Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot

Li Su[1], Liqing Chen[1,2], Martin Egli[3], James M. Berger[4] and Alexander Rich[1]

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 (BWYV) 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 exceeds the number involved in Watson–Crick base pairs. This structure will aid mechanistic analyses of ribosomal frameshifting.

In many viruses, including both tumor viruses and retroviruses, the programmed -1 ribosomal frameshifting of polycistronic mRNA regulates the relative level of structural and enzymatic proteins important for efficient viral assembly[1,2]. Because of the -1 shift in reading frames, the stop codon is avoided, and a single fusion protein is produced. 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[1]. The -1 frameshifting is found not only in retroviruses[3–5] but also in coronaviruses[6], in yeast[7] and in plant viruses[8], as well as in bacterial systems[1]. 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 six to eight nucleotides downstream from the 'slippery sequence', 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[3]. 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 base pairs with the anticodons upon slippage[3].

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[9]. The pseudoknot conformation originates from a hairpin structure, in which a single-stranded region folds back to pair with residues in the hairpin-loop region. The stem at the 5' end is named stem 1, and the 3' stem is stem 2. Loop 1 and loop 2 are unpaired regions that cross the major groove of stem 2 and the minor groove of stem 1, respectively[10]. Many mutational studies have been performed on frameshifting pseudoknots to correlate structure and function[4–7], and certain mutations at the junction of the two stems or loop regions can have deleterious effects on frameshifting efficiencies[11,12]. These findings suggest that there are special structural features involving tertiary interactions in a frameshifting pseudoknot that are important for function.

Several NMR studies of pseudoknots have been reported during the past few years[13–18]. Those pseudoknots active in ribosomal frameshifting revealed that the two base-paired stems are stacked upon each other[14,16,17], and are frequently in a bent conformation[14,17]. In addition, loop nucleotides may be restrained by base-stacking interactions[14]. However, the junction of the two stems in frameshifting pseudoknots is not well defined, and very little is known about loop–stem interactions. Here we present a 1.6 Å resolution crystal structure of a 28-nucleotide pseudoknot from beet western yellow virus (BWYV), a plant luteovirus that is active in ribosomal frameshifting and regulates the production of an RNA-dependent RNA polymerase[8]. The structure reveals an RNA triplex formed by loop 2 in the minor groove of stem 1, and quadruple interactions between a loop 1 base and the major groove of stem 2. In addition, there is a marked rotation and bending at the junction of the two stems.

Overview of the structure

The crystallographic analysis of the structure is summarized in Methods. Data have been collected to 1.6 Å resolution and the structure refined to an R-factor of 20.7%, free R-factor of 25.4% (Table 1). Experimental electron density maps are also included (Fig. 1*a*,*b*). The overall structure is very compact (Fig. 2*a*), with van der Waals dimensions of approximately 32 Å × 36 Å × 22 Å. The first dimension is in the 5' to 3' direc-

[1]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA. [2]Current address: Laboratory for Structural Biology, Department of Chemistry and Materials Science, University of Alabama, Huntsville, Alabama 35899 USA. [3]Department of Molecular Pharmacology & Biological Chemistry and the Drug Discovery Program, Northwestern University Medical School, Chicago, Illinois 60611 USA. [4]Whitehead Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA. Current address: Department of Molecular and Cell Biology, University of California, Berkeley, California 94720 USA.

