Crystal Structure and Functional Studies of a Viral RNA Pseudoknot Involved in Ribosomal Frameshifting

by
Li Su

B. S., Chemistry
Beijing University of China, 1992

SUBMITTED TO THE DEPARTMENT OF CHEMISTRY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
February 1999

Signature of Author_____
Department of Chemistry, September 18, 1998

Certified by_____
Alexander Rich, William Thompson Sedgwick Professor of Biophysics
Thesis Supervisor

Certified by_____
Lawrence J. Stern, Professor of Chemistry, Co-supervisor

Accepted by_____
Dietmar Seyferth, Professor of Chemistry
Chairman, Departmental Committee on Graduate Students

Science

Certified by_____

Lawrence J. Stern, Professor of Chemistry

Chairman, Thesis Committee

Certified by_____

Alexander Rich, William Thompson Sedgwick Professor of Biophysics

Supervisor, Thesis Committee

Certified by_____

Uttam L. RajBhandary, Lester Wolfe Professor of Molecular Biology

Member, Thesis Committee

Crystal Structure and Functional Studies of a Viral RNA Pseudoknot Involved in Ribosomal Frameshifting

by
Li Su

Submitted to the Department of Chemistry on September 18, 1998 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Abstract

This thesis presents the first crystal structure of a viral RNA pseudoknot involved in ribosomal frameshifting. Many viruses undergo a programmed –1 ribosomal frameshift event to regulate the translation levels of structural and catalytic proteins important for viral assembly. This process is often induced by a slippery sequence and pseudoknot downstream in the polycistronic mRNA. A pseudoknot originates from a hairpin loop, in which a single-stranded region folds back and base-pairs with the loop region of the hairpin, forming a quasi-continuous helix with two stems and two loops.

The 1.6 Å resolution crystal structure of the beet western yellow virus pseudoknot reveals rotation and a bend at the junction of the two stems. A loop base is inserted in the major groove of one stem with quadruple base interactions. The second loop forms a new minor groove triplex motif with the other stem, involving 2′-OH and triple base interactions, as well as sodium ion coordination. Overall, the number of hydrogen bonds stabilizing the tertiary interactions exceed those involved in Watson-Crick basepairs.

Mutational studies based on this new information show that the adenosine-rich extended minor groove triplex motif is an important feature in frameshifting. The quadruple base interaction at the junction of the two stems is a key element that maintains frameshifting. Furthermore, mutation of the stem-loop turn region near the 5′-end of the molecule can increase frameshifting efficiencies by as much as three fold. The studies confirm the tertiary structural organization of the BWYV pseudoknot and further our understanding of the mechanism of ribosomal frameshifting.

Thesis supervisor: Alexander Rich
Title: William Thompson Sedgwick Professor of Biophysics

Acknowledgement

The work presented here in this thesis relied on help and support from many people. I would like to thank my supervisor, Dr. Alexander Rich, for his support, advice, patience and generosity. I am grateful to Alex for assigning me this challenging and exciting project, which inspired me daily in graduate school. I clearly remember this day six years ago, when he described to me the topology of this intriguing yet unknown molecule, and now the elucidation of this structure has brought us both joy and excitement.

I wish to thank my co-supervisor Dr. Lawrence J. Stern for his guidance and advice in crystallogaphy, his technical input on a good organization of the X-ray facility and cryocooling system. I also thank my former co-supervisor Dr. Jamie Williamson. His nearby laboratory provided a constant resource in the field of RNA, and I enjoyed attending his lively group meetings. I appreciate my thesis committee member, Dr. Uttam RajBhandary for his advice and encouragement on my project and manuscript preparation.

I owe special thanks to my three collaborators. Dr. Liqing Chen taught me the basic techniques and principles in crystallography, and accompanied me on several trips to the synchrotron; Dr. Martin Egli, as well as his postdoctoral fellows George Minasov, Valya Tereshko provided extensive help on the refinement process and in depth knowledge on nucleic acid crystallography and chemistry; Dr. James Berger was a constant source of help in heavy atom phasing and introduced me to various programs. I am grateful to Dr. Eric de la Fortelle, who genuinely instructed me on the program SHARP at the MAD Data Collection and Analysis course in Grenoble. Dr. Imre Berger deserves my special thanks for his constant personal encouragement and scientific inspiration as well

as teaching me the molecular replacement method during our attempt to solve the structure by this method.

I would also like to thank Dr. Craig Ogatta from the NSLS X4A beamline, for his technical assistance in MAD data collection, Mike Soltis from the SSRL for his enthusiasm in trying the Xenon diffusion heavy atom phasing technique, and the staff at CHESS F2 beamline. I acknowledge help from Thomas Schwartz and Dr. David Chan, Markus Schade and undergraduate student Kuan-Jie Wu for data collection at the synchrotron. Dr. Carl Correll and Dr. Amy Anderson were helpful in providing tips on MAD data collection.

Dr. Yang-Gyun Kim had performed the mutational experiments mentioned in this work. His expertise and proficiency in molecular biology allowed us to look into the biological implications of this structure in a timely manner. I also appreciate helpful discussions from Dr. David Bartel, Dr. Jamie Cate and Dr. Mark Rould on the structure.

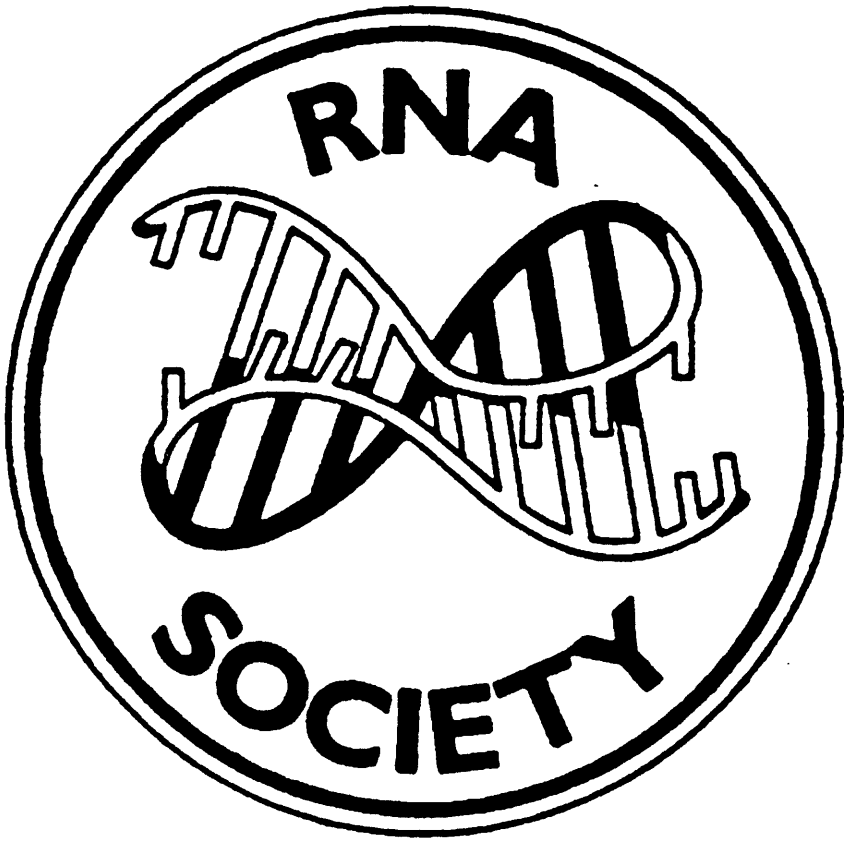
I am indebted to Dr. Henry Taube and Dr. Steve Holbrook for generously providing us with the osmium hexammine triflate compound. Dr. Lee Gerhke and Dr. Felicia Houser-Scott had introduced me to the basic techniques in molecular biology and assisted me in developing the T7 transcription system. Dr. James McCloskey and Dr. Pamela Crain had assisted me in the Bromine Mass spectroscopy analysis.

Within the Rich lab, I have benefited from the help and expertise of many members. Shuguang Zhang constantly looked over my progress and genuinely assisted in setting up my outside collaborations. Ky Lowenhaupt had provided technical assistance in the T7 RNA polymerase purification process. Curtis Lockshin had always been helpful in setting up the computer system and maintaining the X-ray facility. Yanyan Xing, Alan Herbert and Bernard Brown

were a constant source for consultation. I also acknowledge Dr. Douglas Lauffenburger from the Center of Biomedical Engineering for generously providing us with the computer facility.

Finally, I would like to thank my father Jilan Su, a scientist that continuously encouraged and inspired me, my husband Sean Hui for his love, understanding and encouragement.

To the RNA Society




RNA
SOCIETY

To my parents and my husband

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Note: Legends of figures and tables are at the end of each chapter.

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 1

The RNA Pseudoknot Motif

Introduction

RNA Structure

The many biological functions of the RNA molecule such as catalytic ability, ribosomal function, and translational regulation are achieved by its adoption to various three dimensional conformations. Similar to proteins, the RNA molecule folds into these complex structures characterized by secondary and tertiary structural elements. Using the principles of Watson-Crick base-pairing, the RNA forms classical secondary structures such as hairpins, bulge loops, internal loops and multi-branched junctions. Additionally, non-canonical basepairs such as mismatches, water-mediated, and protonated basepairs can form to retain helical regions. These types of interactions have been observed in quite a few NMR studies, such as the UUCG hairpin loops (Heus and Pardi, 1991); RNA aptamers that bind AMP (Jiang et al., 1996), arginine, and citrulline (Yang et al., 1996); the nucleotide bulges formed in the REV-RRE complex (Battiste et al., 1996) and TAR RNA–Arginine complex (Puglisi et al., 1992). In addition, the crystal structure of the 5S rRNA domain has revealed cross-strand purine stacks in a helical region of G•A mismatches (Correll et al., 1997).

The tertiary interactions of the secondary structure fold the RNA into its ultimate three dimensional structure to accomplish important tasks. The transfer RNA series represents a single tertiary structural motif, displaying loop-loop, loop-single strand and stem-nucleotide interactions. The interactions are stabilized by non-canonical base-pairing and base intercalation (Quigley and

Rich, 1976). A few new tertiary motifs have been observed in recent RNA crystal structures. The Hammerhead and group I intron structures both reveal a tetraloop motif which can be involved in RNA docking and domain organization (Pley et al., 1994b; Cate et al., 1996). In this motif, the tetraloop is analogous to the U-turn observed in tRNA (Quigley and Rich, 1976), and interacts with the minor groove of the tetraloop receptor stem region. Other motifs found in the group I intron structure include ribose zippers, which involve 2'-OH to 2'-OH contacts (Cate et al., 1996). These recent crystal structures provide the initial models for tertiary interactions and RNA packing.

The pseudoknot tertiary motif represents an independent entity that can be predicted based on secondary sequence. This conformation originates from a hairpin loop, in which an adjacent single-stranded region folds back and base-pairs with the loop region of the hairpin, forming a quasi-continuous helix with two stems and two loops (Figure 1).

RNA Pseudoknots

Topology

The RNA pseudoknot motif was first proposed upon the finding of aminoacylation properties at the 3'-end of the turnip yellow mosaic virus (TYMV) RNA (Rietveld et al., 1982). Pleij et al. had proposed that utilizing the pseudoknot folding principle the genomic RNA was able to form a quasi-continuous helix that resembles the acceptor arm of the tRNA molecule (Pleij et

al., 1985). As in Figure 1, the stem at the 5′-end is named Stem 1, and the 3′ stem is termed Stem 2; Loop 1 and Loop 2 are unpaired regions. Due to the right-handedness of the helix, the two loops are inequivalent. Loop 1 crosses the major groove of Stem 2 and Loop 2 crosses the minor groove of Stem 1. The topology is unique as in this single configuration, the directionality of the chain is reversed twice. Application of force on the 5′- and 3′- end will result in the unraveling of this configuration without forming a real knot.

Stability

Cations

Similar to the transfer RNA , the folding of the pseudoknot is stabilized by divalent ions, Mg^{2+} in particular. High concentrations of monovalent ions like Na^{+} mimic the effect of low concentrations of Mg^{2+} (Wyatt et al., 1990). The close electrostatic contacts between the phosphate groups at the loop-stem junctions may account for the requirement of ions in stabilization (Puglisi et al., 1990). In the moloney murine leukemia virus pseudoknot, divalent ions are not required to fold the pseudoknot but do stabilize it further (Gluick et al., 1997). In this stop codon read-through pseudoknot, Gluick et al. also showed that the structure was preferentially stabilized by smaller alkali metal ions. It is possible that at regions of closely spaced negative charges, the dehydrated ions with a smaller ionic radius can approach more closely and recover a larger electrostatic energy (Eisenman and Horn, 1983). NMR and biochemical experiments show that the pseudoknot is in equilibrium with the corresponding 5′ and 3′ hairpins

(Wyatt et al., 1990) (Figure 1). Minor changes in salt and sequence can radically alter the equilibrium among the possible conformations (Wyatt et al., 1990).

Thermal stability

The enthalpy of stacking is the main driving force for helix formation (Saenger, 1984). Thermodynamic studies on a synthetic pseudoknot by Wyatt et al. show that the pseudoknot has only a marginal free energy gain (1.5-2 $kcalmol^{-1}$) over the unfolded hairpins (Wyatt et al., 1990). The measured enthalpy of formation is smaller in absolute value then the expected enthalpy calculated from comparable number of basepairs using nearest neighbor interactions (Turner et al., 1988). The smaller than predicted enthalpy may be due to distortions in the stacking region between the two stems as well as positive enthalpic contributions of the loop regions (Puglisi et al., 1990). At optimal salt concentrations, the unfolding transitions of the pseudoknot can be two-state (Wyatt et al., 1990). But, if the pseudoknots unfold in two closely spaced transitions, the two state analysis may underestimate the total unfolding enthalpy (Gluick et al., 1997).

Loop sequence and length

Various pseudoknots that form in mRNA and rRNA can have very large and structured loop regions, as the two stems may not be adjacent. However, the two inequivalent loops have minimal length requirements based on A-form geometry calculations (Pleij et al., 1985). It has been predicted that at least three nucleotides in Loop 2 are needed for crossing the wide minor groove, and two

nucleotides in Loop 1 are required for crossing the narrow major groove (Wyatt et al., 1990). Although, biochemical and biophysical studies on several frameshifting pseudoknots that contain only one base in Loop 1 have been shown to form stable pseudoknots (Du et al., 1996 and 1997). In these cases, a 6-7 base-paired Stem 2 allows a short Loop 1 to cross its major groove, as the distance across the groove reaches its minimum due to the helical turn (Pleij et al., 1985). It is conceivable that changes in ribose puckering, torsion angle (C4′-C5′) and perturbation in the helix may also facilitate the pseudoknot conformation to form (Pleij et al., 1985).

Biological Role

The pseudoknot motif has been found in messenger, ribosomal and catalytic RNA. The pseudoknots found in the previously mentioned viral RNAs (TYMV) suggest an alternative construction of similar RNA structure and functions. Whereas, other pseudoknots tend to facilitate formation of compact tertiary structures by forcing RNA strands into close proximity. In addition, the pseudoknot can be viewed as a unique structure recognized by proteins (reviewed in ten Dam et al., 1992). The structure provides an unusual arrangement of single strands, which is distinguished from hairpin structures.

The diverse functions of the pseudoknot can be categorized in four classes. The first class is found in the 5′ and 3′ untranslated regions in non-polyadenylated plant viral RNAs, such as TYMV (Rietveld et al., 1982) and the

brome mosaic virus (Felden et al., 1994) to control replication. Using the pseudoknot folding motif, the messenger RNA is folded into a tRNA-like structure. The second class of pseudoknots are found in the core region of catalytic RNAs for structure stabilization. For example, in the substrate binding core of the group I intron (Michel et al., 1990), and in the hepatitis delta virus ribozyme (Perrotta et al., 1993). Phylogenetically conserved pseudoknot foldings present in certain subunit ribosomal RNAs comprise the third class of pseudoknots. Three different pseudoknots cluster at the junction of three distinct domains. Pseudoknot I assists in the formation of the 70S ribosomal complex (Dammel and Noller, 1993), a second pseudoknot involves a 530 hairpin loop (Powers and Noller, 1991), and another pseudoknot in domain II of the 23S rRNA is essential for ribosome function (Gunnar et al., 1995). It is postulated that those pseudoknots serve as conformational switches in the functioning ribosome (Dammel and Noller 1993). The most important and exciting fourth class of pseudoknots are involved in translational control. Interestingly, the ribosomal binding sites of several translationally regulated mRNAs are positioned within a pseudoknot structure. The gene 32 protein of T4 phage is postulated to bind to a high-affinity pseudoknot site and thus trigger cooperative binding for repression of its own translation (McPheeters et al., 1988). The ribosomal protein S15 also stabilizes the pseudoknot-containing conformation of its coding mRNA, and blocks the ribosome in a preinitiation complex (Bénard et al., 1996). Ribosomal frameshifting and translational readthrough in viruses (reviewed by Weiss 1991)

are the clearest examples of the functional role of pseudoknot structures, as this shall be our main focus in the following sections.

Ribosomal Frameshifting

In many viruses, including tumor- and retro- viruses, the programmed –1 ribosomal frameshifting of polycistronic mRNA regulates the relative level of structural and enzymatic proteins important for efficient viral assembly (Gesteland and Atkins, 1996; Farabaugh, 1996). The –1 shift in reading frames causes stop codon readthrough, and results in production of a single fusion protein. For example, in the Rous sarcoma retrovirus, the *pol* gene that encodes integrase, protease and reverse transcriptase is expressed with the upstream *gag* gene (encoding virus core proteins) through a *gag-pol* fusion protein. The mature products are later obtained by processing the poly-protein precursor (Jacks et al., 1988). The –1 frameshifting is not only found in retroviruses (Jacks et al., 1988; Chamorro et al., 1992; ten Dam et al., 1994) but also found in coronaviruses (Brierley et al., 1991), yeast (Tzeng et al., 1992) and plant viruses (Miller et al., 1995), as well as bacterial systems (Gesteland and Atkins, 1996). Frameshifting levels can range from 1 to over 30% in different systems to produce gene products in a functionally appropriate ratio. However, the mechanism of ribosomal frameshifting is not understood. It is postulated that a complex mRNA structure 6-8 nucleotides downstream from the "slippery sequence" (Figure 2A), in many cases a pseudoknot, leads to ribosomal pausing and the simultaneous

slippage of both aminoacyl and peptidyl tRNAs toward the 5′ direction by one base (Jacks et al., 1988) (Figure 2B). The slippery shift site on the messenger RNA has an X XXY YYN consensus sequence (the initial reading frame is indicated, and bases X and Y can be identical). This homopolymeric sequence can maintain at least two out of three basepairs with the anticodons upon slippage (Jacks et al., 1988).

In systems that use a pseudoknot for frameshifting, substituting a hairpin with equivalent base-paired stems also induces ribosomal pausing, but does not promote frameshifting (Somogyi et al., 1993). Many mutational studies have been performed on frameshifting pseudoknots to correlate structure and function (Brierley et al., 1991; Chamorro et al., 1992; Tzeng et al., 1992; ten Dam et al., 1994), and certain mutations at the junction of the two stems or loop regions can have deleterious effects on frameshifting efficiencies (Chen et al., 1995; ten Dam et al., 1995). These findings suggest that there are special structural features involving tertiary interactions in a frameshifting pseudoknot that are important for function.

Structural Studies on Pseudoknots

There have been several NMR studies of pseudoknots reported during the past few years (Puglisi et al., 1990; Shen et al., 1995; Du et al., 1997; Kang et al., 1997; Kolk et al., 1998). The first NMR study of a synthetic pseudoknot (Puglisi et al., 1990) reported that the two stems stacked to form a continuous helix with

minor distortions; nucleotide conformations in the stem were C3′-endo sugar pucker (N-type) and in the *anti* glycosidic torsion angle, those at the stem-loop junction were affected by terminal fraying and in equilibrium with the C2′-endo (S-type) conformation; loop nucleotides were poorly defined, and some S-type conformations were suggested; Loop 1 residues were suggested to be stacked in the major groove.

A recent NMR study on the pseudoknotted T arm and acceptor arm of the tRNA-like structure of TYMV has shown an interplay between the two S-type Loop 2 nucleotides and Stem 1 (Kolk et al., 1998) (Figure 1). A conserved adenine base in Loop 2 is tilted to an angle of almost 90° with respect to the plane of the opposing basepairs in Stem 1, allowing it to hydrogen bond to two guanines in adjacent layers; the cytosine residue which stacks on the adenosine in Loop 2 can also be within hydrogen bonding distance to a 2′-OH in Stem 1; Loop 1 nucleotides are stacked in the major groove and some adopt S-type sugar conformation; co-linearity is maintained at the helical junction and slight torsional and swaying motions between these helices are most likely responsible for the observed line broadening for residues near the interface of Stems 1 and 2.

A 4.8 Å resolution crystal structure of a HIV-reverse transcriptase complexed with an RNA pseudoknot inhibitor revealed some general features in the unrefined model, only some parts of the pseudoknot were observed (Jaeger et al. 1998). The pseudoknot is kinked by 59° between the two stems, thereby optimizing its contacts with subunits of the heterodimer; the minor groove of

Stem 2 and Loop 1 stabilize the 'closed' conformation of the polymerase through extensive electrostatic interactions with several basic residues; Loop 2 interacts with basic residues in a cleft and Stem 1 interacts with the connection domain of both subunits.

NMR studies on pseudoknots active in ribosomal frameshifting were usually not of wild-type sequence, but had C•G basepair inversions introduced in Stem 2 to assist structural analysis (Shen et al., 1995; Du et al., 1997). The solution studies of the mouse mammary tumor virus (MMTV) pseudoknot revealed that the two base-paired stems are in a bent conformation (Shen et al., 1995; Kang et al., 1997;). A single adenylate residue is intercalated between the two stems so that direct coaxial stacking of the two stems is not possible. In addition, loop nucleotides may be restrained by base-stacking interactions and are in equilibrium with the C2'-endo conformation. The NMR studies of the simian retrovirus –1 (SRV-1) pseudoknot confirmed base-pairings within the stems (Du et al., 1997), but could not provide unambiguous evidence for or against coaxial stacking of the two stems.

In general, there has been inadequate structural information in pseudoknots for understanding this motif, and this has hindered further studies regarding their functions, in particular in frameshifting. In the previously described NMR studies, the junctions of the two stems are not well defined and very little is known about loop-stem interactions. A high resolution crystal

structure will expand our knowledge in understanding this intriguing motif, and give us one part of the picture in ribosomal frameshifting.

Figure Legends

Figure 1. Pseudoknot diagram.

Transformation from hairpin loop structure to pseudoknot. The classical H-type pseudoknot forms a quasi-continuous helix of two stems with two connecting loops. Loop 1 crosses the major groove of Stem 2 and Loop 2 crosses the minor groove of Stem 1 (Pleij et al., 1985).

Figure 2.. Mechanism of ribosomal frameshifting.

(A) The homopolymeric hepta-nucleotide slippery sequence and a downstream pseudoknot are signals for frameshifting. The spacer region has to be precisely positioned.

(B) Simultaneous-slippage model (Jacks et al., 1988). $tRNA^{Asn}$ carrying the nascent peptide (jagged line) and $tRNA^{Leu}$ are shown bound to the codons AAU and UUA in the ribosomal peptidyl (P) and amino-acyl (A) sites. When frameshifting occurs, both tRNAs slip back by one nucleotide in the 5′ direction, so that they pair with AAA and UUU, and form basepairs only with the first two nucleotides of each codon. Following peptidyl transfer and translocation, the next tRNA is brought into the A-site.

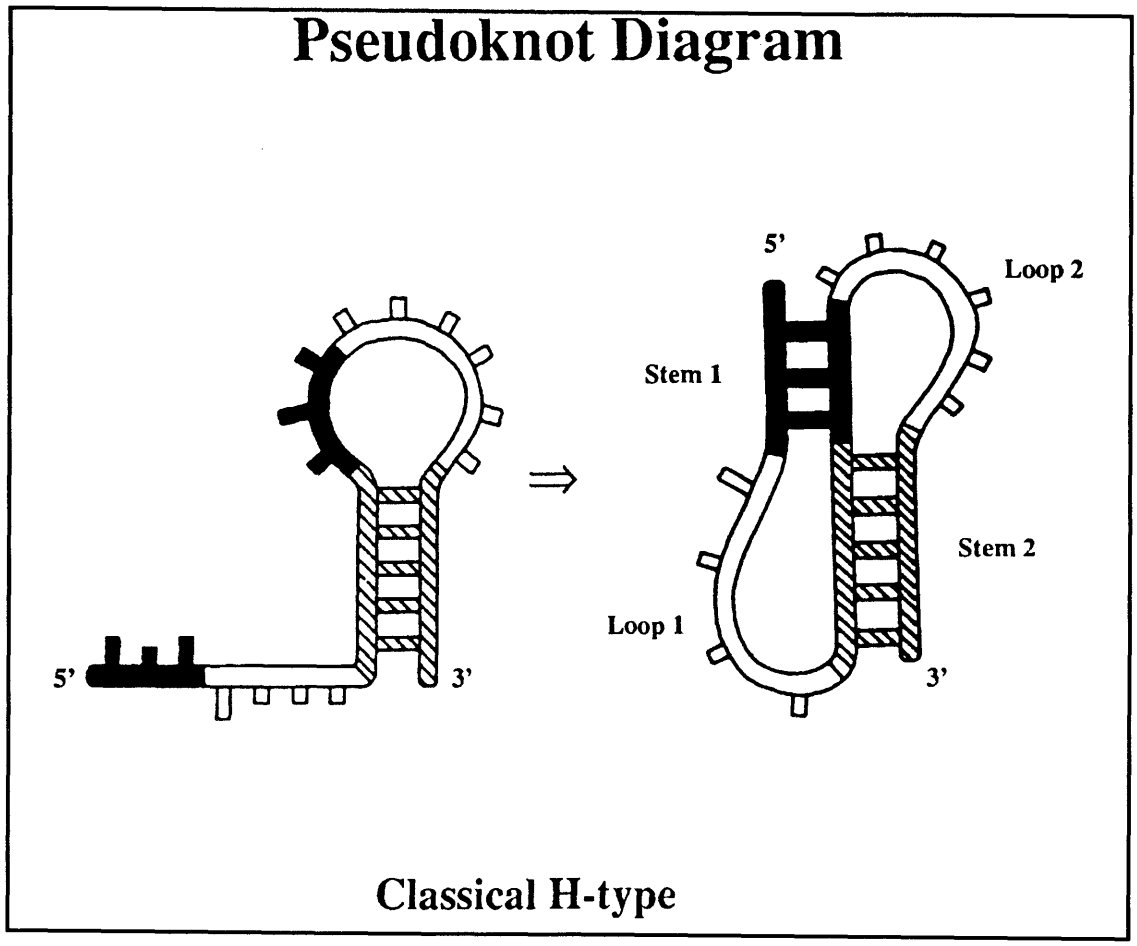

Pseudoknot Diagram
5'
3'
⇒
5'
Loop 2
Stem 1
Stem 2
Loop 1
3'
Classical H-type

Figure 1

Mechanism of Frameshifting

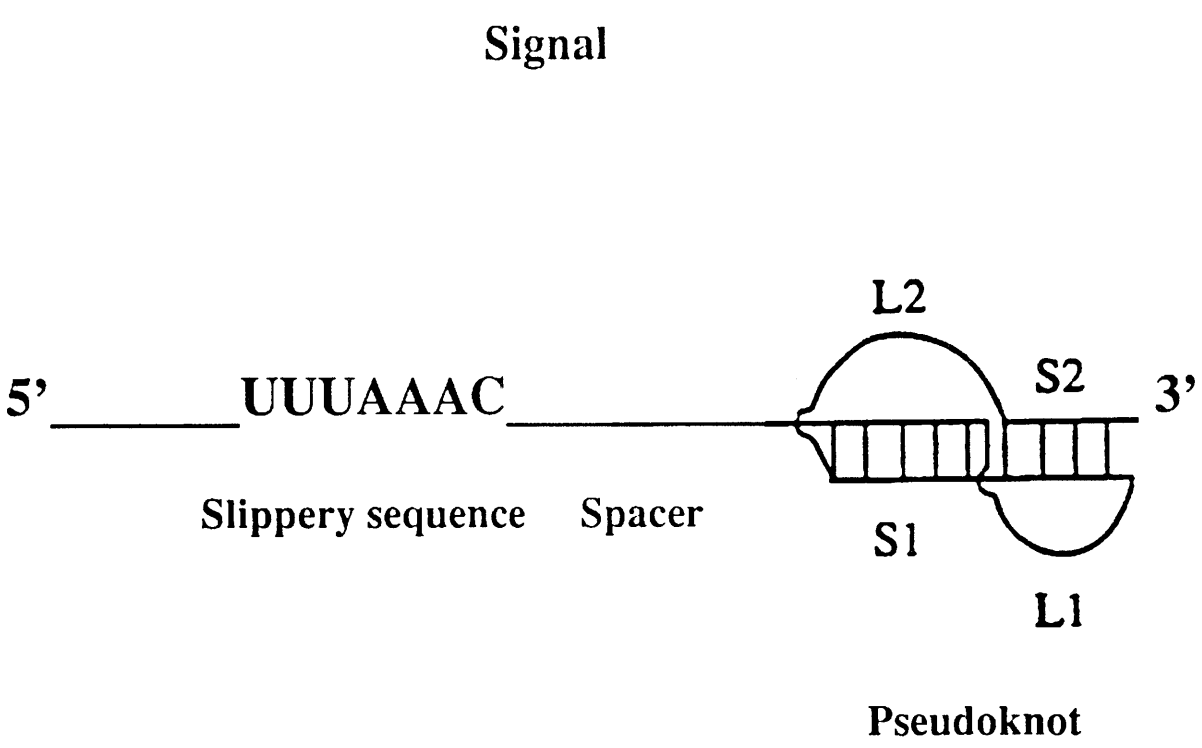
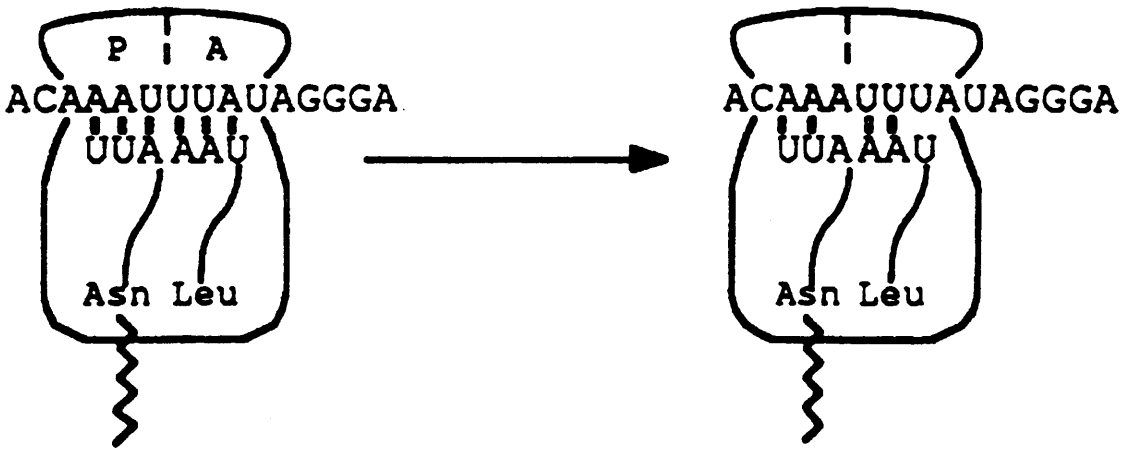


Figure 2A

Mechanism of Frameshifting

-1 Simultaneous Slip



Jacks et. al. (1988) Cell 55

Figure 2B

Chapter 2

Transcription, Crystallization and Structure Determination of the Pseudoknot from Beet Western YellowVirus

Methods

In vitro T7 RNA Polymerase Transcription

Milligram amounts of homogeneous RNA are required in the crystallization process. The *in vitro* T7 RNA polymerase transcription system is a convenient method developed by Milligan for synthesizing large quantities of RNA that range from 9 to over 100 nucleotides (Milligan et al., 1989). In comparison to chemical synthesis, the heterogeneity of the product is solely in the length, and the RNA is free of deprotection side products. In addition, for large RNA fragments, producing 20-30 mg of the full length product is feasible at a lower cost.

T7 RNA polymerase purification

To synthesize large amounts of RNA, large quantities of T7 RNA polymerase are consumed in this process. Commercial sources of the enzyme are very expensive, therefore a rapid method of obtaining the enzyme was carried out with a plasmid carrying the T7 RNA polymerase gene. We had been provided with the *E. coli* strain BL21/pAR1219 by Dr. James Williamson for overexpression of the enzyme. This plasmid contains the ampicillin resistant gene, therefore, cells were grown in M9TB media with 40 ug/ml of ampicillin for cell selection. Addition of the isopropyl-ß-thiogalactoside inducer at mid-log phase cell growth relieved suppression of the lacI supressor and allowed overexpression of the T7 RNA polymerase . Cells were harvested after additional hours of incubation by centrifugation for 10 minutes at speed 6000 rpm, and washed twice with 100 ml of 20 mM Tris HCl pH 8, 20 mM NaCl, 2 mM

trisodium EDTA, and stored as a frozen pellet at –70 °C. SDS protein gels were ran on samples before and after induction to check if overexpression was successful.

