Volume 9 Number 2 1981 Nucleic Acids Research

The use of synthetic oligodeoxynucleotide primers in cloning and sequencing segment 8 of influenza virus (A/PR/8/34)

Greg Winter, Stan Fields, Michael J.Gait
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, and

George G.Brownlee
The Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

Received 21 November 1980

ABSTRACT

Complete double-stranded DNA copies of the RNA genes of the human influenza virus A/PR/8/34 have been synthesized by using two synthetic oligodeoxynucleotide primers. The gene encoding the non-structural proteins NS1 and NS2, prepared with these primers, has been cloned into the bacteriophage M13mp7 and sequenced. The sequence is compared with that from another human strain and from an avian strain.

INTRODUCTION

Influenza virus has a segmented genome of eight single-stranded RNAs [1] totalling about 14,000 nucleotides [2,3]. Each gene has a region of 12 nucleotides in common with the others at the 3' end, and another region of 13 nucleotides in common at the 5' end [4,5]. As part of a strategy to clone and sequence the entire genome of influenza virus, we used a dodecanucleotide complementary to the 3' end of each vRNA as this had been shown by others to prime full length single-stranded (ss) DNA copies of all the RNA genes [6,7]. A tridecanucleotide complementary to the 3' end of the ssDNA was then used to prime second strand synthesis and thus produce full length double-stranded (ds) DNA copies. Using these two primers we have cloned and obtained sequences from full length copies of the gene encoding the non-structural proteins NS1 and NS2. Together with sequences derived from a "shotgun" of total viral dsDNA [8], this completes the sequence of the gene.

MATERIALS AND METHODS

Primer synthesis

The dodecanucleotide d(A-G-C-A-A-A-A-G-C-A-G-G) was chemically synthesized by the solid phase phosphotriester method on a polydimethylacrylamide support as described previously [9,10]. In the synthesis of the tridecanucleotide d(A-G-T-A-G-A-A-A-C-A-A-G-G) the following significant

improvements were made: (1) Acidic removal of terminal 5'-O-dimethoxytrityl protecting groups was effected by 2 x 2 min treatments with 10% trichloroacetic acid in chloroform. Methanol was omitted from this reaction since a very slow esterification of the acid was found to occur simultaneously leading to the formation of water droplets. The water contaminating the acidic reagent became absorbed to the resin during deprotection and in small scale synthesis was found to inhibit coupling reactions. (2) The coupling agent, mesitylenesulphonyl-3-nitro-1,2,4-triazole, was used in coupling reactions instead of the previously used triisopropylbenzenesulphonyl tetrazole owing to greater stability and faster coupling rate [11]. The tridecanucleotide was purified by ion exchange and reverse phase high performance liquid chromatography as described previously [9,10].

Preparation of dsDNA

DNA for the M13 shotgun was prepared as described previously [8]. Full length ssDNAs were prepared from 50 µg viral RNA (about 80 pmole of 3' ends), 400 pmole dodecanucleotide, 90 u reverse transcriptase (generously supplied by Dr. J. Beard) in 250 µl 50 mM Tris-Cl pH 8.3, 10 mM $MgCl_2$, 70 mM KCl, 80 mM 2-mercaptoethanol and 0.5 mM deoxynucleotide triphosphates (dNTPs). The mixture was incubated at 37°C for 1 h, boiled for 1 min, chilled at -70°C, the protein precipitate removed by centrifugation, 5 µl (10 mg/ml) RNase A (Worthington) added, incubated at 37°C for 30 min and ssDNA purified by phenol extraction and ethanol precipitation. A sealed capillary containing the ssDNA and 400 pmole tridecanucleotide in 200 µl 10 mM Tris-Cl pH 8.0, 10 mM $MgCl_2$, 50 mM NaCl, 0.1 mM dithiothreitol (DTT) was boiled for 1 min and left to cool slowly to room temperature. The second strand was synthesized by adding dNTPs to 50 µM, 50 µCi α-^{32}P-dATP (Amersham, U.K.) and 4 u of the Klenow subfragment of DNA polymerase (Boehringer). After 1 h at room temperature the dsDNA was extracted with phenol and collected as the breakthrough of a 1 ml column of Sepharose 4B, precipitated with ethanol and dissolved in 50 µl water.

