Polyprotein Processing By Potato Leafroll Virus

A thesis submitted in partial fulfilment of

the requirements for the degree of

Doctor of Philosophy

By

Xuejun Li (BSc., MSc.)

School of Biology

University of St. Andrews

November 2000

Declarations

I, Xuejun Li, hereby certify that this thesis, which is approximately 50,000 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 30/11/00 Signature of candidate

I was admitted as a research student in September, 1995 and as a candidate for the degree of Doctor of Philosophy in September, 1995; the higher study for which this is a record was carried out in the University of St. Andrews between 1995 and 2000.

Date 30/11/00 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 1/12/00 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 .....

Abstract

P1 and P2 are the only two viral genes essential for the RNA replication of potato leafroll virus (PLRV). Multiple domains have been identified within these two proteins on the basis of sequence homology. The hypothesis that the protease motif within the P1 protein was involved in the processing of P1 and P2 proteins was confirmed for the first time, after being suspected for some ten years.

Part of the P1 protein was expressed in *E. coli* and used to raise antibodies, which were used to identify products of proteolytic processing. A baculovirus system was developed to study the processing of P1 and P1/2 transframe proteins. Two cleavage sites were identified. The cleavage at the N-terminus of VPg was shown to be mediated by the P1 protein, and this cleavage could occur *in trans*. Site-directed mutagenesis has identified four conserved residues (His-255, Asp-286, Ser-354 and Gly-355), consistent with the proposal that a serine protease motif is essential for the catalysis. The processing of P1 and P2 proteins was not observed *in vitro*. A model for the processing of P1 and P2 proteins and the RNA replication of PLRV is proposed.

Acknowledgements

I take this opportunity to express my indebtedness to my supervisors, Dr. Joe Lamb and Dr. Martin Ryan, for their invaluable supervision, advice, support and encouragement throughout this investigation and the writing of this thesis. I would also like to thank Professor Ron Hay, Dr. Rick Randall and Dr. Graham Kemp for their help in baculovirus, protein purification, antibody production, protein-tagging and viral protease work. I thank Dr. Mike Mayo at SCRI for providing plant materials and information on plant virus research. I thank John Nicolson, Ellis Jaffray, Dan Young and all the guys in the old Irvine Building and the BMS Building, who help me through the course of this Ph.D. Thanks also to Dr. Gary Luke and the rest of the group, whose friendliness has made my days in St. Andrews memorable.

Special thanks go to Dr. Huangting Liu, for his critical comments and enormous help, especially in plant virology and mutagenesis area, and to Mr. Changjiang Dong, for help and advice.

I owe a huge gratitude to my parents for their constant encouragement.

Thanks to the University of St. Andrews and the Department for Education and Employment (DfEE) for giving me a studentship and ORS award.

Dedicated to my wife Hongfan Chi

Abbreviations

AVP	adenovirus protease
BSN	Bjerrum and Schafer-Nielsen transfer buffer
BV	budded virus
CMM	canine pancreatic microsomal membranes
CPA	carboxypeptidase A
CPK	Corey, Pauling and Koltun colour scheme
3,4-DCI	3,4-dichloroisocoumarin
DFP	diisopropl fluorophosphate
DTT	dithiothreitol
E64	3-carboxy-trans-2,3-epoxypropyl-leucylamido(4-granidino)butane
ECL	enhanced chemiluminescence
E. coli	*Esherichia coli*
EDTA	ethylenediaminetetra-acetic acid
ELISA	enzyme linked immunosorbent assay
ER	endoplasmic reticulum
FCS	fetal cow serum
fs	frameshift
gag	group-specific antigens
GST	glutathione S-transferase
GUS	β-glucuronidase
HC-pro	helper component protease
IRES	internal ribosome entry site
IPTG	isopropylthio-β-D-galactoside
kDa	kilo Daltons

LB	Luria-Bertani medium
MCS	multiple cloning site
MOI	multiplicity of infection
MW	molecular weight
NCR	non-coding region
NLS	nuclear localization signal
n.t.	nucleotide(s)
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PIB	polyhedral inclusion body
PMSF	phenylmethylsulfonyl fluoride
PVP	polyvinylpyrrolidone
RdRp	RNA-dependent RNA polymerase
RRL	rabbit reticulocyte lysate
SDS	sodium dodecyl sulphate
sgRNA	subgenomic RNA
SGPA, B	*Streptomyces griseus* protease A, B
RTD	readthrough domain
TMB	3,3',5,5'-tetramethylbenzidine
TnT	coupled transcription /translation
UTR	untranslated region
VPg	virion protein, genome
WGE	wheat germ extract

Virus Abbreviations:

AcMNPV	*Autographa californica* multiple nuclear polyhedrosis virus
Ad	adenovirus
BDV	border disease virus
BmNPV	*Bombyx mori* nuclear polyhedrosis virus
BMYV	beet mild yellowing virus
BVDV	bovine viral diarrhea virus
BWYV	beet western yellows virus
BYDV	barley yellow dwarf virus
CABYV	cucurbit aphid-borne yellows virus
CMV	cytomeglovirus
CPMV	cowpea mosaic virus
CSFV	classical swine fever virus
EMCV	encephalomyocarditis virus
ERV	equine rhinovirus
FMDV	foot-and-mouth disease virus
GFLV	grapevine fanleaf virus
HAV	hepatitis A virus
HCV	hepatitis C virus
HIV	human immunodeficiency virus
PEMV	pea enation mosaic virus
PLRV	potato leafroll virus
RHDV	rabbit hemeorrhagic disease virus
RYMV	rice yellow mottle virus
SBMV	bean southern mosaic virus
SDV	soybean dwarf virus

SFV	Semliki Forest virus
SIN	Sindbis virus
SV5	simian virus 5
TEV	tobacco etch virus
TMEV	Theiler's murine encephalomyelitis virus
TMV	tobacco mosaic virus
ToRSV	tomato ringspot virus
TVMV	tobacco vein mottling virus
YF	yellow fever virus

Amino Acid Abbreviations

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

Universal Genetic Code

First	Second				Third
	T	C	A	G	
T	TTT Phe (F)	TCT Ser (S)	TAT Tyr (Y)	TGT Cys (C)	T
	TTC Phe (F)	TCC Ser (S)	TAC Tyr (Y)	TGC Cys (C)	C
	TTA Leu (L)	TCA Ser (S)	TAA Ter	TGA Ter	A
	TTG Leu (L)	TCG Ser (S)	TAG Ter	TGG Trp (W)	G
C	CTT Leu (L)	CCT Pro (P)	CAT His (H)	CGT Arg (R)	T
	CTC Leu (L)	CCC Pro (P)	CAC His (H)	CGC Arg (R)	C
	CTA Leu (L)	CCA Pro (P)	CAA Gln (Q)	CGA Arg (R)	A
	CTG Leu (L)	CCG Pro (P)	CAG Gln (Q)	CGG Arg (R)	G
A	ATT Ile (I)	ACT Thr (T)	AAT Asn (N)	AGT Ser (S)	T
	ATC Ile (I)	ACC Thr (T)	AAC Asn (N)	AGC Ser (S)	C
	ATA Ile (I)	ACA Thr (T)	AAA Lys (K)	AGA Arg (R)	A
	ATG Met (M)	ACG Thr (T)	AAG Lys (K)	AGG Arg (R)	G
G	GTT Val (V)	GCT Ala (A)	GAT Asp (D)	GGT Gly (G)	T
	GTC Val (V)	GCC Ala (A)	GAC Asp (D)	GGC Gly (G)	C
	GTA Val (V)	GCA Ala (A)	GAA Glu (E)	GGA Gly (G)	A
	GTG Val (V)	GCG Ala (A)	GAG Glu (E)	GGG Gly (G)	G

A: Adenosine

C: Cytidine

G: Guanosine

T: Thymidine

Ter: Termination Signal

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Potato Leafroll Virus: An Overview

Potato leafroll virus (PLRV) is the causative agent of potato leafroll disease, which occurs worldwide and is responsible for large crop losses. In 1916, Quanjer *et al.* described the infectious nature of this disease (see Harrison, 1984). The virus particles were purified about fifty years later (Peters, 1967). The whole sequence of the PLRV genome was published in 1989 (Mayo *et al*; van der Wilk *et al*) and 1990 (Keese *et al*).

PLRV has a narrow range of hosts, most of which belongs to the family *Solanaceae*. Important diagnostic hosts include *Datura stramonium*, *Physalis floridana* and *Solanum tuberosum* ssp. *tuberosum* (potato). Infected leaves from the latter two species are often the source for virus purification. In Western Europe, PLRV seems to be confined to the potato (Harrison, 1984).

Unlike other viruses of the potato (potato virus X and potato virus Y), which can be transmitted mechanically, PLRV is obligately transmitted by aphids, *Myzus persicae* being the most important vector. Transmission is circulative, persistent and non-propagative. Ingested virus particles are transported from the digestive tube into the haemolymph (acquisition), where the acquired virus particles are retained in an infective form, without replication, for a long period of time. Virions enter the cells of the accessory salivary glands by diffusion, and are eventually excreted into the plant phloem sap during feeding. Two barriers are present in this circulative path: one is the epithelial cell lining of the intestine, the other is the basal lamina of the accessory salivary glands,

which seems to be more selective because only virions of transmissible strains can pass through (Gildow, 1993a, b).

The primary infection usually leads to pallor, or sometimes reddening of the tip leaves, which may become rolled and assume an erect habit. Plants grown from infected tubes show secondary symptoms, i.e. stunting of the shoots, upward rolling of leaflets and marginal necrosis on lower leaves (Harrison, 1984).

In plants, PLRV is confined to the phloem tissue, mainly in the cytoplasm of phloem parenchyma and companion cells as unstructured aggregates, or in the vacuoles as crystalline aggregates. Cell walls may become thickened, and callose accumulates in some sieve tubes, which impairs phloem transportation and is the basis of various staining tests.

Strains of PLRV differ in the severity of symptoms they cause in plants, and the ease of transmission by *Myzus persicae*; but they cannot be distinguished by immunological methods.

Under the electron microscope, PLRV virions appear to be non-enveloped, isometric particles with a diameter of 24 nm. Viral particles contain a single-stranded positive sense genomic RNA of some 5.8 kb in length.

On the basis of symptoms, cytopathological effects, aphid vector specificity, serological relationships, and more recently the genome organization, PLRV is classified as a member of the luteovirus family, which displays a mosaic of non-canonical strategies in gene expression, as is described below.

1.2 Gene Expression of Luteoviruses

1.2.1 Classification of Luteoviruses

Luteoviruses were formally recognized as a group in 1976 (Fenner), though the "yellowing" diseases caused by member viruses had been known for a long time. Currently luteoviruses are classified as a family, comprising three genera (Table 1-1). The genera *Luteovirus* and *Polerovirus* approximately correspond to former subgroup I and II of the luteovirus group, and are more closely related to each other than they are to the third genus, *Enamovirus* (Mayo, personal communication). In this thesis, the term 'luteovirus' refers to the family, rather than the genus, unless denoted otherwise.

1.2.2 Luteovirus genome structure

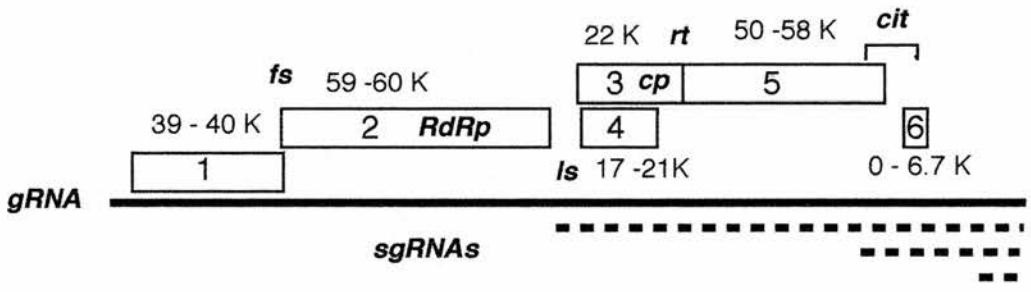
The three genera of the luteovirus family have similar yet distinct genome organizations (Figure 1-1). All members have a monopartite single-stranded positive-sense genomic RNA, encoding 5-7 genes, which are arranged into 2 blocks separated by a small intergenic region. Replicative proteins are encoded in and translated from the 5' proximal half of the genomic RNA, while capsid and movement proteins are encoded in the 3' proximal half and are translated from a subgenomic RNA. The putative RNA-dependent RNA polymerase (RdRp) is expressed as a fusion protein to P1 via -1 frameshift. P5 is translated by readthrough of the stop codon at the end of P3 (coat protein) gene. Apart from the enamovirus, viruses in the other two genera encode a movement protein P4 within the P3 coding sequence, but in a different reading frame. No RNA molecules have poly(A) sequence at the 3' terminus.

Table 1-1 Classification of Luteoviruses

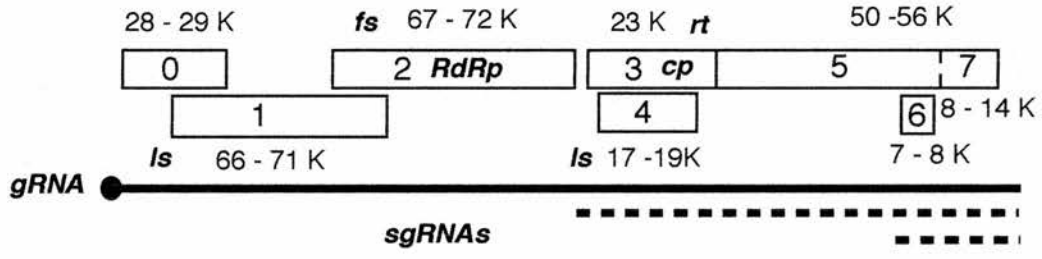
Genus	Species in the genus
Luteovirus	barley yellow dwarf virus – PAV **type species**
	barley yellow dwarf virus – MAV
Polerovirus	potato leafroll virus **type species**
	beet mild yellowing virus
	beet western yellows virus
	cereal yellow dwarf virus – RPV (formerly BYDV-RPV)
	cucurbit aphid-borne yellows virus
Enamovirus	pea enation mosaic virus **type species**
(unassigned)	barley yellow dwarf virus – GPV
	barley yellow dwarf virus – SGV
	bean leafroll virus
	carrot red leaf virus
	cereal yellow dwarf virus - RMV (formerly BYDV-RMV)
	chickpea stunt associated virus
	groundnut rosette assistor virus
	Indonesian soybean dwarf virus
	sweet potato leaf speckling virus
	soybean dwarf virus
	tobacco necrotic dwarf virus

Viruses in the genus *Luteovirus* do not contain ORF0, but contain a long sequence (600-867 n.t.) downstream of ORF5, which has been demonstrated to be important in the cap-independent translation of viral genes. ORF1 is much shorter than that of viruses in

Genus *Luteovirus*



Genus *Polerovirus*



Genus *Enamovirus*

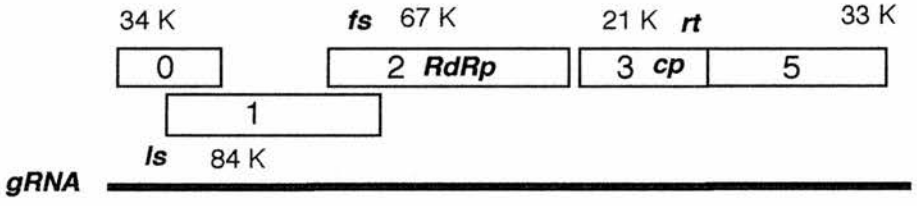


Figure 1-1 Genome organization of the luteovirus family.
Boxes indicate ORFs, numbered as Martin *et al.* (1990). ***cp,*** coat protein; ***RdRp,*** RNA-dependent RNA polymerase. Known or proposed sites of frameshifting (***fs***), leaky scanning (***ls***), readthrough (***rt***) , and the sequence required for cap-independent translation (***cit***) are indicated. Subgenomic RNAs (***sgRNAs***) are shown by dashed lines, while genomic RNAs (***gRNAs***) are shown by solid lines.

other genera, and the overlap between ORF1 and ORF2 is only 7-13 n.t. long. The RdRp is believed to be homologous to that of the "carmo-like" viruses (Koonin and Dolja, 1993). The central intergenic region is about 100 n.t. long. No VPg is attached to the genomic RNA.

Poleroviruses have a VPg covalently linked to the 5' end of the genome RNA. There is an open reading frame (ORF0) upstream of ORF1. The P1 protein is 66-71 kDa, and the overlap between ORF1 and ORF2 is much longer. The intergenic region is about 200 n.t. in length. Recently, two small ORFs were found in the 3' end region of the PLRV genomic RNA, which are translated from a smaller subgenomic RNA (Ashoub *et al.*, 1998). It is uncertain whether this is true for other viruses in this genus.

Pea enation mosaic virus -1 (PEMV-1) is the only member in the genus *Enamovirus*. Like poleroviruses, it contains a VPg and ORF0. The polymerase (RdRp) of both poleroviruses and enamoviruses is similar to that of "sobemo-like" viruses (Koonin and Dolja, 1993). However, PEMV-1 does not encode ORF4 (movement protein) and appears to be deficient in cell-to-cell movement when inoculated alone, though it can replicate in protoplasts. Naturally, PEMV-1 is always associated with PEMV-2, an umbravirus, and consequently can be transmitted both mechanically and by aphids, and can replicate and accumulate in mesophyll tissue (Demler *et al.*, 1994). The ORF5 of PEMV-1 is smaller than that of other luteoviruses.

1.2.3 Gene functions

In the past twenty years, developments in nucleotide sequencing techniques have allowed the determination of the complete sequences of representatives of the majority of virus genera and families. For many viruses, we now have established the number,

size, sequences, and genomic locations of viral gene products. The more demanding task remaining is the unravelling of gene functions. Reverse Genetics allows the manipulation of the genome of RNA viruses at the cDNA stage, and thus allows several approaches in tackling the question: (1) *in vitro* transcription/ translation system; (2) transfection of protoplasts with RNA from *in vitro* transcription as well as purified viral RNA; (3) agroinfection, in which plant tissues are injected with agrobacteria carrying a DNA copy of the viral genome in a Ti plasmid. Current knowledge of the functions of luteoviral genes is described here.

1.2.3.1 Replicase

All luteoviruses encode a putative RNA-dependent RNA polymerase (RdRp) in ORF2. The signature motif, GxxxTxxxN(x_{25-40})GDD, is present about 100-200 amino acids upstream of the C-terminus of P2. All the positive-strand RNA viruses encode a RdRp containing this conserved motif.

Sequence alignment studies have shown that members of former subgroup I luteoviruses, SDV and BYDV, show a high homology in their RdRps with carmoviruses and tombusviruses; while the RdRps of subgroup II luteoviruses (PLRV and BWYV) and PEMV-1 are significantly similar to that of southern bean mosaic virus (SBMV), the type member of the sobemoviruses.

Most positive-strand RNA viruses with a genome larger than 6 kb are thought to encode an RNA helicase, which, it is proposed, would unwind RNA duplex as part of the viral replication apparatus (Gorbalenya and Koonin, 1989). Habili and Symons (1989) found helicase motifs in ORF 1 and 2 of luteoviruses; but the existence of a helicase was questioned by Koonin and Dolja (1993).

1.2.3.2 VPg

Mayo *et al.* (1982) found that a VPg protein of about 7 kDa was covalently linked to the genomic RNA of PLRV. The VPg of CYDV-RPV is 17 kDa (Murphy *et al.*, 1989). For a time, it was proposed that ORF4 encoded the VPg on the basis of its size. This is not consistent, however, with other observations. PEMV has a VPg but lacks an equivalent of ORF4. In the picornaviruses, the VPg appears to play an essential role in the initiation of viral RNA synthesis (Rueckert, 1996). On the other hand, ORF1 and ORF2 have been demonstrated to be the only two viral genes necessary for BWYV RNA replication. The VPg is not likely to be encoded by ORF4 because deletion of this gene did not block viral replication (Reutenauer, 1993).

Recently, The N-terminus of VPg of PLRV has been mapped to position 400 of the ORF1 product (van der Wilk, 1997). Although the C-terminus of VPg has not been determined precisely, motifs essential for viral replication are apparently arranged in an unusual order in the genome in comparison with other positive-strand RNA viruses: protease-VPg-polymerase.

1.2.3.3 Protease

The original work of Gorbalenya *et al.*(1986) found that the picornavirus 3C proteases shared sequence homology with chymotrypsin-like serine proteases in regions surrounding the putative residues of catalytic nucleophile. Subsequently, further homologies were identified between the cysteine proteases of a number of positive-strand RNA viruses and the cellular serine proteases (Bazan and Fletterick, 1988;

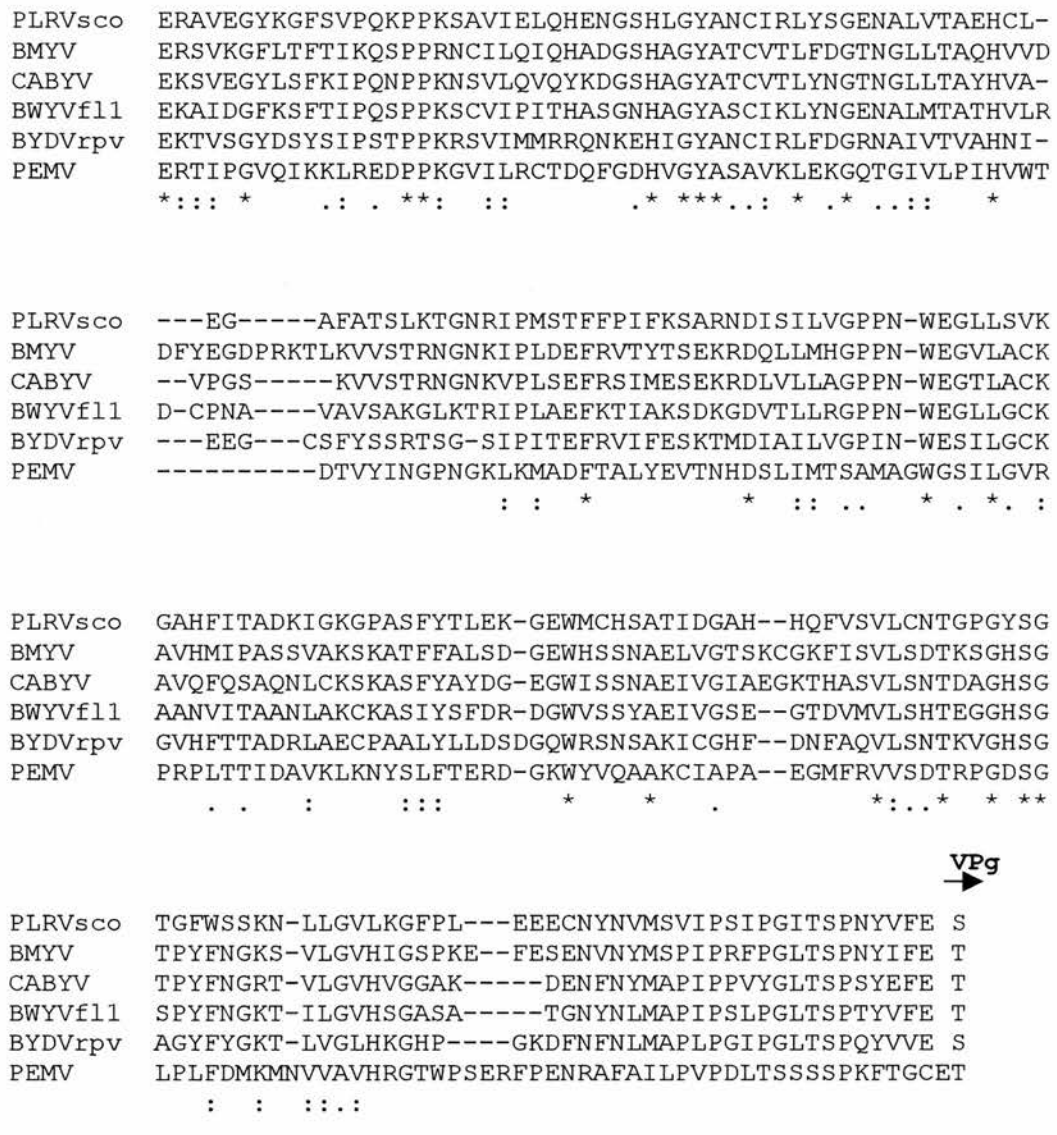


Figure 1-2 Alignment of luteovirus P1 amino acid sequences comprising the putative protease domain

PLRVsco – potato leafroll virus, Scottish isolate; BMYV – beet mild yellowing virus; CABYV – cucurbit aphid-borne yellows virus; BWYVfl1 – beet western yellows virus, isolate FL-1; BYDVrpv - cereal yellow dwarf virus, strain RPV; and PEMV – pea enation mosaic virus. The proposed active site residues are shown in reverse font, the sites of mutations are shown together with the protease/VPg cleavage site.

Gorbalenya *et al.*, 1989). The consensus sequence of the catalytic triad identified by Gorbalenya is $Hx_{\sim 25}[D/E]x_{70\text{-}80}T[R/K]xGxSG$.

In the case of the luteoviruses, Demler and de Zoeten (1991) found the protease motif T[R/ K]xGxSG in the ORF1 of PEMV RNA-1. Koonin and Dolja (1993), and Miller *et al.* (1995) predicted the residues that formed the catalytic triad of a putative protease domain in the ORF1 of several subgroup II luteoviruses (poleroviruses). Figure 1-2 shows the alignment of part of the ORF1 sequences of poleroviruses and PEMV-1. Not every virus contains the arginine or lysine residue four amino acids upstream of the catalytic serine. Both Koonin and Miller predicted that His-255, Asp-286, and Ser-354 form the catalytic triad. Gly-355 might be part of the oxyanion hole that helps to stabilize the transition state during the nucleophilic attack by Ser-354.

Experimental evidence demonstrating the proteolytic activity of this domain is lacking. *In vitro* translation of purified PLRV RNA in wheat germ extracts yielded two major products of 28 kDa and 70 kDa, which had the same sizes as expected products of ORF0 and ORF1 (Mayo *et al.*, 1989). A minor product of 125 kDa was suggested to be the ORF1/ORF2 transframe product. Several minor products of intermediate sizes also appeared, which might be the result of premature termination, internal initiation, or proteolytic cleavage. Even if those minor products arise from proteolytic processing, evidence is lacking to show that the protease domain in ORF1 carries out the cleavages.

1.2.3.4 Movement protein

P4 is the viral movement protein responsible for cell-to-cell spread. Like the tobacco mosaic virus (TMV) 30-kDa movement protein, the 17-kDa P4 protein of PLRV binds to single-stranded nucleic acids in a co-operative, non-specific manner, and is

phosphorylated *in planta* (Tacke *et al.*, 1991). The nucleic acid-binding domain was shown to reside in the basic C-terminal portion. The acidic N-terminal region contains two to three amphipathic α-helices separated by a proline residue. The amphipathic nature of the helices allows the protein to interact with each other to form homo- or hetero-oligomers. PLRV P4 protein has been shown to exist in infected or transgenic plants mainly as a homodimer (Tacke *et al.*, 1993). Introducing the helix-breaking residue proline, or altering the amphipathic feature by substitution of hydrophilic residues for hydrophobic ones, abolished or significantly reduced the dimerization ability.

In infected plants, P4 is expressed with high efficiency and accumulates to high levels (Tacke *et al.*, 1990). Subcellular localization studies revealed that P4 was mainly associated with fractions enriched for membrane structures. Based on these features, Tacke *et al.* (1990) have suggested that P4 protein is the phloem-specific movement protein of luteoviruses postulated by Hull (1989).

Interestingly, PEMV-1 does not encode an ORF4 and is deficient in cell-to-cell movement when it infects plants alone. In association with PEMV-2 (an umbravirus), PEMV-1 infects plants systemically and even accumulates in mesophyll tissue, which poleroviruses do not infect. PEMV RNA-2 may, therefore, encode a different movement protein.

1.2.3.5 Capsid protein

The coat proteins are encoded by ORF3, as shown by the association of antibodies specific to virus particles with the P3 proteins expressed *in vitro* (Veidt *et al.*, 1988), expressed in *E. coli* (Miller *et al.*, 1988), in Hela cells (Mayo *et al.*, 1989), or in insect

cells (Lamb *et al.*, 1997). Sequences of luteoviral coat proteins are highly conserved and surface determinants can be mapped to particular regions by secondary structure modeling (Mayo and Zielger-Graff, 1996).

Particle integrity is apparently necessary for aphid transmission. Four monoclonal antibodies specific for readily transmissible PLRV particles have been shown to react with P3. It appears that, in addition to protecting genomic RNA by encapsidation, the coat proteins might be involved in vector specificity as well.

Although coat protein is necessary for BWYV to infect the whole plant, ORF3 is not required for RNA replication in protoplasts (Reutenauer *et al.*, 1993).

1.2.3.6 ORF5

P5 is expressed as a fusion protein with the coat protein by readthrough of the stop codon of ORF3. The readthrough protein has been detected in purified virions, infected protoplasts and plants (Bahner *et al.*, 1990; Reutenauer *et al*, 1993). The species in purified virions is smaller, some 53 kDa, presumably being truncated at the 3 terminus of ORF5.

ORF5 can be divided into three regions. (1) The short region immediately following the stop codon of ORF3 is rich in proline. This proline hinge separates the coat protein domain from the readthrough domain, so that the former assembles into the capsid and the latter protrudes from the virus particle. (2) Downstream of the proline hinge is a region conserved in all luteoviruses. (3) The C-terminal region is variable among luteoviruses, but within this region there is a motif of about 45 amino acids, showing high homology among PLRV, BWYV and CABYV. This motif might be involved in

vector specificity, as these three viruses are all transmitted by *Myzus persicae*. The luteovirus homology domain might be involved in the recognition events at the hindgut (or intestine), which shows little selectivity among luteoviruses; whilst the PLRV/ BWYV/ CABYV homology domain might contain signals required for the more specific uptake at the basal plasmalemma of the salivary gland of *Myzus persicae* (reviewed by Mayo and Zielger-Graff, 1996).

The involvement of the readthrough domain (RTD) in aphid transmission is supported by many observations. Two PLRV isolates poorly transmitted by aphids were found to lack a conformation-sensitive epitope present in readily transmissible isolates (Massalski and Harrison, 1987). Either or both of the two amino acid changes in the C-terminal part of the readthrough protein are responsible for the poor transmission and the loss of the epitope (Jolly and Mayo, 1994). Experiments using BWYV containing mutations in the readthrough domain revealed that these mutant viruses could not be transmitted by aphids fed on infected plants or protoplasts (Brault *et al.*, 1995). It has been shown that the primary endosymbiotic bacteria of *Myzus persicae* produce a 60 kDa GroEL-like protein (MpB groE) and release this protein into the haemolymph. This protein readily binds to PLRV particles and possibly helps the latter to resist the hostile environment in the vector (van den Heuvel *et al.*, 1994; Hogenhout *et al.*, 1998). van den Heuvel *et al.*(1997) has reported that the readthrough domain, which is protruding from the surface of virions, contains the determinants for GroEL-binding.

ORF5 is not required for the replication of BWYV, PLRV, and BYDV-PAV in protoplasts (Reutenauer *et al.*, 1993; Miller *et al.*, 1995; Mayo and Ziegler-Graff, 1996), though an earlier report suggested otherwise (Young *et al.*, 1991).

1.2.3.7 ORF0

ORF0 is present in the polero- and enamoviruses, but shows little homology among different viruses. Its function remains obscure. BWYV RNA transcripts with the start codon of ORF0 having been removed can replicate in protoplasts, and the progeny RNAs appear to be encapsidated (Veidt *et al.*, 1992). Tobacco plants infected with a BWYV construct containing N-terminal deletion of ORF0 develop wild-type symptoms and produce aphid-transmissible virions. Therefore, ORF0 of BWYV is not required for viral RNA replication, symptom induction and aphid transmission (Mayo and Ziegler-Graff, 1996).

In the case of PLRV, sequence analysis has predicted that the N-terminal region of P0 is very hydrophobic and has some characteristics of membrane-associated proteins (Mayo *et al.*, 1989). Transgenic potato plants expressing only ORF0 show classic symptoms of potato leafroll disease. When ORF0 was removed, protoplast transfection or leaf-disk agroinfection was not affected (Mayo and Ziegler-Graff, 1996).

1.2.3.8 Small ORFs

Small ORFs near the 3' end of genomic RNA of BYDV-PAV and MAV have been identified for many years (Kelly *et al.*, 1994; Miller *et al.*, 1988; Ueng *et al.*, 1992). ORF6 encodes a 4.3-6.7 kDa protein downstream and in a different reading frame of ORF5, and is presumably translated from a smaller subgenomic RNA than that for the expression of P3 and P3/P5 readthrough proteins. A frameshift mutation in the ORF6 of BYDV-PAV rendered the virus non-infectious (Young *et al.*, 1991).

Recently Ashoub *et al.* (1998) reported that two small proteins (ORF6 and ORF7) were translated from an 800 n.t. subgenomic RNA of PLRV. The ORF6 was in a different open reading frame with ORF3, but ORF7 encoded a 14 kDa protein C-coterminal with P5. *In vivo* transient expression showed that ORF6 was translated about 16-fold more efficiently than ORF7 in potato protoplasts.

The functions of these small ORFs are still unknown.

1.2.4 Strategies of gene expression

Luteoviruses utilize a wide range of noncanonical strategies to regulate their gene expression transcriptionally, translationally and post-translationally. The amazing usage of their small genome is exceptional even among RNA viruses.

1.2.4.1 Subgenomic RNA

Like many positive-sense RNA viruses, the 3'-proximal gene cluster of the luteoviruses is translated from subgenomic RNAs (sgRNAs), which are not encapsidated in virions. Tacke *et al.* (1990) first reported that the 5' end of the major PLRV sgRNA to be 40 n.t. upstream of the start codon of coat protein. Miller and Mayo (1991), however, mapped it to be 172 nucleotides further upstream, within the 3' end coding region of the RdRp. The 5' end of subgenomic RNA determined by Mayo is identical to the 5' end of the genomic RNA: 5'-ACAAAAGA-3'. The subgenomic RNAs of BWYV and CABYV also start with 5'-ACAAAAGA-3' (Guilley *et al.*, 1994; Mayo and Ziegler-Graff, 1996;). In addition, two A-U rich regions were identified within the intergenic regions of luteoviruses (Mayo *et al.*, 1989; Vincent *et al*, 1991).

The above observations are consistent with the mechanism of subgenomic RNA synthesis, originally demonstrated in bromoviruses (Marsh *et al.*, 1988; Miller *et al.*, 1985). In infected cells, the viral replicase first synthesizes the full-length (-) strand from the (+) template. It would then bind the (-) strand RNA either at the 3' end or internally at a subgenomic promoter to initiate (+) strand synthesis. The sequence 5'-UUUUGU-3' in the (-) strand might act as a signal for replicase recognition.

Recently, Ashoub *et al.* (1998) reported a second 0.8 kb subgenomic RNA of PLRV, whose 5' end was mapped to n.t. 5190. The 5' end sequence is 5'-AGCGGCGG-3', different from that of the genomic RNA (5'-ACAAAAGA-3'), or that of the major subgenomic RNA (5'-GUUUACCU-3').

BYDV-PAV has three subgenomic RNAs. The 5' end of the largest subgenomic RNA was reported to start at either n.t. 2769 (Dinesh-Kumar 1992) or n.t. 2670 (Kelly, 1994). If it starts at n.t. 2670, then all the three subgenomic RNAs will show similar 5' end: 5'-AGUGAAG-3'.

1.2.4.2 Internal initiation of translation

Initiation of protein synthesis on eukaryotic mRNAs is thought to follow the scanning model. The 40S ribosomal subunit complex binds at the 5'end of the mRNA and scans until it reaches the first AUG, where the 60S subunit binds to form the 80S ribosome and protein synthesis begins (Kozak, 1989). If the first AUG is not in an optimal context, then a portion of the scanning 40S ribosomes may bypass that AUG and initiate at a downstream AUG if it is in a more favoured context. The optimal context around the start codon identified by Kozak is CCA(or G)CC**AUG**G (the start codon AUG is in bold). In luteoviruses, such leaky scanning can occur at two sites.

ORF0 and ORF1

In polero- and enamoviruses, the start codon of ORF0 is flanked by less optimal bases than that of ORF1 (Miller *et al.*, 1995). Products of expected sizes from both ORFs were observed after *in vitro* translation of the genomic RNA of PLRV (Mayo *et al.*, 1989) and BWYV (Veidt *et al.*, 1992).

ORF3 and ORF4

Another example is ORF3 and ORF4, both of which are expressed from a subgenomic RNA and their start codons are the first two AUGs. In most luteoviruses, the ORF4 start codon is in a better context than the ORF3 AUG. Using GUS reporter gene, Tacke *et al.* (1990) found that initiation at the P4 start codon was seven-fold more efficient than at the CP start codon.

An extra feature of BYDV-PAV is that the start codon of coat protein is embedded in a stem-loop structure (Dinesh-Kumar *et al.*, 1992). Disruption of this secondary structure enhanced expression of coat protein and movement protein proportionally. Furthermore, changing the context of the movement protein start codon reduced initiation at the coat protein start codon, which is 43 bases upstream. Dinesh-Kumar and Miller (1993) proposed a ribosomal pausing model in which the leaky 40S ribosomal complex pauses at the AUG of movement protein to associate with the 60S subunit. The 80S initiation complex can melt the upstream stem-loop so that the trailing 40S ribosome is stalled in a place near the AUG of coat protein and has a better chance to initiate translation of coat protein. Suboptimal context at AUG of ORF4 means less pausing of ribosomes and less expression of coat protein.

1.2.4.3 Ribosomal frameshifting

All luteoviruses express their polymerase as a fusion protein with the P1 protein via a -1 frameshift during translation. In the well accepted simultaneous slippery model, the frameshift occurs at a 'shifty' heptanucleotide (or so-called 'slippery' site), where the tRNAs in the peptidyl and aminoacyl sites on the ribosome slide back simultaneously by one base along the mRNA with two of the three bases of the tRNA anticodon still base-paired to the mRNA. The ribosome then continues translation in the -1 frame relative to the first one. Downstream pseudoknot (ten Dam *et al.,* 1990) or stem-loop (Parkin *et al.,* 1993) RNA structures have been shown to be important for frameshifting: these structures can stall the ribosome at the shifty site, allowing more time for shifting.

In BYDV-PAV, the overlap of ORF1 and ORF2 is only 13 nucleotides long and the position of frameshift has been mapped to the two codons immediately before the UAG of ORF1 (Di *et al.,* 1993). The stop codon of ORF1 was necessary for frameshifting *in vivo* (Brault and Miller, 1992) but was not when the full-length BYDV-PAV RNA was translated *in vitro* (Dinesh-Kumar, 1993). On the contrary, the stem-loop upstream of the shifty site is required *in vitro* but not *in vivo* (Brault and Miller, 1992). The region downstream of the shifty site can exist in two conformations: a long stem-loop or a pseudoknot with two stem-loops base-paired between the loops ("kissing stem-loops"). The long stem-loop is favoured because it is also conserved in soybean dwarf virus (SDV). Disruption of the putative secondary structure abolished frameshifting. Another striking feature of BYDV-PAV is that the 3'-terminal 600 nucleotides of the genomic RNA are required for frameshifting *in vitro* (Miller *et al.,* 1995; Mayo and Ziegler-Graff, 1996).

BWYV and PLRV have much longer overlap between ORF1 and ORF2. Prüfer *et al.* (1992) found that a 214 n.t. sequence within the ORF1/ORF2 overlap of PLRV was sufficient to direct frameshifting, which contained the shifty heptanucleotide UUUAAAU. Using a GUS reporter gene, frameshifting was detected to occur at a rate of about 1%. Prüfer *et al.* (1992) identified a downstream stem-loop that influenced the efficiency of frameshifting, but Kujawa *et al.* (1993) found that a pseudoknot was essential. Deletion that disrupted the stem-loop but maintained the pseudoknot structure increased the rate of frameshifting a bit rather than reduced it. The situation of BWYV seems complex in that the frameshifting might occur at more than one place in the overlapping region (Mayo and Ziegler-Graff, 1996).

1.2.4.4 Stop Codon Suppression

When the ribosome reaches the stop codon that separates two in-frame ORFs, occasionally a suppressor tRNA can bind to the stop codon, incorporate an amino acid in the nascent peptide to suppress the stop codon, and result in the translation of the downstream ORF as a fusion protein with the first ORF.

In luteoviruses, ORF5 is translated as a P3/P5 fusion protein by readthrough of the amber codon of ORF3. The sequence around the ORF3 stop codon of luteoviruses is highly conserved: AAA**UAG**GUAGAC. A series of 5-16 tandem repeats of CCNNNN are present 12-21 n.t. downstream of the stop codon, and are likely to be important in promoting the readthrough. No conserved secondary structure or pseudoknot is found in this region. In BYDV-PAV, another sequence 700 n.t. downstream of the ORF3 stop codon is required for efficient readthrough *in vitro* and in protoplast (Brown *et al.*, 1996). This distant element might exist in BWYV and PEMV as well (Demler *et al.*, 1997; Bruyere *et al.*, 1997).

When tobacco or potato protoplasts were electroporated with a construct in which the region 18 n.t. upstream and 21 n.t. downstream of the stop codon of PLRV ORF3 were inserted between CaMV 35S promoter and GUS gene, the readthrough efficiency was about 1% (Tacke *et al.*, 1990). *In vitro* suppression of ORF3 UAG in BYDV-PAV occurs at a rate of 7-15%, depending on the salt concentration within the reaction mixture (Dinesh-Kumar *et al.*, 1992).

1.2.4.5 Cap-independent translation

In BYDV-PAV, a 109 n.t. cap-independent translation signal is present between ORF5 and ORF6. This signal enhances translation from uncapped viral genomic or subgenomic RNA 30-100 fold in wheat germ extracts and is called 3' translation enhancer (3' TE). 3' TE mediated cap-independent translation also requires the 5' UTR of BYDV-PAV genomic and subgenomic RNA 1. This led Wang *et al.* (1997) to propose a model in which the 3' TE interacts with the 5' UTR and mimics a 5' cap. Uncapped transcripts containing the 3' TE are translated with similar efficiency to capped transcripts without the 3' TE; deletion of the 3' TE can be compensated by capping.

In plant cells, the stimulation by the 3' TE needs additional 3' end sequence from the BYDV-PAV, which mimics a poly (A) tail not required in wheat germ extract.
On the other hand, 3' TE-containing sequences, like subgenomic RNA 2 of BYDV-PAV, act as an inhibitor to translation from genomic RNA and subgenomic RNA 1, when added in excess *in trans*. The inhibition is stronger for genomic RNA than for subgenomic RNA 1, suggesting a role of subgenomic RNA in the switch from early to late translation (Wang *et al.*, 1999).

1.2.4.6 Proteolytic processing

ORF1 and ORF2 play an essential role in the replication of luteoviruses. By sequence alignment, multiple functional domains have been identified within their coding sequences: protease, helicase, and RNA-dependent RNA polymerase. In PLRV, VPg has been mapped to the C-proximal region of ORF1 (van der Wilk *et al.*, 1997), which apparently must be released by proteolytic cleavage. Recently, Prüfer *et al.* (1999) showed that P1 was processed in plants infected with PLRV. A 25 kDa product (P1-C25) represented the C-terminal of P1, with its N-terminus being either the VPg domain or closely adjacent to it. The P1-C25 was not detected when ORF1 was translated in cell-free system.

In all known RNA viruses, the processing of replicative precursors is mediated by viral proteases. A chymotrypsin-like protease domain is identified in poleroviruses and PEMV-1, which also have VPg attached to their genomic RNAs. This domain is a good candidate for the protease that is responsible for the processing of VPg, though experimental evidence of the proteolytic activity of this domain is absent.

1.3 Polyprotein Processing in RNA Viruses

1.3.1 Overview

All RNA viruses need to perform some basic functions: (1) synthesis of viral polymerase or at least a component of it; (2) replication of viral genomes; (3) synthesis of structural proteins and assembly of virions. Obviously, the synthesis of viral proteins plays a key role in viral replication. Since all viruses utilize the translation machinery of their host cells, one constraint is imposed on viruses, in that the eukaryotic translation machinery is organized to translate only one protein from one mRNA. To solve this problem, some viruses synthesize a subgenomic RNA to translate downstream proteins, but this strategy also has limitations. Many RNA viruses have evolved another strategy, in which a large precursor protein encompassing several distinct domains, a polyprotein, is synthesized from one mRNA and then proteolytically processed to generate intermediate and mature proteins.

With the polyprotein strategy, every mature protein is generated in an equimolar fashion. By including all proteins of a multi-component complex in a polyprotein, it is easier for individual components to assemble after being synthesized. That is required for the different subunits of picornaviral capsid. Some proteins are, however, needed in different amounts. Much more VPg is required, for example, than the replicase. To address these problems, many viruses use a combination of various strategies, such as polyprotein processing plus ribosomal frameshift and/or stop codon readthrough.

Polyprotein processing is essential for many viruses. The cleavages are highly specific and may exhibit elegant temporal and spatial control. To achieve this, most (if not all)

viruses encode their own proteases to process their polyproteins, especially those polyproteins involved in replication of viral genome. Precursors of viral envelope proteins, however, usually are translocated to the endoplasmic reticulum (ER) and processed by cellular enzymes (e.g. signalase), which seem to have enough specificity to cleave only at the correct sites.

Proteolytic processing can take place by two mechanisms: an intramolecular (*cis*) reaction or an intermolecular (*trans*) reaction. In a *cis* reaction, the protease domain and the cleavage site reside in the same molecule, so zero-order kinetics is followed, which means the reaction is insensitive to dilution. *Cis* cleavages are often rapid and occur co-translationally. A *trans* reaction, in which the protease domain and cleavage site are present in two separate molecules, is usually slower, following second-order kinetics, sensitive to dilution, and generally more sensitive to protease inhibitors and sequence variations at cleavage sites. A *cis* reaction is also referred to as autocatalytic or monomolecular, while a *trans* reaction as bimolecular.

Viral polyproteins were first reported in 1968 (Summers and Maizel). In the past 30 years, this strategy of gene expression has been found in many animal and plant viruses. Five well-characterized families of RNA viruses are chosen here as representatives, with a focus on their polyprotein organizations and cleavage cascades. In next section, more details are discussed on the structure, catalytic mechanism, cleavage specificity, and regulation of activity of viral proteases.

1.3.2 *Picornaviridae*

Picornaviruses are one of the largest and most important families of human and animal pathogens. They are currently divided into 9 genera: *Rhinovirus*, *Enterovirus*,

Aphthovirus, *Cardiovirus*, *Hepatovirus*, *Parechovirus*, *Erbovirus*, *Kobuvirus* and *Teschovirus* (Update to the 7th ICTV Report, in the Picornavirus Homepage at the Institute for Animal Health, U.K.). They have small (28-30 mm in diameter), non-enveloped, icosahedral virions with a monopartite, linear, positive-stranded genomic RNA of 7.2-8.4 kb.

Picornaviruses have similar genome organizations (Figure 1-3). The genomic RNA is polyadenylated at the 3' terminus and has a small virus-encoded protein (virion protein, genome; VPg) covalently attached to its 5' terminus. All RNAs encode a single large polyprotein of 200-300 kDa, flanked by untranslated regions (UTRs). The 5' UTRs are 600-1200 n.t. long and rich in secondary structures. Unlike 5' capped mRNA, translation is initiated via an internal ribosome entry site (IRES) in the 5' UTRs. Aphthoviruses and cardioviruses contain a poly(C) tract upstream of the IRES.

The polyprotein is cleaved during and after translation by viral proteases to release all the structural and non-structural proteins of the virus (Jacobson and Baltimore, 1986). In the L-4-3-4 nomenclature system, the leader (L) proteins of cardioviruses are some 7 kDa, while the aphthoviruses have two nested L proteins (23 kDa and 16 kDa), translated from two in-frame start codons. The entero-, rhino- and hepatoviruses do not have L proteins. The P1 region encodes the capsid proteins, VP1 (1D), VP2 (1B), VP3 (1C) and VP4 (1A). Non-structural proteins are present in P2 and P3 regions. The 2A protein mediates the co-translational separation of structural and non-structural proteins (except in the hepatoviruses). The 2B is a host-range determinant (Yin and Lomax, 1983), whereas 2C is involved in the RNA synthesis since the guanidine resistance marker maps to 2C. The VPg is encoded by the 3B region of the genome (Pallansch *et al.*, 1980). Protein 3C is the major viral protease, responsible for most of the cleavages of the viral polyprotein, either as a free protein, or as part of the 3CD protein. The 3D is

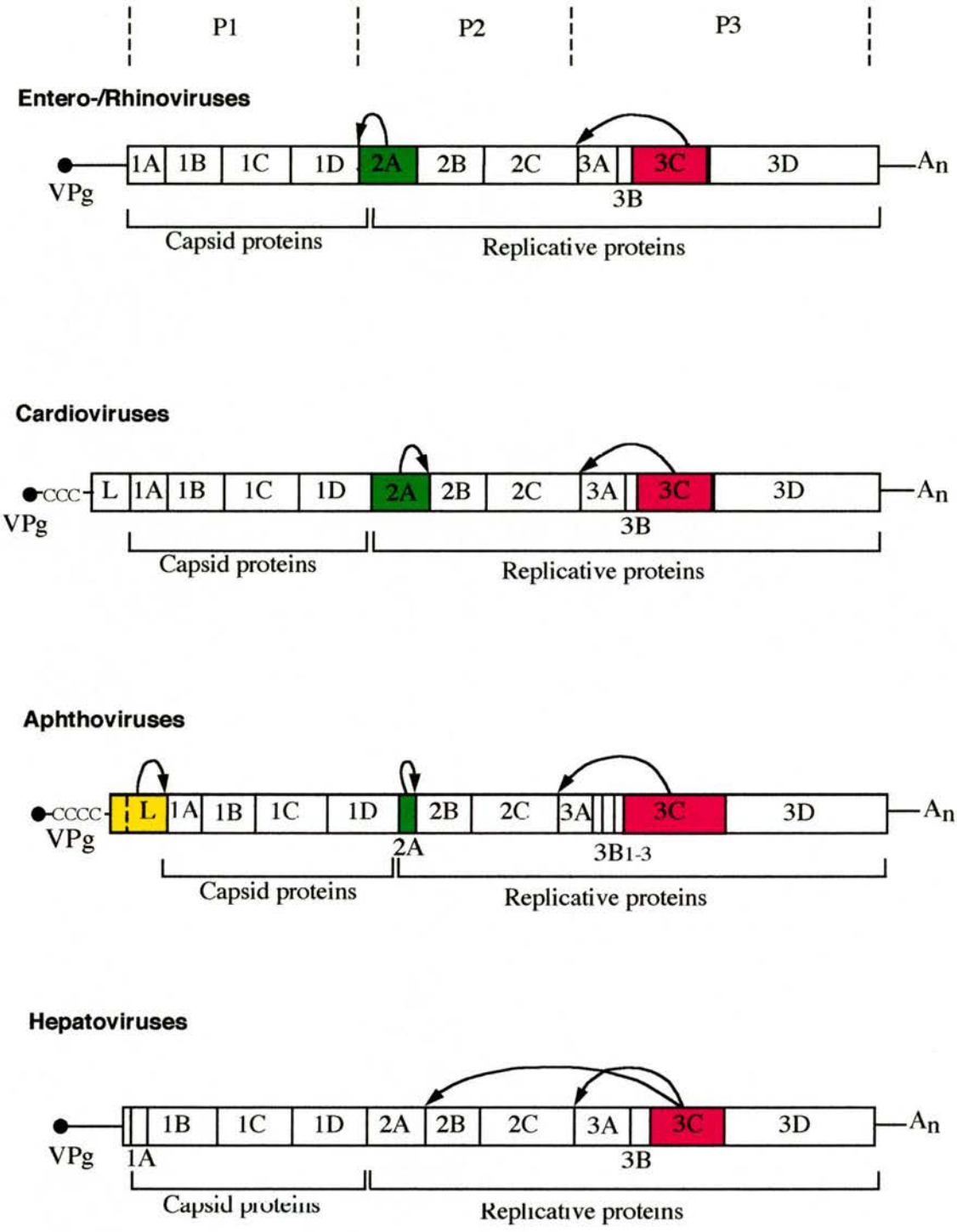


Figure 1-3 Genome organization of picornaviruses
The proteolytic enzymes encoded by the polyproteins are shown by colours. Green indicates a 2A protease, yellow an L protease and pink a 3C protease. Arrows indicate the cleavages that produce P1, P2 and P3 intermediate precursors.

the polymerase, capable of elongating nascent RNA chains from an RNA template (Flanegan and Baltimore, 1977).

The processing of the picornaviral polyprotein can be divided into three steps: separation of the capsid precursor from the growing nascent polypeptide chain (primary cleavage); processing of the structural and non-structural precursor proteins (secondary cleavages), and maturation cleavage of the capsid (Krausslich and Wimmer, 1988).

Primary Cleavage

Primary cleavages occur rapidly and efficiently, when the polyprotein is *statu nascendi* (co-translationally). No precursor polypeptides spanning the cleavage site are detected under normal conditions.

Different genera use distinct mechanisms to accomplish the primary cleavage between the capsid and replicative proteins (Figure 1-3). In entero- and rhinoviruses, the 2A protease ($2A^{pro}$), a cysteine protease, cleaves at its own N-terminus in a *cis* reaction as soon as the ribosome reaches the middle of the genome. The primary cleavage in cardio- and aphthoviruses is also mediated by 2A, but the mechanism seems quite novel. The cleavage site is between 2A and 2B, instead of P1 and 2A in entero- and rhinoviruses. The 2A sequences of EMCV and TMEV show no similarity to either those of entero- and rhinovirus, or any other known protease motifs. However, the C-terminal region of cardiovirus 2A is highly similar to the 2A of FMDV, which is only 18 amino acid long. The tetrapeptide spanning the 2A/2B junction (-NPG↓P-) is completely conserved. Ryan *et al.* (1991) has shown that the sequence of the 18 amino acids of FMDV 2A plus the N-terminal proline of 2B was capable of the cleavage. Expression of truncated 2A/2B region of EMCV *in vitro* has shown similar results. The mechanism of this kind

of cleavage is not fully understood yet. It seems that the 2A of cardio- and aphthoviruses has no proteolytic activity but mediates a novel co-translational reaction that couples the release of nascent L-P1-2A (cardioviruses) or P1-2A (aphthoviruses) precursor and the initiation of translation from the proline of 2B. The primary cleavage in hepatoviruses is not mediated by 2A. In hepatitis A virus, the cleavage is mediated by 3C protease at the 2A/2B junction (Jia *et al.*, 1993; Schultheiss *et al.*, 1994).

Secondary Cleavage

Most of the subsequent cleavages of the picornaviral structural and non-structural precursors are mediated by the viral protease 3C, which are present in all picornaviruses. However, in FMDV, one additional protease, the L protease, is involved in the release of itself from P1 protein in a co-translational intramolecular reaction. In addition, the 2A protease of entero- and rhinoviruses also carries out a cleavage within 3CD to give rise to alternate products, 3C' and 3D' (McLean *et al.*, 1976; Hanecak *et al.*, 1982). The biological function of this cleavage is not clear, and it has been shown not essential to viral replication in tissue culture (Lee and Wimmer, 1988). It also should be noted that in some literature, the cleavages between L/1A (release of the L protein) and 2C/3A (to separate the P2 and P3 precursors) are thought to be primary. Here, we use the definition of Krausslich & Wimmer and classify these cleavages as 'secondary'.

Generally, 3C releases itself and its precursors through stepwise autocatalytic reactions. The order and rate of the multiple cleavages carried out by 3C are not precisely determined. For EMCV, cleavages at 2B/2C and 2C/3A are rapid and coordinated (Jackson, 1986, 1989). In poliovirus, cleavages at 2A/2B and 2C/3A are probably carried out early in the cascade, resulting in early appearance of 2BC when the entire polyprotein is expressed *in vitro*. Lawson and Semler (1992) have proposed that two

alternative pathways might exist: one rapid, *cis*-like cascade is detected *in vitro* and in the membranous fraction of infected cell cytoplasm, in which the 3C first cleaves at 2A/2B junction and then 2C/3A junction; another slower, *trans*-like cascade occurs in the soluble fraction, in which the 3C first cleaves at 2C/3A junction. Similar events were observed in FMDV by Flint (personal communication). The cleavages at 3A/3B and 3C/3D seem slow, so 3AB and 3CD are quite stable. The addition of 3AB greatly enhances the release of 3C from 3CD precursor (Molla *et al.*, 1994).

The capsid precursor released by primary cleavage is also processed by 3C proteolytic activity, apparently in a *trans* reaction. In EMCV, all proteins containing 3C, i.e. 3ABCD, 3ABC, 3CD and 3C, are capable of processing the L-P1-2A precursor. The processing is carried out in a defined order: L↓1ABCD2A, 1ABCD↓2A, 1ABC↓1D, and then 1AB↓1C (Shih, 1979; Jackson, 1986). In poliovirus, however, the P1 (1ABCD) protein is processed by 3CD, instead of 3C, which can not cleave at the 1C/1D junction.

Maturation Cleavage

The secondary processing of capsid precursor generates three proteins, VP0 (1AB), VP1 (1D) and VP3 (1C). One 5S protomer consists of one copy of each of the 3 proteins, while 5 protomers form a 14S pentamer. Twelve pentamers are assembled into a provirion, in which a molecule of genomic RNA is packaged. Concomitant with the RNA encapsidation, VP0 is cleaved into VP2 (1B) and VP4 (1A), rendering the viral particle infectious. Since no protease responsible for this cleavage has ever been isolated, and the cleavage site is deeply buried within the virion, this final maturation cleavage seems to be an autocatalytic reaction (Rossmann *et al.*, 1985). However, the mechanism is still unclear (Harber *et al.*, 1991; Lee *et al.*, 1993).

1.3.3 *Comoviridae*

Comovirus and *Nepovirus* are two genera within the comovirus family of plant viruses. Both of them have bipartite positive-sense single-stranded RNA genomes packaged in different polyhedral particles. The B-particle (Bottom) contains the larger genomic RNA, called B-RNA (in comoviruses) or RNA-1 (in nepoviruses), while the M-particle (Middle) contains the smaller RNA, called M-RNA or RNA-2. Both RNAs have a VPg attached to the 5' termini, a 3' poly(A) tail, and contain a single open reading frame, which is expressed as a large polyprotein and subsequently processed into individual mature proteins by viral proteases. The genome organization and expression strategies of como- and nepoviruses are so similar to that of the picornaviruses that these viruses are assigned to the picornavirus superfamily.

Cowpea mosaic virus (CPMV) is the type member of the comoviruses (Figure 1-4). The M-RNA encodes 3 proteins, the movement protein at the N-terminus of the polyprotein, followed by two capsid proteins, 37kDa and 23 kDa, which are not homologous to the counterpart of picornaviruses. Translation of M-RNA can start at either of the two in-frame start codons, resulting in a polyprotein of 105 kDa or 95 kDa. The 200 kDa polyprotein coded for by the B-RNA of CPMV contains 5 non-structural proteins, starting from the N-terminus, which are 32kDa, 58 kDa (putative helicase), 4 kDa (VPg), 24 kDa (protease) and 87 kDa (polymerase). The 58 kDa, 4 kDa, 24 kDa and 87 kDa proteins show significant homology to the 2C, 3B, 3C and 3D proteins of picornaviruses respectively (Franssen *et al.*, 1984; Argos *et al.*, 1984; Koonin and Dolja, 1993). The order of domains, Vpg-protease-polymerase, is identical to that of the picornavirus family.

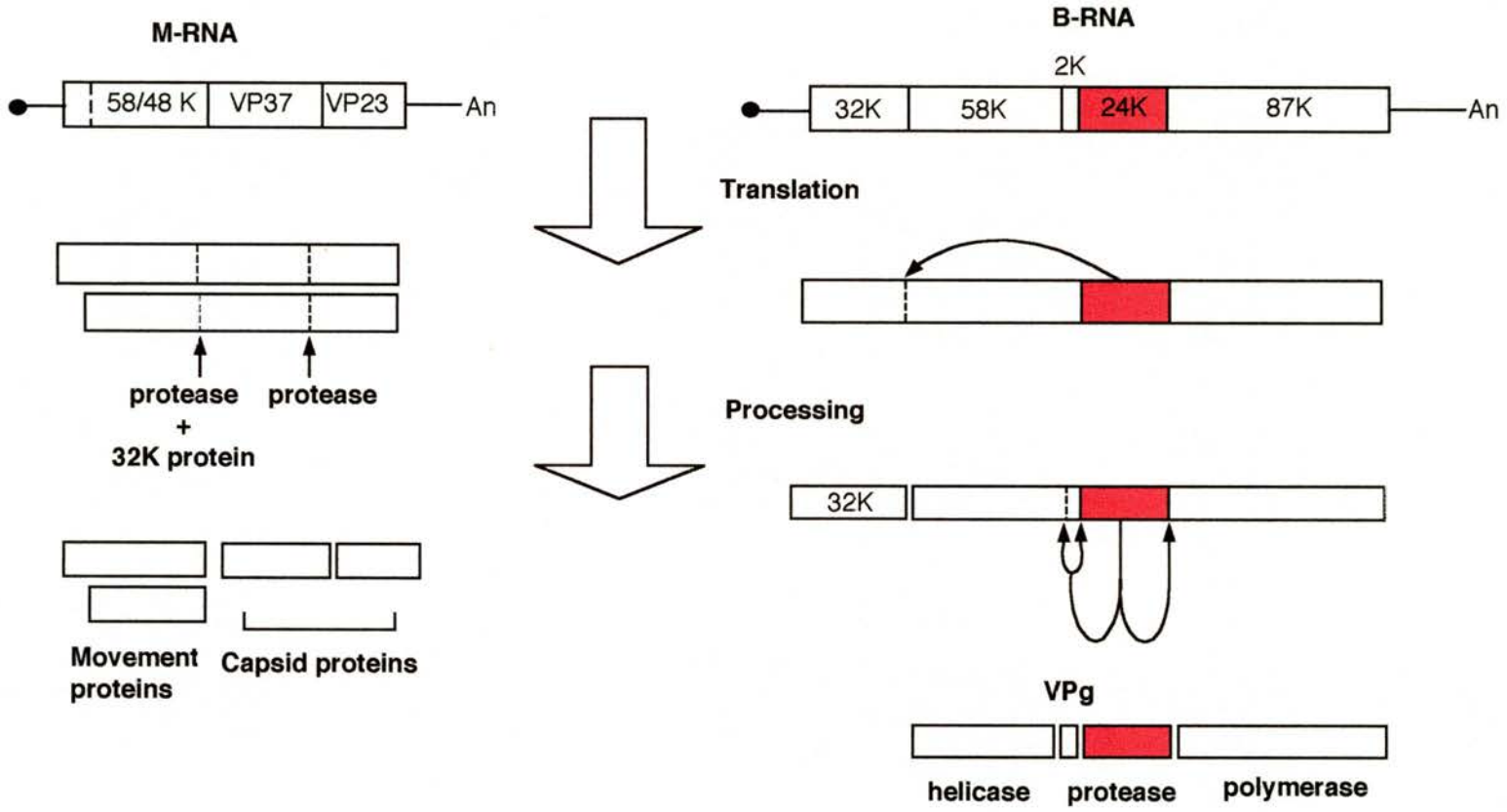


Figure 1-4 CPMV genome organization and polyprotein processing
The 24 kDa serine protease is indicated by the pink box. Translation of M-RNA can initiate at two in-frame AUG codons, giving rise to two products. The 32 kDa protein has an important regulatory role in processing.

As mentioned above, poliovirus encodes two proteases: (1) $2A^{pro}$, which separates the structural and non-structural precursors in a primary cleavage; (2) $3C^{pro}$, which is responsible for the remaining cleavages except the maturation cleavage. CPMV expresses its structural and non-structural proteins from two genomic RNAs, so it is not surprising that CPMV encodes only one protease, the 24 kDa protease, structurally and functionally similar to the $3C^{pro}$, to carry out all the cleavages.

The processing starts with a *cis* cleavage performed by the 24 kDa protease at the C terminus of 32 kDa protein, which still associates with the 170 kDa polyprotein after the cleavage, and slows down further processing. This results in temporal accumulation of the 170 kDa and other intermediate precursors, one of which, the 110 kDa (24 kDa + 87 kDa) has been shown to have polymerase activity (Dorssers *et al.*, 1984). If the 170 kDa polyprotein is expressed from B-RNA with the 32 kDa coding sequence deleted, it is processed rapidly and completely (Peng and Shih, 1984; Peter *et al.*, 1992). On the other hand, the 32 kDa protein is required for the 24 kDa protease to cleave between the movement protein and the 66 kDa capsid precursor, though further processing of the 66 kDa precursor can proceed by the 24 kDa protease alone.

Nepoviruses are similar to comoviruses in genome organization and expression strategies, though there is no protein that regulates the activity of the viral protease like the CPMV 32 kDa protein.

1.3.4 *Potyviridae*

The potyvirus family, a member of the picornavirus superfamily, contains many economically important pathogens. It has three genera: the **poty-** and **rymoviruses** have

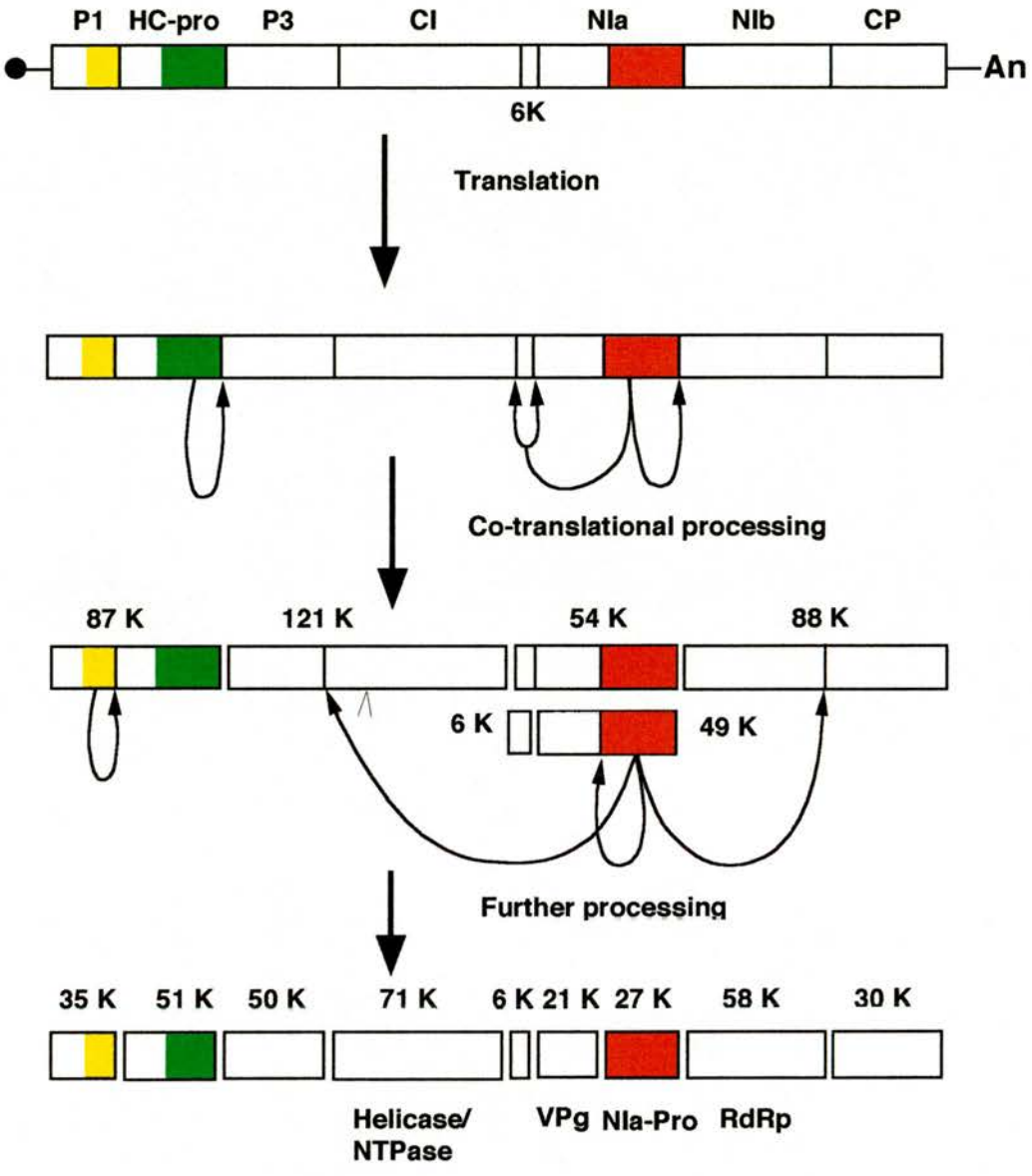


Figure 1-5 Genome organization and polyprotein processing of TEV
Yellow indicates the P1 serine protease, green the HC-protease and pink the NIa protease.

monopartite single-stranded, positive-sense RNA genome, while the **bymoviruses** are a bipartite version. All RNAs have a VPg covalently linked to the 5' terminus and a poly(A) tail at the 3' end.

