JOURNAL OF VIROLOGY, Aug. 1996, p. 5083–5099 Vol. 70, No. 8
0022-538X/96/$04.00+0

Structure, Function, and Evolution of Adenovirus-Associated RNA: a Phylogenetic Approach

YULIANG MA[1,2†] AND MICHAEL B. MATHEWS[1*]

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724,[1] and Graduate Program of Molecular Genetics and Microbiology, State University of New York at Stony Brook, Stony Brook, New York 11790[2]

Received 16 February 1996/Accepted 23 April 1996

To explore the structure and function of a small regulatory RNA, we examined the virus-associated (VA) RNA species of all 47 known human adenovirus serotypes and of one simian virus, SA7. The VA RNA gene regions of 43 human adenoviruses were amplified and sequenced, and the structures of 10 representative VA RNAs were probed by nuclease sensitivity analysis. Most human viruses have two VA RNA species, VA RNA_I and VA RNA_{II}, but nine viruses (19%) have a single VA RNA gene. Sequence alignments classified the RNAs into eight families, corresponding broadly to the known virus groups, and three superfamilies. One superfamily contains the single VA RNAs of groups A and F and the VA RNA_I species of group C; the second contains the VA RNA_I species of groups B1, D, and E and the unclassified viruses (adenovirus types 42 to 47), as well as the single VA RNAs of group B2; and the third contains all VA RNA_{II} species. Fourteen regions of homology occur throughout the molecule. The longest of these correspond to transcription signals; most of the others participate in RNA secondary structure. The previously identified tetranucleotide pair, GGGU:ACCC, is nearly invariant, diverging slightly (to GGGU:ACCU) only in the two group F viruses and forming a stem in the central domain that is critical for VA RNA structure and function. Secondary structure models which accommodate the nuclease sensitivity data and sequence variations within each family were generated. The major structural features—the terminal stem, apical stem-loop, and central domain—are conserved in all VA RNAs, but differences exist in the apical stem and central domains, especially of the VA RNA_{II} species. Sequence analysis suggests that an ancestral VA RNA gene underwent duplication during the evolution of viruses containing two VA RNA genes. Although the VA RNA_{II} gene seems to have been lost or inactivated by secondary deletion events in some viruses, the high degree of homology among the VA RNA_{II} species implies that this RNA may play an undiscovered role in virus survival. We speculate that the VA RNA genes originated from cellular sequences containing multiple tRNA genes.

Virus-associated (VA) RNA is the name given to a class of abundant, low-molecular-weight, regulatory RNA species found in the cytoplasm of adenovirus-infected cells (57). The synthesis, structure, and function of VA RNA have attracted considerable attention (for a review, see reference 47). In the mammalian adenoviruses studied, there are either one or two genes for VA RNA, known as VA RNA_I and VA RNA_{II} when two species are present. The genes are located near map unit 30 in the viral genome (Fig. 1A) and are transcribed by RNA polymerase III. Those RNAs that have been examined are highly structured (39). Some of the VA RNA species contribute to infectivity, but others do not seem to function effectively in this way (39, 73). An examination of the similarities and differences among a family of related genes can illuminate their structural and functional features, as well as their evolutionary origins and interrelationships. To these ends, we have compared the sequences of the VA RNA genes of the human adenoviruses.

Most work to date has concentrated on the prototypical VA RNAs of adenovirus type 2 (Ad2) and the closely related serotype Ad5. Studies with mutant viruses showed that inactivation of the VA RNA_I gene caused a 10- to 20-fold decrease in viral growth and that deletions in both VA RNA genes caused a 60-fold decrease, although elimination of VA RNA_{II} alone gave no detectable phenotype (7, 34, 73). Further investigation demonstrated that Ad2 VA RNA_I stimulates protein synthesis in the infected cell (54, 58, 63, 66) and serves to antagonize the interferon-induced cellular antiviral defense system (34, 43, 44, 54, 62). It can also stimulate the expression of transfected genes (32, 69, 70) and may contribute to the selective translation of viral mRNAs relative to host messages late in infection (31, 53). All of these actions appear to be mediated by the protein kinase PKR, the double-stranded RNA (dsRNA)-activated inhibitor of translation (also called DAI, P1, p68, and PKds). Activation of PKR is accompanied by, and possibly caused by, its autophosphorylation. Once activated, PKR phosphorylates initiation factor eIF2, leading to the inhibition of protein synthesis. VA RNA_I binds to PKR and competes with the dsRNA activator, apparently produced by symmetrical transcription of the viral genome (41), therefore preventing PKR from being activated (34, 35, 47).

Three structural features of Ad2 VA RNA_I, the apical stem-loop, central domain, and terminal stem (Fig. 1C), were identified by nuclease sensitivity analysis and related methods for probing RNA secondary structure (18, 49). Mutagenesis showed that the central domain and the adjacent part of the apical stem-loop are required for VA RNA_I to bind PKR and inhibit its activation (11, 18, 24, 25, 48–50, 55, 56). These conclusions have been refined and extended by protection studies (9, 10) and examination of the RNA-binding site of PKR (27, 60). The functional requirement at the base of the apical stem can be satisfied by a short base-paired duplex without apparent sequence specificity (11, 50). The structure of the central domain is complex, and recent work has led to a

* Corresponding author. Mailing address: Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, NY 11724. Phone: (516) 367-8374. Fax: (516) 367-8815. Electronic mail address: mathews@cshl.org.

† Present address: University of California, San Diego, Cancer Center, La Jolla, CA 92093.

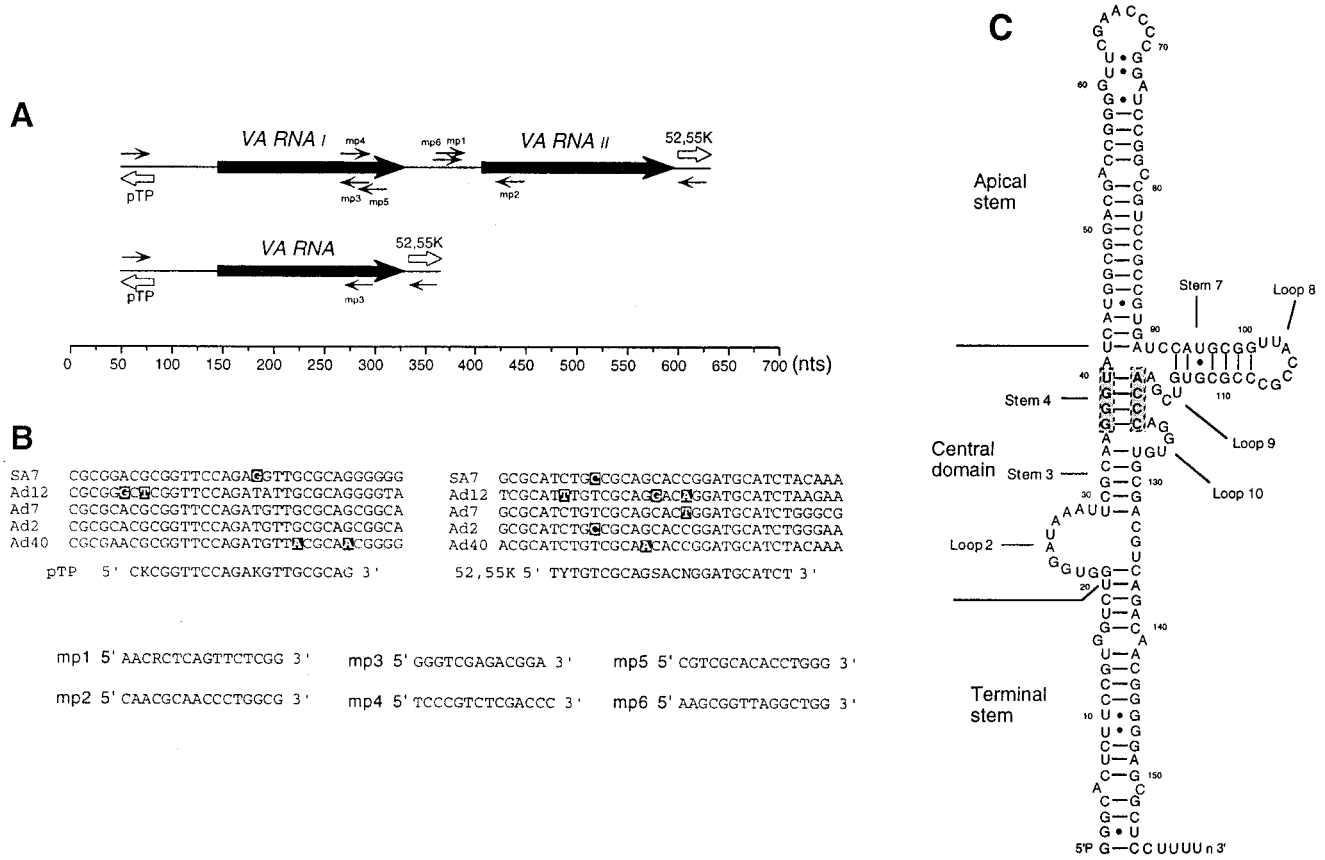

FIG. 1. Organization of the VA RNA region. (A) Diagram of the gene organization in adenoviruses containing two (top) or one (bottom) VA RNA gene. The VA RNAs and flanking protein coding regions (pTP and 52,55K) are highlighted with solid and open arrows, respectively. Primers for PCR and sequencing and internal primers for sequencing are marked with thick and thin arrows, respectively. (B) Nucleotide sequences of the primers. The five viral fragments (from GenBank) used to design the PCR primers (pTP and 52,55K) are shown as sequence alignments above the sequences of the PCR primers. Nucleotides that are not conserved are highlighted. (C) Ad2 VA RNA_I model. The three structural elements, the apical stem-loop, central domain, and terminal stem, are marked on the secondary structure model of Clarke et al. (11). Individual stems and loops discussed in the text are labeled. Stem 4 is composed of the conserved tetranucleotides ACCC and GGGU, shown shaded and boxed.

revision of the model in this region (10, 39, 55), with the suggestion that tertiary interactions in the central domain may be required for VA RNA_I function (40).

The compactness of the VA RNA structure imposes limitations on the mutagenic approach: instead of restricted local changes, many mutations cause wide-ranging structural alterations in the RNA that are difficult to interpret (48). As an alternative approach, we have used comparative sequence analysis to identify common features and deduce structural information (39). The underlying idea is that the VA RNA species carried by different adenoviruses all probably contribute to viral growth, so nucleotide positions that are unimportant for function should accommodate more variations than others that play an important role. Human adenoviruses have been classified into six groups or subgenera, A to F (20, 64, 74). An initial study of VA RNAs from Ad12, Ad7, and Ad2 (representing human adenovirus groups A, B, and C), as well as from SA7 and CELO (simian and avian adenoviruses, respectively), suggested that the overall structural plan of the molecule retains the three chief features characterizing the Ad2 VA RNA_I model (39). In addition, a pair of mutually complementary tetranucleotides, GGGU and ACCC, was found in all of the mammalian VA RNAs, predicting the existence of a hitherto unsuspected short stem in the center of the molecule. The GC-rich apical stem contained a second pair of conserved tetranucleotides, CCGG:(C/U)CGG. Despite these structural similarities, the seven VA RNAs displayed wide variations in the ability to inhibit PKR: Ad7 VA RNA_I was at least as active as Ad2 VA RNA_I, but their VA RNA_{II} counterparts and CELO VA RNA were essentially inactive; the single VA RNAs of Ad12 and SA7 exhibited intermediate levels of function. Thus, the relationship between VA RNA structure and function is not obvious. Furthermore, these RNAs appeared to be so divergent over most of their length that sequence comparisons among them revealed little information. Within group C, on the other hand, the sequences are so similar that it was not possible to discern compensating base changes which might confirm suspected pairing schemes (46).

Assuming that common sequences and structures might be discernible from a larger datum set, we have extended the phylogenetic analysis to include all of the human adenovirus VA RNAs. We collected published DNA sequences for the VA RNA region of six viruses (Ad2, Ad5, Ad7, Ad12, Ad40, and SA7) and amplified this region from all the other human adenoviruses by PCR. The sequences thus obtained were analyzed, and the secondary structures of representative VA RNA species were probed by the nuclease sensitivity technique. The previously discovered tetranucleotide pairs proved to be conserved not only in sequence but also in the structural elements that they form. Furthermore, several additional con-

served sequences were uncovered. Secondary structure models were deduced by combining structural data with sequence information (including compensating changes, neutral substitutions, and base conservation). All of the VA RNAs have structures that are broadly similar to that proposed for Ad2 VA RNA_I, but functionally significant differences were evident, especially in VA RNA_{II}. Despite evidence that this species is dispensable, its sequences were among the most highly conserved, suggestive of a separate and unique function. The sequence comparisons also led to inferences concerning transcription signals as well as the evolutionary relationships and origins of the VA RNAs. Surprisingly, the greatest sequence variability was found in functional parts of the VA RNA molecule, the apical stem-loop and central domain, suggesting that VA RNAs from different groups may have incorporated sequences of different origins to mold similar structures.

MATERIALS AND METHODS

Adenovirus DNA. All of the human adenoviruses, Ad1 to Ad47, except Ad2, Ad5, Ad7, Ad12, and Ad40, were obtained from the American Type Culture Collection (Table 1). Ad10 and Ad14 were shipped in a dried state and were dissolved in 1 ml of Dulbecco's modified Eagle medium without serum. Ad7, Ad16, and Ad21 were also provided by C. Tibbetts (Vanderbilt University). To isolate virion DNA, 200 μl of virus suspension was incubated for 30 min at 37°C with 1 mg of proteinase K per ml in the presence of 0.5% sodium dodecyl sulfate to digest viral proteins. The samples were then adjusted to 0.1 M NaCl and extracted several times with phenol and with chloroform-isoamyl alcohol (24:1). DNA was precipitated with ethanol and dissolved in 50 μl of 10 mM Tris-HCl–1 mM EDTA buffer, pH 7.6 (TE).

