JOURNAL OF VIROLOGY, Apr. 1982, p. 55–63
0022-538X/82/040055-09$02.00/0

Vol. 42, No. 1

Complete Nucleotide Sequence of the Polymerase 3 Gene of Human Influenza Virus A/WSN/33

JOHN S. KAPTEIN AND DEBI P. NAYAK*

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

Received 19 November 1981/Accepted 14 December 1981

The complete nucleotide sequence of polymerase 3 (P3) gene of a human influenza virus (A/WSN/33) has been determined using cDNA clones except for the last 11 nucleotides which were obtained by direct RNA sequencing. The WSN P3 gene contains 2,341 nucleotides and codes for a protein of 759 amino acids (molecular weight 85,800). The WSN P3 protein, as deduced from the plus-strand DNA sequence, is basic and enriched in positively charged amino acids. In addition, it contains clusters of basic amino acids which may provide sites for the interaction of P3 protein with the capped primer, template, and/or other polymerase proteins during the transcriptive and replicative processes of influenza viral RNA.

Influenza A viruses contain eight RNA segments. After infection these RNA segments are transcribed into mRNA's which are translated into proteins. The proteins encoded by the corresponding individual RNA segments have been identified (11, 22, 29). Seven of these RNAs code for structural proteins which are eventually found in mature virions, and one (RNA8) codes for nonstructural proteins (NS1 and NS2) which are found only in infected cells.

Influenza RNA segments vary considerably in size. The largest ones are over 2,200 nucleotides long and code for the three polymerase (P1, P2, P3) proteins (24). The smaller RNAs range approximately from 1,800 to 900 nucleotides in length and code for hemagglutinin, neuraminidase, nucleoprotein, two or possibly more membrane proteins (17, 23a), and two nonstructural proteins (16). Using recombinant DNA technology and DNA sequencing procedures, the sequences of these five smaller viral RNA segments from one or more influenza A viruses were determined and the primary structures of the corresponding proteins were deduced.

Three polymerase proteins encoded by the three largest RNA segments are required for replication and transcription (15, 29). Both replication and transcription are complex processes which involve the interaction of these polymerase proteins with RNAs, other viral proteins, and host factors (12). Among the three polymerase proteins, P3 has recently been shown to recognize the 5′ Cap 1 structure of cellular mRNA which is used as the primer during viral mRNA synthesis (33a). In addition, defective interfering viral RNA segments that have been examined to date also originate only from polymerase genes (20). However, although the three polymerase genes constitute at least 50% of the total RNA mass of influenza virus, virtually no information is available concerning the nucleotide or amino acid sequences of different polymerase genes or their variation among different subtypes or strains of influenza virus.

In an attempt to provide a basis for understanding the structure and function of polymerase proteins and to elucidate their role in the process of viral replication and transcription, we have undertaken DNA cloning and sequence analysis of polymerase genes. In this report we present the complete sequence of P3 gene of A/WSN/33 virus.

MATERIALS AND METHODS

Virus and cells. Viral RNA used for cloning was obtained from ts52 virus (a group II temperature-sensitive mutant of A/WSN/33) which was grown in MDBK cells at 34°C. Viral RNA was isolated from purified virus preparation and enriched for polymerase genes by fractionating in sucrose velocity gradients (3).

Recombinant DNA cloning and DNA sequencing of P3 gene. The P3 gene of influenza A/WSN/33 ts52 (21) was cloned as a double-stranded DNA copy in the *Pst*I site of pBR322 (3). Accordingly, RNA enriched in polymerase genes was reverse transcribed into a plus-strand DNA copy using the 5′-specific primer 5′ dAGCGAAAGCAGG 3′. Approximately full-length plus-strand DNA copies were isolated on 1.4% alkaline agarose gels, and copied into double-stranded DNA using fold-back of the 3′ end as the self-primer. After S1 nuclease treatment, double-stranded DNA fragments were size fractionated on neutral agarose gels, and approximately 20 dC residues were added to their 3′ ends. Finally the dC-tailed double-stranded

