PROTEOLYTIC PROCESSING AND MEMBRANE ASSOCIATION OF TOMATO RINGSPOT NEPOVIRUS RNA-1-ENCODED PROTEINS

By

AIMING WANG

B.Sc., Jiangsu Agricultural College, Yangzhou University, Yangzhou, China, July 1983

M.Sc., Nanjing Agricultural University, Nanjing, China, June 1988

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Department of Botany, The Faculty of Science)

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

Vancouver, B.C., Canada, December 1998

©Aiming Wang, 1998
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Botany
The University of British Columbia
Vancouver, Canada

Date Jan 21, 1999
GENERAL ABSTRACT

Tomato ringspot nepovirus (TomRSV) encapsidates a bipartite, messenger sense RNA genome. RNA-1 is likely to encode all the viral proteins required for RNA replication. The expression and processing strategies of the RNA-1-encoded polyprotein (P1) were studied. It was found that P1 is intramolecularly proteolytically processed by a virus-encoded protease at five cleavage sites to release six end products, i.e. X1, X2, NTB, VPg, Pro and Pol. All P1 cleavage sites were identified. The genomic organization of TomRSV RNA-1 is distinct from that of the closely related comoviruses and of other nepoviruses. In TomRSV, two mature proteins (X1 and X2) are released from the region upstream of NTB whereas only one mature protein is released from the corresponding region in all other characterized nepo- and comoviruses.

Viral RNA replication is thought to be carried out by a replication complex which is secured on cellular membranes by a virus-encoded protein. To test if the NTB-VPg protein which contains a predicted trans-membrane domain is the anchor protein, association of this protein with microsomal membranes was studied in vitro. It was found that the NTB-VPg protein is an integral membrane protein and that the predicted transmembrane domain at the C-terminus of NTB is required for the membrane-association. The association of the NTB-VPg protein with microsomal membranes results in the generation of glycosylated and signal peptidase-processed proteins. The importance of these modifications for viral RNA replication needs to be elucidated.

The molecular variability of the RNA-1-encoded replication-related viral proteins including VPg, Pro and Pol, and of the RNA-2-encoded coat protein among five TomRSV isolates was studied. It was found that the nucleotide sequences and the deduced amino acid
sequences of RNA-1 were more conserved than those of RNA-2. Amino acid substitutions were not found in all conserved motifs identified previously in the protease and polymerase domains. Furthermore, the VPg amino acid sequence was identical in all isolates studied despite some variations at the nucleotide sequence level.

The above results are discussed in light of the possible strategies for processing of the P1 polyprotein and for the formation of the replication complex on cellular membranes.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>iv</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>vii</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>viii</td>
</tr>
<tr>
<td>Preface</td>
<td>ix</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>x</td>
</tr>
<tr>
<td>Chapter 1. <strong>Overview: Tomato Ringspot Virus</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>2</td>
</tr>
<tr>
<td>1.2 General Properties of TomRSV</td>
<td>2</td>
</tr>
<tr>
<td>1.2.1 Taxonomy and relationship</td>
<td>2</td>
</tr>
<tr>
<td>1.2.2 Transmission, geographical distribution and diseases</td>
<td>3</td>
</tr>
<tr>
<td>1.2.3 Detection and diagnosis</td>
<td>4</td>
</tr>
<tr>
<td>1.2.4 Cytopathology</td>
<td>5</td>
</tr>
<tr>
<td>1.2.5 Control</td>
<td>5</td>
</tr>
<tr>
<td>1.3 Molecular Biology of TomRSV</td>
<td>6</td>
</tr>
<tr>
<td>1.3.1 Purification and viral particles</td>
<td>6</td>
</tr>
<tr>
<td>1.3.2 Sequence and predicted genomic organizations of TomRSV</td>
<td>7</td>
</tr>
<tr>
<td>1.3.3 Characterization of the movement protein, protease domain and cleavage specificity</td>
<td>8</td>
</tr>
<tr>
<td>1.3.4 Expression of the movement and coat proteins in protoplasts</td>
<td>10</td>
</tr>
<tr>
<td>1.4 Research Objective of This Thesis</td>
<td>11</td>
</tr>
<tr>
<td>Chapter 2. <strong>Literature Review: Expression and Replication of the Picorna-like Viruses and Other Single-Stranded Positive Sense RNA Viruses</strong></td>
<td>12</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>13</td>
</tr>
<tr>
<td>2.2 Classification of ss(+) RNA Viruses</td>
<td>13</td>
</tr>
<tr>
<td>2.3 Expression of the Picorna-like Viruses and Other ss(+) RNA viruses</td>
<td>16</td>
</tr>
<tr>
<td>2.3.1 Expression strategies</td>
<td>16</td>
</tr>
<tr>
<td>2.3.2 Genomic organization of picorna-like viruses</td>
<td>17</td>
</tr>
<tr>
<td>2.3.3 Polypeptide processing in the picorna-like supergroup</td>
<td>22</td>
</tr>
<tr>
<td>2.4 Viral Proteins Associated with Viral RNA Replication</td>
<td>32</td>
</tr>
<tr>
<td>2.4.1 RNA-dependent RNA polymerase</td>
<td>32</td>
</tr>
<tr>
<td>2.4.2 NTP-binding protein, helicase and membrane binding protein</td>
<td>35</td>
</tr>
<tr>
<td>2.4.3 Methyltransferase</td>
<td>37</td>
</tr>
<tr>
<td>2.4.4 Genome-linked viral protein (VPg)</td>
<td>37</td>
</tr>
<tr>
<td>2.4.5 Protease</td>
<td>40</td>
</tr>
<tr>
<td>2.5 Cis-Acting Viral RNA Sequences Required for RNA Replication</td>
<td>42</td>
</tr>
<tr>
<td>2.5.1 3'-Terminal sequences of the genomic RNA in the picorna-like viruses</td>
<td>42</td>
</tr>
<tr>
<td>2.5.2 5'-Terminal sequences of the genomic RNA in the picorna-like viruses</td>
<td>43</td>
</tr>
</tbody>
</table>
2.6 Host Factors and Cellular Membranes Involved in Viral RNA Replication .... 44
2.6.1 Host proteins involved in viral RNA replication .......................... 44
2.6.2 Cellular membranes involved in RNA replication ......................... 46
2.7 Summary ................................................................................. 48

Chapter 3. Genomic Organization of Tomato Ringspot Nepovirus RNA-1: Proteolytic Processing of the Carboxy-Terminus of the Polyprotein .................. 50
3.1 Abstract .................................................................................. 51
3.2 Introduction ............................................................................. 51
3.3 Materials and Methods ................................................................. 53
3.3.1 Plasmid constructions ............................................................. 53
3.3.2 Induction in E. coli, cell lysis and purification of fusion proteins ...... 57
3.3.3 Chemical crosslinking of synthetic VPg peptide to ovalbumin ......... 58
3.3.4 Antisera production ................................................................. 59
3.3.5 SDS-PAGE, gel-staining and western blots ............................... 60
3.3.6 Microsequencing of the N-terminal protease ............................... 60
3.3.7 In vitro transcription and translation and quantitative measurements .. 61
3.4 Results ...................................................................................... 61
3.4.1 Production of monoclonal and polyclonal antibodies against viral proteins .................................................................................. 61
3.4.2 Proteolytic processing of protease precursors by the TomRSV protease in E. coli and in vitro ......................................................... 62
3.4.3 Temperature dependence of proteolytic processing in E coli and in vitro ...................................................................................... 68
3.4.4 Presence of VPg domain on precursors increases cleavage efficiency at Pro-Pol cleavage site .............................................................. 71
3.4.5 Expression and processing of a complex protease precursor in E. coli .. 74
3.4.6 Mutation of the cleavage sites at both sides of the putative VPg and the protease domains ........................................................................ 76
3.5 Discussion .................................................................................. 79

Chapter 4. Genomic Organization of Tomato Ringspot Nepovirus RNA-1: Proteolytic Processing of the Amino Acid-Terminus of the Polyprotein and Detection of P1 Proteins in Plants .................. 84
4.1 Abstract .................................................................................... 85
4.2 Introduction .............................................................................. 85
4.3 Materials and Methods ................................................................. 86
4.3.1 Plasmid construction ............................................................... 86
4.3.2 Antisera ................................................................................ 90
4.3.3 Viral RNA purification and microsequencing of the VPg protein .... 90
4.3.4 Purification of the recombinant active protease .......................... 91
4.3.5 In vitro transcription and translation and quantitative measurements .. 91
4.3.6 SDS-PAGE and western immunoblotting ................................. 91
4.4 Results ...................................................................................... 91
4.4.1 Expression and processing of the N-terminal portion of P1 .......... 91
4.4.2 P1 cleavage sites are processed intramolecularly ........................................... 98
4.4.3 Detection of P1 mature and precursor proteins in infected plants .................. 100
4.4.4 Microsequencing of VPg ................................................................. 103
4.5 Discussion ......................................................................................... 104

Chapter 5. Membrane-Associated Glycosylation and Signal Peptidase Processing of the Tomato Ringspot Virus NTB-VPg Protein in vitro ................................................. 109
5.1 Abstract ............................................................................................ 110
5.2 Introduction ....................................................................................... 110
5.3 Materials and Methods ....................................................................... 112
5.3.1 Plasmid construction and site-directed mutagenesis .............................. 112
5.3.2 In vitro transcription and translation .................................................... 113
5.3.3 Deglycosylation assay .................................................................... 114
5.3.4 Direct membrane binding assay ........................................................ 114
5.3.5 Biochemical treatment of the membrane fraction .................................. 114
5.4 Results ............................................................................................... 115
5.4.1 The NTB-VPg protein associates with microsomal membranes in vitro ...... 115
5.4.2 The VPg sequence and a predicted transmembrane domain at the C-terminus of NTB are required for the membrane-induced modifications ............... 117
5.4.3 The NTB-VPg precursor is an integral membrane protein .................... 119
5.4.4 The glycosylated protein is deglycosylated by N-glycosidase F ............. 121
5.4.5 The glycosylation consensus sequence on VPg is required for the glycosylation and the signal peptidase processing probably takes place at the junction of the NTB and VPg proteins .................................................. 121
5.5 Discussion ........................................................................................ 126

Chapter 6. Molecular Variability in the Carboxy-Terminal Regions of Polyproteins Encoded by Tomato Ringspot Nepovirus Isolates .................................................. 131
6.1 Abstract ............................................................................................ 131
6.2 Introduction ....................................................................................... 132
6.3 Materials and Methods ....................................................................... 133
6.3.1 Source of virus isolates .................................................................... 133
6.3.2 Virus propagation and viral RNA extraction ....................................... 134
6.3.3 cDNA synthesis, PCR amplification and DNA purification .................. 134
6.3.4 Sequencing and sequence analysis ..................................................... 136
6.4 Results ............................................................................................... 136
6.4.1 Host plant responses ....................................................................... 136
6.4.2 Detection of TomRSV by RT-PCR ..................................................... 138
6.4.3 Comparison of the RNA-1-encoded VPg, Pro and Pol coding regions of different TomRSV isolates .......................................................... 140
6.4.4 Comparison of the RNA-2 CP gene of different TomRSV isolates ....... 143
6.5 Discussion ........................................................................................ 146

Chapter 7. Summary, General Discussion and Future Prospects .......................... 149
7.1 Summary and General Discussion ............................................................ 150
LIST OF TABLES

Table 3.1 Primers used for construction of plasmids .................................. 53
Table 4.1 Primers used for construction of plasmids and site-directed mutagenesis .......... 87
Table 5.1 Primers used for construction of plasmids ........................................ 112
Table 6.1 Primers used for cDNA synthesis, PCR amplification and DNA sequencing .... 135
Table 6.2 Percent identities among the nucleotide and amino acid sequences of the VPg-Pro-Pol precursor, of VPg, of Pro and of Pol of TomRSV isolates ......................... 140
Table 6.3 Percent identities among the nucleotide and amino acid sequences of the coat protein of TomRSV isolates ...................................................... 144
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Predicted genomic organization of TomRSV</td>
</tr>
<tr>
<td>2.1</td>
<td>Similarity and variability in genomic organization of the picorna-like superfamily</td>
</tr>
<tr>
<td>2.2</td>
<td>Schematic representation of the expression strategy of the CPMV genome</td>
</tr>
<tr>
<td>3.1</td>
<td>Schematic diagram of different constructs derived from TomRSV RNA-1 cDNA</td>
</tr>
<tr>
<td>3.2</td>
<td>Production of monoclonal and polyclonal antibodies against fusion proteins</td>
</tr>
<tr>
<td>3.3</td>
<td>Processing of the VPG-Pro-N-Pol precursor polypeptide in <em>E. coli</em></td>
</tr>
<tr>
<td>3.4</td>
<td>Processing of the NTB-VPG-Pro precursor polypeptide and mutation of the NTB-VPG cleavage site <em>in vitro</em></td>
</tr>
<tr>
<td>3.5</td>
<td>Temperature dependence of VPG-Pro-N-Pol and Pro-N-Pol auto-cleavage in <em>E. coli</em></td>
</tr>
<tr>
<td>3.6</td>
<td>Temperature dependence of the proteolytic activity of the protease on P1 proteins <em>in vitro</em></td>
</tr>
<tr>
<td>3.7</td>
<td>Presence of VPG is critical to the protease activity on the Pro-Pol cleavage site <em>in vitro</em></td>
</tr>
<tr>
<td>3.8</td>
<td>Processing of the Thio-NTB-VPG-Pro-Pol precursor polypeptide in <em>E. coli</em></td>
</tr>
<tr>
<td>3.9</td>
<td>Mutation of the predicted cleavage sites between Pro and Pol</td>
</tr>
<tr>
<td>4.1</td>
<td>Schematic diagram of the different constructs derived from TomRSV RNA-1 cDNA, which are used for analyses on the processing strategy at the N-terminus of P1</td>
</tr>
<tr>
<td>4.2</td>
<td>Processing of the X2-NTB-VPG-Pro precursor polypeptide <em>in vitro</em></td>
</tr>
<tr>
<td>4.3</td>
<td>Processing of the X2.1-X2.2-NTB-VPG-Pro precursor polypeptide <em>in vitro</em></td>
</tr>
<tr>
<td>4.4</td>
<td>Processing of the X1-X2-NTB-VPG-Pro precursor polypeptide <em>in vitro</em></td>
</tr>
<tr>
<td>4.5</td>
<td>The protease does not process P1 proteins <em>in trans in vitro</em></td>
</tr>
<tr>
<td>4.6</td>
<td>Immunoblot analysis of TomRSV proteins in infected plants</td>
</tr>
<tr>
<td>4.7</td>
<td>A model for processing of the RNA-1-encoded polyprotein P1</td>
</tr>
<tr>
<td>5.1</td>
<td>The NTP-VPG precursor undergoes modification in the presence of microsomal membranes <em>in vitro</em></td>
</tr>
<tr>
<td>5.2</td>
<td>The C-terminus of NTB and the VPG sequence are required for the membrane-induced modification</td>
</tr>
<tr>
<td>5.3</td>
<td>Direct membrane binding assay</td>
</tr>
<tr>
<td>5.4</td>
<td>Biochemical treatment of membrane-associated proteins</td>
</tr>
<tr>
<td>5.5</td>
<td>Deglycosylation assay</td>
</tr>
<tr>
<td>5.6</td>
<td>Glycosylation and signal peptidase cleavage of a NTB-VPG precursor (C-NTB-VPG)</td>
</tr>
<tr>
<td>5.7</td>
<td>Time-course and pulse-chase analyses of glycosylation and signal peptidase cleavage of the C-NTB-VPG precursor</td>
</tr>
<tr>
<td>6.1</td>
<td>Symptoms caused in cucumber plants by TomRSV isolates</td>
</tr>
<tr>
<td>6.2</td>
<td>Agarose gel electrophoretic analysis of RT-PCR amplified TomRSV cDNA</td>
</tr>
<tr>
<td>6.3</td>
<td>Multiple amino acid alignments of VPG, Pro and Pol of TomRSV isolates</td>
</tr>
<tr>
<td>6.4</td>
<td>Multiple amino acid alignments of the coat protein of TomRSV isolates</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATION

3'  three prime
5'  five prime
A   adenosine in the context of nucleotide sequence
    alanine in the context of amino acid sequence
AIMV alfalfa mosaic virus
ArMV arabis mosaic virus
ATP adenosine-5'-triphosphate
AUG triplet codon as a start codon and/or for amino acid methionine
1A   poliovirus 1A coat protein
2A   poliovirus 2A protease
3A   poliovirus 3A protein which contains a putative membrane binding domain
3AB  poliovirus 3AB protein, precursor of 3A and 3B
B    bottom component in the context of virus particle sediment
BMV brome mosaic virus
bp   base pairs
BSA bovine serum albumin
BRL Bethesda Research Laboratories
1B   poliovirus 1B coat protein
2B   poliovirus 2B protein
2BC  poliovirus 2BC protein
3B   poliovirus viral genome-linked protein (VPg)
C    cytidine in the context of nucleotide sequence
cysteine in the context of amino acid sequence
ca. approximately
CaMV 35S cauliflower mosaic 35S
cDNA complementary DNA
Cl   cylindrical inclusion
CLRV cherry leaf roll virus
CMV   cucumber mosaic virus
CNV   cucumber necrosis virus
como- comovirus
Co-Pro comovirus protein as a co-factor of protease
CP   coat protein
CP1  coat protein one of comoviruses
CP2  coat protein two of comoviruses
CPMV cowpea mosaic virus
C-terminal carboxy-terminal
C-termini carboxy-termini
C-terminus carboxy-terminus
°C degrees Celsius
1C poliovirus 1C coat protein
2C poliovirus 2C protein (helicase- or NTB-like protein)
3C poliovirus 3C protease
3CD poliovirus 3C protease and 3D polymerase precursor
D aspartic acid
dATP deoxyadenosine triphosphate
dCTP deoxycytidine triphosphate
dGTP deoxyguanosine triphosphate
DNA deoxyribonucleic acid
DTT dithiothreitol
dTTP deoxothymidine triphosphate
1D poliovirus 1D coat protein
3D poliovirus 3D RNA-dependent RNA polymerase
E glutamic acid
E. coli Escherichia coli
EDTA ethylenediaminetetraacetic acid
EF elongation factor
ELISA enzyme-linked immunosorbent assays
EtBr ethidium bromide
F phenylalanine
FMDV foot-and-mouth disease virus
g gram
G guanosine in the context of nucleotide sequence
G glycine in the context of amino acid sequence
GCMV grapevine chrome mosaic virus
GDD tripeptide sequence of G, D and D, a conserved amino acid sequence in RNA-dependent RNA polymerase
GFLV grapevine fanleaf virus
Gln glutamine
Glu glutamic acid
Gly glycine
H histidine
HC potyvirus HC protease
HC-Pro potyvirus HC protease
hr hour
HRV-14 human rhinovirus-14
I isoleucine
IRES internal ribosome entry site
K lysine in the context of amino acid sequence
K thousand in the context of size
kb kilobase
kDa kilodalton
6k potyvirus 6k or 6k2 protein containing a membrane binding domain located between CI and N1a-VPg
6k1 potyvirus 6k1 protein located between P3 and CI
6k2 potyvirus 6k2 protein located between CI and N1a-VPg
l litre
L leucine in the context of amino acids
L L protease in the context of proteins encoded by cardio- and aphthoviruses
LB Luria broth
μ micro
(-) minus sense
m mini or milli
M methionine in the context of amino acid sequence
M middle component in the context of virus particle sediment
M molar in the context of concentration
MES 2[N-morpholino]ethanesulphonic acid
min minute
MLRSV myrobalan latent ringspot virus
MP movement protein
mRNA messenger RNA
n nano
N asparagine
nepo- nepovirus
N1a potyvirus VPg and 3C-like protease domains
N1a-Pro potyvirus 3C-like protease domain
N1a-VPg potyvirus VPg domain
N1b potyvirus RNA-dependent polymerase domain
nt nucleotide
N-terminal amino-terminal or NH$_2$-terminal
N-termini amino-termini or NH$_2$-termini
N-terminus amino-terminus or NH$_2$-terminus
NTB nucleotide triphosphate-binding protein
NTB-VPg precursor protein of NTB and VPg
NTP nucleoside triphosphate
NTPM nucleotide triphosphate-binding motif
NXT/S conserved amino acid sequence of glycosylation (N: asparagine, X: any amino acid, T/S: threonine or serine
OLRSV olive latent ringspot virus
ORF open reading frame
(+ ) positive sense
P proline
P1 polyprotein encoded by como- and nepovirus RNA-1
P1 P1 protease in the context of proteins encoded by potyviruses
P1 region encoding 1ABCD protein in the context of picornaviruses
P2 polyprotein encoded by como- and nepovirus RNA-2
P2 region encoding 2ABC protein in the context of picornaviruses
P3 potyvirus P3 protein located between HC and 6k1
P3 region encoding 3ABCD in the context of picornaviruses
PAGE polyacrylamide gel electrophoresis
PCR polymerase chain reaction
PEG polyethylene glycol
picorna- picornaviruses
picorna-like picornavirus superfamily or supergroup
POL or Pol polymerase
poly(A) polyadenylate or polyadenylic acid
poly(U) polyuridylate or polyuridylic acid
poty- potyvirus
PPV plum pox virus
PRO or Pro protease in the context of proteins
Pro-Pol precursor of protease and polymerase
PVX potato virus X
Q glutamine
R arginine
RNA ribonucleic acid
RNP ribonucleoprotein
RRSV raspberry ringspot virus
RdRp RNA-dependent RNA polymerase
RNA ribonucleic acid
RRV raspberry ringspot virus
RT-PCR reverse-transcription polymerase chain reaction
S serine
SDS sodium dodecyl sulfate
SDS-PAGE sodium dodecyl sulfate-polyacrylamide gel electrophoresis
ss(+) single-stranded positive sense
T threonine in the context of amino acid
T thymidine in the context of nucleotide
T top component in the context of virus particle sediment
TaMV tamarillo mosaic potyvirus
TBRV tomato blackring virus
TEV tobacco etch virus
TMV tobacco mosaic virus
TobRSV tobacco ringspot virus
TomRSV tomato ringspot virus
Tris hydroxymethyl amino methane
tRNA transfer RNA
TRSV tobacco ringspot virus
TVMV  tobacco vein mottling virus
TYMV  turnip yellow mosaic virus
X1    N-terminal protein of P1 encoded by TomRSV RNA-1
X2    the protein between X1 and NTB of P1 encoded by TomRSV RNA-1
U     uridine
V     valine
VPg   viral protein genome-linked
VPg-Pro-Pol precursor protein of VPg, protease and polymerase
VS    protein encoded by vector sequence
v/v   volume over volume
W     tryptophan
Y     tyrosine
PREFACE

This thesis is submitted in partial fulfilment of the requirements for the degree of doctor of philosophy in the Faculty of Graduate Studies and Department of Botany, the Faculty of Science, the University of British Columbia (UBC). All the requirements and auditing courses were completed on UBC campus and all the experiments presented in this thesis were conducted at Dr. Hélène Sanfaçon's laboratory located at Pacific Agriculture Research Centre (PARC), Vancouver before Aug. 1996, and relocated to Pacific Agri-Food Research Center (PARC), Summerland after Aug. 1996.

Due to his commitment to agricultural sciences and molecular biology, the author chose molecular virology as his thesis project. Virus diseases are called "plant cancers" by farmers in South-eastern China, which vividly describes the damage of the diseases and the difficulty to control them. It has been well documented that viral diseases cause huge losses every year in agricultural production in the world. Resistance sources to virus diseases in nature are rare. The most common measure to control these viral diseases is the application of pesticides or nematicides to prevent virus transmission. However, these chemicals usually are too expensive and cause public concerns because of their potential threat to the environment. Recently, genetic engineering has emerged as an alternative method to reduce the loss caused by virus diseases. Although the approach is very promising, the candidate genes for transformation are limited to some viral genes (such as coat protein, replicase and protease). Moreover, concerns about the potential risks of recombination of inserted viral genes in transgenic plants with the viral genome of infected viruses have been raised in the scientific community and in the general public. To develop new strategies to engineer resistance against plant virus diseases, it is necessary to extend the current knowledge of the virus
replication cycle at the molecular level. Using tomato ringspot nepovirus (TomRSV) as a model system, the objective of this thesis is to elucidate how the viral proteins involved in replication are produced through proteolytic processing by a viral encoded protease, and to determine which of these proteins associate with cellular membranes to form the membrane-bound replication complex, two key features for RNA replication. This thesis consists of seven chapters.

Chapter 1 introduces the virus which has been studied in this thesis. The general properties and the results of previous studies on the virus are presented. The research goals of this thesis are given.

TomRSV is a positive sense, single stranded RNA virus that uses a polyprotein strategy for genome expression and replication. A literature review about expression and replication of single-stranded positive sense RNA viruses is presented in Chapter 2. Since the knowledge of the polyprotein processing of nepoviruses is very limited, this chapter also introduces progress in the expression and processing strategies of nepovirus related viruses (picornavirus superfamily).

Chapter 3 consists of experiments designed to determine if the TomRSV serine-like protease is responsible for processing of the polyprotein encoded by RNA-1 (P1). By means of western blots, time course studies, mutagenesis and protein micro-sequencing, it was found that the TomRSV protease located on P1 could process P1 precursor proteins in E. coli and in vitro, and that the protease activity on P1 precursors was temperature dependent. Since the protease prefers some cleavage sites to others, a step-wise model was proposed for the protease to process its precursors. The dipeptide sequences of the NTB-VPg, VPg-Pro and Pro-Pol cleavage sites were identified. Therefore, the VPg, Pro and Pol were precisely localized in the P1 polyprotein.

Chapter 4 attempts to determine how the protease processes P1 and how many mature proteins are located in P1. Expression and processing of the N- and C-terminal portions of P1 in vitro reveals that there is one additional protein located at the N-terminus of P1 polyprotein
compared to other nepo- and comoviruses. The sequence of the putative VPg was determined by direct microsequencing of the VPg which is linked to viral genome. Along with the results in chapter 3, all P1 cleavage sites were identified. According to these in vitro and in vivo results, the possible pathway for the protease to process the P1 polyprotein was discussed.

As mentioned earlier in this preface, TomRSV like other positive sense RNA viruses is believed to replicate its viral RNA on cellular membranes. In Chapter 5, P1 mature and intermediate precursor proteins obtained from chapter 4 were examined for their association with microsomal membranes. It was found that NTB-VPg but not other P1 proteins was glycosylated and signal peptidase-processed in the presence of microsomal membranes. A potential transmembrane domain located on the C-terminal region of NTB was responsible for the membrane binding while a potential glycosylation conserved sequence on the middle part of the VPg was determined to be the site for glycosylation.

Chapter 6 aims to compare the sequence of replication-related viral proteins including the C-terminal portion of NTB, the entire VPg, the entire protease and the entire polymerase, and a structural protein (coat protein) among five TomRSV isolates. From the sequencing results, it was found that the VPg sequence was extremely conserved, indicating that VPg may be critical for viral replication or infectivity.

The results from preceding chapters are discussed in Chapter 7. Future experiments which should be conducted to further characterize the replication mechanism of TomRSV are suggested.

Summerland, B.C. Canada
December, 1998

Aiming Wang
ACKNOWLEDGMENTS

The author would like to express his sincere appreciation to his major supervisor, Dr. Hélène Sanfaçon, for her inspiring guidance, sound advice, tremendous encouragement, financial support and indispensable enthusiastic help throughout this study. He would like to thank Drs. Carl Douglas (his co-supervisor), Dave Theilmann and Ljerka Kunst for serving in the supervisory committee and for providing valuable knowledge, generous advice and helpful discussions as well as for their critical reviews of manuscripts to the successful fulfilment of his Ph.D. program.

The author also would like to acknowledge Dr. D'Ann Rochon for her very helpful suggestions and critical review of part of this thesis. He is grateful to Drs. Dean Struble [former director of Pacific Agriculture Research Centre (PARC)], Gordon Neish (director), Thierry Vrain (head of Biotechnology Section) and many other people at PARC for their help and generous permission to use facilities and supplies; Dr. Iain Taylor (Head), Lebby Balakshin (Head Secretary), Fiona Craig (former Head Secretary) and many other staff of Department of Botany, UBC for their assistance in many aspects; Drs. Adib Rowhani at the University of California at Davis and Robert Martin at PARC for generously providing TomRSV isolates; Professors Helin Xu, Hualen Zhao and Xinhan Wang, former directors of Institute of Vegetable Crops, Jiangsu Academy of Agricultural Sciences, China for their constant support and encouragement.

The author specifically wants to thank Joan Chisholm for her assistance in obtaining lab supplies, critical reading of the whole thesis and valuable discussions and suggestions. In Chapter 3, he wants to acknowledge the following lab colleagues for their contributions and help, Joan Chisholm for a western blot showing the temperature dependence of processing of the VPg-Pro-Pol precursor in E. coli, Karma Carrier for identification of the cleavage site between the protease and the putative RNA-dependent RNA polymerase, Andrew Wieczorek for construction of clone pT7-xviii
MPCATΔATG and for cloning and purification of the middle part of the putative NTP-binding protein and the protease, Dr. Claire Huguenot for technical supervision in the production of monoclonal and polyclonal antibodies against NTB and Dr. Fabienne Hans for construction of clone pT7-VPg-Pro-N-Pol-II.

He also would like to thank Michael Weis for his excellent job in photographing and scanning, Peggy Watson for her help in literature retrieval, Anita Quail for her assistance in DNA sequencing, Jim Wild for computer assistance, Zhencai Wu for his help in DNA analysis and Wenshuang Xie, Fengchen Sun and many other great people he met at UBC and PARC for their help and friendship.

The author is much indebted to his wife Yun Zhang and their son David Jueyu Wang for their constant love and tremendous support and understanding during the past four years, which were essential for bringing this work to fruition.

This work was supported in part by a grant from Natural Sciences and Engineering Research Council to Dr. Hélène Sanfaçon. The author successfully won yearly competitions for University Graduate Fellowships (UGF) at UBC for four successive years from 1995 to 1998.
CHAPTER 1

OVERVIEW: TOMATO RINGSPOT NEPOVIRUS
1.1 Introduction

Tomato ringspot virus (TomRSV) is a nepovirus that contains a bipartite, single-stranded, positive-sense RNA genome. Nepoviruses belong to the picornavirus superfamily. TomRSV is therefore evolutionarily related to the animal picornaviruses, including poliovirus, and to the plant como- and potyviruses (Goldbach, 1987; Koonin and Dolja, 1993; Chapter 2). TomRSV causes economically important diseases in North America (see below). Since the 1960's, its general biological properties such as taxonomy, transmission methods, geographical distributions and detection approaches have been extensively studied. More recently, the entire nucleotide sequence of its genomic RNAs has been determined allowing further study of the virus at the molecular level. Studies of the virus replication cycle at the molecular level is of fundamental importance towards the development of new strategies to control the diseases caused by TomRSV and related viruses including nepo-, como- and potyviruses.

1.2 General Properties of TomRSV

1.2.1 Taxonomy and relationship

TomRSV does not infect tomato naturally, but causes severe diseases in many other economically important crops. The designation “tomato ringspot virus” was first applied to two viruses. One isolate was obtained from field tomatoes in Indiana which caused ringspot symptoms (Imle and Samson, 1937), and the second isolate, originally described as tobacco ringspot no. 2, was found in tobacco seedlings in New Jersey (Price, 1936). It was later found that these two viruses were unrelated (Tall et al., 1949). The former so called “tomato ringspot virus” is probably the same as the “tomato top necrosis” virus (Bancroft, 1968). The later “tobacco ringspot no. 2 virus” has been replaced by the vernacular name tomato ringspot virus though the virus does not occur naturally in tomato. TomRSV has been classified into the nepovirus genus and comoviridae family.
Nepoviruses have been subdivided into three subgroups on the basis of their serological relationships, and length and packaging of RNA-2 (Murant, 1981; Mayo and Robinson, 1996). TomRSV and thirteen other nepoviruses including peach rosette mosaic, cherry leaf roll (CLRV) and myrobalan latent ringspot (MLRSV) viruses are the definitive members of subgroup c, the largest subgroup of nepoviruses. Tobacco ringspot virus (TRSV) and seven other definitive nepovirus members including grapevine fanleaf virus (GFLV), arabis mosaic virus (ArMV) and raspberry ringspot virus (RRSV) belong to the subgroup a. The eight remaining definitive members of nepoviruses, including tomato black ring virus (TBRV), grapevine chrome mosaic virus (GCMV) and olive latent ringspot virus (OLRSV), constitute the subgroup b nepoviruses. Serologically TomRSV is unrelated to other members of nepoviruses. However, it is very difficult to distinguish it from other nepoviruses based only on host range and symptomatology.

1.2.2 Transmission, geographical distribution and diseases caused

TomRSV is mainly transmitted by the nematode *Xiphinema americanum* and other closely related *Xiphinema* sp. (Teliz et al., 1966; Harrison and Murant, 1977). The virus can also be spread by dodder (*Cuscuta campestris*) (Puffinberger and Corbett, 1973) or through seed and pollen. Seed transmission of TomRSV occurs in soybeans (Kahn, 1956), strawberries (Mellor and Stace-Smith, 1963), raspberries (Braun and Keplinger, 1973), *Pelargonium* (Scarborough and Smith, 1977) and dandelion (Mountain et al., 1983). The virus is widespread in dandelion and can be disseminated over considerable distances through windblown seeds (Rosenberger et al., 1983). Dandelions and other perennial weeds are not only hosts of the virus, but also provide reservoirs for the virus. Pollen transmission of TomRSV has been reported in *Pelargonium X hortorum* (Scarborough, 1974). TomRSV is also readily transmissible by mechanical inoculation of the infected leaf sap.

TomRSV has been found in infected plants in many countries including Australia, Bulgaria, Canada, Chile, China, Cyprus, Germany, Italy, Japan, Korea Republic, New Zealand, Peru, Puerto Rico, Turkey, the USA, the former USSR and the former Yugoslavia (Stace-Smith, 1984b). But
natural spread is largely confined to North America around the Great Lakes and along the Pacific coast where populations of its nematode vector Xiphinema sp. occur (Harrison and Murant, 1977; Francki et al., 1985). The virus causes serious diseases in many economically important crops, especially perennial crops. In nature, TomRSV infects raspberry (ring spot disease: Stace-Smith, 1962) (yellow blotch-curl disease: Chamberlain, 1938), strawberry (ring spot disease: Mellor and Stance-Smith, 1963), grapevine (grapevine yellow vein virus: Gooding, 1963), lily (yellow stripe: Kim and Choi, 1990), prunus (yellow bud mosaic in peach nectarine, sweet cherry, almond and plum: Schlocker and Traylor, 1976), blueberry (Converse and Ramsdell, 1982), apple (apple union necrosis and decline: Lana et al., 1983) and cucumber (McKeen, 1960). In raspberry, it is considered as the most damaging virus disease (Stace-Smith, 1984a). In strawberry, the disease is lethal in about 90% of commercial cultivars that are graft-inoculated (Mellor and Stace-Smith, 1963). The fact that the virus is prevalent in perennial weeds and crops, that the transmission methods are various and that the virus infection results in severe loss of yields has lead to the suggestion that TomRSV is one of the most damaging plant viruses in North America (Stace-Smith, 1984b). Experimentally, the virus can infect a wide variety of monocots and dicots but principally it infects dicots. It has been reported that TomRSV can infect 285 species in 159 genera of 55 families (Edwardson and Christie, 1997).

1.2.3 Detection and diagnosis

Since the host range and symptoms of TomRSV resemble those of other nepoviruses, serological assays have been suggested to be the most reliable tests for identification (Stace-Smith, 1984b). Several immuno-related methods have been reported to successfully detect the virus. However, some methods which can detect the virus in some crops do not identify the virus in other crops. For example, enzyme-linked immunosorbent assay (ELISA), the most commonly used assay, can readily detect some virus isolates/strains in some infected woody hosts and herbal weeds (Converse, 1978; Rosenberger et al., 1983) but miss some other strains (Stace-Smith, 1984a). This
may result from the high serological specificity of the double sandwich ELISA method (Stace-Smith, 1984b). Therefore, other serological tests or alternative techniques need to be developed to provide more sensitive and reliable diagnoses. In some stone fruit crops, radioimmunoassay has shown to be a preferred way for the virus detection (Hoy et al., 1981). Serologically distinguishable strains have also been successfully observed in agar gel double diffusion tests (Hoy et al., 1984).

1.2.4 Cytopathology

After infection, the virus can spread throughout all parts of infected plants including the mesophyll (Stace-Smith, 1984b; Stace-Smith, 1996). In infected cells, rows of particles are observed in tubules present in the cytoplasm near or traversing the cell wall but not in the nucleus (De Zoeten and Gaard, 1969; Murant, 1981; Piazzolla et al., 1985; Corbett and Podleckis, 1985; Podleckis and Corbett, 1987; Kim and Choi, 1990; Wieczorek and Sanfaçon, 1995; Stace-Smith, 1996). Vacuolate-vesiculate inclusions, which are similar to those induced by tobacco ringspot virus are present in infected Vicia faba and grapevine cells (Smith and Mcwhorter, 1957; Piazzolla et al., 1985). Large cytoplasmic crystalline inclusions consisting of virus particles adjacent to the plasmodesmata are found in infected plant cells as well (Gordejchuk et al., 1977). Pollen grain abortion, abnormal ovules, aborted ovules and tissue disintegration have been observed in infected florets (Murdock et al., 1976). In red raspberry, reduced drupelet set is common in some cultivars after TomRSV infection, which may involve reductions in male and female fertility (Daubeny et al., 1975).

1.2.5 Control

Usually, plant viruses are controlled by eliminating virus sources, preventing transmission from plant to plant and growing virus-resistant cultivars. However, so far no resistant cultivars have been reported for TomRSV. The continuing supply of the “seed” virus in the form of infected perennial weed seeds makes it impossible to eliminate the sources. Moreover, it is very difficult to immediately remove infected perennial plants and replant in the established plantings since the disease symptoms at early stages of development of these crops are usually not obvious. Provided
that the infected plants are accurately identified and removed, the existing virus-carrying nematode is still a threat to both the uninfected and newly replanted plants. Under such circumstances, the effective course would be that when the productivity reduction of the infected crops is severe enough, the crops are completely removed, then the soil is treated with nematicides to control the TomRSV vector, and finally, virus-free stock can be replanted (Stace-Smith, 1984b).

1.3 Molecular Biology of Tomato Ringspot Virus

1.3.1 Purification and virus particles

The virus can be purified from several different host species using a variety of procedures. The most successful virus source for purification has been reported to be infected cucumber tissues (Stace-Smith, 1966; Allen and Dias, 1977) but tobacco and petunia have also been proven to be good hosts (Goff and Corbett, 1977; Powell and Derr, 1983). It appears that the extraction buffer is not critical since any of these buffer systems, namely, 0.5 M sodium borate-boric acid buffer (pH 6.7), 0.05 M potassium citrate (pH 6.5) and 0.05 M potassium phosphate (pH 7.0) have been successfully used for the purification. Purification procedures usually includes partial clarification, particle concentration and separation by density-gradient-centrifugation. Partial clarification is achieved either by freezing buffered extracts or by freezing infected tissue before extraction. After further clarification of the extract by adding granular ammonium sulphate, the virus is concentrated by several rounds of differential centrifugation. Sucrose or cesium chloride density gradient centrifugation is used to separate virus particles and contaminating host materials.