Tobacco etch virus (TEV) is the best-studied member of the potyviruses. Its 9596 n.t. genomic RNA encodes a large polyprotein of 351kDa, which is cleaved by three viral proteases into a series of intermediate precursors and mature proteins (Riechmann *et al.*, 1992; Dougherty and Semler, 1993). During the translation, the HC-protease cleaves at its C-terminus *in cis* to release the 87 kDa polyprotein, which is further processed by the 35 kDa P1 protease autocatalytically. The remaining processing is mediated by the 27 kDa NIa protease, which resides between the VPg and polymerase domains. Therefore, the arrangement of VPg-protease-polymerase is also present in TEV. At first, the 27 kDa NIa protease cleaves at its C-terminus and the N-terminus of the VPg to release itself as a 49 kDa (major) or 55 kDa (minor) precursor. The 49 kDa precursor (VPg-Pro) is the equivalent of the 3BC of the picornaviruses and capable of processing the 121 kDa protein, the 88 kDa protein and itself (Carrington *et al.*, 1988; Dougherty and Parks, 1991).

The genome structure and expression strategies of bymoviruses are similar to those of the potyviruses. However, bymoviruses do not encode the P1 protease (Dougherty and Semler, 1993).

1.3.5 *Flaviviridae*

The family *Flaviviridae* contains three genera: **flavi-**, **pesti-** and **hepatitis C-like viruses**. Most of them are arthropod-borne, and transmitted to vertebrates by chronically infected mosquito or tick vectors. The small, enveloped virions contain a single-stranded

positive sense genomic RNA with a size of about 9.5-12.5 kb. Most viruses lack a 3' end poly(A) tail. For flaviviruses, the 5' end of the RNA is capped and the initiation of translation is believed to occur by ribosome scanning; however, for pestiviruses and HCV, the genomes are thought to be translated by a cap-independent internal ribosome binding mechanism, due to the complex secondary structures in the long 5' NCR and the existence of multiple AUGs upstream of the actual start codon (Rice, 1996; Poole *et al.*, 1995). The most notable feature of the viruses in this family is the presence of a single, long open reading frame with an organization similar to that of the picornaviruses. However, the polyproteins of the *Flaviviridae* are processed co-translationally and post-transitionally by both host proteases (signalases) and viral proteases (reviewed by Rice, 1996; Ryan *et al.*, 1998).

Flaviviruses

Yellow fever (YF) and Dengue (DEN) viruses are well-studied flaviviruses. The long polyprotein contains about 10 proteins in the order of NH_2-anchC(C)-prM(M)-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5-COOH. Three structural proteins are located at the N-terminus: capsid protein C (its precursor is called anchored C or anchC), membrane protein M (its precursor is called prM) and envelope protein E. The remaining proteins are non-structural, including the viral protease NS3 and polymerase NS5 (Figure 1-6).

The translation and processing of the polyprotein show a delicate membrane topology (for a review, see Rice, 1996). Translation is associated with the rough ER. The hydrophobic region at the C terminus of nascent C protein acts as an internal signal sequence for translocation of prM into the ER lumen, where prM is glycosylated. Two adjacent hydrophobic stretches separated by a charged residue function as a stop-

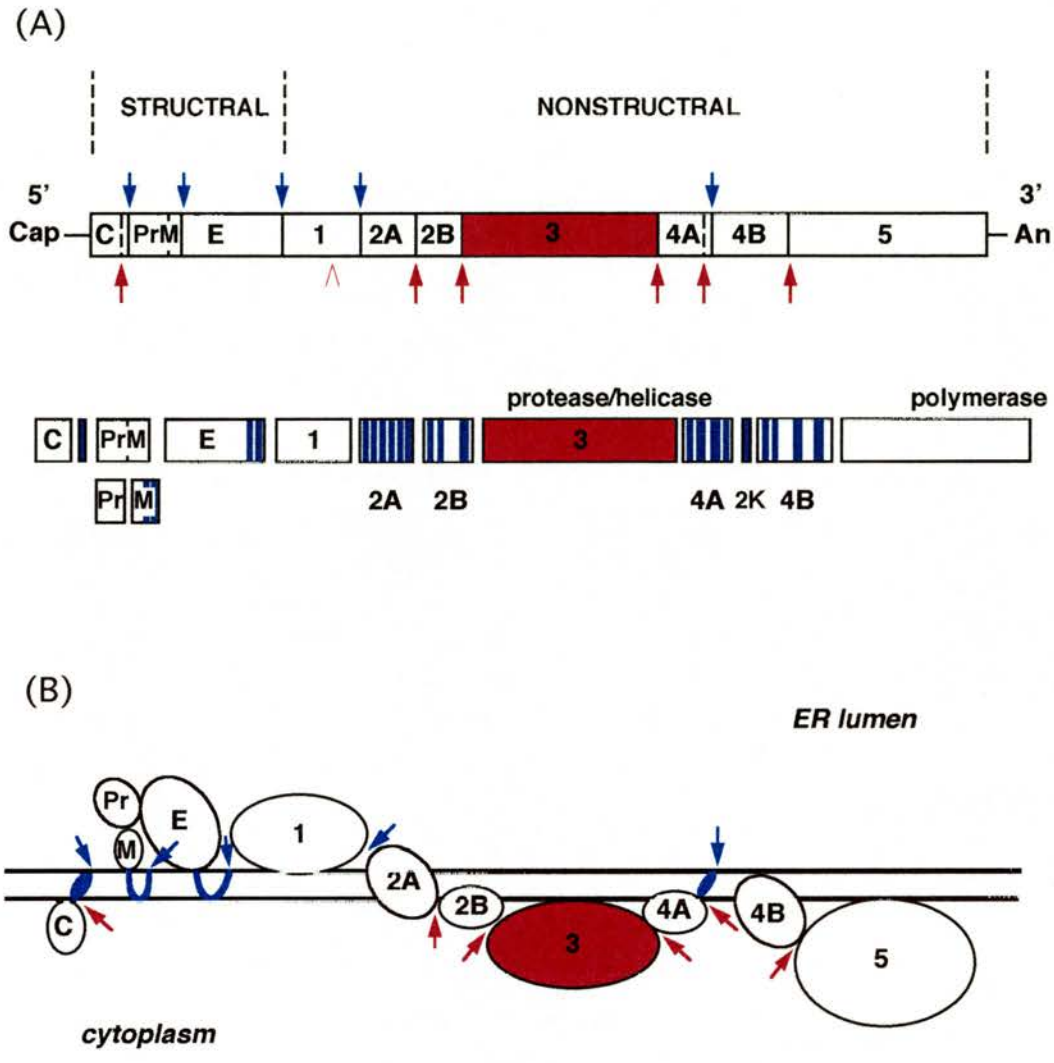


Figure 1-6 Polyprotein processing of the flavivirus
(A) Genome organization and processing scheme. Viral protease NS3 is shown in pink, while continuous stretches of uncharged amino acids are shown in blue bars. Cleavage sites for host and viral proteases are indicated by blue and pink arrows, respectively.
(B) A model for the membrane topology of the flavivirus proteins. C-terminal hydrophobic regions are shown in blue colour

transfer signal for prM and start-transfer signal for the E protein. The hydrophobic regions at the C-terminus of E protein have similar function. Therefore, C protein is anchored on the cytoplasmic side of ER, while prM, E and NS1 are translocated into the lumen of ER. NS2A, NS2B, NS4A and NS4B are quite hydrophobic and embedded in the membrane. NS3 and NS5 are present at the cytoplasmic side.

The release of C from its membrane anchor is mediated by the protease complex of NS3 plus NS2B (called NS2B-3 protease) and seems to be required for efficient cleavage at the anchC/prM junction by the signalase. The signalase in the ER is also responsible for the generation of the N-termini of the proteins E, NS1, NS4B, and probably NS2A. The cleavage at NS1/NS2A is unique because it occurs in the ER lumen and requires the eight C-terminal residues of NS1 plus a large part of NS2A sequences. The cleavage site follows the rule for a signalase site but is not at the C-terminus of a hydrophobic signal peptide. Interestingly, the downstream NS2A is hydrophobic but is not a protease. The maturation cleavage of prM to generate M protein occurs in the ER lumen and is delayed (shortly before or during virion release). The remainder of the cleavages is mediated by the NS2B-3, including the release of NS4A from its C-terminal hydrophobic segment of 2kDa, as shown in Figure 1-6.

Pestiviruses

This genus comprises three species: classical swine fever virus (CSFV), bovine viral diarrhea virus (BVDV) and border disease virus (BDV) of sheep. The processing of pestiviral polyprotein seems similar to that of flaviviruses though some important differences exist (Figure 1-7).

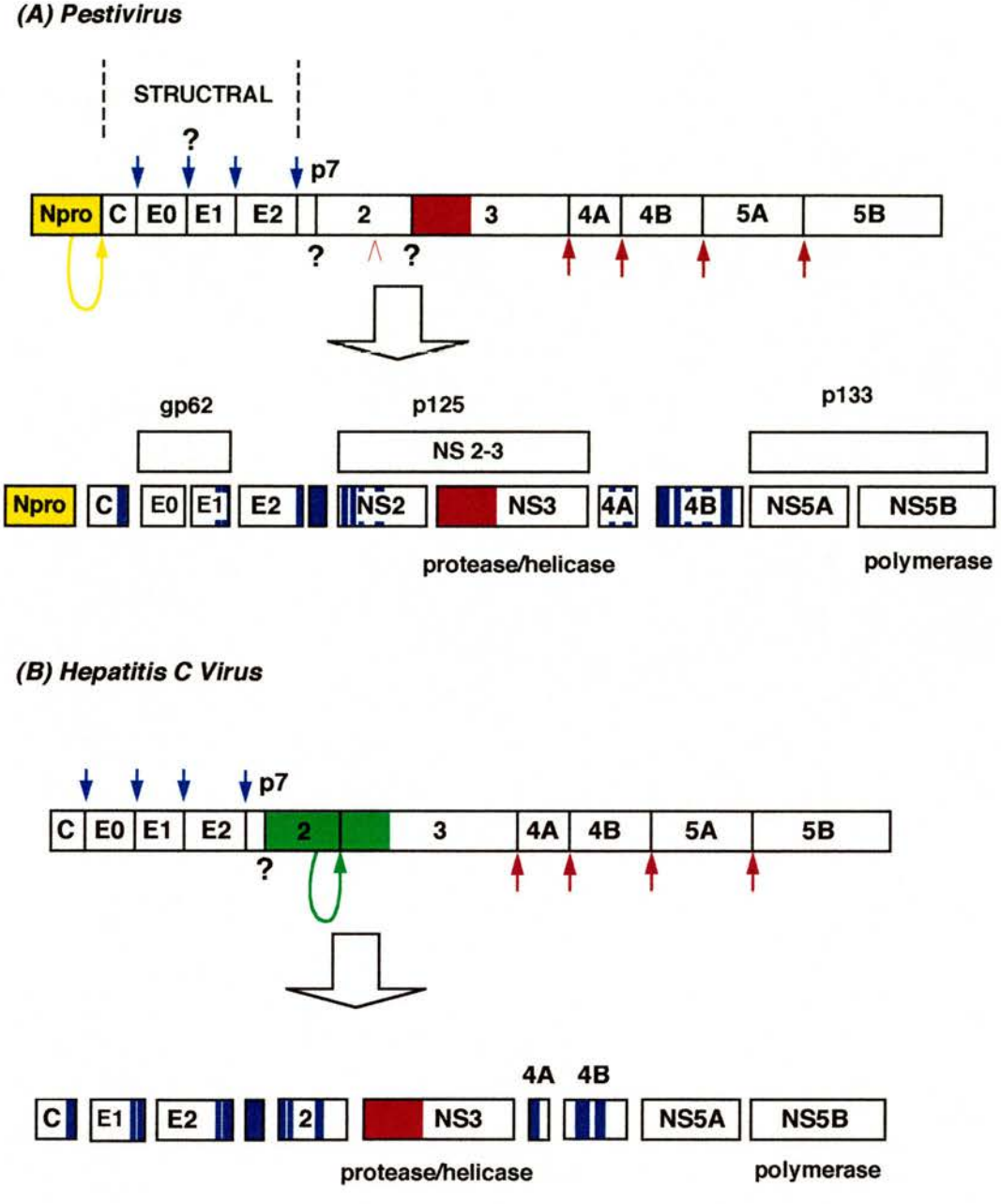


Figure 1-7 Polyprotein processing of the pestivirus (A) and the Hepatitis C virus (B)
Cleavage sites for host protease (signalase), viral papain-like cysteine protease, viral serine protease, and proposed viral metallo-protease are indicted by blue, yellow, pink and green arrows, while corresponding viral proteases are shown in boxes of the same colours. Hydrophobic regions are shown as blue bars. For HCV, the NS3 serine protease and NS2-3 metallo-protease are partly overlapped with each other.

The first cleavage event is the autocatalytic self-release of the N-terminal protease (N^{pro}) from the nascent polyprotein. The N^{pro} is a papain-like cysteine protease and has no counterpart in HCV and flaviviruses. The signal sequence at the C-terminal of capsid protein C starts the translocation of E proteins (E0, E1 and E2), which are glycosylated in the lumen. The cleavages at C/E0, E1/E2 and E2/p7 by signalase are rapid, but cleavage at E0/E1 is delayed (Rice, 1996; Meyers and Thiel, 1996).

The following two thirds of the ORF encode non-structural proteins. The mechanism to release p7 is unknown. The next released product is the polyprotein NS2-3, which for most pestiviruses is processed to some degree to generate small amounts of NS2 and NS3. However, in cells infected with cytopathogenic BVDV (cpBVDV), at least the same amount of NS3 as NS2-3 is produced. In this case, the NS3 is generated via novel RNA recombination event rather than cleavage from the NS2-3 precursor (Greiser-Wilke *et al.*, 1993; Meyers *et al.*, 1989, 1991, 1992).

Pestiviral NS3 contains motifs of serine protease, NTPase and helicase, the activities of which have be demonstrated (Wiskerchen and Collett, 1991; Tamura *et al.*, 1993; Warrener and Collett, 1995). It is believed that NS3 or NS2-3 is responsible for cleavages at four downstream sites to generate the N-termini of NS4A, NS4B, NS5A and NS5B.

Hepatitis C Virus

The genome organization and polyprotein processing of HCV are similar to those of pestiviruses. Host signalase processes the structural-NS2 region. The most striking feature of HCV is that two proteases of different catalytic types are encoded within an

overlapping region. The N-terminal part of NS3 encodes a serine protease, which mediates the downstream cleavages at NS3/NS4A, NS4A/NS4B, NS4B/NS5A and NS5A/NS5B, similar to flaviviruses and pestiviruses. Unlike flaviviruses, this protease activity does not require upstream NS2 sequences (Eckart *et al.*, 1993; Grakoui *et al.*, 1993; Bartenschlager *et al.*, 1993). The other protease is associated with the NS2 plus N-terminal one third of NS3 and mediates the *cis* cleavage at the C-terminus of NS2. This protease is not affected when the serine protease activity of NS3 is abolished by site-directed mutagenesis and is thought to be the only example of viral Zn^{2+}-dependent metalloprotease (Rice, 1996; Ryan *et al.* 1998).

1.3.6 *Togaviridae*

This family comprises two genera: the **alpha-** and **rubiviruses**. They are simple enveloped animal viruses with a single-stranded, positive -sense genomic RNA. Many members can replicate in and be transmitted by mosquitoes. A subgenomic RNA, corresponding to 3'-terminal one third of the genomic RNA, is synthesized to express structural proteins. Non-structural proteins are translated from genomic RNA. Both genomic and subgenomic RNAs are capped at the 5' end and polyadenylated at the 3' end. Both structural and non-structural proteins are translated as polyproteins and processed into individual proteins (for a review, ten Dam *et al.*, 1999).

Sindbis virus (SIN), the type member of alphaviruses, has been studied extensively. Another well-studied member is Semliki Forest virus (SFV). A model of transcription, translation and replication of viral genome has been proposed, based on the studies of these two viruses (Hardy and Strauss, 1989; Shirako and Strauss, 1990; de Groot *et al.*, 1990; ; Sawicki and Sawicki, 1993; Lemm and Rice, 1993a, b; Lemm *et al.*, 1994; Strauss and Strauss, 1994; Wang *et al.*, 1994; Dé *et al.*, 1996). Generally, non-structural

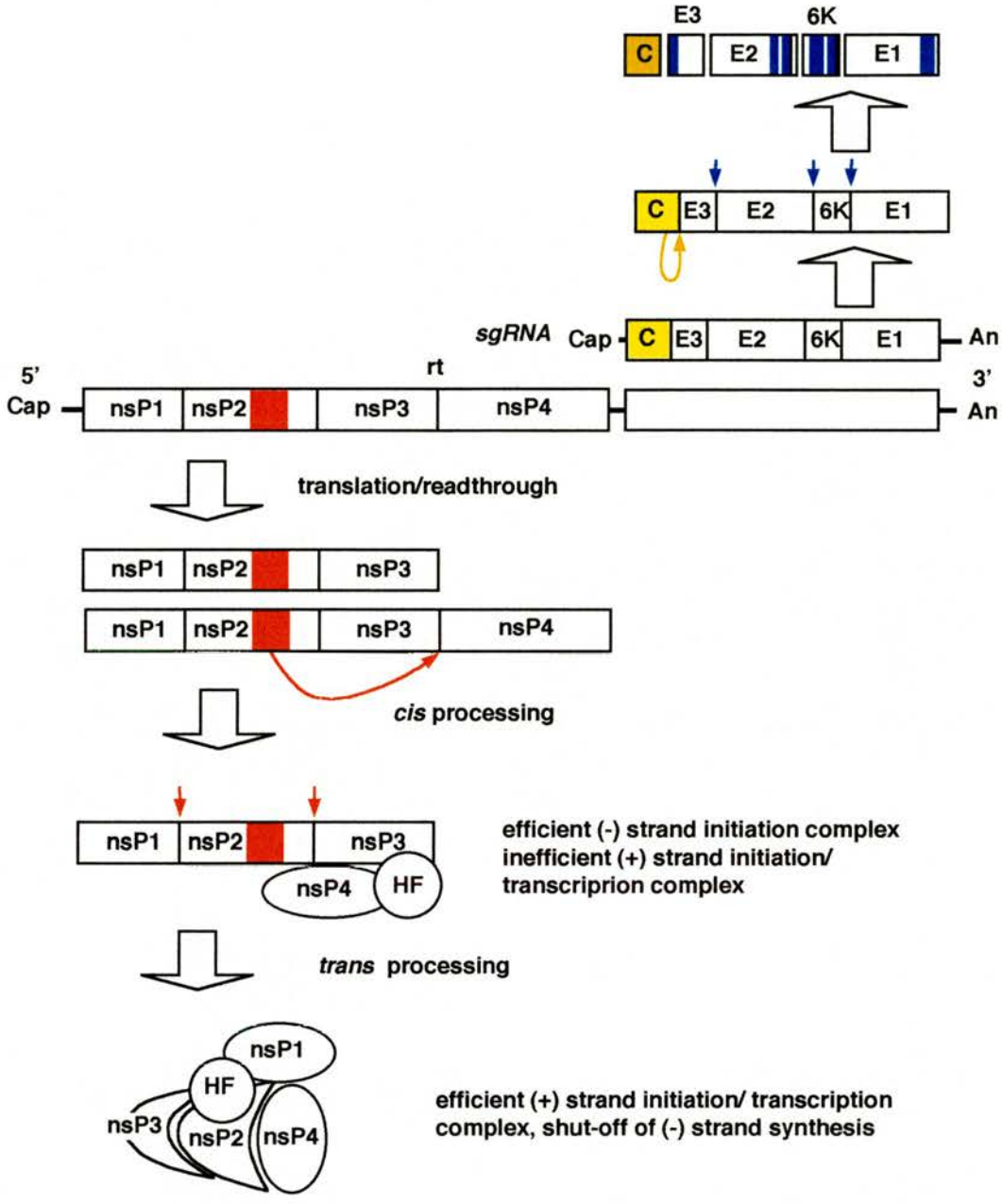


Figure 1-8 A model for gene expression of the alphavirus
Proteolytic domains are shown in pink for the nsP2 cysteine protease, or yellow for the capsid serine protease. Host factors (HF) are probably involved in the formation of replicase complex. Cleavage sites for host proteases (furin-like protease and signalase) are indicated by blue arrows.

proteins are synthesized from genomic RNA and form a replicase complex that synthesizes the complementary minus strand RNA, which is used as template for synthesis of plus strand genomic and subgenomic RNA. Genomic RNA is encapsidated into nucleocapsids. Viral transmembrane glycoproteins are transferred from ER to the host cell plasma membrane, where nucleocapsids are enveloped and released via budding (Figure 1-8).

In Sindbis virus, translation of genomic RNA is initiated at a single AUG near the 5' end and stops at the opal stop codon before nsP4 to yield polyprotein P123, or reads through this codon to yield P1234. A cysteine protease activity is found in the C-terminal portion of nsP2 and is responsible for the processing of non-structural polyproteins. The nsP3/4 site is cleaved rapidly in a *cis* reaction to form a complex of P123 plus nsP4, which initiates minus-strand synthesis efficiently. This complex is gradually processed at the nsP1/nsP2 and nsP2/nsP3 sites. The fully processed complex is efficient in plus-strand initiation/transcription, but cannot carry out minus-strand initiation/transcription. In Semliki Forest virus, there is no stop codon before nsP4, so initially nsP4 is translated in the same amount as nsP1-3. The Semliki Forest virus nsP4 is reported to be an autoprotease, which cleaves at nsP3/nsP4 site and control the level of nsP4 (Takkinen *et al.*, 1990).

Structural proteins are translated 2-3 hours post-infection, upon the synthesis of subgenomic RNA. The first protein, the capsid protein C, has serine protease activity and release itself to the cytoplasm via a co-translational, intramolecular mechanism. The remainder of the polyprotein is translocated into ER lumen by the N-terminal signal sequence of E3. Signalase in the ER lumen cleaves at E2/6k and 6k/E1 sites. Various post-translational modifications, i.e. glycosylation, acylation and phosphorylation, are carried out during the course of transport from ER to plasma membrane. The maturation

cleavage at E3/E2 is mediated by a host furin-type protease in the trans-Golgi vesicle, as is important to generate infectious progeny virions.

1.3.7 *Retroviridae*

The retroviruses are a large family, currently divided into eleven genera: **alpha-, beta-, gamma-, delta-, epsilonretrovirus, avian type C, mammalian type C, type D retroviruses, intracisternal A-particles, lentiviruses** and **spumaviruses**. Some of them are associated with a wide range of serious diseases in vertebrates, including acute or long-latency malignancies, neurological disorders, and AIDS.

The retroviruses have a unique life cycle. The enveloped virions have two copies of single-stranded positive-sense genomic RNA, the only diploid virus. After entering the cell, reverse transcription of the genomic RNA into double-stranded DNA occurs in the virion core. The viral DNA is then transported into the nucleus, where it is integrated into cell chromosome. The provirus is then transcribed by cellular RNA polymerase II into viral genomic RNA, some of which is further spliced into at least one kind of subgenomic RNA (Figure 1-9). Both genomic and subgenomic RNAs are capped and polyadenylated, from which structural and non-structural viral proteins are translated as several polyproteins. Viral membrane-associated glycoproteins (encoded in *env* gene) are translocated into ER and processed by cellular proteases on the secretory path to the plasma membrane. Two polyproteins are translated from the genomic RNA, *gag* (or *gag-pro*) and *gag-pro-pol*, via frameshifting or readthrough of stop codon (Figure 1-10). The *gag* is the precursor of 3 to 5 capsid proteins, always including the matrix (MA) protein, the capsid (CA) protein and the nucleic acid-binding (NC) protein. The *pro* gene encodes the protease (PR) that processes the *gag* and *pol* polyproteins, and

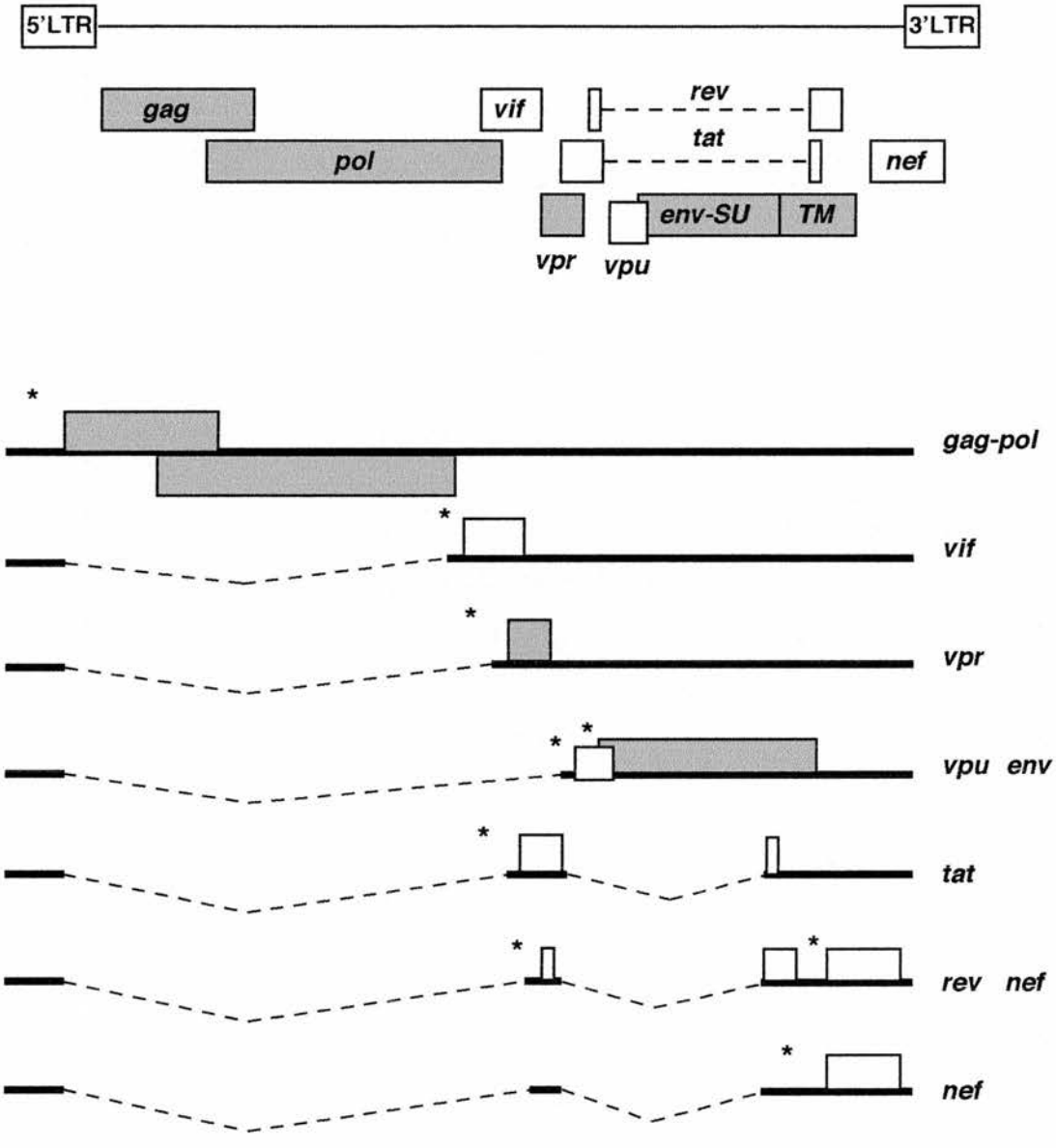


Figure 1-9 Genome organization and RNA splicing pattern of HIV-1
Open reading frames are shown as proviral DNA forms. Structural genes are shaded. Sites of translation initiation are indicated by asterisks.

sometimes part of *env*. The *pol* gene encodes two proteins, the reverse transcriptase (RT) and the integrase

(IN). Virus assembly and budding usually take place simultaneously. The NC region in the *gag* precursor protein binds to the viral genomic RNA, the middle CA region interacts with the CA in other precursors (mainly *gag* or *gag-pro*, small amount of *gag-pro-pol*), while the MA region interacts with the cell membrane. The cleavage products of *env* polyprotein, the surface (SU) protein and the transmembrane (TM) protein, are incorporated into virion envelope, probably via interaction with the MA protein. After assembly, the PR protease in the virion is activated and cleaves the precursors into mature proteins.

Although all retroviruses share the basic strategies of replication mentioned above, the details are distinctive. Human immunodeficiency virus-1 (HIV-1) has been subject of intense scrutiny for many years, and is given here as an example. The genome organization, RNA splicing and polyprotein processing schemes are shown in Figure 1-9 and 1-10. Apart from the *gag*, *pro*, *pol*, and *env* genes, HIV also encodes 6 additional proteins, which have essential (*tat* and *rev*) or accessory (*vif*, *vpr*, *vpu* and *nef*) functions in viral replication. For HIV-1, the PR protease gene is in frame with the *pol* gene and is only translated in the *gag-pol* polyprotein, via ribosomal frameshift. Both *gag* and *gag-pol* polyproteins are translated from the genomic RNA, while *env* is translated from a singly spliced subgenomic RNA. PR protease cleaves at 8 specific sites, 4 in the *gag* region, 4 in the *pol* region (Coffin, 1996; Luciw, 1996).

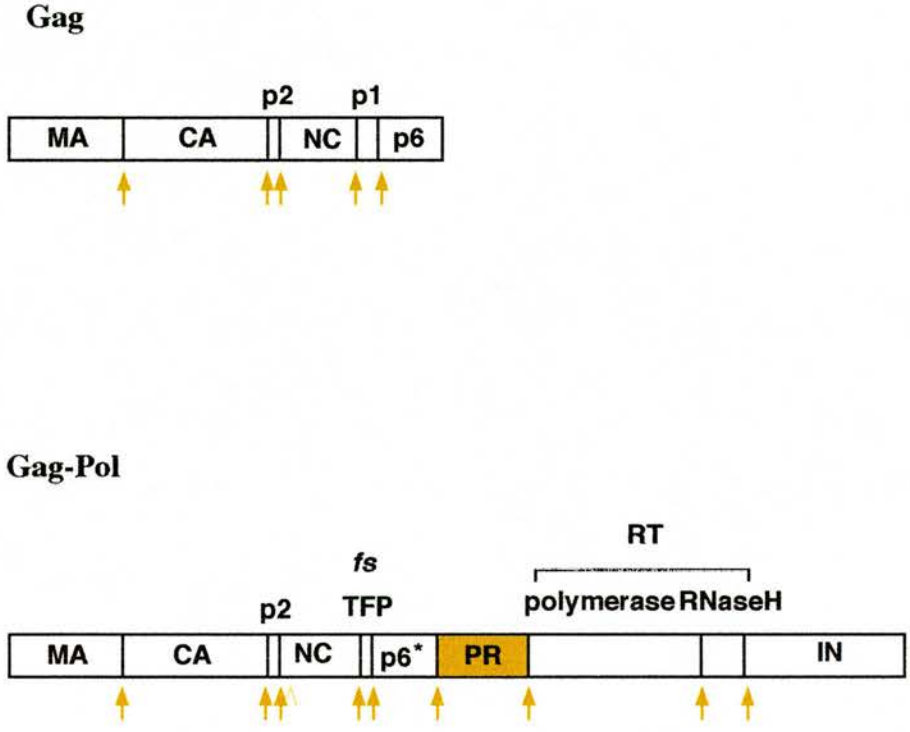


Figure 1-10 Gag and Gag-Pol polyproteins of HIV-1
Viral aspartic acid protease (PR) is shown in yellow box. Arrows of the same colour indicate the its cleavage sites. *Pol* sequences are translated via a frameshift within the region encoding the transframe octapeptide (TFP). Other abbreviations: MA, matrix protein; CA, capsid protein; NC, nucleocapsid; RT, reverse transcriptase; IN, integrase

1.4. Proteases in RNA Viruses

1.4.1 Protease Overview

In the past, three terms, *protease*, *proteinase* and *peptidase*, have been used for the enzymes that hydrolyze peptide bonds. To avoid confusion, the scheme used by Alan Barrett (1986) is adopted here: *protease* (or peptidase) is the general term for 'peptide bond hydrolase', whilst *proteinase* refers to enzyme that catalyzes the hydrolysis of internal peptide bonds of proteins or polypeptides (endopeptidase), in contrast to exopeptidase, which hydrolyzes terminal peptide bonds.

Hydrolysis of a peptide bond requires a nucleophile that attacks the carbonyl carbon of the scissile bond. According to the nature of nucleophiles, proteases have been classified into four major types: serine, cysteine (or thiol), aspartic (or acidic), and metallo-proteases (Hartley, 1960; Barrett and McDonald, 1980). Recently, the fifth type of proteases has been proposed, which use a threonine residue as the nucleophile (Seemuller *et al.*, 1995). Within each type of protease, the type and spatial position of amino acids participating in catalysis are conserved, though the primary sequences of the protease might show little homology. To achieve the conserved orientation of residues at the active centre, different proteases have evolved distinctive structures, which constitute the basis for further classification. The term *family* is used to describe a group of proteases that have arisen from a single ancestral protein, as shown by homology in primary structure. If several families have evolved from a common ancestor, as indicated by the linear order of active site residues, clusters of conserved amino acids around the catalytic residues, and the three-dimensional structures, without

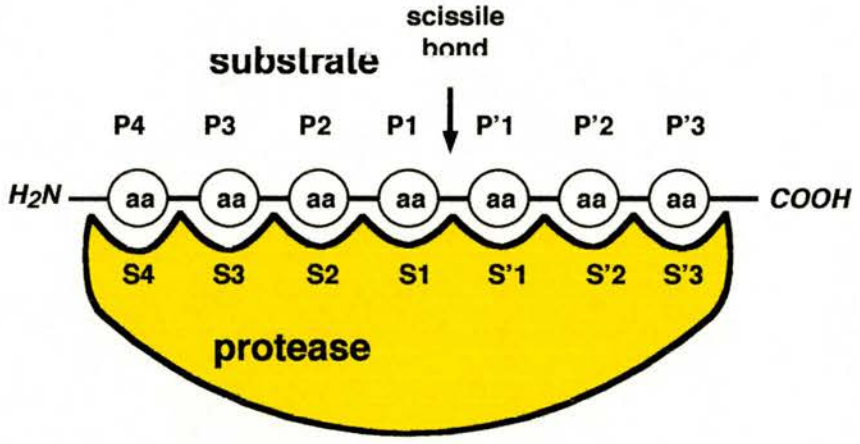


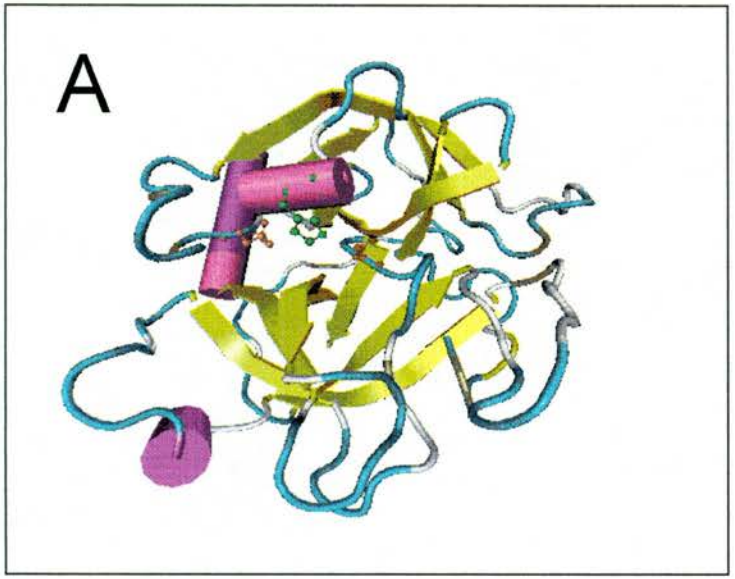
Figure 1-11 The nomenclature (Schechter and Berger, 1967) describing a protease and its substrate
Residues in the peptide substrate are labeled P4 through P'3, whilst subsites on the protease, which bind P4 through P'3 and collectively make up the substrate-binding pocket, are labeled S4 through S'3.

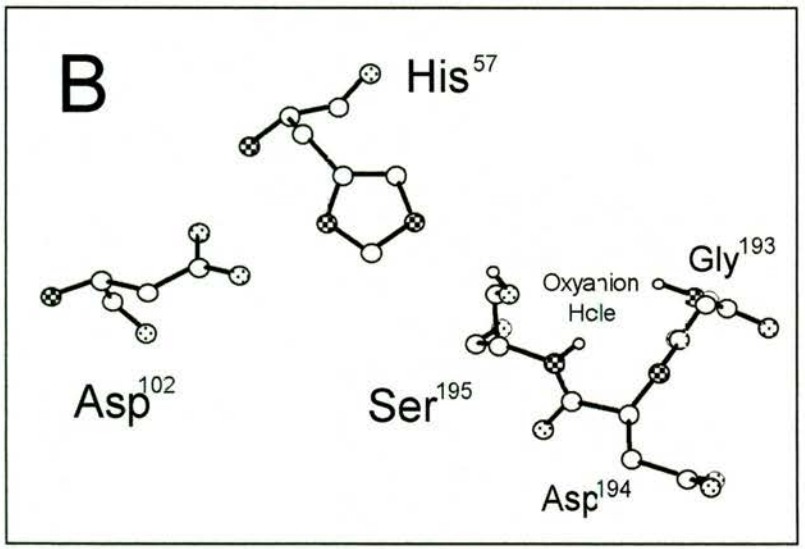
The basic structure of a protease can be viewed as a catalytic centre plus an adjacent substrate-binding pocket. The substrate-binding pocket is believed to consist of a series of subsites, each accommodating the side chain of a single residue on substrate protein. These subsites and corresponding residues encompassing the cleavage site are numbered with a scheme proposed by Schechter and Berger (1967), as shown in Figure 1-11. Substrate specificity, determined by features in the substrate-binding pocket, is the most idiosyncratic for a protease. There is a great diversity even within a family.

1.4.1.1 Serine Proteases

Over 20 families of proteases dependent on a serine residue for catalytic activity are found in animals, bacteria and viruses, and can be grouped into 6 clans: SA (chymotrypsin), SB (subtilisin), SC (Carboxypeptidase C), SE (*Escherichia* D-Ala-D-Ala peptidase A), SF (repressor LexA), and SG (clp endopeptidase). Structures have been known for at least one member of the first four clans, which have been shown to be evolutionarily unrelated. In clan SA, the chymotrypsin family (S1, including chymotrypsin, elastase, thrombin, and trypsin) and α-lytic endopeptidase family (S2) are the two biggest families (Rawlings and Barrett, 1994a).

The arrangement of the catalytic residues in the primary sequence shows the differences at the clan level, e.g. chymotrypsin clan (His-Asp-Ser), subtilisin clan (Asp-His-Ser), carboxypeptidase clan (Ser-Asp-His), *Escherichia* D-Ala-D-Ala peptidase A clan (Ser-Lys). The first three have a classic catalytic triad, which is similarly orientated in the three-dimensional structure.





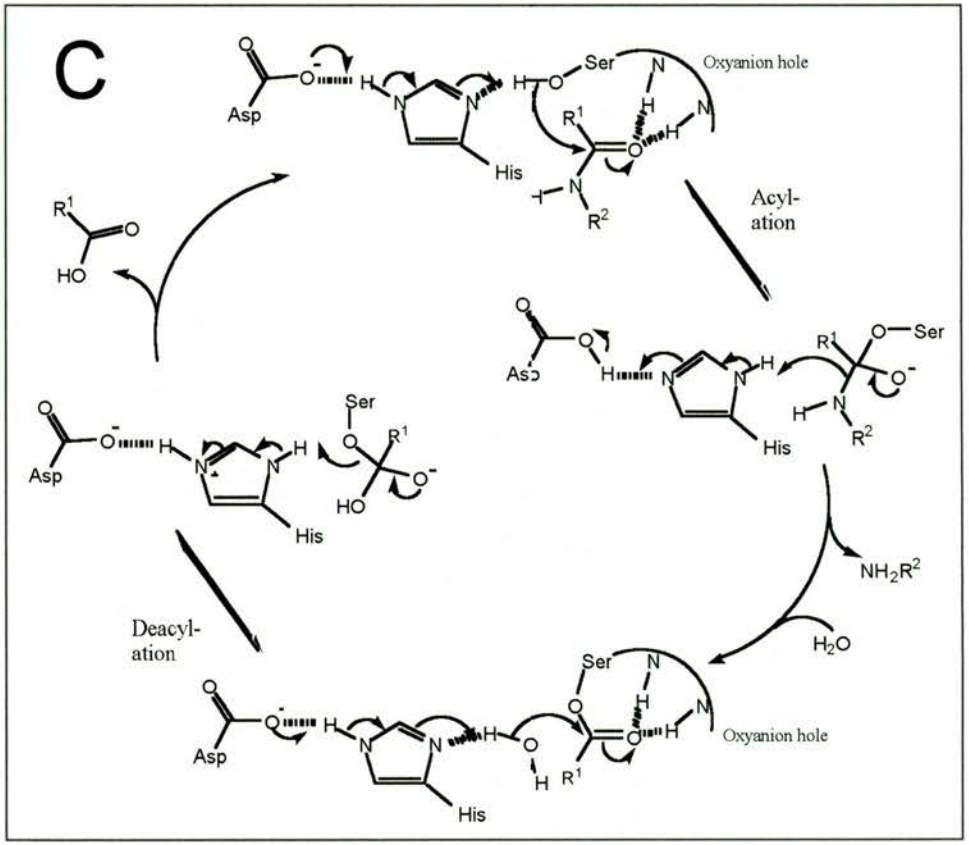


Figure 1-12 Serine protease (Trypsin)

A. Three-dimentional representation of the protease, the catalytic residues are shown in CPK mode.

B. The active site: the catalytic triad (Asp102, His57 and Ser195) and the oxyanion hole.

C. Mechanism of catalysis.

The catalytic mechanism of this type of proteases is shown in Figure 1-12. The side-chain hydroxyl group of Ser-195, polarized by His-57, attacks the carbonyl carbon of the scissile bond, forming a tetrahedral intermediate stabilized by the oxyanion hole, which is an important structural feature in promoting the formation of the tetrahedral intermediate. The leaving NH group acquires a proton from His-57 as the scissile bond breaks. Subsequently, a molecule of water attacks the acyl-enzyme intermediate to complete the catalytic cycle.

Chymotrypsin is the best-studied serine protease, whose crystal structure is one of the earliest to be determined (Matthews *et al.*, 1967). It is composed of two homologous domains, each containing a six-stranded β-barrel. The two domains pack together asymmetrically, with the catalytic residues situated between them (Figure 1-12). The tertiary structure and catalytic site geometry of chymotrypsin are conserved in other members of family S1. In this sense, "chymotrypsin-like" and "trypsin-like" have similar meaning.

The inhibitor of choice to identify a serine protease is 3,4-DCI, which is safe and reacts rapidly and irreversibly with a wide range of serine proteases. Other inhibitors include DFP and PMSF, which are toxic and also inhibit cysteine proteases (but can be reversed by DTT). Besides, DFP is volatile, while PMSF tends to react more slowly. There are reports that certain serine proteases are inhibited by one serine protease inhibitor, but are insensitive to another one, so results of inhibition experiments should be interpreted with caution (Barrett, 1994).

1.4.1.2 Cysteine Proteases

About 20 families of proteases require a cysteine residue at the active centre. The CA clan consists of the best-known **papain** family and other two distantly related families (**calpain** and **streptopain**). The CC clan contains viral cysteine proteases, like HC-pro of TEV and p29 of chestnut blight virus. The clan CB contains many well-known viral cysteine proteases, e.g. picornaviral 3C, and NIa of TEV. Clan CB is distinct in that members of this clan have a chymotrypsin-like fold with significant homology across the β-barrel domains and the motifs forming the catalytic centre. They also have a catalytic triad of His, Asp (or Glu) and Cys, similar to that of chymotrypsin, with a Cys in place of the nucleophilic Ser. Some researchers call them serine-like proteases to indicate this feature (Dougherty and Semler, 1993).

The crystal structure of papain has been elucidated (Kamphuis *et al.*, 1984). The molecule consists of one polypeptide chain of 212 residues, folded into two lobes (L and R) with the active site in the groove between them (Figure 1-13). The L domain has three α-helices: helix LI has the catalytic Cys25 at its N-terminus, LII is buried, and LIII has one side exposed to solvent. The R domain has two helices (RII and RIII) located in the surface region at opposite ends of the anti-parallel β-sheet core, where the active site residue His-159 is located. Other residues have also been identified to play various roles in the catalytic mechanism.

The active form of papain contains a thiolate-imidazolium **ion pair** at neutral pH, which consists of Cys-25 and His-159. The amide oxygen $O_{\delta 1}$ of the Asn-175 side-chain is hydrogen bonded to the $N_{\varepsilon 2}$ of His159. The Cys-His-Asn triad is analogous to the Ser-His-Asp in chymotrypsin. The Asn-175 presumably modulates the rotation of the imidazole ring about the Cγ-Cβ bond, so that the imidazole ring is in

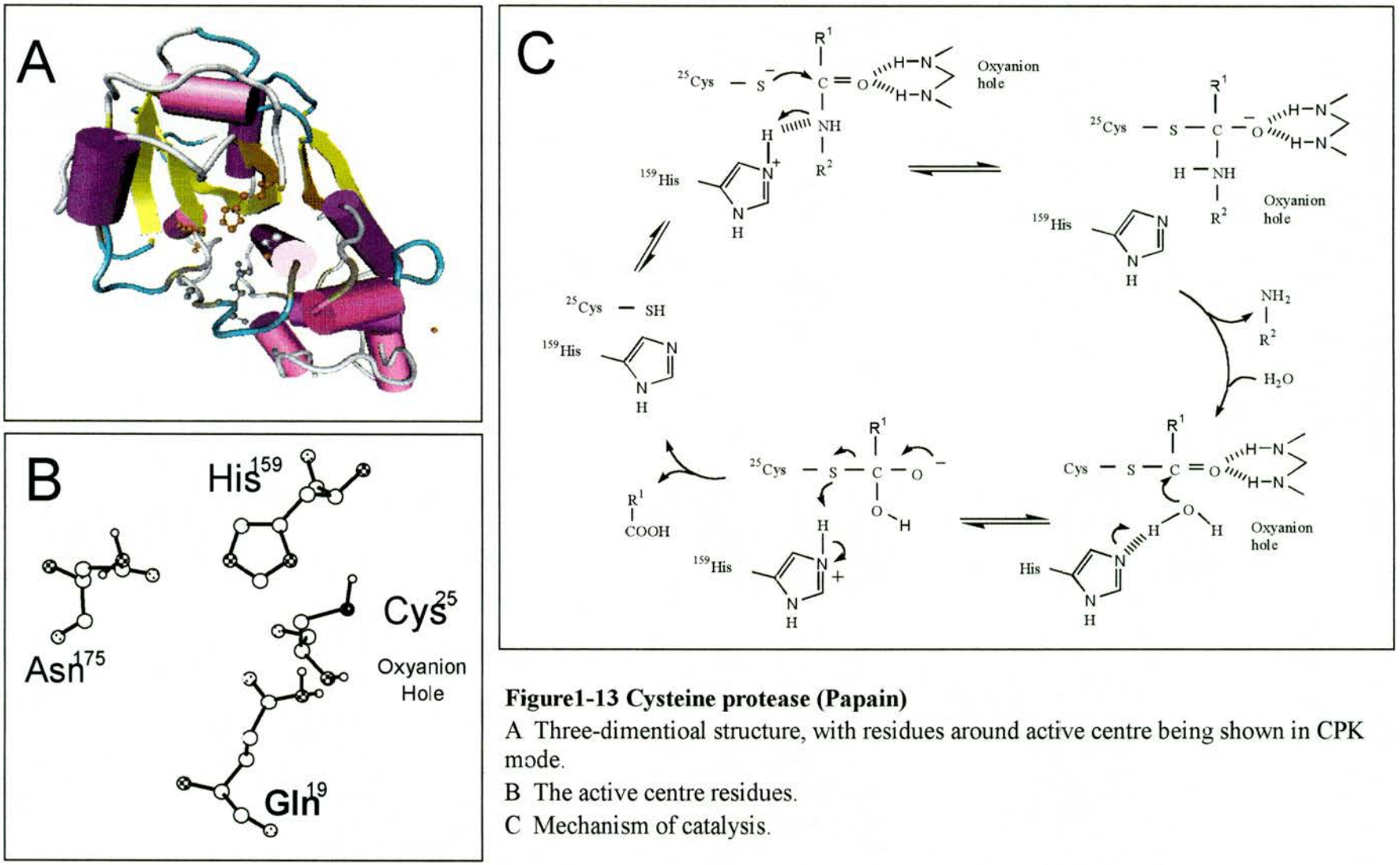


Figure1-13 Cysteine protease (Papain)

A Three-dimentioal structure, with residues around active centre being shown in CPK mode.

B The active centre residues.

C Mechanism of catalysis.

optimal orientations at different steps during catalysis. The oxyanion hole comprises dipoles from the main-chain amide of Cys-25 and the side-chain amide of Gln-19. The catalytic thiol group is unusually nucleophilic. The basic reaction of hydrolysis is shown in Figure 1-13, which includes nucleophilic attack on carbonyl carbon of the scissile bond, formation of acyl-enzyme and release of the amine RNH_2, transfer of the acyl group to a water molecule and regeneration of the enzyme.

Cysteine proteases in the papain family are characteristically susceptible to rapid and irreversible inactivation by E64, an epoxide inhibitor. More general thiol reagents, like iodoacetate, iodoacetamide, and N-ethylmaleimide (1mM) are required to recognize some cysteine protease of other families, e.g. clostripain and streptopain. Many serine and metallo- proteases show significant thiol dependence: they are activated by thiol compounds (e.g. DTT) and/or inhibited by thiol-blocking reagents. Therefore, assignment of catalytic type by using these general thiol reagents is often difficult (Barrett, 1994).

1.4.1.3 Aspartic Acid Proteases

Three families of proteases are known to depend directly on aspartic acid residues for catalysis: pepsin (A1), retropepsin (A2) and cauliflower mosaic virus peptidase (A3). A1 and A2 are homologous, and A3 shows some signs of relationship with A1 and A2, so they are all included in one clan (AA).

All aspartic acid proteases have similar three-dimensional structures. The atomic structure of chymosin is shown in Figure 1-14 (Gilliland, 1984). It contains one polypeptide chain of 323 residues folded into two homologous lobes (N- and C-domains). The folding pattern displays a pseudo 2-fold symmetry, an indication of gene

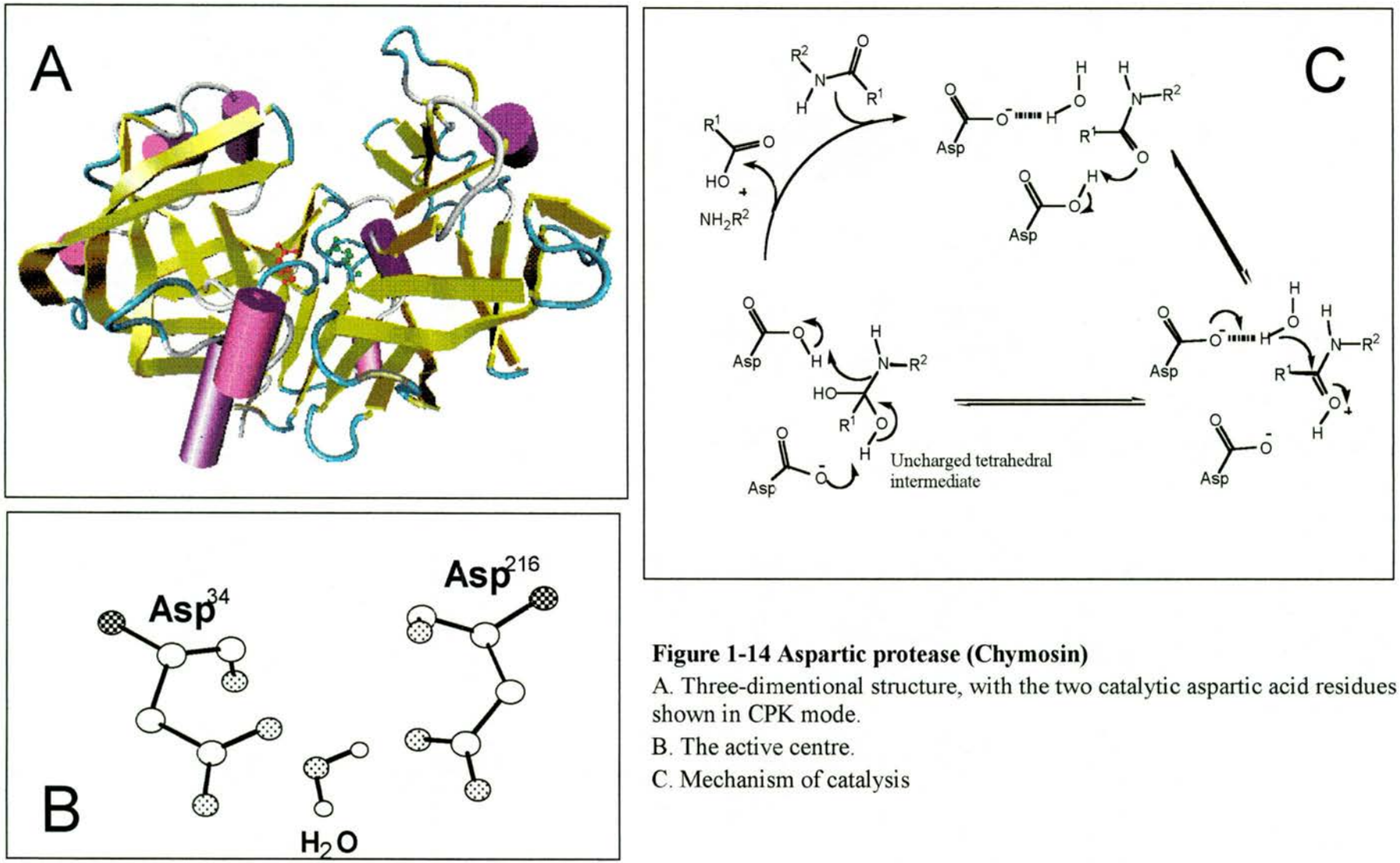


Figure 1-14 Aspartic protease (Chymosin)

A. Three-dimentional structure, with the two catalytic aspartic acid residues shown in CPK mode.

B. The active centre.

C. Mechanism of catalysis

duplication. Each lobe is primarily composed of parallel and antiparallel β-strands and a sort of α-helix connects the two lobes.

Each lobe contributes one of the two catalytic Asp residues (Asp34 and Asp216), which are located at the base of the deep cleft between the two lobes. The catalytic Asp residues reside in the motif Asp-Thr (or Ser)-Gly, which is conserved in all aspartic acid proteases. The side chain carboxylate groups are nearly co-planar and are orientated toward each other with a water molecule between them. Catalysis is carried out by a general acid-base mechanism (Figure 1-14). A molecule of water attacks the carbonyl carbon of the scissile peptide bond with the assistance from the active site carboxylates in proton transfer. A tetrahedral intermediate is formed, but no covalent bond forms between the substrate and enzyme at any stage.

Aspartic proteases homologous to pepsin are specifically inhibited by pepstatin, a reversible inhibitor that seldom affects other type of proteases. Although many aspartic proteases have an acidic pH optimum, some do not (e.g. HIV-2 PR). On the other hand, some serine or cysteine proteases show maximal activity at acidic pH. An acidic pH optimum, therefore, does not necessarily mean aspartic-type (Rawlings, 1995).

1.4.1.4 Metalloproteases

About 30 families of proteases have been recognized as metalloproteases, which have a catalytically essential metal atom at the active centre. Half of the families consists of enzymes containing the His-Glu-Xaa-Xaa-His motif present in a helix. The two His residues form two ligands of the metal ion (usually Zn^{2+}). The third ligand can be a Glu (in clan MA) or His (in clan MB); while the fourth ligand is an activated molecule of

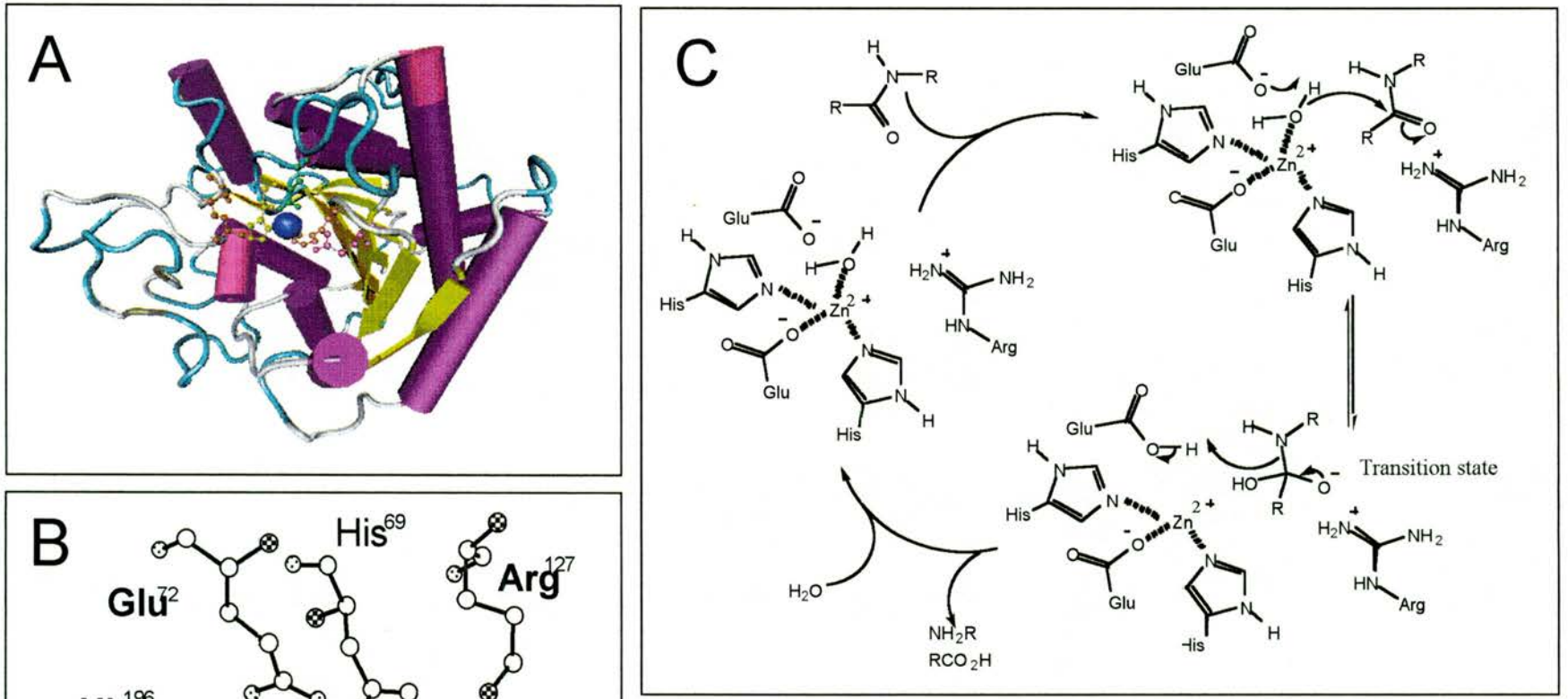


Figure 1-15 Metallo-protease (Caboxypeptidase A)

A. Three-dimensional structure of Caboxypeptidase A, 5 residues and the zinc ion involved in catalysis are shown in CPK and VDW modes, respectively.

B. The active centre.

C. Mechanism of catalysis.

A family. Although the three-dimensional structures of metalloproteases show great diversity, their catalytic centres are in a similar tetrahedral geometry (Kester and Matthews, 1977).

The carboxypeptidase A (CPA) is one of the best-studied metalloprotease (Rees *et al.* 1983). The ellipsoidal molecule has the basic architecture of a three layer alpha-beta-alpha sandwich with a twisted β-sheet of 8 strands running through the centre of the molecule (Figure 1-15). Apart from the relatively short helices, the left half of the molecule is mainly random coil and quite flexible. The zinc is adjacent to the β-sheet and co-ordinated by His-69, Glu-72, and His-196. The fourth position is occupied by a water molecule, which is hydrogen-bonded with Glu-270.

The catalysis reaction is shown in Figure 1-15. Glu-270 promotes the nucleophilic attack on the carbonyl carbon of the scissile bond by the Zn^{2+}-bound water, forming a tetrahedral intermediate. The positively charged Arg-127 helps to accommodate the oxyanion. Subsequent decomposition of the tetrahedral intermediate generates the two products.

The most useful inhibitor to recognize metalloproteases is 1,10-phenanthroline, which has much higher affinity for zinc ($2.5 \times 10^6 M^{-1}$) than for calcium ($3.2\ M^{-1}$). It can be used in the presence of 10 mM Ca^{2+}, as is required for the stability or activity of many proteases. 1,7- and 4,7-phenanthroline are often used as negative controls to show that the inhibition is due to chelation of Zn^{2+} instead of non-specific binding of enzyme. EDTA is often used as well, but it should be noted that Ca^{2+} is also chelated (Barrett, 1994).

1.4.2 Introduction to Viral Proteases

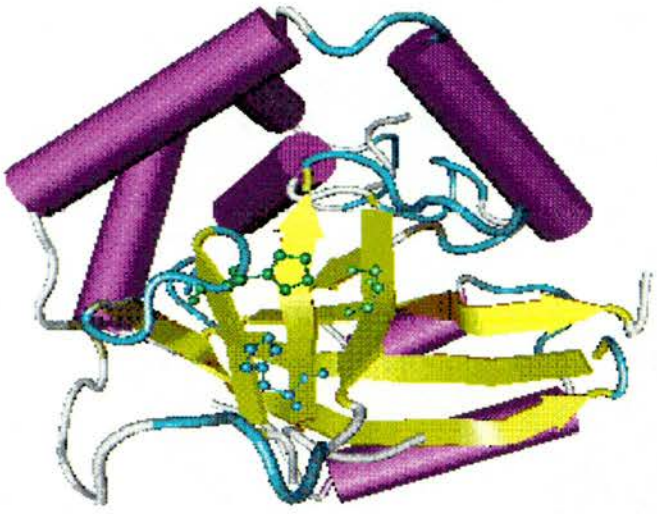
Viral proteins were first found to be generated by limited proteolysis from a polyprotein precursor of poliovirus more than 30 years ago (Summer and Maizel, 1968). That a viral protease was responsible for the processing was first demonstrated in EMCV some ten years later. Since then, viral proteases have been identified in a large number of virus families and groups. Several other virus families are predicted to encode their own proteases, though the proteolytic activities are still to be proven (Table 1-2).

It appears that viral proteases are particularly abundant in the single-stranded positive sense RNA viruses. Viruses of this type start their genome expression with translation rather than transcription, and have developed a wide range of mechanisms to control their genome expression co- and post-translationally. Limited proteolysis by viral proteases is a widely utilized strategy. In the following sections, proteases of single-stranded (+) RNA viruses will be discussed in more details, though the discussion is representative rather than exhaustive: only those best-characterized proteases are chosen.

Most viral proteases, if not all, have been shown to be essential for the life cycle of the viruses. Therefore, viral proteases become the " Achilles' heel " of many viruses. One magnificent example is the protease of HIV, which belongs to the retrovirus family. The properties of HIV protease are also described in this review, but the strategies used to design its inhibitors are omitted.

Proteases are also encoded in DNA viruses. Recently, striking progresses have been made in the studies of two DNA viral proteases: human cytomeglovirus protease and adenovirus protease (AVP) (Figure 1-16). The cytomeglovirus protease is involved in

A



B

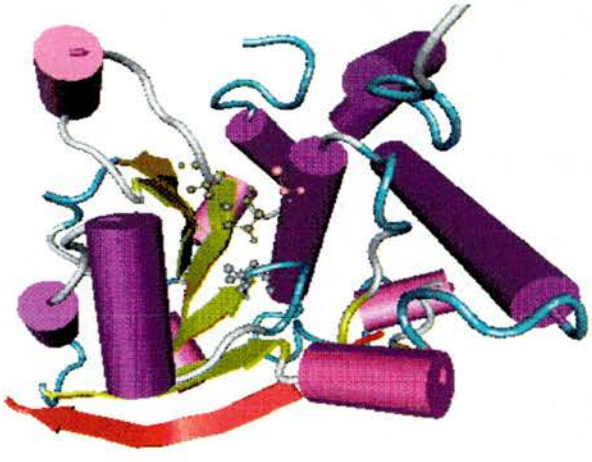


Figure 1-16 Structures of proteases of two DNA viruses

A. Human cytomegalovirus (CMV) protease, catalytic residues (His-157, His-63 and Ser-132) are shown in CPK mode.

B. Adenovirus protease, together with the 11-amino acid peptide cofactor (pVIc, shown in red). Active-site residues (Glu-71, His-54 and Cys-122) are shown in CPK view.

the processing of scaffold proteins during capsid assembling. It has been shown to be a serine protease. The resolution of its atomic structure revealed a unique fold and a novel active centre (Qiu *et al.*, 1996). The seven-stranded β-barrel is surrounded by seven α-helices, and the catalytic triad comprises His/His/Ser, different from the two six-stranded β-barrels and Asp/His/Ser triad of chymotrypsin. The 23 kDa AVP is responsible for the processing of the core proteins within an assembled virion.

