

University of St Andrews



Full metadata for this thesis is available in
St Andrews Research Repository
at:

<http://research-repository.st-andrews.ac.uk/>

This thesis is protected by original copyright

**CO-ORDINATED EXPRESSION VIA ARTIFICIAL
SELF-PROCESSING POLYPROTEIN SYSTEMS.**

By

Vanessa M. Cowton BSc.(Hons)

School of Biology

University of St.Andrews

A thesis submitted in partial fulfillment

of the requirements for the

degree of Doctor of Philosophy

July 2000



DECLARATIONS

I, Vanessa Cowton, hereby certify that this thesis, which is approximately 70,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date..... Signature of Candidate.....

I was admitted as a research student in October 1996 and as a candidate for the degree of Doctor of Philosophy in October 1996; the higher study for which this is a record was carried out in the University of St. Andrews between 1996 and 2000.

Date..... Signature of Candidate.....

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

Date..... Signature of Supervisor.....

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

Date.....

Signature of Candidate.....

ACKNOWLEDGEMENTS

I would like to thank Dr. Martin Ryan for his enthusiastic supervision during the course of this PhD and for his advice and comments throughout the writing of this thesis. I would also like to thank Dr. Claire Halpin and Dr. Abdellah Barakate at Dundee University for all their help and supervision during the plant transformation work. Thanks also to Miss Jill McVee for help with the microscope preparation.

Special thanks to my labmates Emma Bryne and Michelle Donnelly for their help and advice during the early days of this PhD. Thanks also to the rest of the group both past and present.

I would like to dedicate this thesis in memory of my grandparents Edgar and Annie Fletcher, the founder members of the Fletcher clan whom have all given me love and support throughout my life. In particular, a huge thank-you to my parents, my constant source of support and encouragement in everything I've done, and to my sister Sandra who provided a sympathetic ear when things went wrong and then told me to get on with it!

Finally, thanks also to Michael for his endless patience and support during the days when there was no end in sight.

ABSTRACT

The *Picornaviridae* encode a single, long open reading frame that is translated to produce a polyprotein. The viral polyprotein is subsequently processed by viral proteases encoded within the polyprotein to yield the full complement of viral proteins. This strategy of encoding proteins in the format of a self-processing polyprotein has been adapted to create novel artificial self-processing polyprotein systems to allow the introduction and co-ordinated expression of multiple genes in various cellular systems.

Two artificial polyprotein systems were developed using alternative picornaviral proteases, FMDV Lb^{PRO}, a cysteine protease and HRV14 2A^{PRO}, a serine protease, to process the polyprotein. Vectors encoding the artificial polyproteins were constructed and analyzed *in vitro* in rabbit reticulocyte lysate and wheatgerm extract. The polyprotein systems were subsequently expressed and analyzed in prokaryotes and *in planta*.

In all systems both picornaviral proteases processed the polyprotein efficiently *in cis*. *Trans* processing activity however, was not observed *in vitro*. The HRV14 2A^{PRO} processed the polyprotein in prokaryotic and eukaryotic systems completely *in cis* and *in trans* to produce the mature proteins. The FMDV Lb^{PRO} was not as efficient in processing the polyprotein *in trans* in prokaryotes. The activity of FMDV Lb^{PRO} *in planta* was more ambiguous as the downstream reporter protein GUS could not be detected.

CONTENTS

Declarations	ii
Acknowledgements	iv
Abstract	v
Contents	vi
List of figures	xx
List of tables	xxvii
Abbreviations	xxviii

Chapter 1- Introduction

1.1. RNA Viruses	1
1.1.1. Non-enveloped viruses with positive sense RNA genomes	1
1.1.2. Enveloped viruses with positive sense genomes	1
1.1.3. Viruses with negative sense genomes	2
1.1.4. Viruses with double-stranded RNA genomes	2
1.1.5. Viruses that use reverse transcriptase during replication	2
1.2. Picornaviridae	3
1.2.1. Classification of picornaviridae	3
1.2.2. Picornaviral genome	4
1.2.3. Overview of picornavirus replication	7
1.2.4. Virion morphogenesis	10
1.3. Novel RNA processing strategies	10

1.3.1. Production of a polyprotein	11
1.3.2. Leaky scanning.....	11
1.3.3. Readthrough.....	12
1.3.4. Ribosomal frameshift.....	13
1.4. Polyproteins	13
1.4.1. Prokaryotic polyproteins.....	14
1.4.2. Eukaryotic polyproteins.....	15
1.4.3. Viral polyproteins.....	17
1.4.3.1. RNA viruses	18
1.4.3.2. Comovirus polyprotein processing.....	18
1.4.3.3. Potyvirus polyprotein processing.....	19
1.4.3.4. Hepatitis C virus polyprotein processing.....	20
1.4.3.5. Retrovirus polyprotein processing.....	20
1.4.3.6. DNA viruses	21
1.4.4. Picornaviral processing.....	22
1.4.4.1. Primary processing	22
1.4.4.2. Secondary processing.....	24
1.4.4.3. Maturation cleavage.....	24
1.5. Proteases	25
1.5.1. Cellular proteases.....	26
1.5.1.1. Serine proteases	26
1.5.1.2. Cysteine proteases.....	28
1.5.1.3. Metalloproteases.....	29

1.5.1.4. Aspartic proteases.....	31
1.5.2. Viral proteases	32
1.5.2.1. Flavivirus proteases: a serine protease and a metalloprotease?.....	33
1.5.2.2. Potyvirus HC-Pro: a cysteine protease.....	34
1.5.2.3. Retrovirus proteases: the viral aspartic proteases.....	36
1.5.3. <i>Cis</i> and <i>trans</i> cleavage reactions.....	37
1.6. Picornaviral Proteases	38
1.6.1. 3C protease.....	38
1.6.1.1. Cellular targets of 3C ^{Pro}	39
1.6.2. Picornavirus 2A protein.....	41
1.6.2.1. The rhinovirus and enterovirus 2A protein.....	41
1.6.2.2. Other cleavage sites for the 2A ^{Pro}	44
1.6.2.3. The cardiovirus and aphthovirus 2A protein.....	46
1.6.2.4. Parechovirus, Avian Encephalomyelitis virus, and Aichi virus 2A proteins.....	46
1.6.3. Aphthovirus leader protease.....	47
1.6.3.1. Cleavage activities of L ^{Pro}	48
1.7. Overview of translation	50
1.7.1. Principle differences between prokaryotic and eukaryotic translation.....	50
1.7.1.1. Comparison of prokaryotic and eukaryotic mRNA structure.....	52

1.7.2. Eukaryotic translation.....	53
1.7.2.1. Formation of the initiation complex.....	53
1.7.2.2. Elongation.....	58
1.7.2.3. Termination	58
1.8. The eIF4F complex.....	58
1.8.1. eIF4A.....	59
1.8.2. eIF4E.....	59
1.8.3. eIF4G.....	60
1.8.3.1. eIF4G binding activities.....	61
1.8.4. Plant eIF4F complex.....	62
1.9. Host-cell shut-off.....	62
1.9.1. Picornavirus host-cell shut-off.....	63
1.9.1.1. Site of eIF4G cleavage	65
1.9.2. Alternative shut-off mechanisms.....	66
1.9.2.1. Cardiovirus shut-off.....	66
1.9.2.2. Adenovirus shut-off.....	66
1.9.3. Shut-off in plant systems	66
1.10. Cap-independent translation.....	67
1.10.1. Picornaviral internal ribosome entry sites.....	67
1.10.2. Mechanism for internal initiation.....	68
1.10.3. The role of cellular proteins during picornavirus IRES-dependent translation.....	70
1.10.4. Host-cell mRNAs containing IRES elements	70

1.11. Aims of the project	72
1.11.1. Background.....	72
1.11.2. Overview of project aims.....	75

Chapter 2- Experimental

2.1. General manipulation of DNA	78
2.1.1. Materials and solutions.....	78
2.1.2. Agarose-gel electrophoresis	78
2.1.3. Restriction enzyme digests.....	79
2.1.4. Purification of DNA fragments from agarose.....	79
2.1.5. DNA Preparations.....	79
2.1.6. Isolation of Genomic DNA from gram negative bacteria.....	82
2.1.7. Nucleotide dideoxy sequencing of recombinant DNA clones.....	82
2.2. Construction of DNA cassettes	
2.2.1. Materials.....	84
2.2.2. Determination of DNA concentration.....	84
2.2.3. Polymerase chain reaction (PCR).....	84
2.2.4. Direct purification of PCR products.....	85
2.2.5. Annealing of oligonucleotide adapters.....	85
2.2.6. Ligation of vector and insert.....	85
2.3. Transformation of <i>Escherichia coli</i>	
2.3.1. Materials and solutions.....	86

2.3.2. Preparation of competent <i>E.coli</i>	86
2.3.3. Screening for recombinant clones	88
2.4. Transcription and translation <i>in vitro</i>	
2.4.1. Materials	89
2.4.2. Preparation of template DNA for <i>in vitro</i> translation	89
2.4.3. Coupled transcription and translation (TnT) reactions.	90
2.4.4. Two-step transcription and translation system	90
2.5. Protein analysis and detection	
2.5.1. Materials and solutions	91
2.5.2. Protein concentration assay	91
2.5.3. Denaturing PAGE analysis	91
2.5.4. Coomassie brilliant blue staining	92
2.5.5. Fixing radiolabelled gels	92
2.5.6. Autoradiography	93
2.5.7. Phosphorimaging analysis	93
2.5.8. Western blotting	93
2.6. Protein expression in <i>Escherichia coli</i>	94
2.7. Transformation of <i>Agrobacterium tumefaciens</i>	95
2.8. <i>Agrobacterium</i>-mediated transformation of <i>Nicotiana tabacum</i>	
2.8.1. Materials	96
2.8.2. Transformation of <i>Nicotiana tabacum</i> plants	96
2.8.3. Sterile micropropagation and harvesting	97
2.8.4. GUS assay of leaf sections	98

2.8.5. Preparation of plant tissue sections for microscopy.....	99
---	----

Chapter 3- Molecular cloning and *in vitro* analysis of an artificial self-processing polyprotein system

3.1. Part 1. Construction and *in vitro* characterization of control plasmids and the artificial reporter polyproteins

3.1.1. Introduction.....	100
3.1.2. Cloning strategies used to construct the artificial polyprotein system.....	101
3.1.3. Construction of the control plasmid pwtGFP.....	104
3.1.4. Construction of the pLb ^{PRO} GFP plasmid via an overlap PCR strategy.....	106
3.1.5. Construction of the control plasmid pLbGFP.....	108
3.1.6. Construction of the artificial polyprotein plasmid pGFP2A ^{PRO}	110
3.1.7. Construction of the control plasmid pGFP2A.....	112
3.1.8. Translation of pwtGFP <i>in vitro</i> in rabbit reticulocyte lysate.....	116
3.1.9. Translation of pLb ^{PRO} GFP <i>in vitro</i> in rabbit reticulocyte lysate.....	117
3.1.10. Translation of pLbGFP <i>in vitro</i> in rabbit	

reticulocyte lysate.....	118
3.1.11. Comparison of coupled TnT reactions in WGE and RRL of FMDV Lb-containing plasmids.....	119
3.1.12. Translation of pGFP2A ^{PRO} <i>in vitro</i> in rabbit reticulocyte lysate.....	121
3.1.13. Translation of plasmid pGFP2A <i>in vitro</i> in rabbit reticulocyte lysate.....	122
3.1.14. Comparison of translation in WGE and RRL of GFP2A polyproteins.....	125
3.1.15. Comparison of cleavage efficiency of Lb ^{PRO} GFP and GFP2A ^{PRO}	126
3.1.16. Conclusions.....	129
3.1.17. Summary.....	131

3.2. Part 2. Construction of an artificial polyprotein system using FMDV Lb^{PRO} and HRV14 2A^{PRO} to investigate *trans* processing *in vivo*.

3.2.1. Introduction.....	132
3.2.2. Construction of the plasmid pLb ^{PRO} BLG.....	133
3.2.3. Translation of pLb ^{PRO} BLG <i>in vitro</i> in RRL.....	136
3.2.4. Construction of plasmid pGLB2A ^{PRO}	137
3.2.5. Expression of plasmid pGLB2A ^{PRO} <i>in vitro</i> in RRL.....	141
3.2.6. Conclusions.....	142
3.2.7. Summary.....	144

Chapter 4- Investigation into the effect of including an internal ribosome entry sites on the *in vitro* expression of the polyprotein cassettes.

4.1. Part 1: *In vitro* analysis of translation in rabbit reticulocyte lysate and wheatgerm extract using a coupled TnT system

4.1.1. Introduction.....	145
4.1.2. Construction of the IRES plasmids pHI and pFI.....	146
4.1.3. Construction of the HRV14 IRES and FMDV IRES artificial polyprotein plasmids.....	150
4.1.4. Translation of the control plasmid series pwtGFP, pHI.GFP and pFI.GFP <i>in vitro</i> in rabbit reticulocyte lysate.....	154
4.1.5. HRV14 IRES-driven translation of the artificial reporter polyproteins <i>in vitro</i> in rabbit reticulocyte lysate.....	155
4.1.6. Translation of the FMDV IRES-containing series of plasmids <i>in vitro</i> in RRL.....	157
4.1.7. A comparison of the relative translation efficiency of HRV14 IRES-driven translation and FMDV IRES-driven translation of artificial polyproteins.....	159
4.1.8. The effect of the HRV14 2A ^{PRO} on IRES-driven translation.....	162
4.1.9. Translation of the control IRES-driven wtGFP series <i>in vitro</i> in wheatgerm extract.....	165
4.1.10. Translation of the IRES-driven artificial polyproteins	

<i>in vitro</i> in wheatgerm extract.....	166
4.1.11. Translation of pLb ^{PRO} GFP and pLbGFP <i>in vitro</i> in RRL and WGE.....	170
4.1.12. Conclusions.....	174
4.1.13. Summary.....	176

4.2. Part 2: *In vitro* analysis of translation in rabbit reticulocyte lysate using a two-step coupled system.

4.2.1. Introduction.....	177
4.2.2. <i>In vitro</i> analysis of the wtGFP plasmid series using the Novagen STP3 system.....	178
4.2.3. <i>In vitro</i> analysis of the Lb ^{PRO} GFP plasmid series using the Novagen STP3 system.	180
4.2.4. <i>In vitro</i> analysis of the GFP2A ^{PRO} plasmid series using the Novagen STP3 system.	183
4.2.5. Conclusions.....	185
4.2.6. Summary.....	187

Chapter 5- A study of artificial polyprotein processing in prokaryotes.

5.1. Part 1: An investigation of the processing activity of artificial polyproteins containing picornaviral proteases.

5.1.1. Introduction.....	188
--------------------------	-----

5.1.2. Construction of prokaryotic expression vectors	
pTrc.Lb ^{PRO} BFP and pTrc.BFP2A ^{PRO}	190
5.1.3. Construction of pTrc.Lb ^{PRO} BLG and pTrc.GLB2A ^{PRO}	192
5.1.4. A time course study of artificial polyprotein expression in <i>Escherichia coli</i>	195
5.1.5. Western blot analysis of protein expression from pTrc.Lb ^{PRO} BFP and pTrc.BFP2A ^{PRO}	198
5.1.6. Western blot analysis of protein expressed from pTrc.Lb ^{PRO} BLG	200
5.1.7. Western blot analysis of proteins expressed from pTrc.GLB2A ^{PRO}	202
5.1.8. Conclusions.....	204
5.1.9. Summary.....	205

**5.2. Part 2: An investigation into the effect of an N- or C-terminal extension on -
glucuronidase activity in *Escherichia coli***

5.2.1. Introduction.....	206
5.2.2. Construction of the prokaryotic expression plasmid pTrc.GUS2AGFP.....	208
5.2.3. The β -glucuronidase activity of N- and C-terminal fusion proteins in <i>E.coli</i>	210
5.2.4. Investigation of the cleavage activity of FMDV 2A in <i>E.coli</i>	210
5.2.5. Western blot analysis of CAT2AGUS and GUS2AGFP	

expressed in <i>E.coli</i>	212
5.2.6. Conclusions.....	213
5.2.7. Summary.....	214

Chapter 6: An investigation into the potential of the artificial self-processing polyprotein system *in planta*.

6.1. Part 1: Construction of the binary vectors and analysis of the artificial self-processing dual reporter system *in planta*.

6.1.1. Introduction.....	215
6.1.2. Construction of the control binary vector pPZP.BFP.....	217
6.1.3. Construction of the artificial polyprotein binary vectors pPZP.Lb ^{PRO} BLG and pPZP.GLB2A ^{PRO}	219
6.1.4. Transformation of <i>Agrobacterium tumefaciens</i>	222
6.1.5. Optimization of <i>Agrobacterium</i> transformation.....	222
6.1.6. Transformation of <i>Agrobacterium tumefaciens</i> with pPZP.Lb ^{PRO} BLG.....	224
6.1.7. Further analysis of the pPZP.Lb ^{PRO} BLG2 binary vector.....	227
6.1.8. <i>Agrobacterium</i> --mediated transformation of <i>Nicotiana tabacum</i>	230
6.1.9. Initial analysis of pPZP.Lb ^{PRO} BLG and pPZP.GLB2A ^{PRO} transformants from the <i>Agrobacterium</i> -mediated seedling transformation experiment.....	235

6.1.10. Western blot analysis of putative Lb ^{PRO} BLG transformants.....	239
6.1.11. Western blot analysis of GLB2A ^{PRO} transformants.....	242
6.1.12. Visualization of the BFP cleavage product via UV illumination.....	244
6.1.13. Conclusions.....	249
6.1.14. Summary.....	252

6.2. Part 2: investigation of the toxic effect of the FMDV Lb protease to *Agrobacterium tumefaciens*.

6.2.1. Introduction.....	253
6.2.2. Investigation of the transformation efficiency of the pPZP.Lb ^{PRO} BLG plasmid series.....	253
6.2.3. Further analysis of the plasmid DNA extracted from pPZP.Lb ^{PRO} BLG transformed <i>Agrobacterium tumefaciens</i>	256
6.2.4. Investigation of the DNA insertion within FMDV Lb ^{PRO} in pPZP.Lb ^{PRO} BLG1.....	258
6.2.5. Investigation of the origin of IS10.....	261
6.2.6. Conclusions.....	265
6.2.7. Summary.....	268

Chapter 7: Discussion

7.1. Introduction.....	269
7.2. The effect of an N- or C-terminal extension on β -glucuronidase activity in <i>Escherichia coli</i>	269
7.3. Investigation of IRES-driven <i>in vitro</i> translation.....	271
7.3.1. IRES translation in the coupled TnT system.....	271
7.3.2. <i>In vitro</i> TnT translation using STP3 system.....	275
7.3.3. Future work.....	278
7.3.4. Conclusions.....	278
7.4. The FMDV Lb ^{PRO} -based polyprotein system.....	279
7.4.1. <i>In vitro</i> analysis.....	279
7.4.2. Expression in prokaryotes.....	280
7.4.3. Expression <i>in planta</i>	281
7.4.4. Summary of the FMDV Lb ^{PRO} -based polyprotein system.....	286
7.4.5. Further work.....	287
7.5. The HRV14 2A ^{PRO} -based polyprotein system.....	288
7.5.1. <i>In vitro</i> analysis.....	288
7.5.2. Expression in prokaryotes.....	291
7.5.3. Expression <i>in planta</i>	292
7.5.4. Summary of the HRV14 2A ^{PRO} -based polyprotein system.....	292
7.5.5. Further work.....	293

7.6. Conclusions.....	294
7.7. Applications of the artificial polyprotein system.....	296
7.8. Concluding remarks.....	297
APPENDIX	298
BIBLIOGRAPHY	300

LIST OF FIGURES

Chapter 1

1.1. Overview of picornavirus genome organisation	7
1.2. The picornavirus icosahedral capsid.....	8
1.3. Overview of the picornavirus life-cycle.....	9
1.4. Polyprotein production.....	11
1.5. Leaky scanning.....	12
1.6. Readthrough	12
1.7. Ribosomal frameshift.....	13
1.8. Prokaryotic polyproteins.....	15
1.9. Eukaryotic polyproteins.....	17
1.10. CPMV polyprotein.....	19
1.11. TMV polyprotein.....	19
1.12. HCV polyprotein.....	20

1.13.	MMLV polyprotein	21
1.14.	Polyproteins from ASFV.....	22
1.15.	Overview of the picornaviral polyprotein and processing.....	23
1.16.	Chemical reactions catalyzed by proteases.....	25
1.17.	Schechter and Berger nomenclature.....	26
1.18.	The serine protease catalytic mechanism	28
1.19.	Crystal structure of papain	29
1.20.	Catalytic mechanism of the metalloproteases.....	30
1.21.	Structure of the HIV1 protease.....	32
1.22.	Crystal structure of HCV NS3 protease.....	35
1.23.	Coding regions of retrovirus polyproteins.....	37
1.24.	Crystal structure of HRV14 2A ^{pro}	42
1.25.	Sequence alignment of the VP1-2A cleavage site.....	44
1.26.	Crystal structure of FMDV Lb ^{pro}	49
1.27.	The eukaryotic cell.....	51
1.28.	The prokaryotic cell.....	52
1.29.	Step 1 of initiation of translation.....	53
1.30.	Step 2 of initiation of translation.....	54
1.31.	Step 3 of initiation of translation.....	55
1.32.	Step 4 of initiation of translation.....	56
1.33.	The mammalian eIF4F complex.....	59
1.34.	The binding sites of the eIF4G protein.....	61
1.35.	The 2A ^{pro} and Lb ^{pro} mediated cleavage of eIF4G.....	65

1.36. Type I and Type II IRES structure	69
---	----

Chapter 3

3.1. Addition of novel restriction sites to constructs by the PCR.....	101
3.2. The overlap PCR strategy.....	103
3.3. Construction of pwtGFP	105
3.4. Construction of pLb ^{PRO} GFP.....	107
3.5. Construction of pLbGFP.....	109
3.6. Construction of pGFP2A ^{PRO}	111
3.7. Initial strategy for construction of pGFP2A	113
3.8. Alternative strategy for the construction of pGFP2A.....	115
3.9. <i>In vitro</i> transcription and translation in RRL of pwtGFP.....	116
3.10. <i>In vitro</i> transcription and translation of pLb ^{PRO} GFP in RRL.....	118
3.11. <i>In vitro</i> transcription and translation of pLbGFP in RRL.....	119
3.12. Comparison of TnT reactions in RRL and WGE systems.....	120
3.13. <i>In vitro</i> transcription and translation of pGFP2A ^{PRO} in RRL.....	122
3.14. <i>In vitro</i> transcription and translation of pGFP2A.....	123
3.15. Comparison of <i>in vitro</i> TnT in RRL and WGE.....	126
3.16. Comparison of cleavage efficiency of self-processing polyproteins in RRL and WGE.....	127
3.17. Lb ^{PRO} linker region.....	134
3.18. Construction of plasmid pLb ^{PRO} BLG.....	135
3.19. <i>In vitro</i> transcription and translation of pLb ^{PRO} BLG in RRL.....	137

3.20.	HRV14 2A ^{PRO} linker region.....	138
3.21.	Construction of the plasmid pGLB2A ^{PRO}	140
3.22.	Transcription and translation <i>in vitro</i> of pGLB2A ^{PRO} in RRL.....	142

Chapter 4

4.1.	Construction of the pHI plasmid.....	148
4.2.	Construction of the pFI plasmid.....	149
4.3.	Construction of the HI.polyprotein plasmid series.....	152
4.4.	Construction of the FI.polyprotein plasmid series.....	153
4.5.	<i>In vitro</i> translation of the wtGFP plasmid series in RRL.....	154
4.6.	<i>In vitro</i> translation of HRV14 IRES-driven transcripts in RRL.....	156
4.7.	<i>In vitro</i> translation of FMDV IRES-driven transcripts in RRL.....	158
4.8.	<i>In vitro</i> translation of uncapped and IRES-driven transcripts in RRL.....	160
4.9.	A prolonged exposure of the complete plasmid series translation profiles in RRL.....	163
4.10.	IRES-driven polyprotein translation products in RRL.....	164
4.11.	Comparison of <i>in vitro</i> TnT of GFP plasmid series in RRL and WGE.....	165
4.12.	TnT analysis of complete series in RRL and WGE.....	167
4.13.	Translation of the pLbGFP plasmid series in RRL and WGE.....	171
4.14.	Translation of GFP plasmid series using the STP3 system.....	179
4.15.	TnT analysis of the Lb ^{PRO} GFP plasmid series using the STP3 system.....	181

4.16. TnT analysis of the GFP2A ^{pro} plasmid series using the STP3 system.....	183
--	-----

Chapter 5

5.1. Co-ordinated expression of multiple genes.....	188
5.2. Construction of pTrc.Lb ^{pro} BFP and pTrc.BFP2A ^{pro}	191
5.3. Construction of pTrc.Lb ^{pro} BLG and pTrc.GLB2A ^{pro}	194
5.4. Time course of Lb ^{pro} BFP and BFP2A ^{pro} expression in <i>E.coli</i>	196
5.5. Time course of Lb ^{pro} BLG and GLB2A ^{pro} expression in <i>E.coli</i>	197
5.6. Western blot probing with polyclonal GFP antisera.....	198
5.7. Western blot of Lb ^{pro} BFP and BFP2A ^{pro} with BFP antisera.....	199
5.8. Western blot analysis with GUS antisera.....	201
5.9. Western blot analysis of pTrc.Lb ^{pro} BLG.....	202
5.10. Western blot analysis of pTrc.GLB2A ^{pro}	203
5.11. Monitoring GUS expression in <i>E.coli</i>	207
5.12. Construction of pTrc.GUS2AGFP.....	209
5.13. Monitoring GUS expression in FMDV 2A polyproteins.....	211
5.14. Western blot analysis of FMDV 2A polyproteins expressed in <i>E.coli</i>	213

Chapter 6

6.1. Construction of the control binary vector pPZP.BFP.....	218
6.2. Construction of pPZP.Lb ^{pro} BLG.....	220
6.3. Construction of pPZP.GLB2A ^{pro}	221

6.4.	Restriction analysis of pPZP.Lb ^{PRO} BLG.....	226
6.5.	Construction of pGEM.Lb ^{PRO} BLG2.....	228
6.6.	<i>In vitro</i> translation of pGEM.Lb ^{PRO} BLG2 in RRL.....	229
6.7.	Four-week <i>Nicotiana tabacum</i> seedlings.....	232
6.8.	Lb ^{PRO} BLG GUS assay.....	237
6.9.	GLB2A ^{PRO} GUS assay.....	238
6.10.	Western blot of Lb ^{PRO} BLG with BFP antisera.....	240
6.11.	Western blot of Lb ^{PRO} BLG with GUS antisera(5'→3').....	241
6.12.	Western blot of Lb ^{PRO} BLG with GUS antisera (Zeneca).....	242
6.13.	Western blot of GLB2A ^{PRO} with BFP antisera.....	243
6.14.	Western blot of GLB2A ^{PRO} with GUS antisera (Zeneca).....	244
6.15.	Spectra of GFP variants.....	245
6.16.	UV illumination of root extracts.....	246
6.17.	Microscopy of stem cell transverse sections.....	248
6.18.	Proposed GLB2A ^{PRO} polyprotein processing.....	250
6.19.	Proposed Lb ^{PRO} BLG polyprotein processing.....	251
6.20.	<i>Agrobacterium</i> transformation on Kan ^R plates.....	254
6.21.	<i>Agrobacterium</i> transformation on Chl ^R plates.....	255
6.22.	Graph of <i>Agrobacterium tumefaciens</i> colony number.....	256
6.23.	Alignment of CaMV promoter region.....	257
6.24.	Amplification of the FMDV Lb ^{PRO} by the PCR.....	259
6.25.	Alignment of the ORF of the insertion sequence.....	260

6.26.	The FMDV Lb ^{PRO} region from pPZP.Lb ^{PRO} BLG1.....	261
6.27.	Analytical PCR of <i>Agrobacterium tumefaciens</i> populations.....	263
6.28.	Analytical PCR of vector DNA.....	264
6.29.	Analytical PCR of XL1-Blue <i>E.coli</i>	265

Chapter 7

7.1.	Proposed model of uncapped translation.....	274
7.2.	Site of the -FFF- sequence in the Lb ^{PRO} BLG polyprotein.....	284
7.3.	TnT analysis of FFF-containing constructs in WGE.....	285
7.4.	Summary of FMDV Lb ^{PRO} -based system.....	287
7.5.	Summary of HRV14 2A ^{PRO} -based system.....	293
7.6.	Composition of polyprotein encoding enzymes A, B and C.....	295

LIST OF TABLES

Chapter 1

1.1.	Classification of Picornaviridae.....	3
1.2.	List of eukaryotic initiation factors.....	57

Chapter 2

2.1.	Oligonucleotide sequencing primers.....	83
2.2.	Recipes for SDS-PAGE analysis.....	92

Chapter 3

3.1.	Molecular weight of full-length and cleaved Lb ^{PRO} GFP.....	117
3.2.	Molecular weight of full-length and cleaved GFP2A ^{PRO}	121
3.3.	Molecular weight and sequence context of putative internal initiation products from pGFP2A ^{PRO}	125
3.4.	Methionine residues in polyproteins.....	128
3.5.	Phosphorimaging analysis of polyproteins.....	128
3.6.	Radioactivity in uncleaved polyprotein and cleavage products.....	129
3.7.	Predicted molecular weight for full-length and cleavage products of pLb ^{PRO} BLG.....	136
3.8.	Predicted molecular weight for full-length and cleavage products of pGLB2A ^{PRO}	141

Chapter 4

4.1.	Complete list of basic reporter system plasmids.....	151
4.2.	Densitometric analysis of wtGFP plasmid series in RRL.....	155
4.3.	Densitometric analysis of polyprotein plasmid series in RRL.....	162
4.4.	Densitometric analysis of complete plasmid series in WGE.....	169

4.5.	Putative initiation codons present within FMDV Lb ^{PRO}	172
4.6.	Densitometric analysis of pGFP2A plasmid series in RRL and WGE.....	173
4.7.	Phosphorimaging analysis of wtGFP in the STP3 TnT system.....	180
4.8.	Phosphorimaging analysis of Lb ^{PRO} GFP in the STP3 TnT system.....	182
4.9.	Phosphorimaging analysis of GFP2A ^{PRO} in the STP3 TnT system.....	184
 Chapter 6		
6.1.	<i>Agrobacterium</i> resistance to chloramphenicol.....	223
6.2.	<i>Agrobacterium</i> resistance to kanamycin.....	223
6.3.	Non-transformed <i>Agrobacterium</i> antibiotic resistance in liquid culture.....	224
6.4.	Observations of plantlets from seedling transformation.....	233
6.5.	Observations of plantlets from leaf-disc transformation.....	235
 Chapter 7		
7.1.	Ratios of cleavage products for –FFF- constructs.....	286
 Appendix		
8.1.	List of oligonucleotide primers for PCR products.....	298

ABBREVIATIONS

ACTH- Adrenocorticotrophic hormone	EDTA- Disodium ethylenediaminetetra acetate
AFGP- Antifreeze glycoproteins	EF- Elongation factor
AIDS- Acquired immune deficiency syndrome	EGTA- Ethylene glycol-bis(β -aminoethyl ether) N,N,N',N',-Tetraacetic acid
ALV- Avian leukosis virus	eIF- Eukaryotic initiation factor
ASFV- African swine fever virus	EMCV- Encephalomyocarditis virus
Atm- Atmospheres	FMDV- Foot-and-mouth disease virus
BA- 6-benzylaminopurine	4E-BP- 4E-binding protein
BFP- Blue fluorescent protein	GDP- Guanosine 5' diphosphate
BiP- immunoglobulin heavy-chain binding protein	GFP- Green fluorescent protein
BSA- Bovine serum albumin	GTP- Guanosine 5' triphosphate
C- Carboxy-	GUS- β -glucuronidase
CaMV- Cauliflower mosaic virus	HAV- Hepatitis A virus
CAT- Chloramphenicol acetyl transferase	HC-pro- Helper component protease
CLIP- Corticotropin-like intermediate lobe protein	HCV- Hepatitis C virus
CPMV- Cowpea mosaic virus	HPeV- Human parechovirus
CREB- cAMP-responsive element binding protein	HRP- Horseradish peroxidase
CTE- C-terminal extension	HRV- Human rhinovirus
dH₂O- Distilled water	HIV- Human immunodeficiency virus
dNTP- 2'Deoxy nucleoside 5'triphosphate	HSV- Herpes simplex virus
DNA - Deoxyribonucleic acid	ICAM-1- Intercellular adhesion molecule-1
DMF- Dimethyl formamide	IL-12- Interleukin-12
dsRNA- Double-stranded RNA	IPTG- Isopropylthio- β -D-galactoside
eBFP- Enhanced blue fluorescent protein	IRES- Internal ribosome entry site
E.coli- <i>Eschericia coli</i>	kDa- Kilo Daltons
	LB- Luria-Bertani medium
	LDLR- Low-density lipoprotein receptor

LMP- Low melting point
LPH- Lipotropin hormone
MAP-4- Microtubule associated protein
MCS- Multiple cloning site
MLV- Murine leukaemia virus
MMTV- Mouse mammary tumour virus
MPMV- Mason-Pfizer monkey virus
mRNA- messenger RNA
MS- Murashige and Skooge media
MSH- Melanocyte stimulating hormone
m7G- 7-methyl guanosine
N- Amino-
NAA- naphthalene acetic acid
N1a-pro- Nuclear inclusion a protease
ORF- Open reading frame
PABP- Poly(A) binding protein
PAGE- Polyacrylamide gel electrophoresis
PBS- Phosphate buffered saline
PCBP-2- Poly (rC) binding protein 2
PCR- Polymerase chain reaction
Pol- Polymerase
POMC- Proopiomelanocortin
PSL- Photo-stimulated luminescence
PTB- Poly tract binding protein
PVR- Poliovirus receptor
PVX- Potato virus X
RNA - Ribonucleic acid
RRL- Rabbit reticulocyte lysate
RRM- RNA recognition motif
RSV- Rous sarcoma virus
SDS- Sodium dodecyl sulphate
TAE - Tris-acetate/EDTA electrophoresis buffer
Taq Pol - *Thermus aquaticus* polymerase
TBP- TATA binding protein
TE - Tris-EDTA buffer
TEMED- N,N,N',N'-tetramethylethylenediamine
TEV- Tobacco etch virus
TF- Transcription factor
TnT- Transcription and translation
tRNA- Transfer RNA
TYMV- Turnip yellow mosaic virus
UTR- Untranslated region
UV- Ultraviolet
VPg- Genome-linked viral protein
vRNA- Viral RNA
WGE- Wheatgerm extract
wt- Wildtype
X-GAL- 5-Bromo-4-Chloro-3-Indolyl- β -D-Galactoside
X-GlcA- 5-Bromo-4-Chloro-3-Indolyl- β -D-Glucuronide

Nucleotide abbreviations-

A- Adenosine
C- Cytidine
G- Guanosine
T- Thymidine

Amino Acid	Three letter code	Single letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamic Acid	Glu	E
Glutamine	Gln	Q
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

CHAPTER 1: INTRODUCTION

1.1 Overview of RNA viruses.

A disease causing agent smaller than bacteria that could pass through bacterial filters was first suggested in 1898. A huge number of viruses have been identified since, these fall into two principle divisions, *i.e.* RNA-containing and DNA-containing viruses. RNA viruses may be considered in five distinct groups.

1.1.1 Non-enveloped viruses with positive-sense RNA genomes

Viruses within this group do not possess a lipid-containing envelope, the viral genome that is composed of positive-sense RNA is enclosed within a protein capsid. The viral positive-sense RNA acts as a template for the synthesis of negative-sense strands which in turn act as templates for positive-strand RNA. The positive-sense RNA can either function as mRNA for translation of viral proteins or as virion RNA that is packaged to form new virus.

Positive-sense RNA viruses are typified by the *Picornaviridae* which will be discussed in greater depth in section 1.2. Other members include the caliciviruses, astroviruses and a large number of plant viruses.

1.1.2. Enveloped viruses with positive-sense RNA genomes

Viruses in this group also contain a positive-strand RNA genome contained within a protein capsid this is encased in a lipid bilayer envelope derived from host-cell membrane. The envelope contains glycoproteins encoded by the virus, often the glycoproteins form characteristic projections or 'spikes' in the envelope. Replication occurs via a negative-sense RNA intermediate like the non-enveloped positive-sense RNA viruses. Families within this group include the *Togaviridae*, *Flaviviridae*, *Coronaviridae* and *Toroviridae*.

1.1.3. Viruses with negative-sense RNA genomes

Unlike positive-strand RNA viruses the virion RNA is not infectious, the virus must first transcribe the RNA genome into a positive RNA strand that can act as mRNA. To achieve this the nucleocapsid core of a negative-sense RNA virus also contains a viral protein that has transcriptase activity that performs this function. The nucleocapsid core is enclosed within a lipid-containing bilayer. This group contains the causative agents of many important diseases such as influenza, measles and rabies.

1.1.4. Viruses with double-stranded RNA genomes

The viruses within this group are non-enveloped but often the double-strand RNA genome is encased in a double shell of protein. The structure of the reovirus core has recently been elucidated (Reinisch, Nibert and Harrison, 2000). There are several families within this group the principle family is the *Reoviridae*, members of this family contain between ten and twelve segments of double-stranded RNA that make up the viral genome. The negative-sense strand is transcribed initially to give mRNA that is capped and methylated by viral core proteins. The capped mRNA is translated to produce the full complement of viral proteins. The proteins form new viral cores where the mRNA is used as a template to replicate the double-strand RNA viral genome. This family of viruses is unique in that members utilize a conservative replication process; other families of viruses within the double-strand virus group use the semi-conservative mechanism of replication.

1.1.5. RNA Viruses that use reverse transcriptase during replication

The viruses in this group have a positive-sense RNA genome that is initially copied into DNA by reverse transcriptase - an enzyme with an RNA-dependent DNA polymerase function. This family of viruses is known as retroviruses there has been

much interest in this group because they are associated with important diseases including cancer and AIDS.

The *Retroviridae* are enveloped viruses with an icosahedral nucleocapsid that contains two copies of RNA and the reverse transcriptase enzyme. The reverse transcriptase is responsible for producing DNA complementary to viral RNA, digesting the RNA strand of the resulting DNA/RNA hybrid and finally producing a complementary DNA strand. The provirus double-strand DNA can integrate into the host cell genome and transform the cell. *Retroviridae* are unique because they have a diploid genome of two identical RNA molecules.

1.2 PICORNAVIRIDAE

The picornavirus family includes several of the most important pathogens that affect man and animals in both a medical and economic sense. For instance, some of the more notable members of this family includes poliovirus, rhinoviruses which are the most important etiologic agents of the common cold and foot-and-mouth disease virus (FMDV) that affects cloven-footed animals and hence has a severe impact on the agricultural economy (reviewed by Rueckert, 1996).

1.2.1 Classification of Picornaviridae -

The picornavirus family is at present divided into nine genera (refer to Table 1.1) Although a large number of viruses remain unclassified.

Genus	Species	Serotypes
ENTEROVIRUS	Poliovirus	Human poliovirus 1-3
	Human enterovirus A	Human coxsackievirus A 2,3,5, 7,8,10,12,14,16 Human enterovirus 71

	Human enterovirus B	Human coxsackievirus B 1-6 Human echovirus 1-9,11-21,24-27,29-33 Human enterovirus 69
	Human enterovirus C	Human coxsackievirus A 1,11, 13,15,17-22,24
	Human enterovirus D	Human enterovirus 68,70
	Bovine enterovirus	Bovine enterovirus 1,2 Ovine enterovirus
	Porcine enterovirus A	Porcine enterovirus 8
	Porcine enterovirus B	Porcine enterovirus 9,10
	Unassigned to species	Human coxsackievirus A 4,6 Simian enterovirus 1-18, N125, N203
RHINOVIRUS	Human rhinovirus A	Human rhinovirus 1,2,7,9,11,15, 16,21,29,36,39,49,50,58,62,65,85, 89,
	Human rhinovirus B	Human rhinovirus 3,14,72
	Unassigned to species	Human rhinovirus 4-6,10-13,17- 20,22-28,30-38,40-48,51-57,59-61,63,64,66-71,73-84,86-88,90-100 Bovine rhinovirus 1-3
CARDIOVIRUS	Encephalomyocarditis virus Theilovirus	Encephalomyocarditis virus Theiler's murine encephalomyelitis virus Vilyuisk human encephalitis virus
APHTHOVIRUS	Foot-and-mouth disease	O, A, C, Sat 1-3, Asia-1

	virus Equine rhinitis A virus	Equine rhinitis virus A
HEPATOVIUS	Hepatitis A virus Avian encephalomyelitis-like virus	Human hepatitis A virus Simian hepatitis A virus
PARECHOVIRUS	Human parechovirus	Human parechovirus 1,2
ERBOVIRUS	Equine rhinitis B virus	Equine rhinitis B virus
KOBUVIRUS	Aichi virus	Aichi virus
TESCHOVIRUS	Porcine teschovirus	Porcine teschovirus 1
UNCLASSIFIED	Includes mammalian, avian and piscine viruses	

Table 1.1. Classification of the picornaviruses.

1.2.2 The Picornaviral Genome

Picornaviruses, as the name suggests, are small (pico) RNA viruses. The viral genome is composed of single-strand, positive-sense RNA. The genome length ranges from 7100 to 8400 nucleotides excluding the polyadenylate tail located at the 3' end of the genome that is some 100-150 nucleotides in length.

The picornaviral genome is unusual in that the 5' end of the genome lacks the 7-methylguanosine cap structure that is covalently linked via a 5'-5' linkage to the terminal base of the RNA in the majority of cellular mRNA (Lee *et al.*, 1977). This is replaced by a small viral protein (VPg) that is covalently linked to the 5' terminal base of the RNA via a phosphodiester linkage to the phenolic hydroxyl group of a tyrosine residue (Rothberg *et al.*, 1978). The role of the VPg protein is not known but it may be involved in the initiation of positive and negative strand RNA synthesis. The 5' end

of the genome contains an untranslated region (UTR) that is long compared to other cellular and viral RNA. This region contains 8-12% of total virus genetic material and is the most similar region of the genome within genera. The region has a high G/C content and is predicted to form complex secondary structures that are involved in initiating viral protein synthesis. These regions are termed internal ribosome entry sites or IRES and will be discussed in section 1.10. (Pelletier *et al.*, 1988; Nicholson *et al.*, 1991).

Picornaviral genomes share a common organizational pattern although there are slight variations among the nine genera. Enteroviruses and rhinoviruses conform to the basic pattern of a long untranslated region followed by a single open reading frame encoding the capsid proteins (P1) and non-structural proteins (P2 and P3). An additional leader (L) protein is present in cardio- and aphthoviruses. The cardio- and aphthoviruses also contain a poly-C tract in the 5'-UTR. The aphthovirus genome encodes three copies of the 3B (VPg) protein. There are also variations in the subsequent processing of the picornaviral genome between genera (Palmenberg, 1987a).

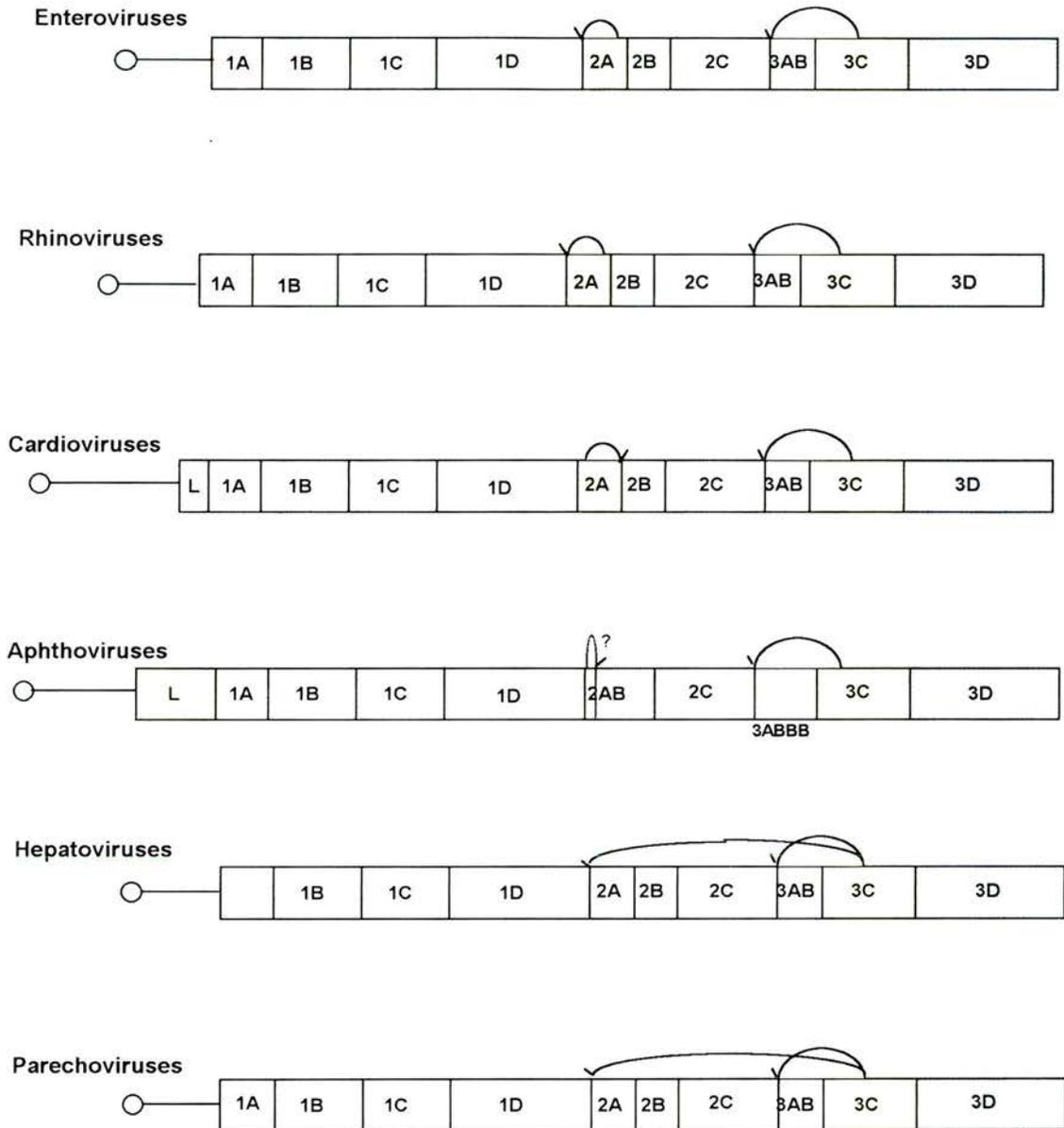


Fig.1.1. Overview of the picornavirus genome organization. The viral RNA-dependent RNA polymerase is shown in purple and the viral proteases are shown in yellow (Stanway, 1990; Stanway and Hyypia, 1999). The principle cleavage events are indicated by the arrows.

The RNA genome is packed within an icosahedral capsid, 20-30 nm in diameter. The capsid is composed of 12 pentamers of the basic protomer subunit. Each protomer contains a copy of each of the four viral structural proteins, VP1 (1D), VP2 (1B) and

VP3 (1C) are exposed on the surface of the virion, VP4 (1A) is buried beneath the surface and is associated with the viral RNA.

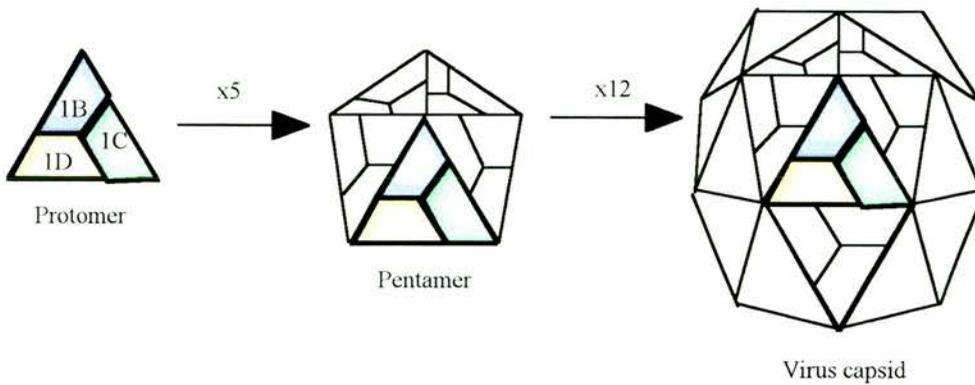


Fig.1.2. The picornavirus icosahedral capsid. The P1 precursor containing the capsid proteins is processed producing the protomer subunit. An intermediate pentamer subunit is produced from five protomer subunits. The final capsid is composed of twelve pentamer subunits.

1.2.3. Overview of picornavirus replication

Picornaviral replication occurs in the cytoplasm of the host cell. The first step in the process is attachment of the virion to the cell surface, the virion binds to a specific receptor on the plasma membrane. The receptor varies between different picornaviruses and helps to determine the host-range and tissue-tropism of the individual viruses. Picornaviral receptors include ICAM-1 (Intercellular adhesion molecule-1) for the major group of rhinoviruses and Coxsackievirus A, LDLR (Low-density lipoprotein receptor) for the minor group of rhinoviruses and PVR (Poliovirus receptor) for poliovirus (Tomassini *et al.*, 1989; Uncapher, DeWitt and Colonno, 1991). The receptor mediates the transfer of the viral genome through the lipid bilayer of the plasma membrane into the cell. Following attachment to the receptor a conformational change occurs in the virion, VP4 is lost and the RNA genome is transferred into the cytosol. Studies have shown that picornaviruses have different requirements for cell penetration and uncoating and therefore must use different cellular processes to enter the cell (Zeichardt *et al.*, 1985; DeTulleo and Kirchhausen,

1998; Schober *et al.*, 1998). The precise details of the mechanisms employed have not yet been determined.

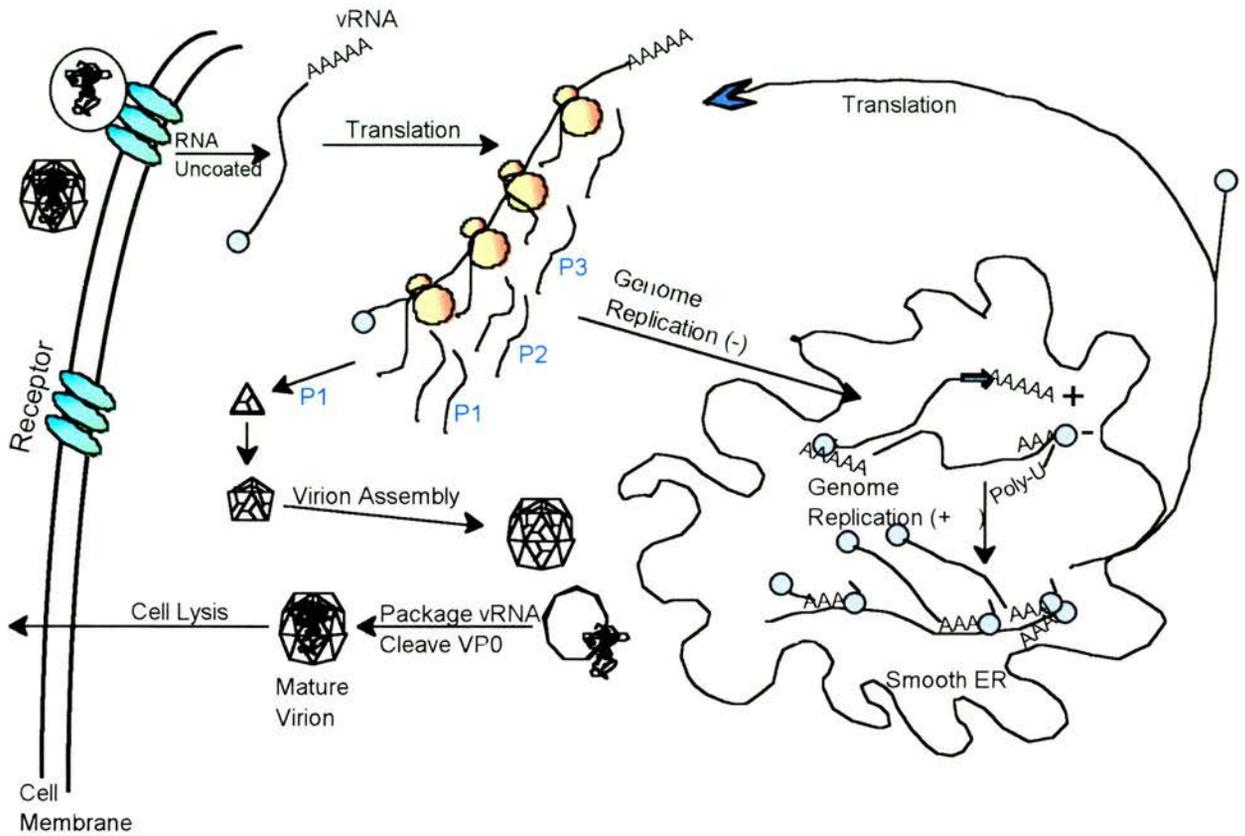


Fig. 1.3. Overview of the picornavirus life-cycle. Adapted from (Rueckert, 1996).

The next stage is translation of the infecting vRNA to produce the picornaviral RNA synthesis machinery required to replicate new vRNA. The picornavirus utilizes the host-cell machinery for protein synthesis and actively inhibits translation of host mRNA (Rose *et al.*, 1978; Wyckoff, 1993). The vRNA encodes a single open reading frame that is translated to produce a large polyprotein. The polyprotein is processed during translation by viral-encoded proteases to yield the full complement of viral proteins (Hellen, Kräusslich and Wimmer, 1989; Palmenberg, 1990).

Synthesis of new vRNA occurs on the smooth endoplasmic reticulum. The infecting vRNA is copied to produce negative-strand RNA that acts as a template for new positive-strand vRNA. Multi-strand replicative intermediates are formed as each full-length negative-strand RNA is thought to act as a template for nascent positive-

strands. The replication process continues until there is a large number of positive-strand RNA molecules that are subsequently packaged into virions.

1.2.4 Virion morphogenesis

Virion assembly is controlled by cleavage of the P1 polypeptide that encodes the coat proteins VP0, VP1 and VP3 that form immature protomers. P1 cleavage is slow during the early stages of infection as the levels of the P1 precursor and the viral proteins 3C^{pro} and 3CD^{pro} that process P1 are low. As concentrations of P1 increase the protomers form pentamers that assemble and package positive VPg-RNA to form the provirion. It is not known if the pentamers form around the VPg-RNA or if the VPg-RNA is transferred into the 80S protein shell.

Provirus are not infectious. The final step to mature virus particles is cleavage of the capsid protein precursor VP0 to VP4 and VP2, this is the maturation cleavage. An interesting observation is that it has now been shown that the maturation cleavage does not occur in the parechoviruses and yet the virus is infectious (reviewed in Stanway and Hyypia, 1999). Following maturation the infectious particles are released upon infection mediated disintegration of the host cell. The complete replication cycle takes between 5 and 10 hours.

1.3.NOVEL RNA PROCESSING STRATEGIES

An early hypothesis in molecular biology was the one gene-one protein hypothesis proposed by Beadle and Tatum (1941). This stated that each gene encoded a single protein, this is valid for the majority of eukaryotic genes but in recent years an increasing number of cases have been identified where multiple proteins are encoded by the same mRNA transcript. Several unique mechanisms are employed at the translational level and perhaps unsurprisingly the mechanisms were first identified in viruses (reviewed by Fütterer and Hohn, 1996; Maia *et al.*, 1996). The viral genome has a finite size due to constraints imposed by the size of the icosahedral capsid.

Therefore the virus must encode all the proteins required for successful viral replication in a small amount of genetic material. Viruses may employ one or more of the following strategies (reviewed by Gallie, 1996; Zeccomer, Haenni and Macaya, 1995).

1.3.1 Production of a polyprotein

A single ORF is translated to produce a large protein precursor that is proteolytically processed to yield an array of functional proteins. Summers and Maizel (1968) first identified this polyprotein strategy in poliovirus but the production of a polyprotein has now been shown to be a widespread phenomenon occurring in prokaryotic and eukaryotic systems. This is discussed in section 1.4.

1. Polyprotein processing

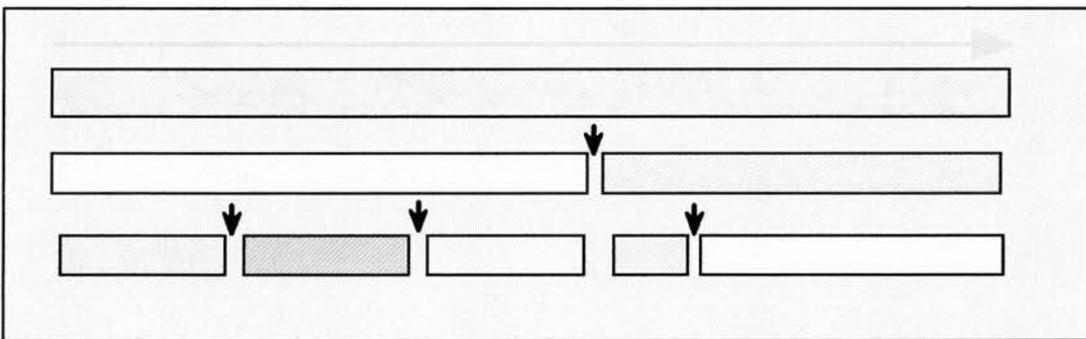
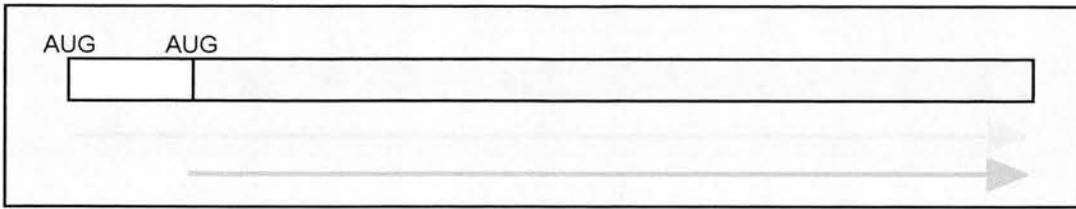


Fig.1.4. Outline of the production of a polyprotein from a single open reading frame. The arrows indicate cleavage sites that are processed to produce the mature proteins.

1.3.2 Leaky scanning

In this case the ORF contains two possible initiation codons. The first is in a suboptimal sequence context thus the site is leaky as a proportion of ribosomes will not initiate translation at this site but will continue scanning until they reach the next AUG codon that is in a more favourable sequence context. The two initiation codons may be in the same reading frame hence the two alternative proteins differ only at the N-terminal region or they may be in different reading frames resulting in two completely different proteins.

2a. Leaky initiation, in-frame



2b. Leaky initiation, out-of-frame

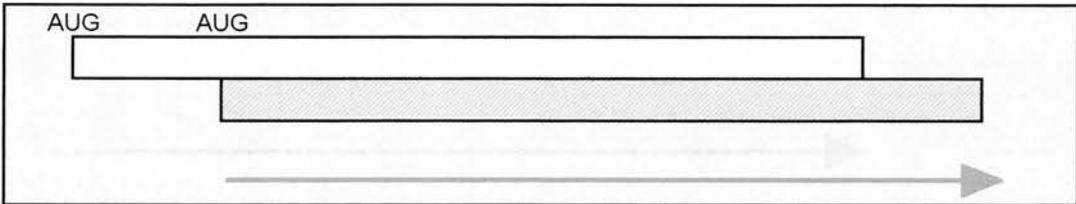


Fig.1.5. Leaky scanning. Alternative proteins can be produced that have different 5' ends if the AUG codons are in-frame (2a) for example, FMDV producing Lab^{PRO} and Lb^{PRO}. Two completely different proteins are produced if the AUG codons are in different reading frames (2b) for example, turnip yellow mosaic virus (TYMV).

1.3.3 Readthrough

This is similar in principle to leaky initiation but in this case it is the termination step that is leaky the ribosomes do not stop at the first termination codon but continue translation producing an extended protein. Suppression of termination is dependent on the sequence context surrounding the termination codons and the presence of a suppressor tRNA.

3. Readthrough

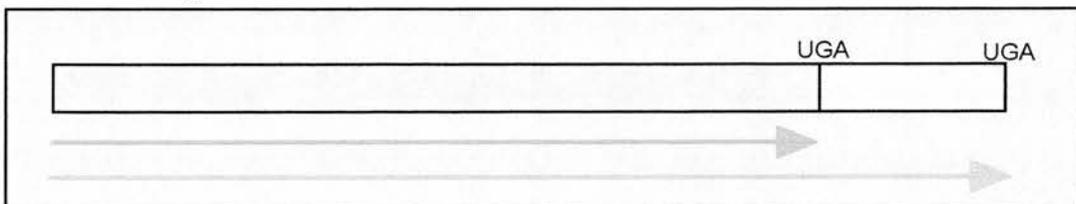


Fig.1.6. A readthrough event at the stop codon produces a second protein with an alternative 3' terminus. The alphaviruses employ this strategy during translation.

1.3.4 Ribosomal frameshift

A frameshift event occurs when the ribosome 'slips' during the elongation process thereby changing the reading frame. The frameshift mechanism has two requirements. The first is an RNA structural element downstream of the frameshift site that causes the ribosome to pause. The second requirement is a region of “slippery” sequence where the codons in the A and P sites are such that the frameshift of +1 or -1 results in a mismatch at the wobble position, thus the two tRNAs in the A and P sites can still interact with the new codons (Pande *et al.*, 1995).

4. Ribosomal frameshift

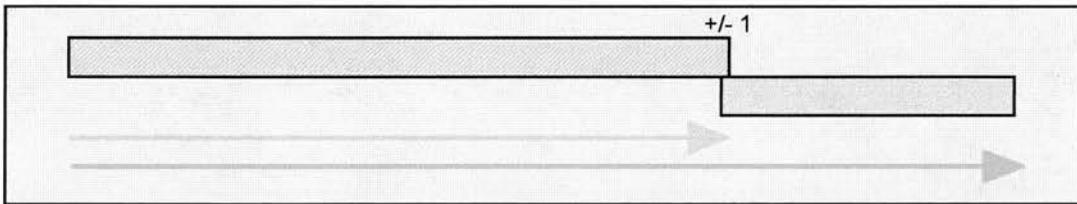


Fig.1.7. A ribosomal frameshift alters the reading frame by +1 or -1 producing an alternative protein. Examples of viruses using a frameshift to produce viral proteins are the retro- and coronaviruses.

1.4 POLYPROTEINS

As previously mentioned polyproteins are large protein precursors that are proteolytically cleaved to produce functional proteins. Polyproteins were primarily identified in viral systems but have now been identified in a wide array of eukaryotic and in some prokaryotic systems. Although polyproteins have been identified in many different systems the production of a polyprotein is still a very rare event. Perhaps the simplest examples are found in regulatory peptides of eukaryotic systems (Herbert and Uhler, 1982). The activity of the functional peptide can be regulated by the cell by controlling processing of an inactive precursor; cleavage of the precursor releases the active peptide. Examples of this are the digestive enzymes including trypsin,

chymotrypsin and pepsin that are initially produced as an inactive zymogen precursor. A polyprotein precursor may contain a single copy of a functional protein as described for the digestive enzymes. Alternatively a polyprotein may encode multiple copies of the same protein as found for the yeast α -mating factor or multiple different proteins as in the viral polyproteins.

1.4.1 PROKARYOTIC POLYPROTEINS

Polyproteins are a rarity in prokaryotic systems but following the discovery of the *E. coli* penicillin G acylase polyprotein (Böck *et al.*, 1983), a select group of polyproteins has been identified in different prokaryotes including *Bacillus subtilis*, *Pseudomonas* sp. and *Bradyrhizobium japonicum* (Thöny-Meyer, Böck and Hennecke, 1992).

E. coli penicillin G acylase (*pac*) catalyses the conversion of penicillin G to phenylacetic acid and 6-aminopenicillanic acid. The enzyme is composed of two subunits α (23kDa) and β (69kDa) that are both encoded by the *pac* gene. Translation of the *pac* gene yields an 840aa polypeptide. The polypeptide is targeted via the N-terminal signal sequence to the periplasm where the signal sequence is catalytically removed, the next step is cleavage at the N-terminus of the β -subunit to release the β -subunit and the α -subunit-spacer, the final step is to remove the spacer peptide releasing the α -subunit. Post-translational processing of the polyprotein occurred in several gram-negative bacterial strains suggesting that the proteolytic activity involved is either autocatalytic or present in bacteria that do not contain a *pac* gene (Schumacher *et al.*, 1986). Homologous acylase polyproteins have been discovered in *Kluyvera citrophila* (Barbero *et al.*, 1986) and *pseudomonas* sp. (Matsuda and Komatsu, 1985).

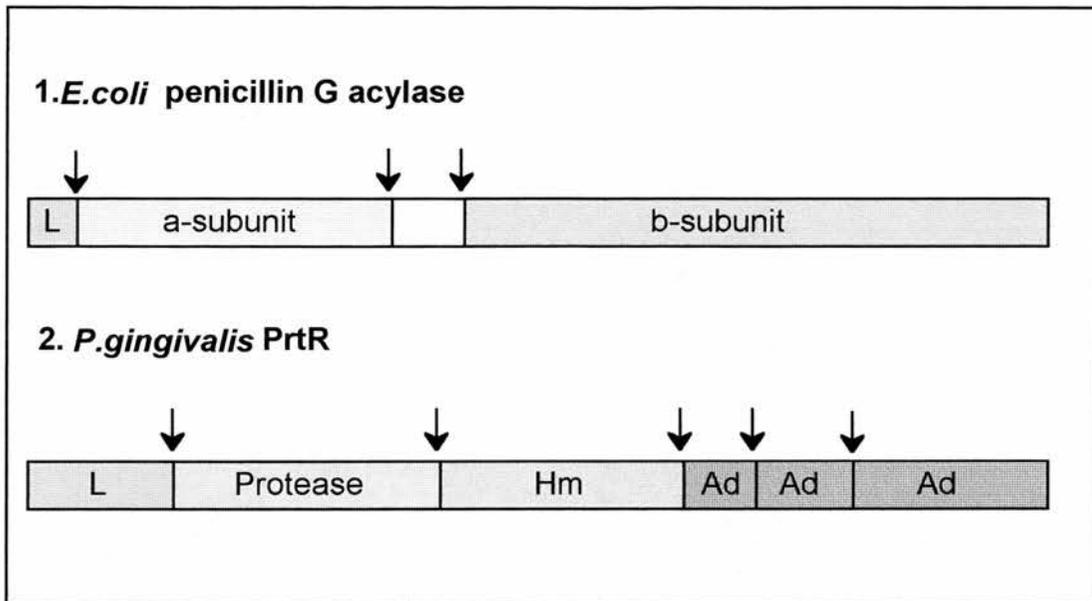


Fig.1.8. Prokaryotic polyproteins.

Porphyromonas gingivalis prtR gene. *PrtR* has an ORF of 5166bp and encodes a polyprotein composed of a short leader sequence, a prosequence followed by an Arg-specific protease and four haemagglutinin/adhesin domains. A lysyl or arginyl residue precedes each domain. The cleaved proteins form a large multiprotein cell-associated complex that combines the ability to bind red-blood cells with a proteolytic activity (Slakeski, Cleal and Reynolds, 1996; Pavloff *et al.*, 1997).

1.4.2 EUKARYOTIC POLYPROTEINS

The majority of polyproteins in eukaryotes encode regulatory proteins such as digestive enzymes, hormones and blood-clotting factors. Many of the polyproteins are very simple involving the production of an inactive precursor which is cleaved to release the functional protein giving the cell a degree of regulation of these potent factors. A number of more complex polyproteins have been identified in eukaryotes they can encode several proteins often involved in the same pathway alternatively the polyprotein may contain multiple copies of a protein. These cellular polyproteins do not have the ability to self-process, they are usually cleaved by cellular proteases. An

exception to this is the gastric enzyme pepsin that is autocatalytic and cleaves itself from the inactive pepsinogen precursor.

Polyproteins, although relatively rare on the whole, have been discovered in a diverse array of eukaryotic organisms from simple yeasts to man. In *Saccharomyces* the prepro- α mating factor gene contains several copies (3-5 depending on species) of a 63bp sequence encoding a 13aa α -factor and an 8aa spacer region (Douglass, Civelli and Herbert, 1984). The silkworm *Bombyx mori* produces a polyprotein precursor that contains the diapause hormone and three other neuropeptides of the FXPRL amide peptide family; homologous genes in other insects have also been identified (Xu *et al.*, 1995; Yamashita, 1996). Antarctic notothenoid fish and Arctic cod both produce antifreeze glycoproteins (AFGP) that enable the fish to withstand the freezing environment that they inhabit. The AFGPs are polymers of a simple glycotriptide. They are encoded by polyproteins containing multiple copies, up to 46, of the glycoproteins linked by tripeptide spacer regions (Chen, DeVries and Cheng, 1997). Polyproteins have also been discovered within the plant kingdom. For example, an *Arabidopsis thaliana* cDNA encodes a polyprotein that contains two thylakoid membrane proteins found to be associated with photosystem II (Mant and Robinson, 1998).

Perhaps one of the most complex eukaryotic polyproteins is a hormone precursor proopiomelanocortin (POMC) that is found in mammals. The prohormone POMC is processed to produce several hormones including adrenocorticotrophic hormone (ACTH), β -endorphin, α -, β -, and γ -melanocyte stimulating hormone (MSH). The polyprotein is processed initially to release the hormones β -lipotropin (β -LPH) and ACTH. These hormonal products can be further processed. ACTH is cleaved to yield α -MSH and corticotropin-like intermediate lobe protein (CLIP). The β -lipotropin precursor can be cleaved to release γ -lipotropin and β -endorphin. POMC mRNA was also found to encode γ -MSH. The pattern of cleavages of the POMC precursor and

therefore the mature products was found to be tissue-dependent (Douglass, Civelli and Herbert, 1984).

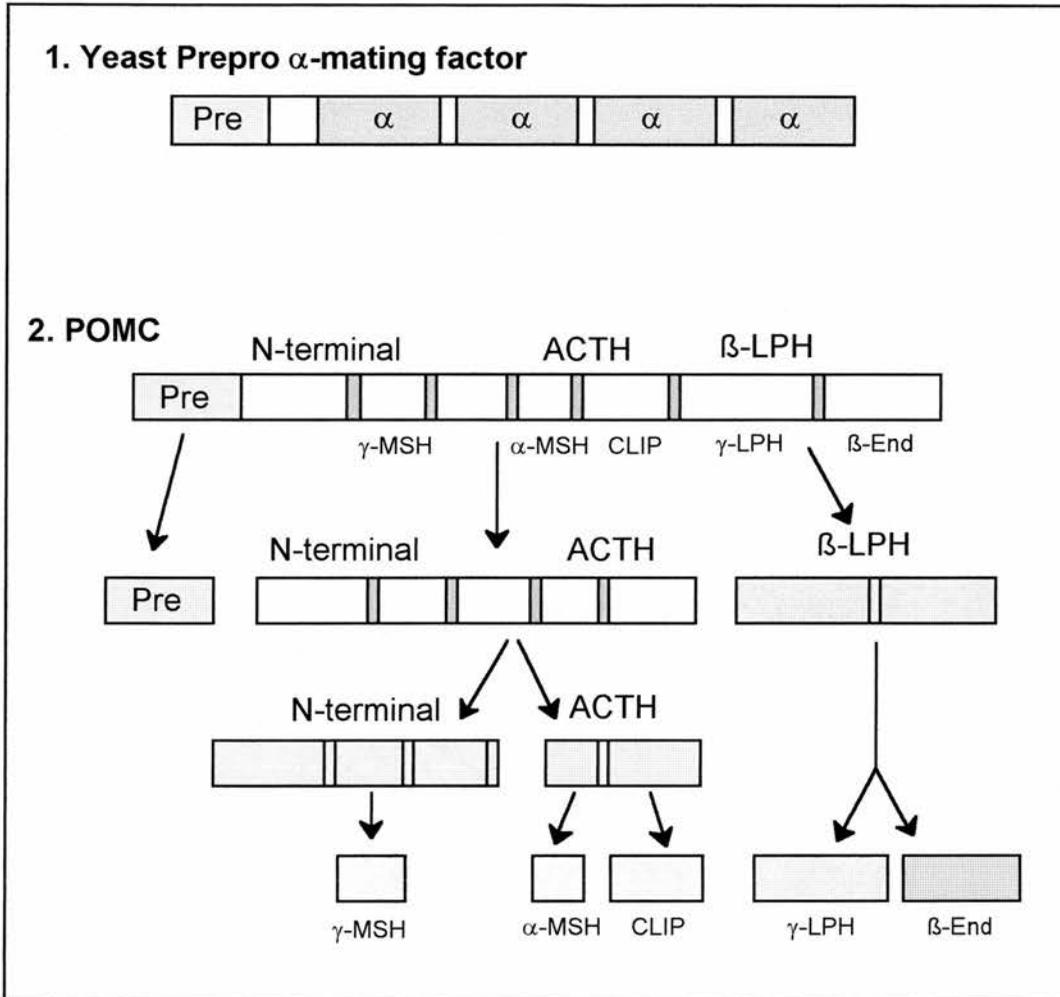


Fig.1.9. Eukaryotic polyproteins. 1. The prepro α -mating factor from yeast. 2. The mammalian prohormone POMC.

1.4.3 VIRAL POLYPROTEINS

The use of polyproteins as a strategy for genome expression is exemplified by the *Picornaviridae* but is also used by many positive-strand RNA viruses, double-strand RNA viruses and retroviruses. The polyprotein strategy allows the virus to encode a large amount of genetic information in a minimal amount of nucleic acid. Production of a single polyprotein eliminates extraneous genetic data including promoter regions, multiple initiation and termination codons and intergenic regions. This is vital for viruses as during infection the virus must produce sufficient functional proteins to

replicate the viral genome and package new viral genomes into virions within the host cell but the amount of genetic material available to the virus is limited by the size of the viral capsid. Processing of the polyprotein permits a degree of control of viral protein production throughout the viral life-cycle. The production of a polyprotein perhaps allows the virus to more readily assemble “macromolecules”. For instance, the capsid proteins can be produced as a polypeptide and transported as a single unit throughout the cell thus the capsid components are readily available for virion morphogenesis when required. Viral polyproteins can be processed by cellular enzymes or by proteases encoded within the polyprotein. A disadvantage of polyprotein processing is that the viral proteins are produced in equimolar amounts even though the virus may only require small amounts of certain proteins, this is overcome by many viruses by combining the production of a polyprotein with alternative strategies (Spall, Shanks and Lomonosoff, 1997).

1.4.3.1 RNA viruses

The use of polyprotein processing as a method of genome expression is very widespread within this group of viruses. The *Flaviviridae* and the picornavirus supergroup all use polyprotein processing to express the viral genome. Other groups of RNA viruses use polyprotein processing in combination with alternative RNA processing strategies to express the viral genome.

1.4.3.2 Comovirus polyprotein processing

Cowpea mosaic virus (CPMV) has a bipartite genome, RNA-1 encodes the replicative proteins and RNA-2 encodes structural proteins and movement proteins. RNA-1 is translated into a 200kDa polyprotein that is co-translationally processed *in cis* by the 24kDa viral protease (Dessens and Lomonosoff, 1991). The RNA-2 has two alternative initiation codons thus producing polyproteins of 105kDa and 95kDa respectively, that differ at the amino-terminus (Holness *et al.*, 1989). The 105kDa (95kDa) polyprotein is also processed by the 24kDa viral protease. Some polyprotein

cleavage reactions require the 32kDa protein product from the 200kDa polyprotein as an accessory protein (Vos *et al.*, 1988). The precursor of the 24kDa protease and the 87kDa protein is stable and is thought to act as the viral replicase.

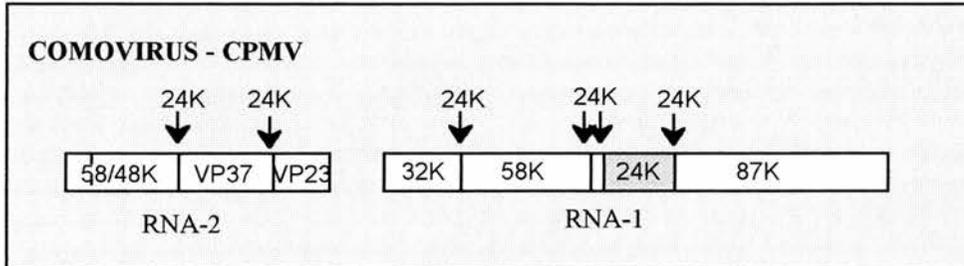


Fig.1.10. CPMV polyprotein. The arrows indicate the cleavage sites processed by the 24K viral protease.

1.4.3.3 Potyvirus polyprotein processing

Tobacco etch virus (TEV) has a monopartite genome that encodes a 351kDa polyprotein. The polyprotein is processed to produce eight mature proteins (Reichmann, Lain and Garcia, 1992). Processing of the polyprotein is controlled by three viral proteases: the P1 protease, the Helper-component (HC) protease and the nuclear-inclusion a (Nla) protease. The P1- and HC-proteases are each responsible for a single autocatalytic cleavage (Carrington *et al.*, 1989; Verchot, Herndon and Carrington, 1991) thus the majority of cleavage events are by the 3C-like Nla protease (Carrington and Dougherty, 1987).

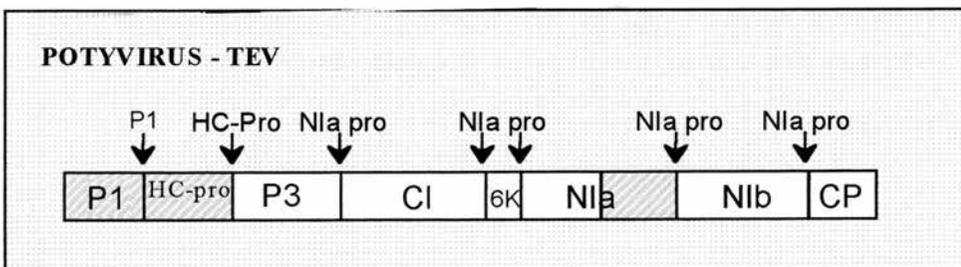


Fig.1.11. TEV Polyprotein. The viral proteases are indicated in orange and the positions of the cleavage sites are shown by the arrows.

1.4.3.4 Hepatitis C virus polyprotein processing

The hepatitis C virus (HCV) genome encodes a single polyprotein that is processed by cellular and virus-encoded proteases (Chambers, Grakoui and Rice, 1990). Cleavage of the structural proteins encoded by the N-terminal domain of the polyprotein occurs via two virus-encoded proteases. The NS2-3 protease catalyses a single autocatalytic event. The N-terminal portion of the NS3 protein contains a serine protease that catalyses the majority of cleavage events. Certain NS3 cleavage events require the viral NS4A protein as a cofactor (Failla, Tomei and DeFrancesco, 1994). The processing events of the C-terminal domain of the polyprotein that encodes the capsid proteins are controlled by cellular proteases.

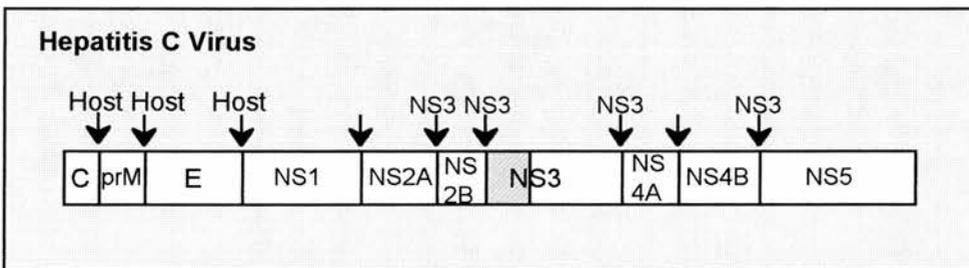


Fig.1.12. HCV Polyprotein. The NS3 proteolytic region is indicated in orange. The sites of the cleavage events mediated by NS3 are indicated. The N-terminal region of the polyprotein is processed by an unidentified host protein.

1.4.3.5 Retroviral polyprotein processing

Retroviruses are RNA viruses that replicate via a DNA intermediate. The retroviral genome encodes three genes in the order 5'-*gag-pol-env*-3'. The *gag* gene encodes the structural proteins. The *pol* region encodes the reverse transcriptase and integrase required for viral replication and the *env* gene encodes the surface and transmembrane proteins. The 35S mRNA is translated to yield the *gag* precursor polyprotein or alternatively a *gag-pol* fusion polyprotein precursor resulting from a ribosomal frameshift at the *gag* termination codon. The *env* gene is translated to produce a polyprotein that is processed by a cellular enzyme. The *gag* and *gag-pol* polyproteins

are processed by the retroviral protease during virion morphogenesis (reviewed by Lowry, 1996).

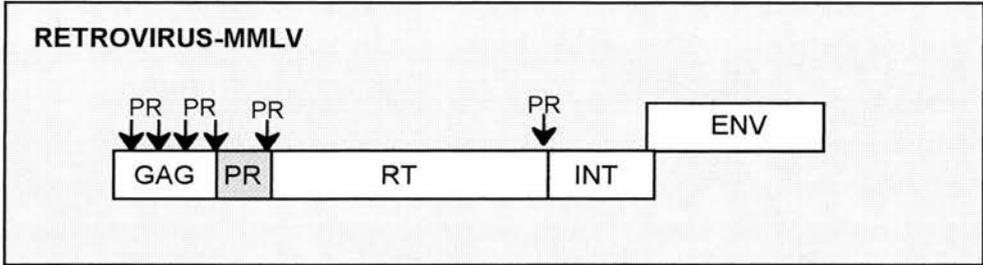


Fig.1.13. MMLV Polyprotein. The viral protease is shown in orange and the cleavage sites are indicated by the arrows.

1.4.3.6 DNA viruses

The first DNA virus to be identified using polyprotein processing as a method of genome expression was African Swine Fever Virus (ASFV). ASFV is a complex DNA virus with a genome of around 170 kbp that encodes about one hundred viral polypeptides (Esteves, Marques and Costa, 1986). Several ASFV structural proteins are produced from a polyprotein precursor. The largest polyprotein precursor pp220 is encoded by a single ORF of 2475aa and is cleaved to produce four mature structural proteins p150, p37, p34 and p14 via an ordered cascade of cleavages (Simón-Mateo, Andrés and Viñuela, 1993). A second polyprotein pp62 is also produced that is cleaved to yield two mature structural proteins p35 and p15 (Simón-Mateo *et al.*, 1997).

In both instances cleavage occurs after the second Gly residue of the consensus sequence Gly-Gly-X. This consensus sequence is also recognized in adenovirus structural proteins and some cellular proteins such as polyubiquitin (López-Otín *et al.*, 1989). In ASFV only some of the putative Gly-Gly-X cleavage sites are cleaved, no further homology between the surrounding amino acid sequence is apparent suggesting that the secondary and tertiary structure of the precursor protein is

important for cleavage. All of the mature proteins produced from polyprotein precursors are located in the core shell of the ASFV particle. The products of pp220 have been shown to be present in equimolar amounts thus they may be required in a strict 1:1 stoichiometry that would be ensured by processing of a polyprotein (Andrés, Simón-Mateo and Viñuela, 1997).

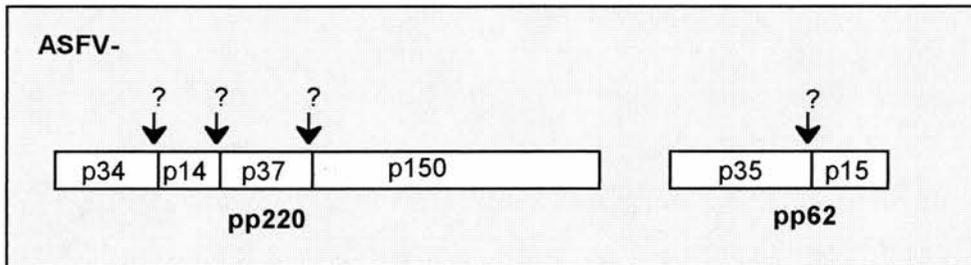


Fig.1.14. Polyproteins from ASFV. The arrows indicate the sites of cleavage but the proteins responsible for the cleavage events are not known.

1.4.4 Picornaviral processing

Processing of the picornavirus polyprotein has three stages; the first is the separation of the structural and nonstructural precursors; the second stage is the release of functional proteins from the polyprotein. The final stage is vital for formation of infectious particles and is the cleavage of VP0 to VP4 and VP2; this step occurs during the final steps of virion morphogenesis. Each stage of processing is catalyzed by different virus-encoded proteolytic activities (Palmenberg, 1987b).

1.4.4.1 Primary processing

The primary cleavage differs between the rhino-, enterovirus group and the cardio- and aphthovirus group of picornaviruses. In both cases the primary cleavage occurs co-translationally, while the polyprotein is nascent on the ribosome before the P3 region has been translated (Jacobson, Asso and Baltimore, 1970). In the rhino- and enteroviruses cleavage occurs at the P1-P2 junction releasing P1, cleavage is catalyzed by the 2A protease *in cis* (Toyoda *et al.*, 1986).

In the cardio- and aphthovirus group cleavage occurs at the 2A/2B junction thus releasing the precursor [P1-2A] (Grubman and Baxt, 1982). The precise cleavage mechanism is still a question of debate. Several lines of evidence suggest that this is not a simple proteolytic cleavage reaction but a novel self-cleavage mechanism. In aphthoviruses the 2A region is very short only 18aa in length, this region is homologous to the C-terminal region of the cardioviruses. Analyses of the cardio- and aphthovirus 2AB regions have not detected any identifiable catalytic motifs. The “cleavage” of the 2A/2B junction occurs at an N-P-G↓P motif (reviewed in Ryan *et al.*, In press).

In other members of the picornavirus family namely the hepatovirus and parechoviruses the primary cleavage appears to be catalyzed by the 3C protease and the 2A region does not have any proteolytic or cleavage capabilities (Jia, Ehrenfeld and Summers, 1991; Schulthies, Kusov and Gauss-Muller, 1994).

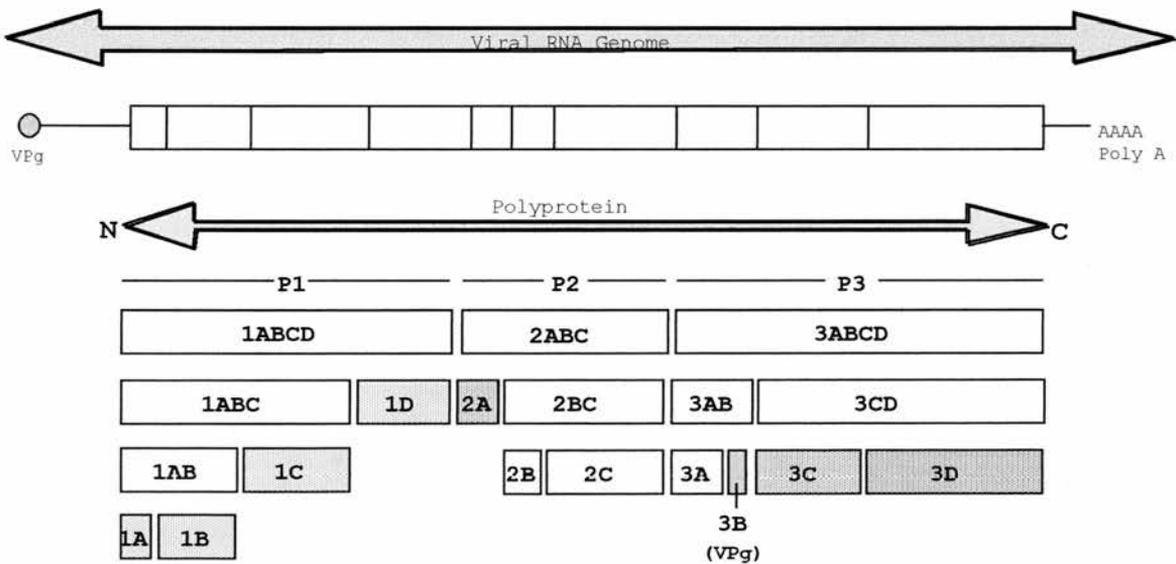


Fig.1.15. Overview of the picornaviral polyprotein and the processing events required to produce the full complement of viral proteins.

1.4.4.2 Secondary processing

In all picornaviruses the majority of cleavage events are processed by the 3C protease or a 3C polypeptide precursor. The 3C protease is highly homologous throughout the Picornaviridae albeit that each 3C protease has a strong preference for homologous substrates. The 3C protease catalyses cleavage of P1, P2 and P3 in a cascade of cleavage events, the precise sequence of cleavage events is not yet fully determined. In cardioviruses, all 3C-containing P3 precursors, i.e. 3ABCD, 3ABC, 3CD and 3C are capable of cleaving the P1 precursor (Parks, Baker and Palmenberg, 1989). The processing of rhino- and enterovirus P1 requires 3CD rather than 3C for efficient cleavage of the VP0-VP3 junction (Jore *et al.*, 1988; Ypma-Wong *et al.*, 1988).

As previously mentioned FMDV possesses an additional proteolytic activity, both cardio- and aphthoviruses contain a leader protein therefore this protein must be separated from the rest of the polyprotein. In cardioviruses, scission at the L-P1 junction is also accomplished via the 3C protease but in aphthoviruses, it has been determined that the L protease has an intrinsic catalytic activity and self-processes its release at the L-1A junction (Strebel and Beck, 1986).

1.4.4.3 Maturation cleavage

The final cleavage event of picornaviruses, the cleavage of 1AB to 1A and 1B is required for production of infectious particles. The maturation cleavage is not catalyzed by any of the viral proteolytic activities identified so far and may be a self-cleaving process. Cleavage occurs upon encapsidation of RNA during the final stages of virion morphogenesis at an N-S dipeptide and occurs within all sixty subunits simultaneously. Production of 1A and 1B is believed to stabilize the capsid structure. One proposed mechanism (although unsubstantiated experimentally) is that a serine residue of 1B and aspartate and asparagine residues located nearby form an analogous conformation to a serine protease active site with the proton-accepting role of the catalytic histidine fulfilled by the viral RNA (Fout *et al.*, 1984). *In vitro* experiments to

test this theory were undertaken by Arnold *et al.* (1987). The potential involvement of viral RNA as the proton abstracting component in the cleavage reaction was upheld as it was found that the addition of diamino compounds resulted in VP0 cleavage.

1.5 PROTEASES

Proteases are enzymes that catalyze the hydrolysis of a peptide bond, many proteases are also capable of catalyzing hydrolysis of an ester bond, this has been used extensively to study the catalytic mechanism.

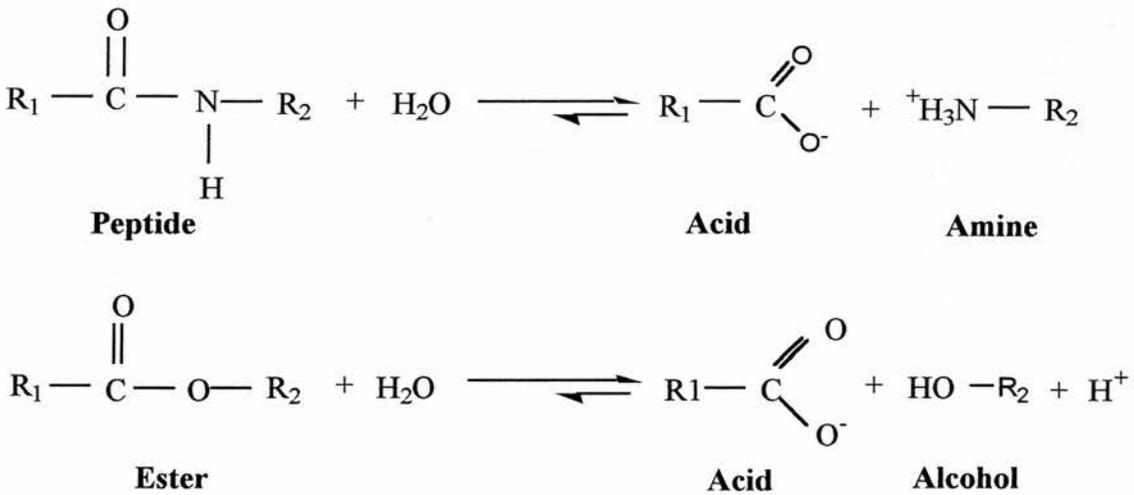


Fig.1.16. Chemical reactions catalyzed by proteases.

Proteases are broadly grouped in accordance with the site of activity; the protease may be an exoprotease that removes residues usually in a progressive manner from the N- or C-terminus of the protein. Alternatively, the protease may be an endoprotease that cleaves between specific amino-acids within the protein substrate.

Protease specificity is defined by the substrate binding pocket that interacts with the substrate residues surrounding the scissile bond, some proteases tolerate a wide array of substrates with very little substrate preference others are highly specific and only cleave a single substrate. The nomenclature used to describe the cleavage site and substrate binding pocket was proposed in 1967 by Schechter and Berger (1967).

Substrate residues on the N-terminal side of the scissile bond are labelled progressively P1, P2, etc. and substrate residues on the C-terminal side of the scissile bond are labelled P1', P2', etc. Correspondingly, residues composing the substrate binding pocket that binds P1, P1' etc. are termed S1, S1' etc.

