in which the 3' end of the tRNA has been shifted by approximately 20Å into the E-site. Peptidyl-tRNA is bound to the A-site in an A/P hybrid state.

Translocation of the mRNA is now required in order to move the mRNA by one codon and shift the anticodon ends of both tRNA molecules together with the mRNA towards the next site. It appears that the movement of the mRNA is dependent upon, and coupled to, that of the tRNA with the tRNA sites providing the precision for movement of the mRNA by precisely one codon. The elongation factor eEF-2 and GTP are required for the translocation step. The eEF-2.GTP complex binds to the ribosome and it is generally accepted that this induces a conformational change, allowing translocation to occur. The hydrolysis of GTP reduces the affinity of the complex for the ribosome and it dissociates. However, it has recently been suggested that the hydrolysis of GTP drives translocation (Rodnina *et al.*,1997). Previous observations that GTP hydrolysis was not required for translocation, but for the release of EF-G after translocation, had led to the conclusion that GTP hydrolysis must have taken place after translocation. However no kinetic experiments were performed which would have shown that GTP hydrolysis substantially accelerates translocation and proceeds five times faster than translocation as demonstrated

by Rodnina *et al.* By analogy to the structural change of EF-Tu brought about by GTP hydrolysis, the loss of the γ-phosphate from GTP probably induces a conformational strain which influences the overall architecture of eEF-2. This in turn drives rearrangement of the ribosome. The results of Rodnina *et al.* suggest that domain IV of eEF-2 has a role in coupling the conformational change of the factor to the ribosome (1997). The main target of this domain is likely to be the small ribosomal subunit. Translocation may be promoted by eEF-2 inducing a rearrangement of the small and large subunits relative to one another. A ribosomal state may then be created in which the tRNA molecules can move together with the mRNA. The function of eEF-2 would thus be to translate the chemical energy stored in GTP into a directional molecular movement of the ribosome. It seems logical, given the universal bipartite structure of ribosomes with each tRNA interacting with both subunits, that the relative movement of the two subunits is required for the ribosomal function. Deacylated tRNA now spontaneously dissociates from the E-site and the peptidyl-tRNA exists in the P/P state with the A-site available for the entry of the next aminoacyl-tRNA.

A stereochemical analysis of ribosomal peptide bond formation has suggested that an α-helical conformation is adopted at the C-terminal end of the nascent peptide due to the angle of attack of the aminoacyl-tRNA on the peptidyl-tRNA (Lim and Spirin, 1986). Some 30 to 40 amino acid residues at the C-terminal end of the nascent polypeptide have consistently been shown to be protected by the ribosome from protease treatments (Malkin and Rich, 1967; Blobel and Sabatini, 1970). The consistent protected length suggests a special channel for the nascent polypeptide exists and that the conformation of all polypeptides (which may be α-helical) is universal within it.

The presence of stop codons UAA, UAG or UGA, for which no tRNA molecules with the appropriate anticodon exist, results in the binding of a release factor to the ribosome. A single release factor functions with all three stop codons, unlike prokaryotes for which two release factors exist to recognise all stop codons. A second eukaryotic release factor,

eRF3, was recently discovered which forms a complex with eRF1 and is capable of binding GTP. An eRF3.eRF1.GTP complex can bind to the A-site of the ribosome, perhaps mimicking the eEF-1α complex, and trigger a peptidyl transferase reaction. However on this occasion there is no aminoacyl-tRNA nucleophile available to attack the ester bond to the nascent peptide and instead water attacks this bond to release the peptide.

4.6.3 A model for 2A-mediated "cleavage".

The critical observation which indicated that the 2A-mediated cleavage was not a proteolytic cleavage was the excess of the first cleavage product. It was also noticed that in the wheat germ system the excess was always larger than in the rabbit reticulocyte system. The cleavage reaction was clearly not dependent solely on the 2A sequence but also on the translating ribosomes.

Our data shows an excess of the first cleavage product, which could occur by the termination of protein synthesis after translation of the 2A region. Hydrolysis of the ester linkage between the nascent peptide and the tRNA to release the first "cleavage" product is one mechanism by which this could happen. Such an ester hydrolysis could be mediated by either (i) the 2A protein, or (ii) peptidyl transferase - *via* the same mechanism by which termination of translation occurs when triggered by a release factor. The 2A region itself could directly catalyse the hydrolysis reaction; the hydrolysis of an ester requiring less energy than that of an amide. If this was the case it would explain why no 2A activity was seen in experiments involving chemically synthesised peptides. The ribosome itself would also be expected to participate at least in providing the catalytic environment.

The hydrolysis of the ester linking the nascent peptide to the tRNA would result in a ribosome with deacylated tRNAs in either the A/P and P/E sites or E and P/P sites. Figure 4.6.3 demonstrates the possible ribosomal states within which the hydrolysis could occur and the possible outcomes. These may be unstable states and the ribosome may dissociate rapidly, resulting in termination of translation. Further experiments would be required to investigate the stability of ribosomal complexes containing deacylated tRNA molecules. Alternatively, if the ester hydrolysis occurs after translocation, a prolyl-tRNA may also be in the A-site when cleavage occurs. Such a prolyl-tRNA would not have a peptidyl-tRNA electrophile to attack in the peptidyl transferase reaction and would remain an aminoacyl-tRNA. Whether a deacylated tRNA or prolyl-tRNA occupies the A/P site, the efficiency of the translocation reaction may be impaired by the unusual nature of the tRNA in this position and the stability of such a complex may be low.

We propose that the 2A region is, however, not only capable of terminating protein synthesis after the 2A region, but can also cause an *apparent* proteolytic cleavage at the C-terminus of the 2A region by allowing synthesis to proceed past the "cleavage" site in a proportion of cases. Since the 2A region is only 18 amino acids in length and is inactive post-translationally it is unlikely that it possesses a separate proteolytic activity. It is more probable that the same activity which results in termination of translation also causes this apparent proteolysis. This would imply that the nascent 2A region again hydrolyses the ester linkage to its tRNA after translation of its C-terminal glycine residue, but on this occasion the ribosome continues to translate the next proline codon and the rest of the encoded protein. Perhaps the final outcome of the proposed ester hydrolysis - termination or apparent cleavage - is determined by the relative kinetics of the translocation reaction and the 2A ester hydrolysis reaction, thereby determining the ribosomal state in which cleavage occurs and the stability of the resultant ribosomal complex containing the deacylated tRNA. Alternatively, the relative rates of complex dissociation and the translocation of the next prolyl-tRNA may be in competition.
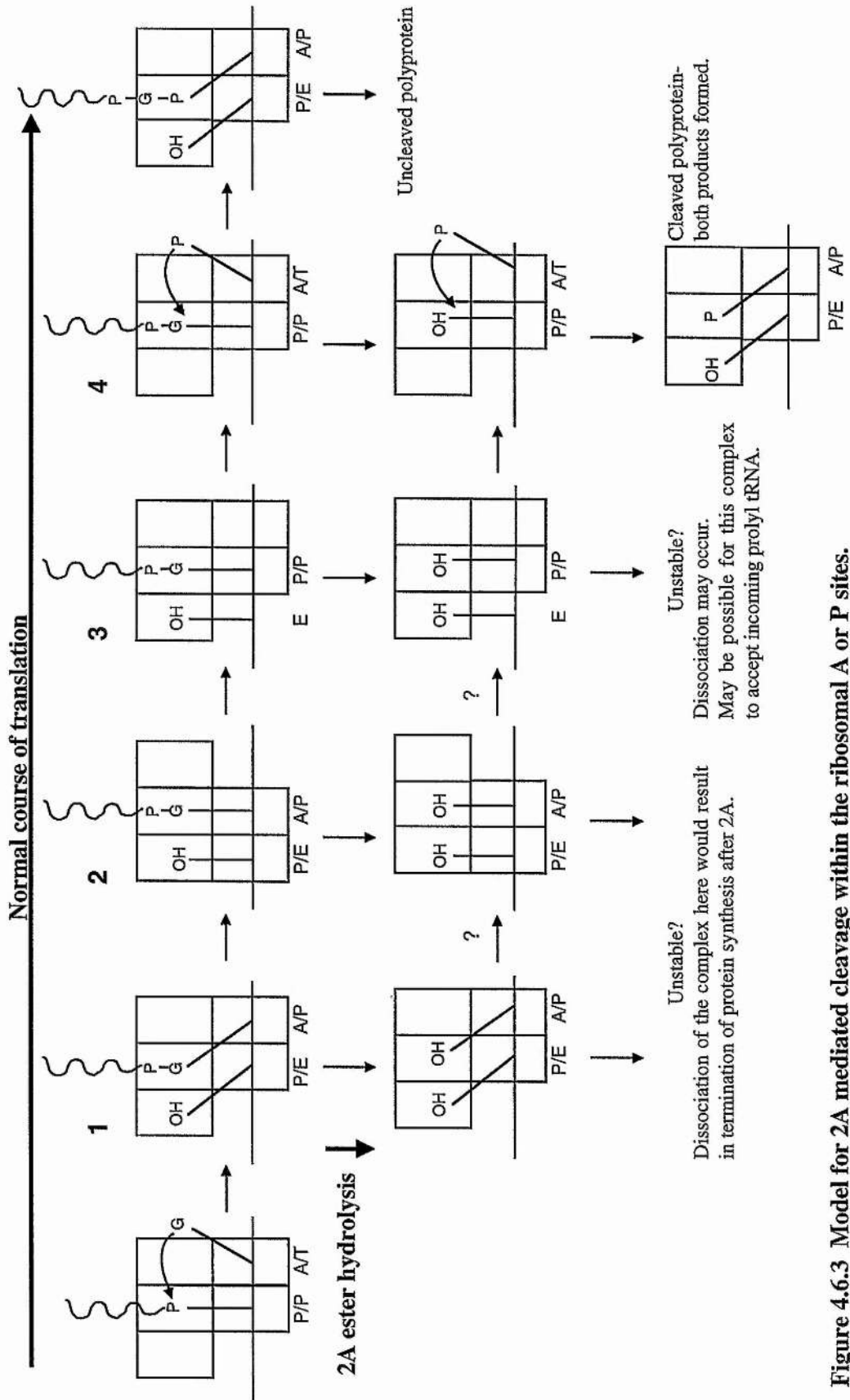
193