Correspondence should be addressed to A.R. *fax: 617-253-8699*

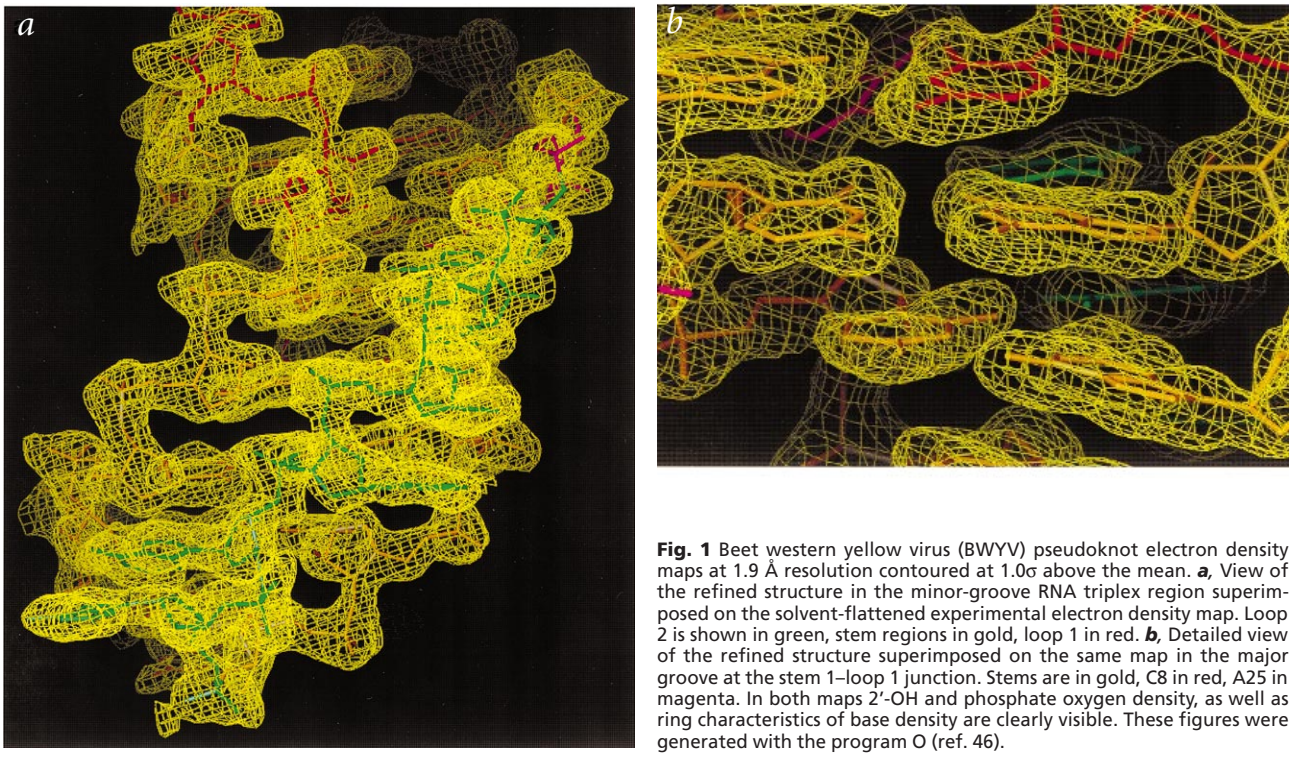

Fig. 1 Beet western yellow virus (BWYV) pseudoknot electron density maps at 1.9 Å resolution contoured at 1.0σ above the mean. ***a***, View of the refined structure in the minor-groove RNA triplex region superimposed on the solvent-flattened experimental electron density map. Loop 2 is shown in green, 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 (ref. 46).

tion 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 (Fig. 2*b*). 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 (Fig. 2*a*,*c*).

The two stems are noncoaxially stacked and appear to hinge at the junction. Continuous stacking of loop 2 onto A25 (magenta) of stem 2 (Fig. 2*d*) and the short length of loop 1 induce the rotation of stem 2 in relation to stem 1 (Fig. 2*c*). 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 (Fig. 2*c*). The helical axes of the two stems are shifted in relation to one another and displaced by ~5 Å, and the average tilt angle between the two stems is ~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[10]. 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 (Fig. 2*a*,*c*,*d*). This considerable stretch is facilitated by overtwisting the base pairs near loop 1 in both stems.

The geometry of stems 1 and 2 differ significantly from standard A-form duplex RNA (Fig. 2*c*). In both stems, the base pairs are largely propeller twisted; stem 1 has an average propeller twist of 15°, while it is 26° for stem 2. In addition, the base pairs 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° (ref. 19). Stem 1 is curved toward 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 ~0.7 Å wider. From top to bottom, stem 2 bends moderately toward 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[8]. 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 (Fig. 3*a*). 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[20,21]. 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 coplanar 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 toward the major groove.

The first base of loop 2 (G19) projects away from stem 1 (Fig. 2*a*). However, the adjacent residue (A20) is 30° out of plane from the G4–C17 base pair (Fig. 2*a*), leading to a stabilizing stacking interaction with the ribose O4' of residue C5, as found in many nucleic acid structures (reviewed in ref. 22). This anchorlike adenosine spans the minor groove and interacts with both strands of stem 1. It is held in this position by an extensive network of seven hydrogen bonds (Fig. 3*b*), using both base and sugar interactions. The 2'-OH of A20 contributes four of the hydrogen bonds by interacting with two layers of bases (Table 2). Two of the hydrogen bonds are bifurcated, involving O2 atoms of residues C5 and C17, resulting in the tilt of these bases (Fig. 2*c*). Some of these adenosine interactions have also been seen in the crystal structures of the hammerhead ribozyme[23] and group I intron[24]. In both cases, the adenosine in an RNA tetraloop contacts the minor-groove side of a G–C base pair to form four of the seven hydrogen bonds

