Proc. Natl. Acad. Sci. USA
Vol. 73, No. 9, pp. 3045–3049, September 1976
Biochemistry

Influenza virus genome consists of eight distinct RNA species

(polyacrylamide gel electrophoresis/oligonucleotide mapping)

DUNCAN MCGEOCH*, PETER FELLNER†, AND CLIVE NEWTON†

*Division of Virology, Department of Pathology, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ; and †Searle Research Laboratories, Lane End Road, High Wycombe, Buckinghamshire, England

Communicated by F. Sanger, June 14, 1976

ABSTRACT **The genomic RNA of the avian influenza A virus, fowl plague, was fractionated into eight species by electrophoresis in polyacrylamide–agarose gels containing 6 M urea. The separated ^{32}P-labeled RNA species were characterized by digestion with RNase T_1 and fractionation of the resulting oligonucleotides by two-dimensional gel electrophoresis; this demonstrated that each species has a distinct nucleotide sequence. A tentative correlation of each genome RNA species with the virus protein that it encodes was made.**

A number of years ago, studies on the genetic and biological properties of influenza virus showed that the genetic material of the virus behaved as though it were composed of several individual units (1–3). More recently, physical studies of the genome RNA of influenza viruses, by sucrose density gradient centrifugation and by polyacrylamide gel electrophoresis, demonstrated that the RNA consists of a number of size classes with molecular weights between about 200,000 and 1,000,000, none of them large enough to represent the entire genetic material of the virus (4–8). It was also demonstrated by oligonucleotide mapping that three size classes of virus RNA contained distinct sequences, supporting the idea that fractions obtained by gel electrophoresis represented unique species rather than fragments of a larger RNA molecule (9). However, until now it has not been possible to separate the various RNA species completely, so that a complete description of influenza virus RNA in terms of the number of species, their relative representation, their molecular weights, and their nucleotide sequence relationships has not been possible. Despite this lack of resolution, the present, generally accepted view is that each RNA species contains the genetic information for one virus polypeptide (10). This independent nature of the virus genes is then held to account for the characteristic phenomena of high recombination frequencies in mixed infections (1, 3) and multiplicity reactivation of UV-irradiated virus (2) and is the basis for one theory concerning the origin of new, pandemic strains of influenza virus (11, 12).

We have found that a greatly improved fractionation of influenza virus RNA can be achieved by the use of polyacrylamide–agarose gels containing a high level of urea. In this paper we describe the separation of the virus RNA into eight species by this method. We also report our subsequent characterization of each RNA species by oligonucleotide mapping using two-dimensional polyacrylamide gel electrophoresis. We conclude that the influenza virus genome is composed of eight RNA species containing distinct nucleotide sequences, and we attempt to correlate the genome RNA species with the virus proteins that they specify.

METHODS

Polyacrylamide–Agarose Gel Electrophoresis of RNA. Gels were prepared, essentially as described by Floyd *et al.* (13), in slabs 1.5 mm thick by 14 cm wide, and contained polyacrylamide (2.0–2.8%), 0.6% agarose, 6 M urea, 36 mM Tris–phosphate at pH 7.8, 1 mM EDTA. RNA samples for electrophoresis were dissolved in 36 mM Tris–phosphate at pH 7.8, 6 M urea, 1 mM EDTA, 10% (vol/vol) glycerol, 0.2% sodium dodecyl sulfate, 0.1% bromophenol blue, and were heated at 80° for 30 sec. Electrophoresis was at room temperature, with 36 mM Tris–phosphate at pH 7.8, 1 mM EDTA, 0.2% sodium dodecyl sulfate in the electrode tanks. Other conditions are specified for individual experiments.

Oligonucleotide Mapping of RNA. ^{32}P-Labeled fowl plague virus RNA (14) containing 2×10^7 cpm was fractionated at 180 V for 25 hr in a 32 cm long gel containing 2.2% acrylamide, 0.6% agarose. The RNA bands were then located by autoradiography of the cellophane-covered wet gel for 20 min. The bands were excised and the RNA was extracted essentially by the method of Jeppesen *et al.* (15). Each RNA was recovered finally in 50 μl 10 mM Tris·HCl at pH 7.5, 1 mM EDTA containing 80 μg of carrier RNA, and digested with 4 μg of RNase T_1 for 30 min at 37°. The digests were then fractionated by two-dimensional gel electrophoresis (16, 17)—first dimension 10% polyacrylamide, 6 M urea, 25 mM citric acid, pH 3.5; second dimension 20% polyacrylamide, 0.1 M Tris–borate at pH 8.3, 2.5 mM EDTA. The gels were autoradiographed for 4–8 days. The separation of oligonucleotides in the first dimension, at pH 3.5, depends mainly on their charge, i.e., base composition. The fractionation in the second dimension, at pH 8.3, is by chain length. Sequence complexities for RNA species were estimated using the fingerprint gels as follows. Several large oligonucleotides whose chain lengths were known from gel mobility and/or from further analysis of the RNase A digestion products were quantitatively recovered from the gel by electrophoresis onto a DEAE-paper disc (17) and their ^{32}P contents were measured to give ^{32}P cpm for a mononucleotide unit. The total ^{32}P content of the gel was obtained by drying the gel and scintillation counting. This latter value divided by the ^{32}P content per mononucleotide gave an estimate of the chain length.

RESULTS

Separation of virus RNA into its constituent species

All the detailed work described here used the avian influenza A virus, fowl plague [A/chicken/Rostock/34(HAV1N1)]. The virus was propagated in primary cultures of chick cells, in the presence of inorganic [^{32}P]phosphate. Virus was purified and the ^{32}P-labeled RNA was extracted (14), then fractionated by electrophoresis on slab gels of polyacrylamide–agarose containing 6 M urea (13). The RNA was separated into eight species by this procedure, as shown in Fig. 1. Band 1, and sometimes Band 4, tended to be more diffuse than the others. The same pattern was obtained with RNA from virus grown in fertile hen eggs, and with RNA from newly cloned virus. No discrete species of RNA moving faster than Band 8 was detectable. No further fractionation of the eight bands was ever observed using

gels containing 2.0–2.8% polyacrylamide and extended conditions of electrophoresis (6 V/cm for 36 hr).

When RNAs extracted from the human influenza A virus strains A/WS/33(HON1) and A/Bel/43(HON1) and from influenza B virus strain B/Lee/40, were fractionated in this way, similar separation into eight bands was observed (data not shown). We therefore consider that our findings represent a general feature of the RNA of influenza viruses.

We estimated the chain lengths of the RNA species by comparison of their mobilities in gels with the mobilities of chick and *Escherichia coli* ribosomal RNAs. Fig. 2 illustrates the results obtained with a 2.5% polyacrylamide gel. A straight-line relation between mobility of the rRNA species and logarithm of the chain length was obtained for chain length less than 3000 nucleotides. This relation broke down for higher chain lengths. The deviation was more severe for gels of lower acrylamide concentration. The chain lengths and molecular weights estimated from Fig. 2 for each virus RNA species are listed in Table 1. These fall within the range of molecular weights previously estimated for influenza virus RNAs. The proportion of the total radioactivity in each RNA band was estimated by cutting the dried gel into strips, which were subjected to scintillation counting. Taken together with the molecular weight estimates, this allows calculation of the representation, on a molar basis, of each species (Table 1). When the molar abundances are expressed relative to a value of unity for Band 5, we estimate those of the other bands to vary from 0.7 for the largest species to 1.3 for Band 7.

Oligonucleotide mapping of the RNA species

We examined the eight RNA species by fingerprinting to ascertain whether they contained related or independent sequences, and whether each species was homogeneous. The fingerprints of the RNA species are presented in Fig. 3. By two criteria each RNA species consists of a distinct sequence. First, each RNA gave a characteristic and unique pattern of oligonucleotides; the complexity decreased from Band 1 to Band 8. Second, from each fingerprint, at least four of the larger oligonucleotides were recovered and then digested with RNase A and fractionated by electrophoresis on DEAE-paper at pH 3.5. The products arising from RNase A digestion demonstrated clearly that none of the large oligonucleotides that we examined are common to two or more RNA bands. We conclude that each band contains a distinct set of RNA sequences. None of the smaller bands could be derived by breakdown from any of the larger species.

We estimated the sequence complexity for each species, with

Table 1. RNA species of influenza virus

RNA species	Chain length from gel mobility	Chain length from fingerprint	^{32}P Content, % of total	Relative molar abundance
1	3500	5100	16.2	0.7
2	3000	3670	13.8	0.7
3	2950	4230	13.8	0.7
4	2450	2490	13.7	0.8
5	2000	2370	13.5	1.00
6	1720	2400	12.5	1.1
7	1080	930	9.6	1.3
8	870	820	6.8	1.2

The estimated chain length of each RNA is listed, as obtained by mobility in gel electrophoresis (Fig. 2) and from analysis of oligonucleotides. For RNAs 1–6, the two-dimensional oligonucleotide gels were used as detailed in *Methods*. Chain lengths for RNAs 7 and 8 were obtained from extensive analysis of two-dimensional oligonucleotide maps on DEAE-paper (unpublished work). The ^{32}P content of each RNA in a dried gel was estimated by scintillation counting, allowing for background ^{32}P between bands, and is expressed as a percentage of the total ^{32}P in the RNA bands. Using the relative ^{32}P contents and the chain lengths estimated by mobility in gel electrophoresis, the molar representation of each band was calculated, relative to a value of unity for Band 5 (underlined).

