JOURNAL OF VIROLOGY, Oct. 1982, p. 321-329
0022-538X/82/100321-09$02.00/0
Copyright © 1982, American Society for Microbiology

Vol. 44, No. 1

Sequence Analysis of the Polymerase 1 Gene and the Secondary Structure Prediction of Polymerase 1 Protein of Human Influenza Virus A/WSN/33

N. SIVASUBRAMANIAN AND DEBI P. NAYAK*

Jonsson Comprehensive Cancer Center and Department of Microbiology and Immunology, School of Medicine, University of California, Los Angeles, California 90024

Received 29 March 1982/Accepted 18 June 1982

The nucleotide sequence of polymerase 1 (P1) gene of a human influenza virus (A/WSN/33) has been determined by using cDNA clones, except for the last 83 nucleotides, which were obtained by primer extension. The WSN P1 gene contains 2,341 nucleotides and codes for a protein of 757 amino acids (M_r = 86,500). P1 gene possesses a striking tandem repeat of 12 nucleotides (nucleotide position 2,188 to 2,199, 2,200 to 2,211) and a corresponding tandem repeat of tetrapeptide in the P1 protein. The deduced sequence of P1 protein is enriched in basic amino acids, particularly arginine. In addition, it also contains clusters of basic amino acids which may provide sites for the interaction with the template virion RNA capped primer as well as with other proteins involved in viral replication and transcription. A secondary structure prediction, using Chou and Fasman analyses (Annu. Rev. Biochem. 47:251-276, 1978), shows that the P1 protein possesses some unique features, viz., one "four-helical supersecondary structure" and four "polypeptide double helices" (antiparallel β-pleated sheets) which are considered important in RNA binding.

It is now well known that the segmented genome of influenza virus is transcribed and replicated by using the gene products of three polymerase (P1, P2, and P3) genes and also, possibly, of the nucleoprotein (NP) gene (41, 42). These events, especially the primary transcription process, have been reported to occur in the nucleus of the host cell just after infection (22, 34) and do not require either the host or viral protein synthesis (20, 46). Also, it has been reported that since influenza transcripts use the 5′ end of the capped host RNAs as primers (6, 14, 30, 45), the virus-specific transcription process requires a continuous function of the host RNA polymerase II (22). Furthermore, the involvement of splicing enzyme in the processing and maturation of some viral messengers has also been reported (32, 33). Additionally, polymerase genes have been found to play another important role in the biology of influenza virus, namely, all defective interfering (DI) influenza viral RNAs studied to date appear to originate from the polymerase genes (11, 12, 38).

Clearly, an understanding of the structure and function of polymerase proteins will be required to elucidate their role in the processes of viral transcription and replication and in the formation of DI RNAs. As a first step towards this objective, we have already determined the primary sequences of the P3 gene (28) of WSN virus. In this report, we present the complete sequence of the P1 gene as well as the predicted primary and secondary structures of P1 protein of A/WSN/33 virus.

MATERIALS AND METHODS

Virus and cells. The procedures for growing WSN virus by using MDBK cells, for purifying the virus by using sucrose velocity gradients, and for isolating the viral RNA used for cloning have been described previously (10). *ts*52 virus (a group II temperature-sensitive mutant of A/WSN/33 virus) grown in MDBK cells at 34°C was used in these studies.

Recombinant DNA cloning and DNA sequencing of P1 gene. The procedures for DNA cloning and for identifying P1 clones have been reported (10, 28). Briefly, virion RNA enriched in polymerase genes was reverse transcribed with the avian myeloblastosis virus reverse transcriptase into cDNA (plus strand). cDNAs of full length were isolated on 1.4% alkaline agarose gels and used for the synthesis of double-stranded DNA, using the foldback loop at the 3′ end as the self-primer. Subsequently, double-stranded DNA fragments were treated with S1 nuclease and fractionated on neutral agarose gels to determine their size. Finally, approximately 20 deoxycytidine residues were added to their 3′ ends. These double-stranded DNAs were then inserted into the *Pst*I site of pBR322 DNA to which approximately 20 deoxyguanidine residues had been added. *Escherichia coli* χ1774 cells were transformed. Clones which were resistant to tetracycline but sensitive to ampicillin were analyzed for insert size. Clones containing inserts of approximately 2.2 to 2.4 kilobases were tentatively designated as clones of polymerase genes, and analyzed for identification as being of P1, P2, or P3 origin.

The nucleotide sequence of the insert DNA was carried out by the methods of Maxam and Gilbert (36, 37), employing asymmetric cleavage by a second restriction enzyme to obtain DNA fragments uniquely labeled at one 5′ end. Some doubly labeled fragments were strand separated according to Maxam and Gilbert (37) and then sequenced. The sequence at the 3′ end of cRNA (plus strand) was completed by using a primer extension procedure (23).

Computer analysis of the sequence and secondary structure prediction. Computer analysis of the nucleotide and amino acid sequence was performed by using the program of Queen and Korn (48). Secondary structure prediction of the P1 protein from the amino acid sequence was done according to Chou and Fasman (9), utilizing the computer programs provided by Nancy Woods (University of California, Los Angeles [UCLA]). The helical hydrophobic moment (⟨μH⟩) and the mean hydrophobicity (⟨H⟩) were determined by using the hydrophobic values of amino acids (27) according to Eisenberg, Weiss, and Terwilliger of UCLA (personal communication), utilizing the computer program provided by Robert M. Weiss (UCLA).

RESULTS

Identification of DNA clones of the P1 gene. Several selection criteria were employed to identify clones containing an insert of P1 origin. (i) All clones belonging to this group contained inserts of approximately 2.2 to 2.4 kilobases, which is larger than the expected size of any influenza gene except the polymerase genes. (ii) Only the combined polymerase gene RNAs isolated from gels—and no other viral RNA segments—hybridized to these clones, demonstrating that these clones were of polymerase gene origin. (iii) Furthermore, these clones were classified into three groups by restriction analyses, as expected for three polymerase genes. (iv) Hybridization to specific DI RNAs originating from known polymerase genes was used to identify clones of specific polymerase genes. For example, DI RNAs L3 and L2b of P1 origin hybridized only to the DNA from 1-39b and 1-72b clones. These DI RNAs are easily separable by gel electrophoresis and have been extensively characterized (11, 12, 39). (v) Finally, the sequences at the 5′ and 3′ ends of the plus strands of these clones were compared with the previously reported end sequence of P1 gene to confirm clones of P1 origin (49). Thus, 1-39b and 1-72b clones were identified as clones of P1 origin and used for detailed sequence analyses.

Sequencing strategy. The sequencing strategy and restriction sites which were used in sequencing are shown in Fig. 1. All of these sites that were 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. Additionally, all *Eco*RII (*Bst*NI) sites were verified by sequencing through these sites on both strands as well as by mapping the *Bst*NI sites.

