Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 2216–2220, April 1982
Biochemistry

Complete sequence analyses show that two defective interfering influenza viral RNAs contain a single internal deletion of a polymerase gene

(viral interference/DNA sequence analysis/influenza virus/defective interfering virus)

DEBI P. NAYAK, N. SIVASUBRAMANIAN, ALAN R. DAVIS, RITA CORTINI, AND JAE SUNG

Department of Microbiology and Immunology, UCLA School of Medicine, Los Angeles, California 90024

Communicated by Norman Davidson, January 4, 1982

ABSTRACT **Defective interfering (DI) influenza viral RNAs arise by internal deletion of progenitor RNAs. By using recombinant DNA cloning and DNA sequence analysis techniques, we have deduced the complete sequence of two such RNAs (L2b and L3), both arising from the same polymerase (*P1*) gene of WSN influenza virus. We have also partially determined the sequence of the *P1* polymerase gene, including the sequence at the point of deletion and the flanking regions. Our sequence study shows the following. (*i*) Both L2b and L3 arise by a simple deletion in the *P1* gene. (*ii*) L2b and L3 are 683 and 441 nucleotides long, respectively. (*iii*) The first 413 and 244 nucleotides of the 5′ ends of L2b and L3, respectively, are identical to those of the 5′ end of the *P1* gene. (*iv*) The last 270 nucleotides of L2b and 197 nucleotides of L3 are the same as those of the 3′ end of the *P1* gene. (*v*) The entire sequence of L3 is present in the sequence of L2b. (*vi*) Both the 5′ and the 3′ termini, including the transcription stop and poly(A) addition signals of the progenitor *P1* gene, are present in both L2b and L3. (*vii*) The sequences at the deletion point and the flanking region of the *P1* gene do not resemble the consensus splicing sequence of spliced mRNA suggesting that a replicational event rather than splicing is involved in the formation of influenza defective interfering RNAs.**

Influenza viruses produce defective interfering (DI) virus particles when passaged at high multiplicity (1, 2, 3). Influenza DI viruses contain, in addition to other viral RNA segments, small RNA molecules (DI RNA) not present in standard viruses (4). These DI RNA molecules have been shown to be responsible for interference (5) and have been found to be of widespread occurrence and present in many stock virus preparations (6). All influenza DI RNAs studied to date have been shown to arise from either *P1*, *P2*, or *P3* polymerase genes (4, 7). A single polymerase gene can give rise to DI RNAs of multiple size (7, 8). Furthermore, each influenza DI RNA studied by us has been shown to arise from internal deletion of a single polymerase gene in which both the 3′ and the 5′ termini are preserved (9).

In our attempt to understand the biological and biochemical properties of DI influenza viruses, including the generation of influenza DI RNA, DI virus-mediated interference, and the role of DI influenza viruses in the natural history of virus evolution including antigenic variation, we have undertaken a detailed structural analysis of both DI and viral RNA segments. In this paper, we report the complete sequences of two DI RNAs (L2b and L3) and compare them to the partial sequence of the *P1* gene (the progenitor of both L2b and L3).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Virus and Cells. A/WSN/33 strain of influenza virus was grown as in Madin–Darby bovine kidney (MDBK) cells as described (4, 10). RNA was isolated from ts52 virus, a group 2 temperature-sensitive mutant (11) of the WSN strain and used for recombinant DNA cloning of viral polymerase genes (*P1*, *P2*, and *P3*). To obtain DNA clones of influenza DI RNA segments, RNA was isolated from purified DI virus prepared from virus clone L, which gives rise to four major DI RNAs, L1, L2a, L2b, and L3 (8–10). L2b and L3, the two DI RNA segments reported in this paper, are both of *P1* origin and are ≈700 and 400 nucleotides long, respectively (8, 9).

Recombinant DNA Cloning. Procedures used for cloning of double-stranded DNA forms of polymerase RNA and DI RNAs were essentially as described (12). Avian myeloblastosis virus reverse transcriptase (Life Sciences, St. Petersburg, FL) was used to prepare a cDNA copy of either ts52 viral RNA that had been enriched for the polymerase genes (12) or of total DI/L RNA (8, 9). Single-stranded cDNA species corresponding in size to the polymerase genes or L2b or L3 RNAs were isolated from alkaline agarose gels (13). Double-stranded DNA was prepared from single-stranded DNA by using the Klenow fragment of DNA polymerase, treated with nuclease S1, fractionated on a 1.4% neutral agarose gel as described (12), and cloned in *Escherichia coli* 294 using pBR322 as vector (14). Plasmid DNA from individual clones was isolated (15). Subsequently, DNA from each clone was denatured, applied to nitrocellulose filters, and hybridized with either a ^{32}P-labeled specific DI RNA or a nick-translated insert of a known polymerase or other influenza gene (16).