Ligation into M13

*Taq*I, *Alu*I and *Sau*3A fragments from restriction digests of total viral dsDNA were ligated as described previously [8], except that the *Taq*I fragments were cloned into the *Acc*I site of M13mp7 [12]. Full length dsDNA was phosphorylated and blunt end ligated into the *Hind*II site of M13mp7. In the absence of dsDNA, the *Hind*II cut vector gave a high proportion (10%) of white plaques [13,14] on religation. These plaques, presumably resulting from exonuclease in the *Hind*II which destroys the

reading frame of the α-peptide of β-galactosidase, cannot be distinguished from the white plaques produced by the insertional inactivation of the α-peptide of β-galactosidase. Religation of the vector was therefore minimised by pretreating the HindII cut M13mp7 with calf intestinal phosphatase (Worthington). A typical ligation contained dsDNA derived from 1 μg viral RNA, 10 ng HindII cut phosphatase treated M13mp7 vector, 0.4 u T4 ligase in 50 mM Tris-Cl pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, 1 mM rATP and was incubated for 24 h at 14°C.

Transformation and sequencing

Transformation and purification of recombinant phage was as described earlier [8], except that plaques were grown for 6 h at 37°C. Clones were sequenced by the dideoxy chain termination method [15] using the cloned primer from the plasmid pSP14 [16]. Sequences were stored on computer files and aligned using the DBUTIL programs [17].

RESULTS

Use of the tridecanucleotide primer

The dodecamer primer yielded full length ssDNA copies of all the influenza genes and these could be back-copied successfully with the tridecanucleotide primer to give apparently full length dsDNA copies of at least genes 4-8 (Fig. 1). Presumably genes 1-3 were also back-copied although the sensitivity here is not sufficient to show this.

Cloning of the mixture of dsDNA derived from 5 μg influenza virus vRNA yielded about 50 recombinant phage. The first four characterised were shown by dideoxy sequencing to contain either one or the other end of segment 8. "Clone turn-around", in which the replicative form of a clone is cut and religated so as to turn the insert around, [8] was used to prove that each of these clones was complete at both ends, and therefore that each represented a full length copy of the band 8 gene.

The sequencing of segment 8

The nucleotide sequence of the 5' and 3' ends of segment 8 was determined from the two orientations of the cloned full length dsDNA. These sequences were extended and overlapped by sequences from the TaqI, AluI and Sau3A shotgun of total influenza dsDNA to give the sequence of the complete gene (Fig. 2). Most regions of the sequence (65%) are covered by two or more independent clones and most of these regions (85%) are covered on both strands.

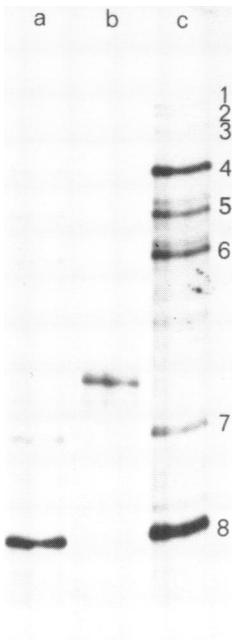

Figure 1. Electrophoresis of influenza complementary DNAs on a denaturing polyacrylamide-7 M urea thin gel [18]. Samples were loaded after boiling for 1 min in formamide dyes. In condition (A), the total cold ssDNA was primed with tridecanucleotide and labelled with α-^{32}P-dATP: the bands in (A) therefore correspond to the second strand DNA. A control (B) was included in which the cold ssDNA was primed in the absence of the tridecanucleotide and labelled with α-^{32}P-dATP: the bands in (B) therefore correspond to self-primed ssDNA. In condition (C), total vRNA was primed with the dodecanucleotide and labelled with α-^{32}P-dATP: the bands in (C) therefore correspond to the first strand DNA.

DISCUSSION

Use of tridecanucleotide primer

The use of two synthetic oligodeoxynucleotides to prime both first and second strand synthesis yields full length flush ended dsDNA copies of most or all of the influenza genes. Short dsDNA copies should also be expected from this technique and could derive from two sources. Firstly, self-priming of the ssDNA by 3' loopback [19] could result in short dsDNA copies with a loop at one end of the DNA. Secondly, non-specific priming of the dodecanucleotide on the vRNA could yield short transcripts with a common 3' end: the tridecanucleotide would then prime specifically on these strands to yield short dsDNA copies with flush ends. Assuming such non-specific priming can be minimised by optimising the primer to template ratio in the first strand synthesis, blunt end ligation of the dsDNA into the vector should generally select the full length flush ended genes from the incomplete transcripts. A limited analysis of the recombinant clones did, nevertheless, identify several incomplete copies of some influenza virus genes.

