Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

J. Mol. Biol. (1991) **220**, 889–902

Mutational Analysis of the RNA Pseudoknot Component of a Coronavirus Ribosomal Frameshifting Signal

Ian Brierley†, Nicola J. Rolley, Alison J. Jenner and Stephen C. Inglis

Division of Virology
Department of Pathology
University of Cambridge
Tennis Court Road, Cambridge CB2 1QP, U.K.

(Received 12 February 1991; accepted 18 April 1991)

The genomic RNA of the coronavirus IBV contains an efficient ribosomal frameshift signal at the junction of the overlapping 1a and 1b open reading frames. The signal is comprised of two elements, a heptanucleotide "slip-site" and a downstream tertiary RNA structure in the form of an RNA pseudoknot. We have investigated the structure of the pseudoknot and its contribution to the frameshift process by analysing the frameshifting properties of a series of pseudoknot mutants. Our results show that the pseudoknot structure closely resembles that which can be predicted from current building rules, although base-pair formation at the region where the two pseudoknot stems are thought to stack co-axially is not a pre-requisite for efficient frameshifting. The stems, however, must be in close proximity to generate a functional structure. In general, the removal of a single base-pair contact in either stem is sufficient to reduce or abolish frameshifting. No primary sequence determinants in the stems or loops appear to be involved in the frameshift process; as long as the overall structure is maintained, frameshifting is highly efficient. Thus, small insertions into the pseudoknot loops and a deletion in loop 2 that reduced its length to the predicted functional minimum did not influence frameshifting. However, a large insertion (467 nucleotides) into loop 2 abolished frameshifting. A simple stem–loop structure with a base-paired stem of the same length and nucleotide composition as the stacked stems of the pseudoknot could not functionally replace the pseudoknot, suggesting that some particular conformational feature of the pseudoknot determines its ability to promote frameshifting.

Keywords: RNA pseudoknot; ribosomal frameshifting; coronavirus; translational regulation; RNA structure

1. Introduction

Viruses employ a wide variety of strategies to co-ordinate and control gene expression. Over the last few years it has been recognized that a number of viruses of higher eukaryotic organisms, particularly the retroviruses, utilize ribosomal frameshifting to control expression of their replicases (for a review, see Craigen & Caskey, 1987; ten Dam *et al.*, 1990). Ribosomal frameshifting is a directed change in translational reading frame which allows the production of a single protein from two (or more) overlapping genes, and was first described for the vertebrate retroviruses, Rous sarcoma virus (RSV‡) (Jacks & Varmus, 1985) and mouse mammary tumour virus (MMTV) (Moore *et al.*, 1987; Jacks *et al.*, 1987). Retroviral frameshifting appears to be a mechanism for regulating the expression of the viral RNA-dependent DNA polymerase; one termination codon in RSV and two in MMTV are suppressed by −1 ribosomal frameshifts to generate the Gag/Pol (RSV) and Gag/Pro/Pol (MMTV) polyproteins from which the viral polymerases are subsequently

† Author to whom all correspondence should be addressed.

‡ Abbreviations used: RSV, Rous sarcoma virus; MMTV, mouse mammary tumour virus; IBV, infectious bronchitis virus; ORF, open reading frame; n.m.r., nuclear magnetic resonance; bp, base-pair(s); kb, 10^3 base-pairs; MHV, mouse hepatitis virus.

0022-2836/91/160889-14 $03.00/0

derived. A similar kind of translational strategy is employed by the avian coronavirus infectious bronchitis virus (IBV). The 5′ end of the IBV genomic RNA contains two briefly overlapping open reading frames (ORF) 1a and 1b (formerly F1 and F2), with 1b in the −1 reading frame with respect to 1a (Boursnell *et al.*, 1987). We have shown that the 1b ORF is expressed as a fusion with the upstream 1a ORF following a highly efficient (30%) ribosomal frameshift event that takes place within the overlap region (Brierley *et al.*, 1987). More recently, we investigated the precise sequence components of the frameshift signal and defined an 86 nucleotide stretch that is in itself sufficient to direct frameshifting in a heterologous genetic context (Brierley *et al.*, 1989). The signal is composed of two distinct elements; a heptanucleotide "slippery" sequence, UUUAAAC, the probable site of the ribosomal slip (Jacks *et al.*, 1988) and a tertiary RNA structure in the form of an RNA pseudoknot downstream from this sequence. We were able to establish the presence of this tertiary RNA structure by creating, using site-directed mutagenesis, a number of complementary and compensatory base changes within the predicted RNA helices of the pseudoknot and then testing the ability of synthetic RNA transcripts containing the mutant sequences to promote frameshifting in a cell-free translation system. These experiments indicated that the pseudoknot was essential for high efficiency frameshifting and, in addition, had to be positioned at a precise distance downstream from the slippery sequence for frameshifting to occur.

Although there is little experimental evidence, it is highly likely that pseudoknots are more generally involved in the process of eukaryotic ribosomal frameshifting, since sequences capable of forming such structures can be found at the frameshift sites of many retroviruses (Brierley *et al.*, 1989; ten Dam *et al.*, 1990) and a number of other viruses suspected to use frameshifting (ten Dam *et al.*, 1990). Indeed, a pseudoknot structure has been implicated in the frameshift signal of RSV (Jacks *et al.*, 1988). Pseudoknots were first identified as structural elements at the 3′ end of certain plant viral RNAs (Rietveld *et al.*, 1982, 1983, 1984; Joshi *et al.*, 1983; van Belkum *et al.*, 1985). It has been suggested that the formation of such structures in RNA may not be uncommon (Pleij *et al.*, 1985) and examples have been found in viral and messenger RNAs, in ribosomal RNAs and potentially in the catalytic sites of some ribozymes (for reviews, see Pleij & Bosch, 1989; Pleij, 1990). Almost all pseudoknots identified to date, including the IBV pseudoknot, are of the hairpin-loop type (ten Dam *et al.*, 1990), which form when nucleotides in the single-stranded loop of a hairpin-loop base-pair with a complementary sequence elsewhere in the RNA chain (see Fig. 1(a)). The resulting configuration contains two base-paired stems, S1 and S2, which are thought to stack coaxially to form a quasi-continuous, extended double-helix. The stacked helices are assumed to adopt the conformation of the A-type RNA helix and are connected by single-stranded loops L1 and L2, which span the major and minor grooves of the helix, respectively. Stacking of the two stems in a

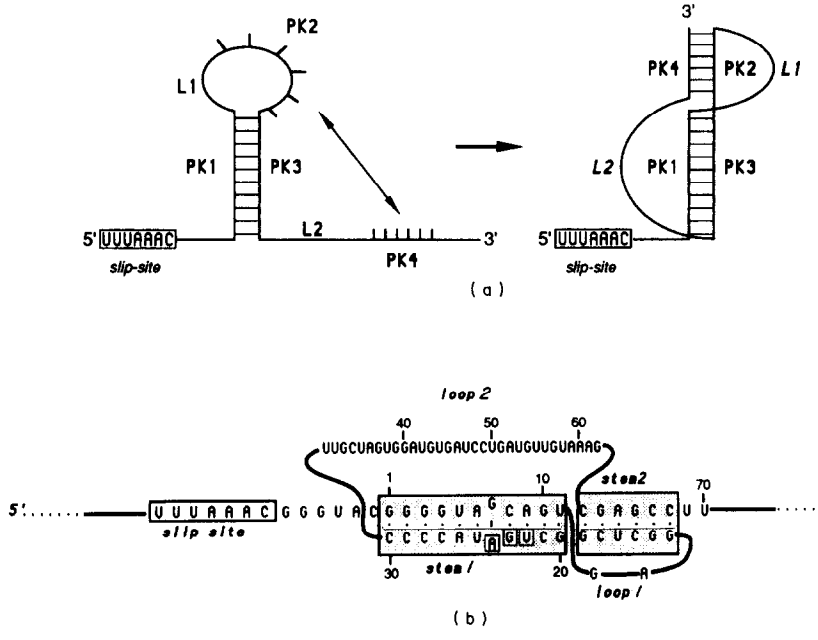

Figure 1. (a) Formation of an RNA pseudoknot at the IBV frameshift site. Nucleotides in the loop of the stem–loop structure (PK2) base-pair with a region downstream (PK4) to generate an extended quasi-continuous double helix. The helical regions are connected by single-stranded loops L1 and L2. (b) IBV pseudoknot model. The pseudoknot contains a quasi-continuous double helix of 16 base-pairs (shaded) with a mismatched pair (G7-A24) in stem 1. The single-stranded loops contain 2 (loop 1) and 32 (loop 2) nucleotides, respectively. The termination codon of the 1a ORF (UGA) is boxed, as is the slip site (UUUAAAC).

pseudoknotted, short (24 nucleotides) oligoribonucleotide has recently been demonstrated by nuclear magnetic resonance (n.m.r.) spectroscopy (Puglisi *et al.*, 1990). Shown in Figure 1(b) is the proposed structure of the IBV pseudoknot, based on nucleotide sequence analysis and our initial mutagenesis data (Brierley *et al.*, 1989) and conforming to current building principles (Pleij *et al.*, 1985; Pleij & Bosch, 1989). In this model, the stems formed by base-pairing between the PK1·PK3 and PK2·PK4 sequences have been stacked coaxially to generate a 16 base-pair quasi-continuous helix with one mismatched pair (G7-A24) in S1. The connecting loops L1 and L2 contain two and 32 nucleotides, respectively.

