Volume 10 Number 6 1982 Nucleic Acids Research

Nucleotide sequence of human influenza A/PR/8/34 segment 2

Greg Winter and Stan Fields*

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Received 25 January 1982; Accepted 26 February 1982

ABSTRACT

The nucleotide sequence of RNA segment 2 of human influenza strain A/PR/8/34 has been determined. Segment 2 is 2341 nucleotides long and encodes a protein of 757 amino acids (86,500 daltons molecular weight) which is involved in RNA synthesis. Although segment 2 is identical in size to segment 1, which encodes a protein of related function, neither the nucleotide sequences of these two RNA segments nor the amino acid sequences of the encoded proteins appear to be homologous. The sequence of segment 2 completes the sequence of the virus (total 13,588 nucleotides).

INTRODUCTION

Influenza (A) virus contains eight single stranded RNA segments (1) which encode at least ten polypeptides (2-4). Each viral RNA (vRNA) segment in the nucleus of the host cell (5) serves as template for the synthesis of two types of complementary RNA (cRNA): the polyadenylated viral messenger RNA (A(+)cRNA) is synthesised first and subsequently the complementary RNAs (A(-)cRNA) necessary for viral replication (6). By contrast with the full length copies of A(-)cRNA, the A(+)cRNA is an incomplete copy of the vRNA, synthesis terminating at a stretch of uridine residues close to the 5' end of the vRNA (7,8). In addition, the 5' cap structure and the first 10-15 nucleotides of the A(+)cRNA are derived from cannibalised host cell mRNAs (9). The synthesis of vRNA and cRNA is catalysed by a viral transcriptase which, in the virion, can be identified with three large proteins (P1, P2 and P3) physically attached to a viral core of RNA and nucleoprotein (10-12). The P proteins are encoded by the three largest vRNA segments 1-3 (13,14) and temperature-sensitive mutations in these segments have been used to delineate their roles in synthesising vRNA and cRNA in the cell (15-18).

A priori, the involvement of the P proteins in RNA synthesis and the similarity of their sizes suggests that these molecules may have evolved from a single ancestral polymerase gene, and therefore we hoped that a

0305-1048/82/1006-2135$2.00/0

comparison of their sequences might reveal conserved residues critical for their structure or catalytic mechanism. We have undertaken a nucleotide sequence analysis of the human influenza strain A/PR/8/34 (19-24) and, with the sequence of segment 2 presented below, have now completed the sequence of the virus. However, as detailed below, a systematic comparison of the primary sequences of the P proteins with each other reveals no evidence of homology.

METHODS

Cloning of the segment 2 gene into pBR322

From 50 μg viral RNA, double-stranded DNA (dsDNA) corresponding to segments 1 and 2 was prepared essentially as in (24) except using dodecanucleotide and tridecanucleotide primers (20) which had been phosphorylated with T4 polynucleotide kinase (P-L. Biochemicals). The dsDNA was dissolved in 50 μl water and a 5 μl portion ligated to 20 ng of pBR322 that had been cut with PvuII and treated with phosphatase (25). After 6 h ligation *E. coli* strain 803 (26) was transfected (27) and transformants selected on ampicillin plates. Filters were prepared and hybridised (28) to a segment 2 probe (2 x 10^7 dpm Cerenkov). The probe was prepared by nick translation of 0.2 μg replicative form of an M13 clone which contained an insert corresponding to bases 1638-1938 of segment 2. Positive colonies, identified after 6 h exposure of the filters to pre-flashed film at -70° C (29), were grown in 20 ml culture and about 10 μg plasmid prepared (30). An analytical BamHI digest of the plasmid was used to further characterise the recombinants.