The frozen cells were resuspended in 35 ml of lysis buffer containing 50 mM Tris pH 8, 20 mM NaCl, 2 mM EDTA, 1 mM DTT. Phenylmethylsulfonyl (PMSF) and leupeptin were added to prevent proteolysis. Lysis was initiated with egg white lysozyme, and completed by adding sodium deoxycholate. The viscosity of the lysate was reduced by sonication for four times at 15 second bursts. 100 ug/ml Rnase free Dnase and 20 mM MgCl2 were required for DNA digestion. 5 ml of 2 M ammonium sulfate was added to disrupt protein-DNA binding.

Polymin P is a positively charged polymer, and is used to precipitate nucleic acids. 2.5 ml of 10% Polymin P was slowly added to the crude lysate with stirring on ice, and the procedure was repeated once. The nucleic acid precipitate was then removed by centrifugation for 15 minutes at 16,000 rpm. 0.82 volume of saturated ammonium sulfate was slowly added to the supernatant on ice, and incubated for 20 minutes with gentle stirring at 4 °C. The protein was precipitated by centrifugation at 9500 rpm for 10 minutes, and dissolved in 15 ml of buffer C containing 20 mM sodium phosphate, pH 7.7, 100 mM NaCl, 1 mM DTT, 1 mM EDTA and 5% glycerol and 0.25 mM PMSF. The dissolved protein fraction was dialyzed against buffer C using 3000 molecular weight cut

off dialysis tubing. The dialyzed solution was then clarified from aggregation products through centrifugation at 10,000 rpm for 10 minutes.

Three ion exchange columns were used to obtain the final fraction of T7 RNA polymerase. SP Trisacryl Plus-M is a strong acidic cation exchanger, the matrix is poly(N-tris[hydroxymethyl]methylmethacrylamide. The protein solution was diluted with an equal volume of buffer C without NaCl, and passed through the Tris acryl SP column at a flow rate of 2 ml/min. The column was then washed with 4 volumes of starting buffer to remove proteins that fail to bind the column. The bound T7 RNA polymerase is eluted with buffer C containing 200 mM NaCl. The T7 fractions were dialyzed 4 times against 500 ml of 25 mM NaCl buffer for 1.5 hours. The TSK CM-650M is a weak cation exchange column, and the matrix is carboxymethyl. The dialyzed fraction was applied to the TSK CM column at a flow rate of 1 ml/min, and the T7 RNA polymerase was recovered through the flow through. The last column used is a TSK DEAE 650M weak anion exchange column, the matrix is diethylaminoethyl. The flow through from the previous column was applied to the TSK DEAE column in buffer C with 25 mM NaCl. Finally, the T7 RNA polymerase was eluted with a 200 ml NaCl gradient from 25 to 250 mM at 1 ml/min. Peak fractions of polymerase eluted near 100 mM NaCl, and were concentrated by dialysis against 100 mM NaCl buffer C containing 50% glycerol. The protein was checked by SDS-page, assayed for transcription activity, and then stored at

–20 °C at 10 mg/ml concentration. The purification procedure produced approximately 50 mg of pure , highly active T7 RNA polymerase per 1 Liter culture.

***In vitro* transcription of RNA oligonucleotides**

The transcription is performed on an annealed template, in which the non-coding strand contains the promoter region and the bottom strand contains the complementary promoter and template region. It is known that the addition of G residues at the 5'-end of the RNA transcript can greatly increase yields (Milligan et al., 1989). Only two G residues were added at the 5'-end of the 28 nucleotide BWYV transcript (Figure 1) (Miller et al., 1995) in consideration of possible aggregation in G-rich sequences during crystallization. The DNA oligonucleotides were synthesized by the Biopolymers Lab at MIT, and purified to homogeneity using polyacrylamide gel electrophoresis. Initial optimization trials were carried out to find a condition that produces maximum yield of the full length product. The concentrations of the T7 RNA polymerase, NTPs and $MgCl_2$ were varied. Adding large amounts of DTT that prevented oxidation of the enzyme and providing access GTP in the reaction increased the yields significantly. The large scale transcriptions were carried out with a reaction mixture consisting of 80 mM Hepes pH 8 , 1 mM spermidine, 0.01% Triton X-100, 5 mM DTT, 80 mg/ml polyethylene glycol, 2.5 mM each of ATP/CTP/UTP, 23.4 mM $MgCl_2$, 300 nM DNA template and 0.015 mg/ml T7 RNA polymerase,

180mM DTT, and pyrophosphatase. Transcription reactions were incubated at 37 °C for 3.5 hours.

EDTA was added to a final concentration of 50 mM to quench the reaction and to dissolve the magnesium pyrophosphate precipitate produced during the reaction. The reaction mixture was extracted with equal volume of phenol/chloroform, and centrifuged at 10 minutes at 4000 rpm. The RNA aqueous phase was collected, and the bottom phenol phase was back extracted with 100 mM Tris pH 6.5. The combined aqueous phase was then extracted with 24:1 chloroform/isoamyl alcohol to eliminate the remaining phenol. The RNA was then precipitated in 2.5 volumes of ethanol with 0.1 volume of 5 M sodium acetate at pH 5.2. The solution was left at –20 °C for 4 hours prior to centrifugation at 10,000 rpm for half an hour. The RNA pellet was then washed twice with 70% ethanol to remove salts, and lyophilized.

RNA sequencing

The transcription produces abortive , n+1 or n-1 products and in some cases several major bands, as shown in Figure 2B. The identity of the full length product was resolved by RNA sequencing with nucleotide specific ribonucleases T1/U2/PhyM/B on 5′ phosphorylated RNA under denaturing conditions. Ribonuclease T1 cleaves after G, U2 cleaves after A, PhyM cleaves after U and A, and B cleaves after C and U. An alkaline hydrolysis ladder of the RNA was generated to locate each nucleotide in the sequencing gel. A control lane of the full length product was also run on the 20% polyacrylamide gel, and minor

degradation products were detected that had cleavage between C22 and A23 in Loop 2 (Figure 2A). The results showed that the major products in the reaction were the n mer and n+1 mer, the latter which runs slower on the gel.

Purification of the full length RNA product

20% denaturing polyacrylamide gels were run on the ethanol precipitated RNA mixture. 1 mg of RNA (final product) was loaded on a 3 mm thick sequencing sized gel. As the gel lanes were run to the bottom, the n mer and n+1 mer gel bands were clearly separated. The RNA gel bands were visualized by UV shadowing over a TLC plate. For the light sensitive brominated RNA, the gel was stained with 0.02% of methylene blue for band detection. The band for the full length product was cut out and gel slices were eluted overnight via crush and soak method, and applied to Amicon Centriplus (10kd cutoff) units for desalting and concentrating. Finally, the RNA was microdialyzed over double distilled water for 12 hours, lyophilized and stored at –20 °C . A yield of 3mg per 50ml reaction was obtained for the BWYV pseudoknot.

Crystallization

The conformation of the RNA pseudoknot is stabilized from the unfolded 5′ or 3′ hairpin species by magnesium ions and high concentrations of monovalent ions (Wyatt et al., 1990). Annealing and slow cooling the RNA under appropriate salt concentrations will increase the likelihood of maintaining one homogenous pseudoknot conformation upon crystallization (Doudna et al.,

1993). Therefore, the RNA was neutralized with 300 mM sodium cacodylate (pH 6.5), and incubated at 60 °C for 10 minutes in the presence of 15 mM $MgCl_2$, followed by slow cooling to 25 °C on the heat block and transferred to a 4 °C cold room. The RNA solution was immediately used for making set up trials, since freeze and thawing the RNA would cause gradual degradation. Native I crystals grew to the size of 0.6x0.15x0.15 mm^3 within one week at 25 °C from a sitting drop set up (Figure 3A). The sitting drop contained 4 ul of 4 mg/ml RNA, and 6ul of stock solution containing 5 mM $MgCl_2$, 5% sec-butanol, 2 mM spermidine, 100 mM potassium MOPS buffer (pH 7.0). The sitting drop was equilibrated over a reservoir of 18% sec-butanol. The rod-shaped crystals in Figure 2A diffracted to 1.6 Å, and crystallized in the spacegroup $P3_221$ (a=b=30.08 Å, c=140.08 Å). Because of the volatile nature of the precipitant sec-butanol, the crystals were stored at 4 °C after maximum crystal growth. Another crystal form of the cubic space group was obtained in a high salt condition with 6ul of 4mgl/ml of RNA and 3 ul of stock solution containing 3.2 M ammonium sulfate and 100 mM sodium citrate, pH 5.0. The sitting drop was equilibrated over 1.9-2.0M ammonium sulfate at 4 °C for 1-2 weeks. The crystals grew to the size of 0.7x0.7x0.5 mm^3 and were very sensitive to temperature change or handling (Figure 3B). The data collection of these crystals required cryo-cooling conditions at 100K to prevent serious decay and the diffraction limit was 2.85 Å. The cryo-protectant used contained 2.4 M lithium sulfate (ammonium sulfate would crystallize out), 0.1 M sodium citrate and 20%

glycerol. Due to the fragility and the lower diffraction limit of these cubic form crystals, we focused on the trigonal form crystals in the following work.

Heavy atom derivatization

Initial trials of heavy metal soaking were carried out using samarium, platinum, mercury and osmium containing compounds. But none of them were high affinity derivatives that gave interpretable Patterson maps to locate the heavy atom sites. Incorporation of a single bromine site in the RNA was possible through introducing 5-bromo-UTP in the transcription reaction. The brominated RNA samples were subjected to Mass-spectroscopy analysis to check for the presence of bromine. Brominated crystals took 1-2 months to grow from stock solutions containing 5% 2-methyl-2,4-pentanediol (MPD), 5 mM $MgCl_2$, 1 mM spermine, 100 mM potassium PIPES buffer (pH 6.5), and equilibrated over a 25% MPD reservoir at 25 °C. The brominated crystals were slightly non-isomorphous with the Native I crystals grown out of sec-butanol, as indicated from the higher R_{deriv} and the noisy Patterson maps. However, native crystals (Native II) later grew from the MPD conditions and were fully isomorphous with the brominated crystals.

Osmium derivatives were prepared by soaking the Native II crystals in 10 mM Osmium (III) hexammine trifluoromethanesulfonate for 15 hours and backsoaking them in mother liquor overnight. Non-optimal concentrations and soak times would influence cell parameters and produce a high R_{sym} in the crystal data.

Data Collection and Reduction

All data were collected at 100K, and crystals were soaked in mother liquor containing 25% MPD as a cryoprotectant for less than 1 minute before freezing. Data collection on Native II crystals was carried out on a Rigaku R-AXIS IIC Imaging plate detector with a Rigaku RU200 copper rotating anode generator, equipped with double focusing mirrors. Native I data were collected at Brookhaven beamline X4A with a Fuji plate detector. The Native I crystal displayed some anisotropic characteristics in diffraction, as higher resolution data were observed along the long axis c direction. In addition, strong overloading diffraction spots were observed around 3.0-3.4 Å along the long axis, indicating possible base stacking along c in the lattice (Figure 4). This continuous stacking pattern was also observed in the native self Patterson maps along the Z axis. The program *hklplot* (CCP4, 1994) was used to visualize the native data processed in P1 along the reciprocal lattices to find the symmetry axes. A self rotation search with program POLARRFN (CCP4, 1994) on the Native II data processed as P1 was performed to verify a crystallographic two-fold axis along a or b (Figure 5A), and a three-fold axis along c (Figure 5B) in the $P3_221$ space group.

The osmium and bromine derivatives both contain suitable anomalous scatterers for Multiple Anomalous Dispersion (MAD) data collection. The anomalous scattering plots of Br and Os are illustrated in Figure 6A and 6B, respectively. The tunable wavelengths of the synchrotron light source can help optimize the real (f′) and imaginary (f″) components of the heavy atom scattering

factor, so that the $\Delta f'$ between different wavelengths and $\Delta f''$ from the peak wavelength can provide phasing information within one derivative (Hendrickson 1991). Usually four wavelengths are required for a MAD data set: low energy remote reference ($|f'|$ small, f'' small), inflection point (maximum $|f'|$), peak (absorption edge, maximum f''), and high energy remote reference ($|f'|$ small, f'' considerable). When limited beam time is under consideration, the low rather than the high energy remote can be eliminated first, because it does not contribute to any anomalous signal. In extreme circumstances of time consideration, the inflection point and high energy remote can produce reasonable phasing. But, it is always better to have more data points so that the set of simultaneous equations for MAD phasing is over-determined (Ramakrishnan et al., 1997).

In the fluorescence curve, the f'' drastically increases at peak wavelength and plateaus at higher energy. In comparison, the abrupt change in f' is within an energy range of 3-4 eV, this requires high precision in the synchrotron tunable monochromator and in the experimental procedures (Figure 6). Due to the fluctuations in energy between beam fills at the synchrotron light source, fluorescence scans and calibrations are required during each cycle. In addition, because the dispersive and anomalous differences (around 5%) are small compared to the measurement error, care has to be taken in collecting an accurate data set to maximize the signal. Since it was difficult to align a mirror perpendicular to the spindle axis in the $P3_221$ spacegroup, Bijvoet pairs were collected in the inverse

beam mode. All wavelengths of one derivative were collected consecutively at 8° small block intervals close in time.

The 2.0 Å MAD data for the bromine derivative were collected at the Brookhaven X4A beamline. A bigger brominated test crystal was used to obtain an accurate fluorescence scan for the absorption curve. Three wavelengths in an oscillation range of 60° were collected: inflection, λ=0.9201 Å; peak, λ=0.9199 Å; high energy remote, λ=0.8550 Å. The data quality was not very satisfactory due to a solvent ring in the diffraction pattern, and the close spacing of diffraction spots caused some difficulty in the data processing. In addition, the anomalous signal at the collected white line peak wavelength was actually not at its maximum.

The 2.4 Å MAD data for osmium derivative crystals were collected at the CHESS F2 beamline. A big native crystal soaked heavily with osmium hexamine triflate was used for fluorescence scanning. Four wavelengths were collected with 2° oscillations: inflection, λ=1.1403 Å; peak, λ=1.1395 Å; second inflection λ=1.1391 Å, high energy remote, λ=1.117 Å. All data were processed using the programs DENZO, SCALEPACK (Otwinowski and Minor, 1997), and the CCP4 suite (CCP4, 1994). Data statistics are presented in Table 1A.

Phase Determination

There is one RNA pseudoknot per asymmetric unit, this was initially predicted by calculations based on the long axis cell length (140.08 Å, standard base stacking interval (3.4Å), number of basepairs in the coaxial stacking (9) and

multiplicity in the $P3_221$ space group (6). This prediction was confirmed by the single bromine site located from the 2.1 Å isomorphous difference Patterson map of Native II and the bromine derivative data (Figure 7A).

Initially, when osmium synchrotron data had not been collected, we attempted to use the bromine MAD or MAD combined with single isomorphous replacement for heavy atom phasing. The local scaling technique is often used towards MAD data or isomorphous data to obtain the accurate differences between data sets. In this method, for each reflection in the derivative data set, a sphere of its neighbor reflections in reciprocal space and the corresponding sphere of reflections in the native data set are used to determine a scale factor that is only applied to that central derivative reflection (Rould, 1997). Although local scaling procedures such as NEWLSC were implemented, the lower phasing power of the bromine element in addition to the sub-optimal data quality yielded only partial stacking and connectivity in the electron density maps. The lower refined occupancy (around 0.5) of the bromine site indicated that the uracil might be situated in a mobile loop region that would decrease the bromine phasing power. In addition, the anomalous signal did not resolve the ambiguity for the handedness of the spacegroup in the electron density maps.

The isomorphous difference Patterson maps between the Native II and osmium data collected at home source were not interpretable, despite the repeated peak patterns in different data sets. Using difference Fourier techniques with the bromine phase, we located one major osmium site within the Os derivative data

collected at the synchrotron absorption edge (peak wavelength). This site was confirmed in the Os anomalous difference Patterson maps (Figure 7B). Four other weaker Os sites were later found by difference Fourier techniques, using the bromine and initial osmium phases. The enantiomorph between $P3_221$ and $P3_121$ was indicated early in the difference Fourier step, in which the bromine phase gave a higher cross Fourier peak for the osmium major site in the $P3_221$ space group. This was later confirmed in the better quality electron density maps generated with $P3_221$.

We used a combination of multiple isomorphous replacement and anomalous scattering (MIRAS) methods for phasing with the program SHARP (de La Fortelle and Bricogne, 1997) (Statistical Heavy Atom Refinement Program). The isomorphous or anomalous residual map functions in SHARP are essentially Fourier syntheses calculated from inverse-variant weighted difference coefficients between the derivative and native data. These maps enable the detection of minor sites, and perform this task in an optimal fashion because they take into account the full unbiased phase information available from the data at the current stage of refinement (de La Fortelle and Bricogne, 1997). The enhanced sensitivity to any departure from the current heavy-atom model makes them reveal other subtle features of the heavy atom sites. The residual maps for the MIRAS heavy atom refinement suggested an alternative position for the major osmium binding site as well as anisotropic temperature factors for the bromine and the major osmium site. The electron density maps were greatly improved through these modifications in

refinement. The MIRAS electron density maps were subjected to 130 cycles of solvent flattening (solvent content 43%) by the program SOLOMON (Abrahams and Leslie, 1996) with a mask automatically generated by the program. Phasing statistics are presented in Table 1B.

Model Building

The solvent flattened electron density maps generated from Native II and the bromine/osmium derivative data were of good quality. The initial maps showed good connectivity noted by the high contour phosphate density, ring characteristics of some bases and 2′-OH density. After careful inspection of 5′ 3′ characteristics of nucleic acid geometry, the 5′ 3′ directionality was determined by the following clues: in helical regions, the relationship between the sugar and the two adjacent phosphates is that the 5′-end of the nucleotide has a further distance to the corresponding phosphate and the connection to the phosphate is usually in a horizontal position relative to the sugar. Using the program O (Jones et al., 1991), a poly(rC) model with appropriate sugar puckers was built into the initial map. Although Native I was slightly non-isomorphous with the derivative data, due to the better data quality of Native I and that the program SHARP can account for lack of isomorphism parameters, an excellent 1.9 Å electron density map was generated from Native I and the bromine/osmium derivative data. This map confirmed the trace, clarified the identity of bases, and revealed the position of the

5′-end β-phosphate. All of the RNA molecule except the 5′ triphosphate was built into the initial model.

Structure Refinement

Initial refinement cycles were performed with XPLOR 3.851 (Brünger 1992b) using the multi-endo nucleic acid parameter and topology files. 10% of the diffraction data were selected for the free set for R_{free} calculations (Brünger 1992a). The initial statistics for the unrefined model was: R=0.376, R_{free}=0.337, 8-3 Å. 8-3 Å data were used for initial rounds of rigid body and positional refinement, and 8-1.8 Å data were later used for individual temperature factor refinement, water picking and reiterations of refinement. Early observation of unaccounted density in the MIRAS maps and the later appearance of connected density in the simulated-annealing omit maps suggested two conformations of nucleotide U13 and part of the phosphate backbone of residue C14. The two conformations were included in refinement and relative occupancies calculated (C2′-endo/C3′-endo 60%:40%). During the later stages of refinement, the electron density of the 5′ triphosphate regions improved and this segment was built into the model. The α and β phosphates were well ordered, but the γ phosphate displayed some flexibility. The R factor converged at 25%, R_{free}=29% in the X-PLOR refinement, and the addition of solvent or application of simulated annealing techniques did not lower the R_{free}. Final rounds of refinement to 1.6 Å were performed using CNS 0.3c (Pann and Read, 1996), which rapidly decreased both the R-factor and R-free. Bulk solvent and

anisotropic temperature factor corrections were carried out in the final process. Refinement statistics are presented in Table 1C.

A high density water peak that appeared in the solvent flattened maps was tentatively identified as a sodium ion, with the coordination distance criteria to water molecules or donors of the RNA as 2.4-2.8 Å. The penta-coordination contacts meet the requirements for alkali metal coordination, with ligand distances of 2.77 Å to an N3, 2.81 Å to a 2'-OH, 2.60 Å to a phosphate oxygen, 2.77 Å to a water molecule, and 2.91 Å to an N7. These contacts that contain three hydrogen bond acceptors and two hydrogen bond donors with a square pyramid-like geometry are not found in regular water hydrogen bonding. Although there were also K^+ ions in the crystallization buffer, considering the higher number of electrons and the ionic radius (1.33 Å) of potassium, we can rule out this possibility. During the early stages of refinement, this high density peak was assigned as water which has similar number of electrons as sodium, and a normal temperature factor that assimilated to the neighboring atoms was observed. The magnesium ion was identified by its strict octahedral coordination with ligand distances between 2.1 and 2.3 Å to water molecules and phosphate oxygens. This density peak was assigned as a magnesium ion early in the refinement with the appropriate parameters for electric charge and mass. Throughout the refinement, the octahedral coordination geometry was maintained. Helical parameters for the structure were generated by the program CURVES (Lavery and Sklenar, 1988).

Tables

Table 1. Data Collection, Phase Determination, and Refinement Statistics

A. Crystallographic Data

Data Set	Resolution (Å)	Completeness[a](%)	R_{sym}[ba](%)	I/σ[ca]	Redundancy
Native I	30-1.6	93.0/69.9	7.0/16.5	18.7/3.5	7.1
Native II	25-2.1	91.8/67.2	6.8/18.4	13.1/3.4	5.8
Os(λ 1.1395 Å)	25-2.2	86.8/54.0	5.7/16.9	18.5/5.4	5.4
Br(λ 0.8550 Å)	25-2.0	89.2/88.5	5.3/32.1	16.9/2.4	5.8

Heavy Atom Sites (fractional)	Br	x= 0.7536	y= 0.0683	z= 0.0990
	Os (major)	x= 0.2045	y= 0.9397	z= 0.0607

B. MIRAS Phase Determination (Native I as native)

	Res. Range(Å)	R_{deriv}[d](sites)	R_{cullis}[e]iso(c/ac)	R_{cullis} ano	Phasing Power[f] iso(c/ac),ano
Os	25-2.5	22.7% (5)	0.82/0.84	0.58	0.96/1.3, 2.5
Br	25-2.5	13% (1)	0.87/0.85	0.86	1.10/1.3, 1.2
Overall Figure of Merit after solvent flattening (43% solvent content)					0.896 (1.9 Å)

C. Refinement (Native I, 766 non-hydrogen atoms, 124 water molecules, one magnesium ion, and one sodium ion)

Res. (Å)	Reflections /(free set)	R_{cryst}[g]/R_{free}(%) (2σ)	R.m.s.d.[h] Bonds(Å)	Angles(°)	Ave. B-factor (Å^2) B-factor (Å^2)	(RNA)
8-1.6	8793/916	20.7/25.4	0.017	1.87	7.0	24.6

[a] All data/last shell, the last shell corresponds to 1.66-1.6 Å, 2.2-2.1 Å, 2.28-2.2 Å and 2.06-2.00 Å for Native I, Native II, Os and Br data sets, respectively.

[b] $R_{sym}= \Sigma|I-\langle I\rangle|/\Sigma I$, where I is the observed intensity, $\langle I\rangle$ is the statistically weighted average intensity of multiple observations of symmetry-related reflections.

[c] I/σ= Average intensity/error.

[d] $R_{deriv}= \Sigma\ ||F_{PH}|-|F_P||/\Sigma|F_P|$, where $|F_P|$ is the nucleic acid structure factor amplitude and $|F_{PH}|$ is the heavy-atom derivative structure factor amplitude.

[e] R_{cullis} = <phase-integrated lack of closure>/$\langle|F_{PH}-F_P|\rangle$, c is centric data and ac is acentric data. Iso, isomorphous; Ano, anomalous.

[f] Phasing Power= $\langle[|F_H(calc)|$/phase-integrated lack of closure]>, where F_H is the heavy-atom structure factor amplitude.

[g] $R_{crys}=\Sigma||F_o|-|F_c||/\Sigma|F_o|$, where F_o and F_c are observed and calculated structure factor amplitudes, respectively. R_{free} is calculated for a randomly chosen 10% of reflections; R_{cryst} is calculated for the remaining 90% of reflections used for structure refinement.

[h] R.m.s.d. is the root mean square deviation from ideal geometry.

Figure Legends

Figure 1. Secondary sequence of the pseudoknot from beet western yellow virus. The pseudoknot sequence starts from C3. Two 5′ G residues were added to maximize the yield of the T7 transcription. The second G is in the original mRNA sequence. Loop 1 crosses the major groove of Stem 2 and Loop 2 crosses the minor groove of Stem 1.

Figure 2. BWYV RNA 20% polyacrylamide gel electrophoresis.

(A) After purification, the BWYV RNA full length product is susceptible to degradation at C22-A23. The RNA is 99% pure.

(B) Gel ran on the T7 reaction during initial optimization of yields. The two major bands on the top are n+1 mer and n mer, there are also several bands from abortive products.

Figure 3.. Crystals of the BWYV pseudoknot.

(A) Rod-shaped trigonal crystals (0.6x0.15x0.15 mm^3) that diffract to 1.6 Å.

(B) Cubic form crystals (0.7x0.7x0.5 mm^3) that diffract to 2.85 Å.

Figure 4. 1.6 Å diffraction pattern of Native I trigonal pseudoknot crystals.

2° oscillation frame collected at Broohaven beamline X4A with a Fuji plate detector. There are overloaded reflections at 3.0-3.4 Å along the long axis , which are caused

by continuous base stacking along the c axis. The space group is $P3_221$, cell a=b=30.08 Å, c=140.08 Å, α=β=90.0°, γ=120.0°.

Figure 5. Self rotation search of Native II data.

(A) Section κ=180° reflects the 2-fold symmetry axes in the lattice, generated with a 20 Å integration radius in Patterson space. The native data were processed in spacegroup P1. Data between 15 and 3 Å were used and numbers in the drawings are values for angle φ. Thus the crystallographic 2-fold x axis runs along the horizontal (φ=0°), the crystallographic 2-fold y axis is at φ=120°, and there is a crystallographic 2-fold axis on the diagonal at φ=60°. Since there is only one RNA molecule in the asymmetric unit, the noise peaks at φ= 30°, 90°, 150° indicate a pseudo 2-fold symmetry within the pseudoknot.

(B) The crystallographic 3-fold z axis is perpendicular to the plane of projection and appears on κ=120°.

Figure 6. Anomalous scattering plots of heavy atom elements.

(A) Br anomalous scattering curve, plotting theoretical values of f'' (top line, positive value) and f' (bottom line, negative value) as a function of X-ray energy. X-ray energy in keV=12.398/λ in Å. The f'' value makes a jump (from 0.5 to 3.8) between 13471 to 13476 eV, and reaches its maximum at the white line peak 13476 eV; the inflection point with maximum f' (-8.3) is in theory the wavelength at half height of the peak f''. Br MAD data were collected at 13474 eV (inflection), 13478.1

eV (peak) and 14500 eV (high energy remote, f'= -1.63, f''=3.3). Due to the higher resolution and better quality of the high energy remote data, this data were used in subsequent MIRAS methods.

(B) Os anomalous scattering plot. The f'' value makes a jump (from 3.8 to 10.2) between 10868 to 10878 eV, and reaches its maximum at the white line peak 10878 eV; the inflection point with maximum f' (-19.53) is in theory the wavelength at half height of the peak f''. Os MAD data were collected at 10873eV (inflection), 10881 eV (peak), 10884 eV (second inflection), and high energy remote at 11100 eV (f'= -9.96, f''=9.86). The experimental X-ray energy for inflection and peak may vary depending on the calibration and chemical environment for each compound. The peak wavelength data were used in subsequent MIRAS analysis.

Figure 7. Patterson maps of derivative data

(A) Isomorphous Patterson map between bromine data and Native II at 15-2.1 Å. There is one Br atom per asymmetric unit. The sections are on Z= 0.15, 0.333 (Harker), 0.81, 0.48 respectively.

(B) Anomalous Patterson map of osmium derivative data collected at the absorption edge at 20-2.5 Å. This Harker section is on Z=0.333. Site 1 is the major osmium binding site, Site 2 and 3 are medium binding sites. The Patterson peaks of the remaining two weak sites are not visible on this map.