The mechanism by which the pseudoknot promotes frameshifting is not yet clear, but it has been suggested that the elongating ribosome encounters and is required to unwind the pseudoknot whilst translating the slippery sequence codons, and that this interaction may slow or stall the passage of the ribosome along the mRNA, promoting a (−1) frameshift at the slippery sequence (Jacks *et al.*, 1988; Brierley *et al.*, 1989). A clearer understanding of how such events could occur, however, has been hindered by uncertainties over the precise structure of the pseudoknot. In addition, the potential contributions of individual elements of the pseudoknot structure and, indeed, of particular nucleotides within the structure to the frameshift process remain poorly characterized. We therefore set out to investigate the structure of the IBV pseudoknot by site-directed mutagenesis on the premise that nucleotide changes predicted to destabilize the structure should be inhibitory to frameshifting. Our results largely confirm the model shown in Figure 1(b) and suggest that frameshifting in the IBV system is not dependent on any primary sequence determinants within the pseudoknot; as long as the overall structure is maintained, frameshifting occurs at high efficiency. We further show that the pseudoknot cannot be replaced in the frameshift signal by a simple hair-pin. Thus, the contribution of this tertiary structure is not simply due to its energetic stability, but rather to its unusual conformation.

2. Materials and Methods

(a) *Site-specific mutagenesis*

Site-specific mutations within the IBV frameshift region were prepared using a procedure based on the method described by Kunkel (1985) (Brierley *et al.*, 1989). Plasmid pFS8 (or mutant derivatives, see below) contains the intergenic region of the filamentous bacteriophage f1 (Dotto *et al.*, 1981) enabling single-stranded pFS8 DNA to be generated following infection of plasmid-carrying bacteria with bacteriophage R408 (Russel *et al.*, 1986). Uracil-containing, single-stranded DNA substrates for mutagenesis were prepared by R408-superinfection of plasmid-carrying *Escherichia coli* RZ1032 cells (*dut*$^-$ *ung*$^-$: Kunkel, 1985). Mutagenic oligonucleotides were synthesized using an Applied Biosystems 381A DNA synthesizer and the mutagenesis reactions performed as before (Brierley *et al.*, 1989). Mutants were identified by dideoxy sequencing (Sanger *et al.*, 1977) of single-stranded DNA templates rescued from *E. coli* JM101 (Yanisch-Perron *et al.*, 1985).

(b) *Construction of plasmids*

Plasmid pFS8 (see the text) was constructed from plasmid pFS7 (Brierley *et al.*, 1989) by introducing an oligonucleotide (sequence 5′ AATTAATACGACTCACTATAGGGAGA 3′) containing the bacteriophage T7 RNA polymerase promoter just downstream from the bacteriophage SP6 RNA polymerase promoter of pFS7, by site-directed mutagenesis. Plasmid pFS8.47, a derivative of pFS8 was constructed as follows. Firstly, a unique *Xho*I restriction endonuclease cut-site was introduced into loop 2 of the IBV pseudoknot (see the text) at position 12,406 in the IBV genome (Boursnell *et al.*, 1987) by insertion mutagenesis to create plasmid pFS8.1. Plasmid pVB2+ (Brierley *et al.*, 1987) was digested with *Rsa*I and *Pvu*II and a 467 bp fragment from the influenza A/PR8/34 PB2 gene (sequence information from position 710 to 1176 bp: Young *et al.*, 1983) isolated. Plasmid pFS8.47 was created by ligating this fragment into plasmid pFS8.1, which had previously been digested with *Xho*I and end-filled using DNA polymerase Klenow fragment. The PB2 fragment was chosen such that the ORFs of the IBV and influenza portions of the construct within loop 2 were contiguous. The junction sequences were confirmed by dideoxy nucleotide sequencing (Sanger *et al.*, 1977) of single-stranded pFS8.47. Plasmids were maintained in *E. coli* JM101.

(c) *Preparation of plasmid DNA template for* in vitro *transcription*

Plasmids for transcription were prepared by the alkaline lysis mini-preparation method (Birnboim & Doly, 1979) and linearized by digestion with *Sma*I. Digests were extracted once with a mixture of phenol and chloroform (1 : 1, by vol.) and the aqueous phase passed through a Sephadex G-50 spin column (Maniatis *et al.*, 1982) equilibrated with water. Linearized template was concentrated by precipitation with ethanol and transcribed as described (Brierley *et al.*, 1987), except that T7 RNA polymerase (Gibco-Bethesda Research Laboratories) replaced SP6 RNA polymerase and the concentration of each ribonucleotide in the reaction was doubled (to 1 mM each of ATP, CTP, UTP and 0·05 mM-GTP).

(d) *Translation of synthetic mRNAs* in vitro

Purified mRNAs were translated in rabbit reticulocyte lysates as described (Brierley *et al.*, 1987) and translation products analysed on SDS/10% (w/v) polyacrylamide gels according to standard procedures (Hames, 1981). The relative abundance of non-frameshifted or frameshifted products on the gels was estimated by scanning densitometry and adjusted to take into account the differential methionine content of the products.

3. Results

(a) *Ribosomal frameshift assay*

Our approach to the analysis of the contribution of structural features of the pseudoknot to the

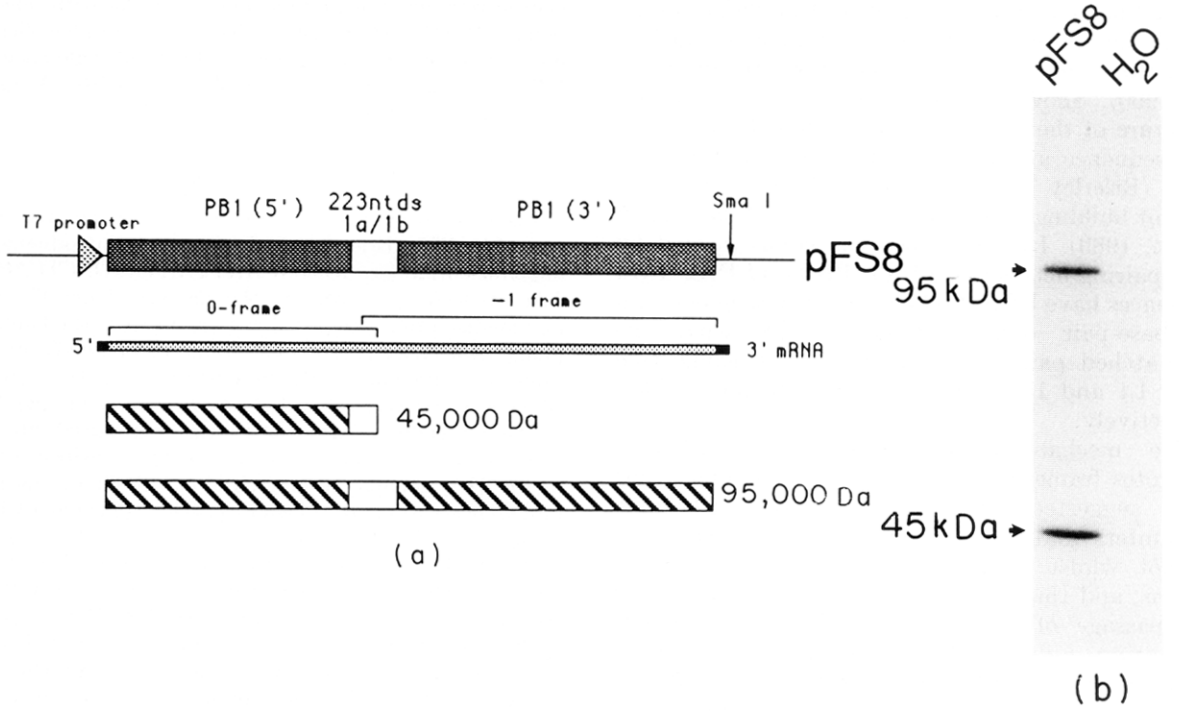

Figure 2. Ribosomal frameshifting assay. (a) The diagram of plasmid pFS8 shows the IBV 1a/1b ORF overlap region of 223 nucleotides containing the frameshift signal (empty box) flanked by the influenza PB1 reporter gene (shaded). Linearization of the plasmid with *Sma*I and *in vitro* transcription using T7 RNA polymerase yields an mRNA (2·8 kb) which, when translated in rabbit reticulocyte lysates, produces a 45,000 Da product corresponding to ribosomes that terminate at the 1a termination codon within the 1a/1b overlap region, and a 95,000 Da (−1) frameshift product corresponding to a PB1 (5′)-1a-1b-PB1-(3′) fusion protein. (b) Reticulocyte lysate translation products synthesized in response to mRNA derived from *Sma*I-digested pFS8. The RNA was translated and products were labelled with [^{35}S]methionine as described in Materials and Methods. Polypeptides were separated on an SDS/10% polyacrylamide gel and detected by autoradiography.

frameshift process was to create defined mutations within the structure and to assess the effect of these changes on the efficiency of frameshifting in a cell-free translation system. The frameshift assay shown in Figure 2 was almost exactly as described (Brierley *et al.*, 1989), except that synthetic mRNAs were generated using bacteriophage T7 RNA polymerase. Plasmid pFS8 has a 230 bp cDNA region derived from the IBV 1a/1b overlap region (and containing the essential 86 nucleotides) cloned within a reporter gene (PB1 of influenza virus A/PR8/34) which, in turn, is flanked by the 5′ and 3′ non-coding regions of the *Xenopus* β-globin gene (Krieg & Melton, 1984) downstream from a T7 promoter. Linearization of the plasmid with *Sma*I followed by transcription results in the production of a capped and polyadenylated 2·8 kb mRNA designed such that on translation in rabbit reticulocyte lysates, ribosomes which terminate at the 1a ORF stop codon produce a 45,000 Da product (the "stopped" product) and those that frameshift, a 95,000 Da product (Brierley *et al.*, 1989).

