PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Nucleotide Sequences from the Terminal Regions of Fowl Plague Virus Genome RNA

J. S. Robertson

Phil. Trans. R. Soc. Lond. B 1980 **288**, 371-374
doi: 10.1098/rstb.1980.0013

Phil. Trans. R. Soc. Lond. B **288**, 371–374 (1980)
Printed in Great Britain

Nucleotide sequences from the terminal regions of fowl plague virus genome RNA

By J. S. Robertson
Division of Virology, Department of Pathology, University of Cambridge, Laboratories Block, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, U.K.

The genome of influenza virus consists of eight discrete single-stranded RNA segments each of which codes for a unique polypeptide species (McGeoch *et al.* 1976; Inglis *et al.* 1977; Inglis & Almond, this symposium). These viral polypeptides are synthesized from mRNA molecules that are complementary to the viral RNA (vRNA) and the 3′ ends of which lack sequences corresponding to the 5′ terminus of vRNA (Hay *et al.* 1977). I have determined the nucleotide sequences of the 5′ and 3′ termini of the eight RNA segments of fowl plague virus (FPV), an avian influenza A strain, by recently developed methods for direct RNA sequencing. These methods involve radiolabelling of the RNA segments *in vitro* at either the 5′ end or the 3′ end followed by partial digestion of individual segments with specific endoribonucleases and analysis of the products by one-dimensional or two-dimensional polyacrylamide gel electrophoresis (Donis-Keller *et al.* 1977; Simoncsits *et al.* 1977; England & Uhlenbeck 1978; Lockard *et al.* 1978).

Table 1 shows the nucleotide sequence of the 5′ terminal regions of the eight RNA segments of FPV. Each segment contains a common region of 13 nucleotides at the 5′ end as has been reported previously by Skehel & Hay (1978), followed by a distinctive triplet sequence (underlined in table 1) which is unique in each segment, with the exception of segments 4 and 8 (the HA and NS genes respectively) which contain an identical triplet. These triplets are followed by a region of 5–7 uridine residues after which each segment contains a unique nucleotide sequence for the additional 40–60 nucleotides analysed. The 5′ terminal region of vRNA, with the exception of the first 25–30 residues, corresponds to the 3′ end of mRNA (Skehel & Hay 1978) and several protein termination codons can be identified (not shown); however, whether or not any of these termination codons are utilized *in vivo* remains to be determined.

The nucleotide sequences derived from the 3′ termini of the RNA segments of FPV are shown in table 2 and reveal a common region of 12 nucleotides at the extreme 3′ end of each segment. These sequences complement those reported by Skehel & Hay (1978) who found a common sequence of 12 nucleotides at the 5′ end of *in-vitro* transcripts of vRNA. Beyond this common region each segment contains a unique sequence. It was first observed by Skehel & Hay (1978) that in any one segment a hexanucleotide comprising the distinctive triplet underlined in table 1 plus the adjacent three nucleotides of the common region (i.e. residues 11–16 of the 5′ end) was exactly complementary to a hexanucleotide comprising the triplet underlined in table 2 plus the adjacent three nucleotides of the common region (i.e. residues 10–15 of the 3′ end). Tables 1 and 2 indicate that this complementarity exists in segments 3, 5, 6, 7 and 8 but not in segment 4. Preliminary data (not shown) suggest that this complementarity also exists in segments 1 and 2.

The 5′ terminal nucleotide sequence of influenza mRNA can be deduced by complementarity