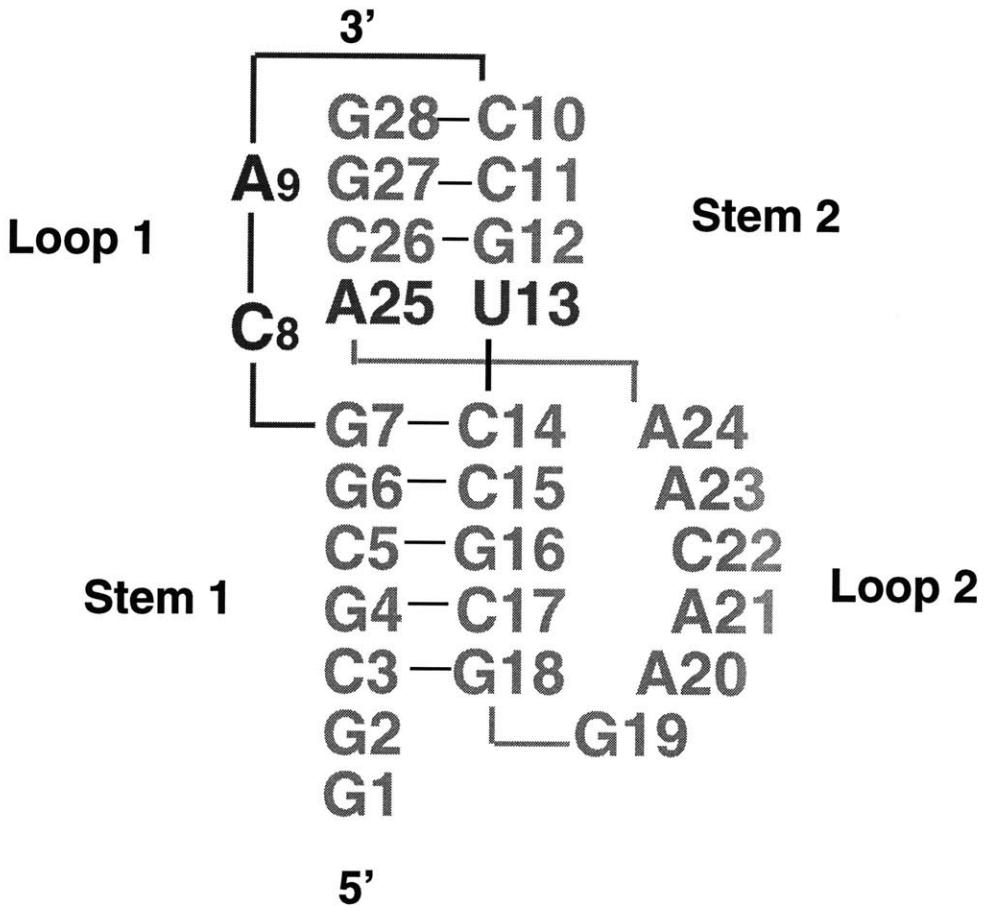

3'
G28–C10
A9 G27–C11
Loop 1
C26–G12
Stem 2
A25 U13
C8
G7–C14 A24
G6–C15 A23
C5–G16 C22
Stem 1
G4–C17 A21
Loop 2
C3–G18 A20
G2 G19
G1
5'

Figure 1

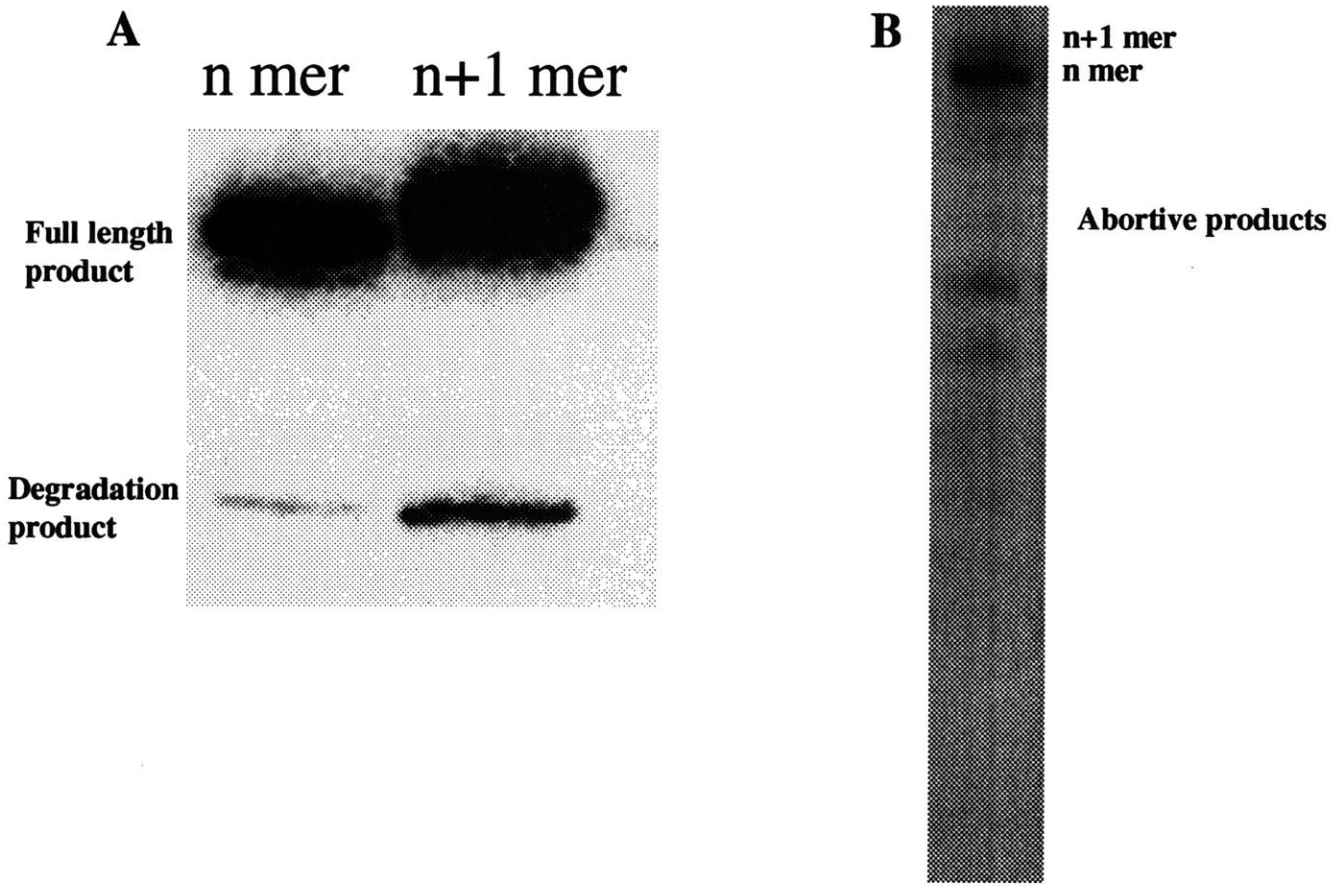

A
n mer
n+1 mer
Full length product
Degradation product
B
n+1 mer
n mer
Abortive products

Figure 2

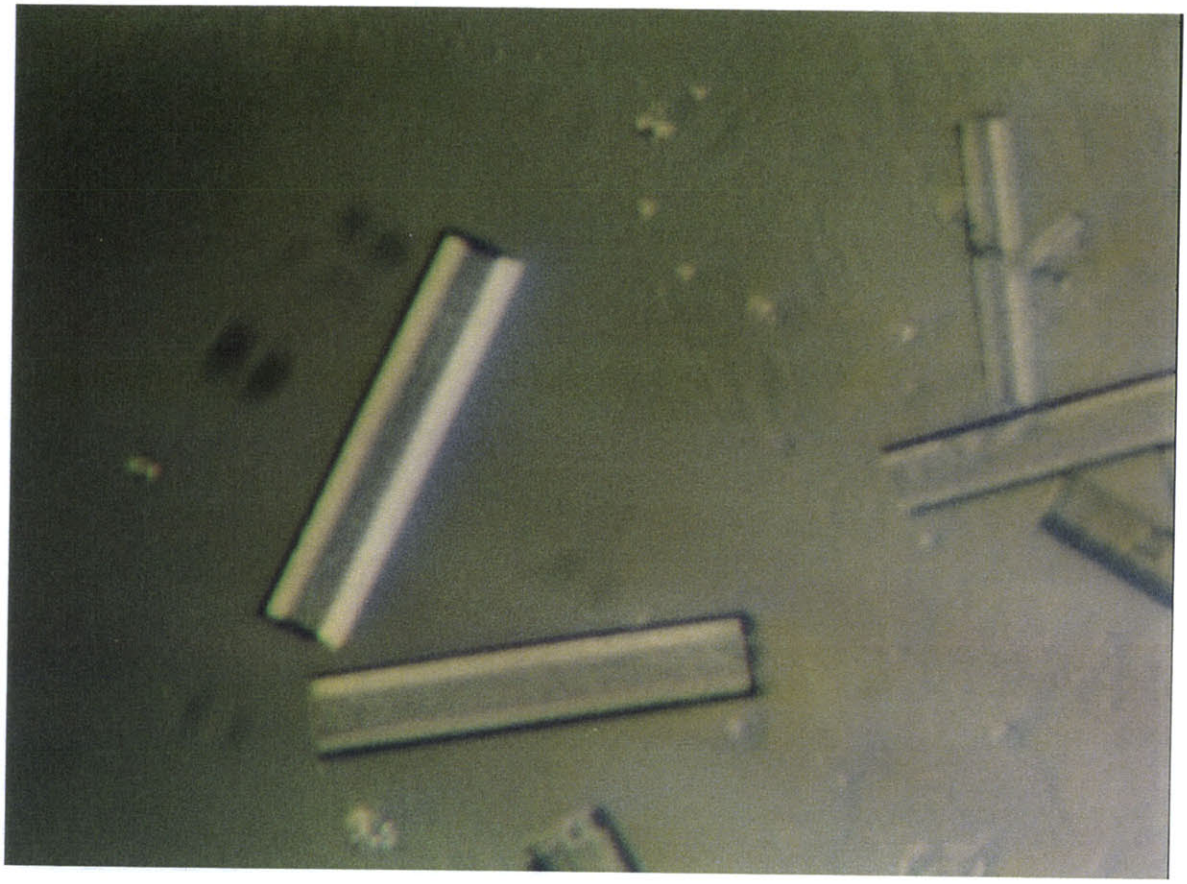


Figure 3A

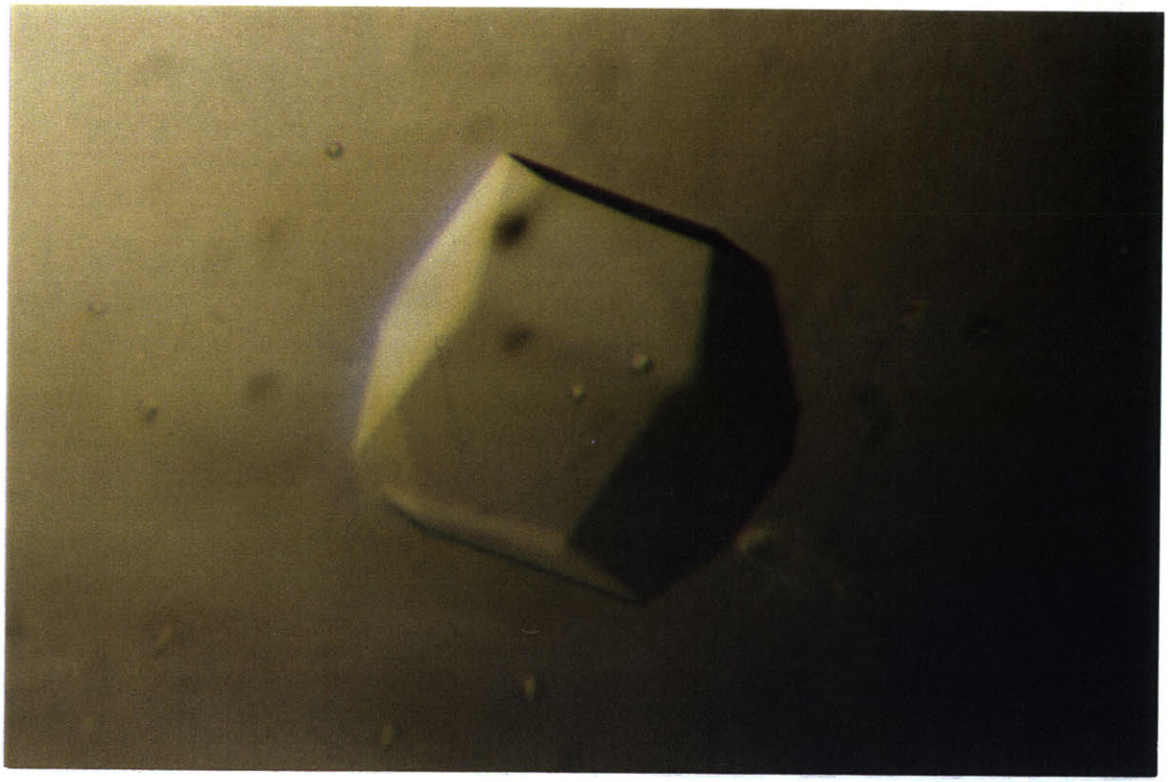


Figure 3B

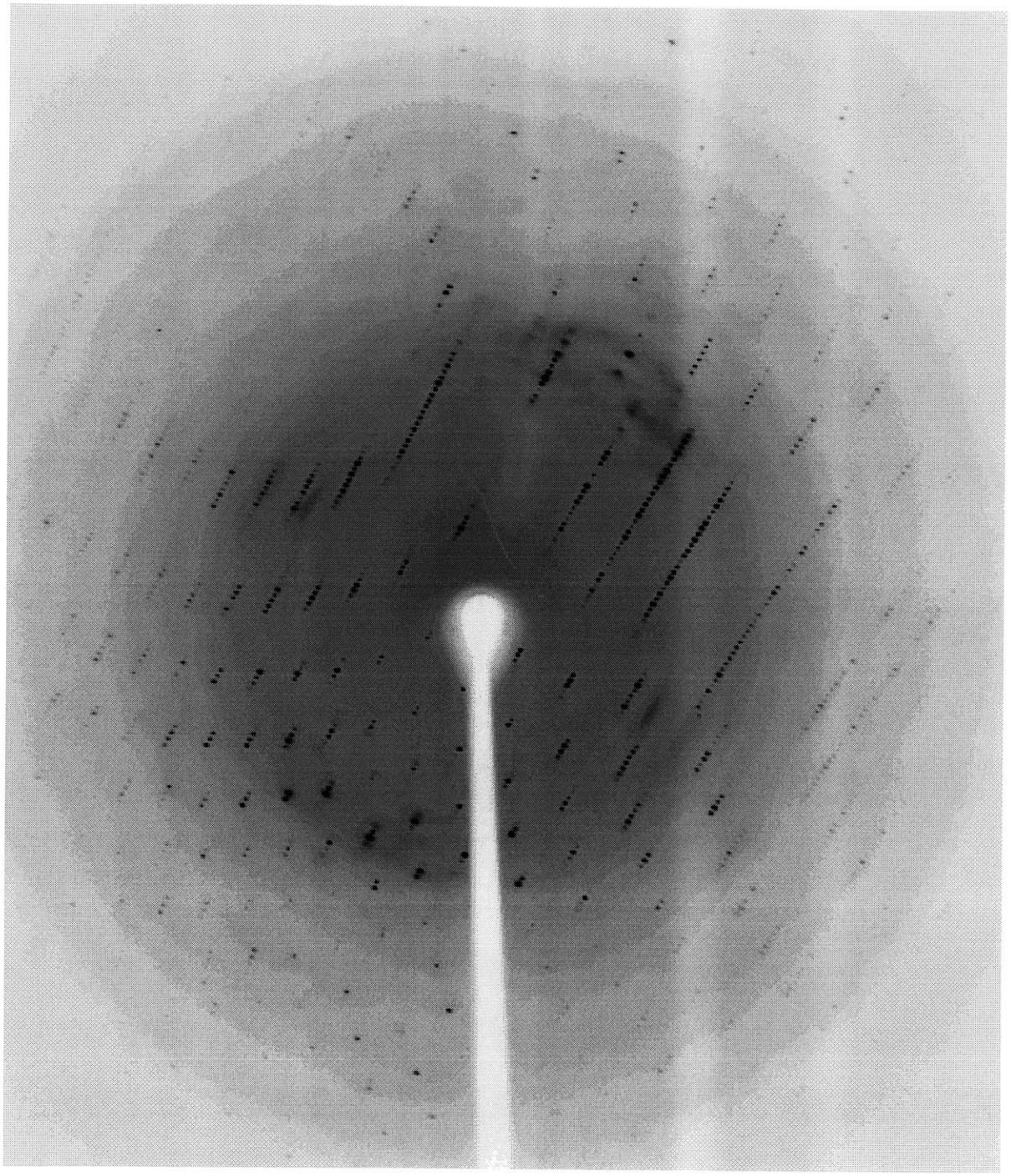


Figure 4

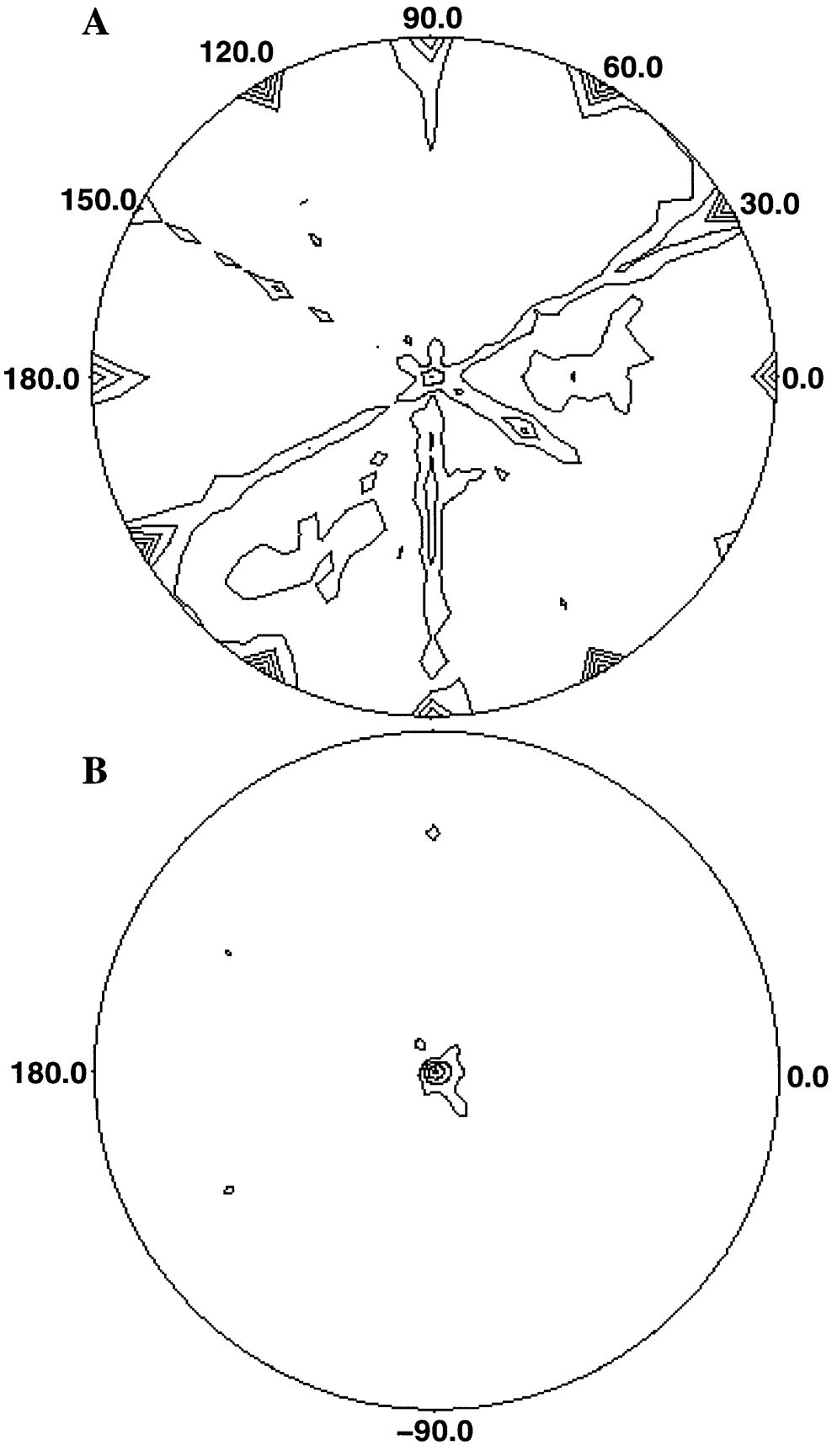


Figure 5

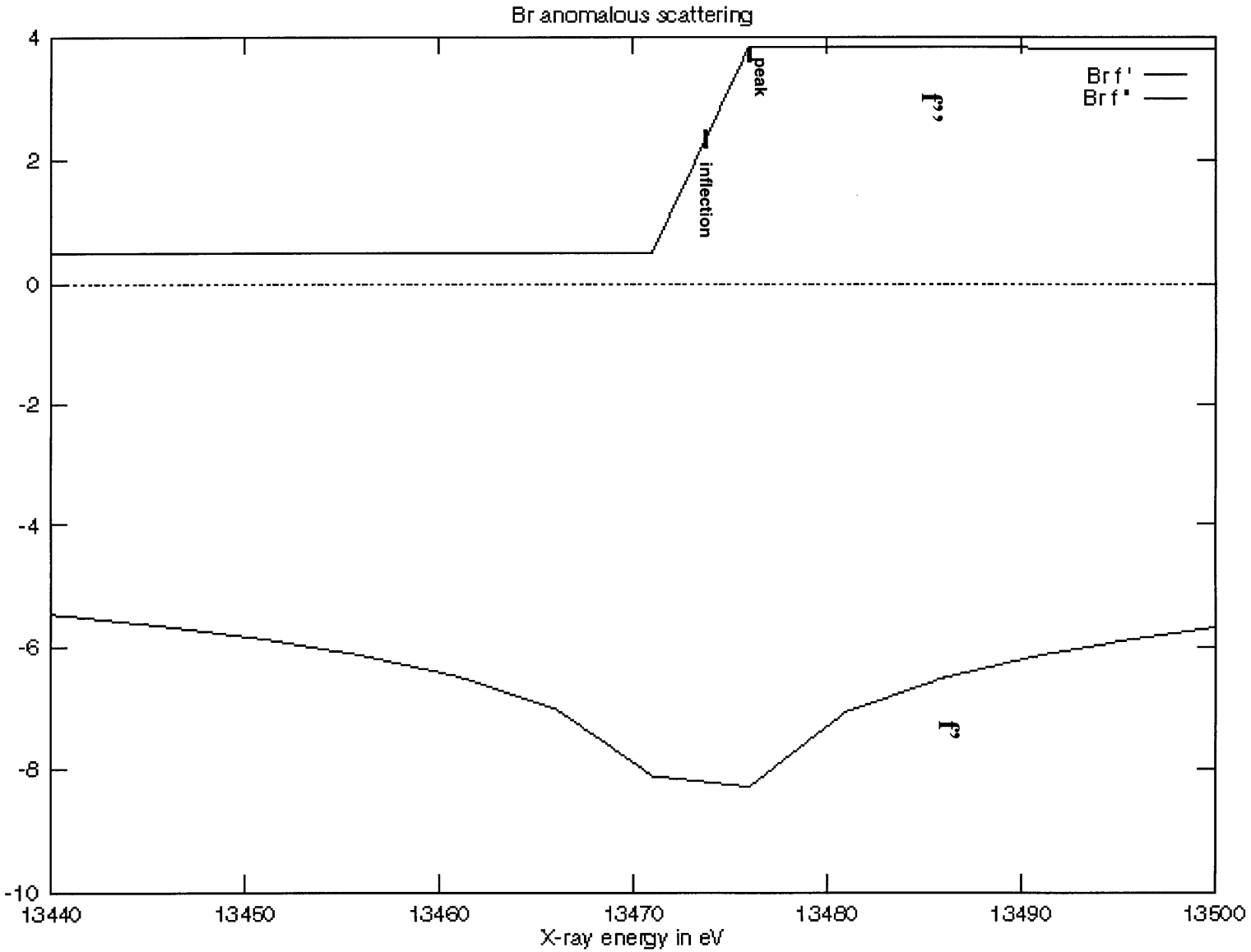

Br anomalous scattering
Br f'
Br f"
peak
inflection
f''
f'
4
2
0
-2
-4
-6
-8
-10
13440
13450
13460
13470
13480
13490
13500
X-ray energy in eV

Figure 6A

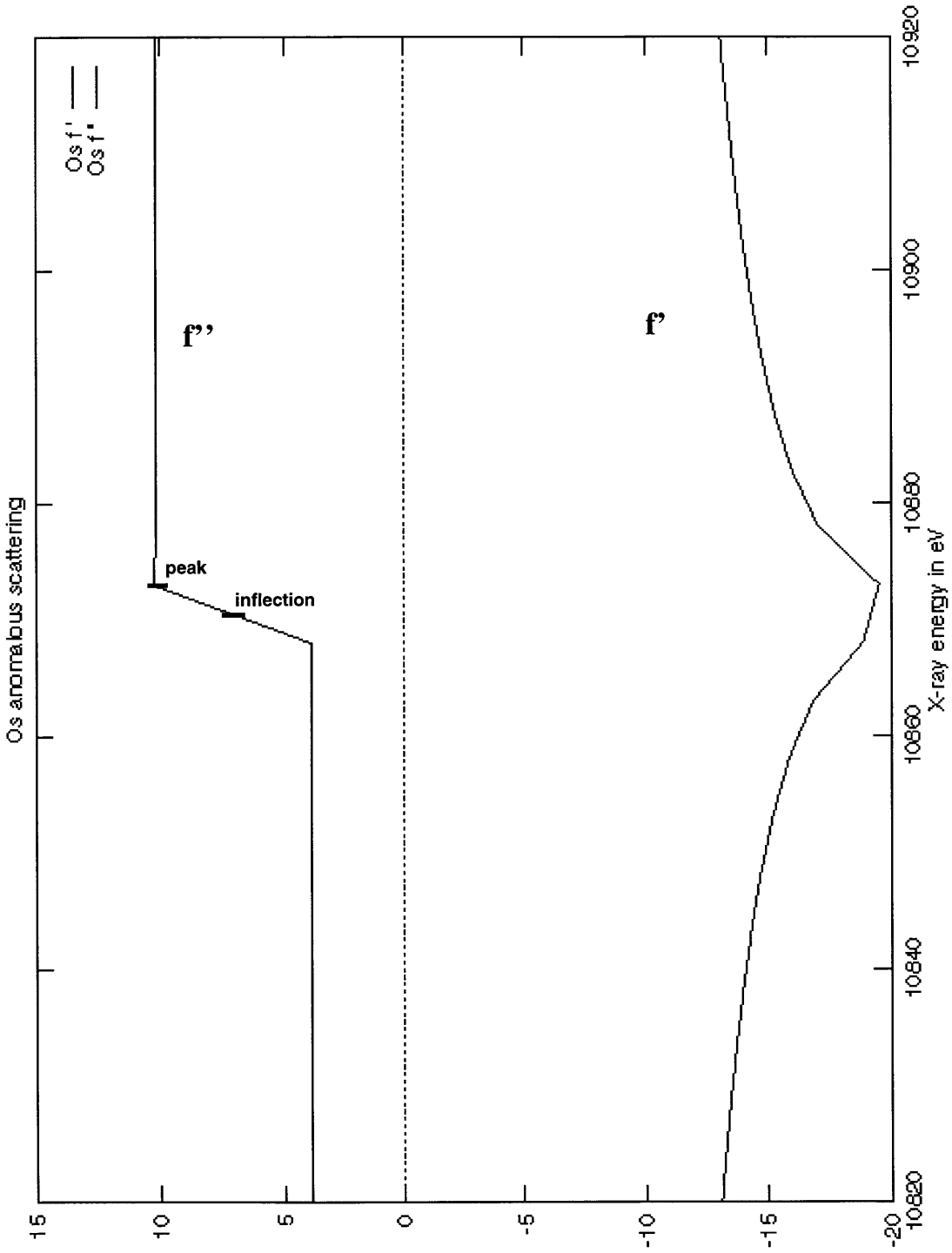

Os anomalous scattering
Os f'
Os f"
f''
f'
peak
inflection
15
10
5
0
-5
-10
-15
-20
10820
10840
10860
10880
10900
10920
X-ray energy in eV

Figure 6B

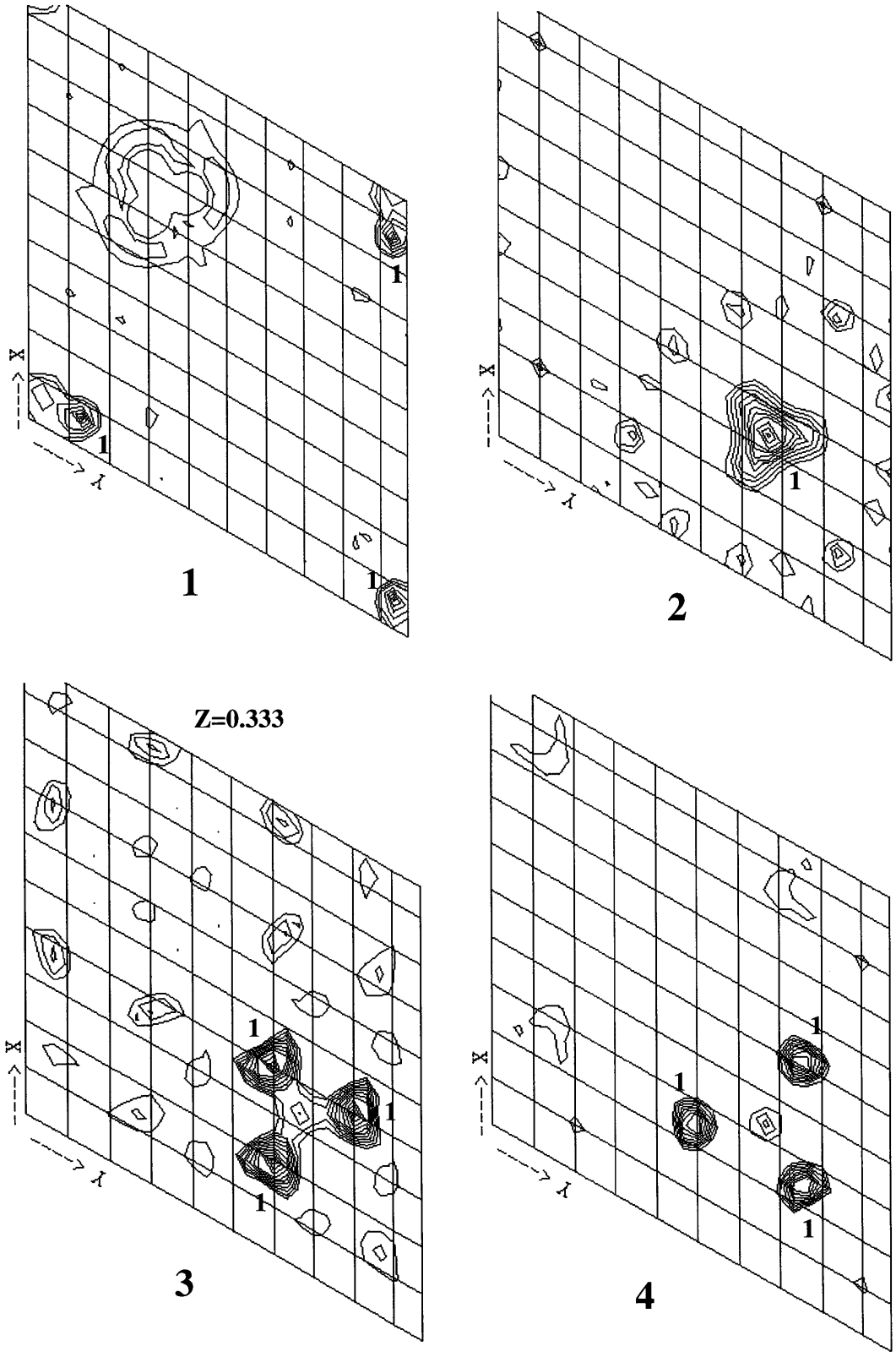


Figure 7A

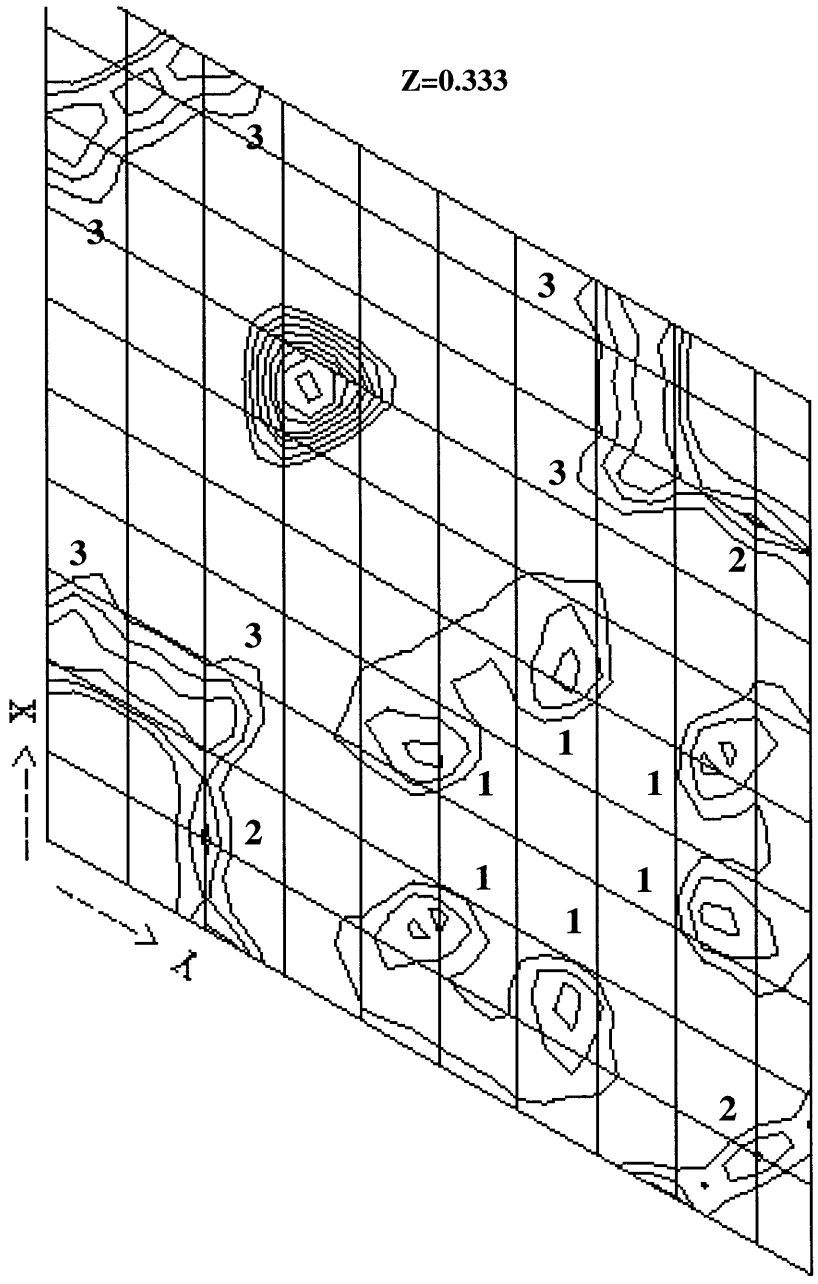


Figure 7B

Chapter 3

Minor Groove RNA Triplex in the BWYV Pseudoknot

Part of this chapter describing the BWYV pseudoknot structure will be submitted as a manuscript titled " A Minor Groove RNA Triplex in the Crystal Structure of a Viral Pseudoknot involved in Ribsomal Frameshifting ", Authors include Li Su, Liqing Chen, Martin Egli, James M. Berger and Alexander Rich.

Summary

Many viruses regulate translation of polycistronic mRNA using a –1 ribosomal frameshift induced by an RNA pseudoknot. A pseudoknot has two stems that form a quasi-continuous helix and two connecting loops. A 1.6 Å crystal structure of the beet western yellow virus pseudoknot reveals rotation and a bend at the junction of the two stems. A loop base is inserted in the major groove of one stem with quadruple base interactions. The second loop forms a new minor groove triplex motif with the other stem, involving 2′-OH and triple base interactions, as well as sodium ion coordination. Overall, the number of hydrogen bonds stabilizing the tertiary interactions exceed those involved in Watson-Crick basepairs. This structure will aid mechanistic analyses of ribosomal frameshifting.

Results

Overview of the Structure

Here we present a 1.6 Å resolution crystal structure of a 28 nucleotide pseudoknot from beet western yellow virus (BWYV), a plant luteovirus, which is active in ribosomal frameshifting and regulates the production of an RNA-dependent RNA polymerase (Miller et al., 1995).Experimental electron density maps are shown in Figure 1. The overall structure is very compact (Figure 2), with Van der Waals dimensions of approximately 32x36x22 $Å^3$. The first dimension is in the 5′ to 3′ direction of the molecule, and the longest dimension incorporates the peripheral Loop 2 region. Some aspects of the structure deviate from the conventional secondary structure diagram, which was assumed in the absence of structural data (Figure 2B). The junctional A25•U13 pair predicted at the bottom of Stem 2 is not formed. Instead, the U13 nucleotide bulges out of the helical region, and A25 is tilted and stacked upon Loop 2 (Figure 2A).