This approach compares favourably with the main alternatives of either 3' terminal loopback of the ssDNA followed by S_1 nuclease [19], or of hybridising cDNA transcripts of vRNA and viral mRNA [6,20]. The first alternative necessarily yields incomplete dsDNA while the latter requires

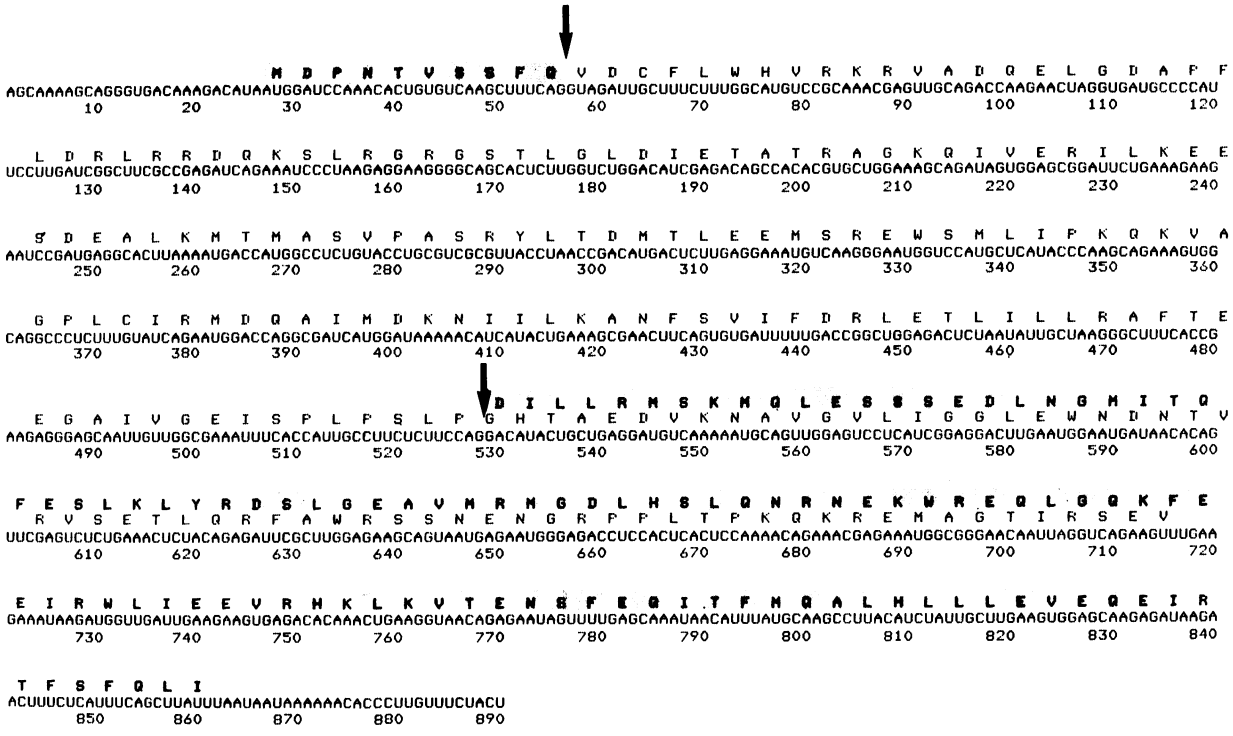

Figure 2. The nucleotide sequence of A/PR/8/34 band 8 in the mRNA sense (complementary to vRNA). The amino acid sequences predicted by the two open reading frames are depicted in the one letter amino acid code. The second frame, and the common N-terminus to NS1 and NS2, are cross-hatched, while the splice junctions are indicated by arrows.

isolation of both vRNA and mRNA.

The sequence of segment 8

The nucleotide sequence (Fig. 2) contains two overlapping reading frames, the first encoding the NS1 protein and the second encoding most of the NS2 protein [21-25]. The proteins are produced from two different mRNAs, the smaller NS2 mRNA apparently generated from the NS1 mRNA by RNA splicing between positions 56-57 and 528-529 [25].

Comparison of the nucleotide sequence of A/PR/8/34 band 8 with that from an avian strain (A/FPV/Rostock/34) [26] and a more recent human strain (A/Udorn/72) [25] demonstrates 91% nucleotide conservation (Table 1) and so generally confirms the high level of conservation estimated by hybridising mRNA and vRNA from different influenza strains [27]. However, a suggestion from the hybridisation studies [27] that band 8 from human and avian strains falls into two separate groups is not confirmed by the sequence comparison. Band 8 from A/PR/8/34 is, in fact, more closely related to that from A/FPV/Rostock/34 than to band 8 from the other human strain A/Udorn/72. The nucleotide sequence at the splice junction is precisely conserved and the sequence involved in the overlap of the NS1 and NS2 reading frames is more conserved than is the rest of the gene (Table 1).