Cloning of segment 2 fragments into M13

dsDNA restriction fragments corresponding to all the segments were prepared and cloned into M13mp2 as described (19) and dsDNA restriction fragments corresponding to purified segments 1, 2 and 3 were prepared and cloned into M13mp7 (31) as described (24). A segment 2-pBR322 recombinant plasmid was subcloned into M13mp93 as follows: a 2 μg aliquot of plasmid was digested with BamHI and another 2 μg aliquot with PstI. 100 ng of each digest was ligated to 20 ng BamHI cut phosphatase-treated M13mp93 vector or 20 ng PstI cut M13mp93 as appropriate and used to transfect *E. coli* strain JM101 (32). The vector M13mp93 was kindly provided by Dr J. Messing and contains single BamHI and PstI sites.

DNA sequencing

Single-strand template was prepared from recombinant M13 plaques (33)

and sequenced by the dideoxy method (34) using either a 17-mer primer complementary to the sequence of M13 flanking the insert (35) or, if appropriate, the tridecanucleotide or the dodecanucleotide primers (20) complementary to the 5' or 3' ends of the influenza virus segments. Sequence data was compiled as in (36).

Comparison of segments 1, 2 and 3

An unpublished programme (DIAGON) was kindly provided by R. Staden in which all sections of segment 1 were scored systematically for the percentage of nucleotide identities with all sections of segment 2. These analyses were displayed on a graphics terminal, any matches appearing as diagonal lines. The sequences of the proteins encoded by segments 1 and 2 were also compared by DIAGON in a manner similar to McLachlan (37). Segments 1 and 3 and segments 2 and 3 were compared in the same manner.

RESULTS

Cloning and sequence strategies

Shotgun cloning of dsDNA restriction fragments derived from all the viral segments (19) yielded seven clones derived from segment 2. Shotgun cloning of dsDNA fragments from a mixture of polymerase segments yielded 89, 9 and 74 clones derived respectively from segments 1, 2 and 3. In view of the low proportion of segment 2 clones, dsDNA corresponding mainly to segment 2 was prepared by priming on total vRNA with suitable restriction fragment primers and shotgun cloning (details not supplied - see [38]). This yielded only seven more clones and the sequence of segment 2 remained incomplete with two gaps, bases 819-964 and bases 2042-2216 (Fig. 1).

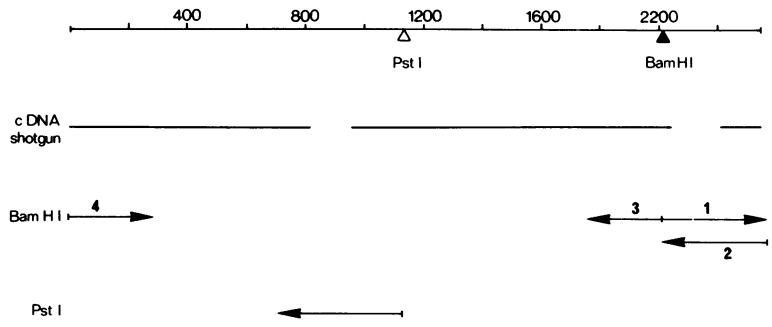

Fig. 1 Summary of sequence evidence for segment 2. Arrows marked 1-4 are derived from sequencing of clones 1-4 respectively, as in Fig. 2.

A new strategy was therefore used which involved ligating the full length gene into pBR322 (25), transfecting E. coli strain 803 and identifying the recombinants with a segment 2-M13mp7 probe. Although the lac gene in the probe does hybridise to the chromosomal DNA of E. coli 803 to produce a weak background signal, seven positive colonies were readily identified from the 500 screened. A BamHI digest of the recombinant plasmids confirmed that the segment had a single internal BamHI site and revealed that both orientations had been cloned.

