ABSTRACT

Title of dissertation: CIS-ACTING ELEMENTS AND A CONFORMATIONAL SWITCH INVOLVED IN REPLICATION OF A POSITIVE-STRAND RNA VIRUS

Jiuchun Zhang, Doctor of Philosophy, 2006

Dissertation directed by: Professor Anne E. Simon
Department of Cell Biology and Molecular Genetics

Replication of positive (+)-strand RNA virus genomes is a fundamental process in a virus's life cycle. *Turnip crinkle virus* (TCV) and an associated satellite RNA (satC) share 151 nt of 3' terminal sequence, which are predicted to fold into four phylogenetically-inferred hairpins (from the 3' end: Pr hairpin, H5, H4b and H4a). Pr hairpin is part of the core promoter (Pr) required for satC (Song and Simon, 1995) and TCV (Sun and Simon, in press) negative (-)-strand synthesis. To identify other regulatory cis-acting elements throughout satC, individual deletions of six other predicted hairpins (H5, H4b, H4a, M1H, H6, and H2) were performed. The results show that these deletions significantly reduce accumulation of (+)-strand monomers and differentially affect accumulation of (+)-strand dimers and (-)-strands in Arabidopsis protoplasts.

Results from in vivo genetic selection and mutational analyses of satC H5 indicate that robust satC accumulation in vivo requires specific sequences in the large symmetrical internal loop (LSL) and a stable stem and specific base pairs in the lower stem. The upper stem-loop has considerable plasticity. Moreover, H5 may be involved in accumulation of both strands. Mutational analyses also suggest that the LSL and/or the 3'

terminus may have other functions in addition to forming a pseudoknot (Ψ_1), which is required for replication.

An RNA conformational switch from a pre-active structure to an active structure appears required to regulate initiation of satC (-)-strand synthesis (Zhang et al., 2006). Results from mutational analyses suggest that H4a and H4b function as a unit and Ψ_2, formed between H4b and sequence flanking the 3' side of H5 and whose disruption reduces satC accumulation in vivo, stabilizes the pre-active satC structure. In addition, an upstream element (DR) may help to promote the switch.

Step-wise conversion of satC and TCV 3' terminal homologous sequences into the counterpart's sequence revealed the importance of having the cognate Pr. The satC Pr is a substantially better promoter than the TCV Pr when assayed in vitro. These results suggest that the TCV Pr requires upstream elements for full functionality and that evolution of satC generated a Pr that functions more efficiently by itself.

CIS-ACTING ELEMENTS AND A CONFORMATIONAL SWITCH INVOLVED IN REPLICATION OF A POSITIVE-STRAND RNA VIRUS

by

Jiuchun Zhang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2006

Advisory Committee:

Professor Anne E. Simon, Chair
Professor Dorothy Beckett
Associate Professor James N. Culver
Associate Professor Jeffrey J. DeStefano
Senior Investigator Kim Y. Green

ACKNOWLEDGEMENTS

My parents, who are professors in Microbiology, are the most important influence on my career choice. I still remember how I was amazed when I, as a small kid, saw those teeny-tiny creatures under a microscope. I am grateful to them for their care, love and support. Their expectations have been inspiring me to do my best.

I would like to acknowledge my advisor Dr. Anne E. Simon for providing me with the opportunity to work in her lab. Her enthusiasm and dedication to science and her encouragement turned me from reluctantly to willingly doing my research. I greatly appreciate that she switched me to a new project in my third year. This made my next few years fruitful. Anne has been passing on her valuable experience and knowledge in research and many other aspects to me. She also offered enormous help in improving my spoken and written English as well as presentation skills. I want to thank her for being patient and understanding during my dissertation writing. I am also deeply indebted to my committee members, Dr. Dorothy Beckett, Dr. James N. Culver, Dr. Jeffrey J. DeStefano, and Dr. Kim Y. Green, for their support and advice. I would like to thank Dr. Bruce A. Shapiro for generating the secondary structure prediction for satC by using *MPGAfold* program. I want to thank Robert M. Stuntz for performing mutagenesis in the 3' side of the large symmetrical internal loop of satC H5 and helping me with sequencing. I would also like to express my gratitude to my colleagues Sohrab Bodaghi, Rong Guo, John C. McCormack, Vera Stupina, Xiaoping Sun and Fengli Zhang for their valuable discussion, friendship and sharing some unforgettable moments together. I want to thank Catherine P. Gretschel and Nancee D. Soni for their help in managing lab

business. Special thanks go to my friends Jianlong Wang and his wife Hongwei Zhou, Hancheng guan and his wife Yanlin Wang, Ben Liu and his wife Jinyun Chen, Wenling Bao and her husband Qiming He, Junlin Wu and his wife Pu Li, Lizhen Tao, Zhengliang Gao, Meihong Liu and Haichen Song for their friendship and help through the years.

I am grateful to my husband Guohua Zhang for his love, understanding and support, which have accompanied me in sunny and rainy days. I would also like to thank my daughters Tianyue and Tianchen, whose sweet smiles and warm hugs have always cheered me up when I felt blue and encouraged me to go through hardships. Heartfelt thanks go to many other people who contributed to the completion of this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2,4 D: 2,4-Dichlorophenoxyacetic acid

3'CSS: 3'-Complementary silencer sequence

3'PE: 3'-Proximal element

3'TE: 3' Cap independent translation element

5'PE: 5'-Proximal element

ADSL: Adjacent downstream stem loop

AMV: Alfalfa mosaic virus

BaMV: Bamboo mosaic virus

BMV: Brome mosaic virus

BNYVV: Beet necrotic yellow vein virus

BSA: Bovine serum albumin

BTE: BYDV cap-independent translation element

BVDV: Bovine viral diarrhea virus

BYDV: Barley yellow dwarf virus

CAM: Clamped adenine motif

CarMV: Carnation mottle virus

CCFV: Cardamine chlorotic fleck virus

CCS: Carmovirus consensus sequence

CMV: Cucumber mosaic virus

CNV: Cucumber necrosis virus

CP: Coat protein

CPB: Coat protein binding

CPMoV: Cowpea mottle virus

cre: Cis-acting replication element

CS: Circularization sequence

CSE: Conserved sequence element

CyRSV: Cymbidium ringspot tombusvirus

D: Dimer

diG: Defective interfering RNA G

DI RNA: Defective interfering RNA

DNA: Deoxyribonucleic acid

dpi: Days post inoculation

DTT: Dithiothreitol

EAV: Equine arteritis virus

E. coli: Escherichia coli

EDTA: Ethylene diamine tetraacetic acid

FHV: Flock house virus

FMDV: Foot-and-mouth disease virus

GaMV: Galingsoga mosaic virus

H2: Hairpin 2

H4a: Hairpin 4a

H4b: Hairpin 4b

H5: Hairpin 5

H6: Hairpin 6

HCRSV: Hibiscus chlorotic virus

HDV: Hepatitis delta virus

HIV: Human immunodeficiency virus

hpi: Hours post inoculation

HRV: Human rhinovirus

IRES: Internal ribosome entry site

JINRV: Japanese iris necrosis virus

LDFE: Long-distance frameshift element

LS: Lower stem

LSL: Large symmetrical internal loop

M: Monomer

M1H: Motif 1 hairpin

MDV: Qβ bacteriophage-associated midivariant RNA

MHV: Mouse hepatitis virus

miRNAs: MicroRNAs

MNSV: Melon necrotic spot virus

MP: Movement protein

MS: Murashige and Skoog

ncRNAs: Non-coding RNAs

NLS: Nuclear localization signal

NMR: Nuclear magnetic resonance

ORF: Open reading frame

PABP: Poly(A) binding protein

PCBP: Poly(rC)-binding protein

PCM: Protoplast culture medium

PEG: Polyethylene glycol

PFBV: Pelargonium flower break virus

PIM: Protoplast isolation medium

PNRSV: Prunus necrotic ringspot virus

Pr: Core promoter for negative-strand initiation

PRRSV: Porcine reproductive and respiratory syndrome virus

Ψ: Pseudoknot

PSNV: Pea stem necrosis virus

PSTVd: Potato spindle tuber viroid

PTB: Polypyrimidine tract-binding protein

PVP: Polyvinylpyrrolidone

PVX: Potato virus X

RCNMV: Red clover necrotic mosaic virus

RdRp: RNA-dependent RNA polymerase

RHDV: Rabbit hemorrhagic disease virus

RNA: Ribonucleic acid

RNP: Ribonucleoprotein

rRNA: Ribosomal RNA

RSE: Replication silencer element

satC: Satellite RNA C

satRNAs: Satellite RNAs

satRPV: Satellite RNA of Cereal yellow dwarf virus-RPV

SCV, Saguaro cactus virus

SD: Shine-Dalgarno

SDS: Sodium dodecyl sulfate

SELEX: Systematic Evolution of Ligands by Exponential Enrichment

sgRNA: Subgenomic RNA

siRNAs: Small-interfering RNAs

SINV: Sindbis virus

TA: Trans-activator

TABS: Trans-activator binding sequence

TBSV: Tomato bushy stunt virus

TCV: Turnip crinkle virus

TGGE: Temperature-gradient gel electrophoresis

TLS: tRNA-like structure

TMV: Tobacco mosaic virus

TRS: Transcription-regulating sequence

TYMV: Turnip yellow mosaic virus

UCS: Upstream complementary sequence

US: Upper stem-loop

UTR: Untranslated region

UV: Ultraviolet

YFV: Yellow fever virus

CHAPTER I

CIS-ACTING ELEMENTS AND CONFORMATIONAL SWITCHES INVOLVED IN RNA REPLICATION OF POSITIVE-STRAND RNA VIRUSES: AN OVERVIEW

Introduction

Positive (+)-strand RNA viruses include many important vertebrate, bacterial, fungal, insect and more than 75% of plant viruses (Hull and Davies, 1983; van Regenmortel et al., 2000). These viruses can cause serious diseases such as encephalitis, hemorrhagic fever, hepatitis and SARS in human and animals and severe crop losses. Studying the replication of (+)-strand RNA viruses will help us to control the diseases caused by these viruses.

The multiplication of (+)-strand RNA virus involves six fundamental steps, which can overlap chronologically: (i) Entry, a process mediated by virus-receptor interaction for animal viruses and mechanical destruction of the cell wall and perforation of the plasma membrane for plant viruses (Poranen et al., 2002); (ii) Uncoating, the events that either couple with or immediately follow viral penetration into the host cell to release the viral genome (Poranen et al., 2002); (iii) Translation, a process in which the uncoated viral genome serves as mRNA to produce structural and nonstructural proteins. For picornaviruses, flaviviruses and potyviruses, a single polypeptide precursor is synthesized

first and then is proteolytically processed into single structural and nonstructural proteins. For many other (+)-strand RNA virus, only nonstructural proteins such as the RNA-dependent RNA polymerase (RdRp) are synthesized from the genomic RNA, while structural proteins are synthesized from subgenomic RNAs (sgRNA); (iv) Replication of the viral genome, a process producing progeny RNAs (Figure 1.1); (v) Assembly of virions with the encapsidation of progeny RNA; and (vi) Release of progeny virus from the infected cell. For animal viruses, this process is mediated by virus-induced cell lysis, cell death followed by membrane degradation or budding. For plant viruses, this process requires movement proteins encoded by the viral RNA and usually involves transport of a viral nucleoprotein complex, and not an assembled capsid, from cell-to-cell through plasmodesmata.

Although much effort has been made to explore how (+)-strand RNA viruses replicate their genomes, our knowledge of viral RNA replication is still quite limited. Replication starts with the assembly of the viral replication complex on intracellular membranes, which usually induces membrane alterations such as formation of vesicles (Ahlquist et al., 2003; Miller et al., 2001; Rust et al., 2001). The viral replication complex is composed of the viral-encoded RdRp, other viral proteins, and host factors (Ahlquist et al., 2003; Barton et al., 1995; Buck, 1996; Chong et al., 2004; David et al., 1992; Goh, 2004; Lai, 1998; Noueiry and Ahlquist, 2003; Strauss and Strauss, 1999; Tomita, et al., 2003; van der Heijden et al., 2002). Assembly of the replicase may require assistance from cis-acting elements on viral RNA (Panaviene et al., 2004; Panaviene et al., 2005; Quadt et al., 1995; Vlot et al., 2001). Infecting (+)-strand RNA is recruited to the replication site by viral encoded proteins and may require facilitation host factors.

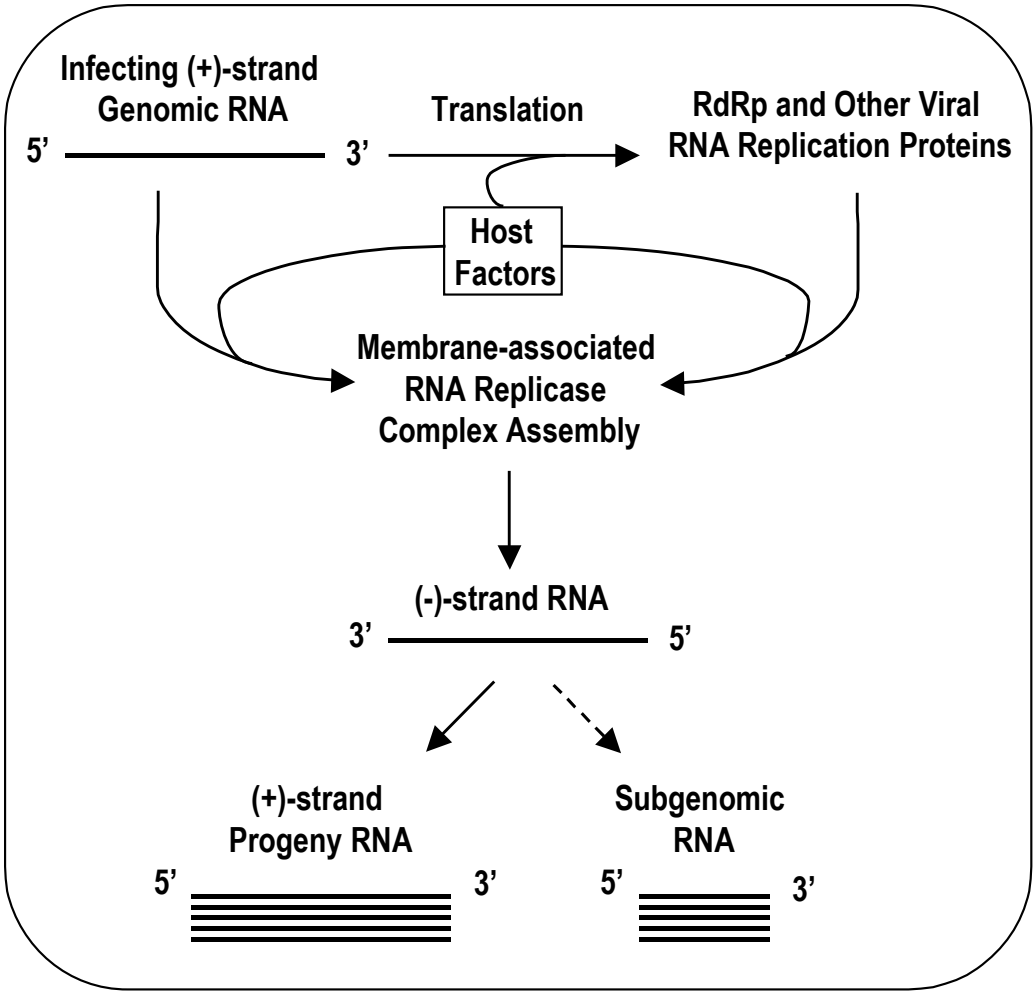


Figure 1.1 A general scheme of positive-strand RNA virus RNA replication. All steps shared by all positive-strand RNA viruses are depicted by solid arrows. Steps not shared by all (+)-strand RNA viruses are depicted with dashed arrows. Steps that might involved host factors are indicated.

RNA replication can be regarded as a two-step process (Buck, 1996; David et al., 1992): First, a complementary full-length (-)-strand intermediate is synthesized from the infecting (+)-strand RNA template; Second, large quantities of full-length (+)-strand RNA are synthesized from (-)-strand RNAs, which are either complexed with (+)-strands (Khromykh et al., 2003) or single-stranded (Axelrod et al., 1991; Garnier et al., 1980). Many (+)-strand RNA viruses also synthesize 3' co-terminal sgRNAs to express internal or 3' proximal genes (Miller and Koev, 2000; White, 2002).

All (+)-strand RNA viruses regardless of superfamily division, which are defined by common elements in their viral replicases (van Regenmortel et al., 2000), share common replication features: (i) the genomic RNA encodes RdRp that is not packaged within the virion; (ii) the genomic RNA serves as template for both translation and replication; (iii) RNA replication takes place on intracellular membranes; (iv) replication produces more (+)- than (-)-strand RNA, with ratio usually between 1:10 to 1:100 (Ahlquist et al., 2003). Therefore, positive-strand RNA viruses have to evolve mechanisms to coordinate translation and replication, recruit viral RNA template to particular intracellular sites, and efficiently synthesize asymmetric levels of (-)- and (+)-strand RNAs. Cis-acting elements located on viral RNA molecules play indispensable roles in these processes. Cis-acting elements may comprise specific sequences, secondary and tertiary structures and function through interactions with viral or cellular proteins (Dreher, 1999; Duggal et al., 1994). Recently, emerging evidence shows that RNA conformational switches can mask or expose cis-acting elements (Huthoff and Berkhout, 2001; Khromykh et al., 2001; Klovins et al., 1998).

In this chapter, I will present the current knowledge of two aspects of RNA

replication: (i) cis-acting elements involved in RNA replication; (ii) replication-specific RNA conformational switches.

Cis-acting Elements Involved in Viral RNA Replication

Cis-acting elements have been identified throughout viral genomes. In addition to core promoters located at the 3' ends of both strands to recruit viral RNA replicases to the transcription initiation sites for accurate RNA synthesis, many other elements appear required for different viruses to replicate efficiently. These include enhancers and silencers as well as RNA elements that are required for genome circularization, replication complex assembly, template recruitment by viral protein and host factors, primer synthesis or function in trans.

Core promoters

Core promoters are defined as the minimal sequence capable of directing detectable complementary strand synthesis. Core promoters are therefore able to recruit viral RNA replicases to the template for transcription initiation but may not be the sole element involved in RdRp recruitment. Using a reductionist approach, which combines deletion mutagenesis and in vitro RdRp transcription assays, core promoters for (-)-strand and (+)-strand synthesis have been identified for many (+)-strand RNA viruses.

Core Promoters for initiation of (-)-strand synthesis

Core promoters for initiation of (-)-strand synthesis are usually located within the 3' untranslated region (3' UTR). The 3' UTR of (+)-strand RNA viruses can terminate in a tRNA-like structure (TLS), poly (A) tail, or non-TLS heteropolymeric sequence (Table 1.1).

Core promoter for initiation of (-)-strand synthesis in viruses containing 3' TLS

The genome of the bromovirus, *Brome mosaic virus* (BMV) is composed of three RNAs. The 3'UTR of each genomic RNAs fold into a TLS (Felden et al., 1994) that can direct (-)-strand synthesis in vitro (Chapman and Kao, 1999; Dreher and Hall, 1988) and is required for viral RNA replication in vivo (Sivakumaran et al., 2003), and participates in regulation of translation (Barends et al., 2004). It also serves as a nucleation site for coat protein assembly and RNA encapsidation (Choi et al., 2002). The BMV TLS contains an extra stem-loop (SLC) compared to canonical tRNAs. Nuclear magnetic resonance (NMR) revealed SLC contains a flexible stem with an internal loop and a stable stem with a 5'-AUA-3' triloop in which the 5'A is fixed to the stem forming a clamped adenine motif (CAM) (Kim et al., 2000; Kim and Tinoco, 2001). Mutational analysis showed that the SLC is necessary and sufficient to bind the BMV replicase in vitro (Chapman and Kao, 1999; Choi et al., 2004; Dreher and Hall, 1988). CAM was shown to play a key role in this interaction since mutations disrupting the CAM were detrimental for RNA synthesis both in vitro and in vivo (Choi et al., 2004; Kim et al., 2000; Kim and Tinoco, 2001). Moreover, the SLC along with an 8 nt sequence containing 3'terminal CCA could direct de novo (-)-strand RNA synthesis. Mutations in

Table 1.1 Structure of the 3'UTR of (+)-strand RNA viruses

Structure of 3' end	Example of viruses
TLC	alfamovirus, bromoviruses, cucumoviruses, some furo-like viruses, hordeiviruses, tobamoviruses, tobraviruses, tobraviruses, tymoviruses
Poly(A)	benyviruses, caliciviruses, capilloviruses, carlaviruses, comoviruses, coronaviruses, potexviruses, potyviruses, picornaviruses, togaviruses
Stem-loop	carmoviruses, closteroviruses, coliphages, dianthoviruses, flaviviruses, luteoviruses, necroviruses, pestiviruses, sobemoviruses, tombusviruses, umbraviruses

either SLC or the 8 nt sequence abolish RNA synthesis, suggesting that the RdRp binding element and initiation element can function independently when present individually, but need to coordinate with each other when present together (Chapman and Kao, 1999). In addition, the BMV TLS can also interact with cellular proteins such as tyrosyl-tRNA synthetase, (ATP, CTP)-tRNA nucleotidyl transferase, and translation elongation factor EF-1a (Noueiry and Ahlquist, 2003). Aminoacylation is critical for RNA1 and RNA2 replication in vivo but not for RNA3 (Dreher et al., 1989; Duggal et al., 1994; Rao and Hall, 1991). Mutational analysis with chemeric RNAs showed that exchange of the 3' 200 bases among the three BMV RNAs led to abnormal ratios of the three RNAs although these sequences share high sequence and structure similarity: RNA1 and RNA2 differ from RNA3 at 11 positions and 1 base, respectively (Duggal et al., 1992). In vitro RdRp transcription analyses showed that RNA1 and RNA3 TLSs can direct similar levels of RNA synthesis (Chapman et al., 1998). Exchange of 5'UTR among BMV RNAs also severely reduced (-)-strand synthesis in vivo while a mutant containing both the 5' and 3' UTR from RNA2 and coding region from RNA1 could replicate as well as wild-type. Moreover, replication of RNA1 may also require sequences within the coding region (Choi et al., 2004). These results suggest that the replication of BMV genome requires interaction of TLC and upstream elements.

SLC-like structures (C-SLC) were also found in *Cucumber mosaic virus* (CMV), a cucumovirus that also terminates with a 3' TLS (Sivakumaran et al., 2000). Interestingly, BMV replicase can synthesize RNA from C-SLC in the context of the whole CMV TLC but not the minimal C-SLC motif suggesting that other features of the CMV TLC are also required for BMV replicase to recognize the CMV promoter.

The 3'UTR of *Tobacco mosaic virus* (TMV) contains a TLS at the 3' terminus preceded by three pseudoknots arranged in tandem (Felden et al., 1996). The minimal promoter for initiation of (-)-strand synthesis in vitro contains of the TLS and the 3'-most pseudoknot (Chandrika et al., 2000; Osman et al., 2000; Takamatsu et al., 1990).

The 3' end 82 nt of *Turnip yellow mosaic virus* (TYMV, a member of *Tymovirus*) genomic RNA also fold into a TLS that can be valylated by host valyl-tRNA synthetase (Dreher et al., 1992) and can regulate translation of the replicase domain-containing polyprotein by entrapping ribosomes (Barends et al., 2003). Unlike BMV, no specific promoter elements for (-)-strand synthesis have been identified within the TYMV TLC (Deiman et al. 1997; Singh and Dreher, 1997; Singh and Dreher, 1998). The minimal element controlling (-)-strand initiation is the 3'-terminal sequence –CCA-OH (Deiman et al. 1998; Singh and Dreher, 1998), which directs (-)-strand initiation opposite of the penultimate nucleotide (Singh and Dreher, 1997). The specificity for initiation may only require that a –CCA-sequence be sterically accessible by the viral replicase (Singh and Dreher, 1998). Recently, it was reported that interaction of eEF1A·GTP with the aminoacylated TYMV TLC strongly represses (-)-strand synthesis (Matsuda et al., 2004) while enhancing translation in vitro (Matsuda and Dreher, 2004). The authors proposed that this might occur at an early stage of viral infection to coordinate translation and replication (Matsuda et al., 2004).

The 3' terminal 145 nt region is conserved among the three genomic RNAs of *Alfalfa mosaic virus* (AMV) and was shown to contain promoter activity for (-)-strand synthesis (van Rossum et al., 1997). This region can fold into a TLS, which cannot be aminoacylated, and an immediate upstream triloop hairpin (hpE) with a 4 nt bulge

(Olsthoorn and Bol, 2002; Olsthoorn et al., 1999). Mutational analysis in RNA3 demonstrated that hpE is critical for (-)-strand synthesis in vitro and both the sequence and structure of the stem below the triloop are essential for function. Deletions constructed in the bulge of hpE and TLS or reciprocal changes in the position of hpE and TLS resulted in internal initiation (Olsthoorn and Bol, 2002). Recently, Olsthoorn and colleagues found that hpE by itself can functionally replace the sgRNA promoter both in vitro and in vivo (Olsthoorn et al., 2004). These results suggest that the AMV (-)-strand promoter and sgRNA promoter (described as below) are highly similar. The authors proposed that when hpE attaches to the TLC, the viral replicase might be forced to initiate (-)-strand synthesis at the 3' end of the template by the tertiary structure formed between the bulge of hpE and the TLS (Olsthoorn et al., 2004). A homologue of hpE can also be found at a similar position in the RNAs of ilarviruses, suggesting a similar function (Olsthoorn and Bol, 2002).

Core promoters for initiation of (-)-strand synthesis in viruses containing a poly(A) tail

Viruses containing poly(A) tails usually require both the poly(A) tail and additional elements in the 3'UTR for replication. *Sindbis virus* (SINV) is the type species of the *Alphavirus* genus (Strauss and Strauss, 1994). A 19 nt, AU-rich conserved sequence element (3′ CSE) located within the 3' UTR and the adjacent poly(A) tail have been identified as the core promoter for (-)-strand RNA synthesis (Hardy and Rice, 2005). Further mutational studies showed that the minimal length requirement of the poly(A) tail is at least 11 residues and the poly(A) tail must immediately proceed the 3' CSE (Hardy and Rice, 2005).

The 3' terminal 55 nt of the 3' UTR plus poly(A) tail were identified as the minimal cis-acting signal for (-)-strand synthesis in *Mouse hepatitis virus* (MHV) using a defective interfering RNA (DI RNA) containing a reporter gene (Lin et al., 1994). The terminal 42 nt that includes a stem-loop structure interact with four host proteins: mitochondrial aconitase, mitochondrial HSP70, HSP60 and HSP40 (Liu et al., 1997; Nanda and Leibowitz, 2001; Nanda et al., 2004). Both the primary and secondary structures are important for protein binding and mutations within this region that affected protein binding also reduced viral RNA accumulation in vivo (Johnson et al., 2005).

Core promoter for initiation of (-)-strand synthesis in viruses containing non-TLS, non-poly(A) sequence

The 3'-termini of many viruses fold into a series of hairpins. *Tomato bushy stunt virus* (TBSV) is the prototype of both genus *Tombusvirus* and family *Tombusviridae*. *Turnip crinkle virus* (TCV), which I have been studying, belongs to the same family but a different genus, *Carmovirus*. The core promoter for TBSV (-)-strand synthesis (gPR) is 19 nt containing a hairpin, 2 nt upstream sequence and a 3 nt single-stranded tail (Panavas et al., 2002). Similar to TYMV, –CCA can direct in vitro transcription by the RdRp of TCV and Qβ bacteriophge (Yoshinari et al., 2000) although the 3'-termini of these two viruses form a series of hairpins that are very different from TYMV, which terminate with a TLC.

Core promoters for initiation of (+)-strand synthesis

Promoters for initiation of (+)-strand synthesis are usually located within the 3' UTR of the (-)-strand. Therefore, the 5' untranslated region (5'UTR) of viral genome is important for (+)-strand synthesis. The minimal promoter for BMV (+)-strand synthesis is located within the 3' terminal 26 nt of the (-)-strand (Sivakumaran and Kao, 1999). In vitro studies showed that the initiation of genomic (+)-strand synthesis from the 3' end of BMV (-)-strands needs one nucleotide, prefer a guanylate, 3' of the initiation nucleotide. This nucleotide along with the +1 C and +2 A is crucial for replicase interaction while the secondary structure is not important (Sivakumaran and Kao, 1999; Sivakumaran et al., 1999). A +3 U in addition to +1 C and +2 A is also required for RNA accumulation in barley protoplasts (Hema and Kao, 2004). A sequence called cB box (complementary to the B box motif, discussed below) within this region in (-)-strands of RNA1 and RNA2 is important for replicase binding (Choi et al., 2004) and initiation of (+)-strand RNA synthesis in vitro (Sivakumaran and Kao, 2000). A single mutation in the core sequence 3' CCAA of cB box eliminated accumulation of both strands in barley protoplasts (Choi et al., 2004).

In the case of the TBSV DI RNA, the core promoter for (+)-strand synthesis (cPR11) was found to comprise only the 3'-terminal 11 nucleotides (Panavas et al., 2002; Wu and White, 1998). In vitro RdRp transcription assays showed that cPR11 less efficiently promoted RNA synthesis than the core promoter for (-)-strand synthesis (gPR) (Panavas et al., 2002). However, in vitro competition assays demonstrated that template containing cPR11 is a better competitor than template containing gPR. The authors proposed that the biased synthesis of (+)- and (-)- strand in vivo may be at least partially

due to the different ability of the initiation promoters to compete for limited replicase (Panavas et al., 2002). It is noteworthy that in these assays, the promoters were taken out of context, which excluded all other regulatory elements. Actually, a later study showed that the replicase assembly on TBSV requires not only the gPR, but also two other cis-acting elements (discussed below; Panaviene et al., 2005).

Subgenomic RNA Promoters

Many (+)-strand RNA viruses transcribe sgRNAs to express genes located downstream of the 5' proximal gene. Up to now, three mechanisms have been proposed for sgRNA synthesis (Miller and Koev, 2000; White, 2002): (i) Internal initiation, in which the viral replicase initiates sgRNA transcription internally on a full-length genomic (-)-strand RNA; (ii) Discontinuous transcription, in which the replicase stalls at an internal termination site during genomic (-)-strand synthesis, jumps to a more 5'-proximal site and continues (-)-strand synthesis to the 5' end of the genomic RNA. This product is then used as the template for sgRNA transcription; and (iii) Premature termination, in which a subgenomic-length (-)-strand is synthesized due to premature termination during genomic (-)-strand synthesis and serves as the template for sgRNA transcription.

The BMV sgRNA4 is transcripted from RNA3. The sgRNA promoter was originally identified in the intercistronic untranslated region in RNA3 containing an upstream A-U rich sequence, a poly(U) tract, the core promoter and a downstream A-U rich sequence (Marsh et al., 1988). In vitro RdRp transcription assays showed that only the 20 nt core promoter is required to accurately and efficiently direct sgRNA synthesis

(Adkins et al., 1997; Siegel et al., 1997). Within the core promoter, four nucleotides (–17, -14, -13, –11) along with the +1 initiation site were essential (Siegel et al., 1997; Siegel et al., 1998). Spatial changes between nucleotides –17 and –11 decreased RNA synthesis while changes between nucleotides –11 and +1 led to initiation at alternative positions in addition to the +1 initiation site in vitro, suggesting that the RdRp has some flexibility in its RNA binding site (Stawicki and Kao, 1999). As with genomic RNA (+)-strands, sgRNA accumulation in barley protoplasts specifically required +1C, +2A and +3U (Hema and Kao, 2004). Haasnoot et al. **(**2002) proposed that the sgRNA promoter of BMV is functionally equivalent to SLC because the sgRNA promoter can form a triloop hairpin and SLC can direct internal initiation. However, a more recent study showed that sgRNA promoter cannot replace SLC for (-)-strand synthesis (Ranjith-Kumar et al., 2003). On the contrary, the 15 nt terminal loop of SLC could functionally replace the sgRNA promoter (Sivakumaran et al., 2004). The essential nucleotides were involved in replicase binding while the hairpin structure was involved in a later step of replication (Sivakumaran et al., 2004).

The AMV sgRNA core promoter, a 37 nt sequence upstream of the initiation cytidylate, also contains a 24 nt triloop hairpin with a bulged adenylate flanked by a single stranded region (Haasnoot et al., 2000). Mutational analysis showed that the 3' A in the triloop and formation of the upper stem were essential for core promoter activity. In addition, nucleotide identity in the lower stem was also important since exchanging the two sides of the lower stem reduced RNA synthesis by more than 40% (Haasnoot et al., 2000).

The CMV sgRNA4 core promoter was defined to be within the 28 nt sequence upstream of the initiation cytidylate. It contains a hairpin with a 9 nt stem, a 6 nt purine-rich loop and a 4 nt single stranded region (Chen et al., 2000). Unlike the AMV sgRNA promoter, the stem-loop but not nucleotide identity is required, with the exception of one or more adenylates in the 5′ portion of the loop (Chen et al., 2000).

The promoters of the three sgRNAs of *Barley yellow dwarf virus* (BYDV) are 98, 143 and 44 nt long, respectively, and fold into very different secondary structures (Koev and Miller, 2000). In contrast to sgRNA1 and most of sgRNA promoters identified in other RNA viruses, sgRNA2 and sgRNA3 promoters are located down-stream of the transcription initiation site (Koev et al., 1999; Koev and Miller, 2000). Upstream sequences can affect the transcription start site from the sgRNA2 promoter (Moon et al., 2001).

It was proposed that TBSV sgRNAs are synthesized by a premature termination mechanism (Choi et al., 2001; Choi and White, 2002a; Lin and White, 2004; Zhang et al., 1999). The sgRNA2 promoter comprises 11 nt that are complementary to the 5' terminus of sgRNAs. This promoter is similar in structure to the promoter for genomic (+)-strand RNA synthesis, which it can functionally replace (Lin and White, 2004).

SgRNA promoters have also been identified for *Beet necrotic yellow vein virus* (BNYVV) RNA 3 (Balmori et al., 1993), *Cucumber necrosis virus* (CNV) (Johnston and Rochon, 1995), *Rabbit hemorrhagic disease virus* (RHDV) (Morales, et al., 2004), TMV (Grdzelishvili et al., 2000), TCV (Wang and Simon, 1997), TYMV (Schirawski, et al., 2000) and satellite RNA of *Bamboo mosaic virus* (BaMV) (Lee et al., 2000).

How the viral replication complex recognizes and interacts with different promoters

An open question concerning the replication of (+)-strand RNA viruses is how the viral replication complex recognizes and interacts with so many promoter elements that differ in sequence and structure. At least three possibilities have been proposed: (i) Viral RdRp has multiple RNA recognition domains. This is consistent with the observation that the N-terminal amino acid sequences within the viral RdRp 3D and/or the $3D^{Pro}$ domain of the protease/polymerase precursor $3CD^{Pro}$ are required for poliovirus (+)-strand RNA synthesis (Cornell et al., 2004); (ii) Different factors associated with RdRp are promoter specific. For example, synthesis of Qβ (+)- and (-)-strand RNAs requires different host factors (Barrera et al., 1993) and poliovirus protein 3A is required for (+)-strand but not (-)-strand RNA synthesis (Teterina et al., 2003); and (iii) Different processing of viral RdRp-containing polyprotein precursors. The four nonstructural proteins of SINV are translated as a polyprotein P1234, with P123 subsequently cleaved by nsP2 to form three replication complexes: P123+nsP4, which is unstable and functions in (-)-strand synthesis; nsP1+P23+nsP4, which is also unstable and functions in (-)-strand and genomic (+)-strand synthesis; and the fully processed stable nsP1+nsP2+nsP3+nsP4, which functions in genomic and subgenomic (+)-strand RNA synthesis (van der Heijden, 2002).

Replication enhancers and silencer

Although core promoters contain sequence necessary for the RdRp to initiate complementary strand synthesis, transcriptional efficacy is poor. In addition, core promoters are identified as the minimum template needed for transcription by viral RdRp

in vitro, which may not recapitulate the requirements for viral replication in vivo. Other RNA elements such as replication enhancers and silencers are usually required to modulate core promoter efficacy. These regulatory elements can be located both proximal and distal.

Viral RNA replication enhancers exhibit properties similar to DNA transcriptional enhancers such as the ability to increase basal levels of RNA accumulation significantly while function independently of orientation and position (Nagy et al., 1999). Replication enhancers are generally found on viral (-)-strands, contain sequence and/or structural features of core promoters, and can promote transcription in the presence of sequences resembling the transcription initiation site (Nagy et al., 1999; Panavas and Nagy, 2003, 2005; Ray and White, 2003).

Based on in vivo and in vitro studies, an 82 nt cis-acting element (region III) located within the 3'end of TBSV was identified as a replication enhancer (Ray and White, 1999, 2003; Panavas and Nagy, 2003, 2005). Region III is derived from both viral coding and noncoding regions and is composed of two stem-loops linked by a 23 nt single-strand region. Deletion of region III reduced prototypical DI RNA accumulation by ~10-fold in protoplasts. The function of region III was position and orientation independent and duplication of region III had no additive effect on RNA accumulation (Ray and White, 1999, 2003). In vitro RdRp transcription assays showed that the (-)-strand region III [RIII(-)] stimulated RNA synthesis by 10- to 20-fold while the (+)-strand region III stimulated RNA synthesis by 3-fold. In addition, the efficiency of stimulation from RIII(-) is higher from (+)-strand initiation promoter cPR11 than from (-) strand initiation promoter gPR. Therefore, this enhancer mainly functions on the (-)-

strand. Interestingly, the two stem-loops of RIII(-) were functionally redundant and interchangeable (Panavas and Nagy, 2003) and both of them could bind to the replicase (Panavas and Nagy, 2005). A 6 nt sequence (5'ACCUCU) within the single-stranded linker region was shown to interact with the cPR11 promoter (Panavas and Nagy, 2005). Based on these results, it was proposed that RIII(-) enhance (+)-strand synthesis by binding the replicase and then positioning the RdRp at the 3' end of the RNA through its interaction with the cPR11 promoter, thereby leading to accurate initiation of synthesis (Panavas and Nagy, 2005).

An interesting finding in BMV is the SLC within the TLS could function position-independently in vitro (Ranjith-Kumar et al., 2003). In addition, the initiation site was not required by the SLC for binding to the BMV replicase (Chapman and Kao, 1999). These data suggest that the SLC has enhancer-like activity: after binding to SLC, the BMV replicase complex can recognize the initiation site regardless of the distance between these two sites.

In addition to replication enhancers which up-regulate viral RNA replication, a replication silencer (repressor) which down-regulates viral RNA replication has been identified in vitro for TBSV (Pogany et al., 2003). This element (5'GGGCU) is located within the asymmetrical internal loop of a stem-loop structure (SL3) located on (+)-strands just upstream from gPR, which composes the extreme 3' terminal sequence AGCCC-OH of viral RNAs. In vitro RdRp transcription assays and oligonucleotide-based inhibition studies showed that the RNA-RNA interaction between the silencer and the gPR promoter inhibits (-)-strand RNA synthesis. This interaction is also required for DI RNA accumulation in protoplasts. The TBSV replication silencer may specifically

mask the promoter initiation sequences from the RdRp. The authors also propose that the function of replication silencers could be to promote asymmetrical synthesis of (-)- and (+)-strand RNA, provide recognition signal for viral replicase protein or host factors, protect 3' end of the RNA from degradation and coordinate translation and replication (Pogany et al., 2003).

RNA elements that may be required for genome circularization

Genome circularization, either through RNA-RNA interactions or protein bridges, might be a common feature of many (+)-stranded RNA viruses (Herold and Andino, 2001). It has been shown that genome circularization is involved in the regulation of translation, RNA replication and sgRNA synthesis.

Cis-acting elements involved in 5' and 3'end interactions

The long and highly structured 5'UTR of the picornavirus genome contains cis-acting elements involved in RNA replication as well as an internal ribosome entry site (IRES) that directs cap-independent translation initiation (Rohll et al., 1994). The 5'end 90 nt of poliovirus folds into a cloverleaf structure, which is conserved in enterovirus and rhinovirus (Revera et al., 1988; Zell et al., 1999). The 5' cloverleaf performs several functions depending on the interacting protein (Bedard and Semler, 2004). This sequence serves as the binding site for $3CD^{pro}$ and host poly(rC)-binding protein (PCBP) (Gamarnik and Andino, 1997; Gamarnik and Andino, 2000; Parsley et al., 1997). The binding of PCBP to the 5' cloverleaf may stabilize the genomic RNA after cleavage of VPg and enhance viral translation (Barton et al., 2001; Simoes and Sarnow, 1991). The

binding of 3CD to the 5' cloverleaf, in contrast, downregulates translation thereby switching the genomic RNA from its role as a mRNA to a template for replication (Gamarnik and Andino, 1998). The ternary ribonucleoprotein (RNP) complex was initially showed to be required in trans in (+)-strand synthesis (Andino et al., 1993). Recently, evidence demonstrated that this RNP complex is also required for initiation of (-)-strand synthesis (Barton et al., 2001; Herold and Andino, 2001; Lyons et al., 2001). Both 3CD and PCBP interact with cellular poly(A) binding protein (PABP), which binds to the 3' poly(A) tail of the viral RNA (Wang et al., 1999; Herold and Andino, 2001). It has been proposed that the genomic RNA of poliovirus circularizes through this RNA-protein-protein-RNA interaction to deliver 3CD, which binds to the 5' end of genomic RNA, to its functional site within the 3' poly(A) tail, thereby facilitating replication initiation (Barton et al., 2001; Herold and Andino, 2001).

While the 5' cloverleaf structure is involved in (-)-strand RNA synthesis, the poliovirus 3'UTR also plays a cell-dependent role in (+)-strand RNA synthesis. Deletion of the 3'UTR, which only slightly affected genomic RNA replication in HeLa cells, reduced (+)-strand accumulation to 10% of wild-type in SK-N-SH cells. More importantly, the ratio of (+)- to (-)-strands was reduced 6 fold compared to wild-type and there was no difference in stability between the wild-type and the mutant. This also suggested that cellular factor(s) may interact with the 3'UTR during (+)-strand synthesis (Brown et al., 2004; Todd et al., 1997).

In SINV, the 5'UTR was reported to be involved in both (+)- and (-)-strand synthesis (Frolov et al., 2001). The authors proposed that the viral replicase may first interact with the 5'UTR, and then shuttle to the initiation site at the 3' end by genome

circularization facilitated by the subunits of the eIF4F complex. The genome of coronaviruses may also circularized through a cap-eIF4E-eIF4G-PABP-poly(A) interaction since a minimum length of the 3' poly(A) tail, which correlates with the length required for efficient PABP binding, is required for achieving efficient replication (Spagnolo and Hogue, 2000). A similar strategy was also proposed for RNA replication of BMV by Diez and colleagues (Diez et al., 2000). It was found that the function of Lsm1p, a yeast protein required for efficient template selection during BMV viral RNA replication, could be replaced by addition of a poly(A) tail to the 3'end of the genomic RNA.

Different mechanisms were proposed for viruses bearing neither a 5' cap nor a 3' poly(A) tail. Conserved complementary circularization sequences (CS) in the 5' region of the capsid gene and the 3'UTR were predicted for many viruses in genus *Flavivirus* by *mfold* (Khromykh et al., 2001). Taking a reverse genetics approach, Khromykh and colleague (Khromykh et al., 2001) showed that separate mutations in the 5' or 3' CS of a *Kunjin virus* replicon RNA abolished viral replication in BHK cells while compensatory mutations in both CS restored viral replication. Similar results were also shown in *Yellow fever virus* (YFV) (Cover et al., 2003; Hahn et al., 1987). In the case of *Dengue virus*, in vitro studies demonstrated that the long-range interaction between 5' and 3' ends of the genome is required for (-)-strand RNA synthesis (You et al., 2001; You and Padmanabhan, 1999). Recently, Alvarez and collegues (Alvarez et al., 2005) visualized individual *Dengue virus* viral RNA in circular conformations using atomic force microscopy. Moreover, compensatory mutagenesis showed that additional sequences (named 5' and 3' UAR) as well as 5' and 3' CS were required for formation of RNA-

RNA complexes in vitro and viral RNA replication in vivo. Complex formation also requires the presence of Mg^{2+} suggesting that the 5'-3' interaction may also involve tertiary structures (Alvarez et al., 2005).

A set of cellular proteins called NFAR proteins bind to both the 5' and 3' ends of a pestivirus, *Bovine viral diarrhea virus* (BVDV). The bindings involved a stem-loop structure (hairpin Ia) at the 5' end and the variable portion (3'V) of the 3'UTR, both of which were thought to modulate translation and RNA synthesis (Isken et al., 2003, 2004; Yu et al., 2000). Therefore, the 5'-3' end interaction of BVDV may be mediated by NFAR proteins (Isken et al., 2004). Requirement of 5'-proximal sequences and structural elements for (-)-strand synthesis was also reported in *Aichi virus* (Nagashima et al., 2005), AMV (Vlot and Bol, 2003), and TBSV (Ray et al., 2003, 2004; Wu et al., 2001).

Cis-acting elements involved in other long-distance interactions

In addition to communication between the 5' and 3' ends, long-distance interactions between two RNA elements necessary for efficient RNA replication have also been revealed for some viruses. These interactions involve at least one RNA element located within the coding region.

Two long-distance interactions were found in the (+)-stranded RNA bacteriophage Qβ: an 8 base pairing that closes the ~1000 nt replicase domain RD2 (Klovins et al., 1998) and a pseudoknot formed by base pairing between 8 nt of the 15 nt single strand region linking RD1 and RD2 and the loop of the 3' terminal hairpin located ~1200 nt downstream (Klovins and van Duin, 1999). It is thought that these long-distance

interactions, especially the pseudoknot, bring the 3' end to the vicinity of the internal replicase-binding site (M site).

A kissing-loop interaction was found in arterivirus, *Porcine reproductive and respiratory syndrome virus* (PRRSV) (Verheije et al., 2002). A 34 nt conserved sequence within ORF7 folds into a putative hairpin with a 12 base terminal loop. Seven nucleotides of the loop base pairs with the loop sequence of a hairpin located ~300 nt downstream within the 3' UTR. This kissing-loop interaction was shown to be required for RNA replication.

A similar kissing-loop interaction essential for RNA replication was found recently for HCV (Friebe et al., 2005). This interaction forms between 7 nt of the loop of SL2 in a highly conserved 98 nt sequence within the 3' UTR (region X, X tail or 3'X), and its complementary sequence in the loop region of 5BSL3.2, a hairpin located ~250 nt upstream within the NS5B coding sequence.

Long-distance RNA-RNA interactions were also found for *Flock house virus* (FHV) (Lindenbach et al., 2002), *Potato virus* X (PVX) (Kim and Hemenway, 1999) and TBSV (Choi et al., 2001; Choi and White, 2002; Lin and White, 2004; Zhang et al., 1999) that regulate sgRNA synthesis. In the case of TBSV, synthesis of sgRNA1 requires a long-distance interaction between the terminal loop of a stem-loop located within the RdRp coding region and a linear element located ~1000 nt downstream that is only 3 nt 5' of the sgRNA1 initiation site (Choi and White, 2002). Synthesis of sgRNA2 requires a more complex network of at least three RNA-RNA interactions: a long-distance interaction similar to what is required for synthesis of sgRNA1 (Lin and White, 2004); a second long-distance interaction, located within the ~2000 nt loop-out region that

facilitates the first interaction by positioning its two base-pairing partners into close proximity and also stabilizes the first interaction via coaxial stacking; and a third one that helps to stabilize the second interaction (Choi et al., 2001; Zhang et al., 1999). The authors proposed that long-distance interactions involving elements located outside the sgRNAs for sgRNA synthesis might help control sgRNA synthesis, thereby properly regulating the production of viral proteins encoded by these sgRNAs (Lin and White, 2004).

RNA elements involved in replicase assembly

Some cis-acting elements are involved in replicase assembly. For example, active BMV RdRp can only be obtained from yeast expressing replication protein 1a and 2a along with RNA3 derivatives containing both the 3'UTR and intercistronic noncoding region (Quadt *et al*., 1995). The 3'UTR of either RNA1, RNA2 or RNA 3 of AMV was required to obtain RdRp activity from agroinfiltrated leaves of *Nicotiana benthamiana*, which was transiently expressing viral replication proteins P1 and P2 from T-DNA. In the absence of the 3'UTR, the RdRp lost activity without a significant reduction in RdRp levels (Vlot *et al*., 2001). Coexpression of a DI RNA along with the p33 and p92 viral replicase proteins was required to obtain highly active CNV replicase from yeast (Panaviene et al., 2004). Further mutational analysis revealed that three RNA elements were involved in assembly of the CNV replicase: an internal replication element, named the p33 recognition element [p33RE, also called RII(+)-SL or RII core] located within the p92 RdRp coding region; a 3' proximal replication silencer element (RSE) and the (-)-strand initiation promoter (gPR). These elements could form two alternative structures

[base pairing between RSE and gPR (RSE-gPR) or RSE and p33RE (RSE-p33RE)] that along with some primary sequences within these elements were thought to be important for replicase assembly (Panaviene et al., 2005). A model was proposed that the replicase assembly is initiated by binding of p33 dimers/multimers to p33RE (Pogany et al., 2005), then recruiting p92 and perhaps host factors (Panaviene et al., 2005). After p33 binding to RSE-gPR, the complex is associated with membrane that might induce structural changes in the RNA and perhaps also in the proteins. These changes may disrupt RSE-gPR and form RSE-p33RE, thus assembling the activated RdRp. The requirement for template RNA in RdRp assembly would enhance template specificity of the RdRp.

RNA elements for template recruitment

Positive-strand RNA viruses replicate in the host cell cytoplasm. Therefore, the ability to specifically and efficiently select viral RNA templates out of the pools of host RNAs is important for viral RNA replication. In BMV, template specificity in RNA3 replication depends on viral protein 1a (with domains homologous to methyltransferases and helicases), host factors (Lsm1p in yeast, Diez et al., 2000) and an approximately 150 nt region, which only functions in the (+)-sense orientation (French and Ahlquist, 1987), within the intergenic untranslated region (Pacha and Ahlquist, 1991; Traynor et al., 1990). This region was proposed to be involved in RdRp assembly (as discussed above; Quadt et al., 1995). Deletion of this region severely inhibits RNA3 accumulation in vivo and (-)-strand synthesis in vitro (French and Ahlquist, 1987; Quadt et al., 1995). In the absence of protein 2a (contains RdRp activity), this region functions with 1a to increase RNA3 stability but not translation in yeast (Janda and Ahlquist, 1998; Sullivan and

Ahlquist, 1999). Solution structure probing for this region revealed a stem-loop structure presenting a box B motif, which is conserved within the TΨC loop of tRNAs and is essential for RNA3 replication and 1a-induced stabilization, at its apex (Baumstark and Ahlquist, 2001; Sullivan and Ahlquist, 1999). Highly conserved box B motifs that are also present in the 5' UTRs of BMV RNA1 and RNA2 are critical for RNA2 replication and 1a-induced stabilization and membrane association of RNA2. Moreover, there was a good correlation between the level of 1a-induced membrane association and the level of accumulation in yeast for RNA2 mutants. These results were interpreted as 1a-induced viral RNA stabilization and membrane association reflect the process of template recruitment from translation to replication at an early stage in viral RNA replication (Chen et al., 2001).

Similarly, the internal replication element RII(+)-SL thought to be involved in TBSV RdRp assembly was also proposed to play a major role in recruitment of viral RNAs for replication (Monkewich et al., 2005; Pogany et al., 2005). Evidence provided to support this proposal include: (i) RII(+)-SL functions in (+)-strands, and is absent in sgRNAs; (ii) RII(+)-SL is required at an early step in TBSV RNA replication; (iii) translation inhibits RII(+)-SL activity (Monkewich et al., 2005); (iv) the replication protein p33 specifically binds to RII(+)-SL, which depends on the presence of a $C_{99}.C_{143}$ mismatch within a RNA helix in RII(+)-SL and the p33:p33/p92 interaction domain in p33; and (v) p33: RII(+)-SL complex formation in vitro correlates well with viral RNA accumulation in vivo (Pogany et al., 2005).

RNA element (cre) involved in primer synthesis

An internal cis-acting replication element (cre), proposed as a ~50-60 nt hairpin with a conserved AAACA in the loop segment, has been identified in some picornavirus genomes (Gerber et al., 2001; Goodfellow et al., 2000; Lobert et al., 1999; Mason et al., 2002; Mcknight and Lemon, 1996; Mcknight and Lemon 1998). In vitro evidence showed that the cres of poliovirus, *Human rhinovirus* 2 (HRV-2), HRV-14 and *Foot-and-mouth disease virus* (FMDV) serve as templates for uridylylation of VPg by the 3D RNA polymerase (Gerber et al., 2001; Nayak et al, 2005; Paul et al., 2000; Yang et al., 2002) possibly through a "slide-back" mechanism on the AAACA motif (Paul et al., 2003). This reaction is stimulated by $3CD^{Pro}$ or 3C (Nayak et al, 2005; Pathak et al., 2002; Paul et al., 2003). In vitro mutational analyses of poliovirus suggested that the cre-dependent uridylylation (forming VPgpUpU) is only required for the synthesis of (+)-strand RNAs and poly(A)-dependent uridylylation [forming VPg-poly(U)] may be involved in the synthesis of (-)-strand RNAs (Goodfellow et al., 2003b; Morasco et al., 2003; Murray and Barton, 2003). The importance of cre in viral RNA replication is also supported by in vivo studies (Mason et al., 2002; Rieder et al., 2000). The function of cre is position-independent (Goodfellow et al., 2000; Mason et al., 2002; Yin et al., 2003) and can perform in trans (Goodfellow et al., 2003a).

Cis-acting elements that function in trans

The first example of RNA-mediated trans-activation of transcription from a viral RNA was reported for *Red clover necrotic mosaic virus* (RCNMV) in 1998 (Sit et al., 1998). A 34-nt sequence within the MP ORF on RNA-2 was predicted to form a stem-

loop that is important for the replication of RNA-2 (Tatsuta et al., 2005). Compensatory mutations showed that the 8 nt loop region of this stem loop [trans-activator (TA)] can base-pair with an 8-nt element within the CP subgenomic promoter on RNA-1 [TA binding sequence (TABS)] (Sit et al., 1998). Structural models generated by *MC-Sym* computer program suggested that the stem of the RNA-2 TA hairpin stacked with the helix (5-6 bp) formed by intermolecular base pairing between RNA-1 and RNA-2 (Guenther et al., 2004). It was proposed that this TA-TABS interaction might promote premature termination during synthesis of complementary strands of RNA-1, thereby activating sgRNA synthesis from RNA-1 (Sit et al., 1998; Guenther et al., 2004).

In the case of BYDV, three sgRNAs are synthesized from the genomic RNA. A 105-nt BYDV cap-independent translation element (BTE) located in the 5'UTR of sgRNA2, which is unlikely a template for translation in vivo, inhibits BYDV RNA accumulation in protoplasts by trans-inhibiting translation of BYDV genomic RNA. BTE also slightly trans-inhibits translation of sgRNA1. The authors proposed that this trans-inhibitory effect leads to switching BYDV gRNA from its role as a mRNA to a template for replication and also making it available for encapsidation (Shen and Miller, 2004).

FHV contains a bipartite (+)-sense RNA genome. Replication of FHV RNA2 is transactivated by the sgRNA3, which is transcribed from RNA1 (Albarino et al., 2003; Eckerle and Ball, 2002). Mutations that prevented RNA3 replication also prevented transactivation, suggesting that replication of RNA3 may be required for transactivation of RNA2. The RNA3-dependent replication signal in RNA2 was mapped to the last 50 nt at the 3' end (Albarino et al., 2003). After being transactivated, RNA2 then suppresses RNA3 transcription from RNA1 and RNA3 replication (Eckerle et al., 2003; Zhong and

Rueckert, 1993). This counterregulation between RNA2 and RNA3 was implicated in coordinating the replication of the two viral genome segments (Eckerle and Ball, 2002).

Replication-specific Conformational Switches

In recent years, a prominent topic in RNA research is how RNA plays a regulatory role in cellular processes. Important findings include the discovery of: (i) microRNAs (miRNAs) and small-interfering RNAs (siRNAs), which are involved in gene-silencing and development (Bartel, 2004; Carrington and Ambros, 2003; Du and Zamore, 2005; Dykxhoorn et al., 2003; Hannon, 2002; He and Hannon, 2004; McManus and Sharp, 2002; Zamore and Harly, 2005); (ii) non-coding RNAs (ncRNAs), which also function in posttranscriptional regulation of gene expression to control cell growth and differentiation (Erdmann et al., 2001; Szymanski et al., 2005); and (iii) riboswitches, which control metabolism by specifically binding their target metabolite in the absence of protein factors leading to RNA conformational changes that affect transcription termination, translation initiation, RNA cleavage, or other aspects of protein production (Brantl, 2004; Mandal and Breaker, 2004; Nagel and Pleij, 2002; Nudler and Mironov, 2004; Sudarsan, et al., 2003; Tucker and Breaker, 2005).

The need for RNA viruses to switch between mutually exclusive processes for genome amplification suggests that RNA switches may also control different steps in the virus life cycle. RNA conformational switches control RNA dimerization in retroviruses, (Abbink et al., 2005; Berkhout et al., 2002; Dey et al., 2005; D'Souza and Summers, 2004; D'Souza and Summers, 2005; Greatorex, 2004; Huthoff and Berkhout, 2001;

Ooms et al., 2004a; Ooms et al., 2004b) and in vitro ribozyme activity in *Hepatitis delta virus* (HDV) (Harris et al., 2004; Ke et al., 2004; Tinsley et al., 2004), *Potato spindle tuber viroid* (PSTVd) (Baumstark et al., 1997) and the satellite RNA of *Cereal yellow dwarf virus-RPV* (satRPV) (Song et al., 1999; Song and Miller, 2004).

As described earlier, genomic RNAs of (+)-strand RNA viruses must initially assume a conformation that is recognized by cellular ribosomes for translation of viral products such as the RdRp. At some point, the RNA must switch to a form that is not available for translation but contains cis-acting elements recognized by the RdRp leading to initiation of (-)-strand synthesis (Gamarnik and Andino, 1998; van Dijk et al., 2004). Following reiterative synthesis of (+)-strands from (-)-strand templates, newly synthesized (+)-strands of some viruses may not be templates for further (-)-strand synthesis (Brown et al., 2004; Chao et al., 2002; Ishikawa et al., 1991), suggesting that these strands may need to adopt a structure that is incompatible with RdRp recognition (Figure 1.1). Below I discuss some examples of emerging evidence indicating that conformational switches play important roles in (+)-strand RNA virus replication.

Coordination of translation and replication

The translation of protein A of the single-stranded phage MS2 is regulated by the conformation of the untranslated leader. At equilibrium, this leader folds into a cloverleaf structure in which the Shine-Dalgarno (SD) sequence is base paired with an upstream complementary sequence (UCS) thus inhibiting translation of protein A (Groeneveld et al., 1995). During (+)-strand RNA synthesis, the RNA is trapped in a metastable structure (Poot et al., 1997; van Meerten et al., 2001) by formation of a small hairpin that prevents

formation of the 5’ end hairpin found in the equilibrium structure. The small hairpin presents the UCS sequence in its loop region, which delays the base pairing between the SD and UCS, allowing translation of the protein A (van Meerten et al., 2001). Later, the metastable structure is disrupted and the RNA folds to cloverleaf structure (van Meerten et al., 2001).

In the case of BYDV, cap-independent translation requires a kissing-loop interaction between stem-loop IV (SL-IV) in the 5’UTR and the 3’ cap independent translation element (3’TE) in the 3’UTR (Guo et al., 2001). Programmed –1 ribosomal frameshifting requires another long-distance base-pairing interaction between sequences within the bulge of the adjacent downstream stem loop (ADSL) and the loop of the long-distance frameshift element (LDFE), which is approximately 4 kb downstream from the shifty site (Barry et al., 2002). RNA molecules containing these two sets of long-distance base-pairing interactions are proposed to be the template for translation of ORF 1 and ORF 2, which encodes subunits of the viral RdRp (Barry et al., 2002). The newly translated RdRp initiates (-)-strand synthesis from the 3’ end and melts out the ADSL:LDFE and 3’TE:5’UTR base pairing, shutting down the frameshift and translation initiation sequentially. The ribosome-free RNA molecules are then used as the reiteration template for replication. When there is an excess of (+)-strand RNA compared with RdRp, the two long-distance base-pairing interactions reform, stimulating translation and frameshifting (Barry et al., 2002).

For some other viruses, coordination of translation and replication likely involves changes in viral RNA conformation in 3' regions of the genome, a process mediated by one or more unstable base-pairs between complementary short sequences located within

and outside of hairpins, that hide or expose the 3' terminus or permit the formation of important cis-acting structures. For example, Olsthoorn and colleagues (1999) proposed that the 3'UTR of AMV adopts two different structures. In one structure, the 3'UTR folds into a series hairpins that contain multiple CP binding sites (Bol, 1999; Bol, 2005). In the absence of CP, the 3'UTR adopts a pseudoknot structure that resembles a TLS (TLS conformer) by base pairing between sequences within hairpin A and D. This TLS conformer was required for (-)-strand RNA synthesis in vitro and RNA3 and RNA4 accumulation in P12 transgenic tobacco plants and protoplasts. Binding of CP to the 3' termini disrupts the formation of the pseudoknot resulting in a linear conformation (CP-binding conformer, CPB conformer) thereby inhibiting (-)-strand synthesis (Olsthoorn et al., 1999). CPB conformer is also involved in translation (Bol, 2005; Krab et al., 2005). It was proposed that the CP-triggered conformational switch in the 3'UTR of AMV might coordinate translation and replication in early stages of infection, and regulate asymmetric (+)-strand RNA synthesis in late stages of infection (Olsthoorn et al., 1999). A similar CP induced conformational switch model was also proposed for *Prunus necrotic ringspot virus* (PNRSV, Ilarvirus) (Aparicio et al., 2003). However, the Gehrke laboratory (Guogas et al., 2005; Petrillo et al., 2005) has recently presented evidence against this model based on the following results: (i) biochemical and crystal structural data indicate that CP binding leads to a more compact RNA conformation (Baer et al., 1994; Guogas et al., 2004); (ii) in a subgenomic luciferase reporter system, viral RNA replication was activated at low concentrations of CP but inhibited at higher concentrations of CP (Guogas et al., 2005); (iii) mutations that eliminated CP production or affected CP's RNA binding ability severely reduced viral RNA accumulation in vivo

(Guogas et al., 2005); (iv) CP can still bind to viral RNA in the presence of high concentration of Mg^{2+} proposed to stabilize the pseudoknot structure (Petrillo et al., 2005); and (v) compensatory mutations proposed to reestablish the pseudoknot do not restore viral RNA accumulation in nontransgenic tobacco protoplasts (Petrillo et al., 2005). Based on these findings, a 3' organization model was proposed: binding of CP to the 3' UTR results in a compactly organized 3' end that is functionally equivalent to the pseudoknotted TLS of other bromoviruses and is required for viral RNA replication (Petrillo et al., 2005). It is noteworthy that some experiments were done in different systems or using different conditions by these two laboratories although the same constructs were used. An explanation for these seemingly contradictory data is that they reflect different stages during viral replication. For example, a low ratio of CP to RNA (1:4, Olsthoorn et al., 1999) may reflect an early stage in infection while high concentrations of Mg^{2+} may prevent CP binding in vitro. A high ratio of CP to RNA (1:6.25-50, Petrillo et al., 2005) may reflect a later stage in infection and the presence of high concentration of Mg^{2+} may not be able to prevent CP binding in vitro. This is also consistent with the finding that lower concentrations of CP stimulate RNA replication while higher concentrations of CP inhibit RNA replication (Guogas et al., 2005). Similarly, activation of AMV (-)-strand synthesis may include two steps in the more natural system with limited amounts of virally-expressed RdRp: first, the 3' regions of AMV genomic RNAs assume an initial compactly organized structure mediated by CP; second, the initial structure converts to the pseudoknot conformer. Participation of one or both of the pseudoknot partners in required alternative pairings in the initial structure would preclude re-establishment of replication in the compensatory mutant. However, in

the more artificial transgenic system, the absence of CP would preclude formation of the initial compact structure, which may also not be required in the presence of high levels of nuclear-expressed RdRp.

The 3' UTR of BVDV contains a variable region (3'V) upstream of a conserved 3' region (3'C), which has been identified as a critical component of the (-)-strand initiation promoter (Yu et al., 1999). 3'V fold into two unstable stem-loop structures presenting multiple UGA box motifs (Isken et al., 2003). Structure probing and mutational analysis show that the correct formation of 3'V is important for IRES-mediated translation initiation, accurate termination of translation and replication (Isken et al., 2003, 2004). In addition, a stem-loop (hairpin Ia) located at the 5' terminus is also involved in both viral translation and replication (Yu et al., 2000). Moreover, cellular NFAR proteins interact with both hairpin Ia and the 3'V region (Isken et al., 2003). Based on these findings, Isken and colleagues (Isken et al., 2004) proposed that the 3'V region is involved in coordination of translation and transcription of BVDV: early in viral infection, circularization of the BVDV genome through the stem-loop Ia-NFAR-3'V interaction promotes translation initiation; 3'V then facilitate translation termination and the viral replication complex start to assembly at the 3' end, possibly changing the conformation of 3'V and further affecting the 5'UTR resulted in inhibition of translation initiation (Isken et al., 2004).

The 3' UTR of flaviviruses are predicted to fold into a series of stem-loop and pseudoknot structures (Olsthoorn and Bol, 2001; Shi et al., 1996). It has been shown that both the 3' terminal structure and circularization of the viral genome by long-distance interactions are important for viral replication (Bredenbeek et al., 2003; Cover et al.,

2003; Khromykh et al., 2001; Lo et al., 2003; Nomaguchi et al., 2004; Tilgner and Shi, 2004; You et al., 2001; Zeng et al., 1998). However, long-distance interactions could affect the 3' terminal structure (Khromykh et al., 2001; You et al., 2001). This implies that a conformational change must occur during replication. It is proposed that circularization of the viral genome may occur during or after the assembly of replication complex on the 3' terminal structure to inhibit translation initiation thereby coordinating translation and replication (Khromykh et al., 2001).

Synthesis of asymmetric levels of (+)- and (-)-strands

Positive-strand RNA viruses accumulate much more (+)- than (-)-strand RNA (Ahlquist et al., 2003). Asymmetric RNA synthesis may be achieved by several mechanisms: (i) the differential ability of (+)- and (-)-strand initiation promoters to compete for viral and/or host proteins that are involved for replication (Panavas et al., 2002); (ii) replication enhancers that are usually stronger in their (-)-strand orientation than in their (+)-strand orientation (Panavas and Nagy, 2003); (iii) trans-factors that either facilitate the enhancement of (+)-strand synthesis and/or the reduction in (-)-strand synthesis (De et al., 1996; Marsh et al., 1991; Satyanarayana et al., 2002; van der Kuyl et al., 1991); and (iv) shut-down of (-)-strand synthesis at an early stage in viral infection (Ishikawa et al., 1991).

Recent evidence obtained using several unrelated viruses suggests that RNA conformational switches may be needed to hide and expose viral 3' ends, which may temporally regulate (+)- and (-)-strand synthesis. For example, the 3' terminal 109 nt of BYDV contain three or four stem-loops that form a "pocket" structure in which the last 4

nt are embedded in a coaxially stacked helix thus making the 3’ end unavailable to the RdRp (Koev et al., 2002). It was proposed that this conformation might repress (-)-strand synthesis and favor (+)-strand RNA synthesis while an alternative structure with a free 3’ end is required for (-)-strand RNA synthesis (Koev et al., 2002).