Clones containing the *P1* gene were identified by their size [≈2200 base pairs (bp)] and by hybridization to L2b or L3, as these RNAs are easily separable and of *P1* origin (8). In this way, clones 1-39b and 1-72b, containing inserts of ≈2200 bp and 2400 bp, respectively, were found to be of *P1* gene origin. Similarly, clones D2-8 and D2-62 were identified as clones of L3 and clone D1-3 was of L2b.

All work was done in accordance with National Institutes of Health guidelines for recombinant DNA research.

DNA Sequence Analyses. DNA sequence analyses were carried out by the procedure of Maxam and Gilbert (17). The sequence at the 5′ end of the viral RNA that was absent in the DNA insert was determined by primer extension (18). In the case of the *P1* gene, a standard viral RNA preparation free from visible DI RNA was used as a template to extend a primer

Abbreviations: DI, defective interfering (RNA); bp, base pair(s).

(*Hinf*I–*Bgl* II isolated from 1-72b) uniquely labeled at the *Hinf*I site (see Fig. 1A). Similarly, the sequence of L3 RNA was completed by using a *Taq* I–*Alu* I restriction fragment (see Fig. 1C) as the primer and DI/L RNA as the template. cDNA prepared by this method was analyzed by Maxam and Gilbert procedure (17).

RESULTS

Sequence Analyses of L2b, L3, and *P1* Gene RNAs. The *P1* gene is ≈2450 nucleotides long and its sequence has been partially determined (Figs. 1A and 2). The inserts of clones 1-39b and 1-72b used in the sequence analyses are ≈2200 and 2400 bp long, respectively. Both clones are complete at the 3′ end of the viral RNA but lack various amounts of the 5′ end. The sequence at the 5′ end was completed by extending the *Hinf*I–*Bgl* II fragment labeled at the *Hinf*I (Fig. 1A) site.

The L2b sequence was determined essentially as described above but using the insert of clone D1-3. The restriction sites and the strategy used in the sequence analysis are shown in Fig. 1B and the complete sequence is shown in Fig. 2. Clone D1-3 is complete at the 3′ end and contains all but the last 13 nucleotides from the 5′ end, which was obtained by directly determining the RNA sequence (9).

The L3 sequence was determined by analyzing inserts of clones D2-8 and D2-62 and completed by primer extension (Figs. 1C and 2). Both clones were complete at the 3′ end of the viral RNA but clone D2-8 lacked 32 nucleotides and clone D2-62 lacked 31 nucleotides at the 5′ end. The complete sequence at the 5′ end of the viral RNA was obtained by extension of the *Taq* I–*Alu* I primer uniquely labeled at the 5′ end of the *Taq* I site (Fig. 1C). Thus, both the 5′ and the 3′ termini and the rest of the sequences of *P1*, L2b, and L3 were determined independently. The terminal sequences in all cases contain the common 5′- and 3′-end sequences reported previously (19) and are similar to the terminal sequences of DI RNA obtained previously by direct RNA sequence analyses (9) and to the partial sequence of the *P1* gene of fowl plague virus (20).

Sequence Relationships Among L2b and L3 RNAs and the *P1* Gene. L2b and L3 RNAs are 683 and 441 nucleotides long and correlate fairly well to the size expected from gel analysis

