Proc. Natl. Acad. Sci. USA
Vol. 77, No. 1, pp. 210–214, January 1980
Biochemistry

Cloning DNA sequences from influenza viral RNA segments

(cDNA strands/cleavage maps/nonstructural protein/matrix protein/hemagglutinin)

CHING-JUH LAI, LEWIS J. MARKOFF, SUSAN ZIMMERMAN, BRONNA COHEN, JO ANN BERNDT, AND ROBERT M. CHANOCK

Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20205

Contributed by Robert M. Chanock, October 5, 1979

ABSTRACT **DNA sequences corresponding to gene segments that code for the nonstructural protein, the matrix protein, and the hemagglutinin of influenza A virus [strain A/Udorn/72 (H3N2)] were cloned in *Escherichia coli* pBR 322. Initially, positive and negative cDNA strands were prepared separately by reverse transcription. The positive strands of cDNA were transcribed from genomic RNA segments by using a specific dodecamer DNA sequence as a primer; the negative strands of cDNA were transcribed from cytoplasmic viral mRNA segments by using an oligo(dT) primer. DNA duplexes corresponding in size to the virus RNA segments were then purified, inserted into the plasmid DNA, and used for transformation of *E. coli*. The influenza virus-specific DNA sequences isolated from recombinant plasmid molecules were characterized by mapping restriction enzyme cleavage sites. In addition, the orientation of cloned DNA was determined with reference to the 3′ terminus of viral RNA.**

The negative-stranded RNA genome of influenza A virus contains eight segments that vary from 0.2 to 0.9×10^6 daltons (1–4). Analysis of virion RNA from influenza A strains diverse in antigenic subtype indicates that all eight RNA segments contain a common sequence of 13 nucleotides at the 3′ terminus and another common sequence of 12 nucleotides at the 5′ terminus (5, 6). In the infected cell, these RNA segments are transcribed and polyadenylylated to generate the corresponding positive-stranded mRNA species (7, 8). With the exception of viral RNA segment 8 which encodes two nonstructural proteins (NS_1 and NS_2), each RNA segment seems to code for a specific viral component and to be responsible for a specific viral function (9–11). The virion components include the polymerase proteins (P1, P2, P3), nucleoprotein (NP), matrix protein (M), and the two surface proteins, the hemagglutinin (H) and the neuraminidase (N) (12, 13). The functional role which these viral proteins play in viral replication and maturation is only partially understood. It is also unclear how the requisite eight RNA segments are selected for assembly into infectious virions. Above all, many questions remain to be answered concerning genetic variation that results in the emergence of new viral subtypes responsible for pandemics or epidemics. To approach this problem we have used recombinant DNA techniques to examine the nucleotide sequences of the viral genomic segments. In this paper, we describe the isolation of influenza virus recombinant DNA clones corresponding to several viral genes and their characterization by restriction enzyme cleavage.

MATERIALS AND METHODS

Virus Strains. Influenza A/Udorn/72 (H3N2) virus was used as the source of viral RNA (vRNA) for cloning gene segments. In this study we also analyzed cDNA products from reverse transcription of vRNAs of two other subtypes, A/FM-1/47 (H1N1) and A/Taiwan/62 (H2N2).

Preparation of Influenza Virus RNA Segments. Infection of 8- to 9-day embryonated eggs with influenza virus, purification of virus particles from the allantoic fluid, and extraction of vRNA were carried out as described (14). Infection of MDCK cells with virus and subsequent treatment of cells to enhance cytoplasmic RNA synthesis were performed as described (15). Procedures for preparation of cytoplasmic RNA and separation of the poly(A)-containing RNA fraction by affinity chromatography on oligo(dT)-cellulose columns were as described (16).

Synthesis of cDNA Strands. Conditions for the reverse transcriptase reaction were as detailed by Myers *et al.* (17). The reaction mixture (1 ml) contained: vRNA, 50 μg; dodecamer DNA [d(A-G-C-A-A-A-A-G-C-A-G-G)] (prepared by Collaborative Research), 5 μg; dCTP, dTTP, dATP, and dGTP, each at 0.2 mM; [α-^{32}P]dCTP at a final specific activity of 0.5 Ci/mmol (1 Ci = 3.7×10^{10} becquerels); sodium pyrophosphate, 4 mM; AMV reverse transcriptase (obtained from J. Beard, Life Sciences, Gulfport, FL), 200 units; and buffer (final concentrations) made of 50 mM Tris·HCl at pH 8.3, 45 mM KCl, 8 mM $MgCl_2$, and 0.4 mM dithiothreitol. Cytoplasmic influenza mRNA (50 μg) was reverse transcribed under similar conditions by using 10 μg of oligo(dT) as primer instead of the DNA dodecamer. Incubation was carried out at 37°C for 1 hr, followed by heating at 70°C in 0.2 M NaOH for 30 min to remove RNA. After neutralization and phenol extraction the labeled cDNA products were purified by gel filtration on Sephadex G-50.