The protease has a Glu/His/Cys catalytic triad in a geometry similar to that of papain (Asn/His/Cys). This is remarkable, because all other known papain-like viral cysteine proteases have a catalytic dyad of His/Cys. Seven α-helices and one five-stranded β-sheet form a unique sandwich architecture. An 11-amino-acid peptide can stimulate proteolytic activity more than 300-fold. This cofactor comes from the C terminus of viral protein pVI (a substrate of AVP) and functions as the sixth β-strand of the core β-sheet. The binding of viral DNA can enhance the activity by another 20-fold (Ding *et al.*, 1996).

Table 2 List of Viral Proteases

Virus family	Virus representative	Protease	Catalytic type	Protein folding
Single-stranded (+) RNA viruses				
Picornaviridae				
Enterovirus	PV	2A	cysteine	CHL
		3C	cysteine	CHL
Rhinovirus	HRV	2A	cysteine	CHL
		3C	cysteine	CHL
Cardiovirus	EMCV	3C	cysteine	CHL
Aphthovirus	FDMV	L	cysteine	PL
		3C	cysteine	CHL
Hepatovirus	HAV	3C	cysteine	CHL
Potyviridae				
Potyvirus	TEV	NIa	cysteine	CHL
		HC-pro	cysteine	PL
		P1	serine	CHL
Bymovirus	BaYMV	NIa	cysteine	CHL
		p28?	cysteine?	PL?
Comoviridae				
Comovirus	CPMV	p24	cysteine	CHL
Nepovirus	GFLV	p23	cysteine	CHL
Flaviviridae				
Flavivirus	YF	NS3	serine	CHL
Pestivirus	BVDV	NS3	serine	CHL
		Npro	cysteine	PL

Table 2 List of Viral Protease (continue)

Hepatitis C-like viruses	HCV	NS3	serine	CHL
		NS2-3	Zn^{2+}?	
Togaviridae				
Alphavirus	SIN	nsP2	cysteine	PL
		C	serine	CHL
	SFV	nsP2	cysteine	PL
		nsP4	aspartic	
		C	serine	CHL
Rubivirus	RuV	p150	cysteine	PL
Caliciviridae	RHDV	3CL	cysteine	CHL
Coronaviridae	MHV	3CL	cysteine	CHL
		PL1	cysteine	PL
		PL2?	cysteine?	PL?
Arteriviridae	PRRSV	nsp1α	cysteine	PL
		nsp1β	cysteine	PL
		nsp2	cysteine?	
	EAV	nsp2	cysteine?	
Tymovirus	TYMV	p150	cysteine	PL
Capillovirus	AGSV	p24?	cysteine	PL
		p36?	serine	CHL
Carlavirus	BBScV	p166	cysteine	PL
	ASPV	?	cysteine?	
Closteroviridae				
Closterovirus	BYV	p65	cysteine	PL
	CTV	PL^{pro}?	cysteine?	

Table 2 List of Viral Protease (continue)

Furovirus	TYMV	p150?	cysteine?	PL?
Luteoviridae				
Polerovirus	PLRV	P1?	serine?	CHL?
Enamovirus	PEMV-1	P1?	serine?	CHL?
Sobemovirus	SBMV	P2?		
Sequiviridae	RTSV	?	cysteine?	CHL?
	PYFV	?	cysteine?	CHL?
Trichovirus	APCLV	?	cysteine?	
Retroid viruses				
Retroviridae	HIV-1	PR	aspartic	
Caulimoviridae	CaMV	p20	aspartic	
Badnavirus	RTBV	PR?	aspartic?	
Double-stranded DNA				
Adenoviridae	Ad	AVP	cysteine	novel
Herpesviridae	CMV*	PR	serine	novel
Double-stranded RNA				
Birnaviridae	IBDV	VP4	?	
Hypoviridae	HyAV	p29	cysteine	PL
		p48	cysteine	PL

1. Uncertainty is indicated by ?
2. Abbreviations: ASGV, apple stem grooving virus; ASPV, apple stem pitting virus; BYV, beet yellows virus; CMV, cytomeglovirus (* in plant virology, CMV refers to cucumber mosaic virus, which is not discussed in this thesis); CTV, citrus tristeza virus; PYFV, parsnip yellow fleck virus; RTBV, rice tungro bacilliform virus; RTSV, rice tungro spherical virus. For other abbreviations, see the list on page viii.
3. CHL, chymotrypsin-like; PL, papain-like.

1.4.3 Viral Chymotrypsin-like Serine Proteases

1.4.3.1 Sinbis virus capsid protease

As mention in Section 1.3.6, Sinbis virus structural proteins are first translated from a subgenomic RNA (26S) as a polyprotein (p130). The N-terminal capsid protein releases itself from the nascent polyprotein autocatalytically. Sequence alignment suggested that the capsid protein is a chymotrypsin-like protease (Hahn *et al.*, 1985). Site-directed mutagenesis confirmed that His-141 and Ser-215 are part of the catalytic triad (Melancon and Garoff, 1987; Hahn and Strauss, 1990), while the third one (Asp-163) was identified after the resolution of the crystal structure (Figure 1-17) (Choi *et al.*, 1991; Tong *et al.*, 1993). Interestingly, when Ser-215 was mutated to Cys, the enzyme retained 60% of the proteolytic activity *in vitro*, but transfection of RNA transcripts containing this mutation did not produce infectious virus particles (Hahn and Strauss, 1990).

Sinbis virus capsid protein only cleaves at a conserved Trp-264↓Ser-265 site at its own C terminus. The released capsid protein shows no proteolytic activity. The crystal structure of purified capsid protein shows the reason for this: the C terminal Trp-264 remains at the active centre and the essential Asp-163 is exposed to the solvent, instead of being buried in the cleft of the active centre, like in other chymotrypsin-like proteases. This presumably inactivates the protease effectively.

1.4.3.2 Flavi-, Pesti- and Hepatitis C virus NS3 proteases

Sequence alignment showed that the N-terminal part (about 180 aa) of the NS3 proteins

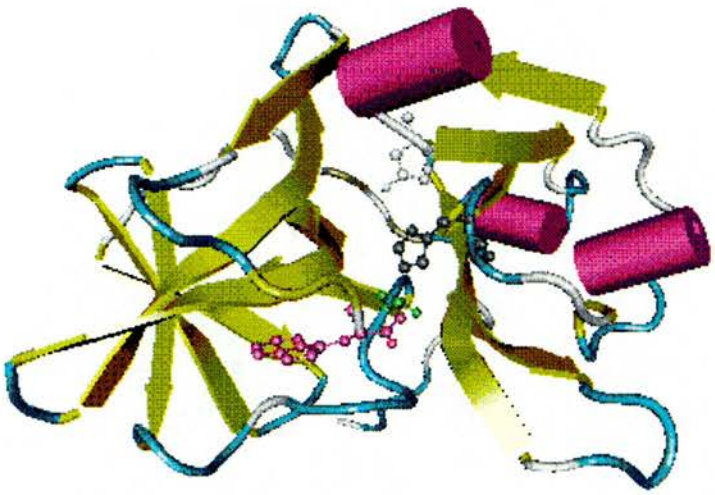


Figure 1-17 Sindbis virus capsid protein monomer

This chymotrypsin-like serine protease cleaves at its own C terminus in a *cis* fashion. The liberated C terminal Trp-264 remains at the active centre (shown in purple and CPK mode), resulting in the inactivation of the protease. The catalytic triad, His-141, Asp-163, and Ser-215, are also shown in CPK mode (PDB, 1KXA).

of flavi-, pesti- and Hepatitis C viruses contains a trysin-like serine protease domain (Bazan and Fletterick, 1989; Gorbalenya *et al.*, 1989; Miller and Purcell, 1990). The remaining sequences contain typical NTPase/helicase, whose activities has been confirmed experimentally (Wengler, 1991; Warremer and Collett, 1995; Kim *et al.*, 1997).

The catalytic triad of NS3 of yellow fever virus (YF) was determined to be His-53, Asp-77 and Ser-138 (Chambers *et al.*, 1990). The sequence around the catalytic serine , GXSGXP, is conserved among flaviviruses.

The NS3 protease has a preference to cleave after two basic residues and before a residue with a small side chain. In the YF polyprotein, NS3 cleaves at the C/anchor, 2A/2B, 2B/3, 3/4A, 4A/2K, and 4B/5 junctions, which follow the pattern of (R, K, or Q)-(R, or K)↓(G, S, A, or T). It has been proposed that the conserved Asp-131 and motif GLYGNG (residues 151 to 156) in the substrate-binding pocket are important to cleavage site specificity. In infected cells, cleavages at 2A/2B, 2B/3, 3/4A, and 4B/5 junctions occur rapidly and efficiently (Chambers *et al.*, 1990). *In vitro* experiments showed that cleavages at 2A/2B and 2B/3 were apparently in a *cis* fashion.

Remarkably, the flavivirus NS3 protease needs **NS2B** protein as a cofactor, which can be supplied *in cis* or *in trans* (Chambers *et al.*, 1991; Falgout *et al.*, 1991, 1993; Cahour *et al.*, 1992; Arias *et al.*, 1993). The NS3 protease, though highly hydrophilic, is localized to membrane structures (Wengler *et al.*, 1990), presumably via association with hydrophobic NS2B protein (Arias *et al.*, 1993; Chambers *et al.*, 1993). In contrast to flaviviruses, pestivirus BVDV $NS3^{pro}$ needs another cofactor **NS4A** to cleave at 5A/5B site. The activity of the hepatitis C virus $NS3^{pro}$ is stimulated by up to 100-fold by **NS4A** sequences (Wiskerchen and Collett, 1991; Failla *et al.*, 1994, 1995;

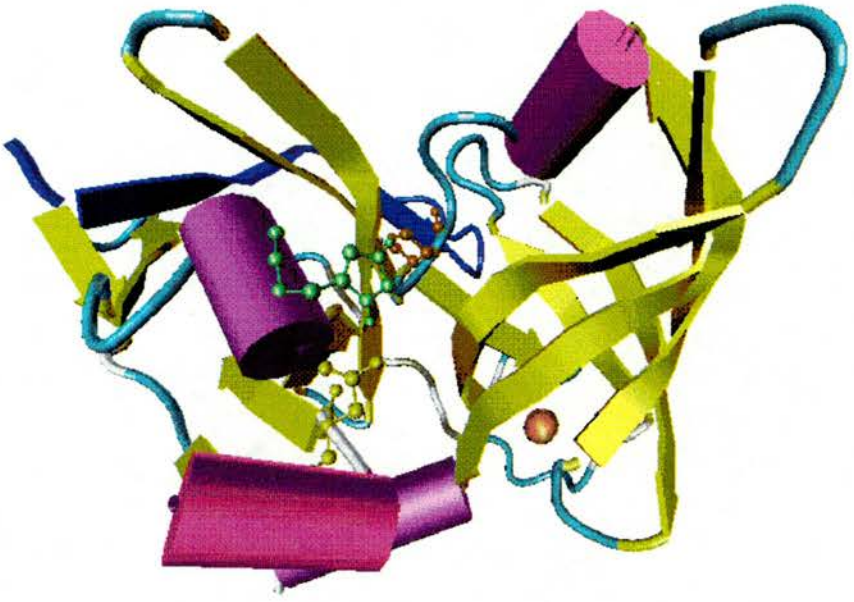


Figure 1-18 Structure of the NS3:NS4A complex of HCV

The synthetic NS4A activator peptide is shown in dark blue, while the zinc ion in brown. The active-site residues, His-83, Asp-107 and Ser-165, are displayed in CPK mode.

Bartenschlager *et al.*, 1995; Lin and Rice, 1995; Lin *et al.*, 1995; Shimizu *et al.*, 1996; Steinkuhler *et al.*, 1996). Steady-state kinetic analysis has shown that the HCV $NS3^{pro}$ binds to the activating NS4A first, then the substrate (Landro *et al.*, 1997).

In the case of HCV, the N-terminal 20 residues of NS4A are hydrophobic and putatively target the $NS3^{pro}$ to membranes via interaction between $NS3^{pro}$ and downstream NS4A sequences. Crystal structure of the hepatitis C virus $NS3^{pro}$ complexed with a synthetic NS4A peptide showed that 17 out of the 19 residues of NS4A (Gly21-Pro39) form hydrogen bonds with $NS3^{pro}$, in addition to the hydrophobic interaction between side chains of NS4A and N-terminal region of $NS3^{pro}$ (Kim *et al.*, 1996) (Figure 1-18).

The cleavage sites of HCV $NS3^{pro}$ conform to the conserved sequence: (Cys or Thr)↓(Ser or Ala). The Phe-154 at the S1 subsite in the substrate-binding pocket is proposed to interact with the hydroxyl or sulphydryl groups of the P1 (Thr or Cys) residue.

The HCV NS3 protease has a Zn^{2+} binding site composed of Cys-97, Cys-99, Cys-145, and a water molecule hydrogen-bonded to His-149. Since this site is far away from the catalytic centre at the middle cleft, the bound Zn^{2+} ion is likely to play a structural role rather than participating the catalysis of the serine protease.

1.4.3.3 Potyvirus P1 protease

In TEV, the N-terminal 35 kDa protein P1 has a protease domain (~148 aa), which mediates the release of P1 from the 87 kDa precursor in a *cis* fashion (See Section 1.3.4). It is predicted that the cleavage occurs between Tyr-304 and Ser-305, and the catalytic triad is composed of His-215, Asp-224, and Ser-256. The putative nucleophile

Ser is present in the Gly-Xaa-**Ser**-Gly motif, which frequently occurs around the catalytic Ser of cellular serine proteases. Site-directed mutagenesis has confirmed the essential role of His-215 and Ser-257 in the proteolytic activity (Verchot *et al.*, 1991).

1.4.4 Viral chymotrypsin-like cysteine proteases

1.4.4.1 Entero- and rhinovirus 2A proteases

The 2A proteases of entero- and rhinoviruses cleave two sites in the viral polyproteins. One primary cleavage (1D/2A) separates the capsid protein precursor from nascent non-structural precursor in a *cis* fashion (Toyoda *et al.*, 1986; Sommergruber *et al.*, 1989). Another site (within the 3D sequence) yields the alternative products, 3C' and 3D'. The latter cleavage is observed only in some entero- and rhinoviruses and appears not essential for virus viability (Lee and Wimmer, 1988).

Notably, the 2A proteases also induce the cleavage of the eIF4G subunit of the cap-binding protein complex eIF4F, resulting in the inhibition of host cellular mRNA translation (Bernstein *et al.*, 1985; Lloyd *et al.*, 1988). Some reports (Krausslich *et al.*,1987; Lloyd *et al.*, 1986; Wyckoff *et al.*, 1990) suggested that polioviral 2A protease activated a cellular protease, which carried out the cleavage, but other experiments showed that $2A^{pro}$ of rhinovirus and coxsackievirus B4 cleaved eIF4G directly (Lamphear *et al.*, 1993; Liebig *et al.*, 1993). Haghighat *et al* (1996) found that purified rhinoviral $2A^{pro}$ cleaved eIF4G-eIF4E efficiently, but did not cleave eIF4G alone. In addition, 2A of poliovirus, acting as a translational *trans*-activator, enhances the translation initiated from the IRES in the 5' non-coding region, when translation from host mRNA (cap-dependent) is not inhibited (Hambidge and Sarnow, 1992).

The poliovirus 2A protease has been shown to be inhibited by some thiol protease inhibitors (iodoacetamide and N-ethylmaleimide), but insensitive to inhibitors **E-64,** the specific inhibitor of papain (Konig and Rosenwirth, 1988). Site-directed mutagenesis studies have confirmed that $2A^{pro}$ contains a conserved catalytic triad (His-20, Asp-38, Cys-109, poliovirus numbering), similar to that of trypsin-like serine proteases. Sequence alignment and tertiary structure modeling have demonstrated that 2A is homologous to the subclass of small trypsin-like serine proteases (SGPA and SGPB). Identical residues were mapped to the interface between the two β-barrels, and found to be either catalytically important or structurally vital. Another unusual feature of 2A is that it contains an equimolar Zn^{2+}, which is essential for structural integrity. The Zn^{2+} binding site is on the side opposite to the catalytic triad, and inhibition studies have shown that the Zn^{2+} ion seems not involved in catalysis (Sommergruber *et al.*, 1994, 1997; Voss *et al.*, 1995).

P4, P2, P'1 and P'2 have been shown to be determinants of cleavage sites (Skern *et al.*, 1991; Hellen *et al.*, 1992; Sommergruber *et al.*, 1992). The 2A proteases do not have strict site specificity for *cis* cleavage, though a bulky residue (Val or Ile) at P'1 is not preferred (Skern *et al.*, 1991; Hellen *et al.*, 1992). Gly appears at the P'1 position of all 1D/2A junction in entero-/rhinoviruses. In contrast, mutation of the Thr at P2 position abolished the cleavages *in trans* completely, showing different specificity.

1.4.4.2 Picornavirus 3C proteases

3C, the major protease of all picornaviruses, is responsible for a series of cleavages in the structural and non-structural precursors. The 3C can exist in many forms, particularly 3C and 3CD, which show different site specificities in poliovirus. Both 3C and 3CD can process the non-structural precursor, but only 3CD can cleave both the

1AB(VP0)/1C and 1C/1D junctions (Ypma-Wong *et al.*, 1988; Ypma-Wong and Semler, 1987; Jore *et al.*, 1988). In addition to processing viral proteins, 3C also cleaves many host cellular proteins, including histone H3, TATA-binding protein, transcription factor IIIC, and microtubule-associated protein 4 (Fall *et al.*, 1990; Clark *et al.*, 1993; Das and Dasgupta, 1993; Joachims *et al.*, 1995). However, the significance of these cleavages has not been well established.

Early inhibition studies showed that picornavirus 3C was inhibited by both serine and thiol protease inhibitors (reviewed by Lawson and Semler, 1990). Sequence alignment and structural modeling, independently carried out by Gorbalenya *et al.* (1986, 1988 and 1989) and Bazan and Fletterick (1988), identified homology between picornavirus 3C and trypsin-like serine proteases. The 3C protease was predicted to assume a trypsin-like fold and have a catalytic triad of His-40, Asp-85 or Glu-71 (Asp71 for aphtho-, cardioviruses), and Cys-147 (poliovirus numbering), corresponding to His-57, Asp-102, and Ser-195 of trypsin. Site-directed mutagenesis experiments support the role of His-40, Glu-71 (Asp-71), and Cys-147 in mediating proteolytic cleavage (Ivanoff *et al.*, 1986; Cheah *et al.*, 1990; Hammerle *et al.*, 1991; Kean *et al.*, 1991, 1993; Grubman *et al.*, 1995).

The atomic structures of HRV-14 (Matthews *et al.*, 1994) and HAV (Allaire *et al.*, 1994) 3C proteases confirmed the trypsin-like fold, composed of two six-stranded β-barrels with a groove between them for substrate binding. In the case of HRV-14 3C protease (Figure 1-19), Cys-146, His-40, and Glu-71 form the catalytic triad, as predicted by Gorbalenya. Since the side chain of Glu-71 is longer than that of a canonical Asp, it is folded in such a way that the carboxylate group accepts a hydrogen bond from the Nδ1 of the catalytic His-40, using anti-oriented lone pair electrons, in contrast with Ser-His-Asp serine proteases, which use syn-oriented carboxyls. However,

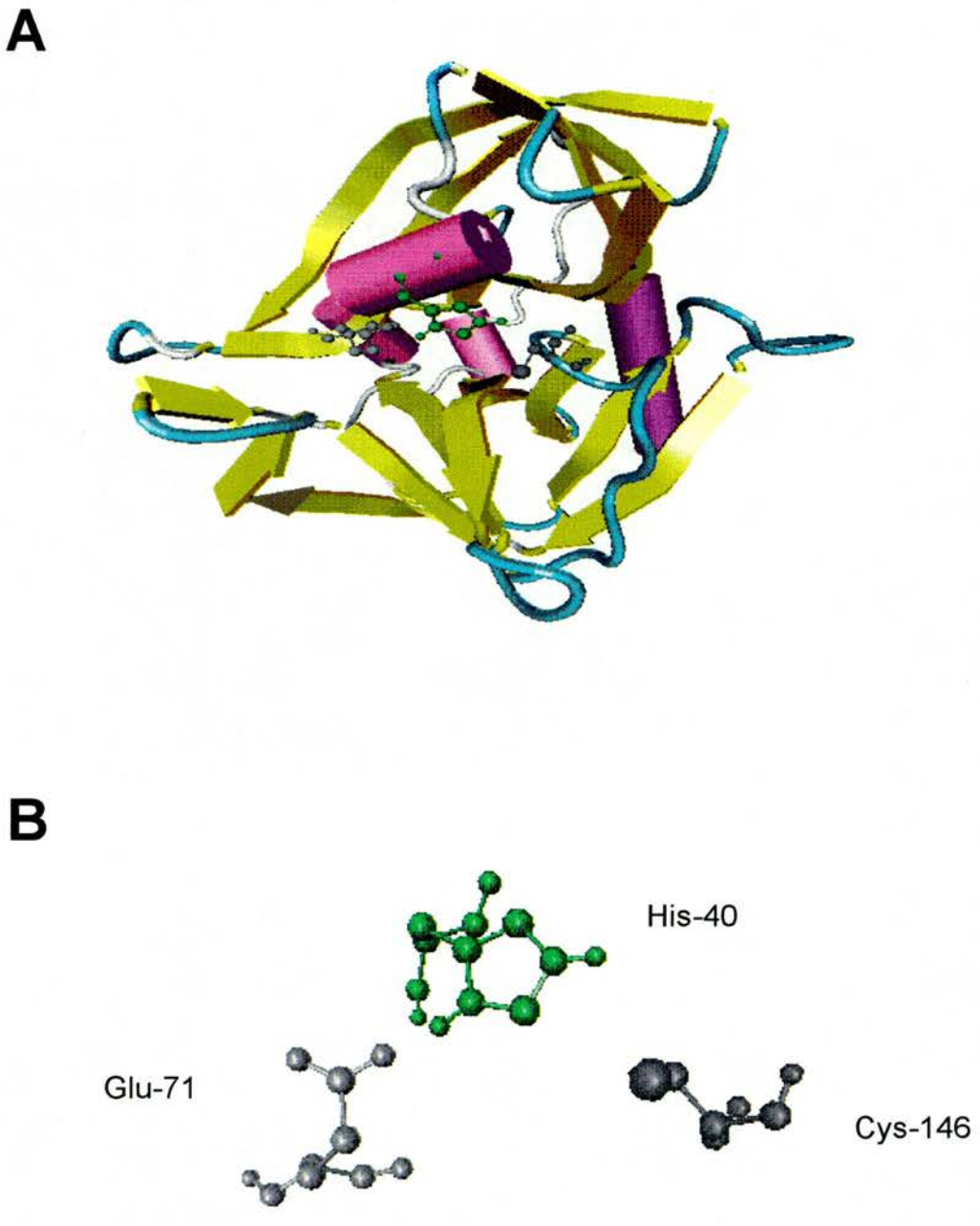


Figure1-19 The 3C protease of human rhinovirus-14

A. Structure of the 3C protease,which shows a typical bilobal structure with active centre in the middle groove. Each lobe folds into a 6-stranded beta-barrel.

B. The catalytic residues (His-40, Glu-71 and Cys-146) are shown in CPK mode.

the active site of HAV shows a significantly different geometry (Figure 1-20). In HAV, the Cys-172, His-44, and Asp-84 correspond to Ser-195, His-57, and Asp-102 of trypsin, but the side chain of Asp-84 turns 180° away from the imidazole ring of His-44, and is in strong interaction with the side chain of Lys-202 and main-chain amide of Asp-158. The authors proposed a dyad of Cys/His, instead of triad, for the catalytic apparatus of HAV 3C protease.

The picornavirus 3C proteases have much higher substrate specificity than cellular serine proteases (reviewed by Lawson and Semler, 1990; Seipelt *et al.*, 1999). Poliovirus 3C cleaves exclusively at Gln/Gly junctions in authentic viral polyproteins, and an Ala residue at P4 position is preferred. For HRV-14 $3C^{pro}$, the preferred substrate is P2(bulky)-Gln↓Gly-Pro (Long *et al.*, 1989). HAV $3C^{pro}$ has a prevalence for a large hydrophobic residue (Ile or Leu) at P4, a small hydrophilic residue (Ser or Thr) at P2, and a Gln at P1, but the requirement at P'1 is less restricted, with Met, Val, Ala, Gly, Ser and Thr all being accepted. Specific subsites at the binding pockets are believed to account for the specificities. The bulge in β-strand βI of HAV $3C^{pro}$ allows several carbonyl groups to accept hydrogen bond from substrate residue at P' positions, while in HRV-14 $3C^{pro}$, the lack of this bulge means Gly-Pro is preferred at P'1 and P'2, which induces a turn-away from the active site (Bergmann *et al.*, 1997; Matthews *et al.*, 1994).

Another important feature of picornavirus $3C^{pro}$ is the RNA binding site on the side opposite to the catalytic centre. Residues important for the binding were identified (Andino *et al.*, 1990a, b; 1993; Leong *et al.*, 1993; Walker *et al.*, 1995) and mapped to a region connecting the two lobes of the protease, lying between the two helices of the N and C termini (Figure 1-20). Although both 3C and 3CD are capable of RNA binding, 3CD shows higher affinity. 3C and 3CD have been shown to bind to the cloverleaf

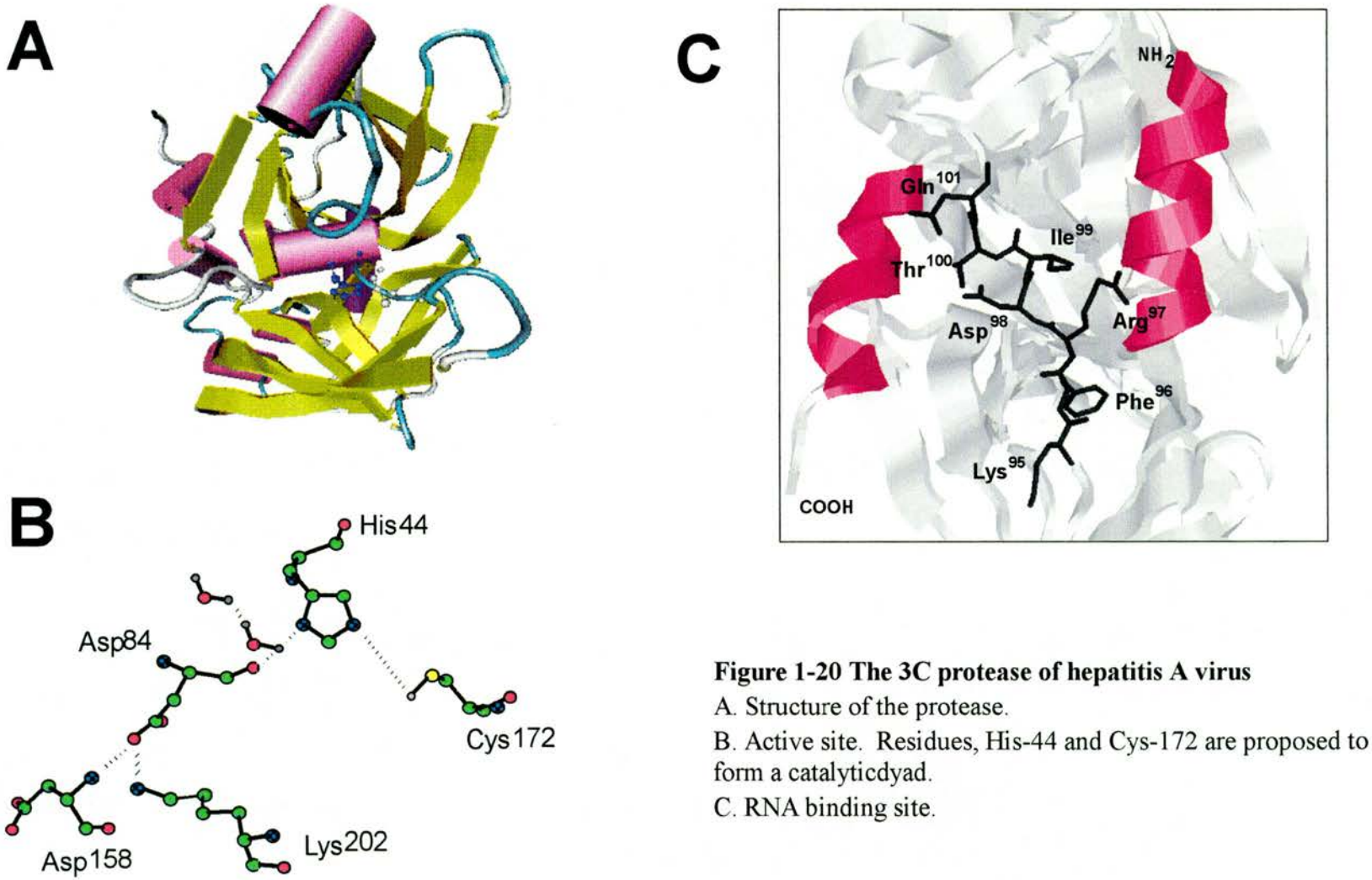


Figure 1-20 The 3C protease of hepatitis A virus

A. Structure of the protease.

B. Active site. Residues, His-44 and Cys-172 are proposed to form a catalyticdyad.

C. RNA binding site.

structure in the 5'NCR of viral RNA, forming a RNP complex essential for viral replication.

1.4.4.3 Potyvirus NIa protease

In TEV, the 27 kDa protease domain of NIa, equivalent to the picornaviral 3C protease, mediates a series of cleavages both *in cis* and *in trans* (Section 1.3.4). Like picornaviral proteases, TEV NIa is inhibited by Zn^{2+}, iodoacetamide, and N-ethylmaleimide, but not by most other protease inhibitors at their normal effective concentration. Mutagenesis work confirmed the hypothesis that His-46, Asp-81, and Cys-151 form the catalytic triad, whereas the Asp could be replaced by Glu, and Cys by Ser, with significant retention of proteolytic activity (Dougherty *et al.*, 1989b).

The cleavage site is defined by a consensus heptapeptide, Glu-Xaa-Ile (or Val, Leu)-Tyr-Xaa-Gln ↓ Ser (or Gly). Substitution of residues at P6, P4, P3, P1, and P'1 abolishes or greatly reduces proteolytic activity, while residues at P5 and P2 positions affect cleavage rates (Dougherty *et al.*, 1989a). Similar, but distinct, consensus sequences are found in the cleavage sites of other potyviruses (Ghabrial *et al.*, 1990). Using hybrid proteases containing both TEV and TVMV (another potyvirus) NIa sequences, three domains within the C-terminal 150 amino acids of the NIa protease were identified to contain sequences important to recognize a specific cleavage site (Parks and Dougherty, 1991). Particularly, His-167 was predicted to interact with the Gln at P1 position. Interestingly, this residue is present in many 3C proteases of viruses in the picornavirus superfamily, which often cleave at a Gln↓Gly (or Ser) site.

Effective protease concentration can be regulated by differential targeting of different precursor forms of proteases. In TEV-infected cells, two viral proteins, 49 kDa NIa and

58 kDa NIb (polymerase), both of which contain nuclear localization signals (NLSs), are transported to the nucleus and aggregate into crystal-like nuclear inclusion body. The 49 kDa NIa has an N-terminal VPg domain (21 kDa) and a C terminal protease domain (27 kDa). A proportion of NIa protein undergoes an inefficient self-cleavage between the VPg and protease domains *in vivo*. The free 27kDa NIa protease is localized in the cytoplasm. Interestingly, some of the 49 kDa NIa also exists as a larger 55 kDa precursor with a 6 kDa protein adjacent to its N-terminus. The membrane-binding signal in the 6 kDa protein can mask the NLS in the VPg domain of NIa, and target the protein to cytoplasmic membranes, where viral RNA replication is postulated to occur (Restrepo-Hartwig and Carrington, 1992).

1.4.4.4 Como- and nepovirus proteases

Como- and nepoviruses are two genera in the family *Comoviridae*, a member of the picornaviral superfamily. Both of them have a bipartite single-stranded RNA genome. The larger RNA molecule, RNA-1 in nepoviruses or B-RNA in comoviruses, encodes a 3C-like protease in a VPg-protease-polymerase cassette, which is the only virus-coded protease and mediates all the cleavages in the viral polyproteins.

The cowpea mosaic virus protease is a 24 kDa protein. Its inhibition profile is rather confusing (Pelham, 1979). Under the guidance of sequence alignment and structural modeling, site-directed mutagenesis has confirmed that the catalytic triad is composed of His-40, Glu-76, and Cys-166. Replacement of the nucleophilic Cys-166 by Ser causes a 10-fold decrease of the proteolytic activity (Dessens and Lomonossoff, 1991). Similarly, the catalytic triad of nepovirus GFLV 24 kDa protease was identified to be His-43, Glu-87, and Cys-179; but the Cys to Ser mutation generates a mutant protease with comparable proteolytic activity (Margis and Pinck, 1992).

Proteases of some nepoviruses (subgroup a and b) have a distinct substrate specificity. The 3C-like proteases of picornavirus, potyvirus, and comovirus have a conserved His 12-18 amino acid downstream of the nucleophilic Cys, which is believed to be a key residue in the interaction of S1 binding site with the Gln at the P1 position. In the 24 kDa protease of grapevine fanleaf nepovirus, the conserved histidine is replaced by a leucine; meanwhile this protease often cleaves after a Cys or Arg, instead of a Gln (Serghini *et al.*, 1990; Pinck *et al.*, 1991; Margis *et al.*, 1993). The 27 kDa 3C-like protease of tomato ringspot nepovirus (ToRSV, a member of nepovirus subgroup c), which cleaves at Gln↓Gly and Gln↓Ser pairs, also has a His residue at the S1 substrate-binding site (Carrier *et al.*, 1999).

The proteolytic activity of the 24 kDa protease domain of CPMV seems to be regulated by a 32 kDa protein, residing at the N-terminus of the B-RNA-encoded polyprotein. The 32 kDa protein remains associated with the C-terminal 170 kDa polyprotein after they are separated in a co-translational, *cis* reaction, and further processing on the 170 kDa polyprotein is slowed down. On the other hand, the 32 kDa protein is essential for the 24 kDa protease to cleave the Gln↓Met site between the movement protein and capsid precursor encoded by the M-RNA (Peters *et al.*, 1992).

1.4.5 Viral papain-like cysteine proteases

1.4.5.1 Alphavirus nsP2 proteases

The non-structural proteins of alphaviruses are translated from the genomic RNA as one or two polyproteins, which are processed by the nsP2 protease located within the polyproteins. For Sindbis virus, nsP2 carries out all the 3 cleavage events, while in the

case of Semliki Forest virus, an aspartic acid-type proteolytic activity mediates a *cis* cleavage at the nsP3/nsP4 site (Section 1.3.6).

Homology studies have suggested that nsP2 protease domain has an overall structure similar to that of a cellular cysteine protease papain (Hardy and Strauss, 1989; Gorbalenya *et al.*, 1991). Cys-481 and His-558 (Sindbis virus numbering) were predicted to form the catalytic dyad, and these two residues were confirmed to be crucial for proteolytic activity by site-directed mutagenesis (Strauss *et al.*, 1992). Unlike many cysteine proteases, which have an Asn residue participating in catalysis, nsP2 doesn't contain an Asn residue essential for activity. Interestingly, Trp-559, the residue immediately C-terminal to catalytic His-558, has been shown to be absolutely required for proteolytic activity, and is proposed to have a similar role to the Asn in other cysteine proteases in maintaining the conformation of catalytic residues.

The nsP2 of alphaviruses is about 90 kDa, and probably a polyprotein itself. A helicase domain is contained in the N-terminal region, whilst a protease domain is present in the central part. Compared with other cysteine proteases, the nsP2 has an extra domain of about 210 amino acids at its C-terminus. Mutants with deletions in this region are deficient in processing (Hardy and Strauss 1989). It is possible that this region co-ordinates the interaction of the protease with its substrates, like the 3D part in the 3CD protease in processing structural precursor of poliovirus.

Sequences around the cleavage sites are recognized to be highly conserved among alphaviruses. A typical sequence, (Ala, Val or Ile)-Gly-(Ala or Gly) ↓(Ala, Gly or Tyr), has been identified (Shirako and Strauss, 1990). Sites at the same place of different alphaviruses show even higher conservation than different sites of the same virus (ten Dam *et al.*, 1999); correspondingly, each site shows distinct kinetics and particular

requirement of nsP2-containing form for processing. A delicate model of virus replication has been developed over several years (Section 1.3.6). In this model, nsP3/nsP4 junction is cleaved rapidly in a *cis* fashion, forming an efficient minus-strand initiation complex of P123/nsP4; cleavage at nsP1/nsP2 follows, but subsequent cleavage at nsP2/nsP3 will trigger the switch from minus-strand synthesis to plus-strand production. This model is consistent with all the following observations (de Groot *et al.*, 1990):

(a) All nsP2-containing proteins are proteolytically active, and able to cleavage the nsP1/nsP2 site, *in trans*.
(b) nsP2/nsP3 site cannot be cleaved by nsP1-containing forms
(c) nsP3/nsP4 can only be cleaved by nsP3-containing forms

1.4.5.2 Aphthovirus L protease

Foot-and-mouth disease virus and equine rhinovirus (ERV) 1 and 2 encode a protease at the N-terminus of the polyprotein (leader protease or L^{pro}) (Strebel and Beck, 1986). This protease releases itself from the rest of the polyprotein in a co-translational event. At least three forms of L protein are found in infected cells: Lab (20 kDa) and Lb (16 kDa) arise from alternative initiation of translation at two in-frame start codons, whilst Lb' is derived from Lb via a post-translational modification by a carboxypeptidase activity (Sangar *et al.*, 1988). Both Lab and Lb exhibit proteolytic activity at the L/P1 junction *in cis* and *in trans* (Medina *et al.*, 1993; Cao *et al.*, 1995).

The L protease has been suggested to be a thiol protease of the papain type (Gorbalenya *et al.*, 1991). Although sequence similarity between the L protease and papain is low and predominantly present around the nucleophilic Cys residue, their tertiary structures

are closely related: both contain an α-helical domain and a mainly β-sheet domain with an averaged rms of 1.3 Å for 76 residues (Guarne *et al.*, 1998).

Site-directed mutagenesis has shown that Cys-51 and His-148 are active-site residues of FMDV (Piccone, 1995; Roberts 1995). ERV-1 also has the conserved Cys-51 and His-148, but the candidate histidine residue of ERV-2 is misaligned by 9 residues (Ryan and Flint, 1997). Like papain, L^{pro} has the catalytic Cys-51 at the top of the central α-helix, opposing the His-148 on the other domain (Guarne *et al.*, 1998). Residues Glu-76 and Asp-164 were suggested to participate substrate binding, probably by accommodating basic residues at substrate cleavage sites.

Another function of the L protease is shutting down the host cap-dependent translation by cleaving the eukaryotic initiation factor eIF4G, a role played by the 2A protease in entero- and rhinoviruses. However, these two proteases cleave at different sites in eIF4G: L^{pro} cleaves at Gly479↓Arg480, while $2A^{pro}$ cleaves at Arg486↓Gly487 (Kirchweger *et al.*, 1994).

1.4.5.3 Potyvirus HC-protease

In addition to NIa and P1 proteases, the third proteolytic activity encoded in the potyviral polyprotein is the helper component-protease (HC-pro). The HC-pro of TEV is a ~51 kDa multifunctional protein. The HC activity is implicated in aphid transmission, while the proteolytic activity , present in the C-terminal 154 amino acids is responsible for only one co-translational, *cis* cleavage event, resulting in the release of a 87 kDa N-terminal precursor, which is subsequently processed by another viral protease P1 (Figure 1-5). No *trans* cleavage has been observed for this protease (Carrington, 1989).

Site-directed mutagenesis studies identified that Cys-649 and His-772 were essential for the proteolytic activity of HC-pro, consistent with the hypothesis that HC-pro is a papain-like cysteine protease (Oh 1989).

The HC-pro shows a highly strict requirement for specific amino acids at its cleavage site. Autocatalytic processing was assayed using HC-pro polyprotein containing altered cleavage sites at its C-terminus. The results showed strong preference for Tyr at P4, Leu (or Val) at P2, and Gly at P1 and P'1 positions (Carrington 1992).

1.4.6 Retrovirus: aspartic proteases

In retroviruses, the processing of *gag* and *pol* precursors is mediated by a viral protease (PR) just before or after budding. Viruses deficient in PR are able to assemble into immature, non-infectious particles.

Retroviral proteases were first related to microbial and mammalian aspartic proteases by sequence alignment and structural modeling, which identified the conserved Asp-[Thr/Ser]-Gly motif in the active site of both viral and cellular enzymes (Power *et al.*, 1986; Pearl and Taylor, 1987). High-resolution crystal structures are available for proteases of SIV, HIV-1 and 2, and avian sarcoma-leukemia virus (ASLV), which show similar features. The active enzyme of retrovirus PRs is a homodimer and has an overall structure similar to that of cellular aspartic proteases (Figure1-21). Each monomer is equivalent to one domain of the bilobal cellular enzyme: contains one copy of Asp-Thr-Gly motif and contributes one catalytic Asp residue. Besides, two large flaps, one from each monomer, cover the entrance to the binding cleft, and are presumably involved in interaction with substrates and inhibitors. In monomeric aspartic proteases like pepsin, there is only one N-terminal flap.

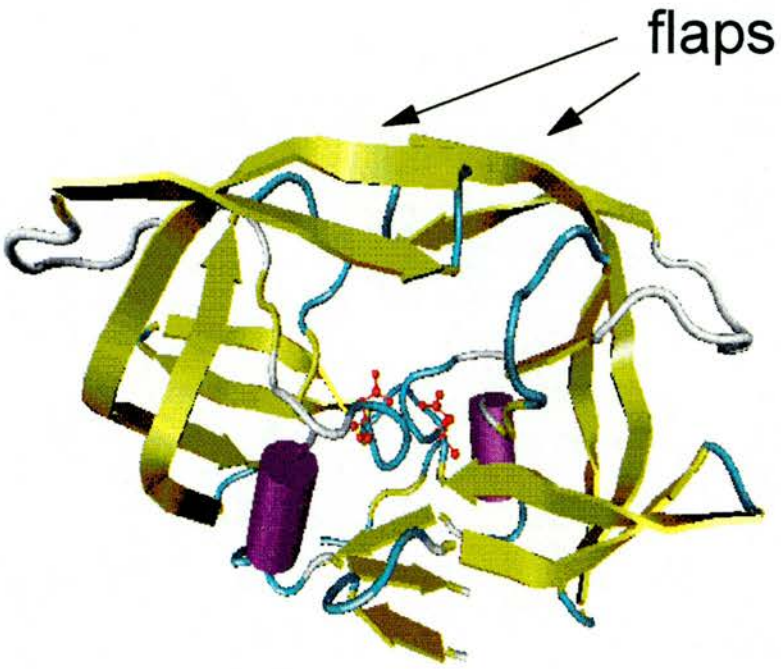


Figure1-21 Structure of HIV-1 protease

The active protease is a homodimer of the 99 amino acid polypeptide chain. The two catalytic Asp-25s are shown in red, one from each chain. The two flexible 'flaps' covering the substrate binding site and the active centre are also indicated.

The mature form of HIV protease has 99 residues. Complete mutagenesis of each residue has been carried out and proteolytic activity of each mutant protease has been assessed in bacterial expression system (Loeb *et al.*, 1989). At least three regions were shown to be sensitive to mutation: one at the flap region and two at/near the active centre topologically.

The retroviral proteases are specific for protein substrates. The cleavage sites in the gag and gag-pol polyproteins of HIV-1 and 2 have been identified (Tomaszek *et al.*, 1992; Freed, 1998). It is difficult to find a consensus sequence of amino acids around the scissile bond, but some preference toward certain amino acids has been observed. The HIV protease cleaves two type of junctions: between hydrophobic-hydrophobic (Leu-Ala, Met-Met, Phe-Leu, Leu-Phe) or aromatic (Tyr or Phe)-Pro residue pairs (Debouck *et al.*, 1987; Mizrahi *et al.*, 1989). The binding site of the enzyme is extended and interactions with residues in positions from P3 to P'3 are important (Grinde *et al.*, 1992; Cameron *et al.*, 1993). There is a strong preference for Glu or Gln at P'2; whilst some amino acids appear to be forbidden for certain positions, e.g. no Lys from P2 to P'2, no aromatic residues at P2 or P'2, and no β-branched residue at P1. Synthetic peptides that mimic the cleavage site in viral polyproteins have been shown to be suitable substrate for kinetic studies (Tozser *et al.*, 1991; Kay and Dunn, 1990) and inhibitor designing (Roberts *et al.*, 1990; Meek *et al.*, 1990; McQuade *et al.*, 1990). Other studies using viral polyprotein with altered residues proximal and distal to the known scissile bond revealed the importance of the context of the cleavage sites. The three cleavage sites in the HIV pol region (p6/PR, PR/RT, and RT/IN) have different levels of flexibility in amino acid substitutions at P1 and P'1 (Loeb *et al.*, 1989).

The maturation cleavage mediated by retroviral proteases takes place late in the life cycle. Premature intracellular cleavage of *gag* would prevent assembly and budding (Karacostas *et al.*, 1993; Krausslich, 1992). It is essential to keep the protease inactive before the assembly starts. Polyprotein precursors containing HIV PR undergo autoprocessing when expressed in bacteria (Mous *et al.*, 1988), and are capable of spontaneous processing *in vitro* (Louis *et al.*, 1991, 1994), suggesting that no cellular protease is required for the activation of HIV protease. In infected cells, the concentration of the *gag-pol* precursor might be so low that dimerization , which is necessary for the protease activity, is kept to minimum level until the polyprotein reaches assembly sites on the plasma membrane. In freshly budded virions, the protease releases itself from the precursor in either a *cis* or a *trans* reaction, first the N-terminus then the C-terminus. After this initiation processing, the released PR will cleave out more protease, and the processing of the whole capsid is carried out rapidly (Co *et al.*, 1994; Luciw, 1996).

1.4.7 HCV NS2-3 protease: a putative metalloprotease

Processing of the non-structural precursor of HCV is carried out by two viral proteases, encoded in the NS2-3 region. As discussed before, the serine protease NS3 mediates the cleavage at four downstream sites: NS3/NS4A, NS4A/NS4B, NS4B/NS5A, and NS5A/NS5B. Cleavage at NS2/NS3 site requires both NS2 and NS3 sequences and occurs in an intramolecular fashion. Interestingly, this cleavage is not abolished by mutation of the nucleophilic Ser-139 of the NS3 serine protease, which blocks cleavage at known NS3 sites - suggesting a second proteolytic activity, referred to as NS2-3 protease, which might be a metalloprotease. The metalloprotease hypothesis is consistent with other observations. *In vitro* transcription/translation studies showed that NS2-3 protease was inhibited by EDTA and phenanthroline, and stimulated by

exogenous Zn^{2+} (Grakoui *et al.*, 1993; Hijikata *et al.*, 1993; Reed *et al.*, 1995; for a review, see Wu *et al.*, 1998).

The resolution of the atomic structure of HCV NS3 protein revealed that a Zn^{2+} ion was co-ordinated by three cysteine residues and a molecule of water hydrogen-bonded to His-1175 (Kim *et al.*, 1996; Love *et al.*, 1996). Since a Zn^{2+} ion coordinated to three amino acid residues and an activated water is generally accepted as a catalytic zinc, and the Zn^{2+} binding site is close to the N-terminus of NS3 protein (within 11Å), the Zn^{2+} might play a direct role in the catalysis of the cleavage at NS2/NS3 junction. Mutation of the three cysteine residues at the Zn^{2+} binding site (Cys-1123, Cys-1125, Cys-1171) significantly reduced the *cis* cleavage (Hijikata *et al.*, 1993).

However, one problem remains with the zinc protease model: the mutation of His-1175 to Ala did not affect the proteolytic activity of NS2-3. Other researchers have proposed a different mechanism, in which NS2 functions as a cysteine protease (Rawlings and Barrett, 1994b). Apparently, further studies are required to address this issue.

Concluding Remarks

From the review above, we can see that viral proteases play a key role in the replication of many virus, especially for the (+) stranded RNA viruses. The precursors of replicative proteins, e.g. polymerase, are invariably processed by viral proteases – sometimes in combination with host cellular proteases. In other viruses (e.g. picornavirus superfamily) viral structural proteins are also processed by viral proteases. In some cases, host proteases are involved in the processing of viral structural proteins but this normally takes place in the lumen of endoplasmic reticulum and/or along the translocation pathway.

Generally, the replication of RNA viruses starts by translating one or more polyproteins from genomic RNA. These polyproteins are then processed to give rise to proteins essential for viral RNA synthesis. Coat proteins can be translated as part of the above polyprotein, or translated from subgenomic RNA synthesized by viral RNA-dependent RNA polymerase. Retroviruses use a different strategy, in which viral polyproteins are processed after the assembly of virions.

For the four catalytic types of proteases, it seems cysteine is the most common nucleophile used by viral proteases, with serine being the second. As to the structure, the chymotrypsin-like folding is the most common. Therefore the chymotrypsin-like cysteine protease makes the largest group of viral proteases found so far.

Cleavage specificity is one of the most variable and elusive features of viral proteases. Unlike many cellular proteases (e.g. trypsin, proteinase K), which have rather broad substrate range, viral proteases normally have very strict cleavage specificity. Every

viral protease shows unique requirements for its substrates. Apparently, the primary amino acid sequence around the scissile bond is important, and some common patterns have been identified for 3C and 2A proteases within the picornaviruses. However, prediction of cleavage site based on primary sequences is only preliminary at this stage. For example, our understanding of the specificity of HIV protease is still limited, in spite of huge effort being put into this area. The secondary and tertiary structures may also be involved in the determination of site specificity and cleavage efficiency. Site-directed mutagenesis and resolution of atomic structures of viral proteases definitely can provide key information.

1.5 Aims of the Project

The aim of the project is to investigate the processing of the PLRV P1 and P1/2 (transframe) proteins.

P1 and P2 are the only two viral proteins required for viral replication, and multiple function domains have been identified within the P1 and P2 region, including a serine protease motif, VPg, and RdRp. The processing of P1 has been observed in plants. The elucidation of the pattern of the processing of P1 and P2 proteins and the role of the putative protease motif in this processing will enhance our understanding of the virus replication greatly.

The expression of P1 protein in bacteria will be carried out to raise polyclonal antibody specific to P1 protein. Due to the difficulties to study PLRV in plants (non-transmissible mechanically, confinement to phloem, and low virus concentration in plants), an *in vivo* system using the baculovirus as vector will be developed to study the proteolytic activity of PLRV P1 protease and the processing of P1 and P2 proteins. *In vitro* transcription and translation system will also be used in this study, in parallel to the *in vivo* system.

Chapter 2

Materials and Methods

2.1 Materials

Oligonucleotids

Oligonucleotides used in this study were synthesized by Ian Armitt in the University of St. Andrews, or purchased from OSWEL (Southampton, UK) or MWG-Biotech UK Ltd.

Plasmids

pRSETa, b, and c were obtained from Invitrogen Corporation. pGEX-3X was purchased from Pharmacia. pGEX-KG and pGEX-KT were obtained from J. E. Dixon (Dept. of Biological Chemistry, University of Michigan). pVL1392/1393 were purchased from Invitrogen. pFastBac1 was purchased from LifeTechnologies. pGEM-3zf(+) and pGEM-T easy were purchased from Promega. pSAB53, which contains a cDNA copy of the whole PLRV genome, was constructed in this lab by J.W. Lamb. pVL140Pk was a kind gift from Dr. R. E. Randall.

Insect Cells and Baculoviruses

SF9 cells were obtained from Invitrogen. Purified viral DNA of baculovirus AcRAK6 (Possee) was a kind gift from R.T. Hay. *E. coli* DH10Bac, which was used in the Bac-to-Bac system, was obtained from Life Technologies.

Plant Tissue

Leaves from healthy and PLRV-infected (Scottish isolate, strain 1) potato plants (*Solanum tuberosum*, cultivar Maris Piper) were kindly provided by M. A. Mayo (SCRI).

Antibody

Monoclonal antibody specific to Pk tag was kindly provided by R. E. Randall.

Table 2-1 List of oligonucleotids used in this study

NAME	SEQUENCE
304	M N R F T A Y 5′ GACT**AGATCT**CCACC ATG AAC AGA TTT ACC GCA TAT G 3′ Forward primer encoding the N-terminus of P1 of PLRV with a *Bgl* II site and the Kozak motif before the start ATG.
305	K L N S K A * g aag ctg aac tcc aaa gcc tga 3′ C TTC GAC TTG AGG TTT CGG ACT **CTTAAG**GTCA 5′ Reverse primer encoding the C-terminus of P1 of PLRV with an *Eco*R I site after the stop codon.
208	P G Y S G T G ccc gga tat **tcc gga** aca ggg 3′ CCT ATA AGG CCT TGT 5′ Reverse primer correcting the Val-355 mutation back to Gly. The *Acc* III site was also recovered.
3329	P G Y S G T G F 5′ GA CCC GGA TAT **TCC GGA** ACA GGG TTT G 3′ Forward primer correction the Val-355 mutation back to Gly. The *Acc* III site was also recovered.
155	D L K L N S K A gat ctg aag ctg aac tcc aaa gcc 3′ A GAC TTC GAC TTG AGG TTT CGG **CTC GAG** GTC A 5′ Reverse primer encoding the C-terminal of P1 with the stop codon replaced by a *Sac* I.

5637	A L T N K P L N G (P1) gct cta aca aac aag cct tta aat ggg ca 3′ GAT TGT TTG TTC GGA AAT 5′ Reverse primer to insert an extra A within the frameshift site so that the downstream P2 sequences is in-frame with P1 sequences.
390	Actagccaagcatacacgagttgcaa 3′ CGGTTCGTATGTGCTCAA 5′ Reverse primer with the intergenic region of PLRV genome, about 20 n.t. downstream of the P2 gene.
PkMfe	G K P I P N P L L G L D Gga aag ccg atc cca aac cct ttg ctg gga ttg gac 3′ CCT TTC GGC TAG GGT TTG GGA AAC GAC CCT AAC CTG S T * tcc acc tga **caattg** AGG TGG ACT GTTAAC 5′ Reverse primer encoding the Pk tag (14 aa) followed by a stop codon and a *Mfe* I site.
2HA	A L G Y A N C 5′ GCT CTC GGG TAC GCG AAC TG 3′ **tcg cga** gag ccc atg cgc ttg acg Forward primer encoding mutation H^{234}→A. A new *Eco*47 III site is created.
AHA	E N G S A L G Y 5′ GAA AAC GGC AGC GCT CTC GGG TAC 3′ ctt ttg ccg **tcg cga** gag ccc atg Forward primer encoding mutation H^{234}→A. A new *Eco*47 III site is created. Used to replace 2HA.
BHW	L V T A E W C L E G ttg gtg aca gct gaa tgg tg**t cta ga**a ggc 3′ C CAC TGT CGA CTT ACC ACA GAT CTT CC 5′ Reverse primer encoding mutation H^{255}→W. The down-stream *Xba* I site was used for cloning.
BHA	L V T A E A C L E G ttg gtg aca gct gaa gcg tg**t cta ga**a ggc 3′ C CAC TGT CGA CTT CGC ACA GAT CTT CC 5′ Reverse primer encoding mutation H^{255}→A. The down-stream *Xba* I site was used for cloning.

5DE	F K S A R N E I ttc aaa agt gcc cgt aat **gaa atc** 3′ G TTT TCA CGG GCA TTA CTT 5′ Reverse primer encoding mutation D^{286}→E. The *Eco*R V site (5′GATATC3′) in the wild-type sequence was eliminated.
6DA	F K S A R N A I ttc aaa agt gcc cgt aat **gct atc** 3′ G TTT TCA CGG GCA TTA CGA 5′ Reverse primer encoding mutation D^{286}→A. The *Eco*R V site (5′GATATC3′) in wild-type gene was eliminated.
7SC	C G T G F W S 5′ **TGC GGA** ACA GGG TTT TGG TC 3′ acg cct tgt ccc aaa acc aga Forward primer encoding mutation S^{354}→C. The *Acc* III site (5′TCCGGA3′) in the wild-type sequence was eliminated.
8SA	A G T G F W S 5′ **GCC GGA** ACA GGG TTT TGG TC 3′ cgg cct tgt ccc aaa acc aga Forward primer encoding mutation S^{354}→A. The *Acc* III site (5′TCCGGA3′) in the wild-type sequence was eliminated.
9SA	A P N Y V F E 5′ GCC CCA AAT TAT GTG TTT GA 3′ cgg ggt tta ata cac aaa ctc Forward primer encoding mutation S^{393}→A.

The sequences of oligonucleotides were shown in upper case and blue colour. Mutations introduced were indicated by red colour.

2.2 Methods

2.2.1 Cloning Techniques

PCR and Site-directed Mutagenesis

The polymerase chain reaction (PCR) was used to amplify specific DNA fragment for cloning purposes. It provides a convenient way to generate/modify DNA sequences of choice. A typical PCR was carried out in a 100 μl volume containing: 50 mM KCl, 10 mM Tris·Cl (pH9.0, 25°C), 0.1% (v/v) Triton X-100, 2.0 mM $MgCl_2$, 0.2 mM of each dNTP, 0.5 μM of each primer, template DNA (50 ng for plasmid; 1 μg for genomic DNA) and 1.25 unit of *Taq* DNA polymerase (Promega, 0.25 μl). Depending on the type of thermocycling machine, the reaction may or may not be overlaid with 100 μl mineral oil. Cycling parameters were variable and the following was only a guideline:

1. Pre-denature: 94°C, 5 min
2. 25-35 cycles of denaturing (94°C, 45 sec), annealing (50°C, 90 sec, variable) and extension (72°C, 1 min if product < 1.0 kb; 1 min/kb if > 1.0 kb)
3. Final extension: 72 °C, 7 min; then held at 4°C indefinitely

PCR products were then purified from agarose with QIAGEN kit or extracted with phenol/chloroform and precipitated in ethanol before further procedures. For QIAGEN HotStar *Taq* DNA polymerase, the pre-denaturing was carried out at 95°C for 15 min as per the manufacturer's instructions. The concentration of $MgCl_2$ may be optimized (1.5 to 4.5 mM).

Site-directed mutagenesis was carried out by a novel two-step PCR. The 1st PCR reaction was carried out using Vent^R DNA polymerase (NEB) under conditions specified by the manufacturer. The product was purified from agarose gel and 10-20% of it was used as one primer in the 2nd PCR reaction using *Taq* DNA polymerase (Promega), which was not proofreading.

Restriction Enzyme Digestions

All restriction of DNA fragments were carried out under conditions specified by the supplier. In general, 0.5-2.0 μg of DNA was digested at 3-5 units enzyme per μg DNA in a volume of 20 μl for 3-5 hours.

Agarose Gel Purification of DNA fragments

Agarose gel (0.7-2% w/v) electrophoresis was carried out in 1xTAE. Ethidium bromide was added to the gel or buffer at 0.5 μg/ml (final). DNA bands were visualized by UV (keep total exposure time < 1 min), excised with a clean scalpel and weighed. The DNA was extracted with QIAquick Gel Extraction Kit (QIAGEN), according to the manufacturer's instructions.

Ligations

DNA fragments may be blunted with Klenow Polymerase (5'-overhang) or T4 DNA polymerase (3'-overhang), or dephosphorylated with calf intestinal alkaline phosphatase (CIAP) after digestion by restriction enzymes. These end-modification reactions may be carried out directly in the restriction enzyme buffers; otherwise, DNA fragments were extracted with phenol/chloroform and precipitated in ethanol prior to modification.

Normally all the DNA fragments in a ligation reaction were gel purified. The vector DNA was added at a concentration so that the ratio *j/i* was larger than 10: *j* is the effective concentration of one end of a DNA molecule in the neighbourhood of the other end of the same molecule, while *i* is a measure of complementary termini in the solution (Sambrook *et al.*, 1989). The molar concentration of insert DNA was routinely about 5 times as that of the vector. This was particular important for a three-way ligation, in which three DNA fragments, each containing two different ends, were ligated together in one reaction. For a routine ligation, 20 ng vector DNA and 20 ng insert DNA in a 10 µl volume containing 1 Weiss unit of ligase (Promega, 0.3 µl) gave satisfactory results. Ligation was carried out at 4°C overnight (sticky-end ligation), or at room temperature for 1-3 hours (blunt-end ligation). The concentration of ATP may be reduced to 0.5 mM for blunt-end ligation. To ligate PCR products into pGEM-T easy vector, a 2xRapid ligation (Promega) buffer gave better results, which contained PEG and favoured intermolecular ligation.

Transformation of *E. coli*

Competent *E. coli* cells were prepared with the calcium chloride method (Modified from Current Protocols in Molecular Biology 1.8.1-3). 500 ml of LB broth were inoculated with 1 ml overnight culture of *E. coli* in LB, and incubated at 37°C on a shaker (220 rpm) until early- or mid-log phase was reached (about 2.5 hours, OD_{590} about 0.375), then the cells were cooled to 0 °C for 30 minutes, pelleted at 3000 rpm (1400 x g), 4 °C for 7 min, and the supernatant was discarded. Each pellet (from 250 ml culture) was resuspended in 125 ml of ice cold 0.1 M calcium chloride and stored on ice for 30 minutes. The cells were pelleted at 3000 rpm, 4°C for 5 min, the supernatant was discarded and then each pellet was resuspended in 5 ml ice cold 0.1 M calcium chloride containing 15% glycerol. Cells were dispensed into prechilled sterile tubes at 200 µl

aliquots, frozen in liquid nitrogen and stored at -70°C. Alternatively, the cell pellet was resuspended in ice cold 0.1 M calcium chloride, store at 4°C and used within one week.

To 50 µl of cells 1 ng of DNA, or 3-5 µl of ligation reaction, was added and incubated on ice for 30 minutes. The cells were then heat shocked by heating to 42 °C for 45-55 seconds and rapidly transferred to an ice bath for 1 minute, before adding 0.45 ml LB. Cells were incubated at 37°C for 1 hour, and plated out on LB agar plates containing the appropriate antibiotics. Bacterial colonies appeared within 12-16 hours.

Preparation of Plasmid DNA

(a) Mini-preparation

Mini-preparation of DNA was carried out by different methods; the Promega method was used for preparation of DNA for automated sequencing.

Alkaline lysis method (Sambrook *et al.*, 1989): 10 ml of LB broth containing 60 µg/ml ampicillin was inoculated with a single colony and incubated in an orbital incubator overnight. The cells were pelleted, the supernatant was removed, and the cells were resuspended in 200 µl of solution I (50 mM glucose, 25 mM Tris·Cl, pH8.0, 10 mM EDTA, 4°C). 400 µl of solution II (0.2 M NaOH, 1% SDS, freshly made, at room temperature) was added and mixed, followed by 300 µl of solution III (3M potassium acetate, 2M acetic acid, pH5.5). The sample was mixed and incubated on ice for 5 minutes, the solution was centrifuged and the supernatant was transferred to fresh tubes. The sample was then extracted by adding an equal volume of phenol/chloroform, vortexing, centrifuging and removing the upper aqueous layer to a fresh eppendorf tube. Two volumes of cold ethanol were then added, mixed, and the sample was incubated at

room temperature for 2 min. The sample was centrifuged and the DNA precipitate was washed with 70 % (v/v) ethanol, dried, and then resuspended in 50 μl of distilled TE (pH8.0) containing 20 μg/ml RNaseA.

Promega Wizard SV DNA miniprep kit: as instructed by the manufacturer. Briefly, 2-10 ml overnight (12-16 hours) culture of *E. Coli* was pelleted, resuspended in a solution containing 100 μg/ml RNase A, lysed with NaOH/SDS, and neutralized with a solution of guanidine hydrochloride/ potassium acetate/ acetic acid. After centrifugation, the supernatant was loaded onto the Spin Column. After two steps of washing, the plasmid DNA was eluted with 100 μl nuclease-free water or 10 mM Tris·Cl, pH8.3.

(b) Maxi-preparation

Maxi-preparation of plasmid DNA was carried out by two different methods; plasmids prepared by the 2nd method are suitable for *in vitro* transcription and translation.

Alkali/PEG method: A single colony was inoculated into 5-10 ml LB containing appropriate antibiotics and grown for about 8 hours or overnight. 250 ml of LB containing 60 μg/ml ampicillin was inoculated with 1 ml saturated culture and incubated in an orbital incubator until saturation (12-16 hours, $OD_{600} > 2$). The cells were pelleted and the supernatant was decanted. The cells were then resuspended in 2 ml of solution I. Then, 0.5 ml of 25 mg/ml lysozyme (in solution I) was added, mixed, and stood for 10 min at room temperature. 5 ml of solution II was added, mixed and stood on ice for 10 min, followed by adding 7.5 ml of solution III, mixing and standing on ice for another 10 min. The mixture was then centrifuged for 10 minutes at 20,000 x g, 4°C. The supernatant was poured through muslin into fresh tubes. Isopropanol (0.6 volumes) was added, left at room temperature for 10 minutes and then centrifuged for 10 minutes

at 15,000 g, room temperature. The supernatant was removed by pipetting and the nucleic acid was resuspended in 1 ml of TE.

An equal volume of 5M LiCl was added, mixed and centrifuged for at 10,000 x g, 4°C. The plasmid was precipitated with an equal volume of isopropanol at room temperature, washed with 70% ethanol and resuspended in 0.5 ml TE containing 20 μg/ml RNAase. After standing 30 min at room temperature, 0.5 ml 1.6M NaCl/13%(w/v) PEG8000 was added. After incubation 12-15 hours at 0-4°C, the plasmid was recovered by centrifugation at 12,000 g, 4°C for 20 min, dissolved in 0.4 ml TE, extracted with phenol/chloroform, precipitated in ethanol/ammonium acetate and resuspended in 500 ml 10 mM Tris·Cl, pH8.0.

Qiagen method: Maxi-preparations were carried out according to the Qiagen Plasmid Maxi Kit protocol, as per manufacturer's instructions (Qiagen). Briefly, 100 ml (500 ml for low-copy-number plasmids) overnight culture of *E. coli* was pelleted by centrifugation, resuspended in Buffer P1 (containing 100 μg/ml RNase A), lysed with Buffer P2 (containing NaOH/SDS), and neutralized with Buffer P3 (3M KOAc/ HOAc, pH 5.5). After centrifugation, the supernatant was loaded onto the column QIAGEN-tip 500, equilibrated with Buffer QBT. After two steps of washing with Buffer QC, the plasmid DNA was eluted with Buffer QF, precipitated with 0.7 volumes of isopropanol. The plasmid pellet was washed with 70% ethanol and resuspended with 100 μl nuclease-free water or 10 mM Tris·Cl, pH8.3.

2.2.2 Expression/Purification of His-tagged Proteins

Expression Using *E. coli* Strain BL21(DE3)

The *E. coli* strain BL21(DE3), containing a chromosomal copy of T7 RNA polymerase under the control of the *lacUV5* promoter, was transformed with pRSET plasmid with target gene inserted downstream of the T7 promoter. 10 ml LB/Amp was inoculated with a single colony and incubated at 37°C with shaking until OD600 was 0.5-1.0. Expression was induced by adding IPTG to 0.5 mM. After further 2-3 hours of incubation, cells were harvested by centrifugation.