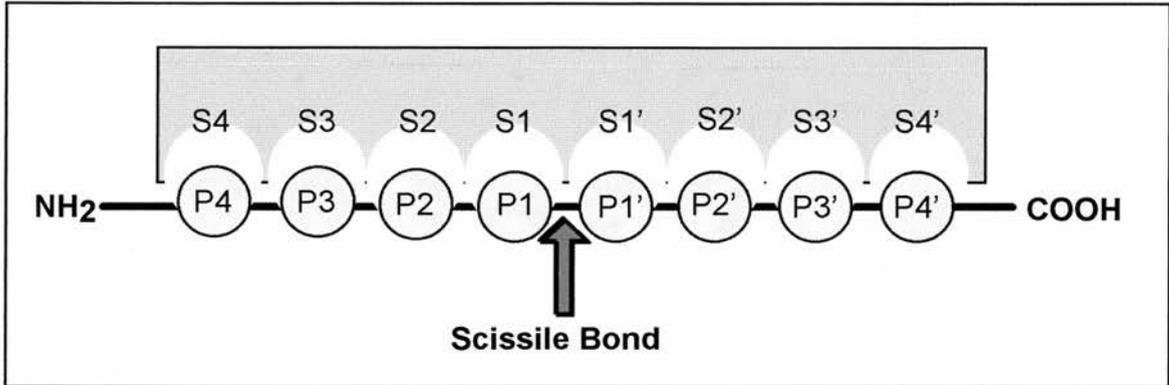


Fig.1.17. Figure demonstrating the nomenclature used to describe the cleavage site and substrate binding pocket.

1.5.1 Cellular proteases

Cellular proteases have been classified into four main categories according to the active site nucleophile these are the serine proteases, cysteine (or thiol) proteases, metalloproteases and aspartic (or acid) proteases. However, a fifth type of protease containing a threonine residue as the active site nucleophile has been discovered (Seemüller, Lupas and Baumeister, 1996). During catalysis the serine and cysteine proteases become covalently bound to the substrate forming intermediary complexes. The metallo- and aspartic protease groups catalyze hydrolysis of the peptide bond without forming covalent bonds with the substrate.

1.5.1.1 Serine proteases

This class of proteases is the most prevalent of the four major classes of proteolytic enzymes. Serine proteases contain a catalytic triad composed of a serine, histidine and aspartate residue; the catalytic mechanism of all serine proteases is constant. An interesting evolutionary aspect is the discovery of two distinct groups of serine protease that share the same catalytic triad and mechanism but the overall sequence

and structure of the two families is very different. Within the chymotrypsin-like family the catalytic triad residues occur in the order His57- Asp102- Ser195. The domain structure is composed primarily of β -structure, turns and loops (Lesk and Fordham, 1996). The smaller subtilisin-like family of serine proteases contain the catalytic residues in the order Asp32 - His64 - Ser221. The overall structure is also different in that it is largely composed of α -helices.

The catalytic mechanism of the serine proteases involves a catalytic relay system composed of the catalytic triad and the 'oxyanion hole'. Hydrolysis of the peptide bond starts with the serine hydroxyl group oxygen atom attacking the scissile bond carbonyl carbon atom. The carbonyl group carbon-oxygen bond becomes a single bond with a negatively charged oxygen atom (an oxyanion). His-57 accepts a proton from Ser-195 thus becoming positively charged. Protonated His-57 is stabilized by negatively-charged Asp-102. His-57 relays the proton to the amine group of the scissile bond that subsequently diffuses away leaving an acyl-enzyme intermediate. This is the acylation step. Water-catalyzed hydrolysis of the acyl-enzyme intermediate reverses the process releasing the enzyme and products (reviewed by Wharton, 1998).

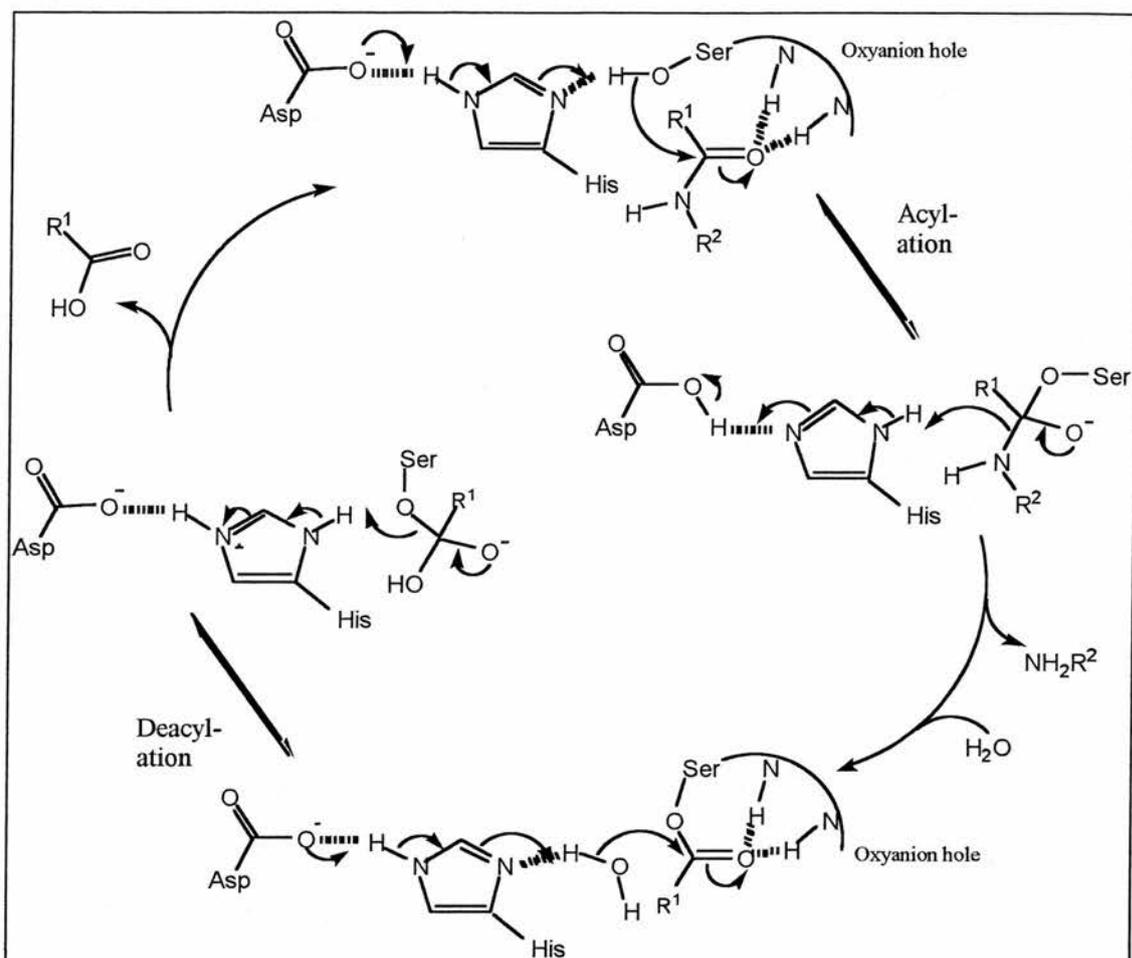


Fig.1.18. *The serine protease catalytic mechanism.*

1.5.1.2 Cysteine proteases

The cysteine protease family can be divided into different groups with related structure and functions. The archetypal cysteine protease is papain. Papain and members of the papain-like family of cysteine proteases are well characterized (reviewed by Berti and Storer, 1995). The active site contains a catalytic dyad of a cysteine and histidine residue. Papain is separated into two domains, R and L, with a deep cleft dividing them. The R domain forms an antiparallel β -sheet barrel with two α -helices lying across the top and bottom of the barrel. The R domain wall of the cleft contains the active site His residue. The active site Cys residue lies on the opposite side of the cleft in the L domain. The L domain is composed mainly of three α -helices and a hydrophobic core. The catalytic mechanism of cysteine proteases is similar to that of

serine proteases with the active site cysteine thiol group as the nucleophile (reviewed by Brocklehurst *et al.*, 1998).

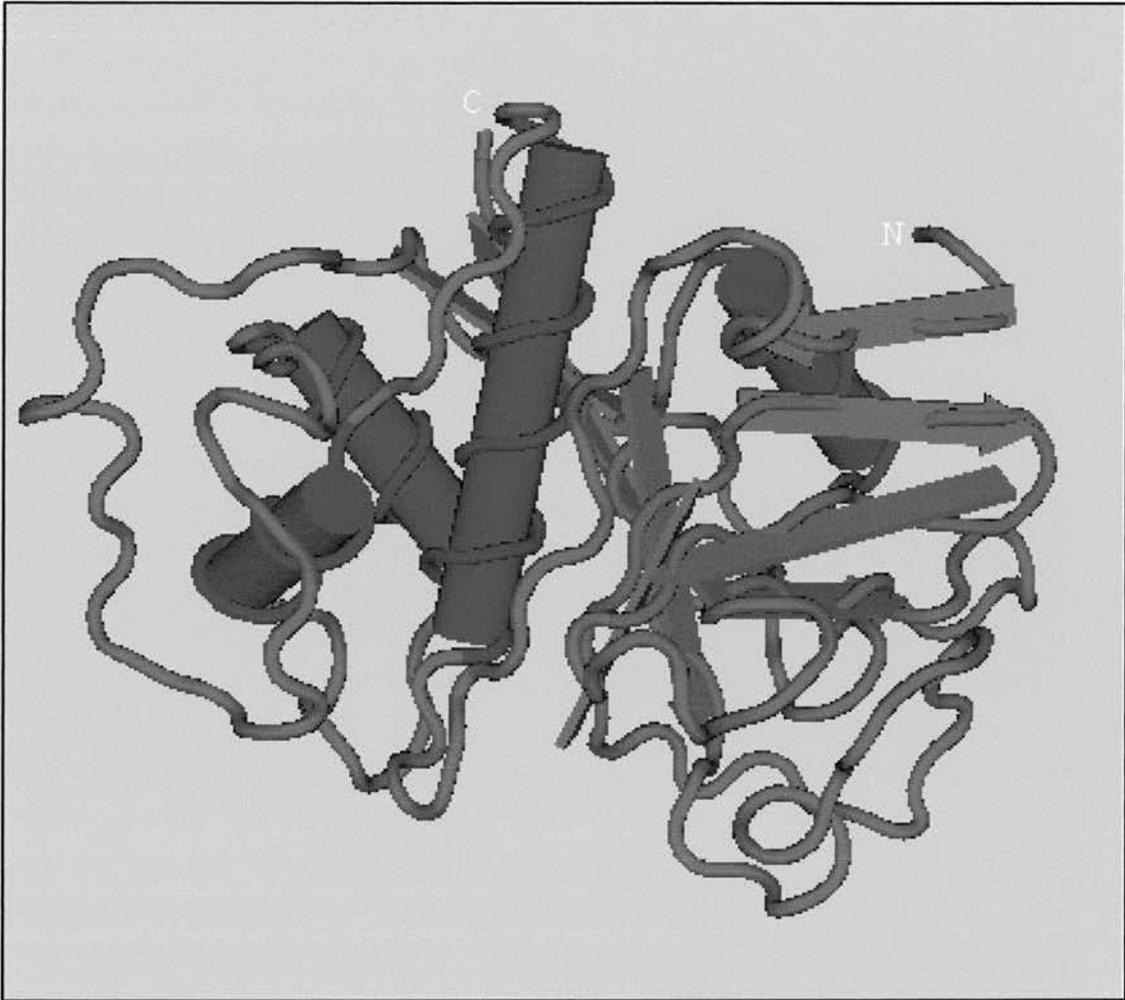


Fig.1.19. Structure of Papain. The secondary structural elements are colour-coded purple for α -helices, blue for β -strands. The L and R domains are easily distinguished, the L domain is composed of three α -helices and the R domain is composed of the anti-parallel β -barrel typical of this group of proteases.

1.5.1.3 Metalloproteases

This class of proteases require a tightly bound metal ion, usually zinc for enzymatic activity. Metalloproteases can be classified according to the nature of the zinc binding site. The zinc ion can be substituted by other divalent cations with no loss of activity, in some cases the presence of an alternative cation enhances enzymatic activity. Therefore the prevalence for zinc is partly due to its biological availability.

In the case of carboxypeptidase A, the model metalloproteinase, the Zn^{2+} ion lies within a groove near the enzyme surface and is bound to two histidine side-chains, one glutamate side-chain and a water molecule. The zinc ion co-ordinates to the carbonyl-group oxygen atom thus making it more susceptible to nucleophilic addition. Glutamate 270 removes a proton from the zinc-bound water molecule. The hydroxyl group attacks the scissile carbonyl group forming a tetrahedral oxyanion that is stabilized by the positively-charged arginine residue (reviewed by Mock, 1998).

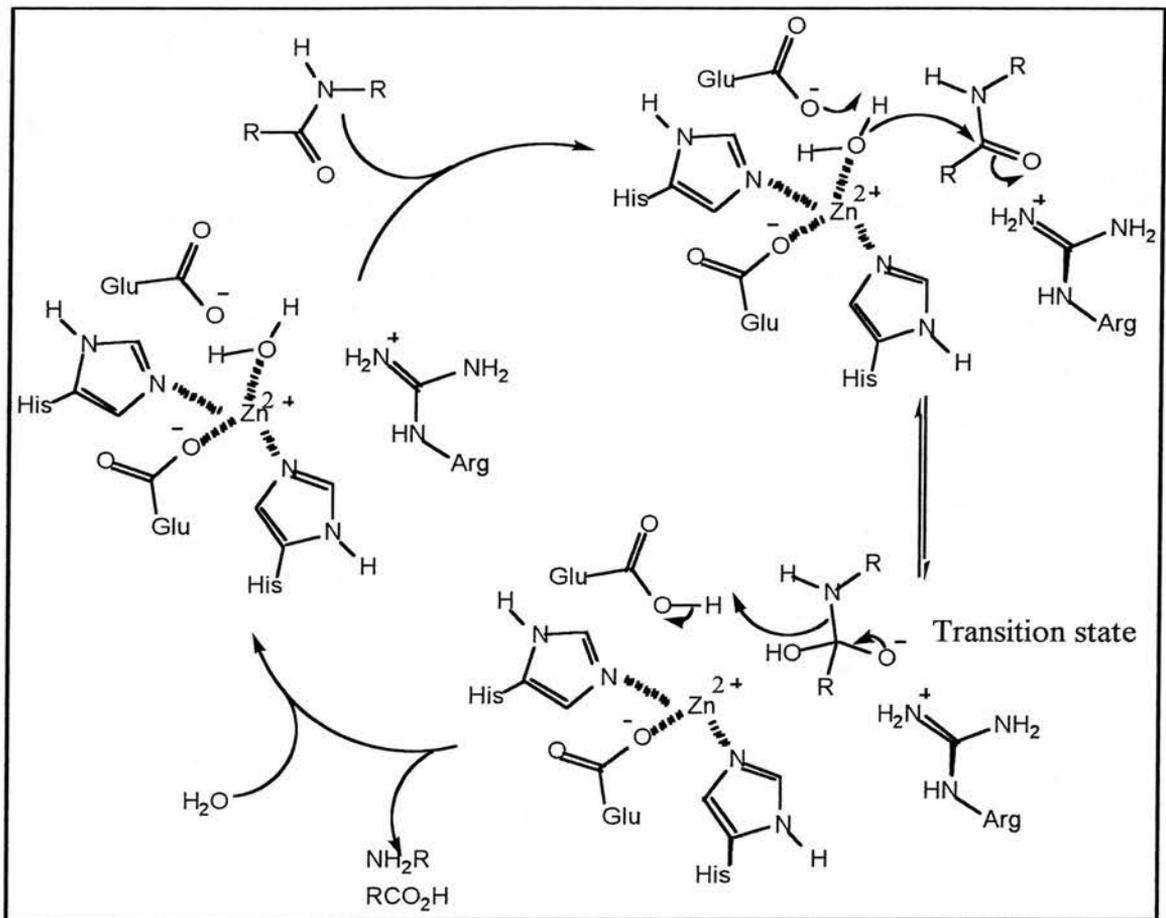


Fig.1.20. Catalytic mechanism of the metalloprotease Carboxypeptidase A.

1.5.1.4 Aspartic proteases

Aspartic proteases are also known as acid proteases because they are usually active in an acidic environment. For example, the classic aspartic protease pepsin is a major digestive enzyme in the stomach with an optimum pH of 2-3. The catalytic dyad is composed of two highly-conserved aspartate residues in the consensus sequence -Asp-Thr (Ser)-Gly-. The proteases are bilobal with an Asp residue in each lobe, the substrate binds in the cleft between the two lobes. In the active site only one of the Asp residues is protonated. The protonated aspartic group protonates the carbonyl oxygen of the scissile bond. The unprotonated aspartic residue removes a proton from a water molecule and transfers the proton to the amine group releasing enzyme and products (reviewed by Meek, 1998).

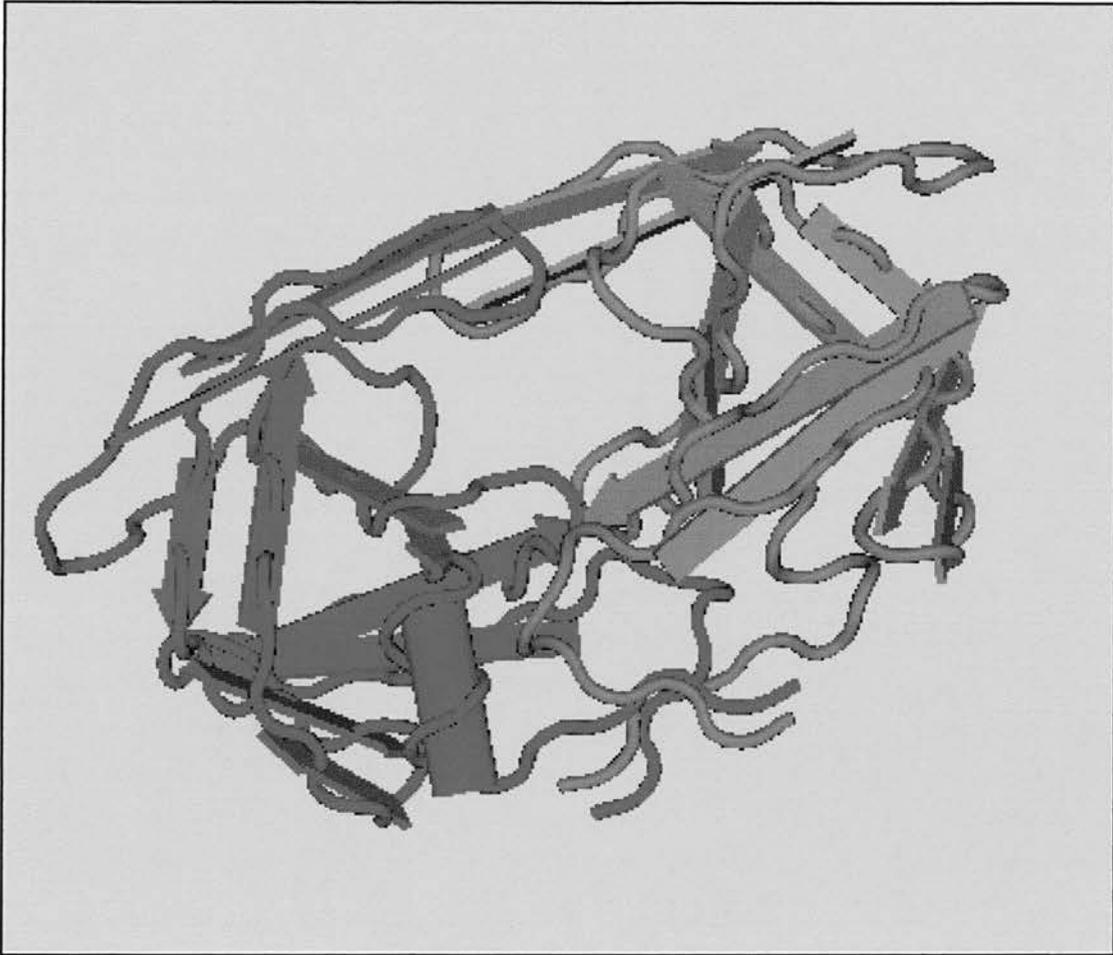


Fig.1.21. Structure of the HIV1 protease, an acid protease. The figure clearly shows the two domains typical of the acid proteases separated by the active site cleft. The secondary structural elements are indicated.

1.5.2 Viral proteases

Virus-encoded proteases are responsible principally for processing large polyprotein precursors into mature, functional proteins. Some viral proteases may also have specific cellular targets that allow the virus to modify the host-cell machinery for viral replication. Another advantage of viral protease production is that the virus is not reliant on the host-cell providing a specific proteolytic activity. Such dependency may influence the host-range capacity of the virus (Babé and Craik, 1997).

Virus-encoded proteases identified to date include examples of all the four major classes of protease. Reviewed in Dougherty and Semler (1993), and more recently in

Ryan and Flint, (1997) and Ryan, Monaghan and Flint, (1998). An example of each type is outlined below.

1.5.2.1 Flavivirus proteases: a serine protease and a metalloprotease?

The flavivirus polyprotein is processed by cellular and virus-encoded proteases. Cellular proteases process the structural proteins and the nonstructural proteins are processed *in cis* and *in trans* by virus-encoded proteases. Hepatitis C virus (HCV) appears to possess two proteolytic activities which may both be associated with the NS3 protein (Gallinari *et al.*, 1998). Sequence analysis suggested that the proteolytic activity was associated with the N-terminal 180aa of NS3 (Bazan and Fletterick, 1989), this region contained the serine protease catalytic triad His-57, Asp-81 and Ser-139 (Miller and Purcell, 1990). NS3 contains the typical chymotrypsin fold composed of two six-stranded β -barrels. It has two unusual features. These are a long N-terminus that interacts with other proteins and a zinc binding site. Crystal structure analysis (Kim *et al.*, 1996; Love *et al.*, 1996) positioned the zinc binding site remote from the active site suggesting a structural rather than a catalytic role for the zinc. The strongly conserved zinc-binding residues Cys-97, Cys-99, Cys-145 and His-149 are mimicked in the entero- and rhinovirus 2A protease that has also been shown to bind a zinc ion (Sommergruber *et al.*, 1994). HCV NS3 protease activity is enhanced by a protein cofactor that maps to a short 14aa region within NS4A (Lin, Thomson and Rice, 1995; Tanji *et al.*, 1995). The cleavage of the NS3/NS4A junction is processed *in cis* by the NS3/NS4A protease. The remaining cleavage reactions at the NS4A/NS4B, NS4B/NS5A, NS5A/NS5B junctions are processed *in trans*.

Mutation of the active site serine residue abolished proteolysis of all the NS3 cleavage sites except the *cis* cleavage of the NS2B/NS3 junction. Deletion studies mapped the proteolytic activity to NS3 and the C-terminal region of NS2 (Wu *et al.*, 1998). Proteolysis of NS2B/NS3 was inhibited by addition of EDTA and stimulated by zinc thus suggesting a zinc protease catalytic mechanism may process this cleavage event

(De Francesco *et al.*, 1996). If the zinc protease activity is confirmed then HCV will possess the only viral metalloprotease identified to date.

1.5.2.2 Potyvirus HC-Pro: a cysteine protease.

The potyvirus HC-Pro protease is one of three potyviral proteolytic activities. The 50-55kDa HC-pro protein is bifunctional; the N-terminal portion contains the helper component activity required for aphid transmission of the virus (Pirone and Thornbury, 1983). The C-terminal region contains the proteolytic activity that processes an autocatalytic reaction releasing the N-terminal domain of the growing polyprotein (Oh and Carrington, 1989). HC-pro has been identified as a cysteine protease that cleaves a Gly-Gly dipeptide. Studies have shown that the TEV HC-pro active site residues are Cys-649 and His-772, analyses of the substrate requirements have revealed that like the potyvirus NIa protease HC-pro also needs an extensive cleavage site of the sequence (Carrington and Herndon, 1992):



There is no evidence to suggest that HC-pro protease cleaves any other substrates and HC-pro does not appear to be able to cleave its substrate *in trans*.



Fig.1.22. Crystal structure of the HCV NS3 protease. The secondary structural elements are colour coded, α -helices in purple and the β -strands in blue. In monomer 1 the Zinc is tetrahedrally coordinated by three Cys residues and a His residue indicated in yellow, in ball-and-stick conformation. In monomer 2 the catalytic residues Asp, Ser and His are shown in green. Note that the Zn-binding site and the active site are positioned on opposite sides of the protease.

1.5.2.3 Retrovirus proteases: the viral aspartic proteases.

The retrovirus family contains all known examples of viral aspartyl proteases. These were initially identified by sequence alignment studies that found a conserved Asp-Thr (Ser)-Gly motif known to be indicative of cellular aspartyl proteases (Tang and Wong, 1987). Unlike cellular aspartyl proteases that are bilobular proteins of about 325aa, the smaller viral aspartyl proteases are only active in a dimer complex. The aspartyl residues of both subunits are required for activity; the active site lies between the two domains (Pearl and Taylor, 1987). In Rous Sarcoma Virus (RSV) the retroviral protease is encoded at the carboxy-terminus of the *gag* polypeptide but in the majority of retroviruses the protease is located at the amino-terminus of the *pol* polypeptide. Therefore production of the viral protease is dependent on the frameshift or termination readthrough event that produces the *gag-pol* polypeptide. This may play a role in regulating cleavage of the capsid proteins that are encoded in the *gag* polypeptide.

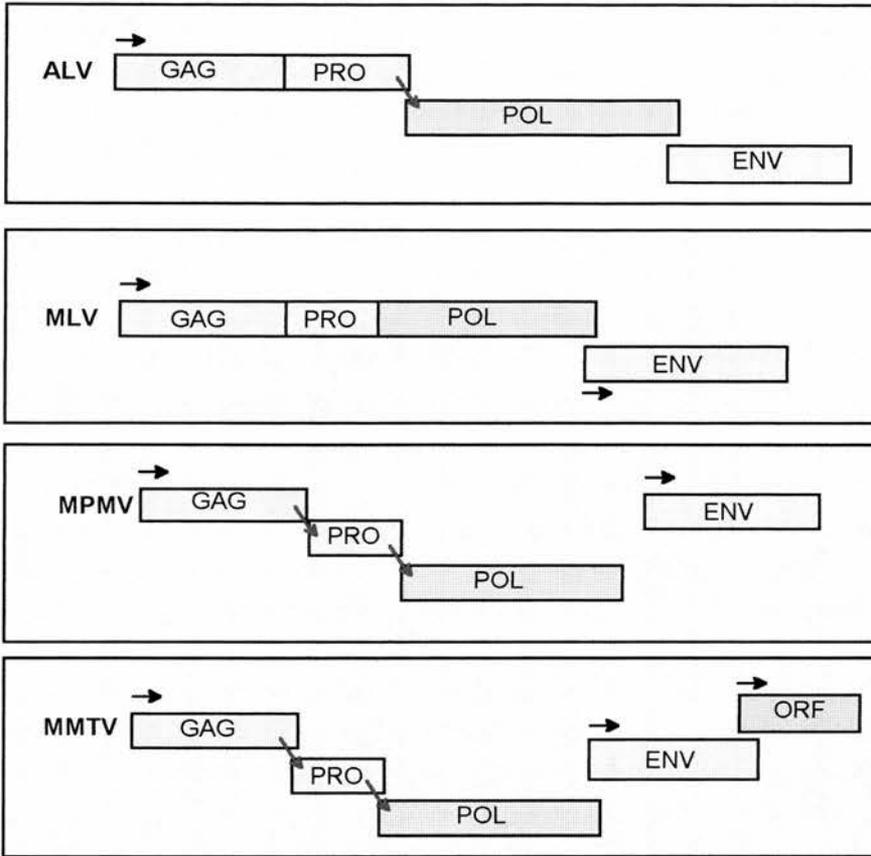


Fig. 1.23. Coding regions of retrovirus genomes. The horizontal arrows indicate sites of translational initiation. The diagonal arrows indicate frameshift sites.

1.5.3. Cis and trans cleavage reactions

Processing of the polyprotein may occur either *in cis* (monomolecular) or *in trans* (bimolecular). *Cis* cleavages are also referred to as autocatalytic; the reactions are usually very rapid and occur co-translationally. *Cis* reactions occur when the protease is attached to the substrate; thus *cis* cleavages are insensitive to dilution and follow zero-order kinetics. *In trans* cleavage reactions occur when the protease cleaves a second protein at a substrate cleavage site therefore, *trans* reactions are sensitive to dilution and follow second-order kinetics. The use of *cis* and *trans* cleavage reactions during processing of the polyprotein may play a role in controlling polyprotein processing. Autocatalytic cleavage reactions do not depend on the concentrations of protease and substrate therefore can occur in the early stages of viral infection when the levels of viral proteins are very low. In the latter stages of infection the relative

concentrations of viral proteins will have greatly increased and thus be more favourable for *trans* processing reactions to occur. Thus the virus may be able to control levels of specific proteins and polypeptide precursors during the different stages of viral replication.

1.6 PICORNAVIRAL PROTEASES

Three virus-encoded proteases have been identified within the picornavirus family. Only one of the proteases is conserved in all members of the picornavirus family this is the 3C protease. Examples of 3C-like proteases are also found in plant viruses that are members of the picornavirus superfamily these include the potyvirus NIa protease and the comovirus 24K protease.

1.6.1 3C protease

The picornavirus 3C protease performs the majority of the cleavage events required for polyprotein processing. Cleavage occurs predominantly at Gln-Gly dipeptides, primary sequence alone is not sufficient for a putative cleavage site as some Gln-Gly bonds in the polyprotein precursor are not cleaved by 3C protease therefore other factors play a role in substrate specificity (Nicklin *et al.*, 1988; Long *et al.*, 1989). The picornavirus 3C protease is inhibited by serine and cysteine protease inhibitors this was one of the first indications that 3C^{PRO} was a novel-type of protease (Pelham, 1978). Sequence alignments indicated that the 3C proteases were similar to the chymotrypsin-like serine proteases (Gorbelenya, Blinov and Donchenko, 1986). Structural predictions suggested a chymotrypsin-like fold with a similar catalytic triad to the serine proteases but the active-site nucleophile was a cysteine residue (reviewed in Malcolm, 1995). Two residues of the predicted catalytic triad His-40 and Cys-147 were confirmed by site-directed mutagenesis these are conserved across picornavirus members. Conclusive evidence of the third member of the catalytic triad remained elusive, two possible residues Asp-85 and Glu-71 (poliovirus numbering) were implicated (Kean *et al.*, 1991; Chernaia *et al.*, 1993; Grubman *et al.*, 1995). The

crystal structures of HRV 3C (Matthews *et al.*, 1994) and HAV 3C (Allaire *et al.*, 1994) did not resolve the issue. The HRV 3C structure indicated that Glu-71 is the third member of the catalytic triad. In contrast, the HAV 3C structure indicates that the active site contains a catalytic dyad of Cys-147 and His-40 similar to the papain group of thiol proteases. The picornavirus 3C proteases fall into two groups with the rhino- and enteroviruses forming a closely related group containing the catalytic triad Cys-147, His-40 and Glu-71. In the second more diverse group that includes the aphtho-, cardio- and hepatoviruses the novel catalytic mechanism remains unresolved.

The 3C protease catalyses the secondary cleavage events of picornaviral polyprotein processing. However, *in vivo* relatively little 3C protease is found. The majority of 3C^{Pro} remains uncleaved from 3D the RNA-dependent polymerase. 3CD was found to have proteolytic activity but no polymerase activity (Harris *et al.*, 1992). In poliovirus, the precursor 3CD^{Pro} is an absolute requirement for efficient processing of the P1 polypeptide but P2 and P3 processing can be catalyzed by 3C^{Pro} alone. 3CD^{Pro} is cleaved at a Gln-Gly pair by 3C^{Pro} *in trans* to release 3C^{Pro} and 3DP^{ol}. An alternative cleavage pathway catalyzed by 2A^{Pro} cleaves within the 3D region at a Tyr-Gly pair to produce the alternative products 3C' and 3D'. The *in vivo* levels of these alternative cleavage products are relatively high and while 3C' and 3D' are not required for viral replication the role of these alternative products has not been elucidated.

1.6.1.1 Cellular targets of 3C^{Pro}

Comparison of poliovirus and mock-infected cells by two-dimensional gel analysis identified less than ten proteins altered by infection (Urzainqui and Carrasco, 1989). Poliovirus infection inhibits transcription by all three polymerase systems. Pol I inhibition is observed 1 to 2h post-infection, pol II-mediated transcription is apparent 4h post-infection and finally pol III-mediated transcription is inhibited at 5h post-

infection (Rubinstein and Dasgupta, 1989; Fradkin *et al.*, 1987). 3C^{pro} has been implicated in the inhibition of two of these transcription systems.

The pol III transcription system requires polymerase III and two transcription factors TFIIB and TFIIC. In poliovirus-infected HeLa cells pol III activity is unchanged but there is a slight decrease in TFIIB activity and a dramatic decrease in TFIIC activity. 3C^{pro} has been shown to cleave transcriptionally active TFIIC to an inactive form *in vivo* and *in vitro* (Clark *et al.*, 1991). In order for 3C^{pro} to cleave TFIIC it must enter the nucleus, 3C^{pro} is a small protein of 20kDa therefore it may diffuse into the nucleus alternatively the precursor 3CD^{pro} is known to enter the nucleus following infection (Fernandez-Tomas, 1982). TFIIC cleavage is not observed in FMDV-infected cells although inhibition of transcription is observed due to degradation of histone H3 (Falk *et al.*, 1990).

The pol II transcription system also requires several transcription factors, pol II inhibition by poliovirus correlates with reduced TFIID activity. The 3C^{pro} has been shown to directly cleave the DNA-binding component of TFIID the TATA-binding protein (TBP) *in vivo* and *in vitro*. Cleavage of TBP is detected 4h postinfection that correlates with the time-scale for pol II inhibition (Clark *et al.*, 1993). The 3C^{pro} has also been shown to cleave the transcription factor CREB (Yalamanchili, Datta and Dasgupta, 1997).

Another cellular target for 3C^{pro} and 3CD^{pro} cleavage is the microtubule-associated protein-4 (MAP-4). *In vitro* cleavage products of MAP-4 correspond to the products seen *in vivo*. Mutation of 3C^{pro} abolishes MAP-4 cleavage. Cleavage of MAP-4 is associated with breakdown of the microtubular network (Joachims, Harris and Etchison, 1995). Finally, Belsham *et al.* have also recently proposed that FMDV 3C^{pro} can cleave the eukaryotic initiation factors eIF4A and eIF4G (Belsham, McInerney and Ross-Smith, 2000).

1.6.2 Picornavirus 2A protein

The picornaviral proteins are mostly homologous across the different genera, an exception to this is the 2A region. The picornaviral 2A region can be grouped into three distinct classes that share homology within the 2A locus.

1.6.2.1 The rhinovirus and enterovirus 2A protein.

The 2A protease catalyses the primary cleavage event at the P1-2A junction that separates the structural and nonstructural protein precursors. The 2A protease cleaves a tyrosine-glycine bond. The cleavage event was proposed to occur co-translationally while the protein is still on the ribosome but recent evidence from IRES-screening studies suggests that this may not be the case (Paul *et al.*, 1998).

Sequence alignments indicated that 2A^{Pro} is similar to the subtilisin-like group of serine proteases (Bazan and Fletterick, 1989). Inhibitor studies found that 2A^{Pro} is inhibited by cysteine protease inhibitors including iodoacetamide and N-ethylmaleimide (Yu and Lloyd, 1991). Therefore it has been proposed that like picornaviral 3C proteases the 2A^{Pro} is structurally a serine protease but the active site serine nucleophile is replaced by a cysteine residue. The catalytic triad is composed of His-20, Asp-38 and Cys-109 this has been confirmed by the crystal structure of HRV2 2A protease (Petersen *et al.*, 1999) and site-directed mutagenesis studies (Yu and Lloyd, 1992). The 2A protease contains a tightly bound zinc ion that is co-ordinately bound by four highly conserved residues (Cys-52, Cys-54, Cys-112 and His-114) (Sommergruber *et al.*, 1994). The crystal structure confirms that the zinc ion plays a structural rather than a catalytic role stabilizing the N-terminal domain. A similar zinc binding site is present in the HCV NS3 protease, discussed in section 1.5.2.1 refer to Fig 1.22.

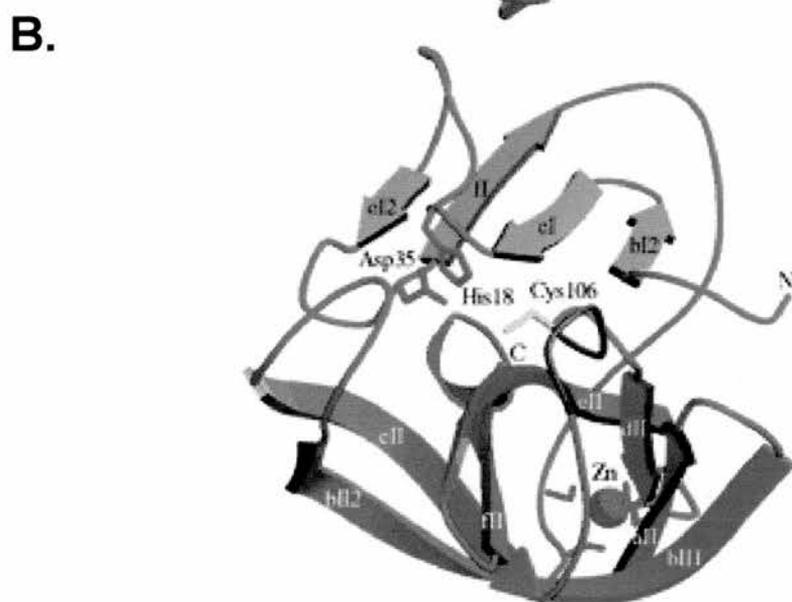
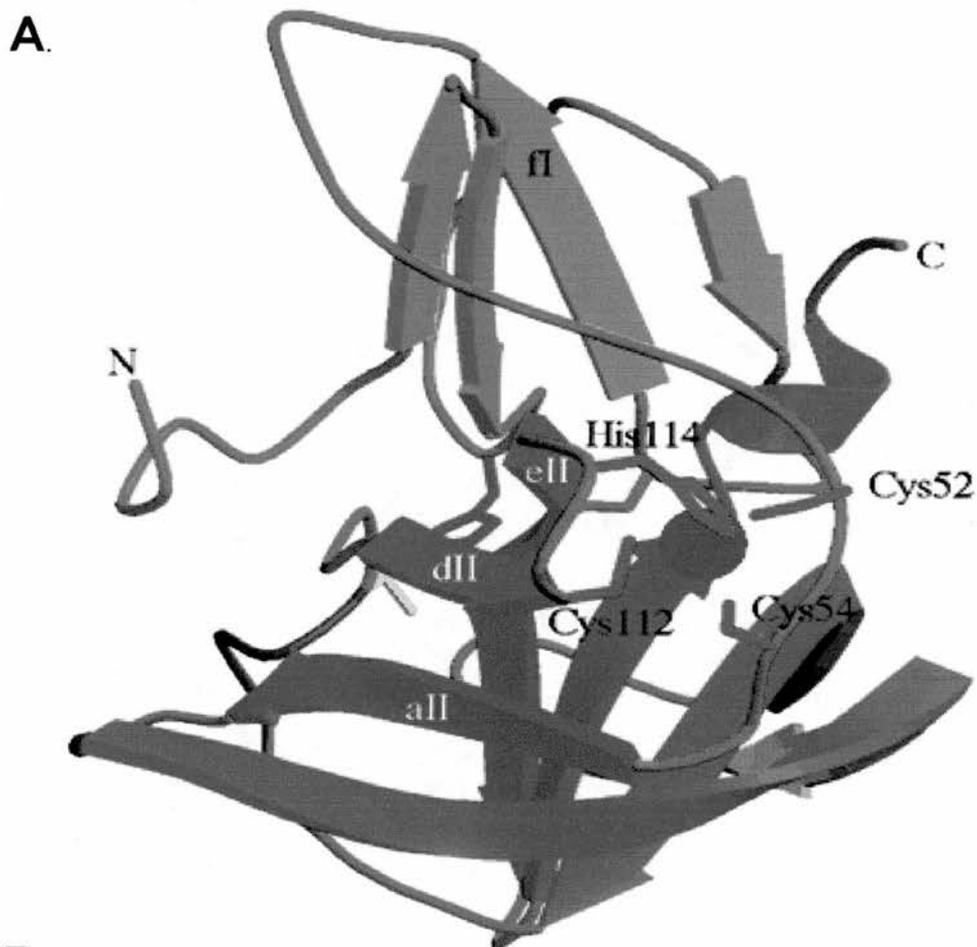


Fig.1.24. Structure of HRV2 2A protease. A shows the zinc binding site, the co-ordinating residues are labelled. B shows the overall structure of HRV2 2A protease, the β -strands of the N-terminal domain are shown in turquoise and the β -strands of the C-terminal domain in blue. The catalytic residues are labelled and shown in ball-and-stick form. Adapted from Petersen et al., 1999.

The 2A protease has a preferred rather than an absolute substrate requirement therefore it is proposed that the substrate structure is of primary importance rather than the surrounding primary sequence. The substrate requirement for *cis* processing is different to that for *trans* processing. The *cis* cleavage reaction has a strict requirement for Gly at the P1' site but is tolerant of residues in other positions surrounding the scissile bond. Mutagenic studies show that only Thr/Ser/Asn or Arg can be tolerated at the P2 position, there is also a preference for an aliphatic residue in the P4 position (Skern *et al.*, 1991; Wang, Sommergruber and Johnson, 1997). The general consensus sequence for a 2A^{PRO} cleavage site is X-L-X-T-X-G, with X-G as the scissile bond (Ventoso, Barco and Carrasco, 1999).

HRV 2	1	- I V T R P I I T A G P S D M Y V H V G	19
HRV 89	1	R P D V F T V T N V G P S S M F V H V G	20
HRV 1B	1	I V P R A S M K T V G P S D L Y V H V G	20
HRV 9	1	R D N V R A V K N V G P S D M Y V H V G	20
HRV 85	1	L K E R A S L T T A G P S D M Y V H V G	20
HRV 14	1	K K R K G D I K S Y G L G P R Y G G I Y	20
POLIO 1	1	P L S T K D L T T Y G F G H Q N K A V Y	20
POLIO 2	1	P L P E K G L T T Y G F G H Q N K A V Y	20
POLIO 3	1	P L S E K G L T T Y G F G H Q N K A V Y	20
COX B1	1	T T T R S N I T T T G A F G Q Q S G A V	20
COX B3	1	R Q S I Y Y M I N T G A I W T T I R G S	20
COX B4	1	T A E R A S L I T T G P Y G H Q S G A V	20

Fig.1.25. Sequence alignment of the VP1-2A cleavage site.

1.6.2.2 Other cleavage sites for the 2A^{PRO}.

The entero- and rhinovirus 2A protease also processes a *trans* cleavage reaction in the polyprotein. The polypeptide 3CD^{PRO} can be processed by 3C^{PRO} *in trans* to release 3C^{PRO} and 3DP^{ol} or 2A^{PRO} can cleave at a Tyr-Gly dipeptide at a site within the N-terminal region of 3D to produce the alternative cleavage products 3C' and 3D' (Lee and Wimmer, 1988). The reason for the existence of this alternative cleavage pathway is not understood. The 3C' and 3D' products are not required for viral replication and a role for these products has not been elucidated.

2A^{PRO} has been implicated in cleaving a number of host-cell proteins (Ventoso, Barco and Carrasco, 1998). The most important of these cellular substrates is the initiation factor eIF4G, a component of the eIF4F initiation complex. Cleavage of the initiation

factor eIF4G leads to shut-down of the host cells' translational machinery (Kräusslich *et al.*, 1987; Liebig *et al.*, 1993). This will be discussed in section 1.9.1.

Expression of the 2A protease in COS-1 cells found that translation and transcription were inhibited indicating that 2A^{PRO} may interfere with cellular gene expression at both levels (Davies *et al.*, 1991). Inhibition of transcription by the 2A protease may be direct via cleavage of a cellular protein or it may be a by-product of translational inhibition. Yalamanchili, Banerjee and Dasgupta (1997) demonstrated that 2A^{PRO} cleaves the TATA-binding protein (TBP) *in vitro*. TBP is cleaved by the picornaviral 3C protease resulting in inhibition of RNA pol II-mediated transcription, refer to section 1.6.1.1. Interestingly, 2A^{PRO} cleavage of TBP did not inhibit transcription by RNA Pol II. The 2A^{PRO} cleaves TBP at position 34 thus only the N-terminal residues are removed. The role of the TBP N-terminal region is not clear, these results suggest that it is not required for pol II transcription. It is possible that 2A^{PRO}-mediated TBP cleavage may have a role in inhibition of RNA pol I or RNA pol III transcription as TBP is a component of both complexes.

During poliovirus infection the translational regulator protein poly(A) binding protein (PABP) is degraded. Preliminary studies *in vitro* indicate that the viral proteases 2A^{PRO} and 3C^{PRO} are both capable of cleaving PABP. The 2A^{PRO} and 3C^{PRO} cleavage sites both map to the C-terminal region of PABP (Joachims, Van Breuchel and Lloyd, 1999; Kerekatte *et al.*, 1999). PABP interacts with the eIF4F complex during translation thus it has been postulated that PABP cleavage may be involved in inhibition of cellular translation. The eIF4G binding domain is within the N-terminal region and is not disrupted by 2A/3C protease cleavage. The role of PABP cleavage during poliovirus infection and the effects on the cell remains unclear.

A novel target of coxsackievirus 2A protease has recently been identified. Infection by coxsackievirus can cause dilated cardiomyopathy although the viral mechanism involved is not clear. Sequence analysis predicted two possible 2A^{PRO} cleavage sites

within the cytoskeletal protein dystrophin. Studies *in vitro* and *in vivo* confirmed that 2A^{PRO} cleaves dystrophin at these sites. Dystrophin is a large protein (427kDa) composed of four domains separated by hinge regions. Cleavage by 2A^{PRO} separates the N-terminal actin-binding site from the rod domain and the C-terminal β -dystroglycan binding domain thereby disrupting the interaction between the internal cytoskeleton and the external basement membrane (Badorff *et al.*, 1999).

1.6.2.3 *The cardiovirus and aphthovirus 2A protein.*

The cardio- and aphthovirus primary cleavage occurs at the 2A/2B junction thereby generating P1-2A in aphthovirus and L-P1-2A in cardiovirus. P1 is released by the 3C protease that subsequently cleaves the P1-2A junction. The aphthovirus 2A region is very short only about 18aa in length but the cardiovirus 2A region is similar in size to the rhino- and enterovirus 2A region. Interestingly, only the C-terminal region of 2A is strongly conserved in cardioviruses. The C-terminal region was found to be very similar to FMDV 2A with complete conservation of the C-terminal residues -NPG- and the 2B proline N-terminal residue (Donnelly *et al.*, 1997). The conserved region of cardiovirus 2A does not contain any known protease motifs.

Experimental evidence suggests that no other picornaviral proteins are involved in cleavage of the 2A/2B junction. The 18aa FMDV 2A region and the C-terminal residues of cardiovirus in conjunction with the N-terminal 2B residue are able to mediate co-translational cleavage when inserted into an artificial polyprotein system (Ryan and Drew, 1994).

1.6.2.4 *Parechovirus, Avian Encephalomyelitis Virus, and Aichi virus 2A proteins.*

The 2A proteins in this group of viruses are not believed to be involved in polyprotein processing. The primary cleavage in hepatoviruses separating the P1 precursor is mediated by the 3C protease. Until recently there was no evidence of any similarity between the 2A proteins of these viruses. Three regions of similarity have been

identified, His-24 and a KNCET motif at positions' 91-95 (HPeV1 numbering) are strongly conserved, the motif is equidistant from a long stretch of hydrophobic residues that is a putative transmembrane domain. Database analysis has revealed similar patterns in four cellular proteins (Hughes and Stanway, 2000). The precise function of the cellular proteins is not fully understood although three of the cellular proteins have been described as class II tumour suppressors that act to reduce cell proliferation thus it is postulated that the viral 2A protein may also manipulate cell functions to aid viral replication.

1.6.3 Aphthovirus Leader protease.

Aphthoviruses and equine rhinitis viruses serotypes 1 and 2 are unique among the picornavirus family in that they possess a protease at the N-terminus of the viral genome. Cardioviruses also encode a protein at this position but it does not contain any proteolytic activity. The aphthovirus leader or L protease is found in two forms termed Lab^{PRO} and Lb^{PRO}. This is due to the presence of two in-frame AUG codons 84aa apart producing Lab^{PRO} and Lb^{PRO} respectively (Sangar *et al.*, 1987). The sequence context surrounding the first AUG codon is suboptimal thus a number of ribosomes fail to initiate translation at this position and continue scanning until the second AUG codon that is within a favourable sequence context.

The two forms of the leader protease have both been demonstrated to be functional for the activities attributed to the L protease (Medina *et al.*, 1993). Intriguingly, studies indicate that the Lb form is necessary for production of live virus (Cao *et al.*, 1995) although FMDV Lb^{PRO} is not required for viral replication (Piccone *et al.*, 1995a). Initial sequence alignment proposed that L^{PRO} was a papain-like cysteine protease this was supported by inhibitor studies that found that L protease was inhibited by E-64 a characteristic inhibitor of this group of proteases (Kleina and Grubman, 1992). The catalytic dyad is composed of residues Cys-51 and His-148 (Roberts and Belsham, 1995; Piccone *et al.*, 1995b). The His residue is in a similar context (EHAV) to the

active site Cys residue of papain (DHAV). Sequence analysis of all known L protease cleavage sites has failed to identify a consensus primary sequence requirement therefore it is postulated that L^{PRO} substrate requirements are predominantly structural.

The crystal structure of Lb^{PRO} solved by Guarné *et al.* (1998) identified two subdomains that are equivalent to the left- and right-hand domains of the papain-fold. The N-terminal domain is composed of four α -helices and two short antiparallel β -strands and the C-terminal domain is composed of a mixed β -sheet structure that is common to all β -fold proteins (Guarné *et al.*, 1998; Skern, Fita and Guarné, 1998). Due to the discrepancy between the sizes of the FMDV L protease and papain the linking strands between the conserved secondary structure elements are shorter in the Lb^{PRO}. Unlike papain the Lb^{PRO} structure is unusual in that it contains a long, flexible C-terminal extension.

1.6.3.1 Cleavage activities of L^{PRO}

The L protease autocatalytically separates itself from the N-terminus of the polyprotein thus exposing the myristoylation site of the P1 capsid precursor. Inefficient cleavage of the L-P1 junction could interfere with formation of the capsid and virion assembly. Cleavage of the L-P1 junction is efficiently processed *in cis* and *in trans*. Both forms of the FMDV L^{PRO} also cleave a cellular substrate the eIF4G protein thereby resulting in shut-off of host-cell translation (Devaney *et al.*, 1988; Kirchweger *et al.*, 1994). This is discussed in section 1.9.1.

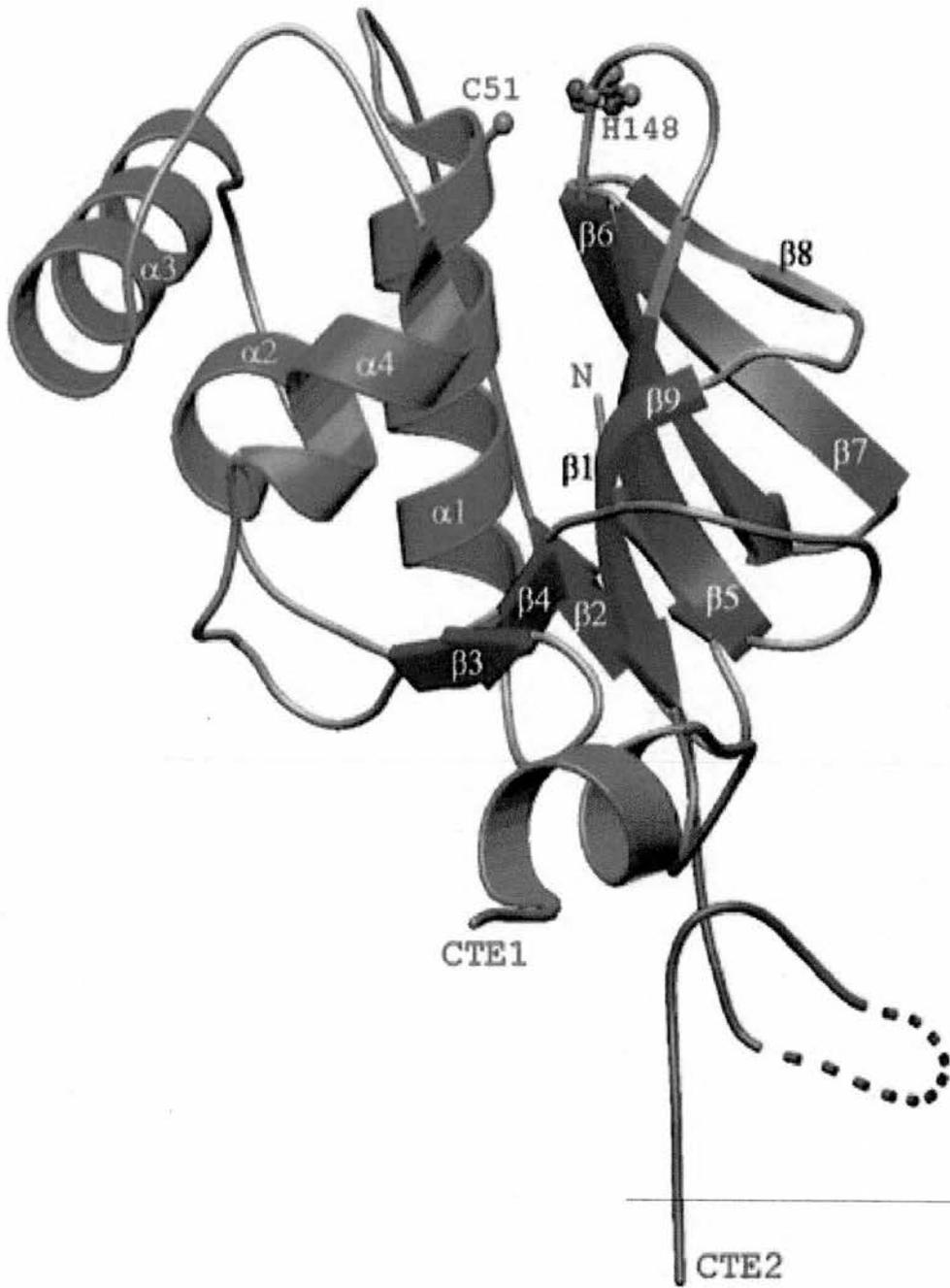


Fig.1.26. Structure of FMDV Lbpro. The secondary structure elements are shown, α -helices in green and β -strands in purple. The ordered C-terminal element (CTE 1) is indicated in orange with the disordered C-terminal element (CTE 2) is shown in blue. The active site residues Cys51 and His148 are shown as ball and sticks. Adapted from Guarné et al., 1998.

1.7 OVERVIEW OF TRANSLATION

The process of decoding the mRNA to synthesize protein occurs in three phases initiation, elongation and termination. During the initiation phase the initiation codon (AUG) is identified and the ribosomal complex is formed. In the elongation phase all the codons are translated sequentially thus extending the nascent polypeptide. Finally, when a termination codon (UAA, UGA or UAG) is encountered, elongation ceases, the complete polypeptide is released and the ribosomal complex dissociates.

1.7.1 Principle differences between prokaryotic and eukaryotic translation.

The overall process of translation is very similar in prokaryotic and eukaryotic cells although the precise mechanisms are somewhat different. By definition prokaryotic cells lack a membrane bound nucleus thus the processes of transcription and translation are not separated. Transcription and translation are closely coupled and translation can begin before the transcript is fully synthesized. In contrast, the transcription process in eukaryotic cells occurs in the nucleus. The nascent RNA is extensively modified to produce the final mRNA that is transported across the nuclear membrane into the cytoplasm where the translation process occurs. The elongation and termination phases are largely similar between prokaryotes and eukaryotes, the major differences occur in the initiation phase this is possibly due to the difference in the starting material (mRNA) for translation between the two systems. Comprehensive reviews include Moldave (1985), Merrick, (1992) and Browning, (1996).

EUKARYOTIC CELL

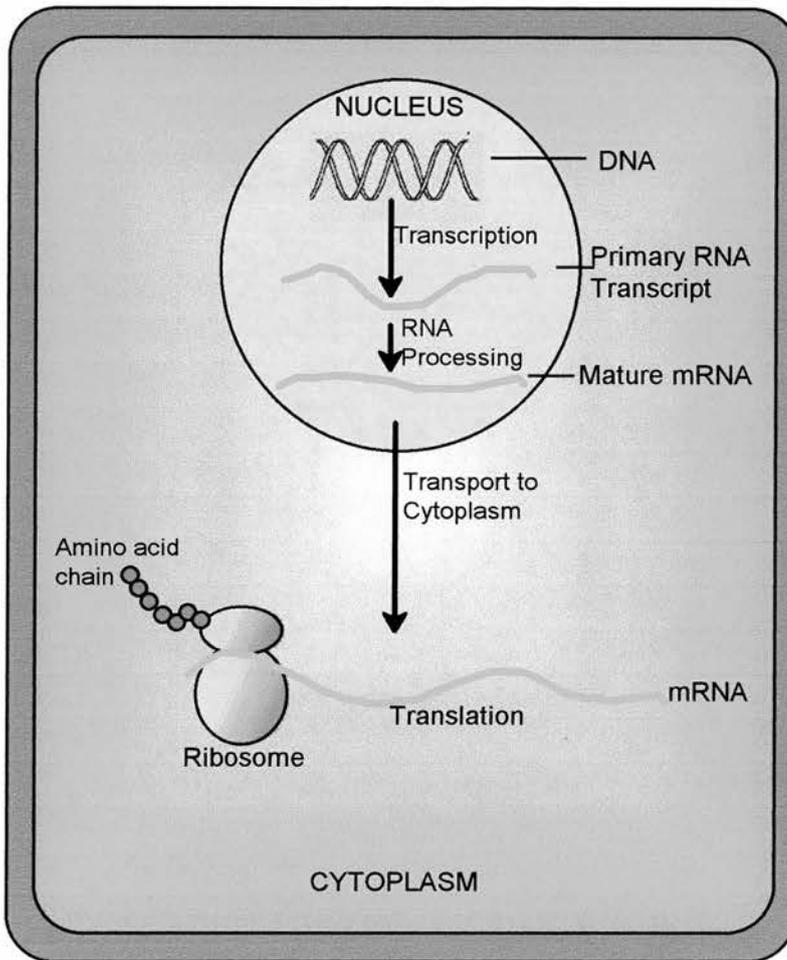


Fig.1.27. The eukaryotic cell. Transcription occurs in the nucleus, the RNA is processed to produce mature mRNA. The mRNA is transported across the nuclear membrane to the cytoplasm the site of translation.

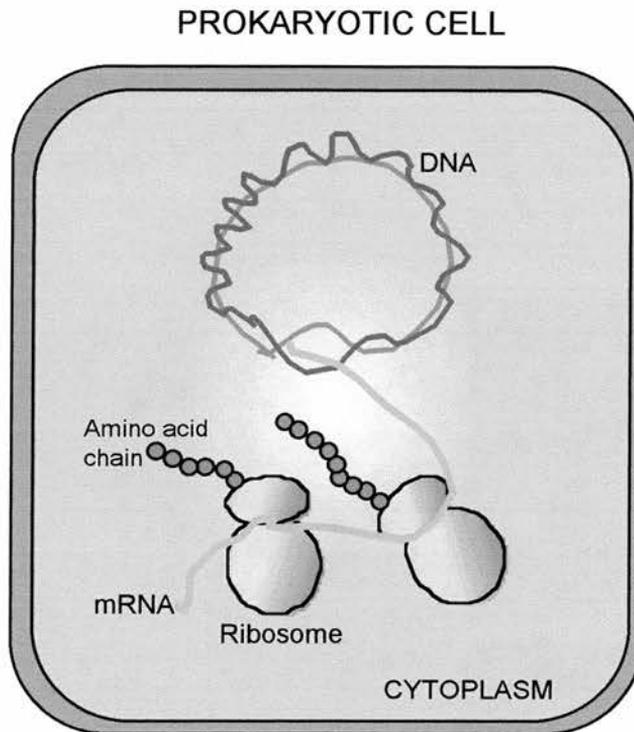


Fig.1.28. The prokaryotic cell. The DNA is transcribed producing mRNA. The ribosomes bind to the mRNA and initiate translation.

1.7.1.1 Comparison of prokaryotic and eukaryotic mRNA structure.

Eukaryotic mRNAs are monocistronic with a 7-methyl guanosine triphosphate cap structure linked to the 5' end this is followed by a short (50-100nts), unstructured untranslated region (UTR). The authentic initiation codon is usually the first AUG codon from the 5' end. Following the ORF stop codon there is a 3' UTR and a poly (A) tail.

Prokaryotic mRNAs like eukaryotic mRNA contain a short untranslated region at the 5' end and at the 3' end. Unlike eukaryotic messages prokaryotic mRNAs are not capped and usually lack a 3' poly(A) tail. The major difference is that prokaryotic mRNAs are polycistronic encoding several proteins therefore each mRNA contains multiple initiation and termination signals that must be recognized by the ribosome. Short intercistronic nontranslated regions usually separate the coding sequences (Kozak, 1983). The prokaryotic initiation codon is also AUG although a few examples of alternative initiation codons have been identified. A purine-rich sequence termed the

Shine-Dalgarno sequence situated about ten nucleotides upstream of the initiation codon interacts with the 16S rRNA.

1.7.2 Eukaryotic translation

1.7.2.1 Formation of the initiation complex

For a comprehensive review refer to Maitra, Stringer and Chaudhuri (1982) and Pain (1996), for a shorter version refer to Sachs, Sarnow and Hentze (1997).

(i) Dissociation of ribosomal subunits-

The first step of initiation is to generate 40S ribosome subunits, this requires a shift in the equilibrium of the reaction that normally favours the formation of inactive 80S ribosome complexes. Three initiation factors are involved in this step to prevent reassociation of the ribosomal subunits. The most active is eIF6 that binds to the 60S subunit, eIF3 binds to the 40S subunit. The activity of eIF3 is enhanced by the presence of eIF1A.

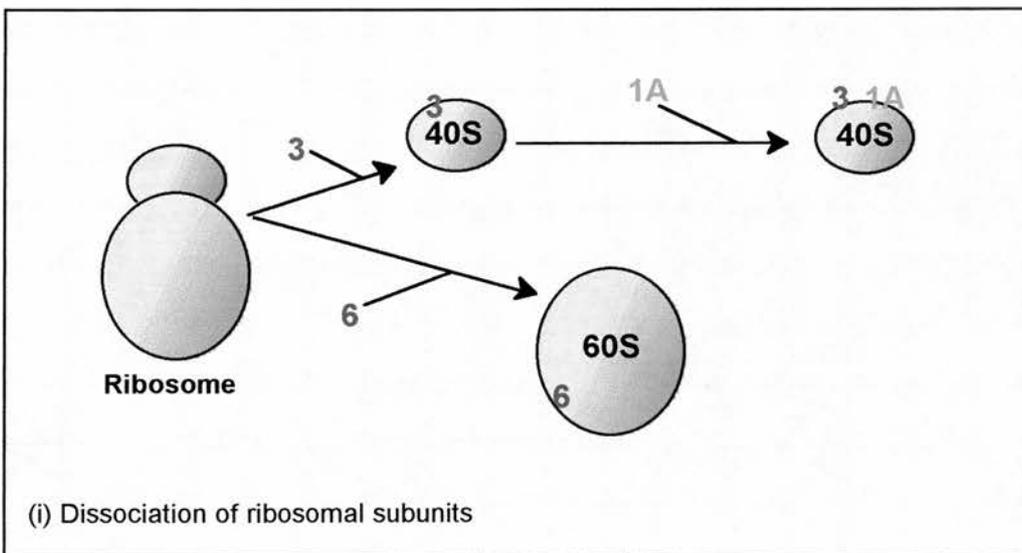


Fig.1.29. Step 1 of translation initiation.

(ii) Formation of the 43S ternary complex -

The next stage involves the formation of a ternary complex composed of eIF2, GTP and the initiator tRNA (Met-tRNA_i). The ternary complex binds to the 40S subunit forming the 43S preinitiation complex.

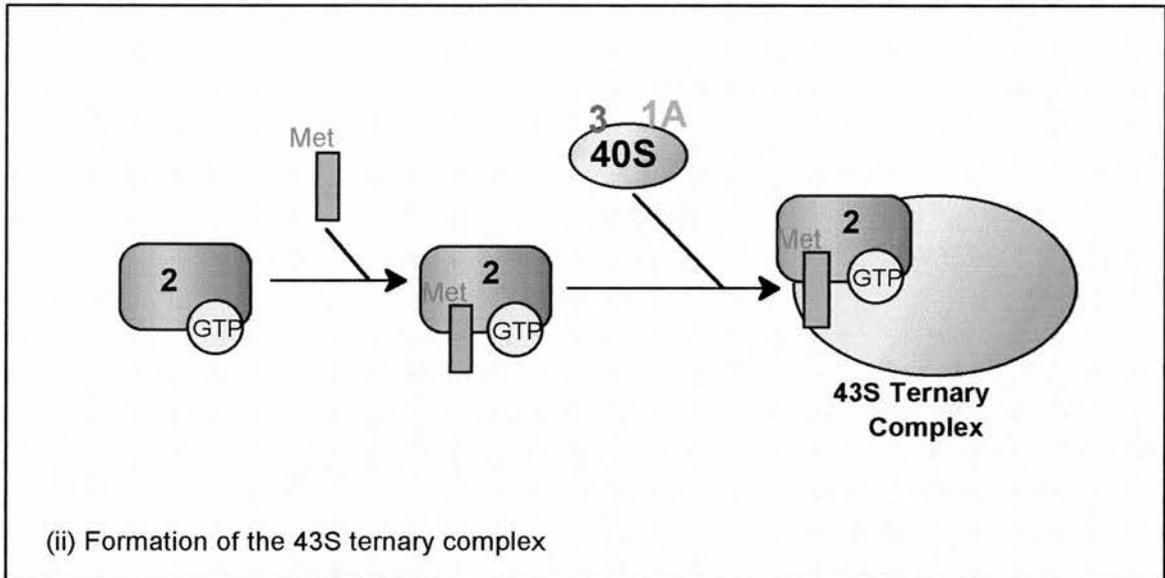


Fig.1.30. Step 2 of translation initiation.

(iii) mRNA recruitment and identification of the initiation codon.

The subsequent step involves the recruitment of mRNA and identification of the initiation AUG codon. The precise details of the process involved is still a matter for debate, one view will be presented here. The function of the eIF4F complex is to recruit the mRNA to the ribosome. The eIF4E subunit binds to the mRNA 5' cap structure. The eIF4A component of the complex has an RNA helicase activity and thus unwinds any secondary structure present allowing the 43S preinitiation complex to bind. The unwinding activity of eIF4A is enhanced by the presence of eIF4B. The individual functions of the initiation factors involved in recruiting the mRNA are coordinated via the eIF4G protein that acts as a scaffolding protein uniting the activities of eIF4E, eIF4A and the ribosome binding protein eIF3.

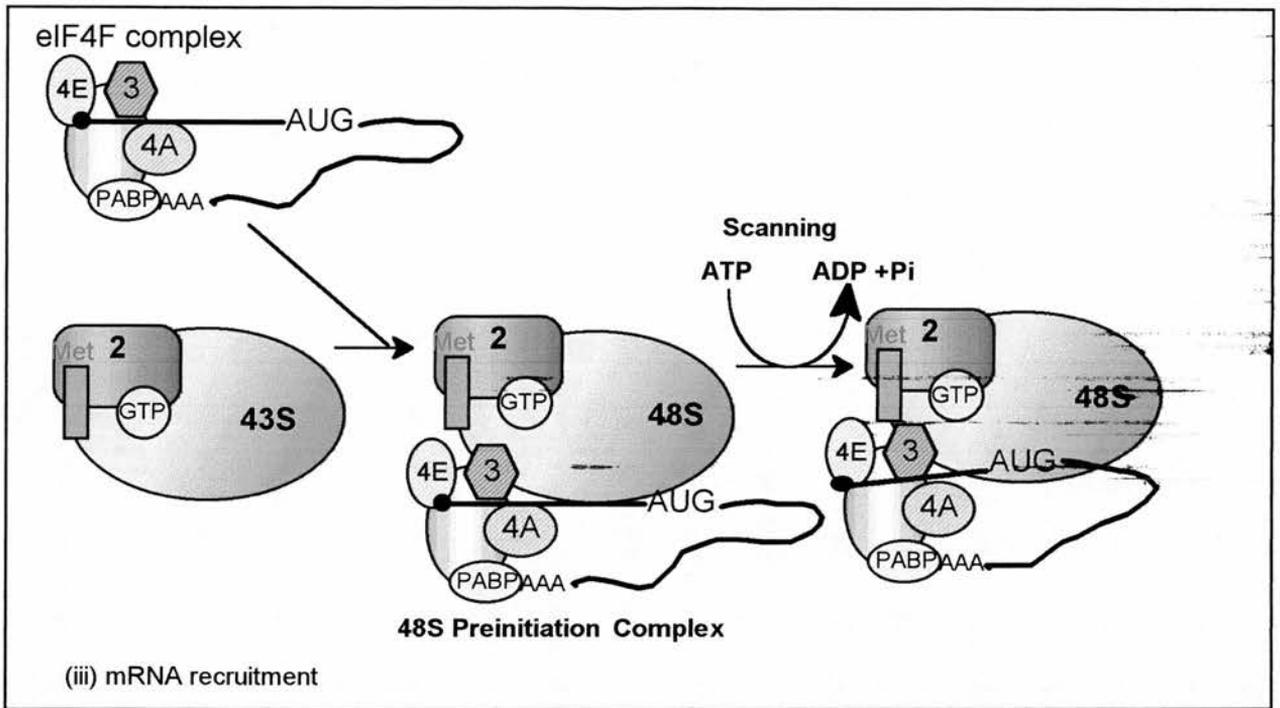


Fig.1.31. Step 3 of translation initiation.

The 43S preinitiation complex must identify the initiation AUG codon, following binding to the 5' cap the 43S complex scans along the mRNA. The primary factor involved in identifying the initiation codon is recognition by the anticodon of the initiator tRNA. To aid in correct identification two other factors may be involved. The first is the sequence context surrounding the AUG codon. The second involves slowing down the scanning process due to the secondary structure found within the coding region.

(iv) Formation of the 80S ribosome complex and initiation of protein synthesis.

The recognition of the correct AUG codon by the met-tRNAⁱ triggers eIF5 to stimulate hydrolysis of bound GTP by eIF2. The conversion to eIF2.GDP releases the initiation factors from the 40S subunit at the same time eIF6 is released from the 60S subunit. The ribosomal subunits bind forming the 80S complex. eIF2.GDP is released from the ribosome and recycled in readiness for another round of initiation by eIF2B. The final role of the initiation factors occurs during formation of the first peptide bond,

a conformational change is required that is believed to be mediated by eIF5A. With the initiator tRNA in the P site, the ribosome is ready to begin the elongation phase.

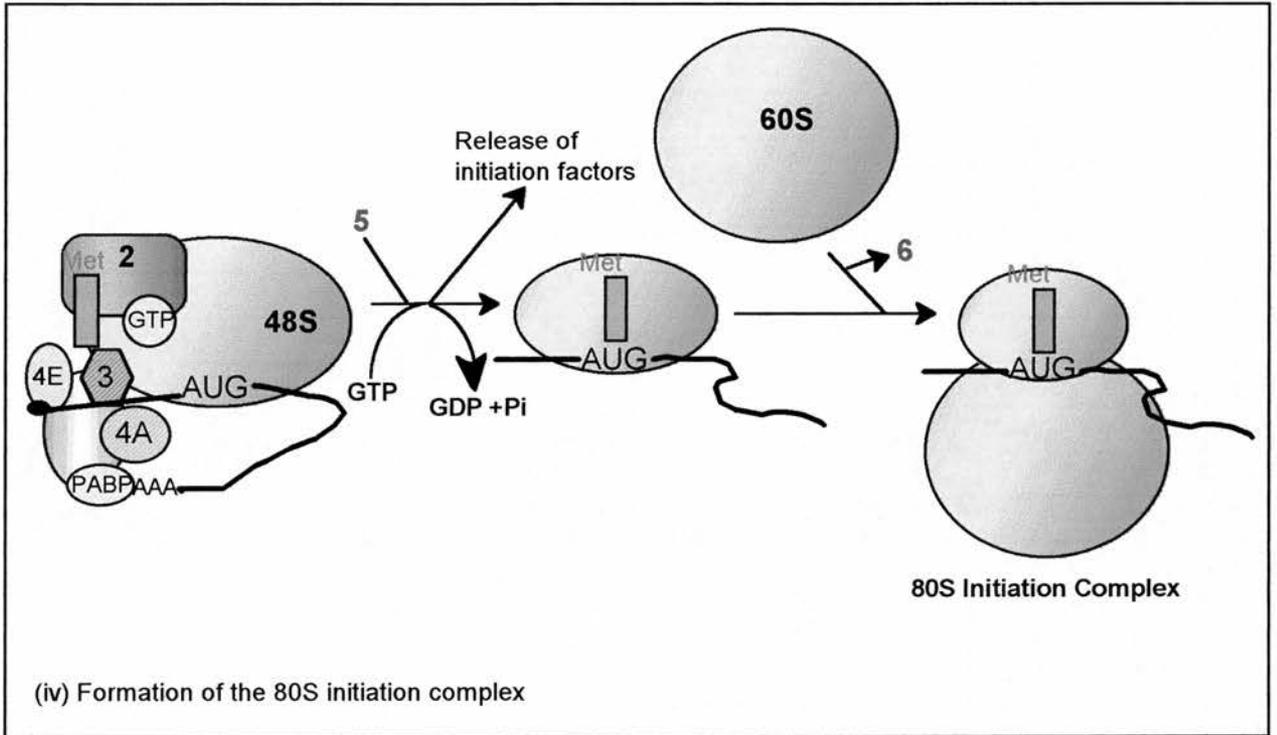


Fig.1.32. Step 4 of translation initiation.

New Nomenclature	Old Nomenclature	Function
eIF1	eIF-1	Unknown
eIF1A	eIF-4C	Stimulates subunit joining.
eIF2	eIF-2	Forms ternary complex with GTP and Met-tRNA _i . Binds Met-tRNA _i to 40S subunit.
eIF2B	eIF-2B, GEF	Recycles eIF2.GDP
eIF3	eIF-3	Subunit dissociation. Binding mRNA to 40S subunit.
eIF4A	eIF-4A	ATP-dependent unwinding of mRNA.
eIF4B	eIF-4B	Stimulates ATP-dependent unwinding of mRNA.
eIF4F	eIF4	Composed of eIF4E, eIF4G, eIF4A. Recognizes mRNA cap, unwinds mRNA, recruits mRNA to ribosome.
eIFiso4F		Functions as eIF4F identified in plants.
eIF4H		Recently discovered in 1998 (Richter-Cook <i>et al.</i> , 1998), function unknown.
eIF5	eIF-5	Joining of 60S subunit.
eIF5A	eIF-4D	Stimulates formation of first peptide bond.
eIF6	eIF-3A	Prevents association of 60S and 40S subunits.

Table 1.2. Eukaryotic initiation factors.

1.7.2.2 Elongation

The elongation process adds one amino acid to the C-terminus of the nascent polypeptide every round of the cycle (reviewed by Wilson and Noller, 1998). EF1 α .GTP binds to aminoacyl-tRNA and delivers it to the A site of the ribosome. Once the codon: anticodon has been correctly matched, GTP hydrolysis is triggered releasing EF1 α .GDP. A peptide bond is formed between the aminoacyl-tRNA in the A site and the peptidyl-tRNA in the P site transferring the peptide from the P to the A site. Formation of the peptide bond is catalyzed by the peptidyl transferase in the 60S subunit. The 3' end of the new peptidyl-tRNA shifts to the P-site but the anticodon stays in the A site. In the same way the 3' end of the deacylated tRNA shifts to the E site. The final step of the cycle translocation involves eEF2.GTP, the mRNA is moved by three nucleotides placing a new codon in the A site. The peptidyl-tRNA and the deacylated-tRNA codons are also shifted placing the two tRNAs fully in the P and E sites. The ribosome is ready to repeat the elongation cycle.

1.7.2.3 Termination

When a stop codon (UAA, UAG or UGA) enters the A site there are no available matching tRNAs. The stop codon is recognized by the release factor (eRF1.GTP). eRF1.GTP forms a complex with eRF3 that is thought to mimic the eEF1 α complex thereby triggering a peptidyl transferase event. This results in the release of the peptide as there is no aminoacyl-tRNA in the A site to receive the nascent peptide (reviewed by Stansfield, Jones and Tuite, 1995).

1.8 THE eIF4F COMPLEX

The eIF4F complex has a crucial role during initiation of translation, the complex is responsible for uniting the mRNA and the ribosome. The eIF4F complex is composed of three main initiation factors eIF4A, eIF4E and eIF4G (reviewed by Gingras, Raught

and Sonenberg, 1999). Overexpression of eIF4E or eIF4G results in transformation of the cell (Rousseau *et al.*, 1996).

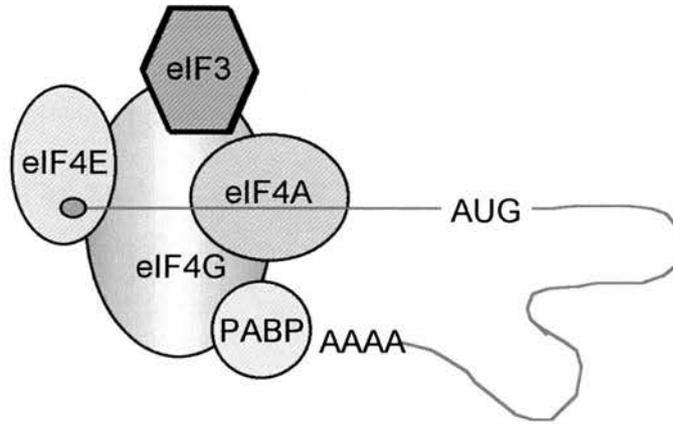


Fig.1.33. The mammalian eIF4F complex. The eIF4G protein acts as a scaffolding protein, eIF4E binds to the 5' cap of the mRNA, PABP binds to the poly A tail.

1.8.1 eIF4A

eIF4A is a 46kDa protein and is a member of the DEAD-box family of proteins (Linder *et al.*, 1989). The DEAD-box proteins all exhibit NTPase activity. In addition to this eIF4A has a double-strand bidirectional RNA unwinding activity. In mammalian cells there are three isoforms of eIF4A that exhibit different tissue and developmental expression (Nielsen and Trachsel, 1988; Li *et al.*, 1999). The helicase activity of eIF4A is stimulated by a 69kDa RNA-binding protein eIF4B.

1.8.2 eIF4E

eIF4E is a 24kDa protein responsible for recognition of the methyl-7-guanosine cap structure located at the 5' end of mRNA thus eIF4E is vital for cap-dependent translation. eIF4E is a prime target for regulation of cap-dependent translation. To date three alternative mechanisms have been discovered that regulate eIF4E activity: -

1.8.2.1 eIF4E transcription - The activation of eIF4E transcription can be triggered by external stimuli. A regulator of cell growth (MYC) was strongly implicated in the up-regulation of eIF4E (Rosenwald *et al.*, 1993). Furthermore two MYC-binding sites

that are important for promoter activity have been located within the eIF4E promoter region (Jones *et al.*, 1996).

1.8.2.2 Phosphorylation - several external stimuli including hormones and growth factors induce phosphorylation of eIF4E. Studies suggest that eIF4E can be phosphorylated by numerous kinases but recent work has identified the primary kinase involved is Map kinase-interacting protein kinase-1 (MNK1). The eIF4E phosphorylation site was initially incorrectly identified as residue Ser-53 (Rychlik, Russ and Rhoads, 1987) further work has mapped the phosphorylation site to residue Ser-209 (Flynn and Proud, 1995; Joshi *et al.*, 1995). Due to this the full effects of eIF4E phosphorylation are not clear although there is some evidence that phosphorylated eIF4E has an increased affinity for the 5' m7G cap.

1.8.2.3 Translational inhibitors - A family of small proteins (10-12kDa) that contain an eIF4E-binding site homologous to the eIF4E-binding site of eIF4G (consensus sequence YXXXXLL/M/F) were discovered in mammalian cells. The 4E-binding proteins directly compete with eIF4G for the eIF4E binding site (Mader *et al.*, 1995; Lawrence and Abraham, 1997). As a result the available eIF4E is sequestered by the 4E-binding proteins thus preventing association with eIF4G and the ribosome. The effect of 4E-binding proteins on eIF4E activity is modulated by phosphorylation of the 4E-BPs. The 4E-BPs can be phosphorylated at multiple serine and threonine residues. The level of phosphorylation and the specific residues involved alters the affinity for eIF4E thereby regulating binding of 4E-BP (Gingras *et al.*, 1998).

1.8.3 eIF4G

eIF4G (previously known as p220) is the central protein of the eIF4F complex. eIF4G possesses binding sites for several proteins and acts as a scaffolding protein uniting the functional activities of the bound proteins. In mammalian cells two eIF4G isoforms have been identified eIF4GI (Morley, Curtis and Pain, 1997) and eIF4GII (Gradi *et al.*, 1998a). The proteins have similar predicted molecular weights (171kDa and 176kDa).

Two smaller related proteins are also found in mammalian cells p97 (Imataka, Olsen and Sonenberg, 1997; Levy-Strumpf *et al.*, 1997; Shaughnessy, Jenkins and Copeland, 1997) and PA1P-1 (Craig *et al.*, 1998).

1.8.3.1 eIF4G binding activities

eIF4G has binding sites for initiation factors eIF4A, eIF4E and eIF3. Other sites identified so far include a binding site for the Poly(A)-tail binding protein (PABP) (Imataka, Gradi and Sonenberg, 1998), a binding site for the kinase MNK-1 (Pyronnet *et al.*, 1999) and an RNA-recognition-motif-like RNA binding domain. The eIF4E-binding site is located at position 572-578 in the N-terminal region, the consensus eIF4E binding site is YXXXXLL/M/F. Mammalian eIF4G has two eIF4A binding sites, one in the central region, the other in the C-terminal region. The eIF3 binding site is also located within the central region but is independent of the central eIF4A binding site. Thus eIF4G unites the mRNA and ribosome through binding to eIF4E and eIF3 respectively. Finally through eIF4A helicase activity the mRNA is prepared for the addition of the 60S ribosome subunit.

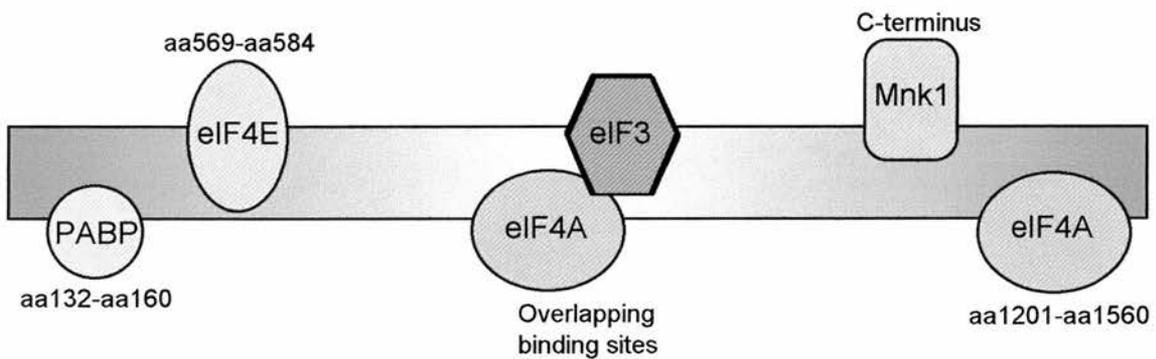


Fig.1.34. The binding sites of the eIF4G protein.

Via the interaction with PABP, eIF4G effectively circularizes the mRNA bringing the 5' and 3' together. This provides an explanation for how the 3'UTR can influence translation. Another possibility is that the ribosomes can pass directly from the 3' to the 5' end of the mRNA and reinitiate without dissociating from the ribosome, this has

yet to be substantiated experimentally although circular polysomes have been observed under the electron microscope.

As mentioned previously the MNK1 kinase is the primary candidate for the eIF4E kinase. MNK1 binds directly to eIF4G and disruption of the binding site was found to result in a dramatic decrease in phosphorylation of eIF4E. Thus it is postulated that eIF4G facilitates eIF4E phosphorylation by uniting the kinase and substrate.

1.8.4 Plant eIF4F complex

In many ways the plant eIF4F complex is analogous to the mammalian counterpart. The major difference is the existence of a smaller complex designated eIFiso4F composed of two subunits p28 and p86 (Browning *et al.*, 1992). p28 (eIFiso4E) binds to the 5'cap. p86 (eIFiso4G) is much smaller than eIF4G but sequence analysis shows that it is homologous to eIF4G particularly within the central region (Allen *et al.*, 1992). eIFiso4G also contains binding sites for eIF4E and PABP. Studies demonstrate that both eIF4F and eIFiso4F are not required for protein synthesis, the complexes can substitute for each other. Plant eIFiso4F can also substitute for the mammalian eIF4F complex (Metz and Browning, 1996). Plant eIF4F and eIFiso4F demonstrate different affinities for m7G-containing cap analogues suggesting that these complexes may be involved in regulating translation via mRNA selection (Carberry, Darzynkiewicz and Goss, 1991; Sha *et al.*, 1995).

1.9 HOST-CELL SHUT-OFF

Following viral infection of a cell many viruses cause host-cell 'shut-off'. During viral replication the virus must utilize part of the host-cell machinery therefore it must compete with cellular genes. To gain an advantage over cellular genes many viruses

have developed mechanisms to target parts of the host-cell gene expression pathway not required by the virus.

Widespread mechanisms include an increased tolerance by viral RNA for changes in ion concentration that occur during virus infection therefore the virus avoids any inhibition due to activation of a cellular response to the viral infection. In response to the presence of dsRNA the host activates a dsRNA-dependent protein kinase (PKR) that phosphorylates the α -subunit of eIF2 inactivating it (Crum *et al.*, 1988; Samuel, 1993). Viral RNA is a stronger competitor for the residual non-phosphorylated eIF2 than cellular messages.

Some viruses have also developed specific mechanisms to inactivate the PKR response. For example the herpes simplex virus-1 (HSV-1) actively causes dephosphorylation of eIF2 α (He, Gross and Roizman, 1997). Viruses interfere with host-gene expression at all stages of the pathway including transcription, RNA maturation, RNA transport and translation.

1.9.1 Picornavirus host-cell shut-off

Early observations noted that following picornavirus infection there was a dramatic shut-off of host-cell protein synthesis but viral protein synthesis was not affected (Rose *et al.*, 1978). The level of cellular mRNA in infected cells remains normal but no proteins are produced suggesting that inhibition is due to a defect in translation of cellular messages. Further studies identified that inhibition occurs at the level of initiation of translation. Subsequent analysis of cellular initiation factors from infected cells found that a component of the eIF4F complex, eIF4G was cleaved (Etchison *et al.*, 1982; Ehrenfeld, 1982; Wyckoff, 1993).

The primary candidates for mediators of eIF4G cleavage activity were the viral proteases 2A and 3C and indeed mutations within 2A^{Pro} abolished eIF4G cleavage (Kräusslich *et al.*, 1987; Liebig *et al.*, 1993). Although picornaviral 2A^{Pro} is widely

accepted as the viral protease involved in eIF4G cleavage the mechanism involved has been the subject of some debate. There are two proposed models of 2A^{Pro} activated eIF4G cleavage that 2A^{Pro} directly cleaves eIF4G or that 2A^{Pro} promotes eIF4G cleavage through activation of an unknown cellular protease.

In favour of the direct cleavage hypothesis *in vitro* studies indicate that rhinovirus, coxsackievirus and poliovirus 2A^{Pro} cleave eIF4G to give products of the same size found *in vivo* in infected cells. Cleavage of eIF4G within the eIF4F complex is 100-fold more efficient than lone eIF4G. Therefore formation of the complex may alter the conformation of eIF4G making the cleavage site more accessible to 2A^{Pro} (Haghighat *et al.*, 1996; Ventoso *et al.*, 1998). Other factors that support a direct cleavage include the mapped cleavage site of eIF4G that conforms to the consensus sequence required for 2A^{Pro} cleavage (Lamphear *et al.*, 1993; Lamphear *et al.*, 1995) and analysis of the active site cleft from the crystal structure of 2A^{Pro} (Petersen *et al.*, 1999).

The indirect cleavage theory is still supported primarily due to the kinetics of eIF4G cleavage (Bovee *et al.*, 1998ab). *In vitro* eIF4G cleavage is relatively slow and inefficient this does not support the observations *in vivo* of rapid and complete cleavage of eIF4G. Experimental data that also upholds this theory is that 2A^{Pro} does not co-purify with eIF4G activity and that inactivation of 2A^{Pro} with 2A^{Pro}-antibodies does not inhibit eIF4G cleavage (Lloyd *et al.*, 1986; Kräusslich *et al.*, 1987). So far identification of a putative cellular eIF4Gase activated by 2A^{Pro} has failed.

Some experimental evidence that contradicts the involvement of eIF4G cleavage in the process of host-cell shut-off has recently been explained by the discovery of the eIF4G isoform (eIF4GII) in mammalian cells. Under certain conditions where eIF4GI cleavage was complete there was incomplete inhibition of host-cell protein synthesis. There was also a significant time lapse between complete cleavage of eIF4GI and the onset of host-cell shut-off. It was found that both of these discrepancies were solved

by the discovery of eIF4GII. eIF4GII is more resistant to cleavage than eIF4GI, therefore, conditions that permitted complete eIF4GI cleavage dramatically reduced cleavage of eIF4GII. The cleavage of eIF4GII coincided with the onset of host-cell shut-off. Thus cleavage of both forms of eIF4G are required for inhibition of cellular protein synthesis (Gradi *et al.*, 1998b).

An interesting evolutionary aspect of eIF4G cleavage is that in FMDV (lacking a 2A protease), eIF4G cleavage activity is retained. In FMDV eIF4G cleavage is a function of the L protease (Devaney *et al.*, 1988). The L^{PRO} cleavage site maps to the same region of eIF4G as the 2A^{PRO} cleavage site (Kirchweger *et al.*, 1994).

1.9.1.2. Site of eIF4G cleavage

eIF4GI is cleaved between an Arg-Gly pair at position 642 by entero- and rhinovirus 2A^{PRO} and at a Gly-Arg pair at position 635 by FMDV L^{PRO}. Cleavage separates the N-terminal region containing the PABP and eIF4E binding sites from the C-terminal region that contains the eIF4A, eIF3 and Mnk1 binding sites. As a result eIF4G cleavage prevents cap-dependent translation by physically separating the cap-binding and ribosome-binding activities of the eIF4F complex. Remarkably eIF4G cleavage not only inhibited cap-dependent translation but actively promoted cap-independent translation (Ohlmann *et al.*, 1995; Borman *et al.*, 1997a). The stimulatory effect on cap-independent translation was assigned to the C-terminal cleavage product of eIF4G. Further studies demonstrate that the middle region of eIF4G that contains the eIF4A, eIF3 and RRM-binding sites is sufficient for translation of uncapped and IRES-driven translation (Ohlmann *et al.*, 1996; Morino *et al.*, 2000).

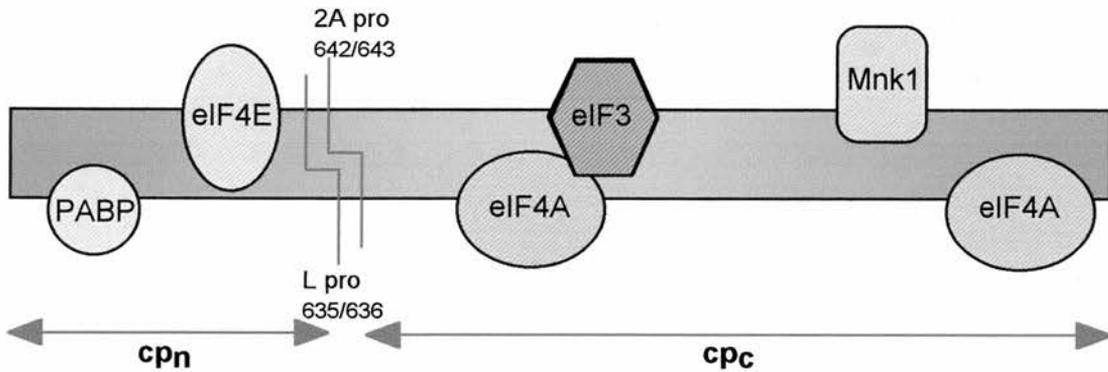


Fig.1.35. The site of *2A^{pro}* and *L^{pro}* mediated cleavage of *eIF4G*.

1.9.2 Alternative shut-off mechanisms

1.9.2.1 Cardiovirus shut-off

Cardiovirus 2A protease does not cleave either of the eIF4G isoforms but still causes host-cell shut-off. The mechanism employed by cardioviruses to inhibit host-cell protein synthesis also targets the eIF4F complex. Cardiovirus infection induces dephosphorylation of 4E-BP1 thereby increasing its binding affinity for eIF4E (Gingras *et al.*, 1996). Thus dephosphorylated 4E-BP1 sequesters available eIF4E effectively inhibiting cap-dependent translation.

1.9.2.2 Adenovirus shut-off

The 4E-BP proteins are targeted by adenovirus infection but in contrast to cardioviruses, the adenoviruses induce phosphorylation of 4E-BP1 and 4E-BP2 resulting in deactivation of these proteins and hence an increase in available eIF4E (Gingras and Sonenberg, 1997). Therefore adenovirus infection initially promotes cap-dependent translation as the early adenovirus genes are translated in a cap-dependent manner. During the latter stages of adenovirus infection when the viral late

genes are expressed the virus induces host-cell shut-off via dephosphorylation of eIF4E. This process effectively reduces the binding affinity of eIF4E to the 5' cap and increases the likelihood for ribosome jumping, the method employed by the adenovirus late genes. This demonstrates the adeptness of viruses to manipulate cellular processes to suit the viral replicative requirements throughout infection.