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

TABLE 1. 5′ TERMINAL SEQUENCES OF FOWL PLAGUE VIRUS

segment	25	50
8	AGUAGAAACAAGGGUGUUUUUUAUCAUUAAAUAAGCUGAAACGAGAAAGUUCUUAUCUCUUGCUCCA	
7	AGUAGAAACAAGGUAGUUUUUUACUCCAGCUCUAUGUUGACAAAAUGACCAUCGUCAACAUCCACAGCACUCUGCUGUUCCUGCCGA	
6	AGUAGAAACAAGGAGUUUUUUUGAACAAACUACUUGUCAAUGGUGAAUGGCAACUCAGCACCGUCUGG	
5	AGUAGAAACAAGGGUAUUUUUCUUUAAUUGUCAUACUCCUCUGCAUUGUCUCCGAAGAAAUAAGAUCCCUCAUUA	
4	AGUAGAAACAAGGGUGUUUUUUUCCAAACUUAUAUACAAAUAGUGCACCGCAUGUUUCCGUUCUUCACACA	
3	AGUAGAAACAAGGUACUUUUUUGGACAGUAUGGAUAGCAAAUAGUAGCAUUGCCACAACUAUUUCAGUGCAUG	
2	AGUAGAAACAAGGCAUUUUUUUCAUGAAGGACAAGCUAAAUUCACUAUUUCUGCCGUCUGAGCUCUU	
1	AGUAGAAACAAGGUCGUUUUUAAACAAUUCGACAUUAAUUGAUGGCCAUCCGAAUUCUUUUGG	

TABLE 2. 3′ TERMINAL SEQUENCES OF FOWL PLAGUE VIRUS

50	25	segment
AAAGAAAGCAGUCUACCUGAAAGCUUGACACAGUGUUGGAAUCCAUUAUGUUUUUGUCACCCUGCUUUUGCU		8
CGGGACGACAGAGAGAACGUACGUUCCAACCUCGGUUAGAAGACUCAUCUUUAAAUAUCUACCUGCUUUUGCU		7
GACCCAAUGGUUAUUAUUUUCUGAUUUGGAUUCAUUUUGAACUCCUGCUUUUGCU		6
GUCCUGAGACGCCAUGAUAUGGACGCCACUCAGUGAGUGAUUAUAUACCCUGCUUUUGCU		5
GCAAGGGCGAAAACCAGGAUUUGAGUGUUCAUUUUGUAACCCCUGCUUUUGCU		4
GGAUUGAAGCAUUGACGCACAAAUUCUUCCAUUUUGGAUCAGUACCUGCUUUUGCU		3
	CCUGCUUUUGCU	2
	CCUGCUUUUGCU	1

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

TABLE 3. 5′ TERMINAL SEQUENCES OF THE mRNA OF FOWL PLAGUE VIRUS†

segment		25														
8	AGCAAAAGCAGGGUGAGAAAAACAUAAUG	Asp GAU	Ser UCC	Asn AAC	Thr ACU	Val GUG	Ser UCA	Ser AGC	Phe UUU	Glu CAG	Val GUA	Asp GAC	Cys UGC	Phe UUU	Leu CUU	
7	AGCAAAAGCAGGUAGAUAUUUAAAGAUG	Ser AGU	Leu CUU	Leu CUA	Thr ACC	Glu GAG	Val GUU	Gly GGA	Thr ACG	Tyr UAC	Val GUU	Leu CUC	Ser UCU	Val GUC	Val GUC	Pro CCG
6	AGCAAAAGCAGGAGUUCAAAAUG	Asn AAU	Pro CCA	Asn AAU	Gln CAG	Lys AAA	Ile AUA	Ile AUA	Thr ACC	Ile AUU	Gly GGG	Ser UC				
5	AGCAAAAGCAGGGUAUAUAAUCACUCACUGAGUGGCGUCCAUAUCAUG						Ala GCG	Ser UCU	Gln CAG	Asp GAC						
4	AGCAAAAGCAGGGGUUACAAAAUG	Asn AAC	Thr ACU	Gln CAA	Ile AUC	Leu CUG	Val GUU	Phe UUC	Ala GCC	Leu CUU	Ala GC					
3	AGCAAAAGCAGGUACUGAUCCAAAAUG		Glu GAA	Glu GAA	Phe UUU	Val GUG	Arg CGU	Gln CAA	Cys UGC	Phe UUC	Asn AAU	Pro CC				

