INVESTIGATION OF RNA STRUCTURAL MOTIFS IN A VIROID
REQUIRED FOR ITS INTERCELLULAR TRAFFICKING
AND
CHARACTERIZATION OF A PLANT RNA LIGASE
ESSENTIAL FOR VIROID REPLICATION

DISSEVERATION

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

by

Ryuta Takeda

Graduate Program in Molecular, Cellular and Developmental Biology

The Ohio State University

2011

Dissertation Committee:
Biao Ding, Ph.D., Adviser
Juan D. Alfonzo, Ph.D.
Venkat Gopalan, Ph.D.
Daniel R. Schoenberg, Ph.D.
ABSTRACT

Increasing evidence shows that RNA intercellular trafficking plays a fundamental role for gene regulation at the whole organism level, viral infection, and anti-viral response in plants. Despite the essential and versatile functions of RNA trafficking, the underlying mechanism for this phenomenon has not yet been fully understood. As a possible mechanism, our laboratory hypothesized that unique RNA structural motifs mediate trafficking across different cellular layers. To test this postulate, we used a model Potato spindle tuber viroid (PSTVd) RNA. Viroids are small, circular, non-coding RNAs which infect plants. Since they do not encode any protein, viroids serve as an excellent model to investigate the direct role of RNA in intercellular trafficking. Among all viroids, PSTVd serves as the best specimen to investigate the structure-function relationships of RNA in vivo because of its well understood secondary (two-dimensional, 2-D) structure and infection process.

In this thesis, identification and functional analyses of PSTVd structures required for the intercellular trafficking are described. First, our study identified a three-dimensional (3-D) structural motif, called loop 6, required for PSTVd trafficking from palisade mesophyll to spongy mesophyll in Nicotiana benthamiana leaves. We present a 3-D structural model of loop 6 that specifies all non-Watson-Crick base pair interactions,
derived by isostericity-based sequence comparisons with 3-D RNA motifs in the database. This model is supported by available chemical modification patterns, natural sequence conservation/variations in PSTVd isolates and related species, and functional characterization of all possible mutants for each of the loop 6 base pairs. Second, through the analyses of trafficking-defective mutants of PSTVd, we observed rapid in vivo evolution of an RNA loop required for systemic trafficking of PSTVd in N. benthamiana. These trafficking-defective mutants were generated by mutating a wild type PSTVd 2-D structure called loop 19. The new variations that evolved from this mutant acquired a de novo structure, which we termed loop 19*, similar to loop 19 and in its one base-pair shifted genomic location. Genetic analyses showed that loop 19* mutants are trafficking-competent. To gain insights into loop 19 structure-function relationships, a series of loop 19 and loop 19* mutants were further analyzed. Taken together, these studies not only support our hypothesis of RNA structural motif mediated intercellular trafficking, but also demonstrate PSTVd as an excellent model to investigate RNA intercellular trafficking.

We also investigated the mechanism of PSTVd intra-molecular ligation, which is critical to accomplish viroid replication as well as those of other subviral agents; however, the RNA ligase required for this process has not yet been identified for any of them. Though we could not identify the RNA ligase for PSTVd in this study, our simple assay system and observed strong ligation activity demonstrated the potential of PSTVd as an experimental model to investigate this biological process.

iii
Dedicated to my grandmother and grandaunt
ACKNOWLEDGEMENTS

This thesis would not be possible without ample support from the people who shared time with me during my years of research in the Ding laboratory.

First, I express my sincere gratitude to my advisor Dr. Biao Ding for his continuous support of my Ph.D. study. His scientific enthusiasm and inspiration, continuous encouragement and patience, and his invaluable advice have lifted me up to this stage. I could not think of a better mentor for a graduate student. The knowledge, wisdom, and, most importantly, patience I learned from him during my Ph.D. study will definitely benefit my future career.

I thank Dr. Neocles Leontis and his graduate students Anton Petrov and Jesse Stombaugh at Bowling Green State University for our wonderful collaboration. Their crucial contribution to three-dimensional structural analyses on PSTVd RNA helped in the formulation of my research described in Chapter 2. Figures 2.2, 2.3, 2.4, 2.5, and 2.6, Tables 2.1, 2.2, and 2.3, and Appendices A and B were kindly provided by them.

I am grateful to my advisory committee members, Drs. Juan Alfonzo, Venkat Gopalan, and Daniel Schoenberg for their invaluable input on my research. I was always delighted and inspired by the discussions we had, and also motivated by their high expectations. I am really honored to have had all of you as my advisory committee
members. I would especially like to thank Dr. Gopalan for his guidance on the FPLC system and column chromatography.

I would like to thank the fellow Ding laboratory members: Dr. Farah Khan, Dr. Woong June Park, Yuan Tian, Dr. Ying Wang, and Xiaorui Yang for their continuous support and invaluable suggestions regarding my research and thesis manuscript. I would especially like to thank Dr. Ying Wang for his continuous encouragement, technical suggestions, and consistent advice on experiment design. I am grateful to all the undergraduate students in the Ding laboratory for their conducting lab maintenance and experiments for my projects. Especially, I thank Elizabeth Luscher and Casey Smith for the RNA extractions they performed. Also, I am indebted to the Ding lab alumni: Drs. Asuka Itaya, Xiaorong Tao, and Xuehua Zhong. Their patience, support, and training in my initial years in the laboratory provided me with a solid basis to conduct independent research. I am thankful to Carter Mason, a rotation student in the laboratory, for his suggestions on this thesis manuscript.

I would like to thank following researchers in The Ohio State University: The Gopalan laboratory members, Drs. Pradip Biswas and Lien Lai, for their guidance in operating the FPLC system, and Dr. Mark Foster for his insights into PSTVd structure. I also would like to thank all the members of the Plant Biotechnology Center for their suggestions and for their sharing of chemicals and equipment.

I am also thankful to the staff in the Plant Biotechnology Center − Diane Furtney, Melinda Parker, Scott Hines, David Long, and Joe Takayama − for their efforts in maintaining a great working environment. I would like to thank Jan Zinach, the program
manager in the Molecular, Cellular and Developmental Biology Program, for her continuous support.

I would like to thank my friend, Dr. Eric Lehoux, for sharing his knowledge about protein purification.

I would like to thank the Ding family, Yan Xun, Arthur, and Adeline for their support and hospitality throughout my Ph.D. study.

Last, but not certainly least, I would like to thank my parents and my fiancée, Yuko Uejima, for their spiritual support throughout my journey toward my Ph.D. degree.
VITA

December 2, 1980..........................Born in Tokyo, Japan

2004.................................B.A. International Christian University

2004 to present .........................Graduate Teaching and Research Associate,
                               Graduate Program in Molecular, Cellular and
                               Developmental Biology, The Ohio State
                               University

PUBLICATIONS

motif in *Potato spindle tuber viroid* mediates trafficking from palisade mesophyll

factors and broad impacts. Viruses *1,* 210-221.

C., Nelson, R.S., and Ding, B. (2007). A structured viroid RNA serves as a
substrate for dicer-like cleavage to produce biologically active small RNAs but is
resistant to RNA-induced silencing complex-mediated degradation. J. Virol. *81,*
2980-2994.
FIELDS OF STUDY

Major Field: Molecular, Cellular and Developmental Biology
CHAPTER 1 INTRODUCTION: VIROID AS AN EXCELLENT MODEL SYSTEM TO ELUCIDATE RNA MOTIFS FOR RNA INTERCELLULAR TRAFFICKING............................................................................. 1

1.1 Importance of intercellular RNA trafficking...................................................... 2

1.2 Intercellular RNA trafficking pathways in plants.............................................. 3
CHAPTER 2  A THREE-DIMENSIONAL RNA MOTIF IN POTATO SPINDLE TUBER VIROID MEDIATES TRAFFICKING FROM PALISADE MESOPHYLL TO SPONGY MESOPHYLL IN NICOTIANA BENTHAMIANA ................................................................. 15

2.1 ABSTRACT ........................................................................................................ 16

2.2 INTRODUCTION ............................................................................................... 17

2.3 MATERIALS AND METHODS ............................................................................ 22

2.3.1 cDNA construction of PSTVd loop 6 mutants .............................................. 22

2.3.2 Plant growth and infection ........................................................................... 23

2.3.3 In vitro transcription ................................................................................... 23

2.3.4 RNA extraction and RNA gel blotting ......................................................... 24

2.3.5 RT-PCR ........................................................................................................ 24

2.3.6 Sequencing of PSTVd progenies .................................................................. 25

2.3.7 Tissue processing and in situ hybridization ............................................... 25
2.3.8 Quantitative analysis of PSTVd-accumulating cells .......................... 26

2.4 RESULTS AND DISCUSSION........................................................................ 27

  2.4.1 A three-dimensional structural model for PSTVd loop 6............... 27

  2.4.2 The proposed PSTVd loop 6 structural model is compatible with chemical probing data................................................................................................. 31

  2.4.3 Pospiviroid and 23S rRNA sequence alignment supports the PSTVd loop 6 structural model.................................................................................................................. 32

  2.4.4 Mutational analysis provided genetic evidence in support of the PSTVd loop 6 structural model.................................................................................................................. 34

  2.4.5 All systemic infection-defective mutants of loop 6 fail to exit the inoculated leaves ............................................................................................................................... 37

  2.4.6 The failure of some loop 6 mutants to establish systemic infection is not due to deficiency in replication .......................................................................................... 38

  2.4.7 Loop 6 is required for trafficking from palisade to spongy mesophyll.... 40

CHAPTER 3 RAPID IN VIVO EVOLUTION OF AN RNA LOOP FOR SYSTEMIC TRAFFICKING OF POTATO SPIDLE TUBER VIROID.................. 71

  3.1 ABSTRACT ................................................................................................. 72

  3.2 INTRODUCTION ....................................................................................... 72

  3.3 MATERIALS AND METHODS .................................................................. 75
3.3.1 cDNA construction of PSTVd mutants ............................................................... 75
3.3.2 Plant growth and infection ................................................................................. 76
3.3.3 In vitro transcription .......................................................................................... 76
3.3.4 RNA extraction and RNA gel blotting .............................................................. 77
3.3.5 Sequencing of PSTVd progenies ....................................................................... 77
3.4 RESULTS ................................................................................................................ 78
  3.4.1 Rapid in vivo evolution of an RNA loop from loop 19-closed PSTVd mutant ............................................................... 78
  3.4.2 Trafficking-competency of loop 19* mutants ................................................. 80
  3.4.3 Systemic trafficking analyses of loop 19 mutants ............................................. 81
  3.4.4 Replication analyses of loop 19 mutants ......................................................... 85
3.5 DISCUSSION .......................................................................................................... 87
  3.5.1 A working model: loop 19/loop 19* as critical loops for PSTVd systemic trafficking ........................................................................................................... 87
  3.5.2 General conservation of loop 19 secondary structure in PSTVd natural variants .................................................................................................................. 88
  3.5.3 Loop 19 and loop 19* as a region of high nucleotide variation ................. 89

CHAPTER 4 CHARACTERIZATION OF A PLANT RNA LIGASE ESSENTIAL FOR VIROID REPLICATION .............................................................................. 106
4.1 ABSTRACT .................................................................................................................. 107

4.2 INTRODUCTION ......................................................................................................... 107

4.3 MATERIALS AND METHODS ............................................................................... 111

4.3.1 Preparation of a unit-length PSTVd substrate ................................................. 111

4.3.2 PSTVd ligation assay ......................................................................................... 111

4.3.3 Adenyltransferase assay ..................................................................................... 112

4.3.4 Preparation for the cauliflower protein extraction ........................................... 113

4.3.5 Large scale purification protocols of PSTVd ligase from cauliflower flower buds ........................................................................................................ 114

4.3.6 Mass spectrometry analyses in Midwest Bio Services .................................. 117

4.4 RESULTS ..................................................................................................................... 118

4.4.1 Optimization of the protein extraction buffer condition ............................. 118

4.4.2 Purification of PSTVd ligase by multi-step column chromatography ... 119

4.4.3 A 130 kDa band in tRNA fractions that possess PSTVd ligase activity was adenylated .............................................................................................................. 120

4.4.4 Failure of detecting a PSTVd ligase candidate in tRNA fractions by mass spectrometry analyses ............................................................. 121

4.5 DISCUSSION ............................................................................................................. 122

4.5.1 Revisiting the purification strategy ................................................................. 122

xiv
4.5.1 The possibility of the plant tRNA ligase as a PSTVd ligase .............. 124

4.5.2 Structures of 1-PSTVd 5' and 3' termini as substrates for the ligase ..... 125

CHAPTER 5 FUTURE PROSPECTS ................................................................. 136

5.1 SUMMARY ............................................................................................ 137

5.2 FUTURE PROSPECTS ......................................................................... 138

5.2.1 Cellular factors to mediate PSTVd intercellular trafficking ............ 138

5.2.2 Tertiary structure of PSTVd functional motifs .............................. 139

5.2.3 Evolution of RNA motifs for intercellular trafficking .................... 141

5.2.4 Cataloguing RNA motifs for intercellular trafficking .................... 142

REFERENCES ............................................................................................ 144

APPENDIX A RNA THREE-DIMENSIONAL MOTIFS USED FOR MODELING OF
PSTVD LOOP 6 ......................................................................................... 156

APPENDIX B THE ALIGNMENT OF PSTVD NATURAL VARIANTS SHOWING
CONSERVATION OF LOOP 6 SEQUENCE ............................................. 158

APPENDIX C MAFFT ALIGNMENT OF POSPIVROID REFSEQ SEQUENCES
IN CLUSTAL FORMAT ............................................................................... 176
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Three-dimensional (3-D) motifs in crystal structures of rRNAs used to model three-dimensional structure of PSTVd loop 6</td>
<td>47</td>
</tr>
<tr>
<td>2.2 Sequence variants of loop 6 based on multiple sequence alignment of eight pospiviroid species</td>
<td>48</td>
</tr>
<tr>
<td>2.3 Comparison of sequence variability observed for H89 motif in 23S bacterial rRNA alignments with the sequence of PSTVd loop 6</td>
<td>49</td>
</tr>
<tr>
<td>2.4 Sequencing of PSTVd loop 6 mutant progenies in systemically infected plants</td>
<td>50</td>
</tr>
<tr>
<td>3.1 Sequencing of C227U progenies from systemic leaves of infected plants</td>
<td>91</td>
</tr>
<tr>
<td>3.2 Sequencing of loop 19* mutant progenies from systemic leaves of infected plants</td>
<td>92</td>
</tr>
<tr>
<td>3.3 Sequencing of trafficking-competent loop 19 mutant progenies from systemic leaves of infected plants</td>
<td>93</td>
</tr>
<tr>
<td>3.4 Sequencing of trafficking-alteration loop 19 mutant progenies from systemic leaves of infected plants</td>
<td>94</td>
</tr>
</tbody>
</table>
3.5 Sequence variants of loop 19/loop 19* and its equivalent regions in PSTVd natural variations………………………………………………………………………………95

4.1 Purification fold of PSTVd ligase………………………………………………127

4.2 The identified proteins from LC/MS/MS analyses………………………………128
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>A simplified view of cell-to-cell and long-distance intercellular RNA trafficking pathways in a plant</td>
<td>10</td>
</tr>
<tr>
<td>1.2</td>
<td>PSTVd 2-D structure and previously identified trafficking RNA motifs and regions</td>
<td>12</td>
</tr>
<tr>
<td>1.3</td>
<td>Asymmetric rolling circle replication model of Pospiviroidae</td>
<td>14</td>
</tr>
<tr>
<td>2.1</td>
<td>Secondary structure of PSTVd</td>
<td>51</td>
</tr>
<tr>
<td>2.2</td>
<td>RNA nucleobase interactions</td>
<td>52</td>
</tr>
<tr>
<td>2.3</td>
<td>Proposed non-Watson-Crick base pairing for PSTVd loop 6 based on the crystal structure of a 23S rRNA internal loop as template</td>
<td>53</td>
</tr>
<tr>
<td>2.4</td>
<td>Docking site for Lys$_{123}$ from L16 ribosomal protein created by the cytosine-cytosine bifurcated pair</td>
<td>54</td>
</tr>
<tr>
<td>2.5</td>
<td>Chemical probing data in support of the proposed three-dimensional structure of PSTVd loop 6</td>
<td>55</td>
</tr>
<tr>
<td>2.6</td>
<td>Pospiviroid sequence alignments provide support to the PSTVd loop 6 structural model</td>
<td>56</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>2.7</td>
<td>Systemic infection analyses on PSTVd loop 6 mutants</td>
<td>58</td>
</tr>
<tr>
<td>2.8</td>
<td>Representative RNA gel blots showing systemic infection of PSTVd loop 6 mutants</td>
<td>60</td>
</tr>
<tr>
<td>2.9</td>
<td>Detection of wild type and mutants of PSTVd loop 6 in petioles of inoculated leaves by RT-PCR</td>
<td>61</td>
</tr>
<tr>
<td>2.10</td>
<td>Representative images of whole mount in situ hybridization to assay replication of loop 6 mutants</td>
<td>62</td>
</tr>
<tr>
<td>2.11</td>
<td>Infection efficiencies of PSTVd loop 6 mutants as percentages of wild type (set to 100%) as determined by whole mount in situ hybridization</td>
<td>63</td>
</tr>
<tr>
<td>2.12</td>
<td><em>In situ</em> hybridization localization of wild type and A325G mutant PSTVd in inoculated leaves of <em>N. Benthamiana</em></td>
<td>64</td>
</tr>
<tr>
<td>2.13</td>
<td><em>In situ</em> hybridization on 14-μm paraffin sections of <em>N. benthamiana</em> leaves inoculated with PSTVd or water</td>
<td>66</td>
</tr>
<tr>
<td>2.14</td>
<td><em>In situ</em> hybridization 14-μm paraffin sections of <em>N. benthamiana</em> leaves inoculated with A37U mutant</td>
<td>68</td>
</tr>
<tr>
<td>2.15</td>
<td>Comparison of infection efficiencies of WT, A325G and A37U in leaf epidermis of inoculated <em>N. benthamiana</em></td>
<td>70</td>
</tr>
<tr>
<td>3.1</td>
<td>Secondary structures of PSTVd loop 19 and its mutants</td>
<td>96</td>
</tr>
<tr>
<td>3.2</td>
<td>Systemic trafficking analyses on C227U mutant</td>
<td>98</td>
</tr>
<tr>
<td>3.3</td>
<td>Systemic trafficking analyses on loop 19* mutants</td>
<td>99</td>
</tr>
<tr>
<td>3.4</td>
<td>Secondary structures of loop 19/loop19* variations observed in the systemic trafficking analyses</td>
<td>100</td>
</tr>
</tbody>
</table>
3.5 Systemic trafficking analyses on PSTVd loop 19 mutants
3.6 RNA gel blots for systemic trafficking analyses on PSTVd loop 19 mutants
3.7 Replication efficiencies of loop 19 mutants
3.8 c-PSTVd signal on the inoculated leaves
4.1 A schematic presentation of the shared three step enzymatic reaction of RNA ligases
4.2 A schematic presentation of the PSTVd ligase assay and the quantification of the ligation efficiency
4.3 PSTVd ligation assays to optimize the condition of the PSTVd ligase extraction
4.4 Purification of PSTVd ligase
4.5 Adenyltransferase activity of a 130 kDa band in the tRNA fractions having the PSTVd ligase activity
4.6 The protein profile of a tRNA fraction analyzed in LC/MS/MS analyses
4.7 Detection of a band migrating more slowly than c-PSTVd
ABBREVIATIONS

AMP, adenosine monophosphate
AS, ammonium sulfate
ATP, adenosine-5’-triphosphate
AtAGO1, Arabidopsis thaliana ARGONAUTE 1
AtAGO5, Arabidopsis thaliana ARGONAUTE 5
AtTrl, Arabidopsis thaliana tRNA ligase
CBB, Coomassie Brilliant Blue
C, (+)-circular PSTVd
c-PSTVd, (+)-circular PSTVd
CE, crude protein extract
CEVd, Citrus exocortis viroid
CSVd, Chrysanthemum stunt viroid
cWC, cis-Watson-Crick
cWC bif, cis Watson-Crick bifurcated
DIG, digoxigenin
DMS, dimethylsulfide
dpi, days post-inoculation
DTT, dithiothreitol
EB, extraction buffer
EIF3-gamma, initiation factor 3 gamma
ER, endoplasmic reticulum
FR3-D, Find RNA 3-D
FT, flow through
GFP, green fluorescent protein
H, Hoogsteen
HDV, hepatitis delta virus
HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
h41, helix 41
H89, helix 89
IP, input
L, unit length linear PSTVd
1-PSTVd, unit length linear PSTVd
LC, liquid chromatography
M, marker
miRNA, microRNAs
MS/MS, tandem mass spectrometry
NMR, nuclear magnetic resonance
non-WC, non-Watson-Crick
PAGE, polyacrylamide gel electrophoresis
PB, purification buffer
PB50, purification buffer containing 50 mM NaCl
PD, plasmodesmata
PDB, Protein Database
PM, plasma membrane
PMSF, phenylmethanesulfonylfuoride
Pol II, DNA-dependent RNA polymerase II
pRZ6-2-Int, Plasmid pRZ6-2 containing the cDNA of PSTVd<sup>int</sup>
PSTVd, <i>Potato spindle tuber viroid</i>
PSTVd<sup>int</sup>, Intermediate strain of PSTVd (treated as wild type PSTVd in all analyses)
rRNAs, ribosomal RNAs
S, Sugar
ScTrl1, <i>Saccharomyces cerevisiae</i> tRNA ligase
SDS, sodium dodecyl sulfate
siRNAs, short interfering RNAs
tHS, <i>trans</i> Hoogsteen-Sugar
Tris, tris(hydroxymethyl)aminomethane
tSH, <i>trans</i> Sugar-Hoogsteen
T4Rnl1, T4 RNA ligase 1
UTR, untranslated region
W, wash
WC, Watson-Crick
WT, wild type
2'-PO<sub>4</sub>, 2'-phosphate
2-D, two-dimensional or secondary
2'3'-cyclic PO<sub>4</sub>, 2', 3'-cyclic phosphate
3'-OH, 3'-hydroxyl
3-D, three-dimensional
5'-PO₄, 5'-phosphate
CHAPTER 1

INTRODUCTION:

VIROID AS AN EXCELLENT MODEL SYSTEM

TO ELUCIDATE RNA MOTIFS FOR RNA INTERCELLULAR TRAFFICKING
1.1 Importance of intercellular RNA trafficking

The careful interchange of molecules between different cells, tissues, and organs is vital for multi-cellular organisms orchestrating their system at the whole-organism level. For the coordination of a multi-cellular-system in the processes of development, adaptation to environmental change, and defense against pathogens, specific molecules generated in a certain cell are transported elsewhere. Recent studies in plants have showed that various RNAs traffic to and regulate gene expression in cellular locations different from their site of biogenesis (Examples in detail are described in Section 2.2) (Giakountis and Coupland, 2008; Kehr and Buhtz, 2008; Lucas et al., 2009; Turgeon and Wolf, 2009; Chitwood and Timmermans, 2010; Chuck and O'Connor, 2010; Hannapel, 2010; Lehesranta et al., 2010). These non-cell-autonomous gene regulations of cellular RNAs as well as the long recognized systemic trafficking phenomena of RNA-based pathogens, such as viruses and viroids (Flores et al., 2005; Scholthof, 2005; Lucas, 2006; Ding and Itaya, 2007; Taliansky et al., 2008; Ding, 2009), expand our understanding of gene regulation from the single cell to the whole organism level. To further understand the intercellular trafficking phenomena, the following unsolved fundamental mechanistic questions need to be addressed: 1) How is a specific RNA recognized and/or how does it play an active role in trafficking? and 2) How is a trafficking destination of a specific RNA determined? To answer these questions, it is critical to identify and investigate RNA motif(s) regulating intercellular RNA trafficking. In this chapter, I briefly review
general RNA trafficking pathways in a plant body and introduce viroids as an excellent experimental model to investigate RNA motifs for intercellular trafficking.