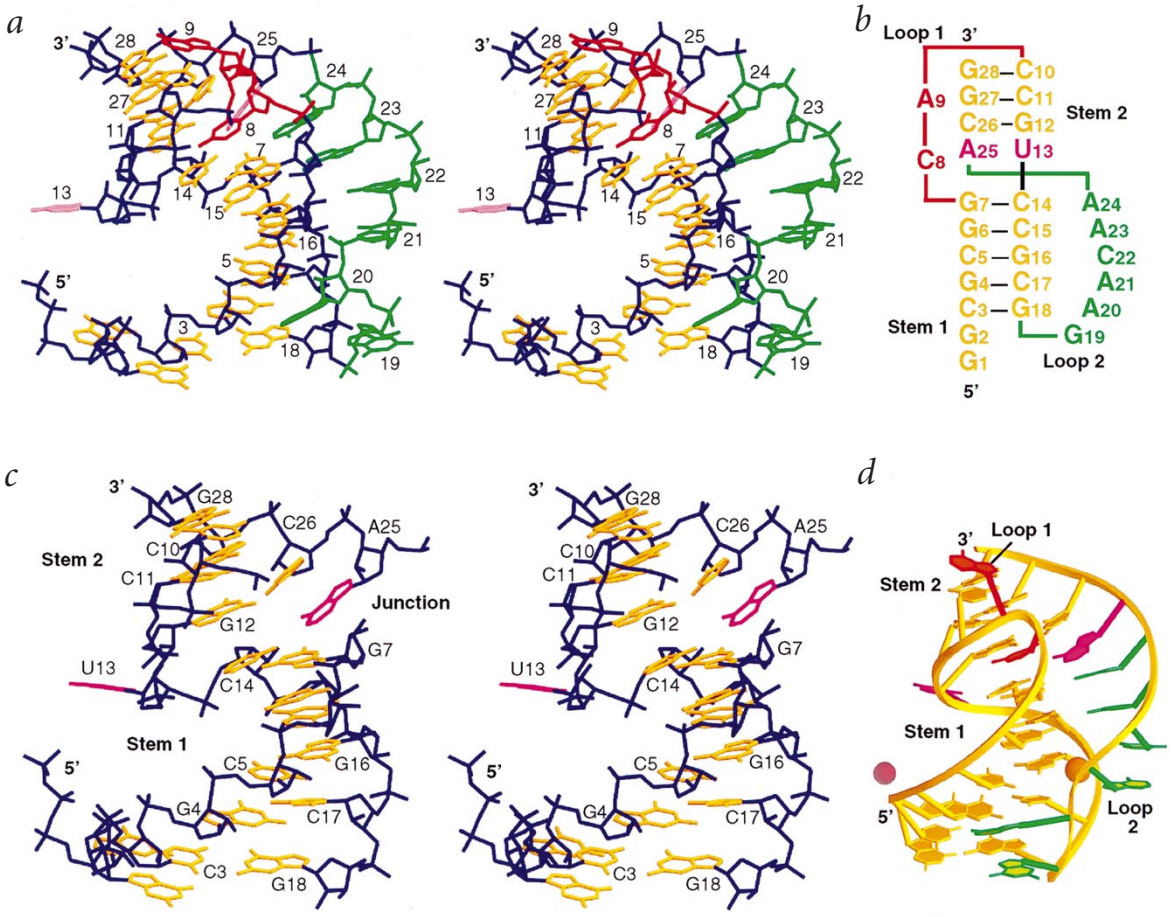

Fig. 2 The beet western yellow virus (BWYV) pseudoknot crystal structure and secondary sequence. ***a,*** Stereo view of the pseudoknot crystal structure. The color scheme of bases corresponds to that used in the secondary-structure diagram in (*b*). Loop 1 (red) crosses the major groove of stem 2, and 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[8]. The crystal structure has significant differences from the secondary prediction: 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 from (*a*) of stem 1 and stem 2 without the loop nucleotides. Stem 2 is rotated in relation to stem 1 and is noncolinearly stacked; all base pairs 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.

observed here. In contrast to A20, the adenines there are not tilted and are largely in the plane of the base pair.

The next two residues A21 and C22 (Fig. 2*a*) 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 (Fig. 3*c*) and by a sodium ion that mediates base–base interactions (Fig. 5*c*). The adjacent A23 is stacked on the exocyclic amino group of C22 (Fig. 3*a*) and utilizes its Watson–Crick face to interact with the neighboring C15 (Fig. 3*d*). A24, the last base of loop 2, stacks on A23 (Fig. 3*a*) and interacts with guanosine G7 through three hydrogen bonds (Fig. 3*e*). This minor groove-specific interaction is unlike other G–A interactions described previously[23–27]. 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 2). 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 (Figs 2*a*, 3). In addition, it is interesting that the temperature factors in the backbone of loop 2, in particular A21 through A25, are 60% higher than those in stem 1, suggesting increased mobility.

The stem 1–stem 2 junction

The first residue of stem 2, A25, stacks on the edge of the A24 base (Fig. 3*a*). There, it is tilted and interacts with bases in different layers (Figs 2*a*, 4). Rather than pairing with U13 as predicted (Fig. 2*b*), the A25 adenine forms three hydrogen bonds at the junction through its Watson–Crick face (Fig. 4*b*). The interactions with the O2 and 2'-OH of C14 are similar to those observed between A23 and C15 (Fig. 3*d*). On the strand opposite to A25 (Fig. 2*a*,*c*), the extruded U13 has a rotated and reversed sugar