1.9.3 Shut-off in plant systems

In plant systems viruses do not cause host-gene shut-off in the same way as animal viruses. Instead the viruses manipulate cellular processes in such a way as to allow viral replication but to maintain a viable cell by preserving host-gene expression (Wang and Maule, 1995; Aranda and Maule, 1998). The difference between the two systems is probably due to the requirements for viral transmission. In animal cells viruses are released upon lysis of an infected cell, the virus can bind to receptors on adjacent cells and thus gain access to new host cells. This method is not viable in plant tissues where the cells are bound by rigid cell walls and transmission via this method would be severely limited. Plant viruses move from cell-to-cell through the plasmodesmata that connects adjacent cells, to utilize this method of transmission the virus must preserve a healthy cell.

1.10 CAP-INDEPENDENT TRANSLATION

Several lines of evidence suggest that picornaviruses initiate translation in a cap-independent manner. The 5' end of picornaviral RNA is not capped instead it is linked to a small protein VPg therefore it cannot be recognized by eIF4E. The 5' region of the picornavirus genome contains a long (600-1200 nt) non-coded region (NCR) that is highly structured this would be expected to inhibit ribosomal scanning. The final piece of evidence that supports the theory that picornaviruses employ a novel mechanism of initiation is the process of picornaviral-induced host-cell shut-off that inhibits cap-dependent translation yet picornavirus RNA is successfully translated in infected cells. It was proposed that picornavirus RNA could bind ribosomes internally

that would subsequently scan the RNA to locate the initiation codon (Jang *et al.*, 1989; Jang *et al.*, 1990). The region involved was mapped to a 450nt segment within the 5' NCR, this was designated the internal ribosome entry site or IRES (Jackson, Howell and Kaminski, 1990; Stewart and Semler, 1997).

1.10.1 Picornaviral internal ribosome entry sites.

Picornaviral IRESes have been divided into three groups according to the primary sequence and predicted secondary structure of the IRES. Rhino- and enteroviruses have type I IRESes, cardio- and aphthoviruses contain type II IRESes and hepatoviruses have type III IRESes. Investigations into the requirements for translational efficiency of the IRESes demonstrated that the three types of IRES function differently.

Type I IRESes are inefficient in rabbit reticulocyte lysates and require additional factors present in HeLa cell extracts to initiate translation. This group of IRESes is also very sensitive to fluctuations in salt concentration. The efficiency of type I IRESes is greatly enhanced by the addition of the entero- and rhinovirus 2A protease or FMDV L protease.

In contrast, type II IRESes are very efficient in rabbit reticulocyte lysates and do not require supplementing with HeLa cell extracts. The translation efficiency of type II IRESes is much more stable; they are not significantly affected by alterations in salt concentrations or by addition of the 2A or L proteases.

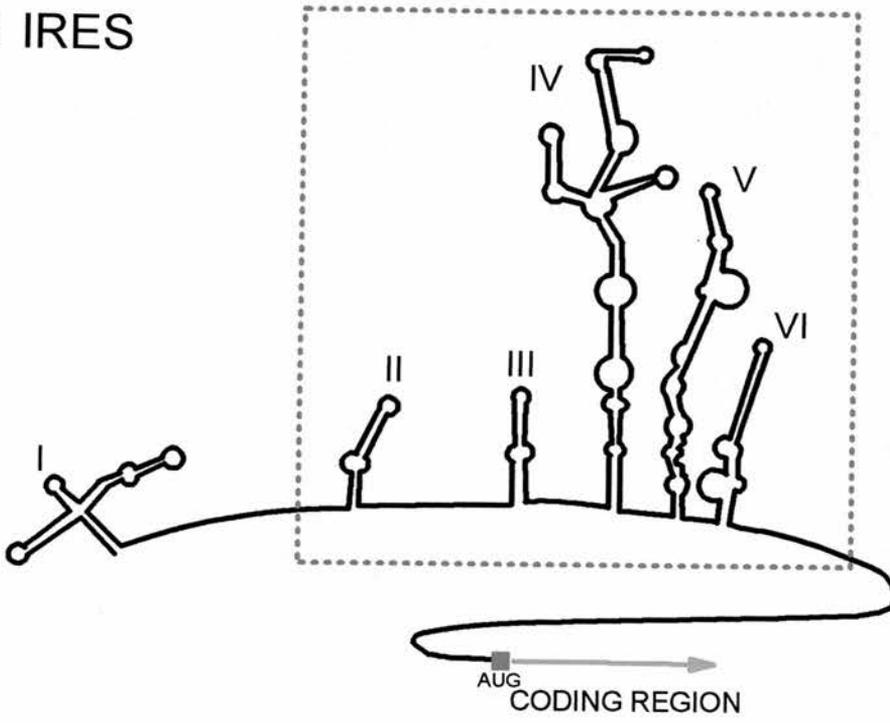
The hepatovirus type III IRES like the type I IRES is very inefficient *in vitro* but it is not stimulated by addition of HeLa cell extracts. The type III IRES is tolerant of a broad range of salt concentrations but interestingly was inhibited by the addition of the 2A or L proteases. Therefore the requirements for the different groups of IRES appear to vary dramatically (Borman *et al.*, 1995; Hunt *et al.*, 1999b). There is some

indication that the specific needs of the IRES may play a role in determining viral tissue tropism.

1.10.2 Mechanism for internal initiation

The process utilized by picornaviral IRESes to recruit ribosomes directly is not yet understood. Pestova *et al.* (1996) have demonstrated that the assembly of the 48S complex on the EMCV IRES *in vitro* required a similar set of canonical initiation factors as cap-dependent initiation specifically eIF2, eIF4F, eIF3 and eIF4B (Pestova, Hellen and Shatsky, 1996). IRES-mediated initiation may also require a number of novel cellular factors.

Type 1 IRES



Type 2 IRES

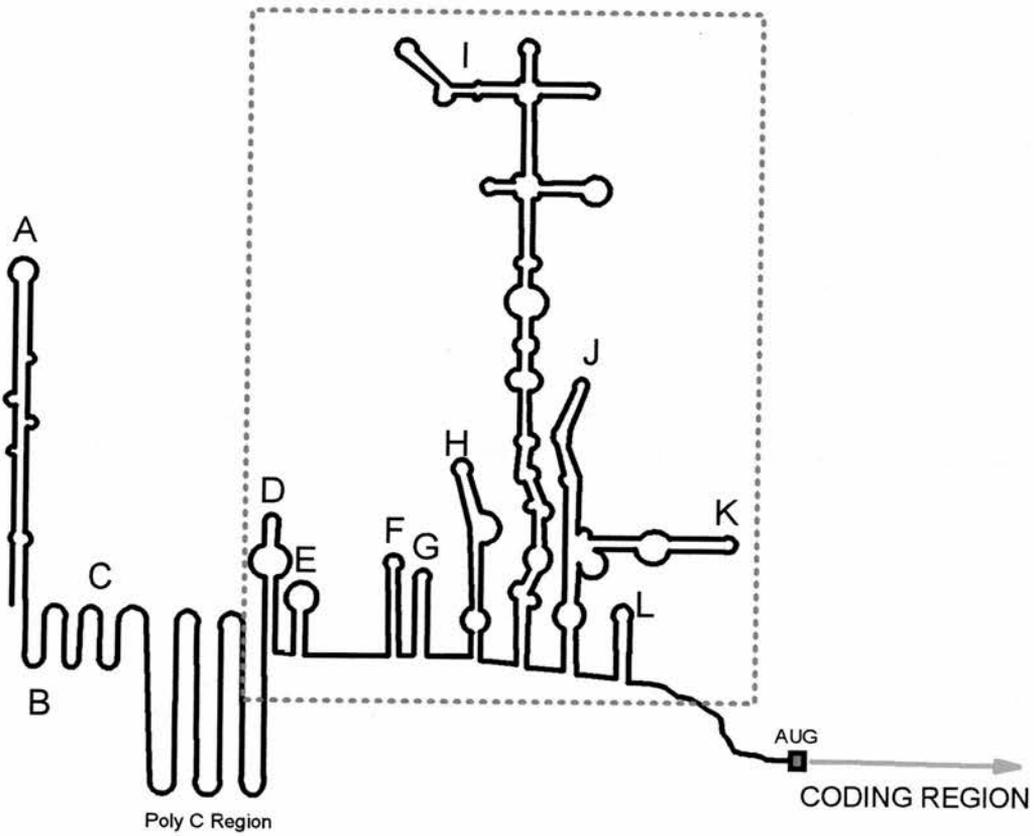


Fig. 1.36. Predicted RNA secondary structure of picornavirus 5'NCRs. Type I IRES i.e. Rhinoviruses and Type II IRES i.e. Aphthoviruses

1.10.3 The role of cellular proteins during picornavirus IRES-dependent translation.

Several cellular proteins have been implicated in IRES-mediated initiation these include pyrimidine-tract-binding (PTB) protein (Niepmann, 1996; Niepmann *et al.*, 1997; Kaminski and Jackson, 1998; Hunt and Jackson, 1999), La-autoantigen (Luz and Beck, 1991; Isoyama *et al.*, 1999; Kim and Jang, 1999), unr (Hunt *et al.*, 1999a) and poly-C-binding protein 2 (PCBP-2) (Spångberg and Schwartz, 1999). All of these proteins contain multiple RNA-binding sites therefore it has been postulated that these proteins bind several regions of the IRES and stabilize the complex tertiary structure of the IRES. The precise role of these cellular proteins is still a matter of debate. Studies to date seem to suggest that each specific IRES may have a different requirement for cellular factors (Borman *et al.*, 1997b; Roberts, Seamons and Belsham, 1998) and employ a slightly different mechanism for initiation. Experiments indicate that the EMCV IRES delivers the ribosome precisely to the correct initiation codon (Kaminski, Howell and Jackson, 1990). In comparison, the poliovirus IRES binds the ribosome at a point upstream of the initiation codon then as in cap-dependent initiation the ribosome scans the RNA to locate the initiation codon (Pelletier and Sonenberg, 1988).

1.10.4 Host-cell mRNAs containing IRES elements.

The process of binding the ribosome internally is not unique to viruses a number of eukaryotic cellular mRNAs have been identified that can recruit ribosomes in a cap-independent manner. The cellular mRNAs identified have no primary sequence homology but share several features that are incompatible with ribosomal scanning these include a long 5' UTR with a high G+C content that can potentially form RNA secondary structures (Le and Maizel, 1997). Cellular mRNAs that contain IRES elements include growth factors (Huez *et al.*, 1998; Vagner *et al.*, 1995; Teerink, Voorma, and Thomas, 1995; Stein *et al.*, 1998), the proto-oncogene *c-myc* (Nanbru *et*

al., 1997) and the immunoglobulin heavy-chain binding protein (BiP) (Yang and Sarnow, 1997). It has been postulated that these mRNAs are normally translated in a cap-dependent manner but when this pathway is inhibited during conditions of stress or cell-growth arrest an alternative IRES-mediated pathway is used to translate this group of cellular mRNAs.

1.11 AIMS OF THE PROJECT

1.11.1. Background

Viruses have developed extremely efficient methods of modifying host-cell systems to suit their requirements. Due to the nature of viruses and the limited coding capacities available, the viral genetic elements responsible are often very effective and small making them a valuable source of sequences to molecular biologists for use in genetic engineering.

The use of viral genetic elements as genetic tools is widespread (Mushegian and Shepherd, 1995). The simplest example is the incorporation of viral promoter sequences to enhance gene expression. A wide range of promoters are available, selection of an appropriate promoter can allow a degree of control over both the level and tissue-specificity of gene expression.

A rapidly expanding area of interest is the use of viruses as vectors to transport foreign genes into a cell. There are two basic types of viral vector, the replacement vector where a region of the viral genome is deleted and replaced with a foreign gene. The second type is the insertion vector where the foreign gene is simply added to the original genome.

The expression of multiple foreign genes is a desirable prospect. The introduction of multiple genes into a system is relatively simple this can be achieved in a variety of ways. The principle difficulty is synchronizing the expression of the genes. To overcome these problems, the viral strategy of initially producing the viral proteins as a single polyprotein has been adopted. A drawback of polyprotein production from the point of view of the virus is that the viral proteins are produced in equimolar amounts this is the goal for researchers attempting to produce co-ordinated expression.

The development of a viable polyprotein system for more complex applications of genetic engineering requires that a method of processing the artificial polyprotein must be included. The initial focus of this research group was to utilize the short peptide 2A sequence from FMDV as a means to process artificial polyproteins. Ryan and Drew (1994) demonstrated that the FMDV 2A peptide could function effectively when inserted between two reporter genes.

This strategy of using FMDV 2A as a genetic “splicing” tool is now well established and 2A has been used in a range of applications. One of the earliest examples was the modification of influenza virus to develop an influenza viral vector to introduce foreign genes into a mammalian system (Percy *et al.*, 1994). Inclusion of the FMDV 2A peptide separated the foreign gene from the viral genes thus maintaining viable virus. Similar viral vector systems utilizing FMDV 2A have been developed using paramyxovirus (Precious *et al.*, 1995), poliovirus (Mattion *et al.*, 1996), retrovirus (de Felipe *et al.*, 1999), flavivirus (Varnavski and Khromykh, 1999) and semliki forest virus (Smerdou and Liljestrom, 1999).

The autocatalytic cleavage function of FMDV 2A has been used to overcome difficulties encountered in investigating the cytokine interleukin-12 (IL-12). Active IL-12 is a heterodimer composed of two subunits (p40 and p35). A separate gene encodes each subunit therefore to successfully produce recombinant IL-12 both genes must be expressed and the heterodimer complex formed. Use of more established methods to express both genes such as transformation with two vectors or transformation with a vector containing both genes controlled by separate promoters produced unequal expression of the genes resulting in formation of a p40 homodimer. Expression of an artificial polyprotein encoding both subunits separated by the FMDV 2A sequence produced a functional recombinant IL-12 heterodimer (Chaplin *et al.*, 1999; Kokuho *et al.*, 1999).

The primary target for the use of artificial polyproteins is the plant system due to the particular difficulties associated with gene expression in this area (Stam, Mol and Kooter, 1997). The problems associated with multiple gene expression are compounded in plant systems by additional factors caused by the method of *Agrobacterium*-mediated transformation. The foreign gene contained between the T-DNA borders is inserted into the host genome. The level of expression is influenced by several factors including the adjacent host sequences, transgene copy number and DNA methylation of the introduced gene. These problems are amplified when the expression of multiple genes is required. Attempts to overcome these problems by closely linking two genes contained within the same T-DNA by using divergent or tandem promoters failed to produce co-ordinated levels of expression (Peach and Velten, 1991).

Marcos and Beachy performed the first introduction of an artificial self-processing polyprotein into a plant system (Marcos and Beachy, 1994). The polyprotein utilized the proteolytic activity of a plant virus protease: the TEV NIa protease. The TEV NIa protease was selected because it has a well-characterized cleavage site and was known to be active when expressed as a transgene in plants. The polyprotein cassette contained two viral coat proteins the artificial polyprotein was properly processed by the NIa protease *in vivo*. Expression of the viral coat proteins was effective in mediating an enhanced protection against these viruses (Marcos and Beachy, 1997; Ceriani *et al.*, 1998).

The successful use of the FMDV 2A sequence to process an artificial polyprotein within transgenic plants was demonstrated by Halpin *et al.* (1999) The FMDV 2A sequence efficiently processed the reporter polyprotein. A modified FMDV 2A sequence that cleaves inefficiently has been utilized by Santa Cruz *et al.* (1996) to 'decorate' viral capsids with foreign peptides. This is achieved by using modified potato virus X (PVX) with a foreign gene fused to the coat protein. To aid efficient virus assembly the presence of native coat protein is required. Therefore the

introduction of the inefficient FMDV 2A sequence between the coat protein and the foreign gene produced native coat protein and fused coat protein. This strategy has been applied to produce viruses coated with a single-chain antibody to the herbicide diuron (Smolenska *et al.*, 1998).

1.11.2. Overview of Project Aims.

The aim of this project is to develop the established technology within this laboratory that utilizes the novel cleavage mechanism of the FMDV 2A protein to process artificial polyproteins. The principle aim is to expand the genetic 'tool-box' available for the production of artificial polyproteins by developing an artificial self-processing polyprotein system exploiting the proteolytic activities of two picornaviral proteases HRV14 2A^{PRO} and FMDV Lb^{PRO}. The addition of these proteases to the artificial polyprotein repertoire would broaden the potential applications of the current artificial polyprotein technology.

The TEV NIa protease and the FMDV 2A protein both have limitations. The main problem associated with the use of the TEV NIa protease is the large size of the protein (49kDa). The FMDV 2A protein does not function in bacteria and thus cannot be used for production of polyproteins in prokaryotes. The novel cleavage mechanism of FMDV 2A puts several constraints on the structure of the polyprotein. Each FMDV 2A protein can only process one cleavage site therefore as each gene is added to the polyprotein another copy of FMDV 2A is also required. FMDV 2A cleaves C-terminally thus the FMDV 2A protein remains attached to the upstream protein and may disrupt the function of the protein of interest.

The FMDV Lb protease and HRV14 2A protease are small proteins approximately 20 kDa and 16 kDa respectively. Both proteases have the ability to process cleavage sites *in cis* and *in trans*, this ability can be exploited in the creation of artificial polyproteins to cleave multiple sites within the polyprotein using a single copy of the protease.

Aims.

- Construction of an artificial self-processing reporter polyprotein system containing the FMDV Lb protease to monitor the *cis* and *trans* processing activity of the protease.
- Construction of an artificial self-processing reporter polyprotein system using the HRV14 2A protease to monitor the *cis* and *trans* processing activity of the protease.
- To investigate the effect on translation of the addition of an IRES element to the artificial polyprotein system.
- To investigate the artificial polyprotein system *in vitro*.
- To express the artificial polyprotein system *in vivo* in prokaryotic and eukaryotic systems.

CHAPTER 2: EXPERIMENTAL

2.1. GENERAL MANIPULATION OF DNA.

2.1.1. Materials and solutions. The 1kb DNA ladder, restriction enzymes and corresponding buffer solutions were obtained from Promega. T4 DNA ligase was obtained from New England Biolabs. Alex Houston performed the DNA sequencing on a Perkin-Elmer ABI Prism™ 377 DNA sequencer using multicolour ABI dRhodamine 'BigDye' terminators.

Agarose gel loading buffer (x6)	2 x TAE
	50% [w/v] glycerol
	0.005% [w/v] bromophenol blue
	0.004% [w/v] ethidium bromide

2.1.2. Agarose-gel electrophoresis. Flat bed agarose gels, of concentration 0.7-2% [w/v] were prepared with 1xTAE (0.04M Tris-acetate, 0.001M EDTA), ethidium bromide was added to a final concentration of 0.5µg/ml. DNA samples in agarose gel loading buffer were applied to the gel. Electrophoresis was carried out at 100-180V in 1xTAE containing 0.5µg/ml ethidium bromide. The DNA bands were subsequently visualized by illumination from an UV transilluminator (UVP).

2.1.3. Restriction enzyme digestions. Restriction enzyme digests of plasmid DNA were conducted following the conditions recommended by the supplier. Typically, 1µg of DNA was digested with 1 unit of enzyme in a total volume of 20µl, containing 2µl of 10X restriction buffer. The reaction was incubated at the optimum temperature for enzyme activity, in general 37°C.

2.1.4. Purification of DNA fragments from agarose. The desired DNA band was run into 1% low-melting point (LMP) agarose, the band of DNA was excised from the LMP agarose and placed in a 1.5ml microcentrifuge tube. The sample was heated to 70°C to melt the agarose, DNA was isolated from the agarose using the Wizard® Preps DNA Purification system (Promega). Briefly, a Wizard® minicolumn was attached to a syringe barrel. The melted agarose was mixed with 1ml of DNA purification resin for 20 seconds then pushed through the minicolumn. The minicolumn was washed with 2ml of 80% isopropanol. The column was dried by centrifugation for 2 minutes at 10000g. The bound DNA was eluted from the column by the addition of 50µl of dH₂O and incubated for 1 minute. Finally, the minicolumn was centrifuged at 10000g for 20 seconds.

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

2.1.5.a. Small-scale preparation of plasmid DNA. Small quantities of plasmid DNA were prepared by two methods as outlined below. DNA prepared using the Wizard SV® miniprep kit (Promega) could be used directly in automated sequencing.

(i) Alkaline lysis method. 10ml of LB broth containing 100µg/ml ampicillin was inoculated with a single colony and incubated on a shaker at 37°C for 16 hours. A 1-5ml aliquot of cells was pelleted and the supernatant was discarded. The cells were resuspended in 100µl of solution 1 (50mM glucose, 25mM Tris.Cl (pH 8.0), 10mM EDTA) by vigorous vortexing, 200µl of solution 2 (1% SDS, 0.2M NaOH) was added followed by 150µl of solution 3 (3M KAc, 5M glacial acetic acid). The sample was mixed gently and incubated on ice for 5 minutes. The sample was centrifuged (12000g, 5 minutes, 4°C) and the supernatant was transferred to a fresh tube. The nucleic acid was extracted by the addition

of an equal volume of phenol:chloroform, the solution was vortexed and centrifuged (12000g, 2 minutes, 4°C). The upper aqueous layer was transferred to a fresh tube, 2 volumes of ethanol were added to precipitate the DNA. The sample was vortexed and incubated at room temperature for 2 minutes then centrifuged (12000g, 5 minutes, 4°C). The DNA pellet was washed with 70% [v/v] ethanol, dried and resuspended in 50µl of dH₂O with 2µl RNaseA (1mg/ml).

(ii) Wizard SV™miniprep kit (Promega). Minipreparation of DNA was performed according to the manufacturers instructions. Briefly, 10ml of antibiotic selective LB broth was inoculated with a single colony of bacteria. The culture was incubated overnight at 37°C on an orbital shaker. 1.5ml of culture was transferred to a clean microcentrifuge tube and the bacteria are pelleted by centrifugation (14000g, 5 minutes). The supernatant was discarded. The cells were resuspended in 250µl of cell resuspension solution (50mM Tris-HCl, (pH 7.5), 10mM EDTA, 100µg/ml RNase A) and 250µl of cell lysis buffer (0.2M NaOH, 1% SDS) was added and mixed gently. 10µl of alkaline protease solution was added and gently mixed with the sample and incubated at room temperature for 5 minutes. 350µl of neutralization buffer (4.09M guanidine hydrochloride, 7.59M KAc, 2.12M glacial acetic acid, pH 4.2) was added and the sample was gently mixed prior to centrifugation at 14000g for 10 minutes. The cleared bacterial lysate was transferred to a spin column and centrifuged at 14000g for 1 minute. The column was washed with 750µl of column wash solution (60mM KAc, 10mM Tris-HCl (pH 7.5), 60% ethanol) and dried by centrifugation at 14000g for 1 minute. The spin column was washed again with 250µl of column wash and dried by centrifugation at 14000g for 2 minutes. The spin column was transferred to a clean microcentrifuge tube and 100µl of nuclease-free dH₂O added. The DNA was eluted by centrifugation at 14000g for 1 minute.

An additional step as outlined in the manufacturers guidelines was required to prepare the DNA for use in automated sequencing. 50µl of 7.5M ammonium acetate and 375µl of 95% ethanol were added to the 100µl sample. The sample was centrifuged at 14000g for 15 minutes. The supernatant was carefully removed and the DNA pellet was washed in 250µl of 70% ethanol and repelleted by centrifugation at 14000g for 5 minutes. The ethanol was removed and the pellet was allowed to air-dry for 3-5 minutes. Finally, the DNA pellet was resuspended in 25µl of nuclease-free dH₂O.

2.1.5.b. Midi- and maxi-preparation of DNA. Midi- and maxi-preparations of plasmid DNA were produced using the QIAfilter Plasmid Maxi or Midi Kit (Qiagen). Bacteria from a 100ml bacterial culture were harvested by centrifugation (6000g, 4°C) for 15 minutes. The bacterial pellet was resuspended in resuspension buffer (50mM Tris-Cl, pH 8.0, 10mM EDTA, 100µg/ml RNase A) and lysis buffer (200mM NaOH, 1% SDS) was added and gently mixed. The bacterial lysate was incubated at room temperature for 5 minutes. Chilled neutralization buffer (3.0M potassium acetate, pH5.5) was added, mixed gently and the bacterial lysate was transferred to a QIAfilter cartridge and incubated at room temperature for 10 minutes. Meanwhile, a QIAGEN-tip was equilibrated with equilibration buffer (750mM NaCl, 50mM MOPS, pH 7.0, 15% isopropanol, 0.15% Triton® X-100). The bacterial lysate was gently pushed through the QIAfilter cartridge onto the column. The column was washed twice with wash buffer (1.0M NaCl, 50mM Tris, Tris-Cl, pH8.5, 15% isopropanol). The DNA was eluted from the column by the addition of elution buffer (1.6M NaCl, 50mM MOPS, pH 7.0, 15% isopropanol) and collected in a Corex™ tube. The DNA was precipitated with 0.7 volumes of isopropanol and pelleted by centrifugation (15000g, 4°C) for 30 minutes. The pellet was resuspended in nuclease-free dH₂O.

2.1.6. Isolation of Genomic DNA from Gram Negative bacteria. Genomic DNA was extracted from gram negative bacteria using the Wizard® Genomic DNA Purification kit (Promega). The protocol was adapted for use with gram negative bacteria. Briefly, the cells from 1ml of an overnight culture were pelleted by centrifugation (14000g, 2 minutes). The supernatant was discarded and the cells were resuspended in 600µl Nuclei Lysis solution. The cells were incubated at 80°C for 5 minutes and cooled to room temperature. 3µl of RNase solution was added to the lysed cells and mixed gently. The cells were incubated at 37°C for 15-60 minutes. The sample was cooled to room temperature prior to the addition of 200µl Protein precipitation buffer. The sample was vortexed vigorously for 20 seconds and incubated on ice for 5 minutes. The sample was centrifuged at 14000g for 3 minutes to pellet the proteins and the supernatant was transferred to a clean tube. The DNA was precipitated by the addition of 600µl of room temperature isopropanol and gently mixed by inversion. The DNA was pelleted by centrifugation (14000g, 2 minutes). The supernatant was removed and 600µl of 70% ethanol was added to wash the pellet. The sample was centrifuged at 14000g for 2 minutes and the ethanol was removed. The pellet was air-dried for 10-15 minutes. The DNA was resuspended in 100µl of DNA Rehydration solution and incubated at 65°C for 1 hour.

2.1.7. Nucleotide dideoxy sequencing of recombinant DNA clones. Alex Houston performed automated sequencing on a Perkin Elmer ABI Prism™ 377 DNA sequencer. Sequencing primers were designed for the T7 and SP6 polymerase promoters, a region immediately preceding the multiple cloning site in the pTrc99a vector and internal regions of eBFP (see table 2.1.). A sample containing 500ng of plasmid DNA and 4-5pmol of the appropriate oligonucleotide primer was submitted for sequencing. The sequencing data was viewed using Editview software from Applied Biosystems.

T7 promoter primer (20mer)	5'-d(TAATACGACTCACTATAGGG)-3'
SP6 promoter primer (19mer)	5'-d(TATTTAGGTGACACTATAG)-3'
BFP-N primer	5'-d(CGTCGCCGTCCAGCTCGACCAG)-3'
BFP-C primer	5'-d(CATGGTCCTGCTGGAGTTCGTG)-3'
pTrc99a sequencing primer (18mer)	5'-d(AGCGGATAACAATTCAC)-3'

Table 2.1. Nucleotide sequences of oligonucleotide sequencing primers used in automated DNA sequencing.

2.2. CONSTRUCTION OF DNA CASSETTES.

2.2.1. Materials. Oligonucleotides were synthesized by phosphoramidite chemistry on an Applied Biosystems 381A by Ian Armit or obtained from Oswel Ltd. The Taq polymerase was obtained from Promega. T4 DNA ligase was supplied by New England Biolabs.

2.2.2. Determination of DNA concentration. The concentration of DNA was calculated by monitoring the absorbance levels at 260nm in a spectrophotometer. The purity of the sample was estimated by monitoring the A260/A280 ratio.

2.2.3. Polymerase Chain Reaction (PCR). The PCR was used to amplify regions of DNA for cloning and also to introduce specific mutations within a gene. A typical reaction was performed in a 100µl volume and contained 20ng of template DNA, 200pmol of each primer, 25mM of each dNTP, 2.5mM MgCl₂ and 10µl of x10 Taq polymerase reaction buffer (50mM KCl, 10mM Tris.HCl (pH 9.0), 0.1% [v/v] Triton X-100, 1.5mM MgCl₂). The reaction mix was overlaid with an equal volume of mineral oil to prevent evaporation during thermal cycling. The reactions were initially heated to 94° for 2 minutes then held at 85°C for 1 minute while 2 units of Taq polymerase (Promega) were added to the aqueous phase of each reaction. Amplification of DNA was carried out on the thermal cycler following the subsequent parameters: 94°C for 1 minute, to denature the DNA; 50°C for 1 minute; to allow primers to anneal to the template DNA; 72°C for 1 minute for every thousand base pairs to be amplified. The amplification was performed for 25 to 30 cycles, finally the reaction was held at 72°C for 5 minutes to ensure that the majority of final product was full-length double-stranded DNA. The annealing temperature of the reaction was varied according to the particular base composition of the primers involved.

2.2.4. Direct purification of PCR products. PCR products were purified using the Wizard® PCR preps kit (Promega). One Wizard® minicolumn was prepared per sample, the minicolumn was attached to a syringe barrel. 100µl of direct purification buffer (50mM KCl, 10mM Tris-HCl (pH 8.8), 1.5mM MgCl₂, 0.1% Triton® X-100) was added to the PCR reaction and vortexed. 1ml of DNA purification resin was added to the sample and mixed by vortexing briefly. The DNA/resin mix was transferred to the syringe and pushed through the minicolumn. The minicolumn was washed with 2ml of 80% isopropanol. The column was dried by centrifugation for 2 minutes at 10000g. The bound DNA was eluted from the column by the addition of 50µl of dH₂O, incubated for 1 minute and centrifugation at 10000g for 20 seconds.

2.2.5. Annealing of oligonucleotide adapters. 100mM stock solutions of the oligonucleotide adapters were prepared. 20µl of the 100mM stock of each oligonucleotide adapter were mixed together. A series of dilutions of this mixture were produced. To anneal the oligonucleotide adapters they were boiled at 100°C for 2 minutes. The annealed oligonucleotide adapters were diluted to give a series of concentrations and ligated into suitably restricted vector.

2.2.6. Ligation of vector and insert DNA. Ligation reactions were generally carried out in a final volume of 25µl. The reaction mix contained 1µl T4 DNA ligase, 0.5µg of vector DNA and insert DNA at concentrations of 2 fold, 5 fold or 10 fold molar ratios in T4 DNA ligase buffer (50mM Tris-HCl (pH 7.5), 10mM MgCl₂, 10mM dithiothreitol, 1mM ATP, 25µg/ml BSA). The reactions were incubated for 16 hours at 16°C.

2.3. TRANSFORMATION OF *Escherichia coli*.

2.3.1. Materials and solutions. The *E.coli* strains JM109 (*rec A1 sup E44 end A1 hsd R17 gyr A96 rel A1 thi Δ(lac-proAB)*); F'[*traD36 proAB⁺ lac Iq lac ZΔM15*] JM105 (*sup E end A sbc B15 hsd R4 rps L thi Δ(lac-proAB)*); F'[*tra D36 proAB⁺ lac Iq lac ZΔM15*] and XL1-Blue (*sup E44 hsd R17 rec A1 end A1 gyr A46 thi relA1 lac⁻*; F'[*proAB⁺ lac Iq lac ZΔM15 Tn10(tet^r)*]) were routinely used for transformations. The chromogenic lacZ substrate X-Gal and the β-glucuronidase substrate X-GlcA were supplied by Melford Laboratories.

Luria Broth (LB)	1% [w/v] bacto-tryptone
	0.5% [w/v] bacto-yeast extract
	1% [w/v] NaCl pH 7.0

2.3.2. Preparation of competent *E.coli*. Competent *E.coli* were prepared and transformed by one of the following methods. Alternatively, high-efficiency competent JM109 cells were obtained from Promega and transformed.

2.3.2.1. Calcium chloride method. This method was used in the majority of cases for both transformation of plasmids and ligation reactions. 10ml of LB was inoculated with 100μl from an overnight culture of LB inoculated with *E.coli*, and incubated on a shaker at 37°C until the exponential phase (OD₆₀₀=0.4) was reached. The cells were cooled to 0°C, pelleted at 4°C and the supernatant was discarded. The pellet was resuspended in 10ml of ice-cold 0.1M calcium chloride and incubated on ice for 30 minutes. The cells were pelleted at 0°C, the supernatant discarded and the pellet resuspended in 1ml ice-cold 0.1M calcium chloride. The cells were stored on ice until required. The transformation efficiency of cells prepared by this method was at an optimum between 12 and 24 hours after

preparation. To transform the cells 0.5µg of DNA, or a ligation reaction was added to a 200µl aliquot of cells and incubated on ice for 30 minutes. The cells were subsequently heat-shocked at 42°C for 90 seconds then placed on ice immediately for 2 minutes prior to plating out onto LB agar plates containing the appropriate antibiotics.

2.3.2.2. Modified Hanahan method. This method was used when cells with a higher transformation efficiency were required. 5mM MgCl₂ and 5mM MgSO₄ were added to 25ml SOB (2% [w/v] bacto-tryptone, 0.5% [w/v] bacto-yeast extract, 0.05% [w/v] NaCl, 0.25M KCl, 10mM MgCl₂), this was inoculated with 250µl bacterial culture and incubated on a shaker at 37°C overnight. 250µl of the overnight culture was used to inoculate 25ml SOB plus 5mM MgCl₂ and 5mM MgSO₄, this was incubated at 37°C on a shaker until the cells reached log phase. The cells were cooled to 0°C for 10 minutes, pelleted at 4°C and the supernatant discarded. The cells were resuspended in 2ml cold TFB (10mM MES (2-[N-morpholino]ethansulphonic acid), 45mM MnCl₂.4H₂O, 10mM CaCl₂.2H₂O, 100mM KCl, 3mM Hexaminecobaltchloride), incubated on ice for 30 minutes, pelleted at 4°C and the supernatant discarded. The cells were resuspended in 2ml cold TFB and incubated on ice for 5 minutes. 70µl Dimethylformamide (DMF) was added and the cells were incubated on ice for 5 minutes. 5.2µl β-mercaptoethanol was added to 100µl KMES (1M MES, pH to 6.3 with 5M KOH), 70µl of this solution was added to the cells that were incubated on ice for a further 5 minutes. 200µl aliquots of cells were added to the DNA in pre-cooled tubes and incubated on ice for 30 minutes. The cells were heat-shocked at 42°C for 2 minutes then plated out onto LB agar plates containing the appropriate antibiotics.

2.3.2.3. Promega high efficiency JM109 competent cells. This method was used for transformations of ligations of the plant vector pPZP111 that has very low transformation efficiency. The cells are stored at -70°C therefore, prior to use they were thawed on ice

for 5 minutes. The DNA to be transformed was put into pre-cooled tubes and 100µl of cells were added. The cells were incubated on ice for 10 minutes then heat-shocked at 42°C for 45 seconds. The cells were immediately placed on ice for 2 minutes prior to plating out onto LB agar plates containing the appropriate antibiotics.

2.3.3. Screening recombinant clones.

2.3.3.1. Screening for recombinant clones using α -complementation. The multiple cloning site in the pGEM vector series occurs within the coding sequence for the lacZ gene. LacZ encodes the β -galactosidase that converts the chromogenic substrate X-gal producing a blue colony. Therefore if the lacZ coding sequence is interrupted by the introduction of an insert β -galactosidase is not produced thus recombinant colonies are white.

The JM109 *E.coli* transformed with ligations of pGEM vectors and insert were plated onto antibiotic selective LB agar plates containing IPTG (40mg/ml) to induce lacZ expression and the chromogenic substrate X-gal (40mg/ml).

2.3.3.2. Screening for GUS expression from plasmid DNA. Competent JM105 *E.coli* cells were prepared using the CaCl₂ method. An aliquot of competent cells was transformed with 1µl of maxiprep plasmid DNA. The transformed cells were plated out onto LB-agar plates containing 100mg/ml ampicillin, 5mM IPTG and 50mg/ml X-GlcA and incubated overnight at 37°C. The plates were stored at 4°C to allow the blue colour to develop.

2.4. TRANSCRIPTION AND TRANSLATION *IN VITRO*.

2.4.1. *Materials.* Sephadex® G-50 was obtained from Pharmacia biotech. The radiolabelled Redivue ³⁵S-L-Methionine was supplied by Amersham Pharmacia Biotech.

2.4.2. *Preparation of template DNA for in vitro translation.* The DNA used to program the transcription and translation reactions was purified to remove any contaminants that may adversely affect the transcription or translation process. Several methods were used as outlined below.

2.4.2.1. *Phenol/chloroform extraction of DNA.* An equal volume of phenol/chloroform (v/v) was added to the DNA sample, this was vortexed and centrifuged (14000g, 5 minutes). The upper aqueous layer was transferred to a clean tube, an equal volume of phenol/chloroform was added and the sample was vortexed and centrifuged (14000g, 5 minutes). The upper aqueous layer was transferred to a fresh tube and the DNA was precipitated by the addition of 2.5 volumes of ethanol and one 20th volume of 2M sodium acetate. The sample was stored at -70°C for a minimum of 30 minutes. The DNA was pelleted by centrifugation at 0°C (14000g, 30 minutes). The ethanol was removed, the sample was dried and resuspended in distilled water.

2.4.2.2. *Spin column chromatography.* A Sephadex® G-50 column was prepared by plugging a disposable 1ml syringe with sterile glass wool. The column was filled with swollen Sephadex® G-50. The column was placed in a 25ml centrifuge tube and centrifuged at 2500g for four minutes to pack the column. The process was repeated until the packed Sephadex® G-50 level was above 0.9ml on the syringe. The packed column was washed three times by the addition of 1 volume of distilled water and centrifugation at 2500g for four minutes. The DNA sample (one volume) was applied to the column. The

column was centrifuged and the eluted sample collected in a de-capped, sterile microcentrifuge tube.

2.4.2.3. Wizard® DNA Clean-up kit. (Promega) One Wizard minicolumn was attached to a disposable syringe per DNA sample. 1 ml of Wizard DNA clean-up resin was added to the DNA sample and gently mixed. The DNA/resin was transferred to the syringe and slowly pushed through the minicolumn. The minicolumn was washed with 2ml of 80% isopropanol. The minicolumn was dried by centrifugation at 10000g for 2 minutes. The DNA was eluted by the addition of 50µl of warmed (65°C-70°C) dH₂O to the minicolumn. The minicolumn was incubated for 1 minute prior to centrifugation at 10000g for 20 seconds.

2.4.3. Coupled transcription and translation (TnT) reactions. Proteins were expressed *in vitro* using coupled transcription/translation kits (Promega). Proteins were expressed in both wheatgerm extract and rabbit reticulocyte lysate systems according to the manufacturers instructions (Promega). The proteins were radiolabelled with ³⁵S-methionine (Amersham). The reactions were incubated at 30°C for 90 minutes then stopped by the addition of an equal volume of 2xSDS protein loading buffer. Aliquots of the translation reactions were analyzed by denaturing PAGE.

2.4.4. Two-step transcription and translation system. Proteins were expressed *in vitro* from a DNA template using the single tube protein system 3 supplied by Novagen following the manufacturers protocol. The transcription reactions were incubated at 30°C for 15 minutes prior to addition of the translation reaction components. Proteins were labelled with ³⁵S-methionine (Amersham). The translation reactions were incubated at 30°C for 60 minutes then stopped by the addition of an equal volume of 2xSDS loading buffer. Aliquots were analyzed by denaturing PAGE.

2.5. PROTEIN ANALYSIS AND DETECTION.

2.5.1. Materials and solutions. The HybondTMECLTM nitrocellulose membrane was supplied by Amersham Life Science. The polyclonal GFP antibody and the Living Colours peptide antibody were obtained from Clontech. The GUS and CAT antibodies were both supplied by 5 Prime -> 3 Prime. The horseradish peroxidase conjugated anti-rabbit antibody was provided by Amersham life science.

2 x SDS-PAGE loading buffer	124mM Tris.HCl (pH6.8)
	4% SDS
	20% glycerol
	10% β -mercaptoethanol
	0.2% bromophenol blue

2.5.2. Protein concentration assay. The concentration of protein was assayed using a protein assay kit supplied by Bio-rad. The dye reagent was prepared by diluting 1 part of the dye reagent concentrate with 4 parts of dH₂O. A series of dilutions of the protein standard (bovine serum albumin) were prepared. 5 μ l of each standard and sample solution were assayed in 1ml of diluted dye reagent. The samples were thoroughly mixed and incubated at room temperature for 5 minutes. The absorbance at 595nm was measured on a spectrophotometer. The protein concentration of the samples was calculated from the standard curve.

2.5.3. Denaturing PAGE 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 stated gels were routinely constructed with a 4% polyacrylamide stacking gel and a 10% polyacrylamide resolving gel. The gels were prepared in a minigel Protean-II system (Hoefer®).

Solution	10% resolving gel	4% stacking gel
40% acrylamide	1.25ml	0.2ml
2% bis-acrylamide	0.863ml	0.33ml
1M Tris	(PH 8.8) 0.92ml	1.176ml
10% SDS	1.875ml	(pH 6.8) 0.25ml
10% Aps	50 μ l	20 μ l
TEMED	8.3 μ l	4 μ l
dH ₂ O	33.3 μ l	20 μ l

Table 2.2. Recipes for SDS-PAGE analysis.

Typically electrophoresis of each gel was carried out in 1X Tris-glycine buffer (0.1% [w/v] SDS, 25mM Tris, 250mM Glycine) at a constant current of 15-20mA throughout the stacking gel and 30-40mA through the resolving gel.

2.5.4. Coomassie brilliant blue staining. The proteins separated by denaturing-PAGE were visualized by staining with Coomassie brilliant blue. The gel was soaked in coomassie brilliant blue stain (20% [v/v] methanol, 20% [v/v] glacial acetic acid, 0.2% [w/v] coomassie brilliant blue) on an orbital shaker for 30 minutes. The staining solution was discarded. The gel was destained in (20% [v/v] methanol, 10% [v/v] glacial acetic acid) until the background was clear. The destaining solution was changed frequently.

2.5.5. Fixing radiolabelled gels. The denaturing-PAGE gels containing proteins labelled with ³⁵S-methionine were fixed by soaking the gel in fixing solution (10% [v/v] glacial acetic acid, 50% [v/v] methanol) on an orbital shaker for 30 minutes. The fixing solution was replaced with fresh solution and incubated for a further 30 minutes. The high methanol content of the fixing solution reduces the size of the gel, therefore, to resize the gel it was soaked in destain solution (20% [v/v] methanol, 10% [v/v] glacial acetic acid) for 15 minutes.

2.5.6. Autoradiography. The fixed gels were dried onto Whatmann 3MM paper at 80°C under vacuum on a gel drier for 1 hour. The gels were exposed to X-ray film (Kodak Biomax™MR film) at room temperature for the required time period (usually overnight). The autoradiograph was developed using a Kodak M35 X-OMAT Processor.

2.5.7. Phosphorimaging analysis. Phosphorimaging plates were exposed to the dried radioactive gel for 1 hour. The plate was scanned by the FujiX Bas1000 phosphorimager. The scanned gel was analyzed using MacBas v2.0 image analysis software.

2.5.8. Western blotting. The proteins were separated by SDS-PAGE, the gel was soaked in transfer buffer (25mM Tris, 200mM Glycine, 20% [v/v] methanol) for 15 minutes. The nitrocellulose membrane and filter paper were presoaked in transfer buffer for 5 minutes. The membrane and gel were placed in the blot between six layers of soaked filter paper taking care to remove all air bubbles. The blot was placed in the semidry blotting apparatus and an appropriate current (0.5mA/cm²) was applied for 90 minutes. The gel was stained with Coomassie brilliant blue stain to verify that all the proteins had been transferred to the membrane.

Western analysis was performed on the membrane using the ECL™ kit following the protocol outlined by the manufacturer. Briefly the blot was blocked in blocking solution (5% dried milk in 0.1% PBS-T) on an orbital shaker for 1 hour. The membrane was washed three times in washing buffer (0.1 % Tween™ 20 in PBS) for 10 minutes. The blot was incubated in primary antibody on an orbital shaker for 1 hour then washed as before. The blot was incubated in secondary antibody on an orbital shaker for 45 minutes. The membrane was washed four times in washing buffer for ten minutes.

The following antibody concentrations were used, antibodies were diluted in blocking solution (5% dried milk in 0.1% PBS-T):

Polyclonal GFP antibody	1:1000
Living colours peptide antibody	1:200
Anti-GUS antibody	1:3000
Anti-CAT antibody	1:3000
Anti-rabbit antibody	1:4000

The secondary antibody was conjugated to horseradish peroxidase. The antibody was detected by chemiluminescence. Equal volumes of detection reagent 1 and detection reagent 2 were mixed together. The membrane was covered with the mixed detection solution (0.125ml/cm² membrane) and incubated at room temperature for 1 minute. The excess detection reagent was drained and the membrane was wrapped in SaranWrap™. The blot was exposed to film for 15 seconds initially and the exposure time was adapted accordingly.

2.6. PROTEIN EXPRESSION IN *Escherichia coli*.

An aliquot of competent JM105 cells was transformed with 1µl of plasmid DNA. 10ml of LB broth containing 100µg/ml ampicillin was inoculated with a single colony and incubated on a shaker at 37°C for 16 hours. 10ml of LB broth containing 100µg/ml ampicillin was inoculated with 100µl of culture and incubated on a shaker at 37°C for 3 hours. Protein expression was induced by the addition of 5mM IPTG. The cells were incubated on a shaker at 30°C for 5 hours. A 1ml aliquot of culture was removed for analysis. The cells were pelleted by centrifugation (10000rpm, 3 minutes) and the

supernatant was removed. The pellet was resuspended in 1ml of 1xSDS protein loading buffer and boiled for four minutes. The proteins were analyzed by denaturing PAGE.

2.7. TRANSFORMATION OF *Agrobacterium tumefaciens*.

2.7.1. *A. tumefaciens* transformation. *Agrobacterium tumefaciens* strain LBA4404 that contains the *vir* helper plasmid L4404 was transformed by one of the following methods.

2.7.1.1. Freeze-thaw transformation. 10ml of LB medium was inoculated with a colony of *Agrobacterium tumefaciens* and grown on a shaker for 48 hours at 28°C. The culture was cooled on ice before pelleting by centrifugation. The pellet was resuspended in 250µl ice-cold 20mM CaCl₂ and dispensed into 100µl aliquots. One µg of DNA was added to each aliquot of bacteria. The cells were frozen in liquid Nitrogen and thawed at 37°C for 5 minutes. 1ml of fresh LB medium was added to each transformation and the culture was incubated at 28°C for 3 hours. Aliquots of the transformation were spread onto LB agar plates containing the appropriate antibiotics and incubated at 28°C for 2-3 days.

2.7.1.2. Electroporation of *Agrobacterium tumefaciens*. 10ml of LB medium was inoculated with a colony of *Agrobacterium tumefaciens* and grown on a shaker for 48 hours at 28°C. 4ml of the bacterial culture were pelleted by centrifugation (3000g, 10 minutes, 4°C). The pellet was resuspended in 1ml of sterile dH₂O and centrifuging at 15000g for 2 minutes repelleted the bacteria, this was repeated three times. The bacteria were resuspended in 100µl of dH₂O. 1µg of plasmid DNA was added to a 40µl aliquot of the washed bacteria and transferred to an electroporation cuvette. An electrical pulse of 180V was applied for 2 seconds. 1ml of LB media was added to the shocked bacteria. The bacteria were incubated at 28°C for 3 hours prior to plating onto LB agar plates containing the appropriate antibiotics and incubated at 28°C for 2-3 days.

2.8. *Agrobacterium*-MEDIATED TRANSFORMATION OF *Nicotiana tabacum*.

2.8.1. *Materials.* Murashige and Skoog medium (MSO) was prepared from a complete medium concentrate supplied by GibcoBRL. The plant hormones and carbenicillin were also obtained from GibcoBRL.

2.8.2. *Transformation of *Nicotiana tabacum* plants.* Two alternative methods were used. The traditional leaf-disc method was the preferred method.

2.8.2.1. *Transformation of *Nicotiana tabacum* seedlings.* The protocol used was adapted from protocols outlined by Tinland *et al.* (1995). *Nicotiana tabacum* seeds were sterilized in 10% bleach solution for 20 minutes. The sterilized seeds were washed twice in sterile dH₂O and spread onto sterile 3MM Whatmann filter paper to dry. The seeds were transferred to MSO solid media plates. The plates were sealed with parafilm and incubated at 25°C to allow the seedlings to grow for 1-2 weeks. 10ml cultures of *Agrobacterium* were grown at 25°C under antibiotic selection for 48 hours. A 3ml aliquot of bacteria was pelleted and washed twice in sterile 10mM MgSO₄ solution. The bacteria were pelleted and resuspended in 10ml of sterile MSO media. Approximately 100 two week old seedlings were added to each bacterial culture. The seedlings were exposed to a reduced pressure of 0.15 atm for 5 minutes. The seedlings were plated onto fresh MSO solid media plates containing the hormones 6-benzylaminopurine (BA) and α -naphthaleneacetic acid (NAA) and incubated at 25°C for 3 days. The plantlets were washed in 10mM MgSO₄ containing carbenicillin (500 μ g/ml), kanamycin (100 μ g/ml), NAA (5 μ g/ml) and BA (50 μ g/ml) and plated onto fresh MS solid media plates containing the same cocktail of antibiotics and plant hormones. The plantlets were transferred to fresh plates weekly for 4-6 weeks until calli had formed. The calli were removed and a section of the plantlet was

transferred to fresh MSO solid media containing kanamycin only and incubated at 25°C to allow roots to develop. After 4-6 weeks the plant was harvested.

2.8.2.2. Transformation of *Nicotiana tabacum* leaf-discs. Leaves were harvested from sterile plants and cut into 0.5cm² discs. The discs were placed on MSO solid media plates containing the plant hormones BA (50µg/ml) and NAA (5µg/ml). The plates were sealed with parafilm and incubated at 25-28°C for 48 hours. Cultures of LB containing rifampicin (50µg/ml) and kanamycin (10µg/ml) were inoculated with the transformed *Agrobacterium* strains and incubated on a shaker at 28°C for 48 hours. After incubation for 48 hours the leaf discs were submerged in the *Agrobacterium* culture, drained on sterile filter paper and returned to the MSO plates. The plates were resealed with parafilm and incubated at 25-28°C for a further 48 hours. The leaf discs were transferred to selective MSO solid media plates containing the antibiotics carbenicillin (500µg/ml) and kanamycin (100µg/ml) and plant hormones BA (50µg/ml) and NAA (5µg/ml). The plates were sealed with parafilm and incubated at 25-28°C under low light intensity for 4-6 weeks. The transgenic shoots were excised and individually placed into pots of MSO media containing kanamycin (100µg/ml). After 4-6 weeks the rooted plants were micropropagated and samples for analysis were harvested.

2.8.3. Sterile micropropagation and harvesting. The plant was dissected at root level and the stem was cut into sections containing one leaf. The leaf was removed and the stem section was placed in a pot of fresh selective MSD4x2 media containing kanamycin. After 4-6 weeks a new plant developed.

The remaining leaves were collected and stored at -70°C. A crude protein extract from a 5g tissue sample was prepared. A pestle and mortar was cooled with liquid Nitrogen prior to addition of the sample. The sample was ground to a powder in liquid Nitrogen,

transferred to a microcentrifuge tube and 200 μ l of protein extraction buffer (0.1M Tris pH7.5, 2% polyvinylpyrrolidone, 2% polyethylene glycol, 10mM β -mercaptoethanol) was added. The sample was centrifuged at 4°C, 15000rpm for 10 minutes. The supernatant was transferred to a clean microcentrifuge tube and centrifuged as before. The supernatant was transferred to a clean microcentrifuge tube and stored at -20°C.

2.8.4. GUS assay of leaf sections. The GUS assay extraction buffer (50mM NaPO₄ (pH 7.0), 10mM EDTA, 0.1% Triton® X-100, 0.1% Sarkosyl, 10mM β -mercaptoethanol) was prepared by adding 1mg of the substrate X-GlcA per ml of buffer. A section of leaf was removed from the plant and placed into GUS assay extraction buffer. The samples were placed under vacuum for 30 minutes. The samples were incubated at 37°C overnight to allow the blue colour to develop. The leaf sections were treated with ethanol to remove the chlorophyll.

2.8.5. Preparation of plant tissue sections for microscopy.

2.8.5.1. Protocol adapted for fixing and paraffin embedding plant tissue. The protocol outlined below was adapted for use with plant tissue by Pamela Gray. The wax sections were prepared by Jill McVee.

Day 1. The plant tissue was dissected and placed immediately into the fixative solution (20g dm⁻³ formaldehyde, 7.5g dm⁻³ sucrose in 100mM NaH₂PO₄, pH 7.2). The tissue was submerged by vacuum infiltrating the solution. The fixative solution was renewed and the tissue was incubated at 25°C for 24 hours.

Day 2. The material was washed in 0.85% NaCl for 30 minutes on ice and dehydrated in a series of ethanol in saline solutions (30% (v/v), 50%(v/v), 75%(v/v), 85%(v/v), 95%(v/v),

100%) for 90 minutes per solution. The tissue was stored overnight in 100% ethanol at 4°C.

Day 3. The tissue was placed in fresh 100% ethanol for 1 hour at room temperature and transferred to a solution of 1 part 100% ethanol and 1 part Histo-clear for 1 hour at room temperature. The tissue was transferred to 100% Histo-clear solution and incubated at room temperature for 1 hour, this step was repeated twice. The tissue was moved to fresh Histo-clear solution and one half volume of paraplast chippings was added. The tissue was incubated overnight at 40°C. The material was transferred to molten wax at 60°C and the molten wax was changed twice daily for 3 days. The material was put into flexible plastic moulds containing molten paraplast X-tra. The mould was floated on water to solidify the wax and the wax blocks were stored at 4°C. 5-6µm sections of the wax block were prepared using a microtome. Finally, the cut sections were dewaxed by rinsing twice in 100% Histo-clear solution for ten minutes.

CHAPTER 3: MOLECULAR CLONING AND *IN VITRO* ANALYSIS OF AN ARTIFICIAL SELF-PROCESSING POLYPROTEIN SYSTEM.

3.1. Part 1: Construction and *in vitro* characterization of control plasmids and the artificial reporter polyproteins.

3.1.1. Introduction.

In order to investigate the potential of developing an artificial polyprotein expression system utilizing the picornaviral proteases FMDV Lb^{PRO} and HRV14 2A^{PRO} a simple reporter system containing the reporter gene GFP linked to the viral protease maintaining an appropriate protease cleavage site was required. These proteases share similar proteolytic activities during the viral life-cycle they are both responsible for an autocatalytic *cis* cleavage event during polyprotein processing and also cleave the cellular protein eIF4G *in trans*. Although the activities of the proteases are similar they are structurally very different. FMDV Lb^{PRO} is a papain-like cysteine protease. HRV14 2A^{PRO} is a novel type of protease with a subtilisin-like serine protease fold but containing a cysteine residue as the active site nucleophile. The artificial reporter polyprotein was developed to test the individual cleavage activities of FMDV Lb^{PRO} and HRV14 2A^{PRO} when removed from the native polyprotein and attached to a reporter gene. Using this system we could investigate whether the proteases were still capable of autocatalytic cleavage activity without the presence of other viral sequences and also monitor the relative specificity and efficiency of the cleavage reaction.

The reporter gene wtGFP was selected for use within the artificial polyprotein system due to its small size (about 27.5kDa) in a bid to avoid previous difficulties of *in vitro* expression of internal initiation products from alternative reporter genes.

Since the basic artificial polyprotein system composed of the picornaviral protease and wtGFP reporter gene would be initially tested *in vitro* a series of transcription vectors containing the artificial polyprotein cassettes would be required. The vector

pGEM7zf+ (Promega) was selected for use as an appropriate vector permitting transcription of the cloned gene(s) from the bacteriophage T7 promoter.

3.1.2. Cloning strategies used to construct the artificial polyprotein system.

Addition of novel restriction enzyme sites by the polymerase chain reaction (PCR).

Novel restriction enzyme sites were introduced to the termini of the artificial polyprotein cassettes via the PCR as an aid for molecular cloning. Oligonucleotide primers were designed for the 5' and 3' ends of the cassette. The forward (5') primer contained a region of non-coding nucleotides preceding the restriction site to ensure correct trimming of the ends of the PCR fragment prior to ligation. The restriction site was followed by the initiation ATG codon and the first six codons of the 5' end of the gene which would anneal to the template.

The reverse (3') primer was designed in the same way beginning with the last six codons of the 3' end of the gene. Following the coding region a stop codon (TGA) was introduced prior to the restriction site. Finally, a length of non-coding nucleotides was included at the end of the primer. The sequence of the 3' reverse primer must be reversed and complemented to direct extension of the sequence by the polymerase from the 3' end of the template gene.

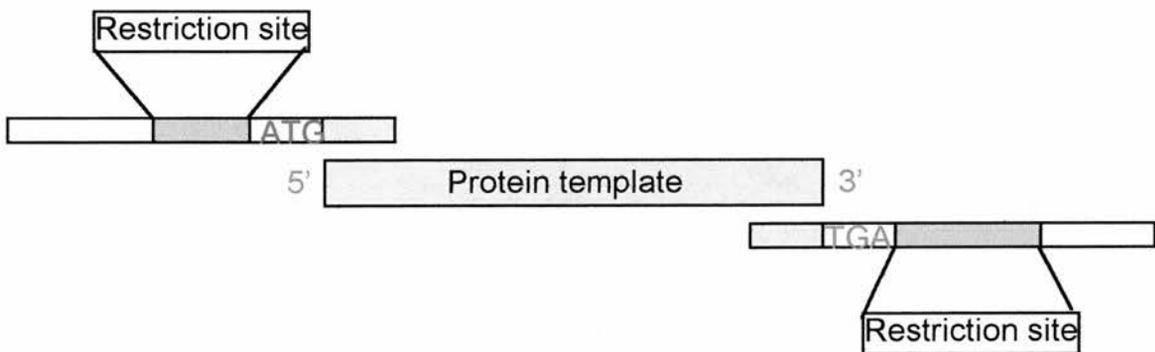


Fig.3.1. Addition of novel restriction sites to constructs by PCR. The basic primer design containing a restriction site, initiation/stop codon and a complementary region of the protein template is shown.

Tailoring of gene constructs using an overlap PCR strategy. An “overlap” PCR method was employed to create the polyprotein cassettes and to introduce single point mutations. Creation of the artificial polyprotein cassettes requires the individual genes to be precisely linked such that the open reading frame is maintained throughout the polyprotein. If a frameshift was introduced between the two genes the second protein would not be translated and premature termination would occur.

The overlap PCR strategy requires a set of four primers. The first primer pair contains a forward primer (Primer 1) to anneal to the 5' terminus of the gene this also introduces a novel restriction enzyme site to the 5'end. The reverse primer (Primer 2) anneals at a downstream position. To link two genes together the reverse primer is designed to anneal to the 3'end of the first gene and to add a nucleotide tail that is complimentary to the first six amino-acids of the second gene. Alternatively to introduce a mutation the reverse primer would anneal to the region of the template surrounding the mutation site, the codon of interest would be changed in the primer. The second primer pair is designed in the same way. The forward primer (Primer 3) would anneal to the 3' end of the second gene adding a 5' tail complementary to the C-terminal region of the first gene, or anneal to the region surrounding the mutation site. The reverse primer (Primer 4) anneals to the 3'end of the final product adding a second restriction site.

The first-round of PCR requires two separate PCR reactions using primers 1+2 and primers 3+4 to amplify the first round products. The first round products share a region of complementarity and can anneal together therefore they can be used as templates in the second-round PCR reaction which uses the primer pair (1+4) to generate the full-length product.

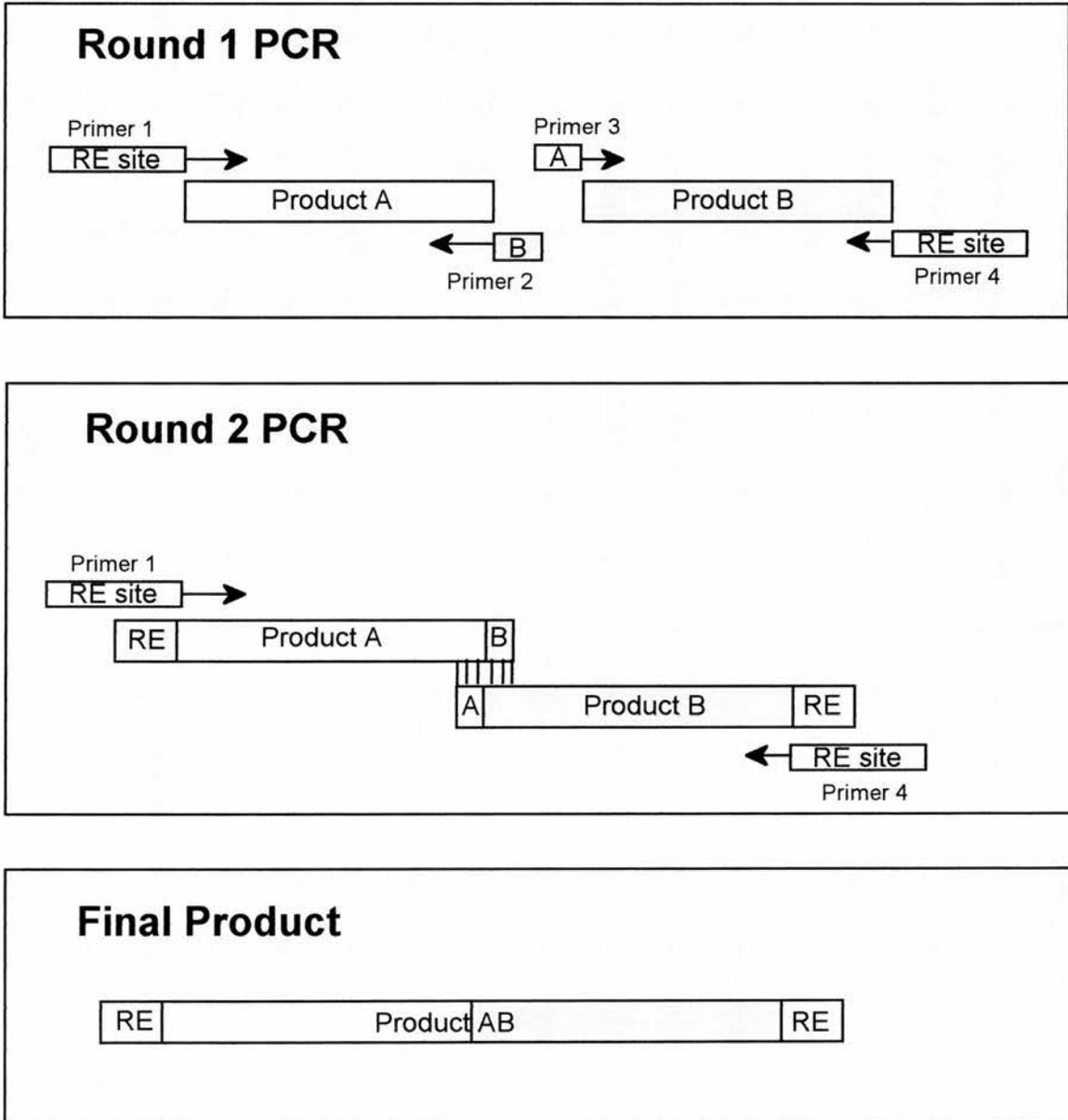


Fig.3.2. The overlap PCR strategy. This outlines the requirements for each stage of the overlap PCR method to create a polyprotein construct.

3.1.3. Construction of the control plasmid pwtGFP.

The plasmid pwtGFP was required as a control plasmid for the reporter gene wtGFP. This would permit translation of wtGFP *in vitro* and provide a control for the size of the wtGFP protein during PAGE analysis. Translation of the control plasmid that lacked the proteolytic enzymes could also indicate any effects on the *in vitro* system due to the proteolytic activities of FMDV Lb^{PRO} and HRV14 2A^{PRO}. Two complementary oligonucleotide primers were designed as described above, to anneal to the terminal amino acids of the wtGFP protein and to introduce restriction enzyme sites to the termini of wtGFP. The forward primer added a BamHI site adjacent to the ATG initiation codon of wtGFP. The reverse primer introduced a stop codon and a Sac I site downstream of the 3' end of wtGFP.

The primers were used to amplify by the PCR wtGFP from the plasmid PTXS-GFP2ACP donated by Simon Santa Cruz. The amplified wtGFP fragment was purified. The wtGFP fragment and the pGEM7zf⁺ vector were prepared for ligation by digestion with the BamHI and Sac I restriction enzymes. The restricted insert and vector were purified from agarose and ligated together. The ligation reaction was used to transform competent JM109 *E.coli* cells. Transformed cells were plated onto antibiotic selective media and screened for α -complementation. DNA was prepared from the putative clones. Analytical digests with restriction enzymes determined the presence of wtGFP.

To verify the insertion of wtGFP into pGEM7zf⁺ and to detect any mutations within wtGFP that may have occurred during the PCR process the plasmid DNA from the positive clones was sequenced by automated sequencing. The sequencing results confirmed that wtGFP had been successfully cloned into the pGEM7zf⁺ vector to create pwtGFP.

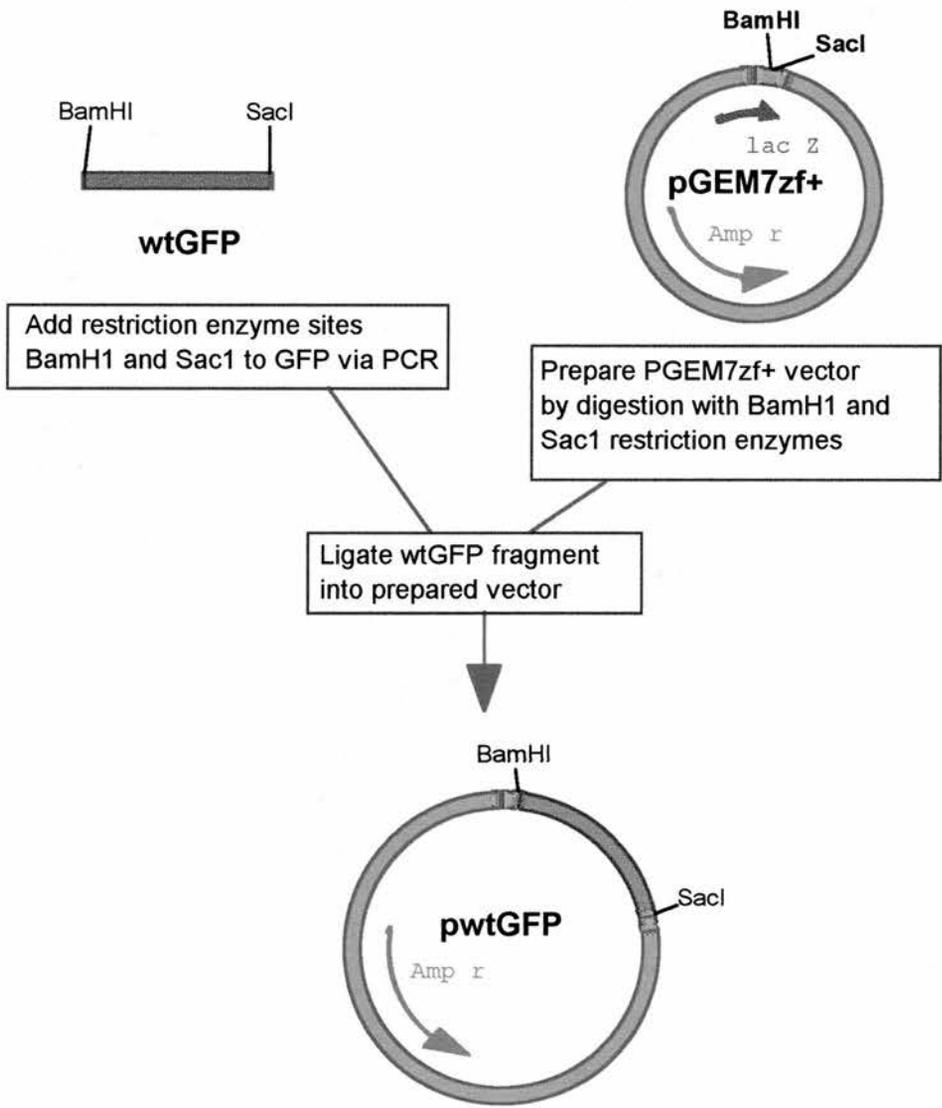


Fig.3.3. Construction of pwtGFP. The wtGFP coding sequence was amplified by PCR adding the restriction sites BamHI and Sac I. The PCR fragment and the vector pGEM7zf+ were digested with BamHI and Sac I and ligated to form pwtGFP.

3.1.4. Construction of the *pLb^{PRO}GFP* plasmid via an overlap PCR strategy.

The native FMDV Lb^{PRO} catalyses the *cis* cleavage of the L/1A junction at a Lys-Gly pair. Therefore to mimic the native polyprotein structure the reporter gene was added to the C-terminus of Lb^{PRO}. An overlap PCR strategy was developed to link together the FMDV Lb^{PRO} protein and the reporter gene wtGFP to produce the polyprotein Lb^{PRO}GFP. For the FMDV Lb^{PRO} protein an oligonucleotide primer was designed that added a BamHI restriction enzyme site to the 5' end of the protein. At the 5' end in order to maintain an appropriate Lb^{PRO} cleavage site the first four amino-acids of FMDV 1A were maintained. The oligonucleotide primer was designed to add a complementary region of the first six amino-acids of wtGFP following the cleavage site. In the same way a pair of oligonucleotide primers were designed that added the terminal cleavage site residues of FMDV Lb^{PRO}/1A to the N-terminus of wtGFP and a stop codon followed by a Sac I restriction site to the C-terminus.

The FMDV Lb^{PRO}/1A product was amplified by the PCR from plasmid pMR30 encoding FMDV L-P1-P2 Δ 3946-4693-P3 Δ 3D previously constructed by M. Ryan (M.Ryan pers.comm). The wtGFP product was also amplified by PCR using the PTXS-GFP2ACP plasmid as a template. The PCR products from the separate reactions were purified from agarose. In the second-round PCR reaction the additional complementary regions introduced during the first-round of PCR permits the individual products to anneal forming a template to generate full-length Lb^{PRO}GFP. The full-length product was amplified using the primers designed to add restriction enzyme sites to Lb^{PRO} and wtGFP.

The second-round PCR product and the vector pGEM7zf⁺ were prepared for ligation by digestion with BamHI and Sac I. The cut vector and insert were ligated and used to transform competent JM109 *E.coli* cells. The transformed cells were plated onto antibiotic selective media. Plasmid DNA from the putative clones was extracted and analyzed by restriction enzyme digestion to confirm the presence of Lb^{PRO}GFP. The

Lb^{PRO}GFP sequence was verified and checked for mutations by automated DNA sequencing using T7 and SP6 sequencing primers that anneal to the promoter regions of the pGEM7zf⁺ multiple cloning site.

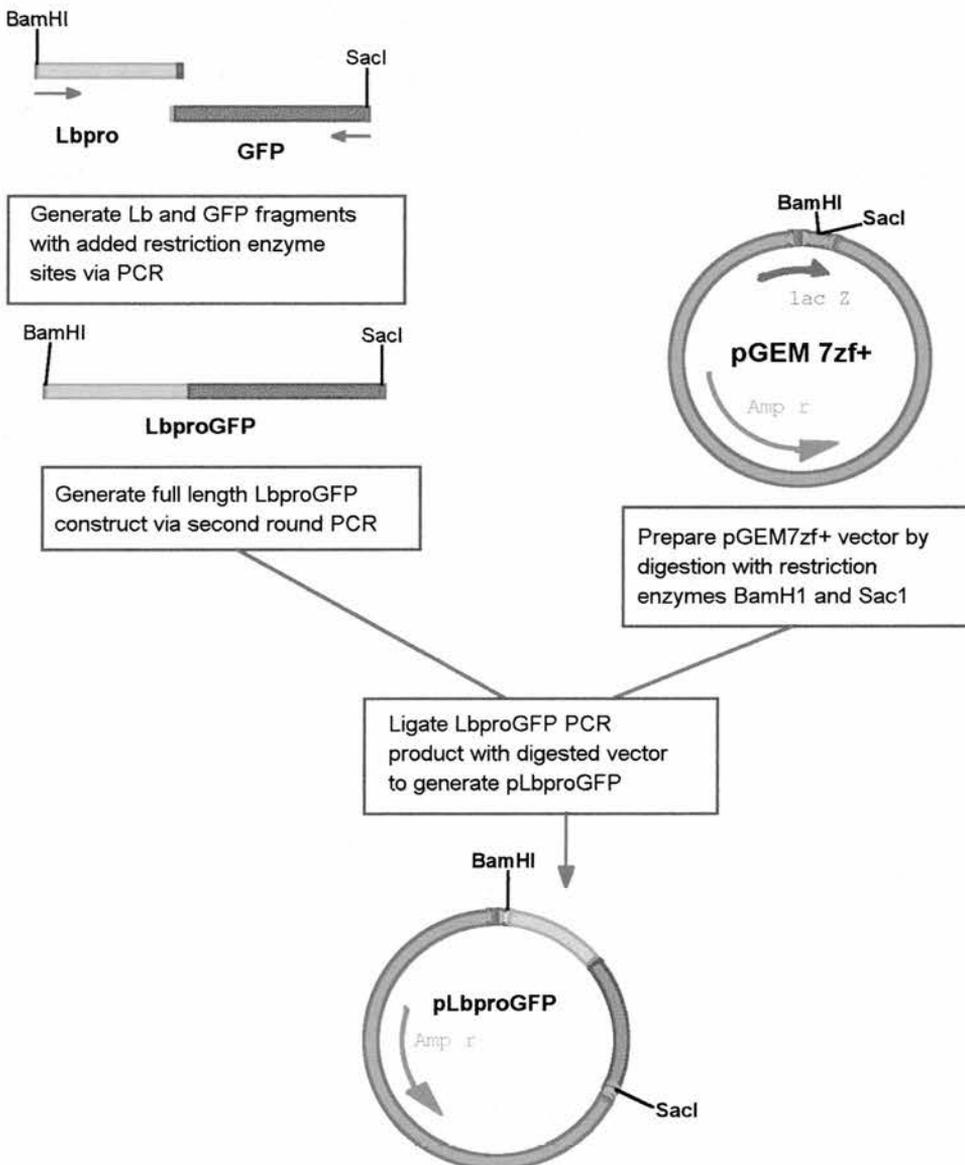


Fig.3.4. Construction of pLb^{PRO}GFP. The Lb^{PRO} and wtGFP PCR products were amplified in the first-round of PCR. The products were purified and used as templates in the second-round PCR reaction to generate Lb^{PRO}GFP. The Lb^{PRO}GFP fragment was digested with BamHI and Sac I and ligated into similarly restricted pGEM7zf⁺ to form pLb^{PRO}GFP.

3.1.5. Construction of the control plasmid pLbGFP.

A plasmid containing an inactive FMDV Lb^{PRO} was created as a control. Translation of the artificial polyprotein containing an inactive protease would produce a full-length product to be used during PAGE analysis to identify any uncleaved product. Mutating the active site nucleophile at position 23 from cysteine to an alanine residue destroyed the proteolytic activity of the enzyme. An overlap PCR strategy using the pLb^{PRO}GFP plasmid as a template was designed to introduce a single amino acid change at this position.

A series of four oligonucleotide primers was designed to generate two products from pLb^{PRO}GFP overlapping in the region of the mutation. During the first round PCR reactions the individual products Lb* and *LbGFP were amplified. The resulting PCR products were purified and used as template DNA for the second-round PCR reaction to generate a product containing the Cys to Ala amino-acid mutation.

The ends of the PCR product and the pGEM7zf⁺ vector were restricted by digestion with BamHI and Sac I and purified from agarose. The cut vector and insert were ligated and the ligation mix was used to transform competent JM109 *E.coli* cells. The transformed cells were plated onto antibiotic selective agar and screened for α -complementation. Plasmid DNA from putative clones was extracted and submitted for automated DNA sequencing analysis which confirmed that a single amino-acid mutation of the active site cysteine residue had been achieved.

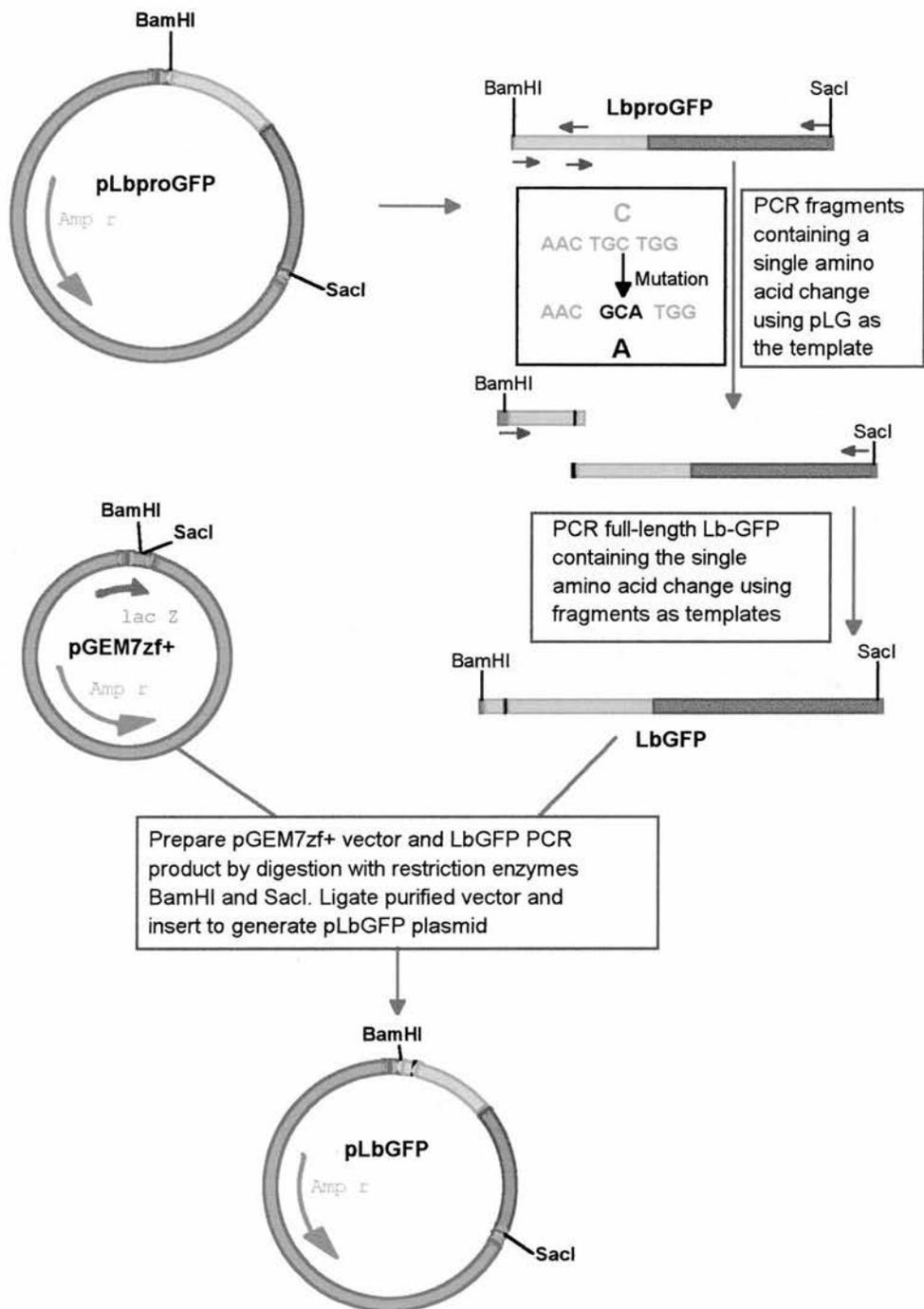


Fig.3.5. Construction of pLbGFP. Using pLb^{pro}GFP as a template an overlap PCR strategy was developed to introduce a single point mutation of the active site Cys-53 to Ala. The final PCR product LbGFP was digested with BamHI and Sac I and ligated into similarly restricted pGEM7zf+ to generate pLbGFP.

3.1.6. Construction of the artificial polyprotein plasmid pGFP2A^{pro}.

To investigate the viability of creating an artificial polyprotein system using the 2A^{pro} from HRV14 the transcription plasmid pGFP2A^{pro} was constructed. The HRV14 2A^{pro} cleaves the native polyprotein *in cis* at a Tyr-Gly dipeptide at its N-terminus thus separating P1 from P2-P3. Therefore in the artificial reporter polyprotein wtGFP was added upstream of the HRV14 2A^{pro}. The cassette was constructed to include the last six residues of ID in order to maintain the native cleavage site.

An overlap PCR strategy was designed to create the artificial polyprotein cassette sequence. A set of four oligonucleotide primers was devised that was used to amplify a wtGFP product and an HRV14 2A^{pro} product during the first round of PCR. The wtGFP product had a BamHI restriction site added to the 5' end and a region complementary to the 5' end of HRV14 2A^{pro} at the 3' end. The HRV14 2A^{pro} product amplified from a full-length HRV14 cDNA clone had an additional Sac I restriction site at the 3' end and a wtGFP complementary region at the 5' end. The complete polyprotein was amplified during the second-round of PCR using the 5'wtGFP and 3'HRV14 2A^{pro} primer pair with the first round products as templates.

The ends of the second-round 1209bp product and the pGEM7zf+ vector were restricted in preparation for ligation by digestion with BamHI and Sac I. The cut vector and restricted PCR product were purified from agarose and ligated. The ligation mix was used to transform competent JM109 *E.coli* cells and plated onto selective agar. Putative clones were identified on the plates by antibiotic resistance and α -complementation. Plasmid DNA was extracted from the putative clones and analyzed by restriction enzyme digests. Positive clones were confirmed by automated DNA sequencing.

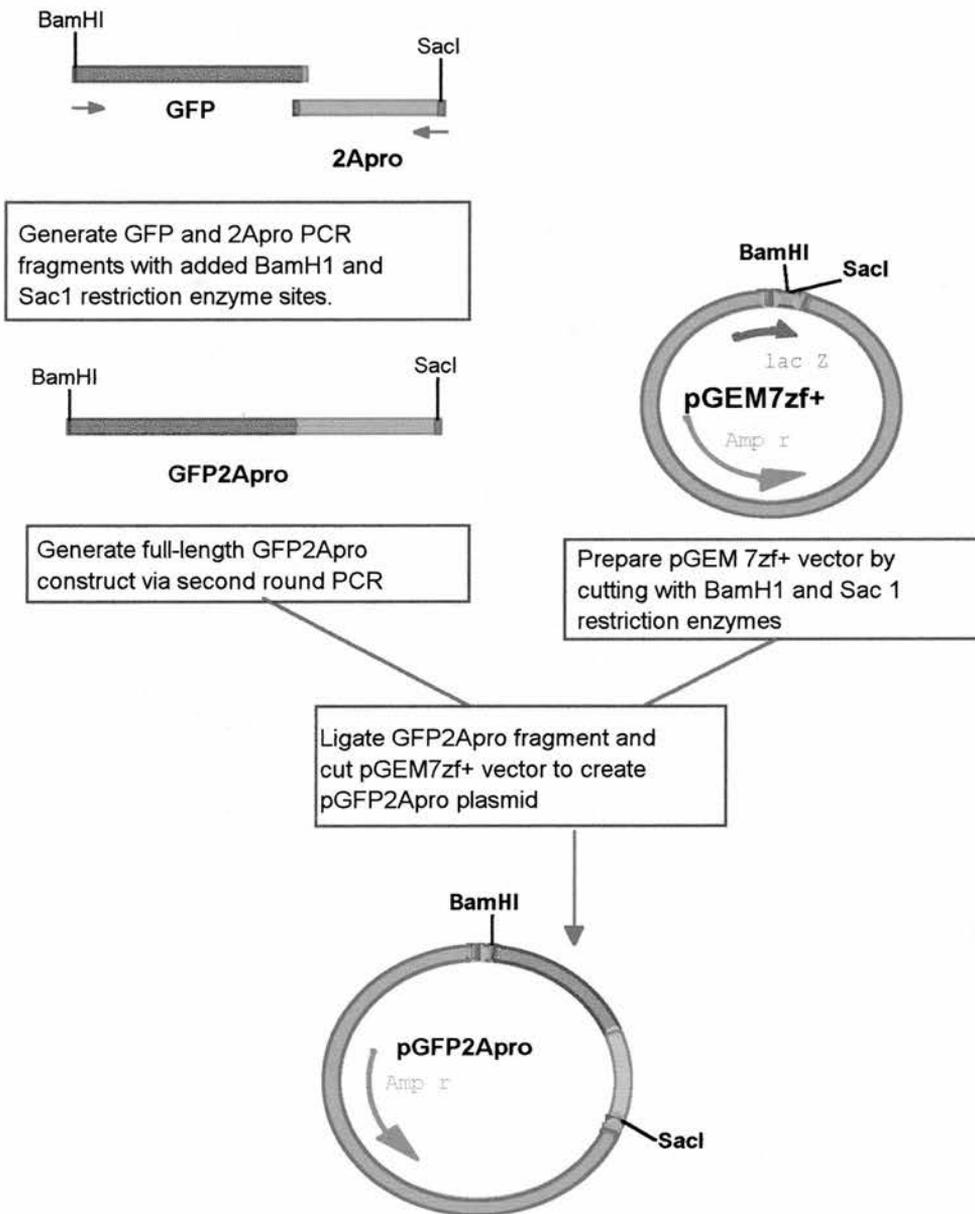


Fig.3.6. Construction of pGFP2A^{pro}. The first-round PCR products wtGFP and 2A^{pro} were used as templates in the second-round of PCR to generate the PCR product GFP2A^{pro}. The final product was digested with BamHI and Sac I and ligated into similarly restricted pGEM7zf⁺ to create pGFP2A^{pro}.

3.1.7. Construction of the control plasmid pGFP2A.

The control plasmid pGFP2A was required as a size marker control during PAGE analysis to identify any uncleaved GFP2A polyprotein. The initial strategy developed to construct pGFP2A was along the same lines as the method used to construct the plasmid pLbGFP. A series of oligonucleotide primers was designed to amplify two products GFP2A* and *2A introducing a single amino-acid mutation of the active site cysteine residue at position 109 of HRV14 2A^{PRO} to an alanine residue. The two first-round PCR products would overlap in the region of the mutation thus could anneal and be used as a template during the second-round of PCR to amplify a full-length GFP2A product.

Both first-round PCR products were produced successfully but problems arose during the second-round PCR reaction due to the large discrepancy between the sizes of the GFP2A*(1098bp) and *2A(129bp) products. Only a very small amount of full-length product was amplified during the second-round PCR reaction and attempts to improve the yield by using the full-length product as a template in a subsequent round of PCR failed. To avoid any loss of material the crude PCR products were used as templates for the second-round PCR. A PCR product corresponding to the full-length product (1209bp) was produced. The PCR fragment and vector were digested with BamHI and Sac I and ligated together. Competent JM109 *E.coli* cells were transformed with the ligation mix and plated onto selective media. Plasmid DNA from putative clones was prepared and submitted for automated DNA sequencing but the results indicated that the active site cysteine residue had not been mutated.

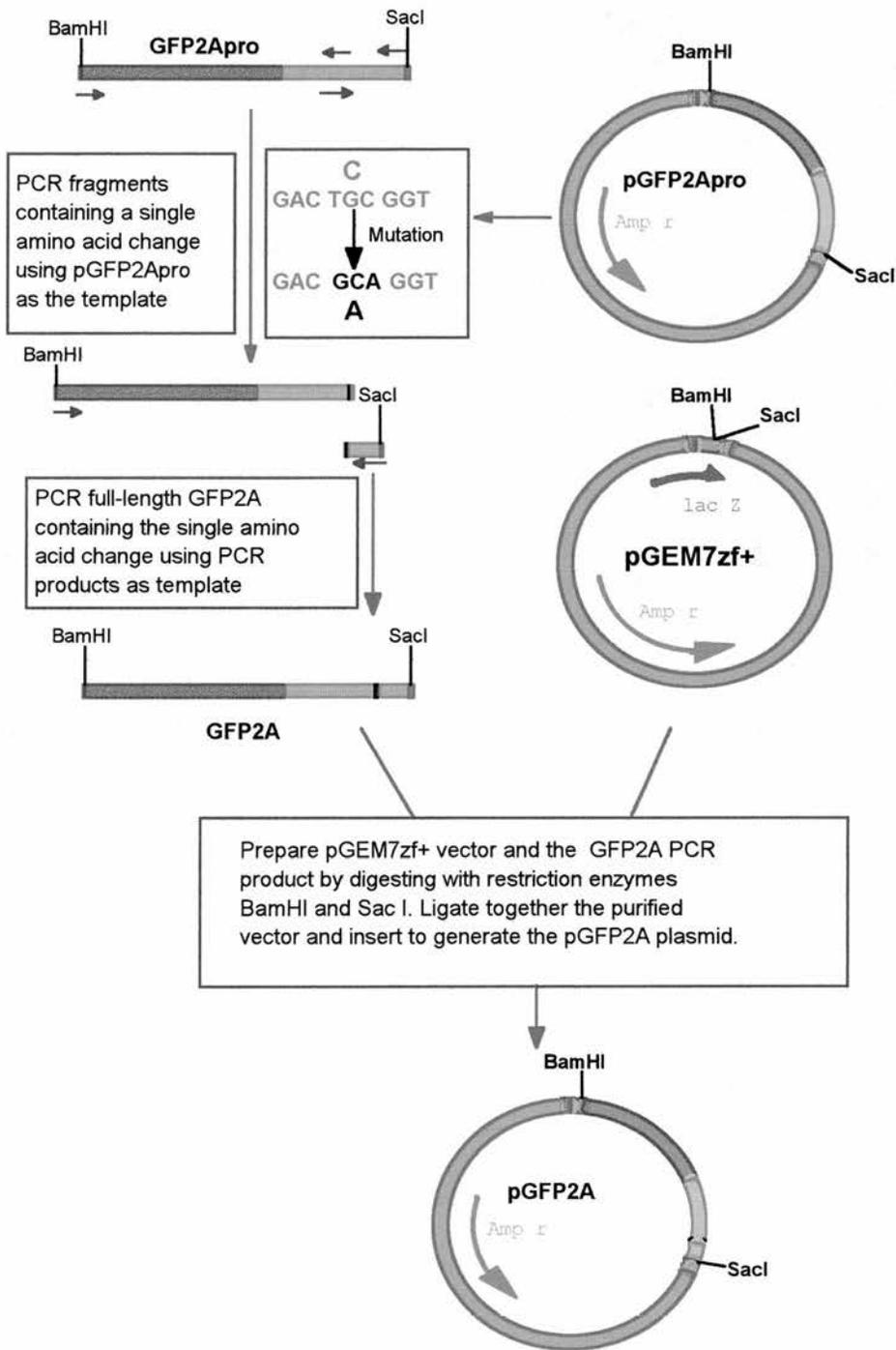


Fig.3.7. Initial strategy for construction of pGFP2A. The pGFP2A^{pro} was used as template for the overlap PCR designed to introduce a single point mutation of the active site Cys-109 to Ala. The (*) in GFP2A* and *2A indicates the mutation site.

A second cloning strategy was developed also using overlap PCR. Instead of creating a full-length wtGFP2A containing a single point mutation the aim of this method was to create an HRV14 2A product containing the mutation. An alternative first-round PCR product of 2A*(465bp) was amplified from the cDNA sequence of HRV14. During the second-round PCR reaction this 2A* PCR product and the *2A PCR product from the initial strategy were used as templates to create a mutated 2A product. The yield of the mutated HRV14 2A PCR product was much greater than had been achieved for the full-length wtGFP2A.

In order to create the desired pGFP2A plasmid the mutated HRV14 2A PCR product and the pGFP2A^{PRO} plasmid were digested with AccB7I and Sac I. AccB7I restricts HRV14 2A at a site upstream of the active site cysteine residue. Therefore the C-terminal region of HRV14 2A^{PRO} was removed from the plasmid pGFP2A^{PRO} and replaced with the C-terminal region of the 2A PCR product that contained the single-point mutation. Competent JM109 *E.coli* cells were transformed with the ligation reaction and plated onto antibiotic selective media. The plasmid DNA from putative clones was extracted and submitted for automated DNA sequencing. The sequencing results confirmed that the active site cysteine residue of HRV14 2A^{PRO} had been successfully mutated to an alanine residue.