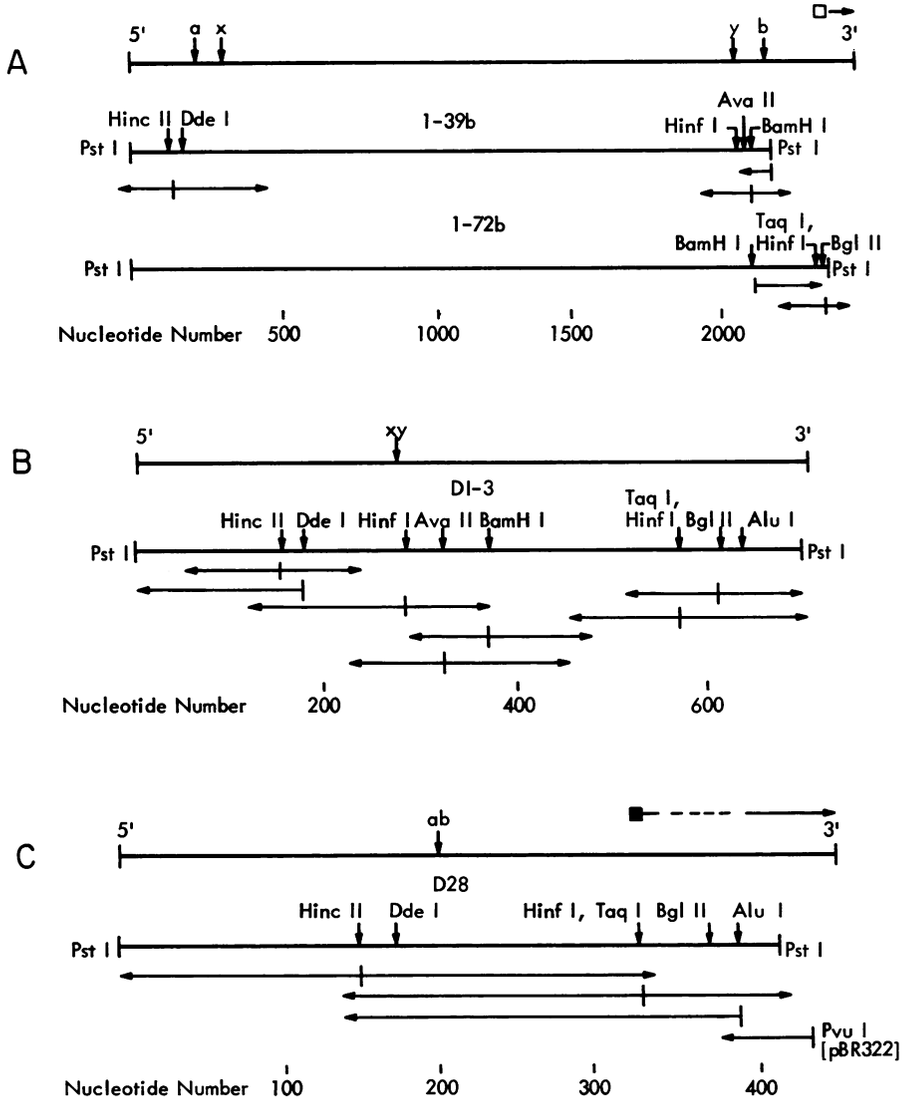

FIG. 1. Strategy used in sequence analyses of cloned *P1* RNA (*A*), L2b RNA (*B*), and L3 RNA (*C*). ↔, Composite sequence information obtained from multiple gels. Sequences at the 5′ ends of *P1* and L3 RNAs were completed by primer extension. The *Hinf*I–*Bgl* II primer (□→) was extended by using reverse transcriptase to complete the 5′-*P1* sequences missing in the DNA clones. The *Taq* I–*Alu* I primer (■→) was extended to complete the 5′-L3 sequence that was absent in DNA clones (*Hinf*I/*Taq* I is a part of the same site.) ↓: a and b, junction points in the *P1* gene of DI L3 sequences; x and y, junction points of DI L2b sequences.