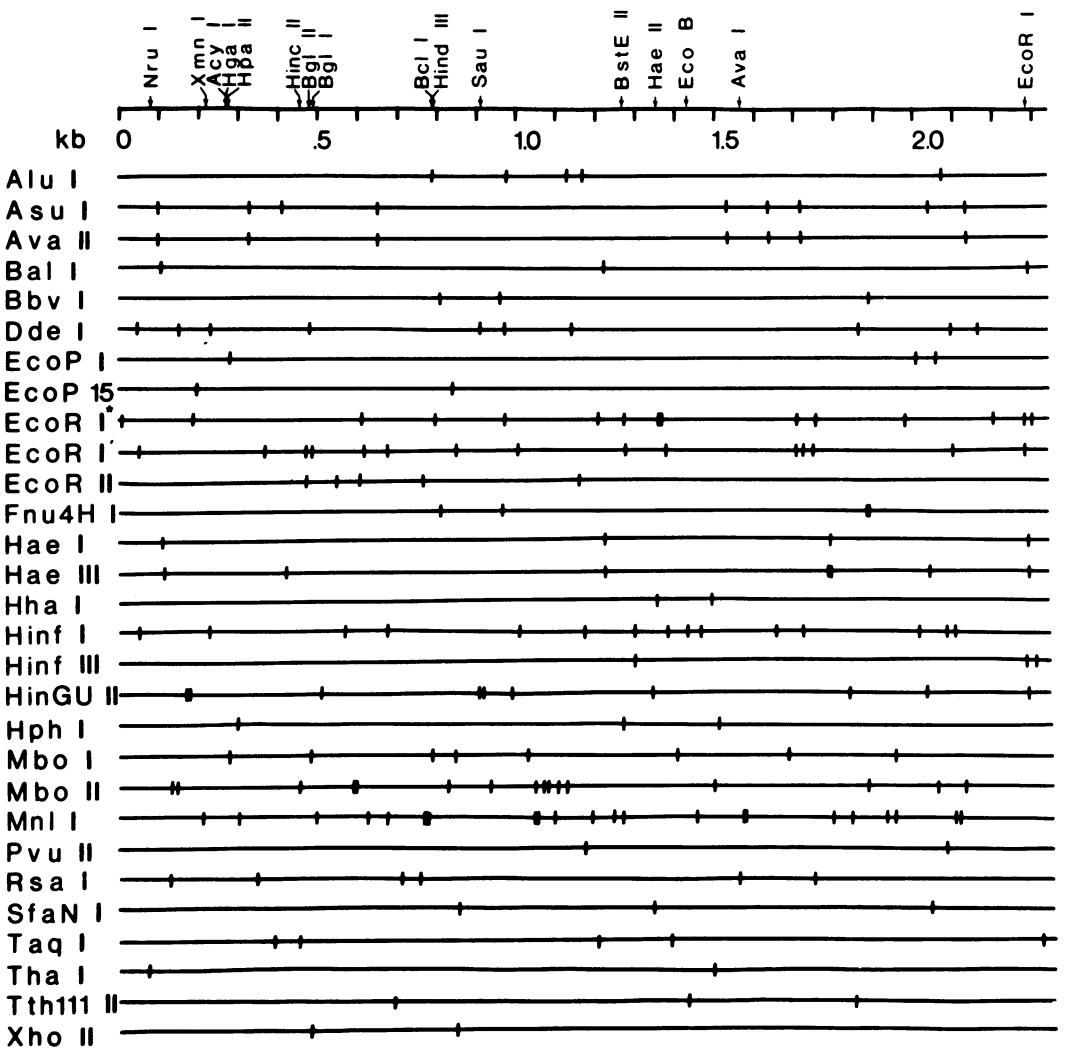


FIG. 1. Restriction map of A/WSN/33 P3 gene. Sites listed across the top of the diagram represent recognition sequences which are found only once in the P3 insert. Those listed along the side represent sequences found more than once. The sites of cleavage are designated by the small vertical bars. The following recognition sequences are not present: *Acc*I, *Asu*II, *Ava*III, *Avr*II, *BamH*I, *Cau*II, *Cla*I, *Eco*K, *Gdi*II, *HgiA*I, *HgiC*I, *HgiE*II, *Hpa*I, *Kpn*I, *Mst*I, *Nae*I, *Nar*I, *Pst*I, *Pvu*I, *Rru*I, *Sal*I, *Sma*I, *Sna*I, *Sph*I, *Sst*I, *Sst*II, *Stu*I, *Tthlll*I, *Xba*I, *Xho*I, *Xma*III.

DNA was inserted into the *Pst*I site of pBR322 to which approximately 20 dG residues had been added. *Escherichia coli* χ1776 cells were transformed, screened for tetracycline resistance, and characterized for insert size. Clones with inserts of approximately 2.2 to 2.4 kilobases (kb) were tentatively designated as clones of polymerase genes and analyzed to identify them as either of P1, P2, or P3 origin.

DNA sequencing of the P3 inserts was carried out by the methods of Maxam and Gilbert (18). In all cases, asymmetric cleavage by a second restriction enzyme was used for isolating DNA fragments uniquely labeled at one 5' end and, thereby, allowing sequence analysis of a fragment from its labeled ends.

RESULTS

Identification of DNA clones of the P3 gene. Before extensive sequence analysis, a number of approaches were used to identify clones containing an insert of P3 origin. (i) As mentioned before, inserts of all clones in this group were approximately 2.2 to 2.4 kb and were thus larger than the expected size of any influenza gene except for the polymerase genes. (ii) All of these clones hybridized to combined polymerase gene RNAs isolated from gels, but not to other viral RNA segments. (iii) Restriction analyses classi-

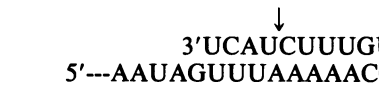
fied these clones into three groups as expected for three polymerase genes. (iv) Clones of specific polymerase genes were identified by hybridization to specific defective interfering RNAs originating from known polymerase genes. For example, defective interfering RNA L1 and L2a of P3 origin hybridized only to the DNA from 1-14b and 1-26b clones. These defective interfering RNAs are easily separable by gel electrophoresis and have been extensively characterized (2). (v) Finally, similarity of the sequences at the 5′ and 3′ ends of the plus strands of these clones to the previously reported end sequences of P3 gene confirmed the identity of these clones (6a, 26).