1.2 Intercellular RNA trafficking pathways in plants

Plants, having less mobility compared to animals, need to develop elaborate pathways to communicate between different cells, tissues and organs (a simplified view is presented in Figure 1.1). Plasmodesmata (PD) are plant specific micro-channels, which span through the cell membrane and cell wall between neighboring cells (Figure 1.1B) (Maule, 2008; Lucas, 2009; Benitez-Alfonso et al., 2010). Inside of PD, a modified endoplasmic reticulum (ER) spans through each pore and plasma membranes. Proteins are embedded in PD, and may play structural and/or regulatory roles to maintain this micro-channel. A variety of macromolecules traffic cell-to-cell through PD: such as proteins and RNAs as well as photo-assimilates (Lough and Lucas, 2006; Kehr and Buhtz, 2008).

In addition to this local cell-to-cell molecular exchange, a higher plant body develops vascular tissue, phloem, for long-distance trafficking. Given that there is no detectable RNase activity in phloem sap, the phloem conduit is considered an ideal route for RNAs to be transported (Sasaki et al., 1998; Doering-Saad et al., 2002). In support of this claim, numerous RNAs have been detected in phloem sap, indicating that they traffic in phloem (Sasaki et al., 1998; Ruiz-Medrano et al., 1999; Yoo et al., 2004; Doering-Saad et al., 2006; Omid et al., 2007; Deeken et al., 2008; Buhtz et al., 2008; Buhtz et al., 2010;
Varkonyi-Gasic et al., 2010). Sieve tube, a conducting cell type in phloem tissue, interconnects end-to-end to form long transport channels (Turgeon and Wolf, 2009); therefore, molecules in sieve tubes can traffic easily in long distance as bulk flow from source to sink organs (Figure 1.1A, red broken arrows).

1.3 RNA motif mediated RNA intercellular trafficking

Increasing evidence shows that RNA intercellular trafficking is regulated. For example, the genomic RNA of *Potato spindle tuber viroid* (PSTVd) is selectively transported to the sepal but not to other floral organs in *Nicotiana benthamiana* and in tomato (Zhu et al., 2001; 2002). As another example, Qi and his colleagues showed that four nucleotides distantly located in a PSTVd are required in the trafficking from bundle sheath to mesophyll cells in tobacco (Qi et al., 2004). Furthermore, Zhong and her colleagues showed that the three-dimensional (3-D) structure of the specific two nucleotide-loop motif in PSTVd is required for PSTVd phloem entry in *N. benthamiana* (Zhong et al., 2007). These results have led our laboratory to hypothesize that a unique RNA structural motif mediate trafficking across different cellular layers. Each cellular boundary and RNA motif(s) required to pass through each boundary can be imagined as a different doors and different key(s), respectively. Are there any master keys? Or do different doors always need different keys? If so, which key is required for which door?

To further test the above hypothesis, viroids provide an excellent model to study RNA structural motifs for intercellular RNA trafficking. Below, I briefly review the
biological, chemical and structural properties of a viroid species, PSTVd, in relation to its excellence as a model system.

1.4 General properties of viroids

Viroids are single-stranded, circular, and infectious RNAs in plants that have genome sizes ranging from about 250 to 400 nucleotides (Tabler and Tsagris, 2004; Flores et al., 2005; Ding, 2009). Without encoding any protein and without encapsidation, these tiny naked RNAs must usurp the host plant machinery to expand their area of territory. The non-coding and non-encapsidation nature of viroids distinguish them from viruses. The first identification of viroid as a disease-causative small RNA was in 1971 (Diener, 1971). Since then, more than 40 viroid species and 1,700 unique sequences have been identified and registered in the Subviral RNA Database (http://subviral.med.uottawa.ca) (Rocheleau and Pelchat, 2006). The symptoms of viroid infection vary from symptomless, mild to severe stunting, deformation of various organs and tissues, and even lead to plant death (Randles, 2003; Singh et al., 2003). Viroid infection in some crop and ornamental plants causes severe economic losses, as has been the case in potatoes, tomatoes, hops, coconuts, and chrysanthemums (Randles, 2003).

All viroids currently known to date are classified into two families: Pospiviroidae and Avsunviroidae, based on their sequences, secondary structures and biological properties (Flores et al., 2005). Briefly, the members of Avsunviroidae have a quasi-rod-like secondary (two-dimensional, 2-D) structure with branching, possess intrinsic
ribozyme activities and replicate in the chloroplast. The members of Pospiviroidae, typified by PSTVd, have a rod-like 2-D structure (Figure 1.2A), lack evident ribozyme activities and replicate in the nucleus (Woo et al., 1999; Qi and Ding, 2003).

1.5 Infection processes of Pospiviroidae

Viroids in the Pospiviroidae family need to accomplish many processes in their hosts to establish systemic infection: 1) entry into the plant body, 2) replication in the nucleus, 3) local cell-to-cell trafficking, 4) entry into phloem, 5) long-distance trafficking in phloem, 6) exit from phloem, and 7) local cell-to-cell trafficking upon exiting from phloem to the new region of the plant (Figure 1.1A) (Ding, 2009; Takeda and Ding, 2009; Wang and Ding 2010). Overall, viroid trafficking follows the general RNA trafficking pathway described above. The local cell-to-cell trafficking of a viroid species, PSTVd, occurs through PD (Ding et al., 1997), and the long-distance trafficking occurs through phloem (Palukaitis, 1987 and Zhu et al., 2001). In addition, the replication process makes PSTVd a unique trafficking RNA. In Pospiviroidae, (+)-circular, unit length, RNA is transcribed to a (-)-strand multimer in linear form via an asymmetric rolling circle mechanism (Figure 1.3) (Branch and Robertson, 1984). Then, the (-)-strand multimer is used as a template to generate multimeric (+)-strand linear RNA. These processes are catalyzed by a host DNA-dependent RNA polymerase II (Mühlbach and Sänger 1979; Rackwitz et al., 1981; Schindler and Mühlbach, 1992). Finally, the multimeric (+)-strand RNA is cleaved and circularized into unit-length (+)-circular RNA.
The non-coding nature of viroids suggests that they utilize their plant endonuclease and RNA ligase for their own processing; however, the specific enzymes responsible for these processes have not yet been identified in the case of Pospiviroidae.

1.6 PSTVd as an excellent model system to investigate RNA motifs required for RNA intercellular trafficking

To investigate/identify both known and unknown RNA motifs, detailed mutational analyses on PSTVd genomic RNA are feasible and informative for the following reasons. First, the availability of a well-supported rod-like 2-D structure of PSTVd genomic RNA makes it feasible to identify unknown RNA motifs within the genome and to perform detailed analysis of its structure-function relationships. The rod-like 2-D structure has been supported by microscopic (Sogo et al., 1973), biophysical, (Riesner et al., 1979), chemical/enzymatic probing (Gast et al., 1996), and mutational studies (Zhong et al, 2008). A nuclear magnetic resonance (NMR) study of the PSTVd left terminal region also supports the rod-like 2-D structure of this region (Dingley et al., 2003).

Second, the small genome size of PSTVd makes the genome-wide mutational analyses feasible. Various 2-D structural features exist in the PSTVd genome whose 3-D structure and biological functions remain unknown. These properties can be fully investigated by analyzing PSTVd mutants for the intended structures. Third, the general intercellular trafficking pathway of PSTVd has been well understood; therefore,
trafficking mutants can easily be screened by checking the existence of each mutant in
different cellular locations. Fourth, the high accumulation level of PSTVd in host plant
nuclei makes the easy detection of its cellular localization possible. Fifth, without
encoding protein, PSTVd must interact and usurp existing cellular machinery, suggesting
that its RNA motifs are also functionally important in host cellular RNAs. Therefore, it
may be possible to elucidate the trafficking mechanism of cellular RNAs, and induce the
general principle of trafficking shared in both exogenous and endogenous RNAs. In fact,
the Ding laboratory successfully identified several RNA motifs/regions required for
PSTVd intercellular trafficking as follows.

As shown in Figure 1.2A, the 2-D structure of PSTVd consists of 27
loops/bulges (collectively referred to as 'loops' hereafter for simplicity of description)
flanked by short helices, indicating canonical cis-Watson-Crick (cWC) and/or cWC G/U
wobble pairs. X-ray crystallographic and NMR studies have shown that most loops in
RNA 2-D structures often form recurrent 3-D structures via non-Watson-Crick base
pairing, base stacking, and other base-base interactions (Leontis et al., 2006), and that
such structures often function in RNA-RNA, RNA–protein, and RNA-small ligand
interactions (Leontis et al., 2002a, b; Noller, 2005; Leontis et al., 2006; Steitz, 2008).
Thus, it is reasonable to postulate that PSTVd loops act as functional motifs. Indeed, 11
out of 27 loops were identified as necessary for PSTVd systemic trafficking in N.
benthamiana (Fig 1.2 highlighted by red color) (Zhong et al., 2008). This serves as a
solid starting point to further investigate the structure-function relationships of each loop.
Zhong and her colleagues further analyzed one of these loops, loop 7 (Figure 1.2B; U43/C318) (Zhong et al., 2007). Based upon comparative sequence analysis and comparison with X-ray crystal structures of similar motifs in ribosomal RNAs, loop 7 is modeled to be composed of a water-inserted cWC base pair. Specific mutations that disrupt the 3-D structure of this motif were not detected in phloem, suggesting that the trafficking incapability of loop 7-disrupted mutants to traverse the cellular boundary between bundle sheath and phloem cells. This provided an example of an RNA motif where 3-D structure governs PSTVd intercellular trafficking to traverse a specific cellular boundary.

Qi and his colleagues showed that 4 nucleotides (Figure 1.2C; U201, U309, U47/A313, collectively named as the bipartite motif) separated into two distant locations in the genome of PSTVd^{NB} strain are required for PSTVd trafficking from bundle sheath to mesophyll cell (Qi et al., 2004). These sequences/regions are not required for the opposite direction of trafficking, suggesting the uni-directionality of the function of these sequences/regions. The 3-D structure of these sequences and the requirement of 3-D structure in the above trafficking phenomenon have not yet been elucidated.

Thus, PSTVd serves as an excellent experimental system to identify RNA motifs/regions required for intercellular trafficking. In the following chapters, investigations of RNA structural motifs required for PSTVd intercellular trafficking are described as well as studies on the host RNA ligase required for PSTVd replication.
Figure 1.1 A simplified view of cell-to-cell and long-distance intercellular RNA trafficking pathways in a plant. A) The bulk flow of long-distance RNA trafficking from source leaf to sink leaf through phloem is shown as red broken lines. Cell-to-cell trafficking pathways are shown in the schematic of transverse leaf sections. As the black arrows show, RNA can traffic from one cell to another between the same type of cells and different type of cells. Once RNA enters the phloem (red arrow), it embarks on long-distance trafficking. Subsequently, RNA is unloaded from phloem (red arrow). Then, in a different location, it can conduct cell-to-cell trafficking. B) Two cells separated by cell wall with a plasmodesma that allows direct cell-to-cell trafficking. The RNA trafficking path through the plasmodesma is illustrated as a dashed line. ER: endoplasmic reticulum. PM: plasma membrane.
Figure 1.2 PSTVd secondary structure and previously identified trafficking RNA motifs and regions. A) PSTVd secondary (2-D) structure generated from mfold (Zucker et al., 2003) with 27 loops numbered from the left to the right. The loops required for trafficking are highlighted in red based on Zhong et al (2008). The nucleotide numbers of the PSTVd genome are highlighted for nucleotides C1, A50, A100, U179, U180, U250, U300, and U359. Positions of loop 7 and bipartite motifs are indicated. B) Loop 7 (U43/C318 highlighted in red), which comprises a U/C cWC base pair with water insertion and which is required for bundle sheath-to-phloem trafficking, based on Zhong et al (2007). C) The 2-D structure of the bipartite motif (U201, U309, and U47/313A highlighted in blue) in strain PSTVd_{NB} for bundle sheath-to-mesophyll trafficking, based on Qi et al (2004).
Figure 1.2
Figure 1.3 Asymmetric rolling circle replication model of Pospiviroidea. The simplified view of Pospiviroidea replication based on Branch and Robertson (1984) is shown. The proposed host enzyme responsible for each replication process is also provided. Pol II: DNA-dependent RNA polymerase II.
CHAPTER 2

A THREE-DIMENSIONAL RNA MOTIF
IN POTATO SPINDLE TUBER VIORID MEDIATES TRAFFICKING
FROM PALISADE MESOPHYLL TO SPONGY MESOPHYLL
IN NICOTIANA BENTHAMIANA

Data presented here were published as Takeda R., Petorov A.I., Leontis N.B., and Ding B. (2011). A three-dimensional RNA motif in Potato spindle tuber viroid mediates Trafficking from palisade mesophyll to spongy mesophyll in Nicotiana benthamiana. Plant Cell, In press. The data are presented with modifications with permissions from the American Society of Plant Biologists.
2.1 ABSTRACT

Cell-to-cell trafficking of RNA is an emerging biological principle that integrates systemic gene regulation, viral infection, antiviral response and cell-to-cell communication. A key mechanistic question is how an RNA is specifically selected for trafficking from one type of cell into another type. Here we report the identification of an RNA motif in *Potato spindle tuber viroid* (PSTVd) required for trafficking from palisade mesophyll to spongy mesophyll in *Nicotiana benthamiana* leaves. This motif, called loop 6, has the sequence 5'-CGA-3’...5’-GAC-3’ flanked on both sides by *cis* Watson-Crick (cWC) G/C and G/U wobble base pairs. We present a three-dimensional (3-D) structural model of loop 6 that specifies all non-WC base pair interactions, derived by isostericity-based sequence comparisons with 3-D RNA motifs from the RNA X-ray crystal structure database. This model is supported by available chemical modification patterns, natural sequence conservation/variations in PSTVd isolates and related species, and functional characterization of all possible mutants for each of the loop 6 base pairs. Our findings and approaches have broad implications for studying the 3-D RNA structural motifs mediating trafficking of diverse RNA species across specific cellular boundaries and for studying the structure-function relationships of RNA motifs in other biological processes.
2.2 INTRODUCTION

A central question in current biology concerns how basic processes in individual cells are integrated to support development and function of the whole multicellular organism. Cellular boundaries play a pivotal role in this integration by maintaining a certain level of cellular autonomy while enabling communication between cells to achieve coordinated gene expression and metabolism within an organism. Cell-to-cell trafficking of specific RNA and protein molecules is emerging as a new paradigm of gene regulation at the whole-organism level in plants, in addition to its fundamental role in the systemic spread of viral infection and defense responses (Giakountis and Coupland, 2008; Kehr and Buhtz, 2008; Lucas et al., 2009; Turgeon and Wolf, 2009; Chitwood and Timmermans, 2010; Chuck and O'Connor, 2010; Hannapel, 2010; Lehesranta et al., 2010). Such trafficking requires rethinking of RNAs and proteins as functioning solely within the cells in which they are synthesized.

Infectious agents, such as viroids and viruses, must move from cell to cell and throughout a plant to establish systemic infections. All viroids and many plant viruses have RNA genomes, so that their systemic infections represent classical examples of cell-to-cell and long-distance RNA trafficking (Flores et al., 2005; Scholthof, 2005; Lucas, 2006; Ding and Itaya, 2007; Ding, 2009). Many cellular RNAs also traffic between cells and between organs. Analyses of phloem sap collected from various plant species revealed many different mRNAs (Sasaki et al., 1998; Ruiz-Medrano et al., 1999; Doering-Saad et al., 2006; Omid et al., 2007; Deeken et al., 2008), small RNAs including
microRNAs (miRNAs) and short interfering RNAs (siRNAs) (Yoo et al., 2004; Buhtz et al., 2008; Buhtz et al., 2010; Varkonyi-Gasic et al., 2010) and some other types of RNAs (Zhang et al., 2009). Grafting experiments have shown that some mRNA species are transported over long distances in the phloem to regulate developmental processes such as leaf morphogenesis in tomato (Kim et al., 2001; Haywood et al., 2005) and tuber formation in potato (Banerjee et al., 2006). Trafficking of a miRNA from shoot to root has been implicated in the regulation of a gene involved in phosphate homeostasis in Arabidopsis thaliana (Lin et al., 2008; Pant et al., 2008). Intercellular trafficking of some miRNAs contributes to regulation of root vascular development (Carlsbecker et al., 2010). siRNAs can act as mobile signals that traffic from cell to cell and from organ to organ to mediate gene silencing including RNA-dependent DNA methylation (Chitwood et al., 2009; Schwab et al., 2009; Dunoyer et al., 2010a, b; Molnar et al., 2010). Gene silencing signals containing siRNAs also traffic within a plant to mediate systemic RNA silencing as a means of antiviral defense (Ding and Voinnet, 2007; Kalantidis et al., 2008). RNA trafficking also occurs between some parasitic plants and their hosts, with biological functions yet to be understood (Roney et al., 2007; David-Schwartz et al., 2008).

Cell-to-cell RNA trafficking also occurs in animals. Numerous circulating nucleic acids including RNAs have been found in human plasma and serum under healthy and diseased conditions (Fleischhacker and Schmidt, 2007; Vlassov et al., 2007). Gene silencing signals in animals traffic intercellularly as a means of gene regulation and antiviral defense (Ding and Voinnet, 2007; Jose and Hunter, 2007). Exosomes, which are
membrane vesicles released into extracellular spaces by many types of mammalian cells, enclose and mobilize mRNAs and miRNAs between different cells (Valadi et al., 2007; Kosaka et al., 2010). Viral miRNAs can also be transferred via exosomes between certain cells of the immune system and function therein (Pegtel et al., 2010). In addition to this secretory pathway, cell contact also enables intercellular transport of small RNAs between mammalian immune cells (Rechavi et al., 2009). These and other examples support the idea that intercellular RNA trafficking plays a role in signaling (Vlassov et al., 2007; Dinger et al., 2008; Simons and Raposo, 2009).

An outstanding question is how different RNAs are recognized for transport between distinct types of cells, which is essential for controlling the specificity of trafficking. We have been using Potato spindle tuber viroid (PSTVd) infection as a model system to test the hypothesis that unique RNA structural motifs mediate trafficking across different cellular boundaries. Viroids are single-stranded, circular and noncoding RNAs that infect plants. They are the smallest plant pathogens known to date, with sizes ranging from 250 to 400 nucleotides (Flores et al., 2005; Ding and Itaya, 2007; Tsagris et al., 2008; Ding, 2009). Lacking protein-coding capacity and helper viruses, viroid RNAs most likely interact directly with pre-existing cellular machineries to move from cell to cell and from organ to organ to establish systemic infections. Therefore, viroids provide simple models to investigate the RNA structural elements critical for cell-to-cell trafficking (Ding and Wang, 2009; Wang and Ding, 2010). The 359-nucleotide genome of PSTVd folds into a rod-like secondary structure that has been supported by biophysical, structural and mutational studies (Flores et al., 2005; Owens, 2007; Ding,
As shown in Figure 2.1, this secondary structure comprises 27 loops/bulges (collectively referred to as 'loops' hereafter, for simplicity of description), numbered 1 to 27, starting from the left in Figure 2.1, that are flanked by short double-stranded helices. With the PSTVd model, our previous studies obtained genetic evidence for the role of RNA loops in mediating RNA trafficking between specific cells (Qi et al., 2004; Zhong et al., 2007).

Computational analyses such as minimum free energy calculations (e.g., mfold) (Zuker, 2003) generally show RNA secondary structures as comprising short double helices punctuated by “loops.” In the absence of additional information, such loops are usually depicted as single-stranded regions lacking structure. However, many X-ray crystallographic and NMR studies in the last decade have provided evidence that nucleotides within most “loops” form distinct three-dimensional (3-D) motifs via non-Watson-Crick (non-WC) base pairing and base stacking, and that such motifs comprise the primary binding sites for RNA-RNA, RNA-protein and RNA-small ligand interactions (Leontis et al., 2002a, b; Noller, 2005; Leontis et al., 2006; Steitz, 2008). RNA nucleobases interact with each other by hydrogen-bonding at any one of three edges, the Watson-Crick (WC), Hoogsteen (H), and Sugar (S) Edges (Figure 2.2). Non-WC base pairs are categorized into 12 geometric base pairing families according to the interacting edges and relative orientation (cis or trans) of glycosidic bonds of the paired bases (Figure 2.2) (Leontis and Westhof, 2001).

Many RNA 3-D motifs recur in non-homologous RNA molecules or in distinct sites of the same RNA molecule (Simons and Grunberg-Manago, 1998; Leontis et al.,
Recurrent 3-D motifs comprise sets of nucleotides with similar spatial arrangements including the same non-WC base pairs. The 3-D structures of recurrent motifs are more conserved than their sequences; in other words, different RNA sequences can fold into the same 3-D structure by forming the same geometric non-WC base pairs (Leontis and Westhof, 1998a, b; Leontis et al., 2002b; Schudoma et al., 2010). The set of base substitutions compatible with the 3-D structure of a motif is its sequence signature. Thus, certain non-WC base pairs can substitute for each other in a motif without distorting the 3-D structure of the motif; such base pairs are isosteric or nearly so (Michel and Westhof, 1990; Leontis and Westhof, 1998b; Leontis et al., 2002b). More specifically, isosteric pairs form hydrogen bonds by using the same base edges and have the same glycosidic bond orientation and identical or similar C1’-C1’ (i.e., first carbon of ribose) distances. Isostericity matrices summarize the isosteric relationships for each geometric base pairing family and provide the basis for analyzing the sequence signatures of RNA motifs (Leontis et al., 2002b). Furthermore, isostericity matrices allow predictions of base substitutions in an RNA motif that will either disrupt or maintain the isostericity of a base pair, and consequently the structural integrity of the motif. We have used these principles to successfully predict, and then experimentally validate by mutational and functional analyses, the 3-D structures of PSTVd loop E, which is critical for replication (Zhong et al., 2006), and loop 7 which mediates trafficking from bundle sheath to phloem (Zhong et al., 2007).