Expression Using M13/T7 Phage

E. coli Strain JM109 cells were transformed with recombinant pRSET plasmids. 10 ml LB/amp was inoculated with a single colony and incubated at 37°C to OD_{600}=0.3, when 1 mM IPTG was added. The cells were infected one hour later with M13/T7 phage at an MOI of 5. Cells were harvested 4 hours later unless indicated otherwise.

Preparation of M13/T7 Stock: 5 ml LB was inoculated with a single JM109 colony picked from an M9 minimal plate containing thiamine (0.5 ng/ml) and incubated at 37°C for 6-8 hours (plating cells). M13/T7 phage stock was diluted serially (10^{-5} to 10^{-10}) and 0.1 ml each dilution was added to 3 ml of molten LB top-agar (0.7% agar, kept at 47°C), followed by 0.1 ml JM109 plating cells. The contents were vortexed briefly and poured onto an LB plate (no antibiotics) and incubated at 37°C for 12-16 hours. Plaques were picked into 2ml LB/50 µl plating cells and incubated at 37°C for 4-5 hours. Cells were collected for mini-prep of the replicative form (RF) of M13/T7 phage DNA, which was

purified as a plasmid and restricted with *Hind* III to confirm the existence of the 2.7 kb fragment containing the T7 RNA polymerase gene. The supernatant was kept at 4°C as phage stock after titration.

Analytical Purification

For purification under denaturing conditions, cells were resuspend in 0.4 ml Buffer B (8M urea, 0.1M NaH_2PO_4, 10 mM Tris·Cl, pH 8.0) and lysed by gently vortexing. Cell debris were removed by centrifugation at 15,000 g for 20 min. 30-50 µl slurry of Ni-NTA beads (QIAGEN) was added to the supernatant, and mixed gently for 30 min at room temperature. Beads were collected by spinning for 10 sec in a microfuge, washed 3 times with 1 ml Buffer B containing 10 mM imidazole. His-tagged proteins were eluted with Buffer B containing 250 mM imidazole and analyzed by SDS-PAGE and Coomassie blue staining.

For purification under native conditions, cells were resuspended in 1 ml ice cold lysis buffer (50 mM NaH2PO4, pH8.0, 0.3 M NaCl, 10 mM imidazole), and lysed by sonication (3x 10 sec). After centrifugation at 10,000 g, 4°C for 20 min, the supernatant was mixed with 30-50 µl slurry of Ni-NTA beads at 4°C for 60 min. Beads were collected by centrifugation and washed 3 times with lysis buffer containing 20 mM imidazole. Proteins were eluted with lysis buffer containing 250 mM imidazole.

Preparative Purification of HP1PE (Native Conditions)

His-tagged protein HP1PE was expressed with the plasmid pHP1PE using the M13/T7 system. 1 litre culture was prepared similarly to the analytic purification. Cells were collected and resuspended in 20 ml lysis buffer containing 1 mM PMSF. Lysozyme was

added to 1 mg/ml. After incubation on ice for 30 min, the cells were lysed by sonication (9x 10 sec). Cell debris was removed. 2 ml Ni-NTA beads were added to the supernatant and mixed gently by shaking at 4°C for 60 min. The beads were loaded onto a column and washed 20 ml lysis buffer containing 20 mM imidazole. His-tagged protein was eluted with 20 ml lysis buffer containing 40-250 mM imidazole/1.0 M NaCl.

2.2.3 Expression/Purification of GST-fusion Proteins

Screening of transformants

E. coli strain DH5α was transformed with the ligation mixture. Transformant colonies (6-8) were picked into 10 ml LB/Amp and incubated at 37°C overnight. Overnight culture (1 ml) was diluted 1:10 with LB/Amp and incubated for 1 hour, while the remaining cells were saved for mini-prep of plasmids. Fusion protein expression was induced by the addition of IPTG to a final concentration of 0.5 mM. Incubation continued at appropriate temperature (37°C or 25°C) for another 2 hours. Cells were collected by centrifugation, resuspended in 1 ml PBS and lysed by sonication (3x 10 sec). Cell debris was removed and 20 μl 50% glutathione agarose beads was added to the supernatant and mixed gently at room temperature for 20-30 min. The beads were washed with 1 ml PBS, and fusion protein was eluted with 50 μl SDS loading buffer.

Preparative Expression of GST-P1SP

LB/Amp (500 ml) was inoculated with 5 ml overnight culture of *E. coli* strain JM101 transformed with pGEX-P1SP and grown at 37°C with shaking until OD_{600} reached 0.5. IPTG was added to a final concentration of 0.5 mM and the culture was incubated for another 4 hours. Cells were collected and resuspended in 20 ml PBS containing 1%

Triton X-100, 1 mM PMSF and 5 mM benzamidine. The cells were then lysed by passing through a French press twice. Cell debris was removed by centrifugation at 15,000 g, 4°C for 15 min. 1 ml glutathione-agarose beads were added to the supernatant and mixed gently by shaking at 4°C for 60 min. The beads were washed with 5x 10 ml PBS/1% Triton X-100/0.5 M NaCl and 5x 10 ml PBS/0.15M NaCl and resuspended in 1 ml PBS/0.15M NaCl. 20 μl thrombin (Sigma, 1 U/μl) was added and mixed gently at 25°C for 1-3 hours. The mixture was loaded onto a column and released P1SP was eluted with 10 ml PBS/0.15M NaCl.

2.2.4 Generation of Antisera

Pre-bleed serum (5 ml) was collected from two rabbits in the Animal House at the University of St. Andrews, one day before the primary inoculation. For each rabbit, 50 μgP1SP (in 0.5 ml PBS) was mixed with 0.5 ml Freund's complete adjuvant, completely emulsified by sonication, and then injected into multiple intramuscular and subcutaneous sites. Booster injections (100 μg antigen) were administered at 2 weeks intervals using incomplete Freund's adjuvant. The titters were monitored by ELISA 10 days after each booster. The final bleeding was carried out after the 4th booster (500 μg antigen). The serum was stored in two ways: (1) storing at -20°C in 50% glycerol, for routine use; (2) lyophilization, for long term storage

2.2.5 ELISA

Indirect ELISA was carried out essentially as described in Current Protocols in Immunology (2.1.3-6). Purified protein P1SP (2 μg/ml stock in PBS) was used as antigen, while cell extract (2 μg/ml stock in PBS) of *E. coli* JM101, harbouring pGEX-KT and induced with IPTG, was used as negative control. Each well of a 96-well plate

was coated with 100 µl of P1SP or *E. coli* JM101 extract overnight at 4°C or 1 hour at 37°C, washed 5 times with PBS/0.1% Tween-20, and then blocked with 200 PBS/5%dried milk/0.1% Tween-20 for 1 hour at 37°C. The plate was washed 5 times with PBS/0.1% Tween-20. Test antiserum diluted in blocking buffer was added 100 µl /well and incubated at room temperature for 1 hour and then washed 5 times with PBS/0.1% Tween-20. Horseradish peroxidase conjugate (goat anti-rabbit IgG) was diluted 1:4000 with blocking buffer and applied to each well (100 µl), incubated at room temperature for 1 hour, and washed 5 times with PBS/0.1% Tween-20. Solution A containing substrate TMB and Solution B containing H_2O_2 (both purchased from KPL) were mixed at 1:1 (v/v) just before use, and 100 of the mixture was applied to each well. The plate was left at room temperature 10-20 min, and the reaction was stopped by adding 100 µl 1N HCl. The plate was read on a Dynatech MR5000 machine (dual wavelength, test filter 450 nm, reference filter 630 nm).

2.2.6 SDS-PAGE

The system of Laemmli (1970) was used. Gels were constructed with a 5% stacking and a 10% or 13% separating polyacrylamide gel, and were run on the Bio-Rad Mini-PROTEAN II system according to the manufacturer's instruction.

Sample Preparation:

Protein solution: 1 volume 2x SDS sample buffer (125 mM Tris·Cl, pH6.8, 4%SDS, 2% 2-mecaptoethanol, 25% glycerol, 0.001% bromophenol blue) was added and the sample was heated to 95°C for 3 min. For a complex mixture, 20-25 µg protein in 10 µl SDS sample buffer gave a strongly stained Coomassie blue pattern, and 0.5-1.0 µg was required for highly purified proteins

E. coli cells: 1 OD_{600} (about 10^9 cells in 1 ml) cells was resuspended in 100 µl TE buffer (on ice), mixed with 100 µl 2x SDS sample buffer and immediately heated to 95°C for 3 min.

Insect cells: 10^6 cells was washed with PBS, resuspended in 50 µl TE buffer (on ice), mixed with 2x SDS sample buffer and immediately heated to 95°C for 3 min; alternatively, washed monolayer cells with ice-cold PBS on the plate, left the plate on ice, added 100 µl 1x SDS sample buffer, transferred the lysate into an eppendorf tube and immediately heated to 95°C for 3 min.

Plant leaves: cut the leaves into small pieces, grind in 4 vol. 2xSDS buffer on ice, transfer to an eppendorf tube and immediately heated to 95°C for 5 min. After centrifugation at 12,000 g for 10 min, 5-10 µl of the supernatant is loaded to the gel (the buffer of the sample has 2-fold strength).

2.2.7 Western Blotting

Proteins were resolved by SDS-PAGE (10% or 13%) and the gel was soaked in BSN buffer (48 mM Tris·Cl, 39 mM glycine, 20% methanol, pH9.2) containing 0.0375% SDS for about 15 min. The proteins were transferred onto PVDF (polyvinylidene difluoride) membrane at 10V for 30 min in a Bio-Rad SD semi-dry transfer cell, as described (Bjerrum and Schafer-Nielson, 1986). Membranes were blocked with PBS/0.1% Tween 20/ 0.1%BSA/ 0.1%Ficoll/ 0.1% PVP for 4 hours at room temperature or overnight at 4°C. The rabbit polyclonal anti-P1SP primary antibodies (1:2000 dilution in blocking buffer) were incubated with the membranes for 1 hour. The mouse monoclonal anti-Pk tag antibodies (1:1000 dilution in blocking buffer) were

incubated with the membranes for 1 hour. Membranes were then washed with PBS/0.1% Tween 20 and incubated for 1 hour with peroxidase-conjugated goat anti-rabbit or mouse IgG (1:4000 dilution in blocking buffer). Bound secondary antibody was detected using enhanced chemiluminescence (ECL, from Amersham).

2.2.8 *In Vitro* Transcription and Translation

A coupled transcription/translation wheat germ extract system (TnT; Promega) was programmed with 1 μg unrestricted pGEM-derived plasmid DNA and incubated at 30°C for 90 min, in the presence of T7 RNA polymerase and [^{35}S] methionine, as recommended by the manufacturer. About 5 μl of the reaction was analyzed by SDS-PAGE and autoradiography. TnT reaction using rabbit reticulocyte lysate was carried as recommended by the manufacturer as well.

Extended incubations were performed following the arrest of translation by the addition of RNase A (0.5mg/ml) and cycloheximide (0.8mg/ml). The reaction might be overlaid with mineral oil to prevent evaporation and protein precipitation.

For cotranslational processing using canine pancreatic microsomal membranes (CMM) in TnT reaction, the following recipe was used:

TnT Lysate	12.5 μl
TnT Reaction Buffer	0.5 μl
Amino Acid Mixture (-Met)	0.5 μl
RNasin Ribonuclease Inhibitor (40U/μl)	0.5 μl
TnT T7 RNA Polymerase	0.5 μl
[^{35}S]-Met (1000 Ci/mmol) at 10 mCi/ml	2.0 μl

Nuclease-free water	6.5 µl
Plasmid DNA (1 µg/µl)	0.5 µl
CMM	1.5 µl

(1) Membrane association experiment (Promega UK):

TnT reaction was performed as normal and then spun at 70,000 rpm (200,000 g), 4°C in a tabletop ultracentrifuge (Beckman, TL-100) for 30 min. The supernatant was removed and the pellet was resolubilized in 10 µl 1x SDS sample buffer.

(2) Proteinase K protection experiment (Alperin *et al.*, 1997):

TnT reactions were performed with CMM, then stored on ice. Proteinase K was diluted to 1 mg/ml with 50 mM Tris·Cl pH7.5, 10 mM $CaCl_2$, and incubated at 37°C for 15 min to inactivate potential lipases. Proteinase K digestions were set up with and without 0.8% Triton X-100: added 1 µl 100 mM $CaCl_2$ and 1 µl proteinase K to 10 µl TnT reaction. After incubation for 1 hour on ice, PMSF was added to 1 mM to stop the digestion. 1 vol. of 2x SDS sample buffer was then added, and solution was heated to 95°C for 3 min.

2.2.9 Protein Assay

Routinely, the concentration of protein in solutions was assayed with the method of Bradford (1976): 100 µl solution (containing protein 1-10 µg) was mixed with 1 ml G-250 reagent (dissolve 100 mg Coomassie Brilliant Blue G-250 in 50 ml 95% ethanol, mixed with 100 ml 85% phosphoric acid, diluted to 1 litre). OD_{595} was measured within 20 min.

SDS and Triton X-100 interfere in the above assay system. For solutions containing these substances, the method of Brown *et al.* (1989) was used: diluted protein solution to 1 ml, added 0.1 ml sodium deoxycholate (DOC) (0.15% w/v), and stood at room temperature for 10 min. After spinning at 3000g for 15 min at room temperature, the pellet was dissolved with 50 µl 5%SDS/0.1N NaOH. BCA stock solution (Solution A, from Pierce) and 4% $CuSO_4 \cdot 5H_2O$ (Solution B) was mixed at 50:1 (v/v). This working solution (1 ml) was added to the protein solutions, vortexed immediately and incubated at room temperature (2 hours), 37°C (30 min) or 60°C (30 min), then cooled to room temperature. OD_{562} was then measured.

2.2.10 Tissue Culture and Baculovirus Methods

The methods described by King and Possee (1992) and O'Reilly *et al.* (1992) were used.

Routine SF9 Cell Culture Procedures

SF9 cells were routinely maintained and passaged as monolayer culture in 75 cm^2 flasks containing 25 ml TC-100/7%FCS (as described by O'Reilly *et al.*, 1992). When reaching 80-90% confluence, the cells were subcultured. Cells were washed off by gently pipetting and diluted at 1:8 into a new flask. The cells were incubated at 28°C for 4-5 days when the next subculturing was due. When the suspension culture was needed, the monolayer cells in one 75cm^2 flask was washed off the surface and transferred to a spinner flask containing 200 ml TC-100/7%FCS. The cells were incubated at 28°C with stirring at 60-80 rpm until cell density reached $1x10^6$ cells/ml (3-4 days). The cells were diluted 5 fold and were ready to use when the density reached $1x10^6$ cells/ml.

Transfection

Two methods were used to transfect SF9 cells with viral DNA. For Bac-to-Bac system (GibcoBRL), the methods recommended by the manufacturer were followed. For the pVL1392/1393 system, the following co-transfection method was used:

Each petri dish (35mm) was seeded with 1-2 ml SF9 spinner cells. It took about 30min to form monolayer (about 70-90% confluent). In a polyethylene tube, 1 μl baculovirus DNA (AcRAK6, 40 ng/μl, linearized with *Bsu*36 I) and 0.32 μg Qiagen purified transfer plasmid were mixed with water in a total volume of 15 μl. Lipofectin (5 μl) was then added and mixed gently by moving pipette tip up and down. Heavy precipitate should not form; but ideally you might see very fine one. The mixture was then left at room temperature for about 10 min. Media were removed and cells were washed with 1ml TC-100 (no serum) before adding 1ml fresh TC-100 (no serum). Lipofectin was added to the cells, dropwise into TC-100 at different points on plate. The dish was rocked gently back and forward and then placed in an incubator at 28°C for 3-4 hours (no longer). Medium was removed and cells were gently washed with fresh TC-100/ 7%FCS. After adding 2 ml fresh TC-100/7%FCS, the cells were returned to the incubator. Four days later, the cells were harvested by centrifugation at 1500rpm for 5 min and store at 4°C.

Plaque Assay/Purification

Dishes (35 mm) were seeded with SF9 (1.5×10^6 in 1.5-2.0 ml medium) and left for 1-2 h at 28°C to allow the cells to form a fairly sparse, even monolayer (50% confluence). Serial log dilutions of the virus were prepared (recombinant virus, 10^{-5}-10^{-7} dilutions; co-transfection S/N, 10^{-1}-10^{-4} dilutions; plaque-pick, 10^{-1}-10^{-3} dilutions). The medium

was removed with a Pasteur pipette attached to a vacuum line. Virus dilution (100 µl) was overlaid (dropwise) to the centre of the dish, left at room temperature for 1 h on a level surface and rocked very gently at 15-20 min intervals.

3% LGT agarose (in water, autoclaved) was melted with microwave and cooled to 37°C before adding 2 vol. of TC100/10%FCS (37°C) carefully (avoid bubbles). The mixture was stored at 37°C. The inoculum was removed and 2 ml agarose overlay was pipetted down the side wall. The agarose was allowed to set at room temperature and 1 ml TC100/10%FCS liquid medium was added. The dish was placed in a moist box and incubated at 28°C for 3-4 days.

Neutral red (0.5%) was diluted 1 in 20 with TC-100/10%FCS. Each dished was stained with 1 ml neutral red solution for 2-4 hours, the stain was then removed, the dish was inverted and stored at room temperature for 2 h or O/N, in the dark. As soon as possible, well isolated plaques were marked with a marker pen. After calculating the titre, a plug of agarose from the plaque was picked up with a glass Pasteur pipette and washed out into 0.5 ml TC-100 growth medium. The virus was released by vortexing and stored at 4°C.

Infection

Infection of SF Cells in Monolayer Culture

35 mm dishes were seeded with 1x 10^6 cells/dish in 1.5-2 ml growth medium and incubated overnight at 28 °C. The cells were checked under microscope (healthy, 50%confluent). The medium was removed, and 0.1 ml virus inoculum was added (dropwise) very gently to the centre of the dish, at an appropriate MOI. The cells were

incubated for 1 h at room temperature (ensure the surface is level; rock dishes every 15-20 min to distribute virus evenly). The dishes were tilted to aspirate the inoculum. Immediately, 1.5-2 ml medium was added gently. The dishes were incubated in a humidified sandwich box lined with moist tissue paper.

Dish/Flask	Cell Number (2h:O/N)	Inoculum Volume	Medium Volume
35mm dish	$1.5{:}1.0 \times 10^6$	0.1:0.5 ml	1.5-2.0 ml
60mm dish	$2.5{:}2.0 \times 10^6$	0.4:1.0 ml	3.0-5.0 ml
$25cm^2$ flask	$2.0{:}1.5 \times 10^6$	0.4:1.0 ml	3.0-5.0 ml
$75cm^2$ flask	$6.0{:}4.0 \times 10^6$	1.0:2.0 ml	10-15 ml

NB: All the protocols above were for analysis of proteins or DNA. For the preparation of virus stock, less SF9 cells were seeded, and the cells were infected at a low MOI of 0.1-0.2. The supernatant was harvested 3-6 days later, until cells were well infected.

Infection of SF Cells in Suspension Culture

For preparing virus stock, a suspension was seeded at $1\text{-}2 \times 10^5$ cells/ml and grew to 5×10^5/ml. Required amount of virus inoculum (MOI 0.1-0.2) was added directly. Concentration by centrifugation was not necessary. The medium was harvested after 4-5 days, or when the cells were well infected. Virus stock was aliqoted and stored at 4°C (up to 1 year) or -70°C. The titre was determined by plaque assay.

For analysis/purification of DNA/protein, a suspension was grown to 1×10^6 cells/ml. The cells were collected by centrifugation at 800 x g, room temperature for 5 min, and resuspended gently in fresh TC-100/7%FCS (0.1 original volume). The required amount

of virus inoculum was added, and the cells were incubated at 28°C for 1 hour, swirling periodically to mix. The medium/inoculum was removed by centrifugation and the cells were resuspended with TC-100/7%FCS and returned to the spinner flask. The volume was brought back to the original one with fresh TC-100/7%FCS. The cells were incubated at 28°C and harvested when required.

Chapter 3

Results

3.1 Expression of PLRV Protein P1 in *Escherichia coli*

3.1.1 Introduction

Like other luteoviruses, PLRV is difficult to study for a few reasons. Firstly, the virus particles are confined within the phloem tissue and can only be transmitted by aphids (a procedure difficult to control). Secondly, the concentration of virus particles in plants is very low, while the concentration of viral non-structural proteins is even lower. In fact, the first report on detection of viral non-structural proteins was published in 1999 (Prüfer *et al.*), some ten years after the publication of the full sequence of the genome. Furthermore, studies on replication of PLRV were hampered by two related problems: the lack of purified viral non-structural proteins and the lack of antibodies against them.

Previously, P1 and P2 were expressed as ß-galactosidase fusion proteins (J. Lamb, personal communication). This project set out to address those problems mentioned above by expressing PLRV P1 protein in *E. coli* using other vector systems. At first, expression of full-length P1 as a GST-fusion or a His-tagged protein proved unsuccessful. Subsequently, expression of different parts of the P1 gene was attempted, and some success was achieved. Antibody was raised against purified P1 derived protein (see Section 3.2). Finally, based on the results obtained from baculovirus system, we tried to express the active protease domain of P1.

Two prokaryotic expression systems, pRSET and pGEX, were used to express P1 protein, which are described briefly here. Both systems have wide application in expressing various prokaryotic and eukaryotic proteins

The pRSET Expression System

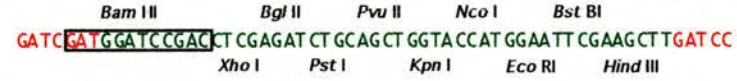
The pRSET plasmids are pUC-derived expression vectors to express recombinant protein in *E. coli* (Figure 3-1). The target gene is inserted downstream of and in frame with sequences coding for an initiation Met, a series of six histidine residues for nickel binding, a transcript stabilizing sequence from gene 10 of phage T7, and an enterokinase cleavage site. Downstream of the enterokinase cleavage site is a multicloning site (MCS), containing 10 unique restriction sites. Appropriate vector can be chosen from a set of three versions of pRSET plasmids (pRSET A, B and C), so that the target gene is in frame with the initial ATG codon.

High levels of expression of target proteins are based on the T7 promoter transcription-translation system. After providing a source of T7 RNA polymerase, which is very selective and active, most of the host's resources are sequestered to express target proteins. In hosts like *E. coli* BL21(DE3), the T7 RNA polymerase (under the control of the lacUV5 promoter) can be induced by IPTG from a copy of phage λDE3 lysogenized in the host chromosome. There is some basal transcription of T7 RNA polymerase in the uninduced state, which in turn allows some expression of target proteins. For expression of toxic proteins, hosts carrying plasmid pLysS or pLysE provide more stringent control, because pLys plasmids encode T7 lysozyme, which is a natural inhibitor of T7 RNA polymerase. The most stringent control of expression of target protein might be achieved by using an M13 phage to provide the T7 RNA polymerase. There is no T7

pRSET A

pRSET B

pRSET C



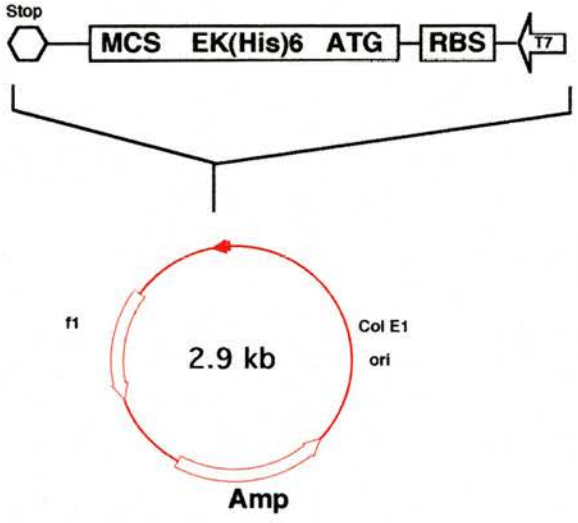


Figure 3-1 Plasmids pRSET A, B and C
Multiple cloning site (MCS) is shown in green colour. Variable regions are indicated by boxes. RBS, ribosome binding site; EK, enterokinase cleavage site. Note that pRSET C does not contain the *Sac* I site.

RNA polymerase in the host before infection. Infection is carried out in the presence of IPTG, which induces the expression of T7 RNA polymerase from the lac promoter in the M13 phage.

The tag of six histidine residues allows purification of target protein using immobilized nickel beads in a single step. The six consecutive histidine residues bind to the nickel ions on the matrix via their imidazole rings. The interaction between them is independent of the tertiary structure of the proteins, as long as the His tag is exposed. Therefore, target proteins can be purified under native or denaturing conditions. This is particularly useful to purify proteins that form inclusion bodies, which can be solubilized with 8M urea or even 6M guanidine hydrochloride.

Furthermore, the presence of an enterokinase cleavage site between the His tag and the target protein makes it possible to remove the N-terminal fusion peptide after the affinity purification of the tagged protein.

The pGEX Expression System

The pGEX vectors (Figure 3-2) are pBR322-derived plasmids designed to express recombinant target proteins fused to the C-terminus of glutathione-S-transferase (GST), an eukaryotic cytoplasmic protein of 26 kDa. High levels of expression are achieved using IPTG inducible, strong tac promoter. Since the plasmids also encode the lac repressor ($lacI^q$) required to repress basal expression, a wide range of suitable hosts are available.

The main advantage of this system is that the fusion proteins typically remain soluble within the *E. coli* and can be purified using immobilized glutathione at mild conditions.

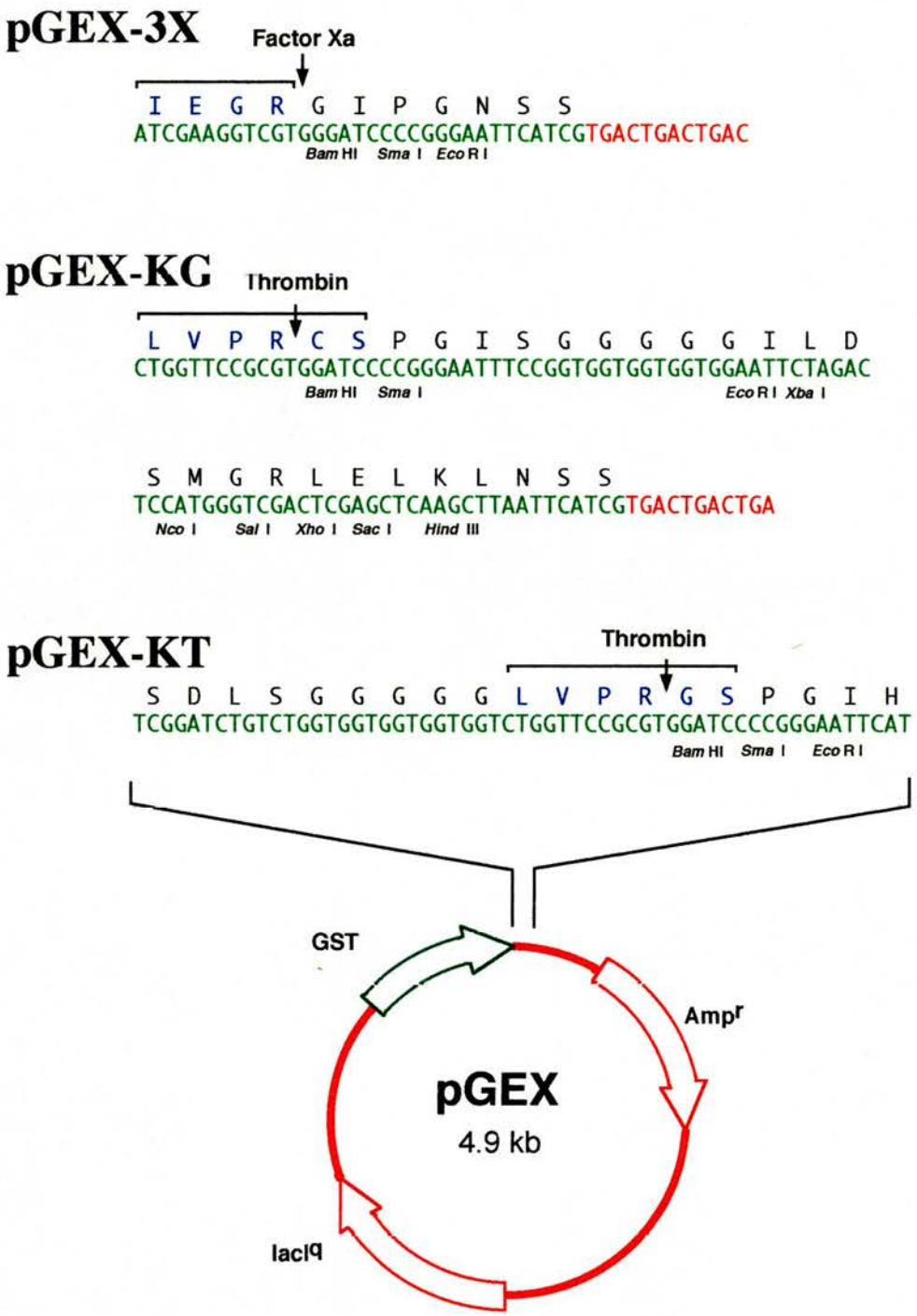


Figure 3-2 pGEX plasmids used in this study

Therefore, foreign proteins normally retain their antigenicity and biological activities. Other remarkable features include the efficiency and rapidity of purification, especially when target protein moiety can be released from the affinity column with site specific protease thrombin or Xa, leaving the GST handle still attached to the beads. Pure protein can be obtained in a single step.

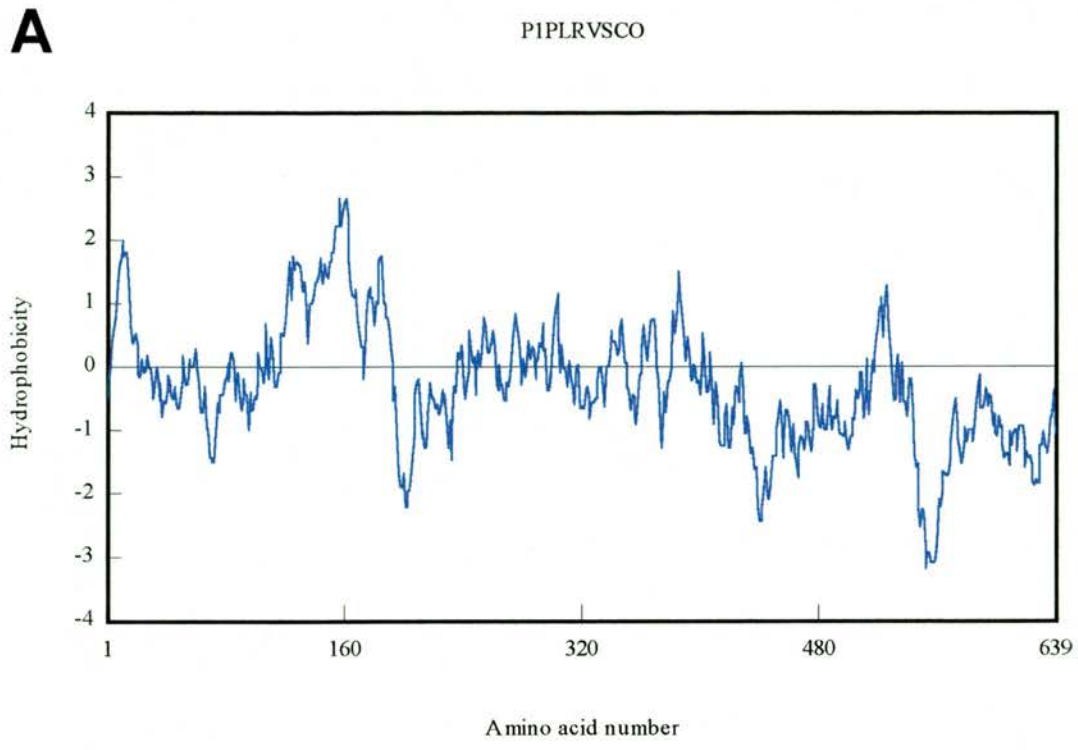
Three pGEX vectors were used in this project, pGEX-3X, pGEX-KG and pGEX-KT (Figure 3-2). pGEX-3X encodes a cleavage site of factor Xa between the GST coding sequence and the multiple cloning site (MCS), while pGEX-KG and pGEX-KT encode a thrombin cleavage site instead. pGEX-KG contains more unique restriction sites within its MCS, and encodes a glycine-rich linker (glycine kinker) located immediately downstream of the thrombin cleavage site, which facilitates the cleavage of fusion proteins. In pGEX-KT, the glycine kinker is placed upstream of the thrombin cleavage site to minimize the unrelated N-terminal extension of the released protein.

3.1.2 Computer analysis of P1 protein

Before construction of recombination plasmids encoding the P1 protein, the peptide sequence of P1 was analyzed with computer software, GCG package and DNAman, in order to visualize basic features of this protein (Figure 3-3 and 3-4). Since the algorithms are not entirely reliable, the results should be viewed with caution.

P1 protein is 639 amino acids long. When this work was carried out, the function of P1 protein was unclear but it was believed to be involved in viral RNA replication. At that time, a chymotrypsin-like serine protease domain in the middle one-third of the protein was predicted, but the VPg domain had not yet been located at the C-terminal part of P1.

A



B

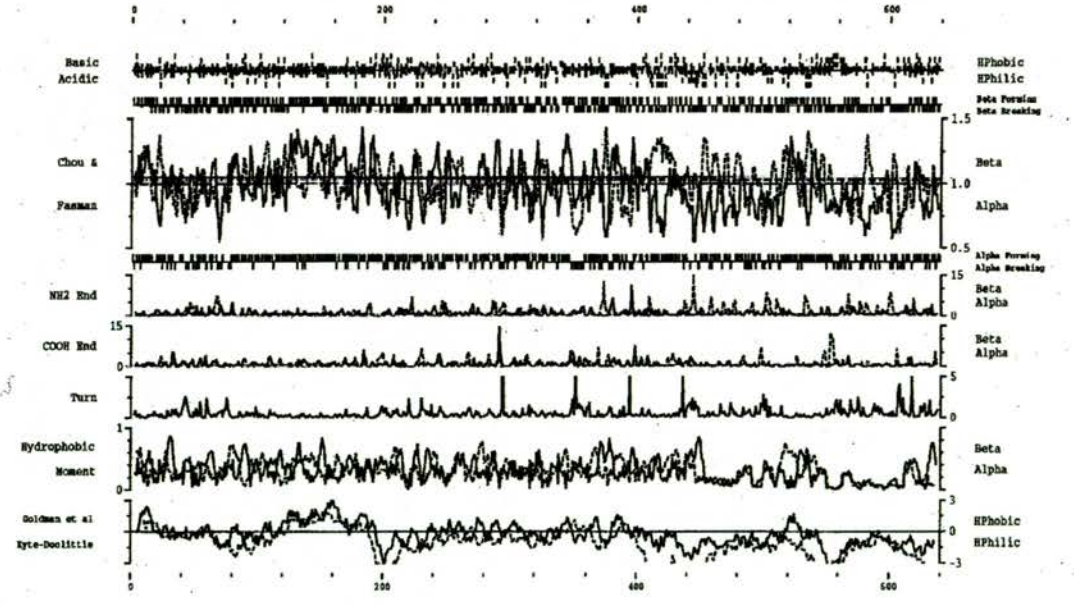


Figure 3-3 Sequence analysis of the P1 protein

A. Hydrophobic profile (plotted with DNAMAN)

B. Predictions of basic/acidic profile, secondary structure, and hydrophobic/ hydrophilic profile (plotted with PEPPLOT of GCG)

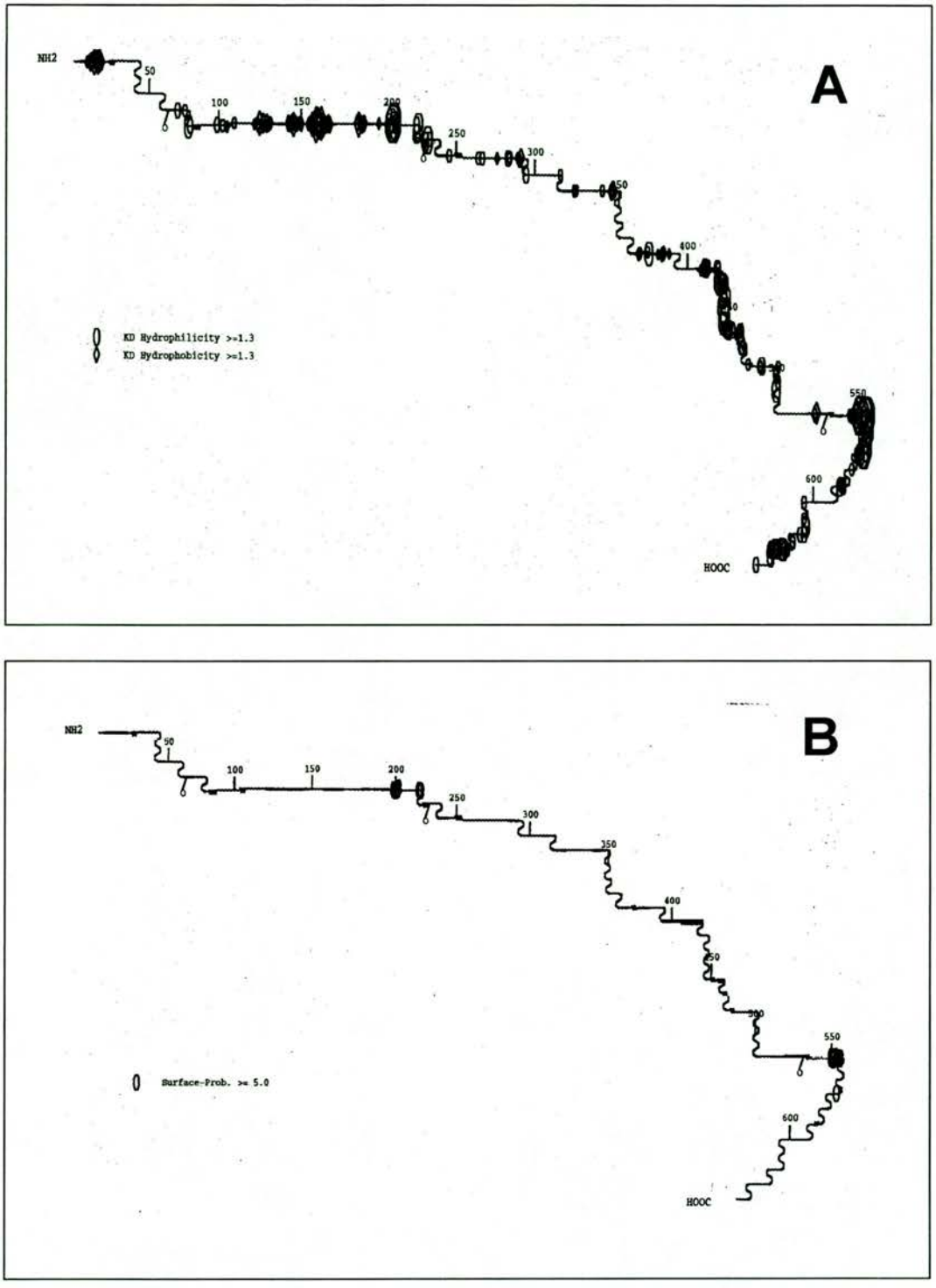


Figure 3-4 Sequence analysis of PLRV P1 protein: Chou-Fasman Prediction

A. Hydrophilic/hydrophobic Profile

B. Surface Probability

Analysis using DNAman predicted that P1 protein had a MW of 69550, a pI of 9.44-9.88 (depending on the number of disulfide bond). In terms of hydrophilic/ hydrophobic character, the N-terminus of P1 has a short segment of hydrophobic residues before a small moderately hydrophilic region, followed by a long highly hydrophobic region. There is a hydrophilic region immediately downstream of this long hydrophobic region. The middle one-third of P1, which contains the protease domain, is not particularly hydrophilic or hydrophobic; whilst the C-terminal one-third is highly hydrophilic, containing a lot of charged residues. This region is particularly rich in basic amino acids, and might have nucleic acids binding ability.

The GCG package gave more comprehensive and complicated analysis, including prediction of secondary structure, mapping of basic/acidic residues, and surface probability. Overall, the two software packages gave compatible results.

3.1.3 Expression of full-length P1 proteins

3.1.3.1 Construction of plasmids pHP1 and pGEX-P1

The DNA fragment encoding the entire P1 protein was amplified from plasmid pSAB53 using PCR. The forward primer was 304, containing a *Bgl* II site and the Kozak motif at its 5' end. The Kozak sequence, CC[A/G]CCAUG(G), is a consensus sequence for eukaryotic initiation of translation (Kozak, 1984). The reverse primer was 305, with an *Eco*R I site following the stop codon of P1.

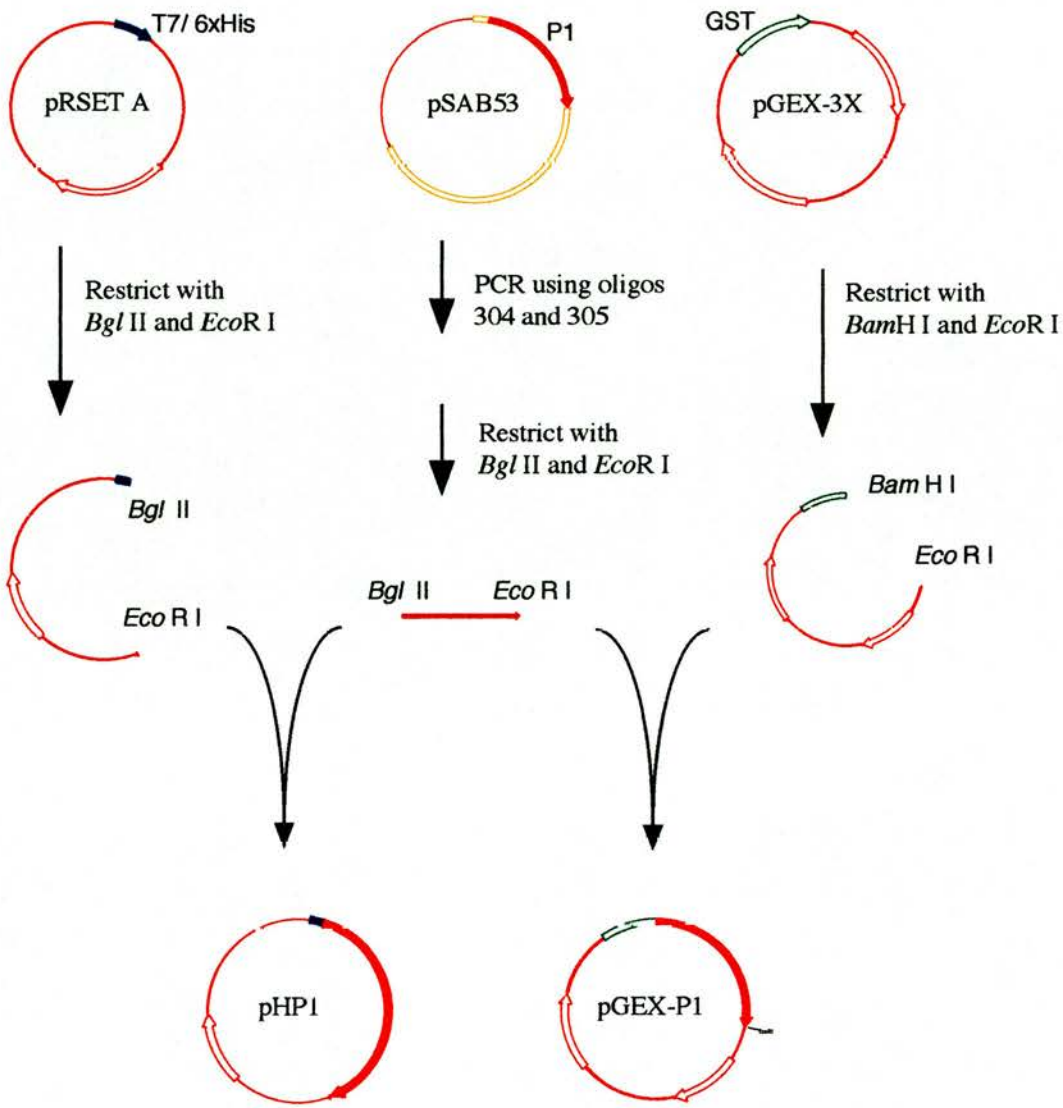


Figure 3-5 Construction of pHP1 and pGEX-P1

304:

```
                          M   N   R   F   T   A   Y
5'GACTAGATCTCCACC ATG AAC AGA TTT ACC GCA TAT G 3'
      Bgl II
```

305:

```
      K   L   N   S   K   A   *
  g aag ctg aac tcc aaa gcc tga
3'C TTC GAC TTG AGG TTT CGG ACT CTTAAGGTCA 5'
                                 EcoR I
```

The PCR product was restricted with *Bgl* II and *Eco*R I, purified from agarose gel, then ligated into (1) pRSET A, restricted with *Bgl* II and *Eco*R I, to generate plasmid pHP1; (2) pGEX-3X, restricted with *Bam*H I and *Eco*R I, to generate plasmid pGEX-P1 (Figure 3-5). In pHP1, the sequences encoding the N-terminal peptide and the P1 sequences formed an ORF of a 74 kDa fusion protein. In pGEX-P1, the fusion ORF encoded a protein of 96 kDa.

It was found that pSAB53 contained a point mutation causing a coding change (Gly-355 to Val) within an important motif (GXSG) of the putative serine protease domain.

3.1.3.2 Expression of full-length P1 fusion proteins in *E. coli*

pHP1

E. coli BL21(DE3) was transformed with pHP1 and grown to an OD_{600} of about 0.5. Expression was induced by adding IPTG to a final concentration of 0.5 mM. Cells were harvested at 0, 1, 2 and 4 hours after induction, and the proteins were then examined by SDS-PAGE. It was found that his-tagged full-length P1 protein was not expressed in sufficient amounts to be identified on Coomassie blue stained gel (Figure 3-6, panel A). Since the cells carrying pHP1 grew very slowly, it was suspected that P1 protein might

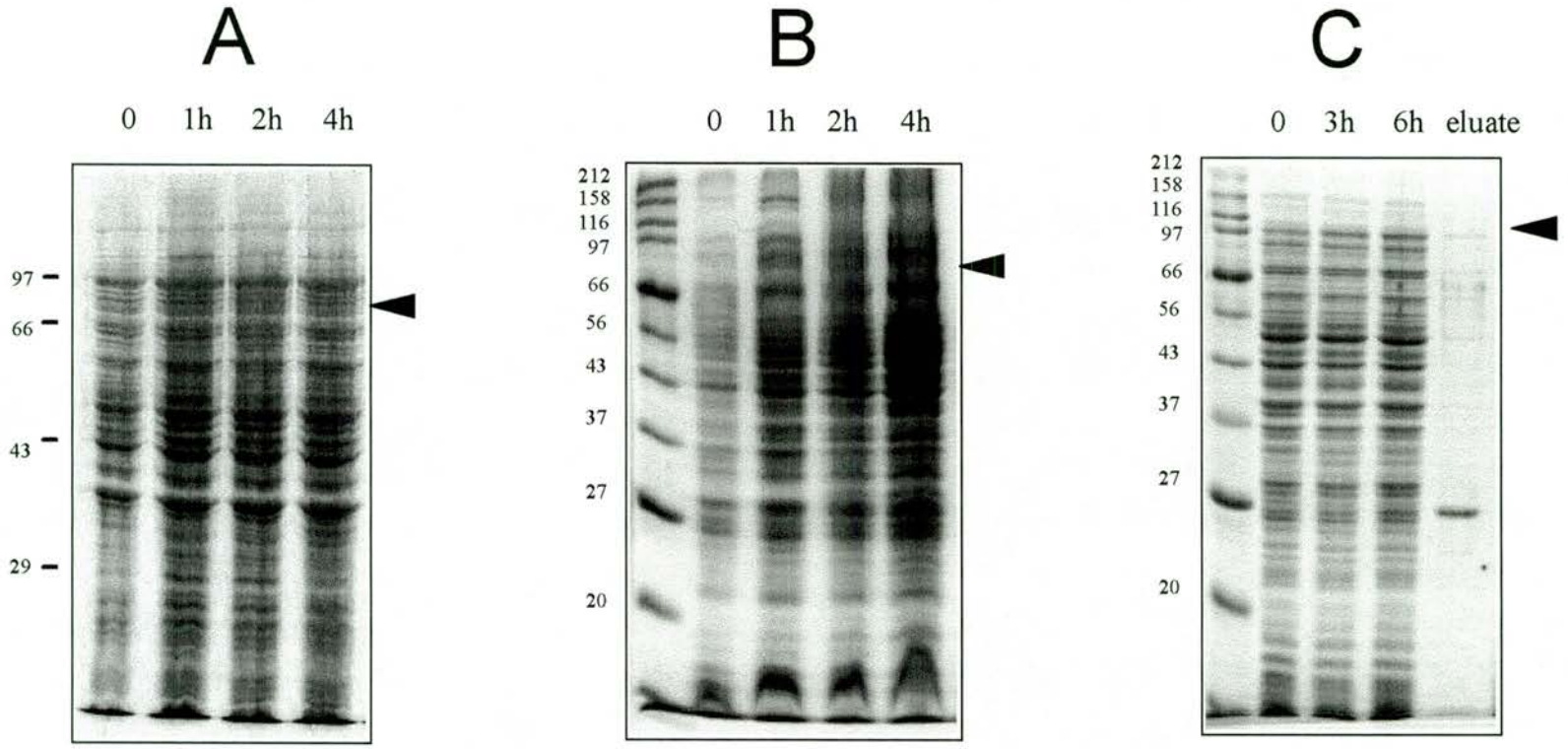


Figure 3-6 Expression of full-length P1 in bacteria

A. Expression of his-tagged P1 protein from plasmid pHP1 in BL21(DE3). Cells were harvested at 0, 1, 2 and 4 hours after addition of 0.5mM IPTG, and analyzed with SDS-PAGE.

B. Expression of his-tagged P1 protein using M13/T7 phage system.

C. Expression of P1 fused to GST from plasmid pGEX-P1 in DH5α. Proteins purified with glutathione-agarose beads (eluate) are also shown here.

The arrows show where the full-length fusion proteins should be.

be toxic to these cells. Therefore, an expression system with a more stringent control was chosen, in which *E. coli* JM109 was the host that harboured the plasmid pHP1. IPTG was added to a final concentration of 1 mM when OD_{600} reached 0.3. Expression was initiated one hour later by infecting cells with M13/T7 phage at an MOI of 5. The results are shown in Figure 3-6 (panel B).

pGEX-P1

E. coli DH5α was transformed with plasmid pGEX-P1 and grown to OD_{600}= 0.5. Expression was induced by adding IPTG to a final concentration of 0.2 mM. Upon analysis of SDS-PAGE, a very faint band of a size corresponding to the GST-P1 fusion protein was observed, which was absent before IPTG induction (Figure 3-6, panel C). The yield of full-length fusion protein was so low that it was barely visible on a Coomassie blue stained gel. This protein could be purified from cell lysate with glutathione beads, though the major protein eluted from beads had a size similar to that of GST. The identity of the 96 kDa protein was later confirmed to be the GST-P1 fusion protein, when antibody against P1 became available (Section 3.2.4). Unfortunately, the expression level was too low for a practical purification. Different *E. coli* strains and induction conditions (incubation temperature, IPTG concentrations, cell density when IPTG was added) were tested, but not much difference was observed (data not shown).

3.1.4 Expression of truncated forms of P1 protein in *E. coli*

In order to obtain enough protein to raise P1-specific antibody, it was decided to express only part of P1 to overcome the difficulty of expressing full-length P1 protein.

3.1.4.1 Construction of plasmids

A series of plasmids that encoded truncated forms of P1 with His-tag at their N-termini were constructed (Figure 3-7).

pHP1XE: pHP1 was restricted with *Xho* I and *Eco*R I. The 1.8 kb fragment was purified from agarose gel and ligated into pRSET B, similarly restricted.

HP1XbE: pHP1 was restricted with *Xba* I, treated with T4 DNA polymerase, and then restricted with *Eco*R I. The 1.15 kb fragment was purified and ligated into pRSET C, restricted with *Pvu* II and *Eco*R I.

pHP1HE: pHP1 was restricted with *Hind* III, treated with T4 DNA polymerase, and then restricted with *Eco*R I. The 0.64 kb fragment was purified and ligated into pRSET B, restricted with *Pvu* II and *Eco*R I.

pHP1PE: pHP1 was restricted with *Pst* I and *Eco*R I. The 0.28 kb fragment was purified and ligated into pRSET C, restricted with *Pst* I and *Eco*R I.

pHP1BH, pHP1XH and pHP1XbH: pHP1, pHP1XE and pHP1XbE were all restricted with *Hind* III, and then re-circularized with T4 DNA ligase.

3.1.4.2 Expression of truncated P1 proteins fused to His-tag

Expression of truncated forms of P1 protein using these plasmids was carried out with the *E. coli* JM109-M13/T7 system, described above. Generally, *E. coli* cells carrying these plasmids grew slowly, suggesting a toxic effect of recombinant proteins on the

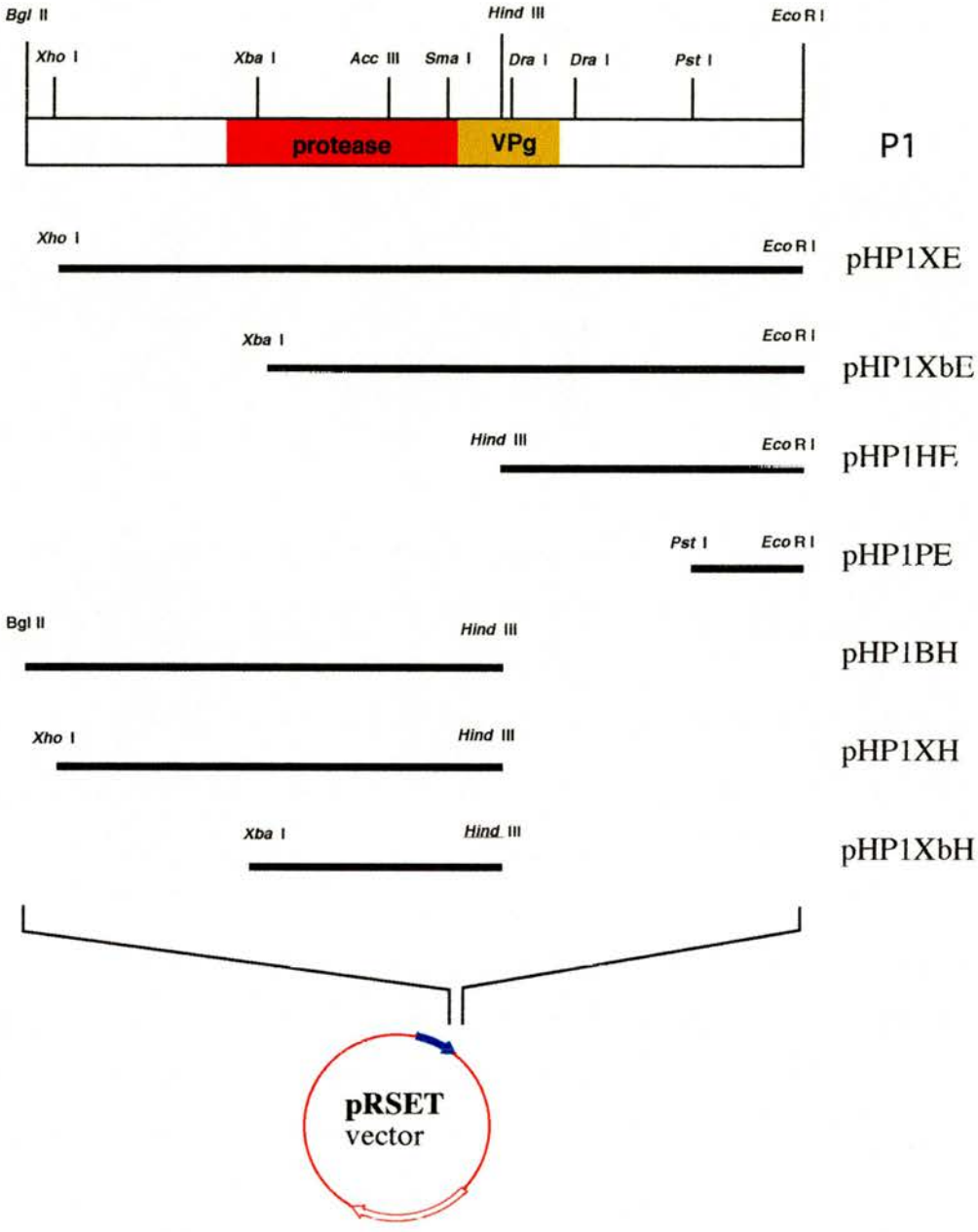


Figure 3-7 Plasmids expressing different regions of P1 as His-tagged proteins

cells. As a precaution, only freshly transformed cells were used in expression experiments. As shown in Figure 3-8a, for most constructs, the His-tagged proteins were not expressed sufficiently to be identified on Coomassie blue stained gels. The His-tagged protein HP1PE (expressed from pHP1PE plasmid), with an apparent molecular weight of ~18 kDa, had the highest expression level and could be purified under both native and denaturing conditions (Figure 3-8b).

3.1.4.3 Expression of truncated P1 proteins fused to GST

It was suspected that the poor expression of His-tagged P1-derived proteins was, at least partially, due to the toxicity of the proteins. Fusion to GST might reduce the toxic effects, and thus improve the expression. To this end, four plasmids encoding truncated P1 proteins fused to GST were constructed (Figure 3-9).

pGEX-P1A: pGEX-P1 was restricted with *Xba* I and *Eco*R I, blunted with Klenow fragment, and then re-circularized. The polypeptide encoded in this construct contained the N-terminal third of P1 and was mainly hydrophobic.

pGEX-P1B: P1 sequence was amplified from pSAB53 in a PCR, using oligonucleotides 304 and 305. The PCR product was restricted with *Xba* I and *Hind* III. The 0.5 kb fragment was purified and ligated into pGEX-KG, similarly restricted. The DNA insert in this plasmid contains most of the putative protease domain and the G355 to V mutation at the GXS**G** motif.

pGEX-P1C: P1 sequence was amplified from pSAB53 in a PCR, using oligonucleotides 304 and 305. The PCR product was restricted with *Dra* I and *Eco*R I. The 0.46 kb fragment was purified and ligated into pGEX-KG, restricted with *Sma* I and

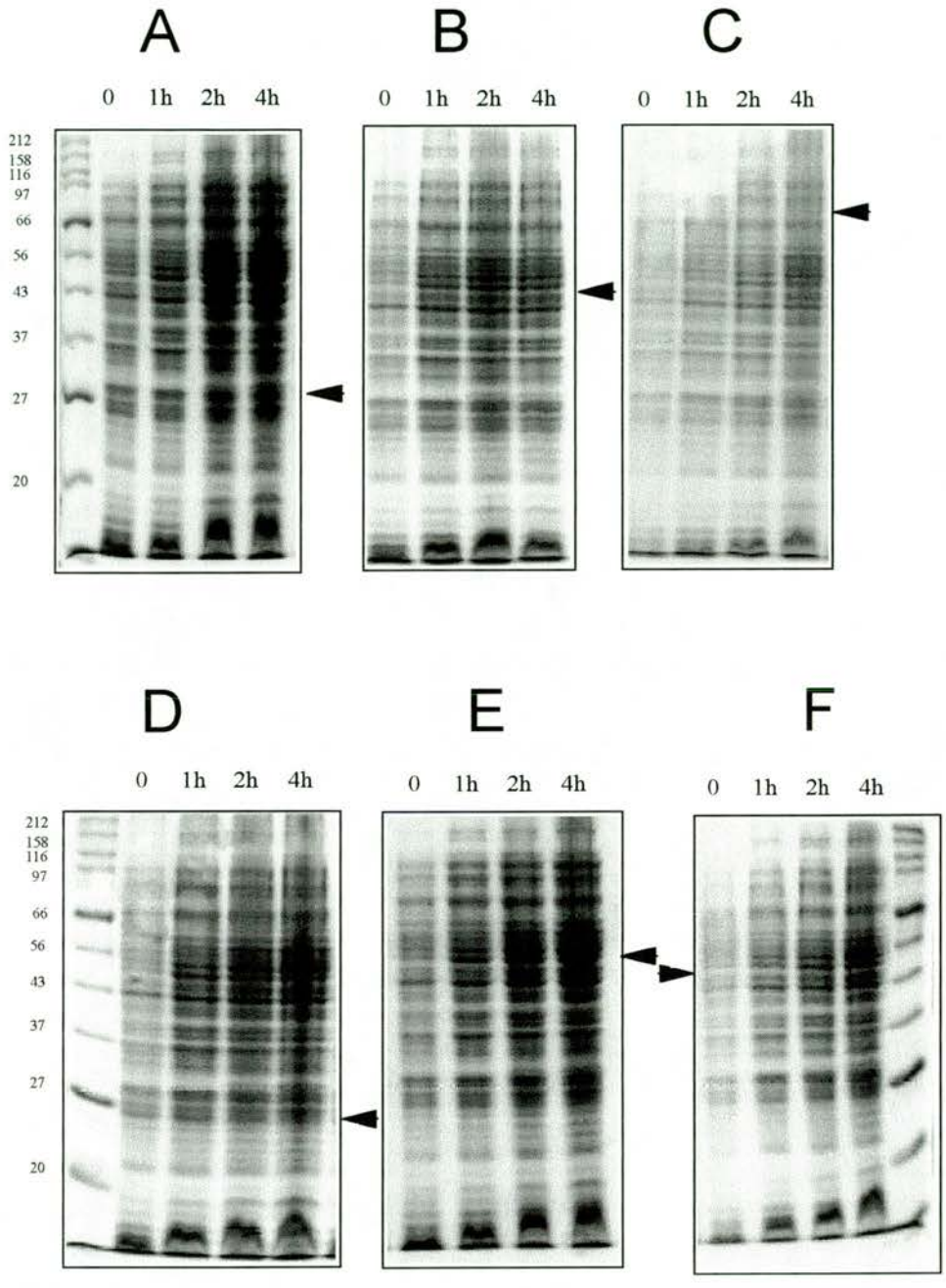


Figure 3-8a Expression truncated forms of his-tagged P1 proteins
A.pHP1HE; B. pHP1XbE; C. pHP1XE; D. pHP1XbH; E. pHP1BH; F. pHP1XH (JM109-M13/T7 system was used). The arrows show where his-tagged proteins should be.

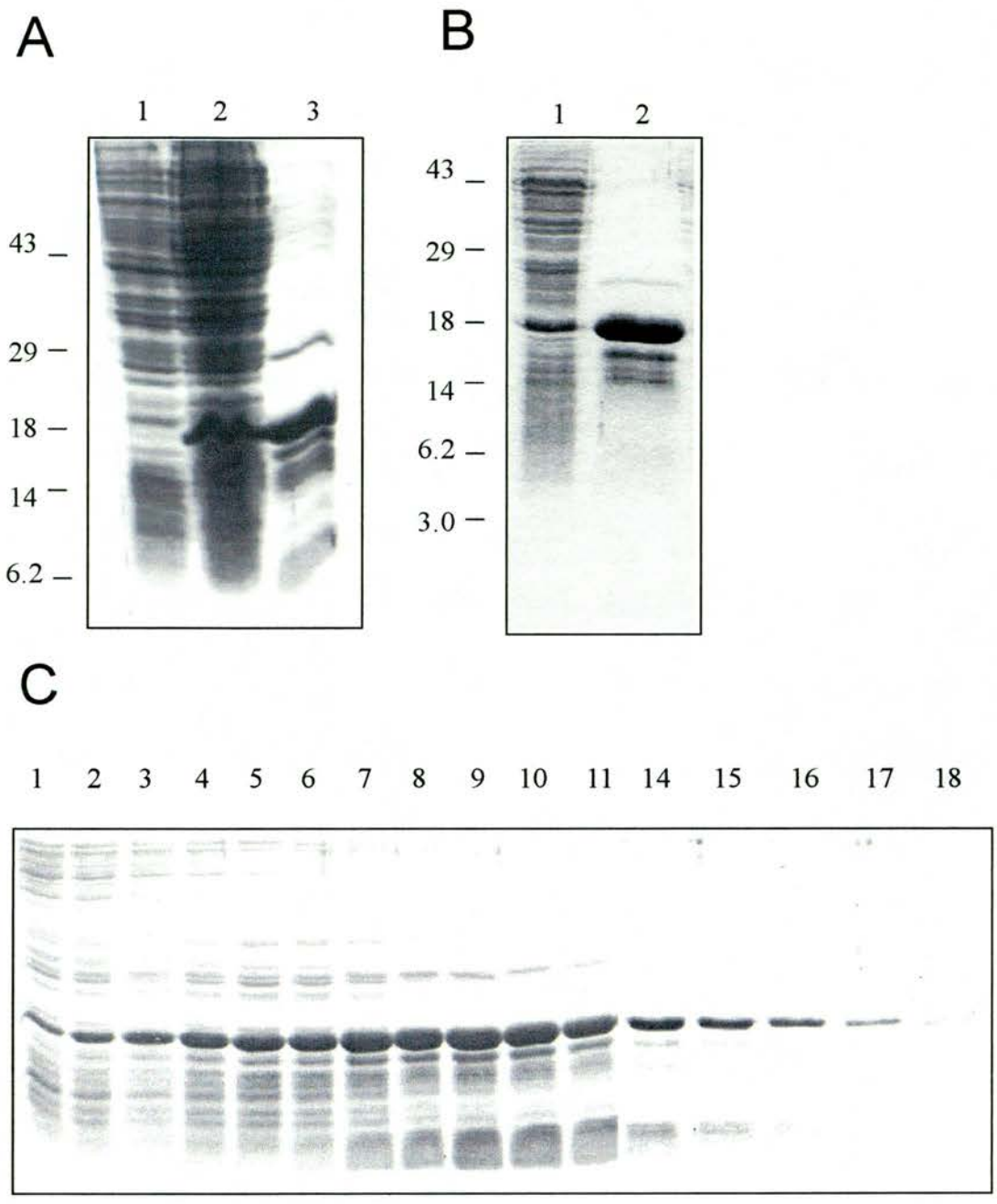


Figure 3-8b Expression and purification of HP1PE

His-tagged protein HP1PE was expressed from pHP1PE in JM109-M13/T7 system and analytically purified under native (A) and denaturing (B) conditions.

A. Lane 1, uninduced cells; lane 2, after addition of IPTG; lane 3, purification on Ni-beads.

B. Lane 1, extract of cells induced with IPTG; lane 2, eluate from Ni-beads.

C. Large scale purification under native conditions. Lane 1-18, different fractions eluted with 40-250 mM imidazole/1.0 M NaCl.