The nucleotide sequence was first obtained from two P1 gene clones, viz., 1-39b and 1-72b, and completed by primer extension (23). The entire sequence of 1-39b insert was first determined. It has the entire 5′ end of the complementary DNA, including the dodecadeoxynucleotide primer used for reverse transcription, and ends at position 2,103 at the 3′ end. Hence, this clone is incomplete and is missing 238 nucleotides. The clone 1-72b has the entire sequence of 1-39b and 157 additional nucleotides at the 3′ end. Finally, the sequence of the P1 gene was completed by isolating a primer frag-

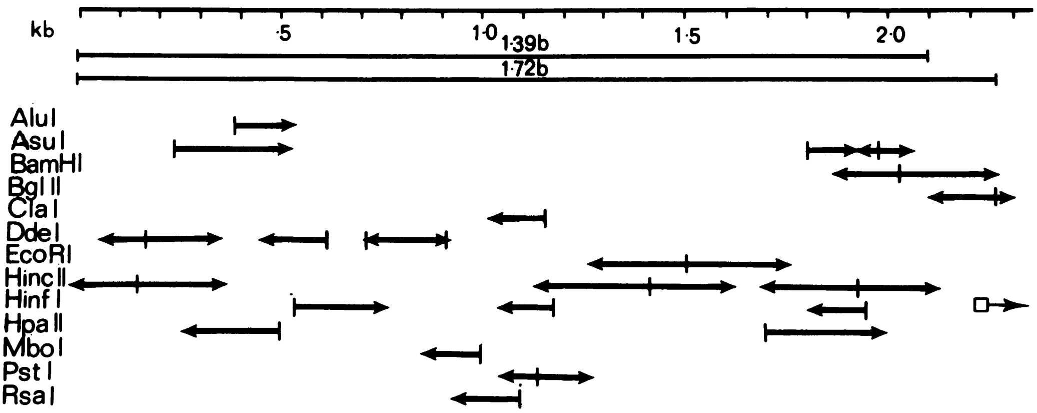

FIG. 1. Sequencing strategy of cloned P1 DNA. Vertical bars represent the restriction sites which were used for end labeling. Solid line arrows represent the length of sequences obtained from the corresponding restriction sites through overlapping gels.

ment, *Hin*fI to *Bgl*II (nucleotide position 2,215 to 2,258), from the 1-72b insert uniquely labeled at the *Hin*fI site and extended with avian myeloblastosis virus reverse transcriptase, using the total virion RNA as the template (23). The sequence obtained by the primer extension was 81 nucleotides. Later, these sequences were confirmed with the direct P1 gene virion RNA end sequences and also with the DI RNA (L2b and L3) end sequences which we obtained independently from different DNA clones (39).

Nucleotide sequence of A/WSN/33 P1 gene. The complete nucleotide sequence of the viral RNA and the complementary RNA of the WSN P1 gene are shown in Fig. 2. It contains 2,341 nucleotides, including conserved sequences of 13 nucleotides at the 5′ and 3′ ends. The plus strand at the 5′ region contains 24 untranslated nucleotides before the first AUG. From the nucleotide position 25 to 2,295, there is an open reading frame of 2.271 kilobases with a coding capacity of 757 amino acids ending with two consecutive in-phase termination codons (UAG, UGA). The other two reading frames contain numerous termination codons which are rather evenly scattered throughout the entire sequence. The 46 nucleotides at the 3′ end are not translated and contain the proposed polyadenation site (2,321 to 2,325) of the mRNA (50).

An analysis of the frequency of codon usage in the P1 mRNA showed that 60 codons, with the exception of CGC, are used to translate the P1 protein. Although 36 of 61 codons are used more than 10 times, the frequency of CG-containing codons and of CG dinucleotide even outside the codon, as reported for other eucaryotic genes, is low (28).

Amino acid sequence of P1 protein. The P1 protein, as predicted from our sequence data, is probably the largest protein of influenza virus. Although it contains 757 amino acids and is 2 amino acids shorter than the predicted P3 protein, the P1 protein has a slightly larger molecular weight (86,500) than the WSN P3 protein (85,800). The size of the P1 protein predicted from our sequence is somewhat smaller than the estimated molecular weight (M_r = 96,000) of the P1 protein by polyacrylamide gel electrophoresis analysis (43, 53). An analysis of the predicted amino acid composition (Table 1) indicates that it is a basic protein, as has been previously reported (25). Among the basic amino acids, the arginine content is high. Among the hydrophobic amino acids, the content of the alanine and valine is low, whereas the content of isoleucine and methionine is high when compared with the average composition of proteins (13). It is also low in cysteine.

Of the 757 amino acids of the P1 protein, 113 amino acid residues are basic (53 Arg, 48 Lys, 12 His) and 80 are acidic (32 Asp, 48 Glu). Charge calculations indicate that the P1 protein is more basic than nucleoprotein and matrix proteins but slightly less basic than the P3 protein. At pH 6.5, the WSN P1 protein has a net charge of +27, compared with +29 for A/WSN/33 P3 (28), +14 for PR/8 NP (60), and +9.5 for PR/8 M proteins (59).

A striking feature of the P1 amino acid sequence is an iterative tetrapeptide beginning at the amino acid residue 722 (Ala-Arg-Ile-Asp-Ala-Arg-Ile-Asp). The RNA which codes this region is equally iterative, with only a single nucleotide mismatch, and is suggestive of a duplication event in the history of the P1 gene. This octapeptide is predicted to form an α-helix. Iterative tetrapeptide has also been found in the amino acid sequence of *Eco*RI endonuclease (18, 40).