We used mutational analysis to identify at least 11 loops in PSTVd critical for systemic trafficking in *Nicotiana benthamiana* plants (Zhong et al., 2007; Zhong et al.,
2008). The present study investigates loop 6, which comprises the six nucleotides G36, A37, C38, C323, G324, and A325 in strain PSTVd\textsuperscript{Int} (accession: NC_002030; Figure 2.1). When we mutated three nucleotides (G36U/A37C/C38G) to form WC base pairs and thereby 'closed' loop 6, as predicted by mfold (Zuker, 2003), systemic trafficking of the mutant PSTVd, but not its replication, was abolished (Zhong et al., 2008). To gain mechanistic insights into the function of loop 6, we investigated its 3-D structure (i.e., specific non-WC base pairing) as well as the cellular boundary at which it functions to mediate trafficking. Here we provide evidence that loop 6 plays a decisive role in PSTVd trafficking from palisade mesophyll to spongy mesophyll cells of \emph{N. benthamiana} leaves. We also present a 3-D structural model of loop 6 that specifies all non-WC base pair interactions. Our findings and approaches have broad implications for studying the 3-D RNA structural motifs mediating trafficking of diverse RNA species across cellular boundaries, and for studying the structure-function relationships of RNA motifs in other biological processes.

2.3 MATERIALS AND METHODS

2.3.1 cDNA construction of PSTVd loop 6 mutants

Plasmid pRZ6-2 containing the cDNA of PSTVd\textsuperscript{Int} (pRZ6-2-Int) was constructed by Hu and colleagues (Hu et al., 1997) and was a gift from Dr. Robert Owens. All PSTVd mutants described here were generated by site-directed mutagenesis based on this template, as previously described (Zhong et al., 2008). All introduced mutations were
verified by DNA sequencing with a 3730 DNA Analyzer (Applied Biosystems, Carlsbad, CA) at the Plant-Microbe Genomics Facility (PMGF) of The Ohio State University. Plasmid pInter (-) was constructed as previously described (Qi and Ding, 2002).

2.3.2 Plant growth and infection

*N. benthamiana* plants were grown in a growth chamber maintained at 16 h light/8 h dark (28°C) cycles. The *in vitro* transcripts of PSTVd mutants were mechanically inoculated onto the first two true leaves of two-week-old *N. benthamiana* plants (200 ng/leaf) dusted with carborundum powder. Diethylpyrocarbonate (DEPC)-treated H2O was used for mock inoculation. All inocula were applied exclusively to the upper surfaces of leaves.

2.3.3 In vitro transcription

To prepare *in vitro* transcripts of WT and mutant PSTVd for systemic infection and replication experiments, *Hind*III-linearized pRZ6-2-Int and all mutant derivatives were transcribed by using T7 Megascript kit according to the instructions provided by the manufacturer (Ambion, Austin, TX). Antisense PSTVd riboprobes for *in situ* hybridization and RNA gel blots were prepared by *in vitro* transcription of *SpeI*-linearized pInter (-) plasmid using T7 Maxiscript kit (Ambion). [α-32P] UTP-labeled riboprobes were used for RNA gel blots and digoxigenin (DIG)-UTP labeled riboprobes
were used for in situ hybridization. All RNA transcripts were purified by MEGAClear kit (Ambion) after DNase I digestion.

2.3.4 RNA extraction and RNA gel blotting

Total RNA from leaves of PSTVd- or mock-inoculated N. benthamiana plants was extracted by Ribozol reagent (Amresco, Solon, OH) by following the manufacturer’s instructions. RNA gel blotting was performed essentially as described in Zhong et al (Zhong et al., 2006).

2.3.5 RT-PCR

Total RNA from petioles of PSTVd- or mock-inoculated N. benthamiana leaves was extracted by using Ribozol reagent. The extracted RNA was purified by RNeasy mini (Qiagen, Valencia, California) followed by “RNA clean up” by using the manufacturer’s protocol. A total of 500 ng of pooled RNA was reverse-transcribed by using Superscript III (Invitrogen, Carlsbad, CA). The synthesized cDNAs were amplified with Taq polymerase by using PSTVd-specific forward primer (5’-CGGAACCTAAACTCGTGGTTCCT-3’) and reverse primer (5’-AGGAACCAACTGCGGTTCCA-3’). As an internal control, actin mRNA was RT-PCR amplified by using actin-specific forward primer (5’-CACCATGGCAGATGGAGGATATTCCAGCC -3’) and reverse primer (5’-
AGGAACCAACTGCGGTTCCA -3'). The PCR products were run on 1.2% (w/v) agarose gel and visualized with ethidium bromide staining.

2.3.6 Sequencing of PSTVd progenies

The PSTVd progenies were sequenced in a manner similar to that described in Qi and Ding (2002). Briefly, after the PSTVd signal detection either by RNA gel blotting or RT-PCR, the PSTVd RNA was RT-PCR amplified as described above except using Pfu Turbo DNA polymerase (Agilent Technologies, Santa Clara CA) instead of Taq polymerase. Next, the Pfu-amplified product was incubated with Taq polymerase for 20 min at 72 °C and ligated into pGEM-T vector (Promega, Madison, WI). The inserted DNA was sequenced as described above.

2.3.7 Tissue processing and in situ hybridization

The tissue fixation and processing followed the protocols previously described (Zhu et al., 2001; Qi et al., 2004) with slight modifications. Briefly, 8 days post-inoculation (dpi)-leaf samples were fixed in a solution of 50% (v/v) ethanol/5% (v/v) formaldehyde/5% (v/v) acetic acid for 30 min and then dehydrated by a step-wise gradient of ethanol solutions from 50 to 100% (v/v). The samples for whole mount in situ hybridization were washed by phosphate buffer saline and treated with 10 µg/ml of proteinase K for 20 min at 37°C. Then, the samples were hybridized with DIG-labeled PSTVd antisense riboprobes at 55°C overnight. The samples were washed and incubated
with alkaline phosphatase-conjugated DIG antibodies (Roche, Basel, Switzerland) followed by visualization with NBT/BCIP (Roche) color reaction.

To obtain thin sections, fixed and dehydrated samples were infiltrated with paraffin and embedded in paraffin blocks. Sections of 14 μm in thickness were obtained with a HM 340 E rotary microtome (Microm International GmbH, Walldorf, Germany). PSTVd accumulation was detected by in situ hybridization followed by the color reaction as describe above.

In situ hybridized-samples were examined and photographed under a Nikon Eclipse E600 light microscope (Nikon, Tokyo, Japan). Images were captured and processed with a “SPOT RT KE Slider” camera and the associated software, “SPOT Advanced” (Diagnostics instruments, Sterling Heights, MI). The size bar in each picture was stamped on each image after magnification calibration by using “SPOT Advanced,” and cut and pasted to the intended location of the same picture, if needed, by Adobe Photoshop.

2.3.8 Quantitative analysis of PSTVd-accumulating cells

PSTVd-accumulating cells were counted by visual inspection under the light microscope. For whole mount in situ hybridization samples, the counting was conducted in a 2.5 × 2.5 mm area per leaf sample. For thin section in situ hybridization samples, the number of PSTVd-accumulating cells was determined separately for different cell types: epidermis, palisade mesophyll and spongy mesophyll. The number was counted per
section, which was ~1 × 1 mm. For the WT, 103 sections from 14 leaves (each from an individual plant) were used. For mutant A325G, 88 sections from 17 leaves were used. For each leaf, at least 4 different sections were analyzed, and the average of numbers from these sections was used to represent the number from each leaf for statistical analysis, so that comparisons were made among biological replicates, rather than among individual sections. The data were analyzed by t test using Microsoft Excel (two-tailed distribution, two-sample unequal variance). When the p-value < 0.05, it was treated as a significantly different value, and is indicated in the text.

2.4 RESULTS AND DISCUSSION

2.4.1 A 3-D structural model for PSTVd loop 6

To infer the 3-D structure of PSTVd loop 6, we used the Find RNA 3D (FR3D) program suite (Sarver et al., 2008) to search a non-redundant list of atomic-resolution Protein Database (PDB) files, updated on May 5th, 2010. The goal was to find internal loop sequence matches of the PSTVd loop 6 sequence 5’-CGA-3’...5’-GAC-3’ flanked on both sides by WC (including near-isosteric G/U wobble) base pairs as templates for modeling the structure of loop 6. Although we did not find an exact sequence match in the 3-D database, we did find two internal loops that each differs from PSTVd loop 6 at single base positions. One of these loops occurs in helix 89 (H89) of bacterial and archaeal 23S ribosomal RNAs (rRNAs). The sequence and 3-D structure of this motif
appear to be broadly conserved in bacterial and archaeal 23S rRNAs. The sequence of this internal loop is 5’-CAA-3’..5’-GAC-3’ in the 23S rRNAs of both the archaeon *Haloarcula marismortui* and the distantly related bacterium *Escherichia coli*, having the same crystal structure (Table 2.1). The boldfaced A in this bacterial and archaeal loop sequence differs from G at the equivalent position of PSTVd loop 6. The corresponding H89 motif in the 23S rRNAs of bacterium *Deinococcus radiodurans* has the sequence 5’-CGA-3’..5’-GAC-3’, with G being different from the A at position 37 of the PSTVd loop 6 (Table 2.1). The structure of this H89 motif is nearly identical to the *E. coli* and *H. marismortui* versions. In the following paragraphs, a boldfaced base in a loop sequence indicates that it is different from the base at the equivalent position in PSTVd loop 6. The other internal loop we found in the database has the sequence 5’-CAA-3’..5’-GAC-3’ in helix 41 (h41) of *E. coli* 16S rRNA (Table 2.1). It has the same 3-D structure as the H89 internal loop, but is not conserved in other 16S rRNAs.

To further explore the sequence variability that can produce the same geometry, we used the H89 internal loop structure to carry out geometric searches using FR3-D to find geometrically similar 3-D motifs. In addition to recovering the original motifs, we identified an internal loop in the 3-D structure of the *Azoarcus* group I intron that adopts a 3-D structure similar to that of the 23S H89 motif, even though its sequence, 5’-CAA-3’..5’-AAA-3’, is considerably different (3 base changes) (Table 2.1). The results of this search, including 3-D structure superpositions and detailed structure annotations, can be accessed at the interactive website http://rna.bgsu.edu/WebFR3-D/Results/pstvd. A view of the weblink is shown in Appendix A.
All of these motifs comprise a conserved set of three stacked non-WC base pairs flanked by WC base pairs, as shown in Figure 2.3. According to the Leontis-Westhof nomenclature of base pair classification (Leontis and Westhof, 2001), the first non-WC base pair is a cis Watson-Crick bifurcated (cWC bif) pair, the second is a trans Sugar-Hoogsteen (tSH) pair, and the third is a trans Hoogsteen-Sugar (tHS) pair. Note that “tSH GA” and “tHS AG” indicate the same base pair geometry, with hydrogen-bonding between the Hoogsteen edge of A and the Sugar edge of G.

The sequence variability observed in the instances of this recurrent 3-D motif recovered from the structure database is consistent with the base pair isostericity matrices defined for base pair families (Leontis et al., 2002b). GA, AA and GG are mutually isosteric or near-isosteric tSH base pairs, and CC and AC are near-isosteric cWC bifurcated pairs. Most importantly, these same interactions can also be formed by the PSTVd sequence 5’-CGA-3’...5’-GAC-3’. Thus, we hypothesize that PSTVd loop 6 follows the same base pair pattern as seen in this 3-D motif. Specifically, we predict that a GA tSH base pair forms between nucleotides 36 and 325, a tHS base pair between A37 and G324, and a bifurcated cWC pair between C38 and C323 (Figure 2.3). As can be seen in Figure 2.3, the only difference between the PSTVd sequence 5’-CGA-3’...5’-GAC-3’ and the 23S rRNA sequence 5’-CAA-3’...5’-GAC-3’ is the substitution of G for A in the second base pair (tSH). This implies the presence of a GA tSH pair in the viroid where an AA tSH pair is found in the H89 23S motif. As tSH AA and GA base pairs are isosteric (Stombaugh et al., 2009), which means that they occupy the same 3-D space, this change does not significantly perturb the 3-D structure of the motif. In fact, AA
substitutions for GA tSH base pairs are very common in homologous RNA molecules (Stombaugh et al., 2009). Furthermore, the rRNA alignments show that GA is found at this position in some 23S rRNAs, as will be discussed below. Finally, we will show that a PSTVd loop 6 mutant with G to A substitution at this position, a sequence identical to the 23S rRNA motif, is indeed capable of systemic infection (see below). Figure 2.3C shows the structural similarity of the AA and GA tSH base pairs.

As mentioned above, the internal loop that we propose as a model for PSTVd loop 6 is highly conserved in H89 of archaeal and bacterial 23S rRNAs. The crystal structure shows that this motif provides a specific protein-binding site for the conserved ribosomal protein L16 (called “L10e” in archaeal and eukaryal ribosomes). More specifically, Lys123 from L16 protein in E.coli and the equivalent Lys156 from L10e protein in H. marismortui interact with the conserved CC bifurcated pair, which provides an ideal docking site for the positively charged Lys amino group. Figure 2.4 shows the electrostatic complementarity between the terminal -NH₃(+) group of lysine and the minor groove of the bifurcated CC pair. Examination of the L16/L10e alignment in the Pfam database (Finn et al., 2010) (Pfam id PF00252) showed this lysine to be almost universally conserved. Conservation of this docking interaction in bacterial and archaeal rRNA 3-D structures as well as in sequence alignments indicates the biological significance of this docking interaction, which could also be relevant to other RNAs, including PSTVd.
2.4.2 The proposed PSTVd loop 6 structural model is compatible with chemical probing data

Our structural model for PSTVd loop 6 is also consistent with existing chemical probing data, which show that A37, A325, and C38 are highly modified by dimethylsulfide (DMS) (Gast et al., 1996). Figure 2.5A shows the proposed base-pairing model for loop 6 with the bases accessible to DMS marked by red arrows. DMS attacks the imino nitrogens of adenine (atom A-N1) and cytosine (atom C-N3) when the WC edges of these bases are exposed. If the WC edges are involved in base pairing, these imino nitrogens would not be DMS reactive. In the proposed PSTVd loop 6 structural model, the A’s that form tHS base pairs using their Hoogsteen edges have exposed WC edges, as shown in Figure 2.5C, which are predicted to react with DMS, consistent with the chemical probing data. In fact, tHS pairs are among the most common non-WC base pairs, and A’s forming these base pairs are generally highly reactive to DMS probing unless such A's use their WC edges to form another interaction, such as a base triple. Furthermore, the fact that C38 is DMS-reactive whereas C323 is not is also consistent with the proposed loop 6 model. As shown in Figure 2.5D, the bifurcated base pair between C38 and C323 is not symmetrical, as the WC edge of C323 forms bifurcated hydrogen-bonds with the amino group of C38. This interaction protects the WC edge of C323 from DMS-modification, while leaving the WC edge of C38 exposed and accessible to the probe.

Loop 6 is also susceptible to RNase T1 cleavage between G36 and A37 (Gast et al., 1996). RNase T1 cleaves on the 3’-side of guanines that are not engaged in WC base
pairing. This is also consistent with the proposed base-pairing model of PSTVd loop 6, as G36 forms a non-WC tHS base-pair with A325.

We also compared the chemical modification pattern of the PSTVd loop 6 with that observed in the h41 motif of 16S *E. coli* rRNA (Moazed et al., 1986) in the region corresponding to the proposed structural motif (nucleotides 1273:1275, 1260:1262, see Table 2.1). The pattern of reactivity in the h41 motif of 16S rRNA is essentially identical to that of PSTVd loop 6 (Figure 2.5B). All bases, except C1273, are accessible to chemical probes targeting atoms on the WC edges of the nucleotides. C1273 is the only base making an interaction that protects its N3 atom in this structure. These data support our PSTVd loop 6 model, especially for the bifurcated CC base pair. Unlike the h41 motif, PSTVd was not probed with kethoxal. Our model predicts that both G36 and G324 of PSTVd loop 6 should be accessible to kethoxal.

2.4.3 Pospiviroid and 23S rRNA sequence alignment supports the PSTVd loop 6 structural model

To further test the proposed PSTVd loop 6 structural model, we analyzed the sequence variability of this loop in all natural PSTVd variants as well as in different pospiviroid species. The question was whether the proposed model could accommodate the observed sequence variations. First, in the 138 unique sequences of PSTVd natural variants (among 156 isolates) registered in Subviral RNA Database (Rocheleau and Pelchat, 2006) (http://subviral.med.uottawa.ca), PSTVd loop 6 sequence 5’-GAC-3’…5’-
CGA-3' is conserved in the equivalent genomic positions of all the variants. The alignment of PSTVd sequences was conducted by ClustalW program available at the Subviral RNA Database website. The result of this search is provide in Appendix B and also can be accessed at the following url: http://subviral.med.uottawa.ca/cgi-bin/clustalw-results.cgi?jobID=clustalw-1283877994). Second, we downloaded sequences of 8 species of the pospiviroid group from the NCBI Refseq collection (Pruitt et al., 2007) and aligned them using MAFFT (Katoh and Toh, 2008). Sequence variants identified in the pospiviroid alignment are summarized in Table 2.2. The sequence alignments are summarized in Figure 2.6, and the details are shown in Appendix C. We excluded Pepper chat fruit viroid (Verhoeven et al., 2009) from the analysis because its sequence was poorly aligned with other pospiviroid sequences in the loop 6 region due to weak nucleotide sequence identity.

We analyzed the alignment using the Jalview program (Waterhouse et al., 2009). By looking at the columns corresponding to the location of PSTVd loop 6, we observed that the only difference in the pospiviroid sequences of loop 6 was a change from G to C in Chrysanthemum stunt viroid (CSVd), at a position corresponding to G324 in PSTVd (Table 2.1). This substitution results in a change of an AG tHS pair to an isosteric AC tHS pair (Stombaugh et al., 2009), suggesting that the 3-D structure of loop 6 may be conserved across all or most pospiviroid species. Intriguingly, the sequence of loop 6-equivalent from CSVd is the same as the sequence of an essential part of L-Trp-binding motif (Majerfeld et al., 2010), again suggesting the possibility that PSTVd loop 6 and its equivalent in other pospiviroid species serve as protein-binding sites.
We also examined rRNA alignments to gain additional insight into sequence variants capable of forming the proposed 3-D motif. We used the Ribosomal program (Mokdad and Leontis, 2006) to analyze non-redundant alignments of rRNA sequences (Stombaugh et al., 2009). We focused on the H89 motif from 23S rRNA because its structure is conserved in archaea and bacteria and because it binds a conserved protein. The h41 motif from 16S rRNA is not conserved, even within Bacteria, and the sequence alignments show high variability indicating variation in its 3-D structure.

Table 2.3 summarizes the sequence variability observed for the H89 motif. Most species have the same sequence as that observed for *E. coli*, 5’-CAA-3’…5’-GAC-3’, which differs from PSTVd loop 6 at one position, as discussed above. However, five species have exactly the same sequence as PSTVd loop 6, providing additional support for the loop 6 structural model.

2.4.4 Mutational analysis provided genetic evidence in support of the PSTVd loop 6 structural model

To experimentally test the 3-D structural model of PSTVd loop 6, we generated all possible mutations for each base pair in the PSTVd\textsuperscript{WT} (wild type) template and then assayed their systemic infection capabilities. The goal was to determine whether the functional consequence of a mutation could be accounted for by the structural model. For the systemic infection assay, *in vitro* transcripts were prepared for each mutant and used to inoculate the first two true leaves of *N. benthamiana* seedlings as described in
MATERIALS AND METHODS in this chapter (see Figure 2.7A). The systemic infection ability of each mutant was assayed by RNA gel blot analysis of RNA extracted from upper, non-inoculated leaves (10th and 11th true leaves, collectively termed “systemic leaves” in this study) at 28 days post-inoculation. The presence of (+)-circular PSTVd in the upper leaves, followed by sequencing to verify maintenance of the mutant sequence, demonstrates the ability of a mutant to mount a systemic infection. In contrast, absence of a mutant viroid in systemic leaves indicates a failed systemic infection. The experiments were repeated with at least 8 biological replicates (i.e., individually inoculated plants) for each mutant. Inoculation with WT PSTVd and water served as positive and negative controls, respectively.

A total of 45 mutants were analyzed. The systemic infection system and representative RNA gel blots are shown in Figure 2.7A. The results from multiple experiments are summarized in the three 4 × 4 matrices shown in Figure 2.7B, one matrix each for all mutants derived from each proposed base pair of the loop 6 3-D model. The sequencing results of mutant progenies extracted from systemic leaves are presented in Table 2.4. We categorized a mutant as viable only when the mutation(s) was maintained in systemic leaves. Additional representative RNA gel blots showing the systemic infection capacities of different mutants are presented in Figure 2.8.

The first notable feature is that no viable variants (i.e., those capable of systemic infection) were observed for base combinations that cannot form a base pair of the indicated geometric type (“no BP” in the matrices in Figure 2.7). The second attribute is
that in each matrix, all viable sequence variants are able to form base pairs belonging to the same isosteric group as those predicted to occur in the WT PSTVd loop 6 sequence.

For the G36/A325 tSH base pair, only those mutants in which G36 was changed while retaining A at position 325 were capable of systemic infection. In all of these variants, the bases at positions 36 and 325 can form tSH base pairs, isosteric to the wild type tSH GA base pair (Stombaugh et al., 2009). While tSH pairs in which C replaces A325 are also isosteric to GA, they are not viable. Evidently, A is required at position 325, perhaps for sequence-specific protein binding in the minor groove. Thus, besides the 3-D structure of a motif, its primary nucleotide sequence can also be important for function.

Sequence variation is even more restricted for the A37/G324 tHS base pair, which only tolerates AA and AG. We observed reversion of three mutants (A37G, A37U, and G324U) to WT in infected plants in all the progeny clones that were sequenced (Table 2.4). This suggested that these mutants were incapable of systemic infection. In terms of the proposed model, we note that the tHS AG base pair is calculated to be the most energetically favorable pair of this type (Mládek et al., 2009). Interestingly, although the AA combination did not appear in the pospiviroid alignments, it was common among the bacterial 23S rRNA sequence alignments at the corresponding location. Here it should be noted that G324U was reported as a viable sequence variant, and its reversion to WT also was observed, in a previous study (Zhong et al., 2008).

The third pair C38/C323 is required for PSTVd to traffic throughout the plant. All pospiviroid sequences analyzed also have two C's at these positions, which, taken
together with the fact that in the 23S rRNA the H89 motif interacts with a conserved protein, suggests that PSTVd loop 6 serves as a binding site for a protein responsible for trafficking.