PCRs. The VA RNA gene region was amplified by using two degenerate primers, pTP and 52,55K, designed to hybridize to conserved flanking sequences (Fig. 1A and B). In each 50-μl PCR, 2 μl of the viral DNA sample (about 20 pg) was incubated with 0.2 mM both primers, 0.2 mM all four deoxynucleoside triphosphates, and 1 U of *Taq* polymerase (Perkin-Elmer) in the presence of 10 mM Tris-HCl (pH 8.3), 1.5 mM $MgCl_2$, 50 mM KCl, and 100 μg of gelatin per ml. In a Hybrid OmniGene thermal cycler, the mixture was first heated at 94°C for 30 s and then incubated for 30 cycles at three temperatures (94°C for 30 s, 60°C for 1 min, and 72°C for 30 s). After a final incubation at 72°C for 10 min, EDTA was added to a concentration of 10 mM. Positive controls included DNA from Ad2 and Ad7 purified by CsCl banding. Water was used for negative controls.

DNA sequencing. The amplified DNA fragments were purified by electrophoresis through 1.5 or 2% NuSieve (FMC BioProducts) agarose gels in 50 mM Tris-acetate buffer (pH 8.0) containing 0.5% ethidium bromide. Gel slices containing the desired bands were excised under UV light and melted at 68°C for 5 min in 400 μl of TE. DNA was extracted twice with phenol and then once with chloroform-isoamyl alcohol mixture (24:1), precipitated, and dissolved in 60 μl of TE. Sequence analysis was performed with 18 to 27 μl of DNA sample for each set of four reactions with either the regular or the deaza Taq-Track kit (Promega) as instructed by the manufacturer. The reaction mixtures were heated at 80°C for 2 min before loading on a 8% polyacrylamide–8 M urea sequencing gel made with a 0.4-mm or a wedged (GIBCO BRL) spacer. Each sample was run for two or three different times. All of the PCR fragments have a GC-rich region which causes compressions in the sequencing gels, as well as runs of A or T residues which cause band shifts that generate multiple bands for the subsequent nucleotides. Therefore, we synthesized additional internal primers (mp1 to mp6 [Fig. 1]) to bypass difficult points and improve resolution. Sequences were read from both strands to confirm assignments and resolve uncertainties. The sequences have been deposited in GenBank (accession numbers listed in Table 1) and concur with those reported previously (33, 39) apart from scattered differences in Ad3, Ad4, Ad8, Ad19, Ad37, and Ad41. In Ad41, the differences are all outside the transcribed sequence. Discrepancies are unlikely to be due to PCR artifacts since sequences were read from DNA directly amplified from a pool of viral DNA equivalent to about 10^5 molecules.

Sequence analysis. Stretches of overlapping and complementary strand sequence from each viral DNA fragment were assembled into a coherent sequence by using the SEQ program (IntelliGenetics suite). The nucleotide sequences of the DNA fragments were analyzed with the aid of several computer programs to extract information on the organization of the region as a whole (Fig. 2) and to study the evolution of VA RNA (Fig. 4) using TreeAlign (29, 30) with a default gap penalty of 8 and length penalty of 3. Individual VA RNA genes (Fig. 3) were also examined by using PILEUP, from the Genetics Computer Group (GCG) package, with a weight parameter setting of 3 and gap penalty setting of 0.1. Sequence alignments were edited manually with LINEUP (GCG package) and with the text editors vi and EMACS and then input to the programs DNApar, DNAdist, and Fitch in the PHYLIP package (16) for the analysis of evolutionary trees. The evolutionary tree was displayed by using the Drawtree program in the PHYLIP package. RNA secondary structure predictions were made with the help of the RNA folding program MulFold (26) or MFOLD (GCG package), both using Zuker's algorithm (77). The top 10% of optimal structures were collected and compared. In some analyses, nuclease sensitivity data were incorporated by substituting B, D, H, and V for unpaired A, C, G, and U nucleotides, respectively, in the sequences examined by the computer folding programs.

Nuclease sensitivity analysis. The secondary structure of VA RNA synthesized in vitro was probed with a battery of enzymes of different specificities:

TABLE 1. Sources of adenoviruses and sizes of VA RNA region PCR fragments

Virus	ATCC catalog no.	Subgroup	Size (nt)	Accession no.
Ad1	VR-1078	C	568	U52532
Ad2	—[a]	C	567	
Ad3	VR-3	B	582	U52534
Ad4	VR-4	E	548	U52535
Ad5	—[b]	C	567	
Ad6	VR-1803	C	567	U52533
Ad7	—[c]	B	584	U52574
Ad8	VR-1085	D	520	U52563
Ad9	VR-1086	D	521	U52536
Ad10	VR-1087	D	493	U52566
Ad11	VR-12	B	354	U52569
Ad12	—[d]	A	307	
Ad13	VR-1090	D	521	U52537
Ad14	VR-15	B	353	U52570
Ad15	VR-1092	D	521	U52538
Ad16	VR-1093[e]	B	571	U52564
Ad17	VR-1094	D	521	U52539
Ad18	VR-1095	A	306	U52567
Ad19	VR-1096	D	521	U52540
Ad20	VR-1097	D	521	U52541
Ad21	VR-1098[e]	B	571	U52565
Ad22	VR-1100	D	521	U52542
Ad23	VR-1101	D	521	U52543
Ad24	VR-1102	D	521	U52544
Ad25	VR-1103	D	521	U52545
Ad26	VR-1104	D	521	U52546
Ad27	VR-1105	D	521	U52547
Ad28	VR-1106	D	521	U52548
Ad29	VR-1107	D	521	U52549
Ad30	VR-273	D	521	U52550
Ad31	VR-1109	A	299	U52568
Ad32	VR-625	D	521	U52551
Ad33	VR-626	D	521	U52552
Ad34	VR-716	B	354	U52571
Ad35	VR-718	B	357	U52572
Ad36	VR-913	D	521	U52555
Ad37	VR-929	D	521	U52553
Ad38	VR-988	D	520	U52562
Ad39	VR-932	D	521	U52554
Ad40	—[f]	F	323	
Ad41	VR-930	F	326	U52573
Ad42	VR-1304	D[g]	521	U52556
Ad43	VR-1305	D	521	U52557
Ad44	VR-1306	D	521	U52558
Ad45	VR-1307	D	521	U52559
Ad46	VR-1308	D	521	U52560
Ad47	VR-1309	D	521	U52561
SA7	—[h]		322	

[a] Viral DNA provided by B. Stillman; GenBank locus name ADRCG.
[b] GenBank locus name AD5001.
[c] Virus provided by C. Tibbetts and cloned viral DNA fragment by J. Engler; GenBank locus name ADR7001.
[d] Viral fragment provided by J. Engler; GenBank locus name ADRE2B.
[e] Virus also provided by C. Tibbetts.
[f] GenBank locus name ADRVA.
[g] Ad42 to Ad47 are classified in group D in this study.
[h] Cloned viral fragment provided by G. Akusjärvi; GenBank locus name AESVARNA.

RNases T_1, U_2, Bc (from *Bacillus cereus*), and T_2 (specific for G, A, pyrimidine, and any nucleotide, respectively, in single-stranded regions) and nuclease V_1 (specific for stacked or duplexed bases). Genes encoding the VA RNA_I and VA RNA_{II} species of Ad4 and Ad15 were subcloned under the control of the T7 RNA polymerase promoter as described previously (39). The T7 construct encoding Ad2 VA RNA_I was described by Mellits et al. (50). Constructs encoding Ad2 VA RNA_{II} and the VA RNAs of Ad7, Ad12, and SA7 were described previously (39). RNA was synthesized by T7 RNA polymerase from the T7 constructs and labeled at either its 5′ end with [γ-^{32}P]ATP (49) or its 3′ end with [5′-^{32}p]pCp (14). After purification by electrophoresis through a polyacrylamide-urea gel, nuclease sensitivity analysis was conducted as described previously (10, 39) such that <10% of the molecules were cleaved. For RNases T_1, Bc, and T_2 and nuclease V_1, the reaction mixtures contained 50 mM Tris-HCl (pH 7.5), 2 mM $MgCl_2$, and 150 mM KCl. RNase U_2 digests were buffered at pH 5.0 with 50 mM citric acid-potassium citrate plus 2 mM $MgCl_2$. Digestion with RNase T_1 at 68°C and alkaline digestions were conducted as described previously (50). The partial digests were resolved in a 12% polyacrylamide-urea gel containing 45 mM Tris-boric acid (pH 8.0) and 10 mM EDTA with wedged spacers (GIBCO BRL) at 1,400 V for 4 h.

RESULTS

Strategy. The VA RNA genes of three human adenoviruses, Ad2, Ad7, and Ad12, and of the simian virus SA7 are located between regions encoding the conserved viral proteins pTP and 52,55K (2, 15, 37, 39, 65). In SA7 and Ad12, there is a single VA RNA gene, whereas Ad2 and Ad7 have two distinct genes, VA RNA_I and VA RNA_{II}, in tandem (Fig. 1A). Comparison of the six RNAs revealed common features as well as differences in both structure and function (39). To assess the generality of these findings, we studied the VA RNA genes of the other human serotypes. DNA containing these genes was amplified by using primers designed to hybridize to flanking protein coding regions. From an alignment of the DNA sequences of five viruses whose sequences were available in GenBank, we selected two stretches of 22 to 24 nucleotides (nt) which displayed the greatest degree of conservation and synthesized the degenerate primers pTP and 52,55K (Fig. 1B). Since the sequences are drawn from a broad sample of human adenoviruses, including members of groups A, B, C, and F as well as a simian virus, we hoped that these two primers would serve for the remaining serotypes. Indeed, by varying the annealing temperature between 60 and 65°C, a PCR fragment was amplified from each of the human adenovirus samples. We obtained fragments of about 300 bp (range, 299 to 357 bp) from 10 viruses (including SA7) and of about 550 bp (range, 493 to 584 bp) from the others (Table 1). Although several of the VA RNAs were identical, each PCR fragment differed by at least one nucleotide from all of the others, thereby eliminating the possibility of cross-contamination of virus stocks or PCR products.

Organization of the VA RNA gene region. As expected, the short PCR fragments listed in Table 1 proved to encode one VA RNA whereas the long fragments contained two tandemly arranged homologous sequences (Fig. 2). Following the convention established for Ad2 (42), we refer to the upstream transcript as VA RNA_I and the downstream transcript as VA RNA_{II}. All members of groups A and F, as well as SA7, have a single VA RNA gene, whereas all viruses in groups C, D, and E have two VA RNA genes. The unclassified viruses Ad42 to Ad47, also have two genes. Only group B contains viruses which differ in the number of VA RNA genes encoded in their genomes. This feature, as well as sequence homology (see below), divides group B into two subgroups which correspond to groupings recognized by Wadell et al. (74) on other criteria; ironically, the subgroup B2 viruses (serotypes 11, 14, 34, and 35) have one VA RNA, while subgroup B1 viruses (serotypes 3, 7, 16, and 21) have two.

The VA RNA genes were located by assuming that their 5′ termini correspond to the major start sites identified for the VA RNAs studied previously (39) and that their 3′ termini are at the T residues in the 3′-terminal thymidine run. An additional signature is the presence of a B-box homology starting at approximately +60 relative to the start site. The subgroup B2 sequences contain two runs of T residues separated by about 40 nt, and some VA RNA_{II} genes have several TTT triplets. To locate the 3′ end of these VA RNAs, we used as a further criterion the ability of the 5′ and 3′ ends of VA RNA to pair with one another. As summarized in Table 2, all of the VA RNAs are expected to be 163 ± 14 nt in length, apart from Ad10 VA RNA_{II}, which is unusually short. It would seem that this gene has suffered a deletion of 27 nt as a result of looping out between flanking direct repeats (CGCGGC). With this exception, the RNAs within each group are tightly clustered in size, to the extent that the number and length of the species could be used as a criterion for classification. In this respect, as well as in their sequence (Fig. 3), the six unclassified viruses (Ad42 to Ad47) are very similar to the 23 viruses classified in group D, and we therefore consider them additional members of this group. Many of the expanded group D viruses differ in sequence at only one or two positions and therefore are represented by a consensus sequence in the alignment shown in Fig. 2.

The VA RNA coding sequences closely abut the adjacent protein coding sequences (Fig. 2). Upstream of the VA RNAs, 28 to 33 nt separate the VA RNA initiation site from the pTP splice site. The region between about −10 and −20 (relative to the 5′ end at +1) is remarkably rich in adenosine, which often occurs in two runs of three or four consecutive A residues separated by two other bases. Downstream of the VA RNAs, there are 12 to 23 nt between the beginning of the run of T residues of the VA RNA termination signal and the 52,55K protein splice site. This stretch is very rich in T residues, and in some viruses it consists almost exclusively of thymidines. Exceptionally, in the four subgroup B2 viruses, this downstream span contains about 50 nt. The extra nucleotides are not especially T rich and display patchy homology with the subgroup B1 VA RNA_{II} gene, suggesting that they correspond to remnants of a VA RNA_{II} gene. The view that the subgroup B2 viruses suffered deletions resulting in the loss of their VA RNA_{II} species is strengthened by inspection of the B2 VA RNA sequences; these RNAs are very similar to the Ad3 and Ad7 VA RNA_I species (subgroup B1), from which they differ mainly in a 13-nt deletion (Fig. 3).