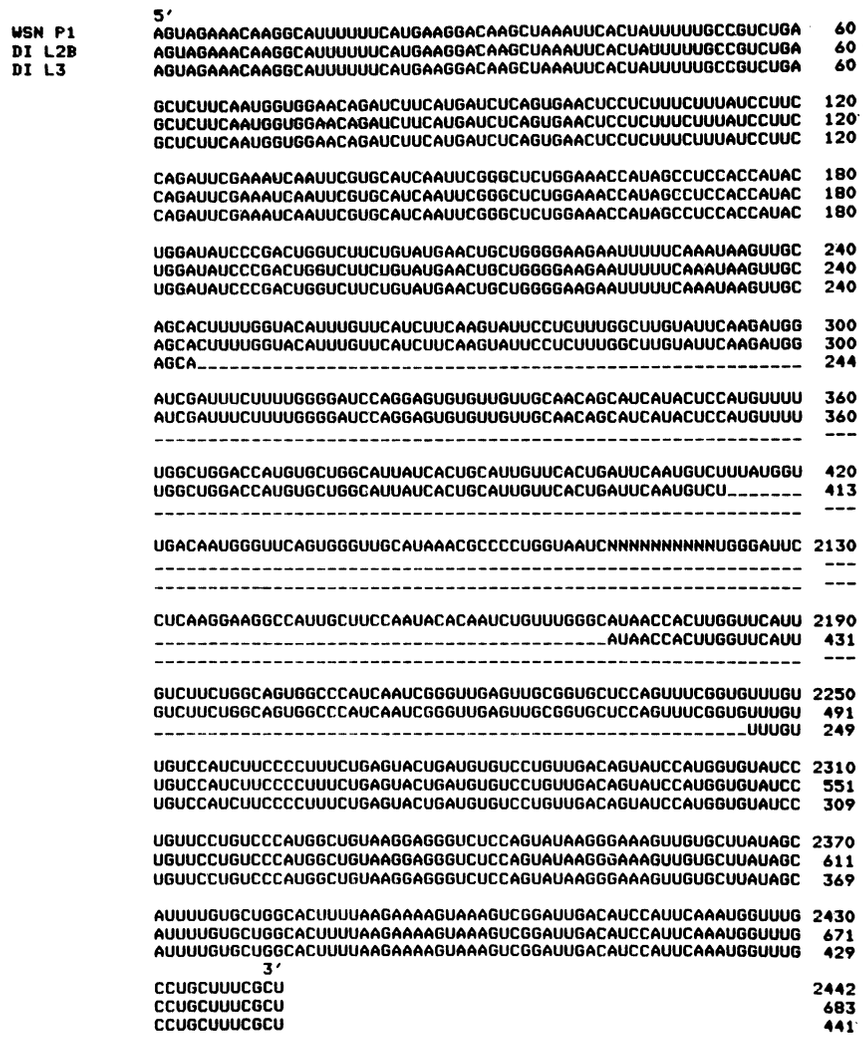

FIG. 2. Complete nucleotide sequences of L2b and L3 and partial sequence of *P1* gene (minus strand) viral RNAs. ---, Regions that are absent in L2b and L3 RNAs; N-N-N-N-N-N-N, undetermined sequence of ≈1660 nucleotides. Numbers indicate numbers of nucleotides; those for the *P1* gene are approximate. (See *Note Added in Proof.*)

(8). We have shown earlier by oligonucleotide mapping and direct sequence analyses of 5′ and 3′ RNA termini that L2b and L3 RNAs contain the 5′ and 3′ ends of the *P1* gene and that L2b RNA contains all the oligonucleotides present in L3 RNA. However, these earlier studies would not have recognized any altered RNA sequence due to either base change or deletion unless it affected the unique T1 oligonucleotides. In addition, the exact location of deletion, the extent of deletion, or the number of deletions could not be determined.

The present study shows that, in both cases, a single deletion of varying size gives rise to L2b and L3 RNAs. In L2b viral RNA, 413 nucleotides at the 5′ end and 270 nucleotides at the 3′ end are identical to those of the *P1* gene. Similarly, L3 RNA arises from the fusion of 244 nucleotides of 5′ end to 197 nucleotides of 3′ end of the *P1* gene. Our sequence data also show that there is not a single mismatch in the homologous region of L2b, L3, and *P1* gene RNAs—even though the *P1* clones were obtained from a different preparation.

Based on the size, charge, and sequence information we can identify and order six unique oligonucleotide spots in L3 RNA (8). These are 5′-I11-H5-↓-I6-J15-D1-G4-3′ (↓, fusion point). These oligonucleotides are present in all DI RNAs—i.e., B1, P1, L2b, and ts$^+$1—of *P1* gene origin (8). Thus, as in L2b RNA, the entire sequence of L3 RNA is likely to be present in all larger DI RNAs arising from the *P1* gene, suggesting that some of these sequences may have an important functional role in either the viral RNA replication or assembly and maturation.