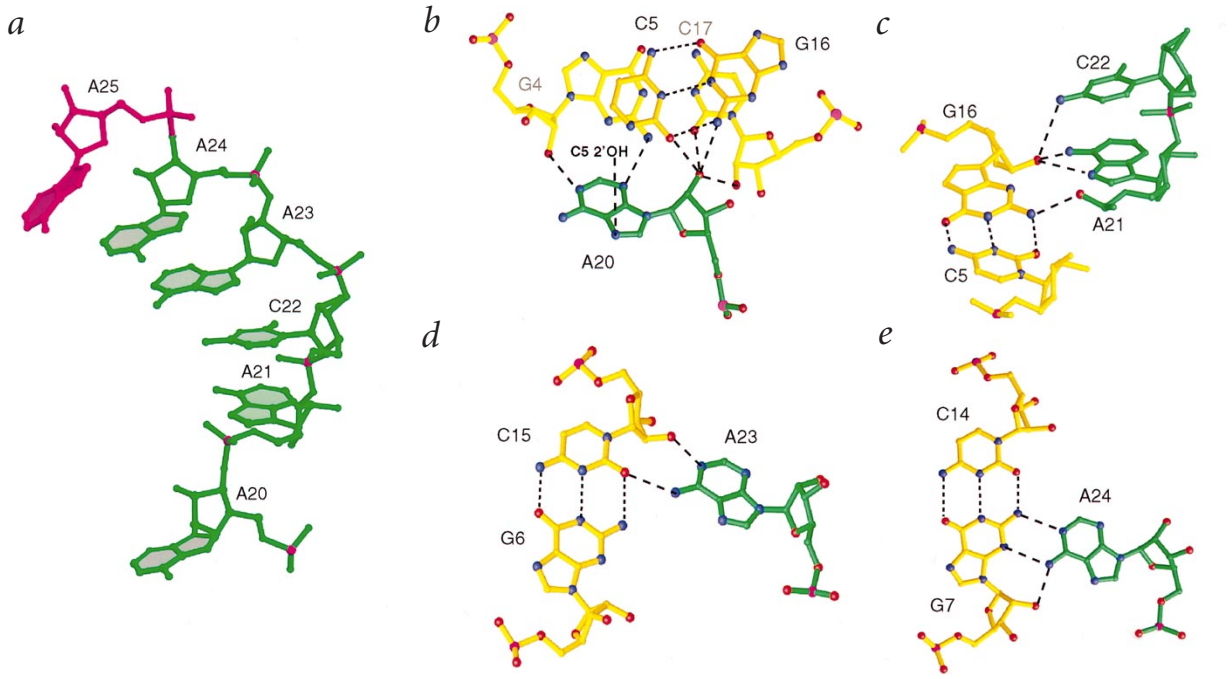

Fig. 3 RNA triplex interactions of loop 2 in the minor groove of stem 1. ***a***, Separate view of the conserved predominantly adenosine ladder. Despite the systematic stacking, each loop base is rotated in different orientations to maximize interactions with the groove nucleotides. ***b***, The tilted A20 interacts with two layers of base pairs 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.

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 stack directly upon the base plane and furanose oxygen of C14 (Fig. 2*c*). Other bulge nucleotide conformations have been observed, although their sugar rotations are less marked[28–30]. In the crystal structure, U13 makes lattice contacts through stacking and base-pairing interactions. Symmetry-related bases that are bulged or unpaired, namely U13, G19 and G1, form a continuous stacking core of six layers. In addition, the U13 nucleotides in this stacking core use their Watson–Crick faces to form Hoogsteen base pairs with symmetry-related adenosine 20.

Quadruple interactions of loop 1

Most frameshifting pseudoknots have a short loop 1, usually one to two nucleotides in length[4–6,8,11]. In the BWYV pseudoknot, a C8–A9 sequence crosses the major groove, and the C residue is conserved in some luteovirus pseudoknots[8]. In the BWYV structure, C8 is inserted into the major groove of stem 2, forming a quadruple-base interaction with G12, A25 and C26 (Figs 2*a*,4*a*). This is the first example of one base hydrogen bonding to three other bases, two of which form a base pair. The G12–C26 base pair has a strong propeller twist, which is needed for C26 to stack upon the tilted A25 (Figs 2*c*,4*a*). As noted by the asterisk in Fig. 4*a*, 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 is also found in other nucleic acid structures[31–33]. Two water molecules stabilize this core interaction by mediating the 2'-OH of C8 to base contacts (A25 and C26) (Fig. 5*b*). 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[28], the hammerhead ribozyme[26,27] and group I intron[24], which are stabilized by multiple interactions. In the BWYV structure, we have named the loop 1 to stem 2 transition a 'C turn' (Fig. 5*a*), 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 in relation to its base-paired G28 (Fig. 2*c*). An organized network of water molecules facilitates the turn by bridging phosphate or base contacts (Fig. 5*a*). The array of water molecules on the other side of loop 1 stabilizes its conformation in the major groove (Fig. 5*b*).

A sharp turn is also found at the junction between stem 1 and loop 2 at G18-G19-A20 (Fig. 2*a*,*d*). This turn is facilitated by C2'-endo sugar puckers (the remainder of the RNA is C3'-endo, except for U13, which has both conformations) at residues G18

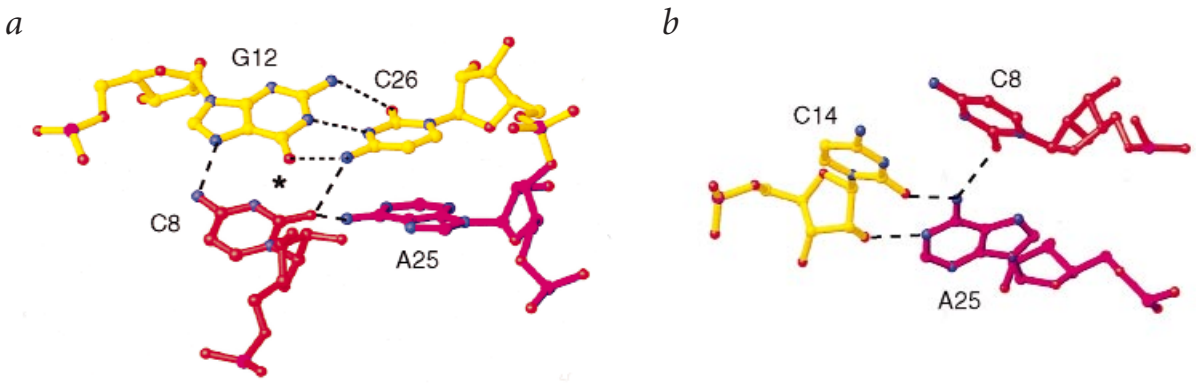

Fig. 4 Quadruple-base interactions of loop 1. ***a***, The C8 organizer base of loop 1 inserts deeply into the major groove of stem 2 and interacts simultaneously with bases of three other nucleotides. The asterisk next to N3 of C8 indicates probable protonation with a hydrogen bond to O6 of G12. C26 is propeller twisted in the C26–G12 base pair 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.

and G19, which lead to an extended backbone conformation. 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.