The two gaps in the sequence were then covered by a directed strategy which took advantage, in turn, of a nearby BamHI site and a nearby PstI site in segment 2. BamHI cuts the recombinant plasmid once within the insert and once within the tetracycline gene of the parental pBR322 (39) to generate two large fragments whose exact sizes depend on the orientation of the insert. These fragments were cloned into the BamHI site of the vector M13mp93 and 12 recombinant plaques grown up and screened by dideoxy-T tracks (33) to identify the four types of recombinant clones (Fig. 2). Using these clones the gap between bases 2042 and 2216 was sequenced on both strands by priming with the 17-mer M13 primer (35) on clone 1 and with the tridecanucleotide flu primer (20) on clone 2 (Figs 1,2). Existing sequences derived from the M13 shotgun approach were double checked by priming with the 17-mer primer on clone 3 and the dodecanucleotide flu primer on clone 4 (Figs 1,2). Similarly the final gap (bases 819-964) was closed by cloning a

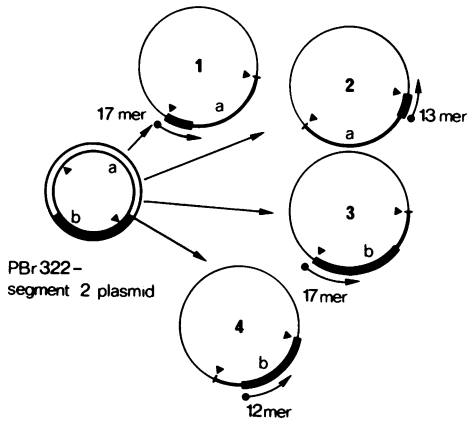

Fig. 2 Subcloning the BamHI fragments of pBR322-segment 2 plasmid into M13mp93 and sequencing with a 17-mer synthetic M13 primer and a 12-mer and 13-mer synthetic flu primer.

PstI digest of the recombinant plasmid into M13mp93 and sequencing a suitable clone (Fig. 1). Most of the sequence of segment 2 was covered on both strands (57%) or at least by two or more independent clones (82%).

Sequence of segment 2

Segment 2 of influenza virus A/PR/8/34 is 2341 nucleotides long (Fig. 3) and therefore has the same length as segment 1 (24). Segment 2 contains a single open reading frame of 757 amino acids beginning at the first AUG codon (base 25) and ending at a UAG codon (base 2295). The polypeptide encoded by segment 2 is predicted to have a basic charge of 28^{+} at pH 6.5, a molecular weight of 86,500 daltons and the amino acid composition detailed in Table 1.

A systematic comparison of the nucleotide sequences of segments 1, 2 and 3 using the DIAGON computer programme revealed no homologies apart from