DISCUSSION

Based on structure, DI RNAs can be classified as 5′ DI RNA, 3′ DI RNA, or 5′,3′ DI RNA depending on whether one or both termini of the genomic RNA is preserved in the DI RNA (21). In addition, a class of mosaic (or compound) DI RNA that contains additional structural rearrangements has been demonstrated (22, 23, 24). The majority if not all of influenza DI RNAs are of the 5′,3′ type (3), and the sequences reported here establish that two influenza DI RNAs are overlapping DI RNAs and arise from a simple deletion in the genomic RNA without a single mismatch or a structural rearrangement. The primary structures of these influenza DI RNAs are, therefore, clearly different from the majority of other positive- and negative-strand DI RNAs. An α virus (positive-strand) DI RNA, whose complete nucleotide sequence has been published recently, contains multiple deletions, duplications, and structural rearrangement (24). On the other hand, the majority of the rhabdovirus or paramyxovirus (negative-strand) DI RNAs are 5′ DI RNAs of panhandle or snapback type while a few belong to the 5′,3′ type, or the mosaic type (21, 25, 26, 27). However, the precise locations and numbers of deletions, as well as the nu-

Table 1. Deletion sites, junction sites, and 5′ and 3′ termini of L2b and L3 RNAs

Site	RNA	Sequence
Deletion	L3 Viral	5′ U-G-G-A-G-C-A ↓ C-U-U-U-U-G-G . . . U-U-U-C-G-G-U-G ↓ U-U-U-G-U-U-G 3′
	L3 Complementary	3′ A-C-C-U-C-G-U ↓ G-A-A-A-A-C-C . . . A-A-A-G-C-C-A-C ↓ A-A-A-C-A-A -C 5′
Junction	L3 Viral	5′ U-G-G-A-G-C-A ↓ U-U-U-G-U-U-G 3′
	L3 Complementary	3′ A-C-C-U-C-G-U ↓ A-A-A-C-A-A -C 5′
Deletion	L2b Viral	5′ A-A-U-G-U-C-U ↓ U-U-A-U-G-G-U . . . U-U-U-G-G-G-C ↓ A-U-A-A-C-C- A 3′
	L2b Complementary	3′ U-U-A-C-A-G-A ↓ A-A-U-A-C-C-A . . . A-A-A-C-C-C-G ↓ U-A-U-U-G-G-U 5′
Junction	L2b Viral	5′ A-A-U-G-U-C-U ↓ A-U-A-A-C-C -A 3′
	L2b Complementary	3′ U-U-A-C-A-G-A ↓ U-A-U-U-G-G-U 5′
Termini	Viral	5′ A-G-U-A-G-A-A-A-C-A-A- G . . . C-C-U-G-C-U-U-U-C-G-C-U 3′
	Complementary	3′ U-C-A-U-C-U-U-U-G-U-U-G . . . G-G-A-C-G-A-A-A-G-C-G-A 5′

cleotide sequences at the junction points or at the flanking regions of the parenteral genome involved in deletion have not been determined.

A complete sequence analysis also provides some insight into the possible mechanism of the genesis of influenza DI RNAs and the mode of interference. For example, it is unlikely that RNA splicing is involved in the generation of influenza DI RNAs because the consensus donor or acceptor sequences (5′ . . . A-G ↓ G-U . . . A-G ↓ G-N; ref. 28) are absent at the deletion points of either viral or complementary RNA of *P1* gene (Table 1). On the other hand, an aberrant replicational event that involves omission of a portion of a template RNA during the replication of either the viral or the complementary RNA is a likely possibility (21). In this model, a polymerase(s) that has an attached nascent chain may detach and reinitiate either at a specific sequence or at a secondary structure or may simply roll over to another site due to the juxtaposition of distant RNA sequences produced by secondary structure (21). Furthermore, the sequence data rules out that influenza DI RNAs must contain a constant amount of either the 5′ end or the 3′ end of the genomic RNA (Fig. 3 *A* and *B*) but supports that both the 5′ end and the 3′ end can vary (Fig. 3*C*). Although both L2b and L3 RNAs contain a single deletion in the genomic RNA, the possibility of multiple deletions that have been reported for positive-strand viral DI RNAs (29) cannot be ruled out for other DI RNAs. In addition, although such events are rare, this model does not rule out strand switch (i.e., plus to minus or vice versa) or a segment switch (i.e., from one viral RNA segment to another viral RNA segment in a segmented viral genome) in the generation of DI RNA (22, 23).