Nucleotide sequence analysis. A preliminary restriction map was constructed using several of the enzymes which have six-nucleotide recognition sequences. The orientation of the P3 gene with respect to the pBR322 DNA was determined and found to be the same for both clones with the 3′ end of the plus-strand DNA in close proximity to the *Pvu*I site of pBR322. Initially, a cleavage map of the insert DNA was determined with a number of restriction enzymes. Appropriate cleavages were then employed to obtain the fragments used for sequencing. A detailed restriction map obtained from the complete sequence information and also partly confirmed by actual restriction enzyme analyses is shown in Fig. 1.

The series of fragments and restriction sites which were used in sequencing is shown in Fig. 2. All sites used as either the site of labeling or the site of second cleavage were also read through from another site to verify the continuity of overlaps. The sequence through *Eco*RII (*Bst*NI) sites was verified by sequencing through the site from both strands, mapping of *Bst*NI sites, and kinasing and sequencing from *Bst*NI sites. Thus all gaps in the sequencing ladder, due to the presence of methylcytosine (22), were resolved.

The sequence presented was obtained from the clone 1-14b except for the nucleotides from 2322 to 2341 (Fig. 3). At the 5′ end of the plus strand, clone 1-14b contains the complete sequence of the oligonucleotide primer used for priming DNA synthesis on the viral RNA (vRNA) template and is linked to pBR322 through 15 dG residues. At the 3′ end, clone 1-14b ends at nucleotide 2321 followed by a tail of 11 dC residues. Clone 1-26b contains an additional viral sequence of 9 nucleotides, extending to position 2330 followed by a tail of 19 dC residues. However, it is not possible to ascertain whether the two cytosines at position 2329 and 2330 originated from reverse transcription of the viral RNA or from the addition of G:C residues. Finally, the last 11 nucleotides (position 2331-2341) were obtained by direct RNA sequencing of the vRNA (2). Confirmation of the correct overlap was also obtained by comparison of this sequence to the partial sequence of P3 gene of fowl plague virus (26).

The sequence at the 3′ end of full length cDNA,



shows the likely loop structure involved in priming the double-stranded DNA synthesis. Arrows show the position of possible S1 nuclease cleavage sites which could generate the insert of the 1-14b and 1-26b clones. This fortuitous self-complementarity at the 3′ end of the cDNA may have enabled us to obtain nearly full-length clones of this gene. The other genes of influenza virus differ in the nucleotide sequence beyond the last 13 nucleotides and therefore would not be expected to generate clones with the same degree of completeness by this procedure.

Characterization of the A/WSN/33 P3 (WSN P3) gene. The entire nucleotide sequence of the WSN P3 gene and the amino acid sequence of

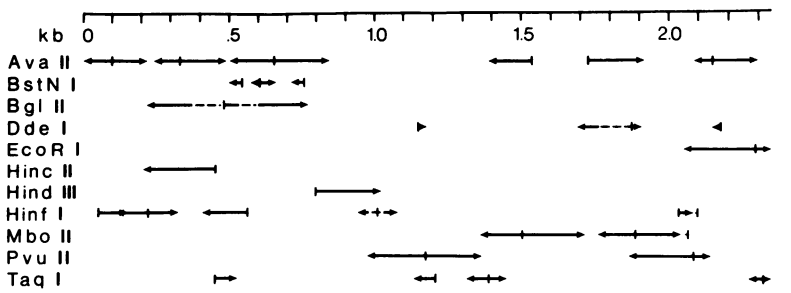


FIG. 2. Sequencing strategy of cloned P3 DNA. Listed along the side are the restriction endonuclease sites which were kinased as shown by vertical bars. The site of cleavage by a second restriction enzyme is not shown. The solid lines represent the sum of the nucleotide sequence as determined from a series of overlapping gels. The dashed lines represent nucleotides which were run off the bottom of the gel and were not determined from those sites.