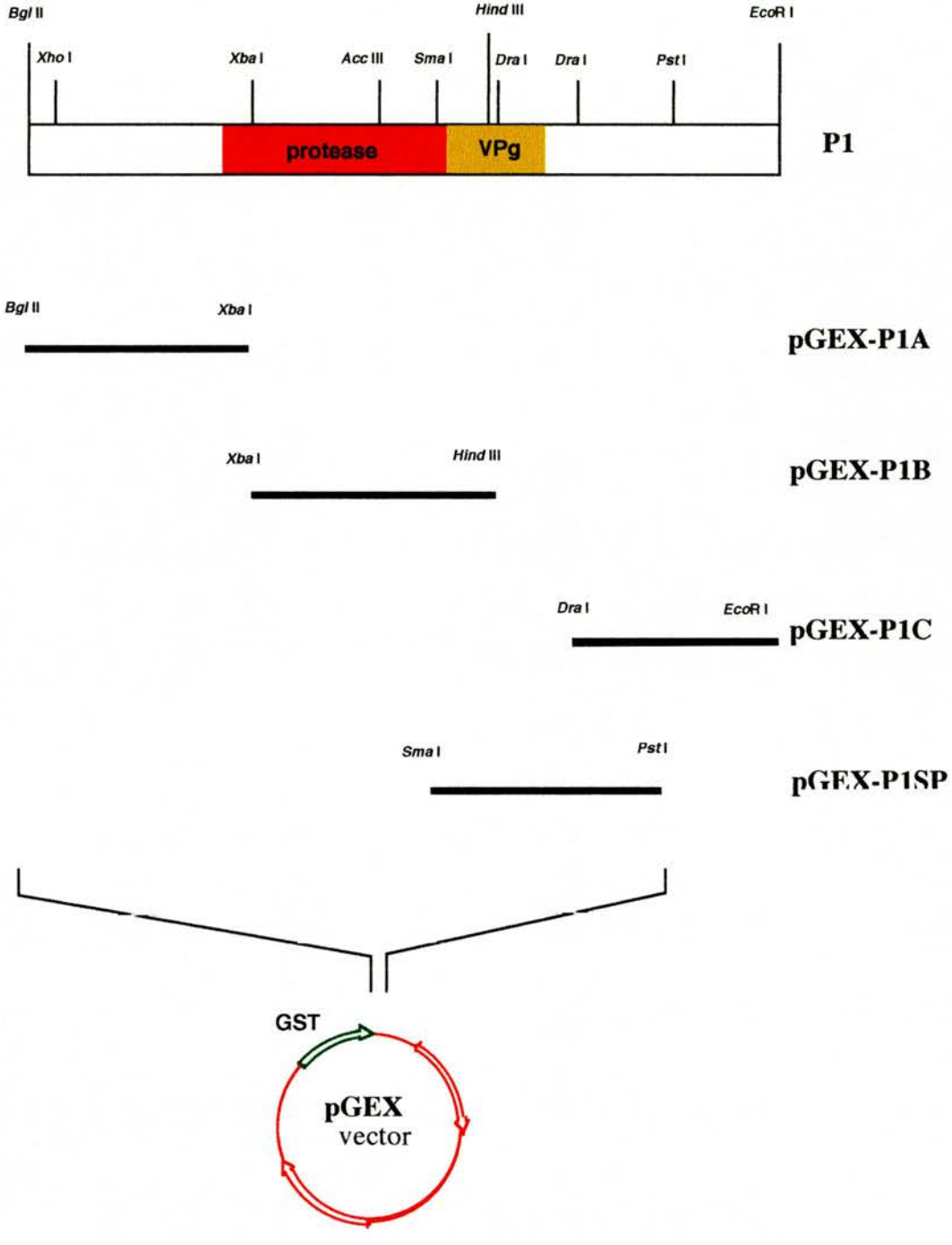


Figure 3-9 Plasmids expressing different regions of P1 as GST-fusion proteins

*Eco*R I. The DNA insert in this plasmid encodes the C-terminal **17** kDa of P1, from the frameshift site to the C-terminus of P1.

pGEX-P1SP: pHP1XE was restricted with *Xho* I and *Pst* I. The 1.5 kb P1 sequence was purified and ligated into pRSET B, similarly restricted. The plasmid obtained, named pHP1XP, has an *Eco*R I site downstream of the *Pst* I site. pHP1XP was subsequently restricted with *Sma* I and *Eco*R I and the restriction fragment encoding the C-terminal region of PLRV P1 (delimited by the *Sma* I and *Pst* I sites within P1) was purified. This fragment was ligated together with the *E. coli* expression vector pGEX-KT, similarly restricted.

For pGEX-P1A, pGEX-P1B and pGEX-P1C, the screening of recombinant plasmids was performed by mini-preps of plasmids and subsequent digestion with restriction enzymes. For pGEX-P1SP, this was carried out by directly identifying transformants that expressed fusion proteins. The ligation reaction was used to transform competent *E. coli* cells, and transformant colonies were randomly picked into 10 ml LB/ampicillin medium and incubated with agitation until visibly turbid (3-5 hours). IPTG was then added to a final concentration of 0.4 mM. Incubation continued for a further 4 hours, when cells were harvested and fusion proteins purified with glutathione beads. This method can identify transformants that express fusion proteins rapidly and reveal whether these proteins can be purified by glutathione beads. Figure 3-10 Panel B shows the eluted proteins analyzed with SDS-PAGE in the screening of pGEX-P1SP. Apparently, the GST fusion protein is readily purified from the cells. The high background was due to the minimal washing step carried out during purification.

The expression of fusion proteins GST-P1A, GST-P1B and GST-P1C are shown in Figure 3-10 (panel A). GST-P1B and GST-P1C were synthesized efficiently at 25 °C,

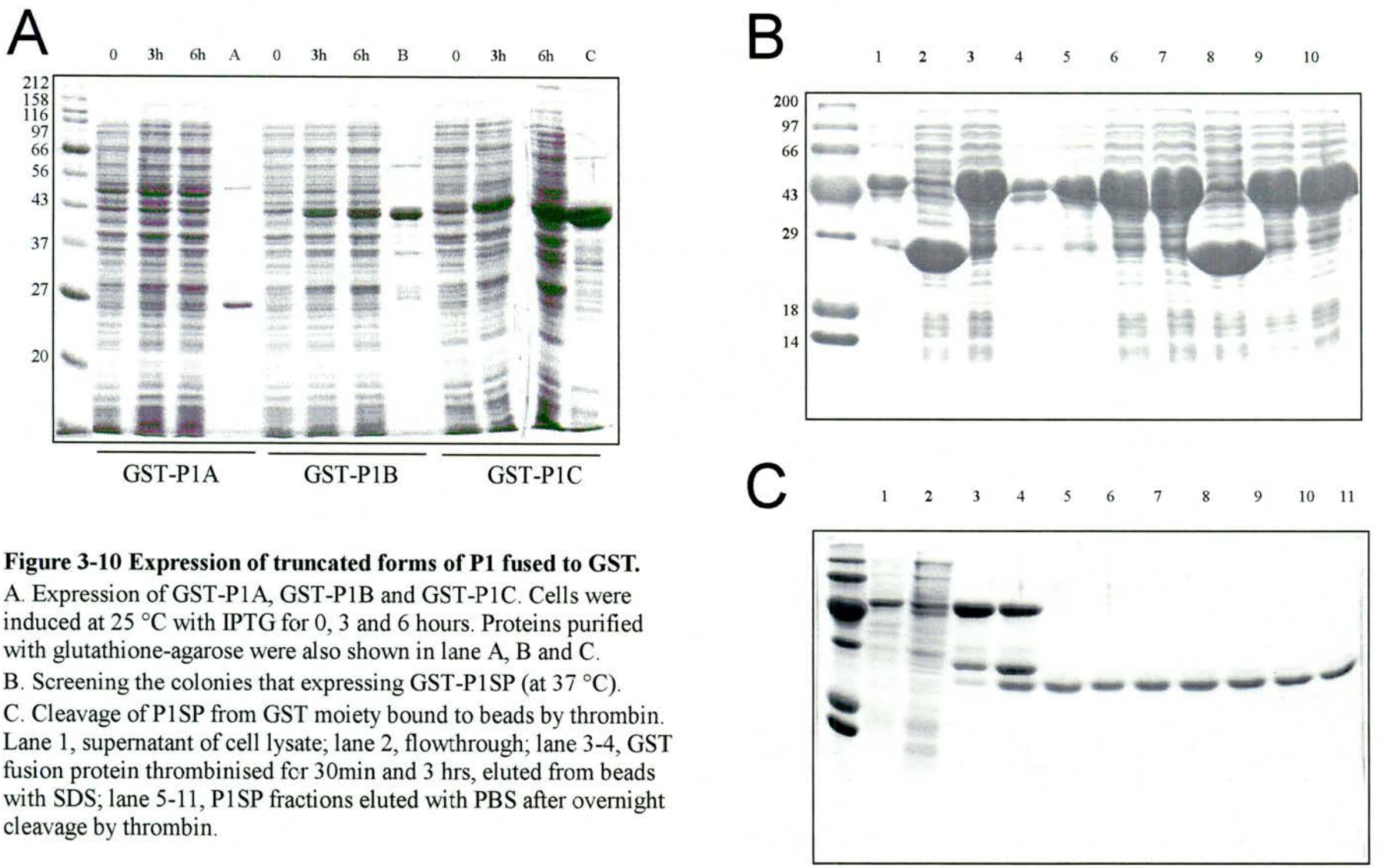


Figure 3-10 Expression of truncated forms of P1 fused to GST.

A. Expression of GST-P1A, GST-P1B and GST-P1C. Cells were induced at 25 °C with IPTG for 0, 3 and 6 hours. Proteins purified with glutathione-agarose were also shown in lane A, B and C.

B. Screening the colonies that expressing GST-P1SP (at 37 °C).

C. Cleavage of P1SP from GST moiety bound to beads by thrombin. Lane 1, supernatant of cell lysate; lane 2, flowthrough; lane 3-4, GST fusion protein thrombinised for 30min and 3 hrs, eluted from beads with SDS; lane 5-11, P1SP fractions eluted with PBS after overnight cleavage by thrombin.

and could be purified with glutathione beads. Expression at 37 was attempted, but expressed GST-P1B formed inclusion bodies, while the expression of GST-P1C at 37°C was poor (data not shown). Expression of GST-P1A was quite low, with GST being the main product.

3.1.4.4 Large scale preparation of P1SP protein

Large-scale purification of fusion protein was carried out using affinity chromatography. The P1-derived moiety, P1SP of 18 kDa, was released from the GST part by protease thrombin, when the fusion protein was still bound to the beads. Figure 3-10 (panel C) shows the results of the purification. The cleavage was completed after overnight incubation in PBS buffer. The P1SP protein was eluted by passing through 10-15 bed volumes of PBS/0.5 M NaCl. The yield of P1SP was about 3-5 mg per litre of culture.

3.1.5 Expression the active domain of P1 protease

The experiment described below was based on the data discussed in Section 3.3-3.5. It is included here only because the nature of the work is similar to the rest of this section: expressing P1-derived proteins in *E. coli*.

With the baculovirus system, we demonstrated that the serine protease domain was responsible for the processing of both P1 and P1/2 fusion proteins (Section 3.3). Site-directed mutagenesis has identified the essential residues at the active centre (Section 3.5). Both the type of catalytic residues and the distance between them were typical of viral chymotrypsin-like serine proteases, which normally have an active domain of about 20-30 kDa. Based on these data, a plasmid was constructed in an effort to express the active domain of the P1 serine protease.

The P1 sequence between the *Bss*H II and *Hind* III was inserted into pGEX-KT. This segment of DNA covers the proposed protease domain with small extensions at both N- (presumably) and C- termini of the domain. The His-234 has been shown not to be required by the proteolytic activity, so probably the domain N-terminal to this residue is not essential for proteolytic activity. Computer analysis has shown that polypeptide encoded by sequences upstream of the *Bss*H II site are very hydrophobic, which might form a hydrophobic core too big to be included into the protease domain. Sequences downstream of the *Bss*H II site encode a series of hydrophilic residues, which have a high surface probability and possibly form the region connecting the proteolytic domain with N-terminal part of P1 protein. The 3' end of the DNA insert extends into the VPg motif, so the C-terminus of the protease domain is definitely encoded within it.

The construction of plasmid pGEX-P1BsH is shown in Figure 3-11. Plasmid pGEM-P1 was restricted with *Bss*H II and *Hind* III, then treated with Klenow fragment. In pGEM-P1, entire P1 sequences were inserted into pGEM-T Easy vector (Section 3.4). The blunted *Bss*H II-*Hind* III fragment was purified from agarose gel and ligated into pGEX-KT, restricted with *Sma* I and treated with calf alkaline phosphatase. Because all ends in the ligation reaction were blunt, the insert could have been inserted in two orientations. Screening was carried out with the mini-expression method, which had the extra advantage of identifying the colonies having the right orientation of the insert, because plasmids carrying inserts in wrong orientation produce a GST protein with a small C-terminal extension.

E. coli cells were directly lysed with Laemmli buffer after 4 hours induction with IPTG (0.4 mM final concentration) at 37 °C. When a pilot purification was carried out, the fusion protein expressed at 37 °C was found mainly in the insoluble pellet after

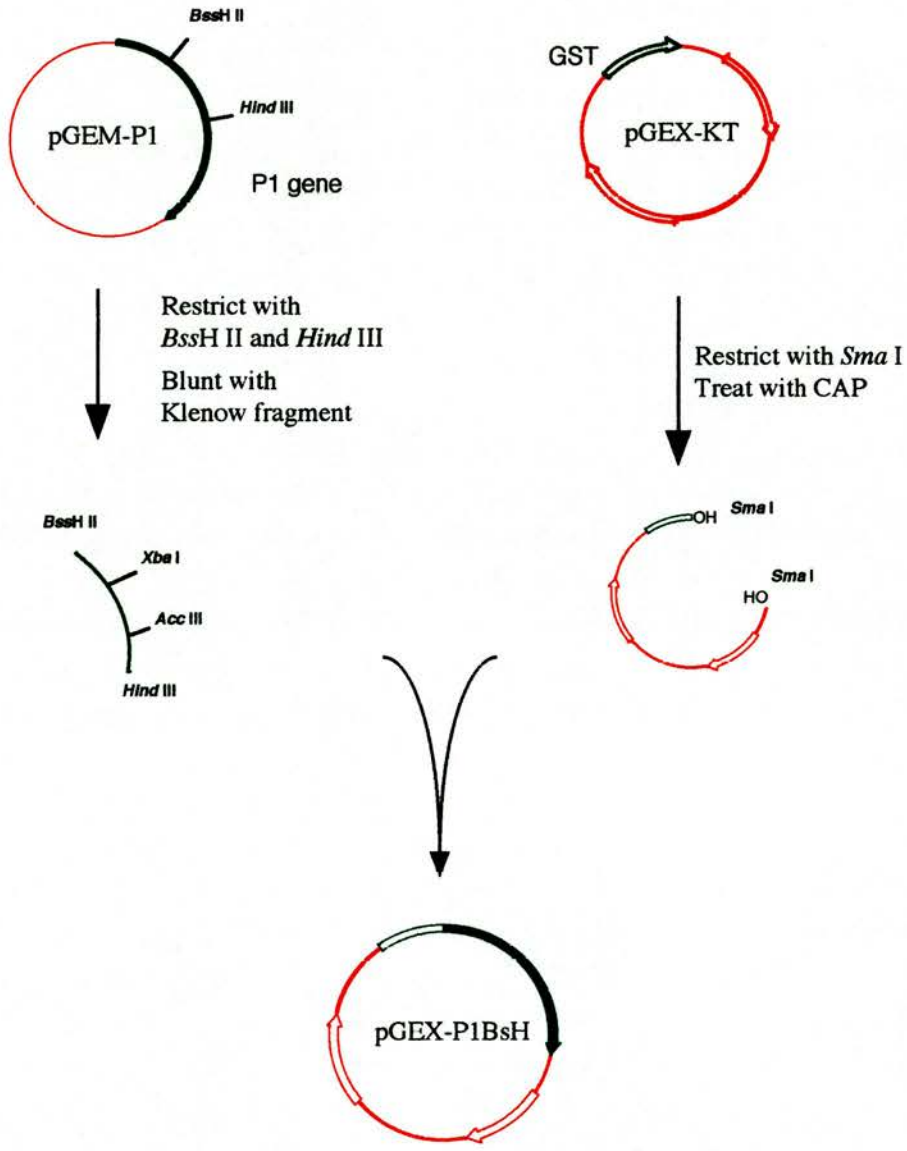


Figure 3-11 Construction of pGEX-P1BsH

sonication, and could not be purified with glutathione beads (Data not shown). However, expression at 25 °C gave interesting results. As shown in Figure 3-12, three hours after adding IPTG (0.2 mM final concentration), the recombinant protein produced was mainly the full-length GST-P1BsH (55 kDa); as the expression went on for a further 3 hours, the majority of the full-length GST-P1BsH disappeared, and the amount of two smaller protein of 29 kDa and 25 kDa increased considerably. Test purification showed that the 29 kDa protein could be purified with glutathione beads, whereas the 25 kDa protein could not. We speculate that the 29 kDa and 25 kDa proteins were the cleavage products of the 55 kDa full-length fusion protein. Since the GST itself is about 26 kDa, the 29 kDa protein should contain a short C-terminal extension derived from P1 protein.

The protease responsible for the cleavage may be either the proteolytic domain within the P1BsH, or a host protease. This is similar to the situation when the full-length P1 is processed in insect cells (Section 3.3). To address this question, a similar approach was used. DNA fragment encoding a Gly→Val mutation at the catalytic centre which inactivates the proteolytic activity of P1 protease, was released from pHP1XH and inserted into pGEX-KT to generate plasmid pGEX-P1GVBsH, in a way similar to the construction of pGEX-P1BsH. When *E. coli* carrying pGEX-P1GVBsH was induced with IPTG, only full-length fusion protein was produced at 25°C. This observation hints that the cleavage of full-length P1BsH is mediated by the serine protease domain of P1(See Section 3.3 and Discussion in this Section).

The effects of IPTG concentration and cell density on expression were also investigated. Expression was better when IPTG was added at OD_{600} of 0.7 than at 1.5, and 0.3 mM IPTG induced higher levels of expression than 0.1 mM IPTG.

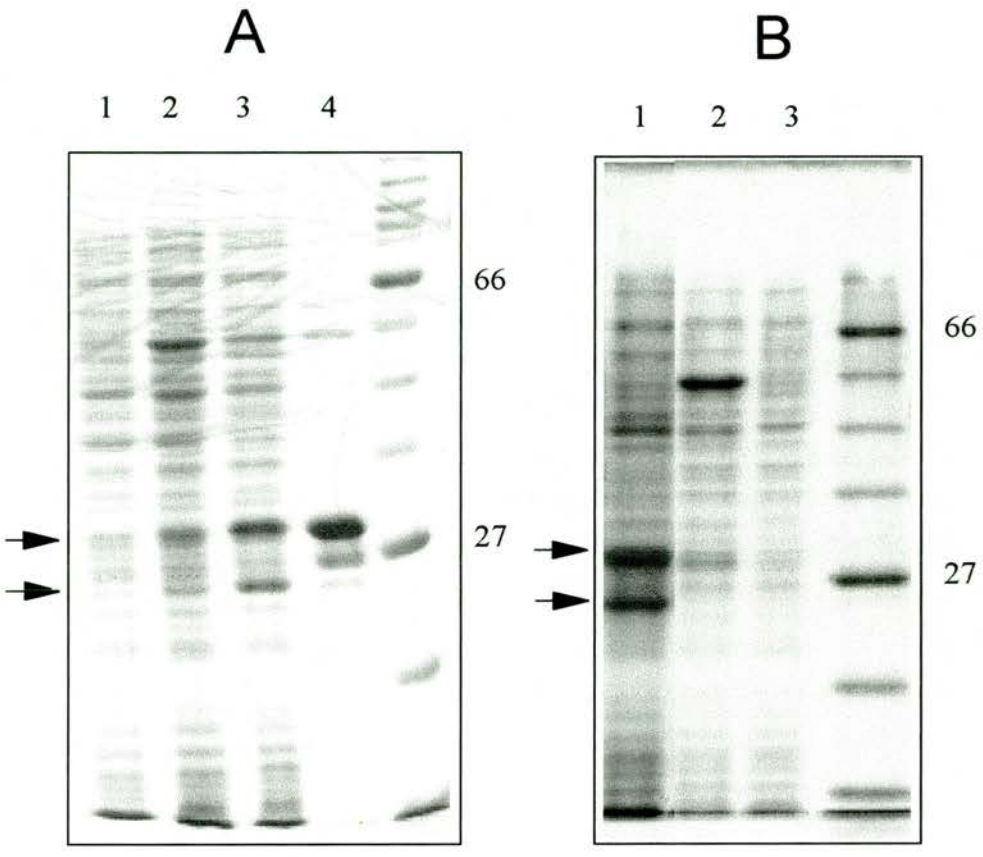


Figure 3-12 Expression of active domain of P1 protease (25°C)

A. Expression of wild-type GST-P1BsH. Lane 1, uninduced cells; lane 2 and 3, induced by IPTG for 3 and 6 hrs; lane 4, glutathione-agarose purification. The two cleavage products are indicated by arrows.

B. Expression of mutant form (Gly-355 to Val) of GST-P1BsH. Lane 1, wild-type GST-P1BsH control; lane 2, after IPTG induction; lane 3, uninduced cells. No cleavage product was observed in lane 2.

3.1.6 Discussion

This section describes the expression of proteins coded by P1 sequences in *E. coli*, which was carried out throughout the project. At first, the main objective was to produce enough protein for antibody production. After the proteolytic activity had been established in the baculovirus system, attempts were made to express the active domain of P1 protease.

3.1.6.1 pGEX system vs. pRSET system

Two prokaryotic expression systems were used in this study. In the pRSET system, target proteins were synthesized with a 6-histidine tag at the N-termini, under the control of T7 phage promoter. This system has the advantage of purifying recombinant proteins under both native and denaturing conditions. Purification under denaturing conditions is best suited when target proteins are accumulated in inclusion bodies. However, in my case, the main problem was the low level of expression. Only HP1PE, a truncated form of P1 (encoded by the sequence between *Pst* I and *Eco*R I sites), was expressed at a reasonably high level to allow purification on Ni-beads.

The pGEX system achieved most success in expressing P1-derived proteins. This might be due to several reasons. The 26 kDa GST moiety in the fusion protein usually helps to keep it soluble, and stabilizes the target protein if it is small. If a protein is toxic, it is more likely to be tolerated when fused to GST. Generally, purification using the glutathione column has very low background. Because purer proteins can be obtained using the pGEX system, it became the method of choice.

3.1.6.2 Factors that affect the expression

Probably, the most important single factor is the nature of the protein to be expressed. It is very difficult to predict how well a novel protein will be synthesized in any expression system. Full-length P1 protein was not expressed well in either of the two systems, so different parts of the P1 gene were tested, with varying results. Only the results with pGEX system are discussed here.

The expression of the N-terminal one-third of P1 was not detected in Coomassie blue stained gel. No protein that contains this part of P1 was ever found to be expressed well. This region of P1 is very hydrophobic and might be toxic to the cells. It is also interesting to note that codon usage might affect the expression level. Table 3-1 shows the comparison of the codons used by *E. coli* and *Solanum tuberosum* and the frequencies of each codon used in the N-terminal region of P1 (the first 208 residues, designated N-P1). Codons for arginine are used with severe bias by PLRV. This part of P1 seems to have evolved for expression in the host plant, but not in *E. coli*.

Table 3-1 Codon usage of arginine

codons of arginine	CGU	CGC	CGA	CGG	AGA	AGG
usage in *E. coli*	37%	38%	7%	10%	5%	3%
usage in *S. tuberosum*	20%	8%	12%	6%	32%	23%
frequency in N-P1	0	0	1	2	1	5

The middle one-third of P1 was the protease domain. Plasmid pGEX-P1B contains sequences encoding most of this domain and the Gly355 to Val mutation. The GST-

fusion protein was produced to high level, as insoluble inclusion bodies at 37°C, becoming soluble when expressed at 25°C. The result with pGEX-P1BsH was similar, but an active viral protease was probably synthesized this time. The VPg part of P1 was most successfully expressed as a GST fusion protein. It was produced in large amounts, was soluble, and was easily purified.

Expression conditions were also found to be important. Temperature of incubation after addition of IPTG had the most dramatic effects on the solubility of fusion proteins. At lower temperatures, recombinant proteins were synthesized more slowly, had more time to fold correctly and were more likely to be soluble. Similar but lesser effects were found for the concentration of IPTG. Low concentrations of IPTG improve the solubility of fusion protein, but they also lowered the total expression levels (data not shown).

3.1.6.3 An active protease domain expressed in *E. coli*

Interesting results were obtained with plasmid pGEX-P1BsH, which contains the wild type protease domain. At 37°C, the fusion protein was produced as inclusion body; but at 25°C, a processing event was clearly observed during the 6- hour expression. Two lines of evidence suggest that the protease domain of P1 was responsible for the cleavage. It was found in insect cells, that the N-terminal part of P1 was not required by the proteolytic activity of P1 protease, as virus BacP1BsE (expressing truncated P1 protein with the N-terminal 184 residues being deleted) could still carry out the self-cleavage at the Pro/VPg site (data not shown). Plasmid pGEX-P1GVBsH, which encodes the mutant version of P1 protease (Gly355 to Val), only produced full length GST-fusion protein at 25°C, no processing being observed. Final confirmation will be the purification of the active P1 protease and testing its activity *in vitro*.

3.2 Characterization of Antibody

3.2.1 Introduction

The use of antisera plays a very important role in the studies of polyprotein processing. Viral proteins produced *in vivo* or *in vitro* can be identified by immunoprecipitation, immunoblotting, immunofluorescence/ immunogold-labelling, ELISA. For studies of polyprotein processing of foot-and-mouth disease virus (FMDV), immuno-precipitation has been shown to be very useful (Harris *et al.*, 1981; Grubman *et al.*, 1984; Strebel *et al.*, 1986; Vakharia *et al.*, 1987). Monospecific antisera were able to ascertain precursor-product relationships in the Sindbis virus polyprotein processing pathway (Hardy and Strauss, 1988, 1989).

Most antisera specific for PLRV were raised against purified viral particles or viral structural proteins. Using poly- and monoclonal antisera, Bahner *et al.* (1990) demonstrated that the 80 kDa P3/5 readthrough protein gave rise to a minor 53 kDa component of viral particles, presumably by the loss of C-terminal sequences. The processing of non-structural proteins of PLRV has been suspected for a long time (Mayo *et al.*, 1989; Miller *et al.*, 1995). Some protein bands were observed when viral RNA was translated in wheat germ extract, but whether they were the products of protein processing was unknown due to the lack of antibody specific for the three non-structural proteins encoded in the 5-proximal half of the RNA genome (Mayo *et al.*, 1989). Later, it was shown (van der Wilk *et al.*, 1997) that P1 protein was a polyprotein, which must be processed to give rise to the VPg. Production of an antiserum specific for P1 protein should help address some of these questions.

The protein sequence of P1 was analyzed using PEPTIDESTRUCTURE of GCG (Figure 3-13). It shows the C-terminal region of P1, starting from the VPg domain, is highly antigenic, and presumably immunogenic. To prepare antisera, four candidates were available as possible antigen, namely: (His)HP1PE, (GST)P1B, (GST)P1C and (GST)P1SP. The His-tagged HP1PE was the C-terminal 14.5 kDa of P1, whose function was unknown. Similar is the GST-fusion protein (GST)P1C. The GST-fusion P1B contained most of the proposed protease domain, but it contained the Gly→Val mutation at the GXS**G** motif, so P1SP was finally chosen as antigen. Since P1SP contains the important VPg domain and encompasses the frameshift site, the antibody raised against it should recognize both P1 protein and P1/2 transframe protein.

After the preparation of the antibody used in this project had finished, the first report about the processing of P1 *in planta* was published (Prüfer *et al.*, 1999). In their studies, different domains of the P1 were synthesized as GST-fusion proteins and used to raise poly- and monoclonal antisera.

3.2.2 The raising of antibody specific for P1 protein

Protein P1SP, which was released from the GST moiety, was used to inoculate two rabbits by a standard procedure (See Materials and Methods). Prior to inoculation, a pre-bleed serum sample was collected from each rabbit. The titres of the sera were monitored by ELISA, using purified P1SP as antigen and total cell extract from *E. coli* JM101 (harbouring the plasmid pGEM-3X and induced with IPTG) as negative control. Sera were collected 3 weeks after the fourth boost injection.

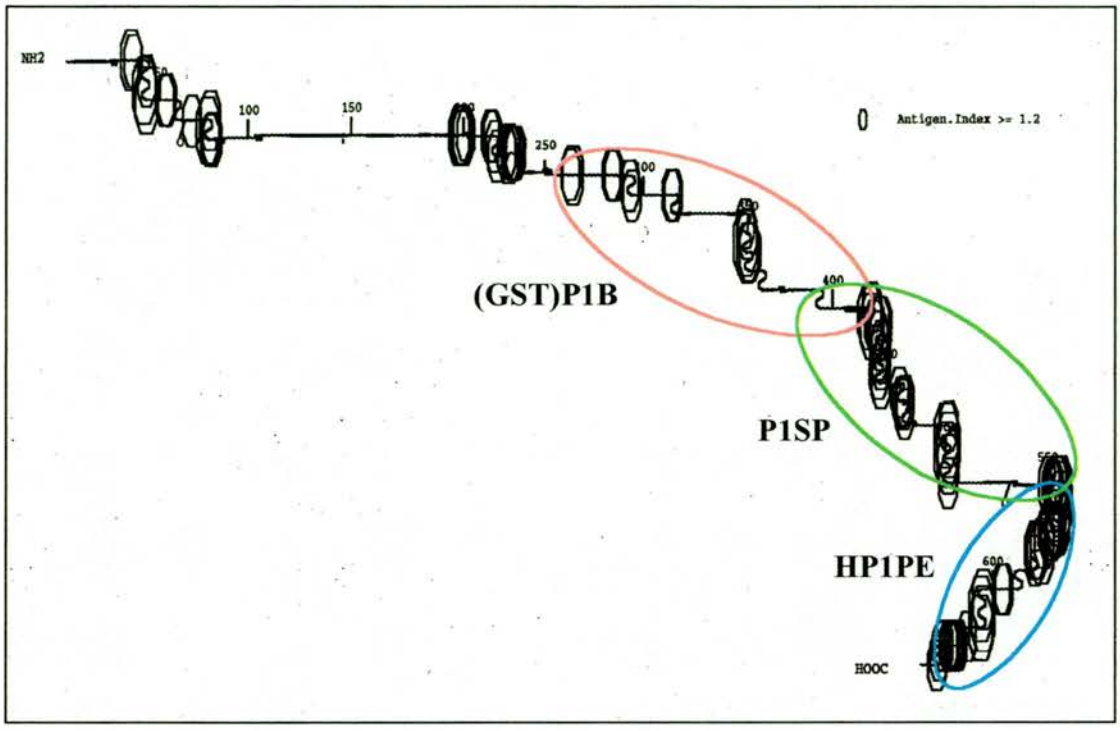


Figure 3-13 Prediction of antigenic regions of PLRV P1 protein
The amino acid sequence of P1 was analysed with PEPTIDESTRUCTURE of GCG package. Regions expressed in bacteria are shown by coloured circles.

3.2.3 Titres of antisera

The titres of the antisera were assayed using an indirect ELISA. The plates were coated with a saturating amount (200 ng per well) of P1SP or cell extract from *E. coli* JM101, followed by blocking with PBS/0.1%Tween-20/5%skimmed milk. A series of dilutions of antisera were added, followed by the addition of horseradish peroxidase-coupled goat anti-rabbit antibody. TMB detection system was used and the results are shown in Figure 3-14a, b. Antisera from both rabbits react strongly with purified P1SP, and the reaction with cell extract from *E. coli* JM101 was very weak. When the antisera from rabbit A and B were diluted at 512,000-fold and 256,000 fold respectively, the reading for the P1SP was still more than twice that for the control. The reaction between pre-bleeding sera and P1SP or cell extract of *E. coli* JM101 was very low as well.

The detection limit of P1SP in ELISA using anti-P1SP (rabbit A+B) was also determined (Figure 3-14c). The plate was coated with a series of dilutions of purified P1SP. The control was cell extract from *E. coli* JM101, similarly diluted. The limit was estimated to be about 0.4 ng (signal/noise > 2.0). When P1SP was added to the plate in the presence of a saturating amount of cell extract from *E. coli* JM101 (200ng per well), the limit was about 1.6 ng. The cell extract from *E. coli* JM101 had two effects, both of which reduced the signal/noise ratio. First, it competed with P1SP for the binding to the well (reducing the signal); second, it would bind to antisera non-specifically (increasing the noise).

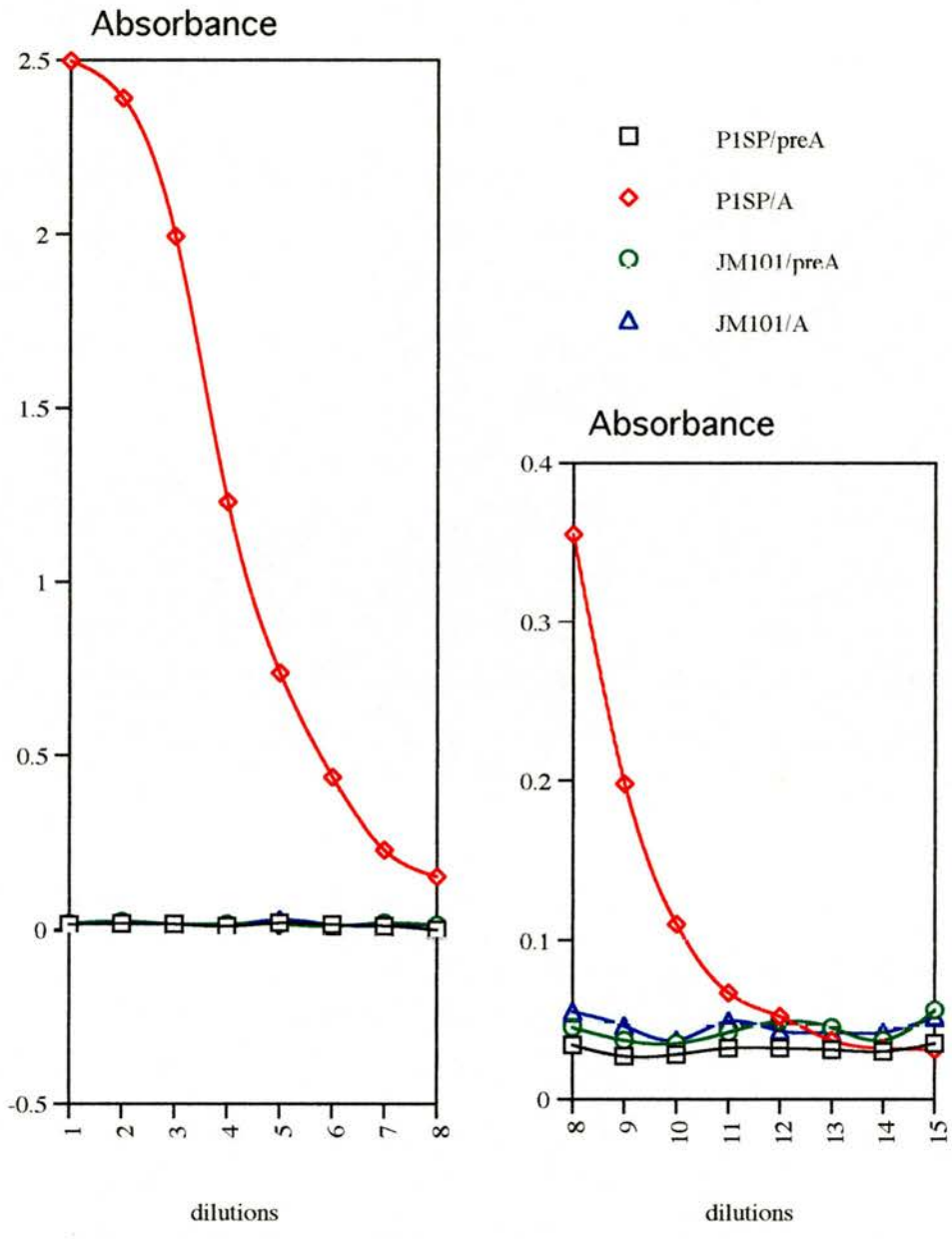


Figure 3-14a ELISA assay of anti-P1SP serum A
Purified P1SP and cell lysate from JM101 (control) were used as antigen (200 ng/well). The anti-P1SP serum A (A) and pre-bleed serum A (preA) were serially diluted (1:2): no. 1-15, from 1:1000 to 1:32,768,000. The colour was allowed to develop for 10 min (left) and 20 min (right) after the addition of the substrate.

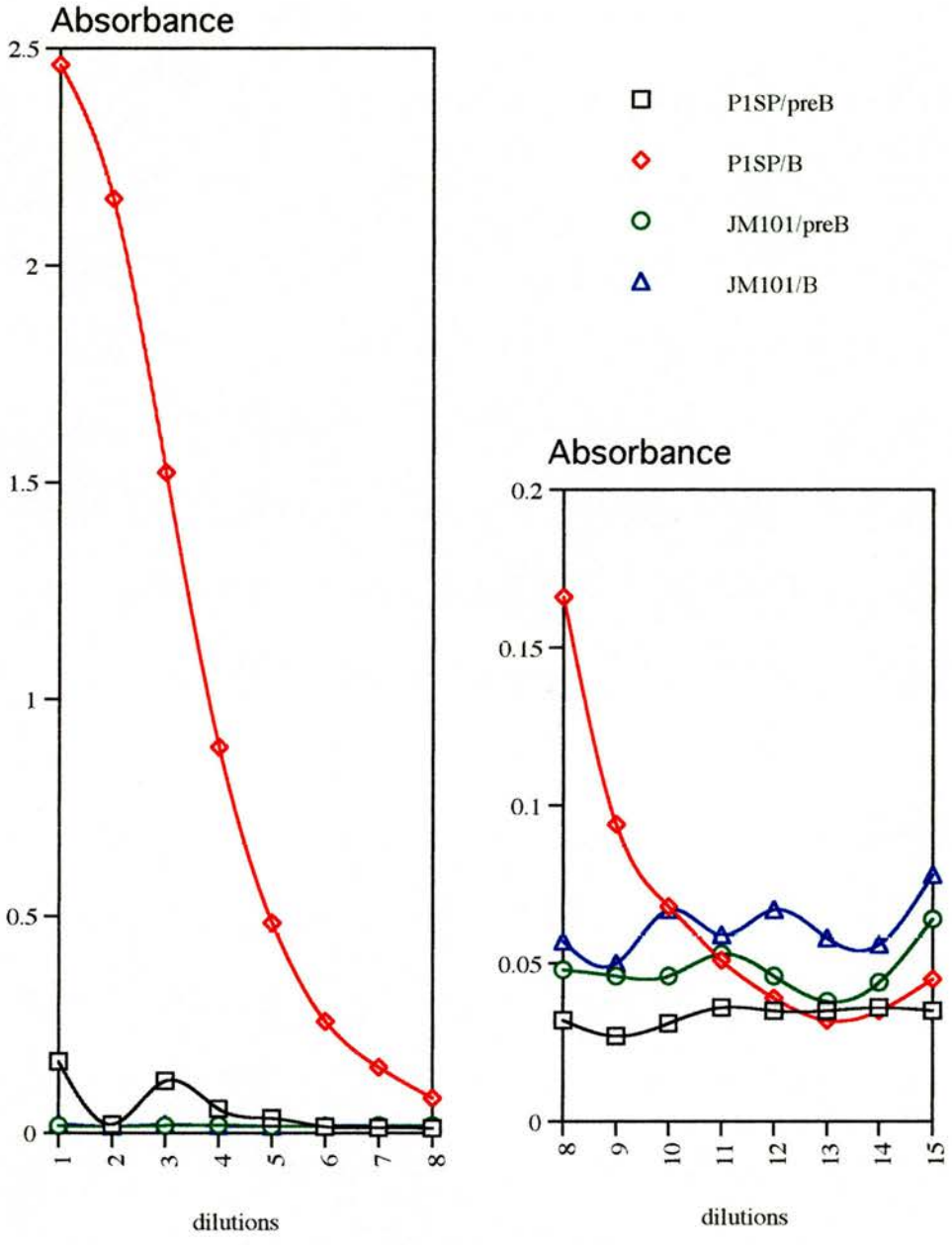


Figure 3-14b ELISA assay of anti-P1SP serum B
Purified P1SP and cell lysate from JM101 (control) were used as antigen (200 ng/well). The anti-P1SP serum B (B) and pre-bleed serum B (preB) were serially diluted (1:2): no. 1-15, from 1:1000 to 1:32,768,000. The colour was allowed to develop for 10 min (left) and 20 min (right) after the addition of the substrate.

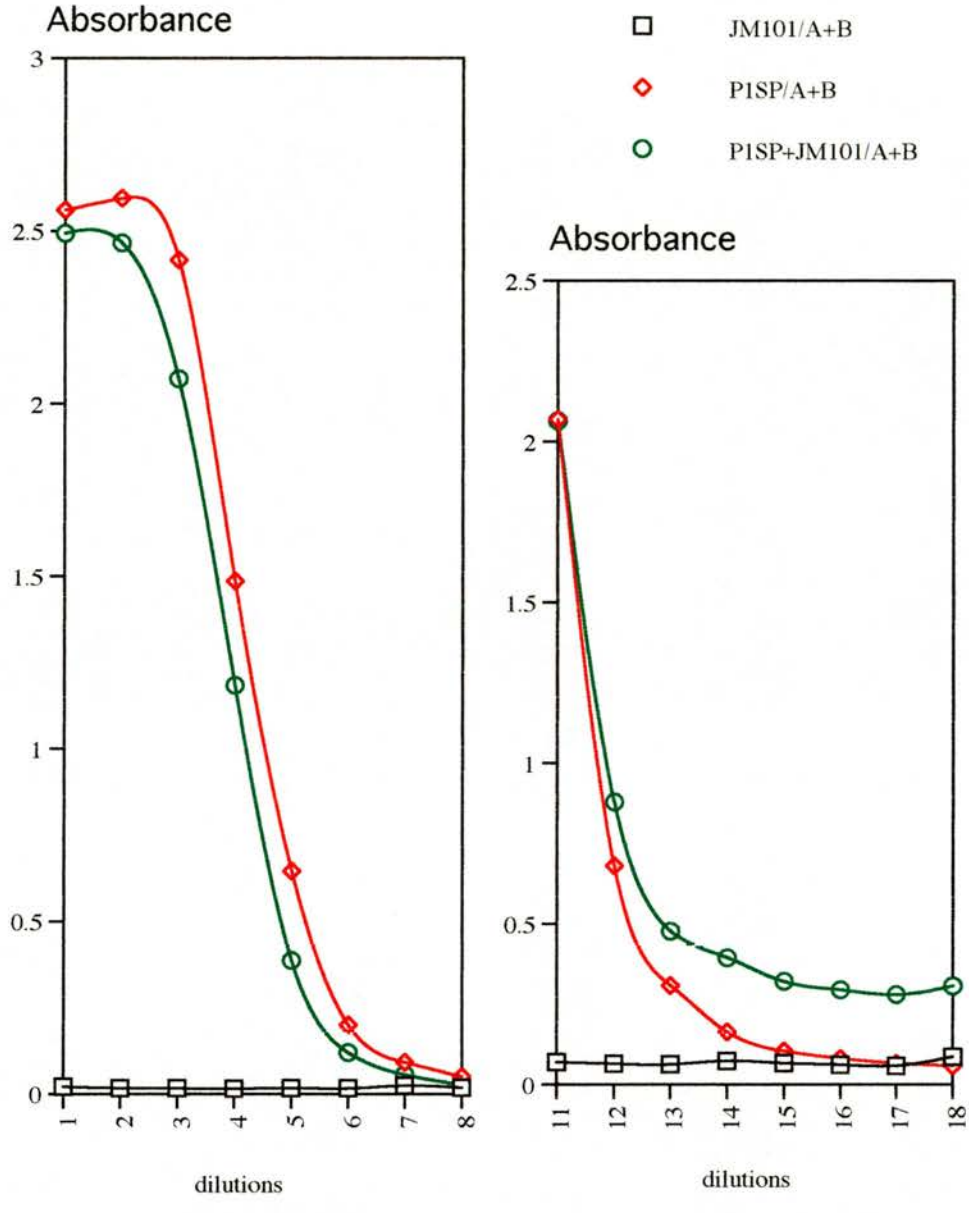


Figure 3-14c Detection limit of P1SP with anti-P1SP antisera (A+B)
Antisera A and B were pooled and used at 1:2000 dilution. The cell lysate from *E. coli* JM101 (control, black) and purified P1SP (red) were 1:2 diluted with PBS and used as antigen (no. 1-8, from 200 ng to 1.04 ng; no. 11-18, from 3.1 ng to 0.024 ng). For the green curve, purified P1SP was diluted as for the red curve except that 200 ng proteins from cell lysate of JM101 were also added into each well. The colour was allowed to develop for 10 min (left) and 20 min (right) after the addition of the substrate.

3.2.4 Detecting P1-derived proteins from *E. coli*

The antisera were also tested for reaction with P1-derived proteins synthesized in *E. coli* using Western blotting. The antiserum used in all Western-blotting experiments was a mixture of antisera A and B. As shown in Figure 3-15, the anti-P1SP reacted strongly with GST-P1SP and P1SP. It should be noted that the electro-blotting property of P1SP is unusual. P1SP passes easily through the membrane (nitrocellulose or PVDF), even when smaller marker proteins are retained on the membrane. This might be partly due to the low pI of P1SP protein, which carries more negative charges than most other proteins. As a result, both the voltage and time of transfer need to be reduced. For the semi-dry method, transfer at 6-8 V for 10-15 min gave good results.

Cell extract from *E. coli* transformed with pGEX-P1, pGEX-P1B and pGEX-P1C were also tested with immunoblotting (Figure 3-15, Panel C). These three plasmids encode GST fusion proteins that contain the whole or part of the domain that was used to raise anti-P1SP. For each plasmid, anti-P1SP reacted with a protein with the calculated molecular weight of the GST-fusion. For pGEM-P1, the signal was not strong, consistent with the low level of expression. Although GST-P1(GV)B was produced at high levels, the signal was not very strong, possibly because it only had 40 amino acids in common with P1SP. GST-P1C had a longer segment of peptide identical to P1SP and gave a strong reaction with the anti-P1SP. Some proteins smaller than the GST-fusions were also detected, and they could be the results of degradation, or premature termination.

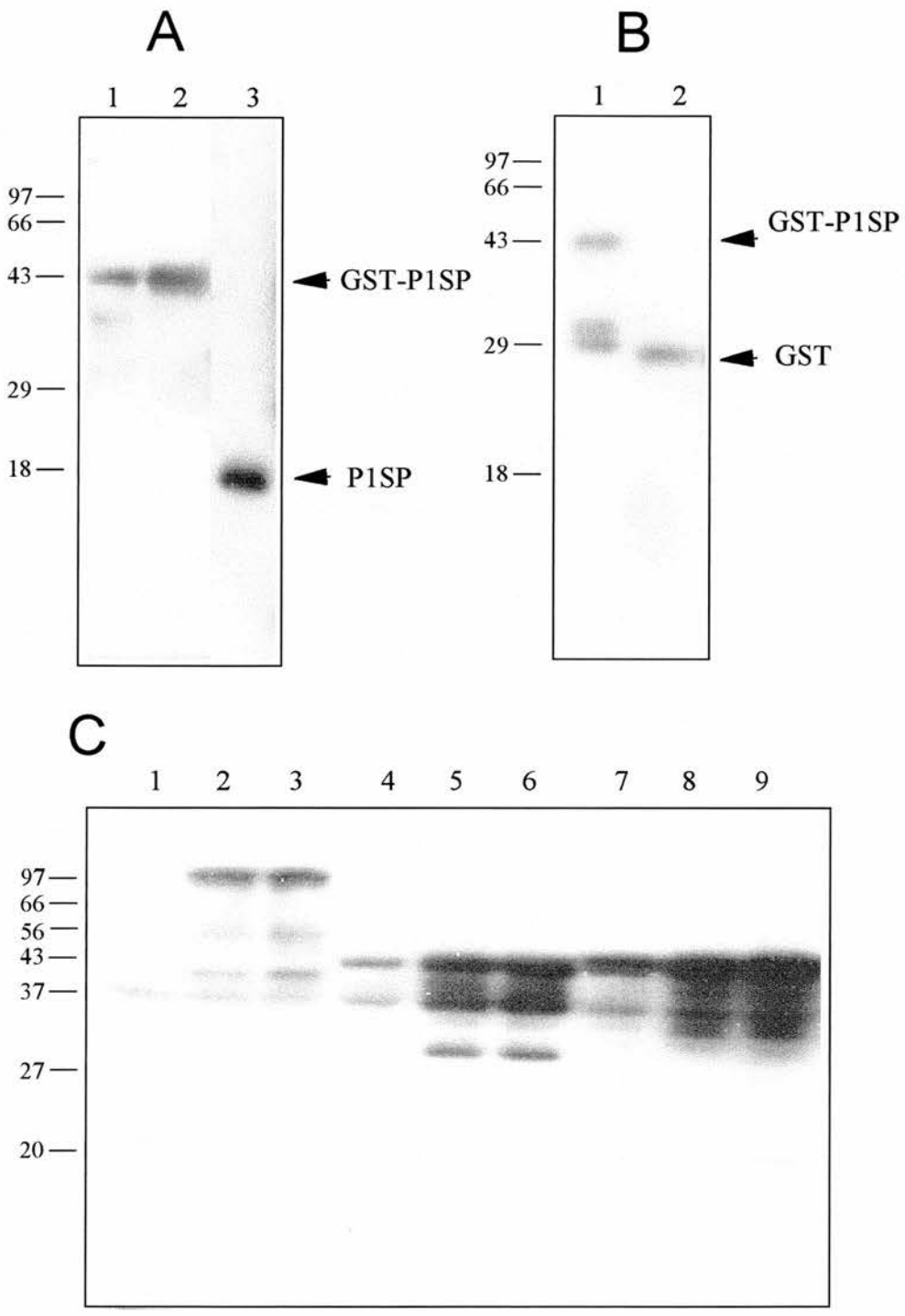


Figure 3-15 Detection of P1-derived proteins expressed in bacteria

A. Probed with anti-P1SP. Lane 1, lysate of cells induced to express GST-P1SP; lane 2, purified GST-P1SP; lane 3, thrombinised GST-P1SP.

B. Probed with antisera specific to GST. Lane 1, partially thrombinised GST-P1SP; lane 2, completely thrombinised GST-P1SP.

C. Probed with anti-P1SP. Lane 1-3, cells expressing GST-P1(0, 3 and 6 hrs after adding IPTG); lane 4-6, GST-P1B; lane 7-9, GST-P1C.

3.2.5 Detection P1-derived proteins from insect cells

The reaction of anti-P1SP to the cell extracts of SF9 infected with recombinant baculoviruses, BacP1GV and BacP1GVPk, were also tested by Western blotting. The construction of these baculoviruses is described in the Section 3.3.2. BacP1GV was constructed to express full-length P1 with a Gly to Val mutation at Ser-355, P1GV. BacP1GVPk encodes a protein that has a small (14 residues) tag (Pk), attached to the C-terminus of P1GV. As shown in Figure 3-16, anti-P1SP can detect a single band of about 70 kDa (71 kDa for P1GVPk). The identity of P1GVPk was further confirmed by Western blotting with monoclonal antibody against the Pk tag. The 43 kDa band detected by monoclonal anti-Pk appeared to be non-specific, because it also appeared in SF9 cells infected by BacP1 and other baculoviruses (Data not shown). It is clear that anti-P1SP can be used to detect P1-derived proteins expressed in the baculovirus system.

3.2.6 Detection of P1-derived proteins from plants

Leaf samples from healthy and PLRV-infected potato were also tested by Western blotting using anti-P1SP. Fresh leaves were ground in SDS sample buffer and extracts corresponding to 1mg leaf were loaded onto SDS-polyacrylamide gel. The results are shown in Figure 3-17, anti-P1SP could detect four bands in infected potato. The 55 kDa and 37 kDa bands also appeared in healthy potato, but the 27 kDa and 24 kDa proteins were absent. This 27 kDa band appeared to correspond to the protein detected in insect cells, and could be the 25 kDa protein detected by Prüfer (1999), which is believed to represent a product of processing P1 protein in plants, though the 24 kDa protein is a little closer in size.

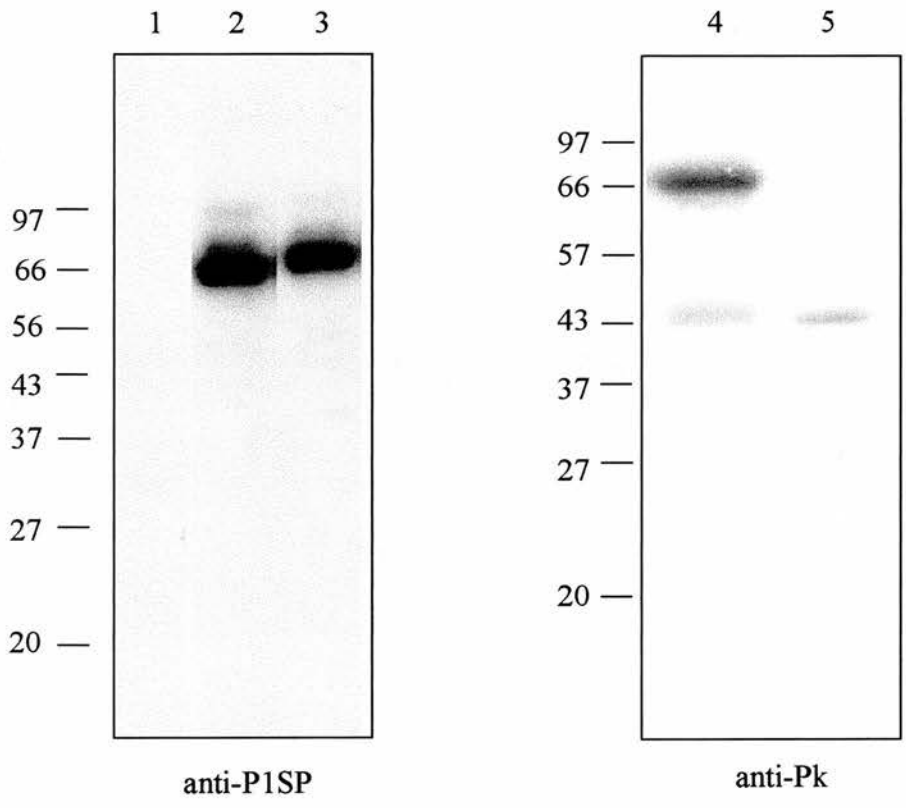


Figure 3-16 Detection of P1 proteins expressed in insect cells

Lane 1-3, probed with anti-P1SP; lane 4-5, probed with monoclonal antibody anti-Pk.

Lane 1, SF9 cells infected with BaclacZ (negative control); lane 2 and 5, cells infected with BacP1GV; lane 3 and 4, cells infected with BacP1GVPk.

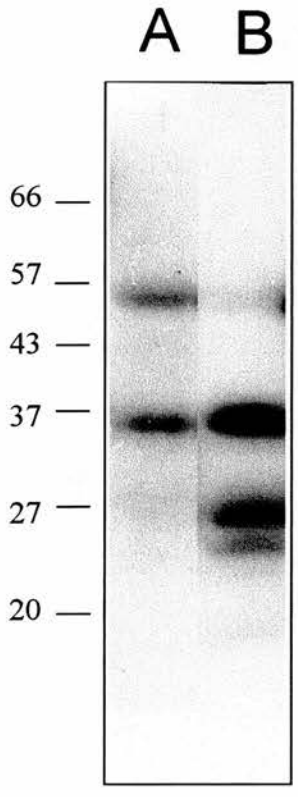


Figure 3-17 Detection of P1-derived proteins in plants

A. Leaves from healthy *M. piper*

B. Leaves from PLRV-infected *M. piper*

Leaves were ground in SDS-loading buffer. A sample equivalent to 2 mg leaves was loaded into each well. Anti-P1SP, 1:2000; HRP-anti-rabbit Ig, 1:4000.

3.2.7 Discussion

Polyclonal antisera were raised against the region of P1 encompassing the VPg domain, which was expressed in *E. coli*. This region of P1 was predicted to be highly antigenic, and my data support the view that it is highly immunogenic. With the pre-immune sera as controls in ELISA, the anti-P1SP sera still reacted specifically (signal/noise > 2) when diluted 0.5-1.0 x 10^6 fold.

The titres of the antisera were assayed in ELISA using extracts of *E. coli* JM101 expressing the GST protein as controls. The titres were quite high, $50x10^6$ and $25x10^6$ for antiserum A and B, respectively, suggesting the antisera were highly specific to the recombinant protein used to inoculate the rabbits. The low detection limit of P1SP in ELISA (3.2 ng), in the presence of saturating amounts of control proteins, is consistent with the high specificity of the antisera.

The antisera were also tested in Western blotting for their reaction with proteins expressed in *E. coli*, insect cells and PLRV-infected plants. As expected, the anti-P1SP reacted with proteins expressed in *E. coli*, which contained part or all of the regions used to raise the antibody. Several proteins with smaller molecular weights also reacted with anti-P1SP (weakly). If they were proteins encoded by *E. coli* or the vector sequences, they should appear in all lanes, which seems not to be the case. These proteins were more likely to be the results of partial degradation by *E. coli* proteases. The reaction of anti-P1 with P1 protein expressed in insect cells was very specific with very low background. Therefore, anti-P1SP is suitable to be used in baculovirus system.

The results of Western blotting using extract from infected plants showed that anti-P1 could react with 4 proteins. The two larger proteins appeared in both healthy and PLRV-infected plants. Since the antigen used to raise anti-P1SP was synthesized in *E. coli*, and not purified from plants, it was impossible to be contamination by plant proteins. These two proteins might contain structures that mimic the determinants identified by anti-P1SP. The smaller proteins, about 27 and 24 kDa, were present in infected plants only, and were believed to be P1-derived. Unfortunately, anti-P1SP was not suitable for detecting P1 protein in plants with ELISA, due to the reaction with the two larger proteins.

3.3 Studies on the Proteolytic Activity of P1 in Insect Cells

3.3.1 Introduction

When this project started, little was known about the function of the P1 protein of luteoviruses except that (1) molecular genetic analysis showed that P1 was essential for viral RNA replication in BWYV (Reutenauer *et al.*, 1993); (2) sequence alignments identified helicase (Habili and Symons, 1989) and protease motifs (Koonin and Dolja, 1993) in P1 coding sequences. The original aim of the study described in this section was to purify large amounts of P1 protein from insect cells to determine its biochemical features and test its possible functions. Later, it became clear that wild-type P1 protein was processed in insect cells, which meant no full-length P1 accumulated. This prompted us to use baculovirus as an *in vivo* system to answer some of the questions about the functions of the P1 protein, i.e. the processing of P1 and P1/2 proteins.

The major characteristics of the baculovirus expression system are briefly described here.

Baculoviruses

Baculoviruses have rod-shaped capsids 40-50 nm in diameter and 200-400 nm in length. The genomes of baculoviruses are double-stranded, circular and supercoiled DNA molecules about 80-200 kb long. The most often used baculoviruses are *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV) and *Bombyx mori* nuclear polyhedrosis virus (BmNPV). In cell culture of *Spodoptera frugiperda*, the infection cycle of AcMNPV comprises three phases:

1. **Early Phase**: Infection is mediated by the budded virus (BV), a rod-shaped nucleocapsid surrounded by lipid envelope containing peplomer-like structure (gp64) at one end. After adsorptive endocytosis, the nucleocapsid migrates into the nucleus, and releases the DNA core. Early viral gene expression occurs 0.5 to 6 hours post infection (pi). Host gene expression is shut off. Infected cells undergo cytoskeletal rearrangement, host chromatin dispersion and nucleus enlargement.

2. **Late Phase**: This phase extends from 6 h pi to about 20-24 h pi and includes late gene expression, viral DNA replication and BV production. The characteristic virogenic stroma appears in the nucleus. Production rate of BV peaks from 12 to 20 hours pi, decreasing thereafter.

3. **Very Late Phase**: This phase starts from about 20 h pi and features by production of occluded viruses (OVs), also called polyhedral inclusion bodies (PIBs). Polyhedrin and fibrous p10 protein are synthesized from polh and p10 promoters. The polh promoter becomes highly active at about 18 h pi. By 27-48 h pi, polh mRNA makes up 20% of the total polyadenylated RNA in the cell. Cell lysis begins about 60 h pi, and protein synthesis ceases by 72 h pi. The process of cell dying and lysing can last until 4 to 5 days pi.

Baculoviruses as Expression Vectors

Baculoviruses are transmitted horizontally in cell culture. Polyhedrin and p10 are required in vertical transmission only, so their gene can be replaced with foreign genes, which are expressed via the strong polh or p10 promoter. Advantages of baculovirus expression system include:

1. Eukaryotic system of protein synthesis: Eukaryotic proteins expressed in baculovirus system generally fold correctly, form the correct disulfide bonds, and are post-translationally modified. Insect cells can identify mammalian signals for glycosylation, acylation, phosphorylation, amidation, and signal peptide cleavage. Recombinant proteins are often soluble and easily purified from infected cells. The baculovirus system is particular useful to express unspliced genes (cDNA version, no introns), though it was reported to carry out some splicing (Iatrou *et al.*, 1989).

High level of expression: Expression of polyhedrin can reach about 1 gram per 10^9 cells (~1 litre culture), equivalent to 25-50% of the total cellular protein. Most recombinant proteins are expressed at levels from 10 mg to 100 mg per litre culture, higher than most other eukaryotic expression systems. However, it is difficult to predict the yield of a new gene in any expression system.

Tolerance of cytotoxic effect: Recombinant proteins are expressed after the production of BV, that means an advantage for expression of genes with a negative effect on essential cell functions, and less pressure for the virus to delete or inactivate the heterologous gene.

Simplicity of technology: Recombinant baculoviruses are relatively easy to construct, especially with the site-specific transposition method. No helper viruses are required. Viruses are usually propagated in monolayer and/or suspension of SF cell lines. The medium is readily available and the production of recombinant proteins is easy to scale up.

Safety: Baculoviruses have a narrow host range and each baculovirus only infect a few related insect species. Insect cell lines are not transformed by pathogenic or infectious viruses, so containment is minimal.

There are some limitations of baculovirus system. Some post-translational modifications are slightly different from that in natural system. Due to the decline in host cell functions, the efficiency of post-translational modification in very late phase might decrease. Expression in baculovirus is still a lengthier procedure than *E. coli* systems.

Generation of Recombinant Baculoviruses

There are two methods to generate recombinant baculoviruses, both of which have been used in this study.

Homologous Recombination (e.g. pVL1392/1393): The gene of interest is inserted downstream of a polh promoter (or p10 promoter) flanked by baculovirus non-essential sequences (i.e. encompassing polyhedrin or p10 gene) in a transfer plasmid. The recombinant transfer plasmid is co-transfected into insect cells with purified parental AcMNPV DNA. The gene of interest is inserted into the viral genome by allelic replacement, which involves two distinct homologous recombination events at each side of the foreign gene (double crossover). To distinguish double crossover recombinant from single crossover recombinant, in which the whole transfer plasmid is inserted into the viral genome, viral DNA usually contains a marker gene (e.g. β-galactosidase) between the two crossover regions. In double crossover recombinants, the marker gene is replaced by the gene of interest. Linearization of virus DNA at the target site (e.g. cleavage at a unique *Bsu*36 I site in the β-galactosidase gene) will greatly increase the

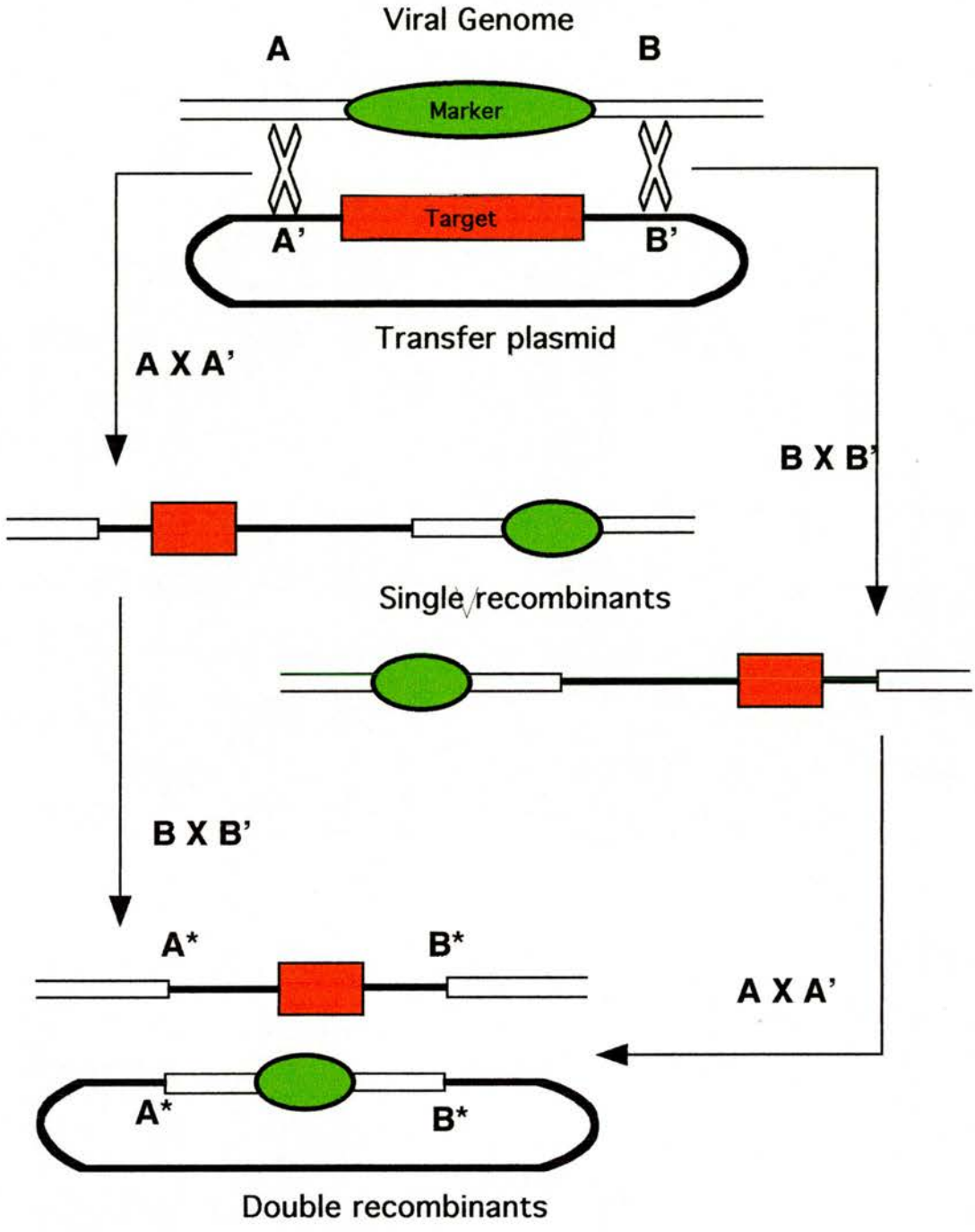


Figure 3-18a Generation of recombinant virus by homologous recombination

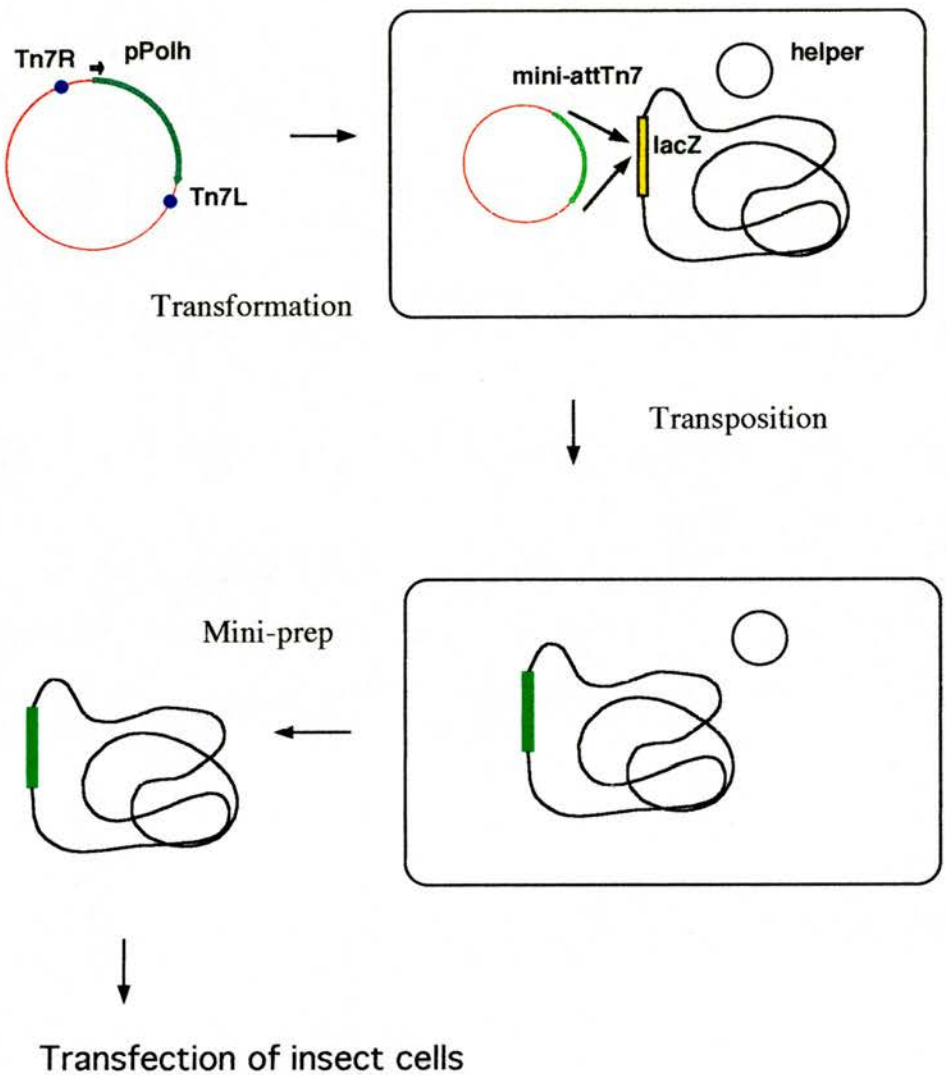


Figure 3-18b Generation of recombinant baculovirus by site-specific transposition (Bac-to-Bac)

proportion of double crossover recombinants, because only double crossover at both side of the cleavage site will produce a circular recombinant viral genome, which is the substrate of viral replication apparatus (Figure 3-18a).