2.4.5 All systemic infection-defective mutants of loop 6 fail to exit the inoculated leaves

The experiments described above showed that most mutants of loop 6 were incapable of systemic infection. There could be two possible explanations: either the mutants failed to exit the inoculated leaves or they failed to traffic long-distances after they left the inoculated leaves. To investigate this, we conducted RT-PCR (40 cycles) with RNA samples collected from the petioles of inoculated leaves for all mutants. The RNA samples were collected at 8 dpi, the same time point when RNA was collected from inoculated leaves for in planta replication assay as described below. Use of the same time point for these two types of assays ensured consistency in data interpretation. As shown in Figure 2.9, the progenies of these mutants were not detected in the petioles except for three mutants, G324C, A37G, and A37U. As described above, sequencing of the progenies of these three mutants from systemic leaves revealed reversion back to the WT PSTVd sequences. As expected, the four loop 6 mutants which showed systemic infection while maintaining their mutated sequences (blue color in Figure 2.9 and Figure 2.7) were detected in the petioles.
2.4.6 The failure of some loop 6 mutants to establish systemic infection is not due to deficiency in replication

Given that the systemic infection-defective loop 6 mutants did not exit the inoculated leaves, the next question was whether these mutants failed to replicate in the inoculated leaves. To address this question, *in vitro* transcripts of each loop 6 mutant were inoculated onto young *N. benthamiana* leaves as for the systemic infection assays described above. WT PSTVd- and water-inoculations served as positive and negative controls, respectively. The inoculated leaves were harvested at 8 dpi and chemically fixed for whole-mount *in situ* hybridization using digoxigenin-labelled PSTVd-specific riboprobes. Here whole-mount refers to the fact that pieces of a leaf, rather than thin sections, were used for *in situ* hybridization. We established that WT PSTVd could be detected in the petioles of inoculated leaves at 5 dpi; therefore, 8 dpi was a sufficiently long time interval to analyze the replication patterns. The assay was repeated with 7-8 biological replicates (i.e., individual plants inoculated) for each mutant.

Figure 2.10 shows typical images of PSTVd localization patterns from such analyses. The dark purple spots represent PSTVd hybridization signals in the nuclei. Because of overlap of cells and compression of the tissues incurred during the fixation/*in situ* hybridization processes, it was not feasible to identify clearly the types of cells containing the PSTVd signals. This issue will be addressed below by *in situ* hybridization on thin sections. However, our objective here was to detect signs of viroid replication/accumulation, regardless of cell type. Most mutants showed hybridization signals, suggesting that they were capable of replication. However, there were generally
fewer cells exhibiting hybridization signals for the mutants than for the WT PSTVd. This could be due to low infection efficiency (i.e., low capacity to initiate replication in all inoculated cells), low levels of replication in some cells, or limited trafficking between cells.

The next question was whether the failed systemic infection of a mutant was due to fewer cells supporting its replication/accumulation. To address this, we determined the percentages of cells showing visible accumulation of all mutants as well as the WT. We counted the number of PSTVd-accumulating cells per 2.5 × 2.5 mm area of inoculated leaves. The number of cells accumulating a mutant was calculated as a percentage of that accumulating the WT, to represent a measure of relative replication efficiency. The data are shown in Figure 2.11. A key observation is that mutant G36C, which is capable of systemic infection (see Figure 2.7), was detected in less than 10% of the cells relative to the WT. We therefore postulated that a 10% infection efficiency was sufficient to enable a mutant to achieve systemic trafficking. Then, for any mutant showing 10% or higher infection efficiency in the inoculated leaves, their absence from systemic leaves could be interpreted as chiefly the consequence of failed cell-to-cell trafficking. Fifteen mutants fit this criterion, including G36C/A325C, A325C, A325G, G36U/A325G, G324C, G324U, A37C/G324A, A37C/G324C, A37C, A37C/G324U, A37G/G324A, A37G, A37U, C323U, and C38G/C323U. For mutants that exhibited less than 10% of the WT infection efficiency, it is unclear whether their failure to develop systemic infection is due to impaired intracellular trafficking such as nuclear import/export, replication, in vivo stability, or cell-to-cell trafficking.
We recognize that using 10% of infection efficiency as a threshold level sufficient for systemic infection is arbitrary. However, it is a useful and practical approach in the context of the present investigation. What is important is that a relatively low infection efficiency as compared to the WT, in terms of the percentage of cells infected, could still enable a viroid mutant to achieve successful systemic infection if trafficking is not impaired.

2.4.7 Loop 6 is required for trafficking from palisade to spongy mesophyll

As described above, the whole mount in situ hybridization images do not allow assessment of infection efficiencies of loop 6 mutants in distinct cell types, and therefore do not reveal the cellular boundary at which loop 6 functions to mediate trafficking. The data, however, allowed us to choose particular mutants for further in situ hybridization on thin sections of inoculated leaves to identify this cellular boundary. Mutant A325G was first chosen because it exhibits 15% of the WT infection efficiency and represents nearly the median level among all mutants. Furthermore, we did not observe any reversion of this mutant to WT in our extensive infection experiments, which ensures consistency of observations in multiple samples.

The A325G-inoculated N. benthamiana leaves were chemically fixed at 8 dpi to obtain paraffin sections for in situ hybridization. Figure 2.12A illustrates the leaf cell types in a transverse view and shows that viroid inocula were applied to the upper epidermis. Figure 2.12B shows typical in situ hybridization images of transverse sections
of *N. benthamiana* leaves inoculated with WT PSTVd (positive control), mutant A325G and water (mock negative control). The WT PSTVd hybridization signal was detected in the nuclei in upper epidermal, palisade mesophyll and spongy mesophyll cells. It was also detected in the phloem cells (Figure 2.13). Mutant A325G was localized in the nuclei of upper epidermal and palisade mesophyll cells, but very rarely in spongy mesophyll cells. Furthermore, the mutant was not observed in any bundle sheath or phloem cells (Figure 2.13). Hybridization signal was completely absent from any cells in mock-inoculated leaves (Figure 2.12B). These observations suggest preliminarily that loop 6 plays a critical role in PSTVd trafficking from the palisade mesophyll to spongy mesophyll.

To further test this possibility, we analyzed the number of epidermal, palisade and spongy mesophyll cells exhibiting mutant A325G signal as compared to the WT control. We analyzed 88 sections from 17 leaves individually inoculated with mutant A325G and showing viroid hybridization signals, and 103 sections from 14 leaves individually inoculated with WT PSTVd and showing viroid hybridization signals. As demonstrated in Figure 2.12C, the numbers of viroid-accumulating epidermal cells per section (~1 x 1 mm for each section) are not significantly different between A325G- and WT-inoculated leaves (*t* test, *p* > 0.05). However, the numbers of A325G-infected cells are significantly lower for palisade (*P* < 0.05) and spongy mesophyll (*P* < 0.05) compared to the WT-infected cells. These quantitative data demonstrate, importantly, that mutant A325G has similar replication/accumulation capacity as the WT PSTVd, as shown in epidermal cells. The lower number of palisade mesophyll cells showing A325G accumulation may then
be due to impaired trafficking from the epidermis into these cells. It is possible that comprised lateral trafficking between palisade mesophyll cells also contributed to the lower number of infected cells. Most significantly, however, the rare presence of mutant A325G in spongy mesophyll cells is best explained by the requirement of loop 6 for PSTVd trafficking from the palisade to spongy mesophyll.

In addition to A325G, we analyzed A37U, which had a relatively high infection efficiency (50% of the WT level) as shown by whole mount in situ hybridization. In some leaves, A37U was localized only in the upper epidermal and palisade mesophyll cells, similar to mutant A325G (Figure 2.14). In other leaves, it was detected in a few spongy mesophyll as well as upper epidermal and palisade mesophyll cells (Figure 2.14). Because mutant A37U often reverted to the WT (Table 2.4), we postulate that this instability may well account for the localization of A37U in the spongy mesophyll in some cases, although direct evidence was difficult to obtain from the fixed leaf samples. This reversion may also lead to an arbitrarily high infection efficiency of A37U among all mutants. Overall, exclusive localization of A37U in epidermal and palisade mesophyll cells in many cases is consistent with a decisive role of loop 6 in trafficking from palisade to spongy mesophyll as revealed by the A325G mutant.

Although our current data support a critical role of loop 6 in PSTVd trafficking from palisade to spongy mesophyll cells in young N. benthamiana leaves, we do not know whether this motif is also required for trafficking in the reverse direction (i.e., from spongy to palisade) or whether leaf development impacts loop 6 function at a specific cellular boundary. This is an important question to be addressed in future studies, given
our previous finding that a bipartitie PSTVd motif, whose 3-D structure remains to be determined, is required for unidirectional trafficking from mesophyll to bundle sheath in a leaf development-dependent manner in tobacco (*N. tabacum*) (Qi et al., 2004).

These thin section *in situ* hybridization data may provide an explanation for the low infection efficiencies for many mutants as shown from whole mount *in situ* hybridization. Specifically, some mutants, like A325G, may have similar infection efficiency as the WT in the epidermis. Their impaired trafficking into mesophyll cells, especially the spongy mesophyll, would then lead to the overall lower infection efficiency when infected cells are counted from whole mount *in situ* hybridization samples. In support of this interpretation, our counting of epidermal cells infected by A37U, on *in situ* hybridized thin sections of leaf sample in which this mutant was absent from the spongy mesophyll, led to a number that not significantly different from those for A325G- and WT-inoculated leaves (Figure 2.15). For other mutants, other possibilities may also be considered. First, their low infection efficiency may suggest an additional role of loop 6 in replication. However, it is possible that any mutations in a loop, while disrupting the major biological function of the loop, leads to disturbance to some degree the overall viroid structure so as to indirectly impact other biological functions of the RNA as a whole. This would not be surprising if we consider any one motif to function as an integral part of the whole RNA, rather than an isolated entity. Second, for those mutations that caused reduction in infection efficiency to barely detectable levels (e.g., A37G/G324C, A37G/G324U, A37U/G324C, A37U/G324U, C323A/C38A, C323A/C38G, C323G/C38G and C323U/C38U), we cannot exclude the possibility that
some of these mutations produce a dominant negative effect by interacting with some cellular factors to interfere with intracellular trafficking or replication. Finally, whether any mutations cause or enhance RNA instability \textit{in vivo} is an open question. These issues may provide opportunities for new learning about RNA structures in relation to function and for developing more comprehensive research tools to advance studies on RNA structure-function relationships.

In summary, our work provides evidence that the PSTVd loop 6 plays a decisive role in mediating RNA trafficking from palisade mesophyll to spongy mesophyll in \textit{N. benthamiana} leaves. Our structural and bioinformatic analyses, together with the experimental data including structure probing data of PSTVd and the functional data on the loop 6 mutants, support the proposed 3-D structural model of PSTVd loop 6. While it is formally possible that other 3-D models could also be consistent with these data, at present we are not aware of any such models. Whether the primary sequence and proposed 3-D structure of this loop also functions similarly in other host plant species remains to be investigated. Together with previous identification of PSTVd U43/C318 (loop 7) 3-D motif mediating trafficking from the bundle-sheath to phloem in \textit{N. benthamiana} (Zhong et al., 2007), our findings provide compelling evidence to support a hypothesis that distinct 3-D RNA motifs mediate trafficking across specific cellular boundaries. Further studies will determine whether one or more motifs are involved in trafficking across a particular boundary. This hypothesis and our experimental approaches may be useful for identifying the 3-D motifs in other RNAs that mediate trafficking between specific cells. A \textit{cis} element in the 5'‐untranslated region (UTR) of
potexviral RNA, which plays a role in replication (Miller et al., 1998), was found to mediate cell-to-cell transport of a fused green fluorescent protein (GFP) reporter RNA (Lough et al., 2006). A GFP reporter system also was used to show the existence of RNA trafficking signals in nucleotides 1-102 of the *Arabidopsis FLOWERING LOCUS T* mRNA (Li et al., 2009) and in the 3'-UTR of *GIBBERELLIC ACID-INSSENSITIVE* mRNA (Huang and Yu, 2009). The UTRs of potato *BEL5* mRNA are important for long-distance movement (Banerjee et al., 2006; Banerjee et al., 2009). Systemic spread of *Brome mosaic virus* RNAs in the absence of replication suggests that they have sequence or structural elements recognized by cellular factors (Gopinath and Kao, 2007). The identification of 3-D structural motifs in these RNAs mediating trafficking between specific cells should significantly advance our mechanistic knowledge of RNA trafficking.

How might PSTVd loop 6 or other motifs mediate trafficking between specific cells? A simple working model predicts that these motifs are recognized by distinct cellular factors that are components of the cell-specific RNA trafficking machinery. There is evidence for cellular proteins involved in trafficking cellular RNAs. Xoconostle-Cázarés et al. (Xoconostle-Cazares et al., 1999) isolated CmPP16-1 from pumpkin phloem exudates that binds and traffics endogenous RNAs. Yoo et al. (Yoo et al., 2004) identified *Cucurbita maxima* Phloem SMALL RNA BINDING PROTEIN 1 (PSRP1) which binds synthetic single-stranded siRNAs and potentiates their trafficking in *N. benthamiana* mesophyll as shown by microinjection. Recent studies identified a mobile 50-kD RNA binding protein (RBP50) from the pumpkin phloem exudates that binds
polypyrimidine tract in mRNAs, providing a platform for additional protein interactions to form a ribonucleoprotein complex in the phloem translocation stream (Ham et al., 2009). While the putative cellular factors recognizing any particular PSTVd motifs are yet to be identified, the observations that rRNA 3-D motifs similar or identical to PSTVd loop 6 and loop 7 represent protein-binding sites support the proposition that these motifs are recognized by palisade- and bundle sheath-specific factors, respectively, to initiate trafficking.
Table 2.1 Three-dimensional (3-D) motifs in crystal structures of rRNAs used to model 3-D structure of PSTVd loop 6.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Molecule /location</th>
<th>Sequence</th>
<th>Nucleotide numbers</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSTVd</td>
<td>Loop 6</td>
<td>5'-CGA-3'..5'-GAC-3'</td>
<td>323:325/36:38</td>
<td>NA</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>23S rRNA, helix 89</td>
<td>5'-CAA-3'..5'-GAC-3'</td>
<td>2467:2469/2481:2483</td>
<td>2QBE</td>
</tr>
<tr>
<td><em>H. marismortui</em></td>
<td>23S rRNA, helix 89</td>
<td>5'-CAA-3'..5'-GAC-3'</td>
<td>2502:2504/2516:2518</td>
<td>1S72</td>
</tr>
<tr>
<td><em>D. radiodurans</em></td>
<td>23S rRNA, helix 89</td>
<td>5'-CGA-3'..5'-GGC-3'</td>
<td>2446:2448/2516:2518</td>
<td>2ZJR</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>16S rRNA, helix 41</td>
<td>5'-CAA-3'..5'-GAC-3'</td>
<td>1273:1275/1260:1262</td>
<td>2QAN</td>
</tr>
<tr>
<td><em>Azoarcus sp.</em></td>
<td>Group I intron</td>
<td>5'-CAA-3'..5'-AAA-3'</td>
<td>85:87/57:59</td>
<td>1U6B</td>
</tr>
</tbody>
</table>

The bold-faced nucleotide differs from that in the PSTVd loop 6.
Table 2.2 Sequence variants of loop 6 based on a multiple sequence alignment of eight pospiviroid species.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Viroid</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PSTVd</strong></td>
<td></td>
</tr>
<tr>
<td>Tomato planta macho viroid</td>
<td></td>
</tr>
<tr>
<td>Mexican papita viroid</td>
<td></td>
</tr>
<tr>
<td>5’-CGA-3’..5’-GAC-3’</td>
<td>Iresine viroid 1</td>
</tr>
<tr>
<td></td>
<td>Columnnea latent viroid</td>
</tr>
<tr>
<td></td>
<td>Citrus exocortis viroid</td>
</tr>
<tr>
<td></td>
<td>Tomato apical stunt viroid</td>
</tr>
<tr>
<td>5’-CCA-3’..5’-GAC-3’</td>
<td>Chrysanthemum stunt viroid</td>
</tr>
</tbody>
</table>

The bold-faced nucleotide differs from that in the PSTVd loop 6.
Table 2.3 Comparison of sequence variability observed for H89 motif in 23S bacterial rRNA alignments with the sequence of PSTVd loop 6.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form same structure:</td>
<td></td>
</tr>
<tr>
<td>5’-CGA-3’..5’-GAC-3’</td>
<td>PSTVd</td>
</tr>
<tr>
<td>5’-CAA-3’..5’-GAC-3’</td>
<td>129 species</td>
</tr>
<tr>
<td>5’-CGA-3’..5’GAC-3’</td>
<td>Aquifex sp.</td>
</tr>
<tr>
<td></td>
<td>Propionobacterium sp.</td>
</tr>
<tr>
<td></td>
<td>Wolbachia sp.</td>
</tr>
<tr>
<td></td>
<td>Thermotoga sp.</td>
</tr>
<tr>
<td></td>
<td>Chlorobium sp.</td>
</tr>
<tr>
<td>Form different structure:</td>
<td></td>
</tr>
<tr>
<td>5’-GCA-3’..5’-GAC-3’</td>
<td>Pirellula sp.</td>
</tr>
<tr>
<td></td>
<td>Thermus sp.</td>
</tr>
</tbody>
</table>

The bold-faced nucleotide differs from that in the PSTVd loop 6.
Table 2.4 Sequencing of PSTVd loop 6 mutant progenies in systemically infected plants.

<table>
<thead>
<tr>
<th>Original mutations</th>
<th>Trafficking</th>
<th>No. of plants sequenced</th>
<th>Progeny sequences (No. of plants)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G36A</td>
<td>8/8</td>
<td>3</td>
<td>G36A (3)</td>
</tr>
<tr>
<td>G36C</td>
<td>2/8</td>
<td>2</td>
<td>G36C (1), G36A (1)</td>
</tr>
<tr>
<td>G36U</td>
<td>8/8</td>
<td>3</td>
<td>G36U (3)</td>
</tr>
<tr>
<td>G324A</td>
<td>6/8</td>
<td>3</td>
<td>G324A (3)</td>
</tr>
<tr>
<td>G324U</td>
<td>1/8</td>
<td>1</td>
<td>WT (1)</td>
</tr>
<tr>
<td>A37G</td>
<td>6/16</td>
<td>3</td>
<td>WT (3)</td>
</tr>
<tr>
<td>A37U</td>
<td>8/16</td>
<td>3</td>
<td>WT (3)</td>
</tr>
</tbody>
</table>
Figure 2.1 Secondary structure of PSTVd. The loops/bulges are numbered 1-27 from left to right. The inset shows loop 6, flanked by Watson-Crick base pairs in adjoining stems.
Figure 2.2 RNA nucleobase interactions. A) Each base has three edges, the Watson-Crick, Hoogsteen and Sugar edges available for hydrogen-bonding with other bases. B) Examples of non-Watson-Crick base pairs with cis or trans orientation of the glycosidic bond.
Figure 2.3 Proposed non-Watson-Crick base pairing for PSTVd loop 6 based on the crystal structure of a 23S rRNA internal loop as template. A) and B) Two-dimensional (2-D) base pair annotations A) and stereo view B) of an observed RNA 3-D motif from 23S E.coli rRNA (top) used as a template to manually build a structural model for PSTVd loop 6 (bottom). According to this loop 6 model, bases G322 and C39 make a cWC base pair, C323 and C38 make a bifurcated cWC base pair, bases G324 and A37 make a tSH pair, A325 and G36 make a tHS pair, and, finally, G326 and U35 make a cWC wobble pair. The only difference between the 23S rRNA loop and PSTVd loop 6 structures is the isosteric substitution of the AA tSH pair in the former loop by the GA tSH pair in the latter loop (the A and G bases are highlighted with the colored background in 2-D annotations and with dots in the 3-D structures). The 2-D annotations are based on the Leontis-Westhof nomenclature (Leontis and Westhof, 2001). The 3-D structure for this figure was taken from PDB 2QBE (23S rRNA of E. coli) (Borovinskaya et al., 2007). C) Isostericity of AA and GA tSH base pairs, with identical C1’-C1’ distances in both pairs (dashed lines indicated by arrows) and the same orientation of the bases. The isodiscrepancy index (IDI) calculated between exemplars for AA and AG is 1.56, typical for very similar base pairs (Stombaugh et al., 2009).
Figure 2.4 Docking site for Lys_{123} from L16 ribosomal protein created by the cytosine-cytosine bifurcated pair. The surface is colored by the electrostatic charge calculated using the PDB2PQR webserver (Dolinsky et al., 2004) and displayed with PyMOL (Schrodinger, 2010). Red color indicates negative charge and blue color indicates positive charge. The 3-D structure is based on the high resolution structure of 23S *E. coli* rRNA (PDB 2QBE).
Figure 2.5 Chemical probing data in support of the proposed three-dimensional structure of PSTVd loop 6. A) PSTVd loop 6 structure probing data (Gast et al., 1996) is consistent with the proposed model. Blue arrow denotes the RNase T1 cleavage site; red arrows show the DMS modification sites. B) Chemical probing data of a loop in the 16S rRNA helix 41 (Moazed et al., 1986) used as a model for the PSTVd loop 6. C) The N1 atoms of A325 and A37 of PSTVd loop 6 are accessible to DMS according to the proposed model because the Watson-Crick face of the adenines is not engaged in base pairing. D) The N3 atom of C38 in PSTVd loop 6 is also accessible for DMS in the proposed model.
Figure 2.6 Pospiviroid sequence alignments provide support to the PSTVd loop 6 structural model. Alignment of Pospiviroid Refseq sequences using MAFFT.
Figure 2.6
Figure 2.7 Systemic infection analyses on PSTVd loop 6 mutants. A) The upper left panel illustration shows the systemic infection assay system. The red-colored leaves were inoculated. Total RNA was collected from upper leaves and subjected to RNA gel blot analyses. Representative RNA gel blots are shown. Each lane indicates pooled sample from 8 individual plants. Ethidium bromide staining of the corresponding samples are shown as a loading control. Representative RNA gel blots for individual plants are shown in Figure 2.8. B) Summary of the systemic infection results of PSTVd loop 6 mutants presented in 4 × 4 matrices. Each matrix summarizes data for all mutants derived from one wild type (WT) base pair of the proposed 3-D model of loop 6. For each matrix, the WT sequence is indicated with a green background. All viable mutants capable of systemic infection are indicated with a blue background. Non-viable mutants (i.e., systemic infection-defective) are indicated with a red background. The geometric type ("tHS" or "bif") and isosteric group ("I1" or "I2") of a base pair formed by a sequence variant is indicated in each matrix. Base combinations that cannot form base pairs of the indicated geometric type are indicated with "no BP." The number in parenthesis indicates the number of systemically infected plants out of the total number of inoculated plants.
**Figure 2.7**