In sucrose density gradients, the virus particles sediment as three components, designated top (T), middle (M) and bottom (B) (Schneider et al., 1974; Allen and Dias, 1977). The virus particles are isometric, about 28 nm in diameter with angular outlines, and consist of protein subunits (coat proteins). In the denatured protein gel, the coat protein migrates as one component of an apparent molecular mass of 58 kDa. The T component is the empty particle and does not encapsidate any
genomic RNA. The ratio of M to B particles is approximately 1:1. Each M or B particle harbors one RNA molecule which is single-stranded positive sense. The B component contains an RNA-1 molecule which is of molecular mass $5.5 \times 10^6$ daltons and the M component harbors an RNA-2 molecule which is of molecular mass $3.2 \times 10^6$ daltons (Stace-Smith, 1984b). Like other nepoviruses, each of the two RNA species is covalently linked with a virus-encoded protein (VPg). Treatment of the purified genomic RNA with proteinase K destroys RNA infectivity, demonstrating the importance of this small protein (Mayo et al., 1979; Mayo et al., 1982).

1.3.2 Sequence of genomic RNA and predicted genomic organization

Of the eight definitive members of nepovirus subgroup c, TomRSV is the only member whose genome has been fully sequenced (Rott et al., 1991b; 1995). The RNA-1 is 8214 nt in length, excluding a 3' poly(A) tail (32 nt) and contains a single long open reading frame (ORF) of 6591 nt beginning at the first AUG codon at nucleotide position 78. This would give rise to a polyprotein (P1) with a predicted molecular mass of 244 kDa. The RNA-2 is 7273 nt in length excluding the 3' poly(A) tail and contains a single long ORF of 5646 nt beginning at position 78 and terminating at position 5723. Therefore, the RNA-2 encodes a polyprotein (P2) of a molecular mass of 207 kDa.

Computer-assisted comparisons have revealed that there is extensive nucleotide sequence similarity between both the 5' and 3' terminal regions of RNA-1 and those of RNA-2. The 3' terminal 1533 nt of RNA-1 and RNA-2 are identical with the exception of only 3 nt positions. The 5' terminal 459 nt of both RNA molecules are identical at all positions and the following 447 are identical at 75.8% of nt positions. This extensive identity of nucleotide sequence at the 5' and 3' ends of TomRSV RNA-1 and RNA-2 is rare among the small RNA viruses, and is suggested to result from duplication of the nt sequence from one RNA species to the other via RNA recombination (Rott et al., 1991a). RNA viruses with more than one genomic component often display some nt sequence similarity in their 5' and 3' non-coding regions. These sequences, which are generally imperfectly repeated and less than 300 nt in length are thought to be an important replicase recognition site.
Therefore, the additional non-coding sequence of TomRSV RNA may have a function other than just for replicase recognition.

Besides sequence identity of long non-coding regions at the 5' and 3' termini of RNA-1 and RNA-2, the 5' termini of RNA-1 and RNA-2 also share sequence identity in the coding regions and therefore, the N-terminal polypeptides encoded by both RNA molecules have extensive amino acid sequence similarity. Beginning at the first AUG codon at nucleotide position 78 of both RNA-1 and RNA-2, the first 132 amino acids are identical followed by another 145 amino acids of which 75.3% of the positions are identical.

The deduced amino acid sequences of polyproteins from the TomRSV RNA-1 and RNA-2 cDNA sequences have been compared with those of other nepo- and como-viruses. Based on sequence homology, the genomic organization of TomRSV RNA-1 and RNA-2 has been predicted (Fig. 1.1) (Rott et al., 1991a, 1991b and 1995). The TomRSV RNA-1 encoded polyprotein (P1) contains sequence characteristic of a viral protease cofactor (Vos et al., 1988; Ritzenthaler et al., 1991), an NTP-binding protein (Gorbalenya and Koonin, 1989; Gorbalenya et al., 1989b), a VPg protein (Mayo and Frisch, 1994), a 3C-like cysteine protease (Bazan and Fletterick, 1988) and an RNA-dependent RNA-polymerase (Kamer and Argos, 1984; Argos et al., 1984; Argos, 1988). The TomRSV RNA-2 encoded polyprotein (P2) includes a movement protein, a coat protein, and protein(s) with unknown functions at its N-terminus (Rott et al., 1991b).

1.3.3 Characterization of movement protein, protease domain and cleavage specificity

To characterize the putative movement protein encoded by the TomRSV RNA-2, the predicted movement protein domain which is located upstream of the predicted coat protein domain was cloned and expressed in E. coli (Wieczorek and Sanfaçon, 1993). The bacterium-produced fusion protein was used to raise monoclonal antibodies against the putative movement protein. A viral protein with an apparent molecular mass of 45 kDa, the predicted size of the putative movement protein, was detected by the antibodies in TomRSV-infected plant extracts. In infected cells, this
Fig. 1.1. **Predicted genomic organization of TomRSV.** The boxes represent open reading frames and the single short lines show non-coding sequences. The VPG protein covalently attached to the 5’ end of the genomic RNA is shown. For clarity, coding regions are not drawn to scale. Abbreviations: ?, unknown protein(s); NTB, NTP-binding protein; VPG, viral encoded genome-linked protein; Pro, protease; Pol, RNA-dependent RNA polymerase; MP, movement protein; CP, coat protein.
protein was found to be associated with tubular structures containing virus-like particles. The tubular structures were localized near or protruding from and traversing the cell wall. These results suggest that the putative movement protein is likely to be involved in the cell-to-cell movement via the formation of these tubular structures.

Recently, the TomRSV RNA-1 encoded P1 protein region containing the predicted VPg, Pro and N-terminal portion of Pol has been cloned into an \textit{in vitro} transcription vector and into a transient expression vector (Hans and Sanfaçon, 1995). Proteins generated from these clones could cleave the P2 derived precursors containing the cleavage site between the movement protein and the coat protein \textit{in vitro} and \textit{in vivo}. Site-directed mutagenesis experiments have shown that this proteolytic activity was abolished by mutation of either the predicted catalytic triad (H1283D mutation) or the putative substrate binding pocket (H1451L mutation) of the protease domain. By direct N-terminal sequencing of the coat protein and deletion mutagenesis, the cleavage site between the movement protein and the coat protein (MP-CP) was determined to be Q/G, a typical cleavage site recognized by picornavirus 3C-like proteases. These results suggest that there is an RNA-1-encoded 3C-like protease which is responsible for processing P2 at the MP-CP cleavage site, and that the MP-CP cleavage site shares characteristics of cleavage sites of subgroup c nepoviruses and of poty- and comoviruses. However, it is not known if this 3C-like protease is also responsible for processing of cleavage sites on the P1 polyprotein.

\textbf{1.3.4 Expression of the TomRSV movement and coat proteins in protoplasts}

To further elucidate the mechanism of the cell-to-cell movement and encapsidation of TomRSV \textit{in vivo}, the accumulation of the TomRSV movement protein and coat protein was studied in protoplasts (Sanfaçon \textit{et al.}, 1995). It was shown that TomRSV movement protein was detected at late stages of infection, which is different from most movement proteins of other single-stranded positive sense RNA viruses. Although the movement protein was less stable than the coat protein in protoplasts, considerable amounts of the movement protein could be detected at 72 h post-
infection, a time at which viral protein synthesis had already decreased (Sanfaçon, et al., 1995).
These data suggest that the mechanism of the TomRSV cell-to-cell movement may be similar to that of caulimo-, como-, and other nepoviruses, which are thought to move as encapsidated particles through tubular structures composed at least in part of the movement protein.

1.4 Research Objectives of This Thesis

Although knowledge of TomRSV at the molecular level has significantly increased in the last decade, some basic features of the virus replication cycle are still far beyond our understanding. For example, it is not known yet if and how the predicted protease domain processes the P1 precursor polyprotein to release mature or intermediate precursor proteins which are required for RNA replication. In addition, it is not clear which viral protein(s) associated with cellular membranes where the replication is likely to occur.

The overall goal of this thesis is to elucidate expression and replication strategies of TomRSV RNA-1. To realize this goal, this thesis has been designed to achieve the following objectives:

(i) To determine if the predicted 3C-like protease domain, in addition to its activity on the P2 cleavage site, is also responsible for P1 polyprotein processing and to precisely locate the protease;

(ii) To establish P1 polyprotein processing strategy;

(iii) To identify the protein(s) which are associated with cellular membranes where the virus replication is thought to occur;

(iv) To study the molecular variability in the C-terminal regions of polyproteins encoded by different TomRSV isolates.
CHAPTER 2

LITERATURE REVIEW: EXPRESSION AND REPLICATION OF PICORNA-LIKE AND OTHER SINGLE-STRANDED, POSITIVE SENSE RNA VIRUSES
2.1 Introduction

The replication of the viral genome is a key step in the biology of single-stranded, positive sense [ss(+)] RNA viruses. The importance of understanding the replication mechanism is obvious. For example, it may lead to the development of new strategies to control virus diseases. It is envisaged that the replication of ss(+) RNA viruses in infected cells takes place in three successive stages. First, upon entry into host cells, the positive strand RNA genome acts as an mRNA to enable the synthesis of viral proteins which are required for the replication of the viral RNA. Then the negative strand RNA molecule is synthesized by the virus-encoded RNA-dependent RNA polymerase using the virus genome as the template. Subsequent transcription of this negative strand RNA allows the production of new progeny copies of the virus genomic RNA. The first stage is dependent on the host translation machinery while the subsequent stages are thought to be catalyzed by a complex defined as the replication complex which contains RNA-dependent RNA polymerases, other viral proteins such as helicases and membrane binding proteins, endogenous RNA templates, host factors and cellular membranous structures essential for RNA synthesis.

In this chapter, the taxonomy of ss(+) RNA viruses will be discussed, in particular the characteristics of the picornavirus-like supergroup. In addition, the expression strategies used by these viruses will be outlined, especially the polyprotein strategy which is employed by picorna-like viruses. The roles or putative functions of specific viral proteins (mature proteins or polyprotein precursors) in viral RNA replication will be presented. Finally, the possible functions of cis-acting viral RNA sequences, host proteins and cellular membranes which are also components of replication complexes will be discussed.

2.2 Classification of ss(+) RNA Viruses

Since the term "virus" was proposed in the late nineteenth and early twentieth century (Beijerinck, 1898; Baur, 1904) to distinguish the filterable causative agents of some plant diseases
from bacterial agents, about 4000 viruses which infect plants, animals, fungi and bacteria have been characterized and classified into 71 families, 11 subfamilies and 164 genera on the basis of their genome which is either DNA or RNA, double-stranded or single-stranded and of positive- or negative-polarity and other characteristics (Murphy et al., 1995). Most viruses consist of ss(+) RNA genomes (Francki et al., 1994; Murphy et al., 1995).

Over the last decade the complete nucleotide sequences of a large number of ss(+) RNA virus genomes have been determined. From the sequencing data, the number, size and amino acid sequence of putative gene products can be predicted. The amino acid sequence of these putative translation products can be further compared to other viral and cellular protein sequences which have putative or characterized functions. Conserved motifs have been identified in some non-structural proteins encoded by different ss(+) RNA viruses. Proteins with such motifs include RNA-dependent RNA polymerases (replicases), putative RNA helicases, chymotrypsin-like and papain-like proteases and methyltransferases. Of these viral non-structural proteins, the replicases are most highly conserved and all ss(+) RNA virus replicases can be grouped unambiguously into one of three classes, i.e. POL 1, POL 2 and POL 3. Based on this finding, all ss(+) RNA viruses have been placed into these three classes, i.e. class 1 (containing POL 1), class 2 (containing POL 2) and class 3 (containing POL 3) (Steinhauer and Holland, 1987; Koonin and Dolja, 1993; Buck, 1996). Despite a wide variety of virion morphologies and hosts within the same class, viruses within this class share extensive similarities in their replicase and to a lesser extent in their genomic organization and gene expression and may, therefore, have derived from a common evolutionary origin (Koonin and Dolja, 1993; Strauss and Strauss, 1994; Buck, 1996). Class 1 contains picornavirus-like (or picorna-like), potyvirus-like, sobemo-like, and arteri-like superfamilies (superfamilies). Class 2 has fewer members (the carmo-like supergroup including bacteriophage, carmoviruses and tombusviruses; and the flavi-like supergroup including flaviviruses) but includes viruses with divergent virion morphologies and wide range of hosts. Class 3 is divided into 3 superfamilies, i.e., tymovirus-like, alphavirus-like (rubi-
like or sindbis-like) and tobamovirus-like supergroups.

The four supergroups in class 1, containing rod-shaped or icosahedral viral particles, are all eukaryotic viruses (plants and mammals as their hosts) and are distinguished from one another by their helicase type or lack thereof. It has been suggested that during evolution, each supergroup, starting with a common polymerase module, has acquired its own unique helicase independently (Goldbach and De Haan, 1994). The picorna-like supergroup consists of the Picornaviridae family including enteroviruses (poliovirus), hepatoviruses [hepatitis A virus (HAV)], cardioviruses (encephalomyocarditis virus), rhinoviruses [foot-and-mouth disease (FMDV)]; and the Comoviridae family including comoviruses [cowpea mosaic virus (CPMV)], nepoviruses [grapevine fanleaf virus (GFLV), tomato ringspot virus (TomRSV) and tobacco black ring virus (TBRV)], and fabaviruses [broad bean wilt virus (BBWV)]. The Potyviridae family, including potyviruses [tobacco vein mottling virus (TVMV) and tobacco etch virus (TEV)], has been suggested to constitute an independent supergroup under class 1 since the helicases of the family are different from those of the picorna-like supergroup (Goldbach and Van Kammen, 1985; Domier et al., 1987; Goldbach, 1987; Steinhauer and Holland, 1987; King et al., 1991; Koonin and Dolja, 1993; Buck, 1996; Strauss et al., 1996). The family is, however, closely related to the families Picornaviridae and Comoviridae in the features of genomic organization, gene expression and replication. In the literature (and this chapter), the Potyviridae are often referred to as a member of the picorna-like supergroup. The viruses of the picornavirus-like supergroup (including potyviruses) have the following features in common: (1) they are eukaryotic viruses (mammals and plants as hosts); (2) they are ss(+) RNA genomes; (3) a viral encoded genome-linked protein (VPg) is covalently linked to the 5' terminus of the genomic RNA and a poly(A) tract is present at the 3' terminus of the genomic RNA; (4) single long open reading frames (ORFs) coding for polyproteins are processed by viral encoded proteases to give the functional gene products; (5) they encode several non-structural proteins that have similar functions and significant amino acid sequence similarity (>20%); (6) they do not produce
subgenomic RNAs with exceptions of a nepovirus, cherry leaf roll virus (CLRV) (Brooks and Bruening, 1995), and of viruses in the family *Caliciviridae*; (7) they contain the same type of viral-encoded replicases (class 1); (8) they encode the same type of helicases (type 3) (type 2 for potyviruses); (9) they consist of icosahedral particles (rod-shaped for potyviruses) with separate encapsidation (each viral particle contains one viral RNA molecule); and (10) the genomic organization of non-structural proteins is similar.

### 2.3 Expression of Picorna-like Viruses and Other ss(+) RNA Viruses

#### 2.3.1 Expression strategies

To initiate viral RNA replication in the virus-infected living cells, the infecting ss(+) RNA viruses must be uncoated, at least to the extent of allowing the first ORF of the viral RNA to be translated. After uncoating, the naked ss(+) RNA molecules are recognized by ribosomes of the host cells, and translation starts.

ss(+) RNA viruses have developed the following strategies to enable the effective synthesis of virus-encoded proteins by the host protein-synthesizing systems (Davies and Hull, 1982; Kestelev *et al.*, 1982; Dougherty and Heibert, 1985; Matthews, 1991; Verver *et al.*, 1991). These strategies include: (1) Multipartite genomes. Viral genomes consist of a few RNA species, and each of them is independently replicated. The 5' ORF on each RNA segment is translated; (2) Subgenomic RNAs. One or more subgenomic RNAs are synthesized, and the 5' ORF on the genomic and subgenomic RNAs is translated; (3) Read-through proteins. Some ribosomes can read through the termination codon to allow translation of a second longer functional polypeptide; (4) Transframe proteins. A longer “transframe” protein is synthesized by initiation of translation at the same 5' AUG of the 5' ORF but switching of the reading frame near the first termination codon; (5) Internal initiation. Ribosomes recognize a specific sequence [internal ribosome entry site (IRES)] near an internally located AUG and initiate translation from that AUG; (6) Leaky ribosomal scanning. Ribosomes scan
from the 5' end of RNA, skip the 5' proximal start codon due to an unfavorable context surrounding the AUG codon, and start translation at the next downstream AUG codon; (7) Premature termination. Translation is terminated in the absence of a termination codon; (8) Polyprotein processing. A larger polyprotein is translated from a long ORF, and then cleaved at specific sites by a viral encoded protease or proteases (in some cases, host proteases are involved) to release the mature or intermediate precursor proteins. Most ss(+) viruses make use of two or three different strategies in combination. For example, animal viruses in the families Togaviridae and Caliciviridae and plant viruses in the genus Tymovirus use both polyprotein processing and subgenomic RNA strategies (Spall et al., 1997). In the picornavirus-like supergroup, plantcomo- and nepoviruses employ at least two strategies, i.e. multipartite genome and polyprotein strategies.

2.3.2 Genomic organization of picorna-like viruses

Picornaviridae The Picornaviridae represent one of the largest and most important families of human and agricultural viruses. The genome of the family is a single-stranded (+) RNA molecule of 7,200-8,500 nt excluding the poly(A) tail, which contains a single long reading frame that encodes a long polyprotein. The family is divided into five genera and contains over 217 serotypes. All picornaviruses employ a polyprotein strategy. In spite of differences in the number of mature proteins, all genera share a very similar arrangement of genome organization (Fig. 2.1). Poliovirus, the prototype of the family has been extensively studied. Excluding a 62 nt poly(A) tail, poliovirus has a genome of 7,433 nt which encodes a 247 kDa polyprotein (Rueckert, 1996). This polyprotein is cleaved by two virus-encoded proteases (see below) to ultimately generate 11 end products and 16 intermediate precursors (Harris et al., 1990, Wimmer et al., 1993). The P1 region encoding the coat protein is ultimately cleaved into 1A, 1B, 1C and 1D (Kitamura et al., 1981; Hellen and Wimmer, 1992). The P2 and P3 regions encode non-structural proteins. The P2 precursor includes 2A, 2B and 2C proteins and the P3 precursor contains 3A, 3B, 3C and 3D proteins. 2A is a protease and is also involved in the shutoff of host cell translation (Cuconati et al., 1998a). It may function
Fig. 2.1 Similarity and variability in genomic organization of the picorna-like supergroup. For clarity, coding regions are not drawn to scale. The genome maps of coding sequences are shown as boxes. The individual coding regions are indicated either above or in the corresponding boxes. Same patterns or symbols represent similar functions. Conserved regions are indicated by discontinuous lines. Abbreviations: 1A, 1B, 1C, 1D, CP, CP1, CP2, and CP, coat protein or capsid protein; 2A, 2A protease; 2B, no definite functions given (probably required for RNA replication); 2C, NTB and CI, helicase (NTB also contains a transmembrane domain indicated as a star at its C-terminus); 3A and 6k2, membrane anchor protein (star); 3B, VPG and NiaVPG; VPG (covalently linked to 5' end of each genomic RNA); 3C, Pro and NiaPro, 3C or 3C-like protease with similar structures; Pol and Nib, RNA-dependent RNA-polymersase (class 1); L, L protease; Y, unknown protein; X, unknown protein with certain homology with Co-Pro; Co-Pro, protease co-factor; P1, P1 protease; HC-Pro, HC protease; P3, possible movement protein; 6k1, unknown function.
directly in the replication of the viral genome as well (Lu et al., 1995). The functions of 2BC and its cleavage products, 2B and 2C are less defined. 2C contains a functional nucleotide triphosphate-binding (NTB) and hydrolysis domain (Cuconati et al., 1998a). 3AB is a membrane-associated protein (3A has a transmembrane domain) and the precursor to VPg (3B) (Semler et al., 1982). 3C is a protease and 3D, an RNA-dependent RNA polymerase (Xiang et al., 1998).

**Comoviridae (Comoviruses and Nepoviruses)** The genus *Comovirus* comprises 14 definitive species. All comoviruses use a multipartite genome (two RNA molecules, i.e. RNA-1 and RNA-2) and polyprotein as their translation strategy. Each RNA is linked to a VPg at its 5' end and includes a poly(A) tail at its 3' end. Excluding the poly(A) tails, the RNA-1 molecule has 5,850-6,050 nt and the RNA-2 molecule has 3450-3750 nt (Chen and Bruening, 1992 and 1994; Goldbach and Wellink, 1996). Each RNA molecule contains a single ORF and has 5'- and 3'-terminal nontranslated regions (NTRs). The type species, CPMV has been thoroughly studied with respect to genomic structure and translation strategy. The RNA-1 of CPMV has 5,889 nt excluding the poly(A) tail and encodes a 200 kDa polyprotein. This polyprotein is cleaved by a virus-encoded protease to release five end products, in the order of N-terminus, a protease co-factor (Co-Pro), an NTP-binding protein (NTB) which contains a nucleotide triphosphate (NTP)-binding domain and a transmembrane helix corresponding to 2C and 3A in poliovirus, respectively, a viral genome linked protein (VPg) corresponding to 3B, a protease (Pro) corresponding to 3C and an RNA-dependent RNA polymerase (Pol) corresponding to 3D (Fig. 2.1). The RNA-2 of CPMV has 3,481 nt excluding the 3' poly(A) tract and encodes a 116 kDa polyprotein from the first AUG of the 5' ORF or a 106 kDa polyprotein from the second AUG of the same ORF, which are processed by the RNA-1-encoded protease. The RNA-2 encoded polyprotein contains a movement protein (MP) and two coat proteins, i.e. coat protein 1 (CP1) and coat protein 2(CP2) (Fig. 2.1). Initiation of translation at two different AUG codons results in two proteins, 58 kDa and 48 kDa in size. The 48 kDa protein is a movement protein while the 58 kDa protein is probably required for RNA-2 replication (Van Bokhoven et al., 1993).
Nepoviruses are currently classified as the genus *Nepovirus* in the family *Comoviridae* (Goldbach *et al.*, 1995). This group includes 27 definitive members and 8 possible members (Chapter 1). Tobacco ringspot virus (TobRSV) is identified as the type member (Sanfaçon, 1995). Like comoviruses, nepoviruses adopt multipartite genome (bipartite) and polyprotein strategies. Nepoviruses have been subclassified into three subgroups, i.e. subgroup a, b and c (Chapter 1). Recently, the genomes of a few nepoviruses including TBRV, grapevine chrome mosaic virus (GCMV), GFLV and TomRSV have been cloned and sequenced (Meyer *et al.*, 1986; Grief *et al.*, 1988; Brault *et al.*, 1989; Le Gall *et al.*, 1989; Serghini *et al.*, 1990; Ritzenthaler *et al.*, 1991; Rott *et al.*, 1991b; Rott *et al.*, 1995). The arrangement of the protein domains in the nepovirus RNA-1-encoded polyprotein P1 closely resembles that in the RNA-1-encoded polyprotein of CPMV (Fig. 2.1). Starting from its C-terminus, P1 contains a Pol corresponding to the poliovirus 3D protein, a Pro corresponding to 3C, a VPg corresponding to 3B, an NTB (containing NTP-binding and transmembrane domains) corresponding to 2C and 3A, and an X protein. Although the X protein has certain homology with the 32 kDa protease-cofactor of CPMV, the experimental data obtained so far does not suggest that this protein is not a protease co-factor (Mayo and Robinson, 1996). In nepovirus subgroup c (e.g. TomRSV), there may be two or more proteins located upstream of the NTB domain since this region is extremely long and contains several predicted cleavage sites. The functions of these potential N-terminal proteins are not known. The arrangement of protein domains in the nepovirus RNA-2-encoded polyprotein (P2) differs from that of CPMV. Unlike the CPMV RNA-2-encoded polyprotein which contains a movement protein and two coat proteins, the nepovirus RNA-2-encoded polyprotein contains a single CP (coat protein), an MP (movement protein), and at least one unknown protein upstream of MP in the N-terminal region of P2 (Fig. 2.1).

**Potyviridae** The *Potyviridae* constitute the largest group of plant viruses. At least 180 definitive and putative members of this family are generally grouped into 3 definitive genera, *Potyvirus*, *Bymovirus* and *Rymovirus*, and one possible genus, *Ipomovirus* (Barnett, 1991 and 1992; Riechmann *et al.*,...
Since much information regarding this family is obtained from studies on the genus *Potyvirus*, this section will focus on this genus. Similar to picornaviruses, the genome of potyviruses contains a single-stranded (+) RNA molecule with a VPg protein covalently linked to its 5' terminus and a poly(A) tract at its 3' end. The genome of the sequenced viruses usually consists of 9,400-9,920 nt and encodes an extremely long polyprotein, over 300 kDa which is proteolytically processed by three viral encoded proteases. Tobacco etch potyvirus (TEV) is a well-characterized member of the family. The RNA sequence of TEV is 9,496 nt in length excluding the poly(A) tail and encodes a large polyprotein of 346 kDa (Allison *et al.*, 1986). At least 10 mature proteins, P1, HCPro, P3, 6k1, CI, 6k2 (or 6k), NIaVPg, NIaPro, Nlb and CP arise by proteolytic processing of this polyprotein (Dougherty and Semler, 1993) (Fig. 2.1). Except for the two 6k1 and 6k2 proteins, all other mature proteins have been detected in infected plants (Hari, 1995). The P1, HCPro and NIa are proteases which are responsible for the proteolytic processing. The P1 protein also has been proposed to be involved in viral RNA replication (Verchot and Carrington, 1995). The P3 protein is a possible movement protein (Dougherty and Semler, 1993). The function of the 6k1 protein has not been determined. The CI protein contains conserved domains known as helicase domains corresponding to 2C in poliovirus (Lain *et al.*, 1991). The 6k protein corresponding to 3A is associated with membranes and probably serves to anchor RNA replication complexes to the membranous structures (Restrepo-Hartwig and Carrington, 1994; Schaad *et al.*, 1997). Besides its function as a protease, the NIa protein along with its cleavage product, NIaVPg serves as the genome-linked protein (Murphy *et al.*, 1990). The Nlb protein corresponding to 3D is probably the RNA-dependent RNA polymerase (Allison *et al.*, 1986). The CP protein is the capsid protein. In addition to its role in the assembly and structure of the virion, it is probably involved in the movement of the virus in plants and in aphid-mediated transmission as well (Atreya *et al.*, 1990; Dolja *et al.*, 1994). Therefore, the CI-6k2-NIa-Nlb region of the polyprotein contains sequence and functional similarity with the 2C-3A-3B-3C-3D region of the picornavirus polyprotein and NTB-
VPg-Pro-Pol region of the comovirus- and nepovirus-encoded polyprotein.

2.3.3 Polyprotein processing in the picornavirus-like supergroup

Polyprotein processing as an important strategy of genome expression for ss(+)) RNA viruses was first described for poliovirus (Summer and Maizel, 1968; Jacobson and Baltimore, 1968). Since then, it has been shown that all members of picornavirus-like supergroups and many other ss(+) RNA viruses employ this strategy either as their principal mode of genome expression or in conjunction with other strategies such as multipartite genomes and subgenomic RNAs. In fact, almost all ss(+) RNA viruses require proteolytic processing during their replication cycle (Dougherty and Semler, 1993). Viruses which use this strategy encode one or more large precursor polypeptides that contain several domains which are proteolytically processed by viral encoded protease or cellular proteases to produce functional gene products.

Proteases play a key role in polyprotein processing. The virus-encoded proteases are highly specific for their substrates, and have been organized into four types on the basis of their active site nucleophiles, i.e. serine, cysteine, aspartic and metallo-proteases (Dougherty and Semler, 1993; Ryan and Flint, 1997).

Polyprotein processing in the Picornaviridae The genome of all picornaviruses consist of one (+)ss RNA molecule which encodes all viral proteins in a single long open reading frame. However, the unprocessed full-length polyprotein is barely detectable in vivo and in vitro. The polyprotein has been suggested to be processed into three intermediate precursors (primary products) P1, P2 and P3, by extremely rapid co-translational, intramolecular (in cis) cleavages (Ryan and Flint, 1997). Then these three primary products, following an ordered proteolytic cascade, are further processed either in cis or intermolecularly (in trans) to release different sub-sets of functional mature and intermediate precursor proteins required for RNA transcription and translation, and RNA encapsidation. There are four proteases, L, 2A, an oligopeptide, and 3C, which have been identified to be mainly responsible for processing of picornavirus-encoded polyproteins. The 2A protease
encoded by the entero- and rhinoviruses and the 3C protease encoded by all picornaviruses are serine-like proteases, while the L protease exclusively encoded by the genera *aphthovirus* of the family is a cysteine or thiol enzyme (Ryan and Flint, 1997).

The L protease of aphthoviruses, located at the N-terminus of the polyprotein (Fig. 2.1) cleaves co-translationally or post-translationally at the L-1A junction either *in cis* or *in trans* (Strebel and Beck, 1986; Medina *et al.*, 1993; Cao *et al.*, 1995). A relationship between the L protease and the cellular thiol proteinase papain has been suggested (Gorbalenya *et al.*, 1991), and confirmed by site-directed mutagenesis (Piccone *et al.*, 1995; Roberts and Belsham, 1995). Besides cleaving itself from the nascent polyprotein, L has been suggested to cleave the host cell protein eIF-4F as shown also for the 2A proteinase of the entero- and rhinoviruses (Devaney *et al.*, 1988; Kirchweger *et al.*, 1994). The cleavage of eIF-4F by the L and 2A proteases may be involved in host gene shut-off (Devaney *et al.*, 1988; Thatch, 1992).

In the entero- and rhinoviruses, the L protease is not present (Fig. 2.1). The 2A protease directs cleavage between the P1 capsid protein precursor and the P2-P3 replication-related protein precursor. This co-translational cleavage at the junction between 1D and 2A results in the generation of primary intermediate precursors P1 and P2-P3 (Toyoda *et al.*, 1986). Although the 2A protease has an active site thiol group, its structure has been predicted to be similar to that of the serine protease (Bazen and Fletterick, 1988). The aphtho- and cardioviruses do not have a 2A protease between the P1 and P2 regions. Instead, they have an oligopeptide which contains a self-cleaving activity (Palmenberg, 1990; Palmenberg *et al.*, 1992). The oligopeptide in the FMDV polyprotein consists of only 19 amino acids (Ryan *et al.*, 1991). This oligopeptide is able to cleave at its C-terminus in a co-translational manner. After this cleavage, the oligopeptide remains attached to the P1 region, and is subsequently cleaved *in trans* by the 3C protease (Ryan *et al.*, 1991). The mechanism of this cleavage has not been solved.

Unlike the above proteases which are only involved in a single primary cleavage, the 3C
protease is responsible both for the primary intramolecular cleavage at the 2C-3A site to produce P2 (2ABC) and P3 (3ABCD), and for all secondary cleavages to process the P1 capsid, and P2 and P3 replication-related protein precursors. This secondary processing occurs both in cis (to process the P2-P3 regions) and in trans (to release the mature capsid proteins from the primary intermediate precursor P1 region). However, an alternative primary cleavage by the 3C protease at the 2A-2B site for poliovirus (Lawson and Semler, 1992), and the 2B-2C site for FMDV (Ryan and Flint, 1997) has also been suggested. Efficient processing of the poliovirus capsid protein precursor P1 to release 1AB, 1C and 1D requires 3CD, an intermediate precursor from the P3 region (3ABCD) produced by rapid self-cleavage at the 3B-3C site. Both the 3C and 3CD forms of the protease are capable of processing the nonstructural polypeptides (P2 and P3 regions) (Jore et al., 1988; Ypma-Wong et al., 1988; Hellen and Wimmer, 1992). However, the requirement for 3CD but not 3C for P1 processing may not be common to all picornaviruses (Parks et al., 1989; Harmon et al., 1992). For example, in encephalomyocarditis virus, the 3C protease rather than 3CD is responsible for P1 processing (Parks et al., 1989). Processing of P2 yields three end products (2A, 2B and 2C) and one long-lived precursor (2BC) (Harris et al., 1990; Cuconati et al., 1998b). The four end products of P3, i.e. 3A, 3B, 3C and 3D, are obtained by slow processing of 3AB in trans and of 3CD in cis (Xiang et al., 1998). This sequential processing cascade may be, at least in part, regulated by intermediate precursor proteins and different forms of the 3C protease. In addition to the differential activities of 3C and 3CD on P1 cleavage sites, there are several other lines of evidence to support this suggestion. For example, the presence of 3AB protein stimulates auto-processing of 3CD to 3C and 3D. 3AB also stimulates processing of protein 2BC in trans by 3C or 3CD (Molla et al., 1994). The 3CD protease is more efficient than the 3C protease in cleaving 3AB in trans (Lama et al., 1994). The final maturation cleavage step during the poliovirus replication cycle is to cleave the intermediate capsid precursor protein 1AB into 1A and 1B, which are components of four mature coat proteins. This processing probably needs the presence of the viral RNA since it occurs late in capsid assembly.
and concomitantly with the encapsidation of the viral RNA (Hellen and Wimmer, 1992). Besides its proteolytic activities on viral proteins, the 3C protease is capable of cleaving a number of host-cell proteins such as the TATA-binding protein (Clark et al., 1993), microtubule-associated protein 4 (Joachims et al., 1995), transcription factor IIIC (Clark et al., 1991) and Histone H3 (Falk et al., 1990). Such disturbances of host-cell macromolecular processes by the 3C protease and other viral proteases mentioned earlier may constitute the sole causative factor of the host gene shut-off (Yalamanchili et al., 1996; Aranda and Maule, 1998).

Based on inhibitor studies, amino acid sequence comparisons, and a site-directed mutagenesis study, the active-site nucleophile of 3C is presumed to be the highly conserved Cys-147 (poliovirus 3C numbering) (Ivanoff et al., 1986). Subsequent studies have revealed significant homology between the Cys-containing proteases of a larger number of animal and plant ss(+) RNA viruses and the cellular chymotrypsin-like serine proteases (Gorbalenya et al., 1986; Bazan and Fletterick, 1988; Gorbalenya et al., 1989a; Bazan and Fletterick, 1990; Allaire et al., 1994; Matthews et al., 1994). His-40, Asp-85 or Glu-71, and Cys-147 of poliovirus 3C protease have been suggested to constitute the putative catalytic triad (Hämerle et al., 1991; Kean et al., 1991; Lawson and Semler, 1991). In addition, several lines of evidence suggest that the 3C protease contains a putative substrate binding pocket which recognizes cleavage sites of viral polyprotein substrates (Gorbalenya et al., 1986; Bazan and Fletterick, 1988; Gorbalenya et al., 1989a; Bazan and Fletterick, 1990; Nienaber et al., 1993; Allaire et al., 1994; Matthews et al., 1994). The specificity of cleavage by the picornavirus 3C proteases is for Q/(G,S,A,N). A His located within the C-terminal region of the 3C protease has been suggested to be an essential component of the pocket. Mutation of this His (at position 164) results in inactivation of the poliovirus protease (Ivanoff et al., 1986). The atomic structures of the recombinant 3C protease from rhinovirus-14 and hepatitis A virus show shallow elongated substrate pockets that could easily accommodate 8 amino acids from the -4 to +4 positions of the cleavage sites (Allaire et al., 1994; Matthews et al., 1994). Gln in the -1 position has been
suggested to be recognized primarily by the His of the protease substrate binding pocket.

**Polyprotein processing in the Comoviridae: Comoviruses and Nepoviruses** Upon translation of the RNA-1 and RNA-2 of comoviruses, two large polyproteins are produced that are processed into several intermediate and end cleavage products by an RNA-1-encoded 3C-like protease, the only protease of the virus. This process has been studied extensively for CPMV both *in vivo* and *in vitro*. Processing data for other comoviruses are scarce, but sequence comparisons suggest that processing follows pathways similar to those found for CPMV. The CPMV RNA-1 encodes a 200 kDa polyprotein which is co-translationally cleaved into 32 kDa (Co-Pro) and 170 kDa (NTB-VPg-Pro-Pol) proteins *in cis* (Fig. 2.2) (Pelham, 1979; Frassen *et al.*, 1984; Peng and Shih, 1984). The 32 kDa protein associates with the 170 kDa precursor and slows down the processing of the 170 kDa polyprotein, resulting in the temporal accumulation of the 170 kDa protein. In the absence of the 32 kDa protein, the 170 kDa is efficiently processed into 112 (VPg-Pro-Pol), 110 (Pro-Pol), 87 (Pol), 84 (NTB-VPg-Pro), 60 (NTB-VPg), 58 (NTB) and 24 (Pro) kDa proteins (Peters *et al.*, 1992a). Conversely, the 32 kDa protein is essential for the proteolytic activity of the 24 kDa protease on the RNA-2-encoded polyprotein at the Q/M cleavage site between the movement protein and the 66 kDa structural protein precursor (Vos *et al.*, 1988; Peters *et al.*, 1992a). Further cleavage of the 66 kDa structural protein precursor into the two capsid proteins is not affected by the 32 kDa protein. Studies with cleavage site mutants have shown that all cleavage sites in the RNA-1-encoded polyprotein occur most efficiently *in cis* (Peters *et al.*, 1992b). The 170 kDa protein can be processed *in cis* into three sets of proteins: 58 and 112, 60 and 110, and 84 and 87 kDa proteins. Further cleavage of the 112 kDa protein (VPg-Pro-Pol) occurs via two alternative pathways either into the VPg and 110 kDa proteins or into the 26 and 87 kDa proteins (Goldbach and Wellink, 1996). The 26 kDa protein is further quickly cleaved into the VPg and the 24 kDa protease. Processing of the 110 kDa protein corresponding to 3CD in poliovirus at the junction of the 24 kDa and 87 kDa proteins (Pro-Pol) is, however, extremely inefficient (Dessens and Lomonossoff, 1992). Cleavage at this site is greatly
Fig. 2.2. **Schematic representation of the expression strategy of the CPMV genome.** Open reading frames are indicated with open boxes. The VPg protein covalently attached to the 5' end of the genomic RNA is shown. The possible polyprotein processing pathways are described in the text. The final gene products, their molecular masses and possible functions are indicated. Abbreviations are the same as those in Fig. 2.1.
enhanced by sequences (NTB-VPg or VPg) upstream of the 24 kDa protease, suggesting that the mature 87 kDa (Pol) may arise only through direct processing of the 170 kDa protein. Unlike the 3C or 3CD proteases in poliovirus which can process 3AB \textit{in trans} to release 3A corresponding to the C-terminal portion of NTB in CPMV and 3B equivalent to VPg in CPMV (Lama \textit{et al.}, 1994), the CPMV 24 kDa protease can not cleave the 60 kDa protein (NTB-VPg) \textit{in trans} (Peters \textit{et al.}, 1992b).

The dipeptide sequences of CPMV cleavage sites have been determined to be Q/M (two sites), Q/S (two sites), and Q/G (two sites) (Wellink \textit{et al.}, 1986), which are very similar to those in picornavirus polyproteins.

Sequence alignments have shown that there is considerable amino acid sequence homology between the 24 kDa protease of CPMV and the 3C proteases of picornaviruses (Bazan and Fletterick, 1988; Gorbunova \textit{et al.}, 1989a). The His-40, Glu-75 and Cys-166 of the 24 kDa protein have been suggested to form the catalytic triad (Gorbunova \textit{et al.}, 1989). This prediction is supported by results obtained from mutational analyses on the predicted catalytic triad (Dessens and Lomonossoff, 1991).

Like comoviruses, each of the two RNA molecules of the nepovirus genome encodes one large polypeptide which is cleaved to release functional intermediate and end products by the RNA-1-encoded protease, the only known protease which is ca. 25 kDa. However, polyprotein processing has been considerably less studied for nepoviruses than for como- and picornaviruses. Unlike como-, poty- (see below), and picornavirus-encoded full-length polyproteins which are co-translationally processed into several primary products, nepovirus-encoded full-length polyproteins are at most only partially cleaved during translation (Demangeat \textit{et al.}, 1990; Hemmer \textit{et al.}, 1995). Translation of the TBRV RNA-1 \textit{in vitro} leads to synthesis of a 250 kDa polyprotein (Hemmer \textit{et al.}, 1995). This polyprotein is cleaved \textit{in cis} by the 23 kDa protease into intermediate precursor and mature proteins including 50 kDa (X protein containing certain sequence homology to the 32 kDa Co-Pro of CPMV), 60 kDa (NTB), 120 kDa (VPg-Pro-Pol) and 190 kDa (NTB-VPg-Pro-Pol) proteins. The 120 kDa protein is very stable and is not further cleaved \textit{in vitro}. Both the 120 kDa (VPg-Pro-Pol) and 117
kDa (Pro-Pol) proteins can, without addition of any other viral proteins cleave the RNA-2-encoded 150 kDa polyprotein *in trans* (Hemmer *et al.*, 1995) to generate a coat protein and the 96 kDa protein which is further processed into a movement protein and a 50 kDa protein with unknown function (Demangeat *et al.*, 1991 and 1992). Processing on the cleavage site between the 96 kDa protein and the coat protein by Pro-Pol was more efficient than by VPg-Pro-Pol (Hemmer *et al.*, 1995). This result suggests that nepovirus proteases, unlike comovirus proteases, are probably not regulated by a protease cofactor. A similar conclusion is reached from another experiment on a subgroup c nepovirus, TomRSV. The TomRSV protease can efficiently process a precursor protein which contains the cleavage site between the movement protein and the coat protein without the presence of the predicted protease co-factor (Hans and Sanfaçon, 1995). Study on processing of GFLV RNA-1-encoded proteins has shown that VPg-Pro-Pol, a presumed intermediate precursor in the RNA-1-encoded polyprotein cleavage cascade, can catalyze its auto-cleavage to produce Pol and the VPg-Pro protein which is very stable (Margis *et al.*, 1994). The 122 kDa polyprotein encoded by the GFLV RNA-2 can be cleaved by the RNA-1-encoded 24 kDa protease alone into a 56 kDa coat protein and a 66 kDa intermediate precursor which is subsequently processed into a 38 kDa movement protein and a 28 kDa protein with unassigned function (Margis and Pinck, 1992). It is also found that both forms of the protease either mature or as a VPg precursor can cleave the amino terminus of the RNA-1-encoded polyprotein *in trans* to release a 45 kDa protein which is located upstream of NTB and has unknown functions. Both proteases can also cleave an RNA-2-encoded polyprotein. Efficiency of processing is regulated by the maturation of the protease itself. VPg-Pro processes the RNA-1-encoded proteins faster than the mature Pro, while Pro processes the RNA-2-encoded polyprotein faster than VPg-Pro (Margis *et al.*, 1994).

Nepovirus polyproteins have a wide range of dipeptide sequences at the cleavage sites. The characterized dipeptide sequences of nepovirus subgroup a and b cleavage sites are R/G, C/A, C/S and G/E in the GFLV polyprotein, R/A in the GCMV and K/A and K/S in the TBRV polyprotein,
which are different from (Q, E)/(S, G, E, M, A, N) in picorna-, poty- (see below) and comovirus polyproteins (Wellink et al., 1986; Brault et al., 1989; Serghini et al., 1990; Pinck et al., 1991; Demangeat et al., 1992; Margis et al., 1993; Hemmer et al., 1995). However, all characterized cleavage sites of nepovirus group c, such as Q/G in the TomRSV polyprotein (Hans and Sanfaçon, 1995), N/S in the polyprotein encoded by blueberry mottle virus (BLMV) (Bacher et al., 1994), D/S in the polyprotein encoded by mite-transmitted blackcurrant reversion associated virus (BRAV) (Latvala et al., 1998) and Q/S in the polyprotein encoded by cherry leaf roll virus (Scott et al., 1993) resemble the picorna-, como- and potyvirus polyprotein cleavage sites.

Alignment of nepovirus protease sequences with the comovirus 24 kDa protease and picornavirus 3C protease sequences suggest that nepovirus proteases may have a similar secondary structure and a similar catalytic triad (Margis and Pinck, 1992; Hans and Sanfaçon, 1995). However, nepoviruses of subgroup a and b (GCMV, GFLV and TBRV) possess a Leu in the putative substrate binding pocket, instead of the typical His found at this position in picorna-, como-, and potyviruses and nepovirus subgroup c (TomRSV). This may explain why subgroup a and b nepovirus proteases have different cleavage specificity from subgroup c nepoviruses, and como-, poty-, and picornaviruses. As mentioned earlier, the conserved His in the putative substrate binding pocket of the proteases is probably interacting directly with the Gln at -1 position of the cleavage site (Bazen and Fletterick, 1988; Allaire et al., 1994; Matthews et al., 1994). Results of site-directed mutagenesis of nepovirus proteases are consistent with the suggested composition of the putative catalytic triad and the putative substrate binding pocket (Margis and Pinck, 1992; Hans and Sanfaçon, 1995).

**Polyprotein processing in the Poatyviridae** The genome of the viruses of the Poty-, Bymo-, Rymo- and presumably Ipomovirus genera contains a single open reading frame resulting in the translation of a large polyprotein. This polyprotein is 346 kDa in TEV (Allison et al., 1986), 358 kDa in turnip mosaic virus (Nicolas and Laliberte, 1992), 368 kDa in potato virus Y (Robaglia et al., 1989) and 355-357 kDa in plum pox virus (Lain et al., 1989a). This polyprotein is cleaved into at least ten end
products by three virus-encoded proteases, P1, HCPro and Nla. A model of TEV polyprotein processing has been proposed (Carrington et al., 1989; Verchot et al., 1991; Dougherty and Semler, 1993). Upon translation of the TEV polyprotein, four or five primary products are cotranslationally proteolytically produced, 87 kDa (P1-HCPro), 121 kDa (P3-6k1-Cl), 54 kDa (6k2-Nla) or 6k2 plus 49 kDa (Nla) and 88 kDa (NIb-CP). The 87 kDa P1-HCPro protein is obtained by self-cleavage of HC-Pro at its C-terminus at the junction between HCPro and P3. To generate the 121 kDa protein, Nla cleaves at the Cl-6k2 cleavage site. The Nla further cleaves at the Nla-Nlb cleavage site and optionally at the 6k2-Nla cleavage site to obtain either the 88 and 54 kDa or the 88, 6 and 49 kDa proteins. Then, post-translational processing takes place. The 87 kDa protein is auto-cleaved by the P1 protease to mature itself (35 kDa) and HCPro (52 kDa). The 121 kDa protein is processed by Nla \textit{in trans} to generate P3 (44 kDa), 6k1 (6 kDa) and CI (71 kDa). The Nla protein cleaves the 88 kDa protein \textit{in trans} at the Nlb-CP cleavage site to release Nlb and CP (30 kDa). Self-cleavage of the 6k2-Nla protein by the Nla protease releases 6k2 and Nla. At last, self-cleavage of Nla occurs to release the VIaVPg (22 kDa) and VIaPro (27 kDa). Recently, it has been found that the VIaPro C-terminal twenty amino acids are cleaved autocatalytically (Kim et al., 1995; Parks et al., 1995).

P1 and HCPro proteases are apparently only involved in one cleavage at their own C-terminus (Carrington et al., 1989; Verchot et al., 1991). Sequence comparison and site-directed mutagenesis analysis suggest that the P1 protease is similar to the smaller bacterial serine protease while the HC protease shares certain structural similarity with cellular cysteine proteases (Dougherty and Semler, 1993).

The main source of proteolytic activities in potyviruses is from Nla. Like como- and nepovirus group c 3C-like proteases, picornavirus 3C proteases and cellular trypsin-like serine proteases, Nla has a putative catalytic triad composed of His, Asp and Cys (Bazan and Fletterick, 1988; Gorbalenya et al., 1989a). The importance of these three amino acids to the proteolytic activity of Nla has been confirmed by site-directed mutagenesis (Carrington et al., 1988; Dougherty et al.,
1989; Garcia et al., 1990). Similar to subgroup c nepovirus proteases and como- and picornavirus proteases, a His located on the C-terminal sequence has been found to be a base component of the putative substrate binding pocket. Therefore, it is not surprising that the dipeptide sequences of most NIa-processed cleavage sites within the potyvirus-encoded polyproteins are Q/(A,S,G).

2.4 Viral Proteins Associated with Viral RNA Replication

Translation of ss(+)-virus RNA is absolutely dependent on existing cellular translation machineries. Viral RNA transcription and replication is relatively more dependent upon viral gene products. It is believed that normal cells do not have preexisting transcription and replication machinery to support viral RNA-dependent RNA transcription and replication. Increasing evidence suggests that this process is accomplished by a so-called replication complex, which is mainly composed of viral proteins including an RNA-dependent polymerase (RdRp), a helicase (and membrane binding protein), capping and methylation enzymes (for viruses with 5' capped genomic RNA), a VPg (for the picorna-like supergroup) and a protease (for the picorna-like supergroup). During the past several years, active viral replication complexes have been successfully isolated from infected cells for bromo mosaic virus (BMV) (Quadt and Jaspars, 1990; Kao et al., 1992; Dinant et al., 1993; Quadt et al., 1993; Quadt et al., 1995), for CPMV (Dorssers et al., 1984; Peters et al., 1994), for cucumber mosaic virus (CMV) (Hays and Buck, 1990; Hayes et al., 1994), for poliovirus (Bienz et al., 1990; Giachetti and Semler, 1991; Bienz et al., 1992), for potato virus X (PVX) (Doronin and Hemenway, 1996) and for tobacco mosaic virus (TMV) (Osman and Buck, 1996 and 1997). The viral proteins are assembled in the complexes either as precursors or as mature proteins by protein-protein interactions.

2.4.1 RNA-dependent RNA polymerase

Of the viral proteins involved in replication, RdRp has been characterized as a critical enzyme for RNA replication since it polymerises minus and plus viral RNA strands (Richards and
Ehrenfeld, 1990; Porter, 1993; Duggal et al., 1994; De Graaff and Jaspars, 1994; Buck, 1996). Sequence comparison studies on ss(+) RNA viruses have shown that all sequenced ss(+) RNA viruses encode a putative RdRp which contains a consensus sequence motif Gly-Asp-Asp (GDD) embedded in a stretch of hydrophobic amino acids and a few other highly conserved motifs (Kamer and Argos, 1984; Franssen et al., 1984; Koonin and Dolja, 1993).

Poliovirus, a member of picornaviruses was the first ss(+) animal RNA virus for which an RdRp was isolated (Giachetti and Semler, 1991; Bienz et al., 1992). A partially purified replication complex, solubilized from cellular membranes was shown to have the ability to extend an oligo(U) primer hybridized to any RNA with a 3'-terminal poly(A) tail. The elongation did not stop until the end of the template was reached. Recombinant 3D protein expressed in E. coli can catalyze primer-dependent poly(U) synthesis, which is stimulated by addition of exogenous 3AB expressed and purified from bacterial cells (Lama et al., 1994). Genetic analyses have shown that 3D rather than 3CD has polymerase activity (Harris et al., 1994). Complete replication of poliovirus (both minus and VPg-capped genomic RNA) in vitro requires the replication complex purified from infected cells, which contains soluble cellular factors, cellular membranes and replicative-intermediate RNA, and is enriched in viral protein 3CD (protease and RdRp) and the membrane-associated viral proteins 2C (helicase), 2BC (helicase precursor protein) and 3AB (membrane binding protein and VPg) (Barton et al., 1995). Mutation of any single amino acid of the conserved sequence, GDD in the RdRp abolishes RNA replication (Jablonski and Morrow, 1995).

For plant comoviruses, a partial replication complex was successfully purified from CPMV-infected plant cells (Dorssers et al., 1984). After solubilization from cellular membranes with detergents, the purified complex contained a 110 kDa viral protein (Pro-Pol) corresponding to 3CD in poliovirus, two cellular proteins (68 kDa and 57 kDa) and viral RNA. This complex was able to elongate in vitro preexisting nascent chains to full length viral RNAs. Unlike the poliovirus 3D, the CPMV Pol or its precursors expressed in bacterium cells or baculoviruses were inactive in a poly(A)-
oligo(U) polymerase assay. This suggests that the polymerase from como- and picornaviruses may differ in the mechanisms leading to initiation of RNA replication (Richards et al., 1989; Van Bokhoven et al., 1991).

For plant potyviruses, substitutions of the conserved GDD in the TEV Nib (the putative RdRp) have been shown to be lethal. The Nib-defective mutants can be rescued in trans by the wild-type Nib expressed in transgenic plant cells (Li and Carrington, 1995). These results suggest that the mature Nib is essential for RNA amplification and that Nib precursor polyproteins may not perform essential roles at the later stages of replication, i.e. after polyprotein processing. It is also possible that the mature Nib and other viral proteins including Nib precursors are all required for formation of a replication complex in which the mature Nib plays a central role for RNA polymerization.

The replication process of the plant viruses in the picorna-like supergroup (includingcomo-, nepo- and potyviruses) is poorly understood. Characterization of RdRp has been more successful for plant viruses of sindbis-like supergroup such as turnip mosaic virus (TYMV) (Mouches et al., 1984; Candresse et al., 1986; Singh and Dreher, 1997), brome mosaic virus (BMV) (Quadt and Jaspars, 1990; Kao and Ahlquist, 1992; Kao et al., 1992; Janda and Ahlquist, 1993; Quadt et al., 1993; Quadt et al., 1995; Kao and Sun, 1995; Restrepo-Hartwig and Ahlquist, 1996; Smirynagina et al., 1996; Sun and Cao, 1997) and cucumber mosaic virus (CMV) (Hays and Buck, 1990 and 1993; Hays et al., 1994). Taking CMV as an example, complete replication can be catalyzed in vitro by a soluble RdRp preparation, which is purified by differential centrifugation, detergent treatment and chromatography (Hays and Buck, 1990). The preparation contains two viral proteins (1a and 2a) and a cellular protein. After removal of the endogenous RNA, the RdRp activity of the preparation is completely and specifically dependent on addition of the CMV RNA as a template. All the three genomic RNA species and one subgenomic RNA molecule are produced. Furthermore, the preparation catalyzes synthesis of both stages of the replication process, i.e. synthesis of the positive strand RNA, as well as the negative strand RNA (Hays and Buck, 1990). A monoclonal antibody
which specifically recognizes the GDD conserved motif of the 2a protein (RdRp) can inhibit the RdRp activity of the preparation in vitro (Hays et al., 1994), demonstrating the functional importance of the identified motif GDD in CMV replication.

2.4.2 NTP-binding protein, helicase and membrane binding protein

A second viral protein required for RNA replication is the helicase-like NTP-binding protein (NTB), which is also conserved among ss(+) RNA viruses (Franssen et al., 1984; Gorbunova et al., 1989b; Gorbunova et al., 1990; Neill, 1990; King et al., 1991, Peters et al., 1994; Rott et al., 1995). All ss(+) RNA viruses with genomes over ca. 6 kb encode an NTB motif (NTBM). NTBM is a characteristic amino acid sequence motif for ribonucleotide-utilizing proteins, consisting of two separate parts referred to as the A and B sites (Walker et al., 1982). The NTBM-containing proteins are the 2C protein in picornaviruses and the CI protein in potyviruses. Computer-assisted sequence analysis among the NTBMs of all ss(+) RNA viruses has led to group them into three major classes (Gorbunova et al., 1988 and 1989b). The NTBM-containing proteins of sindbis-like viruses and potyviruses bear resemblance to helicases that are related to eukaryotic translation initiation factor eIF-4A (Class II), and prokaryotic helicases, i.e. UvrD, Rep, RecB and RecD. The NTBM-containing proteins of picorna-, como- and nepoviruses are related to Simian virus-40 (S40) larger T antigen (Class III), a protein containing RNA and DNA helicase activity (Hodgman, 1988a and 1988b; Gorbunova and Koonin, 1989; Lain et al., 1989b, 1990 and 1991; Gorbunova et al., 1990). Therefore, the NTBM-containing proteins probably act as a helicase.

The poliovirus 2C protein is required for RNA replication (Bienz et al., 1992; Wimmer et al., 1993). It is present in the membranous replication complex isolated from infected cells (Bienz et al., 1990). Mutations in the conserved amino acids of the NTP-binding motifs A and B abolish or greatly reduce RNA synthesis (Mirzayan and Wimmer, 1992, Teterina et al., 1992). The 2C protein, expressed in E. coli or insect cells has been shown to have RNA binding activity and RNA-dependent ATPase activity, though its putative RNA helicase activity has not yet been demonstrated.
(Mirzayan and Wimmer, 1992; Rodriguez and Carrasco, 1993 and 1995). In CPMV, a comovirus, the importance of the principal NTP-binding motif for virus replication has been demonstrated by site-directed mutagenesis (Peters et al., 1994). In CMV, a member of sindbis-like viruses, the helicase-like protein (Class I) is an essential subunit of the replication complex that is purified from infected cells and can catalyze complete replication of CMV RNAs in vitro (Hays and Buck, 1990). Duplex unwinding in a 3'-5' direction has been reported for a purified replication complex, isolated from plants infected with alfalfa mosaic virus, another member of sindbis-like viruses (De Graaf et al., 1995). The replication complex contains the helicase-like 1a protein (Class I). However, it is not clear whether the unwinding is due to the 1a protein, or 2a protein (RdRp), or some other protein components of the complex (Quadt et al., 1991). RNA helicase, RNA-stimulated ATPase, and RNA binding activities have been observed for the viral proteins of a few poty- and flaviviruses, i.e. the CI protein of plum pox (PPV) and tamarillo mosaic (TaMV) potyviruses (Lain et al., 1990 and 1991; Eagles et al., 1994; Fernandez et al., 1995) and the NS3 proteins of yellow fever, West Nile and hepatitis C flaviviruses (Wengler and Wengler, 1991; Warrener et al., 1993; Gwack et al., 1997). The NTB proteins in como- and nepoviruses have been proposed to be the anchor proteins responsible for the attachment of the replication complex to membranes (Eggen and Van Kammen, 1988; Sanfaçon, 1995). Unlike NTB proteins in como- and nepoviruses which contain a predicted transmembrane domain at their C-terminus, CI and 2C apparently lack a defined membrane binding domain. A small protein immediately downstream of the CI and 2C proteins containing a membrane-binding domain, denoted as 6K2 and 3A, respectively, has been shown to contain membrane binding abilities in vitro and in vivo (Lama et al., 1994; Xiang et al., 1995; Schaad et al., 1997; Xiang et al., 1998). This protein is suggested to act as the anchor protein for binding the partial or whole replication complex to membranous structures (Towner et al., 1996; Schaad et al., 1997). It is possible that in the replication complex, CI and 2C interact with membrane binding proteins such as 6K2 and 3A, or exist as precursor polyproteins such as CI-6k2 and 2C-3AB. Indeed, the CI-6k2
precursor is detected in TEV-infected cells (Restrepo-Hartwig and Carrington, 1994). Taken together these results suggest that the NTBM-containing proteins encoded by ss(+) RNA viruses might in fact be NTPase that are involved in processes such as duplex unwinding during RNA replication and transcription, mRNA translation, signal transduction and membrane transport (Walker et al., 1982; Gorbalenya and Koonin, 1989; Gorbalenya et al., 1989b).

2.4.3 Methyltransferase

The alpha-like, corona-like, flavi-like and carmo-like viruses of the ss(+) RNA viruses possess capped genomic RNA. Replication of these viruses requires a virus-encoded methyltransferase. In alpha-like viruses, a protein domain with sequence similarity to the Sindbis virus nsP1 protein is conserved and located at the N-terminus of the nonstructural polyprotein (Mi et al., 1989; Mi and Stollar, 1991; Scheidel and Stollar, 1991). The sindbis virus nsP1 protein has methyltransferase and possibly guanylytransferase activities thought to be involved in viral RNA capping. These domains have unique conserved motifs different from those found in cellular methyltransferases (Rozanov et al., 1992). In BMV, this domain is contained in the N-terminal region of the RNA-1-encoded 1a protein, which also contains a helicase-like domain at its C-terminus (Ahlquist et al., 1994). The 1a and the RdRp-like 2a proteins interact in vitro and in vivo, and are the required components of the functional replication complex purified from infected cells (Quadt and Jaspars, 1990; Kao et al., 1992; Quadt et al., 1995). Mutations of all three conserved domains, i.e. helicase-like, RdRp-like and methyltransferase-like domains result in destruction of genomic RNA synthesis and accumulation (Kroner et al., 1990; Traynor et al., 1991), suggesting that all three are required for RNA replication. It is noteworthy that the methyltransferase-containing protein is also a required subunit of the replication complexes purified from cells infected by BMV-related alpha-like viruses (Hays and Buck, 1990; Quadt et al., 1991).

2.4.4 Genome-linked viral protein

VPg (viral protein genome-linked) is a virus-encoded protein that is covalently linked by a
phosphodiester linkage to the 5'-terminal nucleotide of the virus genomic RNA. The ss(+) RNA viruses with VPg capped genomic RNAs include the picorna-like and sobemo-like supergroups. The function of VPg has not been definitely established. For poliovirus, a picornavirus, it is believed that 3AB, a membrane-associated VPg-precursor protein is used for the delivery of VPg to the 5' ends of plus- and minus-strand RNAs during RNA replication (Semler et al., 1982). However, the mechanism by which the covalent linkage of VPg to the genomic RNA occurs remains controversial.

It has been suggested that the VPg is involved in RNA replication as a primer (Lee et al., 1977; Takeda et al., 1987; Reuer et al., 1990; Cao et al., 1993; Paul et al., 1998). This hypothesis is supported by the following observations: (i) all newly synthesized picorna-viral RNAs recovered from infected cells are linked to VPg (Lee et al., 1977); (ii) anti-VPg antibodies can specifically inhibit initiation of poliovirus RNA synthesis (Baron et al., 1982; Morrow and Dasgupta, 1983); (iii) in the replication complex, the endogenous VPg can be uridylylated in vitro to VPg-pUpU and subsequently extended to longer RNA molecules (Takeda et al., 1986); (iv) mutation in a hydrophobic domain just amino terminal of VPg leads to a defect in viral plus-strand RNA synthesis of polioviruses (Giachetti and Semler, 1991); and (v) antibodies raised against VPg do immunoprecipitate VPg as well as a uridylylated form from poliovirus-infected cells (Crawford and Baltimore, 1983). Most recently and decisively, it has been shown that a synthetic VPg peptide can be uridylylated by the poliovirus 3D protein expressed and purified from E. coli, and the uridylylated VPg can prime the transcription of a poly(A) oligonucleotide by 3D to produce VPg-linked poly(U) (Paul et al., 1998). Another model has been proposed for the involvement of VPg in synthesis of minus-strand RNA. In this model, poliovirus 3D and a host factor (Dasgupta et al., 1980; Dasgupta, 1983) catalyze self-priming of minus-strand synthesis on a plus-strand template (Andrews and Baltimore, 1986). The hairpin generated at the junction of plus- and minus-strand RNAs is then cleaved by VPg or a VPg precursor with concomitant linkage of VPg to the 5' end of newly made minus-strand RNA (Tobin et al., 1989). The poliovirus VPg is a basic, 22-amino-acid oligopeptide
that is linked to the RNA by a phosphodiester bond between the 5'-terminal uridine residue of the RNA and the O\(^4\)-hydroxyl group of the tyrosine, the third residue of the VPg (Ambros and Baltimore, 1978). Substitutions of this tyrosine block RNA replication (Reuer et al., 1990).

The involvement of VPg in RNA replication of other picorna-like viruses is not well understood. In potyviruses, replacement of the tyrosine residue that links the tobacco vein mottling virus (TVMV) VPg to the viral RNA is lethal (Murphy et al., 1996). The 6k2-Nla polyprotein has been proposed to serve as the critical membrane-bound VPg precursor during initiation of RNA synthesis (Hong et al., 1995; Li et al., 1997; Daros and Carrington, 1997). In nepoviruses, the presence of a covalently linked VPg protein at the 5' end of the genomic RNAs is necessary for their infectivity but not for their translation in vitro (Chu et al., 1981; Mayo et al., 1982; Hellen and Cooper, 1987). It is not clear, however, whether the covalent linkage of VPg to the genomic RNA is required solely to protect the RNA genome from degradation, to direct the genomic RNA for encapsidation and cell-to-cell movement or to be a primer for RNA replication. In comoviruses, the NTB-VPg precursor corresponding to 3AB in poliovirus, is abundant in CPMV-infected plant cells (Peters et al., 1992a). However, it is not likely that NTB-VPg is the precursor protein that delivers VPg to the genomic RNA since cleavage at the NTB-VPg site in trans is extremely inefficient. In contrast, 3AB can be cleaved in trans by the poliovirus 3C or 3CD protease (Lama et al., 1994), as discussed earlier. In line with this suggestion, the 112 kDa protein (VPg-Pro-Pol) has been proposed as a possible donor of VPg based on the finding that the 112 kDa protein can be processed by its protease to release the VPg in vivo (Peters et al., 1995). Direct experimental proof is still lacking to support this suggestion.

Despite poor understanding of the functions of the VPg proteins in comoviruses and nepoviruses, the sequences of several VPg proteins have been established by direct microsequencing of the genomic RNA-linked VPg proteins (Jaegle et al., 1987; Pinck et al., 1991; Hemmer et al., 1995; Zalloua et al., 1996). In contrast to the picorna- and potyviruses, the amino acid which links VPg
to the genomic RNA has been determined to be a serine near or at the N-termini of the VPg proteins for nepo- and comoviruses. Alignment of the amino acid sequences of these VPg proteins has shown that viruses within the family Comoviridae are highly conserved (Mayo and Fritsch, 1994).

2.4.5 Protease

The critical function of the 3C or 3C-like proteases as the enzyme responsible for the majority of maturation cleavage in the precursor polyprotein has been discussed earlier in this chapter. In recent years, other possible functions of these proteases in RNA replication have been investigated in picorna- and potyviruses but not in como- and nepoviruses. The poliovirus 3C protease is a critical component of the replication complex (Andino et al., 1990; Clark et al., 1991; Andino et al., 1993). Although picornavirus proteases clearly lack recognizable conserved RNA-binding motifs (Burd and Dreyfuss, 1994), the poliovirus 3C protease possesses the ability to bind RNA (Andino et al., 1993). This binding ability can be strengthened by the presence of 3D as in the 3CD precursor. The poliovirus 3C protease in the 3CD precursor specifically binds to a computer-predicted cloverleaf structure at the 5' end of poliovirus RNA in the presence of the host factor hnRNP E to form a ribonucleoprotein (RNP) complex which is essential for RNA replication (Andino et al., 1990 and 1993; Xiang et al., 1995; Gamarnik and Adino; 1997; Parsley et al., 1997). Disruptions of the RNP complex abolish RNA replication but do not affect RNA translation or stability (Andino et al., 1993). In another study, 3CD has been found to interact with another host protein EF-1α and binds to the 5' and 3' termini of the poliovirus genome (Harris et al., 1994). Crystallographic analysis of human rhinovirus-14 (HRV-14) 3C protease suggests that Arg87 is involved in the RNA-binding activity (Matthews et al., 1994). The conserved domain involved in the RNA-binding activity in the HRV-14 3C protease has been mapped to amino acids 74-90 by mutagenesis (Walker et al., 1995). Indeed, a substitution affecting Arg87 of the poliovirus 3C eliminates virus viability (Hammerle et al., 1992). The 3CD protein is also found to interact with 3AB, the membrane anchor protein and the precursor protein of VPg (Xiang et al., 1995 and 1998).
Based on these findings, a model for poliovirus RNA synthesis has been suggested. In this model, 3AB interacts with 2C, 3CD, and the host factor EF-1α to bind to the 3' terminus of the viral genome and form a replication complex on cellular membrane (Harris et al., 1994; Molla et al., 1994; Xiang et al., 1998). Proteolytic cleavage by 3CD releases 3D and 3B. 3B is eventually uridylylated by 3D in the presence of the viral RNA attached to 3C (Lama et al., 1994; Molla et al., 1994; Xiang et al., 1998). Then the uridylylated VPg serves as a primer for 3D to catalyze (-) RNA synthesis (Paul et al., 1998). Once 3D has transcribed the portion of the 5' non-coding region of the viral genome corresponding to the cloverleaf structure, a new complex containing the host factor hnRNP E may be assembled on the 5'-end of the viral genome (Andino et al., 1993; Xiang et al., 1995; Gamarnik and Adino; 1997; Parsley et al., 1997). A similar mechanism is used to produce VPg-pUpU which will prime to the 3' end of the newly synthesized (-) strand. The (+) strand genome is synthesized by 3D.

Potyviral proteins NiaPro (a 3C-like protease) and Nia (containing VPg and protease domains) do not have a clearly recognizable conserved RNA-binding motif either. But both viral proteins, like the two other potyviral proteases, P1 and HCPro, possess nonspecific RNA-binding activity (Brantley and Hunt, 1993; Soumounou and Laliberte, 1994; Maia and Bernardi, 1996; Daros and Carrington, 1997). Nia has been determined to interact with Nlb (RdRp) in vivo (Hong et al., 1995; Li et al., 1997). The Nlb protein is required for RNA replication and can be supplied in trans outside of the context of the TEV polyprotein (Li et al., 1995). Therefore, a model similar to the model for poliovirus replication has been suggested to describe potyvirus RNA replication. In this model, the 6k2-Nia precursor interacts with Nlb to initiate RNA replication (Li et al., 1997).

These models are inferred from the current understanding of viral proteins mainly implicated in (-) RNA transcription. How the replication complex synthesizes the genomic, i.e. positive RNA remains to be elucidated. Nevertheless, it is clear that in picornavirus RNA replication, the protease play essential roles not only on proteolytic processing but also on RNA synthesis.
2. 5 *Cis*-Acting Viral RNA Sequence Required for Viral RNA Replication

Essential *cis*-acting sequences required for RNA transcription and replication include promoters and additional sequences involved in the assembly of the replication complex. Significant progress in this regard has been made for a number of viruses, in particular, many of the plant alapha-like viruses such as BMV (Quadt *et al.*, 1995; Sun and Kao, 1996; Siegel *et al.*, 1997). Excellent reviews have been written by Florentz and Gierge (1995), Buck (1996) and Xiang *et al.* (1997). Here the emphasis will be put on picorna-like viruses.

2.5.1 3'-Terminal sequences of genomic(+) RNA in picorna-like viruses

The 3'-terminal sequences in all picorna-like viruses are poly(A). The poly(A) tail is required for the infectivity of poliovirus and CPMV since the viral RNA without a poly(A) tail is not infectious (Sarnow, 1989; Eggen *et al.*, 1989a). The poly(A) sequence should be transcribed from a poly(U) sequence at the 5' end of the (-) strand. It is very interesting that the poly(A) tail of poliovirus genomic RNA is longer than the poly(U) tract of the (-) strand (Larsen *et al.*, 1980). Presumably, when the (-) strand is transcribed, the VPg-pUpU hybridizes to the poly(A) of genomic RNA either at random or defined positions which are distinct from the very 3' end of the genomic RNA. When the progeny (+) genomic RNA is synthesized, the additional A residues may be added by a terminal nucleotidyl transferase or by slippage of RdRp on a poly(U) template (Neufeld *et al.*, 1994). In CPMV, it has been shown that infectious transcripts with shorter 3' poly(A) tail produce progeny genomic RNA with poly(A) tails of variable length (Eggen *et al.*, 1989b). Therefore, in picorna-like viruses, it is likely that the poly(A) tail is not copied exactly from the poly(U) stretch during plus-strand synthesis.

Sequences upstream of the 3' poly(A) are also important for RNA replication. The region upstream of the poly(A) tail of poliovirus RNA is predicted to form a tRNA-like secondary structure (Sarnow, 1989; Pilipenko *et al.*, 1992; Rohll *et al.*, 1995). A pseudoknot between the 3' untranslated region and sequences upstream of the translational terminator has been confirmed and shown to be
essential for replication (Jacobson et al., 1993). The 3C domain of the 3CD precursor binds to this pseudoknot structure in the presence of 3AB (Harris et al., 1994). A mutation that affects the pseudoknot structure confers a temperature-sensitive phenotype (Sarnow et al., 1986). Based on the above genetic and biochemical data, it is suggested that this structure may be involved in the initiation of (-) strand RNA synthesis (Harris et al., 1994). In CPMV, the 3'-terminal 151 nt of RNA-2 are required for its replication (Rohll et al., 1993). The 65 nt at the 3' untranslated region upstream of the poly(A) tail are homologous in RNA-1 and RNA-2 and can be folded to form a stem-loop structure (Eggen et al., 1989b; Rohll et al., 1993). Site-directed mutagenesis experiments have shown that these secondary structures rather than their sequences are the important determinants for efficient replication (Eggen et al., 1989a and 1989b; Rohll et al., 1993). Indeed, exchange of the 3' noncoding regions of RNA-2 with those of RNA-1, which fold into similar structures, has little effect on RNA replication (Van Bokhoven et al., 1993).

2.5.2 5'-terminal sequences of genomic RNA in picorna-like viruses

The 5' untranslated region of the genomic RNA of poliovirus can form two secondary structures which appear to act independently for different purposes in the replication cycle. The first one is a cloverleaf-like secondary structure formed by the 5' terminal 88 nt, which has been suggested to play a central role in organizing viral and cellular proteins involved in positive strand production (Andino et al., 1993; Rohll et al., 1994). The second one is a complex secondary structure, formed by nt 120 to 640, which directs translation by internal ribosome entry (Nicholson et al., 1991; Rohll et al., 1994). Mutations that modify the cloverleaf in the positive strand but not in the negative strand are lethal to the virus (Andino et al., 1993). A viral protein precursor, 3CD, is found to bind to this RNA structure (Andino et al., 1990). Mutations either in the specific region of the RNA cloverleaf or in 3CD that prevent this binding inhibit virus growth and RNA accumulation (Andino et al., 1993).

The 5' untranslated regions of the genomic RNA-1 and RNA-2 of CPMV have extensive
sequence homology. The first 50 nt in the 5' leader sequence show 86% homology which can fold into a similar stem-loop structure as the 65 nt at the 3' untranslated region upstream of the poly(A) (Eggen et al., 1989a). Exchange of the 5' terminal 44 nt of RNA-1 with that of RNA-2 does not affect viral infectivity (Van Bokhoven et al., 1993). Further experiments will be needed to test if the replication complex binds to these structures. These results are consistent with the suggestion that secondary structures of noncoding regions of virus RNA but not their nucleotide sequence are important for RNA replication.

2.6 Host Factors and Cellular Membranes Involved in Viral RNA Replication

Viral RNA replication requires not only viral proteins and cis-acting viral sequences, but also cellular proteins and cellular membranes. Many of these proteins are normal components of cellular RNA processing or translation machineries, which are subverted to play an integral or regulatory role in viral RNA replication. In fact it is well known that infectivity and replication of many viruses and their mutants are species-, variety- and even cell type-dependent. Specific species, variety and cell type may provide specific proteins (host factors) which are required for viral RNA replication. Once host cells are compatible to a certain virus, the virus initiates translation and replication. The virus will induce proliferation or reorganization of cellular membranes and vesicles, possibly as a mechanism to increase the available surface area for RNA synthesis. Increasing evidence suggests that viral proteins, host factors and RNA templates are assembled on cellular membrane-derived structures to form the replication complex which is responsible for the replication.

2.6.1 Host proteins involved in viral RNA replication

Cellular proteins involved in viral RNA replication can be divided into two groups, virus-RdRp binding proteins and virus-RNA binding proteins (Lai, 1998). The first group includes the cellular proteins which bind to the viral RdRp and are usually present in purified RNA replication complexes. The second group contains the cellular proteins which directly bind to the RNA template
to direct the RdRp holoenzyme to the template.

**Group I: virus-RdRp binding protein** A well-characterized eukaryotic protein, which is present in the RNA replication complex isolated from BMV-infected cells, is the eukaryotic translation initiation factor eIF-3 (Quadt et al., 1993). The factor is tightly bound to the viral 2a protein (RdRp), and the addition of the exogenous eIF-3 stimulates the RdRp activity of the complex. The known functions of eIF-3 in protein synthesis include binding mRNA to the 40S preinitiation complex and stabilization of Met-tRNA binding to the 40S ribosomal subunit. Although the exact role of eIF-3 in the replication complex has not been defined, these results suggest a correlation between translation and replication (Quadt et al., 1993). The eIF-3 factor is also a component of the replication complex purified from tobacco mosaic virus (TMV)-infected cells (Osman and Buck, 1997). This replication complex is able to initiate RNA synthesis on templates containing the 3' terminal sequences of the TMV positive or negative strands (Osman and Buck, 1996). TMV and BMV are members of alpha-like viruses and both have a cap at the 5' end and a tRNA-like structure at the 3' end of their genomic RNA molecules. So, the association with translation factors may represent a general feature of RNA replication by the alpha-like viruses.