Cloning Influenza Virus Gene Segments in *Escherichia coli* Plasmid pBR 322. Positive (+) cDNA strands transcribed from vRNA and negative (−) cDNA strands transcribed from mRNA, approximately 10 μg each, were mixed and hybridized by heating at 68°C for 3 hr in 10 ml of 0.14 M phosphate, pH 6.8/1 mM EDTA/0.1% sodium dodecyl sulfate. The DNA duplexes containing influenza virus-specific sequences were selected by hydroxyapatite chromatography and separated according to size by electrophoresis in 3.5% polyacrylamide gel (acrylamide/bisacrylamide = 20:1) at 3.5 V/cm for 40 hr. Ten to 30 deoxycytidylyl residues were added to the 3′ end of the influenza-specific DNA fragments eluted from the gel, by using terminal transferase (18). Similarly, deoxyguanidylyl residues of the same length were added to the *Pst* I-cleaved pBR 322 DNA. The recombinant DNA molecules were constructed by hybridization of the two components and were used to transform *E. coli* K-12 strain HB101 (19). Tetracycline-resistant colonies were analyzed for the presence of influenza-specific sequences by the technique of Grunstein and Hogness (20) with ^{32}P-labeled influenza (+) cDNA as probe. These cloning experiments were conducted under P1 containment as prescribed in the National Institutes of Health guidelines.

Experimental Design. The approach used to obtain double-stranded DNA molecules from bacterial phage Qβ RNA

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.

Abbreviation: vRNA, viral RNA.

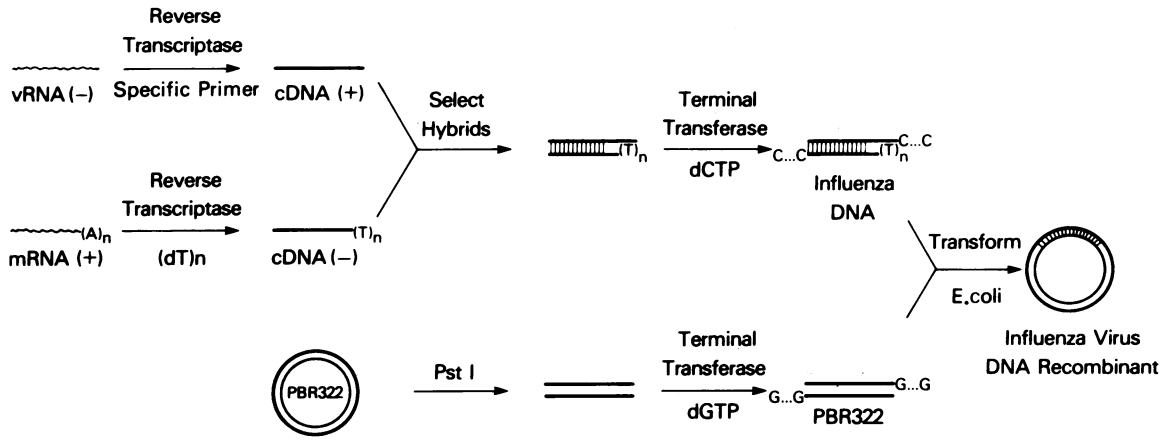

FIG. 1. Scheme for construction and cloning of influenza virus DNA.

was adapted for cloning influenza viral gene segments (21). cDNA strands of influenza virus were separately prepared from the virion RNA and cytoplasmic viral mRNA segments by using the AMV reverse transcriptase (Fig. 1). After hybridization, the duplex DNA molecules were purified and separated by electrophoresis according to size. The selected influenza-specific DNA fragments were inserted into plasmid pBR 322 at the *Pst* I site by utilizing the dG/dC joining technique (22). The recombinant DNA molecules were used for transformation of *E. coli* and the transformants were analyzed for the presence of influenza viral DNA-specific sequences by using ^{32}P-labeled total (+) cDNA as a probe. We assumed that cellular repair enzymes would convert the unpaired regions at the termini of the constructed DNA to yield a complete duplex molecule. DNA recombinants corresponding to several individual influenza genes were isolated by using this approach.