In viruses that encode two VA RNAs, there is also a spacer sequence intervening between the two RNA coding genes. Unlike the 5′ and 3′ flanking regions, which are all of approximately the same length and display sequence similarities, the spacer varies greatly between virus groups in both length and sequence (Table 2; Fig. 2). It consists of about 96 nt in group C, 56 nt in groups D and E, and 72 nt in group B1. The spacer sequences are homologous within each group, but little homology is evident when the four groups are compared (Fig. 2). Some homology exists between the spacer regions immediately upstream of VA RNA_{II} and the sequence upstream of VA RNA_I. This is especially evident in the group C viruses, which appear to contain stretches of sequence in their spacer regions that are homologous to the pTP coding sequence and the A-rich stretch upstream of the VA RNA_I gene.

Sequence relationships among the VA RNAs. The unique VA RNA sequences are listed in Table 3 and compared in Fig. 3 and 5. The VA RNA_I and VA RNA_{II} species of a particular virus are substantially different from one another, but equivalent VA RNA species within a serotype group resemble one another closely in sequence. Indeed, many of the RNAs are identical; in such cases, we have selected one example to stand as prototype (Table 3). A total of 37 unique RNAs derive from

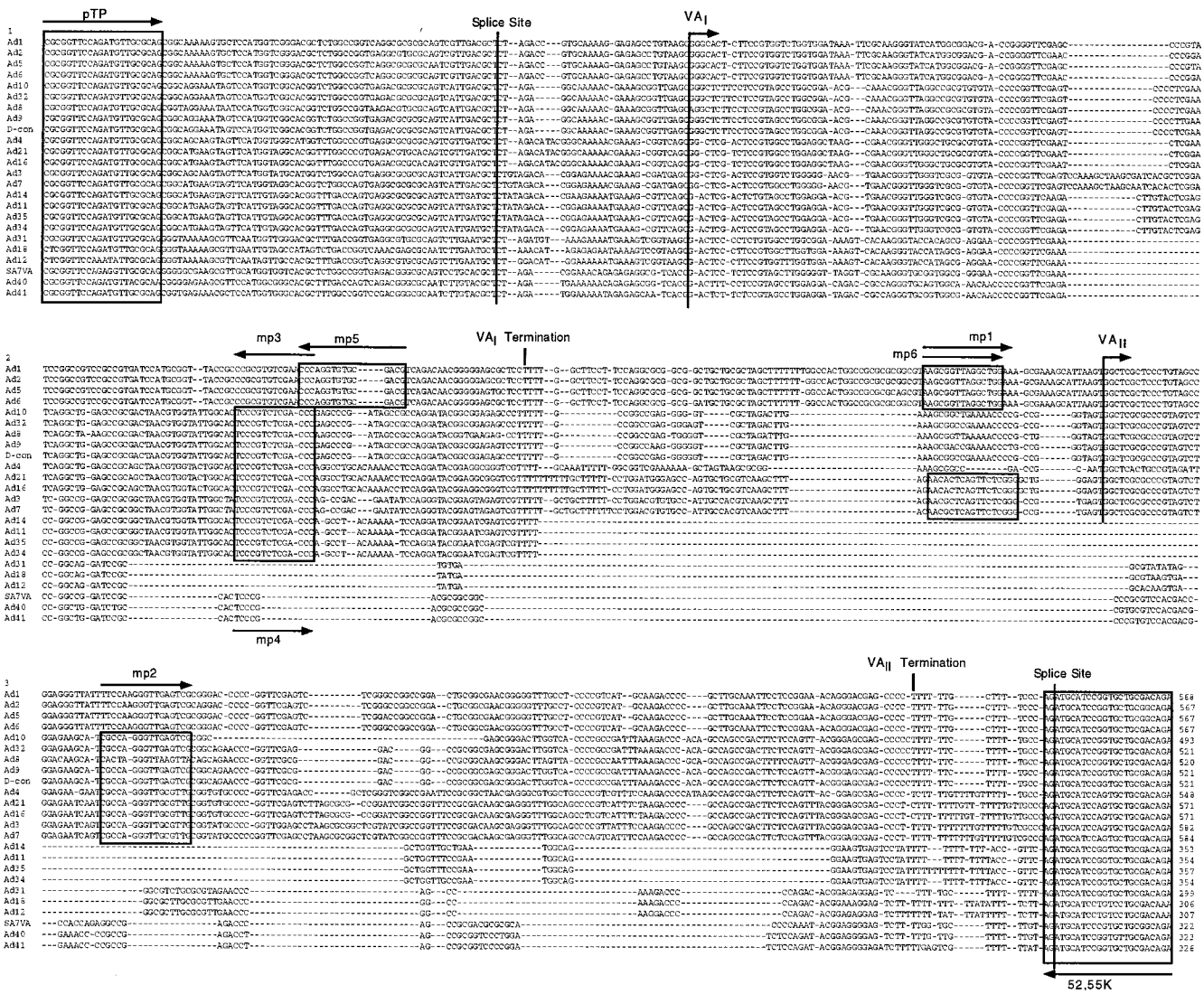

FIG. 2. Sequence alignment of the VA RNA gene region. All human adenovirus sequences and that of SA7 are shown, except that group D is represented by the four most divergent sequences (Ad10, Ad32, Ad8, and Ad9) and a consensus sequence (D-con) for all group D viruses. Initiation and termination sites of the VA RNAs and the splice sites of the terminal protein precursor and 52,55K protein mRNAs are indicated. The regions to which the PCR primers and the internal sequencing primers hybridize are marked with arrows and boxed.

the 47 human adenoviruses; given that all but nine of the viruses encode two species, the maximum possible number of distinct sequences is 85, which provides a measure of the redundancy in the datum set. In contrast to the relative homogeneity within a group, sequence differences between groups are usually substantial. An exception is the VA RNA$_I$ species of the group E virus Ad4, which differs in only two nucleotides from the corresponding species of the group B virus Ad21. Their VA RNA$_{II}$ species are also related, albeit less closely, but the spacer sequences differ considerably (Fig. 2).

Relationships among the prototype sequences are illustrated by the alignment shown in Fig. 3. Three superfamilies can be recognized. The first of these (A, F, C$_I$) contains the single VA RNA species of groups A and F (including SA7), together with the VA RNA$_I$ species of the group C viruses; the second (BDE$_I$) contains the VA RNA$_I$ species of groups D, E, and B1, together with the single VA RNA species of subgroup B2; the third (VA$_{II}$) contains all of the VA RNA$_{II}$ species which, despite significant differences among groups, form a more cohesive set than the VA RNA$_I$ species. Within each superfamily, sequences from each serotype group were clustered, forming eight families. The families correspond to the serotype groups with the VA RNA$_I$/VA RNA$_{II}$ distinction superimposed and the relationships among the B1, B2, and E VA RNAs recognized. Thus, the families are A (the single VA RNAs of group A and SA7); BE$_I$ (the VA RNA$_I$ species of groups B1 and E, as well as the single VA RNAs of group B2); BE$_{II}$; C$_I$; C$_{II}$; D$_I$; D$_{II}$; and F.

This organization of the sequences into three superfamilies and eight families correlates well with the clustering of sequences when their relationships are displayed as either a dendrogram (not shown) or a phylogenetic tree. Figure 4 gives an example of an unrooted phylogenetic tree generated on the basis of a combination of distance matrix and approximate parsimony methods (29, 30). In superfamily A, F, C$_I$, the group C VA RNA$_I$ species and group A VA RNA species are closer to each other than to group F VA RNAs. The BDE$_I$ superfamily has the least amount of sequence variation, despite the

presence of a 13-nt insertion in the apical loop of Ad3 and Ad7 VA RNA_I. These two group B1 VA RNA_I species and the group B2 VA RNAs are in a single subbranch. The other two group B1 VA RNA_I species (Ad16 and Ad21) form a unique subbranch with Ad4 VA RNA_I (group E). The group D VA RNA_I species form the third subbranch. In superfamily VA_{II}, three branches are formed. The branch composed of the VA RNA_{II} species of group C is separated from the other two branches, which contain the VA RNA_{II} species of group D and of groups B and E. Similar trees were generated by several other programs and in a study including a broader selection of simian viruses (33), lending credibility to the phylogenetic relationships deduced above and the consensus sequences described below.

Sequence conservation. The more detailed sequence alignment shown in Fig. 5 has been optimized to emphasize sequence conservation. Three tiers of consensus sequences were deduced. Tight consensus sequences were obtained for the eight families of VA RNAs; comparison of these family consensus sequences allowed the derivation of looser consensus sequences for the superfamilies and of a more tentative general consensus for all of the human VA RNAs (shown at the bottom of Fig. 5). The conservation of individual nucleotides is highlighted and indicated quantitatively, on a scale from 1 to 8, above the general consensus sequence. Strict sequence conservation across the entire ensemble of VA RNAs is limited to scattered regions and is largely associated with transcriptional functions of the gene (Fig. 5) or with secondary structural features of the RNA (see below), most notably the pairing of sequences in the apical and terminal stems and in stem 4 (depicted in Fig. 1C and 3).

The A and B boxes, which constitute the internal control regions (or promoter) for RNA polymerase III and have been defined experimentally for Ad2 VA RNA_I (6, 17, 28, 59, 67, 76), are well conserved. All of the VA RNAs have both boxes at similar positions, except for Ad10 VA RNA_{II}, which lacks a B box at the usual position as a result of the deletion noted above. The sequences match well the consensus sequences derived from tRNA genes (19, 67). Divergence from the consensus, when present, is usually toward the 3′ edge of the box. On the other hand, flanking nucleotides, especially on the 5′ sides of the boxes, are also highly conserved. The B-box homology is very high, with 8 nt matching the 9-nt consensus. The A boxes, located 34 to 40 nt upstream of the B box, are slightly less homologous to the consensus in some of the VA RNAs. In group A and the group B2 VA RNA_I species, 9 nt match the 10-nt A-box consensus; 8 nt match in group F and SA7 and the VA RNA_{II} species of groups C and D; and only 7 nt match in the group B1 VA RNA_{II} species.

Highly conserved sequences, 2 to 4 nt in length, are present throughout the VA RNA molecule (Fig. 5). Although they are short, compelling argument for their significance comes from the observation that many of them are mutually complementary and participate in the formation of duplexed stems forming the framework of the RNA secondary structure as described below. Thus, the stretches of sequence extending 20 nt into the RNA from the 5′ and 3′ ends display six regions of patchy conservation (Fig. 3 and 5). At the extreme termini, this conservation is probably influenced by the requirements for initiation and termination, which are not fully defined but include a preference for a 5′-terminal G (or possibly A) and for a 3′ run of T residues surrounded by a GC-rich stretch char-

TABLE 2. Organization of VA RNA genes

Group	Serotype(s) or virus	Sequence length (nt)					
		5′ spacer[a]	VA RNA_I[b]	Spacer	VA RNA_{II}	3′ spacer[c]	Total
A	12	31	149			20	307
	18	31	149			19	306
	31	31	149			12	299
B1	3	32	174	70	176	23	582
	7	32	174	71	177	23	584
B2	11	32	161			54	354
	14	32	161			53	353
B1	16, 21	33	162	73	176	20	571
B2	34	32	161			54	354
	35	32	161			57	357
C	1	30	160	97	161	13	568
	2, 5, 6	30	160	96	161	13	567
D	8	29	161	56	154	13	520
	9, 13, 15, 17, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 32, 33, 36, 37, 39, 42, 43, 44, 45, 46, 47	28	163	56	154	13	521
	10	28	163	55	127	13	493
	38	28	163	55	154	13	520
E	4	33	162	56	173	17	548
F	40	28	173			15	323
	41	28	174			17	326
	SA7	28	172			15	322

[a] Length does not include 85 nt of preterminal protein coding sequences.
[b] Including the single VA RNAs of viruses which contain only one VA RNA gene.
[c] Does not include 22 nt of 52,55K protein coding sequences.

FIG. 3. Structural features in VA RNA sequences. Alignments of prototype VA RNAs (see Table 3) generated with the PILEUP program are arranged into three superfamilies, which are separated by blank lines. The serotype groups to which these RNAs belong are labeled on the left. The conserved tetranucleotide sequences, GGGU and ACCC, are boxed and shaded. Regions which can pair and form the duplex structures of the apical stem (AS) and terminal stem (TS) are boxed.

TABLE 3. VA RNA prototypes and their equivalents

Prototype	Identical RNA species[a]
VA RNA_I	
Ad1 VA_I	5
Ad2 VA_I	6
Ad3 VA_I	—
Ad4 VA_I	—
Ad7 VA_I	—
Ad8 VA_I	—
Ad9 VA_I	—
Ad13 VA_I	17, 39
Ad15 VA_I	20, 22, 37, 46
Ad19 VA_I	—
Ad21 VA_I	16
Ad23 VA_I	—
Ad44 VA_I	10, 24, 25, 26, 27, 28, 29, 30, 32, 33, 36, 38, 42, 43, 45, 47
VA RNA	
Ad11 VA	34, 35
Ad12 VA	—
Ad14 VA	—
Ad18 VA	—
Ad31 VA	—
Ad40 VA	—
Ad41 VA	—
SA7 VA	—
VA RNAII	
Ad1 VA_{II}	6
Ad2 VA_{II}	—
Ad3 VA_{II}	—
Ad4 VA_{II}	—
Ad5 VA_{II}	—
Ad7 VA_{II}	—
Ad8 VA_{II}	—
Ad10 VA_{II}	—
Ad13 VA_{II}	—
Ad15 VA_{II}	9, 17, 20, 23, 28, 30, 33, 36, 38, 42, 43, 19, 27, 29, 32, 39, 44, 45, 46
Ad16 VA_{II}	—
Ad21 VA_{II}	—
Ad22 VA_{II}	—
Ad24 VA_{II}	—
Ad25 VA_{II}	26
Ad37 VA_{II}	—
Ad47 VA_{II}	—

[a] The serotype numbers of viruses which contain a VA RNA species identical to the prototypes. —, no identical species.

acterizing efficient polymerase III termination sequences (8). As these requirements appear to be rather relaxed, it is likely that the terminal conservation owes more to constraints imposed by the base pairing of the 5′ and 3′ sequences to form the terminal stem (Fig. 3). Indeed, the six conserved blocks form three discontinuous segments of terminal stem duplex in secondary structure models (Fig. 6 to 8).