RNA pseudoknots are generally stabilized from unfolding by magnesium ions[34]. However, in many cases, high concentrations of monovalent ions can also stabilize the conformation[34,35]. 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 because of the gap (Figs 2*d*,3*c*). Water molecules are also involved in stabilizing the conformation around the sodium ion (Fig. 5*c*). An octahedrally coordinated magnesium ion is bound to the triphosphate at the 5' end (Fig. 2*d*), coordinating to phosphate oxygens of the β- and γ-phosphate, the pro-S_p oxygen of the G2 phosphate and three water molecules.

Relevance to frameshifting and RNA packing

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. Pseudoknots are usually in some type of equilibrium between the folded and the unfolded form[34], 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, Table 2) than those engaged in Watson–Crick base pairs (24 total). As the pseudoknot barrier induces ribosomal pausing, it is likely that the ribosome contacts both the slippery site and the pseudoknot simultaneously[36]. 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.

As the ribosome travels along the messenger RNA in translation, the RNA triplex formed at the 5' end of the molecule is likely to be the first feature that the ribosome encounters[36], and the third strand may prevent the unraveling of stem 1 by cross-strand interactions. This 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[9,11]. The triple-base contacts in the structure show a preference for adenosine residues in loop 2, and appear to be sequence specific. Mutations in other pseudoknots that invert base pairs in stem 1[4,5,7] or eliminate the high adenosine content in loop 2 (ref. 12) are reported to have adverse effects on frameshifting. These mutations may disrupt triplex interactions and cause the pseudoknot to unravel.

A second stabilizing feature in the pseudoknot is the array of interactions at the stem 1–stem 2 junction. C8 seems to play an organizing role in assembling the junction between the stems. The junctional A25 base not only provides a continuous stacking platform into stem 2 but also plays an important role in stabilizing the C8 base-specific quadruple interactions. Probable unpairing of a junctional base pair in frameshifting pseudoknots as observed here with U13 has also been reported in several biochemical[11,37] and NMR studies[14]. The ribosomal frameshifting efficiency of the luteovirus pseudoknot is relatively low compared to those that have different junctional sequences[4–6]. 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[6,11,12]. In those systems, the identity of the loop 1 base analogous to C8 is not crucial for frameshifting activity. Nevertheless, it has been proposed that a certain bent conformation induced by an adenosine wedged base between the two stems may also be important for high-level frameshifting[17]. Recently, a 4.8 Å crystal structure of a synthetic SELEX pseudoknot complexed with HIV-1 reverse transcriptase was analyzed. Although not all of the pseudoknot was observed, the structure showed a pronounced kink between the two stems, thereby optimizing its contacts with subunits of the heterodimer[38].

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, A.A., Ortleva-Donnelly, L., Ryder, S.P. Moncoeur, E. & Strobel, S.A., pers. comm.). That study suggested that 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

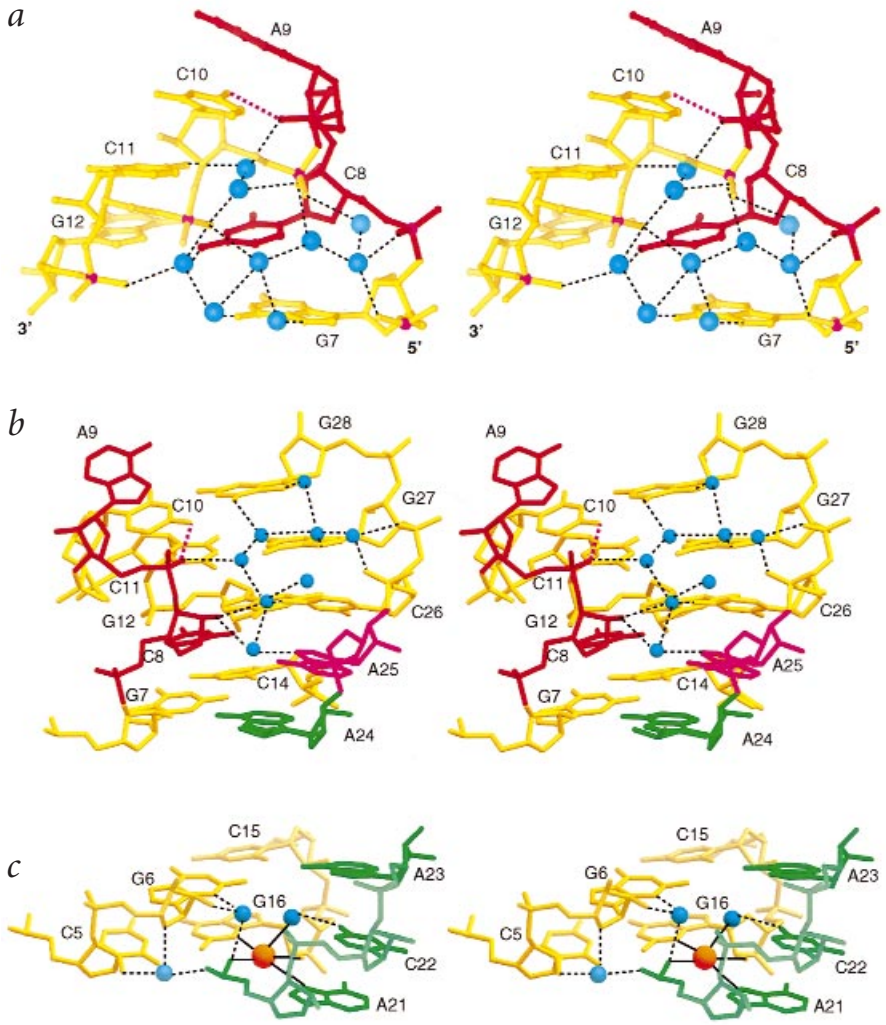

Fig. 5 Stereo view of stabilizing interactions at sharp turns, ions and water molecules. ***a***, The 'C turn' from loop 1 (red) to C10 of stem 2 (gold) is stabilized by a base-to-phosphate hydrogen bond (magenta dashed lines) and an organized water network (cyan spheres). ***b***, An array of water 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 in magenta, and A24 of loop 2 is in green. ***c***, A sodium ion in the minor groove mediates base-to-base contacts 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 Fig. 1 were generated with the program RIBBONS[50].

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[39]. For example, three hydrogen-bonding contacts between loop 2 and stem 1 have been reported in the NMR study of a pseudoknot formed by the turnip yellow mosaic virus genomic RNA[18].

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 (Fig. 3*a*). 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[40]. Other regions of loop 2 that are not involved in tertiary interactions may be fully accessible for ribosomal contact, for example, near G19.

Insight into the requirements for ribosomal frameshifting has been obtained using a reticulocyte lysate translation system to measure the activity of the BWYV pseudoknot and its mutants (Kim, Y.G., Su, L., Maas, S. & Rich, A., pers. comm.). Changing the base pair C3–G18 at the bottom of stem 1 has little effect, but substitution of A25 or C8 abolishes activity. The studies also show that frameshifting is dependent upon loop 2 adenosines. However, the extruded U13 can be deleted, and most of the wild-type frameshifting activity is retained. These findings suggest the importance of the RNA triplex and quaternary interactions for ribosomal frameshifting.

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. In addition, this structure may help design drugs for antiviral therapy that interfere with frameshifting, an event critical for the viral life cycle[41].

Methods

Crystallization and data collection. The 28-nucleotide RNA was prepared by *in vitro* T7 transcription[42] and purified to homogeneity by polyacrylamide gel electrophoresis. Brominated RNA was prepared by substituting UTP with 5-Br-UTP, leading to the incorporation of one bromo-uracil in the RNA sequence. The RNA was extracted from the gel by the crush-and-soak method and concentrated with Centriplus (10 kDa) (Amicon) units. After extensive dialysis against double-distilled water, the RNA was neutralized with 300 mM sodium cacodylate (pH 6.5), incubated at 60 °C for 10 min in the presence of 15 mM $MgCl_2$, followed by slow cooling to 25 °C and

Table 1 Data collection, phase determination, and refinement statistics

Crystallographic data

Data set	Resolution (Å)	Completeness (%)[1]	R_{sym} (%)[1,2]	I/σ[1,3]	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

MIRAS phase determination (native I as native)

	Resolution range (Å)	R_{deriv} (sites)[4]	R_{cullis} iso(c/ac)[5]	R_{cullis} ano[5]	Phasing power iso(c/ac), ano[5,6]
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)

Refinement (native I, 766 nonhydrogen atoms, 124 water molecules, one magnesium ion, and one sodium ion)

Resolution (Å)	Reflections/free set	R_{cryst}/R_{free} (%)[7] ,2σ	R.m.s.d.[8]			Average B-factor (Å²), RNA
			Bonds (Å)	Angles (°)	B-factor (Å²)	
8–1.6	8,793/916	20.7/25.4	0.017	1.87	7.0	24.6

[1]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.
[2]$R_{sym} = \Sigma|I-<I>|/\Sigma I$, where I is the observed intensity, <I> is the statistically weighted average intensity of multiple observations of symmetry-related reflections.
[3]I/σ = Average intensity/error.
[4]$R_{deriv} = \Sigma\,||F_{PH}| - |F_P||\,/\,\Sigma\,|F_P|$, where $|F_P|$ is the nucleic acid structure factor amplitude and $|F_{PH}|$ is the heavy-atom derivative structure factor amplitude.
[5]R_{cullis} = <phase-integrated lack of closure> / $<|F_{PH} - F_P|>$; c is centric data and ac is acentric data; iso is isomorphous and ano is anomalous.
[6]Phasing power = $<[|F_H|\ (calc) / $ phase-integrated lack of closure$]>$, where F_H is the heavy-atom structure factor amplitude.
[7]$R_{cryst} = \Sigma\,||F_o| - |F_c||\,/\,\Sigma|F_o|$, 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.
[8] R.m.s.d. is the root mean square deviation from ideal geometry.

4 °C. Native I crystals (0.6 × 0.15 × 0.15 mm³) were grown at 25 °C from a sitting drop containing 4 mg ml^{-1} RNA, 5 mM $MgCl_2$, 5% sec-butanol, 2 mM spermidine, 100 mM potassium MOPS buffer (pH 7.0) over a reservoir of 18% sec-butanol. The crystals diffracted to 1.6 Å, and crystallized in the space group $P3_221$ (a = b = 30.08 Å, c = 140.08 Å). Brominated crystals could only be grown from solutions containing 5% 2-methyl-2,4-pentanediol (MPD), 5 mM $MgCl_2$, 1 mM spermidine, 100 mM potassium PIPES buffer (pH 6.5). These brominated crystals were slightly non-isomorphous with the native I crystals grown out of sec-butanol. 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 h and backsoaking them in mother liquor overnight.

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. Multiple anomalous dispersion (MAD) data for the bromine and osmium derivative crystals were collected at the Brookhaven X4A and CHESS F2 beamlines, respectively. All data were collected at 100 K, and crystals were soaked in mother liquor containing 25% MPD as a cryoprotectant for 1 min before freezing. Data were processed using the programs DENZO, SCALEPACK[43], and the CCP4 suite[44].