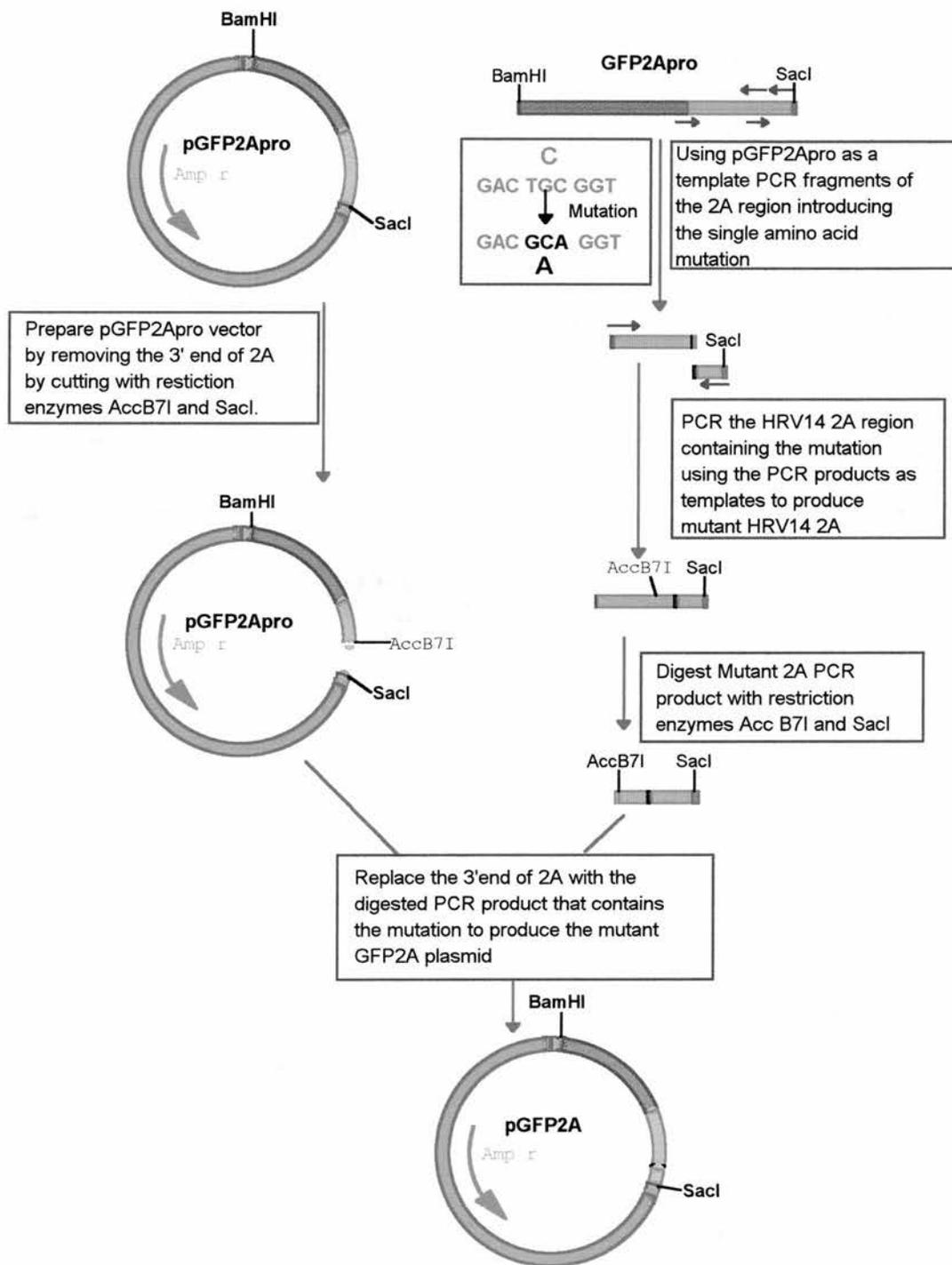


Fig.3.8. Alternative strategy for the construction of pGFP2A. 2A^{pro} with a single point mutation of Cys-109 to Ala was produced via overlap-PCR. The PCR product and pGFP2A^{pro} were digested with AccB7I and Sac I to replace the C-terminal of 2A^{pro} with the mutated 2A PCR fragment to form pGFP2A.

3.1.8. Translation of *pwtGFP* in vitro in rabbit reticulocyte lysate (RRL).

The control plasmid *pwtGFP* was used to program a coupled transcription and translation (TnT) system in RRL to verify that the control plasmid could be successfully translated to yield a product of the correct size. The computer program DNA Strider was used to calculate the predicted molecular weight of the *pwtGFP* translated product. The predicted molecular weight was 27.3kDa. The protein products of the TnT reaction were radiolabelled with ^{35}S -methionine. The products of the TnT reaction were separated by denaturing-PAGE and visualized by autoradiography. A single major band corresponding to the predicted molecular weight for wtGFP was produced.

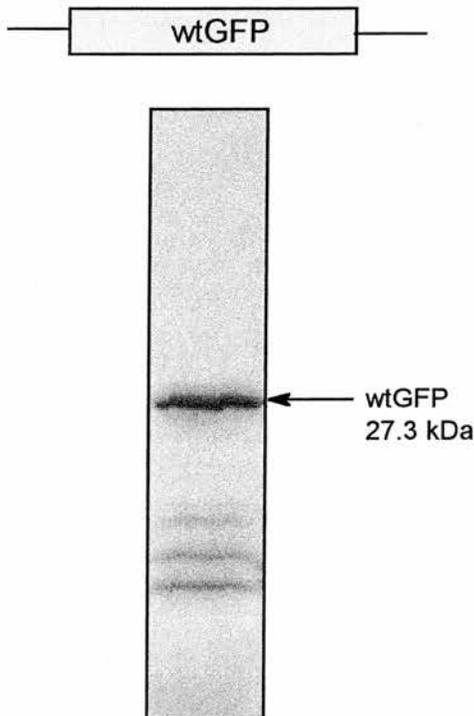


Fig. 3.9. In vitro transcription and translation in RRL of *pwtGFP*. Denaturing PAGE analysis of the coupled reaction programmed with *pwtGFP*. A single major product of the predicted size of wtGFP is present.

3.1.9. Translation of pLb^{PRO}GFP *in vitro* in rabbit reticulocyte lysate (RRL).

The computer program DNA Strider was used to predict the molecular weight of the uncleaved Lb^{PRO}GFP product and the sizes of the cleavage products.

Protein product	Predicted Mol.Wt (kDa)
Full-length Lb ^{PRO} GFP	47.5
Cleaved wtGFP	27.7
Cleaved Lb ^{PRO}	19.8

Table 3.1. The predicted molecular weight of full-length and cleaved Lb^{PRO}GFP.

The *in vitro* rabbit reticulocyte lysate TnT system was programmed with the pLb^{PRO}GFP plasmid. The proteins were radiolabelled with ³⁵S-methionine. The products of pLb^{PRO}GFP were separated by denaturing-PAGE and detected by autoradiography. A sample of the translation products of pwtGFP was run in an adjacent lane to provide a size marker for the GFP cleavage product.

Translation of pLb^{PRO}GFP yields three products corresponding to the uncleaved Lb^{PRO}GFP, wtGFP and Lb^{PRO} respectively. The production of the cleavage products confirms that *in vitro* in rabbit reticulocyte lysate the artificial reporter polyprotein Lb^{PRO}GFP is successfully processed *in cis* at the Lys-Gly cleavage site.

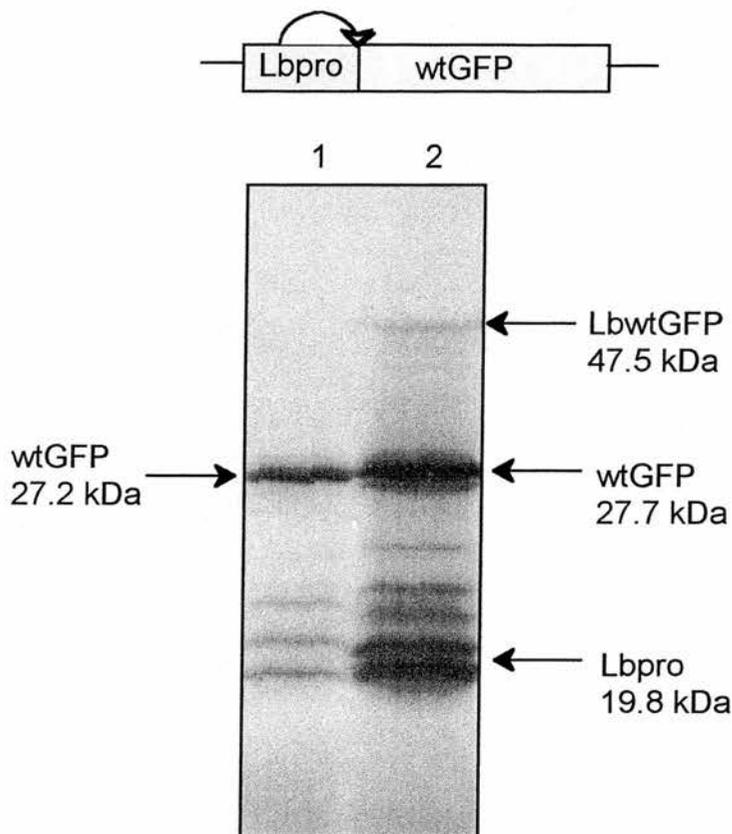


Fig.3.10. *In vitro* transcription and translation of pLb^{PRO}GFP in RRL. Denaturing-PAGE gel of coupled reactions. Lane 1- pwtGFP, Lane 2- pLb^{PRO}GFP. Lane-2 has three bands corresponding to uncleaved Lb^{PRO}GFP, cleaved GFP and cleaved Lb^{PRO}.

3.1.10. Translation of pLbGFP *in vitro* in rabbit reticulocyte lysate (RRL).

The plasmid pLbGFP was used to program the *in vitro* rabbit reticulocyte lysate TnT system. The protein products were radiolabelled with ³⁵S-methionine. The products of the translation were analyzed by denaturing-PAGE and visualized by autoradiography. To aid identification of the products the translation products of pwtGFP and pLb^{PRO}GFP were run in adjacent lanes on the gel.

Translation of pLbGFP yields a single band corresponding to the uncleaved full-length LbGFP product. Therefore the single amino-acid mutation of the active site cysteine residue to alanine has successfully abolished the proteolytic *cis* cleavage activity of FMDV Lb^{PRO}.

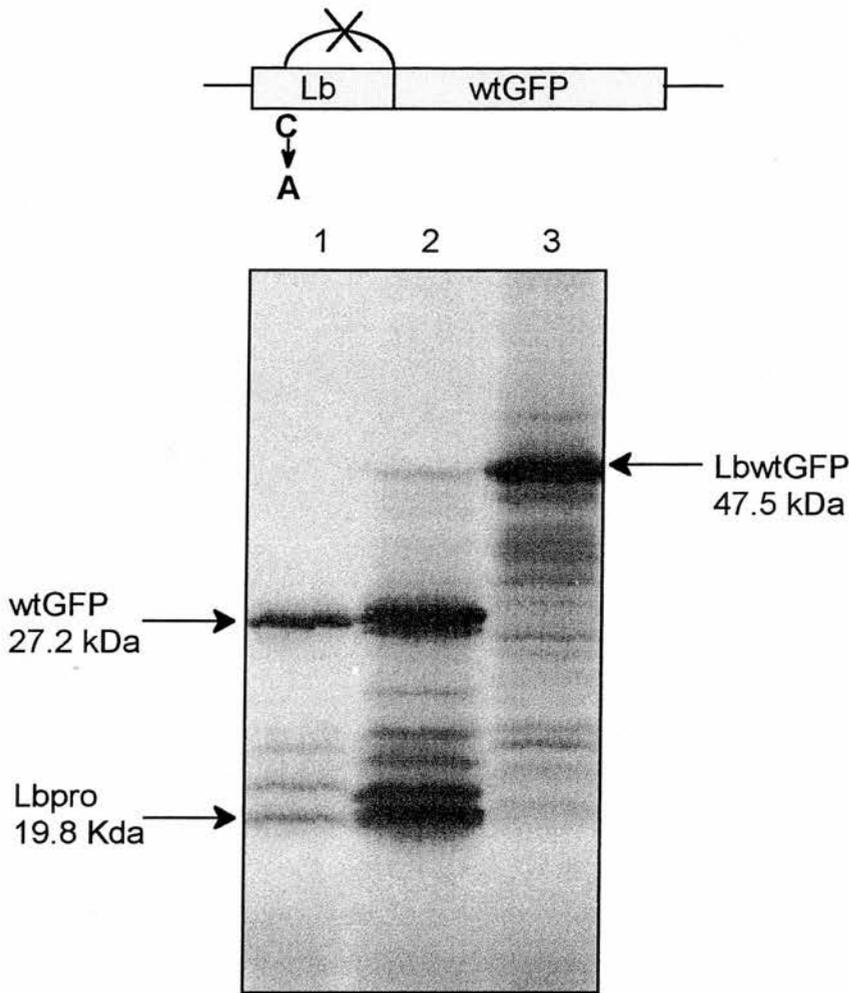


Fig.3.11. In vitro transcription and translation of pLbGFP in RRL. Autoradiograph of denaturing-PAGE of TnT samples, Lane 1- pwtGFP, Lane 2- pLb^{pro}GFP, Lane 3- pLbGFP.

3.1.11. Comparison of coupled TnT reactions in wheatgerm extract and rabbit reticulocyte lysate of FMDV Lb-containing plasmids.

In order to investigate the potential of using the artificial polyprotein system in plants the reporter polyprotein plasmid pLb^{pro}GFP and the control plasmids pwtGFP and pLbGFP were used to program an *in vitro* coupled TnT system using wheatgerm extract. The proteins were radiolabelled with ³⁵S-methionine. The translation products of each plasmid from both systems were analyzed by denaturing-PAGE and viewed using autoradiography.

All the plasmids were correctly translated in wheatgerm extract. The Lb^{PRO}GFP was correctly processed *in cis* to yield the cleavage products wtGFP and Lb^{PRO} in the wheatgerm extract system. Finally, the single point mutation at position 53 of Lb^{PRO} in the LbGFP plasmid efficiently abolished the proteolytic *cis* cleavage activity of Lb^{PRO} in both systems. In this translation of pLbGFP in rabbit reticulocyte lysate an internal initiation product was present (refer to section 3.1.13).

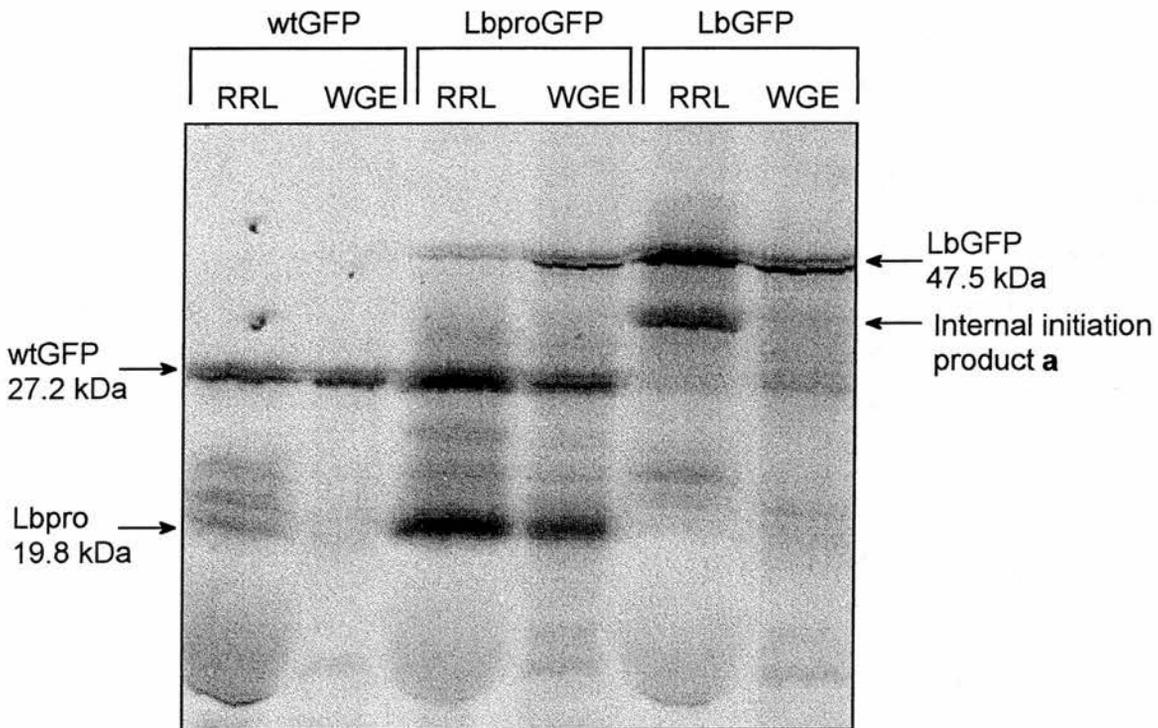


Fig.3.12. Comparison of TnT reactions in RRL and WGE systems. The Lb-containing plasmids and the control pwtGFP were transcribed and translated in both systems. The figure shows the autoradiograph of SDS-PAGE analysis of the reactions.

3.1.12. Translation of pGFP2A^{PRO} *in vitro* in rabbit reticulocyte lysate.

The computer program DNA Strider was used to predict the molecular weight of the expected products from the plasmid pGFP2A^{PRO}.

Product	Predicted Mol.Wt (kDa)
Full-length wtGFP2A ^{PRO}	43.98
Cleaved wtGFP	27.9
Cleaved 2A ^{PRO}	16.08

Table 3.2. The predicted Mol.Wt. of full-length and cleaved GFP2A^{PRO}.

The plasmid pGFP2A^{PRO} was used to program the *in vitro* TnT system in RRL radiolabelled with ³⁵S-methionine. The products were analyzed by denaturing-PAGE and viewed by autoradiography. A sample of the pwtGFP translation products was included on the gel in an adjacent lane as a size marker for the wtGFP cleavage product.

Transcription and translation of pGFP2A^{PRO} yields three main bands corresponding to full-length GFP2A^{PRO} and the cleavage products wtGFP and 2A^{PRO}. These results confirm that the 2A^{PRO} is active *in vitro* and can successfully cleave the GFP2A^{PRO} polyprotein *in cis* at the Tyr-Gly bond to yield GFP and 2A^{PRO}.

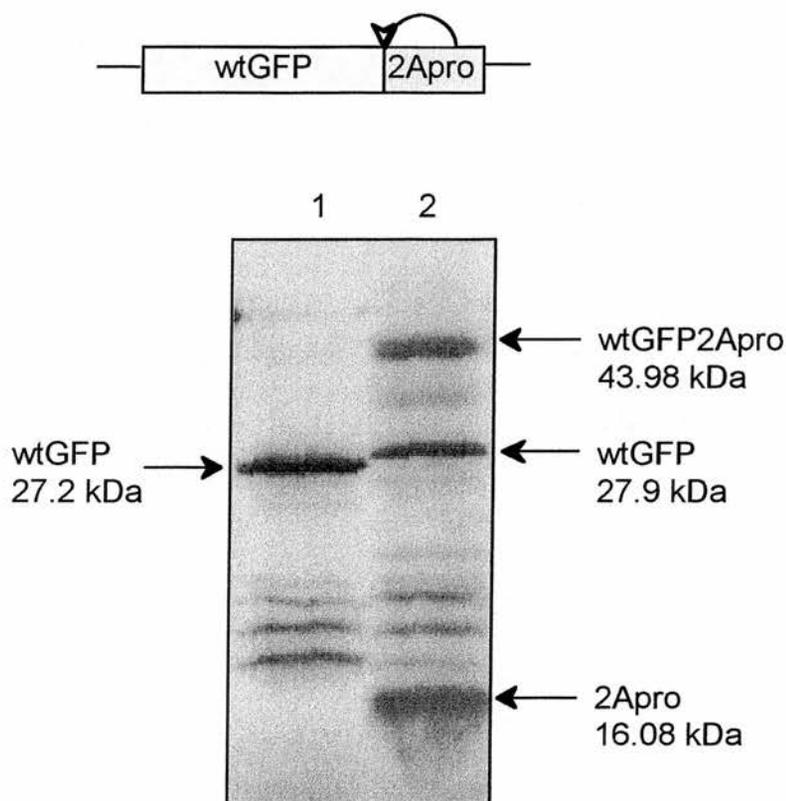


Fig.3.13. *In vitro* transcription and translation of pGFP2A^{pro} in RRL. A coupled TnT system was programmed with pwtGFP and pGFP2A^{pro}. The figure shows the autoradiograph of SDS-PAGE analysis of the reactions in Lane 1- pwtGFP and Lane 2 - pGFP2A^{pro}.

3.1.13. Translation of plasmid pGFP2A *in vitro* in rabbit reticulocyte lysate.

The rabbit reticulocyte lysate TnT system was programmed with the plasmid pGFP2A and radiolabelled with ³⁵S-methionine. The products were identified by denaturing PAGE analysis and visualized by autoradiography. Lanes containing the products of pwtGFP and pGFP2A^{pro} were included on the gel to aid in identification of the bands.

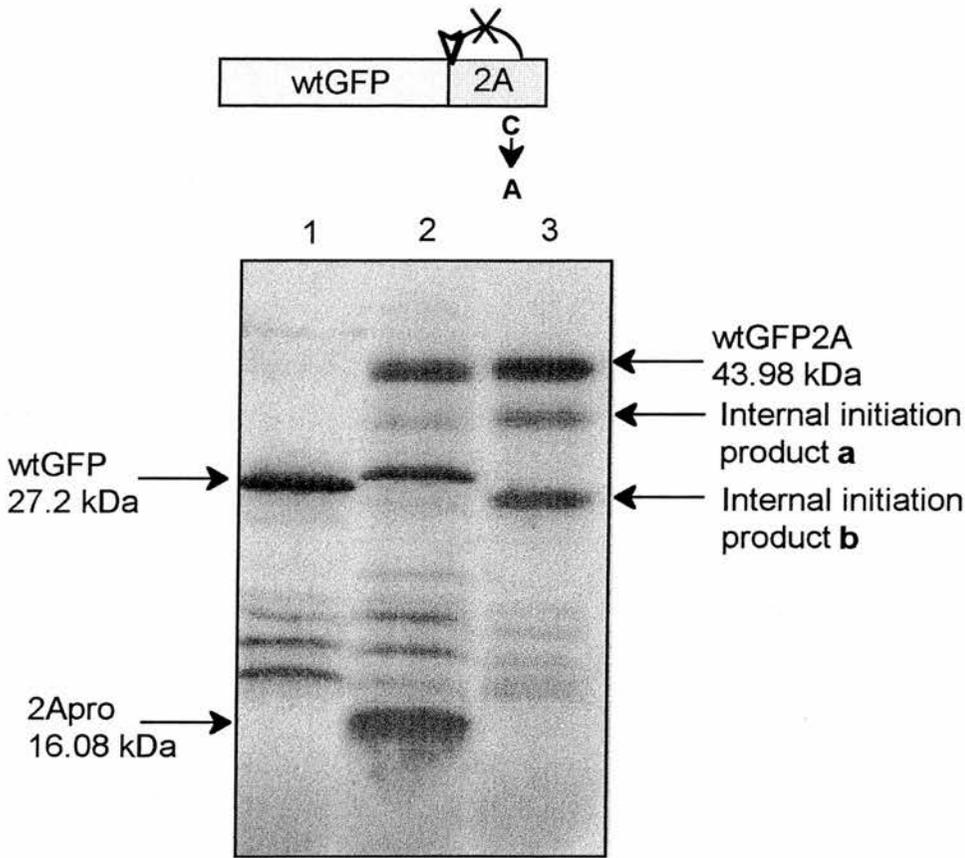


Fig.3.14. In vitro transcription and translation of pGFP2A. in RRL The autoradiograph of SDS-PAGE analysis of the samples, Lane 1- pwtGFP, Lane 2- pGFP2A^{Pro} and Lane 3- pGFP2A.

Translation of the plasmid pGFP2A did not yield bands corresponding to the cleaved products wtGFP (27.9kDa) and 2A (16.08kDa). Therefore the single point mutation of the active site cysteine residue to an alanine residue has abolished the proteolytic activity of the HRV14 2A protease. Translation of pGFP2A does yield three bands. The largest band was identified as the full-length GFP2A (43.98kDa). Two additional smaller products were also present labelled initiation product “a” and initiation product “b”. A small amount of the protein labelled initiation product “a” is present in some translations of pGFP2A^{Pro}. The third protein however, labelled initiation product “b”, is only seen in translations of the inactive polyprotein from pGFP2A. The three alternative explanations for the additional strong bands are that the proteins are:

(i) a result of protein degradation in the system – this is not very plausible as it would be expected that protein degradation by the ubiquitin pathway would not be so specific and would completely degrade the protein.

(ii) shorter products due to internal initiation of translation - during the initiation of translation the ribosome scans along the mRNA from the 5'end and must identify the initiation codon. The nucleotide sequence surrounding the initiation AUG codon influences the recognition of the AUG codon by the ribosomes, if the initiation codon is in a suboptimal context some ribosomes may fail to initiate translation at this position and continue scanning to initiate at a subsequent AUG codon. The optimum sequence context in animals was identified as CACCAUGG, the most important factor influencing initiation efficiency is the presence of a pyrimidine residue at the -3 position (Lütcke *et al.*, 1987; Joshi *et al.*, 1997).

(iii) shorter products due to premature termination of translation – translation may terminate if the levels of a specific amino-acid or tRNA are limiting.

The most plausible explanation for the extra bands present in the translation of pGFP2A is that they are internal initiation products. Examination of the nucleotide sequence identified several internal methionine residues. The surrounding nucleotide sequence and the predicted molecular weight are shown below in table 3.3. From the approximate size of the band on the gel and comparing the sequence context with the optimum sequence context reported by Kozak initiation product “a” is translated from Met 2 and initiation product “b” is translated from Met 4.

In-frame Met. residue	Predicted Mol. Wt. (kDa)	Sequence context
Met 1 (Authentic AUG)	43.98	TTCCATGA
Met 2	35.63	TCATATGA
Met 3	34.41	TGCCATGC
Met 4	26.76	CATCATGG
Met 5	19.45	CCACATGC

Table 3.3. The predicted molecular weight and sequence context of internal AUG codons within pGFP2A.

3.1.14. Comparison of translation in rabbit reticulocyte lysate and wheatgerm extract of GFP2A polyproteins.

The plasmids pGFP2A^{PRO} and pGFP2A were used to program an *in vitro* coupled TnT system using wheatgerm extract to monitor the viability of the artificial polyprotein system in plants. The translated products were radiolabelled with ³⁵S-methionine. The translated products of the plasmids in rabbit reticulocyte lysate and wheatgerm extract systems were analyzed by denaturing PAGE and viewed by autoradiography.

The translation products of the plasmids were all the predicted size confirming that the plasmids pGFP2A^{PRO} and pGFP2A are accurately translated in the wheatgerm extract system. The artificial polyprotein GFP2A^{PRO} was efficiently processed *in cis* by HRV14 2A^{PRO} in the wheatgerm extract system. Translation of the plasmid pGFP2A in wheatgerm extract demonstrates that mutation of the active site cysteine residue at position 109 effectively abolishes the proteolytic activity of HRV14 2A^{PRO} in both systems.

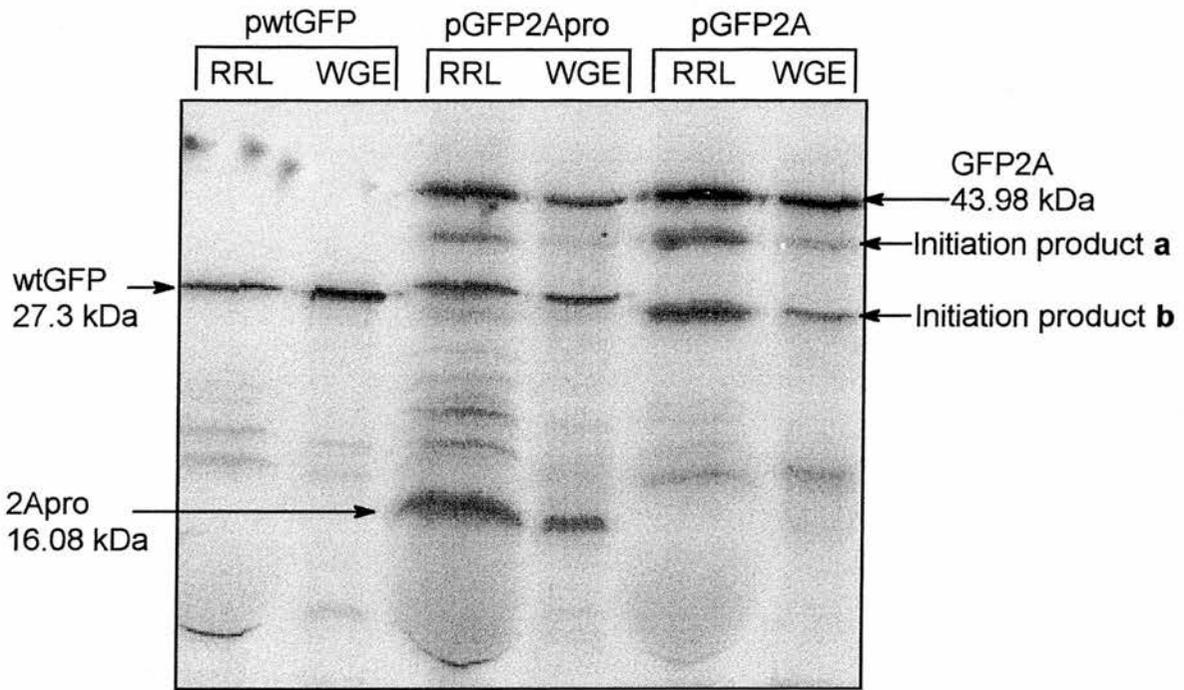


Fig.3.15. Comparison of *in vitro* transcription and translation in RRL and WGE. The HRV14 2A-containing plasmids and the control pwtGFP were used to program coupled TnT reactions in RRL and WGE. The products from both systems were analyzed by SDS-PAGE.

3.1.15. Comparison of cleavage efficiency of Lb^{PRO}GFP and GFP2A^{PRO}.

The specific cellular components of the rabbit reticulocyte lysate and wheatgerm extract TnT systems are by definition very different. The rabbit reticulocyte lysate is prepared from mammalian cells. In contrast, the source of wheatgerm extract is a higher, monocotyledon plant. The artificial polyprotein reporter plasmids pLb^{PRO}GFP and pGFP2A^{PRO} were expressed *in vitro* in both systems to monitor any effect on polyprotein processing due to the cellular makeup of the two systems. The products were radiolabelled with ³⁵S-methionine and analyzed by denaturing PAGE. The protein products were viewed by autoradiography.

The figure shows that the artificial polyproteins using the FMDV Lb^{PRO} and HRV14 2A^{PRO} are correctly processed *in cis* at the respective cleavage sites to produce the cleaved reporter protein wtGFP and the active protease.

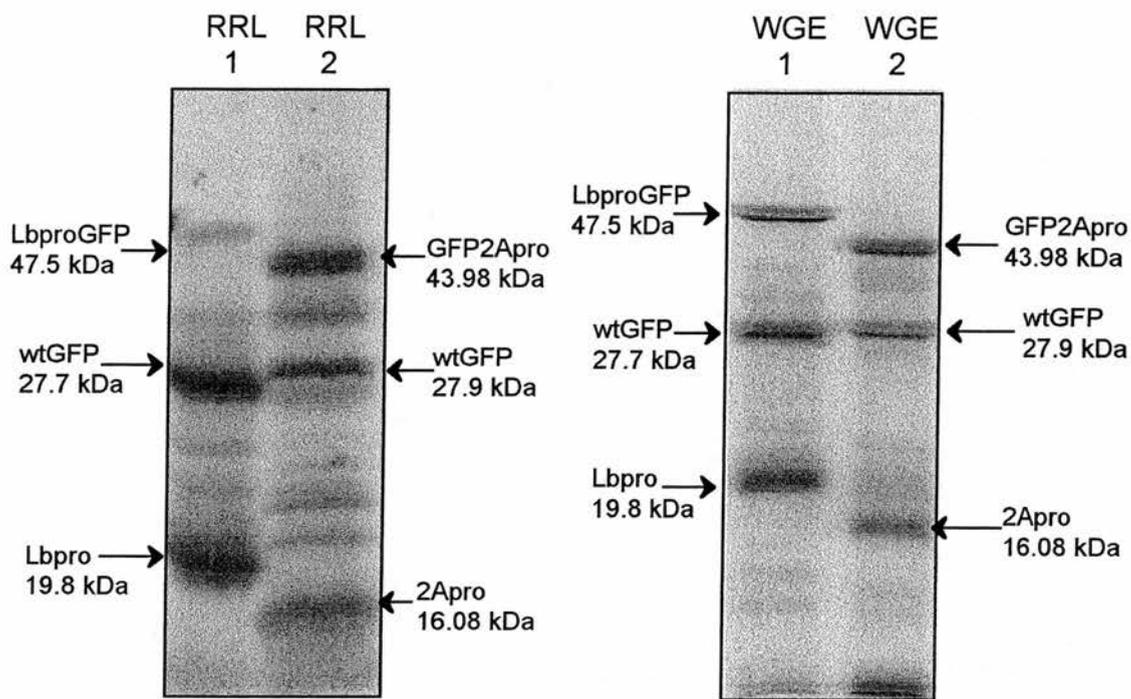


Fig.3.16. Comparison of cleavage efficiency of self-processing polyproteins in RRL and WGE. The plasmids $pLb^{pro}GFP$ and $pGFP2A^{pro}$ were used to program coupled TnT reactions in RRL and WGE. The RRL and WGE samples were analyzed on separate gels. **RRL gel** - Lane 1- $pGFP2A^{pro}$, Lane 2- $pLb^{pro}GFP$. **WGE gel** - Lane 1- $pLb^{pro}GFP$, Lane 2- $pGFP2A^{pro}$.

Densitometric analysis of the translation products of $pLb^{pro}GFP$ and $pGFP2A^{pro}$ was used to calculate the relative cleavage efficiency of the polyprotein in RRL and WGE. The denaturing-PAGE gel was visualized by phosphorimaging. The intensity of each protein product was quantified using MacBas version 2 software. The initial PSL values for each band were corrected for the background level of radioactivity to give a PSL-BG value. The TnT protein products are radioactively labelled with ^{35}S -methionine thus the band intensity is related to the number of methionine residues in each protein product. Therefore the PSL-BG value for each product is divided by the number of methionine residues to give an adjusted PSL-BG value.

Protein Product	No. Methionine Residues
Lb ^{pro} GFP	9
GFP2A ^{pro}	9
GFP	6
Lb ^{pro}	3
2A ^{pro}	3

Table 3.4. *The number of methionine residues in the polyproteins Lb^{pro}GFP and GFP2A^{pro} and their respective cleavage products.*

In order to monitor the relative cleavage efficiency of the polyprotein the proportion of the total radioactivity present in the uncleaved polyprotein and the combined cleavage products is calculated using the following method.

$$\% \text{ Radioactivity} = \frac{\text{Met. adjusted PSL-BG}}{\text{Total Radioactivity}} \times 100$$

	Adjusted PSL-BG (in RRL system)	Adjusted PSL-BG (in WGE system)
Lb ^{pro} GFP	3.33	23.64
GFP	92.87	40.57
Lb ^{pro}	199.27	81.9
TOTAL	295.47	146.11
GFP2A ^{pro}	63.89	18.9
GFP	107.97	20.58
2A ^{pro}	175.43	45.03
TOTAL	347.29	84.51

Table 3.5. *The quantitated intensity of the protein products detected by phosphorimaging and corrected for the number of methionine residues.*

	% radioactivity uncleaved product	in % radioactivity in cleavage products
Lb ^{PRO} GFP in RRL	1.1	98.9
Lb ^{PRO} GFP in WGE	16.2	83.8
GFP2A ^{PRO} in RRL	18.4	81.6
GFP2A ^{PRO} in WGE	22.4	77.6

Table 3.6. The calculated value of the % radioactivity present in uncleaved polyprotein and the combined cleavage products.

Therefore overall the cleavage efficiency of both self-processing polyproteins is higher in rabbit reticulocyte lysate than in the wheatgerm extract system. There is quite a large difference in the relative cleavage efficiency of Lb^{PRO}GFP and GFP2A^{PRO} in the rabbit reticulocyte lysate system, 98.9% of total Lb^{PRO}GFP is cleaved compared to 81.6% of total GFP2A^{PRO}. The difference is still apparent albeit dramatically reduced in wheatgerm extract with 83.8% of total Lb^{PRO}GFP cleaved compared to 77.6% of total GFP2A^{PRO} polyprotein.

3.1.16. Conclusions

A simple self-processing artificial polyprotein system and related control plasmids have been successfully constructed in pGEM transcription vectors. Analysis of the translation products *in vitro* in rabbit reticulocyte lysate and wheatgerm extract systems confirms that the correct products are translated.

The picornaviral proteases efficiently process the artificial polyproteins at the protease/wtGFP cleavage site *in vitro*. The cleavage reaction occurs in *cis*.

A single-point mutation of the active site cysteine residue to alanine in FMDV Lb^{PRO} and HRV14 2A^{PRO} completely abolishes the *cis* processing activity of the protein. An interesting observation was that this correlated with an increase in the internal

initiation products particularly for the translation of pGFP2A. The concentration and the number of internal initiation products both increased. The effect was apparent in both rabbit reticulocyte lysate and wheatgerm extract systems. From this it appears that mutation of the protease interferes with the scanning mechanism of the ribosome and identification of the initiation codon.

Although the proteases were functioning *in vitro* the artificial polyproteins were not completely processed in either *in vitro* system. Polyprotein processing was more efficient in the rabbit reticulocyte lysate system than in wheatgerm extract. The papain-like protease FMDV Lb^{PRO} was more efficient *in vitro* than the subtilisin-like protease HRV14 2A^{PRO}.

3.1.17. Summary

- An artificial self-processing polyprotein system has been constructed and analyzed *in vitro*.
- The FMDV Lb^{PRO} functions *in vitro* to process the artificial polyprotein Lb^{PRO}GFP. In RRL 99% of total polyprotein is processed.
- The FMDV Lb^{PRO} cleaves the artificial polyprotein *in cis* at the C-terminus of Lb^{PRO}.
- Mutation of the active site Cys 53 to Ala abolishes the proteolytic activity of FMDV Lb^{PRO}.
- The HRV14 2A^{PRO} functions *in vitro* to process the artificial polyprotein GFP2A^{PRO}. In RRL 82% of total polyprotein is processed.
- The HRV14 2A^{PRO} processes the artificial protein at the N-terminus of 2A^{PRO} *in cis*.
- Mutation of the active site Cys 109 to Ala destroys the proteolytic activity of HRV14 2A^{PRO}.
- The artificial polyproteins are efficiently processed *in vitro* in rabbit reticulocyte lysate and in wheatgerm extract.
- FMDV Lb^{PRO} processing is more efficient than HRV14 2A^{PRO} in both *in vitro* systems.
- The cleavage efficiency of both proteases is slightly higher in rabbit reticulocyte lysates.
- The presence of functional FMDV Lb^{PRO} and HRV14 2A^{PRO} improves the translation efficiency of both systems.

3.2. Part 2. Construction of an artificial polyprotein system using FMDV Lb^{Pro} and HRV14 2A^{Pro} to investigate *trans* processing *in vivo*.

3.2.1. Introduction.

The potential of the artificial polyprotein system as a method to introduce multiple genes is dependent on the ability to build in additional cleavage sites that would be processed by HRV14 2A^{Pro} or FMDV Lb^{Pro} in a bimolecular or *trans* reaction. The basic artificial polyprotein system was developed to add a second reporter gene and another cleavage site to the polyprotein. The reporter gene β -glucuronidase (GUS) was selected because it is widely used as a reporter gene *in planta* and antibodies were readily available in our laboratory. To ensure that an authentic cleavage site was maintained the built-in cleavage region was composed of a total of 14 amino-acids surrounding the scissile bond.

The *trans* processing function of the artificial polyproteins would be monitored *in vivo* in prokaryotes and in eukaryotes. The primary target for use of the artificial polyprotein system was in plants there was some evidence indicating a potential problem with using the reporter gene wtGFP in plants. The wtGFP sequence contains a region recognized by *Arabidopsis thaliana* as a cryptic plant intron that would result in bases 400-483 of the sequence being spliced out (Haselhoff and Amos, 1995). Therefore in case a similar problem also occurred in higher plants the decision was made to replace wtGFP with the alternative variant enhanced blue fluorescent protein (eBFP). The codons of eBFP have been altered to conform to the preferred mammalian codon usage this conversion has a convenient byproduct in that it removes the intron-coding sequence recognised by *Arabidopsis thaliana*.

The extended artificial polyproteins were initially constructed in the pGEM vector series rather than an expression vector system. This was primarily for cloning purposes as these vectors have large multiple cloning sites thus extending the available restriction sites and also plasmids based on this system containing the GUS gene were

available in the laboratory. Another aspect was that construction of the artificial polyproteins in a transcription vector allowed the polyproteins to be screened *in vitro* to verify that the complete polyprotein was produced and that the protease was active *in cis*.

3.2.2. Construction of the plasmid pLb^{PRO}BLG.

To develop a system in which to investigate the *trans* processing ability of Lb^{PRO} within an artificial polyprotein pLb^{PRO}BLG was constructed. The Lb^{PRO} cleaves the native polyprotein C-terminally thus the additional cleavage site and the second reporter gene GUS were added to the C-terminus of the artificial polyprotein. The additional cleavage site was based on the native L-1A cleavage site.

The first step to create Lb^{PRO}BLG was to replace wtGFP with BFP. The amino-acid sequence of the variants is very similar but due to the converted codon usage the nucleotide sequence is very different therefore it wasn't possible to simply substitute BFP for wtGFP in the plasmid pLb^{PRO}GFP. Thus the artificial polyprotein Lb^{PRO}BFP was constructed by the same overlap PCR strategy used to create the original pLb^{PRO}GFP. The BFP fragment was amplified by the PCR from the pEBFP vector supplied by Clontech. One important point was that the stop codon at the end of BFP was replaced in the reverse primer with a Xba I restriction site. The Lb^{PRO}BFP cassette from the second-round PCR was purified and directly cloned into the vector pGEM-T. The pGEM-T vector system exploits the fact that Taq polymerase adds a single adenosine nucleotide to the 5'end of the amplified fragment. The pGEM-T vector is linearized and has a single thymidine base added to each strand of the vector. Therefore the thymidine from the vector can anneal with the adenosine nucleotide added to the PCR fragment. The cloning is not directional thus the orientation of the insert was determined by automated DNA sequencing of the plasmid using the T7 and SP6 primers.

The second stage involved linking together the second reporter gene and the *trans* cleavage site. The starting point for this was the plasmid pGFP2AGUS created by M.Donnely. The FMDV 2A fragment was removed by digestion with Xba I and Apa I. Two large oligonucleotide adapters coding for the *trans* cleavage site and the required restriction enzyme sites were designed. The oligonucleotide adapters were annealed and the ends were restricted with Xba I and Apa I in preparation for ligation. The cut vector pGFP2AGUS and the annealed oligonucleotide adapters were ligated together to form the plasmid pGFPlinkerGUS.

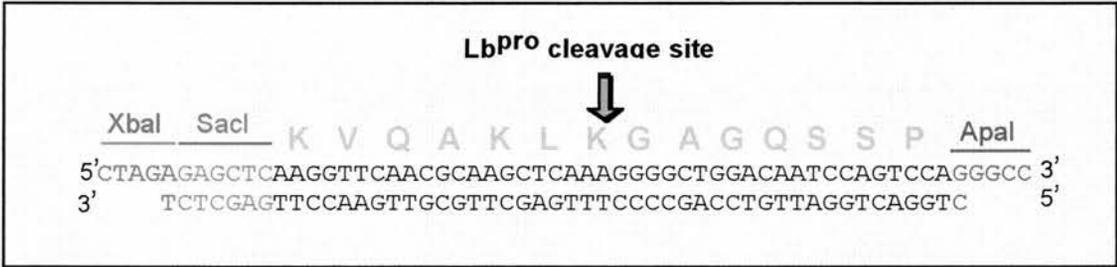


Fig.3.17. Lb^{PRO} linker region. Oligonucleotide adapters designed to introduce the linker region containing the FMDV L-1A cleavage site.

The final step was to link together the two fragments to form the complete artificial polyprotein. The linkerGUS fragment was removed from pGFPlinkerGUS by digestion with Xba I and Nsi I and purified from agarose. The pLb^{PRO}BFP vector was prepared for ligation by digestion with Xba I and Nsi I. The vector and linkerGUS insert were ligated together to create pLb^{PRO}BFP-linker-GUS (pLb^{PRO}BLG). Plasmid DNA from putative clones was analyzed by restriction digests and the insert was confirmed by automated DNA sequencing.

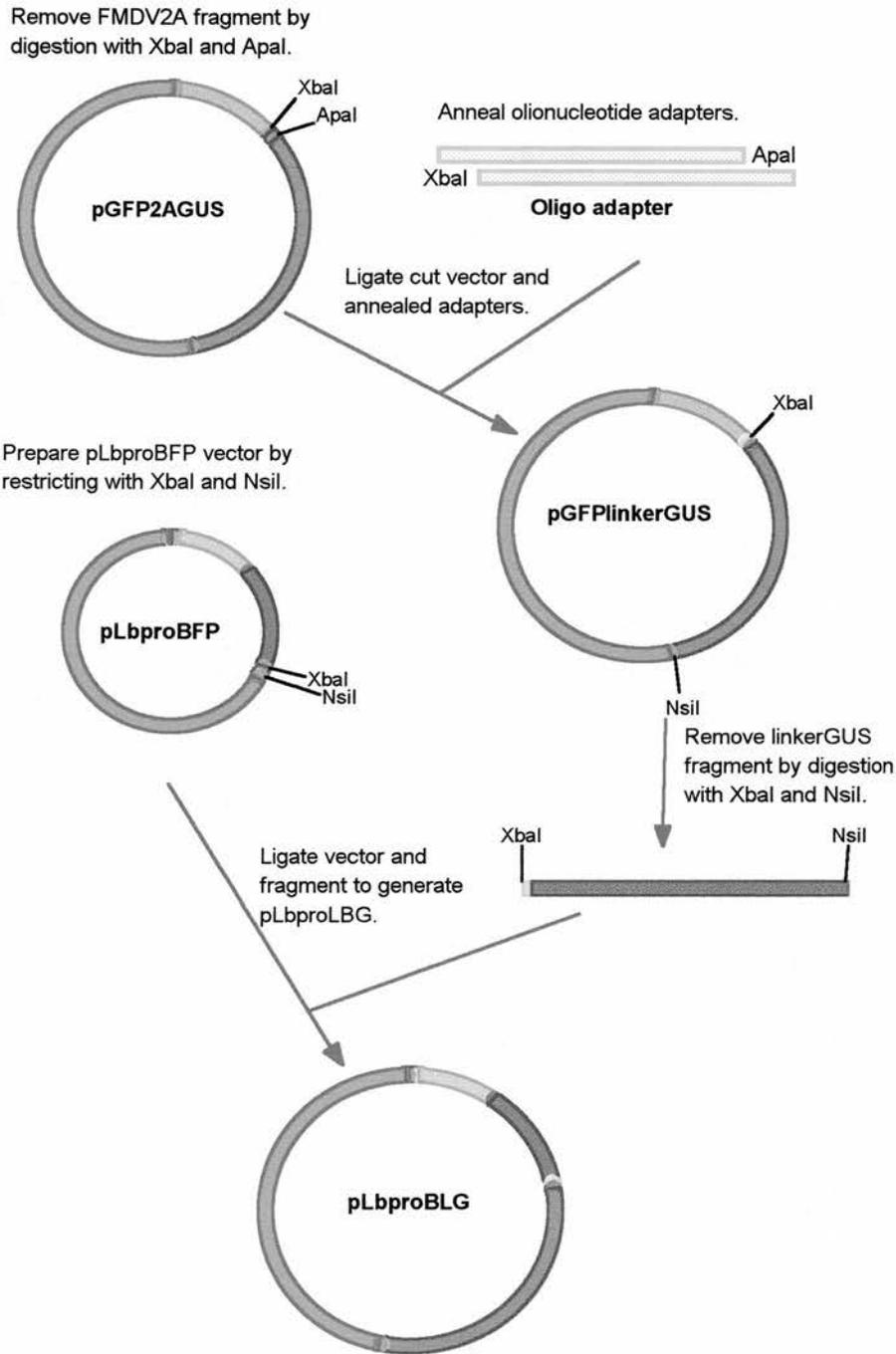


Fig.3.18. Construction of plasmid pLb^{pro}BLG. The plasmid was constructed in two steps. The first step was to produce GFPlinkerGUS from pGFP2AGUS by removing FMDV 2A and replacing it with the linker region oligo adapter. The linkerGUS fragment was excised by digestion with Xba I and Nsi I and ligated into the similarly digested pLb^{pro}BFP vector to produce pLb^{pro}BLG.

3.2.3. Translation of pLb^{PRO}BLG *in vitro* in rabbit reticulocyte lysate.

To confirm that the full-length polyprotein was translated from pLb^{PRO}BLG and that Lb^{PRO} was active *in cis* plasmid DNA of pLb^{PRO}BLG was used to program an *in vitro* TnT system in rabbit reticulocyte lysate. The *in vitro* rabbit reticulocyte lysate system was selected as opposed to the wheatgerm extract system because large proteins (>60kDa) are not translated in wheatgerm extract very efficiently. A number of control plasmids were also expressed in the same system to help identify the various bands. The controls included pLb^{PRO}BFP that had a stop codon after BFP to provide a control for the Lb^{PRO} band and BFP. A range of plasmids created by Michelle Donnelly provided controls for bands containing GUS. The translation products were analyzed by denaturing-PAGE and viewed by autoradiography. The molecular weights of the products of the *cis* and *trans* cleavage events of the artificial polyprotein Lb^{PRO}BLG were predicted using the computer program DNA Strider.

Predicted Product (processing events)	Molecular Weight (kDa)
Lb ^{PRO} BLG (full-length)	116.48
BFPlinkerGUS (<i>cis</i> cleavage)	96.69
GUS (<i>trans/cis</i> and <i>trans</i> cleavage)	67.97
Lb ^{PRO} BFP (<i>trans</i> cleavage only)	48.53
BFP (<i>cis</i> and <i>trans</i> cleavage)	28.74
Lb ^{PRO} (<i>cis/cis</i> and <i>trans</i> cleavage)	19.81

Table 3.7. The predicted molecular weight of full-length Lb^{PRO}BLG polyprotein and all possible cleavage products.

The product from pLb^{PRO}BFP was efficiently processed by Lb^{PRO} *in cis* to yield the cleavage products Lb^{PRO} and BFP. Comparison of lanes containing pLb^{PRO}BFP and pLb^{PRO}BLG confirms that Lb^{PRO} is active *in cis* in the larger artificial polyprotein and has cleaved the Lb^{PRO}/BFP junction to yield the cleavage products Lb^{PRO} and BFPlinkerGUS. There were no bands present corresponding to the predicted sizes of the *trans* cleavage products BFP and GUS.

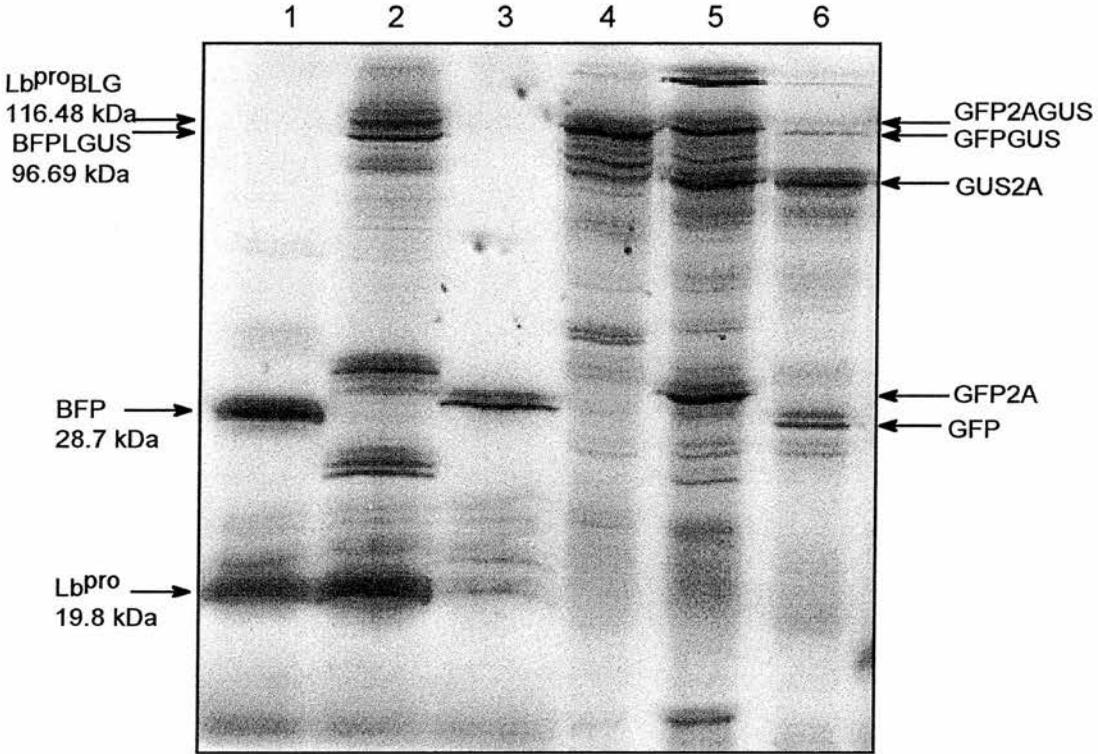


Fig.3.19. *In vitro* transcription and translation of pLb^{PRO}BLG in RRL. The autoradiograph of SDS-PAGE analysis of Lane 1- pLb^{PRO}BFP, Lane 2- pLb^{PRO}BLG, Lane 3- peBFP, Lane 4- pGFPGUS, Lane 5- pGFP2AGUS, Lane 6- pGUS2AGFP.

3.2.4. Construction of plasmid pGLB2A^{PRO}.

To create an artificial polyprotein system in which to study the *trans* processing of HRV14 2A^{PRO} the plasmid pGLB2A^{PRO} was constructed using a similar strategy to that employed to create pLb^{PRO}BLG. Unlike the Lb^{PRO}BLG polyprotein that duplicated the native FMDV L/1A cleavage site, the second cleavage site in pGLB2A^{PRO} was based on the 2A^{PRO} cleavage site of rabbit eIF4G, a cellular target of 2A^{PRO}.

The first stage was to create pBFP2A^{PRO} using the overlap PCR strategy developed for construction of pGFP2A^{PRO}. The BFP product was amplified using the vector pEBFP (Clontech) as a template. The second-round PCR product was purified and directly ligated into the pGEM-T vector. The plasmid DNA from putative clones was prepared and analyzed by restriction digests. The orientation of the insert was identified by automated DNA sequencing using T7 and SP6 primers.

The second stage to create the fragment containing the *trans* cleavage site and the GUS reporter gene used the plasmid pGUS2AGFP produced by M. Donnelly. A set of oligonucleotide adapters that coded for the cleavage site and additional restriction enzyme sites were designed. The oligonucleotides were annealed and restricted in preparation for ligation with the restriction enzymes Xba I and Nsi I. The plasmid pGUS2AGFP was also prepared for ligation by digestion with Xba I and Nsi I thus removing the -2AGFP- fragment. The cut vector containing GUS was ligated with the annealed oligonucleotide adapter to produce pGUSlinker. To verify that GUS and the cleavage region had been correctly linked together the plasmid DNA from putative clones was sequenced.

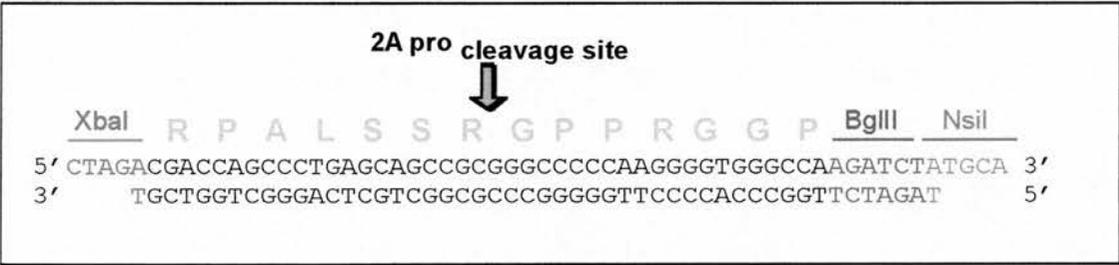


Fig.3.20. HRV14 2A^{PRO} linker region. Oligonucleotide adapters designed to encode the linker region containing the 2A^{PRO} cleavage site from rabbit eIF4G.

Finally, to create a plasmid encoding the full-length polyprotein GUSlinkerBFP2A^{PRO} the GUSlinker fragment was removed from pGUSlinker by digestion with Sfi I and Bgl II. The pBFP2A^{PRO} vector was digested with Sfi I and BamHI. The purified cut vector and GUSlinker fragment were ligated together. The Sfi I site used was not in

the multiple cloning site of the pGEM vectors but slightly upstream in the T7 promoter region. The restriction sites BamHI and Bgl II have compatible ends therefore the vector-BamHI and insert-Bgl II could anneal but a restriction site was not reformed. The plasmid DNA from putative clones was extracted and analyzed by restriction digests. The insert was sequenced by automated DNA sequencing using a series of primers including T7, SP6, BFP-C and BFP-N that anneal to the N- and C-termini of eBFP.

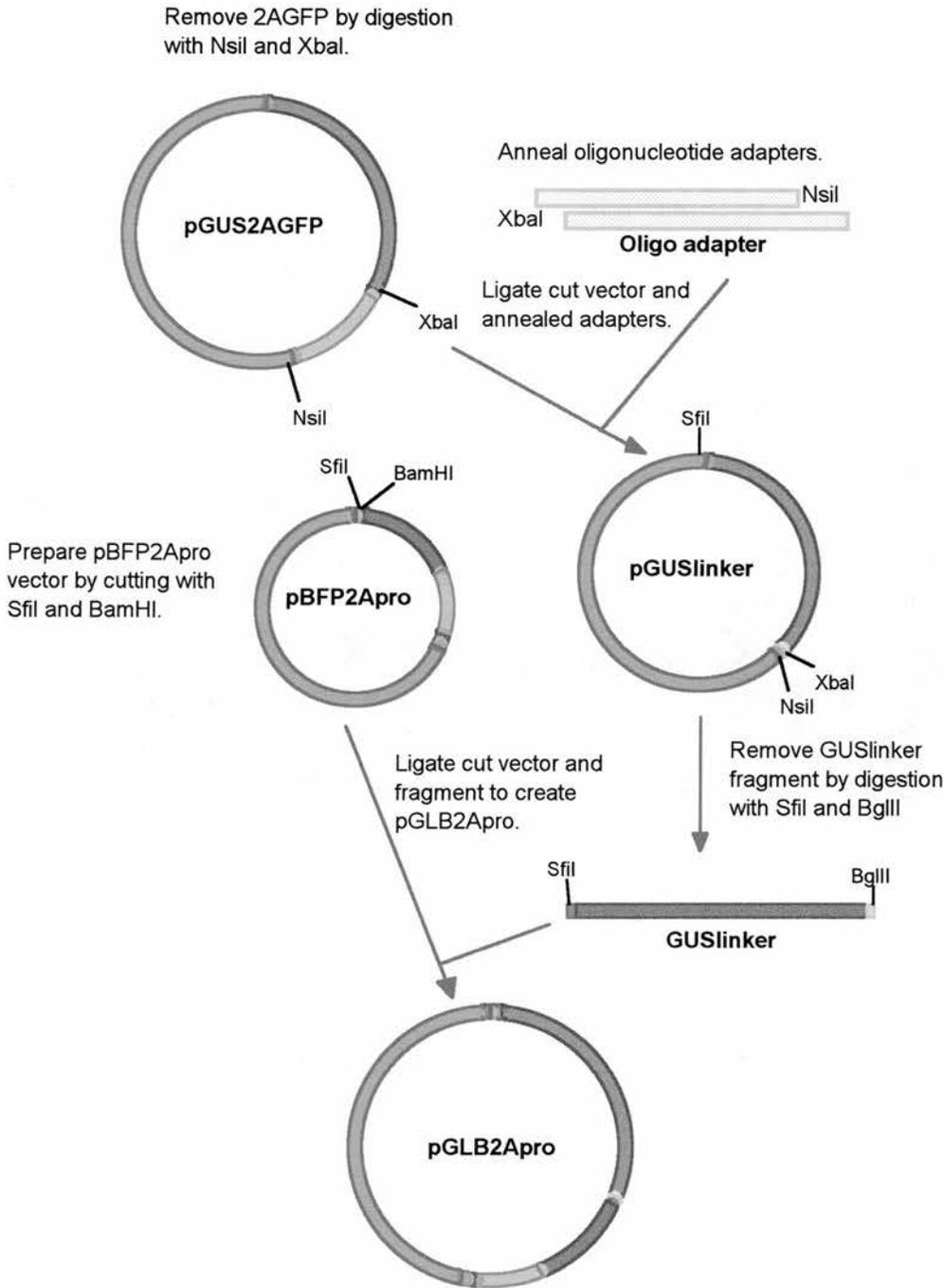


Fig. 3.21. Construction of the plasmid pGLB2Apro. pGLB2Apro was constructed in two steps. Step 1- 2AGFP of pGUS2AGFP was excised and replaced with the annealed oligonucleotide adapter of the rabbit eIF4G 2Apro cleavage site to form pGUSlinker. Step 2- GUSlinker was excised by digestion with Sfi I and Bgl II. The plasmid pBFP2Apro was prepared by digestion with Sfi I and BamHI. The cut pBFP2Apro and the GUSlinker fragment were ligated forming pGLB2Apro.

3.2.5. Expression of plasmid pGLB2A^{PRO} in vitro in rabbit reticulocyte lysate.

To confirm that the plasmid sequence of pGLB2A^{PRO} did not encode any premature stop codons and that the translated polyprotein product was the predicted size a sample of pGLB2A^{PRO} DNA was used to program an *in vitro* TnT system in rabbit reticulocyte lysate. The molecular weights of the translation products of pGLB2A^{PRO} were predicted using the computer program DNA Strider.

Protein product (cleavage event)	Molecular Weight (kDa)
GLB2A ^{PRO} (full-length)	113.68
GUSlinkerBFP (<i>cis</i> cleavage)	97.62
GUS (<i>trans/cis</i> and <i>trans</i> cleavage)	69.23
BFP2A ^{PRO} (<i>trans</i> cleavage)	44.47
BFP (<i>cis</i> and <i>trans</i> cleavage)	28.41
2A ^{PRO} (<i>cis/cis</i> and <i>trans</i> cleavage)	16.08

Table 3.8. The predicted molecular weight of full-length polyprotein GLB2A^{PRO} and all possible cleavage products.

The TnT system was also programmed with several control plasmids pBFP2A^{PRO}, pGFPGUS, pGFP2AGUS and pGUS2AGFP. The translation products were radiolabelled with ³⁵S-methionine. An aliquot of the translation products from each reaction was analyzed by denaturing-PAGE and the radioactive proteins were visualized by autoradiography.

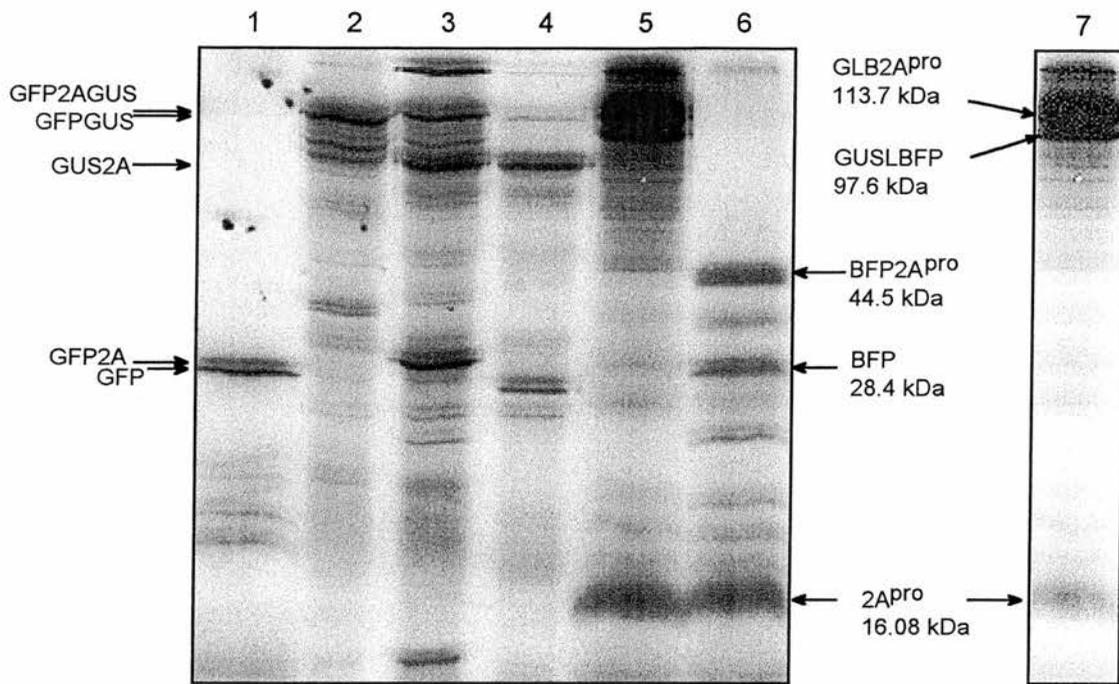


Fig.3.22. Transcription and translation *in vitro* of pGLB2AP^{ro} in RRL. The autoradiograph of SDS-PAGE analysis of TnT reactions in RRL. Lane 1-pwtGFP, Lane 2-pGFPGUS, Lane 3- pGFP2AGUS, Lane 4-pGUS2AGFP, Lane 5- pGLB2AP^{ro}, Lane 6- pBFP2AP^{ro}. Lane 7- Duplicate lane 5 with altered contrast to improve visualization of larger bands.

Translation of pGLB2AP^{ro} confirmed that the complete polyprotein was translated *in vitro*. A band correlating to 2AP^{ro} was identified for the plasmids pBFP2AP^{ro} and pGLB2AP^{ro} verifying that HRV14 2AP^{ro} is active *in cis in vitro*. There were no bands correlating to the predicted sizes of GUS and BFP therefore the artificial polyprotein was not processed *in trans* by HRV14 2AP^{ro}.

3.2.6. Conclusions

Two artificial self-processing polyprotein systems using the proteolytic function of picornaviral proteases FMDV Lb^{Pro} and HRV14 2AP^{ro} have been constructed in a transcription vector. The polyprotein construct can be moved to appropriate vectors to allow expression of the polyprotein in different *in vivo* systems where the expression and subsequent processing of the polyprotein can be monitored.

Transcription and translation of the artificial polyproteins was conducted *in vitro* using a coupled system in rabbit reticulocyte lysate that can translate large proteins. Translation confirmed that the full-length polyprotein was produced from the coding sequence. The polyproteins were efficiently processed *in cis* by the encoded picornaviral protease confirming that the proteases are functional.

The *in vitro* translated polyproteins were not processed at the second cleavage site. The primary cleavage of the artificial polyprotein occurs at the BFP/protease junction thus processing of the second cleavage site in the linker region must be a bimolecular or *trans* cleavage event. Previous work within our laboratory suggests that the coupled *in vitro* TnT system is not a suitable system to monitor *trans* processing events (E.Byrne pers.comm.). A *trans* cleavage event is bimolecular and follows first-order kinetics therefore the concentrations of substrate and protease are rate-limiting. The level of protease produced in the TnT system is too low for *trans* processing.

3.2.7. Summary

- A self-processing polyprotein system to introduce two reporter genes has been produced using the cleavage activity FMDV Lb^{PRO}.
- Translation of pLb^{PRO}BLG *in vitro* in RRL yields full-length polyprotein.
- The Lb^{PRO} in pLb^{PRO}BLG is active and cleaves the Lb^{PRO}/BFP junction *in cis*.
- *Trans* processing activity of Lb^{PRO} is not evident in the *in vitro* system.
- A self-processing polyprotein system using the cleavage specificity of HRV14 2A^{PRO} to introduce two genes has been produced.
- Translation of the pGLB2A^{PRO} *in vitro* in RRL yields full-length polyprotein.
- The HRV14 2A^{PRO} in pGLB2A^{PRO} is active *in cis* and cleaves the polyprotein at the 2A^{PRO}/BFP junction.
- The *trans* processing activity of HRV14 2A^{PRO} is not evident in the *in vitro* system.

CHAPTER 4: INVESTIGATION INTO THE EFFECT OF INCLUDING AN INTERNAL RIBOSOME ENTRY SITE ON THE *IN VITRO* EXPRESSION OF THE ARTIFICIAL POLYPROTEIN CASSETTES.

4.1. Part 1: *In vitro* analysis of translation in rabbit reticulocyte lysate and wheatgerm extract using a coupled TnT system.

4.1.1. Introduction.

The positive strand picornavirus viral RNA differs from cellular mRNA in that the 7-methyl guanosine cap structure at the 5' terminus is replaced by a viral protein VPg that is linked to the 5' end of the RNA. The cellular mRNA is recruited to the ribosome by the eIF4F initiation complex (refer to section 1.8). The eIF4E component binds to the 5' cap of the cellular mRNA, then subsequently binds to the scaffolding protein eIF4G. The 40S ribosome preinitiation complex is bound by eIF3 which also binds to eIF4G thus the cellular mRNA is recruited to the ribosome in preparation for translation.

Two major features preclude picornaviruses using the cellular initiation machinery; the most obvious is simply the lack of a 5' cap structure on the viral RNA. The second reason is due to the action of the FMDV L protease and the rhino- and enterovirus 2A protease that cleave the scaffolding protein eIF4G separating the binding sites for eIF4E and eIF3. Infected cells, therefore, are unable to recruit capped messenger RNA to the ribosome resulting in the host-cell shutoff phenomenon. Picornaviruses circumvent the requirement for a 5' cap structure to recruit the cellular ribosomal machinery via the eIF4F complex through the ability of a tract of RNA within the 5' untranslated region of the genome to recruit the ribosomal machinery directly thus allowing translation of viral RNA. The stretch of RNA involved is highly structured and has been designated as the internal ribosome entry site or IRES.

The aim was to develop the basic artificial reporter polyproteins by including the picornavirus IRES regions upstream of the polyprotein cassette to improve the expression of the artificial reporter polyproteins within the *in vitro* translation system. The IRES-containing reporter systems would be able to recruit the ribosomal machinery thus gaining an advantage over the uncapped reporter system constructs. The inclusion of an IRES within the reporter system would be required if the reporter system was used within mammalian cells due to the toxicity of the FMDV Lb protease and HRV14 2A protease.

The IRES regions from FMDV and HRV14 were selected for use, as the proteases used within the artificial self-processing reporter system were cloned from these viruses. The picornavirus IRESes can be classed into three groups; the HRV14 IRES (type I IRES) and the FMDV IRES (type II IRES) thus we can compare the function of different types of IRES. The basic artificial self-processing polyprotein reporter system was modified by the addition of a picornavirus IRES at the 5' terminus of the polyprotein cassette.

4.1.2. Construction of the IRES plasmids pHI and pFI.

The IRES plasmids were produced initially as a source of the FMDV and HRV14 IRESes. The IRES cassettes could be subsequently subcloned into the various control and artificial polyprotein plasmids.

A set of oligonucleotide primers was designed to anneal to the boundaries of the regions which confer IRES function. Novel restriction enzyme sites were introduced at both ends of the IRES region for cloning purposes, a Bbu I site was added to the 5' terminus and a Xho I was added to the 3' terminus. The HRV14 IRES was amplified by the PCR from the full-length HRV14 cDNA coding sequence. The FMDV IRES was amplified by the PCR from the plasmid pMR111 encoding the FMDV IRES -P12A-NEO- previously constructed by Martin Ryan. The amplified IRES fragments were purified and restricted with Bbu I and Xho I. The pGEM7zf+ vector was similarly restricted with Bbu I and Xho

I. The restricted IRES fragments and cut vector were purified from agarose and ligated together to create the plasmids pHI and pFI. Competent JM109 *E.coli* cells were prepared and transformed with the ligation reactions. The transformed cells were plated onto antibiotic selective media and screened for α -complementation. The plasmid DNA was extracted from the putative clones. The presence of the IRES regions was determined by a series of analytical restriction digests. Finally, the positive clones were sequenced by automated sequencing to monitor the integrity of the sequence identifying any mutations arising from the amplification process.

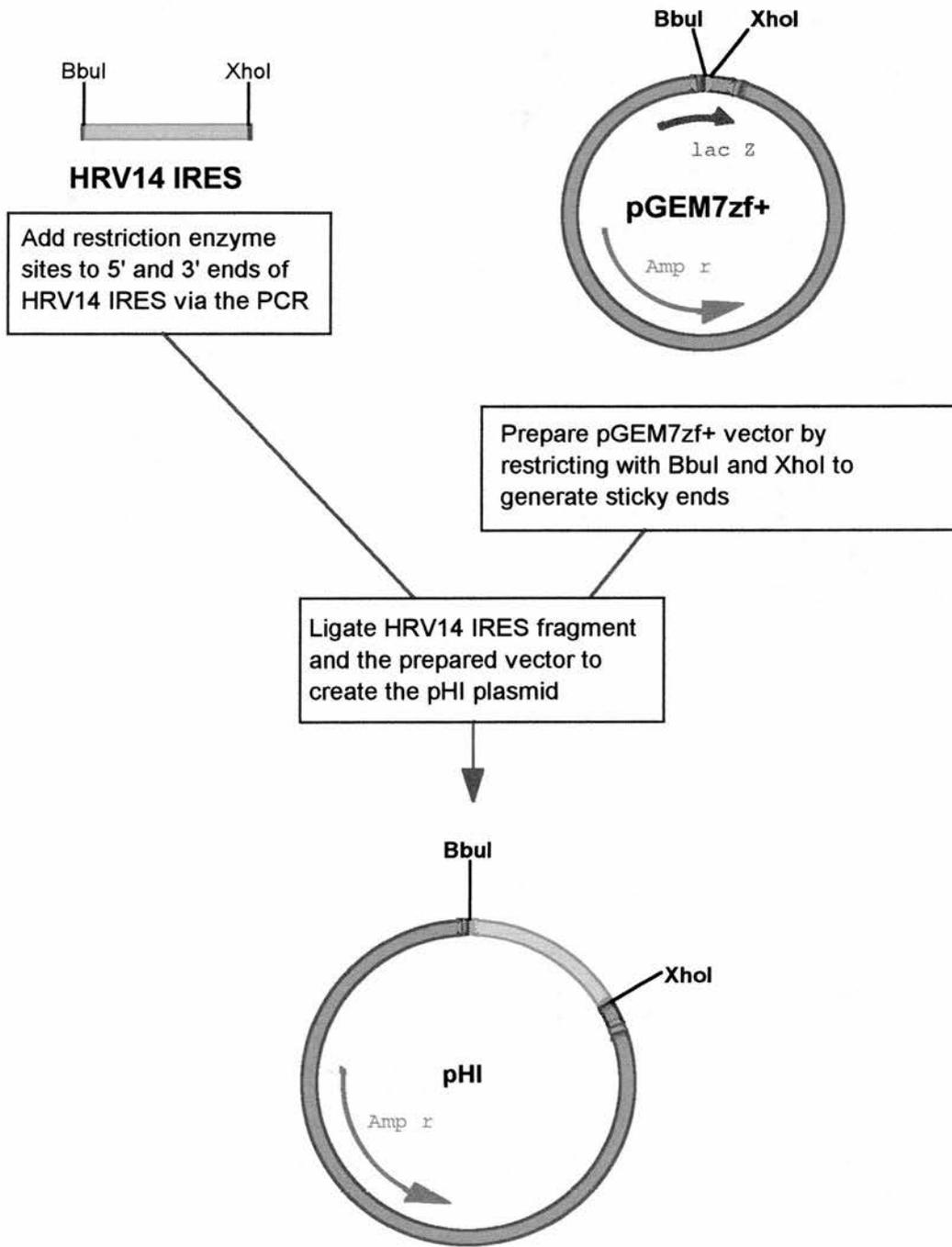


Fig.4.1. Construction of the pHI plasmid.

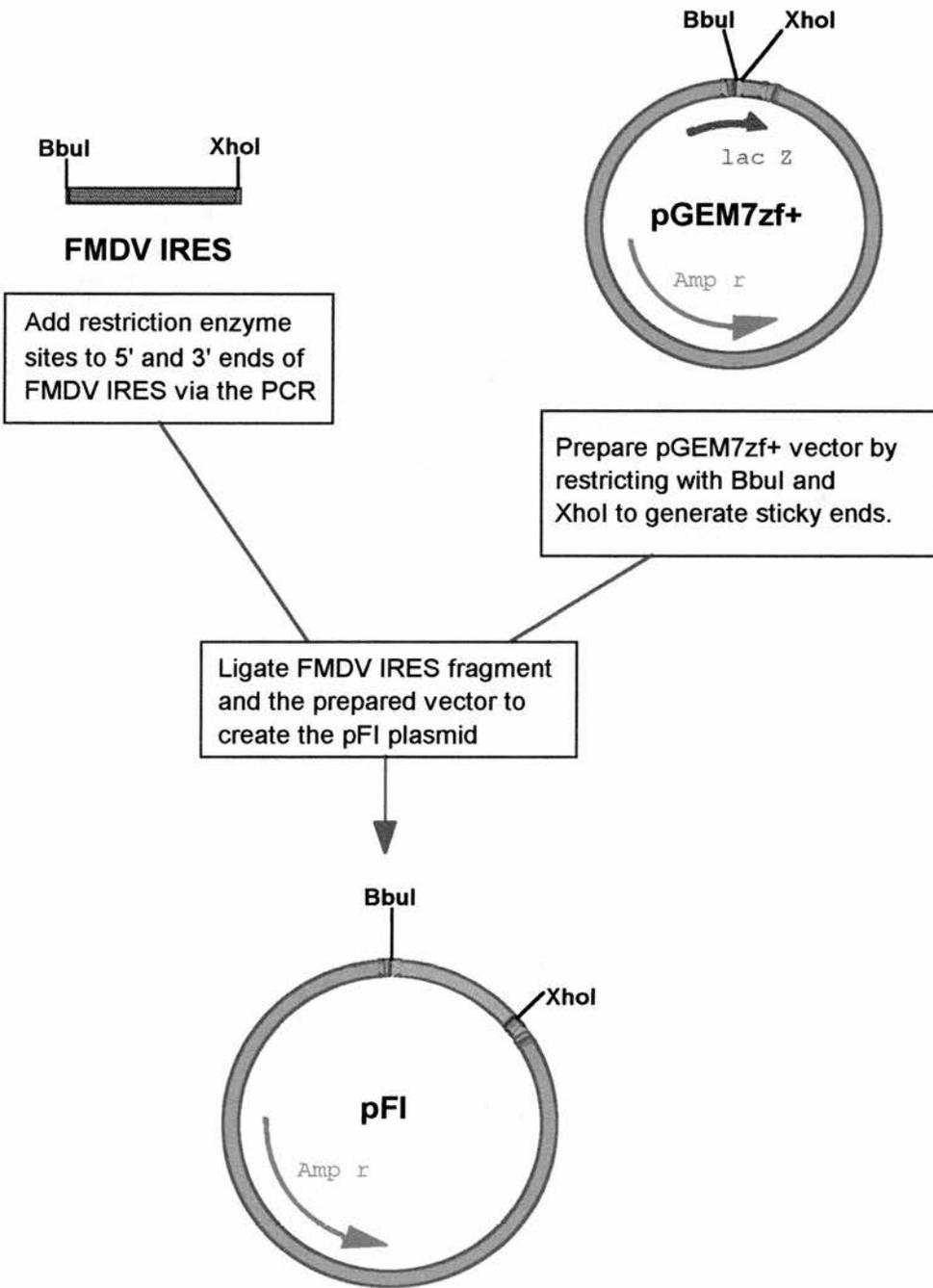


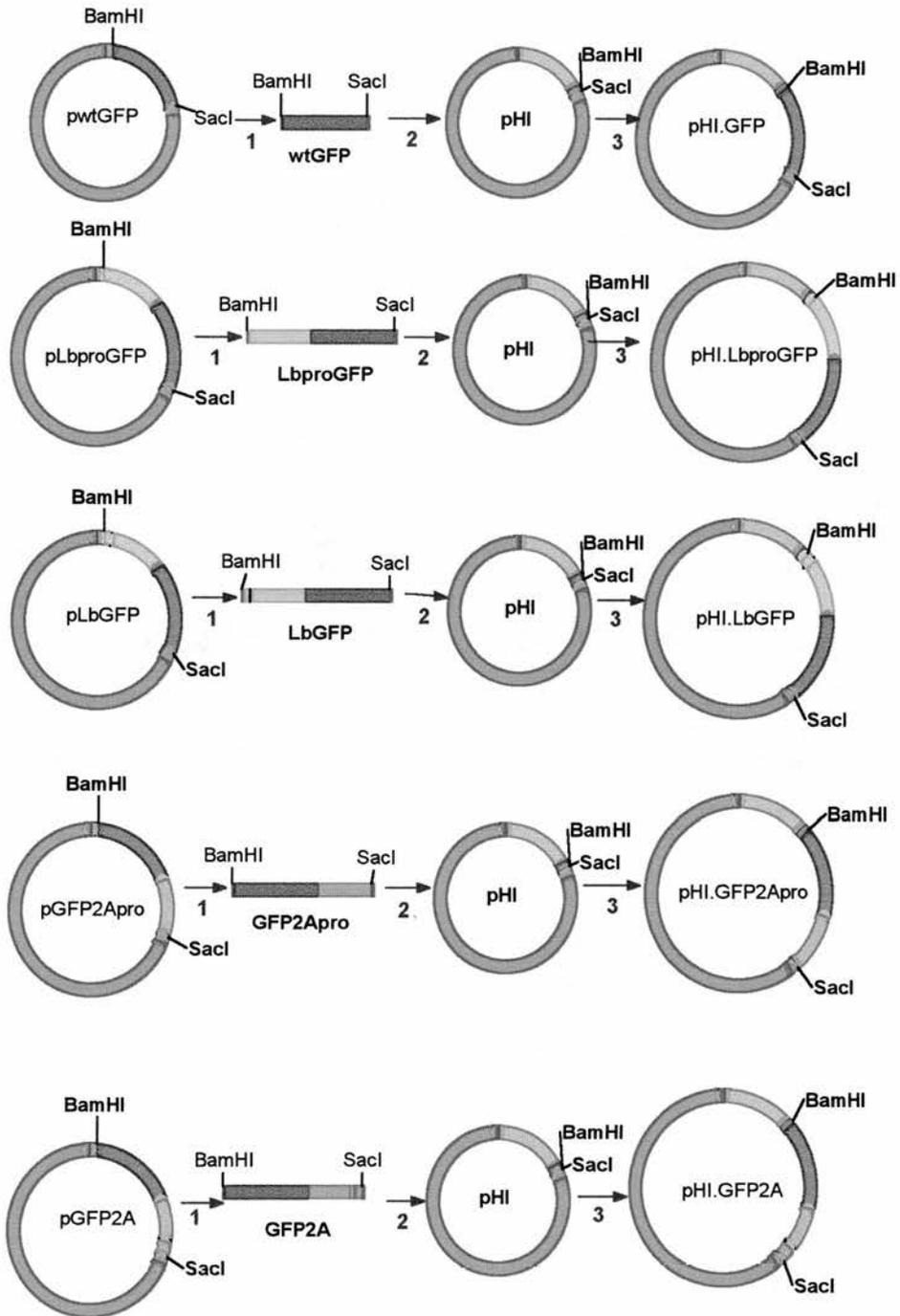
Fig.4.2. Construction of the pFI plasmid.

4.1.3. Construction of the HRV14 IRES and FMDV IRES artificial polyprotein plasmids.

To monitor the effect of the HRV14 and FMDV IRESes on the translation levels of the artificial polyproteins *in vitro* two series of plasmids containing the single reporter gene artificial polyprotein cassettes, outlined in chapter 3, downstream of the HRV14 IRES or the FMDV IRES were created. The gene cassettes from the basic reporter plasmid series of pwtGFP, pLb^{PRO}GFP, pLbGFP, pGFP2A^{PRO} and pGFP2A were excised by restricting the plasmids with BamHI and Sac I. The insert fragment was purified from agarose. The IRES-containing plasmids pHI and pFI were restricted with BamHI and Sac I generating sticky ends complementary to the insert fragments. The restricted IRES-containing vectors were purified from agarose. A set of ligation reactions was prepared for each IRES:artificial polyprotein pair. The ligation reactions were used to transform competent JM109 *E.coli* cells and plated onto antibiotic selective media. The putative clones could not be screened by α -complementation as the *lacZ* gene had already been disrupted by the insertion of the IRES. The putative clones were screened by analytical restriction digests designed to monitor the presence of the IRES and the artificial polyprotein cassette. The positive clones were also verified by sequencing using the T7 and SP6 sequencing primers. The final complement of the *cis* processing reporter plasmids is summarized below in table 4.1.

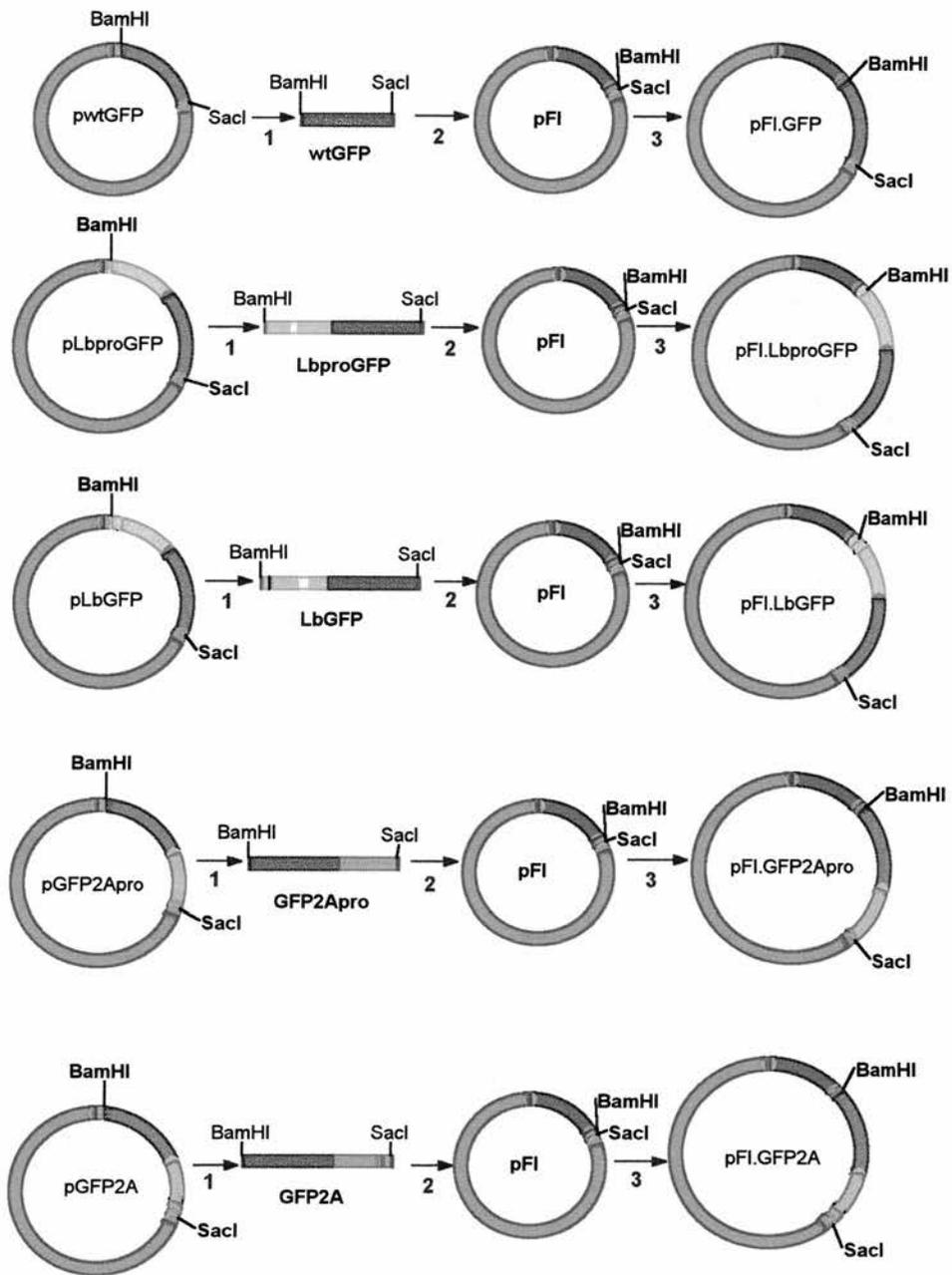
5' terminus	Protein
Uncapped	wtGFP
HRV14 IRES	wtGFP
FMDV IRES	wtGFP
Uncapped	Lb ^{pro} GFP
HRV14 IRES	Lb ^{pro} GFP
FMDV IRES	Lb ^{pro} GFP
Uncapped	LbGFP
HRV14 IRES	LbGFP
FMDV IRES	LbGFP
Uncapped	GFP2A ^{pro}
HRV14 IRES	GFP2A ^{pro}
FMDV IRES	GFP2A ^{pro}
Uncapped	GFP2A
HRV14 IRES	GFP2A
FMDV IRES	GFP2A

Table 4.1. A complete list of the basic reporter system plasmid series.



Stage 1 - Restrict the polyprotein plasmids with BamHI and SacI to generate insert fragments.
Stage 2 - Prepare the pHl plasmid by restricting with BamHI and SacI.
Stage 3 - Ligate the insert fragments and the restricted pHl plasmid together to generate the HI.polyprotein cassette plasmid series.

Fig.4.3. Construction of the HI.polyprotein plasmid series.



Stage 1 - Restrict polyprotein plasmids with BamHI and SacI to generate insert fragments.

Stage 2 - Prepare the pFI plasmid by restricting with BamHI and SacI.

Stage 3 - Ligate the insert and restricted pFI plasmid to generate the final plasmid series.

Fig.4.4. Construction of the FI polyprotein plasmid series.

4.1.4. Translation of the control plasmid series pwtGFP, pHI.GFP and pFI.GFP in vitro in rabbit reticulocyte lysate.

The plasmid DNA of the control wtGFP series pwtGFP, pHI.GFP and pFI.GFP was used to program a coupled transcription and translation *in vitro* rabbit reticulocyte system. The translation products were radiolabelled with ³⁵S-methionine and separated by denaturing-PAGE. The translation products were visualized by autoradiography. The translation profile of all three plasmids yielded a main band of about 27.5kDa this was identified as wtGFP. The plasmid pwtGFP was included on the gel as a control for the level of translation produced from an uncapped transcript, the yield of wtGFP from uncapped pwtGFP was markedly higher than from either of the IRES plasmids.

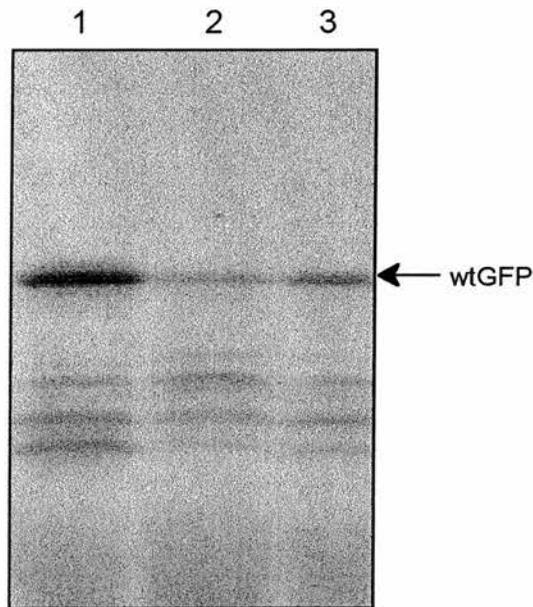


Fig.4.5. In vitro translation of the wtGFP plasmid series in RRL. Lane 1- pwtGFP, Lane 2- pHI.GFP(HRV14 IRES), Lane 3-pFI.GFP(FMDV IRES).

Observation of the translation levels from the IRES-driven transcripts indicated that the yield from pFI.GFP was marginally greater than the yield from pHI.GFP. This was

confirmed by densitometric analysis of the wtGFP protein bands on the autoradiograph using the computer program Quantiscan.

Plasmid	Protein product	Peak Height	Net Area
pwtGFP	wtGFP	200.30	558.66
pHI.GFP	wtGFP	150.80	109.67
pFI.GFP	wtGFP	162.97	179.00

Table 4.2. Densitometric analysis of wtGFP plasmid series in RRL.

4.1.5. HRV14 IRES-driven translation of the artificial reporter polyproteins *in vitro* in rabbit reticulocyte lysate.

The HRV14 IRES-driven artificial reporter polyprotein plasmids pHI.wtGFP, pHI.Lb^{pro}GFP, pHI.LbGFP, pGFP2A^{pro} and pGFP2A were used to program the *in vitro* coupled TnT system using rabbit reticulocyte lysate. The translation products were radiolabelled with ³⁵S methionine and separated by denaturing PAGE analysis. A sample of the corresponding ‘uncapped’ TnT reaction was run in the adjacent lane to allow a direct comparison of the translational efficiency of the IRES. The translation products were visualized by autoradiography.

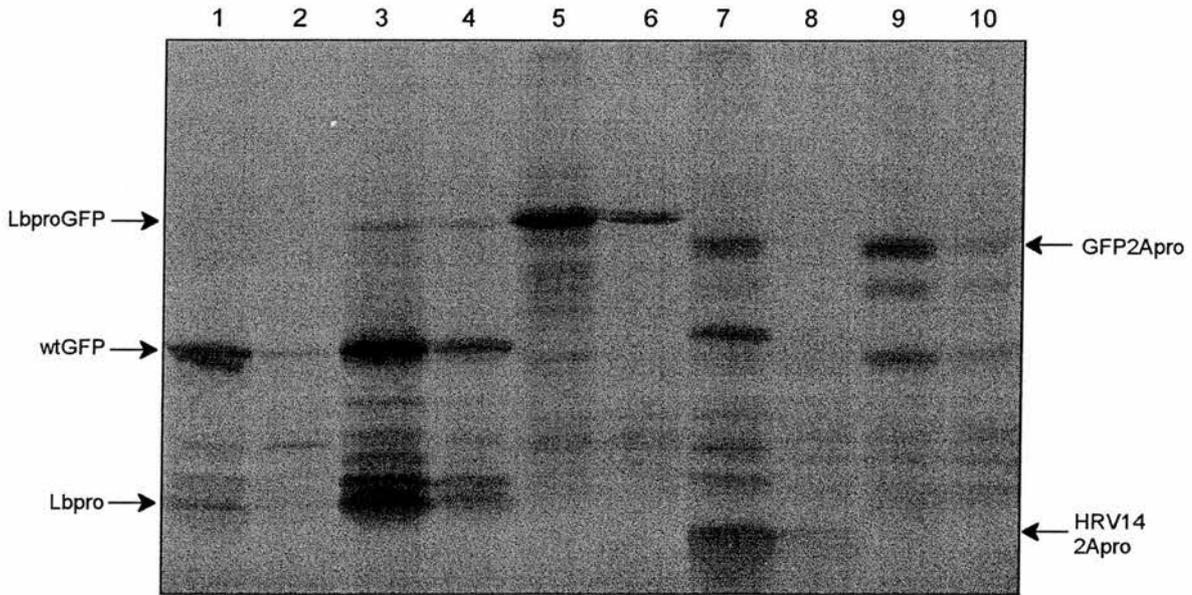


Fig.4.6. In vitro translation of HRV14 IRES-driven transcripts in RRL. Lane 1- *pwtGFP*, Lane 2- *pHI.GFP*, Lane 3- *pLb^{PRO}GFP*, Lane 4- *pHI.Lb^{PRO}GFP*, Lane 5- *pLbGFP*, Lane 6- *pHI.LbGFP*, Lane 7- *pGFP2A^{PRO}*, Lane 8- *pHI.GFP2A^{PRO}*, Lane 9- *pGFP2A*, Lane 10- *pHI.GFP2A*.