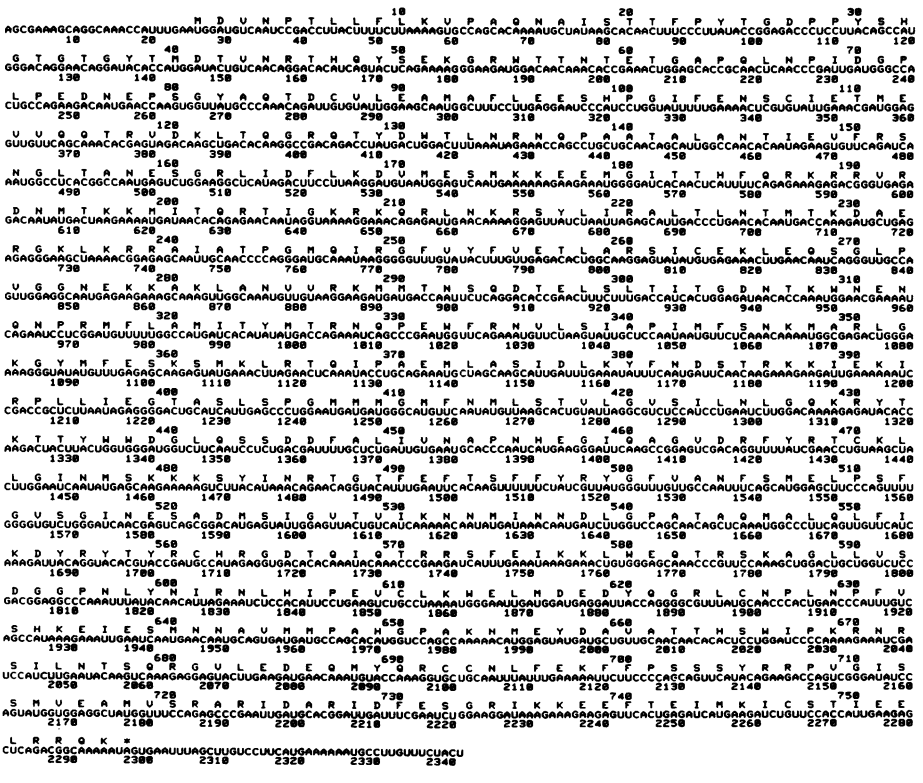

Fig. 3 The nucleotide sequence of A/PR/8/34 segment 2 with the amino acid sequence of the encoded protein P1.

Table 1 Amino acid composition of the protein encoded by A/PR/8/34 segment 2

Amino acid	Number	Amino acid	Number	Amino acid	Number
Alanine (A)	41	Glycine (G)	47	Proline (P)	32
Arginine (R)	54	Histidine (H)	10	Serine (S)	50
Asparagine (N)	49	Isoleucine (I)	46	Threonine (T)	61
Aspartic acid (D)	31	Leucine (L)	58	Tryptophan (W)	9
Cysteine (C)	10	Lysine (K)	49	Tyrosine (Y)	25
Glutamic acid (E)	49	Methionine (M)	40	Valine (V)	33
Glutamine (Q)	31	Phenylalanine (F)	32		757

the conserved 5' and 3' ends. Likewise a comparison of the sequences of the encoded proteins revealed no significant homologies. The best match was obtained with residues 45-57 of the segment 1 protein and residues 573-585 of the segment 3 protein (24). The calculated double matching probability (37) for this match was about 10^{-5} and is not significant.

DISCUSSION

Cloning and sequencing strategies

The sequences of human influenza virus A/PR/8/34 segments 1, 3, 4, 5, 6, 7 and 8 (19-24) were determined mainly by the shotgun cloning of influenza virus cDNA restriction fragments into the bacteriophage vector M13. Gaps in the sequences of segments 4 (22), 5 (23) and 7 (19) were covered by dideoxy sequencing using restriction fragments from suitable M13 clones as primers on total vRNA. The gaps in segments 5 and 7 were probably attributable to EcoK sites since the host JM101 (32) is EcoK+ and EcoK sites are located within the single gap of segment 5 (23) and within both gaps of segment 7 (19). However, the gaps in segment 2 cannot be attributed to EcoK sites or even to a lack of suitable restriction sites in the cDNA. The gaps exist because we were unable to generate sufficient clones to ensure complete coverage. This may reflect much smaller quantities of segment 2 compared with the other segments in the viral RNA preparation. We therefore employed an alternative strategy in which we first cloned the full length gene into pBR322 (25) and then subcloned the recombinant plasmid into M13 for sequencing. The strategy is effective and is recommended for cloning and

sequencing segments from other strains of influenza, especially if the amount of RNA is limited.

Sequence of segment 2

Segment 2 from A/PR/8/34 is exactly the same length as segment 1 (24) and was initially identified by its close homology to the terminal sequences of segment 2 of the strain A/FPV/Rostock/34 (40). Since a matching of the 3' terminal sequences presented for the vRNA segments of strain FPV (40) with those of PR8 (41) clearly identifies segments 1 and 3 of FPV with segments 1 and 3 respectively of PR8, segment 2 of FPV should, by inference, correspond to segment 2 of PR8. We therefore believe that our sequence corresponds to segment 2 of PR8 although this is at variance with some genetic evidence (reviewed in Refs 42, 43).