A comparison of the amino acid sequence of the NS1 and NS2 proteins from the three strains (Fig. 3) shows that the NS2 protein is more conserved than the NS1 protein, and that in the overlap of NS1 and NS2 reading frames, the NS2 protein is conserved at the expense of the NS1 protein [25] (Table 1). The third position of each codon in the NS2 frame aligns

Table 1. Conservation of nucleotide and amino acid sequence in a comparison of segment 8 from strains A/PR/8/34, A/FPV/Rostock/34 [26] and A/Udorn/72 [25].

	PR8/FPV	PR8/Udorn	Udorn/FPV
Nucleotides conserved (%) in the region spanning nucleotides:			
1-890 (complete band)	92	91	89
27-527 (N-terminus NS1)	90	90	87
528-717 (NS1, NS2 overlap)	93	93	92
718-861 (C-terminus NS2)	93	91	90
Amino acids conserved (%) in the region spanning nucleotides:			
27-716 (NS1)	91	88	85
528-716 (NS1, in region of overlap)	86	87	84
27-56, 529-861 (NS2)	92	93	92
529-717 (NS2, in region of overlap)	92	95	94

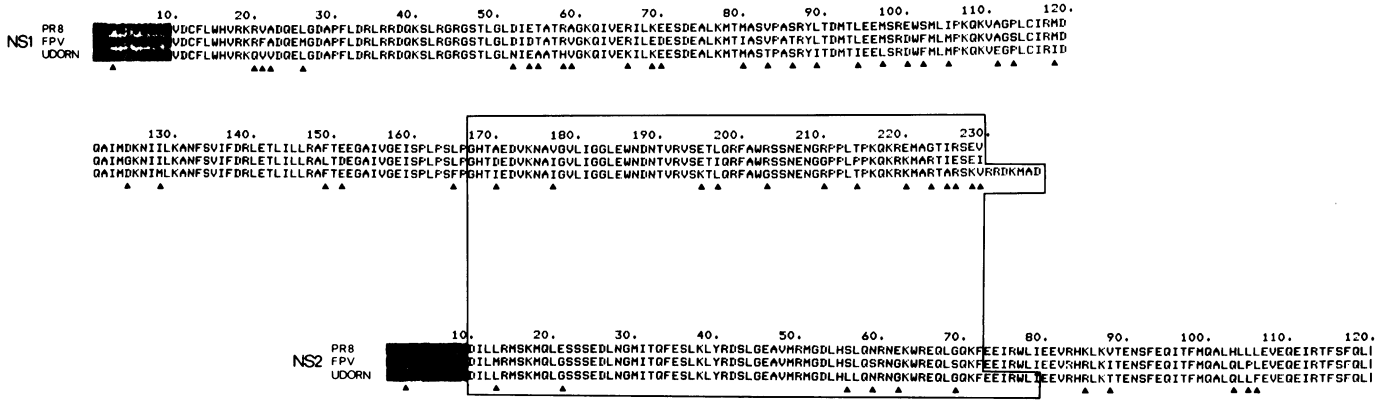

Figure 3. The amino acid sequences of NS1 and NS2 from A/PR/8/34, A/FPV/Rostock/34 [26] and A/Udorn/72 [25] are aligned and the portion of polypeptide encoded by the overlapping reading frames are boxed. Sequence differences between the three strains are marked thus: ▲, and the common N-terminus to NS1 and NS2 [25] is cross-hatched.

with the first position of each codon in the NS1 frame, with the result that third position changes in the NS2 frame, which would usually leave an amino acid unchanged in character, are likely to provoke radical changes in the NS1 frame. Thus the character of substitution in the NS1 protein is more radical in the overlap region, as in the arginine to glycine changes at residues 204, 211 and 224 (Fig. 3).

It is interesting to speculate how the overlapping NS1 and NS2 genes evolved. Two simple models are suggested. In the first model the NS1 and NS2 genes are originally colinear on the vRNA, but not overlapping, and with the NS2 mRNA produced by splicing of the larger NS1 mRNA as at present. Readthrough of a terminator at the end of NS1 then allows the NS1 protein to encroach into an unused frame of the NS2 gene. The extra C-terminal appendage of the NS1 protein in the A/Udorn/72 strain might therefore reflect a further stage in the encroachment of NS1 into the NS2 frame and the model is consistent with the conservation of NS2 protein at the expense of the NS1. In the second model the NS1 protein originally occupies the entire vRNA but as a result of mutations which fortuitously generate a splice site in the NS1 mRNA, a second shorter mRNA is produced which encodes the ancestral NS2 protein. In time, the NS2 protein becomes more critical for the function of the virus than the C-terminal portion of the NS1 protein, which is then progressively lost. It is not easy to choose between these two models.