(b) *Mutational analysis of the pseudoknot region*

(i) *Pseudoknot stem analysis*

In our analysis of the pseudoknot, we selected four particular features for study; the stems, the nucleotides where the stems are thought to stack, the G7-A24 mismatched pair and the loops (see Fig. 1(b) and Fig. 3). We first examined the effect on frameshifting of disrupting and reforming the base-paired regions within S1 and S2 predicted by our model (Fig. 1(b)) and the results of these changes are shown in Figure 3(a). For each region studied, our strategy was to change the nucleotides of each strand of the predicted base-paired regions to their complementary nucleotides in separate constructs, and then to create double-mutant, pseudo-wild-type constructs in which both changes are made and so should be compensatory. We also introduced a number of additional point mutations into the stems, and these are shown in Figure 4. In the main, the results of the stem analysis support strongly the idea that the overall stability of each of the pseudoknot stems is related to its ability to promote frameshifting. As can be seen in Figure 3, all the complementary changes predicted to destabilize the stems reduced or abolished frameshifting. In the double-mutant, pseudo-wild-type constructs, in which the stems are predicted to be restabilized, frameshifting was restored to high levels (15 to 30%) in all cases. Consistent with the importance of stem-stability to the frameshift process is the observation that the efficiency of frameshifting was less dramatically reduced (to about 15%) in

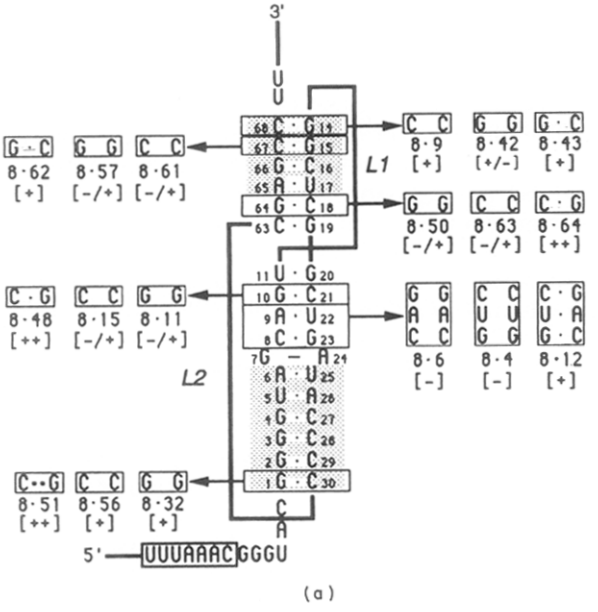

Figure 3. (a) Complementary and compensatory changes within the pseudoknot region. In this representation of the pseudoknot, the nucleotides in the stems are arranged vertically, and the loops (L1 and L2) are shown as thick lines. For each base-pair studied, the 2 complementary changes (no base-pairing) and the compensatory change (base-pairing restored) are boxed and labelled with a mutant number. Beneath this number is an indication of the frameshift efficiency of the mutant; (−) no detectable frameshift; (−/+) 1% or less; (+/−) 5% or less; (+) 10 to 20%; (++) wild-type (25 to 30%). The shaded nucleotides in the stems represent the 2 blocks of nucleotides that were tested by this method and shown to be base-paired in a previous analysis (Brierley *et al.*, 1989).

constructs where the substitutions were at the very ends of the S1 helix (pFS8.32: 8·56, Fig. 3(a)). These changes would be expected to destabilize the structure only slightly and this is supported by estimations of the expected change in free energy upon formation of the mutant structures compared to that of the wild-type structure using the base-pair stacking rules of Turner *et al.* (1988). As it is not possible to apply these rules to pseudoknots as a whole (since the contribution of the loops and the stacking of S2 upon S1 to the free energy of formation of the pseudoknot is not known), we considered only the S1 stem–loop structure in isolation. The changes in free energy calculated were $-12{\cdot}4$ kcal mol^{-1} for wild-type S1, $-11{\cdot}4$ kcal mol^{-1} for S1 in pFS8.32 (G1-G30 at the base of S1) and $-10{\cdot}6$ kcal mol^{-1} for S1 in pFS8.56 (C1-C30 at the base of S1). More central changes, which are expected to be highly destabilizing, greatly reduced frameshifting. This was particularly apparent when G·C base-pairs were changed (pFS8.11: 8·15, 8·31, 8·33 and 8·52 in S1; pFS8.50: 8·57, 8·61 and 8·63 in S2) where the frameshift efficiency was 2% or less. Changes in A·U base-pairs were also inhibitory, although in the examples studied, less dramatic in S1 (pFS8.49, 10%) than in S2 (pFS8.53, 1%). Calculations of the predicted stability of S1 in the more central S1 mutants using the Turner rules support the hypothesis that in these mutants, S1 is considerably less stable. The changes in free energy calculated were -8 kcal mol^{-1} (8·11 and 8·15), $-5{\cdot}8$ kcal mol^{-1} (8·31 and 8·33) and $-8{\cdot}4$ kcal mol^{-1} (8·49).

A point mutation analysis of the mismatched G7-A24 pair in S1 (Fig. 4) indicated that the iden-

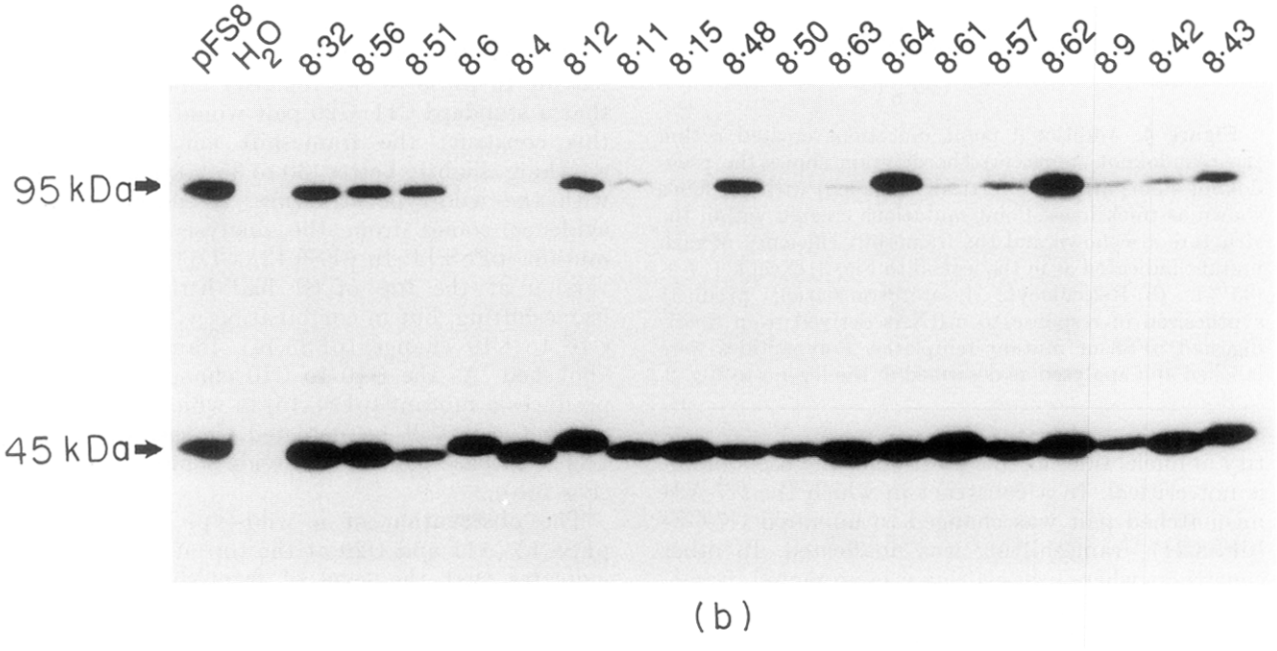

Figure 3. (b) Reticulocyte lysate translation products synthesized in response to mRNAs derived from *Sma*I-digested pFS8 or mutant templates. H_2O, indicates a no RNA control translation. Polypeptides were labelled and analysed as described in the Legend to Fig. 2.

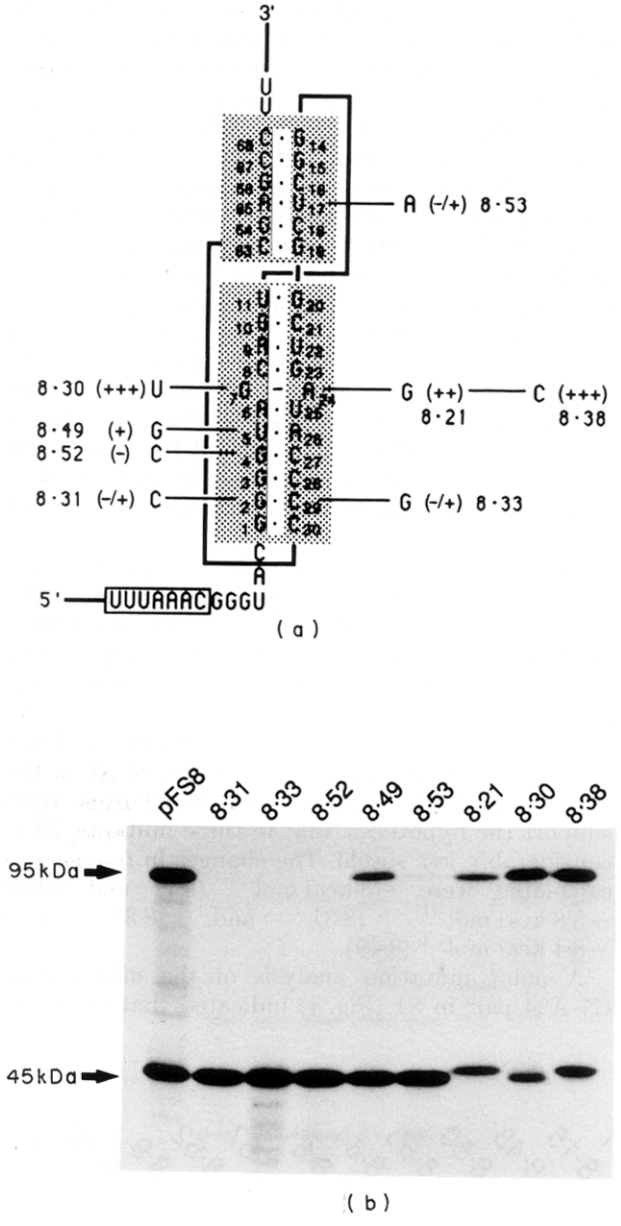

Figure 4. Additional point mutations created within the pseudoknot stems. (a) The diagram shows the pseudoknot stems arranged vertically (shaded) with the loops shown as thick lines. Point mutations created within the structure are shown and the frameshift efficiency of each mutant indicated as in the legend to Fig. 3, except + + + (35%). (b) Reticulocyte lysate translation products synthesized in response to mRNAs derived from *Sma*I-digested pFS8 or mutant templates. Polypeptides were labelled and analysed as described in the legend to Fig. 2.

tity of nucleotides at this position of the pseudoknot is not critical. In a construct in which the G7-A24 mismatched pair was changed to unpaired G7-G24 (pFS8.21), frameshifting was unaffected. In other constructs where base-pairing was promoted, frameshifting was at slightly greater than the wild-type level (35%) (pFS8.30: U7·A24; pFS8.38: G7·C34). Thus, forming a canonical Watson–Crick base-pair actually improves the frameshift process and this seems once again to be related to an increase in the overall stability of S1. Calculations of the predicted stability of S1 in these mutants using the Turner rules support this increase in stability (pFS8.30: $-16{\cdot}4$ kcal mol^{-1}; pFS8.38: $-18{\cdot}3$ kcal mol^{-1}).

The final feature studied in the stem analysis concerned the nucleotides at the junction where S1 and S2 are thought to stack coaxially. Such an arrangement of stems is strongly suggested from models of the pseudoknots in plant viral tRNA-like structures, and is one of the central features of the pseudoknot building principle (Pleij *et al.*, 1985). The results of a mutational analysis of the "stacking region" of the IBV pseudoknot are shown in Figure 5. We expected that the introduction of a mismatched nucleotide pair at the top of S1 or bottom of S2 would destabilize the pseudoknot and inhibit frameshifting, since the stacking of S2 upon S1 in such mutants would be energetically less favourable. Thus, when we changed G20 to C20 (pFS8.10) such that the predicted U11·G20 base-pair at the top of S1 would be replaced by unpaired U11-C20, we were surprised to find that frameshifting was still efficient (15%). This result could be explained by suggesting that in this mutant, U11 may be displaced into loop 1 and replaced by G62 from loop 2 such that a new G62·C20 base-pair could form at the junction between S1 and S2 (Fig. 5(a)). In order to test this possibility, G62 was changed to C62 in a pFS8.10 background (to create pFS8.34) such that we could be confident that only mismatched base-pairs were present at the top of S1 (either U11-C20, or C62-C20 if U11 is displaced). Once again, this mutant displayed efficient frameshifting (15%). A similar efficiency was seen in pFS8.39, in which U11 was replaced by G11 such that an unmatched G11-G20 pair was present at the top of S1 (Fig. 5(b)). Thus, a standard base-pair at the top of S1 is not absolutely required. It appears, however, that base-pairing at this location can contribute to frameshifting, albeit to a limited extent. In pFS8.40, we replaced U11 by C11 such that a standard C11·G20 pair would be present; in this construct the frameshift efficiency was, if anything, slightly better (30 to 35%) than that seen with the wild-type structure. A second piece of evidence comes from the analysis of a double mutant, pFS8.14. In pFS8.13, a U11 to A11 transversion at the top of S1 had little effect upon frameshifting, but in combination with an adjacent G10 to C10 change (pFS8.14), frameshifting was abolished. As the G10 to C10 change in isolation produces a mutant (pFS8.15) in which a frameshift product can still be detected, it seems that the U11·G20 base-pair does provide some stabilization (Fig. 5(b)).

The observation of a wild-type frameshift in pFS8.13 (A11 and G20 at the top of S1), however, indicates that the level of frameshifting seen in mutants created at the top of S1 may, perhaps, not be determined simply by the presence or absence of a base-pair at this position. This possibility is supported by our analysis of the contribution of the C63·G19 base-pair at the bottom of S2 to the