Such structural constraints may also contribute to the patchy conservation evident in the region encoding the apical stem (Fig. 1 and 3). A pair of mutually complementary tetranucleotides, CCGG and (U/C)CCGG, that forms part of the apical stem as noted previously (39) is reasonably well conserved in the full set of VA RNA sequences. The first CCGG, which includes the first two bases of the B box, is invariant. An additional variation on the (U/C)CGG sequence, CCAGG, is found in the BDE_I superfamily of VA RNAs, but the A residue seems to loop out to allow pairing to occur (Fig. 6 to 8). Thus, for the human VA RNAs as a whole, the consensus pairing is CCGG:(U/C)C(A)GG. An additional pair of tetranucleotides, GCGN and CCGC (Fig. 3), contributes to the lower part of the apical stem as seen in the Ad2 VA RNA_I secondary structure model (Fig. 1C).

The middle of the molecule is generally variable in sequence (Fig. 5). Within a superfamily, the alignments are convincing but contain some gaps, implying that several deletion, insertion, or duplication events have occurred. Between superfamilies, homologies are highly dispersed, raising the possibility that the virus superfamilies have derived their central sequences from different sources. In contrast to most of the region, two mutually complementary tetranucleotides in the central domain, ACCC and GGGU, are strikingly conserved (Fig. 3 and 5). These two tetranucleotides, which were previously reported to be common to seven species of VA RNA (39), are invariant in human VA RNAs apart from those of group F (Ad40 and Ad41), which diverge in a single position. In these RNAs, the sequence ACCU replaces ACCC; nevertheless, this tetranucleotide is still able to pair with the GGGU sequence by substituting a G:U base pair for one of the G:C pairs. Direct evidence that the tetranucleotides pair with one another has been obtained through generation of compensating mutants (40). This base pairing makes a critical structural contribution to the central domain and to the duplexed axis of the molecule (see below); the lack of compensating base changes in natural isolates, and the deleterious effect of introducing such mutations, suggests that its sequence may also be important.

The central domain also contains two additional short conserved stretches, two and three bases in length, which cannot pair with one another (Fig. 5). In most cases, the two-base sequence is unpaired and located in loop 10, but in the VA RNA_{II} species of groups B2 and D, it is in a short stem below loop 10 (Fig. 6 to 8). The three-base sequence is usually paired in stem 7, but its partners are variable. The significance of these nucleotides is uncertain at present, but they may be involved in tertiary structure formation or interactions with PKR.

VA RNA secondary structure. Examination of the primary sequence revealed the existence of several regions where stretches of nucleotides exhibited substantial or partial complementarity. These regions, boxed in Fig. 3, correspond to the apical stem, terminal stem, and GGGU:ACCC stem (stem 4) of Ad2 VA RNA_I (Fig. 1C), suggesting that the overall structures of all VA RNAs may be similar. Because no single method of analysis can reliably predict RNA secondary structure, we have combined the results from three independent approaches to arrive at models for VA RNA.

Comparative data obtained from phylogenetic surveys provide valuable information about higher-order structure, especially when used to test the robustness of models obtained by other means. One criterion for a satisfactory model is that it should be able to accommodate the sequence changes observed in closely related species. Potential structural models can be derived by combining sequence information with computer-assisted predictions of folding patterns made by using the Zuker algorithm (77, 78). This procedure typically gives a series of models differing in calculated stability, but the correct structure is not necessarily the one with the lowest free-energy value. Consequently, it is often necessary to consider several such models and to take into account experimental data derived from secondary structure probing experiments (summarized in Fig. 6 to 8). The structure of Ad2 VA RNA_I has been intensively examined with chemical and enzymatic probes, as well as by RNase protection and mutagenic approaches; therefore, this species formed the starting point for these analyses. Somewhat different strategies were required in the three superfamilies, which will be discussed in turn.

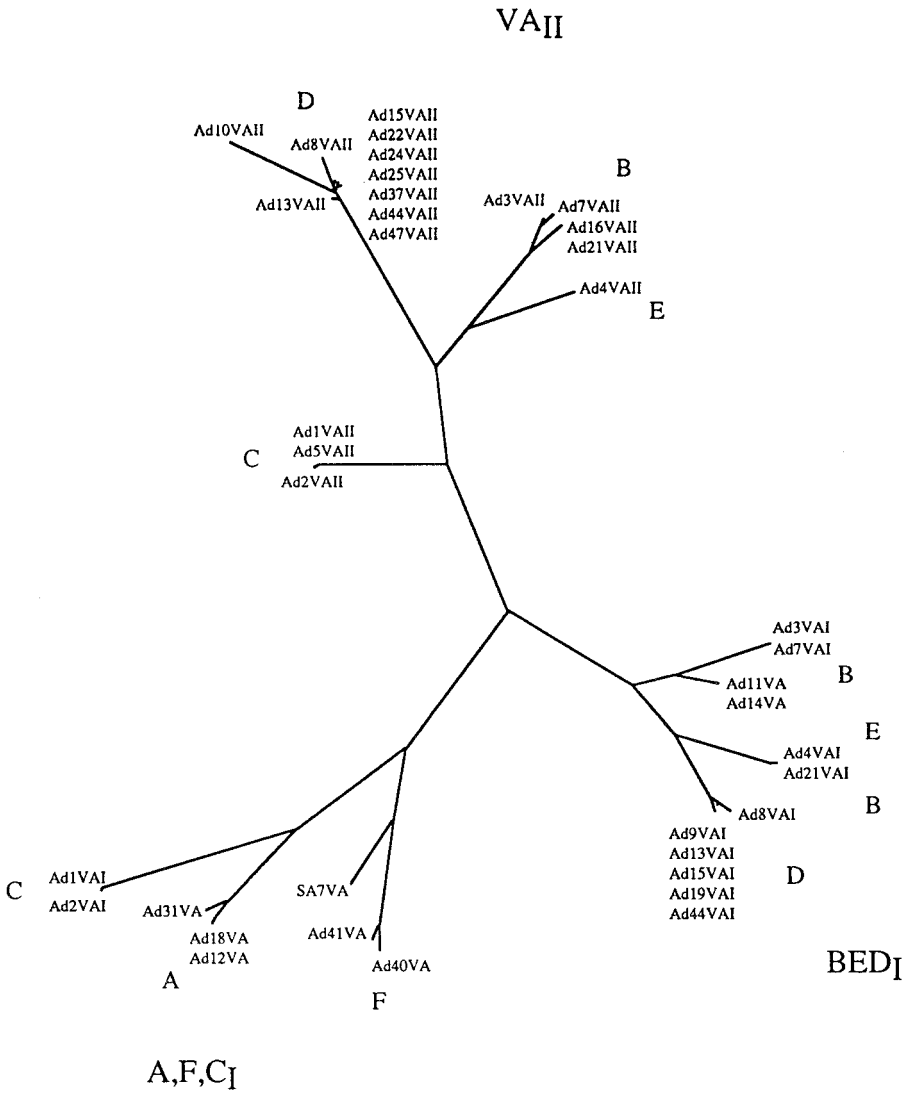

FIG. 4. Phylogenetic tree for VA RNA. An unrooted tree was generated for the representative VA RNA species (Table 3) with the program TreeAlign. Each species of VA RNA is labeled at the tip of its respective branch.

The A, F, C_I superfamily. The model shown in Fig. 6A is that proposed for Ad2 VA RNA_I by Clarke and Mathews (10). The two single nucleotide changes observed in other group C viruses (Ad1 and Ad5) are marked on the Ad2 VA RNA_I diagram. An alternative conformation of the apical stem region (40), which in some ways resembles those of other VA RNA species discussed below, is also illustrated. Neither base substitutions within group C nor comparisons with VA RNAs of other groups allow a decisive choice to be made between these alternatives, and evidence from nuclease sensitivity and mutagenesis suggests that the two forms may be in equilibrium with one another. Because of the paucity of changes with the group C VA RNA_I sequences, the phylogenetic approach yielded little refinement and only limited confirmation of the structure in these cases.

On the other hand, comparison with the VA RNAs of group A and F was informative. To model these species, we selected structures predicted by the folding program that are compatible with nuclease sensitivity data obtained for Ad12 VA RNA (group A) and SA7 (representing group F). Both RNAs adopt structures similar to that for group C VA RNA_I despite considerable differences in sequence (Fig. 6B and C). All three structures consist of a terminal stem, an apical stem-loop, and a central domain containing the GGGU:ACCC stem as a key element (boxed in Fig. 6). Within group A and group F, several examples of compensatory and conservative base changes occur in paired regions, lending support to the stem structures diagrammed. The group A and group F models are similar to each other in the apical stem-loop: both have a large internal loop halfway up the apical stem. In this respect, they differ slightly from both possible forms of group C VA RNA_I, which has either a large terminal loop or a bulge (depending on the conformation) ostensibly resulting from a 7-nt insertion in the stem. Nevertheless, in all members of the superfamily, the GC-rich paired section at the base of this stem is preserved. All of the conserved nucleotides from Fig. 5 (circled in Fig. 6) occupied equivalent positions in the three models, and they engaged in identical pairing schemes. Although its base-pairing pattern changes slightly as a result of sequence changes, the terminal stem is conserved throughout the superfamily in its

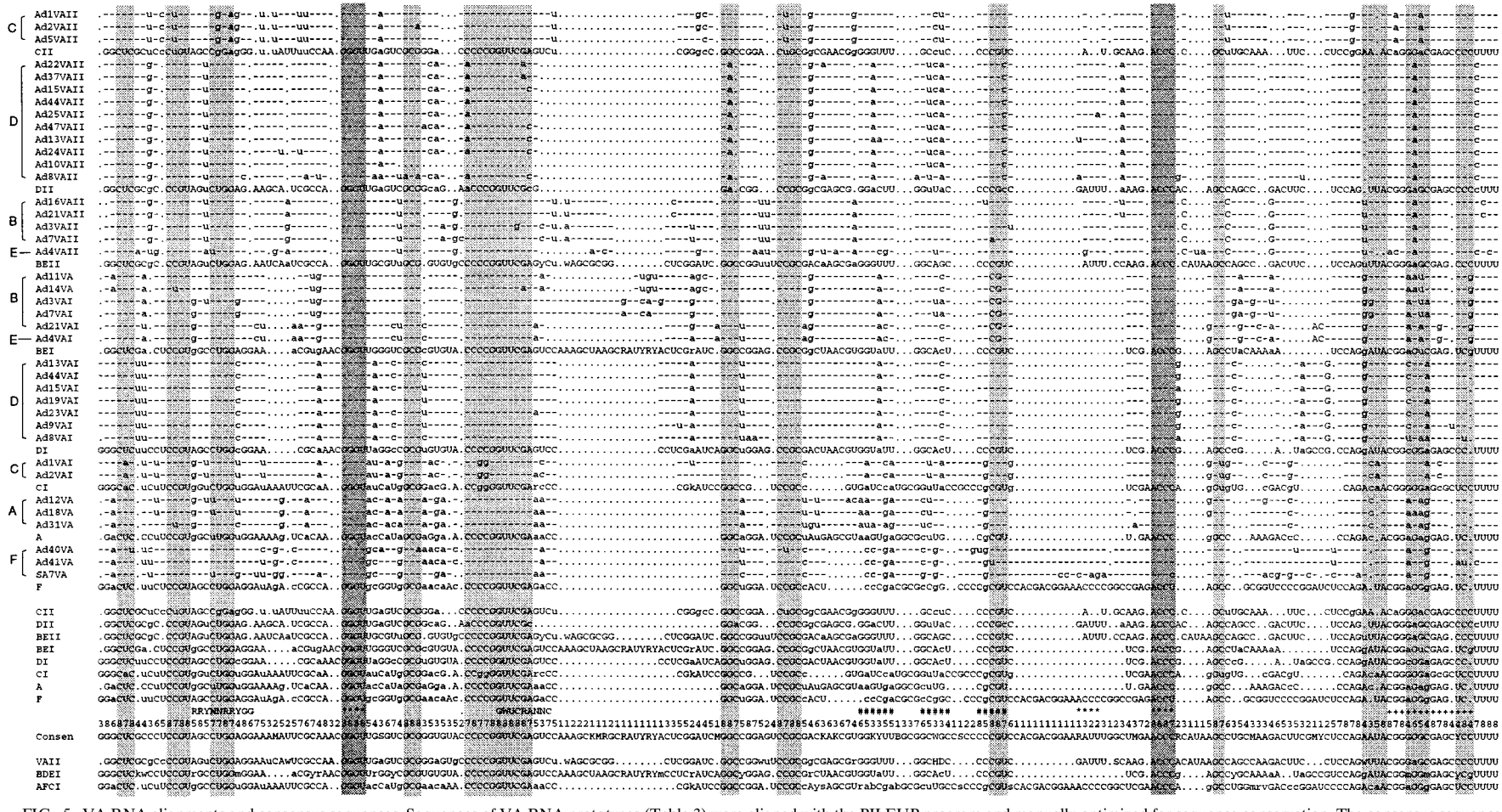

FIG. 5. VA RNA alignments and consensus sequences. Sequences of VA RNA prototypes (Table 3) were aligned with the PILEUP program and manually optimized for sequence conservation. The consensus sequence for each family is shown below the prototype VA RNA sequences. The consensus sequences for the eight families are arranged to generate a general consensus (Consen) and three superfamily consensus sequences at the bottom. Numbers shown on the line above the general consensus sequence represent the number of occurrences of each consensus nucleotide in the eight family consensus sequences. Asterisks and letters two lines above the general consensus sequence indicate the positions of the A box, GGGU sequence, B box, and ACCC sequence. Nucleotides that do not match the consensus are shown in lowercase letters. Nucleotides that match the general consensus are represented by dashes in individual VA RNA sequences and are capitalized in the family consensus sequences. Gaps are represented by dots. The code used to represent degenerate nucleotides is as follows: M for A or C; R for A or G; W for A or U; S for C or G; Y for C or U; K for G or U; V for A, C, or G; H for A, C, or U; D for A, G, or U; B for C, G, or U; and N for A, C, G, or U.