Secondary structure of the P1 protein. The secondary and supersecondary structure of P1 protein was determined according to the analyses of Chou and Fasman (9). Although the accuracy of these analyses is about 80%, the procedure has been used to predict the secondary structure of a number of proteins (3, 8, 18, 40) and is a first step towards understanding the structure-function relationship of a protein. Figure 3 shows the predicted secondary structure of P1 protein. It consists of 33% predicted α-helices, 26% β-pleated sheets, 23% β-reverse turn, and 18% undefined structure. Furthermore, it contains four antiparallel α-helices, known as "four-helical supersecondary structure" (1) between the amino acid residues 341 and 415, and four pairs of antiparallel β-pleated sheets (otherwise known as "polypeptide double helix" [5]) at the amino acid residues 17 to 49, 434 to 451, 447 to 477, and 555 to 572 (Fig. 3). Figure 4 shows the mean plot of helical hydrophobic moments of predicted α-helices against their average hydrophobicities. The mean helical hydrophobic moments are defined as the mean vector sum of the side chains of a helix, and the values were estimated by the method of Eisenberg, Weiss, and Terwilliger (personal communication). None of the α-helices of the P1 protein are of transmembrane type (the transmembrane helices of hemagglutinin are plotted for comparison). The majority of the α-helices of P1 protein have medium hydrophobic moments and low mean hydrophobicities, which are characteristic of soluble globular proteins, except for one α-helix (amino acid residue 695 to 700) which possesses a larger hydrophobic moment than that of typical globular protein α-helices and has the characteristics of amphiphilic or surface-seeking helices, i.e., one face is moderately hydrophilic but the other is moderately hydrophobic.

```
A/WSN/33  P1 POLYMERASE GENE
                                      30                                      60                                      90
VRNA 3' UCGCUUUCGUCCGUUUGGUAAACU UAC CUA CAG UUA GGC UGA AAU GAA AAG AAU UUU CAC GGU CGU GUU UUA CGA UAU UCG UGU UGA AAG
CRNA 5' AGCGAAAGCAGGCAAACCAUUUGA AUG GAU GUC AAU CCG ACU UUA CUU UUC UUA AAA GUG CCA GCA CAA AAU GCU AUA AGC ACA ACU UUC
                                                                      10                                      20
                                 MET ASP VAL ASN PRO THR LEU LEU PHE LEU LYS VAL PRO ALA GLN ASN ALA ILE SER THR THR PHE

                                    120                                     150                                     180
GGA AUA UGA CCU CUG GGA GGA AUG UCG GUA CCC UGU CCU UGU CCU AUG UGG UAC CUA UGA CAG UUG UCC UGU GUA GUC AUG AGU CUU UCC
CCU UAU ACU GGA GAC CCU CCU UAC AGC CAU GGG ACA GGA ACA GGA UAC ACC AUG GAU ACU GUC AAC AGG ACA CAU CAG UAC UCA GAA AGG
                             30                                      40                                      50
PRO TYR THR GLY ASP PRO PRO TYR SER HIS GLY THR GLY THR GLY TYR THR MET ASP THR VAL ASN ARG THR HIS GLN TYR SER GLU ARG

                                    210                                     240                                     270
CCU UCU ACC UGU UGU UUG UGG CUU UGA CCU CGU GGC GUU GAG UUG GGC UAA CUA CCC GGU GAC GGU CUU CUG UUA CUU GGU UCA CCA AUA
GGA AGA UGG ACA ACA AAC ACC GAA ACU GGA GCA CCG CAA CUC AAC CCG AUU GAU GGG CCA CUG CCA GAA GAC AAU GAA CCA AGU GGU UAU
                             60                                      70                                      80
GLY ARG TRP THR THR ASN THR GLU THR GLY ALA PRO GLN LEU ASN PRO ILE ASP GLY PRO LEU PRO GLU ASP ASN GLU PRO SER GLY TYR

                                    300                                     330                                     360
CGG GUU UGU CUA ACA CAU AAC CUU CGU UAC CGG AAG GAA CUC CUU AGG GUA GGA CCA UAG AAA CUC UGG AGC ACA GAA CUU UGC UAC CUC
GCC CAA ACA GAU UGU GUA UUG GAA GCA AUG GCC UUC CUU GAG GAA UCC CAU CCU GGU AUC UUU GAG ACC UCG UGU CUU GAA ACG AUG GAG
                             90                                     100                                     110
ALA GLN THR ASP CYS VAL LEU GLU ALA MET ALA PHE LEU GLU GLU SER HIS PRO GLY ILE PHE GLU THR SER CYS LEU GLU THR MET GLU

                                    390                                     420                                     450
CAA CAA GUC GUU UGU GCU CAC CUG UUC GAC UGU GUU CCG GCU GUC UGG AUA CUG ACC UGA GAU UUA UCC UUG GUC GGA CGA CGU UGU CGU
GUU GUU CAG CAA ACA CGA GUG GAC AAG CUG ACA CAA GGC CGA CAG ACC UAU GAC UGG ACU CUA AAU AGG AAC CAG CCU GCU GCA ACA GCA
                            120                                     130                                     140
VAL VAL GLN GLN THR ARG VAL ASP LYS LEU THR GLN GLY ARG GLN THR TYR ASP TRP THR LEU ASN ARG ASN GLN PRO ALA ALA THR ALA