### A

- PSTVd inoculation → RNA extraction

### G36

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>11 tHS (6/8)</td>
<td>11 tHS (6/8)</td>
<td>11 tHS (6/8)</td>
<td>11 tHS (6/8)</td>
</tr>
<tr>
<td>C</td>
<td>11 tHS</td>
<td>11 tHS</td>
<td>no BP</td>
<td>11 tHS</td>
</tr>
<tr>
<td>G</td>
<td>no BP</td>
<td>no BP</td>
<td>12 tHS (16/16)</td>
<td>no BP</td>
</tr>
<tr>
<td>U</td>
<td>12 tHS</td>
<td>no BP</td>
<td>12 tHS (16/16)</td>
<td>no BP</td>
</tr>
</tbody>
</table>

### G324

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>11 tHS (6/8)</td>
<td>11 tHS (6/8)</td>
<td>11 tHS (6/8)</td>
<td>11 tHS (6/8)</td>
</tr>
<tr>
<td>C</td>
<td>11 tHS</td>
<td>11 tHS</td>
<td>no BP</td>
<td>11 tHS</td>
</tr>
<tr>
<td>G</td>
<td>no BP</td>
<td>no BP</td>
<td>12 tHS (16/16)</td>
<td>no BP</td>
</tr>
<tr>
<td>U</td>
<td>12 tHS</td>
<td>no BP</td>
<td>12 tHS (16/16)</td>
<td>no BP</td>
</tr>
</tbody>
</table>

### C38

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>11 bif</td>
<td>11 bif</td>
<td>no BP</td>
<td>no BP</td>
</tr>
<tr>
<td>C</td>
<td>I2 bif (8/8)</td>
<td>no BP</td>
<td>no BP</td>
<td>no BP</td>
</tr>
<tr>
<td>G</td>
<td>no BP</td>
<td>no BP</td>
<td>I1 bif</td>
<td>I1 bif</td>
</tr>
<tr>
<td>U</td>
<td>no BP</td>
<td>no BP</td>
<td>no BP</td>
<td>no BP</td>
</tr>
</tbody>
</table>

### C323

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>11 bif</td>
<td>11 bif</td>
<td>no BP</td>
<td>no BP</td>
</tr>
<tr>
<td>C</td>
<td>I2 bif (8/8)</td>
<td>no BP</td>
<td>no BP</td>
<td>no BP</td>
</tr>
<tr>
<td>G</td>
<td>no BP</td>
<td>no BP</td>
<td>I1 bif</td>
<td>I1 bif</td>
</tr>
<tr>
<td>U</td>
<td>no BP</td>
<td>no BP</td>
<td>no BP</td>
<td>no BP</td>
</tr>
</tbody>
</table>
Figure 2.8 Representative RNA gel blots showing systemic infection of PSTVd loop 6 mutants. A) Each lane represents RNA sample from an individual plant inoculated with WT-PSTVd or mutant as labeled. The number of plants exhibiting systemic infection out of the total number of inoculated plants is given in parenthesis. B) Each lane indicates RNA sample pooled from 8 individual plants. In this batch A37G and A37U did not show systemic infection. The two images were taken from the same gel. In all experiments, rRNA serves as loading control.
Figure 2.9 Detection of wild type and mutants of PSTVd loop 6 in petioles of inoculated leaves by RT-PCR. The wild-type base pair is indicated by green color, mutant pairs capable of systemic infection by blue and those incapable of systemic infection by red. M indicates mock inoculation. Asterisks indicate mutants that reverted to WT during infection. RT-PCR amplified actin RNA serves as loading control.
Figure 2.10 Representative images of whole mount in situ hybridization to assay replication of loop 6 mutants. The purple dots, some indicated by arrows, represent viroid hybridization signals in the nuclei. Bars = 100 µm. A) Wild type PSTVd. B) G36A (systemic infection-competent mutant). C) A325G (systemic infection-defective mutant). D) Mock inoculation.
Figure 2.11 Infection efficiencies of PSTVd loop 6 mutants as percentages of wild type (set to 100%) as determined by whole mount in situ hybridization. Mutant pairs capable of systemic infection by blue and those incapable of systemic infection by red. M indicates mock inoculation. Each error bar indicates standard deviation from the mean.
Figure 2.12 In situ hybridization localization of wild type and A325G mutant PSTVd in inoculated leaves of *N. benthamiana*. A) Schematic representation of a transverse section of a young *N. benthamiana* leaf. B) *In situ* hybridization on 14-µm thin sections of *N. benthamiana* leaves inoculated by WT-PSTVd, A325G mutant, and water as a mock. Purple dots indicate viroid hybridization signals in the nuclei. Bars = 100 µm. C) Quantitative analysis of the cellular distribution of mutant A325G (n = 17) as well as WT PSTVd (n = 14). The “n” number represents the number of individual leaves (and therefore individual plants) sampled. See MATERIALS AND METHODS in this chapter for details on sampling and analysis. The data in the two columns indicated by an asterisk are significantly different (t-test, P < 0.05). The numbers in the two columns indicated by double asterisks also are significantly different (P < 0.05). The non-highlighted data in the two columns are not significantly different (P > 0.05) Each error bar indicates standard deviation from the mean.
Figure 2.12
Figure 2.13 *In situ* hybridization on 14-μm paraffin sections of *N. benthamiana* leaves inoculated with PSTVd or water. Representative images of both transverse and oblique sections are shown. The red arrows indicate vascular cells. The black arrows indicate PSTVd signals in phloem. uEp, upper epidermis; Pm, palisade mesophyll; Sm, spongy mesophyll; BS, bundle sheath. Bar=100 μm.
Figure 2.13
Figure 2.14 *In situ* hybridization 14-μm paraffin sections of *N. benthamiana* leaves inoculated with A37U mutant. Representative images of both transverse and oblique sections are shown. The cellular distribution patterns of A37U progenies vary in different inoculated leaves, consistent with the observed reversion of A37U mutant back to WT in some cases. The red arrows indicate the vascular cells. uEP, upper epidermis; Pm, palisade mesophyll; Sm, spongy mesophyll; BS, bundle sheath. Bar=100 μm.
Figure 2.14
Figure 2.15 Comparison of infection efficiencies of WT, A325G and A37U in leaf epidermis of inoculated *N. benthamiana*. For A37U, only sections from leaves showing epidermal and palisade infections were used. Those showing spongy mesophyll infection were excluded for concerns of reversion to WT to cause data bias. WT: n =14 leaves (103 sections); A325G: n=17 leaves (88 sections); A37U: n=10 leaves (53 sections). T-test results: differences WT and A325G or between WT and A37U are not significant. p > 0.05 for each test. Details of cell counting and analysis are described in MATERIALS AND METHODS in this chapter.
CHAPTER 3

RAPID IN VIVO EVOLUTION OF
AN RNA LOOP FOR SYSTEMIC TRAFFICKING OF
POTATO SPIDLE TUBER VIROID
3.1 ABSTRACT

Systemic trafficking of RNA, achieved by the combination of cell-to-cell and long-distance RNA trafficking is a key process for RNA-based pathogens to broaden their territory inside a host; however, its regulatory mechanism is still not fully elucidated. Analyses on a trafficking-defective mutant of *Potato spindle tuber viroid* (PSTVd) led us to detect rapid *in vivo* evolution by which the mutant gained capability of systemic trafficking in *Nicotiana benthamiana*. The original trafficking-defective mutant was generated by obliterating a wild type-PSTVd secondary (two-dimensional, 2-D) structure called loop 19. Analyses on the gain-of-function variants of PSTVd evolved from this mutant revealed that the variants acquired a *de novo* 2-D structure, named loop 19*, similar to loop 19 and in its one base-pair shifted genomic location. In this chapter, identification of these gain-of-function variants, their characteristic 2-D structures, and their sequences are presented through analyses of PSTVd loop 19 and loop 19* mutants. These findings and approaches have broad implications for understanding RNA motifs essential for intercellular trafficking as well as for the host adaptation.

3.2 INTRODUCTION

Various RNAs can systemically traffic in a plant body by passing through different cellular layers from one cell to another (Giakountis and Coupland, 2008; Kehr and Buhtz, 2008; Lucas et al., 2009; Takeda and Ding; 2009; Turgeon and Wolf, 2009;
Chitwood and Timmermans, 2010; Chuck and O’Connor, 2010; Ding and Wang 2010; Hannapel, 2010; Lehesranta et al., 2010). During their trafficking, RNAs enter into a vascular tissue, phloem, where they can traffic in long-distance through continuous vascular tube (Lough and Lucas, 2006; Turgeon and Wolf, 2009). Thus, systemic trafficking is achieved by both cell-to-cell and long-distance trafficking. Systemic trafficking phenomena of both endogenous and exogenous RNAs have been reported in plants (various examples are described in Section 2.2). A striking example of endogenous RNA trafficking is *BEL5* mRNA in potato. The full-length *BEL5* mRNA, produced in potato shoot, traffics to the roots through phloem, resulting in promotion of potato growth (Banerjee et al., 2006; Yu et al., 2007). This work elegantly demonstrated the developmental and physiological regulation triggered by the long-distance trafficking of RNA. Exogenous RNAs, such as viruses and viroids, can also traffic to expand their area of infection (Gopinath and Kao, 2007; Benitez-Alfonso et al., 2010), and cause various symptoms in plants. Despite the pivotal roles that RNA systemic trafficking plays in plant physiology and development, the mechanism regulating this process remains to be fully elucidated.

As one of the possible regulatory mechanisms, RNA structural motif-mediated RNA intercellular trafficking is gaining experimental support (Banerjee et al., 2006; Gopinath and Kao, 2007; Lough et al., 2006; Zhong et al. 2007; 2008). To understand the versatile RNA trafficking phenomena, it is critical to identify yet unknown RNA motifs regulating intercellular RNA trafficking. Also, the accumulation of knowledge of known motifs helps to increase our understanding of RNA structure-function relationships in
In the role of structural motifs in RNA trafficking, PSTVd provides a unique platform to identify and investigate the direct functions of RNA motifs because of its noncoding property, and characteristic loop and stem secondary (two-dimensional, 2-D) structures. Indeed, most loop structures often form recurrent three dimensional (3-D) structures via non-Watson-Crick (non-WC) base pairing, base stacking, and other base-base interactions (Leontis et al., 2006), such structures often function in RNA-RNA, RNA–protein, and RNA-small ligand interactions (Leontis et al., 2002a, b; Noller, 2005; Leontis et al., 2006; Steitz, 2008). Our lab has identified a total 11 loop/bulge (collectively referred to as 'loops') regions in the Potato spindle tuber viroid (PSTVd) genomic RNA that are required for its systemic trafficking in Nicotiana benthamiana (Zhong et al., 2007; 2008). The 3-D structural motifs of the two loops, loop 6 (Chapter 2) and loop 7 (Zhong et al., 2007), were identified as requirements for PSTVd trafficking through specific cellular layers.

Loop 19, in PSTVd genomic RNA, was categorized as a 2-D loop required for PSTVd systemic trafficking (Zhong et al., 2008). This loop is composed of two nucleotides, A135 and C227, flanked by canonical cis-Watson-Crick (cWC) base pairs (Figure 3.1A). Zhong and her colleagues reported that the obliteration of loop 19 structure by C227U mutation (Figure 3.1B) abolished PSTVd systemic trafficking in N. benthamiana, while retaining replication efficiency in 46% of wildtype (WT: strain PSTVd\textsuperscript{Int}, accession NC002030) in a N. benthamiana single-cell assay system (Zhong et al., 2008). The trafficking capability of the A74-deleted mutant, which retained 20 % replication efficiency in the single-cell replication assay, ensures that 46 % replication-
efficiency of C227U should be sufficient for trafficking (Zhong et al., 2008). Thus, Zhong and her colleagues concluded that loop 19 is required for PSTVd systemic trafficking. This C227U mutation changes the 2-D-loop structure to the “closed” status in mfold (Zucker, 2003); therefore, it is possible that loop 19 3-D structure is required for PSTVd systemic trafficking. At the same time, we cannot rule out the possibility of loop 19 sequence, and/or other unidentified factor(s) as the primary cause of PSTVd trafficking.

In this study, in order to elucidate the detailed mechanism of loop 19 in PSTVd systemic trafficking, analyses on a series of loop 19 mutants were conducted. These led to the unexpected observation that C227U mutant evolved to gain systemic trafficking ability in a plant by acquiring a de novo 2-D-loop structure similar to loop 19. In this chapter, the observation of this rapid gain-of-function evolution of PSTVd mutants and the analyses on the novel PSTVd variations are presented and discussed, in addition to the detailed analyses on loop 19 mutants.

3.3 MATERIALS AND METHODS

3.3.1 cDNA construction of PSTVd mutants

Plasmid pRZ6-2 containing the cDNA of PSTVd<sup>Int</sup> (pRZ6-2-Int) was constructed by Hu and colleagues (Hu et al., 1997) and was a gift from Dr. Robert Owens. All PSTVd mutants described in this paper were generated by site-directed mutagenesis based on this template, as previously described (Zhong et al., 2008). All introduced
mutations were verified by DNA sequencing with a 3730 DNA Analyzer (Applied Biosystems, Carlsbad, CA) at the Plant-Microbe Genomics Facility (PMGF) of The Ohio State University. Plasmid pInter (-) was constructed as previously described (Qi and Ding, 2002).

3.3.2 Plant growth and infection

*N. benthamiana* plants were grown in a growth chamber maintained at 16 h light/8 h dark (28°C) cycles. The *in vitro* transcripts of PSTVd mutants were mechanically inoculated onto the first two true leaves of two-week-old *N. benthamiana* plants (200 ng/leaf) dusted with carborundum powder. Diethylpyrocarbonate (DEPC)-treated H₂O was used for mock inoculation. All inocula were applied to the upper surfaces of leaves.

3.3.3 *In vitro* transcription

To prepare *in vitro* transcripts of WT and mutant PSTVd for systemic infection and replication experiments, *Hind*III-linearized pRZ6-2-Int and all mutant derivatives were transcribed by using T7 Megascript kit according to the manufacturer's instruction (Ambion, Austin, TX). Antisense PSTVd riboprobes for *in situ* hybridization and RNA gel blots were prepared by *in vitro* transcription of SpeI-linearized pInter (-) plasmid using T7 Maxiscript kit (Ambion). [α-³²P] UTP-labeled riboprobes were used for RNA gel blots. All RNA transcripts were purified by MEGAClear kit (Ambion) after DNase I digestion.
3.3.4 RNA extraction and RNA gel blotting

Total RNA from leaves of PSTVd- or mock-inoculated *N. benthamiana* plants was extracted by Ribozol reagent (Amresco, Solon, OH) by following the manufacturer’s instructions. RNA gel blotting was performed essentially as described in Zhong et al (Zhong et al., 2006).

3.3.5 Sequencing of PSTVd progenies

The PSTVd progenies were sequenced in a manner similar to that described in Qi and Ding (2002). Briefly, after the PSTVd signal detection by RNA gel blotting, the PSTVd RNA was RT-PCR amplified. In this RT-PCR reaction, total RNA (1 µg) was reverse-transcribed by using Superscript III (Invitrogen, Carlsbad, CA). The synthesized cDNAs were amplified with Pfu Turbo DNA polymerase (Agilent Technologies, Santa Clara CA) by using PSTVd-specific forward primer (5’-CGGAACTAAACTCGTGTCCT-3’) and reverse primer (5’-AGGAACCAACTGCCTGGCTCA-3’). Next, the Pfu-amplified product was incubated with Taq polymerase for 20 min at 72 °C and ligated into pGEM-T vector (Promega, Madison, WI). The cloned DNA was sequenced as described above.
3.4 RESULTS

3.4.1 Rapid in vivo evolution of an RNA loop from loop 19-closed PSTVd mutant

The systemic trafficking ability of C227U mutant, which has a loop 19-closed 2-D structure, was re-examined. The unit-length C227U mutant transcribed in vitro was inoculated onto the first two true leaves of N. benthamiana seedlings to test its trafficking ability as described in MATERIALS AND METHODS in this chapter. This was assayed by RNA gel blot analysis of RNA extracted from upper, non-inoculated leaves (10th and 11th true leaves, collectively named as “systemic leaves” in this study) at 28 days post-inoculation (dpi). The presence of (+)-circular PSTVd (c-PSTVd) in the systemic leaves, followed by sequencing to verify maintenance of the mutant sequence, demonstrates the ability of a mutant to systemically traffic.

In more than 80% of the 320 inoculated plants, PSTVd did not exist in systemic leaves (Figure 3.2A), suggesting trafficking-defectiveness of C227U as reported earlier (Zhong et al., 2008). However, at the same time, we observed the existence of C227U progenies in a small proportion of plants. Figure 3.2B shows PSTVd existence in pooled RNA from a total of 320 individual plants individually inoculated. To estimate the systemic trafficking efficiency, we further examined the existence of PSTVd in smaller pools or individual plants and estimated that the systemic trafficking occurred in a maximum of 20% inoculated plants.

To elucidate the characteristics of the population which systemically traffics, the progenies of C227U inoculum extracted from the systemic leaves of 7 randomly chosen
plants were sequenced, and the results are summarized in Table 3.1. It is noteworthy that all the sequenced progenies acquired a *de novo* mutation in G134 position in addition to C227U maintained from the sequence of the inoculums. Five of the sequenced progenies acquired a G134A mutation, and the other two progenies acquired a G134U mutation. Several additional mutations (U27C, G67C, U200A, U200C, and G319U) were observed in four of the sequenced progenies, but they occurred only with G134A or G134U mutation. These additional mutations that occurred either with G134A or G134U do not affect the 2-D structure of loop 19 or overall rod-like structure of each PSTVd-mutant in mfold.

Among these newly acquired mutations, G134A caught our attention because of its high frequency of occurrence, occurring in 5 out of 7 plants sequenced. To understand the effect of G134A mutation on the overall loop 19-closed PSTVd 2-D structure in C227U mutant, we conducted mfold on G134A/C227U double mutant. The predicted 2-D structure revealed the generation of new loop structure (loop 19*: A134/C228, Figure 3.1C). This loop 19* structure is essentially one base pair shift from the original loop 19-A/C loop (A135/C227) observed in WT PSTVd. This phenomenon implies that the loop 19* is a *de novo* loop-structure acquired to gain systemic trafficking ability. Interestingly, G134U/C227U mutant was also folded to a similar two-nucleotide-loop structure in the loop19* position (Figure 3.1D). Hereafter, for simplicity, all the PSTVd sequence variations which generate a two-nucleotide-loop in 134/228 position by mfold are called loop 19* mutants.
3.4.2 Trafficking-competency of loop 19* mutants

To confirm that loop 19* mutants, G134A/C227U and G134U/C227U, are truly trafficking-competent, we generated these mutants by *in vitro* transcription and tested their systemic trafficking ability in *N. benthamiana* by the systemic trafficking assay described above.

In 5 out of the 7 plants individually inoculated with G134A/C227U mutant, the progenies showed systemic trafficking (Fig. 3.3A). The progenies extracted from 5 individually infected plants were sequenced and the results are summarized in Table 3.2. Three of the progenies maintained exactly same sequence as the inoculums, indicating *bona fide* trafficking-competency of G134A/C227U mutant. Two of the sequenced progenies acquired additional mutations during their infection. One of the progenies acquired U240C mutation and another one acquired C148U mutation. Though the effect of these mutations in systemic trafficking remains to be answered, each mutation did not affect the loop 19* and overall rod-like 2-D structure of PSTVd in mfold.

In 7 out of 9 plants individually inoculated with G134U/C227U mutant, progenies showed systemic trafficking (Figure 3.3B). The progenies extracted from 3 individually infected plants were sequenced, and the results are summarized in Table 3.2. One progeny-sequence maintained G134U/C227U – the original inoculum sequence; therefore, G134U/C227U was also confirmed as a *bona fide* trafficking-competent mutant. Another two progeny sequences exhibited G134A sequence (U to A change at position 134 and U to C change at position 227 from U134/U227) as well as G insertion
between nucleotide C93 and C94 (Table 3.2). This G134A sequence variation provides a larger loop structure fusing loop 19 and loop 19* (Figure 3.4B).

To understand the effect of the larger loop structure having both loop 19 and loop 19* in PSTVd, systemic trafficking ability of this mutant was analyzed. In 9 out of 10 plants individually inoculated with G134A mutant, progenies showed systemic trafficking (Figure 3.3C). Two of the sequenced progenies maintained G134A sequence in systemic leaves, showing the *bona fide* trafficking-competency of this mutant.

In summary, it was shown that two loop 19* mutants, G134A/C227U and G134U/C227U, are trafficking-competent. In addition, G134A mutant, which has both loop 19 and loop 19* A/C sequence, generates a fused larger loop structure and is also trafficking-competent. Thus, these mutants were shown as *bona fide* gain-of-function mutants that rapidly evolved within 30 dpi from a trafficking-defective PSTVd mutant during their infection process.

### **3.4.3 Systemic trafficking analyses of loop 19 mutants**

Next, a series of loop 19 mutants (total of 15 mutants) was created, and their systemic trafficking abilities were analyzed by RNA gel blots. The experiment was repeated with at least 8 biological replicates (i.e., individually inoculated plants). Inoculation with WT PSTVd and water served as positive and negative controls, respectively.
In addition to simply cataloguing the systemic trafficking ability of each mutant, these analyses were conducted with the expectation of gaining the following results and insights into the systemic trafficking of loop 19 mutants: 1) generating different variation of loop 19* mutants by rapid evolution from various loop 19 mutants, 2) comparing the systemic trafficking ability of loop 19 and loop 19* mutants to understand structural similarity or difference in these two loops, and 3) deducing 3-D structure of loop 19 from the systemic trafficking pattern of each mutant based on the isostericity matrices. Isostericity matrices provide the basis for analyzing the sequence signatures of RNA 3-D motifs (Leontis et al., 2002b). Certain non-WC base pairs can substitute for each other in a motif without distorting the 3-D structure of the motif; such base pairs are isosteric or nearly so (Michel and Westhof, 1990; Leontis and Westhof, 1998b; Leontis et al., 2002b). If the 3-D structure of loop 19 is the primary determinant of PSTVd systemic trafficking ability, the non-isosteric (structurally different) type of mutant should be functionally null. Given this, functionality of mutants can be compared to the all available structural patterns. The structural model of one of the PSTVd loops, loop 7, was successfully deduced by this methodology (Zhong et al., 2007).

The results from multiple experiments are summarized in the 4 × 4 matrix shown in Figure 3.5A. The representative RNA gel blots are shown in Figure 3.5B, and additional RNA gel blots are shown in Figure 3.6. The background color of the matrix indicates systemic trafficking ability of each loop 19 mutant. In the matrix, the WT sequence is indicated with a green background. Blue background indicates a mutant that maintains the systemic trafficking ability (i.e., trafficking-competent mutant). The red
background indicates the mutant was not detected in systemic leaves. The yellow 
background indicates a mutant whose progenies were detected in the systemic leaves with 
sequence alteration(s) from the original inoculum sequence (called trafficking-alteration 
mutant). The number of observed trafficking events is listed in each box in the matrix. For example, “11/12” for the sequence A135/A227 indicates that 11 individual 
trafficking events were observed out of 12 independent inoculations. The sequencing 
results of the progenies are summarized in Table 3.3 for trafficking-competent mutants 
and in Table 3.4 for trafficking-alteration mutants.