Similarly, the 3’ terminal 5 nt (AGCCC-OH) of TBSV [termed 3’-complementary silencer sequence (3’CSS)] can base pair with RSE located in a 3’ proximal stem-loop (Na and White, 2006; Pogany et al., 2003). This interaction inhibited (-)-strand RNA synthesis in vitro (Pogany et al., 2003) and is essential for DI RNA accumulation in vivo (Na and White, 2006; Pogany et al., 2003). As discussed above, RSE can also base pair with p33RE, an element involved in replicase complex assembly (Pogany et al., 2005). During replicase complex assembly and/or membrane association, the RSE-3’CSS interaction may be disrupted and p33RE-RSE formed, thus making the 3’ terminus accessible by the RdRp (Pogany et al., 2005). *Mfold* analysis revealed that potential RSE-3’CSS interactions may be a common feature in the *Tombusviridae* (Na and White, 2006).

In addition to cis-acting sequences, trans-acting cellular factors or viral encoded proteins may affect the balance between alternative structural conformations. For example, the 3' terminal five bases of Qβ bacteriophage are involved in long distance base-pairing that does not permit efficient access to the polymerase. Either the host protein Hfq or a series of mutations including alterations to the 3' end and interacting sequence, is needed to destabilize the secondary structure in the region and allow access of the polymerase to the 3' end (Schuppli et al., 1997; Schuppli et al., 2000).

Coordination of genomic and subgenomic RNA synthesis

Many (+)-strand RNA viruses generate sgRNAs to regulate the expression of genes located downstream of the 5' proximal gene. For coronaviruses and arteriviruses, a nested set of sgRNAs is generated by joining a common 5' leader sequence, derived from the 5' end of the genome, to each of the sgRNA coding sequences (mRNA body) through a discontinuous minus-strand RNA synthesis mechanism (Sawichi and Sawichi, 2005). A key step in this process is base pairing between RNA elements known as transcription-regulating sequences (TRSs), which are located at the 3' end of the leader sequence (leader TRS) and the 5' end of each of the mRNA bodies (body TRSs) (Pasternak et al., 2003; van Marle et al., 1999; Zuniga et al., 2004). In *Equine arteritis virus* (EAV), the leader TRS is located in the terminal loop of a 5' proximal hairpin (LTH) (van den Born et al., 2004; van den Born et al., 2005). Mutational analysis showed that LTH and its flanking sequences form a functional unit to facilitate efficient sgRNA synthesis (van den Born et al., 2005). LTH may also be involved in genomic RNA replication and/or translation (van den Born et al., 2004). Since the 5' and 3' flanking sequences can potentially base pair with each other resulting in extension of the LTH stem, it was hypothesized that this region may require a conformational switch to regulate the function of the 5' proximal region of EAV (van den Born et al., 2005). Similar LTH structure may also be present in the 5' proximal region of all arteriviruses and most coronaviruses genomes (van den Born et al., 2004; van den Born et al., 2005).

Evidence for important alternative structures with no known function

A molecular switch in the MHV genome involves sequences within a 68 nt stem-

loop and a downstream pseudoknot in the 3'-untranslated region (Goebel et al., 2004). These two structures are mutually exclusive because the last 8 nt of the stem-loop are also involved in stem 1 of the pseudoknot. This switch was proposed to regulate RNA replication although a definite function has not been assigned (Goebel et al., 2004).

Structural probing data suggested that the region X of HCV, which is required for RNA replication (Friebe and Bartenschlager, 2002; Yi and Lemon, 2003), can fold into a 46 nt stem-loop (SL1) at the 3' terminus, a 20 nt stem-loop (SL3) at the 5' end and the structure of the middle region that remains controversial. To this end, two models have been proposed: first, a two hairpin (SL2ab)/pseudoknot model that claims the middle region folds into two small hairpins that could potentially form a pseudoknot (Dutkiewicz and Ciesiolka, 2005); second, a single hairpin (SL2)/kissing-loop interaction model that describes a kissing-loop interaction between the loops of SL2 and another stem-loop located within the NS5B coding region essential for viral RNA replication (Blight and Rice, 1997; Friebe et al., 2005). Both of these models suggest that this region may contain multiple conformations. It has been shown that region X can specifically interact with the viral NS3 protein (Banerjee and Dasgupta, 2001), some ribosomal proteins (Wood et al., 2001), and two other host proteins of unknown function (Inoue et al., 1998). In addition, the primary and secondary structures of the SL2 and SL3 regions interact with polypyrimidine tract-binding protein (PTB) (Ito and Lai, 1997; Tsuchihara et al., 1997). Viral RdRp NS5B has also been reported to interact with the 3' end of HCV, including the 3′-terminal NS5B coding region (Cheng et al., 1999; Yamashita et al., 1998). Whether these proteins are involved in the conformational changes and what the functions of the conformational changes might be requires further investigation.

Since conformational switches usually involve tertiary structures such as pseudoknots and kissing interactions, we can expect that more conformational switches will be discovered in the future. Studying how conformational switches affect viral RNA function will deepen our insight into viral replication.

Turnip Crinkle Virus as a Model System for Studying Positive-strand RNA Virus Replication

The discovery of replication-associated elements throughout viral genomes suggests that the viral RNA replication is a much more complicated process than previously thought. These findings suggest that analysis of cis-acting elements removed from their natural context may lead to an oversimplification of many viral processes. However, efforts to identify cis-acting elements and conformational switches in intact viruses is complicated by large genome sizes and dual roles of the genome (i.e., template for both translation and replication). Subviral RNAs such as satellite (sat) RNAs and DI RNAs, whose replication depends on the assistance of a specific helper virus to provide viral replication proteins, have been widely used to define cis-acting elements and identify replication-specific conformational switches because they have limited genome sizes while containing all cis-elements necessary to utilize the replication components provided by their helper viruses. In addition, these RNAs usually do not contain ORFs (Simon et al., 2004). However, findings in these small RNA systems may not always recapitulate what is going on in viral genomic RNA systems.

The association of TCV with several subviral RNAs makes it an ideal system with which to identify cis-acting elements and conformational switches specifically involved in RNA replication. TCV is a member of the family *Tombusviridae*, genus *Carmovirus* (van Regenmortel et al., 2000). The 4054 nt genomic RNA (Carrington et al., 1989; Oh et al., 1995) encodes p28 and its readthrough product p88, which are required for virus replication (Hacker et al., 1992; White et al., 1995). p88 contains the conserved polymerase active site motif GDD and *E. coli* expressed p88 alone is capable of directing complementary strand synthesis of exogenously added (+)- and (-)-strands of TCV subviral RNAs in vitro, producing double-stranded products that are not templates for further transcription (Rajendran et al., 2002). Two 3' co-terminal sgRNAs are synthesized from genomic TCV (Carrington et al., 1989; Figure 1.2A): i) a 1.45 kb sgRNA that directs the synthesis of 38 Kda coat protein (CP), which is necessary for systemic movement (Hacker et al., 1992; Heaton et al, 1991; Cohen et al., 2000a) and suppression of posttranscriptional gene silencing (Qu et al., 2003; Thomas et al., 2003); and ii) a 1.7 kb sgRNA encoding p8 and p9, which function in trans to mediate cell-to-cell movement (Hacker et al., 1992; Li et al., 1998). p8 was shown to interact with an *Arabidopsis thaliana* protein (Atp8) in yeast cells and in vitro. Atp8 carries the characteristics of host proteins participating in plant virus cell-to-cell movement (Lin and Heaton, 2001). In addition, p8 also contains two nuclear localization signals (NLSs) suggested it may have an unknown function (Cohen et al., 2000b).

TCV is associated with several dispensable, non-coding DI RNAs and satRNAs. While DI RNAs are derived from the genomic RNA, most satRNAs share little sequence similarity with the helper viral genomes. SatC is an unusual chimeric RNA. The 5' 190 nt

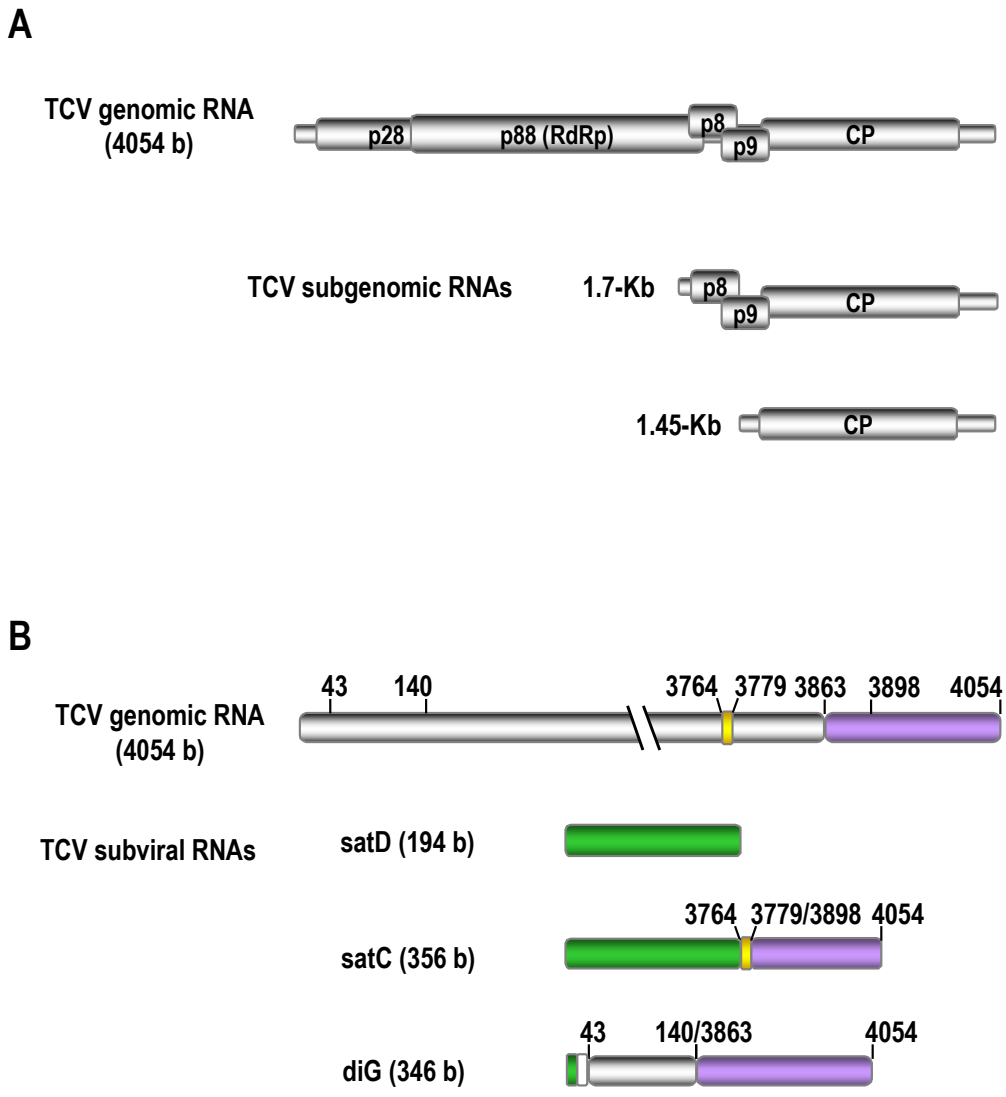


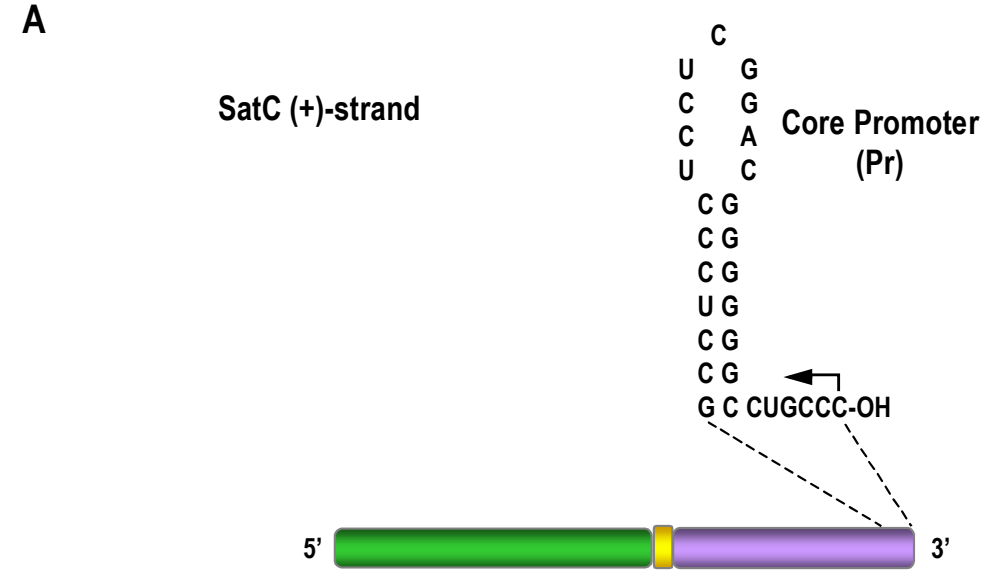
Figure 1.2 Schematic representation of TCV and its subviral RNAs. (A) Genomic organization of TCV. The two sgRNAs are shown below the genomic RNA. (B) Schematic representation of TCV and three associated subviral RNA, satD, satC and diG. Similar regions are color-coded. Numbers refer to positions of sequences in the TCV genome. The sizes of RNAs are given in parenthesis.

of satC share 88% similarity with nearly full-length satD, which is a typical satRNA, while the 3' 166 nt share 94% similarity with two regions at the 3' end of TCV (Simon and Howell, 1986) (Figure 1.2B). In addition to sat RNAs, TCV is also associated with diG. From the 5' end, diG is made up of 10 nt from the 5' end of satD, 12 nt of unknown origin, 99 nt from near the 5' end of TCV, and the last 225 nt from the TCV 3'UTR including one repeated region (shares 94% similarity with TCV) (Li et al., 1989; Kong et al., 1997).

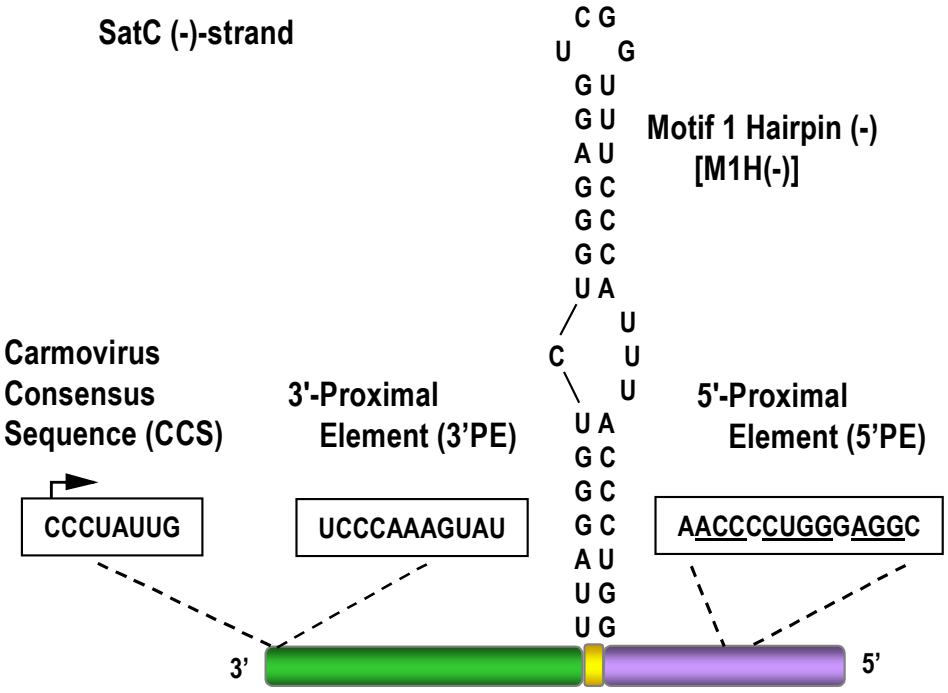
A number of *cis*-acting elements implicated in satC replication have been identified using whole plant, protoplasts, and in vitro systems (Figure 1.3). On the (+)-strand, satC terminates with a single-stranded tail (C<u>CUGCCC</u>-3') that is conserved among TCV RNAs. Deletions up to 6 nt from the 3' end were repaired to wild-type in vivo (Nagy et al., 1997) suggesting this motif is important for satC replication. The core promoter for (-)-strand initiation (Pr) is located within 29 bases at the 3'-terminus composed of a hairpin flanking the 3'-terminal motif (Song and Simon, 1995; Figure 1.3A). Mutagenesis and in vivo SELEX (Systematic Evolution of Ligands by Exponential Enrichment) results showed that both sequence and structure of the Pr are important for satC accumulation in plants (Carpenter and Simon, 1998, Stupina and Simon, 1997). However, recent results from RNA solution structure probing suggest that alternative structures may exist within this region (Zhang et al, 2004, 2006). On the (-)-strand, four cis-acting elements were identified before my project initiated: (i) The 3'-terminal carmovirus consensus sequence (CCS, 3' $C_{1\text{-}3}A/U_{3\text{-}9}$) is required in vivo but dispensable in vitro (Guan et al., 1997; Guan et al., 2000a); (ii) The 3'-proximal element (3'PE), which also contains a CCS, can function as a promoter in vitro but the sequence

Figure 1.3 Cis-acting elements involved in replication of satC that were identified prior to this study. (A) Core promoter (Pr) required for (-)-strand synthesis of satC. Arrow denotes the transcription initiation site. (B) Sequences on satC (-)- strand implicated in (+)-strand synthesis. Conserved nucleotides in 5'PE are underlined. See text for detail.

A


SatC (+)-strand
Core Promoter
(Pr)
C
U G
C G
C A
U C
C G
C G
C G
U G
C G
C G
G C CUGCCC-OH
5'
3'

B


SatC (-)-strand
C G
U G
G U
G U
A U
G C
G C
G C
U A
U
C U
U
U A
G C
G C
G C
A U
U G
U G
Motif 1 Hairpin (-)
[M1H(-)]
Carmovirus
Consensus
Sequence (CCS)
3'-Proximal
Element (3'PE)
5'-Proximal
Element (5'PE)
CCCUAUUG
UCCCAAAGUAU
AACCCCUGGGAGGC
3'
5'

is highly variable according to in vivo SELEX results (Guan et al., 1997; Guan et al., 2000a); (iii) The 5'-proximal element (5'PE), which is required for satC accumulation in vivo and can serve as a promoter in vitro. In vivo SELEX results indicated that the 5'PE is highly sequence specific (Guan et al., 1997; Guan et al., 2000b); and (iv) The motif 1 hairpin (-) [M1H(-)], which is a recombination hotspot and transcriptional enhancer (Cascone et al., 1993; Nagy et al., 1998, 1999, 2001). In addition, recent results demonstrated that the complementary form of M1H(-) on plus-strands (M1H) functions to juxtapose single-stranded sequences at its base, which interferes with TCV virion formation enhancing virus movement (Zhang and Simon, 2003b).

Since Pr is the only cis-acting element involved in replication that had been identified on satC (+)-strands before I initiated my study, it was crucial to identify other regulatory cis-acting elements. As described above, the 3' end region of satC share high sequence similarity with that of TCV genomic RNA, and this region might contain other important cis-acting elements in addition to the Pr. Using computer RNA *mfold* analysis and phylogenetic comparisons of carmoviral 3' end regions, four hairpins were revealed to be structurally and spatially conserved (to varying extent) among carmoviruses (Figure 1.4, Figure 1.5). Ten of 11 carmoviruses contain a very stable 3' terminal hairpin (Pr hairpin), which has been identified as part of the core promoter for (-)-strand synthesis in TCV genomic RNA and satC (Song and Simon, 1995; Sun and Simon, in press). In addition, a second hairpin (H5) was found within 16 to 27 bases upstream of all carmoviruses with a Pr hairpin-like terminal hairpin. H5 contains a central large symmetrical or nearly symmetrical internal loop (LSL) with a highly conserved sequence. Four carmoviruses, including TCV, had identical LSL sequence. Interestingly,

Galingsoga mosaic virus (GaMV), which does not contain a Pr hairpin-like stem-loop, contains a similar hairpin with the highly conserved sequence presented in an asymmetric internal loop. The 3' side of the various LSL can potentially base pair with 3' terminal residues to form a pseudoknot (Ψ_1). H5 is very similar to a hairpin recently found to be a strong suppressor of (-)-strand synthesis in vitro in the *Tombusvirus* genus (Pogany et al., 2003). The SL3 hairpin, described earlier, is located in a similar upstream position relative to a 3' terminal hairpin (Fabian et al., 2003; Figure 1.5, boxed structure). Five potential base-pairs are possible between the 3' terminal AGCCC of tombusviruses (and viruses in related genera) and the asymmetrical loop sequence, which was proposed to sequester the 3' terminus from the RdRp by pairing up the 3' terminal nucleotide (Pogany et al., 2003).

The third hairpin from the 3' end is H4b. Seven carmoviruses have a conserved UGG in the terminal loop of H4b. Potential base pairing between the H4b terminal loop and the sequence flanking the 3' side of H5 (in the case of GaMV, the terminal single-stranded tail) to form a second pseudoknot (Ψ_2) is observed for all carmoviruses regardless of whether they contain the conserved UGG. The most 5' of these hairpins, H4a, is the least conserved. It directly flanks H4b in 6 of 11 carmoviruses. The functions of H4a, H4b, H5 and the presence of two potential pseudoknots form the basis of my thesis.

Figure 1.4 Sequence and structure of the 3' related regions of TCV and satC. Numbering of bases is from the 5' end. Different nucleotides between TCV and satC are denoted in red. Names of the hairpins are boxed. Pr, 3' terminal stem-loop that is a portion of the core promoter required for TCV and satC (-)-strand synthesis. Red triangle indicates a four base deletion in the satC Pr compared to the TCV Pr. H5 and H4b, hairpins of unknown function structurally and positionally conserved in all carmoviruses. Conserved UGG in H4b is in green. H4a, hairpin of unknown function that directly flanks H4b in 6 of 11 carmoviruses. Ψ_1 and Ψ_2 are conserved among carmoviruses.

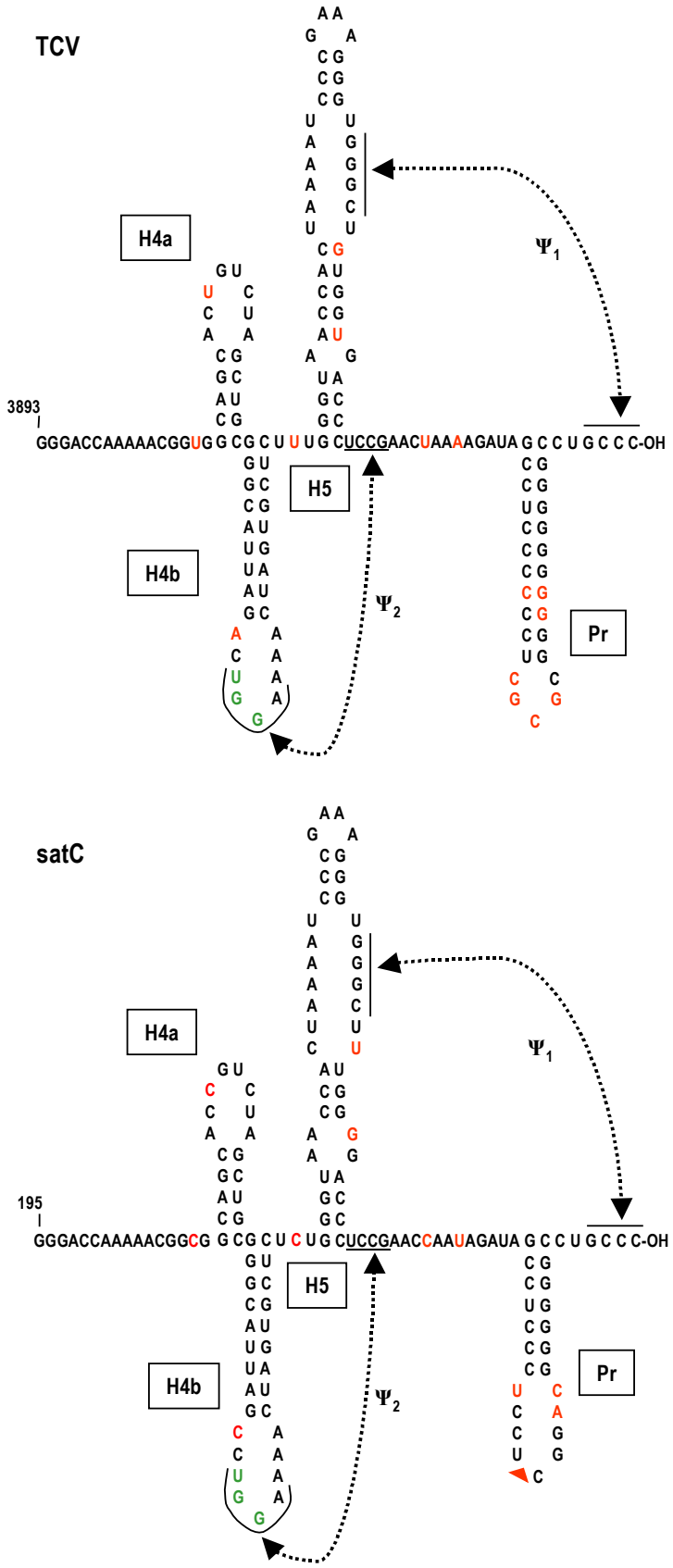
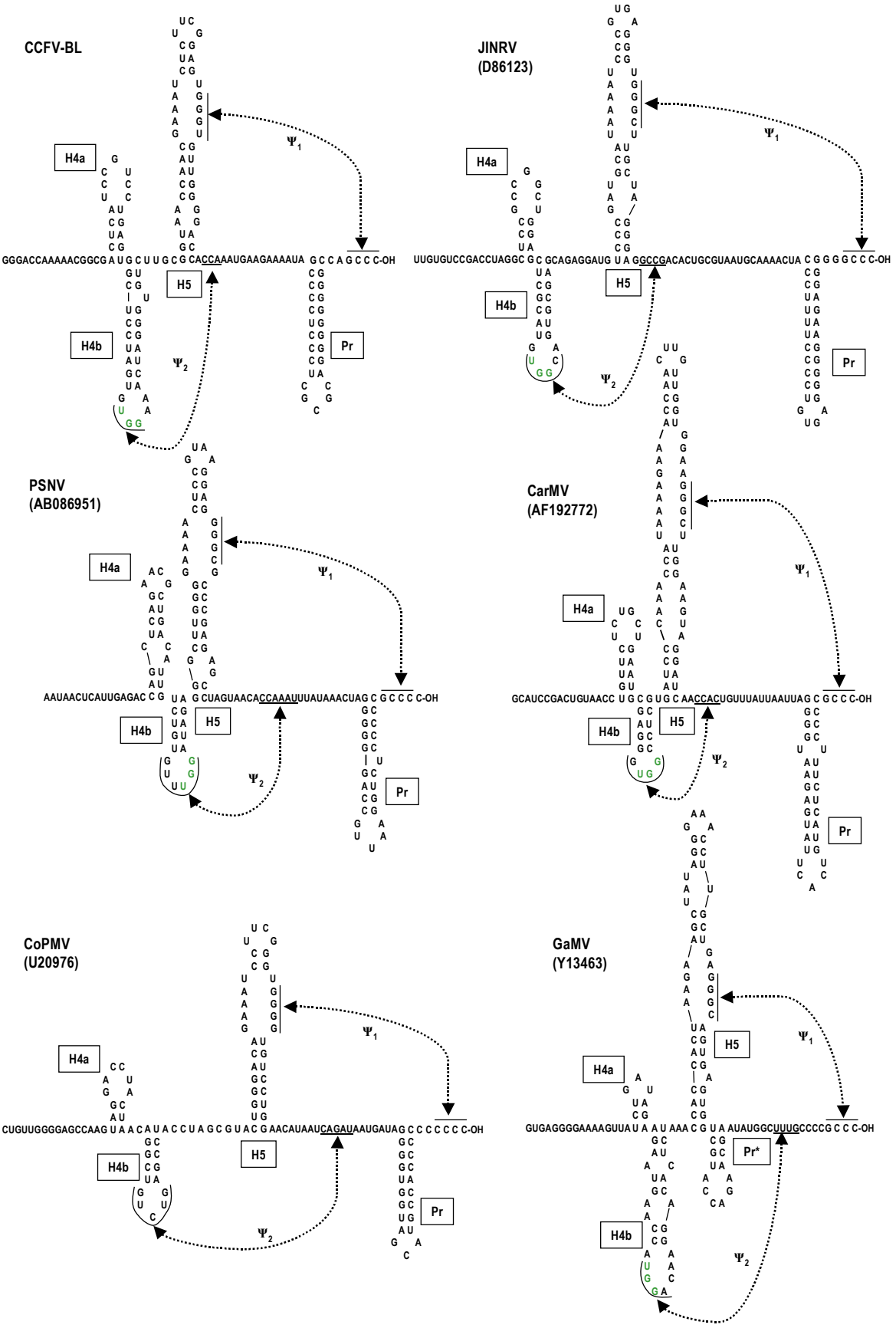
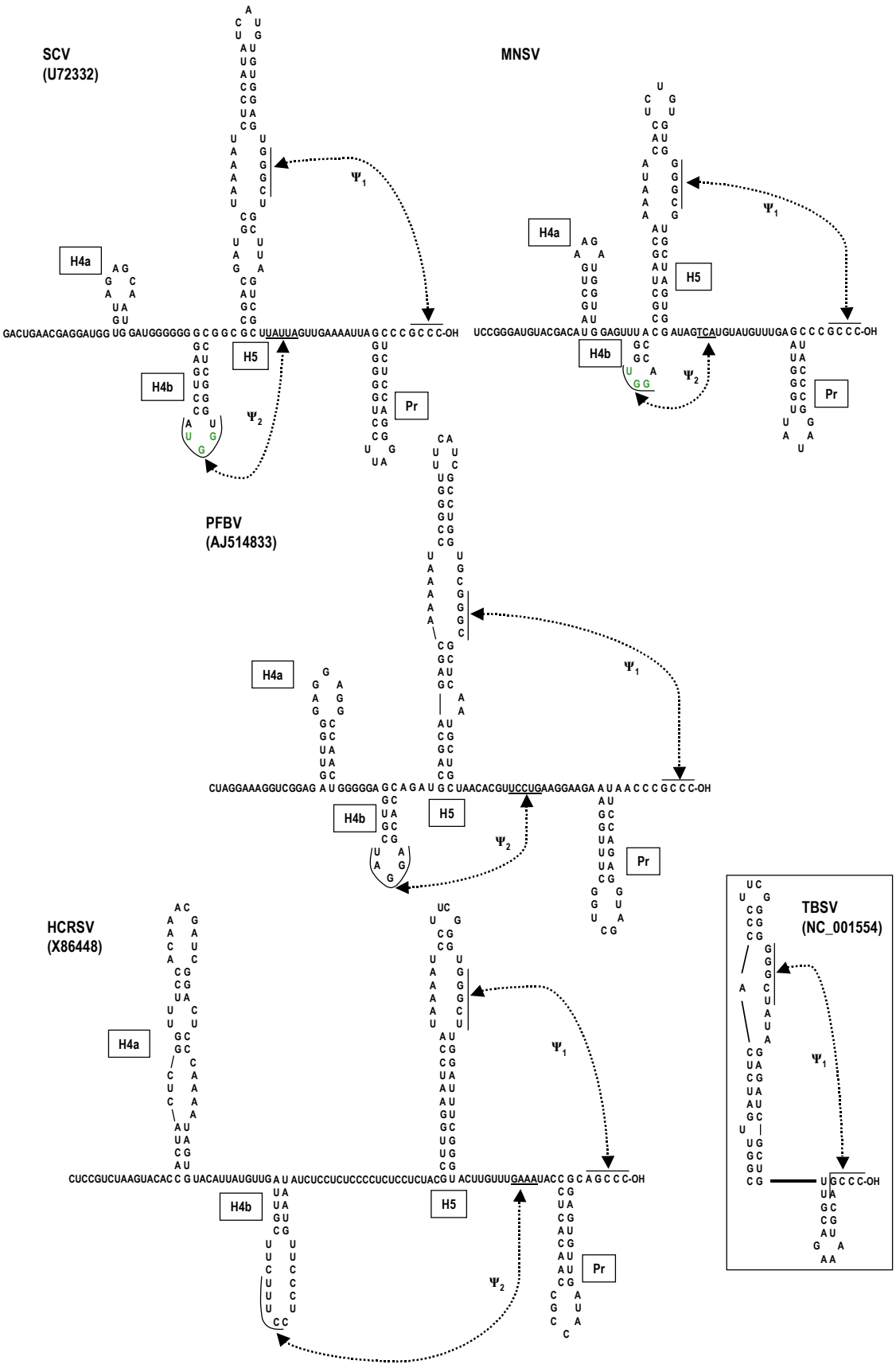

TCV
H4a
H5
H4b
Pr
Ψ1
Ψ2
3893
GGGACCAAAAACGGUGGCGCUUUGCUCCGAACUAAAAGAUAGCCUGCCC-OH
satC
H4a
H5
H4b
Pr
Ψ1
Ψ2
195
GGGACCAAAAACGGCGGCGCUCUGCUCCGAACCAAUAGAUAGCCUGCCC-OH

Figure 1.5 Sequence and structure of the 3' ends of other carmoviruses. The names of the hairpins are given. See legend to Figure 1.4 and text for detail. Pr*, the 3' most hairpin in GaMV. Genebank accession numbers of the sequences are indicated in parentheses. Putative interacting bases are indicated. Asterisks denote covariation in the LSL/3' end interaction in CPMoV. Conserved UGG within H4b loop are denoted in green. CCFV, *Cardamine chlorotic fleck virus* (Skotnicki et al., 1993); JINRV, *Japanese iris necrosis virus* (Takemoto et al., 2000); SCV, *Saguaro cactus virus* (Weng and Xiong, 1997); HCRSV, *Hibiscus chlorotic virus* (Huang et al., 2000); CPMoV, *Cowpea mottle virus* (You et al., 1995); MNSV, *Melon necrotic spot virus* (Riviere and Rochon, 1990); PSNV, *Pea stem necrosis virus* (Suzuki et al, 2002); PFBV, *Pelargonium flower break virus* (Rico and Hernandez, 2004); CarMV, *Carnation mottle virus* (Guilley et al., 1985); GaMV, *Galinsoga mosaic virus* (Ciuffreda et al., 1998). The boxed structure is the replication silencer and 3' terminal hairpin of *Tomato bushy stunt virus*, a member of Genus *Tombusvirus*.


CCFV-BL
H4a
H5
H4b
Pr
Ψ1
Ψ2
JINRV
(D86123)
PSNV
(AB086951)
CarMV
(AF192772)
CoPMV
(U20976)
GaMV
(Y13463)
Pr*


SCV
(U72332)
MNSV
PFBV
(AJ514833)
HCRSV
(X86448)
TBSV
(NC_001554)
H4a
H4b
H5
Pr
Ψ1
Ψ2

Thesis Plan

In this thesis, I describe my studies on the cis-acting elements and a conformational switch involved in satC accumulation in vivo. In Chapter II, I describe the importance of secondary structures predicted by computer programs on (+)-strands of satC for RNA accumulations in protoplasts. I provide data that does not support the existence of a possible third pseudoknot predicted by computer programs. I show that H5 is required for satC accumulation and is position-dependent. In Chapter III, I describe the sequence requirements for the satC H5 LSL, and the results from analyzing possible interactions between the LSL and other regions of satC. In Chapter IV, I further describe the sequence and structural requirements for satC H5 upper stem-loop and lower stem, and show that H5 may be involved in the synthesis of both strands of satC. In Chapter V, I present data indicating a pseudoknot in the pre-active form of satC is part of a structural switch activating (-)-strand synthesis. In Chapter VI, I report that having the cognate Pr is important for replication of TCV and its subviral RNAs.

CHAPTER II

SECONDARY STRUCTURES ON POSITIVE-STRANDS OF SATC ARE IMPORTANT FOR RNA ACCUMULATION IN PROTOPLASTS

Introduction

Positive-strand RNA viruses contain cis-acting elements that play indispensable roles in replication of the viral genome (Buck, 1996; Duggal et al. 1994). Core promoters, which recruit viral encoded RdRp to transcription initiation sites, are often located at the 3' ends of both strands (Buck, 1996; Chapman and Kao, 1999; Deiman et al. 1998; Duggal et al. 1994; Sivakumaran et al. 1999; Song and Simon, 1995). Elements that regulate promoter efficacy, including enhancers and repressors, can be located either proximal or distal to the core promoter (Barton et al., 2001; Herold and Andino, 2001; Khromykh et al., 2001; Klovins et al. 1998; Nagy et al. 1999, 2001; Panavas and Nagy, 2003; Pogany et al., 2003; Ray and White, 1999, 2003; You et al., 2001; Zhang and Simon, 2003b; Zhang et al., 2004). Additional cis-acting elements have also been identified that assist in replicase assembly (Quadt et al., 1995; Vlot et al., 2001). Most of these cis-acting elements share little sequence similarity or structural features between unrelated viruses and, where examined, appear to function through direct interaction with other sequence elements or viral or cellular proteins (Dreher, 1999; Duggal et al., 1994; Fabian et al., 2003; Haldeman-Cahill et al., 1998; Klovins et al., 1998; Melchers et al.,

1997; Pillai-Nair et al., 2003; Singh and Dreher, 1998; Sivakumaran et al., 1999; Williams et al., 1999; Yu et al., 1999; White et al., 1992; Zhang et al., 2004).

As stated in Chapter I, satC is a satRNA associated with TCV. SatC shares high sequence similarity with a second satRNA (satD) at its 5'end and two regions from TCV at its 3' end. The secondary structure predictions for satC generated by *MPGAfold* (the massively parallel genetic algorithm for RNA folding, Kasprzak and Shapiro, 1999; Shapiro *et al*, 2001) and RNA *mfold* (Zucker, 2003) computer programs revealed an identical series of hairpins with only one exception, Hairpin 4a (H4a, Figure 2.1, Figure 2.2). H4a was predicted by *MPGAfold* to be a hairpin with a three base-pair stem capped with a nine-base terminal loop. Five bases (5'CCGUC) in the terminal loop of H4a were proposed to interact with the 5' side flanking sequence (5'GGCGG) to form a pseudoknot (Figure 2.1; Figure 2.5A, left). On the other hand, *mfold* predicted that H4a was a hairpin with a five base-pair stem and a eight-base terminal loop. Four bases (5'CCGU) in the terminal loop of H4a also could potentially base pair with the 5' side flanking sequence (5'ACGG) to form a pseudoknot (Figure 2.2; Figure 2.5A, right). For convenience, these two forms of H4a are named H4a-G (*MPGAfold*) and H4a (*mfold*), respectively.

Two of these hairpins were previously identified. The 3' terminal stem-loop (Pr hairpin) along with the 3' terminal six bases of satC is the core promoter required for satC (-)-strand synthesis (Carpenter and Simon, 1998; Song and Simon, 1995; Stupina and Simon, 1997), although alternative structures within this 29 base region have been suggested by recent results from RNA solution structure probing (Zhang et al., 2004,

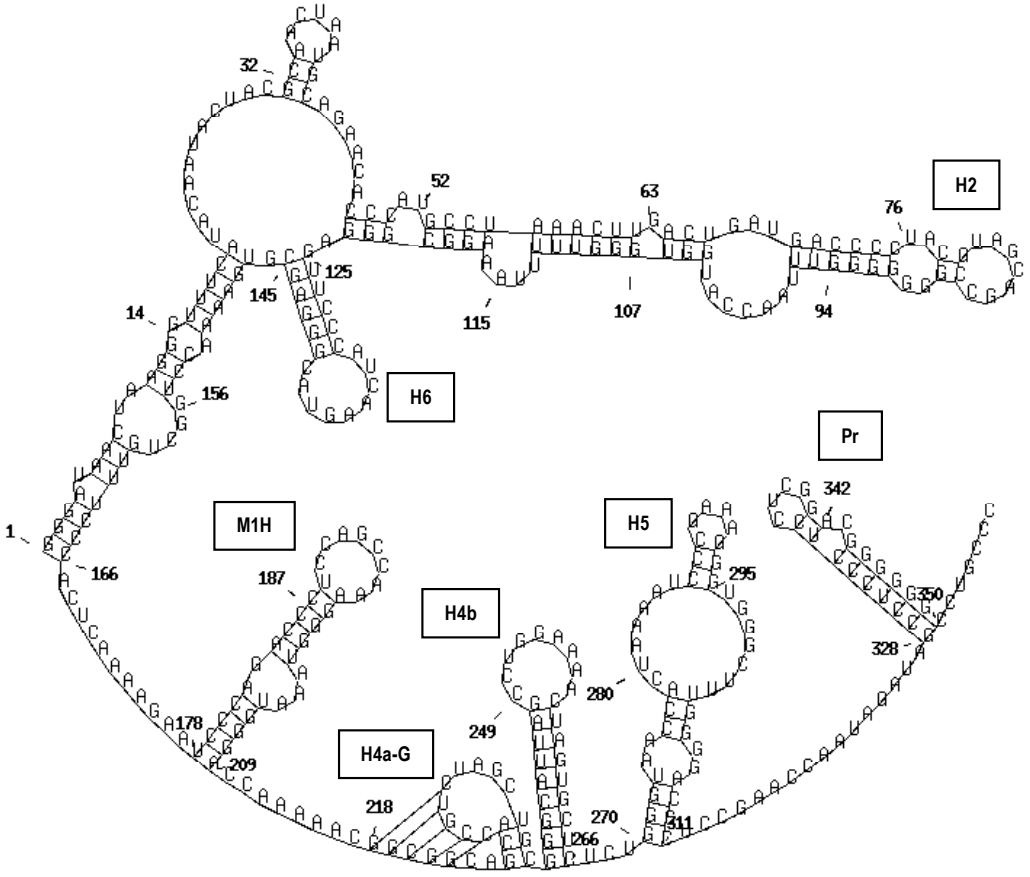


Figure 2.1 The secondary and tertiary structure of satC predicted by *MPGAfold* (the massively parallel genetic algorithm for RNA folding) computer program. STEM TRACE is from 20 runs of *MPGAfold* at a 16K population. Numbering of bases is from the 5' end. The names of the hairpins are shown. Pr, 3' terminal hairpin that is part of the core promoter required for satC (-)-strand synthesis. M1H, hairpin required on (+)-strands to juxtapose single-stranded sequences at its base, which interferes with TCV virion formation enhancing virus movement (Zhang and Simon, 2003). H2, hairpin involved in the normal ratio of monomeric to multimeric forms of satC (Simon et al., 1988; Carpenter et al., 1991a, 1991b). H4a-G, H4b, and H5, hairpins of unknown function structurally and positionally conserved in TCV and several other carmoviruses. H6, hairpin of unknown function. A predicted pseudoknot involving H4a-G and its 5' side flanking sequence is shown.

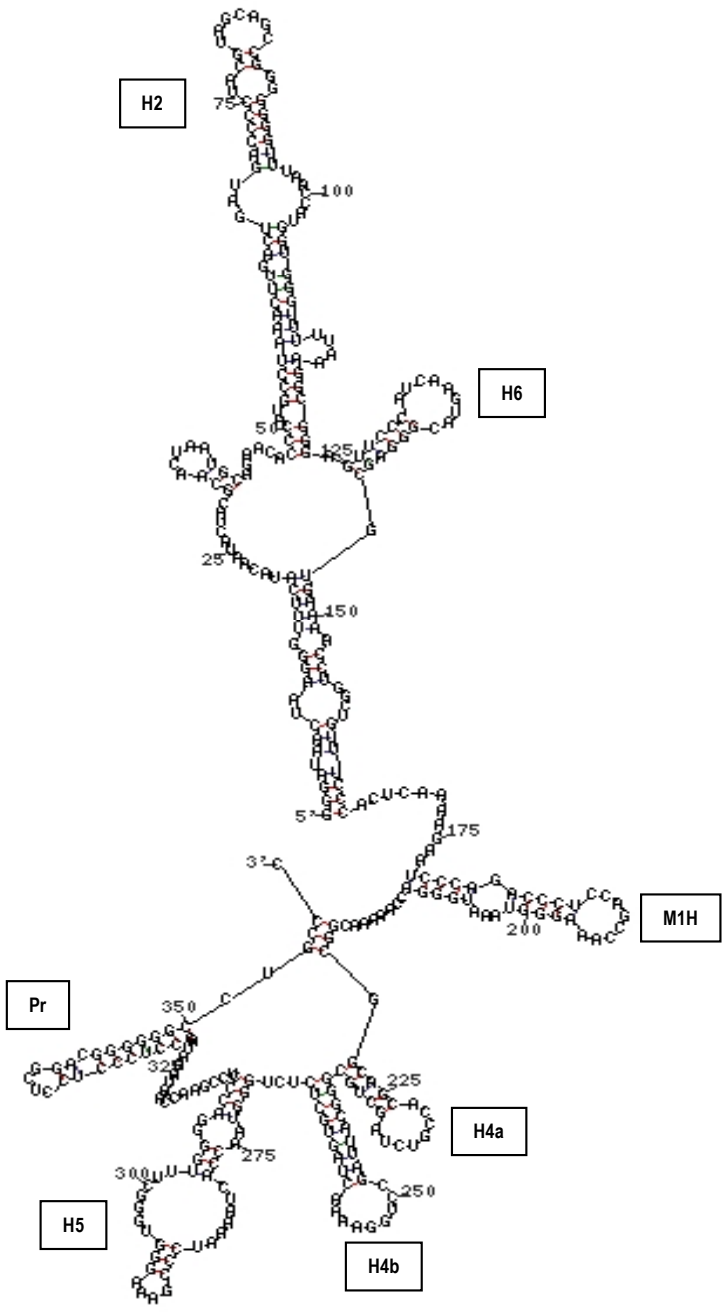


Figure 2.2 The secondary structure of satC predicted by RNA *mfold* computer program. The names of the hairpins are given. All hairpins are identical to those presented in Figure 2.1 except H4a. See Figure 2.1 and text for more information on these elements.

2006). M1H is a hairpin required on (+)-strands to juxtapose single-stranded sequences at its base, which interferes with TCV virion formation enhancing virus movement (Zhang and Simon, 2003b). The complementary form of M1H on the (-)-strand of satC [M1H(-)] is a recombination hotspot and replication enhancer (Cascone et al., 1993; Nagy et al., 1998, 1999, 2001). In addition to these two hairpins, deletions within a 22 base sequence (positions 79 to 100) in H2 resulted in changes in the normal ratio of monomeric to multimeric forms of satC that accumulated in plants, and some specific mutations also led to addition of poly(U) at a downstream position which originally contained five uridylates (Simon et al., 1988; Carpenter et al., 1991a, 1991b). H4a, H4b, H5 are also predicted to be conserved within the *Carmovirus* genus by *mfold* computer program and phylogenetic analysis (Figure 1.4; Figure 1.5).

In this chapter, I report that: (i) individual deletions of these hairpins reduced satC accumulation in Arabidopsis protoplasts; (ii) deletion of M1H or H2 increased satC dimer accumulation while all other deletions decreased satC dimer accumulation; (iii) deletion of either M1H, H2 or H6, which are located upstream of H4a, had much less effect on (-)-strand accumulation compared to deletion of H4a or downstream hairpins; (iv) mutational analysis did not support the existence of a pseudoknot in the H4a region; (v) the function of H5 is position-dependent; and (vi) replacing H4a, H4b and H6 with their reverse complement severely reduced satC accumulation.

Materials and Methods

Small-scale plasmid DNA preparation

E. coli bacteria were grown in a 3 ml L-broth culture overnight at 37°C with continuous shaking. 1.5 ml of cells were transferred to an eppendorf tube and centrifuged at 13000 rpm for 12 sec in a microfuge. The pellet was resuspended in 150 µl of STET buffer [8% (w/v) sucrose, 5% (v/v) Triton X-100, 50 mM EDTA, 50 mM Tris-HCl, pH 8.0] and mixed well. The mixture was boiled for 1 min in a water bath and then centrifuged at 13000 rpm for 10 min at room temperature. The pellet of cell debris was removed by using a toothpick. The plasmid DNA was precipitated with 150 µl of isopropanol followed by centrifugation at 13000 rpm for 5 min at 4°C. The pellet of DNA was rinsed with 75% ethanol, dried, and dissolved in 25 µl distilled H_2O.

Large-scale plasmid DNA preparation

150 ml of L-broth culture was inoculated with 1.5 ml of overnight *E. coli* bacteria culture and incubated overnight at 37°C with continuous shaking. The cells were collected by centrifugation at 6000 rpm for 10 min (Sorvall GSA rotor). The pellet was resuspended in 2.5 ml of suspension buffer [25% (v/v) sucrose and 50 mM Tris-HCl, pH 7.5]. The cell wall was broken by addition of 0.4 ml of 10 mg/ml lysozyme and incubation on ice for 10 min. The mixture was then mixed well with 0.7 ml of 0.5 M EDTA (pH 8.0) and re-incubated on ice for 10 min. Following addition of 5.3 ml of lysis buffer (0.1% Triton X-100, 62.5 mM EDTA, 50 mM Tris-HCl, pH 8.0) and incubation at 42°C for 5 min, the mixture was centrifuged at 17,000 rpm for 20 min (Sorvall SS34

rotor). The supernatant was collected and adjusted to 11 ml with distilled H_2O after addition of 8.7 g CsCl. The solution was mixed with 0.2 ml of 10 mg/ml ethidium bromide and transferred to a OptiSeal Polyallomer centrifuge tube (Beckman). After centrifugation at 65,000 rpm for 4.2 hours at 20°C (Beckman Vti 65.1 rotor), the lower DNA band was collected with a 5 ml syringe and 18 gauge needle and the ethidium bromide extracted 3 times with NaCl-saturated isopropanol. The solution was then mixed with 2 volumes (original volume) of distilled H_2O and 6 volumes of ethanol and incubated at -20°C for at least 2 hours. Following centrifugation at 3000 rpm for 10 min (Beckmam GPR-type swinging bucket), the precipitated plasmid DNA was re-dissolved in 0.4 ml of distilled H_2O and transferred to a 1.5 ml eppendorf tube followed by precipitation with ethanol. After centrifugation at 13,000 rpm for 10 min at 4°C, the pellet was washed with 70% ethanol, dried, and dissolved in 0.5 ml of distilled H_2O. The DNA concentration was estimated by measuring the absorbance at 260 nm.

Construction of satC mutants

Generation of CΔM1H was described elsewhere (also named Δmot1, Nagy et al., 1999). All constructs used in this chapter are listed in Table 2.1. For construction of plasmid CΔH2, pT7C+, a plasmid containing wild-type satC cDNA downstream from a T7 RNA polymerase promoter (Song and Simon, 1994, Figure 2.3), was digested with *Sna*BI and *Nco*I followed by treatment with T4 DNA polymerase. The fragments were self-ligated. To generate plasmid CΔ5P and CΔ5P/M1H, PCR was performed with primers T7C5' (all primer sequences are shown in Table 2.2) and CΔ5P and templates pT7C+ and CΔM1H, respectively. PCR products were cloned into the *Sma*I site of

TABLE 2.1 Summary of satC mutants used in Chapter II

Name	Description
CΔM1H	SatC with deletion of positions 181 to 208 (Nagy et al., 1999)
CΔH2	SatC with deletion of positions 79 to 100
CΔ5P	SatC with deletion of positions 304-315
CΔ5P/M1H	SatC with deletion of positions 181 to 208 and positions 304-315
CΔ5P/H2	SatC with deletion of positions 79 to 100 and positions 304-315
CΔM1H/H2	SatC with deletion of positions 79 to 100 and positions 181 to 208
CΔ5P/M1H/H2	SatC with deletion of positions 79 to 100, positions 181 to 208 and positions 304-315
CΔH4a	SatC with deletion of positions 222-239
CH4aR	SatC with the reverse complement of H4a (positions 222 to 239)
CΔH4b	SatC with deletion of positions 240-266
CH4bR	SatC with the reverse complement of H4b (positions 240 to 266)
CΔH5	SatC with deletion of positions 270 to 311
CΔH6	SatC with deletion of positions 125 to 145
CH6R	SatC with the reverse complement of H6 (positions 125 to 145)
G230C	SatC with a G to C change at position 230
C220G	SatC with a C to G change at position 220
C220G/ G230C	SatC with a C to G change at position 220 and a G to C change at position 230
G218C	SatC with a G to C change at position 218
C229G	SatC with a C to G change at position 229
G218C/ C229G	SatC with a G to C change at position 218 and a C to G change at position 229
U231C	SatC with a U to C change at position 231
G218C/ U231C	SatC with a G to C change at position 218 and a U to C change at position 231
M216-219	SatC with a ACGG to UGCC change at positions 216 to 219
M219-222	SatC with a GCGG to UGCC change at positions 219 to 222
M228-231	SatC with a CCGU to GGCA change at positions 228 to 231
M216-219/ M228-231	SatC with a ACGG to UGCC change at positions 216 to 219, and a CCGU to GGCA change at positions 228 to 231
M219-222/ M228-231	SatC with a GCGG to UGCC change at positions 219 to 222, and a CCGU to GGCA change at positions 228 to 231
$satC_E$	SatC with insertions of one U after position 316 and another U after position 317, thereby generating a *Eco*RV site between positions 315 and 318
2xH5	$SatC_E$ with insertion of a second copy of H5 (positions 258 to 316 and an extra U) in the same orientation at *Nco*I site
H5H5R	$SatC_E$ with insertion of a second copy of H5 (positions 258 to 316 and an extra U) in the reverse orientation at *Nco*I site
H5-Nco	$SatC_E$ with an ectopically positioned H5 (positions 258 to 316 and an extra U) in the same orientation at *Nco*I site
H5R-Nco	$SatC_E$ with an ectopically positioned H5 (positions 258 to 316 and an extra U) in the reverse orientation at *Nco*I site

TABLE 2.1 continued

Name	Description
H5-I	SatC with a U to C change at position 280 and an A to U change at position 281
H5-L	SatC with a G to U change at position 304 (Guan et al., 2000b)
2xH5-I	SatC with insertion of a second copy of H5 (positions 258 to 316 and an extra U) in the same orientation at *Nco*I site, a U to C change at position 280 and an A to U change at position 281 (numbering of bases is from the 5' end of wild-type satC)
2xH5-L	SatC with insertion of a second copy of H5 (positions 258 to 316 and an extra U) in the same orientation at *Nco*I site and a G to U change at position 304 (numbering of bases is from the 5' end of wild-type satC)
H5-H5Rc	SatC with insertion of a mutated copy of H5 (positions 258 to 316 and an extra U with a U to C change at position 280 and an A to U change at position 281, numbering of bases is from the 5' end of wild-type satC) in the reverse orientation at *Nco*I site,
H5-H5Rp	SatC with insertion of a mutated copy of H5 (positions 258 to 316 and an extra U with a G to U change at position 304 (numbering of bases is from the 5' end of wild-type satC) in the reverse orientation at *Nco*I site,

TABLE 2.2 Summary of the oligonucleotides used in Chapter II

Application/ construct	Name	Position[a]	Sequence[b]	Polarity[c]
Mutagenesis in satC	T7C5'	1-19	5'-*GTAATACGACTCACTATA*GGGAUAACUAAGGGTTTCA	+
	Oligo 7	338-356	5'-GGGCAGGCCCCCCGTCCGA	-
	CΔ5P	286-356	5'-GGGCAGGCCCCCCGTCCGAGGAGGGAGGCTATCTA TTGGTTAGCCCACCCTTTCGGG	-
	CΔH5	248-356	5'-GGGCAGGCCCCCCGTCCGAGGAGGGAGGCTATCTA TTGGTTCGGAAGAGAGCACTAGTTTTCCAGGC	-
	ΔH4a	206-278	5'-ccTGGTTACCCAGAGAGCACTAGTTTTCCAGGCTAAT GCCCCGCCGTTTTTGGTCCC	-
	Oligo H4aR	206-263	5'-CACTAGTTTTCCAGGCTAATGCCCGCAGCACCGTCT AGCTGCCGCCGTTTTTGGTCCC	-
	ΔH4b	225-279	5'-ggGTGGTTACCCAGAGCAGCTAGACGGTGC	-
	Oligo H4bR	225-279	5'-ggGTGGTTACCCAGAGGGCATTAGCCTGGAAAACTA GTGCTCGCAGCTAGACGGTGC	-
	ΔH6	97-160	5'-TAACCATGGTGGGTTTTTAAAGGCGGGAGTGAAAAC CTGGCTG	+
	Oligo H6R	96-160	5'-TTAACCATGGTGGGTTTTTAAAGGCGGGAGCTCCCG TACTTGATGGGAACGTGAAAACCTGGCTG 3'	+
	C220G	208-263	5'-gACTAGTTTTCCAGGCTAATGCCCGCAGCTAGACGG TGCTGCCCCCGTTTTTGGTC	-
	G230C	220-263	5'-gACTAGTTTTCCAGGCTAATGCCCGCAGCTAGAGGG TGCTGCCG	+
	C220G/G230C	206-264	5'-CACTAGTTTTCCAGGCTAATGCCCGCAGCTAGAGGG TGCTGCCCCCGTTTTTGGTCCC	-
	G218S/C229S	206-263	5'-gACTAGTTTTCCAGGCTAATGCCCGCAGCTAGACSGT GCTGCCGCSGTTTTTGGTCCC	-
	U231CA	220-263	5'-GACTAGTTTTCCAGGCTAATGCCCGCAGCTAGKCGG TGCTGCCG	+
	M216-219	199-262	5'-tACTAGTTTTCCAGGCTAATGCCCGCAGCTAGACGGT GCTGCCGGGCATTTTGGTCCCATTTACC	-
	M219-222	199-262	5'-tACTAGTTTTCCAGGCTAATGCCCGCAGCTAGACGGT GCTGGGCACGTTTTTGGTCCCATTTACC	-
	M228-231	217-262	5'-tACTAGTTTTCCAGGCTAATGCCCGCAGCTAGTGCCT GCTGCCGCCG	-
	Comp1	199-262	5'-tACTAGTTTTCCAGGCTAATGCCCGCAGCTAGTGCCT GCTGCCGGGCATTTTGGTCCCATTTACC	-
	Comp2	199-262	5'-tACTAGTTTTCCAGGCTAATGCCCGCAGCTAGTGCCT GCTGGGCACGTTTTTGGTCCCATTTACC	-
	CEcV	300-356	5'-GGGCAGGCCCCCCGTCCGAGGAGGGAGGCTATCTA TTG (G**A**T**A**TC) GGAGGGTCCCCAAAG	-
	MC279-285	257-300	5'-ggacACTAGTGCTCTCTGGGTAACCANNNNNNNCCCG AAAGGGTGGGC	+
RNA gel blots	Oligo 13	249-269	5'-AGAGAGCACTAGTTTTCCAGG	-

[a] Coordinates correspond to those of satC. Positions 304 to 315, 270 to 311, 222 to 239, 240 to 266 and 125 to145 are deleted in oligonucleotides CΔ5P, CΔH5, ΔH4a, ΔH4b and ΔH6, respectively. Oligo 13 is also complementary to positions 3950 to 3970 of TCV genomic RNA.

[b] Italic letters indicate T7 promoter sequence. Lowercase letters indicate extra bases added to achieve efficient digestion. Bold residues denote bases inserted to generate an *Eco*RV site (in parentheses). Mutations are denoted by underlined letters.

[c] "+" and "-" polarities refer to homology and complementarity with satC positive strand, respectively.

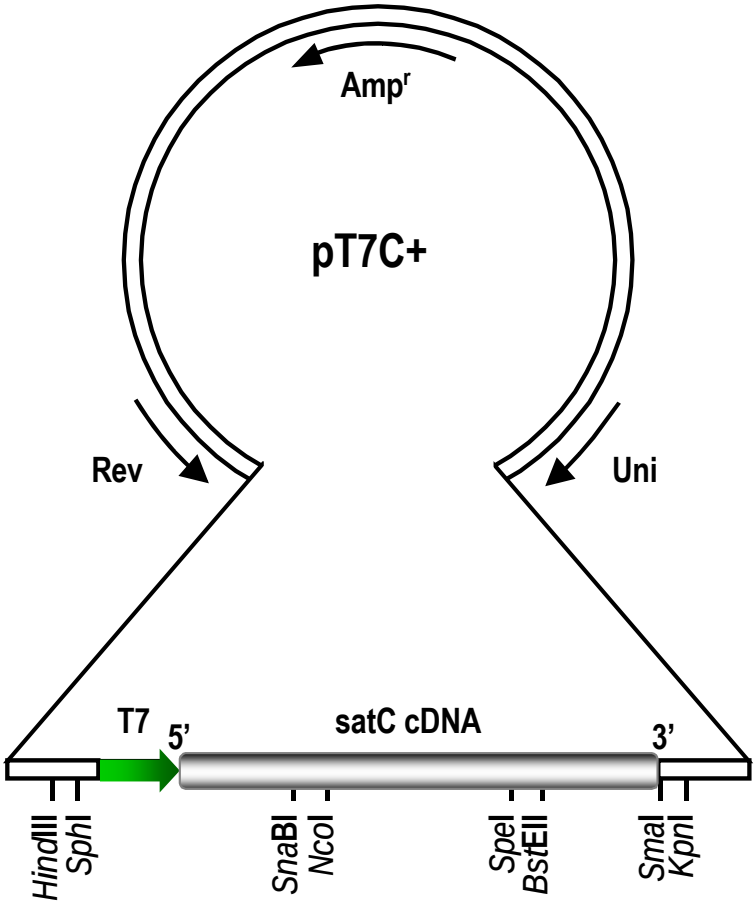


Figure 2.3 Map of pT7C+. The blank and filled bars represent the pUC19 and satC cDNA sequences, respectively. The 5' and 3' ends of satC (+)-strand sequence are shown. The green arrow represents the T7 RNA polymerase promoter sequence. The restriction endonuclease sites used for cloning in this study are shown. Rev and Uni refer to sequence complementary to the reverse and universal primers that are used to sequence constructs (Modified from Song, 1995).

pUC19. Plasmids CΔ5P/H2 and CΔM1H/H2 were constructed by digestion of plasmids CΔ5P and CΔM1H, respectively, with *Sna*BI and *Nco*I. The fragments were treated with T4 DNA polymerase and self-ligated. To construct plasmid CΔ5P/M1H/H2, oligonucleotides T7C5' and CΔ5P were used as primers with template CΔM1H/H2 in a PCR. PCR products were cloned into the *Sma*I site of pUC19.

To generate plasmid CΔH4a, oligonucleotides T7C5' and CΔH4a were used as primers along with pT7C+ as template in a PCR. Following digestion with *Spe*I and *Nco*I, the fragment was inserted into the analogous location in pT7C+, which had been treated with the same restriction enzymes. Plasmid CH4aR was generated in a similar fashion except oligonucleotide ΔH4a was replaced by Oligo H4aR. Plasmids CΔH4b and CH4bR were generated by PCR using T7C5' along with oligonucleotides ΔH4b and Oligo H4bR, respectively, as primers and pT7C+ as template. PCR products were subsequently treated with *Nco*I and *Bst*EII, and cloned into the analogous location in pT7C+, which had been treated with the same restriction enzymes. For construction of plasmid CΔH5, oligonucleotides T7C5' and CΔH5 were used as primers with template pT7C+ in a PCR. PCR products were cloned into the *Sma*I site of pUC19. Plasmid CΔH6 was generated by PCR using oligonucleotides ΔH6 and Oligo 7 as primers and pT7C+ as template. Following digestion with *Spe*I and *Nco*I, the fragment was inserted into the analogous location in pT7C+, which had been treated with the same restriction enzymes. Plasmid CH6R was obtained similarly except that oligonucleotide ΔH6 was replaced with Oligo H6R.

For construction of plasmid G230C, oligonucleotides T7C5' and G230C were used as primers with template pT7C+ in a PCR. Following treatment with *Spe*I and

*Sna*BI, the fragments were inserted into the analogous location in pT7C+, which had been treated with the same restriction enzymes. To generate plasmid C220G, PCR was performed with primers T7C5' and C220G. PCR products were digested with *Spe*I and *Nco*I and ligated into similarly digested pT7C+, replacing the endogenous fragment. Plasmid C220G/G230C was generated similarly except that primer C220G was replaced with C220G/G230C. Plasmids G218C, C229G and G218C/C229G were also generated in a similar fashion except primer G218S/C229S was used instead of C220G. Plasmid U231C was obtained by PCR using primers T7C5' and U231AC and template pT7C+. PCR products were digested with *Spe*I and *Sna*BI and ligated into similarly digested pT7C+, replacing the endogenous fragment. Plasmid G218C/U231C was generated similarly except that template was plasmid G218C and *Nco*I was used instead of *Sna*BI. To construct plasmids M216-219, M219-222, M228-231, M216-219/M228-231 and M219-222/M228-231, oligonucleotide T7C5' was used along with oligonucleotides M216-219, M219-222, M228-231, comp1 and comp2, respectively, as primers in a PCR. The template was pT7C+. Following digestion with *Spe*I and *Nco*I, the fragment was inserted into the analogous location in pT7C+, which had been treated with the same restriction enzymes.

To generate plasmids 2xH5 and H5H5R, an *Eco*RV restriction site was generated downstream of H5 in pT7C+. Oligonucleotides CEcV and T7C5' were used as primers with template pT7C+ in a PCR, and products were cloned into the *Sma*I site of pUC19 generating plasmid satC_E. Plasmid satC_E was digested with *Spe*I and *Eco*RV. Following treatment with T4 DNA polymerase, the fragment was inserted into the *Nco*I site of plasmid satC_E, which had been treated with *Nco*I and T4 DNA polymerase. Plasmids H5-

Nco and H5R-Nco were generated by deletion of the sequence between *Spe*I and *Eco*RV sites in plasmids 2xH5 and H5H5R, respectively. Plasmid H5-I was constructed by PCR using primers M279-285 and Oligo 7 with template pT7C+. Following digestion with *Spe*I and *Sma*I, the fragment was inserted into the analogous location in pT7C+, which had been treated with the same restriction enzymes. Generation of plasmid H5-L was described elsewhere (Guan et al. 2000b). For construction of plasmids 2xH5-I and 2xH5-L, plasmid 2xH5 was digested with *Spe*I and *Sna*BI and the fragment was respectively cloned into the analogous location in plasmids H5-I and H5-L, which had been treated with the same restriction enzymes. Plasmids H5-H5Rc and H5-H5Rp were generated by PCR using primers T7C5' and CEcV. Templates were H5-I and H5-L, respectively. PCR products were digested with *Spe*I and *Eco*RV. Following treatment with T4 DNA polymerase, the small fragments were inserted into the *Nco*I site of plasmid satC_E, which had been treated with *Nco*I and T4 DNA polymerase. All constructs were identified by DNA sequencing.

DNA sequencing

Five microliters out of the 25 μl of small-scale prepared plasmid DNA as described above was mixed with 1 μl of primer (2 pmol/μl), 4 μl of 1N NaOH, and 12 μl of distilled H_2O. The mixture was boiled in a water bath for 2 min and then mixed with 3 μl of 3 M NaOAc and 50 μl of ethanol. After incubation at –80°C for 10 min, the mixture was centrifuged at 13,000 rpm for 10 min at 4°C. The precipitated DNA was washed with 75% ethanol, dried, and subjected to dideoxynucleotide sequencing, which was performed using Sequenase Version 2.0 DNA sequencing kit from USB.

In vitro transcription using T7 RNA polymerase

Plasmid DNA was digested with *Sma*I, extracted with phenol/chloroform, and precipitated with ethanol. The dried DNA was dissolved in an appropriate amount of distilled H_2O and the concentration was estimated by measuring the absorbance at 260 nm. Linearized plasmid DNA (8 μg) was mixed with 6 μl of 100 mM dithiothreitol, 12 μl of 5 mM rNTP mix (contains 5 mM each of ATP, CTP, GTP, UTP), 12 μl of 5 x T7 buffer (125 mM NaCl, 40 mM $MgCl_2$, 10 mM Spermidine, 200 mM Tris-HCl, pH 8.0), 240 U of T7 RNA polymerase and distilled H_2O to bring the final volume up to 60 μl. The reaction was carried out at 37°C for 2 hours or until the solution became turbid. The RNA transcripts were then extracted with phenol/chloroform, precipitated with NH_4OAC and isopropanol, dried, and dissolved in an appropriate amount of distilled H_2O. The RNA concentration was estimated by measuring the absorbance at 260 nm. The transcripts generated from *Sma*I linearized plasmid DNA contained precise wild-type 5' and 3' ends.