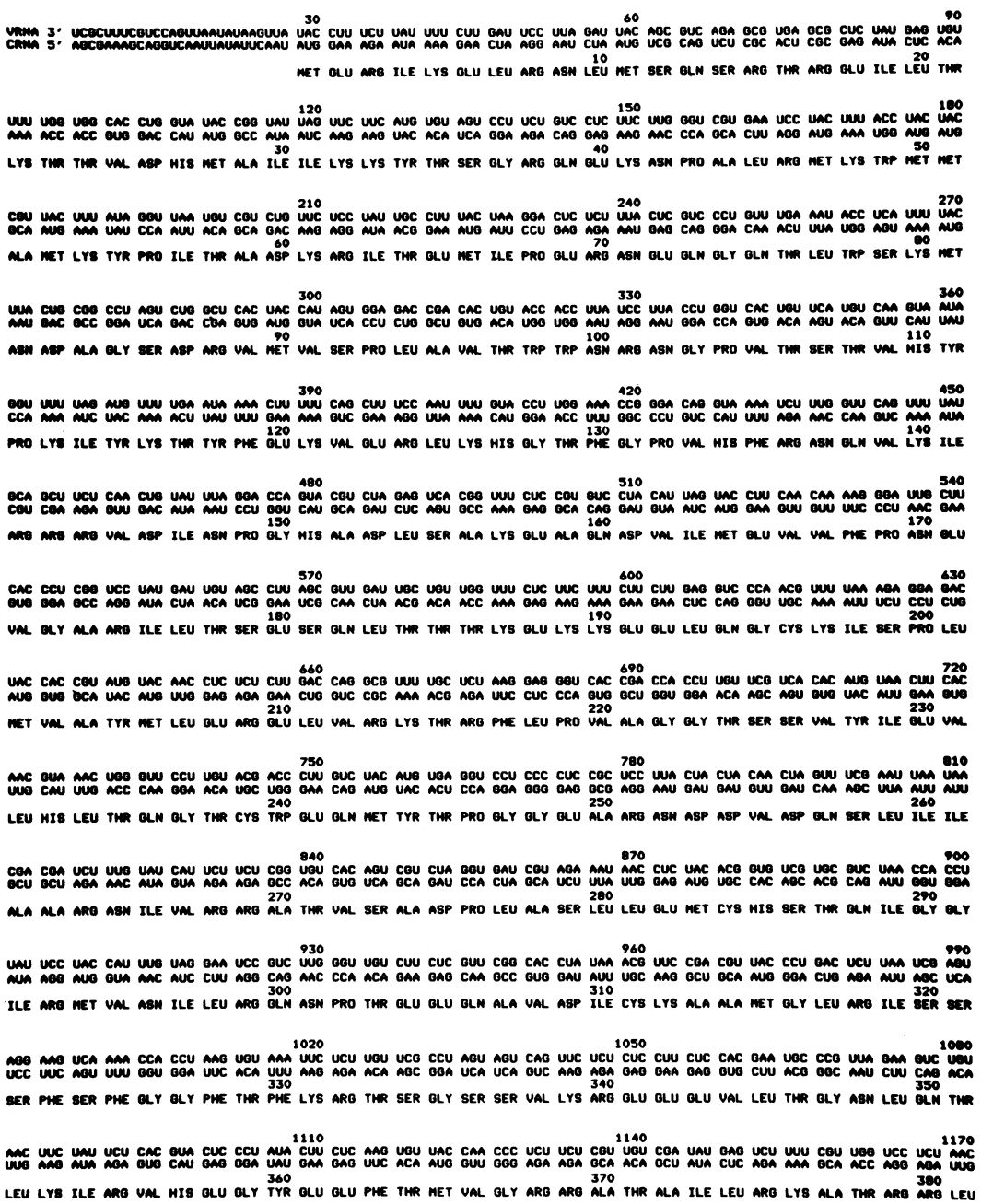


FIG. 3. P3 gene of A/WSN/33. The nucleotide sequence of both the minus (vRNA) strand and the plus (cRNA) strand is shown. Numbering of the nucleotides is from the 5′ end of the plus strand. Also shown is the amino acid sequence of the P3 protein as deduced from translation of the nucleotide sequence starting from the first AUG of the plus strand.

the P3 protein, as deduced from the plus-strand sequence starting from the first AUG (12, 13), are presented in Fig. 3. The WSN P3 gene is 2,341 nucleotides long. It is initiated and terminated by the known 13 nucleotide conserved sequences at the 5′ and 3′ ends.

The plus strand at the 5′ region contains 27 untranslated nucleotides prior to the first AUG. This reading frame is then open for almost the entire gene and ends with a termination codon (UAG) at nucleotide position 2305 followed by a second in-phase termination codon (UAA) at