**Trafficking-competent mutants:** We observed six trafficking-competent 
mutants of this category, the original inoculum sequences were recovered in systemic 
leaves, suggesting maintained sequences throughout trafficking. This sequence 
maintenance was observed in at least one of the plants inoculated (Table 3.3, bold letters 
represent the maintained sequences). At the same time, we did observe additional 
mutations, deletions, or substitutions in this category of mutants. Most of the mutations 
occurred in non-loop 19 areas, and they did not either alter loop 19 or overall rod like 
structure of PSTVd in mfold. One of the sequences had an extra mutation in nucleotide 
G134 (i.e., loop 19* position) from A135U/C227U inoculums, generating 
G134A/A135U/C227U mutant. However, this mutant did not generate either loop 19 or 
loop 19* structure in mfold (Figure 3.4C). The effects of these additional mutations for 
trafficking-competency remain to be elucidated. We also observed mutations in loop 19 
nucleotides (A135 and C227). Two of the sequences changed from A135G/C227A to
C227A. Since C227A is trafficking-competent, systemic trafficking of this progeny is accountable. However, C227U, which is generally trafficking-defective (see above) was detected in a plant inoculated by C227A mutant. Since C227U mutant did not show trafficking capability in any of 320 inoculations with its sequence maintained, it is plausible to think that this mutation occurred after systemic trafficking.

**Non-trafficking-competent mutants:** The progenies of five other mutants were detected in the upper leaves, but their sequences were always altered from those of the inoculums. The mutants in this category (A135U, A135C/C227A, A135C/C227U, A135U/C227A, and C227U, collectively called trafficking-alteration mutants) are shown as yellow background in Figure 3.5A. The list of the sequencing results from this type of loop 19 mutants is shown in Table 3.4. We observed G134A *de novo* mutation, again. A135C/C227A was mutated to G134A/A135C/C227U. A135C/C227U was also mutated to G134A/A135C/C227U. This mutant sequence did not contain loop 19* 2-D structure in mfold (Figure 3.4D), and the effects of these additional mutations for trafficking-competency remain to be elucidated. We also observed A135U reverting back to WT (U in position 135 to A). This same mutation occurred in A135U/C227A mutant, generating C227A in systemic leaves. C227A is categorized as trafficking-competent; therefore, this change probably appears to make A135U/C227U gain trafficking-competent status. The rest of the mutants that we did not detect in systemic leaves were A135C/C227G, A135G/C227G, A135G/C227U, and A135U/C227G (highlighted by red color in Figure 3.5A).
Based on the trafficking pattern above, any 3-D structure model of loop 19 could not be deduced. Also, any particular sequence or 2-D structure of loop 19 required to maintain the systemic trafficking could not be deduced. These issues will be discussed in more detail later in this chapter.

3.4.4 Replication analyses of loop 19 mutants

For all the loop 19 mutants, we conducted in planta replication assay to explore the possibility of structure- and/or sequence-replication function. Each mutant was in vitro transcribed and inoculated onto young N. benthamiana leaves as in the systemic trafficking assay described above. WT PSTVd- and water-inoculations served as positive and negative controls, respectively. The inoculated leaves were collected at 8 dpi, and total RNA was extracted for RNA gel blot analyses. We established that WT PSTVd could be detected in the petioles of inoculated leaves at 5 dpi; therefore, 8 dpi was a sufficiently long enough time interval to analyze the replication patterns in the inoculated leaves. The assay was repeated with 3 biological replicates (i.e., individual plants inoculated) for each mutant. The signal intensity of c-PSTVd was quantified and normalized by that of WT. WT signal intensity was set as 100% replication-efficiency. Figure 3.7 shows a summary of replication-efficiency based on the 3 individual RNA gel blots for all the mutants. The RNA gel blot results are provided in Figure 3.8.

A key observation is that one of the trafficking-competent mutants, A135G, showed about 20% replication-efficiency on average relative to WT. Therefore, it could
be postulated that a 20% replication-efficiency is sufficient to enable a mutant to achieve systemic trafficking. All the mutants except A135G/C227G are comparable or higher than this replication-efficiency, suggesting that most of the mutants maintained the level of PSTVd high enough to achieve systemic trafficking. Importantly, 3 out of 4 mutants which we did not detect in systemic leaves (A135C/C227G, A135G/C227U, and A135U/C227G) had comparable or higher levels of replication-efficiency. Therefore, the absence of these mutants in systemic leaves could be interpreted chiefly as the consequence of failed cell-to-cell trafficking. One mutant, A135G/C227G, had level of replication-efficiency lower than 20%; therefore, it is unclear whether the mutant’s failure to develop systemic infection is due to impaired intracellular trafficking such as nuclear import/export, replication, in vivo stability, or intercellular trafficking.

Thus, in general, trafficking-competent and trafficking-alteration mutants had relatively high replication-efficiencies. In the case of trafficking-alternation mutants, it remains to be elucidated if these high replication-efficiencies are due to acquired mutations. From the results of the replication analyses, any correlation between loop 19 structure/sequence and replication-efficiencies could not be deduced.
3.5 DISCUSSION

3.5.1 A working model: loop 19/loop 19* as critical loops for PSTVd systemic trafficking

In this chapter, rapid evolution of trafficking-competent PSTVd loop 19* mutants is reported. These mutants have a *de novo* 2-D loop structure similar to WT loop 19, but it is located one basepair shift from the original genomic position. These mutants evolved from a trafficking-defective C227U mutant during its infection to *N. benthamiana*. Taken together, the data suggests the importance of loop 19* structure and/or sequence in PSTVd systemic trafficking; however, more detailed mutational and structural analyses of loop 19* should be carried out to conclude this issue. In addition, trafficking-alteration and trafficking-defective mutants show that certain mutations in loop 19 abolish PSTVd systemic trafficking, suggesting the importance of loop 19 in PSTVd systemic trafficking. Thus, it is plausible to postulate that both loop 19 and loop 19* are critical loops for PSTVd systemic trafficking. However, the primary factor in switching trafficking ability on and off in loop 19/loop 19* has not yet been identified.

In this study, any correlation could not be deduced between loop 19 structure/sequence and systemic trafficking. One of the possibilities is that the structure of this loop is not the primary determinant of PSTVd systemic trafficking ability. It is tempting to postulate that the specific host factor(s) can interact with loop 19/loop19*, such as yet unidentified proteins. These proteins may physically bind to loop 19 or loop
*19* in *N. benthamiana*, and play decisive roles in systemic trafficking together with loop 19/loop 19*. Further studies need to be conducted to resolve this issue.

### 3.5.2 General conservation of loop 19 secondary structure in PSTVd natural variants

Loop 19 2-D structure and sequence is conserved in most of the 156 PSTVd natural variants registered in the Subviral RNA Database (Rocheleau and Pelchat, 2006) (http://subviral.med.uottawa.ca) except for four variants having the accession numbers AF483470, AY372398, EF459701, and X17268. The sequences of all the natural variants were extracted from the website and aligned by using ClustalW available at the Subviral RNA Database website. Loop 19 sequence of PSTVd\textsuperscript{Int} (WT in this study) and equivalent positions in other natural variants were visually inspected. The three variants have “A” in the equivalent position of G134 in PSTVd\textsuperscript{Int}. This G to A change generates a larger loop structure, fusing loop 19 A/C and loop 19* A/C sequences (Table 3.5). Interestingly, this larger loop structure/sequence was observed as one of the evolved mutants during systemic trafficking in *N. benthamiana* (G134A; Figure 3.4A), and this mutation is trafficking-competent. Other than the G134A type variants, there is only one natural variant, X17268, which does not have loop 19 nor loop 19* structure (Table 3.5). The mechanism of how this non-loop 19 natural variant manages its infection remains to be answered. This general conservation suggests the importance of loop 19 in PSTVd infection to host plants.
3.5.3 Loop 19 and loop 19* as a region of high nucleotide variation

Contrary to the general conservation of loop 19/loop 19* region in almost all PSTVd natural variants, many de novo sequence variations were observed in this study. A systemic trafficking study of various loop 19 mutants revealed its high susceptibility to mutation. Among 47 sequenced progenies of various loop 19 mutants, 13 had additional mutations in loop 19* position, and 10 showed mutations in loop 19 position. Thus, about 50% of the sequenced progenies in loop 19 mutants acquired mutation(s) in loop 19/loop19* position. In a similar analysis in loop 6, 8 out of 18 sequenced progenies had additional mutations in loop 6 position (Chapter 2). Thus, the mutation rate is comparable between the two different loops. However, most of the mutations in loop 6 were simple reversion to WT sequence (in 7 out of 8 mutations). Whereas, in loop 19, WT reversion was observed only in 5 out of 23 newly acquired loop19/loop19* mutations, and various mutations evolved in vivo were observed in loop19/loop 19* region. With this small number of sequences, it is difficult to conclude the occurrence frequencies of each mutation and the mutation hot spot. Sequencing of larger number of progenies will be required to quantify and understand this issue more clearly.

In a different species of viroids in different hosts, certain reproducible mutation-hot-spots were reported (Kawaguchi-Ito et al., 2010). Hop stunt viroid isolates from different host plants acquired convergent de novo mutations in hop plants during 15 years of infection, suggesting that positive host-selection pressure existed. Loop 19/loop 19*
region might also be the case since PSTVd was infectious with various loop 19/loop 19* sequences in this study, but at the same time this region is highly conserved with low variation in the existing PSTVd natural variants. This implies the existence of positive selection pressure for this loop region. It also suggests the importance of the loop 19/loop 19* region in PSTVd infection. It is tempting to postulate that loop 19/loop19* interacts with host specific factors, and this interaction works as a host-selection pressure so that the existence of loop 19/loop 19* can be a determinant of PSTVd host adaptation.

As a support of this claim, another research group showed an example of PSTVd using loop 19 to adapt in a new host environment. Vachev and his colleagues isolated C227U natural sequence (loop 19-closed) variation of PSTVd, strain KF-440-2, during trafficking from WT infected tomato to a holoparasitic plant, Orobanche ramose (Vachev et al., 2009). They did not find loop 19* mutation in tomato by C227U inoculation as identified in this study by using N. benthamiana as the experimental host plant. Taken together, it suggests a usage of loop 19 variation for PSTVd adaptation in different host plants.

Accumulating knowledge of these gain-of-function mutants in different hosts may help to determine and characterize host specific factors for systemic trafficking of PSTVd as a mechanism underlying PSTVd host adaptation. In addition, cataloging both gain-of-function and loss-of-function mutants might be useful to shed light on the not-yet-determined 3-D structure of loop 19, and a host factor that recognizes loop 19 and/or loop 19* sequence/structure.
Table 3.1 Sequencing of C227U progenies from systemic leaves of infected plants.

<table>
<thead>
<tr>
<th>No. of plants sequenced</th>
<th>Progeny sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>G134A/C227U</td>
</tr>
<tr>
<td>1</td>
<td>G134A/C227U/U27C/U200C</td>
</tr>
<tr>
<td>1</td>
<td>G134A/C227U/G67C</td>
</tr>
<tr>
<td>1</td>
<td>G134A/C227U/G319U</td>
</tr>
<tr>
<td>1</td>
<td>G134U/C227U/U200A</td>
</tr>
<tr>
<td>1</td>
<td>G134U/C227U</td>
</tr>
</tbody>
</table>

Newly acquired mutations through systemic trafficking are highlighted by bold letters. G134A and G134U mutations are highlighted with red and blue colors.
Table 3.2 Sequencing of loop 19* mutant progenies from systemic leaves of infected plants.

<table>
<thead>
<tr>
<th>Original mutations</th>
<th>Trafficking</th>
<th>Progeny sequences (No. of plants)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G134A/C227U</td>
<td>6/8</td>
<td>G134A/C227U (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G134A/C227U/U240C (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G134A/C227U/C148U (1)</td>
</tr>
<tr>
<td>G134U/C227U</td>
<td>7/8</td>
<td>G134U/C227U (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G134A (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G134A/C93-G-C94 (1)</td>
</tr>
<tr>
<td>G134A</td>
<td>9/10</td>
<td>G134A (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U24C/A107G (1)</td>
</tr>
</tbody>
</table>

Original inoculum-sequences recovered from upper leaves are highlighted by bold letters.
Table 3.3 Sequencing of trafficking-competent loop 19 mutant progenies from systemic leaves of infected plants.

<table>
<thead>
<tr>
<th>Inoculum sequence</th>
<th>Trafficking</th>
<th>No. of plants sequenced</th>
<th>Progeny sequences (No. of plants)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A135C</td>
<td>12/12</td>
<td>4</td>
<td>A135C (4)</td>
</tr>
<tr>
<td>A135G</td>
<td>2/12</td>
<td>2</td>
<td>A135G (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A135G/U137A/138C139 (1)</td>
</tr>
<tr>
<td>A135G/C227A</td>
<td>8/8</td>
<td>3</td>
<td>A135G/C227A (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C227A (2)</td>
</tr>
<tr>
<td>A135U/C227U</td>
<td>4/8</td>
<td>3</td>
<td>A135U/C227U (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G134A/A135U/C227U (1)</td>
</tr>
<tr>
<td>C227A</td>
<td>11/12</td>
<td>4</td>
<td>C227A (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C227U (1)</td>
</tr>
<tr>
<td>C227G</td>
<td>4/12</td>
<td>3</td>
<td>C227G (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C227G/G98A/G241A (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C227G/C113-D (1)</td>
</tr>
</tbody>
</table>

Original inoculum-sequences recovered from upper leaves are highlighted by bold letters. Red color indicates the acquired mutations in loop 19* position.
Table 3.4 Sequencing of trafficking-alteration loop 19 mutant progenies from systemic leaves of infected plants.

<table>
<thead>
<tr>
<th>Inoculum sequence</th>
<th>Trafficking</th>
<th>No. of plants sequenced</th>
<th>Progeny sequences (No. of plants)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A135U</td>
<td>5/8</td>
<td>5</td>
<td>WT (5)</td>
</tr>
<tr>
<td>A135C/C227A</td>
<td>9/19</td>
<td>3</td>
<td>A135C/C227A/U355C (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G134A/A135C/C227U (2)</td>
</tr>
<tr>
<td>A135C/C227U</td>
<td>9/11</td>
<td>4</td>
<td>G134A/A135C/C227U(3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C227U/A170G/U187C</td>
</tr>
<tr>
<td>A135U/C227A</td>
<td>6/12</td>
<td>2</td>
<td>C227A (2)</td>
</tr>
<tr>
<td>C227U</td>
<td>&lt; 68/304</td>
<td>7</td>
<td>See Table 3.1</td>
</tr>
</tbody>
</table>

Red color indicates the acquired mutations in loop 19* position.
Table 3.5 Sequence variants of loop 19/loop 19* and its equivalent regions in PSTVd natural variations.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Accession</th>
<th>Number of sequences</th>
<th>2-D structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' GGGAGUG … CACCCU 3'</td>
<td>N/A</td>
<td>152</td>
<td>5'-GGG 3'-UCC CAC-5'</td>
</tr>
<tr>
<td></td>
<td>AF483470</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5' GGAAGUG … CACCCU 3'</td>
<td>AY372398</td>
<td>3</td>
<td>5'-GG 3'-UC CAC-5'</td>
</tr>
<tr>
<td></td>
<td>EF459701</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5' GGGAGC … CACCCU 3'</td>
<td>X17268</td>
<td>1</td>
<td>5'-GG 3'-CC GA C-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Loop 19 (A135/C227) and equivalent nucleotides are highlighted by red color. Loop 19* (G134/C228) and equivalent nucleotides are highlighted by green color.
Figure 3.1 Secondary structures of PSTVd loop 19 and its mutants. The figure on the top shows the most energetically favorable secondary (2-D) structure of WT PSTVd predicted by mfold. Loop 19 is highlighted by a red bar. 2-D structures of A) WT-loop 19 (nucleotide A135/C227) and B) C227U mutant are shown. 2-D structures of loop 19* (nucleotide 134/228) in C) G134A/C227U mutant and D) G134U/C227U mutant are shown. A newly acquired mutation through systemic trafficking is highlighted by green color. U in position 227 is highlighted by red color.
Figure 3.1
Figure 3.2 Systemic trafficking analyses on C227U mutant. A) The illustration shows the systemic trafficking assay system and the summary of the results on C227U mutant. WT PSTVd traffics systemically in N. benthamiana without changing its sequence, whereas about 80% of C227U inoculated plants did not show systemic trafficking. Systemically trafficking populations of C227U progenies were observed with a sequence alteration in G134A to A or U. B) Representative RNA gel blots are shown. Total RNAs extracted from systemic leaves were pooled from 32 plants inoculated by WT (lane 1) or C227U-PSTVd (lane 2-11) and loaded into each lane. In the upper row of the figure, the bands show c-PSTVd. In the lower row of the figure, rRNAs are shown as a loading control.
Figure 3.3 Systemic trafficking analyses on loop 19* mutants. RNA gel blots to detect A) G134A/C227U, B) G134U/C227U, and C) G134A progenies in systemic leaves. Total RNA extracted from the systemic leaves of an individually inoculated plant was loaded into each lane. The leftmost lane shows the sample from a WT-PSTVd inoculated plant, and other lanes show mutant inoculated samples. The number of systemic trafficking out of inoculated plants is shown in the parenthesis. In the lower row of the figure, rRNAs are shown as a loading control.
**Figure 3.4** Secondary structures of loop 19/loop19* variations observed in the systemic trafficking analyses. Secondary structures of indicated mutants are shown: A) WT, B) G134A, C) G134A/A135U/C227U, and D) G134A/A135C/C227U. Nucleotides that are different from WT are highlighted by red color.
Figure 3.5 Systemic trafficking analyses on PSTVd loop 19 mutants. A) The 4 × 4-matrix shows a series of 15 mutants in loop 19. The WT sequence (A135/C227) is highlighted by a green background. Blue background indicates trafficking-competent mutants, in which the inoculum sequence was maintained in systemic leaves. Yellow background indicates trafficking-alteration mutants. Red background indicates the mutants that were never observed in systemic leaves in this study. The number in each box indicates the number of plants in which PSTVd exhibited systemic trafficking out of the total number of plants inoculated. The percentage “< 20%” for the sequence A135/U227 (C227U mutant) is the result from 320 inoculated plants. B) Representative RNA gel blots are shown. Each lane represents total RNA extracted from systemic leaves pooled from 4-8 plants inoculated with WT or each mutant indicated. The white spots for the A135C and A135U/C227U lanes indicate a saturated signal. rRNAs are shown as a loading control.
Figure 3.5
Figure 3.6 RNA gel blots for systemic trafficking analyses on PSTVd loop 19 mutants. Each lane represents an RNA sample from systemic leaves of a plant individually inoculated with WT or each mutant indicated. The number in the parenthesis indicates the number of trafficking events observed in this study out of the number of inoculated plants (see Figure 3.5A). The asterisk indicates the RNA extracted from WT inoculated plants. In all experiments, rRNA bands are shown as a loading control.
Figure 3.7 Replication efficiencies of loop 19 mutants. The bar graph represents the replication-efficiency of each loop 19 mutant as percentages of WT set to 100%. Blue columns represent trafficking-competent mutants. Yellow columns represent trafficking-alteration mutants. Red columns represent that the mutant was never detected in systemic leaves. The intensities of c-PSTVd bands of each mutant in 3 independent RNA gel blot experiments shown in Figure 3.8 were quantified, and the mean values and standard deviations are presented.
Figure 3.8 c-PSTVd signal on the inoculated leaves. RNA gel blots show c-PSTVd signal of each of the loop 19 mutant progenies in inoculated leaves as an evidence of replication. The results from 3 biological replicates are shown. Total RNA extracted from an individually inoculated plants was loaded onto each lane. In the upper row of the figure, the bands show c-PSTVd. In all experiments, rRNA serves as a loading control.
CHAPTER 4

CHARACTERIZATION OF A PLANT RNA LIGASE

ESSENTIAL FOR VIROID REPLICATION
4.1 ABSTRACT

Viroids are the smallest plant pathogens known to date. They are single-stranded and covalently closed RNA molecules that are circular in nature. They do not encode any proteins and are not encapsidated; therefore, they must rely on host factors to establish infection (i.e., replication and intercellular trafficking). The final step in viroid replication is the intra-molecular ligation of the linear genomic RNA into circular form, which requires a host derived RNA ligation activity. However, the RNA ligase required for viroid circularization as well as other subviral agents has not yet been identified. This chapter describes a chromatographic purification scheme used to identify the plant derived RNA ligase by using Potato spindle tuber viroid (PSTVd) as a substrate. The optimized purification condition, the multi-step chromatography processes, and the identified candidates by mass spectrometry are presented. Identification of the RNA ligase for viroids not only will enhance the understanding of the replication of viroids, but also will elucidate yet unidentified subviral substrates for higher eukaryotic RNA ligase.

4.2 INTRODUCTION

RNA ligases play critical roles in various biological processes, such as repairing tRNA breakage (Amitsur et al., 1987), splicing tRNA intron (Abelson et al., 1998; Englert and Beier, 2005), non-spliceosomal splicing of mRNA (Sidrauski et al., 1996),
and RNA editing (Schaufer et al., 2001). RNA ligases, together with the DNA ligases
and RNA capping enzymes, are grouped in the superfamily of covalent nucleotidyl
transferases (Shuman and Schwer, 1995). Knowledge of the reaction mechanisms and
substrate specificity of this group of enzymes has been applied to the field of current
biotechnology and molecular biology. As a result, these enzymes are widely utilized as
an essential part of polynucleotide probing and sequencing.

Although RNA ligases vary in their sequences, functional domains, and
substrates, they share catalyze the same reaction that creates a phosphoester bond
between 3'-hydroxyl (3'-OH) and 5'-phosphate (5'-PO₄) termini of two different RNA
strands (or one RNA strand). This enzymatic reaction is composed of three steps. First,
the enzyme is adenylated to form an enzyme-adenosine monophosphate (AMP)
intermediate (Figure 4.1A). Second, the AMP transfers from the intermediate to the 5'-
PO₄ end of a polynucleotide (Figure 4.1B). Third, the 3'-OH attacks the 5'-PO₄, which
joins two ends, and releases AMP (Figure 4.1C).

