Phylogenetic and physical analysis of the 5' leader RNA sequences of avian retroviruses

Perry B.Hackett*, Mark W.Dalton¹, Darrin P.Johnson¹ and Robert B.Petersen⁺

Department of Genetics and Cell Biology and Institute of Human Genetics and ¹Molecular Biology Computing Center, University of Minnesota, St Paul, MN 55108-1095, USA

Received June 14, 1991; Revised and Accepted November 25, 1991

ABSTRACT

A study of the secondary structures of the 5'-leader RNA sequences of avian leukosis/sarcoma viruses was conducted using phylogenetic sequence alignment, theoretical structures calculated from base-pairing interactions involving the calculated minimal ΔG values, and RNaseT1 sensitivity. The results suggest that all of the avian retroviral RNA leaders may be able to adopt similar conformations. Open reading frames in the leader RNAs may be positioned to facilitate viral activities such as translation and packaging of the genomic RNA into virus particles.

INTRODUCTION

The full-length RNA of every animal retrovirus has four roles during the viral life cycle: 1) replication—the RNA is a template for RNA-dependent DNA synthesis by reverse transcriptase; 2) splicing—the RNA is the primary transcript for the subgenomic mRNAs; 3) translation—the RNA is the message for the *gag* and *pol* genes in replication-competent viruses and for fusion gene products in defective, transforming viruses; and 4) virion formation—the RNA is packaged by the *gag* proteins to initiate virus production. The last three processes persist for many cell generations in virus-infected cells. For each of these operations, either viral or cellular proteins and RNA interact with the full length RNA to catalyze the roles outlined above. All of the functions described utilize regions in the 5' leader RNA that precede the *gag* gene sequence, the first conventional gene in retroviral genomes.

The avian retroviruses have an additional complexity. Their 5'-leader sequences all have three short open reading frames (ORFs) that are conserved with respect to their positions in the leader RNA, their lengths, and the nucleotide contexts (1) of their AUG initiation codons. However, the encoded amino acid sequences in the ORFs are variable (2). Since the entire 5' leader sequence is contained on all spliced and unspliced avian retroviral mRNAs (3), we have studied the effects of the ORFs and other 5'-leader sequences on translational regulation and viral propagation of one avian retrovirus RNA, Rous sarcoma virus,

strain SR-A(SF) (3). On the basis of ribosome binding (4–6), peptide isolation (2), and mutation of the leader ORFs (6; Moustakas et al., submitted) we have shown that the leader ORFs affect translation of downstream genes and appear to be translated.

In general, short ORFs in leader sequences of other animal and viral mRNAs attenuate translation of downstream genes. In yeast there is an exception to this general rule; while most upstream ORFs lower translation of downstream genes, two very short ORFs in the GCN4 mRNA of *S. cerevisiae* act to enhance translation while two others repress translation (7). Recently we have shown that the ORFs in the RSV 5'-leader sequence can also enhance (Moustakas et al., submitted) as well as repress translation (8).

Putative RNA secondary structures (9) with ΔG 's greater than –30 to –50 kcal/mole correlate with inhibition of ribosomal subunit scanning and repression of mRNA translation (10,11). Studies of the translational efficiencies of globin (12) and insulin (13) mRNAs, alfalfa mosaic virus RNAs (14,15), and SV40 mRNAs (16), suggest that RNA structure modulates translation. Protein binding to mRNA can affect translation, as has been demonstrated for ferritin mRNA (17). RNA structure has particular importance in the expression of retroviral RNAs at the steps of *pol* gene translation by frameshifting (8) and Tat regulation of HIV gene expression (11). Thus, the available evidence suggests that the three ORFs and secondary structure may regulate translation of avian retroviral RNAs.

We have characterized the secondary structure of the 5'-leader RNAs of Rous sarcoma virus (RSV), and other avian retroviruses, by 1) using phylogenetic comparisons of their sequences, 2) calculating the lowest theoretical energy structures for each RNA to determine a consensus folding pattern, and 3) mapping RNaseT1 sensitive sites to determine the validity of the theoretical structures. The results suggest possible roles for the ORFs in the regulation of viral replication and gene expression.