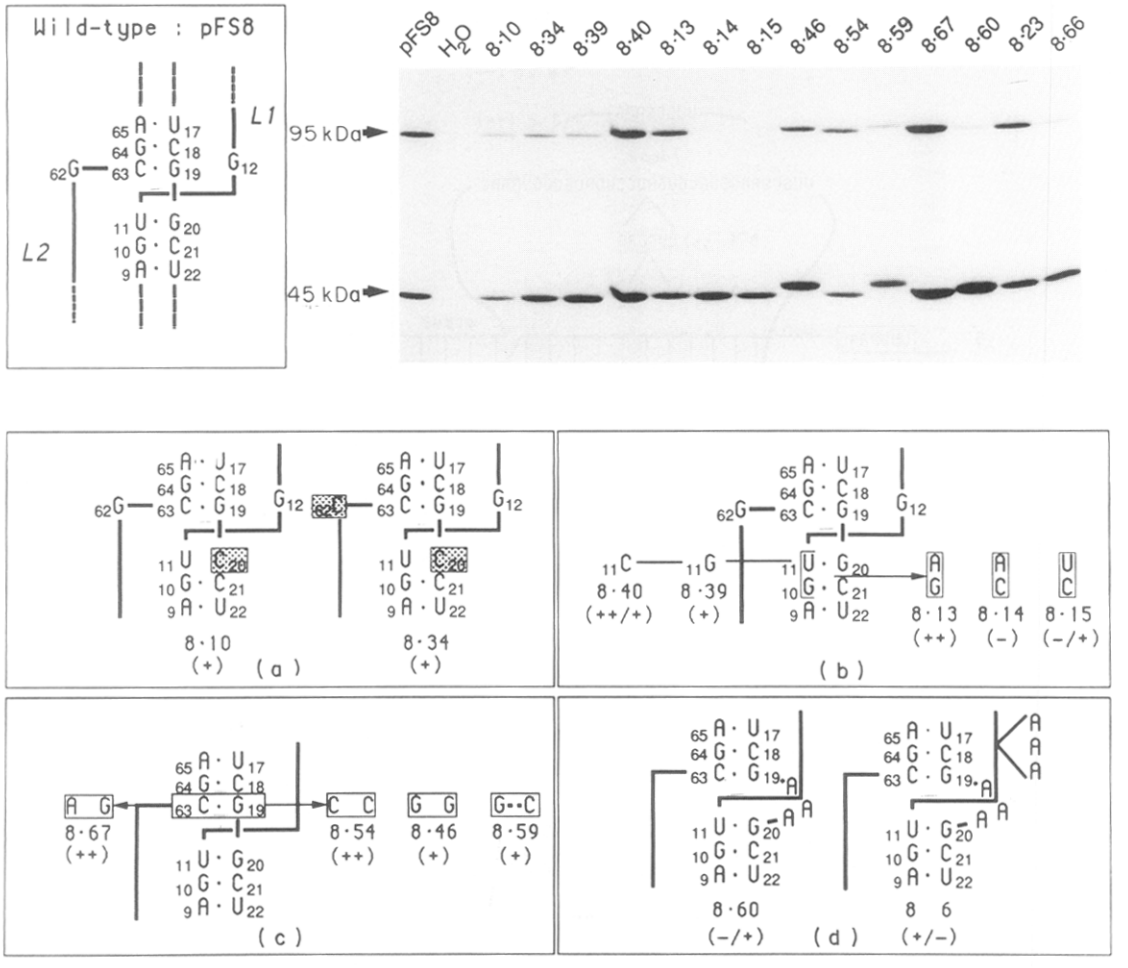

Figure 5. Analysis of the stacking region. The diagram shows mutants created in the nucleotides where stem 1 and stem 2 are thought to stack coaxially. Only 3 bp from each stem are shown. (a) Mutant nucleotides are boxed and shaded. (b) Changes at the top of stem 1 are shown; the frameshift efficiency of each mutant is indicated as in the legend to Fig. 3, except + +/+, (30 to 35%). (c) Changes in the C63·G19 pair at the base of stem 2 are shown. (d) Insertion of 3 nucleotides (AAA) between stem 1 and stem 2, with a loop 1 length of either 2 nucleotides (GA) in 8·60 or 5 nucleotides (GAAAA) in 8·66. The gel shows reticulocyte lysate translation products synthesized in response to mRNAs derived from *Sma*I-digested pFS8 or mutant templates. H_2O indicates a no RNA control translation. Polypeptides were labelled and analysed as described in the legend to Fig. 2.

frameshift process (Fig. 5(c)). In pFS8.46, a point mutation that created unpaired G63-G19 at this position reduced the frameshift efficiency by half to about 15%, a level similar to that seen with the equivalent change at the top of S1 (G11-G20, pFS8.39). In a mutant in which the corresponding C63-C19 mismatch was created (pF8.54), frameshifting was unaffected (30%), yet surprisingly, in a pseudo-wild-type double mutant (G63·C19, pFS8.59), frameshifting was once again reduced (to 15%). This was an unexpected observation, since in the analysis of the U11 and G20 pair at the top of S1, wild-type frameshifting was seen with all the constructs in which the bases at this position were paired and raised the possibility that there may be a specific requirement for the C63 nucleotide at the bottom of S2. However, in an additional mutant (pFS8.67) in which C63 was changed to A63 to create unpaired A63-G19, frameshifting was at the wild-type level. The apparent inconsistencies in the correlation of frameshift efficiency with base-pair formation at the stacking region is not fully understood, but may be related to the possibility that the structure of the RNA in this region of the pseudoknot is unusual. The n.m.r. analysis of a short synthetic RNA pseudoknot performed by Puglisi *et al.* (1990) revealed that although the two pseudoknot stems did, indeed, stack, the A-form geometry of the RNA helix was distorted at the junction of the loops and the stacked stems. Thus, the biological effect of point mutations which influence the particular bases at this region (particularly, in the case of the IBV pseudoknot, U11 and C63) probably cannot be interpreted simply on the basis of the ability to form base-pairs. Clearly, more detailed information on the three-dimensional structure of the stacking region of the IBV pseudoknot is needed before the effects of changes in this region can be fully interpreted. Nevertheless, the finding that preventing base-pair formation in the stacking

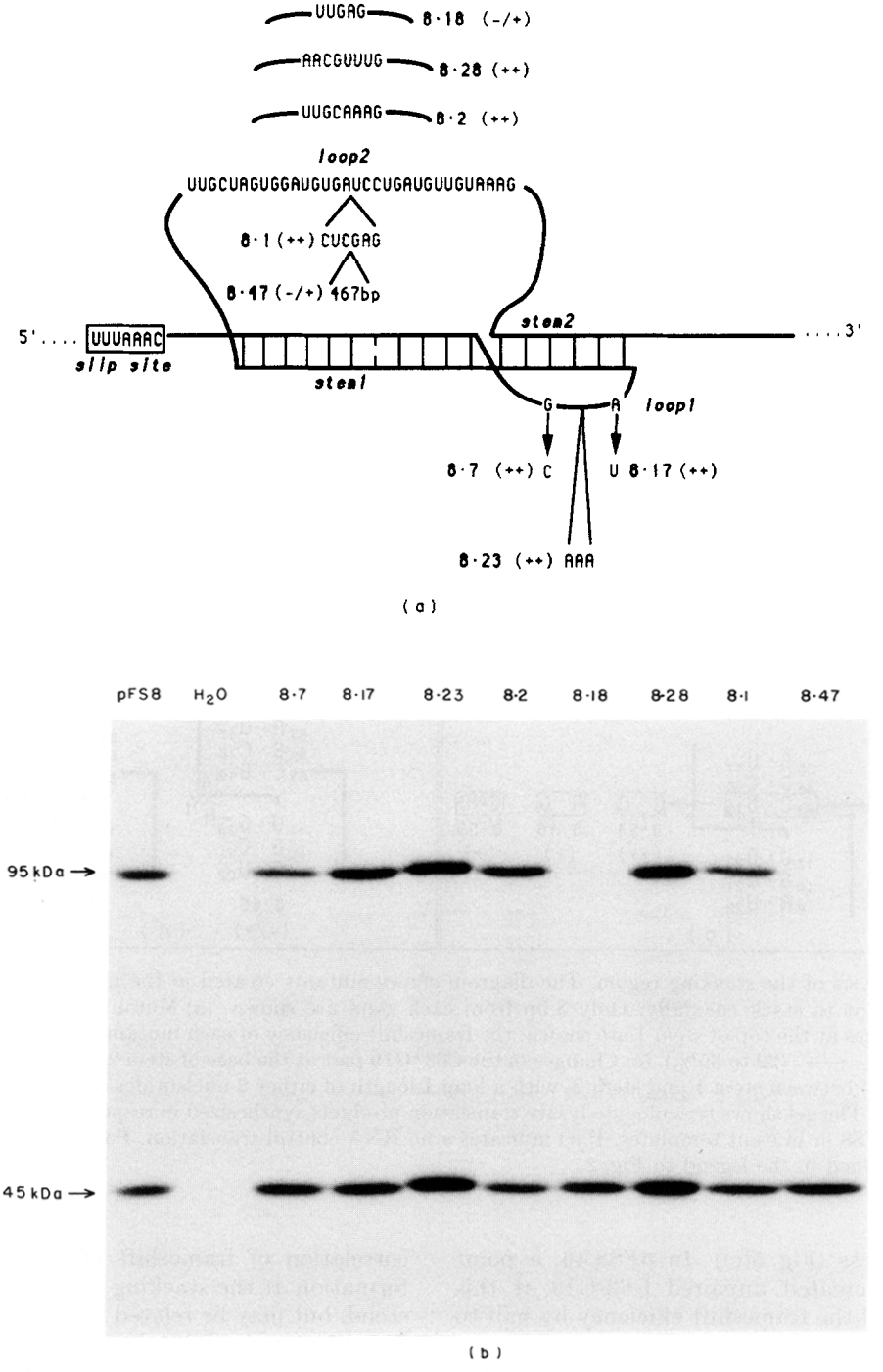

Figure 6. Analysis of the IBV pseudoknot loops. (a) The diagram shows the point mutations, insertions and deletions created within the pseudoknot loops. Mutants are labelled and the frameshift efficiencies indicates as in the legend to Fig. 3. (b) Reticulocyte lysate translation products synthesized in response to mRNAs derived from *Sma*I-digested pFS8 or mutant templates. H_2O indicates a no RNA control translation. Polypeptides were labelled and analysed as described in the legend to Fig. 2.

region had only a limited effect on the frameshift process raised the possibility that direct stacking of the two stems was not an essential requirement of the process. To investigate this, we sought to separate S1 and S2 by inserting three nucleotides between G20 of S1 and G19 of S2 (pFS8.60: AAA insertion) such that stacking could only occur if this insertion was looped out of the helix (Fig. 5(d)). In order to rule out the possibility that the two nucleotides of loop 1 were insufficient to span the increased distance between the two helices, the insertion was also introduced into a variant construct (pFS8.23) in which a three nucleotide insertion (AAA insertion) had previously been made in L1 (and shown to be functional: see Fig. 6, pseudoknot loop analysis) to create pFS8.66. As can be seen in Figure 5(d), the