RNA ligases have two distinct families - Rnl1 family and Rnl2 family - that differ
in their polynucleotide substrate specificity (Nandakumar et al., 2006). Rnl1 can ligate
the broken ends of single stranded RNA, while Rnl2 repairs them in a double stranded
RNA. The type member of Rnl1, bacteriophage T4 RNA ligase 1 (T4Rnl1) repairs
broken tRNA anticodon loop (Amitsur et al., 1987). The other member of Rnl1, which is
found in yeast (Saccharomyces cerevisiae), and called tRNA ligase (ScTrl1), is involved
in intron splicing of tRNA (Abelson et al., 1997). ScTrl1 can use 2', 3'-cyclic phosphate
(2'3'-cyclic PO₄) and 5'-OH as substrates because it has the additional activities of cyclic
phosphodiesterase and GTP dependent kinase to generate 2'-phosphate (2'-PO$_4$), 3'-OH, and 5'-PO$_4$ ends. The central polynucleotide kinase domain of ScTrl1, which resembles T4Rnl1, joins 3'-OH and 5'-PO$_4$ ends together while leaving 2'-PO$_4$ in one end. It has long been known that both human HeLa cell extract (Zillmann et al., 1991) and plant wheat germ extract (Konarska et al., 1981; 1982) possess yeast-like tRNA ligation ability. However, no RNA ligase has been identified until the recent identification of Arabidopsis thaliana tRNA ligase (AtTrl), which is involved in tRNA intron splicing (Englert and Beier, 2005). Currently, AtTrl and its orthologue in rice are the only RNA ligases identified in higher eukaryotes.

Several eukaryotic RNA pathogens are supposed to rely on yet unidentified host RNA ligases. For example, a subviral agent, the RNA genome of hepatitis delta virus (HDV) needs to be circularized to complete its replication cycle (Reid and Lazinski, 2000; Taylor, 2006). Plant RNA pathogens, viroids, also need circularization of its genome to complete their replication cycle (Ding, 2009). Ligation of viroids by using either wheat germ or Chlamydomonas reinhardtii protein extract has been shown (Kikuchi et al, 1982). However, the corresponding RNA ligase has not yet been identified for HDV and viroids.

Potato spindle tuber viroid (PSTVd) is an excellent experimental system to identify yet unknown RNA ligase for the eukaryotic RNA pathogens. PSTVd is a single-stranded RNA pathogen, which does not encode any protein. Upon its infection, PSTVd enters into the host-plant nucleus to replicate by asymmetric rolling circle mechanism (Branch and Robertson, 1984). The schematic representation and the detailed explanation
of this process were shown in Chapter 1 of this thesis. Briefly, the mature PSTVd is a covalently circular, single-stranded RNA, which is arbitrarily assigned as (+)-strand polarity. By using the (+)-circular form as an initial template, the complementary (-)-strand multimer is synthesized. Then, this (-)-multimer is used as an intermediate template to synthesize the (+)-multimer. Synthesis of both (+)- and (-)-multimers are supposed to be catalyzed by host DNA dependent RNA polymerase II (Mühlbach and Sänger 1979; Rackwitz et al., 1981; Schindler and Mühlbach, 1992). Finally, the (+)-multimer is cleaved and circularized by an unidentified host endonuclease and RNA ligase. Thus, the circularization of PSTVd is the last step of its replication.

By using cauliflower protein extract, our group observed strong ligation activity of unit length linear PSTVd (l-PSTVd) to circular-PSTVd (c-PSTVd) \textit{in vitro}, suggesting the presence of the RNA ligase for PSTVd circularization. This chapter describes a chromatographic purification to identify a host-plant RNA ligase required for PSTVd circularization. The optimized purification condition, multi-step chromatography processes, and the identified candidates by mass spectrometry are presented. Identification of the RNA ligase for viroids not only will enhance the understanding of the viroid replication, but also will elucidate yet unidentified subviral substrates for higher eukaryotic RNA ligase.
4.3 MATERIALS AND METHODS

4.3.1 Preparation of a unit-length PSTVd substrate

The substrate for the ligation assay, a unit length 1-PSTVd, was *in vitro* transcribed from *HindIII*-linearized pRZ6-2-Int template plasmid. This plasmid was constructed by Hu and colleagues (Hu et al., 1997), and was a gift from Dr. Robert Owens. *In vitro* transcription was performed by using the T7 Maxiscript kit (Ambion, Austin, TX) in which the dNTPs were doped with a small amount of \( \alpha^{-32}P \) UTP. The synthesized 1-PSTVd was purified by NucAway Spin Columns (Ambion) after DNase I digestion. The synthesized unit length 1-PSTVd substrate has 5'-OH and 2'3'-cyclic PO₄ termini due to spontaneous ribozyme cleavages at both the 5' and 3' termini of the transcript (Figure 4.2).

4.3.2 PSTVd ligation assay

The PSTVd ligation assay system was modified from tRNA ligation assay described by Englert and Beier (Englert and Beier, 2005). In 10 µl ligation buffer [20 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), pH 7.5, 50 mM NaCl, 4 mM MgCl₂, 5 mM dithiothreitol (DTT), 1 mM ATP (adenosine-5’-triphosphate), 2 mM phenylmethanesulfonylfluoride (PMSF)], the synthesized 1-PSTVd substrate was incubated with 2.5 µl protein fraction of each purification step described below. The incubation was performed for 30 min at 37°C. This ligation mixture was directly loaded onto 5% (w/v) polyacrylamide gel containing 8M urea (SequaGel system; National
Diagnostics, Atlanta, Georgia), and separated by polyacrylamide gel electrophoresis (PAGE) on constant 200 V for 1.5 h. Immediately after PAGE, the gel was exposed to Storage Phosphor Screen (Kodak Rochester, NY). The exposed screen was scanned by Molecular Imager FX (Bio-Rad, Hercules, CA), and the c-PSTVd signal was visualized and quantified using the software Quantity One-4.1.1 (Bio-Rad).

The ligation efficiency of each fraction was calculated by the signal intensity of c-PSTVd divided by that of the input. The total activity of the ligase was arbitrarily assigned as total activity = ligation efficiency × fraction volume (ml). The specific activity of the ligase was calculated as total activity divided by the total protein amount (mg). The purification fold was calculated by comparing the specific activity of each purification step. The protein amount was measured by “Bio-Rad Protein Assay” reagent (Bio-Rad) as followed by the manufacturer’s protocol. The schematic representation of the PSTVd ligase assay and the quantification method are presented in Figure 4.2.

4.3.3 Adenyltransferase assay

Adenylation of the purified protein fraction was performed essentially following the protocol used in the adenylation of wheat germ tRNA ligase (Englert and Beier, 2005). Briefly, 12 µl reaction mixtures [50 mM Tris [(tris(hydroxymethyl)aminomethane)-HCl (pH 7.5), 3 mM MgCl₂, 0.1% (v/v) Triton X-100, 5 mM DTT, 0.2 mM [α-³²P] ATP] were incubated with 10 µl of chromatographic fractions for 15 min at 37°C. The incubated samples were separated by Tris-glycine SDS
(sodium dodecyl sulfate)-PAGE (acrylamide:bis-acrylamide=29:1). The gel was made by following Molecular Cloning (Sambrook et al., 1989). Immediately after separation, the gel was exposed to a Storage Phophor Screen. The exposed screen was scanned by Molecular Imager FX, and the adenylated portion of the gel was visualized by the software Quantity One-4.1.1.

4.3.4 Preparation for the cauliflower protein extraction

Cauliflower (Brassica oleracea) florets were obtained from local stores. Flower buds (600 g) were excised using a razor blade, and immediately frozen by liquid nitrogen. The frozen samples were mixed with 60 g polyvinylpyrrolidone (Sigma, St. Louis, MO), and were ground to a powder by using a mortar and pestle. The ground mixture was suspended into 2.4 liter of extraction buffer (EB) [100 mM HEPES (pH 7.5), 50 mM NaCl, 3 mM MgCl2, 10% (v/v) glycerol, 0.1% (v/v) Triton X-100, 5 mM DTT, 0.2 mM PMSF, 0.2 mM benzamidine, 0.4 mg/ml pepstatin A, and 0.5 mg/ml leupeptin]. The homogenate was centrifuged at 10,000 g for 10 min at 4°C. The supernatant was collected and filtered by cheese cloth, and centrifuged at 10,000 g for 10 min at 4°C, again. The supernatant was collected as a crude protein extract (CE fraction) and further purified as described below.
4.3.5 Large scale purification protocols of PSTVd ligase from cauliflower flower buds

All the purification steps described below were conducted at 4°C except the following procedures conducted at room temperature: each column fractionation, the SDS-PAGE separation of the protein fractions, and the gel staining. All column chromatography steps were conducted by using an ÄKTA FPLC (GE Healthcare, Piscataway, NJ). The flow rate was 1 ml/min except for the HiPrep 16/60 Sephacryl S-200 HR column (GE Healthcare).

Step 1: Fractionation by ammonium sulfate precipitation

The CE fraction extracted from 300 g cauliflower flower buds was brought to 40% ammonium sulfate, and incubated for 30 with by gentle rocking. The precipitated proteins were removed by centrifugation for 15 min at 10,000 g. The supernatant was brought to 70% ammonium sulfate, incubated and centrifuged as above. The precipitated proteins were suspended by 200 ml of purification buffer (PB: [20 mM HEPES (pH 7.5), 3 mM MgCl2, 10% (v/v) glycerol, 0.1% (v/v) Triton X-100, 5 mM DTT, 0.2 mM PMSF, 0.2 mM benzamidine, 0.4 mg/ml pepstatin A, and 0.5 mg/ml leupeptin]) containing 50 mM NaCl (PB50), and filtered through 0.45 μm² and 0.22 μm² pore-size membranes (DURAPORE Membrane Filters; Millipore, Billerica, MA). This filtered fraction was named as the ammonium sulfate (AS) fraction.
Step 2: Multiple-step column chromatography purification by commercially available columns

All the columns used below were purchased from GE Healthcare. Before loading each fraction as indicated below, HiTrap Q FF (Q) column and HiTrap SP FF (SP) column were equilibrated with PB$_{50}$. HiTrap Heparin HP (Heparin) column was equilibrated with PB containing 200 mM NaCl. HiPrep 16/60 Sephacryl S-200 HR (S-200) column was equilibrated with S-column buffer [20 mM HEPES (pH 7.5), 200 mM NaCl, 3 mM MgCl$_2$, 5 mM DTT].

**Step 2-1: Q column purification.** The AS fraction was divided into five batches (40 ml each), and each of them was separately loaded into five 5 ml Q columns in tandem. These tandem Q columns were treated as a 25 ml column. After loading, the column was washed with 50 ml of PB$_{50}$, and fractionated with 70 ml linear gradient 100-800 mM NaCl in PB. The peak fractions containing PSTVd ligation activity were eluted between 400-600 mM NaCl. These fractions containing the ligase peak activity were pooled (Q fraction), and brought to 70% ammonium sulfate followed by a centrifugation for 30 min at 10,000 g. The precipitated protein pellet was temporarily stored at -80°C overnight.

**Step 2-2: SP column purification.** The Q fraction pellet was suspended with PB$_{50}$, and loaded into 3 × 5 ml SP columns. The three columns connected in tandem (treated as a 15 ml column) were washed with 30 ml of PB$_{50}$, and fractionated with a 60 ml linear gradient of 50-650 mM NaCl in PB. The peak fractions containing ligase activity (SP fraction) were eluted between 250-350 mM NaCl.
**Step 2-3: S-200 column purification.** The SP fraction was directly loaded into 2 × 1 mL heparin-sepharose columns connected in tandem. The heparin-sepharose columns were washed with 4 ml of PB containing 300 mM, and directly connected to S-200 gel filtration column through the FPLC system. The bound proteins were directly eluted into the S-200 column by 3 ml of PB containing 1 M NaCl, and fractionated by using a buffer containing 0.2 mM PMSF, 0.2 mM benzamidine, 0.4 mg/ml pepstatin A, and 0.5 mg/ml leupeptin. The PSTVd ligase activity eluted in fractions corresponding to the 100-150 kDa range (S-200 fractions). The flow rate of the fractionation was 0.7 ml/min.

**Step 3: tRNA column purification**

The tRNA column was made by coupling wheat tRNA (Sigma) to a 1 ml HiTrap NHS-activated Sepharose column (GE Healthcare) based on the manufacturer’s instructions. Before loading the protein fraction, the tRNA column was equilibrated with PB_{50}. Then, the pooled S-200 fraction was loaded, washed with 2 ml of EB_{50}, and fractionated with a 2.5 ml linear gradient of 50-550 mM NaCl in PB. The peak fractions containing ligase activity (tRNA fractions) eluted between 290-350 mM NaCl. About 5 µg proteins were collected after this purification step. The fractions containing PSTVd ligase activity were brought to 80% ammonium sulfate precipitation followed by a centrifugation for 30 min at 10,000 g. The precipitated protein pellet was stored in -80°C. Steps 1-3 was repeated to scale up the amount of tRNA fractions.
Step 4: Separation of the purified proteins by SDS-PAGE

About 10 µg of the tRNA fractions were collected by the previous step. The ammonium sulfate precipitated fractions were re-suspended in EB$_{50}$, and concentrated by Microcon YM-50 spin column (Millipore). The concentrated protein sample was separated by 7% SDS-PAGE. The gel was stained by Coomassie Brilliant Blue (CBB) staining (SimplyBlue SafeStain; Invitrogen, Carlsbad, CA). A fraction from each purification step was also run on a different 7% SDS-PAGE gel, and the purification profile was visualized by Silver Stain Plus Kit (Bio-Rad). Prestained SDS-PAGE Standards Broad Range (Bio-Rad) was used as a size marker.

4.3.6 Mass spectrometry analyses in Midwest Bio Services

The CBB stained bands were manually excised and sent to Midwest Bio Services, LLC (Overland Park, KS) for mass spectrometry analysis. In Midwest Bio Services, the proteins in these gel slices were digested by trypsin, and the peptides were extracted and detected by nano-Liquid Chromatography (LC)/ Tandem Mass Spectrometry (MS/MS). The data were analyzed by TURBOSEQUEST software to search available protein databases. Above protein identification procedures are described in the following website: http://www.midwestbioservices.com/.
4.4 RESULTS

4.4.1 Optimization of the protein extraction buffer condition

Initially, the buffer condition used by Englert and Beier to purify wheat germ tRNA ligase was tested [50 mM Tris–HCl (pH 7.5), 3 mM MgCl₂, 0.1% (w/v) Triton X-100, 4 mM, β-mercaptoethanol, 0.2 mM PMSF, 0.2 mM benzamidine, 0.4 mg/ml peptatin A and 0.5 mg/ml leupeptin] (Englert and Beier, 2005). However, the c-PSTVd signal was barely detectable by our PSTVd ligation assay (Figure 4.3A, lane 1-3).

To improve the protein extraction and the following purification, the ligation efficiencies of the fractions extracted by the different pH conditions were examined. Since the pH of Tris buffers change as a function of the different temperature, HEPES (which shows less variation) was chosen. Also, the concentration of HEPES was increased to 100 mM since 50 mM was not enough to buffer the acidic extracts from cauliflowers. As a result, the ligation efficiency increased by 2-fold compared to the initial condition (Figure 4.3A). Figure 4.3B shows that the 100 mM HEPES buffer in pH 7.5 had the highest ligation activity in this assay system between 6.5-8.5 buffer pH range.

Next, different concentrations of a reducing reagent, DTT, were tested. Five mM DTT had the highest ligation efficiency in the 2-15 mM range tested (Figure 4.3C). Compared to the initial condition, 1.5-fold increase of ligation efficiency was observed (Figure 4.2C, lane 1-3 and 4-6). In addition, 100 mM NaCl and 5% glycerol were added to the buffer in order to gain general stability of the protein. A slight increase of the ligation efficiency was observed in these conditions (Figure 4.2-D).
Based on the above optimizations, the extraction buffer condition was fixed as 100 mM HEPES (pH 7.5), 50 mM NaCl, 3 mM MgCl2, 10% (v/v) glycerol, 0.1% (v/v) Triton X-100, 10 mM DTT, 0.2 mM PMSF, 0.2 mM benzamidine, 0.4 mg/ml pepstatin A and 0.5 mg/ml leupeptin. Changing the condition improved the detectable ligation signal intensity in the CR fraction by 4.4-fold.

4.4.2 Purification of PSTVd ligase by multi-step column chromatography

The simplified flow chart of the purification scheme described in Section 4.3.5 is shown in Figure 4.4A. All the purification conditions (e.g., ammonium sulfate concentration, NaCl buffer concentration used to elute bound protein in each column, and the sequence of columns) were optimized based on small scale tests using the same matrix in a smaller column volume. Figure 4.4B shows the results of the each ligation assay with each purification fraction. In each lane, c-PSTVd bands were observed as the ligated product of the l-PSTVd substrate, suggesting the existence of PSTVd ligase in each fraction. Also, the results of the ligation assay using tRNA column fraction is shown in Figure 4.4C. The ligase activity peak from tRNA column chromatography was observed in fractions eluting around 300 mM NaCl. Table 4.1 shows the specific activity of the PSTVd ligase in each purification fraction, the total protein amount, the total activity, and purification fold calculated from the specific activity. The purification fold from the CR fraction to the S-200 fraction was calculated by setting the specific activity of the CR fraction as 1 (Fold 1). The S-200 fraction and the tRNA input fraction were
essentially the same material; however, the ligase activity was significantly decreased in the latter due to the storage in -80 °C. Therefore, the purification fold was calculated separately for the tRNA column chromatography, setting the specific activity of the tRNA input fraction as 1 (Fold 2). Total purification fold was calculated by the multiplication of Fold 1 and Fold 2. Thus, by this purification scheme, PSTVd ligase was purified by 800-fold. Figure 4.4D shows the protein profile of each purification fraction visualized by silver staining.

4.4.3 A 130 kDa band in tRNA fractions that possess PSTVd ligase activity was adenylated

In the tRNA fraction protein profile (Figure 4.4D, tRNA lane), many minor protein bands were still observed. To determine which bands were to be analyzed by mass spectrometry, an adenyltransferase assay was performed on the tRNA fractions. Since all the identified RNA ligases can be adenylated by this assay, the unidentified PSTVd ligase was also expected to be adenylated. Figure 4.5 shows the result of the adenylated protein profile on a 7% SDS-PAGE gel. A 130 kDa band was detected by autoradiography as an [α-32P] ATP adduct. The 130 kDa band showed up only in the fractions which had the detectable PSTVd ligation activity, indicating its containing PSTVd ligase. Interestingly, a 110 kDa band in the input fraction (S-200 fraction), which had weaker ligase activity compared to the eluted tRNA fractions was also adenylated. However, the 130 kDa band could not be detected in the input fraction probably due to
the relatively low abundance of the adenylated protein before tRNA column fractionation.

4.4.4 Failure of detecting a PSTVd ligase candidate in tRNA fractions by mass spectrometry analyses

The peak tRNA fractions of PSTVd ligase activity were pooled, and separated by 7% SDS-PAGE. About 10 µg of protein was separated on the gel, and visualized by the CBB staining compatible with the LC/MS/MS analysis (Figure 4.6A). The three bands highlighted by the arrow were excised and separately analyzed by LC/MS/MS. A faint band around 130 kDa, a band around 110 kDa, and a very strong band around 70 kDa were excised. The 130 kDa band was expected to contain PSTVd ligase since it corresponded to the size of the adenylated band. The other two bands were analyzed because of their abundance. In Figure 4.6B, the band pattern of a tRNA fraction in a different SDS-PAGE gel, visualized by the silver staining, is shown. The three bands equivalent to the ones analyzed by the LC/MS/MS analysis are indicated.

The proteins identified by the LC/MS/MS are summarized in Table 4.2. All of the listed proteins were identified by at least two peptides detected in the LC/MS/MS analyses. No known ligase was identified from the three bands. Also, a 130 kDa or similar-sized protein could not be identified from the 130 kDa faint band. The only protein identified from this band is A. thaliana ARGONAUTE 5 (AtAGO5; At2g27880), which was annotated to be 110 kDa by its genomic sequence. From the 110 kDa band, three proteins were identified: 1) AGO5, 2) A. thaliana NOL1/NOP2/sun family protein
(At2g22400, 78.6 kDa), and 3) A. thaliana pseudouridylate synthase (At3g04820, 89.8 kDa). From the 70 kDa band, A. thaliana eukaryotic initiation factor 3 gamma (EIF3-gamma) subunit family protein (At2g45730, 49.4 kDa) was identified. No protein listed above has been reported to contain ligase activity, ligase motif, or adenylation activity.

A. thaliana and cauliflower are both in the Brassicaceae family; therefore, many proteins are expected to be conserved and are similar in size between the two species. Given this, it is difficult to rationalize that AGO5, annotated as 110 kDa, is a *bona fide* 130 kDa protein which can be adenylated. At the same time, AGO5 was also identified in the 110 kDa band, which is more corresponding to its annotated size. Therefore, it is reasonable to rationalize that the AGO5 identified from the 130 kDa faint band was composed of a trace amount of trailed portion of 110 kDa product. Thus, at this point, I am not in favor of claiming the possibility that AGO5 is the PSTVd ligase. However, thinking about the high isoelectric point value of AGO5 (pl = 9.92), it is possible that AGO5 migrates slower than its expected size in a SDS-PAGE gel (V. Gopalan, pers. comm.). Still, the possibility of AGO5 to be the PSTVd ligase cannot be ruled out.

4.5 DISCUSSION

4.5.1 Revisiting the purification strategy

In this chapter, a purification scheme to identify PSTVd ligase by using multi-step column chromatography followed by LC/MS/MS is presented. Though the PSTVd ligation assay, a system to effectively test the ligase activity in a plant extract was
successfully established and optimized, the plant derived PSTVd ligase could not be identified. One of the plausible reasons of this failure is due to the insufficient purification in this trial. As described in the Section 4.4.2, the PSTVd ligase activity was purified by 800-fold, whereas the wheat germ tRNA ligase was purified by 6,000-fold to be identified (Englert and Beier, 2005). Even though the adenylationase assay suggests that the ligase exists as a 130 kDa protein, we could not detect any protein similar to this size by LC/MS/MS.

As described above, the possibility of AGO5 being the PSTVd ligase in question cannot be ruled out; however, there is no report for ligase or adenylationase activity of this protein. AGO proteins are generally known as integral players in small RNA-directed pathways. Though the function of AtAGO5 has not yet been characterized, its closed paralog, AtAGO1 (A. thaliana ARGONAUTE 1), has been well characterized. AtAGO1 is required to maintain the integrity of microRNA (miRNA) pathway (Vaucheret et al., 2004). AtAGO1 binds to miRNA and has intrinsic RNaseH activity to cleave mRNAs directed by the associated miRNA (Baumberger and Baulcombe, 2005; Qi et al., 2005). The RNA binding motif and RNaseH motif are conserved in AtAGO5, suggesting the shared mechanistic properties of these proteins. Given its RNA-binding property, it is possible for AtAGO5 to be associated with viroids; however, its ligase activity is less likely the function of AtAGO5. Further characterization of AGO5 needs to be done to fully understand this point.