Site-specific Transposition (e.g. Bac-to-Bac): Recombinant baculoviruses can be generated much more quickly by using Tn7 sponsored site-specific transposition. The foreign gene is cloned into a mini-Tn7 element within a pFastBac donor plasmid. The recombinant plasmid is transformed into *E. coli* cells DH10Bac, which contain the bacmid with a mini-att Tn7 site and the helper plasmid. The mini-Tn7 element on the donor plasmid can transpose to the mini-att Tn7 site on the bacmid and disrupt the *lacZ* gene in it, resulting in a white colony. The recombinant bacmid is purified from *E. coli* and transfected into insect cells (Figure 3-18b).

3.3.2 Processing of P1 and P1/2 proteins in SF9 cells

3.3.2.1 Correction of a point mutation

The plasmid pSAB53 contains the whole genome of PLRV inserted into the vector pUC19 with the insertion of a T7 promoter immediately upstream of the 5' terminus of the PLRV sequences. This plasmid was used as the source of P1 sequence, but was found to contain a point mutation (GGA to GTA) within an important motif of the putative P1 serine protease domain, which caused a coding change (Gly-355 to Val). Therefore, it was decided to correct this mutation by site-directed mutagenesis.

Since the mutation abolished an *Acc* III site (5'TCCGGA3') and its correction re-creates the site, the mutagenesis experiment was carried out in a simplified way. Sequences encoding the N-terminal half of the P1 protein were amplified using oligonucleotide

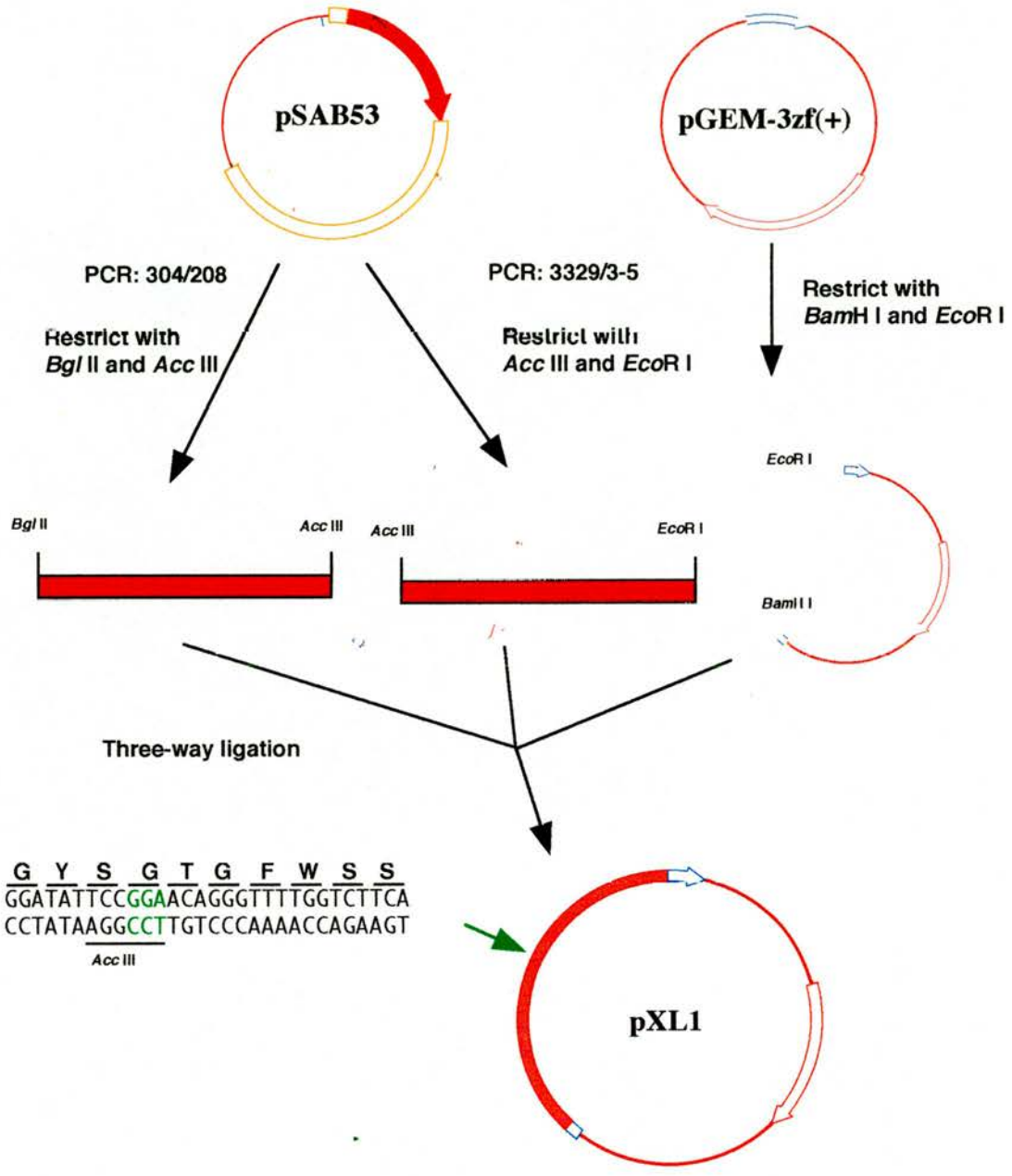


Figure 3-19 Construction of plasmid pXL1: correct a point mutation
Two fragments containing wild-type P1 sequences were amplified with PCR, restricted with enzymes, and ligated into pGEM-3zf(+).

primers 304 (5′-GACTAGATCTCCACCATGAACAGATTTACCGCATATG-3') and 208 (5′-TGTT**TCCGGA**ATATCC-3'), the reverse primer 208 correcting the point mutation. The sequences encoding the C-terminal half of P1 were amplified using primers 3329 (5′-GACCCG GATAT**TCCGGA**ACAGGGTTTTG-3') and 305 (5′-ACTGGAATTCTCA GGCTTTGGAGTTCAGCTTC-3'), the forward primer 3329 (overlapping primer 208) also encoding the correction. In the former case the PCR product was restricted with *Bgl* II and *Acc* III, whilst in the latter the PCR product was restricted with *Acc* III and *Eco*R I. The purified two restriction fragments were ligated together with the vector pGEM3zf(+) (Promega) restricted with *Bam*H I and *Eco*R I (Figure 3-19). This plasmid was designated pXL1.

Sequences encoding the P1 region were amplified by PCR using the template plasmid pXL1 and primers 304 and 305. The PCR product was restricted with *Bgl* II and *Eco*R I, and ligated into the vector pFastBac1 (GibcoBRL), restricted with *Bam*H I and *Eco*R I, to generate the recombinant donor plasmid pBacP1.

On the other hand, the mutant version (containing Gly355 to Val mutation) of P1 was cloned into pFastBac1 similarly by using pSAB53 as template of PCR, instead of pXL1. The plasmid is designated pBacP1GV.

The donor plasmids were used to generate corresponding recombinant baculoviruses using the method described in Materials and Methods.

3.3.2.2 Introducing a C-terminal Pk tag

Pk tag is a 14 amino acid sequence oligopeptide in the common region of the P and V proteins of simian virus 5 (SV5), which has shown to be a valuable universal affinity

tag. pVL140Pk, the plasmid containing the Pk tag coding sequence, and monoclonal antibody specific to the Pk tag, were kindly provided by Dr. Randall (Hanke *et al.*, 1995). The sequences encoding the Pk tag (in blue colour) in pVL140Pk is:

```
 Sac I   G   K   P   I   P   N   P   L   L   G
gag ctc GGA AAG CCG ATC CCA AAC CCT TTG CTG GGA

 L   D   S   T   *   Pst I
TTG GAC TCC ACC tga ctgcagtgat
```

Sequences encoding the entire P1 region of pSAB53 were amplified using oligonucleotides 304 and 155 (5'-ACTGGAGCTCGGCTTTGGAGTTCAGCTT CAGA-3'). The PCR product was restricted with *Hind* III, then doubly restricted with either *Bgl* II, or, *Sac* I. The *Bgl* II–*Hind* III and *Hind* III-*Sac* I restriction fragments were ligated together with plasmid pVL140Pk restricted with *Bam*H I and *Sac* I (Figure 3-20a). This three-way ligation strategy was adopted due to the presence of an internal *Sac* I restriction enzyme site. This plasmid, designated pVLP1GVPk, encodes the P1 region ($G^{355}\rightarrow V$) of PLRV fused to the Pk tag (supplied from plasmid pVL140Pk).

Plasmid pVLP1GVPk was initially restricted with *Kpn* I (present in the vector sequences downstream of the insert) and subsequently *Sma* I. The small restriction fragment (encoding sequences downstream of the proteinase domain together with the Pk tag) was purified and ligated with *Kpn* I/*Sma* I restricted pBacP1 and pBacP1GV, to generate donor plasmids pBacP1Pk and pBacP1GVPk, respectively (Figure 3-20b).

3.3.2.3 Fusing the P2 reading frame to that of P1

Naturally ORF P2 is expressed as a P1/2 transframe protein via a -1 frameshift at the slippery heptanucleotide UUUAAAU at a frequency of about 1%. In order to translate

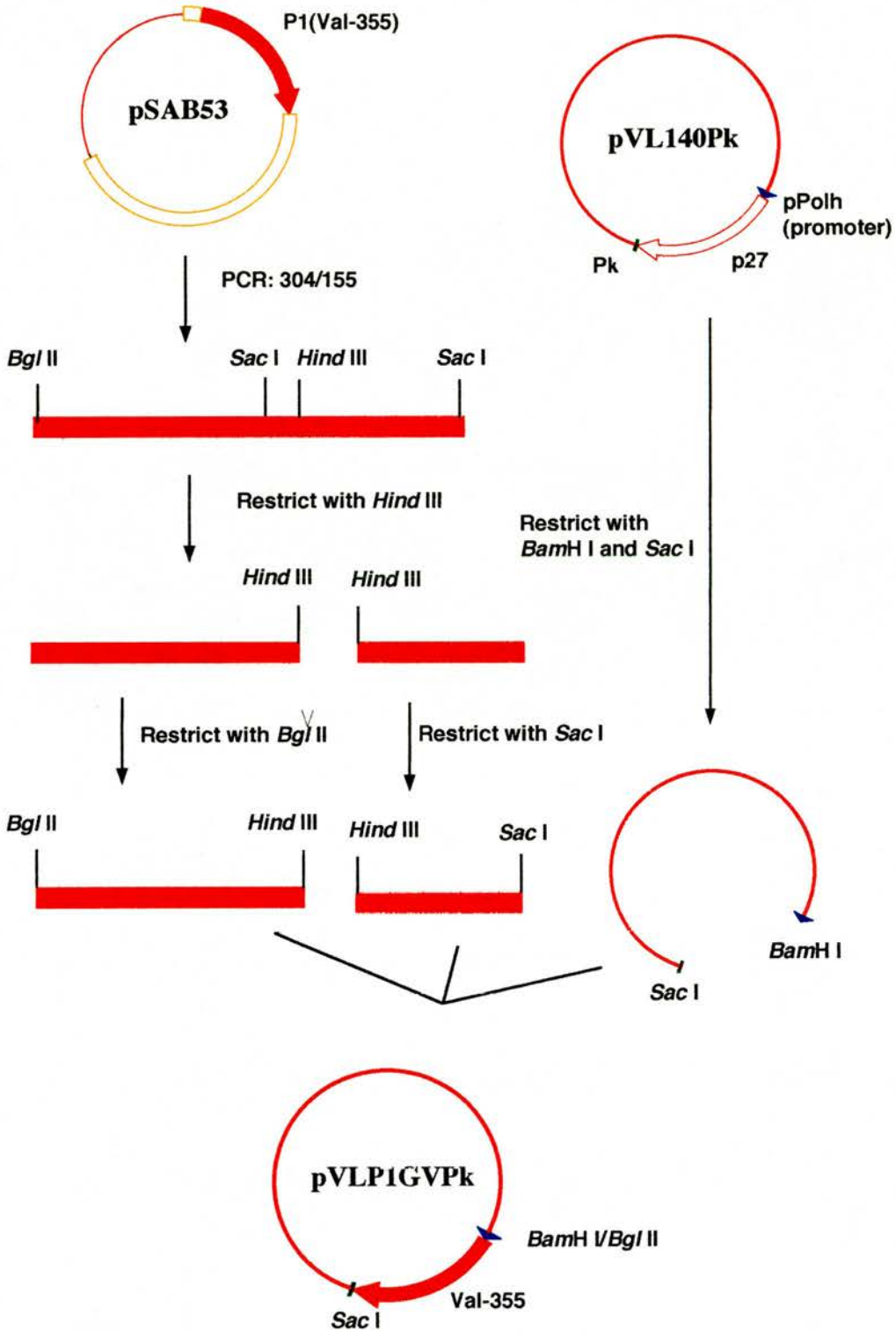


Figure 3-20a Construction of plasmid pVLP1GVPk

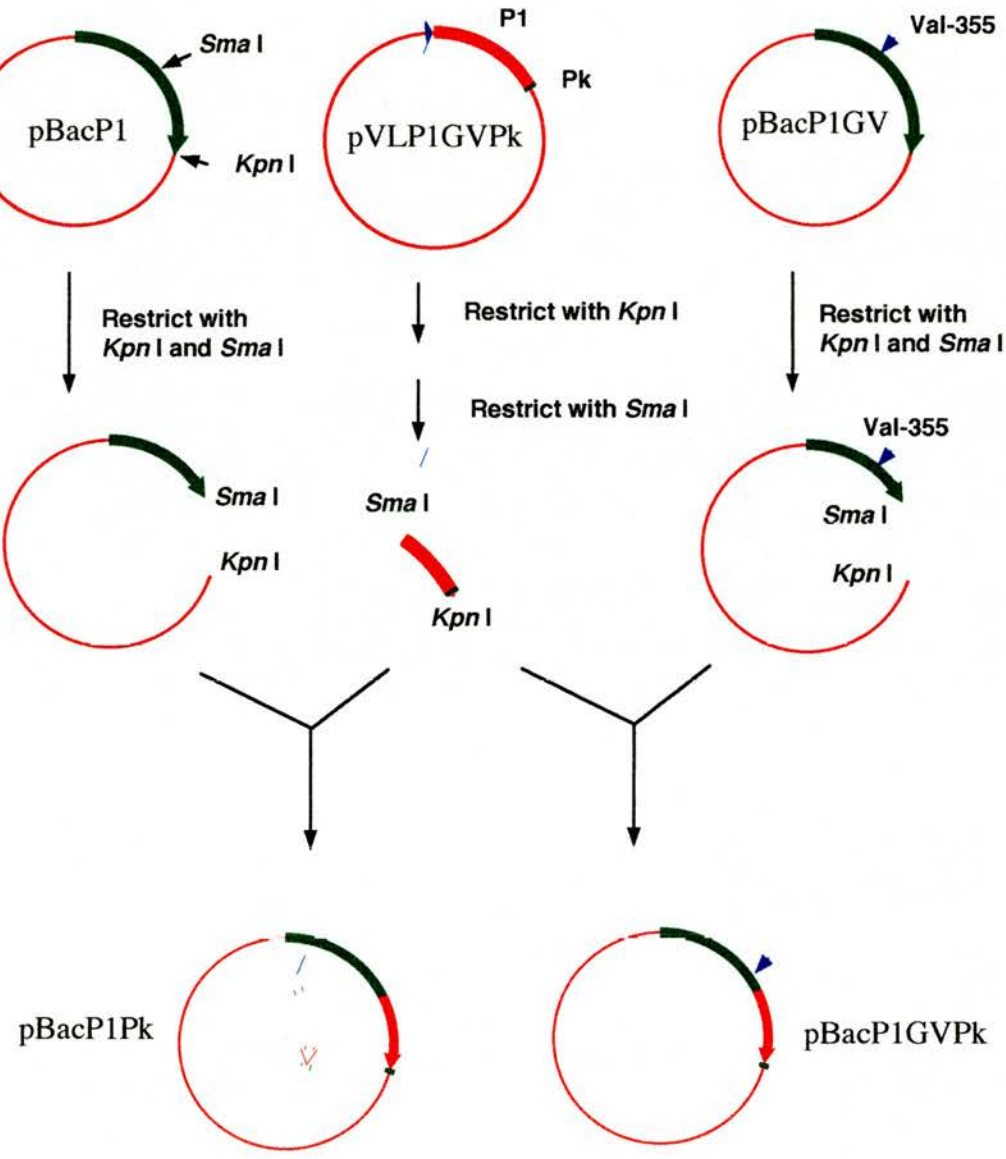


Figure 3-20b Construction of plasmids pBacP1Pk and pBacP1GVPk

the P1/2 protein without frameshifting, an extra nucleotide A was inserted into the shifty heptanucleotide so that downstream ORF P2 is in frame with ORF P1 (Figure 3-21).

The sequence upstream of the frameshifting site was amplified using oligonucleotides 304 and 5637, treated with Klenow fragment, purified from agarose gel and digested with *Bgl* II. The sequence downstream of the site was released from pSAB53 with restriction enzymes *Dra* I and *Mfe* I and purified from an agarose gel. These two fragments were ligated into the *Bam*H I and *Eco*R I restricted vector plasmid pFastBac1. The plasmid thus obtained was designated pBacP2.

3.3.2.4 Construction of control baculoviruses

Plasmid pSAB53 was restricted with *Kpn* I, treated with mung bean nuclease to produce blunt ends, then restricted with *Bgl* II. The 5.8kb restriction fragment, which contained the whole PLRV genome, was ligated into the vector pVL1393 (Invitrogen) restricted with *Sma* I and *Bgl* II. This plasmid, designated pVLTV, was constructed by Dr. Joe Lamb in this lab.

Plasmid pVLTV was restricted with *Xba* I and then re-circularised to give rise to plasmid pVLTVΔX, in which the majority of P1 and P2 sequences (protease, VPg and polymerase domains) were removed (Figure 3-22). This plasmid was used to generate the baculovirus BacTVΔX via *in vivo* homologous recombination (see Materials and Methods). Baculovirus BacP1GVPK was also constructed using this *in vivo* homologous recombination method with plasmid pVLP1GVPk, but no difference has been detected between the baculoviruses generated by the two methods (Data not shown).

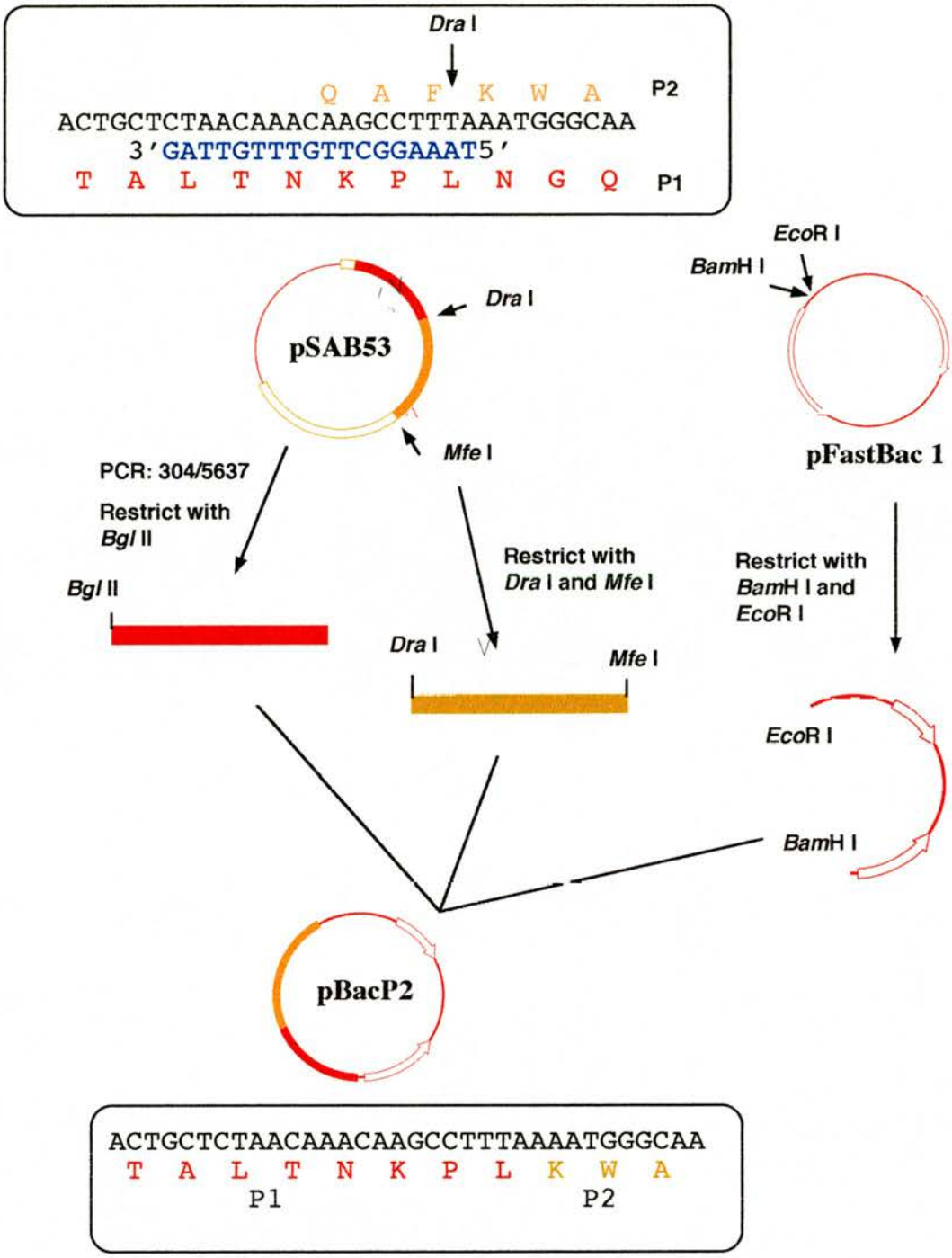


Figure 3-21 Construction of pBacP2
By using the oligo 5637 (in blue colour), an extra nucleotide was inserted into the frameshift site so that the downstream P2 sequence was in frame with the P1 sequence.

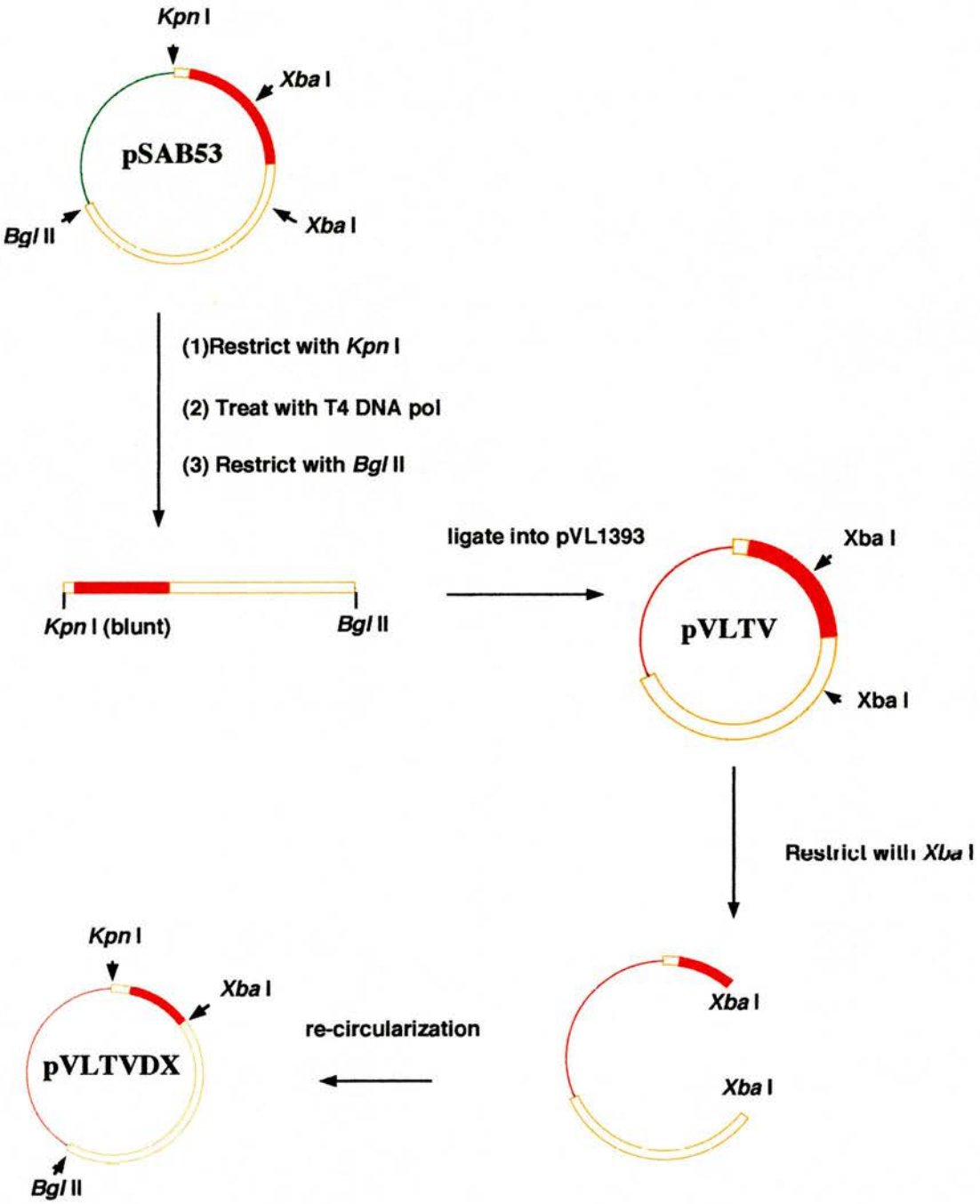


Figure 3-22 Construction of control plasmid pVLTVDX

3.3.3 Processing of P1 and P1/2 in insect cells

Three recombinant baculoviruses were constructed to express P1 containing sequences directed by the polyhedrin promoter. Baculovirus BacP1 expressed the native P1 protein, whilst the P1 protein encoded by BacP1Pk had a C-terminal 16 aa extension corresponding to the highly antigenic Pk tag from Simian virus 5 (Hanke *et al.*, 1995). The P1 and P2 coding sequences in BacP2 were engineered so that they were in the same reading frame, so BacP2 expressed the P1/P2 fusion protein, which can only be expressed via a -1 frameshift in infected plants.

Figure 3-23 shows the time course of the expression of these proteins in insect cells. SF9 cells were infected at MOI of 2 and harvested at the time points indicated. The cells were lysed in Laemmli buffer, resolved by SDS-PAGE, and analyzed by Western blotting. The predicted molecular masses of P1, P1Pk and P1/2 proteins are 70, 71 and 119 kDa respectively. Western blot analyses using anti-P1 antisera showed, however, that P1, P1Pk and P1/2 proteins were all processed to smaller products (Figure 3-23). The VPg-containing cleavage products of these proteins had similar size, about 27 kDa. To produce a VPg-containing protein of such a size from the P1/2 precursor, there should be at least two processing sites, one at each side of the VPg domain. In addition, the C-terminal Pk tag extension appeared to have no discernible effect on processing.

The expression of P1, P1Pk and P1/2 in SF9 cells infected at low MOI was also examined by Western blotting. As shown in Figure 3-24, full-length P1, P1Pk and P1/2 precursor proteins could be detected in SF9 cells three days after infection at MOI of 0.1, in addition to the products of processing. For P1/2 protein, it seemed an

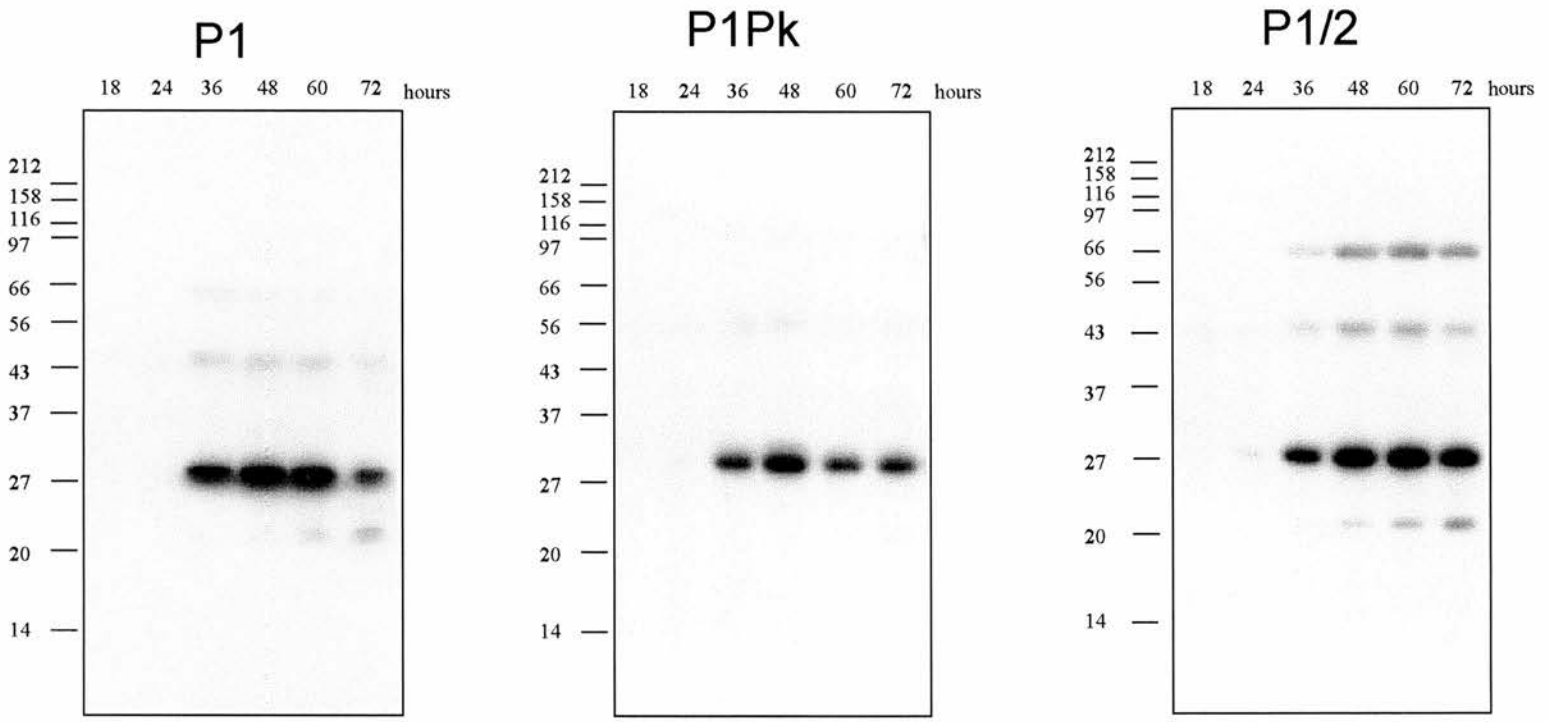


Figure 3-23 Time course of expression of P1, P1Pk and P1/2 proteins in SF9 cells

SF9 cells were infected with baculoviruses BacP1, BacP1Pk and BacP2 at MOI of 2, and harvested at 18, 24, 36, 48, 60 and 72 hours pi. Cell lystes were resolved by SDS-PAGE. Expression of recombinant proteins was probed with anti-P1SP after Western blotting. In all cases, the main protein reacting with the antisera was a product of protein processing.

intermediate of about 73 kDa was produced, which appeared as a strong band when cells were infected at low MOI but was a faint band at high MOI (Figure 2-23 and 2-24).

When probed with monoclonal antibody against the Pk tag, the cleavage product containing Pk tag had a molecular weight of about 27-28 kDa (Figure 3-24 panel A, lane 2). Since the Pk tag is at the C-terminus of P1, the cleavage site should reside at the N-terminal region of the VPg domain, which means this C-terminal cleavage product should be detected by both polyclonal anti-P1SP and monoclonal anti-Pk.

3.3.4 The effects of Gly355 to Val mutation

During the construction of the recombinant baculoviruses BacP1 and BacP1Pk, it was found that the plasmid pSAB53 contains a single point mutation resulting in the coding change $G^{355}\rightarrow V$. This mutation is at a conserved proteinase motif (GxS**G**) with G^{355} corresponding to the glycine residue immediately downstream of the putative catalytic serine of the protease domain. This Gly is conserved not only among poleroviruses, but also in almost all other chymotrypsin-like serine proteases. It is thought to be involved in the formation of the oxyanion 'hole' of the putative proteolytic domain. To investigate the effect of the $G^{355}\rightarrow V$ mutation two further recombinant baculoviruses were constructed (BacP1GV and BacP1GVPk), each bearing this point mutation. Western blot analyses of SF9 cells infected with BacP1GV and BacP1GVPk viruses showed only full-length proteins were produced (Figure 3-25). These full-length proteins accumulated in the cell from 1 day pi in parallel to the progress of viral infection. They began to suffer slight non-specific degradation, which could be seen by the smear of the bands, from 3 days pi, when infected cells already started to lyse. This non-specific degradation was clearly different from the processing we observed before, which happened much earlier.

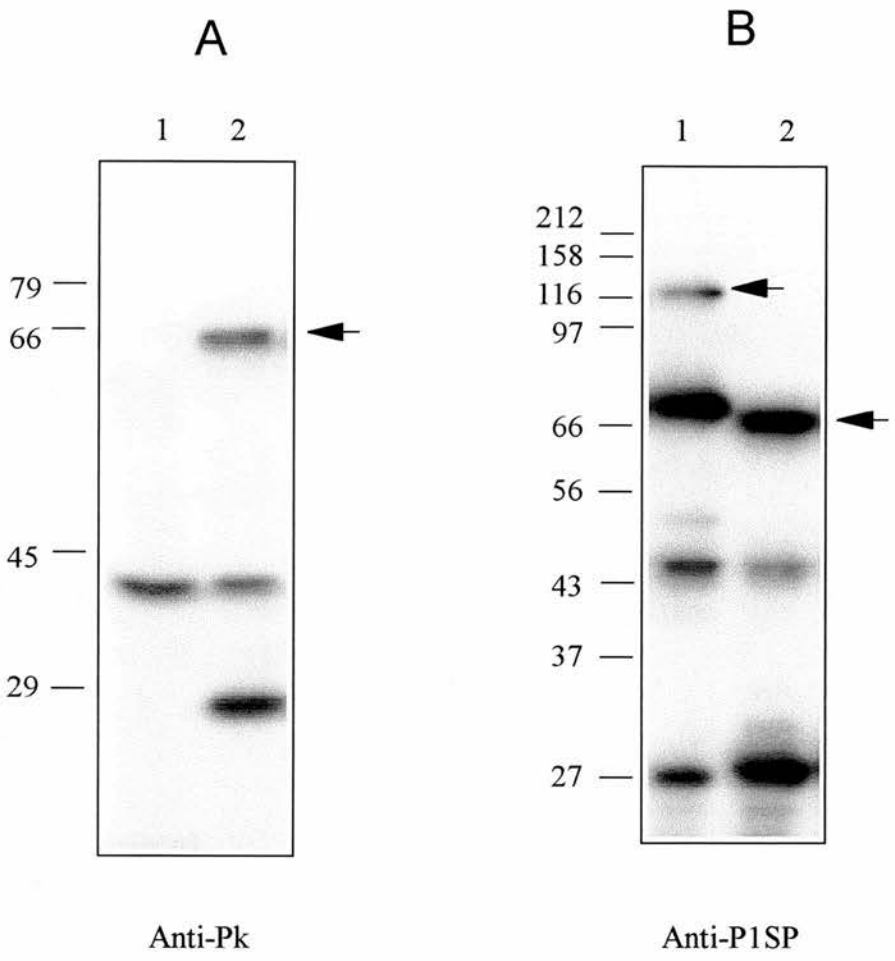


Figure 3-24 Expression of P1-derived proteins at low MOI

SF9 cells were infected by baculoviruses at low MOI (about 0.1) and harvested 3 days pi. Cells lysate was analyzed by immunoblotting. Full-length precursors (indicated by arrows) and cleavage products (including intermediates) were observed.

A. probed with monoclonal anti-Pk. Lane 1, BacTVΔX (control); lane 2, BacP1Pk.

B. probed with polyclonal anti-P1SP. Lane1, BacP2; lane 2, BacP1.

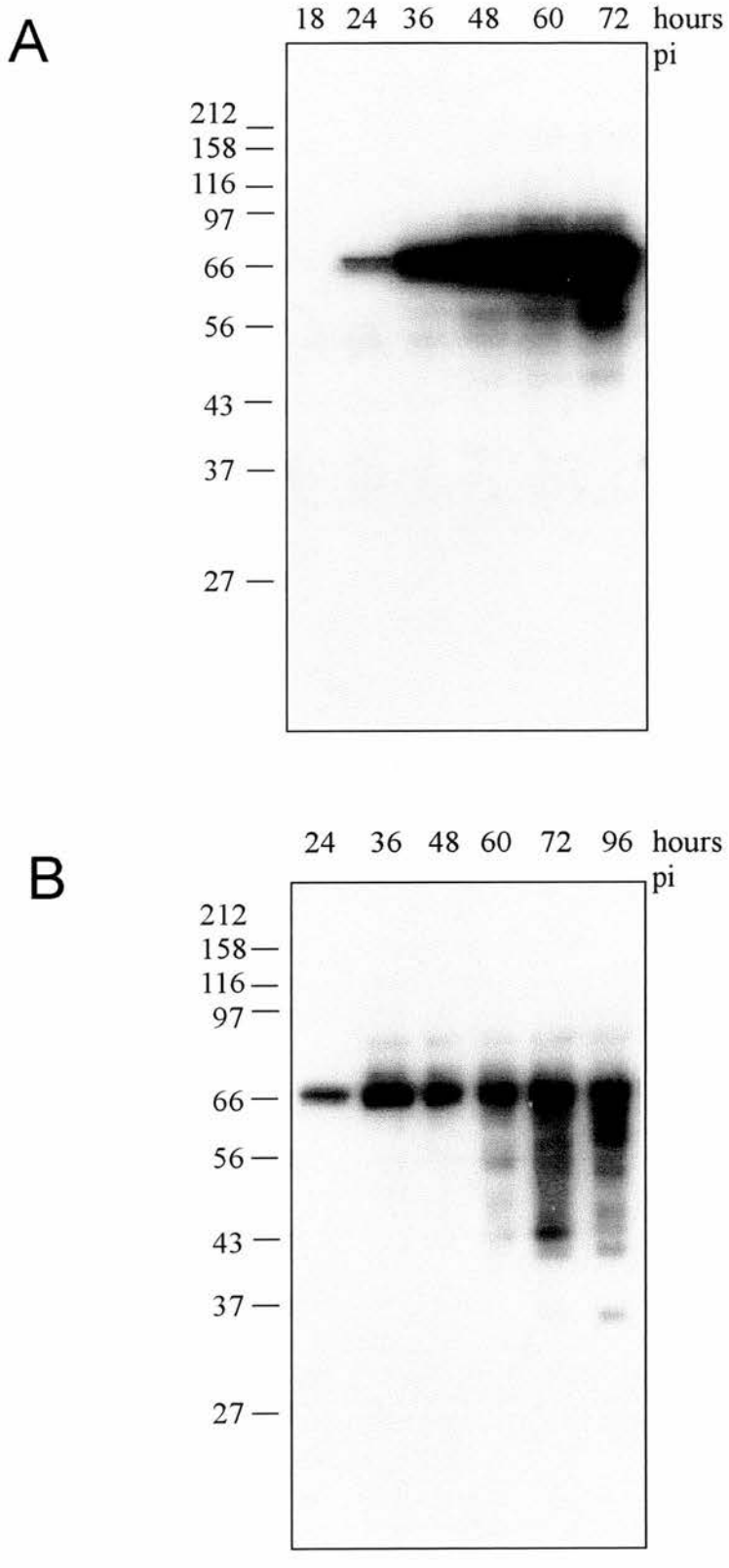


Figure 3-25 Expression of P1 and P1Pk proteins containing Val-355
SF9 cells were infected with baculoviruses BacP1GV (panel A) and BacP1GVPk (panel B), respectively, and harvested at different times. Cell lysate was probed with anti-P1SP.

These observations are consistent with residue G^{355} being an important component of the proposed serine protease domain (Figure 1-2) and directly contradict the notion that the processing we observed with BacP1 and BacP1Pk was due to non-specific degradation.

3.3.5 *Trans* activity of P1 protease in SF9 cells

The processing observed in baculovirus-infected SF9 cells could be accounted for in two ways: (1) the processing was mediated by the protease domain in the P1 protein, which was inactive in the $G^{355}\rightarrow V$ version; (2) or an undefined cellular or baculovirus-specific proteinase was responsible for the cleavages. To address this question, a co-infection experiment was designed to test if the cleavage could be carried out by the P1 protein *in trans, in vivo*.

As shown above, when SF9 cells were infected by BacP1GVPk, a Pk-tagged full-length P1protein ($G^{355}\rightarrow V$ version) was produced, not being cleaved in this system. Baculovirus BacP1, however, produced untagged P1 protein, which was cleaved. Figure 3-26 shows the results when SF9 cells were co-infected with both BacP1 and BacP1GVPk. Western blotting with a monoclonal antibody directed against the Pk tag showed that protein P1GVPk was cleaved upon co-infection with BacP1. Taken together, these data show protein P1GVPk, functioning as a substrate (alone), was processed *in trans* by the wild-type P1 protein (expressed from BacP1) and not by a cellular nor baculovirus-specific proteinase. By increasing the multiplicity of infection (MOI) ratio BacP1vs BacP1GVPk, increased P1 processing *in trans* is observed. Similar observations are made with cells singly infected with BacP1Pk: at MOI = 0.1 processing

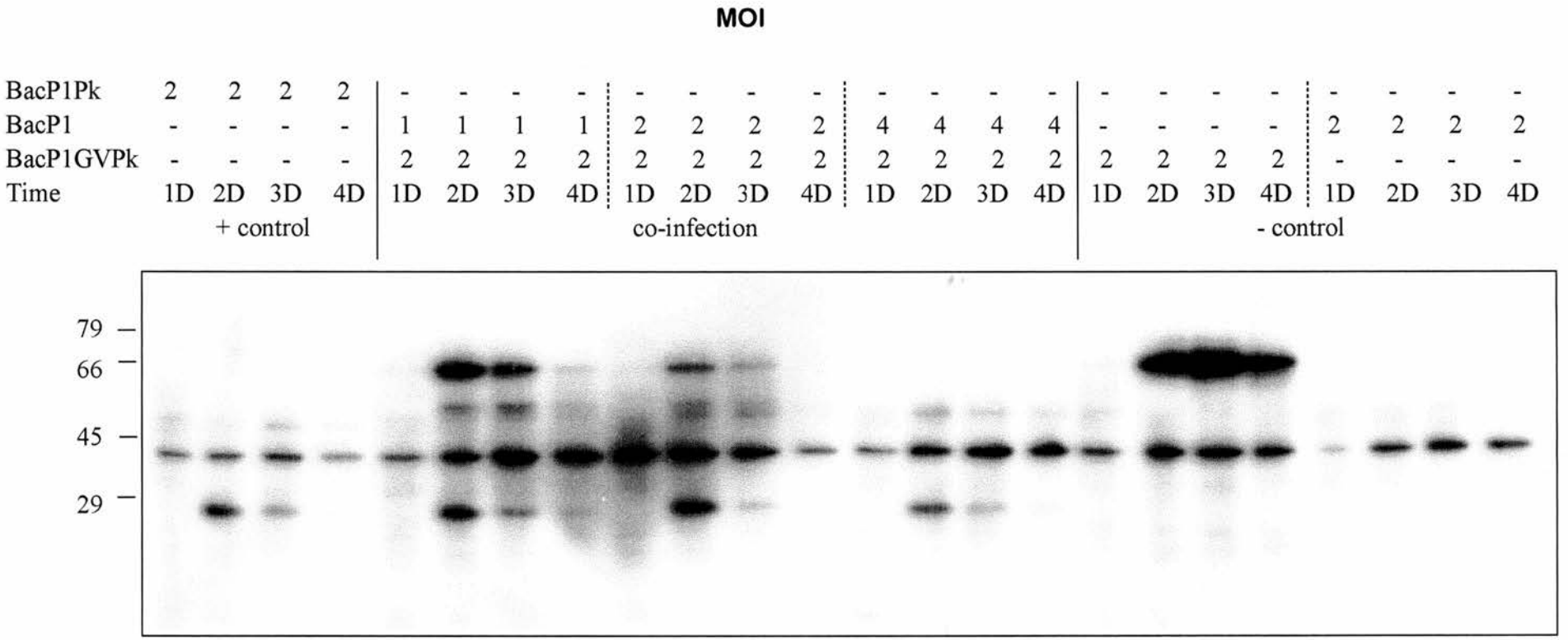


Figure 3-26 Protein processing *in trans*

SF9 cells were infected with recombinant baculoviruses at the indicated MOIs and harvested at 1, 2, 3, and 4 days pi. Cell lysate was probed with anti-Pk monoclonal antibodies. The 43 kDa band was non-specific.

is partial with substantial amounts of uncleaved P1 observed (Figure 3-24), whereas at MOI = 2 processing of P1 appeared to be complete.

3.3.6 Discussion

3.3.6.1 Development of an *in vivo* system

Non-structural proteins of RNA viruses are usually difficult to study. They are often involved in viral RNA replication and interfere with transcription and translation of host cells; they might damage the membrane system of host cells; or they could be toxic to cells by unknown mechanisms. All these factors could result in poor or no expression in prokaryotic or eukaryotic expression systems usually used by researchers, which in turn hampers *in vitro* biochemical studies of these proteins. In the case of PLRV, additional difficulties are present due to the inability to infect the plant mechanically, the containment of the virus within phloem tissue, and the low concentration of the virus in plants.

The problems have been alleviated by the use of a range of new techniques over the few past years. Remarkable progress has been made by application of *in vitro* transcription and translation systems, the protoplast technique, and agroinfection. Here, we describe the development of an *in vivo* system to study the PLRV non-structural proteins, specifically the processing of the P1 and P1/2 proteins. This system is actually a combination of several ideas/techniques already used in the studies of many human and animal viruses. My work shows that this system has some features suitable for the study of PLRV protein processing.

Baculovirus system

Fundamental features of baculovirus have been introduced at the start of the section. Several points critical to the work described in this thesis are emphasized here.

Compared with transmission by aphids, transfection of protoplast, or even the infection of plants by agrobacteria, the infection of cultured cells by baculovirus is a very efficient way to deliver foreign genes into the expression system. The MOI can be adjusted easily and precisely. Within the range of MOI (0.1 to 10.0), you can control how many cells are infected, and calculate the percentage of cells infected by 0, 1, 2, 3... viruses. Furthermore, it is easier to achieve synchronized infection and expression in the baculovirus system than in aphid transmission or agroinfection. The life cycle of baculoviruses lasts about 4 - 5 days, and the uninfected insect cells double every 18-24 hours. Experiments can, therefore, be carried out relatively quickly.

The activity of the polyhedrin promoter is well characterized. Being a very late promoter, it becomes highly active from about 18 hours post infection, when host cells have stopped dividing and are more likely to tolerate toxic proteins. By 27 h pi to 48 h pi, mRNA transcribed from polyhedrin promoter makes up about 20% of total polyadenylated RNA. By 72 h pi, most cells cease to synthesize proteins. Therefore, the expression from the polh promoter lasts about 2 days - a relatively long period if compared with *in vitro* expression system. As discussed later, this could allow enough time for the *trans* cleavage event to occur.

Another advantage of the baculovirus system is the possibility to carry out co-infection experiments. This means two recombinant baculoviruses expressing different proteins can be introduced into one cell at the same time, which allows the study of protein-

protein interaction *in vivo*. In the case of protein processing, it is possible to express the substrate protein and the protease at the same time, and study the processing *in trans*. Not all systems allow this kind of experiments. For example, different plasmids are stable in one *E. coli* cell only when they belong to different incompatible groups. Normally *E. coli* cannot be transformed with two plasmids at the same time, because most of the vectors used in molecular biology have the ColE1 *ori* and belong to the same incompatibility group. Furthermore, it is also possible to adjust the ratio of the two baculoviruses in the co-infection experiment, and control the relative levels of their expression.

Baculovirus system has its own limitations. Since PLRV only replicates in plants, and *Autographa californica* is not a vector of PLRV, PLRV proteins expressed in this system are present in an entirely artificial environment; it is very important, therefore, to demonstrate that the cleavage is mediated by a protease encoded by PLRV, not by proteases of baculoviral or insect cell origin, if a processing of PLRV protein is observed in insect cells. Co-factors in potato plants might not be present in insect cells.

The Pk tag

The study of protein processing was greatly facilitated by the Pk tag. Due to its highly specific antigenicity, it provided a reliable way to trace the products of protein processing. It can be recognized by monoclonal anti-Pk under both denaturing and native conditions when linked to proteins of various sizes (Hanke *et al.*, 1992).

The short Pk tag (14aa) has been shown on a number of occasions not to affect the function of the target protein it is attached to. When Pk was linked to the C-terminus of simian immunodeficiency virus (SIV) envelope protein gp160, gp160Pk was shown to

have the same expression level, stability, CD4 binding curve and susceptibility to furin cleavage as the parental gp160 (Hanke *et al.*, 1995). Attachment of the Pk did not prevent secretion of the Fv antibody fragment into the periplasmic space of *E. coli*, or the reaction of the Fv with its natural antigen in ELISA. In my studies, the attachment of Pk to the C-terminus of P1 did not affect the cleavage at the site between protease and VPg domains, and the same could be true for other possible cleavage sites in P1 and P1/2 proteins.

The Gly355 to Val mutation

The idea of introducing a mutation into the protease motif of a self-processing polyprotein to inactivate the protease domain and produce a precursor substrate for the assay of the protease is not new. Nicklin *et al.* (1987) constructed a plasmid containing the P1-P2 region of poliovirus with a mutation in the $2A^{pro}$ domain. This plasmid produced a stable P1-P2 precursor that could be processed specifically by exogenous $2A^{pro}$ provided by extracts of poliovirus-infected cells, by $2A^{pro}$ translated *in vitro*, and by $2A^{pro}$ expressed in *E. coli* (Krausslich *et al.*, 1987; Krausslich and Wimmer, 1988). Using this approach, Nicklin confirmed that the 2A functioned as a protease in the cleavage of the junction between P1 and P2. In a similar way, we demonstrated that the serine protease domain was able to cleave at the site N-terminal to the VPg, though this time an *in vivo* system was used instead of *in vitro* one.

For this approach to work well, it should satisfy two conditions: (1) the mutation introduced into the protease domain should abolish the protease activity; (2) the mutation does not affect the accessibility of the cleavage site, so that it can be cleaved by an exogenous protease source. Although the mutation of Gly355 to Val was found by accident, it fortunately meets the above requirements. Gly-355 is located in the active

centre and probably participates in the formation of the oxyanion hole that stabilizes the transition intermediate during the catalysis. Substitution by a bulky Val seems to block the active centre effectively. Gly355 is 64 amino acids upstream of the Glu-Ser site between protease and VPg domains, and the mutation is unlikely to change the conformation of the P1 polypeptide chain or disturb the cleavage site in other subtle ways.

3.3.6.2 Processing of P1 in insect cells

The first recombinant baculovirus containing P1 sequences was BacP1GV, which has the Gly355 to Val mutation. This virus synthesized full-length mutant P1 protein in SF9 cells, as expected. After correcting the mutation and infecting the insect cells with baculovirus encoding wild type P1, however, the major protein that reacted with anti-P1SP was about 27 kDa. As no processing of the P1 protein had been reported then, and baculovirus system was an entirely artificial system to PLRV, the first question to be addressed was whether the cleavage represented the processing of the P1 protein or just a degradation by a protease from the baculovirus or the host insect cells.

Several lines of evidence favour the former assumption. Firstly, P1 and P1GV have only one amino acid different. Although it is possible that this difference makes P1 susceptible to degradation by proteases of baculovirus or insect cells, while P1GV resistant, the probability is extremely low. Secondly, the non-specific degradation of P1GV (and P1) was observed (as a smear band in the Western blotting experiment), which only occurred after 4 day pi, when insect cells had already begun to lyse. At the same MOI, the site-specific cleavage of P1 happened about 2 days earlier, suggesting that the cleavage of P1 was different from the degradation usually expected when the cells had begun to lyse. Thirdly, when P1 was expressed with the Pk tag at its C-

terminus, it was cleaved in a similar way. The size of the Pk tagged product allowed the determination of the cleavage site to be the N-terminal region of VPg domain, showing the possible significance of the cleavage in releasing the VPg from P1 protein.

It was the co-infection experiment that conclusively demonstrated that the P1 protein itself was responsible for the cleavage observed in insect cells. BacP1GV produced full-length P1GV protein, which was a suitable substrate, and wild-type P1 protease could be provided by BacP1 (if the hypothesis that the P1 protein contained a proteolytic domain was correct). Since the processing of P1 itself also produced a product that reacted with anti-P1SP antibodies, a method to differentially detect the product from cleavage of mutant precursor was required. The Pk tag and the monoclonal antibody against it provided the solution. The substrate actually used in the co-infection experiment was P1GVPk, provided by virus BacP1GVPk.

Prüfer *et al.* (1999) demonstrated that the P1 protein was processed in plant cells, and predicted that the P1 protein was self-cleaved. One product of the processing in plants was about 25 kDa, similar to the protein we detected in insect cells. The little difference in size might come from different post-translational modification, e.g. phosphorylation; or even from the different protein markers being used.

3.3.6.3 Processing of P1/2 protein in insect cells

In addition to the processing of P1, we also found that the P1/2 transframe protein was also processed in insect cells. Until now, no report of the processing of P1/2 has been published.

To study the P1/2 protein, which is expressed via -1 frameshift in nature, we introduced an extra A at the frameshift site so that the protein encoded by the new ORF (no frameshift) was the equivalent to the protein produced from the wild type genome by frameshifting. When SF9 cells were infected by baculovirus (BacP2) encoding this new ORF at MOI of 2, the main product that reacted with anti-P1SP was a protein of about 28 kDa. Since the region identified by anti-P1SP was in the middle of the P1/2 protein, it is clear that P1/2 protein was cleaved at least twice in insect cells, once at the N-terminal region of VPg domain (probably the Glu-Ser site at the start of VPg), and again at the end of the overlapping region of P1 and P2. The first site appears in both P1 and P1/2 proteins, but the second site should be unique to P2.

When SF9 cells were infected by BacP2 at low MOI (about 0.1), in addition to the 27 kDa protein, two other proteins were also identified by anti-P1SP: one had the size of the full-length P1/2 protein, the other was about 73 kDa. The identity of this 73 kDa intermediate product was uncertain. If the first cleavage is at the site upstream of the VPg, the anti-P1SP will detect a 74 kDa protein; if the downstream site is cleaved first, a 71 kDa protein will be identified. Considering the error in estimating the size of the protein, both assumptions are possible. Introducing the Pk tag at the C terminus of P1/2 should answer this question.

At present, we have not demonstrated that the cleavage of P1/2 in insect cells was mediated by the P1 serine protease, though this is the favoured hypothesis, at least for the cleavage at the upstream site.

3.4 Studies of PLRV polyproteins *in vitro*

3.4.1 Introduction

In vitro translation studies of PLRV RNA were first carried out by Mayo *et al.* (1982). When programmed with PLRV genomic RNA, two prominent bands of 70 kDa and 28 kDa, corresponding to ORF0 and ORF1 respectively, appeared in a wheat germ extract translation system. The largest product was 125 kDa, presumably being the P1/2 transframe protein. Several minor bands with intermediate sizes were also detected, and might be the result of premature termination (Mayo *et al.*, 1989) or proteolytic cleavage (Miller *et al.*, 1995).

Uncoupled *in vitro* transcription and translation systems were used to investigate the frameshift within the overlap of ORF1 and ORF2 of PLRV and BWYV (Prüfer *et al.*, 1992; Kujawa *et al.*, 1993; Garcia *et al.*, 1993). A slippery heptanucleotide motif and downstream stem-loop or pseudoknot has been shown to be involved in the frameshifting.

In Section 3.3, the P1 and P1/2 proteins have been shown to be processed into several smaller proteins, which presumably carry out different functions (i.e. protease, VPg, RdRp) in the life cycle of PLRV. This section describes the work studying the processing of these polyproteins using coupled *in vitro* transcription and translation systems (TnT).

The recently developed coupled *in vitro* transcription and translation system bypasses many of the steps in standard *in vitro* translation systems, by incorporating transcription

directly in the translation mix. Plasmid DNA containing a T7, T3 or SP6 RNA polymerase promoter is used as template. In combination with the pGEM-T Easy system, analysis of target genes can be carried out conveniently. Target gene is amplified by Taq polymerase in a PCR. The product has a single overhanging A at its 3' ends and anneals to the vector DNA, which has 3' end T's, in a ligation reaction. Recombinant plasmid with the right orientation is selected and used in TnT reaction.

3.4.2 Construction of plasmids

The construction of plasmids pGEM-P1, pGEM-P1Pk, pGEM-P2 and pGEM-P1GV are shown in Figure 3-27. The proteins encoded in these plasmids were named P1, P1Pk, P1/2 and P1GV, and represented the native P1 protein, P1 with a C-terminal Pk tag, P1/2 transframe protein and P1 protein with Gly-355 to Val mutation, respectively. DNA fragments encoding P1 and P1/2 sequences were amplified using corresponding oligonucleotide primers and template plasmids, purified from agarose gel and ligated into pGEM-T Easy vectors. White colonies were selected and the presence and orientation of inserts were confirmed by digestion with appropriate restriction enzymes.

The oligonucleotide PkMfe is the reverse primer encoding the Pk tag. The oligonucleotide 390 complements the PLRV sequence 3531-3514, about 20 n.t. downstream of the stop codon of P2 (see Materials and Methods for their sequences).

3.4.3. Rabbit reticulocyte lysate vs. wheat germ extract

Both rabbit reticulocyte lysate and wheat germ extract systems were tested using pGEM-P1 to program TnT reactions, as shown in Figure 3-28. In both systems, the major product was a protein of about 66 kDa, similar to the protein we detected in the

Plasmid	Protein Encoded	Forward Primer	Backward Primer	Template plasmid
pGEM-P1	P1	304	305	pBac-P1
pGEM-P1Pk	P1Pk	304	PkMfe	pBac-P1Pk
pGEM-P2	P1/2	304	390	pBac-P2
pGEM-P1GV	P1GV	304	305	pSAB53

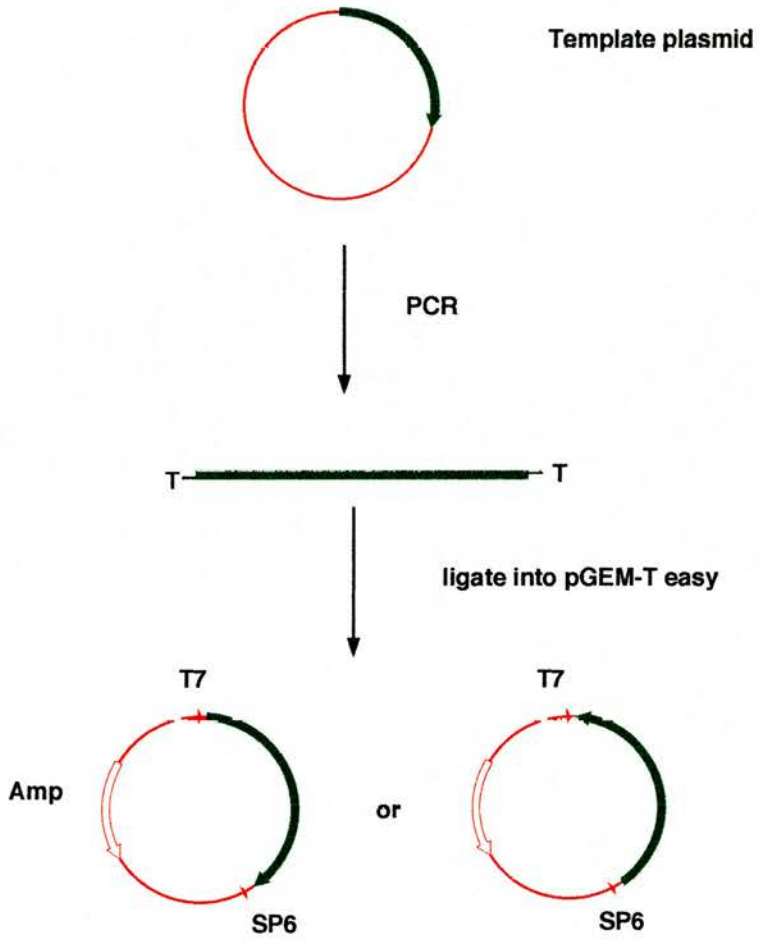


Figure 3-27 Construction of plasmids pGEM-P1, pGEM-P1Pk, pGEM-P2 and pGEM-P1GV

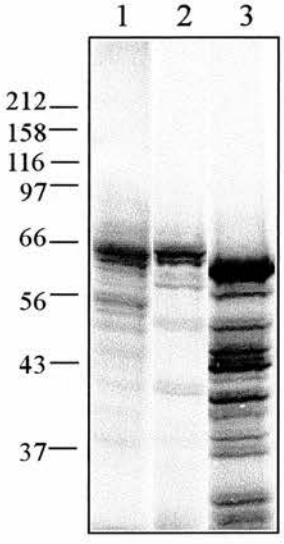


Figure 3-28 *In vitro* translation: rabbit reticulocyte lysate vs. wheat germ extract

pGEM-P1 was used to program TnT reaction. Lane 1, rabbit reticulocyte lysate; lane 2, wheatgerm extract.

Lane 3, control plasmid expressing luciferase (wheat germ extract). The expression of luciferase in rabbit reticulocyte lysate was similar to that in wheat germ extract (data not shown).

baculovirus system, but about 3 kDa smaller than calculated molecular weight of full-length P1 protein. It was possible that this band was the N-terminal truncated form of full-length P1, with the N-terminal 20-30 hydrophobic residues being removed by signalase or other protease. However, without further evidence, this 67 kDa band is interpreted as the full-length P1 protein. It is not unusual that a protein shows an apparent molecular weight on SDS-PAGE somewhat different from the calculated molecular weight.

Some minor products with smaller molecular weights were also synthesized, and might arise from internal initiation, which is a common problem of many TnT reactions. Other explanations include premature termination, and proteolytic cleavage. Since this problem was more serious in rabbit reticulocyte lysate than in wheat germ extract, all subsequent experiments were done using the latter system.

3.4.4 *In vitro* translation of PLRV proteins

Expression of individual proteins

The ORFs encoding P1, P1Pk, P1/2 and P1GV proteins were expressed in the WGE TnT system programmed with plasmids pGEM-P1, pGEM-P1Pk, pGEM-P2 and pGEM-P1GV (Figure 3-29). For plasmids pGEM-P1, pGEM-P1Pk and pGEM-P1GV, the main product was a band of some 67 kDa. All three plasmids also synthesized several minor products of smaller molecular weights. It was interesting that most of the products synthesized by pGEM-P1Pk were a little larger than equivalent products synthesized by pGEM-P1 and pGEM-P1GV, and suggests that those proteins synthesized by pGEM-P1Pk shared the same N-terminus with others, but had an extra Pk tag at its C-terminus.

For plasmid pGEM-P2, the main product was about 116 kDa. Similar to other plasmids, pGEM-P2 also produced several minor products, whose origin was not clear.

Expression using full-length PLRV genome

Plasmid pSAB53 contains the whole PLRV genome immediately downstream of a T7 promoter. The results of a TnT reaction programmed with this plasmid are shown in Figure 3-29 (lane 5). Although the 'background' was higher, the result was similar to that of the *in vitro* translation programmed with viral genomic RNA (Mayo *et al.*, 1989). Three major products were synthesized, being 28 kDa, 67 kDa and 116 kDa. The 28 kDa protein was encoded by the ORF0, the 67 kDa P1 protein was believed to be synthesized according to the leaky scanning model, and the 116 kDa protein co-migrated with the P1/2 transframe protein synthesized by pGEM-P2. This protein is equivalent to the 125 kDa protein observed by Mayo *et al.* (1989) and supports the prediction that the 125 kDa protein is the P1/2 transframe protein (Mayo *et al.*, 1989).