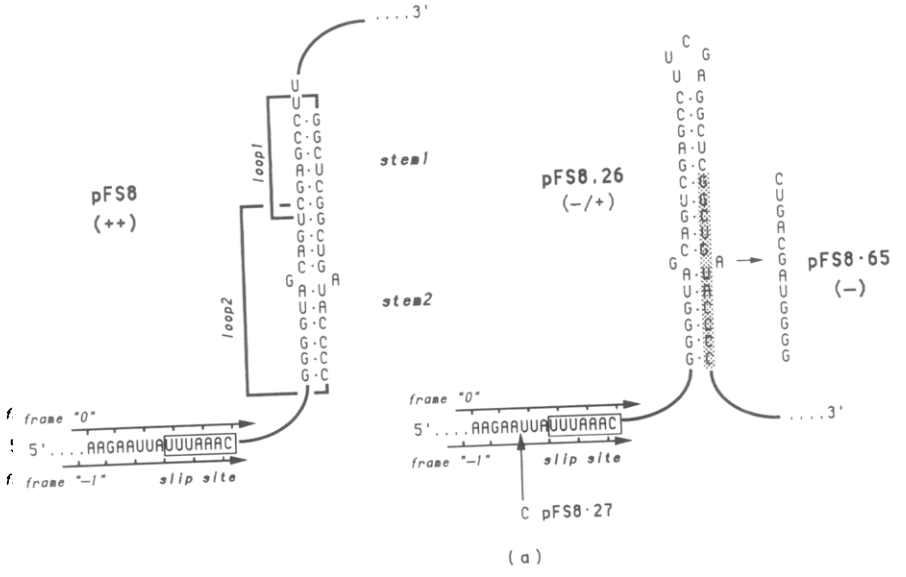

Figure 7. Simple stem–loop experiment. (a) The diagram shows the IBV pseudoknot (pFS8) and a stem–loop construct (pFS8.26) in which the stem nucleotides are of the same length and nucleotide composition as the stacked stems of the pseudoknot. Mutant pFS8.27 has a C nucleotide insertion 3 nucleotides upstream from the slip site. The nucleotides of pFS8.26 that were changed to complementary nucleotides in pFS8.65 are shaded.

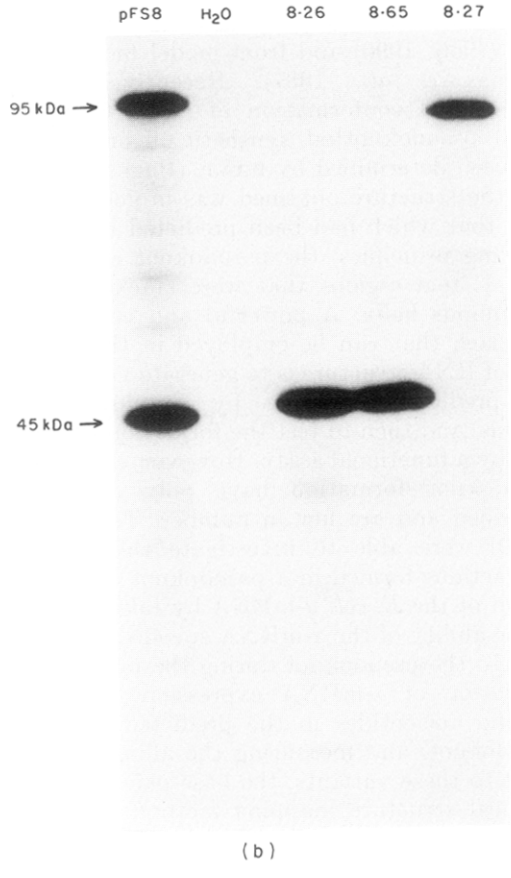

Figure 7. (b) Reticulocyte lysate translation products synthesized in response to mRNAs derived from *Sma*I-digested pFS8 or mutant templates. H_2O indicates a no RNA control translation. Polypeptides were labelled and analysed as described in the legend to Fig. 2.

AAA insertion between S1 and S2 reduced the efficiency of frameshifting in both pFS8.60 (wild-type background, 1%) and pFS8.66 (pFS8.23 background, 5%) supporting the view that "intact" individual stems are not sufficient for high efficiency frameshifting, and that S1 and S2 need be in close proximity.

(ii) *Pseudoknot loop analysis*

On the basis of the known co-ordinates for synthetic RNA double helices (Arnott *et al.*, 1972) and from model building studies of pseudoknots (summarized in Pleij *et al.*, 1985; Dumas *et al.*, 1987) it can be estimated that loop 1 and loop 2 need to span approximately 11 Å and 40 Å, respectively (1 Å = 0·1 nm), in order to bridge the deep and shallow grooves of the RNA helix, respectively, and connect S1 and S2. Assuming that a single nucleotide can span at most a distance of about 8 Å (Saenger, 1984), loop 1 and loop 2 would have to have a minimum length of two and six nucleotides, respectively, in order to connect the helices without distorting the idealized pseudoknot structure. If frameshifting requires only the formation of the correct tertiary structure and does not depend on the presence of particular nucleotides within the loops, altering the sequence and length of the loops within the constraints of the required length ought not to affect frameshifting. This was indeed the case (Fig. 6). Either of the nucleotides proposed to comprise loop 1 could be changed without affecting frameshifting (pFS8.7: G12 to C12; pFS8.17: A13 to U13) and an insertion of an extra three nucleotides in the loop was tolerated (pFS8.23: AAA insertion). As the correct reading frames have to be maintained

in the pFS8 construct in order to monitor the frameshift, only insertions and deletions in multiples of three nucleotides were possible. Thus, we were not able to delete proposed L1 nucleotides. Deletion analysis was, however, possible with loop 2, and the results obtained were in good agreement with the predicted minimal length of this loop (6 nucleotides). In pFS8.2, all but eight nucleotides of L2 were deleted without effect, but in pFS8.18, in which a further three nucleotides were deleted (leaving just 5 nucleotides to span the helix), frameshifting was greatly inhibited. We interpret this observation in terms either of a distortion or an abolition of the pseudoknot structure in this mutant. As was the case of loop 1, the identity of loop 2 nucleotides was unimportant in the frameshift process. When seven of the eight nucleotides of L2 in pFS8.2 were changed to their complements (in construct pFS8.28), frameshifting occurred with wild-type efficiency. (The 8th nucleotide, G62, the last nucleotide in L2 was changed to C62 without dramatic effect in mutant pFS8.34, described earlier.) In principle, there are no upper limits for the length of L2 (or L1) (Pleij *et al.*, 1985) and we have been able to insert six nucleotides corresponding to a unique *Xho*I restriction site into L2 without effect (pFS8.1). However, when a 467 bp DNA fragment derived from the influenza PB2 gene (see Materials and Methods) was inserted in-frame into this *Xho*I site, frameshifting was greatly reduced (pFS8.47; see Fig. 6). The reason for this is not clear, but may be a consequence of competition for the PK2 sequence between the authentic PK4 sequence and sequences within the inserted PB2 segment (see Discussion, below).

(c) *Efficient frameshifting is pseudoknot specific*

The mechanism by which the pseudoknot promotes ribosomal slippage is not known, but one possibility is that the translating ribosome slows or stalls as it reaches this structure and that this can promote slippage at the adjacent slippery site. Consistent with this idea is the observation that the pseudoknot must be correctly positioned with respect to the slip site; insertion or deletion of three nucleotides in the six nucleotide intervening sequence severely inhibits frameshifting (Brierley *et al.*, 1989). The way in which the structure might slow the ribosome is uncertain, but the effect may be simply dependent on the stability of the pseudoknot and the overall energy required to unwind it. If this were the case, it might be expected that a simple stem–loop, comprising an equivalent set of base-pairs, would promote frameshifting as well, if not better, than the pseudoknot if encountered in the same genetic context. To test this, the pseudoknot in pFS8 was deleted and a simple stem–loop structure containing a base-paired stem, of the same length and base-pair composition as the stacked stems of the pseudoknot was introduced at a position six nucleotides downstream from the slippery sequence (pFS8.26). This construct did not display efficient frameshifting, however (0·5%, Fig. 7), suggesting that the pseudoknot has some particular structural feature which is required for the effect. It might be argued that, in this case, the failure to detect high levels of a frameshifted product was due not to a reduction in slippage, but to a general inability of ribosomes to translate through the hairpin structure. However, when the upstream and downstream PB1 ORFs were placed in the same reading frame by introduction of a single nucleotide (C) just upstream from the slippery sequence in a control construct (pFS8.27), only the full-length translation product was observed. That the stem–loop structure actually forms in pFS8.26 is supported by the observation that in an additional construct (pFS8.65), in which the predicted stem–loop was destabilized by a complementary change in the stem, we were not able to detect a frameshift product. Thus, the simple stem–loop cannot functionally replace the pseudoknot.