RESULTS

Synthesis of Influenza Virus DNA Strands. We used a dodecamer DNA sequence, d(A-G-C-A-A-A-A-G-C-A-G-G), complementary to the 3′ terminus of vRNA segments as a specific primer to initiate reverse transcription to yield the (+) cDNA segments. The (−) strands of cDNA were similarly transcribed from poly(A)-containing mRNA, isolated from the virus-infected cells, in an oligo(dT)-primed synthesis. Both cDNA products were analyzed by electrophoresis on alkaline agarose gels.

The electrophoregram presented in Fig. 2 shows the (+) cDNA sequences synthesized from vRNAs of three major human influenza A virus subtypes including A/FM-1/47 (H1N1), A/Taiwan/62 (H2N2), and A/Udorn/72 (H3N2). Reverse transcription in the presence of the specific primer yielded eight major (+) cDNA fragments, each of which corresponded to a genomic RNA segment in its gel migration pattern (3). One additional prominent cDNA band (approximately 480 nucleotides), probably derived from defective genomic sequences, was produced in the reverse transcription of A/FM-1/47 vRNA. Gel electrophoresis of (−) cDNA synthesized from the mRNA of A/Udorn/72 wild-type virus yielded a pattern that resembled that seen with the (+) cDNA segments. The first three bands in this gel (Fig. 2, lane a) were faint but could be discerned with certainty. The (−) cDNA segments corresponded in size to the cDNA transcripts of vRNA. Analysis by hybridization with a (+) cDNA probe showed that these labeled (−) cDNA bands were virus-specific. The estimated molecular size of the (+) cDNAs from this gel analysis is shown in Table 1, and comparison of these values with those reported by Emtage *et al.* (24) and Sleigh *et al.* (25) suggest that our (+) cDNA species may be complete or almost complete transcripts of all eight separate viral genomic segments.

Cloning of Influenza DNA Segments in *E. coli* Plasmid pBR 322. Duplex DNA molecules of influenza genes were obtained after hybridizing (+) and (−) cDNA preparations and isolating double-stranded molecules on a hydroxyapatite column. The DNA duplexes were further separated according to size by polyacrylamide gel electrophoresis. These DNA seg-