The frameshifting signal was apparently recognized by the wheat germ extract system, since P1/2 protein must be synthesized via a -1 frameshift. The efficiency of the frameshift event, however, seems higher than reported in protoplasts (1%): it was estimated to be about 10% in wheat germ extracts.

3.4.5 Detecting protein processing in TnT reaction

As shown above, the main products of TnT reaction programmed with plasmids containing P1 and/or P2 sequences were the full-length proteins encoded in those ORFs. No processing products were clearly detected, though some minor bands of smaller sizes

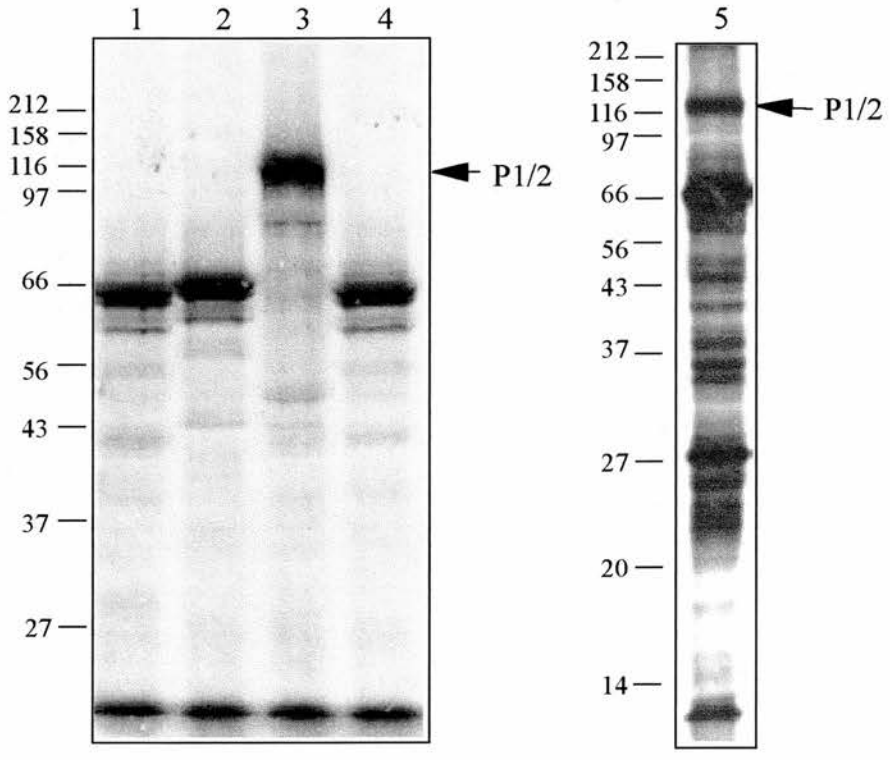


Figure 3-29 *In vitro* expression of PLRV proteins

TnT reactions were programmed with pGEM-P1 (lane 1), pGEM-P1Pk (lane 2), pGEM-P2 (lane 3), pGEM-P1GV (lane 4) and pSAB53 (lane 5). The transframe protein P1/2 is indicated by arrows.

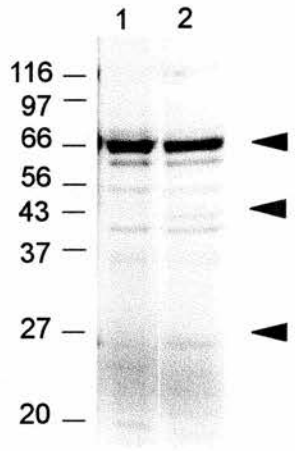


Figure 3-30 Processing after longer incubation

Coupled transcription/translation reactions were programmed with plasmid pGEMP1 and incubated for 90min (lane 1), or, the translation was arrested and subsequently incubated for an additional 2hr (lane 2). The positions of P1 (70kDa) and the cleavage products (43kDa and 27kDa) are indicated.

appeared, but they could also be the results of internal initiation or premature termination.

It is possible that processing by P1 protease might be a slow reaction. Therefore, the TnT reactions were left at 30°C for a longer incubation period after adding RNase A and cycloheximide to arrest further transcription and translation. As shown in Figure 3-30, after the longer incubation, two new bands of about 43 kDa and 27 kDa did appear in the TnT reaction programmed with pGEM-P1. The 27 kDa product co-migrated with the product observed by Western blotting of the baculovirus extracts. This result suggests a slow proteolytic processing did occur in wheat germ extract system, possibly mediated by the P1 protease.

It is possible that the processing is mediated by the P1 protease *in trans*. Since a very small amount of protein is synthesized in most *in vitro* translation systems, such processing is usually difficult to observe. In my case, both the enzyme and the substrate were P1 protein, whose concentration was very low. To tackle this, P1 protein from insect cells was used as an extraneous protease source. Labeled P1 protein synthesized in TnT was incubated with extract of SF9 cells infected with BacP1. Unfortunately, the lysis of insect cells released a lot of proteases, which degraded the TnT products non-specifically.

3.4.6 Membranc-anchoring of P1

The N-terminal region of P1 was very hydrophobic and was thought to be involved in membrane binding. To test this hypothesis, a TnT reaction was carried out in the presence of canine microsome membranes (CMM). The CMMs were derived from endoplasmic reticulocyte and were able to carry out membrane insertion, translocation,

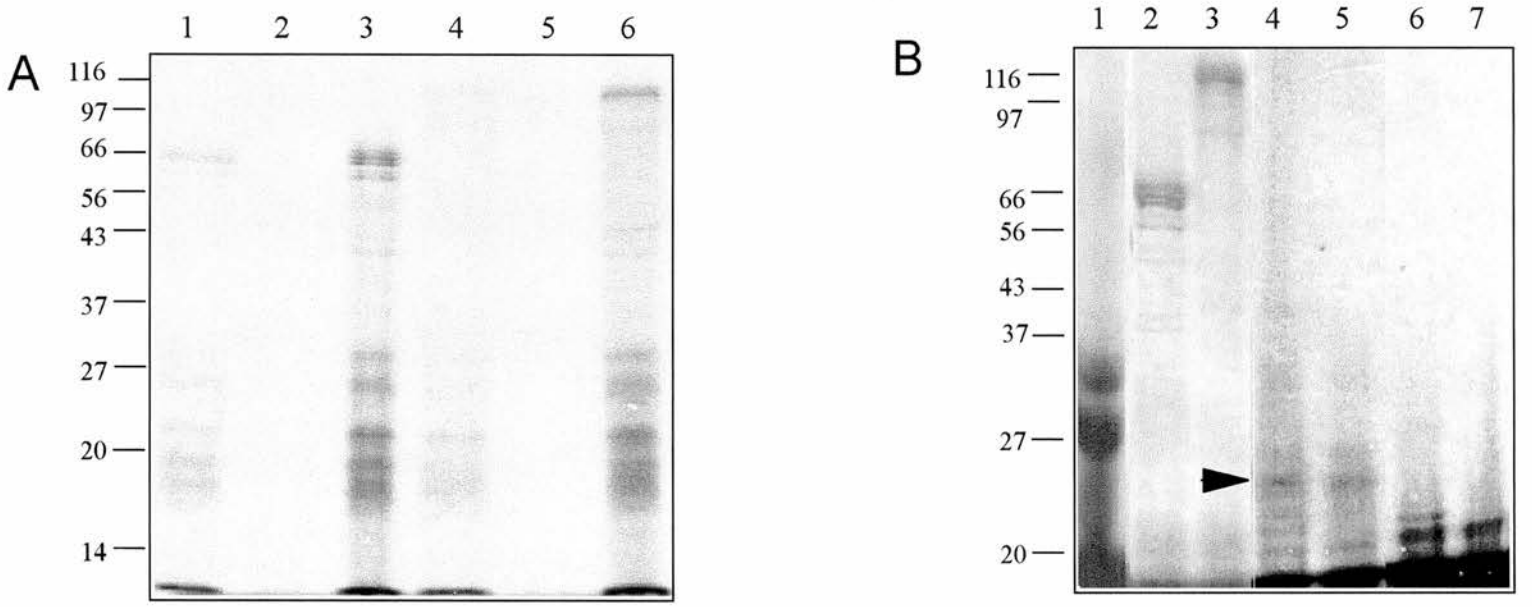


Figure 3-31 Membrane association of P1 and P1/2 proteins

A. TnT reaction was programmed with pGEM-P1 (lane 1-3) and pGEM-P2 (lane 4-6) in the presence of canine pancreatic microsomal membrane (CMM). Lane1, 4: TnT reaction; lane2, 5: supernatant after centrifugation at 200,000xg, 30 min; lane 3, 6: microsomal membrane pellet.

B. Proteinase K protection. Lane 1, control translation of *E. coli* β -lactamase showing that CMM was functional. Lane2, 4 and 6, TnT reaction programmed with pGEM-P1; lane 3, 5 and 7, pGEM-P2. Lane 2-3, TnT reaction in the presence of CMM; lane 4-5, digestion with proteinase K on ice for 1 hour, the 22-23 kDa protein resistant to digestion were shown by arrow; lane 6-7, digestion with proteinase K in the presence of 0.7% Triton X-100 (to dissolve membrane).

signal peptide cleavage and glycosylation. After the TnT reaction was completed, the membranes were collected by ultracentrifugation at 70 000 rpm (200,000xg) for 30 min. As shown in Figure 3-31, both P1 and P1/2 proteins associated with the membranes.

Protease protection experiments were also carried out to determine the orientation of the proteins inserted into /associated with the membrane. A TnT reaction programmed with pGEM-P1 or pGEM-P2 was carried out in the presence of CMM. The translation products were treated with proteinase K for 1 hour on ice with and without the addition of Triton X-100. If P1 or P1/2 protein enters the lumen of the membrane, it should be protected from digestion by proteinase K in the absence of Triton X-100. The results showed that P1 and P1/2 did not enter the lumen and were degraded by proteinase K. In the absence of Triton X-100, a cleavage product of 22-23 kDa appeared and was resistant to further degradation. This protein was degraded in the presence of Triton X-100, which could dissolve the membrane. The explanation is that P1 and P1/2 were inserted into the membrane in such an orientation that they were facing outside. The region resistant to degradation by proteinase K might represent the membrane anchor. No signal peptide cleavage or glycosylation was observed.

3.4.7 Discussion

3.4.7.1 Expression of PLRV nonstructural proteins in the TnT system

In vitro transcription and translation have made a tremendous contribution to our knowledge of RNA viruses. In this section, we described the expression of PLRV non-structural protein, P1, P1Pk, P1/2 and P1GV, in TnT systems programmed with recombinant plasmids derived from pGEM-T easy vector. In each case, the main product was the ' full-length' protein, in addition to several smaller proteins. For P1,

P1Pk or P1GV, the ' full-length' product has an apparent molecular weight 2-3 kDa smaller than predicted.

The protein products of smaller sizes could be the result of initiation of translation at different AUGs within the N-terminal region of ORF1. Two candidates of such AUG are present: one is 12 residues downstream of the first AUG, the other is 37 residues further downstream. According to the leaky scanning model, initiation of translation is preferred at AUG in a context consensus to Kozak motif, CC[A/G]CC AUGG. Comparison of the contexts of the first three AUGs of ORF1 shows that the third AUG is in a nearly optimal context, but the first AUG is also in a good context.

first	UAAUCAUGA
second	UCUUCAUGU
third	CCACUAUGA

In nature, P1 is translated from genomic RNA, in which ORF1 overlaps with upstream ORF0. It is believed P1 is translated by the leaky scanning mechanism. Further experiments are required to answer which of the above three AUGs is the start codon actually being used by ORF1. It is interesting to note that in insect cells, the protein produced in SF9 cells infected with BacP1GV is 67 kDa, as is the protein produced in coupled transcription and translation reaction programmed with the plasmid containing entire PLRV genome. Currently, it is assumed that the ' full length' P1 protein has an apparent molecular weight of 67 kDa.

3.4.7.2 Difficulty in detecting protein processing in TnT system

Cell-free transcription and translation systems (rabbit reticulocyte lysate or wheat germ extract; coupled or uncoupled) are frequently used tools in the study of viral proteases and polyprotein processing. If a precursor-product pattern, similar to that found in infected cells, can be established, or similar patterns are obtained in different cell-free system, a viral protease is usually believed to be responsible for the cleavage. Cell-free systems also provide an easy way to test the effects of different inhibitors, and thus shed light on what type of protease is in action.

Although most of the viral proteases listed in Table 1-2 show proteolytic activities in at least one of the *in vitro* systems, it was difficult to detect the self-cleavage of PLRV P1 protein in such *in vitro* systems. Internal initiation was a severe problem in rabbit reticulocyte lysate. In wheat germ extract, the main product was the full-length P1 protein, though minor amounts of smaller protein corresponding to the product observed in insect cells did appear after long incubations. Extracts in BacP1 infected insect cells were tested as an extraneous source of P1 protease, but cellular or baculoviral proteases released in the extracts may invalidate its usage.

It has been reported that such cell-free systems contain inhibitors or lack co-factors required for the proteolytic activity of some viral proteases. It was shown that the cleavage of the 200 kDa polyprotein of cowpea mosaic virus into the 35 kDa and 170 kDa products required DTT and ATP (Pelhem, 1979; Tian *et al.*, 1986;). Shih *et al.* (1987) found that cleavage of this 200 kDa precursor was effectively inhibited by inhibitor activity in wheat germ. Verchot *et al.* (1991) have shown that the 35 kDa P1 protease of tobacco etch virus (TEV) is able to function in wheat germ extract, but not in

reticulocyte lysate, and the author proposed that an inhibitor is likely to exist in the latter. Due to the limitation of time, the possibility of the presence of inhibitors or the lack of co-factors for the processing of P1 protein in the wheat germ extract was not investigated.

Another possibility is that the P1 protease might lack *cis* activity and mainly function in a *trans* fashion, which requires an interaction between two or more P1 molecules. Generally, *trans* action is more difficult to observe in cell-free systems and needs longer incubations, due to the extremely low concentrations of proteins synthesized in such systems. Picornavirus 3C protease produced in reticulocyte lysate has been shown to cleave precursors of coat proteins (*in trans*) efficiently. The 3C protease might have a much higher k_{cat}/K_m than PLRV P1 protease.

3.5 Mutational Analysis of P1 Proteolytic Domain

3.5.1 Introduction

The presence of a chymotrypsin-like proteolytic domain in PLRV P1 protein has long been suspected (Koonin and Dolja, 1993; Miller *et al.*, 1995; Mayo and Ziegler-Graff, 1996). Now that the P1 protein shows proteolytic activity in insect cells, site-directed mutagenesis experiments were carried out to confirm that the proposed protease domain was responsible for the cleavage observed, and to determine the essential residues at the active site. This approach has been used to characterize almost all the known proteases.

In the classic triad of chymotrypsin, Asp102-His57-Ser195, forms an electron relay system at the active centre. In the case of PLRV, sequence alignment analysis conducted by Koonin and Dolja (1993) and Miller *et al.* (1995) gave similar predictions: residues His-255, Asp-286 and Ser-354 form the putative catalytic triad. These three residues are conserved not only among P1 protein of subgroup II luteoviruses, but also in a number of viral chymotrypsin-like proteases. Therefore, they are suitable targets of mutation. Another two residues, His-234 and Ser-393, are also conserved among subgroup II luteoviruses and at positions near the putative catalytic His and Ser, so the possibility that they may be part of the catalytic triad was also investigated.

All the five residues were mutated to alanine, which has a small neutral side chain (-CH_3) and presumably does not interfere very much with the secondary structure of the polypeptide chain. Ser-354 was mutated to cysteine as well as alanine, because in some chymotrypsin-like proteases, the cysteine can function as a nucleophile and render the mutant protease full or part activity. Asp-286 was converted to glutamic acid, which is

the electrophile of the triad in some chymotrypsin-like proteases. For His-255, tryptophan was chosen only because both histidine and tryptophan have aromatic side chains.

Table 3-2 Oligonucleotides and cloning strategies used in the generation of site-directed P1 mutants

Plasmid	Mutation	Primer1	Primer 2	Primer 3	RE
pBacP1-2	$H^{234} \rightarrow A$	aHA	5637	304	*Xho* I / *Xba* I
pBacP1-3	$H^{255} \rightarrow W$	bHW	304		*Xho* I / *Xba* I
pBacP1-4	$H^{255} \rightarrow A$	bHA	304		*Xho* I / *Xba* I
pBacP1-5	$D^{286} \rightarrow E$	5DE	304	5637	*Xba* I / *Swa* I
pBacP1-6	$D^{286} \rightarrow A$	6DA	304	5637	*Xba* I / *Swa* I
pBacP1-7	$S^{354} \rightarrow C$	7SC	5637	304	*Xba* I / *Swa* I
pBacP1-8	$S^{354} \rightarrow A$	8SA	5637	304	*Xba* I / *Swa* I
pBacP1-9	$S^{393} \rightarrow A$	9SA	5637	304	*Xba* I / *Swa* I

3.5.2 Construction of plasmid pBacP1Z

pBacP1 has some extra restriction enzyme sites (*Xba* I, *Pst* I and *Xho* I.), which were downstream of the *Eco*R I site and derived from the multiple cloning site of pFastBac1. To remove these sites, pBacP1was restricted with *Eco*R I and *Kpn* I, blunted with T4 DNA polymerase, and then recircularized to re-generate the *Eco*R I site. The new plasmid, pBacP1Z, has only one site for *Xho* I, *Xba* I and *Swa* I, which are present

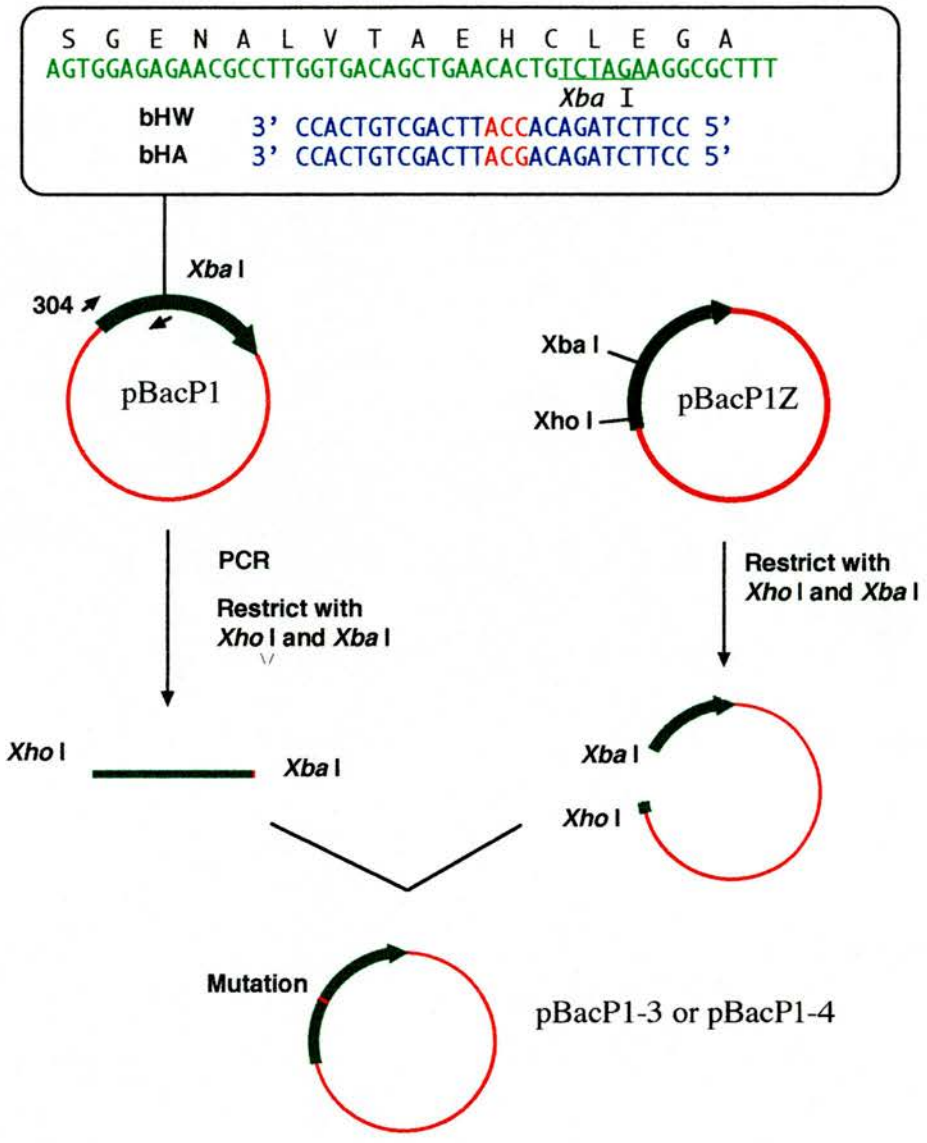


Figure 3-32 Construction of pBacP1-3 and pBacP1-4

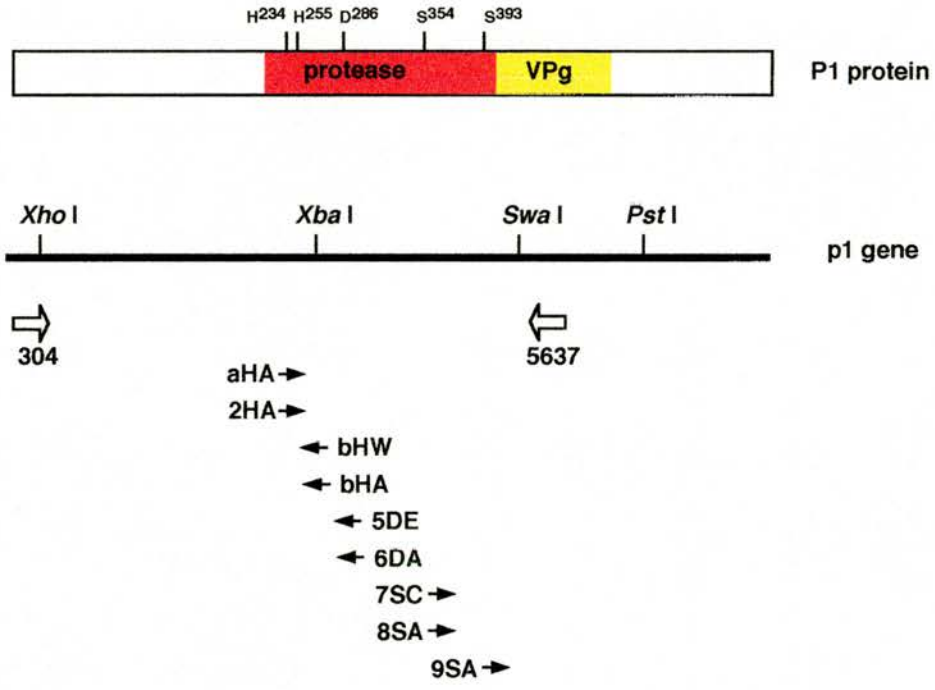


Figure 3-33 Site-directed mutagenesis of P1
Positions of residues being mutated, together with locations and orientations of primers used in the mutagenesis are shown. The N-terminus of protease domain and the C-terminus of VPg have not been precisely determined yet.

in the PLRV P1 sequences (Figure 3-32). In the following mutagenesis work, the wild-type P1 sequences in this plasmid were replaced by mutant sequences generated by PCR, as summarized in Table 3-2 and Figure 3-33.

3.5.3 Construction of plasmids pBacP1-3 and pBacP1-4

To convert His-255 to tryptophan or alanine, PCR was carried out using oligonucleotide 304 as forward primer, and oligonucleotide bHW or bHA as reverse primer, respectively (Figure 3-34). pBacP1 was used as template throughout the mutagenesis work. The sequences of oligonucleotides bHW and bHA, and the amino acid sequences they encode are shown in List of oligonucleotides. Both bHW and bHA have the sense of the (-) strand.

Both oligonucleotides bHW and bHA have the *Xba* I site at its 5'end, which is two nucleotides downstream of the His-255 codon in the P1 sequences and greatly facilitated the cloning work. DNA fragments amplified from PCR were restricted with *Xho* I and *Xba* I, and then ligated into similarly restricted pBacP1Z. pBacP1-3 and pBacP1-4 contain the mutation of His-255→Trp and His-255→Ala respectively.

3.5.4 Construction of plasmids pBacP1-2, and pBacP1-5 to pBacP1-9

Site-directed mutagenesis by PCR

The introduction of other mutations was carried out by a two-step PCR method, modified from the sequential PCR method (Ausebel *et al.*, 1994) according to Dr. H. Liu (personal communication). For each mutation, only one mutant oligonucleotide was

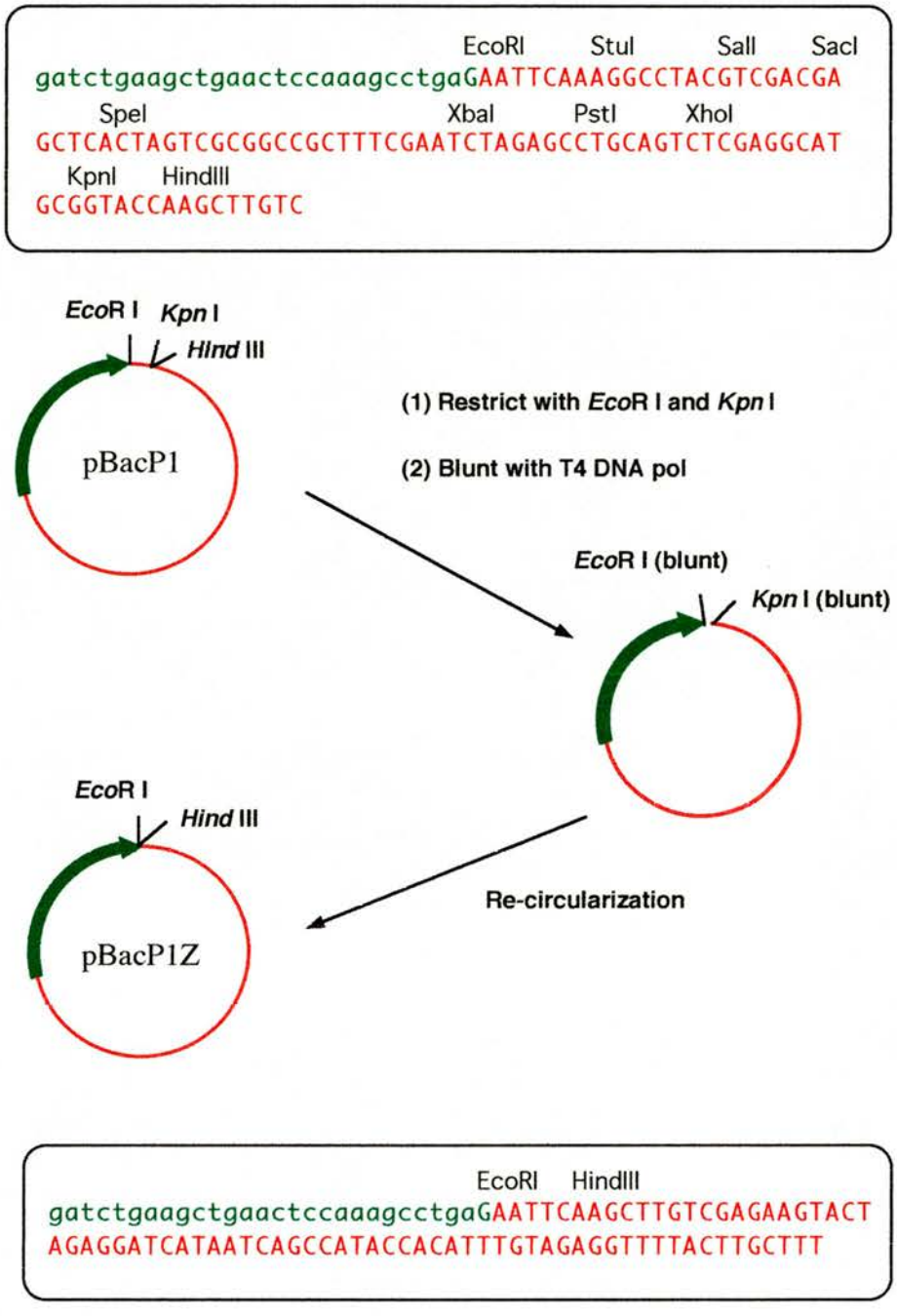


Figure 3-34 Construction of pBacP1Z
The restriction sites between *Eco*R I and *Hind* III were removed so that the *Xho* I and *Xba* I sites within P1 sequence became unique.

used, in addition to the upstream and downstream primers flanking the region to be mutated.

The procedure is presented in Figure 3-35. In the first round PCR, the mutant oligonucleotide (Primer 1) and downstream reverse primer (Primer 2) were used to amplify a relative small DNA fragment that has the point mutation at one end. This DNA fragment was purified from agarose gel and used as a primer in the second PCR: actually only one strand of the product of the first round PCR was used as the reverse primer in the second PCR. The product of the second PCR was restricted with restriction enzymes (RE1 and RE2), purified and then ligated with similarly digested vector plasmid.

In the first round PCR, the $Vent_R$ DNA polymerase (New England Biolab) was used to take advantage of its 3' to 5' proofreading exonuclease activity. The fidelity of the $Vent_R$ DNA polymerase is 5-15 times higher than that of *Taq* DNA polymerase (Mattila *et al.*, 1991). Also due to the 3' to 5' exonuclease activity, the products amplified by the $Vent_R$ DNA polymerase are blunt-ended with no overhanging nucleotide at the 3' end. Therefore, it is unnecessary to treat the DNA fragment amplified in the first round PCR with Klenow fragment.

In the oligonucleotides 5DE, 6DA, 7SC, 8SA and 9SA, the mutant codons were present at the 5' end of the oligonucleotides, it would appear at the 3' end of the DNA fragment used as the reverse primer in the second PCR. This is where this novel method is different from the widely used sequential PCR method, in which two DNA fragments containing the same point mutation are used as template mutually in the second PCR, so there is no mismatch when these two fragments anneal with each other. However, in the procedure used in this study, the plasmid containing the wild type P1 sequences was

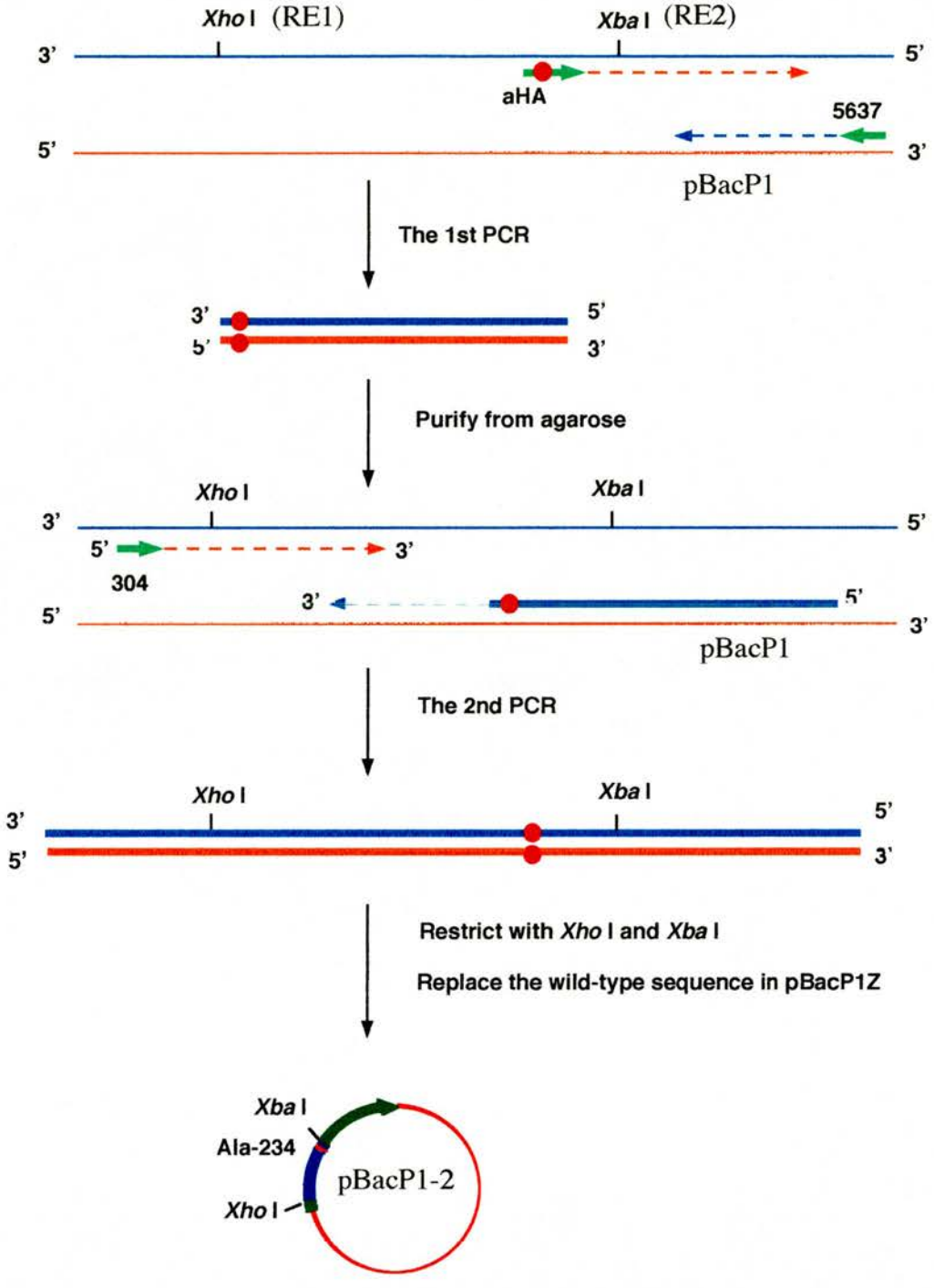


Figure 3-35 Construction of pBacP1-2: introduction of a mutation by a novel PCR strategy
Primer aHA containing the mutation (Ala-234) and the downstream primer 5637 were used in the 1st round PCR, the product of which was then used as one primer in the second round PCR.

used as template in the second PCR, resulting in a mismatch when the backward primer annealed with the plasmid (Figure 3-36a).

Fortunately, the *Taq* DNA polymerase does not have the 3' to 5' exonuclease activity and is not sensitive to the 3' end mismatch. Huang *et al.* (1992) has systematically studied the extension efficiency of mismatched base pairs at template-primer 3' termini with *Taq* DNA polymerase. Their studies have shown that the mispair C(primer)·T is extended with high efficiency, about 10^{-2} compared to a correct A·T base pair. The relative efficiency is about 10^{-6} for A·A, and less than 10^{-6} for C·C. After an examination of the mispairs at the 3' end of the reverse primer in the second PCR, it was estimated that most mispairs had a good chance to be extended by *Taq* polymerase. To increase the chance, the following protocol was carried out:

1. 15 min pre-denaturation at 95 °C
2. (5 cycles) 45 sec denaturation at 94 °C
   90 sec annealing at 40 °C
   90 sec extension at 72 °C
3. (20 cycles)45 sec denaturation at 94 °C
   90 sec annealing at 55 °C
   90 sec extension at 72 °C
4. 7 min final extension at 72 °C

The annealing temperature in the first 5 cycles was lowered to help the extension of the mispair, while the higher annealing temperature in the following cycles was chosen to increase the specificity of the final products. The results of the second PCR was consistent with the studies of Huang *et al.* (1992): the extension efficiency was high for mispair C(primer)·T (e.g. 8SA and 9SA), and low for A·A (e.g. 5DE). In the case of

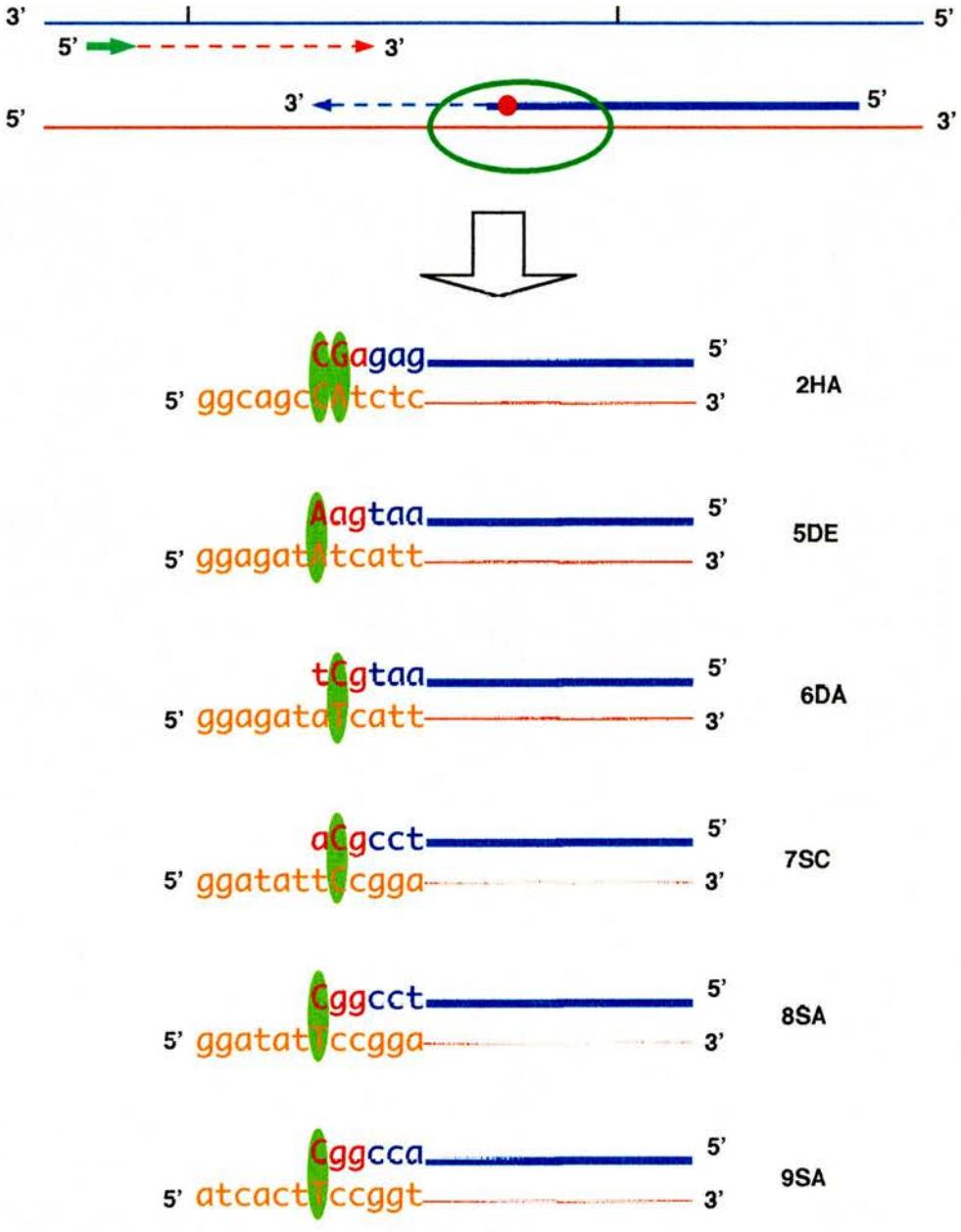


Figure 3-36a Mismatched 3' ends in the 2nd PCR
For the six mutations using the two-step PCR, the 3'ends at the mutation point are shown. For the first mutation (His-234 to Ala), there are two mismatched nucleotides. In other five cases, only one mismatched nucleotide. The efficiency of extending these mismatched 3' end by Taq DNA polymerase is shown in Figure 3-36b.

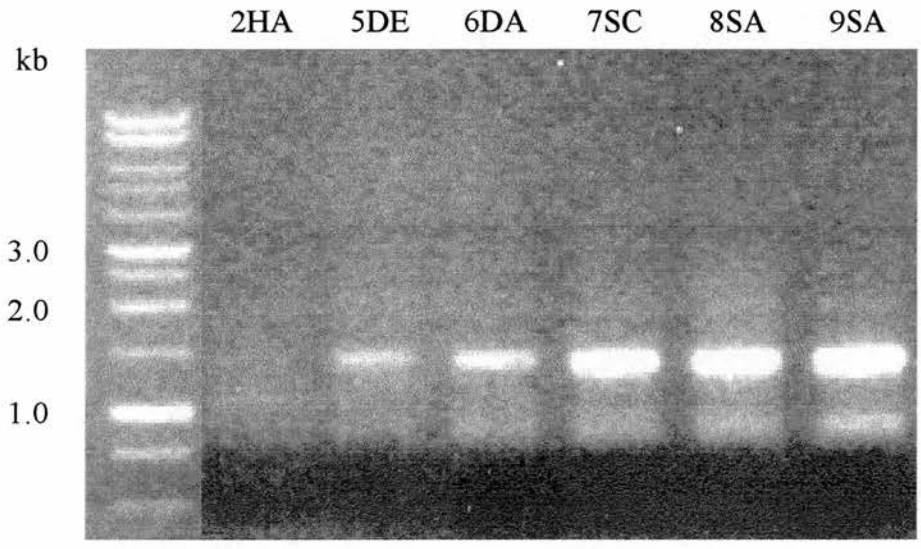


Figure 3-36b Forced extension of mismatched 3' ends in the 2nd PCR
The products of the 1st PCR were used as primers in the 2nd PCR, the 3'ends of which contained mismatched nucleotides with the template plasmid. The efficiency of the extension depended on the number of mismatched nucleotides and the length of the PCR product in the first PCR.

2HA, there were two mispairs at the 3' end, so the extension was too difficult to produce any detectable amounts of product. Besides, the extension could be carried out with high efficiency if the difficult mispair (C·C) was not at the position of the ultimate nucleotide (e.g. 7SC). The size of the DNA fragment amplified in the first round PCR seems to influence the efficiency as well: the larger the fragment, the lower the yield (Figure 3-36b). For 5DE and 6DA, the products of the first PCR were about 870 bp, while in 7SC, 8SA and 9SA, it was about 400bp, 400bp and 290 bp respectively. For 2HA, it was 770 bp, also quite long.

For the mutation of His-234 to alanine, a new oligo (aHA) was designed with a further 12 nucleotides at its 5' end. The second round PCR was carried out with high efficiency under normal conditions (annealing temp. 53°C).

Apparently, the $Vent_R$ DNA polymerase cannot be used in the second PCR. The 3' to 5' proofreading exonuclease activity will let the polymerase detect the mispair and excise the mismatched nucleotides before it carries out the extension. Therefore, the products will have only the wild type sequences.

Introduction of mutant sequences into pBacP1Z

To construct pBacP1-2, the PCR product containing the His-234 to alanine mutation (using oligo aHA) was digested with *Xho* I and *Xba* I, purified from agarose, and then ligated into pBacP1Z, similarly restricted.

In the case of pBacP1-5 to pBacP1-9, the PCR products were restricted with *Xba* I and *Swa* I, purified from agarose, and ligated into *Xba* I and *Swa* I restricted pBacP1Z.

One of the most difficult thing in site-directed mutagenesis is the selection of the transformants that contain the required mutations. If the mutation point is part of the site of a restriction enzyme, the mutant and wild type plasmids will have different physical map, so the positive transformants are easy to identify. For examples, (1)the mutation His-234 (CAT)→Ala (GCT) creates a new *Eco47* III site; (2) mutation of Asp286 (TGA)→Glu (GAA) or Ala (GCT) abolishes the *Eco*R V site; (3) mutation of Ser-354 (TTC) →Cys (TGC) or Ala (GCC) abolishes the *Acc* III site.

In cases where the mutation does not change the physical map, a clean preparation of the vector plasmid is of paramount importance, which can be indicated by few or no transformants in the control using the ligation mixture containing the vector only.

3.5.5 Proteolytic activity of mutant P1 proteins in SF9 cells

Recombinant baculoviruses containing mutant P1 proteins were constructed using the above recombinant "donor" plasmids with the Bac-to-Bac baculovirus expression system described previously.

SF9 cells were infected with recombinant baculoviruses at a MOI of 2, and harvested at 1, 2, and 3 days p.i. The total protein of the cell lysate was analyzed by Western blotting.

The conversion of His-234→Ala did not abolish the proteolytic activity in that the 27 kDa was readily observed. However, this mutation did influence the cleavage because another protein of about 50 kDa was produced in similar amounts. Only trace amounts of protein with similar size were detected when SF9 cells were infected by wild type BacP1. There were also a couple of minor bands of intermediate sizes that were not

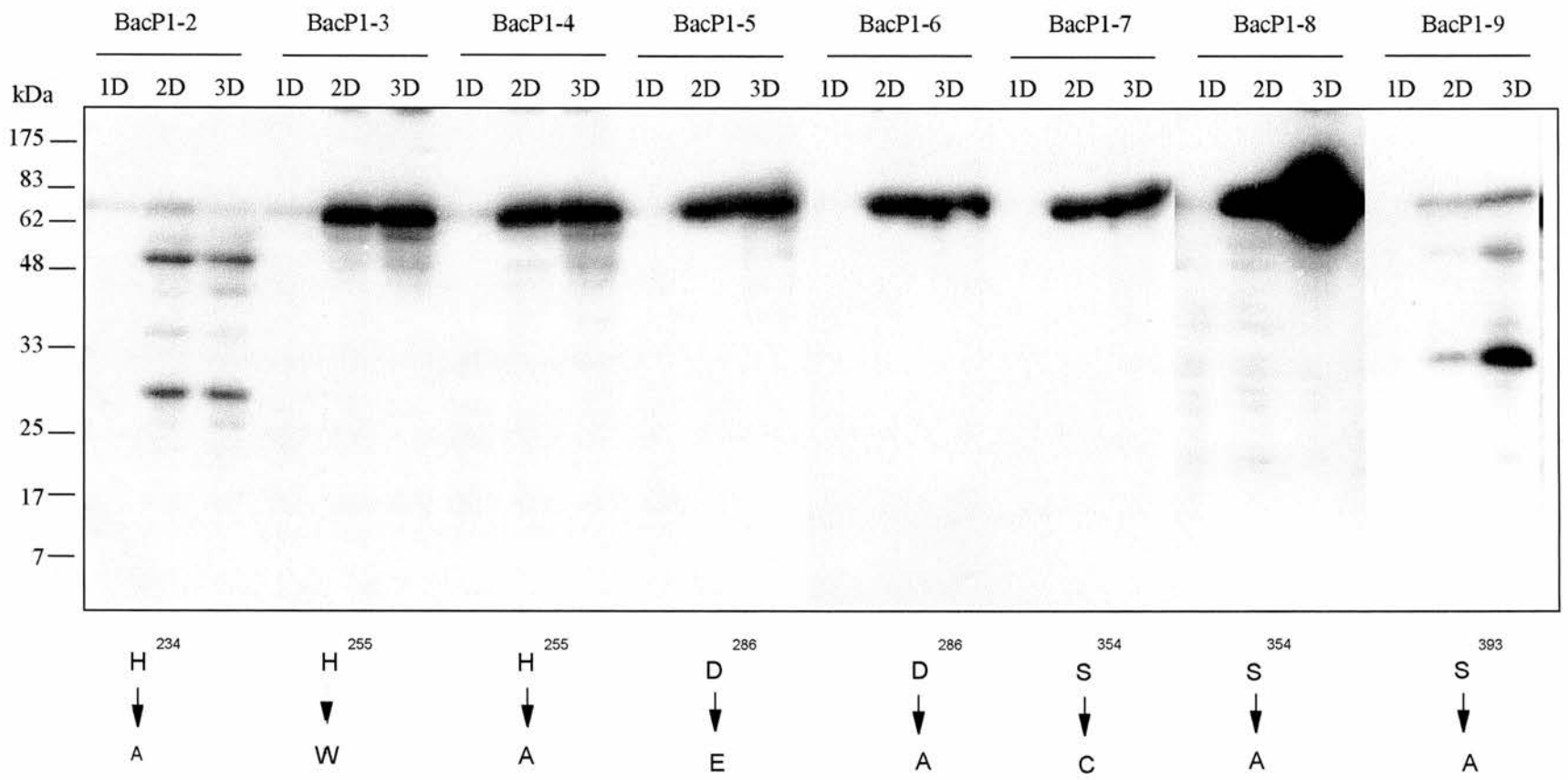


Figure 3-37 Analysis of P1 site-directed mutants

SF9 cells were infected with baculoviruses encoding mutant forms of the P1 protease domain, and were harvested at 1, 2 and 3 days pi. Cell extracts were probed with anti-P1 antisera.

observed in BacP1-infected SF9 cells. For the mutation of Ser-393 to Ala, a proteolytic active enzyme was produced with a cleavage profile similar to that of wild type protease except that slightly more 50 kDa cleavage product was observed and the appearance of the 27 kDa product was delayed a little.

The proteolytic activity was abolished when the residues of putative catalytic triad were mutated (His-255 to Trp or Ala, Asp-286 to Glu or Ala, Ser-354 to Cys or Ala). Only the full-length mutant P1 proteins were observed.

3.5.6 Discussion

3.5.6.1 Mutagenesis using the two-step PCR method

Using DNA fragment as a primer in PCR has been described before (Ausebel *et al.*, 1994). In the sequential PCR method for site-directed mutagenesis, the two PCR products of the first round PCR anneal with each other in the second round PCR and function as primer mutually. The method used in this study is simple, and, in my experience, very efficient. Its advantages over the sequential PCR method includes:

(1) Only one mutant oligonucleotide is needed for each point mutation, instead of two.
(2) Only two reactions are carried out, instead of three (two in the first round, one in the second round).
(3) The overlapping region of the two annealing DNA molecules (in the final PCR) is much longer (normally more than 200 n.t.), instead of about 20 n.t. or less. Therefore, the annealing temperature can be raised to increase specificity.

The limitation of this method is that the molar amount of the DNA fragment used as primer (in the second PCR) is relatively small, compared with oligonucleotide primers, and this might limit the final yield in the second PCR. In my experience, one-tenth of the DNA product (about 0.2 μg) of the first PCR (100 μl volume) is enough. The design of oligonucleotides and choice of restriction enzymes RE1 (upstream of mutation site) and RE2 (downstream of mutation site) is very important (Figure3-35). The mutant oligo can be either strand of the target sequence. The reverse primer (Primer 2) should be chosen so that the product of the first round PCR is about 200-400 n.t. long: it should be long enough to be purified from agarose gel, and short enough for the *Taq* polymerase to generate large (molar) amounts of it. It is better to choose RE1 and RE2 so that the region between them is short to reduce the chance that unwanted mutation is incorporated during PCR, and can be sequenced in one reaction (less than 700 n.t. long).

The choice of $Vent_R$ DNA polymerase in the first PCR and *Taq* DNA polymerase in the second was originally intended to deal with the flaw in the design of oligonucleotides, but it worked well in the normal protocol.

3.5.6.2 Catalytic residues of P1 protease

The mutagenesis experiment shows that the proteolytic activity observed in the insect cells was indeed encoded by the chymotrypsin-like protease domain in the middle one-third of the P1 sequence. My results conforms to the prediction made by sequence alignment and show that His-255, Asp-286, Ser-354 are essential for the catalysis.

Ser-354 is found within the sequence Gly-Tyr-Ser-Gly, which conforms to the conserved motif Gly-Xaa-Ser-Gly around the catalytic serine residue of serine proteases in the Chymotrypsin clan, which includes many cellular serine proteases (e.g.

Chymotrypsin family, α-Lytic endopeptidase family) and some viral serine proteases (e.g. Tobacco etch virus 35 kDa protease and Sindbis virus core protease). This motif is conserved in all the members of subgroup II luteoviruses (poleroviruses and enamovirus). The 3C-like proteases, encoded in picornavirus super-group and making up the majority of viral chymotrypsin-like proteases, have their catalytic cysteine residue in a similar motif of Gly-Xaa-Cys-Gly.

The other two residues, His-255 and Asp-286 were also found to be necessary for the proteolytic activity. The order of the three essential residues in the primary structure of PLRV P1 protein resembles that of cellular chymotrypsin-like serine protease (e.g. chymotrypsin His57-Asp102-Ser195). However, the distance between these residues is shorter than that of cellular counterparts, and similar to that of poliovirus 3C protease (His40-Glu71-Cys147) and TEV NIa protease (His234-Asp269-Cys339). This is not unusual since viral proteases normally have shorter loop regions and more compact structures.

It is conceivable that these three residues probably form a catalytic triad and catalyze the hydrolysis of the peptide bond in a similar manner to that of a classic chymotrypsin-like serine protease, though the final confirmation has to wait for the resolution of the atomic structure of P1 protease.

Many viral chymotrypsin-like proteases, e.g. poliovirus 3C, contain a catalytic triad of His-Glu-Cys with the substitutions of cysteine for serine, and glutamic acid for aspartic acid, as the nucleophile and the base, respectively, during proteolysis. This suggests that serine or cysteine, and aspartic acid or glutamic acid may function interchangeably. Poliovirus 3C protease with the substitution of Cys-147 with serine retains considerable proteolytic activity (Lawson and Semler, 1990). In tobacco etch virus NIa protease, the

catalytic Asp-269 could be replaced with Glu (but with altered cleavage activity), while reduced proteolytic activity (1-2%) was retained when the nucleophilic Cys-339 was replaced with serine (Dougherty *et al.*, 1989). As a conversion of the above mutations, Sindbis capsid protease retains 60% of its autoproteolytic activity when its active site serine is mutated to cysteine (Hahn and Strauss, 1990). For rat trypsin, the k_{cat} of mutant protease (Ser-195 to cysteine) is reduced by a factor of 6.4×10^5, while the k_{cat} of doubly mutated variant (Ser-195 to cysteine and Asp-102 to Asn) was lowered by a factor of 3.4×10^7. In the case of PLRV P1 protease, no proteolytic activity could be detected when active site serine or Asp was mutated to cysteine or glutamic acid, respectively. At the moment, it is uncertain whether the mutation renders the protease inactive or the remnant activity was too low to be detected in the baculovirus system. Experiments with purified active protease should address this question.

3.5.6.3 Non-catalytic residues

Two other residues within the protease domain, His-234 and Ser-393, which are completely conserved among subgroup II luteoviruses, were shown to be non-essential for the proteolytic activity. The significance of their conservation is not known. It is interesting to note that the mutation of His-234→Ala altered the proteolytic activity to some extent: the cleavage seemed to be carried out more slowly, and considerable amounts of a 50 kDa protein were produced. This 50 kDa protein was produced in trace amount by wild type P1 protein and might represent an intermediate product.

Ser-393 is the only conserved serine at the C-terminal region of the protease domain other than Ser-354. It occupies the P7 position relative to the cleavage site at the N-terminus of VPg, and appears not to affect the catalysis of hydrolysis or the site specificity.

Chapter 4

General Discussion

PLRV is an RNA virus with a genome of about 6 kb. It use a variety of strategies in its gene expression, which include:

(1) Synthesis of subgenomic RNA to express downstream genes

(2) Internal initiation of translation by leaky scanning

(3) Ribosomal frameshifting to express overlapping genes

(4) Stop codon suppression

Based on the work described in this thesis, and data published by other researchers, another strategy, which has long been suspected, is finally added to the list:

(5) Polyprotein processing by viral protease

Although none of the above strategies are unique to PLRV or the luteovirus family, it is unusual that so many are present in one virus with a tiny genome.

4.1 P1 and P1/2 are polyproteins

ORF1 and ORF2 are the only two viral genes essential for the replication of luteovirus RNA. Multiple function motifs, i.e. RNA-dependent-RNA polymerase, helicase, and protease, have been identified within these two genes by sequence analysis (Mayo *et al.*, 1989; Habili and Symons, 1989; Koonin and Dolja, 1993). Prüfer *et al.* (1992) has shown that the PLRV transframe protein P1/2 contains a basic nucleic acid-binding

domain within the P2 region, in addition to the –GDD- motif of the polymerase. The hydrophobic sequence -SIEAFCLILLGCITSLI- at the N-terminus of P1 (residues 152-168) is thought to be characteristic for membrane-associated proteins, and the C-terminal part of P1 has been found to exhibit nucleic acid-binding properties. A strongly basic region, -KxKxKKRxRRxxRxK-, is identified in this region (Prüfer *et al.*, 1999). The gene encoding VPg was mapped to the C-proximal part of P1 protein (van der Wilk *et al.*, 1997). The release of the VPg requires a proteolytic activity, and it has been shown that the P1 protein was processed in plants (Prüfer *et al.*, 1999).

In addition to the processing of P1 protein, the transframe protein P1/2 was also found to be processed in insect cells (Section 3.3). In insect cells, P1 was found to be cleaved at the N-terminal region of the VPg domain; whilst the P1/2 was cleaved at both sides of the VPg.

Within the P1 protein, there might be more than one processing site. The work using *E. coli* expression system suggested that another site of processing is present N-terminal to the protease domain, though we can't demonstrate this conclusively. Unfortunately, it cannot be confirmed whether this site is cleaved in insect cells or not at present, since the anti-P1SP antibodies only detected the region of P1 that is C-terminal to the protease domain.

Considering the data collected in this study, and those previously published, I believe that both P1 and P1/2 are polyproteins, which are processed to give rise to several proteins with distinct functions. Figure 4-1 shows a scheme proposed here for the processing of the two polyproteins.

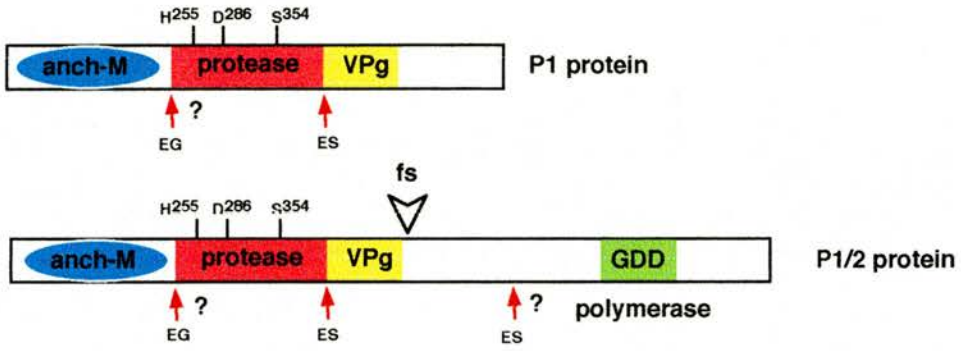


Figure 4-1 A proposed processing scheme of the P1 and P1/2 proteins
The locations of the cleavage sites are predicted on the basis of the sizes of the processing product and the possible similarity of site specificity among the PLRV P1 protease, the EAV serine protease and some 3C-like viral proteases. The cleavage between the protease and VPg domains has been shown to be mediated by the P1 protease, while the protease responsible for the other two cleavages remains to be confirmed. The postulated membrane achor (anch-M) is indicated in blue colour.

P1 and P1/2 proteins share a common N-terminal region, which includes, from N-termini to C-termini, hydrophobic membrane-anchoring domain, protease domain and VPg domain. The C-terminus of VPg domain remains uncertain. The VPg was reported to be about 8 kDa (Mayo *et al.*, 1982), so both P1 and P1/2 protein should have the entire VPg domain, since the frameshift site is some 10 kDa downstream of the start of VPg. Furthermore, there should be a cleavage site before the frameshift site, which is cleaved at some stage of the life cycle of PLRV to release the C-terminus of VPg. However, we have no evidence that such a site is cleaved in insect cells, *E. coli* or wheat germ extract, though it should be noted that the 5' end of the PLRV genomic RNA was not present.

The overlapping region downstream of the frameshift site encodes different polypeptide sequences for P1 and P2. Both segments are rich in basic amino acids, and have been shown to bind to nucleic acid non-specifically, and are thought to be involved in RNA replication. It is also worth to note that the RNA corresponding to this overlapping region contains stem-loop and pseudoknot structures, proposed to be involved in frameshifting (Prüfer *et al.*, 1990; Garcia *et al.*, 1993; Kjuawa *et al.*, 1993).

Within the P1/2 transframe protein, a cleavage site was found to reside just downstream of the overlapping region. This will separate the nucleic-acid-binding motif from the downstream part of P2, which includes the proposed helicase motif and the GDD polymerase motif. Therefore, the polymerase domain could exist as several C-coterminal forms. No evidence is available as to which form is the active one, but it obviously provides a possible way to regulate the viral polymerase activity.

The N-terminus of VPg was determined to be Ser-400 (van der Wilk *et al.*, 1997). It is proposed that P1 protease cleaves at a dipeptide, Glu-Ser. The P1 protease seems to have a site specificity similar to the 3C proteases of the picornaviruses, which preferably cleave at dipeptide Gln (or Glu)/Gly (Ala or Ser). Suppose the P1 protease does cleave at the Glu-Ser site between the protease and VPg domains, according to the sizes of cleavage products, putative cleavage sites are predicted within the two polyproteins, as shown in Figure 4-1. In this scheme, the PLRV P1 protease cleaves at one Glu/Gly and two Glu/Ser sites, showing even more similarity in site specificity to another 3C-like seine protease, the nsp4 of equine arthritis virus, which also cleaves at Glu/Ser and Glu/Gly sites (Snijder *et al.*, 1996).

All viral proteases found to date are synthesized as a domain in a polyprotein. For (+) stranded viruses, the majority of the chymotrypsin-like proteases occupy the central position in the conserved layout of (N) helicase-VPg-protease-polymerase (C). This organization is typical of the picornavirus-like supergroup, which includes picorna-, como-, poty-, sequi-, and caliciviruses, and the proteases are referred as 3C-like. For PLRV, the serine protease domain was once predicted to be 3C-like and reside within the array VPg-protease-polymerase (Koonin and Dolja, 1993; Miller *et al.*, 1995; Mayo and Ziegler-Graff, 1996). Recently, it is clear that in PLRV, the protease is in a rare (if not unique) array: protease-VPg-polymerase. Another difference between PLRV and picornaviruses is that in picornaviruses, the separation of the VPg and protease domain is normally carried out by a quick, cotranslational, *cis* cleavage by 3C; whilst in PLRV, our data suggest that the separation might be mediated by a slow, *trans* cleavage. However, whether there is a true relation between the arrangement of conserved domains and the lack of *cis* activity of P1 protease in PLRV is still uncertain.

4.2 A new member of viral proteases

Among the viral proteases list in Table 1-2, the largest group is the so-called Chymotrypsin-like proteases (CHL), which adopt a Chymotrypsin-like fold containing two β-barrels and most of them have a nucleophilic Cys at active centre (e.g. 3C proteases). The P1 protease of PLRV is the first protease found in the luteovirus family with some properties similar to that of 3C proteases, as well as some unique characteristics.

A canonical catalytic system, composed of His-255, Asp-286 and Ser-354, is employed in the P1 protease. This form of catalytic triad appears in P1 protease of TEV, NS3 protease of yellow fever virus (YF) and HCV, capsid protease of SIN, and nsp4 protease of HEV. It has been proposed that an ancestral Cys protease may have preceded the chymotrypsin-like serine proteases in evolution. It is interesting that P1 protease of PLRV probably shares similar site specificity with the 3C-like proteases (see discussion above). The 3C-like proteases have conserved threonine and histidine in their substrate-binding pocket, while in PLRV P1 protease, the threonine is also conserved (Figure 4-2). However, the histidine residue, which is conserved in other 3C-like proteases and believed to be involved in substrate binding, is replaced by a leucine residue. The significance of this substitution is not clear, but it may well be that this substitution may affect the cleavage site specificity of the PLRV P1 protease.

Remarkably, it was not possible to detect *cis* activity of PLRV P1 protease. One characteristic of a *cis* cleavage is the insensitivity to dilution when it is assayed *in vitro*. Indirect evidence suggests that P1 protease might predominantly cleave at the downstream Pro/VPg site *in trans* in insect cells. In the co-infection experiment (Section 3.3), the cleavage at this site, which was carried out *in trans*, was sensitive to the MOI

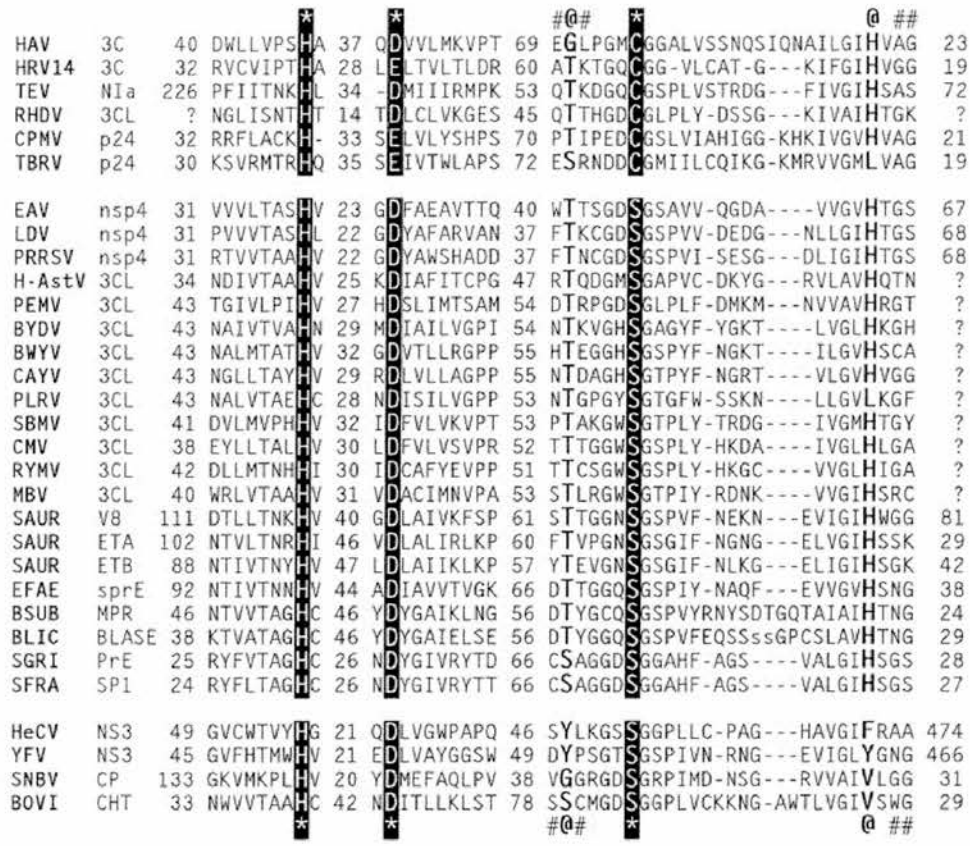


Figure 4-2 Sequence alignment of representative chymotrypsin-like proteases (Snijder *et al.*, 1996)

The upper block contains viral 3C-like cysteine proteases. The middle block contains (putative) 3C-like serine proteases. Proteases in these two blocks have conserved Thr and/or His in their substrate-binding pocket (indicated with @). The lower block contains non-3C-like viral serine protease and bovine chymotrypsin, the prototype enzyme. HAV, hepatitis A virus; HRV14, human rhinovirus 14; TEV, tobacco etch virus; RHDV, rabbit hemeorrhagic disease virus; CPMV, cowpea mosaic virus; TBRV, tomato black ring virus; H-AstV, human astrovirus; PEMV, pea enation mosaic virus; BYDV, barley yellow dwarf virus; BWYV, beet Western yellows virus; CAYV, cucurbit aphid-transmitted yellow virus (now named cucurbit aphid-borne yellow virus, CABYV); PLRV, potato leafroll virus; SBMV, Southern bean mosaic virus; CMV, cocksfoot mottle virus; RYMV, rice yellow mottle virus; MBV, mushroom bacilliform virus (Australian isolate); SAUR ETA, ETB, and V8, *Staphylococcus aureus* exotoxins A and B and V8 protease, respectively; EFAE, *Enterococcus faecalis*; sprE, serine proteinase E; BSUB, *Bacillus subtilis*; MPR, metalloprotease; BLIC, *Bacillus licheniformis*; BLASE, *B. licheniformis* protease; SGRI, *S. griseus*; PrE, protease E; SFRA, *Streptomyces fradiae*; SP1, serine protease 1; HeCV, hepatitis C virus; YFV, yellow fever virus; SNBV, Sindbis virus; BOVI CHT, bovine chymotrypsin.

of the baculovirus that provided active P1 protease. The explanation is that the higher the MOI, the more P1 protease being synthesized, and the quicker the processing. On the other hand, when SF9 cells was infected by BacP1 only, the self-processing of P1 protein was observed to be slower at low MOI (0.1) than at high MOI (2). Therefore, the processing of P1 was sensitive to MOI, as could be easily explained if the processing happened mainly *in trans*.