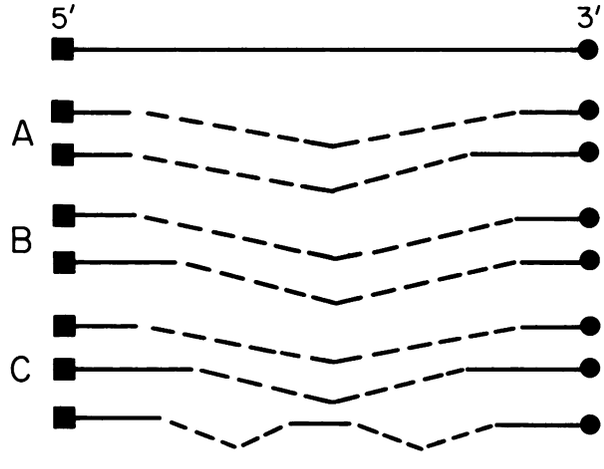

FIG. 3. Generation of 5′,3′ DI RNAs. ■ and ●, 5′ and 3′ ends of progenitor viral RNA (*Upper*); ——, region of progenitor viral RNA that has been deleted; ——, region of progenitor RNA that has been preserved. (*A*) Constant amount of RNA 5′ sequence and varying amount of RNA 3′ sequence. (*B*) Varying amount of RNA 5′ sequence and constant amount of RNA 3′ sequence. (*C*) Varying amounts of both RNA 5′ and RNA 3′ sequences.

The mechanism by which DI viruses interfere remains unclear. It has been proposed that, since 5′ DI RNAs lack the proper 3′ leader sequence required for transcribing messages both *in vitro* (30) and *in vivo* (31), they replicate more efficiently and, therefore, interfere competitively with the genomic RNA (32, 33). However, both influenza DI RNAs (L2b and L3) contain a significant amount of the 5′ and 3′ terminal sequences of the *P1* gene, including the proposed transcription termination and poly(A) addition signals (U-U-U-U-U-U, position 16–21; ref. 34). Whether these DI RNAs can be transcribed in infected cells is as yet unknown. If DI mRNAs are transcribed, these and other 5′,3′ DI RNAs may interfere by a mechanism different from the one proposed for 5′ DI RNAs (32). If, on the other hand, they are not transcribed, structures other than the termini of viral RNA are required for transcription.

The coding potentials of L2b and L3 plus-strand RNAs in all three reading frames are shown in Fig. 4. Only one reading frame (1, Fig. 4) beginning with AUG (methionine) codon (position 657–659, DI L2b, and position 415–417, DI L3; see Fig. 2) remains open up to the junction point. Similarly, after the junction point, only one reading frame is open (3, DI L3, and 2, DI L2b; Fig. 4). A complete sequence analysis of *P1* gene would determine whether these uninterrupted reading frames really belong to a single reading frame and code for the P1 protein. However, since L2b and L3 RNAs do not maintain the same reading frame after junction point, they are unlikely to produce a common functional protein.

The complete sequence analysis of the influenza DI RNAs reported here and that of the α virus DI RNA (24) show that the primary structure of the DI viral genome may vary widely and, furthermore, suggest that this variation in the structure

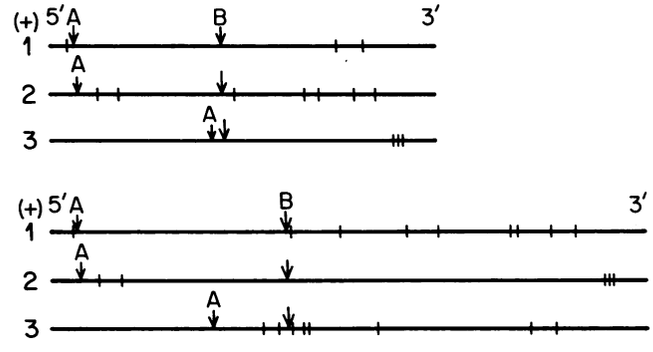

FIG. 4. Putative translation reading frames of L2b (*Lower*) and L3 (*Upper*) DI complementary RNAs. |, Position of terminator on DI complementary RNA. ↓: A, first AUG codon in reading frame; B, point of deletion.

may reflect different modes in the genesis of DI RNAs and in the mechanism of interference.