Preparation and inoculation of Arabidopsis protoplasts

Protoplasts were prepared from callus cultures of *Arabidopsis thaliana* ecotype Col-0. To generate the calli, *A. thaliana* seeds were washed in 70% ethanol 2 times followed by incubation in a bleach solution (4-6% NaOCl) containing 0.5% (w/v) SDS for 8 to 10 min at room temperature. The seeds were then washed with sterile H_2O five times before sprinkling on Callus maintenance medium (pH 5.8) containing 1.0% (w/v) Bacto agar (DIFCO products), 3% (w/v) sucrose, 4.4 mg/ml Murashige-Skoog (MS) salts

(Gibco BRL or Sigma), 0.5 μg/ml 2,4 D (2,4-Dichlorophenoxyacetic acid, Sigma), 0.5 μg/ml Kinetin (Sigma), and 1 x Vitamins/Glycine (1 μg/ml nicotinic acid, 10 μg/ml thiamine HCl, 1 μg/ml pyrodoxine HCl, 100 μg/ml myo-inositol, 4 μg/ml glycine, Sigma) in 100 x 15 mm petri dishes. The plates were incubated in a growth chamber (Percival Scientific) at 20°C under 35 μmol/m^2S lights with a photoperiod of 16-hour light and 8-hour dark. Cultures were transferred onto fresh media every 3 weeks.

To prepare protoplasts, calli from 1 to 3 plates at passage 4 to 6 were minced into small pieces and soaked in 30 ml of 0.6 M mannitol at room temperature for 20 min with gentle shaking. Following centrifugation at 2000 rpm for 5 min (Beckman GPR-type swing bucket), the recovered calli were resuspended in 50 to 100 ml of PIM (Protoplast isolation medium, pH 5.8, adjusted by 1N KOH) containing 10 mg/ml (11,900 U/g) cellulose and 2 mg/ml (3,140 U/g) macerase/pectinase (Calbiochem, La Jolla, CA). One liter PIM contained 1 ml of 1000 x vitamin stock (0.5 mg/ml nicotinic acid, 0.5 mg/ml pyrodoxine HCl, 1 mg/ml thiamine-HCl, 100 mg/ml myo-inositol), 0.5 ml of 2000 x hormone stock (0.4 mg/ml 2,4 D, 0.4 mg/ml kinetin, 50 mM KOH), 4.4 g of MS salts, 34.2 g sucrose (0.1 M), 0.585 g of MES (3 mM), 91 g of mannitol (0.5 M) and 0.555 g $CaCl_2$ (5 mM). The mixture was incubated at 26°C for 3.5 to 4 hours with gentle shaking (slowly increased the speed until all the calli started floating). The solution should become turbid and protoplasts were recovered by filtration through sterile nylon mesh (53 μm, Small Parts) followed by centrifugation at 900 to 1000 rpm for 5 min at 4°C (turn brake off on the centrifuge). The protoplasts were then washed 3 times with 40 ml of 0.6 M mannitol (pre-cooled at 4°C) and number of protoplasts was calculated using a haemocytometer. 5 x 10^6 protoplasts were aliquoted into each 50 ml centrifuge tube and

centrifuged at 900 to 1000 rpm for 5 min at 4°C. The supernatant was decanted as much as possible (the leftover liquid was about 100 μl per tube).

TCV genomic RNA and satC transcripts were synthesized using T7 RNA polymerase from plasmids linearized with *Sma*I. Protoplasts (5×10^6) prepared as described above were mixed with 20 μg of TCV genomic RNA transcripts with or without 2 μg of each satC RNA transcripts, 8 μl of 1.0 M $CaCl_2$ and 400 μl of sterile distilled H_2O. Following addition of 2.17 ml of 50% PEG 1450 (prepared in 50 mM Tris-HCl, pH 7.4), the tube was shaken gently to mix the solution well and incubated at room temperature for 15-20 sec. Pre-cooled (4°C) 0.6 M mannitol containing 1 mM $CaCl_2$ was added to bring the volume to 30 ml and the mixture was incubated on ice for 15 min. Protoplasts were recovered by centrifugation at 900 to 1000 rpm for 5 min at 4°C and washed 3 times with 20 ml of cold 0.6 M mannitol/1 mM $CaCl_2$. Protoplasts were resuspended in 3 ml of PCM (Protoplast culture medium, pH 5.8, adjusted by 1 N KOH; PCM differed from PIM in containing no $CaCl_2$ and the concentration of mannitol was 0.4 M) and transferred into a 60x15 mm petri dish. Plates were covered with aluminum foil and incubated at room temperature for desired times (usually 40 hours).

Extraction of total RNA from Arabidopsis protoplasts

Protoplasts were collected at the desired time (hours post-inoculation [hpi]) by aliquoting into two 1.5 ml eppendorf tubes and centrifuging at 2000 rpm for 4 min at 4°C. After decanting the supernatant, 150 μl each of RNA extraction buffer (5 mM EDTA, 100 mM NaCl, 50 mM Tris-HCl, pH 7.5) and phenol/chloroform were added into one tube (the other one was frozen at -80°C for backup). Following vigorous vortexing

for 20 sec, the solution was centrifuged at 13,000 rpm for 3 min. The aqueous layer was recovered and total RNA was precipitated with ethanol and dissolved in 30 µl of distilled H_2O. Total RNA concentration was estimated by measuring the absorbance at 260 nm.

RNA gel blots

Equal amounts (2 to 3 µl) of total RNA extracted from protoplasts at 40 hpi were denatured by heating in 50 to 75% formamide at 75°C for 3 min and then subjected to electrophoresis through a 1.5% non-denaturing agarose gel. Following soaking of the gel in 6% formaldehyde for 1 hour with shaking, the gel was soaked in 10 x SSC (1.5 M NaCl, 0.15 M sodium citrate) for 10 min and then another 15 min with the water-prewetted Nitroplus membrane (Micron Separations Inc.) in the same container. RNA was transferred to Nitroplus membrane (traditional paper tower soaked method) and crosslinked by placing the blot on a UV light box (310 nm) for 2 min followed by drying the membrane at 80°C for 10 min.

The (+)-strand RNA was probed with a [γ-^{32}P]-ATP-labeled oligonucleotide (Oligo 13) complementary to both positions 3950 to 3970 of TCV genomic RNA and positions 249 to 269 of satC. The membrane was prehybridized for 1 hr and then hybridized with the probe for at least 2 hr at 42°C in a solution containing 5 x SSPE [1 liter of 20 x SSPE containing 175.3 g NaCl, 27.6 g NaH_2PO_4, 40 ml of 0.5 M EDTA (pH 8.0) and is adjusted to pH 7.4 with 10 N NaOH], 10 x Denhard's reagent [100 x Denhard's reagent containing 2% (w/v) Ficoll, 2% (w/v) PVP and 2% (w/v) BSA], 0.2 mg/ml denatured salmon sperm DNA, 0.2% SDS, and 30% formamide. After hybridization, the membrane was washed once in a high salt solution containing 6 x

SSPE and 0.1% SDS, then washed twice with a low salt solution containing 0.1 x SSPE and 0.1% SDS. Each time the membrane was incubated at 41°C for 10 min. The membrane was then briefly blotted between two sheets of blotting paper, wrapped with saran wrap and subjected to autoradiography.

For analysis of (-)-strand accumulation, the RNA was probed with an [α-^{32}P]-UTP-labeled riboprobe obtained from full-length or *Nco*I-digested pT7C+ by transcription with T7 RNA polymerase. The membrane was prehybridized for 1 to 3 hr and then hybridized with the probe overnight at 42°C in a solution containing 5 x SSPE, 10 x Denhard's reagent, 0.2 mg/ml denatured salmon sperm DNA, 0.2% SDS, and 50% formamide. After hybridization, the membrane was washed twice in a high salt solution containing 2 x SSC (0.3 M NaCl, 0.03 M sodium citrate) and 0.1% SDS, then washed twice with a low salt solution containing 0.1 x SSC (15 mM NaCl, 1.5 mM sodium citrate) and 0.1% SDS. Each time the membrane was incubated at 42°C for 15 min. The membrane was then briefly blotted between two sheets of blotting paper, wrapped with saran wrap and subjected to autoradiography.

Results

Deletions of hairpins predicted by computer structure programs affected satC accumulation in Arabidopsis protoplasts

To determine whether some of the hairpins predicted by *MPGAfold* and RNA *mfold* computer programs are functionally relevant, satC mutants were constructed with individual deletions of H2, H4a, H4b, H5 and H6. The construct containing deletion of

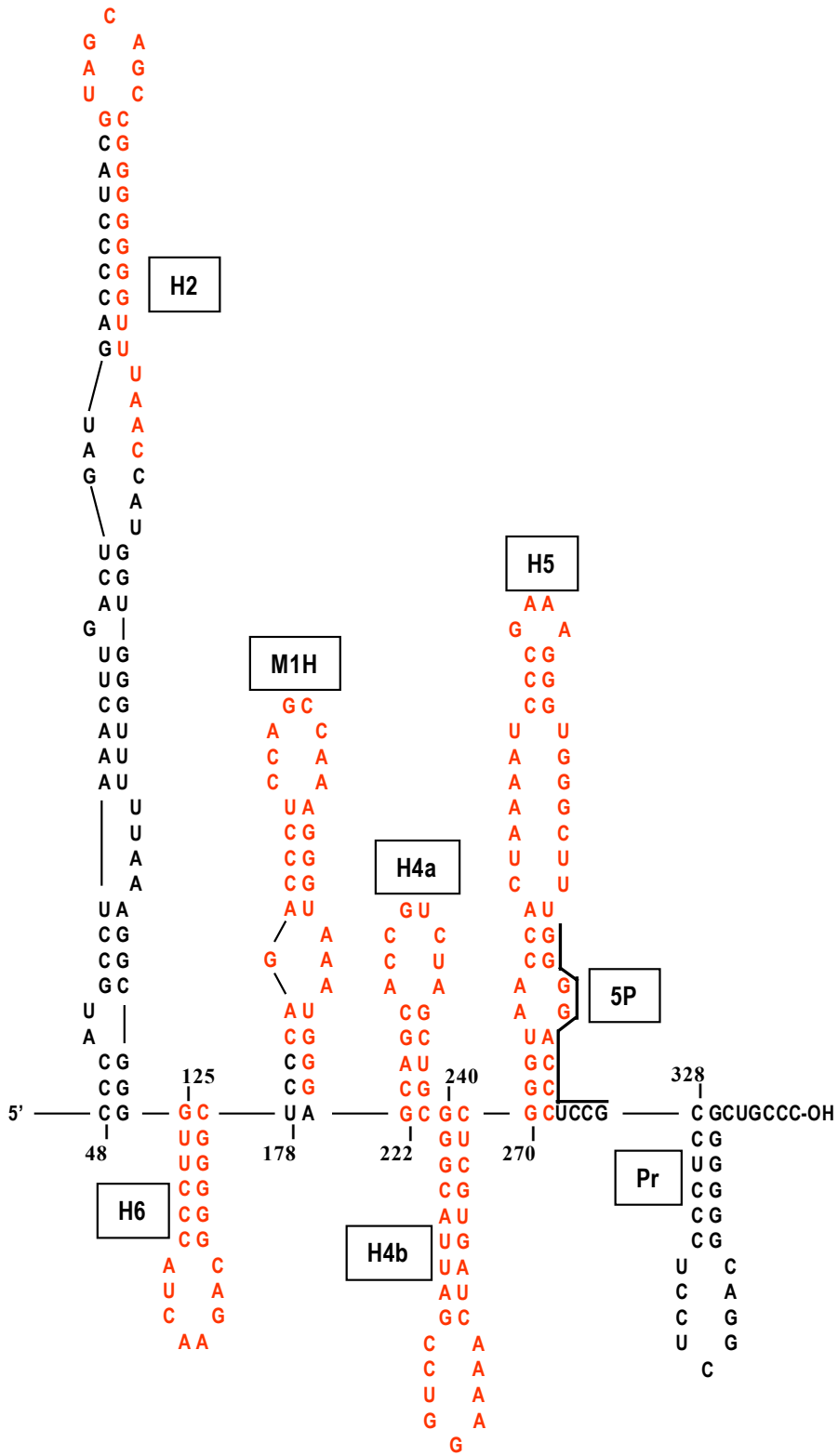
M1H (CΔM1H) was described previously (Nagy et al., 1999). Transcripts of satC variants synthesized in vitro using T7 RNA polymerase were inoculated onto Arabidopsis protoplasts along with TCV genomic RNA. Total RNA was extracted at 40 hpi and subjected to RNA gel blot. At 40 hpi, (+)-strand monomers of CΔH4a (denoted as satC, deletion, name of hairpin), CΔM1H and CΔH2 accumulated to 6-7% of wild-type levels (Figure 2.4). Deletion of H6 (CΔH6) reduced satC levels to 3% of wild-type. The most detrimental consequences were from deletion of H4b and H5. CΔH4b accumulated to less than 1% of wild-type levels, while CΔH5 was undetectable even when the blot was over-exposed. Deletion of the 3' side of the lower stem of H5 only (construct CΔ5P, Figure 2.4A) or combining this mutation with deletion of M1H and/or H2 (constructs CΔ5P/M1H, CΔ5P/H2 and CΔ5P/M1H/H2) also reduced satC accumulation to undetectable levels. Double deletion of M1H and H2 (construct CΔM1H/H2) gave a similar result. These results indicate that deletion of any of the predicted hairpins strongly affects the accumulation of satC (+)-strand monomers in protoplasts.

Deletion of M1H or H2 increased satC dimer accumulation while all other deletions decreased satC dimer accumulation

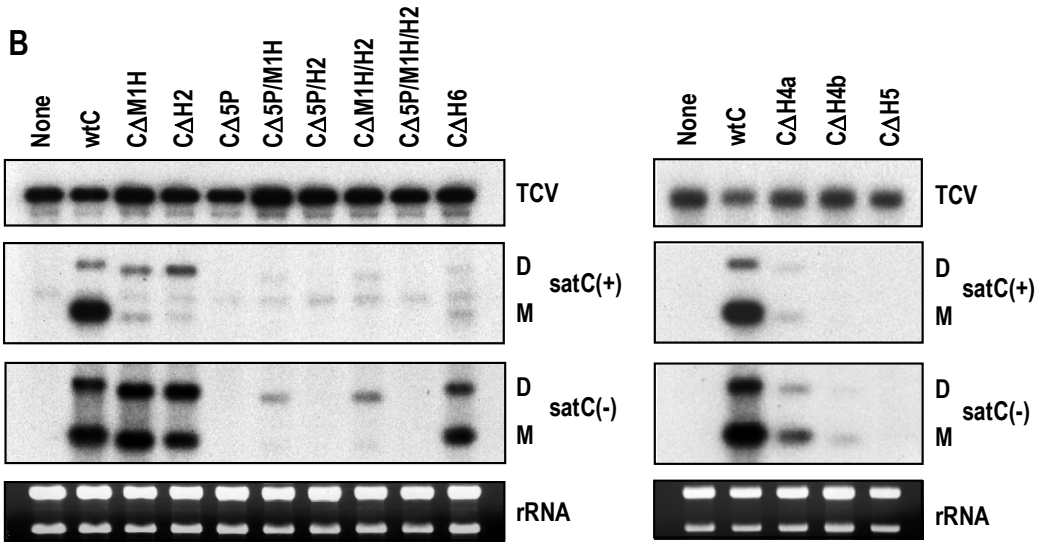
In addition to monomers, multimeric forms of satC have been identified in infected plants and protoplasts. Previous studies revealed that satC multimers are head-to-tail repeats of unit-length monomers with deletion of 5' and 3' end nucleotides and/or addition of non-template nucleotides at the junctions (Carpenter et al., 1991a; Simon and Howell, 1986).

Figure 2.4 Deletion analysis of satC secondary structures. (A) Secondary structures of satC used in this study. Deleted sequences in each corresponding constructs are denoted in red. Sequence deleted in construct CΔ5P is overlined. Solid lines indicate linker regions between hairpins. (B) Representative RNA gel blot of viral RNAs accumulating in protoplasts at 40 hpi. TCV, TCV genomic RNA (+)-strand. satC(+), satC (+)-strand. satC(-), satC (-)-strand. D, dimer. M, monomer. None, no satC in the inoculum. wtC, wild-type satC. The blots were probed with (+)- and (-)-strand specific probes. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot). (C) Summary of accumulation levels of satC variants in protoplasts at 40 hpi. Values given in the table are the averages of two independent experiments, with the wild-type satC levels of monomers arbitrarily assigned a value of 100. ND, not determined. See text for details.

A


H2
M1H
H4a
H5
5P
Pr
H6
H4b
5'
48
125
178
222
240
270
328
GCUCCG
C GCUGCCC-OH

B



C

	(+)-strand			(-)-strand		
	M	D	M:D	M	D	M:D
wtC	100	10	10:1	100	52	1.9:1
CΔM1H	7	15	1:2	84	64	1.3:1
CΔH2	6	23	1:3.8	55	65	1:1.1
CΔ5P	<1	<1	ND	1	<1	ND
CΔ5P/M1H	<1	1	ND	1	7	1:7
CΔ5P/H2	<1	<1	ND	<1	<1	ND
CΔM1H/H2	<1	4	ND	2	14	1:7
CΔ5P/M1H/H2	<1	<1	ND	<1	<1	ND
CΔH6	5	1	5:1	70	30	2.3:1
CΔH4a	6	1	6:1	43	12	3.6:1
CΔH4b	<1	<1	ND	11	3	3.7:1
CΔH5	<1	<1	ND	1	<1	ND

While all deletions severely reduced the accumulation of satC (+)-strand monomers, they had different effects on the accumulation of satC (+)-strand dimers. Deletion of M1H and H2 increased the accumulation of satC (+)-strand dimers by 1.5- and 2.3-fold, respectively, thereby changing the ratio of monomeric to multimeric forms of satC from 10:1 (wild-type) to 1:2 (CΔM1H) and 1:3.8 (CΔH2), respectively. These results were consistent with previous findings (Simon et al., 1988; Carpenter et al., 1991a). Double deletion of M1H and H2 retained 38% of dimers. The accumulation of CΔ5P/M1H, CΔH4a and CΔH6 dimers were 8- to 12-fold less than wild-type levels. The ratios of monomer to dimer forms of CΔH4a and CΔH6 were 6:1 and 5:1, respectively. All other mutants accumulated dimers to less than 2% of wild-type levels. It is noteworthy that one satC dimer can bind two probe molecules while one satC monomer can bind only one probe molecule. For convenience, the ratio of monomeric to multimeric forms of satC was calculated based on the density of the radiograph, which reflects the different accumulation levels but not the molecular ratio of monomeric to multimeric forms of satC variants. In summary, these results suggested that different mechanisms might be involved in the synthesis of satC monomers and multimers.

Deletions of hairpins located upstream of H4a had much less effect on (-)-strand accumulation compared to deletion of H4a or downstream hairpins

Positive-strand RNA virus replication involves two phases: first, synthesis of (-)-strand intermediates; second, synthesis of large amount of (+)-strand progeny from (-)-strand intermediates. To further understand at which phase each hairpin functions, accumulation levels of (-)-strand satC were also examined (Figure 2.4). CΔM1H, CΔH2

and CΔH6 accumulated to 84, 55 and 70% of wild-type (-)-strand monomers, respectively. CΔH4a reduced the accumulation levels of (-)-strand monomers to 43% of wild-type levels. CΔH4b and CΔH5 accumulated 11% and 1% of wild-type levels. All other mutants accumulated to less than 2% of wild-type levels. Individual deletion of hairpins also slightly affected the ratio of (-)-strand monomeric to multimeric forms of satC within a two-fold range (the ratio is 1.9:1 for wild-type). For constructs with deletions of more than one hairpin, CΔ5P/H2 and CΔ5P/M1H/H2, monomers and dimers accumulated to barely detectable levels, CΔ5P/M1H and CΔM1H/H2 had the lowest ratio of (-)-strand monomeric to multimeric forms among all the mutants examined (1:7). These results suggested that deletion of hairpins located upstream of H4a (M1H, H2 and H6) had much less effect on (-)-strand accumulation compared to deletion of H4a or downstream hairpins (H4b and H5). In addition, these results imply that M1H, H2 and H6 may be primarily involved in (+)-strand synthesis.

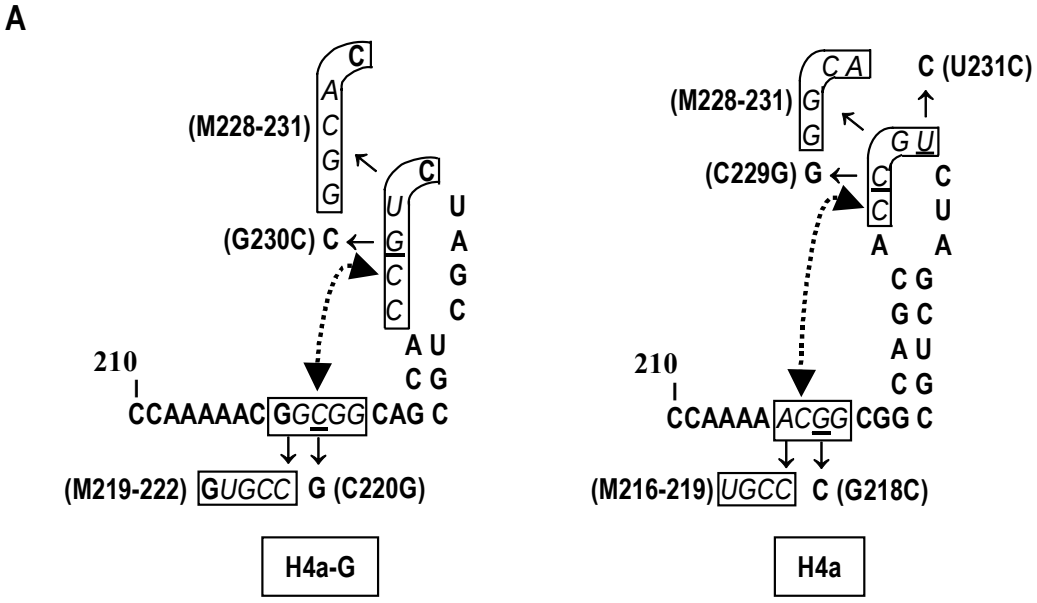
Mutational analysis did not support the existence of a pseudoknot in the H4a region

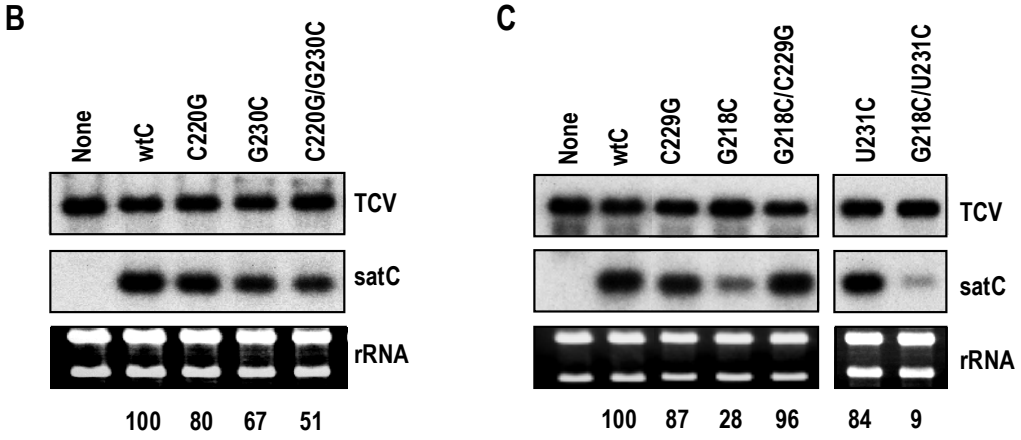
As mentioned above, the structure in the H4a region was predicted differently by *MPGAfold* and RNA *mfold* computer programs. For both predictions, the terminal loop of H4a (or H4a-G) can potentially pair with 5' side flanking sequences to form a pseudoknot (Figure 2.5A). To examine whether a pseudoknot exists and in which form, single and compensatory mutations were introduced into H4a and its flanking sequence. The accumulation levels of mutants were compared with that of wild-type at 40 hpi.

Position C220 and G230 are located centrally within the potential five base pair interacting region in the H4a-G pseudoknot (Figure 2.5A, left). C220G (mutants are

named by their wild-type base, position, mutant base) and G230C accumulated to 80 and 67% of wild-type levels, respectively. C220G/G230C, in which the proposed base-pairing was restored, had reduced accumulation to 51% (Figure 2.5B). Single mutations at position 229 and 231 (constructs C229G and U231C) located within the H4a terminal loop retained 87 and 84% of wild-type levels, respectively, while G218C reduced satC accumulation to 28% of wild-type levels indicating that this residue is important for normal satC function. A non-compensatory mutation for G218C (generating G218C/U231C) further decreased satC accumulation to 9% of wild-type levels. On the contrary, the compensatory mutation (construct G218C/C229G) enhanced satC accumulation to 96% of wild-type levels (Figure 2.5C). These results suggested that the pseudoknot might form through a four base pair interaction. Consequently, the structure prediction generated by *mfold* program might reflect the actual structure of H4a.

To confirm this conclusion, four base mutations were introduced into H4a and its flanking sequence. M228-231, in which positions 228 to 231 were converted from 5'CCGU to 5'GGCA, accumulated to 46% of wild-type (Figure 4D). These four bases, located within H4a, are potentially able to base pair with positions 219 to 222 (5'GCGG) in the H4a-G pseudoknot or positions 216 to 219 (5'ACGG) in the H4a pseudoknot. Both M219-222 (5'UGCC) and M216-219 (5'UGCC) accumulated to very low levels in protoplasts, suggesting that the sequence between H4a and M1H is important for satC accumulation. Neither of the two mutants containing compensatory mutations (M216-219/M228-231 and M219-222/ M228-231) had accumulation restored in protoplasts. Therefore, these latter data do not support the existence of a pseudoknot in the H4a region (see discussion).


A
(M228-231)
(G230C) C
210
CCAAAAAC GGCGG CAG C
(M219-222) GUGCC G (C220G)
H4a-G
C (U231C)
(C229G) G
CCAAAA ACGG CGG C
(M216-219) UGCC C (G218C)
H4a


B
None
wtC
C220G
G230C
C220G/G230C
TCV
satC
rRNA
100 80 67 51
C
None
wtC
C229G
G218C
G218C/C229G
U231C
G218C/U231C
TCV
satC
rRNA
100 87 28 96 84 9

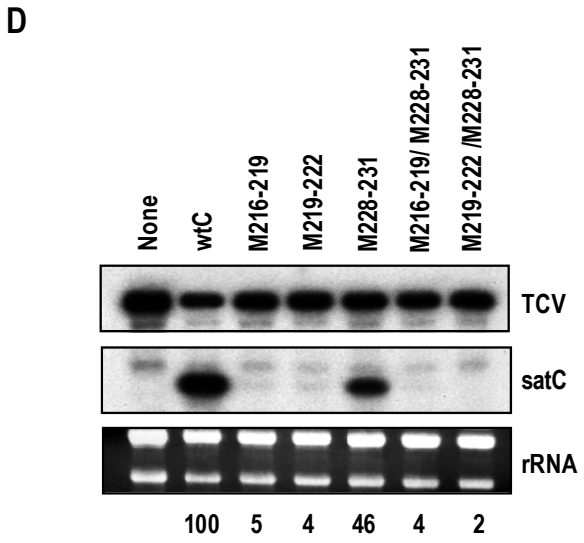


Figure 2.5 Analysis of a predicted pseudoknot between H4a and its 5' side flanking sequence. (A) Structures of H4a and its 5' side flanking sequence predicted by *MPGAfold* and *mfold* computer programs. Potential base-pairing between boxed residues is denoted by connected arrowheads. Mutations generated in putative base-paired partners are shown. Italic letters indicate 4-base mutations. (B) (C) (D) RNA gel blot of viral RNAs accumulating in protoplasts at 40 hpi. TCV, TCV genomic RNA (+)-strand. satC, satC (+)-strand monomer. None, no satC in the inoculum. wtC, wild-type satC. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot). Values given below the blots are the averages of at least two independent experiments, with the satC level arbitrarily assigned a value of 100.

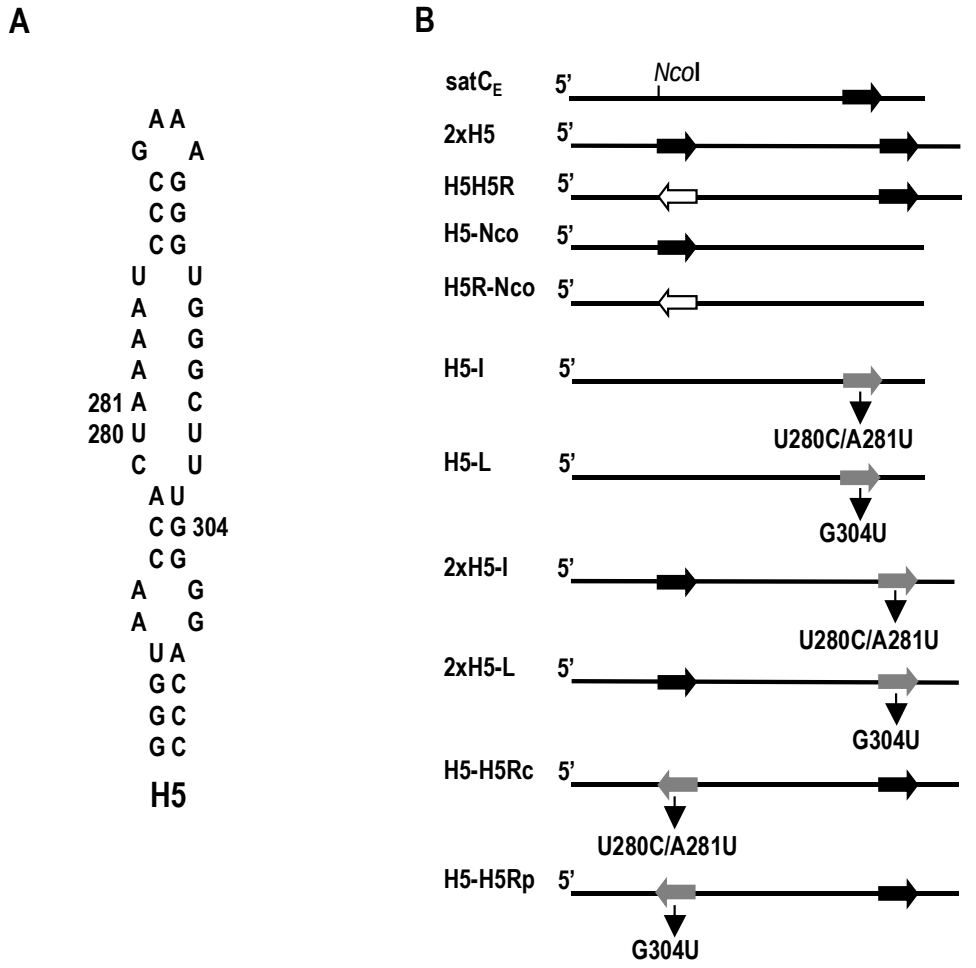
The function of H5 is position-dependent

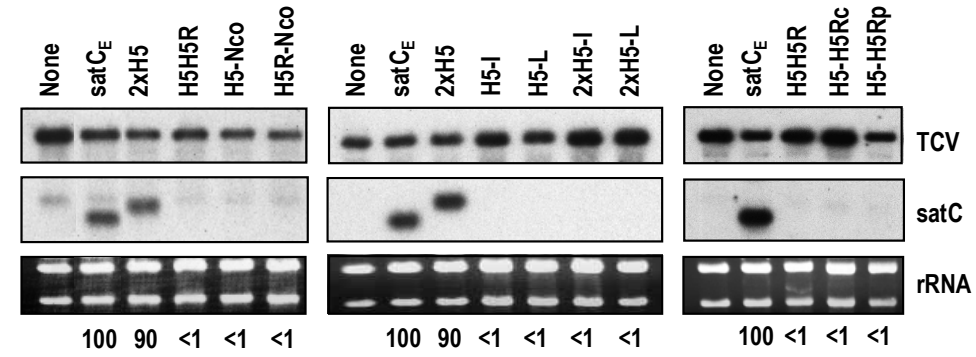
Deletion analysis suggested that deletion of H5 had the most detrimental effect on satC accumulation in protoplasts (Figure 2.4). To further determine the function of satC H5, duplication and ectopic positioning of H5 were performed at the *Nco*I site located about 170 bases upstream of the natural H5 site. This location was previously demonstrated to transiently tolerate up to a 60 base insertion (Simon et al., 1988). The parental satRNA for these constructs, $satC_E$, contains a two base alternation in the linker sequence between H5 and Pr, producing a new *Eco*RV site in the plasmid, to aid in cloning. $SatC_E$ accumulated to 67% of wild-type levels in protoplasts.

Insertion of a second copy of H5 in the same orientation at the *Nco*I site (generating 2xH5) did not substantially affect satC accumulation in protoplasts while insertion of a second copy in the reverse orientation (H5H5R) eliminated detection of satC. Ectopic positioning of H5 in either orientation (H5-Nco and H5R-Nco) resulted in undetectable levels of satC in protoplasts (Figure 2.6). These data suggested that a second copy of H5 did not enhance satC accumulation and that satC H5 was defunct when ectopically positioned.

Since insertion of a second copy of H5 in the opposite orientation differently affected satC accumulation, I wanted to know if the second copy of H5 in construct 2xH5 could complement the original H5 when containing a mutation that eliminated function and why the second copy of H5 in H5H5R is detrimental to satC accumulation. H5-I and H5-L contained alterations within the large symmetrical loop and lower stem of H5, respectively. Both mutations eliminated detection of satC (+)-strand at 40 hpi suggesting that both sequence and structure are important for H5 function (Figure 2.6). These

Figure 2.6 Duplication and ectopic positioning of H5. (A) Sequence and structure of H5 is shown. The positions of mutations are indicated. (B) Schematic representation of constructs used in duplication and ectopic positioning study of H5. Black line represents wild-type satC sequence. The corresponding location of the *Nco*I site is shown. The black arrowhead indicates wild-type satC H5 sequence. The blank arrowhead represents satC H5 sequence in reverse orientation. The grey arrowhead represents H5 with mutations as indicated below. (C) RNA gel blot of viral RNAs accumulating in protoplasts at 40 hpi. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot). Values given below the blots are the averages of at least two independent experiments, with the $satC_E$ level arbitrarily assigned a value of 100. TCV, TCV genomic RNA (+)-strand. satC, satC (+)-strand monomer. None, no satC in the inoculum.


A
A A
G A
C G
C G
C G
U U
A G
A G
A G
281 A C
280 U U
C U
A U
C G 304
C G
A G
A G
U A
G C
G C
G C
H5
B
satC$_E$ 5' NcoI
2xH5 5'
H5H5R 5'
H5-Nco 5'
H5R-Nco 5'
H5-I 5'
U280C/A281U
H5-L 5'
G304U
2xH5-I 5'
U280C/A281U
2xH5-L 5'
G304U
H5-H5Rc 5'
U280C/A281U
H5-H5Rp 5'
G304U


C
None
satC$_E$
2xH5
H5H5R
H5-Nco
H5R-Nco
None
satC$_E$
2xH5
H5-I
H5-L
2xH5-I
2xH5-L
None
satC$_E$
H5H5R
H5-H5Rc
H5-H5Rp
TCV
satC
rRNA
100 90 <1 <1 <1
100 90 <1 <1 <1 <1
100 <1 <1 <1

alterations were introduced into the original H5 in construct 2xH5 (generating 2xH5-I, 2xH5-L) and the second copy of H5 in construct H5H5R (generating H5-H5Rc, H5-H5Rp), respectively. All these mutants accumulated to undetectable levels in protoplasts, suggesting that the second copy of H5 in 2xH5 could not functionally complement mutations in the original H5, and that the detrimental effect of the second copy of H5 in H5H5R was not due to insertion of a functional H5 in (-)-strands of satC. A possible explanation for why insertion of H5 in reverse orientation was detrimental to satC accumulation is that the ectopically positioned H5 may cause formation of long double strand structure with the natural H5, which would eliminate function of H5.

Replacement of H4a, H4b and H6 with their reverse complement severely reduced satC accumulation

To better understand the structure and function of H4a, H4b and H6, satC constructs were generated containing replacement of each hairpin by their reverse complement (Figure 2.7A). SatC with the reverse complement of H4a (CH4aR), which did not disrupt the stem but altered each position in the loop (Figure 2.7A), accumulated to 42% of wild-type satC levels (Figure 2.7B). In contrast, satC with the reverse complement of H4b (CH4bR) and H6 (CH6R), which contained both sequence and structural alterations, accumulated to only 3% and 10% of wild-type levels. These results confirmed that these hairpins were important for satC accumulation and suggested that both stem and loop sequences contributed to the function of H4a.

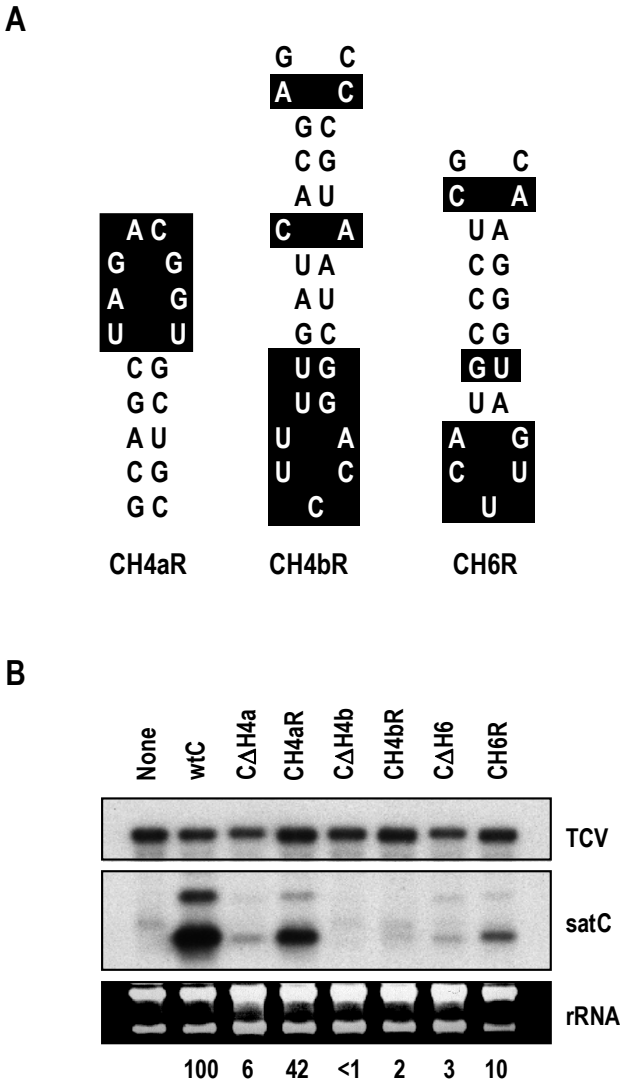


Figure 2.7 H4a, H4b and H6 are important for satC accumulation in protoplasts. (A) SatC containing either precise deletions of H4a (CΔH4a), H4b (CΔH4b), or H6 (CΔH6) or with H4a, H4b or H6 in their reverse complement orientations (CH4aR, CH4bR, CH6R, respectively) were inoculated onto protoplasts with TCV genomic RNA. Total extracted RNAs were examined by RNA-gel blot analysis at 40 hpi. The blot was probed with an oligonucleotide specific for both TCV genomic RNA and satC. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot). Numbers below the blot represent average values for three independent experiments. None, no satRNA added to the innoculm; wtC, wild-type satC. (B) Sequences and predicted structures of the reverse complement of H4a (H4aR), H4b (H4bR) and H6 (H6R). Sequences that differ with wild-type H4a, H4b and H6 are boxed.

Discussion

Deletion analysis showed that none of the hairpins on satC (+)-strands tested in this chapter could be deleted without affecting satC accumulation in protoplasts (Figure 2.4). These results suggested that all these hairpins might be cis-acting elements required for efficient satC replication. However, previous studies on TCV DI RNA suggested that accumulation of diG in plants (Li and Simon, 1991) and in protoplasts (Zhang and Simon, 1994) was affected by the size of RNA. In the case of satC, while decrease in size might contribute to the effect of deletions on satC accumulation in protoplasts, it seems not the principal cause for the following reasons: (i) although replacement of a 22 nt sequence between position 79 to 100 (the sequence which has been deleted in construct CΔH2) with unrelated sequences did not affected satC accumulation in planta (Carpenter et al., 1991a), a recent study showed this change seriously reduced satC accumulation in protoplast (F. Zhang and A.E. Simon, unpublished results); (ii) relocation of M1H (Nagy et al., 1999) or replacement of M1H with random sequences (Zhang and Simon, 2003b) reduced satC accumulation to less than 30% of wild-type; (iii) relocation of H5 as well as mutations in H5 eliminated satC accumulation in protoplasts (Figure 2.6); (iv) satC constructs containing replacement of H4a, H4b and H6 by their reverse complement accumulated to 42, 3 and 10% of wild-type levels, respectively (Figure 2.7).

While all individual hairpin deletions significantly reduced accumulation of satC (+)-strand monomers by over 16-fold compared to wild-type, they differentially affected accumulation of satC (+)-strand dimers. Deletion of M1H and H2 enhanced accumulation of satC (+)-strand dimers by 1.5-and 2.3-fold, respectively, and led to a 20-(CΔM1H) and

38-fold (CΔH2) decrease in the ratio of (+)-strand monomeric to multimeric forms. These deletions had a more limited effect on accumulation of (-)-strand monomers and dimers as well as the ratio of (-)-strand monomers and dimers (up to 2-fold). These results suggest that enhanced dimer synthesis likely happened during (+)-strand synthesis. This supports the model proposed previously in the Simon lab that satC multimers are formed during (+)-strand synthesis by a replicase-driven copy-choice mechanism and not a rolling circle mechanism (Carpenter et al., 1991a). Unlike the generation of DI RNA dimers of *Cymbidium ringspot tombusvirus* (CyRSV), which is size-dependent (Dalmay et al., 1995), generation of satC dimers seems not correlated to size since deletion of H4b (27 nt) reduced dimer accumulation to 2% while deletion of M1H (28 nt) increased dimer accumulation by 1.5- fold (Figure 2.4). In addition, while CΔ5P/M1H (40 nt deletion) and CΔM1H/H2 (50 nt deletion) retained 10 to 40% dimers, CΔ5P/H2 (34 nt deletion), CΔH5 (42 nt deletion) and CΔ5P/M1H/H2 (62 nt deletion) accumulated barely detectable levels of dimers. It is possible that deletion of M1H and H2 might lead to conformational changes in satC (-)-strand RNAs that resulted in a structure such that the 5' and 3' termini are much closer to each other compared to wild-type, thereby favoring replicase resumption of synthesis before release of the nascent strand.

An interesting finding is that deletions of 3' proximal hairpins (H4a, H4b and H5) were more detrimental to (-)-strand RNA accumulation compared to deletions of 5' proximal hairpins (M1H, H2 and H6) (Figure 2.4). These results suggested that H4a, H4b and H5, along with Pr, might function together to achieve wild-type levels of (-)-strand synthesis while the other hairpins are mainly involved in (+)-strand synthesis. Since H4a, H4b and H5 are also predicted to form in other members of the *Carmovirus* genus, these

findings provide additional evidence that phylogenetically conserved structures are often relevant to function.

Deletion of H5 had the most detrimental effect on satC replication. The function of H5 is position-dependent. H5 contains a large symmetrical loop (LSL) and four of the seven bases on the 3' side of LSL (5'GGGC) can potentially pair with the 3' terminal nucleotides of satC (GCCC-OH). Recently, this interaction was demonstrated to repress (-)-strand synthesis in vitro by sequestering the satC 3' terminus from the RdRp (Zhang et al., 2004), which is similar to the replication silencer found in TBSV.

Results from mutational analyses did not support the existence of a pseudoknot between H4a and its 5' side flanking sequence. However, a single mutation (G218C) in the sequence flanking the 5' side of H4a significantly reduced satC accumulation in protoplasts. These results suggest that this region is important for satC replication. Why does C229G compensate for G218C? One possibility is that G218 is located within a ACGG motif at positions 216 to 219. A cytidylate to guanylate change at position 229 forms a second ACGG motif (positions 227 to 230) within the H4a loop. In G218C/C229G, the second ACGG motif may functionally compensate for the mutated motif at positions 216 to 219.

My studies in this chapter suggest that replication of satC may involve several cis-acting elements and be more complicated than expected. H5 plays an essential role in satC replication. The sequence and structure requirements for satC H5 function and possible interactions between H5 and other regions of satC will be further described in the chapters III-V.

CHAPTER III

ANALYSIS OF A VIRAL REPLICATION ELEMENT: SEQUENCE REQUIREMENTS FOR A LARGE SYMMETRICAL INTERNAL LOOP

Introduction

Positive-strand RNA viruses contain *cis*-acting elements in their 5' and 3' untranslated regions (UTRs) that play indispensable roles in replication of the viral genome (Duggal et al. 1994; Buck, 1996). 3' UTR elements such as tRNA-like structures or other hairpins, poly(A) tails, or heteropolymeric single-stranded sequences have important roles in diverse viral functions including promotion of minus-strand synthesis, regulation of replicase access to the site of transcription initiation, telomeric-type protection of 3' end sequences, packaging nucleation signals, modulation of translation, and targeting of RNA to specific subcellular sites (Buck, 1996; Dreher, 1999). Since many of these functions have related effects on genome amplification, subviral RNA replicons have proven useful for identifying elements involved primarily in replication.

The association of TCV with several subviral RNA replicons makes it an ideal system with which to identify sequences specifically involved in RNA replication. These subviral RNA replicons include satC (356 bases) and satD (194 bases). As discussed earlier, satC is an unusual chimeric RNA sharing 88% similarity with nearly full-length satD at its 5' end. The 3' portion of satC originated from two regions in the 3' end of TCV, with the related regions sharing 94% sequence similarity (Simon and Howell,

1986) (Figure 1.2B). Using a combination of computer *mfold* modeling and phylogenetic comparisons of carmoviral 3' end region, four hairpins (Pr hairpin, H5, H4a and H4b) were revealed to be structurally and spatially conserved (to varying extents) among carmoviruses (Figure 1.4, Figure 1.5). Deletion and mutational analysis suggested that H5, located proximal to the Pr of satC, is crucial for satC accumulation in protoplasts (Figure 2.4) and its function is position-dependent (Figure 2.6). H5 contains a large, 14 base symmetrical internal loop (LSL). Four of the seven bases on the 3' side of the LSL (5'GGGC) can potentially base pair with the satC 3' terminal bases (GCCC-OH) (Zhang et al., 2004, Figure 1.4). Deletion of the satC 3' terminal three cytidylates or mutations that altered the 3' side of the LSL or H5 structure enhanced synthesis of both full-length and aberrantly initiated complementary strands in an in vitro transcription assay using purified TCV p88. Compensatory exchanges of putative base pairs between the 3' side of the LSL and 3' terminal bases restored near normal accumulation of complementary strands in this assay. In addition, solution structure analysis indicated that deletion of the 3' terminal three bases (CCC-OH) had a substantial effect on the structure of H5 and 3' proximal sequences without major modification of other sequences (Zhang et al., 2004).

In this study, I performed site-specific mutagenesis and in vivo SELEX to analyze sequence requirements of the satC H5 LSL for satC accumulation in vivo. My findings demonstrate that nearly all positions in the middle and upper regions of the LSL are crucial for replication of satC in protoplasts and fitness to accumulate in vivo. My results also suggest that H5 might have additional functions besides 3' end interaction. I also examined possible interactions between the H5 LSL and other regions of satC.

Materials and Methods

Construction of satC mutants

Mutations were introduced into the satC H5 LSL by PCR using pT7C+ (Figure. 2.3). All constructs used in this chapter are listed in Table 3.1. For mutations in the 5' side of the LSL, primers were MC279-285 and Oligo7 (primers used in this chapter are listed in Table 3.2). For construction of additional mutations in positions 279 to 281, primers were Oligo 7 and C279M, C280M or C281M, respectively. Following digestion with *Spe*I and *Sma*I, the fragment was inserted into the analogous location in pT7C+, which had been treated with the same restriction enzymes. For mutations in the 3' side of the LSL, PCR was performed with primers T7C5' and PSC81 and template pT7C+. For additional mutations at position 296 and 302, primers used were T7C5' and PSC82. PCR products were ligated into the *Sma*I site of pUC19.

Mutations were introduced into the terminal loop of satC H4a by PCR using pT7C+ and primers C5' and C230G, U231CA or C232AU, respectively. PCR products were subsequently digested with *Spe*I and *Sna*BI and cloned into pT7C+ replacing the analogous wild-type fragment. To generate satC containing mutations in both H5 and H4a, construct G230C was digested with *Spe*I and *Sna*BI, and the small fragment was cloned into construct C300G that had been treated with the same restriction enzymes.

To obtain plasmid C_B-C300G, plasmid C300G was digested with *Spe*I and *Sma*I, and the small fragment was inserted into pT7C_B replacing the analogous wild-type fragment. pT7C_B contains an alteration at position 176 (A to G), creating a *Bam*HI site that permitted distinguishing between satC containing reversions of the original mutations and possible contamination of wild-type satC (Zhang and Simon, 2003b).

TABLE 3.1 Summary of satC mutants used in Chapter III

Name	Description
U285A	SatC with a U to A change at position 285
U285G	SatC with a U to G change at position 285
U285C	SatC with a U to C change at position 285
A284C	SatC with an A to C change at position 284
A283G	SatC with an A to G change at position 283
A283C	SatC with an A to C change at position 283
A283U	SatC with an A to U change at position 283
A282U	SatC with an A to U change at position 282
A281G	SatC with an A to G change at position 281
A281C	SatC with an A to C change at position 281
A281U	SatC with an A to U change at position 281
U280A	SatC with a U to A change at position 280
U280G	SatC with a U to G change at position 280
U280C	SatC with a U to C change at position 280
C279A	SatC with a C to A change at position 279
C279G	SatC with a C to G change at position 279
C279U	SatC with a C to U change at position 279
U296C	SatC with a U to C change at position 296
G297A	SatC with a G to A change at position 297
G297C	SatC with a G to C change at position 297
G297C-1	SatC with a C to G change at position 300 and a G to A change at position 230
G297C-2	SatC with a C to G change at position 300, a G to A change at position 230, and a CU deletion between position 264 and 269
G298A	SatC with a G to A change at position 298
G298C	SatC with a G to C change at position 298
G299A	SatC with a G to A change at position 299
C300G	SatC with a C to G change at position 300
C300G-1	SatC with a U to G change at position 302 and a G to A change at position 230
C300U	SatC with a C to U change at position 300
U301A	SatC with a U to A change at position 301
U302A	SatC with a U to A change at position 302
U302G	SatC with a U to G change at position 302
G230C	SatC with a G to C change at position 230
U231A	SatC with a U to A change at position 231

TABLE 3.1 continued

Name	Description
U231C	SatC with a U to C change at position 231
C232A	SatC with a C to A change at position 232
C232U	SatC with a C to U change at position 232
C300G/G230C	SatC with a C to G change at position 300 and a G to C change at position 230
C_B	SatC with an A to G change at position 176, creating a *Bam*HI site between positions 175 and 180 (Zhang et al., 2003b)
C_B-C300G	C_B with a C to G change at position 300
C_B-G353C	C_B with a G to C change at position 353
C_B-C300G/ G353C	C_B with a C to G change at position 300 and a G to C change at position 353
C_B-G353A	C_B with a G to A change at position 353
C_B-C300G/ G353A	C_B with a C to G change at position 300 and a G to A change at position 353
C_B-C300G/ G353U	C_B with a C to G change at position 300 and a G to U change at position 353
G353A	SatC with a G to A change at position 353
C300U/G353A	SatC with a C to U change at position 300 and a G to A change at position 353
G353C	SatC with a G to C change at position 353
C220G	SatC with a C to G change at position 220
C300G/ G353C	SatC with a C to G change at position 300 and a G to C change at position 353
G353C/C220G	SatC with a G to C change at position 353 and a C to G change at position 220
G353C/C300G/C220G	SatC with a G to C change at position 353, a C to G change at position 300, and a C to G change at position 220
C50G	SatC with a C to G change at position 50
C129G	SatC with a C to G change at position 129
C165G	SatC with a C to G change at position 165
G218C/C165G	SatC with a G to C change at position 218 and a C to G change at position 165
A167U	SatC with an A to U change at position 167
G218C/ A167U	SatC with a G to C change at position 218 and an A to U change at position 167
C165G/ A167U	SatC with a C to G change at position 165 and an A to U change at position 167
G218C/ C165G/ A167U	SatC with a G to C change at position 218, a C to G change at position 165 and an A to U change at position 167
U296A	SatC with a U to A change at position 296
U296A/U285C	SatC with a U to A change at position 296 and a U to C change at position 285
A167U/U296A	SatC with an A to U change at position 167 and a U to A change at position 296
A167U/U296A/U285C	SatC with an A to U change at position 167, a U to A change at position 296 and a U to C change at position 285
A167U/U285C	SatC with an A to U change at position 167 and a U to C change at position 285
pNco-C277	Truncated satC containing the 5' 277 nt and a 4 nt (CAUG) insertion between positions 104 and 105

TABLE 3.2 Summary of the oligonucleotides used in Chapter III

Application/ construct	Name	Position[a]	Sequence[b]	Polarity[c]
Mutagenesis in satC	T7C5'	1-19	5'-*GTAATACGACTCACTATA*GGGAUAACUAAGGGTTTCA	+
	Oligo 7	338-356	5'-GGGCAGGCCCCCCGTCCGA	-
	MC279-285	257-300	5'-ggacACTAGTGCTCTCTGGGTAACCANNNNNNNCCCGAAAGGGTGGGC	+
	C279M	257-294	5'-gcACTAGTGCTCTCTGGGTAACCANTAAAATCCCGAAAGG	+
	C280M	257-294	5'-gcACTAGTGCTCTCTGGGTAACCACNAAAATCCCGAAAGG	+
	C281M	257-294	5'-gcACTAGTGCTCTCTGGGTAACCACTNAAATCCCGAAAGG	+
	PSC81	276-356	5'-GGGCAGGCCCCCCGTCCGAGGAGGGAGGCTATCTATTGGTTCGGAGGGTCCCCANNNNNNNCCCTTTCGGGATTTTAGTGG	-
	PSC82	275-356	5'-GGGCAGGCCCCCCGTCCGAGGAGGGAGGCTATCTATTGGTTCGGAGGGTCCCCANAGCCCNCCCTTTCGGGATTTTAGTG	-
	G230C	220-263	5'-GACTAGTTTTCCAGGCTAATGCCCGCAGCTAGAGGGTGCTGCCG	+
	U231CA	220-263	5'-GACTAGTTTTCCAGGCTAATGCCCGCAGCTAGKCGGTGCTGCCG	+
	C232AU	220-263	5'-GACTAGTTTTCCAGGCTAATGCCCGCAGCTAWACGGTGCTGCCG	+
	G353N	337-356	5'-GGGNAGGCCCCCCGTCCGAG	-
	C50G	30-85	5'- CTGCTACGTAGGGGTCATCAGTCAAGTTTAGGCATCGGTGTTCTGCATTAGTTGCG	-
	C129G	97-142	5'- gTAACCATGGTGGGTTTTTAAAGGCGGGAGTTCGCATCAAGTACGGG	+
	C165S/ A167W	99-181	5'- ccgACCAUGGUGGGUUUUUAAAGGCGGGAGUUCCCAUCAAGUACGGGAGCGUGAAAACCUGGCUGUUUCSCWCTCAAAAGAAUCCC	+
	Oligo277	261-277	5'-GGTTACCCAGAGAGCAC	-
SELEX	3CLS	261-356	5'-GGGCAGGCCCCCCGTCCGAGGAGGGAGGCTATCTATTGGTTCGGAGGGTCCCCANNNNNNNCCCTTTCGGGNNNNNNNTGGTTACCCAGAGAGCAC	-
	C5'	1-19	5'-GGGAUAACUAAGGGTTTCA	+
RNA gel blots	Oligo 13	249-269	5'-AGAGAGCACTAGTTTTCCAGG	-

[a] Coordinates correspond to those of satC. Oligo 13 is also complementary to positions 3950 to 3970 of TCV genomic RNA.

[b] Italic letters indicate T7 promoter sequence. Lowercase letters indicate extra bases added to achieve efficient digestion. "K" represents mixed base G, T. "N" represents randomized base. "W" represents mixed base A, T. "S" represents mixed base G, C. Mutations are denoted by underlined letters.

[c] "+" and "-" polarities refer to homology and complementarity with satC positive strand, respectively.

Plasmids C_B-G353C, C_B-G353A were generated by PCR using template pT7C_B and primers T7C5' and G353N. Following treatment with T4 polynucleotide kinase and T4 DNA polymerase, PCR products were ligated into the *Sma*I site of pUC19. Plasmids C_B-C300G/G353C, C_B-C300G/G353A and C_B-C300G/G353U were generated similarly except that plasmid C300G was used as template. Plasmid C300U/G353A was also generated in a similar fashion except that plasmid C300U was used as template. For construction of plasmids G353A, G353C, C300G/G353C, pT7C+ was digested with *Spe*I and *Nco*I, and the small fragment was cloned into plasmids C_B-G353A, C_B-G353C and C_B-C300G/G353C that had been treated with the same restriction enzymes, respectively. Plasmids G353C/C220G, G353C/C300G/C220G were constructed similarly except that plasmids G353C and G353C/C300G were used as vectors.

For construction of plasmid C50G, M13/pUC Reverse sequencing primer (New England BioLab) and oligonucleotide C50G were used in a PCR. The template was pT7C+. PCR products were digested with *Sna*BI and *Sph*I and ligated into similarly digested pT7C+, replacing the endogenous fragment. Plasmid C129G was obtained by PCR using oligonucleotides C129G and Oligo 277. PCR products were digested with *Spe*I and *Nco*I and inserted into pT7C+ to replace the similar sequence. Plasmids C165G, A167U and C165G/A167U were generated in a similar fashion except that oligonucleotides T7C5' and C165S/A167W were used as primers. To generate plasmids U296A and U296A/U285C, PCR was performed with primers T7C5' and U296A and templates pT7C+ and U285C, respectively. PCR products were cloned into the *Sma*I site of pUC19. Plasmids A167U/U296A and A167U/ U296A/U285C were generated

similarly except that the PCR products were digested with *Spe*I and the small fragments were cloned into plasmid A167U that had been treated with *Spe*I and *Sma*I.

pNco-C277, which was used for in vivo SELEX, was generated by PCR using pNco-C (Simon et al., 1988) and primers C5' and Oligo 277. PCR products were cloned into the *Sma*I site of pUC19. All mutations were identified and confirmed by sequencing.

Turnip plant inoculations, and cloning and sequencing of progeny virus

Two leaves each of three two-week old turnip seedlings were mechanically inoculated with 0.4 μg satC wild-type or mutant transcripts along with 4 μg of TCV genomic RNA [dissolved in 10 μl of distilled H_2O and then mixed with 10 μl of 2 x infection buffer containing 0.05 M glycine, 1% (w/v) bentonite, 1% (w/v) celite, and 0.03 M K_2HPO_4, pH 9.2). Total RNA was extracted from uninoculated leaves (described below) at 21 days post-inoculation (dpi). SatC-sized species were amplified by RT-PCR as described below and cloned into the *Sma*I site of pUC19. Complete full-length sequences were determined for each clone.

Extraction of RNA from turnip leaves

Plant tissue (without midrib) from one 2 inches long uninoculated leaf was ground in liquid nitrogen in a 50 ml beaker with a small pestle. The frozen leaf powder was transferred into a 1.5 ml eppendorf tube (should be less than 0.5 ml of tube volume) and extracted with 0.55 ml of RNA extraction buffer [25 mM EDTA, 0.4 M LiCl, 1% (w/v) SDS, 0.2 M Tris-HCl, pH 9.0] and 0.55 ml of H_2O-saturated phenol by vigorously

vortexing for 20 sec. The mixture was centrifuged at 13,000 rpm for 3 min. The aqueous layer was recovered and re-extracted with 0.5 ml of phenol/chloroform. Total RNA was then precipitated with ethanol and re-suspended in 0.3 ml of 2 M LiCl. Following centrifugation at 13,000 rpm for 5 min at 4°C, the pellet was dissolved in 0.3 ml distilled H_2O and precipitated with ethanol again. After washing with 70% ethanol, the pellet was dissolved in 50 μl of distilled H_2O. RNA concentration was estimated by measuring the absorbance at 260 nm.

RT-PCR using M-MLV reverse transcriptase

To generate first strand cDNA, 2.5 μg of total RNA was mixed with 10 pmol of Oligo 7 and distilled H_2O to bring the volume up to 10 μl. The mixture was heated at 75°C for 10 min and incubated immediately on ice to denature the RNA. The solution was then mixed with 4 μl of 5x reverse transcription buffer provided by the manufacturer, 2 μl of 5 mM dNTP mixture (containing 5 mM each of dATP, dGTP, dCTP and dTTP), 2 μl of 100 mM DTT and 1 μl of RNase Out (Invitrogen, 40 U/μl) and incubated at 37°C for 2 min before adding 1 μl of M-MLV reverse transcriptase (USB, 200 U/μl). The reaction was carried out at 37°C for 1 hour and stopped by heating at 75°C for 15 min and placed immediately on ice. PCR was performed using 8 μl of cDNA synthesized as described above with primers C5' and Oligo 7 in a 100 μl reaction.

In vivo SELEX

In vivo SELEX was performed as previously described (Guan et al., 2000a, b; Zhang and Simon 2003b; Figure 3.1, Figure 3.2). Full-length satC cDNAs containing

randomized bases in the H5 LSL were generated by PCR using pNco-C277 as template. Primers used were T7C5' and 3CLS, which is complementary to positions 261 to 356. Within this primer, positions 279 to 285 and 296 to 302 contained randomized sequence. Full-length PCR products were purified and directly subjected to in vitro transcription using T7 RNA polymerase. The number of cDNA molecules used for in vitro transcription of RNA to infect one plant was 4×10^{12}.

For the first SELEX round, 5 μg of satC transcripts containing randomized LSL sequence were mechanically inoculated onto each of 60 turnip seedlings along with 4 μg of TCV transcripts. Total RNA was extracted from uninoculated leaves at 21 dpi as described above. Viable satC species accumulating in two plants showing satC symptom and 8 otherwise randomly selected plants were recovered by RT-PCR using primers C5' and Oligo 7, cloned into the *Sma*I site of pUC19 and sequenced. For the second round, equal amounts of leaf tissue from the 60 plants were combined, total RNA extracted and then inoculated (~5 μg/plant) onto six turnip seedlings. For the third round, equal amounts of total RNA, extracted from each plant of the previous round, were pooled and then inoculated onto six turnip seedlings (~5 μg/plant). SatC species at 21 dpi were cloned and the full-length sequence determined.

Competition of SELEX winners in plants

For competition between wild-type satC and the most fit winner, equal amounts of transcripts of wild-type satC and the most fit winner were combined and used to inoculate three turnip seedlings (0.4 μg/plant) along with TCV genomic RNA transcripts (4 μg /plant). For competition among the four SELEX winners, equal amounts of

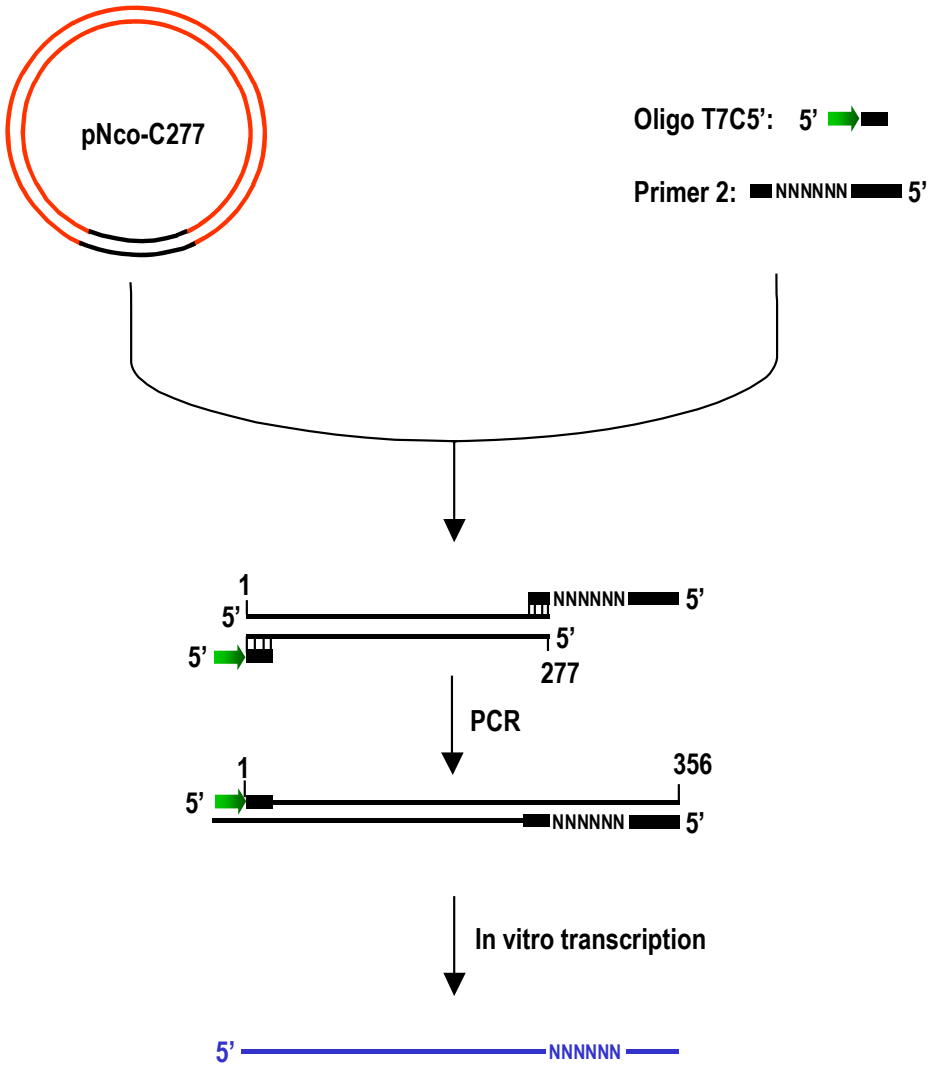


Figure 3.1 Preparation of in vitro transcripts used for in vivo SELEX. Plasmid pNco-C277 contains truncated satC cDNA (5' 277 nt and a 4 nt insertion between position 104 to 105). Sequence in the target region was randomized by PCR using a oligonucleotide containing mixed four nucleotides at each position in the target sequence. Base-pairings between satC cDNA and primers are shown. Full-length PCR products were purified and directly subjected to in vitro transcription using T7 RNA polymerase. The red, black, and blue lines represent the pUC19, satC cDNA and in vitro transcript sequences, respectively. The thick black lines represent oligonucleotides used as primers in PCR. The green arrow represents the T7 RNA polymerase promoter sequence. "N" denotes randomized nucleotide.