Figure 4.6.3 Model for 2A mediated cleavage within the ribosomal A or P sites.

194

Figure 4.6.3 demonstrates the outcomes of ester hydrolysis occurring in each of the ribosomal states between the peptidyl transferase reaction to form the peptide bond to the amino group of glycine and the peptidyl transferase reaction to form the next peptide bond to proline. In a small percentage of translating polyproteins containing 2A neither "cleavage" nor termination occurs and translation continues through and after 2A as normal. The percentage of non-cleaving polyproteins increased with the introduction of mutations into the 2A region (see Section 3.4). Presumably this was due to a slowing of the ester hydrolysis reaction such that the attack by the incoming prolyl-tRNA usually occurred and thus no cleavage products were observed. The formation of uncleaved polyprotein *in vitro* must have been an artifact of the experimental system since a virus would not be viable if "cleavage" did not occur. This was probably caused by the suboptimal protein context of the 2A peptide and a decrease in the concentration of the translation system components with respect to *in vivo* systems.

For the ester hydrolysis to occur before translocation of the peptidyl-tRNA was complete (in the states labelled 1 and 2 in Figure 4.6.3), the hydrolysis reaction would have to be very fast. Mutation of the P1'-proline to other amino acids would not be expected to effect cleavage activity, if it occurred at this stage, since this proline residue would not yet be within the ribosome. Mutation of this proline residue however led to almost complete abolition of the "cleavage" reaction and only the products resulting from termination after 2A or normal protein synthesis were seen. Since the termination of protein synthesis still occurs in constructs bearing a mutation of the P1'-proline, it is still conceivable that the hydrolysis reaction which results in termination of translation occurs at this stage. This would proceed *via* deacylated tRNAs in both A/P and P/E sites of the ribosome. It is not known whether translocation of such a ribosomal complex would occur, nor whether this would be a stable state. It seems possible that hydrolysis of the ester at this stage in the elongation cycle might have led to an unstable ribosomal state which could dissociate. Equally plausible is the theory that translocation of such unusual tRNA species may have

led to incorrect positioning of the mRNA in the A-site, and the early occurrence of a stop codon resulting in termination of translation.

Hydrolysis of the ester was also possible in the post-translocation stage but before entry of the next aminoacyl tRNA (state 3 of Figure 4.6.3). This is perhaps the most likely state for the ribosome to exist in when hydrolysis occurs since the following step is the rate limiting step in the elongation cycle. The result of this would have been a deacylated tRNA within the P/P site. Again, it is unknown whether this would be a stable situation, but it does not seem improbable that the prolyl-tRNA may then be able to enter the A-site, and upon dissociation of the eEIF-1α.GDP complex from the ribosome, the eEF-2.GTP complex may be able to translocate the prolyl-tRNA to the P-site. Therefore if the product of hydrolysis was stable and transloction was allowed then this would have led to cleavage of the polyprotein after the 2A region. On the other hand a resulting unstable complex would have led to the termination of translation after 2A.

The P1'-proline mutation data once again appears difficult to explain if the cleavage reaction is occurring at this stage in the elongation cycle. However it could be explained by examining the P1'-proline codon usage within both the rabbit (*Oryctolagus cuniculus*) and wheat (*Triticum aestivum*) systems. Firstly the wild-type sequence contains the codon CCC in aphthoviruses and CCC or CCT in cardioviruses. In the natural aphthovirus host (*Bos taurus*) CCC is used for one third of proline residues. In *E. coli*, which does not support 2A-mediated activity, CCC and CCT occur in 12 % and 17 % of codons encoding proline, and thus one might expect translation of this codon to be retarded . In *E. coli*, however, it seems that a structural ribosomal difference does not support activity of the 2A region. In wheatgerm extract the occurrence of these codons (CCC - 19 %, CCT - 17 %) is much lower than for either the rabbit (CCC - 39 %, CCT - 24 %) or cow (CCC - 36 %, CCT - 26 %). This may be explain the increase in termination in translation at the C-terminus of 2A in the wheatgerm system.

Hydrolysis of the ester bond whilst a prolyl-tRNA occupies the A-site but before it has taken part in the peptidyl transferase reaction (state 4 of Figure 4.6.3) would be most likely option from the model to be capable of producing both cleavage products. This stage of translation would be most affected by the nature of the incoming nucleophile and, hence, mutation of the proline to any other less hindered nucleophile would result in faster formation of the peptide bond and non-cleavage of the polyprotein.

The existence of three outcomes (cleavage, termination and non-cleavage), from what is probably a single 2A-mediated activity, suggests that the activity of 2A may be influenced by the normal activity of the ribosome. The termination of protein synthesis at the 2A region of EMCV within fractionated Krebs-2 cell free extracts and the restoration of translation of 2BC by the addition of eEF-2 suggests that the action of 2A resulting in termination of synthesis is in competition with that resulting in apparent cleavage (Svitkin and Agol, 1983). Elongation factor eEF-2, which increases the rate of translocation (thereby increasing the likelihood of the ribosome existing in a post-translocation state), enhances the "cleavage" reaction. This experiment therefore seems to indicate that in the pre-translocation state if the ester bond between the nascent peptide and the tRNA is hydrolysed termination of translation occurs. Hydrolysis of the ester bond whilst in the post-translocation state (3 or 4 in Figure 4.6.3) results in an *apparent proteolytic cleavage* with the nascent protein being released and the ribosome continuing to translate the next proline codon and the rest of the encoded protein. Alternatively it may indicate that the state of the ribosomal complex after the proposed 2A hydrolysis of the ester linkage is unstable, and rapid translocation of either the deacylated tRNA or prolyl-tRNA from the A/P state is required for continued protein synthesis.

The final outcome of 2A activity for a polyprotein may then be dependent on the relative rates of the 2A ester hydrolysis reaction, the peptidyl transferase reaction and the rate of translocation of the peptidyl-tRNA.

The 2A RNA sequence cannot be directly involved in the hydrolysis reaction since it is localised in the small ribosomal subunit far from the 3'-end of the tRNA. However the sequence following 2A clearly has some effect on 2A activity (*re.* the differences between mutations in pMD2 and pCAT2AGUS). This effect may be to pause translation and thereby make non-cleavage less likely. However further work is required to elucidate the role of the mRNA.

Since the 2A region is acting in close association with the ribosome in this model, mutations within the 2A sequence which disrupt "cleavage" may be either directly disrupting 2A activity, or, interfering with interactions between the peptide and the ribosome which are required to correctly orient the 2A peptide for "2A-mediated cleavage". The site directed mutational analysis of 2A therefore revealed little regarding the precise mechanism of 2A action at a molecular level. However, we would postulate that the large dipole moment caused by the α-helical conformation of the nascent peptide may enable a main-chain carbonyl-oxygen to activate a water molecule to attack the ester bond linking the glycine residue to its peptidyl-tRNA. This would explain why insertions within the α-helix, which disrupted the stabilising side-chain / side-chain interactions, destroyed 2A cleavage activity. Mono- and divalent cations are required for the elongation steps of translation and have been implicated in the peptidyl transferase reaction (Moldave, 1985; Noller *et al.*, 1992). Such cations might also play a role in the 2A-cleavage reaction.

The function of the RNA sequence following the 2A region may have been to slow the ribosome at the C-terminus of 2A. Interestingly, the reaction of peptidyl-tRNA with 2'(3')-O-L-glycyladenosine and 2'(3')-O-L-prolyladenosine molecules, catalysed by peptidyl transferase, is slow compared to reactions with other natural 2'(3')-O-L-aminoacyl adenosine molecules (Rychlik *et al.*, 1970). This may also indicate that the movement of the ribosome is slowed during translation of the glycine and proline residues flanking the 2A-cleavage site.

Section 4.7 Implications of excess capsid precursor in FMDV?

Since the addition of wild-type 1D sequences between CAT and 2A did not rectify the imbalance in quantities of cleavage products, but did increase the efficacy of 2A-mediated cleavage it seemed possible that, in *in vitro* systems at least, the wild-type sequence of P1 might also lead to the increased production of the capsid proteins with comparison to the rest of the viral proteins.

Both transcriptional and translational regulation of viral genome expression are well documented in alphaviruses, retroviruses and RNA bacteriophages for the differential production of structural and non-structural proteins.

In alphaviruses, such as Sindbis virus, a 42S virion RNA is transcribed into 42S and subgenomic 26S mRNA. The structural proteins are encoded in this 26S mRNA. In Sinbis infected cells there is a molar excess of 26S RNA throughout the infectious cycle. Not surprisingly, therefore, virtually all the virus specific proteins detected in the infected cell are the structural proteins. The 26S RNA appears to be a transcriptional method of regulating the production of large quantities of structural proteins without equimolar production of non-structural proteins.

Translational frameshifting is used by most retroviruses and retrotransposons to express the *pol* gene encoding reverse transcriptase and integrase. Retroviruses and retrotransposons transcribe *gag* (group specific antigen - encoding the structural proteins) and *pol* genes within two separate reading frames but into a single mRNA with the *gag* gene preceding the *pol* gene. The function of retrovirus and retrotransposon proteins depends on assembly conditions which include a molar excess of *gag* protein molecules. A termination codon follows the *gag* gene such that only *gag* is normally synthesised, but occasionally frame-shifting occurs at the overlap region of the two genes and results in the translation of a *gag-pol* fusion protein. Thus, the requirement for over-expression of the structural proteins is fulfilled. A rare exception is the retrovirus Murine Leukemia

Virus which uses a nonsense suppression mechanism to overexpress *gag* (Jamjoom *et al.*, 1977). The retrotransposon TF1 exceptionally contains a single reading frame, but uses specific degradation of integrase (In) to produce a molar excess of gag to In (Levin *et al.*, 1994).

The genomes of small RNA bacteriophages, such as MS2 and R17, code for only 3 proteins: phage coat protein, maturation protein and the RNA synthetase enzyme. With such a small genome it would be very surprising if the phages were capable of exerting either transcriptional or translational control over their genome expression. However differences in the frequency of initiation of translation for the three proteins and inhibition of translation of the synthetase by the coat protein results in a ratio of 20:5:1 for coat protein : maturation protein : RNA synthetase.

Our current understanding of picornavirus polyprotein processing predicts that all picornavirus proteins should be present in equimolar quantities, although this has not been found in practice (Agol, 1980). It has been suggested that specific degradation of some viral proteins and some sort barrier to translation after P1 are responsible for these discrepancies.

The capsid of a picornavirus is composed of 60 copies of the protomer unit which is formed by the translation and processing of the P1 region. Given the recyclable nature of the viral enzymes a mechanism for over expression of their capsid proteins might represent a more efficient replication strategy. The complex processing of picornaviruses *via* alternative pathways has made it rather difficult to quantify relative proportions of viral proteins. However, the removal or inactivation of the 3C protease considerably simplifies the situation, as it has done in our experiments.

Early studies of FMDV polyprotein processing suggested that the primary processing products were formed in approximately equimolar quantities (Sangar *et al.*, 1977; Doel *et al.*, 1978). However, these studies were complicated by secondary processing by 3C[pro] and hampered by the contempary understanding of the genome order.

Close examination of evidence presented by Ryan *et al.* (1991) for 2A cleavage in the context of recombinant FMDV polyproteins, from which 3C^{pro} region was absent, revealed that consistent with our results, the products were not always produced in stoichiometric amounts. Figure 4.7.1 shows the translation products from rabbit reticulocyte reactions programmed with DNA encoding the recombinant polyproteins described in Table 4.7.1.

Both P12A-Δ3D and P12A-Δ1B1C1D2A polyproteins were cleaved by the 2A region in approximately 1:1 ratios. In contrast, the polyproteins P1P2-Δ3A, 1A1B1C-Δ2A2B2CΔ3A from both A10 and 01K serotypes, and 1B1C1D2A-2B2CΔ3A cleaved at the 2A cleavage site to yield more of the amino-terminal product than the C-terminal product. It appears notable that when the 2A region was followed by the 2B sequence, whether or not it was from the same serotype, then there was a very apparent difference in the quantities of the two cleavage products.

"+ P" Indicates incubation of the translation products with processing extract from FMDV-infected cells.
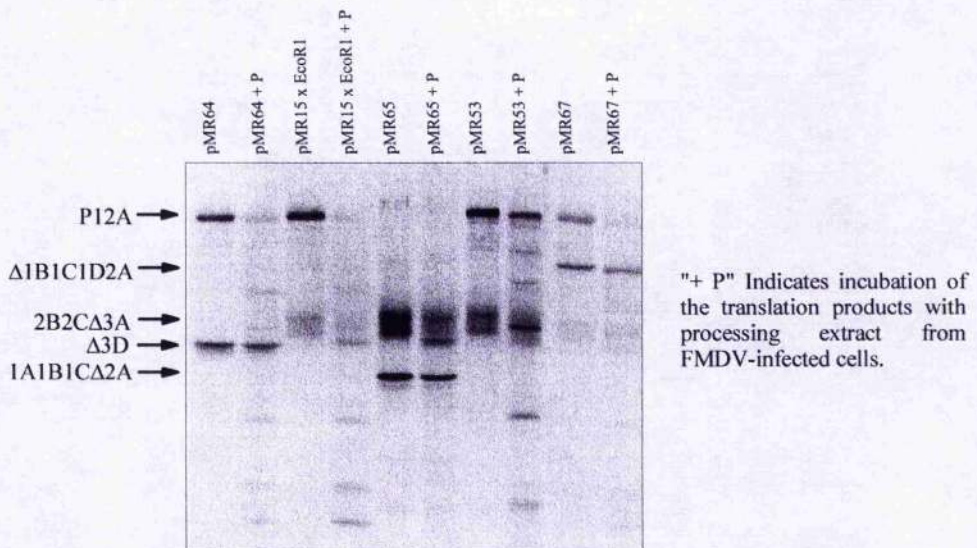
Figure 4.7.1 Translation products of rabbit reticulocyte lysates programmed with pMR transcripts (Ryan *et al.*, 1991). The FMDV protein products encoded were; pMR64 - P12AΔ3D, pMR15 x EcoRI - P1P2Δ3A, pMR65 - L1A1B1CΔ2A2B2CΔ3A, pMR53 - LP12A2B2CΔ3A and pMR67 - P12AΔ1B1C1D2A2B2CΔ3A.

Plasmid	Polyprotein encoded	Number of methionines			Qualitative densitometric analysis
		1st product	2nd product	3rd product	
pMR64	LP12AAΔ3D	14	12	–	P12A≈Δ3D
pMR15xEcoRI	LP1P2Δ3A	14	14	–	P12A»2B2CΔ3A
pMR65	L1A1B1CA2A2B2CΔ3A	6	14	–	1A1B1CΔ2A>2B2CΔ3A
pMR53	LP12A2B2CΔ3A	14	14	–	P12A»2B2C
pMR67	LP12AAΔ1B1C1D2A2B2CΔ3A	14	12	14	P12A≈Δ1B1C1D2A»2B2CΔ3A

Table 4.7.1 Analysis of the translation reactions of recombinant FMDV polyproteins. The plasmids encoding recombinant FMDV polyproteins as detailed were translated in rabbit reticulocyte lysate with [^{35}S]-methionine and an analysis by eye of the darkness of the bands representing cleavage products was compared to the number of methionines in each product. Sequences derived from serotype 01K are shown in black, whereas those from serotype A10 are shown in blue.

In recent experiments by Hahn and Palmenberg (1996) the 2A mediated cleavage of EMCV was studied using site directed mutagenesis of a recombinant EMCV polyprotein, Δ1C1D2A2BΔ2C. The appearance of both [^{35}S]-methionine labelled cleavage products, Δ1C1D2A and 2BΔ2C, was studied. The products would be expected to contain 5 and 7 methionines, respectively, and bearing this in mind they appeared, by eye, to have been formed in the ratio of 2 : 1. Unexpectedly, they also found that, when supposedly inactivating mutations (P1-G->A, P3-N->K, and P7D->N) were inserted in the full length genome, translation of the genome resulted in a product co-migrating with L-P1-2A. It was suggested that this might be as a result of spurious 3C^{pro} mediated cleavage near the 2A cleavage site or the co-migration of a different product of similar size. However taking into account the evidence presented in this thesis and obtained from the paper of Ryan *et al.*, it seems likely that this product is evidence of a 2A specific effect in which an interruption in translation occurs after the 2A region.

Further evidence of an excess of picornavirus capsid precursor polyprotein is seen in the work of Svitkin and Agol (1983). Here fractionated cell extracts, deficient in an elongation factor, eEF2, were programmed with EMCV RNA. A barrier to translation within the P2 region was observed which resulted in the inability to synthesis polypeptide 2BC or the other non-structural proteins. Addition of eEF-2 then restored the synthesis of 2BC and P3.

If our hypothesis is correct it appears, therefore, that aphtho- and cardioviruses may, indeed, be capable of synthesising an excess of their structural proteins over non-structural proteins, which may well provide a more efficient replication strategy. Indeed, the relative proportions of termination of translation or "cleavage" occuring after 2A and may vary throughout infection depending upon the "translational status" of the cell. This would provide a very subtle and elegant means of controlling the relative levels of structural and non-structural proteins; an ideal situation being that early in infection equal quantities of structural and replicative proteins may be formed, whereas late in infection

an excess of capsid proteins could be synthesised. If this was the case then the aphtho- and cardioviruses, being capable exerting such a fine regulation, could be regarded as more "advanced" than the entero- and rhinoviruses which have no such means of control.

Section 4.8 Concluding Remarks

On commencing this work it was thought that the 2A region was mediating its effect *via* some novel mechanism of proteolysis. This was a entirely reasonable hypothesis since it had been shown, to date, that all of the processing of picornavirus polyproteins was mediated by virus-encoded proteases. However, a detailed characterisation of the *in vitro* translation products of constructs encoding 2A within a reporter gene cassette has demonstrated that the mechanism of 2A-mediated cleavage cannot be *via* proteolysis. A model by which the 2A region could exert its effect has been presented here, and is supported by the site-directed mutagenesis studies, densitometric analyses, and reports of anomalies in the literature. I have proposed a new hypothesis which implicates the 2A region in the hydrolysis of the ester bond between itself and the tRNA, before the formation of the "cleaved" peptide bond. Further experiments to investigate the strength of this model by manipulation of the conditions of translation *in vitro* are required. However, if this new model is supported, many exciting experiments can be envisioned to investigate this completely novel method of viral polyprotein processing and its consequences on viral replication.

Appendix

I Additional samples of the translation products of pCATGUS and pCAT2AGUS in rabbit reticulocyte lysate (shown below) were required for the calculation of mean values for internal initiation and percentage of cleavage.
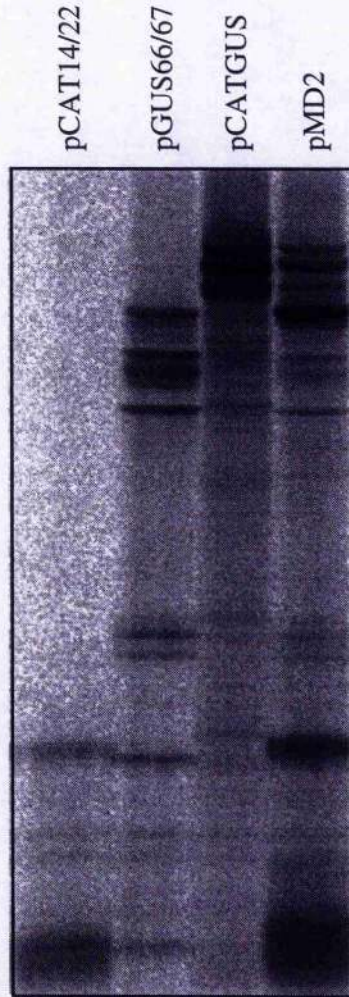
References

Agol, V. I. (1980) Structure, translation, and replication of picornaviral genomes., *Prog. Med. Virol. 26*, 119-157.

Allaire, M., Chernaia, M. M., Malcolm, B. A., & James, M. (1994) Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases., *Nature 369*, 72-76.

Anteunis, M. J. O., Borremans, F. A. M., Stewart, J. M., & London, R. E. (1981) 360-MHz ^{1}H NMR Conformational analysis of Gly-Pro-X peptides (X=Ala, Cha, Phe)., *J. Am. Chem. Soc. 103*, 2187-2191.

Bachrach, H. L., Swaney, J. B., & Vande Woude, G. F. (1973) Isolation of the structural polypeptides of foot-and-mouth disease virus and analysis of their C-terminal sequences., *Virology 52*, 520-528.

Batson, S., & Rundell, K. (1991) Proteolysis at the 2A/2B junction in Theiler's murine encephalomyelitis virus., *Virology 181*, 140-150.

Bazan, J. F., & Fletterick, R. J. (1988) Viral cysteine proteases are homologous to the trypsin-like serine proteases:Structural and functional implications., *Proc. Natl. Acad. Sci. USA 85*, 7872-7876.

Beadle, G. W., & Tatum, E. L. (1941) Genetic control of biochemical reactions in *Neurospora.*, *Proc. Natl. Acad. Sci. USA 27*, 499-506.

Bernstein, H. D., Sonenberg, N., & Baltimore, D. (1985) Poliovirus mutant that does not selectively inhibit host cell protein synthesis., *Mol. Cell. Biol. 60*, 1040-1049.

Bischoff, R., & Kolbe, H. V. J. (1994) Deamidation of asparagine and glutamine residues in proteins and peptides - structural determinants and analytical methodology., *Journal Of Chromatography B Biomedical Applications 662*, 261-278.

Blobel, G., & Sabatini, D. D. (1970) Controlled proteolysis of nascent polypeptides in rat liver cell fractions., *J. Cell. Biol. 45*, 130-145.

Blodgett, J. K., Loudon, G. M., & Collins, K. D. (1985) Specific cleavage of peptides containing an aspartic acid (β-hydroxamic acid) residue., *J. Am. Chem. Soc. 107*, 4305-4313.

Böck, A., Wirth, R., Schmid, G., Schumacher, G., Lang, G., & Buckel, P. (1983a) The two subunits of penicillin acylase are processed from a single precursor., *FEBS Lett. 20*, 141-144.

Böck, A., Wirth, R., Schmid, G., Schumacher, G., Lang, G., & Buckel, P. (1983b) The penicillin acylase from Escherichia coli ATCC11105 consists of two dissimilar subunits., *FEMS Microbiol. Lett. 20*, 135-139.

Bonner, D. M., DeMoss, J. A., & Mills, S. E. (1965) The evolution of an enzyme. In: *Evolving genes proteins* , Symp. Rutgers, State University, pp 305-318.

Brandts, J. F., Halvorson, H. R., & Brennan, M. (1975) Consideration of the possibility that the slow step in protein denaturation is due to *cis-trans* isomerism in proline residues., *Biochemistry 14*, 4953-4963.

Bryan, P., Pantoliano, W., Quill, S. G., Hsiao, H.-Y., & Poulos, T. (1986) Site-directed mutagenesis and the role of the oxyanion hole in subtilisin., *Proc. Natl. Acad. Sci. USA 83*, 3742-3745.

Capasso, S., Mazzarella, L., Sica, F., Zagari, A., & Salvadori, S. (1993) Kinetics and mechanism of succinimide ring formation in the deamidation process of asparagine residues., *J. Chem. Soc.; Perkin Trans. II*, 679-682.

Chen, H.-H., Kong, W.-P., Zhang, L., Ward, P. L., & Roos, R. P. (1995) A picornaviral protein synthesized out of frame with the polyprotein plays a key role in a virus-induced immune mediated demyelinating disease., *Nature Medicine 1*, 927-931.

Chernaia, M. M., Malcolm, B. A., Allaire, M., & James, M. N. G. (1993) Hepatitis A Virus 3C Proteinase: Some properties, crystallisation and preliminary crystallographic characterisation., *J. Mol. Biol. 234*, 890-893.

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

Clarke, B. E., & Sangar, D. V. (1988) Processing and assembly of foot-and-mouth disease virus proteins using subgenomic RNA., *J. Gen. Virol. 69*, 2313-2325.

Clarke, S. (1987) Propensity for spontaneous succinamide formation from aspartyl and asparaginyl residues in cellular proteins., *Int. J. Pept. and Prot. Res. 30*, 808-821.

Cooper, A. A., Chen, Y. J., Lindorfer, M. A., & Stevens, T. H. (1993) Protein splicing of the yeast TFP1 intervening protein-sequence - a model for self-excision., *EMBO J. 12*, 2575-2583.

Cooper, A. A., & Stevens, T. H. (1993) Protein splicing - excision of intervening sequences at the protein level., *Bioessays 15*, 667-674.

Cooper, A. A., & Stevens, T. H. (1995) Protein splicing - self-splicing of genetically mobile elements at the protein level., *TIBS 20*, 351-356.

Cutting, S., Zheng, L., & Losick, R. (1991) Gene encoding two alkali-soluble components of the spore coat from *Bacillus subtillis.*, *J. Bacteriol. 173*, 2915-2919.

Dasso, M. C., & Jackson, R. J. (1989) On the fidelity of mRNA translation in th nuclease-treated rabbit reticulocyte lysate system., *Nucleic Acids Res. 17*, 3129-3144.

Davies, D. R. (1990) The structure and function of the aspartic proteinases., *Ann. Rev. Biophys. Biophys. Chem. 19*, 189-215.

Davis, E. O., Sedgwick, S. G., & Colston, M. J. (1991) Novel structure of the *recA* locus of *Mycobacterium tuberculosis* implies processing of the gene product., *J. Bacteriol. 173*, 5653-5662.

Davis, E. O., Jenner, P. J., Brooks, P. C., Colston, M. J., & Sedgwick, S. G. (1992) Protein splicing in the maturation of mycobacterium-tuberculosis recA protein - a mechanism for tolerating a novel class of intervening sequence., *Cell 71*, 201-210.

Dean, C., Favreau, M., Tamaki, S., Bond-Nutter, D., Dunsmuir, P., & Bedbrook, J. (1988) Expression of tandem gene fusions in transgenic tobacco plants., *Nucleic Acids Res. 16*, 7601-7617.

Devaney, M. A., Vakharia, V. N., Lloyd, R. E., Ehrenfeld, E., & Grubman, M. J. (1988) Leader protein of foot-and-mouth disease is required for cleavage of the p220 component of the cap-binding protein complex., *J. Virol. 62*, 4407-4409.

Doel, T. R., Sangar, D. V., Rowlands, D. J., & Brown, F. (1978) A re-appraisal of the biochemical map of foot-and-mouth disease virus RNA., *J. Gen. Virol. 41*, 395-404.

Domogatsky, S. P., Vlasik, T. N., Seryakova, T. A., Ovchinnkov, L. P., & Spirin, A. S. (1978) Difference in RNA-binding ability between eukaryotic and prokaryotic elongation factors of translation., *FEBS Lett. 96*, 207-215.

Donaghy, G. & Jupp, R. (1995) Characterisation of the Epstein-Barr virus proteinase and comparison with the human cytomegalovirus proteinase., *J. Virol. 69*, 1265-1270.

Donnelly, M. L. L., Gani, D., Flint, M., Monaghan, S., & Ryan, M. D. (1997) The cleavage activities of aphtho- and cardiovirus 2A proteins., *J. Gen. Virol. 78*, 13-21.

Dougherty, W. G., & Semler, B. L. (1993) Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes., *Microbiol. Rev. 57*, 781-822.

Douglass, J., Civelli, O., & Herbert, E. (1984) Polyprotein gene-expression - generation of diversity of neuro- endocrine peptides., *Ann. Rev. Biochem. 53*, 665-715.

Duncan, K., Edwards, R. M., & Coggins J. R. (1987) The pentafunctional *arom* enzyme of *Saccharomyces cerevisiae* is a mosaic of monofunctional domains., *Biochem. J. 246*, 375-386

Eberhardt, E. S., Loh, S. N., & Raines, R. T. (1993) Thermodynamic origin of prolyl peptide bond isomers., *Tet. Lett. 34*, 3055-3056.

Fitzgerald, P. M. D. (1993), HIV protease-ligand complexes., *Curr. Opin. Struct. Biol. 3*, 868-874.

Flint, M. (1995) A study of foot-and-mouth disease virus polyprotein processing., Ph.D. Thesis, Department of Molecular Biology, University of Reading, Reading.

Frank, J. (1997) The ribosome at higher resolution - the donut takes shape., *Curr. Opin. Struct. Biol. 7*, 266-272.

Geiger, T., & Clarke, S. (1987) Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides - succinimide-linked reactions that contribute to protein-degradation., *J. Biol. Chem. 262*, 785-794.

Gessert, S. F., Kim, J. H., Nargang, F. E., & Weiss, R. L. (1994) Polyprotein precursor of two mitochondrial-enzymes in neurospora-crassa - gene structure and precursor processing., *J. Biol. Chem. 269*, 8189-8203.

Gidoni, D., Bond-Nutter, D., Brosio, P., Jones, J., Bedbrook, J., & Dunsmuir, P. (1988) Coordinated expression between two photosynthetic petunia genes in transgenic plants., *Mol. Gen. Genet. 211*, 507-514.

Gorbalenya, A. E., Donchencko, A. P., Blinov, V. M., & Koomin, E. V. (1989) Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases., *FEBS Lett. 243*, 103-114.

Gorbalenya, A. E., Koomin, E. V., & Lai, M. M.-C. (1991) Putative papain-related thiol protease of positive strand RNA viruses: identification of rubi- and aphthovirus proteases and delinaeation o a novel conserved domain associated with proteases of rubi- alpha- and coronaviruses., *FEBS Lett. 288*, 201-205.

de Groot, R. J., Hardy, W. R., Shirako, Y., & Strauss, J. H. (1990) Cleavage-site preferences of Sindbis virus polyproteins containing the non-structural proteinase. Evidence for temporal regulation of polyproteein processing *in vivo.*, *EMBO J. 9*, 2631-2638.

Grubman, M. J., Zellner, M., Bablanian, G., Mason, P. W., & Piccone, M. E. (1995) Identification of the active-site residues of the 3C proteinase of Foot-and Mouth Disease Virus., *Virology 213*, 581-9.

Gu, H. H., Xu, J., Gallagher, M., & Dean, G. E. (1993) Peptide splicing in the yeast vacuolar ATPase subunit A from *Candida tropicalis.*, *J. Biol. Chem. 268*, 7372-7381.

Hahn, H., & Palmenberg, A. C. (1996) Mutational analysis of the encephalomyocarditis virus primary cleavage., *J. Virol. 70*, 6870-6875.

Harris, J. I., & Roos, P. (1956) Amino-acid sequence of melanophore stimulating peptide., *Nature 178*, 90.

Hellen, C. U. T., Krausslich, H.-G., & Wimmer, E. (1989) Proteolytic processing of polyproteins in the replication of RNA viruses., *Biochemistry 28*, 9881-90.

Hellen, C. U. T., Facke, M., Krausslich, H. G., Lee, C. K., & Wimmer, E. (1991) Characterization of poliovirus-2A proteinase by mutational analysis - residues required for autocatalytic activity are essential for induction of cleavage of eukaryotic initiation factor-4F polypeptide p220., *J. Virol. 65*, 4226-4231.

Hellen, C. U. T., & Wimmer, E. (1992) The role of proteolytic processing in the morphogenesis of virus- particles., *Experientia 48*, 201-215.

Hodges, R. A., Perler, F. B., Noren, C. J., & Jack, W. E. (1992) Protein splicing removes intervening sequences in an archaea dna- polymerase., *Nucleic Acids Res. 20*, 6153-6157.

Hooper, N. M. (1994) Families of zinc metalloproteases., *FEBS Lett. 354*, 1-6.

Hsiao, K.-C., Cheng, C.-H., Fernandes, I. E., Detrich, H. W., & DeVries, A. L. (1990) An antifreeze glycopeptide gene from the antarctic cod *Notothenia coriceps neglecta* encodes a polyprotein of high peptide copy number., *Proc. Natl. Acad. Sci. USA 87*, 9265-9269.

Jackson, R. J. (1986) A detailed kinetic analysis of the in vitro synthesis and processing of encephalomyocarditis virus products., *Virology 149*, 114-127.

Jacobson, M. F., & Baltimore, D. (1968) Morphogenesis of poliovirus., *J. Mol. Biol. 33*, 369-78.

James, M. N. G., & Sielecki, A. R. (1984) Stereochemical analysis of peptide bond hydrolysis catalysed by the aspartic proteinase penicillopepsin., *Biochemistry 24*, 3701-3713.

Jamjoom G. A., Naso, .B., Arlinghaus, R.B. (1977) Further characterisation of intracellular precursor polyproteins of Rauscher leukemia virus., *Virology 78*, 11-34.

Jia, X.-Y., Ehrenfeld, E., & Summers, D. F. (1991) Proteolytic activity of Hepatitis A virus 3C protein., *J. Virol. 65*, 2595-2600.

Jore, J., De Geus, B., Jackson, R. J., Pouwels, P. H., & Enger-Valk, B. E. (1988) Poliovirus protein 3CD is the active protease for processing of the precursor protein P1 *in vitro., J. Gen.Virol. 69*, 1627-1636.

Jorgensen, W. L., & Gao, J. (1988) Cis-trans energy difference for the peptide bond in the gas phase and in aqueous solution., *J. Am. Chem. Soc. 110*, 4212-4216.

Juvvadi, P., Dooley, D. J., Humblet, C. C., Lu, G. H., Lunney, E. A., Panek, R. L., Skeean, R., & Marshall, G. R. (1992) Bradykinin and angiotensin II analogs containing a conformationally constrained proline analog., Int. J. Pept. Prot. Res. *40*, 163-170.

Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M., & Stevens, T. H. (1990) Protein splicing converts the yeast tfp1 gene-product to the 69-kd subunit of the vacuolar H+-adenosine triphosphatase., *Science 250*, 651-657.

Kaziro, Y. (1978) The role of guanosine 5'-triphosphate in polypeptide chain elongation., *Biochim. Biophys. Acta. 505*, 95-127.

van Kleef, F. S. M., de Jong, W. W., & Hoenders, H. J. (1975) Stepwise degradation and deamidation of eye lens protein - α crystalin in aging., *Nature 258*, 264-266.

Kleina, L. G., & Grubman, M. J. (1992) Antiviral effects of a thiol protease inhibitor on foot-and-mouth disease virus., *J. Virol. 66*, 7168-7175.

Klotz, A. V., & Thomas, B. A. (1993) N^5-Methylasparagine and asparagine as nucleophiles in peptides: main-chian vs side-chain amide cleavage., *J. Org. Chem. 58*, 6985-6989.

Klump, H., Auer, H., Liebig, H. D., Kuechler, E., & Skern, T. (1996) Proteolytically active 2A proteinase of human rhinovirus-2 is toxic for *saccharomyces-cerevisiae* but does not cleave the homologs of eIF- 4-gamma *in-vivo* or *in-vitro., Virology 220*, 109-118.

Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A. F., Scolnick, E. M., Sigal, I. S. (1988) Active human immunodeficiency virus protease is required for viral infectivity., *Proc. Natl. Acad. Sci. USA 85*, 4686-4690.

Kollman, P. A. (1992) Theory of enzyme mechanisms., *Curr. Opin. Struct. Biol. 2*, 765-761.

Konig, H., & Rosenwirth, B. (1988) Purification and partial charactrisation of poliovirus protease 2A by means of a functional assay., *J. Virol. 62*, 1243-1250.

Kossiakoff, A. A. (1988) Tertiary structure is a principal determinant to protein deamidation, *Science 240*, 191-194.

Krausslich, H. G., & Wimmer, E. (1988) Viral proteinases., *Ann. Rev. Biochem. 57*, 701-754.

Kraut, J. (1977) Serine proteases: structure and mechanism of catalysis., *Ann. Rev. Biochem. 46*, 331-358.

Kuchler, K., Richter, K., Trnovsky, J., Egger, R., & Kreil, G. (1990) Two precursors of thyrotropin-releasing hormonefrom skin of *Xenopus laevis.*, *J. Biol. Chem. 265*, 11731-11733.

Kurjan, J., & Herskowitz, I. (1982) Structure of a yeast pheromone gene (MFα): a putative α-factor precursor contains four tandem copies of mature α-factor., *Cell 30*, 933-943.

Kurz, C., Forss, S., Küpper, H., Strohmaier, K., & Schaller, H. (1981) Nucleotide sequence and corresponding amino acid sequence of the gene for the major antigen of foot-and-mouth disease virus., *Nucleic Acids Res. 9*, 1919-1931.

Law, K. M., & Brown, T. D. K. (1990) The complete nucleotide sequence of the GDVII strain of Theiler's murine encephalomyelitis virus (TMEV)., *Nucleic Acids Res. 18*, 1919-1931.

Lawson, M. A., & Semler, B. L. (1991) Poliovirus thiol proteinase 3C can utilize a serine nucleophile within the putative catalytic triad., *Proc. Natl. Acad. Sci. USA 88*, 9919-9923.

Lawson, M. A., & Semler, B. L. (1992) Alternate poliovirus nonstructural protein processing generated by primary sites of 3C proteinase cleavage., *Virology 191*, 309-320.

Lea, S., Hernández, J., Blakemore, W., Brocchi, E., Curry, S., Domingo, E., Fry, E., Abu-Ghazaleh, R., King, A., Newman, J., Stuart, D., & Mateu, M. G. (1994) The structure and antigenicity of a type C foot-and-mouth disease virus., *Structure 2*, 123-139.

Lee, C.-K., & Wimmer, E. (1988) Proteolytic processing of poliovirus polyprotein: elimination of 2A^{pro}-mediated, alternative cleavage of polypeptide 3CD by *in vitro* mutagenesis., *Virology 166*, 405-414.

Li, C. H., & Chung, D. (1976) Isolation and structure of an untriakontapeptide with opiate activity from camel pituitary glands., *Proc. Natl. Acad. Sci. USA 73*, 1145-1148.

Li, F., Browning, G. F., Studdert, M. J., & Crabb, B. S. (1996) Equine rhinovirus-1 is more closely-related to foot-and-mouth-disease virus than to other picornaviruses., *Proc. Natl. Acad. Sci. USA 93*, 990-995.

Lim, V. I., & Spirin, A. S. (1986) Stereochemical analysis of ribosomal transpeptidation: conformation of nascent peptide., *J. Mol. Biol. 188*, 565-577.

Ling, N., Burgus, R., & Guillemin, R. (1976) Isolation, primary structure, and synthesis of α-endorphin and γ-endorphin, two peptides of hypothalamic-hypophysial origin with morphinomimetic activity., *Proc. Natl. Acad. Sci. USA 73*, 3942-3946.

Lohmann, V., Koch, J. O. & Bartenschlager, R. (1996) Processing of the hepatitis-C virus proteins., *J. Hepatology 24*, 11-19.

London, R. E., & Stewart, J. M. (1979) Carbon-13 NMR spectroscopy of [20%-1,2-^{13}C$_2$-Gly^6]Bradykinin. Role of serine in reducing structural heterogeneity., *J. Am. Chem. Soc. 101*, 2455-2462.

MacArthur, M. W., & Thornton, J. M. (1991) Influence of proline residues on protein conformation., *J. Mol. Biol. 218*, 397-412.

Malkin, L. I., & Rich, A. (1967) Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal sheilding., *J. Mol. Biol. 26*, 329-346.

di Marzo Veronese, F., Copeland, T. D., Oroszlan, S., Gallo, R. C., & Sarngadharan, M G. (1988) Biochemical and immunological analysis of human immunodeficiency disease virus *gag* gene products p17 and p24., *J. Virol. 62*, 795-801.

215

Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, R., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., Gribskov, C. L., & Worland, S. (1994) Structure of a Human Rhinovirus 3C Protease reveals a trypsin-like polypeptide fold, RNA binding site, and means for cleaving precursor polyprotein., *Cell* 77, 761-71.

Mattion, N. M., Harnish, E. C., Crowley, J. C., & Reilly, P. A. (1996) Foot-and-mouth disease virus 2A protease mediates cleavage in attenuated Sabin 3 poliovirus vectors engineered for delivery of foreign antigens., *J. Virol.* 70, 8124-8127.

McDonald, C. K., & Kuritzkes, D. R. (1997) Human immunodeficiency virus type 1 protease inhibitors., *Arch. Intern. Med.* 157, 951-959.

Medina, M., Domingo, E., Brangwyn, J. K., & Belsham, G. J. (1993) The two species of the foot-and-mouth disease virus leader protein, expressed individually, exhibit the same activities., *Virology* 194, 355-359.