ACKNOWLEDGEMENTS

We would like to thank M. Singh for assistance in oligonucleotide synthesis, Dr. B.G. Barrell for discussions, M. Robertson for preparing the virus, R. Staden for help with the DBUTIL computer programs and Drs. R.A. Lamb and A. Porter for distributing copies of their band 8 sequences before publication. S.F. acknowledges support of an NSF pre-doctoral fellowship.

REFERENCES

1. McGeoch, D., Fellner, P. and Newton, C. (1976) Proc. Nat. Acad. Sci. USA 73, 3045-3049.
2. Desselberger, U. and Palese, P. (1978) Virology 88, 394-399.
3. Sleigh, M.J., Both, G.W. and Brownlee, G.G. (1979) Nucleic Acids Res. 6, 1309-1321.
4. Skehel, J.J. and Hay, A.J. (1978) Nucleic Acids Res. 5, 1207-1218.
5. Robertson, J.S. (1979) Nucleic Acids Res. 6, 3745-3757.
6. Lai, C.J., Markoff, L.J., Zimmerman, S., Cohen, B., Berndt, J.A. and Chanock, R.M. (1980) Proc. Nat. Acad. Sci. USA 77, 210-214.
7. Cummings, I. and Salser, W. (1980) in "Structure and variation in influenza virus", eds. G. Laver and G. Air, Developments in Cell Biology, Vol. 5, pp. 147-155, Elsevier/North-Holland, Amsterdam.

8. Winter, G. and Fields, S. (1980) Nucleic Acids Res. 8, 1965-1974.
9. Gait, M.J., Singh, M., Sheppard, R.C., Edge, M.D., Greene, A.R., Heathcliffe, G.R., Atkinson, T.C., Newton, C.R. and Markham, A.F. (1980) Nucleic Acids Res. 8, 1081-1096.
10. Gait, M.J., Popov, S.G., Singh, M. and Titmas, R.C. (1980) Nucleic Acids Research Symposium Series No. 7 (1980) 243-258.
11. Gait, M.J. and Popov, S.G. (1980) Tetrahedron Letters 21, 2841-2842.
12. Messing, J., Crea, R. and Seeburg, P.H., Proc. Nat. Acad. Sci., in press.
13. Messing, J., Gronenborn, B., Müller-Hill, B. and Hofschneider, P.H. (1977) Proc. Nat Acad. Sci. USA 74, 3642-3646.
14. Gronenborn, B. and Messing, J. (1978) Nature (London) 272, 375-377.
15. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. USA 74, 5463-5467.
16. Anderson, S., Gait, M.J., Mayol, L. and Young, I.G. (1980) Nucleic Acids Res. 8, 1731-1743.
17. Staden, R. (1980) Nucleic Acids Res. 8, 3673-3694.
18. Sanger, F. and Coulson, A.R. (1978) FEBS Lett. 87, 107-110.
19. Efstratiadis, A., Kafatos, F.C., Maxam, A.M. and Maniatis, T. (1976) Cell 7, 279-288.
20. Caton, A.J. and Robertson, J.S. (1980) Nucleic Acids Res. 8, 2591-2603.
21. Inglis, S.C., Barrett, T., Brown, C.M. and Almond, J.W. (1979) Proc. Nat. Acad. Sci. USA 76, 3790-3794.
22. Lamb, R.A. and Choppin, P.W. (1979) Proc. Nat. Acad. Sci. USA 76, 4908-4912.
23. Lamb, R.A., Choppin, P.W., Chanock, R.M. and Lai, C.J. (1980) Proc. Nat. Acad. Sci. 77, 1857-1861.
24. Inglis, S.C., Gething, M.-J. and Brown, C.M. (1980) Nucleic Acids Res. 8, 3575-3589.
25. Lamb, R.A. and Lai, C.J. (1980) Cell 21, 475-485.
26. Porter, A.G., Smith, J.C. and Emtage, J.S. (1980) Proc. Nat. Acad. Sci. USA 77, 5074-5078.
27. Scholtissek, C. and Von Hoyningen-Huene, V. (1980) Virology 102, 13-20.