The two stems are non-coaxially stacked and appear to hinge at the junction. Continuous stacking of Loop 2 onto A25 (magenta) of Stem 2 (Figure 2D) and the short length of Loop 1 induce the rotation of Stem 2 relative to Stem 1 (Figure 2C). There is a 48° rotation between the bottom base of Stem 2 (A25) and the top base of Stem 1 (G7), therefore, they are not stacked (Figure 2C). The helical axes of the two stems are shifted relative to one another and displaced by approximately 5 Å, and the average tilt angle between the two stems is approximately 25°. It has been previously suggested that pseudoknots with loop

lengths shorter than the minimum length required for crossing the grooves of an A-form helix may impose distortion in the stem regions (Pleij et al., 1985). The length of Loop 1 is relatively short for stretching across a major groove opening between Stem 1 and Stem 2 that is enlarged by their relative rotation and bend (Figure 2A and 2C). This considerable stretch is facilitated by overtwisting the basepairs near Loop 1 in both stems.

The geometry of Stems 1 and 2 differ significantly from standard A-form duplex RNA (Figure 2C). In both stems, the basepairs are largely propeller twisted; Stem 1 has an average propeller twist of 15°, while it is 26° for Stem 2. In addition, the basepairs in Stem 1 are tilted in the range of -3° to 5° from the helix axis, while the A-form RNA duplex usually tilts 16-19° (Saenger, 1984). Stem 1 is curved towards the major groove, as the helical axis gradually bends 13° from bottom to top. The minor groove of Stem 1 is narrower by 2 Å on average and 1.8 Å more shallow than standard A-form; the major groove is approximately 0.7 Å wider. From top to bottom, Stem 2 bends moderately towards the major groove by 7.5°.

An Extended Minor Groove RNA Triplex

In several luteovirus frameshifting pseudoknots, the sequence of AACAAA (Loop 2 A20-A24 and the continuing junctional base A25) is conserved (Miller et al., 1995). As shown in Figure 3A, Loop 2 forms a stacking ladder with each adenine (or cytosine) rotated in somewhat different orientations

to yield a variety of interactions with the minor groove of Stem 1. This extended RNA triplex, which has six layers of consecutive interactions in the minor groove has not been previously seen, and differs substantially from the nucleic acid triplexes with the third strand in the major groove (Felsenfeld et al., 1957; Rajagopal and Feigon, 1989). The triplex is RNA specific, as every interaction between a Loop 2 nucleotide and Stem 1 involves a 2'-OH group. The triple base interactions are not co-planar in many cases. Several interactions involve the N1 and N6 of loop adenines, and N3 and N2 of guanines in the stem. These interactions with the third strand contribute to the large propeller twists in Stem 1, and the gradual bend towards the major groove.

The first base of Loop 2 (G19) projects away from Stem 1 (Figure 2A). However, the adjacent residue (A20) is 30° out of plane from the G4•C17 basepair (Figure 2A), leading to a stabilizing stacking interaction with the ribose O4' of residue C5, as found in many nucleic acid structures (Egli et al., 1995). This anchor-like adenosine spans the minor groove and interacts with both strands of Stem 1. It is held in this position by an extensive network of 7 hydrogen bonds (Figure 3B), using both base and sugar interactions. The 2'-OH of A20 contributes 4 of the hydrogen bonds by interacting with two layers of bases (Table 1). Two of the hydrogen bonds are bifurcated, involving O2 atoms of residues C5 and C17, resulting in the tilt of these bases (Figure 2A or 2C). Some of these adenosine interactions have also been seen in the crystal structures of the hammerhead ribozyme (Pley et al., 1994b) and group I intron (Cate et al., 1996).

In both cases, the adenosine in an RNA tetraloop contacts the minor groove side of a G•C basepair to form 4 of the 7 hydrogen bonds observed here. In contrast to A20, the adenines there are not tilted and are largely in the plane of the basepair.

The next two residues A21 and C22 (Figure 2A) are somewhat rotated away from the minor groove but the bases are stacked. The backbone conformation is extended and appears to be stabilized by hydrogen bonding between bases to the 2'-OH of G16 (Figure 3C) and by a sodium ion that mediates base-base interactions (Figure 5C). The adjacent A23 is stacked on the exocyclic amino group of C22 (Figure 3A) and utilizes its Watson-Crick face to interact with the neighboring C15 (Figure 3D). A24, the last base of Loop 2 stacks on A23 (Figure 3A) and interacts with guanosine G7 through three hydrogen bonds (Figure 3E). This minor groove specific interaction is unlike other G•A interactions described previously (Pley et al., 1994a and 1994b; Scott et al., 1995; Cate et al., 1996; Brown and Hunter, 1997). The junction between A24 and A25 is stabilized by a 2.82 Å hydrogen bond between the 2'-OH of A24 and the O4' of A25 (not shown). With the exception of A20, Loop 2 is not involved in lattice contacts.

The minor groove triplex formed by Loop 2 and Stem 1 has a total of 16 hydrogen bonding interactions, 7 to one strand of Stem 1 and 9 to the other (Table 1). The extended backbone conformations in Loop 2 are related to the interactions with Stem 1 that switch from one strand to the other. For instance, P-P distances of 6.8-6.9 Å are found between A20-A21 and A23-A24, where the

first base interacts with one strand of Stem 1, and the next base interacts with the opposite strand (Figure 2A). In addition, temperature factors in the backbone of Loop 2, in particular A21 through A25, are 60% higher than those in Stem 1. This observation may be associated with the variation in P-P distances.

The Stem 1-Stem 2 Junction

The first residue of Stem 2, A25, stacks on the edge of the A24 base (Figure 3A). There, it is tilted and interacts with bases in different layers (Figure 2A and Figure 4). Rather than pairing with U13 as predicted (Figure 2B), the A25 adenine forms three hydrogen bonds at the junction through its Watson-Crick face (Figure 4B). The interactions with the O2 and 2′-OH of C14 are similar to those observed between A23 and C15 (Figure 3D). On the strand opposite to A25 (Figure 2A and 2C), the extruded U13 has a rotated and reversed sugar direction in relation to the adjacent nucleotides. This results in considerable compression of the backbone with a close U13-C14 P-P distance of 4.65 Å, which allows G12 to directly stack upon the base plane and furanose oxygen of C14 (Figure 2C). Other bulge nucleotide conformations have been observed (Quigley and Rich, 1976; Klimasauskas et al., 1994; Portman et al., 1996), although their sugar rotations are less marked. In the crystal structure, U13 adopts two conformations with C2′- or C3′-endo sugar pucker. However, the uridine bases in both conformations are nearly superimposable, possibly due to lattice contacts that include hydrogen bonding to the N6 and N7 of a symmetry-related A20 and additional stacking

interactions. For clarity, we have included only the dominant C2′-endo conformation in the figures.

Quadruple Interactions of Loop 1

Most frameshifting pseudoknots have a short Loop 1, usually 1 to 2 nucleotides in length (Brierley et al., 1991; Chamorro et al., 1992; ten Dam et al., 1994; ; Chen et al., 1995; Miller et al., 1995). In the BWYV pseudoknot, a C8-A9 sequence crosses the major groove, and the C residue is conserved in some luteovirus pseudoknots (Miller et al., 1995). In the BWYV structure, C8 is inserted into the major groove of Stem 2, forming a quadruple base interaction with G12, A25 and C26 (Figure 2A and Figure 4A). This is the first example of one base hydrogen bonding to three other bases, two of which form a basepair. The G12•C26 basepair has a strong propeller twist, which is needed for C26 to stack upon the tilted A25 (Figure 2C and Figure 4A). As noted by the asterisk in Figure 4A, the distance between N3 of C8 and O6 of G12 is 2.82 Å, and it is likely that C8 is protonated in the N3 position to form a hydrogen bond. Protonation of cytosines are also found in other nucleic acid structures (Hartman and Rich, 1965; Gehring et al., 1993; Chen et al., 1994). Two water molecules stabilize this core interaction by mediating C8 2'-OH to base contacts (A25 and C26) (Figure 5B). The fact that a cytosine at position 8 is crucial for frameshifting activity (Kim et al., unpublished data) suggests that C8 may have an organizing role in assembling the junction between the stems. Overall, these major groove

interactions and the consecutive minor groove triplex appear to drive the rotation and bend at the Stem 1-Stem 2 junction.

Sharp Turns, Ions and Water Molecule Stabilization

In complex RNA structures, changes in backbone direction are achieved by sharp turns such as the U turn seen in tRNA (Quigley and Rich, 1976), the hammerhead ribozyme (Pley et al., 1994a; Scott et al., 1995) and group I intron (Cate et al., 1996), which are stabilized by multiple interactions. In the BWYV structure, we have named the Loop 1 to Stem 2 transition a "C turn" (Figure 5A), which has some features in common with the U turn. The "C turn" involves the change in direction between Loop 1 and the top of Stem 2, through three nucleotides C8, A9, and C10, the last C being base paired. The turn is less sharp than the 180° U turn, but is stabilized by a hydrogen bond from N4 of C10 to the phosphate oxygen of A9, which is the second phosphate in the 5′ direction. In comparison, the U turn uses a uracil N3 hydrogen bonded to the fourth phosphate group in the 3′ direction. In addition, adenine 9 is partially stacked at an angle on cytosine 10, which is associated with a parallel tilting of C10, propeller twisted relative to its base-paired G28 (Figure 2C). An organized network of water molecules facilitates the turn by bridging phosphate or base contacts (Figure 5A). The array of water molecules on the other side of Loop 1 stabilizes its conformation in the major groove (Figure 5B and 6).

A sharp turn is also found at the junction between Stem 1 and Loop 2 at G18-G19-A20 (Figure 2A and Figure 2D). This turn is facilitated by C2'-endo sugar puckers (the remainder of the RNA is C3'-endo, except for U13) at residues G18 and G19, which lead to an extended backbone conformation (Figure 7). In contrast to the C turn, the only interaction at this turn is a self base (N3) to 2'-OH interaction at G19 fixing the base orientation. However, G18 is base-paired and there are multiple interactions binding A20 to Stem 1 (Figure 3B). Thus, the turn is stabilized by fixing both ends.

RNA pseudoknots are generally stabilized from unfolding by magnesium ions (Wyatt et al., 1990). However, in many cases, high concentrations of monovalent ions can also stabilize the conformation (Wyatt et al., 1990; Gluick et al., 1997). The sodium ion found in the minor groove is located at a key position to tie together Stem 1 and Loop 2, where a base-base contact is not possible due to the gap (Figure 3C). Water molecules are also involved in stabilizing the conformation around the sodium ion. An octahedrally coordinated magnesium ion is bound to the triphosphate at the 5'-end (Figure 2D), coordinating to phosphate oxygens of the β and γ phosphate, the pro-S_p oxygen of the G2 phosphate and three water molecules.

Lattice Packing and Heavy Atom Binding

There is one RNA pseudoknot molecule in the asymmetric unit. The molecule is roughly positioned at the corner of the ab plane. Six molecules

related by two-fold and 3_2 symmetry wind up the three-fold c axis to form an infinite helix by head-head or tail-tail terminal basepair purine stacking. At the 5′-end, G18 is stacked on a 2-fold related G18′; at the 3′-end, G28 is stacked on a 2-fold related G28′. In addition, at the 5′-end, the 2-fold symmetry related G2′ forms a Hoogsteen pair with G18 in the major groove. Furthermore, hydrogen bond donor and acceptor groups on the adenine 9 make multiple contacts to the 3′-end of a 2-fold related molecule. The contacts are: N6 of A9 to N1 of G28′ and O2 of C11′; N7 of A9 to the 2′-OH of G28′; N1 of A9 to 2′-OH of C11′.

There are large solvent channels in the crystal, in which exposed nucleotide bases from four symmetry related molecules form a stacking core. These nucleotides include a bulge G19 from one molecule, extruded U13′ and G1′ from the second molecule, another U13″ and G1″ from the third molecule, and G19‴ from the fourth molecule. The continuous stack from bottom to top is in the order: U13″ G19‴ G1′ G1″ G19 U13′. In addition, the U13 nucleotides in this stacking core use their Watson-Crick faces to form Hoogsteen basepairs with symmetry related A20 nucleotides.

The osmium hexammine triflate compound has been previously used in the structure determination of the group I intron P4/P6 domain (Cate et al., 1996a). There, the Os sites were found in the major groove near non-canonical basepairs G•U, G•G and U•U (Cate et al, 1996b). Likewise, the structurally similar cobalt hexammine was found directly coordinating to G•G pairs in $tRNA^{Phe}$ (Jack et al., 1977) as well as in A-form DNA (Nunn and Neidle, 1996). In

the BWYV structure, the two stronger osmium binding sites with higher occupancy and lower temperature factors were also found in the major groove near G6 and G7. Other low affinity binding sites were found near clusters of phosphates in the solvent channel between symmetry related molecules or near sharp turns. For example, in the region of the A20/U13 lattice contact; the Stem 1-Loop 2 junction involving residues 18,19 and 20. As the distance between the osmium sites and RNA ranged from 4-5 Å, this indicated that the coordinated osmium may contact the RNA indirectly through the hexammine groups.

Discussion

Structural Comparison with other Frameshifting Pseudoknots

The BWYV crystal structure provides a first detailed look at a frameshifting pseudoknot with ions and water molecules, and completely defines the interactions of loops. The geometry at the junction of the stems is fully defined, with an average 25° tilt angle between the two stems, this is significantly less than the 60° angle reported in the MMTV pseudoknot (Shen et al., 1995). The smaller tilt angle observed in BWYV may be due to the absence of an adenosine to serve as a hinge between the two stems, as well as different sequences and loop lengths. In the MMTV pseudoknot, the major groove opening between Stem 1 and Stem 2 is also widened, which is due to the stacked adenosine. Similar to the A25•U13 junctional pair, the corresponding G•U pair in MMTV is referred to as a breathing basepair. In contrast, in the SRV-1

pseudoknot, the assignment of the imino proton resonance of the junctional base-paired U was unambiguous (Du et al., 1997), although controversy remains in the literature (Chen et al., 1996).

In the BWYV crystal structure, we observe in detail the loop-stem interactions that were not unambiguously assessed in other structures due to the limited resolution. In the MMTV NMR converged structures, the bases in Loop 1 point away from the groove with only one exception (Shen et al., 1995). In one of the structures, a Loop 1 guanosine resides in the major groove and can interact with a residue in Stem 2, which is analogous to that observed in BWYV. Having a longer Loop 2, the 5′ bases in the MMTV Loop 2 are stacked under Stem 1 and the remaining adenine bases are more or less stacked in the loop; several Loop 2 and stem-loop junction nucleotides are suggested to be in the C2′-endo sugar conformation. In BWYV, although extended backbone conformations are observed, C3′-endo conformation is predominant except at the G19 sharp turn.

A comparison is made in Figure 8 between the BWYV conventional secondary sequence and schematic presentation of the crystal structure, we note several major differences: Due to the continuous stacking of Loop 2 into Stem 2 through junctional base A25, this segment of the pseudoknot is rotated to the other side so that the two stems are non-coaxial; A25 is not stacked on G7 to form a continuous helix, this difference from the conventional prediction is clearly illustrated in Figure 9. A25 is on a higher level than U13, and does not base-pair with the extruded U13 as predicted, G12 directly stacks upon C14; C8 stacks on

G7 and forms a base quadruple interaction with G12•C26 in the same layer, and A25 from below. These major deviations from the secondary sequence were not apparent from previous NMR structures. In addition, we were also able to observe the detailed deviations from standard A-form helix in Stem 1 and Stem 2 induced by the majority of tertiary interactions, which could not be accurately determined through other lower resolution methods.

Relevance to RNA Packing

The BWYV structure presents the first detailed view of an extended minor groove RNA triplex, a new motif for mediating intramolecular and possibly intermolecular interactions. Interactions of this type have been predicted to form between the P1 substrate helix and the single-stranded J8/7 segment in the active site of the *Tetrahymena* intron (Szewczak et al., personal communications). Their study suggested four consecutive base-triples and 2'-OH mediated tertiary interactions form from the P1 helix. In BWYV, 10 of the 16 hydrogen bonding contacts in the minor groove are mediated by 2'-OH groups. The 2'-OH groups extend from both edges of the minor groove and are involved in every Loop 2 nucleotide interaction. Therefore, it is not surprising that the minor groove is narrowed to clamp the third strand into close contact. The Loop 2-Stem 1 interactions may also be a general feature in other pseudoknots that are not involved in frameshifting (Pleij, 1994), as shown in the solution structure of the turnip yellow mosaic virus genomic RNA (Kolk et al., 1998). Furthermore, in

the BWYV Loop 2 conformation, the 2'-OH and certain hydrogen bond donor/acceptor groups of the adenine bases are positioned on the outside surface where they are available for higher-order contact (Figure 3A). It is possible that this interface is recognized by proteins or can have dynamic interactions with the ribosome. An interface of this kind could also be involved in domain organization for packing of large RNA molecules (Ortoleva-Donnelly et al., 1998).

Tables

Table 1. Tertiary Hydrogen Bonds

Residue		Distance (Å)	Residue		Distance (Å)
Loop 2					
A20 2'OH[a]	C5 O2	2.83	A21 N7	G16 2'OH	3.05
A20 2'OH[a]	C17 O2	2.81	A21 N6	G16 2'OH	2.90
A20 2'OH	G16 N2	3.17	A21 O1P	G16 N2	2.89
A20 2'OH	C17 2'OH	2.57	C22 N4	G16 2'OH	3.30
A20 N1	G4 2'OH	2.70	A23 N1	C15 2'OH	2.69
A20 N3	G4 N2	3.02	A23 N6	C15 O2	2.98
A20 N7	C5 2'OH	3.30			
A24 N1	G7 N2	3.16	A24 N6	G7 N3	3.01
A24 N6	G7 2'OH	2.91	A24 2'OH	A25 O4'	2.82
Loop 1/Junction/Sharp turn					
C8 N4	G12 N7	3.10	A25 N6	C8 O2	2.97
C8 O2	C26 N4	3.23	A25 N6	C14 O2	2.84
C8 O2	A25 N6	2.97	A25 N1	C14 2'OH	2.70
C8 N3[b]	G12 O6	2.82			
A9 O2P	C10 N4	2.80	G19 2'OH	G19 N3	2.77

[a] Bifurcated hydrogen bonds

[b] Probably protonated

Figure Legends

Figure 1. BWYV pseudoknot electron density maps at 1.9 Å resolution contoured at 1.0 σ above the mean.

(A) View of the refined structure in the minor groove RNA triplex region superimposed on the solvent flattened experimental electron density map. Loop 2 is shown in green, stem regions in gold, Loop 1 in red.

(B) Detailed view of the refined structure superimposed on the same map in the major groove at the Stem 1-Loop 1 junction. Stems are in gold, C8 in red, A25 in magenta. In both maps 2'-OH and phosphate oxygen density, as well as ring characteristics of base density are clearly visible. These figures were generated with the program O (Jones et al., 1991).

Figure 2. The beet western yellow virus (BWYV) pseudoknot secondary sequence and crystal structure.

(A) Stereo view of the pseudoknot crystal structure. The color scheme of bases corresponds to that used in the secondary structure diagram in B. Loop 1 (red) crosses the major groove of Stem 2 and Loop 2 (green) stacks in the minor groove of Stem 1. The stem backbone is blue; bases are gold.

(B) The conventional representation of the BWYV pseudoknot secondary sequence. Stem 1 of the wild type pseudoknot sequence starts at C3, G1 was added at the 5'-end to assist transcription and G2 is in the original mRNA

sequence (Miller et al., 1995). The actual structure has significant differences: A25•U13 is not paired, and the A25-G28 strand in Stem 2 is flipped to the other side.

(C) Stereo view at a slightly different angle of Stem 1 and Stem 2 without the loop nucleotides. Stem 2 is rotated relative to Stem 1 and is non-colinearly stacked, all basepairs are highly propeller twisted. At the junction, G12 is stacked on C14, but on the opposite strand the bottom base A25 of Stem 2 (magenta) is not stacked on G7, the top base of Stem 1.

(D) The general fold of the pseudoknot and metal ions. A magnesium ion in rose color binds at the 5′ triphosphate region. One sodium ion in orange is coordinated in the minor groove.

Figure 3. RNA triplex interactions of Loop 2 in the minor groove of Stem 1.

(A) Separate view of the conserved predominantly adenosine ladder. Despite the systematic stacking, each loop base is rotated in different orientations to maximize interactions with the groove nucleotides.

(B) The tilted A20 interacts with 2 layers of basepairs through a base triplet, and a 2′-OH multiple hydrogen bonding network. For clarity, the ribose of C5 is omitted, but the weaker hydrogen bond between N7 of A20 and C5-2′-OH is indicated.

(C) A21 and C22 contact G16 through another 2′-OH multiple interaction. The Hydrogen bond from the C22 amino group is relatively weaker.

(D) A23 forms a triple base interaction where the 2′-OH of C15 interacts with N1 of A23.

(E)A24 forms a unique interaction with G7. Not shown is a hydrogen bond between the 2′-OH of A24 and the O4′ of the A25 furanose ring.

Figure 4. Quadruple base interactions of Loop 1.

(A) The C8 organizer base of Loop 1 inserts deeply into the major groove of Stem 2 and interacts simultaneously with bases of three other nucleotides. The asterisk next to N3 of C8 indicates probable protonation with a hydrogen bond to O6 of G12. C26 is propeller twisted in the C26•G12 basepair as a result of stacking on A25.

(B) A25 does not pair with its predicted basemate U13, but is involved in the quadruple base interaction, and tilts between the C8 and C14 layer.

Figure 5. Stabilizing interactions at sharp turns with ions and water molecules.

(A) The " C " turn from Loop 1 (red) to C10 of Stem 2 (gold) is stabilized by a base to phosphate hydrogen bond and an organized water network (cyan spheres).

(B) An array of water molecules in the major groove of Stem 2 (gold) stabilize Loop 1 (red) on the other side. The water molecules stabilize the insertion of C8 into the major groove to form quadruple base interactions. Junctional base A25 is magenta, and Loop 2 A24 is green.

(C) A sodium ion in the minor groove mediates base to base contact between Stem 1 (gold) and Loop 2 (green). The sodium ion (orange sphere) coordinates to N3 and 2'-OH of G16, pro-R_p phosphate oxygen and more weakly to N7 of A21. A water molecule that hydrogen bonds to N4 of C22 is also coordinated. Two other water molecules mediate 2'-OH, phosphate and base contact. Dashed lines are hydrogen bonds, and solid lines are metal coordination. All diagrams except Figure 1 were generated with the program RIBBONS (Carson et al., 1991).

Figure 6. Clusters of water molecules in the pseudoknot structure

This view of the pseudoknot is looking into the minor groove of Stem 1 and major groove of Stem 2, with the inequivalent loops crossing the same side of the helix. There is an intricate network of water molecules in the major groove of Stem 2 on both sides of Loop 1, stabilizing its conformation.

Figure 7. The Stem 1-Loop 2 sharp turn involving G19.

This sharp turn is facilitated by C2'-endo sugar puckers at residue G18 and G19, the base orientation of G19 is stabilized by a self base (N3)-2'OH interaction. The turn is further stabilized by having both ends fixed, G18 is base-paired and A20 is involved in a 7 hydrogen bond network (Figure 3A).

Figure 8. Comparison of the conventional BWYV pseudoknot secondary sequence and schematic representation of crystal structure.

The most apparent differences are: U13•A25 is not base-paired as predicted, the continuous stacking of A25 into Stem 2 induces a rotation in the 3'-end strand of Stem 2 so that A25 is not stacked on G7.

Figure 9. Illustration of the difference in pseudoknot conventional prediction and crystal structure representation.

Due to the minor groove triplex interactions and the junctional core , there is a rotation between the two stems, so that they are non-coaxial.

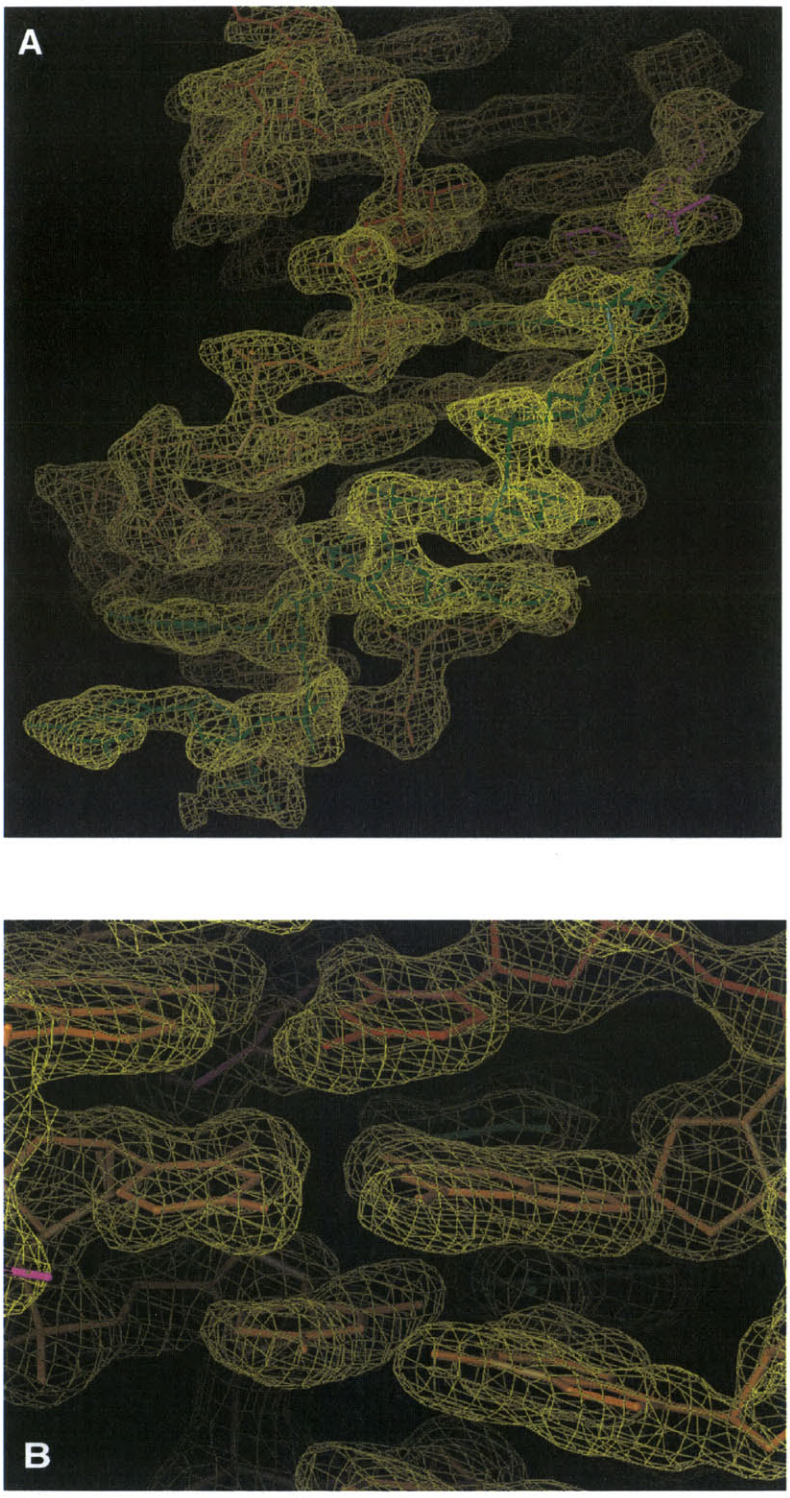


Figure 1

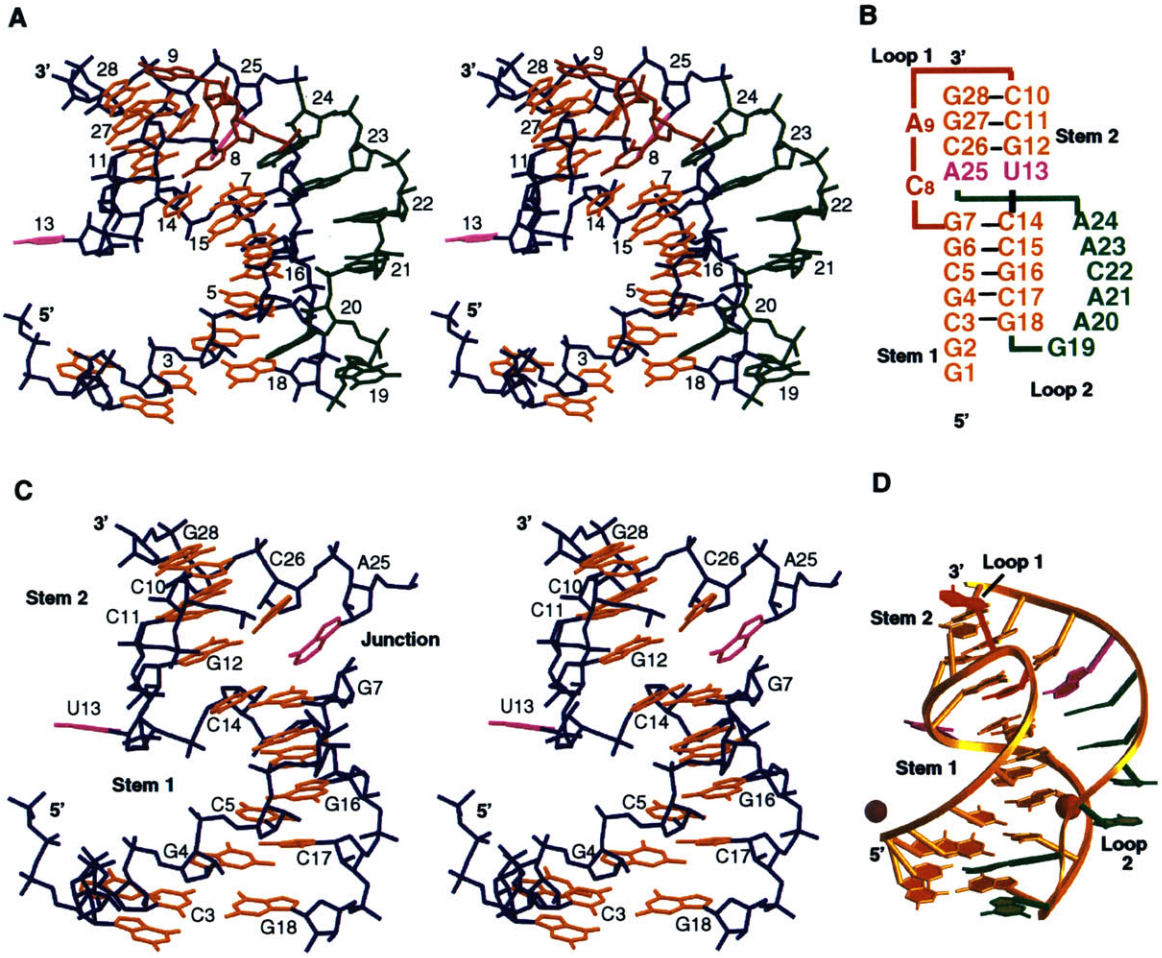

A
3'
28
9
25
24
27
23
11
8
7
22
13
14
15
21
16
5
20
5'
3
18
19
B
Loop 1
3'
G28–C10
A9 G27–C11
C26–G12
Stem 2
A25 U13
C8
G7 –C14 A24
G6 –C15 A23
C5 –G16 C22
G4 –C17 A21
C3 –G18 A20
Stem 1
G2
G19
G1
Loop 2
5'
C
3'
G28
C26
A25
Stem 2
C10
C11
Junction
G12
G7
U13
C14
Stem 1
5'
C5
G16
G4
C17
C3
G18
D
3'
Loop 1
Stem 2
Stem 1
5'
Loop 2