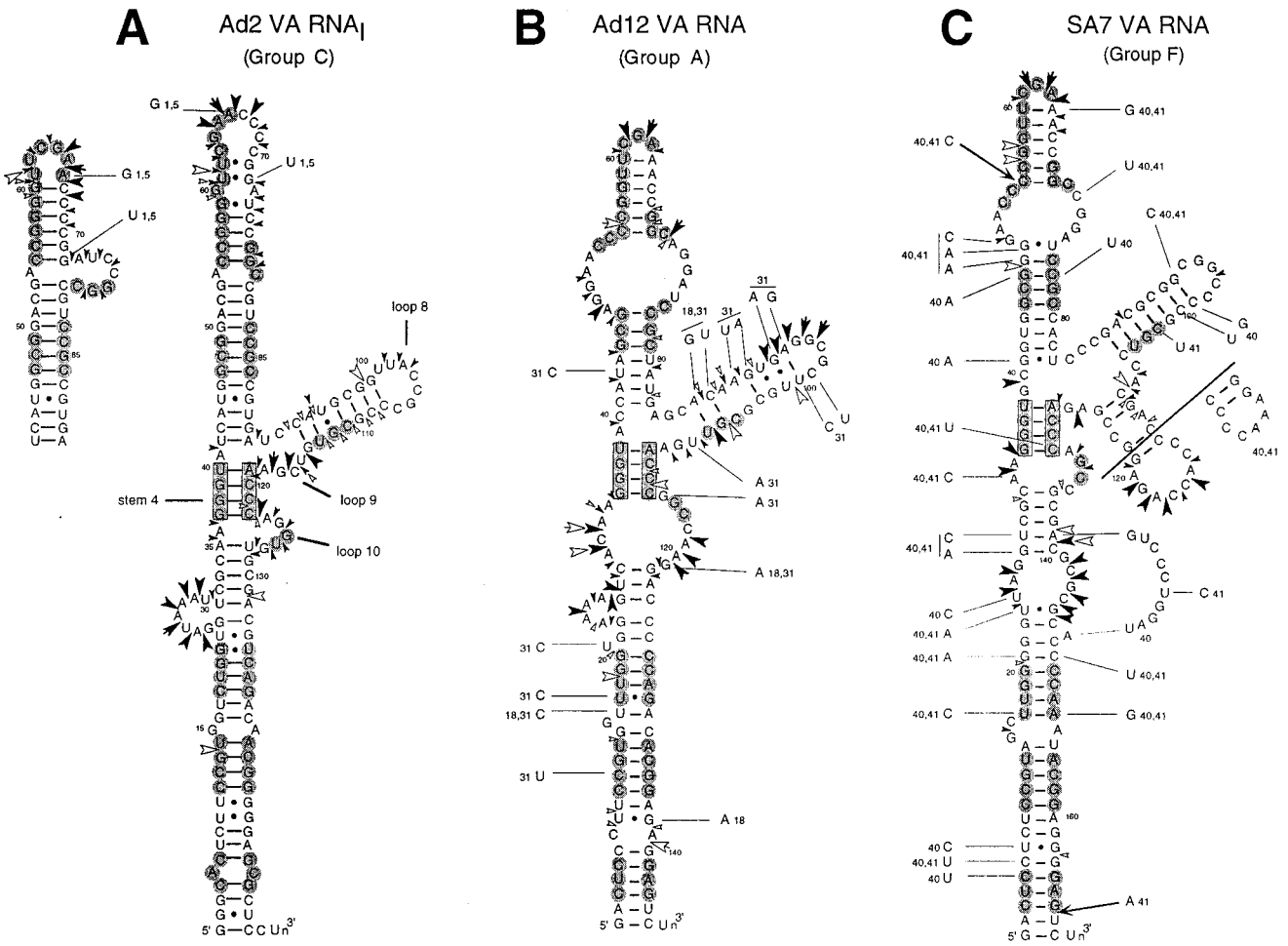

FIG. 6. Secondary structure models for VA RNA species in the A, F, C_I superfamily. Solid arrowheads indicate sites that are sensitive to single-strand-specific nucleases, and open arrowheads indicates sites that are cut by nuclease V_1. The degree of sensitivity is depicted by the size of the arrowheads; a shaft on the arrowhead indicates very strong cleavage. Nucleotide substitutions in members of the virus group are marked with a line connecting to the alternative nucleotides, which are labeled to indicate the serotype containing the substitution. Deletions, insertions, or other changes involving several nucleotides are indicated by linking the nucleotides involved with a bar.

length (20 bp) and in its largely (but not perfectly) duplexed nature.

The fundamental organization of the central domain is recognizably conserved in the A, F, C_I superfamily. Common features include (i) the conserved GGGU:ACCC stem (stem 4 [Fig. 1C]) discussed above; (ii) a stem-loop on the right of the molecule followed by a bulge (groups A and C_I) or stem-loop (group F) between the base of the apical stem and stem 4; (iii) two internal loops (loops 9 and 10) separated by a short stem of variable sequence between stem 4 and the terminal stem; and (iv) identical placement of conserved residues. However, despite the overall similarity, individual elements of the central domain vary substantially in size and sequence (Fig. 6).

The BDE_I superfamily. Members of the BDE_I superfamily exhibit much less sequence variation than those of groups A and F, so the consensus structural models derived for these RNAs (Fig. 7) are considerably simpler than those for the A, F, C_I superfamily. Their derivation required a different approach because numerous structures of similar stabilities were predicted by the folding program, and the nuclease sensitivity data derived for the VA RNA_I species of Ad7, Ad4, and Ad15 (representing groups B, E, and D, respectively) could fit several models equally well. We therefore compared all predicted structures with calculated stabilities within 10% of the most stable structure and selected for further consideration one that could accommodate all of the sequence variations within a group as well as the nuclease sensitivity data.

The resultant models (Fig. 7) closely resemble the Ad2 VA RNA_I model in overall structure and are so similar to each other that they can be combined into a single model (which is too encumbered by base substitution for illustration). In the apical and terminal stems, compensating and conservative base changes maintain base pairing and support the existence of these regions as duplexes. As in the A, F, C_I superfamily, the stems contain the conserved sequences (circled) in equivalent positions. Above the internal loop in the apical stem, the conserved tetranucleotide CCGG pairs with a downstream UCGG (Ad3 and Ad7), CCGG (group B2), or UCAGG (groups D and E and Ad16 and Ad21) in which the A is bulged out. The base of the apical stem retains its GC-rich character. In the central domain, the GGGU:ACCC tetranucleotide pair (boxed) again forms a short stem separated from the apical stem by a stem-loop on the right side of the model. The variable bulge or stem-loop that lies between these structures on the right side of the model in the A, F, C_I superfamily (loop 9 in Ad2 VA RNA_I [Fig. 1C]) is absent from the BDE_I superfamily. Beneath the

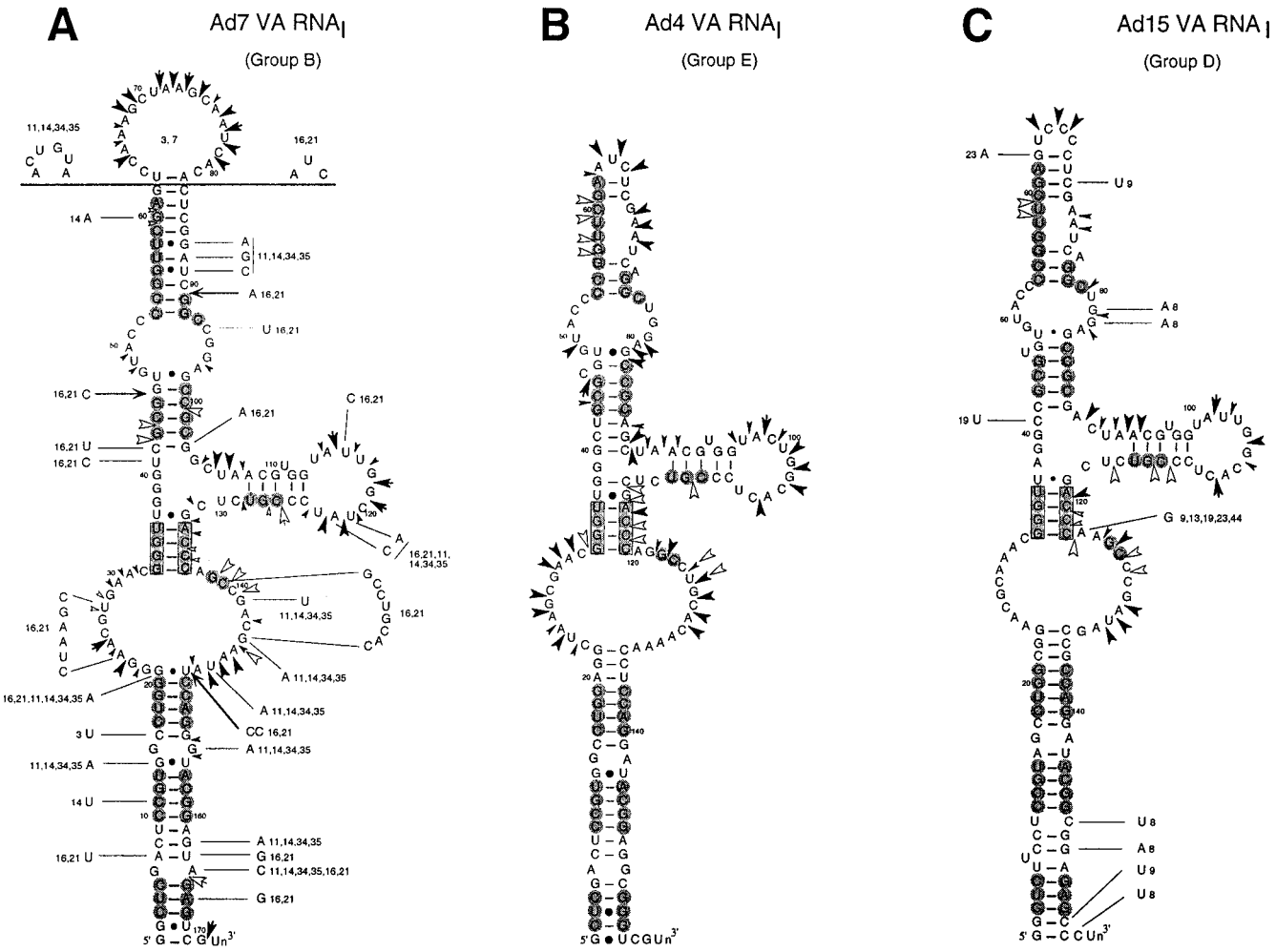

FIG. 7. Secondary structure model for VA RNA species in the BDE_I superfamily. The nuclease sensitivity data and sequence variations are labeled as in Fig. 6.

tetranucleotide pair, a large internal loop joins the central domain to the terminal stem. This internal loop occupies the same position as two smaller loops that are separated by a short stem (stem 3) in the A, F, C_I superfamily: the large internal loop appears to result from sequence changes that eliminate the pairing of stem 3 in this superfamily. The other internal and terminal loops occupy similar positions in all of the RNAs but display considerable variation in sequence and size.

The VA_{II} superfamily. Examination of the VA RNA_{II} sequences indicated that, in principle, they can all form two different structures in the central domain. One resembles the structure deduced for the other VA RNAs, as depicted in Fig. 6 and 7. In the alternative possibility, the conserved ACCC pairs with a different GGGU sequence (groups B, C, and E) or GGG sequence (group D) located about 20 nt upstream of the conserved ACCC, thereby forming a different stem and side stem-loop. Structural analysis suggests that the original structure is the correct one, however.

For the VA RNA_{II} species of Ad2, Ad4, and Ad15 (groups C, E, and D, respectively), computer-assisted secondary structure prediction, conducted with the nuclease sensitivity data taken into account, gave the models shown in Fig. 8A, B, and D. These are similar to each other and to the other VA RNAs and are compatible with most of the sequence variations observed. Alternative potential structures predicted by folding programs were much less stable and were very different from one another and from those described above (Fig. 6 and 7). On the other hand, the alternative structure was predicted for Ad7 VA RNA_{II}. This appears to be a vagary of the program, however, because the related VA RNA_{II} species of Ad16 and Ad21 were predicted to adopt the structure shown in Fig. 8C. Furthermore, the model of Fig. 8C fits the nuclease sensitivity data for Ad7 VA RNA_{II} quite well and is also compatible with the sequence variations in group B.