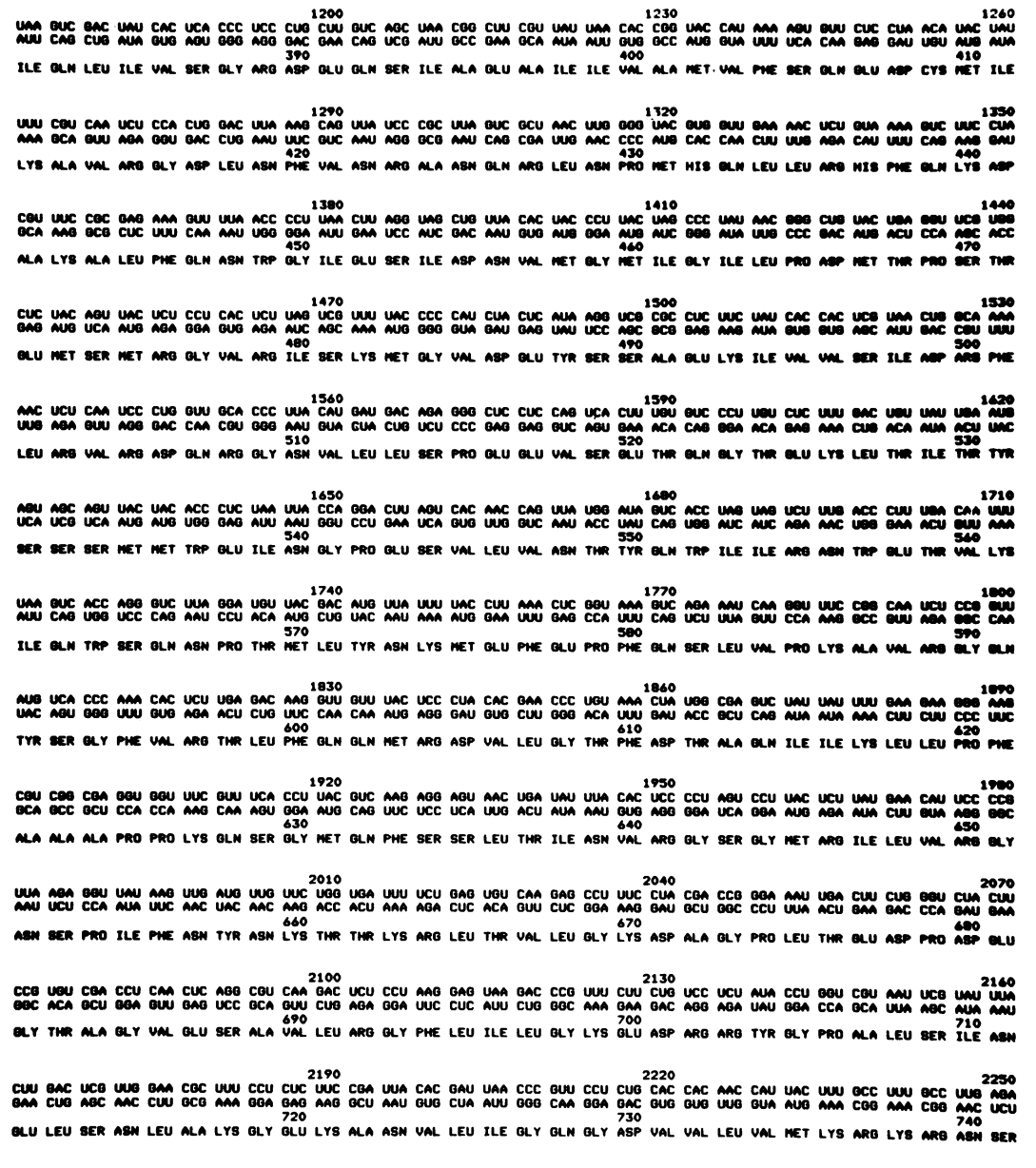


FIG. 3—*Continued*

position 2320 (Fig. 3 and 4). The 37 nucleotides at the 3′ end which are not translated contain the proposed polyadenylation site (positions 2321 to 2325) of the mRNA (27).

The frequency of codon usage is shown in Table 1. As found for other eucaryotic genes, CG-containing codons are relatively few. On a random basis considering the base composition of P3 plus-strand RNA, one would expect 83 CG-containing codons, but only 28 occur in the P3 gene. This bias against CG is particularly evident in the usage of CCG and CGN coding for proline and arginine, respectively. Furthermore, the occurrence of CG dinucleotide irrespective of its presence in codons is also low (2.4% compared to expected 4.7%). A similar deficien-

cy in CG dinucleotide whether present in codons or between adjacent codons has been reported for other genes of influenza virus and for other viruses as well. Thus the bias against CG dinucleotide appears to be operating at the level of both DNA (33) as well as RNA genomes.

The portion of the gene which can be translated extends from the nucleotide position 28 to 2304 and codes for 759 amino acids. The other two reading frames of the plus strand and all three reading frames of the minus strand are blocked repeatedly by termination codons and are unlikely to be used in synthesizing functional proteins (Fig. 4). However, sequences over 300 nucleotides without any termination codons are present in these reading frames. Of interest in this respect is the first AUG of the minus strand (phase 3) at nucleotide position 40, which is present approximately in the same relative position as the first AUG of the plus strand. This AUG is followed by an open reading frame extending for 300 nucleotides. However, the significance of an open reading frame in the minus strand is unknown.

Sites which closely resemble consensus donor or acceptor sites for splicing

↓----------intron---------- ↓
(....AG GUPuAG.....PyNPyPyPyNCAG ;

reference 30) can be found in both the plus and minus strands (in the plus strand, donor sites at positions 12, 399, 1057, 1274, and 1579, and acceptor sites at positions 1344 and 1894; and in the minus strand, donor sites at positions 413, 1278, 1673, 2172, and 2329, and acceptor sites at positions 230 and 965). Since neither altered mRNA's nor altered P3 proteins have been demonstrated, the significance of these potential splicing sites is unknown. Furthermore, splicing sites do not appear to be involved in generating influenza defective interfering RNAs (D. P. Nayak, N. Sivasubramanian, A. R. Davis, R. Cortini, and J. Sung, Proc. Natl. Acad. Sci. U.S.A., in press).

The P3 protein predicted from our sequence data contains 759 amino acids and has a molecular weight of 85,800. This compares favorably with previous estimates of the size of the P3 protein, the smallest of the three polymerase proteins with molecular weights ranging from 80 $\times 10^3$ to 100 $\times 10^3$ (24, 29). Our data therefore suggest that the primary translation product is probably the functional protein. However, in the absence of either the amino- or carboxyterminal amino acid sequences, any minor proteolytic cleavage or additional modification cannot be ruled out.

In polyacrylamide gel electrophoresis, P3 RNA runs as the largest RNA segment whereas P3 protein migrates as the smallest of three polymerase proteins (24). Our sequence data, however, do not support a reduced coding capacity of P3 mRNA or a major cleavage of the primary translation product to generate the P3 protein. It is therefore likely that this represents an anomalous migratory behavior of either polymerase RNAs or polymerase proteins, or both, in polyacrylamide gel electrophoresis or a post-translational modification of P1 and P2 proteins or a combination of these factors.

Table 2 shows the amino acid composition of the predicted P3 protein. Clearly it is a basic protein. Horisberger (10) has also reported that P1 and P3 are basic proteins and that P2 is acidic. In addition we find that the P3 protein contains a large excess of methionine (36 residues) and fewer cysteine residues (5 residues) when compared to the average composition of proteins (5). P3 protein has a very hydrophilic amino end and does not contain any large clusters of hydrophobic or nonpolar amino acids at either amino- or carboxy-terminus. Therefore, this protein is unlikely to be attached to membranes during its biosynthesis, transport, or assembly into virions.