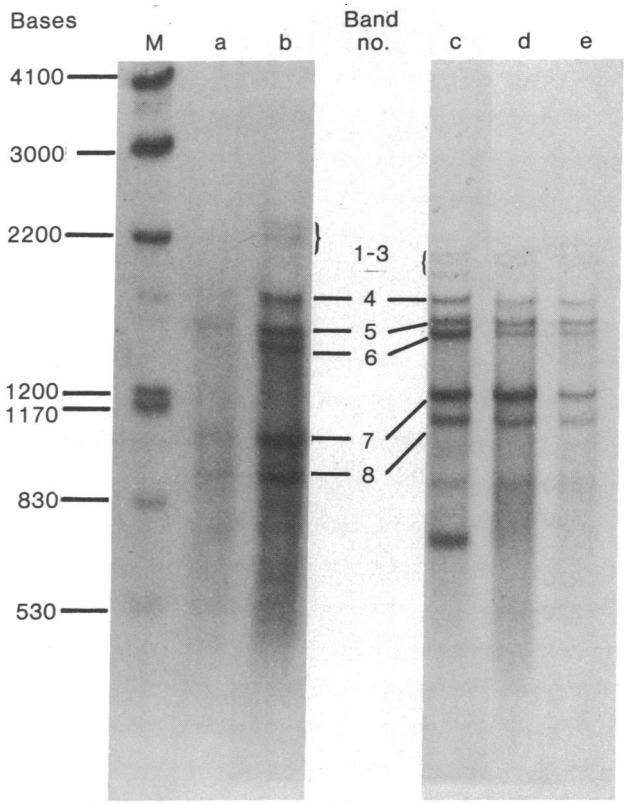

FIG. 2. Synthesis of cDNA strands of influenza A viruses. Influenza virus strain A/Udorn/72 (H3N2) was used to prepare cytoplasmic viral mRNA and virion RNA segments for reverse transcription with oligo(dT) or specific dodecamer DNA primer. The ^{32}P-labeled cDNA products were subjected to electrophoresis on a 1.4% alkaline agarose gel at 3 V/cm for 10–12 hr (23) and autoradiographed. Lanes: a, (−) cDNA from mRNA; b and e, (+) cDNA from vRNA; c, (+) cDNA of A/FM/1/47 (H1N1); d, (+) cDNA of A/Taiwan/62 (H2N2); M, simian virus 40 DNA fragments used as size standards.

Table 1. Estimated molecular size of cDNA segments of influenza A/Udorn/72 virus

Gene	Band no.	cDNA, no. of bases A/Udorn/72	A/Vic/75*	A/NT/68†
P1-P3	1	2300	2500	2390
	2	2200	2400	2390
	3	2100	—	2290
H	4	1850	1920	1760
NP	5	1550	1680	1560
N	6	1450	1580	1480
M	7	1050	1100	1060
NS	8	900	930	890

The size of (+) cDNA species (bands 1 to 8 in Fig. 2) was estimated by gel electrophoresis using restriction enzyme segments of simian virus 40 as standards.
* Data from Emtage *et al.* (24).
† Data from Sleigh *et al.* (25).

ments corresponded in size to the genes of the NS, M, and H proteins. Each of these DNA fragments was inserted into the *Pst* I site of pBR 322 DNA and the recombinant DNA was subsequently used to transform *E. coli* K-12. Plasmid DNA was isolated from the transformants and influenza-specific sequences were identified by hybridization of this DNA to specific vRNA segments (data not shown). In this manner it was shown that recombinant plasmid pFV 26 contained the NS gene, pFV 45 contained the M gene, and pFV 88 contained the H gene. Fig. 3 shows the *Pst* I digests of representative recombinant DNA molecules in which two *Pst* I sites were generated. The NS gene contained approximately 950 base pairs including linker G/C sequences (lane a); the DNA that codes for hemagglutinin (lane c) contained 1950 base pairs including the G/C sequences. The cloned DNA fragment coding for the M protein was cleaved internally by *Pst* I to yield an 830- and a 290-base-pair segment.

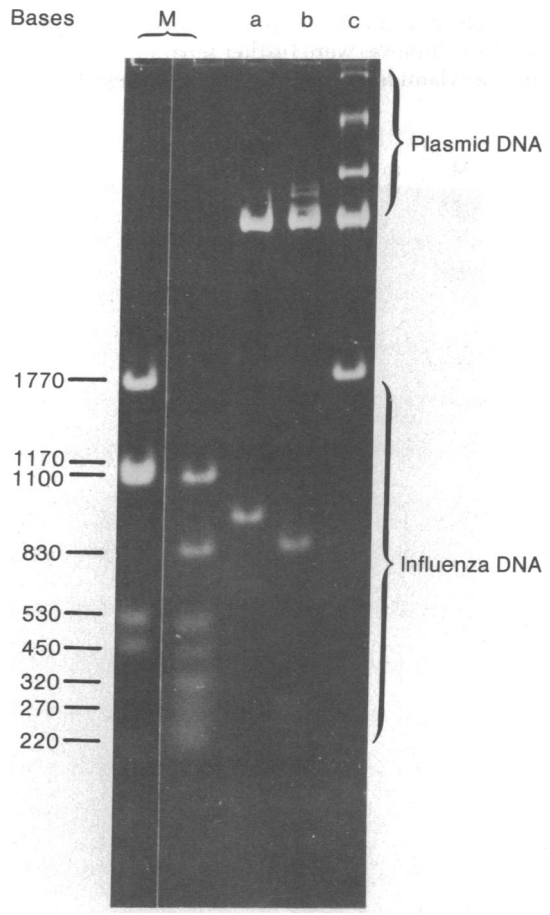

FIG. 3. *Pst* I cleavage of influenza virus recombinant plasmids. Plasmid DNA (form I) for digestion with restriction enzyme *Pst* I was prepared from *E. coli* transformants containing individual influenza virus gene segments (26). The digests were analyzed by electrophoresis on 1.4% agarose gel (27). Three plasmids that contained influenza virus genes are shown: a, pFV 26 which contains NS; b, pFV 45 which contains M; and c, pFV 88 which contains H. The multiple plasmid bands were probably products of incomplete digestion. Lane M contains simian virus 40 DNA fragments as standard markers. Note the cloned influenza M gene DNA was cleaved at one site by *Pst* I whereas the cloned NS and H DNAs were not.

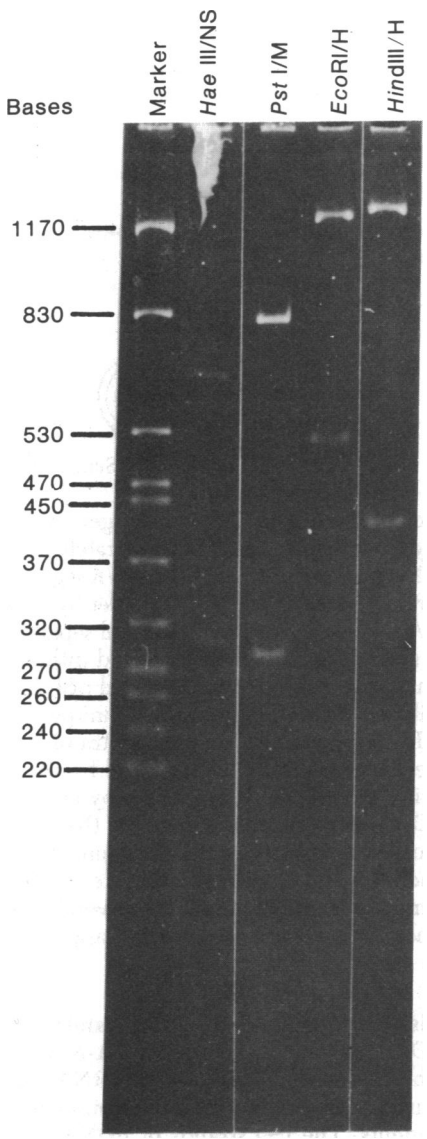

FIG. 4. Restriction enzyme analysis of cloned influenza virus DNA segments. Influenza virus-specific DNA segments were digested with various restriction enzymes and the digests were analyzed by polyacrylamide gel electrophoresis (4%, acrylamide/bisacrylamide = 20:1). Presented are digestions used in part to establish the cleavage maps shown in Fig. 5. Markers are restriction enzyme fragments of simian virus 40 DNA.

Strand Orientation and Restriction Enzyme Maps. The cloned influenza virus DNA sequences isolated after *Pst* I cleavage were characterized further by restriction enzyme

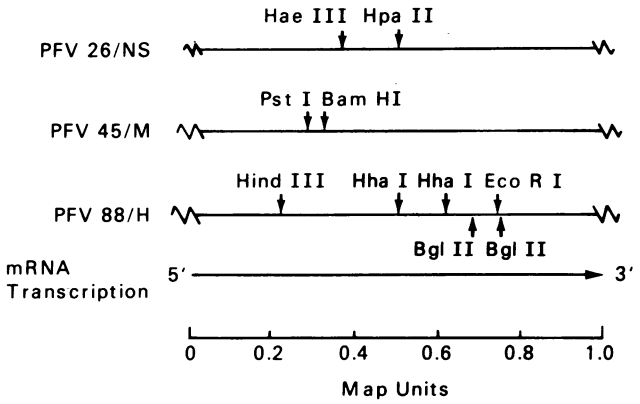

FIG. 5. Cleavage maps and strand orientation of cloned influenza virus DNA segments. The cleavage sites in the influenza virus-specific DNA sequences for restriction enzymes were located by analysis of the DNA fragments. One map unit is the total length of a cloned gene segment.

digestion. The digests of these DNA species resulting from treatment with several enzymes that produced a single cleavage were analyzed by polyacrylamide gel electrophoresis. Representative results are presented in Fig. 4. The NS DNA was cleaved by *Hae* III into two fragments of 640 and 310 nucleotides, and the M gene was cleaved by *Pst* I into fragments of 830 and 290 nucleotides. Similarly, the H gene was also cleaved once by *Eco*RI or *Hin*dIII to yield fragment pairs of 1430 and 520 or 1520 and 430 nucleotides, respectively. It should be noted that these cloned DNA segments contained oligo(G/C) sequences at both termini. NS DNA was cleaved once by *Hpa* II to yield fragments containing 470 and 480 nucleotides (not shown in Fig. 4). Similarly, *Bam*HI cleaved the 830-nucleotide *Pst* I fragment of M DNA to yield approximately 800 and 30 nucleotides. *Bgl* II and *Hha* I each cleaved H DNA at two sites to yield fragments of approximately 1200, 550, and 200 nucleotides and of 870, 850, and 220 nucleotides, respectively. These sites were located by identifying the labeled fragments from a digest of 3'-^{32}P-labeled H DNA duplexes previously cleaved with *Eco*RI.