Although the VA RNA_{II} structures resemble one another and the other VA RNAs in overall structure and in the placement of conserved bases, there are some notable differences and discrepancies in the apical stem region. First, the paired region at the top of the apical stem is short in several of the VA RNA_{II} species. In Ad4 VA RNA_{II}, a 10-bp apical stem can be drawn, but nuclease sensitivity analysis indicated that most of these nucleotides are not paired (Fig. 8B); in group B, nucleotides are deleted from the stem in Ad16 and Ad21 (Fig. 8C); and in group D, the paired region contains only 3 bp and two of the conserved nucleotides of the CCGG sequence are unpaired (Fig. 8D). Second, there are some anomalies in the lower part of the apical stem. Unexpected single-strand-specific cleavages were observed at the base of the apical stem in Ad2 VA RNA_{II} (Fig. 8A), and the base changes in Ad8 and

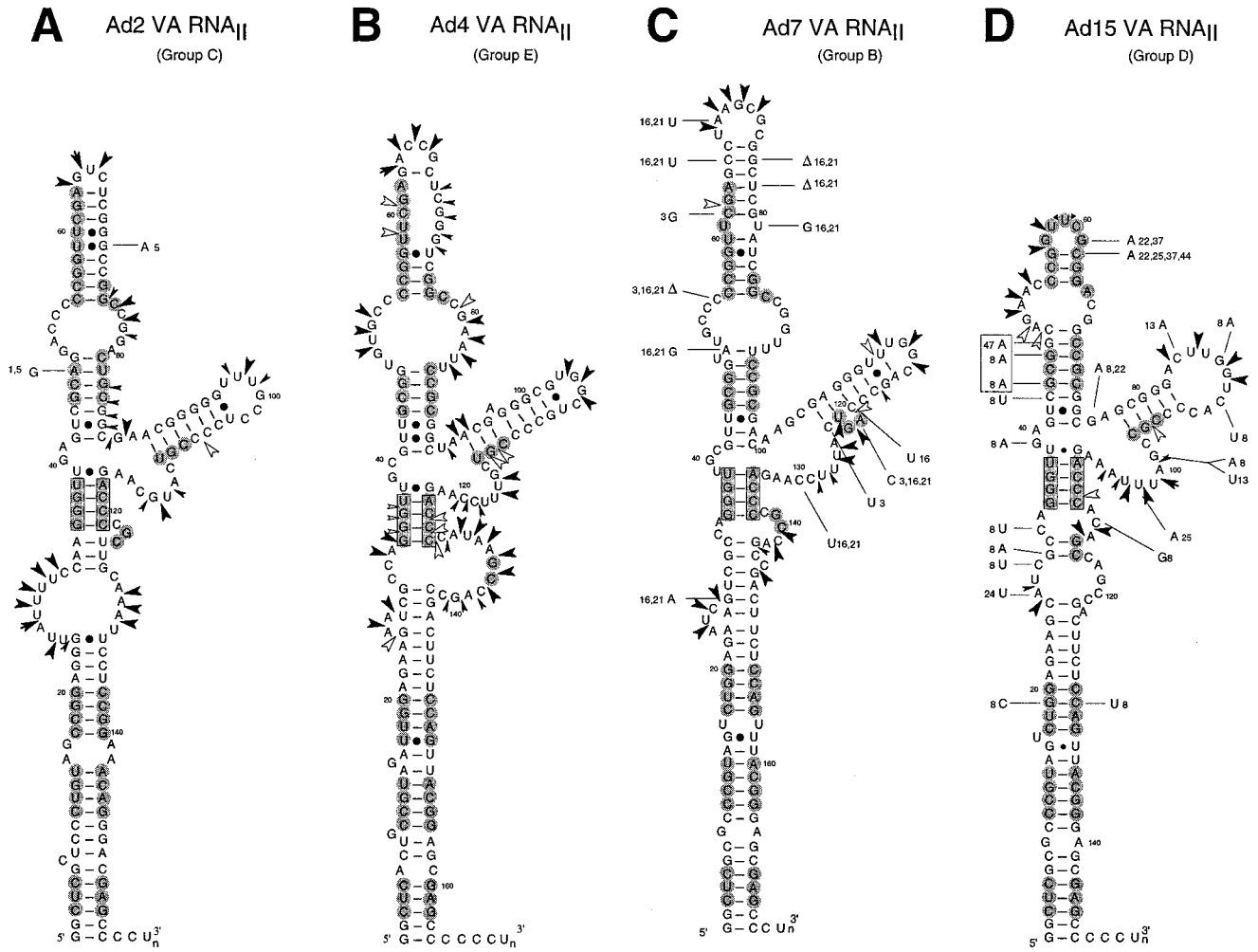

FIG. 8. Secondary structure model of VA RNA_{II} species. The nuclease sensitivity data and sequence variations are labeled as in Fig. 6.

Ad47 (boxed in Fig. 8D) are not consistent with the pairing scheme deduced for the group D species. Furthermore, in this region the VA RNA_{II} apical stems are alone in possessing G:U pairs which might weaken the duplex relative to G:C pairs. Thus, the VA RNA_{II} species do not conform as closely as the VA RNA_I and single VA RNA species to the canonical Ad2 VA RNA_I structure in the apical stem region; conceivably, these deviations in the stem might interfere with binding to PKR and provide an explanation for the inability of VA RNA_{II} to substitute for VA RNA_I in functional assays (39).

DISCUSSION

The human VA RNA genes form a rather homogeneous set in most respects (in size and location within the viral genome, for example), but the differences among them are revealing and informative. This comprehensive phylogenetic analysis of the adenovirus VA RNA genes, supplemented with nuclease sensitivity analysis of representative RNA species, was undertaken to elucidate the structure and function of these regulatory RNAs. In addition, the study sheds light on the transcription and origin of these genes as well as the evolution of adenoviruses.

Polymerase III transcriptions signals. All of the VA RNA genes but one have sequences in the 5′ half homologous to the tRNA transcription initiation elements, the A- and B-box consensus sequences RRYNNARYGG and GWTCRANNC, respectively (17, 19). The A-box homology in the VA RNA_{II} genes is generally weaker than that in the VA RNA_I genes, in accord with the finding that the A box is less important for VA RNA transcription than the B box (76). This may make these genes less competitive for binding the transcription factor TFIIIC, accounting for the lower transcription efficiency of VA RNA_{II} than of VA RNA_I as found in Ad2 (68). Supporting this idea, mutations that inactivate the VA RNA_I promoter increase the transcription of VA RNA_{II} (7). The extension of conservation of the 5′ borders of the A and B boxes may imply that the polymerase III signals are slightly more extensive than previously thought. Consistent with this interpretation, an extended box B, termed B+, has been proposed (23), but mutagenesis of Ad2 VA RNA_I did not indicate a transcriptional role for the flanking nucleotides in Ad2 VA RNA_I (76). Probably they are of structural significance (see below). The Ad10 VA RNA_{II} gene, which lacks a B box at the usual position, is apparently not transcribed (38).

At the end of the VA RNA coding sequences is a run of T residues flanked by the nucleotides C and G, typical of polymerase III termination sites (8). The number of thymidines varies from a minimum of 4 to more than 10, and A residues

are absent for at least 3 nt on either side of the T-rich run (except in Ad12 and Ad18, which have A residues in the middle of very long T runs), but no consensus sequence was discernible. Presumably the VA RNAs all terminate with several consecutive uridine residues and thereby interact with the La antigen (45). In the 5′ and 3′ ends of VA RNA sequences, the nucleotides that form the terminal stem are conserved. This could well be accounted for by the intramolecular base pairing that is responsible for the formation of the terminal stem structure, but the existence of sequence constraints imposed by transcription elements in the 5′ end is not ruled out.

In contrast to these internal transcription signals, there are few homologies in the flanking regions, suggesting that they do not harbor essential functions. Between the pTP splice site and the VA RNA genes, however, is a common A-rich sequence also noted by Kidd et al. (33). These nucleotides might be required for proper transcription initiation, since a deletion in this upstream area alters the selection of initiation sites (72). In wild-type virus, a minority of transcripts initiate 3 nt upstream from the major start site, but utilization of this site is eliminated by the deletion of nt -22 and −23 in Ad5*dl*309. The A-rich stretch is less pronounced in the spacer region upstream of the VA RNA_{II} genes, which may contribute to the weaker transcription of these species.

Structures and functions of the VA RNAs. The similarities discernible among the VA RNAs with respect to primary and secondary structure stand in contrast to the seeming sequence diversity of the first few members of the family that were examined. When their sequences are optimally aligned, no fewer than one-third of the nucleotides (54 of approximately 160) are fully or nearly fully conserved across all human adenovirus serotypes (Fig. 5). The conserved bases are clustered in blocks of 2 to 11 (typically 3 or 4) bases, are present in constant relative positions and in most cases are in mutually complementary pairs. Many of them contribute to the skeleton of the secondary structure of the molecule by forming the duplex scaffold of the terminal stem (three pairs), apical stem (two pairs), and central domain (one pair). Two additional small clusters in the central domain are not paired to one another.

The molecules all consist of the three elements of secondary structure first identified in Ad2 VA RNA_I (49). Of these, the terminal stem is the most constant in structure,length, and base composition, but no specific function has yet been assigned to it at the RNA level. The apical stem is divided by an internal loop of variable size into upper and lower portions, each containing a conserved pair of sequences. The central domain is recognizably conserved in overall structure despite little sequence conservation. Both the apical stem and central domain are critical to the function of Ad2 VA RNA_I.

Terminal stem. As described above, the terminal stem region contains transcription signals (A box; initiation and termination sites). Mutagenesis of Ad2 VA RNA_I suggests that the top part of the stem, adjacent to the central domain, is important for inhibiting PKR activation in vitro, while sequences near to the termini can be altered without loss of function (18, 24, 56). Possibly the duplex adjacent to the central domain acts as a clamp to secure the bottom part of the central domain as in the three-dimensional structure recently proposed (40); the terminal part of the duplex may play a stabilizing role in vivo by protecting the molecule from attack by exonucleases.

Apical stem. The apical stem region varies considerably in size and structure. While the lower region of duplex adjacent to the central domain is usually 7 or 8 bp long (ranging from as short as 6 bp in Ad7 VA RNA_I to as long as 12 bp in Ad2 VA RNA_I), the upper region of the apical stem-loop is highly variable in length of duplex and loop size. The internal loop is atypically small in Ad2 VA RNA_I, in which it comprises a single unmatched nucleotide on each side. In an alternative conformation of the apical stem of Ad2 VA RNA_I, which may be in equilibrium with the usual form (40), there is a 7-nt bulge loop on the 3′ side of the stem. Previous work showed that Ad2 variant 3 (46) has a 4-nt insertion into the bulge loop in this model or the apical loop in the more usual model. VA RNAs of all other groups have at least six ostensibly unmatched nucleotides on each side, but it is likely that these nucleotides are organized in some kind of compact structure because in most cases they are relatively insensitive to attack by single-strand-specific nucleases. Precedents for such compact arrangements exist in 5S RNA (21, 75) and the Rev response element RNA (4).

Consistent with its duplexed structure, the apical stem is involved in PKR binding (11, 25, 48). Protection experiments show that PKR, specifically its RNA-binding domain, makes contacts with the basal part of the apical stem, proximal to the central domain as well as the central domain itself (9, 10). Mutagenic studies implied that at least 8 bp at the base of the stem are required for the efficient binding of Ad2 VA RNA_I to PKR (11), an estimate which agrees closely with the minimum length of duplex observed in this phylogenetic survey. In mutant *dl*1, the RNA re-formed a 7-bp duplex at the bottom of the apical stem, and it is functional in a viral rescue assay but is not functional in vitro. In mutant *sub*742, nucleotides in the base of apical stem are altered, but the mutant is functional in a viral rescue assay. In this mutant, VA RNA may reorganize to generate a structure similar to that of wild-type VA RNA_I in the apical stem and central domain (18). Similarly, the sequence of the stem is less important than its duplexed nature (11, 24, 50), suggesting that its role is to provide a binding site for PKR and help stabilize the central domain structure (40). In many of the VA RNA_{II} species, features of the lower part of the apical stem raise the suspicion that it might not be as firmly duplexed as in Ad2 VA RNA_I; however, this suggestion must be considered tentative since unexpected light cleavage by single-strand-specific nucleases was also observed in both Ad12 VA RNA and Ad4 VA RNA_I.

The upper part of the apical stem can tolerate some, but not all, mutational changes (5, 18, 48, 50). Deletion or substitution of the conserved GGC nucleotides is tolerated in mutants *dl*712, A2 dl2, and *dl*1 (5, 17, 18, 49), although these bases always form part of the internal loop. The loop itself can also be eliminated without loss of function in A2 dl2 (5, 18, 24, 49).

Central domain. The GGGU:ACCC stem (stem 4), which has been shown to comprise a key element that maintains the structure of the central domain in Ad2 VA RNA_I (40), is a prominent feature of all VA RNA models (Fig. 6 to 8). Two other elements, the GC-rich side stem-loop (stem 7 and loop 8) and the loop on the 3′ border of the ACCC sequence (loop 10), are located close to stem 4, although their detailed structures vary among superfamilies. In Ad2 VA RNA_I, the side stem-loop is thought to fold and twist toward stem 4 and loop 10, interacting with the nucleotides in these structural elements, thereby adopting an upside-down ampersand structure (40). Examination of the secondary structures and sequences of other VA RNAs suggests that such interactions are possible in all cases. Consistent with the involvement of loop 10 in the interactions, several nucleotides in the 5′ end of loop 10 of all VA RNAs are insensitive to single-strand-specific nucleases. The conserved sequence GC in loop 10 of all VA RNAs (except the group C VA RNA_I species, which have GU) could be important for the interactions by pairing with the GC in loop 8. In group A and group F VA RNAs, the nucleotides GGC in

loop 8 could pair with the GCC in loop 10 and form a pseudoknot. The group C VA RNA_I species differ from other VA RNAs in having an ACC in loop 8 and a GGU in loop 10. These observations all support the idea that loops 8 and 10 interact. VA RNAs in superfamily BDE_I differ in that loop 8 is highly sensitive to single-strand-specific nucleases, suggesting that it may not be involved in the interaction. Instead, the interaction could involve nucleotides in the 3′ side of stem 7. RNAs in the VA_{II} superfamily appear to form a structure similar to those of the other VA RNAs, but there are some inconsistencies in both the apical stem and the central domain, as discussed below.