DISCUSSION

The P3 gene reported here contains 2341 nucleotides and is the largest of the influenza viral genes sequenced to date. The influenza WSN P3 gene contains sequences which are similar to the known partial sequences of this gene from other influenza viruses. For example, in the plus strand, a comparison with A/PR/8 P3 gene (6a) shows only two changes in the first 110 nucleotides (position 51, G→A, and position 75, U→C). Similarly, at the 5′ end, the partial sequence of fowl plague virus segment one RNA (26) differs at two positions out of 49 (position 17, U→A, and position 33, A→G), and at the 3′ end two positions out of 63 vary (position 2307, G→A, and position 2316, A→U). These changes occur in either the noncoding region or in the last position of a codon without altering any of the amino acid residues. However, additional sequence studies of other A viruses will be required before assessing the diversity and lineage of the P3 genes among subtype A viruses.

Studies of temperature-sensitive mutants have shown that all three polymerase genes are involved in viral RNA (both plus and minus strand) synthesis (15, 29). Moreover, Krug and his colleagues (33a) have recently shown that the P3 protein recognizes the 5′ terminal Cap I structure of host mRNA which is used as a primer in viral transcription. They have also shown that the P3 protein remains associated

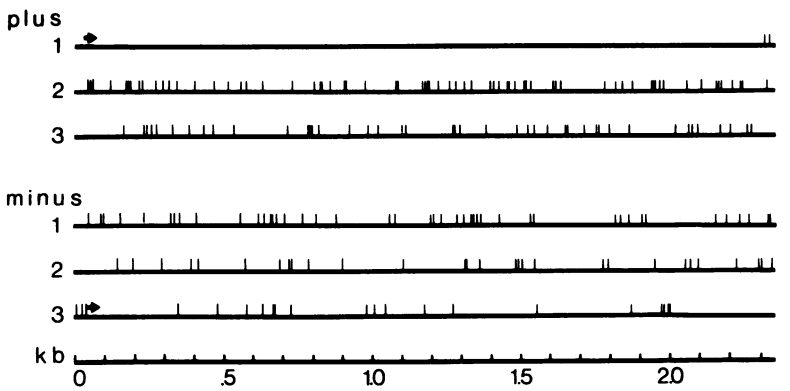


FIG. 4. Termination codons in the sequence of the P3 gene of A/WSN/33. The vertical bars represent termination codons found in the plus strand (top) and minus strand (bottom). 1, 2, and 3 represent the reading frames phased from the first, second, and third nucleotides, respectively. The arrows represent the position of the first AUG in the plus and minus strands.

with this cap structure throughout the transcription process.

As mentioned earlier, P3 is a basic protein and contains 115 basic amino acid residues (60 Arg, 45 Lys, 10 His) compared to 81 acidic amino acid residues (29 Asp, 52 Glu). Charge calculations indicate that P3 is more basic than nucleoprotein and membrane proteins, the other two basic influenza proteins for which the primary structure has been deduced (35, 36a). At pH 6.5, each molecule of WSN P3 protein has a net charge of +29 compared to +14 for PR/8 nucleoprotein (36) and +9.5 for PR/8 membrane proteins (36).

When a secondary structure analysis was performed using Chou-Fasman (1) analysis supplemented by the helix wheel plot (28), P3 protein was found to contain several clusters of basic amino acids in non-α-helical regions (e.g., amino acid residues 140, 213, 375, 736, and 752). These clusters contain three to four arginine and lysine residues in close proximity without being interrupted by acidic residues. These clusters of basic amino acids are much more pronounced than those reported for PR/8 nucleoprotein (36a) and PR/8 membrane (35, 36) proteins. In addition, the P3 protein also contains an α-helical region where basic amino acids are spaced three

TABLE 1. Frequency of codon usage in A/WSN/33 P3 gene

	Frequency						
	U		C		A		G
U	Phe 13	Ser	7	Tyr	7	Cys	1 U
	11		6		9		4 C
	Leu 7		14	Term	0	Term	0 A
	13		5		1	Trp	10 G
C	Leu 11	Pro	9	His	8	Arg	3 U
	9		4		2		3 C
	7		16	Gln	15		3 A
	13		0		21		3 G
A	Ile 19	Thr	12	Asn	23	Ser	9 U
	10		11		12		13 C
	24		25	Lys	28	Arg	31 A
	Met 36		5		17		17 G
G	Val 14	Ala	11	Asp	14	Gly	7 U
	8		11		15		7 C
	9		17	Glu	27		27 A
	26		6		25		9 G

TABLE 2. Amino acid composition of P3 protein (A/WSN/33) as deduced from the nucleic acid sequence

Amino acid	No. of residues
Ala	45
Arg	60
Asn	35
Asp	29
Cys	5
Gln	36
Glu	52
Gly	50
His	10
Ile	53
Leu	60
Lys	45
Met	36
Phe	24
Pro	29
Ser	54
Thr	53
Trp	10
Tyr	16
Val	57

to four amino acids apart (amino acid residues 431 to 448). In this region one side of the α-helix presents a cluster of basic groups. Since the P3 protein interacts with the capped primer RNA (33a), and probably also with the template viral RNA and other polymerase proteins as well (e.g., P2), the clusters of basic amino acid residues may provide areas of interaction. Similar RNA-protein interaction via clusters of basic amino acids has been proposed for influenza nucleoprotein (36a),Semliki forest virus nucleocapsid (7), VP1 of simian virus 40 (34) and polyoma virus (31), and the core antigen of hepatitis B virus (25).