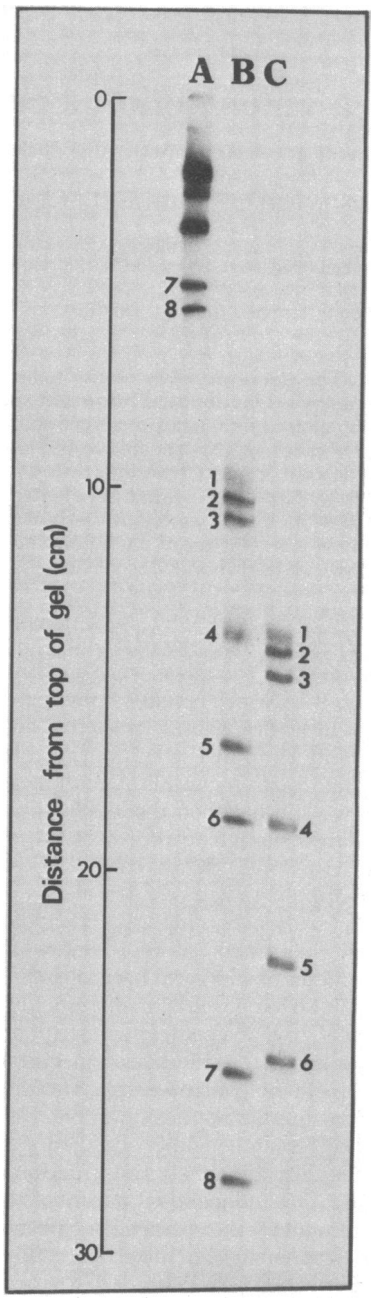

FIG. 1. Polyacrylamide–agarose gel electrophoresis of influenza virus RNA. ^{32}P-Labeled fowl plague virus RNA was prepared as described previously (14). Samples (10 μl) of RNA were loaded into 1 cm slots in a 32 cm long slab gel containing 2.2% acrylamide, 0.6% agarose, and fractionated by electrophoresis at 180 V. (A) Sample run for 6 hr. (B) Sample run for 25 hr. (C) Sample run for 32 hr. After completion of the run, the gel was washed in water (10 min), dried onto Whatman 3 MM paper, and autoradiographed (3 hr). The RNA bands are numbered as in the *text*.