A translational initiation factor, eIF 4E was also found to interact with the N-terminal part (the VPg domain) of the NIa protein of turnip mosaic potyvirus, a member of the picorna-like supergroup (Wittmann et al., 1997). As mentioned earlier, the NIa protein is a component of the replication complex. Therefore, this interaction may be essential for RNA replication or alternatively for translation of the viral polyprotein. In poliovirus-infected cells, a complex of a cellular protein and 3CD has been found to bind to the cloverleaf structure at the 5' end of the genomic RNA (Andino et al., 1990 and 1993). The cellular protein is identified as an N-terminal fragment of EF-1α (a translation elongation factor) (Harris et al., 1994). But the significance of this complex is not clear since the 3AB-3CD complex also binds the cloverleaf structure. In addition to translation factors, another host protein, Sam68 has been found to be associated with the viral 3D protein (McBride et
Sam68 is a nuclear protein and is involved in cellular mitosis. In the poliovirus-infected cells, this protein interacts with 3D and then is relocalized from the nucleus to the cytoplasm where poliovirus replication occurs. In spite of lack of direct evidence, these results strongly suggest that Sam68 is a host protein with an unknown but probably important functional role in poliovirus replication.

**Group II: virus-RNA binding protein** A number of host proteins have been reported to bind to 3' or 5' untranslational regions of viral genomes (Lai, 1998). Some of them such as La antigen, PTB and EF-1α can bind several different, unrelated viral RNAs. This non-specificity could be explained in two ways. The host proteins may bind a variety of RNAs through specific interactions with other viral or host proteins present in the replication complex. Alternatively, these apparently unrelated viral RNAs may form similar unknown secondary structures which are recognized by the host factors.

Some host proteins in this group can bind multiple sites of viral genomic RNA to form different RNPs which may play different roles during the replication cycle. In poliovirus, poly(rC)-binding protein (hnRNP E) has been found to be able to bind both the 5' end secondary structure formed by nt 120 to 640 and defined as the internal ribosome entry site (IRES) (Blyn *et al*., 1996), and the 5' terminal cloverleaf structure formed by the 5' terminal 88 nt (Garmarinik and Andino, 1997; Parsley *et al*., 1997). Binding to IRES may affect viral RNA translation while binding to the cloverleaf structure may alter the RNA conformation to facilitate the formation of the replication complex for initiation of RNA synthesis (Lai, 1998).

### 2.6.2 Cellular membranes involved in RNA replication

Although there is no direct evidence that the vesicularization of intracellular membranes is required for efficient genomic replication, cellular membranes always appear to be essential for the replication of poliovirus RNA *in vitro* in a coupled translation and replication system (Molla *et al*., 1991 and 1994; Barton, 1995). Infection of cells with poliovirus results in the rearrangement of the
intracellular membranous organelles of the secretory system (Golgi apparatus and endoplasmic reticulum) and the formation of vesicular structures that accumulate in the cytoplasm (Bienz et al., 1983, 1987 and 1992; Troxler et al., 1992). Lysosomes, trans-Golgi, and the trans-Golgi network also contribute to the virus-induced membranous structures (Schlegel et al., 1996). The perinuclear region of the cell becomes crowded with membranous vesicles of heterogeneous sizes. These virus-induced membranous structures are bound by a double lipid bilayer (Schlegel et al., 1996). Continuous synthesis of lipid appears to be required for poliovirus replication (Maynell et al., 1992). Inhibition of lipid biosynthesis or cellular secretory pathway inhibits poliovirus replication, perhaps by preventing formation of these membranous structures (Maynell et al., 1992; Cuconati et al., 1998a). Viral protein 2C or 2BC has been shown to be responsible for the formation of the membranous structures but not for an increase in lipid synthesis (Cho et al., 1994; Barco and Carrasco, 1995; Aldabe et al., 1996). Viral proteins 2B or 3A can inhibit cellular protein secretion (Doesdens and Kirkegaard, 1995). 2B also has been reported to be associated with permeabilization of the plasma membrane (Van Kuppeveld et al., 1997b) and disassembly of the Golgi apparatus (Sandoval and Carrasco, 1997). 2C or 2BC may attach to cellular membranes via the hydrophobic side of an amphipathic helix of 2C (Paul et al., 1994), and 2B and 3A use their hydrophobic domains to bind to membranes (Data and Dasgupta, 1994; Van Kuppeveld et al., 1997a and 1997b). The premature replication complex containing 3CD has been suggested to be located on the surface of these membrane structures (Bienz et al., 1990; Troxler et al., 1992). Viral proteins 2C and its precursors 2BC and P2 (2ABC) are found to be exclusively associated with the complex in the infected cells, and therefore may be directly responsible for its organization or activation (Bienz et al., 1990).

Similar membranous vesicles have been observed in plant nepovirus- and comovirus-infected cells (Wellink et al., 1988; Piazzolla et al., 1985). Expression of the CPMV RNA-1-encoded P1 polyprotein (containing the NTB protein) or the NTB-VPg precursor protein induced the
proliferation of membranous structures in insect cells (Van Bokhoven et al., 1992). Other viral proteins which did not contain the NTB protein were not able to induce the membrane proliferation. As mentioned earlier in this chapter, the NTB protein is equivalent to poliovirus 2C and 3A and may induce membrane vesicles by a similar mechanism.

Characteristic cytopathological structures in the cytoplasm of potyvirus-infected plant cells include pinwheels formed by the CI protein, invaginations in the nuclear membrane and a series of discrete aggregated membranous structures derived from the endoplasmic reticulum (Lesemann, 1988; Restrepo-Hartwig and Carrington, 1994; Schaad et al., 1997). Viral replication complex in infected cells are associated with the endoplasmic reticulum-like membranes. The 6k (6k2) protein can associate membranes as an integral protein via a central 19 amino acid hydrophobic domain. It has also been suggested to induce membrane proliferations in the periphery of the nucleus (Restrepo-Hartwig and Carrington, 1994; Schaad et al., 1997). The 6k protein is equivalent to the 3A protein of poliovirus and C-terminal portion of the NTB protein of CPMV, and accordingly, may possess corresponding functions.

Membrane localization of replication complexes in alpha-like viruses and viruses of other supergroups shows diversities. For example, the chloroplast outer membranes have been suggested to be the RNA replication site of AMV (De Graaf et al., 1993). Replication of CMV is believed to take place on the vacuolar membranes (tonoplast) (Hatta and Francki, 1981). Synthesis of alphavirus genomes such as sindbis occurs on the surface of membranous structures called type I cytopathic vacuoles derived from endosomes and lysosomes (Froshauer et al., 1988). The mechanism used by different viruses to localize their replication complexes in different regions of cells and on different membranous structures which are derived from different cellular membranes remains unknown.

2.7 Summary

During the last decade, significant progress has been made towards understanding virus
replication mechanisms. All ss(+) RNA viruses are most likely to establish a viral factory (the replication complex), containing viral RdRp as a core, other viral proteins, viral RNA, cellular proteins and other cellular factors for genomic replication. Replication complexes are located on cellular membranes, which may function not just as a site for replication, but also have a central role in the organization, activation and functioning of the replication complex. Similar schemes have emerged in the replication of related viruses within a supergroup. This can be used to extrapolate some of the replication strategies employed by less characterized viruses. However, the details of the replication mechanisms employed by ss(+) viruses, even within supergroups, families and groups may be considerably divergent. Several key questions have yet to be answered. For example, it is not clear whether the replication complex assembles on the RNA template and cellular membranes during translation or after translation, how the replication complex is modified to bind to the 3' end of positive and negative RNA templates to initiate transcription and replication, and how viruses regulate the copy numbers of the negative and positive RNA. Elucidating the nature, function, organization and interaction of various participants in RNA transcription and replication remains a challenging frontier.
CHAPTER 3

GENOMIC ORGANIZATION OF TOMATO RINGSPOT NEPOVIRUS RNA-1: PROTEOLYTIC PROCESSING OF THE CARBOXY-TERMINUS OF THE POLYPROTEIN
3.1 Abstract

Tomato ringspot nepovirus (TomRSV) RNA-1 encodes polyprotein P1. The C-terminal portion of the polyprotein contains a putative NTP-binding protein (NTB), a putative VPg, a putative RNA-dependent RNA polymerase (Pol) and a serine-like protease (Pro), which have been suggested to be involved in viral RNA replication. Proteolytic processing of protease precursors containing these proteins was studied in *Escherichia coli* and in vitro. The TomRSV protease cleaved the precursor proteins producing the predicted mature proteins or intermediate precursors. Processing was more efficient as the temperature was lowered from 37 to 20°C during induction of protein expression in *E. coli*. The optimal catalytic temperature in vitro was 15°C. Although processing was detected at all three predicted cleavage sites (NTB-VPg, VPg-Pro and Pro-Pol) in *E. coli*, processing at the VPg-Pro cleavage site was inefficient in *E. coli* and not detectable in vitro, resulting in accumulation of the VPg-Pro intermediate precursor. In addition, the presence of the VPg sequence in the precursor resulted in increased cleavage at the Pro-Pol cleavage site in *E. coli* and in vitro. The sequence of the NTB-VPg, VPg-Pro and Pro-Pol cleavage sites was determined, allowing the localization of the putative VPg and of the protease on the RNA-1 encoded polyprotein.

3.2 Introduction

Tomato ringspot nepovirus (TomRSV) RNA-1 is 8214 nucleotides in length, excluding the 3' poly(A) tail and contains a single long open reading frame of 6591 nucleotides beginning at the first AUG codon at nucleotide position 78 (Rott et al., 1995). This would give rise to a polyprotein (P1) with a predicted molecular mass of 244 kDa. Based on amino acid sequence comparisons between the C-terminal portion of the P1 polyprotein and proteins encoded by TomRSV-related picorna-like viruses, it has been predicted that the TomRSV RNA-1 encodes an NTP binding protein (NTB), a VPg, a protease (Pro) and an RNA dependent RNA polymerase (Pol) (Rott et al., 1995),
which are likely involved in RNA replication (Viry et al., 1993). Of these P1 proteins, the TomRSV protease domain has been found to be related to the 3C protease of picornaviruses and to be responsible for cleavage of precursor polyproteins encoded by RNA-2 (Hans and Sanfaçon, 1995). Substitution of the amino acid H in the predicted catalytic triad or the amino acid H, the base component of the predicted substrate binding pocket abolishes the protease activity on the RNA-2-encoded precursor protein (Hans and Sanfaçon, 1995). However, it is not known if this protease is also responsible for processing of the RNA-1-encoded P1 protein including the NTB-VPg-Pro-Pol precursor.

Mainly by means of in vitro analyses, proteolytic processing of precursor proteins by viral protease has been extensively studied for many members of the picorna-like superfamily such as poliovirus, potyviruses and comoviruses but to a lesser extent nepoviruses (Chapter 2). Proteolytic activity of 3C-like proteases has been observed in E. coli for proteases from poliovirus (Ivanoff et al., 1986; Baum et al., 1991), hepatitis A virus (Gauss-Müller et al., 1991; Harmon et al., 1992; Schultheiss et al., 1994), potyvirus (Laliberté et al., 1992; Kim et al., 1995; Ménard et al., 1995; Parks et al., 1995) and comovirus (Garcia et al., 1987). Proteolytic activity of nepovirus proteases in E. coli has not been reported. Recently, the turnip mosaic potyvirus protease has been shown to have a low temperature optimum in vitro (15°C) (Kim et al., 1996). In contrast, the optimal temperature for an animal virus-encoded 3C protease (Mengo virus) is 30-37°C (Hall and Palmenberg, 1996). The effects of temperature on the proteolytic activity on the comovirus- and nepovirus-encoded proteases have not been established.

In this chapter, we report that the TomRSV protease can process the C-terminal portion of P1 in E. coli and in vitro. The efficiency of the proteolytic processing is temperature dependent. The optimum temperature for the proteolytic processing is determined. The sequences of the NTB-VPg, VPg-Pro and Pro-Pol cleavage sites are suggested.
3.3 Materials and Methods

3.3.1 Plasmid constructions

Plasmid pMR10 containing the cDNA clone for the entire RNA-1 has been described (Rott et al., 1995). Plasmid pTrxFusNTB-VPg-Pro-Pol (Fig. 3.1) containing the entire coding region for NTB, VPg, Pro and Pol fused at its N-terminus to the thioredoxin gene, was constructed by amplifying a 4752 nt fragment of pMR10 (TomRSV RNA-1 nt 1932-6678) with Taq polymerase and Pfu polymerase (1:1) (Stratagene). Primers W07 (Table 3.1) and W08 were used for amplification. The amplified fragment was digested with Xba I and Kpn I and then inserted into the corresponding sites of the plasmid pTrxFus polylinker (Invitrogen).

Table 3.1 Primers used for construction of plasmids

<table>
<thead>
<tr>
<th>Name</th>
<th>Polarity</th>
<th>Sequence (5' to 3')</th>
<th>Corresponding sequence of RNA-1</th>
<th>Restriction site or mutation introduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>W01</td>
<td>+</td>
<td>CCTCTAGAGTCTAGTTATCAATCTC</td>
<td>4526-4544</td>
<td>Xba I</td>
</tr>
<tr>
<td>W02</td>
<td>-</td>
<td>TGCTAACCTGCAGTCTGCTGCTCAAGAG</td>
<td>6695-6676</td>
<td>Pst I</td>
</tr>
<tr>
<td>W07</td>
<td>+</td>
<td>CGGGGTAACCTCAAGGGCTCACTGAGCTT3</td>
<td>1932-1951</td>
<td>Kpn I</td>
</tr>
<tr>
<td>W08</td>
<td>-</td>
<td>GCTCTAGAGGATTTAGCTGCAAT</td>
<td>6678-6659</td>
<td>Xba I</td>
</tr>
<tr>
<td>W023</td>
<td>-</td>
<td>ACGCGTCCAGCTTGGCGAAGAGCAAG</td>
<td>4534-4518</td>
<td>Sal I</td>
</tr>
<tr>
<td>W030</td>
<td>+</td>
<td>ACGCCCATGTTCCGCCTGTATCGAAG</td>
<td>1959-1977</td>
<td>Nco I</td>
</tr>
<tr>
<td>37</td>
<td>-</td>
<td>GGGTTCCATGGTTCTTCTTGTGCGG</td>
<td>3811-3794</td>
<td>Nco I</td>
</tr>
<tr>
<td>W045</td>
<td>+</td>
<td>TCGACGATTCCCTCGTGAT3</td>
<td>3714-3732</td>
<td></td>
</tr>
<tr>
<td>W046</td>
<td>-</td>
<td>AAGCTCAATTTTCCCGACAG</td>
<td>3710-3691</td>
<td></td>
</tr>
<tr>
<td>W050</td>
<td>+</td>
<td>CATGCGATTCGGATCGACATCCCTCCCG</td>
<td>3712-3730</td>
<td>Nco I</td>
</tr>
<tr>
<td>W051</td>
<td>+</td>
<td>CATGCCATGCTTCTTTTGGCGAAGA</td>
<td>3794-3812</td>
<td>Nco I</td>
</tr>
<tr>
<td>W052</td>
<td>-</td>
<td>ACGCGTCCAGCCTTAAGCGAAGAAGCTTC</td>
<td>5141-5121</td>
<td>Sal I</td>
</tr>
<tr>
<td>K4535</td>
<td>+</td>
<td>TTTGACCTCCTGCGCGTACAGTTATC</td>
<td>4533-4535</td>
<td>CAG-GCG</td>
</tr>
<tr>
<td>K4470</td>
<td>+</td>
<td>GTTGTAGAGTGGCGATCCAGATAC</td>
<td>4470-4472</td>
<td>CAG-GCG</td>
</tr>
</tbody>
</table>

The sequences for engineering restriction site are underlined and for mutagenesis are in bold.

Plasmid pTrxFusPol containing the entire Pol coding region fused at its N-terminus to the thioredoxin gene was constructed by amplifying a 2160 nt fragment encompassing the entire pol domain from clone pMR10 (RNA-1 nt 4526-6695) with Taq polymerase and Pfu polymerase (1:1) (Stratagene). Primers W01 and W02 were used to generate the fragment. The PCR-amplified
Fig. 3.1. **Schematic diagram of the different constructs derived from TomRSV RNA-1 cDNA.** TomRSV RNA-1 is represented as a solid black bar. The VPg is indicated by a circle at the 5'-end and the predicted open reading frame is indicated by the initiation (AUG) and termination (UAA) codons. The corresponding polyprotein P1 is represented underneath as boxes. The putative functions of the mature products and their predicted molecular weights are shown inside or above the corresponding boxes. The cDNA clones containing the different portions of the RNA-1 open reading frame are presented and their names given at the right. The promoter used to drive the transcription of the cDNA fragments in *E. coli* or in cell free system is indicated for each clone (T7, T7 promoter; P_L, P_L promoter from a bacteriophage). Thio (shaded with lines) and VS (shaded with dots) represent thioredoxin protein and vector sequence, respectively.
fragment was digested by *XbaI* and *PstI* and inserted into the corresponding sites of plasmid pTrxFus.

To construct plasmid pT7-NTB-VPg-Pro encompassing the entire NTB, VPg and Pro domains, primers W030 and W023 were used to amplify the corresponding coding region of plasmid pMR10 (RNA-1 nt 1959-4534) with Pfu polymerase. The amplified fragments were digested with *Nco I* and *Sal I* and inserted into the corresponding sites of plasmid pCITE-4a(+) (Novagen). To construct plasmid pT7-NTB-VPg-Pro<sup>H1283D</sup>, the large *Pst I-BamH I* fragment of plasmid pT7-NTB-VPg-Pro was ligated with the small *Pst I-BamH I* fragment of plasmid pET15bVPg-Pro-N-Pol<sup>H1283D</sup> (see below). To obtain the pT7-NTB-VPg-Pro cleavage mutant, primers W045 and W046 were used to amplify plasmid pT7-NTB-VPg-Pro with Pfu polymerase. The amplified fragments were self-ligated to give plasmid pT7-NTB-ΔQ-VPg-Pro. Presence of the deletion was confirmed by sequence analysis.

To obtain clone pET21aMid-NTB containing the middle portion of the NTB domain, an *EcoR I-XhoI* fragment of pMR10 (TomRSV RNA-1 nt 2658-3393), encoding a 31 kDa polypeptide between the highly conserved “A” and “B” sites and the C-terminal hydrophobic stretch of NTB, was ligated into the corresponding sites of vector pET 21a (Novagen).

Plasmids pET15bVPg-Pro-N-Pol, pET15bVPg-Pro<sup>H1283D</sup>-N-Pol (protease mutated in the putative catalytic triad) and pET15bVPg-Pro<sup>H1451L</sup>-N-Pol (protease mutated in the putative substrate binding pocket) containing the coding region for VPg, wild type or mutated Pro and N-terminal Pol (TomRSV RNA-1 nt 3714-4601) were obtained in the following manner. The *Nco I-EcoR I* fragments of plasmids pT7-Pro, pT7-Pro<sup>H1283D</sup> and pT7-Pro<sup>H1451L</sup> (Hans and Sanfaçon, 1995) were first inserted into the corresponding sites of plasmid pMTL23 (Chambers *et al.*, 1988) to allow the introduction of a stop codon. The *Nco I-Nde I* fragments were then extracted from those intermediate plasmids and inserted into the corresponding sites of the pET15b polylinker (Novagen).

To construct clone pET15bPro-N-Pol containing the coding region for Pro and N-terminal
Pol, a modified version of plasmid pET15bVPg-Pro-N-Pol (TomRSV RNA-1 nt 3714-4601) was obtained in which the VPg sequence had been precisely deleted. An 816 nt fragment (TomRSV RNA-1 nt 3794-4601) was amplified from plasmid pT7-Pro with Taq DNA polymerase (Gibco-BRL) using primers 18 (Hans and Sanfaçon, 1995) (containing an EcoRI restriction site) and primers 37. The amplified fragment was digested with Nco I and EcoRI and used to replace the corresponding fragment of plasmid pET15bVPg-Pro-N-Pol.

Plasmids pT7-VPg-Pro-N-Pol-II and pT7-VPg-ProH1283D-N-Pol-II containing the entire VPg and Pro domains and an extended region of the N-terminal portion of Pol were constructed by ligating a 417 nt BamHI-EcoRI fragment from clone pMR10 (TomRSV RNA-1 nt 4436-4852) with the large BamHI-EcoRI fragment of plasmid pT7-Pro and pT7-ProH1283D (Hans and Sanfaçon, 1995). To construct pT7-VPg-Pro-N-Pol-II cleavage site mutants, site directed mutagenesis was performed with mutant oligonucleotides priming synthesis on single-stranded templates of pT7-VPg-Pro-N-Pol-II generated by the M13 bacteriophage and enriched with uridine to allow selection of the mutant strand (Kunkel, 1985). To construct plasmids pT7-VPg-ProQSAS-N-Pol-II and pT7-VPg-ProQMAM-N-Pol-II, mutations were made simultaneously in single in vitro reactions by adding equimolar amounts of the mutagenic oligonucleotide K4535 and K4470 to pT7-VPg-Pro-N-Pol-II single-stranded DNA. Positive clones were identified by dideoxy sequencing of purified DNA (Sanger et al. 1977).

Primers W050 and W052 were used to amplify the coding region of pMR10 for the VPg, Pro and N-terminus of Pol with Pfu polymerase. Plasmid pT7-VPg-Pro-N-Pol-III was obtained by ligation of three fragments: the large Nco I-SalI fragment of vector pCITE-4a(+), a 0.95 kb fragment obtained by digestion of the PCR products with Nco I and HindIII and a 0.5 kb fragment obtained by digestion of the PCR products with HindIII and SalI. A similar strategy was used to obtain plasmid pT7-Pro-N-Pol-III containing Pro and the N-terminal portion of Pol except that primers W052 and W051 were used to amplify the coding region for Pro and the N-terminus of Pol.
To obtain pT7-VPg-Pro\(^{H1283D}\)-N-Pol-III and pT7-Pro\(^{H1283D}\)-N-Pol-III, the large \(Pst\ I-BamH\) I fragments of plasmids pT7-VPg-Pro-N-Pol-III and pT7-Pro-N-Pol-III were ligated with the small \(Pst\ I-BamH\) I fragment of plasmid pET15bVPg-Pro\(^{H1283D}\)-N-Pol.

### 3.3.2 Induction in \textit{E. coli}, cell lysis and purification of fusion proteins

Plasmids pET21aMid-NTB, pET15bPro-N-Pol, pET15bVPg-Pro-N-Pol, pET15bVPg-Pro\(^{H1283D}\)-N-Pol and pET15bVPg-Pro\(^{H1451L}\)-N-Pol were transformed into \textit{E. coli} strain JM109 (DE3) or BL21 (DE3) (Novagen). The bacterial cultures were grown to an OD\(_{550}\) of 0.5 in Luria broth containing 100 \(\mu\)g/ml ampicillin at 37°C unless otherwise stated. Isopropyl thio-\(\beta\)-D-galactoside (IPTG) was then added to a final concentration of 1 mM. The bacteria were harvested by centrifugation (4000 g for 10 min) 3 hr after induction. For protein analysis, the bacteria were resuspended directly in 2X protein loading buffer (0.125 M Tris-HCl, pH 6.8, 20 % glycerol, 4% SDS, 10% \(\beta\)-mercaptoethanol and 0.0025% bromophenol blue). The bacterial suspension was frozen in liquid nitrogen and dissolved at 45°C twice to allow lysis of the bacteria. Before loading onto protein gels, the lysed solution was heated at 95°C for 5 min and centrifuged at 14,000 g for 10 min. For partial protein purification, the bacterial pellet was washed, centrifuged and resuspended in 10 volumes of lysis buffer [50 mM Tris-HCl (pH 8), 100 mM NaCl, 1 mM ethylenediaminetetracetic acid (EDTA)], lysed in a French press and centrifuged at 5000 g for 10 min. All the induced proteins were found predominantly in the insoluble fraction as determined by denaturing sodium dodecyl sulfate-polyacrylamide gels (SDS-PAGE) analysis (Laemmli, 1970) of the pellet and supernatant fractions. The pellet containing the inclusion bodies was further purified by several rounds of washing in the ice cold lysis buffer containing 0.1% sodium deoxycholate and 1% NP-40. To solubilize the proteins, the purified inclusion bodies were resuspended in 20 volumes of the lysis buffer containing 8 M urea. The suspension was stirred at 4°C for 3 hr and centrifuged at 30,000 g for 45 min. To allow protein renaturation, the soluble fraction was dialysed against successively lower concentrations of urea (4 M, 2 M, 1 M, 0.5 M and 0.25 M) in 50 mM Tris-HCl (pH 8) for 4
hr each time and finally against 50 mM Tris-HCl (pH 8) overnight at 4°C.

Plasmids pTrxFusPol and pTrxFusNTB-VPg-Pro-Pol were transformed into *E. coli* strain GI 724 (Invitrogen). The bacteria harboring the plasmids were grown in RM medium (Invitrogen) containing 100 µg/ml ampicillin at 20°C or 30°C and shaken at 225 rpm overnight. One volume of the culture was used to inoculate 20 volumes of induction medium (Invitrogen). These cultures were grown to an OD$_{550}$ of 0.5, then L-tryptophan was added to a final concentration of 100 µg/ml. The cultures were grown at 20°C (for processing studies) or transferred to 37°C (for protein purification) for 3 hr to allow expression. The bacteria containing plasmid pTrxFusNTB-VPg-Pro-Pol were lysed for protein analysis as above while the bacteria containing plasmid pTrxFusPol were used for further protein purification. Purification of the inclusion bodies was essentially as described above. The overexpressed chimeric protein in the inclusion bodies was washed three times with lysis buffer plus 1% NP-40 and further purified by SDS-PAGE. The Pol fusion protein was extracted from the gel by electro-elution. The eluted protein was dialysed overnight against phosphate-buffered saline (PBS: 136.9 mM NaCl, 2.69 mM KCl, 10.14 mM Na$_2$HPO$_4$ and 1.76 mM KH$_2$PO$_4$, pH 7.4).

### 3.3.3 Chemical crosslinking of synthetic VPg peptide to ovalbumin

A 16-mer peptide composed of amino acids (SGSYADVYNARNMTRV) was obtained from Peptide Synthesis Laboratory, the University of British Columbia. This peptide corresponds to a part of VPg predicted to be immunogenic by the method of Hopp and Woods (1981). Six mg of lyophilized chicken egg albumin (Sigma) and 6 mg of the synthetic peptide were dissolved in 600 µl and 1500 µl of conjugation buffer (0.1 M MES (2-[N-morpholino]ethane sulphonic acid, pH 4.8)), respectively. Thirty mg of EDC (1-ethyl-3-[3-dimethylaminopropyl]carbodimide hydrochloride) was added to the mixture of peptide and carrier protein by gentle mixing. The mixed solution was allowed to react for 2 hr at room temperature, and then was dialysed against PBS buffer at 4°C overnight.
3.3.4 Antisera production

To produce monoclonal antibodies against Pro (anti-Pro antibodies), three Balb/c mice were immunized by subcutaneous injection using the purified fusion protein (60 μg) obtained from clone pET15bPro-N-Pol, which was emulsified with an equal volume of Freund’s incomplete adjuvant. After resting for 30 days, each mouse was boosted with an additional 60 μg of protein intraperitoneally. Two weeks after the second injection, the mice were tail-bleed and blood samples were tested in an ELISA. The mouse with the strongest ELISA reaction was used for subsequent monoclonal antibody production. The fusion protocol, subsequent screening techniques and ascites fluid production were essentially as described (Harlow and Lane, 1988; Wieczorek and Sanfaçon, 1993). All three cell lines of monoclonal antibodies were subclassed into IgG1 isotypes using a Hybridoma Sub-Isotyping Kit (Calbiochem Immunochemicals).

Immunization of the mice with gel-purified Pol fusion protein (Thio-Pol) and production of the polyclonal antibodies against Pol (anti-Pol antibodies) were as described above. To prevent immunoreaction of the antibodies with the thioredoxin fusion proteins in Western blots, 10 μg IgG was preabsorbed with 500 μg of thioredoxin protein which was induced from clone pTrxFus (Invitrogen) and purified by affinity chromatography using ThioBond Resin (Invitrogen).

The monoclonal antibody against thioredoxin (anti-Thio antibodies) was supplied by Invitrogen.

To raise polyclonal antibodies against NTB (anti-NTB antibodies), one mg of the fusion protein obtained from clone pET21aMID-NTB was emulsified with an equal volume of Freund’s incomplete adjuvant (Gibco-BRL) and injected into a rabbit intramuscularly. After three injections, antisera were harvested and IgG was purified according to the protocol of Clark and Adams (1977).

To raise polyclonal antibodies against VPg (anti-VPg antibodies), one mg of ovalbumin crosslinked synthetic VPg peptides emulsified with Freund’s incomplete adjuvant was injected into a rabbit intramuscularly. The rabbit was injected three times at 4-week intervals with the same dose.
Antisera were collected two weeks after the last injection. IgG was purified as above. To improve its specificity, the purified IgG of the antisera was further purified by affinity chromatography using a carbohydrate/glycoprotein gel covalently linked with ovalbumin (E Y Laboratories, Inc.).

3.3.5 SDS-PAGE, gel-staining and western blots

Proteins were separated by SDS-PAGE and visualized by Coomassie brilliant blue R-250 (Sigma) staining (Harlow and Lane, 1988).

For protein blots, the proteins separated by SDS-PAGE were transferred to polyvinylidene fluoride (PVDF) membranes (Bio-RAD) using a Bio-RAD miniblotter (125 V for 2 hr or 80 V overnight plus 125 V for 1 hr at 4°C in a Tris/MeOH/glycine buffer (25 mM Tris, 192 mM glycine and 20% methanol at pH 8.3). The membranes were incubated in blocking solution [3% bovine serum albumin in TBS (20 mM Tris-HCl, 500 mM NaCl, pH 7.5)] with gentle agitation for 30 min, and then washed twice with TBS for 5 min each. After incubation with the first antibodies in incubation buffer (TBS plus 1% bovine serum albumin and 0.05% Tween-20) for 1 hr, the membranes were washed with washing buffer (TBS plus 0.05% Tween-20) for 5 min three times. Secondary antibodies conjugated with alkaline phosphatase (goat anti-mouse IgG or goat anti-rabbit IgG, Sigma) were incubated with the membranes in incubation buffer for 1 hr, and washed three times with washing buffer and twice with alkaline phosphatase buffer (AP buffer, 100 mM Tris, 100 mM NaCl, 50 mM MgCl₂, pH 9.5). The antibodies recognizing protein bands were visualized with enzyme substrates, BCIP (bromochloroindolyl phosphate, Gibco-BRL) and NBT (nitro blue tetrazolium, Gibco-BRL) in AP buffer. The reaction was stopped by addition of stop buffer (20 mM Tris-HCl, 5 mM EDTA, pH 7.5).

3.3.6 Microsequencing of the N-terminal amino acids of the protease

Expression products of plasmid pET15bVPg-Pro-N-Pol at 25°C were separated by SDS-PAGE and blotted onto a PVDF sequencing membrane. Proteins were identified by Western blots using anti-VPg polyclonal antibodies and anti-Pro monoclonal antibodies, or staining with 0.01%
Coomassie brilliant blue dissolved in 50% methanol. The stained band corresponding to the position of the band recognized only by anti-Pro monoclonal antibodies but not by anti-VPg polyclonal antibodies was excised and subjected to N-terminal amino acid sequencing (Protein Microsequencing Laboratory, Department of Biochemistry and Microbiology, University of Victoria, Victoria, Canada).

3.3.7 In vitro transcription and translation and quantitative measurements

In vitro transcription and translation of the cDNA clones were performed using a TNT coupled transcription translation system (Promega). Labelled translation products were obtained by adding 1 μl of purified plasmid DNA (1 μg/μl) to 8 μl of TNT coupled transcription translation mixture containing ³⁵S-methionine. After 1 hr incubation at 30°C, the translation was arrested by the addition of 10 μl of Tris buffer (0.1 M, pH 8) containing 5 units RNase (Promega), 2 units DNase (Promega) and excess non-radioactive L-methionine (20 mM, Sigma). Aliquots of the mixture were added to an equal volume of 2X protein sample buffer either immediately or after incubation at 16°C for 20 hr unless otherwise stated. Samples were then separated by SDS-PAGE and visualized by autoradiography of dried gels.

For quantitative measurements, the amount of radioactivity in the precursor and the cleaved products were measured using a phosphorimager (Storm 860, Molecular Dynamics). The concentration of precursor and products were calculated taking into account the number of methionines present in each protein. Protease activity was calculated as the slope of percentage of conversion of the precursor to the product plotted against the time of incubation.

3.4 Results

3.4.1 Production of monoclonal and polyclonal antibodies against viral proteins

To study the proteolytic processing of the C-terminal portion of the P1 polyprotein, a series of antibodies were raised against the NTB, VPg, Pro and Pol proteins. The middle part of NTB and
the entire Pro regions were inserted into bacterial expression vectors pET21a and pET15b, respectively (Fig. 3.1). The entire Pol region was fused in frame with the thioredoxin protein of the vector pTrxFus (13.5 kDa) (Fig. 3.1). Following expression of the fusion proteins in *E. coli*, prominent polypeptides of an apparent molecular mass corresponding to the expected size for the middle portion of NTB (31 kDa), for Pro (32 kDa) and for the thioredoxin-polymerase fusion protein (Thio-Pol) (93.5 kDa) were detected (Fig. 3.2 a, b and c, lanes 2 and 3). All three chimeric proteins were found predominantly in the insoluble fraction (Fig. 3.2 a, b and c, lanes 4 and 5). After purification and resolubilization of inclusion bodies, about 95% of the resolubilized proteins were the fusion proteins (Fig. 3.2 a, b and c, lane 6). These preparations were used for the production of monoclonal and polyclonal antibodies. Three monoclonal antibody (mAb) cell lines directed against the purified Pro were obtained. The mAb 3A10 which had the strongest immunoreaction with Pro (Fig. 3.2 a, lane 7) was used in subsequent experiments. The epitope recognized by mAb 3A10 was mapped to the middle part of Pro (results not shown). Rabbit anti-NTB polyclonal antibodies and mouse anti-Pol polyclonal antibodies were obtained that recognized the corresponding fusion proteins in Western immunoblots (Fig. 3.2 b and c, lane 7).

Rabbit anti-VPg polyclonal antibodies were produced against a synthetic peptide cross-linked with ovalbumin. The peptide (SGSYADVYNARNMTRV) corresponds to the middle part of the putative VPg. The purified IgG recognized VPg-Pro, the peptide cross-linked ovalbumin and ovalbumin (data not shown). To remove the antibody against ovalbumin, the purified IgG was further purified by affinity chromatography using an ovalbumin linked carbohydrate/glycoprotein gel. The purified antisera reacted strongly with the VPg-Pro and the synthetic peptide cross-linked with ovalbumin, but did not detect the ovalbumin alone (Fig. 3.2 d, lanes 5, 6 and 7).

### 3.4.2 Proteolytic processing of protease precursors by the TomRSV protease in *E. coli* and *in vitro*

To determine if the protease activity could be detected in *E. coli*, a small polyprotein
Fig. 3.2. Production of monoclonal and polyclonal antibodies against fusion proteins.
(a) Expression of the viral protease from clone pET15bPro-N-Pol. Proteins were separated on 12% polyacrylamide-SDS gels and stained with Coomassie brilliant blue R-250 or transferred onto PVDF membranes for western detection. Mr: molecular weight markers, Unind. and Ind.: extracts from uninduced and induced E. coli cells respectively, Sol. and Insol.: soluble and insoluble fractions from induced E. coli cell extracts respectively, Pur.: purified protein resolubilized from inclusion bodies, Western: immunodetection of the fusion protein by the monoclonal antibodies against the protease. (b) Expression of the middle portion of NTB from clone pET21aMid-NTB. Proteins were analysed as in (a). Lanes 1 to 6 as in (a), Western: immunodetection of the fusion protein by the polyclonal antibodies against the middle portion of NTB. (c) Expression of the putative polymerase from clone pTrxFusPol. Proteins were separated on 7.5% polyacrylamide-SDS gels. Lanes 1 to 5 as in (a) and (b), Pur: fusion protein purified by gel elution, Western: immunodetection of the fusion protein by the polyclonal antibodies against pol. (d) Test of polyclonal antibodies against VPg. Proteins were separated on 15% polyacrylamide-SDS gels and blotted onto PVDF membranes. The membrane was subjected to staining as in (a), (b) and (c) or immunodetection. Mr: molecular weight markers, ova: ovalbumin, p-ova: VPg derived peptide crosslinked with ovalbumin, VPg-Pro: fusion protein expressed from clone pET15bVPg-Pro-N-Pol containing the VPg domain. Lanes 1 to 4: staining of the blotted proteins. Lanes 5 to 7: immunodetection of the blotted proteins by the purified polyclonal antibodies against VPg.
precursor was designed that contains the VPg, the entire Pro and the N-terminal portion of Pol. A cDNA fragment containing the corresponding coding region was inserted into an expression vector to give plasmid pET15bVPg-Pro-N-Pol (Fig. 3.1). Production of a fusion protein with a calculated size of 34.5 kDa was predicted upon induction of bacteria containing this plasmid (Fig. 3.3 a). Indeed, after 3 hr induction at 30°C a protein with an apparent molecular mass of 34.5 kDa was detected in extracts from induced cells (Fig. 3.3 b, lanes 1 and 2). However, an additional protein with an apparent molecular mass of 30 kDa was also observed in the extracts. To determine the nature of these proteins, Western immunoblotting was conducted with anti-VPg polyclonal antibodies and anti-Pro monoclonal antibodies. As expected, the 34.5 kDa protein was recognized by both anti-VPg and anti-Pro antibodies, confirming that it was the precursor containing VPg (3 kDa), Pro (27 kDa) and likely also N-terminal pol (4.5 kDa) (Fig. 3.3 b, lanes 5 and 8). The 30 kDa protein was also recognized by both anti-VPg polyclonal antibodies and anti-Pro monoclonal antibodies and corresponded to the predicted size for the VPg-Pro, suggesting it was the VPg-Pro precursor resulting from cleavage between Pro and N-terminal Pol. A small amount of an additional protein with an apparent molecular mass of 27 kDa was detected by anti-Pro monoclonal antibodies but not by anti-VPg antibodies. This protein is of the expected molecular mass for the mature Pro, suggesting that it was the result of further proteolytic processing between VPg and Pro. Microsequencing the N-terminus of the 27 kDa protein supports this conclusion (see below). The predicted Pro-N-Pol precursor (31.5 kDa) which could have arisen by cleavage of the VPg-Pro-N-Pol at the VPg-Pro cleavage site was not detected by anti-Pro antibodies. To confirm that this processing resulted from the TomRSV protease rather than the proteolytic activity from E. coli, a mutation was introduced into clone pET15bVPg-Pro-N-Pol. In clone pET15bVPg-Pro\textsuperscript{H1283D}-N-Pol, His\textsuperscript{1283}, which is postulated to be the base component of the catalytic triad, was replaced with Asp. This mutation was previously found to inactivate the protease activity on P2 cleavage sites \textit{in vitro} (Hans and Sanfaçon, 1995). After induction of the bacteria containing this plasmid, anti-VPg
Fig. 3.3. Processing of the VPg-Pro-N-Pol precursor polypeptide in *E. coli*. (a) Schematic representation of processing of the VPg-Pro-N-Pol precursor protein. The expected molecular mass of precursor polyproteins and predicted mature proteins are shown. Cleavage sites present on the precursor proteins are indicated by the vertical lines. (b) Detection of precursors and processed products in *E. coli* extracts. The bacteria containing clones pET15bVPg-Pro-N-Pol (wt: wild type protease) and pET15bVPg-Pro<sup>H1283D</sup>-N-Pol (H-D: protease mutated in the putative catalytic triad) were induced for 3 hr at 30°C and lysed in the protein loading buffer. The denatured proteins were separated by 12% polyacrylamide-SDS gels, and then stained with Coomassie brilliant blue R-250 or electro-blotted onto PVDF membranes for western detection with the monoclonal antibodies against protease (anti-Pro) and the polyclonal antibodies against VPg (anti-VPg). The expected position for the different precursors and mature proteins is indicated on the right side of the gels. The position of molecular weight markers is indicated on the left side of the gels. Lanes 1 to 4: Commissie blue staining. Ind: extracts from induced bacterial cells, Unind: extracts from uninduced bacterial cells. Lanes 5 to 9: Immunoreaction. Anti-Pro Abs: immunoreacted with monoclonal antibodies against the protease, Anti-VPg Abs: immunoreacted with polyclonal antibodies against VPg, Wt: containing wild type protease, H-D: containing mutated protease, -: control (bacterial cells containing pET15b vector).
polyclonal antibodies and anti-Pro monoclonal antibodies detected only the 34.5 kDa precursor protein in the protein extracts (Fig. 3.3b, lanes 6 and 9). Therefore, processing only occurred in the Pro precursor which contained the wild-type Pro and did not occur in the precursor with the mutated Pro. These results suggest that the VPg-Pro-N-Pol precursor is processed by the TomRSV protease in E. coli.

To study if the protease could process P1 precursor proteins in vitro, another P1 polyprotein precursor was designed that contained the entire NTB, the VPg and the entire Pro. A cDNA fragment containing the corresponding coding region was inserted into vector pCITE-4a(+) to obtain plasmid pT7-NTB-VPg-Pro (Fig. 3.1). As expected, a precursor protein with an apparent molecular mass of 99 kDa corresponding to the predicted size for the NTB-VPg-Pro-VS precursor protein [66 kDa NTB, 3 kDa VPg, 27 kDa Pro and 3 kDa VS (VS: vector sequence)] was generated by the cell-free translation system (Fig. 3.4a and b, lane 1). The translation products were incubated at 16°C for 20 hr to allow processing. After incubation, the amount of the 99 kDa precursor protein decreased while the amount of two smaller proteins with apparent molecular mass 66 kDa and 33 kDa increased (Fig. 3.4a and b, lanes 1 and 2). These two proteins correspond to the predicted sizes for the mature NTB (66 kDa) and VPg-Pro-VS (3 kDa VPg, 27 kDa Pro and 3 kDa VS), respectively. Indeed, the 66 kDa protein could be immunoprecipitated with anti-NTB antibodies and the 33 kDa could be immunoprecipitated with anti-VPg and anti-Pro antibodies (data not shown). To eliminate the possibility that the 66 kDa and 33 kDa proteins were generated by degradation, the translation products of plasmid pT7-NTB-VPg-Pro\textsuperscript{H1283D}, in which the protease was mutated by substitution of a His with an Asp at position 1283, were used as a control. After incubation at 16°C, the 66 kDa and 33 kDa proteins were not detected (Fig. 3.4b, lanes 3 and 4). In addition, mutation of the predicted NTB-VPg cleavage site eliminated the processing (see below), suggesting that the 66 kDa and 33 kDa proteins are likely to be produced by cleavage of the NTB-VPg-Pro precursor at the NTB-VPg cleavage site. Taken together these results suggest that the protease can process a P1 precursor
Fig. 3.4. Processing of the NTB-VPg-Pro precursor polypeptide and mutation of the predicted NTB-VPg cleavage site *in vitro*. (a) Schematic representation of processing of the NTB-VPg-Pro(VS) precursor protein. The expected molecular mass of precursor polypeptide and cleaved proteins is shown. Cleavage sites are indicated by short vertical lines. The predicted dipeptide cleavage site between the putative NTP-binding protein and the putative VPg is shown (VS: vector sequence). (b) Auto-cleavage of the NTB-VPg-Pro(VS) precursor and mutation of the predicted NTB-VPg cleavage site. [S$^{35}$]methionine-labelled *in vitro* translation products of plasmids pT7-NTB-VPg-Pro (wt: containing the wild type protease), pT7-NTB-VPg-Pro$^{H1283D}$ (H-D: containing the inactive protease), pT7-NTB-$\Delta Q$-VPg-Pro($\Delta Q$: containing a deletion at the Q$^{1212}$ at the predicted NTB-VPg cleavage site) were incubated for 20 hr at 16°C post-translation (hpt: hours post-translation), separated on 10% polyacrylamide-SDS gels and visualized as described in Material and Methods. The positions of uncleaved NTB-VPg-Pro(VS) and the Pro cleavage products are indicated. The position of molecular weight markers is indicated on the left side of the gel.
protein in vitro. The predicted NTB-VPg precursor (69 kDa) and the mature Pro (Pro-VS, 30 kDa) which would have resulted from cleavage of the precursor at the VPg-Pro cleavage site were not detected upon incubation at 16°C of the translation products from either the wild type clone or from the mutated clone in which the predicted NTB-VPg cleavage site was deleted.

3.4.3 Temperature dependence of proteolytic processing in *E. coli* and in vitro

To determine the effects of temperature on processing of protease precursors in *E. coli*, bacteria containing plasmid pET15bVPg-Pro-N-Pol (Fig. 3.1) were induced at 37°C, 25°C and 20°C. Western immunoblotting experiments showed that at 37°C, the 34.5 kDa protein (VPg-Pro-N-Pol) was the main protein detected, suggesting that very little cleavage occurred at the two cleavage sites (Fig. 3.5, lanes 1 and 14). As shown above (Fig. 3.3), the 34.5 kDa (VPg-Pro-N-Pol) and the 30 kDa (VPg-Pro) proteins were the main proteins detected at 30°C, indicating that the 34.5 kDa precursor was processed at the Pro-Pol cleavage site. Small amounts of the 27 kDa protein (Pro) were detected, indicating that some processing occurred at the VPg-Pro site. At 25°C, anti-Pro antibodies detected three main proteins, the 34.5 kDa precursor, the 30 kDa VPg-Pro and the 27 kDa mature Pro, suggesting that the VPg-Pro (30 kDa) was further cleaved to release the 27 kDa mature Pro (Fig. 3.5, lanes 5 and 18). At 20°C, the 34.5 kDa precursor was not detected, but the 30 kDa (VPg-Pro) and the 27 kDa (Pro) proteins were predominant, suggesting that the precursors were completely processed into the 30 kDa VPg-Pro and the 27 kDa Pro (Fig. 3.5, lanes 10 and 23). Thus, low temperatures favours the processing at both the VPg-Pro cleavage site and the Pro-Pol cleavage site. The predicted Pro-Pol precursor (31.5 kDa) could not be detected at any of the temperatures tested. Precursors containing the mutated protease described above (pET15bVPg-Pro<sup>H1283D</sup>-N-Pol) and an additional mutated protease (pET15bVPg-Pro<sup>H1451L</sup>-N-Pol) in which the His<sup>1451</sup> predicted to be located in the putative substrate binding pocket of the protease was replaced by Leu were tested for proteolytic cleavage in *E. coli* at all the temperatures tested. Only the 34.5 kDa precursor was detected confirming our previous observation at 30°C that precursors containing mutated protease
Fig. 3.5. **Temperature dependence of VPg-Pro-N-Pol and Pro-N-Pol auto-cleavage in *E. coli*.** The bacteria containing plasmids pET15bVPg-Pro-N-Pol (wt), pET15bPro-N-Pol (ΔVPg), pET15bVPg-ProH1283D-N-Pol (H-D), and pET15bVPg-ProH1451L-N-Pol (H-L) and pET15b (-: control) were induced for 3 hr at 37°C or 4 hr at 25°C and 20°C. Proteins to be analysed were separated on 12% polyacrylamide-SDS gels, electro-blotted onto PVDF membranes and immunoreacted with the monoclonal antibodies against protease (anti-Pro Abs) or the polyclonal antibodies against VPg (anti-VPg Abs). The expected position for the different precursors and mature proteins is indicated on the right side of the gels. The position of molecular weight markers is indicated on the left side of the gels.
Fig. 3.6. Temperature dependence of the proteolytic activity of the protease on P1 proteins in vitro. (a) The effects of temperatures on the protease activity. The reactions of $[^{35}S]$methionine-labelled *in vitro* translation products generated from plasmid pT7-NTB-VPg-Pro were prepared as described in Material and Methods. Aliquots of the reactions were incubated at different temperatures for 3 and 6 hr. The protease activities were measured as described in Material and Methods. The protease activities were arbitrarily set at 100%. The relative protease activities are given as the percentage of the calculated activity against the activity at 16°C. Three set of data from three independent experiments are given. (b) SDS-PAGE analysis of proteins after processing. The aliquots of the reactions obtained as (a) were incubated under the indicated conditions (hpt: hours post-translation), separated on 10% polyacrylamide-SDS gels and visualized by autoradiography. The positions of the precursor NTB-VPg-Pro(VS) (99 kDa) and the Pro cleavage products [66 kDa NTB and 33 kDa VPg-Pro(VS) (VS: vector sequence)] are indicated. The position of molecular weight markers is indicated on the left side of the gel.
were not cleaved (Fig. 3.5, lanes 3, 7, 12, 16, 20 and 25).

To uncouple protein expression and accumulation from the protease proteolytic activity, a time course assay was conducted on the in vitro translation products of plasmid pT7-NTB-VPg-Pro at different temperatures. The TNT coupled transcription and translation system (Promega) was used to produce the translation products (30°C, 1 hr). The translation was arrested by addition of RNase, DNase and cold methionine. The translation products were incubated at various temperatures and the processing of the precursor was analyzed in a time-course experiment (Fig. 3.6 a and b). The protease was found to be highly active at relatively low temperatures with its optimum proteolytic activity at 16°C (Fig. 3.6 a). Indeed at 16°C, the amounts of the 99 kDa precursor protein (NTB-VPg-Pro) progressively decreased during the course of the incubation while the amount of the mature NTB (66 kDa) and of the intermediate precursor VPg-Pro(VS) (33 kDa) progressively increased (Fig. 3.6 b, lanes 1-5). Only trace amounts of 66 kDa and 33 kDa proteins were detected after incubation at higher temperatures (Fig. 3.6 b, lanes 11-20). These results suggest that the optimum temperature for processing of protease precursors is relatively low both in in vitro and in E. coli.

3.4.4 Presence of VPg domain on precursors increases cleavage efficiency at the Pro-Pol cleavage site

The presence of VPg domains on cowpea mosaic comovirus has been shown to influence cleavage at the Pro-Pol cleavage site (Dessens and Lomonossoff, 1992). To determine if the presence of the VPg domain on a TomRSV precursor has an effect on processing at the Pro-Pol cleavage site in E. coli, the clone pET15bPro-N-Pol (Fig. 3.1) was constructed by deletion of the cDNA region encoding VPg in the plasmid pET15bVPg-Pro-N-Pol. As expected, anti-VPg antibodies could not detect the Pro-N-Pol precursor polypeptide in Western blotting experiments (Fig. 3.5, lanes 15, 19 and 24). However, anti-Pro antibodies could detect a protein with an apparent molecular mass of 31.5 kDa in extracts from the bacteria induced at 37°C or 25°C, which was of the predicted size for the Pro-N-Pol precursor (Fig. 3.5, lanes 2 and 6). If induction was carried out at 20°C, an additional
protein of an apparent molecular mass of 27 kDa corresponding to the predicted size for the mature Pro was detected by anti-Pro antibodies (Fig. 3.5, lane 11). Therefore, cleavage at the Pro-Pol cleavage site did not occur in the bacteria induced at 37°C or 25°C but did at 20°C. These results support the above suggestion that processing of protease precursors is temperature dependent. At 20°C, approximately half of the Pro-N-Pol precursor was cleaved (Fig. 3.5, lane 11). However, at 20°C, all the detectable VPg-Pro-N-Pol precursor produced by the bacteria containing the clone pET15bVPg-Pro-N-Pol was processed at the Pro-Pol cleavage site (Fig. 3.5, lanes 10 and 23).

A time course analysis was performed to study if the presence of VPg on the precursor would affect processing at the Pro-Pol cleavage site in vitro. For this purpose, plasmid pT7-VPg-Pro-N-Pol-III containing the coding region for the entire VPg, the entire Pro and the N-terminal portion of Pol was created (Fig. 3.1). A derivative of this plasmid in which the VPg coding region was precisely deleted was also produced (plasmid pT7-Pro-N-Pol-III, Fig. 3.1). Upon translation in the coupled transcription and translation system, plasmid pT7-VPg-Pro-N-Pol-III generated a protein of an apparent molecular mass of 57 kDa corresponding to the predicted size for the VPg-Pro-N-Pol (Fig. 3.7a; Fig. 3.7b, lane 1). Plasmid pT7-Pro-N-Pol-III produced a protein of a molecular mass of 54 kDa corresponding to the predicted size for Pro-N-pol III (Fig. 3.7b, lane 8). After incubation of the 57 kDa precursor at 16°C, a small protein with an apparent molecular mass of 33 kDa corresponding to the predicted size for VS-VPg-Pro (3 kDa VS, 3 kDa VPg and 27 kDa Pro) was detected (Fig. 3.7a and b, lanes 1-5). This protein could not be observed upon incubation at 16°C of a precursor containing a protease mutated in the catalytic triad (Fig. 3.7, lanes 6 and 7), demonstrating that generation of the 33 kDa protein was the result of the Pro activity. If the VPg was deleted from the 57 kDa precursor, auto-processing of the 54 kDa precursor (Pro-N-Pol-III) was not detected at 16°C (Fig. 3.7, lanes 8-12). To test whether the 54 kDa protein (Pro-N-Pol-III) had retained its proteolytic activity, non-labelled translation products of plasmid pT7-Pro-N-Pol-III were incubated with the MPCAT substrate which contains a cleavage site from the P2 polyprotein (Hans
Fig. 3.7. *Presence of VPg is critical to the protease activity on the Pro-Pol cleavage site in vitro.* (a) Schematic representation of processing the (VS)VPg-Pro-N-Pol-III precursor protein. The cleavage sites present are shown by short vertical lines. The expected molecular mass of the predicted precursor protein, the predicted intermediate precursor [(VS)VPg-Pro (VS: vector sequence)] and the mature N-terminal portion of Pol (N-Pol-III) are shown. (b) Processing of (VS)VPg-Pro-N-Pol-III *in vitro.* [S\(^{35}\)]methionine-labelled *in vitro* translation products of plasmids pT7-VPg-Pro-N-Pol-III (wt: containing the wild type protease), pT7-VPg-Pro\(^{H1283D}\)-N-Pol-III (H-D: containing the inactive protease), pT7-Pro-N-Pol-III (wt: containing the wild type protease), pT7-Pro\(^{H1283D}\)-N-Pol-III (H-D: containing the inactive protease) were prepared as described in Material and Method. The aliquots were incubated at 16°C for the times indicated (htp: hours post-translation), separated on 10% polyacrylamide-SDS gels and visualized by autorraphy. The position of the 57 kDa (VS)VPg-Pro-N-Pol-III precursor, the 54 kDa (VS)Pro-N-Pol-III precursor and the 33 kDa (VS)VPg-Pro processed intermediate precursor of the 57 kDa precursor is indicated. The position of molecular weight markers is shown on the left side of the gels.
and Sanfaçon, 1995). Efficient cleavage was detected (data not shown). Taken together, these results suggest that the presence of the VPg domain on the precursor results in increased processing at the Pro-Pol cleavage site in *E. coli* and *in vitro*.

### 3.4.5 Expression and processing of a complex protease precursor in *E. coli*

To observe the TomRSV protease proteolytic activity in a larger precursor containing multiple cleavage sites, the entire NTB-VPg-Pro-Pol domains were cloned into vector pTrxFus in frame with the thioredoxin gene to give plasmid pTrxFusNTB-VPg-Pro-Pol (Fig. 3.1). The bacteria containing this plasmid were induced at 20°C to maximize processing efficiency. Presence of the predicted precursor, potential mature products and intermediate precursors in extracts from induced cells was analysed by Western immunoblot, using anti-Thio monoclonal antibodies, anti-NTB polyclonal antibodies, anti-VPg polyclonal antibodies, anti-Pro monoclonal antibodies and anti-Pol polyclonal antibodies. All antibodies recognized a protein with an apparent molecular mass of 190 kDa, which is of the expected size for Thio-NTB-VPg-Pro-Pol precursor protein containing the thioredoxin (Thio) (13 kDa), NTB (66.5 kDa), VPg (3 kDa), Pro (27 kDa) and Pol (80.5 kDa) (Fig. 3.8 b, lanes 1, 3, 5, 7, 9). All the antibodies also detected a protein of apparent molecular mass 155 kDa. The origin of this protein is not clear. It may have been produced by premature termination resulting in a truncated precursor polyprotein. A protein of apparent molecular mass 110 kDa was also recognized by all the antibodies tested. This protein is of the expected size for the following two intermediate precursors: the Thio-NTB-VPg-Pro which would be recognized by anti-Thio, anti-NTB, anti-VPg and anti-Pro antibodies, and the VPg-Pro-Pol which would be detected by anti-VPg, anti-Pro and anti-Pol antibodies and therefore may correspond to a mixture of these two intermediate precursors. Anti-Thio antibodies and anti-NTB antibodies recognized a protein of apparent molecular mass 79.5 kDa which is of the estimated size for the Thio-NTB mature fusion protein (Fig. 3.8 b, lanes 1 and 3). Anti-Pol polyclonal antibodies recognized a protein of apparent molecular mass 80.5 kDa, which corresponds to the predicted size for the mature Pol (Fig. 3.8 b, lanes 9 and
Fig. 3.8. Processing of the NTB-VPg-Pro-Pol precursor polypeptide in *E. coli*. (a) Schematic diagrams of precursor polyproteins processed in bacterial cells. The open solid line boxes represent the cDNA region of TomRSV RNA-1 and the discontinuous line box represents the thioredoxin protein fused in frame to the open reading frame. The expected precursor polyprotein [(Thio)NTB-VPg-Pro-Pol] is shown with the predicted molecular mass. The cleavage sites are indicated as vertical bars. Two possible pathways for the protease to process the precursor are labelled as (1) and (2). The predicted molecular mass is indicated under each processed protein. (b) Immunodetection of precursors and processed products in *E. coli* extracts. The bacteria containing plasmid pTrxFusNTB-VPg-Pro-Pol (Ind: extracts from induced bacterial cells including lanes 1, 3, 5, 7, 9) and pTrxFus (-: lanes 2, 4, 6, 8, 10 as controls) were induced 4 hr at 20°C. The proteins in the *E. coli* extract were separated on 7.5% polyacrylamide-SDS gels and transferred onto PVDF membranes. In Western blots, the following antibodies were used: monoclonal antibodies against thioredoxin (anti-thio); polyclonal antibodies against NTB (anti-NTB); polyclonal antibodies against VPg (anti-VPg); monoclonal antibodies against Pro (anti-Pro); and polyclonal antibodies against Pol (anti-Pol). Molecular weight markers are shown on the left. The nature of the detected proteins are shown on the right. (c) Immunodetection of precursors and processed products in *E. coli* extracts. The induced and processed proteins as in (b) were separated on 12% polyacrylamide-SDS gels, blotted onto PVDF membranes and immunodetected using polyclonal antibodies against VPg (anti-VPg) or monoclonal antibodies against the protease (anti-pro). Only the area of interest is shown. The expected positions for the precursors and mature proteins are shown on the right side of the gels. The position of molecular weight markers is indicated on the left side of the gels.
10). A protein of apparent molecular mass 30 kDa, which is of the expected size for VPg-Pro, immuno-reacted with both anti-VPg antibody and anti-Pro antibody (Fig. 3.8 c, lanes 1 and 3). A protein of apparent molecular mass 27 kDa corresponding to the predicted size for the mature Pro could only be detected at very low level by anti-Pro antibodies (Fig. 3.8 c, lane 3). A predicted Thio-NTB-VPg (82.5 kDa) precursor that would have arisen by cleavage at the VPg-Pro site was not detected by anti-Thio, anti-VPg or anti-NTB antibodies.

3.4.6 Mutation of the cleavage sites at both sides of the putative VPg and protease domains

Site-directed mutagenesis was performed on the putative the NTB-VPg and Pro-Pol cleavage sites. Based on the finding that the cleavage sites of TomRSV are more similar to those of picorna-, como-, and potyviruses than to those of other nepoviruses (Hans and Sanfaçon, 1995), one potential cleavage site, Q/S (1212/1213) and two potential cleavage sites, Q/M (1465/1466) and Q/S (1486/1487), were predicted between NTB and VPg and between Pro and Pol, respectively (Rott et al., 1995).

For the NTB-VPg cleavage site, the cDNA clone pT7-NTP-VPg-Pro (Fig. 3.1) was mutated to delete the codon for the Q at position 1212. After incubation of *in vitro* translation products, the precursor protein generated by the wild type clone was cleaved into two smaller proteins, 66 kDa NTB and 33 kDa VPg-Pro-VS (Fig. 3.4 b, lanes 1 and 2). The precursor produced by the mutated clone was not cleaved (Fig. 3.4 b, lanes 5-6), indicating that deletion of the amino acid at the potential cleavage site abolished the processing at the NTB-VPg cleavage site. This result suggests that the sequence of the NTB-VPg cleavage site is likely to be GKMTVQ/STIPSG, deduced from RNA-1 cDNA clone (Rott et al., 1995). Direct microsequencing of the VPg purified from the infected plants was consistent with this suggestion (Chapter 4).

For the Pro-Pol cleavage site, the cDNA clones pT7-VPg-Pro-N-Pol-II containing the wild type protease and pT7-VPg-Pro\[^{H1283D}\]-N-Pol-II containing a protease mutated in the catalytic triad were created to express *in vitro* the region of the TomRSV polyprotein P1 containing the VPg, Pro,
Fig. 3.9. **Mutation of the predicted cleavage sites between Pro and Pol.** (a) Schematic representation of processing of the VPg-Pro-N-Pol-II precursor protein. The cleavage sites present are shown by short vertical bars. The expected molecular mass of the predicted polyproteins (VPg-Pro-N-Pol-II), the predicted intermediate precursors (VPg-Pro) and the mature N-terminal portion of Pol (N-Pol-II) are shown. (b) Suggestion of the cleavage sites by mutations of the predicted cleavage sites. [S\(^{35}\)]methionine-labelled *in vitro* translation products of clones pT7-VPg-Pro-N-Pol-II (wt: containing the wild type protease), pT7-VPg-Pro\(^{H1283D}\).N-Pol-II (H-D: containing the inactive protease), pT7-VPg-Pro\(^{QS-AS}\).N-Pol-II (QS-AS: containing a mutation at the Q\(^{1486}\)-S\(^{1487}\) predicted cleavage site) and pT7-VPg-Pro\(^{QM-AM}\).N-Pol-II (QM-AM: containing a mutation at the Q\(^{1465}\).M\(^{1466}\) predicted cleavage site) were separated on 10% polyacrylamide-SDS gels and visualized as described in Material and Methods. The times of incubation are indicated (htp: hours post-translation). The positions of uncleaved VPg-Pro-N-Pol-II and the Pro cleavage products are indicated. The position of molecular weight markers is indicated on the left side of the gel.
and N-terminal Pol (Fig. 3.9 a). After incubation of in vitro translation proteins from pT7-VPg-Pro-N-Pol-II, a protein with an apparent molecular mass of 50 kDa close to the predicted size for the precursor polyprotein (3 kDa VPg, 27 kDa Pro and 18 kDa N-Pol-II) was processed to give two proteins with apparent molecular masses of 30 kDa (corresponding to the predicted size for VPg-Pro) and 22 kDa (Fig. 3.9 a and b, lanes 1 and 2). However, the 50 kDa precursor containing an inactive protease was not processed after incubation (Fig. 3.9 b, lanes 3 and 4). These results confirm that the clone pT7-VPg-Pro-N-Pol-II encodes a protease cleavage site between the protease and the polymerase that was cleaved by the endogenous protease. Plasmids pT7-VPg-Pro-<sup>QS-AS</sup>-N-Pol-II (Q in position 1486 mutated to an A) and pT7-VPg-Pro-<sup>QM-AM</sup>-N-Pol-II (Q in position 1465 mutated to an A) were created to determine which of the two potential cleavage sites was cleaved by the protease in vitro. Only the 50 kDa precursor was observed in the radio-labeled polypeptides generated from pT7-VPg-Pro-<sup>QS-AS</sup>-N-Pol-II (Fig. 3.9 b, lanes 5 and 6). The 50 kDa, the 30 kDa and the 22 kDa proteins could be visualized from clone pT7-VPg-Pro-<sup>QM-AM</sup>-N-Pol-II (Fig. 3.9 b, lanes 7-8). Therefore, mutation of the Q/S potential cleavage site abolished processing between Pro and Pol while mutation of the Q/M potential cleavage site did not affect processing. Hence, these results indicate that the cleavage site between Pro and Pol is S-F-A-P-C-Q/S-S-S-V-I-K. Results obtained from direct microsequencing of the N-terminus of the 22 kDa protein confirmed these results (K. Carrier, unpublished data).

To determine the VPg-Pro cleavage site, microsequencing of the mature protease processed in E. coli was conducted. The clone pET15bVPg-Pro-N-Pol was induced at 25 °C for 3 hr. Anti-pro monoclonal antibodies recognized three proteins with apparent molecular masses of 27 kDa, 30 kDa and 34.5 kDa while anti-VPg antibodies recognized two proteins (30 kDa and 34.5 kDa) (Fig. 3.5 b, lanes 5 and 18). The band stained by Coomassie blue corresponding to the position of 27 kDa protein recognized only by anti-pro monoclonal antibodies was subjected to microsequencing. The sequence of the N-terminus of Pro was determined to be G-S-S-L-A-E-A-Q, which is identical to
the sequence deduced from the TomRSV cDNA clone. The cleavage site between VPg and Pro was determined to be R-P-Q-S-V-Q/G.

3.5 Discussion

In this chapter we report proteolytic processing of TomRSV protease precursors in *E. coli* as verified by protease mutation experiments, N-terminal amino acid sequencing of the mature products and Western immunoblots with specific antibodies. We have studied processing of truncated VPg-Pro-Pol and Pro-Pol precursors and of the complete NTB-VPg-Pro-Pol precursor. We have observed cleavage at the NTB-VPg, VPg-Pro and Pro-Pol cleavage sites. Proteolytic processing of 3C protease precursors of animal viruses has been demonstrated in *E. coli*, e.g. on poliovirus 3CD (Pro-Pol) precursors (Baum *et al.*, 1991; Ivanoff *et al.*, 1986) and on hepatitis A virus 3ABCD (NTB-VPg-Pro-Pol) precursors (Gauss-Müller *et al.*, 1991; Harmon, *et al.*, 1992; Schultheiss *et al.*, 1994). For plant viruses, slow processing of potyvirus NIa (VPg-pro) protein has been shown to occur in *E. coli* (Laliberté *et al.*, 1992; Kim *et al.*, 1995; Ménard *et al.*, 1995; Parks *et al.*, 1995). In comoviruses, a complex hybrid protease precursor containing sequences of B-RNA and M-RNA was processed in *E. coli* resulting in the release of the mature coat protein contained in the M-RNA sequences (Garcia *et al.*, 1987). Our results provide the first evidence of proteolytic processing of nepovirus protease precursors in *E. coli*.

In this chapter, we show that proteolytic processing of protease precursor is temperature dependent in *E. coli* and *in vitro*. Our kinetic studies using the *in vitro* translation system which uncouples expression and processing suggest that the activity of the protease itself is affected by the processing temperature. Alternatively, folding of the protease precursors may be altered at higher temperature resulting in sub-optimal secondary or tertiary structure around the cleavage sites precluding their optimal recognition. However, our results indicate that processing of two different precursors (VPg-Pro-Pol and NTB-VPg-Pro) is temperature dependent at all cleavage sites studied.
in *E. coli* and *in vitro*. In *E. coli*, it is possible that the expressed proteins were less soluble at higher temperature as suggested by Laliberté *et al* (1992) for their observation that internal cleavage of NIa protease of turnip mosaic virus (TuMV) in *E. coli* occurred at room temperature but not at 37°C. We have analysed the relative solubility of the expressed proteins at different temperatures. The protease (precursors and mature products) was found to be mainly in the insoluble fraction regardless of the temperature of induction (data not shown). It seems therefore unlikely that the temperature dependence of processing of the TomRSV protease precursors in *E. coli* could be due solely to the different solubility of the fusion proteins.

Purified potyvirus NIa protease was shown to have an optimum *trans* catalytic activity on a synthetic peptide at approximately 15°C *in vitro* while autocatalytic activity was shown to be higher at 25°C than at 12°C (Kim *et al*., 1996). The authors did not provide additional results at higher temperatures. They concluded that the low temperature optimum of the protease *trans*-activity was related to the highly flexible structure of the NIa protease at lower temperatures. In an animal virus, Mengo virus, the optimal temperature for the 3C protease on a synthetic peptide encompassing the 2C-3A cleavage site was 30-37°C (Hall and Palmenberg, 1996). According to our findings and to the above observations, we propose that the preference for low temperature may be a property unique to plant virus proteases, which may have arisen as an adaption to the plant cell environment.

Although we could detect processing at all cleavage sites in *E. coli*, several lines of evidence lead us to suggest that cleavage at the VPg-Pro cleavage site is inefficient both in *E. coli* and *in vitro*. First, proteolytic processing of a number of precursors: NTP-VPg-Pro, VPg-Pro-N-Pol, VPg-Pro-N-Pol-II, VPg-Pro-N-Pol-III and Thio-NTB-VPg-Pro-Pol resulted in accumulation of the VPg-Pro precursor in *E. coli* and *in vitro*. Release of mature protease could not be detected *in vitro* and was detected only in small precursors at low temperatures in *E. coli*. Second, intermediate precursors resulting from cleavage at the VPg-Pro cleavage site were not detected (such as the Pro-N-Pol intermediate precursor for clones pET15bVPg-Pro-N-Pol, pT7-VPg-Pro-N-Pol-II, pT7-VPg-Pro-N-
Pol-III, or the NTB-VPg intermediate precursor for clone pT7-NTB-VPg-Pro, or the Thio-NTB-VPg intermediate precursor for clone pTrxFusNTB-VPg-Pro-Pol). Our inability to detect these intermediate precursors was not likely due to their intrinsic instability as demonstrated by the stable expression of the Pro-N-Pol precursor in E. coli (Fig. 3.5, lanes 2, 6 and 11) and in vitro (Fig. 3.7 b, lanes 8-14). Based on the above observations, we conclude that processing is likely to occur stepwise. In the VPg-Pro-N-Pol precursor, our results suggest that processing occurs predominantly at the Pro-Pol site releasing the VPg-Pro intermediate precursor. Under certain conditions (in E. coli at low temperatures) the VPg-Pro is further cleaved to give rise to the mature protease. The possible pathways by which the TomRSV protease could process a large precursor (Thio-NTB-VPg-Pro-Pol) is suggested (Fig. 3.8 a): the protease would first cleave either at the NTB-VPg site to release the Thio-NTB fusion protein or at the Pro-Pol cleavage site to release the mature polymerase. The VPg-Pro would then be produced from the Thio-NTB-VPg-Pro or the VPg-Pro-Pol smaller intermediate precursors. According to this model, a third step would then be required to release the mature protease from the VPg-Pro precursor. Indeed only very small amounts of mature protease were detected in extracts from cells expressing this precursor. Similar to our results, Margis et al. (1994) reported that maturation of the VPg-Pro intermediate precursor of GFLV occurred at very low rates in vitro. Also, the potyvirus Nla protein was very inefficiently cleaved at the VPg-Pro site in vitro and in E. coli (Dougherty and Parks, 1991; Laliberte et al, 1992). In contrast, the VPg-Pro-Pol precursor of tomato blackring nepovirus (TBRV) was found to accumulate in vitro and in vivo, indicating inefficient cleavage at both the VPg-Pro and Pro-Pol cleavage sites (Demangeat et al., 1992; Hemmer et al., 1995). TomRSV VPg-Pro has been shown to contain protease activity on a P2 substrate recognized in trans (Hans and Sanfaçon, 1995). Differential activities of the VPg-Pro and Pro on P2 substrates were observed for two other nepoviruses (Margis et al, 1994; Hemmer et al, 1995). Further experiments will be necessary to determine if the TomRSV VPg-Pro also accumulates in infected plants and if the presence of the VPg on the TomRSV protease influences its activity in
trans on P2 substrates.

We show that the presence of VPg sequences on the precursor resulted in an increase of processing efficiency at the Pro-Pol cleavage site both in *E. coli* and *in vitro*. Previously, the presence of the VPg on CPMV B-RNA encoded polyprotein precursors was found to enhance processing at the Pro-Pol cleavage site (Dessens and Lomonossoff, 1992). In contrast, presence of the VPg domain on the VPg-Pro-Pol precursor of GFLV did not influence cleavage at the Pro-Pol site (Margis et al., 1994). One likely explanation is that the presence of VPg sequences on the precursor results in a conformational change that favours processing at the TomRSV and CPMV Pro-Pol cleavage sites. It is also possible that other sequences on P1 not included in the precursors studied here, may influence cleavage at the VPg-Pro or other cleavage sites.

We have identified one and suggested two new dipeptide cleavage sites for TomRSV, i.e. Q/G between VPg and Pro, Q/S between NTB and VPg, and Q/S between Pro and Pol. Previously, the sequence of the TomRSV polyprotein cleavage site between the movement protein and the coat protein (Q/G) was determined by this laboratory (Hans and Sanfaçon, 1995). The specificity of cleavage by the picorna-, poty- and comovirus proteases is for peptide bonds between glutamine (or glutamic acid) and residues with small side chains, such as Q/S, Q/M, Q/G, E/G and E/S (Wellink et al., 1986; Hellen et al., 1989; Palmenberg, 1990). Our results are fully consistent with those known cleavage sites and support our previous suggestion that TomRSV protease cleavage sites are more related to the cleavage sites of the poty-, como-, and picornavirus proteases than to the cleavage sites recognized by other nepovirus proteases (Hans and Sanfaçon, 1995). Recently, the atomic structures of recombinant 3C proteases from rhinovirus-14 (HAR14) and hepatitis A virus (HAV) show shallow elongated substrate pockets that could easily accommodate 8 amino acids from the -4 position to the +4 position of the cleavage site (Allaire et al., 1994; Matthews, et al., 1994). Gln in the -1 position has been suggested to be recognized primarily by the histidine of the protease substrate binding pocket. Indeed, mutation of this histidine results in inactivation of proteases as
described previously (Bazan and Fletterick, 1988; Gorbalenya et al., 1989, Hans and Sanfaçon, 1995) and in this study. Suggestion of the NTB-VPg, VPg-Pro and Pro-Pol cleavage sites allowed the tentative localization of the putative VPg domain and protease domain on the TomRSV RNA-1 encoded P1 polyprotein. Deduced from the RNA-1 cDNA clone, the putative VPg has a calculated molecular mass of 3.02 kDa and contains 27 amino acids located between amino acids 1213 and 1239. The protease has a calculated molecular mass of 26.80 kDa and contains 247 amino acids located between amino acids 1240 and 1486.
CHAPTER 4

GENOMIC ORGANIZATION OF TOMATO RINGSPOT NEPOVIRUS RNA-1: PROTEOLYTIC PROCESSING OF THE AMINO ACID-TERMINUS OF THE POLYPROTEIN AND DETECTION OF P1 PROTEINS IN PLANTS
4.1 Abstract

Tomato ringspot virus (TomRSV) RNA-1 encodes a polyprotein (P1) with a predicted molecular mass of 244 kDa. The C-terminal region of the polyprotein has been shown to contain the putative NTP-binding protein (NTB), a VPg, the protease (Pro) and the putative RNA-dependent RNA polymerase (Pol). The N-terminal region (upstream of NTB) of P1, with a coding capacity for a protein (or precursor protein) of 67 kDa, however, has not been characterized. To study the expression and processing strategy of the N-terminal portion of the P1 polyprotein, a series of overlapping RNA-1 cDNA clones were constructed and transcribed and translated in vitro. It was found that the TomRSV RNA-1-encoded 3C-like protease could process the N-terminal P1 region at a predicted cleavage site to release two proteins, i.e. X1 and X2. Therefore, there is one additional protein (X1) located at the N-terminus of P1, which is absent in other sequenced members of the family comoviridae. To detect all P1 mature proteins including X1, X2, NTB, VPg, Pro and Pol, Western immunoblotting was performed using antibodies against these mature proteins. A number of viral mature and intermediate precursor proteins were detected, including NTB, NTB-VPg, Pro, Pol, VPg-Pro-Pol and two proteins with molecular masses of 35 and 66 kDa which may correspond to the X1 protein and the X1-X2 precursor, respectively. The amino acid sequence of the VPg purified from the infected plants was determined, allowing the identification of the NTB-VPg cleavage site. Dipeptide sequences of the X1-X2 and X2-NTB cleavage sites were inferred from analysis of in vitro site-directed mutagenesis of the predicted cleavage sites. In vitro processing of P1 polyprotein occurred only intramolecularly. A model for the expression and processing of P1 is proposed.

4.2 Introduction

TomRSV, along with peach rosette mosaic, cherry leaf roll and myrobalan latent ringspot
viruses has been classified as subgroup c nepoviruses. Of the members in the subgroup c, TomRSV is the only one whose genome has been completely sequenced. RNA-1 is 8214 nucleotides (nt) in length excluding the 3' poly (A) tail (Rott et al., 1995) and contains a single long open reading frame (ORF) of 6591 nucleotides beginning at the first AUG codon at nucleotide position 78 (Rott et al., 1995). Translation of this ORF from the first AUG codon would lead to the synthesis of a polyprotein (P1) with a predicted molecular mass of 244 kDa (Rott et al., 1995). Previously we have demonstrated that the TomRSV RNA-1-encoded 3C-like protease is responsible for processing of the C-terminal portion of P1 at the NTB-VPg, VPg-Pro and Pro-Pol cleavage sites (Chapter 3). The N-terminal domain of the TomRSV P1 polyprotein with a total coding capacity for a protein or precursor protein of 67 kDa has not been characterized. In other members of the Comoviridae family, only one mature protein is released from the region upstream of NTB. It is not known whether one or several mature proteins are produced from the corresponding region of the TomRSV P1 polyprotein. In this study, we have analyzed the processing strategy at the N-terminus of the P1 polyprotein and provided evidence for the presence of an additional cleavage site at this region. Along with our previous processing data at the C-terminus of the P1 polyprotein (Chapter 3), the genomic organization of TomRSV RNA-1 has been established.

4.3 Materials and Methods

4.3.1 Plasmid construction

Plasmid pMR10, a full cDNA clone of TomRSV RNA1 was obtained from Rott et al. (1995).

To obtain plasmid pTrxFusC-X1 (Fig. 4.1), a 454 bp fragment of pMR10 (RNA-1 nt 602-1055) was amplified with Pfu polymerase (Stratagene). Primers W016 (Table 4.1) and W017 were used for the amplification. The amplified fragment was digested with Kpn I and Xba I, and inserted into the corresponding sites of the pTrxFus vector (Invitrogen).

To construct plasmid pT7-X1-Pro (Fig.4.1), a 3748 nt fragment of pMR10 (RNA-1 nt 786-
4534) embracing all potential cleavage sites upstream of Pro was amplified with primers W034 and W023, and digested with Nco I and Sal I. The digested fragment was ligated into the Nco I-Sal I sites of the vector pCITE-4b(+) (Novagen). For the construction of plasmid pT7-X1-Pro<sup>H1283D</sup>, the following three fragments were ligated: a 3.8 kb Nco I-Bam H I fragment from plasmid pT7-X1-Pro, a 609 bp Nco I-Xho I fragment from plasmid pX1-Pro and a 1043 bp Xho I-Bam H I fragment from plasmid pT7-NTB-VPg-Pro<sup>H1283D</sup> (Chapter 3). Plasmid pT7-X1-Pro<sup>ΔQ326</sup>, in which the codon encoding the amino acid Q at position 326 (amino acid sequence of P1) was deleted, was constructed by self-ligation of a fragment amplified by priming plasmid pT7-X1-Pro with phosphorylated primers W038 and W039.

Table 4.1 Primers used for construction of plasmids and site-directed mutagenesis<sup>a</sup>

<table>
<thead>
<tr>
<th>Name</th>
<th>Polarity</th>
<th>Sequence (5' to 3')</th>
<th>Corresponding sequence of RNA-1</th>
<th>Restriction site engineered</th>
</tr>
</thead>
<tbody>
<tr>
<td>W016</td>
<td>+</td>
<td>CGCGGTACCCGAGCTCATCCGAGCTCGC</td>
<td>602-621</td>
<td>Kpn I</td>
</tr>
<tr>
<td>W017</td>
<td>-</td>
<td>CGTCTAGACTACTGATGGGAAGCACACTTGCC</td>
<td>1055-1035</td>
<td>Xba I</td>
</tr>
<tr>
<td>W023</td>
<td>-</td>
<td>ACGCGTCGACCTGCAAGGAGCACAATGGGC</td>
<td>4534-4518</td>
<td>Sal I</td>
</tr>
<tr>
<td>W024</td>
<td>+</td>
<td>ACGCCCATGGGAAGCATATGATCTCTTGTC</td>
<td>1055-1073</td>
<td>Nco I</td>
</tr>
<tr>
<td>W025</td>
<td>-</td>
<td>ACGCGTCGACCTGACATGATTTGTGGC</td>
<td>3794-3778</td>
<td>Sal I</td>
</tr>
<tr>
<td>W026</td>
<td>+</td>
<td>ACGCCCATGGGAATATTGCTAGTGTGTG</td>
<td>2948-2967</td>
<td>Nco I</td>
</tr>
<tr>
<td>W028</td>
<td>+</td>
<td>ACGCCCATGGGAATTTGAGTGCTCATG</td>
<td>1959-1977</td>
<td>Nco I</td>
</tr>
<tr>
<td>W030</td>
<td>+</td>
<td>CGCCCATGGGGCTGAGGAGGAGAGGCTG</td>
<td>786-801</td>
<td>Nco I</td>
</tr>
<tr>
<td>W034</td>
<td>+</td>
<td>CGCGGTACCCGAGCTCATCCGAGCTCGC</td>
<td>6696-6682</td>
<td>Sal I</td>
</tr>
<tr>
<td>W035</td>
<td>-</td>
<td>GGCGCTCACTGAGCTTTCG</td>
<td>1938-1956</td>
<td>N/A</td>
</tr>
<tr>
<td>W036</td>
<td>-</td>
<td>AGTTGGGCTGCTGGACACCACCGT</td>
<td>1934-1917</td>
<td>N/A</td>
</tr>
<tr>
<td>W038</td>
<td>+</td>
<td>GAAAGCATATGATCTCTTTCG</td>
<td>1056-1075</td>
<td>N/A</td>
</tr>
<tr>
<td>W039</td>
<td>-</td>
<td>ATGGGAAGCAGCAGCTTCGGC</td>
<td>1052-1035</td>
<td>N/A</td>
</tr>
<tr>
<td>KC13</td>
<td>+</td>
<td>TTTGCTCATCTGGCAGTCTATG</td>
<td>4521-4550</td>
<td>N/A</td>
</tr>
<tr>
<td>34</td>
<td>+</td>
<td>GCCGCGGTACCCGAGCTCATCCGAGCTCGC</td>
<td>6696-6682</td>
<td>N/A</td>
</tr>
</tbody>
</table>

<sup>a</sup> The sequences for engineering restriction site are underlined. In KC13, three nucleotides, i.e. c, c, t (underlined) were deleted.

Plasmid pT7-X2.1-Pro (Fig. 4.1) was obtained by amplifying of a DNA fragment from clone pMR10 (RNA-1 nt 1055-4534) with primers W024 and W023, digesting the amplified fragment with Nco I and Sal I and inserting it into the corresponding sites of vector pCITE-4a(+) (Novagen). The
Fig. 4.1. Schematic diagram of the different constructs derived from TomRSV RNA-1 cDNA, which are used for analyses on the processing strategy at the N-terminus of P1. The predicted genomic organization of TomRSV RNA-1 is shown. The VPg is indicated by a circle at the 5'-end and the predicted open reading frame is indicated by the initiation (AUG) and termination (UAA) codons. The corresponding polyprotein P1 is represented underneath as a solid bar. The putative functions of the mature products and their predicted molecular mass are shown inside the corresponding boxes of RNA-1 and under the corresponding regions of P1, respectively. The determined cleavage sites are indicated in bold and by solid vertical lines. The predicted cleavage sites are shown in italic and by dashed vertical lines. The cDNA clones containing the different portions of the RNA-1 open reading frame are presented and their names given at the right. T7 represents T7 promoter.
corresponding clone pT7-X2.1-Pro\textsuperscript{H1283D}, which encodes a defective protease, was generated by the three piece-ligation strategy as described above for the construction of clone pT7-X1-Pro\textsuperscript{H1283D}.

Primers W026 and W023 were used to amplify a DNA fragment of pMR10 (RNA-1 nt 1366-4534) for the construction of plasmid pT7-X2.2-Pro (Fig. 4.1). The amplified fragment was digested with Nco I and Sal I, and then ligated into the same sites of vector pCITE-4a(+). The three piece-ligation strategy described above was used to produce plasmid pT7-X2.2-Pro\textsuperscript{H1283D} containing a mutated protease. To create the cleavage site mutant pT7-X2.2-Pro\textsuperscript{AQ620}, plasmid pT7-X2.2-Pro was amplified with primers W036 and W037, and the amplified fragment was self-ligated. In this mutant clone, the codon encoding amino acid Q at position 620 (amino acid sequence of P1), which is the -1 position of the predicted X2.2-NTB cleavage site, was deleted.

Plasmid pT7-NTB-VPg was obtained by amplifying a fragment of TomRSV RNA-1 cDNA clone pMR10 (TomRSV RNA-1 nt 1959-3794) with primers W025 and W030. The amplified fragment was cleaved with Nco I and Sal I, and inserted into the corresponding sites of the pCITE-4a(+) polylinker. To clone the predicted NTB-VPg cleavage site with flanking sequences (plasmid pT7-C-NTB-VPg), a fragment of pT7-NTB-VPg containing the coding sequence for the C-terminal part of NTB and the VPg protein (TomRSV RNA-1 nt 2948-3794) was amplified using primers W025 and W028. The amplified fragment was digested with Nco I and Sal I and inserted into the corresponding sites of vector pCITE-4a(+).

To clone the coding sequence for the C-terminal portion of P1, i.e. the entire NTB-VPg-Pro-Pol region, primers W028 and W035 were used to amplify a fragment of pMR10 (TomRSV nt 2948-6696). The amplified fragment was digested with BamH I and Sal I. The 2.3 kb fragment was inserted into the corresponding sites of plasmid pT7-NTB-Pro\textsuperscript{H1283D} (Chapter 3) to give plasmid pT7-NTB-Pol\textsuperscript{H1283D}. Plasmid pT7-MPCAT\Delta ATG was obtained by site-directed mutagenesis using primers 34 to prime synthesis on single-stranded plasmid pT7-MPCAT (Hans and Sanfaçon, 1995) as described (Chapter 3).
*Escherichia coli* strain DH5α (Novagen) was used as the host for plasmid cloning and amplification.

4.3.2 Antisera

The polyclonal antibodies against the putative polymerase, the putative NTB binding protein and the putative VPg, and the monoclonal antibodies against the protease were described previously (Chapter 3). To raise polyclonal antibodies against the C-terminal portion of the X1 protein, plasmid pTrxFusX1 was transformed into *E. coli* strain GI 724 (Invitrogen). The bacteria harboring the plasmid were grown at 30°C in RM medium (Invitrogen) containing 100 μg/ml ampicillin and shaken at 225 rpm overnight. One volume of the culture was used to inoculate 20 volumes of induction medium (Invitrogen). These cultures were grown to an OD$_{550}$ of 0.5, then L-tryptophan was added to a final concentration of 100 μg/ml. The cultures were transferred to 37°C for 3 hr to allow expression. Inclusion bodies were purified essentially as described (Chapter 3). The solubilized protein was dialysed overnight against phosphate-buffered saline (PBS: 136.9 mM NaCl, 2.69 mM KCl, 10.14 mM Na$_2$HPO$_4$ and 1.76 mM KH$_2$PO$_4$, pH 7.4). The purified Thio-X1 fusion protein was injected into mice and the polyclonal antibodies against the Thio-X1 protein were isolated as described (Chapter 3).

4.3.3 Viral RNA purification and microsequencing of the VPg protein

A field isolate of TomRSV originally purified from raspberry was obtained from Dr. R. Stace-Smith (Vancouver Research Station, Agriculture Canada) and maintained on *Nicotiana clevelandii*. The virus was propagated on *Cucumis sativus* var straight 8 infected by mechanical inoculation with extracts from *N. clevelandii* infected leaves and virus particles were purified as described (Wieczorek and Sanfaçon, 1993) except that one additional clarification step (10 min, 8000 g) was included between the two ultracentrifugation steps. The viral RNA was isolated by mixing 500 μl of purified virions with 100 μl of 1 M Tris-Cl pH 8.9, 25 μl of 20% SDS, and 5 μl of 100 mM EDTA pH 8.0, extracting twice with 300 μl phenol (saturated with TE buffer) and 300
µl chloroform /octanol (v/v 24:1) and precipitating with 70% ethanol and 1.2 M ammonium acetate. The precipitated RNA was resuspended in DEPC (diethyl pyrocarbonate) dd-H₂O and stored at -80°C. One mg of virus RNA was hydrolysed in 150 µl of 20% trifluoroacetic acid (TFA, Sigma) for 48 hrs at room temperature. The hydrolysed material was directly subjected to microsequencing (Pinck et al., 1991; Chapter 3).

4.3.4 Purification of recombinant active protease

Recombinant active protease was purified from the expression products of plasmid pET15bPro-N-Pol essentially as described previously (Chapter 3). Upon expression, the protease was insoluble. Solubilization of the protease from the purified inclusion bodies with urea and renaturation of the protease by gradual dialysis was as described (Chapter 3). Purified protease was stored in Tris buffer (50 mM Tris, pH 8; 1mM DTT; 10% glycerol) at -70°C.

4.3.5 In vitro transcription-translation and quantitative measurements

In vitro transcription-translation of the cDNA clones was carried out using a TNT coupled transcription-translation system (Promega). The in vitro synthesized proteins were labelled with 35S-methionine. Protein analysis was essentially as described (Chapter 3).

4.3.6 SDS-PAGE and western immunoblotting

Proteins from in vitro translation were separated by denaturing sodium dodecyl sulphate-polyacrylamide gels electrophoresis (SDS-PAGE) (Laemmli, 1970). The plant sap from TomRSV infected Cucumis sativus or Nicotiana clevelandii leaves was extracted using a Pollähne Press with an equal volume of 0.1 M Tris-HCl, pH 8.0. SDS-PAGE and Western immunoblotting were performed as described (Wieczorek and Sanfaçon, 1993; Chapter 3).

4.4 Results

4.4.1 Expression and processing of the N-terminal portions of P1

Based on our previous results that the TomRSV cleavage sites were consistent with the
conserved consensus dipeptide sequences of the picorna-like viruses [(Q, E)/(G, S, M)] (Hans and Sanfaçon, 1995; Chapter 3), three potential cleavage sites, i.e. Q^{326}/E, Q^{423}/G and Q^{620}/G were predicted upstream of NTB (Fig. 4.1). Cleavage of P1 at these 3 predicted cleavage sites and at 3 previously identified cleavage sites (NTB-VPg, VPg-Pro and Pro-Pol cleavage sites) would give rise to seven end products, i.e. from the N-terminus, X1, X2.1, X2.2, NTB, VPg, Pro and Pol. To elucidate the processing strategy of the N-terminal moiety (upstream of NTB) of P1, a series of precursors in the corresponding region were studied in vitro (Fig. 4.1). In clone pT7-X2.2-Pro, the encoded viral polyprotein X2.2-NTB-VPg-Pro contained one potential cleavage site between X2.2 and NTB (Rott et al., 1995), and two other identified cleavage sites (NTB-VPg and VPg-Pro) (Chapter 3). This clone was translated using a TNT coupled transcription-translation system (Promega). The translation was arrested by addition of an excess of non-radioactive methionine and RNase. The translation products were incubated for 0, 2.5, 5 and 20 hr at 16°C to allow processing. In vitro translation resulted in the synthesis of a protein with an apparent molecular mass of 122 kDa, the predicted size for the precursor polyprotein containing, from the N-terminus, VS (amino acids encoded by the vector sequence, 3 kDa), X2.2 (20 kDa), NTB (66 kDa), VPg (3 kDa), Pro (27 kDa) and VS (3 kDa) (Fig. 4.2 A; Fig. 4.2 B, lane 1). Upon incubation, the amount of the 122 kDa precursor protein progressively decreased while a number of smaller proteins accumulated. These proteins had apparent molecular masses of 89 kDa corresponding to the predicted size for (VS)X2.2-NTB, 66 kDa corresponding to the predicted size for NTB, 33 kDa corresponding to the predicted size for VPg-Pro(VS), and 21.5 kDa approximately corresponding to the predicted size for (VS)X2.2 (23 kDa) (Fig. 4.2 B, lanes 2, 3, 4; Fig. 4.2 C, lane 6). To confirm that these smaller proteins resulted from proteolytic cleavage of the precursor polyprotein by the viral encoded protease, a mutation in the putative catalytic triad of the protease (His at amino acid position 1283 of the P1 sequence replaced with Asp: H1283D) was introduced into the wild type clone. This mutation was previously found to abolish protease activity (Hans and Sanfaçon, 1995; Chapter 3). Translation of
Fig. 4.2. Processing of the X2-NTB-VPG-Pro precursor polypeptide in vitro. (A) Schematic representation of processing of the X2.2-NTB-VPG-Pro precursor protein. The expected molecular masses of the precursor polypeptide and cleaved proteins are shown. Predicted and determined cleavage sites are indicated by short dashed and solid vertical lines, respectively. The predicted dipeptide cleavage site between X2.2-NTB is shown in italic (VS: vector sequence). (B) Time course analysis of processing of the X2.2-NTB-VPG-Pro precursor. [35S]methionine-labelled in vitro translation products of clone pT7-X2.2-Pro (WT) were prepared as described. Aliquots of the reactions were incubated for 0, 2.5, 5 and 20 hr at 16°C (hpt: hours post-translation), separated on 10% polyacrylamide-SDS gels and visualized as described in Material and Methods. The positions of precursor and cleavage products are indicated. (C) Mutation of the predicted X2.2-NTB cleavage site and of the protease on the X2.2-NTB-VPG-Pro precursor. Three plasmids: pT7-X2.2-Pro, pT7-X2.2-ProH128D (HD) and pT7-X2.2-ProG62D (ΔQ62D) were translated, incubated (two time points only) and separated as in (B).
Fig. 4.3. Processing of the X2.1-X2.2-NTB-VPg-Pro precursor polypeptide \textit{in vitro} (A) Schematic representation of processing of the X2.1-X2.2-NTB-VPg-Pro precursor polypeptide. The expected molecular masses of precursor polypeptide and cleaved proteins are shown. The predicted dipeptide cleavage site and identified cleavage sites are indicated by short dashed and solid vertical lines, respectively. The predicted dipeptide cleavage site between X2.1-X2.2 is shown in italic (VS: vector sequence). (B) Auto-cleavage of the X2.1-X2.2-NTB-VPg-Pro precursor. [\textsuperscript{35}S]methionine-labelled \textit{in vitro} translation products of plasmids pT7-X2.1-Pro (WP) and pT7-X2.1-Pro\textsuperscript{H1283D} (HD) were prepared as described in Material and Methods. Aliquots of the reactions were incubated at 16\degree C for the time length indicated above the gel, then separated on a 10\% polyacrylamide-SDS gel and visualized as described in Material and Methods. The positions of precursor and cleavage products are indicated.
this clone gave rise to the predicted 122 kDa precursor polyprotein (Fig. 4.2 C, lane 9). However, none of the smaller proteins which were observed for the wild-type clone were detected after a 20 hr incubation of the precursor containing a defective protease. To further analyze processing of the X2.2-NTB-VPg-Pro precursor and verify the nature of the smaller proteins released, site-directed mutagenesis was conducted on the predicted X2.2-NTB cleavage site. Plasmid pT7-X2.2-Pro\textsuperscript{AQM20} was constructed, in which Gln at amino acid position 620 of the P1 sequence was precisely deleted. Upon destruction of the predicted cleavage site, the 122 kDa precursor was processed into the 89 and 33 kDa proteins, whereas, the 21.5 and 66 kDa proteins could not be detected. These results suggest that the dipeptide sequence Q\textsuperscript{620}/G was the X2.2-NTB cleavage site, and that the 21.5, 33, 66 and 89 proteins were (VS)X2.2, VPg-Pro(VS), NTB and (VS)X2.2-NTB, respectively. Immunoprecipitation experiments with antibodies against NTB, VPg and Pro were consistent with the suggested nature of these proteins (data not shown).

To determine if proteolytic processing could be detected at the predicted X2.1-X2.2 cleavage site \textit{in vitro}, clone pT7-X2.1-Pro was constructed (Fig. 4.1). This clone contained an extended region at the N-terminus of the polyprotein and included one more potential protein X2.1. As a control, a derivative clone, which encoded a polyprotein containing an inactive protease (H1283D mutation) was also constructed. After translation of these two clones in the cell-free system as above, a protein with an apparent molecular mass of 134 kDa corresponding to the calculated size for (VS)X2.1-X2.2-NTB-VPg-Pro(VS) was generated (Fig. 4.3 A; Fig. 4.3 B, lanes 1 and 3). After a 20 hr incubation, several smaller proteins with apparent molecular masses of 33, 66 and 99 kDa were detected in the translation products generated by the wild-type clone but not in those generated by the mutated clone (Fig. 4.3 B, lane 2). The band corresponding to a protein of 33 kDa was diffuse and possibly contained both the VPg-Pro(VS) protein (predicted molecular mass of 33 kDa) and the (VS)X2.1-X2.2 protein (predicted molecular mass of 35 kDa). The 66 kDa protein was likely to be the mature NTB protein (predicted molecular mass of 66 kDa). The 99 kDa protein could consist of
either the (VS)X2.1-X2.2-NTB precursor (predicted molecular mass of 101 kDa) or the NTB-VPg-Pro(VS) precursor (predicted molecular mass of 99 kDa). Presumably, the smaller proteins detected above resulted from cleavage at the X2-NTB, and NTB-VPg cleavage sites. However, the predicted products, which would have resulted from cleavage at the predicted X2.1-X2.2 cleavage site alone or in combination with cleavage at other sites were not detected. These products could have included proteins with the following predicted molecular masses: 123 kDa [X2.2-NTB-VPg-Pro(VS)], 88 kDa (X2.2-NTB), 22 kDa (X2.2), and 13 kDa [(VS)X2.1]. These experiments were repeated several times using protein gels with lower or higher concentrations of polyacrylamide to ensure a maximum separation of the larger precursors and smaller proteins, respectively. Therefore, processing of the potential X2.1-X2.2 cleavage site (dipeptide sequence Q/G) was not detected in vitro. Hereafter, the X2.1 and X2.2 proteins were considered as one protein X2.

To test if proteolytic processing could be detected at the predicted XI-X2 cleavage site in vitro, plasmid pT7-X1-Pro was constructed, which contains the coding region for the C-terminal moiety of X1 (C-X1) (14 kDa) and the entire X2, NTB, VPg and Pro. As expected, a polyprotein with an apparent molecular mass of 143 kDa was observed upon in vitro transcription and translation of this clone and of a derivative clone which contains an inactive protease (H1283D mutation) (Fig. 4.4 B, lanes 1 and 5). Upon incubation of the translation products, the wild-type polyprotein was processed to release the smaller proteins with apparent molecular masses of 29, 33, 44, 66, 97, 112 and 143 kDa (Fig. 4.4 B, lanes 1 and 2; Fig. 4.4 C, lanes 7 and 8). The 29 kDa protein corresponded to the predicted size for X2 (32 kDa) or mature Pro(VS) (30 kDa). The 33 kDa protein was of the predicted size for VPg-Pro(VS) (33 kDa) or X2 (32 kDa). The 44 kDa protein was probably the XI-X2 intermediate precursor (predicted molecular mass of 46 kDa). The 66 kDa protein was likely to be NTB (66 kDa). The 97 kDa protein corresponded to the predicted sizes for X2-NTB (98 kDa) or NTB-VPg-Pro(VS) (99 kDa). The 112 kDa protein was of the predicted molecular mass for X1-X2-NTB (112 kDa). These smaller proteins were not produced by incubation of the protease
Fig. 4.4. Processing of the X1-X2-NTB-VPg-Pro precursor polypeptide in vitro. (A) Schematic representation of processing of the X1-X2-NTB-VPg-Pro polyprotein. The expected molecular masses of precursor polypeptide and cleaved proteins are shown. The predicted dipeptide cleavage site and identified cleavage sites are indicated by short dashed and solid vertical lines, respectively. The predicted dipeptide cleavage site between X1-X2 is shown in italic (VS: vector sequence). (B) Auto-cleavage of the X1-X2-NTB-VPg-Pro precursor. Three plasmids: pT7-X1-Pro, pT7-X1-Pro_{Ha283D} (HD) and pT7-X1-Pro_{Q326D} (ΔQ326) were transcribed and translated as described in Material and Methods. The translation products were incubated for 20 hr at 16°C, separated on a 7.5% polyacrylamide-SDS gel and visualized as described in Material and Methods. The positions of the precursor and cleavage products are indicated. (C) Auto-cleavage of the X1-X2-NTB-VPg-Pro precursor. The translation products obtained as in (B) were separated on a 10% polyacrylamide-SDS gel. Only the area of interest is shown. The expected positions of the precursor and mature proteins are shown on the right side of the gel. The position of molecular weight markers is indicated on the left side of the gel.
precursor generated by clone pT7-X1-Pro^{H128D} in which the protease was mutated (Fig. 4.4 B, lanes 5 and 6; Fig. 4.4 C, lanes 11 and 12), suggesting that all the smaller proteins were proteolytic products of the protease. To verify the nature of the 29 kDa protein, site-directed mutagenesis was conducted on the predicted X1-X2 cleavage site to give clone pT7-X1-Pro^{ΔQ326}. The potential X1-X2 cleavage site of the encoded polyprotein X1-X2-NTB-VPg-Pro was destroyed by deletion of the amino acid Q (P1 amino acid sequence 326) at the -1 position of the predicted dipeptide sequence. Translation of this mutant clone led to synthesis of a precursor polyprotein with the same molecular mass as that produced by the wild type clone (Fig. 4.4 B, lanes 1 and 3). Upon incubation of the mutated polyprotein, the smaller proteins, which were detected in the wild-type clone, were observed with the exception of the 29 kDa protein (Fig. 4.4 B, lanes 1-4; Fig. 4.4 C, lanes 7-10). This suggests that this protein was produced by processing at the X1-X2 cleavage site in the wild-type polyprotein and, accordingly, corresponded to the mature X2 protein. Mutation of the predicted X1-X2 cleavage site increased the production of all the cleavage products including the 44 kDa (X1-X2), 33 kDa [VPg-Pro(VS)], 66 kDa (NTB), 97 kDa [NTB-VPg-Pro(VS)] and 112 kDa (X1-X2-NTB) proteins. Immunoprecipitation experiments with antibodies against X1, antibodies against the N-terminal portion of X2, antibodies against NTB, antibodies against VPg and antibodies against Pro confirmed the nature of these smaller proteins including the 29 kDa protein (data not shown). These results suggest that the dipeptide sequence Q^{326}/E corresponds to the X1-X2 cleavage site.

4.4.2 P1 cleavage sites are processed intramolecularly in vitro

To investigate if P1 cleavage sites could be processed by the protease in trans, several clones containing different regions of P1 were constructed. Plasmid pT7-MPCATΔATG was used as a control for the protease activity. Translation of this plasmid led to generation of a precursor with a predicted molecular mass of 59 kDa containing the RNA-2-encoded MP-CP (movement protein and coat protein) cleavage site. Translation of all clones was arrested by addition of an excess of non-radioactive methionine, RNase and DNase, and incubated in the presence of the recombinant
Fig. 4.5. **The protease does not process P1 proteins in trans in vitro.** Plasmids pT7-X1-Pro$^{H128D}$, pT7-NTB-Pol$^{H128D}$, pT7-NTB-VPg, pT7-C-NTB-VPg and pT7-MPCAT were translated in the presence of [S$^{35}$]methionine as described in Material and Methods. The aliquots of the translations products were added either with 0.1 M Tris buffer (pH 8.0) (indicated as: -) or with 1 μl protease (indicated as: +) expressed and purified from *E. coli*. The samples were incubated for 20 hr at 16ºC, separated on 10% polyacrylamide-SDS gels, and visualized as described in Material and Methods. The positions of the precursor proteins and cleavage products (MP and CP-CAT) are indicated.
protease purified from *E. coli* (see Materials and Methods). After 20 hr incubation, the MPCAT precursor (59 kDa) was processed by the exogenously produced protease into two small proteins with apparent molecular masses of 33 and 26 kDa corresponding to the predicted sizes for the C-terminal moiety of the movement protein (MP) (33 kDa) and a fusion protein (CP-CAT) (26 kDa) containing the C-terminal moiety of the coat protein and the CAT protein (Fig. 4.5, lanes 9 and 10). However, cleavage by the exogenous protease was not detected in large polyprotein precursors containing an inactive protease generated by plasmids pT7-X1-Pro<sup>H1283D</sup> (containing the potential cleavage sites upstream of Pro) and pT7-NTB-Pol<sup>H1283D</sup> (containing all the potential cleavage sites downstream of NTB). To exclude the possibility that the mutated protease in the precursor generated by plasmid pT7-X1-Pro<sup>H1283D</sup> blocked the access of the exogenous protease to the substrate, other precursors including X1-X2, X1-X2-NTB and X1-X2-NTB-VPg were used in the protease trans-processing assay. Cleavage was not detected on these precursors (data not shown). The NTB-VPg present in infected plants (see below) may represent a precursor for the VPg protein. To test this hypothesis directly, two precursors were designed containing either the entire NTB domain or the C-terminal moiety of the NTB domain followed by the VPg protein (generated by plasmid pT7-NTB-VPg and pT7-C-NTB-VPg, respectively) (Fig. 4.5, lanes 1-8). Cleavage was not detected upon addition of the exogenous protease to those precursors. These results suggest that processing at P1 polyprotein cleavage sites is predominantly an intramolecular event.

4.4.3 Detection of P1 mature and precursor proteins in infected plant leaves

To study processing of P1 proteins *in vivo*, antibodies directed against fusion proteins containing different domains at the C-terminus of the P1 polyprotein, the middle portion of NTB, the middle sequence of VPg, the entire Pro and the entire Pol were used to identify viral proteins in leaf extracts of infected cucumber plants. Anti-NTB antibodies identified three viral proteins with apparent molecular masses of 66, 70 and 130 kDa (Fig. 4.6 B). The 66 and 70 kDa proteins corresponded to the predicted size for the mature NTB (66 kDa) and the NTB-VPg precursor (69 kDa).
Fig. 4.6. **Immunoblots analysis of TomRSV proteins in infected plants.** Leaf extracts (Infected: from infected plants, Mock: from healthy plants) were collected 2-3 weeks post-inoculation. Positions of standard protein markers are shown on the left and of the detected proteins are indicated on the right. Protein separation (12% polyacrylamide-SDS gel for A and 10% for B, C, D and E) and transfer were described in Material and Methods. Detection of the viral proteins was with A) anti-X1 antibodies, B) anti-NTB antibodies, C) anti-VPg antibodies, D) anti-Pro antibodies and E) anti-Pol antibodies.
The 130 kDa protein could either be X1-X2-NTB (predicted size: 132 kDa) or the X2-NTB-Pro (predicted size: 136 kDa). Two proteins of approximately 110 and 70 kDa corresponding to the predicted molecular mass for VPg-Pro-Pol precursor (111 kDa) and NTB-VPg (69 kDa), respectively, were recognized by anti-VPg antibodies (Fig. 4.6 C). Anti-Pro antibodies were immuno-reactive with two proteins with apparent molecular masses of 27 and 110 kDa corresponding to the predicted size for the mature Pro (27 kDa) and the VPg-Pro-Pol precursor (111 kDa) (Fig. 4.6 D). Two proteins of approximately 80 and 110 kDa were detected by anti-Pol antibodies (Fig. 4.6 E). The 80 kDa protein is of the expected size for the mature Pol while the 110 kDa protein is of the calculated molecular mass for the VPg-Pro-Pol precursor protein. Taken together these results suggest that the 70 kDa protein recognized by anti-NTB and anti-VPg antisera is the NTB-VPg precursor protein. The 110 kDa detected by anti-VPg, anti-Pro and anti-Pol is the VPg-Pro-Pol precursor protein. The 66 kDa protein identified by anti-NTB sera, the 27 kDa protein recognized by anti-Pro antibodies, and the 80 kDa protein detected by anti-Pol antibodies are the mature NTB, Pro and Pol proteins, respectively. The 130 kDa protein occasionally recognized by anti-NTB antibodies was never detected by other available antibodies. It was, therefore not possible to clearly determine its nature. The 66 kDa (NTB) and 70 kDa (NTB-VPg) viral proteins were consistently detected with the corresponding antibodies in extracts from a number of independent infected plants. The other viral proteins were occasionally detectable. To determine if these viral proteins are associated with membranes, sub-cellular fractionation experiments were carried out essentially as described (Ritzenthaler et al., 1995). It was found that all P1 proteins detected by the available antibodies were present in the Pe-30 fraction which mainly contains crude cellular membranes (data not shown).

To detect the viral proteins produced from the N-terminal part of P1 in infected plants, polyclonal antibodies were raised against a fusion protein containing the C-terminal half of the putative X1 protein fused at its N-terminus with the thioredoxin protein. The X1 domain included
in this fusion protein contained a region of 12 kDa with 77% amino acid identity with the corresponding region of the polyprotein encoded by RNA-2, and a region of 7 kDa at its C-terminus which is unique to the polyprotein encoded by RNA-1. Anti-X1 antibodies detected a protein with an apparent molecular mass of 36 kDa, which corresponds to the predicted size for the mature X1, and small amounts of a protein with an apparent molecular mass of 66 kDa, close to the predicted molecular mass (68 kDa) for a potential X1-X2 precursor (Fig. 4.6 A). The size of the proteins detected was consistent with the position of the cleavage sites detected in vitro on the P1 polyprotein. However, we can not exclude the possibility that these proteins might be cleavage products of the RNA-2 encoded polyprotein that could have been detected by anti-X1 antibodies through the region of sequence homology. To detect the X2 protein in infected plants, polyclonal antibodies against the N-terminal (X2.1) and C-terminal (X2.2) portions X2 protein were raised. The X2 protein or its precursors were not detected.

4.4.4 Microsequencing of VPg

To localize precisely the VPg in the P1 polyprotein, the VPg protein linked to the TomRSV RNA genome was purified from infected plants and subjected to N-terminal microsequencing. In Chapter 3, the VPg sequence was suggested to be 27 amino acids in length based on direct N-terminal microsequencing of the mature protease and by site-directed mutagenesis of the only potential cleavage site at its N-terminus. Direct microsequencing of the VPg protein identified a 26-amino-acid sequence from its N-terminus. The last amino acid was deduced to be a Q inferred from the TomRSV RNA-1 cDNA clone (Rott et al., 1995). Therefore, the VPg sequence was established to be:

```
1  5  18  24  25  27
STIPSGLSYADVYNARNMARRPRQPSVq
```

Amino acid sequence alignment analysis revealed that this sequence closely matches one portion of
amino acid sequence (P1 a.a. 1213-1239) deduced from the RNA-1 nucleotide sequence (TomRSV nt 3714 to nt 3794) (Rott et al., 1995). The only difference from the published data was a replacement of the Thr at position 18 (P1 a.a. 1230) with an Ala. This sequence is consistent with the previous findings that the dipeptide sequences of the NTB-VPg and VPg-Pro cleavage sites are Q^{1212}/S and Q^{1239}/G, respectively (Chapter 3).

4.5 Discussion

In this study, we have shown that the TomRSV RNA-1-encoded 27 kDa protease, the only known TomRSV protease of the virus, was responsible for cis-cleavage of P1 at five cleavage sites in vitro, i.e. the X1-X2, X2-NTB, NTB-VPg, VPg-Pro and Pro-Pol cleavage sites. Cleavage at an additional predicted cleavage site (X2.1-X2.2) was not detected in two different precursors (X2.1-X2.2-NTB-VPg-Pro and X1-X2.1-X2.2-NTB-VPg-Pro). Cleavage at the five identified cleavage sites is predicted to release the following mature proteins, starting at the N-terminus of P1, XI (35 kDa), X2 (32 kDa), NTB (66 kDa), VPg (3 kDa), 27 kDa (Pro) and 81 kDa (Pol) proteins. Of them NTB, VPg, Pro, Pol and possibly XI were detected in vivo. The X2 protein was not detected in infected plants using antibodies raised against the N-terminal or C-terminal moieties of X2. This could result from a low sensitivity of the antibodies or from a undetectable level of the protein in the plants. Further analysis of plant extracts with additional antibodies will be necessary to confirm the presence of the predicted XI and X2 in infected plants and to exclude the possibility that other cleavage sites not recognized in vitro may be used in vivo.

The overall genomic organization of TomRSV RNA-1 is very similar to that of the closely related como- and other nepovirus RNA-1 with the following order of the protein domains, X (or Co-Pro, protease co-factor)-NTB-VPg-Pro-Pol (Rott et al., 1991a, 1991b and 1995). TomRSV, however, has two protein domains, X1 and X2 upstream of NTB on the RNA-1-encoded polyprotein. The N-terminal portion of the P2 polyprotein contains a large region of amino acid sequence identity
with that of X1. It would be reasonable to suggest that the X1 protein and the N-terminal protein of P2, similar to the 58 kDa protein of CPMV RNA-2 (Van Bokhoven et al., 1993), may play a cis-role in RNA replication. The amino acid sequence of X1 does not present sequence homology with other proteins with known or putative functions (Rott et al., 1995). However, homologies are found between the middle part of the TomRSV X1 protein and the N-terminal portion of the X protein located upstream of NTB on the TBRV and GCMV RNA-1-encoded polyprotein. No homologies are found between TomRSV X1 and the X protein on the GFLV RNA-1-encoded polyprotein. TomRSV, TBRV and GCMV are members of nepovirus subgroup b and c while GFLV is a member of nepovirus subgroup a. It is not known whether other members of subgroup b and c encode similar sequences. Conserved sequences are found among the X2 protein of TomRSV, the 32 kDa protein (Co-Pro) of CPMV and the X protein (N-terminal protein) of all other nepovirus RNA-1-encoded P1 polyproteins sequenced to date including GFLV, GCMV and TBRV (Rott et al., 1995). We would therefore like to suggest that there are two distinct protein domains upstream of NTB on the polyprotein encoded by RNA-1. These domains probably serve distinct functions in the replication of nepoviruses. In TomRSV, a subgroup c nepovirus, these domains are present as two separate mature proteins (X1 and X2) produced via proteolytic processing although the X1-X2 precursor was also detected in vitro and possibly in vivo. In GCMV and TBRV (two subgroup b nepoviruses), these domains are probably present on the same mature protein (Hemmer et al., 1995). In GFLV (a subgroup a nepovirus) and in CPMV (a comovirus), only the second domain (X2) is present on the polyprotein.

The 32 kDa protein of CPMV has been shown to act as a protease co-factor. The CPMV RNA-1-encoded 200 kDa polyprotein is co-translationally processed into a 32 kDa protein (Co-Pro) and a 170 kDa intermediate precursor protein (NTB-VPg-Pro-Pol) by the 3C-like protease (Frassen et al., 1984; Peng and Shih, 1984). Subsequent post-translational processing of the 170 kDa is slowed by the 32 kDa protein (Co-Pro) (Peters et al., 1992a). The 32 kDa protein enhances
processing efficiency at the cleavage site between the movement protein and the coat protein (MP-CP) (Vos et al., 1988). The X2 protein of TomRSV is, however, unlikely to play a similar role. In this study, release of NTB and VPg-Pro was efficient in all the precursors either in the presence or absence of the X2 protein. Previous results also showed that the protease efficiently processed the MP-CP cleavage site in the absence of the X2 domain (Hans and Sanfaçon, 1995). Similar results were obtained in other nepoviruses (Margis et al., 1994; Hemmer et al., 1995). The conserved sequences among the X2 protein of TomRSV, the 32 kDa protein of CPMV and the X protein of TBRV, GCMV and GFLV may therefore possess an alternative unknown function perhaps in RNA replication.

In picorna-, poty- and comoviruses, full length polyproteins are difficult to detect due to rapid co-translational processing into primary products, which are subsequently post-translationally cleaved to release functional mature and intermediate precursor proteins (Chapter 2). In this study, processing of either N-terminal or C-terminal portions of the TomRSV P1 polyprotein, however, was very slow during in vitro translation, indicating processing of P1 may be mainly post-translational at least in vitro. Inefficient processing of polyproteins during translation was also observed in vitro for another nepovirus, TBRV (Demangeat et al., 1990; Hemmer et al., 1995; Mayo and Robinson, 1996).

We show that the TomRSV NTB-VPg precursor protein accumulated at a high level in infected plants. Nepo- and comovirus NTB proteins contain a transmembrane domain similar to the 3A protein of poliovirus and the 6K protein of potyviruses (Frassen et al., 1984; Gorbalenya et al., 1989; Peters et al., 1994; Rott et al., 1995). Virus replication is thought to occur on cellular membranous structures through the formation of a replication complex. The poliovirus 3A (or 3A precursors), potyvirus 6K (or 6K precursors) and the como- and nepovirus NTB (or NTB precursors) have been proposed to be responsible for the attachment of the replication complex to cellular membranes (Sanfaçon, 1995; Goldbach and Wellink, 1996; Towner et al., 1996; Schaad et al.,
Fig. 4.7. A model for processing of the RNA-1-encoded polyprotein P1. The precursor polyprotein P1 is represented in the top solid line. The identified cleavage sites are indicated by short vertical lines and the corresponding dipeptide sequences are given. Each predicted or determined domain and its calculated molecular mass are indicated. All predicted intermediate precursors and mature proteins are shown.
Like the TomRSV NTB-VPg precursor, 3AB of poliovirus and NTB-VPg of CPMV are also abundant in infected cells (Semler et al., 1982; Peters et al., 1992b). 3AB has been suggested to be the precursor protein which delivers the VPg, the primer for RNA replication (Paul et al., 1998). However, the NTB-VPg of CPMV and of two nepoviruses, i.e. TomRSV and TBRV is not likely to be the donor of VPg. Unlike 3AB, trans-proteolytic processing of NTB-VPg by the CPMV and TomRSV proteases is not detectable (Peters et al., 1992b; Hemmer et al., 1995; this study). VPg-Pro-Pol has been suggested as a possible donor of VPg (Peters et al., 1995). Consistent with this suggestion, we have previously found that the TomRSV VPg-Pro-Pol precursor could be processed in E. coli into VPg-Pro and Pol, and that the VPg-Pro intermediate precursor could be further cleaved into VPg and Pro (Chapter 3), indicating that the possible precursor for VPg is VPg-Pro. The biological functions of the abundant NTB-VPg precursors in infected plants need to be investigated.

In this Chapter, we have shown that P1 cleavage sites were not efficiently cleaved in trans by the protease. Similar results were obtained for TBRV, a nepovirus of subgroup b (Hemmer et al., 1995). In contrast, trans-cleavage was observed at the X-NTB cleavage site of GFLV, a nepovirus of subgroup a (Margis et al., 1994). In the TomRSV NTB-VPg-Pro-Pol precursor, the protease was likely to release NTB, then Pol, and finally VPg and itself. In this sequential order of processing, the final cleavage was to auto-cleave the VPg-Pro precursor (Chapter 3). Since the protease can only efficiently process P1 in cis, complete processing of the N-terminal portion of P1 is probably stepwise too. Taken the above in vitro and in vivo results together, a possible model for the expression and processing of TomRSV RNA-1 is shown (Fig. 4. 7). This model is based on the in vitro processing results obtained from partial P1 precursor proteins including the C- and N-terminal portions of the P1 polyprotein. It would be interesting to directly test if the processing of the whole P1 polyprotein in vivo follows this model.
CHAPTER 5

MEMBRANE-ASSOCIATED GLYCOSYLATION AND SIGNAL PEPTIDASE PROCESSING OF TOMATO RINGSPOT VIRUS NTB-VPg PROTEIN IN VITRO
5.1 Abstract

Single-stranded positive sense RNA viruses are thought to assemble their replication complexes on cellular membranes for genome replication. The mechanism by which the replication complex associates with membranes is, however, poorly understood. The association of the NTB-VPg protein with cellular membranes was studied in vitro. This protein contains a predicted transmembrane motif at the C-terminus of the NTB domain and is abundant in infected plants. It was found that the NTB-VPg protein was an integral membrane protein and associated with microsomal membranes in vitro, resulting in the generation of glycosylated and signal peptidase-processed NTB-VPg proteins. Site-directed mutagenesis experiments revealed that the membrane association required the presence of the predicted transmembrane domain at the C-terminal portion of NTB and that the glycosylation occurred at a predicted glycosylation site on the VPg sequence. These results suggest that the NTB-VPg precursor protein may play a role as an anchor responsible for the attachment of the putative replication complex to cellular membranes as suggested previously for the 6k protein of potyviruses and the 3A protein of poliovirus.

5.2 Introduction

We have shown in previous chapters that the TomRSV P1 polyprotein is processed by the 3C-like protease to release at least six end products, starting from the N-terminus of P1, X1 (36 kDa), X2 (32 kDa), the putative NTP-binding protein (NTB, 66 kDa), the VPg (VPg, 3 kDa), the protease (Pro, 27 kDa) and the putative RNA-dependent RNA polymerase (Pol, 81 kDa), and several intermediate precursor proteins including X1-X2, NTB-VPg and VPg-Pro. Of these viral proteins, the NTB-VPg protein is relatively abundant in infected plants and is found in the Pe-30 fraction (mainly containing crude membranes) upon subcellular fractionation (Chapter 4). The predicted
NTB domain contains 592 amino acids and has a calculated molecular mass of 66.4 kDa (Chapter 4). The N-terminal part of NTB embraces the highly conserved “A” and “B” motifs, which are present in NTP-binding proteins (Gorbalenya and Koonin, 1989; Gorbalenya et al., 1989; Rott et al., 1995). The C-terminal portion of NTB has a stretch of 14 hydrophobic amino acids (amino acids 1169-1182 of P1) which are predicted to constitute a transmembrane domain (Argos et al., 1984; Rott et al., 1995). The 27 amino-acid VPg protein is located between the NTB protein and the 3C-like protease (Chapter 4). The VPg protein is very basic and hydrophilic and contains a potential glycosylation site.

Studies on related viruses have shown that NTB (equivalent to the 2C-3A precursor in picornaviruses and to the CI-6K precursor in potyviruses), VPg, Pro and Pol are required for virus replication (Dorssersers et al., 1984; Bienz et al., 1990; Giachetti and Semler, 1991; Bienz et al., 1992; Klein et al., 1994; Peters et al., 1994; Li and Carrington, 1995). These viral proteins are present in a membrane-bound replication complex either as mature proteins or in the form of precursor proteins, along with host factors and other required components. The NTB protein in the comoviruses and nepoviruses, the 3A protein in picornaviruses and the 6k protein in potyviruses all contain a predicted membrane-binding domain and have been suggested to act as an anchor for binding of the replication complex to membranous structures (Eggen and Van Kammen, 1988; Van Bokhoven et al., 1992; Sanfaçon, 1995; Towner et al., 1996; Schaad et al., 1997). Direct evidence for association of the potyvirus 6k protein and the poliovirus 3A protein with cellular membranes has recently been published (Towner et al., 1996; Schaad et al., 1997). However, the possible association of the nepovirus-encoded NTB proteins with membranes has not been investigated.

In this study, we present evidence that the NTB-VPg precursor protein associates with microsomal membranes in vitro. This association leads to glycosylation and signal peptidase cleavage of the protein. We show that the predicted transmembrane domain and the VPg sequence are both required for the modifications. We demonstrate that the NTB-VPg is an integral membrane
protein.

5.3 Materials and Methods

5.3.1 Plasmid construction and site-directed mutagenesis

Plasmid pMR10, a full cDNA clone of TomRSV RNA1 was obtained from Rott et al. (1995).

Table 5.1 Primers used for construction of plasmids and site-directed mutagenesis

<table>
<thead>
<tr>
<th>Name</th>
<th>Polarity</th>
<th>Sequence (5' to 3')</th>
<th>Corresponding Sequence of RNA-1</th>
<th>Restriction site introduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>W027</td>
<td>-</td>
<td>ACGCGTCGACCTGAGTTGGGCTCGTC</td>
<td>31937-1922</td>
<td>Sal I</td>
</tr>
<tr>
<td>W028</td>
<td>+</td>
<td>ACGCCCATGGAATTAAGTGCTGAGTTGT</td>
<td>2948-2967</td>
<td>Nco I</td>
</tr>
<tr>
<td>W029</td>
<td>+</td>
<td>ACGCCCATGGAGTTTCTTTCTCTTG</td>
<td>441-459</td>
<td>Nco I</td>
</tr>
<tr>
<td>W030</td>
<td>+</td>
<td>ACGCCCATGGTTCTCGAGATATCATGA</td>
<td>1959-1977</td>
<td>Nco I</td>
</tr>
<tr>
<td>W031</td>
<td>-</td>
<td>ACGCGTCGACCATTTCCTCCACACGAGCC</td>
<td>3704-3686</td>
<td>Sal I</td>
</tr>
<tr>
<td>W035</td>
<td>-</td>
<td>AACATGGCACGCGTTTTCCGCAC</td>
<td>6696-6682</td>
<td>Sal I</td>
</tr>
<tr>
<td>W040</td>
<td>+</td>
<td>AACATGCGACCGGTGTTGCCCAC</td>
<td>3759-3783</td>
<td></td>
</tr>
<tr>
<td>W041</td>
<td>-</td>
<td>GCACACATGGACACGTCT</td>
<td>3758-3740</td>
<td></td>
</tr>
<tr>
<td>W042</td>
<td>+</td>
<td>ATGACACCGGTGTTGCACGCA</td>
<td>3762-3780</td>
<td></td>
</tr>
<tr>
<td>W043</td>
<td>+</td>
<td>TCTGCTTGTATAAGTTGATGC</td>
<td>3624-3645</td>
<td></td>
</tr>
<tr>
<td>W044</td>
<td>-</td>
<td>GAGCTTACACCGGTGAGGC</td>
<td>3581-3564</td>
<td></td>
</tr>
</tbody>
</table>

The sequences for engineering restriction site are underlined and A-G substitution in primer W040 is indicated in italic.

To obtain plasmid pT7-X1-X2 containing the coding sequence for the X1-X2 precursor protein, primers W029 and W027 were used to amplify a fragment of pMR10 (TomRSV RNA-1 nt 441-1937). The amplified fragment was digested with Nco I and Sal I, and ligated into the Nco I-Sal I sites of pCITE-4a(+) (Novagen).

The NTB coding region from pMR10 (TomRSV RNA-1 nt 1959-3704) was amplified using PCR primers W030 and W031. The amplified fragment was cleaved with Nco I and Sal I, and cloned into the corresponding sites of pCITE-4a(+) to generate plasmid pT7-NTB. To delete the N-terminal
portion of NTB, primers W028 and W031 were used to amplify a fragment of pMR10 (TomRSV RNA-1 nt 2948-3704). The amplified fragment was digested with Nco I and Sal I and inserted into the Nco I-Sal I sites of pCITE-4a(+) to produce plasmid pT7-C-NTB. To further delete the predicted transmembrane domain of C-NTB, plasmid pT7-C-NTB(dC) containing the coding sequence of TomRSV RNA-1 nt 2948-3393 was constructed by self-ligation of the large Xho I fragment of plasmid pT7-C-NTB.

Plasmids pT7-NTB-VPg and pT7-C-NTB-VPg were described previously (Chapter 4). To substitute amino-acid T (a.a. 1230 of the P1 sequence) with an A at the predicted glycosylation site of the VPg sequence (pT7-C-NTB-VPg\textsuperscript{T1230A}), phosphorylated primers W040 and W041 were used to amplify plasmid pT7-C-NTB-VPg. The amplified fragment was self-ligated to produce the clone. To delete the amino acid N (a.a. 1228 of the P1 sequence) at the predicted glycosylation site, plasmid pT7-C-NTB-VPg was amplified using phosphorylated primers W041 and W042. The amplified fragment was self-ligated to produce plasmid pT7-C-NTB-VPg\textsuperscript{AN122S}. Removal of the entire region of the predicted transmembrane domain (a.a. 1169-1182 of the P1 sequence) to produce clones pT7-C-NTB-VPg\textsuperscript{A1169-1182} was achieved by a similar PCR strategy using oligonucleotides W043 and W044 as primers and plasmid pT7-C-NTB-VPg as a template.

To construct plasmid pT7-Pol, a fragment of pMR10 (TomRSV RNA nt 2948-6696) was amplified with primers W028 and W035. This fragment was digested with BamH I and Sal I to generate a new fragment of 2.3 kb (TomRSV RNA-1 nt 4436-6696) encoding the C-terminus of the protease (33 amino acids) and the entire Pol. The 2.3 kb fragment was ligated into the corresponding sites of vector pCITE-4a(+) (Novagen).

5.3.2 \textit{In vitro} transcription and translation

Approximately 1 µg of each plasmid DNA was transcribed and translated at 30°C for 1 hr unless otherwise stated using a TNT rabbit reticulocyte lysate system (Promega) with or without 3.6 equivalents of microsomal membranes (Promega) in 12.5 µl reaction solutions. The synthesized
proteins were labeled with $^{35}$Smethionine (NEW). The reaction was separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and visualized as described (Chapters 3 and 4).

5.3.3 Deglycosylation assay

*In vitro* translation products (25 μl reaction) were microcentrifuged at 4°C for 15 min at 7000 g. The pellet fraction (membrane-bound fraction) was denatured at 95°C for 3 min in the presence of 10 μl of 1X protein loading buffer (Chapter 3), and then diluted with 90 μl of deglycosylation assay buffer [200 mM potassium phosphate buffer, pH 7, 25 mM ethylenediaminetetraacetic acid (EDTA), and 1% Triton-100]. Aliquots of 25 μl of the solution were supplemented with 0, 5, 25 and 50 mU of N-Glycosidase F (Boehringer), and incubated at 30°C overnight. Equal amounts of the deglycosylated products from each treatment were denatured with protein loading buffer and separated by SDS-PAGE for visualization. The quantitative measurements were as described (Chapter 3).

5.3.4 Direct membrane binding assay

*In vitro* translation products synthesized in the presence of microsomal membranes were obtained as above. Each reaction was arrested as described (Chapter 4) and microcentrifuged at 4°C for 15 min at 7000 g to separate soluble and particulate (membrane-bound) fraction. The pellet fraction was resuspended in 0.1 M Tris-HCl (pH 8.0) buffer and brought to a volume equal to that of the soluble fraction. Both fractions were subjected to SDS-PAGE for separation and visualization of the proteins.

5.3.5 Biochemical treatment of the membrane-bound fraction

Biochemical treatment of the membrane-bound fraction obtained as above was essentially as described (Schaad et al., 1997). The fraction was resuspended with the following solutions: alkaline buffer [0.1 M Na$_2$CO$_3$, pH 10.5, 4 mM EDTA and 4 mM phenylmethylsulfonyl fluoride (PMSF)], urea (25 mM HEPES, pH 6.8, 4 mM EDTA, 4 mM PMSF and 4 M urea) and salt (50 mM
Tris-HCl, pH 7.4, 15 mM MgCl₂, 10 mM KCl, 20% glycerol, 0.1% β-mercaptoethanol, 5 µg/ml leupeptin, 2 µg/ml aprotinin and 1 M KCl). The resuspended preparations were incubated on ice for 30 min and then subjected to centrifugation at 20,000 g at 4°C for 30 min. The pellets (particulate fraction) were resuspended in protein loading buffer in volumes equal to those of the corresponding supernatants. Equivalent amounts of samples were analyzed by SDS-PAGE.

Triton X-114 treatment of the membrane-bound fraction was carried out following the method of Bordier (1981). The final aqueous and detergent-soluble phases were brought to an equal volume with protein loading buffer and subjected to SDS-PAGE.

5.4 Results

5.4.1 The NTB-VPg protein associates with microsomal membranes in vitro

We have previously shown that the NTB-VPg precursor and the NTB mature protein are abundant in infected plants. These proteins contain a predicted transmembrane domain. To test the ability of proteins to associate with membranes in vitro, canine microsomal membranes were used. These membrane preparations contain membranes derived from the endoplasmic reticulum (ER) and represent authentic versions of ER (Dallner, 1974). Plasmid pT7-NTB-VPg was expressed in vitro in the presence of [³⁵S]-methionine using a coupled transcription-translation system. Translation was performed in the absence or presence of canine microsomal membranes. The expected protein was produced in the absence of the microsomal membranes with an apparent molecular mass corresponding to the predicted size for NTB-VPg (Fig. 5.1, lane 3). When the NTB-VPg precursor was synthesized in the presence of microsomal membranes, two additional proteins were observed (Fig. 5.1, lane 4). The mobility of the larger one, which was glycosylated, was about 3 kDa more than the predicted molecular mass of NTB-VPg and that of the smaller one, which was likely to be signal peptidase-processed, was about 5 kDa less than the precursor. To test whether these membrane-induced modifications were specific to the NTB-VPg precursor, other viral proteins
Fig. 5.1. The NTB-VPg precursor undergoes modification in the presence of microsomal membranes in vitro. [S35]methionine-labelled in vitro translation products of plasmids pT7-X1-X2, pT7-NTB-VPg, pT7-VPg-Pro, pT7-Pro and pT7-Pol were prepared in the presence (M) or absence (−) of canine microsomal membranes as described in Material and Methods. Translation products were separated on a 10% polyacrylamide-SDS gels and visualized as described in Material and Methods. The positions of unmodified proteins are indicated on the left side of the gel. The modified proteins from NTB-VPg precursor are shown (NTB-VPg: glycosylated NTB-VPg, pNTB-VPg: signal peptidase-processed NTB-VPg).
including X1-X2, VPg-Pro, Pro and Pol were tested. These viral proteins including the VPg-Pro precursor were not modified in the presence of microsomal membranes (Fig. 5.1, lanes 1-2 and 5-10). These results suggest that NTB-VPg can associate with membranes, resulting in modifications of the protein and that the NTB sequence is required for such an association.

5.4.2 The VPg sequence and a predicted transmembrane domain at the C-terminus of NTB are required for the membrane induced modifications

To determine which regions of the NTB-VPg precursor are required for the modifications, a series of site-directed mutagenesis experiments were performed. To test if the VPg sequence is required, plasmid pT7-NTB was translated in the presence or absence of the membrane preparations. Modified proteins were not produced in the presence of membranes, suggesting that NTB alone could not be modified by canine membrane preparations (Fig. 5.2, lanes 3 and 4). To test if the N-terminal region of NTB was required for the membrane-induced modifications, plasmid C-NTB-VPg containing the C-terminal region of NTB and the entire VPg sequence was constructed. Synthesis of the C-NTB-VPg precursor in the absence of the membranes resulted in the production of only one primary protein with an apparent molecular mass of 36 kDa (Fig. 5.2, lane 5). In contrast, when this precursor was synthesized in the presence of the membranes, three prominent proteins with apparently molecular masses of 39, 36 and 31 kDa were produced (Fig. 5.2, lane 6). The 36 kDa protein was of the predicted size for the C-NTB-VPg. Consistent with the size of the modified proteins derived from the entire NTB-VPg precursor, the 39 kDa protein was probably a glycosylated protein while the 31 kDa was probably generated by signal peptidase cleavage at C-terminus (near the NTB-VPg cleavage site) of the C-NTB-VPg precursor protein. To further test whether the VPg sequence was required for the modifications of this precursor, a new plasmid was constructed in which the VPg domain of the C-NTB-VPg precursor was deleted. In an additional construct, the VPg domain along with a region of 7 kDa of the C-terminus of NTB was removed from the C-NTB-VPg precursor. *In vitro* transcription and translation of these two constructs led to
Fig. 5.2. The C-terminus of NTB and the VPg sequence are required for the membrane-induced modification. (A) Schematic diagram of the different constructs derived from TomRSV RNA-1 cDNA. Clones are indicated on the left (the small box represents the deleted transmembrane domain). The corresponding proteins are represented on the right. The expected molecular mass of each protein domain is indicated (VS: vector sequence). (B) [S\textsuperscript{35}]methionine-labelled in vitro translation products of plasmids were prepared in the presence (M) or absence (−) of membranes as described in Material and Methods. Translation products were separated on a 10% polyacrylamide-SDS gel and visualized as described in Material and Methods. The positions of unmodified proteins and modified proteins from NTB-VPg or C-NTB-VPg precursors are shown (NTB-VPg or gC-NTB-VPg : glycosylated NTB-VPg or glycosylated C-NTB-VPg, pNTB-VPg or pC-NTB-VPg : signal peptidase-processed NTB-VPg or signal peptidase-processed C-NTB-VPg) on the right side of the gel. The position of molecular weight markers is indicated on the left side of the gel.
synthesis of the proteins with the predicted sizes (Fig. 5.2, lanes 7 and 11). However, modification was not detected when these deletion proteins were translated in the presence of the membranes (Fig. 5.2, lanes 8 and 12). These results support the suggestion that the VPG protein is required for the modification.

Previous sequence analysis showed that NTB contained a predicted transmembrane domain of 14 hydrophobic residues (P1 sequence 1169-1182) at its C-terminus (NTB: P1 sequence 621-1212) (Argos et al., 1984; Rott et al., 1995). The finding that the deletion of the N-terminal portion of the NTB did not affect the association of the NTB-VPG precursor with the membranes suggests that the predicted transmembrane was probably responsible for membrane association and therefore essential for the modification. To test this hypothesis, a new plasmid was constructed in which the 14 hydrophobic residues were deleted from the precursor C-NTB-VPG. Synthesis of this precursor in the presence of the membranes did not result in the modification (Fig. 5.2, lanes 9 and 10), suggesting that the predicted transmembrane domain is essential for the membrane-induced modification.

5.4.3 The NTB-VPG precursor protein is an integral membrane protein

A direct binding assay was conducted to examine if the modified proteins were bound to the membranes. In this assay, plasmid pT7-C-NTB-VPG was translated in the presence of microsomal membranes. After translation, the synthesized products were microcentrifuged to separate the membrane-bound and soluble fractions. These fractions were analyzed by SDS-PAGE. All modified proteins were detected in the whole or the membrane-bound fraction but not in the soluble faction (Fig. 5.3, lanes 1-3). Therefore, the modified proteins were distributed only to the membrane-containing pellet fraction. This suggests that all modified proteins were associated with the membranes.

To test whether the NTB-VPG protein is an integral or peripheral membrane protein, biochemical analyses were conducted using the membrane fraction obtained above (Fig. 5.3, lane
Fig. 5.3. Direct membrane binding assay. *In vitro* transcription-translation of pT7-C-NTB-VPg was performed in the presence of [S\(^{35}\)]methionine and canine microsomal membranes as described in Material and Methods. Particulate (membrane associated) and soluble fractions were separated by centrifugation and analyzed by SDS-PAGE (10% gel). W, S and P indicate non-separated lysates, soluble and particulate fraction, respectively. The glycosylated (g-subscript), signal peptidase-processed (p-subscript) and unmodified proteins are indicated.
3). After treatment with 0.1 M Na₂CO₃ (pH 10.5), 4 M urea or 1 M KC1, all the membrane-associated proteins were found predominantly in the pellet fraction (membrane fraction) (Fig. 5.4, lanes 1-6), suggesting that the modified or unmodified NTB-VPg proteins were still tightly associated with the membranes. A phase separation experiment was performed using Triton X-114 to further determine the nature of this association. It was found that the majority of the proteins partitioned to the detergent phase (Fig. 5.4, lanes 7-9). Taken together these results suggest that the NTB-VPg protein is an integral protein.

5.4.4 The glycosylated protein is deglycosylated by N-glycosidase F

To confirm that the larger modified protein is a glycosylated protein, a deglycosylation assay was conducted using N-glycosidase F. N-glycosidase F can remove all types of asparagine bound (N-linked) carbohydrate chains provided that the amino-group as well as the carboxyl group is present in a peptide linkage and that the oligosaccharide has the minimum length of the chitobiose core unit. After overnight incubation with 50 mU of N-glycosidase F, the putative glycosylated protein was not detected (Fig. 5.5, lane 4). Trace amounts of the putative glycosylated protein were observed upon incubation with 5 mU or 25 mU of N-glycosidase F (Fig. 5.5, lanes 2 and 3), whereas, the intensity of the unmodified precursor protein significantly increased (5.5, lanes 1-4). Quantitative analysis of the amount of each protein, i.e. the putative glycosylated protein, the signal peptidase-processed protein and the unmodified precursor protein, revealed that after incubation with N-glycosidase F, the amount of the predicted signal peptidase-processed protein remained unchanged (data not shown). The increase in the amount of the precursor protein corresponded to the decrease in the amount of the putative glycosylated protein (quantitative data not shown, Fig. 5.5, lanes 1-4). These results suggest that glycosylation took place during the membrane association and that the signal peptidase-processed protein was not glycosylated.

5.4.5 The glycosylation consensus sequence on VPg is required for the glycosylation and the signal peptidase processing probably takes place at the junction of the NTB and VPg proteins
Fig. 5.4. Biochemical treatment of membrane-associated proteins. *In vitro* transcription-translation of plasmid pT7-C-NTB-VPg was performed in the presence of [S\(^{35}\)]methionine and canine microsomal membranes as described in Material and Methods. The membrane fraction was obtained by centrifugation and treated with biochemical solutions as described in Material and Methods. P and S represent particulate (membrane) and soluble fractions, respectively. T, DP and AP represent total fraction, detergent-soluble and aqueous phases after treatment with Triton X-114, respectively. Each fraction was analyzed by SDS-PAGE (10% gel). The glycosylated (g-subscript), signal peptidase-processed (p-subscript) and unmodified proteins are indicated.
Fig. 5.5. Deglycosylation assay. *In vitro* transcription-translation of plasmid pT7-C-NTB-VPg was performed in the presence of [S$^{35}$]methionine and canine microsomal membranes as described in Material and Methods. The translation products were denatured and diluted with deglycosylation buffer. N-Glycosidase F (N-Gly-F) was added to the solutions at the concentration indicated above each lane. The reactions were analyzed by SDS-PAGE (10% gel). The glycosylated (g-subscript), signal peptidase-processed (p-subscript) and unmodified proteins are indicated.
Computer-assisted analysis of the NTB-VPg sequence deduced from the published TomRSV sequence (Rott et al., 1995) revealed that there was a unique glycosylation consensus sequence (NXT/S) on the VPg sequence. To determine if the glycosylation occurred at the conserved asparagine, plasmid pT7-C-NTB-VPg<sup>ΔN1228</sup> was produced in which amino acid N at the predicted glycosylation site was deleted from the C-NTB-VPg precursor. Translation of this clone in the presence of the membranes did not result in the production of the glycosylated protein, suggesting that the deletion of N at the unique glycosylation consensus sequence of VPg prevented the glycosylation (Fig. 5.6 A, lane 4). Similar results were also obtained when another mutant precursor C-NTB-VPg(T/A) generated from plasmid pT7-C-NTB-VPg<sup>T1230A</sup> was used in which T (a.a. 1230 of PI sequence), the other amino acid of the glycosylation consensus sequence was substituted with an A (Fig. 5.6 A, lane 6). Signal peptidase-processed proteins were detected in both mutants, suggesting that signal peptidase cleavage was not coupled with glycosylation. To further detect all the cleavage products resulting from signal peptidase-processing, a time course experiment was performed using plasmids pT7-C-NTB-VPg and pT7-C-NTB-VPg<sup>T1230A</sup>. The translation products were analyzed on a 15%-polyacrylamide SDS gel. Upon translation of the wild-type clone for 15 min, two primary products, the precursor protein (36 kDa) and the glycosylated protein (39 kDa) were produced while the signal peptidase-processed protein (31 kDa) was barely detected (Fig. 5.6 A lane 7). In the mutated clone, only one primary protein was detected at this time point (Fig. 5.6 B, lane 13). As the translation time increased, the signal peptidase-processed protein (31 kDa) and another smaller protein with an apparent molecular mass of 8 kDa corresponding to the predicted size for the glycosylated VPg(VS) [VPg: 3 kDa, VS (vector sequence): 2 kDa] were detected in the wild-type clone (Fig. 5.6 B, lanes 7-10). In the mutated clone, the signal peptidase-processed protein (31 kDa) and another protein with an apparent molecular mass of 5 kDa corresponding to the predicted size for the unmodified VPg(VS) accumulated (Fig. 5.6 B, lanes 14-17). These modified proteins were not produced when these two clones were translated in the absence of microsomal
Fig. 5.6. Glycosylation and signal peptidase cleavage of a NTRB-Vpg precursor (C-NTNB-Vpg). In vitro transcription-translation of plasmids p17-CNTNB-Vpg and in the presence of [35S]methionine as described in Material and Methods. Translation products were analyzed on a 10% polyacrylamide-SDS gel. A time-course analysis of glycosylation and signal peptidase cleavage of the C-NTNB-Vpg precursor was performed for 15, 30, 60, and 120 min at each time point. Translation products were analyzed on a 15% polyacrylamide-SDS gel.
membranes (Fig. 5.6 B, lanes 11-13, 18-21). The signal peptidase-processing resulted in the release of two cleavage products, the 31 and 8 kDa proteins in the wild-type clone, and the 31 and 5 kDa proteins in the mutated clone (mutated at the glycosylation site), suggesting the 8 kDa protein was the glycosylated 5 kDa protein. Since our previous experiments have shown that glycosylation occurred on the VPg sequence, the signal peptidase-cleavage is likely to occur at or near the junction between the NTB and the VPg proteins. This experiment also showed that glycosylation occurred early during translation while the signal peptidase cleavage took place at a relatively later stage, raising the possibility that glycosylation may be a co-translational event while the signal peptidase cleavage may be a post-translational event. To test this possibility, a time-course pulse-chase experiment was performed. Upon translation of the C-NTB-VPg precursor from 3 to 15 min, the amounts of glycosylated and unmodified proteins progressively increased. Quantitative measurement results revealed that the ratio of glycosylated and unmodified proteins remained essentially the same during the course of translation (data not shown), suggesting that glycosylation is predominantly a co-translational event. The signal peptidase cleavage was not detectable after 15 min translation (Fig. 5.7 A). At this time point, translation was arrested by addition of cold membrane, RNase and DNase. Further incubation of the translation products allowed detection of the signal peptidase-processed protein (Fig. 5.7 B). The amounts of the signal peptidase-processed protein increased during the incubation, indicating that the signal peptidase cleavage is a post-translational event.

5.5 Discussion

In this study, we show that the NTB-VPg protein associates with microsomal membranes as an integral membrane protein in vitro. We demonstrate that during the process of membrane association, this protein is modified. The NTB-VPg protein contains an anchor sequence of 14 amino acids at the C-terminus of the NTB protein followed with a small VPg protein which is very basic and hydrophilic. Therefore, the NTB-VPg protein probably associates with membranes through the
Fig. 5.7. **Time-course and pulse-chase analyses of glycosylation and signal peptidase cleavage of the C-NTB-VPg precursor.** (A) Time-course analysis. *In vitro* transcription-translation of plasmid pT7-C-NTB-VPg was performed in the presence of [S\(^{35}\)]methionine and canine microsomal membranes as described in Material and Methods. The translation was arrested at 3, 6, 9, 12 and 15 min by addition of protein denaturing buffer (lanes 1-5). The reactions were analyzed by SDS-PAGE (10% gel) and autoradiography. The glycosylated (g-subscript) and unmodified proteins are indicated. (B) Pulse-chase analysis. [S\(^{35}\)]methionine labelling of the *in vitro* translation products was conducted for 15 min, after which excess nonlabeled methionine, RNase and DNase were added. Aliquots were removed at the times indicated (in min) after initiation of the chase and analyzed by SDS-PAGE (10% gel) and autoradiography. The glycosylated (g-subscript), signal peptidase-processed (p-subscript) and unmodified proteins are indicated.
hydrophobic anchor sequence of the C-terminus of the NTB protein, leaving most of the terminal polypeptide chain including the conserved “A” and “B” motif of the NTB protein and the VPg in the cytoplasm. Indeed, this hypothesis is consistent with our observation that association of the NTB-VPg protein with membranes did not protect the protein from proteinase K digestion (data not shown).

In eukaryotic cells, there are two possible pathways for proteins to target to the cytoplasmic membranes. The first, more commonly used pathway is a co-translational process in which a signal near the N-terminus of the nascent polypeptide is recognized by a 54-kDa polypeptide of the signal recognition particle (SRP) (Rapoport, 1992). Prior to termination of the translation, the nascent polypeptide is transported to the endoplasmic reticulum (ER) and translocated. The second process, independent of SRP, can occur both co- and post-translationally. This class of membrane proteins (tail-anchored proteins) lacks a signal sequence but contains a single hydrophobic domain near its C-terminus (Kutay et al., 1993 and 1995). In some of these proteins including synaptobrevin and cytochrome b5, the N-terminus (most of the polypeptide chains) is on the cytoplasm face of the membranes while the C-terminal anchor sequence is embedded within the membranes (Mitoma and Ito, 1992, Kutay et al., 1995). In other proteins of this class, both the C- and N-termini are on the cytoplasm face of the membranes while the transmembrane domain of the protein is inserted into the membranes. After targeting to the membranes, the proteins can be subject to modifications, i.e. glycosylation or signal-processing by the two known enzymes in the translocation site, the signal peptidase and the oligosaccharyltransferase (Rapoport, 1992). Our results suggest that the TomRSV NTB-VPg precursor may follow the SRP-independent mechanism of membrane targeting.

Our previous results have shown that the NTB-VPg precursor accumulates in infected plants. Similarly, the comovirus NTB-VPg precursor, the potyvirus 6K-NlαVPg and the poliovirus 3AB precursor proteins corresponding to the TomRSV C-NTB-VPg protein (C-terminal part of NTB and VPg) were detected in infected cells (Datta and Dasgupta, 1994; Peters et al., 1994; Restrepo-
Hartwig and Carrington, 1994). Recent studies on the 6 kDa protein have revealed that it is an integral protein and associates with the ER-derived membrane structures via its central 19 amino acid hydrophobic domain (Schaad et al., 1997). Targeting of the TEV replication complex to the sites of membranes is probably through post-translational interactions between the 6 kDa protein and ER. The poliovirus 3AB protein was also shown to bind to membranes in vitro and in vivo (Datta and Dasgupta, 1994; Towner et al., 1996). 3AB possibly in the form of larger 2C- and 3AB-containing precursor proteins may be the lipophilic anchor for the assembly of the RNA replication complex which carries out the replication (Datta and Dasgupta, 1994; Towner et al., 1996; Paul et al., 1998). The association of NTB-VPg precursors with membranes may therefore be a general feature among picorna-like viruses and may be essential for virus replication.

We show that upon association with microsomal membranes, the TomRSV NTB-VPg protein is glycosylated in vitro and that the glycosylation is probably a co-translational event. Previously the poliovirus 3AB protein was also found to be glycosylated in vitro. Moreover, treatment of infected cells with 6-diazo-5-oxo-L-norleucine (DON), an inhibitor of glycoprotein synthesis, significantly inhibits poliovirus RNA synthesis in vivo (Data and Dasgupta, 1994). A point mutation in the glycosylation site of 3AB is lethal to the virus (Datta et al., 1996). These results suggest that the glycosylation may be critical in poliovirus replication although the potential role for glycosylation is not clear. Examination of the NTB-VPg sequences of poty-, como- and nepoviruses revealed the presence of a potential glycosylation site in the proximity of the transmembrane domain. The predicted glycosylation sites were found either in the NTB or in the VPg domains (results not shown). The potential glycosylation sites were at least 12 residues away from the predicted membrane-anchor domain, corresponding to the required conditions for glycosylation (Nilsson and Von Heijne, 1993). This conservation among the picorna-like supergroup may indicate potential functional significance of glycosylation of a non-structural protein in virus replication. It is not known whether the TomRSV NTB-VPg protein undergoes glycosylation in infected cells. Further
experiments will be targeted to this question. It would also be interesting to determine whether the 6K-VPg protein of potyviruses and the NTB-VPg polypeptides of como- and other nepoviruses can be glycosylated *in vitro* and *in vivo*.

In this study, the NTB-VPg precursor was post-translationally processed at its C-terminus at or near the junction between the NTB and VPg by a host signal peptidase. Previously, signal peptidase cleavage was found to be involved in the maturation of several viral proteins including a membrane protein in flaviviruses (Yamshchikov and Compans, 1993; Stocks and Lobigs, 1995 and 1998). This signal peptidase cleavage plays an essential role on the regulation of processing of the flavivirus-encoded polyprotein, which is required to release the functional mature and intermediate proteins required for RNA replication. We have shown previously that the NTB-VPg precursor was not cleaved *in trans* by the TomRSV protease (Chapter 4). The potential biological importance of the signal peptidase cleavage needs to be elucidated.
CHAPTER 6

MOLECULAR VARIABILITY IN THE C-TERMINAL REGIONS OF POLYPROTEINS ENCODED BY TOMATO RINGSPOT NEPOVIRUS ISOLATES
6.1 Abstract

Tomato ringspot nepovirus (TomRSV) isolates differ in their biological properties such as natural hosts, symptom expression, geographical distribution and immunological reactions. This biological variability depends on the molecular genetic diversity among the isolates. In this study, we have analyzed the extent of genetic diversity at the level of the nucleotide sequence in four TomRSV isolates which exhibit different symptoms on cucumber. Using a reverse-transcription polymerase chain reaction (RT-PCR) method, fragments corresponding to the 3' terminal regions of RNA-1 and RNA-2 were amplified and sequenced. The 3' terminal region of RNA-1 encodes a virus genome-linked protein (VPg), a 3C-like protease (Pro) and a putative RNA-dependent RNA polymerase (Pol). The 3' terminal region of RNA-2 encodes the coat protein (CP). The nucleotide sequences and the deduced amino acid sequences of RNA-1 were more conserved than those of RNA-2. In particular, the VPg amino acid sequence was identical in all isolates studied despite some variations at the nucleotide sequence level. The overall nucleotide and amino acid sequences of two raspberry isolates and one peach isolate were highly similar (over 96% identity) to each other. In contrast, a grape isolate was less related to the other isolates (identity at the nucleotide level was 80% and at the amino acid level was 86%). The different levels of genetic diversity among the isolates correlated with the range of symptoms on cucumber plants.

6.2 Introduction

Since TomRSV was first isolated from field tomatoes in Indiana in the 1930's (Imle and Samson, 1937), a number of isolates have been reported. These isolates differ from each other in their natural hosts, symptom expression and geographical origins (Stace-Smith, 1984a; Edwardson and Christie, 1997). Different TomRSV isolates not only cause divergent symptoms in different
hosts, but in the same host also (Cadman and Lister, 1961). This biological variability depends on the molecular genetic diversity among the isolates.