Ménard, R., Carrière, J., Laflamme, P., Plouffe, C., Khouri, H. E., Vernet, T., Tessier, D. C., Thomas, D. Y., & Storer, A. C. (1991) Contribution of the Glutamine-19 side chain to transition-state stabalisation in the oxyanion hole of papain., *Biochemistry* 30, 8924-8928.

Moldave, K. (1985) Eukaryotic protein synthesis., *Ann. Rev. Biochem.* 54, 1109-1149.

Nakanishi, S., Inoue, A., Taii, S., & Numa, S. (1977) Cell-free translation product containing corticotropin and ß-endorphin encoded by messenger RNA from anterior lobe and intermediate lobe of bovine pituitary., *FEBS Lett.* 84, 105-109.

Nakanishi, S., Teranishi, Y., Noda, M., Notake, M., Watanabe, Y., Kakidani, H., Jingami, H., & Numa, S. (1980) The protein-coding sequence of the bovine ACTH-ß-LPH precursor is split near the signal peptide region., *Nature* 287, 752-755.

Navia, M. A., Fitzgerald, P. M. D., McKeever, B. M., Leu, C., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., & Springer, J. P (1989) Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1., *Nature* 337, 615-620.

Nicklin, M. J. H., Krausslich, H., Toyoda, H., Dunn, J. J., & Wimmer, E. (1987) Poliovirus polypeptide precursors:Expression in vitro and processing by exogenous 3C and 2A proteinases., *Proc. Natl. Acad. Sci. USA 84*, 4002-6.

Noller, H. F., Hoffarth, V., & Zimniak, L. (1992) Unusual resistance of peptidyl transferase to protein extraction procedures., *Science 256*, 1416-1419.

Nomark, S., Bergström, Edlund, T., Grundström, T., Bengtåke, J., Lindberg, F. P., & Olsson, O. (1983) Overlapping genes., *Ann. Rev. Gen. 17*, 499-525.

Oh, C-S., & Carrington, J. C. (1989) Identification of essential residues in potyvirus proteinase HC-Pro by site-directed mutagenesis., *Virology 173*, 692-699.

Paetzel, M., & Dalbey, R. E. (1997) Catalytic hydroxyl / amine dyads within serine proteases., *TIBS 22*, 28-31.

Palmenberg, A. (1990) Proteolytic processing of picornaviral polyprotein., *Ann. Rev. Microbiol. 44*, 603-623.

Palmenberg, A. C., Parks, G. D., Hall, D. J., Ingraham, R. H., Seng, T. W., & Pallai, P. V. (1992) Proteolytic processing of the cardioviral P2 region: primary 2A/2B cleavage in clone derived precursors., *Virology 190*, 754-762.

Palmenberg, A. C., & Reukert, R. R. (1982) Evidence for intramolecular self-cleavage of picornaviral replicase precursors., *J. Virol. 41*, 244-249.

Pauling, L. (1948) Nature of forces between large molecules of biological interest., *Nature 161*, 707-709.

Pauling, L., & Corey, R. B. (1953) Fundamental dimensions of polypeptide chains., *Proc. Royal Soc. London B141*, 10-20.

Peach, C., & Velten, J. (1991) Transgene expression variability (position effect) of CAT and GUS reporter genes driven by linked T-DNA promoters., *Plant Mol. Biol. 17*, 49-60.

Pearl, L. H. (1987) The catalytic mechanism of aspartic proteinases., *FEBS Lett. 214*, 8-12.

Pelham, H. R. B. (1978) Translation of encephalomyocarditis virus RNA in vitro yields an active proteolytic processing enzyme., *Eur. J. Biochem. 85*, 457-462.

Percy, N., Barclay, W. E., Garcìa-Sastre, A., & Palese, P. (1994) Expression of a foreign protein by Influenza A virus., *J. Virol. 68*, 4486-4492.

Perler, F. B., Comb, D. G., Jack, W. E., Moran, L. S., Qiang, B. Q., Kucera, R. B., Benner, J., Slatko, B. E., Nwankwo, D. O., Hempstead, S. K., Carlow, C. K. S., & Jannasch, H. (1992) Intervening sequences in an archaea DNA-polymerase gene., *Proc. Natl. Acad. Sci. USA 89*, 5577-5581.

Phillips, M. A., & Fletterick, R. J. (1992) Proteases., *Curr. Opin. Struct. Biol. 2*, 713-720.

Pierce, J. G., & Parsons, T. F. (1981) Glycoprotein hormones: structure and function., *Ann. Rev. Biochem. 50*, 465-495.

Porse, B. T., Rodriguez-Fonseca, C., Leviev, I., & Garrett, R. A. (1995) Antibiotic inhibition of the movement of tRNA substrates through a peptidyl transferase cavity., *Biochem. Cell Biol. 73*, 877-885.

Reuckert, R. R. (1996) Picornaviridae and their replication. In: *Virology* (Fields, B. N., Knipe, D. M., *et al.*, Eds.) 3rd Edition, pp 609-654, Lippincott - Raven Publishers, Philadelphia.

Richter, K., Egger, R., & Kreil, G. (1986) Sequence of preprocaerulein cDNAs cloned from skin of *Xenopus laevis.*, *J. Biol. Chem. 261*, 3676-3680.

Roberts, P. J., & Belsham, G. J. (1995) Identification of critical amino acids within the foot-and-mouth disease virus leader protein, a cysteine protease., *Virology 213*, 140-146.

Robertson, B. H., Grubman, M. J., Weddell, G. N., Moore, D. M., Welsh, J. D., Fischer, T., Dowbenko, D. J., Yansura, D. G., Small, B., & Kleid, D. G. (1985) Nucleotide and amino acid sequence coding for polypeptides of foot-and-mouth disease virus type A12., *J. Virol. 54*, 651-660.

Rodnina, M. V., Savelsbergh, A., Katunin, V. I., & Wintermeyer, W. (1997) Hydrolysis of GTP by elongation factor G drives tRNA movement by the ribosome., *Nature 385*, 37-41.

Roos, R. P., Kong, W., & Semler, B. L. (1989) Polyprotein processing of Theiler's murine encephalomyelitis virus., *J. Virol. 63*, 5344-5353.

Roosien, J., Belsham, G. J., Ryan, M. D., King, A. M. Q., & Vlak, J. M. (1990) Synthesis of foot-and-mouth disease virus capsid proteins in insect cells using baculovirus expression vectors., *J. Gen. Virol. 71*, 1703-1711.

Ryan, M. D., G.J., B., & King, A. M. Q. (1989) Specificities of enzyme substrate interactions in foot-and-mouth disease virus polyprotein processing., *Virology 173*, 35-45.

Ryan, M. D., King, A. M. Q., & Thomas, G. P. (1991) Cleavage by foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence., *J. Gen. Virol. 72*, 2727-2732.

Ryan, M. D., & Drew, J. (1994) Foot-and-Mouth Disease Virus 2A oligopeptide mediated cleavage of an artificial polyprotein., *EMBO J. 13*, 928-33.

Ryan, M. D., & Flint, M. (1997) Virus-encoded proteases of the picornavirus super-group., *J. Gen. Virol. 78*, 699-723.

Rychlìk, I., Cernà, J., Chládek, S., Pulkrábek, P., Zemlicka, J. (1970) Substrate specificity of ribosomal peptidyl transferase., *Eur. J. Biochem. 16*, 136-142.

Saleski, N., Cleal, S. M., & Reynolds, E. C. (1996) Characterisation of a *Porphyromonas gingivalis* gene prtR that encodes an arginine-specific thiol protease., *Biochem. Biophys. Res. Comm. 224*, 605-610.

Samaha, R. R., Green, R., & Noller, H. F. (1995) A base pair between tRNA and 23S rRNA in the peptidyl transferase centre of the ribosome., *Nature 377*, 309-314.

Sangar, D. V., Black, D. N., Rowlands, D. J., & Brown, F. (1977) Biochemical mapping of the foot-and-mouth disease virus genome., *J. Gen. Virol. 35*, 281-297.

Sangar, D. V., Newton, S. E., Rowlands, D. J., & Clarke, B. E. (1987) All FMDV serotypes initiate protein synthesis at two seperate AUGs., *Nucleic Acids Res. 15*, 3305-3315.

Scheuchter, I., & Berger, A. (1967) On the size of the active site in proteases. I. Papain., *Biochem. Biophys. Res. Comm. 27*, 157.

Scott, A. P., Ratcliffe, J. G., Rees, L. H., London, J., Bennett, H. P. J., Lowry, P. J., & McMartin, C. (1973) Pituitary peptide., *Nature New Biol. 244*, 65-67.