Figure 2

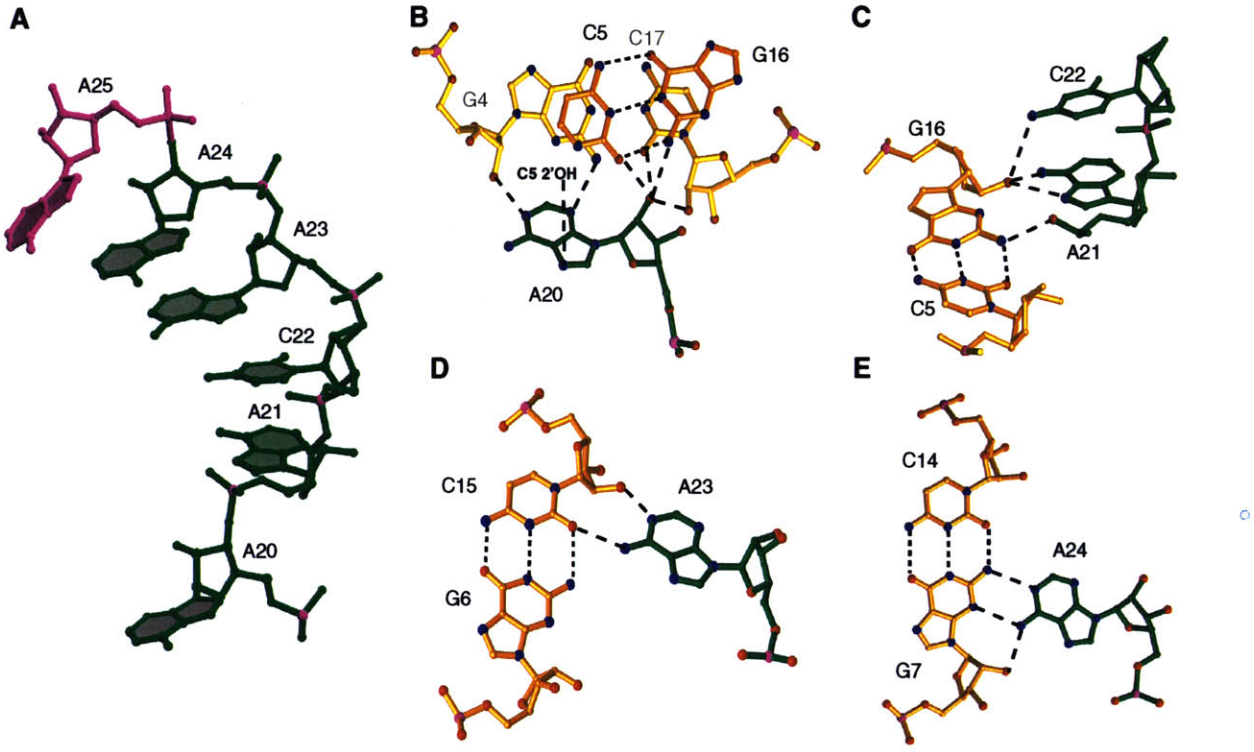

A
A25
A24
A23
C22
A21
A20
B
C5
C17
G16
G4
C5 2'OH
A20
C
C22
G16
A21
C5
D
C15
A23
G6
E
C14
A24
G7

Figure 3

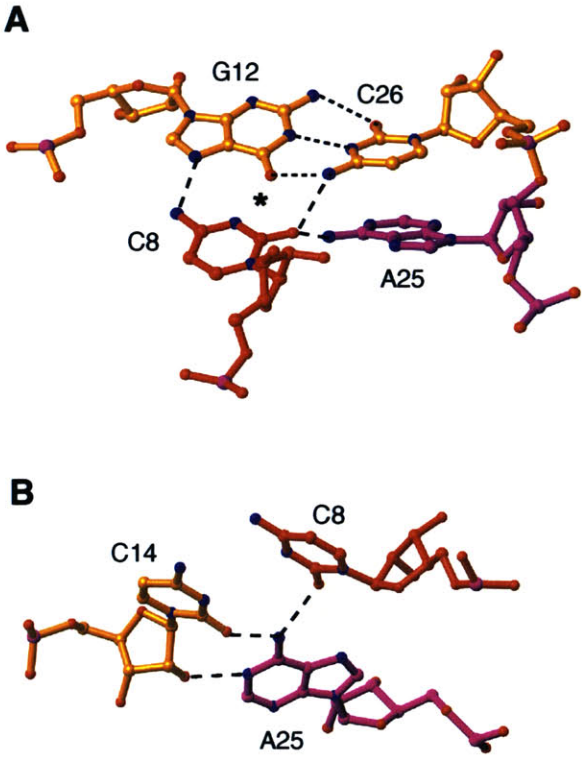


Figure 4

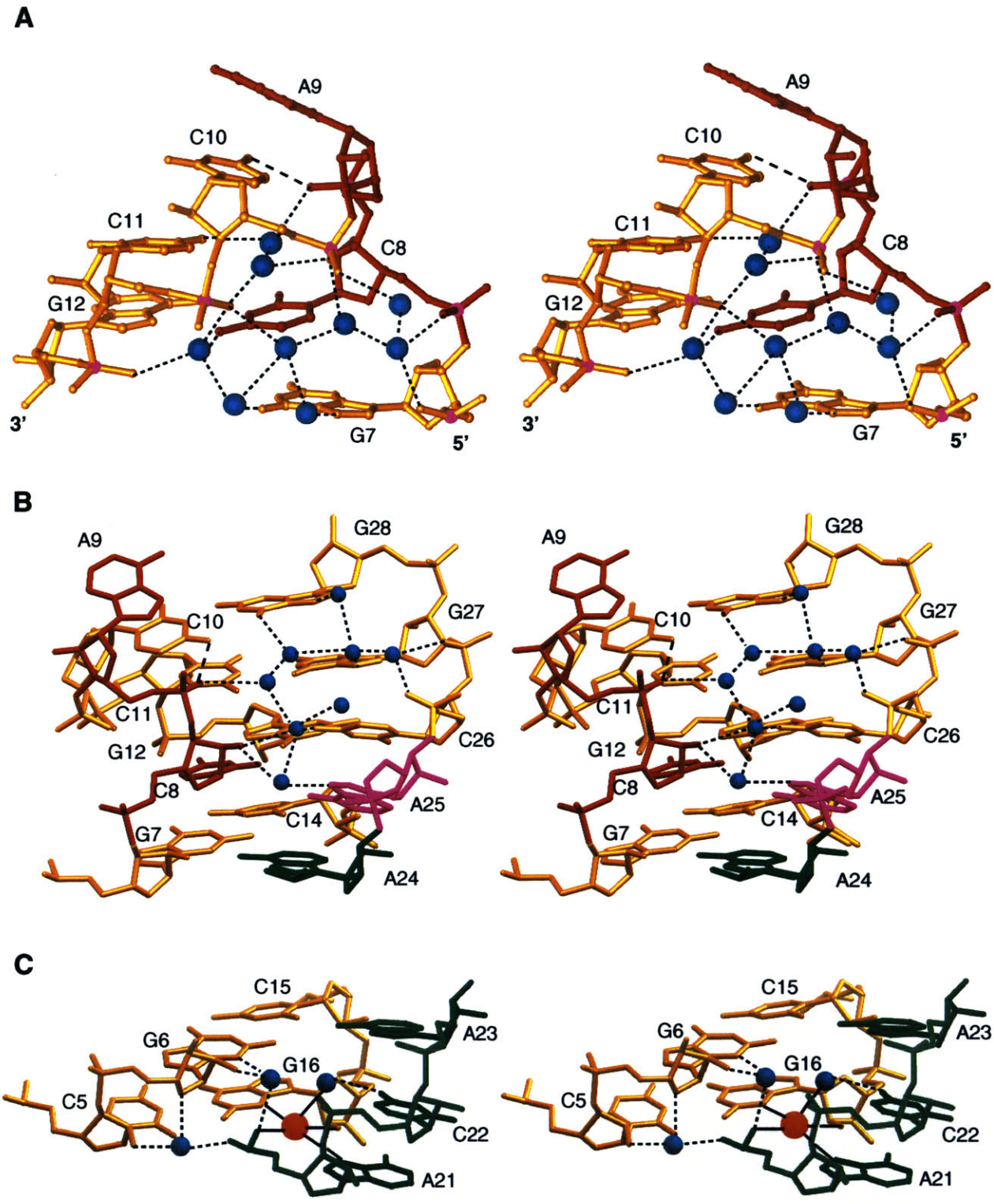

A
A9
C10
C11
C8
G12
3'
G7
5'
A9
C10
C11
C8
G12
3'
G7
5'
B
A9
G28
G27
C10
C11
G12
C26
C8
A25
C14
G7
A24
A9
G28
G27
C10
C11
G12
C26
C8
A25
C14
G7
A24
C
C15
A23
G6
G16
C5
C22
A21
C15
A23
G6
G16
C5
C22
A21

Figure 5

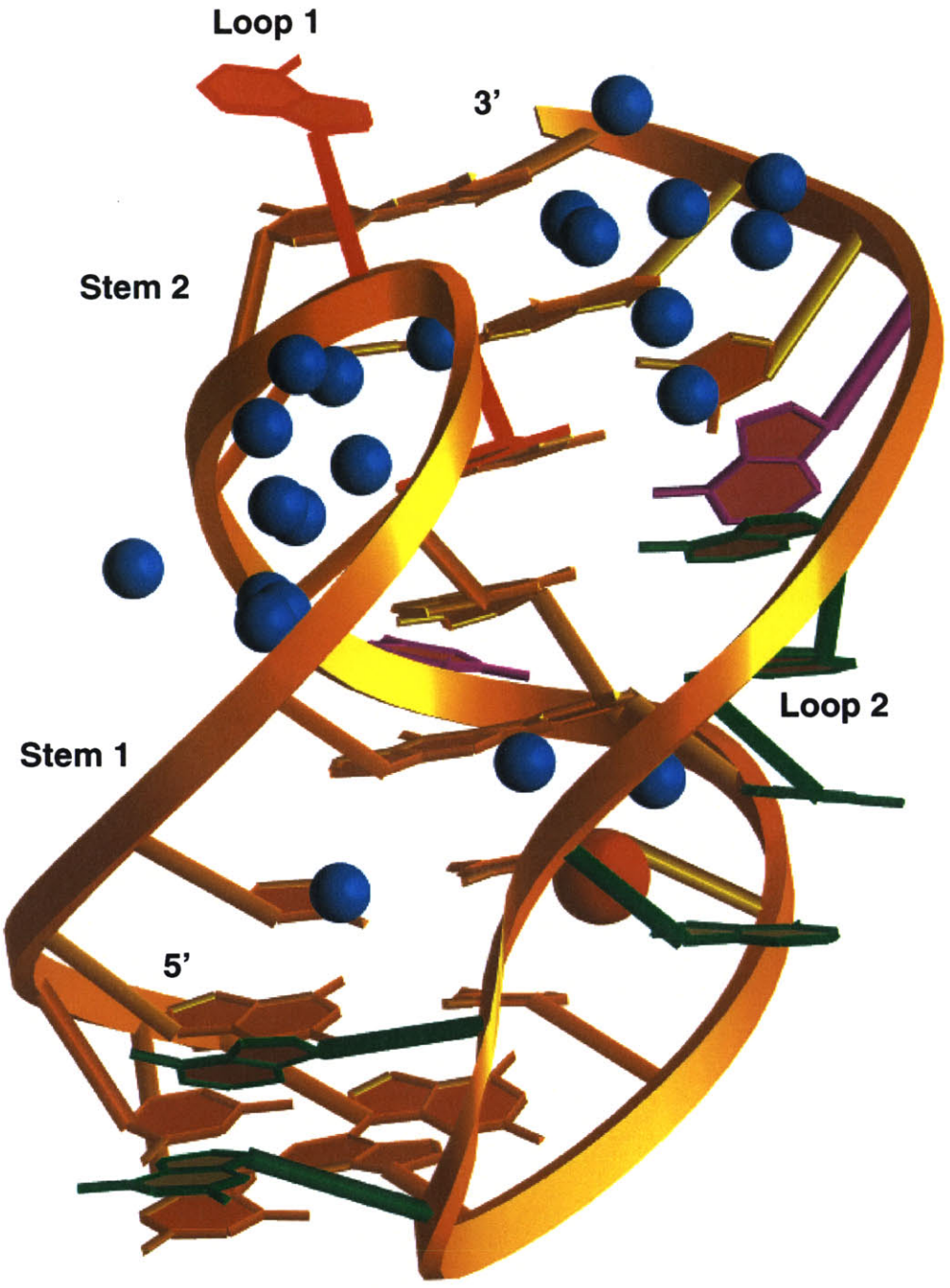


Figure 6

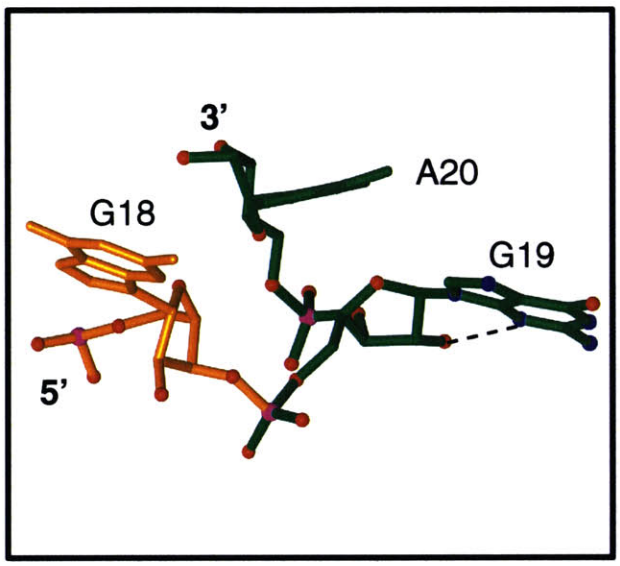


Figure 7

Secondary Structure of BWYV Pseudoknot

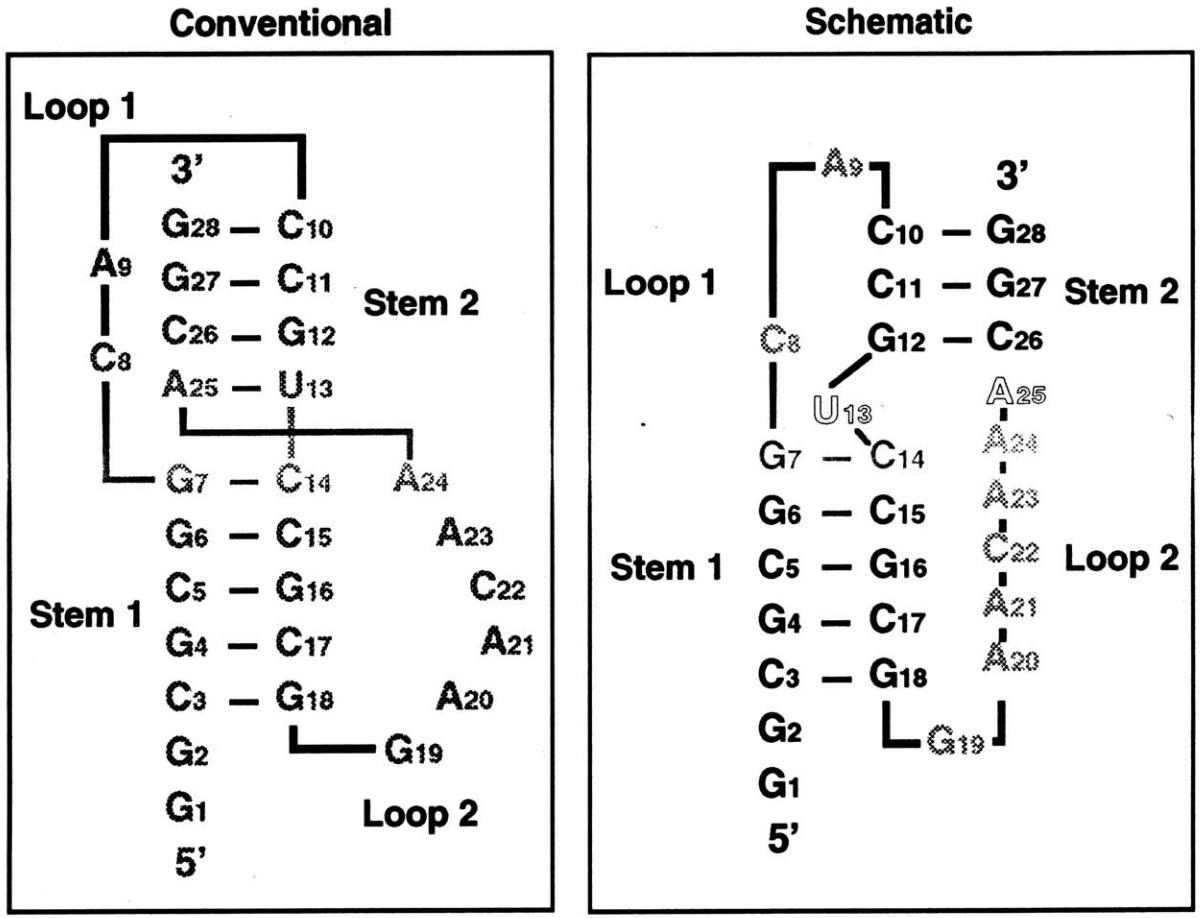


Figure 8

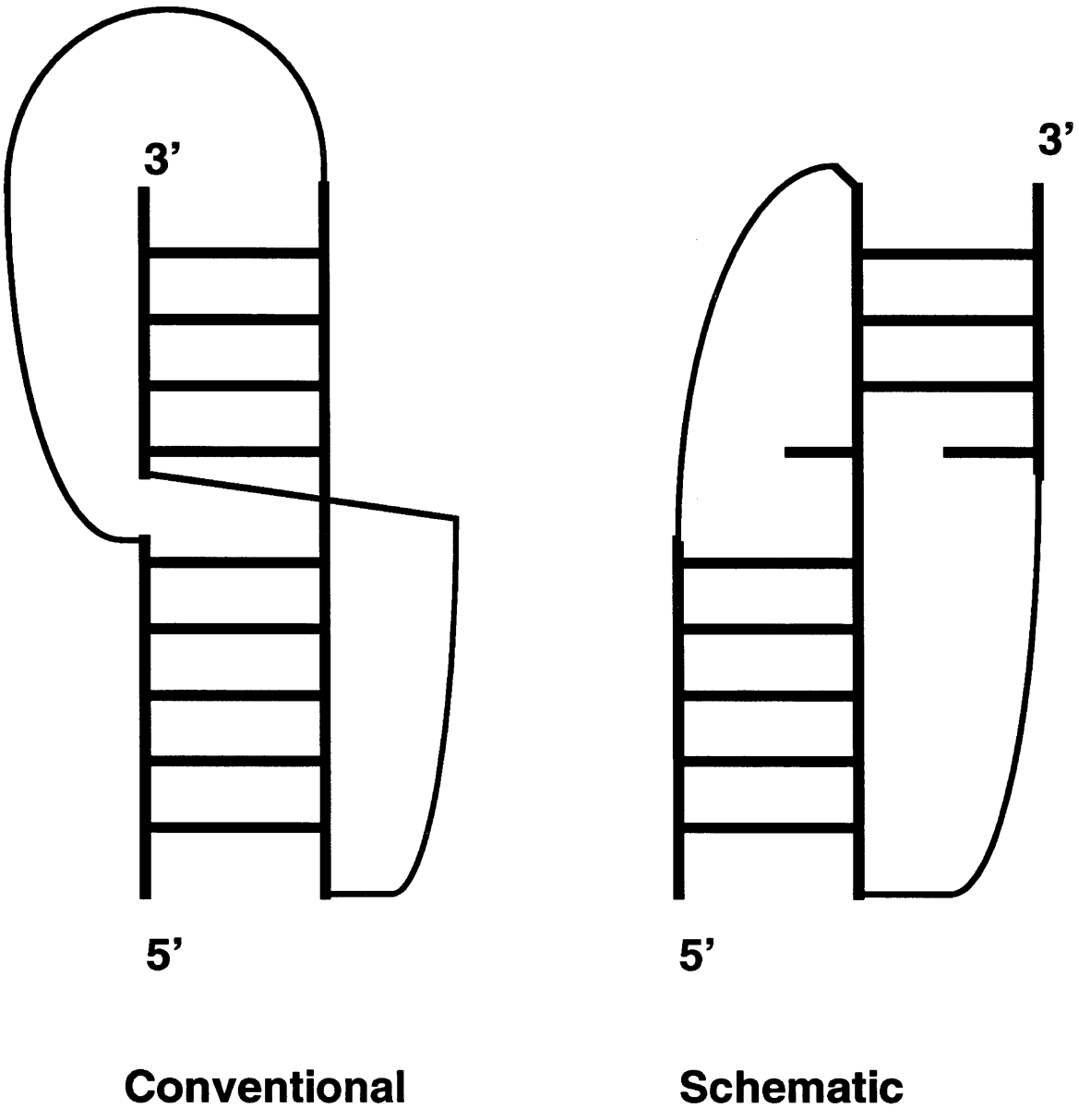


Figure 9

Chapter 4

Structural Basis for Ribosomal Frameshifting

The mutational experiments in this chapter were performed by Dr. Yang-Gyun Kim. Part of this Chapter will be submitted as a manuscript titled " The Structural Basis of Ribosomal Frameshifting for Pseudoknots ". Authors include Yang-Gyun Kim, Li Su and Alexander Rich.

Introduction

Elements in Frameshifting

The ribosomal frameshifting found in retroviruses (Rous sarcoma virus RSV, mouse mammary tumor virus MMTV, simian retrovirus –1 SRV), coronaviruses (infectious brochitis virus IBV) often have higher efficiencies, ranging from 5% to over 30% (Figure 1) (Farabaugh et al., 1996). Frameshifting efficiencies can be influenced by slippery sequence, spacer length and pseudoknot sequence. The three elements are relatively independent, and the shift site of one species can be inserted into the context of another system, although efficiencies may change (Chamorro et al., 1992). The shifty heptanucleotide sequence X XXY YYZ usually consists of a run of three adenine, uracil, or guanine residues followed by the tetranucleotide UUUA, UUUU, AAAC, or AAAU (Miller et al., 1995). This allows two out of three tRNA anticodon bases to base-pair to the mRNA after slippage (Jacks et al., 1988). It has been found that mutations which reduce the homopolymeric nature of the slip-sequence in RSV (Jacks et al., 1988), IBV (Brierley et al., 1989) and the yeast virus (Dinman et al., 1991) frequently reduce frameshifting, presumably by reducing the post-slippage mRNA-tRNA complementarity (Brierley et al., 1992). However, in certain cases only minimal post-slippage mRNA-tRNA pairing is required for efficient frameshifting (Brierley et al., 1992). In addition, certain mutations of slippery sequence can even change the nature of slippage to –2 or +1 frameshifting (Brierley et al., 1992).

Although, triplets of A, C, G, U are functional in the ribosomal P site, C-triplets are least efficient in promoting slippage (Brierley et al., 1992). In several viruses, only triplets of A and U are functional in the ribosomal A-site (Jacks et al., 1988; Dinman et al., 1991; Brierley et al., 1992), and there are only rare exceptions to the use of AAC or UUA as the YYZ codon. Since only a limited number of base triplets can function as slip sites, this raises the possibility that only certain 'shifty' tRNAs are capable in this process (Brierley, 1993). This was suggested by Chamorro et al. from mutational studies of the Z nucleotide in the pre-slippage YYZ codon (Chamorro et al., 1992). In mammalian cells, the preferred pre-slippage codon AAC (asparagine) is recognized by a single $tRNA^{Asn}$, most of which have a modified base queosine in place of G at the wobble position (Chamorro et al., 1992). In addition, $tRNA^{Lys}_{SUU}$ with a 5-modified-2-thiouridine at the wobble position facilitates –1 frameshift for the *E. coli* DNA polymerase gamma subunit (Agris et al., 1997). Brierley has proposed that tRNA anticodon hypermodification probably increases frameshifting by weakening the interaction between tRNA and its cognate codon prior to slippage (Brierley, 1993).

In addition, the pseudoknot and the slippery sequence needs to be precisely positioned, usually in the range of four to eight nucleotides, depending on the species. Insertion or deletion of a single nucleotide within the spacer region greatly reduces frameshifting (Brierley et al., 1989 and 1992; ten Dam et al., 1994). Sequence comparison of spacers show little primary sequence conservation in this region (ten Dam et al., 1994). The length requirement for the spacer has lead to the postulation

that the ribosome may require simultaneous contact of the slippery sequence and pseudoknot to achieve frameshifting (Tu et al., 1992). It has been speculated that one role of the pseudoknot is to act as a barrier in translation, perhaps pausing ribosomes over the slippery sequence and increasing the likelihood that the ribosome-bound tRNAs can realign in the –1 phase (Brierley 1993). Somogyi et al. has shown by time-pulse experiments that there was a true paused intermediate in translation and not a 'dead end product' (Somogyi et al., 1993).

Pseudoknots are often found to be the structured elements downstream the slippage site for frameshift signaling. For example, in mouse mammary tumor virus, feline immunodeficiency virus, simian retrovirus -1 and the *Saccharomyces cervisiae* double stranded RNA viruses L-A and Sc-V (Farabaugh 1996). However, in the HIV-1 system, a stem loop structure is used to regulate the *gag-pol* synthesis (Bidou et al., 1997). In systems that use a pseudoknot, the requirement is absolute, and a stem-loop containing a base-paired stem of the same length and basepair composition as the stacked stems of the pseudoknot lead to five-fold lower ribosomal pausing, but cannot replace the pseudoknot in frameshifting (Somogyi et al., 1993).

Luteovirus Pseudoknots

The genomes of the luteovirus group of plant viruses consist of a single positive sense RNA encoding five or six genes encapsidated in a spherical particle (Miller et al., 1995). Like many other RNA viruses, expression of these genes is

largely regulated at the level of translation, resulting in subgenomic mRNAs (Miller et al., 1995).

The putative RNA polymerase genes of the luteovirus is expressed via a –1 shift in the overlapping ORF1 and ORF2 (polymerase gene) reading frames (Miller et al., 1995). Thus, the stop codon which is usually downstream the structured element is readthrough and the polymerase (P2) is expressed as a fusion protein with P1 (Miller et al., 1995). A defined ratio of viral proteins are produced in the luteovirus system varying around 1%, which is significantly lower than the level observed in vertebrate viruses. The slippery site found in the luteovirus family is usually of the G GGA AAC consensus, which is a lower frameshifting slippery sequence in the luteovirus context. Changing the wild type slippery sequence in BWYV to the IBV sequence can actually boost frameshifting levels to 10.8% (Kim et al, unpublished results).

About six bases downstream of the slippery sequence is a structured region comprising a pseudoknot or stem-loop structure. From phylogenetic and mutagenic evidence, the BWYV (Garcia et al., 1993)and potato leafroll virus (PLRV) (Kujawa et al., 1993) favor formation of a pseudoknot. As in Figure 1, the secondary sequence of PLRV is nearly identical to that of BWYV, except for the elimination of the 5'-end basepair of Stem 1 and the substitution of CAA at the G19 Loop 2 position. Although, the primary secondary sequence of barley yellow dwarf virus (BYDV) has homology with those that form pseudoknots, it is still uncertain whether it forms "kissing" stem-loops or a single, larger stem-loop (Brault and

Miller, 1992). Nevertheless, the AACAAA sequence in Loop 2 and the junction region is found in all the luteovirus structured elements.

Abstract

The BWYV structure elucidates the unique features of a frameshifting pseudoknot and gives us new insight on how the conformation is organized to stabilize its folding. Pseudoknots are usually in some type of equilibrium between the folded and the unfolded form (Wyatt et al., 1990) , and it is likely that frameshifting occurs when the pseudoknot does not unravel during translocation. Therefore, features that stabilize the compact form of the pseudoknot may be important for function. It is notable in this regard that the BWYV pseudoknot is stabilized by more hydrogen bonds involved in tertiary interactions (26 total) than those engaged in Watson-Crick basepairs (24 total). The detailed knowledge arising from the crystallographic analysis has stimulated us to carry out a series of mutations on a known structure that will target each tertiary element and provide conclusive information on structure and function. The results clearly show that the minor groove triplex formed at the 5′-half of the molecule is an important feature in frameshifting; maintaining the specific quadruple base interactions and the conformation at the junction region are crucial. This implies that tertiary interactions in a frameshifting pseudoknot play an active role in maintaining stability and conformation. In addition, some mutations at the Stem 1-Loop 2 junction enhance the efficiency of frameshifting in a dramatic fashion. The studies also suggest parts of the molecule that could be involved in ribosome or unknown protein factor contacts in the frameshifting process.

Methods and Materials

Template construct for frameshifting assay

The Glutathione S-transferase (GST) gene was PCR amplified from the pGEX-5X1 (Pharmacia biotech.) vector with two primers containing the EcoRI site and BamHI site in flanking regions. The amplified GST DNA was digested with the EcoRI, BamHI restriction enzymes and gel-purified. The digested GST DNA was then ligated to a EcoRI/BamHI digested pGEM-3Z vector (Promega) containing a T7 promoter sequence upstream of the EcoRI restriction site. The resulting plasmid was named pGEM-GST. In addition, in order to place the green fluorescent protein (GFP) gene downstream of the GST gene, the GFP gene was amplified from a pEGFP-c2 (Clonetech) vector with primers containing PstI and HindIII in the flanking regions. After the PstI and HindIII digestions, the GFP DNA was ligated to the PstI/HindIII digested pGEM-GST vector. This ligated vector containing the GST and GFP genes was named pGEM-GG.

The slippery sequence and pseudoknot sequence was inserted into pGEM-GG at the BamHI and PstI restriction sites using annealed duplex DNA oligomers. This duplex DNA template contained a SpeI restriction site between the slippery sequence and pseudoknot sequence. Separate mutations in the pseudoknot sequence or the slippery sequence were introduced by inserting oligomers containing mutant sequences at SpeI/PstI for the pseudoknot and BamHI/SpeI for the slippery sequence. The DNA template construct is shown in Figure 2. All wild type and mutant constructs were confirmed by the dideoxy

DNA sequencing methods (Figure 3). In order to compare the frameshifting effects on a quantitative scale, the wild type GGGAAAC slippery sequence that yielded 3.9% frameshifting in this in vitro system was not used. A more efficient UUUAAAC slippery sequence from infectious bronchitis virus was used.

The UAG termination codon of the GST gene is immediately after the slippery sequence. If a -1 frameshift occurs at the slippery sequence, the termination codon of the GST gene is not read and further translation through the GFP gene proceeds, resulting in the production of a GST-GFP fusion protein. Corresponding changes were made in the construct after the pseudoknot sequence to compensate for mutations that involved insertions and deletions in the pseudoknot in order to maintain the correct GFP reading frame.

Frameshifting Assay

All plasmids were isolated by the alkaline lysis method using the Wizard plus DNA isolation kit (Promega). The plasmid DNA were further purified through phenol/chloroform extraction and ethanol precipitation. The lyophilized DNA were dissolved in TE buffer (Tris-Cl pH 8.0 and 1 mM EDTA). For the frameshifting assay, we used rabbit reticulocyte lysates in the *in vitro* T7 translation system coupled with transcription (Promega). 200 ng of template DNA were used in a 20 ul reaction containing 10 ul reticulocyte lysates and 0.8 ul of 10 uCi/ul ^{35}S labeled methionine. Reactions were incubated for 90 minutes at 30 °C following the manufacture's description and quenched by freezing at -20 °C.

The GST-GFP fusion product is 58 kDa and contains 14 methionine residues, the non-frameshifting GST product is 30 kDa and contains 9 methionines. In order to separate the larger fusion protein from the GST product, the samples were run on 12 % SDS polyacrylamide gels. After electrophoresis, gels were dried and exposed to phospho-imaging cassettes (Figure 4). The amounts of the two translation products were measured by the software program Phosphoimager (Molecular Dynamics). The frameshifting efficiencies were calculated with the formula (I[FS]/14) / ((I[FS]/14) + (I[NFS]/9)) , where I[FS] is the radiolabel intensity of the frameshifting product and I[UFS] is the radiolabel intensity of the non-frameshifting product. The frameshifting and non-frameshifting products were confirmed by Western blot analysis with a polyclonal anti-GST antibody (Sigma). All reaction procedures were accompanied by the unmutated translation and repeated three times to determine the average frameshifting efficiencies.

Results and Discussion

Adenosine-rich Minor Groove Triplex is an Important Feature

The exposed adenosine-rich Loop 2 region forms a minor groove triplex with Stem 1, involving triple base and 2'-OH interactions (Figure 5A). A20 has a specific base-base interaction with G14, and a 2'-OH mediated multiple sugar-base network (Figure 6A); A21 and C22 contact G16 at the 2'-OH group; A23 and A24 form specific base to base hydrogen bonds with C15 and G7 of different

strands, respectively (Figure 6B). The triplex third strand is mainly in a base stacked conformation, which continues the stacking fashion into Stem 2 through the junctional base A25 (Figure 5C) (Su et al., in press). The observed triplex strengthens the stability of Stem 1 by interacting with both strands of the stem, and is the major tertiary stabilizing factor in the pseudoknot structure. The pseudoknot stability and specific conformation involving certain conserved sequences are likely to be features that manipulate frameshifting frequency. The conservation of adenosine residues in the luteovirus Loop 2 sequence AACAA (Miller et al., 1995) and high adenosine contents found in other frameshifting pseudoknots (Chamorro et al., 1992; ten Dam et al., 1994) suggests that there might be sequence specificity in the triplex interactions. Therefore, we made a set of mutations in Loop 2 to substitute adenosines for guanosines or uridines (Figure 7A). In general, the individual and accumulative mutations to G or U show that there is a preference for adenosine residues at each position. Adenosine residues are very common in loop regions of complex RNA structure (Pley et al., 1994a and 1994b; Scott et al., 1995; Cate et al., 1996; Ortoleva-Donnelly et al., 1998). Compared to pyrimidines, adenosines can form stacking platforms and interact with the shallow minor groove, yet still have hydrogen donors and acceptors available for other contacts. Adenosines also have the advantage over the stackable guanosines, in which steric problems are caused by the additional O6 group (Pley et al., 1994b).

Substitution of adenosine residues with uridines may result in the loss of optimal stacking or disruption of triplex interactions. The replacement with G residues will result in the disruption of specific triple base interactions due to steric clash of the O6 group (A20, Figure 6A) or alterations in the hydrogen bond donor/acceptor scheme (A23, A24, Figure 6B). In comparison, the 2'-OH mediated interactions will not be affected. Other triplex interactions with different geometry may form with the substitution of U or G, but would require rearrangement in the backbone or stacking conformation. On the other hand, the mutations in Loop 2 may also interfere with possible ribosome or protein factor contacts, in which the free donor/acceptor groups of the bases may engage. This is suggested from mutations in A21 (to G21) and C22 (to U22), in which no loss of tertiary interactions are anticipated, but substitutions show a negative effect on frameshifting (Figure 6A and 7A). In contrast, the mutation C22 to A22 (9%) compared to the G22 (2.6%) substitution, has little impact on frameshifting.

The experiments that invert the C•G basepairs or substitute for U•A basepairs in Stem 1 give us complementary information on the triplex interactions (Figure 7C). The U5•A16 substitution results in the loss of two hydrogen bonds in the A20, A21 tertiary interactions, and therefore the frameshifting decreases 50%. In the C•G basepair inversions, due to the intrinsic two-fold symmetry at the guanine N2 position of a G•C basepair, inverting these basepairs in Stem 1 should not alter the hydrogen bond donor acceptor theme in the minor groove (Rosenberg et al., 1976). The inversion would

result in: no change in the guanine N2 position, but a slight variation in directionality of the hydrogen donor; the O2 group of the cytosine would replace the hydrogen acceptor at the N3 position of guanine. Therefore, these types of mutations are likely to have only minor effects on the triplex stabilizing interactions. This is evidenced by the C5•G16 and G4•C17 revertants, of which little impact on frameshifting efficiencies are detected. An interesting observation is that the frameshifting efficiency is very sensitive to mutations in stem revertants and in Loop 2 nucleotides positioned near the junction region, compared to those near the 5'-end of the molecule. The mutations at A23, A24 and their interacting stem basepairs G6•C15 and G7•C14 have detrimental effects on frameshifting, whereas similar loop mutations in A20 still retain 30% efficiency (Figure 7A and 7C). The stem revertants may lead to perturbations in adjacent base-stacking interactions in the stem region (Saenger 1984), or indirectly introduce subtle conformational changes in the interacting Loop 2 nucleotides (Figure 5C). For example, the invert mutation of G7•C14 may result in the loss of optimal intra-strand pyrimidine-purine stacking in A-form RNA between C8 and G7 (Figure 5C). In addition, the water network near these regions may undergo rearrangement. These findings imply that resistance of the pseudoknot against the deformation, driven by the translocation machinery, can be greatly affected by the junction region, or a specific conformation there is in contact with the ribosomal machinery.

Junctional Core Interactions are Crucial in Frameshifting

The other striking feature in the BWYV pseudoknot is the core interactions at the junction region between C8 of Loop 1 and the G12•C26 basepair of Stem 2, as well as A25 (Figure 6C). The Loop 1 base C8 is deeply inserted into the major groove of Stem 2 and stacks on G7 (Figure 5A). At this junctional core, the organizer base C8 simultaneously hydrogen bonds with the other three bases, forming a quadruple base interaction (Su et al., in press) (Figure 6C). This interaction brings together the nucleotides from Loop 1, Stem 2 and the Stem 1-Loop 2 minor groove triplex. Therefore, this region may determine the equilibrium between the folded and the unfolded form. The observed quadruple base interaction is sequence specific, and every base in this junctional core is essential for function. As in Figure 7B and 7C, the deletion or substitution of C8 or A25, or inversion of the G12•C26 basepair completely destroys frameshifting. All of these mutations, except C8 to U8, are expected to disrupt the observed junctional interactions in the structure due to steric clashes (C8 to A8 or G8), or unfavorable hydrogen donor acceptor contacts, and would not maintain the same number of hydrogen bonds without variations in geometry (C12•G26, A25 to G25 or U25) (Figure 6C and 7B). The severe decrease in frameshifting is unexpected for the mutation C8 to U8, which only results in the loss of one hydrogen bond to the N7 group of G12. In the structure, C8 is deeply buried in the major groove of Stem 2 and is unlikely to be accessible for higher order contact (Figure 5A). Considering the accumulative negative charges from the

surrounding phosphate groups, the positive charge on the protonated cytosine residue may play a role in electrostatics (Figure 6C). Therefore, a cytosine at position 8 is crucial for the overall pseudoknot stability.

On the other hand, the transitional adenosine 25 platform is a key element at the junction, as it is not only involved in the minor groove triplex interactions, but also continues the stacking of the triplex third strand into Stem 2 (Figure 5C and 6C). In Figure 5C, A25 is stacked between C26 of Stem 2 and A24, the last base of Loop 2. It is tilted between the C8 and C14 layer with multiple interactions (Su et al., in press). The stability of Stem 1 can be stabilized indirectly by the junctional interactions, since the cross-strand triplex interactions are likely to stabilize the Stem 1 helix. Recently, thermodynamics studies on a T4 gene 32 pseudoknot which closely resembles the frameshifting pseudoknots, has shown that sequence substitution or length increase in Loop 1 destabilizes Stem 1, with only a minor effect on Stem 2 (Theimer et al., 1998). It is possible that similar tertiary interactions are also present in the T4 pseudoknot, as the 3′ nucleotide in Loop 2 (analogous to A24 in BWYV) shows a significant increase in RNase T1 cleavage upon mutations in Loop 1 (Theimer et al., 1998). The mutation of Loop 1 may disrupt junctional interactions and translate the effect into Stem 1 through Loop 2.

Comparing the BWYV junctional sequences with other frameshifting pseudoknots: in the mouse mammary tumor virus (MMTV) (Chamorro et al., 1992) there is a U•G pair at the U13•A25 position; in simian retrovirus –1 (SRV)

(ten Dam et al., 1994), the junctional pair is inverted to A•U (Figure 1). In addition, the identity of the Loop 1 base analogous to C8 is not crucial for frameshifting (Brierley et al., 1991; Chen et al., 1995; ten Dam et al., 1995). Therefore, the junctional interactions might be quite different in those systems. Several biochemical and NMR studies suggest that these junctional basepairs are also not formed for the frameshifting function (Sung et al., 1998; Chen et al., 1995). Tinoco and co-workers had proposed that in the MMTV system, a bent conformation induced by an intercalated adenosine between the two stems may be important for frameshifting (Kang et al., 1997). In BWYV, mutations were carried out to mimick the higher efficiency frameshifting pseudoknots in attempt to improve frameshifting. A single adenosine (A13a) was inserted between the two stems before U13 and an inversion of the U13•A25 pair was performed, but both resulted in severe decrease in frameshifting efficiencies (Figure 7B). This suggests that the junction region is an integrated part of the molecule, which interactions are related to the overall geometry of the stems and neighboring sequences.

Exposed Regions not Involved in Tertiary Contact can Effect Frameshifting

There are three nucleotides in the structure which project away from the core without interaction. These include the extruded U13, the Stem 1-Loop 2 junction at G19, and A9 that caps the continuous strand of the quasi-continuous

helix (Figure 5C and 8). Mutations were also performed on these exposed regions.

The Extruded U13

In the crystal structure, U13 is not base-paired to A25 as predicted. The extrusion of U13 from the stem region widens the major groove to allow insertion of the organizer base C8 into Stem 2 (Figure 5C). In comparison to the previous mutations at the junction, the deletion or substitutions of U13 still maintains about 60% of wild-type frameshifting (Figure 7B). This is consistent with the observation from the structure that U13 is neither involved in secondary interactions with A25 nor engaged in tertiary interactions. However, U13 seems to play some kind of role in the frameshifting process. The deletion of U13 is likely to introduce tension in the backbone of the corresponding strand and result in subtle changes in the junctional interactions. Surprisingly, compared to the deletion, the substitution of U13 to A or C equally reduces the efficiency (Figure 7B). This suggest that U13 may be in contact with the ribosomal machinery or it could promote the transition from an unstructured hairpin to the pseudoknot conformation by transiently base-pairing with A25.

The A9 capping region

A9 of Loop 1 partially stacks on the cytosine 10 base in the so-called " C turn" (Figure 9). This "C turn" is stabilized by a hydrogen bond from the N4 of C10 to the phosphate oxygen of A9 in the minor groove of Stem 2, as well as an intricate network of water molecules (Su et al., in press). The substitution of this

adenine decreased frameshifting by 40-50% depending on the sequence (Figure 7B). A related observation was reported in the mutation on the helix cap at the 3′-end of the T4 gene 32 pseudoknot that affected its stability (Theimer et al., 1998). The decrease in frameshifting observed in the A9 to C9 or U9 mutation may be explained by the loss in optimal purine stacking which can stabilize the overwinding of basepairs in Stem 2 near the Loop 1 region. The G9 mutation does not alter stacking, and the efficiency decrease is unexpected. This suggests the likelihood that the pseudoknot stability is not the only factor that contributes to frameshifting efficiency. The 3′ region near A9 might be a potential site for ribosomal contact, further mutations were performed on the adjacent C10•G28 pair (Figure 7C). The combined mutations of A9 (C9/G9/U9) and the G10•C28 inversion also implicate that an adenosine sequence in this region is preferred (Figure 7C). As there are certain loop length requirements for crossing the major groove of an A-form helix (Wyatt et al., 1990), it is not surprising that deletion of A9 abolishes frameshifting (Figure 7B). In the crystal structure, the two nucleotides in Loop 1 make a considerable stretch across the major groove opening widened by the rotation and bend at the junction; the overwinding of basepairs near Loop 1 facilitate in shortening the gap (Su et al., in press). The deletion of A9 will make it difficult for C8 to cross the groove, and it is likely to cause disruption in the junctional interactions or the pseudoknot may not even form.

It is interesting to compare the results of the various mutations on the 3'-end basepairs: G28•U10 (0.4%), G10•C28(3.9%) and C10•C28 (4.3%) (Figure 7C). In the G28•U10 mutation, if a G•U mismatch basepair forms, displacement of the guanine and uridine bases towards to the major and minor groove, respectively, are required in order to form two hydrogen bonds, and this may alter helical stacking. Furthermore, the "C turn" will not be stabilized due to the loss of the amino group in a cytosine (Figure 9). This may have serious effects on the junctional core interaction organized by the adjacent C8 and may lead to the abolished frameshifting. Likewise, the G10•C28 inversion will also result in the loss of the "C turn" stabilization.

It is also interesting that the disruption of a Stem 2 basepair (C10•C28 mutation) does not entirely abolish frameshifting (Figure 7C). Although, it is possible that C10•C28 can form protonated basepairs, the general observation is consistent with previous studies, in which the stability of Stem 2 was reported to be less crucial in frameshifting than that of Stem 1 (ten Dam et al., 1995). The formation of the cross-strand minor groove triplex may be dependent on the pre-existence of Watson-Crick basepairs in the Stem 1 region. In addition, it seems that the effects of inversions on non-tertiary involved stem basepairs can also vary in different regions of the structure. The G3•C18 inversion or U3•A18 substitution almost has no effect on frameshifting, and the C27•G11 inversion that is situated near the junctional core abolishes frameshifting (Figure 7C).

Stem 1-Loop2 junction

In the BWYV conformation, Loop 2 residues are quite compact and form a stacking ladder, progressing upward to interact with each basepair layer of Stem 1 (Figure 5A). Each nucleotide contributes to the stability of the structure and seems indispensable. Adenosine residues are involved in sequence specific interactions and other residues such as G19 and C22 may serve as space linkers. A sharp turn occurs at residue G19 at the Stem 1-Loop 2 junction, and G19 is bulged out in the crystal structure (Figure 5A and 5C). In comparison to the " C turn", the G19 turn is more abrupt (Figure 5C), and not found to be stabilized by inter-nucleotide contacts or a water network (Su et al, in press). The only stabilizing element is that both ends are fixed, G18 is base-paired and A20 is involved in a 7 hydrogen bonding network (Figure 6A).

As expected, deletion of G19 decreases frameshifting by 50%, possibly disrupting the A20 interaction. In addition, the insertion of an adenosine (A19a) or pyrimidines (C19a or U19a) after G19 increased the efficiency, and the insertion of 2-4 pyrimidines gradually decreased the efficiency (Figure 7D). This implicates that while one insertion can relieve some tension at the sharp turn or provide an additional loop base to interact with the free C3•G18 basepair (Figure 5C), additional insertions may introduce flanking nucleotides that are energetically unfavorable. Similar observations have been observed in the simian retrovirus –1, in which deletion of non-conserved nucleotides in a longer Loop 2 increased frameshifting notably (ten Dam et al., 1995). Likewise, the insertion of

an adenosine (A24a) between A24 and A25 decreased the frameshifting to 3.6% (Figure 7A).

The Stem 1-Loop 2 junction nucleotide G19 is neither conserved in the luteovirus family nor found frequently in other pseudoknots. Base substitutions were also carried out on G19, and the 30% increase in frameshifting for mutation U19 was not anticipated, while A/C mutations gave wild type efficiency (Figure 7D). Strikingly, the G19 to U19C19a substitution and insertion lead to a unprecedented significant three fold (30.5%) increase in frameshifting. This phenomenon was also observed in the translation system using wheat germ abstract (data not shown). The comparison in mutation of G19C19a (13.3%), U19C19a (30.5%), and U19A19a (25.5%) conveys that a single U at position 19 can induce a dramatic effect in frameshifting (Figure 7D). This Stem 1-Loop 2 junction mutation and insertion may stabilize the pseudoknot structure, but the increase in frameshifting of this magnitude suggests that this region which is at the 5'-end of the molecule may be within ribosomal contact.

Conclusion

The pseudoknot is a unique tertiary motif that has a complicated topology and folding characteristics. In our studies, mutations that are likely to disrupt or alter the observed tertiary interactions in the BWYV crystal structure show a decrease in frameshifting efficiencies. This suggests that tertiary stabilizing factors that influence the overall stability or conformation of the

pseudoknot are important for frameshifting. As the hypothetical ribosome-associated RNA helicase attempts to unwind the pseudoknot, the 5'-end minor groove triplex, junctional core and A9 capping helps to resists this deformation.

The ribosome may first encounter the minor groove triplex that is positioned near the 5'-end of the pseudoknot, which may impede the process of unraveling by cross-strand interactions with Stem 1. This triplex feature would generally be absent in large stem-loop structures, which are known to delay ribosomal movement but are unable to induce frameshifting in pseudoknot systems (Somogyi et al., 1993; Chen et al., 1995). This adenosine-specific triplex motif seems to be a general feature in other frameshifting pseudoknots as well. Individual and accumulative substitutions for G or U nucleotides have been carried out on the Loop 2 adenosines of the SRV-1 and MMTV pseudoknots. Results also show a significant decrease in frameshifting (data not shown). This is consistent with other pseudoknot studies, in which deletion of the adenosine-rich sequence in SRV-1 (ten Dam et al., 1995) and murine leukemia virus (Wills et al., 1994) significantly decreased frameshifting or readthrough, respectively. These mutations may disrupt triplex interactions and cause the pseudoknot to unravel.

The junctional core is the most essential region that can influence frameshifting. As a general trend in the negative mutations, the drastic effect on frameshifting is measured by their proximity to the junction. The mutations near the junction region at A23, A24, A25, C8 and inversions at G12•C26, C11•G27,

G7•C14, G6•C15 basepairs all destroy frameshifting. The intricacies of the quadruple base and stacking interactions at the junction may be an important stabilizing element that organizes the scaffold of the pseudoknot structure or specific conformation. As in the mutational studies of the MMTV system, a single G nucleotide deletion (analogous to A25) in the junctional G•U pair significantly decreased frameshifting and induced the pseudoknot into a different conformation (Kang et al., 1996). The ribosomal frameshifting efficiency of the luteovirus pseudoknot is relatively low compared to those which have different junctional sequences (Brierley et al., 1991; Chamorro et al., 1992; ten Dam et al., 1994). While the stabilization in Loop 2 appears to be a general feature utilized by frameshifting pseudoknots, the junctional interactions may be different in high-efficiency frameshifting systems (Brierley et al., 1991; Chen et al., 1995; ten Dam et al., 1995). Nevertheless, it has been proposed that a certain bent conformation between the two stems may also be important for high level frameshifting (Kang et al., 1997).

As the pseudoknot barrier induces ribosomal pausing, it is likely that the ribosome contacts both the slippery site and the pseudoknot simultaneously (Tu et al., 1992). Continued activity of the mRNA translocation machinery probably induces deformation of the pseudoknot, transmitting the "tug" through Stem 1 into Loop 1. Understanding how the tertiary stabilized pseudoknot dynamically responds to deformation in this process may shed light on the –1 slippage mechanism. Although the pseudoknot stability is one element that plays a role in

this process, the resistance towards deformation by the translocation machinery from a topological point of view should also be considered. Compared to the hairpin loop structure, the pseudoknot has two junction regions where there are changes in 5'-3' directionality. This along with the intricate loop-stem interactions perhaps makes the pseudoknot topologically more difficult to unravel in a linear fashion. On the other hand, a specific conformation or certain residues may have dynamic interactions with the ribosome or protein factors to trigger the –1 slippage. The junctional core, in particular, accessible Loop 2 nucleotides, Stem 1-Loop 2 junction near the 5'-end, A9 capping region, and extruded U13 are possible sites for higher order contact.

In our studies, we also show an extreme case where minor modifications in the pseudoknot sequence at the Stem 1-Loop 2 junction can dramatically increase frameshifting efficiencies. This raises the possibility that throughout evolution, viral systems could have manipulated the translation levels of proteins through the addition of nucleotides in the pseudoknot sequence which may interact with the ribosomal machinery. Systematic experiments on other pseudoknot systems based on the results of the BWYV structure and mutations will further our understanding on the structural requirements for a high frameshifting pseudoknot.

Figure Legends

Figure 1. Secondary sequences of frameshifting pseudoknots

MMTV and SRV-1 are of the vertebrate family, there is a single A-bulge between the junction of the two stems in MMTV. In SRV-1 the junctional A•U pair is not predicted to form. BWYV and PLV are from the luteovirus family, and have the conserved AACAAA sequence in Loop 2 and the junctional base. The frameshifting efficiencies were measured *in vitro* in rabbit reticulocyte lysates with wild type pseudoknot and slippery sequences (Chamorro et al., 1992; ten Dam et al., 1994; Garcia et al., 1993; Kujawa et al., 1993) .

Figure 2. Template construct for ribosomal frameshifting.

The slippery sequence used is the higher frameshifting UUUAAAC sequence from infectious bronchitis virus. There is a stop codon immediately after the slippery sequence. Due to the requirement of restriction sites, the spacer sequence is different from the wild type. The non-frameshifting product is the GST protein, and the –1 frameshift produces the GST-GFP fusion protein.

Figure 3. Sequencing the DNA construct.

The Sanger DNA sequencing was performed to verify all constructs for the desired pseudoknot mutations. This gel shows the mutations in Loop 2, substituting the adenosines for uridines.

Figure 4. SDS-PAGE analysis of translation products.

The translation products were labeled by ^{35}S-methionine, and the 58kd frameshifting and 30kd non-frameshifting products were separated by 12% SDS-PAGE. The unmutated frameshifting efficiency is 10.8%.

Figure 5. Beet western yellow virus pseudoknot structure.

(A) BWYV pseudoknot crystal structure in stereo view. The orientation shows the minor groove side of Stem 1 and Stem 2 (blue backbone and gold bases), Loop 2 in green stacks in the minor groove of Stem 1 forming a triplex structure, mediated by 2'-OH and triple base interactions. Loop 1 in red crosses the major groove of Stem 2, and C8 is inserted to form a junctional core interaction. U13 in magenta is extruded from the helical region, and does not base-pair with A25 (magenta) as predicted.

(B) Schematic secondary structure diagram of the BWYV pseudoknot representing the organization in the crystal structure. The 5'-end G1 was added to assist transcription and G2 was in the original mRNA sequence. The two stems are non-coaxial, A25, the bottom base of Stem 2 rotates away, and does not stack on G7, the top base of Stem 1. The Loop 1 base C8 is stacked on G7 and is in the same layer with the G12•C26 basepair, A9 is partially stacked on C10. Loop 2 continuously stacks into Stem 2 through junctional base A25.

(C) Ribbons diagram of the overall fold of the pseudoknot in another view. This view has the pseudoknot rotated to the major groove side relative to Figure 2A,

showing the major groove opening between Stems 1 and 2. G19, at the Stem 1-Loop 2 junction is bulged out with no interaction. The Stem 1-Loop 2 turn is more abrupt then the Loop 1-Stem 2 "C turn" (Figure 7), and contains no inter-nucleotide hydrogen bonds at the turn for further stabilization.

Figure 6. Tertiary interactions in the BWYV pseudoknot structure.

(A) Triplex interactions in the minor groove involving A20, A21 and C22. Loop 2 is in green, Stem 1 is in gold. The adenine base of A20 contacts G4 at the 2'-OH and N2 position. The 2'-OH of A20 forms a multiple hydrogen bonding network with two layers of basepairs. A21 and C22 both contact G16 at the 2'-OH position, and the amino group of G16 makes a phosphate contact to A21. In the crystal structure, a sodium ion was found making base to base contact between G16 and A21, C22 (not shown).

(B) Triplex interaction of A23 and A24 close to the junction region. A23 and A24 use their Watson-Crick faces to interact with the bases and 2'-OH groups of G7 and C15 of different strands, respectively.

(C) Junctional core interaction involving the quadruple bases. The protonated C8 (indicated by "+") simultaneously interacts with three other bases G12, C26, A25. C8 is on the same level as the G12•C26 basepair. Thick dashed lines represent the hydrogen bonding on the top layer, and gray dashed lines represent hydrogen bonds in the lower layer. Junctional base A25 tilts between C14 and C8, and C26 propeller twists in the basepair scheme to stack on A25.

Figure 7. Mutations in the BWYV pseudoknot and the effects on frameshifting. Triangle indicates deletion, pentahedral indicates insertion (with abcd footnote), circle represents substitutions. The unmutated frameshifting efficiency in this context is 10.8%. The numbers near the mutations correspond to the frameshifting efficiency after mutation.

(A) Mutations in Loop 2.

(B) Mutations at the junction and Loop 1.

(C) Mutations and inversions in the Stem 1, Stem 2 basepairs.

(D) Mutations at the Stem 1-Loop 2 junction.

Figure 8. Van der Waals presentation of the pseudoknot structure.

(A) Minor groove view of Stem 1 (gold) with Loop 2 (green) bound. G19 (green), U13 (magenta), and A9 (red) are exposed regions in the molecule with no engagement in inter-nucleotide tertiary interactions. The 5′-end is behind the molecule. In these VDW figures, the additional 5′G in the crystal structure that was added to assist transcription is eliminated.

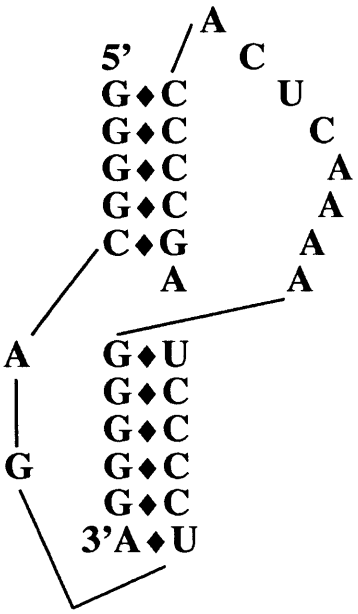
(B) Major groove view of the pseudoknot.

(C) Viewing the pseudoknot from top to bottom (3′ to 5′)

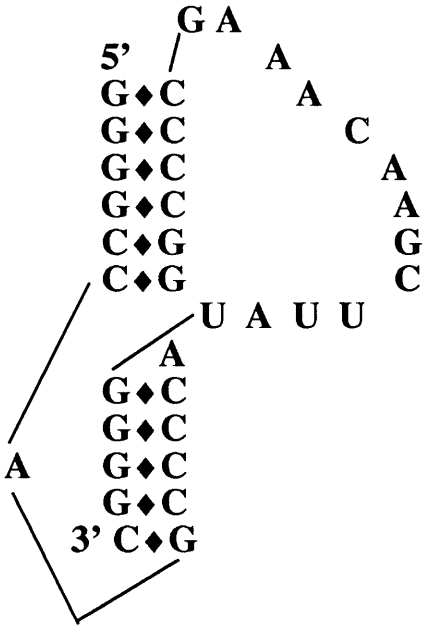
(D) Viewing the pseudoknot from bottom to top (5′ to 3′), the 3′-end is under the molecule.

Figure 9. The "C turn".

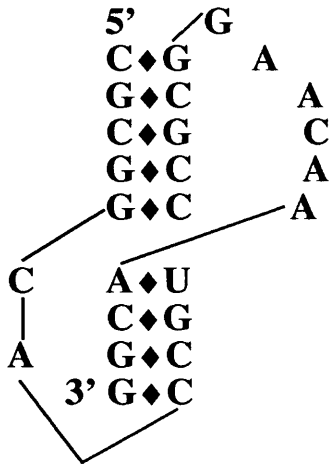
This turn involves a change in chain direction through residues C8, A9 and C10. A9 is partially stacked on C10, and N4 of C10 hydrogen bonds to the phosphate oxygen of A9. Three water molecules (cyan spheres) stabilize the turn by hydrogen bonding to the phosphates.



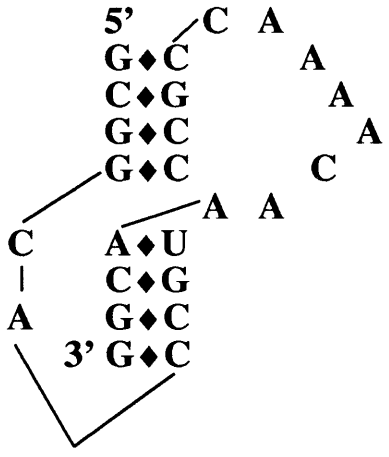
Mouse Mammary Tumor Virus
Efficiency 20%



Simian Retrovirus -1
Efficiency 23%



Beet Western Yellow Virus
Efficiency 1.4%



Potato Leafroll Virus
Efficiency 1%

Figure 1

Template Construct for Ribosomal Frameshifting

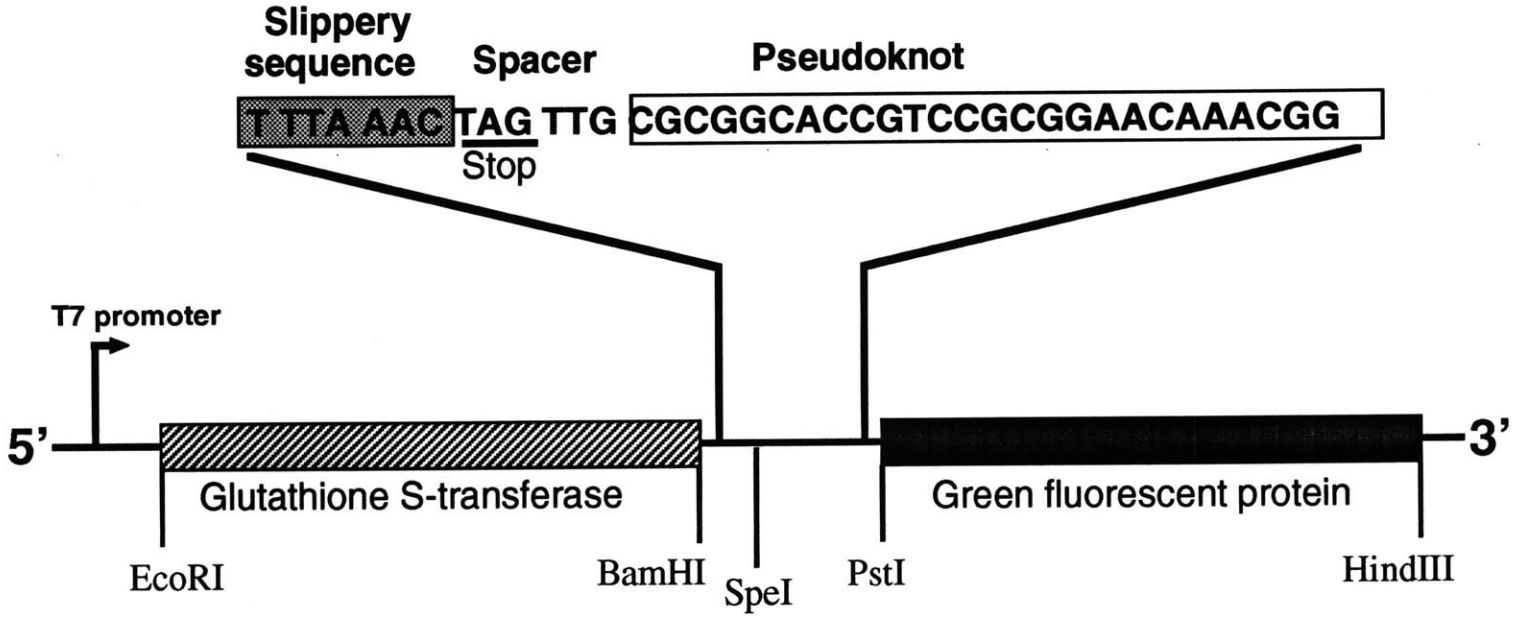


Figure 2

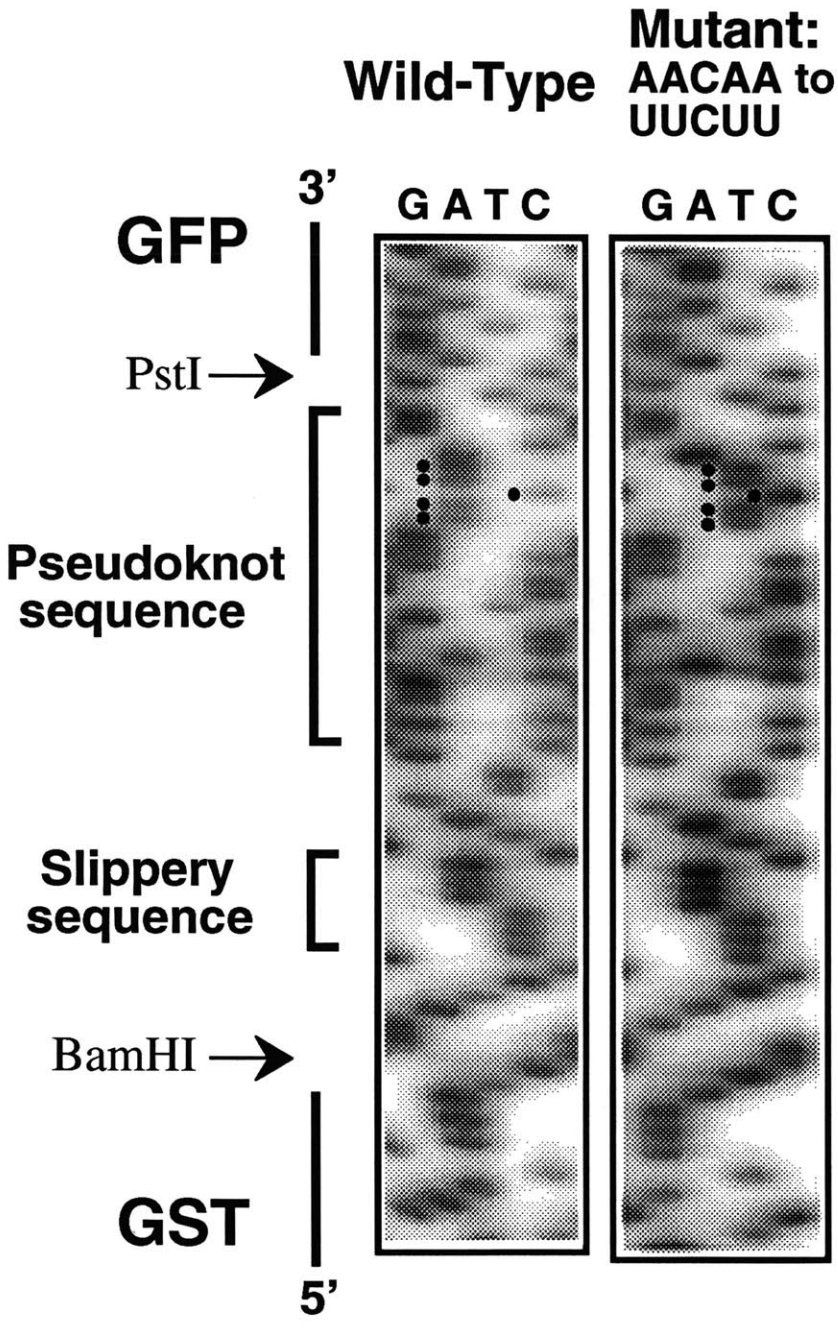
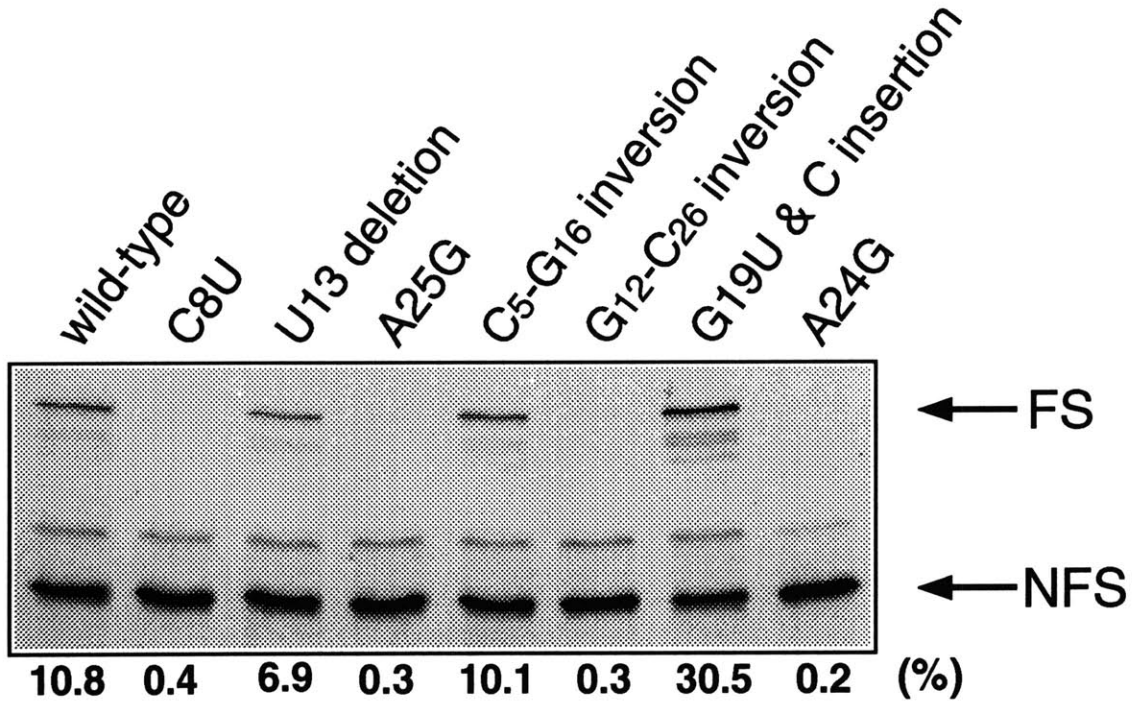

Wild-Type
Mutant:
AACAA to
UUCUU
3'
GATC
GATC
GFP
PstI
Pseudoknot
sequence
Slippery
sequence
BamHI
GST
5'

Figure 3

SDS-PAGE Analysis of ^{35}S-methionine labeled Translation Products



FS: Frameshifting product
NFS: Non-frameshifting product

Figure 4

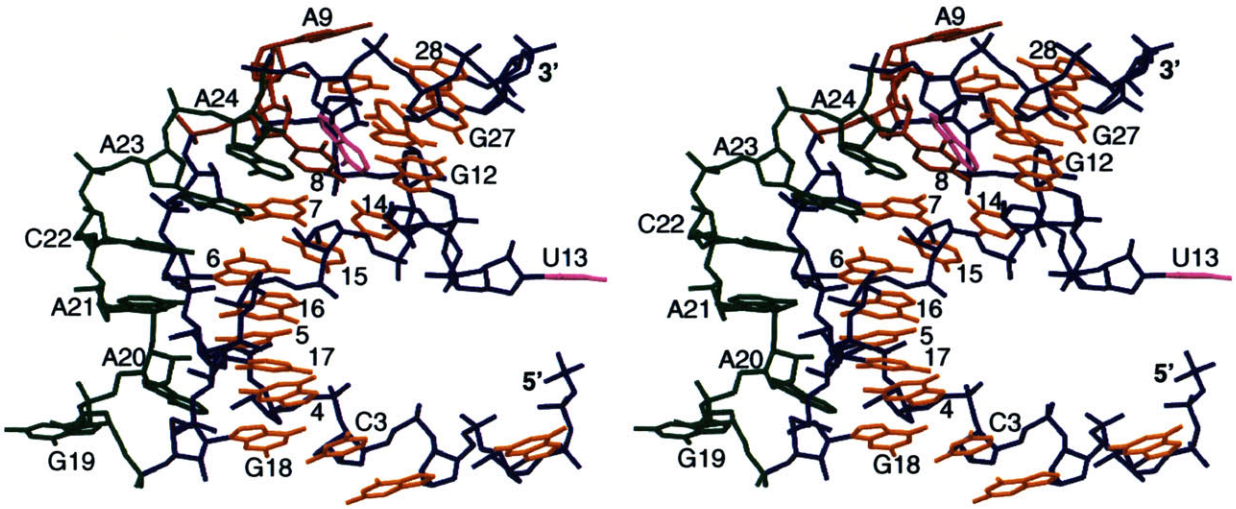

A9
28
3'
A24
A23
G27
G12
8
7
14
C22
6
15
U13
A21
16
5
A20
17
5'
4
C3
G19
G18
A9
28
3'
A24
A23
G27
G12
8
7
14
C22
6
15
U13
A21
16
5
A20
17
5'
4
C3
G19
G18

Figure 5A

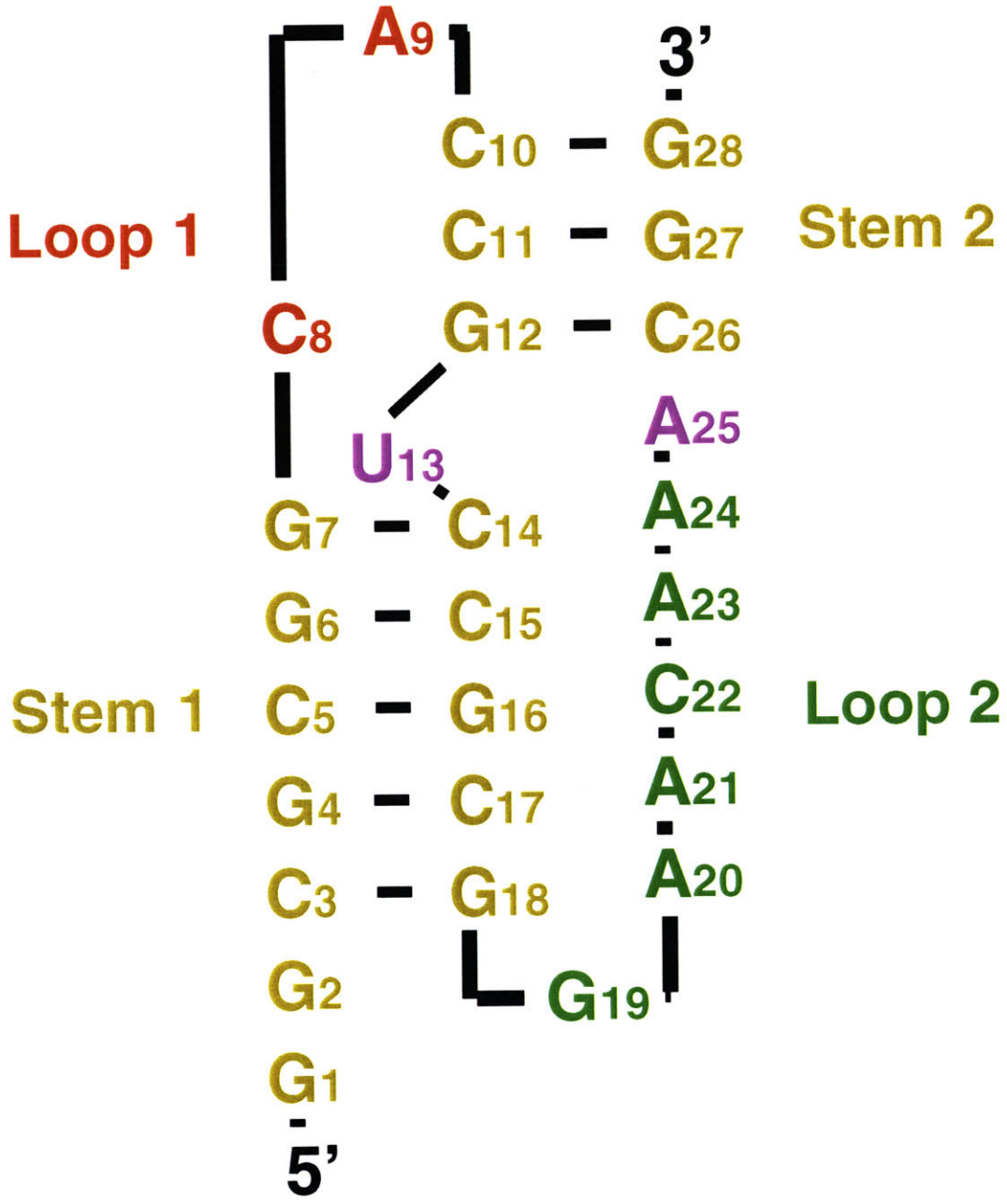

A9
3'
C10 – G28
Loop 1
C11 – G27
Stem 2
C8
G12 – C26
A25
U13
G7 – C14
A24
G6 – C15
A23
Stem 1
C5 – G16
C22
Loop 2
G4 – C17
A21
C3 – G18
A20
G2
G19
G1
5'

Figure 5B

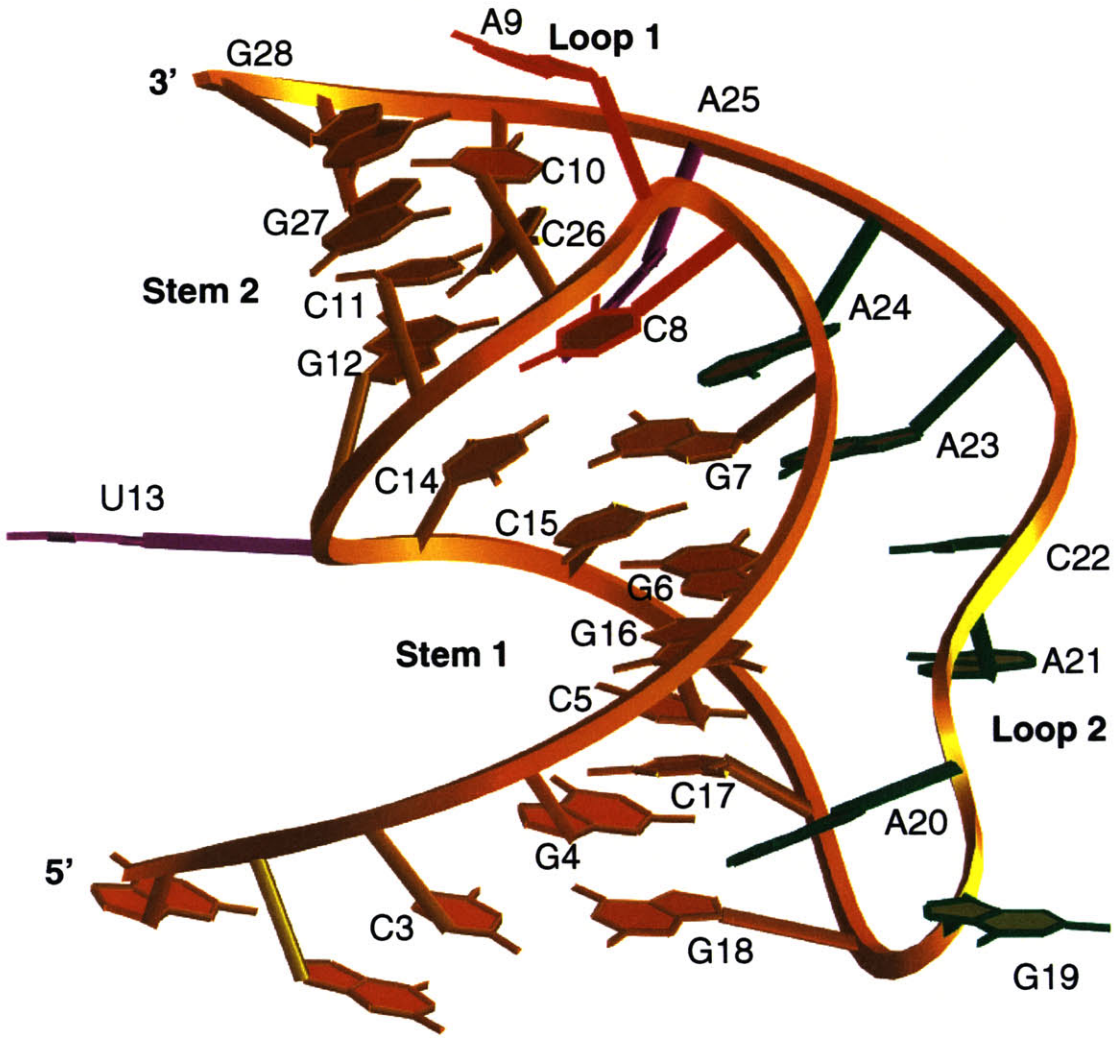

A9
Loop 1
G28
3'
A25
C10
G27
C26
Stem 2
C11
C8
A24
G12
A23
G7
C14
U13
C15
C22
G6
G16
Stem 1
A21
C5
Loop 2
C17
A20
G4
5'
C3
G18
G19

Figure 5C

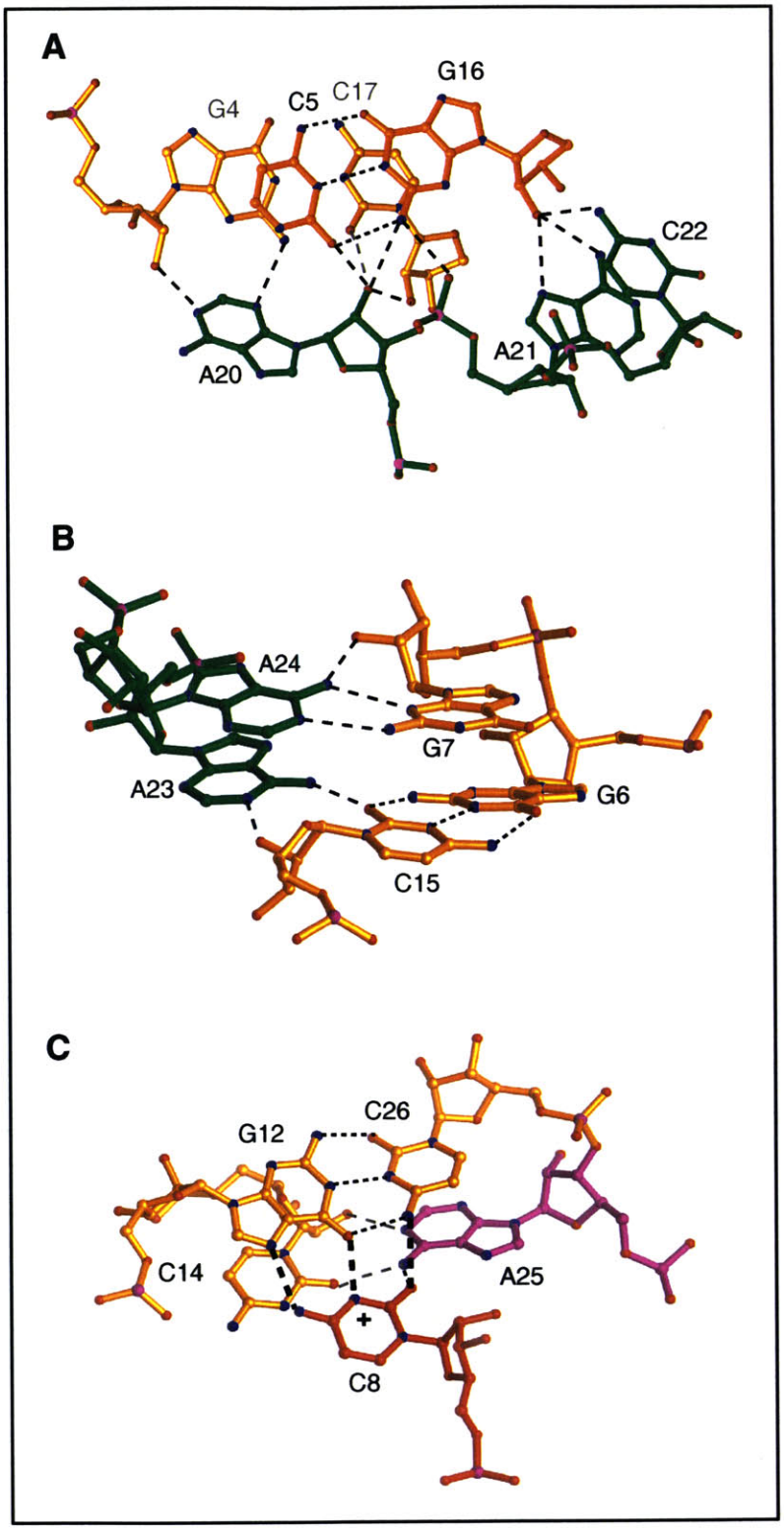


Figure 6

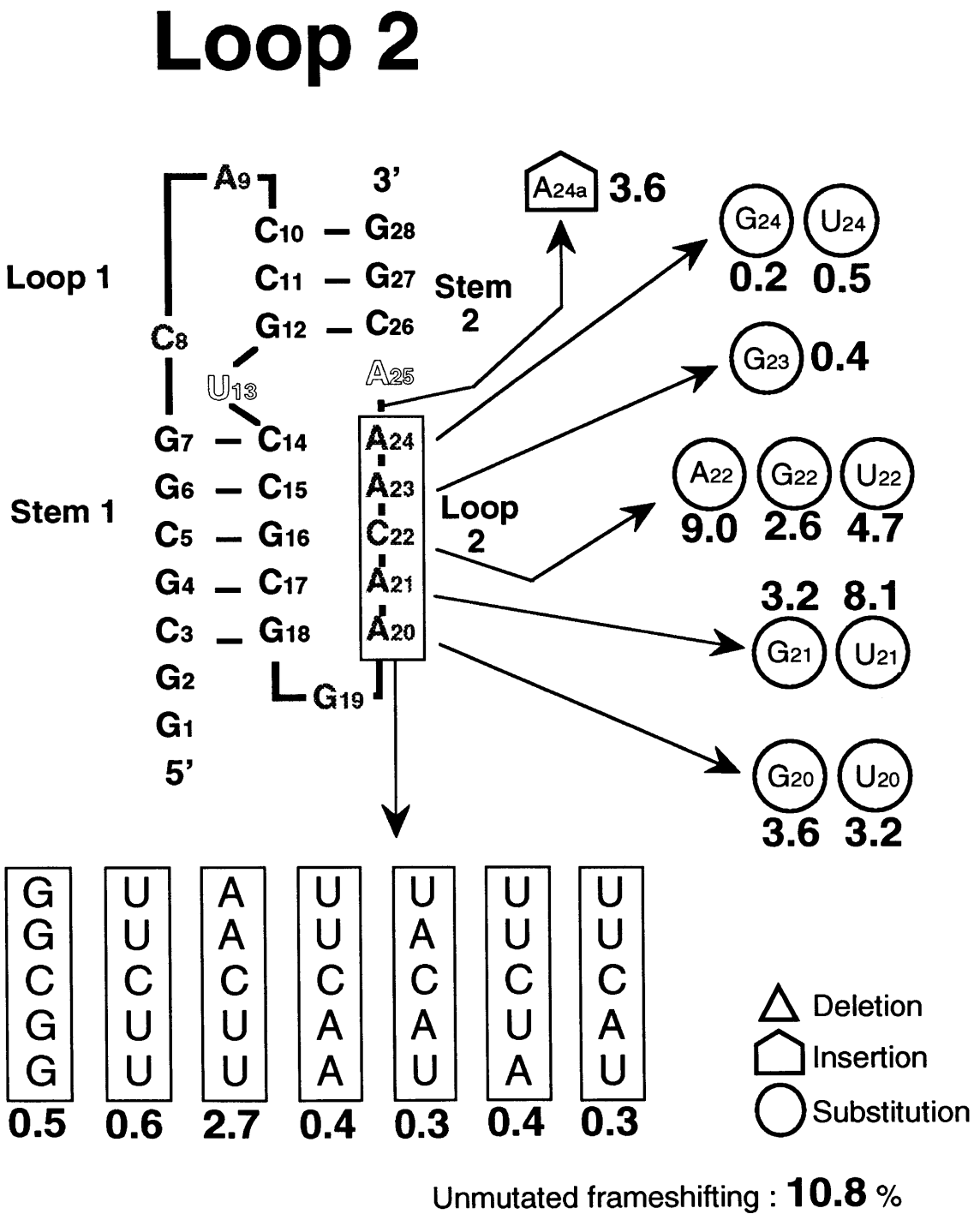

Loop 2
Loop 1
Stem 1
Stem 2
Loop 2
A9
C8
U13
G7
G6
C5
G4
C3
G2
G1
5'
3'
C10
C11
G12
G28
G27
C26
A25
C14
C15
G16
G17
G18
G19
A24
A23
C22
A21
A20
A24a 3.6
G24 U24
0.2 0.5
G23 0.4
A22 G22 U22
9.0 2.6 4.7
3.2 8.1
G21 U21
G20 U20
3.6 3.2
G G G C G G
U U C C U U
A A C C U U
U U C C A A
U A C C A U
U U C C U A
U U C C A U
0.5 0.6 2.7 0.4 0.3 0.4 0.3
Deletion
Insertion
Substitution
Unmutated frameshifting : 10.8 %

Figure 7A

Loop 1+Junction

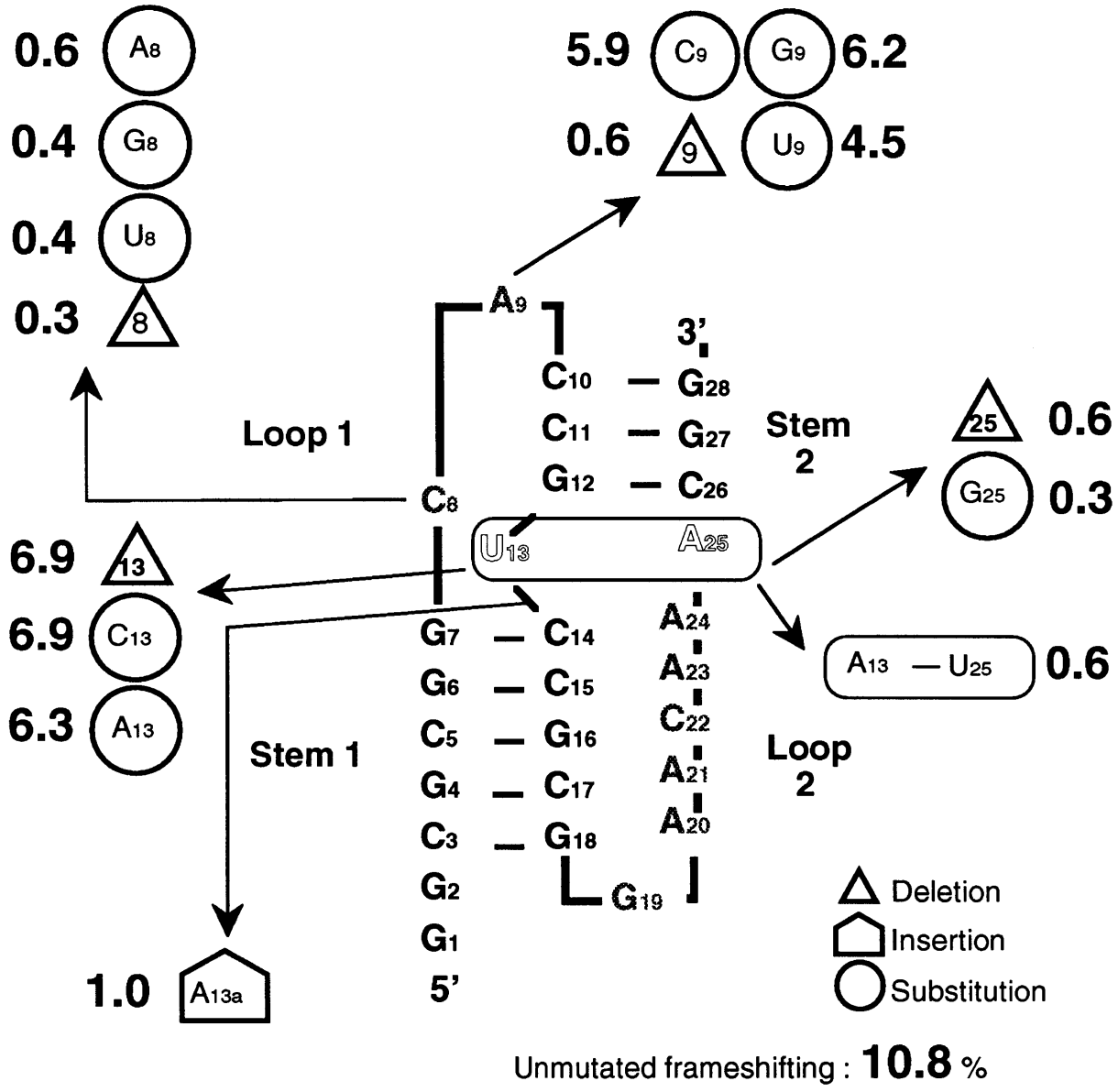


Figure 7B

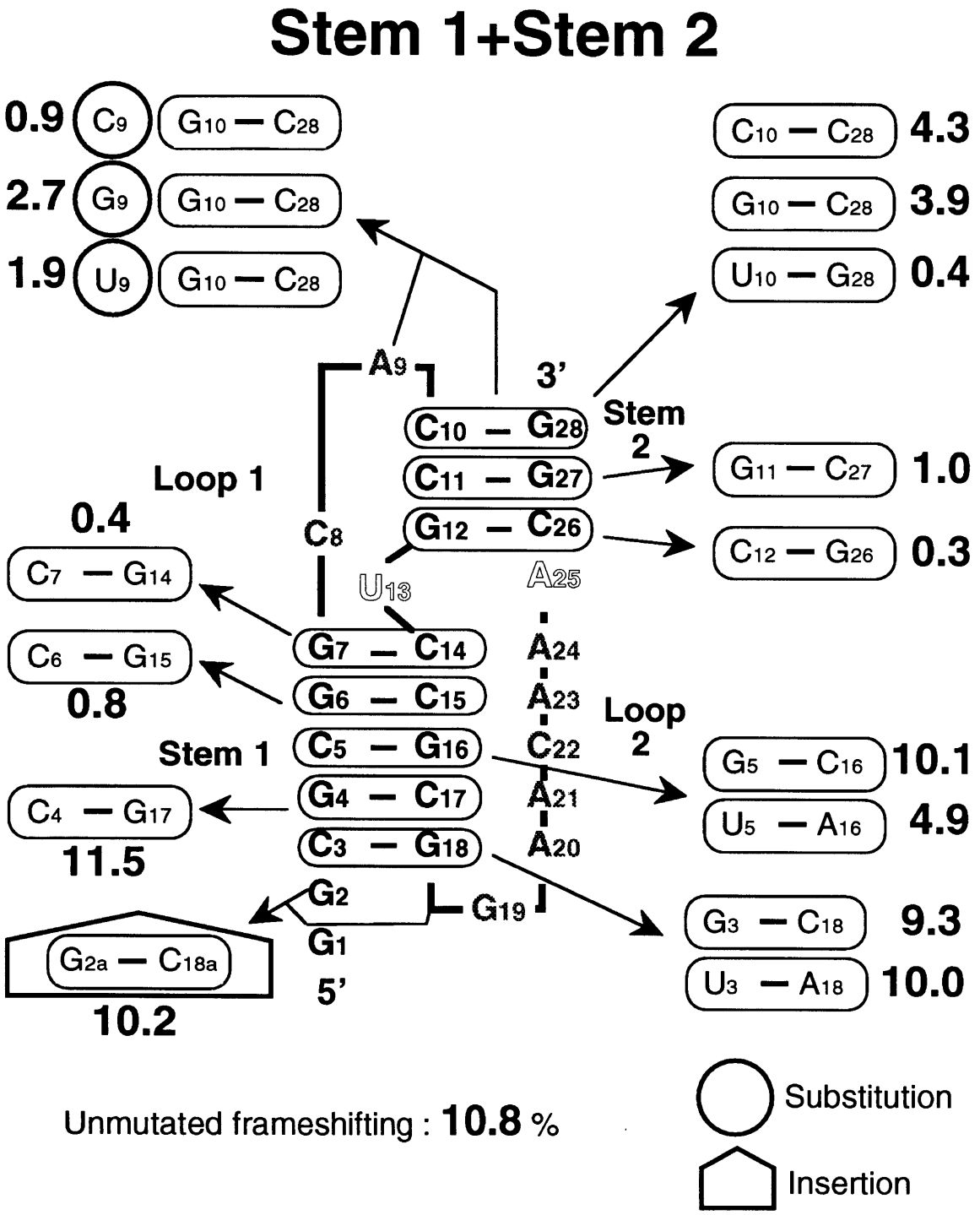

Stem 1+Stem 2
0.9 C9 G10 — C28
2.7 G9 G10 — C28
1.9 U9 G10 — C28
A9
3'
C10 — G28
C11 — G27
G12 — C26
Stem 2
C10 — C28 4.3
G10 — C28 3.9
U10 — G28 0.4
G11 — C27 1.0
C12 — G26 0.3
Loop 1
0.4
C8
U13
A25
C7 — G14
C6 — G15
0.8
G7 — C14
G6 — C15
C5 — G16
G4 — C17
C3 — G18
A24
A23
C22
A21
A20
Loop 2
Stem 1
C4 — G17
11.5
G5 — C16 10.1
U5 — A16 4.9
G2
G19
G1
5'
G2a — C18a
10.2
G3 — C18 9.3
U3 — A18 10.0
Unmutated frameshifting : 10.8 %
Substitution
Insertion

Figure 7C

Stem 1-Loop 2 Junction

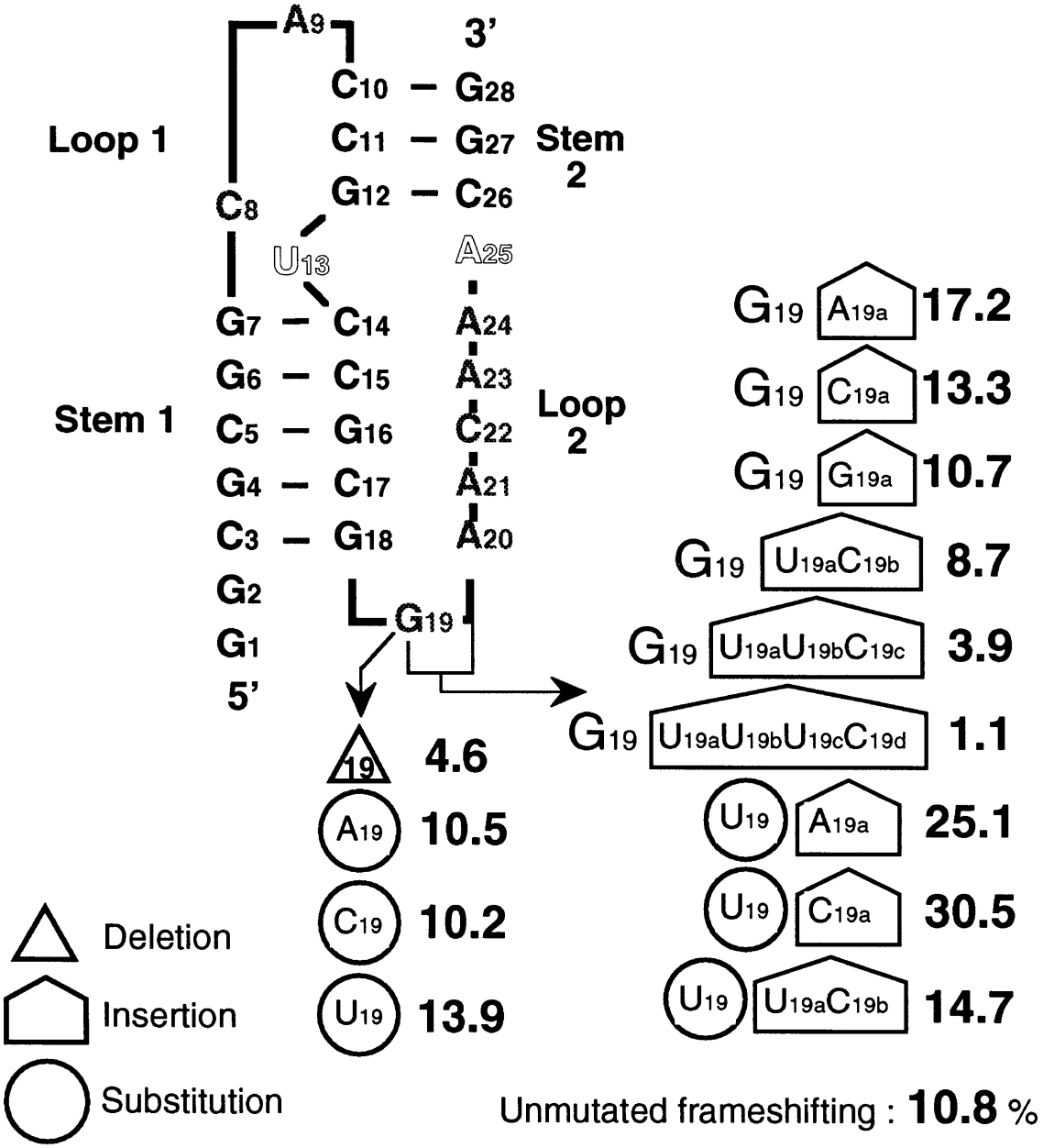


Figure 7D

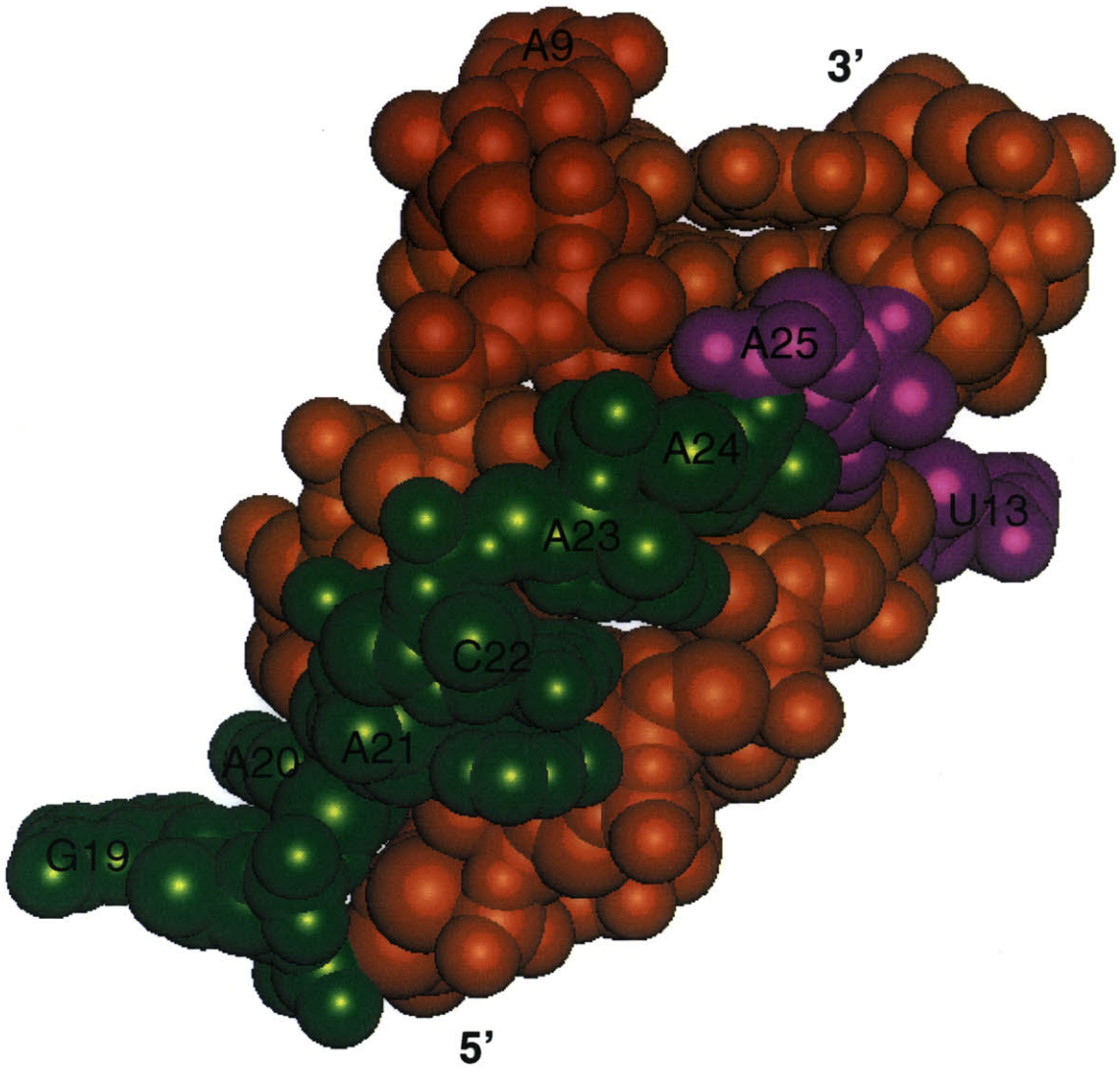


Figure 8A

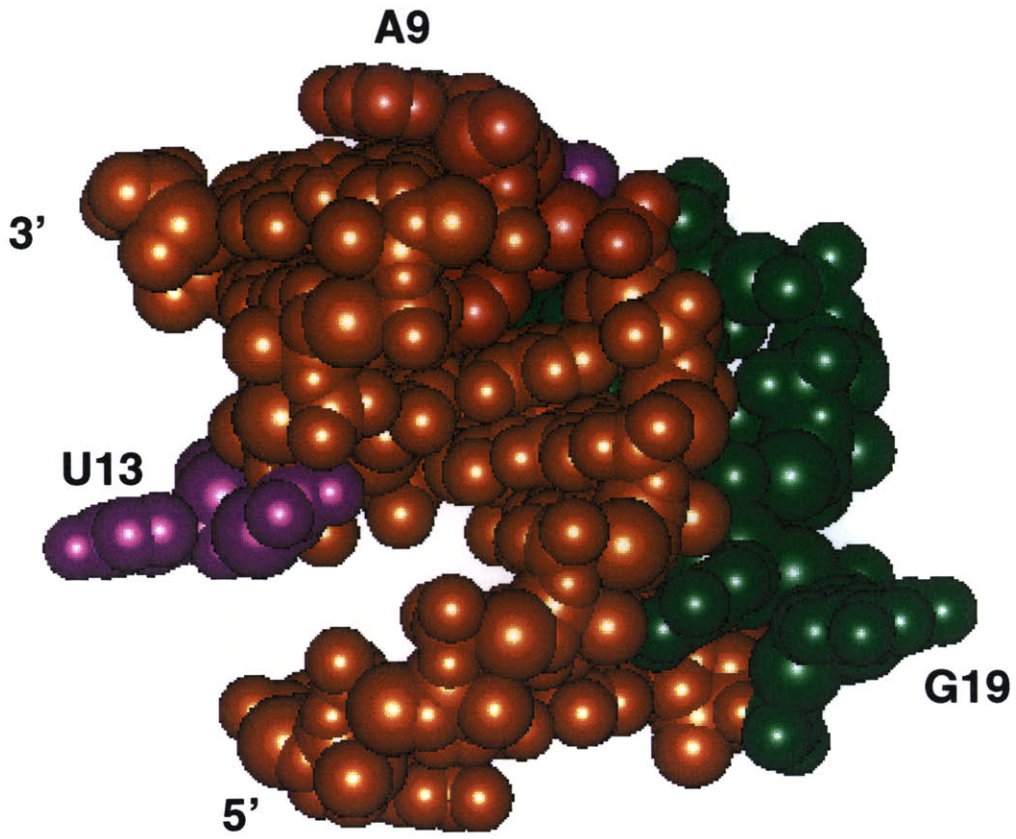


Figure 8B

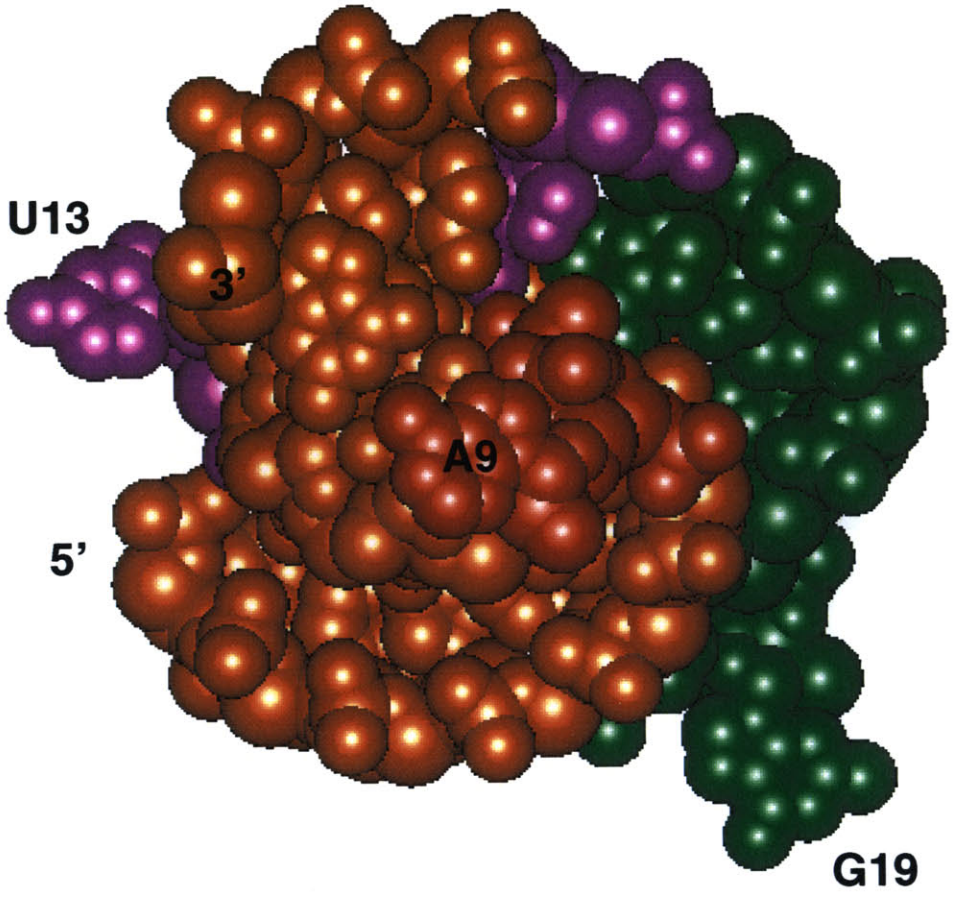


Figure 8C

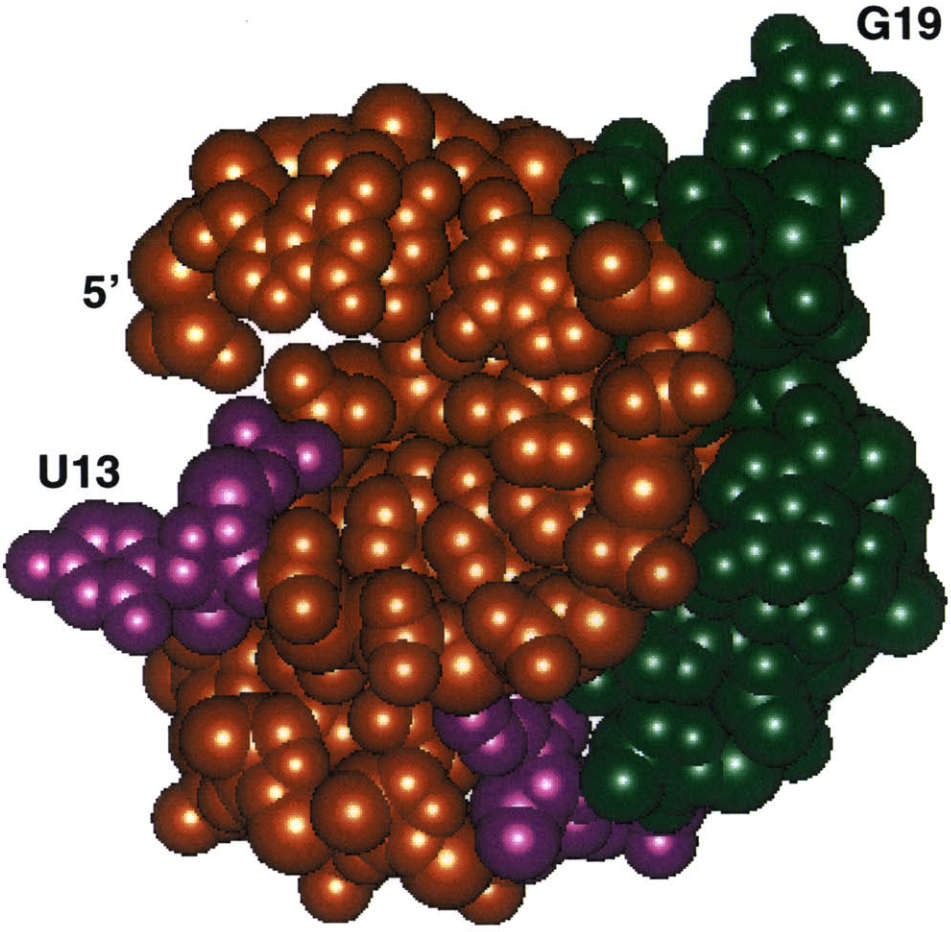


Figure 8D

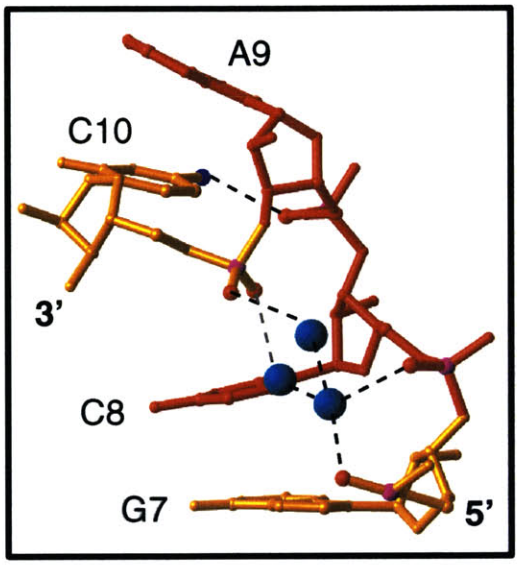


Figure 9

Future Outlooks

The BWYV pseudoknot structure at atomic resolution identifies several features that appear to be important for –1 ribosomal frameshifting, and will facilitate further analysis regarding its mechanism. Future work that focuses on the importance and nature of the ribosome-pseudoknot contact as well as the possible participation of proteins or other factors will be required to provide a full picture in this process. Suggested experiments include cross linking studies of ribosome-pseudoknot interaction, modification interference studies on Loop 2 residues to investigate whether the minor groove triplex interface is involved in higher order contact.

The pseudoknot is not likely to be a rigid entity, and may be constantly cycling between the folded and unfolded states during translation involving ribosomal frameshifting. When the ribosome encounters the slippery shift site, it pauses at the folded pseudoknot barrier; after the –1 slippage event, the pseudoknot unfolds to allow translation through the pseudoknot sequence; then, the pseudoknot perhaps refolds and awaits the next ribosome. A high frameshifting pseudoknot would perhaps require a rapid refolding process that may be facilitated by protein factors. It might be of interest to carry out systematic thermodynamic studies on frameshifting pseudoknots to see whether there is a correlation between higher frameshifters and pseudoknot stability. Another interesting question remains: Why does the pseudoknot unravel after the –1 slippage but not before slippage?

Other experiments that focus on the in vitro SELEX system for selecting high frameshifting pseudoknots can be carried out to provide complementary information on the structural and sequence requirements for efficient frameshifting.

In addition, it is interesting to compare systems that use a stem-loop structure (HIV-1) to achieve frameshifting. There has been some controversy in the field over whether HIV-1 uses a hairpin or pseudoknot (secondary sequence predictions) in the process. Further work in this field may help design drugs for antiviral therapy that interfere with frameshifting, an event critical for the viral life cycle (Dinman et al., 1998).

References

Agris, P.F., Guenther, R., Ingram, P.C., Basti, M.M., Stuart, J.W., Sochacka, E. and Malkiewicz, A. (1997). Unconventional structure of $tRNA^{Lys}_{SUU}$ anticodon explains tRNA's role in bacterial and mammalian ribosomal frameshifting and primer selection by HIV-1.

Abrahams, J.P. and Leslie, A.G.W. (1996) Acta Cryst. D52, 30-42.

Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E. Frankel, A. D., Williamson, J. R. (1996). α helix-RNA major groove recognition in an HIV-1 Rev Peptide-RRE RNA complex. Science 273, 1547-1551.

Bénard, L., Philippe, C., Ehresmann, B., Ehresmann, C. and Portier, C. (1996). Pseudoknot and translational control in the expression of the S15 ribosomal protein. Biochimie 78, 568-576.

Bidou, L., Stahl, G., Grima, B., Liu, H., Cassan, M. and Rousset J.-P. (1997). In vivo HIV-1 frameshifting efficiency is directly related to the stability of the stem-loop stimulatory signal. RNA 3, 1153-1158.

Brault, V. and Miller W.A. (1992). Translational frameshifting mediated by a viral sequence in plant cells. Proc. Natl. Acad. Sci. USA 89, 2262-226.

Brierley, I., Digard, P. and Inglis, S. C. (1989). Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell, 57, 537-547.

Brierley, I., Rolley, N.J., Jenner, A.J. and Inglis, S.C. (1991). Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. J. Mol. Biol. 220, 889-902.

Brierley, I., Jenner, A.J. and Inglis, S.C. (1992). Mutational Analysis of the "Slippery-sequence" component of a coronavirus ribosomal frameshifting signal. J. Mol. Biol. 227, 463-479.

Brierley, I. (1993). Probing the mechanism of ribosomal frameshifting on viral RNAs. Biochem. Soc. Trans. 21, 822-826.

Brown, T. and Hunter, W.N. (1997). Non-Watson-Crick base associations in DNA and RNA revealed by single crystal x-ray diffraction methods: mismatches, modified bases, and
nonduplex DNA. Biopoly. 44, 91-103.

Brünger, A.T. (1992a). The free R value: a novel statistical quality for assessing the accuracy of crystal structures. Nature 355, 472-474.

Brünger, A.T. (1992b). X-PLOR (version 3.1): A System for X-Ray crystallography and NMR (New Haven, CT: Yale University).

Carson, M. (1991). Ribbons 2.0. J. Appl. Crystallogr. 24, 958-961.

Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R. and Doudna, J.A. Crystal structure of a group I ribozyme domain: principles of RNA packing. (1996a). Science 273, 1678-1685.

Cate, J.H., Doudna, J.A. (1996b). Metal binding sites in the major groove of a large ribozyme domain. Structure 4, 1221-1229.

CCP4. (1994). Collaborative computing project number 4. The CCP4 suite: programs for protein crystallography. Acta Cryst. D50, 760-763.

Chamorro, M., Parkin, N. and Varmus, H.E. (1992). An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. Proc. Natl. Acad. Sci. USA 89, 713-717.

Chen, L., Cai, L., Zhang, X. and Rich, A. (1994). Crystal structure of a four-stranded intercalated DNA: d(C4). Biochemistry 33, 13540-13546.

Chen, X., Chamorro, M., Lee, S.I., Shen, L.X., Hines, J.V., Tinoco, I. Jr., and Varmus, H.E. (1995). Structural and functional studies of retroviral RNA pseudoknots involved in ribosomal frameshifting: nucleotides at the junction of the two stems are important for efficient ribosomal frameshifting. EMBO J. 14, 842-852.

Chen, X. Kang, H., Chen, L.X., Chamorro, M., Varmus, H.E. and Tinoco I. Jr. (1996). A characteristic bent conformation of RNA pseudoknots promotes –1 frameshifting during translation of retroviral RNA. J. Mol. Biol. 260, 479-483.

Correll, C. C., Freeborn, B., Moore, P.B. and Steitz, T.A. (1997). Metals, motifs and recognition in the crystal structure of 5s rRNA Domain. Cell 91, 705-712.

Dammel, C.S. and Noller H.F. (1993). A cold-sensitive mutation in 16S rRNA provides for helical switching in ribosome assembly. Genes Dev. 7, 660-670.

de La Fortelle, E. and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter

refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Meth. Enzymol. 276, 472-494.

Dinman, J.D., Icho, T. and Wickner, R.B. (1991). A –1 ribosomal frameshift in a double stranded RNA virus of yeast forms a *gag-pol* fusion protein. Proc. Natl. Acad. Sci. U.S.A. 88, 174-178.

Dinman, J.D., Ruiz-Echevarria, M.J. and Peltz, S.W. (1998). Translating old drugs into new treatments: ribosomal frameshifting as a target for antiviral agents. Trends. Biotechnol. 16, 190-196.

Doudna, J.A., Grosshans, C., Gooding, A. and Kundrot C.E. (1993). Crystallization of ribozymes and small RNA motifs by a sparse matrix approach. Proc. Natl. Acad. Sci. USA 90, 7829-7833.

Du, Z., Giedroc, D.P. and Hoffman, D.W. (1996). Structure of the autoregulatory pseudoknot within the gene 32 messenger RNA of bacteriophage T2 and T6: a model for a possible family of structurally related pseudoknots. Biochemistry 35, 4187-4198.

Du, Z., Holland, J.A., Hansen, M. R., Giedroc, D.P. and Hoffman, D.W. (1997). Basepairings within the RNA pseudoknot associated with simian retrovirus -1 *gag-pro* frameshift site. J. Mol. Biol. 270, 464-470.

Egli, M. and Gessner, R.V. (1995). Stereoelectronic effects of deoxyribose O4′ on DNA conformation. Proc. Natl. Acad. Sci. USA 92, 180-184.

Eisenman, G. and Horn, R. (1983). Ionic selectivity revisited: the role of kinetic and equilibrium processes in ion permeation through channels. J. Membr. Biol. 76, 197-225.

Farabaugh, P.J. (1996). Programmed translational frameshifting. Microbiol. Rev. 60, 103-134.

Felden, B., Florentz, C., Giegé, 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.

Felsenfeld, G., Davies, D.R. and Rich, A. (1957). Formation of a three-stranded polynucleotide molecule. J. Amer. Chem. Soc. 79, 2023-2024.

Garcia, A., van Duin, J. and Pleij C.W.A. (1993). Differential response to frameshift signals in eukaryotic and prokaryotic translational systems. Nucleic Acids Res. 21, 401-406.

Gehring, K., Leroy, J.-L., Guéron, M. (1993). A tetrameric DNA structure with protonated cytosine.cytosine basepairs. Nature 363, 561-565.

Gesteland, R.F. and Atkins, J.F. (1996). Recoding: dynamic reprogramming of translation. Annu. Rev. Biochem. 65, 741-768.

Gluick, T.C., Wills, N.M., Gesteland, R.F. and Draper, D.E. (1997). Folding of an mRNA pseudoknot required for stop codon readthrough: effects of mono- and divalent ions on stability. Biochemistry 36, 16173-16186.

Gunnar, R., Hansen, L.H. and Douthwaite, S. (1995). Pseudoknot in domain II of 23S rRNA is essential for ribosome function. J. Mol. Biol. 249, 59-68.

Hartman, K. A. and Rich, A. (1965). The tautomeric form of helical polyribocytidylic acids. J. Am. Chem. Soc. 87, 2033-2039.

Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science 254, 51-58.

Heus, H. and Pardi, A. (1991). Structural features that give rise to the unusual stability for RNA hairpins containing GNRA loops. Science 253, 191-194.

Jack, A., Ladner, J.E., Rhodes, D., R.S. and Klug, A. (1977). A crystallographic study of metal-binding to yeast phenylalanine transfer RNA. J. Mol. Biol. 111, 315-328.

Jacks, T., Madhani, H.D., Masiarz, F.R., and Varmus, H.E. (1988). Signals for ribosomal frameshifting in the Rous sarcoma virus *gag-pol* region. Cell 55, 447-458.

Jaeger, J., Restle T. and Steitz, T.A. (1998). The structure of HIV-1 reverse transcriptase complexed with an RNA pseudoknot inhibitor. EMBO J. 17, 4535-4542.

Jiang, F., Kumar, R. A., Jones, R. A. and Patel, D. J. (1996) Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex. Nature 382, 183-186.

Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991). Improved methods for building models in electron density maps and location of errors in these models. Acta Cryst. A47, 110-119.

Kang, H., Hines, J.V. and Tinoco I. Jr. (1996) Conformation of a non-frameshifting RNA pseudoknot from mouse mammary tumor virus. J. Mol. Biol. 259, 135-147.

Kang, H. and Tinoco, I. Jr. (1997). A mutant RNA pseudoknot that promotes ribosomal frameshifting in mouse mammary tumor virus. Nucleic Acids Res. 25, 1943-1949.

Klimasauskas, S., Kumar, S., Roberts, R.J. and Cheng, X. (1994). Hhal methyltransferase flips its target base out of the DNA helix. Cell 76, 357-369.

Kolk, M.H., van der Graaf, M., Wijmenga, S.S., Pleij, C.W.A., Heus, H.A., Hilbers, C.W. (1998). NMR structure of a classical pseudoknot: interplay of single- and double-stranded RNA. Science 280, 434-438.

Kujawa, A.B., Drugeon, G., Hulanicka, D. and Haenni A.-L. (1993). Structural requirements for efficient translational frameshifting in the synthesis of putative viral RNA-dependent RNA polymerase of potato leafroll virus. Nucleic Acids Res. 21, 2165-2171.

Lavery, R. and Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. J. Biomol. Struct. Dynam. 6, 63-91.

McPheeters, D.S., Stormo G.D. and Gold, L. (1988). Autogenous regulatory site on the bacteriophage T4 gene 32 messenger RNA. J. Mol. Biol. 201, 517-535.

Michel, F. and Westhof, E. (1990). Modeling of the three-dimensional architecture of Group I catalytic introns based on comparative sequence analysis. J. Mol. Biol. 216, 585-610.

Miller, W.A., Dinesh-Kumar, S.P. and Paul, C.P. (1995). Luteovirus gene expression. Critical Reviews in Plant Sciences 14, 179-211.

Milligan, J.F. and Uhlenbeck, O.C. (1989). Synthesis of small RNAs using T7 RNA polymerase. Meth. Enzymol. 180, 51-62.

Nunn, C.M. and Neidle, S. (1996). The high resolution crystal structure of the DNA decamer d(AGGCATGCCT). J. Mol. Biol. 256, 340-351.

Ortoleva-Donnelly, L., Szewczak, A.A., Gutell, R.R. and Strobel S.A. (1998). The chemical basis of adenosine conservation throughout the *Tetrahymena* ribozyme. RNA 4, 498-519.

Otwinowski, Z. and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Meth. Enzymol. 276, 307-326.

Perrotta A. and Been M.D. (1993). Assessment of disparate structural features in three models of the hepatitis delta virus ribozyme. Nucleic Acids Res. 21, 3959-3965.

Pann, N.S. and Read, R.J. (1996). Improved structure refinement through maximum likelihood. Acta Cryst. A52, 659-668.

Pleij, C.W.A., Rietveld, K. and Bosch, L. (1985). A new principle of RNA folding on pseudoknotting. Nucleic Acids Res. 13, 1717-1731.

Pleij, C.W.A. (1994). RNA Pseudoknots. Current Opinion in Structural Biology 4, 337-344.

Pley, H.W., Flaherty, K.M. and McKay D.B. (1994a). Three-dimensional structure of a hammerhead ribozyme. Nature 372, 68-74 .

Pley, H.W., Flaherty, K.M. and McKay, D.B. (1994b). Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. Nature 372, 111-113.

Portmann, S., Grimm, S., Workman, C., Usman, N. and Egli, M. (1996). Crystal structures of an A-form duplex with single-adenosine bulges and conformational basis for site specific RNA self-cleavage. Chem. Biol. 3, 173-184.

Powers, T. and Noller H.F. (1991). A functional pseudoknot in 16S ribosomal RNA. EMBO J 10, 2203-2214.

Puglisi, J.D., Wyatt, J.R. and Tinoco, I. Jr. (1990). Conformation of an RNA pseudoknot. J. Mol. Biol. 214, 437-453.

Quigley, G.J. and Rich, A. (1976). Structural domains of transfer RNA molecules. Science 194, 796-806.

Rajagopal, P. and Feigon, J. (1989). Triple-strand formation in the homopurine: homopyrimidine DNA oligonucleotides d(G-A)$_4$ and d(T-C)$_4$. Nature 339, 637-640.

Ramakrishnan, V. and Biou, V. (1997). Treatment of MAD data as a special case of MIR. Methods in Enzymology 276, 538-557.

Reitveld, K., van Poelgeest, R., Pleij, C.W.A., van Boom, J.H. and Bosch, L. (1982). The tRNA-like structure at the 3′ terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA. Nucleic Acids Res. 10, 1929-1946.

Rould, M. (1997). Screening for heavy-atom derivatives and obtaining accurate isomorphous differences. Methods in Enzymology 276, 461-471.

Rosenberg, J.M., Seeman, N.C., Day, R.O. and Rich, A. (1976). RNA double helical fragment at atomic resolution. II. The crystal structure of sodium guanyl-3', 5'-cytidine nonahydrate. J. Mol. Biol. 104, 145-167.

Saenger, W. (1984). RNA structure. In Principles of Nucleic acid Structure. (New York, NY: Springer-Verlag New York Inc.) pp 242-252.

Scott, W.G., Finch, J.T. and Klug, A. (1995). The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for catalytic cleavage. Cell 81, 991-1002.

Shen, L.X. and Tinoco, I.Jr. (1995). The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. J. Mol. Biol. 247, 963-978.

Somogyi, P., Jenner, A.J., Brierley, I. and Inglis, S.C. (1993). Ribosomal pausing during translation of an RNA pseudoknot. Mol. Cell. Biol. 13, 6931-6940.

Sung, D. and Kang, H. (1998). Mutational analysis of the RNA pseudoknot involved in efficient ribosomal frameshifting in simian retrovirus -1. Nucleic Acids Res. 26, 1369-1372.

Tanner, M.A. and Cech, T.R. (1997). Joining the two domains of a group I ribozyme to form the catalytic core. Science 275, 847-849.

ten Dam, E., Pleij, K. and Draper, D. (1992). Structural and functional aspects of RNA pseudoknots. Biochemistry 31, 11665-11676.

ten Dam, E., Brierley, I., Inglis, S. and Pleij, C.W.A. (1994). Identification and analysis of the pseudoknot-containing gag-pro ribosomal frameshift signal of simian retrovirus -1. Nucleic Acids Res. 22, 2304-2310.

ten Dam, E. B., Verlaan, P.W.G. and Pleij, C.W.A. (1995). Analysis of the role of the pseudoknot component in the SRV-1 gag-pro ribosomal frameshift signal: Loop lengths and stability of the stem regions. RNA 1, 146-154.

Theimer, C.A., Wang, Y., Hoffman, D.W., Krisch, H.M. and Giedroc, D. P. (1998). Non-nearest neighbor effects on the thermodynamics of unfolding of a model mRNA pseudoknot. J. Mol. Biol. 279, 545-564.

Tu, C., Tzeng, T.-H. and Bruenn, J.A. (1992). Ribosomal movement impeded at a pseudoknot required for frameshifting. Proc. Natl. Acad. Sci. USA 89, 8636-8640.

Turner, D.H., Sugimoto, N. and Freier, S.M. (1988). RNA structure prediction. Annu. Rev. Biophys. Biophys. Chem. 17, 167-192.

Tzeng T.-H., Tu, C.-L. and Bruenn J.A. (1992). Ribosomal frameshifting requires a pseudoknot in the *Saccharomyces cerevisiae* double-stranded RNA virus. J. Virol. 66, 999-1006.

Weiss R.B. (1991). Ribosomal frameshifting, jumping and readthrough. Curr. Opinion. in Cell Biol. 3, 1051-1055.

Wills, N.M., Gesteland, R.F. and Atkins, J.F. (1994). Pseudoknot-dependent read-through of retroviral gag termination codons: importance of sequences in the spacer and loop 2. EMBO J. 13, 4134-4144.

Wyatt, J.R., Puglisi, J.D. and Tinoco, I. Jr. (1990). RNA pseudoknots. Stability and loop size requirements. J. Mol. Biol. 214, 455-470.

Yang, Y., Kochoyan, M., Burgstaller, P. Westhof, E., Famulok, M. (1996). Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. Science 272, 1343-1346.

Biographical Note

Education

09/92-09/98 **Ph.D. Massachusetts Institute of Technology** Cambridge MA

Biochemistry Division, Department of Chemistry.
Thesis " Crystal Structure and Functional Studies of a Viral RNA Pseudoknot Involved in Ribosomal Frameshifting "
with Professor Alexander Rich in the Department of Biology.

09/88-07/92 **B. S. Beijing University** Beijing, China
Structural Chemistry, Department of Chemistry.
Undergraduate thesis on structural analysis of organic compounds.

Professional Activities

05/98 **Third Annual Meeting of the RNA Society** Madison, WI
Oral Presentation " Crystal structure of an RNA pseudoknot involved in ribosomal frameshifting " in the Platform session of RNA structure.

06/97 **EMBO workshop** Grenoble, France
"MAD data collection and analysis"

08/96 **International Summer School** Spetsai, Greece
" Molecular Mechanisms of Signaling and Targeting "

03/96 **Conference series Instituto Juan March** Madrid, Spain
" From Transcript to Protein "
Poster Presentation " RNA pseudoknot at 1.9 Å resolution "

Experience

1994-1996 **Undergraduate Research Supervisor, MIT** Cambridge, MA

1992-1993 **Teaching Assistant, MIT** Cambridge, MA

1989-1992 **Translation Coordinator, ABC News** Beijing, China

Publications

Su, L., Chen, L., Egli, M., Berger, J. & Rich, A. (1998). Minor groove RNA triplex in the crystal structure of a viral pseudoknot involved in ribosomal frameshifting. Submitted.

Kim, Y.G., **Su., L.** and Rich, A. (1998). Structural basis for ribosomal frameshifting. Manuscript in preparation.

Berger, I., **Su, L.**, Spitzner, J.R., Kang, C.H., Burke, T. & Rich, A. (1995). Molecular structure of the halogenated anti-cancer drug iododoxorubicin complexed with d(TGTACA) and d (CGATCG) Nucleic Acids Res. 23, 4488-4494.