One of the major difficulties in studying the polymerase proteins has been the small amount of P proteins present either in infected cells or in virions (probably one to two molecules per RNA segment). Therefore, although a peptide mapping analysis of fowl plague P3 protein (11) has been performed, neither the amino acid composition nor a direct amino acid sequencing of polymerase proteins of any influenza virus has been possible. However, since cloned influenza genes can now be expressed in eucaryotic (8, 8a, 31) as well as procaryotic (4, 5, 9) systems, it should be possible to express P3 clones and to produce relatively large amounts of functional protein. This would then help in defining the role of the P3 protein in the transcription/replication process of influenza viruses.

ACKNOWLEDGMENTS

We thank Keiichi Itakura for providing us with the synthetic primers and Alan R. Davis for his help in cloning. We also thank David Londo for independently reading the gels.

This work was supported by a National Research Service Award (CA-6091) and by a grant from the National Science Foundation (PCM 7823220) and Public Health Service grants from the National Institute of Allergy and Infectious Diseases, (R01AI12749, R01AI16348).

LITERATURE CITED

1. **Chou, P. Y., and G. D. Fasman.** 1978. Empirical predictions of protein conformation. Annu. Rev. Biochem. **47**:251–276.
2. **Davis, A. R., A. L. Hiti, and D. P. Nayak.** 1980. Influenza defective interfering viral RNA is formed by internal deletion of genomic RNA. Proc. Natl. Acad. Sci. U.S.A. **77**:215–219.
3. **Davis, A. R., A. L. Hiti, and D. P. Nayak.** 1980. Construction and characterization of a bacterial clone containing the hemagglutinin gene of the WSN strain (H0N1) of influenza virus. Gene **10**:205–218.
4. **Davis, A. R., D. P. Nayak, M. Ueda, A. L. Hiti, D. Dowbenko, and D. G. Kleid.** 1981. Expression of antigenic determinants of the hemagglutinin gene of a human influenza virus in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **78**:5376–5380.
5. **Dayhoff, M. O., L. T. Hunt, and S. Hurst-Calderone.** 1978. Composition of proteins, p. 363. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5, supplement 3, chapter 25. National Biomedical Research Foundation, Washington, D.C.
6. **Emtage, J. S., W. C. A. Tacon, G. H. Catlin, B. Jenkins, A. G. Porter, and N. H. Carey.** 1980. Influenza antigenic determinants are expressed from hemagglutinin gene 3 cloned in *Escherichia coli*. Nature (London) **283**:171–174.

6a. **Fields, S., and G. Winter.** 1981. Influenza virus A/PR/8/34 genes: sequencing by a shotgun approach, p. 55–63. *In* D. P. Nayak (ed.), Genetic variation among influenza viruses. Academic Press, Inc., New York.

7. **Garoff, H., A. M. Frischauf, K. Simons, H. Lehrach, and H. Delius.** 1980. The capsid protein of Semliki Forest virus has clusters of basic amino acids and prolines in its amino terminal region. Proc. Natl. Acad. Sci. U.S.A. **77**:6376–6380.
8. **Gething, M-J., and J. Sambrook.** 1981. Cell-surface expression of influenza hemagglutinin from a cloned DNA copy of the RNA gene. Nature (London) **293**:620–625.

8a. **Hartman, J. R., D. P. Nayak, and G. C. Fareed.** 1981. Human influenza virus hemagglutin is expressed in monkey cells using simian virus 40 vectors. Proc. Natl. Acad. Sci. U.S.A. **79**:233–237.