MATERIALS AND METHODS

Computer alignments of primary structures

The Intelligenetics Genalign program, with the Multiple Sequence Algorithm (19) was used on sequences with homology to the SR-

* To whom correspondence should be addressed

⁺ Present address: Institute of Pathology, Case Western Reserve University, Cleveland, OH 44106, USA

A(SF) strain of avian retrovirus (GenBank release 45). Table 1 lists the viral RNA sequences that were analyzed in this study and their GenBank identification numbers.

Calculations of secondary structure

RNA secondary structure was predicted using an optimized version (D.J., unpub.) of the Zuker algorithm that displays optimal and suboptimal folding patterns (20). The program was run on a Silicon Graphics Iris 4D/240S using the Freier et al. thermodynamic parameters and Jaeger loop parameters for predicting secondary structures (21).

RNA isolation and analysis

RNA was obtained from chick embryo fibroblasts infected with the SR-A(SF) strain of RSV as previously described (3). Essentially, the procedure we used is a novel application of our ribosome binding site protocol (4,5). 70S viral RNA was denatured with 10mM methyl mercuric hydroxide, which was subsequently quenched with dithiothreitol to allow renaturation (22) as judged by the electrophoretic mobility of renatured rRNA, and resuspended in 100mM NaCl, 10 mM Na citrate (pH 5.0) for RNase T1 digestion as previously described (29). Viral RNA was digested with RNase T1 (Pharmacia-P.L. Biochemicals)

Figure 1. Alignment of the 5'-leader RNA sequences of avian retroviruses. Abbreviations for the viral sequences are in Table I. The nts of the consensus sequence are capitalized if maintained in all of the viruses and in lower case if not uniformly conserved. Dashes (—) indicate gaps in the alignment of the sequences. The ORFs are boxed; the PBS is shown by brackets; the approximate boundaries of the sequence required for RNA packaging is indicated by the dotted line, and the double arrows around nt 350 indicate a duplication of a short stretch of RNA sequence, and the 5' end of the gag gene is shown by a long arrow.

0.01, 0.001, and 0.0001 U/ml for 10 min at 0°C, digested with Proteinase K, and extracted with phenol and chloroform. The resulting RNA fragments were labelled at their 5'-ends using polynucleotide kinase and [γ - 32 P]ATP and hybridized to cDNA from particular regions of the RSV leader. The cDNA subclones were covalently coupled to diazobenzoyloxymethyl-paper which permitted hybrid-selection of specific RSV RNA fragments. The hybrid-selected RNA fragments were separated by electrophoresis through a denaturing sequencing gel and recovered for detailed RNaseT1 analysis as described (5).

RESULTS AND DISCUSSION

Sequence comparison of the 5' RNA leaders of avian retroviruses

Phylogenetic comparisons of primary sequences have been shown to be the most reliable method for determining secondary RNA structures (23,24). Fig. 1 shows the alignment of the 13 different retroviruses used in this study and the derived consensus sequence. Landmarks for sequence alignment included the three conserved ORFs, the initiation site for the *gag* gene, and the tRNA^{Trp} Primer Binding Site (PBS). Additionally, an approximate region required for virion RNA packaging, Ψ , is shown. The alignment is modified from that generated by the computer algorithm which introduced many more gaps in the sequence than are shown. Inspection of the computer alignments revealed several sites where adjustments based on landmarks or allowance of single nucleotide mismatches brought the sequences into greater alignment.

The percent similarity of each of the 5'-leader RNA sequences to the consensus sequence was calculated (Table I). They show that the viral sequences varied from the consensus by 3 to 17%. Comparison of the sequences reveals a high degree of conservation of the three leader ORFs in terms of position and length but not in the encoded peptides, absolute conservation of the primer binding site, and a variable region between ORF-3 and the initiation site for *gag*. A duplication of 13 nucleotides in the AMV and RAV-0 RNA leader has been noted previously (25) and is indicated by the arrows in Fig. 1 between consensus nucleotides 350–370. Nucleotide changes occur throughout the sequences of all of the viruses.