The flexibility allowing stem 7 and loop 8 in superfamilies A, F, C_I and VA_{II} to twist and bend is provided by two linking sequences. One of these is located between stems 4 and 7, corresponding to loop 9 in Ad2 VA RNA_I, and varies in size. In superfamily A, F, C_I, the group C VA RNA_I and group A VA RNA species form a bulge at this place, whereas group F VA RNA species have a stem-loop. The bulge and the stem-loop are both sensitive to nucleases, suggesting that the nucleotides might not involved in other interaction. In superfamily VA_{II}, loop 9 is 7 to 9 nt in length, but only half of the nucleotides near stem 7 are sensitive to nucleases (Fig. 8): the remaining nucleotides may be involved in different interactions. Although VA RNAs in superfamily BDE_I lack loop 9, they have a big internal loop below stem 4 (loop 10) which could provide the flexibility for nucleotides in loop 10 to interact with loop 8. The second linking sequence is located between stem 7 and the base of the apical stem. This sequence is short and heterogeneous, and insertion of two bases between nt 92 and 93 in Ad2 VA RNA_I (VAI-CB [24]) is not deleterious to VA RNA function in vitro. In contrast, mutant *pm*91 was nonfunctional, presumably because it disrupted the base of the apical stem (56).

The fourth conserved sequence in the central domain is CGU, located on the 3′ side of stem 7 at the edge of the central domain in the tertiary structure model (40). Nuclease protection experiments indicated that this sequence in Ad2 VA RNA_I is not protected by PKR, and so these nucleotides probably do not interact directly with PKR (10). There are five stretches of sequences which are protected by PKR (10), and none of them is conserved. Probably the position of a nucleotide in the three-dimensional structure, rather than the primary sequence, determines its interaction with PKR. For example, in Ad2 VA RNA_I, changing C-111, which is not protected by PKR, to an A abolished VA RNA_I function (10, 56). However, substituting the adjacent nucleotides had no effect on VA RNA_I function (56). It seems that C-111 is important for the tertiary structure of the central domain, which is required for binding to PKR, or for the inhibition of PKR. These two properties of the central domain may be contributed by two separate subdomains, as suggested by certain mutations. Some mutations, such as *dl*103-109, reduced the binding activity to PKR and lead to an increase in the concentration of VA RNA required to prevent PKR activation, but the autophosphorylation of PKR drops sharply in response to the elevated concentration of the mutant RNA, similar to the response to wild-type Ad2 VA RNA_I (11). Other mutants, such as *is*3 and *is*5, were not significantly reduced in binding ability but inhibited PKR activation only at very high concentrations (11, 50); in these cases, the autophosphorylation of PKR gradually declines with the increased VA RNA input, and the mutants seem to exert their function by competing with the dsRNA activator as seen with high concentrations of short dsRNA molecules that are unable to activate the enzyme.

The enigma of VA RNA_{II}. While the majority of adnoviruses carry two VA RNA genes, a substantial minority (9 of 47) have only one. When there is only a single species, it is invariably of the VA RNA_I type, implying that the VA RNA_{II} function, if there is one, is less essential. Supporting this view, there is evidence that in several viruses the VA RNA_{II} genes have been lost or inactivated. In group B, four viruses (Ad11, Ad14, Ad34, and Ad35; subgroup B2) encode only one VA RNA, whereas the other four (Ad3, Ad7, Ad16, and Ad21; subgroup B1) each have two VA RNAs. The VA RNAs in the B2 group are highly homologous to the VA RNA_I species of the B1 group. It seems that the VA RNA_{II} gene was lost from the B2 group almost in its entirety by deletion. The VA RNA_{II} gene of the group D virus Ad10 is present but inactivated by a B-box deletion (38), and similar VA RNA_{II} deletion is also found in a group E virus, Ad4a (33). Thus, VA RNA_{II} is apparently dispensable in certain viruses.

VA RNA_{II} suppresses PKR activation poorly. Ad2 VA RNA_{II} has little influence on virus growth in the absence of VA RNA_I (73), and both Ad2 and Ad7 VA RNA_{II} are only weakly active, if at all, in assays for PKR function (39). Correspondingly, the VA RNA_{II} species differ from the other VA RNAs in several structural respects. First, nuclease sensitivity at the base of the Ad2 VA RNA_{II} apical stem suggests that this region may not be stably paired. Second, although several nucleotides in loop 8 of VA RNA_{II} could in principle interact with loop 10, they are sensitive to single-strand-specific nucleases. Third, the size of loop 10 in group C and group D VA RNA_{II} species is only 3 nt, which may be insufficient to accommodate a stable interaction with loop 8. Fourth, in Ad4 VA RNA_{II}, the loop 10 nucleotides which could be involved in base pairing are 6 nt away from the conserved ACCC, instead of 1 or 2 nt as in the other two superfamilies. Finally, although the nuclease sensitivity patterns of loops 8 and 10 in Ad7 VA RNA_{II} are similar to those in group A and F VA RNAs, suggesting that these RNAs may form similar tertiary interactions, the insensitivity of some the loop 9 sequences (CCAGG) to nucleases indicates that these nucleotides might be involved in other interactions. Taken together, these observations indicate that VA RNA_{II} may have a different structure in those regions of the molecule that are important for interactions with PKR. This difference might explain the inability of VA RNA_{II} to inhibit PKR, but could suit it for another function.

Consistent with this hypothesis, the three branches of the VA_{II} superfamily are relatively close to one another in comparison with the branches within the other VA RNA superfamilies (Fig. 4). This affinity could implicate these RNAs in an as yet unidentified function that enhances virus growth, perhaps in a tissue-specific fashion. Viruses in groups A and F, which contain one VA RNA, infect the gastroenteric tract (1, 22) whereas the other adenoviruses cause respiratory tract infection or keratoconjunctivitis (52, 61, 71, 74). Furthermore, subgroup B2 viruses seem to infect a broader spectrum of organs than subgroup B1 viruses (64). Therefore, it seems likely that VA RNA_{II} subserves a function that is different from that of VA RNA_I and is dispensable in the gastroenteric tract. The nature of this hypothetical function is a matter of speculation.

Origin and evolution of VA RNA. The VA RNA sequences fall into eight closely related families and three more distantly related superfamilies (Fig. 3 to 5). Within a superfamily, the individual serotype groups retain their identities and display greater or lesser degrees of similarity with other groups in the same superfamily. A detailed analysis of the sequence alignments indicates that the homology is not uniform throughout the VA RNA molecule. Most of the sequence divergences are

located in its 3′ half, especially in the sequences forming the central domain and the apical stem-loop, whereas the 5′ side contains highly conserved transcription elements. An evolutionary tree produced by comparing the more highly homologous 5′ half sequences had a general topology similar to that in Fig. 4, but the lengths and relative positions of the branches differed (data not shown). From this analysis, it appears that the ancestors of the VA RNA_I and VA RNA_{II} species of groups B, D, and E are evolutionarily relatively close to each other, suggesting that the VA RNA_I and VA RNA_{II} species might be the products of a duplication which occurred before the segregation of groups B, D, and E. During evolution, mutations may have accumulated in the 3′ halves of these VA RNAs, causing them to separate further in the tree generated with full-length VA RNAs. Ad4 (group E) is closely related to groups B and D in its VA RNAs, as in other respects (3), but the spacers between the VA RNA genes are not clearly homologous. Presumably mutations have accumulated more rapidly in the functionally less essential spacer sequences since the segregation of group E from group B and D. Gene conversion and recombination mechanisms cannot be ruled out, however Such events may explain why groups C and D are closely allied in a phylogenetic tree generated with fiber gene sequences (3) although their relationship is more distant in the VA RNA comparisons.

The two VA RNAs in group C may also be products of gene duplication, perhaps an event separate from that in groups B, D, and E, because a stretch of nucleotides immediately upstream of VA RNA_{II} is homologous to a stretch upstream of VA RNA_I. Since the homology between these two RNAs in group C is relatively weak, the presumptive duplication may have occurred early in adenovirus evolution. Indeed, the closeness of the group C VA RNA_{II} species to other VA RNA_{II} species in the evolutionary tree suggests that they share a recent common ancestor, whereas the group C VA RNA_I diverged at an earlier time. Kidd et al. (33) have proposed that the VA RNA genes of group C evolved in separate virus lineages and were brought together by recombination. While the VA RNA_I species of group C and the VA RNAs of groups A and F resemble the single VA RNA of monkey adenoviruses, the VA RNA_{II} species of group C and the VA RNAs of groups B, D, and E resemble chimpanzee VA RNAs. Thus, it is possible that the upstream homology noted above is a relic of putative substitution event that replaced the original VA RNA_I gene in a progenitor of the group C viruses.

A clue to the origin of the VA RNAs comes from the observation that the transcription control region in the 5′ half of Ad2 VA RNA_I is similar to the corresponding region of several tRNAs (6, 17). To various extents, the 5′ portions of other VA RNAs are homologous to different tRNA genes (not shown). These homologies are limited to the sequence CCGGTTCGA around the B box and the T stem-loop of the homologous tRNA. Similar homologies have been found in the family B repetitive sequence in the mouse (36), the ID repeat in the rat (51), and the short repetitive elements in the primate *Galago* sp. (12), each of which is thought to have arisen through the integration of a tRNA pseudogene into an *Alu* repeat (12, 13). There are two weaker B-box homologies in all VA RNA genes, one in the middle of the gene and another near its 3′ end, especially evident in the VA RNA_{II} genes. This finding raises the possibility that the VA RNAs originated from a tandem repeat of tRNA genes through a series of deletion or unequal recombination events. In such tandem repeats, the nucleotides between the tRNA-homologous patches are highly heterogeneous, suggesting that different VA RNAs may come from different origins. Compatible with this hypothesis, the pairing of the nucleotides near the 5′ and 3′ ends to form the terminal stem is a property that may be inherited from tRNA, whereby nucleotides in the 5′ end pair to the B-box region near the 3′ end to form the acceptor stem.

In summary, the evolution of VA RNA may have occurred in three steps, the first involving the merger of several tRNA genes to form an ancestral VA RNA and the second leading to the formation of VA RNA_I and VA RNA_{II} genes by gene duplication. These events could have taken place on more than one occasion during the evolution of adenoviruses. Finally, in some viruses, the VA RNA_{II} gene seems to have been lost or inactivated as a result of deletions.

ACKNOWLEDGMENTS

We thank Tsafrira Pe'ery for advice and critical discussions.

This work was supported by grant CA13106 from the National Cancer Institute.

REFERENCES

1. **Adrian, T., and R. Wigand.** 1989. Genome type analysis of adenovirus 31, a potential causative agent of infants' enteritis. Arch. Virol. **105:**81–87.
2. **Akusjärvi, G., M. B. Mathews, P. Andersson, B. Vennström, and U. Pettersson.** 1980. Structure of genes for virus-associated RNA_I and RNA_{II} of adenovirus type 2. Proc. Natl. Acad. Sci. USA **77:**2424–2428.
3. **Bailey, A., and V. Mautner.** 1994. Phylogenetic relationships among adenovirus serotypes. Virology **205:**438–452.
4. **Bartel, D. P., M. L. Zapp, M. R. Green, and J. W. Szostak.** 1991. HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. Cell **67:**529–536.
5. **Bhat, R. A., P. H. Domer, and B. Thimmappaya.** 1985. Structural requirements of adenovirus VAI RNA for its translation enhancement function. Mol. Cell. Biol. **5:**187–196.
6. **Bhat, R. A., B. Metz, and B. Thimmappaya.** 1983. Organization of the noncontiguous promoter components of adenovirus VAI RNA gene is strikingly similar to that of eucaryotic genes. Mol. Cell. Biol. **3:**1996–2005.
7. **Bhat, R. A., and B. Thimmappaya.** 1984. Adenovirus mutants with DNA sequence perturbations in the intragenic promoter of VAI RNA gene allow the enhanced transcription of VAII RNA gene in HeLa cells. Nucleic Acids Res. **12:**7377–7388.
8. **Bogenhagen, D. F., and D. D. Brown.** 1981. Nucleotide sequences in Xenopus 5S DNA required for transcription termination. Cell **24:**261–270.
9. **Circle, D. A., O. D. Neel, P. Clarke, L. Manche, M. B. Mathews, and H. D. Robertson.** Unpublished data.
10. **Clarke, P. A., and M. B. Mathews.** 1995. Interactions between the double-stranded RNA binding motif and RNA: definition of the binding site for the interferon-induced protein kinase DAI (PKR) on adenovirus VA RNA. RNA **1:**7–20.
11. **Clarke, P. A., T. Pe'ery, Y. Ma, and M. B. Mathews.** 1994. Structural features of adenovirus 2 virus-associated RNA required for binding to the protein kinase DAI. Nucleic Acids Res. **22:**4364–4374.
12. **Daniels, G. R., and P. L. Deininger.** 1983. A second major class of Alu family derived from mammalian tRNA genes. Nucleic Acids Res. **11:**7595–7610.
13. **Daniels, G. R., and P. L. Deininger.** 1991. Characterization of a third major SINE family of repetitive sequences in the galago genome. Nucleic Acids Res. **19:**1649–1656.
14. **England, T. E., and O. C. Uhlenbeck.** 1978. 3′-terminal labelling of RNA with T4 RNA ligase. Nature (London) **275:**560–561.
15. **Engler, J. A., M. S. Hoppe, and M. P. van Bree.** 1983. The nucleotide sequence of the genes encoded in early region 2b of human adenovirus type 7. Gene **21:**145–159.
16. **Felsenstein, J.** 1993. PHYLIP (phylogeny interface package) version 3.5c. Distributed by the author and available through anonymous ftp to evolution.genetics.washington.edu.
17. **Fowlkes, D. M., and T. Shenk.** 1980. Transcriptional control regions of the adenovirus VAI RNA gene. Cell **22:**405–413.
18. **Furtado, M. R., S. Subramanian, R. A. Bhat, D. M. Fowlkes, B. Safer, and B. Thimmappaya.** 1989. Functional dissection of adenovirus VAI RNA. J. Virol. **63:**3423–3434.
19. **Galli, G., H. Hofstetter, and M. L. Birnstiel.** 1981. Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. Nature (London) **294:**626–631.
20. **Garon, C. F., K. W. Berry, J. C. Hierholzer, and J. A. Rose.** 1973. Mapping of base sequence heterologies between genomes from different adenovirus serotypes. Virology **54:**414–426.
21. **Garrett, R. A., and S. O. Olesen.** 1982. Structure of eukaryotic 5S ribonucleic acid: a study of *Saccharomyces cerevisiae* 5S ribonucleic acid with ribonucleases. Biochemistry **21:**4823–4830.