† Deduced by complementarity from the 3′ terminal nucleotide sequence of genome RNA (table 2).

from the 3′ end of vRNA and is shown in table 3. These data do not include any primer sequences found *in vivo* at the 5′ end of mRNA and which are not complementary to vRNA (Bouloy *et al.* 1978; Krug *et al.*, this symposium). The first and only AUG protein initiation codon located in each mRNA species is underlined in table 3 and is located 20–30 residues from the 5′ end in all species except mRNA 5. Preliminary data (not shown) also locate the first AUG codon in mRNAs 1 and 2 in this region. In mRNA 5, the first AUG codon occurs at residues 46–48. It is of interest that the AUG codons in mRNAs 3, 4 and 6 are all located in the unique sequence 5′CAAAAUG-3′. A comparison of all known eukaryotic 5′ mRNA sequences indicates that in nearly all cases the AUG closest to the 5′ end is responsible for initiation and a model for ribosome binding and protein synthesis initiation incorporating this observation has been proposed (Kozak 1978). Thus although there is no direct evidence, by comparison with other eukaryotic mRNAs it is highly probable that these AUG codons identified in table 3 are utilized for protein synthesis initiation. The putative terminal amino acid sequences of the corresponding viral proteins can be deduced from the nucleotide sequences and are indicated in table 3. The predicted amino acid sequence for segments 4 and 6, which code for the haemagglutinin and neuraminidase glycoproteins respectively, contain a high proportion of hydrophobic amino acid residues. This is compatible with the proposed 'signal' sequence of amino acids which should be found in the extreme amino-terminus of membrane glycoproteins. Several termination codons can be identified (not shown), none of which are located in the same reading frame as the predicted amino acid sequence and thus do not contradict the proposed structures. More direct evidence either from complete nucleotide sequence analysis of the segments or from direct amino acid sequencing will be required to determine whether or not the mRNA species direct protein synthesis as shown in table 3.

This project was supported by a grant from the Medical Research Council, U.K.

References (Robertson)

Bouloy, M., Plotch, S. J. & Krug, R. M. 1978 Globin mRNAs are primers for the transcription of influenza viral RNA *in vitro*. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4886–4890.

Donis-Keller, H., Maxam, A. M. & Gilbert, W. 1977 Mapping adenines, guanines, and pyrimidines in RNA. *Nucl. Acids Res.* **4**, 2527–2538.

England, T. E. & Uhlenbeck, O. C. 1978 3′-Terminal labelling of RNA with T4 RNA ligase. *Nature, Lond.* **275**, 560–561.

Hay, A. J., Abraham, G., Skehel, J. J., Smith, J. C. & Fellner, P. 1977 Influenza virus messenger RNAs are incomplete transcripts of the genome RNAs. *Nucl. Acids Res.* **4**, 4197–4209.

Inglis, S. C., McGeoch, D. J. & Mahy, B. W. J. 1977 Polypeptides specified by the influenza virus genome. 2. Assignment of protein coding functions to individual genome segments by *in vitro* translation. *Virology* **78**, 522–536.

Kozak, M. 1978 How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15**, 1109–1123.

Lockard, R. E., Alzner-Deweerd, B., Heckman, J. E., MacGee, J., Tabor, M. W. & RajBhandary, U. L. 1978 Sequence analysis of 5′[^{32}P]labelled mRNA and tRNA using polyacrylamide gel electrophoresis. *Nucl. Acids Res.* **5**, 37–56.

McGeoch, D., Fellner, P. & Newton, C. 1976 Influenza virus genome consists of eight distinct RNA species. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3045–3049.

Simoncsits, A., Brownlee, G. G., Brown, R. S., Rubin, J. R. & Guilley, H. 1977 New rapid gel sequencing method for RNA. *Nature, Lond.* **269**, 833–836.

Skehel, J. J. & Hay, A. J. 1978 Nucleotide sequences at the 5′ termini of influenza virus RNAs and their transcripts. *Nucl. Acids Res.* **5**, 1207–1219.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES