TERTIARY STRUCTURAL AND FUNCTIONAL ANALYSES OF RNA MOTIFS THAT MEDIATE VIROID REPLICATION AND SYSTEMIC TRAFFICKING

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

by

Xuehua Zhong, M.S.

The Ohio State University

2007

Dissertation Committee:

Professor Biao Ding, Adviser
Professor Jyan-Chyun Jang
Professor Iris Meier
Professor Randy Scholl

Approved by

______________________________
Professor Biao Ding
Adviser

______________________________
Graduate Program of Plant Cellular and Molecular Biology

The Ohio State University
RNA-templated RNA replication is essential for viral or viroid infection, as well as for gene regulation. Besides its regulation at the cellular levels, systemic RNA trafficking is emerging as a new paradigm in gene regulation at the whole plant level. Knowledge of how an RNA replicates in a cell and further traffics across specific cellular boundaries and eventually into and out of the vascular tissue to reach specific target cells/tissues is crucial for understanding mechanisms that regulate developmental and physiological processes at both cellular and organism levels. Viroids of the *Pospiviroidae* family, as represented by the *Potato spindle tuber viroid* (PSTVd), replicate in the nucleus and traffic intercellularly across different cellular boundaries to establish systemic infection. PSTVd is a powerful system to elucidate structural motifs that regulate RNA replication and systemic trafficking because of its small RNA genome, diverse structural motifs, high accumulation levels in its hosts and complete dependence on cellular factors for systemic infection. I performed structural and functional analyses and identified essential role of the Loop E motif in replication and the U43/C318 motif for vascular entry. I have also identified multiple and distinct motifs that are important for replication and systemic trafficking by genome-wide mutational analyses.

The loop E motif, located in the central conserved region of PSTVd, has been implicated in replication, host adaptation and symptom expression based on
gain-of-function mutational analyses. However, loss-of-function genetic analysis is necessary to provide the direct evidence for the role of the loop E. Here, I investigated the tertiary structure and function of the loop E in replication by combining isostericity matrix analysis with genetic and functional assays. A tertiary-structural model of the loop E motif, inferred by comparative sequence analysis and comparison with nuclear magnetic resonance and X-ray crystal structures of the loop E motifs in other RNAs, is presented with core non-Watson-Crick base pairs precisely specified. Isostericity matrix analysis of these base pairs showed that the model accounts for the reported natural sequence variations and viable experimental mutations in the loop E motifs of PSTVd and other viroids. Furthermore, isostericity matrix analysis allowed me to design disruptive, as well as compensatory, mutations of PSTVd loop E. Functional analyses of such mutants by in vitro and in vivo experiments demonstrated that the loop E structural integrity is crucial for replication, specifically during transcription. My results suggest that the PSTVd the loop E motif exists and functions in vivo and provide loss-of-function genetic evidence for the essential role of a viroid RNA three-dimensional motif in rolling-circle replication. The use of isostericity matrix analysis of non-Watson-Crick base pairing to rationalize mutagenesis of tertiary motifs and systematic in vitro and in vivo functional assays of mutants offers a novel, comprehensive approach to elucidate the tertiary structure-function relationships for RNA motifs of general biological significance.

By combining single cell replication assay in cultured cells and systemic infection assay in plants of *Nicotiana benthamiana*, I identified an RNA motif for PSTVd systemic trafficking. In situ hybridization demonstrated that this motif is required for PSTVd to traffic from nonvascular into the vascular tissue phloem to initiate systemic
infection. This motif consists of nucleotides U/C that form a water-inserted cis Watson-Crick/Watson-Crick base pair flanked by short helices that comprise canonical Watson-Crick/Watson-Crick base pairs. This tertiary structural model was inferred by comparison with X-ray crystal structures of similar motifs in rRNAs and is supported by combined mutagenesis and covariation analyses. Hydration pattern analysis suggests that water insertion induces a widened minor groove conducive to protein and/or RNA interactions. My model and approaches have broad implications to investigate the RNA structural motifs in other RNAs for vascular entry and to study the basic principles of RNA structure-function relationships.

Having shown the critical role of the loop E in replication and the U43/C318 motif in systemic trafficking, I next asked whether replication and trafficking each involve single or multiple motifs. I conducted a genome-wide mutational analysis of the role of each of all PSTVd individual loops/bulges in replication in single cells and systemic trafficking in a plant. I identified multiple and distinct motifs essential or important for replication and systemic trafficking. Here, I present a genomic map of the PSTVd replication/trafficking motifs which shows that, in the PSTVd secondary structure, the motifs most critical for replication are clustered in the distal end of the left terminal domain and central region, whereas the trafficking motifs are mostly clustered in the variable and right terminal domains and around the pathogenicity domain. This map may guide further studies on the tertiary structure and functional mechanisms of each motif and identification of the cognate cellular factors.

All together, my findings from PSTVd structure and function studies demonstrate that a combination of tertiary structural analyses and genetic and biological experiments
can provide a foundation to elucidate the elaborate RNA motif-cellular factor interactions that regulate diverse biological processes, and to address the fundamental question of how sophisticated features of RNA structures have evolved to achieve optimal functions.
Dedicated to all members of my family, especially my daughter Grace
ACKNOWLEDGMENTS

No single word can express my sincere gratitude for my adviser, Dr. Biao Ding, for his intellectual guidance, endless encouragement and consistent support through this unforgettable journey. He has been the most amazing adviser, shared the knowledge he possesses and helped me to develop a greater understanding of subjects I am passionate about. The knowledge I gained from him is something that will stay with me for a lifetime and benefit me greatly in my future career.

I am very indebted to Dr. Neocles Leontis and his student Jesse Stombaugh for their extensive help and contributions with tertiary structural analyses on PSTVd RNA. They taught me about non-Watson-Crick base pairs and the concept of isostericity and made me familiar with the chemistry and RNA structure. Figures 1.2, 1.3, 1.4 and 1.5 are provided by Neocles. Figures 2.8, 2.10 and 2.13 are prepared by Neocles and Jesse.

I am grateful to my advisory committee members, Dr. Iris Meier, Dr. Venkat Gopalan, Dr. Randy Scholl and Dr. Jyan-Chyun Jang for their invaluable suggestions and discussions on my projects and my thesis. Especially, I would like to thank Dr. Venkat Gopalan for sharing many chemicals and allowing me to use his FPLC.

My special appreciation goes to Dr. Asuka Itaya for her friendship and numerous discussions on my research projects and advises on my experiment designs. I thank Dr. Xiaorong Tao for in situ hybridization experiments in chapter 2 and for figures 2.5, 2.6 and 2.7. I also thank Dr. Yijun Qi for his great help during my first year in the lab. I would also
like to thank Ying Wang and Ryuta Takeda for technical help. I appreciate all present and past undergraduates in Ding’s lab for all maintenance work. Especially, I thank Anthony Archual and Amy Amin for their help in RNA, DNA extraction and other experiments.

I am also grateful to our collaborator Dr. Kathleen Boris-Lawire and Shuiming Qian for their help with realtime RT-PCR experiments and figure 1.10.

I thank Dr. David Bisaro, Dr. David Somers, Dr. Desh Pal Verma, and Dr. Erich Grotewold for their advice on my research. Especially, I would like to thank Dr. David Bisaro for providing me with the *Nicotiana benthamiana* cultured cells. I also thank Dr. Patrice Hamel for sharing his knowledge about the science and his invaluable suggestions for my future career. I appreciate all the lab members in Plant Biotech Center for sharing equipments and chemicals.

I thank the staff of the Plant Biotechnology Center - Diane Furtney, Melinda Parker, Dave Long, Scott Hines and Joe Takayama and the staff of PCMB -Jill Hartman, Denise Blackburn-Smith, Debra Gamble, Laurel Shannon and Rene Madsen.

Finally and most importantly, I would like to thank my parents, my mother-in-law, my sisters, my husband Shuiming and my daughter Grace for their love, support and encouragements. They have been my spiritual support throughout my whole career. Especially, I want to thank my mother-in-law for her tremendous help in taking care of Grace. Shuiming has been my best friend and always there whenever I need him. He also provides many invaluable suggestions and ideas on my research projects. Grace has always been my source of energy and motivation. There are no words to say how much I love and cherish all of them.

路漫漫其修远兮，吾将上下而求索！
VITA

March 11, 1977 .................................................. Born – Wenling, Zhejiang, China

1999 ................................................................. B.S. Microbiology and Immunology
Wuhan University, Wuhan, China

1999 - 2002 .......................................................... M.S.
Plant Molecular and Developmental Biology
Wuhan University, Wuhan, China

2002 - Present ..................................................... Graduate Teaching and Research Associate,
Plant Cellular and Molecular Biology
The Ohio State University

PUBLICATIONS

function of an RNA motif required for plant vascular entry to initiate systemic trafficking.
EMBO J 26: 3836-3846.

cleavage to produce biologically active small RNAs but is resistant to RNA-induced


structural and functional analyses of a viroid RNA motif by isostericity matrix and


**FIELDS OF STUDY**

Major Field: Plant Cellular and Molecular Biology
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Tertiary structural and functional analyses of a viroid RNA motif by isostericity matrix and mutagenesis reveals its essential role in replication</td>
</tr>
<tr>
<td>1.1 Abstract</td>
</tr>
<tr>
<td>1.2 Introduction</td>
</tr>
<tr>
<td>1.3 Materials and Methods</td>
</tr>
<tr>
<td>1.3.1 Isostericity matrix analysis of PSTVd loop E</td>
</tr>
<tr>
<td>1.3.2 Plasmid construction</td>
</tr>
<tr>
<td>1.3.3 <em>In vitro</em> transcription</td>
</tr>
<tr>
<td>1.3.4 Protoplast inoculation</td>
</tr>
<tr>
<td>1.3.5 RNA extraction and Northern blot</td>
</tr>
<tr>
<td>1.3.6 Sequencing of RNA progeny</td>
</tr>
<tr>
<td>1.3.7 Preparation of <em>N. benthamiana</em> nuclear extract</td>
</tr>
</tbody>
</table>
1.3.8 RNA circularization assay ................................................................. 14
1.3.9 UV-crosslinking to analyze PSTVd loop E ................................. 14
1.3.10 Microinjection and microscopy .................................................... 15
1.3.11 Quantitative real-time RT-PCR ..................................................... 15
1.4 Results ............................................................................................... 17
1.4.1 Tertiary structural model for the PSTVd loop E motif .............. 17
1.4.2 The tertiary structural model of PSTVd loop E accounts for natural sequence variations and viable mutation .................................... 20
1.4.3 Application of isostericity matrix analysis to design disruptive mutations in PSTVd loop E motif .................................................... 24
1.4.4 Maintaining the tertiary structure of loop E is necessary for PSTVd accumulation in inoculated protoplasts ............................................ 27
1.4.5 A99C and A261C mutations do not impair nuclear import function ...... 28
1.4.6 A99C and A261C mutations do not affect circularization of the unit-length inoculum transcripts ......................................................... 29
1.4.7 A99C and A261C mutations inhibit the accumulation of (-)-strand PSTVd ................................................................. 30
1.5 Discussion ......................................................................................... 31
1.5.1 A tertiary structural model for viroid loop E motif that can guide functional studies ................................................................. 32
1.5.2 PSTVd loop E plays a crucial role in RNA-templated RNA transcription ................................................................. 33
1.5.3 PSTVd loop E is a useful model to study RNA structure-function relationships of broad significance ......................................................... 35
1.5.4 An integrative approach to address structure-function relationships of RNA motifs ................................................................. 36

2. Tertiary structure and function of an RNA motif required for plant vascular entry to initiate systemic trafficking ................................................................. 55

2.1 Abstract ............................................................................................. 56
2.2 Introduction....................................................................................................... 56
2.3 Materials and Methods.................................................................................... 59
  2.3.1 Plant materials and growth condition.......................................................... 59
  2.3.2 PSTVd cDNA construction......................................................................... 59
  2.3.3 Plant and protoplast transfection................................................................. 60
  2.3.4 In vitro transcription ................................................................................. 60
  2.3.5 Tissue processing, in situ hybridization...................................................... 60
  2.3.6 RNA extraction and RNA gel blot............................................................. 61
  2.3.7 RT-PCR.................................................................................................... 61
  2.3.8 Sequencing of RNA progeny...................................................................... 62
2.4 Results............................................................................................................. 62
  2.4.1 PSTVd infection of Nicotiana benthamiana as an experimental system 62
  2.4.2 Identification of PSTVd mutants defective in systemic trafficking............ 63
  2.4.3 Mutants C318A and U43G were capable of trafficking between
       non-vascular cells....................................................................................... 64
  2.4.4 Mutants C318A and U43G failed to traffic from bundle sheath into
       phloem........................................................................................................ 65
  2.4.5 Predicted tertiary structure of U43/C318 motif.......................................... 66
  2.4.6 Mutational analyses showed that local distortion of helix was critical for
       trafficking.................................................................................................... 67
  2.4.7 Covariation analyses support the tertiary structural model....................... 70
  2.4.8 Hydration patterns suggest a functional mechanism ............................. 72
2.5 Discussion....................................................................................................... 73

3.  Genome-wide mutational analyses identify multiple and distinct viroid RNA motif
    critical for replication and systemic trafficking ............................................. 95

  3.1 Abstract......................................................................................................... 96
  3.2 Introduction.................................................................................................. 96
  3.3 Materials and Methods................................................................................. 100
    3.3.1 Plant materials and growth condition..................................................... 100
3.3.2 PSTVd cDNA construction ................................................................. 100
3.3.3 In vitro transcription ................................................................. 101
3.3.4 Plant and protoplast infection ......................................................... 101
3.3.5 RNA extraction and RNA gel blot .................................................. 101
3.3.6 Sequencing of RNA progeny .......................................................... 102

3.4 Results ................................................................................................. 102
3.4.1 Generation of PSTVd mutants with disrupted loops ......................... 102
3.4.2 Multiple PSTVd loops were critical for replication in single cells ...... 104
3.4.3 Multiple loops were essential for PSTVd systemic trafficking .......... 105
3.4.4 Mutations in helical regions altered neighboring loop structures and inhibited PSTVd systemic trafficking .............................................. 107
3.4.5 Disruption of HPII and HPI did not inhibit PSTVd replication in single cells ......................................................................................... 108

3.5 Discussion ............................................................................................. 109

4. Future outlook ......................................................................................... 124

Bibliography ............................................................................................... 128
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Effects of nucleotide sequence variations on the predicted structure of loop E motifs and the viability of viroids</td>
</tr>
<tr>
<td>1.2</td>
<td>Nuclear import of PSTVd RNAs</td>
</tr>
<tr>
<td>3.1</td>
<td>Trafficking-defective mutants</td>
</tr>
<tr>
<td>3.2</td>
<td>Trafficking-impaired mutants</td>
</tr>
<tr>
<td>3.3</td>
<td>Trafficking-unknown mutants</td>
</tr>
<tr>
<td>3.4</td>
<td>Replication and trafficking of helical, HPI and HPII mutants</td>
</tr>
<tr>
<td>3.5</td>
<td>Sequencing of viroid progeny from a fraction of systemically infected plants</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>41</td>
</tr>
<tr>
<td>1.2</td>
<td>43</td>
</tr>
<tr>
<td>1.3</td>
<td>44</td>
</tr>
<tr>
<td>1.4</td>
<td>46</td>
</tr>
<tr>
<td>1.5</td>
<td>48</td>
</tr>
<tr>
<td>1.6</td>
<td>49</td>
</tr>
<tr>
<td>1.7</td>
<td>50</td>
</tr>
<tr>
<td>1.8</td>
<td>51</td>
</tr>
<tr>
<td>1.9</td>
<td>52</td>
</tr>
<tr>
<td>1.10</td>
<td>54</td>
</tr>
<tr>
<td>2.1</td>
<td>78</td>
</tr>
<tr>
<td>2.2</td>
<td>79</td>
</tr>
<tr>
<td>2.3</td>
<td>80</td>
</tr>
</tbody>
</table>
2.4 RNA blots showing defects of C318A and U43G mutations in systemic infection in *N. benthamiana* ................................................................. 81

2.5 Mutants C318A and U43G fail to traffic into the vascular tissue phloem at 12 days post inoculation ................................................................. 83

2.6 In situ hybridization reveals failure of mutants C318A and U43G to traffic into the vascular tissue phloem at 6 days post inoculation .................. 84

2.7 In situ hybridization reveals failure of mutants C318A and U43G to traffic into the vascular tissue phloem at 18 days post inoculation .............. 85

2.8 Local distortion of an RNA helix due to the presence of a water-inserted cWW U/C basepair ................................................................. 87

2.9 Mutational analyses of the trafficking motif ........................................... 89

2.10 Hydration patterns for the cWW and water-inserted cWW basepairs ....... 91

2.11 Covariation analyses of U/C motif in viroids of family *Posipiviroidae* .... 92

2.12 Secondary structures of PSTVd and other viroids in the family *Posipiviroidae* ......................................................................................... 93

2.13 Covariation analysis in 23S bacterial sequences for three conserved U/C basepairs ......................................................................................... 94

3.1 Functional analysis of PSTVd structural motifs ....................................... 120

3.2 Thermodynamic prediction of secondary structures of PSTVd loop mutants ............................................................................................... 121

3.3 Thermodynamic prediction of secondary structures of helical, HPII and HPI mutants .................................................................................. 122

3.4 A genomic map of PSTVd loop motifs that are essential/critical for replication and essential systemic trafficking ........................................ 123
ABBREVIATIONS

c-PSTVd, Circular PSTVd
CCR, Central conserved region
CCCVd, *Coconut cadang cadang viroid*
cWW, cis Watson-Crick/Watson-Crick
CEVd, Citrus exocortis viroid
DAPI, 4′,6′-Diamidino-2-phenylindole
EM, Engineered mutations
EDTA; Ethylene diamine tetraacetic acid
FR3D, Find RNA 3D
GFP, Green fluorescent protein
H-bonds, Hydrogen bonds
HDV, Hepatitis delta virus
H, hoogsteen edge
HPII, Hairpin II
HPI, Hairpin I
l-PSTVd, Linear PSTVd
M, Mock
MPVd, Mexican papita viroid
NMR, Nuclear Magnetic resonance
NbNE, *Nicotiana Benthamiana* nuclear extract
NV, Natural variation
PDB, Protein data bank
PSTVd, *Potato spindle tuber viroid*
PSTVd\textsubscript{Int}, Intermediate strain of PSTVd
RISC, RNA-induced silencing complex
RT-PCR, Reverse transcription-polymerase chain reaction
SE, Sugar edge
SDS, Sodium dodecyl sulfate
SM, Spontaneous mutations
SWS, Solvation web service
T\textsubscript{L}, Left terminal domain
T\textsubscript{R}, Right terminal domain
TCDVd, *Tomato chlorotic dwarf viroid*
TPMVd, *Tomato planta macho viroid*
TASVd, *Tomato apical stunt viroid*
WC, Watson-Crick edge
WT, wild type
WW, Watson-Crick/Watson-Crick
3D, Three-dimension
CHAPTER 1

TERTIARY STRUCTURAL AND FUNCTIONAL ANALYSES OF A VIROID RNA MOTIF BY ISOSTERICITY MATRIX AND MUTAGENESIS REVEAL ITS ESSENTIAL ROLE IN REPLICATION

1.1 ABSTRACT

RNA-templated RNA replication is essential for viral/viroid infection as well as for regulation of cellular gene expression. Specific RNA motifs likely regulate various aspects of the replication. Viroids of the Pospiviroidae family, as represented by the Potato spindle tuber viroid (PSTVd), replicate in the nucleus by utilizing the DNA-dependent RNA polymerase II. How an RNA template is replicated by the nuclear transcription machinery is poorly understood. We investigated the role of the loop E (sarcin/ricin) motif of the PSTVd genomic RNA in replication. A tertiary structural model of this motif, inferred by comparative sequence analysis and comparison with NMR and X-ray crystal structures of loop E motifs in other RNAs, is presented in which core non-Watson-Crick basepairs are precisely specified. Isostericity Matrix analysis of these basepairs showed that the model accounts for the reported natural sequence variations and viable experimental mutations in loop E motifs of PSTVd and other viroids. Furthermore, Isostericity Matrix analysis allowed us to design disruptive as well as compensatory mutations of the PSTVd loop E. Functional analyses of such mutants by in vitro and in vivo experiments demonstrated that the loop E structural integrity is crucial for replication, specifically during transcription. Our results provide loss-of-function genetic evidence for the essential role of a viroid RNA 3D motif in the rolling circle replication. The use of Isostericity Matrix analysis of non-Watson-Crick basepairing to rationalize mutagenesis of tertiary motifs and systematic in vitro and in vivo functional assays of mutants offer a novel, comprehensive approach to elucidate the tertiary structure-function relationships for RNA motifs of general biological significance.
1.2 INTRODUCTION

According to the “RNA World” scenario, the appearance of RNA molecules simultaneously capable of self-replication and information storage signaled a major milestone in the evolution of life (Gilbert, 1986; Joyce and Orgel, 1999; Joyce, 2002). In modern biology, RNA replication is central to viral/viroid infection as well as to the regulation of cellular gene expression. Virus-encoded RNA-dependent RNA polymerases play a major role in the replication of RNA viruses (Hull, 2002; Flint et al., 2004). Cellular RNA-dependent RNA polymerases generate double-stranded RNAs as triggers for RNA silencing (Baulcombe, 2004). Intriguingly, the DNA-dependent cellular RNA polymerases can also transcribe at least two types of RNA templates: viroid RNAs (Gora-Sochacka, 2004; Tabler and Tsagris, 2004; Flores et al., 2005) and the human hepatitis delta virus (HDV) RNA (Taylor, 2003; Lai, 2005). The replication of viroid and HDV RNAs raises the question of whether the DNA-templated transcription machinery also replicates other cellular RNAs yet to be identified. Elucidating the replication mechanisms of these infectious RNAs should help address this question of profound biological interest.

Viroids are the smallest known nucleic acid-based infectious agents and self-replicating genetic units. Their “genomes” consist of single-stranded, circular RNAs ranging in size from 250 to 400 nucleotides (Hull, 2002). Viroids can replicate and spread throughout an infected plant, although they do not encode proteins, do not have encapsidation mechanisms and do not require helper viruses. Furthermore, they cause devastating diseases by altering host gene expression and developmental processes (Tabler and Tsagris, 2004; Flores et al., 2005). Evidently, viroid RNA genomes contain all the
sequence and structural information needed to mediate or trigger the various functions associated with infection.

Potato spindle tuber viroid (PSTVd) is the type member of the Pospiviroidae family (Diener, 1971; Flores et al., 2000). The PSTVd genome consists of 359 nucleotides and assumes a rod-shaped secondary structure in the native state (Sanger et al., 1976) with five structural domains as shown in Figure 1.1A (Keese and Symons, 1985). This secondary structure, which is typical of viroids in the Pospiviroidae family, comprises many loops and bulges flanked by short Watson-Crick helices. Formation of this secondary structure is necessary for infection (Wassenegger et al., 1994). During asymmetric rolling-circle replication of PSTVd (Branch and Robertson, 1984), the (+)-circular strands serve as templates for the synthesis of concatemeric, linear (-)-strands, which then function as the replication intermediates for the synthesis of concatemeric, linear (+)-strands. These are subsequently cleaved into monomers and ligated into circular molecules (Fig. 1.1B). Without encoding proteins, PSTVd replicates in the nucleus of a host cell and therefore presents an ideal model to address the question of how the plant nuclear transcription machinery replicates an RNA template.

Previous work has attempted to associate particular parts of the viroid RNA with specific steps of the replication cycle. In vitro transcription using potato nuclear extracts and (+)-circular PSTVd RNA template suggested that nucleotides U359/C1 in the left terminal loop are the transcription initiation sites on the circular RNA (Kolonko et al., 2006). Subsequent site-directed mutagenesis in combination with infection studies in tomato showed that the C1G mutation is stably maintained whereas U359G reverted to the wild type, suggesting that perhaps U359 is the bona fide initiation site (Kolonko et al.,
A metastable hairpin II (HPII) encompassing nucleotide sequences 227-237 and 318-328 was predicted to form during thermal denaturation of the PSTVd secondary structure (Henco et al., 1979; Riesner et al., 1979) and has been detected in vitro and in vivo (Schroder and Riesner, 2002). Mutations that are designed to inhibit the formation of HPII were found to always revert to the wild type in mechanically inoculated plants, suggesting that HPII is important for replication, perhaps as a binding site for an unknown cellular factor (Loss et al., 1991; Owens et al., 1991; Qu et al., 1993; Candresse et al., 2001). Many mutations have been mapped throughout the PSTVd genome that inhibit infection in mechanically inoculated plants, and often were broadly interpreted to affect replication (Owens et al., 1986; Hammond and Owens, 1987; Owens et al., 1991; Hu et al., 1996).

The “loop E” motif, located in the central conserved region (CCR; Fig. 1.1A) of PSTVd, has been implicated in replication and other functions. This motif is similar to the loop E motifs of archaeal and eucaryal 5S rRNAs, which are also found in the conserved sarcin/ricin loops of 23S rRNA (Leontis and Westhof, 1998a; Leontis et al., 2002a). In vitro studies with longer than unit-length PSTVd transcripts showed that cleavage and ligation occur between G95 and G96, which appear to involve formation of a metastable tetraloop motif followed by a conformational change into a stable loop E (Baumstark et al., 1997). A single nucleotide substitution, C259U, in loop E enhances replication by 5- to 10-fold in cultured cells of tobacco (Nicotiana tabacum; Qi and Ding, 2002). The enhanced replication in tobacco cells may account for the adaptation of PSTVd to this host (Wassenegger et al., 1996; Zhu et al., 2002). The substitution U257A in the loop E motif also enables PSTVd to infect tobacco (Qi and Ding, 2002; Zhu et al., 2002) and confers
lethal symptoms in tomato (Qi and Ding, 2003). These in vitro and gain-of-function mutational results suggest that loop E plays one or more crucial roles in viroid-plant interactions.

Several critical issues must be addressed in order to assign a specific function of viroid replication in vivo to a particular viroid structural element. First, reversions of mutated nucleotides to the wild type or gain-of-function mutations do not provide direct evidence for the role of the affected RNA structural element in a biological process. Loss-of-function mutations, on the other hand, will provide definitive proof linking the affected structural element to a function. Second, infection assays on whole plants do not reveal whether failure of infection of a mutant results from defects in a specific step of the replication cycle or in intra-/inter-cellular trafficking. To address these issues, it is imperative to integrate structural, genetic, cellular, biochemical and molecular approaches.

Genetic alteration of an RNA structural element for functional studies requires precise understanding of its tertiary structure and the effects of mutations on that structure. However, while the secondary structures of viroids have been well established through thermodynamic, chemical mapping, mutational and NMR studies (Wassenegger et al., 1994; Dingley et al., 2003; Steger and Riesner, 2003), much less is known about their tertiary structures and specific functions. Minimum free energy calculations (e.g., mfold; Zuker, 2003) coupled with conventional comparative sequence analysis reveal RNA secondary structures comprising double-stranded helices formed by Watson-Crick basepairs, punctuated by so-called “loops” and “bulges” (hairpin, internal and junction or multi-helix loops) defined by lack of Watson-Crick basepairs. However, most RNA “loops” in structured, biologically active RNA molecules have been shown by X-ray
crystallography and NMR spectroscopy to form specific, well-structured 3D motifs in which many of the bases form non-Watson-Crick basepairs and/or stacking interactions (Leontis and Westhof, 1998a; Leontis and Westhof, 1998b; Leontis and Westhof, 1998c). Thus, current thermodynamic calculations generally cannot reveal the nature of non-Watson-Crick basepairing within “loops” and offer little help in designing mutations that disrupt or restore such basepairing.

As shown in Figure 1.2, RNA bases pair by hydrogen-bonding at their Watson-Crick, Hoogsteen or sugar edges, and with their glycosidic bonds oriented cis or trans relative to each other (Leontis and Westhof, 2001). Non-Watson-Crick basepairs result when either or both bases interact at their Hoogsteen or sugar edges. All basepairs, including Watson-Crick and non-Watson-Crick, can be classified geometrically into 12 families by noting 1) the base edges used to form hydrogen bonds and 2) the relative orientations (cis or trans) of the glycosidic bonds of the interacting nucleotides (Leontis and Westhof, 1998a; Leontis and Westhof, 2001). Two examples of these families, the cis and trans Hoogsteen/sugar edge basepairs, are shown in the lower panel of Figure 1.2. In non-Watson-Crick basepairs, the Watson-Crick edges may be available for further interactions – with other RNAs, proteins, or small molecules. In fact, one base can potentially pair with up to three other bases at the same time, allowing the formation of complex motifs involving base-triples, quadruples and etc.

Certain RNA 3D motifs, such as loop E, recur in non-homologous RNA molecules or in distinct sites of the same RNA molecules and play important roles by interacting with other cellular factors that regulate diverse biological functions (Simons and Grunberg-Manago, 1998; Leontis et al., 2002a). Recurrent 3D motifs comprise sets of
nucleotides with similar spatial arrangements, including a core of conserved non-Watson-Crick basepairs. The 3D structures of recurrent motifs are more conserved than their sequences (Leontis and Westhof, 1998a; Leontis and Westhof, 1998b; Leontis and Westhof, 1998c; Leontis et al., 2002a). The set of base substitutions compatible with the 3D structure of a motif is its sequence signature. Like the Watson-Crick basepairs, A/U and G/C, which can substitute for each other in double helices, certain non-Watson-Crick basepairs can substitute for each other in motifs without distorting the 3D structure of the motif; such basepairs are isosteric (Michel and Westhof, 1990; Leontis and Westhof, 1998c; Leontis et al., 2002a). Isostericity Matrices summarize the isosteric relationships for each geometric basepairing family and provide the basis for analyzing the sequence signatures of RNA motifs (Leontis et al., 2002b).

In this paper we present our studies on the PSTVd loop E tertiary structure and its function in PSTVd replication by combining Isostericity Matrix analysis with genetic and functional assays. Our results provide the first loss-of-function genetic evidence for the essential role of a viroid 3D motif in replication. They also demonstrate that the integration of rationalized mutagenesis, based on Isostericity Matrix analysis of non-Watson-Crick basepairing, with in vitro and in vivo functional assays provides a novel, comprehensive approach to investigate the precise tertiary structure-function relationships for a wide range of RNA motifs in diverse biological processes.
1.3 MATERIALS AND METHODS

1.3.1 Isostericity matrix analysis of PSTVd loop E

In previous work, basepairs belonging to each geometric base-pairing family were grouped into isosteric subgroups and displayed in 4x4 Isostericity Matrices, one for each geometric family (Leontis et al., 2002b). In the Isostericity Matrices, basepairs belonging to an isosteric subgroup “n” are designated “I_n”. For each non-Watson-Crick basepair composing an RNA motif, the Isostericity Matrices are consulted to determine which mutations result in 1) basepairs belonging to the same isosteric subgroup, 2) basepairs belonging to nearly isosteric groups, 3) base juxtapositions that cannot form basepairs of the same geometric type, or 4) base juxtapositions that can form basepairs but which are non-isosteric. In cases 3) or 4), the 3D structure of the motif will be altered or even disrupted. For cases 1) and 2), if one or more base participates in a second or third basepair, the Isostericity Matrices corresponding to those basepair types must also be consulted to make sure that the mutation is compatible with all possible interactions. Three criteria were used to group basepairs in isosteric subgroups. The first criterion is that all basepairs share the same relative orientation (cis or trans) of the glycosidic bond vectors. The second criterion is that the distance between the C1’ atoms of the paired nucleotides be the same (nominally, within about 1 Å). The third criterion is that hydrogen-bonding takes place between equivalent base positions. Basepairs that meet all three criteria are considered isosteric, whereas basepairs that meet only the first two criteria are called nearly isosteric (Lescoute et al., 2005). Nearly isosteric basepairs, such as the “wobble” basepairs in the cis Watson-Crick/Watson-Crick geometric family, exhibit generally small lateral shifts at the
interacting edges and while not exactly isosteric, can sometimes substitute for each other without major disruption of the structure of a motif.

Additional examples of isosteric basepairs, published since the original Leontis et al. (2002b) compilation, were obtained using the geometrical RNA motif search program, FR3D (Find RNA 3D) (Sarver et al., 2007).

### 1.3.2 Plasmid construction

Plasmid pRZ6-2 carries cDNA of PSTVd<sup>Int</sup> (Gross et al., 1978) that is flanked by ribozymes at both ends. It was constructed by Hu et al (1997) and kindly provided by Dr. Robert Owens. All mutants were generated by site-directed mutagenesis using Quickchange Site-Directed Mutagenesis kit (Stratagene, La Jolla, CA) according to manufacturer’s instructions. The introduced mutations were confirmed by sequencing.

Plasmids pInter(+) and pInter(-), which were used as templates for <em>in vitro</em> synthesis of riboprobes specific for (-) and (+)-PSTVd RNAs, respectively, were described in Qi and Ding (2002).

The full length cDNA of EGFP (enhanced green fluorescent protein) was amplified by PCR using specific primers and pEGFP-1 as the template (Clontech, Mountain View, CA) and cloned into pGEM4Z (Promega, Madison, WI) at <i>Eco</i>R<sub>I</sub> and <i>Bam</i>H<sub>I</sub> sites to give rise to pGEM4Z:EGFP to be used for <em>in vitro</em> transcription.

### 1.3.3 In vitro transcription

To prepare PSTVd inoculum, plasmids containing the PSTVd cDNAs were linearized with HindIII. The linearized plasmids were used as templates for <em>in vitro</em>
transcription using the T7 MEGAscript kit (Ambion, Austin, TX) according to manufacturer’s instructions. Because of the nature of restriction site used for the DNA cloning, the PSTVd transcripts after ribozyme self-cleavage have G88 at the 5’ end and G87 at the 3’ end. These unit-length PSTVd transcripts were used as the inoculum.

To prepare PSTVd RNA substrates for UV crosslinking or ligation analyses, in vitro transcription was performed using the T7 MAXIscript Kit (Ambion) in the presence of [α-32P]-UTP. To make GFP transcripts that served as control substrates in UV crosslinking experiments, plasmid pGEM4Z:EGFP was linearized by EcoRI to serve as the template for in vitro transcription. The 32P-labeled PSTVd in vitro transcripts were separated on 5% polyacrylamide/8 M urea gel, and unit-length monomeric bands were eluted in 0.3 M NaCl and precipitated by ethanol.

For microinjection, PSTVd transcripts were synthesized by using the T7 MAXIscript Kit (Ambion) in the presence of Alexa Fluor 488-5-UTP (Molecular Probes, Eugene, OR). The fluorescently labelled 434 nt vector RNA transcripts were synthesized from NaeI-digested pGEMT (Promega) template by using SP6 MAXIscript Kit (Ambion).

After all transcription reactions, DNA templates were removed by DNase I treatment. RNA transcripts were purified with the MEGAClear kit (Ambion) and further quantified by UV spectrometry or scintillation counter (for radioactive riboprobes).

1.3.4 Protoplast inoculation

*N. benthamiana* suspension cells were cultured as described in Sunter and Bisaro (2003). Protoplasts were prepared and inoculated with PSTVd transcripts by electroporation as described in Qi and Ding (2002). Detailed protocols for maintenance of
N. benthamiana suspension cells, preparation of protoplasts and electroporation are described in Zhong et al. (2005). At 3 days post-inoculation, transfected protoplasts were harvested for RNA extraction and Northern blot analysis.

1.3.5 RNA extraction and Northern blot

Total RNAs were isolated from infected plants and protoplasts using the RNeasy Plant Mini Kit (Qiagen, Valencia, CA). The procedures were essentially the same as described in Qi et al., (2004b). Ten µg of RNA samples were run on 5% polyacrylamide/8 M urea gel, transferred to a Hybond-XL nylon membrane (Amersham Biosciences, Piscataway, NJ) using a vacuum blotting system (Amersham), and immobilized by UV-crosslinking. Hybridization with [α-32P]-UTP-labeled riboprobes was carried out at 65°C in ULTRAhyb reagent (Ambion). After overnight hybridization, the membranes were washed at 65°C in 2XSSC/0.1%SDS and 0.2XSSC/0.1% SDS and exposed to a Storage Phosphor Screen (Kodak, Rochester, NY). Hybridization signals were quantified with Molecular Imager FX using Quantity One-4.1.1 software (Bio-Rad, Hercules, CA).

1.3.6 Sequencing of RNA progeny

The protocols for preparing cDNAs of the PSTVd progeny isolated from protoplasts or plants were performed essentially as described by Qi and Ding (2002). Briefly, cDNAs of PSTVd RNA were RT-PCR amplified and sequenced in both directions using the ABI377 DNA sequencer (Perkin-Elmer, Boston, MA) at the DNA Sequencing Facility at Ohio State University.
1.3.7 Preparation of *N. benthamiana* nuclear extract

The nuclear extract was prepared according to Baumstark and Riesner (1995) and Roberts and Okita (1991) with slight modifications. All steps were performed on ice. The protoplasts prepared from cultured cells of *N. benthamiana* (see above) were first resuspended in 15 ml of buffer A (20 mM MES-KOH of pH 5.8, 20 mM K-acetate, 15% Ficoll 400, 0.15 mM spermine, 0.5 mM spermidine, 10 mM β-mercaptoethanol, 0.5 mM PMSF, 0.15 μM pepstatin, and 0.6 μM leupeptin) and then lysed in a Dounce homogenizer with 10-15 strokes. DAPI staining was used to monitor the release of nuclei. The nuclei were purified by loading the homogenates onto a two-step gradient consisting of 8 ml of buffer B [87.6% (v/v) Percoll, 0.62 x buffer C (buffer A minus Ficoll)] and 12 ml of buffer D (buffer A with 18% instead of 15% Ficoll 400). The mixtures were centrifuged at 4,000 g for 60 min at 4 °C. The interphase containing the nuclei was collected with a glass pipette and 1.5 volume of buffer C was added. The mixtures were centrifuged at 1,400 g for 30 min at 4 °C and the pelleted nuclei were resuspended in buffer E (60 mM HEPES-KOH pH 7.9, 0.12 mM EDTA, 0.84 mM Mg-acetate, 0.72 μM leupeptin, 0.18 μM pepstatin, 12 mM β-mercaptoethanol, and 0.6 M KCl). The sample was stirred gently for 1 h at 4 °C and centrifuged at 36,000 g for 60 min at 4 °C, using a L8-80M Ultracentrifuge with SW41 rotor (Beckman, Fullerton, CA). Saturated (NH₄)₂SO₄ was slowly added to the resulting supernatant to a final concentration of 75% and centrifuged at 36,000 g for 30 min at 4 °C to precipitate proteins. The pellet was resuspended in buffer F (20 mM HEPES-KOH pH 7.9, 10 mM Mg-acetate, 50 mM K-acetate, 5 mM EDTA, 12 mM β-mercaptoethanol and 25% glycerol) and dialyzed overnight at 4 °C in Buffer F. Aliquots were quickly frozen in liquid nitrogen and stored at –80 °C.
1.3.8 RNA circularization assay

The circularization assay was conducted by incubating $^{32}$P-labeled and gel-purified unit-length PSTVd *in vitro* transcripts and control RNAs in *N. benthamiana* nuclear extract (NbNE). In a final volume of 20 µl reaction, $1 \times 10^5$ cpm of linear unit-length (+)-PSTVd transcripts were mixed with 4 µl of 5x ligation buffer (100 mM Tris-HCl of pH 8.0, 30 mM Mg-acetate, 1 mM spermidine and 2 mM EDTA; Owens et al., 1991), 1 µl of RNaseOut (Invitrogen) and 10 µl NbNE and incubated at 37°C for 2-3 h. The reaction was terminated by adding 260 µl of 2x proteinase K buffer (200 mM Tris-HCl pH 8.0, 25 mM EDTA pH 8.0, 1% SDS) and 10 µl of 20 mg/ml proteinase K and incubated at 65°C for 30 min followed by phenol extractions (phenol was saturated with 10 mM Tris-HCl buffer pH 4.0). The reaction products were ethanol-precipitated and analyzed on 5% polyacrylamide/8 M urea gel.

1.3.9 UV-crosslinking to analyze PSTVd loop E

UV crosslinking was performed according to Schrader et al (2003) under conditions that favor formation of the stable PSTVd secondary structure (equivalent to ExL conformation of their substrate) containing loop E. Briefly, $^{32}$P-labeled *in vitro* transcripts of gel purified unit-length PSTVd transcripts ($\sim 10^5$ cpm) were pre-incubated in high ionic strength buffer (500 mM NaCl, 4 M urea, 1 mM Na-cacodylate, 0.1 mM EDTA at pH 7.9) at 40°C for 45 min followed by slow cooling to room temperature overnight. The reaction mixtures in Eppendorf tubes were placed on ice and irradiated with 258 nm-UV at different
time intervals using a Stratalinker (Model 1800, Stratagene) in a cold room. The samples were ethanol-precipitated and analyzed on 5% polyacrylamide/8 M urea gel.

### 1.3.10 Microinjection and microscopy

A small patch of epidermis was removed from the lower surface of a *N. benthamiana* leaf and covered with distilled water immediately. The leaf was mounted onto a glass plate to fit the stage of a Nikon E600 epifluorescence microscope (Nikon, Tokyo, Japan). Fluorescently labeled RNA transcripts were loaded into glass pipettes made from thin-walled glass tubes (World Precision Instruments, Sarasota, FL) on a pipette puller (Model PB-7, Narishige, Tokyo, Japan). Injection was performed with an MMO-203 Micromanipulator (Narishige) and 1M-5B Injector (Narishige). Subcellular localization of the injected RNA transcripts was visualized with a filter set consisting of an EX of 450-490 nm, a DM of 510 nm, and a BA of 520-560 nm. Images were recorded and processed with a SPOT 2 Slider CCD camera and the associated software (Diagnostics Instruments Inc., Sterling Heights, MI). DAPI was injected into the same cells to visualize the nuclei after imaging of RNA localization. DAPI fluorescence was visualized with a filter set consisting of an excitation filter (EX) of 330-380 nm, a dichroic mirror (DM) of 400 nm and a barrier filter (BA) of 435-485 nm and recorded as above.

### 1.3.11 Quantitative real-time RT-PCR

Total RNA isolated from *N. benthamiana* protoplasts was treated with TURBO DNA-free™ kit (Ambion) according to manufacturer's instructions. One µg of total RNA
was reverse-transcribed using ThermoScript™ RT-PCR system (Invitrogen) with primer (+)-97-125 (5’-CCTTTTTTGCCAGTTCGCTCCAGGTTTCC-3’) for (+)-PSTVd, primer (-)-13-39 (5’-CGTGGTTCCTGTGGTTCACACCTGACC-3’) for (-)-PSTVd, and primer 1093-1114R (5’-CCC GagaACCAAAACTTTTG-3’) for 18S rRNA. The ThermoScript reverse transcriptase was inactivated by incubating the mixture at 85 °C for 5 min. The template RNA was removed by RNase H treatment at 37 °C for 1 h. Real-time PCR was performed with QuantiTect SYBR Green PCR kit (Qiagen) in a PCR reaction mixture containing 2 µl of cDNA sample (the RT products diluted by 4 times), 0.5 µM of forward and reverse primers and 1X QuantiTect SYBR Green PCR Master Mix. Primer combinations for real-time PCR are as follows: primer (+) 97-125 and primer (+) 13-39 for amplification of either (+)- or (-)-PSTVd cDNA templates, and primer 1093-1114 R and primer 1000-1020 F (5’-GATCAGATACCAGTCCTAGTC-3’) for amplification of 18S rRNA cDNA template. The real-time cycler conditions for LightCycler system (Roche, Indianapolis, IN) were 10 min at 95 °C for activation of HotStar Taq DNA Polymerase followed by 40 cycles of reactions (95 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s). The melting curve analysis was performed to determine the specificity of the amplified products. Positive (standard curve) and negative (no template) controls were measured in each of the PCR runs. The standard curve was generated by serial dilutions ranging from 10^0 to 10^{10} copies of the pRZ:PSTVd<sub>int</sub>. 18S ribosomal RNA was used for normalization of PSTVd RNA accumulation levels. PSTVd copy numbers were calculated using the standard curve. The PCR results were analyzed with LightCycler Software Version 3.5 (Roche, Indianapolis, IN). For the wild type PSTVd and each of the mutants, results from six replicate RT-PCR experiments (three from samples of each of the two biological
replicates) were used to calculate the mean and standard error. To facilitate comparison, the accumulation level of PSTVd\textsuperscript{int} is arbitrarily set to a value of 1 and the levels of the mutants are presented as relative values.

1.4 RESULTS

1.4.1 Tertiary structural model for the PSTVd loop E motif

To facilitate genetic studies of PSTVd loop E function, we first analyzed the tertiary structure of this motif by comparative sequence analysis using Isostericity Matrices (Leontis et al., 2002b). Sequence analysis has identified recurrent loop E or sarcin/ricin motifs in different RNA molecules (Leontis and Westhof, 1998a; Leontis et al., 2002a). This motif comprises five core non-Watson-Crick basepairs, as shown in Figure 1.3A. On the basis of UV crosslinking (Branch et al., 1985), nuclease and chemical mapping (Gast et al., 1996) and sequence comparison, it was concluded that a loop E (sarcin/ricin) motif also exists in PSTVd at the position indicated in Figure 1.3B. Figure 1.4 shows the inferred basepair structures with hydrogen bonds for the five core non-Watson-Crick basepairs (A261/G98, U260/A99, U260/C259, A258/A100, and A101/U257) in PSTVd loop E. Structures for some common, isosteric sequence variants are also shown. Each non-Watson-Crick basepair and its sequence variations in the PSTVd loop E motif are discussed below. Following the convention (Leontis and Westhof, 2001), the edges that participate in the interactions for each base pair are presented in the following order: Watson-Crick Edge, Hoogsteen Edge and Sugar edge.
**A261/G98 trans Hoogsteen/sugar edge pair:** The Hoogsteen edge of A261 forms hydrogen bonds with the sugar edge of G98 in the *trans* orientation. The *trans* Hoogsteen/sugar edge geometric family comprises two isosteric subfamilies (I₁ and I₂) (Leontis et al., 2002b); see METHODS for specific criteria of basepair classifications). The A261/G98 pair belongs to I₁ (see Figure 1.5). The isosteric C/U double mutant, discussed below, is shown in Figure 1.4.

**U260/C259 cis Hoogsteen/sugar edge pair:** In the PSTVd sequence, U260 and C259 may form a *cis* Hoogsteen/sugar edge basepair. In most loop E (sarcin/ricin) motifs, the corresponding bases are U and G, which form a *cis* Hoogsteen/sugar edge basepair stabilized by a strong hydrogen bond between U (O4) (oxygen at position 4 of uracil) and G (N2) (nitrogen at position 2 of guanine). *Cis* Hoogsteen/sugar edge basepairs between successive U/U, U/C, and U/A are observed infrequently in crystal structures. While they have C1’-C1’ distances comparable to the U/G pair, they lack the strong hydrogen bond. Consequently, these pairs are usually stabilized by interactions with other bases, forming triples, or with proteins. The isosteric U/U *cis* Hoogsteen/sugar edge pair has been observed in several crystal structures, including the 23S rRNA (PDB file 1S72, U831/U832 from Protein Data Bank), where it is part of a base triple. On the basis of observed U/U pairs, the isosteric U/C pair was proposed (Leontis et al., 2002b) and has since been observed (for example, U223/C222, chain A, PDB file 1ET4). The U/U and U/C pairs are stabilized in part by weak hydrogen bonds between U (C5) (carbon atom at position 5 of uracil) and Y (O2) (oxygen at position 2 of pyrimidine).

The U/A pair can form by weak hydrogen-bonding between A (C2) (carbon at position 2 of adenine) and U (O4) or between A (O2’) (the 2’-OH of adenosine) and U (C5).
One example of this type of basepair occurs in association with a protein and indicates that U/A has a C1’-C1’ distance of 6.8Å (PDB file 1EC6, U12/A11 in C chain), and thus can also substitute isosterically with U/U, U/C, and U/G. The RNA structural search programs used at the time of the original compilation of basepairs (Leontis et al., 2002b) did not detect the U/A pair.

**U260/A99 trans Watson-Crick/Hoogsteen edge pair:** The Watson-Crick edge of U260 hydrogen-bonds with the Hoogsteen edge of A99 in the trans orientation. The U at this position is very conserved in most loop E motifs because it also uses its Hoogsteen edge to form a second basepair with G (i.e., a cis Hoogsteen/sugar edge pair). In other contexts, however, C can substitute for U in trans Watson-Crick/Hoogsteen basepairs to form isosteric C/C, or nearly isosteric C/A or C/G basepairs of this type (Leontis et al., 2002b).

**A258/A100 trans Hoogsteen/Hoogsteen edge pair:** The Hoogsteen edges of A258 and A100 interact to form a symmetric trans Hoogsteen/Hoogsteen basepair. Mutation of A100 to G still allows for hydrogen-bonding with A258 by formation of a hydrogen bond between A (N6) (nitrogen at position 6 of adenine) and G (O6) (oxygen at position 6 of guanine). This substitution has been observed in a composite loop E motif in 23S rRNA of *H. marismortui* (PDB file 1S72, A913/G1071) with a C1’-C1’ distance of 12.2 Å. Thus, the A/G basepair is nearly isosteric with A/A.

**A101/U257 trans Hoogsteen/sugar edge pair:** The Hoogsteen edge of A101 hydrogen-bonds with the sugar edge of U257. Note that this basepair is oriented oppositely to the trans Hoogsteen/sugar edge pair A261/G98, as shown by the symbols annotating these basepairs in Figure 1.4. The structures of three isosteric variants, A/G, A/C, and A/A
are shown in Figure 1.4. Very often, the isosteric C/U basepair variant occurs at this position, as shown for the generic sarcin/ricin loop in Figure 1.3A. The structure of the C/U basepair is shown in the upper inset of Figure 1.4.

**C256/C102 cis Watson-Crick/Watson-Crick Bifurcated pair:** This pair is outside the conserved core structure of loop E and sarcin/ricin motifs. For C/C to form an edge-to-edge Watson-Crick basepair, one cytosine must be protonated at the N3 position. While this has been observed in some structures, more commonly the cytosines adopt the Watson-Crick bifurcated geometry in which the N4 amino group of one cytosine hydrogen-bonds with both N3 and O2 of the other cytosine. This bifurcated geometry results in a longer C1’-C1’ distance of about 10.4Å, which is nearly isosteric to cis Watson-Crick A/U and G/C basepairs or with wobble Watson-Crick basepair U/G or C/A (Leontis et al., 2002b).

### 1.4.2 The tertiary structural model of PSTVd loop E accounts for natural sequence variations and viable mutation

To test the validity of the proposed tertiary structural model of the PSTVd loop E, we first examined whether this model could account for the functional state of the reported natural sequence variations or viable mutations in loop E motifs of PSTVd and other viroid species in the *Pospiviroidae* family. The natural sequence variants and viable mutants are summarized in Table 1.1. Figure 1.4 illustrates the molecular structures of selected sequence variants, mapped onto the annotated non-Watson-Crick basepairing structure that forms the core of PSTVd loop E motif. Figure 1.5 shows the functional status of mutants
and natural variants mapped onto the respective Isostericity Matrices for each basepair, according to the geometric family to which it belongs (Leontis et al., 2002b).

The C259U substitution is found in a natural PSTVd variant, together with the C256A substitution, and several other nucleotide changes outside loop E (Owens et al., 1992). The C→U substitution at this position is also detected in the loop E-like motif of a natural variant of Citrus exocortis viroid (CEVd), a member of the Pospiviroidae family (Semancik et al., 1993). These are operationally classified as natural variations (NV) in Table 1.1. Furthermore, the C259U substitution was detected in the PSTVd progeny in transgenic tobacco plants expressing the wild type PSTVd sequences (Wassenegger et al., 1996; Zhu et al., 2002). These are classified as spontaneous mutations (SM) in Table 1.1.

The experimentally constructed C259G and C259A substitutions, designated as engineered mutations (EM) in Table 1.1, retained PSTVd viability as well as the C259U SM in tobacco and N. benthamiana protoplasts (Qi and Ding, 2002). These natural variations and viable mutations can all be accounted for by the fact that U260 will form isosteric cis Hoogsteen/sugar edge pairs with A, C, G, or U at position 259 of PSTVd or at the equivalent position within loop E of CEVd. Cis Hoogsteen/sugar edge basepairs can occur between immediately successive nucleotides, as in the PSTVd loop E motif, or between nucleotides separated by one or more other nucleotides. The first type comprises isosteric subgroup I₁ and the second type, subgroup I₂. At the time of the 2002 compilation of base pairs (Leontis et al., 2002b), the only example of cis Hoogsteen/sugar edge U/A detected by the computation search programs was of type I₂. The recently developed FR3D (Find RNA 3D) search program (Sarver et al., 2007), however, identified type I₁ U/A cis Hoogsteen/sugar edge basepair (in this case in complex with a protein) (PDB file 1EC6,
U12/A11 in C chain). The C1’-C1’ distance 6.8Å makes this basepair isosteric with U/C (7Å), U/G (6.7Å) and U/U (7Å) pairs (Leontis et al., 2002b). Therefore, all of these mutations could maintain the correct tertiary structure of loop E, if the corresponding bases are paired (in some cases the pairing may occur only in the presence of another factor, such as a protein, that stabilizes this interaction). Because loop E motifs in cellular RNAs are well known to serve as important binding sites for RNAs and proteins (Leontis et al., 2002a), it will be of great interest to determine in future studies whether binding of a cellular protein to the PSTVd loop E contributes to the stabilization of U/C and its viable variant U/U, U/G and U/A basepairs.

It should be noted that C259G mutation in loop E was shown to inhibit in vitro 3’ cleavage and ligation of a minicircle RNA derived from the CCR of PSTVd (Schrader et al., 2003). Given that 1) U/G cis Hoogsteen/sugar edge basepair is common at the equivalent position in other loop Es (see Figure 1.3 for the generic sarcin/ricin motif; Leontis et al., 2002b), 2) the C259G mutation is not expected to alter the tertiary structure of PSTVd, and 3) this mutation does not inhibit PSTVd replication in tobacco and N. benthamiana protoplasts (Qi and Ding, 2002) and in potato plants (data not shown), the reason for its effect on the in vitro processing of the minicircle RNA remains to be understood.

Zhu et al. (2002) detected a U257A spontaneous mutation in transgenic tobacco expressing the PSTVd wild type sequence. Experimentally constructed U257G and U257C substitutions retained PSTVd viability in tobacco and N. benthamiana protoplasts (Qi and Ding, 2002). Interestingly, a U→G substitution occurs at the equivalent position in loop E motifs of several CEVd natural variants (Visvader and Symons, 1985; Semancik et al.,
1993; Skoric et al., 2001) as well as *Mexican papita viroid* (MPVd), *Tomato planta macho viroid* (TPMVd) and *Tomato apical stunt viroid* (TASVd) (Singh et al., 2003). Furthermore, a U→C substitution occurs in CEVd-C at this position (Gross et al., 1982). All of these observations can be accounted for by the fact that A101 can form isosteric *trans* Hoogsteen/sugar edge pairs with A, C, G or U at position 257 of PSTVd and the equivalent position in loop Es of all the other viroids (Table 1.1; Fig. 1.4, 1.5). All these basepairs belong to the same isosteric subgroup, I1, of the *trans* Hoogsteen/sugar edge geometric family.

The A100G substitution does not affect *in vitro* cleavage/ligation of a mini-circle RNA containing repeated sequences of the CCR of PSTVd (Schrader et al., 2003). This observation was at first difficult to explain because the A/G *trans* Hoogsteen/Hoogsteen basepair has a C1′-C1′ distance of 12.9Å in Leontis et al. (2002b), which places it in a different isosteric subfamily than A/A *trans* Hoogsteen/Hoogsteen basepair having a C1′-C1′ distance of 11Å (Leontis et al., 2002b). However, since the publication of the original compilation of basepairs (Leontis et al., 2002b), we have identified other examples in higher resolution structures for some of the basepairs reported. Thus we have a higher resolution example (2.4Å vs. 3.1Å) of an A/G *trans* Hoogsteen/Hoogsteen basepair - A913/G1071 in the 50S subunit of *H. marismortui* (PDB file 1S72). In this basepair the C1′-C1′ distance is 12.2Å, which is closer to the 11Å seen for A/A basepair of this type. More significantly, this basepair is actually part of a loop E motif (Leontis et al., 2002b) and thus shows directly that A/G can substitute for A/A at position 258/100. Whether this substitution affects the biological activities of PSTVd *in vivo* remains to be determined.
Outside the non-Watson-Crick basepair core of loop E, the C256/C102 cis Watson-Crick/Watson-Crick bifurcated pair also shows viable sequence variations. Owens et al. (1992) reported C256A substitution in a natural variant of PSTVd (together with C259U substitution). Nucleotide A is also found at the equivalent position in MPVd and TASVd (Singh et al., 2003), and A or C is found at the equivalent position in CEVd (Visvader and Symons, 1985; Semancik et al., 1993). These variants can be accounted for by the fact that A at this position in the motif can form a wobble base pair with C102 of PSTVd and with C at the equivalent position in the other viroids. The C1’-C1’ in the C/A wobble Watson-Crick pair is ~10.4Å, almost identical to the C1’-C1’ distance in the bifurcated Watson-Crick C/C pair, so the A/C and C/C basepairs are nearly isosteric (Leontis et al., 2002b).

In conclusion, we find that all reported biologically viable sequence variations or mutations in viroid loop E motifs can be accounted for by the isosteric basepairing rules as summarized in the Isostericity Matrices, within the scope of the literature of which we are aware. This supports the notion that an intact loop E motif is required for PSTVd viability. Furthermore, this argues for the functional significance of intact loop E in other viroids in which this motif occurs. In particular, the loop Es of Chrysanthemum stunt viroid (CSVd) and PSTVd have identical sequence signatures (Keese and Symons, 1985).

1.4.3 Application of isostericity matrix analysis to design disruptive mutations in PSTVd loop E motif

To investigate the function of the PSTVd loop E motif in replication by loss-of-function genetics, we applied Isostericity Matrix analysis to design mutations that
disrupt this motif, based on the tertiary structural model in Figure 1.3B. Changing A99 to C (mutation A99C) prevents the trans Watson-Crick/Hoogsteen interaction between positions 260 and 99, because the Watson-Crick edge of U260 cannot hydrogen-bond with the Hoogsteen edge of C99 (Leontis et al., 2002b). This is indicated in the Isostericity Matrices for trans Watson-Crick/Hoogsteen basepairs by the red box for the matrix element corresponding to the U row and the C column (Fig. 1.5; Leontis et al., 2002b). Likewise, changing A261 to C prevents the trans Hoogsteen/sugar edge basepair at positions 261/98 from forming because the Hoogsteen edge of C261 cannot hydrogen-bond with the sugar edge of G98. This is indicated by the red box at the corresponding positions (C row, G column) of the Isostericity Matrices for trans Hoogsteen/sugar edge basepairs (Fig. 1.5; Leontis et al., 2002b). Computer programs that calculate minimum free energy secondary structures, such as mfold, do not currently predict the formation of non-Watson-Crick basepairs. Therefore, mutations affecting the structures of motifs consisting of ordered arrays of non-Watson-Crick basepairs, such as loop E, are also not correctly accounted for by mfold and similar programs.

To test physically whether the A99C and A261C mutations each indeed disrupt the tertiary structure of loop E, we subjected the corresponding mutants to UV crosslinking experiments. These mutants are referred to as PSTVd\textsuperscript{Int}A99C and PSTVd\textsuperscript{Int}A261C, respectively. The names indicate that they are derived from the wild type PSTVd Intermediate strain (PSTVd\textsuperscript{Int}; Gross et al., 1978). Previous studies have demonstrated that UV treatment specifically crosslinks the bases G98 and U260 in loop E of PSTVd (Branch et al., 1985; Baumstark et al., 1997) and of a minicircle-RNA derived from the PSTVd CCR (Schrader et al., 2003). It also crosslinks the corresponding G and U bases in the loop
E of human 5S rRNA from HeLa cells (Branch et al., 1985). The crosslinking occurs because of a unique partial cross-strand stack of G on U in loop E motifs that gives rise to photoreactivity. The crosslinking causes RNA mobility retardation on denaturing gels (Fig. 1.6). We used unit-length linear in vitro transcripts as the substrates. These transcripts are predicted to fold into the same rod-shaped secondary structure as the circular RNA, except for the nick between the two ends (G88 and G87; see MATERIALS AND METHODS for details). As shown in Figure 1.6, A99C and A261C mutations greatly reduced the accumulation of crosslinked products. For instance, after 1 min of UV irradiation, the crosslinked product was already evident for PSTVdInt RNA, but barely visible for the two mutants. After 5 min of UV irradiation, 32% of the wild type RNA was crosslinked whereas only 6-7% of the mutant RNAs were crosslinked. The similar sizes of the crosslinked products in the wild type and mutant PSTVd RNAs suggest that UV crosslinking occurred in the same region of these RNAs. Furthermore, the specificity of the G98-U260 crosslinking was demonstrated with two mutants that lack specific G98-U260 base stacking in loop E (Fig. 1.6). The absence of crosslinked products from green fluorescent protein (GFP) RNAs further demonstrates the specificity of UV crosslinking (Fig. 1.6). The quantitative difference indicates that a substantial portion of the PSTVdIntA99C and PSTVdIntA261C mutant RNA molecules failed to form the correct loop E structure so as to place G98 and U260 into the necessary proximity to allow UV crosslinking, in contrast to the wild type transcripts under the same incubation conditions. The band observed between the crosslinked product and the substrate RNA was also observed in previous experiments and its nature remains to be determined (Schrader et al., 2003).
1.4.4 Maintaining the tertiary structure of loop E is necessary for PSTVd accumulation in inoculated protoplasts

The availability of the loop E-defective mutants allowed us to test whether the intact tertiary structure of loop E is crucial for PSTVd infection. To this end, we inoculated *N. benthamiana* protoplasts with *in vitro* transcripts derived from the wild type PSTVd$^{\text{Int}}$ and loop E-defective single mutants PSTVd$^{\text{Int}}$A99C and PSTVd$^{\text{Int}}$A261C, respectively. Northern blots, run to detect the circular and linear (+)-strand PSTVd, showed no visible accumulation of either PSTVd$^{\text{Int}}$A99C or PSTVd$^{\text{Int}}$A261C in protoplasts (Fig. 1.7), in contrast to PSTVd$^{\text{Int}}$.

These data suggest that maintaining the intact tertiary structure of loop E is critical for PSTVd accumulation in the inoculated protoplasts. To further test this, we applied Isostericity Matrix analysis to design a compensatory mutant in which G98 is replaced by U in the PSTVd$^{\text{Int}}$A261C background. In the double mutant (PSTVd$^{\text{Int}}$A261C/G98U), the *trans* Hoogsteen/sugar edge C261/U98 basepair can form. This C261/U98 basepair is isosteric with the original A261/G98 basepair and similar substitutions have been observed in the crystal structures of related motifs (Leontis and Westhof, 1998b; Leontis et al., 2002a; see Fig. 1.4, 1.5). We found that, indeed, this double mutant accumulated as well as PSTVd$^{\text{Int}}$ in the inoculated protoplasts (Fig. 1.7). Sequencing of the PSTVd progeny confirmed that no other mutations occurred during replication of the double mutant. To extend this finding, we generated a second compensatory mutant in which U260 is replaced by C in the PSTVd$^{\text{Int}}$A99C background. In this double mutant, PSTVd$^{\text{Int}}$U260C/A99C, the *trans* Watson-Crick/Hoogsteen C260/C99 basepair that is
isosteric with the original U260/A99 basepair can form (Leontis et al., 2002b). Importantly, C260 and C259 may also form a cis Hoogsteen/sugar edge basepair isosteric with the original U260/C259 basepair (Fig. 1.5; Leontis et al., 2002b). Protoplast infection assay showed that this double mutant also restored replication function (Fig. 1.7). Taken together, all these data provide compelling evidence that the tertiary structure of loop E is essential for PSTVd accumulation in the inoculated protoplasts. Furthermore, these experimental results demonstrate the value of Isostericity Matrix analysis as a tool to design mutations that disrupt or restore non-Watson-Crick basepairs in functional studies of RNA motifs.

1.4.5 A99C and A261C mutations do not impair nuclear import function

Following delivery into the cytoplasm of protoplasts and leaf cells, the viroid transcripts must first enter the nucleus to be replicated. To determine whether the diminished accumulation of mutants PSTVd\textsuperscript{Int}A99C and PSTVd\textsuperscript{Int}A261C in the inoculated protoplasts is due to defects in nuclear import, we injected fluorescently labeled, (+)-strand in vitro transcripts into the cytoplasm of leaf cells. Both single mutants, like the wild type, appeared in the nucleus to visible levels under a fluorescence microscope 20 to 30 min after injection (Fig. 1.8; Table 1.2). Fluorescently labeled in vitro transcripts derived from vector sequences did not show nuclear accumulation (Fig. 1.8; Table 1.2), indicating specificity of nuclear import of the PSTVd RNAs. Therefore, the diminished accumulation of the mutants in the inoculated protoplasts can be attributed to their impaired capacity to replicate within the nucleus, rather than to defects in nuclear import. These results demonstrate that the intact tertiary structure of loop E is required for PSTVd replication in
the nucleus. Furthermore, the fact that the fluorescently labeled transcripts of PSTVd\textsuperscript{Int}A99C, PSTVd\textsuperscript{Int}A261C and PSTVd\textsuperscript{Int} are imported into the nuclei and accumulate there, unlike the cytoplasmic accumulation of vector transcripts, indicate that A99C and A261C mutations do not make PSTVd RNA more vulnerable to degradation.