22. **Gary, G. W. J., J. C. Hierholzer, and R. E. Black.** 1979. Characteristics of noncultivable adenoviruses associated with diarrhea in infants: a new subgroup of human adenoviruses. J. Clin. Microbiol. **10:**96–103.
23. **Geiduschek, E. P., and G. P. Tocchini-Valentini.** 1988. Transcription by RNA polymerase III. Annu. Rev. Biochem. **57:**873–914.
24. **Ghadge, G. D., P. Malhotra, M. R. Furtado, R. Dhar, and B. Thimmapaya.** 1994. In vitro analysis of virus-associated RNA1 (VAI RNA): inhibition of the double-stranded RNA-activated protein kinase PKR by the VAI RNA mutants correlates with the in vivo phenotype and the structural integrity of the central domain. J. Virol. **68:**4137–4151.
25. **Ghadge, G. D., S. Swaminathan, M. G. Katze, and B. Thimmapaya.** 1991. Binding of the adenovirus VAI RNA to the interferon-induced 68 kDA protein kinase correlates with function. Proc. Natl. Acad. Sci. USA **88:**7140–7144.
26. **Gilbert, D. G.** 1990. MulFold, a Macintosh version of MFOLD (version 2.0) by Michael Zuker and John Jaeger for prediction of RNA secondary structure by free energy minimization. Available via anonymous ftp to iubio-.bio.indiana.edu.
27. **Green, S. R., L. Manche, and M. B. Mathews.** 1995. Two functionally distinct RNA binding motifs in the regulatory domain of the protein kinase DAI. Mol. Cell. Biol. **15:**358–364.
28. **Guilfoyle, R., and R. Weinmann.** 1981. Control region for adenovirus VA RNA transcription. Proc. Natl. Acad. Sci. USA **78:**3378–3382.
29. **Hein, J.** 1990. TreeAlign, a multiple sequence alignment program that builds tree as it align DNA or protein sequences. Version 90, May. Available through anonymous ftp to ftp.embl-heidelberg.de.
30. **Hein, J.** 1994. TreeAlign. Methods Mol. Biol **25:**349–364.
31. **Huang, J. T., and R. J. Schneider.** 1990. Adenovirus inhibition of cellular protein synthesis is prevented by the drug 2-aminopurine. Proc. Natl. Acad. Sci. USA **87:**7115–7119.
32. **Kaufman, R. J., and P. Murtha.** 1987. Translational control mediated by eukaryotic initiation factor 2 is restricted to specific mRNAs in transfected cells. Mol. Cell. Biol. **7:**1568–1571.
33. **Kidd, A. H., D. Garwicz, and M. Oberg.** 1995. Human and simian adenoviruses: phylogenetic inferences from analysis of VA RNA genes. Virology **207:**32–45.
34. **Kitajewski, J., R. J. Schneider, B. Safer, S. M. Munemitsu, C. E. Samuel, B. Thimmappaya, and T. Shenk.** 1986. Adenovirus VAI RNA antagonizes the antiviral action of interferon by preventing activation of the interferon-induced eIF-2a kinase. Cell **45:**195–200.
35. **Kostura, M., and M. B. Mathews.** 1989. Purification and activation of the double-stranded RNA-dependent eIF-2 kinase DAI. Mol. Cell. Biol. **9:**1576–1586.
36. **Krayev, A. S., T. V. Markuseva, D. A. Kramerov, A. P. Ryshov, K. G. Skyrabin, A. A. Bayev, and G. P. Georgiev.** 1982. Ubiquitous transposon-like repeats B1 and B2 of mouse genome: B2 sequencing. Nucleic Acids Res. **10:**7461–7475.
37. **Larsson, S., C. Svensson, and G. Akusjärvi.** 1986. Characterization of a low-molecular-weight virus-associated (VA) RNA encoded by simian adenovirus type 7 which functionally can substitute for adenovirus type 5 VA RNAI. J. Virol. **60:**635–644.
38. **Liao, H., and M. B. Mathews.** Unpublished data.
39. **Ma, Y., and M. B. Mathews.** 1993. Comparative analysis of the structure and function of adenovirus virus-associated RNAs. J. Virol. **67:**6605–6617.
40. **Ma, Y., and M. B. Mathews.** Secondary and tertiary structure in the central domain of adenovirus type 2 VA RNA_I. Submitted for publication.
41. **Maran, A., and M. B. Mathews.** 1988. Characterization of the double-stranded RNA implicated in the inhibition of protein synthesis in cells infected with a mutant adenovirus defective for VA RNA_I. Virology **164:**106–113.
42. **Mathews, M. B.** 1975. Genes for VA-RNA in adenovirus 2. Cell **6:**223–229.
43. **Mathews, M. B.** 1993. Viral evasion of cellular defense mechanisms: regulation of the protein kinase DAI by RNA effectors. Semin. Virol. **4:**247–257.
44. **Mathews, M. B.** 1996. Interactions between viruses and the cellular machinery for protein synthesis, p. 505–548. *In* J. W. B. Hershey, M. B. Mathews, and N. Sonenberg (ed.), Translational control. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
45. **Mathews, M. B., and A. M. Francoeur.** 1984. La antigen recognizes and binds to the 3′-oligouridylate tail of a small RNA. Mol. Cell. Biol. **4:**1134–1140.
46. **Mathews, M. B., and T. Grodzicker.** 1981. Virus-associated RNAs of naturally occurring strains and variants of group C adenoviruses. J. Virol. **38:**849–862.
47. **Mathews, M. B., and T. Shenk.** 1991. Adenovirus virus-associated RNA and translational control. J. Virol. **65:**5657–5662.
48. **Mellits, K. H., M. Kostura, and M. B. Mathews.** 1990. Interaction of adenovirus VA RNA_I with the protein kinase DAI: non-equivalence of binding and function. Cell **61:**843–852.
49. **Mellits, K. H., and M. B. Mathews.** 1988. Effects of mutations in stem and loop regions on the structure and function of adenovirus VA RNA_I. EMBO J. **7:**2849–2859.
50. **Mellits, K. H., T. Pe'ery, and M. B. Mathews.** 1992. Role of the apical stem in maintaining the structure and function of adenovirus virus-associated RNA stem. J. Virol. **66:**2369–2377.
51. **Milner, R. J., F. E. Bloom, C. Lai, R. A. Lerner, and J. G. Sutcliffe.** 1984. Brain specific genes have identifier sequences in their introns. Proc. Natl. Acad. Sci. USA **81:**713–717.
52. **O'Donnell, B., E. A. McCruden, and U. Desselberger.** 1993. Molecular epidemiology of adenovirus conjunctivitis in Glasgow 1981–1991. Eye **7:**8–14.
53. **O'Malley, R. P., R. F. Duncan, J. W. B. Hershey, and M. B. Mathews.** 1989. Modification of protein synthesis initiation factors and shut-off of host protein synthesis in adenovirus-infected cells. Virology **168:**112–118.
54. **O'Malley, R. P., T. M. Mariano, J. Siekierka, and M. B. Mathews.** 1986. A mechanism for the control of protein synthesis by adenovirus VA RNA_I. Cell **44:**391–400.
55. **Pe'ery, T., K. H. Mellits, and M. B. Mathews.** 1993. Mutational analysis of the central domain of adenovirus VA RNA mandates a revision of the proposed secondary structure. J. Virol. **67:**3534–3543.
56. **Rahman, A., P. Malhotra, R. Dhar, T. Kewalramani, and B. Thimmapaya.** 1995. Effect of single-base substitutions in the central domain of VAI RNA on its function. J. Virol. **69:**4299–4307.
57. **Reich, P. R., J. Rose, B. Forget, and S. M. Weissman.** 1966. RNA of low molecular weight in KB cells infected with Ad2. J. Mol. Biol. **17:**428–439.
58. **Reichel, P. A., W. C. Merrick, J. Siekierka, and M. B. Mathews.** 1985. Regulation of a protein synthesis initiation factor by adenovirus virus-associated RNAI. Nature (London) **313:**196–200.
59. **Rohan, R. M., and G. Ketner.** 1987. A comprehensive collection of point mutations in the internal promoter of the adenovirus VA_I gene. J. Biol. Chem. **262:**8500–8507.
60. **Schmedt, C., D. R. Taylor, L. Manche, S. R. Green, Y. Ma, and M. B. Mathews.** 1995. Functional characterization of the RNA binding domain and motif of the double-stranded RNA-dependent protein kinase DAI. J. Mol. Biol. **249:**29–44.
61. **Schmitz, H., R. Wigand, and W. Heinrich.** 1983. Worldwide epidemiology of human adenovirus infections. Am. J. Epidemiol. **117:**455–466.
62. **Schneider, R. J.** 1996. Adenovirus and vaccinia virus translational control, p. 575–605. *In* J. W. B. Hershey, M. B. Mathews, and N. Sonenberg (ed.), Translational control. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
63. **Schneider, R. J., C. Weinberger, and T. Shenk.** 1984. Adenovirus VAI RNA facilitates the initiation of translation in virus-infected cells. Cell **37:**291–298.
64. **Sharp, I. R., and G. Wadell.** 1994. Adenoviruses, p. 287–308. *In* A. J. Zuckerman, J. E. Banatvala, and J. R. Pattison (ed.), Principles and practice of clinical virology. John Wiley & Sons, Chichester, England.
65. **Shu, L. M., J. S. Hong, Y. F. Wei, and J. A. Engler.** 1986. Nucleotide sequence of the genes encoded in early region 2b of human adenovirus type 12. Gene **46:**187–195.
66. **Siekierka, J., T. M. Mariano, P. A. Reichel, and M. B. Mathews.** 1985. Translational control by adenovirus: lack of virus-associated RNA_I during adenovirus infection results in phosphorylation of initiation factor eIF-2 and inhibition of protein synthesis. Proc. Natl. Acad. Sci. USA **82:**1959–1963.
67. **Snouwaert, J., D. Bunick, C. Hutchison, and D. M. Fowlkes.** 1987. Large numbers of random point and cluster mutations within the adenovirus VA I gene allow characterization of sequences required for efficient transcription. Nucleic Acids Res. **15:**8293–8330.
68. **Söderlund, H., U. Pettersson, B. Vennström, L. Philipson, and M. B. Mathews.** 1976. A new species of virus-coded low molecular weight RNA from cells infected with Ad2. Cell **7:**585–593.
69. **Svensson, C., and G. Akusjärvi.** 1985. Adenovirus VA RNA_I mediates a translational stimulation which is not restricted to the viral mRNAs. EMBO J. **4:**957–964.
70. **Svensson, C., and G. Akusjärvi.** 1990. A novel effect of adenovirus VA RNA_I on cytoplasmic mRNA abundance. Virology **174:**613–617.
71. **Takafuji, E. T., J. C. Gaydos, R. G. Allen, and F. H. J. Top.** 1979. Simultaneous administration of live, enteric-coated adenovirus types 4, 7 and 21 vaccines: safety and immunogenicity. J. Infect. Dis. **140:**48–53.
72. **Thimmappaya, B., N. Jones, and T. Shenk.** 1979. A mutation which alters initiation of transcription by RNA polymerase III on the Ad5 chromosome. Cell **18:**947–954.
73. **Thimmappaya, B., C. Weinberger, R. J. Schneider, and T. Shenk.** 1982. Adenovirus VAI RNA is required for efficient translation of viral mRNAs at late times after infection. Cell **31:**543–551.
74. **Wadell, G., M. L. Hammarskjöld, G. Winberg, T. M. Varsanyi, and G. Sundell.** 1980. Genetic variability of adenoviruses. Ann. N.Y. Acad. Sci. **354:**16–42.
75. **Wimberly, B., G. Varani, and I. Tinoco, Jr.** 1993. The conformation of loop E of eukaryotic 5S ribosomal RNA. Biochemistry **32:**1078–1087.
76. **Wu, G. J., J. F. Railey, and R. E. Cannon.** 1987. Defining the functional domains in the control region of the adenovirus type 2 specific VA RNA1 gene. J. Mol. Biol. **194:**423–442.
77. **Zuker, M.** 1989. On finding all suboptimal foldings of an RNA molecule. Science **244:**48–52.
78. **Zuker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9:**133–148.