4. Discussion

Much of the available information regarding the structure of RNA pseudoknots has been derived from direct chemical and enzymatic cleavage analysis of pseudoknot-containing RNA fragments, in combination with phylogenetic sequence comparisons (Pleij, 1990) and from model building studies (Dumas *et al.*, 1987). Recently, the three-dimensional conformation of a short (24 nucleotides), pseudoknotted, synthetic oligoribonucleotide has been determined by n.m.r. (Puglisi *et al.*, 1990) and the structure obtained was in good agreement with that which had been predicted from previous building principles; the pseudoknot contained two helical stem regions that were stacked to form a continuous helix. A powerful and complementary approach that can be employed in the determination of RNA structure is to generate variants within the predicted structure by site-directed mutagenesis, and then to test the formation of the structure by a functional assay. However, such assays for pseudoknot formation have only recently been described and are few in number. Tang & Draper (1989) were able to investigate the base-pairing interactions formed in a pseudoknot located at the 5′ end of the *E. coli* α-mRNA by taking advantage of the ability of the α-mRNA-encoded S4 protein to bind to the pseudoknot during the process of autoregulation of α-mRNA expression. By mutating specific nucleotides in the predicted stems of the pseudoknot, and measuring the affinity of binding of S4 to these variants, the base-pairs predicted by classical structure mapping methods (Deckman *et al.*, 1987) were confirmed and additional interactions discovered. Similarly, Mans *et al.* (1990) have studied the tRNA-like structure of turnip yellow mosaic virus RNA by creating mutations within the structure, and testing for the ability of *E. coli* RNase P to cleave the pseudoknot. Here, we have used this approach to characterize the struc-

tural organization of a naturally occurring pseudoknot in the genomic RNA of a coronavirus. By creating a comprehensive series of changes within the pseudoknot region by mutagenesis and testing for the ability of these mutants to promote ribosomal frameshifting in a cell-free translation system, we have been able to investigate the structure of the IBV pseudoknot, and to compare it with that proposed from the currently available building rules. Moreover, the analysis has important implications for the mechanism of ribosomal frameshifting.

(a) *The structure of the IBV pseudoknot*

The results of our mutagenic analysis clearly demonstrate that the RNA downstream from the slippery sequence folds into a pseudoknot of the hairpin-loop type. The model proposed in Figure 1(b) is largely supported by this analysis; point mutations and complementary and compensatory base changes within the stem regions confirm, in the main, the base-pairs predicted. We have interpreted the mutagenesis results in terms of the expected stability of the two pseudoknot stems in mutant constructs on the basis that neither of the constituent hairpins alone is capable of stimulating high efficiency frameshifting. The point mutational analysis has shown that, with certain exceptions, a single mutation in either stem is sufficient to destabilize the structure and reduce frameshifting. Although S1 is some four base-pairs longer than S2, its stability is undoubtedly reduced by the presence of the central G-A mismatched pair, since frameshifting is improved if this pair is replaced by a Watson–Crick pair. A similar situation is observed in a related coronavirus, mouse hepatitis virus (MHV). The frameshift region of MHV contains a pseudoknot and there is considerable sequence covariance within the predicted base-paired regions of the MHV and IBV pseudoknots (Bredenbeek *et al.*, 1990). In MHV, a G-A mismatched pair is present in S2, and in this case, the helix is four base-pairs longer than S2 of IBV. It appears, therefore, that each stem must possess a certain minimum stability in order that the pseudoknot be functional. In the complementary and compensatory base-change analysis (Fig. 3) it was observed that in a number of double-mutant, pseudo-wild-type constructs, the frameshift efficiency was not fully restored to the wild-type level. The reason for this is uncertain, since in at least one of the cases (pFS8.12, Fig. 3) the stability of S1 in the pseudo-wild-type construct ($-12{\cdot}9$ kcal mol^{-1}), as predicted by the Turner *et al.* (1988) rules, is similar to that predicted for the wild-type S1 hairpin loop ($-12{\cdot}4$ kcal mol^{-1}). Clearly, more information on pseudoknot thermodynamics and, indeed, the three-dimensional structure of pseudoknots is needed before these subtle effects on frameshift efficiency can be fully understood.

When the G·C base-pairs at the start of the S1 helix (G1·C30) were mutated to unpaired nucleotides (Fig. 3), we noted a small reduction in frameshift efficiency and have argued that as the changes are at the end of the helix, they may be expected to have a less dramatic effect on the overall stability of the structure. We cannot rule out the possibility, however, that the observed reduction in frameshifting was not a result of a destabilization of S1 below a critical threshold, but rather a result of increasing the effective slippery sequence-pseudoknot spacing distance by one nucleotide.

Another feature of the pseudoknot where simple interpretations based on expected stability are not possible is the stacking region, where base-pair formation does not appear to be required for efficient frameshifting. In a recent analysis of the stability requirements for RNA pseudoknot formation, Wyatt *et al.* (1990) were able to demonstrate that a synthetic pseudoknotted oligonucleotide (26 nucleotides) was only marginally more stable than either of the constituent hairpins (1·5 to 2 kcal mol^{-1} at 37 °C). Moreover, the increase in enthalpy observed upon forming the pseudoknot was less than the predicted gain; this was probably a consequence of both distortions in stacking between the two stems and to positive enthalpic contributions of the loop regions. A number of the mutations introduced into the stacking region of the IBV pseudoknot gave a frameshift profile that was difficult to interpret in simple terms of base-pair formation; this may well be related to a structural distortion in this region. The availability of a large collection of pseudoknot mutants created in this study should allow an investigation into the energetics of pseudoknot formation by biochemical and biophysical methods.

Our analysis of the pseudoknot loops has revealed that, as was the case with the pseudoknot stems, no specific nucleotides are involved in the frameshift process. Each loop can tolerate a small insertion without influencing frameshifting, and theoretical considerations for the minimal predicted lengths of the loops (Pleij *et al.*, 1985) appeared to be substantiated in the case of loop 2. Although there is no theoretical upper limit for loop length (Pleij *et al.*, 1985) we found that, at least in the one example studied (pFS8.47), loop 2 could not tolerate a large (467 bp) insert. Pseudoknots with large loop sizes have been proposed in the structure of RNase P (James *et al.*, 1988) and in 16 S rRNA, where most of the rRNA secondary structure is contained within loop 2 of a pseudoknot (Moazed & Noller, 1987). In addition, a long-range pseudoknot has been proposed in a structural model for mitochondrial group I introns (Davies *et al.*, 1982). In the case of pFS8.47, we have speculated that pseudoknot formation may be prevented by competition between the authentic PK4 sequence and alternative stretches within the inserted information. A second possibility concerns the rate of translation of L2 in this mutant. Having melted S1, the elongating ribosome would take considerably longer to translate L2 in pFS8.47 than in the wild-type situation and this may impede refolding of the pseudoknot to such an extent that the next ribosome

translating the mRNA may not encounter a pseudoknot and would therefore not frameshift. It maybe significant that in the pseudoknots predicted to form at ribosomal frameshift sites in other systems, the longest loop length is only 69 nucleotides, and in the vast majority of cases, the loops are considerably shorter (ten Dam *et al.*, 1990).

(b) *Implications for ribosomal frameshifting*

The mutational analysis so far indicates that the formation of the pseudoknot is critical for high efficiency frameshifting and has helped to provide a detailed picture of the structure involved. A summary of the changes which have been made is shown in Figure 8. The "non-essential" nucleotides of loop 2 (defined by mutant pFS8.2) are not included in the Figure, hence the 44 nucleotides depicted represent our estimate of the minimal number required for the formation of the IBV pseudoknot. As each of the nucleotides has been changed to an alternative nucleotide in one or more of the various mutant constructs, the particular nucleotide sequence of the pseudoknot *per se* is not important in the frameshift process. Nevertheless, efficient frameshifting depends upon the formation of the correct structure

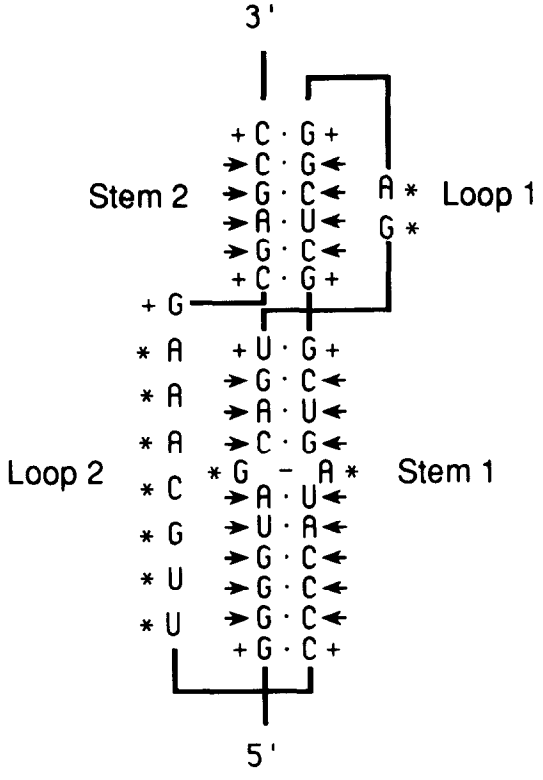

Figure 8. Summary of changes in the pseudoknot region. The "non-essential" nucleotides of loop 2 (see the text) are not included in the Figure; the 44 nucleotides depicted represent our estimate of the minimal number required for the formation of the IBV pseudoknot. *, indicates change not inhibitory; +, change giving intermediate phenotype; →, inhibitory change but can be compensated by complementary change elsewhere.

downstream from the slippery site; the pseudoknot could not be functionally replaced by a stem–loop structure of a similar or greater predicted stability. It remains unclear how the pseudoknot actually causes ribosomal slippage. The most likely explanation is ribosomal pausing (Jacks *et al.*, 1988), and we are investigating possible mechanisms within the framework of two models. In the first model, pausing occurs as a result of a direction interaction of a ribosomal protein(s) or additional component(s) of the translation apparatus with the pseudoknot. The control of mRNA translation through specific recognition by proteins of pseudoknotted RNAs has been documented in a number of prokaryotic systems (McPheeters *et al.*, 1988; Tang & Draper, 1989; Philippe *et al.*, 1990), although in these cases, the proteins concerned bind to, or induce the formation of, pseudoknots in the 5′ non-coding regions of mRNAs and are involved in autoregulation. In this investigation we have been unable to demonstrate any requirement for specific pseudoknot nucleotide sequences in the frameshift process that could not be interpreted in terms of the pseudoknot structure. The inability of the simple stem–loop structure (pFS8.26) to direct efficient frameshifting, however, supports the view that some unique feature of the pseudoknot plays a role in the process and the possibility that sequence-independent recognition of, for example, a pseudoknot loop(s) by a protein factor cannot be discounted. Autogenous regulation of gene *32* protein expression in bacteriophage T4 involves an interaction between the protein and the loop(s) of a pseudoknot in the 5′ non-coding region of the gene *32* mRNA (McPheeters *et al.*, 1988). In the case of the IBV pseudoknot, eight nucleotides would have to suffice for such an interaction with loop 2, since 24 nucleotides of this loop could be deleted without effect (pFS8.2).