Note Added in Proof. The complete (2341 nucleotides) sequence of the *P1* gene has been determined recently. L2b has a 1658-nucleotide-long deletion (position 414–2071) and L3 has a 1900-nucleotide-long deletion (position 245–2144) of the *P1* gene.

We thank Dr. M. Ueda for valuable discussion. This study was supported by research grants from the National Science Foundation (PCM 7823220) and the National Institute of Allergy and Infectious Diseases (AI 12749 and AI 16348).

1. von Magnus, P. (1954) *Adv. Virus Res.* **2**, 59–78.
2. Huang, A. S. & Baltimore, D. (1977) *Compr. Virol.* **10**, 73–116.
3. Nayak, D. P. (1980) *Annu. Rev. Microbiol.* **34**, 619–644.
4. Nayak, D. P., Tobita, K., Janda, J. M., Davis, A. R. & De, B. K. (1978) *J. Virol.* **28**, 375–386.
5. Janda, J. M. & Nayak, D. P. (1979) *J. Virol.* **32**, 697–702.
6. De, B. K. & Nayak, D. P. (1980) *J. Virol.* **36**, 847–859.
7. Ueda, M., Nakajima, K. & Sugiura, A. (1980) *J. Virol.* **34**, 1–8.
8. Davis, A. R. & Nayak, D. P. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3092–3096.
9. Davis, A. R., Hiti, A. L. & Nayak, D. P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 215–219.
10. Janda, J. M., Davis, A. R., Nayak, D. P. & De, B. K. (1979) *Virology* **95**, 48–58.
11. Sugiura, A., Ueda, M., Tobita, K. & Enomoto, C. (1975) *Virology* **65**, 363–375.
12. Davis, A. R., Hiti, A. L. & Nayak, D. P. (1980) *Gene* **10**, 205–218.
13. McDonell, M. W., Simon, M. M. & Studier, F. W. (1977) *J. Mol. Biol.* **110**, 119–146.
14. Seeburg, P. H., Shine, J., Martial, J. D., Baxter, J. D. & Goodman, H. M. (1977) *Nature (London)*, **270**, 486–494.
15. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.
16. Southern, E. (1975) *J. Mol. Biol.* **98**, 503–517.
17. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
18. Hiti, A. L., Davis, A. R. & Nayak, D. P. (1981) *Virology* **111**, 113–124.
19. Desselberger, U., Racaniello, U. R., Zazra, J. J. & Palese, P. (1980) *Gene* **8**, 315–328.
20. Robertson, J. R. (1979) *Nucleic Acids Res.* **6**, 3745–3757.
21. Lazzarini, R. A., Keene, J. D. & Schubert, M. (1981) *Cell* **26**, 145–154.
22. Keene, J. D., Chien, I. M. & Lazzarini, R. A. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2090–2094.
23. Moss, B. A. & Brownlee, G. G. (1981) *Nucleic Acids Res.* **9**, 1941–1947.
24. Lehtovaara, P., Söderlund, H., Keränen, S., Pettersson, R. & Kääriäinen, L. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 5353–5357.
25. Kolakofsky, D. (1976) *Cell* **8**, 547–555.
26. Epistein, D. A., Herman, R. C., Chien, I. & Lazzarini, R. A. (1980) *J. Virol.* **33**, 818–829.
27. Perrault, J. & Semler, B. L. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 6191–6195.
28. Sharp, P. A. (1981) *Cell* **23**, 643–646.
29. Lundquist, R. E., Sullivan, N. & Maizel, J. V. (1979) *Cell* **18**, 759–769.
30. Emerson, S. U., Dierks, P. M. & Parsons, T. J. (1977) *J. Virol.* **23**, 708–716.
31. Rao, R. D. & Huang, A. S. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3742–3745.
32. Perrault, J., Semler, B., Leavitt, R. W. & Holland, J. J. (1978) in *Negative Strand Viruses and Host Cells*, eds. Mahy, B. W. J. & Barry, R. D. (Academic, New York), pp. 527–538.
33. Huang, A. S. (1977) *Bacteriol. Rev.* **41**, 811–821.
34. Robertson, J. R., Schubert, M. & Lazzarini, R. A. (1981) *J. Virol.* **38**, 157–163.