1.4.6 A99C and A261C mutations do not affect circularization of the unit-length inoculum transcripts

According to the rolling circle replication model, the incoming circular PSTVd molecules serve as templates for the synthesis of the (-)-strands. Therefore, upon entering the nucleus, the linear, unit-length \textit{in vitro} transcripts of PSTVd may need to be ligated into circular molecules prior to serving as the initial transcription templates. Alternatively, it may be possible for the PSTVd-transcription enzyme complex to transcribe the linear RNA transcripts by jumping across the nick using a template switching mechanism, as does the HDV-transcription enzyme complex during rolling circle replication in mammalian cells (Chang and Taylor, 2002). Assuming, however, that circularization of the input PSTVd transcripts is necessary for transcription, one possible reason for the failure of single mutants PSTVd\textsuperscript{Int}A99C and PSTVd\textsuperscript{Int}A261C to replicate is that their linear \textit{in vitro} transcripts cannot be ligated at G88 and G87 (the 5’ and 3’ ends, respectively, of the transcripts; see MATERIALS AND METHODS for details) to form circular molecules. To test circularization of PSTVd\textsuperscript{Int}A99C and PSTVd\textsuperscript{Int}A261C, we incubated gel-purified unit-length linear (+)-PSTVd RNA transcripts of the wild type and mutants in \textit{N. benthamiana} nuclear extract (NbNE). As shown in Figure 1.9, all RNA substrates were
properly circularized. As a control, GFP RNAs showed no circularization. Therefore, the two mutations do not affect circularization of PSTVd RNA inocula.

1.4.7 A99C and A261C mutations inhibit the accumulation of (-)-strand PSTVd

Given that PSTVd\textsuperscript{Int}A99C and PSTVd\textsuperscript{Int}A261C proved to be competent for nuclear import and ligation, we asked whether they were defective in transcription. One method to address this question was to examine the accumulation of (-)-PSTVd RNAs, the replication intermediate. Northern blots did not detect the presence of the (-)-PSTVd RNAs for either mutant (data not shown). Because the (-)-strand PSTVd RNA is present at low levels during replication of even the wild type PSTVd\textsuperscript{Int} (Qi and Ding, 2002), we used the more sensitive quantitative real-time reverse transcription-polymerase chain reaction (real-time RT-PCR) to quantify the (-)-strands. The 18S ribosomal RNA served as an internal control template in each reaction. As shown in Figure 1.10, the accumulation levels of the (-)-strand RNAs from PSTVd\textsuperscript{Int}A99C and PSTVd\textsuperscript{Int}A261C were approximately 1,000-fold lower, respectively, than that of the wild type. In contrast, the accumulation level of (-)-strand PSTVd\textsuperscript{Int}A261C/G98U was approximately 5-fold lower than that of the wild type PSTVd\textsuperscript{Int}. The accumulation level of PSTVd\textsuperscript{Int}U260C/A99C was approximately 10-fold higher than that of PSTVd\textsuperscript{Int}A99C, showing notable restoration of accumulation. Similar accumulation patterns were detected for the (+)-PSTVd RNAs, which is consistent with Northern blot analysis. These data suggested that the (+)-strand PSTVd\textsuperscript{Int}A99C and PSTVd\textsuperscript{Int}A261C transcripts inoculated into the protoplasts were inefficiently transcribed into the (-)-strands, which further led to reduced synthesis and accumulation of the (+)-strand RNAs.
Whether transcription of (+)-RNA from concatemeric (-)-RNA template is affected by the mutations remains unknown. This question can be addressed once methods are developed to construct a concatemeric RNA template that contains identical mutations in each of the unit-length monomers for transcription analyses.

1.5 DISCUSSION

Viroid infection represents a remarkable case in which a small infectious RNA molecule, without encoding any proteins, contains all the genetic information and forms the appropriate 3D structure to interact with cellular factors so as to ensure its replication and systemic trafficking throughout a host plant. It is unlikely that plants have evolved mechanisms for RNA-templated RNA replication and systemic RNA trafficking simply to serve viroid RNA infection. Rather, the viroid RNA must have evolved to adapt to and exploit existing cellular mechanisms used for endogenous functions. Therefore, the biology of viroid infection likely mimics the biological roles of endogenous RNAs and mechanisms for their synthesis, processing and intra-/intercellular transport. Elucidating how PSTVd replicates should therefore shed new light on the structure and function of the eucaryal nuclear transcription machinery, on the evolution of RNA structure-function relationships, and on the evolution of RNA-based infectious agents.

We have developed a novel approach that integrates comparative sequence analysis of non-Watson-Crick basepairs using Isostericity Matrices, mutagenesis and in vitro as well as in vivo functional assays to dissect the role of one conserved RNA motif in the rolling circle replication of PSTVd. This systematic approach is necessary to pinpoint the
process(es) affected by mutations in a noncoding and multifunctional RNA so as to determine the specific biological role of an RNA structural motif.

1.5.1 A tertiary structural model for viroid loop E motif that can guide functional studies

Branch et al. (1985) first recognized that PSTVd contains an RNA motif very similar to or identical with loop E of eucaryal (or archaeal) 5S rRNA on the basis of specific UV crosslinking between the conserved U260 and G98 bases. Based on nuclease and chemical mapping data and comparisons with the NMR structures of *Xenopus laevis* (eucaryal) 5S rRNA loop E (Wimberly et al., 1993) and of the sarcin/ricin loop in 28S rRNA (Szewczak and Moore, 1995), Gast et al. (1996) suggested a scheme of non-Watson-Crick basepairing interactions in the PSTVd loop that is similar to these rRNA motifs. The Gast model depicted A261/G98, A99/U260 and A100/A258 basepairs essentially the same as in our model. However, it depicted A101/U257 as a canonical Watson-Crick basepair and did not specify the nature of the C102/C256 basepair or the possible existence of the *cis* Hoogsteen/sugar edge U260/C259 interactions. Our Isostericity Matrix analysis is based on high-resolution X-ray crystal structures of other loop E motifs and the principles of isosteric non-Watson-Crick basepairing (Leontis et al., 2002b). The tertiary structural model we present here specifies the precise nature of the edge-to-edge interactions for all the non-Watson-Crick basepairs. The model is currently well supported by several lines of biological and physical data. First, this model can explain mechanistically why certain mutations in loop E do not affect PSTVd viability (Wassenegger et al., 1996; Qi and Ding, 2002; Zhu et al., 2002; Schrader et al., 2003).
Second, it can account for the naturally occurring nucleotide sequence variations in the loop Es of PSTVd and other viroids in the *Pospiviroidae* family (Gross et al., 1982; Visvader and Symons, 1985; Owens et al., 1992; Semancik et al., 1993; Skoric et al., 2001; Singh et al., 2003). Third, UV crosslinking data provide further evidence that the A99C and A261C substitutions disrupt the loop E structure by preventing the formation of isosteric non-Watson-Crick basepairs. Fourth, as discussed below, the functional analyses pinpointed the deleterious effects of these mutations on PSTVd transcription. Finally, the two compensatory mutations that are predicted to restore the loop E tertiary structure based on isosteric basepairing principles and on the X-ray crystal structures of similar basepairs also restored PSTVd replication. Further studies on the effects of other nucleotide mutations within loop E should provide additional tests of the validity of this structural model.

### 1.5.2 PSTVd loop E plays a crucial role in RNA-templated RNA transcription

Our functional analyses of the PSTVd loop E mutations provide new insights into the role of this motif in PSTVd replication at the cellular level. Our experiments showed that the A99C and A261C mutations, which disrupt the loop E structure, inhibit accumulation of PSTVd progeny in inoculated plants and protoplasts. Further analyses indicated that these mutations do not impair nuclear import function or circularization of the linear inoculum transcripts. Therefore, the failure of the PSTVd mutants to replicate and accumulate in the inoculated plants and protoplasts likely results from a defect(s) in one or more of the steps of rolling circle replication. The requirement for intact loop E in replication is further supported by the observation that compensatory mutations predicted
to restore non-Watson-Crick basepairs within the motif were observed to restore the replication function.

The A99C and A261C loss-of-function mutations could inhibit the transcription, cleavage or ligation steps of rolling circle replication. Although our current data cannot rule out completely the possibility that these mutations affect cleavage/ligation, several observations suggest strongly that they mainly inhibit transcription. First, the A99C and A261C mutations severely inhibit the synthesis of the (-)-strand PSTVd. Second, if these mutations in fact inhibited cleavage rather than transcription, one would expect to see accumulation of high levels of the multimeric (+)-PSTVd RNAs. This was not observed on Northern blots. Third, if these mutations inhibited ligation, rather than transcription or cleavage, one would expect to see accumulation of high levels of the linear monomeric (+)-PSTVd RNAs. This was not observed either on Northern blots. Based on these considerations, we propose that the loop E motif plays a crucial role in transcription, in addition to its well-documented role in cleavage/ligation (Baumstark et al., 1997), during PSTVd replication. The multi-functionality of loop E is discussed below.

How might loop E be involved in transcription? Loop E motifs in cellular RNAs are well known to serve as important binding sites for RNAs and proteins (Leontis et al., 2002a). The loop E motif of PSTVd may serve as a binding site for the RNA polymerase or for a cellular factor(s) that recruits the RNA polymerase for transcription. Alternatively, loop E may interact with a cellular factor to localize the PSTVd RNA to a particular subnuclear site to facilitate access to the nuclear transcription machinery. It is also possible that this motif is required for interacting with a cellular factor that leads to a conformational
change in the viroid RNA, allowing it to be recognized by the transcription machinery. These possibilities can be tested by further experimental studies.

It should be noted that a hairpin I (HPI) is predicted to form as a thermodynamically metastable structure through pairing of nucleotides $^{79}$CGCUUCAGG$^{87}$ and nucleotides $^{110}$GCGAGGUC$^{102}$, with nucleotides 88-101 forming a 14-nt loop as a result of HPI formation (Riesner et al., 1979). Whether the mutations we generated would affect the structure and function of HPI requires further analyses. In fact, it remains to be determined whether HPI exists in vivo and has any biological functions. A putative tetraloop containing A99 is postulated by Baumstark et al. (Baumstark et al., 1997) to be important for in vitro processing. Whether such a tetraloop structure exists and functions in vivo, and whether A99C would affect its structure and function, remain to be investigated.

1.5.3 PSTVd loop E is a useful model to study RNA structure-function relationships of broad significance

Loop E is one of the most extensively studied RNA structural motifs (Leontis and Westhof, 1998c; Leontis et al., 2002a). Besides serving as a model to dissect PSTVd replication and the general mechanisms of RNA-templated replication by the nuclear transcription machinery, the PSTVd loop E motif offers an attractive system to study RNA structure-function relationships of general significance. First, the multi-functionality of this motif is biologically significant. This motif is involved in transcription (This Study), processing (Baumstark et al., 1997), pathogenicity (Qi and Ding, 2003), and host adaptation (Wassenegger et al., 1996; Qi and Ding, 2002; Zhu et al., 2002). To accomplish
such diverse functions, submotifs of loop E may interact with distinct cellular factors. Elucidating the underlying mechanisms will greatly expand our understanding of the capacity of a single RNA motif to regulate multiple biological processes. Such multi-functionality of a single RNA motif may be important for a wide range of RNA-based pathogens to expand their genome functions during evolution without increases in genome sizes.

Second, the loop E is a recurrent motif found in many RNAs including 5S, 16S and 23S rRNAs, group I and group II introns, bacterial RNase P, ribozyme of Tobacco ringspot virus satellite RNA (reviewed in Leontis and Westhof, 1998a) and lysine riboswitches (Grundy et al., 2003; Sudarsan et al., 2003) where it plays critical roles in RNA-RNA and RNA-protein interactions. While mutational analyses of cellular RNA motifs may be limited by the potential lethality of motif-disruptive mutations to an organism or by the high structural complexity of large cellular RNAs, such limitations do not apply to PSTVd, a host parasite. Therefore, further mutagenesis can be performed on PSTVd loop E, and other motifs, to determine their sequence signatures and gain detailed insights about RNA structure-function relationships that may not be feasible with a cellular RNA. Furthermore, the recurrent nature of this motif allows broad application of the knowledge obtained.

1.5.4 An integrative approach to address structure-function relationships of RNA motifs

Our findings from PSTVd loop E structure and function studies demonstrate that integration of Isostericity Matrix analysis, rationalized mutagenesis and systematic
functional studies can provide important insights into the structure-function relationship of an RNA motif and further into the elaborate cellular control over this relationship. Isostericity Matrix analysis can be used to predict base substitutions that will prevent formation of isosteric (or nearly isosteric) basepairs thereby disrupting the 3D structure of an RNA motif. If the structural integrity of the motif is critical for function, these mutations are likely to result in loss-of-function. This is well demonstrated by the replication defects of PSTVd carrying either the A99C or A261C mutation that disrupts loop E. Isostericity Matrix analysis can also predict which base substitutions can retain or restore the tertiary structure of a motif, providing a useful means to infer the tertiary structure of RNA motifs. This analysis alone, however, may not always predict a priori which isosteric or nearly isosteric base substitutions in a motif will retain or restore the function of a given RNA in the cellular environment, because the functioning of the motif requires interactions with other factors that can only be elucidated through biological assays. This is illustrated well by the results from the two compensatory mutants PSTVd\textsuperscript{Int}A261C/G98U and PSTVd\textsuperscript{Int}U260C/A99C. Both mutants are predicted by Isostericity Matrix analysis to have restored loop E tertiary structure. However, while the replication level of PSTVd\textsuperscript{Int}A261C/G98U is close to that of the wild type, the replication level of PSTVd\textsuperscript{Int}U260C/A99C is less than that of the wild type. Presumably, the loop E motif with the C259/C260/C99 base triple in the compensatory mutant PSTVd\textsuperscript{Int}U260C/A99C is sufficient, but not as optimal as the motif with C259/U260/A99 base triple as in the wild type, to be recognized by the cellular machinery for replication. These results therefore demonstrate that a combination of Isostericity Matrix analysis, genetic and biological experiments can provide a foundation to elucidate the elaborate RNA motif-cellular factor
interactions that regulate diverse biological processes and to address the fundamental question of how sophisticated features of RNA structures have evolved to achieve optimal functions.
<table>
<thead>
<tr>
<th>Loop E Basepair (PSTVd WT)</th>
<th>Mutations (Equivalent position in PSTVd)</th>
<th>Origin</th>
<th>Viroid</th>
<th>Predicted Loop E Structure a</th>
<th>Viability b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trans H/SE (A261/G98)</td>
<td>A261C1</td>
<td>EM</td>
<td>PSTVd</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>A261C/G98U1</td>
<td>EM</td>
<td>PSTVd</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Trans WC/H (U260/A99)</td>
<td>A99C1</td>
<td>EM</td>
<td>PSTVd</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>U260C/A99C1</td>
<td>EM</td>
<td>PSTVd</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Cis H/SE (U260/C259)</td>
<td>C259A2</td>
<td>SM</td>
<td>PSTVd</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>C259G1,3</td>
<td>EM</td>
<td>PSTVd</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>C259U1,2,4,5</td>
<td>NV/SM</td>
<td>PSTVd</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>C264U6 (259)</td>
<td>NV</td>
<td>CEVd</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Trans H/H (A258/A100)</td>
<td>A100G7</td>
<td>EM</td>
<td>PSTVd</td>
<td>+</td>
<td>+ c</td>
</tr>
<tr>
<td>Trans H/SE (A101/U257)</td>
<td>U257A2</td>
<td>SM</td>
<td>PSTVd</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>U257C1</td>
<td>EM</td>
<td>PSTVd</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>U257G3</td>
<td>EM</td>
<td>PSTVd</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>U258A6 (257)</td>
<td>NV</td>
<td>TCDVd</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>U262G6 (257)</td>
<td>NV</td>
<td>CEVd</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>U255G6 (257)</td>
<td>NV</td>
<td>TPMVd</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>U174G6 (257)</td>
<td>NV</td>
<td>CCCVd</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>U255G6 (257)</td>
<td>NV</td>
<td>TASVd</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

WC, Watson-Crick Edge
SE, Sugar Edge
H, Hoogsteen Edge
EM, Engineered Mutation
SM, Spontaneous Mutation
NV, Natural Variation
PSTVd, Potato spindle tuber viroid
MPVd, Mexican papita viroid
CEVd, Citrus exocortis viroid
TCDVd, Tomato chlorotic dwarf viroid
CCCVd, Coconut cadang cadang viroid
TPMVd, Tomato planta macho viroid
TASVd, Tomato apical stunt viroid

1 This study
2 Zhu et al., 2002
3 Qi and Ding, 2002
4 Owens et al., 1992
5 Wassenegger et al., 1996
6 Visvader and Symons, 1985
7 Semancik et al., 1993
8 Schrader et al., 2003
9 Singh et al., 2003

Table 1.1 Effects of nucleotide sequence variations on the predicted structure of loop Es and viability of viroids
<table>
<thead>
<tr>
<th>Injected RNA</th>
<th>Number of injected cells showing nuclear import/total number of injected cells (% of cells showing import)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSTVd&lt;sub&gt;Int&lt;/sub&gt;</td>
<td>14/16 (89)</td>
</tr>
<tr>
<td>PSTVd&lt;sub&gt;Int&lt;/sub&gt;A99C</td>
<td>9/9 (100)</td>
</tr>
<tr>
<td>PSTVd&lt;sub&gt;Int&lt;/sub&gt;A261C</td>
<td>10/12 (83)</td>
</tr>
<tr>
<td>Vector</td>
<td>0/15 (0)</td>
</tr>
</tbody>
</table>

**Table 1.2** Nuclear import of PSTVd RNAs
Figure 1.1 Secondary structure (A) and replication model (B) of PSTVd. (A) The rod-like secondary structure of PSTVd as well as the numbering of nucleotides are based on Gross et al. (1978). The five structural domains (Keese and Symons, 1985) include: 1) the left terminal (T_L, 1-46/315-359), 2) the pathogenicity (47-73/286-314), 3) the central (74-120/240-285), 4) the variable (121-148/212-239), and 5) the right terminal (T_R, 149-211). The loop E motif is in the central conserved region. (B) Asymmetric rolling circle replication of PSTVd (Branch et al., 1984). The incoming circular monomeric (+)-PSTVd serves as initial template to synthesize linear, concatemeric (-)-PSTVd. The latter functions as the replication intermediates to direct synthesis of concatemeric (+)-PSTVd, which is cleaved into unit-length monomers and further ligated into circular forms.
Figure 1.2 Geometric classification of RNA basepairing. The upper panel shows that each nucleotide base has three edges (Watson-Crick, Hoogsteen and Sugar) that can potentially hydrogen-bond with one of the three edges of another base. Thus, each base can be represented schematically using a triangle and can potentially pair with up to three other bases. The interacting bases can pair with cis or trans relative orientation of their glycosidic bonds. This is illustrated in the lower panel for the cis and trans orientations of nucleotides pairing at the Hoogsteen edge of one base and the sugar edge of the second base. In these basepairs, the Watson-Crick edges of the interacting bases are available for further interactions – with other RNAs, proteins, or small molecules. The cross and circle in the triangle where the Hoogsteen and sugar edges meet indicate 5’→3’ and 3’→5’ orientations, respectively, of the sugar-phosphate backbones relative to the plane of the page. W-C, Watson-Crick edge; H, Hoogsteen edge; SE, Sugar edge. [Adapted from (Leontis and Westhof, 2001) with permission from the RNA Society. This figure is prepared by Dr. Neocles Leontis].
Figure 1.2
Figure 1.3 Tertiary structural model of PSTVd loop E. (A) The paradigmatic sarcin/ricin motif based on X-ray crystal structures (adapted from (Leontis et al., 2002a)). (B) The inferred PSTVd loop E structural model. The dashed arrows indicate local changes in the strand orientation. All symbols to denote non-Watson-Crick basepairs and strand orientations are based on Leontis and Westhof (Leontis and Westhof, 2001). Circles, squares, and triangles indicate participation of Watson-Crick, Hoogsteen, and Sugar edges, respectively. Open symbols indicate basepairs with trans orientation of the glycosidic bonds and closed symbols indicate basepairs with cis orientations. (This figure is prepared by Dr. Neocles Leontis.)
Figure 1.4 The five non-Watson-Crick basepairs for the core of PSTVd loop E motif. Examples of natural variants or experimentally designed mutations that can form isosteric basepairs are also illustrated. Y – pyrimidine (U or C), W – water molecule. The ribose is represented by a closed circle except when it is involved in hydrogen-bonding. (This figure is prepared by Dr. Neocles Leontis.)
Figure 1.4
Figure 1.5 The functional status of viroid mutants and natural variants mapped onto
the respective Isostericity Matrices for each basepair, according to the geometric
family to which it belongs (Leontis et al., 2002b). I1-I5 denote subfamilies within each
of them the basepairs are isosteric (see MATERIALS AND METHODS for the definition
of isostericity and reference (Leontis et al., 2002b) for details on the classification of
subfamilies). Each blank and red box indicates that non-Watson-Crick basepairing
between the two nucleotides in the family has not been reported in X-ray crystal structures
or cannot be formed based on isosteric predictions (Leontis et al., 2002b). *Chrysanthemum
stunt viroid* (CSVd) has identical sequences with PSTVd in loop E. For sources of data and
abbreviations of all other viroids see lists in Table 1.1. (This figure is prepared by Dr.
Neocles Leontis.)
Figure 1.5