The translation profile of *pHI.Lb^{PRO}GFP* contains three major products corresponding to full-length *Lb^{PRO}GFP* and the cleavage products *wtGFP* and FMDV Lb protease. Therefore the presence of the HRV14 IRES has not affected processing of the polyprotein. The *pHI.LbGFP* TnT reaction yielded a single band of approximately 48kDa that was identified as the full-length *LbGFP* product.

The translation of the plasmid *pHI.GFP2A^{PRO}* was very inefficient although a faint band corresponding to the HRV14 *2A^{PRO}* cleavage product could be detected suggesting that the *GFP2A^{PRO}* polyprotein was processed. The translation profile of the *pHI.GFP2A* plasmid was identical to that of the uncapped *pGFP2A* plasmid containing three main bands that had been identified as the full-length *GFP2A* polyprotein and two internal initiation products.

Comparison of the translation products of the uncapped and HRV14 IRES plasmids confirmed that on the whole the presence of the IRES did not alter the products of the translation reaction but greatly reduced the yield of the translation reaction. The translation of the pHI.GFP2A^{PRO} plasmid was extremely poor thus it was difficult to ascertain if the predicted products full-length GFP2A^{PRO}, wtGFP and HRV14 2A^{PRO} were present.

4.1.6. Translation of the FMDV IRES-containing series of plasmids in vitro in RRL.

The plasmid DNA of the FMDV IRES-containing plasmid series pFI.GFP, pFI.Lb^{PRO}GFP, pFI.LbGFP, pFI.GFP2A^{PRO} and pFI.GFP2A was used to program an *in vitro* coupled TnT system in rabbit reticulocyte lysate. The translation products were radiolabelled with ³⁵S-methionine. The translation products were separated by denaturing-PAGE analysis and visualized by autoradiography. The translation products from the corresponding uncapped transcript were run in the adjacent lane to allow a direct comparison of the uncapped and FMDV IRES-driven translation.

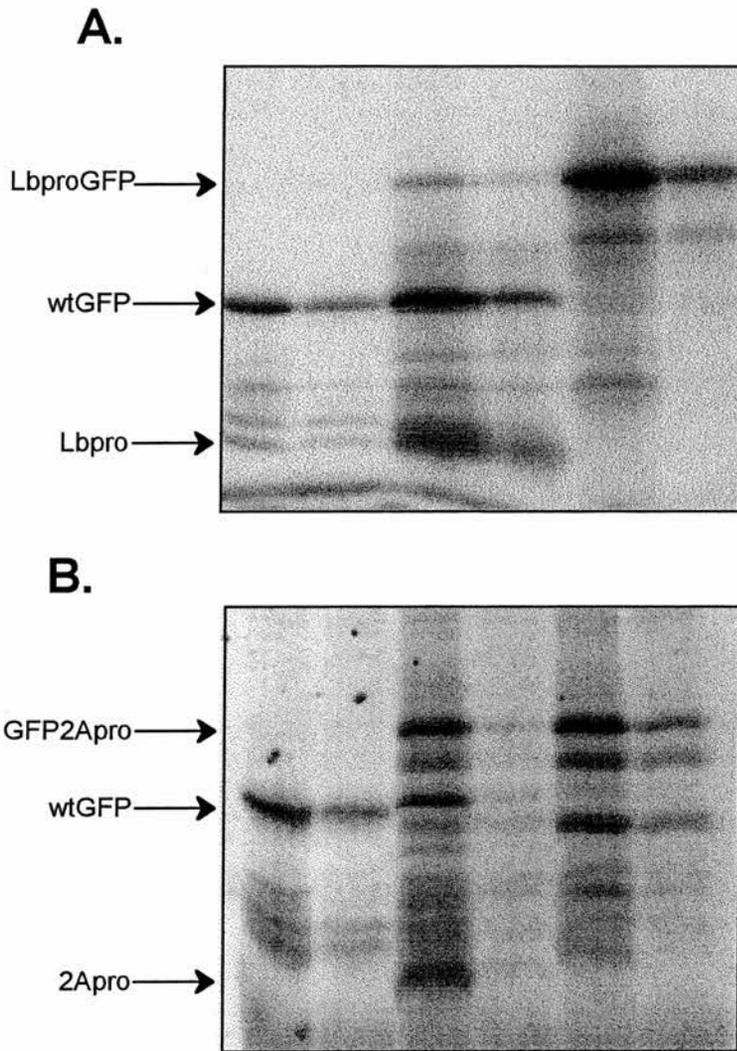


Fig.4.7. In vitro translation of FMDV IRES-driven transcripts in RRL. Gel A. Lane 1- wtGFP, Lane 2- FI.GFP, Lane 3- Lb^{pro}GFP, Lane 4- FI.Lb^{pro}GFP, Lane 5- LbGFP, Lane 6- FI.LbGFP. Gel B. Lane 1- wtGFP, Lane 2- FI.GFP, Lane 3- GFP2A^{pro}, Lane 4- FI.GFP2A^{pro}, Lane 5- GFP2A, Lane 6- FI.GFP2A.

The translation of pFI.Lb^{pro}GFP yielded three main bands corresponding to the full-length Lb^{pro}GFP polyprotein and the two cleavage products wtGFP and FMDV Lb protease. The translation of pFI.LbGFP produced a single band about 48kDa that was identified as the full-length product LbGFP. The presence of the FMDV IRES did not alter the translation profile of the polyproteins Lb^{pro}GFP and LbGFP thus the *cis* processing activity of the FMDV Lb protease was not affected by the FMDV IRES.

The translation of the plasmid pFI.GFP2A^{PRO} was very poor but faint bands corresponding to the full-length GFP2A^{PRO} polyprotein and the wtGFP cleavage product could be identified. Therefore the GFP2A^{PRO} polyprotein is processed *in cis* to produce the cleavage products. The translation of the plasmid pFI.GFP2A produced three main bands previously identified as the full-length GFP2A polyprotein and two internal initiation products.

Analysis of the uncapped translation and the analogous FMDV IRES-driven translation found that although the complement of translation products from each polyprotein was identical the yield of products from the FMDV IRES-driven transcripts was significantly lower than from the uncapped transcript.

4.1.7. A comparison of the relative translation efficiency of HRV14 IRES-driven translation and FMDV IRES-driven translation of artificial polyproteins.

The initial analysis of the translation profiles of the IRES-driven transcripts confirmed that the products of the uncapped and the IRES-driven translations were identical but the yield of the translation products varied greatly. To investigate this further the plasmid DNA of pLb^{PRO}GFP, pH1.Lb^{PRO}GFP, pFI.Lb^{PRO}GFP, pLbGFP, pH1.LbGFP and pFI.LbGFP was used to program the *in vitro* coupled TnT system in rabbit reticulocyte. The products were radiolabelled with ³⁵S-methionine and separated by denaturing-PAGE analysis. The labelled proteins were visualized by autoradiography. The uncapped, HRV14 IRES and FMDV IRES samples for each polyprotein were run in adjacent lanes to allow direct comparison of the translational efficiency of the transcripts.

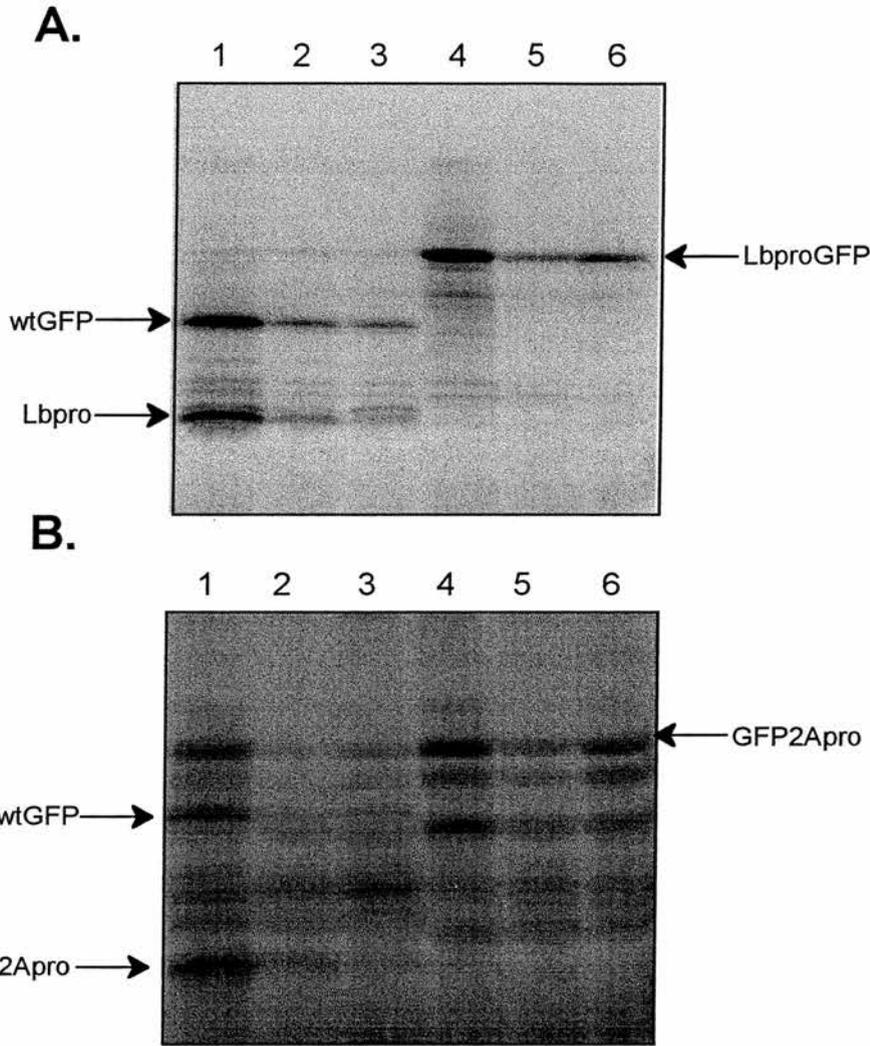


Fig.4.8. In vitro translation of uncapped and IRES-driven transcripts in RRL. Gel A. Lane 1- *pLb^{pro}GFP*, Lane 2- *pHI.Lb^{pro}GFP*, Lane 3- *pFI.Lb^{pro}GFP*, Lane 4- *pLbGFP*, Lane 5- *pHI.LbGFP*, Lane 6- *pFI.LbGFP*. Gel B. Lane 1- *pGFP2A^{pro}*, Lane 2- *pHI.GFP2A^{pro}*, Lane 3- *pFI.GFP2A^{pro}*, Lane 4- *pGFP2A*, Lane 5- *pHI.GFP2A*, Lane 6- *pFI.GFP2A*.

Observation of the translation profiles of the uncapped and IRES-driven reactions indicated that for all the plasmids tested the uncapped plasmid was translated most efficiently. Comparison of the translation profiles of the IRES-driven transcripts indicated that the FMDV IRES was more efficient than the HRV14 IRES at initiating translation in this system. These observations were confirmed by completing a

densitometric analysis of the autoradiograph using the computer program Quantiscan to quantify the products of each reaction.

Construct	Protein Product	Peak height	Net Area
pLb ^{pro} GFP	Lb ^{pro} GFP	161.09	213.61
	wtGFP	216.50	1089.70
	Lb ^{pro}	216.32	2104.05
pHI.Lb ^{pro} GFP	Lb ^{pro} GFP	150.19	176.70
	wtGFP	205.81	1110.11
	Lb ^{pro}	194.59	1127.84
pFI.Lb ^{pro} GFP	Lb ^{pro} GFP	150.02	162.85
	wtGFP	196.76	874.15
	Lb ^{pro}	178.61	513.83
pLbGFP	LbGFP	214.46	1695.85
pHI.LbGFP	LbGFP	202.85	767.78
pFI.LbGFP	LbGFP	198.80	974.24
pGFP2A ^{pro}	GFP2A ^{pro}	182.83	660.28
	wtGFP	196.76	619.90
	2A ^{pro}	186.90	1182.38
pHI.GFP2A ^{pro}	GFP2A ^{pro}	-	-
	wtGFP	-	-
	2A ^{pro}	-	-
pFI.GFP2A ^{pro}	GFP2A ^{pro}	-	-
	wtGFP	-	-
	2A ^{pro}	-	-

pGFP2A	GFP2A	201.00	863.03
	Product A	171.09	531.00
	Product B	184.11	462.37
pHI.GFP2A	GFP2A	147.03	231.21
	Product A	153.10	381.28
	Product B	143.08	130.79
pFI.GFP2A	GFP2A	163.85	559.72
	Product A	152.62	213.44
	Product B	153.97	443.95

Table 4.3. Densitometric analysis of translation products of the *pLb^{PRO}GFP* and *pGFP2A^{PRO}* plasmid series in vitro in rabbit reticulocyte lysate.

The results of the densitometric analysis highlighted an interesting result that was not immediately obvious from observation of the autoradiograph. The results show that translation of the *pHI.Lb^{PRO}GFP* plasmid was more efficient than translation of the *pFI.Lb^{PRO}GFP* plasmid. The order of translational efficiency for all the other plasmid series tested was confirmed as uncapped > FMDV IRES > HRV14 IRES thus the *Lb^{PRO}GFP* plasmid series did not follow the common trend.

4.1.8. The effect of the HRV14 2A^{PRO} on IRES-driven translation.

Initial observations of the translation of the IRES-driven transcripts containing the active HRV14 2A protease noted that the translation levels were reduced in comparison to the level of translation of similar IRES-driven transcripts. By exposing film to the radioactive sample for an extended period the extent of the translation reduction in the IRES-driven 2A^{PRO} transcripts was clearly evident.

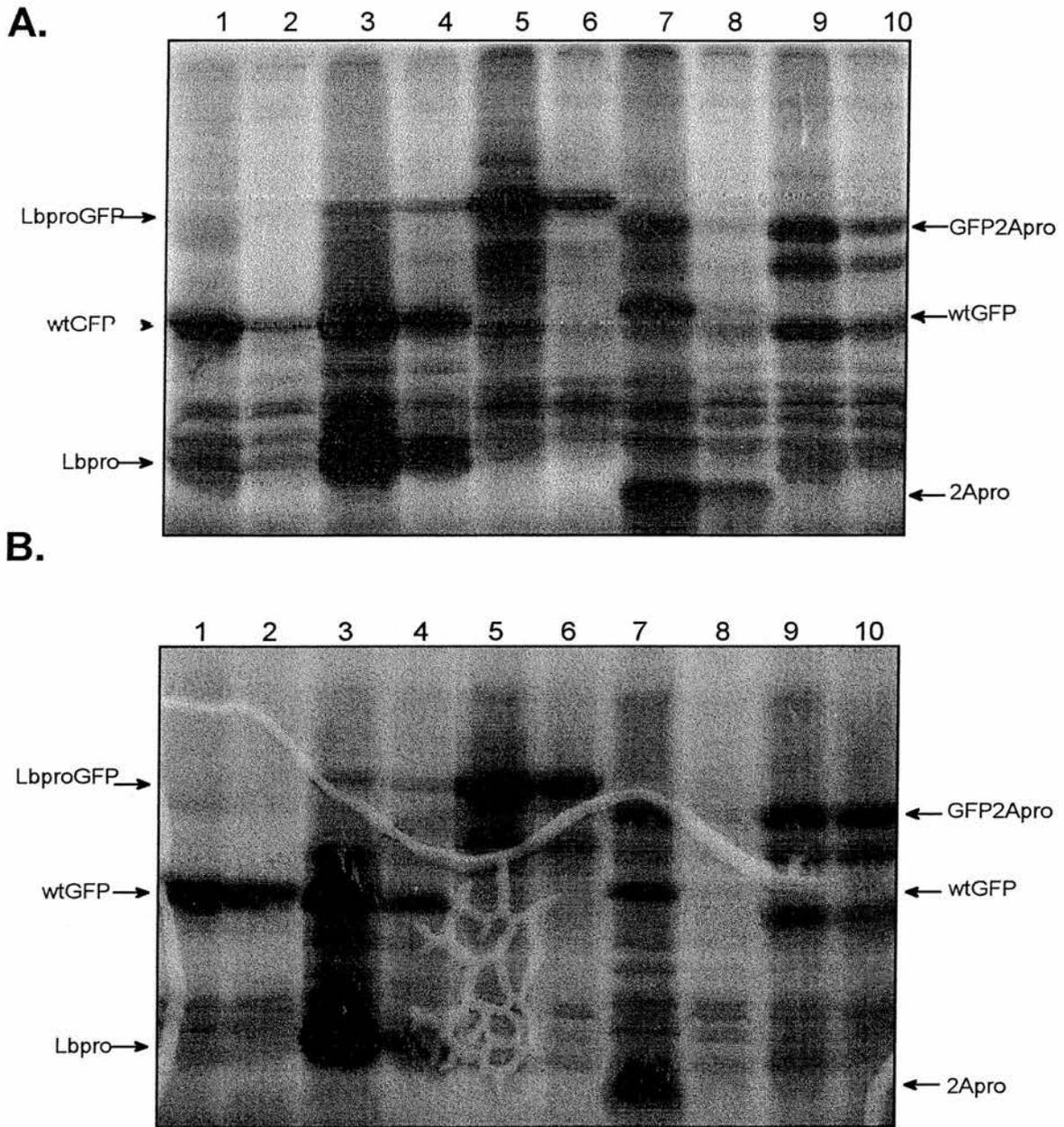


Fig.4.9. A prolonged exposure of the complete plasmid series translation profiles in RRL. Gel A. Lane 1- *pwtGFP*, Lane 2- *pHI.GFP*, Lane 3- *pLb^{pro}GFP*, Lane 4- *pHI.Lb^{pro}GFP*, Lane 5- *pLbGFP*, Lane 6- *pHI.LbGFP*, Lane 7- *pGFP2A^{pro}*, Lane 8- *pHI.GFP2A^{pro}*, Lane 9- *pGFP2A*, Lane 10- *pHI.GFP2A*. Gel B. Lane 1- *pwtGFP*, Lane 2- *pFI.GFP*, Lane 3- *pLb^{pro}GFP*, Lane 4- *pFI.Lb^{pro}GFP*, Lane 5- *pLbGFP*, Lane 6- *pFI.LbGFP*, Lane 7- *pGFP2A^{pro}*, Lane 8- *pFI.GFP2A^{pro}*, Lane 9- *pGFP2A*, Lane 10- *pFI.GFP2A*.

In gel A and gel B the translation level of the sample in lane 8 is severely reduced in contrast to the rest of the samples. Lane 8 contains the translation products of plasmids pHI.GFP2A^{pro} in gel A and pFI.GFP2A^{pro} in gel B. Faint bands correlating to the predicted products full-length GFP2A^{pro} and the cleavage products wtGFP and HRV14 2A^{pro} could be identified for pHI.GFP2A^{pro} and pFI.GFP2A^{pro} thus a small amount of translation has occurred.

The translation products of the IRES-driven GFP2A plasmids and the IRES-driven LbGFP plasmids were loaded onto a single gel to allow direct comparison of the samples. The proteins were separated by denaturing-PAGE and visualized by autoradiography.

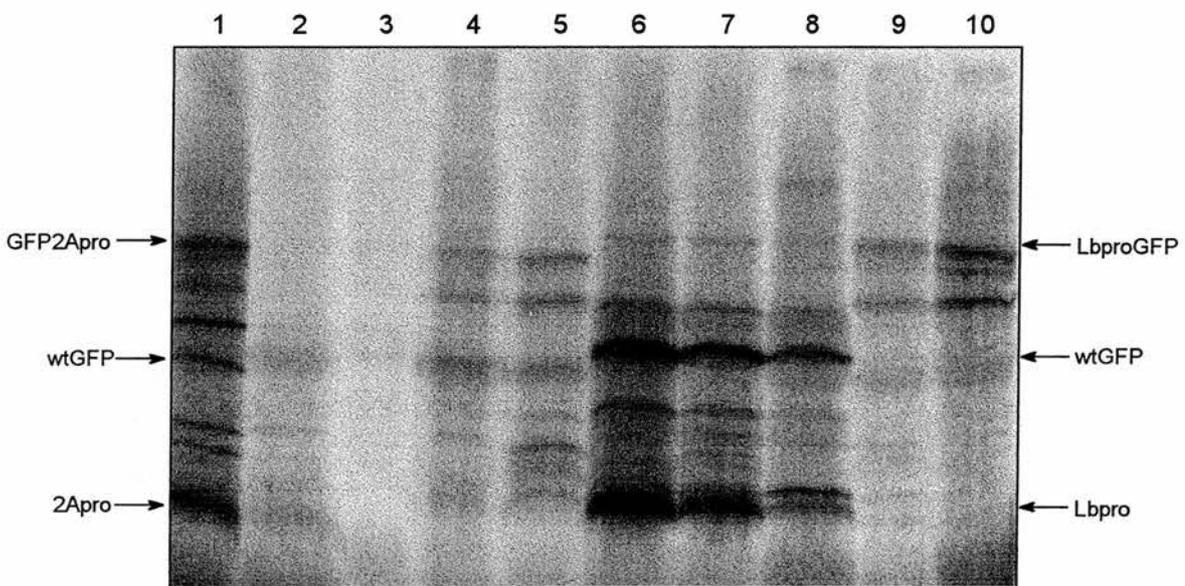


Fig.4.10. IRES-driven polyprotein products in RRL from Lane 1- pGFP2A^{pro}, Lane 2- pHI.GFP2A^{pro}, Lane 3- pFI.GFP2A^{pro}, Lane 4- pHI.GFP2A, Lane 5- pFI.GFP2A, Lane 6- pLb^{pro}GFP, Lane 7- pHI.Lb^{pro}GFP, Lane 8- pFI.Lb^{pro}GFP, Lane 9- pHI.LbGFP, Lane 10- pFI.LbGFP.

The translation products from the IRES-driven GFP2A^{pro} transcripts are not visible. Translation of the IRES-driven GFP2A transcripts yields the full-length GFP2A polyprotein and the internal initiation products, products A and B which can be readily

identified in lanes 4 and 5. Comparison of lanes 1 to 5 with the parallel lanes 6 to 10 containing the Lb^{PRO}GFP plasmid series clearly shows that the same pattern is not evident for the Lb^{PRO}GFP series.

4.1.9. Translation of the control IRES-driven wtGFP series *in vitro* in wheatgerm extract.

The control wtGFP plasmid series was used to program an *in vitro* coupled TnT system in wheat germ extract to investigate the potential use of this system to translate IRES-driven transcripts. The translation products were labelled with ³⁵S-methionine and separated by denaturing-PAGE. The radiolabelled proteins were visualized by autoradiography. Samples of the corresponding plasmid translation products in rabbit reticulocyte lysate were also analyzed by denaturing PAGE. The translation products of the wtGFP plasmid series in rabbit reticulocyte lysate and wheat germ extract are shown below.

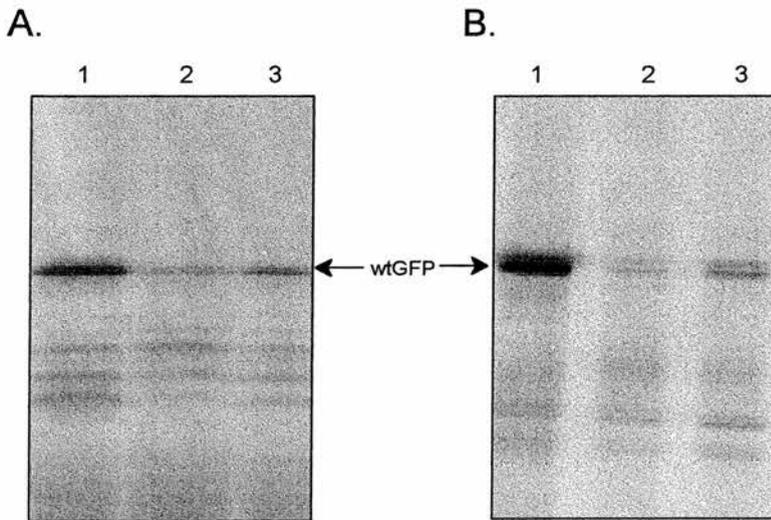


Fig.4.11. Comparison of *in vitro* translation of the *pwtGFP* control plasmid series in RRL and WGE. Gel A- In RRL, Lane 1-*pwtGFP*, lane 2- *pHI.GFP*, Lane 3- *pFI.GFP*. Gel B- In WGE, Lane 1-*pwtGFP*, Lane 2- *pHI.GFP*, Lane 3- *pFI.GFP*.

The main product in all lanes was identified as wtGFP confirming that the IRES-driven transcripts were successfully translated in the wheat germ extract. The product yield from uncapped and IRES-driven transcripts in both systems followed the same trend. The uncapped transcripts were translated more efficiently than transcripts containing either IRES. An observational comparison of the IRES-driven translation products indicates that marginally more wtGFP is produced from the FMDV IRES transcripts than the HRV14 IRES transcripts.

4.1.10. Translation of the IRES-driven artificial polyproteins in vitro in wheatgerm extract.

The control wtGFP plasmid series had been successfully translated *in vitro* in wheatgerm extract therefore the full complement of IRES plasmids were used to program the *in vitro* coupled TnT system in wheat germ extract. The proteins were radiolabelled with ³⁵S-methionine. The samples were analyzed by denaturing PAGE and visualized by autoradiography. Duplicate gels containing samples translated *in vitro* in rabbit reticulocyte lysate were produced to allow a direct comparison of the translation profile of each plasmid in both systems.

Observation of the translation profiles of the FMDV Lb^{PRO} plasmid series and the HRV14 2A^{PRO} plasmid series in rabbit reticulocyte lysate and wheatgerm extract clearly shows that overall the product yield is greater in rabbit reticulocyte lysate for all transcripts. By taking the overall decreased yield into consideration for the wheatgerm extract samples it is apparent that the trend of dramatically reduced translation of IRES-driven transcripts previously noted for the rabbit reticulocyte lysate samples is echoed in the wheatgerm extract samples. Comparison of the different IRES transcripts demonstrates again that translation of the FMDV IRES transcripts is marginally greater than translation of the HRV14 IRES transcripts.

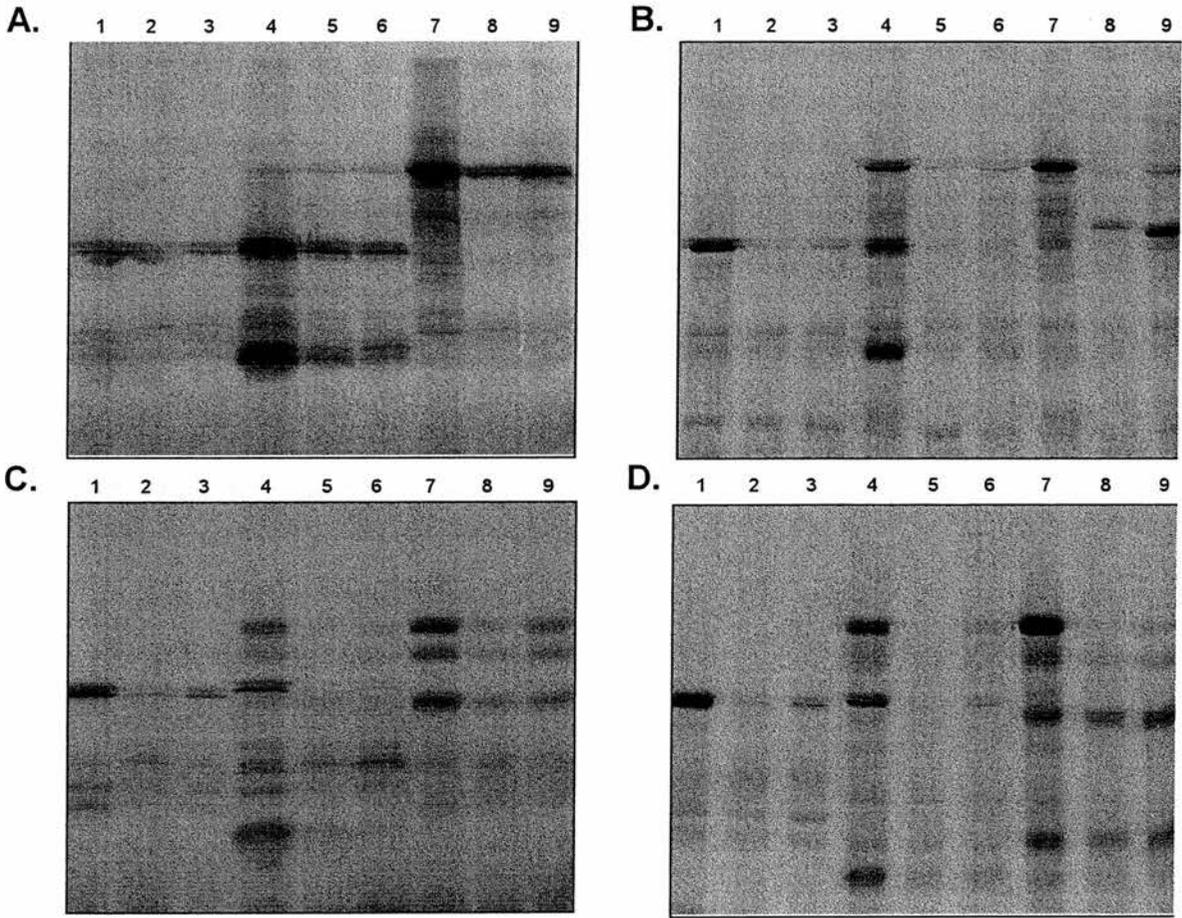


Fig.4.12. Gel A – Lb^{PRO} series in RRL, Gel B – Lb^{PRO} series in WGE. For both gels Lane 1- $pwtGFP$, Lane 2- $pHI.GFP$, Lane 3- $pFI.GFP$, Lane 4- $pLb^{PRO}GFP$, Lane 5- $pHI.Lb^{PRO}GFP$, Lane 6- $pFI.Lb^{PRO}GFP$, Lane 7- $pLbGFP$, Lane 8- $pHI.LbGFP$, Lane 9- $pFI.LbGFP$. Gel C – $2A^{PRO}$ series in RRL, Gel D – $2A^{PRO}$ series in WGE. For both gels Lane 1- $pwtGFP$, Lane 2- $pHI.GFP$, Lane 3- $pFI.GFP$, Lane 4- $pGFP2A^{PRO}$, Lane 5- $pHI.GFP2A^{PRO}$, Lane 6- $pFI.GFP2A^{PRO}$, Lane 7- $pGFP2A$, Lane 8- $pHI.GFP2A$, Lane 9- $pFI.GFP2A$.

Inspection of the $pLbGFP$ and $pGFP2A$ plasmid series in wheatgerm extract and rabbit reticulocyte lysate identified a discrepancy between translation of these plasmids in the alternative systems. In rabbit reticulocyte lysate translation of $pLbGFP$ produced a single

band of approximately 48kDa that was identified as the full-length LbGFP polyprotein. In the wheatgerm extract however, the uncapped transcript produced the major LbGFP band but the major protein produced from the IRES transcripts was significantly smaller (30-40kDa).

A similar trend is evident in the pGFP2A samples. In rabbit reticulocyte lysate the principle protein from translation of the uncapped and IRES transcripts is the full-length GFP2A, smaller amounts of two proteins identified as internal initiation products are also present. The same bands are found in the wheatgerm extract translations but for the IRES transcripts the balance has shifted resulting in a large increase in the level of the smallest internal initiation product.

The shift in the selection of the preferred initiation site is not apparent in rabbit reticulocyte lysate translations or in wheatgerm extract translation of the polyproteins containing an active protease.

Construct	Protein Product	Peak Height	Net Area
pLb ^{pro} GFP	Lb ^{pro} GFP	213.05	906.89
	wtGFP	206.70	1158.20
	Lb ^{pro}	205.91	1557.49
pHI.Lb ^{pro} GFP	Lb ^{pro} GFP	82.02	129.47
	wtGFP	70.48	49.73
	Lb ^{pro}	74.62	49.02
pFI.Lb ^{pro} GFP	Lb ^{pro} GFP	96.25	9.45
	wtGFP	77.50	64.32
	Lb ^{pro}	-	-

pLbGFP	LbGFP	233.97	1121.41
pHI.LbGFP	LbGFP	82.88	98.94
	36kDa product	137.36	429.94
pFI.GFP	LbGFP	135.37	338.71
	35kDa product	216.00	870.51
pGFP2A ^{pro}	GFP2A ^{pro}	227.79	1133.00
	wtGFP	226.86	274.84
	2A ^{pro}	145.59	949.30
pHI.GFP2A ^{pro}	GFP2A ^{pro}	79.72	209.14
	wtGFP	-	-
	2A ^{pro}	-	-
pFI.GFP2A ^{pro}	GFP2A ^{pro}	92.73	51.87
	wtGFP	96.05	62.39
	2A ^{pro}	-	-
pGFP2A	GFP2A	253.64	1796.48
	Product A	168.04	83.11
	Product B	177.56	843.26
pHI.GFP2A	GFP2A	69.68	100.00
	Product A	75.72	2.28
	Product B	138.76	747.40
pFI.GFP2A	GFP2A	88.88	34.76
	Product A	98.44	345.20
	Product B	170.28	1204.83

Table 4.4. Densitometric analysis of the translation profiles in WGE.

The densitometric analysis results confirm the observation that the IRES-driven transcripts are translated, although inefficiently, *in vitro* in wheatgerm extract. The polyproteins are processed efficiently by the active protease to produce the cleavage products in the presence of the IRES. The principle trends that were observed in the rabbit reticulocyte lysate system are also apparent in the wheatgerm extract system. Namely, that the uncapped transcripts are translated highly efficiently *in vitro* in wheatgerm extract, in contrast, the IRES-driven transcripts are poorly translated. Comparison of the ability of the different IRES types to initiate translation in this system identified a marginal distinction between the FMDV IRES and the HRV14 IRES. Thus the order of translational efficiency of transcripts in wheatgerm extracts is uncapped > FMDV IRES > HRV14 IRES.

4.1.11. Translation of pLb^{PRO}GFP and pLbGFP in vitro in RRL and WGE.

As outlined above in section 4.1.9. observation of the translation profiles of the pLbGFP plasmid series in rabbit reticulocyte lysate and wheatgerm extract identified a discrepancy involving the translation of the IRES-driven transcripts in the wheatgerm system. The translation products from the pLbGFP plasmid series in rabbit reticulocyte lysate and wheatgerm extract are shown below.

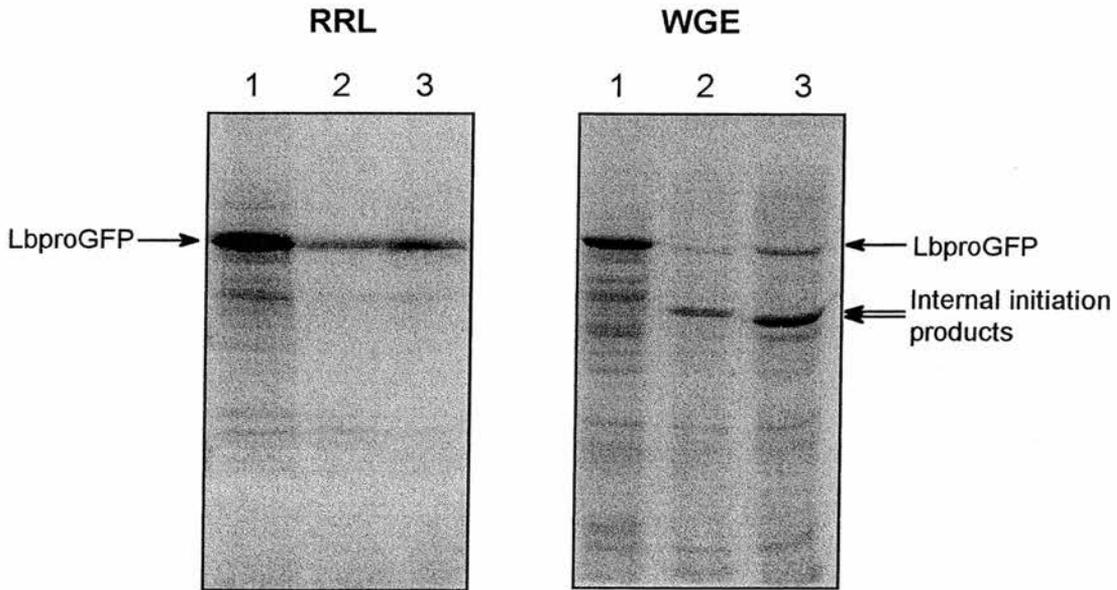


Fig.4.13. Translation of the pLbGFP plasmid series in RRL and WGE. In both gels, Lane 1- pLbGFP, Lane 2- pHI.LbGFP, Lane 3- pFI.LbGFP.

Translation of the LbGFP polyprotein in rabbit reticulocyte lysate from an uncapped or IRES-driven transcript produces a single major protein identified as the full-length LbGFP polyprotein. The FMDV Lb protease in pLbGFP is inactive due to a single point mutation of the active site histidine residue therefore the polyprotein is not processed to produce wtGFP and Lb^{PRO}.

In contrast, *in vitro* translation of the pLbGFP plasmid series in wheatgerm extract produced a different translation profile for each plasmid. The uncapped pLbGFP transcript produced a single product identified as the LbGFP polyprotein. The HRV14 IRES LbGFP transcript produced a small amount of the full-length product but the major protein was about 35kDa. Finally, the FMDV IRES LbGFP transcript also produced a small amount of the full-length polyprotein but the principle band was again about 35kDa. These products were not observed in the translation profiles of the pHI.Lb^{PRO}GFP and pFI.Lb^{PRO}GFP plasmids. A direct comparison of the HRV14 IRES and FMDV IRES-

driven translations revealed that the major product of the FMDV IRES translation was slightly smaller than the HRV14 IRES major band.

Examination of the sequence of FMDV Lb^{PRO} identified eight ATG codons including the authentic initiation codon. The computer program DNA Strider was used to predict the size of the protein produced from these alternative initiation codons (refer to table 4.5).

ATG codon position	Sequence context	Predicted Size (kDa)
ATG (1) authentic codon	GGATCCATGGAA	47.5
ATG (200)	AACTGCATGAGG	0.97
ATG (292)	GTGTGCATGGTG	36.4
ATG (299)	TGGTGGATGGTA	8.9
ATG (310)	ACGGACATGTGC	35.8
ATG (329)	ATTTCCATGCTG	7.7
ATG (407)	TTGACGATGAGG	4.9
ATG (489)	CGGGGAATGGAA	2.1

Table 4.5. Putative initiation codons present within FMDV Lb^{PRO}.

Only two of the seven alternative initiation sites were in frame with the authentic initiation codon thus producing proteins of significant length to be identified by denaturing-PAGE. The ATG codon at position 292 produced a protein product of 36.4 kDa and the ATG codon at position 310 produced a protein product of 35.8 kDa. The molecular weights of these predicted internal initiation products correlate with the approximate sizes of the major bands produced from pHI.LbGFP and pFI.LbGFP.

Therefore the major product of the pHI.LbGFP plasmid in wheatgerm extract is an internal initiation product from the ATG at position 292. The major product from translation of the pFI.LbGFP plasmid in wheatgerm extract is also an internal initiation product resulting from initiation of translation at position 310.

A similar trend was identified for the IRES-driven GFP2A plasmids in wheatgerm extract. Translation of the GFP2A plasmid series in rabbit reticulocyte lysate and wheatgerm extract yields three major products, the full-length GFP2A polyprotein and two internal initiation products (labelled A and B). Examination of the densitometric results for the GFP2A plasmid series in rabbit reticulocyte lysate and wheatgerm extract detected a shift in the balance of the translation products of the pHI.GFP2A and pFI.GFP2A plasmids in wheatgerm extract.

Translation	GFP2A protein	Product A	Product B
pGFP2A RRL	201.00	171.09	184.11
pHI.GFP2A RRL	147.03	153.10	143.08
pFI.GFP2A RRL	163.85	152.65	153.97
pGFP2A WGE	253.64	168.04	177.56
pHI.GFP2A WGE	69.68	75.72	138.76
pPFI.GFP2A WGE	88.88	98.44	170.28

Table 4.6. Peak height of translation products of the pGFP2A plasmid series in RRL and WGE.

The three main products are produced in similar amounts when the GFP2A plasmid series are translated in rabbit reticulocyte lysate. The uncapped GFP2A transcript also

produces similar amounts of the three products. The IRES-driven transcripts when translated in wheatgerm extract however, produced increased amounts of the smallest protein product B.

4.1.12. Conclusions.

The HRV14 IRES and the FMDV IRES were successfully cloned into the pGEM7zf+ vector to create the plasmids pHI and pFI that were subsequently used to insert an IRES upstream of the protein coding region of the plasmids created in chapter 3 to produce the complete set of 15 plasmids.

The IRES-driven plasmids were translated correctly *in vitro* in both rabbit reticulocyte lysate and wheatgerm extract. Overall the product yield was greater in the rabbit reticulocyte lysate than in wheatgerm extract. The Lb^{PRO}GFP and GFP2A^{PRO} polyproteins were processed *in cis* in the presence of the IRES to yield the cleavage products wtGFP and the active protease.

In both systems translation of the uncapped transcript was considerably more efficient than translation of the IRES-driven transcripts for all the plasmid series tested. A comparison of the translational efficiency of the different types of IRES showed that although the level of translation from the IRESes was quite poor overall, the FMDV IRES transcripts were translated marginally better than the parallel HRV14 IRES transcripts. There was one exception densitometric analysis of the Lb^{PRO}GFP plasmid series found that in this case translation of the HRV14 IRES transcript was better than translation of the FMDV IRES Lb^{PRO}GFP transcript.

Interestingly, it was noted that the translation of IRES-driven GFP2A^{PRO} transcripts was severely inhibited in rabbit reticulocyte lysate. Translation of the uncapped GFP2A^{PRO} transcript was not inhibited. A single point mutation within the 2A^{PRO} that renders the

2A^{PRO} inactive alleviated the inhibition as translation of IRES-driven GFP2A transcripts was not impaired. This would suggest that the proteolytic activity of HRV14 2A^{PRO} is responsible for the inhibition. A similar pattern of inhibition was not observed for the IRES-driven Lb^{PRO}GFP transcripts therefore the target of the HRV14 2A protease that results in this effect is not shared by the FMDV Lb protease.

The translation profile of the IRES-driven LbGFP transcripts in wheatgerm extracts demonstrated a dramatic shift in the selection of the principle initiation codon. A reduced amount of the full-length LbGFP polyprotein was produced in parallel with the appearance of two smaller internal initiation products. *In vitro* translation of the pHI.LbGFP transcript favoured initiation at position 292 yielding a 36.4kDa product. Similarly, initiation of translation of the pFI.LbGFP transcript preferentially occurred at position 310 to yield a 35.8kDa protein product. These internal initiation products were not detected in the translation profiles of the pHI.Lb^{PRO}GFP and pFI.Lb^{PRO}GFP transcripts in wheatgerm extract. Thus it is possible that the proteolytic activity of the FMDV Lb protease has a role in the selection of the initiation codon in the wheatgerm system. Abolishing the proteolytic activity of the protease results in increased leaky scanning producing aberrant internal initiation products. A similar increase in the leakiness of the authentic initiation codon for the GFP2A polyprotein from IRES-driven transcripts in wheatgerm extract resulting in larger amounts of internal initiation products was also reported.

4.1.13. Summary

- An IRES-driven artificial self-processing polyprotein system has been constructed and analyzed *in vitro*.
- The efficiency of type I IRES and type II IRES driven translation has been monitored *in vitro* using a coupled transcription and translation system.
- The artificial polyproteins were translated from the IRES-driven transcripts in RRL and WGE to yield the predicted products. Overall translation was more efficient in the RRL system.
- The artificial polyproteins were correctly processed in the presence of the IRES.
- Translation of the artificial polyproteins was most efficient from the uncapped transcripts in both the RRL and WGE systems.
- Overall translation of the polyprotein was marginally better from the type II IRES than the type I IRES.
- The presence of active FMDV LbP^{ro} altered IRES-driven translation. Thus for the LbP^{ro}GFP transcripts translation from the type I IRES was better than translation from the type II IRES.
- The combination of an active HRV14 2A protease and an IRES resulted in severe inhibition of translation in RRL.
- Translation of the polyproteins containing the mutated protease in WGE resulted in increased leaky scanning thus the yield of internal initiation products also increased.

4.2. Part 2: *In vitro* analysis of translation in rabbit reticulocyte lysate using a two-step coupled system.

4.2.1. Introduction.

The observations of the translational efficiency of the relative uncapped and IRES-driven transcripts described in Part 1 of this chapter were somewhat surprising. The observed trend that translation of the uncapped messages was very efficient in this system and that IRES-driven translation was substantially less efficient was at complete variance with the established literature on this subject.

Examination of the literature in finer detail established that one major difference was the type of system used. The reported *in vitro* translation data had been collected using the rabbit reticulocyte lysate translation system that was programmed with an RNA transcript produced in a separate reaction. In this experiment we had used a single step coupled transcription and translation system. Therefore the transcription and translation processes were closely linked in time and space due to the nature of the system transcription no longer occurred within a membrane-bound nucleus. Consequently, although the extract was prepared from eukaryotic cells the reaction environment mimicked the prokaryotic cellular environment. This led to the proposal that, as in the prokaryotic cell, the transcription and translation processes were tightly coupled in this system with the RNA being translated by the ribosomal machinery before formation of the IRES. Thus the nascent RNA structure is 'melted' by the ribosomal machinery preventing development of the IRES complex secondary structure.

A number of other observations were also noted that conflicted with the established viewpoint. The first was that the translational efficiency of the HRV14 IRES-driven transcripts was very similar to the translational efficiency of the FMDV IRES-driven transcripts. In contrast to my observations it was reported that type I IRESes were

extremely inefficient in reticulocyte lysate and that the lysate must be supplemented with HeLa cell extract to obtain any level of translation. In particular the HRV14 IRES was reported to be the least efficient of the type I IRESes tested (Roberts *et al.*, 1998). The second report was that IRES-driven transcripts could not be translated in wheatgerm extract (Jackson and Kaminski, 1995) however, using the coupled system the IRES-containing plasmids were translated in wheatgerm extract albeit inefficiently.

These puzzling observations supported the theory that within the coupled transcription and translation system the IRES was not formed and thus could not recruit the ribosomal machinery. To test this theory an alternative coupled transcription and translation system supplied by Novagen was used. This system separates the transcription and translation stages. The plasmid DNA is added to a transcription mixture and transcribed for 15 minutes, at this point the translation mixture is added that inhibits further transcription. The reaction is incubated for a further 60 minutes to allow translation to occur. Using this alternative system we can separate transcription and translation allowing the RNA secondary structure to form.

4.2.2. In vitro analysis of the wtGFP plasmid series using the Novagen STP3 system.

The plasmid DNA of the control wtGFP plasmid series pwtGFP, pHI.GFP and pFI.GFP was used to program the transcription mix of the STP3 system. The reaction was incubated at 30°C for 15 minutes to allow transcription to occur. The translation mix was added and the reaction was incubated at 30°C for a further 60 minutes. The translation products were labelled with ³⁵S-methionine and separated by denaturing-PAGE. The translation products were visualized by autoradiography. A prolonged exposure time of 60 hours was used to visualize any low intensity bands.

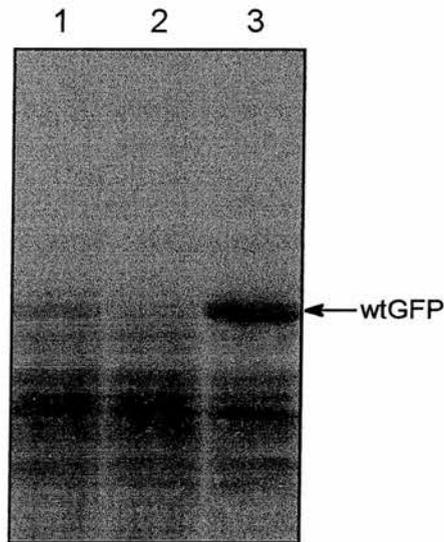


Fig.4.14. Translation of wtGFP plasmid series using the STP3 transcription and translation system from Novagen. Lane 1- pwtGFP, Lane 2- pHI.GFP, Lane 3- pFI.GFP.

A major product of approximately 28kDa that was positively identified as wtGFP was produced from the FI.GFP plasmid. Thus the FMDV IRES-driven wtGFP transcript was very efficiently translated. After a prolonged exposure period very faint bands of the same size could be detected in the translation profiles of pwtGFP and pHI.GFP. The uncapped and HRV14 IRES-driven transcripts were very poorly translated in this system.

In order to quantify the translational efficiency of each plasmid the gel was exposed to a phosphorimaging plate for 2 hours. The results of the phosphorimaging analysis are presented below in table 4.7. For the wtGFP plasmid series translation from the FMDV IRES is approximately twenty times more efficient than translation of the uncapped message.

Plasmid	Protein product	PSL-BG
pwtGFP	wtGFP	64.02
pHI.GFP	wtGFP	-
pFI.GFP	wtGFP	1350.00

Table 4.7. Phosphorimaging results of the translation profiles of the wtGFP plasmid series in the STP3 TnT system.

4.2.3. *In vitro* analysis of the Lb^{PRO}GFP plasmid series using the Novagen STP3 system.

The plasmid DNA of the Lb^{PRO}GFP plasmid series pLb^{PRO}GFP, pHI.Lb^{PRO}GFP, pFI.Lb^{PRO}GFP, pLbGFP, pHI.LbGFP and pFI.LbGFP were used to program the transcription mix of the STP3 system. The reaction was incubated at 30°C for 15 minutes to allow transcription of the DNA. The translation mix was added and the reaction was incubated at 30°C for a further 60 minutes. The protein products were radiolabelled with ³⁵S-methionine. Samples of the reactions were separated by denaturing-Page analysis and visualized by autoradiography using a prolonged exposure period of 60 hours. Aliquots of the wtGFP plasmid TnT reactions were also loaded onto the gel to allow comparison of the translation profiles and to aid in identification of the GFP protein.

Observation of the translation profiles of the Lb^{PRO}GFP plasmid series identified a band in all lanes of approximately 48kDa this was identified as the full-length LbGFP polyprotein. Therefore the plasmids are all translated successfully in the STP3 TnT system. A band corresponding to the wtGFP cleavage product was identified in the translation profiles of pLb^{PRO}GFP, pHI.Lb^{PRO}GFP and pFI.Lb^{PRO}GFP thus the polyprotein is processed *in cis* correctly in this system. Unfortunately due to the large

number of smaller background bands we were unable to positively identify the Lb^{PRO} product.

In accordance with the previous observations of the relative translational efficiency of the wtGFP plasmid series within this system the uncapped and HRV14 IRES-driven transcripts were poorly translated and the FMDV IRES-driven transcripts were very efficiently translated. Interestingly, translation of the IRES-driven Lb^{PRO}GFP transcripts did not conform to this pattern. In this case the HRV14 IRES Lb^{PRO}GFP transcript was translated more efficiently than the FMDV IRES transcript, this observation supports a similar anomaly observed in section 4.1.7.

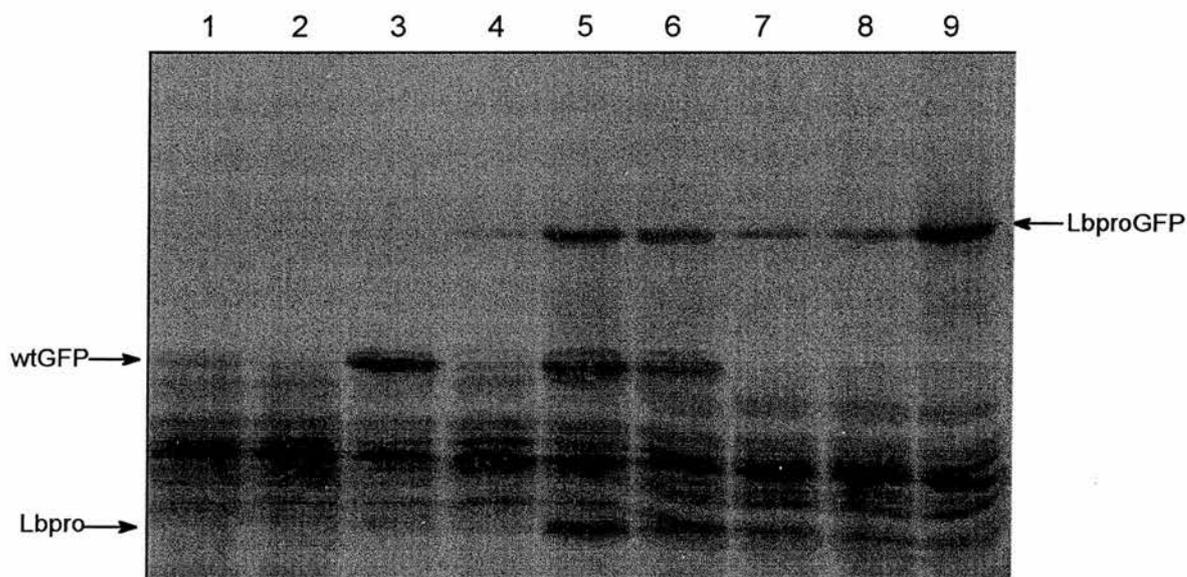


Fig.4.15. *In vitro* analysis of the Lb^{PRO}GFP plasmid series using the STP3 system from Novagen. Lane 1- pwtGFP, Lane 2- pHI.GFP, Lane 3- pFI.GFP, Lane 4- pLb^{PRO}GFP, Lane 5- pHI.Lb^{PRO}GFP, Lane 6- pFI.Lb^{PRO}GFP, Lane 7- pLbGFP, Lane 8- pHI.LbGFP, Lane 9- pFI.LbGFP.

The observations reported above were confirmed by phosphorimaging analysis of the translation profiles. The ratio of the translation products from the FI.Lb^{PRO}GFP and HI.Lb^{PRO}GFP plasmids was calculated to be approximately 1: 2.5.

Plasmid	Protein Product	PSL-BG
pLb ^{pro} GFP	Lb ^{pro} GFP	-
	wtGFP	-
	Lb ^{pro}	-
pHI.Lb ^{pro} GFP	Lb ^{pro} GFP	410.30
	wtGFP	1245.00
	Lb ^{pro}	-
pFI.Lb ^{pro} GFP	Lb ^{pro} GFP	213.00
	wtGFP	442.30
	Lb ^{pro}	-
pLbGFP	LbGFP	95.31
pHI.LbGFP	LbGFP	151.10
pFI.LbGFP	LbGFP	2373.00

Table 4.8. Phosphorimaging analysis of the translation profiles of the Lb^{pro}GFP plasmid series in the STP3 TnT system.

4.2.4. In vitro analysis of the GFP2A^{pro} plasmid series using the Novagen STP3 system.

The STP3 transcription mix was programmed with plasmid DNA of the GFP2A^{pro} plasmid series pGFP2A^{pro}, pHI.GFP2A^{pro}, pFI.GFP2A^{pro}, pGFP2A, pHI.GFP2A and pFI.GFP2A. The reaction was incubated at 30°C for 15 minutes prior to the addition of the translation mixture. The translation reaction was incubated for a further 60 minutes at 30°C. The protein products were labelled with ³⁵S-methionine and separated by denaturing-PAGE. The translation products were visualized by autoradiography using a prolonged exposure period of 60 hours to detect any faint products.

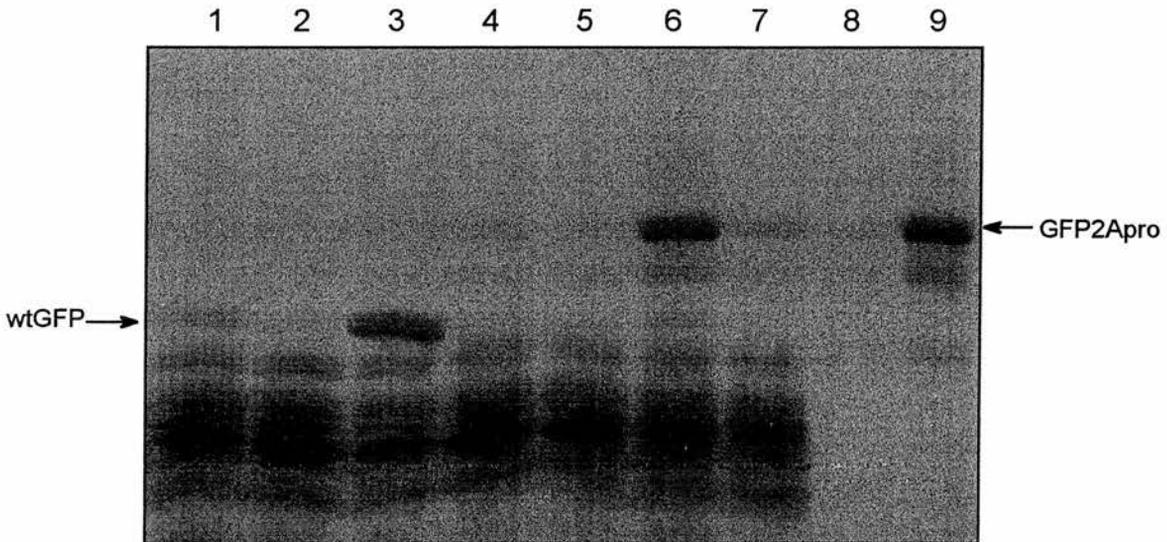


Fig.4.16. In vitro analysis of the GFP2A^{pro} plasmid series using the STP3 transcription and translation system from Novagen. Lane 1- pwtGFP, Lane 2- pHI.GFP, Lane 3- pFI.GFP, Lane 4- pGFP2A^{pro}, Lane 5- pHI.GFP2A^{pro}, Lane 6- pFI.GFP2A^{pro}, Lane 7- pGFP2A, Lane 8- pHI.GFP2A, Lane 9- pFI.GFP2A.

The same pattern of translation was seen for each set of plasmids, the uncapped and HRV14 IRES-driven transcripts were very poorly translated. In comparison the FMDV IRES transcripts were extremely efficiently translated. The plasmids pFI.GFP2A^{pro} and pFI.GFP2A produced a protein of approximately 45kDa that was identified as the full-

length GFP2A polyprotein. Comparison of the FI.GFP and FI.GFP2A^{pro} translation profiles identified a faint band in the GFP2A^{pro} translation corresponding to the wtGFP protein. The HRV14 2A^{pro} cleavage product could not be detected due to the high number of similar sized background bands. Identification of the wtGFP protein confirms that the GFP2A^{pro} polyprotein is processed in this system although this band was very faint indicating that the majority of GFP2A^{pro} remained uncleaved. These observations were confirmed by phosphorimaging analysis of the gel.

Plasmid	Protein Product	PSL-BG
pGFP2A ^{pro}	GFP2A ^{pro}	-
	wtGFP	-
	2A ^{pro}	-
pHI.GFP2A ^{pro}	GFP2A ^{pro}	-
	wtGFP	-
	2A ^{pro}	-
pFI.GFP2A ^{pro}	GFP2A ^{pro}	922.00
	wtGFP	118.10
	2A ^{pro}	-
pGFP2A	GFP2A	91.87
pHI.GFP2A	GFP2A	-
pFI.GFP2A	GFP2A	1821.00

Table 4.9. Phosphorimaging analysis of the GFP2A^{pro} plasmid series in the STP3 TnT system.

4.2.5. Conclusions.

The transcription and translation of the complete plasmid series in the Novagen STP3 system has produced dramatically different results from those reported in Part 1 of this chapter. Using this alternative system translation of uncapped transcripts was very inefficient for all plasmids tested. In the majority of cases this was also true for the HRV14 IRES transcripts, thus in accordance with the literature the HRV14 IRES was unable to initiate translation *in vitro* in rabbit reticulocyte lysate. The FMDV IRES however, was extremely efficient at initiating translation in rabbit reticulocyte lysate again this agreed with the results reported in the literature.

The one exception was the pLb^{PRO}GFP plasmid series where translation of the HRV14 IRES-driven transcript was 2.5 times more efficient than translation of the FMDV IRES-driven transcript. This improved HRV14 IRES translation was not observed for the pHI.LbGFP transcript indicating that the proteolytic activity of FMDV Lb^{PRO} was involved. The presence of an active HRV14 2A^{PRO} did not produce a similar enhancement of HRV14 IRES-driven translation suggesting that the activity responsible was not common to both proteases.

Therefore we can conclude that the hypothesis postulated in the introduction of this section that the IRES was not formed in the coupled system and that initiation of translation did not involve the IRES was correct. Therefore the tight coupled nature of the Promega TnT system prevents formation of RNA structure and as a result is not suitable for *in vitro* analysis of constructs where RNA secondary structure is important *i.e.* IRES formation.

The acceptance of the hypothesis that the IRES is not formed in the coupled transcription system requires an explanation for the discrepancy of translational efficiency between the uncapped and IRES-driven transcripts this can be simply explained by the presence of the

IRES. In the IRES-containing transcripts several hundred additional nucleotides including numerous AUG codons are present between the 5' end of the transcript and the protein initiation codon therefore the ribosomal machinery must scan along this region prior to initiating translation of the polyprotein.

A difference between the translational efficiency of the IRESes was also noted, the HRV14 IRES was slightly less efficient at initiating translation than the FMDV IRES. Examination of the sequences of the different IRESes found that the FMDV IRES contained four putative initiation codons and the HRV14 IRES contained twelve putative initiation codons. The computer program DNA Strider was used to predict the sizes of the proteins produced from these codons. All the predicted protein products were very small and would not be detected by the denaturing PAGE analysis. The mechanism employed to translate uncapped messages is not known but it would be expected that initiation would also occur at these sites this would account for the difference between the efficiency of the different IRES types.

4.2.6. *Summary.*

- Translation of uncapped transcripts was very poor in this system.
- The HRV14 IRES was not capable of initiating translation in RRL using this system.
- Translation from the FMDV IRES was extremely efficient.
- The Lb^{PRO}GFP polyprotein is efficiently processed in this system to produce wtGFP and Lb^{PRO}.
- The GFP2A^{PRO} polyprotein is processed to produce GFP but the majority remains uncleaved.
- The IRES does not form in the coupled system.
- The presence of the FMDV Lb^{PRO} enhances translation from the HRV14 IRES.
- The presence of the HRV14 2A^{PRO} does not alter translational efficiency.

CHAPTER 5: A STUDY OF ARTIFICIAL POLYPROTEIN PROCESSING IN PROKARYOTES.

5.1. Part 1: An investigation of the processing activity of artificial polyproteins containing picornaviral proteases.

5.1.1. Introduction.

Metabolic engineering of enzyme-catalyzed reactions within different systems is a rapidly expanding area of research. Metabolic control analysis of enzyme pathways has demonstrated that flux through a pathway is not limited by a single enzyme, therefore to successfully manipulate a pathway the simultaneous expression of multiple enzymes is required.

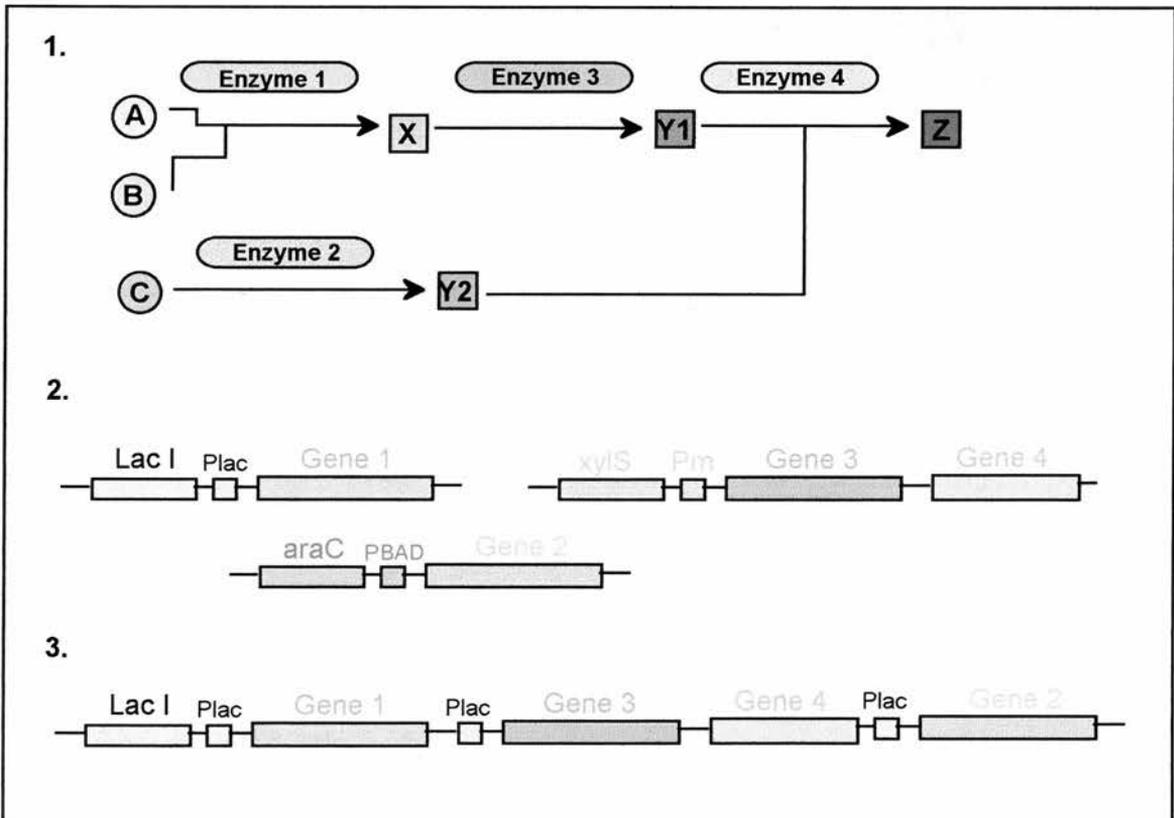


Fig.5.1. Co-ordinated expression of multiple genes encoding a hypothetical pathway (1). Use of three different inducible promoters to express the gene cassettes (2). Alternatively multiple copies of the same promoter can be used to express all the cassettes.

Current methods employed to coordinate the expression of multiple genes include the use of multiple inducible promoters (one per gene) leading to problems as several inducers must be added to the medium. Alternatively, the genes can be controlled by copies of the same inducible promoter this can also lead to problems with the expression levels of individual genes (Keasling, 1999). Thus to investigate the potential of using an artificial self-processing polyprotein system in prokaryotes to coordinate expression of multiple genes the artificial polyprotein were expressed in *E.coli*.

Translation of the artificial polyproteins *in vitro* had shown that the FMDV Lb^{PRO} and HRV14 2A^{PRO} were functional and processed the polyproteins *in cis*. No *trans* processing activity however, was observed in the *in vitro* system. Therefore the *trans* processing activity *in vivo* would be monitored in *E.coli* by the expression of the dual-reporter polyproteins.

5.1.2. Construction of prokaryotic expression vectors *pTrc.Lb^{PRO}BFP* and *pTrc.BFP2A^{PRO}*.

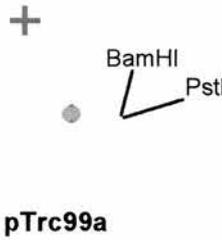
The Lb^{PRO}BFP and BFP2A^{PRO} gene cassettes were previously constructed in the pGEM transcription vector series and analyzed *in vitro*. The gene cassettes could, therefore, be subcloned into a suitable prokaryotic expression vector. The commercially available pTrc99a vector (Pharmacia) was selected. Expression of genes inserted into the multiple cloning site (MCS) from the strong *trc* promoter is induced by the addition of isopropyl- β -D-thiogalactoside (IPTG). Proteins are expressed from the MCS initiation codon therefore the proteins contain additional N-terminal amino acids.

The Lb^{PRO}BFP and BFP2A^{PRO} cassettes were excised from the plasmids pGEM-T.Lb^{PRO}BFP-stop and pGEM-T.BFP2A^{PRO} by digestion with BamHI and Pst I. The pTrc99a prokaryotic expression vector was similarly digested with BamHI and Pst I. The polyprotein-encoding cassettes were ligated into the MCS of the cut pTrc99a vector. The ligation mix was used to transform competent *E.coli* cells and plated onto antibiotic selective media. The plasmid DNA of putative clones was extracted and analytical restriction digests verified the insertion of Lb^{PRO}BFP and BFP2A^{PRO}. A sequencing primer was designed that annealed upstream of the multiple cloning site of pTrc99a to sequence the pTrc99a-based clones. The putative pTrc.Lb^{PRO}BFP and pTrc.BFP2A^{PRO} clones were submitted for automated DNA sequencing that confirmed the presence of the polyprotein cassettes.

1. Excise LbproBFP and BFP2Apro cassettes from pGemTLB and pGemTB2A plasmids by digestion with BamHI and PstI



2. Prepare pTrc99a vector by digestion with BamHI and PstI.



3. Ligate cut vector and gene cassettes together to produce pTrc.LbproBFP and pTrc.BFP2Apro.

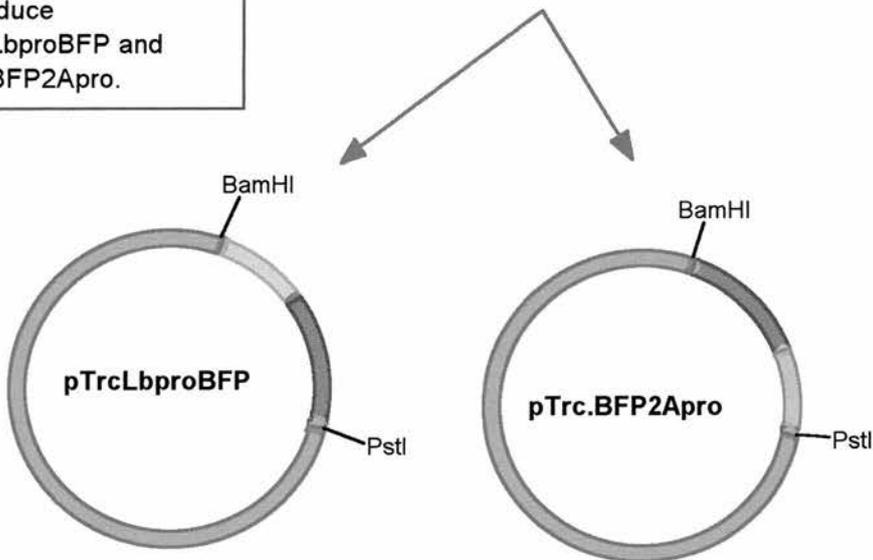


Fig.5.2. Construction of pTrc.Lb^{pro}BFP and pTrc.BFP2A^{pro}. The polyprotein gene cassettes are excised from the pGEM vector plasmids with BamHI and Pst I and ligated into similarly digested pTrc99a vector.

5.1.3. Construction of *pTrc.Lb^{PRO}BLG* and *pTrc.GLB2A^{PRO}*.

The *trans* processing polyprotein gene cassettes were constructed originally in the pGEM vector series. To create the prokaryotic expression vectors the cassettes were subcloned into the pTrc99a vector. The gene cassettes were excised by digestion with BamHI and Pst I. The insert was purified from agarose. The pTrc99a vector was similarly digested. The cut linearized vector and insert cassettes were ligated. The ligation mix was used to transform competent *E.coli* and plated onto antibiotic selective media. Putative clones were selected and the plasmid DNA was extracted. The presence of the gene cassettes was verified by analytical digests and automated DNA sequencing using the pTrc99a sequencing primer.

The expression of proteins from *pTrc.Lb^{PRO}BLG* and *pTrc.GLB2A^{PRO}* in *E.coli* in initial experiments could not be detected. Subsequent analysis of the *pTrc.Lb^{PRO}BLG* and *pTrc.GLB2A^{PRO}* plasmids showed that they were in fact the original pGEM clones. The initial sequencing and analytical digest checks on the putative clones both monitored the presence of the polyprotein gene cassette in the plasmid and did not monitor the vector backbone. The problem was caused by contamination of the purified insert with cut pGEM vector. The cut pGEM vector and the gene cassettes are both about 3kb thus it was very difficult to separate the bands by agarose gel electrophoresis. As a result the ligation mix contained the polyprotein insert and two alternative vectors *pTrc99a* and the original pGEM vectors. Unfortunately the selected clones were ligated pGEM vector and insert.

To avoid this problem in the subsequent cloning experiment the *Lb^{PRO}BLG* and *GLB2A^{PRO}* cassettes were excised from the plant vectors *PZP.Lb^{PRO}BLG* and *PZP.GLB2A^{PRO}* (refer to section 6.1.3.) as the cut vector (10kb) and insert (3kb) could be easily separated by gel electrophoresis. The polyprotein gene cassettes were excised by digestion with BamHI and PstI. The *pTrc99a* vector was similarly digested. The vector and insert were ligated to form *pTrc.Lb^{PRO}BLG* and *pTrc.GLB2A^{PRO}*.

Analytical digests were used to confirm the presence of the pTrc99a vector and the polyprotein gene cassettes. The sequence of the insert was verified via automated DNA sequencing using the pTrc99a sequencing primer.

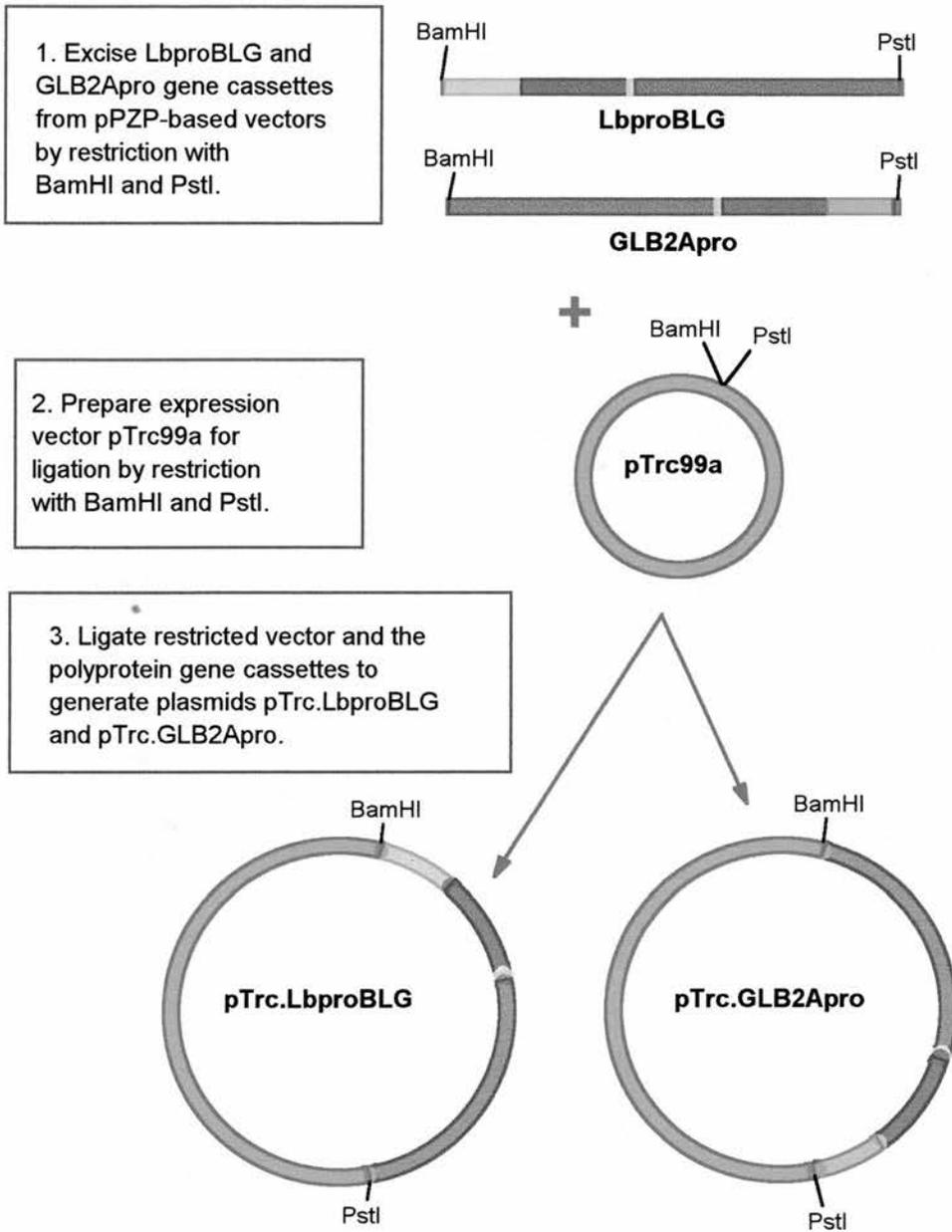


Fig.5.3. Construction of pTrc.Lb^{pro}BLG and pTrc.GLB2A^{pro}. The polyprotein gene cassettes were excised from the PZP plasmid series by digestion with BamHI and Pst I. The pTrc99a prokaryotic expression vector was similarly digested. The cut vector and gene cassettes were ligated forming pTrc.Lb^{pro}BLG and pTrc.GLB2A^{pro}.

5.1.4. A time course study of artificial polyprotein expression in Escherichia coli.

In order to investigate the processing of the artificial polyprotein system *in vivo* the polyprotein cassettes were expressed in the gram-negative bacteria *Escherichia coli*. The JM105 strain (*supE endA sbcB15 hsdR4 rpsL thiΔ(lac-proAB) F'[traD36 proAB⁺ lacI^q lacZΔM15]*) was selected for use as recommended by the supplier of the expression vector pTrc99a. Competent JM105 cells were prepared and transformed with the plasmid DNA. The transformed cells were plated onto antibiotic selective media. Antibiotic selective LB broth was inoculated with a single colony. A negative control of non-transformed JM105 cells was also prepared. The cultures were grown at 37°C for 3 hours prior to induction of protein expression with IPTG. Following induction cultures were incubated at 30°C. Cells were harvested at 1h, 2h, 3h, 4h, 5h and 16h post-induction. The cultures were pelleted by centrifugation and resuspended in 1xSDS protein loading buffer. The total cell fractions for non-induced cells and induced cells were boiled and separated by 10% PAGE. The proteins were stained with Coomassie brilliant blue.

Strong bands correlating to the polyprotein expression products could not be identified on the denaturing-PAGE gel. Therefore the post-induction incubation period for use in subsequent experiments was selected as 5h. Extending the incubation period past 5h did not significantly increase expression levels and this time-period was convenient for the working day.

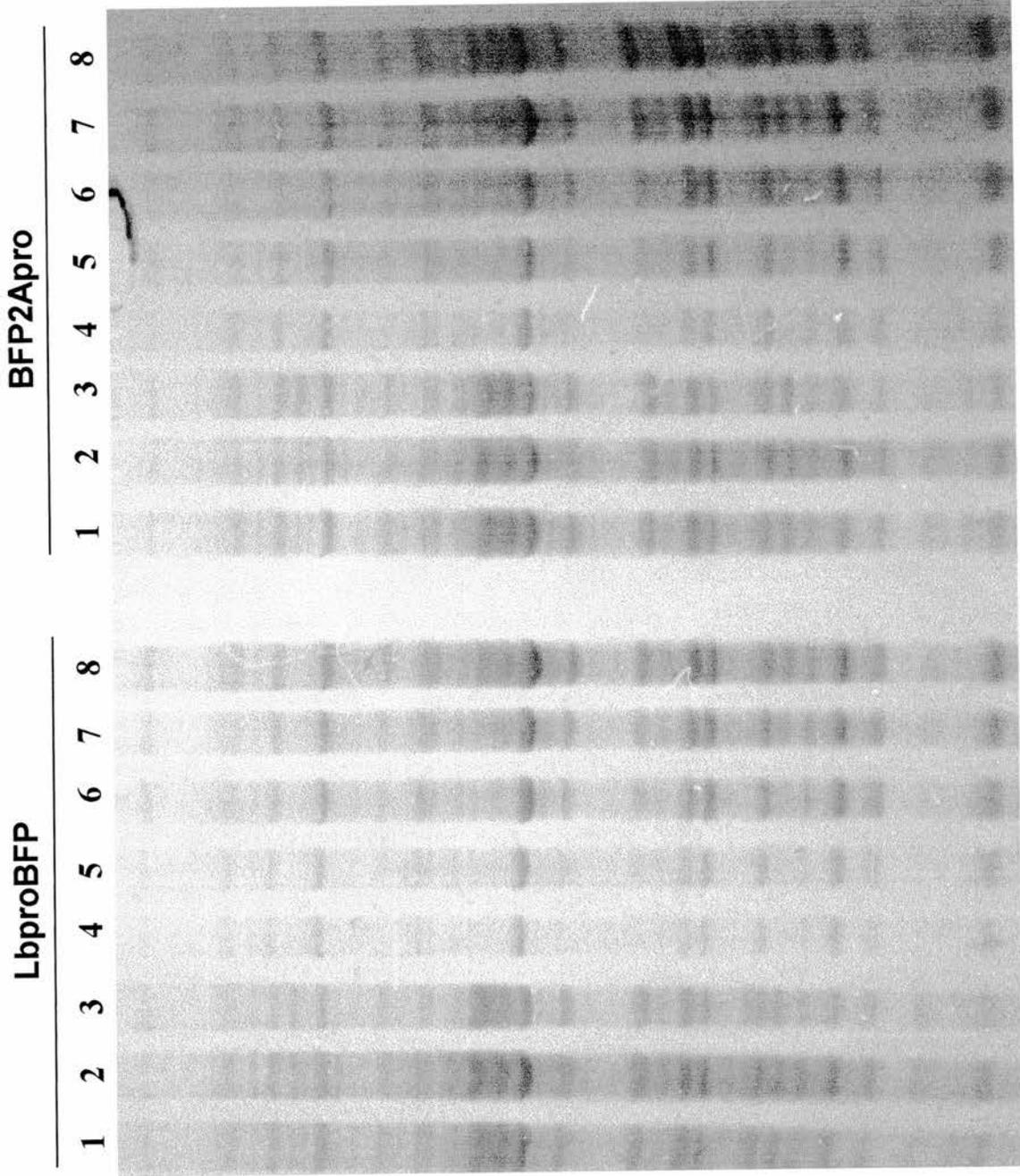


Fig.5.4. Time course of LbproBFP and BFP2Apro polyproteins in *E. coli*. Lane 1- non-induced JM105, Lane 2- IPTG induced pTrc99a, Lane 3- noninduced polyprotein, Lane 4- IPTG induced +1h, Lane 5- IPTG induced +2h, Lane 6- IPTG induced +3h, Lane 7- IPTG induced +4h, Lane 8- IPTG induced +8h.

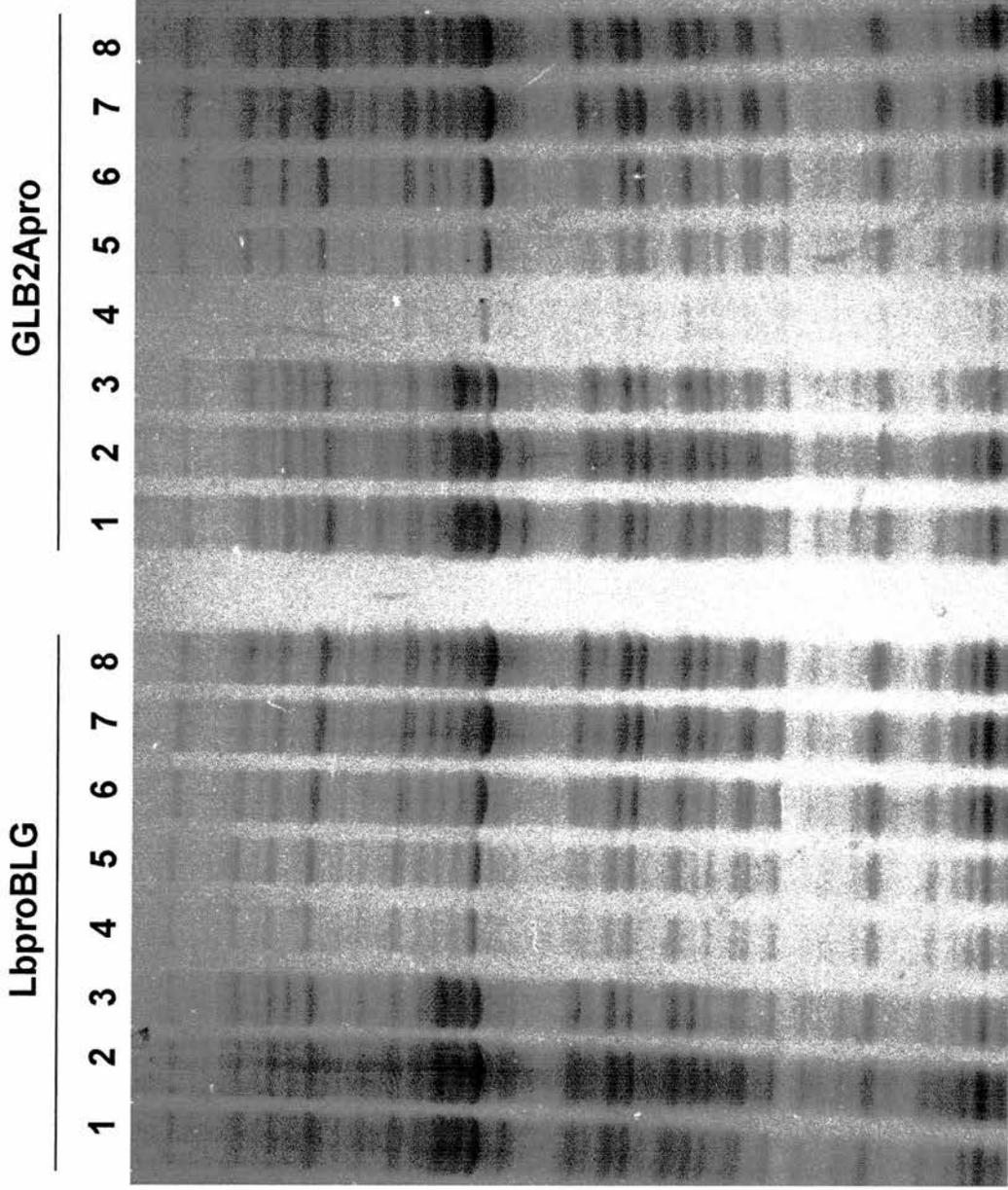


Fig.5.5. Time course of LbproBLG and GLB2Apro polyproteins in JM105 E.coli cells. Lane 1- non-induced JM105, Lane 2- IPTG induced pTrc99a, Lane 3- non-induced polyprotein, Lane 4-IPTG induced +1h, Lane 5- IPTG induced +2h, Lane 6-IPTG induced +3h, Lane 7- IPTG induced +4h, Lane 8-IPTG induced +5h.

5.1.5 Western blot analysis of protein expression from *pTrc.Lb^{pro}BFP* and *pTrc.BFP2A^{pro}*.

Additional bands corresponding to the polyprotein products could not be readily identified by denaturing-PAGE analysis thus western blot analysis as outlined in section 2.5.8. was performed to detect the polyprotein products. The proteins were separated by 11% PAGE and transferred to a nitrocellulose membrane. The membrane was incubated in 5% blocking solution then probed with 1:1000 dilution of polyclonal GFP antibody (Clontech). The membrane was washed and probed with 1:5000 dilution of the HRP-linked anti-rabbit antibody (Amersham). The proteins were visualized by chemiluminescence and autoradiography.

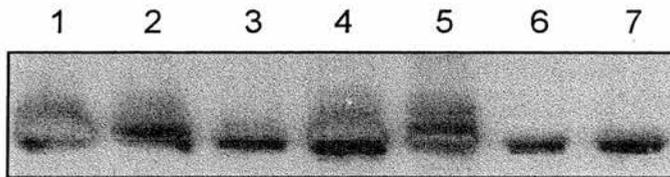


Fig.5.6. Western blot probing with polyclonal GFP antiserum. Protein expression was induced in all samples by the addition of IPTG. Lane 1- *pTrc99a*, Lane 2- *peBFP*, Lane 3- *pRAJ 275*, Lane 4- *pTrc.Lb^{pro}BFP*, Lane 5- *pTrcBFP2A^{pro}*, Lane 6- *pTrc.Lb^{pro}BLG*, Lane 7- *pTrc.GLB2A^{pro}*.

The polyclonal GFP antibody cross-reacted with a protein in all the samples. Unfortunately this was a similar size (about 30kDa) to the predicted BFP protein thus it was difficult to ascertain the presence of authentic BFP protein. Manipulation of the antibody concentrations and blocking conditions failed to alleviate the cross-reactivity of the antibody to this protein therefore the decision was made to use an alternative GFP antibody. The living colours peptide antibody from Clontech was selected as it is highly specific for all GFP variants including eBFP and the blot was repeated.

The conjugated-peptide BFP antibody was used at a 1:200 dilution as instructed by the manufacturer. The problematic background band that was detected with the polyclonal

GFP antibody was not present on blots probed with the conjugated-peptide antibody. Western blot analysis of the total cell fraction of pTrc.Lb^{PRO}BFP and pTrc.BFP2A^{PRO} with BFP antiserum identified a single band of about 30kDa corresponding to cleaved BFP. Larger proteins corresponding to the uncleaved polyprotein were not detected thus the Lb^{PRO}BFP and BFP2A^{PRO} polyproteins are completely processed *in vivo*. Antibodies to the picornaviral proteases FMDV Lb^{PRO} and HRV14 2A^{PRO} were not available thus we were unable to directly confirm the presence of the processed proteases.

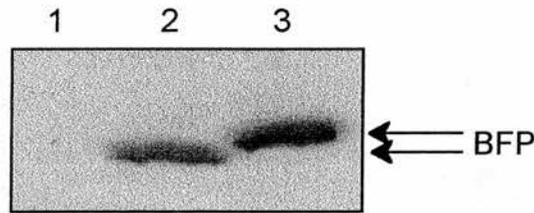


Fig.5.7. Western blot of Lb^{PRO}BFP and BFP2A^{PRO} with BFP antiserum. Protein expression in *E.coli* was induced by the addition of IPTG. Lane 1- pTrc99a control, Lane 2 -pTrc.Lb^{PRO}BFP, Lane 3 - pTrc.BFP2A^{PRO}.

5.1.6. Western blot analysis of protein expressed from pTrc.Lb^{PRO}BLG.

Western blot analysis of pTrc.Lb^{PRO}BFP confirmed that the FMDV Lb protease is functional in *E.coli*. The Lb^{PRO}BFP polyprotein was processed completely *in cis* to yield the reporter gene BFP and the FMDV Lb protease. The dual reporter polyprotein Lb^{PRO}BLG was expressed in *E.coli* to monitor the *trans* processing activity of FMDV Lb^{PRO} in prokaryotes. The total cellular fraction was separated by denaturing-PAGE and transferred to nitrocellulose membrane. Antiserum for both reporter genes was commercially available thus by duplicating the series of lanes on the gel and bisecting the membrane blots could be probed with BFP and GUS antisera and the results compared.

To aid preliminary identification of the polyprotein expression products a positive control for BFP and GUS were included on the blots. For the BFP control recombinant GFP protein from yeast was available in the laboratory. Recombinant GUS was not available therefore the original GUS vector pRAJ 275 (Clontech) was expressed in *E.coli* and the total cell fraction was included on the anti-GUS blots.

Western blot analysis with BFP antiserum was performed using the conditions outlined above. The conditions for probing with GUS antiserum had to be optimized. The GUS antibody from 5 Prime ->3 Prime cross-reacted very strongly with proteins in the pTrc99a expressing cells thus the background reactivity was high and it was impossible to identify the authentic GUS containing products. The primary and secondary antibodies were initially used at 1:5000 dilution in 5% blocking solution. Varying the concentrations of either antibody affected all the bands in the same way thus removal of the background also removed any authentic bands.

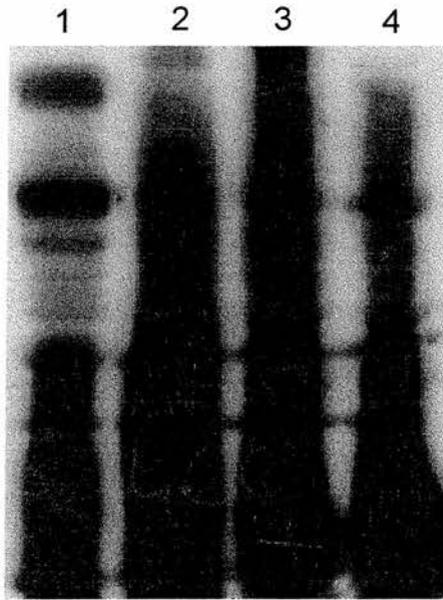


Fig. 5.8. Western blot analysis with GUS antibody using 5% milk blocking solution. Protein expression was induced by the addition of IPTG. Lane 1- pRAJ 275, Lane 2- pTrc99a, Lane 3- pTrc.Lb^{PRO}BLG, Lane 4- pTrc.GLB2A^{PRO}.