The orientation of the cloned DNA (corresponding to synthesis of RNA during infection) was established as follows. M DNA duplexes containing a 5'-^{32}P-labeled (+) DNA strand were obtained by mixing the 5'-^{32}P-labeled (+) cDNA with an excess amount of cloned M DNA followed by dissociation and hybridization of strands. The labeled duplexes yielded a labeled fragment of 0.24 map unit (290 nucleotides) after *Pst* I cleavage. In the case of NS and H, hybridization of vRNA occurred with the labeled strand of a 0.67-unit fragment (640 nucleotides) produced by *Hae* III cleavage of 3'-^{32}P-labeled NS DNA and with the labeled strand of a 0.26-unit fragment (520 nucleotides) produced by *Eco*RI cleavage of 3'-^{32}P-labeled H DNA duplexes (data not shown). From these analyses we established the orientation of the cloned influenza virus DNA segments as well as the positions of the restriction enzyme sites with respect to this orientation (Fig. 5).

DISCUSSION

We obtained cloned DNA sequences corresponding to NS, M, and H genomic RNA segments of influenza A/Udorn/72 (H3N2) virus by using DNA duplexes that were formed between cDNAs copied from (−) virion RNA segments and (+) viral mRNA species derived from infected cells. This approach should be generally applicable to other influenza virus genes as well as to other (−) RNA viruses. Reverse transcription of vRNA primed by a 3'-terminus DNA dodecamer yielded cDNA copies that most likely represented all eight full-length genomic sequences. Similar results have been obtained by using vRNA that had been polyadenylylated and primed with oligo(dT) (24, 25). The same dodecamer also primed efficiently the reverse transcription of vRNA from other influenza A subtypes including A/FM-1/47 (H1N1), A/Taiwan/61 (H2N2), A/Ann Arbor/60 (H2N2), and A/WSN/33 (H0N1) (unpublished observations).

The second strand of the DNA duplex used for cloning was derived from reverse transcription of cytoplasmic poly(A)-containing mRNA. It has been shown that the viral mRNA species are 20–30 bases shorter at their 3' ends and therefore a portion of the 5' end of vRNA is not represented (8, 15). On the other hand, the cytoplasmic viral mRNA may contain additional priming sequences at the 5' ends as suggested by the recent findings of Bouloy *et al.* (28). Therefore, small regions of the influenza virus double-stranded DNA structure formed between the two DNA strands may contain mismatched and unpaired nucleotides. After transformation of *E. coli* with such partial DNA duplexes, the cellular DNA repair enzymes presumably can correct the mismatches and complete the unpaired regions. This appears to be the case because the *Pst* I sites were preserved. Thus, it should be possible to obtain viral DNA clones containing full-genomic information if reverse transcription of the vRNA progresses to completion.

Insertion of the influenza DNA duplexes at the *Pst* I site of pBR 322 by using the oligo(dG) and oligo(dC) joining sequences yielded recombinant plasmids from which the virus-specific DNA fragments could be separated after *Pst* I digestion. The viral components were thus easily isolated for further characterization, such as mapping by restriction enzyme cleavage and transcription orientation.