### Notes:
- Includes spontaneous (SM) and engineered (EM) mutations shown in Table 1
- Includes EM shown in Table 1

The nucleotide numbering is for PSTVd. For the other viroids, each base pair occurs at position equivalent to that of PSTVd base pair.
Figure 1.6 Analysis of loop E formation by UV-crosslinking. Gel purified substrate RNAs were pre-incubated under conditions that favor loop E formation and subjected for UV-crosslinking as described in MATERIALS AND METHODS. (A) Diagram showing structural conformations of the substrate and crosslinked RNAs in their native state and in denaturing gels. G88 and G87 indicate the 5'- and 3'-ends, respectively, of the folded, linear substrate RNA. The established crosslinking site (U260/G98) is marked in red. (B) Autoradiography of gel blots showing presence or absence of UV-crosslinked products from the substrate RNAs as indicated. The efficiency of crosslinking is presented as the percentage of substrate RNAs that is converted into the crosslinked products. Note that UV treatment of the two compensatory mutants PSTVd\textsuperscript{int}A261C/G98U and PSTVd\textsuperscript{int}A99C/U260C, which are predicted by Isostericity Matrix analysis to restore loop E tertiary structure and showed restored replication, resulted in little crosslinked products. This indicates that G98/U260 base stacking, in addition to proper loop E tertiary structure, is required for specific UV crosslinking. The data from these two mutants thus provide evidence for UV crosslinking occurring specifically in loop E. GFP, green fluorescent protein RNA.
Figure 1.7 Northern blots showing accumulation of PSTVd RNAs in protoplasts of *N. benthamiana*. The rRNA serves as loading controls. c-PSTVd and l-PSTVd indicate monomeric circular and monomeric linear PSTVd RNAs, respectively.
Figure 1.8 Nuclear import of fluorescently labeled *in vitro* transcripts of PSTVd. The transcripts of wild type (Int) and the two mutants (A99C and A261C) appear in the nucleus within 30 min of injection into the cytoplasm (upper row). The vector transcripts do not show nuclear accumulation. DAPI staining (lower row) shows the position of nucleus for each injected cell.
Figure 1.9 Circularization of loop E mutants in *N. benthamiana* nuclear extract (NbNE). Gel-purified and radiolabeled unit-length linear (+)-PSTVd transcripts were incubated in NbNE for 2 hr and analyzed by urea-PAGE. c-PSTVd and l-PSTVd indicate monomeric circular and monomeric linear PSTVd RNAs, respectively. GFP, green fluorescent protein RNA.
Figure 1.10 Determination of (-)-PSTVd (A) and (+)-PSTVd (B) RNA levels in infected *N. benthamiana* protoplasts by quantitative real-time RT-PCR. For details see MATERIALS AND METHODS. M, Mock inoculation. To facilitate comparison, the accumulation level of wild type (Int) is arbitrarily set to a value of 1 and the levels of the mutants are presented as relative values on a log scale. (This figure is prepared by Shuiming Qian.)
Figure 1.10

A Accumulation of (-)-PSTVd RNA

B Accumulation of (+)-PSTVd RNA
CHAPTER 2

TERNIARY STRUCTURE AND FUNCTION OF AN RNA MOTIF REQUIRED FOR PLANT VASCULAR ENTRY TO INITIATE SYSTEMIC TRAFFICKING

Zhong X, Tao X, Stombaugh J, Leontis N, Ding B. (2007) Tertiary structure and function of an RNA motif required for plant vascular entry to initiate systemic trafficking. EMBO J 26: 3836-3846. (This paper is used here with permission from Nature Publishing Group.)
2.1 ABSTRACT

Vascular entry is a decisive step for the initiation of long distance movement of infectious and endogenous RNAs, silencing signals, and developmental/defense signals in plants. However, the mechanisms remain poorly understood. We used *Potato spindle tuber viroid* (PSTVd) as a model to investigate the direct role of the RNA itself in vascular entry. We report here the identification of an RNA motif that is required for PSTVd to traffic from nonvascular into the vascular tissue phloem to initiate systemic infection. This motif consists of nucleotides U/C that form a water-inserted cis Watson-Crick/Watson-Crick basepair flanked by short helices that comprise canonical Watson-Crick/Watson-Crick basepairs. This tertiary structural model was inferred by comparison with X-ray crystal structures of similar motifs in rRNAs and is supported by combined mutagenesis and covariation analyses. Hydration pattern analysis suggests that water insertion induces a widened minor groove conducive to protein and/or RNA interactions. Our model and approaches have broad implications to investigate the RNA structural motifs in other RNAs for vascular entry and to study the basic principles of RNA structure-function relationships.

2.2 INTRODUCTION

Increasing evidence indicates that intercellular trafficking of RNAs plays important roles in physiological and developmental processes in plants (Ruiz-Medrano et al., 1999; Kim et al., 2001; Haywood et al., 2005; Banerjee et al., 2006; Lough and Lucas, 2006). Furthermore, viroid and viral RNAs apparently utilize the endogenous trafficking system
to spread within a plant to establish systemic infection (Flores et al., 2005; Lucas, 2006; Ding and Itaya, 2007b). Gene silencing signals, which likely have an RNA component, traffic intercellularly to trigger systemic silencing in plants (Palauqui et al., 1997; Voinnet and Baulcombe, 1997; Voinnet et al., 1998; Yoo et al., 2004), Caenorhabditis elegans (Fire et al., 1998) and Tribolium (Coleoptera) (Bucher et al., 2002). These observations suggest that intercellular trafficking of RNAs plays critical roles in many aspects of biology. Elucidating the underlying mechanisms is crucial to understand gene regulation as well as host-pathogen interactions at the organismal level.

Plasmodesmata and the vascular tissue phloem form continuous cytoplasmic channels for direct cell-to-cell and systemic transport of RNAs and proteins as well as photoassimilates in plants (Lucas et al., 1993, Lucas et al., 2001; Lough and Lucas, 2006). Some plant proteins bind RNAs and facilitate their intercellular trafficking (Xoconostle-Cazares et al., 1999; Yoo et al., 2004). Viral proteins play an important role in viral cell-to-cell movement (Lucas, 2006). There is also evidence suggesting that some viral RNAs have a role in systemic infection (Ding et al., 2005; Lough et al., 2006). In general, the specific role of an RNA as well as the mechanisms of protein-RNA interactions in trafficking remain poorly understood. In particular, there is little knowledge of how an RNA or protein enters the phloem, which is a decisive step to initiate systemic trafficking (Ding, 1998; Lucas et al., 2001; Haywood et al., 2002; Lough and Lucas, 2006).

Viroids provide simple models to investigate the direct role of RNA motifs in systemic RNA trafficking. These single-stranded, circular and noncoding RNAs are the smallest pathogens known to date, with sizes ranging from 250 to 400 nucleotides (Flores et al., 2005; Ding and Itaya, 2007b). We utilize Potato spindle tuber viroid (PSTVd) as our
experimental system. This viroid contains 359 nucleotides and forms a rod-like secondary structure in its native in vitro state (Fig. 2.1A; Gross et al., 1978). There is strong evidence that this structure exists and functions in vivo (Wassenegger et al., 1994; Zhong et al., 2006; Wang et al., 2007; Eiras et al., 2007). PSTVd replicates in the nucleus via a rolling-circle mechanism, which starts with transcription of the incoming monomeric, circular (+)-strand RNA into multimeric, linear (-)-strands. The latter then serve as the replication intermediates for the production of multimeric, linear (+)-strands that are finally processed into unit-length, circular RNAs (Branch et al., 1984). To establish systemic infection, PSTVd traffics across many different cellular boundaries from the epidermis to the phloem to spread throughout the infected plant (Fig. 2.1B; Ding et al., 2005; Ding and Itaya, 2007b). Because PSTVd does not encode proteins, it must have evolved sequence/structural features that are directly recognized by the endogenous cellular machinery to accomplish replication and systemic trafficking. Therefore, investigating PSTVd structural motifs for trafficking can help illuminate the general principles of molecular interactions that control systemic RNA trafficking. Using PSTVd as a model, Qi et al. (2004a) presented genetic evidence for the role of an RNA motif in mediating RNA trafficking across a specific cellular boundary in the nonvascular tissue. However, the tertiary structure and functional mechanism of this motif remain to be elucidated.

Here we report the first loss-of-function genetic identification of an RNA tertiary structural motif in the PSTVd required for vascular entry, specifically from the bundle sheath into the phloem regardless of the developmental stage of an infected leaf. Our results support the hypothesis that tissue-specific factors interact with distinct RNA motifs
to control trafficking across specific cellular boundaries. Furthermore, our approach should have broad implications to investigate the structural motifs in endogenous and infectious RNAs critical for vascular entry and to study the basic principles of RNA structure-function relationships.

2.3 MATERIALS AND METHODS

2.3.1 Plant materials and growth condition

**Nicotiana benthamiana** plants were grown in a growth chamber controlled at 14 hour light (27 °C)/10 hour dark (24 °C) cycles. *N. benthamiana* cells were cultured in Murashige and Skoog medium (MS salts; Life Technologies, Rockville, MD). Cells were maintained and subcultured as previously described in detail by Zhong et al. (2005).

2.3.2 PSTVd cDNA construction

Plasmid pRZ6-2 containing cDNAs of PSTVd\textsuperscript{Int} was constructed by Hu et al. (1997) and was a gift from Dr. Robert Owens. All PSTVd-derived mutants were generated by site-directed mutagenesis using the Quickchange Site-Directed Mutagenesis kit (Stratagene, La Jolla, CA) using pRZ:PSTVd\textsuperscript{Int} as the template. The introduced mutations were verified by sequencing.

Construction of pInter(−) and pInter(+) was described in Qi and Ding (2002). SpeI-linearized pInter(−) and pInter(+) were used as the templates to generate riboprobes specific for (+) and (−)-PSTVd RNAs, respectively.
2.3.3 Plant and protoplast infection

The in vitro transcripts of PSTVd variants were used to inoculate the carborundum-dusted first two true leaves of two week-old *N. benthamiana* plants (300 ng/plant). DEPC-H$_2$O was used for mock inoculation. *N. benthamiana* protoplasts were prepared and transfected with PSTVd transcripts by electroporation as described by Zhong et al. (2006). At three days post inoculation, transfected protoplasts were collected for RNA extraction and gel blot analysis.

2.3.4 In vitro transcription

To prepare riboprobes for RNA gel blots or in situ hybridization, [$\alpha$-$^{32}$P]- or digoxigenin-UTP-labeled antisense riboprobes were prepared by in vitro transcription using T7 Maxiscript kit (Ambion, Austin, Texas, USA) following the methods recommended by the manufacturer, using *Spe*I-linearized pInter(−) or pInter(+) as the templates, respectively. After in vitro transcription, the DNA templates were removed by digestion with RNase-free DNase I. The RNA transcripts were purified with MEGAClear kit (Ambion) and were quantified by UV spectrometry or scintillation counting, respectively.

2.3.5 Tissue processing, in situ hybridization

Samples from infected plants were processed to obtain paraffin sections for in situ hybridization, as described in Zhu et al. (2001). Sections (8-10 μm) were obtained with a rotary microtome (HM340E, Microm International GmbH, Walldorf, Germany). *In situ*
hybridization was performed as described previously (Qi et al., 2004a), using
digoxigenin-labeled antisense PSTVd riboprobes.

2.3.6 RNA extraction and RNA gel blot

Total RNAs from infected plants were isolated using Trizol reagent (Invitrogen,
Carlsbad, California, USA) and total RNAs from protoplasts were extracted using
RNeasy plant mini kit (Qiagen, Valencia, California, USA) according to the manufacture’s
instructions. RNA gel blots were performed as described in Zhong et al. (2006).

2.3.7 RT-PCR

Total RNAs from N. benthamiana petioles were isolated using Trizol reagent
(Invitrogen) and treated with TURBO DNA-free™ kit (Ambion) according to the
manufacture’s instructions. One µg of total RNA was reverse-transcribed using the
ThermoScript™ RT-PCR system (Invitrogen) with a reverse primer
(5’-AGGAACCACTGCGGTCCA-3’) for (+)-PSTVd and 1093-1114R
(5’-CCCGGAACCCAAAAACTTTTG-3’) for 18S rRNA. The reverse transcribed PSTVd
cDNA was amplified by PCR with a forward primer
(5’-CGGAACCTAAGCTCGGTCTCCT-3’) and the reverse primer. The primers
1093-1114R and 1000-1020F (5’-GATCAGATAACGTCTAGTC-3’) were used for
amplification of 18S rRNA cDNA template, which served as an internal control. The PCR
products were run on 1.2% agarose gel for EB staining or 5% polyacrylamide/8 M urea gel
for hybridization analyses using a riboprobe specific for PSTVd.
2.3.8 Sequencing of RNA progeny

The protocols for preparing cDNAs of the PSTVd progeny isolated from protoplasts or plants were performed essentially as described by Qi and Ding (2002). Briefly, cDNAs of PSTVd RNA were RT-PCR amplified and sequenced in both directions using the ABI377 DNA sequencer (Perkin-Elmer, Boston, MA) at the DNA Sequencing Facility at Ohio State University.

2.4 RESULTS

2.4.1 PSTVd infection of Nicotiana benthamiana as an experimental system

We used infection of *N. benthamiana* by the intermediate strain of PSTVd (WT) as our experimental system. PSTVd replicates efficiently in cultured cells of *N. benthamiana* (Qi and Ding, 2002) and readily establishes systemic infection in the plant (Hu et al., 1997; Zhu et al., 2001), making it possible to investigate separately structural motifs involved in the replication and trafficking. To establish the system, we used RNA blots to analyze the accumulation of the (+)-circular and linear viroid RNAs in the youngest systemic leaves at 0, 6, 9, 12, 15, 18, 21, 24 and 27 days post inoculation (dpi). As shown in Fig. 2.2A and B, accumulation of the (+)-circular PSTVd in systemic leaves was first visible at 12 dpi. The level reached a maximum around 21 dpi and dropped slightly afterwards. This time course set the basis for the subsequent trafficking experiments.
2.4.2 Identification of PSTVd mutants defective in systemic trafficking

Numerous PSTVd mutations have minimal or drastic effects on infection in tomato (Hammond and Owens, 1987; Owens et al., 1991; Owens et al., 1995; Owens et al., 1996; Loss et al., 1991; Qu et al., 1993; Hu et al., 1996). We developed a simple and fast assay to screen for PSTVd mutants that are defective in systemic trafficking. We generated many of these mutants and tested their systemic infection in *N. benthamiana*. Several mutants showed accumulation in inoculated leaves but not in systemic leaves (X. Zhong and B. Ding, unpublished data). One such mutant contains C318 to A substitution (C318A) that closes the original U43/C318 “loop” by allowing the formation of a canonical cis Watson-Crick/Watson-Crick U43/A318 basepair, without altering the global secondary structure of the viroid RNA, as predicted by mfold (Zuker, 2003; Fig. 2.1C; 2.3). As shown in Fig. 2.4A, the mutant C318A did not accumulate in systemic leaves at 21 dpi. However, it accumulated in the inoculated leaves (Fig. 2.4B). To further determine its replication capacity, we performed replication assays in protoplasts prepared from cultured cells of *N. benthamiana*. As shown in Fig. 2.4C, mutant C318A accumulated in the protoplasts at 3 dpi. Sequencing confirmed maintenance of the mutant sequence in the RNA progeny in the infected protoplasts and leaves. These data indicate that mutant C318A is defective in systemic trafficking rather than replication. Further analyses showed that the mutant did not show accumulation in systemic leaves up to 7 weeks post-inoculation (data not shown), indicating that systemic trafficking of this mutant is abolished rather than delayed.

To further test whether it is the U43/C318 “loop,” or the requirement of C at position 318, that is critical for systemic trafficking, we substituted U43 with G to close the “loop” (Fig. 2.1C) and tested its effect on systemic infection. As shown in Fig. 2.4A and B,
mutant U43G was absent from systemic leaves but accumulated in inoculated leaves. U43G also replicated in the protoplasts, albeit at reduced levels compared to the WT (Fig. 2.4C). As shown below, this level of reduced accumulation was not the cause for defects in systemic infection. The data provide additional evidence that the U43/C318 “loop” is critical for systemic trafficking of PSTVd in *N. benthamiana*.

### 2.4.3 Mutants C318A and U43G were capable of trafficking between nonvascular cells

To establish systemic infection, PSTVd inoculated into leaf epidermal cells must be capable of replication and movement into mesophyll, bundle sheath and phloem, followed by long distance movement through the vascular system and finally exit from the phloem to invade new tissues (Fig. 2.1B). Failure of trafficking at any of these steps will block systemic infection. To identify the cellular interface(s) at which trafficking is blocked for mutants C318A and U43G, we performed in situ hybridization experiments to visualize their cellular localization in inoculated leaves. Time points of 6, 12 and 18 dpi were chosen because the accumulation of the WT viroid RNA first became visible in systemic leaves at 12 dpi (see Fig. 2.2).

The results from 12 dpi samples are shown in Fig. 2.5 and those from 6 and 18 dpi are shown in Figures 2.6 and 2.7. At 12 dpi, we could detect the accumulation of mutants C318A and U43G as well as the WT in the mesophyll cells (Fig. 2.5A) and in the epidermis (data not shown). Often clusters of cells containing hybridization signals were detected. Identical results were obtained from 6 and 18 dpi samples (Fig. 2.6 and 2.7). These results
indicate that the mutant viroid RNAs were capable of trafficking between epidermal and mesophyll cells.

2.4.4 Mutants C318A and U43G failed to traffic from bundle sheath into phloem

As shown in Figures 2.5B, 2.6 and 2.7, we could also detect the presence of mutants C318A and U43G as well as the WT in the bundle sheath. Strikingly, while the WT viroid RNA was further detected in the phloem cells, the two mutants were never detected there. The same observations applied to leaf samples collected at 6 and 18 dpi, indicating that blockage of vascular entry for the mutants was maintained during leaf growth.

It is possible that mutants C318A and U43G did enter the phloem and trafficked along the phloem, but at a level below the detection limit of in situ hybridization. To test this possibility, we used the highly sensitive reverse transcription polymerase chain reaction (RT-PCR) to detect the presence of mutants in the petioles of inoculated leaves at 12 dpi. As shown in Fig. 2.5C, no signal was detected in any of the two plants infected by C318A or by U43G, whereas bands corresponding to the full-length PSTVd sequence could be detected in WT infection. Amplification of the 18S rRNA served as a control for the quality of RNA in each sample. Hybridization analysis with PSTVd-specific probes confirmed that the RT-PCR products indeed were PSTVd sequences (Fig. 2.5D). Three repeated experiments yielded identical results.

Taken together, all data provided compelling evidence that mutants C318A and U43G had specific defects in entering the phloem from bundle sheath in N. benthamiana plants.
### 2.4.5 Predicted tertiary structure of U43/C318 motif

U43 and C318 are generally drawn unpaired in the secondary structure of PSTVd forming a small 1x1 unpaired “loop” flanked by cis Watson-Crick/Watson-Crick (cWW) basepairs as shown in Fig. 2.1A and C. (In PSTVd, one of the flanking pairs is a G/U which can form a cWW basepair nearly isosteric to cWW A/U or G/C, while in related viroids this pair is a G/C or A/U as shown in Figure 2.12.) However, as shown by X-ray crystallography and nuclear magnetic resonance spectroscopy, most RNA “loops” in structured, biologically active RNA molecules form distinct, well-structured three-dimensional (3D) motifs in which most bases participate either in modified WW or non-WW basepairs and/or stacking interactions (Leontis et al., 2002a). Many RNA 3D motifs are recurrent – they are observed in different RNA molecules (Leontis and westhof, 2003; Leontis et al., 2006). Recurrent 3D motifs comprise sets of nucleotides with similar spatial arrangements and their 3D structures are generally more conserved than their sequences (Lescoute et al., 2005). On the basis of these considerations, we employed the 3D RNA motif search program, “Find RNA 3D” or FR3D (Sarver et al., 2007), to search exhaustively the RNA 3D database for possible motifs formed by U/C base juxtapositions embedded in RNA helices so as to infer plausible tertiary structures for the PSTVd U/C motif.

The FR3D search revealed that in all cases the U and C bases were inserted in the helix with their Watson-Crick edges facing each other and forming one or more hydrogen-bonds (H-bonds). Moreover, for most occurrences in the highest resolution X-ray crystal structures (<2.6 Å), the U and C bases form water-inserted cis WW (cWW) basepairs as shown in figures 2.8 and 2.10. In this type of basepair, a direct H-bond occurs...
between U(O4) and C(N4) and an inserted water molecule forms bridging H-bonds to U(N3) and C(N3). The insertion of the water molecule opens the basepair toward the minor (shallow) groove of the RNA helix and significantly increases the C1’-C1’ distance from the normal ~10.6Å of a canonical A/U or C/G cWW basepair to ~11.8Å. The angle subtended by the glycosidic bonds of the U and C bases also increases. This causes a distortion of the helix, widening the minor groove locally as shown in figure 2.8 when this helix is superimposed on a helix comprising exclusively canonical cWW basepairs. The latter was identified by a structure search using FR3D in which the immediately flanking base-pairs were matched to those of the U/C containing helix. Based on these search results, we propose that U43 and C318 in the PSTVd “loop” also form a cWW basepair with an inserted water molecule. The following analyses provide further evidence in support of this model.

2.4.6 Mutational analyses showed that local distortion of helix was critical for trafficking

To test whether the proposed local helix distortion is critical for trafficking, we generated all 16 possible nucleotide substitutions at positions 43 and 318 of PSTVd and tested their trafficking functions. We also carried out geometric structure searches using FR3D to identify exemplars in the 3D database for each mutant.

The trafficking functions of all mutants are displayed in the 4x4 matrix in figure 2.9A, based on their accumulation in systemic leaves as shown by RNA blots (Fig. 2.9B). The green and red background colors denote basepairs that retained and lost systemic
trafficking functions, respectively. Sequencing confirmed maintenance of the mutant sequences in the progeny. The first striking observation is that all nucleotide changes that allow formation of canonical cWW basepairs (AU, UA, GC, and CG) abolished systemic accumulation. It should be noted that A43/U318 and C43/G318 did not show replication in protoplasts (Lanes 4 and 15, Fig 2.9C), so the current assay cannot determine whether they were truly defective in the trafficking function itself.

The G/G juxtaposition cannot form any type of cWW basepair. Consistent with this, the G43/G318 mutant is not viable in infection.

The C43/U318 substitution is viable in trafficking. It is expected to form a basepair that is very similar to U43/C318 (WT) due to symmetry as is evident in comparing the C/U and U/C panels in figure 2.10. Sequence analysis of conserved C/U water-inserted basepairs in the rRNAs also revealed cases of C/U covarying with U/C, as discussed below.