                                    480                                     510                                     540
AAC CGG UUG UGU UAU CUU CAC AAG UCU AGU UUA CCG GAG UGC CGG UUA CUU AGG CCU UCC GAG UAU CUG AAG GAA UUC CUA CAU UAC CUC
UUG GCC AAC ACA AUA GAA GUG UUC AGA UCA AAU GGC CUC ACG GCC AAU GAA UCC GGA AGG CUC AUA GAC UUC CUU AAG GAU GUA AUG GAG
                            150                                     160                                     170
LEU ALA ASN THR ILE GLU VAL PHE ARG SER ASN GLY LEU THR ALA ASN GLU SER GLY ARG LEU ILE ASP PHE LEU LYS ASP VAL MET GLU

                                    570                                     600                                     630
AGU UAC UUG UUU CUU CUU UAC CUC UAG UGU UGA GUA AAA GUC UCU UUC UCU GCU CAC UCU CUG UUA UAC UGA UUC UUU UAC CAC UGU GUC
UCA AUG AAC AAA GAA GAA AUG GAG AUC ACA ACU CAU UUU CAG AGA AAG AGA CGA GUG AGA GAC AAU AUG ACU AAG AAA AUG GUG ACA CAG
                            180                                     190                                     200
SER MET ASN LYS GLU GLU MET GLU ILE THR THR HIS PHE GLN ARG LYS ARG ARG VAL ARG ASP ASN MET THR LYS LYS MET VAL THR GLN

                                    660                                     690                                     720
UCU UGU UAU CCA UUU UCC UUC GUC UCU AAC UUG UUU UCC UCA AUA GAU UAA UCC CGU AAC UGG GAC UUG UGU UAC UGG UUU CUA CGA CUC
AGA ACA AUA GGU AAA AGG AAG CAG AGA UUG AAC AAA AGG AGU UAU CUA AUU AGG GCA UUG ACC CUG AAC ACA AUG ACC AAA GAU GCU GAG
                            210                                     220                                     230
ARG THR ILE GLY LYS ARG LYS GLN ARG LEU ASN LYS ARG SER TYR LEU ILE ARG ALA LEU THR LEU ASN THR MET THR LYS ASP ALA GLU

                                    750                                     780                                     810
UCU CCC UUC GAU UUU GCC UCU CGU UAA CGU UGG GGU CCC UAC GUU UAU UCC CCC AAA CAU AUG AAA CAA CUC UGU GAU CGU UCC UCA UAU
AGA GGG AAG CUA AAA CGG AGA GCA AUU GCA ACC CCA GGG AUG CAA AUA AGG GGG UUU GUA UAC UUU GUU GAG ACA CUA GCA AGG AGU AUA
                            240                                     250                                     260
ARG GLY LYS LEU LYS ARG ARG ALA ILE ALA THR PRO GLY MET GLN ILE ARG GLY PHE VAL TYR PHE VAL GLU THR LEU ALA ARG SER ILE

                                    840                                     870                                     900
ACA CUC UUU GAA CUU GUU AGU CCU AAC GGU CAA CCU CCG UUA CUC UUC UUU CGU UUC AAC CGU UUA CAA CAU UCC UUC UAC UAC UGG UUA
UGU GAG AAA CUU GAA CAA UCA GGA UUG CCA GUU GGA GGC AAU GAG AAG AAA GCA AAG UUG GCA AAU GUU GUA AGG AAG AUG AUG ACC AAU
                            270                                     280                                     290
CYS GLU LYS LEU GLU GLN SER GLY LEU PRO VAL GLY GLY ASN GLU LYS LYS ALA LYS LEU ALA ASN VAL VAL ARG LYS MET MET THR ASN

                                    930                                     960                                     990
AGA GUC CUG UGA CUU UAA AGA AAG UGG UAG UGA CCU CUA UUG UGG UUU ACC UUG CUU UUA GUC UUG GGA GCC UAC AAA AAC CGG UAC UAG
UCU CAG GAC ACU GAA AUU UCU UUC ACC AUC ACU GGA GAU AAC ACC AAA UGG AAC GAA AAU CAG AAC CCU CGG AUG UUU UUG GCC AUG AUC
                            300                                     310                                     320
SER GLN ASP THR GLU ILE SER PHE THR ILE THR GLY ASP ASN THR LYS TRP ASN GLU ASN GLN ASN PRO ARG MET PHE LEU ALA MET ILE

                                   1020                                    1050                                    1080
UGU AUA UAU UGG UCU UUA GUC GGG CUU ACC AAG UCU UUA CAA GAU UCA UAA CGA GGU UAU UAC AAG AGU UUG UUU UAC CGC UCU GAC CCU
ACA UAU AUA ACC AGA AAU CAG CCC GAA UGG UUC AGA AAU GUU CUA AGU AUU GCU CCA AUA AUG UUC UCA AAC AAA AUG GCG AGA CUG GGA
                            330                                     340                                     350
THR TYR ILE THR ARG ASN GLN PRO GLU TRP PHE ARG ASN VAL LEU SER ILE ALA PRO ILE MET PHE SER ASN LYS MET ALA ARG LEU GLY

                                   1110                                    1140                                    1170
UUC CCC AUG UAC AAA CUC UCG UUC UCA UAC UUU UAA UCU UGA GUU UAU GGA CGU CUU UAC GAU CGU UCG UAG CUA AAC UUU AUG AAG UUA
AAG GGG UAC AUG UUU GAG AGC AAG AGU AUG AAA AUU AGA ACU CAA AUA CCU GCA GAA AUG CUA GCA AGC AUC GAU UUG AAA UAC UUC AAU
                            360                                     370                                     380
LYS GLY TYR MET PHE GLU SER LYS SER MET LYS ILE ARG THR GLN ILE PRO ALA GLU MET LEU ALA SER ILE ASP LEU LYS TYR PHE ASN
```