Phase determination and structure refinement. There is one RNA pseudoknot per asymmetric unit. The single bromine site was located from the 2.1 Å isomorphous-difference Patterson map of native II and the bromine derivative data. Five Os sites were located in the Os derivative data collected at the absorption edge by difference Fourier techniques, using the bromine phases. These sites were also confirmed in the Os anomalous-difference Patterson maps. We used a combination of multiple isomorphous-replacement and anomalous-scattering (MIRAS) methods for phasing with the program SHARP[45]. The Residual maps from SHARP revealed other characteristics of the heavy-atom sites, and the space group enantiomorph was determined. The electron density maps were greatly improved through the addition of an alternative position for the major osmium-binding site along with refinement of

Table 2 Tertiary hydrogen bonds

Residue		Distance (Å)	Residue		Distance (Å)
Loop 2					
A20 2'-OH[1]	C5 O2	2.83	A21 N7	G16 2'-OH	3.05
A20 2'-OH[1]	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[2]	G12 O6	2.82			
A9 O2P	C10 N4	2.80	G19 2'-OH	G19 N3	2.77

[1]Bifurcated hydrogen bonds.
[2]Probably protonated.

anisotropic temperature factors for the bromine site and the major osmium site. Using the program O (ref. 46), a poly(rC) model with appropriate sugar puckers was built into the solvent-flattened electron density maps generated from native II and the bromine/osmium derivative data. Using SHARP, an excellent 1.9 Å electron density map was also generated from native I and the bromine/osmium derivative data, confirming the trace, clarifying the identity of bases and revealing the position of the 5'-end β-phosphate. Initial refinement cycles were performed with X-PLOR[47]; 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%). Final rounds of refinement to 1.6 Å were performed using CNS[48], which rapidly decreased both the R-factor and R_{free}. Bulk solvent and anisotropic temperature factor corrections were carried out in the final process. Statistics are presented in Table 1. A high-density 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 magnesium ion was identified by its strict octahedral coordination with ligand distances between 2.1 and 2.3 Å. Helical parameters for the structure were generated by the program CURVES[49].

Coordinates. Coordinates have been deposited in the Protein Data Bank (accession code 437D).

Acknowledgments

We thank E. de La Fortelle for extensive help on the program SHARP; G. Minasov and I. Berger for help in data processing and refinement, Y.-G. Kim for frameshifting discussions; C. Ogata (X4A-NSLS), M. Soltis (SSRL) and the staff at CHESS F2 for synchrotron assistance. We acknowledge help from T. Schwartz, D. Chan, M. Schade and K.-J. Wu on data collection; V. Tereshko, A. and H. Szoke for technical assistance; J. Cate, D. Bartel, M. Rould, U. RajBhandary, A. Herbert and B. Brown for helpful discussions. We are indebted to H. Taube and S. Holbrook for providing us with the osmium hexammine triflate compound; J. McCloskey and P. Crain for bromine mass spectroscopy analysis. We also thank L. Gerhke, K. Lowenhaupt and F. Houser-Scott for assistance in the T7 transcription system; D. Lauffenburger and S. Zhang at the MIT Center of Biomedical Engineering for providing the computer facility. This research was supported by grants from the National Science Foundation, the National Institutes of Health, the National Foundation for Cancer Research and the National Aeronautics and Space Administration.

Received 24 September, 1998; accepted 7 December, 1998.