The C43/C318 and U43/U318 substitutions are viable in trafficking. We used FR3D to search for C/C and U/U basepairs geometrically similar to the water-inserted U/C pair. The search revealed a similar C/C basepair also occurring between two canonical cWW basepairs in the structure of an siRNA duplex bound to Carnation Italian ringspot virus P19 protein (PDB file: 1rpu) (Vargason et al., 2003). The protein-binding site spans the minor groove of the C/C basepair. The search also revealed a water-inserted cWW U/U basepair between two canonical cWW basepairs in the 16S rRNA A-site bound to Lividomycin A (PDB File: 2esj) (Francois et al., 2005). In this basepair, a direct H-bond occurs between U5(O4) and U42(N3). An inserted water molecule forms bridging H-bonds to U5(N3) and U42(O2), opening the basepair toward the minor groove and
increases the C1’-C1’ distance to 10.7Å. Lividomycin binds in the major groove of the motif that includes the U/U basepair. Similar C/C and U/U basepairs occur in other structures and frequently have bound water molecules as shown in figure 2.10. These pairs assume an asymmetric bifurcated geometry that also locally distort the helix, increasing the angle subtended by the glycosidic bonds of the bases and widening the minor groove with C1’-C1’ distances of ~10.6Å, longer than the ~8.5Å distance observed for “wobble”-type cWW C/C or U/U basepairs.

G43/A318 is functional and can form a cWW basepair with a C1’-C1’ distance of 12.7Å that also locally distorts the helix, as shown by the exemplar in figure 2.10 (G150/A161 from PDB file 1hq1) (Batey et al., 2001). In contrast, the A43/G318 substitution does not show systemic accumulation. Because it fails to replicate (Fig. 2.9C), its trafficking function cannot be determined. The A/G and G/A pairs are not completely symmetric structurally because A/G cWW basepairs tend to be propeller twisted and consequently the G(N2) amino group is not in the same place when the A and G are reversed (Sponer et al., 2003). The G(N2) functional group is most often observed to interact with protein in the minor groove. In the absence of an inserted water molecule available to interact with protein in the minor groove, the G(N2) functional group may play that role. Therefore, this additional level of interaction with different environments may render these two pairs functionally different (see further discussions below).

A43/A318 is viable for trafficking. Examples of A/A are observed forming cWW basepairs with a C1’-C1’ distance of 13.3Å as shown in figure 2.10 (exemplar A675/A715 from 1j5e) (Wimberly et al., 2000). Thus A/A also distorts the helix by association with a water molecule in the minor groove.
G318/U43 is viable for trafficking. It is predicted to form a wobble cWW basepair as shown in figure 2.10 (exemplar G226/U137 from 2avy) (Schuwirth et al., 2005) with a C1’-C1’ distance of 10.5Å. G/U wobble pairs strongly bind a water molecule in the minor groove and present the G(N2) functional group for possible interactions (Mokdad and Leontis, 2006). G43/U318 is not viable likely for reasons similar to those for A43/G318.

A43/C318 and C43/A318 are both viable for trafficking. They are structurally similar to G318/U43 with C1’-C1’ distances of 10.3Å and 10.9Å as shown in figure 2.10 (exemplar A5/C24 from 1hq1 and C63/A20 from 1m5p) (Wild et al., 1999; Barton et al., 2002), and all can form a pocket that allows for the insertion of water molecules to distort the helix.

The pattern that emerges from these analyses is that the mutations that did not disrupt the systemic trafficking function are all predicted to form modified cWW basepairs, some with inserted water molecules, that distort the helix and widen the minor groove, perhaps to facilitate interaction with a protein factor. In contrast, a regular helix without such distortion causes defects in systemic trafficking and possibly also replication for some basepairs.

2.4.7 Covariation analyses support the tertiary structural model

To further test our model, we performed two types of nucleotide covariation analyses, based on the rationale that when two nucleotides interact to form a particular structure, mutation in one nucleotide may result in covariation in its interacting nucleotide to maintain the same or similar structure and thus function (Pace et al., 1999). First, we examined the equivalent positions of other viroid sequences in the genus Pospiviroid that
are predicted to fold into similar secondary structures as PSTVd. All of these viroids have the conserved loop E-like motif (Zhong et al., 2006). As shown in figure 2.11 and 2.12, the U/C motif is conserved for *Tomato planta macho viroid* and *Tomato chlorotic dwarf viroid*, and is replaced by variation C/C in *Citrus exocortic viroid*, *Tomato apical stunt viroid* and *Chrysanthemum stunt viroid* and by variation A/A in *Mexican papita viroid* (Singh et al., 2003); Subviral RNA Database, [http://subviral.med.uottawa.ca/cgi-bin/home.cgi](http://subviral.med.uottawa.ca/cgi-bin/home.cgi). As discussed above, both A/A and C/C base juxtapositions in this context can form cWW basepairs with associated or inserted water molecules to distort the local helix (Fig. 2.10) and are viable for trafficking (Fig. 2.9).

We also performed covariation analyses of U/C basepairs in the structures of the rRNAs by using the Ribostral program for covariation analysis (Mokdad and Leontis, 2006). First, FR3D identified several U/C cWW basepairs in the 16S or 23S rRNA structures of *Escherichia coli* (Schuwirth et al., 2005) and *Thermus thermophilus* (Selmer et al., 2006), and in the 23S structure of *Haloarcula marismortui* (Harms et al., 2001). We focused our attention on those U/C pairs which are flanked on both sides by canonical cWW basepairs (including G/U wobble) and which are conserved in two or more homologous structures. As shown in figure 2.13, the three examples meeting these criteria are 1) U576/C565 in *E. coli* and *T. thermophilus* 23S (PDB files 2aw4 and 2j01); 2) U807/C673 in *E. coli* and *T. thermophilus* 23S and U900/C764 in *H. marismortui* 23S (PDB file 1s72); and 3) U2511/C2575 in *E. coli* and *T. thermophilus* 23S, which covary with U2546/U2610 in *H. marismortui* 23S. Interestingly, the first and third examples interact, at the widened minor groove, with the universally conserved proteins L3 and L4 which are crucial for 23S rRNA folding in the 50S ribosomal subunit (Klein et al., 2004).
Second, sequence covariations were observed at equivalent positions for these U/C pairs in 23S sequence alignments for bacteria (~800 sequences) and archaea (25 sequences) (Lescoue et al., 2005). As shown in figure 2.13, for U576/C565 the primary covariation is C/U while for U2511/C2575 it is U/U. Interestingly, U/U is observed at the corresponding position of the H. marismonrtui 23S structure and in the other archaeal sequences. All covariations of the U/C pair (i.e., C/U, U/U and U/G) in 23S are also functional in PSTVd (Fig. 2.10), therefore supporting the structural model of the U/C pair.

2.4.8 Hydration patterns suggest a functional mechanism

For the reasons described above, we propose that the U/C basepair in the PSTVd forms a water-inserted cWW basepair with a widened minor groove conducive to protein and/or RNA interactions. To test this model, we examined the hydration patterns for all base juxtapositions that can form cWW or water-inserted cWW basepairs using the Solvation Web Service (SWS) (Auffinger and Hashem, 2007) in combination with geometric structure searches using FR3D. SWS provides solvation data for a given basepair type compiled from the entire structure database. SWS displays the solvation density surrounding each basepair type so that likely locations of water molecules can be inferred. In figure 2.10 the dark blue spheres represent high density and the light blue spheres medium density solvation sites reported by SWS. For the pairs for which SWS did not yield such information (A/A, C/C, U/C, C/U), we employed FR3D to identify all examples in high-resolution X-ray crystal structures containing water molecules. For these pairs, the dark blue spheres indicate the sites of water molecules most often observed. Similar to figure 2.9A, the green and red background colors in figure 2.10 indicate base
substitutions that are functional and defective in PSTVd trafficking, respectively. A clear pattern is that most functional base substitutions bind a water molecule using one or more Watson-Crick edges, resulting in opening of the basepair towards the minor groove. The only exception is A/G which has the G(N2) amino group occupying an equivalent position as an inserted water molecule. The G(N2) amino group is the most common site of protein binding in the minor groove (Klein et al., 2004), and examples involving A/G cWW pairs have been observed (Sponer et al., 2003).

2.5 DISCUSSION

In this paper we reported the first loss-of-function genetic identification of an RNA tertiary structural motif that is required for RNA trafficking from the bundle sheath into the phloem to initiate long-distance trafficking. This motif is not required for trafficking from epidermis to mesophyll and then to bundle sheath. Structurally, this motif comprises juxtaposed nucleotides U/C that form a water-inserted cWW basepair in a helical context, i.e. flanked by cWW A/U, G/C or G/U basepairs. The fundamental feature of the motif is the local distortion of the helix caused by the U/C pair. This motif likely represents a protein-binding site, based on the known function of similar conserved motifs in other RNAs that interact with specific proteins. It is interesting that the left basepair C/G flanking the U/C motif in PSTVd, TPMVd and TCDVd changes to G/C when this motif is naturally replaced by C/C in CEVd, TASVd and CSVd. Whether these flanking sequences only support formation of a motif and/or whether they have functional significance in other capacities remains to be determined.
Although mutant U43G accumulates to lower levels than the wild type PSTVd, our analyses of other mutants indicate that the viroid replication/accumulation levels are not a factor for cell-to-cell trafficking including vascular entry. For instance, the double mutants C43/U318 and G43/A318 (lanes 9 and 10, respectively, Fig. 2.9C) accumulate to lower levels than the wild type (lane 1 in Fig. 2.9C) in infected protoplasts, but they are competent in systemic trafficking (lines 9 and 10, respectively, in Fig. 2.9B). Thus, the failure of mutants C318A and U43G, as well as other mutants, to traffic into the phloem most likely results from disrupted molecular interactions with a cellular factor(s) in the trafficking machinery.

The absence of mutants C318A and U43G in the phloem is unlikely attributed to their instability in the phloem. First, the secondary structure of PSTVd is resistant to RISC (RNA Induced Silencing Complex)-mediated degradation (Wang et al., 2004; Itaya et al., 2007). C318A or U43G substitution causes only a mild change in the local but not the overall secondary structure of PSTVd. Therefore, it is unlikely that they will alter dramatically the stability of PSTVd. In fact, Owens et al, (1996) showed that U43G substitution enhances the thermostability of PSTVd. Second, these mutants are stable and replicate in the protoplasts as well as epidermal, mesophyll and bundle sheath cells. There is no evidence yet that a special RNA degradation system exists in the phloem.

We previously showed that a bipartite motif in PSTVd is required for trafficking from bundle sheath to mesophyll, but is not required for trafficking from mesophyll to bundle sheath, in young tobacco leaves (Qi et al., 2004a). We attempted to determine whether the U/C motif is required for trafficking from the phloem to bundle sheath by generating transgenic *N. benthamiana* plants expressing the cDNAs of C318A and U43G.
mutants under the control of a companion cell-specific promoter, thereby allowing primary transcripts derived from the promoter activity would serve as templates to initiate RNA-RNA replication (Qi et al., 2004a). However, in all 9 lines of transgenic plants viroid progeny showed reversion to wild type sequences or mutations elsewhere (data not shown). Therefore, whether U/C or another motif is required for phloem exit remains to be determined. It also remains to be determined whether the U/C motif alone is sufficient or requires participation of another motif to mediate phloem entry.

It is unclear why certain nucleotide substitutions at U/C positions that distort the helix resulted in abolishment of replication. There are several possibilities. First, the U/C motif also plays a role in replication, besides phloem entry. The specific nucleotide substitutions, while maintaining similar structures as other replication-competent mutants, inhibit interactions with a cellular factor(s) important for replication. There is precedence that a single viroid motif plays multiple roles and nucleotide identity is also important. For instance, the PSTVd loop E motif has a role in replication, host adaptation, in vitro processing and symptom expression (Ding and Itaya, 2007b). Second, these nucleotide substitutions create a novel binding site for some cellular factors. Such binding results in mis-localization of the viroid RNAs within a cell or prevents access of the replication machinery. Resolving these issues should lead to new insights into the molecular interactions between an RNA motif and cellular factors.

The U/C motif may function in a host-specific manner. Owens et al (1996) showed that C318A did not affect systemic infection in tomato, consistent with our own observations (data not shown). On the other hand, U43G substitution affected systemic infection to some extent in tomato (Owens et al., 1996). Thus, vascular entry of PSTVd in
tomato may involve an alternative mechanism. This is not surprising, considering the narrow host range of viroids and the observations that single nucleotide changes can alter host range (Wassenegger et al., 1996; Zhu et al., 2002) and host-dependent symptom expression (Qi and Ding, 2003). Consistent with our current study, Loss et al (1991) showed that C318G substitution did not affect systemic infection in tomato. It should be noted that a metastable hairpin II (HPII) structure has been postulated to play a critical role in PSTVd replication (Loss et al., 1991; Owens et al., 1991) and that U43G substitution presumably inhibits the formation of HPII (Owens et al., 1996). The normal replication of U43G in *N. benthamiana* raises the question whether this putative HPII plays a significant role in replication in this plant. This putative structure unlikely has a role in trafficking, because C318A substitution has little effect on the formation of this putative structure.

Our data, together with previous work identifying a PSTVd motif for trafficking from bundle sheath to mesophyll (Qi et al., 2004a) further support the hypothesis that the bundle sheath plays a critical role in regulating macromolecular trafficking between the phloem and nonvascular tissues. An emerging picture is that trafficking from the same cell into different neighboring cells involves different mechanisms. For instance, SUT1 mRNA is transported only from companion cells into sieve elements (Kuhn et al., 1997). The identification of a viroid RNA motif for trafficking from the bundle sheath into phloem (This Study) and a separate motif for trafficking from bundle sheath to mesophyll (Qi et al., 2004a) provides the most compelling evidence that specific interactions between distinct RNA motifs and the yet-to-be identified cellular factors control trafficking of an RNA from the same cell into different neighbors. The cellular factors could be cytosolic factors and/or plasmodesmal components. This model should have important implications for the
transport mechanisms of infectious and endogenous RNAs across various cellular boundaries.

The majority of studies on RNA functions rely on secondary structural analyses. In this study we demonstrated the utility of a combination of 3D motif search, comparison with X-ray crystal structures, mutagenesis and covariation analysis to study the tertiary structure of a functional RNA motif. Our results, together with those from previous work using similar approaches to investigate the tertiary structure and function of another PSTVd motif (Zhong et al., 2006), provide compelling evidence that viroid RNAs share many structurally conserved motifs with other RNAs and that they can be excellent models to investigate the general principles of RNA structure-function relationships. Furthermore, our approaches may be extended to investigate the general principles of RNA structure-function relationships.
Figure 2.1 Mutational analysis of PSTVd systemic infection. (A) Schematic of secondary structure of PSTVd, showing genomic location of the U43/C318 “loop” required for systemic infection. (B) Schematic showing the major cellular boundaries in inoculated and systemic leaves that PSTVd traffics through to establish systemic infection. The open spaces in the lines represent plasmodesmata. The arrows indicate direction of trafficking. For simplicity, the specific cell types in the phloem are not illustrated. (C) Enlarged view of nucleotide sequences of the short helices flanking the U43/C318 “loop” in the wild type (WT), C318A and U43G mutants.
Figure 2.2 Time course for PSTVd systemic infection in *N. benthamiana* plant. (A) RNA blots showing accumulation of both the (+)-circular (c-PSTVd) and linear (l-PSTVd) viroid RNAs at successive days post inoculation (dpi). rRNA serves as loading control. (B) Quantification of the accumulation levels of c-PSTVd from RNA blot results, calculated by using the known amounts of loading size markers in the same gel as references. Each data point represents the mean value of three biological replicates.
Figure 2.3 Secondary structures of PSTVd wild type (WT) and C318A and U43G mutants, predicted by mfold. The red box delineates the region involved in trafficking.
Figure 2.4 RNA blots showing defects of C318A and U43G mutations in systemic infection in *N. benthamiana*. (A) Accumulation of WT, but not the two mutants, in systemic leaves. (B) Accumulation of the two mutants as well as the WT in inoculated leaves. (C) Accumulation of the two mutants as well as the WT in infected protoplasts. rRNA shows loading control. c-PSTVd, (+)-circular PSTVd; M, mock control.
Figure 2.5 Mutants C318A and U43G fail to traffic into the vascular tissue phloem at 12 days post inoculation. (A) In situ hybridization shows trafficking of the two mutants as well as the WT in mesophyll. The purple hybridization signals show the presence of viroid RNAs in the nuclei (arrows). There is no hybridization signal in mock-inoculated samples. (B) In situ hybridization shows trafficking of the two mutants as well as the WT into the bundle sheath (BS). While the WT also accumulates in the phloem (Ph), the two mutants are absent from the phloem. Scale bars=10µm. (C) RT-PCR detects presence of WT PSTVd (lanes 1 and 2), but absence of mutants C318A (lanes 3 and 4) and U43G (lanes 5 and 6), in petioles of inoculated leaves. Amplification of 18S rRNA serves as an internal control. (D) Hybridization with PSTVd specific probe confirms identity of the RT-PCR products as PSTVd. (Figures 2.5A and 2.5B are provided by Dr. Xiaorong Tao.)
A  Mutants traffic in mesophyll

![Mock](image1.png) ![WT](image2.png)

![C318A](image3.png) ![U43G](image4.png)

B  Mutants fail to traffic into phloem

![Mock](image5.png) ![WT](image6.png)

![C318A](image7.png) ![U43G](image8.png)

C  Ladder (kb)  WT  Mutants

<table>
<thead>
<tr>
<th>Ladder (kb)</th>
<th>Mock</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- PSTVd
- 18S rRNA

D  PSTVd

Figure 2.5
6 dpi

Figure 2.6 In situ hybridization reveals failure of mutants C318A and U43G to traffic into the vascular tissue phloem at 6 days post inoculation (dpi). (A) Accumulation of the two mutants as well as the WT in epidermis. The purple hybridization signals show presence of the viroid RNAs in the nuclei (arrows). (B) Accumulation of the two mutants as well as the WT in mesophyll. (C) In situ hybridization shows trafficking of the two mutants as well as the WT into the bundle sheath (BS). While the WT also accumulates in the vascular tissue phloem (Ph), the two mutants are absent from the phloem. Scale bars ==10µm. (This figure is provided by Dr. Xiaorong Tao.)
Figure 2.7 In situ hybridization reveals failure of mutants C318A and U43G to traffic into the vascular tissue phloem at 18 days post inoculation (dpi). (A) Accumulation of the two mutants as well as the WT in epidermis. The purple hybridization signals show presence of the viroid RNAs in the nuclei (arrows). (B) Accumulation of the two mutants as well as the WT in mesophyll. (C) In situ hybridization shows trafficking of the two mutants as well as the WT into the bundle sheath (BS). While the WT also accumulates in the vascular tissue phloem (Ph), the two mutants are absent from the phloem. Scale bars ==10μm. (This figure is provided by Dr. Xiaorong Tao.)
Figure 2.8 Local distortion of an RNA helix due to the presence of a water-inserted cWW U/C basepair. The x-ray crystal structure of an RNA helix containing a U/C water-inserted basepair (Nucleotides 763:765 and 899:901 from PDB File: 1s72) is superimposed on a helix that contains matching cWW G/C flanking basepairs and cWW A/U in place of U/C (Nucleotides 444:446 and 36:38 from PDB file: 1s72). (A) Stereo figure showing the U/C-containing helix (orange), with the inserted water molecule shown as an orange sphere, superimposed on the helix 5’-CAG-3’/3’-GUC-5’ (blue; see panel C). (B) Planar view of the superimposition of the central water-inserted cWW U/C basepair (orange) on the cWW A/U basepair (blue). Insertion of the water opens the U/C basepair toward the minor groove, with little distortion of the neighboring basepairs. (C) Nucleotide sequences of the superimposed nucleotides. (This figure is prepared by Jesse Stombaugh and Dr. Neocles Leontis.)
Figure 2.8
Figure 2.9 Mutational analyses of the trafficking motif. (A) Summary of mutagenesis results, with green and red boxes showing basepairs competent of and defective in systemic trafficking, respectively. Grey stars indicate mutants that do not replicate. WT indicates wild type U43/C318 motif. (B) RNA blots showing presence or absence of the various mutants in systemic leaves of inoculated *N. benthamiana*. (C) RNA blots show presence or absence of the various mutants in inoculated *N. benthamiana* protoplasts. c-PSTVd, circular PSTVd; 1-PSTVd, linear monomeric PSTVd. rRNA serves as loading controls.
Figure 2.9

A  Trafficking functions of mutants

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>U43 C31Δ</td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>U</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Trafficking</td>
</tr>
<tr>
<td>A</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>U</td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

B  Accumulation in systemic leaves

C  Replication in protoplasts
Figure 2.10 Hydration patterns for the cWW and water-inserted cWW basepairs. Dark blue and light blue spheres represent high density and medium density hydration sites identified on the Solvation Web Service site or by structure search using FR3D. Background colors correspond to the experimentally determined trafficking function. (This figure is prepared by Jesse Stombaugh and Dr. Neocles Leontis.)
Figure 2.10
Figure 2.11 Covariation analyses of U/C motif in viroids of family Posipiviroidea, based on conservation of the overall secondary structures predicted by mfold. The analyses reveal conservation of U/C pair at the equivalent positions for Tomato planta macho viroid (TPMVd), Tomato chlorotic dwarf viroid (TCDVd), variation C/C for Citrus Exocortici viroid (CEVd), Tomato apical stunt viroid (TASVd) and Chrysanthemum stunt viroid (CSVd), and variation A/A for Mexican papita viroid (MPVd).
Figure 2.12 Secondary structures of PSTVd and other viroids in the family *Pospiviroidae* that show conserved secondary structures, based on mfold prediction. The arrows indicate the highly conserved loop E. TPMVd, *Tomato planta macho viroid*; TCDVd, *Tomato chlorotic dwarf viroid*; CEVd, *Citrus Exocortis viroid*; TASVd, *Tomato apical stunt viroid*; CSVd, *Chrysanthemum stunt viroid*; MPVd, *Mexican papita viroid*. The red box delineates the region involved in the analyses.
Figure 2.13 Covariation analysis in 23S bacterial sequences for three conserved U/C basepairs found in 23S \textit{E. coli}. Nucleotide numbering is based on the \textit{E. coli} sequence. The background colors indicate trafficking function in PSTVd variants for U43/C318. (This figure is prepared by Jesse Stombaugh and Dr. Neocles Leontis.)
CHAPTER 3

GENOME-WIDE MUTATIONAL ANALYSES IDENTIFY MULTIPLE AND DISTINCT VIROID RNA MOTIFS CRITICAL FOR REPLICATION AND SYSTEMIC TRAFFICKING
3.1 ABSTRACT

RNA replication and systemic trafficking play significant roles in developmental regulation and host-pathogen interactions. Viroids, as the simplest eukaryotic RNA pathogens and genetic units that are capable of autonomous replication and systemic trafficking, offer excellent models to study the underlying mechanisms. Like other RNAs, the predicted secondary structure of a viroid RNA contains many loops and bulges flanked by double-stranded helices, with their biological relevance remaining mostly unknown. To build a basis for comprehensive functional studies of the viroid RNA structure, we conducted a genome-wide mutational analysis of the role of Potato spindle tuber viroid (PSTVd) loops/bulges in replication in single cells and systemic trafficking in a plant. We identified multiple and distinct motifs essential or important for each of these processes. We present a genomic map of the PSTVd replication/trafficking motifs which shows that, in the PSTVd secondary structure, the motifs most critical for replication are clustered in the distal end of the left terminal domain and central region, whereas the trafficking motifs are mostly clustered in the variable and right terminal domains and around the pathogenicity domain. This map may guide further studies on the tertiary structure and functional mechanisms of each motif and identification of the cognate cellular factors.