Previously, the genomic RNA of a raspberry isolate from the Lower Mainland of British Columbia (BC), Canada was completely sequenced (Rott et al., 1991; 1995). As shown in previous chapters, RNA-1 encodes the P1 polyprotein containing the VPg protein, the protease (Pro) and a putative RNA-dependent RNA polymerase (Pol) at its C-terminus (downstream of the putative NTP-binding protein) which are required for RNA replication. RNA-2 encodes the P2 polyprotein containing the coat protein at its C-terminus (downstream of the movement protein) (Hans and Sanfaçon, 1995). Recently, the coat protein region of a peach isolate obtained from California has been sequenced and shown to present 97% identity to that of the raspberry isolate (Yepes et al., 1996). The extent of variability of other isolates in the coat protein region or in other domains of P1 and P2 has not been investigated.

In this chapter, we studied four isolates obtained from different hosts and geographical locations. We examined symptoms on cucumber plants infected with these isolates. Using RT-PCR, we detected viral genomic RNA in the nucleic acid extracts of infected leaf tissues. To study the extent of molecular variability among the isolates, we determined the sequence of the RNA-1 coding regions for the VPg, Pro and Pol proteins and the sequence of RNA-2 encoding for CP. In total, this represents approximately 40% of the coding sequence of the respective genomic RNAs.

6.3 Materials and Methods

6.3.1 Source of virus isolates

TomRSV isolates Rasp-1, GYV, PYB-1, PYB-2, T392 and Rasp-2 were used for this study. Rasp-1, raspberry isolate-1 obtained from Dr. R. Martin of the Pacific Agriculture Research Centre (PARC), Vancouver, B.C., was purified from infected raspberry in the State of Washington. GYV
(grape yellow vein virus), a grape isolate causing yellow vein symptoms was obtained from Dr. A. Rowhani, Department of Plant Pathology, the University of California at Davis (U.C. Davis). It belongs to a grape virus collection at U.C. Davis and was originally isolated from infected grape in California. PYB-1 (PYB: peach yellow bud mosaic virus), a peach isolate inducing yellow bud mosaic symptoms was also obtained from Dr. A. Rowhani at U.C. Davis and was collected from an infected orchard in the State of California. T392, obtained from Dr. R. Hamilton, belongs to a TomRSV collection at PARC with unrecorded origin. Rasp-2, raspberry isolate-2 was purified by Dr. R. Stace-Smith of PARC from infected raspberry plants grown in the Lower Mainland area of B.C. and was sequenced in Dr. D. Rochon’s laboratory (Rott et al., 1991 and 1995). PYB-2, another peach isolate originally obtained from California was partially sequenced (Yepes, et al. 1996).

6.3.2 Virus propagation and viral RNA extraction

Cucumber variety Straight 8 was used as the host for virus propagation. Cucumber plants were grown in a greenhouse until the seedlings reached the two-cotyledon stage. At this time, the cotyledons were dusted with Carborundum and mechanically inoculated with the virus isolates. Three weeks post-inoculation, one gram of the infected leaf tissues was ground in liquid nitrogen. Five hundred μl of a hot 50/50 mixture (80°C) of phenol and extraction buffer [100 mM LiCl, 100 mM Tris (pH 8.0), 10 mM ethylenediaminetetraacetic acid (EDTA) and 1% sodium dodecyl sulfate (SDS)] were added to the frozen samples. The mixture was thoroughly vortexed for 30 sec, and 250 μl chloroform was added. The solutions were microcentrifuged at 15,000 X g for 5 min. RNA in the aqueous phase was precipitated in the presence of 1/10 (v/v) of 3.0 M NaOAc and 2 volumes of 100% ethanol, and resuspended in 15 μl of RNase-free water.

6.3.3 cDNA synthesis, PCR amplification and DNA purification

A 12-μl reaction containing 2 μl of the above RNA extracts, 1 μl oligonucleotide dT_{12-18} (500 μg/ml) and 9 μl of sterile, distilled water was heated to 70°C for 10 min and quickly chilled on ice. Four μl of 5X first strand buffer [250 mM Tris-HCl (pH 8.3), 375 mM KCl and 15 mM
MgCl₂], 2 μl of 0.1 M dithiothreitol (DTT) and 1 μl of 10 mM dNTP mix (Life Technologies) were added to the solution. After incubation at 42°C for 2 min, 1 μl (200 units) of RNase H' Reverse Transcriptase (SUPERSCRIPT™ II) (Life Technologies) was added, and the solution was incubated at 42°C for a further 80 min to synthesize the first strand cDNA. The reaction was arrested by incubation at 70°C for 15 min.

Table 6.1 Primers used for cDNA synthesis, PCR amplification and DNA sequencing

<table>
<thead>
<tr>
<th>Name</th>
<th>Polarity</th>
<th>Sequence (5' to 3')</th>
<th>Corresponding sequence on TomRSV genome&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>W3520</td>
<td>+</td>
<td>GCTCACCCTACAGAATTGTCG</td>
<td>RNA-1 3520-3540</td>
</tr>
<tr>
<td>W048</td>
<td>-</td>
<td>CAATTGTCGACGTACGATTAG</td>
<td>RNA-1 7735-7716, RNA-2 6794-6775</td>
</tr>
<tr>
<td>W049</td>
<td>-</td>
<td>CAGCTTGTCAAAGGCTGGA</td>
<td>RNA-1 6765-6747, RNA-2 5824-5806</td>
</tr>
<tr>
<td>JC39</td>
<td>-</td>
<td>CTGGCAAGAGCAGAAAAGAG</td>
<td>RNA-1 4535-4513</td>
</tr>
<tr>
<td>W053</td>
<td>+</td>
<td>CCAATCTGGAAGGTGGTTACA</td>
<td>RNA-1 4394-4413</td>
</tr>
<tr>
<td>W054</td>
<td>+</td>
<td>ATTCGCCGAACCGGTGGAG</td>
<td>RNA-2 3925-3943</td>
</tr>
<tr>
<td>W055</td>
<td>+</td>
<td>GAAGTTATCAATGGGAGCGA</td>
<td>RNA-1 4896-4915</td>
</tr>
<tr>
<td>W056</td>
<td>+</td>
<td>TTCCTATGCTCGAATACCTT</td>
<td>RNA-1 5218-5237</td>
</tr>
<tr>
<td>W057</td>
<td>+</td>
<td>CAAGCCCCCTTGGATAATT</td>
<td>RNA-1 5644-5663</td>
</tr>
<tr>
<td>W058</td>
<td>-</td>
<td>TCCACGCCCTTGCCAAGGC</td>
<td>RNA-2 5344-5325</td>
</tr>
<tr>
<td>W059</td>
<td>+</td>
<td>AGTCACGTGGACGTGTGATAT</td>
<td>RNA-2 4376-4396</td>
</tr>
<tr>
<td>W062</td>
<td>+</td>
<td>ATGGCAAGAGGGGCTTTTC</td>
<td>RNA-2 4894-4912</td>
</tr>
<tr>
<td>W063</td>
<td>+</td>
<td>CTTGCCCTTGGCGGAAGGGT</td>
<td>RNA-2 5322-5341</td>
</tr>
<tr>
<td>W064</td>
<td>-</td>
<td>AGAGAAAAAGCCCTCCCTCC</td>
<td>RNA-2 4915-4896</td>
</tr>
<tr>
<td>W065</td>
<td>-</td>
<td>CCAATATCAAAGGTCCTCAGT</td>
<td>RNA-2 4399-4980</td>
</tr>
<tr>
<td>W066</td>
<td>+</td>
<td>GTTCCCCTATGAACAATATTC</td>
<td>RNA-1 6107-6127</td>
</tr>
<tr>
<td>W067</td>
<td>-</td>
<td>TCTGGATCTTTATTACGCAA</td>
<td>RNA-1 6277-6258</td>
</tr>
<tr>
<td>W068</td>
<td>-</td>
<td>GGTCTTGCCCTAAAAAAGG</td>
<td>RNA-1 5833-5814</td>
</tr>
<tr>
<td>W069</td>
<td>-</td>
<td>GCAGTCGATGCCTATTAAAC</td>
<td>RNA-1 5299-5280</td>
</tr>
<tr>
<td>W070</td>
<td>-</td>
<td>CTGTCCCTGGAAGATATCCT</td>
<td>RNA-1 4168-4149</td>
</tr>
<tr>
<td>W071</td>
<td>+</td>
<td>CTTGACTAAAAATGACAGGCC</td>
<td>RNA-1 3913-3932</td>
</tr>
<tr>
<td>W072</td>
<td>-</td>
<td>AACCAAAATCGCTGCCATTGAT</td>
<td>RNA-1 4923-4902</td>
</tr>
</tbody>
</table>

<sup>a</sup>a raspberry isolate (Rasp-2) of TomRSV sequenced by Rott <i>et al.</i> (1991, 1995)

To synthesize the second strand of the cDNA and to amplify different regions of double stranded cDNA, the single stranded cDNA (from the first strand reaction) was used for PCR-amplification with three pairs of primers, i.e., W3520 and W048, W3520 and W049, and W054 and
PCR-amplification was carried out using a RoboCycler 40 temperature cycler (Stratagene) in a reaction mixture containing 2 µl of the first strand cDNA solution, 2 µl of Pfu DNA polymerase (2.5 U/µl, Stratagene), 10 µl of 10X Pfu-PCR buffer (Stratagene), 1 µl of dNTP mixture (25 mM each NTP, Life Technologies), 1 µl of each primer (250 ng/µl) and 84 µl of sterile, distilled water. The PCR cycling protocol was 95°C for 1 min, 48°C for 1 min and 72°C for 8 min for 30 cycles.

The PCR-amplified products were separated by agarose gel electrophoresis. DNAs of the expected sizes were excised from the gel and eluted using the Qiaex II Agarose Gel Extraction Kit (Qiagen).

6.3.4 Sequencing and sequence analysis

The gel-eluted DNAs were sequenced with the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer) using an ABI PRISM 310 Genetic Analyser (Perkin Elmer) as described by the supplier. cDNA fragments were sequenced completely in both directions using primers listed in Table 6.1.

Sequence analyses were carried out using a DNA and protein sequence comparison software (Sequence Navigator, Applied Biosystems).

6.4 Results

6.4.1 Host plant responses

Four TomRSV isolates originally collected on raspberry (Rasp-1), peach (PYB-1), grape (GYV) and an unknown host (T392) were inoculated on cucumber. Five to seven days post-inoculation, local chlorotic lesions appeared on the inoculated cotyledons for all the isolates tested. The symptoms on true leaves, however, were distinct for the specific isolate tested. For isolates Rasp-1 and T392, the infected plants showed systemic symptoms including chlorotic lesions, mosaic, intraveinal chlorosis, sharper leaf serrations, necrotic lesions and distorted leaves, and the
Fig. 6.1. Symptoms caused in cucumber plants by TmRSV isolates. Pictures were taken 17 days post-inoculation. A) Healthy control.
plants were significantly stunted (Fig. 6.1 B and D). Three weeks post-inoculation, approximately 80% of the infected plants were dying (Fig. 6.1 B and D). In contrast, approximately 20% of the plants continued to grow and finally became symptom-free (Fig. 6.1 C). This type of host response is termed as recovery (Wingard, 1928; Ratcliff et al., 1997; Al-Kaff et al., 1998). For isolate PYB-1, symptoms of the infected plants were similar to those caused by Rasp-1 and T392. The only notable difference was that the plants infected with PYB-1 were slightly less stunted (Fig. 6.1 F). In sharp contrast, isolate GYV caused systemic symptoms on true leaves, including chlorotic spots, mottle, mosaic and intraveinal chlorosis, but did not induce necrotic lesions (Fig. 6.1 E and G). The plants were not stunted (Fig. 6.1 E). The typical recovery was not observed even after growing for an extended period (up to eight weeks post-inoculation).

6.4.2 Detection of TomRSV by RT-PCR

To detect different virus isolates in infected cucumber plants, total nucleic acids were extracted from healthy (control) and infected leaf tissues. A reverse-transcription polymerase chain reaction (RT-PCR) was used to synthesize and amplify cDNA from the purified RNA. RT-PCR amplification of TomRSV RNA with primers W048 and W3520 was expected to produce a cDNA fragment with a predicted size of 4216 base pair (bp) (TomRSV RNA-1 nt 3520-7735, Rott et al., 1995). Indeed, a DNA fragment of the expected size was amplified from the extracts of Rasp-1-, PYB-1- and T392-infected plants but not from those of GYV-infected and healthy plants (Fig. 6.2 A, lanes 2, 4, 6 and 8). A different pair of primers (oligonucleotides W3520 and W049) was used to attempt the amplification of a cDNA fragment with an expected size of 3246 bp (TomRSV RNA-1 nt 3520-6765). A prominent DNA band of the expected size was amplified from the leaf tissue extracts infected with all isolates (Fig. 6.2 A, lanes 3, 5, 7 and 9). To detect the RNA-2 molecule of the isolates, primers W049 and W054 were used to amplify a DNA product with an expected size of 1800 bp corresponding to TomRSV RNA-2 nt 3925-5824. After RT-PCR amplification, a major DNA band with the predicted size was detected in the infected leaf tissues with all isolates used in
Fig. 6.2. Agarose gel electrophoretic analysis of RT-PCR amplified TomRSV cDNA. Arrows at the right side of each gel indicate the detected DNA products and the calculated sizes. Mr represents DNA molecular weight markers (lanes 1 and 12), with standard sizes indicated at the left side. W048 and W049 represent primers W3520 and W048, and W3520 and W049, respectively. Isolates are indicated above the gel. A) Amplification of RNA-1 cDNA. B) Amplification of RNA-2 cDNA using primers W054 and W049.
this study (Fig. 6.2 B, lanes 13-17). No amplified products were detected using nucleic acids extracted from healthy leaf tissues (Fig. 6.2 A, lane 10 and 11; Fig. 6.2 B, lane 15), suggesting amplification with RNA-1 and RNA-2 derived primers was specific.

Table 6.2. Percent identities among the nucleotide and amino acid sequences of the VPg-protease-polymerase precursor, of the VPg, of the protease and of the putative RNA-dependent polymerase of tomato ringspot virus isolates

<table>
<thead>
<tr>
<th></th>
<th>Rasp-1</th>
<th>GYV</th>
<th>PYB-1</th>
<th>T392</th>
<th>Rasp-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rasp-1</td>
<td>.....</td>
<td>81, 81, 79, 83</td>
<td>96, 97, 93, 97</td>
<td>96, 93, 96, 97</td>
<td>99, 98, 100, 99</td>
</tr>
<tr>
<td>GYV</td>
<td>89, 100, 89</td>
<td>.....</td>
<td>81, 79, 81, 82</td>
<td>82, 80, 82, 82</td>
<td>81, 83, 79, 83</td>
</tr>
<tr>
<td>PYB-1</td>
<td>97, 100, 97</td>
<td>89, 100, 89</td>
<td>.....</td>
<td>97, 93, 96, 98</td>
<td>96, 98, 93, 97</td>
</tr>
<tr>
<td>T392</td>
<td>98, 100, 98</td>
<td>89, 100, 92, 88</td>
<td>98, 100, 97, 98</td>
<td>.....</td>
<td>97, 95, 96, 97</td>
</tr>
<tr>
<td>Rasp-2</td>
<td>99, 100, 100, 99</td>
<td>89, 100, 89, 89</td>
<td>98, 100, 97, 98</td>
<td>98, 100, 98, 98</td>
<td>.....</td>
</tr>
</tbody>
</table>

* The percentage values are shown for the nucleotide sequences (above the diagonal) and for the amino acid sequences (below the diagonal) of the VPg-protease-polymerase precursor (bold), VPg (italic), protease (underlined) and polymerase (normal). Sequences for Rasp-2 were reported previously (Rott et al., 1995).

6.4.3 Comparison of the RNA-1-encoded VPg, Pro and Pol coding regions of different TomRSV isolates

The DNA products obtained from the RT-PCR amplification of the nucleic acid extracts with primers W3520 and W049 were gel-eluted and directly used for sequencing. The sequences of the coding regions for the entire VPg (corresponding to TomRSV RNA-1 nt 3714-3794), Pro (corresponding to TomRSV RNA-1 nt 3795-4535) and Pol (corresponding to TomRSV RNA-1 nt 4536-6668) were determined (data not shown). These nucleotide sequences and the deduced amino acid sequences were compared to the published sequences of the corresponding regions of isolate Rasp-2 (Rott et al., 1995) (Fig. 6.3, Table 6.2). Percent identities among the different isolates are shown in Table 6.2. Remarkably, the overall level of identities at both the nucleotide and amino acid levels between Rasp-1, a raspberry isolate from the state of Washington, U.S.A. and Rasp-2, a raspberry isolate from British Columbia, Canada was 99%. Two other isolates, PYB-1, a peach...
Fig. 6.3. For legend see next page
Fig. 6.3. **Multiple amino acid alignments of the VPg (italic), the protease (underlined) and the putative RNA-dependent RNA polymerase (normal) of tomato ringspot nepovirus isolates.** Identical amino acids are shown as dots. Amino acids different from those of the published TomRSV (isolate Rasp-2) sequence are shown (Rott et al., 1995). Arrows indicate the amino acids suggested in the catalytic pocket and substrate binding pocket of the protease. Diamonds indicate the amino acids suggested to be involved in the substrate binding pocket of the protease (Hans and Sanfacon, 1995). Boxes indicate the conserved region of Pol and asterisks indicate highly conserved amino acids of Pol (Rott et al., 1995)
isolate from California, U.S.A. and T392 demonstrated a very high level of identity to the two raspberry isolates (between 96 and 98% identity). In contrast, analysis of the nucleotide and amino acid sequences of a grape isolate, GYV, from California, U.S.A. revealed a lower level of identity to the other isolates. Identity to the other isolates was approximately 81% at the nucleotide sequence level and 89% at the amino acid sequence level. Comparison of the nucleotide and amino acid sequences of individual genes, i.e. Pro and Pol gave similar patterns as described above. Interestingly, the entire amino acid sequences of VPg revealed 100% identity among all the isolates tested despite some variations at the nucleotide level (Table 6.2, Fig. 6.3). To further examine the differences in amino acid residues among five isolates, multiple amino acid alignments were made, and specific regions of the polyprotein were examined for variability (Fig. 6.3). The dipeptide sequences of the NTB-VPg, VPg-Pro, Pro-Pol cleavage sites were previously determined to be Q/S, Q/G and Q/S, respectively, for isolate Rasp-2 (Chapters 3 and 4). Identical dipeptide sequences were observed in all isolates with the exception of the grape isolate, GYV, which contained Q/S instead of Q/G at the VPg-Pro cleavage site (Fig. 6.3). In the protease, amino acid residues previously suggested to form conserved motifs in the protease such as H$_{81}$, E$_{119}$ (or D$_{142}$) and C$_{221}$ in the catalytic triad, H$_{239}$ in the substrate binding pocket and amino acids potentially important in the formation of the substrate binding pocket (H$_{193}$, I$_{210}$, L$_{224}$ and L$_{229}$) (Hans and Sanfaçon, 1995; Rott et al., 1995) were identical in all isolates (Fig 6.3). In the polymerase, three amino acid motifs were previously found to be common among nepo-, como-, poty- and picornaviruses (Rott et al., 1995). All these motifs (including the G$_{661}$D$_{662}$D$_{663}$ motif) were conserved in the isolates sequenced in this study (Fig. 6.3).

6.4.4 Comparison of the RNA-2 CP gene of different TomRSV isolates

The RNA-2 sequence corresponding to the entire CP coding region was determined for the four TomRSV isolates by direct sequencing of the major PCR products obtained by amplification of the RNA extracts with primers W049 and W054. The nucleotide sequences and deduced amino
acid sequences were compared to the corresponding regions of RNA-2 of the Rasp-2 and PYB-2 isolates published previously (Rott et al., 1991; Yepes et al., 1996). Percent identities among the different isolates are shown in Table 6.3. Two raspberry isolates showed high similarity at both the nucleotide and amino acid levels (98% and 97%, respectively). High similarity was also observed between two peach isolates (PYB-1 and PYB-2) and isolate T392 (nucleotide sequence: 96%, amino acid sequence: 96-97%). However, the nucleotide and amino acid sequences of a grape isolate, GYV, showed significantly lower similarity to those of other isolates (nucleotide: 78%, amino acid 83-84%). Examination of the length of the CP amino acid sequence revealed that all the isolates encoded a 562 amino-acid coat protein. The dipeptide sequence of the MP-CP cleavage site of all the isolates was Q/G (the amino acid sequence of the C-terminal portion of the movement protein not shown, Fig. 6.4).

### Table 6.3. Percent identities among the nucleotide and amino acid sequences of the coat protein of tomato ringspot virus isolates

<table>
<thead>
<tr>
<th></th>
<th>Rasp-1</th>
<th>GYV</th>
<th>PYB-1</th>
<th>T392</th>
<th>Rasp-2</th>
<th>PYB-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rasp-1</td>
<td>...</td>
<td>78</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>97</td>
</tr>
<tr>
<td>GYV</td>
<td>83</td>
<td></td>
<td>78</td>
<td>78</td>
<td>78</td>
<td>78</td>
</tr>
<tr>
<td>PYB-1</td>
<td>97</td>
<td>83</td>
<td></td>
<td>97</td>
<td>97</td>
<td>94</td>
</tr>
<tr>
<td>T392</td>
<td>99</td>
<td>83</td>
<td>96</td>
<td></td>
<td>97</td>
<td>97</td>
</tr>
<tr>
<td>Rasp-2</td>
<td>98</td>
<td>84</td>
<td>97</td>
<td>98</td>
<td></td>
<td>97</td>
</tr>
<tr>
<td>PYB-2</td>
<td>96</td>
<td>83</td>
<td>96</td>
<td>96</td>
<td>97</td>
<td></td>
</tr>
</tbody>
</table>

*The percentage values are shown for the CP nucleotide sequence (above the diagonal) and for the CP amino acid sequence (below the diagonal). Sequences for Rasp-2 and PYB-2 were reported previously (Rott et al., 1991; Yepes et al., 1996).

Recently, coat protein amino acid alignments have revealed that there are five conserved motifs among nepoviruses (Latvala et al., 1998). Four of these motifs were conserved in the TomRSV isolates (Fig. 6.4). However, an amino acid substitution was observed in the proposed, FDA(F/Y)X(R/K). Substitution of F at the first position by L was found in all isolates sequenced in
Fig. 6.4. Multiple alignments of the coat protein amino acid sequences of tomato ringspot nepovirus isolates. Dots indicate amino acids identical to the TomRSV (isolate Rasp-2) sequence published previously (Rott et al., 1991). Boxes of solid lines indicate the amino acids conserved in all nepoviruses and boxes of discontinuous lines indicate the amino acids conserved in the subgroup c nepoviruses suggested by Latva et al. (1998).
this chapter and published previously.

6.5 Discussion

In this study, we demonstrate that RT-PCR can be used to detect TomRSV in extracts from infected plants. Using synthetic oligonucleotides derived from published data of a raspberry isolate, i.e. Rasp-2, the RNA-1 and RNA-2 genomic sequences of all tested isolates were specifically amplified. Recently RT-PCR has emerged as a new powerful, quick and sensitive tool for RNA virus detection. It has been efficiently used for detection of many plant viruses and viroids such as potato spindle tuber viroid (Shamloul et al., 1997), tobacco rattle virus (Robinson, 1992) and citrus tristeza virus (Lopez et al., 1998). For TomRSV diagnosis, immuno-related tests have been suggested to be the most reliable approach. However, a number of experiments have shown that immunological methods including ELISA do not always allow an reliable detection due to different serotypes (Stace-Smith, 1984a and 1984b). Therefore, RT-PCR may provide an alternative to the traditional diagnostic techniques for TomRSV detection. We show that the sensitivity of RT-PCR to detect a range of isolates depends on sequence identities of the primers used to the target RNA. When we used a set of primers containing primer W048, the predicted amplified products were not detected for GYV, probably due to the low level of identity between GYV and Rasp-2, from which the primers were derived. In this study, we have identified several highly conserved nucleotide sequences among isolates which could be used to design universal primers for detection. For example, the sequence identities of the coding regions for amino acids 36 to 82 (in the N-terminal portion of Pro), 225-236 (in the C-terminal portion of Pro), 321-345 (in the N-terminal part of Pol) and 630-652 (in the middle part of Pol) were over 90%. In particular, the nucleotide sequences corresponding to amino acids 65-72 and 201-236 were identical.

Among the four TomRSV isolates studied, GYV, a grape isolate caused very different symptoms on cucumber plants. GYV caused less severe symptoms on cucumber. At the late stages
of infection, no recovery plants were observed. In contrast, cucumber plants exhibited more severe symptoms when inoculated with the other three isolates tested, and some plants recovered from the disease. Previously Cadman and Lister (1961) reported that TomRSV isolates purified from peach and tobacco induced different symptoms on cucumber. But they did not specify the symptoms and the difference. It is not known if the peach isolate studied by Cadman and Lister is closely related to the PYB-1 isolate studied here. These results confirm the previous conclusion that the symptoms caused by TomRSV are diverse and symptomatology is not recommended as the method for detection (Stace-Smith, 1984b).

To initiate studies on the molecular basis of symptomatology of these isolates, the 3' termini of the coding sequences of the genomic RNA-1 and RNA-2 were sequenced. Our analysis revealed that two raspberry isolates are the most similar based on nucleotide and amino acid sequences (Table 6.2 and 6.3). In spite of different host and geographical origins, these raspberry isolates share over 96% identity in nucleotide and amino acid sequences with a peach isolate, PYB-1 and an isolate of unknown origin, T392 (Table 6.2 and 6.3). This suggests that these four isolates are more related. Both amino acid and nucleotide sequences of these four isolates, however, displayed more dissimilarity to those of GYV, a grape isolate. This correlated with the observation that four isolates: Rasp-1, Rasp-2, PYB-1 and T392 induced very similar symptoms on cucumber, whereas, the grape isolate caused different symptoms. To further investigate the possible correlation between divergent phenotypes and molecular variability, more TomRSV isolates need to be studied and more viral genes need to be sequenced.

Sequence comparisons among these four isolates and another isolate sequenced previously (Rott et al., 1995) revealed that the amino acid sequences of the RNA-1-encoded VPg-Pro-Pol precursor and each individual mature protein, i.e. VPg, Pro and Pol, were more conserved than the sequence of the RNA-2-encoded coat protein. Moreover, the deduced amino acid sequences of a small viral genome-linked protein, VPg were identical among all the five isolates. The possible
functions of VPg in TomRSV-related picornaviruses include initiation of virus RNA replication as a primer (Paul et al., 1998) and protection of viral genomic RNA from degradation (Mayo et al., 1982). The presence of VPg has shown to be important for TomRSV as removal of the protein from viral genomic RNA by proteinase K treatment destroys RNA infectivity (Mayo et al., 1979; 1982). The highly conserved amino acid sequences of VPg among TomRSV isolates suggests that it may be an essential requirement to ensure these possible functions in the TomRSV replication cycle.

The critical function of the protease as the enzyme responsible for maturation of polyprotein P1 and P2 has been discussed (Hans and Sanfaçon, 1995; Chapters 3 and 4). The conserved amino acids present either in the predicted catalytic triad or substrate binding pocket (Hans and Sanfaçon, 1995) were examined in this study. No substitution was found among the five isolates compared, further confirming the importance of these conserved amino acids. The putative RNA-dependent RNA polymerase has been recognized as the most important enzyme for viral RNA replication. All the conserved amino acids suggested previously (Rott et al., 1995) remained unchanged though the overall amino acid sequences were divergent, especially for GYV (Fig. 6.3), further suggesting the critical requirement of these amino acids for virus viability.

In this study, the amino acid sequences of the coat protein encoded by TomRSV RNA-2 were more divergent than those of the RNA-1-encoded proteins. Viral coat proteins are involved in RNA packaging and serotype determination and often in differentiation of symptoms and vector transmission. The variability in the coat protein sequences in this study may therefore play a role in the different symptoms induced by different isolates. Further analysis of the function of the coat protein or other viral proteins in the induction of symptoms awaits the construction of infectious transcripts which will allow the exchange of these domains among the different isolates and the determination of the potential role of single amino acid substitutions.
CHAPTER 7

SUMMARY, GENERAL DISCUSSION
AND FUTURE PROSPECTS
7.1 Summary and General Discussion

TomRSV consists of a genome of two RNA species, RNA-1 and RNA-2. TomRSV RNA-1 is likely to encode all the viral proteins required for RNA replication. This suggestion is supported by the following indirect evidence. First, sequence comparisons have shown that the putative gene products of TomRSV RNA-1 share significant sequence homology with the nonstructural, replication-required proteins encoded by the related viruses (Rott et al., 1995). Second, the arrangement of these conserved protein domains in the TomRSV RNA-1-encoded polyprotein closely resembles that of the RNA-1-encoded polyproteins of the related como- and other nepoviruses, and of the corresponding proteins encoded by poty- and picornaviruses (Rott et al., 1995; Sanfaçon, 1995; Chapter 2). Third, in grapevine fanleaf virus (GFLV) and tomato black ring virus (TBRV), two TomRSV closely related nepoviruses, RNA-1 but not RNA-2 can replicate independently in protoplasts (Robinson et al., 1980; Viry et al., 1993).

The expression and processing strategies of the TomRSV RNA-1-encoded polyprotein (P1) were studied in vitro and in vivo (Chapters 3 and 4). Based on these studies, the genomic organization of TomRSV RNA-1 has been established. Translation of RNA-1 is probably initiated at the first AUG codon of the long open reading frame to produce a 244 kDa polyprotein, which is intramolecularly processed into six end products, i.e. from the N-terminus, X1 (35 kDa), X2 (32 kDa), NTB (66 kDa), VPg (3 kDa), Pro (27 kDa) and Pol (81) kDa, and several precursor polypeptides such as NTB-VPg. These intermediate precursors can not be cleaved in trans by the protease. The overall genomic organization of TomRSV RNA-1 is very similar to that of other nepoviruses and the closely related comoviruses (Rott et al., 1995; Sanfaçon, 1995; Goldbach and Wellink, 1996; Mayo and Robinson, 1996). However, the TomRSV RNA-1-encoded P1 contains two proteins, i.e. X1 and X2, upstream of the NTB protein, whereas, only one protein is present at the corresponding position in all other characterized nepo- and comoviruses. Although the X2 protein has certain sequence homology with the 32 kDa protein of CPMV, which presents a protease
cofactor role (Vos et al., 1988; Peters et al., 1992a; Rott et al., 1995), X2 is not likely to play such a role in the polyprotein processing as shown in this thesis and in a previous study (Hans and Sanfaçon, 1995). The potential functions of these two proteins in RNA replication or in other macromolecular processes need to be elucidated.

Because processing of P1 is an intramolecular event, it must follow a sequential order (Fig. 4.8), which is probably regulated by several factors to ensure the production of the replication-required mature and intermediate precursor proteins in the processing cascade. First, the efficiency of processing varies among the different cleavage sites, which may result in accumulation of some intermediate precursors. For example, in the NTB-VPg-Pro-Pol precursor, the protease prefers to cleave the NTB-VPg and Pro-Pol cleavage sites in vitro and in E. coli. As a result, the VPg-Pro intermediate precursor accumulates (Fig. 3.8). Second, the presence of some viral domains on the precursor may affect the processing efficiency at a distant cleavage site. In Chapter 3, the presence of the VPg sequence was shown to result in enhanced cleavage at the Pro-Pol cleavage site. Third, host factors may regulate the protease activity. The VPg-Pro precursor accumulates in vitro and in E. coli but not in vivo (Fig. 3.4, Fig. 3.5, Fig. 3.8 and Fig. 4.2). The NTB-VPg precursor was detected in vivo but not in vitro (Fig. 3.4 and Fig. 4.2). It is possible that some host factors may bind to either the protease itself or sequences around the cleavage sites resulting in a change of conformation of the protease or the substrates, which may favor processing at the VPg-Pro cleavage site in plants.

Upon translation and processing of P1, the replication-related P1 proteins may form a replication complex by protein-protein interactions. The NTB protein (or NTB-containing precursors such as the NTB-VPg protein) is probably the anchor protein to secure the replication complex to host membranous structures. The NTB-VPg precursor is glycosylated and signal peptidase-processed when translated in the presence of microsomal membranes in vitro (Fig. 5.1, Fig. 5.2 and Fig. 5.5), suggesting that it may be an integral membrane protein. This was confirmed by solubilization experiments (Fig. 5.4). The membrane-derived modifications requires the presence of a predicted
transmembrane domain at the C-terminal moiety of the NTB protein (Fig. 5.2). The NTB-VPg precursor was found to be abundant in infected plants (Fig. 4.2) and to be predominantly distributed in the membrane fractions (Chapter 4, data not shown). The NTB-VPg precursor was also found to be abundant in CPMV-infected plants. The poliovirus 3A and potyvirus 6 kDa proteins, showing membrane binding abilities, are believed to be the anchor protein responsible for the attachment of the replication complex to membranous structures and are also found in infected cells as precursors containing the VPg domain [3AB and 6K-NIa(VPg)] (Lama et al., 1994; Restrepo-Hartwig and Carrington, 1994; Towner et al., 1996; Schaad et al., 1997). Therefore, the TomRSV NTB-VPg precursor may play an important role in RNA replication. In vitro glycosylation of the NTB-VPg precursor occurs at a glycosylation site in the VPg sequence (Fig. 5.5). However, it is not clear whether the NTB-VPg precursor undergoes this modification in vivo. Analysis of the sequence of the NTB-VPg precursors in como- and nepoviruses revealed the presence of a predicted transmembrane domain in the C-terminal region of NTB which is adjacent to a conserved glycosylation sequence. Potential glycosylation of the NTB-VPg precursors in CPMV, GFLV and other nepo- and comoviruses has not been studied. Like the NTB-VPg precursor in this thesis, 3AB of poliovirus can be glycosylated in vitro as well (Datta and Dasgupta, 1994). The glycosylation seems to be required for virus viability (Data et al., 1996), though it is difficult to predict the roles it may play in viral RNA replication. The finding that the TomRSV NTB-VPg is signal peptidase-processed by microsomal membrane preparations in vitro may be of biological importance. Although the cleavage site has not been determined, the size of the proteins released by signal peptidase cleavage and the observation that the small protein released is glycosylated at the conserved glycosylation site in the VPg sequence suggests that a region corresponding approximately to the entire VPg sequence is cleaved (Fig. 5.5). As mentioned earlier, the NTB-VPg proteins are present in infected plants, which can not be processed by the exogenous protease in vitro (Fig. 4.5; Fig. 4.6 B and C; Peters et al., 1992b). In contrast, the equivalent protein of poliovirus, 3AB, which delivers
the VPg to the 5' ends of plus- and minus-strand RNAs during RNA replication can be processed \textit{in trans} (Paul \textit{et al.}, 1998; Xiang \textit{et al.}, 1998). If the NTB-VPg protein in TomRSV and CPMV is a donor of VPg, the mechanism by which the NTB-VPg precursor releases the VPg domain should be addressed. It is not known whether the signal peptidase cleavage found in this thesis provides one possible option.

Direct amino-acid sequencing of the TomRSV VPg protein suggests that the amino acid T of the conserved glycosylation site is substituted by an A (Chapter 4). This substitution can block the glycosylation of the NTB-VPg precursor \textit{in vitro} (Fig. 5.6). Sequencing of the VPg coding region of several TomRSV isolates has shown that the deduced amino acid sequences of VPg of all isolates are identical (containing the conserved sequence for glycosylation) despite the fact that the nucleotide sequences are divergent (Chapter 6, Fig. 6.3 and Table 6.2). The reason for this apparent discrepancy must be addressed. Both the amino acid sequencing of the VPg protein and the nucleotide sequencing of cDNA fragments were done on a total population of viral RNA. It is possible that two populations of the viruses are present in infected cells, one of which contains a glycosylation site. If this is the case, then glycosylation of only a subset of the VPg protein in infected plants may play an important biological role. Alternatively, an unknown mechanism such as editing may take place to result in the translation of the VPg protein with a amino acid substitution at the glycosylation site. Sequencing of a large number of TomRSV cDNA clones may help resolve this question.

7.2 Future Prospects

A number of experiments should be conducted to further deepen and widen our knowledge of TomRSV RNA replication. First, isolation of the active replication complex from infected plants should be attempted. The purified active replication complex can be used for various purposes such as identification of host factors, determination of viral proteins present in the complex and
elucidation of the mechanism for initiation of replication. Using published protocols (Quadt and Jaspars, 1990; Hays and Buck, 1990), isolation of the replication complex for TomRSV was attempted in this thesis (data not shown). The purified proteins in the membrane fraction from TomRSV-infected cucumber plants failed to replicate the exogenous TomRSV RNA in the presence or absence of membranes. It is not known if some host factors or viral proteins which might be required for the replication were lost during the purification procedures or if the conditions for the in vitro replication assay may not have been optimal. Therefore, the published protocols probably need to be modified such as salt and detergent concentrations. Using different hosts may also increase the chances of purifying an active replication complex. Second, protein-protein interactions among the RNA-1-encoded proteins and between viral proteins and host factors should be investigated. A replication complex containing viral proteins and host factors must be formed by protein-protein interactions. Methods such as the yeast two-hybrid system, far-western immunoblots and co-immunoprecipitations can be used to unravel the possible protein-protein interactions. Finally, construction of a TomRSV infectious cDNA clone should be attempted. Infectious RNA transcripts from the viral RNA cDNA clone will allow us to dissect the viral genome and to elucidate the function(s) of each viral gene in virus replication.


Cheng, X. and Bruening, G. 1994. Cloned DNA copies of cowpea severe mosaic virus genomic RNAs:


Crawford, N.M. and Baltimore, D. 1983. Genome-linked protein VPg of poliovirus is present as free VPg and VPgp-pUpU in poliovirus-infected cells. *Proceedings of the National Academy of Sciences, USA* 80, 7452-7455.


Datta, U., Banerjee, R., Liu, X. and Dasgupta, A. 1996. A point mutation in the glycosylation site of


Lain, S., Riechmann, J.L., Martin, M.T. and Garcia, J.A. 1989b. Homologous potyvirus and flavivirus
proteins belonging to a superfamily of helicase-like proteins. *Gene* 82, 357-362.


Neil, J.D. 1990. Nucleotide sequence of a region of the feline calicivirus genome which encodes


Sandoval, I.V. and Carrasco, L. 1997. Poliovirus infection and expression of the polio virus protein 2B
provoke the disassembly of the Golgi complex, the organelle target for the antipoliovirus drug Ro-090179. *Journal of Virology* 71, 4679-4693.


subunits of ATP synthase, myosin, kinase and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO Journal* 1, 945-951.


Wengler, G. and Wengler, G. 1991. The carboxyl-terminal part of the NS3 protein of the West Nile flavivirus can be isolated as a soluble protein after proteolytic cleavage and represents an RNA-stimulated NTPase. *Virology* 184, 707-715.


Yamshchikov, V.F. and Compans, R.W. 1993. Regulation of the late events in flavivirus protein processing


Ypma-Wong, M.F., Dewalt, P.G., Johnson, V.H., Lamb, J. and Semler, B.L. 1988. Protein 3CD is the active protease for processing of the major poliovirus proteinase responsible for cleavage of the P1 capsid precursor. *Virology* 166, 265-270.