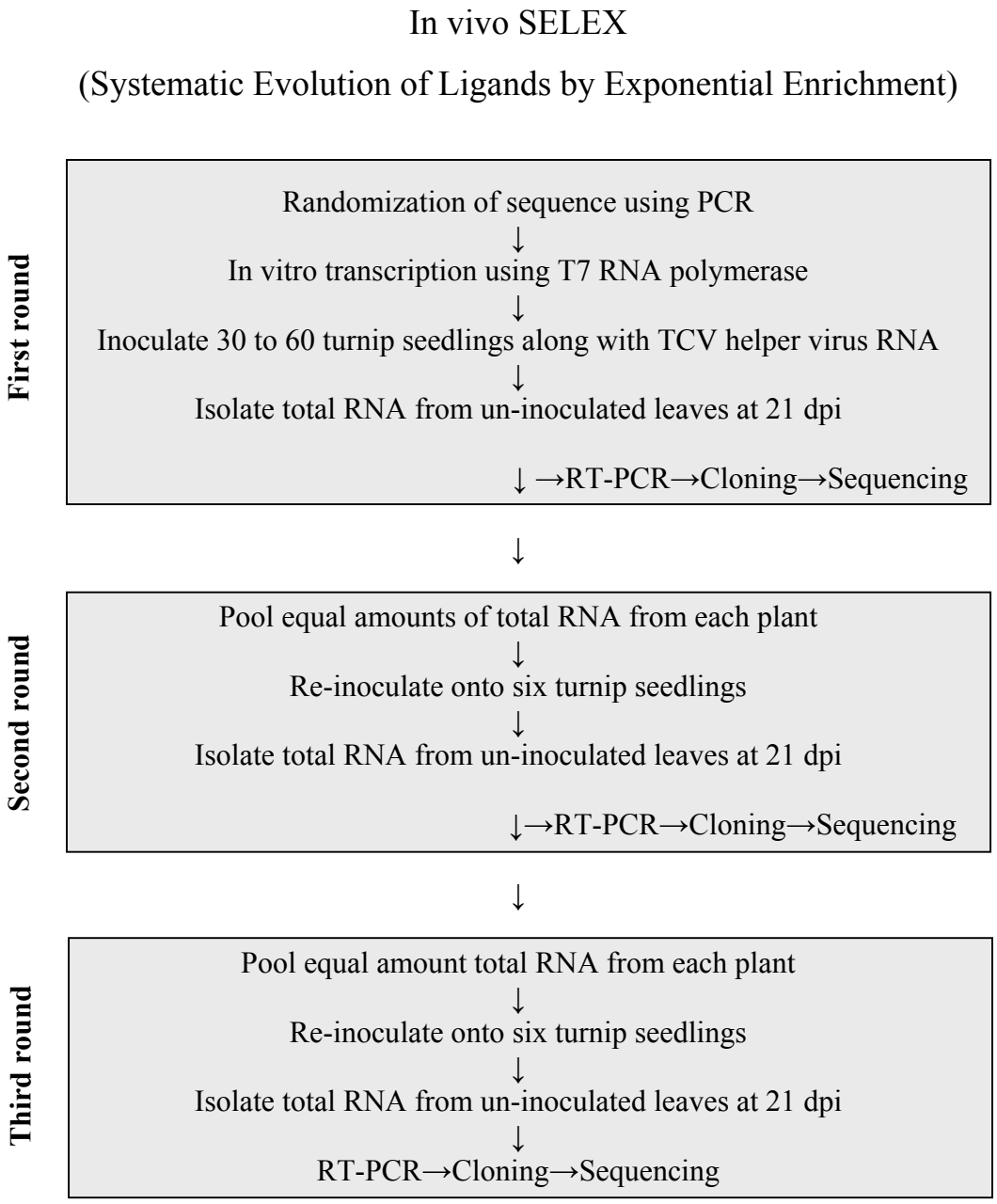


Figure 3.2 Flow chart of in vivo SELEX.

transcripts of each competitor were combined and used to inoculate three turnip seedlings (0.4 μg each/plant) along with TCV genomic RNA transcripts (4 μg/plant). SatC species from all plants at 21 dpi were cloned and assayed as described above.

In vitro transcription, protoplast preparation, inoculation, and RNA gel blots

SatC transcripts were synthesized from plasmids linearized with *Sma*I (for mutations in H5) or directly from PCR products (for SELEX winners, using primers T7C5' and Oligo 7) using T7 RNA polymerase. Protoplast preparation, inoculation, and RNA gel blots were performed as described in Chapter II.

Results

Examination of a possible interaction between the H5 LSL and the 3' terminal nucleotides in vivo

Computer generated RNA structure predictions and phylogenetic analyses indicate that four of the seven bases on the 3' side of LSL (5'GGGC) can potentially base pair with the 3' terminal nucleotides of satC (GCCC-OH) to form Ψ_1 (Figure 1.4). To determine whether this interaction exist, constructs that disrupt (C_B-C300G, C_B-G353C, C_B-G353A, C_B-C300G/G353A, C300G, G353C) or re-establish (C300U, C_B-C300G/G353C, C_B-C300G/G353U, C300U/G353A, C300G/G353C) putative base-pairing between the two sequences were generated in satC (Figure 3.3A, Figure 3.10A). Some of these constructs contained an alteration at position 176 (A to G), creating a *Bam*HI site that allowed for distinguishing between satC containing reversions of the

original mutations and possible contamination of wild-type satC. Three of these constructs (C300G, G353C, C300G/G353C) had been previously subjected to an in vitro RdRp assay (Zhang et al., 2004). C300G and G353C transcripts increased synthesis of full-length complementary strands by similar amounts compared with wild-type satC (10- and 11-fold, respectively) while also increasing aberrantly initiated products. The compensatory construct (C300G/G353C) produced much lower levels of full-length products compared with transcripts containing only G353C or C300G. These results strongly support the existence of an interaction between the 3'CCC and the H5 LSL that represses (-)-strand synthesis in vitro.

To determine the effect of these mutations on satC accumulation in vivo, transcripts were co-inoculated with wild-type TCV onto Arabidopsis protoplasts and satRNA levels examined at 40 hpi by Northern analysis. None of the satC constructs accumulated to detectable levels (Figure 3.3B, C) except for C_B-G353A, which was only revealed when the film was overexposed (Figure 3.3B). This suggests that disrupting the base-pairing interaction strongly inhibits satC accumulation. Furthermore, since the compensatory exchange did not restore satC viability, either C300 or G353 are likely essential for additional functions and thus the compensatory constructs are still non-viable. Alternatively, the compensatory constructs may produce excess (-)-strands as revealed by the in vitro RdRp assay. Transcription of (-)-strands using C300G/G353C resulted in a 3.7-fold elevation compared with wild-type satC (Zhang et al., 2004), which might be sufficient to disrupt satC accumulation in vivo.

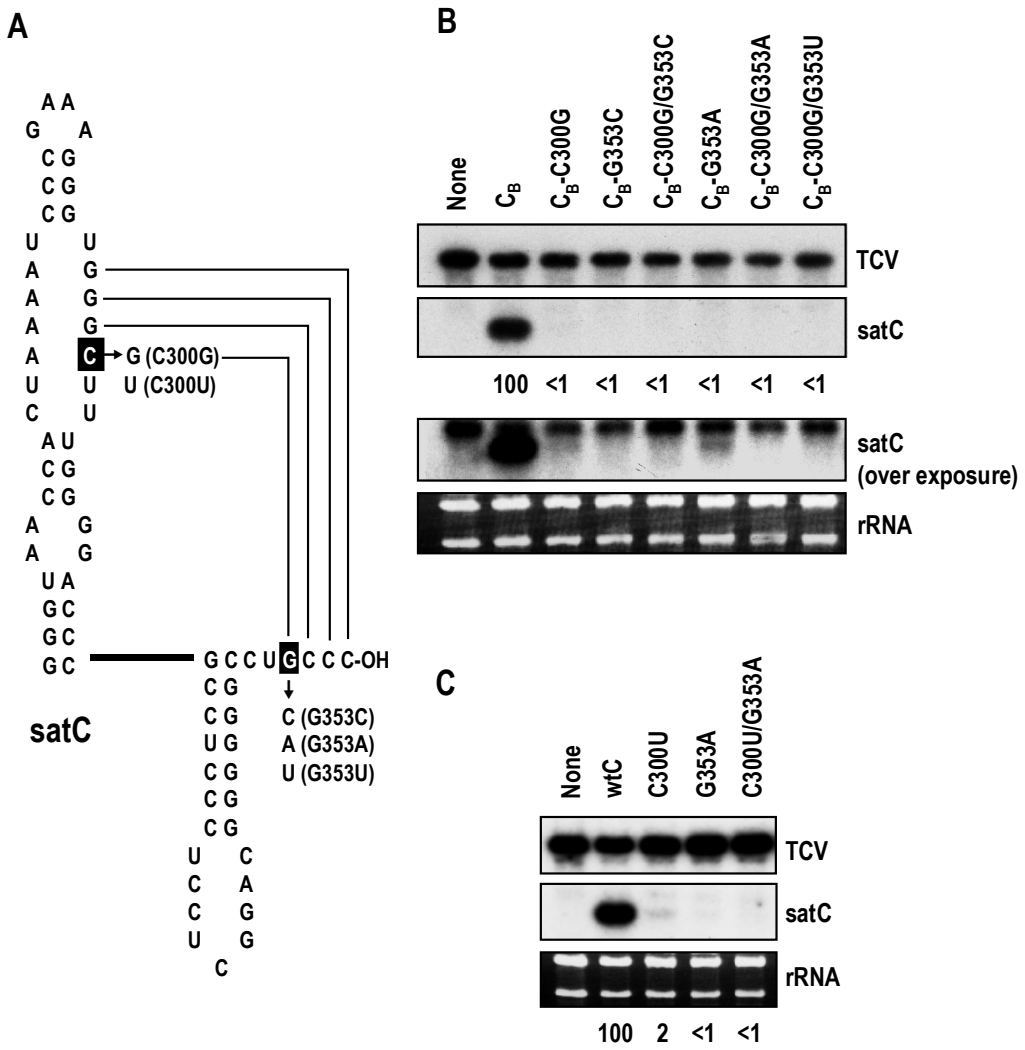


Figure 3.3 Compensatory exchanges between the satC LSL and a base near the 3' end. (A) Location of the point mutations generated in satC. Bases that were altered are boxed and arrows point to the new bases. Names of the altered constructs are shown. Possible base-pairing between the 3' terminal bases and the 3' side of the LSL is shown. (B) (C) Northern blot of mutant satC and helper virus (TCV) (+)-strands. Total RNA was extracted from protoplasts at 40 hpi and probed with an oligonucleotide complementary to both TCV and satC. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot). Values given below the blots are the averages of two independent experiments, with the C_B (B) or wild-type satC (C) level arbitrarily assigned a value of 100. None, no added satC. C_B, satC contains a A to C alteration at position 176, which generate a *Bam*HI site (5'GGAUCC). wtC, wild-type satC.

Single base changes on both sides of the H5 LSL can substantially reduce accumulation of satC in protoplasts

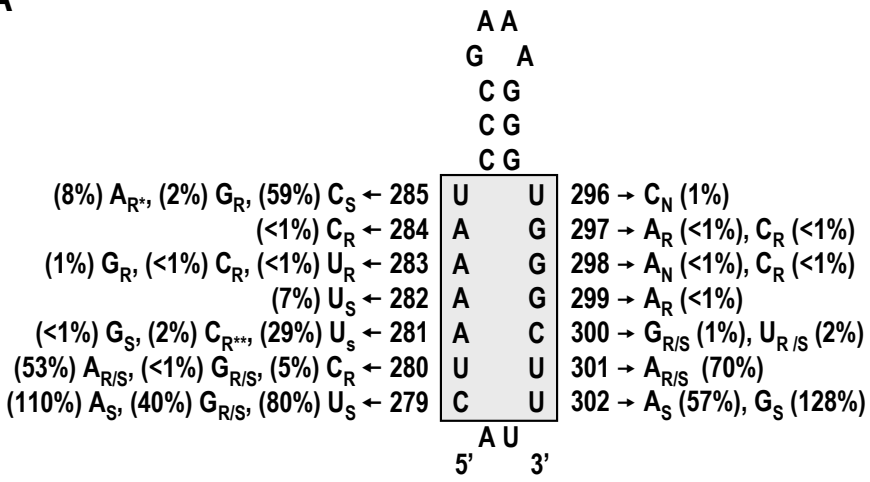
To determine experimentally the importance of specific positions in the LSL for satC accumulation in vivo, single base alterations at each position were generated in a full-length satC clone (Figure. 2.3). Transcripts synthesized from the mutant constructs were inoculated onto Arabidopsis thaliana protoplasts along with TCV helper virus, and accumulation of satC after 40 hours was determined by RNA gel blots (Figure 3.4).

SatC containing alterations at positions 279 and 285, the upper and lowest position in the LSL 5' side, were viable, although different mutations at these positions had different effects on satC accumulation. SatC containing U285A and U285G accumulated to only 8 or 2% of wild-type satC, respectively, while U285C was better tolerated, only reducing satC levels to 59% of wild-type. These results suggested that closing the upper position of the LSL by alterations allowing base-pairing (U285A and U285G) was strongly detrimental to H5 function. In contrast, alterations in the LSL 5' side that permit base-pairing in the lowest position of the LSL (C279G, C279A) accumulated to 40% and 110% of wild-type levels, respectively, indicating that closure of the lowest position of the LSL is permitted. Alteration of the four adenylates on the 5' side of the LSL (positions 281 to 284) was generally detrimental with the exception of A281, the 5' most adenylate, which retained 30% of wild-type accumulation when converted to a uridylate (A281U).

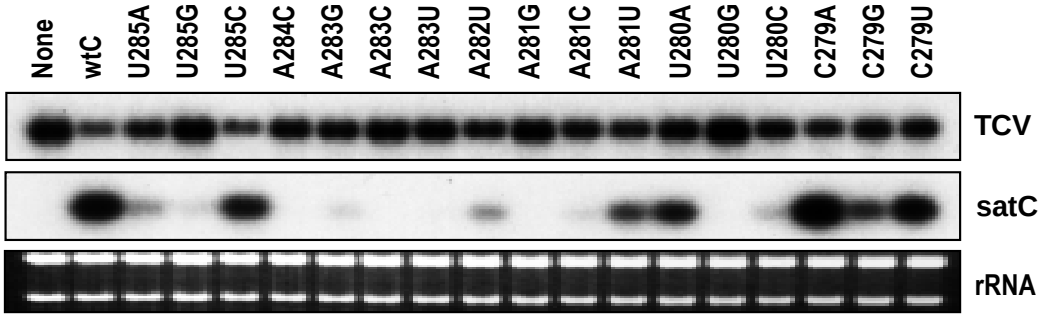
Mutations generated in the 3' LSL GGGC (positions 297 to 300), which were previously determined to base-pair with 3' terminal bases in vitro (Zhang et. al., 2004), were highly detrimental as expected. In addition, U296C and U301A, which flank the

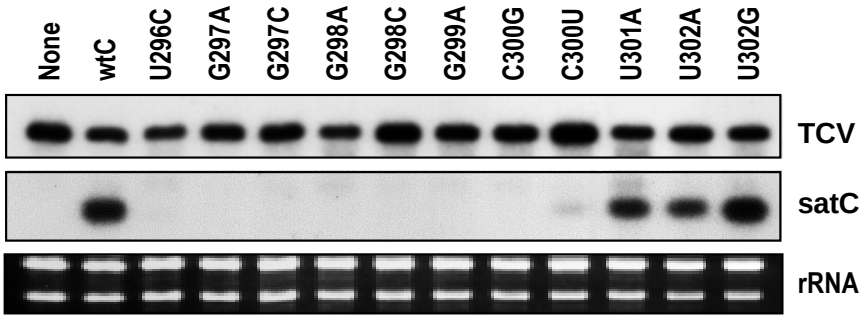
Figure 3.4 Single site mutational analysis of the satC H5 LSL in vivo. (A) Location of the mutations constructed in satC and summary of the effects on satC accumulation in protoplasts and plants. Numbering of bases (from the 5' end) is shown. Accumulation in protoplasts of the individual mutants expressed as a percentage of wild-type satC levels is given in parentheses. Stability of the mutations in plants was assayed at 21 dpi and is denoted by subscripts as follow: N, no satC species detected by PCR; R, mutations reverted to wild-type in all recovered clones; S, mutations were always maintained; R/S mutations either reverted to wild-type or were maintained (see Table 3.3). R* associated with U285A denotes mutation either reverted to the wild-type U or a C. R** associated with A281C denotes mutation either reverted to the wild-type or a U. (B) Representative RNA gel blot of mutant satC and helper virus (TCV) (+)-strands. Total RNA was extracted at 40 hpi from Arabidopsis protoplasts. Gels were stained with ethidium bromide before blotting to reveal rRNA loading control (below the blot). None, no added satC. wtC, wild-type satC.

A


A A
G A
C G
C G
C G
(8%) A$_{R*}$, (2%) G$_R$, (59%) C$_S$ ← 285 U U 296 → C$_N$ (1%)
(<1%) C$_R$ ← 284 A G 297 → A$_R$ (<1%), C$_R$ (<1%)
(1%) G$_R$, (<1%) C$_R$, (<1%) U$_R$ ← 283 A G 298 → A$_N$ (<1%), C$_R$ (<1%)
(7%) U$_S$ ← 282 A G 299 → A$_R$ (<1%)
(<1%) G$_S$, (2%) C$_{R**}$, (29%) U$_s$ ← 281 A C 300 → G$_{R/S}$ (1%), U$_{R/S}$ (2%)
(53%) A$_{R/S}$, (<1%) G$_{R/S}$, (5%) C$_R$ ← 280 U U 301 → A$_{R/S}$ (70%)
(110%) A$_S$, (40%) G$_{R/S}$, (80%) U$_S$ ← 279 C U 302 → A$_S$ (57%), G$_S$ (128%)
A U
5’ 3’

B


None
wtC
U285A
U285G
U285C
A284C
A283G
A283C
A283U
A282U
A281G
A281C
A281U
U280A
U280G
U280C
C279A
C279G
C279U
TCV
satC
rRNA


None
wtC
U296C
G297A
G297C
G298A
G298C
G299A
C300G
C300U
U301A
U302A
U302G
TCV
satC
rRNA

LSL bases involved in the putative 3' end interaction, reduced satC to undetectable levels. All three alterations tested in the two lowest positions on the LSL 3' side still allowed substantial accumulation of satC in protoplasts similar to most of the partner positions on the 5' side. Altogether, these results indicated that most positions within the middle and upper region of the LSL cannot be individually altered without substantially reducing satC accumulation in vivo. It is interesting to note that four carmoviruses (TCV, SCV, JINRV, and HCRSV) have identical LSL sequences in their respective H5 hairpins while most other carmoviruses have related sequences that differ at symmetrical positions from that of TCV, mainly at the base of the LSL (Zhang et al., 2004; Figure 1.4, Figure 1.5).

SatC containing single base changes in the LSL give rise to second site alterations in planta

To examine the effect of H5 LSL mutations on accumulation of satC in a natural host, individual mutants described in Figure 3.4A were inoculated together with TCV onto each of three turnip seedlings. At 21 dpi, satC present in total RNA preparations isolated from uninoculated leaves was amplified by RT-PCR and the resultant cDNAs cloned and sequenced. The results are summarized in Table 3.3 and Figure 3.4A. With only one exception, all mutants that accumulated to less than 7% of wild-type levels in protoplasts either did not accumulate to PCR-detectable levels in plants or some or all clones reverted to wild-type (Figure 3.4A, subscripts N, R and R/S; Table 3.3). The most damaging mutations, for which no progeny were detected, are located on the 3' side of the LSL, in or near the region proposed to interact with the 3' end to repress (-)-strand synthesis (U296C; G298A). SatC containing A281G gave rise to progeny that maintained

TABLE 3.3 Summary of progeny derived from satC containing mutations in the H5 LSL

Name	Maintained mutations	Reversions	Second site mutations[a]	Location of second site mutations[b]
U285A	0	12[c]	1(1)	U285-1: U285C, A282G
U285G	0	5	0	
U285C	2	0	0	
A284C	0	6	0	
A283G	0	3	0	
A283C	0	4	0	
A283U	0	9	0	
A282U	4	0	0	
A281G	2	0	0	
A281C	0	6[d]	0	
A281U	5	0	0	
U280A	1	3	0	
U280G	1	1	0	
U280C	0	2	0	
C279A	6	0	0	
C279G	4	1	0	
C279U	5	0	0	
U296C	0	0	0	
G297A	0	7	0	
G297C	0	5	2(4)	G297C-1: C300G, G230A G297C-2: C300G, G230A, ΔCU between C264 and U269
G298A	0	0	0	
G298C	0	4	0	
G299A	0	4	0	
C300G	4	1	2	C300G-1: U302G, G230A
C300U	2	2	0	
U301A	2	2	1(1)	U301A-1: A215G
U302A	5	0	0	
U302G	5	0	0	

[a] The number of clones that contained second site mutations with the number of second site mutations found in all progeny clones in parentheses.
[b] Name of the clone: bases altered.
[c] Eight clones reverted to wild-type, four had an adenylate to cytidylate transversion.
[d] One clone reverted to wild-type, five clones had a cytidylate to uridylate transition.

the alteration in plants despite having undetectable accumulation in protoplasts (Figure 3.4).

Nine of 10 mutations that permitted accumulation of satC to at least 7% of wild-type levels were either stably maintained in plants or only a portion reverted to wild-type (Figure 3.4A; Table 3.3). SatC containing U285A, which accumulated to 8% of wild-type levels, reverted to either the wild-type base (8 of 12 clones) or to a cytidylate (4 of 12 clones). Similarly, five of six progeny clones isolated from plants inoculated with satC containing A281C had a uridylate at this position while one contained a reversion to the wild-type adenylate (note that A281U was a stable alteration in plants; Figure 3.4A). Altogether, these results support a sequence specific requirement for nearly all residues in the middle and upper portion of the LSL for robust satC accumulation in plants and protoplasts.

Sequencing of 25 full-length satC clones derived from six plants inoculated with wild-type satC revealed no mutations in 7750 bases analyzed. Of the 127 full-length satC clones derived from mutants with alterations in the H5 LSL, six second-site changes were found in five clones isolated from four different plants (Table 3.3; Figure 3.5). One progeny clone derived from satC containing U301A had a reversion of the original mutation and an adenylate to guanylate transition at position 215 in the middle of five consecutive adenylates near the base of M1H. An adenylate to guanylate transition was also found for a single progeny of satC containing U285A, which additionally contained a primary site alteration to a cytidylate. The second site mutation was at position 282 in the LSL (A282G). Since the function of the 5' side of the LSL is unknown, it is not clear how or if this second site change compensated for the original mutation.

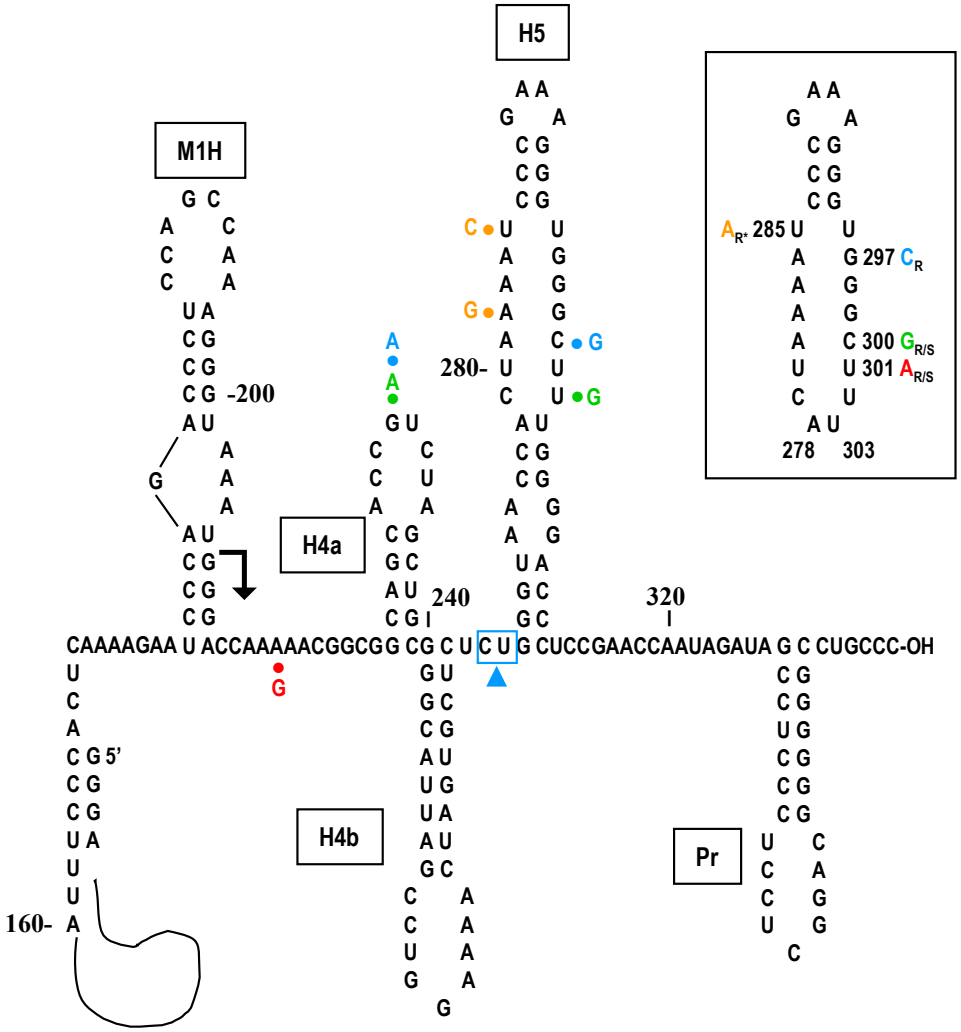


Figure 3.5 Location of second site mutations in the in progeny of satC H5 LSL mutants. The structure of satC 3' region predicted by RNA *mfold* computer program is shown. Arrow denotes contiguous 3' end region shared between satC and TCV. The names of the hairpins are given. Inset, the primary mutations in H5 that gave rise to second site mutations. R, S, R/S and R* subscripts were described in the legend to Figure 3.4. Second site mutations and their primary site progenitors are color-coded. A two base deletion found in progeny G297C-2 (Table 3.3) is marked with a triangle and boxed. Since this region contains three tandem CU bases, which bases were deleted cannot be determined.

Two clones derived from the same plant infected with satC containing G297C had reversions of the original mutations and multiple similar second site alterations (Table 3.3): progeny clone G297C-1 had a cytidylate to guanylate transversion within the LSL at position 300 and a guanylate to adenylate transition at position 230 in the loop of H4a (Figure 3.6A). Progeny clone G297C-2 contained these two alterations as well as a deletion of two bases between H4b and H5. The second site change within the LSL in these two related progeny clones could help to re-establish base-pairing with the 3' end prior to reversion of the original mutation (Figure 3.6A, middle). SatC containing C300G also produced a progeny clone (C300G-1) with similar second site alterations; a uridylate to guanylate transversion within the LSL at position 302 and a second mutation identical to that previously described in the loop of H4a (G230A). The alteration within the LSL could also strengthen base-pairing with the 3' end that would be disrupted by the original mutation (Fig. 3.6A, right).

C300G-1 was the only clone derived from C300G that did not contain a reversion of the original mutation, suggesting that the second site changes were compensatory. To test this possibility, the parental satC containing C300G and progeny clone C300G-1 were assayed for accumulation in protoplasts. At 40 hpi, only C300G-1 accumulated to detectable levels (19% of wild-type satC) (Figure 3.6B), supporting a compensatory effect for the second site alterations. The two clones derived from satC containing G297C were also tested for accumulation in protoplasts. However, neither the parental construct nor the progeny clones accumulated to detectable levels.

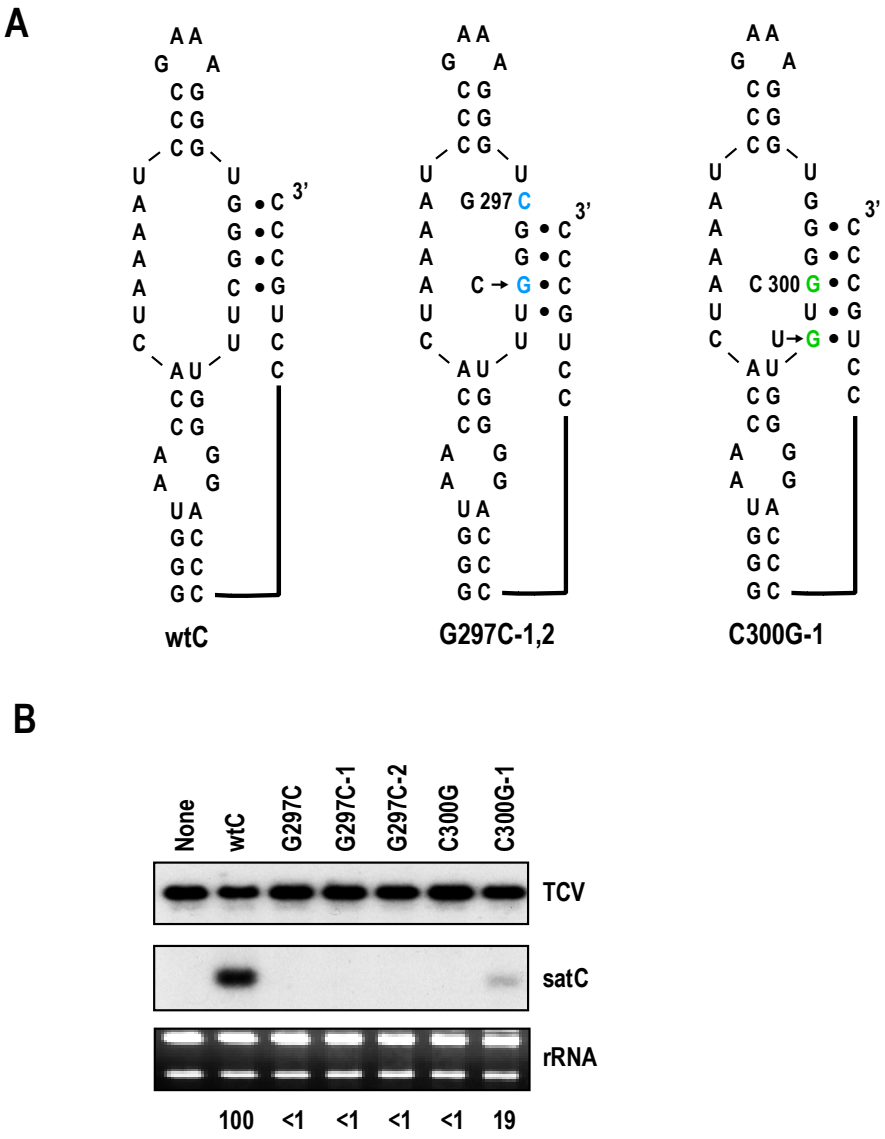


Figure 3.6 Possible contribution of second site changes within H5 to strengthening base-pairing of the LSL 3' side and the 3' terminus of satC. (A) Interaction between the LSL 3' side and the 3' terminus of satC. Progenitor mutations are indicated with their position number. Second site alterations are denoted by arrows preceded by the wild-type base. Names of the mutant clones containing the second site alterations are given below the structures. (B) Representative RNA gel blot of wild-type and mutant parental and progeny satC in protoplasts at 40 hpi. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot). Values given below the blots are the averages of three independent experiments, with the satC level arbitrarily assigned a value of 100. None, no added satC. wtC, wild-type satC.

Examination of a possible interaction between the H5 LSL and the loop of H4a

The finding of identical guanylate to adenylate transitions within the loop of H4a in three clones derived from two different LSL mutants (C300G and G297C) suggested that this alteration might be compensating for the original mutations. H4a is important for satC accumulation in protoplasts (Figure 2.4, Figure 2.5 and Figure 2.7) and exists at the same position relative to H5 and H4b in TCV and the related carmovirus CCFV (Figure 1.4, Figure 1.5). Analysis of the 3' side of the satC, TCV and CCFV LSLs and their respective H4a terminal loop sequences revealed possible base-pairing involving four (CCFV) or five positions (TCV and satC) that would be disrupted by the primary mutations in satC (Figure 3.7A). While these LSL bases have been previously shown to interact with the satC 3' end, it is possible that the loop of H4a might serve to help release the LSL from the 3' end by interacting with the LSL. The second site alteration in the H4a loop could help to re-establish base-pairing disrupted by the primary C300G and G297C mutations (Figure 3.7B).

To provide evidence for or against this possible interaction, mutations were introduced into the loop of satC H4a that either preserved (U231C, C232U, C300G/G230C), disrupted (U231A, C232A) or reduced (G230C) the putative base-pairing shown in Figure 3.8A. SatC with H4a alterations U231A or C232A accumulated to near wild-type levels in protoplasts, suggesting that disruption of this putative H4a/LSL interaction was not detrimental to satC accumulation (Figure 3.8B). However, G230C and U231C reduced accumulation of satC by 33% and 31% respectively, suggesting a role for the H4a loop in satC accumulation in protoplasts. The compensatory exchange between the LSL and H4a (C300G/G230C) was highly detrimental to satC

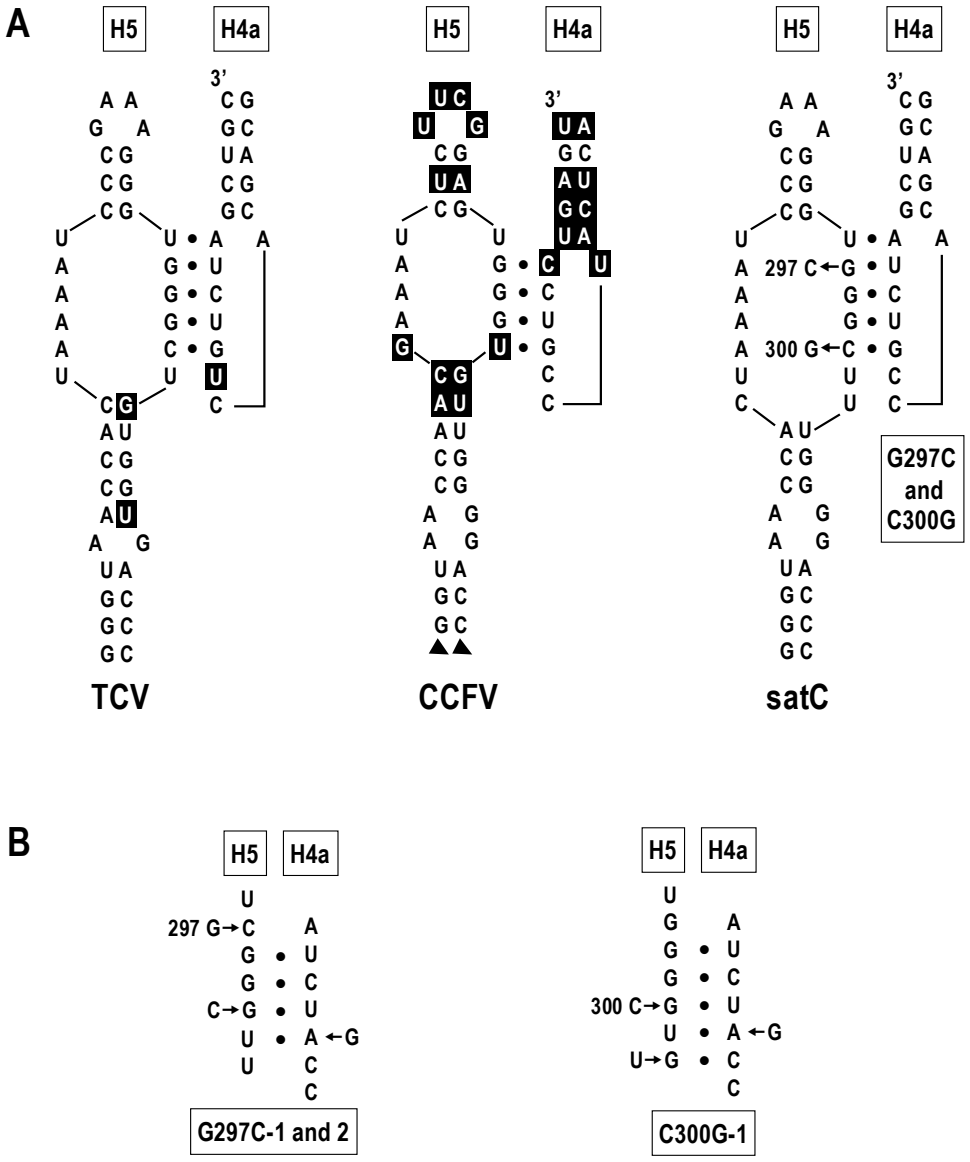


Figure 3.7 A possible H5/H4a interaction in satC. (A) Possible base pairing between H5 LSL and the loop of H4a. The alterations generating parental constructs G297C and C300G, which should disrupt this base-pairing, are shown. Differences between satC H5 and H5 of TCV and CCFV are boxed. Triangles represent absent bases. (B) Predicted base-pairing between H5 LSL and the loop of H4a in progeny of G297C (G297C-1 and G297C-2) and C300G (C300G-1) which contain second site alterations. Wild-type satC bases with positional numbers are shown with arrows pointing to the parental G297C (left) and C300G (right) alterations. Second site changes are similarly designated but without positional numbers. Progeny clones G297C-1 and G297C-2, derived from G297C in the same plant, differed from each other by G297C-2 having an additional two base deletion upstream of H5 (see Figure 3.5). Potential re-establishment of base-pairing due to the second site alterations in G297C-1,-2 and C300G-1 are shown.

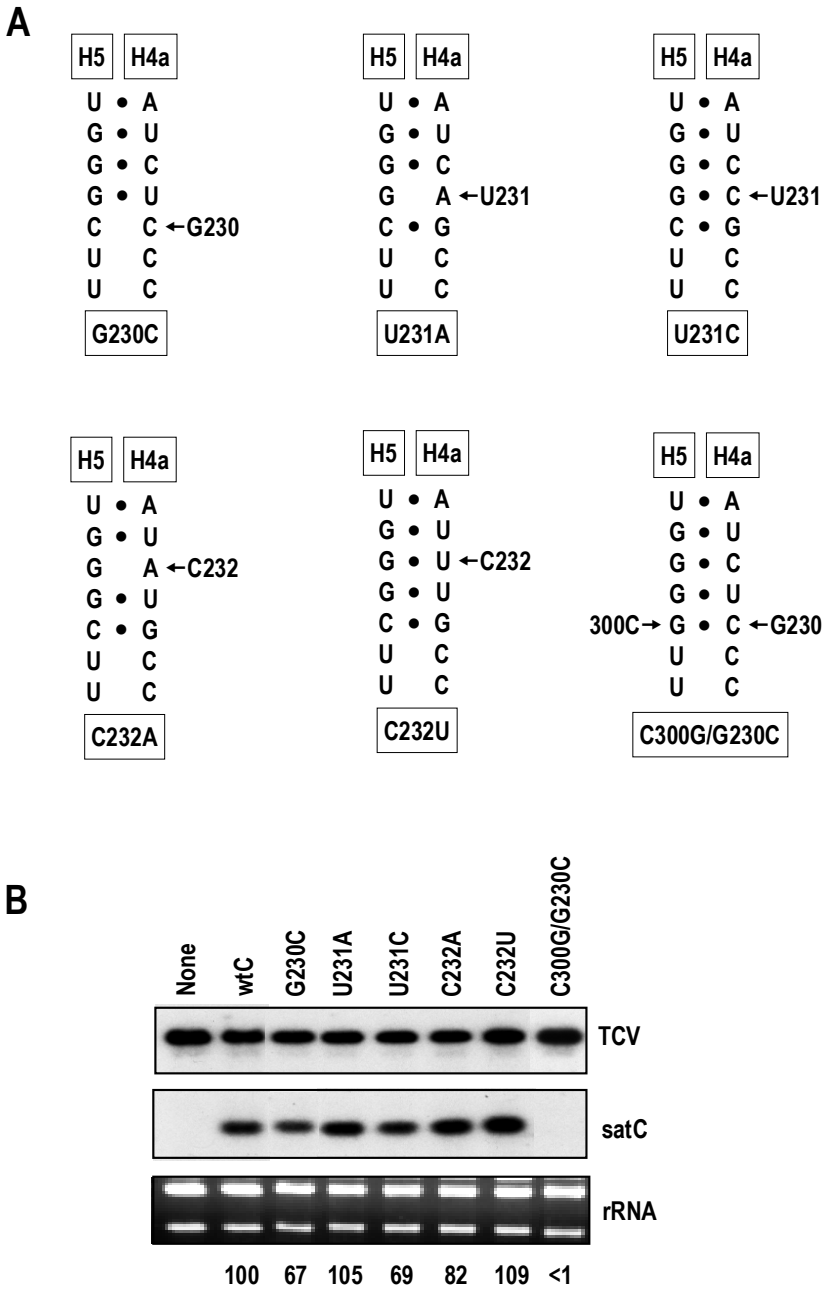


Figure 3.8 Mutational analysis of a possible H5/H4a interaction in satC. (A) Series of satC constructs containing mutations designed to disrupt or maintain base-pairing between H5 and H4a. (B) Representative RNA gel blot of wt and mutant satC accumulation in protoplasts at 40 hpi. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot). Values below the panels are percentages of wt levels of accumulation (average of three independent experiments). None, no added satC. wtC, wild-type satC.

accumulation (Figure 3.8B), indicating that addition of G230C did not compensate for the previously found reduction in satC levels due to C300G (Figure 3.6B). Therefore, while these results suggest a role for the H4a loop in satC accumulation, they do not at this time support the specific interaction presented in Figure 3.8A between H4a loop residues and the H5 LSL.

In vivo SELEX of the satC H5 LSL

Analysis of RNA structures by site-specific mutagenesis provides limited (and sometimes incorrect) information on the importance of specific residues since the alterations are tested in the context of remaining wild-type sequences (Carpenter and Simon, 1998). In other words, specific alterations deemed detrimental in the environment of wild-type sequences may not necessarily be detrimental if other bases in the structure (in addition to compensatory exchanges) are also altered. This may be especially true for internal symmetrical loops, where non-Watson Crick base-pairs likely permit continued interaction between the two strands due to stabilization by one or more hydrogen bonds between partners as well as same or cross strand stacking. If the structure of an internal loop is required for hairpin function, these non-Watson Crick base-pairs might be replaceable by other canonical or non-canonical base-pairs provided that they are isosteric with the original base pairs (Leontis et al., 2002b).

Therefore, to determine more precisely the necessity for specific bases in the LSL, the complete LSL sequence was subjected to in vivo genetic seletion (in vivo SELEX). In vivo SELEX is performed by randomizing all positions in a particular element, and then subjecting the randomized satC population and helper virus to selection for fitness through multiple rounds of competition in host plants. This selection strategy is possible

because satC enhances the fitness of TCV by positively influencing virus movement through interference with virion formation (Zhang and Simon, 2003a); fewer assembled virions increases the amount of free coat protein available to suppress RNA silencing, an endogenous anti-viral defense system (Qu et al., 2003; Thomas et al., 2003). TCV that is isolated from systemic leaves is therefore more likely to be associated with a functional satRNA. In vivo SELEX is advantageous over site-specific mutagenesis in that selection is for fitness regardless of function, and therefore can reveal additional functional roles of elements (Sun and Simon, 2003; Sun et al., 2005; Zhang and Simon, 2003b). In addition, since all positions in an element (or a portion of an element) are randomized, the necessity for a specific base is not judged in a remaining wild-type context. Moreover, in vivo SELEX allows sequence evolution in multiple rounds of infections.

As described in Materials and Methods, plants were inoculated with TCV and a pool of satC sequences containing completely randomized LSL sequence. For the first round SELEX, the 14 bases of the LSL were randomized by PCR using degenerate oligonucleotides (Figure 3.1, Figure 3.2). SatC transcripts synthesized directly from the PCR products were inoculated onto 60 turnip seedlings along with TCV genomic RNA. At 21 dpi, viable satC species were recovered by RT-PCR from RNA extracted from uninoculated leaves and full-length satC was cloned and sequenced.

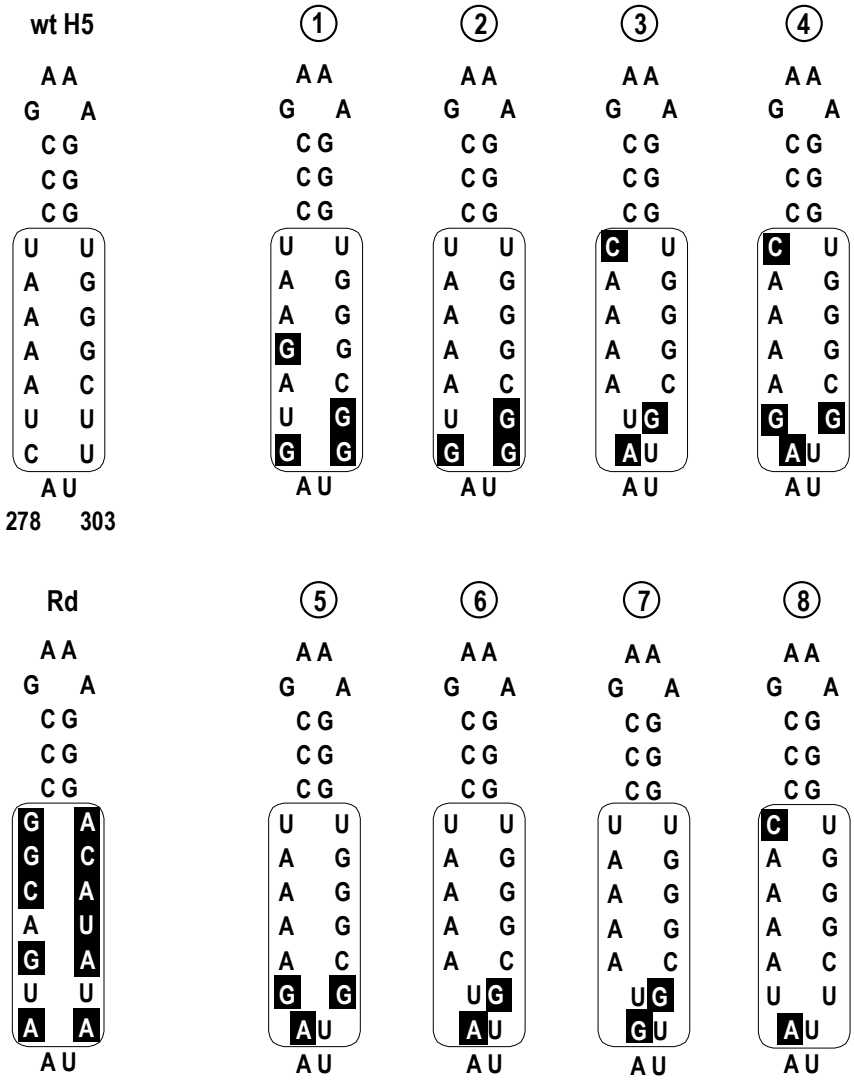
Only 2 of the 60 infected turnip plants displayed satC symptoms (stunting and highly crinkled, dark green leaves) and only those plants contained satC detectable by PCR. Seventeen clones derived from RNA accumulating in the two plants contained five different LSL sequences (Table 3.4, Figure 3.9A). These first round winners displayed strong sequence conservation with the wild-type LSL sequence, especially in the middle

and upper portions of the loop. Eight positions (A281, A283, A284, and 296UGGGC300) were identical to wild-type bases in all clones and A282 was present in 12 of 17 clones. This included the GGGC on the 3' side of the LSL that interacts with the 3' end in vitro (Zhang et al., 2004). Either the wild-type uridylate or a cytidylate was acceptable at position 285. A cytidylate in this position was previously found to be stably maintained in plants and only reduced accumulation of satC by 40% in protoplasts (Figure 3.4A). All clones contained 2 or 3 differences from wild-type at the base of the LSL, the region previously found to be more tolerant to alterations (Figure 3.4A). The guanylate found at position 302 in two clones is also present in the H5 LSL of TCV and was found to enhance satC accumulation in protoplasts by 28% (Figure 3.4A).

For the second round, equal amounts of total RNA isolated from all 60 plants of the previous round were pooled and re-inoculated onto six turnip seedlings, and progeny recovered and sequenced at 21 dpi. Thirty-five of 42 clones recovered were identical to one first round winner (clone 3) (Figure 3.9A). Two additional sequences (clone 6 and 7) had not previously been isolated. All winners contained either G:U, U:G or A:U pairs at the base of the LSL, indicating that canonical base-pairs, while not present in wild-type satC, are acceptable at this location, supporting the conclusions of the site-specific mutagenesis analysis.

Equal amounts of total RNA isolated from second round plants were combined and used to inoculate six more plants. For this third, and final SELEX round, three sequences were recovered, including the first winner from the 2nd round (clone 3). Clone 6, which was a less prevalent 2nd round winner, emerged as the most recovered 3rd round winner. A new sequence was also recovered in the 3rd round (clone 8) that likely

A



Clones	1	2	3	4	5	6	7	8
Round 1	5	4	3	3	2			
Round 2			35			5	2	
Round 3			4			18		10

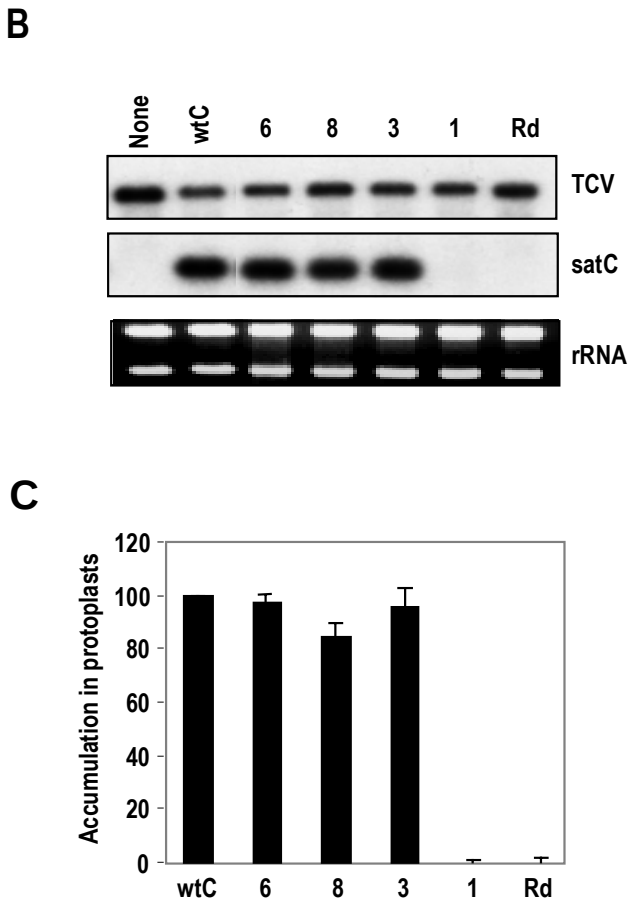


Figure 3.9 In vivo SELEX of the satC H5 LSL. (A) Sequences recovered from three rounds of selection. Sequence designations (numbers) are indicated above each structure. Bases that differ from wt satC in progeny clones 1 to 8 are boxed. Number of clones containing each sequence recovered per round is indicated in the table. (B) Representative RNA gel blot of total RNA extracted from *Arabidopsis* protoplasts at 40 hours after inoculation with TCV and wild-type satC or various SELEX winners. All 3rd round sequences were selected for analysis along with one 1st round winner (sequence 1) and satC containing a randomly selected H5 LSL sequence (Rd). wtC, wild-type satC. Ethidium bromide staining of the gel before blotting shows ribosomal RNA loading control (panel below the blot). (C) Averaged values for accumulation of satC SELEX winners in protoplasts. Values represent results from three independent experiments. Standard deviations are indicated.

TABLE 3.4 Summary of in vivo SELEX of the H5 LSL

	Clone[a]	LSL [b]		Plant[c]						Total
		5' side	3' side	1	2	3	4	5	6	
Round 1	1	GUAGAAU	UGGGCGG	5						5
	2	GUAAAAU	UGGGCGG	4						4
	3	AUAAAAC	UGGGCGU		3					3
	4	AGAAAAC	UGGGCGU		3					3
	5	AGAAAAU	UGGGCGU		2					2
	Total clones assayed from each plant			9	8					17
Round 2	3	AUAAAAC	UGGGCGU	2	7	7	6	9	4	35
	6	AUAAAAU	UGGGCGU		1	1			3	5
	7	GUAAAAU	UGGGCGU					2		2
	Total clones assayed from each plant			2	8	8	6	11	7	42
Round 3	6	AUAAAAU	UGGGCGU	1	2	4	4	4	3	18
	8	AUAAAAC	UGGGCUU	1	2		2	3	2	10
	3	AUAAAAC	UGGGCGU	1	2		1			4
	Total clones assayed from each plant			3	6	4	7	7	5	32

[a] From Figure 3.9A.

[b] Sequence is presented 5' to 3'. Bases that differ from wild-type satC LSL are underlined.

[c] The number of clones found in each plant at 21 dpi.

evolved from clone 3 by a single change from a guanylate to the wild-type uridylate at position 302.

Since the number of satC with a particular sequence recovered in later rounds of in vivo SELEX can be influenced by the starting concentration of each sequence in plants of the previous round, the 3rd round winners were subjected to side-by-side competition for fitness starting with equal amounts of individual transcripts. Also included in this competition was a randomly selected 1st round winner (clone 1). At 21 dpi, satC species were cloned and sequenced from three infected plants. Twenty-nine of 31 clones were identical to clone 6, and the remaining two clones were clone 8 (Table 3.5). The fitness order of the 3rd round winners was therefore: clone 6, clone 8 and then clone 3.

Clone 6 differed from wild-type satC at two positions (C279A and U301G). While the U301G alteration was not tested in the site-specific mutagenesis analysis, satC containing the C279A alteration accumulated 10% better than wild-type satC in protoplasts (Figure 3.4). To determine the fitness of clone 6 compared with wild-type satC in plants, equal amounts of wild-type and mutant transcripts were combined and inoculated onto three plants along with TCV. Twenty-three of the 26 recovered clones were wild-type satC (Table 3.5) indicating that one or both of the base differences in clone 6 reduced fitness of satC in plants.

Fitness of satC variants to accumulate in plants depends on several factors including replication competence, stability, and ability to enhance the movement of TCV (Sun and Simon, 2003; Zhang and Simon 2003b). To determine if the fitness of the in vivo SELEX winners to accumulate in plants correlated with their ability to accumulate in protoplasts, the three 3rd round winners, together with 1st round clone 1 and satC

TABLE 3.5 Competition between 3rd round SELEX winners and wild-type satC

	Clone[a]	Plant[b]			Total
		1	2	3	
Competition 1	6	7	8	14	29
	8	1		1	2
	3				
	1				
Competition 2	Wild-type satC	6	5	12	23
	6	1	2		3

[a] From Figure 3.9A.
[b] The number of clones found in each plant at 21 dpi.

containing a randomly selected sequence from the original randomized population (Rd), were separately inoculated onto protoplasts along with TCV. By assaying accumulation in protoplasts, this allowed for a determination of replication/stability in the absence of movement. All three 3rd round winners accumulated to near wild-type levels in protoplasts (Figure 3.9B and C), while the 1st round winner (clone 1) and the randomly selected satC (Rd) did not accumulate to detectable levels. Since we have never identified mutations in satC that affect the stability of the satRNA, these results suggest that the ability of the SELEX winners to replicate was a primary factor for fitness in plants. However, while all 3rd round winners accumulated to near wild-type levels in protoplasts, clones 3 and 8 were substantially less fit than clone 6 in the competition assay. Furthermore, wild-type satC was considerably more fit in plants than clone 6 yet replication levels in protoplasts were very similar. This suggests that fitness of in vivo SELEX winners with different LSL sequences in plants depends on factors in addition to those that allow for robust replication in protoplasts.

Examination of possible interactions between the H5 LSL and other regions

Based on in vitro studies, phylogenetic comparisons, second site alterations associated with H5 mutations and comparable studies in tombusviruses (Pogany et al., 2003), the carmoviral H5 LSL likely participates in replication by binding to and the 3' end. If correct, an important question is how the 3' end is presented to the RdRp so that it can initiate transcription de novo opposite the terminal cytidylate. At this point in our understanding, a repression-derepression model was proposed, which suggested that an additional sequence, termed as "derepressor", serves to derepress (-)-strand synthesis by

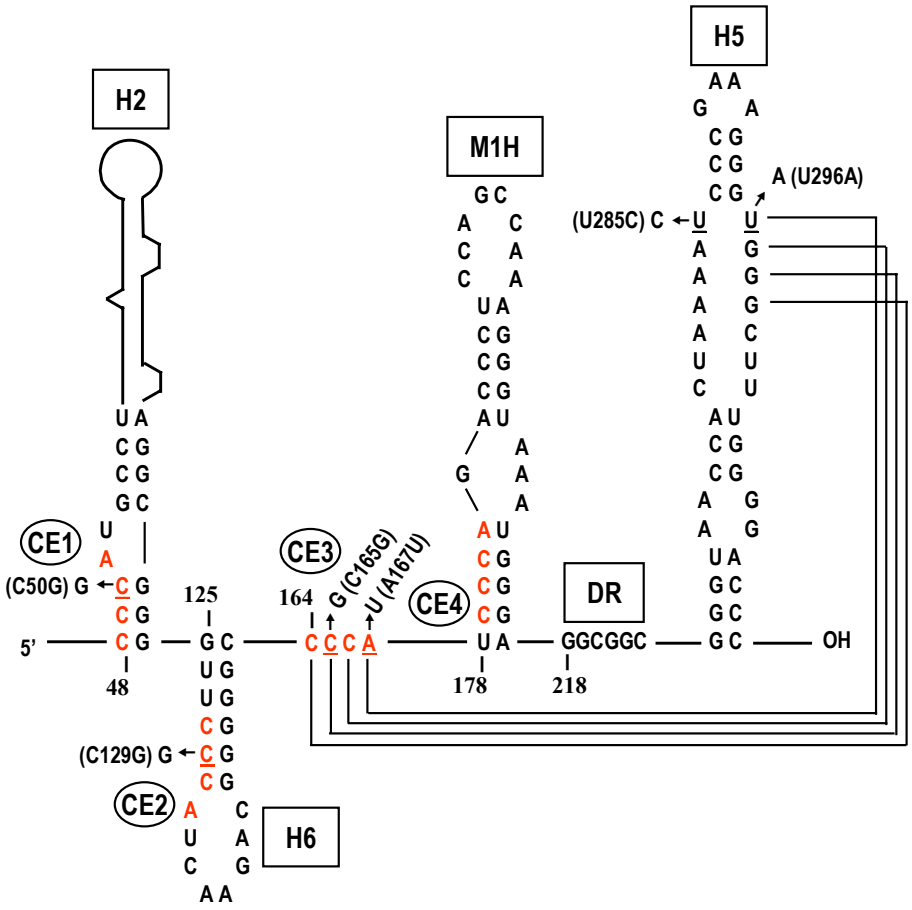
releasing the 3' end from interaction with the LSL. Such a derepressor could be one or more sequences that disrupt LSL/3' end base-pairing by interacting with either the LSL, 3' terminal sequence, or both.

Mfold computational analysis, which is not designed to detect tertiary structure such as the interaction between the 3' end and the LSL of H5, predicts that the UGCCC-OH at the 3' end of satC base-pairs with positions 218-221 (5' CGGCGG, the DR) (See Figure 2.2). The DR is highly sequence-specific as established by in vivo SELEX (Sun et al., 2004) and mutations in this region seriously reduced (-)-strand synthesis in vitro (Zhang et al., 2004, see discussion). As shown in Figure 2.5, two single site mutations had been individually generated in satC at positions 218 (G218C) and 220 (C220G) and tested in protoplasts. Accumulation of G218C was 3.6-fold less than wild-type satC, supporting a role for position 218 in satC replication (Figure 2.5C). However, the alteration in C220G had a more modest effect, with C220G accumulating to 80% of wild-type satC levels (Figure 2.5B, Figure 2.8B). As shown in Figure 3.10A, position G353 could potentially base pair with both position C300 in H5 LSL and C220 in DR. To further investigate possible interactions between the 3' end, DR and H5 LSL, compensatory mutations that combined C220G with either G353C (G353C/C220G) or C300G/G353C (G353C/C300G/C220G) were generated and tested in protoplasts (Figure 3.10B). Neither of these two constructs accumulated to detectable levels, which also suggested that the identity of positions 300 or 353 is important for some additional function besides base-pairing with each other. Sequences that could potentially compete with the 3' terminus for the satC H5 LSL were also examined. As described earlier, mutational analysis did not support an RNA-RNA interaction between the satC H5 LSL

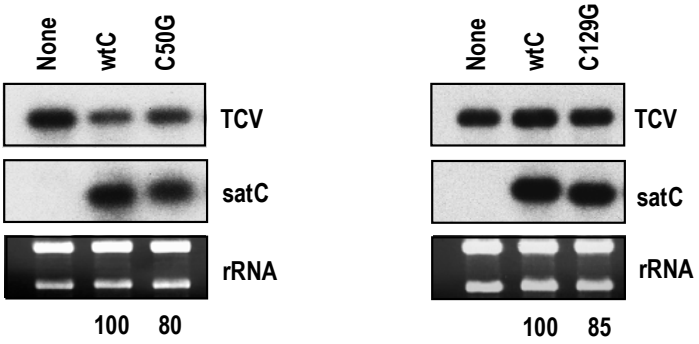
Figure 3.10 Compensatory exchanges between the satC LSL, a nucleotide near the 3' end and a nucleotide within the DR. (A) Location of the point mutations generated in satC. Bases that were altered are underlined and arrows point to the new bases. Names of the altered constructs are shown. Possible base-pairing between the 3' terminal bases and the 3' side of the LSL is shown. (B) Northern blot of mutant satC and helper virus (TCV) plus-strands. Total RNA was extracted from protoplasts at 40 hpi and probed with an oligonucleotide complementary to both TCV and satC. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot). None, no satC added. wtC, wild-type satC.

and H4a loop (Figure 3.7). Second site changes have been found in TCV clones recovered from plants that were inoculated with a TCV variant in which H5 was replaced by JINRV H5 (J.C. McCormack and A.E. Simon, unpublished data). All these second site alterations occurred in 5'CCCA elements located within p28 and CP coding regions. 5'CCCA could potentially base pair with the 5'UGGG at the 3' side of the H5 LSL. These observations suggest that there might be a possible interaction between a 5'CCCA element and H5. SatC contains four 5'CCCA sequences all located within the region derived from satD (labeled as CE1 to CE4 from the 5' end, Figure 3.11A). The most 3' proximal sequence (CE4) is located at the base of M1H. Previous studies showed that the most fit winner from an in vivo SELEX assay of positions 181-208**,** which contains C181A and A182G alterations along with other changes in this region, accumulated to 89% of wild-type levels in protoplasts. This suggests that CE4 may not be important for satC replication. CE1 and CE2 are located within H2 and H6, respectively, with the adenylate presented in either the internal loop or terminal loop. Single site mutations in CE1 (C50G) and CE2 (C129G) did not significantly reduce satC accumulation in protoplasts (Figure 3.11B, C), suggesting that sequence identity in these two elements is not important for satC replication. CE3, located in a linear region flanking the 5' side of satC M1H, is highly conserved and important for satC replication by its ability to repress virion accumulation (Sun et al., 2005). Consistently, single site mutations in CE3 reduced satC accumulation to 34% (C165G) and 66% (A167U), respectively (Figure 3.11C). Combining the two mutations (C165G/A167U) further reduced satC accumulation to 27% of wild-type. The effect on satC accumulation of combining mutations in CE4 and DR (G218C) was also examined (Figure 3.11C). While A167U/G218C accumulated to

A


H2
M1H
H5
A (U296A)
(U285C) C
CE1
(C50G) G
CE3
G (C165G)
U (A167U)
CE4
DR
5'
OH
125
164
48
178
218
(C129G) G
CE2
H6

B


None
wtC
C50G
TCV
satC
rRNA
100
80
None
wtC
C129G
TCV
satC
rRNA
100
85

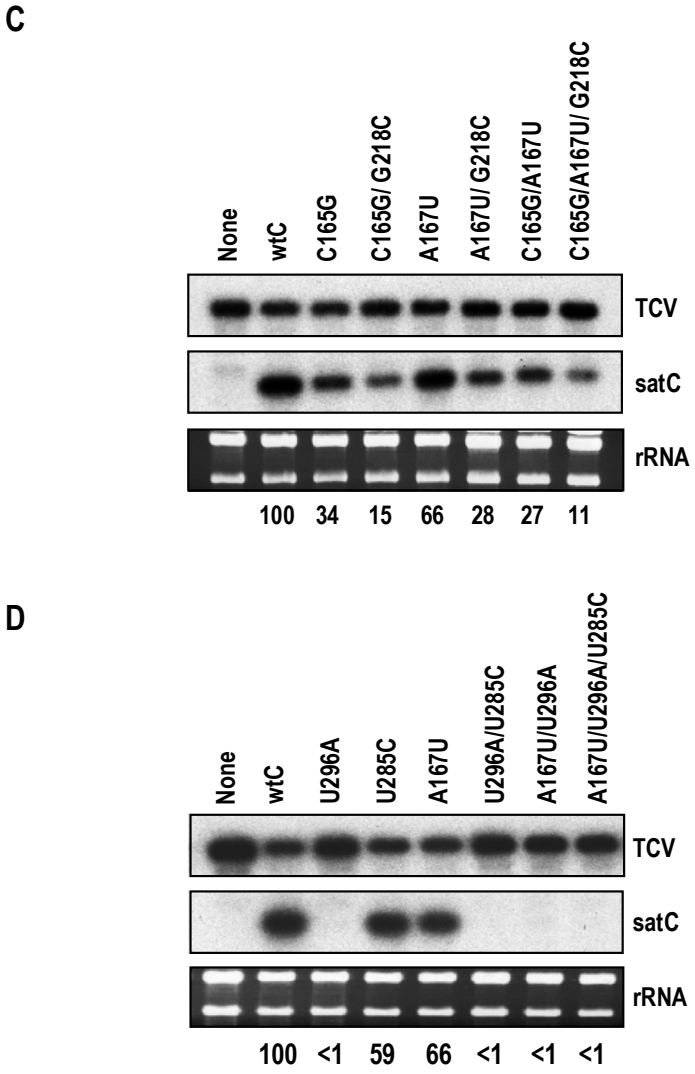


Figure 3.11 Examination of possible interactions between satC 5'CCCA elements and H5 LSL. (A) Location of the point mutations generated in satC. 5'CCCA elements are denoted in red and labeled as CE1 to CE4 from the 5' end. A possible interaction between 5'CCCA at positions 164 to 167 and H5 LSL is shown. Bases that were altered are underlined and arrows point to the new bases. Names of the altered constructs are shown nearby. (B) (C) (D) Northern blots of mutant satC and helper virus (TCV) (+)-strands. Total RNA was extracted from protoplasts at 40 hpi and probed with an oligonucleotide complementary to both TCV and satC. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot). None, no satC added. wtC, wild-type satC.

levels similar to G218C (28%), C165G/G218C and C165G/A167U/G218C further reduced satC accumulation to 15 and 11%, respectively. These results confirm that CE3 is involved in satC accumulation in protoplasts but do not suggest that this role requires an interaction with the DR.

To determine whether there is an interaction between CE3 and the H5 LSL, compensatory mutations between positions A167 and U296 were introduced into these two elements (Figure 3.11A). A uridylate to adenylate change at position 296 may cause base-pairing between U296A and U285. To avoid this possibility, a uridylate to cytidylate change was also introduced at position 285. As shown in Figure 3.4A, U285C accumulated to 59% of wild-type. U296A and U296A/U285A accumulated to barely detectable levels in protoplasts. Compensatory mutations (A167U/U296A and A167U/U296A/U285C) did not restore satC accumulation, suggesting that CE3 may not interact with H5 LSL directly or that position U296 is involved in some other crucial function.

Discussion

Negative-regulation of (-)-strand synthesis in vitro by a cis-acting element that interacts with the 3' terminal nucleotides has been identified for TCV (Zhang et al., 2004), TBSV (Pogany et al., 2003) and predicted for BYDV (Koev et al., 2002). In TCV and satC, the four 3' end terminal bases can form Watson-Crick base-pairs with the 3' side of the H5 LSL, thus presumably sequestering the 3' end from the RdRp (Zhang et al., 2004).

To explore the sequence requirements of the LSL, I performed both single site mutational analyses and in vivo SELEX. The results of these two approaches were very consistent. On the 5' side of the LSL, there was strong conservation of the four consecutive adenylates and for a pyrimidine at position 285. The residues opposite these positions on the 3' side of the LSL were conserved in all recovered SELEX clones and alterations were either poorly or not conserved in satC accumulating in plants. The importance of these 3' side residues supports our model that the 3' terminal satC sequence (GCCC-OH) base pairs with sequence in the 3' side of the LSL. The lower positions of the LSL were considerably more flexible, with both canonical (A:U, G:U and C:G) and non-canonical (G•G) pairings able to replace the wild-type C•U in the lowest position and U:G and G•G pairs able to replace the U•U pairing in the penultimate position. Interestingly, TCV contains a C:G at the base of its LSL while the related carmovirus CCFV has several additional canonical base-pairings in the same region of its LSL.

Since wild-type satC was more fit than the best SELEX winner (clone 6), a question arises as to why clones containing wild-type satC sequence were not recovered from the in vivo SELEX. The initial randomized population of cDNA (4×10^{12} cDNA fragments used for in vitro transcription per plant) was more than sufficient to contain all possible sequences in the 14 positions. However, the actual number of RNAs that enter cells and initiate infection is not known. This would suggest that the vast majority of satRNA sequences inoculated onto plants would not be part of the initial selection, reducing the pool of possible sequence combinations below that required to assay every sequence. In a previous in vivo SELEX of a 14 base satC motif that was determined to

have only 12 specific bases, wild-type sequence was recovered in the 1st round (Guan et al., 2000b). In vivo SELEX thus provides information on what sequences can support accumulation but not always on the identity of the best possible sequence.

Results from in vivo SELEX suggest that H5 has a role in satC fitness in addition to permitting robust replication. All 3rd round winners replicated to near wild-type levels, while fitness in plants varied substantially. A recently completed in vivo SELEX of the M1H, a replication enhancer in its minus-sense orientation, also indicated that sequences were recovered with a role in addition to replication (Zhang and Simon, 2003b). While replication-related motifs were recovered in (-)-sense RNA, a sequence non-specific hairpin in (+)-strands strongly influenced satC fitness by bringing together flanking sequences that interfered with virion formation, which enhanced the ability of TCV to suppress RNA silencing. What additional satRNA property might be influenced by H5 is not known. In a recent investigation of the H5 LSL of TCV, mutations in many positions were also not stable. Surprisingly, 25% of mutant TCV progeny had one to three second site alterations scattered throughout 200 to 300 bases in the 3' UTR in addition to a reversion of the original alterations. Interestingly, the second site changes were strongly biased towards uridylate to cytidylate and adenylate to guanylate transitions (McCormack and Simon, 2004). These results led to the speculation that the TCV H5 also functions as a chaperone required for proper assembly of the RdRp.

The second site changes found in progeny of some satC LSL mutants did not display the same bias or random positioning as the TCV mutants, and some could be hypothetically connected to compensating for the original alterations. The benefit of the G230A second site change found in the loop of H4a in progeny of two LSL mutants,

however, is not understood. While my results do not support possible interactions between H5 and H4a (Figure 3.8), it remains possible that a less obvious relationship exist between H4a and either H5 or sequences that interact with H5.

Other possible interactions between the H5 LSL and other regions in satC were also examined by compensatory mutational analysis. While mutations in the 3' terminus, DR and CE3 significantly reduced satC accumulation in protoplasts, none of the compensatory mutations that re-established potential base-pairing between the 3' side of H5 LSL and either the 3' terminus or CE3 restored satC accumulation. Double compensatory mutations that re-established potential base-pairing between the 3' side of the H5 LSL and the 3' terminus along with the 3' terminus and the DR did not restore satC accumulation either. However, as described above, in vitro RdRp assays confirmed the interaction between the 3' side of the H5 LSL and 3' terminus (Zhang et al., 2004). In addition, results from in vitro RdRp assays of constructs G218C and C220G in the presence or absence of the 3' terminal CCC also suggested that the DR plays a role in activating (-)-strand synthesis (Zhang et al., 2004). Deletion of the 3' terminal CCC, which released the 3' terminus, enhanced synthesis of full-length products in vitro by 3.5-fold and total products by 8-fold over transcription of wild-type satC (Zhang et al., 2004). In contrast, transcripts of C220G and G218C showed an 11- (C220G) or 30-fold (G218C) suppression of full-length (-)-strand synthesis. Deletion of the 3' terminal CCC in combination with G218C or C220G mutations resulted in enhancement of both full-length and aberrant transcription (Zhang et al., 2004). Currently, it is not clear why C220G had only a modest effect on accumulation of satC in vivo. C165G/G218C and C165G/A167U/G218C accumulated to lower levels (15 and 11%, respectively) than

C165C (34%), C165G/A167U (27%) and G218C (28%) suggesting that the effect of mutations in CE3 and DR on satC accumulation are additive. Whether CE3 also plays a role in activating (-)-strand synthesis needs to be further explored.

The similar cis-acting element in TBSV (SL3) contains an internal loop with both similarities and striking differences with the carmoviral H5. Whereas all carmoviral H5 large internal loops are symmetrical or nearly symmetrical, the TBSV SL3 is asymmetrical, with only a single adenylate occupying the 5' side (Pogany et al., 2003). The 3' side of the SL3 and H5 internal loops are similar, with five of eight SL3 bases (GGGCU) identical to their carmoviral counterparts. The requirement for symmetry (or near symmetry) in the H5 internal loop of carmoviruses is not known. The LSL is uncommonly large for an interior symmetrical loop. Other examples of large symmetrical internal loops include loop E of 5S rRNA (14 bases, Specht et al., 1990), the sarcin/ricin loop of 23S rRNA (10 bases; Leontis et al., 2002a), internal loops within ADAR substrates (12 bases, Lehmann and Bass, 1999), and the internal loop in group I introns (10 bases, Cech et al., 1994). Biochemical analyses of the similarly sized loop E suggest that the loop adopts a lightly overwound but still roughly helical structure that is required for ribosomal protein L25 binding (Tang and Draper, 1994). Internal loops play important structural and functional roles in RNA, providing flexibility and allowing RNAs to form more compact structures (Jaeger et al., 1993; Zacharias and Hagerman, 1996). Why carmoviruses appear to require symmetry in the internal loops of their 3' end-binding hairpins while tombusviruses do not will only be answered when detailed three-dimensional structures are available for both hairpins.

CHAPTER IV

IMPORTANCE OF SEQUENCE AND STRUCTURAL ELEMENTS WITHIN A VIRAL REPLICATION ELEMENT

Introduction

Replication of all positive (+)-strand RNA viruses requires a multi-step process that begins with reiterative copying of the infecting genome to generate complementary (-)-sense intermediates, followed by reiterative copying of the intermediates to generate progeny (+)-strand RNAs. This process requires that the viral-encoded RdRp locate the 3' end of its cognate RNA for de novo or primer-dependent initiation of RNA synthesis (Buck, 1996; Kao et al., 2001; van Dijk et al., 2004). Promoter elements that specifically interact with the polymerase for initiation of (-)-strand synthesis are generally located proximal to 3' terminal sequences and usually comprise one or more hairpins with adjoining single-stranded sequence (Dreher, 1999; Duggal et al., 1994). Core promoters can contain multiple sequence and structural features necessary for efficient RdRp recognition (Kim et al., 2000). The ability of some viral RNAs to replicate in the absence of large 3' or 5' terminal fragments (Todd et al., 1997; Wu and White, 1998) also suggests that for some virus-host combinations, promoter sequences may be redundant or additional factors such as close proximity between polymerase and template in replication organelles can obviate the need for distinct molecular recognition (Brown et al., 2004).

While core promoters permit basal levels of RNA transcription, efficient RNA synthesis requires additional viral elements such as structures and sequences at the 5' ends that may be required for genome circularization (Frolov et al., 2001; Herold and Andino, 2001; Isken et al., 2003; Khromykh et al., 2001; You et al., 2001) and internal elements such as repressors, enhancers, and chaperones, which function either in cis (Barton et al., 2001; Herold and Andino, 2001; Khromykh et al., 2001; Klovins et al. 1998; Murray et al., 2003; Nagy et al. 1999, 2001; Panavas and Nagy, 2003; Pogany et al., 2003; Quadt et al., 1995; Ray and White, 1999, 2003; Vlot et al., 2001; You et al., 2001; Zhang et al., 2004) or in trans (Eckerle and Ball, 2002; Sit et al., 1998). Enhancers are generally found on viral (-)-strands, need not be proximal to the core promoter, contain sequence and/or structural features of core promoters, and can promote transcription in the presence of sequences resembling the transcription initiation site (Nagy et al., 1999; Panavas and Nagy, 2003; Ray and White, 2003). Repressors (also known as transcriptional silencers) have recently been identified for members of the family *Tombusviridae* and are located on (+)-strands just upstream from the core promoter. The positioning of transcriptional enhancing and repressing elements on opposite strands has led to the suggestion that these elements function to regulate asymmetric levels of (+)- and (-)-strand synthesis (Pogany et al., 2003).

As described in Chapter I, the 3’ regions of all carmovirus genomic RNAs contain four hairpins (H4a, H4b, H5, Pr hairpin). H5 contains a large internal symmetrical loop (LSL) that is A-rich on the 5' side and G-rich on the 3' side (Zhang et al., 2004; Figure 1.4, 1.5). Synthesis of wild-type levels of satC complementary strands by the RdRp in vitro requires an interaction between the 3' end and the H5 LSL in the template RNA as

assayed by a compensatory mutagenesis approach (Zhang et al., 2004). The upstream 5' sequence, DR, was proposed to have a role in activating (-)-strand synthesis (Zhang et al., 2004). Surprisingly, transcripts containing mutations in the TCV H5 LSL led to the generation of progeny in plants with as much as a 12-fold increase in second site mutations scattered throughout the sequenced region, with most alterations consisting of uridylate to cytidylate or adenylate to guanylate transitions (McCormack and Simon, 2004). This led to the proposal that H5 also functions as a molecular chaperone to aid in correctly assembling the RdRp complex. As described in the previous chapter, members of the family *Tombusviridae* genus *Tombusvirus* also contain a similarly positioned hairpin (but with an internal asymmetrical loop) that is required in vivo (Fabian et al., 2003) and represses (-)-strand synthesis in cell-free assays by interacting with five 3' terminal bases (Pogany et al., 2003).

Analysis of the LSL of satC H5 by in vivo SELEX revealed that nearly all positions in the middle and upper portions of the LSL were crucial for satC accumulation in protoplasts (Chapter III). In this chapter, I extend my analysis of H5 by examining sequence and structural preferences in the upper and lower stems using in vivo SELEX and exchanges with other carmoviral H5. My results indicate that both sequence and structure are important in these H5 regions, and confirm that H5 is a (+)-strand element. I also demonstrate that enhanced fitness of some SELEX winners to accumulate in plants did not correlate with enhanced ability to replicate in protoplasts, suggesting that H5 might have additional, non-replication related functions. Finally, I demonstrate that mutations in H5 strongly affect (+)-strand as well as (-)-strand accumulation, supporting a multifunctional role for this hairpin.

Materials and Methods

Construction of satC mutants

Plasmids $CH5_{JINRV}$ (all constructs used in this chapter are listed in Table 4.1) and $CH5_{CCFV}$ were constructed by PCR using pT7C+ as template and T7C5' (primers used in this chapter are shown in Table 4.2) and JINRVH or CCFVH as primers, respectively. Following digestion with *Eco*RV and *Nco*I, the fragment was inserted into the analogous location in plasmid $satC_E$ (Table 2.1), which had been treated with the same restriction enzymes. LSL/LS and US/LS were generated in a similar fashion except that $CH5_{CCFV}$ was used as template and CCFV-UC and CCFV-IC were used with T7C5' as primers, respectively. For construction of plasmids US/LSL and LSL, oligonucleotide CCFV-LC5 and CCFV-LC3 were used for PCR. The templates were plasmids $CH5_{CCFV}$ and LSL/LS, respectively. PCR products were subsequently digested with *Spe*I and *Eco*RV and cloned into $satC_E$ replacing the analogous wild-type fragment. To generate plasmid US, PCR was performed with primers C-5Uf and CEcV and construct US/LS. PCR products were digested with *Spe*I and *Eco*RV and cloned into $satC_E$ that had been similarly treated. Plasmid LS was generated in a similar fashion except C-5Lf and C-3Lf were used as primers and pT7C+ was used as template. For construction of plasmid $CH5_{TCV}$, oligonucleotides T7C5' and T5PC were used as primers with template pT7C+. PCR products were treated with *Spe*I and T4 polynucleotide kinase and inserted into the analogous location in pT7C+ that had been treated with *Spe*I and *Sma*I. Plasmids G304C and C277G were generated in a similar fashion except that oligonucleotides T7C5' and G304C or G277C and Oligo 7 were used as primers, respectively. Plasmid

TABLE 4.1 Summary of satC and TCV mutants used in Chapter IV

Name	Description
$CH5_{JINRV}$	$satC_E$ with H5 (positions 270 to 311) converted to H5 from JINRV
$CH5_{CCFV}$	$satC_E$ with H5 (positions 270 to 311) converted to H5 from CCFV
US	$satC_E$ with H5 upper stem loop (US, positions 286 to 295) converted to that from CCFV
LSL	$satC_E$ with H5 large internal symmetrical loop (LSL, positions 279 to 285 and 296 to 302) converted to that from CCFV (including two base-pairs directly below the LSL)
LS	$satC_E$ with H5 lower stem (LS, positions 270 to 278 and 303 to 311) converted to that from CCFV
LSL/LS	$satC_E$ with H5 LSL and LS (positions 270 to 285 and 296 to 311) converted to that from CCFV
US/LS	$satC_E$ with H5 US and LS (positions 270 to 278, 303 to 311 and 286 to 295) converted to that from CCFV
US/LSL	$satC_E$ with H5 US and LSL (positions 279 to 302) converted to that from CCFV
G304C	SatC with a G to C change at position 304
C277G	SatC with a C to G change at position 277
C277G/G304C	SatC with a C to G change at position 277 and a G to C change at position 304
G304U	SatC with a G to U change at position 304 (Guan et al., 2000)
C277G/G304U	SatC with a C to G change at position 277 and a G to U change at position 304
G304A	SatC with a G to A change at position 304 (Guan et al., 2000)
$CH5_{TCV}$	SatC with a U to G change at position 302 and a G to U change at position 306, thereby convert satC H5 to TCV H5
$CH5_{TCV}$G304A	$CH5_{TCV}$ with a G to A change at position 304
$CH5_{TCV}$G304U	$CH5_{TCV}$ with a G to U change at position 304
G3998C	TCV with a G to C change at position 3998
G3998U	TCV with a G to U change at position 3998
G3998A	TCV with a G to A change at position 3998

TABLE 4.2 Summary of the oligonucleotides used in Chapter IV

Application/ construct	Name	Position[a]	Sequence[b]	Polarity[c]
Mutagenesis in satC H5	T7C5'	1-19	5'-*GTAATACGACTCACTATA*GGGAUAACUAAGGGTTTCA	+
	C5'	1-19	5'-GGGAUAACUAAGGGTTTCA	+
	CEcV	300-356	5'-GGGCAGGCCCCCCGTCCGAGGAGGGAGGCTATCTAT TG (G**A**T**A**TC) GGAGGGTCCCCAAAG	-
	JINRVH	252-322	5'-ATTG (G**A**T**A**TC) GGATCCCTAGCAAGCCCACCCTCAC GGGATTTTATGCATCGGGAAGAGAGCACTAGTTTTCC	-
	CCFVH	252-322	5'-ATTG (G**A**T**A**TC) GGAGGTCCCCAACACCCACTCCGAA GAGATTTCGTTGGTTACCAGAGAGCACTAGTTTTCC	-
	CCFV-UC	271-322	5'-ATTG (G**A**T**A**TC) GGAGGTCCCCAACACCCACCCTTTC GGGATTTCGTTGGTTACC	-
	CCFV-IC	264-322	5'-ATTG (G**A**T**A**TC) GGAGGTCCCCAAAGCCCACTCCGAA GAGATTTTAGTGGTTACCAGAGAG	-
	CCFV-LC5	257-281	5'-aACTAGTGCTCTCTGGGTAACCAACG	+
	CCFV-LC3	298-322	5'-ATTG (G**A**T**A**TC) GGAGGGTCCCCAACACC	-
	C-5Lf	257-279	5'-gACTAGTGCTCTCTGGTAACCAC	+
	C-3Lf	299-322	5'-ATTG (G**A**T**A**TC) GGAGGTCCCCAAAGC	-
	C-5Uf	257-279	5'-gACTAGTGCTCTCTGGGTAACCAC	+
	T5PC	289-356	5'-GGGCAGGCCCCCCGTCCGAGGAGGGAGGCTATCTA TTGGTTCGGAGGGTCACCACAGCCCACCCTTTC	-
	T5PCm	289-356	5'-GGGCAGGCCCCCCGTCCGAGGAGGGAGGCTATCTA TTGGTTCGGAGGGTCACNACAGCCCACCCTTTC	-
	C277G	257-289	5'-gcacACTAGTGCTCTCTGGGTAACGACTAAAATCCCG	+
	G304C	288-356	5'-GGGCAGGCCCCCCGTCCGAGGAGGGAGGCTATCTAT TGGTTCGGAGGGTCCCGAAAGCCCACCCTTTCG	-
	Oligo 7	338-356	5'-GGGCAGGCCCCCCGTCCGA	-
Mutagenesis in TCV H5	SEQ1	3947-4009	5'-GAAAACTAGTGCTCTTTGGGTAACCACTAAAATCCCG AAAGGGTGGGCTGTHGTGACCCTCCG	+
	KK57	4036-4054	5'-GGGCAGGCCCCCCCCCCGC	-
SELEX	3CTS	266-356	5'-GGGCAGGCCCCCCGTCCGAGGAGGGAGGCTATCTAT TGGTTCGGAGGGTCCCCAAAGCCCANNNNNNNNNNAT TTTAGTGGTTACCCAGAG	-
	CLoS	252-356	5'-GGGCAGGCCCCCCGTCCGAGGAGGGAGGCTATCTAT TGGTTCGGANNNNNNNNNAAGCCCACCCTTTCGGGAT TTTAGNNNNNNNNNAGAGAGCACTAGTTTTCC	-
RNA gel blots	Oligo 13	249-269	5'-AGAGAGCACTAGTTTTCCAGG[d]	-
	T7-229(-)	217-229	5'-*GTAATACGACTCACTATA*GGTGCTGCCGCCG	-

[a] Coordinates correspond to those of the TCV genome (SEQ1 and KK57) or satC (all the other oligonucleotides).
[b] Bases in italics indicate T7 RNA polymerase promoter sequence. Bold residues denote bases inserted to generate an *Eco*RV site (in parentheses). Bases in lowercase were added to achieve efficient digestion. "N" represents randomized base. "H" represents mixed base A, C and T. Mutant bases are underlined.
[c] "+" and "-" polarities refer to homology and complementarity with satC positive strands, respectively.
[d] Oligo 13 is also complementary to positions 3950 to 3970 of TCV genomic RNA.

C277G/G304U was generated in a similar fashion except that pT7M3 (Guan et al., 2000b) was used as template. Plasmid C277G/G304C was also generated similarly with oligonucleotides C277G and G304C as primers. Plasmids $CH5_{TCV}$G304A and $CH5_{TCV}$G304U were generated by PCR using oligonucleotides T7C5' and T5PCm with pT7C+ as template. PCR products were cloned into the *Sma*I of pUC19. Mutants were identified by sequencing.

Construction of TCV mutants

Plasmids G3998C, G3998U and G3998A were generated by PCR using plasmid pT7TCVms as template (Figure 4.1). pT7TCVms contains wild-type TCV cDNA downstream from a T7 RNA polymerase promoter. Oligonucleotides SEQ1 and KK57 were used as primers. Following treatment with T4 polynucleotide kinase and *Spe*I, the fragment was cloned into the analogous location in pT7TCVms, which had been treated with *Spe*I and *Sma*I. Mutants were identified by sequencing.

In vitro transcription, inoculation of Arabidopsis protoplasts and RNA gel blots

TCV genomic RNA was synthesized using T7 RNA polymerase and plasmids previously digested with *Sma*I, which generates transcripts with precise 5' and 3' ends. SatC transcripts were synthesized from plasmids linearized with *Sma*I (for chimeric H5 constructs) or directly from PCR products (for SELEX winners, using primers T7C5' and Oligo 7). Protoplast preparation, inoculation, and RNA gel blots were performed as described in Chapter II. For analysis of (-)-strands, RNA was probed with an [α-^{32}P]-

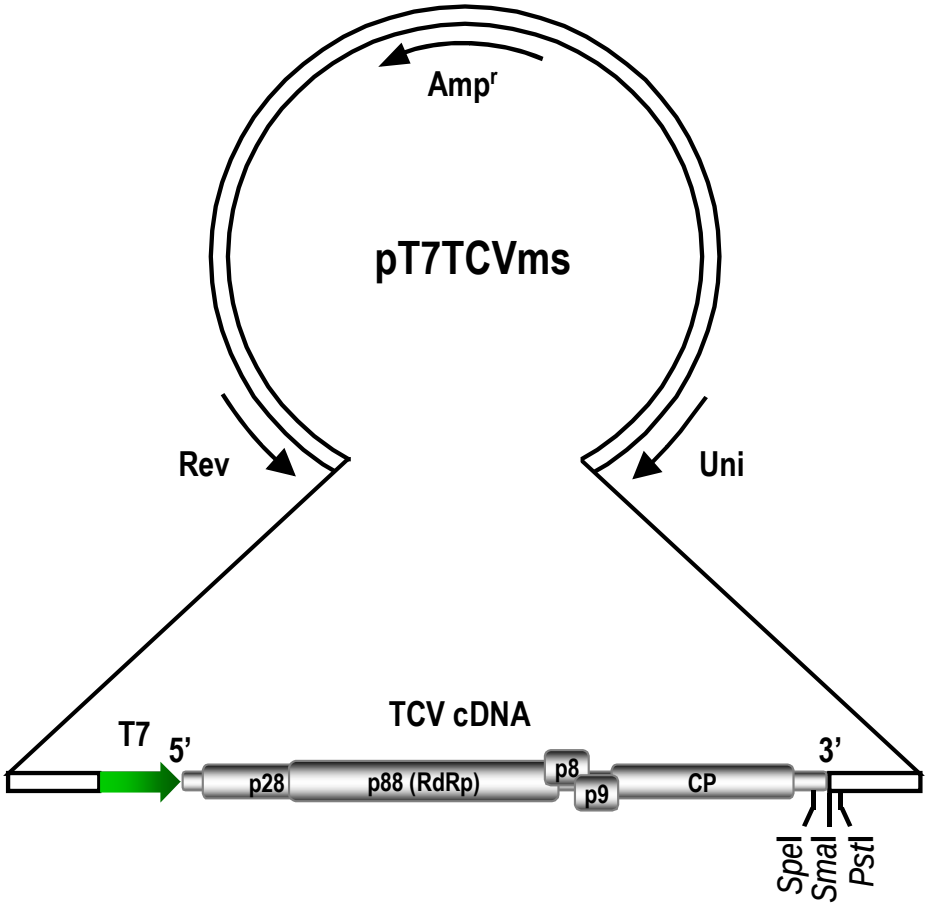


Figure 4.1 Map of pT7TCVms. The open bar represent the pT719E(+)-backbone sequence. The 5' and 3' ends of TCVms positive-strand sequence are shown. The green arrow represents the T7 RNA polymerase promoter sequence. The restriction endonuclease sites used for cloning in this study are shown. (Modified from Oh et al., 1995).

UTP-labeled riboprobe obtained from *Dra*I-digested pT7+ following transcription with T7 RNA polymerase (Nagy et al., 1999).

For RNA stability assay, protoplasts (5×10^6) were inoculated with 5 μg of each satC RNA transcripts. Total RNAs were extracted at 0, 1, 2, 4 and 6 hpi. RNA was probed with an [α-^{32}P]-UTP-labeled riboprobe obtained from PCR using primers T7-299(-) and C5' and template pT7C+ following transcription with T7 RNA polymerase.

In vivo SELEX

In vivo SELEX was performed as described in Chapter III. Full-length satC cDNAs containing randomized bases in the H5 LS and US were generated by PCR using pNco-C277 (Table 3.1) as template. Primers used were T7C5' and either 3CTS or CLoS, respectively. PCR products were purified and directly subjected to in vitro transcription using T7 RNA polymerase. The number of cDNA molecules used for in vitro transcription of RNA to infect one plant was 4×10^{12}.

For the 1st round SELEX, 5 μg of satC transcripts containing randomized LS or US sequences were inoculated onto each of 60 (LS SELEX) or 30 (US SELEX) turnip seedlings along with 4 μg of TCV transcripts. Total RNA was extracted from uninoculated leaves at 21 dpi. Viable satC species were recovered by RT-PCR using oligonucleotides C5' and Oligo 7, cloned into the *Sma*I site of pUC19 or pGEM-T and sequenced. The 2nd and 3rd round SELEX were performed as described in Chapter III.

In planta competition of SELEX winners

Equal amounts of transcripts were combined and used to inoculate three turnip seedlings (0.4 μg/plant) along with TCV genomic RNA transcripts (4 μg each/plant). SatC species from all plants at 21 dpi were cloned and assayed as described in Chapter III.

Results

The LSL and lower stem contribute significantly to H5 function in satC

The importance of H5 in (-)-strand accumulation (Zhang et al., 2004) and RdRp fidelity (McCormack and Simon, 2004) led to the initiation of a detailed characterization of satC H5 sequence and structure. My initial characterization used site-specific mutagenesis and in vivo SELEX to examine the LSL (Chapter III). I determined that satC with various alterations in bases located in the lower two positions on either side of the LSL, including those that caused the bases to pair with their neighbors across the loop, could still accumulate to near wild-type levels or exceed wild-type levels in protoplasts. However, mutations in most other positions on both sides of the loop were extremely detrimental for satC accumulation. These results were not surprising given the strong conservation of the LSL, especially the upper region, among carmoviruses (Figure 1.4, 1.5) and the interaction between the 3' end and right side of the LSL that is necessary for proper initiation of satC in a cell-free assay programmed with RdRp purified from *E. coli* (Zhang et al., 2004). The upper stem (US) and lower stem (LS) of H5 show much less sequence and structural conservation among carmoviruses (Figure 1.4, 1.5). The US range from two to seven base-pairs and are capped by four to six base terminal loops in

the 11 carmoviruses with available sequences. Only GaMV contains an internal asymmetrical loop in the US (Figure 1.5). The LS range in length from eight to 12 bases, and eight of 11 carmoviral LS are interrupted by a single internal symmetrical or asymmetrical loop of varying sequences.

To determine whether the US, LSL, and LS of satC are virus-specific or whether these elements are exchangeable with their counterparts from other carmoviruses, the H5 of satC was precisely replaced with the H5 of JINRV and CCFV (Figure 4.2A). JINRV and CCFV H5 were selected for the following reasons: JINRV contains the identical LSL as TCV, which is identical to the LSL of satC at all but the lowest (and most flexible) position. The US of JINRV H5 is also nearly identical to that of satC, with the single base difference in the terminal loop still maintaining a highly stable GNRA tetraloop configuration (N is any nucleotide and R is a purine) (Moore, 1999). The LS of JINRV, however, differs at nearly every position from the LS of satC. In contrast, CCFV contains a truncated LSL compared with satC with a notable cytidylate to uridylate difference at satC position 300. This positional variance maintains but weakens base-pairing with the 3' end (see Figure 1.5), replacing a C•G pair with a U•G pair. The US of CCFV differs at six of nine positions with the US of satC including a C•G to U•A covariation in the stem and a highly stable UNCG tetraloop replacing the satC GNRA tetraloop. The LS, however, was nearly identical with that of satC, missing only a single base pair at the base and having two additional base-pairs directly below the LSL.

The parental construct for generation mutants in which satC H5 is replaced with H5 from CCFV and JINRV is $satC_E$ (Table 2.1), which contains two single base insertions in the linker sequence between H5 and Pr to ease cloning. $SatC_E$ containing

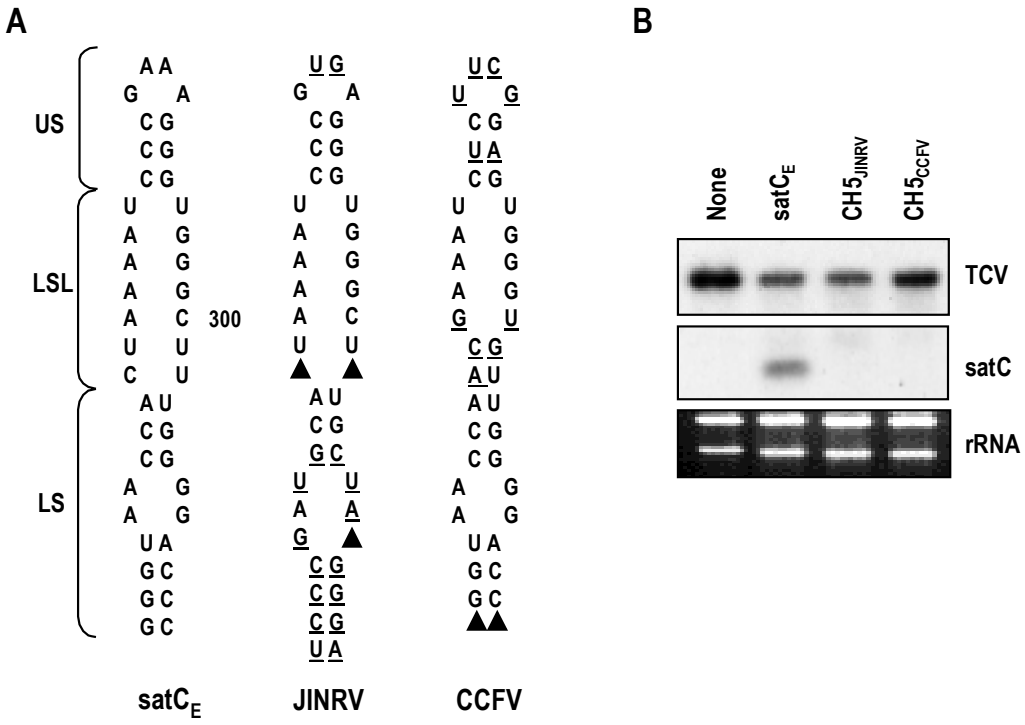


Figure 4.2 Accumulation of satC containing heterologous carmovirus H5 in protoplasts. (A) Sequence differences among the satC, JINRV and CCFV H5. Base differences are underlined. Triangles represent absent bases. SatC$_E$ is derived from satC by addition of two bases near the 3' side of H5, which was required for ease in cloning. (B) Accumulation of satC$_E$ and satC$_E$ containing the H5 of JINRV (CH5$_{JINRV}$) or CCFV (CH5$_{CCFV}$). RNA was extracted from protoplasts at 40 hpi. The RNA gel blot was probed with an oligonucleotide specific for both satC and TCV. None, no satC in the inoculum. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot).

either the H5 of JINRV ($CH5_{JINRV}$) or CCFV ($CH5_{CCFV}$) were assayed for accumulation at 40 hpi in Arabidopsis protoplasts. Positive-strands of both chimeric constructs did not accumulate to detectable levels when assayed by RNA gel blots, even following extensive overexposure of the blots (Figure 4.2B). Since I had previously shown that the lowest position of the satC LSL (C280 and U300) could be base-paired (as in TCV) without substantially altering satC levels in protoplasts (Zhang et al., 2004), the inability of $CH5_{JINRV}$ to accumulate implied that the lower stem (the main difference between the JINRV and satC H5) must contribute significantly to the function of H5. The poor accumulation of $CH5_{CCFV}$ suggested that either the CCFV US or LSL were also incompatible with satC accumulation.

To better understand the relative contribution of the three H5 domains (US, LSL, LS) to H5 function, $satC_E$ constructs were generated where one or two of the satC domains were converted to CCFV domains and the chimeric constructs assayed for accumulation in protoplasts (Figure 4.3A). At 40 hpi, none of the chimeric H5 constructs accumulated to $satC_E$ levels (Figure 4.3B). The CCFV US replacement construct (construct US) was the least debilitated, accumulating to 58% of $satC_E$ levels. Replacement of both the satC US and LSL with that of CCFV (construct US/LSL) or replacement of just the LSL (construct LSL) resulted in a 14- or 8-fold reduction in satC accumulation, respectively. It should be noted that the CCFV LSL domain transferred to satC in both of these constructs arbitrarily included the two base-pairs directly below the LSL. These results suggest that weakening the base-pairing between the 3' end and the CCFV LSL is detrimental to satC. However, additional factors, such as the inclusion of the two base-pairs flanking the LSL, may also have contributed to reducing satC

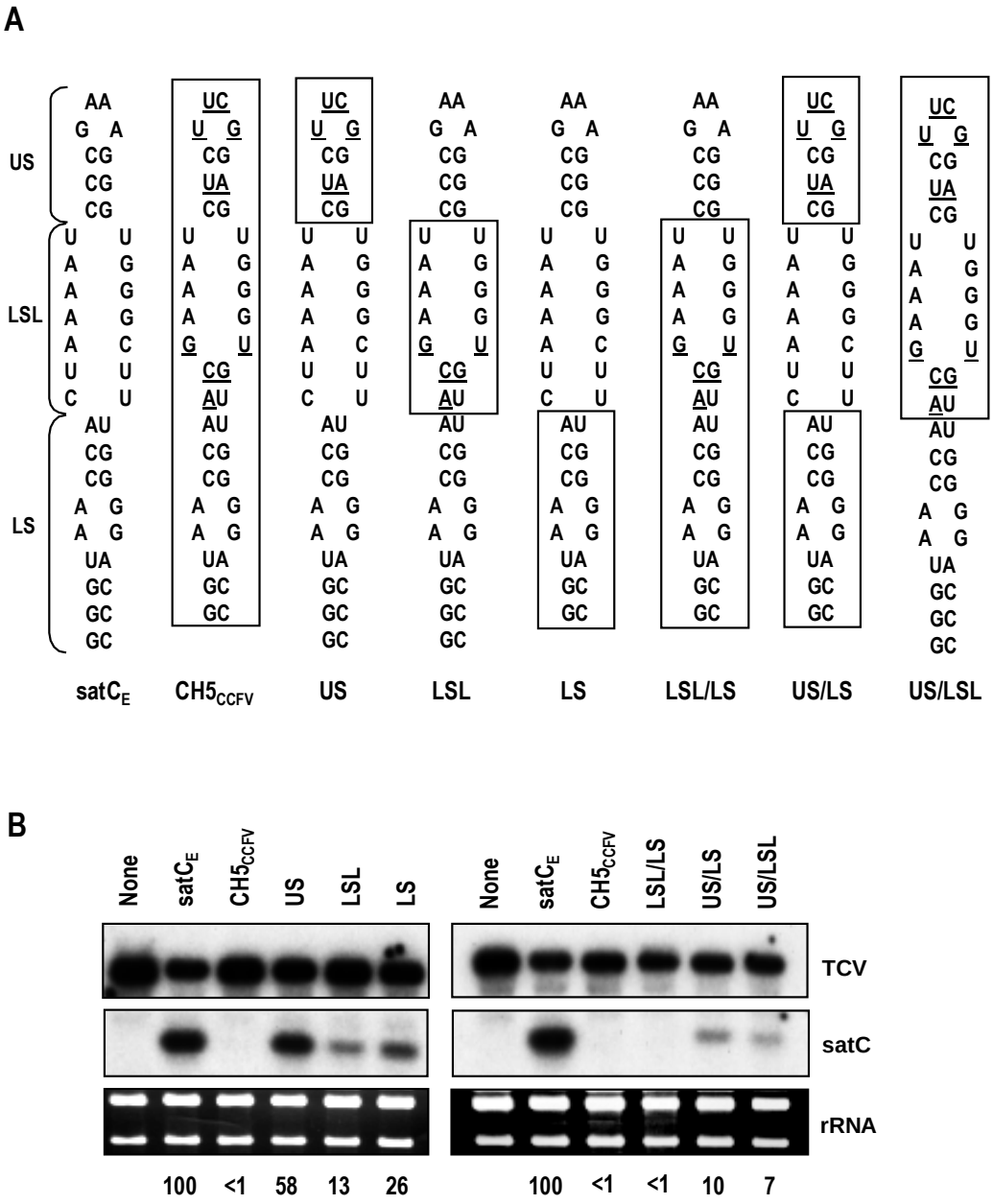


Figure 4.3 Replacing the US, LS and LSL of satC with that of CCFV. (A) Boxed sequences indicates portion of H5 that is derived from CCFV in the chimeric satC constructs. Base differences are underlined. Names of the constructs are given below the sequences. (B) RNA gel blot of viral RNAs accumulating in protoplasts at 40 hpi. The blots were probed with an oligonucleotide complementary to both satC and TCV. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot). Values given below the blots are the averages of two independent experiments, with the satC$_E$ level arbitrarily assigned a value of 100. Names of the constructs are from (A). None, no satC in the inoculum.

accumulation. Unexpectedly, there was a substantial effect of replacing the satC LS with that of CCFV (construct LS). Although construct LS differed from $satC_E$ by only a single G•C base pair at the base of the LS (present in satC [positions G270 and C311] and absent in CCFV), it accumulated to just 26% of $satC_E$ levels. This suggests that the length or stability of the lower stem is important for H5 function in satC. Replacement of both the LS and US with that of CCFV further reduced accumulation to only 10% of $satC_E$ levels. The most debilitating replacement was a combination of the CCFV LSL and LS, with the chimeric satC not accumulating to detectable levels. Altogether, these results suggest that efficient functioning of H5 in vivo requires all three H5 regions, but is most dependent on the cognate LSL and LS.

In vivo SELEX of the satC US

To further characterize the sequence and structural requirements for the satC US and LS, both elements were independently subjected to in vivo SELEX. SatC transcripts containing 10 randomized bases in place of the US were inoculated onto 30 plants and RNA was extracted from systemic leaves three weeks later. A total of 45 satC clones comprising 29 different US sequences were generated from ten randomly selected plants and sequenced (1^{st} round SELEX winners; Figure 4.4, Table 4.3). Up to five different satC were isolated from any one plant and satC with identical US sequence were not isolated from multiple plants. Examination of the sequences revealed a preference for three base-pair stems closed by a terminal four-base loop (20 of 29). The four-base loops varied in sequence with one comprising the wild-type GNRA tetraloop (U2) and four containing the very stable UNCG tetraloop (U5, U6, U18, and U19). Four of 29 clones

had only a two base-pair stem terminating in a six-base loop. Three clones contained sequence that would extend the stem of the US when the remainder of the H5 sequence is considered (boxed in Figure 4.4) and five clones (U21 through U25) would presumably cause the LSL to no longer be symmetrical. Four of the clones contained G•U or U•G pairs, supporting the formation of H5 on (+)-strands. While no fully wild-type sequences were recovered, the compositions of the stem sequences were not random. C•G was the preferred base pair (34 of 81 positions), followed by G•C (21 of 81), U•A (15 of 81), A•U (6 of 81), G•U and U•G (5 of 81). Analysis of the US in natural carmoviral H5 revealed a similar preference for C•G (25 of 44) over G•C (5 of 44) (Figure 1.4, 1.5).

To further explore the composition of US sequences that are beneficial to satC accumulation in plants, equal portions of RNA from the 30 1st round plants were combined and used to inoculate six additional plants. After three weeks, RNA was extracted from systemic leaves of each plant and 37 clones were generated and sequenced (Figure 4.5A, Table 4.3). The sequences were not previously identified among the 1st round winners, indicating the likelihood that many more functional US sequences were present but not isolated from 1st round plants. Of the 2nd round winners, only one clone was isolated from all six plants (U30), comprising 16 of the 37 clones sequenced. This clone contained a UNCG tetraloop and two of three C•G base pairs. All other clones were isolated from only one or two plants and contained four or six base terminal loops. As with the 1st round winners, the 2nd round winners were most likely to have C•G (16 of 32 positions) or U•A (8 of 32) base-pairs in the stem.

RNA from the six 2nd round plants was pooled and used to inoculate six new plants, producing the final 3rd round winners. Out of the 32 clones sequenced from these

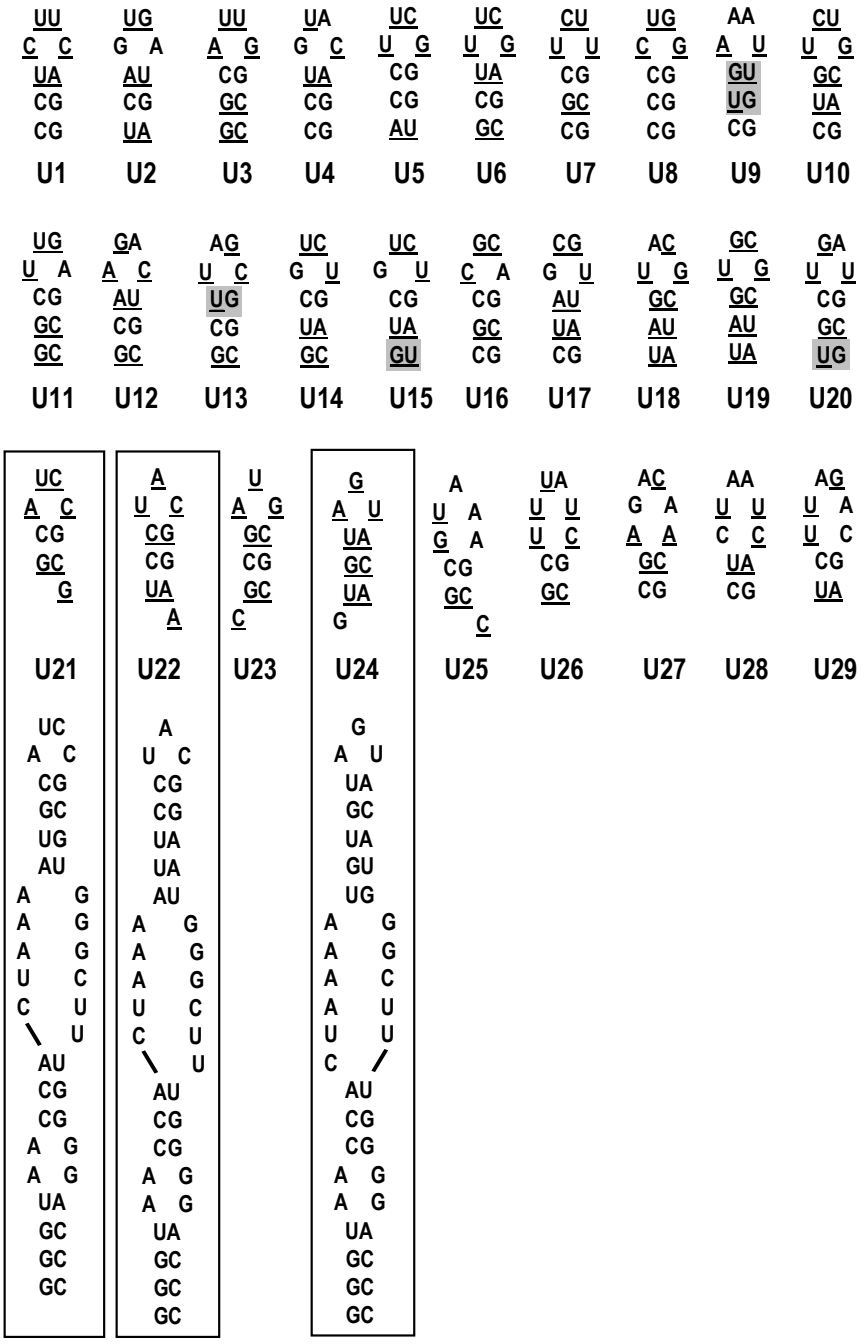


Figure 4.4 First round US SELEX winners. SatC containing randomized bases replacing the US were inoculated with TCV onto 30 turnip plants. SatC species accumulating three weeks later were cloned and sequenced from 10 plants. No particular sequence was found in more than one plant. Differences with wild-type satC US sequences are underlined. Clones containing "orphaned" bases that can putatively base-pair with an LSL base thus expanding the US (as indicated) are boxed. G•U base-pairs (suggesting a plus-strand structure) are shaded. U21 had a single base deletion in the selexed region.

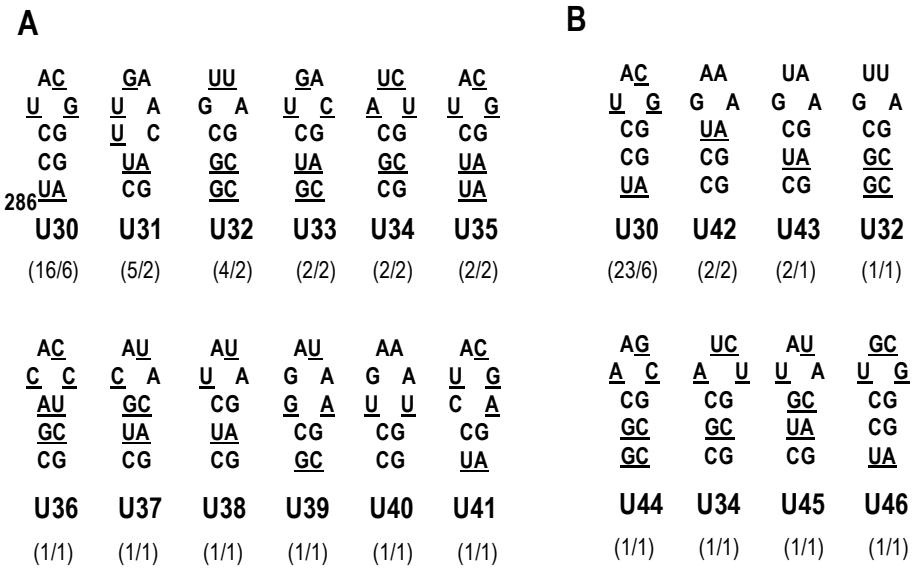


Figure 4.5 Second and 3rd round US SELEX winners. (A) Second round winners. (B) Third round winners. Equal amounts of RNA from 1st or 2nd round plants were combined and used to inoculate six plants. SatC accumulating three weeks later were cloned and sequenced. The first number in the parentheses below each clone represents the total number of that clone in the sequenced population. The second number is the number of plants (out of six total) that contained the particular clone. Underlined bases differ from the wild-type satC US.

TABLE 4.3 Summary of in vivo SELEX of the H5 US

Name	Sequence			Plant										Total
	5' stem	Loop	3' stem	1	2	3	4	5	6	7	8	9	10	
Round 1														
U1	CCU	CUUC	AGG					4						4
U2	UCA	GUGA	UGA								4			4
U3	GGC	AUUG	GCC	3										3
U4	CCU	GUAC	AGG				3							3
U5	ACC	UUCG	GGU										2	2
U6	GCU	UUCG	AGC							2				2
U7	CGC	UCUU	GCG				2							2
U8	CCC	CUGG	GGG		2									2
U9	CUG	AAAU	UGG			2								2
U10	CUG	UCUG	CAG									2		2
U11	GGC	UUGA	GCC									1		1
U12	GCA	AGAC	UGC		1									1
U13	GCU	UAGC	GGC								1			1
U14	GUC	GUCU	GAC							1				1
U15	GUC	GUCU	GAU							1				1
U16	CGC	CGCA	GCG										1	1
U17	CUA	GCGU	UAG			1								1
U18	UAG	UACG	CUA									1		1
U19	UAG	UGCG	CUA									1		1
U20	UGC	UGAU	GCG						1					1
U21	GC	AUCC	GC G		1									1
U22	C GCG	AUG	CGC	1										1
U23	UCC	UAC	GGA A				1							1
U24	GC	GUAAA	GCC			1								1
U25	G UG	UAGUA	CA										1	1
U26	GC	UUUAUC	GC		1									1
U27	CG	AGACAA	CG			1								1
U28	CU	CUAAUC	AG			1								1
U29	UC	UUAGAC	GA		1									1
Total clones assayed for each plant				4	6	6	6	4	1	4	5	5	4	45
Round 2														
U30	UCC	UACG	GGA	5	3	2	2	3	1					16
U31	CU	UUGAAC	AG				2		3					5
U32	GGC	GUUA	GCC			3	1							4
U33	GUC	UGAC	GAC	1		1								2
U34	CGC	AUCU	GCG				1		1					2
U35	UUC	UACG	GAA		1			1						2

TABLE 4.3 continued

Name	Sequence			Plant										Total
	5' stem	Loop	3' stem	1	2	3	4	5	6	7	8	9	10	
U36	CGA	CACC	UCG				1							1
U37	CUG	CAUA	CAG		1									1
U38	CUC	UAUA	GAG						1					1
U39	GC	GGAUAA	GC						1					1
U40	CC	UGAAAU	CG	1										1
U41	UC	CUACGA	GA		1									1
Total clones assayed for each plant				7	6	6	7	4	7					37
Round 3														
U30	UCC	UACG	GGA	1	4	7	7	1		3				23
U42	CCU	GAAA	AGG	1					1					2
U43	CUC	GUAA	GAG	2										2
U32	GGC	GUUA	GCC	1										1
U44	GGC	AAGC	GCC					1						1
U34	CGC	AUCU	GCG			1								1
U45	CUG	UAUA	CAG					1						1
U46	UCC	UGCG	GGA		1									1
Total clones assayed for each plant				5	5	8	7	3	4					32

six plants at 3-weeks postinoculation, U30 again was the only clone found in all six plants (Figure 4.5B, Table 4.3). Two other clones from the 2nd round were also recovered (U32 and U34) along with five new clones (U42 through U46). One of the new clones, U42 differed from 2nd round winner U40 by a single uridylate to adenylate transversion that converted a two base-pair stem/six base loop configuration to a three base-pair stem/GNRA tetraloop. In vivo SELEX frequently results in evolution of particular sequences to more fit sequences in later rounds due to the high frequency of RdRp errors, and thus it is likely that U42 was derived from U40. Sequence evolution was also apparent for most of the other 3rd round winners: U43 differed from 2nd round winner U38 by a uridylate to guanylate alteration in the 5' most position in the four base terminal loop, generating a stable GNRA tetraloop in U43; U45 differed from 2nd round winner U37 by a single position in the four base terminal loop; U46 differed from U30 by a single change that maintained the UNCG tetraloop.

Because of sequence evolution, the number of recovered clones in the 3rd and final round did not necessarily indicate which clones represent the most fit satRNA. To determine an approximate order of fitness of some SELEX winners, 3rd round winners U30 (most prevalent) and U42 (most similar to wild-type satC; differed by only a single C•G to U•A covariation at position 288/293) along with 2nd round winners U31 and U22 were subjected to direct competition. U31 and U22 were selected for inclusion in the competition assay because their US terminated in a six base loop or three base loop, respectively. In addition, U22 differed from U30 by only a single base difference yet contained a substantially altered structure (Figure 4.6A). Three plants were inoculated with equal portions of transcripts derived from the four clones and progeny were

TABLE 4.4 Competition between 3rd round SELEX winners and wild-type satC

	Clone	# Recovered from 3 plants
US:		
Competition 1	U30	8
	U42	18
	U31	0
	U22	0
Competition 2	U42	17
	wild-type satC	14
Competition 3	U30	9
	wild-type satC	18
LS:		
Competition 1	L3	0
	wild-type satC	26

examined from all plants at 3-weeks postinoculation. As shown in Table 4.4, 3rd round winner U42 was the most fit among the clones assayed in plants, comprising 70% of the 26 sequenced clones. All remaining clones were U30. To determine the fitness of U42 and U30 compared with wild-type satC, additional competition assays were performed. U42 and wild-type satC were found to be of similar fitness, comprising 55% and 45% of the cloned population, respectively. Wild-type satC was more fit than U30, comprising 67% of the recovered clones. These results indicate a fitness order for accumulation in plants of U42 and wild-type satC, U30, followed by U31 and U22.

Fitness of satC to accumulate in plants reflects replication competence, stability, trafficking ability, and capacity of enhancing TCV movement. To determine if the satRNAs most fit to accumulate in plants were also the best templates for accumulation in protoplasts (which reflects only replication competence and stability), U42, U30, U31 and U22 were individually inoculated with helper virus onto protoplasts along with wild-type satC and satC containing a randomly selected US sequence (Rd; Figure 4.6A). Although U42 and wild-type satC were equally fit in plants, U42 accumulated to only 46% of wild-type satC levels and reached only 60% of U30 levels. The efficient accumulation of U30 (82% of wild-type satC) suggests that the highly stable satC GNRA tetraloop can be replaced by the equally stable UNCG tetraloop (Figure 4.5B; Figure 4.6C). This argues that the satC GNRA tetraloop is likely present to stabilize the H5 structure and is not involved in tertiary interactions with docking sequences (Abramovitz and Pyle, 1997).

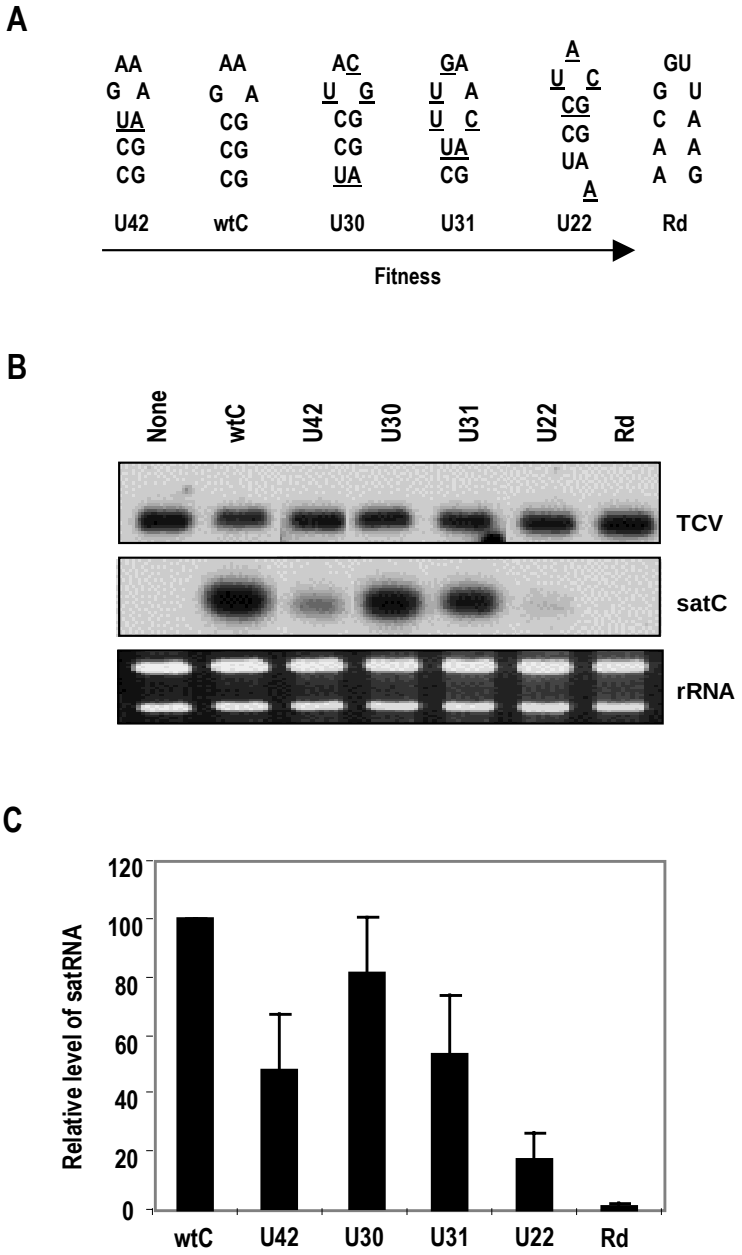


Figure 4.6 Fitness of US SELEX winners to accumulate in protoplasts. (A) Fitness of clones to accumulate in plants (from Table 4.3). U42 and U30 were 3rd round winners. U31 was a 2nd round winner and U22 was a 1st round winner. Rd represents a randomly selected satC from the initial SELEX population generated by PCR. Underlined bases in the SELEX winners are positions that differ from wild-type satC (wtC). (B) Representative RNA gel blot of total RNA extracted at 40 hpi of protoplasts. None, no satC in the inoculum. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot). (C) Quantification of satC accumulation levels. Data are from five independent repetitions.

In vivo SELEX of the satC LS

To investigate sequence requirements in the satC LS, 18 residues were randomized and the satC population subjected to in vivo SELEX (Figure 4.7A). Of the 60 plants inoculated with the satC population, only five contained satC detectable by agarose gel electrophoresis and ethidium bromide staining. Forty-six clones were isolated from these plants and the 10 unique sequences are presented in Figure 4.7B. Clones isolated from the same plant had very similar sequence and likely originated from a single transcript followed by sequence evolution. Interesting features of the 1st round winners included: six to nine base-pair stems in the selexed region; a weak base-pair (U•A, U•G or A•U) at positions 278/303 just below the LSL in most sequences (A•U in wild-type satC); the lack of interruption of the base-paired region in most clones by symmetrical or asymmetrical interior loops; and a G•C pair in the first position of all stems with one exception. An additional common feature was the surprising presence of putatively single-stranded pyrimidines flanking the H5, which were identical pyrimidines in all of the sequences except L7 (satC and TCV H5 are flanked by uridylates). For L7, the base of the LS stem contained a single unpaired 3' side adenylate in the selexed region that could pair with the 5' uridylate flanking the selexed bases. This would place two non-identical pyrimidines flanking the H5. The presence of flanking pyrimidines was surprising given that they would be adjacent to the natural satC uridylates at positions 269 and 312. The presence of identical pyrimidines flanking H5 did not extend to most other carmoviral H5 (Figure 1.5), and thus the significance of this finding is not known.

For the 2nd round, equal portions of RNA isolated from the 60 plants were pooled and used to inoculate six additional plants. Of the 35 clones generated following RNA

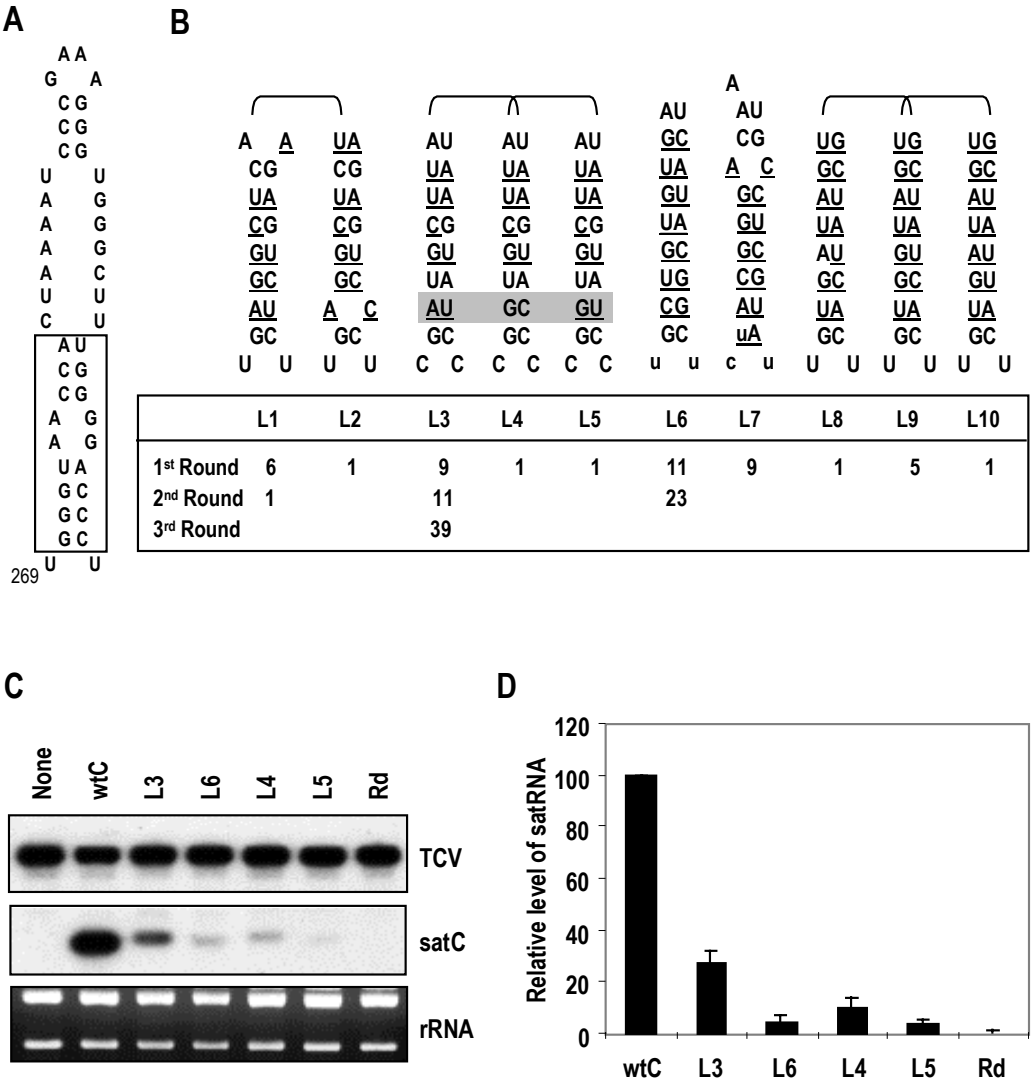


Figure 4.7 In vivo SELEX of the H5 LS. (A) SatC H5. Bases subjected to SELEX analysis are boxed. (B) LS SELEX winners. Sequences of clones found in 1st round plants are shown. Lines connecting sequences indicate isolation from the same plant. Bases that differ from wild-type satC LS are underlined. Names of the winners are shown below the sequences. Numbers of particular clones contained in the sequenced population from three round plants are given below the sequence names. Shading indicates the single positional differences among L3, L4 and L5. Lower case letters denote that these bases were not subjected to selection. (C) Representative RNA gel blot of total RNA extracted at 40 hpi of protoplasts. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot). None, no satC in the inoculum. wtC, wild-type satC. Rd, satC with a randomly selected LS sequence (5' CACACUUAA--AUUCAACUC 3'). (D) Quantification of satC accumulation levels. Data are from three independent repetitions.

extraction at three-weeks postinoculation, only three sequences from the first round were represented and only two of these sequences were found in multiple plants (L3 and L6). L6 was the most prevalent 2nd round winner, comprising 23 of the 35 clones. RNA from the six 2nd round plants was pooled and used to inoculate six additional plants. In the 3rd and final round, only L3 was found out of 39 clones sequenced. L3 differed from 1st round sister clones L4 and L5 at only a single position, the second base-pair from the base of the stem (L3, A•U; L4, G•C; L5, G•U). The enhanced fitness of L3 compared with L4 and L5 suggests that the identity of base pairs in the stem is important and not the absolute strength of the pairings.

L3 was subjected to direct competition in plants along with wild-type satC. All 26 clones isolated three-weeks postinoculation were wild-type satC, indicating a strong preference for the wild-type LS stem. To determine if fitness to accumulate in plants correlated with enhanced accumulation in protoplasts, wild-type satC along with 2nd round winners L3 and L6, and L3 sister clones L4 and L5 were examined for accumulation in protoplasts along with TCV helper virus. Third round winner L3 accumulated 7-fold better than 2nd round winner L6 and 3- to 9-fold better than sister clones L4 and L5. However, L3 only accumulated to 27% of wild-type satC levels (Figure 4.7C and D). The results of the LS SELEX support the need for a base-paired stem with sequence-specific preferences. However, poor accumulation of 3rd round winner L3 compared to wild-type satC indicates that there was insufficient complexity in the initial population of randomized satC that entered plant cells with TCV to efficiently meet these two criteria.

Relationship between H5 and the previously characterized (-)-strand element, 5'PE

Two short linear sequences were previously identified as being redundant elements required for (+)-strand synthesis using (-)-strand satC templates in in vitro transcription system programmed with partially purified TCV RdRp (Guan et al., 1997). SatC containing mutations in one element, the 5'PE, replicated poorly in protoplasts, with a greater reduction in accumulation of (+)-strands compared with (-)-strands (Guan et al., 2000b). Taken together, these results led to the suggestion that the 5'PE was a (-)-strand element that functioned in (+)-strand synthesis. However, 10 of 13 bases of the 5'PE comprise the complement of the 3' side lower stem of H5 (positions 302 to 314; Figure 4.8A). The importance of both sequence and structure of the lower stem of H5 for satC accumulation revealed in the current study required a re-investigation of the 5'PE to determine if it was truly an independent (-)-strand element or whether the sequence functioned as a portion of the H5 LS. This latter explanation would indicate that H5 functions in both (-)-strand and (+)-strand accumulation.

Altering G304 to U (G304U) or A (G304A) was previously demonstrated to be highly detrimental to satC accumulation in protoplasts, reducing (+)-strands to below the level of detection while maintaining low levels of (-)-strands (Guan et al., 2000b). G304 is located in the middle position of the three base stem portion in the H5 LS, and thus these mutations are predicted to substantially alter the structure of H5 (Figure 4.8A). Confirming previous results, satC (+)-strands containing G304A or G304U did not accumulate to detectable levels whereas (-)-strands accumulated to 2% and 5% of wild-type satC levels, respectively (Figure 4.8B). SatC with a new alteration at position 304, G304C, did not accumulate detectable levels of either (+)- or (-)-strands. Alteration of

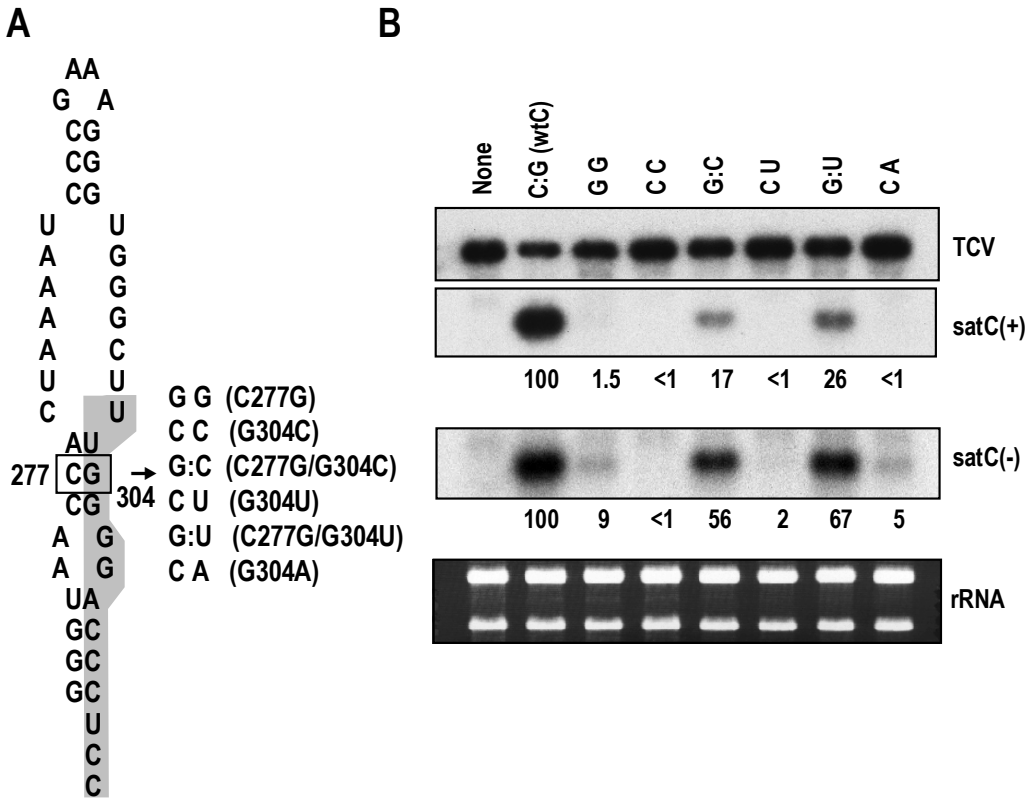


Figure 4.8 Mutational analysis of the upper portion of the H5 lower stem. (A) Single or double mutations were constructed at positions 277 and/or 304 in satC leading to the plus-strand pairings at this position as shown. Names of the mutants are given to the right in parentheses. The sequence complementary to the previously described 5'PE is shaded. (B) SatC containing these mutations were inoculated onto protoplasts along with TCV helper virus and satC accumulating at 40 hpi was examined by RNA gel blots using plus- or minus-strand specific probes. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot). Letters above each lane denote the identity of the residues in position 277 and 304. Numbers below the panels are the average values from two independent experiments. None, no satC in the inoculum; wtC, wild-type satC.

C277, which is proposed to pair with G304 in H5, to a guanylate (C277G) was also detrimental, with (+)-strands accumulating to 1.5% of wild-type satC and (-)-strands accumulating to 9% of wild-type satC. Combining G304C with C277G, which re-establishes base-pairing at this position in the hairpin, enhanced (+)-strand accumulation to 17% of wild-type satC. These combined mutations had an even greater effect on (-)-strand accumulation, enhancing the level of (-)-strands to 56% of wild-type satC. When C277G was combined with G304U, which would allow for a G:U pairing at this position in (+)-strands, satC (+)-strands reached 26% of wild-type levels and (-)-strand levels were enhanced to 67% of wild-type. Since G304A was highly detrimental despite allowing for a G:U pairing in a presumptive (-)-strand H5 structure, these results support H5 as a (+)-strand structure. In addition, these results suggest that the 5'PE is not an independent (-)-strand element but rather that prior mutations in the 5'PE were detrimental because they affected the H5 LS.

All mutations tested resulted in either undetectable levels of (+)- and (-)-strands or in a disproportionate reduction of (+)-strands. One possible explanation for this effect is that the H5 LS mutations reduce the stability of (+)-strands. To test for this possibility, protoplasts were inoculated with wild-type satC or satC containing C277G/G304C, C277G/G304U, or CΔH5 in the absence of TCV and undegraded RNA examined between one and six hours postinoculation by Northern blots (Figure 4.9). Unfortunately, the results showed that the time points of sampling were improperly chosen. However, studies of some other viruses suggested that instability of mutated (+)-strands was not responsible for the disproportionate reduction in (+)-strand levels (see discussion).

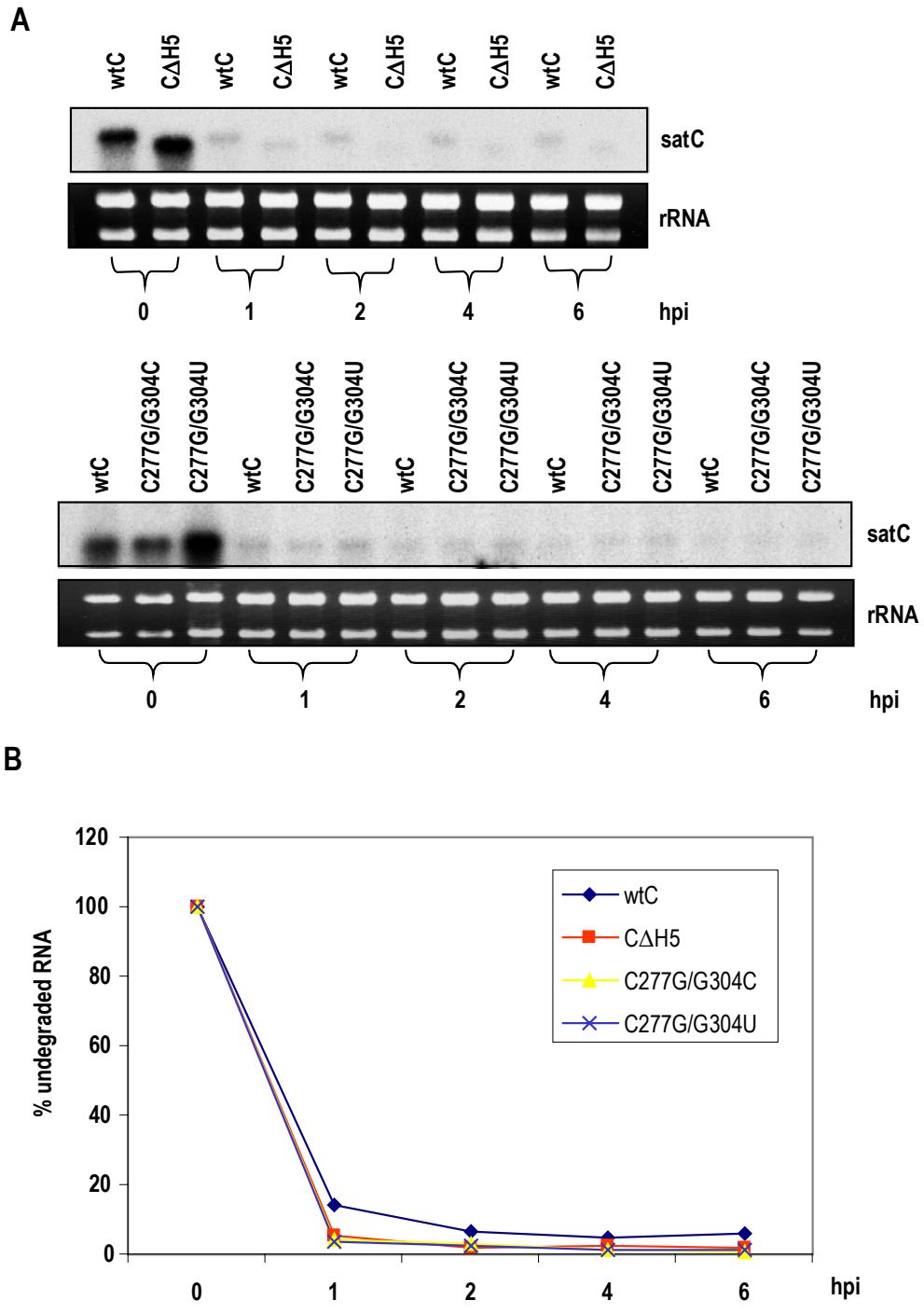
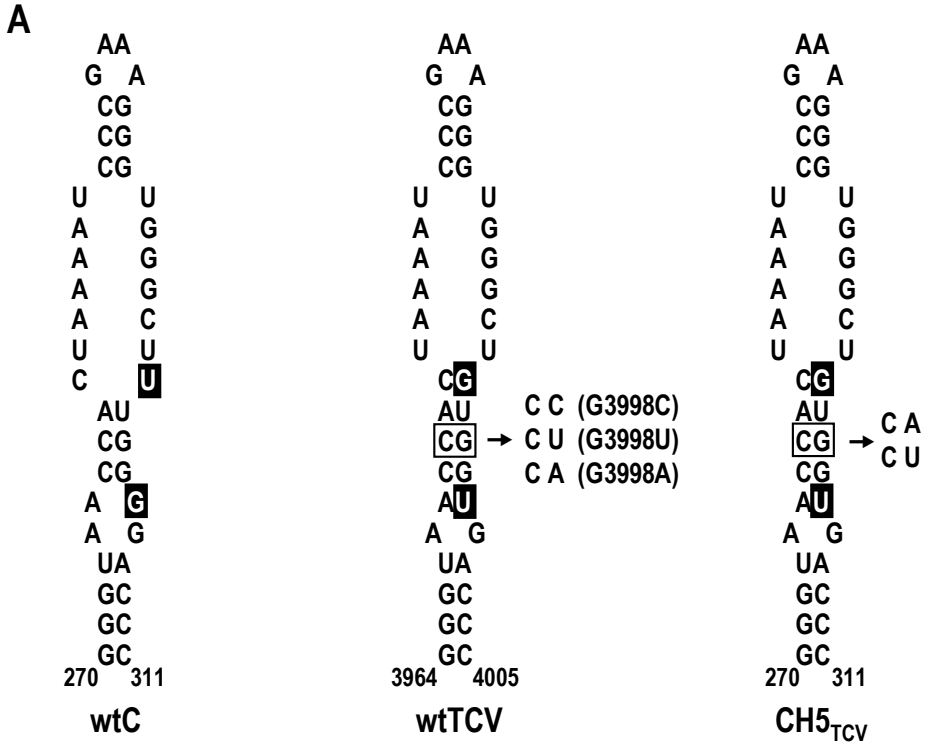


Figure 4.9 Effect of mutations and deletion of satC H5 on RNA stability in protoplasts. (A) Representative RNA gel blot. Time points of sampling are shown. wtC, wild-type satC. (B) Quantification of satC levels.

Figure 4.10 Affect of mutations at position 304 on the analogous location in TCV H5 and in satC containing H5 of TCV. (A) Single mutations were generated in TCV H5 or in satC with H5 of TCV ($CH5_{TCV}$) as shown. Names of the TCV mutants are given in parentheses. Residues boxed in black denote differences between H5 of satC and TCV. (B) $CH5_{TCV}$ containing the alterations shown in (A) were inoculated onto protoplasts along with TCV helper virus and satC accumulation accessed at 40 hpi. Letters above each lane denote the identity of the residues in position 277 and 304. Only satC plus-strands are shown. Numbers below the panels are the average values from two independent experiments. None, no added satRNA; wtC, wild-type satC. (C) TCV containing the mutations shown in (A) were inoculated onto protoplasts and the level of viral RNA accumulating at 40 hpi was examined by RNA gel blots. Letters above each lane denote the identity of the residues in positions 3971 and 3998. Numbers below the panel are the average values from two independent experiments. Mock, no TCV in the inoculum. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot).


A
wtC
wtTCV
CH5TCV
C C (G3998C)
C U (G3998U)
C A (G3998A)
C A
C U
270 311
3964 4005
B
None
C:G (wtC)
C:G (CH5TCV)
C A (CH5TCV)
C U (CH5TCV)
TCV
satC(+)
rRNA
100 108 58 59
C
Mock
C:G (wtTCV)
C C (G398C)
C U (G3998U)
C A (G3998A)
TCV
rRNA
100 90 90 79

To support the conclusion that the 5'PE is not an independent (-)-strand element, H5 of satC was precisely replaced with H5 of TCV, generating $CH5_{TCV}$ (Figure 4.10A). TCV H5 differs from satC H5 at positions 302 and 306 within the complementary 5'PE sequence, which results in a five base-stem just below the H5 LSL compared with a three-base stem for satC H5. The G304 analogous position in TCV H5 is located in the center of the five-base stem and thus altering this position should not be as disruptive to the structure of satC H5. $CH5_{TCV}$ consistently accumulated to slightly higher levels than wild-type satC in protoplasts indicating that the two base differences between the satC and TCV H5 do not negatively impact on the replication of satC (Figure 4.10B). Positive-strands of $CH5_{TCV}$ containing G304A or G304U accumulated to 58 and 59% of wild-type satC levels, respectively, compared with undetectable levels for the analogous satC mutants. The same mutations generated in wild-type TCV H5 (G3971C, G3971U or G3971A) also had only a marginal effect on TCV accumulation, with levels reaching 90%, 90% or 79% of wild-type TCV levels, respectively (Figure 4.10C). Altogether, these results indicate that mutations that disrupt satC H5 impair (+)- and (-)-strand accumulation and that the complementary sequence to the H5 3' side lower stem may not have an independent role in satC replication.

Discussion

Prior reports indicated an important role for H5 in satC and TCV replication in vivo (McCormack and Simon, 2004; Zhang et al., 2004; Chapter III) and in transcription in vitro using purified TCV RdRp (Zhang et al., 2004). It was suggested that H5 is

involved in the correct assembly of the RdRp since mutations in the TCV H5 LSL caused a significant increase in mutation frequency (McCormack and Simon, 2004). While carmoviral H5 have varying degrees of sequence similarity, all are topologically similar and capable of forming four base pairs between their 3' side LSL and the 3' terminus of the genomic RNA, suggesting that this interaction is also necessary for proper viral replication (Zhang et al., 2004). Tombusvirus hairpin SL3, which also interacts with 3' terminal sequences, likely performs a function analogous to the carmovirus H5 (Fabian et al., 2003; Pogany et al., 2003).

The current report indicates, however, that H5 are not functionally interchangeable even when base-pairing between the LSL and 3' end is putatively maintained. Despite similarity between the JINRV and satC LSL, differing only in the lowest and most flexible position, satC with H5 of JINRV did not accumulate to detectable levels in protoplasts. This suggests that the US or LS of JINRV is not compatible with the remaining satC sequence. SELEX of the satC US indicates a preference for a stable tetraloop closed by a C•G base-pair with at least one additional C•G pair in the short stem. However, a variety of other sequence/structural combinations were also functional, indicating substantial plasticity in the upper portion of the hairpin. This suggests that the two positional differences between satC and JINRV H5 tetraloops, which maintain the GNRA configuration, are not likely responsible for the negative effect on satC accumulation and suggest instead that the JINRV LS is incompatible with satC accumulation.

Exchanging the satC US, LSL, and/or LS with equivalent regions from the CCFV H5 support the importance of the cognate LS for H5 function. The absence of a single

G:C pair at the base of the LS (construct LS, Figure 4.3) resulted in a 74% decrease in satC accumulation in protoplasts. The most debilitating exchange of individual H5 regions was the replacement of the satC LSL with the CCFV LSL (construct LSL, Figure 4.3), which reduced accumulation by 87%. However, this construct also extended the LS by two base-pairs, and thus it is not known whether the elongation of the LS or the base differences in the LSL are responsible for the negative effect on satC replication. In vivo SELEX of the satC H5 LS confirmed the importance of both structure and sequence of the LS. With only one exception, all winners contained a G:C pair at the base of the LS flanked by unpaired, identical pyrimidines and a weak base-pair adjacent to the LSL. In addition, the single covariant position in the LS winners L3, L4 and L5 had a substantial effect on satC accumulation in protoplasts that was unrelated to the strength of the paired bases. The importance of both structure and sequence of the LS is also supported by results indicated enhanced satC accumulation with a G•U replacing the C•G at positions 277/304 compared with a G•C (Figure 4.8).

Altogether, these results suggest that the H5 LS functions in more than a purely structural role supporting the phylogenetically conserved structure of H5. While it is possible that factors interacting with H5 may require LS functional groups in specific locations, it is also possible that H5 undergoes a structural rearrangement as part of its role in satC replication. Deletion of the 3' terminal three cytidylates significantly alters the structure of H5 and 3' flanking sequences without substantially affecting the remainder of satC (Zhang et al., 2004). Such structural rearrangement of H5 might involve a secondary interaction between LS sequences and other partner sequences, thus constricting the nature of bases in the stem. Several recent findings support a role for

sequences external to H5 in supporting H5 function. Wang and Wong (2004) determined that the poor ability of TCV H5 to substitute for the H5 of HCRSV could be improved by co-transferring the TCV Pr core promoter hairpin. In addition mutations in the LS of TCV H5 that affected the small interior loop lead to second site alterations in the nearby hairpin H4b (R. Zamora and A.E. Simon, unpublished results). All together, these results suggest complex interactions between H5 and other sequences in satC are likely required for efficient replication.

The importance of the H5 LS for H5 function also explains previous results on an element named the 5'PE. This element was first identified as required for transcription of (+)-strands from (-)-strand templates in vitro in the absence of 3' proximal sequences (Guan et al., 1997). The element was able to independently promote complementary strand synthesis in vitro, and mutations in the sequence within satC resulted in an enhanced reduction of (+)-strands compared with (-)-strands in vivo. Taken together, these results supported the hypothesis that the 5'PE was a (-)-strand element involved in (+)-strand accumulation. However, the mutations constructed in the 5'PE are now predicted to substantially alter the structure of H5 on the complementary (+)-strand. Analogous mutations in satC with H5 of TCV ($CH5_{TCV}$), which disrupt the center of a five base stem in the LS (compared with disrupting the center of a three base stem for satC H5) had a much reduced affect on accumulation of satC (Figure 4.10B). This suggests that the previous mutations in satC were partially or fully disrupting H5 function rather than the complementary sequence. The sequence-specific nature of the winners of our previous in vivo SELEX of the 5'PE (Guan et al., 2000b; shaded sequence in Figure 4.8A) can now be explained by a requirement to maintain the sequence and structure of

the LS. Interestingly, this SELEX revealed that the UCC flanking the 3' side of H5 was conserved in all winners and could be preceded by a random base. The role of the UCC in satC replication will be addressed in the next chapter.

All mutations tested at positions 277 and 304 in the satC LS caused a greater reduction in the accumulation of (+)-strands compared with (-)-strands. This suggests that H5 may function in (+)-strand as well as (-)-strand accumulation. Elements located proximal to the 3' end of (+)-strands that disproportionately reduce the accumulation of (+)-strands compared with (-)-strands have also been found for other viruses. For example, deletion of the 3' UTR of poliovirus resulted in the reduction of (+)-strands to only 10% of wild-type levels in neuronal cells without decreasing (-)-strand levels (Brown et al., 2004). In PVX, mutations that affected either a 3' proximal (+)-strand hairpin or a putative polyadenylation signal reduced progeny (+)-strands by 65 to 80% compared with (-)-strand reductions of only 30 to 40% (Pillai-Nair et al., 2003). In both examples, instability of mutated (+)-strands was not responsible for the reduction in (+)-strand levels (Brown et al., 2004; Hemenway, personal communication). A second possibility for how a (+)-strand element can affect (+)-strand synthesis is if the element alters the structure of the RdRp or assembly of the replicase complex, which may have distinct forms for transcription of (-)- and (+)-strands. Synthesis of (+)- and (-)-strands by two replication complexes with differing stabilities has been shown for SINV (Dé et al., 1996). Our previous suggestion that H5 may be nucleating the TCV replication complex (McCormack and Simon, 2004) supports this possibility.

The results of the in vivo SELEX of the US indicate that enhanced fitness of winner U42 in plants did not correlate with enhanced accumulation in plant cells. This

supports findings from the previous satC LSL SELEX that fitness in plants of H5 mutants does not always correlate with increased accumulation in protoplasts (Chapter III). Currently it is not known what additional role(s) outside of replication might require H5. Past in vivo SELEX of the satC (-)-strand M1H enhancer revealed that one of the most fit winners replicated only marginally better than random sequence. This winner, however, was shown to be exceptionally efficient at reducing virion formation [due to a hairpin that formed on the (+)-strand], thus enhancing the ability of TCV to overcome RNA silencing (Zhang and Simon, 2003a). Whether H5 is also involved in reducing virion levels has not yet been explored.

CHAPTER V

A PSEUDOKNOT IN A PRE-ACTIVE FORM OF A VIRAL RNA IS PART OF A STRUCTURAL SWITCH ACTIVATING MINUS-STRAND SYNTHESIS

Introduction

The inherent ability of RNA to switch conformations in response to different physiological conditions has fundamental implications for regulation of many cellular processes, including transcription termination, protein translation, and RNA cleavage (Nagel and Pleij, 2002; Brantl, 2004). The need for RNA viruses to switch between mutually exclusive processes for genome amplification suggests that RNA switches may also control different steps in the virus life cycle. For example, (+)-strand genomes must initially assume a conformation that is recognized by cellular ribosomes for translation of viral products such as the RdRp. At some point, the RNA must switch to a form that is not available for translation but contains cis-acting elements recognized by the RdRp leading to initiation of (-)-strand synthesis (van Dijk et al., 2004). Following reiterative synthesis of (+)-strands from (-)-strand templates, newly synthesized (+)-strands of some viruses may not be templates for further (-)-strand synthesis (Chao et al., 2002; Brown et al., 2004), suggesting that these strands may need to adopt a structure that is incompatible with RdRp recognition.

RNA conformational switches control RNA dimerization in retroviruses (Huthoff and Berkhout, 2001; Greatorex, 2004) and ribozyme activity in small virus-associated RNAs in vitro (Ke et al., 2004; Song and Miller, 2004). Changes in viral RNA conformation in 3' regions of the genome that hide or expose the 3' terminus (Olsthoorn et al, 1999; Schuppli et al, 2000; Koev et al, 2002) or permit the formation of important cis-acting structures (Khromykh et al., 2001) likely regulate initiation of (-)-strand synthesis. Evidence for important alternative structures with no known function have also been reported (Goebel et al., 2004; Dutkiewicz and Ciesiolka, 2005).

RNA switches usually involve long distance tertiary interactions that stabilize one of the RNA conformations (Klovins et al., 1998; Huthoff and Berkhout, 2001; Khromykh et al., 2001). Therefore removal of cis-acting elements from their natural context for biochemical and biophysical analyses may be leading to an oversimplification of many viral processes. However, efforts to identify conformational switches and equate RNA structure with biological function in intact viruses is complicated by large genome sizes. In addition, switches that may be replication-specific are difficult to identify since both replication and translation occur on the same RNA. These problems can be overcome by analyzing the amplification of small, untranslated subviral RNAs (Simon et al., 2004). Subviral RNAs such as satRNAs and DI RNAs have limited genome sizes while containing all cis-elements necessary to utilize the replication components provided by their helper viruses.

As described in previous chapters, satC contains the 3' terminal 151 nt of TCV genomic RNA, and thus is an excellent model for studying 3' proximal elements required by the TCV RdRp for robust and accurate initiation of (-)-strand synthesis. Using a

combination of computer modeling and phylogenetic comparisons of carmoviral 3' UTR sequences, four (+)-strand hairpins (Pr hairpin, H5, H4a and H4b) were determined to be structurally and spatially conserved among nearly all carmoviruses. A combination of approaches including single site mutagenesis, in vivo SELEX, and sequence replacements with analogous segments from the related carmovirus CCFV indicate that both specific sequence and structural features throughout H5 are necessary for robust satC accumulation in plants and protoplasts (Chapter III, Chapter IV). Synthesis of wild-type levels of satC complementary strands by the RdRp in vitro requires an interaction between the 3' end and the H5 LSL in the template RNA as assayed by a compensatory mutagenesis approach (Zhang et al., 2004). However, in vitro transcription of wild-type satC with T7 RNA polymerase results in transcripts that fold into an initial conformation ("pre-active" structure) that does not contain the phylogenetically inferred H5, Pr structure or the 3' end/H5 interaction (Zhang et al., 2004) bringing into question at what point the 3' end interacts with H5 in the progression of events leading to transcription initiation.

My colleagues and I recently proposed that a conformational switch affecting most of the 3' terminal 140 nt of (+)-strand satC may be an integral step leading to proper initiation of complementary strand synthesis in vitro (Zhang et al., 2006). This model was based on finding that deletion of the 3' terminal three cytidylates, 5' terminal two guanylates or specific mutations in the H5 region caused a rearrangement of the Pr from its initial Pr-1 configuration to the alternative, phylogenetically inferred Pr-2 structure. This conformational rearrangement to Pr-2 correlated with a substantial (>20-fold) increase in the in vitro synthesis of complementary strands by purified, recombinant TCV

RdRp. Structural changes throughout the 3' 140 nt of satC, including the Pr-2 conformation, were also present in transcripts when the RNA was engineered to inhibit 3' terminal base-pairing (Zhang et al, 2006). The proposal was made that a conformational switch converts the pre-active structure to the active structure prior to initiation of (-)-strand synthesis (Zhang et al, 2006).

The DR and the adjacent H4a are also structurally rearranged in transcripts that form Pr-2 and mutations in the DR reduced satC accumulation in vitro and in vivo (Zhang et al, 2006). I now provide evidence for a pseudoknot that stabilizes the pre-active structure and interacts with the DR. Disruption of either pseudoknot partner sequence caused reduced accumulation of satC in vivo and distinctive, nearly identical structural alterations in the pre-active structure in vitro that include specific changes in the DR region (in vitro structure probing was performed by G. Zhang). This result indicates that the pre-active structure identified in vitro has biological relevance in vivo and supports a requirement for this alternative structure and a conformational switch in high level accumulation of satC.

Materials and Methods

Construction of satC mutants

To construct plasmid $H5_{CA}$ (names of plasmids denote which hairpins were replaced in satC by the CCFV equivalent hairpins; all mutants used in this chapter are shown in Table 5.1), oligonucleotides 5'CCFV+C (all oligonucleotide sequences used in this chapter are presented in Table 5.2) and 3'CCFV+A were used as primers, and

plasmid $CH5_{CCFV}$ (Table 4.1) was used as template in a PCR. Following digestion with *Spe*I and *Eco*RV, the fragment was inserted into the analogous location in $satC_E$ (Table 2.1), which had been treated with the same restriction enzymes. Plasmids $H5_{CC}$, $H5_{AC}$, and $H5_{AA}$ were generated in a similar fashion except oligonucleotides were 5'CCFV+C and 3'CCFV+C, 5'CCFV+A and 3'CCFV+C or 5'CCFV+A and 3'CCFV+A, respectively. For construction of $H5_{5'C}$, PCR was performed with primers 5'CCFV+C and Oligo7 and template $CH5_{CCFV.}$ PCR products were digested with *Spe*I and cloned into the analogous location in pT7C+, which had been treated with *Spe*I and *Sma*I. $H5_{3'A}$ was generated by PCR using primers T7C5' and 3'CCFV+A and template $satC_E$. Following digestion with *Eco*RV and *Nco*I, the fragment was inserted into the analogous location in $satC_E$, which had been treated with the same restriction enzymes. Plasmid H4a was constructed by PCR using primers T7C5' and CCH4a and template $satC_{E.}$ Following digestion with *Spe*I and *Nco*I, the fragment was inserted into the analogous location in $satC_E$, which had been treated with the same restriction enzymes. Plasmid H4a/$H5_{CA}$ was generated in a similar fashion except that fragment was inserted into the analogous location in $H5_{CA}$. For construction of plasmid H4b, PCR was performed with oligonucleotides T7C5' and C5CC4b. PCR products were subsequently digested with *Bst*EII and *Nco*I and cloned into the analogous location in $satC_{E.}$ Plasmid H4a/H4b was generated similarly except that primer C5CC4b was replaced with CCFVH4ab and template was H4a. To construct plasmid $H5_{CA}$/H4b, primers T7C5' and CCFVH4ab5 were used with template H4b in a PCR. Following digestion with *Bst*EII and *Nco*I, the fragment was inserted into the analogous location in $H5_{CA}$ which had been treated with the same restriction enzymes. Plasmid H4a/H4b/$H5_{CA}$ was also constructed in a similar

TABLE 5.1 Summary of satC mutants used in Chapter V

Name	Description
$CH5_{CCFV}$G218C	$SatC_E$ with CCFV H5 (positions 270 to 311) and a G to C change at position 218
$H5_{CA}$	$SatC_E$ with CCFV H5 and two bases flanking natural CCFV H5 (C-A pair)
$H5_{AC}$	$SatC_E$ with CCFV H5 flanked by a A-C pair
$H5_{AA}$	$SatC_E$ with CCFV H5 flanked by a A-A pair
$H5_{CC}$	$SatC_E$ with CCFV H5 flanked by a C-C pair
$H5_{5'C}$	$SatC_E$ with CCFV H5 flanked by a 5'C
$H5_{3'A}$	$SatC_E$ with CCFV H5 flanked by a 3'A
H4a	$SatC_E$ with CCFV H4a
H4b	$SatC_E$ with CCFV H4b
H4a/H4b	$SatC_E$ with CCFV H4a and H4b
$H5_{CA}$/H4a	$SatC_E$ with CCFV H5 flanked by a C-A pair and H4a
$H5_{CA}$/H4b	$SatC_E$ with CCFV H5 flanked by a C-A pair and H4b
$H5_{CA}$/H4a/H4b	$SatC_E$ with CCFV H5 flanked by a C-A pair, H4a, and H4b
Pr	$SatC_E$ with CCFV Pr hairpin (positions 328 to 350)
Pr/$H5_{CA}$	$SatC_E$ with CCFV Pr hairpin and H5 flanked by a C-A pair
Pr/H4a	$SatC_E$ with CCFV Pr hairpin and H4a
Pr/H4b	$SatC_E$ with CCFV Pr hairpin and H4b
Pr/H4a/H4b	$SatC_E$ with CCFV Pr hairpin, H4a and H4b
Pr/$H5_{CA}$/H4a	$SatC_E$ with CCFV Pr hairpin, H5 flanked by a C-A pair, and H4a
Pr/$H5_{CA}$/H4b	$SatC_E$ with CCFV Pr hairpin, H5 flanked by a C-A pair and H4b
Pr/$H5_{CA}$/H4a/H4b	$SatC_E$ with CCFV Pr hairpin, H5 flanked by a C-A pair, H4a and H4b
G252C	SatC with a G to C change at position 252
C314G	SatC with a C to G change at position 314
G252C/C314G	SatC with a G to C change at position 252 and a C to G change at position 314
C279A/G252C	SatC with a C to A change at position 279 and a G to C change at position 252
C279A/C314G	SatC with a C to A change at position 279 and a C to G change at position 314
G218C/C279A	SatC with a G to C change at position 218 and a C to A change at position 279

TABLE 5.2 Summary of oligonucleotides used in Chapter V

Application/ construct	Name	Position[a]	Sequence[b]	Polarity[c]
Mutagenesis in satC H5	T7C5'	1-19	5'-*GTAATACGACTCACTATA*GGGAUAACUAAGGGTTTCA	+
	5'CCH5+A	256-281	5'-AACTAGTgCTCTCTAGGTAACCAACG	+
	5'CCH5+C	256-281	5'-AACTAGTGCTCTCTCGGTAACCAACG	+
	3'CCH5+C	252-298	5'-ATTG(G**A**T**A**TC)GGAGGGTCCCCAACACC	-
	3'CCH5+A	252-298	5'-ATTG(G**A**T**A**TC)GGATGGTCCCCAACACC	-
	CCH4a	206-279	5'-ggGTGGTTACCCAGAGAGCACTAGTTTTCCAGGCTA ATGCCCACTCAGGACGG ATGAGTCGCCGTTTTTGG TCCC	-
	C5CC4b	228-279	5'-ggGTGGTTACCCAGAGACACCCTAGTTTCCACACTA GGAGCCGCAGCTAGACGG	-
	CCFVH4ab	221-279	5'-ggGTGGTTACCCAGAGACACCCTAGTTTCCACACTA GGAGCCACTCAGGACGG	-
	CCFVH4ab5	257-279	5'-ggGTGGTTACCGAGAGACACCCTAG	-
	CCFV-Pr	318-356	5'-GGGCAGGCCCCCCCCTGCGCGAGGGAGGGGCTATC TATTGG	-
	G252C	238-263	5'-CACTAGTTTTCGAGGCTAATGCCCGC	-
	C314G	299-356	5' GGGCAGGCCCCCCGTCCGAGGAGGGAGGCTATCTAT TGGTTCCGAGGGTCCCCAAAGC	-
	Oligo 7	338-356	5' GGGCAGGCCCCCCGTCCGA	-
RNA gel blots	Oligo 13	249-269	5'-AGAGAGCACTAGTTTTCCAGG[d]	-
	Oligo 206	206-221	5'-CGCCGTTTTTGGTCCC 3'[e]	-

[a] Coordinates correspond to those of satC.

[b] Bases in italics indicate T7 RNA polymerase promoter sequence. Bold residues denote bases inserted to generate an *Eco*RV site (in parentheses). Bases in lowercase were added to achieve efficient digestion. Mutant bases are underlined.

[c] "+" and "-" polarities refer to homology and complementarity with satC plus-strands, respectively.

[d] Oligo 13 is also complementary to positions 3950 to 3970 of TCV genomic RNA.

[e] Oligo 206 is also complementary to positions 3900 to 3915 of TCV genomic RNA.

fashion except that template was replaced by H4a/H4b. Plasmids Pr and Pr/$H5_{CA}$ were generated by PCR using primers T7C5' and CCFV-Pr with template $satC_E$ or $H5_{CA}$, respectively. PCR products were cloned into the *Sma*I site of pUC19. Plasmids Pr/H4a, Pr/H4b, Pr/H4a/H4b, Pr/H4a/$H5_{CA}$, Pr/H4b/$H5_{CA}$, Pr/H4a/H4b/$H5_{CA}$ were constructed by digestion of plasmids H4a, H4b, H4a/H4b, H4a/$H5_{CA}$, H4b/$H5_{CA}$, H4a/H4b/$H5_{CA}$ with *Eco*RV and *Nco*I and insertion of the fragments into the analogous location in Pr/$H5_{CA}$ that had been treated with the same restriction enzymes.

To generate plasmids G252C and G252C/C279A, primers T7C5' and G252C were used with template pT7C+ in a PCR. Following digestion with *Spe*I and *Nco*I, the fragments were inserted into the analogous location in pT7C+ and C279A, which had been treated with the same restriction enzymes. Plasmids C314G and C314G/C279A were constructed by PCR using primers T7C5' and C314G with template pT7C+ or C279A, respectively. PCR products were cloned into the *Sma*I site of pUC19. Plasmids G252C/C314G and G252C/C314G/C279A were generated by PCR performed with oligonucleotides T7C5' and C314G and template pT7C+ or C279A, respectively. PCR products were treated with *Spe*I. The fragments were cloned into the analogous location in plasmid G252C that had been treated with *Spe*I and *Sma*I.

In vitro transcription, preparation and inoculation of Arabidopsis protoplasts and analysis of viral RNAs

TCV genomic RNA and satC transcripts were synthesized from plasmids containing T7 RNA polymerase promoters and linearized with *Sma*I or directly from

PCR products using T7 RNA polymerase. Preparation and inoculation of Arabidopsis protoplasts and analysis of viral RNAs were performed as described in Chapter II. For analysis of (+)-strand accumulation, the RNA was probed with a [γ-^{32}P]-ATP-labeled oligonucleotide complementary to positions 3950 to 3970 of TCV genomic RNA and positions 249 to 269 of satC (Oligo 13) or complementary to positions 3900 to 3915 of TCV genomic RNA and positions 206 to 221 of satC (Oligo 206). For analysis of (-)-strand, the RNA was probed with an [α-^{32}P]-UTP-labeled riboprobe obtained from *Dra*I-digested pT7C+ by transcription with T7 RNA polymerase (Nagy et al., 1999).

Results

H4a and H4b comprise a single functional unit that is important for satC accumulation in vivo

To determine if H4a and H4b interact with each other, or with H5 or Pr, satC constructs were generated in which one or more hairpins were replaced with the equivalent hairpins from CCFV (Figure 5.2A). The parental satRNA for these constructs, satC$_E$, contains a two base alternation in the linker sequence between H5 and Pr to aid in cloning (Table 2.1). A previous replacement of satC H5 with that of CCFV (construct CH5$_{CCFV}$) resulted in undetectable accumulation in protoplasts, although the 3' end/H5 LSL interaction remained theoretically possible (Figure 4.2, Figure 5.1C). Before making the single and multiple exchanges of the remaining 3' proximal hairpins for the current study, I needed to enhance the accumulation of CH5$_{CCFV}$ to detectable levels, so that the effect of additional replacement hairpins could be more quantitatively evaluated.

As an initial attempt to determine why satC with H5 of CCFV accumulated so poorly in protoplasts, $CH5_{CCFV}$ was inoculated onto six turnip plants along with TCV genomic RNA, and progeny at three weeks postinoculation were cloned and examined for any second site alterations that might have improved fitness. Seventy-two percent of recovered progeny (23 of 32 clones) had an identical transversion at position 218 (G218C) (Table 5.3), which is located within the DR sequence. $CH5_{CCFV}$ containing G218C accumulated to 4% of wild-type levels (compared with undetectable accumulation of $CH5_{CCFV}$), indicating that the second site alteration provided at least modest improvement (Figure 5.1B). While G218C enhanced $CH5_{CCFV}$ accumulation to detectable levels, I did not want to proceed with hairpin exchanges using constructs containing an alteration in the DR region because of its apparent role in conformational changes in vitro (Zhang et al., 2006).

Since poor accumulation of $CH5_{CCFV}$ might be in part caused by topological constraints due to the CCFV H5 lower stem missing two bases (one base-pair) compared with satC H5, the two bases flanking natural CCFV H5 (C-A) were added, producing construct $H5_{CA.}$ As shown in Fig. 3C, $H5_{CA}$ accumulated to 15% of $satC_E$, indicating that the additional non-paired residues were beneficial for satC utilization of the CCFV H5. To determine if the added bases were sequence specific, $CH5_{CCFV}$ was also constructed to contain A-C, A-A and C-C pairs flanking the CCFV H5 (constructs $H5_{AC}$, $H5_{AA}$ and $H5_{CC}$). These constructs accumulated to detectable levels that were 2 to 3-fold lower than that of $CH5_{CCFV}$CA (Figure 5.1C), suggesting a preference for the natural flanking sequence. Constructs with a single insert of either the 5' cytidylate or the 3' adenylate ($H5_{5'C}$, $H5_{CCFV}$3'A) accumulated to levels 7.5-fold lower than $CH5_{CCFV}$CA,

Figure 5.1 Base insertions and alterations improve accumulation of $CH5_{CCFV}$ in protoplasts. (A) Base differences between satC H5 and CCFV H5 are boxed. Location of residues inserted at the base of CCFV H5 in $CH5_{CCFV}$ are indicated. (B) Effect of a second site mutation on accumulation of $CH5_{CCFV}$. $CH5_{CCFV}$ was inoculated onto six turnip plants along with TCV genomic RNA. At three weeks postinoculation, 72% of recovered progeny had a single second site alteration at position 218 (G218C). Protoplasts were inoculated with $satC_E$ (satC with a new restriction site downstream of H5 required for inserting CCFV hairpins into satC), $CH5_{CCFV}$ and $CH5_{CCFV}$-G218C and levels of accumulating satRNA determined at 40 hpi. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot). Values below the lanes denote the averages of at least two replicates. (C) Accumulation of $CH5_{CCFV}$ with insert pairs of adenylates and/or cytidylates flanking the base of H5. (D) Accumulation of $CH5_{CCFV}$ with single inserts upstream or downstream of H5. These results suggest a connection between the DR region and H5, and that spatial positioning of the inserted hairpin in relation to upstream and downstream elements is important for H5 function.

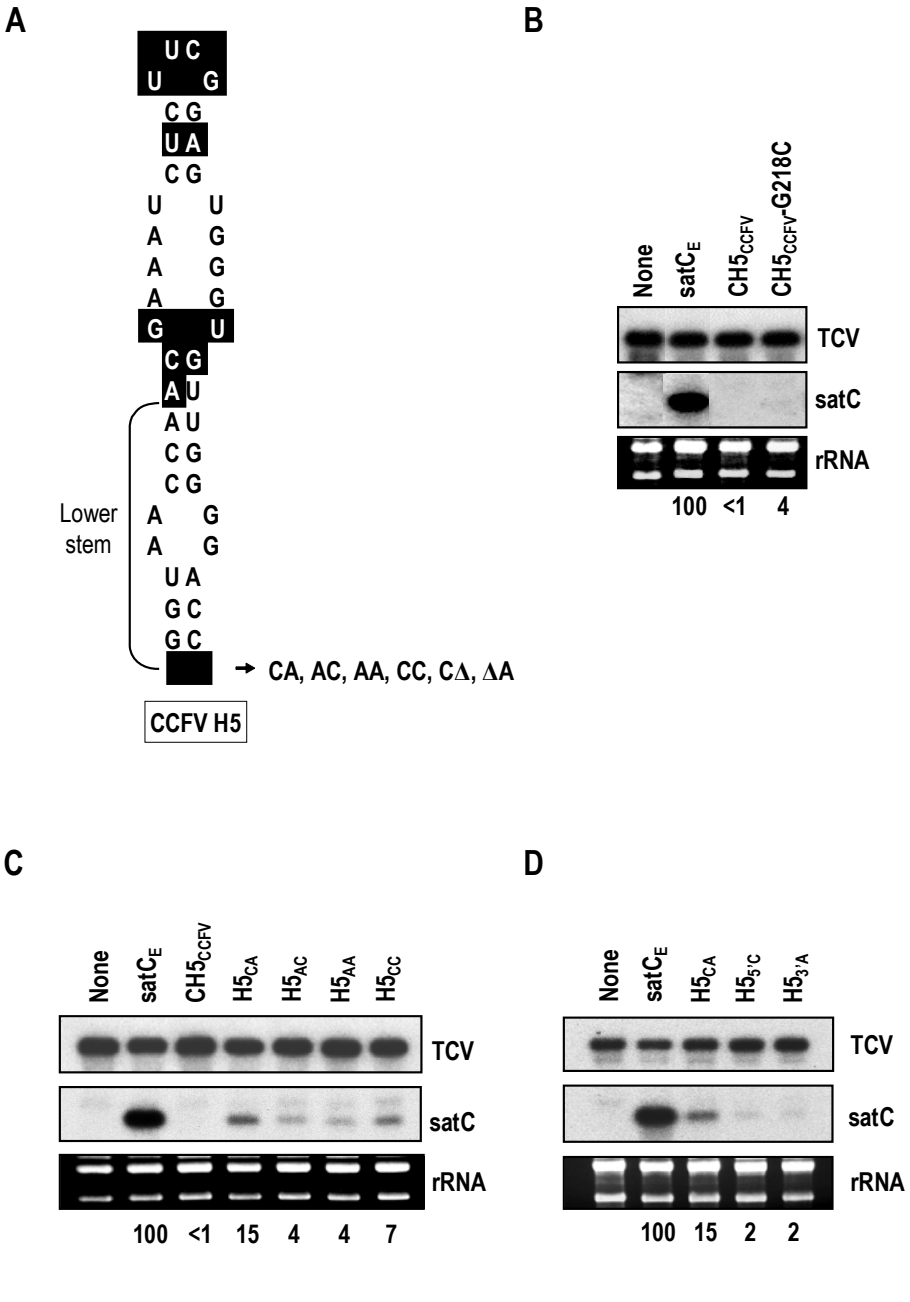
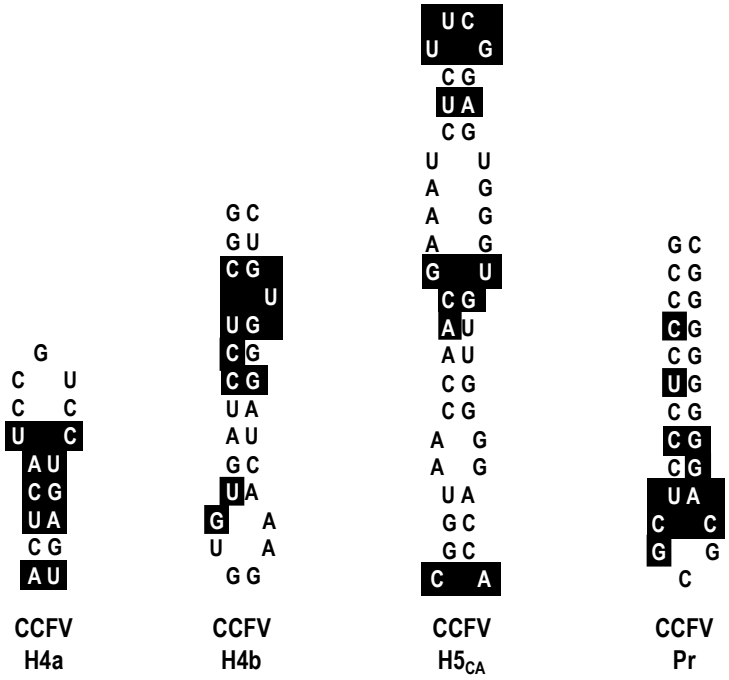
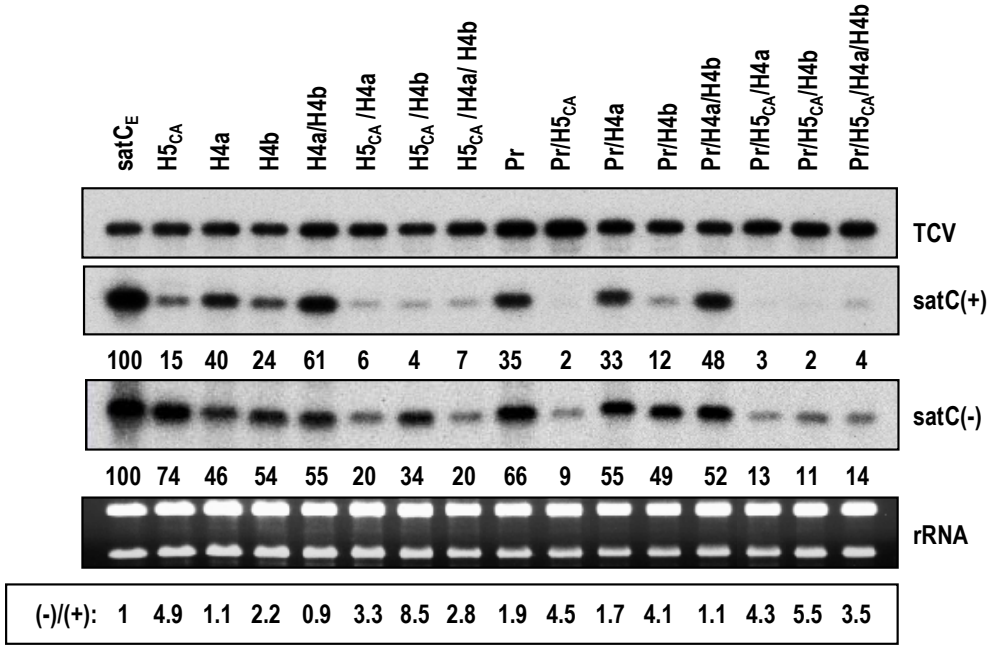

A
B
C
D
Lower stem
CA, AC, AA, CC, CΔ, ΔA
CCFV H5
None
satC$_E$
CH5$_{CCFV}$
CH5$_{CCFV}$-G218C
TCV
satC
rRNA
100 <1 4
H5$_{CA}$
H5$_{AC}$
H5$_{AA}$
H5$_{CC}$
100 <1 15 4 4 7
H5$_{5'C}$
H5$_{3'A}$
100 15 2 2

Figure 5.2 Replcements of satC hairpins with the equivalent hairpins of CCFV. (A) CCFV H4a, H4b, H5CA, and Pr. Sequences that differ with the satC equivalent hairpins are boxed. Hairpins are oriented to more easily compare with the satC hairpins in Figure 1.4B. $H5_{CA}$ includes the cytidylate and adenylate flanking CCFV H5 as addition of these residues increased $CH5_{CCFV}$ accumulation to detectable levels (see text). (B) RNA gel blot of total RNA accumulating at 40 hpi were probed with oligonucleotides specific for either the (+)- or (-)-strands. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot). Names above the lanes indicate which satC hairpins were replaced with those from CCFV. Numbers below the blots represent average levels of accumulation for at least two independent experiments. Boxed values are the ratio of (-)-strands to (+)-strands. SatC (+)-strands accumulate to levels approximately 100-fold higher than (-)-strands.

A


CCFV H4a
CCFV H4b
CCFV H5CA
CCFV Pr

B


satCE
H5CA
H4a
H4b
H4a/H4b
H5CA /H4a
H5CA /H4b
H5CA /H4a/ H4b
Pr
Pr/H5CA
Pr/H4a
Pr/H4b
Pr/H4a/H4b
Pr/H5CA/H4a
Pr/H5CA/H4b
Pr/H5CA/H4a/H4b
TCV
satC(+)
100 15 40 24 61 6 4 7 35 2 33 12 48 3 2 4
satC(-)
100 74 46 54 55 20 34 20 66 9 55 49 52 13 11 14
rRNA
(-)/(+): 1 4.9 1.1 2.2 0.9 3.3 8.5 2.8 1.9 4.5 1.7 4.1 1.1 4.3 5.5 3.5

demonstrating that inserted bases flanking both sides of the hairpin were important for enhanced accumulation. Based on these results, additional hairpin exchanges that included CCFV H5 also contained the CCFV H5 C-A flanking sequences ($H5_{CA}$).

I incorporated CCFV H4a, H4b, $H5_{CA}$, and Pr hairpin in all pair-wise combinations into their respective positions in satC and determined the effect of the heterologous hairpins on satC (+)- and (-)-strand accumulation in protoplasts (Figure 5.2). CCFV H4a and H4b reduced $satC_E$ accumulation to 40% and 24% of wild-type satC, respectively (Figure 5.2B). Replacement of both H4a and H4b with those of CCFV restored accumulation to 61% of $satC_E$. Enhanced accumulation when both CCFV hairpins are present suggests that H4a and H4b comprise a single unit and that function of the unit favors both hairpins originating from the same virus. When satC with CCFV H4a, H4b or H4a/H4b replacements also contained CCFV $H5_{CA}$ or $H5_{CA}$ and Pr hairpin, all mutants accumulated to similar low levels (2-7% of $satC_E$; Figure 5.2B). This indicated that poor satC accumulation due to a heterologous H5 was not improved by the presence of additional CCFV 3' hairpins.

SatC with Pr hairpin of CCFV accumulated to 35% of $satC_E$ levels (Figure 5.2B), and this level was reduced an additional 17.5-fold when CCFV $H5_{CA}$ was also replaced. However, when both Pr hairpin and H4a were replaced, satC accumulated to 33% of $satC_E$, indicating that the negative effects of CCFV Pr hairpin and H4a on satC accumulation were not additive. In contrast, CCFV H4b reduced the accumulation of satC with the CCFV Pr hairpin to 12% of $satC_E$. The poor accumulation of satC with CCFV Pr hairpin/H4b was enhanced 4-fold to 48% of $satC_E$ when H4a also originated from CCFV.

I previously determined that mutations in H5 reduced accumulation of satC (+)-strands more than (-)-strands [when compared with wild-type levels of (+)- and (-)-strands], despite H5 being a (+)-strand element (Figure 4.8). This was suggested to reflect a secondary function of H5 as a scaffold for proper assembly of the RdRp in vivo (McCormack and Simon, 2004). An assembly role was also suggested for the equivalent hairpin (SL3) in viruses from the *Tombusvirus* genus (Panaviene et al, 2005). Enhanced reduction in (+)-strand synthesis due to TCV H5 mutations was attributed to an incorrectly assembled RdRp, which might be exhibiting differential impairment in transcription of (+)- and (-)-strands (Zhang and Simon, 2005; Figure 4.8). When (-)- and (+)-strand levels of the hairpin replacement mutants were compared with those of $satC_E$, the ratio of (-)-strands to (+)-strands was higher than $satC_E$ for nearly all constructs (Figure 5.3B). The ratio was substantially higher for replacements of $H5_{CA}$ (4.9-fold), $H5_{CA}$/H4b (8.5-fold), Pr/$H5_{CA}$ (4.5-fold) and Pr/$H5_{CA}$/H4b (5.5-fold). In contrast, replacement of H4a, H4a/H4b, and Pr/H4a/H4b had near $satC_E$ ratios of (-)- and (+)-strands. These results suggest that CCFV H4b, in the absence of CCFV H4a, is responsible for a similar impairment to satC accumulation as previously found for H5.

H4b sequence forms a pseudoknot with sequence flanking the 3' side of H5

Since alterations to H4b were more detrimental to satC accumulation than similar alterations to H4a, it seemed likely that the H4b portion of the H4a/H4b unit contained an element important for robust RNA replication. Previous results indicated that the conformation of the 3' region of satC contains substantial tertiary structure based on most guanylates in the region losing susceptibility to the single-stranded specific ribonuclease

RNase T1 in the presence of Mg^{2+}, which stabilizes RNA tertiary structure (Woodson, 2005). Six guanylates in the H4a/H4b region had conformations that were altered by Mg^{2+} including G252 and G253 in the H4b loop (Zhang et al., 2006). Examination of the 3' region of satC for sequences that might interact with the H4b loop revealed a possible candidate at positions 312-315 (UCCG), which flanks the 3' side of H5 and could possibly pair with H4b positions 251-254 (UGGA) (Figure 5.3A). In vivo SELEX of positions 312-327 revealed that all satC recovered in the initially inoculated plants contained sequence between positions 312-316 that could maintain pairing with at least the UGG in the H4b loop and up to two additional flanking bases (Performed by R. Guo; Zhang et al., manuscript in revision). Comparative analysis of other carmoviruses also revealed nearly universal conservation of UGG in their H4b loops (7 of 11; Figure 1.4, Figure 1.5), and the existence of possible pairing partners flanking the 3' side of their respective H5 (Figure 1.5). This potential interaction is also observed in carmoviruses that do not contain the conserved UGG (Figure 1.5).

To test for a possible pseudoknot interaction between the H4b UGGA and the UCCG flanking H5, single alterations were constructed at position 252 (G252C) and its proposed base-pair partner, position 314 (C314G) (Figure 5.3A). SatC-G252C accumulated to 35% of wild-type levels in protoplasts (Figure 5.3B), indicating that the residue is important for normal satC function. SatC-C314G accumulated to only 2% of wild-type satC, confirming the importance of the sequence established by in vivo SELEX. SatC-G252C/C314G accumulated to 56% of wild-type satC, strongly suggesting that re-established base-pairing between the two elements is responsible for partially restoring satC accumulation. The differing levels of satC caused by G252C and C314G

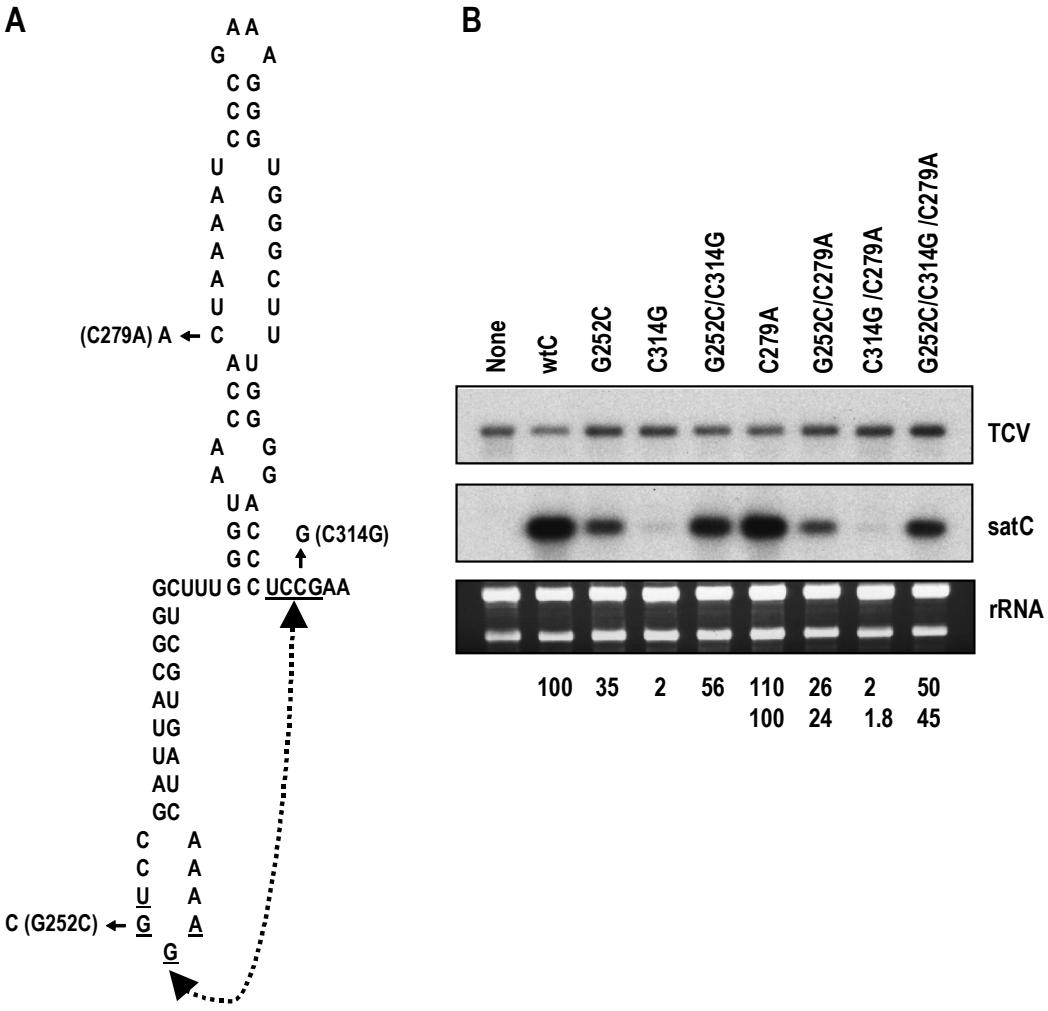


Figure 5.3 Interaction between positions 251-254 and 312-315 is important for satC accumulation in vivo. (A) Phylogenetically inferred structure of the H4b/H5 region. Potential base-pairing between underlined residues is denoted by connected arrowheads. Mutations generated in putative base-paired partners are shown. Location of alteration C279A, which stabilizes H5 by pairing the lower positions in the LSL, is shown. While the interaction is shown in the active form of the 3' region for convenience, this does not imply that the interaction exists in this structural configuration. (B) Accumulation of satC with various mutations indicated above each lane recovered from protoplasts at 40 hpi. Mutations are described in (A). None, no satC added. wtC, wild-type satC. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot). Numbers below the blots are average levels of accumulation for at least three independent experiments. Upper values reflect levels in relation to wild-type satC while lower values are recalculated for ease in comparison with parental C279A, which accumulates 10% better than wild-type satC.

also suggest that the region encompassing 314 has an additional role in satC accumulation beyond simple participation in an interaction with H4b. This new pseudoknot between H4b UGGA and the UCCG flanking H5 has been termed Ψ_2.

While the genetic and phylogenetic evidence support the existence of Ψ_2, this interaction could be topologically incompatible with the long H4b stem and H5 structure. However, previous solution structure probing suggested that the 3' 140 nt of wild-type satC adopts an initial "pre-active" conformation in vitro that does not appear to contain the phylogenetically conserved hairpin structures (Zhang et al., 2004, 2006). In vitro solution structure analysis revealed similar or identical cleavage differences in three regions (UCCG element, DR region and 3' side of H5) of G252C and C314G, which contain disruptions in one or the other Ψ_2 interacting sequences (Performed by G. Zhang; Zhang et al., manuscript in revision). Nearly identical new single-stranded specific cleavages in the UCCG element of G252C and C314G suggested that they result from disruption of Ψ_2 in the satC pre-active structure. In support of this conclusion, G252/G253 were also consistently cleaved more strongly by RNase T1 in C314G than in wild-type satC, suggesting that these residues have also adopted a more single-stranded conformation (Performed by G. Zhang; Zhang et al., manuscript in revision).

Ψ_2 stabilizes the pre-active satC structure

In vitro RdRp transcription assays showed that transcription of G252C and C314G by purified recombinant TCV RdRp (p88) generated 37% and 51% more complementary products, respectively, than produced using wild-type satC (Performed by G. Zhang; Zhang et al., manuscript in revision). This result suggested that disrupting

Ψ_2 reduces the stability of the pre-active structure, thus lowering the activation energy between conversion of the two structures and shifting the equilibrium towards the active structure. A more stable active structure, which might be detrimental for cyclic satC replication in vivo, should enhance transcription in vitro as this assay only reports on initiation of (-)-strand synthesis (products are double-stranded and not templates for further transcription). G252C/C314G had reduced transcriptional activity with an average of 84% of wild-type satC levels, which suggested reformation of Ψ_2 had occurred in the compensatory mutant (Performed by G. Zhang; Zhang et al., manuscript submitted).

Altogether, these results indicate that disrupting Ψ_2 is detrimental to satC accumulation in vivo while increasing transcription of satC in vitro. This suggests that a pre-active structure stabilized by Ψ_2 may be a necessary feature for robust satC accumulation in vivo, which must produce asymmetric levels of (+)- and (-)-strands in the proper cellular location. If this hypothesis is correct, then alterations that stabilize the active structure of satC should further reduce mutant accumulation in vivo while enhancing complementary strand synthesis in vitro. SatC with C279A, which closes the lowest position in the H5 LSL and is predicted to slightly stabilize the H5 stem (Figure 5.3A), accumulated 10% better than wild-type satC in vivo, suggesting that the mutation may be modestly shifting the equilibrium towards the active structure in a non-detrimental fashion [Figure 3.2, Figure 5.3B; a similar enhancement in accumulation (15%) was also found when satC contains the more stable H5 of TCV, Figure 4.10B]. When combined with C279A, G252C, C314G and G252C/C314G accumulation in protoplasts decreased between 20 and 30%. Transcription of C279A by p88 in vitro generated 20% more complementary strands than wild-type satC, supporting a stabilized

active structure for this mutant. When C279A was combined with G252C (C279A/G252C), transcription in vitro was enhanced an additional 30% (Performed by G. Zhang; Zhang et al., manuscript in revision). These results support a requirement for a pre-active structure that is stabilized by Ψ_2 for robust accumulation of satC in vivo.

Possible interaction of the DR with Ψ_2

Solution structure probing revealed strong cleavages in the DR region that accompany G252C or C314G alterations (Performed by G. Zhang; Zhang et al., manuscript in revision), suggesting that the DR may be structurally associated at or near the UGGA/UCCG interacting region in the pre-active structure. An association between the DR and H5 was also indicated by finding that the majority of progeny accumulating in plants inoculated with $CH5_{CCFV}$ had a second site alteration at position 218 (Figure 5.1B, Table 5.3). While this DR mutation enhanced accumulation of $CH5_{CCFV}$, G218C reduced accumulation of wild-type satC by 72% in protoplasts (Figure 5.1B) and transcription in vitro to undetectable levels (Zhang et al., 2006), leading to the suggestion that mutations in the DR inhibited the conformational switch.

If the DR region helps to promote the conformational changes needed to activate satC transcripts, then stabilizing the active structure of satC should reduce the negative effects of G218C. To test this possibility, satC containing C279A, with and without alteration G218C, were assayed for accumulation in protoplasts. At 40 hpi, satC with G218C accumulated to 28% of wild-type. Inclusion of the C279A mutation reduced the negative effect of the DR mutation by 2-fold, with the satRNA now accumulating to 57%

TABLE 5.3 Summary of clones recovered from turnip plants inoculated with $CH5_{CCFV}$

	Plant						Total
	1	2	3	4	5	6	
$CH5_{CCFV}$	0	4	0	0	3	1	8
$CH5_{CCFV}$-G218C	8	6	1	2	2	4	23
Other*		1					1
Total	8	11	1	2	5	5	32

* A recombinant (named K23) contains satC sequence (positions 1 to 192) at the 5' end and TCV sequence (positions 3896 to 4054) at the 3' end. The original recovered clone (named C23) contains satC sequence (positions 338 to 356) instead of TCV sequence at the 3' end because Oligo 7, which is complementary to the positions 338 to 356 of satC sequence, was used in RT-PCR. K23 was obtained by PCR using primers T7C5' and KK57 (complementary to the positions 4036 to 4054 of TCV sequence) and template C23. PCR products were cloned into the *Sma*I site in pUC19.

of the C279A level (Figure 5.4). This result supported a role for the DR region in the conformational changes leading to satC transcription initiation.

Solution structure analysis of G218C indicated structural changes throughout the H4a and H4b regions in vitro (Performed by G. Zhang; Zhang et al., manuscript in revision). Interestingly, G218C transcripts also contained two new RNase V1 cleavages at C314 and G315 in the Ψ_2 UCCG sequence, suggesting that mutating the DR region affects the structure of Ψ_2. The G218C mutation had no effect on the structure downstream of position G315, including the Pr region (Zhang et al., 2006). Thus it is possible that the DR promotes conformational changes through a role in allowing or restricting formation of the important pseudoknot that forms between the H4b loop sequence and the sequence flanking the 3' side of H5.

Discussion

My colleagues and I previously proposed that initiation of satC (-)-strand synthesis in vitro requires two mutually exclusive structures, an initial pre-active structure that contains extensive tertiary interactions and an active structure that contains the phylogenetically inferred hairpins and a required pseudoknot (Ψ_1) between the 3' end and H5 (Zhang et al., 2006). The existence of the pre-active structure was based solely on in vitro solution structure probing, which indicated that satC templates correctly recognized by the TCV RdRp did not appear to contain any of the phylogenetically inferred hairpins. The structure assumed by the transcripts persisted despite a heat/cool

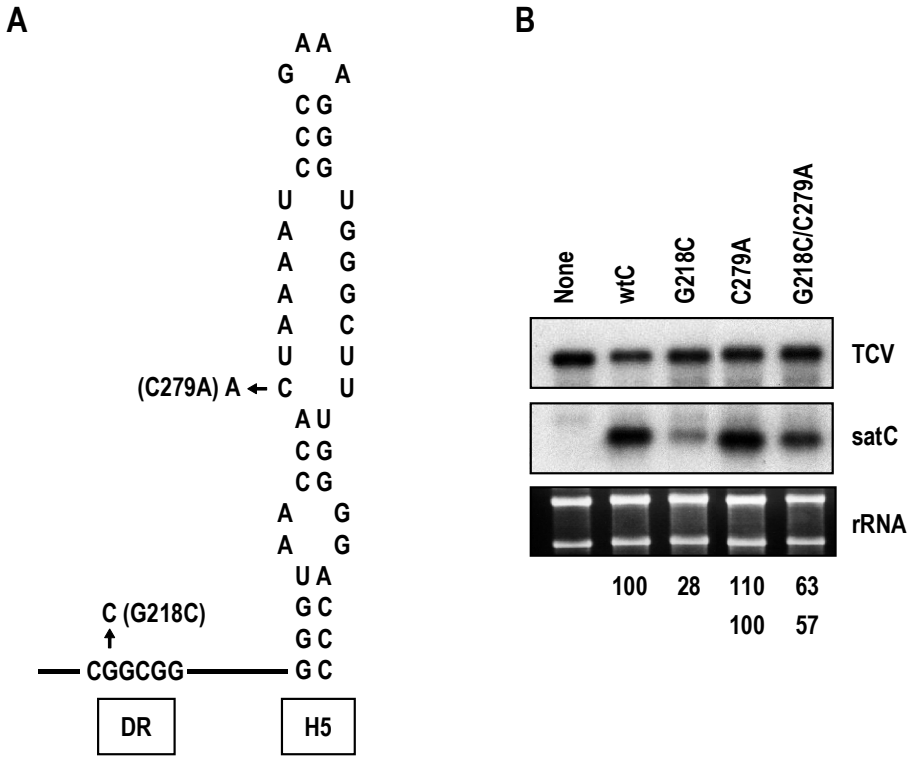


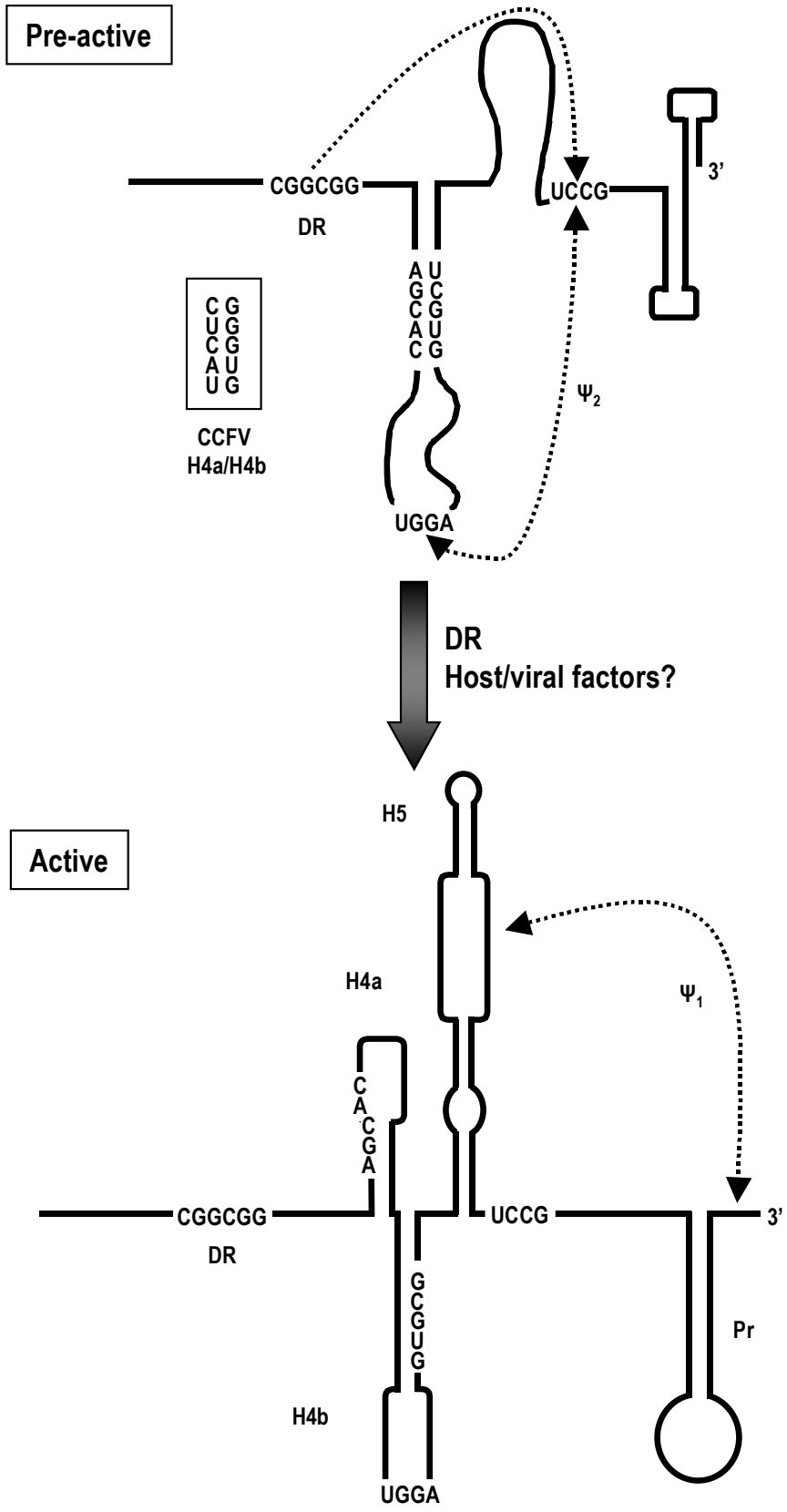
Figure 5.4 Importance of the DR in the satC conformational switch. (A) Sequence and structure of H5 and DR is shown. Solid lines represent satC sequences. The positions of mutation are indicated. (B) RNA gel blot of total RNA isolated at 40 hpi of protoplasts with transcripts of TCV genomic RNA and satC variants. Ethidium bromide staining of the gel before blotting shows rRNA loading control (panel below the blot). Values below each lane are the averages of at least two independent experiments. Since C279A levels were greater than wild-type satC, accumulation of the double mutant (C279A/G218C) is compared with the C279A parental RNA (lower set of values). None, no satC added. wtC, wild-type satC.

treatment, which is routinely used to reduce or eliminate kinetically trapped intermediates (Pan et al., 1997).