3.2 INTRODUCTION

RNA replication not only represents a crucial milestone in the evolution of life based on the “RNA World” scenario (Gilbert, 1986; Joyce, 2002), but also impacts modern life as exemplified by the amplification of infectious RNAs and RNA silencing signals
(Baulcombe, 2004; Ding and Itaya, 2007b). Upon synthesis, some RNAs traffic into neighboring cells or even distant organs to regulate global gene expression that controls development and defense (Lough et al., 2006; Ding and Itaya, 2007a; 2007b). RNA trafficking is also an integral part of viral systemic infection (Lough et al., 2006; Ding and Itaya 2007a; 2007b). Viroids, which may represent living fossils of the pre-cellular RNA world (Diener, 1989) or have evolved later, for instance speculatively from escaped introns (Diener, 1981), are the smallest and simplest eukaryotic RNA pathogens and genetic units that are capable of autonomous replication and systemic trafficking. As such, viroids offer excellent models to study the evolution of RNA structure-function relationships that control RNA replication and trafficking as well as the minimal genetic elements required to make a eukaryotic pathogen.

Structurally, viroids are single-stranded and covalently closed circular RNAs. Without protein-coding capacity, encapsidation and helper viruses, these RNAs must survive the cellular environment, replicate and then spread within an infected plant through direct interactions between their genomes or genome-derived RNAs and cellular factors (Flores et al., 2005; Owens, 2007; Ding and Itaya, 2007b). Evidently, a viroid genome contains structural motifs to mediate diverse functions necessary to establish systemic infection. The secondary structures of viroids have been well studied in the past three decades (Owens, 2007). Like other RNAs, the secondary structure of a viroid RNA contains many loops and bulges flanked by double-stranded helices. There is increasing evidence that loops and bulges are often highly structured three-dimensional motifs formed via non-Watson-Crick base pairing, base stacking and other base interactions. More importantly, these motifs are major sites for RNA-RNA, RNA-protein and RNA-small
ligand interactions (Leontis et al., 2002a; Holbrook, 2005; Noller, 2005; Leontis et al., 2006).

An outstanding issue regarding viroid RNAs is whether all of the predicted loops and bulges have specific biological functions for infection. Except for a few structural motifs that have been shown to play a role in replication and cell-to-cell trafficking (Ding and Itaya, 2007b), the biological significance of most loops and bulges remains unknown. Is every loop/bulge important for viroid function? If so, what is its specific role in the infection process? Do replication and trafficking each involve single or multiple motifs?

We address these fundamental questions with *Potato spindle tuber viroid* (PSTVd) as a model. PSTVd is the type member of the family *Pospiviroidae* that replicates in the nucleus (Flores et al., 2000). Its 359-nucleotide genome is predicted to form a rod-like secondary structure in native in vitro state (Gross et al., 1978) that can be divided into five structural domains: the left and right terminal, pathogenicity, central and variable domains (Fig. 3.1; Keese and Symons, 1985). To develop a systemic infection, the incoming (+)-strand, circular PSTVd RNA enters the nucleus of an initial recipient cell through the nuclear pore complex and is transcribed into multimeric (-)-strand in the nucleoplasm. The latter is in turn transcribed into multimeric (+)-strands, which are presumably transported into the nucleolus to be cleaved into unit length RNAs and circularized. The circular RNAs likely exit the nucleus, move into adjacent cells through plasmodesmata and then enter the vascular tissue phloem to spread into new organs (Ding and Itaya, 2007b). Thus, there are numerous steps during the infection process that likely involve RNA motif-cellular factor interactions.
Transcription of the (+)-strand PSTVd RNA probably initiates at U359 or C1 at the left terminal loop (Kolonko et al., 2006). A putative metastable hairpin II (HPII; see Fig. 3.1A and 3.1B) has been suggested to play a role in replication (Loss et al., 1991; Owens et al., 1991; Schroder et al., 2002). Metastable hairpin I (HPI; see Fig. 3.1A and 3.1B) has also been speculated to be important for infectivity (Hammond and Owens, 1987). The loop E (see Fig. 3.1A), which comprises conserved non-Watson-Crick base pairs (Zhong et al., 2006) and which was recently shown to exist in vivo (Eiras et al., 2007; Wang et al., 2007), is involved in in vitro processing (Baumstark et al., 1997; Schrader et al., 2003), transcription regulation (Qi and Ding, 2002; Zhong et al., 2006), host adaptation (Wassenegger et al., 1996; Qi and Ding, 2002; Zhu et al., 2002) and symptom expression (Qi and Ding, 2003). Specific PSTVd motifs mediate trafficking from bundle sheath to mesophyll (Qi et al., 2004a) and from bundle sheath into the phloem (Zhong et al., 2007). These are only a handful of the many potential motifs in PSTVd that have been identified. For instance, numerous mutations scattered along the PSTVd secondary structure negatively affect infection in inoculated tomato plants (Hammond and Owens, 1987; Loss et al., 1991; Owens et al., 1991; Qu et al., 1993; Owens et al., 1995; Hu et al., 1996). However, the specific step of the infection process (e.g., replication vs. trafficking) affected by a mutation remains unknown, preventing further mechanistic studies.

We have conducted a genome-wide mutational analysis of the function of PSTVd loop/bulge motifs in replication in single cells and systemic trafficking in a plant. We identified multiple and distinct motifs essential or important for each of these processes. We present a genomic map of replication/trafficking motifs for PSTVd that may guide further mechanistic studies on the specific biological function of each motif during
infection. We discuss our results with regard to viroid infection and their broad biological implications.

3.3 MATERIALS AND METHODS

3.3.1 Plant materials and growth condition

*Nicotiana benthamiana* plants were grown in a growth chamber controlled at 14 hour light (27 °C)/10 hour dark (24 °C) cycles. Cultured cells of *N. benthamiana* were maintained in Murashige and Skoog medium (MS salts; Life Technologies, Rockville, MD) supplemented with 30 g/L of sucrose, 256 mg/L of KH₂PO₄, 100 mg/L of myo-inositol, 1 mg/L of thiamine, and 1 mg/L of 2,4-D with a final pH of 5.5. The detailed protocols are described in Zhong et al. (2005).

3.3.2 PSTVd cDNA construction

Plasmid pRZ6-2 containing cDNAs of PSTVd⁻Int was constructed by Hu et al. (1997) and was a gift from Dr. Robert Owens. All PSTVd-derived mutants were generated by site-directed mutagenesis using the Quickchange Site-Directed Mutagenesis kit (Stratagene, La Jolla, CA) using pRZ:PSTVd⁻Int as the template. The introduced mutations were verified by sequencing. Construction of pInter(−) was described in Qi and Ding (2002).
3.3.3 In vitro transcription

To prepare PSTVd inocula, HindIII-linearized plasmid pRZ6-2 containing PSTVd cDNA was used as the template for in vitro transcription with T7 MEGAscript (Ambion, Austin, TX). To prepare riboprobes for RNA gel blotting, [α-\textsuperscript{32}P]-UTP-labeled antisense riboprobes were prepared by in vitro transcription using T7 MAXIscript kit (Ambion) using SpeI-linearized pInter(−) as the template. After in vitro transcription, the DNA templates were removed by digestion with RNase-free DNase I. The RNA transcripts were purified with MEGAClear kit (Ambion). Nonradioactive and radioactive RNA transcripts were quantified by UV spectrometry or scintillation counting, respectively.

3.3.4 Plant and protoplast infection

The in vitro transcripts of PSTVd variants were used to inoculate the carborundum-dusted first two true leaves of two-week old *N. benthamiana* plants (300 ng/plant). DEPC-H\textsubscript{2}O was used for mock inoculation. Four weeks after inoculation, total RNAs were extracted from systemic leaves for RNA gel blot. *N. benthamiana* protoplasts were prepared and transfected with PSTVd transcripts by electroporation as described in Zhong et al. (2005). At three days post inoculation, transfected protoplasts were collected for RNA extraction and gel blot analysis.

3.3.5 RNA extraction and RNA gel blot

Total RNA from infected plants was isolated using Trizol reagent (Invitrogen, Carlsbad, CA) and total RNA from infected protoplasts was extracted using RNeasy plant
mini kit (Qiagen, Valencia, CA). Five μg of total RNA was separated on 5% polyacrylamide/8M urea gel. After electrophoresis, the RNA was transferred to a Hybond-XL nylon membrane (Amersham Biosciences, Piscataway, NJ) using a vacuum blotting system (Amersham) and immobilized by UV-crosslinking. Hybridization with [α-32P]-UTP-labeled riboprobes was carried out at 65°C using ULTRAhyb reagent (Ambion). After overnight hybridization, the membranes were washed twice in 2XSSC/0.1%SDS for 15 min and twice in 0.2XSSC/0.1%SDS for 15 min at 65°C and exposed to a Storage Phosphor Screen (Kodak, Rochester, NY). Hybridization signals were quantified with Molecular Imager FX using Quantity One-4.1.1 software (Bio-Rad, Hercules, CA).

3.3.6 Sequencing of RNA progeny

The protocols for preparing cDNAs of the PSTVd progeny isolated from infected plants were essentially as described by Qi and Ding (2002). Briefly, cDNAs of PSTVd RNA were RT-PCR amplified and sequenced in both directions using the ABI377 DNA sequencer (Perkin-Elmer, Boston, MA) at the DNA Sequencing Facility at Ohio State University.

3.4 RESULTS

3.4.1 Generation of PSTVd mutants with disrupted loops

For simplicity of description, the term “loop” is used in this report to broadly include both loops and nucleotide bulges. Numbered from left to right, there are a total of
27 loops in the PSTVd secondary structure (Fig. 3.1A). We previously showed loop 15 (loop E) to be essential for replication (Zhong et al., 2006) and loop 7 (U43/C318 motif) to be critical for systemic trafficking (Zhong et al., 2007). Therefore, they were not included for detailed analyses in this study. However, we did include A99C, a loop E-disruptive mutant that is defective in replication (Zhong et al., 2006), as a negative control in addition to mock inoculation in all replication experiments.

To examine the function of the other 25 loops in PSTVd replication and systemic trafficking, we generated a series of mutants with each containing one disrupted loop. Specifically, we designed mutations to close each internal loop by introducing canonical Watson-Crick base pairs or delete bulged nucleotides. The two terminal loops cannot simply be closed due to structural constraints. Therefore, we generated mutations to enlarge each of them based on previous studies (Hammond and Owens, 1987). All mutations did not affect the overall PSTVd secondary structure as predicted by mfold (Zuker, 2003; Fig. 3.2).

In addition to the loop-closing or loop-deleting mutants, we investigated the effects of helical mutations on PSTVd infection (see arrows in Fig. 3.1A for specific mutations). These mutations were shown in previous studies to affect PSTVd infection (Hammond and Owens, 1987; Owens et al., 1991). These included G46C (Hammond and Owens, 1987), G26C and G65U (Owens et al., 1991) that completely abolished infectivity in tomato plants. Whether they inhibited replication in single cells or systemic trafficking was unknown.

Finally, we also generated selected mutants with disrupted HPII and HPI to test the function of these metastable structures in replication and systemic trafficking. C229U,
C231A and G324U substitutions are each predicted to disrupt HPII formation (Loss et al., 1991; Qu et al., 1993; see arrows in Fig. 3.1B). G80A and C109U substitutions are each predicted to disrupt HPI (Hammond and Owens, 1987; see arrows in Fig. 3.1B). Double mutations C231A/G324U and G80A/C109U are predicted to restore formation of HPII and HPI, respectively (Hammond and Owens, 1987; Loss et al., 1991; Qu et al., 1993).

For all mutants, we tested their replication in single cells and systemic trafficking in a whole plant of *Nicotiana benthamiana*. Specific results are presented below.

### 3.4.2 Multiple PSTVd loops were critical for replication in single cells

To determine the replication capacity of each mutant, we analyzed its accumulation in protoplasts of *N. benthamiana* cultured cells (Qi and Ding, 2002). Three days post-inoculation, RNA was extracted and analyzed by Northern hybridization using PSTVd-specific riboprobes. The accumulation levels of circular PSTVd RNA were quantified, based on results from three biological replicates, for each mutant and the data are presented as the percentage of the wild type (WT) level (Fig. 3.1C).

In negative control experiments, mutant A99C with disrupted loop E accumulated to less than 5% of the WT level, consistent with previous results (Zhong et al., 2006). The other loop mutants exhibited different levels of accumulation compared to the WT, ranging from less than 20% to over 110% (Fig. 3.1C). Examination of the overall patterns reveals that the loops in the distal end of the left terminal domain are among the most critical for replication. Specifically, enlarging loop 1, deleting loops 2 and 3 and closing loop 4 in this domain reduced replication to below 20%. This is above the 5% level of A99C, suggesting that these mutants still replicated, albeit inefficiently. The variable domain appears to be
the least important for replication. Specifically, closing loops 18, 19, 20, 21 and 22 had only moderate to minimal effects on replication, as shown by their accumulation levels between 40-80% of the WT. The right terminal and the pathogenicity domains are important, but apparently not essential, for replication because closing loops in these domains consistently maintained replication at 30-60% of the WT levels.

The central region showed a more complex pattern. While disruption of loop 15 (loop E) and loop 13 severely inhibited replication to 5% and 10% of the WT level, respectively, deletion of loop 14 consistently enhanced replication to over 110% of the WT level. Disruption of the other loops in this region decreased replication to 20-50% of the WT levels.

3.4.3 Multiple loops were essential for PSTVd systemic trafficking

To identify PSTVd loops that might be important for systemic trafficking, we mechanically inoculated each mutant onto the first two true leaves of two-week-old *N. benthamiana* seedlings. At four-weeks post-inoculation, the presence or absence of PSTVd in systemic leaves was determined by dot blot. Twelve plants, from three biological replicates each involving four plants, were inoculated with each mutant as well as the WT. The results are pooled and presented in Table 3.1-3.3. To determine the trafficking function of a mutant, several factors need to be considered: 1) the replication level of the mutant in protoplasts, 2) the number of inoculated plants that showed systemic infection, and 3) the maintenance of introduced mutations in the RNA progeny. Because reduced replication levels could influence the trafficking capacity of PSTVd as a population, it is important to determine the minimal replication level, relative to the WT, that is still
sufficient to sustain systemic infection. As shown in Fig. 3.1C (black column), deletion of loop 11 reduced PSTVd accumulation to slightly over 20% of the WT level in protoplasts. However, this mutant systemically infected 9 out of the 12 inoculated plants (Table 3.5). Sequencing of viroid progeny from three of the systemically infected plants showed maintenance of the deletion in two plants and additional mutations in the third plant (Table 3.5). Although the effects of additional mutations remain to be understood, it is clear that a 20% of the WT replication level is sufficient for systemic trafficking. Based on these considerations, we operationally classified the trafficking functions of loop mutants into three categories: 1) trafficking-defective, 2) trafficking-impaired and 3) trafficking-unknown. Each is described in detail below.

**Trafficking-defective mutants:** These mutants accumulated to more than 20% of the WT levels in the protoplasts but failed to accumulate in systemic leaves in all of the 12 inoculated plants. These mutants included loops 6, 10, 12, 17, 18, 20, 24, 25 and 26 (Table 3.1; white columns in Fig. 3.1C). These loops are mostly clustered in the variable and right terminal domains and around the pathogenicity domain.

**Trafficking-impaired mutants:** These mutants accumulated to more than 20% of the WT levels in the protoplasts. However, they showed systemic infection in less than 50% of the inoculated plants. As shown in Table 3.2 and grey columns in Fig. 3.1C, loop mutants 5, 8, 9, 14, 16, 19, 21, 22, 23, and 27 belong to this group. We sequenced viroid progeny from a fraction of the systemically infected plants to determine whether the trafficking mutants maintained only the original mutations, acquired new mutations or reverted to the WT. In all cases we detected reversions or additional mutations (Table 3.5). Because the effects of these spontaneous mutations on PSTVd systemic trafficking are not clear, a designation of
“trafficking-impaired” suggests that the original mutations did not abolish, but inhibited to some extent, the trafficking function.

**Trafficking-unknown mutants:** Loop mutants 1, 2, 3, 4 and 13 accumulated to less than 20% of the WT level in protoplasts (Table 3.3; hatched columns in Fig. 3.1C). Systemic accumulation of the mutants in this group was absent in most of the inoculated plants (Table 3.3). Sequencing of RNA progeny from a fraction of the systemically infected plants detected additional mutations (Table 3.5). How these additional mutations affect trafficking and replication is unclear at this stage. Because 20% is below the currently known replication level sufficient to support trafficking, we do not know whether compromised systemic accumulation of these mutants is due to the low replication levels or specific defects in systemic trafficking.

**3.4.4 Mutations in helical regions altered neighboring loop structures and inhibited PSTVd systemic trafficking**

As discussed above, mutations G46C (Hammond and Owens, 1987), G26C and G65U (Owens et al., 1991) completely abolished PSTVd infectivity in tomato plants, but whether they inhibited replication or trafficking is unclear. As summarized in Table 3.4, our analyses showed that these mutants replicated to 20-50% of the WT levels but all failed to accumulate in systemic leaves of *N. benthamiana* plants. Sequencing of viroid progeny from the only plant systemically infected by G65U showed reversion to the WT sequence.

Secondary structural prediction by mfold (Zuker, 2003) showed that all mutations not only disrupted local helices, but also affected the adjacent loops (Fig. 3.3). For example, G26C enlarges loops 4 and 5 and G46C enlarges loops 6 and 7. G65U alters loop 9 and
creates a small bulge between loops 9 and 10. These data suggest that defects in systemic trafficking may result from alteration of the loops, which were shown independently to be important for trafficking by the above loop-disruption mutations. However, we cannot exclude the role of helical structures at this stage.

3.4.5 Disruption of HPII and HPI formation did not inhibit PSTVd replication in single cells

Previous studies showed that mutations predicted to inhibit HPII formation always reverted to WT in infected tomato plants, suggesting the importance of HPII for replication (Loss et al., 1991; Owens et al., 1991; Qu et al., 1993). However, reversion is not direct evidence of HPII functions. To directly test the potential role of HPII in replication, we analyzed the replication and trafficking functions of selected HPII mutants. As shown in Table 3.4, C229U, C231A and G324U mutations, each predicted to disrupt HPII formation (Loss et al., 1991), maintained replication functions in *N. benthamiana* protoplasts. While C231A, G324U and C231A/G324U also trafficked systemically, C229U showed defects in systemic trafficking. Sequencing confirmed maintenance of the introduced mutations in plants inoculated by C231A, G324U and C231A/G324U. We detected several additional mutations in the viroid progeny from plants inoculated by C231A and G324U but not by C231A/G324U (Table 3.5). These mutations are not expected to affect HPII formation but their biological effect is unknown. These data showed that HPII formation is not essential for PSTVd replication in *N. benthamiana* cells. Analysis by mfold showed no effect of C229U substitution on the local secondary structure of PSTVd (Fig. 3.3). The inhibitory effect of this mutation on systemic trafficking suggests either HPII plays a role in
trafficking or more likely the specific nucleotides in the short helix between loops 18 and 19 play a role, together with these loops, in trafficking.

HPI is also suggested to be important for infectivity in tomato plants (Hammond and Owens, 1987). As shown in Table 4, the two HPI-disrupting single mutants, G80A and C109U, and the HPI-restoring double mutant G80A/C109U (Hammond and Owens, 1987) replicated to 50-70% of the WT level but all failed to establish systemic infection in *N. benthamiana*. Therefore, HPI formation is not necessary for PSTVd replication in this plant. The inhibitory effects of G80A on trafficking cannot be simply attributed to the requirement for HPI, because G80A mutation distorts loops 10, 11 and 12 which are essential for trafficking (see Fig. 3.1C). C109U mutation does not alter the helix or neighboring loops. Its trafficking-inhibitory effect could be attributed to a role of specific nucleotides in the helix or to the involvement of HPI. However, since the double mutant G80A/C109U is predicted to restore HPI, but not the native secondary structure (Hammond and Owens, 1987), its inhibition of PSTVd trafficking suggests that HPI unlikely plays a role in trafficking in *N. benthamiana*.

### 3.5 DISCUSSION

By using a combination of mutational and biological experiments that allowed separation of replication and trafficking functions, we have obtained loss-of-function genetic evidence for the role of various PSTVd loops in replication or trafficking. We present in Figure 3.4 a PSTVd genomic map that links specific motifs to replication (R)
and systemic trafficking (T). In this map, we include only loops the disruption of which reduced PSTVd replication to below 20% of the WT level and rendered PSTVd defective in systemic trafficking, respectively. We also include findings on the requirement of loop E (loop 15) for replication (Zhong et al., 2006) and U43/C318 motif (loop 7) for vascular entry (Zhong et al., 2007). This map shows that, in the PSTVd secondary structure, the motifs most critical for replication are clustered in the distal end of the left terminal domain and central region, whereas the trafficking motifs are mostly clustered in the variable and right terminal domains and around the pathogenicity domain. This map represents a major advance over previous mutational studies that defined the negative effect of a mutation as non-infectious at the whole plant level without specific insights into replication or trafficking functions. The map may serve as a guide for further experimental studies to test the model and to gain mechanistic insights.