FIG. 2. P1 gene of A/WSN/33. The nucleotide sequences of both the minus (vRNA) and the plus (cRNA) strands are shown. Numbering of the nucleotides is from the 5′ end of the plus strand. Also shown is the amino acid sequence of the P1 protein as deduced from the nucleotide sequence, starting from the first AUG of the plus strand.

Comparison with other influenza A P1 genes and proteins. A comparison of WSN P1 sequence with those of A/PR/8/34 and A/NT/60/68, which have been recently reported (2, 61), shows a remarkable conservation of the structure of P1 gene and P1 protein. The P1 gene of all three viruses contains 2,341 nucleotides and codes for 757 amino acids. Also, amino acid

```
                                   1200                                    1230                                    1260
CUA AGU UGA UCU UUC UUC UAA CUU UUU UAG GCC GGC GAG AAU UAU CUA CCC UGA CGU AGU AAC UCG GGA CCU UAC UAC UAC CCG UAC AAG
GAU UCA ACU AGA AAG AAG AUU GAA AAA AUC CGG CCG CUC UUA AUA GAU GGG ACU GCA UCA UUG AGC CCU GGA AUG AUG AUG GGC AUG UUC
                            390                                     400                                     410
ASP SER THR ARG LYS LYS ILE GLU LYS ILE ARG PRO LEU LEU ILE ASP GLY THR ALA SER LEU SER PRO GLY MET MET MET GLY MET PHE

                                   1290                                    1320                                    1350
UUA UAC AAU UCA UGA CAU AAU CCG CAG AGG UAG GAC UUA GAA CCU GUU UUC UCU GUG UGG UUC UGA UGA AUG ACC ACC CUA CCA GAA GUU
AAU AUG UUA AGU ACU GUA UUA GGC GUC UCC AUC CUG AAU CUU GGA CAA AAG AGA CAC ACC AAG ACU ACU UAC UGG UGG GAU GGU CUU CAA
                            420                                     430                                     440
ASN MET LEU SER THR VAL LEU GLY VAL SER ILE LEU ASN LEU GLY GLN LYS ARG HIS THR LYS THR THR TYR TRP TRP ASP GLY LEU GLN

                                   1380                                    1410                                    1440
AGA AGA CUA CUA AAA CGA GAC UAA CAC UUA CGU GGG UUA GUA CUU CCC UAA GUU CGG CCU CAG UUG UCC AAA AUA GCU UGG ACA UUC GAU
UCU UCU GAU GAU UUU GCU CUG AUU GUG AAU GCA CCC AAU CAU GAA GGG AUU CAA GCC GGA GUC AAC AGG UUU UAU CGA ACC UGU AAG CUA
                            450                                     460                                     470
SER SER ASP ASP PHE ALA LEU ILE VAL ASN ALA PRO ASN HIS GLU GLY ILE GLN ALA GLY VAL ASN ARG PHE TYR ARG THR CYS LYS LEU

                                   1470                                    1500                                    1530
GAA CCU UAA UUA UAC UCG UUC UUU UUC AGA AUG UAU UUG UCU UGU CCA UGU AAA CUU AAG UGU UCA AAA AAG AUA GCA AUA CCC AAA CAA
CUU GGA AUU AAU AUG AGC AAG AAA AAG UCU UAC AUA AAC AGA ACA GGU ACA UUU GAA UUC ACA AGU UUU UUC UAU CGU UAU GGG UUU GUU
                            480                                     490                                     500
LEU GLY ILE ASN MET SER LYS LYS LYS SER TYR ILE ASN ARG THR GLY THR PHE GLU PHE THR SER PHE PHE TYR ARG TYR GLY PHE VAL

                                   1560                                    1590                                    1620
CGG UUA AAG UCG UAC CUC GAA GGG UCG AAA CCC CAC AGA CCC UAG UUG CUC AGA CGC CUG UAC UCA UAA CCU CAA UGA CAG UAG UUU UUG
GCC AAU UUC AGC AUG GAG CUU CCC AGC UUU GGG GUG UCU GGG AUC AAC GAG UCU GCG GAC AUG AGU AUU GGA GUU ACU GUC AUC AAA AAC
                            510                                     520                                     530
ALA ASN PHE SER MET GLU LEU PRO SER PHE GLY VAL SER GLY ILE ASN GLU SER ALA ASP MET SER ILE GLY VAL THR VAL ILE LYS ASN

                                   1650                                    1680                                    1710
UUA UAC UAU UUG UUA CUA GAA CCA GGU CGU UGG CGA GUU UAC CGG GAA GUC GAC AAG UAG UUU CUA AUG UCC AUG UGC AUG GCC ACG GUA
AAU AUG AUA AAC AAU GAU CUU GGU CCA GCA ACC GCU CAA AUG GCC CUU CAG CUG UUC AUC AAA GAU UAC AGG UAC ACG UAC CGG UGC CAU
                            540                                     550                                     560
ASN MET ILE ASN ASN ASP LEU GLY PRO ALA THR ALA GLN MET ALA LEU GLN LEU PHE ILE LYS ASP TYR ARG TYR THR TYR ARG CYS HIS

                                   1740                                    1770                                    1800
UCU CCA CUG UGU GUU UAU GUU UGG GCU UCU AGU AAA CUU UAU UUC UUU GAC ACC CUC GUU UGG GUA AGG UUU CGA CCU GAC GAC CAG AGG
AGA GGU GAC ACA CAA AUA CAA ACC CGA AGA UCA UUU GAA AUA AAG AAA CUG UGG GAG CAA ACC CAU UCC AAA GCU GGA CUG CUG GUC UCC
                            570                                     580                                     590
ARG GLY ASP THR GLN ILE GLN THR ARG ARG SER PHE GLU ILE LYS LYS LEU TRP GLU GLN THR HIS SER LYS ALA GLY LEU LEU VAL SER

                                   1830                                    1860                                    1890
CUG CCU CCG GGU UUA AAU AUG UUG UAA UCU UUA GAG GUG UAA GGA CUU CAG ACG AAC UUU ACC CUU AAU UAC CUA CUC CUA AUG GUC CCC
GAC GGA GGC CCA AAU UUA UAC AAC AUU AGA AAU CUC CAC AUU CCU GAA GUC UGC UUG AAA UGG GAA UUA AUG GAU GAG GAU UAC CAG GGG
                            600                                     610                                     620
ASP GLY GLY PRO ASN LEU TYR ASN ILE ARG ASN LEU HIS ILE PRO GLU VAL CYS LEU LYS TRP GLU LEU MET ASP GLU ASP TYR GLN GLY

                                   1920                                    1950                                    1980
GCA AAU ACG UUG GGU GAC UUG GGU AAA CAG UUG GUA UUU CUG UAA CUU AGU CAC UUG UUA CGU CAC UAU UAC GGU CGU GUA CCA GGU CGG
CGU UUA UGC AAC CCA CUG AAC CCA UUU GUC AAC CAU AAA GAC AUU GAA UCA GUG AAC AAU GCA GUG AUA AUG CCA GCA CAU GGU CCA GCC
                            630                                     640                                     650
ARG LEU CYS ASN PRO LEU ASN PRO PHE VAL ASN HIS LYS ASP ILE GLU SER VAL ASN ASN ALA VAL ILE MET PRO ALA HIS GLY PRO ALA

                                   2010                                    2040                                    2070
UUU UUG UAC CUC AUA CUA CGA CAA CGU UGU UGU GUG AGG ACC UAG GGG UUU UCU UUA GCU AGG UAG AAC UUA UGU UCG GUU UCU CCU UAU
AAA AAC AUG GAG UAU GAU GCU GUU GCA ACA ACA CAC UCC UGG AUC CCC AAA AGA AAU CGA UCC AUC UUG AAU ACA AGC CAA AGA GGA AUA
                            660                                     670                                     680
LYS ASN MET GLU TYR ASP ALA VAL ALA THR THR HIS SER TRP ILE PRO LYS ARG ASN ARG SER ILE LEU ASN THR SER GLN ARG GLY ILE

                                   2100                                    2130                                    2160
GAA CUU CUA CUU GUU UAC AUG GUU UUC ACG ACG UUG AAU AAA CUU UUU AAG AAG GGG UCG UCA AGU AUG UCU UCU GGU CAG CCC UAU AGG
CUU GAA GAU GAA CAA AUG UAC CAA AAG UGC UGC AAC UUA UUU GAA AAA UUC UUC CCC AGC AGU UCA UAC AGA AGA CCA GUC GGG AUA UCC
                            690                                     700                                     710
LEU GLU ASP GLU GLN MET TYR GLN LYS CYS CYS ASN LEU PHE GLU LYS PHE PHE PRO SER SER SER TYR ARG ARG PRO VAL GLY ILE SER

                                   2190                                    2220                                    2250
UCA UAC CAC CUC CGA UAC CAA AGG UCU CGG GCU UAA CUA CGU GCU UAA CUA AAG CUU AGA CCU UCC UAU UUC UUU CUC CUC AAG UGA CUC
AGU AUG GUG GAG GCU AUG GUU UCC AGA GCC CGA AUU GAU GCA CGA AUU GAU UUC GAA UCU GGA AGG AUA AAG AAA GAG GAG UUC ACU GAG
                            720                                     730                                     740
SER MET VAL GLU ALA MET VAL SER ARG ALA ARG ILE ASP ALA ARG ILE ASP PHE GLU SER GLY ARG ILE LYS LYS GLU GLU PHE THR GLU

                                   2280                                    2310                           2341
UAG UAC UUC UAG ACA AGG UGG UAA CUU CUC GAG UCU GCC GUU UUU AUC ACU UAA AUC GAACAGGAAGUACUUUUUUACGGAACAAAGAUGA 5
AUC AUG AAG AUC UGU UCC ACC AUU GAA GAG CUC AGA CGG CAA AAA UAG UGA AUU UAG CUUGUCCUUCAUGAAAAAAUGCCUUGUUUCUACU 3
                            750                         757
ILE MET LYS ILE CYS SER THR ILE GLU GLU LEU ARG ARG GLN LYS
```

FIG. 2. (Continued.)

changes were relatively few: 15 between A/WSN/33 and A/PR/8/34, 20 between A/PR/8/34 and A/NT/60/68, and 18 between A/WSN/33 and A/NT/60/68. The variation observed between the WSN and the PR/8 sequences probably does not reflect the variation in the original isolates but may be attributed to the varying growth and selection procedures that these two viruses have undergone over the last 50 years in laboratories. Similar changes have been observed in the sequences of P3, hemagglutinin, and neuraminidase of these two viruses (15, 23, 24, 28, 62).

The secondary structures of the P1 proteins of all three viruses are essentially the same. All of the cysteine and proline residues, as well as the basic amino acid clusters, are in identical position. Finally, the supersecondary features such as the four antiparallel α-helices and the ββ antiparallel pleated sheets implicated in RNA binding also remain unaltered.