Except AGO5, the identified proteins functions in tRNA modification. NOL1/NOP2/sun family protein homologues in other organisms are categorized as tRNA
(cytosine-5)-methyltransferase, and the amino acid sequence of EIF3-gamma subunit family protein is similar in that of the tRNA (adenine-N(1))-methyltransferase TRM6 subunit. The pseudouridylate synthase is also involved in tRNA pseudouridylate synthesis. Since the final step of the multi-step column chromatography was the tRNA conjugated column, it is reasonable that these tRNA-modification proteins were purified. As might be expected, a PSTVd conjugated column was tested to purify the ligase activity; however, no successful recovery of PSTVd ligase activity was achieved by using this column. Further optimization of this column will be needed, and may result in the higher purification fold.

4.5.1 The possibility of the plant tRNA ligase as a PSTVd ligase

It is possible that the plant tRNA ligase previously purified is a PSTVd ligase purified in this study. There are several lines of reasons to support this claim. First, AtTrl is 125 kDa, which is similar in size to the 130 kDa band adenylated. Second, though PSTVd cannot systemically infect A. thaliana, PSTVd can replicate in its protoplast single cell system, suggesting that A. thaliana possesses enzymes required for PSTVd replication including the necessary ligase (Daròs and Flores, 2004). Third, the unit-length PSTVd substrate in this study has -OH and 2’3’-cyclic PO₄ termini (see Section 4.3.1), which can be ligated by AtTrl. Fourth, the PSTVd ligase activity was purified by its ability to bind tRNA column. Thus, it is worth testing the AtTrl ligation ability with a
PSTVd substrate in the future. In this case, the PSTVd ligation assay will be conducted with the *bona fide* substrate, l-PSTVd purified from the PSTVd infected plants.

4.5.2 Structures of l-PSTVd 5' and 3' termini as substrates for the ligase

In the classic studies showing PSTVd ligation *in vitro* with the protein extracts from wheat germ and *C. reinhardtii*, it has been proposed that the Pospiviroids (a genus of viroid including PSTVd) have 5'-OH and 2'3'-cyclic PO₄ termini (Branch et al., 1982; Kikuchi et al., 1982). This is the reason for using the unit-length PSTVd substrate with 5'-OH and 2'3'-cyclic PO₄ termini in this study. However, a recent study showed that in *A. thaliana* where dimeric *Citrus exocortis viroid* (CEVd) was overexpressed, unit-length l-CEVd with 5'-PO₄ and 3'-OH termini was generated as a major product (Gas et al., 2008). These termini of linear-CEVd could not be ligated by AtTrl expressed in E. coli. Since CEVd is also a member of Pospiviroids, it is expected to share similar termini structures as PSTVd. The 5'-PO₄ and 3'-OH termini suggest the possibility of another existing RNA ligase yet to be identified in plants. Therefore, it is worth trying to purify PSTVd ligase activity by using 5'-PO₄ and 3'-OH PSTVd substrate in the future.

At the same time, Gas and her colleagues showed the generation of heterogeneous termini of CEVd in its natural host plant, gynura, which could be ligated by AtTrl (Gas et al., 2008). Therefore, in some plants, PSTVd is also expected to possess a variation of termini: either 5'-OH/2'3'-cyclic PO₄ or 5'-PO₄/3'-OH termini. It suggests the possibility that PSTVd might use two different ligases. It is also possible that other enzymes, such as
cyclic phosphodiesterase, polynuclotide kinase or phosphatase, modify PSTVd termini structures into the best substrate structures for the *bona fide* PSTVd ligase yet to be identified.

Interestingly, the band which migrates slower than that of c-PSTVd was occasionally observed as a result of the ligation assay as well as the c-PSTVd band (Figure 4.7). This slowly migrating band might be the multimeric PSTVd ligated together. It could also indicate the longer than unit-length ligation products, which are generated due to the unsuccessful cleavage of ribozymes designed as a portion of the unit-length l-PSTVd transcript (Figure 4.2). The successful cleavages of ribozymes generate the 5'-OH and 2'3'-cyclic PO₄ termini of the unit-length l-PSTVd substrate; otherwise, 5'-triphosphate and 3'-OH termini are generated by the activity of T7 polymerase. It is not clear if these 5'-triphosphate and 3'-OH termini can be used as substrates, or if any modification of the termini structure is required by other enzymes.

These yet unidentified termini structures of l-PSTVd make the investigation of PSTVd ligase complicated. Therefore, further investigation of *bona fide* termini structures of l-PSTVd must be done in the future.
Table 4.1 Purification fold of PSTVd ligase.

<table>
<thead>
<tr>
<th>Purification fraction</th>
<th>Protein amount (mg)</th>
<th>Total activity</th>
<th>Specific activity</th>
<th>Fold 1</th>
<th>Fold 2</th>
<th>Total fold</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR</td>
<td>691</td>
<td>771</td>
<td>1.12</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>27.0</td>
<td>133</td>
<td>4.93</td>
<td>4.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP</td>
<td>2.77</td>
<td>38.3</td>
<td>13.8</td>
<td>12.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-200</td>
<td>0.200</td>
<td>11.6</td>
<td>58.0</td>
<td>52.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA Input</td>
<td>0.200</td>
<td>1.93</td>
<td>9.64</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>0.00442</td>
<td>0.656</td>
<td>148.5</td>
<td>15.39</td>
<td>800</td>
<td></td>
</tr>
</tbody>
</table>


### Table 4.2 The identified proteins from LC/MS/MS analyses.

<table>
<thead>
<tr>
<th>Band size (kDa)</th>
<th>Matched protein sequences</th>
<th>Annotated MW (kDa)</th>
<th>Accession</th>
<th>Locus</th>
</tr>
</thead>
<tbody>
<tr>
<td>132 kDa</td>
<td>AGO5</td>
<td>111</td>
<td>30683679</td>
<td>At2g27880</td>
</tr>
<tr>
<td>110 kDa</td>
<td>AGO5</td>
<td>111</td>
<td>30683679</td>
<td>At2g27880</td>
</tr>
<tr>
<td></td>
<td>pseudouridylate synthase</td>
<td>79</td>
<td>145338120</td>
<td>At3g04820</td>
</tr>
<tr>
<td></td>
<td>NOL1/NOP2/sun family protein</td>
<td>89.8</td>
<td>30681697</td>
<td>At2g22400</td>
</tr>
<tr>
<td>70 kDa</td>
<td>eukaryotic initiation factor 3 gamma subunit family protein</td>
<td>49.4</td>
<td>18406875</td>
<td>At2g45730</td>
</tr>
</tbody>
</table>
Figure 4.1 A schematic presentation of the shared three step enzymatic reaction of RNA ligases. A gray bar represents an RNA ligase, and a black curved line represents an RNA strand with 3’-OH and 5’ PO₄ as indicated. An adenosine- 5’ -triphosphate (ATP) is represented as Appp, and an AMP is represented as Ap in this figure. A) The first step of the ligation: the ligase-AMP adduct is made by the adenylation of the ligase. B) The second step: the AMP is transferred to the 5’ terminus of an RNA strand. C) The third step: Two termini of the RNA strand join together.
Figure 4.2 A schematic presentation of the PSTVd ligase assay and the quantification of the ligation efficiency. A) A unit length l-PSTVd is transcribed from the T7 promoter in the plasmid construct with flanking ribozyme sequences. The termini structure of this transcript are 5'-triphosphate (PPP) and 3'-OH due to T7 polymerase activity. RZ1, 5'-ribozyme is a Hammerhead ribozyme. RZ2, 3' ribozyme, is a Paperclip ribozyme. Each ribozyme is to be spontaneously cleaved upon its transcription, and generate the l-PSTVd termini structures (5'-OH and 2'-3' cyclic PO₄). B) A schematic representation of the results of a ligation assay on a PAGE gel. Each lane represents a mixture of protein and in vitro transcribed PSTVd. Lane 1 shows the representative band pattern observed only with the in vitro transcripts. Each of four bands represents 1) l-PSTVd, 2) l-PSTVd and the uncleaved RZ1, 3) l-PSTVd and the uncleaved RZ2, 4) l-PSTVd with the uncleaved double ribozyme. Lane 2-6 show the ligation activity as indicated by the c-PSTVd band. The ligation efficiency of each lane is calculated by the band intensity of c-PSTVd divided by that of l-PSTVd as an input (Ligation efficiency = c-PSTVd/l-PSTVd).
Figure 4.3 PSTVd ligation assays to optimize the condition of the PSTVd ligase extraction. Proteins were extracted from cauliflower using different buffer conditions. The ligation assay with protein extracts in each of the different conditions was repeated three times and presented as three lanes next to each other. The band positions of c-PSTVd (C) and l-PSTVd (L) were indicated as above. A) Different buffers and buffer strength were tested. Lane 1-3: The ligation assay was conducted with proteins extracted in 50 mM Tris buffer. Lane: 4-6: The buffer was changed to 100 mM HEPES (pH 7.5). The ligation efficiency of the HEPES buffer was increased by 2-fold. B) Different pHs of the 100 mM HEPES buffer were tested. The buffer condition with a pH of 7.5 had the highest ligation efficiency among tested conditions. C) Different concentrations of DTT were tested. The buffer condition with 5 mM DTT had the highest ligation efficiency among tested conditions. D) Different concentrations of NaCl were tested. The buffer condition with 100 mM NaCl had slightly higher ligation efficiency than that of 0 mM NaCl.
Figure 4.4 Purification of PSTVd ligase. A) A simplified flow chart of the purification scheme. B) The PSTVd ligase assay for the each purification fraction. The name of the purification fraction is as indicated on the top of each lane. The band positions of c-PSTVd and l-PSTVd were indicated as C and L. C) The PSTVd ligase assay for tRNA fractions. The positions of input (IP) and the flow through (FT) fractions are indicated. The concentration of the NaCl in the elution fraction is schematically presented on the top of the gel. The three elution fractions corresponding to 290–350 mM NaCl concentration are indicated by asterisks, and these fractions show PSTVd ligase activity. D) The protein profile of each purification step is shown. The position of the size marker (M) and the protein fraction from each purification step are as indicated above. Except the tRNA fraction, each lane contains approx. 500 ng protein. The tRNA lane contains approx. 50 ng protein.
Figure 4.5 Adenyltransferase activity of a 130 kDa band in the tRNA fractions having the PSTVd ligase activity. In lanes 2 and 3, a 130 kDa band was visualized as an [α-32P] ATP adduct. The star indicates the fraction containing PSTVd ligase activity. In input fraction (IP), a 110 kDa band was visualized.
Figure 4.6 The protein profile of a tRNA fraction analyzed in LC/MS/MS analyses. A) The marker (M) and the pooled tRNA fractions having PSTVd ligase activity were separated on a 7% SDS-PAGE gel, and the protein profile was visualized by CBB staining. The three bands highlighted by the arrow were excised and analyzed by LC/MS/MS analyses. B) The protein profile of the tRNA fraction on a different 7% SDS-PAGE gel was visualized by the silver staining. The three bands equivalent to the ones analyzed by LC/MS/MS were indicated.
Figure 4.7 Detection of a band migrating more slowly than c-PSTVd. Occasionally, a band migrating more slowly than that of c-PSTVd was observed. The band position is highlighted by the arrow. The above ligation assay was conducted with the fractions from Q column chromatography. Each lane indicates a different fraction: Input (IP), FT (flow through), W (wash), and 1-11 (elution fractions with a NaCl conc. gradient). This slowly migrating band was often fractionated with the c-PSTVd ligation activity as observed in fraction 3-7. The positions of c-PSTVd (C) and l-PSTVd (L) are indicated as above.
CHAPTER 5

FUTURE PROSPECTS
5.1 SUMMARY

RNA intercellular trafficking plays a fundamental role for orchestrating gene regulation at the whole organism level. To uncover the yet unknown mechanism of RNA trafficking, it is necessary to identify and investigate the RNA motifs required for trafficking. In this thesis, RNA structures required for intercellular trafficking were identified and analyzed by using *Potato spindle tuber viroid* (PSTVd) as the experimental system. First, our study identified the three-dimensional (3-D) structural motif of PSTVd loop 6 and its specific function in mediating PSTVd trafficking between two different mesophyll tissues. Second, through the analyses of PSTVd loop 19 trafficking-defective mutants, we identified the evolution of loop 19*, a secondary (two-dimensional, 2-D) structure similar to loop 19 and in a close proximity to the loop 19 genomic location, by which PSTVd regained trafficking-competency. Taken together, these studies support our hypothesis that a unique RNA structural motif mediates trafficking across different cellular layers. In addition, the identification of loop 19* evolution opened the possibility of PSTVd to be used as an experimental system to track the mechanism of evolution of RNA motifs *in vivo*. Though we could not identify the first RNA ligase for the subviral RNA by using PSTVd as a substrate, our simple assay and the observation of strong ligation activity demonstrated a potential of PSTVd as an experimental system to investigate the mechanism of RNA ligation. Furthermore, our studies open up many outstanding questions to be investigated in the future. Below, these questions are presented with feasible experimental approaches as future prospects.
5.2 FUTURE PROSPECTS

5.2.1 Cellular factors to mediate PSTVd intercellular trafficking

An important question is what cellular factors physically interact with PSTVd to mediate its intercellular trafficking. Since the viroid does not encode any protein, it is necessary for it to usurp host cellular factors to achieve full infection. Therefore, understanding the interaction partners is a necessary step to fully understand the viroid infection process. Moreover, by knowing interacting factors to a specific RNA motif required for the intercellular trafficking of a viroid, the knowledge of the functional role and structure of the motif itself will be enhanced. So far, no cellular factors have been identified as a decisive player of viroid intercellular trafficking. About 20 proteins have been identified to physically interact with PSTVd; however, the functions of most proteins have not yet been elucidated (Ding, 2009). Recently several proteins bound to other viroid species, such as in Peach latent mosaic viroid, have been identified (Dubé et al., 2009); however, each of their biological function is still unclear. In the future, as many host cellular proteins as possible, which interact to PSTVd, need to be identified to understand the nature of RNA intercellular trafficking as well as other infection processes like replication. These proteins might be identified by protein purification using full-length or a part of PSTVd RNA as a bait, then subsequent functional confirmation can be done to find the function of PSTVd interacting proteins.
The ribosomal protein L16 (called “L10e” in archaeal and eukaryal ribosomes) is a promising candidate for protein mediating PSTVd intercellular trafficking. This protein was observed to bind to the highly conserved archaeal and bacterial 3-D structure in 23S ribosomal RNAs (rRNAs), which we proposed as the PSTVd loop 6 structural model (Chapter 2). The physical interaction of PSTVd and the L10e in *Nicotiana benthamiana* can be investigated by co-immunoprecipitation and also be tested by electrophoretic mobility shift assay *in vitro*. Since various loop 6 mutants were already generated, a nucleotide or structure requirement of L10e binding to loop 6, if any, can be easily scrutinized. Upon confirmation of their physical interactions, transgenic *N. benthamiana* for knockout, knockdown, and overexpression of this protein will be investigated to observe the difference in PSTVd intercellular trafficking capability in each plant.

### 5.2.2 Tertiary structure of PSTVd functional motifs

The 3-D structure of any full-length viroid RNA has not yet been reported. Therefore, it is useful and informative to identify and investigate an RNA 3-D structural motif in a viroid by using the available 3-D structural database as a reference with isostericity matrices. The 3-D structure of PSTVd loop 6 (Chapter 2), loop 7 (Zhong et al., 2007) and loop E (also called as loop 15) (Zhong et al., 2006) were deduced by this methodology and supported by the functional analyses using PSTVd mutants of each loop. Needless to say that knowing the 3-D structure of an RNA motif should greatly aid in understanding the mechanism of the functional properties of the motif. Therefore, in
order to further understand the function of each loop, knowing its 3-D structure is essential. Nonetheless, as of the moment, 3-D structures of most loops of PSTVd could not be deduced or supported as in the case of loop 19 and loop 19* (Chapter 3). How can the 3-D structure of loop 19, loop 19*, or other loop be identified?

Difficulty in deducing and/or supporting the 3-D structural model of PSTVd loops is attributed to following limitations. First, our methodology cannot detect the 3-D structure for the sequence variation which does not exist in the database. Therefore, it is difficult to identify both novel sequence variation of the same/similar structure and novel structure. However, if this is the case, such difficulty will be eventually resolved due to the increasing number of available new 3-D structures in the database. Second, the results of the mutational analyses fit the structural criteria only when the structure plays a decisive role in a function. In other words, when the sequence of RNA, binding proteins, and/or factors other than the RNA 3-D structure are the primary determinant of its functional property, the results of mutational analyses do not support the structural model deduced from the available 3-D structural database. To overcome these difficulties, the structure of each loop should be investigated directly by using nuclear magnetic resonance (NMR) and/or X-ray crystallography. Considering the trial and error nature of crystallization of a macromolecule, NMR analysis of a small segment of PSTVd is a reasonable and promising choice as a method to identify the 3-D structure of PSTVd loops. For example, in the case of loop 19, a short RNA strand, which folds into the 2-D structure identical to loop 19 and its flanking stem, will be constructed and subjected to NMR analyses. The structural data will be compared with the results of the functional
analyses (such as systemic trafficking or replication analyses described in Chapter 2 and Chapter 3) and available and/or new generation chemical probing data (Gast et al., 1996). We expect that the combinations of all these available methods will lead to the understanding of the most PSTVd loop structures required for systemic trafficking.

5.2.3 Evolution of RNA motifs for intercellular trafficking

Loop 19 and loop 19* studies lead to several outstanding questions regarding the evolution of this RNA motif required for intercellular trafficking. How was loop 19*, a gain-of-function loop evolved from the trafficking-defective mutant, selected in the cellular machinery as a functional replacement of the loop 19 wild type structure? This specific question addresses the fundamental biological question as to how an infectious RNA evolves and adapts to its host’s environment. Both functional and mechanistic detail of this evolved loop 19* will be further understood by identifying the 3-D structure of loop 19 and loop 19*. Also, re-examination of our trafficking-defective mutant collection (various “loop-closed” mutants as described in Zhong et al., 2008) may result in the generation of novel gain-of-function mutants other than loop 19*. The accumulation of the knowledge in these gain-of-function mutants may lead to the understanding of yet unknown host adaptive mechanisms of viroids.

The variety of observed mutations in loop 19/loop 19* region raised the question as to whether this region is a ‘hot spot” for mutation? If so, is the mutation rate in loop 19/loop 19* significantly higher than in other regions of PSTVd? Due to the small
number of sequences obtained in our study, we could not draw any conclusions on this issue; however, this issue will be resolved by using high-throughput sequencing methods, such as pyrosequencing and Illumina sequencing. Since these methods can obtain a massive number of sequences, it is expected to quantify the occurrence frequencies of the mutations, leading to the identification of a mutation hot-spot. Other than loop 19/loop19*, minor mutations, which will enhance the infectivity of the PSTVd, are expected to be elucidated by the high-throughput sequencing. These mutation hot spots also might be further understood with structural insights. Whether there is a specific structure even in a highly mutated region is in great interest to understanding the host adaptation mechanism of RNA-based pathogens.

5.2.4 Cataloguing RNA motifs for intercellular trafficking

One of the ultimate goals of our studies is to catalogue all RNA motifs mediating intercellular trafficking and to assign detailed function to each of them. Previous studies in our lab identified 11 PSTVd loops required for systemic trafficking (Zhong et al., 2008). Especially, loop 6 was shown to mediate PSTVd trafficking between two different mesophyll tissues (Chapter 2), and loop 7 was shown to mediate the trafficking from bundle sheath to phloem tissue (Zhong et al., 2007). In addition, the bipartite motif was shown to mediate PSTVd trafficking from bundle sheath to mesophyll cells (Qi et al., 2004). Thus, clearly, different RNA motifs can individually function to mediate PSTVd trafficking across different cellular boundaries. Therefore, further investigation of other
loops required for PSTVd systemic trafficking is expected to elucidate an RNA motif which mediates the trafficking through specific cellular boundaries that are yet unidentified. At the same time, it is highly possible to identify RNA motifs which are required to traffic through same cellular boundary. Ultimately, by accumulating the knowledge of various RNA motifs required for intercellular trafficking, it is expected to identify a minimal RNA unit sufficient to go through a certain cellular boundary. This sufficient RNA unit for intercellular trafficking can be used as a molecular tag for a particular RNA to traffic to a specific location, leading to the precise manipulation of RNA trafficking in multicellular organism.
REFERENCES


Cultivated grapevines represent a symptomless reservoir for the transmission of *Hop stunt viroid* to hop crops: 15 years of evolutionary analysis. PLoS One. 4, e8386.


APPENDIX A

RNA THREE-DIMENSIONAL MOTIFS

USED FOR MODELING OF PSTVD LOOP 6
Appendix A RNA three-dimensional motifs used for modeling of PSTVd loop 6. The five motifs share the same three-dimensional (3-D) structure and have the same pattern of base pairing interactions while their sequences are different. The detailed structural annotations and interactive superpositions can be viewed online at the interactive website: http://rna.bgsu.edu/WebFR3-D/Results/pstvd.
APPENDIX B

THE ALIGNMENT OF PSTVD NATURAL VARIANTS

SHOWING CONSERVATION OF LOOP 6 SEQUENCE
Appendix B The alignment of PSTVd natural variants showing conservation of loop 6 sequence. Sequence alignment of PSTVd natural variants shows the conservation of both loop 6 sequences (highlighted by yellow). The nucleotide numbers of are provided based on the PSTVd\textsuperscript{Int} strain. Asterisks denote conserved nucleotides.
Continued
Appendix B continued

Continued
| PSTvd.123 | 61—GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG—115 |
| PSTvd.124 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.125 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.126 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.127 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.128 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.129 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.130 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.131 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.132 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.133 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.134 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.135 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.136 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.137 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.138 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.139 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.140 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.141 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.142 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.143 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.144 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.145 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.146 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.147 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.148 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.149 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.150 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.151 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.152 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.153 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.154 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.155 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.156 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.157 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.158 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.159 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.160 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.161 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.162 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.163 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.164 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.165 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.166 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.167 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.168 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.169 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.170 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.171 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.172 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.173 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.174 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.175 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.176 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.177 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.178 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.179 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.180 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.181 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.182 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.183 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.184 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.185 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.186 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.187 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.188 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.189 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.190 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.191 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.192 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.193 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.194 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.195 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.196 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.197 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.198 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.199 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |
| PSTvd.200 | GAAGCGGCTGTTGGAAGATGCTGAAAATCTGAGGAACTG |

Continued
Appendix B continued
Appendix B continued

Continued
Appendix B continued
Appendix B continued
Appendix B continued

<p>| | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
</tr>
<tr>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
</tr>
<tr>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
</tr>
<tr>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
</tr>
<tr>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
</tr>
<tr>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
<td>351</td>
<td>TGGCTTCTT</td>
</tr>
</tbody>
</table>

175
APPENDIX C
MAFFT ALIGNMENT OF POSPIVIROID RefSeq Sequences
IN CLUSTAL FORMAT
Appendix C MAFFT alignment of Pospiviroid RefSeq sequences in Clustal format.

Pospiviroid sequence alignments provide support to the PSTVd loop 6 structural model. Asterisks denote conserved nucleotides.