The background “noise” was finally removed by improving the blocking conditions. The concentration of the blocking solution was increased to 10% milk (w/v). The blot was blocked prior to probing with the antibodies in 10% blocking solution and to ensure blocking was maintained the antibodies were diluted in the same 10% blocking solution. The primary antibody concentration was 1:3000 and the secondary antibody concentration was 1: 4000.

Probing with anti-BFP antibody identified two bands from Lb^{PRO}BLG corresponding to cleaved BFP and the initial *cis* cleavage product BFPlinkerGUS. Western blot analysis of pTrc.Lb^{PRO}BLG with anti-GUS antibody identified the cleaved GUS product. A faint band corresponding to the larger *cis* cleavage product BFPlinkerGUS was also detected.

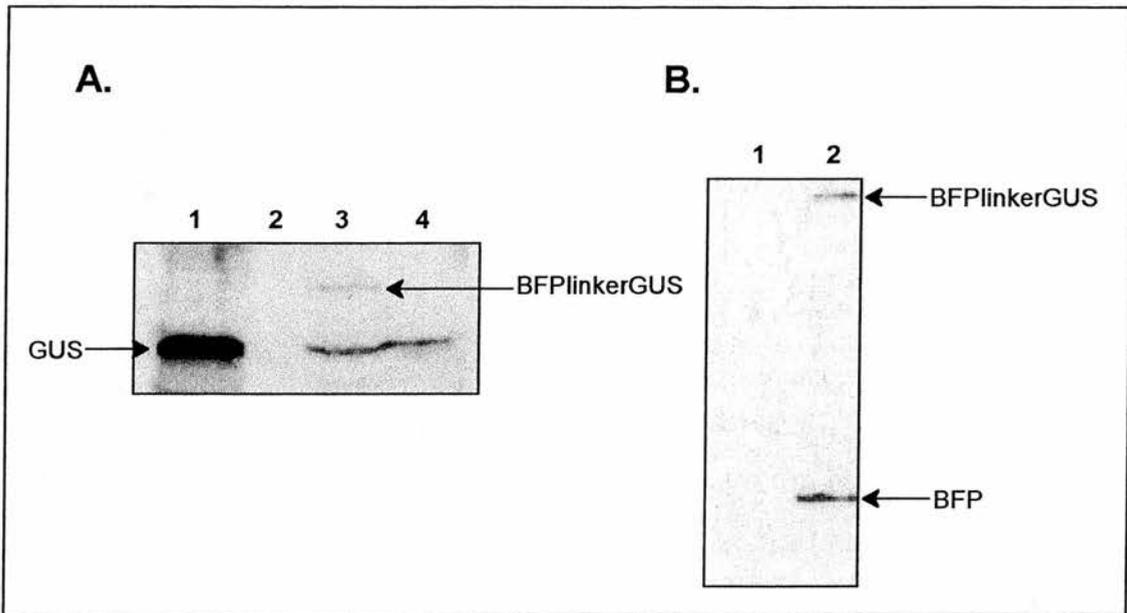


Fig. 5.9. Western blot analysis of pTrc.Lb^{PRO}BLG. Protein expression was induced by the addition of IPTG. (A) Probing with GUS antiserum, Lane 1- pRAJ 275, Lane 2- pTrc99a, Lane 3- pTrc.Lb^{PRO}BLG, Lane 4- pTrc.GLB2A^{PRO}.

Thus Western blot analysis confirms that the full-length polyprotein is expressed in prokaryotes as both reporter proteins were present. The full-length polyprotein Lb^{PRO}BLG was not detected therefore the *cis* cleavage by FMDV Lb^{PRO} at the Lb^{PRO}/BFP junction is 100 percent efficient *in vivo* in *E.coli*. The FMDV Lb protease processed the majority of the *cis* cleavage product BFPlinkerGUS *in trans* at the second FMDV L-1A cleavage site in the linker region. The polyprotein cleavage efficiency *in trans* could not be calculated in the *E.coli* as western blotting is a qualitative rather than a quantitative method.

5.1.7. Western blot analysis of proteins expressed from pTrc.GLB2A^{PRO}.

The western blot analysis of pTrc.BFP2A^{PRO} confirmed that HRV14 2A^{PRO} expressed in *E.coli* is functional. The BFP2A^{PRO} artificial polyprotein was processed *in cis* at the BFP/2A junction. The GLB2A^{PRO} dual reporter polyprotein was expressed in *E.coli* and analyzed by western blot using the conditions optimized for the pTrc.Lb^{PRO}BLG expression products to monitor the *trans* processing activity of HRV14 2A^{PRO} *in vivo*. The total cellular proteins of induced pTrc.GLB2A^{PRO} were separated by 11%

PAGE and transferred to a nitrocellulose membrane. The blots were probed with BFP and GUS antisera. Proteins bound by the antiserum were visualized using chemiluminescence and autoradiography.

Western blotting of the pTrc.GLB2A^{PRO} sample with BFP antiserum identified a single band corresponding to the BFP cleavage product. Western blot analysis with GUS antiserum also identified a single protein in the pTrc.GLB2A^{PRO} sample. Comparison of this band with the control GUS product from pRAJ 275 confirmed that this was the GUS cleavage product. The larger full-length polyprotein or the *cis* processed GUSlinkerBFP polyprotein were not detected in either blot. Therefore these results confirm that the complete polyprotein GLB2A^{PRO} is produced in *E.coli* as both reporter proteins BFP and GUS are present. The GLB2A^{PRO} polyprotein produced in *E.coli* is completely processed by HRV14 2A^{PRO} *in cis* at the 2A/BFP junction and *in trans* at the duplicated rabbit eIF4G cleavage site as no larger polyprotein precursors were detected.

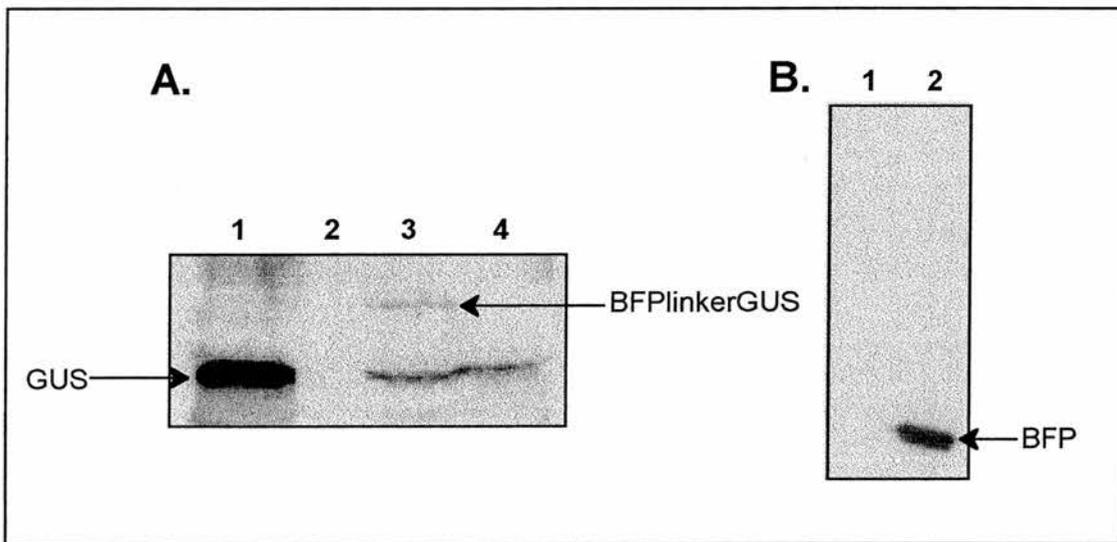


Fig.5.10. Western blot analysis of pTrc.GLB2A^{PRO}. Protein expression was induced by the addition of IPTG. (A) Probing with GUS antiserum, Lane 1- pRAJ 275, Lane 2- pTrc99a, Lane 3- pTrc.Lb^{PRO}BLG, Lane 4- pTrc.GLB2A^{PRO}. (B) Probing with BFP antiserum, Lane 1- pTrc99a, Lane 2- pTrc.GLB2A^{PRO}.

5.1.8. Conclusions

A series of prokaryotic expression vectors to express artificial self-processing polyproteins have been constructed. Therefore the artificial polyproteins can be monitored *in vivo* in the gram-negative bacteria *Escherichia coli*. Full-length polyproteins were successfully produced in the prokaryotic system. Production of the picornaviral proteases FMDV Lb^{PRO} and HRV14 2A^{PRO} was tolerated in *E.coli* cells.

The proteases were correctly synthesized and folded as both expressed proteases were demonstrated to be functional. In *E.coli* the full-length polyproteins were not detected thus the artificial polyproteins were completely processed *in cis* at the protease/BFP cleavage site.

Larger quantities of polyprotein were expressed in *E.coli* therefore *trans* processing of the artificial polyproteins could be monitored for the dual-reporter protein constructs Lb^{PRO}BLG and GLB2A^{PRO}.

Western blot analysis of the expression products of Lb^{PRO}BLG with BFP and GUS antibodies confirmed that the polyprotein was processed at the *trans* cleavage site contained within the linker region. Although the majority of the polyprotein was completely processed to yield BFP, GUS and Lb^{PRO} a small amount of the primary cleavage product BFPlinkerGUS was present. Therefore the polyprotein is completely processed by FMDV Lb^{PRO} *in cis* but *trans* processing of the second L/1A cleavage site is not as efficient. Western blot analysis of the expression products of pTrc.GLB2A^{PRO} with BFP and GUS antibodies only detected the final cleavage products BFP and GUS. In this case larger processing products were not present. Therefore the 1D/2A *cis* cleavage site and the *trans* rabbit eIF4G cleavage site were completely processed by HRV14 2A^{PRO} in *E.coli*.

In conclusion, the self-processing artificial polyprotein system can be used in prokaryotic systems to express multiple genes from a single promoter. The polyproteins are fully processed by the picornaviral proteases.

5.1.9. Summary

- The full-length artificial polyproteins Lb^{PRO}BFP, BFP2A^{PRO}, Lb^{PRO}BLG and GLB2A^{PRO} are expressed in the gram-negative bacteria *E. coli*.
- The picornaviral proteases FMDV Lb^{PRO} and HRV14 2A^{PRO} are active in the prokaryotic environment.
- The artificial polyproteins are completely processed *in cis* by the FMDV Lb^{PRO} and HRV14 2A^{PRO}.
- The majority of the Lb^{PRO}BLG polyprotein is processed *in trans* at the second L/1A cleavage site introduced within the linker region.
- The GLB2A^{PRO} polyprotein is completely processed *in trans* at the rabbit eIF4G cleavage site introduced in the linker region separating GUS and BFP.

5.2. Part 2: An investigation into the effect of an N- or C-terminal extension on β -glucuronidase activity in *Escherichia coli*.

5.2.1. Introduction.

The first stage in the prokaryotic expression of the artificial polyproteins was the transformation of competent JM105 *E.coli*. The transformed bacteria were selected by plating onto antibiotic selective media. Although β -glucuronidase is an *E.coli* gene it was reported that high GUS-expression from a plasmid could be detected by comparing the relative colour intensity of the colony on plates containing a suitable β -glucuronidase substrate (Singh and Sharma, 1991). Intensely blue colonies were expressing high levels of GUS from the plasmid. Therefore it was proposed that this method could be used to select colonies expressing the highest amounts of polyprotein for growth in culture.

Competent JM105 *E.coli* cells were transformed with plasmid DNA and plated onto selective media containing antibiotic, IPTG and glucuronidase substrate. It was observed that pTrc.Lb^{PRO}BLG transformed cells produced a range of colonies from white to deep blue when grown on glucuronidase substrate media. In contrast, the pTrc.GLB2A^{PRO} transformed cells produced white/pale blue colonies consistent with the non-transformed JM105 *E.coli* cells. Western blot analysis of a sample of pTrc.GLB2A^{PRO} confirmed the presence of cleaved GUS in the cells. The reason why GUS from the Lb^{PRO}BLG polyprotein was active but GUS from GLB2A^{PRO} was inactive was not clear. Apart from the presence of the different proteases the components of the polyproteins were the same simply in a different order. The simplest explanation was the possibility that GUS was inactivated by a C-terminal addition. Perusal of the current literature failed to provide examples of C-terminal GUS fusions. GUS is predominantly used as a reporter gene and is usually fused to the N-terminus of the gene of interest. Thus a simple experiment was designed to investigate this question.

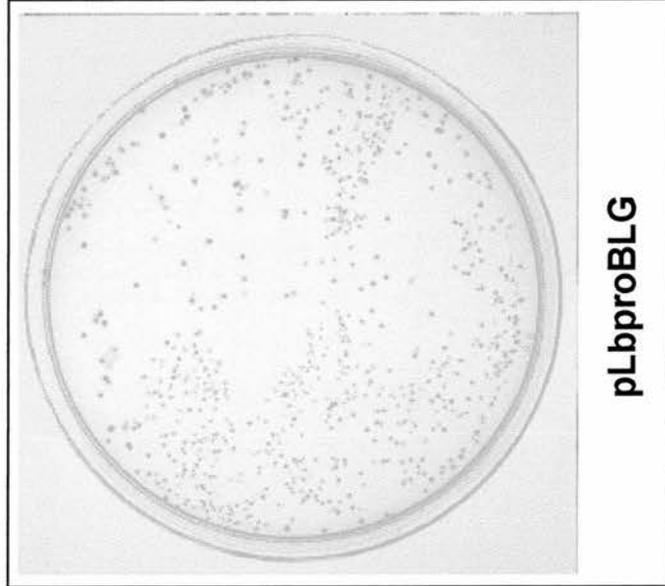
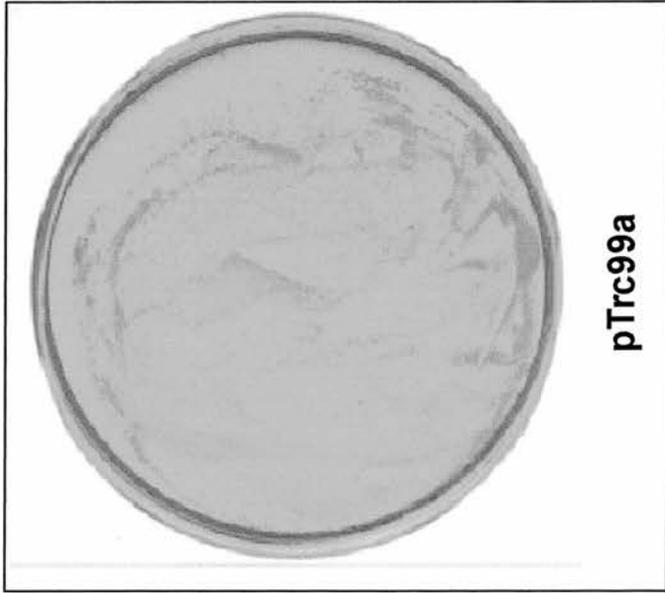


Fig.5.11. Monitoring GUS expression in JM105 *E. coli* cells. Control plate pTrc99a for background native GUS expression in JM105 *E. coli*. Plate pLbproBLG- GUS expression from LbproBLG polyprotein in *E. coli*. Plate GLB2Apro- GUS expression from GLB2Apro polyprotein in *E. coli*.

Alternative prokaryotic expression plasmids containing fusion proteins with GUS added to the N- or C-terminus were required. A suitable prokaryotic expression plasmid pUC:CAT2AGUS containing the N-terminal GUS fusion protein CAT2AGUS (where 2A is FMDV 2A) had been previously constructed in the laboratory by Susan Monaghan. A similar prokaryotic expression plasmid containing a GUS C-terminal fusion protein was not available and had to be constructed.

2.2.2. Construction of the prokaryotic expression plasmid pTrc.GUS2AGFP.

Moving the GUS2AGFP construct from pGUS2AGFP (created by Michelle Donnelly) into the pTrc99a expression vector produced the second plasmid. The GUS2AGFP cassette was cloned into pTrc99a so that GUS2AGFP was in frame with the pTrc99a initiation codon. The GUS2AGFP cassette was excised by digestion with BamHI and Pst I. The pTrc99a vector was similarly digested and the cut vector and insert were ligated together. Competent JM109 *E.coli* cells were transformed with the ligation mix and plated onto antibiotic selective media. Plasmid DNA was prepared from putative clones and analyzed by restriction digests. Putative clones were confirmed by automated sequencing using the pTrc99a sequencing primer.

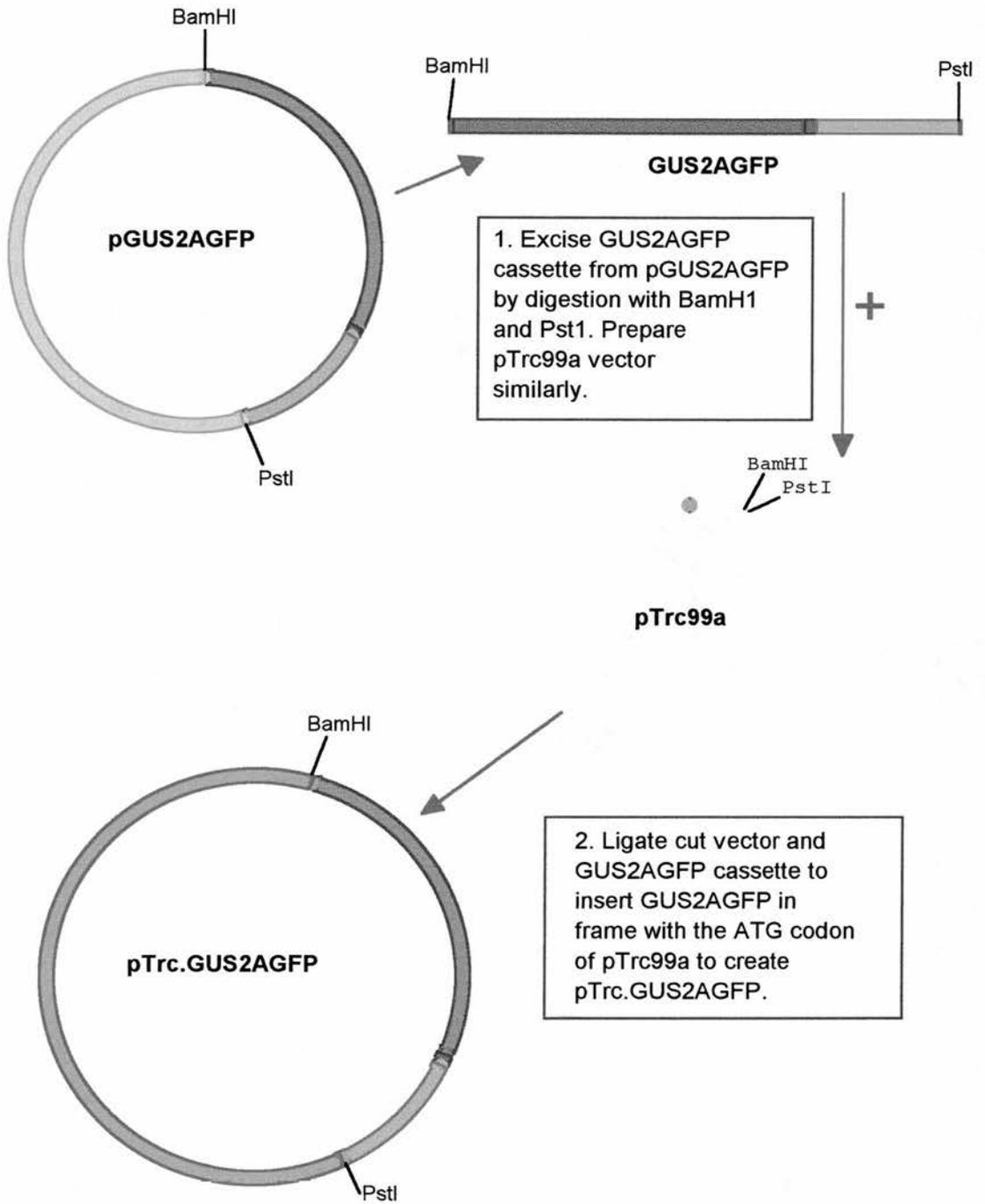


Fig.5.12. Construction of pTrc.GUS2AGFP. The GUS2AGFP cassette was excised from pGUS2AGFP by BamHI and Pst I digestion and inserted into the pTrc99a multiple cloning site.

5.2.3. The β -glucuronidase activity of N- and C-terminal fusion proteins in *E.coli*.

To monitor if β -glucuronidase activity is affected by the presence of an N- or C-terminal extension competent JM105 *E.coli* were transformed with the control vector pTrc99a and pUC:CAT2AGUS and pTrc.GUS2AGFP. The transformed cells were plated onto antibiotic selective media containing IPTG to induce expression and a β -glucuronidase substrate. The plates were incubated at 37°C for 16h to allow the colonies to grow. The plates were incubated at 4°C for 24h to allow the blue colour to develop.

Colonies transformed with pUC:CAT2AGUS and pTrc.GUS2AGFP ranged from white to bright blue. Therefore the expressed GUS protein in both fusion proteins was functional. In this experiment GUS was active with either an N- or C-terminal extension.

5.2.4. Investigation of the cleavage activity of FMDV 2A in *E.coli*.

Previous studies in our laboratory by Susan Monaghan and Michelle Donnelly had investigated the activity of FMDV 2A in prokaryotes. In the initial study pUC:CAT2AGUS was expressed in *E.coli* and analyzed by Western blotting with GUS antiserum. This confirmed the presence of the full-length uncleaved polyprotein CAT2AGUS but no GUS cleavage product was detected. Therefore it was deduced that FMDV 2A is inactive in *E.coli*. No attempt was made to probe for the CAT2A cleavage product.

Subsequent site-directed mutagenesis of the FMDV 2A region by Michelle Donnelly *in vitro* demonstrated that premature termination of CAT2AGUS can occur and the CAT2A protein can be produced in the absence of GUS. The expression of CAT2AGUS in *E.coli* was repeated and the blot was probed for the CAT2A protein. The presence of CAT2A was not detected. Similar results were produced using an alternative polyprotein GFP2AGUS. Thus it was concluded that FMDV 2A was

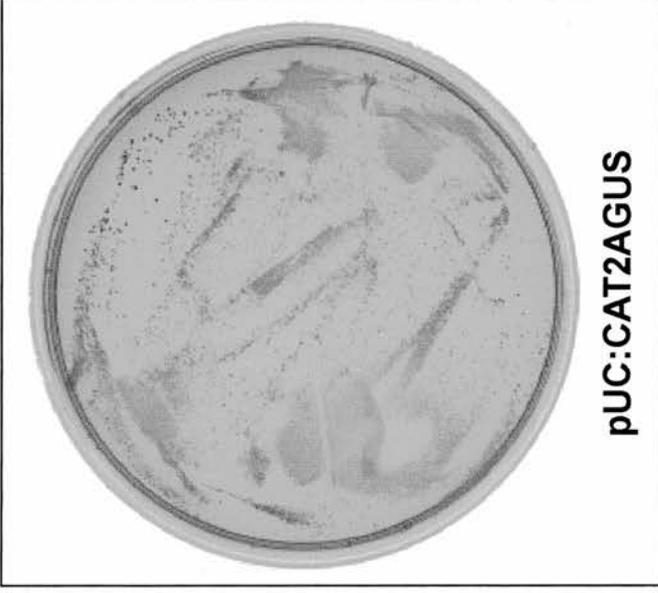
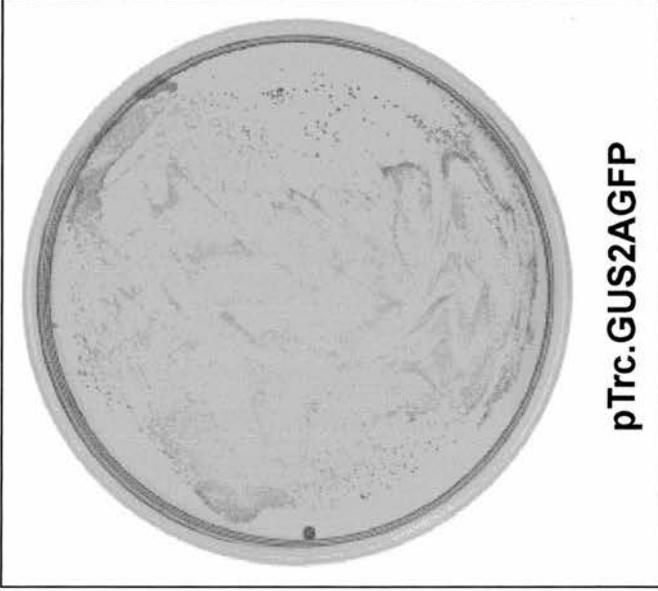
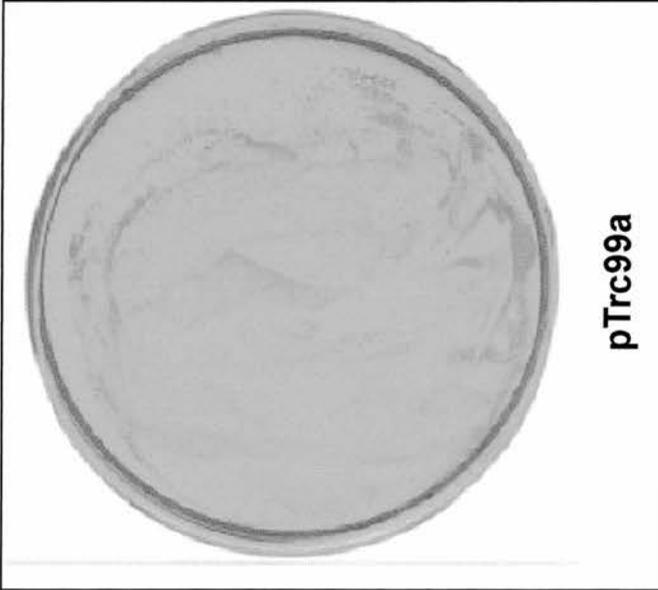


Fig.5.13. Monitoring GUS expression in FMDV 2A polyproteins. Control plate pTrc99a for background native GUS expression in JM105 *E. coli* cells. Plate pTrc.GUS2AGFP- GUS expression from GUS2AGFP polyprotein. Plate pUC:CAT2AGUS- GUS expression from CAT2AGUS polyprotein.

inactive in prokaryotes this was reported by Donnelly *et al.* (1997).

The antisera currently available within the laboratory were proposed to be more sensitive than previous antibodies. Therefore as CAT2AGUS and GUS2AGFP had already been expressed in *E.coli* to monitor the activity of GUS the total cell fractions were analyzed by western blotting.

5.2.5. Western blot analysis of CAT2AGUS and GUS2AGFP expressed in *E.coli*.

The proteins expressed from pUC:CAT2AGUS and pTrc.GUS2AGFP were separated by 11% PAGE and transferred to nitrocellulose membrane. The pTrc.GUS2AGFP blot was probed with GUS and GFP antisera using the conditions optimized for pTrc.Lb^{PRO}BLG. Probing with GFP antiserum identified a single protein correlating to full-length uncleaved GUS2AGFP. Probing with GUS antiserum identified the uncleaved polyprotein and a shorter product correlating to GUS2A. Therefore premature termination of the GUS2AGFP polyprotein occurs at the C-terminus of FMDV 2A.

The proteins expressed from pUC:CAT2AGUS were probed with GUS and CAT antisera using the conditions optimized for anti-GUS western analysis of pTrc.Lb^{PRO}BLG. Probing with GUS antiserum identified a single protein correlating to uncleaved CAT2AGUS. No GUS cleavage product was detected. Probing with anti-CAT antiserum confirmed that this protein was uncleaved CAT2AGUS. A second protein not present in the control sample of approximately 30kDa was identified as the CAT2A protein.

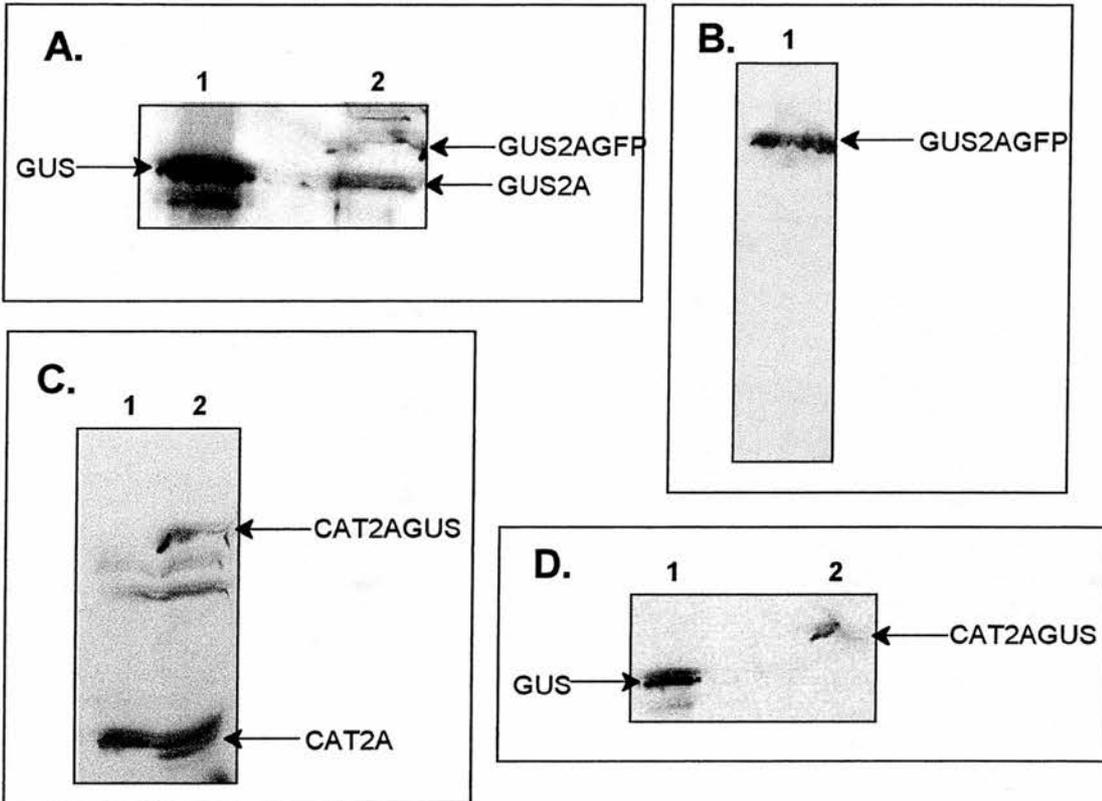


Fig.5.14. Western blot analysis of FMDV 2A polyprotein expressed in *E.coli*. (A) Probing with GUS antiserum, Lane 1- pRAJ 275, Lane 2- pTrc.GUS2AGFP. (B) Probing pTrc.GUS2AGFP with GFP antiserum. (C) Probing with CAT antiserum, Lane 1- pTrc99a, Lane 2- pUC:CAT2AGUS. (D) Probing with GUS antiserum, Lane 1 – pRAJ 275, Lane 2- pUC:CAT2AGUS.

5.2.6. Conclusions.

The expression of CAT2AGUS and GUS2AGFP in *E.coli* in the presence of a glucuronidase substrate confirms that the reporter protein β -glucuronidase is active in the presence of an N- or C-terminal extension.

Western blotting analysis of GUS2AGFP and CAT2AGUS indicates that FMDV 2A is inactive in *E.coli* as no second cleavage product is detected. The presence of the first cleavage product suggests that in accordance with the initial site-directed mutagenesis studies by Michelle Donnelly premature termination of the polyprotein can occur at the C-terminus of FMDV 2A.

5.2.7. Summary

- The reporter gene GUS is active with the addition of an N- or C-terminal extension.
- FMDV 2A has no cleavage activity in *E.coli*.
- Premature termination can occur at the C-terminus of FMDV 2A in *E.coli*.

CHAPTER 6: AN INVESTIGATION INTO THE POTENTIAL OF THE ARTIFICIAL SELF-PROCESSING POLYPROTEIN SYSTEM *IN PLANTA*.

6.1. Part 1: Construction of the binary vectors and analysis of the artificial self-processing dual reporter system *in planta*.

6.1.1. Introduction.

Plant biotechnology is a rapidly expanding area of research ranging from basic manipulation of native plant pathways to enhance specific traits to the more complex and controversial uses of plants as bioreactors to produce, for example, commercially important fatty acids and biodegradable polymers. The introduction of novel pathways composed of multiple genes into the plant system can be achieved via more traditional breeding methods by crossing parent plants each containing a single enzyme from the pathway and selecting the progeny containing both genes. This is a lengthy process for even the simplest plant species but for many target species such as trees that take many years to reach sexual maturity it is not a viable option.

The alternative method using plant transformation to introduce multiple genes into a single plant would involve many steps to create a plant expressing all the genes required. Even at this stage due to the nature of plant transformation and the lack of control over where the genes integrate into the plant chromosome there would be difficulties associated with the expression levels of the enzymes in the pathway. Thus a method that would allow the introduction of multiple genes in one step that are controlled by the same promoter is extremely attractive. The use of an artificial self-processing polyprotein would fulfill these requirements. This has already been tested using the novel cleavage mechanism of FMDV 2A protein to process the polyprotein (Halpin *et al.*, 1999; Santa Cruz *et al.*, 1996). A potential problem of using the FMDV 2A system is that 2A remains attached to the upstream protein thus 2A may potentially interfere with the function of the attached 3' terminus protein. The FMDV Lb^{PRO} and

HRV14 2A^{pro} that are used in the system described in this thesis process all the junctions within the polyprotein thus processing the polyprotein completely. Therefore the self-processing dual-reporter polyproteins were used to transform *Nicotiana tabacum* plants to investigate the potential use of this type of artificial system within plants.

The method of plant transformation selected was *Agrobacterium*-mediated transformation therefore the plasmids had to be transferred into a suitable strain of *Agrobacterium*. The compact binary vector pPZP111 (Hajdukiewicz, Svab and Maliga, 1994) was selected for use. Dr. A. Barakate adapted the pPZP111 vector prior to the introduction of the gene cassettes by excising the region containing the CaMV promoter-MCS-3'NOS sequence from the binary vector pJRIRi by restriction with EcoRI and HindIII. The excised region was inserted between the EcoRI and HindIII sites within the MCS of the pPZP111 vector.

6.1.2. Construction of the control binary vector PZP.BFP.

A control binary vector containing eBFP was required as this GFP variant had not been previously expressed in *Nicotiana tabacum*. Therefore transformation of plants with the control vector would identify any problems associated with the presence of the eBFP reporter protein in a plant system. The eBFP coding region was transferred from the peBFP plasmid (Clontech) by restriction with BamHI and Xba I. The plasmid was restricted with BamHI first due to the presence of two Xba I sites within the MCS of peBFP. The pPZP111 binary vector was similarly restricted with BamHI and Xba I. The eBFP fragment was ligated into the restricted pPZP111 to create the control binary vector pPZP.BFP. 100µl aliquots of high efficiency JM109 *E.coli* (Promega) were transformed with the ligation reactions and plated onto LB agar plates containing 25µg/ml of chloramphenicol. The plasmid DNA from putative colonies was extracted and analyzed by restriction digests to identify positive colonies. The presence of the BFP insert from positive clones was also verified by the PCR.

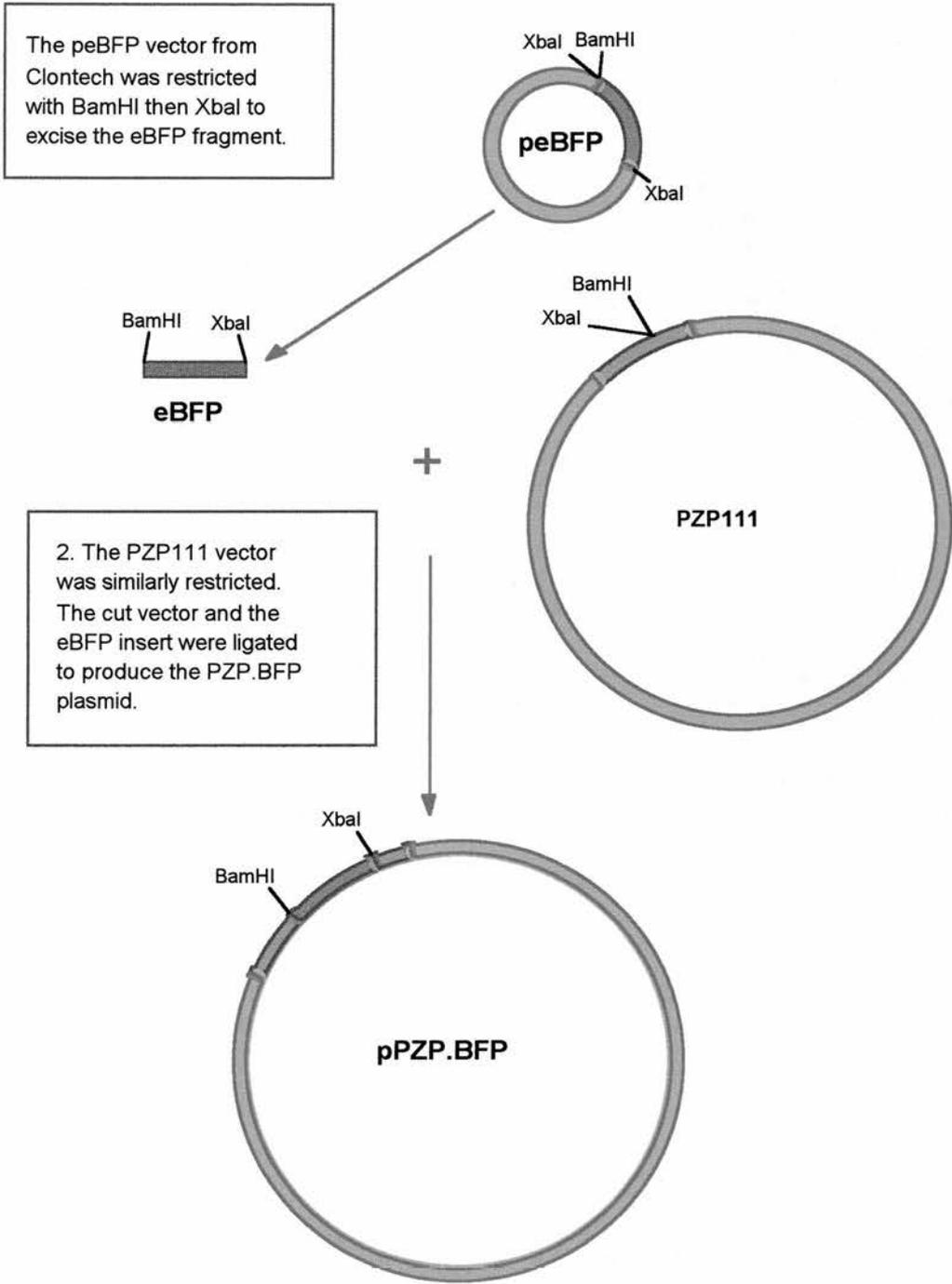


Fig.6.1. Construction of the control binary vector pPZP.BFP. The BFP coding region was excised from the peBFP vector by restriction with BamHI and Xba I and ligated into the similarly restricted pPZP111 vector.

6.1.3. Construction of the artificial polyprotein binary vectors pPZP.Lb^{PRO}BLG and pPZP.GLB2A^{PRO}.

The artificial dual-reporter polyprotein cassettes had been originally constructed in the pGEM vector series and analyzed *in vitro* (refer to section 3.2). In order to investigate the expression of the artificial polyproteins within a plant system the polyprotein encoding region had to be transferred into a suitable binary vector. The polyprotein cassettes were excised from the original pGEM vector by restriction with BamHI and Pst I. The pPZP111 binary vector was also restricted with BamHI and Pst I. The cut vector and the excised polyprotein fragment were ligated together. 200µl aliquots of high efficiency JM109 *E.coli* cells were transformed with the ligation reactions and plated onto LB agar plates containing 25µg/ml of chloramphenicol. The plasmid DNA from putative clones was extracted and analyzed by restriction digests to identify clones containing the insert.

A sequencing primer was designed for the promoter region immediately upstream of the multiple cloning site of the pPZP111 vector. Unfortunately, due to the manipulation of the pPZP111 vector multiple cloning site (outlined in section 6.1.1.) the CaMV promoter region is duplicated in the vector therefore the primer annealed to the second promoter region and the sequence obtained was the kanamycin resistance gene. An alternative sequencing primer was designed that was upstream of the duplicated region, this was used to sequence the 5' end of the insert and confirmed the presence of FMDV Lb^{PRO} and GUS respectively.

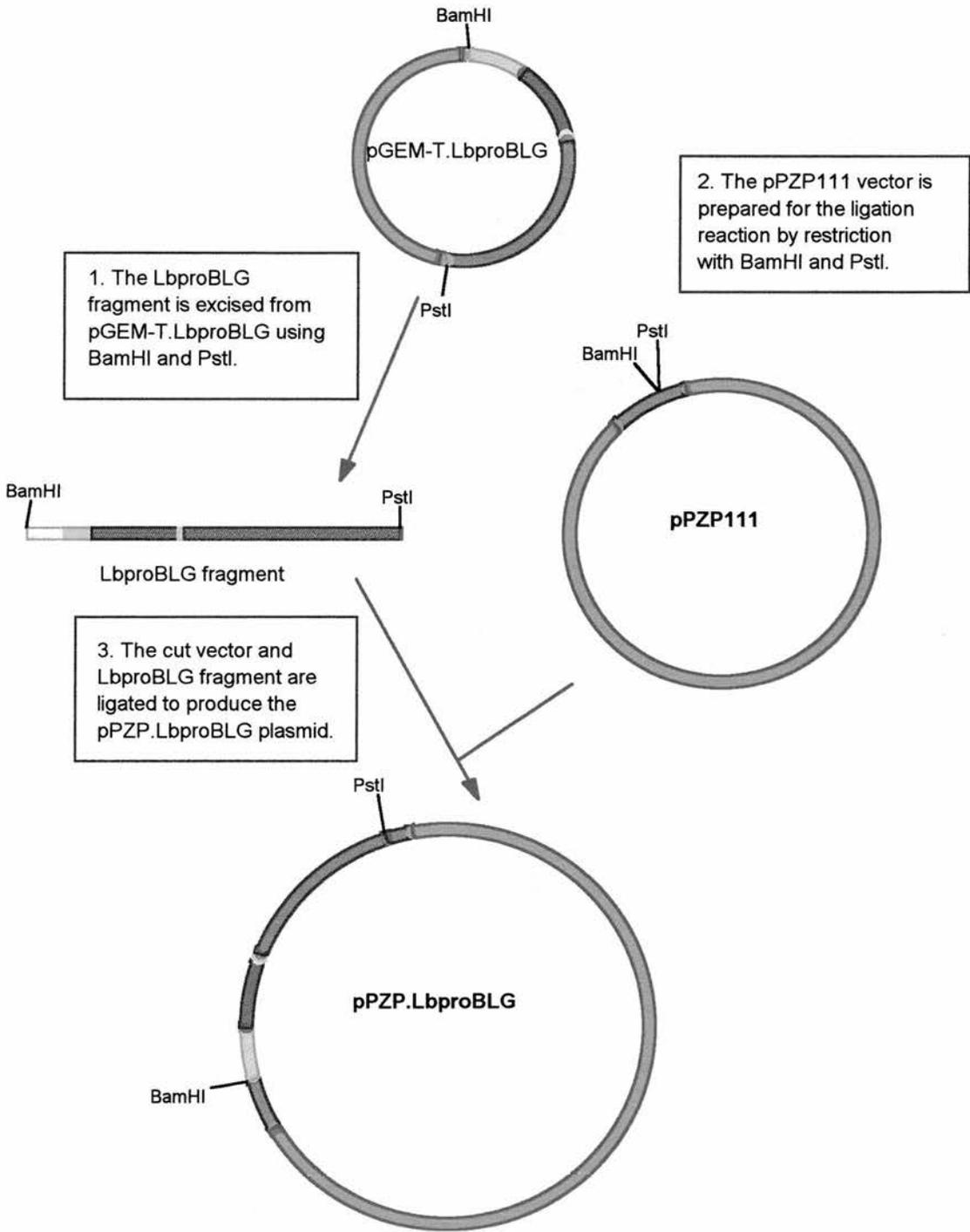


Fig.6.2. Construction of pPZP.Lb^{pro}BLG. The Lb^{pro}BLG fragment was excised from pGEM-T.Lb^{pro}BLG using BamHI and Pst I and ligated into similarly restricted pPZP111 vector.

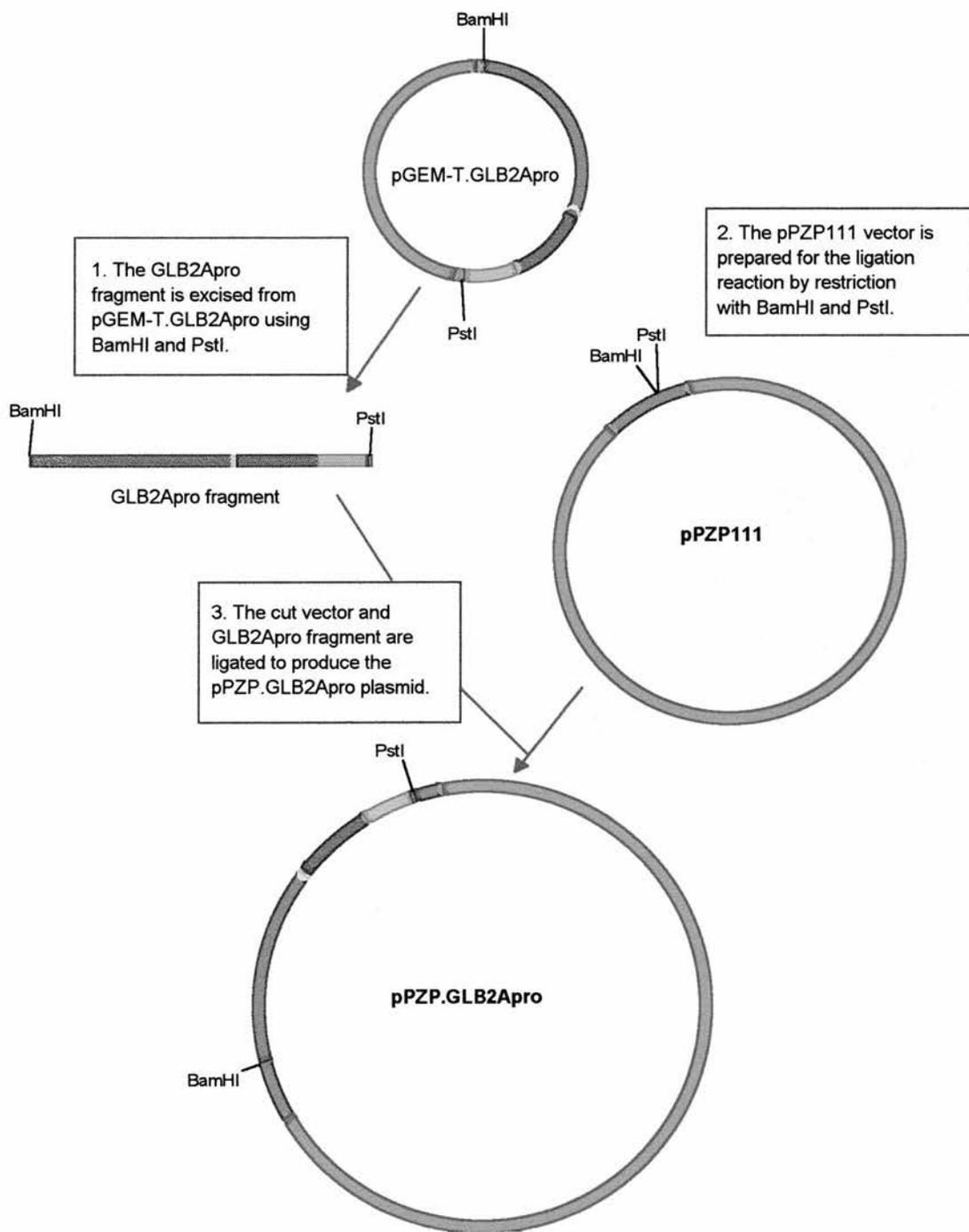


Fig.6.3. Construction of pPZP.GLB2A^{pro}. The GLB2A^{pro} encoding fragment was excised from pGEM-T.GLB2A^{pro} by restriction with BamHI and Pst I and ligated into similarly restricted pPZP.111 to produce the plasmid pPZP.GLB2A^{pro}.

6.1.4. Transformation of *Agrobacterium tumefaciens*.

In preparation for *Agrobacterium*-mediated transformation of *Nicotiana tabacum* the control PZP.BFP and the polyprotein containing plasmids were to be transferred into the *Agrobacterium* strain LBA4404. Competent *Agrobacterium tumefaciens* were prepared by the freeze-thaw method and transformed with 2µl of plasmid DNA of the original pPZP111 vector, pPZP.BFP, pPZP.Lb^{Pr}OBLG and pPZP.GLB2A^{Pr}O. The transformed bacteria were plated onto LB agar plates containing 25µg/ml of chloramphenicol. A sample of non-transformed *Agrobacterium* was also plated onto antibiotic selective media. The plates were incubated at 28°C for 2-3 days. The transformation failed as all the plates were clear. The *Agrobacterium* transformation was repeated but the same result was obtained. As the positive control was also clear this implied that *Agrobacterium* transformed with the pPZP111 vector were not resistant to the concentration of antibiotic in the plates.

6.1.5. Optimization of *Agrobacterium* transformation.

An experiment was designed to monitor the antibiotic resistance of pPZP111 transformed *Agrobacterium*. A series of LB agar plates was prepared with varying concentrations of the antibiotics chloramphenicol and kanamycin. Non-transformed *Agrobacterium* and *Agrobacterium* transformed with pPZP111 were plated onto this series of plates and incubated at 28°C for 2-3 days to allow colonies to develop. The results are shown in Table 6.1. and Table 6.2. Non-transformed *Agrobacterium* do not grow in the presence of chloramphenicol and pPZP111-transformed *Agrobacterium* are not resistant to concentrations above 10µg/ml. pPZP111-transformed *Agrobacterium* were resistant to all the concentrations of kanamycin tested but non-transformed *Agrobacterium* could also grow on low kanamycin plates. Therefore from these results optimum concentrations of antibiotic were selected for use with the pPZP111 series of plasmids. The optimum concentrations were 10µg/ml of chloramphenicol and 25µg/ml of kanamycin.

Conc. Cm. ($\mu\text{g/ml}$).	2.5	5	10	15	25
Transformed(PZP111)	+	+	+	-	-
Non-transformed	-	-	-	-	-

Table 6.1. *Agrobacterium* resistance to the antibiotic chloramphenicol.

Conc. Kan ($\mu\text{g/ml}$)	5	10	15	25
Transformed(PZP111)	+	+	+	+
Non-transformed	+	-	-	-

Table 6.2. *Agrobacterium* resistance to the antibiotic kanamycin.

To confirm that non-transformed *Agrobacterium* are killed by these concentrations of antibiotic, a series of 10ml aliquots of LB broth were inoculated with non-transformed *Agrobacterium tumefaciens* LBA4404 and incubated on an orbital shaker at 28°C for 2-3 days. The results are shown in Table 6.3. There was no bacterial growth in the antibiotic-selective media confirming that non-transformed bacteria were successfully killed by this concentration of antibiotic.

LB only	-
LB + bacteria	+
LB + bacteria + 10µg/ml Cm.	-
LB + bacteria + 25µg/ml Kan.	-

Table 6.3. *Non-transformed Agrobacterium antibiotic resistance in liquid culture.*

Competent *Agrobacterium* were prepared by the freeze-thaw method and transformed with 2µl of vector DNA for pPZP111, pPZP.BFP, pPZP.Lb^{pro}BLG and pPZP.GLB2A^{pro}. 100µl aliquots of transformed *Agrobacterium* were plated onto LB agar plates containing 25µg/ml of chloramphenicol and onto LB agar plates containing 10µg/ml of Kanamycin. The plates were incubated at 28°C for 2-3 days. Colonies grew on the positive control (pPZP111 only) plates and on pPZP.BFP and pPZP.GLB2A^{pro}. There were no colonies on the pPZP.Lb^{pro}BLG plates. Single putative colonies of transformed pPZP.BFP and pPZP.GLB2A^{pro} were picked and used to inoculate 10ml aliquots of antibiotic selective media. The cultures were incubated at 28°C for 2-3 days. The plasmid DNA was extracted from the *Agrobacterium* and analyzed by restriction digest thus confirming that pPZP.BFP and pPZP.GLB2A^{pro} were present.

6.1.6. Transformation of *Agrobacterium tumefaciens* with pPZP.Lb^{pro}BLG.

Competent *Agrobacterium tumefaciens* were produced via the freeze-thaw method and transformed with 1µl, 2µl, 5µl and 10µl of pPZP.Lb^{pro}BLG DNA. 100µl and 200µl aliquots of each transformation were plated onto LB agar plates containing 10µg/ml of chloramphenicol or 25µg/ml of kanamycin. The plates were incubated at 28°C for 2-3 days to allow colonies to develop but the transformation was not successful. *Agrobacterium* were transformed with 1mg of pPZP.Lb^{pro}BLG DNA via electroporation and plated onto LB agar plates containing either 10µg/ml

chloramphenicol or 25µg/ml and incubated at 28°C for 2-3 days. The transformation was still unsuccessful.

The freeze-thaw method was used to prepare competent *Agrobacterium* that were transformed with 2mg of pPZP.Lb^{PRO}BLG DNA. 100µl and 200µl aliquots of the transformation were plated onto two series of LB agar plates containing a range of concentrations of chloramphenicol or kanamycin. The plates were incubated at 28°C for 2-3 days to allow the colonies to develop. After incubation two putative colonies were present on the 15µg/ml kanamycin plate. 10ml aliquots of antibiotic selective LB was inoculated with the putative clones and incubated at 28°C for 2 days. The plasmid DNA was extracted and analyzed by restriction digest with Nru I. There are three Nru I restriction sites within the Lb^{PRO}BLG polyprotein cassette and no sites within the pPZP111 backbone. The fragment sizes of the Nru I restricted pPZP.Lb^{PRO}BLG DNA were predicted using the computer program Gene Construction Kit 2. Fully restricted pPZP.Lb^{PRO}BLG DNA would yield three fragments of 1115bp, 1641bp and 10047bp. The restricted DNA was separated by gel electrophoresis. The putative clone pPZP.Lb^{PRO}BLG2 had three bands corresponding to the predicted fragments thus confirming that the *Agrobacterium* contained the pPZP.Lb^{PRO}BLG binary vector.

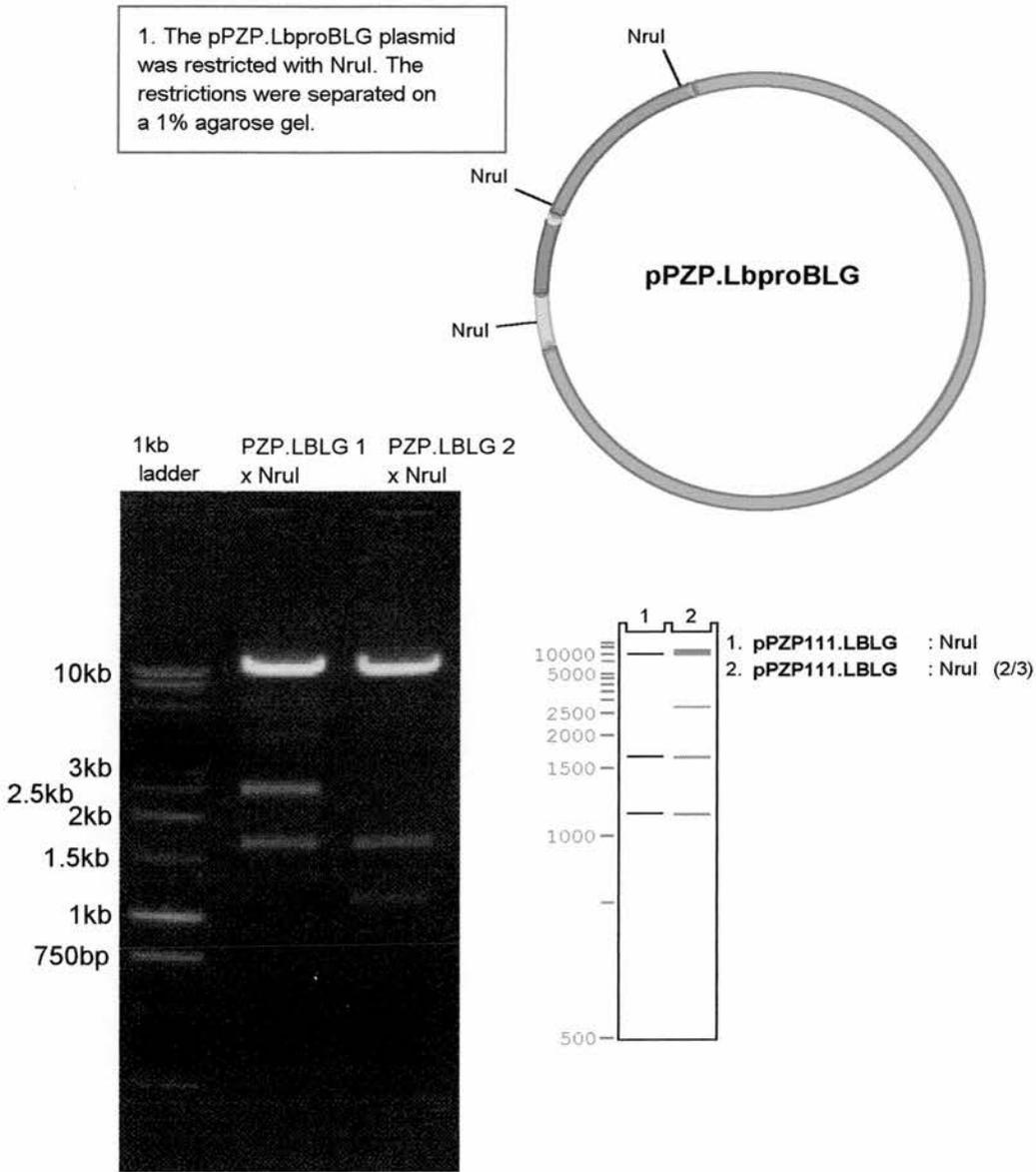


Fig.6.4. Restriction analysis with *NruI* of pPZP.Lb^{pro}BLG plasmid extracted from transformed *Agrobacterium*. The positions of the *NruI* restriction sites on the pPZP.Lb^{pro}BLG plasmid are indicated. The prediction of the restriction fragments is shown with the complete digestion products in Lane 1 and the partial digest products in Lane 2.

The *NruI* restriction of the DNA extracted from the other putative colony pPZP.Lb^{pro}BLG1 was not as clear, three DNA fragments were produced of approximately 1500bp, 2500bp and 10kbp. The 1500bp and the 10kbp band were expected but the third band should be approximately 1100bp. Analysis of the computer prediction of the partial digestion products suggested that the 2500bp band could be a

partial restriction product of pPZP.Lb^{PRO}BLG1. However, if the 1500bp fragment is present the 1100bp fragment would also be expected. The lack of the 1100bp fragment may be due to difficulties with detection of the DNA fragment. Thus at this stage the identification of pPZP.Lb^{PRO}BLG1 is inconclusive.

6.1.7. Further analysis of the pPZP.Lb^{PRO}BLG2 binary vector.

Due to the difficulties encountered with the transformation of *Agrobacterium tumefaciens* with pPZP.Lb^{PRO}BLG further analysis of the polyprotein cassette in pPZP.Lb^{PRO}BLG2 was required to verify that the protease within the polyprotein was functional. The Lb^{PRO}BLG cassette from pPZP.Lb^{PRO}BLG2 was excised by restriction with BamHI and Nsi I and inserted into similarly restricted pGEM7zf+ to yield pGEM.Lb^{PRO}BLG2. Competent JM109 *E.coli* cells were prepared, transformed with the ligations and plated onto LB agar plates containing 100µg/ml ampicillin. The plasmid DNA was extracted from putative colonies and analyzed by restriction analysis to identify positive clones. The pGEM.Lb^{PRO}BLG2 plasmid was sequenced using the T7, SP6 and BFP-N sequencing primers thus confirming that no mutations had been introduced in the polyprotein cassette during the *Agrobacterium tumefaciens* transformation.

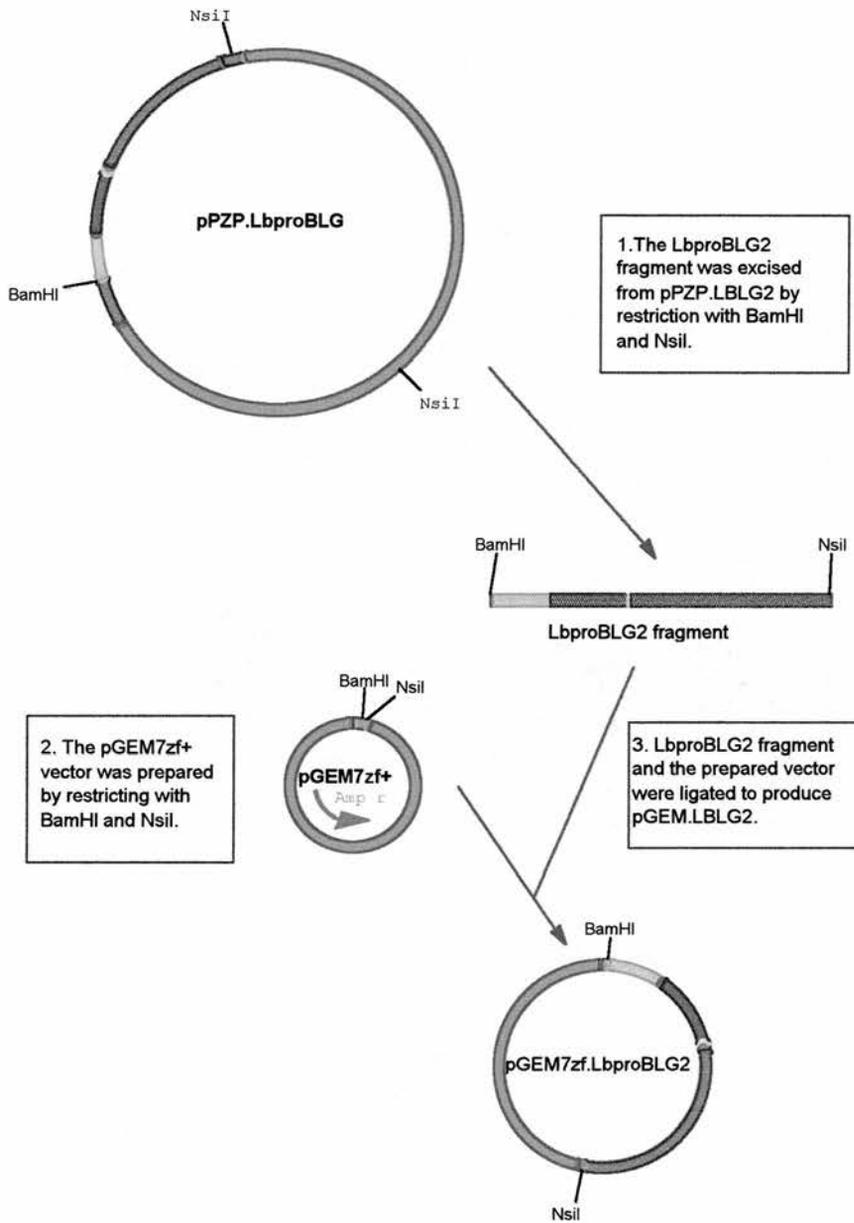


Fig.6.5. Transferral of the *Lb^{PRO}BLG* cassette from the *PZP.Lb^{PRO}BLG2* plasmid recovered from *Agrobacterium* to the *pGEM7zf+* vector. The *Lb^{PRO}BLG* cassette was excised with *Bam*HI and *Nsi*I and ligated into similarly restricted *pGEM7zf+* vector.

Finally, to verify that the FMDV Lb^{pro} within the Lb^{pro}BLG2 polyprotein cassette was active the pGEM.Lb^{pro}BLG2 plasmid was used to program a transcription and translation system *in vitro* in rabbit reticulocyte lysate. The radiolabelled translation products were analyzed by denaturing-PAGE and compared to the translation products obtained from the original pGEM.Lb^{pro}BLG plasmid. The pattern of translation products from pGEM.Lb^{pro}BLG and pGEM.Lb^{pro}BLG2 plasmids were identical. The FMDV Lb^{pro} cleaved *in cis* at the Lb^{pro}/BFP junction releasing the Lb protease and the BFPlinkerGUS polyprotein. The final cleavage products of BFP and GUS were not present indicating that the polyprotein was not processed *in trans*. Therefore the FMDV Lb protease within the pGEM.Lb^{pro}BLG2 polyprotein is functional thus the *Agrobacterium* transformed with PZP.Lb^{pro}BLG2 could be used to transform *Nicotiana tabacum* plants.

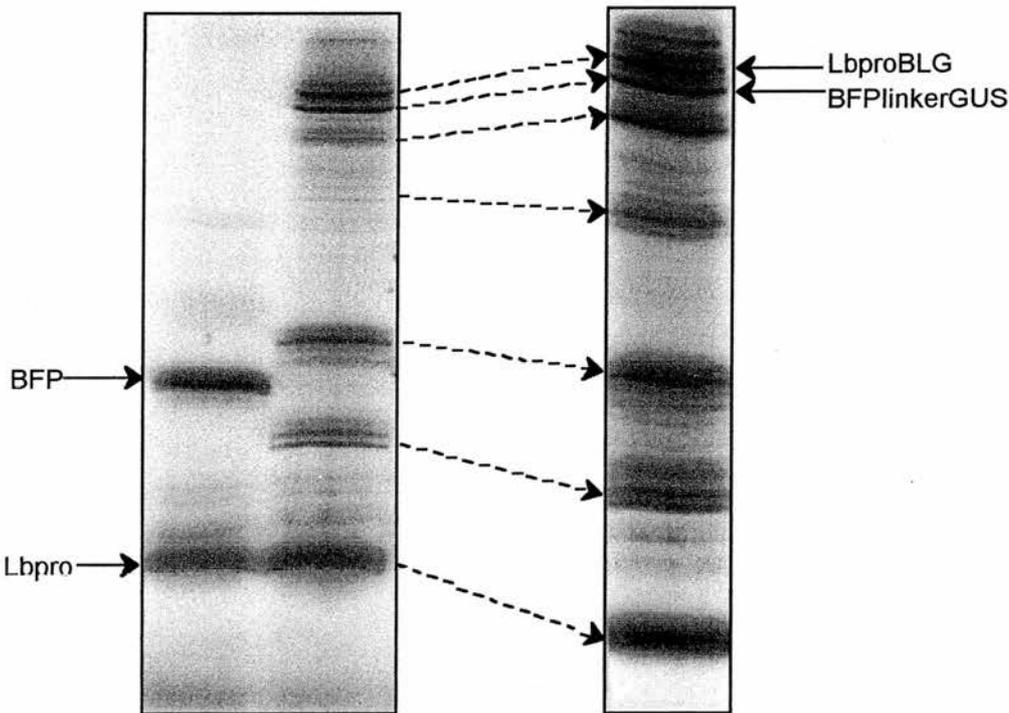


Fig.6.6. *In vitro* translation of pGEM.Lb^{pro}BLG2 in RRL. Lane 1- pGEM.Lb^{pro}BFP, Lane 2- pGEM.Lb^{pro}BLG, Lane 3- pGEM.Lb^{pro}BLG2.

6.1.8. *Agrobacterium*-mediated transformation of *Nicotiana tabacum*.

The pPZP.polyprotein plasmids had been successfully introduced into *Agrobacterium tumefaciens* thus the transformed bacteria were used to introduce the plasmids into the *Nicotiana tabacum* plants.

Wild-type *Nicotiana tabacum* plants were sterilized in 10% bleach for 15 minutes and washed in sterile water to remove the bleach. Leaf-discs were prepared from the sterilized leaves and plated onto solid MS media containing the hormones BA (1µg/ml) and NAA (0.1µg/ml). The plates were sealed with parafilm and incubated at 25°C for 48 hours. 30ml cultures of antibiotic selective media were inoculated with transformed *Agrobacterium* containing pPZP.BFP, pPZP.Lb^{PRO}BLG and pPZP.GLB2A^{PRO} and grown at 28°C for 48 hours. After 48 hours approximately thirty leaf-discs were submerged in diluted bacterial culture, dried on sterile filter paper, returned to the hormone-containing MS plates and incubated at 28°C under constant light for 2 days. The leaf-discs were washed in MS media containing the hormones and the antibiotics carbenicillin (0.5mg/ml) and kanamycin (10µg/ml), dried and plated onto fresh MS agar plates containing antibiotics and hormones. The plates were sealed with parafilm and incubated at 25°C under constant light. The transformation was aborted 2 weeks post-transformation due to contamination of the leaf-discs with various bacteria and fungi. The leaf-discs were prepared from greenhouse-grown *Nicotiana tabacum* plants and unfortunately the sterilization procedures undertaken prior to preparation of the leaf-discs were not sufficiently vigorous to remove contaminating organisms.

Due to the lack of suitable established sterile *Nicotiana tabacum* plants an alternative transformation protocol was attempted. *Nicotiana tabacum* seeds were sterilized and sown onto MS agar plates. The plates were sealed with parafilm and incubated vertically in the growth room at 25°C, constant light for 14 days. After 12 days, 10ml of antibiotic selective media was inoculated with a stab of transformed *Agrobacterium*

tumefaciens. The bacteria were grown at 28°C on an orbital shaker for 48 hours. A 3ml sample of each bacterial culture was pelleted and washed twice with sterile 10mM MgSO₄ solution. The bacteria were resuspended in 10ml of sterile MS media and approximately one hundred 14 day old *Nicotiana tabacum* seedlings were added to each bacterial culture. The bacteria/seedling mixture was subjected to 0.15 atmospheres of pressure for 5 minutes. The seedlings were plated onto fresh MS agar plates containing the plant hormones BA and NAA and incubated at 25°C under constant light for 3 days. The seedlings were washed in sterile 10mM MgSO₄ containing antibiotics and hormones and plated onto fresh antibiotic selective MS agar plates and incubated at 25°C under constant light. The seedlings were washed in MgSO₄ solution and transferred to fresh antibiotic selective MS agar plates once a week. After 4 weeks the *Agrobacterium*-transformed seedlings were dark green and healthy with well-established callus tissue. In contrast, the non-transformed wild-type plants were whitish-green and did not have any callus tissue (refer to Fig.6.7).

The callus tissue was carefully removed from thirty transformed seedlings for each plasmid. The seedlings were micropropagated, transferred to pots containing antibiotic selective MS agar and incubated at 25°C under constant light to allow root development. The following observations were made two weeks after the plantlets were transferred to individual pots.

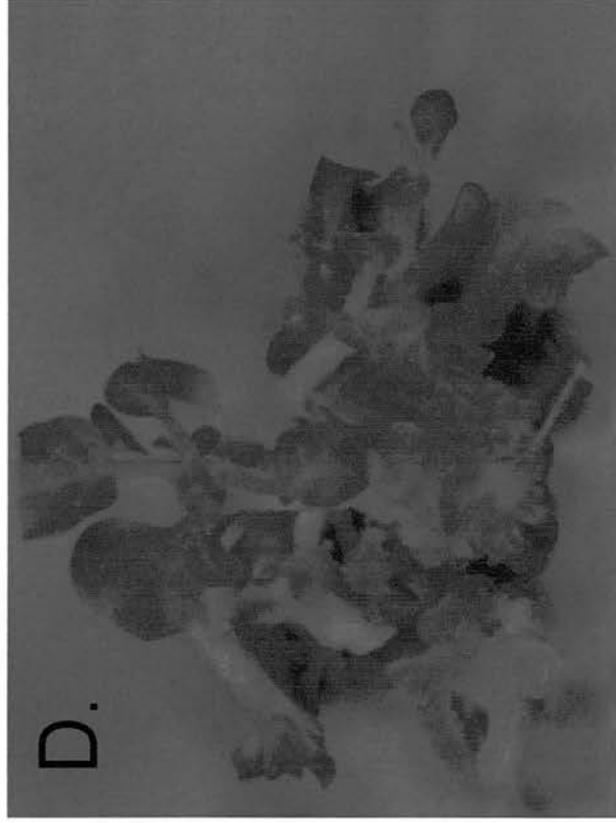
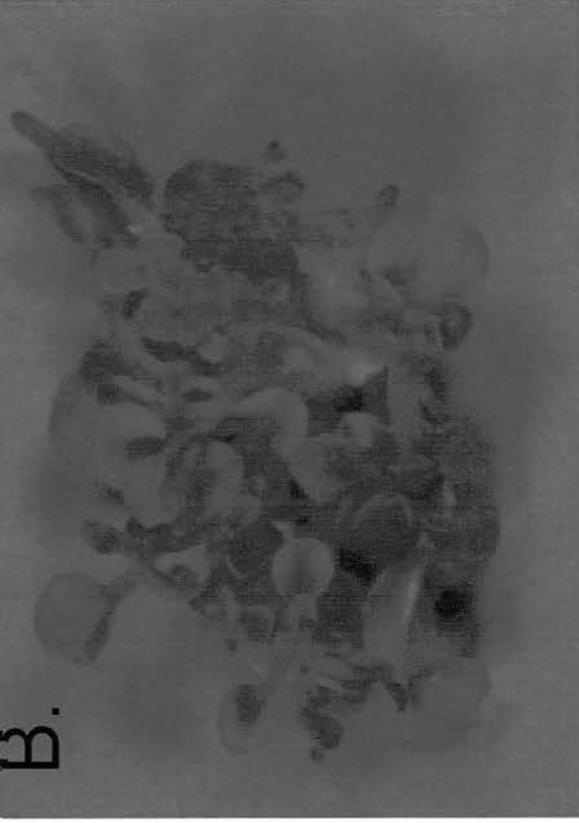


Fig.6.7. Four week old *Nicotiana tabacum* seedlings. A-wt *nicotiana tabacum*, B-pZP:BFP transformant, C-pZP:LbproBLG transformant, D-pZP:GLB2Apro transformant.

Plasmid	Rooted	Healthy, no roots	Dead
pPZP.BFP	6	6	18
pPZP.Lb ^{PRO} BLG	7	7	16
pPZP.GLB2A ^{PRO}	9	9	12

Table 6.4. Observations of plantlets from the *Agrobacterium*-mediated seedling transformation experiment.

The healthy, green plants were micropropagated and transferred to fresh antibiotic selective MS media to try to encourage root development. The rooted plants were micropropagated and tissue samples were harvested for analysis. Of the rooted plants for the polyprotein-containing plasmids four of each type were strong, established plants.

Following the low transformation rate of the seedling transformation the leaf-disc transformation was repeated using *Nicotiana tabacum* plants grown in sterile conditions in an attempt to improve the number of available transformed plants. The leaf-discs were prepared (25 per condition) and plated onto MS agar plates containing NAA and BA. The cut discs were incubated at 25°C under constant light for 2 days. During the incubation a culture of transformed *Agrobacterium* was grown in selective media at 28°C. The leaf-discs were submerged in diluted bacterial culture, dried, returned to the hormone-containing MS plates and incubated at 28°C under constant light for 2 days. The leaf-discs were washed in MS media containing the antibiotics and hormones, dried and plated onto fresh MS agar plates containing antibiotics and hormones. The plates were sealed with parafilm and incubated at 25°C under constant light for 4-6 weeks.

After 2 weeks it was observed that a number of leaf-discs were infected with a single-type of bacterium. The preliminary conclusion was that the bacteria were *Agrobacterium tumefaciens* that had not been killed by the carbenicillin. Therefore a new stock solution of carbenicillin was prepared. The leaf discs were washed in sterile antibiotic selective MS media and transferred to fresh MS agar plates containing antibiotics and hormones and incubated at 25°C, these measures did not remove the bacterial infection.

Other members of the laboratory also encountered similar problems and subsequent analysis of the bacteria identified a strain of coccus. Preliminary investigations involving treating the infected plants with the antibiotic imipenem appeared to successfully remove the infection. Therefore the regenerating shoots were excised from the leaf-discs and transferred individually to pots of MS media containing imipenem (10µg/ml) and kanamycin (10µg/ml). Unfortunately the bacterial infection slowly reestablished and the plants were overgrown by the time they were ready to harvest. The following observations were made for each transformation condition.

Plasmid	Rooted	Healthy, no roots	Dead
pPZP.BFP	15	5	8
pPZP.Lb ^{PRO} BLG	9	3	19
pPZP.GLB2A ^{PRO}	6	10	8

Table 6.5. Observations of plantlets from *Agrobacterium*-mediated leaf-disc transformation experiment.

Despite the problems caused by the bacterial infection a number of plants had rooted and were well-established however, it was decided because the plants were still infected not to proceed any further with this set of transformants.

6.1.9. Initial analysis of pPZP.Lb^{PRO}BLG and pPZP.GLB2A^{PRO} transformants from the *Agrobacterium*-mediated seedling transformation experiment.

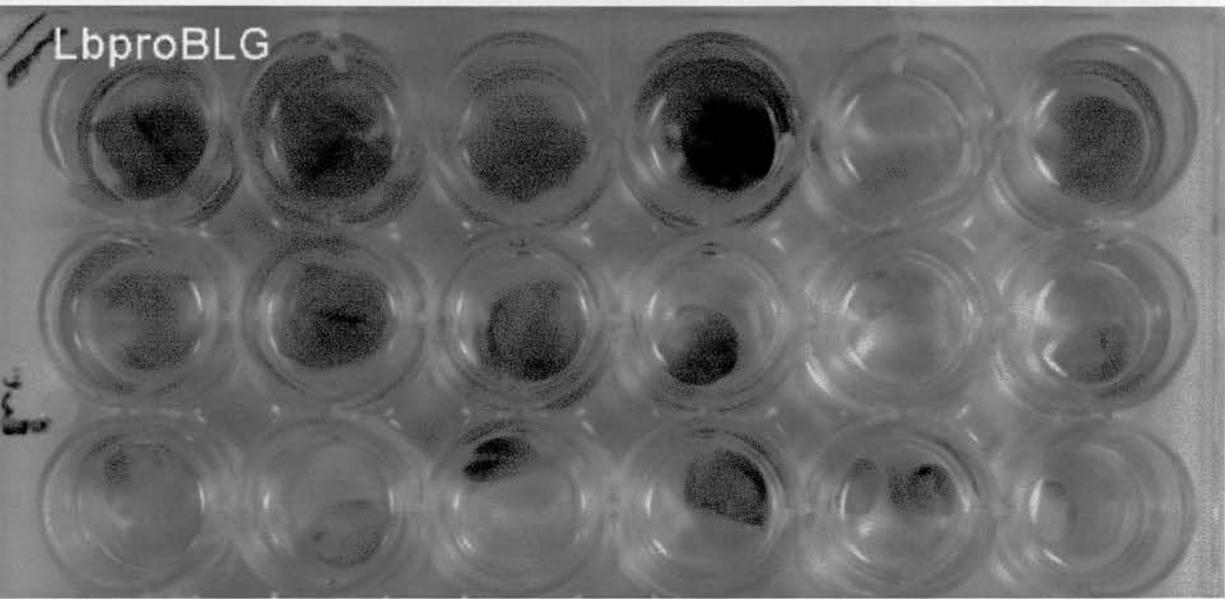
A simple preliminary assay to identify transformed plants containing the reporter protein β -glucuronidase was used to analyze the green rooted and non-rooted plants from the seedling transformation of *Nicotiana tabacum*. A small leaf sample was removed from each plant. The GUS extraction buffer was prepared by adding the chromogenic X-GlcA substrate (1mg/ml). The sample tissue was put into the GUS extraction buffer and incubated under pressure for 30 minutes. The samples were incubated at 37°C for overnight to allow the blue colour to develop. After sufficient colour development the GUS extraction buffer was removed and the leaf samples were treated with ethanol to remove any background chlorophyll. The leaf samples were photographed both under normal white light and backlit with a fluorescent lamp to shine light through the leaf.

Lb^{PRO}BLG GUS assay. The putative Lb^{PRO}BLG transformants did not test positive for β -glucuronidase, although Lb^{PRO}BLG 4 did have a small amount of blue colouration

around the base of the leaf (refer to Fig. 6.8). Therefore positive Lb^{PRO}BLG transformants could not be identified.

GLB2A^{pro} GUS assay. The leaf samples from the putative GLB2A^{PRO} transformants 1 to 4 were deep blue thus indicating that the β -glucuronidase protein was present (refer to Fig.6.9). Plants 1 to 4 were the well-rooted, established plants mentioned above. The remaining rooted plants and the non-rooted green plants did not test positive for β -glucuronidase.

A.



B.

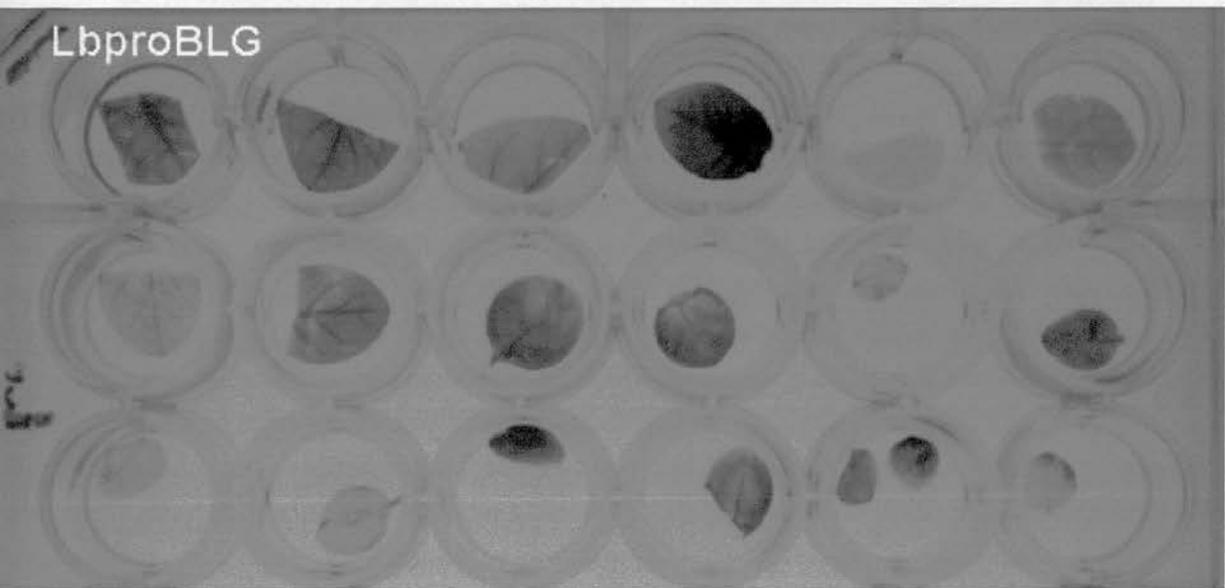
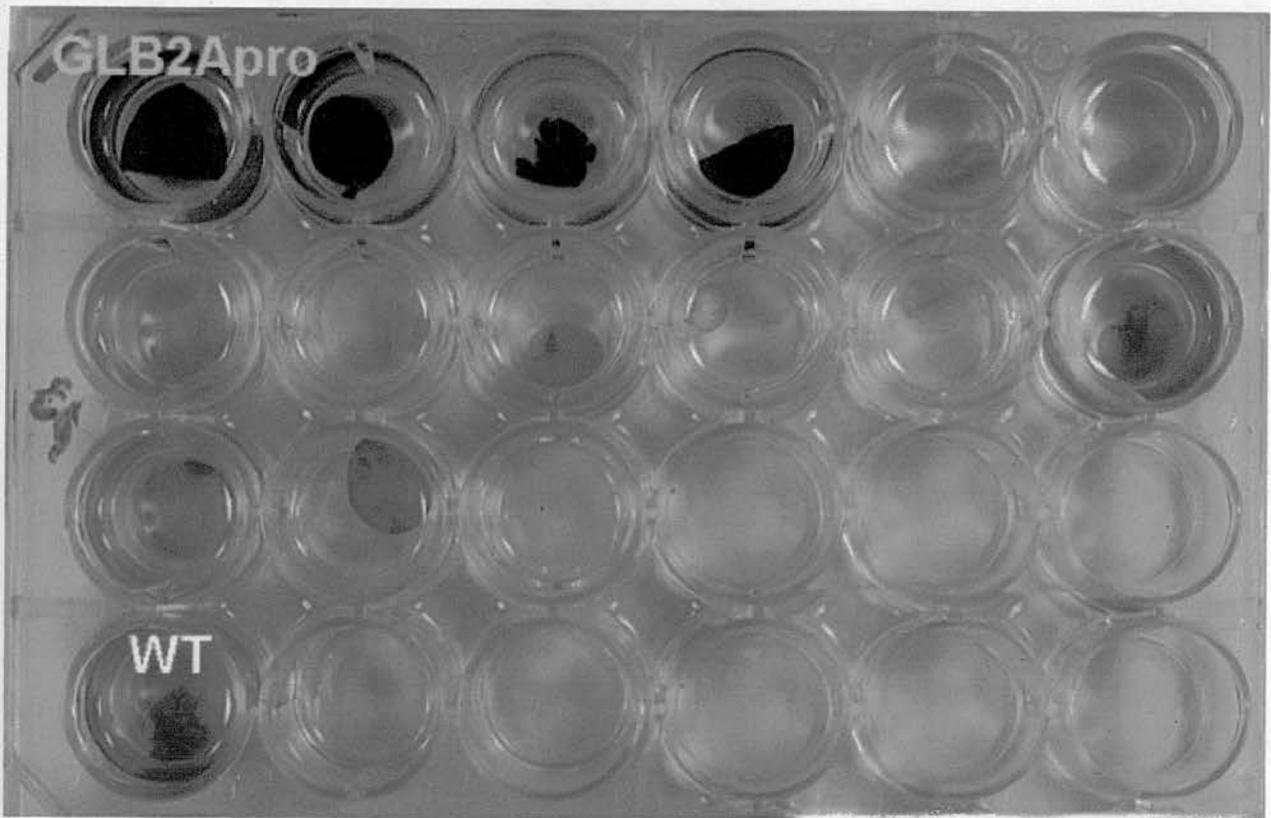


Fig.6.8. LbproBLG GUS assay of putative transformant tobacco plants. Photograph A. was viewed under normal white light. Photograph B. was taken when the samples were backlit.

A.



B.

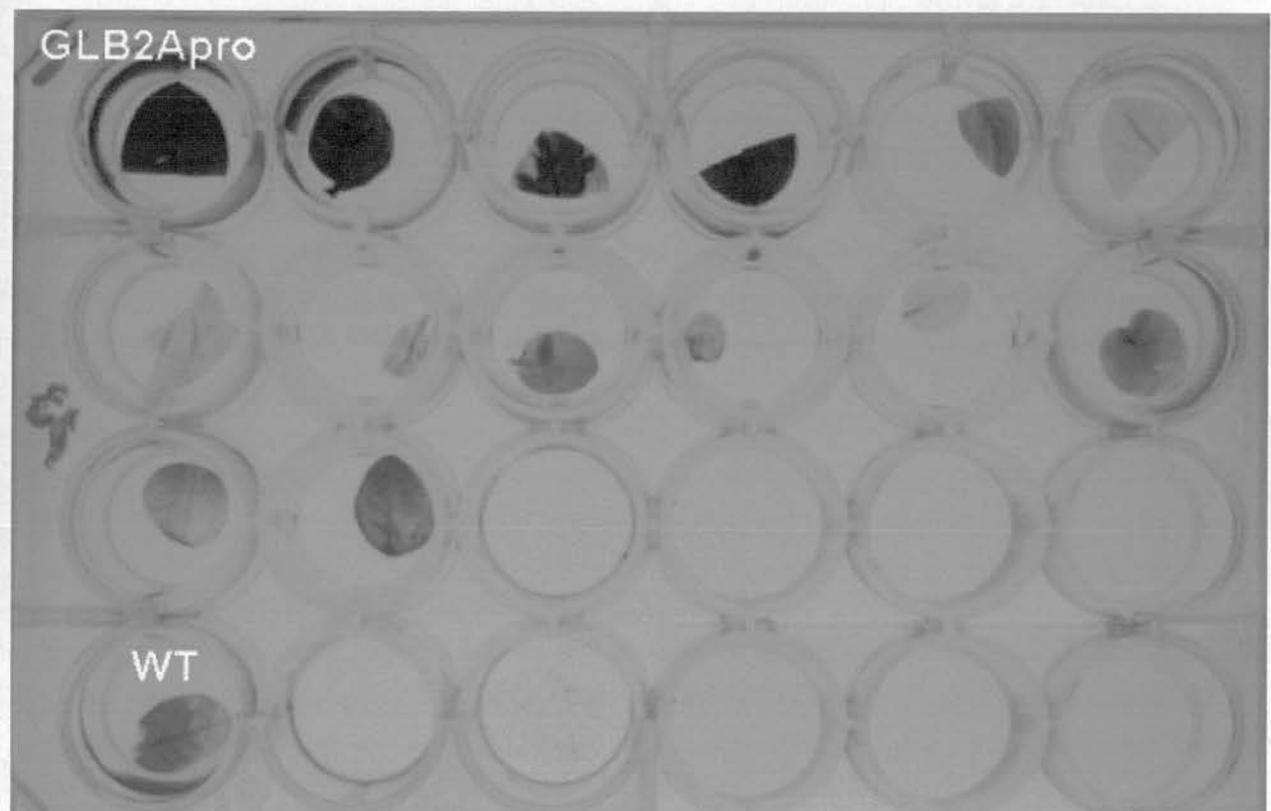


Fig.6.9. GUS assay of putative GLB2Apro transformed tobacco plants. Photograph A. was viewed under normal white light. Photograph B was taken when the samples were backlit.

6.1.10. Western blot analysis of putative Lb^{PRO}BLG transformants.

All of the samples from the putative Lb^{PRO}BLG transformants were negative for the β -glucuronidase reporter protein indicating that either the polyprotein was not expressed or that the β -glucuronidase protein in the polyprotein was not present or inactive. In order to ascertain which of these possibilities was correct the Lb^{PRO}BLG transformants were analyzed by western blotting with antibodies to both reporter proteins. The plants were micropropagated and tissue samples from the four well-established transformants were collected. 5g of leaves were ground to a powder in liquid Nitrogen in a pre-cooled mortar and pestle. The proteins were extracted by adding extraction buffer to the powdered sample. The unground tissue fragments were removed by centrifugation. The protein concentration of each protein extract was determined using the protein assay kit (Bio-rad). 30 μ g of protein from each Lb^{PRO}BLG transformant and the wildtype *Nicotiana tabacum* were separated by denaturing-PAGE. The separated proteins were transferred to a nitrocellulose membrane using a semidry blotting apparatus.

Anti-BFP western blot analysis. The blot was blocked in 5% milk in 0.1% PBS-Tween® for 1 hour and washed three times in a large volume of 0.1% PBS-Tween®. The blot was probed with a 1:200 dilution of the living colours peptide antibody (Clontech) in 5% dried milk in 0.1% PBS-Tween® for 1 hour. The excess antibody was removed by washing in a large volume of 0.1% PBS-Tween® for 10 minutes three times. The blot was probed with a 1:3000 dilution of HRP-linked anti-rabbit antibody in 5% dried milk in 0.1% PBS-Tween® for 1 hour and washed as before. The proteins detected by the antibody were visualized using the ECL™ chemiluminescence system and autoradiography.

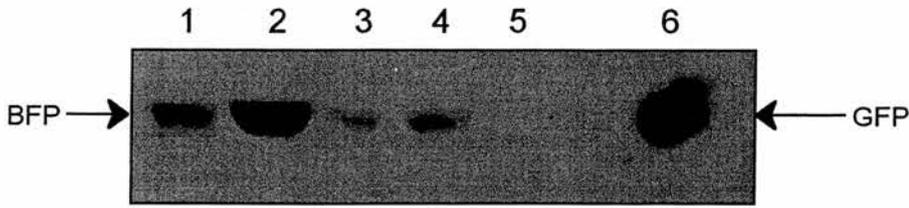


Fig. 6.10. Anti-BFP western blot of *Lb^{PRO}BLG* transformants. Lane 1- *Lb^{PRO}BLG* 1, Lane 2- *Lb^{PRO}BLG* 2, Lane 3- *Lb^{PRO}BLG* 3, Lane 4- *Lb^{PRO}BLG* 4, Lane 5- Wild-type extract, Lane 6- Recombinant GFP.

Probing the protein extract of the *Lb^{PRO}BLG* transformants with BFP antisera identified a single protein of approximately 30kDa this protein was not present in the lane containing the wild-type protein extract. The protein was the same size as the band corresponding to the GFP protein in the positive control lane that contains 1 μ l of a 1:50 dilution of recombinant GFP protein. Therefore all four *Lb^{PRO}BLG* transformants contain the BFP protein cleavage product. The anti-BFP western blot analysis did not detect any larger polypeptide precursor products.

Anti-GUS western blot analysis. A duplicate blot was blocked in 5% dried milk in 0.1% PBS-Tween® for 1 hour and washed three times in a large volume of 0.1% PBS-Tween®. The blot was probed with a 1:3000 dilution of the 5 prime->3 Prime GUS antibody in 5% dried milk in 0.1% PBS-Tween® for 1 hour and the excess antibody was removed by washing in a large volume of 0.1% PBS-Tween® for 10 minutes three times. The blot was probed with a 1:3000 dilution of HRP-linked anti-rabbit antibody in 5% dried milk in 0.1% PBS-Tween® for 1 hour and washed three times in 0.1% PBS-Tween® for 10 minutes. The proteins detected by the antibody were visualized using the ECL™ chemiluminescence system and autoradiography.