My current results provide important evidence that the pre-active structure is biologically relevant, and thus a conformational switch is likely required to initiate satC (-)-strand synthesis (Figure 5.5). This evidence centers on a newly discovered pseudoknot, Ψ_2, which forms by pairing positions $_{251}$UGGA$_{254}$ in the loop of H4b with $_{312}$UCCG$_{315}$, which flanks the 3' side of H5. Evidence for the pseudoknot includes: (i) compensatory alterations between positions 252 and 314 that reestablish pairing enhance accumulation of satC in vivo compared with satC containing the individual mutations (Figure 5.3); (ii) progeny accumulating in plants inoculated with C314G contain new alterations at either the same site (uridylate) or at a secondary site (position 252- G to C), each of which reforms the pseudoknot (Performed by R. Guo; Zhang et al., manuscript in revision); (iii) in vivo SELEX of satC with randomized linker sequences produces progeny that all have between three and five residues at the base of H5 capable of pairing with H4b loop sequence including the phylogenetically conserved "UGG" (Performed by R. Guo; Zhang et al., manuscript in revision); and (iv) satC containing mutations at either position 252 or 314 have identical structural changes in vitro (Performed by G. Zhang; Zhang et al., manuscript in revision). In addition, compensatory mutations that re-establish pairing at either of two positions in TCV Ψ_2 partner sequences also enhance accumulation of the genomic RNA compared with TCV containing the individual mutations (J.C. McCormack and A.E. Simon, unpublished).

Results from the in vitro RdRp transcription assays and solution structure probing indicate that Ψ_2 is present in and stabilizes the pre-active structure (Performed by G.

Figure 5.5 Model for a structural switch in the 3' region of satC that activates the template for (-)-strand synthesis. The complete pre-active structure of satC is not yet known. The pairing interaction between UGGA and UCCG described in this report is tentatively described as a pseudoknot (Ψ_2), however this designation may change when the complete pre-active structure is known. Possible pairing between sequences in H4a and H4b is shown, which can also form from sequences in the identical locations in CCFV H4a and H4b (boxed at left). The DR is proposed to interact in the pseudoknot region, reducing the stability of Ψ_2. It is likely that host or viral factors, such as the viral RdRp, are required to promote the conformational switch. See text for more details.


Pre-active
CGGCGG
DR
UCCG
3'
AGCAC
UCGUG
CUCAU
GGGUG
CCFV
H4a/H4b
Ψ2
UGGA
DR
Host/viral factors?
Active
H5
H4a
Ψ1
CACGA
CGGCGG
DR
UCCG
3'
GCGUG
Pr
H4b
UGGA

Zhang; Zhang et al., manuscript in revision). This interpretation is consistent with the previous finding that release of the 3' end, which led to structural changes throughout the 3' 140 bases in vitro (and is equated with formation of the active structure), resulted in new RNase A cleavage sites in the UCCG element (Zhang et al., 2006). Since disruption of Ψ_2 did not lead to the same structural alterations in the DR/H4a, H5 and Pr regions as release of the 3' end, disruption of Ψ_2 is likely necessary, but not sufficient, for conversion to the active structure.

Results from solution structure probing suggest a structural relationship may exist between DR and Ψ_2 (Performed by G. Zhang; Zhang et al., manuscript in revision). Mutations in either Ψ_2 partner sequences caused identical strong RNase T1 cleavages in the DR sequence. In addition, G218C transcripts contained new RNase V1 sites in the Ψ_2 UCCG sequence. While it is possible that the G218C mutation is stabilizing Ψ_2 and thus the pre-active structure, it is also possible that the structural changes in the DR/H4a/H4b regions caused by the G218C mutation may be the primary basis of conformational switch inhibition, which could explain reduced template activity of G218C transcripts in in vitro RdRp transcription assays. Either explanation is consistent with my finding that G218G together with H5 mutation C279A (which stabilizes H5 and thus the active structure), accumulated 2-fold better than G218C in vivo (Figure 5.3).

The model (Figure 5.5) also suggests that the pre-active structure contains an interaction between H4a and H4b sequences. This is based on finding that H4a and H4b are more functional when originating from the same virus source (Figure 5.2). The structural similarity of CCFV and satC H4a and presence of similar loop sequences suggest that the function of H4a is not merely to act as a scaffold for co-helical stacking

of H4b. Although the CCFV H4b loop contains the UGGA necessary for formation of Ψ_2, poor accumulation of satC with CCFV H4b suggests that maintaining the H4a/H4b interaction may be a necessary requirement for Ψ_2 formation.

In addition to a connection with H4b, H4a may have a separate association with the conformational switch of the core promoter from Pr-1 to Pr-2. This is suggested by finding that accumulation of satC with Pr hairpin of CCFV is not further depressed by co-replacement of H4a (Figure 5.2B). With 10 consecutive base-pairs, 8 of which are GC or CG pairings, the CCFV Pr hairpin may be limited to a single conformation, and thus H4a would not be required to support any Pr conformational switch. An association between H4b/H4a and Pr could explain why all satC mutants with the active Pr-2 structure also contain distinctive structural changes in the H4a region compared with wild-type satC (Zhang et al., 2006).

If satC (+)-strands transcribed in vivo fold into the pre-active structure, an important question is how satC converts to its active form for initiation of (-)-strand synthesis in newly infected cells. Depending on the energy barrier between the two structures, the switch could be self-induced (Nagel and Pleij, 2002.) or require a trans-acting element. In AMV, CP binding to specific 3' end sequences may compact the genomic RNA into its pre-active structure, which would require a conformational switch to an active form consisting of a series of short hairpins (Olsthoorn et al., 1999; Petrillo et al., 2005). Structural rearrangement of the HIV-1 leader RNA to a thermodynamically less stable, dimerization-competent form in vitro is mediated by the viral nucleocapsid protein that is thought to stabilize the branched, multi-hairpin structure (Huthoff and Berkhout, 2001). However, structure probing of the RNA in infected cells and virus

particles did not confirm the dimerization-incompetent structure, suggesting the RNA already exists in dimeric form even while nuclear localized (Paillart et al., 2004). If a trans-acting factor is required to mediate the satC structural switch, the most probable candidate is the viral RdRp, which could also promote an analogous switch in the genomic RNA between translation-competent and replication-competent forms.

In summary, my results suggest that a pre-active structure is a necessary component for robust accumulation of a nontranslated viral RNA in vivo. Formation of a pre-active structure by newly synthesized satC (+)-strands could allow satellite progeny to keep their 3' ends hidden and promoters unavailable to the RdRp. Restricting access of RdRp to progeny (+)-strands would promote the continued transcription of initially infecting (+)-strands into (-)-strands and the continued synthesis of progeny (+)-strands from the (-)-strand intermediates. This process would allow progeny to be "stamped" off of the original parental genome, reducing the amount of potentially deleterious mutations (Chao et al., 2002). Recent evidence for alternative RNA structures in the 5' and 3' regions of plant, animal and bacterial RNA viruses (Goebel et al., 2004; Huthoff and Berkhout, 2001; Koev et al., 2002; Olsthoorn et al., 1999; Schuppli et al., 2000) suggest that conformational switches leading to initiation of (-)-strand synthesis and possibly restricting RdRp access to de novo synthesized (+)-strands may be a contributing factor to the replication of a number of (+)-strand RNA viruses.

CHAPTER VI

EVOLUTION OF VIRUS-DERIVED SEQUENCES FOR HIGH LEVEL REPLICATION OF A SUBVIRAL RNA

Introduction

Replication of genomic and subviral RNAs requires specific interactions between the replicase complex and RNA cis-acting elements. While core promoters located near the 3' ends of (+)- and (-)-strands can independently recruit replication complexes resulting in low levels of de novo synthesized complementary strands (Dreher, 1999), additional elements throughout viral genomes aid in enzyme complex assembly or enhance or repress transcription (e.g., French and Ahlquist, 1987; Frolov et al., 2001; Herold and Andino, 2001; Khromykh et al., 2001; Klovins and van Duin, 1999; McCormack and Simon, 2004; Monkewich et al., 2005; Nagashima et al., 2005; Nagy et al., 1999; Panavas and Nagy, 2003; Panaviene et al., 2005; Ray and White, 2003; Vlot and Bol, 2003; Zhang and Simon, 2003b). In addition, recent evidence suggests that RNA conformational rearrangements play key roles in coordinating translation and replication, regulating subgenomic RNA synthesis, or producing asymmetric levels of (+)- and (-)-strands by masking or exposing elements required for a particular process (Barry et al., 2002; Isken et al., 2004; Khromykh et al., 2001; Olsthoorn et al., 1999; Koev et al., 2002; Na and White, 2006; Pogany et al., 2003; van den Born et al., 2005; Zhang et al., 2004, 2006). For example, conformational changes at the 3’ ends of BYDV

(Koev et al., 2002) and TBSV (Na and White, 2006; Pogany et al., 2003) may control accessibility of the RdRp to the initiation site for (-)-strand synthesis.

The complexities inherent in RNA virus replication has led to the use of untranslated subviral RNAs such as DI RNAs or satRNAs, as models for their larger, multifaceted helper viral genomic RNAs. Many (+)-strand RNA viruses are naturally associated with subviral RNAs, which depend on their helper virus for replication and host trafficking components (David et al., 1992; White and Morris, 1999; Simon et al., 2004). While DI RNAs are mainly derived from 5' and 3' portions of viral genomic RNAs, most satRNAs share little consecutive sequence similarity with their helper virus genomes and may have arisen from an unrelated RNA or a series of recombination events joining short segments of viral and non-viral RNAs that further evolved into a functional molecule (Carpenter and Simon, 1996). As described in Chapter I, satC has features of both DI and satRNAs with its 5' 190 bases originating from nearly full-length TCV satRNA satD (194 bases), and its 3' 166 bases derived from two regions at the 3' end of TCV genomic RNA (Figure 1.3B; Simon and Howell, 1986). TCV is also naturally associated with DI RNAs, such as diG, whose sequence is mainly derived from 5' and 3' regions of the genomic RNA (Figure 1.3B; Li et al., 1989).

The 3' terminal 100 bases shared by TCV and satC differ at only eight positions ("positions" refers to locations with one or more consecutive base differences) and are predicted to be structurally similar by *mfold* (Figure 6.2A; Zhang et al., 2004; Zucker, 2003). This observation suggested that satC would be a good model for determining the function of TCV cis-acting sequences within this region in the replication process. A combination of in vivo studies using Arabidopsis protoplasts, in vitro assays for

transcription initiation using purified recombinant TCV RdRp, and in vitro RNA solution structure probing revealed that this region in satC assumes two very different RNA conformations: an unresolved pre-active conformation stabilized by extensive tertiary structure that includes pseudoknot 2 (Ψ_2) (Figure 5.4) (Zhang et al., 2006; Zhang et al., manuscript in revision); and an active conformation that includes pseudoknot 1 (Ψ_1) and four phylogenically conserved hairpins (Figure 1.4, Figure 1.5; Zhang et al., 2006). These four hairpins have been designated as (from 3' to 5'): (i) Pr hairpin, which is a portion of the core promoter of satC (Song and Simon, 1995) and TCV (Sun and Simon, manuscript in revision) for synthesis of (-)-strands; (ii) H5, which contains a large symmetrical internal loop (LSL) that interacts with 3' terminal bases to form Ψ_1 in satC (Zhang et al., 2004) and TCV (Performed by J. C. McCormack; Zhang et al., manuscript submitted) and is proposed to help organize the replication complex in TCV (McCormack and Simon, 2004); (iii) H4b, which contains terminal loop sequences that forms Ψ_2 with sequence flanking the 3' base of H5 (Chapter V); and (iv) H4a, which forms a single unit with H4b in satC and is flanked by a short GC-rich element (DR) that is proposed to help mediate the conformational switch (Zhang et al., 2004; Zhang et al., manuscript in revision). As described in previous chapters, all four hairpins are required for satC replication in vivo.

Despite sequence and structural similarities between satC and TCV, satC with the 3' terminal 100 bases of TCV (renamed C3'100_T) accumulated very poorly in plants and protoplasts (Wang and Simon, 2000). This result suggested that one or more of the 3' cis-acting elements that are necessary for efficient amplification of TCV (J.C. McCormack and A.E. Simon, unpublished) function poorly when associated with satC sequence or

when separated from the remainder of the viral genome. To gain an understanding of how the 3' region of TCV has evolved in the context of satC to allow for high level satRNA accumulation, satC and TCV 3' terminal sequences were converted in a step-wise fashion into the counterpart's sequence, which revealed the importance of having the cognate Pr. In addition, TCV Pr is a much weaker core promoter than the satC Pr in vitro, even though structural analyses suggested that the TCV Pr assumes a form similar to the highly active satC Pr-2* (Performed by G. Zhang; Zhang et al., in press). These results suggest that the TCV Pr requires additional elements upstream of the region shared between satC and TCV to function optimally in vivo and vitro.

Materials and Methods

Construction of satC, diG and TCV mutants

To construct plasmid $C7_T$ (all mutants used in this chapter are presented in Table 6.1), C268U and Oligo 7 were used as primers (a list of primers used in this study is presented in Table 6.2) and pT7C+ was used as template in a PCR. Following digestion with *Spe*I and *Sma*I, the fragment was inserted into the analogous location in pT7C+, which had been treated with the same restriction enzymes. $C127_T$ was constructed in a similar fashion except using $CH5_{TCV}$ (Table 4.1) as template. Plasmid $C8_T$ was generated by PCR using pT7C+ as template and T7C5' and U335C as primers. PCR products were cloned into the *Sma*I of pUC19. For construction of plasmids $C34_T$, $C347_T$ and $C1234_T$, PCR was performed with primers T7C5' and C34C* and template pT7C+, $C7_T$ or $CH5_{TCV}$, respectively. PCR products were cloned into the *Sma*I of pUC19. To generate

TABLE 6.1 Summary of satC and TCV mutants used in Chapter VI

Name	Description
$C7_T$	SatC with TCV sequence at position 7 (C268U)
$C8_T$	SatC with TCV sequence at position 8 (U335C)
$C3'100_T$	SatC with the 3' end 100 nt (residues 257 to 356) replaced with that from TCV (104 nt)
$C12_T$ ($CH5_{TCV}$)	SatC with TCV sequences at positions 1 (U302G) and 2 (G306U)
$C1234_T$	SatC with TCV sequences at positions 1 (U302G), 2 (G306U), 3 (C319U) and 4 (U322A)
$C1256_T$	SatC with TCV sequences at positions 1 (U302G), 2 (G306U), 5 (a 4 nt [CGCG] insertion after residue 338) and 6 (A342G and C343G)
$C127_T$	SatC with TCV sequences at positions 1 (U302G), 2 (G306U) and 7 (C268U)
$C34_T$	SatC with TCV sequences at positions 3 (C319U) and 4 (U322A)
$C3456_T$	SatC with TCV sequences at positions 3 (C319U) and 4 (U322A), 5 (a 4 nt [CGCG] insertion after residue 338) and 6 (A342G and C343G)
$C347_T$	SatC with TCV sequences at positions 3 (C319U) and 4 (U322A) and 7 (C268U)
$C56_T$	SatC with TCV sequences at positions 5 (a 4 nt [CGCG] insertion after residue 338) and 6 (A342G and C343G)
$C568_T$	SatC with TCV sequences at positions 5 (a 4 nt [CGCG] insertion after residue 338), 6 (A342G and C343G) and 8 (U335C)
$C\text{-}Pr_G$	SatC with diG Pr (a 2 nt [CG] insertion after residue 338, A342G and C343G)
$G\text{-}Pr_C$	DiG with satC Pr (a deletion of residues 335 and 336, G339A, and G340C)
$G\text{-}Pr_T$	DiG with TCV sequence at position 5 (a 2 nt [CG] insertion after residue 335)
$T3'100_C$	TCV with the 3' end 104 nt (residues 3951 to 4054) replaced with that from satC (100 nt)
$T12_C$	TCV with satC sequences at positions 1 (G3996U) and 2 (U4000G)
$T1256_C$	TCV with satC sequences at positions 1 (G3996U), 2 (U4000G), 5 (a deletion of residues 4033 to 4036) and 6 (G4040A and G4041C)
$T56_C$	TCV with satC sequences at positions 5 (a deletion of residues 4033 to 4036) and 6 (G4040A and G4041C)
$T\text{-}Pr_G$	TCV with diG sequence at position 5 (a deletion of residues 4033 and 4044)
$T345678_C$	TCV with satC sequences at positions 3 (U4013C), 4 (A4016U), 5 (a deletion of residues 4033 to 4036), 6 (G4040A and G4041C), 7 (U3962C) and 8 (C4029U).
$T5678_C$	TCV with satC sequences at positions 5 (a deletion of residues 4033 to 4036), 6 (G4040A and G4041C), 7 (U3962C) and 8 (C4029U).
$T3478_C$	TCV with satC sequences at positions 3 (U4013C), 4 (A4016U), 7 (U3962C) and 8 (C4029U).
$T34568_C$	TCV with satC sequences at positions 3 (U4013C), 4 (A4016U), 5 (a deletion of residues 4033 to 4036), 6 (G4040A and G4041C) and 8 (C4029U).
$T125678_C$	TCV with satC sequences at positions 1 (G3996U), 2 (U4000G), 5 (a deletion of residues 4033 to 4036), 6 (G4040A and G4041C), 7 (U3962C) and 8 (C4029U).
$T1278_C$	TCV with satC sequences at positions 1 (G3996U), 2 (U4000G), 7 (U3962C) and 8 (C4029U).
$T12568_C$	TCV with satC sequences at positions 1 (G3996U), 2 (U4000G), 5 (a deletion of residues 4033 to 4036), 6 (G4040A and G4041C) and 8 (C4029U).
$T123478_C$	TCV with satC sequences at positions 1 (G3996U), 2 (U4000G), 3 (U4013C), 4 (A4016U), 7 (U3962C) and 8 (C4029U).
$T12347_C$	TCV with satC sequences at positions 1 (G3996U), 2 (U4000G), 3 (U4013C), 4 (A4016U), and 7 (U3962C).

TABLE 6.1 continued

Name	Description
CPm-T	TCV with mutations at the primary (U2745C) and secondary (C2738U) CP translation initiation codons (Wang and Simon, 1997)
TCVs	TCV with a new EcoRV site in the TCV H5/Pr linker (a 3 nt [CGT] insertion after residue 4014)
C-Pr	MDV with extra 37 nt from satC (residues 320 to 356) at the 3' end
G-Pr	MDV with extra 39 nt from diG (residues 315 to 353) at the 3' end
T-Pr	MDV with extra 41 nt from TCV (residues 4014 to 4054) at the 3' end

TABLE 6.2 Summary of the oligonucleotides used in this study

Application/ construct	Name	Position[a]	Sequence[b]	Polarity[c]
Mutagenesis in satC	T7C5'	1-19	5'-*GTAATACGACTCACTATA*GGGAUAACUAAGGGTTTCA	+
	Oligo 7	338-356	5'-GGGCAGGCCCCCCGTCCGA	-
	C268U	253-279	5'-gaaaACTAGTGCTCTTTGGGTAACCAC	+
	U335C	319-356	5'-GGGCAGGCCCCCCGTCCGAGGGGGGAGGCTATCTA TTG	-
	C34C*	307-356	5'-GGGCAGGCCCCCCGTCCGAGGAGGGAGGCTATCTT TTAGTTCGGAGGGTC	-
	C56C*	324-356	5'-GGGCAGGCCCCCCCCCCG**CGCG**AGGAGGGAGGCTA TC	-
	C568C*	318-356	5'-GGGCAGGCCCCCCCCCCG**CGCG**AGGGGGGAGGCTA TCTATTGG	-
Mutagenesis in TCV	Oligo 3495	3495-4012	5'-GGGACTTCGCAGGTGTTA	+
	T5G	4023-4054	5'-GGGCAGGCCCCCCCCCCGCGΔΔAGGGGGGAGG	-
	KK57	4036-4054	5'-GGGCAGGCCCCCCCCCCGC	-
Mutagenesis in diG	G5T	324-353	5'-GGGCAGGCCCCCCCCCCGCG**CG**AGGAGGGAGG	-
Construction of MDV/Pr chimeras	GPr5	315-322	5'-AAAAGACAGCCTCCCTCC	+
	CPr	320-356	5'-AATAGATAGCCTCCCTCCTCGGACGGGGGGCCTGCCC	+
	TPr	4014-4054	5'-AAAAGATAGCCTCCCCCCTCGCGCGGGGGGGGGGCC	+
	NDE	176-195	5' AGTGCACCATATGCGGTGTG	-
RNA gel blots	Oligo 13	249-269	5'-AGAGAGCACTAGTTTTCCAGG[d]	-

[a] Coordinates correspond to those of the TCV genome, diG, satC or pUC19 (Oligo NDE) as indicated.

[b] Bases in italics indicate T7 RNA polymerase promoter sequence. Bases in lowercase were added to achieve efficient digestion. Mutant bases are underlined. Bold bases denote nucleotides inserted in TCV promoter compared to satC or diG promoter. Bold and underlined bases indicate nucleotides inserted in TCV to generate a *SnaB*I site as shown in parentheses. "ΔΔ" indicates a two base deletion.

[c] "+" and "-" polarities refer to homology and complementarity with sat C plus strands, respectively.

[d] Oligo 13 is also complementary to positions 3950 to 3970 of TCV genomic RNA.

plasmids $C56_T$, $C1256_T$ and $C3456_T$, oligonucleotides T7C5' and C56C* were used as primers and pT7C+, $CH5_{TCV}$, or $C34_T$ were used as template, respectively. PCR products were cloned into the *Sma*I of pUC19. Plasmid $C568_T$ was generated in a similar fashion as $C56_T$ except that oligonucleotide C56C* was replaced by C568C*.

To generate plasmid T12C, PCR was performed with primers T7C5' and C568C* and template $C347_T$. PCR products were subsequently treated with T4 DNA polymerase and *Spe*I and cloned into the analogous location in pT7TCVms that had been treated with *Spe*I and *Sma*I. Plasmid T56C and T5G were generated in a similar fashion except Oligo3495 and Oligo7 or oligonucleotide T5G were used as primers with template pT7TCVms. Plasmid T1256C was also constructed similarly except oligonucleotides T7C5' and Oligo 7 were used as primers with template T12C. Plasmids TC12T, TC1234T, TC1256T, TC127T, TC34T, TC3456T, TC347T, TC56T and TC568T were generated by digesting plasmids $C12_T$, $C1234_T$, $C1256_T$, $C127_T$, $C34_T$, $C3456_T$, $C347_T$, $C56_T$, $C568_T$ with *Spe*I and *Sma*I. The fragments were inserted into the analogous location in pT7TCVms, which had been treated with the same restriction enzymes.

To construct plasmid $G5_T$, primers G5' and G5T were used with template pT7diG (Li et al., 1991). PCR products were subsequently treated with T4 DNA polymerase and *Spe*I and cloned into the analogous location in pT7diG that had been treated with *Spe*I and *Sma*I.

Generation of MDV/Pr constructs

To construct plasmid MDV-GPr, oligonucleotides GPr5 and NDE were used as primers with template $C3456_T$. Following treatment with T4 DNA polymerase and *Nde*I,

the fragment was inserted into pUC19T7MDV (Guan, 2000, Figure 6.1) that had been treated with *Sma*I and *Nde*I. Plasmid MDV-CPr was generated in similar fashion using primers CPr and NDE and template MDV-GPr. For plasmid MDV-TPr, oligonucleotide CPr was replaced by TPr.

In vitro transcription, preparation and inoculation of Arabidopsis protoplasts, and RNA gel blots

TCV genomic RNA, satC and diG transcripts were synthesized by T7 RNA polymerase using plasmids linearized with *Sma*I, which generates transcripts with precise 5' and 3' ends. Preparation and inoculation of Arabidopsis protoplasts, and RNA gel blots were performed as described in Chapter II.

Results

Two positions in the satC/TCV Pr are critical for satC accumulation

To determine which base differences in the 3' 104 bases of C3'100$_T$ were most responsible for reduced satC accumulation, satC constructs were generated with single and multiple positional changes to incorporate TCV-specific bases, and transcripts tested for accumulation in protoplasts (Figure 6.2). The 3' 100 bases of TCV and satC differ at 8 positions, which are labeled 1 through 8 in Figure 6.2A. Most variations are in the Pr, where two bases differences at position 6 and the single variance at position 8 reduce the eleven base-paired stem of the TCV Pr hairpin to seven base-pairs for satC. The satC and TCV Pr hairpin loops also vary substantially (compare the satC Pr in Figure 6.2A to the

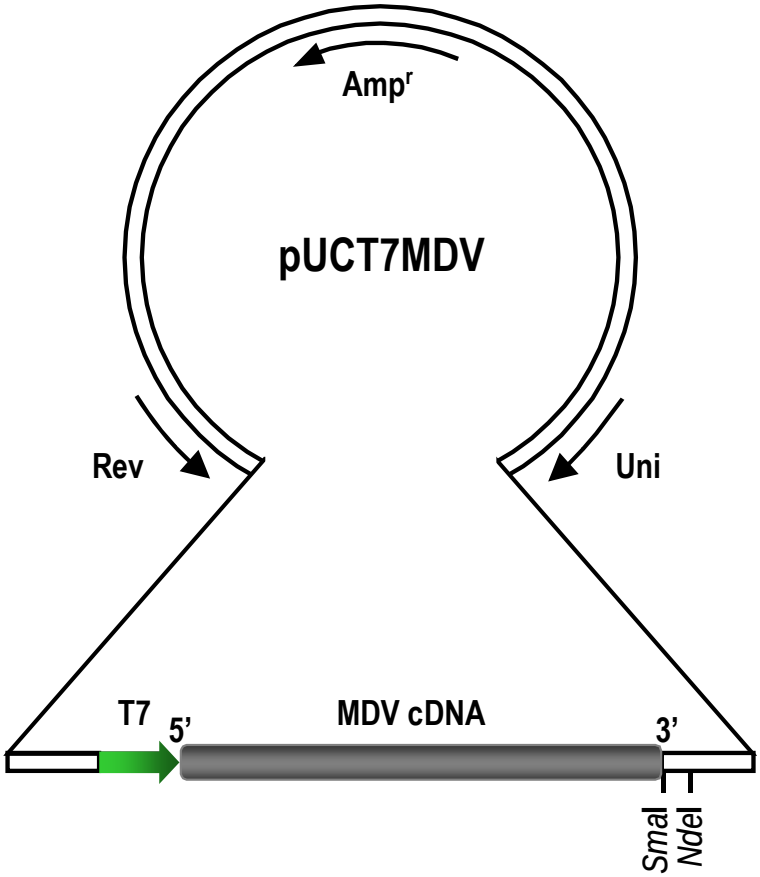
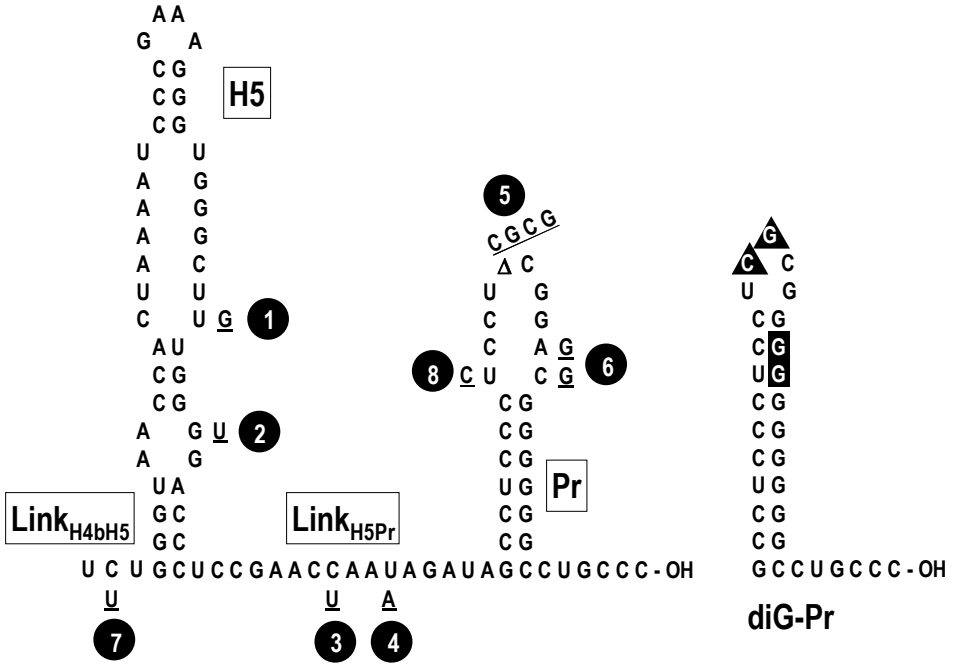


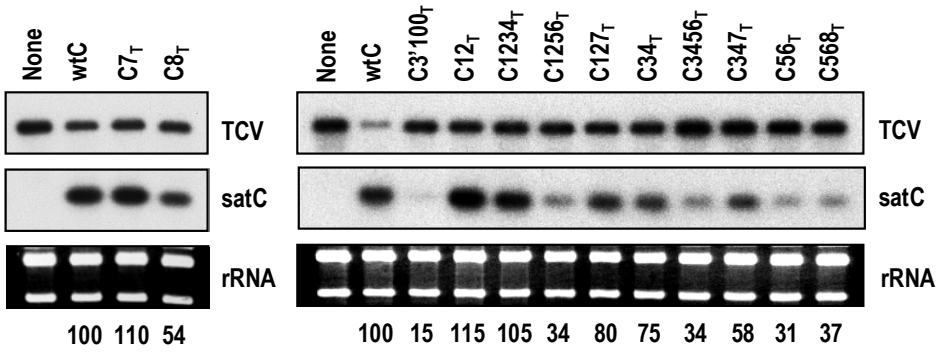
Figure 6.1 Map of pUCT7MDV. The blank and black bars represent the pUC19 and MDV cDNA sequence, respectively. The 5' and 3' ends of MDV positive-strand sequence are shown. The hatched arrow represents the T7 RNA polymerase promoter sequence. The endonuclease restriction sites used for cloning in this study are shown. (Modified from Guan, 2000).

Figure 6.2 Effect on satC accumulation in protoplasts after sequence conversion to residues found in TCV. (A) Left, nomenclature of positional differences between satC and TCV in their 3' 100 bases. The sequence of satC is given and positional variances in TCV are underlined. Locations of differences are identified by circled numbers. TCV has a 4 nt insert at position 5. Right, the Pr of diG. Nucleotides in triangles denote extra bases in diG compared with satC. Base alterations compared with satC are boxed. (B) Arabidopsis protoplasts were inoculated with TCV genomic RNA transcripts and transcripts of constructs shown above each lane. Total RNA was extracted at 40 hpi and probed with an oligonucleotide complementary to TCV, satC and diG. Ethidium bromide staining of the gel before blotting shows ribosomal RNA (rRNA) loading control. Numbers reflect average accumulation levels for two repeats with the exception of the first four lanes, which reflect a single assay. None, no added satC. wtC, wild-type satC; C3'100$_T$, satC with the 3' 104 bases of TCV. Numbers in the names of all other constructs reflect which positions were changed to those of TCV in the background of satC.

A


H5
Pr
Link H4bH5
Link H5Pr
U C U G C U C C G A A C C A A U A G A U A G C C U G C C C - OH
G C C U G C C C - OH
diG-Pr
1
2
3
4
5
6
7
8

B


None
wtC
C7T
C8T
TCV
satC
rRNA
100 110 54
None
wtC
C3'100T
C12T
C1234T
C1256T
C127T
C34T
C3456T
C347T
C56T
C568T
TCV
satC
rRNA
100 15 115 105 34 80 75 34 58 31 37

TCV Pr in Figure 6.5A). The two bases that vary between TCV and satC in H5 (positions 1 and 2) also reduce the stability of the satC H5 stem compared with H5 of TCV. Two base differences (positions 3 and 4) are also located in the linker region ($Link_{H5\text{-}Pr}$) between H5 and Pr and a single base variation (position 7) is in the 3 nt linker ($Link_{H4b\text{-}H5}$) between H4b and H5.

SatC construct $C12_T$, with H5 of TCV [i.e., positions 1 and 2 of TCV; most mutants are labeled with the backbone construct satC (C) or TCV (T) followed by the numerical positions that have been changed to the alternate sequence identified by a subscript T or C], accumulated slightly better than wild-type satC (115%), as previously reported (Zhang et al., 2005). The base change in $Link_{H4b\text{-}H5}$ (position 7; $C7_T$) also did not negatively affect accumulation of the satRNA, indicating that positions 1, 2 and 7 do not independently reduce $C3'100_T$ accumulation. Constructs containing both the H5 and $link_{H4b\text{-}H5}$ variances ($C127_T$), however, accumulated to 80% of wild-type, suggesting that the slight beneficial effect of incorporating the TCV H5 may be eliminated when transcripts contain all three differences. Alteration of positions 3 and 4 in $link_{H5\text{-}Pr}$ ($C34_T$) reduced satRNA accumulation to 75% of wild-type, which was further reduced by inclusion of position 7 ($C347_T$; 58% of wild-type). In contrast, satC containing TCV-specific bases in $link_{H5\text{-}Pr}$ and H5 ($C1234_T$) accumulated to 105% of wild-type.

Previous studies showed that satC with TCV Pr positions 5 or 6 accumulated to wild-type levels in protoplasts (Wang and Simon, 2000). In addition, satC with position 6 of TCV and a single CG dinucleotide at position 5 (this generates the Pr of diG and the construct has been renamed $C\text{-}Pr_G$) accumulated to 64% of wild-type levels (Wang and Simon, 2000; see Figure 6.3B). To examine the effect of additional combinations of

TCV-specific bases in the Pr, satC was altered at position 8 ($C8_T$), positions 5 and 6 ($C56_T$) and at all three locations ($C568_T$). $C8_T$ accumulated to 54% of wild-type, while $C56_T$ and $C568_T$ accumulated to 31 and 34% of wild-type, respectively. Accumulation was not enhanced when alterations at positions 1 and 2 in H5 or 3 and 4 in $link_{H5\text{-}Pr}$ were included with positional changes at 5 and 6 ($C1256_T$ and $C3456_T$, 34%). These results, combined with my earlier results, suggest: (i) satC accumulation is most negatively affected when containing both positions 5 and 6 of TCV; and (ii) no further reduction occurs when satC also contains TCV-specific H5 or the TCV $link_{H5\text{-}Pr}$ region.

Previous studies showed that diG accumulation increases when its Pr included positions 5 or 6 of satC, or both positions of satC, the latter of which generates diG with the Pr of satC (renamed G-Pr_C; Wang and Simon, 2000). To determine if diG accumulation is affected when containing position 5 of TCV [i.e., an insertion of CG at this location producing diG with Pr resembling the TCV Pr (G-Pr_T)], diG, G-Pr_T and G-Pr_C were assayed for accumulation in protoplasts. As previously found, diG was a poor template compared with satC, accumulating to only 11% of wild-type satC (Figure 6.3B). Also similar to prior results, accumulation was enhanced nearly 7-fold when diG contained the Pr of satC (G-Pr_C). In contrast, G-Pr_T accumulated as poorly as diG. Altogether, these results suggest: (i) the TCV Pr is a limited promoter in the context of satC or diG and (ii) base changes in the TCV Pr that occurred after the recombination events that produced satC or diG allow the Pr to function efficiently in the context of satC.

Coat protein ability to bind TCV-like Pr is not a primary factor in reduced accumulation of satC Pr variants

TCV CP binds specifically to the TCV-like Pr of diG but not to the Pr of satC (Wang and Simon, 2000). While efficiency of CP-binding to the TCV Pr was not determined, the possibility existed that CP interaction with the TCV-like Pr was responsible for the reduced accumulation of $C56_T$ and related constructs. To examine whether CP differentially interferes with the accumulation of satC with either the Pr of TCV or diG, wild-type satC, $C3'100_T$, $C56_T$, $C\text{-}Pr_G$, wild-type diG, $G\text{-}Pr_C$, and $G\text{-}Pr_T$ were co-inoculated with either wild-type TCV or CPmT, a TCV variant with mutations at the primary and secondary CP translation initiation codons that eliminate detectable CP (Wang and Simon, 1999, Figure 6.3A). As previously shown, CPmT accumulated to lower levels than wild-type TCV in protoplasts when compared with the levels of control ribosomal RNA (rRNA) (Figure 6.3B). Whether this reflects a negative effect of the mutations on replication/translation of the genomic RNA or a requirement for CP to achieve high level TCV accumulation is not known. Despite the reduction in TCV levels, both wild-type satC and satC variants accumulated to proportionately higher levels compared with rRNA when inoculated with CPmT (Figure 6.3B). Inoculation of wild-type diG and diG variants with either wild-type TCV or CPmT also showed similar results. Since this effect was not restricted to constructs with Pr of TCV or diG, CP binding to TCV-like Pr is apparently not responsible for the reduced accumulation of satC with Pr elements from TCV.

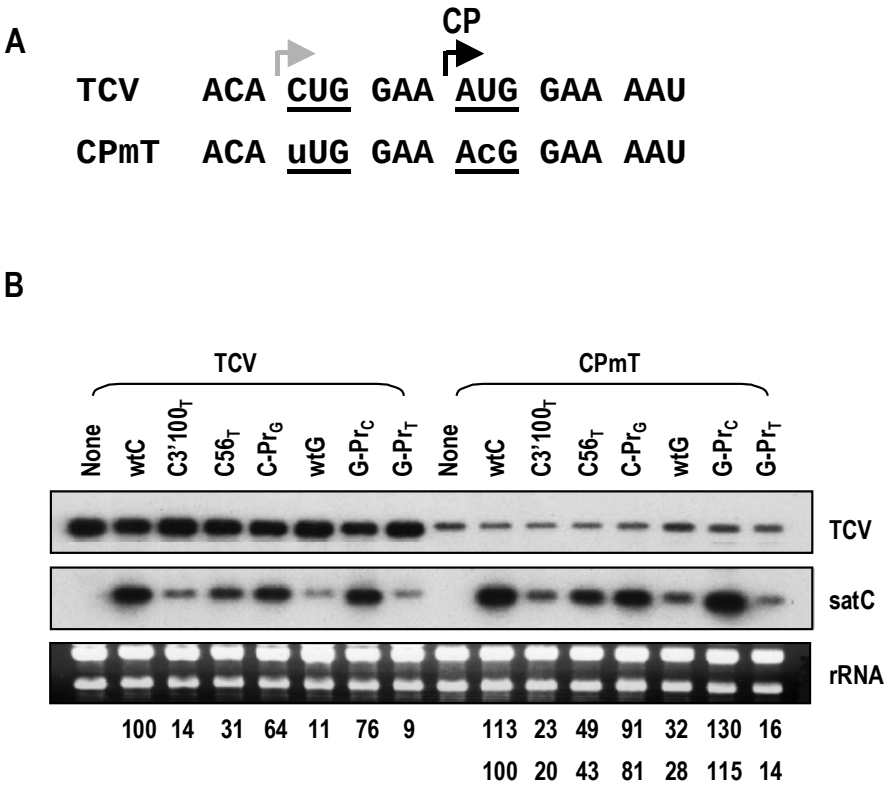


Figure 6.3 Effect of CP on TCV subviral RNA replication in protoplasts. (A) Sequence in the vicinity of the CP translation initiation codon in TCV and its CP-minus derivative CPmT. The normal AUG initiation codon and the alterative initiation codon that is used when the normal codon is mutated are underlined with arrows reflecting positions of translation initiation. The altered nucleotides in CPmT are indicated by lowercase letters. (B) RNA gel blot of mutant satC and TCV genomic RNA (+)-strands. Total RNA was extracted at 40 hpi from Arabidopsis protoplasts. Ethidium bromide staining of the gel before blotting shows ribosomal RNA loading control (below the blot). None, no added satC. wtC, wild-type satC. C3'100$_T$, satC with the 3' 104 bases of TCV. C56T, satC with TCV sequences at position 5 and 6. C-Pr$_G$, satC with the Pr of diG; wtG, wild-type diG; G-Pr$_C$, diG with the Pr of satC; G-Pr$_T$, diG with TCV sequence at position 5.

Efficient TCV accumulation depends on its cognate Pr

To determine if TCV accumulation is affected when containing the satC 3' end region, the 3' terminal 104 bases of TCV were replaced with the analogous region of satC, generating T3'100$_C$. Transcripts of T3'100$_C$ accumulated to barely detectable levels (1% of wild-type TCV), indicating that one or more of the eight positional changes was strongly detrimental to TCV accumulation (Figure 6.4). TCV with the satC H5 (T12$_C$), or satC-specific residues in positions 3, 4, 7 and 8 (T3478$_C$) accumulated to 84 or 81% of wild-type TCV, respectively, suggesting these positions are only marginally involved in reduced accumulation of T3'100$_C$. Additional combinations that retained TCV Pr positions 5 and 6 also had less than a 2-fold reduction in TCV accumulation (T1278$_C$, 66%; T12347$_C$, 56%; T123478$_C$, 61%). In contrast, TCV with positions 5 and 6 of satC, either alone or in combination with other satC-specific bases, accumulated between <1 and 5% of wild-type TCV (Figure 6.4). I also determined the effect of deleting one of the two CG dinucleotide repeats at position 5 in the TCV Pr loop, which generates TCV with Pr resembling the diG Pr (T-Pr$_G$). T-Pr$_G$ accumulated to only 2% of wild-type levels, indicating that altering position 5 strongly impacts on the efficiency of the TCV Pr.

I had previously determined that satC with the Pr hairpin of CCFV accumulates to 35% of the parental construct satC$_E$ (Figure 5.3), similar to satC with the TCV Pr (C568$_T$, 37%; Figure 6.2B). TCV with the Pr hairpin of CCFV accumulates to only 11% of the parental construct TCVs (TCV with a new restriction site created in the TCV H5/Pr linker by insertion of 3 nt, accumulated to identical levels as wild-type TCV in protoplasts, Table 6.2). In addition, TCV with the Pr hairpin of JINRV did not reach detectable levels in protoplasts (Performed by J.C., McCormack; Zhang et al., in press).

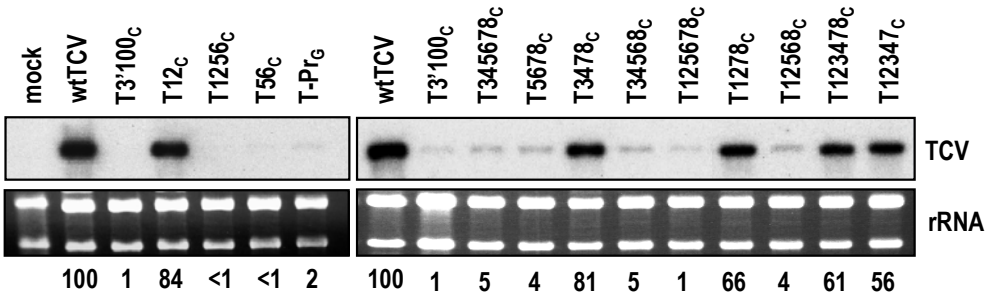


Figure 6.4 Effect on TCV accumulation in protoplasts after sequence conversion to residues found in satC and diG. Arabidopsis protoplasts were inoculated with constructs shown above each lane. Total RNA was extracted at 40 hpi and probed with an oligonucleotide complementary to TCV. Ethidium stained rRNAs were used as a loading control. Numbers reflect average accumulation levels for two repeats. Mock, plants were mock treated; T3'100_C, TCV with the 3' 100 bases of satC; wtTCV, wild-type TCV; T-Pr_G, TCV with diG sequence at position 5. Numbers in the names of all other constructs reflect which positions (described in Figure 6.1A) were changed to those of satC.

These results support a relationship between carmoviral genomic RNA promoters and cognate upstream sequences.

The TCV Pr is a less efficient promoter than the satC Pr when assayed in vitro

SatC transcripts synthesized by T7 RNA polymerase assume an initial pre-active structure that does not contain the phylogenetically conserved hairpins shown in Figure 1.4. Template activity of these transcripts for RdRp-directed complementary strand synthesis is substantially enhanced when specific mutations both within and outside the Pr region disrupt the initial Pr structure (Pr-1), causing the Pr to assume a structure that resembles its phylogenetically inferred form (active structure) known as Pr-2 or Pr-2* (Pr-2 and Pr-2* structures are very similar except that the 3' terminal sequences in Pr-2* remain in their wild-type configuration while these residues in Pr-2 are single-stranded) (Zhang et al., 2006). Although transcription is substantially enhanced for Pr-2*-containing mutant satC transcripts in vitro, none of these transcripts accumulated to detectable levels when inoculated onto protoplasts. These results were interpreted to reflect a satC requirement for both pre-active and active conformations in vivo, while in vitro, the initial presence of the active Pr-2* form enhances transcription without requiring a poorly executed conformational switch.

Solution structure probing showed that $C56_T$ assumes a form similar to the highly active satC Pr-2* in both the Pr region and the DR/H4a region. (Performed by G. Zhang; Zhang et al., in press). However, unlike satC variants with promoters in the form of Pr-2/Pr-2*, transcripts of $C56_T$ are not transcribed at high efficiency (42% of wild-type) by the TCV RdRp in vitro. Since transcription in vitro mainly assays initiation (products are

not templates for further synthesis), these results indicate that the TCV Pr is not as efficient a promoter as the satC Pr in the context of satC.

As a further test of the comparative promoter efficiencies of the TCV and satC Pr, both sequences were removed from their natural location and placed downstream of an unrelated sequence [Qβ bacteriophage-associated midivariant (MDV) RNA (220 nt)] that is not a template for TCV RdRp. The 3'-terminal 37 bases of (+)-strand satC, comprising the Pr hairpin, six flanking 3-terminal bases and eight flanking 5' bases, was previously determined to be able to efficiently direct complementary strand synthesis of MDV using TCV RdRp partially purified from infected turnip plants (Song and Simon, 1995). To directly compare Pr activities in vitro, analogous Pr-containing segments from satC, TCV and diG were joined to MDV, and template activity measured in vitro using p88 RdRp (Performed by G. Zhang). As shown in Figure 6.5, the satC Pr was able to direct 6-fold or 50-fold more complementary strand synthesis than the TCV Pr or the diG Pr, respectively. These results suggest that modifications in the TCV Pr that generated the Pr of satC resulted in a promoter that can function more efficiently in the absence of additional viral sequences in vitro.

Discussion

Efficient accumulation of (+)-strand RNA viruses requires direct or indirect long distance interactions between their 3' and 5' ends (Barton et al., 2001; Frolov et al, 2001; Herold and Andino, 2001; Khromykh et al., 2001; You et al., 2001; Vlot and Bol, 2003). Such long distance interactions may be necessary to coordinate translation and

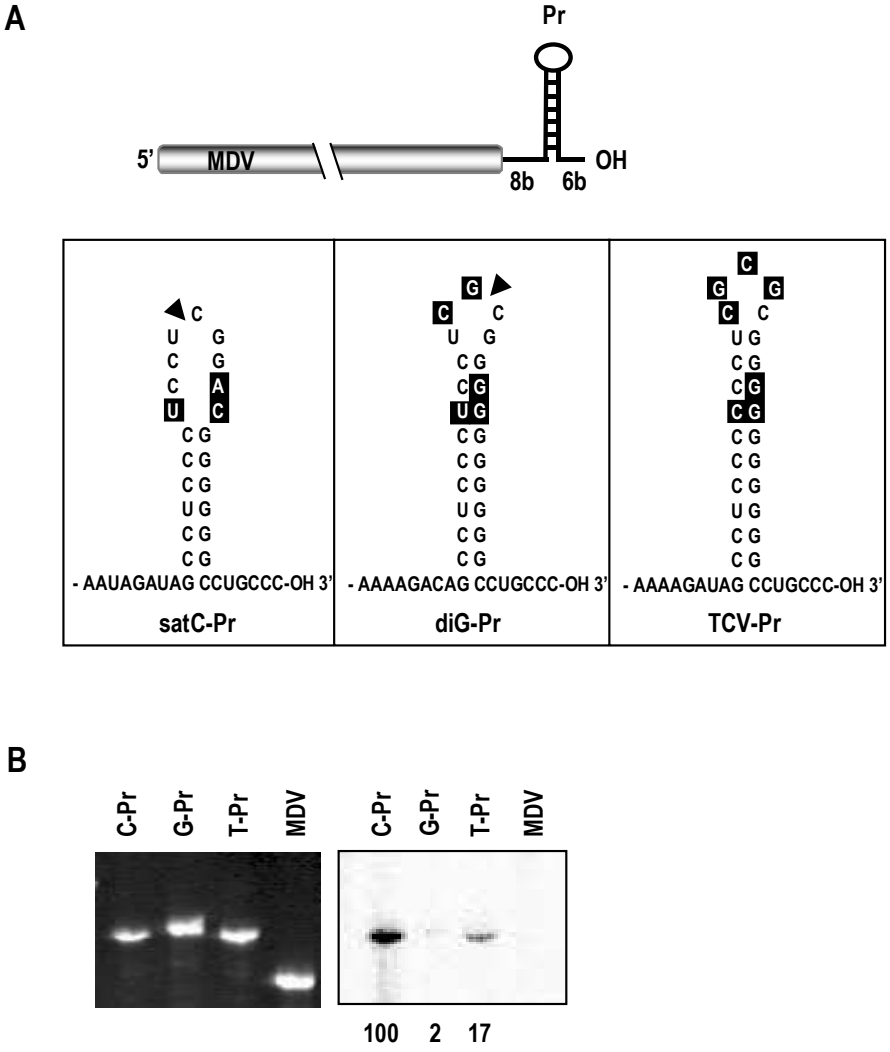


Figure 6.5 Transcription efficiencies of TCV, satC and diG core promoters. (A) Schematic representation of the chimeric RNAs containing Pr hairpin and short flanking sequences that were joined to MDV RNA. Actual core promoter sequences are shown below. Triangles indicate four and two base deletions in the satC Pr and the diG Pr, respectively, compared to the TCV Pr. (B) In vitro RdRp assay (performed by G. Zhang). Transcripts of constructs indicated above each lane were synthesized by T7 RNA polymerase and subjected to complementary strand synthesis by TCV RdRp in the presence of [α^{32}P]-UTP. Identities of the attached Pr are shown above each lane. The gel was stained with ethidium-bromide to indicate template levels. Values below lanes are averages for two independent experiments. C-Pr, Pr of satC; G-Pr, Pr of diG, T-Pr, Pr of TCV; MDV, MDV RNA without any Pr sequence.

replication, present the initiation site to the replicase, and/or stabilize the viral RNA. The initial event that created satC, however, resulted in a molecule that retained only the 3' region of TCV joined to satD, a molecule with limited similarity to the TCV genomic RNA. While modern satC accumulates to levels similar to 5S rRNA, making it one of the most prevalent RNAs in an infected cell, satC with the 3' 100 bases of TCV ($C3'100_T$), which should more closely resemble the progenitor RNA, is a poor template that accumulates to only 15% of modern (wild-type) satC levels.

Mutational analysis revealed that satC constructs were inefficient templates when they contained TCV positions 5 and 6 located in the core Pr promoter. Little difference in accumulation was found when $C56_T$ also contained additional TCV-specific bases in position 8 in the Pr, positions 1 and 2 in the upstream hairpin H5 or positions 3 and 4 in the $link_{H5\text{-}Pr}$ region (34 to 37% of wild-type satC for all constructs). The two-fold difference in accumulation between these constructs and $C3'100_T$ suggests that the three positional variances in the replication-important DR/H4a/H4b regions (Zhang et al., 2006; Zhang et al., manuscript in revision), may also have evolved to enhance satC-specific accumulation. This possibility is supported by recent evidence indicating important functional differences in the H4a region between TCV and satC (J.C. McCormack, R. Guo and A. E. Simon, unpublished). Different functions of related sequences were also found for TMV genomic RNA and its artificially-derived defective RNAs (Chandrika et al., 2000). Replication of the defective RNAs required a smaller 3' element that was more sequence specific than that of its parental genomic RNA, leading to the proposal that replication of RNAs in cis and in trans might use different pools of replicase complexes with different recognition mechanisms.

Weak accumulation of TCV with the Pr of satC (or diG-like) supports the proposition that the TCV core promoter is adapted to function with upstream TCV sequences. The TCV Pr is a weak core promoter when assayed in vitro, either as part of satC or in the absence of any additional TCV sequences, despite the structural form of the Pr resembling the highly active form of the satC Pr (Pr-2*). TCV with the Pr hairpin of CCFV or JINRV also accumulated poorly or undetectably in protoplasts (Performed by J.C. McCormack; Zhang et al., in press), whereas satC with the Pr hairpin of CCFV or TCV accumulated to similar levels (35% and 37% of wild-type satC, respectively; Figure 5.2B, Figure 6.2B). These results suggest that in the absence of upstream interacting elements, the TCV RdRp can similarly recognize the Pr of TCV and CCFV in the context of satC.

The possibility of an interaction between the TCV Pr and upstream sequences is also supported by our previous results showing that TCV full length genomic RNA, but not C3'100$_T$, is subjected to abortive cycling when transcribed by the RdRp (Nagy et al., 1997). The propensity of the RdRp to generate short complementary products (and thus failing to proceed efficiently from initiation to elongation) using TCV genomic RNA, but not C3'100$_T$, suggests that abortive cycling is not caused by the inability of the RdRp to unwind the highly stable TCV Pr. Rather, cycling by the polymerase appears to require additional upstream sequences that together with the Pr produces a structure that causes the RdRp to transition poorly between initiation and elongation.

The increased efficiency of the satC Pr in directing complementary strand synthesis in vitro suggests that evolutionary adaptation has allowed the Pr to function independent of its upstream interacting element. DiG, whose 3' end region closely

resembles that of TCV (Li et al., 1989), accumulates poorly in vivo (11% of wild-type satC; Figure 6.3B), suggesting that the diG Pr is less adapted for high level accumulation than the Pr of satC. However, the 2 nt deletion at position 5 in the diG Pr, while not obviously affecting accumulation of diG (compare wild-type diG with G-Pr_T, Figure 6.3B), results in a nearly 2-fold enhancement in accumulation of satC compared to satC with the TCV Pr [compare $C568_T$ (Figure 6.2B) and C-Pr_G (Figure 6.3B)]. This difference suggests that when associated with a possibly "more evolved" molecule like satC, the diG Pr is able to function with greater independence than the TCV Pr. The importance of these two nucleotides at position 5 in the TCV Pr hairpin loop for TCV accumulation [deletion resulted in a 50-fold decrease (Figure 6.4)] suggests that the Pr hairpin loop region may be the 3' element that interacts with upstream sequences missing in diG and satC.

The recently discovered satC conformational switch involving the 3' terminal 140 nt must also be considered as resulting from the base differences between TCV and satC (Zhang et al., 2006; Zhang et al., manuscript in revision). In TCV genomic RNA, as with other viral genomic RNAs, a conformational switch is probably needed to convert the template from one that is translationally competent to one that is active for replication (van Dijk et al., 2004). Since satC is not translated, its switch may be important to reduce or eliminate the ability of newly synthesized (+)-strands to serve as templates for further (-)-strand synthesis (Chao et al., 2002; Brown et al., 2004). Base differences at positions 1, 2, 5 and 6 all reduce the stability of the satC active structure, which, along with other base changes, may have helped in evolving an alternative structure comprising only available local sequences.

CONCLUSION

In this dissertation, I described my studies on cis-acting elements and a conformational switch proposed to be involved in satC accumulation in protoplasts. My major findings are: (i) The cognate core promoter is important for efficient satC (and also TCV) replication; (ii) Both specific sequence and structural features of H5 are required for robust satC accumulation in plants and protoplasts; (iii) H5 may be involved in synthesis of both strands; (iv) H4a and H4b function as a unit; (v) Formation of Ψ_1 and Ψ_2 are required for satC accumulation; (vi) DR, CE3 and the linker region between H5 and Pr are also important for satC accumulation; (vii) M1H, H2 and H6 may primarily function in (+)-strand synthesis. My results also support a conformational switch from pre-active structure to active structure that appears to be required for satC (-)-strand synthesis. In addition, Ψ_2 may stabilize the pre-active structure, and the DR may facilitate the conformational switch by destabilizing Ψ_2. Furthermore, the H5 LSL and/or the 3' terminus may have a function other than forming Ψ_1. In summary, my results indicate that efficient replication of satC requires cis-acting elements located throughout the RNA molecules and suggest that this process may also require RNA conformational changes. These results provide some important pieces for solving the puzzle of satC replication and also suggest that using conformational switches to regulate replication of virus genomes may be a common feature for (+)-strand RNA viruses.

Replication of such a small viral RNA as satC is so complex that I wonder how much more complicated is replication of the TCV genomic RNA and larger viruses. Some of the findings in satC are also recapitulated in TCV including the importance of

the four phylogenetically conserved hairpins and the formation of Ψ_1 and Ψ_2 for viral RNA replication although differences exist (J.C. McCormack and A.E. Simon, unpublished results; Sun and Simon, in press). However, since satC also contains sequences derived from satD and lacks the 5' thousands of nucleotides of TCV genomic RNA, it contains some unique features in replication that are not shared by TCV, and *vice versa*. For example, H4a and H4b function as a unit in satC (Figure 5.1) but not in TCV (J.C. McCormack and A.E. Simon, unpublished data). A guanylate to cytidylate change at position 230, located in the satC H4a loop, reduced satC accumulation to 67% of wild-type (Figure 3.8) while this same change in TCV eliminated detectable accumulation (J.C. McCormack and A.E. Simon, unpublished data). In addition, satC Pr is a much stronger core promoter than TCV Pr when assayed in vitro, suggesting that the TCV Pr requires upstream elements for full functionality (Chapter VI). Initiation of satC (-)-strand synthesis appears to require conformational changes at the 3' end region including conversion of the pre-active Pr-1 structure into the active Pr-2 structure (Chapter V). Since TCV Pr hairpin is much more stable than satC Pr hairpin, it is unlikely that TCV Pr could comprise part of a conformational switch. Currently it is not clear how the TCV genomic RNA is converted from a template for translation to a template for replication. One possibility is that this conversion also requires a conformational switch involving long-distance RNA-RNA interactions as shown in some other viruses (Isken et al., 2004; Khromykh et al., 2001). In summary, these results suggest that while small replicons derived from the genomic RNA are excellent systems for studying viral RNA replication, they may not reflect the whole picture for genomic RNA replication.

H4a, H4b, H5 and Pr hairpin as well as the formation of Ψ_1 and Ψ_2 are phylogenetically conserved in genus *Carmovirus* suggesting that these elements may have similar functions in other carmoviruses. A hairpin similar to H5 was also found in other members of the *Tombusviridae* (Pogany et al., 2003; Na and White, 2006). As discussed earlier, in TBSV, an interaction between this hairpin (SL3) and the 3' terminal sequence was shown to be important for RNA synthesis both in vivo and in vitro (Pogany et al., 2003) and SL3 is also important for replicase assembly in yeast (Panaviene et al., 2005). A difference between H5 and SL3 is that H5 contains both large and small symmetrical internal loops while SL3 contains an asymmetrical internal loop and a bulged uridylate (Figure 1.4, Figure 1.5). The importance of symmetry to H5 is unknown. In vivo SELEX results suggest that H5 LSL can tolerate slight asymmetry (Chapter IV). While the 3' side of the LSL is involved in interaction with the 3' terminal sequence, the function of the 5' side of the LSL, which is also sequence specific, remains unknown. It will be interesting to see if a satC mutant containing only one adenylate in the 5' side of the LSL, which is SL3-like, can accumulate in vivo and in vitro.

Results from RNA structural probing, in vitro RdRp transcription assays and protoplast experiments suggested that both pre-active and active structures are required for satC replication in vivo and a conformational switch appears required to initiate (-)-strand synthesis. However, many questions await further study. For example, what is the structure of the pre-active RNA? What is the relationship between H4a and H4b, which appear to be one functional unit? How does DR promote the conformational switch? Why does G218C benefit the fitness of satC with CCFV H5 in plants? What is the role of CE3? What triggers the conformational switch? Are RdRp and/or other protein factors

involved in the conformational switch and how? It has been shown that the 5' terminal sequence also affects the structure of the 3' end region and the initiation of (-)-strands (Zhang et al., 2004), so what role does the 5' terminal sequence play in (-)-strand synthesis? Although Ψ_1 is required for replication, it has to be opened at some point, allowing the TCV RdRp to initiate (-)-strand synthesis from the released 3' terminus. How does this happen? While in vitro RdRp transcription assays support the formation of Ψ_1, all these constructs accumulated to undetectable levels in protoplasts despite whether the proposed interaction was disrupted or re-established. Results from protoplast experiments did not support other interactions between H5 LSL, 3' terminus and other regions of satC either. Will in vitro RdRp transcription assays using these constructs as template provide new information? In addition, the formation of Ψ_1 in TCV is supported by protoplast experiments, where mutations disrupt the interaction reduced TCV accumulation to 7 and 29% of wild-type and mutations re-established the proposed interaction restored TCV accumulation to 50% of wild-type (Performed by J.C. McCormack; Zhang et al., in press). This suggested that the sequences involved in satC Ψ_1 might participate in additional activities crucial for satC accumulation in vivo. What are the additional functions of these sequences in satC? Ψ_1 and Ψ_2 may also need to be further defined.

Some other questions that need to be further addressed include: How does H5 affect (+)-strand synthesis? How is H5 involved in the replicase assembly? What is the function of H6? What is the requirement in promoting satC dimer synthesis? Why does TCV CP affect satC accumulation? What is the upstream sequence that may interact with the TCV Pr?

The approaches that may help in the investigation of these problems include mutagenesis, in vitro RdRp transcription assays, protoplast experiments, temperature-gradient gel electrophoresis (TGGE) assays, UV-crosslinking, foot-printing, NMR and crystallography. In addition, atomic force microscopy has already been used to visualize individual *Dengue virus* viral RNA molecule (Alvarez et al., 2005). I hope this technique can be used to observe satC and TCV molecules, which may help to determine RNA structure.

REFERENCE

Abbink, T.E., Ooms, M., Haasnoot, P.C., Berkhout, B., 2005. The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. Biochemistry 44, 9058-9066.

Abramovitz, D. L., Pyle, A. M., 1997. Remarkable morphological variability of a common RNA folding motif: the GNRA tetraloop-receptor interaction. J. Mol. Biol. 266, 493-506.

Adkins., S., Siegel, R.W., Sun, J.H., Kao, C.C., 1997. Minimal templates directing accurate initiation of subgenomic RNA synthesis in vitro by the brome mosaic virus RNA-dependent RNA polymerase. RNA 3, 634-647.

Ahlquist, P., Noueiry, A.O., Lee, W., Kushner, D.B., Dye, B.T., 2003. Host factors in positive-strand RNA virus genome replication. J. Virol. 77, 8181-8186.

Albariño, C.G., Eckerle, L.D., Ball, L.A., 2003. The cis-acting replication signal at the 3′ end of flock house virus RNA2 is RNA3-dependent. Virology 311, 181–191.

Alvarez, D.E., Lodeiro, M.F., Luduena, S.J., Pietrasanta, L.I., Gamarnik, A.V., 2005. Long-range RNA-RNA interactions circularize the Dengue virus genome. J.Virol. 79, 6631-6643.

Andino, R., Rieckhof, G.E., Achacoso, P.L., Baltomore, D., 1993. Poliovirus RNA synthesis utilizes an RNP complex formed around the 5'-end of viral RNA. EMBO J. 12, 3587-3598.

Aparicio, F., Vilar, M., Perez-Paya, E., Pallas, V., 2003. The coat protein of prunus necrotic ringspot virus specifically binds to and regulates the conformation of its genomic RNA. Virology 313, 213-223.

Axelrod, V.D., Brown, E., Priano, C., Mills, D.R., 1991. Coliphage Qβ RNA replication: RNA catalytic for single-strand release. Virology 184, 595-608.

Baer, M., Houser, F., Loesch-Fries, L.S., Gehrke, L., 1994. Specific RNA binding by amino-terminal peptides of alfalfa mosaic virus coat protein. EMBO J. 13, 727-735.

Balmori, E., Gilmer, D., Richards, K., Guilley, H., Jonard, G., 1993. Mapping the promoter for subgenomic RNA synthesis on beet necrotic yellow vein virus RNA 3. Biochimie 75, 517-521.

Banerjee, R., Dasgupta., A., 2001. Specific interaction of hepatitis C virus

protease/helicase NS3 with the 3'-terminal sequences of viral positive- and negative-strand RNA. J. Virol. 75, 1708-1721.

Barends, S., Bink, H.H.J., van den Worm, S.H.E., Pleij, C.W.A., Kraal, B., 2003. Entrapping ribosomes for viral translation: tRNA mimicry as a molecular Trojan horse. Cell 112, 123-129.

Barends, S., Rudinger-Thirion, J., Florentz, C., Giegé, R., Pleij, C.W.A., Kraal, B., 2004. tRNA-like structure regulates translation of brome mosaic virus RNA. J. Virol. 78, 4003-4010.

Barrera, L., Schuppli, D., Sogo, J.M., Weber, H., 1993. Different mechanisms of recognition of bacteriophage Qβ plus and minus strand RNAs by Qβ replicase. J. Mol. Biol. 232, 512-521.

Barry, J.K., Miller, W.A., 2002. A –1 ribosomal frameshift element that requires base pairing across four kilobases suggests a mechanism of regulating ribosome and replicase traffic on a viral RNA. Proc. Natl. Acad. Sci. U. S. A. 99, 11133-11138.

Bartel D.P., 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281-297.

Barton, D.J., Black, E.P., Flanegan, J.B., 1995. Complete replication of poliovirus in vitro: preinitiation RNA replication complexes require soluble cellular factors for the synthesis of VPg-linked RNA. J. Virol. 69, 5516-5527.