The basic charges predicted for the segment 2 and segment 1 proteins and the acidic charge predicted for the segment 3 protein (24) are consistent with the experimental observation that the large P proteins of several strains of influenza (A) viruses are composed of one acidic and two basic polypeptides (44). We would further predict that the larger of the two basic proteins is encoded by segment 2, the smaller by segment 1 and the acidic protein by segment 3 (Table 2). The observation that in A/WSN/33 the P1 and P3 proteins are basic and the P2 protein is acidic (45) agrees entirely with this prediction (see Table 2) since the gene assignment in strain WSN is identical to that of PR8 (17).

It has been suggested that in strain WSN both the P3 and P1 proteins are involved in the synthesis of cRNA (17) and that the P3 protein (segment 1) recognises the 5' terminal cap on mRNAs while the P1 protein (segment 2) catalyses the initiation of transcription (45). Similarly, in strain FPV the P2 protein (segment 1) seems to be directly involved in the mRNA priming reaction (46). In view of their related functions in cRNA synthesis and the evidence that RNA segments 1 and 2 are identical in length and encode two

Table 2 Segments 1, 2 and 3 and their encoded proteins

Segment	Length (nucleotides)	Length of encoded polypeptide (amino acids)	Molecular weight of encoded polypeptide	Net charge on polypeptide at pH 6.5	Gene assignment (13) in A/PR/8/34	Gene assignment (14) in A/FPV/Rostock/34
1	2341	759	85,900	+28	P3	P2
2	2341	757	86,500	+28	P1	P1
3	2233	716	82,400	-13.5	P2	P3

basic proteins of closely similar molecular weight, it is surprising to find such a complete lack of homology in their sequences. An underlying but undetected similarity in the three dimensional structure of these proteins might nonetheless exist. The high rate of mutation of RNA genes (47,48) might have eroded sequence homologies while retaining the overall folding pattern of the protein chain, as can be illustrated by the structural repeats in tobacco mosaic virus coat protein, which are virtually undetectable at the sequence level (49). If there is indeed no underlying homology in these proteins, we must suppose that the identity in lengths of segment 1 and 2 is either pure coincidence or is governed by some other constraint such as the packing of the ribonucleoprotein cores together into the virion.

ACKNOWLEDGEMENTS

It is a pleasure to thank Professor G.G. Brownlee for a preprint describing blunt-end cloning into pBR322, Drs F. Sanger and M.S. Neuberger for practical details, Mr R. Staden for an unpublished computer programme (DIAGON) and Drs P.J.G. Butler, B.W.J. Mahy and J.S. Robertson for helpful discussions.

*Present address: Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143, USA

REFERENCES

1. McGeoch, D., Fellner, P. and Newton, C. (1976) Proc. Nat. Acad. Sci. USA 73, 3045-3049
2. Inglis, S.C., McGeoch, D.J. and Mahy, B.W.J. (1977) Virology 78, 522-536
3. Lamb, R.A. and Lai, C-J. (1980) Cell 21, 475-485
4. Lamb, R.A. and Choppin, P.W. (1981) Virology 112, 729-737
5. Herz, C., Stavnezer, E., Krug, R.M. and Gurney, T. (1981) Cell 26, 391-400
6. Hay, A.J., Lomniczi, B., Bellamy, A.R. and Skehel, J.J. (1977) Virology 83, 337-355
7. Skehel, J.J. and Hay, A.J. (1978) Nucleic Acids Res. 5, 1207-1219
8. Robertson, J.S., Schubert, M. and Lazzarini, R.A. (1981) J. Virol. 38, 157-163
9. Krug, R.M., Broni, B.A. and Bouloy, M. (1979) Cell 18, 329-334
10. Plotch, S.J. and Krug, R.M. (1978) J. Virol. 25, 579-586
11. Plotch, S.J., Bouloy, M. and Krug, R.M. (1979) Proc. Nat. Acad. Sci. USA 76, 1618-1622
12. Bishop, D.H.L., Roy, P., Bean, W.J. and Simpson, R.W. (1972) J. Virol. 10, 689-697
13. Palese, P. (1977) Cell 10, 1-10
14. Almond, J.W. and Barry, R.D. (1979) Virology 92, 407-415
15. Almond, J.W., McGeoch, D. and Barry, R.D. (1978) Virology 92, 416-427