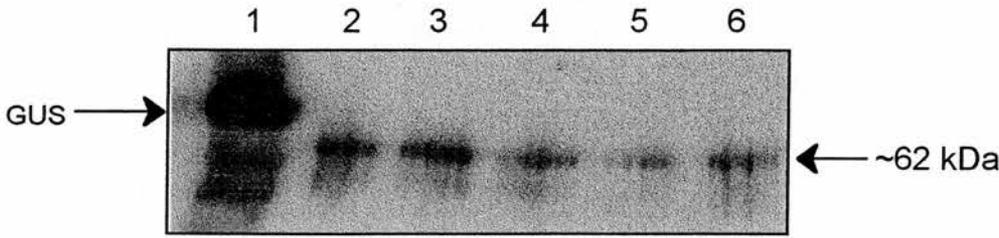


Fig.6.11. Anti-GUS western blot analysis of *Lb^{PRO}BLG* transformants. Lane 1- cellular extract of *E.coli* expressing pRAJ 275, Lane 2- Wild-type extract, Lane 3- *Lb^{PRO}BLG* 1, Lane 4- *Lb^{PRO}BLG* 2, Lane 5- *Lb^{PRO}BLG* 3, Lane 6- *Lb^{PRO}BLG* 4.

A band of approximately 69kDa corresponding to the GUS protein was present in the positive control lane. The western blot analysis did not identify any bands corresponding to the β -glucuronidase protein in the *Lb^{PRO}BLG* protein extracts. A similar sized band of approximately 62kDa in size was detected in all the plant extract samples including the wild-type extract. This 62kDa band was caused by the presence of β -mercaptoethanol in the 2xSDS protein loading buffer therefore in subsequent anti-GUS western blots the samples were loaded in sample buffer without β -mercaptoethanol to remove this band.

Western blot analysis of the *Lb^{PRO}BLG* transformants was repeated using increased antibody concentrations for the primary and secondary antibodies. Alternative GUS antisera donated by Zeneca and SCRI were also used to probe the blot. All the conditions tested failed to identify any proteins corresponding to the β -glucuronidase cleavage product or any larger precursor polypeptides refer to Fig.6.12.

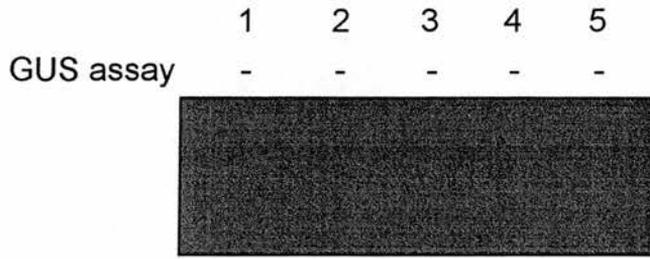


Fig.6.12. Anti-GUS western blot analysis of *Lb^{PRO}BLG* transformants. Lane 1- Wild-type extract, Lane 2- *Lb^{PRO}BLG* 1, Lane 3- *Lb^{PRO}BLG* 2, Lane 4- *Lb^{PRO}BLG* 3, Lane 5- *Lb^{PRO}BLG* 4. The GUS assay results are indicated above each lane.

6.1.11. Western blot analysis of *GLB2A^{PRO}* transformants.

The GUS assay was positive for the four well-established *GLB2A^{PRO}* transformants. The plants were micropropagated and the tissue was harvested. A 5g tissue sample was ground to a powder in liquid Nitrogen and the proteins were extracted. The amount of protein in the crude protein extract was determined using the Bio-rad protein assay kit. A 30µg protein sample from the *GLB2A^{PRO}* transformants and the wild-type *Nicotiana tabacum* were separated by denaturing-PAGE analysis. The proteins were transferred to a nitrocellulose membrane on a semi-dry transfer blotting apparatus.

Anti-BFP western blot analysis. The blot was blocked in 5% dried milk in 0.1% PBS-Tween® for 1 hour and washed in a large volume of 0.1% PBS-Tween® three times. The blot was probed with a 1:200 dilution of the living colours peptide antibody (Clontech) in 5% dried milk in 0.1% PBS-Tween® for 1 hour and the excess antibody was removed by washing three times in a large volume of 0.1% PBS-Tween® for a total of 60 minutes. The blot was probed with a 1:3000 dilution of HRP-linked anti-rabbit antibody in 5% dried milk in 0.1% PBS-Tween® for 1 hour and washed as before. The proteins detected by the BFP antibody were visualized using the ECL™ chemiluminescence system and autoradiography.

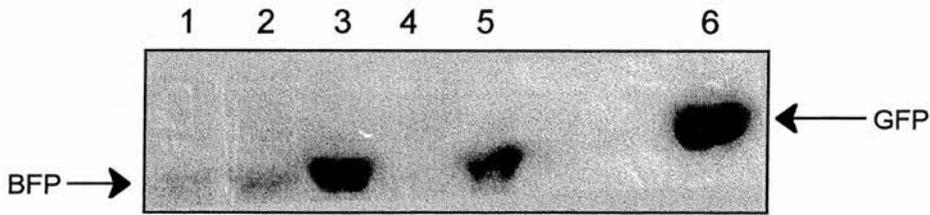


Fig. 6.13. Anti-BFP western blot of GLB2A^{Pro} transformants. Lane 1- Wild-type extract, Lane 2- GLB2A^{Pro} 1, Lane 3- GLB2A^{Pro} 2, Lane 4- GLB2A^{Pro} 3, Lane 5- GLB2A^{Pro} 4, Lane 6- Recombinant GFP.

Probing with BFP antisera detected a protein of approximately 30kDa in the protein extract of GLB2A^{Pro} 1, 2 and 4. The 30kDa protein was not detected in the lane containing the wild-type extract. This protein was the same size as the positive control recombinant GFP protein. Thus the GLB2A^{Pro} transformants 1, 2 and 4 contain the BFP cleavage product. Larger proteins corresponding to polypeptide precursors were not detected.

Anti-GUS western blot analysis. The GUS assays of the leaf tissue from all four GLB2A^{Pro} samples were positive for the β -glucuronidase protein. A 30 μ g sample of crude protein extract from each transformant and the wild-type protein was separated by denaturing PAGE analysis and transferred to a nitrocellulose membrane. The blot was blocked in 5% dried milk in 0.1% PBS-Tween® for 1 hour and washed three times in a large volume of 0.1% PBS-Tween®. The blot was probed with a 1:5000 dilution of the GUS antibody donated by Zeneca in 5% dried milk for 1 hour. The unbound antibody was removed by washing in 0.1% PBS-Tween® for ten minutes three times. The blot was incubated in a 1:4000 dilution of the secondary HRP-linked rabbit antibody in 5% dried milk in 0.1% PBS-Tween® for 1 hour and thoroughly washed as before. The proteins detected by the GUS antibody were visualized via chemiluminescence and autoradiography.

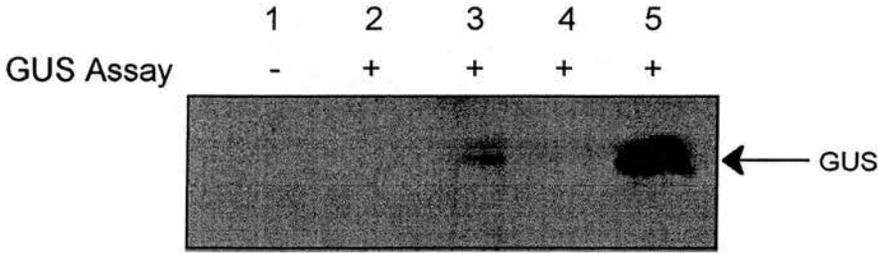


Fig. 6.14. Anti-GUS western blot analysis of GLB2A^{Pro}. Lane 1- Wild-type extract, Lane 2- GLB2A^{Pro} 1, Lane 3- GLB2A^{Pro} 2, Lane 4- GLB2A^{Pro} 3, Lane 5- GLB2A^{Pro} 4. The GUS assay results are indicated above each lane.

The GUS western blot was not as sensitive as the enzyme activity assay. A single band of approximately 69kDa corresponding to the GUS cleavage product was identified in only three samples GLB2A^{Pro} 2, 3 and 4. Probing with GUS antisera did not detect any larger precursor polypeptide products.

6.1.12. Visualization of the BFP cleavage product via UV illumination.

An attractive feature of using BFP as a reporter protein in transformed plants was the possibility of utilizing the intrinsic BFP fluorescence as a simple screening method to identified transformed plants. A brief excursion into the viability of using BFP in this way was undertaken. The BFP protein contains a chromophore that has a fluorescence excitation maxima at 380nm and emission maxima at 440nm (refer to Fig.6.15).

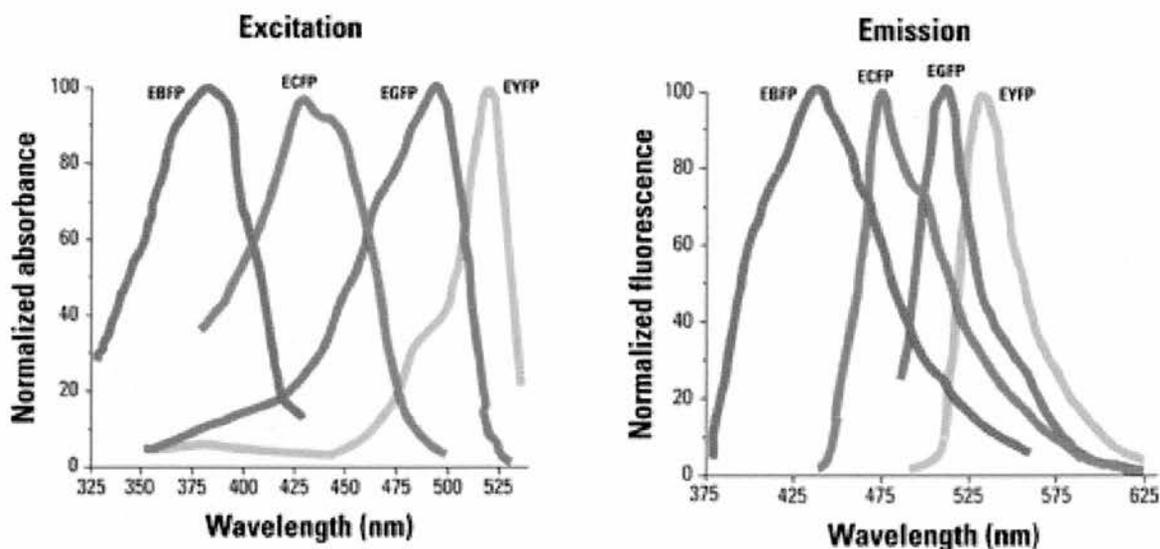


Fig.6.15. The excitation and emission spectra of the enhanced range of GFP protein variants including the eBFP protein used as a reporter protein in this thesis.

UV illumination of the crude root extracts. The roots were collected from an established plant of the wild-type *Nicotiana tabacum* and each transformant. The root extract was used because unlike the leaves there is little or no chlorophyll present to interfere with the BFP fluorescence. A crude protein extract was prepared by grinding the roots to a powder in liquid Nitrogen and adding extraction buffer. The unground tissue was removed by centrifugation.

The root extracts were viewed under UV light from a long-wave UV illuminator that emits light at a wavelength 395nm. The extracts were photographed under UV illumination using a digital camera. The final image was produced by removing the background blue colour present in the wildtype extract from all the samples using the computer program Adobe photoshop 5.5 (refer to Fig.6.16).

The LbP^{PRO}BLG extracts did not fluoresce above the background wild-type level. In contrast, the GLB2A^{PRO} samples 1 and 4 were significantly more fluorescent.

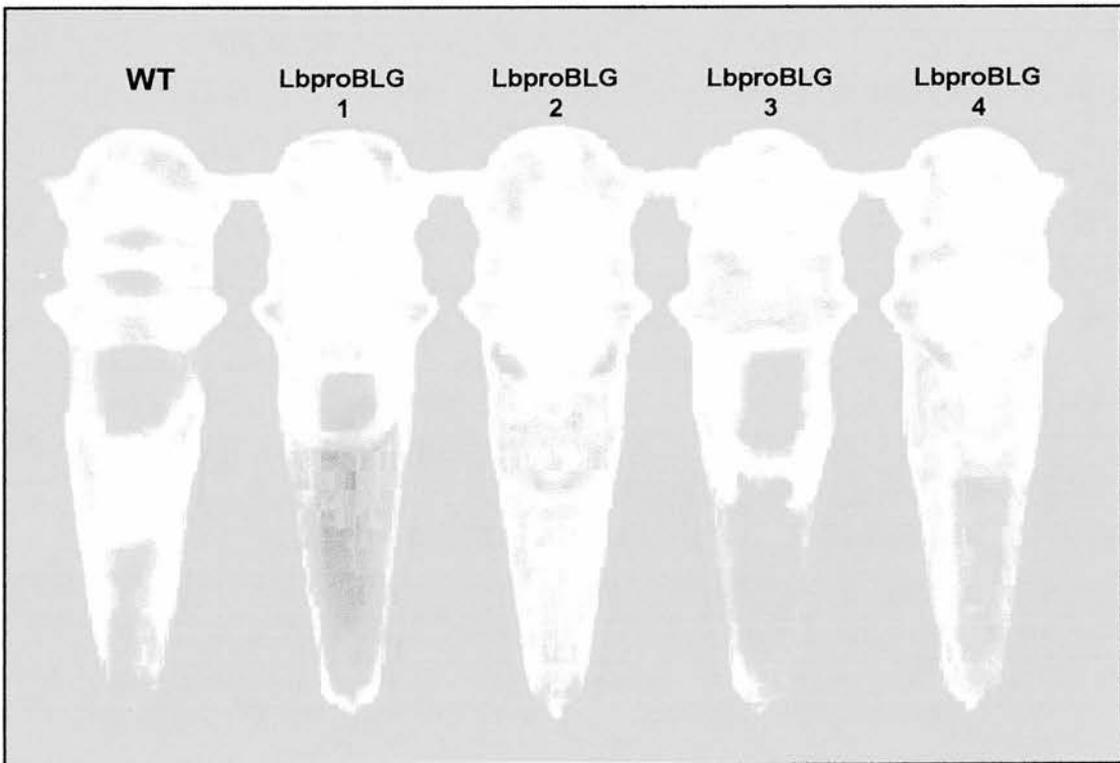
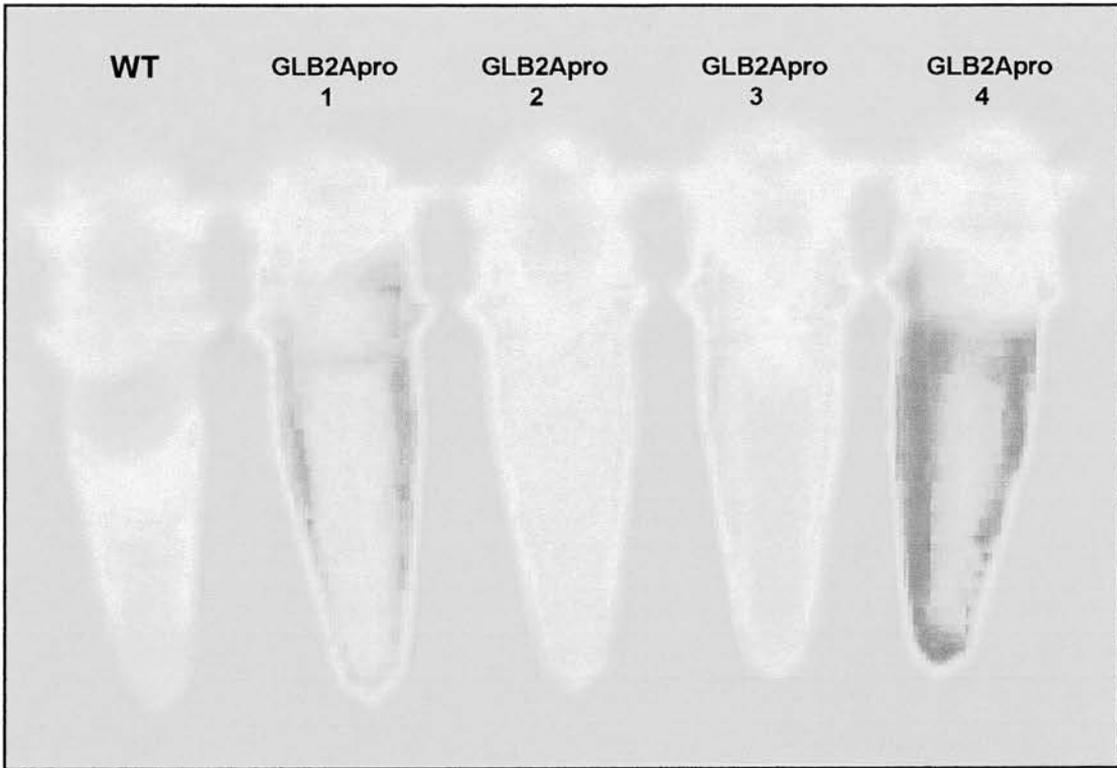


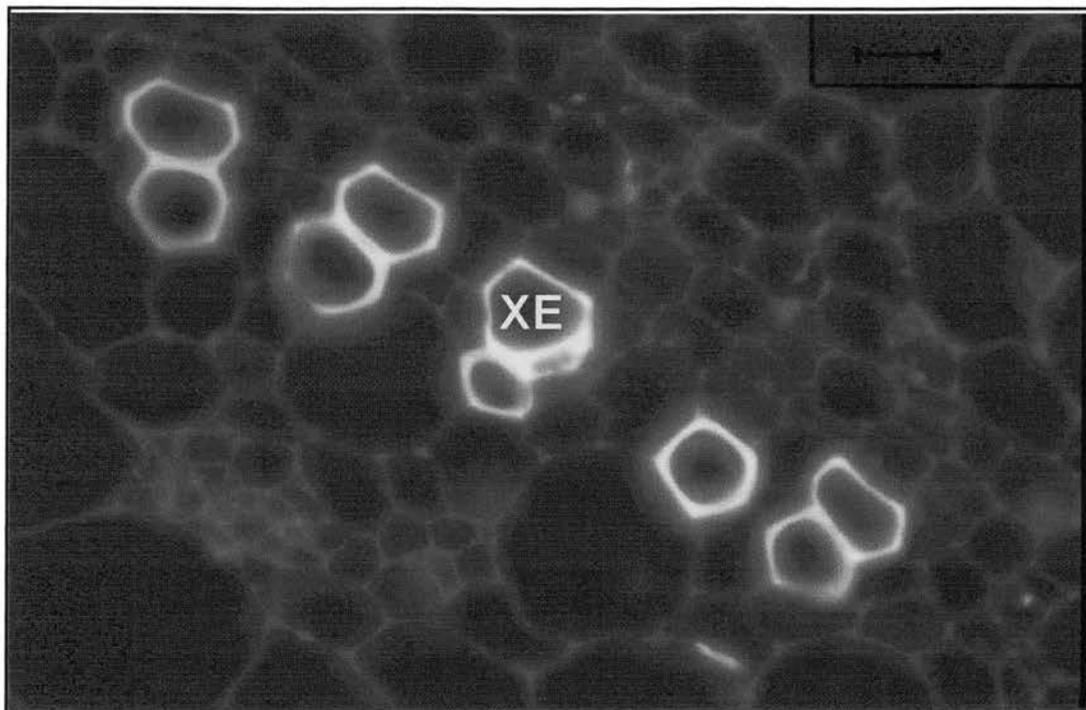
Fig.6.16. The root extracts from the wild-type plant and the transformants were viewed under UV light and photographed using a digital camera. The digital image was modified using Adobe Photoshop 5.5. to remove the background blue light from the UV illumination from all the samples.

Fluorescence microscopy of plant stem sections. Microscope slides of the stem tissue of the wildtype *Nicotiana tabacum* and the transformant plants were prepared by Miss Jill McVee. The tissue was embedded in paraffin blocks and sectioned using a microtome. The stem sections were transferred to a microscope slide and dewaxed.

A series of stem sections were dewaxed and stained with haematoxylin and eosin (H&E) to identify basic cell structures (refer to Fig.6.17B). The nuclei are stained blue and the various tissue components are stained varying shades of red and pink.

The unstained stem sections were viewed under UV light at 365nm using a fluorescence microscope (Zeiss axioplan universal microscope) and photographed. The same pattern of fluorescence was seen in all the plants including the wildtype (refer to Fig. 6.17). The thickened vascular bundles were highly fluorescent at this wavelength. The cell walls could also be clearly picked out. This suggests that eBFP may not be useful in plants for this type of work as there are native compounds present that fluoresce intensely at the optimum excitation wavelength of eBFP.

A.



B.

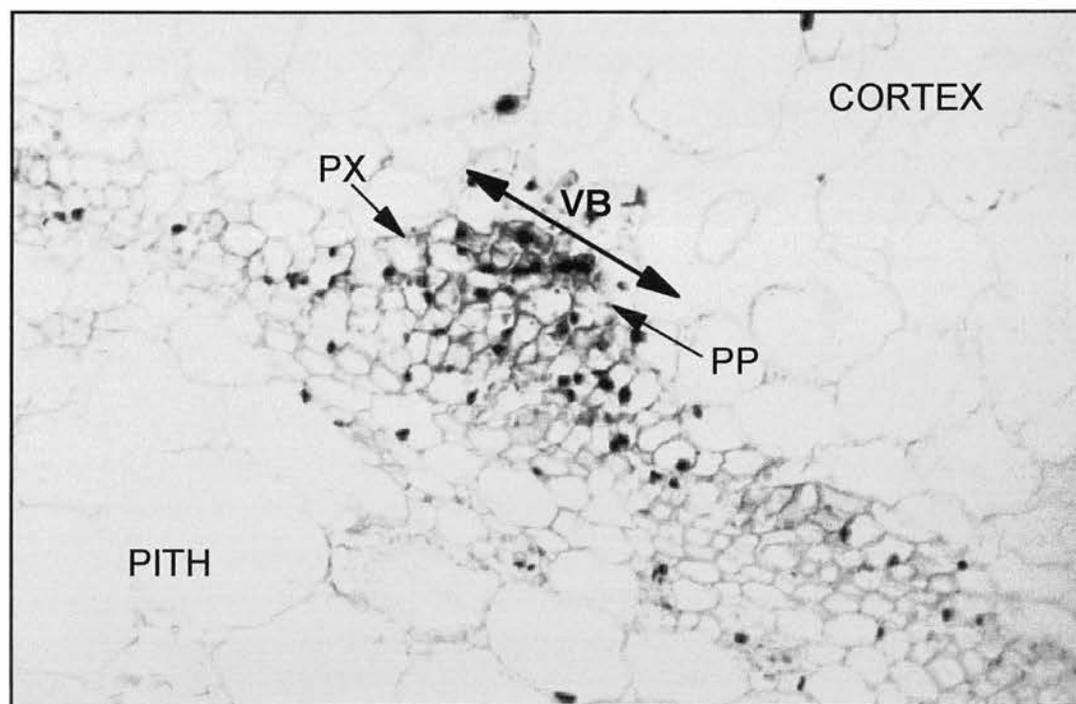


Fig.6.17. Microscopy of stem cell transverse sections. A. Transverse stem section (x67) of GLB2Apro transformant viewed under light at 365nm. XE- Xylem element. B. Transverse stem section (x40) of wt *Nicotiana tabacum* stained with H&E (stains the nuclei blue). VB-Vascular Bundle, PP-Primary Phloem, PX-Primary Xylem.

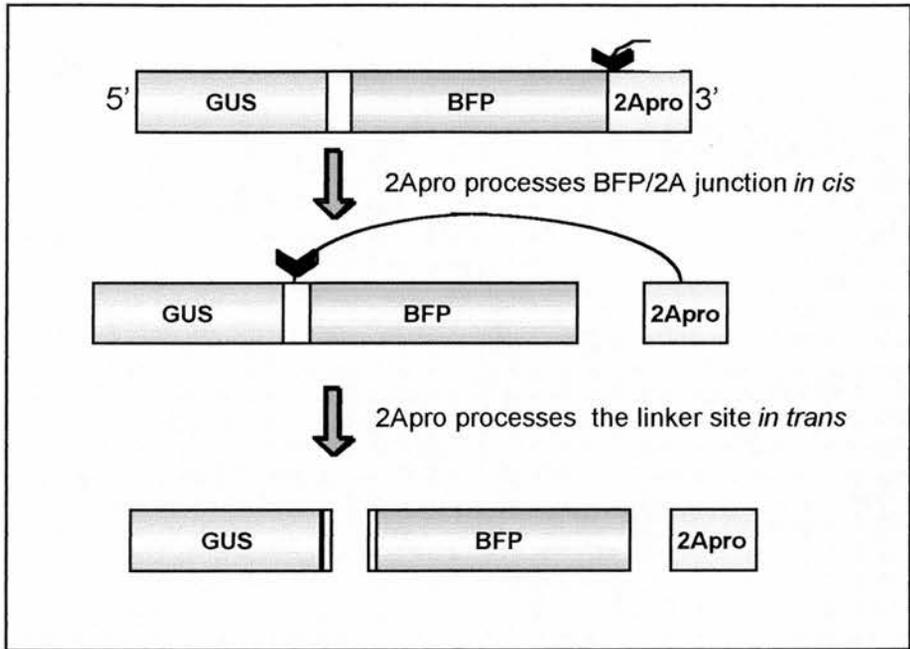
6.1.13. Conclusions

Binary vectors encoding an artificial self-processing dual-reporter system were constructed. The transformation of *Agrobacterium tumefaciens* with the binary vector series was optimized to identify the optimum chloramphenicol concentration for selection of transformed *Agrobacterium*. Kanamycin was also used as an alternative selective antibiotic due to the leaky nature of the CaMV promoter region upstream of the kanamycin resistance gene. The *Agrobacterium* were efficiently transformed with the pPZP.BFP and pPZP.GLB2A^{PRO} vectors but difficulties were experienced with *Agrobacterium* transformation using the pPZP.Lb^{PRO}BLG vector this will be discussed in greater detail in section 6.2.

Nicotiana tabacum seedlings were successfully transformed with the pPZP family of vectors to produce plants expressing the artificial self-processing polyproteins. The yield of transformed plants was low using this method. The main problem encountered was that the seedlings were quite large at the transformation stage and it was difficult to maintain contact of the seedling with the selective media thus the selective pressure on the plants was low hence the high number of false positives. This problem may have been alleviated in part by transforming the seedlings at an earlier stage and transferring the seedlings more frequently to fresh media. A larger number of the original seedlings should have been transferred to individual growth pots to increase the final yield of transformants.

The initial analysis of the putative transformants identified the reporter protein β -glucuronidase in four of the GLB2A^{PRO} plants labelled GLB2A 1,2,3 and 4 thereby confirming that these plants were expressing the GLB2A^{PRO} polyprotein. The presence of the GUS protein does not indicate if the polyprotein has been processed by the HRV14 2A protease as the GUS enzyme is active with a C-terminal extension (refer to section 5.2.3.). Western blot analysis of these plants with antibodies to both reporter proteins confirmed the presence of the GUS cleavage product and the BFP cleavage

product. The western blot analysis did not identify any larger proteins corresponding to precursor polypeptides. Thus using this detection system we can conclude that the GLB2A^{PRO} polyprotein is fully processed *in planta* to yield the final products GUS and BFP. The HRV14 2A protease processes the *cis* and *trans* cleavage sites with



equal efficiency.

Fig. 6.18. Proposed GLB2A^{PRO} polyprotein processing. The full-length polyprotein is processed *in cis* by HRV14 2A^{PRO} at the BFP/2A junction to yield the HRV14 2A protease and the GUS-L-BFP precursor polypeptide. The final step is a *trans* cleavage by HRV14 2A protease within the linker region to yield the final products BFP and GUS.

The initial GUS assay to identify positive Lb^{PRO}BLG transformants failed to identify any plants containing the GUS enzyme. The possible conclusions from this are that none of the plants were expressing the Lb^{PRO}BLG polyprotein or that the Lb^{PRO}BLG polyprotein is being expressed and the GUS enzyme is inactive and thus incapable of using the X-GlcA substrate. The final possibility is that premature termination has occurred either upstream or within the GUS gene resulting in a shortened GUS product or no GUS product. From the results of the GUS assay it was not possible to determine which of these possible explanations was the most probable. Western blot analysis with antibodies to the reporter proteins confirmed the presence of the BFP

product indicating that the polyprotein was produced in the plants and that the polyprotein was fully processed as larger precursor products were not detected. Western blot analysis with the GUS antibodies failed to detect any GUS product therefore either GUS is not being produced and translation is terminating after BFP or the levels of GUS are too low to detect with this system.

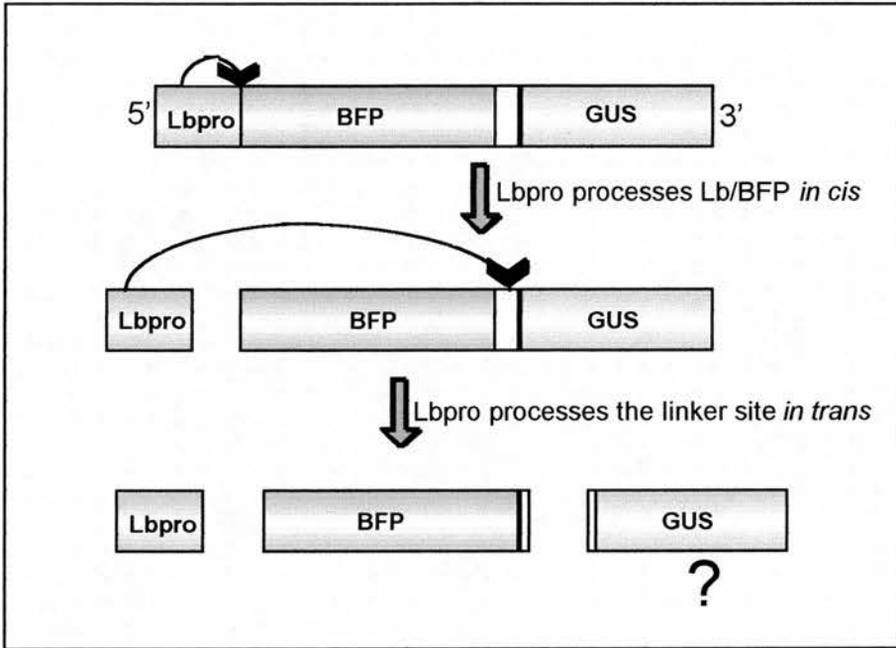


Fig. 6.19. Proposed Lb^{PRO}BLG polyprotein processing. The Lb^{PRO}BLG polyprotein is processed in cis at the Lb^{PRO}/BFP junction to produce Lb^{PRO} and the precursor polypeptide BFP-L-GUS. The BFP-L-GUS precursor is subsequently processed in trans by Lb^{PRO} to produce the final BFP and GUS products.

The crude method used to visualize BFP fluorescence in the root extracts of plants expressing BFP supports the theory that it may be possible to identify positive transformants at an early stage by viewing the roots under UV-illumination. This would require an improved source of UV light to allow excitation at the optimum wavelength.

6.1.14. Summary

- The pPZP series of binary vectors containing the dual reporter polyprotein cassettes were constructed.
- The LBA4404 strain of *Agrobacterium tumefaciens* was transformed with the PZP.polyprotein vectors.
- Transgenic *Nicotiana tabacum* seedlings expressing the artificial dual reporter polyproteins were produced via *Agrobacterium*-mediated transformation.
- The expressed GLB2A^{PRO} polyprotein was processed *in cis* and *in trans* by the HRV14 2A protease to yield the cleavage products BFP and GUS *in planta*.
- The expressed Lb^{PRO}BLG polyprotein was processed *in cis* by FMDV Lb protease to yield the BFP cleavage product. The second cleavage product β -glucuronidase was not detected by western blotting or enzyme assay.

6.2. Part 2: Investigation of the toxic effect of the FMDV Lb protease to *Agrobacterium tumefaciens*.

6.2.1. Introduction.

The difficulties associated with the transformation of *Agrobacterium tumefaciens* with the pPZP.Lb^{PRO}BLG binary vector indicated that the presence of the construct may not be tolerated by the bacteria. These problems were not encountered with the pPZP.GLB2A^{PRO} construct ruling out the vector backbone and both reporter genes as the source of the problem. The most obvious suspect is the FMDV Lb protease.

All the transformation experiments using pPZP.Lb^{PRO}BLG had yielded only two colonies, the DNA from colony 2 had been extracted and verified as pPZP.Lb^{PRO}BLG by restriction digest analysis. The possibility that the protease was inactive in this vector was ruled out by subcloning the polyprotein-encoding fragment into a suitable pGEM vector and analyzing the translation product *in vitro* (refer to section 6.1.7). Thus this did not explain why these two colonies had survived.

6.2.2. Investigation of the transformation efficiency of the pPZP.Lb^{PRO}BLG plasmid series.

A simple experiment was designed to investigate if the transformation efficiency of the pPZP.Lb^{PRO}BLG binary vectors extracted from the transformed *Agrobacterium* was greater than the original pPZP.Lb^{PRO}BLG binary vector. Competent *Agrobacterium tumefaciens* were prepared by the freeze-thaw method and transformed with 1µg of vector DNA. A positive control transformed with the pPZP111 vector and a negative control of non-transformed *Agrobacterium* were included. A 200µl aliquot of each transformation was plated onto LB agar plates containing 10µg/ml

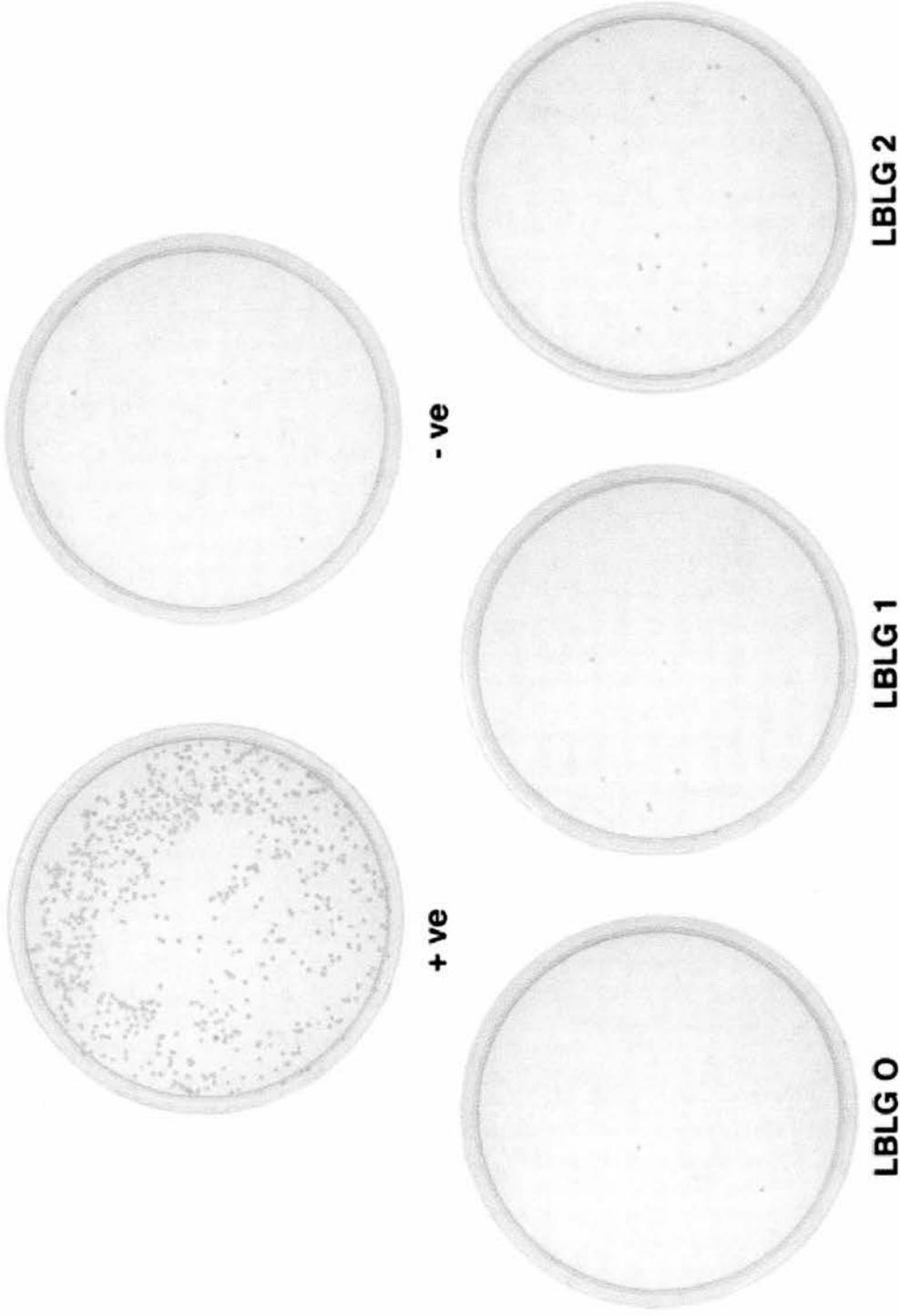


Fig.6.20. Agrobacterium transformation efficiency experiment. Aliquots of competent agrobacterium were transformed with $1\mu\text{g}$ of DNA and plated onto LB agar plates containing $25\mu\text{g/ml}$ Kanamycin. The positive control was transformed with the original pPZP111 binary vector. The negative control was untransformed agrobacterium. LBLG0 is the original pPZP.LbproBLG vector. LBLG 1 and 2 are the pPZP.LbproBLG vector DNA extracted from transformed agrobacterium.

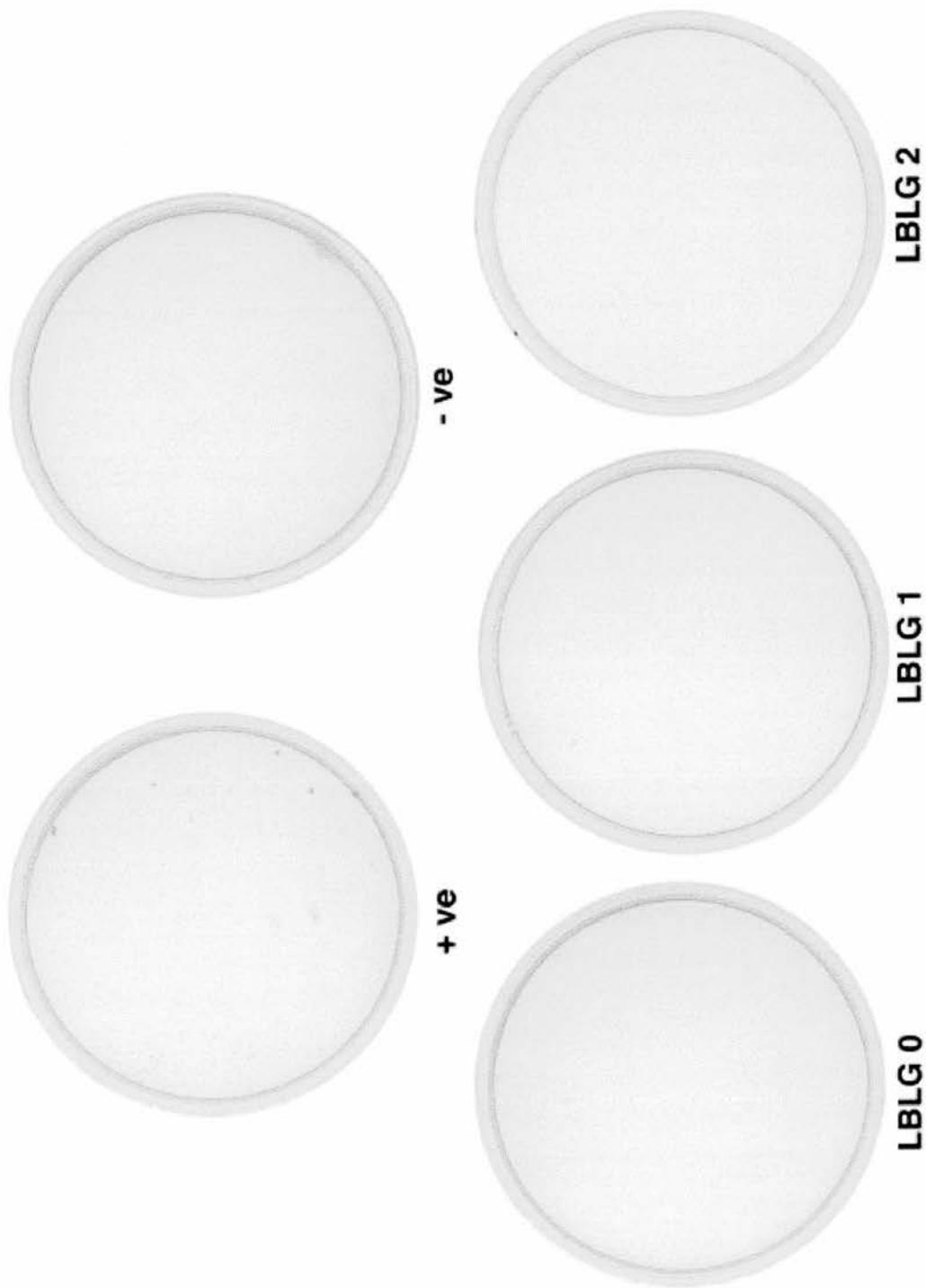


Fig. 6.21. Agrobacterium transformation efficiency experiment. Aliquots of agrobacterium were transformed with $1\mu\text{g}$ of DNA and plated onto LB agar plates containing $10\mu\text{g/ml}$ of Chloramphenicol. The positive control was transformed with the pPZP111 vector. The negative control was non-transformed agrobacterium. LBLG0 was the original PZP.LbproBLG plasmid, LBLG 1 and 2 are the LBLG plasmid extracted from transformed agrobacterium tumefaciens.

chloramphenicol or 25µg/ml of kanamycin. The plates were grown at 28°C for 2-3 days to allow the colonies to develop.

The chloramphenicol selective set of plates is shown in Fig.6.21. The positive control plate contained 11 colonies the remaining plates did not contain any colonies.

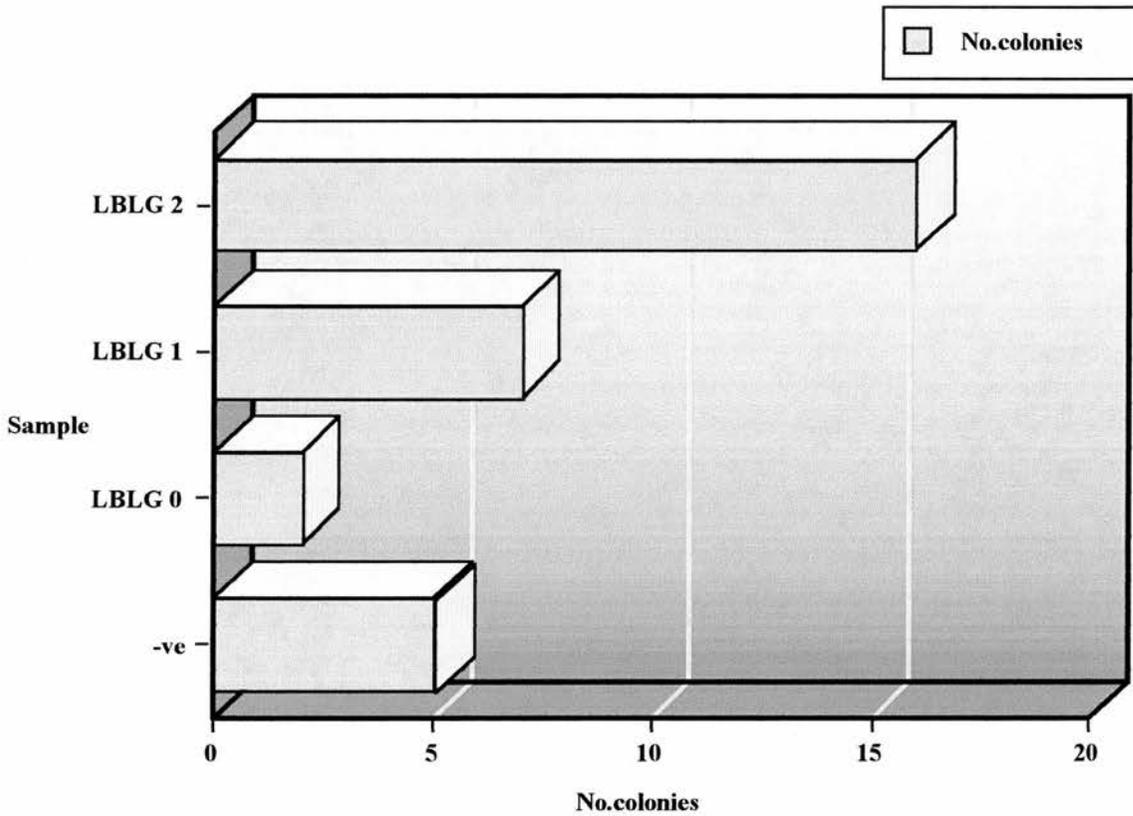


Fig.6.22. Graph of the *Agrobacterium* colony number on 25µg/ml kanamycin LB agar plates.

6.2.3. Further analysis of the plasmid DNA extracted from pPZP.Lb^{PrO}BLG transformed *Agrobacterium tumefaciens* .

The improved transformation efficiency of the pPZP.Lb^{PrO}BLG vector DNA extracted from *Agrobacterium* indicated that a change had occurred within the DNA. The pPZP.Lb^{PrO}BLG cassette from colony 2 had been moved to a pGEM vector to create pGEM.Lb^{PrO}BLG2 DNA sequencing and *in vitro* transcription and translation

analysis had proved that the protease was still active. Therefore mutation of the protease could not account for the altered transformation efficiency of pPZP.Lb^{PRO}BLG2.

An alternative hypothesis that the promoter region had been altered to lower the expression levels of the artificial polyprotein was proposed. To test this hypothesis a reverse sequencing primer was designed within FMDV Lb^{PRO} to allow sequencing of the region upstream of the polyprotein.

Odell *et al.* identified two regions of the DNA sequence important for the activity of the CaMV promoter (Odell, Nagy and Chua, 1985). The first region composed of nucleotides 0 to -46 was critical for the expression of transcripts from the CaMV promoter although the expression level of transcripts from this region was greatly reduced. The second region termed the “enhancer” region extends from nucleotide -46 to -105. Inclusion of the enhancer region dramatically increased the level of expression of transcripts.



Fig. 6.23. Alignment of the CaMV promoter regions of the PZP.Lb^{PRO}BLG binary vectors.

Sequence analysis of the CaMV promoter regions from the pPZP.Lb^{PRO}BLG 1 and pPZP.Lb^{PRO}BLG 2 vectors identified several mutations within these regions. The pPZP.Lb^{PRO}BLG 1 vector contains two mutations within the enhancer region at positions -64 (A→C) and -102 (G→A). The pPZP.Lb^{PRO}BLG 2 vector contains a single mutation within the critical region at position -42 (G→C) and two mutations

within the enhancer region at positions -64 (A→T) and -84 (A→C). It is interesting to note that both promoter regions contain a mutation at position -64.

The FMDV Lb^{pro} region of pPZP.Lb^{pro}BLG 1 vector had not been sequenced previously as this vector had not been used for the plant transformation study. The vector DNA was submitted for automated sequencing using the BFP-N sequencing primer. The sequencing results were surprising the FMDV Lb^{pro} sequence was correct initially but approximately 200 nucleotides into the sequence there was an abrupt change and the sequence after this point was not FMDV Lb^{pro}. The sequencing gel was checked to verify that this was not caused by a problem with the tracking of the gel analysis software. The sequence was submitted for BLAST analysis which confirmed the previous observations. Thus an insertion has occurred within the FMDV Lb^{pro} region of pPZP.Lb^{pro}BLG.

6.2.4. Investigation of the DNA insertion within FMDV Lb^{pro} in pPZP.Lb^{pro}BLG1.

The initial sequencing results of pPZP.Lb^{pro}BLG 1 identified an insertion within the FMDV Lb^{pro}, from the sequence the insertion was at least 400 nucleotides in length. In order to ascertain the approximate size of the insertion the FMDV Lb^{pro} region from pPZP.Lb^{pro}BLG 1 was amplified by the PCR using primers that annealed to the 5' and 3' termini of the FMDV Lb^{pro}. The FMDV Lb^{pro} region of the original pPZP.Lb^{pro}BLG vector was also amplified as a control (refer to Fig.6.24.). FMDV Lb protease is approximately 550bp in length. The fragment amplified from the pPZP.Lb^{pro}BLG vector was approximately 1900bp therefore the insertion within FMDV Lb^{pro} must be approximately 1350bp in total.

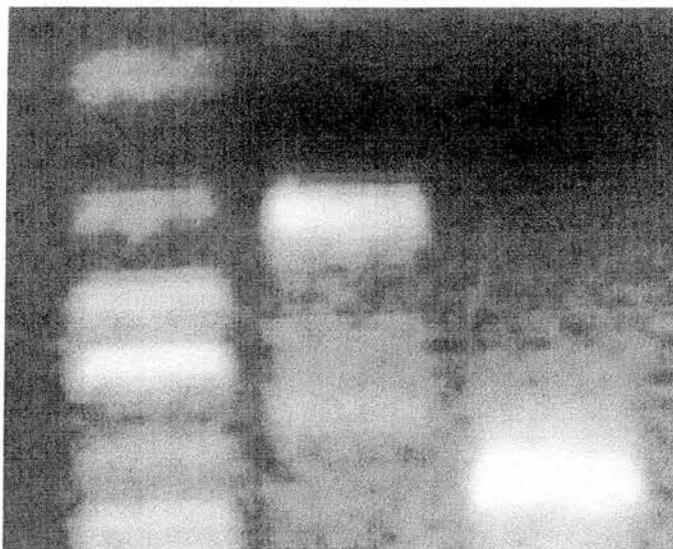


Fig. 6.24. Amplification of the FMDV Lb^{PRO} region by the PCR. Lane 1 – 8 μ l 1Kb DNA ladder, Lane 2 – 10 μ l template pPZP.Lb^{PRO}BLG, Lane 3 – 10 μ l template pPZP.Lb^{PRO}BLG 2.

The insertion within Lb^{PRO} of pPZP.Lb^{PRO}BLG1 was completely sequenced by a series of short primers designed to anneal towards the end of the previous sequence. The sequencing results were linked together until the complete sequence of the insertion was composed. The complete sequence of the insertion was submitted for analysis using the BLAST database (www.ncbi.nlm.nih.gov/BLAST/). The highest scoring sequences were downloaded. The protein sequences were aligned with the insertion sequence using the computer program ClustalX and formatted with SeqVu 1.1. (refer to Fig.6.25). The insertion sequence within Lb^{PRO} was identified as the right border insertion element of the transposon Tn10. The encoded protein is the transposase that promotes the transposition of Tn10.

Insert	1	MCELDILHDSLYQFCPELHLKRLNSLTLACHALLDCKTL
<i>S.flexneri</i>	1	MCELDILHDSLYQFCPELHLKRLNSLTLACHALLDCKTL
5' Tn10	1	MCELDILHDSLYQFCPELHLKRLNSLTLACHALLDCKTL
3' Tn10	1	MCELDILHDSLYQFCPELHLKRLNSLTLACHALLDCKTL
Insert	40	TLTELGARNLPTKARTKHNIKRIDRLLGNRHLHKERLAVY
<i>S.flexneri</i>	40	TLTELGARNLPTKARTKHNIKRIDRLLGNRHLHKERLAVY
5' Tn10	40	TLTELGARNLPTKARTKHNIKRIDRLLGNRHLHKERLAVY
3' Tn10	40	TLTELGARNLPTKARTKHNIKRIDRLLGNRHLHKERLAVY
Insert	79	RWHASFIGSGNTMPIVLVDWSDIREQKRLMVLRASVALH
<i>S.flexneri</i>	79	RWHASFIGSGNTMPIVLVDWSDIREQKRLMVLRASVALH
5' Tn10	79	RWHASFIGSGNTMPIVLVD[S]SDIREQKRLMVLRASVALH
3' Tn10	79	RWHASFIGSGNTMPIVLVDWSDIREQKRLMVLRASVALH
Insert	118	GRSVTLYEKAFFPLSEQCSKKAHDQFLADLASILPSNTTP
<i>S.flexneri</i>	118	GRSVTLYEKAFFPLSEQCSKKAHDQFLADLASILPSNTTP
5' Tn10	118	GRSVTLYEKAFFPLSEQCSK
3' Tn10	118	GRSVTLYEKAFFPLSEQCSKKAHDQFLADLASILPSNTTP
Insert	157	LIVSDAGFKVPWYKSVEKLGWYWLSRVRGKVQYADLGAE
<i>S.flexneri</i>	157	LIVSDAGFKVPWYKSVEKLGWYWLSRVRGKVQYADLGAE
3' Tn10	157	LIVSDAGFKVPWYKSVEKLGWYWLSRVRGKVQYADLGAE
Insert	196	NWKPI SNLHDMSSSHSKTLGYKRLTKSNPISCQILLYKS
<i>S.flexneri</i>	196	NWKPI SNLHDMSSSHSKTLGYKRLTKSNPISCQILLYKS
3' Tn10	196	NWKPI SNLHDMSSSHSKTLGYKRLTKSNPISCQILLYKS
Insert	235	RSKGRKNQRSTRTHCHHPSPKIYSASAKEPWVILATNLPV
<i>S.flexneri</i>	235	RSKGRKNQRSTRTHCHHPSPKIYSASAKEPWVILATNLPV
3' Tn10	235	RSKGRKNQRSTRTHCHHPSPKIYSASAKEPWVILATNLPV
Insert	274	EIRTPKQLVNIYSKRMQIEETFRDLKSPAYGLGLRHSRT
<i>S.flexneri</i>	274	EIRTPKQLVNIYSKRMQIEETFRDLKSPAYGLGLRHSRT
3' Tn10	274	EIRTPKQLVNIYSKRMQIEETFRDLKSPAYGLGLRHSRT
Insert	313	SSSERFDIMLLIALMLQLTCWLAGVHAQKQGWDKHFQAN
<i>S.flexneri</i>	313	SSSERFDIMLLIALMLQLTCWLAGVHAQKQGWDKHFQAN
3' Tn10	313	SSSERFDIMLLIALMLQLTCWLAGVHAQKQGWDKHFQAN
Insert	352	TVRNRNVLSTVRLGMEVLRHSGYITREDSLVAATLLAQ
<i>S.flexneri</i>	352	TVRNRNVLSTVRLGMEVLRHSGYITREDSLVAATLLAQ
3' Tn10	352	TVRNRNVLSTVRLGMEVLRHSGYITREDSLVAATLLAQ
Insert	391	NLFTHGYALGKLVZGDLALC
<i>S.flexneri</i>	391	NLFTHGYV LGKLV
3' Tn10	391	NLFTHGYALGKLV

Fig. 6.25. Alignment of the ORF of the insertion sequence with *Shigella flexneri* Tn10 (Accession No. AF162223) and the 5' (Accession No. JO1827) and 3' (Accession No. JO1829) ends of IS10.

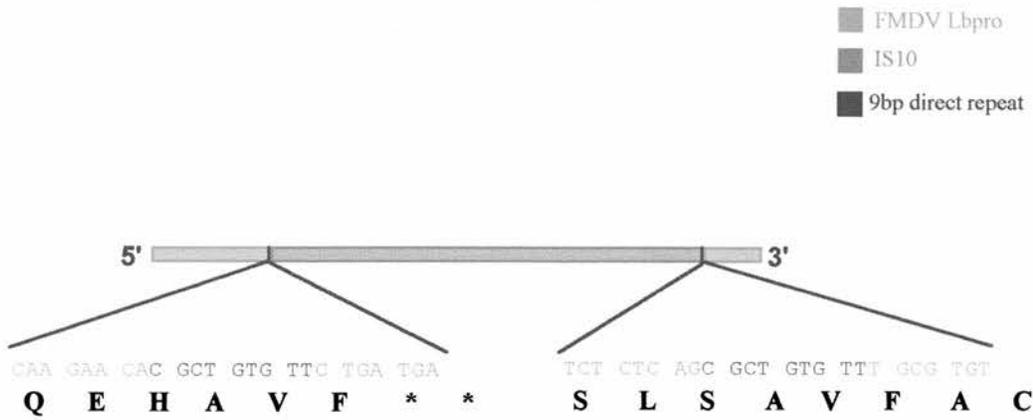


Fig. 6.26. The FMDV Lb^{pro} region from PZP.Lb^{pro}BLG1. The Lb^{pro} region is shown in orange and the inserted IS10 region in green. The duplicated 9bp repeat is highlighted in blue.

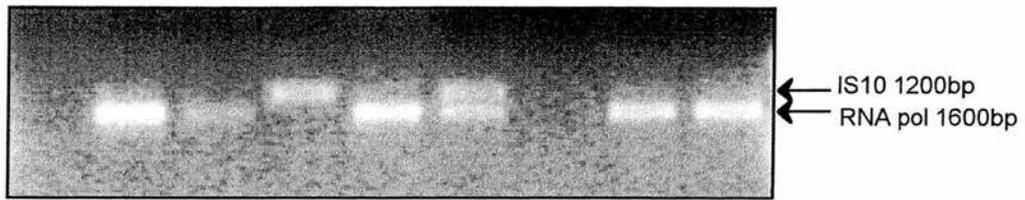
The insertion site of the IS10 element within FMDV Lb^{pro} was identified. The insertion occurred at the active site histidine residue CA*C codon. The transposition event generated a 9bp duplicated repeat thus the inserted DNA was flanked by a 9bp repeat sequence shown in blue in Fig. 6.26.

6.2.5. Investigation of the origin of IS10.

The insertion sequence within FMDV Lb^{pro} of pPZP.Lb^{pro}BLG was sequenced and positively identified as the IS10 transposase from the Tn10 transposon. Therefore the question arose as to the origins of the IS10 transposase. The original pPZP.Lb^{pro}BLG did not contain this insertion thus as the pPZP.Lb^{pro}BLG1 plasmid had been extracted from the transformed *Agrobacterium* the most obvious theory was that the IS10 had originated from the LBA4404 *Agrobacterium tumefaciens*. In order to investigate this theory two sets of oligonucleotide PCR primers were designed to amplify the IS10 transposase and a control encoding the *Agrobacterium* RNA polymerase I gene. The DNA from 1ml of an overnight culture of stock LBA4404 *Agrobacterium* and *Agrobacterium* transformed with pPZP.Lb^{pro}BLG 1 and pPZP.Lb^{pro}BLG 2 was extracted using the Wizard® Genomic DNA extraction kit (Promega). A sample of the extracted DNA was used as a template for a set of three PCR reactions using the IS10 transposase primers, the control primers or a mixture of

both primer sets. A 15 μ l aliquot of each reaction was separated by gel electrophoresis on a 1% agarose gel and visualized on a UV transilluminator. The results are shown in Fig. 6.27.

The *Agrobacterium* RNA polymerase I control was amplified from all three DNA templates but the IS10 transposase was only amplified from DNA extracted from pPZP.Lb^{PRO}BLG 1 transformed *Agrobacterium*. Therefore the LBA4404 *Agrobacterium* does not contain the Tn 10 transposon.



Lane	Template <i>A. tumefaciens</i> genome				PCR primer pair	
	LBA4404	LBA4404 + Lb ^{PRO} BLG1	LBA4404 + Lb ^{PRO} BLG2	IS10	A. tumefaciens control	IS10 + control
1	✓			✓		
2	✓				✓	
3	✓					✓
4		✓		✓		
5		✓			✓	
6		✓				✓
7			✓	✓		
8			✓		✓	
9			✓			✓

Fig. 6.27. Analytical PCR of *Agrobacterium tumefaciens* populations. Refer to the table for the template and primer used in each reaction.

The Tn10 transposon could not be amplified from the stock LBA4404 *Agrobacterium* therefore an alternative theory was proposed that the pPZP.Lb^{PRO}BLG vector DNA

used to transform the *Agrobacterium* could contain a mixed DNA population thereby containing the vector with and without the IS10 transposase.

A series of analytical PCR reactions with the original pPZP.Lb^{PRO}BLG vector or the pPZP.Lb^{PRO}BLG 1 vector DNA as template DNA. The IS10 transposase primers and the *Agrobacterium* RNA polymerase I primers were used in the PCR reactions. A 15µl sample of each PCR reaction was separated by gel electrophoresis on a 1% agarose gel and viewed under UV light, the results are shown in Fig.6.28.

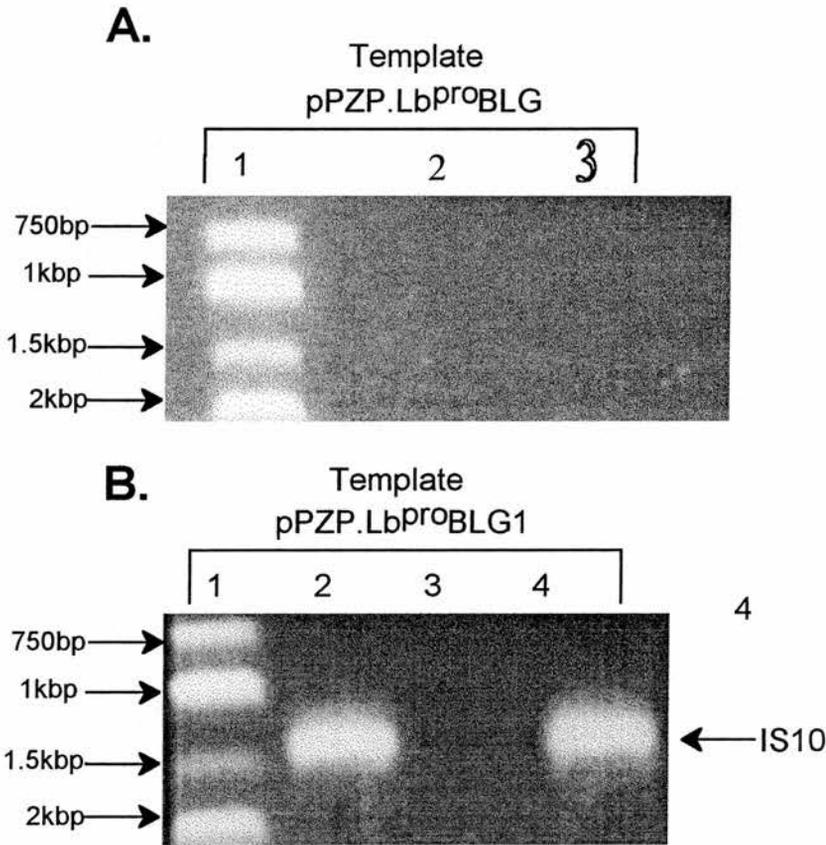


Fig.6.28. Analytical PCR of vector DNA. A. The template used in the PCR was the original pPZP.Lb^{PRO}BLG vector. Lane 1- 8µl 1kb DNA ladder, Lane 2- 15µl of PCR with IS10 Primers, Lane 3- 15µl of PCR with *Agrobacterium* control primers, Lane 4- 15µl of PCR with IS10 + control primers. B. The template used was the pPZP.Lb^{PRO}BLG1 vector extracted from *Agrobacterium*. Lane 1- 8µl 1kb DNA ladder, Lane 2- 15µl of PCR with IS10 Primers, Lane 3- 15µl of PCR with *Agrobacterium* control primers, Lane 4- 15µl of PCR with IS10 + control primers.

The *Agrobacterium* RNA polymerase I was not amplified in any of the PCR reactions indicating that the DNA was not contaminated by *Agrobacterium*. The IS10

transposase was only amplified from the pPZP.Lb^{PRO}BLG 1 vector DNA thus the original pPZP.Lb^{PRO}BLG vector DNA does not appear to contain a mixed population of DNA.

The initial experiment to isolate the IS10 transposase from the *Agrobacterium tumefaciens* stock population failed. The Tn10 transposon carries tetracycline resistance. Therefore in case the stock LBA4404 *agrobacterium* contained a mixed population 100µl aliquots of an overnight *Agrobacterium* culture were plated onto a set of agar plates containing a range of tetracycline concentrations (2.5µg/ml, 5µg/ml, 10µg/ml, 25µg/ml) to select any Tn10-containing *Agrobacterium*. The XL1-Blue strain of *E.coli* carries Tn10 on the F' plasmid. The presence of IS10 within XL1-Blue bacteria was confirmed by the PCR using the IS10 and *Agrobacterium* control primer sets (refer to Fig.6.29). Aliquots of an overnight culture of XL1-Blue *E.coli* were also plated onto a duplicate set of Tetracycline-containing agar plates. The *Agrobacterium* plates were incubated at 28°C and the *E.coli* plates were incubated at 37°C. The XL1-Blue *E.coli* grew on all the concentrations of tetracycline tested. In contrast, there were no *Agrobacterium* colonies on any of the tetracycline-containing plates.

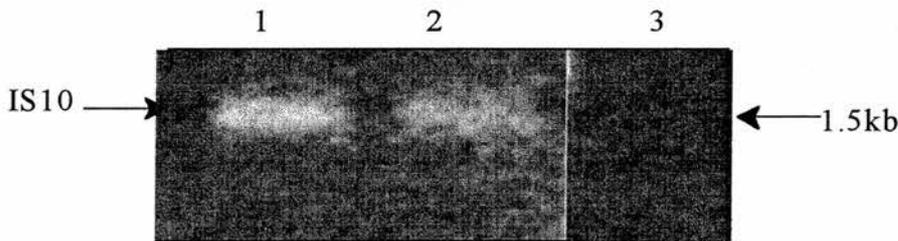


Fig.6.29. Analytical PCR of the *E.coli* XL1-Blue strain genome. Lane 1- 15µl of PCR using IS10 primers, Lane 2- 15µl of PCR using IS10 + control primers, Lane 3- 15µl of PCR using *Agrobacterium* control primers.

6.2.6. Conclusions.

The difficulties experienced with the transformation of *Agrobacterium tumefaciens* with pPZP.Lb^{PRO}BLG outlined in section 6.1.6. led to the conclusion that the FMDV Lb^{PRO} was not tolerated by *Agrobacterium*. The *Agrobacterium* transformation experiments did yield two colonies, the vector DNA was extracted directly from the

Agrobacterium and identified as pPZP.Lb^{PRO}BLG. The Lb protease of pPZP.Lb^{PRO}BLG 2 had been confirmed as active prior to the transformation of the *Nicotiana tabacum* seedlings. The question of whether other changes had occurred within the pPZP.Lb^{PRO}BLG vector that allowed these colonies to survive remained. The improved transformation efficiency of the *Agrobacterium*-extracted vectors in comparison to the original DNA suggested that a change had occurred that reduced the toxicity of the vector to the *Agrobacterium*.

Analysis of the DNA sequence of the CaMV promoter regions of the *Agrobacterium*-extracted vectors identified several mutations. The pPZP.Lb^{PRO}BLG2 vector contained a mutation within the critical region of the promoter this is the region that is vital for expression. There were also mutations within the enhancer region of the CaMV promoter for both vectors most notably at position 64 that had been altered in both cases. These results suggest that the altered CaMV promoter regions lowered the expression level of the polyprotein to a sufficient level that could be tolerated by the *Agrobacterium*. In light of this information the problems associated with the detection of the GUS product from the transformed Lb^{PRO}BLG plants may be explained as the expression levels of the polyprotein in the plant would also be significantly reduced.

The sequence analysis of pPZP.Lb^{PRO}BLG 1 detected another major change in this vector that explained why this *Agrobacterium* colony survived. A large (1350bp) DNA fragment had been inserted into the Lb protease effectively knocking out expression of the protease. The DNA insertion was sequenced and identified as the right IS10 element from the Tn10 transposon. The right IS10 element encodes the transposase protein that is responsible for transposition of the Tn10 or IS10. The transposition site was identified as the active site histidine residue of Lb^{PRO}, during the transposition process a 9-bp sequence originally within Lb^{PRO} had been duplicated and flanked the inserted DNA. The source of the IS10 element was initially presumed to be the *Agrobacterium* but the IS10 could not be amplified from the *Agrobacterium* genome. The second potential source was the original DNA used in the transformation but

attempts to amplify IS10 from this source also failed thus the origins of the IS10 element could not be identified.

The results of the investigation of the pPZP.Lb^{PrO}BLG vectors extracted from *Agrobacterium* has identified in each case discrepancies in the vector DNA that permitted the survival of the transformed *Agrobacterium*. The pPZP.Lb^{PrO}BLG 1 colony survived because Lb^{PrO} expression was knocked out by the transposition of IS10. The mutations within the promoter region of pPZP.Lb^{PrO}BLG2 reduced the level of expression of FMDV Lb^{PrO} to a level tolerated by the *Agrobacterium*. These results strengthen the previous observations that FMDV Lb^{PrO} is toxic to *Agrobacterium* thus an alternative method of transformation bypassing the requirement for *Agrobacterium* would be required to investigate the potential of using this artificial polyprotein system *in planta*.

6.2.7. Summary

- The FMDV Lb protease is not tolerated by *Agrobacterium tumefaciens*.
- The transformation efficiency of the pPZP.Lb^{PRO}BLG vectors extracted from transformed *Agrobacterium* was higher than the original pPZP.Lb^{PRO}BLG vector.
- The CaMV promoter regions of PZP.Lb^{PRO}BLG extracted from *Agrobacterium* contained mutations.
- The pPZP.Lb^{PRO}BLG 2 promoter region contained a single point mutation in the critical region of the CaMV promoter.
- The pPZP.Lb^{PRO}BLG 1 and 2 promoter regions contained mutations in the enhancer region. A single point mutation at position 64 was observed in both vectors.
- A 1350bp DNA insertion was identified in the FMDV Lb^{PRO} of pPZP.Lb^{PRO}BLG1. The inserted DNA was sequenced and identified as the right border IS10 transposase from the Tn10 transposon.
- The IS10 transposase could not be amplified from the original DNA or the *Agrobacterium tumefaciens* LBA4404 thus the origin of IS10 is inconclusive.

CHAPTER 7: DISCUSSION.

7.1.Introduction.

The primary aim at the beginning of this project was to develop artificial self-processing polyprotein systems to allow the expression of multiple proteins from a single transcription unit. The processing activity of two alternative picornavirus proteases HRV14 2A^{PRO} and FMDV Lb^{PRO} was tested. The main body of this thesis describes the construction of artificial self-processing polyprotein systems and the expression of the polyproteins in various *in vitro* and *in vivo* systems. Throughout the course of the study a number of interesting secondary questions arose which will be addressed briefly prior to the main discussion regarding the artificial self-processing polyprotein system.

7.2. The effect of an N- or C-terminal extension on β -glucuronidase activity in *Escherichia coli*.

The observation that *E.coli* expressing the Lb^{PRO}BLG polyprotein but not the GLB2A^{PRO} polyprotein could utilize the chromogenic X-GlcA substrate suggested that the position of GUS within the polyprotein may affect the activity. An answer was not evident in the current literature thus a simple experiment was devised to investigate this further. Two alternative polyproteins GUS2AGFP and CAT2AGUS where 2A is the FMDV 2A peptide were transformed into *E.coli*. The bacteria were plated onto media containing IPTG to induce protein expression and the X-GlcA substrate to monitor GUS activity. Both constructs produced blue colonies indicating that the GUS enzyme was active thus we can conclude that GUS activity is not affected by the presence of an N- or C-terminal extension.

The expression of the FMDV 2A polyproteins in *E.coli* provided the opportunity to redress the question of FMDV 2A cleavage in bacteria. Donnelly *et al.*, (1997) concluded that FMDV 2A did not cleave in *E.coli*. They reported that western blot analysis of the total cellular proteins from *E.coli* expressing pUC:CAT2AGUS with anti-GUS antisera confirmed the presence of the full-length CAT2AGUS polyprotein but no GUS cleavage product was detected. An attempt had been made to probe with anti-CAT antisera but cross-reactivity with *E.coli* proteins had proved problematic. Therefore the expression of the polyprotein pTrc.GUS2AGFP in *E.coli* presented the opportunity to monitor the translation products using GUS and GFP antisera that were known not to cross-react with *E.coli* proteins. Western blot analysis with GFP antisera identified only the full-length polyprotein GUS2AGFP, no GFP cleavage product was detected confirming that FMDV 2A does not “cleave” in *E.coli*. Two bands were detected by western blot analysis with GUS antisera, the bands corresponded to the full-length GUS2AGFP polyprotein and a smaller truncated protein identified as GUS2A. Similar results were obtained for the CAT2AGUS polyprotein only the full-length product and the primary CAT2A product were detected. Thus we can conclude that FMDV 2A is inactive in prokaryotes (in the sense of producing two “cleavage” products) and also that translation is terminated prematurely at the C-terminus of FMDV 2A in a significant proportion of translation events.

7.3. Investigation of IRES-driven *in vitro* translation.

7.3.1. IRES translation in the coupled TnT system.

Translation of uncapped and IRES-driven transcripts within this system produced some unexpected results. Translation of the uncapped transcripts was markedly more efficient than transcripts containing the IRES from FMDV or HRV14, although surprisingly HRV14 IRES transcripts were translated in unsupplemented lysate. Finally, contrary to previous reports that wheatgerm extract does not support IRES-mediated translation the IRES-transcripts were translated in wheatgerm extract. An explanation for these unusual results was the use of the coupled transcription and translation system. Reports of *in vitro* analysis of IRES constructs within the literature produce RNA in a separate reaction that is purified and subsequently used to program a translation reaction. Therefore we propose that the coupled TnT system mimics the transcription/translation process in prokaryotes where the processes are tightly coupled (refer to 1.28.), as a result the complex RNA structure composing the IRES cannot form, the presence of the ribosomes “melts” any RNA structures present. This model would explain the unusual results observed using this system.

Therefore in the coupled system all the transcripts are translated via an uncapped mechanism, the difference in the relative translation efficiency can be accounted for by the presence of several initiation codons within the IRES region upstream of the polyprotein AUG codon. Examination of the sequence of the constructs confirmed that all these initiation codons would result in very short translation products that would not be detected on the gel. The difference in the translational efficiency of the FMDV and HRV14 IRESes is due to the number of initiation codons present within the IRES.

An interesting observation of this system was that IRES-driven transcripts containing an active HRV14 2A^{Pro} were translated very poorly. The uncapped transcript was translated very well therefore the reduced translation is not caused simply by the presence of the protease. Translation of the IRES-driven transcripts containing the inactive mutant protease was not reduced indicating that the proteolytic activity of the protease is required for this effect. Why is translation being reduced so dramatically in the presence of an IRES structure and active 2A^{Pro}?

1. Does HRV14 2A^{Pro} interact with the IRES and block translation?

Macadam *et al.*, (1994) first suggested the possibility that 2A^{Pro} may interact directly with the IRES following the observation that mutations within the 5'NCR of poliovirus that destabilized the IRES were compensated by mutations within 2A^{Pro}. There was no direct evidence for an interaction between the IRES and protease. A recent paper from the same researchers concludes that there is no evidence to support such an interaction (Rowe *et al.*, 2000). Roberts *et al.*, (1998) suggest that the initial observation that mutations within 2A^{Pro} compensated for an unstable IRES was due to improved 2A^{Pro} that cleaved eIF4G more efficiently and not an interaction with the IRES. The IRES is not formed within the coupled transcription/translation systems used here: thus it is highly improbable that the IRES is interacting with 2A^{Pro}.

2. Is 2A^{Pro} inhibiting at the level of transcription not translation?

Davies *et al.*, (1991) first reported that poliovirus 2A^{Pro} is a potent inhibitor of transcription, expression of 2A^{Pro} reduced the level of mRNA within the cell by 25-fold. This evidence has been supported by more recent data from Ventoso *et al.*, (1998) whom reported that expression of poliovirus 2A^{Pro} blocked transcription from a nuclear plasmid but not from a T7 Pol cytoplasmic plasmid. Therefore it seems unlikely that 2A^{Pro} affects T7 Pol transcription within the *in vitro* system.

3. Is inhibition involved with eIF4G cleavage?

Another possibility is that the inhibitory effect of 2A^{Pro} is mediated via the cellular cleavage target eIF4G that is known to bind to the IRES. This explanation does not seem very plausible as 2A^{Pro} is widely attributed to stimulate cap-independent and uncapped translation. Also similar results would be expected for the IRES-Lb^{Pro} constructs as FMDV Lb^{Pro} also cleaves eIF4G.

4. 2A^{Pro} and aberrant internal initiation products.

Yu and Lloyd (1991) reported that translation of poliovirus 2A^{Pro} constructs yielded a high amount of internal initiation products. An interesting observation was that the amount of internal initiation products was inversely proportional to the activity of 2A^{Pro}. The translation profiles from the *in vitro* analysis of pGFP2A^{Pro} and pGFP2A (refer to section 3.1.13.) supports this observation. Translation of the inactive GFP2A polyprotein increased both the overall yield and the species of internal initiation products.

To date the mechanism employed within the coupled *in vitro* TnT system that allows efficient translation of uncapped, non-IRES transcripts has not been elucidated. Svitkin *et al.*, (1996) however proposed a simple model for uncapped translation in RRL proposing that uncapped translation occurs due to a deficiency of RNA-binding proteins. Svitkin *et al.*, propose that ribosomes can bind mRNA in a non-sequence-specific, non cap-dependent manner but that this is prevented in normal cells by general RNA binding proteins that coat the mRNA forcing the ribosomes to bind via the 5' cap structure.

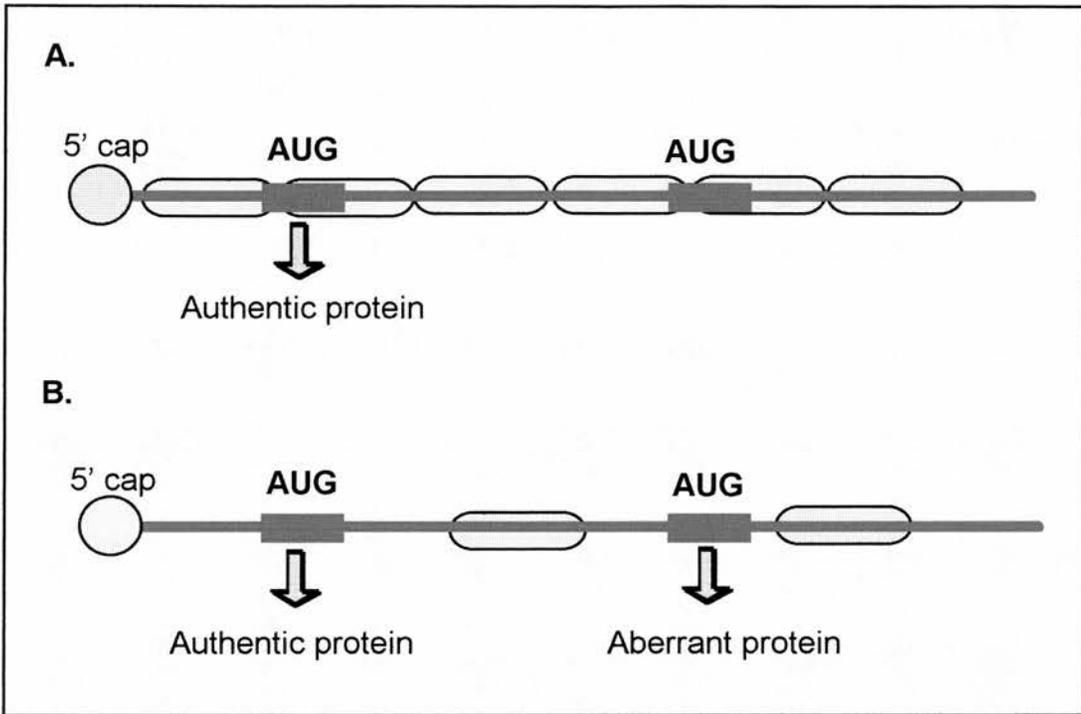


Fig. 7.1. Proposed model of uncapped translation. *A. In normal cells mRNA is coated by general RNA binding proteins preventing non-specific ribosome binding. B. In RRL, RNA binding proteins are limiting especially when levels of RNA are high therefore ribosomes can bind non-specifically.*

In RRL however, general RNA-binding proteins are limiting thus allowing uncapped translation. The observation by Yu and Lloyd (1991) that internal initiation was reduced by the addition of HeLa cell extract that is rich in RNA-binding proteins such as La protein, PTB and unr supports this theory.

The data suggests that the presence of active 2A^{Pro} acts to reduce the amount of internal initiation. Precisely how 2A^{Pro} mediates this effect is unknown but the proteolytic activity of 2A^{Pro} is required. The action of 2A^{Pro} inhibits cap-dependent translation via cleavage of the cellular target eIF4G thus it is unlikely that 2A^{Pro} would also act to promote translation from the 5' cap. At this stage we propose that 2A^{Pro} acts to promote translation from the authentic 5' initiation codon. If this hypothesis is correct then the poor translation of the IRES-GFP2A^{Pro} transcripts is due to initiation of translation at the first initiation codon within the IRES region that

produces a very short translation product not visualized on the gel. Thus this hypothesis predicts that translation of the uncapped GFP2A^{PRO} construct would be efficient because initiation at the GFP initiation codon would be favoured producing full-length polyprotein. Translation from the IRES-GFP2A constructs would also not be affected, as HRV14 2A^{PRO} is inactive in this system. The experimental data follows the predicted pattern. Finally, similar observations for the action of FMDV Lb^{PRO} on internal initiation products have not been reported and were not observed in this study (refer to section 3.1.10).

7.3.2. In vitro TnT analysis using STP3 uncoupled system.

Translation of the full complement of constructs using this system which separates transcription and translation supports the hypothesis that the IRES structure is not formed due to the tight coupling of the Promega system. In this system in accordance with the reports in the literature translation of uncapped and HRV14 IRES transcripts was very poor. The FMDV IRES in contrast, was extremely efficient at initiating translation in this system. It is well-documented that FMDV IRES functions very efficiently in RRL (Luz and Beck, 1991; Borman *et al.*, 1995) but that HRV14 IRES is inefficient at initiating translation in RRL. The HRV14 IRES requires supplementing with HeLa cell extract that supplies the RNA-binding proteins PTB (Hunt *et al.*, 1999b; Hunt and Jackson, 1999) and unr (Hunt *et al.*, 1999a) vital for HRV IRES function, these proteins are found in very low levels in RRL. The protein PCBP2 is also required for HRV IRES function but this protein is abundant in RRL (Walter *et al.*, 1999).

The presence of active FMDV Lb^{PRO} enhanced translation from the HRV14 IRES to a level greater than the corresponding FMDV IRES transcript. A translational enhancement by FMDV Lb^{PRO} was observed to a lesser extent in the coupled system. Translation from the FMDV IRES was not stimulated by the FMDV Lb protease.

Similar translational enhancement of IRES-driven transcripts was not observed for constructs containing the HRV14 2A protease.

The enhancement of translation by the entero- and rhinovirus 2A protease was initially proposed by Hambidge and Sarnow (1992), this effect was demonstrated to occur at the level of translation as the mRNA level was not altered. It is well documented that 2A^{Pro} and Lb^{Pro} cleave eIF4G thus inhibiting capped translation therefore it was suggested that enhancement of IRES-driven translation is due to diminished competition for the translational machinery. This theory has been disproved by the addition of a 4E-binding protein (4E-BP2) that also inhibits capped translation, no increase in IRES translation was observed (Borman *et al.*, 1997a; Ohlmann *et al.*, 1997; Roberts *et al.*, 1998). Therefore it was postulated that the cleavage products of eIF4G actively stimulate IRES-driven translation, this activity has since been attributed to the cpc fragment of eIF4G (Borman *et al.*, 1997a; Morino *et al.*, 2000).

The presence of active Lb^{Pro} stimulated translation from the HRV IRES transcripts but not translation from the FMDV IRES transcripts. Borman *et al.*, (1995) compared the translational efficiency of different types of picornavirus IRES. They reported that the effect of the addition of Lb^{Pro} or 2A^{Pro} was dependent on the type of IRES. Lb^{Pro} stimulated translation from type I IRESes but did not alter type II IRES translation and acted to inhibit translation from a type III IRES. The same group of researchers confirmed these results *in vivo* (Borman *et al.*, 1997b).

Contrary to numerous reports in the literature, the presence of HRV14 2A^{Pro} in the *in vitro* translation reaction did not stimulate translation from the HRV14 IRES. Stimulation of the translational efficiency of the IRES is mediated via the cleavage of the cellular target eIF4G, the cpc fragment stimulates translation. Therefore the most obvious explanation is that HRV14 2A^{Pro} does not cleave eIF4G in this system.

Examination of the current literature found that the majority of experiments monitoring the effect of 2A^{PRO} on IRES translation added recombinant 2A^{PRO} to the translation. Studies using the *in vitro* system frequently pretreated the lysate with recombinant 2A^{PRO} thereby ensuring that all the eIF4G was cleaved prior to translation (Ziegler *et al.*, 1995). Roberts *et al.*, (1998) reported the only example of co-expression of 2A^{PRO} this study was conducted *in vivo* over a period of 20 hours. They reported that 2A^{PRO} was more efficient at enhancing translation than FMDV Lb^{PRO} but that it was more difficult to detect eIF4G cleavage products using the 2A protease.

Studies by Lamphear *et al.*, monitored eIF4G cleavage by Lb^{PRO} (1995) and 2A^{PRO} (1993), the results demonstrate that Lb^{PRO} cleaves eIF4G more efficiently than 2A^{PRO}. Following addition of Lb^{PRO} complete eIF4G could not be detected within 2 minutes, in contrast, following the addition of 2A^{PRO} complete eIF4G was still detected after 20 minutes. Another interesting observation was that Lb^{PRO} continued to degrade the eIF4G cleavage products cpc and cpn. Subsequent degradation of the initial eIF4G cleavage products was not observed for the 2A^{PRO}. This may explain why Roberts *et al.*, (1998) concluded that 2A^{PRO} is more efficient at enhancing translation than Lb^{PRO} as the cpc fragment must remain intact to stimulate IRES translation.

Thus we conclude that co-expression of 2A^{PRO} *in vitro* does not produce sufficient protease to cleave sufficient eIF4G within the 60 minute incubation period to stimulate translation from the HRV IRES.

7.3.3. *Future work.*

In order to confirm the hypothesis concerning the lack of stimulation by HRV14 2A^{Pro} in the *in vitro* system it would be interesting to add recombinant HRV14 2A^{Pro} to the reactions and also to monitor the state of eIF4G within an *in vitro* reaction where 2A^{Pro} is co-expressed.

The next stage would require expression of the IRES:polyprotein constructs in a mammalian system to monitor the translation efficiency of the different IRESes and the efficiency of artificial polyprotein processing within this system. The inclusion of an IRES to the polyprotein construct is required within the mammalian system due to host-cell shutoff mediated via eIF4G cleavage. It would be interesting to investigate how the polyproteins are expressed within this system as both proteases are highly toxic to mammalian cells. The toxic effect of the picornavirus proteases was initially attributed to the cleavage of eIF4G but a recent report demonstrated that expression of inactive 2A^{Pro} that did not cleave eIF4G was still toxic to cells resulting in cell death (Barco, Feduchi and Carrasco, 2000).

7.3.4. *Conclusions.*

As a result of the tight coupling of transcription and translation we can conclude that the coupled *in vitro* TnT system is not suitable for monitoring the translation and subsequent interactions of proteins where the RNA secondary structure is important. The translational efficiency of the IRES *in vitro* is dependent on the type of IRES and the presence of cellular factors including the FMDV Lb protease.

7.4. The FMDV Lb^{PRO}- based polyprotein system.

7.4.1. *In vitro* analysis.

The simple polyprotein Lb^{PRO}GFP was translated *in vitro* in rabbit reticulocyte lysate and wheatgerm extract. In both systems the polyprotein was processed *in cis* at the Lb^{PRO}/GFP junction to yield GFP and Lb^{PRO} confirming that the FMDV Lb^{PRO} is active in these systems and does not require any other viral sequences for cleavage activity. Mutation of the active site cysteine residue to an alanine at position 53 abolished the *cis* processing activity of the protease, thus translation of pLbGFP produced a single band corresponding to the full-length polyprotein.

Densitometric analysis of the translation profile of pLb^{PRO}GFP in RRL and WGE highlighted a discrepancy in the efficiency of the *cis* cleavage reaction between the two systems. In RRL 99% of the polyprotein was processed to produce the final products, in contrast, in WGE this dropped to 84%. A simple explanation of this difference in proteolytic activity is the different environment of the cells from which wheatgerm extract and rabbit reticulocyte lysate are produced. The results indicate that WGE is not an optimal environment for efficient *cis* processing by FMDV Lb^{PRO} this may be due to a deficiency of cellular factors required for correct folding of the protease or alternatively, by the presence of factors that act to inhibit proteolytic activity.

Translation of the dual-reporter polyprotein *in vitro* in RRL found that the polyprotein was only processed at the Lb^{PRO}/BFP junction *in cis*, cleavage *in trans* within the linker region was not detected. *Trans* processing within the *in vitro* system was not expected. *Trans* cleavage assays *in vitro* require an extended incubation period, usually overnight, following the addition of RNase and cycloheximide to terminate the reaction (Yu and Lloyd, 1991). Previous attempts to monitor *trans* processing of polyproteins *in vitro* within this laboratory (E. Byrne pers.comm.) had led to the

conclusion that the level of protease produced by the *in vitro* system was not suitable for monitoring *trans* processing events and an external source of protease was required. Thus the decision was taken not to proceed with this line of investigation.

7.4.2. Expression in prokaryotes.

Western blot analysis of the total cellular extract from *E.coli* expressing the plasmid pTrc.Lb^{PRO}BFP by probing with anti-BFP antisera identified a single protein corresponding to the BFP cleavage product. The larger precursor polyprotein was not detected indicating that within the limits of this detection system cleavage of the polyprotein by FMDV Lb^{PRO} *in cis* is 100% efficient. These results confirm the reports from Kirchweger *et al.*, (1994) that FMDV Lb^{PRO} is proteolytically active in prokaryotic systems. Kirchweger *et al.*, also reported that high levels of FMDV Lb^{PRO} is toxic to the bacteria thus low level expression must be maintained.

Similar analysis of the total cellular extract of *E.coli* expressing the dual-reporter polyprotein Lb^{PRO}BLG, probing with anti-BFP and anti-GUS antisera was conducted. Both antisera identified two bands, the respective cleavage products BFP and GUS were identified as the smallest band in each case. The second larger protein detected by BFP and GUS antisera corresponded to the predicted molecular weight of the precursor polyprotein BFPLGUS. A higher molecular weight band was not identified suggesting that the full-length polyprotein is completely processed *in cis*. The production of the individual cleavage products BFP and GUS confirms that the *trans* cleavage site within the linker region is processed in the majority of proteins but that this cleavage is not as efficient as the *cis* cleavage event at the Lb^{PRO}/BFP junction.

Guarné *et al.*, (1998) recently solved the structure of the FMDV Lb protease, they propose that the long C-terminal element (CTE) of the Lb protease (refer to Fig.1.26) is flexible and folds back into the substrate-binding cleft thus promoting intramolecular

cleavage. Analysis of the surrounding region suggests that the interactions of the CTE with the substrate binding cleft are weak and therefore do not favour an intermolecular reaction. The weak interaction between the CTE and substrate binding cleft ensures that the CTE leaves the binding site following *cis* cleavage making the substrate binding cleft available for the recognition and cleavage of eIF4G, the cellular substrate of FMDV Lb^{PRO}. The weak interaction between the native cleavage site and the substrate binding site may explain why processing of the duplicate cleavage site within the linker region *in trans* was not as efficient as the *cis* reaction. A more favourable reaction may result by replacing the native *cis* cleavage site with the *trans* Lb^{PRO} cleavage site from the cellular substrate eIF4G.

7.4.3. Expression in planta.

The cassette encoding the dual-reporter Lb^{PRO}BLG polyprotein was cloned into the pPZP binary vector and transformed into *Agrobacterium tumefaciens* in preparation for the transformation of tobacco plants. The difficulties encountered during the transformation of *Agrobacterium* with this plasmid are outlined in section 6.1.6. Subsequent analysis of the plasmid DNA extracted from the only two successful transformants (refer to section 6.2.3.) identified mutations in both plasmids within the promoter region and also a large insertion within Lb^{PRO} for the pPZP.Lb^{PRO}BLG1 transformant. As a result we have concluded that FMDV Lb^{PRO} is toxic to *Agrobacterium tumefaciens*.

The *Nicotiana tabacum* seedlings were transformed via *Agrobacterium*-mediated transformation with *Agrobacterium* containing the pPZP.Lb^{PRO}BLG2 binary vector that contains mutations within the CaMV promoter. Only a small number of healthy, well-rooted transformants were produced. One of the main problems with the seedling transformation method used was maintaining the antibiotic selective pressure. 14 day old seedlings were transformed at this stage the seedlings were well-developed and it

was difficult to keep the whole plant in contact with the selective medium, resulting in a large number of false positive plants. In retrospect, transforming the seedlings earlier and changing the medium more frequently to maintain a high level of antibiotic could have alleviated the problem. Also to counteract the higher frequency of false positives resulting from this transformation method a higher number of seedlings should have been transferred to individual pots for rooting.

The healthy, green plants were screened for β -glucuronidase activity to identify plants expressing the Lb^{PRO}BLG polyprotein. The GUS assay was negative for all putative Lb^{PRO}BLG transformants (refer to Fig 6.8.). The simple explanation for this result was that the plants were not expressing the polyprotein. Alternatively it is possible that the GUS protein produced from the polyprotein is enzymatically inactive hence producing a false negative result in the assay. Therefore to investigate this four healthy, well-established plants were harvested and analyzed by western blotting probing with anti-BFP and anti-GUS antisera.

Probing with anti-BFP antibodies identified a single protein in all four plants approximately 30kDa in size, which was identified as the BFP cleavage product thus verifying that the Lb^{PRO}BLG polyprotein was expressed. No larger precursor polypeptides were detected indicating that again within the limits of the detection method in this system the polyprotein processing is 100% efficient.

In contrast, probing with anti-GUS antibodies failed to detect any proteins. There are three possible explanations for this; the first that the polyprotein is not expressed is unlikely as the BFP cleavage product was detected by western blotting. The second possibility is that the second reporter protein GUS is produced but is not detected by the assay or western blotting. The analysis of the Lb^{PRO}BLG2 polyprotein cassette detailed in section 6.2.3. identified mutations within the critical and enhancer regions of the CaMV promoter. The modification of the promoter region was proposed to

down-regulate the expression of the Lb^{PRO}BLG polyprotein to a level that could be tolerated by *Agrobacterium tumefaciens*. Therefore the level of polyprotein expressed in transformed plants would also be greatly reduced due to the weakened CaMV promoter. Therefore the amount of GUS protein may be too low to be detected.