Seemüller, E., Lupas, A., Stock, D., Löwe, J., Huber, R., & Baumeister, W. (1995) Proteasome from *Thermoplasma acidophilum*:, a threonine protease., *Science 268*, 579-581.

Shao, Y., Xu, M. Q., & Paulus, H. (1995) Protein splicing - characterization of the aminosuccinimide residue at the carboxyl-terminus of the excised intervening sequence., *Biochemistry 34*, 10844-10850.

Shao, Y., Xu, M. Q., & Paulus, H. (1996) Protein splicing - evidence for an n-o acyl rearrangement as the initial step in the splicing process., *Biochemistry 35*, 3810-3815.

Shub, D. A., & Goodrich-Blair, H. (1992) Protein Introns: a new home for endonucleases., *Cell 71*, 183-186.

Skern, T., Sommergruber, W., Auer, H., Volkmann, P., Zorn, M., Liebig, H.-D., Fessl, F., Llaas, D., & Kuechler, E. (1991) Substrate requirements of a human rhinovirus 2A proteinase., *Virology 181*, 46-54.

Skern, T., & Liebig, H.-D. (1994) Picornains 2A and 3C., *Methods Enzymol. 241*, 583-595.

Sommergruber, W., Zorn, M., Blaas, D., Fessl, F., Volkmann, P., Maurer-Fogy, I., Pallai, P., Merluzzi, V., Matteo, M., Skern, T., & Kuechler, E. (1989) Polypeptide 2A of Human Rhinovirus type 2: Identification as a protease and characterization by mutational analysis., *Virology 169*, 68-77.

Sommergruber, W., Casari, G., Fessl, F., Seipelt, J., & Skern, T. (1994) The 2A proteinase of human rhinovirus is a zinc-containing enzyme., *Virology 204*, 815-818.

Spirin, A. S. (1985) Ribosomal Translocation., *Prog. Nucl. Acids Res. 32*, 75-114.

Srere, P. A. (1987) Complexes of sequential metabolic enzymes., *Ann. Rev. Biochem. 56*, 89-124.

Stanway, G. (1990) Structure, function and evolution of picornaviruses., *J.Gen.Virol. 71*, 2483-2501.

Strauss, J. H., & Strauss, E. G. (1994) The alphaviruses:gene expression, replication, and evolution., *Microbiol. Rev. 58*, 491-562.

Strauss, E. G., de Groot, R. J., Levinson, R., Strauss, J. H. (1992) Identification of the active site residues in the nsP2 proteinase of sindbis virus., *Virology 191*, 932-940

Strebel, K., & Beck, E. (1986) A second protease of foot-and-mouth disease virus., *J. Virol. 58*, 893-899.

Stryer, L. (1988) *Biochemistry.*, 3rd Edition, pp 609-654, W. H. Freeman and Company, New York.

Sugana, K., Padlan, E. A., Smith, C. W., Carlson, W. D., & Davies, D. R. (1987) Binding of a reduced peptide inhibitor to the aspartic proteinase from *Rhizobium chinensis*: Implications for a mechanism of action., *Proc. Natl. Acad. Sci. USA 84*, 7009-7013.

Svitkin, Y. V., & Agol, V. I. (1983) Translational barrier in central region of Encephalomyocarditis virus genome: modulation by elongation factor 2 (eEF-2)., *Eur. J. Biochem. 133*, 145-154.

Thönymeyer, L., James, P., & Hennecke, H. (1991) From one gene to two proteins: The biogenesis of cytocromes *b* and c_1 in *Bradyrhizobium japonicum.*, *Proc. Natl. Acad. Sci. USA 88*, 5001-5005.

Thönymeyer, L., Bock, A., & Hennecke, H. (1992) Prokaryotic polyprotein processing., *FEBS Lett. 307*, 62-65.

Toyoda, H., Nicklin, M., Murray, M. G., Anderson, C. W., Dunn, J. J., Studier, F. W., & Wimmer, E. (1986) A 2nd virus-encoded proteinase involved in proteolytic processing of poliovirus polyprotein., *Cell 45*, 761-770.

Uozumi, N., Sakurai, K., Sasaki, T., Takekawa, S., Yamagata, N. T., & Udaka, S. (1989) A single gene directs synthesis of a precursor with ß- and α-amylase activities in *Bacillus polymyxa.*, *J. Bacteriol. 171*, 375-382.

Vasquez, M. (1983) Computed conformational states of the 20 naturally ocurring amino-acid residues and of the prototype residue alpha-aminobutyric acid., *Macromolecules* *16*, 1043-1049.

Voorter, C. E. M., de Haard-Hoekman, W.A., van den Oetelaar, P. J. M., Bloemendal, H., & de Jong, W. W. (1988) Spontaneous peptide bond cleavage in aging α-crystallin through a succinamide intermediate., *J. Biol. Chem. 263*, 19020-19023.

Voss, T., Meyer, R., & Sommergruber, W. (1995) Spectroscopic characterisation of rhinoviral protease 2A: Zn is essential for structural integrity., *Protein Science 4*, 2526-2531.

Wellink, J., & van Kammen, A. (1988) Proteases involved in the processing of viral polyproteins: brief review., *Arch. Virol. 98*, 1-26.

West G. (1992) *Black's Veterinary Dictionary*, 17th Edition, pp 316-319, A & C Black (publishers) Ltd.

Whiting, A. K., & Perticolas, W. L. (1994) Details of the acyl-enzyme intermediate and the oxyanion hole in serine protease catalysis., *Biochemistry 33*, 552-561.

Wimmer, E., Hellen, C. U. T., Molla, A., Harris, K. S., & Paul, A. V. (1993) Viral proteinases., *Faseb J. 7*, 1128-1128.

Wright, H. T., & Robinson, A. B. (1982) Cryptic amidase active sites catalyse deamidation in proteins. In: *From cyclotrons to cytochromes* (Kaplan, N. D., & Robinson, A. B., Eds.) pp 727-743, Academic Press, New York.

Wright, H. T. (1991(a)) Nonenzymatic deamidation of asparaginyl and glutaminyl residues in proteins., *Crit. Rev. Biochem. Mol. Biol. 26*, 1-52.

Wright, H. T. (1991(b)) Sequence and structure determinants of the nonenzymatic deamidation of asparagine and glutamine residues in proteins., *Protein Engineering 4*, 283-294.

Wüthrich, K., & Grathwohl, C. (1974) A novel approach for studies of the molecular conformations in flexible polypeptides., *FEBS Lett. 43*, 337-340.

Wutz, G., Auer, H., Nowotny, N., Grosse, B., Skern, T., & Kuechler, E. (1996) Equine rhinovirus serotype-1 and serotype-2 - relationship to each other and to aphthoviruses and cardioviruses., *J. Gen. Virol. 77*, 1719-1730.

Xu, M. Q., & Perler, F. B. (1996) The mechanism of protein splicing and its modulation by mutation., *EMBO J. 115*, 5146-5153.

Xu, M. Q., Southworth, M. W., Mersha, F. B., Hornstra, L. J., & Perler, F. B. (1993) In-vitro protein splicing of purified precursor and the identification of a branched intermediate., *Cell 75*, 1371-1377.

Yaron, A., & Naider, F. (1993) Proline dependent structural and biological properties of peptides and proteins., *Crit. Rev. Biochem. Mol. Biol. 28*, 31-81.

Ypma-Wong, M. F., & Semler, B. L. (1987) In vitro molecular genetics as a tool for determining the differential cleavage specificities of the poliovirus 3C proteinase., *Nucleic Acids Res.*, 2069-2088.

Ypma-Wong, M. F., Dewalt, P. G., Johnson, V. H., Lamb, J. G., & Semler, B. L. (1988) Protein 3CD is the major poliovirus proteinase responsible for cleavage of the P1 capsid precursor., *Virology 166*, 265-270.

Ziegler, E., Borman, A. M., Deliat, F. G., Liebig, H. D., Jugovic, D., Kean, K. M., Skern, T., & Kuechler, E. (1995) Picornavirus 2A proteinase-mediated stimulation of internal initiation of translation is dependent on enzymatic-activity and the cleavage products of cellular proteins., *Virology 213*, 549-557.

Zlotnick, A., Reddy, V. S., Dasgupta, R., Schneemann, A., Ray, W. J., Reuckert, R. R., & Johnson, J. E. (1994) Capsid assembly in a family of animal viruses primes an autoproteolytic maturation that depends on a single aspartic acid residue., *J. Biol. Chem. 269*, 13680-13684.