16. Mahy, B.W.J., Barrett, S.T., Nichol, S.T., Penn, C.R. and Wolstenholme, A.J. (1981) in Developments in Cell Biology, Vol 7, pp 379-387, Elsevier/North Holland, Amsterdam
17. Palese, P., Ritchey, M.B. and Schulman, J.L. (1977) J. Virol. 21, 1187-1195
18. Ritchey, M.B. and Palese, P. (1977) J. Virol. 21, 1196-1204
19. Winter, G. and Fields, S. (1980) Nucleic Acids Res. 8, 1965-1974
20. Winter, G., Fields, S., Gait, M.J. and Brownlee, G.G. (1981) Nucleic Acids Res. 9, 237-245
21. Fields, S., Winter, G. and Brownlee, G.G. (1981) Nature 290, 213-217
22. Winter, G., Fields, S. and Brownlee, G.G. (1981) Nature 292, 72-75
23. Winter, G. and Fields, S. (1981) Virology 114, 423-428
24. Fields, S. and Winter, G. (1982) Cell 28, 303-313
25. Huddleston, J.A. and Brownlee, G.G. (1982) Nucleic Acids Res. 10, 1029-1038
26. Murray, N.E., Brammar, W.J. and Murray, K (1977) Mol. gen. Genetics 150, 53-61
27. Dagert, M. and Ehrlich, S.D. (1979) Gene 6, 23-28
28. Grunstein, M. and Hogness, D.S. (1975) Proc. Nat. Acad. Sci. USA 72, 3961-3965
29. Laskey, R.A. and Mills, A.D. (1977) FEBS Lett. 82, 314-316
30. Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523
31. Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucleic Acids Res. 9, 309-321
32. Messing, J. (1979) Recombinant DNA Technical Bulletin 2, 43-48
33. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178
34. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. USA 74, 5463-5467
35. Duckworth, M.L., Gait, M.J., Goelet, P., Hong, G.F., Singh, M. and Titmas, R.C. (1981) Nucleic Acids Res. 9, 1691-1706
36. Staden, R. (1980) Nucleic Acids Res. 8, 3673-3694
37. McLachlan, A.D. (1971) J. Mol. Biol. 61, 409-424
38. Gething, M.J., Bye, J., Skehel, J. and Waterfield, M. (1980) Nature 287, 301-306
39. Sutcliffe, J.G. (1978) Cold Spring Harbor Symp. Quant. Biol. 43, 77-90
40. Robertson, J.S. (1979) Nucleic Acids Res. 6, 3745-3757
41. Desselberger, U., Racaniello, V.R., Zazra, J.T. and Palese, P. (1980) Gene 8, 315-328
42. Scholtissek, C. (1978) Curr. Topics Microbiol. Immunol. 80, 139-169
43. Scholtissek, C. (1979) Adv. Genetics 20, 1-36
44. Horisberger, M.A. (1980) Virology 107, 302-305
45. Ulmanen, I., Broni, B.A. and Krug, R.M. (1981) Proc. Nat. Acad. Sci. USA 78, 7355-7359
46. Nichol, S.T., Penn, C.R. and Mahy, B.W.J. (1982) J. Gen. Virol. 57, 407-413
47. Ortin, J., Najera, R., Lopez, C., Davila, M. and Domingo, E. (1980) Gene 11, 319-331
48. Fields, S. and Winter, G. (1981) Gene 15, 207-214
49. McLachlan, A.D., Bloomer, A.C. and Butler, P.J.G. (1980) J. Mol. Biol. 136, 203-224