TABLE 1. Amino acid composition (frequency and moles percent) of P1 protein (A/WSN/33) as deduced from the nucleic acid sequence

Amino acid	Frequency	Mol%	Avg protein[a]
Alanine	41	5.4	8.6
Arginine	53	7.0	4.9
Asparagine	51	6.7	4.3
Aspartic acid	32	4.2	5.5
Cysteine	10	1.3	2.9
Glutamine	31	6.3	6.0
Glutamic acid	48	4.1	3.9
Glycine	46	6.1	8.4
Histidine	12	1.6	2.0
Isoleucine	49	6.5	4.5
Leucine	56	7.4	7.4
Lysine	48	6.3	6.6
Methionine	37	4.9	1.7
Phenylalanine	33	4.3	3.6
Proline	32	4.2	5.2
Serine	49	6.5	7.0
Threonine	62	8.2	6.1
Tryptophan	9	1.2	1.3
Tyrosine	24	3.2	3.4
Valine	34	4.5	6.6

[a] The average amino acid composition of proteins (13) is included for comparison.

DISCUSSION

Sequence analysis shows that the WSN P1 gene contains 2,341 nucleotides and is one of the two largest polymerase genes of influenza virus (2, 15, 28, 61). Both P1 and P3 genes contain an identical number of nucleotides and code for basic proteins of essentially similar length (2, 15, 28, 61). However, a comparison of P1 and P3 at the level of nucleotide or amino acid sequences shows no significant homology. This suggests against a possible convergent evolutionary process in the origin of multiple polymerase genes of influenza viruses.

Genetic studies involving temperature-sensitive mutants have shown that P1 and P3 proteins are involved in the complementary RNA synthesis and that the P2 and nucleoproteins are most probably involved in the synthesis of virion RNA (31, 52, 53). Ulmanen and his colleagues (55) have recently divided the viral transcription process into different steps. Firstly, the P3 protein recognizes the *Cap*I structures of host mRNAs, and a viral endonuclease complex (possibly P3 and P1 proteins) cleaves RNA containing the *Cap*I structures at some selective sites to generate primers for the viral transcription process. Secondly, the initiation of transcripts via the addition of a G residue to the primer is possibly catalyzed by the P1 protein. Thus, the P1 protein may be involved in both the cleavage of the primers from the host and also the initiation of viral transcription.

A comparison of the amino acid groups among the basic polymerase proteins of influenza virus, MS2 replicase (16), and poliovirus P3-1b (29) proteins shows that they possess a similar pattern, including short stretches (six residues or less) of amino acid homologies (data not shown). However, although all of these proteins are involved in nucleic acid binding and synthesis and are basic proteins, the content of arginine

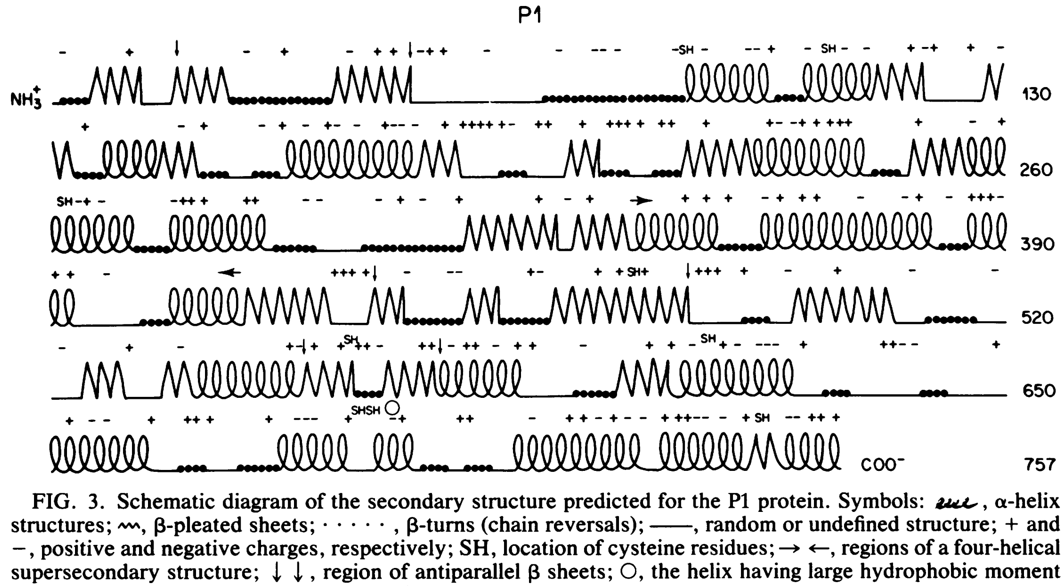

FIG. 3. Schematic diagram of the secondary structure predicted for the P1 protein. Symbols: ℓℓℓ, α-helix structures; ʌʌʌ, β-pleated sheets; · · · · · , β-turns (chain reversals); ——, random or undefined structure; + and −, positive and negative charges, respectively; SH, location of cysteine residues; → ←, regions of a four-helical supersecondary structure; ↓ ↓, region of antiparallel β sheets; ○, the helix having large hydrophobic moment (⟨μH⟩) with moderate hydrophobicity (⟨H⟩) (see Fig. 4).

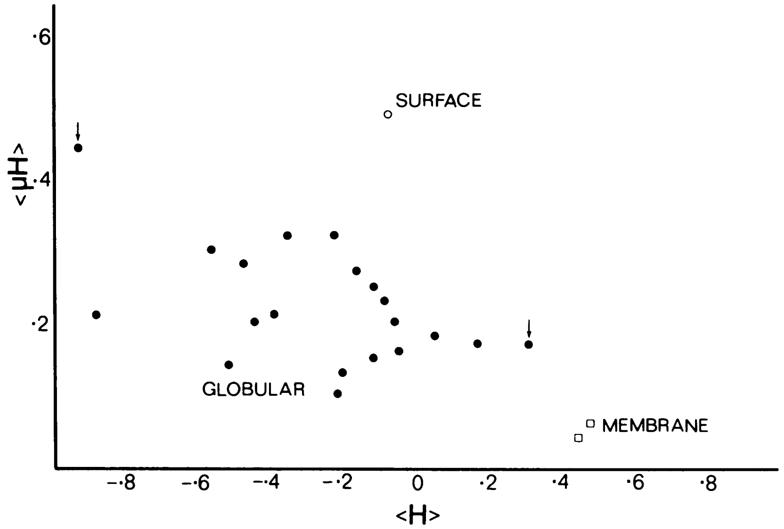

FIG. 4. A hydrophobic moment plot for the predicted α-helical regions of P1 protein of influenza virus. The abscissa gives the mean hydrophobicity (⟨H⟩) of each α-helix, and the ordinate gives the corresponding value of helical hydrophobic moment (⟨μH⟩) as defined in the text. The circles represent the α-helices of the P1 protein; the open circle represents the α-helix with large hydrophobic moment and moderate hydrophobicity. The arrows indicate two of the helices of four-helical supersecondary structure: the one on the right (amino acid position 407 to 415) is the most hydrophobic, and the other, on the far left (amino acid position 386 to 393), is the most hydrophilic. The open squares represent the membrane-penetrating α-helices (185 to 208, 527 to 550) of influenza hemagglutinin (47, 57). "GLOBULAR," "SURFACE," and "MEMBRANE" indicate the regions of the graph where α-helices with corresponding functions plot (Eisenberg, Weiss, and Terwilliger, personal communication).