An alternative and attractive possibility for a pausing mechanism is that RNA pseudoknots may present an unusually resistant structure to a ribosome-associated RNA helicase activity functioning to promote local unwinding of mRNA during elongation. The lack of frameshifting in the simple stem–loop construct is significant in that this structure can be predicted to be much more stable than the pseudoknot, and so a simple energetic barrier to translation is not sufficient for the effect. This hypothesis is based upon the observation of Wyatt *et al.* (1990) that small pseudoknots are only marginally more stable energetically than either of their constituent hairpins, and likely to be less stable than a hairpin-loop with a comparable number of base-pairs.

It should be possible to test these hypotheses through *in vitro* analysis of the interaction of short, pseudoknotted RNAs and components of the translation apparatus.

This work was supported by an AFRC link grant, LRG 171, awarded to S.C.I. We are grateful to Cornelis Pleij, Edwin ten Dam and Jan van Duin for stimulating discussion.

References

Arnott, S., Hukins, D. W. L. & Dover, S. D. (1972). Optimised parameters for RNA double-helices. *Biochem. Biophys. Res. Commun.* **48**, 1392–1399.

Birnboim, H. C. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.* **7**, 1513–1523.

Bredenbeek, P. J., Pachuk, C. J., Noten, A. F. H., Charité, J. Luytjes, W., Weiss, S. R. & Spaan, W. J. M. (1990). The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59; a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. *Nucl. Acids Res.* **18**, 1825–1832.

Brierley, I., Boursnell, M. E. G., Binns, M. M., Bilimoria, B., Blok, V. C., Brown, T. D. K. & Inglis, S. C. (1987). An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO J.* **6**, 3779–3785.

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

Boursnell, M. E. G., Brown, T. D. K., Foulds, I. J., Green, P. F., Tomley, F. M. & Binns, M. M. (1987). Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. *J. Gen. Virol.* **68**, 57–77.

Craigen, W. J. & Caskey, C. T. (1987). Translational frameshifting: where will it stop? *Cell*, **50**, 1–2.

Davies, R. W., Waring, F. B., Ray, J. A., Brown, T. A. & Scazzocchio, C. (1982). Making ends meet: a model for RNA splicing in fungal mitochondria. *Nature (London)*, **300**, 719–724.

Deckman, I. C., Thomas, M. S. & Draper, D. E. (1987). S4-α-mRNA translation regulation complex 1. Thermodynamics of formation. *J. Mol. Biol.* **196**, 313–332.

Dotto, G. P., Enea, V. & Zinder, N. D. (1981). Functional analysis of bacteriophage f1 intergenic region. *Virology*, **114**, 463–473.

Dumas, P., Moras, D., Florentz, C., Geige, R., Verlaan, P., van Belkum, A. & Pleij, C. W. A. (1987). 3-D graphics modelling of the tRNA-like 3′ end of turnip yellow mosaic virus RNA: structural and functional implications. *J. Biomol. Struct. Dynam.* **4**, 707–728.

Hames, B. D. (1981). An introduction to polyacrylamide gel electrophoresis. In *Gel Electrophoresis of Proteins–A Practical Approach* (Hames, B. D. & Rickwood, D., eds), pp. 1–91, IRL Press, Oxford.

Jacks, T. & Varmus, H. E. (1985). Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. *Science*, **230**, 1237–1242.

Jacks, T., Townsley, K., Varmus, H. E. & Majors, J. (1987). Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus *gag*-related polyproteins. *Proc. Nat. Acad. Sci., U.S.A.* **84**, 4298–4302.

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

James, B. D., Olsen, G. J., Liu, J. & Pace, N. R. (1988). The secondary structure of ribonuclease P RNA, the catalytic element of a ribonucleoprotein enzyme. *Cell*, **52**, 19–26.

Joshi, R. L., Joshi, S., Chapeville, F. & Haenni, A. L. (1983). tRNA-like structures of plant viral RNAs: conformational requirements for adenylation and aminoacetylation. *EMBO J.* **2**, 1123–1127.

Krieg, P. A. & Melton, D. A. (1984). Functional messenger RNAs are produced by SP6 *in vitro* transcription of cloned cDNAs. *Nucl. Acids Res.* **12**, 7057–7071.

Kunkel, T. A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Nat. Acad. Sci., U.S.A.* **82**, 488–492.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbour, NY.

Mans, R. M. W., Guerrier-Takada, C., Altman, S. & Pleij, C. W. A. (1990). Interaction of RNase P from *Escherichia coli* with pseudoknotted structures in viral RNAs. *Nucl. Acids Res.* **18**, 3479–3487.

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

Moazed, D. & Noller, H. F. (1987). Interaction of antibiotics with functional sites in 16 S ribosomal RNA. *Nature (London)*, **327**, 389–394.

Moore, R., Dixon, M., Smith, R., Peters, G. & Dickson, C. (1987). Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: two frameshift suppression events required for translation of *gag* and *pol*. *J. Virol.* **61**, 480–490.

Philippe, C., Portier, C., Mougel, M., Grunberg-Manago, M., Ebel, J. P., Ehresmann, B. & Ehresmann, C. (1990). Target site of *Escherichia coli* ribosomal protein S15 on its messenger RNA. *J. Mol. Biol.* **211**, 415–426.

Pleij, C. W. A. (1990). Pseudoknots: a new motif in the RNA game. *Trends Biochem. Sci.* **15**, 143–147.

Pleij, C. W. A. & Bosch, L. (1989). RNA pseudoknots: structure, detection and prediction. *Meth. Enzymol.* **180**, 289–303.

Pleij, C. W. A., Rietveld, K. & Bosch, L. (1985). A new principle of RNA folding based on pseudoknotting. *Nucl. Acids Res.* **13**, 1717–1731.

Puglisi, J. D., Wyatt, J. R. & Tinoco, I. (1990). Conformation of an RNA pseudoknot. *J. Mol. Biol.* **214**, 437–453.

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

Rietveld, K., Pleij, C. W. A. & Bosch, L. (1983). Three dimensional models of the tRNA-like 3′ termini of some plant viral RNAs. *EMBO J.* **2**, 1079–1085.

Rietveld, K., Linschooten, K., Pleij, C. W. A. & Bosch, L. (1984). The three dimensional folding of the tRNA-like structure of tobacco mosaic virus RNA. A new building principle applied twice. *EMBO J.* **3**, 2613–2619.

Russel, M., Kidd, S. & Kelley, M. R. (1986). An improved filamentous helper phage for generating single-stranded plasmid DNA. *Gene*, **45**, 333–338.

Saenger, W. (1984). In *Principles of Nucleic Acids Structures*, chap. 15, Springer Advanced Texts in Chemistry, Springer-Verlag, New York.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Nat. Acad. Sci., U.S.A.* **74**, 5463–5467.

Tang, C. K. & Draper, D. E. (1989). Unusual mRNA pseudoknot structure is recognised by a protein translational repressor. *Cell*, **57**, 531–536.

ten Dam, E. B., Pleij, C. W. A. & Bosch, L. (1990). RNA pseudoknots: translational frameshifting and read-through on viral RNAs. *Virus Genes*, **4**, 121–136.

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

van Belkum, A., Abrahams, J.-P., Pleij, C. W. A. & Bosch, L. (1985). Five pseudoknots are present at the 204 nucleotides long 3′ non-coding region of tobacco mosaic virus RNA. *Nucl. Acids Res.* **13**, 7673–7686.

Wyatt, J. R., Puglisi, J. D. & Tinoco, I. (1990). RNA pseudoknots; stability and loop size requirements. *J. Mol. Biol.* **214**, 455–470.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13 mp18 and pUC19 vectors. *Gene*, **33**, 103–119.

Young, J. F., Desselberger, U., Graves, P., Palese, P., Shatzman, A. & Rosenberg, M. (1983). Cloning and expression of influenza virus genes. In *The Origin of Pandemic Influenza Viruses* (Laver, W. G., ed.), pp. 129–138, Elsevier Science, Amsterdam.

Edited by M. Yaniv