1. Gesteland, R.F. & Atkins, J.F. Recoding: dynamic reprogramming of translation. *Annu. Rev. Biochem*. **65**, 741–768 (1996).
2. Farabaugh, P.J. Programmed translational frameshifting. *Microbiol. Rev*. **60**, 103–134 (1996).
3. Jacks, T., Madhani, H.D., Masiarz, F.R. & Varmus, H.E. Signals for ribosomal frameshifting in the Rous sarcoma virus *gag-pol* region. *Cell* **55**, 447–458 (1988).
4. Chamorro, M., Parkin, N. & Varmus, H.E. 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 (1992).
5. ten Dam, E., Brierley, I., Inglis, S. & Pleij, C. Identification and analysis of the pseudoknot-containing *gag-pro* ribosomal frameshift signal of simian retrovirus-1. *Nucleic Acids Res.* **22**, 2304–2310 (1994).
6. Brierley, I., Rolley, N.J., Jenner, A.J. & Inglis, S.C. Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol*. **220**, 889–902 (1991).
7. Tzeng T.-H., Tu, C.-L. & Bruenn J.A. Ribosomal frameshifting requires a pseudoknot in the *Saccharomyces cerevisiae* double-stranded RNA virus. *J. Virol.* **66**, 999–1006 (1992).
8. Miller, W.A., Dinesh-Kumar, S.P. & Paul, C.P. Luteovirus gene expression. *Crit. Rev. Plant Sci.* **14**, 179–211 (1995).
9. Somogyi, P., Jenner, A.J., Brierley, I. & Inglis, S.C. Ribosomal pausing during translation of an RNA pseudoknot. *Mol. Cell. Biol*. **13**, 6931–6940 (1993).
10. Pleij, C.W.A., Rietveld, K. & Bosch, L. A new principle of RNA folding on pseudoknotting. *Nucleic Acids Res.* **13**, 1717–1731 (1985).
11. Chen, X. *et al*. 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 (1995).
12. ten Dam, E.B., Verlaan, P.W.G. & Pleij, C.W.A. 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 (1995).
13. Puglisi, J.D., Wyatt, J.R. & Tinoco, I. Jr. Conformation of an RNA pseudoknot. *J. Mol. Biol.* **214,** 437–453 (1990).
14. Shen, L.X. & Tinoco, I. Jr. The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *J. Mol. Biol*. **247**, 963–978 (1995).
15. Du, Z., Giedroc, D.P. & Hoffman, D.W. 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 (1996).
16. Du, Z., Holland, J.A., Hansen, M.R., Giedroc, D.P. & Hoffman, D.W. Basepairings within the RNA pseudoknot associated with simian retrovirus-1 *gag-pro* frameshift site. *J. Mol. Biol*. **270**, 464–470 (1997).
17. Kang, H. & Tinoco, I. Jr. A mutant RNA pseudoknot that promotes ribosomal frameshifting in mouse mammary tumor virus. *Nucleic Acids Res.* **25**, 1943–1949 (1997).
18. Kolk, M.H. *et al*. NMR structure of a classical pseudoknot: interplay of single- and double-stranded RNA. *Science* **280**, 434–438 (1998).
19. Saenger, W. *Principles of nucleic acid structure*. (Springer-Verlag, New York; 1984).
20. Felsenfeld, G., Davies, D.R. & Rich, A. Formation of a three-stranded polynucleotide molecule. *J. Amer. Chem. Soc.* **79**, 2023–2024 (1957).
21. Rajagopal, P. & Feigon, J. Triple-strand formation in the homopurine : homopyrimidine DNA oligonucleotides $d(G\text{-}A)_4$ and $d(T\text{-}C)_4$. *Nature* **339**, 637–640 (1989).
22. Egli, M. & Gessner, R.V. Stereoelectronic effects of deoxyribose O4′ on DNA conformation. *Proc. Natl. Acad. Sci. USA* **92**, 180–184 (1995).
23. Pley, H.W., Flaherty, K.M. & McKay, D.B. 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 (1994).
24. Cate, J. *et al*. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678–1685 (1996).
25. Brown, T. & Hunter, W.N. 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 (1997).
26. Pley, H.W., Flaherty, K.M. & McKay D.B. Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68–74 (1994).
27. Scott, W.G., Finch, J.T. & Klug, A. The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for catalytic cleavage. *Cell* **81**, 991–1002 (1995).
28. Quigley, G.J. & Rich, A. Structural domains of transfer RNA molecules. *Science* **194**, 796–806 (1976).
29. Portmann, S., Grimm, S., Workman, C., Usman, N. & Egli, M. 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 (1996).
30. Klimasauskas, S., Kumar, S., Roberts, R.J. & Cheng, X. Hhal methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357–369 (1994).
31. Hartman, K.A. & Rich, A. The tautomeric form of helical polyribocytidylic acids. *J. Am. Chem. Soc.* **87**, 2033–2039 (1965).
32. Gehring, K., Leroy, J.-L., Guéron, M. A tetrameric DNA structure with protonated cytosine.cytosine base pairs. *Nature* **363**, 561–565 (1993).
33. Chen, L., Cai, L., Zhang, X. & Rich, A. Crystal structure of a four-stranded intercalated DNA: d(C4). *Biochemistry* **33**, 13540–13546 (1994).
34. Wyatt, J.R., Puglisi, J.D. & Tinoco, I. Jr. RNA pseudoknots. Stability and loop size requirements. *J. Mol. Biol*. **214**, 455–470 (1990).
35. Gluick, T.C., Wills, N.M., Gesteland, R.F. & Draper, D.E. Folding of an mRNA pseudoknot required for stop codon readthrough: effects of mono- and divalent ions on stability. *Biochemistry* **36**, 16173–16186 (1997).
36. Tu, C., Tzeng, T.-H. & Bruenn, J.A. Ribosomal movement impeded at a pseudoknot required for frameshifting. *Proc. Natl. Acad. Sci. USA* **89**, 8636–8640 (1992).
37. Sung, D. & Kang, H. Mutational analysis of the RNA pseudoknot involved in efficient ribosomal frameshifting in simian retrovirus-1. *Nucleic Acids Res.* **26**, 1369–1372 (1998).
38. Jaeger, J., Restle T. & Steitz, T.A. The structure of HIV-1 reverse transcriptase complexed with an RNA pseudoknot inhibitor. *EMBO J*. **17**, 4535–4542 (1998).
39. Pleij, C.W.A. RNA Pseudoknots. *Curr. Opin. Struct. Biol*. **4**, 337–344 (1994).
40. Ortoleva-Donnelly, L., Szewczak, A.A., Gutell, R.R. & Strobel S.A. The chemical basis of adenosine conservation throughout the *Tetrahymena* ribozyme. *RNA* **4**, 498–519 (1998).
41. Dinman, J.D., Ruiz-Echevarria, M.J. & Peltz, S.W. Translating old drugs into new treatments: ribosomal frameshifting as a target for antiviral agents. *Trends. Biotechnol*. **16**, 190–196 (1998).
42. Milligan, J.F. & Uhlenbeck, O.C. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol*. **180**, 51–62 (1989).
43. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*. **276**, 307–326 (1997).
44. CCP4. Collaborative computing project number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
45. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol*. **276**, 472–494 (1997).
46. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building models in electron density maps and location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
47. Brünger, A.T. *X-PLOR Version 3.1: a system for X-ray crystallography and NMR*. (Yale University Press, New Haven; 1992).
48. Pann, N.S. & Read, R.J. Improved structure refinement through maximum likelihood. *Acta Crystallogr. A* **52**, 659–668 (1996).
49. Lavery, R. & Sklenar, H. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dynam*. **6**, 63–91 (1988).
50. Carson, M. Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 958–961 (1991).