Barton, D.J., O'Donnell, B.J., Flanegan, J.B., 2001. 5' cloverleaf in poliovirus RNA is a cis-acting replication element required for negative-strand synthesis. EMBO J. 20, 1439-1448.

Baumstark, T., Schroder, A.R.W., Riesner, D., 1997. Viroid processing: switch from cleavage to ligation is driven by a change from a tetraloop to a loop E conformation. EMBO J. 16, 599-610.

Baumstark, T., Ahlquist, P., 2001. The brome mosaic virus RNA3 intergenic replication enhancer folds to mimic a tRNA TpsiC-stem loop and is modified in vivo. RNA 7, 1652-1670.

Bedard, K.M., Semler, B.L., 2004. Regulation of picornavirus gene expression. Microbes Infect. 6, 702-713.

Berkhout, B., Ooms, M., Beerens, N., Huthoff, H., Southern, E., Verhoef, K., 2002. In vitro evidence that the untranslated leader of the HIV-1 genome is an RNA checkpoint that regulates multiple functions through conformational changes. J. Biol. Chem. 277, 19967-19975.

Blight, B.J., Rice, C.M., 1997. Secondary structure determination of the conserved 98-base sequence at the 3' terminus of hepatitis C virus genome RNA. J. Virol. 71, 7345-7352.

Bol, J.F., 1999. Alfalfa mosaic virus and ilarviruses: involvement of coat protein in multiple steps of the replication cycle. J. Gen. Virol. 80, 1089-1102.

Bol, J.F., 2005. Replication of Alfamo- and Ilarviruses: role of the coat protein. Annu. Rev. Phytopathol. 43, 39-62.

Brantl, S. 2004. Bacterial gene regulation: from transcription attenuation to riboswitches and ribozymes. Trends Microbiol. 12, 473-475.

Bredenbeek, P.J., Kooi, E.A., Lindenbach, B., Huijkman, N., Rice, C.M., Spaan, W.J., 2003. A stable full-length yellow fever virus cDNA clone and the role of conserved RNA elements in flavivirus replication. J. Gen. Virol. 84, 1261-1268.

Brown, D.M., Kauder, S.E., Cornell, C.T., Jang, G.M., Racaniello, V.R., Semler, B.L., 2004. Cell-dependent role for the poliovirus 3' noncoding region in positive-strand RNA synthesis. J. Virol. 78, 1344-1351.

Buck, K.W., 1996. Comparison of the replication of positive-stranded RNA viruses of plants and animals. Adv. Virus Res. 47, 159-251.

Carpenter, C.D., Cascone, P.J., Simon, A.E., 1991a. Formation of multimers of linear satellite RNAs. Virology 183, 586-594.

Carpenter, C.D., Cascone, P.J., Simon, A.E., 1991b. Mutations in a satellite RNA of turnip crinkle virus result in addition of Poly (U) in vivo. Virology 183, 595-601.

Carpenter, C.D., Simon, A.E., 1998. In vivo restoration of biologically active 3' ends of virus-associated RNAs by nonhomologous RNA recombination and replacement of a terminal motif. . J. Virol. 70, 478-486.

Carpenter, C.D., Simon, A.E., 1998. Analysis of sequences and putative structures required for viral satellite RNA accumulation by in vivo genetic selection. Nucleic Acids Res. 26, 2426-2432.

Carrington, J.C., Heaton, L.A., Zuidema, D., Hillman, B.I., Morris, T.J., 1989. The genome structure of turnip crinkle virus. Virology 170, 219-226.

Carrington, J.C., Ambros, V., 2003. Role of microRNAs in plant and animal development. Science 301, 336-338.

Cascone, P.J., Haydar, T.F., Simon, A.E., 1993. Sequences and structures required for recombination between virus-associated RNAs. Science 260, 801-805.

Cech, T.R., Damberger, S.H., Gutell, R.R., 1994. Representation of the secondary and tertiary structure of group I introns. Nature Struct. Biol. 1, 273-280.

Chandrika, R., Rabindran, S., Lewandowski, D.J., Manjunath, K.L., Dawson, W.O., 2000. Full-length Tobacco mosaic virus RNAs and defective RNAs have different 3' replication signals. Virology 273, 198-209.

Chao, L., Rang, C.U., Wong, L.E., 2002. Distribution of spontaneous mutants and inferences about the replication mode of the RNA bacteriophage □6. J. Virol. 76, 3276-3281.

Chapman, R.M., Rao, A.L., Kao, C.C., 1998. Sequences 5' of the conserved tRNA-like promoter modulate the initiation of minus-strand synthesis by the brome mosaic virus RNA-dependent RNA polymerase. Virology 252, 458-467.

Chapman, R.M., Kao, C.C., 1999. A minimal RNA promoter for minus-strand RNA synthesis by the brome mosaic virus polymerase complex. J. Mol. Biol. 286, 709-720.

Chen, J., Noueiry, A., Ahlquist, P., 2001. Brome mosaic virus Protein 1a recruits viral RNA2 to RNA replication through a 5' proximal RNA2 signal. J. Virol. 75, 3207-19.

Chen, M.H., Roossinck, M.J., Kao, C.C., 2000. Efficient and specific initiation of subgenomic RNA synthesis by cucumber mosaic virus replicase in vitro requires an upstream RNA stem-loop. J. Virol. 74, 11201-11209.

Cheng, J.-U., Chang, M.-F., Chang, S.C., 1999. Specific interaction between the hepatitis C virus NS5B RNA polymerase and the 3′ end of the viral RNA. J. Virol. 73, 7044-7049.

Choi, I.-R., Ostrovsky, M., Zhang, G., White, K.A., 2001. Regulatory activity of distal and core RNA elements in tombusvirus subgenomic mRNA2 transcription. J. Bio. Chem. 276, 41761-41768.

Choi, I.-R., White, K.A., 2002. An RNA activator of subgenomic mRNA1 transcription in tomato bushy stunt virus. J. Bio. Chem. 277, 3760-3766.

Choi, S.-K., Hema, M., Gopinath, K., Santos, J., Kao, C.C., 2004. Replicase-binding sites on plus- and minus-strand brome mosaic virus RNAs and their roles in RNA replication in plant cells. J. Virol. 78, 13420-13429.

Choi, Y.G., Dreher, T.W., Rao, A.L., 2002. tRNA elements mediate the assembly of an icosahedral RNA virus. Proc. Natl. Acad. Sci. U. S. A. 99, 655-660.

Chong, J.L., Chuang, R.Y., Tung, L., Chang, T.H., 2004. Ded1p, a conserved DExD/H-box translation factor, can promote yeast L-A virus negative-strand RNA synthesis in vitro. Nucleic Acid Res. 32, 2031-8.

Ciuffreda, P., Rubino, L., Russo, M., 1998. Molecular cloning and complete nucleotide sequence of galinsoga mosaic virus genomic RNA. Arch. Virol. 143, 173-180.

Cohen, Y., Gisel, A., Zambrski, P.C., 2000a. Cell-to-cell and systemic movement of recombinant green fluorescent protein-tagged turnip crinkle viruses. Virology 273, 258-266.

Cohen, Y., Qu, F., Gisel, A., Morris, T.J., Zambrski, P.C., 2000b. Nuclear localization of turnip crinkle virus movement protein p8. Virology 273, 276-285.

Cornell, C.T., Perera, R., Brunner, J.E., Semler, B.L., 2004. Strand-specific RNA synthesis determinants in the RNA-dependent RNA polymerase of poliovirus. J. Virol. 78, 4397-4407.

Cover, J., Lenches, E., Smith, K., Robison, R.A., Sando, T., Strauss, E.G., Strauss, J.H., 2003. Fine mapping of a cis-acting sequence element in yellow fever virus RNA that is required for RNA replication and cyclization. J. Virol. 77, 2265-2270.

Dalmay, T., Szittya, G., Burgyan, J., 1995. Generation of defective interfering RNA dimmers of cymbidium ringspot tombusvirus. Virology 207, 510-517.

David, C., Gargouri-Bouzid, R., Haenni, A.L. 1992. RNA replication of plant viruses containing an RNA genome. Prog. Nucleic Acid Res. Mol. Biol. 42, 157-227.

De, I., Sawicki, S.G., Sawicki, D.L., 1996. Sindbis virus RNA-negative mutants that fail to convert from minus-strnd to plus-strand synthesis: role of the nsP2 protein. J. Virol. 70, 2706-2719.

Deiman, B.A.L.M., Kortlever, R.M., Pleij, C.W.A., 1997. The role of he pseudoknot at the 3' end of turnip yellow mosaic virus in minus-strand synthesis by the viral RNA-dependent RNA polymerase. J. Virol. 71, 5990-5996.

Deiman, B.A.L.M., Koenen, A.K., Verlaan, P.W.G., Pleij, C.W.A., 1998. Minimal template requirements for initiation of minus-strand synthesis in vitro by the RNA-dependent RNA polymerase of turnip yellow mosaic virus. J. Virol. 72, 3965-3972.

Dey, A., York, D., Smalls-Manty, A., Summers, M.F., 2005. Composition and sequence-dependent binding of RNA to the nucleocapsid protein of moloney murine leukemia virus. Biochemistry 44, 3735-3744.

Diez, J., Ishikawa, M., Kaido, M., Ahlquist, P., 2000. Identification and characterization of a host protein required for efficient template selection in viral RNA replication. Proc. Natl. Acad. Sci. U. S. A. 97, 3913-3918.

Dreher, T.W., Hall, T.C., 1988. Mutational analysis of the sequence and structural requirements in brome mosaic virus RNA for minus-strand promoter activity. J. Mol. Biol. 201, 31-40.

Dreher, T.W., Hall, T.C., 1988. Mutational analysis of the tRNA mimicry of brome mosaic virus RNA. J. Mol. Biol. 201, 41-55.

Dreher, T.W., Rao, A.L., Hall, T.C., 1989. Replication in vivo of mutant brome mosaic virus RNAs defective in aminoacylation. J. Mol. Biol. 206, 425-438.

Dreher, T.W., Tsai, C.H., Florentz, C., Giege, R., 1992. Specific valylation of turnip yellow mosaic virus RNA by wheat germ valyl-tRNA synthetase determined by three anticodon loop nucleotides. Biochemistry 31, 9183-9189.

Dreher, T.W., 1999. Functions of the 3'-untranslated regions of positive strand RNA viral genomes. Annu. Rev. Phytopathol. 37, 151-174.

D'Souza, V., Summers, M.F., 2004. Structural basis for packaging the dimeric genome of moloney murine leukaemia virus. Nature 431, 586-590.

D'Souza, V., Summers, M.F., 2005. How retroviruses select their genomes. Nature Reviews Microbiol. 3, 643-655.

Du, T., Zamore, P.D., 2005. microPrimer: the biogenesis and function of microRNA. Development 132, 4645-4652.

Duggal, R., Rao, A.L., Hall, T.C., 1992. Unique nucleotide differences in the conserved 3' termini of brome mosaic virus RNAs are maintained through their optimization of genome replication. Virology 187, 261-270.

Duggal, R., Lahser, F.C., Hall, T.C., 1994. Cis-Acting sequences in the replication of plant viruses with plus-sense RNA genomes. Annu. Rev. Phytopathol. 32, 287-309.

Dutkiewicz, M., Ciesiolka, J., 2005. Structural characterization of the highly conserved 98-base sequence at the 3 ' end of HCV RNA genome and the complementary sequence located at the 5 ' end of the replicative viral strand. Nucl. Acids Res. 33, 693-703.

Dykxhoorn, D.M., Novina, C.D., Sharp, P.A., 2003. Killing the messenger: short RNAs that silence gene expression. Nature Rev. Mol. Cell Biol. 4, 457-467.

Eckerle, L.D., Ball, L.A., 2002. Replication of the RNA segments of a bipartite viral genome is coordinated by a transactivating subgenomic RNA. Virology 296, 165-176.

Eckerle, L.D., Albariño, C.G., Ball, L.A., 2003. Flock house virus subgenomic RNA3 is replicated and its replication correlates with transactivation of RNA2. Virology 317, 95-108.

Erdmann, V.A., Barciszewska, M.Z., Hochberg, A., de Groot N., Barciszewski, J., 2001. Regulatory RNAs. Cell Mol. Life Sci. 58,960-977.

Fabian, M.R., Na, H., Ray, D., White, K.A., 2003. 3'-Terminal RNA secondary structures are important for accumulation of Tomato bushy virus DI RNAs. Virology 313, 567-80.

Felden, B., Florentz, C., Geige', R., Westhof, E., 1994. Solution structure of the 3'-end of brome mosaic virus genomic RNAs. Conformational mimicry with canonical tRNAs. J. Mol. Biol. 235, 508-531.

Felden, B., Florentz, C., Geige', R., Westhof, E., 1996. A central pseudoknotted three-way junction imposes tRNA-like mimicry and the orientation of three 5' upstream pseudoknots in the 3' terminus of tobacco mosaic virus RNA. RNA 2, 201-212.

French, R., Ahlquist., P., 1987. Intercistronic as well as terminal sequences are required for efficient amplification of Brome mosaic virus RNA3. J. Virol. 61, 1457-1465.

Friebe P., Bartenschlager, R., 2002. Genetic analysis of sequences in the 3' nontranslated region of hepatitis C virus that are important for RNA replication. J. Virol. 76, 5326-5338.

Friebe, P., Boudet, J., Simorre, J.-P., Bartenschlager, R., 2005. Kissing-loop interaction in the 3' end of the hepatitis C virus genome essential for RNA replication. J. Virol. 79, 380-392.

Frolov, I., Hardy, R., Rice, C.M., 2001. Cis-acting RNA elements at the 5' end of Sindbis virus genome RNA regulate minus- and plus-strand RNA synthesis. RNA 7, 1638-1651.

Gamarnik, A.V., Andino, R., 1997. Two functional complexes formed by KH domain containing proteins with the 5' noncoding region of poliovirus RNA. RNA 3, 882-892.

Gamarnik, A.V., Andino, R., 1998. Switch from translation to RNA replication in a positive-stranded RNA virus. Genes Dev. 12, 2293-2304.

Gamarnik, A.V., Andino, R., 2000. Interactions of viral protein 3CD and poly(rC) binding protein with the 5' untranslated region of the poliovirus genome. J. Virol. 74, 2219-2226.

Garnier, M., Mamoun, R., Bove, J.M., 1980. TYMV RNA replication in vivo: replicative intermediate is mainly single stranded. Virology 104, 357-374.

Gerber, K., Wimmer, E., Paul, A.V., 2001. Biochemical and genetic studies of the initiation of human rhinovirus 2 RNA replication: identification of a cis-replicating element in the coding sequence of $2A^{pro}$. J. Virol. 75, 10979-10990.

Goebel, S.J., Hsue, B., Dombrowski, T.F., Masters, P.S., 2004. Characterization of the RNA components of a putative molecular switch in the 3' untranslated region of the murine coronavirus genome. J. Virol. 78, 669-682.

Goh, P.Y., Tan, Y.J., Lim, S.P., Tan, Y.H., Lim, S.G., Fuller-Pace, F., Hong, W., 2004. Cellular RNA helicase p68 relocalization and interaction with the hepatitis C virus (HCV) NS5B protein and the potential role of p68 in HCV RNA replication. J. Virol. 78, 5288-5298.

Goodfellow, I.G., chaudhry, Y., Richardson, A., Meredith, J., Almond, J.W., Barclay, W., Evans, D.J., 2000. Identification of a cis-acting replication element within the Poliovirus coding region. J. Virol. 74, 4590-4600.

Goodfellow, I.G., Kerrigan, D., Evans, D.J., 2003a. Structure and function analysis of the poliovirus cis-acting replication element (CRE). RNA 9, 124-137.

Goodfellow, I.G., Polacek, C., Andino, R., Evans, D., 2003b. The poliovirus 2C cis-acting replication element-mediated uridylylation of VPg is not required for synthesis of negative-sense genomes. J. Gen. Virol. 84, 2359-2363.

Grdzelishvili, V.Z., Chapman, S.N., Dawson, W.O., Lewandowski, D.J., 2000. Mapping of the Tobacco mosaic virus movement protein and coat protein subgenomic RNA promoters in vivo. Virology 275, 177-192.

Greatorex, J., 2004. The retroviral RNA dimer linkage: different structures may reflect different roles. Retrovirology 1, 22

Groeneveld, H., Thimon, K., van Duin, J., 1995. Translational control of maturation-protein synthesis in phage MS2: a role for the kinetics of RNA folding? RNA 1, 79-88.

Guan, H., Song, C., Simon, A.E., 1997. RNA promoters located on (-)-strands of a subviral RNA associated with turnip crinkle virus. RNA 3, 1401-1412.

Guan, H., Carpenter, C.D., Simon, A.E., 2000a. Analysis of cis-acting sequences involved in plus-strand synthesis of a TCV-associated satellite RNA identifies a new carmovirus replication element. Virology 268, 345-354.

Guan, H., Carpenter, C.D., Simon, A.E., 2000b. Requirement of a 5'-proximal linear sequence on minus-strand for plus-strand synthesis of a satellite RNA associated with TCV. Virology 268, 355-363.

Guan, H., 2000. Replication and 3'-end repair of a subviral RNA associated with turnip crinkle virus. PhD dissertation. University of Massachusetts, Amherst.

Guenther, R.H., Sit, T.L., Gracz, H.S., Dolan, M.A., Townsend, H.L., Liu, G., Newman, W.H., Agris, P.F., Lommel, S.A., 2004. Structural characterization of an intermolecular RNA–RNA interaction involved in the transcription regulation element of a bipartite plant virus. Nucleic Acids Res. 32, 2819–2828.

Guilley, H., Carrington, J.C., Balazs, E., Jonard, G., Richards, K., Morris, T.J., 1985. Nucleotide sequence and genome organization of carnation mottle virus RNA. Nucleic Acids Res. 13, 6663-6677.

Guo, L., Allen, E.M., Miller, W.A., 2001. Base-pairing between untranslated regions facilitates translation of uncapped, nonpolyadenylated viral RNA. Mol. Cell 7, 1103-1109.

Guogas, L.M., Filman, D.J., Hogle, J.M. Gehrke, L., 2004. Cofolding organizes alfalfa mosaic virus RNA and coat protein for replication. Science 306, 2108-2111.

Guogas, L.M., Laforest, S.M., Gehrke, L., 2005. Coat protein activation of alfalfa mosaic virus replication is concentration dependent. J. Virol. 79, 5752-5761.

Haasnoot, P.C.J., Brederode, F.T., Olsthoorn, R.C., Bol, J.F., 2000. A conserved hairpin structure in alfamovirus and bromovirus subgenomic promoters is required for efficient RNA synthesis in vitro. RNA 6, 708-716.

Haasnoot, P.C.J., Olsthoorn, R.C., Bol, J.F., 2002. The Brome mosaic virus subgenomic promoter hairpin is structurally similar to the iron-responsive element and functionally equivalent to the minus-strand core promoter stem-loop C. RNA 8, 110-122.

Hacker, D.L., Petty, I.T.D., Wei, N., Morris, T.J., 1992. Turnip crinkle virus genes required for RNA replication and virus movement. Virology 186, 1-8.

Hahn, C.S., Hahn, Y.S., Rice, C.M., Lee, E., Dalgamo, L., Strauss, E.G., Strauss, J.H., 1987. Conserved elements in the 3’ untranslated region of flavivirus RNAs and potential cyclization sequences. J. Mol. Biol. 198, 33-41.

Haldeman-Cahill, R., Daros, J.A., Carrington, J.C., 1998. Secondary structures in the capsid protein coding sequence and 3’ nontranslated region involved in amplification of the tobacco etch virus genome. J. Virol. 72, 4072-4079.

Hannon, G.J., 2002. RNA interference. Nature 418, 244−251.

Hardy, R.W., Rice, C.M., 2005. Requirements at the 3' end of the sindbis virus genome for efficient synthesis of minus-strand RNA. J. Virol. 79, 4630-4639.

Harris, D.A., Tinsley, R.A., Walter, N.G., 2004. Terbium-mediated footprinting probes a catalytic conformational switch in the antigenomic hepatitis delta virus ribozyme. J. Mol. Biol. 341, 389-403.

He, L., Hannon, G.J., 2004. MicroRNAs: small RNAs with a big role in gene regulation. Nat. Rev. Genet. 5, 522-531.

Heaton, L.A., Lee, T.C., Wei, N., Morris, T.J., 1991. Point mutations in the turnip crinkle virus capsid protein affect the symptoms expressed by Nicotiana benthamiana. Virology 183, 143-150.

Hema, M., Kao, C.C., 2004. Template sequence near the initiation nucleotide can modulate Brome mosaic virus RNA accumulation in plant protoplasts. J. Virol. 78, 1169-1180.

Herold, J., Andino, R., 2001. Poliovirus RNA replication requires genome circularization through a protein-protein bridge. Mol. Cell 7, 581-591.

Huang, M., Koh, D.C., Weng, L.J., Chang, M.L., Yap, Y.K., Zhang, L., and Wong, S.M. 2000. Complete nucleotide sequence and genome organization of hibiscus chlorotic ringspot virus, a new member of the genus Carmovirus: evidence for the presence and expression of two novel open reading frames. J. Virol. 74, 3149-3155.

Hull, R., Davies, J.W., 1983. Genetic engineering with plant viruses, and their potential as vectors. Adv. Virus Res. 28, 1-33.

Huthoff, H., Berkhout, B., 2001. Two alternating structures of the HIV-1 leader RNA. RNA 7,143-157.

Inoue, Y., Miyazaki, M., Ohashi, R., Tsuji, T., Fukaya, K., Kouchi, H., Uemura, T., Mihara, K., Namba, M., 1998. Ubiquitous presence of cellular proteins that specifically bind to the 3' terminal region of hepatitis C virus. Biochem. Biophys. Res. Commun. 245, 198-203.

Ishikawa, M., Meshi, T., Ohno, T., Okada, Y., 1991. Specific cessation of minus-strand RNA accumulation at an early stage of tobacco mosaic virus infection. J. Virol. 65, 861-868.

Isken, O., Grassmann, C.W., Sarisky, R.T., Kann, M., Zhang, S., Grosse, F., Kao, P.N., Behrens, S.E., 2003. Members of the NF90/NFAR protein group are involved in the life cycle of a positive-strand RNA virus. EMBO J. 22, 5655-5665.

Isken, O., Grassmann, C.W., Yu, H., Behrens, S.E., 2004. Complex signals in the genomic 3' nontranslated region of bovine viral diarrhea virus coordinate translation and replication of the viral RNA. RNA 10,1637-1652.

Ito, T., Lai, M.M., 1997. Determination of the secondary structure of and cellular protein binding to the 3'-untranslated region of the hepatitis C virus RNA genome. J Virol. 71, 8698-8706.

Jaeger, J.A, Santalucia, J., Tinoco, I., 1993. Determination of RNA structure and thermodynamics. Annu. Rev. Biochem. 62, 255-287.

Janda, M., Ahlquist, P., 1998. Brome mosaic virus RNA replication protein 1a dramatically increases in vivo stability but not translation of viral genomic RNA3. Proc. Natl. Acad. Sci. U. S. A. 95, 2227-2232.

Johnston, J.C., Rochon, D.M., 1995. Deletion analysis of the promoter for the cucumber necrosis virus 0.9-kb subgenomic RNA. Virology 214, 100-109.

Johnson , R.F., Feng, M., Liu, P., Millership, J.J., Yount, B., Baric, R.S., Leibowitz, J.L., 2005. Effect of mutations in the mouse hepatitis virus 3'(+)42 protein binding element on RNA replication. J. Virol. 14570-14585.

Kao, C.C., Singh, P., Ecker, D.J., 2001. De novo initiation of viral RNA-dependent RNA synthesis. Virology 287, 251-260.

Kasprzak, W., Shapiro, B.A., 1999. Stem Trace: An interactive visual tool for comparative RNA structure analysis. Bioinformatics 15, 16-31.

Ke, A., Zhou, K., Ding, F., Cate, J.H.D., Doudna, J.A., 2004. A conformational switch controls hepatitis delta virus ribozyme catalysis. Nature 429, 201-204.

Khromykh, A.A., Meka, H., Guyatt, K.J., Westaway, E.G., 2001. Essential role of cyclization sequences in flavivirus RNA replication. J. Virol. 75, 6719-6728.

Khromykh, A.A., Kondratieva, N., Sgro, J.-Y., Palmenberg, A., Westaway. E.G., 2003. Significance in replication of the terminal nucleotides of the Flavivirus genome. J. Virol. 77, 10623-10629.

Kim, C.H., Kao, C., and Tinoco, I., 2000. RNA motifs that determine specificity between a viral replicase and its promoter. Nat. Struct. Biol. 7, 415-423.

Kim, C.H., Tinoco, I., 2001. Structural and thermodynamic studies on mutant RNA motifs that impair the specificity between a viral replicase and its promoter. J. Mol. Biol. 307, 827-839.

Kim, K.H., Hemenway, C.L., 1999. Long-distance RNA-RNA interactions and conserved sequence elements affect potato virus X plus-strand RNA accumulation. RNA 5, 636-645.

Klovins, J., Berzins, V., van Duin, J., 1998. A long-range interaction in Q-beta RNA that bridges the thousand nucleotides between the M-site and the 3’ end is required for replication. RNA 4, 948-957.

Klovins, J., van Duin, J., 1999. A long-range pseudoknot in Qβ is essential for replication. J. Mol. Biol. 294, 875-884.

Koev, G., Mohan, B.R., Miller, W.A., 1999. Primary and secondary structural elements required for synthesis of barley yellow dwarf virus subgenomic RNA1. J. Virol. 73, 2876-2885.

Koev, G., Miller, W.A., 2000. A positive-strand RNA virus with three very different subgenomic RNA promoters. J. Virol. 74, 5988-5996.

Koev, G., Liu, S., Beckett, R., Miller, W.A., 2002. The 3'-terminal structure required for replication of Barley yellow dwarf virus RNA contains an embedded 3' end. Virology 292, 114-126.

Kong, Q., Wang, J., Simon, A.E., 1997. Satellite RNA-mediated resistance to turnip crinkle virus in Arabidopsis involves a reduction in virus movement. Plant Cell 9, 2051-2063.

Krab, I.M., Caldwell, C., Gallie, D.R., Bol, J.F., 2005. Coat protein enhances translational efficiency of alfalfa mosaic virus RNAs and interacts with the eIF4G component of initiation factor eIF4F. J. Gen. Virol. 86, 1841-1849.

Lai, M.M.C., 1998. Cellular factors in the transcription and replication of viral RNA genomes: a parallel to DNA-dependent RNA transcription. Virology 244, 1-12.

Lee, Y.-S., Hsu, Y.-H., Lin, N.-S., 2000. Generation of subgenomic RNA directed by a satellite RNA associated with bamboo mosaic potexvirus: Analyses of potexvirus subgenomic RNA promoter. J. Virol. 74, 10341-10348.

Lehmann, K.A., Bass, B.L., 1999. The importance of internal loops within RNA substrates of ADAR1. J. Mol. Biol. 291, 1-13.

Leontis, N.B., Stombaugh, J., Westhof, E., 2002a. Motif prediction in ribosomal RNAs. Lessons and prospects for automated motif prediction in homologous RNA molecules. Biochimie 84, 961-973.

Leontis, N.B., Stombaugh, J., and Westhof, E., 2002b. The non-Watson-Crick base pairs and their associated isostericity matrices. Nucleic Acids Res. 30, 3497-3531.

Li , W.-Z, Qu, F., Morris, T.J., 1998. Cell-to-cell movement of turnip crinkle virus is controlled by two small open reading frames that function in trans. Virology 244, 405-416.

Li, X.H., Heaton, L.A., Morris, T.J., Simon, A.E., 1989. Turnip crinkle virus defective interfering RNAs intensify viral symptoms and are generated de novo. Proc. Natl. Acad. Sci. U. S. A. 86, 9173-9177.

Li, X.H., Simon, A.E., 1991. In vivo accumulation of a turnip crinkle virus defective interfering RNA is affected by alterations in size and sequence. J. Virol. 65, 4582-4590.

Lin, B., Heaton, L.A., 2001. An Arabidopsis thaliana protein interacts with a movement protein of turnip crinkle virus in yeast cells and in vitro. J. Gen. Virol. 82, 1245-1251.

Lin H.-X., White, K.A., 2004. A complex network of RNA-RNA interactions controls subgenomic mRNA transcription in a tombusvirus. EMBO J. 23, 3365-3374.

Lin, Y.-J., Liao, C.-L., Lai, M.M.C., 1994. Identification of the cis-acting signal for minus-strand RNA synthesis of a murine coronavirus: implications for the role of minus-strand RNA in RNA replication and transcription. J. Virol. 68, 8131-8140.

Lindenbach, B.D., Sgro, J.-Y., Ahlquist, P., 2002. Long-distance base pairing in flock house virus RNA1 regulates subgenomic RNA3 synthesis and RNA2 replication. J. Virol. 76, 3905-3919.

Liu, Q., Yu, W., Leibowitz, J.L., 1997. A specific host cellular protein binding element near the 3' end of mouse hepatitis virus genomic RNA. Virology 232, 74-85.

Lo, M.K., Tilgner, M., Bernard, K.A., Shi, P.Y., 2003. Functional analysis of mosquito-borne flavivirus conserved sequence elements within 3' untranslated region of West Nile virus by use of a reporting replicon that differentiates between viral translation and RNA replication. J. Virol. 77,10004-10014.

Lobert, P.-E., Escriou, N., Ruelle, J., Michiels, T., 1999. A coding RNA sequence acts as a replication signal in cardioviruses. Proc. Natl. Acad. Sci. U. S. A. 96, 11560-11565.

Lyons, T., Murray, K.E., Roberts, A.W, Barton, E.J., 2001. Poliovirus 5' terminal cloverleaf RNA is required in cis for VPg uridylylation and the initiation of negative-strand RNA synthesis. J. Virol. 75, 10696-10708.

McCormack, J.C., Simon, A.E., 2004. Biased hypermutagensis associated with mutations in an untranslated hairpin of an RNA virus. J. Virol. 78, 7813-7817.

Mandal, M., Breaker, R.R., 2004. Gene regulation by riboswitches. Nature Rev. Mol. Cell Biol. 5, 451-463.

Marsh, L.E., Dreher, T.W., Hall, T.C., 1988. Mutational analysis of the core and modulator sequences of the BMV RNA3 subgenomic promoter. Nucleic Acids Res. 16, 981-995.

Marsh, L.E., Huntley, C.C., Pogue, G.P., Connell, J.P., Hall, T.C., 1991. Regulation of (+):(-)-strand asymmetry in replication of brome mosaic virus RNA. Virology 182, 76-83.

Mason, P.W., Bezborodova, S.V., Henry, T.M., 2002. Identification and characterization of a cis-acting replication element (cre) adjacent to the internal ribosome entry site of Foot-and-mouth disease virus. J. Virol. 76, 9686-9694.

Matsuda, D., Dreher T.W., 2004. The tRNA-like structure of turnip yellow mosaic virusis a 3'-translational enhancer. Virology 321, 36-46.

Matsuda, D., Yoshinari, S., Dreher, T.W., 2004. eEF1A binding to aminoacylated viral RNA represses minus strand synthesis by TYMV RNA-dependent RNA polymerase. Virology 321, 47-56.

Mcknight, K.L., Lemon, S.M., 1996. Capsid coding sequence is required for efficient replication of human rhinovirus 14 RNA. J. Virol. 70, 1941-1952.

Mcknight, K.L., Lemon, S.M., 1998. The rhinovirus type 14 genome contains an internally located RNA structure that is required for viral replication. RNA 4, 1569-1584.

McManus, M.T., Sharp, P.A., 2002. Gene silencing in mammals by small interfering RNAs. Nature Rev. Genetics 3, 737-747.

Melchers, W.J.G., Henderop, J.G.J., Slot, H.J.B., Pleij, C.W.A., Pilipenko, E.V., Agol, V.L., Galama, J.M.D., 1997. Kissing of two predominant hairpin loops in the coxsackie B virus 3' untranslated region is the essential structural feature of the origin of replication required for negative-strand RNA synthesis. J. Virol. 71, 686-696.

Miller, D.J., Schwartz, M.D., and Ahlquist, P., 2001. Flock house virus RNA replicates on outer mitochondrial membranes in Drosophila cells. J. Virol. 75, 11664-11676.

Miller, W.A., Koev, G., 2000. Synthesis of subgenomic RNAs by positive-strand RNA viruses. Virology 273, 1-8.

Monkewich, S., Lin, H.-X., Fabian, M.R., Xu, W., Na, H., Ray, D., Chernysheva, O.A., Nagy, P., White, K.A., 2005. The p92 polymerase coding region contains an internal RNA element required at an early step in Tombusvirus genome replication. J. Virol. 79, 4848-4858.

Moon, J.S., McCoppin, N.K., Domie, L.L., 2001. Effect of mutations in barley yellow dwarf virus genomic RNA on the 5' termini of subgenomic RNAs. Arch. Virol. 146, 1399-1406.

Moore, P.B., 1999. Structural motifs in RNA. Annu. Rev. Biochem. 68, 287-300.

Morales, M., Barcena, J., Ramirez, M.A., Boga, J.A., Parra, F., Torres, J.M., 2004. Synthesis in vitro of rabbit hemorrhagic disease virus subgenomic RNA by internal

initiation on (-)sense genomic RNA: mapping of a subgenomic promoter. J. Biol. Chem. 279, 17013-17018.

Morasco, B.J., Sharma, N., Parilla, J., Flanegan, J.B., 2003. Poliovirus cre(2C)-dependent synthesis of VPgpUpU is required for positive- but not negative-strand RNA synthesis. J. Virol. 77, 5136-5144.

Murray, K.E., Barton, D.J., 2003. Poliovirus CRE-dependent VPg uridylylation is required for positive-strand RNA synthesis but not for negative-strand RNA synthesis. J. Virol. 77, 4739-4750.

Na, H., White, A., 2006. Structure and prevalence of replication silencer-3' terminus RNA interactions in Tombusviridae. Virology 345, 305-316.

Nagashima, S., Sasaki, J., Taniguchi, K., 2005. The 5'-terminal region of the Aichi virus genome encodes cis-acting replication elements required for positive- and negative-strand RNA synthesis. J. Virol. 79, 6918-6931.

Nagel, J.H.A., Pleij, C.W.A., 2002. Self-induced structural switches in RNA. Biochimie 84, 913-923.

Nagy, P.D., Carpenter, C.D., Simon, A.E., 1997. A novel 3'-end repair mechanism in an RNA virus. Proc. Natl. Acad. Sci. U. S. A. 94, 1113-1118.

Nagy, P.D., Zhang, C., Simon, A.E., 1998. Dissecting RNA recombination in vitro: role of RNA sequences and the viral replicase. EMBO J. 17, 2392-2403.

Nagy, P.D., Pogany, J., Simon, A.E., 1999. RNA elements required for RNA recombination function as replication enhancers in vitro and in vivo in a plus-strand RNA virus. EMBO J. 18, 5653-5665.

Nagy, P.D., Pogany, J., Simon, A.E., 2001. In vivo and in vitro characterization of an RNA replication enhancer in a satellite RNA associated with turnip crinkle virus. Virology 288, 315-324.

Nanda, S. K., Leibowitz, J. L., 2001. Mitochondrial aconitase binds to the 3' untranslated region of the mouse hepatitis virus genome. J. Virol. 75, 3352-3362.

Nanda, S. K., Johnson, R.F., Liu, Q., Leibowitz, J.L., 2004. Mitochondrial HSP70, HSP40, and HSP60 bind to the 3' untranslated region of the murine hepatitis virus genome. Arch. Virol. 149, 93-111.

Nayak, A., Goodfellow, I.G., Belsham, G.J., 2005. Factors required for the uridylylation of the foot-and-mouth disease virus 3B1, 3B2, and 3B3 peptides by the RNA-dependent RNA polymerase ($3D^{pol}$) in vitro. J. Virol. 79, 7698-7706.

Nomaguchi, M., Teramoto, T., Yu, L., Markoff, L., Padmanabhan, R., 2004. Requirement for West Nile virus (-)- and (+)-strand subgenomic RNA synthesis in vitro by the viral RNA-dependent RNA polymerase expressed in Escherichia coli. J. Biol. Chem. 279, 12141-12151.

Noueiry, A.O., Ahlquist, P., 2003. Brome mosaic virus RNA replication: revealing the role of the host in RNA virus replication. Annu. Rev. Phytopathol. 41, 01.1-01.22.

Nudler, E., Mironov, A.S., 2004. The riboswitch control of bacterial metabolism. Trends Biochem. Sci. 29, 11-17.

Oh, J.-W., Kong, Q., Song, C., Carpenter, C.D., Simon, A.E., 1995. Open reading frame of turnip crinkle virus involved in satellite symptom expression and incompatibility with Arabidopsis thaliana ecotype Dijon. Mol. Plant-Microbe Interact. 8, 979-987.

Olsthoorn, R.C., Mertens, S., Brederode, F.T., Bol, J.F., 1999. A conformational switch at the 3' end of a plant virus RNA regulates viral replication. EMBO J. 18, 4856-4864.

Olsthoorn, R.C., Bol, J.F., 2001. Sequence comparison and secondary structure analysis of the 3' noncoding region of flavivirus genomes reveals multiple pseudoknots. RNA 7, 1370-1377.

Olsthoorn, R.C., Bol, J.F., 2002. Role of an essential triloop hairpin and flanking structures in the 3' untranslated region of Alfalfa mosaic virus RNA in in vitro transcription. J. Virol. 76, 8747-8756.

Olsthoorn, R.C., Haasnoot, P.C., Bol, J.F., 2004. Similarities and differences between the subgenomic and minus-strand promoters of an RNA plant virus. J. Virol. 78, 4048-4053.

Ooms, M., Huthoff, H., Russell, R., Liang, C., Berkhout, B., 2004a. A riboswitch regulates RNA dimerization and packaging in human immunodeficiency virus type 1 virions. J. Virol. 78, 10814-10819.

Ooms, M., Verhoef, K., Southern, E., Huthoff, H., Berkhout, B., 2004b. Probing alternative foldings of the HIV-1 leader RNA by antisense oligonucleotide scanning arrays. Nucleic Acids Res. 32, 819-827.

Osman, T.A., Hemenway, C.L., Buck, K.W., 2000. Role of the 3’ tRNA-like structure in tobacco mosaic virus minus-strand RNA synthesis by the viral RNA-dependent RNA polymerase in vitro. J. Virol. 74, 11671-11680.

Pacha, R.F., Ahlquist, P., 1991. Use of bromovirus RNA3 hybrids to study template specificity in viral RNA amplification. J. Virol. 65, 3693-3703.

Paillart, J.-C., Dettenhofer, M., Yu, X.-F., Ehresmann, C., Ehresmann, B., Marquet, R., 2004. First snapshots of the HIV-1 RNA structure in infected cells and in virions. J. Biol. Chem. 279, 48397-48403.

Pan, J., Thirumalai, D., Woodson, S.A., 1997. Folding of RNA involves parallel pathways. J. Mol. Biol. 273,7-13.

Panavas, T., Pogany, J., Nagy, P.D., 2002. Analysis of minimal promoter sequences for plus-strand synthesis by the cucumber necrosis virus RNA-dependent RNA polymerase. Virology 296, 263-274.

Panavas, T., Nagy, P.D., 2003. The RNA replication enhancer element of tombusvirus contains two interchangeable hairpins that are functional during plus-strand synthesis. J. Virol. 77, 258-269.

Panavas, T., Nagy, P.D., 2005. Mechanism of stimulation of plus-strand synthesis by an RNA replication enhancer in a tombusvirus. J. Virol. 79, 9777–9785.

Panaviene, Z., Panavas, T., Serva, S., Nagy, P.D., 2004. Purification of the cucumber necrosis virus replicase from yeast cells: role of coexpressed viral RNA in stimulation of replicase activity. J. Virol. 78, 8254-8263.

Panaviene, Z., Panavas, T., Nagy, P.D., 2005. Role of an internal and two 3'-terminal RNA elements in assembly of Tombusvirus replicase. J. Virol. 79, 10608-10618.

Parsley, T.B., Towner, J.S., Blyn, L.B., Ehrenfeld, E., Semler, B.L., 1997. Poly(rC) binding protein 2 forms a ternary complex with the 5'-terminal sequences of poliovirus RNA and the viral 3CD proteinase. RNA 3, 1124-1134.

Pasternak, A.O., van der Born, E., Spaan, W.J.M., Snijder, E.J., 2003. The stability of the duplex between sense and antisense transcription-regulating sequences is a crucial factor in arterivirus subgenomic mRNA synthesis. J. Virol. 77, 1175-1183.

Pathak, H.B., Ghosh, S.K.B., Roberts, A.W., Sharma, S.D., Yoder, J.D., Arnold, J.J., Gohara, D.W., Barton, D.J., Paul, A.V., Cameron, C.E., 2002. Structure-function relationships of the RNA-dependent RNA polymerase from poliovirus ($3D^{pol}$). J. Biol. Chem. 277, 31551-31562

Paul, A. V., Rieder, E., Kim, D.W., van Boom, J. H., Wimmer, E., 2000. Identification of an RNA hairpin in poliovirus RNA that serves as the primary template in the in vitro uridylylation of VPg. J. Virol. 74, 10359-10370

Paul, A.V., Yin, J., Mugavero, J., Rieder, E., Liu, Y., Wimmer, E., 2003. A "slide-back" mechanism for the initiation of protein-primed RNA synthesis by the RNA polymerase of poliovirus. J. Biol. Chem. 278, 43951-43960.

Petrillo, J.E., Rocheleau, G., Kelly-Clarke, B., Gehrke, L., 2005. Evaluation of the conformational switch model for alfalfa mosaic virus RNA replication. J. Virol. 79, 5743-5751.

Pillai-Nair, N., Kim, K.H., Hemenway, C., 2003. Cis-acting regulatory elements in the Potato virus X 3' non-translated region differentially affect minus-strand and plus-strand RNA accumulation. J. Mol. Biol. 326, 701-720.

Pogany, J., Fabian, M.R., White, K.A., Nagy, P.D., 2003. Functions of novel replication enhancer and silencer elements in tombusvirus replication. EMBO J. 22, 5602-5611.

Pogany, J., White, K.A., Nagy, P.D., 2005. Specific binding of tombusvirus replication protein p33 to an internal replication element in the viral RNA is essential for replication. J. Virol. 79, 4859-4869.

Poot, R.A., Tsareva, N.V., Boni, I.V., van Duin, J., 1997. RNA folding kinetics regulates translation of of phage MS2 maturation gene. Proc. Natl. Acad. Sci. U. S. A.. 94, 10110-10115.

Poranen, M.M., Daugelavicius, R., Bamford, D.H., 2002. Common principles in viral entry. Annu. Rev. Microbiol. 56, 521-538.

Qu, F., Ren, T., Morris, T.J., 2003. The coat protein of turnip crinkle virus suppresses posttranscriptional gene silencing at an early initiation step. J. Virol. 77, 511-522.

Quadt, R., Ishikawa, M., Janda, M., Ahlquist, P., 1995. Formation of Brome mosaic virus RNA-dependent RNA polymerase in yeast requires coexpression of viral proteins and viral RNA. Proc. Natl. Acad. Sci. U. S. A.. 92, 4892-4896.

Rajendran, K.S., Pogany, J., Nagy, P.D., 2002. Comparison of turnip crinkle virus RNA-dependent RNA polymerase preparations expressed in Escherichia coli or derived from infected plants. J. Virol. 76, 1707-1717.

Ranjith-Kumar, C.T., Zhang, X., Kao, C.C., 2003. Enhancer-like activity of a brome mosaic virus RNA promoter. J. Virol. 77, 1830-1839.

Rao, A.L., Hall, T.C., 1991. Interference in trans with brome mosaic virus replication by RNA-2 bearing aminoacylation-deficient mutants. Virology 180, 16-22.

Ray, D., White, K.A., 1999. Enhancer-like properties of an RNA element that modulates Tombusvirus RNA accumulation. Virology 256, 162-171.

Ray, D., White, K.A., 2003. An internally located RNA hairpin enhances replication of tomato bushy stunt virus RNAs. J. Virol. 77, 245-257.

Ray, D., Wu, B., White, K.A., 2003. A second functional RNA domain in the 5′ UTR of the Tomato bushy stunt virus genome: intra- and interdomain interactions mediate viral RNA replication. RNA 9, 1232-1245.

Ray, D., Na, H., White, K.A., 2004. Structural properties of a multifunctional T-shaped RNA domain that mediate efficient tomato bushy stunt virus RNA replication. J. Virol. 78, 10490-10500.

Revera, V.M., Welsh, J.D., Maizel Jr., J.V., 1988. Comparative sequence analysis of the 5' noncoding region of the enteroviruses and rhinoviruses. Virology 165, 42-50.

Rico, P., Hernandez, C., 2004. Complete nucleotide sequence and genome organization of Pelargonium flower break virus. Arch. Virol. 149, 641-651.

Rieder, E., Paul, A.V., Kim, D.W., van Boom, J.H., Wimmer, E., 2000. Genetic and biochemical studies of poliovirus cis-acting replication element cre in relation to VPg uridylylation. J. Virol. 74, 10371-10380.

Riviere, C.J., Rochon, D.M., 1990. Nucleotide sequence and genomic organization of melon necrotic spot virus. J. Gen. Virol. 71, 1887-1896.

Rohll, J.B., Percy, N., Ley, R., Evans, D.J., Almond, J.W., Barclay, W.S., 1994. The 5'-untranslated regions of picornavirus RNAs contain independent functional domains essential for RNA replication and translation. J. Virol. 68, 4384-4391.

Rust, R.C., Landmann, L., Gosert, R., Tang, B.L., Hong, W., Hauri, H.P., Egger, D., Bienz, K., 2001. Cellular COPII proteins are involved in production of the vesicles that form the poliovirus replication complex. J. Virol. 75, 9808-9818.

Satyanarayana, T., Gowda, S., Ayllon, M.A., Albiach-Marti, M.R., Rabindran, S., Dawson, W.O., 2002. The p23 protein of citrus tristeza virus controls asymmetrical RNA accumulation. J. Virol. 76, 473-483.

Sawicki, S.G., Sawicki, D.L., 2005. Coronavirus transcription: a perspective. Curr. Top. Microbiol. Immunol. 287, 31-55.

Schirawski, J., Voyatzakis, A., Zaccomer, B., Bernardi, F., Haenni, A.L., 2000. Identification and functional analysis of the turnip yellow mosaic tymovirus subgenomic promoter. J. Virol. 74, 11073-11080.

Schuppli, D., Miranda, G., Tsui, H.C., Winkler, M.E., Sogo, J.M., Weber, H., 1997. Altered 3' terminal RNA structure in phage Qβ adapted to host factor less Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 94, 10239-10242

Schuppli, D., Georgijevic, J., Weber, H., 2000. Synergism of mutations in bacteriophage Qβ RNA affecting host factor dependence of Qβ replicase. J. Mol. Biol. 295, 149-154.

Schwartz, M., Chen, J., Janda, M., Sullivan, M., den Boon, J., Ahlquist, P., 2002. A positive-strand RNA virus replication complex parallels form and function of retrovirus capsids. Mol. Cell 9, 505-514.

Shapiro, B.A., Wu, J.-C., Bengali, D., Potts, M., 2001. The massively parallel genetic algorithm for RNA folding: MIMD implementation and population variation. Bioinformatics 17, 137-148.

Shen, R, Miller, W.A., 2004. Subgenomic RNA as a riboregulator: negative regulation of RNA replication by barley yellow dwarf virus subgenomic RNA 2. Virology 327, 196-205.

Shi, P.Y., Brinton, M.A., Veal, J.M., Zhong, Y.Y., Wilson, W.D., 1996. Evidence for the existence of a pseudoknot structure at the 3' terminus of the flavivirus genomic RNA. Biochemistry 35, 4222-4230.

Siegel, R.W., Adkins, S., Kao, C.C., 1997. Sequence specific recognition of subgenomic promoter by a viral RNA polymerase. Proc. Natl. Acad. Sci. U. S. A. 94, 11238-11243.

Siegel, R.W., Bellon, L., Beigelman, L., Kao, C.C., 1998. Moieties in an RNA promoter specifically recognized by a viral RNA-dependent RNA polymerase. Proc. Natl. Acad. Sci. U. S. A. 95, 11613-11618.

Simoes, E.A., Sarnow, P. 1991. An RNA hairpin at the extreme 5’ end of the poliovirus RNA genome modulates viral translation in human cells. J.Virol. 65, 913-921.

Simon, A.E., Howell, S.H., 1986. The virulent satellite RNA of turnip crinkle virus has a major domain homologous to the 3’ end of the helper virus genome. EMBO J. 5, 3423-3428.

Simon, A.E., Engel, H., Johnson, R.P., Howell, S.H., 1988. Identification of regions affecting virulence, RNA processing and infectivity in the virulent satellite of turnip crinkle virus. EMBO J. 7, 2645-2651.

Simon A.E., Roossinck, M.J., Havelda, Z., 2004. Plant virus satellite and defective interfering RNAs: new paradigms for a new century. Annu. Rev. Phytopathol. 42, 415-447.

Singh, R.N., Dreher, T.W., 1997. Turnip yellow mosaic virus RNA-dependent RNA polymerase: Initiation of negative strand synthesis in vitro. Virology 233, 430-439.

Singh, R.N., Dreher, T.W., 1998. Specific site selection in RNA resulting from a combination of nonspecific secondary structure and -CCR- boxes: initiation of minus strand synthesis by turnip yellow mosaic virus RNA-dependent RNA polymerase. RNA 4, 1083-1095.

Sit, T.L., Vaewhongs, A.A., and Lommel, S.A., 1998. RNA-mediated trans-activation of transcription from a viral RNA. Science 281, 829-832.

Sivakumaran, K., Kao, C.C., 1999. Initiation of genemic plus-strand RNA synthesis from DNA and RNA templates by a viral RNA-dependent RNA polymerase. J.Virol. 73, 6415-6423.

Sivakumaran, K., Kim, C., Tayon Jr., R., Kao, C.C., 1999. RNA sequence and secondary structural determinants in a minimal viral promoter that directs replicase recognition and initiation of genomic plus-strand RNA synthesis. J. Mol. Biol. 294, 667-682.

Sivakumaran, K., Bao, Y., Roossinck, M.J., Kao, C.C., 2000. Recognition of the core RNA promoter for minus-strand RNA synthesis by replicases of brome mosaic virus and cucumber mosaic virus. J.Virol. 74, 10323-10331.

Sivakumaran, K., Kao, C.C., 2000. Genomic plus-strand RNA synthesis by brome mosaic virus (BMV) RNA replicase requires a sequence that is complementary to the binding site of the BMV helicase-like protein. Mol. Plant Pathol. 1, 337-346.

Sivakumaran, K., Hema, M., Kao, C.C., 2003. Brome mosaic virus RNA synthesis in vitro and in barley protoplasts. J.Virol. 77, 5703-5711.

Sivakumaran, K., Choi, S.K., Hema, M., Kao, C.C., 2004. Requirements for brome mosaic virus subgenomic RNA synthesis in vivo and replicase-core promoter interactions in vitro. J Virol. 78, 6091-6101.

Skotnicki, M.L., Mackenzie, A.M., Torronen, M., Gibbs, A.J., 1993. The genomic sequence of Cardamine chlorotic fleck carmovirus. J. Gen. Virol. 74, 1933-1937.

Song, C., Simon, A.E., 1994. RNA-dependent RNA polymerase from plants infected with turnip crinkle virus can transcribe (+)- and (-)-strands of virus-associated RNAs. Proc. Natl. Acad. Sci. U. S. A.. 91, 8792-8796.

Song, C., Simon, A.E., 1995. Requirement of a 3'-terminal stem-loop in in vitro transcription by an RNA dependent RNA polymerase. J. Mol. Biol. 254, 6-14.

Song, C., 1995. In vitro study of RNA replication using isolated viral RNA-dependent RNA polymerase. PhD dissertation, University of Massachusetts Amherst.

Song, S.I., Silver, S.L., Aulik, M.A., Rasochova, L., Mohan, B.R., Miller, W.A., 1999. Satellite Cereal yellow dwarf virus-RPV (satRPV) RNA requires a douXble hammerhead for self-cleavage and an alternative structure for replication. J. Mol. Biol. 293, 781-793.

Song, S.I., Miller, W.A., 2004. Cis and trans requirements of rolling circle replication of a satellite RNA. J. Virol. 78, 3072-3082.

Spagnolo, J.F., Hogue, B.G., 2000. Host protein interactions with the 3' end of bovine coronavirus RNA and the requirement of the poly(A) tail for coronavirus defective genome replication. J. Virol. 74, 5053-5065.

Specht, T., Wolters, J., Erdmann, V.A., 1991. Compilation of 5S rRNA and 5S rRNA gene sequences. Nucleic Acids Res. 19 Suppl., 2189-2191.

Stawicki, S.S., Kao, C.C., 1999. Spatial perturbations within an RNA promoter specifically recognized by a viral RNA-dependent RNA polymerase (RdRp) reveal that RdRp can adjust its promoter binding sites. J. Virol. 73, 198-204.

Strauss, J.H., Strauss, E.G., 1994. The alphaviruses: gene expression, replication, evolution. Microbiol. Rev. 58, 491-562.

Strauss, J.H., Strauss, E.G., 1999. With a little help from the host. Science 283, 802-4.

Stupina, V., Simon, A.E., 1997. Analysis in vivo of turnip crinkle virus satellite RNA c variants with mutations in the 3'-terminal minus-strand promoter. Virology 238, 470-477.

Sudarsan, N., Barrick, J.E., Breaker, R.R. 2003. Metabolite-binding RNA domains are present in the genes of eukaryotes. RNA 9, 644-647.

Sullivan, M., Ahlquist, P., 1999. A brome mosaic virus intergenic RNA3 replication signal functions with viral replication protein 1a to dramatically stabilize RNA in vivo. J. Virol.73, 2622-2632.

Sun, X., Simon, A.E., 2003. Fitness of a turnip crinkle virus satellite RNA correlates with a sequence-nonspecific hairpin and flanking sequences that enhance replication and repress the accumulation of virions. J. Virol. 77, 7880-7889.

Sun, X., Zhang, G., Simon, A.E., 2005. Short internal sequences involved in RNA replication and virion accumulation in a subviral RNA of turnip crinkle virus. J. Virol. 79, 512-524.

Sun, X., Simon, A.E., A cis-acting element functions in both orientations to enhance replication of turnip crinkle virus. In press.

Suzuki, S., Hase, S., Takahashi, H., Ikegami, M., 2002. The genome organization of pea stem necrosis virus and its assignment to the genus Carmovirus. Intervirology 45, 160-163.

Szymanski, M., Barciszewska, M.Z., Erdmann, V.A., Barciszewski, J., 2005. A new frontier for molecular medicine: Noncoding RNAs. Biochim Biophys Acta. 1756, 65-75.

Takamatsu, N., Watanabe, Y., Meshi, T., Okada, Y., 1990. Mutational analysis of the pseudoknot region in the 3' noncoding region of tobacco mosaic virus RNA. J. Virol. 64, 3686-3693.

Takemoto, Y., Kanehira, T., Shinohara, M., Yamashita, S., Hibi, T., 2000. The nucleotide sequence and genome organization of Japanese iris necrotic ring virus, a new species in the genus Carmovirus. Arch. Virol. 145, 651-657.

Tang, R.S., Draper, D.E., 1994. Bend and helical twist associated with a symmetric internal loop from 5S ribosomal RNA. Biochemistry 33, 10089-10093.

Tatsuta, M., Mizumoto, H., Kaido, M., Mise, K., Okuno, T., 2005. The red clover necrotic mosaic virus RNA2 trans-activator is alao a cis-acting RNA2 replication element. J. Virol. 79, 978-986.

Teterina, N.L., Rinaudo, M.S., Ehrenfeld, E., 2003. Strand-specific RNA synthesis defects in a poliovirus with a mutation in protein 3A. J. Virol. 77, 12679-12691.

Thomas, C.L., Leh, V., Lederer, C., Maule, A.J., 2003. Turnip crinkle virus coat protein mediates suppression of RNA silencing in Nicotiana benthamiana. Virology 306, 33-41.

Tilgner, M., Shi, P.-Y., 2004. Structure and function of the 3' terminal six nucleotides of the West Nile virus genome in viral replication. J. Virol. 78, 8159-8171.

Tinsley, R.A., Harris, D.A., Walter, N.G., 2004. Magnesium dependence of the amplified conformational switch in the trans-acting hepatitis delta virus ribozyme. Biochemistry 43, 8935-8945.

Todd, S., Towner, J.S., Brown, D.M., Semler, B.L., 1997. Replication-competent picornaviruses with complete genomic RNA 3' noncoding region deletions. J. Virol. 71, 8868-8874.

Tomita, Y., Mizuno, T., Diez, J., Naito, S., Ahlquist, P., Ishikawa, M., 2003. Mutation of host DnaJ homolog inhibits brome mosaic virus negative-strand RNA synthesis. J. Virol. 77, 2990-2997.

Traynor, P., Young, B.M., Ahlquist, P., 1990. Use of bromovirus RNA2 hybrids to map cis- and trans-acting functions in a conserved RNA replication gene. J. Virol. 64, 69-77.

Tsuchihara, K., Tanaka, T., Hijikata, M., Kuge, S., Toyoda, H., Nomoto, A., Yamamoto, N., Shimotohno, K., 1997. Specific interaction of polypyrimidine tract-binding protein with the extreme 3'-terminal structure of the hepatitis C virus genome, the 3'X. J. Virol. 71, 6720-6726.

Tucker, B.J., Breaker, R.R., 2005. Riboswitches as versatile gene control elements. Curr. Opin. Struct. Biol. 15, 342-348.

van den Born, E., Gultyaev, A.P., Snijder, E.J., 2004. Secondary structure and function of the 5'-proximal region of the equine arteritis virus RNA genome. RNA 10, 424-437.

van den Born, E., Posthuma, C.C., Gultyaev, A.P., Snijder, E.J., 2005. Discontinuous subgenomic RNA synthesis in Arteriviruses is guided by an RNA hairpin structure located in the genomic leader region. J. Virol. 79, 6312-6324.

van der Heijden, M.W., Bol, J.F., 2002. Composition of alphavirus-like replication complexes: involvement of virus and host encoded proteins. Arch. Virol. 147, 875-898.

van der Kuyl, A.C., Neeleman, L., Bol, J.F., 1991. Role of alfalfa mosaic virus coat protein in regulation of the balance between viral plus and minus strand RNA synthesi. Virology 185, 496-499.

van Dijk, A.A., Makeyev, E.V., Bamford, D.H., 2004. Initiation of viral RNA-dependent RNA polymerization. J. Gen. Virol. 85, 1077-1093.

van Marle, G., Dobbe, J.C., Gultyaev, A.P., Luytjes, W., Spaan, W.J.M., Snijder, E.J., 1999. Arterivirus discontinuous mRNA transcription is guided by base pairing between sense and antisense transcription-regulating sequences. Proc. Natl. Acad. Sci. U. S. A. 96, 12056-12061.

van Meerten, D., Girade, G., van duin, J., 2001. Translational control by delayed RNA folding: identification of the kinetic trap. RNA 7, 483-494.

van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R., Wickner, R.B. Eds. 2000. Virus taxonomy: Seventh report of the international committee on taxonomy of viruses (San Diego, CA: Acdemic Press).

van Rossum, C.M.A., Reusken, C.B.E.M., Brederode, F., Bol, J.F., 1997. The 3' untranslated region of alfalfa mosaic virus RNA3 contains a core promoter for minus-strand RNA synthesis and an enhancer element. J. Gen. Virol. 78, 3045-3049.

Verheije, M.H., Olsthoorn, R.C.L., Kroese, M.V., Rottier, P.J.M., Meulenberg, J.J.M., 2002. Kissing interaction between 3' noncoding and coding sequences is essential for porcine arterivirus RNA replication. J. Virol. 76, 1521-1526.

Vlot, A.C., Neeleman, L., Linthorst, H.J.M., Bol, J.F., 2001. Role of the 3'-untranslated regions of Alfalfa mosaic virus RNA in the formation of a transiently expressed replicase in plants and in the assembly of virions. J. Virol. 75, 6440-6449.

Vlot, A.C., Bol, J.F., 2003. The 5′ untranslated region of alfalfa mosaic virus RNA1 is involved in negative-strand RNA synthesis. J. Virol. 77, 11284-11289.

Wang, H.-H., Wong, S.-M., 2004. Significantce of the 3'-terminal region in minus-strand RNA synthesis of Hibiscus chlorotic ringspot virus. J. Gen. Virol. 85, 1763-1776.

Wang, J., Simon, A.E., 1997. Analysis of the two subgenomic RNA promoters for turnip crinkle virus in vivo and in vitro. Virology 26, 174-186.

Wang, J., Simon, A.E., 2000. 3'-End stem-loops of the subviral RNAs associated with turnip crinkle virus are involved in symptom modulation and coat protein binding. J. Virol. 74, 6528-6537.

Wang, Z., Day, N., Trifillis, P., Kiledjian, M., 1999. An mRNA stability complex functions with poly(A)-binding protein to stabilize mRNA in vitro. Mol. Cell. Biol. 19, 4552-4560.

Weng, Z.M., Xiong, Z.G., 1997. Genome organization and gene expression of saguaro cactus carmovirus. J. Gen. Virol. 78, 525-534.

White, K.A., Bancroft, J.B., Mackie, G.A., 1992. Mutagenesis of a hexanucleotide sequence conserved in potexvirus RNAs. Virology 189, 817-820.

White, K.A., Skuzeski, J.M., Li, W., Wei, N., Morris, T.J., 1995. Immunodetection, expression strategy and complementation of turnip crinkle virus p28 and p88 replication components. Virology 211, 525-534.

White, K.A., Morris, T.J., 1999. Defective and defective interfering RNAs of monopartite plus-strand RNA plant viruses. Curr. Top. Microbiol. Immunol. 239, 1-17.

White, K.A., 2002. The premature termination model: a possible third mechanism for subgenomic mRNA transcription in (+)-strand RNA viruses. Virology 304, 147-154.

Williams, G.D., Chang, R.Y., Brian, D.A., 1999. A phylogenetically conserved hairpin-type 3' untranslated region pseudoknot functions in coronavirus RNA replication. J. Virol. 73, 8349-8355.

Wood, J., Frederickson, R.M., Fields, S., Patel, A.H., 2001. Hepatitis C virus 3'X region interacts with human ribosomal proteins. J. Virol. 75, 1348-1358.

Woodson, S.A., 2005. Metal ions and RNA folding: a highly charged topic with a dynamic future. Curr. Opin. Chem. Biol. 9, 104-109.

Wu, B., White, K.A., 1998. Formation and amplification of a novel defective tombusvirus RNA which lacks the 5' nontranslated region of the viral genome. J. Virol. 72, 9897-9905.

Wu, B., Vanti, W.B., White, K.A., 2001. An RNA domain within the 5' untranslated region of tomato bushy stunt virus genome modulates viral RNA replication. J. Mol. Biol. 305, 741-756.

Yamashita, T., Kaneko, S., Shirota, Y., Qin, W., Nomura, T., Kobayashi, K., Murakami, S., 1998. RNA-dependent RNA polymerase activity of the soluble recombinant hepatitis C virus NS5B protein truncated at the C-terminal region. J. Biol. Chem. 273, 15479-15486.

Yang, Y., Rijnbrand, R., McKnight, K.L., Wimmer, E., Paul, A., Martin, A., Lemon, S.M., 2002. Sequence requirements for viral RNA replication and VPg uridylylation directed by the internal cis-acting replication element (cre) of human rhinovirus type 14. J. Virol. 76, 7485-7494.

Yi, M., Lemon, S.M., 2003. 3′ nontranslated RNA signals required for replication of hepatitis C virus RNA. J. Virol. 77, 3557-3568.

Yin, J., Paul, A.V., Wimmer, E., Rieder, E., 2003. Functional dissection of a poliovirus cis-acting replication element [PV-cre(2C)]: analysis of single- and dual-cre viral genomes and proteins that bind specifically to PV-cre RNA. J. Virol. 77, 5152-5166.

Yoshinari, S., Nagy, P.D., Simon, A.E., Dreher, T.W., 2000. CCA initiation boxes without unique promoter elements support in vitro transcription by three viral RNA-dependent RNA polymerase. RNA 6, 698-707.

You, S., Padmanabhan, R., 1999. A novel in vitro replication system for Dengue virus. Initiation of RNA synthesis at the 3’-end of exogenous viral RNA templates requires 5’- and 3’-terminal complementary sequence motifs of the viral RNA. J. Biol. Chem. 274, 33714-33722.

You, S., Falgout, B., Markoff L., Padmanabhan, R., 2001. In vitro RNA synthesis from exogenous Dengue viral RNA templates requires long-range interactions between 5’- and 3’- terminal regions that influence RNA structure. J. Biol. Chem. 276, 15581-15591.

You, X.J., Kim, J.W., Stuart, G.W., Bozarth, R.F., 1995. The nucleotide sequence of cowpea mottle virus and its assignment to the genus Carmovirus. J. Gen. Virol. 76, 2841-2845.

Yu, H., Grassmann, C.W., Behrens, S.-E., 1999. Sequence and structural elements at the 3’ terminus of bovine viral diarrhea virus genomic RNA: Functional role during RNA replication. J. Virol. 73, 3638-3648.

Yu, H., Isken, O., Grassmann, C.W., Behrens, S.-E., 2000. A stem-loop motif formed by the immediate 5’ terminus of the bovine viral diarrhea virus genome modulates translation as well as replication of the viral RNA. J. Virol. 74, 5825-5835.

Zacharias, M., Hagerman, P.J., 1996. The influence of symmetric internal loops on the flexibility of RNA. J. Mol. Biol. 257, 276-289.

Zamore, P.D., Haley, B., 2005. Ribo-gnome: the big world of small RNAs. Science 309, 1519-1524.

Zell, R., Sidigi, K., Henke, A., Schmidt-Brauns, J., Hoey, E., Martin, S., Stelzner, A., 1999. Functional features of the bovine enterovirus 5'-non-translated region. J. Gen. Virol. 80, 2299-2309.

Zeng, L., Falgout, B., Markoff, L., 1998. Identification of specific nucleotide sequences within the conserved 3'-SL in the Dengue type 2 virus genome required for replication. J. Virol. 72, 7510-7522.

Zhang, C., Simon, A.E., 1994. Effect of template size on accumulation of defective interfering RNAs in protoplasts. J. Virol. 68, 8466-8469.

Zhang, F., Simon, A.E., 2003a. Enhanced viral pathogenesis associated with a virulent mutant virus or a virulent satellite RNA correlates with reduced virion accumulation and abundance of free coat protein. Virology 312, 8-13.

Zhang, G., Slowinski, V., White, K.A. 1999. Subgenomic mRNA regulation by a distal RNA element in a (+)-strand RNA virus. RNA 5, 550-561.

Zhang, G., Simon, A.E., 2003b. A multifunctional Turnip crinkle virus replication enhancer revealed by in vivo functional SELEX. J. Mol. Biol. 326, 35-48.

Zhang, G., Zhang, J., Simon, A.E., 2004. Repression and derepression of minus-strand synthesis in a plus-strand RNA virus replicon. J. Virol. 78, 7619-7633.

Zhang, G., Zhang, J., George, A.T., Baumstark, T., Simon, A.E., 2006. Conformational changes involved in initiation of minus-strand synthesis of a virus-associated RNA. RNA 12, 147-162.

Zhang, J., Zhang, G., McCormack, J.C., Simon, A.E., Evolution of virus-derived sequences for high level replication of a subviral RNA. In press.

Zhang, J., Zhang, G., Guo, R., Shapiro, B.A., Simon, A.E., A pseudoknot in a pre-active form of a viral RNA is part of a structural switch activating minus-strand synthesis. Manuscript in revision.

Zhong, W., Rueckert, R.R., 1993. Flock house virus: down-regulation of subgenomic RNA3 synthesis does not involve coat protein and is targeted to synthesis of its positive strand. J. Virol. 67, 2716–2722.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406-3415.

Zuniga, S., Sola, I., Alonso, S., Enjuanes, L., 2004. Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. J. Virol. 78, 980-994.