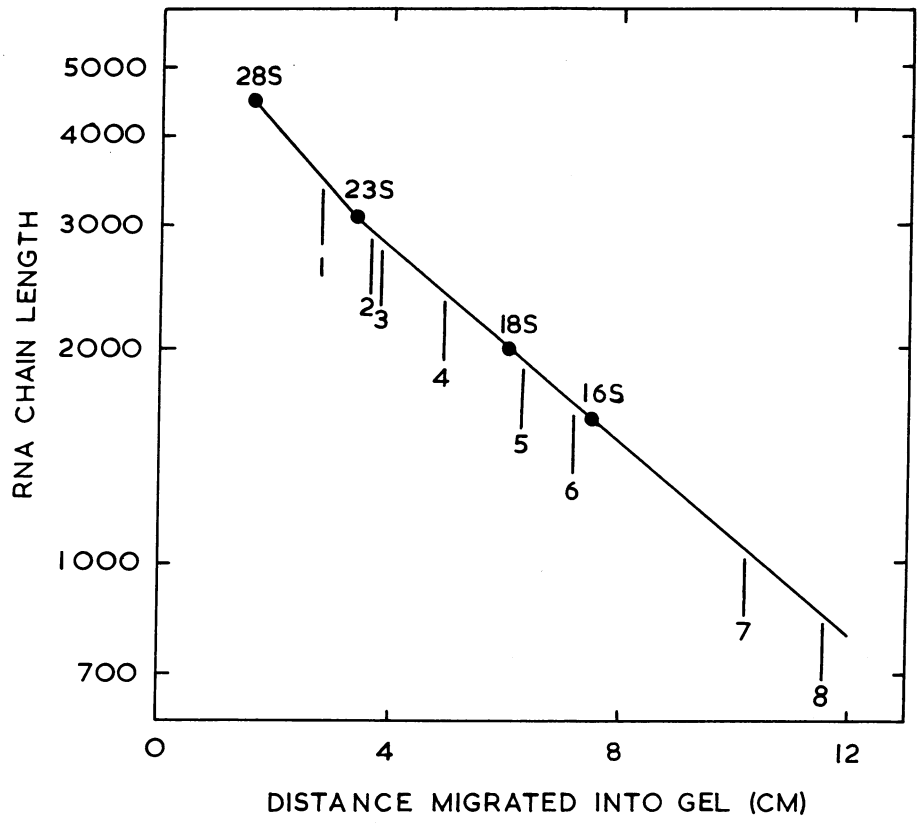

FIG. 2. Estimation of the chain lengths (number of nucleotides) of influenza virus RNAs. ^{32}P-Labeled fowl plague virus RNA was fractionated by electrophoresis for 12 hr at 90 V in a 14-cm-long gel of 2.5% acrylamide, 0.6% agarose. ^{32}P-Labeled chick cell and *E. coli* ribosomal RNAs were run in adjacent slots. The distance migrated by each rRNA species was measured and plotted against the chain length on a logarithmic scale to give the relation shown (18–20). Chain lengths of the influenza virus RNAs were then estimated from the intercepts of the distances migrated with the calibration line, as shown.

the results listed in Table 1. Since we were unable to separate the three largest RNA species completely on preparative gel runs, and since the estimates depend on uniform recoveries of oligonucleotides from the fingerprint gel, we regard the chain lengths obtained by this method as approximate values only. For RNAs 2, 4, 5, 7, and 8 the chain lengths estimated by this method agree well with those obtained by mobility of the RNAs, demonstrating that each RNA band consists of one, homogeneous RNA species. For RNAs 1, 3, and 6 the agreement is less satisfactory. However, since the molarities of the latter species were close to those of their neighbors, and considering the limitations of the complexity estimates, we conclude that these RNAs also consist each of one, largely homogeneous species. Thus, the influenza virus genome consists of eight distinct RNA molecules. The tendency to submolar levels for the large RNAs could result from greater loss during purification, for example by nuclease action, or it could be a real phenomenon in the bulk preparation of virus.

Several other features of the RNA species are apparent from the fingerprints. None of the bands contain a large poly(C) tract, which is a characteristic feature of fingerprints of many picornavirus RNAs (17). The fingerprints also demonstrate directly that influenza virus RNAs do not contain poly(A) sequences; this agrees with previous results obtained by indirect methods (21). [Poly(A) and poly(C) would have run in characteristic positions in the upper part of the gel (17).]

It seems reasonable that all the RNA species might have in common nucleotide sequences involved in regulation of transcription and other aspects of virus replication. We have not detected any such common extended sequences by the fingerprinting technique, but our results in no way exclude their existence. The sensitivity of the fingerprinting method to single base changes makes it unsuitable for detection of sequences that may be related but not identical.

DISCUSSION

The sum of the chain lengths of the eight RNA species is about 17,400 bases, that is, RNA of total molecular weight about 5.7 $\times 10^6$. This is somewhat higher than previous estimates of the size of the influenza virus genome (5–8), a result of the better resolution of RNA species obtained, especially of the highest molecular weight bands. In extensive studies with temperature-sensitive mutants of influenza virus, Hirst (22) was able to identify eight recombination groups. Taken in conjunction with our present results this suggests that mutants have been obtained in every species of the genome RNA and that every RNA specifies a polypeptide essential to the production of infectious virus.

We have attempted to determine which RNA species is the gene for each of the known virus-specified proteins. In doing this we have assumed that each RNA molecule encodes one protein only. This is a reasonable assumption from the available genetic data (22), but is not rigorously proven. We estimated for each RNA the maximum polypeptide coding capacity and compared these data with the catalogue of known virus specified polypeptides, as shown in Table 2. The virion proteins recognized are: three high-molecular-weight proteins P_1, P_2, and P_3 (23), the hemagglutinin, the nucleocapsid protein, the neuraminidase, and the matrix protein (10, 24). In addition, two