The final explanation is that the complete polyprotein is not expressed in the plants due to premature termination of translation downstream of BFP. Several lines of evidence suggest that premature termination does not occur within the GUS protein, the original Lb^{PRO}BLG polyprotein cassette was sequenced and no termination codons within GUS were detected. Secondly, the expression of Lb^{PRO}BLG in prokaryotes demonstrated that full-length active GUS was produced from the polyprotein. Finally, if termination occurred within GUS one would expect to detect a truncated protein probing with anti-GUS antisera. Therefore if premature termination does occur within the polyprotein the data indicates that the termination site must lie within the linker region that separates BFP and GUS. DNA sequencing of this region verified that the linker region did not contain any possible termination codons. A stretch of sequence between the linker and the start codon of GUS however, contained the hydrophobic amino acid sequence –FFF–.

The extra sequence originates from the plasmid pGFP2AGUS used to create pGEM.linkerGUS. In the pGFP2AGUS plasmid the C-terminus of FMDV 2A is followed by the first amino acids from FMDV 2B which contains the –FFF– sequence.

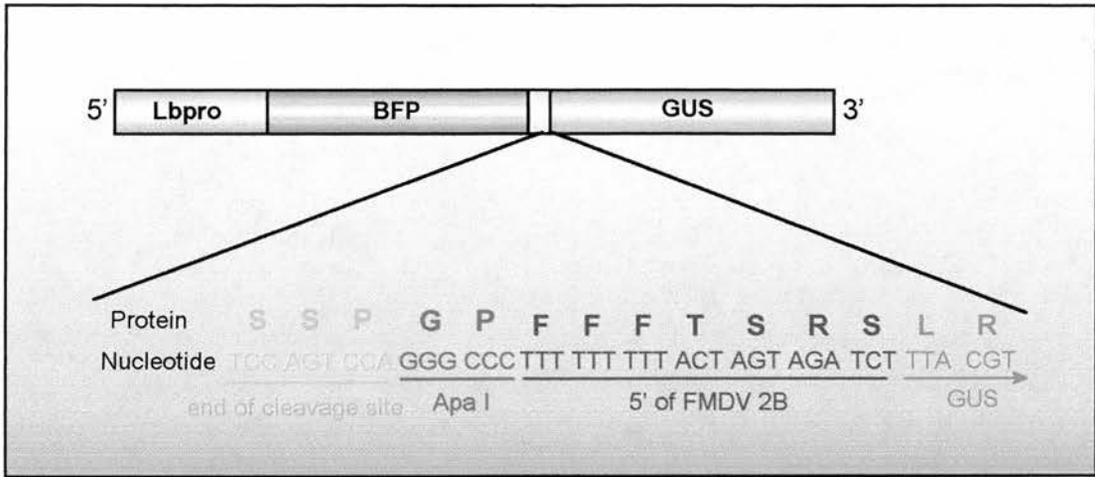


Fig. 7.2. Site of the –FFF– sequence within the *Lb^{pro}BLG* polyprotein construct at the 3' end of the linker region.

Recent work within the laboratory indicates that *in vitro* the –FFF– sequence from FMDV 2B can cause premature termination (unpublished results). Figure 7.3. shows the results of an experiment monitoring the effect of –FFF–, a series of polyprotein cassettes with and without –FFF– were translated *in vitro* in wheatgerm extract. The translation products were analyzed by denaturing-PAGE and visualized by autoradiography.

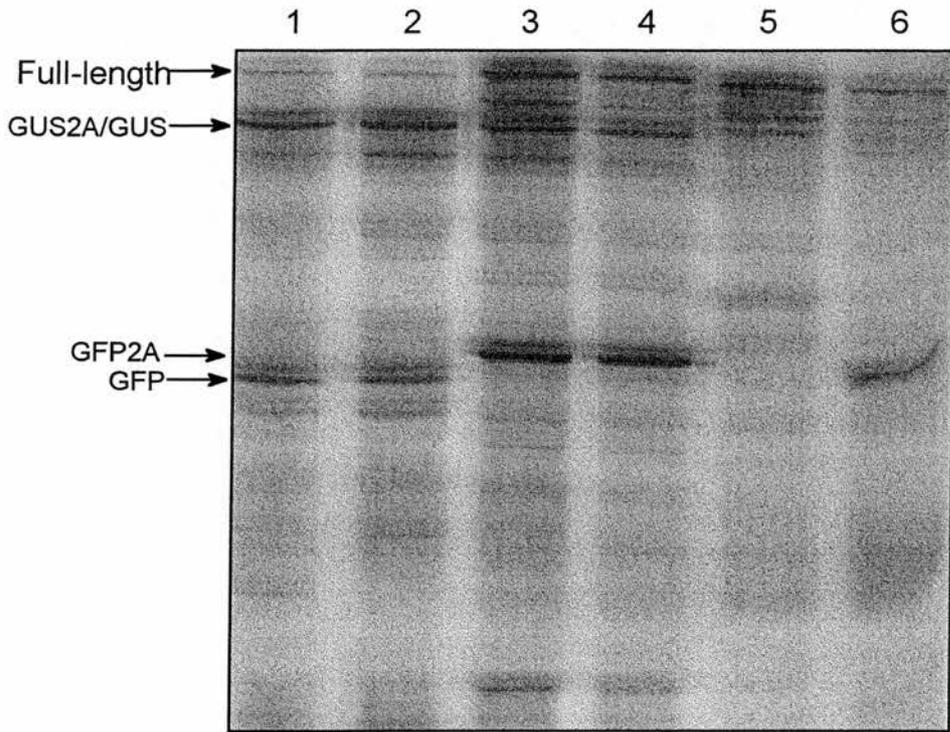


Fig.7.3. TnT analysis of -FFF- containing constructs in WGE. Lane 1- pGUS2AGFP, Lane 2- pGUS2AFFFGFP, Lane 3- pGFP2AGUS, Lane 4- pGFP2AFFFGUS, Lane 5- pGFPGUS, Lane 6- pGFPPFFGUS.

The cleavage products were measured using the phosphorimager and corrected for the respective methionine content of the proteins. The ratio of the first product to the second product was calculated to monitor the effect of the -FFF- sequence on premature termination following FMDV 2A. The imbalance of the first to second product observed for FMDV 2A cleavage within artificial polyproteins is enhanced by the presence of the -FFF- sequence. The insertion of the -FFF- sequence between GFP and GUS in the absence of FMDV 2A clearly demonstrates that the presence of -FFF- alone is sufficient to cause substantial premature termination of translation at this point within the *in vitro* TnT system.

Plasmid	Ratio products A:B
pGUS2AGFP	GUS2A : GFP 1.44 : 1
pGUS2AFFFGFP	GUS2A : FFFGFP 2.2 : 1
pGFP2AGUS	GFP2A : GUS 1.18 : 1
pGFP2AFFFGUS	GFP2A : FFFGUS 2.2 : 1
pGFPGUS	No cleavage
pGFPPFFGUS	GFP : GFPPFFGUS 2.4 : 1

Table 7.1. Calculated ratios of cleavage products compensating for methionine content.

Therefore we cannot rule out the possibility that the presence of the –FFF– sequence within the Lb^{PRO}BLG polyprotein causes some termination of translation following the linker region *in planta* thereby accounting for the lack of GUS protein.

After consideration of the various possibilities at this stage I would favour the modification of the CaMV promoter for the lack of detected GUS product *in planta* because the levels of BFP, although detected using the highly specific BFP antibody, were also low.

7.4.4. Summary of the FMDV Lb^{PRO}-based polyprotein system.

The FMDV Lb^{PRO}-based polyprotein system is very efficient processing intramolecular cleavage events *in vitro* but the major problem with using this protease is the toxicity of the protein *in vivo*. The presence of the FMDV Lb protease was tolerated in *E.coli* perhaps because expression was tightly repressed during the growth phase of the bacteria. Also the level of polyprotein expression within the cells was not high, the polyprotein translation products could not be distinguished visually on a stained denaturing-PAGE gel.

The major difficulty caused by the toxicity of FMDV Lb^{PRO} was the sensitivity of *Agrobacterium tumefaciens*. As a result an alternative method of transformation would be required to introduce FMDV Lb^{PRO} artificial polyproteins into plants. The downregulation of polyprotein expression caused by the mutations that occurred within *Agrobacterium* hampered the subsequent analysis of the function of the Lb^{PRO} artificial polyprotein system *in planta*.

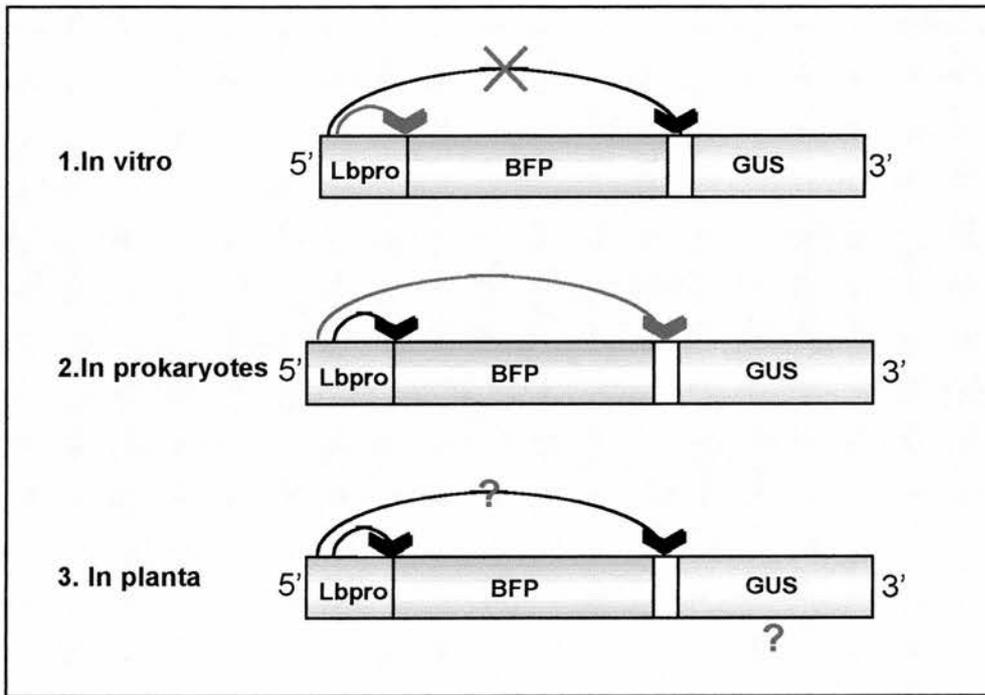


Fig.7.4. Summary of processing of artificial dual-reporter polyprotein by FMDV Lb^{PRO}. Black arrow indicates 100% cleavage, blue arrow indicates less than 100% cleavage.

7.4.5. Further work.

In order to clarify the potential utility of the FMDV Lb^{PRO}-based polyprotein system *in vivo* a few minor alterations to the system are required.

- Replacing the native Lb^{PRO}/1A cleavage site within the linker region with the Lb^{PRO} cleavage site from eIF4G to try to improve the efficiency of the *trans* processing event.

- Deletion of the –FFF- sequence following the linker region.
- Plant transformation via a non-*Agrobacterium* mediated method such as biolistics. Alternatively, transformation using a vector system that tightly represses any translation in *Agrobacterium*. Jahnke *et al.*, (1998) recently reported a repression system involving anti-sense technology that allows *Agrobacterium* to tolerate toxic genes.

7.5. The HRV14 2A^{PRO}- based polyprotein system.

7.5.1. *In vitro* analysis.

The simple polyprotein GFP2A^{PRO} was translated *in vitro* in rabbit reticulocyte and wheatgerm extract, in both systems the polyprotein was processed *in cis* by the HRV14 2A^{PRO} to yield the mature products GFP and HRV14 2A^{PRO}. Thus demonstrating that HRV14 2A^{PRO} does not require any other viral sequences for cleavage. Mutation of the active site cysteine residue to an alanine at position 109 abolished the proteolytic activity of the protease. Translation of pGFP2A produced a major product that was identified as the uncleaved polyprotein GFP2A.

Analysis of the relative quantities of uncleaved and cleaved components indicated that in RRL 18% of the total polyprotein was not processed, in WGE this increased to 22%. The HRV14 2A^{PRO} is markedly less efficient *in vitro* than FMDV Lb^{PRO} which is 99% efficient in RRL. There are a number of possible reasons for this difference in the *in vitro* cleavage efficiency of the proteases outlined below.

1. Does HRV14 2A^{PRO} require other viral sequences?

This is the simplest explanation that another part of the virus is required that enhances the proteolytic activity of HRV14 2A^{pro}. Examination of the literature suggests that this may be the case.

Nicklin *et al.* (1987) translated a plasmid encoding poliovirus P1-2A2B2C' *in vitro* in RRL. They reported that translation yielded two products identified as P1 and 2A2B2C', no full-length polyprotein was detected. Reports by Donnelly *et al.* (in preparation) supported this, they reported that translation of the entire P1P2 region from HRV *in vitro* in RRL yielded two products corresponding to P1 and P2 respectively. Again no uncleaved precursor was observed. Yu and Lloyd (1991) however, reported that the full-length precursor polyprotein was detected upon *in vitro* translation of the poliovirus 1C'1D2A region. Taking these findings into account suggests that complete processing by the entero- and rhinovirus 2A protease requires the presence of other viral sequences within the P1P2 region.

Detailed analysis of the basic requirements for 2A^{pro} cleavage of peptides indicated that the only requirement was the upstream sequence P1-P8 from 1D (May Wang *et al.*, 1997;1998). The constructs analyzed in this thesis contained the residues P1-P6 demonstrating that this minimal requirement can be further reduced.

2. Is the *in vitro* system lacking zinc required by HRV14 2A^{pro}?

The entero- and rhinovirus 2A protease contains a zinc ion that is not involved in the proteolytic activity of the protease but has an important structural role and is vital for folding of the protein. The pretreatment of RRL with the chelating agent EGTA to remove excess calcium ions would also act to deplete the complement of zinc ions present (pers.comm. L. Sheppard, Sigma-Aldrich), the addition of excess Zn²⁺ to the system may rescue the efficiency of the proteolytic activity. Yu and Lloyd (1992) investigated the effect of metal ions on the *trans* processing activity of poliovirus 2A^{pro} by the addition of various chelating agents to the lysate. They reported that the

addition of EGTA or EDTA up to 10mM did not affect processing *in trans* but the presence of a strong, specific Zn^{2+} chelator successfully inhibited the *trans* cleavage activity verifying that Zn^{2+} is important for the activity of 2A^{PRO}. Yu and Lloyd however, did not replace the Zn^{2+} lost as a result of EGTA treatment of RRL therefore the system would already be depleted of Zn^{2+} .

3. Is temperature important for HRV14 2A^{PRO} activity?

Analysis by May Wang *et al.*, (1998) on the cleavage activity of the HRV 2A protease identified a surprising property of the HRV14 2A protease. The HRV14 2A protease activity was found to be very sensitive to temperature and was more efficient at low temperatures, the optimum temperature of HRV14 2A^{PRO} was approximately 20°C. This result was highly unexpected for a viral protease that must process during infection at 37°C. From this analysis the proteolytic activity of HRV14 2A^{PRO} at 37°C is only 65%. This preference for low temperatures was specific for the HRV14 2A protease and was not demonstrated by the HRV2 2A protease which has an optimum temperature of 40°C. The *in vitro* reactions were carried out at 30°C, from the graph of the HRV14 2A^{PRO} activity the protease is less than 100% efficient at this temperature.

4. Is a cellular factor required by HRV14 2A for efficient cleavage?

Another possibility is that a cellular factor required by the 2A protease is deficient in RRL. Comparison with the cleavage efficiency of FMDV Lb^{PRO} in RRL suggests that if this is the case the cellular factor is not required by FMDV Lb^{PRO}. The reduction in cleavage efficiency observed in WGE for both types of proteases suggest that this system is deficient in cellular components required by both proteases.

Translation of the dual-reporter polyprotein GLB2A^{PRO} *in vitro* in RRL confirmed that the HRV14 2A^{PRO} was proteolytically active *in cis*. The primary cleavage products HRV14 2A^{PRO} and GUSlinkerBFP were identified in the translation profile.

Again no *trans* processing within the linker region was detected as predicted for this system.

7.5.2. Expression in prokaryotes.

Western blot analysis of the total cellular extract from *E.coli* expressing the simple BFP2A^{PRO} polyprotein probing with anti-BFP antisera identified a single band corresponding to the BFP cleavage product. The full-length polyprotein was not detected within the limits of this detection system thus we can conclude that *cis* cleavage at the BFP/2A^{PRO} junction by HRV14 2A^{PRO} is complete confirming that HRV14 2A^{PRO} is proteolytically active in *E.coli*.

Western blot analysis of the total cellular extract from *E.coli* expressing the dual-reporter polyprotein GLB2A^{PRO} probing with BFP antisera identified a single band corresponding to the BFP cleavage product. Probing with GUS antisera corroborated this data by identifying a single band corresponding to the GUS cleavage product. Larger precursor polypeptides were not detected indicating that in *E.coli* the GLB2A^{PRO} polyprotein has been completely processed by the HRV14 2A^{PRO} *in cis* and *in trans* to yield the mature cleavage products GUS and BFP.

Comparison of the cleavage of both dual-reporter polyproteins Lb^{PRO}BLG and GLB2A^{PRO} *in vivo* in *E.coli* suggests that contrary to the initial data from the *in vitro* analysis, overall HRV14 2A^{PRO} is more efficient than FMDV Lb^{PRO}, as the intermediate polypeptide BFPlinkerGUS was detected from the Lb^{PRO}BLG expressing *E.coli* extract. Thus HRV14 2A protease cleaves *in trans* within the linker region more efficiently than FMDV Lb^{PRO}. This observation may result from the different types of cleavage site present in the linker region. The linker region from GLB2A^{PRO} contains the 2A^{PRO} *trans* cleavage site from the cellular substrate eIF4G. In contrast, the linker region of Lb^{PRO}BLG contains a duplicate of the native L^{PRO}/1A cleavage

site that is usually processed *in cis*: replacement of this cleavage site with a natural *trans* cleavage site may improve *trans* processing of the polyprotein as detailed above.

7.5.3. Expression in planta.

The binary vector encoding the GLB2A^{PRO} polyprotein was transferred into *Agrobacterium tumefaciens* without difficulty, in preparation for transformation of *Nicotiana tabacum*. Therefore unlike FMDV Lb^{PRO} the HRV14 2A^{PRO} does not appear to be toxic to *Agrobacterium*. The healthy, green plants were screened by an initial GUS assay that identified the plants expressing the GLB2A^{PRO} polyprotein. The positive GUS plants were harvested and analyzed by western blotting probing with GUS and BFP antisera. Probing the total leaf extract with BFP antisera identified a single band approximately 30kDa in size, identified as the BFP cleavage product. Probing with GUS antisera also detected a single band of approximately 69kDa corresponding to the GUS cleavage product. Larger precursor polypeptides were not detected suggesting that processing at the *cis* and *trans* cleavage sites of the polyprotein is completely efficient within the limits of detection in this system.

7.5.4. Summary of HRV14 2A^{PRO}-based polyprotein system.

The HRV14 2A^{PRO}-based polyprotein system is not as 100% efficient at processing intramolecular cleavage *in vitro* but subsequent analysis *in vivo* in both prokaryotes and *in planta* indicate that the polyproteins are completely processed by HRV14 2A^{PRO} to yield the mature cleavage products. Therefore we have demonstrated that HRV14 2A^{PRO} produced within artificial constructs is proteolytically active both *in cis* and *in trans*. Expression of the polyprotein system in *E.coli* confirms previous reports that proteolytically active 2A^{PRO} can be expressed in prokaryotes. More importantly the studies *in planta* demonstrate that healthy plants expressing proteolytically active HRV14 2A^{PRO} can be produced.

Overall HRV14 2A^{PRO} does not appear to be as toxic as its counterpart FMDV Lb^{PRO} most notably demonstrated by the successful transformation of *Agrobacterium tumefaciens*.

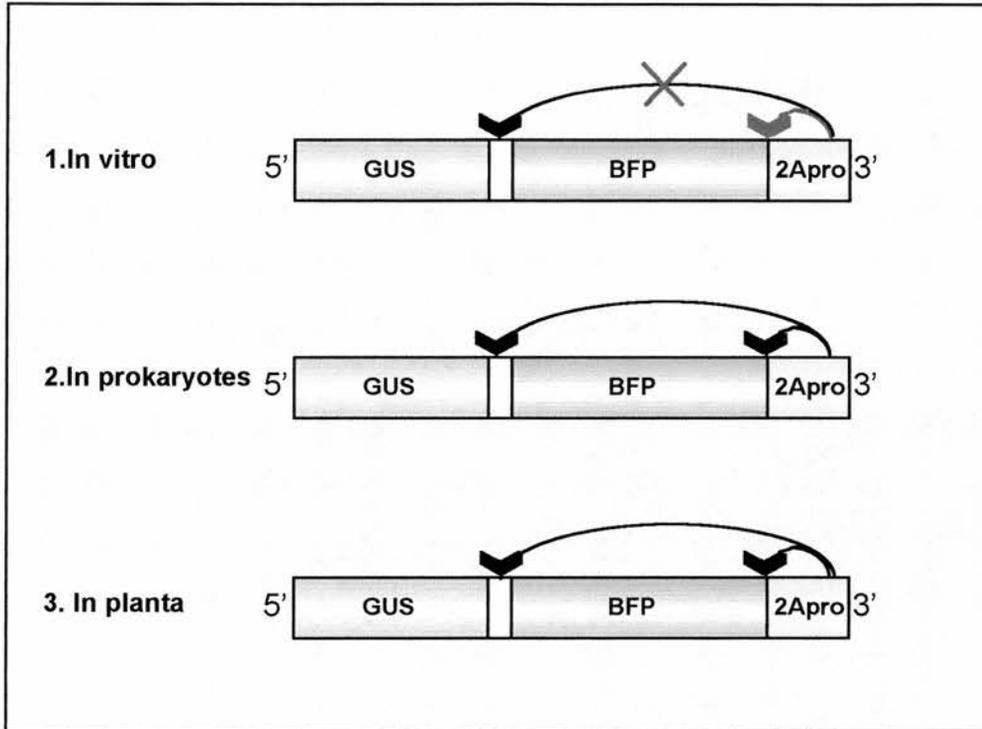


Fig. 7.5. Summary of GLB2A^{PRO} polyprotein processing in various systems. Blue arrow indicates less than 100% cleavage, black arrow indicates complete cleavage.

7.5.5. Further work.

A minor aspect would be to further investigate the reduced cleavage efficiency of HRV14 2A^{PRO} *in vitro* to pinpoint the cause of the inefficient processing. The putative explanations suggested above could be investigated by-

- Including viral sequences within the artificial construct
- Monitoring the addition of Zn²⁺ ions to the lysate
- Monitoring the affect of temperature on the reaction

- The addition of HeLa cell extract to the lysate

The HRV14 2A^{PRO}-based polyprotein system is working efficiently *in vivo*. In order to clarify the system further tagging the proteins would be useful to overcome the limits of the detection system. This would also allow quantification of the polyprotein cleavage products permitting detailed analysis of the relative cleavage efficiency *in cis* and *in trans* and confirmation of the ratios of the mature products.

7.6. Conclusions.

The artificial self-processing systems described in this thesis can be used successfully *in vivo* in both prokaryotes and *in planta*. From this preliminary investigation the protease of choice would be HRV14 2A protease. The 2A protease efficiently processed the polyprotein *in cis* and *in trans* and overall is less toxic than the FMDV Lb protease. The toxicity of FMDV Lb^{PRO} to *Agrobacterium tumefaciens* means that *Agrobacterium*-mediated transformation cannot be used to introduce vectors containing FMDV Lb^{PRO} to plants. *Agrobacterium*-mediated transformation is the most common method of plant transformation thus the requirement for an alternative transformation method for Lb^{PRO} vectors is a definite drawback to the use of this system.

At this stage in development further work is required to clarify the potential use of the polyprotein system *in planta*. An important experiment would be to express the polyproteins in protoplasts, this would allow a direct analysis of the toxicity of the proteases to plant cells. This work is being conducted at the time of writing this thesis. It would also be interesting to transform plants via a non-*Agrobacterium* method thereby alleviating problems due to the sensitivity of *Agrobacterium* to the FMDV Lb protease.

The next step in the development of the artificial polyprotein system would be to attempt to express a novel pathway *in vivo*. The *trans* processing activity of the protease means that the length of the artificial polyprotein is not limited. Therefore the enzymes required for a pathway can be produced within a single polyprotein linked by short protease cleavage sites. Other refinements that would allow targeting the proteins encoded by the polyprotein to specific cell organelles should also be investigated.

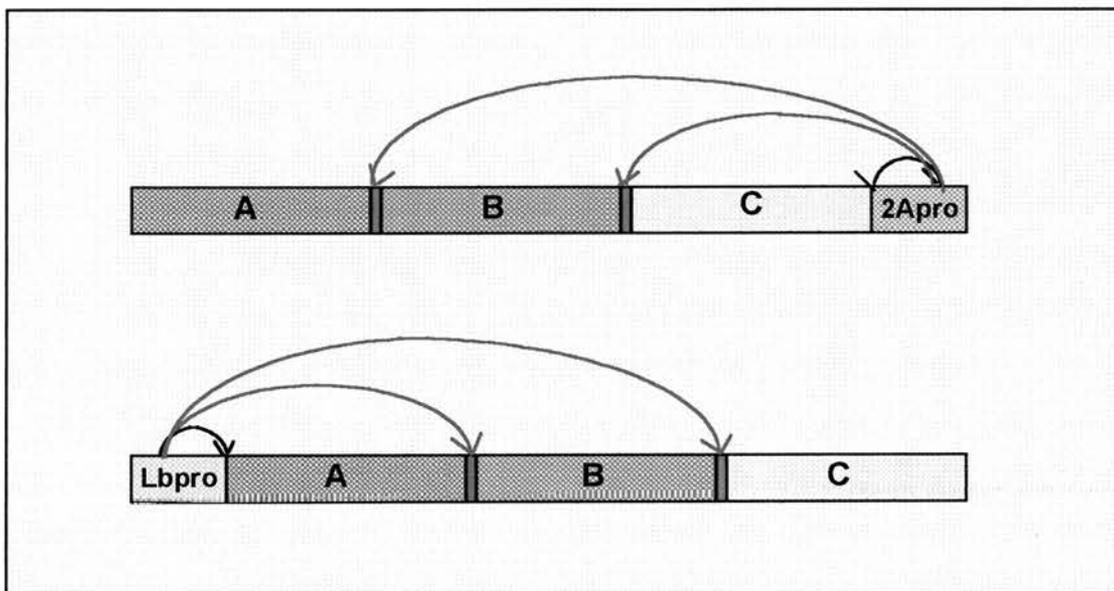


Fig.7.6. Artificial polyproteins containing three pathway enzymes A, B and C. Multiple linker sites (shown in blue) between each protein contain a cleavage site that is processed by the protease (Lb^{Pro}/2A^{Pro}). The black arrows indicate a cis cleavage reaction and the blue arrows a trans cleavage reaction.

7.7. Applications of the artificial polyprotein system.

The field of biotechnology has developed at a staggering pace over recent years particularly in the area of plant biotechnology although more traditional biotechnological applications should not be overlooked. The manipulation of plant cell machinery provides a catalogue of potential uses ranging from the simple alteration of natural plant traits (Kasuga *et al.*, 1999; Hu *et al.*, 1999) important to plant breeders to more bizarre applications such as the production of pharmaceutical antibodies (Stöger *et al.*, 2000), plastics (Slater *et al.*, 1999) or the biodegradation of explosives (French *et al.*, 1999). A major problem associated with this new technology is the introduction and subsequent co-ordination of multiple genes. The artificial polyprotein system borrows a strategy employed by viruses to overcome this problem.

The artificial polyprotein system initially developed by Marcos and Beachy (1994) using the potyvirus NIa protease has been used to produce virus-resistant plants by the expression of different viral coat proteins (Marcos and Beachy, 1997). More recently this system was used by Dasgupta *et al.*, (1998) to target reporter proteins expressed from an artificial polyprotein to the cell chloroplasts. The environment of different cellular organelles can vary dramatically, thus targeting products to specific organelles allows selection of the optimum environment for specific proteins and may help to overcome problems due to the toxicity of the introduced gene products. Thus the potential of the artificial polyprotein system to introduce and target multiple genes in various systems is vast. The limitations of the system will be due to ethical issues rather than of a practical nature.

7.8. Concluding remarks.

A novel artificial self-processing polyprotein system to allow co-ordinated expression of multiple genes has been constructed. The expression of the artificial polyproteins containing the reporter genes GFP and GUS has been monitored to assay the system *in vivo* in prokaryotic and eukaryotic systems.

The artificial polyproteins were processed efficiently by the picornaviral proteases to yield the mature products. Further research must be undertaken to fully investigate the potential use of the system particularly with regards to the FMDV Lb^{PRO}, but the results presented here are an encouraging first assessment of this system. The artificial polyprotein system using the HRV14 2A^{PRO} and FMDV Lb^{PRO} provides an additional alternative to the use of the FMDV 2A peptide that is becoming increasingly popular for the production of artificial polyproteins.

APPENDIX

PCR product	Label	Oligonucleotide Primer
wtGFP	246 254	For. TACCCAGGATCCATGAGTAAAGGAGAAGAACTT Rev. GTGGGAGAGCTCTCATCTAGATCCGGATTTGTATAG
FMDV Lb ^{pro} (GFP)	253 271	For. TACCCAGGATCCATGGAAGTACACTGTACAAC Rev. AAGTTCTTCTCCTTTACTCATTGTCCAGCCCCTTTGAGC TT
GFP (Lb ^{pro})	242 254	For. AAGCTCAAAGGGGCTGGACAAATGAGTAAAGGAGAAGAA CTT Rev. GTGCGAGAGCTCTCATCTAGATCCGGATTTGTATAG
HRV14 2A ^{pro} (GFP)	249 255	For. CTATACAAATCCGGGTCTAGAGGTGACATTAATCCTATG GT Rev. GTGCGAGAGCTCTCACTGTTCTCTGCGATACTC
GFP (2A ^{pro})	246 272	For. TACCCAGGATCCATGAGTAAAGGAGAAGAACTT Rev. ACCATAGGATTTAATGTCACCTCTAGACCCGGATTTGTAT AG
FMDV Lb*	253 A3584	For. TACCCAGGATCCATGGAAGTACACTGTACAAC Rev. GATGGCGTTCAACCATGCGTT
*LbGFP	A3583 254	For. AACGCATGGTTGAACGCCATC Rev. GTGCGAGAGCTCTCATCTAGATCCGGATTTGTATAG
*2A	A3581 255	For. GACGCAGGTGGGATTTTGAAGA Rev. GTGCGAGAGCTCTCACTGTTCTCTGCGATACTC
GFP2A*	246 A3582	For. TACCCAGGATCCATGAGTAAAGGAGAAGAACTT Rev. TCTCAAAATCCCACCTGCGTC
FMDV Lb ^{pro} (BFP)	253 201	For. TACCCAGGATCCATGGAAGTACACTGTACAAC Rev. CTCTCGCCCTTGCTCACCATTGTCCAGCCCCTTTGAGC TT
BFP (Lb ^{pro})	191 190	For. AAGCTCAAAGGGGCTGGACAAATGGTGAGCAAGGGCGAG GAG Rev. GTGCGAGAGCTCTCAGACGCGGCCGCTCTTGTATAG
BFP-No stop (Lb ^{pro})	191 C6765	For. AAGCTCAAAGGGGCTGGACAAATGGTGAGCAAGGGCGAG GAG Rev. GAGCTCTCTAGACTTGTACAGCTCGTCCAT

HRV14 2A ^{pro} (BFP)	242 255	For. GACGAGCTGTACAAGGGTGACATTAATCCTATGGT Rev. GTGCGAGAGCTCTCACTGTTCTCTGCGATACACTC
BFP (2A ^{pro})	C1968 200	For. TACCCAGGATCCATGGTGAGCAAGGGCGAGGAG Rev. ACCATAGGATTTAATGTCACCGTCGCGGCCGCTCTTGTAC AG
HRV14 IRES	251 257	For. GGGTCCGCATGCCGGATGGGTATCCCACCA Rev. ACCCCTCTCGAGAAGACAATATGAAGAAAAATGAGA
FMDV IRES	144 256	For. GGGTCCGCATGCACTGACGCAACTTGAACTCCGCC Rev. ACCCCTCTCGAGGTGTTTTCAGTGGTTATAAAAGG
BFP	C1968 190	For. TACCCAGGATCCATGGTGAGCAAGGGCGAGGAG Rev. GTGCGAGAGCTCTCAGACGCGGCCGCTCTTGTATAG

Table 8.1. List of oligonucleotide primers used to amplify the PCR products. The restriction enzyme sites used for cloning are marked in red, other restriction sites built into the primer are marked in green. The initiation and stop codons are highlighted in blue.

BIBLIOGRAPHY

- Allaire, M., Chernaiia, M. M., Malcolm, B. A. and James, M. N. G. (1994). Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteases. *Nature* **369**, 72-76.
- Allen, M. L., Metz, A. M., Timmer, R. T., Rhoads, R. E. and Browning, K. S. (1992). Isolation and sequence of the cDNAs encoding the subunits of the isozyme form of wheat protein synthesis initiation factor 4F. *Journal of Biological Chemistry* **267**(32), 23232-23236.
- Andrés, G., Simón-Mateo, C. and Viñuela, E. (1997). Assembly of the African Swine Fever virus: role of polyprotein pp220. *Journal of Virology* **71**, 2331-2341.
- Aranda, M. and Maule, A. (1998). Virus-induced host gene shutoff in animals and plants. *Virology* **243**, 261-267.
- Arnold, E., Luo, M., Vriend, G., Rossman, M.G., Palmenberg, A.C., Parks, G.D., Nicklin, M.J.H., Wimmer, E. (1987). Implications of the picornavirus capsid structure for polyprotein processing. *Proc. Natl. Acad. Sci., U.S.A.* **84**, 21-25.
- Babé, L. M. and Craik, C. S. (1997). Viral proteases: Evolution of diverse structural motifs to optimize function. *Cell* **91**, 427-430.
- Badorff, A., Lee, G.-H., Lamphear, B. J., Martone, M. E., Campbell, K. P., Rhoads, R. E. and Knowlton, K. E. (1999). Enteroviral protease 2A cleaves dystrophin: Evidence of cytoskeletal disruption in an acquired cardiomyopathy. *Nature Medicine* **5**(3), 320-326.
- Barbero, J. L., Buesa, J. M., González de Buitrago, G., Méndez, E., Pérez-Aranda, A. and García, J. L. (1986). Complete nucleotide-sequence of the penicillin acylase gene from *Kluyvera-citrophilia*. *Gene* **49**, 69-80.
- Barco, A., Feduchi, E. and Carrasco, L. (2000). A stable HeLa cell line that inducibly expresses poliovirus 2A^{Pro}: Effects on cellular and viral gene expression. *Journal of Virology* **74**(5), 2383-2392.
- Bazan, J.F. and Fletterick, R. J. (1989). Detection of a trypsin-like serine protease domain in flaviviruses and pestiviruses. *Virology* **171**, 637-639.

Beadle, G. W. and Tatum, E. C. (1941). Genetic control of biochemical reactions in *Neurospora*. *Proc. Natl. Acad. Sci., U.S.A.* **27**, 499-506.

Belsham, G. J., McInerney, G. M. and Ross-Smith, N. (2000). Foot-and-mouth disease virus 3C protease induces cleavage of translation initiation factors eIF4A and eIF4G within infected cells. *Journal of Virology* **74**, 272-280.

Berti, P. J. and Storer, A. C. (1995). Alignment/phylogeny of the papain superfamily of cysteine protease. *Journal of Molecular Biology* **246**, 273-283.

Böck, A., Wirth, R., Schmid, G., Scumacher, G., Lang, G. and Buckel, P. (1983). The penicillin acylase from *Escherichia coli* ATCC11105 consists of two dissimilar subunits. *FEMS Microbiology Letters* **20**, 135-139.

Borman, A. M., Bailly, J.-L., Girard, M. and Kean, K. M. (1995). Picornavirus internal ribosome entry segments: comparison of translation efficiency and the requirements for optimal internal initiation of translation *in vitro*. *Nucleic Acids Research* **23**(18), 3656-3663.

Borman, A. M., Kirchweger, R., Ziegler, E., Rhoads, R. E., Skern, T. and Kean, K. M. (1997a). eIF4G and its proteolytic cleavage products: Effect on initiation of protein synthesis from capped, uncapped and IRES-containing mRNAs. *RNA* **3**, 186-196.

Borman, A. M., Le Mercier, P., Girard, M. and Kean, K. M. (1997b). Comparison of picornaviral IRES-driven internal initiation of translation in cultured cells of different origins. *Nucleic Acids Research* **25**(5), 925-932.

Bovee, M. L., Lamphear, B. J., Rhoads, R. E. and Lloyd, R. E. (1998a). Direct cleavage of eIF4G by poliovirus 2A protease is inefficient *in vitro*. *Virology* **245**, 241-249.

Bovee, M. L., Marissen, W. E., Zamora, M. and Lloyd, R. E. (1998b). The predominant eIF4G-specific cleavage activity in poliovirus-infected HeLa cells is distinct from 2A protease. *Virology* **245**, 229-240.

Brocklehurst, K., Watts, A. B., Patel, M., Verma, C. and Thomas, E. W. (1998). Cysteine proteinases. In "Comprehensive biological catalysis", Vol. Volume 1/ Reactions of electrophilic carbon, phosphorus and sulfur, pp. 381-425 4 vols. Academic Press Ltd.

Browning, K. S. (1996). The plant translational apparatus. *Plant Molecular Biology* **32**, 107-144.

- Browning, K. S., Webster, C., Roberts, J. K. M. and Ravel, J. M. (1992). Identification of an isozyme form of protein synthesis initiation factor 4F in plants. *Journal of Biological Chemistry* **267**(14), 10096-10100.
- Cao, X., Bergman, I. E., Fuhlkrieg, R. and Beck, E. (1995). Functional analysis of the two alternative initiation sites of foot-and-mouth disease virus. *Journal of Virology* **69**, 560-563.
- Carberry, S. E., Darzynkiewicz, E. and Goss, D. J. (1991). A comparison of the binding of methylated cap analogues to wheat germ protein synthesis factors 4F and (Iso)4F. *Biochemistry* **1991**(30), 1624-1627.
- Carrington, J. C., Cary, S. M., Parks, T. D. and Dougherty, W. G. (1989). A second proteinase encoded by a plant potyvirus genome. *EMBO Journal* **8**, 365-370.
- Carrington, J. C. and Dougherty, W. G. (1987). Small nuclear inclusion protein encoded by a plant potyvirus genome is a protease. *Journal of Virology* **61**, 2540-2548.
- Carrington, J. C. and Herndon, K. L. (1992). Characterization of the potyviral HC-Pro autoproteolytic cleavage site. *Virology* **187**, 308-315.
- Ceriani, M. F., Marcos, J. F., Hopp, H. E. and Beachy, R. N. (1998). Simultaneous accumulation of multiple viral coat proteins from a TEV-N1a based expression vector. *Plant Molecular Biology* **36**, 239-248.
- Chambers, T. J., Grakoui, A. and Rice, C. M. (1990). Flavivirus genome organization, expression, and replication. *Annual Review of Microbiology* **44**, 649-688.
- Chaplin, P. J., Camon, E. B., Villarreal-Ramos, B., Flint, M., Ryan, M. D. and Collins, R. A. (1999). Production of interleukin-12 as a self-processing 2A polypeptide. *Journal of Interferon and Cytokine research* **19**, 235-241.
- Chen, L., DeVries, A. L. and Cheng, C.-H. C. (1997). Convergent evolution of antifreeze glycoproteins in Antarctic notothenoid fish and Arctic cod. *Proc. Natl. Acad. Sci., U.S.A.* **94**, 3817-3822.
- Chernaia, M. M., Malcolm, B. A., Allaire, M. and James, M. N. G. (1993). Hepatitis A virus 3C proteinase: Some properties, crystallization and preliminary crystallographic characterization. *Journal of Molecular Biology* **234**, 890-893.

Clark, M. E., Hämmerle, T., Wimmer, E. and Dasgupta, A. (1991). Poliovirus proteinase 3C converts an active form of transcription factor IIIC to an inactive form: a mechanism for inhibition of host cell polymerase III transcription by poliovirus. *EMBO Journal* **10**(10), 2941-2947.

Clark, M. E., Lieberman, P. M., Berk, A. J. and Dasgupta, A. (1993). Direct cleavage of human TATA-binding protein by poliovirus protease 3C *in vivo* and *in vitro*. *Molecular and Cellular Biology* **13**(2), 1232-1237.

Craig, A. W. B., Haghghat, A., Yu, A. T. K. and Sonenberg, N. (1998). Interaction of polyadenylate-binding protein with the eIF4G homologue PAIP enhances translation. *Nature* **392**(6675), 520-525.

Crum, C. J., Hu, J., Hiddinga, H. J. and Roth, D. A. (1988). Tobacco mosaic virus infection stimulates the phosphorylation of a plant protein associated with double-stranded RNA-dependent protein kinase activity. *Journal of Biological Chemistry* **263**, 13440-13443.

Dasgupta, S., Collins, G.B. and Hunt, A.G. (1998). Co-ordinated expression of multiple enzymes in different subcellular compartments in plants. *Plant Journal* **16**(1), 107-116.

Davies, M. V., Pelletier, J., Meerovitch, K., Sonenberg, N. and Kaufman, R. J. (1991). The effect of poliovirus proteinase 2A^{PRO} expression on cellular metabolism. *Journal of Biological Chemistry* **266**, 14714-14721.

de Felipe, P., Martín, V., Cortés, M. L., Ryan, M. D. and Izquierdo, M. (1999). Use of the 2A sequence from foot-and-mouth disease virus in the generation of retroviral vectors for gene therapy. *Gene Therapy* **6**, 198-208.

De Francesco, R., Urbani, A., Nardi, M. C., Tomei, L., Steinkühler, C. and Tramontano, A. (1996). A zinc binding site in viral serine proteinases. *Biochemistry* **35**, 13282-13287.

Dessens, J. T. and Lomonosoff, G. P. (1991). Mutational analysis of the putative catalytic triad of the cowpea mosaic virus 24K protease. *Virology* **184**, 738-746.

DeTulleo, L. and Kirchhausen, T. (1998). The clathrin endocytic pathway in viral infection. *EMBO Journal* **17**(16), 4585-4593.

Devaney, M. A., Vakharia, V. N., Lloyd, R. E., Ehrenfeld, E. and Grubman, M. J. (1988). Leader protein of foot-and-mouth disease virus is required for cleavage of the p220 component of the cap-binding protein complex. *Journal of Virology* **62**(11), 4407-4409.

Donnelly, M. L. L., Gani, D., Flint, M., Monaghan, S. and Ryan, M. D. (1997). The cleavage activities of aphthovirus and cardiovirus 2A proteins. *Journal of General Virology* **78**, 13-21.

Dougherty, W. G. and Semler, B. L. (1993). Expression of virus-encoded proteinases: Functional and structural similarities with cellular enzymes. *Microbiological Reviews* **57**(4), 781-822.

Douglass, J., Civelli, O. and Herbert, E. (1984). Polyprotein gene expression: Generation of diversity of neuroendocrine peptides. *Annual Reviews of Biochemistry* **53**, 665-715.

Ehrenfeld, E. (1982). Poliovirus-induced inhibition of host-cell protein synthesis. *Cell* **26**, 435-436.

Esteves, A., Marques, M. I. and Costa, J. V. (1986). Two-dimensional analysis of African Swine Fever virus proteins and proteins induced in infected cells. *Virology* **152**, 181-191.

Etchison, D., Milburn, S. C., Edery, I., Sonenberg, N. and Hershey, J. W. B. (1982). Inhibition of HeLa cell protein synthesis following poliovirus infection correlates with the proteolysis of a 220,000-dalton polypeptide associated with eucaryotic initiation factor 3 and a cap binding complex. *Journal of Biological Chemistry* **257**(24), 14806-14810.

Failla, C., Tomei, L. and DeFrancesco, R. (1994). Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins. *Journal of Virology* **68**, 3753-3760.

Falk, M. M., Griegera, P. R., Bergmann, I. E., Zibert, A., Multhaup, G. and Beck, E. (1990). Foot-and-Mouth disease virus protease 3C induces specific proteolytic cleavage of host cell histone H3. *Journal of Virology* **64**, 748-756.

Fernandez-Tomas, C. (1982). The presence of viral-induced proteins in nuclei from poliovirus-infected HeLa cells. *Virology* **116**, 629-634.

Flynn, A. and Proud, C. G. (1995). Serine-209, not serine-53, is the major site of phosphorylation in initiation-factor eIF4E in serum-treated chinese hamster ovary cells. *Journal of Biological Chemistry* **270**, 21684-21688.

- Fout, G. S., Medappa, K. C., Mapoles, J. E. and Rueckert, R. R. (1984). Radiochemical determination of polyamines in poliovirus and human rhinovirus 14. *Journal of Biological Chemistry* **259**, 3639-3643.
- Fradkin, L. G., Yoshinaga, S. K., Berk, A. J. and Dasgupta, A. (1987). Inhibition of host-cell RNA polymerase III-mediated transcription by poliovirus: In activation of specific transcription factors. *Molecular and Cellular Biology* **7**, 3880-3887.
- French, C.E., Rosser, S.J., Davies, G.J., Nicklin, S. and Bruce, N.C. (1999). Biodegradation of explosives by transgenic plants expressing pentaerythritol tetranitrate reductase. *Nature Biotechnology* **17**, 491-494.
- Fütterer, J. and Hohn, T. (1996). Translation in plants - rules and exceptions. *Plant Molecular Biology* **32**, 159-189.
- Gallie, D. R. (1996). Translational control of cellular and viral mRNAs. *Plant Molecular Biology* **32**, 145-158.
- Gallinari, P., Brennan, D., Nardi, C., Brunetti, M., Tomei, L., Steinkühler, C. and De Francesco, R. (1998). Multiple enzymatic activities associated with recombinant NS3 protein of hepatitis C virus. *Journal of Virology* **72**(8), 6758-6769.
- Gingras, A.-C., Kennedy, S. G., O'Leary, M. A., Sonenberg, N. and Hay, N. (1998). 4E-BP1, a repressor of mRNA translation, is phosphorylated and inactivated by the Akt(PKB) signalling pathway. *Genes and Development* **12**, 502-513.
- Gingras, A.-C., Raught, B. and Sonenberg, N. (1999). eIF4 initiation factors: Effectors of mRNA recruitment to ribosomes and regulators of translation. *Annual Review of Biochemistry* **68**, 913-963.
- Gingras, A.-C. and Sonenberg, N. (1997). Adenovirus infection inactivates the translational inhibitors 4E-BP1 and 4E-BP2. *Virology* **237**, 182-186.
- Gingras, A.-C., Svitkin, Y., Belsham, G. J., Pause, A. and Sonenberg, N. (1996). Activation of the translational suppressor 4E-BP1 following infection with encephalomyocarditis virus and poliovirus. *Proc. Natl. Acad. Sci., U.S.A.* **93**, 5578-5583.
- Gorbelenya, A. E., Blinov, V. M. and Donchenko, A. P. (1986). Poliovirus-encoded proteinase 3C: a possible evolutionary link between cellular serine and cysteine proteinase families. *FEBS Letters* **194**(2), 253-257.

Gradi, A., Imataka, H., Svitkin, Y. V., Rom, E., Raught, B., Morino, S. and Sonenberg, N. (1998a). A novel functional human eukaryotic translation initiation factor 4G. *Molecular and Cellular Biology* **18**(1), 334-342.

Gradi, A., Svitkin, Y. V., Imataka, H. and Sonenberg, N. (1998b). Proteolysis of human eukaryotic translation initiation factor eIF4GII, but not eIF4GI, coincides with the shutoff of host protein synthesis after poliovirus infection. *Proc. Natl. Acad. Sci., U.S.A.* **95**, 11089-11094.

Grubman, M. J. and Baxt, B. (1982). Translation of foot-and-mouth disease virion RNA and processing of the primary cleavage products in a rabbit reticulocyte lysate. *Virology* **116**, 19-30.

Grubman, M. J., Zellner, M., Bablanian, G., Mason, P. W. and Piccone, M. E. (1995). Identification of the active-site residues of the 3C proteinase of foot-and-mouth disease virus. *Virology* **213**, 581-589.

Guarné, A., Tormo, J., Kirchweger, R., Pfistermueller, D., Fita, I. and Skern, T. (1998). Structure of the foot-and-mouth disease virus leader protease: a papain-like fold adapted for self-processing and eIF4G recognition. *EMBO Journal* **17**(24), 7469-7479.

Haghighat, A., Svitkin, Y., Novoa, I., Kuechler, E., Skern, T. and Sonenberg, N. (1996). The eIF4G-eIF4E complex is the target for direct cleavage by the rhinovirus 2A proteinase. *Journal of Virology* **70**, 8444-8450.

Hajdukiewicz, P., Svab, Z. and Maliga, P. (1994). The small, versatile pPZP family of *Agrobacterium* binary vectors for plant transformation. *Plant Molecular Biology* **25**, 989-994.

Halpin, C., Cooke, S. E., Barakate, A., El Amrani, A. and Ryan, M. D. (1999). Self-processing 2A-polypeptides - a system for co-ordinate expression of multiple proteins in transgenic plants. *Plant Journal* **17**(4), 453-459.

Hambidge, S. J. and Sarnow, P. (1992). Translational enhancement of the poliovirus 5' noncoding region mediated by virus-encoded polypeptide 2A. *Proc. Natl. Acad. Sci., U.S.A.* **89**, 10272-10276.

Harris, K. S., Reddigari, S. R., Nicklin, M. J. H., Hämmerle, T. and Wimmer, E. (1992). Purification and characterization of poliovirus polypeptide 3CD, a proteinase and a precursor for RNA polymerase. *Journal of Virology* **66**, 7481-7489.

Haselhoff, J. and Amos, B. (1995). GFP in plants. *Trends in Genetics* **11**, 328-329.

He, B., Gross, M. and Roizman, B. (1997). The g_{134.5} protein of herpes simplex virus 1 complexes with protein phosphatase 1a to dephosphorylate the α subunit of the eukaryotic translation initiation factor 2 and preclude the shutoff of protein synthesis by double-stranded RNA-activated protein kinase. *Proc. Natl. Acad. Sci., U.S.A.* **94**, 843-848.

Hellen, C. U. T., Kräusslich, H.-G. and Wimmer, E. (1989). Proteolytic processing of polyproteins in the replication of RNA viruses. *Biochemistry* **28**(26), 9881-9890.

Herbert, E. and Uhler, M. (1982). Biosynthesis of polyprotein precursors to regulatory peptides. *Cell* **30**, 1-2.

Holness, C. L., Lomonossoff, G. P., Evans, D. and Maule, A. J. (1989). Identification of the initiation codons for translation of cowpea mosaic virus middle component RNA using site-directed mutagenesis of an infectious cDNA clone. *Virology* **172**, 311-320.

Hu, W.-J., Harding, S.A., Lung, J., Popko, J.L., Ralph, J., Stokke, D.D., Tsai, C.-J. and Chiang, V.L. (1999). Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nature Biotechnology* **17**, 808-812.

Huez, I., Créancier, L., Audigier, S., Gensac, M.-C., Prats, A.-C. and Prats, H. (1998). Two independent internal ribosome entry sites are involved in translation initiation of vascular endothelial growth factor mRNA. *Molecular and Cellular Biology* **18**, 6178-6190.

Hughes, P. J. and Stanway, G. (2000). The 2A proteins of three diverse picornaviruses are related to each other and to the H-rev107 family of proteins involved in the control of cell proliferation. *Journal of General Virology* **81**, 201-207.

Hunt, S. L., Hsuan, J. J., Totty, N. and Jackson, R. J. (1999a). unr, a cellular cytoplasmic RNA-binding protein with five cold-shock domains, is required for internal initiation of translation of human rhinovirus RNA. *Genes and Development* **13**, 437-448.

Hunt, S. L. and Jackson, R. J. (1999). Polypyrimidine-tract binding protein (PTB) is necessary, but not sufficient, for efficient internal initiation of translation of human rhinovirus-2 RNA. *RNA* **5**, 344-359.

Hunt, S. L., Skern, T., Liebig, H.-D., Kuechler, E. and Jackson, R. J. (1999b). Rhinovirus 2A proteinase mediated stimulation of rhinovirus RNA translation is additive to the stimulation effected by cellular RNA binding proteins. *Virus Research* **62**, 119-128.

Imataka, H., Gradi, A. and Sonenberg, N. (1998). A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. *EMBO Journal* **17**(24), 7480-7489.

Imataka, H., Olsen, H. S. and Sonenberg, N. (1997). A new translational regulator with homology to eukaryotic translation initiation factor 4G. *EMBO Journal* **16**(4), 817-825.

Isoyama, T., Kamoshita, N., Yasui, K., Iwai, A., Shiroki, K., Toyoda, H., Yamada, A., Takasaki, Y. and Nomoto, A. (1999). Lower concentration of La protein required for internal ribosome entry on hepatitis C virus RNA than on poliovirus RNA. *Journal of General Virology* **80**, 2319-2327.

Jackson, R. J., Howell, M. T. and Kaminski, A. (1990). The novel mechanism of initiation of picornavirus RNA translation. *TIBS* **15**, 477-482.

Jackson, R.J. and Kaminski, A. (1995). Internal initiation of translation in eukaryotes: the picornavirus paradigm and beyond. *RNA* **1**, 985-1000.

Jacobson, M. F., Asso, J. and Baltimore, D. (1970). Further evidence on the formation of poliovirus proteins. *Journal of Molecular Biology* **49**, 657-669.

Jahnke, A., Lörz, H. and Düring, K. (1998). A repression system for genes encoding proteins toxic to bacteria in *Agrobacterium tumefaciens*-mediated plant transformation. *Journal of Biotechnology* **62**, 153-157.

Jang, S. K., Davies, M. V., Kaufman, R. J. and Wimmer, E. (1989). Initiation of protein synthesis by internal entry of ribosomes into the 5' nontranslated region of encephalomyocarditis virus RNA *in vivo*. *Journal of Virology* **63**(4), 1651-1660.

Jang, S. K., Pestova, T. V., Hellen, C. U. T., Witherell, G. W. and Wimmer, E. (1990). Cap-independent translation of picornavirus RNAs: Structure and function of the internal ribosomal entry site. *Enzyme* **44**, 292-309.

Jia, X.-Y., Ehrenfeld, E. and Summers, D. F. (1991). Proteolytic activity of hepatitis A virus 3C protein. *Journal of Virology* **65**(5), 2595-2600.

Joachims, M., Harris, K. S. and Etchison, D. (1995). Poliovirus protease 3C mediates cleavage of microtubule associated protein 4. *Virology* **211**, 451-461.

Joachims, M., Van Breuchel, P. C. and Lloyd, R. E. (1999). Cleavage of poly(A)-binding protein by enterovirus proteases concurrent with inhibition of translation *in vitro*. *Journal of Virology* **73**(1), 718-727.

Jones, R. M., Branda, J., Johnston, K. A., Polymenis, M. and Gadd, M. (1996). An essential E box in the promoter of the gene encoding the mRNA cap-binding protein (eukaryotic initiation factor 4E) is a target for activation by c-myc. *Molecular and Cellular Biology* **16**, 4754-4764.

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

Joshi, B., Cai, A. L., Keiper, B. D., Minich, W. B. and Mendez, R. (1995). Phosphorylation of eukaryotic protein synthesis initiation factor 4E at Ser-209. *Journal of Biological Chemistry* **270**, 14597-14603.

Joshi, C. P., Zhou, H., Huang, X. and Chiang, V. L. (1997). Context sequences of translation initiation codon in plants. *Plant Molecular Biology* **35**, 993-1001.

Kaminski, A., Howell, M. T. and Jackson, R. J. (1990). Initiation of encephalomyocarditis virus RNA translation: the authentic initiation site is not selected by a scanning mechanism. *EMBO Journal* **9**(11), 3753-3759.

Kaminski, A. and Jackson, R. J. (1998). The polypyrimidine tract binding protein (PTB) requirement for internal initiation of translation of cardiovirus RNAs is conditional rather than absolute. *RNA* **4**, 626-638.

Kasuga, M., Liu, Q., Miura, S., Yamaguchi-Shinozaki, K. and Shinosaki, K. (1999). Improving plant drought, salt and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nature Biotechnology* **17**, 287-291.

Kean, K. M., Teterina, N. L., Marc, D. and Girard, M. (1991). Analysis of putative active site residues of the poliovirus 3C protease. *Virology* **181**, 609-619.

Keasling, J. D. (1999). Gene-expression tools for the metabolic engineering of bacteria. *Trends in Biotechnology* **17**, 452-460.

Kerekatte, V., Keiper, B. D., Badorff, C., Cai, A., Knowlton, K. U. and Rhoads, R. E. (1999). Cleavage of poly(A)-binding protein by coxsackievirus 2A protease *in vitro* and *in vivo*: Another mechanism for host protein synthesis shutoff? *Journal of Virology* **73**(1), 709-717.

Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chambers, S. P., Markland, W., Lepre, C. A., O'Malley, E. T., Harbeson, S. L., Rice, C. M., Murcko, M. A., Caron, P. R. and Thomson, J. A. (1996). Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* **87**, 343-355.

Kim, Y. K. and Jang, S. K. (1999). La protein is required for efficient translation driven by encephalomyocarditis virus internal ribosome entry site. *Journal of General Virology* **80**, 3159-3166.

Kirchweger, R., Ziegler, E., Lamphear, B. J., Waters, D., Liebig, H.-D., Sommergruber, W., Sobrino, F., Hohenadl, C., Blaas, D., Rhoads, R. E. and Skern, T. (1994). Foot-and-mouth disease virus leader proteinase: Purification of the Lb form and determination of its cleavage site on eIF-4g. *Journal of Virology* **68**(9), 5677-5684.

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

Kokuho, T., Watanabe, S., Yokomizo, Y. and Inumaru, S. (1999). Production of biologically active, heterodimeric porcine interleukin-12 using a monocistronic baculoviral expression system. *Veterinary Immunology and Immunopathology* **72**, 289-302.

Kozak, M. (1983). Comparison of initiation protein synthesis in procaryotes, eucaryotes and organelles. *Microbiological Reviews* **47**(1), 1-45.

Kräusslich, H. G., Nicklin, M. J. H., Toyoda, H., Etchison, D. and Wimmer, E. (1987). Poliovirus proteinase 2A induces cleavage of eucaryotic initiation factor 4F polypeptide p220. *Journal of Virology* **61**(9), 2711-2718.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of the bacteriophage T4, *Nature* **227**, 680-685.

Lamphear, B., Yan, R., Yang, F., Waters, D., Liebig, H.-D., Klump, H., Kuechler, E., Skern, T. and Rhoads, R. E. (1993). Mapping the cleavage site in protein synthesis initiation factor eIF-4g of the 2A

proteases from human coxsackievirus and rhinovirus, *Journal of Biological Chemistry* **268**(26), 19200-19203.

Lamphear, B. J., Kirchweger, R., Skern, T. and Rhoads, R. E. (1995). Mapping of functional domains in eucaryotic protein synthesis initiation factor 4G (eIF4G) with picornaviral proteases. *Journal of Biological Chemistry* **270**(37), 21975-21983.

Lawrence, J. C. J. and Abraham, R. T. (1997). PHAS/4E-BPs as regulators of mRNA translation and cell proliferation. *TIBS* **22**, 345-349.

Le, S.-Y. and Maizel, J. V. (1997). A common RNA structural motif involved in the internal initiation of translation of cellular mRNAs. *Nucleic Acids Research* **25**(2), 362-369.

Lee, Y. F., Nomoto, A., Detjen, B. M. and Wimmer, E. (1977). A protein covalently linked to poliovirus genome RNA. *Proc. Natl. Acad. Sci., U.S.A.* **74**(1), 59-63.

Lee, C. K. and 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.

Lesk, A. M. and Fordham, W. D. (1996). Conservation and variability in the structures of serine proteinases of the chymotrypsin family. *Journal of Molecular Biology* **258**, 501-537.

Levy-Strumpf, N., Deiss, L. P., Berissi, H. and Kimchi, A. (1997). DAP-5, a novel homologue of eukaryotic translation initiation factor 4G isolated as a putative modulator of gamma interferon-induced programmed cell death. *Molecular and Cellular Biology* **17**(3), 1615-1625.

Li, Q. Y., Imataka, H., Morino, S., Rogers, G. W., Richter-Cook, N. J., Merrick, W. J. and Sonenberg, N. (1999). Eukaryotic translation initiation factor 4AIII, eIF4AIII is functionally distinct from eIF4AI and eIF4AII. *Molecular and Cellular Biology* **19**(11), 7336-7346.

Liebig, H.-D., Ziegler, E., Yan, R., Hartmuth, K., Klump, H., Kowlski, H., Blaas, D., Sommergruber, W., Frasel, L., Lamphear, B., Rhoads, R., Kuechler, E. and Skern, T. (1993). Purification of two picornaviral 2A proteases: Interaction with eIF-4g and influence on *in vitro* translation. *Biochemistry* **32**, 7581-7588.

Lin, C., Thomson, J. A. and Rice, C. M. (1995). A central region in the hepatitis NS4A protein allows formation of an active NS3-NS4A serine proteinase complex *in vivo* and *in vitro*. *Journal of Virology* **69**, 4373-4380.

Linder, P., Lasko, P. F., Ashburner, M., Leroy, P. and Nielsen, P. J. (1989). Birth of the D-E-A-D box. *Nature* **337**, 121-122.

Lloyd, R. E., Toyoda, H., Etchison, D., Wimmer, E. and Ehrenfeld, E. (1986). Cleavage of the cap binding protein complex polypeptide p220 is not effected by the second poliovirus protease 2A. *Virology* **150**, 229-303.

Long, A. C., Orr, D. C., Cameron, J. M., Dunn, B. M. and Kay, J. (1989). A consensus sequence for substrate hydrolysis by rhinovirus 3C proteinase. *FEBS Letters* **258**(1), 75-78.

López-Otín, C., Simón-Mateo, C., Martínez, L. and Viñuela, E. (1989). Gly-Gly-X a novel consensus sequence for the proteolytic processing of viral and cellular proteins. *Journal of Biological Chemistry* **264**, 9107-9110.

Love, R. A., Parge, H. E., Wickersham, J. A., Hostomsky, Z., Habuka, N., Moomaw, E. W., Adachi, T. and Hostomska, Z. (1996). The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell* **87**, 331-342.

Lowry, D. R. (1996). Transformation and oncogenesis: Retroviruses. In "Virology" (B. N. Fields, D. M. Knipe and P. M. Howley, Eds.), Vol. 3rd edition, pp. 235-263. Lippincott-Raven, Philadelphia.

Lütcke, H. A., Chow, K. C., Mickel, F. S., Moss, K. A., Kern, H. F. and Scheele, G. A. (1987). Selection of AUG initiation codons differs in plants and animals. *EMBO Journal* **6**(1), 43-48.

Luz, N. and Beck, E. (1991). Interaction of a cellular 57-kilodalton protein with the internal translation initiation site of foot-and-mouth disease virus. *Journal of Virology* **65**, 6486-6494.

Macadam, A. J., Ferguson, G., Fleming, T., Stone, D. M., Almond, J. W. and Minor, P. D. (1994). Role for poliovirus protease 2A in cap independent translation. *EMBO Journal* **13**(4), 924-927.

Mader, S., Lee, H., Pause, A. and Sonenberg, N. (1995). The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4g and the translational repressors 4E-binding proteins. *Molecular and Cellular Biology* **15**(9), 4990-4997.

Maia, I. G., Séron, K., Haenni, A.-L. and Bernardi, F. (1996). Gene expression from viral RNA genomes. *Plant Molecular Biology* **32**, 367-391.

- Maitra, U., Stringer, E. A. and Chaudhuri, A. (1982). Initiation factors in protein biosynthesis. *Annual Reviews of Biochemistry* **51**, 869-900.
- Malcolm, B. A. (1995). The picornaviral 3C proteinases: Cysteine nucleophiles in serine proteinase folds. *Protein Science* **4**, 1439-1445.
- Mant, A. and Robinson, C. (1998). An arabidopsis cDNA encodes an apparent polyprotein of two non-identical thylakoid membrane proteins that are associated with photosystem II and homologous to algal ycf32 open reading frames. *FEBS Letters* **423**, 183-188.
- Marcos, J. F. and Beachy, R. N. (1994). *In vitro* characterization of a cassette to accumulate multiple proteins through synthesis of a self-processing polypeptide. *Plant Molecular Biology* **24**, 495-503.
- Marcos, J. F. and Beachy, R. N. (1997). Transgenic accumulation of two plant virus coat proteins on a single self-processing polypeptide. *Journal of General Virology* **78**, 1771-1778.
- Matsuda, A. and Komatsu, K.-I. (1985). Molecular-cloning and structure of the gene for 7-beta-(4-carboxybutanamido) cephalosporanic acid acylase from a *pseudomonas* strain. *Journal of Bacteriology* **163**, 1222-1228.
- Matthews, D. A., Smith, W. A., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., Gribskov, C. L. and Worland, S. (1994). Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and a means for cleaving precursor polyprotein. *Cell* **77**, 761-771.
- Mattion, N. M., Harnish, E. C., Crowley, J. C. and 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. *Journal of Virology* **70**, 8124-8127.
- May Wang, Q., Sommergruber, W. and Johnson, R.B. (1997). Cleavage specificity of Human Rhinovirus-2 2A protease for peptide substrates. *Biochemical and biophysical research communications* **235**, 562-566.
- May Wang, Q., Johnson, R.B., Cox, G.A., Villarreal, E.C., Churgay, L.M. and Hale, J.E. (1998). Enzymatic characterization of refolded human rhinovirus type 14 2A protease expressed in *Escherichia coli*. *Journal of Virology* **72**(2), 1683-1687.

Medina, M., Domingo, E., Brangwyn, J. K. and 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.

Meek, T. D. (1998). Catalytic mechanisms of the aspartic proteinases. In "Comprehensive biological catalysis", Vol. Volume 1/Reactions of electrophilic carbon, phosphorus and sulfur, pp. 327-344. 4 vols. Academic Press Ltd.

Merrick, W. C. (1992). Mechanism and regulation of eukaryotic protein synthesis. *Microbiological Reviews* **56**(2), 291-315.

Metz, A. M. and Browning, K. S. (1996). Mutational analysis of the functional domains of the large subunit of the isozyme form of wheat initiation factor eIF-4F. *Journal of Biological Chemistry* **271**(49), 31033-31036.

Miller, R. H. and Purcell, R. H. (1990). Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. *Proc. Natl. Acad. Sci., U.S.A.* **87**, 2057-2061.

Mock, W. L. (1998). Zinc proteinases. In "Comprehensive biological catalysis", Vol. Volume 1/ Reactions of electrophilic carbon, phosphorus and sulfur, pp. 425-455 4 vols. Academic press Ltd.

Moldave, K. (1985). Eukaryotic protein synthesis. *Annual Reviews of Biochemistry* **54**, 1109-1149.

Molla, A., Paul, A. V., Schmid, M., Jang, S. K. and Wimmer, E. (1993). Studies on dicistronic polioviruses implicate viral proteinase 2A^{Pro} in RNA replication. *Virology* **196**, 739-747.

Morino, S., Imataka, H., Svitkin, Y. V., Pestova, T. V. and Sonenberg, N. (2000). Eukaryotic translation initiation factor 4E (eIF4E) binding site and the middle one-third of eIF4GI constitute the core domain for cap-dependent translation, and the C-terminal one-third functions as a modulatory region. *Molecular and Cellular Biology* **20**(2), 468-477.

Morley, S. J., Curtis, P. S. and Pain, V. M. (1997). eIF4G: Translation's mystery factor begins to yield its secrets. *RNA* **3**, 1085-1104.

Mushegian, A. R. and Shepherd, R. J. (1995). Genetic elements of plant viruses as tools for genetic engineering. *Microbiological Reviews* **59**(4), 548-578.

- Nanbru, C., Lafon, I., Audigier, S., Gensac, M.-C., Vagner, S., Huez, G. and Prats, A.-C. (1997). Alternative translation of the proto-oncogene *c-myc* by an internal ribosome entry site. *Journal of Biological Chemistry* **272**(51), 32061-32066.
- Nicholson, R., Pelletier, J., Le, S.-Y. and Sonenberg, N. (1991). Structural and functional analysis of the ribosome landing pad of poliovirus type 2: *in vivo* translation studies. *Journal of Virology* **65**, 5886-5894.
- Nicklin, M.J.H., Kräusslich, H.G., Toyoda, H., Dunn, J.J. and Wimmer, E. (1987). Poliovirus polypeptide precursors: Expression *in vitro* and processing by exogenous 3C and 2A proteinases. *Proc. Natl. Acad. Sci. USA* **84**, 4002-4006.
- Nicklin, M. J. H., Harris, K. S., Pallai, P. V. and Wimmer, E. (1988). Poliovirus proteinase 3C: Large-scale expression, purification and specific cleavage activity on natural and synthetic substrates *in vitro*. *Journal of Virology* **62**(12), 4586-4593.
- Nielsen, P. J. and Trachsel, H. (1988). The mouse protein-synthesis initiation factor-4A gene family includes 2 related functional genes which are differentially expressed. *EMBO Journal* **7**, 2097-2105.
- Niepmann, M. (1996). Porcine polypyrimidine tract-binding protein stimulates translation initiation at the internal ribosome entry site of foot-and-mouth disease virus. *FEBS Letters* **388**, 39-42.
- Niepmann, M., Petersen, A., Meyer, K. and Beck, E. (1997). Functional involvement of polypyrimidine tract-binding protein in translation initiation complexes with the internal ribosome entry site of foot-and-mouth disease virus. *Journal of Virology* **71**(11), 8330-8339.
- Odell, J. T., Nagy, F. and Chua, N.-H. (1985). Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature* **313**, 810-812.
- Oh, C. and Carrington, J. C. (1989). Identification of essential residues in potyvirus proteinase HC-Pro by site-directed mutagenesis. *Virology* **173**, 692-699.
- Ohlmann, T., Pain, V. M., Wood, W., Rau, M. and Morley, S. J. (1997). The proteolytic cleavage of eukaryotic initiation (eIF) 4G is prevented by eIF4E binding protein (PHAS; 4E-BP1) in the reticulocyte lysate. *EMBO Journal* **16**(4), 844-855.

Ohlmann, T., Rau, M., Morley, S. J. and Pain, V. M. (1995). Proteolytic cleavage of initiation factor eIF-4g in the reticulocyte lysate inhibits translation of capped mRNAs but enhances that of uncapped mRNAs. *Nucleic Acids Research* **23**(3), 334-340.

Ohlmann, T., Rau, M., Pain, V. M. and Morley, S. J. (1996). The C-terminal domain of eucaryotic protein synthesis initiation factor (eIF) 4G is sufficient to support cap-independent translation in the absence of eIF4E. *EMBO Journal* **15**(6), 1371-1382.

Pain, V. M. (1996). Initiation of protein synthesis in eukaryotic cells. *European Journal of Biochemistry* **236**, 747-771.

Palmenberg, A. C. (1987a). Genome organization, translation and processing in picornaviruses. In "The molecular biology of the positive strand RNA viruses" (D. J. Rowlands, A. M. Mayo and B. W. J. Mahy, Eds.), Vol. 32, pp. 1-15. Academic Press Ltd.

Palmenberg, A. C. (1987b). Picornaviral processing: Some new ideas. *Journal of Cellular Biochemistry* **33**, 191-198.

Pande, S., Vimaladithan, A., Zhao, H. and Farabaugh, P. J. (1995). Pulling the ribosome out of frame by +1 at a programmed frameshift site by cognate binding of aminoacyl-tRNA. *Molecular and Cellular Biology* **15**(1), 298-304.

Parks, G. D., Baker, J. C. and Palmenberg, A. C. (1989). Proteolytic cleavage of encephalomyocarditis virus capsid region substrates by precursors to the 3C enzyme. *Journal of Virology* **63**, 1054-1058.

Paul, A. V., Mugavero, J., Molla, A. and Wimmer, E. (1998). Internal ribosome entry site scanning of the poliovirus polyprotein: Implications for proteolytic processing. *Virology* **250**, 241-253.

Pavloff, N., Pemberton, P. A., Potempa, J., Chen, W.-C. A., Pike, R. N., Prochazka, V., Kiefer, M. C., Travis, J. and Barr, P. J. (1997). Molecular cloning and characterization of *Porphyromonas gingivalis* lysine-specific gingipain. *Journal of Biological Chemistry* **272**(3), 1595-1600.

Peach, C. and Velten, J. (1991). Transgene expression variability (position effect) of CAT and GUS reporter genes driven by linked divergent T-DNA promoters. *Plant Molecular Biology* **17**, 49-60.

Pearl, L. H. and Taylor, W. R. (1987). A structural model for the retroviral proteases. *Nature* **329**, 351-354.

- Pelham, H. R. (1978). Translation of encephalomyocarditis virus RNA *in vitro* yields an active proteolytic enzyme. *European Journal of Biochemistry* **85**, 457-462.
- Pelletier, J., Kaplan, G., Racaniello, V. R. and Sonenberg, N. (1988). Cap-independent translation of poliovirus mRNA is conferred by sequence elements within the 5' noncoding region. *Molecular and Cellular Biology* **8**(3), 1103-1112.
- Pelletier, J. and Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334**, 320-325.
- Percy, N., Barclay, W. S., García-Sastre, A. and Palese, P. (1994). Expression of a foreign protein by influenza A virus. *Journal of Virology* **68**, 4486-4492.
- Pestova, T. V., Hellen, C. U. T. and Shatsky, I. N. (1996). Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Molecular and Cellular Biology* **16**(12), 6859-6869.
- Petersen, J. F. W., Cherney, M. M., Liebig, H.-D., Skern, T., Kuechler, E. and James, M. N. G. (1999). The structure of the 2A proteinase from a common cold virus: a proteinase responsible for the shut-off of host-cell protein synthesis. *The EMBO Journal* **18**(20), 5463-5475.
- Piccone, M. E., Reider, E., Mason, P. W. and Grubman, M. J. (1995a). The foot-and-mouth disease virus leader proteinase gene is not required for viral replication. *Journal of Virology* **69**(9), 5376-5382.
- Piccone, M. E., Zellner, M., Kumosinski, T. F., Mason, P. W. and Grubman, M. J. (1995b). Identification of the active-site residues of the L-proteinase of foot-and-mouth disease virus. *Journal of Virology* **69**(8), 4950-4956.
- Pirone, T. P. and Thornbury, D. W. (1983). Role of virion and helper component in regulating aphid transmission of tobacco etch virus. *Phytopathology* **73**, 872-875.
- Precious, B., Young, D. F., Bermingham, A., Fearn, R., Ryan, M. D. and Randall, R. E. (1995). Inducible expression of the P, V and NP genes of the paramyxovirus Simian virus 5 in cell lines and an examination of NP-P and NP-V interactions. *Journal of Virology* **69**, 8001-8010.
- Pyronnet, S., Imataka, H., Gingras, A.-C., Fukunaga, R., Hunter, T. and Sonenberg, N. (1999). Human eukaryotic initiation factor 4G (eIF4G) recruits Mnk1 to phosphorylate eIF4E. *EMBO Journal* **18**(1), 270-279.

- Reichmann, J. L., Lain, S. and Garcia, J. A. (1992). Highlights and prospects of potyvirus molecular biology. *Journal of General Virology* **73**, 1-16.
- Reinisch, K.M., Nibert, M. and Harrison, S.C. (2000). Structure of the reovirus core at 3.6 angstrom resolution. *Nature* **404**(6781), 960-967
- Richter-Cook, N. J., Dever, T. E., Hensold, J. O. and Merrick, W. C. (1998). Purification and characterization of a new eukaryotic protein translation factor. *Journal of Biological Chemistry* **273**(13), 7579-7587.
- Roberts, L. O., Seamons, R. A. and Belsham, G. J. (1998). Recognition of picornavirus internal ribosome entry sites within cells; influence of cellular and viral proteins. *RNA*. **4**, 520-529.
- Roberts, P. J. and 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.
- Rose, J. K., Trachsel, H., Leong, K. and Baltimore, D. (1978). Inhibition of translation by poliovirus: Inactivation of a specific initiation factor. *Proc. Natl. Acad. Sci., U.S.A.* **75**(6), 2732-2736.
- Rosenwald, I. B., Rhoads, D. B., Callanan, L. D., Isselbacher, K. J. and Schmidt, E. V. (1993). Increased expression of eukaryotic translation initiation factors eIF4E and eIF-2-alpha in response to growth induction by c-myc. *Proc. Natl. Acad. Sci., U.S.A.* **90**, 6175-6178.
- Rothberg, P. G., Harris, T. J. R., Nomoto, A. and Wimmer, E. (1978). O⁴-(5'uridylyl) tyrosine is the bond between the genome linked protein and the RNA of poliovirus. *Proc. Natl. Acad. Sci., U.S.A.* **75**, 4868-4872.
- Rousseau, D., Kaspar, R., Rosenwald, I., Gehrke, L. and Sonenberg, N. (1996). Translation initiation of ornithine carboxylase and nucleocytoplasmic transport of cyclin D1 mRNA are increased in cells overexpressing eukaryotic initiation factor 4E. *Proc. Natl. Acad. Sci., U.S.A.* **93**, 1065-1070.
- Rowe, A., Ferguson, G.L., Minor, P.D. and Macadam, A.J. (2000). Coding changes in the poliovirus protease 2A compensate for 5'NCR domain V disruptions in a cell-specific manner. *Virology* **269**, 284-293.
- Rubinstein, S. J. and Dasgupta, A. (1989). Inhibition of rRNA synthesis by poliovirus: Specific inactivation of transcription factors. *Journal of Virology* **63**(11), 4689-4696.

Rueckert, R. R. (1996). Picornaviridae: the viruses and their replication. In "Virology" (B. N. Fields, D. M. Knipe and P. M. Howley, Eds.), Vol. 3rd edition, pp. 609-654. Lippincott-Raven, Philadelphia.

Ryan, M. D. and Drew, J. (1994). Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO Journal* **13**(4), 928-933.

Ryan, M. D. and Flint, M. (1997). Virus-encoded proteinases of the picornavirus super-group. *Journal of General Virology* **78**, 699-723.

Ryan, M. D., Monaghan, S. and Flint, M. (1998). Virus-encoded proteinases of the flaviviridae. *Journal of General Virology* **79**, 947-959.

Ryan, M. D., Luke, G., Hughes, L. E., Cowton, V. M., Li, X., Donnelly, M. L. L., Mehorta, A. and Gani, D. (In press). The Aphtho- and Cardiovirus 'primary' 2A/2B polyprotein 'cleavage'. In "The Picornaviruses" (Semler, B., and Wimmer, E., Eds.), Plenum Press.

Rychlik, W., Russ, M. A. and Rhoads, R. E. (1987). Phosphorylation site of eukaryotic initiation factor 4E. *Journal of Biological Chemistry* **262**, 10434-10437.

Sachs, A. B., Sarnow, P. and Hentze, M. W. (1997). Starting at the beginning, middle, and end: Translation initiation in eukaryotes. *Cell* **89**, 831-838.

Samuel, C. E. (1993). The eIF-2a protein kinases, regulators of translation in eukaryotes from yeasts to human. *Journal of Biological Chemistry* **268**, 7603-7606.

Sangar, D. V., Newton, S. E., Rowlands, D. J. and Clarke, B. E. (1987). All foot-and-mouth disease virus serotypes initiate protein synthesis at two separate AUGs. *Nucleic Acids Research* **15**, 3305-3315.

Santa Cruz, S., Chapman, S., Roberts, A. G., Roberts, I. M., Prior, D. A. M. and Oparka, K. J. (1996). Assembly and movement of a plant virus carrying a green fluorescent protein overcoat. *Proc. Natl. Acad. Sci., U.S.A.* **93**, 1-5.

Schechter, I. and Berger, A. (1967). On the size of the active site in proteases. I. papain. *Biochemical and Biophysical Research Communications* **27**, 157-162.

Schober, D., Kronenberger, P., Prchla, E., Blaas, D. and Fuchs, R. (1998). Major and minor receptor group human rhinoviruses penetrate from endosomes by different mechanisms. *Journal of Virology* **72**, 1354-1364.

Schulthies, T., Kusov, Y. Y. and Gauss-Müller, V. (1994). Proteinase 3C of hepatitis A virus (HAV) cleaves the HAV polyprotein P2-P3 at all sites including VP1/2A and 2A/2B. *Virology* **198**, 275-281.

Schumacher, G., Sizman, D., Haug, H., Buckel, P. and Böck, A. (1986). Penicillin acylase from *Escherichia coli* - unique gene protein relation. *Nucleic Acids Research* **14**, 5713-5727.

Seemüller, E., Lupas, A. and Baumeister, W. (1996). Autocatalytic processing of the 20S proteasome. *Nature* **382**, 468-470.

Sha, M., Wang, Y., Xiang, T., van Heerden, A., Browning, K. S. and Goss, D. J. (1995). Interaction of wheat germ protein synthesis initiation factor eIF-(iso)4F and its subunits p28 and p86 with a m7GTP and mRNA analogues. *Journal of Biological Chemistry* **270**(50), 29904-29909.

Shaughnessy, J. D., Jenkins, N. A. and Copeland, N. G. (1997). cDNA cloning, expression analysis and chromosomal localization of a gene with high homology to wheat eIF-(iso)4F and mammalian eIF-4G. *Genomics* **39**, 192-197.

Simón-Mateo, C., Andrés, G., Almazán, F. and Viñuela, E. (1997). Proteolytic processing in African swine fever virus: Evidence for a new structural polyprotein, pp62. *Virology* **71**(8), 5799-5804.

Simón-Mateo, C., Andrés, G. and Viñuela, E. (1993). Polyprotein processing in African swine fever virus: a novel gene expression strategy for a DNA virus. *EMBO Journal* **12**(7), 2977-2987.

Singh, M. and Sharma, C. (1991). Detection of plasmid-encoded gusA gene in GUS-positive *Escherichia coli*. *BioTechniques* **26**, 261-264.

Skern, T., Fita, I. and Guarne, A. (1998). A structural model of picornavirus leader proteinases based on papain and bleomycin hydrolase. *Journal of General Virology* **79**, 301-307.

Skern, T., Sommergruber, W., Auer, H., Volkmann, P., Zorn, M., Liebig, H.-D., Fessl, F., Blaas, D. and Kuechler, E. (1991). Substrate requirements of a human rhinoviral 2A proteinase. *Virology* **181**, 46-54.

Slakeski, N., Cleal, S. M. and Reynolds, E. C. (1996). Characterization of a *Porphyromonas gingivalis* gene prtR that encodes an arginine-specific thiol proteinase and multiple adhesins. *Biochemical and Biophysical Research Communications* **224**, 605-610.

Slater, S., Mitsky, T.A., Houmiel, K.L., Hao, M., Reiser, S.E., Taylor, N.B., Tran, M., Valentin, H.E., Rodriguez, D.J., Stone, D.A., Padgett, S.R., Kishore, G. and Gruys, K.J. (1999). Metabolic engineering of *Arabidopsis* and *Brassica* for poly(3-hydroxybutyrate-co-3-hydroxyvalerate) copolymer production. *Nature Biotechnology* **17**, 1011-1016.

Smerdou, C. and Lilijestrom, P. (1999). Two-helper RNA system for production of recombinant Semliki Forest virus particles. *Journal of Virology* **73**(2), 1092-1098.

Smolenska, L., Roberts, I. M., Learmonth, D., Porter, A. J., Harris, W. J. and Santa Cruz, S. (1998). Production of a functional single chain antibody attached to the surface of a plant virus. *FEBS Letters* **441**, 379-382.

Sommergruber, W., Casari, G., Fessl, F., Seipelt, J. and Skern, T. (1994). The 2A proteinase of human rhinovirus is a zinc containing enzyme. *Virology* **204**, 815-818.

Spall, V. E., Shanks, M. and Lomonosoff, G. P. (1997). Polyprotein processing as a strategy for gene expression in RNA viruses. *Seminars in Virology* **8**, 15-23.

Spångberg, K. and Schwartz, S. (1999). Poly(C)-binding protein interacts with the hepatitis C virus 5'untranslated region. *Journal of General Virology* **80**, 1371-1376.

Stam, M., Mol, J. N. M. and Kooter, J. M. (1997). The silence of genes in transgenic plants. *Annals of Botany* **79**, 3-12.

Stansfield, I., Jones, K. M. and Tuite, M. F. (1995). The end in sight: terminating translation in eukaryotes. *TIBS* **20**, 489-491.

Stanway, G. (1990). Structure, function and evolution of picornaviruses. *Journal of General Virology* **71**, 2483-2501.

Stanway, G. and Hyypia, T. (1999). Parechoviruses. *Journal of Virology* **73**(7), 5249-5254.

Stein, I., Itin, A., Einat, P., Skaliter, R., Grossman, Z. and Keshet, E. (1998). Translation of vascular endothelial growth factor mRNA by internal ribosome entry: Implications for translation under hypoxia. *Molecular and Cellular Biology* **18**, 3112-3119.

Stewart, S. R. and Semler, B. L. (1997). RNA determinants of picornavirus cap-independent translation initiation. *Seminars in Virology* **8**, 242-255.

Stöger, E., Vaquero, C., Torres, E., Sack, M., Nicholson, L., Drossard, J., Williams, S., Keen, D., Perrin, Y., Christou, P. and Fischer, R. (2000). Cereal crops as viable production and storage systems for pharmaceutical ScFv antibodies. *Plant Molecular Biology* **42**, 583-590.

Strebel, K. and Beck, E. (1986). A second protease of foot-and-mouth disease virus. *Journal of Virology* **58**, 893-899.

Summers, D. F. and Maizel, J. V. (1968). Evidence for large precursor proteins in poliovirus synthesis. *Proc. Natl. Acad. Sci., U.S.A.* **59**, 966-971.

Svitkin, Y.V., Ovchinnikov, L.P., Dreyfuss, G. and Sonenberg, N. (1996). General RNA binding proteins render translation cap dependent. *EMBO Journal* **15**(24), 7147-7155.

Tang, J. and Wong, R. N. S. (1987). Evolution in the structure and function of aspartic proteases. *Journal of Cellular Biochemistry* **33**, 53-63.

Tanji, Y., Hijikata, M., Satoh, S., Kaneko, T. and Shimotohno, K. (1995). Hepatitis C virus-encoded nonstructural protein NS4A has versatile functions in viral protein processing. *Journal of Virology* **69**, 1575-1581.

Teerink, H., Voorma, H. O. and Thomas, A. A. M. (1995). The human insulin-like growth factor II leader 1 contains an internal ribosomal entry site. *Biochimica et Biophysica Acta* **1264**, 403-408.

Thöny-Meyer, L., Böck, A. and Hennecke, H. (1992). Prokaryotic polyprotein precursors. *FEBS Letters* **307**, 62-65.

Tinland, B., Schoumacher, F., Gloeckler, V., Bravoangel, A. M. and Hohn, B. (1995). The *Agrobacterium tumefaciens* virulence D2 protein is responsible for precise integration of T-DNA into the plant genome. *EMBO Journal* **14**(14), 3585-3595.

Tomassini, J. E., Graham, D., DeWitt, C. M., Lineberger, D. W., Rodkey, J. A. and Colonno, R. J. (1989). cDNA cloning reveals that the major group rhinovirus receptor on HeLa cells is intercellular adhesion molecule 1. *Proc. Natl. Acad. Sci., U.S.A.* **86**, 4907-4911.

Toyoda, H., Nicklin, M. J. H., Murray, M. G., Anderson, C. W., Dunn, J. J., Studier, F. W. and Wimmer, E. (1986). A second virus-encoded proteinase involved in proteolytic processing of poliovirus protein. *Cell* **45**, 761-770.

Uncapher, C. R., DeWitt, C. M. and Colonno, R. J. (1991). The major and minor receptor families contain all but one human rhinovirus serotype. *Virology* **180**, 814-817.

Urzainqui, A. and Carrasco, L. (1989). Degradation of cellular proteins during poliovirus infection: studies by two-dimensional gel electrophoresis. *Journal of Virology* **63**, 4729-4735.

Vagner, S., Gensac, M.-C., Maret, A., Bayard, F., Amalric, F., Prats, H. and Prats, A.-C. (1995). Alternative translation of human fibroblast growth factor 2 mRNA occurs by internal entry of ribosomes. *Molecular and Cellular Biology* **15**(1), 35-44.

Varnavski, A. N. and Khromykh, A. A. (1999). Noncytopathic flavivirus replicon RNA-based system for expression and delivery of heterologous genes. *Virology* **255**, 366-375.

Ventoso, I., Barco, A. and Carrasco, L. (1998). Mutational analysis of poliovirus 2A^{PRO}. *Journal of Biological Chemistry* **273**(43), 27960-27967.

Ventoso, I., Barco, A. and Carrasco, L. (1999). Genetic selection of poliovirus 2A^{PRO}-binding peptides. *Journal of Virology* **73**(1), 814-818.

Ventoso, I., MacMillan, S. E., Hershey, J. W. B. and Carrasco, L. (1998). Poliovirus 2A proteinase cleaves directly the eIF-4G subunit of eIF4F complex. *FEBS Letters* **435**, 79-83.

Verchot, J., Herndon, K. L. and Carrington, J. C. (1991). The 35-kDa protein from the N-terminus of the potyviral polyprotein functions as a third virus-encoded proteinase. *Virology* **185**, 527-535.

Vos, P., Verver, J., Jaegle, M., Wellink, J., van Kammen, A. and Goldbach, R. (1988). Two viral proteins involved in the proteolytic processing of the cowpea mosaic virus polyproteins. *Nucleic Acids Research* **16**, 1967-1985.

- Walter, B.L., Nguyen, J.H.C., Ehrenfeld, E. and Semler, B.L. (1999). Differential utilization of poly(rC) binding protein 2 in translation directed by picornavirus IRES elements. *RNA* **5**, 1570-1585.
- Wang, D. and Maule, A. J. (1995). Inhibition of host gene expression associated with plant virus replication. *Science* **267**, 229-231.
- Wang, Q. M., Sommergruber, W. and Johnson, R. B. (1997). Cleavage specificity of human rhinovirus-2 2A protease for peptide substrates. *Biochemical and Biophysical Research Communications* **235**, 562-566.
- Wharton, C. W. (1998). The serine proteinases. In "Comprehensive biological catalysis", Vol. Volume 1/ Reactions of electrophilic carbon, phosphorus and sulfur, pp. 345-381 4 vols. Academic Press Ltd.
- Wilson, K. S. and Noller, H. F. (1998). Molecular movement inside the translational engine. *Cell* **92**, 337-349.
- Wu, Z., Yao, N., Le, H. V. and Weber, P. C. (1998). Mechanism of autoproteolysis at the NS2-NS3 junction of the hepatitis C virus polyprotein. *TIBS* **23**, 92-94.
- Wyckoff, E. E. (1993). Inhibition of host cell protein synthesis in poliovirus-infected cells. *Seminars in Virology* **4**, 209-215.
- Xu, W.-H., Sato, Y., Ikeda, M. and Yamashita, O. (1995). Molecular characterization of the gene encoding the precursor protein of diapause hormone and pheromone biosynthesis activating neuropeptide (DH-PBAN) of the silkworm, *Bombyx mori* and its distribution in some insects. *Biochimica et Biophysica Acta* **1261**, 83-89.
- Yalamanchili, P., Banerjee, R. and Dasgupta, A. (1997). Poliovirus-encoded protease 2A^{PRO} cleaves the TATA-binding protein but does not inhibit host cell RNA polymerase II transcription *in vitro*. *Journal of Virology* **71**(9), 6881-6886.
- Yalamanchili, P., Datta, U. and Dasgupta, A. (1997). Inhibition of host cell transcription by poliovirus: Cleavage of transcription factor CREB by poliovirus-encoded protease 3C^{PRO}. *Journal of Virology* **71**(2), 1220-1226.
- Yamashita, O. (1996). Diapause hormone of the silkworm, *Bombyx mori*: Structure, gene expression and function. *Journal of Insect Physiology* **42**(7), 669-679.

Yang, Q. and Sarnow, P. (1997). Location of the internal ribosome entry site in the 5' non-coding region of the immunoglobulin heavy-chain binding protein (BiP) mRNA: Evidence for specific RNA-protein interactions. *Nucleic Acids Research* **25**(14), 2800-2807.

Ypma-Wong, M. F., Dewalt, P. G., Johnson, V. H., Lamb, J. G. and Semler, B. L. (1988). Protein 3CD is the major poliovirus proteinase responsible for cleavage of the P1 capsid precursor. *Virology* **166**, 265-270.

Yu, S. F. and Lloyd, R. E. (1991). Identification of essential amino acid residues in the functional activity of poliovirus 2A protease. *Virology* **182**, 615-625.

Yu, S. F. and Lloyd, R. E. (1992). Characterization of the roles of conserved cysteine and histidine residues in poliovirus 2A protease. *Virology* **186**, 725-735.

Zaccomer, B., Haenni, A.-L. and Macaya, G. (1995). The remarkable variety of plant RNA virus genomes. *Journal of General Virology* **76**, 231-247.

Zeichardt, H., Wetz, K., Willingmann, P. and Habermehl, K.O. (1985). Entry of the poliovirus type 1 and mouse Eberfeld (ME) virus into HEp-2 cells: receptor mediated endocytosis and endosomal or lysosomal uncoating. *Journal of General Virology* **66**, 483-492.

Zeigler, E., Borman, A.M., Deliat, F.G., Liebig, H-D., Jugovic, D., Kean, K.M., Skern, T. and 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.