(the most preferred basic amino acid) is high in the basic proteins of influenza virus. Arginine residues, as in most conserved arginine-rich histones (H3, H4) of eucaryotic cells (4), might play an important role in organizing the nucleoprotein complex in virions as well as in the intracellular replication and the transcription complexes. The possible sites of RNA-protein interaction were further revealed from the secondary structure prediction by using Chou-Fasman analyses (9). (i) The P1 protein showed many clusters of basic amino acids in regions predicted to be devoid of secondary structures (e.g., amino acids in regions starting from 187, 207, 429, and 479) as well as in the α-helical regions. 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 similar to those present in the P3 protein (28) but are much more pronounced than those reported for the PR/8 NP (56, 60) and M (59) proteins and may provide sites for interaction with the template viral RNA during the initiation of transcription. Similar RNA-protein interaction via clusters of basic amino acids has been proposed for influenza P3 (28), influenza NP (60), Semliki Forest virus nucleocapsid (17), VP1 of simian virus 40 (58) and of polyoma virus (54), and the core antigen of hepatitis virus (44). (ii) Additionally, the four-helical supersecondary structure which occurs once in the P1 protein and not in the P3 protein may be involved in RNA protein binding. Similar supersecondary structures have been shown to be present in other proteins involved in either RNA or DNA interaction, e.g., tobacco mosaic virus protein (7, 21), tyrosyl-tRNA synthetase (26), and *E. coli* DNA polymerase I (3). These structures also contain many positive charges, supporting their possible involvement in RNA binding. Furthermore, the most hydrophilic (charged) α-helix (amino acid residues 386 to 393) and the most hydrophobic α-helix (amino acid residues 407 to 415), as determined by "helix wheel" plot (51) and helical hydrophobic moment plot analyses (see Fig. 4) of P1 protein, constitute two of the α-helices of this supersecondary structure. (iii) P1 protein also contains four polypeptide double helices (antiparallel ββ dimer) which are also proposed to be involved in the interaction with the minor groove of RNA helix (5, 19) and are found in DNA polymerase I (3) and Lac repressor (9). Recently, intrasegmental complementation among the temperature-sensitive mutants of P1 gene in A/Udorn/72 (H3N2) virus has been demonstrated (35). Localization of the defect in these mutants may identify the functional domains in the secondary structure of the P1 protein.

ACKNOWLEDGMENTS

We thank K. Itakura for providing us with the synthetic primers and A. R. Davis for his help in cloning. We also thank J. S. Kaptein and Nancy Woods for their help in secondary structure analysis and R. M. Weiss for his help in hydrophobic moment plot analysis.

This work was supported by grants from the National Science Foundation (PCM 7823220) and by Public Health Service grants from the National Institute of Allergy and Infectious Diseases (RO1 AI 12749, RO1 AI 16348).

LITERATURE CITED

1. **Argos, P., M. G. Rossman, and J. E. Johnson.** 1977. A four helical supersecondary structure. Biochem. Biophys. Res. Commun. **75:**83–86.
2. **Bishop, D. H. L., J. A. Huddleston, and G. G. Brownlee.** 1982. The complete sequence of RNA segments of influenza A/NT/60/68 and its encoded P1 protein. Nucleic Acids Res. **10:**1335–1343.
3. **Brown, W. E., K. H. Stump, and W. S. Kelley.** 1982. Escherichia coli DNA polymerase 1 sequence characterization and secondary structure prediction. J. Biol. Chem. **257:**1965–1972.
4. **Camerini-Otero, R. D., B. Sollner-Webb, and G. Felsenfeld.** 1977. The structure of the nucleosome: evidence for an arginine rich histone kernel. *In* H. J. Vogel (ed.), Nucleic acid-protein recognition. Academic Press, Inc., New York.
5. **Carter, C. W., and J. Kraut.** 1974. A proposed model for interaction of polypeptides with RNA. Proc. Natl. Acad. Sci. U.S.A. **71:**283–287.
6. **Caton, A. J., and J. S. Robertson.** 1980. Structure of the host-derived sequences present at the 5′ ends of influenza virus mRNA. Nucleic Acids Res. **8:**2591–2603.
7. **Champness, J. N., A. C. Bloomer, G. Bricogne, P. J. G. Butler, and A. Klug.** 1976. The structure of the protein disk of tobacco mosaic virus to 5 Å resolution. Nature (London) **259:**20–24.
8. **Chou, P. Y., A. J. Adler, and G. D. Fasman.** 1975. Conformational prediction and circular dichroism studies on the *lac* repressor. J. Mol. Biol. **96:**29–45.
9. **Chou, P. Y., and G. D. Fasman.** 1978. Empirical predictions of protein conformation. Annu. Rev. Biochem. **47:**251–276.
10. **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 (Amst.) **10:**205–218.
11. **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.
12. **Davis, A. R., and D. P. Nayak.** 1979. Sequence relationships among defective interfering influenza virus RNAs. Proc. Natl. Acad. Sci. U.S.A. **76:**3092–3096.
13. **Dayhoff, M. O., L. T. Hunt, and S. Hurst-Calderone.** 1978. Composition of proteins, p. 363–373. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5, supplement 3. National Biomedical Research Foundation, Washington, D.C.
14. **Dhar, R., R. M. Chanock, and C. J. Lai.** 1980. Nonviral oligonucleotides at the 5′ terminus of cytoplasmic influenza viral mRNA deduced from cloned complete genomic sequences. Cell **21:**495–500.
15. **Fields, S., and G. Winter.** 1982. Nucleotide sequences of influenza virus segments 1 and 3 reveal mosaic structure of a small viral RNA segment. Cell **28:**303–313.
16. **Fiers, W., R. Contreras, F. Duerinck, G. Haigeman, D. Iserentant, J. Merregaert, W. Minjou, F. Molemans, A. Raeymaekers, A. Van den Berghe, G. Volckaert, and M. Ysebaert.** 1976. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature (London) **260:**500–507.
17. **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.
18. **Greene, P. J., M. Gupta, H. W. Boyer, W. E. Brown, and J. M. Rosenberg.** 1981. Sequence analysis of the DNA encoding the *Eco*RI endonuclease and methylase. J. Biol. Chem. **256:**2143–2153.
19. **Gutte, B., M. Däumigen, and E. Wittschieber.** 1979. Design, synthesis and characterization of a 34-residue polypeptide that interacts with nucleic acids. Nature (London) **281:**650–655.
20. **Hay, A. J., B. Lomniczi, A. R. Bellamy, and J. J. Skehel.** 1977. Transcription of the influenza virus genome. Virology **83:**337–355.
21. **Helene, C., and G. Lancelot.** 1982. Interactions between functional groups in protein-nucleic acid associations. Prog. Biophys. Mol. Biol. **39:**1–68.
22. **Herz, C., E. Stavnezer, R. M. Krug, and T. Gurney, Jr.** 1981. Influenza virus, an RNA virus, synthesizes its messenger RNA in the nucleus of infected cells. Cell **26:**391–400.
23. **Hiti, A. L., A. R. Davis, and D. P. Nayak.** 1981. Complete sequence analysis shows that the hemagglutinin of the H0 and H2 subtypes of human influenza virus are closely related. Virology **111:**113–124.
24. **Hiti, A. L, and D. P. Nayak.** 1982. Complete nucleotide sequence of the neuraminidase gene of human influenza virus A/WSN/33. J. Virol. **41:**730–734.
25. **Horisberger, M. A.** 1980. The large P proteins of influenza A viruses are composed of one acidic and two basic polypeptides. Virology **107:**302–305.
26. **Irwin, M. J., J. Nyborg, B. R. Reid, and D. M. Blow.** 1976. The crystal structure of tyrosyl-transfer RNA synthetase at 2.7 Å resolution. J. Mol. Biol. **105:**577–586.
27. **Janin, J.** 1979. Surface and inside volumes of globular proteins. Nature (London) **277:**491–492.
28. **Kaptein, J. S., and D. P. Nayak.** 1982. Complete nucleotide sequence of the polymerase 3 gene of human influenza virus A/WSN/33. J. Virol. **42:**55–63.
29. **Kitamura, N., B. L. Semer, P. G. Rothberg, G. R. Larsen, C. J. Adler, A. J. Dorner, E. A. Emini, R. Hanecak, C. W. Anderson, and E. Wimmer.** 1981. Primary structure, gene organization and polypeptide expression of poliovirus RNA. Nature (London) **291:**547–553.
30. **Krug, R. M., B. A. Broni, and M. Bouloy.** 1979. Are the 5′ ends of influenza viral mRNAs synthesized in vivo donated by host mRNAs? Cell **18:**329–334.
31. **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.
32. **Lamb, R. A., and C. J. Lai.** 1980. Sequence of interrupted and uninterrupted mRNAs and cloned DNA coding for the two overlapping non-structural proteins of influenza virus. Cell **21:**475–485.
33. **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.
34. **Mahy, B. W. J., C. R. Penn, S. T. Nichol, and D. J. Briedis.** 1981. Influenza virus specific proteins regulating genome transcription. *In* D. P. Nayak (ed.), Genetic variation among influenza viruses. Academic Press, Inc., New York.
35. **Massicot, J. G., K. Van Wyke, R. M. Chanock, and B. R. Murphy.** 1982. Evidence for intrasegmental complementation between two influenza A viruses having ts mutants on their P1 genes. Virology **117:**496–500.
36. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74:**560–564.

37. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.

38. **Nayak, D. P.** 1980. Defective interfering influenza viruses. Annu. Rev. Microbiol. **34**:619–644.

39. **Nayak, D. P., N. Sivasubramanian, A. R. Davis, R. Cortini, and J. Sung.** 1982. Complete sequence analyses show that two defective interfering influenza viral RNAs contain a single internal deletion of a polymerase gene. Proc. Natl. Acad. Sci. U.S.A. **79**:2216–2220.

40. **Newman, A. K., R. A. Rubin, S. H. Kim, and P. Modrich.** 1981. DNA sequences of structural genes for *Eco*RI DNA restriction and modification enzymes. J. Biol. Chem. **256**:2131–2139.

41. **Palese, P., and M. B. Ritchey.** 1977. Live attenuated influenza virus vaccines. Strains with temperature sensitive defects in P3 protein and nucleoprotein. Virology **78**:183–191.

42. **Palese, P., M. B. Ritchey, and J. L. Schulman.** 1977. P1 and P3 proteins of influenza virus are required for complementary RNA synthesis. J. Virol. **21**:1187–1195.

43. **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.

44. **Pasek, M., T. Goto, W. Gilbert, B. Zink, H. Schaller, P. Mackey, G. Leadbetter, and K. Murray.** 1979. Hepatitis B virus genome and their expression in *Escherichia coli*. Nature (London) **282**:575–579.

45. **Plotch, S. J., M. Bouloy, and R. M. Krug.** 1979. Transfer of 5′ terminal cap of globin mRNA to influenza viral complementary RNA during transcription *in vitro*. Proc. Natl. Acad. Sci. U.S.A. **76**:1618–1622.

46. **Pons, M. W.** 1973. The inhibition of influenza virus RNA synthesis by actinomycin D and cycloheximide. Virology **51**:120–128.

47. **Porter, A. G., C. Barber, N. H. Carey, R. A. Hallewell, G. Threlfall, and J. S. Emtage.** 1979. Complete nucleotide sequence of an influenza virus hemagglutinin from cloned DNA. Nature (London) **282**:471–477.

48. **Queen, C. L., and L. J. Korn.** 1980. Computer analysis of nucleic acids and proteins. Methods Enzymol. **65**:595–609.

49. **Robertson, J. S.** 1979. 5′ and 3′ terminal nucleotide sequence of the RNA genome segments of influenza virus. Nucleic Acids Res. **6**:3745–3757.

50. **Robertson, J. S., M. Schubert, and R. A. Lazzarini.** 1981. Polyadenylation sites for influenza virus mRNA. J. Virol. **38**:157–163.

51. **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.

52. **Scholtissek, C.** 1979. Influenza virus genetics. Adv. Genet. **20**:1–36.

53. **Scholtissek, C.** 1978. The genome of the influenza virus. Curr. Top. Microbiol. Immunol. **80**:139–169.

54. **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.

55. **Ulmanen, I., B. A. Broni, and R. M. Krug.** 1981. Role of two of the influenza virus core P proteins in recognizing Cap 1 structures (m^{7}GpppNm) on RNAs and in initiating viral RNA transcription. Proc. Natl. Acad. Sci. U.S.A. **78**:7355–7359.

56. **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.

57. **Verhoryen, M., R. Fang, W. Min Jou, R. Devos, D. Huylebroeck, E. Saman, and W. Fiers.** 1980. Antigenic drift between the hemagglutinin of the Hong Kong influenza strains A/Aichi/2/68 and A/Victoria/3/75. Nature (London) **286**:771–775.

58. **Von Heuverswyn, H., A. Van de Voorde, and W. Fiers.** 1978. Nucleotide sequence of the Simian virus 40 *Hin*dII + III restriction fragment J and the total amino acid sequence of the major structural protein VP1. Eur. J. Biochem. **91**:415–430.

59. **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.

60. **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.

61. **Winter, G., and S. Fields.** 1982. Nucleotide sequence of human influenza A/PR/8/34 segment 2. Nucleic Acids Res. **10**:2135–2143.

62. **Winter, G., S. Fields, and G. G. Brownlee.** 1981. Nucleotide sequence of the hemagglutinin gene of a human influenza virus H1 subtype. Nature (London) **292**:72–75.