Folding of the avian retrovirus leader RNA

Owing to the constraints of the algorithms for RNA folding (9), accurate folding of an RNA sequence generally requires that the sequence be longer than the region of interest in order to ensure

that suboptimal interactions are not forced. We examined the theoretical secondary structures of all of the avian retroviral RNA leader sequences listed in Table I. We used the Zuker (20) folding algorithm that permits analysis of suboptimal folding patterns as well as that with the lowest ΔG and incorporated the latest energy parameters (21). The structures were plotted using the DRAW program of Duvall (Hackett et al., submitted) that identifies ORFs and the relative strengths of AUG initiation codons according to the Kozak rules (10).

Figure 2a and b shows two folding patterns of the first 499 nt of the consensus sequence. The structure in panel a has the lowest free energy and shows where divergences of the SR-A sequence have occurred. Fig. 2b shows an alternative structure for the consensus leader sequence; it has a calculated ΔG 0.7% higher than the structure in (a) and includes an upaired 5' terminus and an extended PBL loop. The conformation in Fig 2c represents the lowest free energy folding pattern of the RNA from the SR-A strain of RSV which we have examined in detail (see below), and structure (d) is a suboptimal pattern that more closely matches the consensus structure in Fig. 2a. A summary of the folding patterns of the first 499 nt of the ten viruses examined is given in Table II. In general the consensus sequence appears to be representative of the lowest energy secondary structures of the avian retroviruses, within 3 kcal/mole (1%). In particular, the long arm designated PBL (Primer Binding Loop), has a $\Delta G = -47$ to -60 kcal/mole, depending on where the terminus of the stem is positioned. PBL contains most of ORF1, ORF2 and the tRNA^{Trp} primer binding site and is common to all viral leader RNA structures. The PBL stem length varies in different structures (compare Fig 2a and 2b; Table II) which probably reflects alternative conformations the RNA may assume in its different roles.

The SR-A RNA has the greatest primary sequence divergence from the consensus. Fig. 2a shows the sites of variation between the consensus and the SR-A strain. In several cases there is co-variation of base pairs such that either one base pair in a stem is replaced by another or G:U (G:C) pairing is replaced by G:C (G:U) bonding. In most cases the secondary structure predicted by the consensus sequence is not grossly altered by the different nucleotides in the SR-A sequence. For instance, the unpaired loop in L4 of the consensus (-313 kcal) folding is extended but is pinched off by two base pairs in the SR-A (-303 kcal) structure even though the difference is a single A:U pair that is unstable under physiological conditions. Two alternative structures for the

TABLE I

Virus	i.d.	Length	%Consensus	Ref
1) SR-A(SF)	ALRD1	1000 nt	83%	30
2) UR2	ACSUR2CG	1000 nt	92%	31
3) CT-10	YOO302	1000 nt	97%	32
4) Y73	ACSY73SG	1000 nt	96%	33
5) PrC	ALRCG	1000 nt	97%	34
6) FSV	ACF	1000 nt	96%	35
7) AEV	ACBAEV5TR	1000 nt	94%	36
8) MC29	ACM	1000 nt	96%	37
9) AMV	ACMHBIPB	1000 nt	92%	38
10)RAV-2	ALEENV	1000 nt	89%	39
11)RAV-0	ALELTR0	260 nt	97%	40
12)EV-2	AC2EV2LTR	536 nt	96%	26
13)EV-1	AC2EV1LTR	536 nt	96%	26

The sequences of the 13 avian leukosis/sarcoma viruses used in this study. i.d., GenBank identification code; Length, length of sequence examined for sequence comparison and folding (sequences 1-10) or just sequence comparison (11-13); % consensus, homology to the consensus leader sequence (Fig. 1), from the cap site to the initiation site for Pr76*gag*.

TABLE II

Viruses	L1	PBL	L2	Q3	L3	L4	L5	G1	G2	G3
RSVCON (-315)	+	+	+	+	+	+	+	+	+	+
RSVCON (-313)	a	a	a	+	+	+	+	+	+	+
SR-A (-303)	+	+	+	b	+	+	a	b	+	b
SR-A (-300)	+	+	+	+	+	+	+	+	+	b
UR2	a	+	+	b	+	+	+	+	-	b
CT-10	+	a	a	b	+	+	+	b	+	-
Y73	a	a	+	-	+	+	+	+	+	+
PrC	+	+	+	+	+	+	-	-	+	b
FSV	+	+	+	+	+	+	+	+	+	-
AEV	a	+	+	+	+	+	+	+	+	-
MC29	+	+	+	+	+	+	+	+	+	-
AMV	-	a	+	b	+	+	+	b	+	-
RAV-2	+	+	+	+	+	+	+	+	+	+

Summary of folding results for avian leukosis/sarcoma viruses. Consensus [RSVCON(-315 kcal/mole)] and an alternative folding of the consensus sequence [RSVCON(-313 kcal/mole)], (see Fig. 2a,b), are compared to two alternative folding patterns for the SR-A strain, shown in Fig. 2c/d. +, structure same as RSVCON(-315); a, alternative structure similar to RSVCON(-313) -type folding pattern; b, SR-A specific structure; +*, an alternative structure within 2% of minimal energy structure exists that is the same as RSVCON(-315).

Figure 2. Computer generated RNA secondary structures of the first 499 nt of avian retroviral RNAs. ORFs in the leader RNA sequence and the initiation codon for the *gag* gene are boxed. Structural motifs include: L1-L5, Leader Loops 1-5; PBL, tRNA^{Trp} Primer Binding Loop; O3, ORF-3 Loop; G1-G3; *gag* Loops 1-3; I; *gag* initiation codon; 5'ss, 5' splice site. The 5' end of each sequence is marked with a large 5'. Dots are placed every 10 nt and each fifty is marked. Identified sites of RNaseT1 cleavage are identified in panels (c) and (d) by arrows. a) Structure for the consensus leader RNA sequence shown in Fig. 1. Lower case letters at certain positions indicate replacement nts in the SR-A sequence; -, deleted base (bases 333-341 are deleted in SR-A); /, added base. b) Alternative consensus RNA structure which has an unpaired 5' terminus and a longer PBL structure. c) Proposed structure of the lowest calculated ΔG value for the RSV strain SR-A sequence. d) Alternative SR-A leader RNA structure which conforms to the consensus structure in panel (a).

SR-A RNA leader are shown in Fig. 2c,d; the less stable structure was selected to show that a conformation around the ORF-3 region (ORF-3 Loop) similar to that of the consensus sequence could be obtained with less than a 1% variation in overall ΔG . In general the same regions remain unpaired in both structures although pairing of the 5' end may be broken during some events, *e.g.*, translation. Accordingly, we initiated preliminary tests of the theoretical SR-A structures to determine their validity.

RNaseT1 sensitivity of SR-A RSV RNA

RNaseT1 cleaves RNA after unpaired G-residues and can be used to determine the sites of unpaired G's in folded RNA. SR-A RNA was subjected to mild hydrolysis with RNaseT1 and the viral RNA fragments were analyzed by a variation of the ribosome mapping procedure (5). The RNA fragments from RNase T1

digestion were labelled at their 5'-ends with [³²P], isolated by selective hybridization to cDNA immobilized on filters, eluted from the filter, and separated by electrophoresis through a polyacrylamide sequencing gel. Selected fragments were further analyzed by partial RNaseT1 digestion which produced a ladder that marked the positions of each G-residue in the RNA sequence. Fig. 3A is a map depicting the cDNAs used for RNA fragment isolation. Fig. 3B shows the results of the first round of electrophoresis, and Fig. 3C displays the results of the second round of electrophoresis for an analysis of several of the intense bands from the first gel. Fig. 4 shows the positions of several of the identified bands in the gels. The sizes of the bands were determined using a poly(A) ladder with adjustments required to compensate for the differential migration between the homopolymer and the heteromeric RNA fragments (4). The initial

Figure 3. RNaseT1 sensitivity analysis. SR-A RNA was mildly hydrolyzed with RNaseT1 and the fragments were hybridized to cloned DNA fragments specific to subregions of the 5' region of the viral RNA and electrophoresed on polyacrylamide sequencing gels. a) Map of the 5'-end of RSV RNA with the nucleotide scale at the top, the three short ORFs and the beginning of the *gag* gene in stippled boxes with the AUG codons indicated by circles, and the subclones used for selections of fragments identified on the bottom. b) Distribution of fragments resulting from mild hydrolysis by RNaseT1. Each band is identified by the clone used for selection (lane numbers) and a letter to the left. Sizes of the bands were determined using a poly(A) ladder (not shown); a scale is on the left. c) Partial RNaseT1 digestion products of RNA fragments isolated from the gel shown in panel b. The lane headings designate the bands from panel b. A, poly(A) markers.

TABLE III		
Site	(-303)	(-300)
24/25	+/-	+/-
29	+	+
93	+	+
112	+	+
184/186	-	+
193	+	+
207	-	+
256	+/-	+/-
298/300	+	+
318/322	+	+
346	+	+
374	+	+
398	+	+
440	+	+

Summary of RNaseT1 sensitive sites. The sensitive sites identified in Fig. 3 were applied to the folded SR-A RNAs depicted in Fig 2c and 2d. +, site is unpaired in structure; -, site is paired in structure; +/-, site is weakly paired in structure.

RNase digestion was intended to identify some of the most sensitive sites in the native RNA conformation. Due to the high G-content (40%) of the sequence in subclone 4 and the dense packing of the bands, only three of the RNase T1-sensitive sites in this region were analyzed further. In some instances where only one major band appeared on the initial gel further resolution indicated that multiple bands migrated closely together, e.g., band 2b which separates into two sequences in Fig. 3b. Generally multiple bands were nested as a result of adjacent G residues

Figure 4. Detailed map of the oligoribonucleotide fragments generated by partial RNaseT1 cleavage of selected RNA fragments. Six subclones 1-6 of the SR-A sequence used for fragment selection (Fig. 3a,b) are shown. The ORFs are boxed, the RNA fragments are indicated by the dashed lines, and the vertical lines indicate sites of partial cleavage by RNase T1.

at one terminus of the fragment. Further clues for placement of the fragments could often be obtained by noting the relative intensities of those bands that come from fragments which overlap the boundaries of the cDNA clones.

The results of the RNaseT1 mapping were applied to the theoretical secondary structures shown in Fig. 2c/d. and are evaluated in Table III. Two considerations were taken into account in reading the data in Fig. 3c. First, sequences rich with runs of G's have compression sites that altered the mobility of the bands, and second, several of the bands expected in the RNase mapping lanes were weak because of base pairing that can occur within the fragment. We note that although both SR-A structures predict most of the identified RNase T1-sensitive sites the conformation in Fig. 2d, which is similar to the consensus structure (2a), is slightly more compatible with the identified RNaseT1-sensitive sites. The finding that most of the unpaired G's are predicted by both structures may be in part due to oscillations between the conformations. In both structures a number of unpaired G residues are not as sensitive as predicted (e.g., nts 327-330 in both structures). This may be due to either tertiary folding and/or generation of small fragments not selected for second stage RNase mapping in this preliminary study. Additionally, G(24 and G(25) which appear to be strongly base-paired in Fig. 2 are probably in a stem of slightly less energy than suggested since there are just 3 unpaired residues in the loop region whereas most loops require four or more unpaired bases. Thus, the A(19):U(23) base pair is probably broken and the stem structure containing G(24 and G(25) is probably semi-stable as noted in Table III.

Further analysis is also required to determine more accurately the structures of the regions between ORF-3 and the *gag* gene. The variation in this region (Table II) which is involved in genome packaging suggests that the RNA is prone to assuming alternative conformations in this region. Just downstream of the *gag* AUG initiation codon, where splicing of the subgenomic mRNAs occurs, is another region of instability (Table II) that may reflect the need for a single-stranded conformation for splicing factors. The structure in panel 2c shows both the *gag* AUG initiation codon and the splice site are unpaired whereas the structure in 2d has a 5' splice site that is masked in a stem conformation. In conclusion, the small differences of about 1% in ΔG values by the different conformations suggest that the avian retroviral RNAs may not have a *secondary structure* but instead may have several alternative structures which may be stabilized by protein/factor interactions for particular viral functions such as RNA packaging and splicing. These multiple considerations must be taken into account in phylogenetic analysis of mRNAs. For rigorous phylogenetic analysis, there should be dual base changes such that alteration of the primary sequence at two residues in a stem will be compensatory to retain structure. In mRNAs the primary sequence is often quite important thus removing the possibility of finding compensatory changes in all loops. This limitation becomes major when the mRNA has multiple functions; retroviral RNA such as RSV mRNA is not only a messenger, but also a template for reverse transcription, a substrate for packaging, and a substrate for RNA splicing. Each function requires specific signals of primary sequence and/or secondary structure.

Consequences of the derived RNA secondary structures

The models presented above should be valuable for suggesting mechanisms to explain previously observed phenomena, and for designing experiments involving RNA-protein interactions that are responsible for every step of the viral life cycle. Katz (27) noted that mutations of A(373) and A(376), conserved in all the avian retroviruses, to G and U respectively, in the *ev-1* sequence lowered the infectivity of the virus by 12-fold. Examination of the structures in Fig. 2a suggests that the mutations would destabilize secondary structure around the *gag* initiation codon and the putative region involved in virion RNA packaging. The structures shown in Fig. 2 conform to the models of the PBL proposed by others (27,28) and fit with their data. This motif may be critical for reverse transcription of avian retroviral RNA.

Tentative structures derived from theoretical and experimental considerations suggest that the structure of the leader RNA for avian retroviruses is independent of the the rest of the RNA. This result is important for investigations of translational control and other regulatory roles of the leader RNA. Specifically, the same structural motifs required for reverse transcriptase, splicing factors, and packaging agents must also be traversed by scanning ribosomes whose progress is inhibited by secondary structures with energies above -30 to -40 kcal/mole (1,29). The identified structures should attenuate or abolish translation of the retroviral RNA. However, ribosomes are not stopped by secondary structure, which might explain the conservation of ORFs in the leader. The ORFs, which are not sufficient to inhibit downstream translation, may act positively by maintaining the 5'-proximal sequence in an open configuration, enabling ribosomes to scan the entire RNA leader up to the *gag* initiation codon (Moustakas and Hackett, submitted). In this regard, the structures presented in Fig. 2 indicate that the conserved ORFs are positioned in

regions rich in secondary structure that are inhibitory to scanning ribosomes. Clearly these results and the suggested consequences of the RNA structures require further investigation at both the structural and functional levels.

ACKNOWLEDGEMENTS

We thank M. Zuker, C. Hensel, A. Moustakas, and T. Sonstegard for discussions and reviewing the manuscript and an anonymous reviewer for suggestions. This work was supported by NIH grant CA 43881 and ACS grant INC/NP-790.

REFERENCES

1. Kozak, M. (1988) *Mol. Cell. Biol.* **8**, 2737-2744.
2. Hackett, P.B., Petersen, R.B., Hensel, C.H., Albericio, F., Gunderson, S.I., Palmenberg, A.C., and Barany, G. (1986) *J. Mol. Biol.* **190**, 45-57.
3. Hackett, P.B., Swanstrom, R., Varmus, H.E., and Bishop, J.M. (1982) *J. Virol.* **41**, 527-534.
4. Petersen, R.B., Hensel, C.H., and Hackett, P.B. (1984) *J. Virol.* **51**, 722-729.
5. Petersen, R.B. and Hackett, P.B. (1985) *J. Virol.* **56**, 683-690.
6. Petersen, R.B., Moustakas, A., and Hackett, P.B. (1989) *J. Virol.* **63**, 4787-4796.
7. Mueller, P. and Hinnebusch, A.G. (1986) *Cell* **45**, 201-207.
8. Hensel, C.H., Petersen, R.B., and Hackett, P.B. (1989) *J. Virol.* **63**, 4986-4990.
9. Turner, D., Sugimoto, N., and Freier, S.M. (1988) *Ann. Rev. Biophys. Biophys. Chem.* **17**, 167-192.
10. Kozak, M. (1989) *J. Cell. Biol.* **108**, 229-241.
11. Pavlakis, G.N. and Felber, B.K. (1990) *New Biol.* **2**, 20-31.
12. Pavlakis, G.N., Lockard, R.E., Vamvakopoulos, N., Rieser, L., RajBhandary, U.L. and Vournakis, J.N. (1980) *Cell* **19**, 91-102.
13. Lowe, W.L., Roberts, C.T., Lasky, S.R. and LeRoith, D. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8946-8950.
14. Godefroy-Colburn, T., Ravelonandro, M., and Pinck, L. (1985) *Eur. J. Biochem.* **147**, 549-552.
15. Jobling, S.A., and Gehrke, L. (1987) *Nature* **325**, 322-325.
16. Dabrowski, C. and Alwine, J.C. (1988) *J. Virol.* **62**, 3182-3192.
17. Klausner, R.D. and Harford, J.B. (1989) *Science* **246**, 870-872.
18. Jacks, T., Power, M.D., Masiarz, F.R., Luciw, P.A., Barr, P.J., Varmus, H.E. (1988) *Nature* **331**, 280-283.
19. Sobel, E. and Martinez, H.M. (1985) *Nuc Acids Res.* **14**, 363-374.
20. Zuker, M. (1989) *Science* **244**, 48-52.
21. Jaeger, J.A., Turner, D.H. and Zuker, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7706-7710.
22. Bailey, J.M. and Davidson, N. (1976) *Anal. Biochem.* **70**, 75-85.
23. Woese, C.R. (1987) *Microbiol. Rev.* **51**, 221-271.
24. Pace, N.R., Smith, D.K., Olsen, G.J. and James, B.D. (1989) *Gene* **82**, 65-75.
25. Bizub, D., Katz, R.A. and Skalka, A.M. (1984) *J. Virol.* **49**, 557-565.
26. Katz, R.A., Cullen, B.R., Malavarca, R., and Skalka, A.M. (1986) *Mol. Cell. Biol.* **6**, 372-379.
27. Cobrinik, D., Soskey, L., and Leis, J. (1988) *J. Virol.* **62**, 3622-3630.
28. Darlix, J.-L. (1986) *J. Mol. Biol.* **189**, 421-434.
29. Pelletier, J. and Sonnenberg, N. (1987) *Biochem. Cell Biol.* **65**, 576-581.
30. Swanstrom, R., DeLorbe, W.J., Bishop, J.M. and Varmus, H.E. (1981) *J. Virol.* **41**, 535-541.
31. Neckameyer, W.S. and Wang, L.-H. (1985) *J. Virol.* **53**, 879-884.
32. Mayer, B.J., Hamaguchi, M. and Hanafusa, H. (1988) *Nature* **332**, 272-275.
33. Kitamura, N., Kitamura, A., Toyoshima, K., Hirayama, Y. and Yoshida, M. (1982) *Nature* **297**, 205-208.
34. Schwartz, D.E., R. Tizard, and W. Gilbert (1983) *Cell* **32**, 853-869.
35. Shibuya, M. and Hanafusa, H. (1982) *Cell* **30**, 787-795.
36. Thoraval, P., Savatier, P., Xiao, J.H., Mallet, F., Samarut, J., Verdier, G., and Nigion, V.M. (1987) *Nuc. Acids Res.* **15**, 9612.
37. Reddy, E.P., Reynolds, R.K., Watson, D.K., Schultz, R.A., Lautenberger, J. and Papas, T.S. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2500-2504.
38. Smith, D.R., Vennstrom, B., Hayman, M.J. and Enrietto, P.J. (1985) *J. Virol.* **56**, 969-977.
39. Bizub, D., Katz, R.A., and Skalka, A.M. (1984) *J. Virol.* **49**, 557-565.
40. Hughes, S.D. (1982) *J. Virol.* **43**, 191-200.