Four to five independent recombinant plasmids were recovered for each gene segment cloned. Most of the corresponding DNA species and their restriction fragments appeared to be identical by gel analysis. Molecular size estimates of the DNA fragments including the G/C linker sequences were slightly greater than those of corresponding (+) cDNA strands which appeared to be complete or almost complete transcripts. This suggests that the cloned DNA segments probably represent full-length molecules. However, it will be necessary to determine nucleotide sequences at both termini in order to identify which, if any, of the cloned DNAs contain complete genomic sequences. Complete influenza viral DNA segments corresponding to each of the eight genes should be useful in elucidating a number of properties of the virus. First, rapid DNA sequence determination methods are available for obtaining the entire nucleotide sequences which should yield information regarding the encoded protein structures. Comparison of sequences unique in each influenza A subtype should provide insight into the origin of antigenic variability which is a prominent characteristic of this virus. Second, the availability of DNA molecules should provide a useful probe for further examination of the control of influenza virus replication at both the transcriptional and translational levels. Third, cloned viral DNA can be used to study translation of individual genes after introduction, in an appropriate vector, into prokaryotic or eukaryotic cells. Fourth, modification of cloned DNA offers one approach to the construction of defined influenza mutants. Success of this approach is of course dependent upon development of a technique for converting cloned DNA back into virion RNA.

We thank Drs. G. Khoury, M. Martin, B. Murphy, and M. Israel for critical reading of this manuscript and helpful discussions.

1. Pons, M. W. & Hirst, G. K. (1968) *Virology* **34,** 385–388.
2. McGeoch, D., Felner, P. & Newton, C. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 3045–3049.
3. Desselberger, U. & Palese, P. (1978) *Virology* **88,** 394–399.
4. Scholtissek, C., Harms, E., Rohde, W., Orlich, M. & Rott, R. (1976) *Virology* **74,** 332–344.
5. Skehel, J. J. & Hay, A. J. (1978) *Nucleic Acids Res.* **5,** 1207–1219.
6. Robertson, J. S. (1979) *Nucleic Acids Res.* **6,** 3745–3757.
7. Etkind, P. R., Buchhagen, D. L., Herz, C., Broni, B. & Krug, R. M. (1977) *J. Virol.* **22,** 346–352.
8. Hay, A. J., Lomniczi, B., Bellamy, A. R. & Skehel, J. J. (1977) *Virology* **83,** 337–355.
9. Lamb, R. A., & Choppin, P. W. (1979) *Proc. Natl. Acad. Sci. USA* **75,** 4908–4912.
10. Lazarowitz, S. G., Compans, R. N. & Choppin, P. W. (1971) *Virology* **46,** 830–843.
11. Inglis, S. C., Barrett, T., Brown, C. M. & Almond, J. W. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 3790–3794.
12. Ritchey, M. B., Palese, P. & Schulman, J. L. (1976) *J. Virol.* **20,** 307–313.
13. Skehel, J. J. (1972) *Virology* **49,** 23–36.
14. Markoff, L. W., Thierry, F., Murphy, B. R. & Chanock, R. M. (1979) *Infect. Immun.* **26,** 280–286.
15. Etkind, P. R. & Krug, R. M. (1975) *J. Virol.* **16,** 1464–1475.
16. Lai, C. J., Dhar, R. & Khoury, G. (1978) *Cell* **14,** 971–982.
17. Myers, J. C., Spiegelman, S. & Kacian, D. L. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 2840–2843.
18. Roychoudhury, R., Jay, E. & Wu, R. (1976) *Nucleic Acids Res.* **3,** 101–116.
19. Enea, V., Voris, G. F. & Zinder, N. D. (1975) *J. Mol. Biol.* **96,** 495–509.
20. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 3961–3965.
21. Taniguchi, T., Palmieri, M. & Weissmann, C. (1978) *Nature (London)* **274,** 223–228.
22. Boyer, H. W., Betlach, M., Boliver, F., Rodriguez, R. L., Heyneker, H. L., Shine, J. & Goodman, H. M. (1977) in *Impact on Science and Society, Proceedings of the 10th Miles International Symposium,* eds. Beers, R. F., Jr. & Bassett, E. G. (Raven, New York), pp. 9–20.
23. McDonnell, M. W., Simon, M. N. & Studier, F. W. (1977) *J. Mol. Biol.* **110,** 119–146.
24. Emtage, J. S., Catlin, G. H. & Carey, N. H. (1979) *Nucleic Acids Res.* **6,** 1221–1239.
25. Sleigh, M. J., Both, G. W. & Brownlee, G. G. (1979) *Nucleic Acids Res.* **6,** 1309–1321.
26. Clewell, D. B. (1972) *J. Bacteriol.* **110,** 667–676.
27. Sharp, P. A., Sugden, B. & Sambrook, J. (1973) *Biochemistry* **12,** 3055–3063.
28. Bouloy, M., Plotch, S. J. & Krug, R. M. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4886–4890.