A question basic to both replication and trafficking is whether accomplishment of a particular process involves one or more motifs and whether a motif is involved in one or more processes. For instance, our results suggest that the distal end of the left terminal domain plays an important role in replication. It will be of great interest to determine whether loops 1, 2, 3 and 4 in this domain function at the same or distinct steps of replication. The importance of loop 1 for replication is consistent with the in vitro mapping of transcription initiation sites at the left terminal loop (Kolonko et al., 2006). Interestingly, it was shown previously that RNA polymerase II, which appears to be involved in PSTVd replication (Schindler and Mühlbach, 1992), binds the terminal regions in vitro (Goodman et al., 1984). The left terminal domain therefore deserves further studies for its role in replication. The other region that is evidently involved in regulating replication is the
central region. Its role in replication has long been suggested (Keese and Symons, 1985) with clear demonstration for in vitro processing (Baumstark et al., 1997; Schrader et al., 2003). We previously provided loss-of-function genetic evidence for the requirement of loop E (loop 15) in this region for replication in vivo (Zhong et al., 2006). Here, we showed that loop 13 is also essential for replication. Strikingly, deletion of loop 14 (bulged U) consistently led to enhanced replication, albeit at a modest level. The significance of these observations remains to be understood. They do, however, raise the question of whether positive and negative regulatory elements exist in PSTVd to maintain an optimal level of replication beneficial to the viroid population. Further studies on the specific role of these loops in RNA stability, transcription, subcellular localization and in vivo processing, together with identification of the cellular factors that interact with these loops, should provide mechanistic insights into the functions of these loops.

For the other loops the disruption of which reduced PSTVd replication to 20-80% of the WT levels, whether they are directly or indirectly involved in replication remains a matter of speculation. We favor the idea that many if not all of these loops have an indirect involvement, because they are apparently not essential. The mild to moderate inhibitory effects of their disruptions could result from unknown effects of slight structural disturbances.

The identification of distinct PSTVd motifs that mediate trafficking from bundle sheath to mesophyll (Qi et al., 2004a) and from bundle sheath to phloem (Zhong et al., 2007) led to the hypothesis that trafficking of the PSTVd RNA across different cellular boundaries requires different structural motifs (Ding et al., 2005; Ding and Itaya, 2007b). In this study we have identified several loops that are clearly essential for systemic
trafficking (Fig. 3.4). Those loops the disruption of which lead to impaired trafficking (Table 3.2) may also play a critical role in trafficking. These findings raise the question of whether each loop motif or a combination of some is required and sufficient for trafficking across various cellular boundaries. For instance, we previously showed that loop 7 (U/C motif) is required for phloem entry (Zhong et al., 2007), but whether it alone is sufficient remains unclear. Obviously, further investigations using in situ hybridization should provide important insights into the specific role of each of the identified loops in trafficking across various cellular boundaries. It should be noted that the identification of loops 24, 25, 26 at the right terminal domain as essential for trafficking is consistent with previous findings that mutations in the right terminal loop inhibited PSTVd systemic infection when inoculated onto tomato (Hammond, 1994). This is also the region that has been shown to interact with VIRP1 (Maniataki et al., 2003; Gozmanova et al., 2003). Thus, the right terminal region can be a focus for further studies on the RNA-protein interactions that direct trafficking across some cellular boundary yet to be identified. Motif-mediated trafficking across specific cellular boundaries has broader significance beyond viroid infection. There is now compelling evidence that viral RNAs contain motifs to mediate trafficking (Lough et al., 2006; Gopinath and Kao, 2007). Specific RNA motifs yet to be defined also regulate trafficking of cellular RNAs (Haywood et al., 2005; Banerjee et al., 2006).

In discussing our results in the context of structural motifs, we need to consider alternative or metastable structures that may form during PSTVd replication. In such cases, mutations that are designed to close a loop or delete a bulge may affect the formation of an alternative, functional motif. Thus far, HPII has been the most extensively investigated. In
this study, we showed that all HPII-disruptive mutations, as well as HPI-mutations, did not abolish PSTVd replication in *N. benthamiana*. Thus, the potential functions of these putative metastable structures remain equivocal. Here, we cannot exclude the possibility that their functions are host-specific. Overall, the data from our current work support the notion that the native secondary structure of PSTVd regulates many aspects of replication/trafficking.

We wish to point out that our current mutagenesis approach relies on thermodynamic calculations of the secondary RNA structures using mfold (Zuker, 2003), which understandably may not always predict every structure that truly exists and functions in vivo (Leontis et al., 2006). We previously inferred the tertiary structures of loop E (Zhong et al., 2006) and U43/C318 motif (Zhong et al., 2007) by comparison with crystal structures of similar motifs in other RNAs. This approach also allows rational mutagenesis to generate disruptive and compensatory mutations targeting specific non-Watson-Crick base pairs for functional analyses. An important goal for future work is to apply similar approaches to investigate the tertiary structure and specific function of the replication/trafficking motifs identified in this study. Their in vivo existence must also be tested. In the broad context of replication, the specific role a motif can play include maintaining RNA stability, nuclear import, transcription, nucleolar import, cleavage, ligation, nucleolar export, and nuclear export. In terms of systemic trafficking, specific functions include trafficking across different cellular boundaries, entry into the phloem, long-distance movement in the phloem, and exit the phloem to traffic across different cellular boundaries. Taken together, our findings reported here and their implications
support the notion that PSTVd can be a productive model to investigate comprehensively how distinct RNA motifs interact with cellular factors to regulate a wide range of cellular processes of general biological significance.
<table>
<thead>
<tr>
<th>loop</th>
<th>Mutations</th>
<th>Domain $^a$</th>
<th>Replication $^b$</th>
<th>Trafficking $^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td></td>
<td></td>
<td>100%</td>
<td>11/12</td>
</tr>
<tr>
<td>6</td>
<td>C36U/A37C/C38G</td>
<td>$T_L$</td>
<td>30%</td>
<td>0/12</td>
</tr>
<tr>
<td>10</td>
<td>A290Δ/A291Δ</td>
<td>P</td>
<td>32%</td>
<td>0/12</td>
</tr>
<tr>
<td>12</td>
<td>A77Δ/G78Δ/C79Δ</td>
<td>CR</td>
<td>29%</td>
<td>0/12</td>
</tr>
<tr>
<td>17</td>
<td>A118G/A119C/A220Δ</td>
<td>CR</td>
<td>52%</td>
<td>0/12</td>
</tr>
<tr>
<td>18</td>
<td>A126G/C127Δ</td>
<td>V</td>
<td>63%</td>
<td>0/12</td>
</tr>
<tr>
<td>20</td>
<td>A222G/C223Δ</td>
<td>V</td>
<td>73%</td>
<td>0/12</td>
</tr>
<tr>
<td>24</td>
<td>U157G/A158G/U161A</td>
<td>$T_R$</td>
<td>25%</td>
<td>0/12</td>
</tr>
<tr>
<td>25</td>
<td>C166A/C167A</td>
<td>$T_R$</td>
<td>29%</td>
<td>0/12</td>
</tr>
<tr>
<td>26</td>
<td>C172GGA*</td>
<td>$T_R$</td>
<td>46%</td>
<td>0/12</td>
</tr>
</tbody>
</table>

$^a$ Structural domains in native secondary structure
   $T_L$, Left terminal;
   P, Pathogenicity
   CR, Central region;
   V, Variable
   $T_R$, Right terminal

$^b$ Replication efficiency of mutants expressed as percentage of wild type (WT) PSTVd

$^c$ Trafficking function is expressed as number of plants showing systemic infection, determined by dot blot, over total number of plants inoculated

* Insertion of GA between residues 172 and 173

**Table 3.1** Trafficking-defective mutants
<table>
<thead>
<tr>
<th>loop</th>
<th>Mutations</th>
<th>Domain</th>
<th>Replication</th>
<th>Trafficking</th>
<th>Genetic stability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WT</td>
<td></td>
<td>100%</td>
<td>11/12</td>
<td>S</td>
</tr>
<tr>
<td>5</td>
<td>U332Δ/U333G</td>
<td>T_L</td>
<td>36%</td>
<td>2/12</td>
<td>M/R</td>
</tr>
<tr>
<td>8</td>
<td>A50G/A51U</td>
<td>P</td>
<td>24%</td>
<td>1/12</td>
<td>R</td>
</tr>
<tr>
<td>9</td>
<td>C301Δ/C303U/A305U</td>
<td>P</td>
<td>37%</td>
<td>2/12</td>
<td>M/R</td>
</tr>
<tr>
<td>14</td>
<td>U267Δ</td>
<td>CR</td>
<td>110%</td>
<td>4/12</td>
<td>M</td>
</tr>
<tr>
<td>16</td>
<td>A112Δ/C113Δ</td>
<td>CR</td>
<td>27%</td>
<td>2/12</td>
<td>M</td>
</tr>
<tr>
<td>19</td>
<td>C227U</td>
<td>V</td>
<td>46%</td>
<td>1/12</td>
<td>NS</td>
</tr>
<tr>
<td>21</td>
<td>A142U</td>
<td>V</td>
<td>92%</td>
<td>4/12</td>
<td>M</td>
</tr>
<tr>
<td>22</td>
<td>C147G</td>
<td>V</td>
<td>41%</td>
<td>5/12</td>
<td>M/R</td>
</tr>
<tr>
<td>23</td>
<td>A150C/C151G</td>
<td>T_R</td>
<td>35%</td>
<td>4/12</td>
<td>M</td>
</tr>
<tr>
<td>27</td>
<td>U177A/U178A</td>
<td>T_R</td>
<td>53%</td>
<td>1/12</td>
<td>M</td>
</tr>
</tbody>
</table>

a,b,c: See Table 3.1 for legends  
S: Stable  
R: Reversion to WT  
M: Additional mutations  
NS: Not sequenced

**Table 3.2** Trafficking-impaired mutants
<table>
<thead>
<tr>
<th>loop</th>
<th>Mutations</th>
<th>Domain</th>
<th>Replication</th>
<th>Trafficking</th>
<th>Genetic stability</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td></td>
<td></td>
<td>100%</td>
<td>11/12</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>G2U/A4C/C6G</td>
<td>T_L</td>
<td>16%</td>
<td>1/12</td>
<td>R</td>
</tr>
<tr>
<td>2</td>
<td>A8Δ</td>
<td>T_L</td>
<td>16%</td>
<td>3/12</td>
<td>M</td>
</tr>
<tr>
<td>3</td>
<td>C13Δ</td>
<td>T_L</td>
<td>12%</td>
<td>2/12</td>
<td>M</td>
</tr>
<tr>
<td>4</td>
<td>U22A/U24G</td>
<td>T_L</td>
<td>130%</td>
<td>0/12</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>A271Δ/A272Δ/A273U/A274C/A275C</td>
<td>CR</td>
<td>10%</td>
<td>1/12</td>
<td>M</td>
</tr>
</tbody>
</table>

a,b,c: See Table 3.1 for legends
S: Stable
R: Reversion to WT
M: Additional mutations

**Table 3.3** Trafficking-impaired mutants
<table>
<thead>
<tr>
<th>Mutation</th>
<th>Metastable Structure</th>
<th>Domain&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Replication&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Trafficking&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Genetic stability</th>
<th>Tomato infectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td></td>
<td></td>
<td>100%</td>
<td>11/12</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>G26C</td>
<td>T&lt;sub&gt;L&lt;/sub&gt;</td>
<td></td>
<td>23%</td>
<td>0/12</td>
<td>N&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>G46C</td>
<td>T&lt;sub&gt;L&lt;/sub&gt;</td>
<td></td>
<td>42%</td>
<td>0/12</td>
<td></td>
<td>N&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>G65U</td>
<td>P</td>
<td></td>
<td>46%</td>
<td>1/12</td>
<td>R</td>
<td>N&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>C229U</td>
<td>HPII V</td>
<td></td>
<td>54%</td>
<td>3/12</td>
<td>R/M</td>
<td>R&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
<tr>
<td>C231A</td>
<td>HPII V</td>
<td></td>
<td>95%</td>
<td>9/12</td>
<td>S/R/M</td>
<td>N&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
<tr>
<td>G324U</td>
<td>HPII T&lt;sub&gt;L&lt;/sub&gt;</td>
<td></td>
<td>82%</td>
<td>8/12</td>
<td>S/R/M</td>
<td>N&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
<tr>
<td>C231A/G324U</td>
<td>HPII T&lt;sub&gt;L&lt;/sub&gt;,V</td>
<td></td>
<td>86%</td>
<td>8/12</td>
<td>S</td>
<td>N&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
<tr>
<td>G80A</td>
<td>HPI CR</td>
<td></td>
<td>53%</td>
<td>0/12</td>
<td>N&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>C109U</td>
<td>HPI CR</td>
<td></td>
<td>72%</td>
<td>0/12</td>
<td>N&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>G80A/C109U</td>
<td>HPI CR</td>
<td></td>
<td>66%</td>
<td>0/12</td>
<td>N&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

a,b,c: See Table 3.1 for legends
S: Stable
R: Reversion to WT
M: Additional mutations
N: Noninfectious
1: Hammond and Owens (1987)
2: Owens et al., (1991)
3: Qu et al., (1993)

**Table 3.4** Replication and trafficking of helical, HPI and HPII mutants
<table>
<thead>
<tr>
<th>loop</th>
<th>Original mutations</th>
<th>Trafficking</th>
<th>No. of plants sequenced</th>
<th>Progeny sequences (# of plants)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>A74Δ</td>
<td>9/12</td>
<td>4</td>
<td>A74Δ(2) A74G/U312C(1) U257A(1)</td>
</tr>
<tr>
<td>5</td>
<td>U332Δ/U333G</td>
<td>2/12</td>
<td>1</td>
<td>WT+C42U/U355C(1)*</td>
</tr>
<tr>
<td>9</td>
<td>C301Δ/C303U/A305U</td>
<td>2/12</td>
<td>1</td>
<td>WT+U257A(1)*</td>
</tr>
<tr>
<td>14</td>
<td>U267Δ</td>
<td>4/12</td>
<td>2</td>
<td>U276Δ/C216U(1) U267Δ/U316C(1)</td>
</tr>
<tr>
<td>16</td>
<td>A112Δ/C113Δ</td>
<td>2/12</td>
<td>2</td>
<td>A112Δ/C113Δ/G243U(1) A112Δ/C113Δ/U247A(1)</td>
</tr>
<tr>
<td>19</td>
<td>C227U</td>
<td>1/12</td>
<td>0</td>
<td>A142U/G221U(3) A142U/A219U(1)</td>
</tr>
<tr>
<td>21</td>
<td>A142U</td>
<td>4/12</td>
<td>4</td>
<td>WT(1) C231A(1) C167A(1)</td>
</tr>
<tr>
<td>22</td>
<td>C147G</td>
<td>5/12</td>
<td>3</td>
<td>WT(1) A150C/C151G/A135G(1) U177A/U178A/A182U(1)</td>
</tr>
<tr>
<td>23</td>
<td>A150C/C151G</td>
<td>4/12</td>
<td>2</td>
<td>A8Δ/A171G(2) C13Δ/U316C(2) A273U/A274C/A275C/U316C(1)</td>
</tr>
<tr>
<td>27</td>
<td>U177A/U178A</td>
<td>1/12</td>
<td>1</td>
<td>WT(2) C229U/C216U(1) C231A(2) A273U/A274C/A275C/U316C(1)</td>
</tr>
<tr>
<td>2</td>
<td>A8Δ</td>
<td>3/12</td>
<td>2</td>
<td>A8Δ/A171G(2) C13Δ/U316C(2) A273U/A274C/A275C/U316C(1)</td>
</tr>
<tr>
<td>3</td>
<td>C13Δ</td>
<td>2/12</td>
<td>2</td>
<td>C13Δ/U316C(2)</td>
</tr>
<tr>
<td>13</td>
<td>A271Δ/A272Δ/A273U/A274C/A275C</td>
<td>1/12</td>
<td>1</td>
<td>A8Δ/A171G(2) C13Δ/U316C(2) A273U/A274C/A275C/U316C(1)</td>
</tr>
<tr>
<td>HPII</td>
<td>C229U</td>
<td>3/12</td>
<td>3</td>
<td>WT(2) C229U/C216U(1) C231A(2) A273U/A274C/A275C/U316C(1)</td>
</tr>
<tr>
<td>HPII</td>
<td>C231A</td>
<td>9/12</td>
<td>5</td>
<td>WT(1) C231A/C216U(2) G324U(2)</td>
</tr>
<tr>
<td>HPII</td>
<td>G324U</td>
<td>8/12</td>
<td>4</td>
<td>WT(1) G324U/U329G(1)</td>
</tr>
</tbody>
</table>

WT: Reversion to wild type sequences
Red letters: New mutations
* Mixture of WT and additional mutations

**Table 3.5 Sequencing of viroid progeny from a fraction of systemically infected plants**
Figure 3.1 Functional analysis of PSTVd structural motifs. (A) The in vitro native rod-like secondary structure of PSTVd showing the five structural domains. T_L, left terminal domain; T_R, right terminal domain. The 27 loops in this secondary structure are numbered from left to right. The arrows indicate introduced mutations in the helical regions. The bars show the positions of nucleotides that are proposed to basepair to form metastable hairpin I (HPI) and II (HPII) structures as shown in (B). (B) The nucleotide sequences and secondary structures of HPII and HPI. The arrows indicate the nucleotide substitutions used in this study. (C) Summary of replication and systemic trafficking functions of loop mutants. The replication level of a mutant is presented as the percentage of the WT level. The systemic trafficking function of a loop mutant is indicated by column shading, with black denoting trafficking-normal mutants, white denoting trafficking-defective mutants, grey denoting trafficking-impaired mutants and hatched denoting trafficking-unknown mutants.
Figure 3.2 Thermodynamic prediction of secondary structures of PSTVd loop mutants (38). Only partial sequences are shown. Red letters denotes introduced mutations.
Figure 3.3 Thermodynamic prediction of secondary structures of helical, HPII and HPI mutants (Zuker, 2003). Only partial sequences are shown. Red letters denotes introduced mutations. The blue numbers indicate no change and the red numbers indicate structural changes in the loops. The dashed lines represent unchanged structures.
Figure 3.4 A genomic map of PSTVd loop motifs that are essential/critical for replication (R) and essential systemic trafficking (T).\(^1\) Data from Zhong et al (2006).\(^2\) Data from Zhong et al. (2007). \(T_L\), left terminal domain; \(T_R\), right terminal domain.
RNA-templated RNA replication is essential for viral or viroid infection, as well as for gene regulation at the cellular level. Systemic RNA trafficking is emerging as a new paradigm in gene regulation at the whole plant level. Here, we have used PSTVd as a model system to investigate the molecular mechanisms of RNA replication and systemic trafficking. These studies demonstrated that an RNA may contain multiple and distinct motifs for replication and for trafficking across different cellular boundaries.

A number of outstanding questions remain to be answered. First, we demonstrated a critical role of the loop E in replication, specifically in transcription (Zhong et al., 2006). How the loop E functions in transcription is not clear. The loop E motifs in cellular RNAs are well known to serve as important binding sites for RNAs and proteins (Leontis et al., 2002a). The loop E motif of PSTVd may serve as a binding site for the RNA polymerase or for a cellular factor(s) that recruits the RNA polymerase for transcription. Alternatively, the loop E may interact with a cellular factor to localize the PSTVd RNA to a particular subnuclear site to facilitate access to the nuclear transcription machinery. Given the critical role of the distal end of the left terminal domain of PSTVd in replication (Zhong and Ding, unpublished data) together with in vitro mapping of the transcription initiation sites at the left terminal end (Kolonko et al., 2006), it is also possible that the loop E function together with the left terminal end in regulating replication via interaction with a cellular factor or by a conformational change within the viroid RNA. These possibilities should be tested by further experimental studies.

Second, we propose a tertiary-structural model for PSTVd loop E motif on the basis of sequence comparison of 5S rRNA. It will be of great interest to determine the tertiary structure of PSTVd loop E by X-ray crystallography or NMR. Furthermore, additional
studies on the functions of other disruptive or compensatory mutations that are predicted on the basis of isosteric base-pairing principles should further validate this structural model.

Third, we have shown a critical role of the U/C motif in mediating vascular entry (Zhong et al., 2007). This motif is likely a binding site for some cellular factors based on the known function of the similar motif in rRNAs that interacts with specific proteins. Further detailed analyses of the role of the U/C base pair as well as the surrounding sequences and identification of the protein factor(s) that binds it should lead to new insights into the mechanism on the molecular regulation of RNA trafficking. Whether this U/C motif works alone or together with another motif to mediate PSTVd bundle sheath-to-phloem trafficking is another outstanding question remains to be answered. It will be of great interest to address these issues when experimental approaches become available. It will be also interesting in the future to verify the tertiary structure of this U/C trafficking motif by physical means such as X-ray crystallography or NMR.

Fourth, our mutational analyses identified multiple and distinct PSTVd motifs critical for replication and systemic trafficking. This finding raises the question of whether each motif or a combination of some is required or sufficient for replication and systemic trafficking. Further studies on the specific role of these motifs in one or more step(s) of replication and systemic trafficking using the combination of tertiary structural analyses and genetic, biochemical, cellular, and molecular tools that we have developed previously (Zhong et al., 2006; 2007) together with identification of the cognate cellular factors should provide mechanistic insights into the functions of these motifs. In the broader context of replication, the specific roles that a motif can play include maintaining RNA
stability, nuclear import, transcription, nucleolar import, cleavage, ligation, nucleolar export and nuclear export. For systemic trafficking, specific functions include trafficking across different cellular boundaries, entry into the phloem, long-distance movement in the phloem and exit the phloem to traffic across different cellular boundaries.

Taken together, our findings and their implications support the notion that PSTVd can be a productive model to investigate comprehensively how distinct RNA motifs interact with cellular factors to regulate a wide range of cellular processes of general biological significance.


Leontis NB and Westhof E (1998c) The 5S rRNA loop E: chemical probing and phylogenetic data versus crystal structure. RNA, 4, 1134-1153.


Qi Y and Ding B (2003) Inhibition of cell growth and shoot development by a specific nucleotide sequence in a noncoding viroid RNA. Plant Cell, 15, 1360-1374.