4.3 Implications of protein processing in the replication of PLRV

Considering the data collected in baculovirus, *E. coli* and *in vitro* systems, the model proposed by Prüfer for the maturation of VPg, and current knowledge on replication of (+) RNA viruses, it is possible to develop a hypothesis on protein processing in the replication of PLRV, which includes a series of events (Figure 4-3):

Initial translation: P1 and P1/2 transframe proteins are translated from genomic RNA by a mechanism involved leaky scanning and -1 frameshifting. The ratio of P1/2 to P1 is about 1-2:100.

Translocation: The hydrophobic N-terminal region within P1 and P1/2 proteins targets them to membrane structures, where viral replication is presumed to occur. PLRV genomic RNA is also carried there by the nucleic acid-binding capacity of P1 and/or P1/2.

Proteolytic Processing: Enrichment of P1 and P1/2 in membrane structures facilitates their processing by the serine protease domain within P1 sequences, which occurs mainly *in trans*. In P1 protein, the protease domain is separated from VPg*, which contains the C-terminal nucleic acid-binding domain of P1. The RNA-dependent RNA polymerase domain is cleaved from the rest of P1/2 protein, and may or may not associate with the 28 kDa product containing VPg sequences. The protease domain may be released from membrane anchor as well.

RNA synthesis: With the help of other viral or cellular proteins, the viral polymerase starts to synthesize (-) strand genomic-length RNA. The VPg* then binds to the 3' end of (-) RNA, and function as a primer during the synthesis of (+) genomic RNA.

Virus assembly: Subgenomic RNA is transcribed from (-)RNA and viral structural proteins are synthesized. Viral genomic RNA is then encapsidated and the C-terminal nucleic acid-binding domain is replaced by coat protein and released from VPg by an unknown mechanism.

Above hypothesis is consistent with data available so far, but many aspects of it are speculative. More experiments are required to modify, refine and confirm them. For example, currently no evidence is available as to how and when the C-terminus of VPg is generated. Many other possibilities can be conceived. The polymerase domain might be active within the P1/2 polyprotein, and minus strand RNA might be synthesized before P1/2 is targeted to membrane.

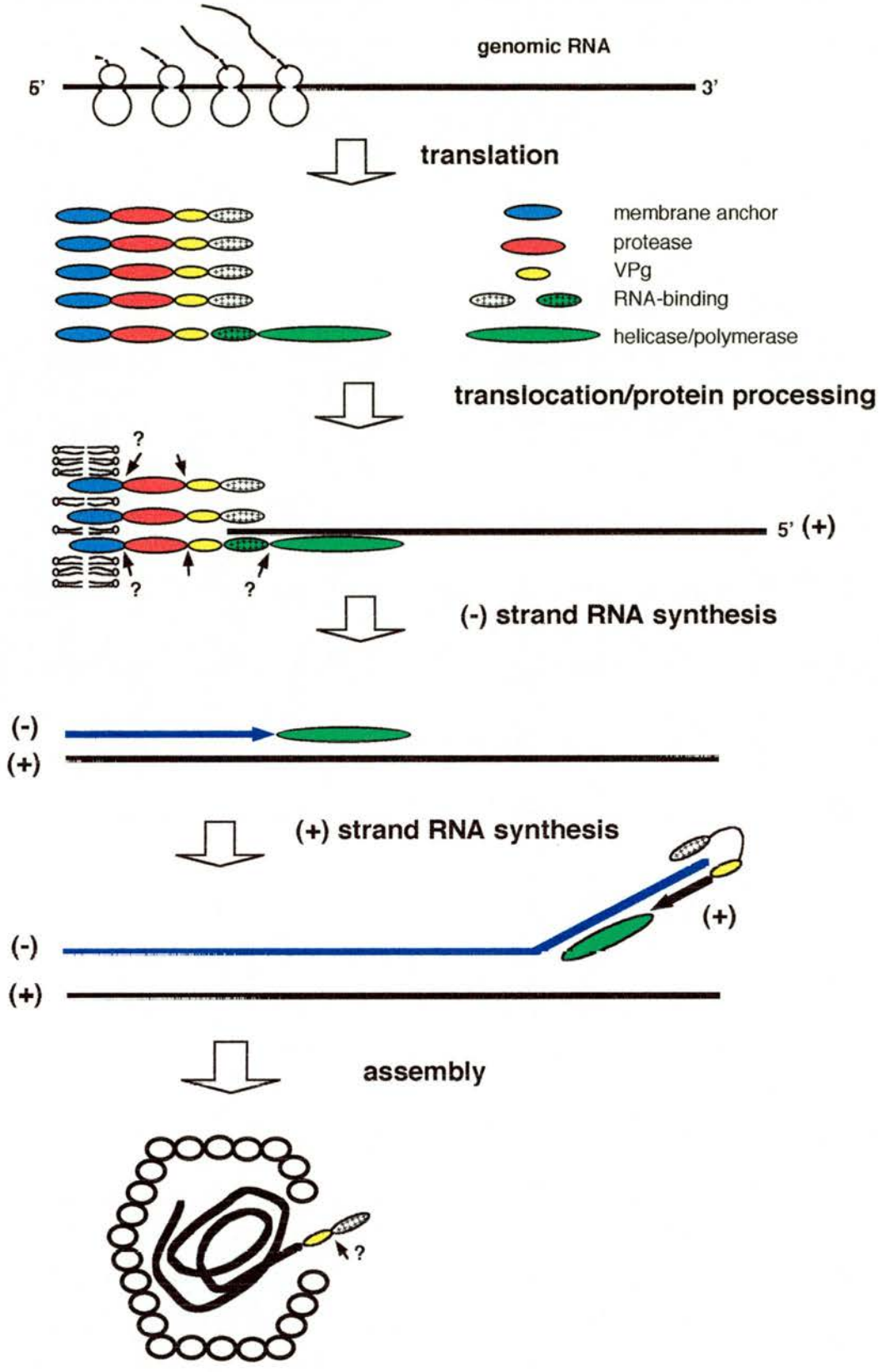


Figure 4-3 A model of P1 and P1/2 expression during the replication of PLRV

4.4 Future prospects

Many questions about the protein processing and P1 protease remained to be answered. The baculovirus, *E. coli* and cell-free systems can be further exploited; in addition, it is also possible to employ protoplast transfection and agroinfection to investigate the effects of protein processing on virus replication.

Only the middle one-third of the P1 protein is necessary for the proteolytic activity. The work on *E. coli* expression shows that it is possible to express an active form of P1 protease at high levels, so the next step would be to purify it by conventional chromatography. If this could be achieved, it will facilitate a range of biochemical and enzymological experiments. Purified protease can be used to investigate the processing of the P1 and P1/2 proteins, the inhibitor profile and the effects of cellular factors from plant tissue. The possible substrate of the protease can also be expressed and purified from *E. coli*. In this regard, the GST fusion protein, P1SP, would be a good candidate. Biochemical properties and enzymological parameters can be determined. A mini-precursor containing the protease and downstream VPg domains is also possible to be synthesized in *E. coli* and used to study the self-processing. Important features of the processing, e.g. whether the protease acts *in cis* or *in trans*, can be determined.

Currently, none of the three cleavage sites we predicted have been confirmed by experimentation. If the purified protease and corresponding substrates are available, the N-termini of the cleavage products can be sequenced, or the product can be analyzed by mass spectrography. It is also possible to determine the cleavage sites in an indirect way. Point mutations can be introduced into the proposed sites, and processing in baculovirus system can be analyzed by Western blotting. Other tags, like 6-histidine,

can also be inserted into the N- or C- termini of P1 and P1/2, to help trace the cleavage products. The cascade of the processing at different sites could be investigated in baculovirus system, or, protoplasts.

P1 and P1/2 proteins are essential for the replication of PLRV. Their processing has a central role in generating mature proteins for RNA synthesis. Mutant versions of P1 and P1/2 can be introduced into PLRV genome to investigate the effects of processing on replication in transfected protoplasts or agro-infected plants. In our model, VPg is required for the synthesis of (+)-strand genomic RNA, but might not be necessary for the synthesis of minus strand RNA. It would be of interest to delete a large part of VPg, but retain its N-terminal cleavage site and downstream frameshifting signal, to determine if minus or plus strand RNA is synthesized.

References

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

Alperin, E.S. and L.J. Shapiro. 1997. Characterization of point mutations in patients with X-linked ichthyosis. Effects on the structure and function of the steroid sulfatase protein. *J Biol Chem 272*:20756.

Andino, R., G.E. Rieckhof, D. Trono, and D. Baltimore. 1990. Substitutions in the protease (3Cpro) gene of poliovirus can suppress a mutation in the 5' noncoding region. *J Virol 64*:607.

Andino, R., G.E. Rieckhof, and D. Baltimore. 1990. A functional ribonucleoprotein complex forms around the 5' end of poliovirus RNA. *Cell 63*:369.

Andino, R., G.E. Rieckhof, P.L. Achacoso, and D. Baltimore. 1993. Poliovirus RNA synthesis utilizes an RNP complex formed around the 5'- end of viral RNA. *EMBO J 12*:3587.

Argos, P., G. Kamer, M.J. Nicklin, and E. Wimmer. 1984. Similarity in gene organization and homology between proteins of animal picornaviruses and a plant comovirus suggest common ancestry of these virus families. *Nucleic Acids Res 12*:7251.

Arias, C.F., F. Preugschat, and J.H. Strauss. 1993. Dengue 2 virus NS2B and NS3 form a stable complex that can cleave NS3 within the helicase domain. *Virology 193*:888.

Ashoub, A., W. Rohde, and D. Prüfer. 1998. In planta *trans*cription of a second subgenomic RNA increases the complexity of the subgroup 2 luteovirus genome. *Nucleic Acids Research 26*:420.

Ausebel, F.M., R. Brent, R.E. Kingston, *et al*. 1994. *Current Protocols in Molecular Biology*. John Wiley & Sons, New York.

Bahner, I., J. Lamb, M.A. Mayo, and R.T. Hay. 1990. Expression of the genome of potato leafroll virus: readthrough of the coat protein termination codon *in vivo*. *J Gen Virol 71*:2251.

Barrett, A.J. 1994. Classification of Peptidases. In *Methods in Enzymology. Volume 244 Proteolytic Enzymes: Serine and Cysteine Peptidases*. A.J. Barrett, ed. Academic Press, Inc., London,

Barrett, A.J. and J.K. McDonald. 1980. Mammalian Proteinases: A Glossary and Bibliography. Vol 1. Endopeptideases. Academic Press, Inc., London,

Barrett, A.J. 1986. Chapter 1 An Introduction to the Proteinases. In *Proteinase Inhibitors*. A.J. Barrett and G. Salvesen, eds. Elsevier Science Publishers BV (Biomedical Division), Amsterdam, p. 3.

Bartenschlager, R., L. Ahlborn-Laake, J. Mous, and H. Jacobsen. 1993. Non-structural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. *J Virol 67*:3835.

Bartenschlager, R., V. Lohmann, T. Wilkinson, and J.O. Koch. 1995. Complex formation between the NS3 serine-type proteinase of the hepatitis C virus and NS4A and its importance for polyprotein maturation. *J Virol 69*:7519.

Bazan, J.F. and R.J. Fletterick. 1988. Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. *Proc Natl Acad Sci U S A 85*:7872.

Bazan, J.F. and R.J. Fletterick. 1989. Detection of a trypsin-like serine protease domain in flaviviruses and pestiviruses. *Virology 171*:637.

Bergmann, E.M., S.C. Mosimann, M.M. Chernaia, B.A. Malcolm, and M.N. James. 1997. The refined crystal structure of the 3C gene product from hepatitis A virus: specific proteinase activity and RNA recognition. *J Virol 71*:2436.

Bernstein, H.D., N. Sonenberg, and D. Baltimore. 1985. Poliovirus mutant that does not selectively inhibit host cell protein synthesis. *Mol Cell Biol 5*:2913.

Bjerrum, O.J. and C. Schafer-Nielson. 1986. *Analytical Electrophoresis*. M.J. Dunn, ed. Verlag Chemie, Weinheim, p. 315.

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

Brault, V. and W.A. Miller. 1992. Translational frameshifting mediated by a viral sequence in plant cells. *Proc Natl Acad Sci U S A 89*:2262.

Brault, V., J.F. van den Heuvel, M. Verbeek, V. Ziegler-Graff, A. Reutenauer, E. Herrbach, J.C. Garaud, H. Guilley, K. Richards, and G. Jonard. 1995. Aphid transmission of beet western yellows luteovirus requires the minor capsid read-through protein P74. *EMBO J 14*:650.

Brown, C.M., S.P. Dinesh-Kumar, and W.A. Miller. 1996. Local and distant sequences are required for efficient readthrough of the barley yellow dwarf virus PAV coat protein gene stop codon. *J Virol 70*:5884.

Brown, R.E., K.L. Jarvis, and K.J. Hyland. 1989. Protein measurement using bicinchoninic acid: elimination of interfering substances. *Anal Biochem 180*:136.

Bruyere, A., V. Brault, V. ZieglerGraff, M.T. Simonis, J. van den Heuvel, K. Richards, H. Guilley, G. Jonard, and E. Herrbach. 1997. Effects of mutations in the beet western yellows virus readthrough protein on its expression and packaging and on virus accumulation, symptoms, and aphid transmission . *Virology 230*:323.

Cahour, A., B. Falgout, and C.J. Lai. 1992. Cleavage of the dengue virus polyprotein at the NS3/NS4A and NS4B/NS5 junctions is mediated by viral protease NS2B-NS3, whereas NS4A/NS4B may be processed by a cellular protease. *J Virol 66*:1535.

Cameron, C.E., B. Grinde, P. Jacques, J. Jentoft, J. Leis, A. Wlodawer, and I.T. Weber. 1993. Comparison of the substrate-binding pockets of the Rous sarcoma virus and human immunodeficiency virus type 1 proteases. *J Biol Chem 268*:11711.

Cao, X., I.E. Bergmann, R. Fullkrug, and E. Beck. 1995. Functional analysis of the two alternative translation initiation sites of foot-and-mouth disease virus. *J Virol 69*:560.

Carrington, J.C., S.M. Cary, and W.G. Dougherty. 1988. Mutational analysis of tobacco etch virus polyprotein processing: *cis* and *trans* proteolytic activities of polyproteins containing the 49- kilodalton proteinase. *J Virol 62*:2313.

Carrington, J.C., D.D. Freed, and T.C. Sanders. 1989. Autocatalytic processing of the potyvirus helper component proteinase in Escherichia coli and in vitro. *J Virol 63*:4459.

Carrington, J.C. and K.L. Herndon. 1992. Characterization of the potyviral HC-pro autoproteolytic cleavage site. *Virology 187*:308.

Chambers, T.J., R.C. Weir, A. Grakoui, D.W. McCourt, J.F. Bazan, R.J. Fletterick, and C.M. Rice. 1990. Evidence that the N-terminal domain of non-structural protein NS3 from yellow fever virus is a serine protease responsible for site-specific cleavages in the viral polyprotein. *Proc Natl Acad Sci U S A 87*:8898.

Chambers, T.J., A. Grakoui, and C.M. Rice. 1991. Processing of the yellow fever virus non-structural polyprotein: a catalytically active NS3 proteinase domain and NS2B are required for cleavages at dibasic sites. *J Virol 65*:6042.

Chambers, T.J., A. Nestorowicz, S.M. Amberg, and C.M. Rice. 1993. Mutagenesis of the yellow fever virus NS2B protein: effects on proteolytic processing, NS2B-NS3 complex formation, and viral replication. *J Virol 67*:6797.

Cheah, K.C., L.E. Leong, and A.G. Porter. 1990. Site-directed mutagenesis suggests close functional relationship between a human rhinovirus 3C cysteine protease and cellular trypsin- like serine proteases. *J Biol Chem 265*:7180.

Choi, H.K., L. Tong, W. Minor, P. Dumas, U. Boege, M.G. Rossmann, and G. Wengler. 1991. Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. *Nature 354*:37.

Clark, M.E., P.M. Lieberman, A.J. Berk, and A. Dasgupta. 1993. Direct cleavage of human TATA-binding protein by poliovirus protease 3C in vivo and in vitro. *Mol Cell Biol 13*:1232.

Co, E., G. Koelsch, Y. Lin, E. Ido, J.A. Hartsuck, and J. Tang. 1994. Proteolytic processing mechanisms of a miniprecursor of the aspartic protease of human immunodeficiency virus type 1. *Biochemistry 33*:1248.

Coffin, J.M. 1996. Chapter 58 *Retroviridae*: The Viruses and Their Replication. In *Fields Virology*. 3rd ed. B.N. Fields, D.M. Knipe and P.M. Howley, eds. Lippincott-Raven, Philadelphia, p. 1767.

De Groot, R.J., W.R. Hardy, Y. Shirako, and J.H. Strauss . 1990. Cleavage-site preferences of Sindbis virus polyproteins containing the non-structural proteinase. Evidence for temporal regulation of polyprotein processing in vivo. *EMBO J 9*:2631.

De, I., S.G. Sawicki, and D.L. Sawicki. 1996. Sindbis virus RNA-negative mutants that fail to convert from minus- strand to plus-strand synthesis: role of the nsP2 protein. *J Virol 70*:2706.

Debouck, C., J.G. Gorniak, J.E. Strickler, T.D. Meek, B.W. Metcalf, and M. Rosenberg. 1987. Human immunodeficiency virus protease expressed in Escherichia coli exhibits autoprocessing and specific maturation of the gag precursor. *Proc Natl Acad Sci U S A 84*:8903.

Demler, S.A. and G.A. de Zoeten. 1991. The nucleotide sequence and luteovirus-like nature of RNA 1 of an aphid non-transmissible strain of pea enation mosaic virus. *J Gen Virol 72*:1819.

Demler, S.A., O.N. Borkhsenious, D.G. Rucker, and G.A. de Zoeten. 1994. Assessment of the autonomy of replicative and structural functions encoded by the luteo-phase of pea enation mosaic virus. *J Gen Virol 75*:997.

Demler, S.A., D.G. Rucker-Feeney, J.S. Skaf, and G.A. de Zoeten. 1997. Expression and suppression of circulative aphid transmission in pea enation mosaic virus. *J Gen Virol 78*:511.

Dessens, J.T. and G.P. Lomonossoff. 1991. Mutational analysis of the putative catalytic triad of the cowpea mosaic virus 24K protease. *Virology 184*:738.

Di, R., S.P. Dinesh-Kumar, and W.A. Miller. 1993. Translational frameshifting by barley yellow dwarf virus RNA (PAV serotype) in Escherichia coli and in eukaryotic cell-free extracts. *Mol Plant Microbe Interact 6*:444.

Dinesh-Kumar, S.P. and W.A. Miller. 1993. Control of start codon choice on a plant viral RNA encoding overlapping genes. *Plant Cell 5*:679.

Ding, J., W.J. McGrath, R.M. Sweet, and W.F. Mangel. 1996. Crystal structure of the human adenovirus proteinase with its 11 amino acid cofactor. *EMBO J 15*:1778.

Dorssers, L., S. Vanderkrol, J. Vandermeer, A. Vankammen, and P. Zabel. 1984. Purification of Cowpea Mosaic-Virus RNA Replication Complex - Identification of a Virus-Encoded 110,000-Dalton Polypeptide Responsible for RNA Chain Elongation. *Proc Natl Acad Sci U S A 81*:1951.

Dougherty, W.G., T.D. Parks, S.M. Cary, J.F. Bazan, and R.J. Fletterick. 1989. Characterization of the catalytic residues of the tobacco etch virus 49- kDa proteinase. *Virology 172*:302.

Dougherty, W.G., S.M. Cary, and T.D. Parks. 1989. Molecular genetic analysis of a plant virus polyprotein cleavage site: a model. *Virology 171*:356.

Dougherty, W.G. and T.D. Parks. 1991. Post-translational processing of the tobacco etch virus 49-kDa small nuclear inclusion polyprotein: identification of an internal cleavage site and delimitation of VPg and proteinase domains. *Virology 183*:449.

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

Eckart, M.R., M. Selby, F. Masiarz, C. Lee, K. Berger, K. Crawford, C. Kuo, G. Kuo, M. Houghton, and Q.L. Choo. 1993. The hepatitis C virus encodes a serine protease involved in processing of the putative non-structural proteins from the viral polyprotein precursor. *Biochem Biophys Res Commun 192*:399.

Failla, C., L. Tomei, and R. De Francesco. 1994. Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus non-structural proteins. *J Virol 68*:3753.

Failla, C., L. Tomei, and R. De Francesco. 1995. An amino-terminal domain of the hepatitis C virus NS3 protease is essential for interaction with NS4A. *J Virol 69*:1769.

Falgout, B., M. Pethel, Y.M. Zhang, and C.J. Lai. 1991. Both non-structural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus non-structural proteins. *J Virol 65*:2467.

Falgout, B., R.H. Miller, and C.J. Lai. 1993. Deletion analysis of dengue virus type 4 non-structural protein NS2B: identification of a domain required for NS2B-NS3 protease activity. *J Virol 67*:2034.

Fenner, F. 1976. The Classification and Nomenclature of Viruses. Second Report of the International Committee on Taxonomy of Viruses. *Intervirology 7*:1.

Flanegan, J.B. and D. Baltimore. 1977. Poliovirus-specific primer-dependent RNA polymerase able to copy poly(A). *Proc Natl Acad Sci U S A 74*:3677.

Franssen, H., J. Leunissen, R. Goldbach, G. Lomonossoff, and D. Zimmern. 1984. Homologous Sequences in Non-Structural Proteins from Cowpea Mosaic-Virus and Picornaviruses. *EMBO Journal 3*:855.

Freed, E.O. 1998. HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology 251*:1.

Garcia, A., J. van Duin, and C.W. Pleij. 1993. Differential response to frameshift signals in eukaryotic and prokaryotic translational systems. *Nucleic Acids Res 21*:401.

Ghabrial, S.A., H.A. Smith, T.D. Parks, and W.G. Dougherty. 1990. Molecular genetic analyses of the soybean mosaic virus NIa proteinase. *J Gen Virol 71*:1921.

Gildow, F.E. and S.M. Gray. 1993. The Aphid Salivary-Gland Basal Lamina As a Selective Barrier Associated With Vector-Specific Transmission of Barley Yellow Dwarf Luteoviruses. *Phytopathology 83*:1293.

Gildow, F.E. 1993. Evidence For Receptor-Mediated Endocytosis Regulating Luteovirus Acquisition By Aphids. *Phytopathology 83*:270.

Gilliland, G.L., E.L. Winborne, J. Nachman, and A. Wlodawer. 1990. The three-dimensional structure of recombinant bovine chymosin at 2.3 A resolution. *Proteins 8* :82.

Gorbalenya, A.E., V.M. Blinov, and A.P. Donchenko. 1986. Poliovirus-encoded proteinase 3C: a possible evolutionary link between cellular serine and cysteine proteinase families [published erratum appears in FEBS Lett 1986 May 26;201(1):179]. *FEBS Lett 194*:253.

Gorbalenya, A.E., E.V. Koonin, V.M. Blinov, and A.P. Donchenko. 1988. Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses. *FEBS Lett 236*:287.

Gorbalenya, A.E., A.P. Donchenko, V.M. Blinov, and E.V. Koonin. 1989. Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold . *FEBS Lett 243*:103.

Gorbalenya, A.E. and E.V. Koonin. 1989. Viral proteins containing the purine NTP-binding sequence pattern. *Nucleic Acids Res 17*:8413.

Gorbalenya, A.E., A.P. Donchenko, E.V. Koonin, and V.M. Blinov. 1989. N-terminal domains of putative helicases of flavi- and pestiviruses may be serine proteases. *Nucleic Acids Res 17*:3889.

Gorbalenya, A.E., E.V. Koonin, and M.M. Lai. 1991. Putative papain-related thiol proteases of positive-strand RNA viruses. Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, alpha- and coronaviruses. *FEBS Lett 288*:201.

Grakoui, A., D.W. McCourt, C. Wychowski, S.M. Feinstone, and C.M. Rice. 1993. A second hepatitis C virus-encoded proteinase. *Proc Natl Acad Sci U S A 90*:10583.

Grakoui, A., D.W. McCourt, C. Wychowski, S.M. Feinstone, and C.M. Rice. 1993. Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. *J Virol 67*:2832.

Greiser-Wilke, I., L. Haas, K. Dittmar, B. Liess, and V. Moennig. 1993. RNA insertions and gene duplications in the non-structural protein p125 region of pestivirus strains and isolates in vitro and in vivo. *Virology 193*:977.

Grinde, B., C.E. Cameron, J. Leis, I.T. Weber, A. Wlodawer, H. Burstein, and A.M. Skalka. 1992. Analysis of substrate interactions of the Rous sarcoma virus wild type and mutant proteases and human immunodeficiency virus-1 protease using a set of systematically altered peptide substrates. *J Biol Chem 267*:9491.

Grubman, M.J., B.H. Robertson, D.O. Morgan, D.M. Moore, and D. Dowbenko. 1984. Biochemical map of polypeptides specified by foot-and-mouth disease virus. *J Virol 50*:579.

Grubman, M.J., M. Zellner, G. Bablanian, P.W. Mason, and M.E. Piccone. 1995. Identification of the active-site residues of the 3C proteinase of foot- and-mouth disease virus. *Virology 213*:581.

Guarne, A., J. Tormo, R. Kirchweger, D. Pfistermueller, I. Fita, and T. Skern. 1998. Structure of the foot-and-mouth disease virus leader protease: a papain-like fold adapted for self-processing and eIF4G recognition. *EMBO J 17*:7469.

Guilley, H., C. Wipf-Scheibel, K. Richards, H. Lecoq, and G. Jonard. 1994. Nucleotide sequence of cucurbit aphid-borne yellows luteovirus. *Virology 202*:1012.

Habili, N. and R.H. Symons. 1989. Evolutionary relationship between luteoviruses and other RNA plant viruses based on sequence motifs in their putative RNA polymerases and nucleic acid helicases. *Nucleic Acids Res 17*:9543.

Haghighat, A., Y. Svitkin, I. Novoa, E. Kuechler, T. Skern, and N. Sonenberg. 1996. The eIF4G-eIF4E complex is the target for direct cleavage by the rhinovirus 2A proteinase. *J Virol 70*:8444.

Hahn, C.S., E.G. Strauss, and J.H. Strauss. 1985. Sequence analysis of three Sindbis virus mutants temperature-sensitive in the capsid protein autoprotease. *Proc Natl Acad Sci U S A 82*:4648.

Hahn, C.S. and J.H. Strauss. 1990. Site-directed mutagenesis of the proposed catalytic amino acids of the Sindbis virus capsid protein autoprotease. *J Virol 64*:3069.

Hambidge, S.J. and P. Sarnow. 1992. Translational enhancement of the poliovirus 5' noncoding region mediated by virus-encoded polypeptide 2A. *Proc Natl Acad Sci U S A 89*:10272.

Hammerle, T., C.U. Hellen, and E. Wimmer. 1991. Site-directed mutagenesis of the putative catalytic triad of poliovirus 3C proteinase. *J Biol Chem 266*:5412.

Hanecak, R., B.L. Semler, C.W. Anderson, and E. Wimmer. 1982. Proteolytic processing of poliovirus polypeptides: antibodies to polypeptide P3-7c inhibit cleavage at glutamine-glycine pairs. *Proc Natl Acad Sci U S A 79*:3973.

Hanke, T., P. Szawlowski, and R.E. Randall. 1992. Construction of solid matrix-antibody-antigen complexes containing simian immunodeficiency virus p27 using tag-specific monoclonal antibody and tag-linked antigen. *J Gen Virol 73*:653.

Hanke, T., D.F. Young, C. Doyle, I. Jones, and R.E. Randall. 1995. Attachment of an oligopeptide epitope to the C-terminus of recombinant SIV gp160 facilitates the construction of SMAA complexes while preserving CD4 binding. *J Virol Methods 53*:149.

Hardy, W.R. and J.H. Strauss. 1988. Processing the non-structural polyproteins of Sindbis virus: study of the kinetics in vivo by using monospecific antibodies. *J Virol 62*:998.

Hardy, W.R. and J.H. Strauss. 1989. Processing the non-structural polyproteins of Sindbis virus: non-structural proteinase is in the C-terminal half of nsP2 and functions both in *cis* and in trans. *J Virol 63*:4653.

Harris, T.J., F. Brown, and D.V. Sangar. 1981. Differential precipitation of foot and mouth disease virus proteins made in vivo and in vitro by hyperimmune and virus particle guinea pig antisera. *Virology 112*:91.

Harrison, B. D. 1984. Potato leafroll virus. CMI/AAB Descriptions of Plant Viruses. No. 291.

Hartley, B.S. 1960. Proteolytic Enzymes. *Annu.Rev.Biochem. 29*:45.

Hellen, C.U., C.K. Lee, and E. Wimmer. 1992. Determinants of substrate recognition by poliovirus 2A proteinase. *J Virol 66*:3330.

Hijikata, M., H. Mizushima, T. Akagi, S. Mori, N. Kakiuchi, N. Kato, T. Tanaka, K. Kimura, and K. Shimotohno. 1993. Two distinct proteinase activities required for the processing of a putative non-structural precursor protein of hepatitis C virus. *J Virol 67*:4665.

Hogenhout, S.A., F. vanderWilk, M. Verbeek, R.W. Goldbach, and J. van den Heuvel. 1998. Potato leafroll virus binds to the equatorial domain of the aphid endosymbiotic GroEL homolog. *J Virol 72*:358.

Huang, M.M., N. Arnheim, and M.F. Goodman. 1992. Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR. *Nucleic Acids Res 20*:4567.

Hull, R. 1989. The Movement of Viruses in Plants. *Annual Review of Phytopathology 27*:213.

Iatrou, K., R.G. Meidinger, and M.R. Goldsmith. 1989. Recombinant baculoviruses as vectors for identifying proteins encoded by intron-containing members of complex multigene families. *Proc Natl Acad Sci U S A 86*:9129.

Ivanoff, L.A., T. Towatari, J. Ray, B.D. Korant, and S.R. Petteway, Jr. 1986. Expression and site-specific mutagenesis of the poliovirus 3C protease in Escherichia coli. *Proc Natl Acad Sci U S A 83*:5392.

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

Jackson, R.J. 1989. An unusual coordinated cleavage event in the processing of encephalomyocarditis virus polypeptides. *Virology 172*:363.

Jacobson, M.F. and D. Baltimore. 1968. Polypeptide cleavages in the formation of poliovirus proteins. *Proc Natl Acad Sci U S A 61*:77.

Jia, X.Y., D.F. Summers, and E. Ehrenfeld. 1993. Primary cleavage of the HAV capsid protein precursor in the middle of the proposed 2A coding region. *Virology 193*:515.

Joachims, M., K.S. Harris, and D. Etchison. 1995. Poliovirus protease 3C mediates cleavage of microtubule-associated protein 4. *Virology 211*:451.

Jolly, C.A. and M.A. Mayo. 1994. Changes in the amino acid sequence of the coat protein readthrough domain of potato leafroll luteovirus affect the formation of an epitope and aphid transmission. *Virology 201*:182.

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

Kamphuis, I.G., K.H. Kalk, M.B. Swarte, and J. Drenth. 1984. Structure of papain refined at 1.65 A resolution. *J Mol Biol 179*:233.

Karacostas, V., E.J. Wolffe, K. Nagashima, M.A. Gonda, and B. Moss. 1993. Overexpression of the HIV-1 gag-pol polyprotein results in intracellular activation of HIV-1 protease and inhibition of assembly and budding of virus-like particles. *Virology 193*:661.

Kay, J. and B.M. Dunn. 1990. Viral proteinases: weakness in strength. *Biochim Biophys Acta 1048*:1.

Kean, K.M., N.L. Teterina, D. Marc, and M. Girard. 1991. Analysis of putative active site residues of the poliovirus 3C protease. *Virology 181*:609.

Kean, K.M., M.T. Howell, S. Grunert, M. Girard, and R.J. Jackson. 1993. Substitution mutations at the putative catalytic triad of the poliovirus 3C protease have differential effects on cleavage at different sites. *Virology 194*:360.

Keese, P., R.R. Martin, L.M. Kawchuk, P.M. Waterhouse, and W.L. Gerlach. 1990. Nucleotide sequences of an Australian and a Canadian isolate of potato leafroll luteovirus and their relationships with two European isolates. *J Gen Virol 71*:719.

Kelly, L., W.L. Gerlach, and P.M. Waterhouse. 1994. Characterisation of the subgenomic RNAs of an Australian isolate of barley yellow dwarf luteovirus. *Virology 202*:565.

Kester, W.R. and B.W. Matthews. 1977. Comparison of the structures of carboxypeptidase A and thermolysin. *J Biol Chem 252*:7704.

Kim, D.W., Y. Gwack, J.H. Han, and J. Choe. 1997. Towards defining a minimal functional domain for NTPase and RNA helicase activities of the hepatitis C virus NS3 protein. *Virus Res 49*:17.

Kim, J.L., K.A. Morgenstern, C. Lin, T. Fox, M.D. Dwyer, J.A. Landro, S.P. Chambers, W. Markland, C.A. Lepre, E.T. O'Malley, S.L. Harbeson, C.M. Rice, M.A. Murcko, P.R. Caron, and J.A. Thomson. 1996. Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide [published erratum appears in Cell 1997 Apr 4;89(1):159]. *Cell 87*:343.

King, L.A. and R.D. Possee. 1992. *The baculovirus expression system: a laboratory guide*. Chapman & Hall, New York,

Kirchweger, R., E. Ziegler, B.J. Lamphear, D. Waters, H.D. Liebig, W. Sommergruber, F. Sobrino, C. Hohenadl, D. Blaas, R.E. Rhoads, and *et al.* 1994. Foot-and-mouth disease virus leader proteinase: purification of the Lb form and determination of its cleavage site on eIF-4 gamma. *J Virol 68*:5677.

Konig, H. and B. Rosenwirth. 1988. Purification and partial characterization of poliovirus protease 2A by means of a functional assay. *J Virol 62*:1243.

Koonin, E.V. and V.V. Dolja. 1993. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences [published erratum appears in Crit Rev Biochem Mol Biol 1993;28(6):546]. *Crit Rev Biochem Mol Biol 28*:375.

Kozak, M. 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res 12*:857.

Kozak, M. 1989. The scanning model for translation: an update. *J Cell Biol 108*:229.

Krausslich, H.G., M.J. Nicklin, H. Toyoda, D. Etchison, and E. Wimmer. 1987. Poliovirus proteinase 2A induces cleavage of eucaryotic initiation factor 4F polypeptide p220. *J Virol 61*:2711.

Krausslich, H.G. and E. Wimmer. 1988. Viral proteinases. *Annu Rev Biochem 57*:701.

Krausslich, H.G. 1992. Specific inhibitor of human immunodeficiency virus proteinase prevents the cytotoxic effects of a single-chain proteinase dimer and restores particle formation. *J Virol 66*:567.

Kujawa, A.B., G. Drugeon, D. Hulanicka, and A.L. Haenni. 1993. Structural requirements for efficient translational frameshifting in the synthesis of the putative viral RNA-dependent RNA polymerase of potato leafroll virus. *Nucleic Acids Res 21* :2165.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature 227*:680.

Lamb, J.W., G.H. Duncan, B. Reavy, F.E. Gildow, M.A. Mayo, and R.T. Hay. 1996. Assembly of virus-like particles in insect cells infected with a baculovirus containing a modified coat protein gene of potato leafroll luteovirus. *J Gen Virol 77*:1349.

Lamphear, B.J., R. Yan, F. Yang, D. Waters, H.D. Liebig, H. Klump, E. Kuechler, T. Skern, and R.E. Rhoads. 1993. Mapping the cleavage site in protein synthesis initiation factor eIF-4 gamma of the 2A proteases from human Coxsackievirus and rhinovirus. *J Biol Chem 268*:19200.

Landro, J.A., S.A. Raybuck, Y.P.C. Luong, E.T. O'Malley, S.L. Harbeson, K.A. Morgenstern, G. Rao, and D.J. Livingston. 1997. Mechanistic role of an NS4A peptide cofactor with the truncated NS3 protease of hepatitis C virus: elucidation of the NS4A stimulatory effect via kinetic analysis and inhibitor mapping. *Biochemistry 36*:9340.

Lawson, M.A. and B.L. Semler. 1990. Picornavirus protein processing--enzymes, substrates, and genetic regulation. *Curr Top Microbiol Immunol 161*:49.

Lawson, M.A. and B.L. Semler. 1991. Poliovirus thiol proteinase 3C can utilize a serine nucleophile within the putative catalytic triad. *Proc Natl Acad Sci U S A 88*:9919.

Lawson, M.A. and B.L. Semler. 1992. Alternate poliovirus non-structural protein processing cascades generated by primary sites of 3C proteinase cleavage. *Virology 191*:309.

Lee, C.K. and E. Wimmer. 1988. Proteolytic processing of poliovirus polyprotein: elimination of 2Apro- mediated, alternative cleavage of polypeptide 3CD by in vitro mutagenesis. *Virology 166*:405.

Lee, W.M., S.S. Monroe, and R.R. Rueckert. 1993. Role of maturation cleavage in infectivity of picornaviruses: activation of an infectosome. *J Virol 67*:2110.

Lemm, J.A. and C.M. Rice. 1993. Roles of non-structural polyproteins and cleavage products in regulating Sindbis virus RNA replication and transcription. *J Virol 67* :1916.

Lemm, J.A. and C.M. Rice. 1993. Assembly of functional Sindbis virus RNA replication complexes: requirement for coexpression of P123 and P34. *J Virol 67*:1905.

Lemm, J.A., T. Rumenapf, E.G. Strauss, J.H. Strauss, and C.M. Rice. 1994. Polypeptide requirements for assembly of functional Sindbis virus replication

complexes: a model for the temporal regulation of minus- and plus-strand RNA synthesis. *EMBO J 13*:2925.

Leong, L.E., P.A. Walker, and A.G. Porter. 1993. Human rhinovirus-14 protease 3C (3Cpro) binds specifically to the 5'- noncoding region of the viral RNA. Evidence that 3Cpro has different domains for the RNA binding and proteolytic activities. *J Biol Chem 268*:25735.

Liebig, H.D., E. Ziegler, R. Yan, K. Hartmuth, H. Klump, H. Kowalski, D. Blaas, W. Sommergruber, L. Frasel, B. Lamphear, and *et al.* 1993. Purification of two picornaviral 2A proteinases: interaction with eIF-4 gamma and influence on in vitro translation. *Biochemistry 32*:7581.

Lin, C. and C.M. Rice. 1995. The hepatitis C virus NS3 serine proteinase and NS4A cofactor: establishment of a cell-free trans-processing assay. *Proc Natl Acad Sci U S A 92*:7622.

Lin, C., J.A. Thomson, and C.M. Rice. 1995. A central region in the hepatitis C virus NS4A protein allows formation of an active NS3-NS4A serine proteinase complex in vivo and in vitro. *J Virol 69*:4373.

Lloyd, R.E., H. Toyoda, D. Etchison, E. Wimmer, and E. Ehrenfeld. 1986. Cleavage of the cap binding protein complex polypeptide p220 is not effected by the second poliovirus protease 2A. *Virology 150*:299.

Lloyd, R.E., M.J. Grubman, and E. Ehrenfeld. 1988. Relationship of p220 cleavage during picornavirus infection to 2A proteinase sequencing. *J Virol 62*:4216.

Loeb, D.D., R. Swanstrom, L. Everitt, M. Manchester, S.E. Stamper, and C.A.d. Hutchison. 1989. Complete mutagenesis of the HIV-1 protease. *Nature 340*:397.

Loeb, D.D., C.A.d. Hutchison, M.H. Edgell, W.G. Farmerie, and R. Swanstrom. 1989. Mutational analysis of human immunodeficiency virus type 1 protease suggests functional homology with aspartic proteinases. *J Virol 63*:111.

Long, A.C., D.C. Orr, J.M. Cameron, B.M. Dunn, and J. Kay. 1989. A consensus sequence for substrate hydrolysis by rhinovirus 3C proteinase. *FEBS Lett 258*:75.

Louis, J.M., R.A. McDonald, N.T. Nashed, E.M. Wondrak, D.M. Jerina, S. Oroszlan, and P.T. Mora. 1991. Autoprocessing of the HIV-1 protease using purified wild-type and mutated fusion proteins expressed at high levels in Escherichia coli. *Eur J Biochem 199*:361.

Louis, J.M., N.T. Nashed, K.D. Parris, A.R. Kimmel, and D.M. Jerina. 1994. Kinetics and mechanism of autoprocessing of human immunodeficiency virus

type 1 protease from an analog of the Gag-Pol polyprotein. *Proc Natl Acad Sci U S A 91*:7970.

Love, R.A., H.E. Parge, J.A. Wickersham, Z. Hostomsky, N. Habuka, E.W. Moomaw, T. Adachi, and Z. Hostomska. 1996. The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell 87* :331.

Luciw, P.A. 1996. Chapter 60 Human Immunodeficiency Viruses and Their Replication. In *Fields Virology*. 3rd ed. B.N. Fields, D.M. Knipe and P.M. Howley, eds. Lippincott-Raven, Philadelphia, p. 1881.

Margis, R. and L. Pinck. 1992. Effects of site-directed mutagenesis on the presumed catalytic triad and substrate-binding pocket of grapevine fanleaf nepovirus 24-kDa proteinase. *Virology 190*:884.

Margis, R., C. Ritzenthaler, J. Reinbolt, M. Pinck, and L. Pinck. 1993. Genome organization of grapevine fanleaf nepovirus RNA2 deduced from the 122K polyprotein P2 in vitro cleavage products. *J Gen Virol 74*:1919.

Marsh, L.E., T.W. Dreher, and T.C. Hall. 1988. Mutational analysis of the core and modulator sequences of the BMV RNA3 subgenomic promoter. *Nucleic Acids Res 16*:981.

Massalski, P.R. and B.D. Harrison. 1987. Properties of Monoclonal-Antibodies to Potato Leafroll Luteovirus and Their Use to Distinguish Virus Isolates Differing in Aphid Transmissibility. *J Gen Virol 68*:1813.

Matthews, B.W., P.B. Sigler, R. Henderson, and D.M. Blow . 1967. Three-dimensional structure of tosyl-alpha-chymotrypsin. *Nature 214*:652.

Matthews, D.A., W.W. Smith, R.A. Ferre, B. Condon, G. Budahazi, W. Sisson, J.E. Villafranca, C.A. Janson, H.E. McElroy, C.L. Gribskov, and *et al.* 1994. Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell 77*:761.

Mattila, P., J. Korpela, T. Tenkanen, and K. Pitkanen. 1991. Fidelity of DNA synthesis by the Thermococcus litoralis DNA polymerase-- an extremely heat stable enzyme with proofreading activity. *Nucleic Acids Res 19*:4967.

Mayo, M.A., H. Barker, D.J. Robinson, T. Tamada, and B.D. Harrison. 1982. Evidence That Potato Leafroll Virus-RNA Is Positive-Stranded, Is Linked to a Small Protein and Does Not Contain Polyadenylate. *J Gen Virol 59*:163.

Mayo, M.A., D.J. Robinson, C.A. Jolly, and L. Hyman. 1989. Nucleotide sequence of potato leafroll luteovirus RNA. *J Gen Virol 70*:1037.

Mayo, M.A. and V. ZieglerGraff. 1996. Molecular biology of luteoviruses. *Advances in Virus Research 46*:413.

McLean, C., T.J. Matthews, and R.R. Rueckert. 1976. Evidence of ambiguous processing and selective degradation in the noncapsid proteins of rhinovirus 1A. *J Virol 19*:903.

McQuade, T.J., A.G. Tomasselli, L. Liu, V. Karacostas, B. Moss, T.K. Sawyer, R.L. Heinrikson, and W.G. Tarpley. 1990. A synthetic HIV-1 protease inhibitor with antiviral activity arrests HIV-like particle maturation. *Science 247*:454.

Medina, M., E. Domingo, J.K. Brangwyn, and G.J. Belsham. 1993. The two species of the foot-and-mouth disease virus leader protein, expressed individually, exhibit the same activities. *Virology 194*:355.

Meek, T.D., D.M. Lambert, G.B. Dreyer, T.J. Carr, T.A. Tomaszek, Jr., M.L. Moore, J.E. Strickler, C. Debouck, L.J. Hyland, T.J. Matthews, and *et al.* 1990. Inhibition of HIV-1 protease in infected T-lymphocytes by synthetic peptide analogues. *Nature 343*:90.

Melancon, P. and H. Garoff. 1987. Processing of the Semliki Forest virus structural polyprotein: role of the capsid protease. *J Virol 61*:1301.

Meyers, G., T. Rumenapf, and H.J. Thiel. 1989. Ubiquitin in a togavirus. *Nature 341*:491.

Meyers, G., N. Tautz, E.J. Dubovi, and H.J. Thiel. 1991. Viral cytopathogenicity correlated with integration of ubiquitin-coding sequences. *Virology 180*:602.

Meyers, G., N. Tautz, R. Stark, J. Brownlie, E.J. Dubovi, M.S. Collett, and H.J. Thiel. 1992. Rearrangement of viral sequences in cytopathogenic pestiviruses. *Virology 191*:368.

Meyers, G. and H.J. Thiel. 1996. Molecular characterization of pestiviruses. *Adv Virus Res 47*:53.

Miller, J.S. and M.A. Mayo. 1991. The location of the 5' end of the potato leafroll luteovirus subgenomic coat protein mRNA. *J Gen Virol 72*:2633.

Miller, R.H. and R.H. Purcell. 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.

Miller, W.A., T.W. Dreher, and T.C. Hall. 1985. Synthesis of brome mosaic virus subgenomic RNA in vitro by internal initiation on (-)-sense genomic RNA. *Nature 313* :68.

Miller, W.A., P.M. Waterhouse, A.A. Kortt, and W.L. Gerlach. 1988. Sequence and identification of the barley yellow dwarf virus coat protein gene. *Virology 165*:306.

Miller, W.A., P.M. Waterhouse, and W.L. Gerlach. 1988. Sequence and organization of barley yellow dwarf virus genomic RNA. *Nucleic Acids Res 16*:6097.

Miller, W.A., S.P. Dineshkumar, and C.P. Paul. 1995. Luteovirus Gene-Expression. *Critical Reviews in Plant Sciences 14*:179.

Mizrahi, V., G.M. Lazarus, L.M. Miles, C.A. Meyers, and C. Debouck. 1989. Recombinant HIV-1 reverse transcriptase: purification, primary structure, and polymerase/ribonuclease H activities. *Arch Biochem Biophys 273*:347.

Mous, J., E.P. Heimer, and S.F. Le Grice. 1988. Processing protease and reverse transcriptase from human immunodeficiency virus type I polyprotein in Escherichia coli. *J Virol 62*:1433.

Murphy, J.F., C.J. Darcy, and J.M. Clark. 1989. Barley Yellow Dwarf Virus-RNA Has a 5'-Terminal Genome-Linked Protein. *J Gen Virol 70*:2253.

Nicklin, M.J., H.G. Krausslich, H. Toyoda, J.J. Dunn, and E. Wimmer. 1987. Poliovirus polypeptide precursors: expression in vitro and processing by exogenous 3C and 2A proteinases. *Proc Natl Acad Sci U S A 84*:4002.

O'Reilly, D.R., L.K. Miller, and V.A. Luckow. 1992. *Baculovirus Expression Vectors: A Laboratory Manual*. W. H. Freeman, New York,

Oh, C.S. and J.C. Carrington. 1989. Identification of essential residues in potyvirus proteinase HC-Pro by site-directed mutagenesis. *Virology 173*:692.

Pallansch, M.A., O.M. Kew, A.C. Palmenberg, F. Golini, E. Wimmer, and R.R. Rueckert. 1980. Picornaviral VPg sequences are contained in the replicase precursor. *J Virol 35*:414.

Parkin, N.T., M. Chamorro, and H.E. Varmus. 1992. Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo. *J Virol 66*:5147.

Parks, T.D. and W.G. Dougherty. 1991. Substrate recognition by the NIa proteinase of two potyviruses involves multiple domains: characterization using genetically engineered hybrid proteinase molecules. *Virology 182*:17.

Pearl, L.H. and W.R. Taylor. 1987. A structural model for the retroviral proteases. *Nature 329*:351.

Pelham, H.R. 1979. Synthesis and proteolytic processing of cowpea mosaic virus proteins in reticulocyte lysates. *Virology 96*:463.

Peng, X.X. and D.S. Shih. 1984. Proteolytic processing of the proteins translated from the bottom component RNA of cowpea mosaic virus. The primary and secondary cleavage reactions. *J Biol Chem 259*:3197.

Peters, D. 1967. The purification of potato leafroll virus from its vector Myzus persicae. *Virology 31*:46.

Peters, S.A., W.G. Voorhorst, J. Wery, J. Wellink, and A. van Kammen. 1992. A regulatory role for the 32K protein in proteolytic processing of cowpea mosaic virus polyproteins. *Virology 191*:81.

Pinck, M., J. Reinbolt, A.M. Loudes, M. Le Ret, and L. Pinck. 1991. Primary structure and location of the genome-linked protein (VPg) of grapevine fanleaf nepovirus. *FEBS Lett 284*:117.

Poole, T.L., C. Wang, R.A. Popp, L.N. Potgieter, A. Siddiqui, and M.S. Collett. 1995. Pestivirus translation initiation occurs by internal ribosome entry. *Virology 206*:750.

Power, M.D., P.A. Marx, M.L. Bryant, M.B. Gardner, P.J. Barr, and P.A. Luciw. 1986. Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. *Science 231*:1567.

Prüfer, D., E. Tacke, J. Schmitz, B. Kull, A. Kaufmann, and W. Rohde. 1992. Ribosomal frameshifting in plants: a novel signal directs the -1 frameshift in the synthesis of the putative viral replicase of potato leafroll luteovirus. *EMBO J 11*:1111.

Prüfer, D., L. Kawchuk, M. Monecke, S. Nowok, R. Fischer, and W. Rohde. 1999. Immunological analysis of potato leaf roll luteovirus (PLRV) P1 expression identifies a 25 kDa RNA-binding protein derived via P1 processing. *Nucleic Acids Research 27*:421.

Qiu, X., J.S. Culp, A.G. DiLella, B. Hellmig, S.S. Hoog, C.A. Janson, W.W. Smith, and S.S. Abdel-Meguid. 1996. Unique fold and active site in cytomegalovirus protease. *Nature 383*:275.

Rawlings, N.D. and A.J. Barrett. 1993. Evolutionary families of peptidases. *Biochem J 290*:205.

Rawlings, N.D. and A.J. Barrett. 1994. Families of Serine Peptidases. In *Methods in Enzymology. Volume 244 Proteolytic Enzymes: Serine and Cysteine Peptidases*. A.J. Barrett, ed. Academic Press, Inc., London, p. 19.

Rawlings, N.D. and A.J. Barrett. 1994. Families of Cysteine Peptidases. In *Methods in Enzymology. Volume 244 Proteolytic Enzymes: Serine and Cysteine Peptidases*. A.J. Barrett, ed. Academic Press, Inc., London, p. 461.

Rawlings, N.D. and A.J. Barrett. 1995. Families of Aspartic Peptidases, and Those of Unknown Catalytic Mechanism. In *Methods in Enzymology. Volume 248 Proteolytic Enzymes: Aspartic and Metallo Peptidases*. A.J. Barrett, ed. Academic Press, Inc., London, p. 105.

Reed, K.E., A. Grakoui, and C.M. Rice. 1995. Hepatitis C virus-encoded NS2-3 protease: cleavage-site mutagenesis and requirements for bimolecular cleavage. *J Virol 69*:4127.

Rees, D.C., M. Lewis, and W.N. Lipscomb. 1983. Refined crystal structure of carboxypeptidase A at 1.54 A resolution. *J Mol Biol 168*:367.

Restrepo-Hartwig, M.A. and J.C. Carrington. 1992. Regulation of nuclear transport of a plant potyvirus protein by autoproteolysis. *J Virol 66*:5662.

Reutenauer, A., V. Ziegler-Graff, H. Lot, D. Scheidecker, H. Guilley, K. Richards, and G. Jonard. 1993. Identification of beet western yellows luteovirus genes implicated in viral replication and particle morphogenesis. *Virology 195*:692.

Rice, C.M. 1996. Chapter 30 *Flaviviridae*: The Viruses and Their Replication. In *Fields Virology*. 3rd ed. B.N. Fields, D.M. Knipe and P.M. Howley, eds. Lippincott-Raven, Philadelphia, p. 931.

Riechmann, J.L., S. Lain, and J.A. Garcia. 1992. Highlights and Prospects of Potyvirus Molecular-Biology. *J Gen Virol 73*:1.

Roberts, N.A., J.A. Martin, D. Kinchington, A.V. Broadhurst, J.C. Craig, I.B. Duncan, S.A. Galpin, B.K. Handa, J. Kay, A. Krohn, and *et al.* 1990. Rational design of peptide-based HIV proteinase inhibitors. *Science 248*:358.

Rueckert, R.R. 1996. Chapter 21 *Picornaviridae*: The Viruses and Their Replication. In *Fields Virology*. 3rd ed. B.N. Fields, D.M. Knipe and P.M. Howley, eds. Lippincott-Raven, Philadelphia, p. 609.

Ryan, M.D. and M. Flint. 1997. Virus-encoded proteinases of the picornavirus super-group. *J Gen Virol 78*:699.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. In 2nd ed. Cold Spring Harbor press, Cold Spring Harbor, NY,

Sangar, D.V., R.P. Clark, A.R. Carroll, D.J. Rowlands, and B.E. Clarke. 1988. Modification of the leader protein (Lb) of foot-and-mouth disease virus. *J Gen Virol 69*:2327.

Sawicki, D.L. and S.G. Sawicki. 1993. A second non-structural protein functions in the regulation of alphavirus negative-strand RNA synthesis. *J Virol 67*:3605.

Schechter, I. and A. Berger. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun 27*:157.

Schultheiss, T., Y.Y. Kusov, and V. Gauss-Muller. 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.

Seemuller, E., A. Lupas, D. Stock, J. Lowe, R. Huber, and W. Baumeister. 1995. Proteasome from Thermoplasma acidophilum: a threonine protease. *Science 268*:579.

Seipelt, J., A. Guarne, E. Bergmann, M. James, W. Sommergruber, I. Fita, and T. Skern. 1999. The structures of picornaviral proteinases. *Virus Res 62*:159.

Serghini, M.A., M. Fuchs, M. Pinck, J. Reinbolt, B. Walter, and L. Pinck. 1990. RNA2 of grapevine fanleaf virus: sequence analysis and coat protein cistron location. *J Gen Virol 71*:1433.

Shih, D.S., B. Ming, M.A. Price, and C.Y.T. Shih. 1987. Inhibition of Cleavage of a Plant Viral Polyprotein by an Inhibitor Activity Present in Wheat-Germ and Cowpea Embryos. *J Virol 61*:912.

Shimizu, Y., K. Yamaji, Y. Masuho, T. Yokota, H. Inoue, K. Sudo, S. Satoh, and K. Shimotohno. 1996. Identification of the sequence on NS4A required for enhanced cleavage of the NS5A/5B site by hepatitis C virus NS3 protease. *J Virol 70*:127.

Skern, T., W. Sommergruber, H. Auer, P. Volkmann, M. Zorn, H.D. Liebig, F. Fessl, D. Blaas, and E. Kuechler. 1991. Substrate requirements of a human rhinoviral 2A proteinase. *Virology 181*:46.

Snijder, E.J., A.L. Wassenaar, L.C. van Dinten, W.J. Spaan, and A.E. Gorbalenya. 1996. The arterivirus nsp4 protease is the prototype of a novel group of chymotrypsin-like enzymes, the 3C-like serine proteases. *J Biol Chem 271*:4864.

Sommergruber, W., M. Zorn, D. Blaas, F. Fessl, P. Volkmann, I. Maurer-Fogy, P. Pallai, V. Merluzzi, M. Matteo, T. Skern, and *et al.* 1989. Polypeptide 2A of human rhinovirus type 2: identification as a protease and characterization by mutational analysis. *Virology 169*:68.

Sommergruber, W., H. Ahorn, A. Zophel, I. Maurer-Fogy, F. Fessl, G. Schnorrenberg, H.D. Liebig, D. Blaas, E. Kuechler, and T. Skern. 1992. Cleavage specificity on synthetic peptide substrates of human rhinovirus 2 proteinase 2A. *J Biol Chem 267*:22639.

Sommergruber, W., G. Casari, F. Fessl, J. Seipelt, and T. Skern. 1994. The 2A proteinase of human rhinovirus is a zinc containing enzyme. *Virology 204*:815.

Sommergruber, W., J. Seipelt, F. Fessl, T. Skern, H.D. Liebig, and G. Casari. 1997. Mutational analyses support a model for the HRV2 2A proteinase. *Virology 234*:203.

Steinkuhler, C., A. Urbani, L. Tomei, G. Biasiol, M. Sardana, E. Bianchi, A. Pessi, and R. De Francesco. 1996. Activity of purified hepatitis C virus protease NS3 on peptide substrates. *J Virol 70*:6694.

Strauss, E.G., R.J. De Groot, R. Levinson, and J.H. Strauss. 1992. Identification of the active site residues in the nsP2 proteinase of Sindbis virus. *Virology 191*:932.

Strauss, J.H. and E.G. Strauss. 1994. The alphaviruses: gene expression, replication, and evolution [published erratum appears in Microbiol Rev 1994 Dec;58(4):806]. *Microbiol Rev 58*:491.

Strebel, K. and E. Beck. 1986. A second protease of foot-and-mouth disease virus. *J Virol 58*:893.

Strebel, K., E. Beck, K. Strohmaier, and H. Schaller. 1986. Characterization of foot-and-mouth disease virus gene products with antisera against bacterially synthesized fusion proteins. *J Virol 57*:983.

Summers, D.F. and J.V. Maizel, Jr. 1968. Evidence for large precursor proteins in poliovirus synthesis. *Proc Natl Acad Sci U S A 59*:966.

Tacke, E., D. Prüfer, F. Salamini, and W. Rohde. 1990. Characterization of a potato leafroll luteovirus subgenomic RNA: differential expression by internal translation initiation and UAG suppression. *J Gen Virol 71*:2265.

Tacke, E., D. Prüfer, J. Schmitz, and W. Rohde. 1991. The potato leafroll luteovirus 17K protein is a single-stranded nucleic acid-binding protein. *J Gen Virol 72*:2035.

Tacke, E., J. Schmitz, D. Prüfer, and W. Rohde. 1993. Mutational analysis of the nucleic acid-binding 17 kDa phosphoprotein of potato leafroll luteovirus identifies an amphipathic alpha-helix as the domain for protein/protein interactions. *Virology 197*:274.

Takkinen, K., J. Peranen, S. Keranen, H. Soderlund, and L. Kaariainen. 1990. The Semliki-Forest-virus-specific non-structural protein nsP4 is an autoproteinase. *Eur J Biochem 189*:33.

ten Dam, E., M. Flint, and M.D. Ryan. 1999. Virus-encoded proteinases of the *Togaviridae*. *J Gen Virol 80*:1879.

ten Dam, E.B., C.W. Pleij, and L. Bosch. 1990. RNA pseudoknots: translational frameshifting and readthrough on viral RNAs. *Virus Genes 4*:121.

Tian, Y.C. and D.S. Shih. 1986. Cleavage of a viral polyprotein by a cellular proteolytic activity. *J Virol 57*:547.

Tomaszek, T.A., Jr., M.L. Moore, J.E. Strickler, R.L. Sanchez, J.S. Dixon, B.W. Metcalf, A. Hassell, G.B. Dreyer, I. Brooks, C. Debouck, and *et al.* 1992. Proteolysis of an active site peptide of lactate dehydrogenase by human immunodeficiency virus type 1 protease. *Biochemistry 31*:10153.

Tong, L., G. Wengler, and M.G. Rossmann. 1993. Refined structure of Sindbis virus core protein and comparison with other chymotrypsin-like serine proteinase structures. *J Mol Biol 230*:228.

Toyoda, H., M.J. Nicklin, M.G. Murray, C.W. Anderson, J.J. Dunn, F.W. Studier, and E. Wimmer. 1986. A second virus-encoded proteinase involved in proteolytic processing of poliovirus polyprotein. *Cell 45*:761.

Tozser, J., I. Blaha, T.D. Copeland, E.M. Wondrak, and S. Oroszlan. 1991. Comparison of the HIV-1 and HIV-2 proteinases using oligopeptide substrates representing cleavage sites in Gag and Gag-Pol polyproteins. *FEBS Lett 281*:77.

Ueng, P.P., J.R. Vincent, E.E. Kawata, C.H. Lei, R.M. Lister, and B.A. Larkins. 1992. Nucleotide sequence analysis of the genomes of the MAV-PS1 and P-PAV isolates of barley yellow dwarf virus. *J Gen Virol 73*:487.

Vakharia, V.N., M.A. Devaney, D.M. Moore, J.J. Dunn, and M.J. Grubman. 1987. Proteolytic processing of foot-and-mouth disease virus polyproteins expressed in a cell-free system from clone-derived transcripts. *J Virol 61*:3199.

van den Heuvel, J.F., M. Verbeek, and F. van der Wilk. 1994. Endosymbiotic bacteria associated with circulative transmission of potato leafroll virus by Myzus persicae. *J Gen Virol 75*:2559.

van den Heuvel, J., A. Bruyere, A. Hogenhout, V. Ziegler-Graff, V. Brault, M. Verbeek, F. van der Wilk, and K. Richards. 1997. The N-terminal region of the luteovirus readthrough domain determines virus binding to Buchnera GroEL and is essential for virus persistence in the aphid. *J Virol 71*:7258.

van der Wilk, F., M.J. Huisman, B.J. Cornelissen, H. Huttinga, and R. Goldbach. 1989. Nucleotide sequence and organization of potato leafroll virus genomic RNA. *FEBS Lett 245*:51.

van der Wilk, F., M. Verbeek, A.M. Dullemans, and J.F. van den Heuvel. 1997. The genome-linked protein of potato leafroll virus is located downstream of the putative protease domain of the ORF1 product. *Virology 234*:300.

Veidt, I., H. Lot, M. Leiser, D. Scheidecker, H. Guilley, K. Richards, and G. Jonard. 1988. Nucleotide sequence of beet western yellows virus RNA. *Nucleic Acids Res 16*:9917.

Veidt, I., S.E. Bouzoubaa, R.M. Leiser, V. Ziegler-Graff, H. Guilley, K. Richards, and G. Jonard. 1992. Synthesis of full-length transcripts of beet western yellows virus RNA: messenger properties and biological activity in protoplasts. *Virology 186*:192.

Verchot, J., E.V. Koonin, and J.C. Carrington. 1991. The 35-kDa protein from the N-terminus of the potyviral polyprotein functions as a third virus-encoded proteinase. *Virology 185*:527.

Vincent, J.R., R.M. Lister, and B.A. Larkins. 1991. Nucleotide sequence analysis and genomic organization of the NY-RPV isolate of barley yellow dwarf virus. *J Gen Virol 72*:2347.

Voss, T., R. Meyer, and W. Sommergruber. 1995. Spectroscopic characterization of rhinoviral protease 2A: Zn is essential for the structural integrity. *Protein Sci 4*:2526.

Walker, P.A., L.E. Leong, and A.G. Porter. 1995. Sequence and structural determinants of the interaction between the 5'- noncoding region of picornavirus RNA and rhinovirus protease 3C. *J Biol Chem 270*:14510.

Wang, S., L. Guo, E. Allen, and W.A. Miller. 1999. A potential mechanism for selective control of cap-independent translation by a viral RNA sequence *in cis* and in trans. *RNA 5*:728.

Wang, S.P., K.S. Browning, and W.A. Miller. 1997. A viral sequence in the 3'-untranslated region mimics a 5' cap in facilitating translation of uncapped mRNA. *EMBO Journal 16*:4107.

Wang, Y.F., S.G. Sawicki, and D.L. Sawicki. 1994. Alphavirus nsP3 functions to form replication complexes transcribing negative-strand RNA. *J Virol 68*:6466.

Warrener, P. and M.S. Collett. 1995. Pestivirus NS3 (p80) protein possesses RNA helicase activity. *J Virol 69*:1720.

Wengler, G., T. Nowak, and E. Castle. 1990. Description of a procedure which allows isolation of viral non-structural proteins from BHK vertebrate cells infected with the West Nile flavivirus in a state which allows their direct chemical characterization. *Virology 177*:795.

Wengler, G. 1991. The carboxy-terminal part of the NS 3 protein of the West Nile flavivirus can be isolated as a soluble protein after proteolytic cleavage and represents an RNA-stimulated NTPase. *Virology 184*:707.

Wiskerchen, M. and M.S. Collett. 1991. Pestivirus gene expression: protein p80 of bovine viral diarrhea virus is a proteinase involved in polyprotein processing. *Virology 184*:341.

Wu, Z., N. Yao, H.V. Le, and P.C. Weber. 1998. Mechanism of autoproteolysis at the NS2-NS3 junction of the hepatitis C virus polyprotein. *Trends Biochem Sci 23*:92.

Wyckoff, E.E., J.W. Hershey, and E. Ehrenfeld. 1990. Eukaryotic initiation factor 3 is required for poliovirus 2A protease- induced cleavage of the p220 component of eukaryotic initiation factor 4F. *Proc Natl Acad Sci U S A 87*:9529.

Yin, F.H. and N.B. Lomax. 1983. Host range mutants of human rhinovirus in which non-structural proteins are altered. *J Virol 48*:410.

Young, M.J., L. Kelly, P.J. Larkin, P.M. Waterhouse, and W.L. Gerlach. 1991. Infectious in vitro transcripts from a cloned cDNA of barley yellow dwarf virus. *Virology 180*:372.

Ypma-Wong, M.F. and B.L. Semler. 1987. In vitro molecular genetics as a tool for determining the differential cleavage specificities of the poliovirus 3C proteinase. *Nucleic Acids Res 15*:2069.

Ypma-Wong, M.F., P.G. Dewalt, V.H. Johnson, J.G. Lamb, and B.L. Semler. 1988. Protein 3CD is the major poliovirus proteinase responsible for cleavage of the P1 capsid precursor. *Virology 166*:265.