9. **Heiland, I., and M-J. Gething.** 1981. Cloned copy of the hemagglutinin gene codes for human influenza antigenic determinants in *E. coli*. Nature (London) **292**:851–852.
10. **Horisberger, M. A.** 1980. The large P proteins of influenza viruses are composed of one acidic and two basic polypeptides. Virology **107**:302–305.
11. **Inglis, S. C., A. R. Carroll, R. A. Lamb, and B. W. J. Mahy.** 1976. Polypeptides specified by the influenza virus genome. 1. Evidence for eight distinct gene products specified by fowl plague virus. Virology **74**:489–503.
12. **Kozak, M.** 1980. Binding of wheat germ ribosomes to bisulfite-modified reovirus messenger RNA: evidence for a scanning mechanism. J. Mol. Biol. **144**:291–304.
13. **Kozak, M., and A. J. Shatkin.** 1978. Migration of 40 S ribosomal subunits on messenger RNA in the presence of edeine. J. Biol. Chem. **253**:6568–6577.
14. **Krug, R. M., S. J. Plotch, I. Ulmanen, C. Herz, and M. Bouloy.** 1981. The mechanism of initiation of viral RNA transcription by capped RNA primers, p. 291–302. *In* D. H. L. Bishop and R. A. Compans (ed.), The replication of negative strand viruses. Elsevier/North Holland Publishing Co., Amsterdam.
15. **Krug, R. M., M. Ueda, and P. Palese.** 1975. Temperature-sensitive mutants of influenza WSN virus defective in virus-specific RNA synthesis. J. Virol. **16**:790–796.
16. **Lamb, R. A., P. W. Choppin, R. M. Chanock, and C-J. Lai.** 1980. Mapping of the two overlapping genes for polypeptides NS1 and NS2 on RNA segment 8 of influenza virus genome. Proc. Natl. Acad. Sci. U.S.A. **77**:1857–1861.
17. **Lamb, R. A., C-J. Lai, and P. W. Choppin.** 1981. Sequences of mRNAs derived from genome RNA segment 7 of influenza virus: colinear and interrupted mRNAs code for overlapping proteins. Proc. Natl. Acad. Sci. U.S.A. **78**:4170–4174.
18. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
19. **McGeoch, D., P. Fellner, and C. Newton.** 1976. Influenza virus genome consists of eight distinct RNA species. Proc. Natl. Acad. Sci. U.S.A. **73**:3045–3049.
20. **Nayak, D. P.** 1980. Defective interfering influenza viruses. Annu. Rev. Microbiol. **34**:619–644.
21. **Nayak, D. P., K. Tobita, J. M. Janda, A. R. Davis, and B. De.** 1978. Homologous interference mediated by defective interfering influenza virus derived from a temperature-sensitive mutant of influenza virus. J. Virol. **73**:375–386.
22. **Ohmori, H., J. I. Tomizawa, and A. M. Maxam.** 1978. Detection of 5-methylcytosine in DNA sequences. Nucleic Acids Res. **5**:1479–1485.
23. **Palese, P.** 1977. The genes of influenza virus. Cell **10**:1–10.

23a. **Palese, P., R. M. Elliott, M. Baez, J. J. Zasra, and J. F. Young.** 1981. Genome diversity among influenza A, B, and C viruses and genetic structure of RNA 7 and RNA 8 of influenza A viruses, p. 127–140. *In* D. P. Nayak (ed.), Genetic variation among influenza viruses. Academic Press, Inc., New York.

24. **Palese, P., M. B. Ritchey, and J. L. Schulman.** 1977. Mapping of the influenza virus genome. II. Identification of the P1, P2 and P3 genes. Virology **76**:114–121.
25. **Pasek, M., T. Goto, W. Gilbert, B. Zink, H. Schaller, P. Mackay, G. Leadbetter, and K. Murray.** 1979. Hepatitis B virus genes and their expression in *E. coli*. Nature (London) **282**:575–579.
26. **Robertson, J.** 1979. 5′ and 3′ terminal nucleotide sequences of the RNA genome segments of influenza virus. Nucleic Acids Res. **6**:3745–3757.
27. **Robertson, J. S., M. Schubert, and R. A. Lazzarini.** 1981. Polyadenylation sites for influenza virus mRNA. J. Virol. **38**:157–163.
28. **Schiffer, M., and A. B. Edmundson.** 1967. Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. Biophys. J. **7**:121–135.
29. **Scholtissek, C.** 1978. The genome of the influenza virus. Curr. Top. Microbiol. Immunol. **80**:139–169.
30. **Sharp, P. A.** 1981. Speculations on RNA splicing. Cell **23**:643–646.
31. **Soeda, E., J. R. Arrand, and B. E. Griffin.** 1980. Polyoma virus DNA—complete nucleotide sequence of the gene which codes for polyoma virus capsid protein VP1 and overlaps the VP2-VP3 genes. J. Virol. **33**:619–630.
32. **Sveda, M. M., and C-J. Lai.** 1981. Functional expression in primate cells of cloned DNA coding for the hemagglutinin surface glycoprotein of influenza virus. Proc. Natl. Acad. Sci. U.S.A. **78**:5488–5492.
33. **Swartz, M. N., T. A. Trautner, and A. Kornberg.** 1962. Enzymatic synthesis of DNA: further studies on nearest neighbor base sequences in deoxyribonucleic acids. J. Biol. Chem. **237**:1961–1967.
33a. **Ulmanen, I., B. A. Broni, and R. M. Krug.** 1981. A role of two of the influenza virus core P proteins in recognizing cap 1 structures ($m^7GpppNm$) on RNAs and in initiating viral RNA transcription. Proc. Natl. Acad. Sci. U.S.A. **78**:7355–7359.
34. **von Heuverswyn, H., A. van de Voorde, and W. Fiers.** 1978. Nucleotide sequence of the simian virus 40 Hind II + III restriction fragment J and the total amino acid sequence of the major structural protein VP1. Eur. J. Biochem. **91**:415–430.
35. **van Rompuy, L., W. Min Jou, D. Huylebroeck, R. Devos, and W. Fiers.** 1981. Complete nucleotide sequence of the nucleoprotein gene from the human influenza strain A/PR/8/34 (H0N1). Eur. J. Biochem. **116**:347–353.
36. **Winter, G., and S. Fields.** 1980. Cloning of influenza cDNA into M13: the sequence of the RNA segment encoding the A/PR/8/34 matrix protein. Nucleic Acids Res. **8**:1965–1974.
36a. **Winter, G., and S. Fields.** 1981. The structure of the gene encoding the nucleoprotein of human influenza virus A/PR/8/34. Virology **114**:423–428.