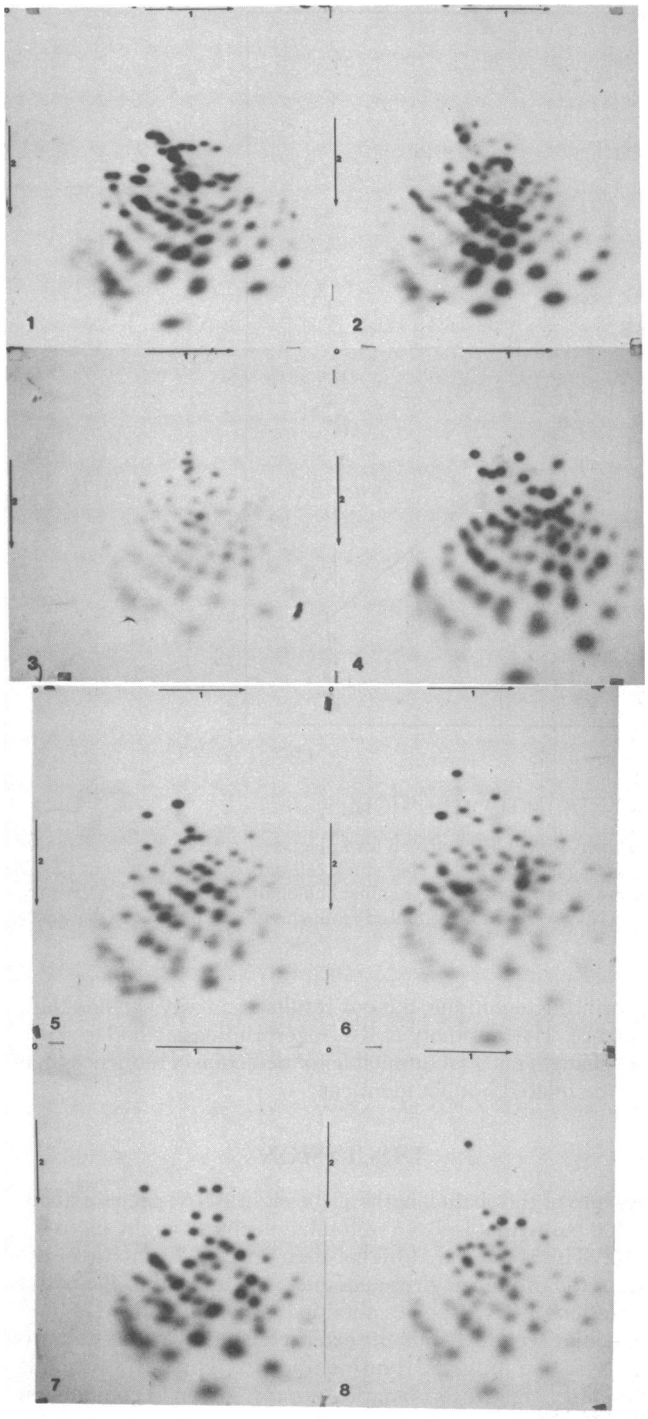

FIG. 3. Oligonucleotide fingerprinting of influenza virus RNAs. ^{32}P RNA species 1–8 of fowl plague virus were digested with T_1 RNase and fractionated by two-dimensional gel electrophoresis as described in *Methods*. First dimension, left to right; second dimension, top to bottom.

proteins, apparently virus coded, are found in infected cells but not in virus particles. These are the nonstructural proteins NS_1 and NS_2 (24). Table 2 shows that RNAs 1, 2, and 3 are the only RNAs large enough to code for P_1, P_2, and P_3. Considering the remainder of the virus proteins, we then suggest that Band 4 codes for hemagglutinin, Band 5 for the nucleocapsid protein, Band 6 for neuraminidase, Band 7 for the matrix protein and Band 8 for the large nonstructural protein NS_1. This assignment leaves no separate RNA species to code for the smaller non-

Table 2. RNAs and proteins of influenza virus

Genome RNAs			Virus proteins	
RNA	RNA chain length	Maximum polypeptide chain length	Polypeptide chain length	Protein
1	3500	1170	960	P_1
2	3000	1000	870	P_2
3	2950	980	850	P_3
4	2450	820	750	Hemagglutinin HA
5	2000	670	530	Nucleocapsid NP
6	1720	570	450	Neuraminidase NA
7	930	310	250	Matrix M
8	820	270	230	Non-structural NS_1
			110	Non-structural NS_2

For each RNA species the nucleotide chain length is shown, together with the maximum size of polypeptide, as amino acid chain length, which could be specified by each RNA. These data are compared with the estimated amino acid chain lengths of the known virus-specified proteins (10, 23, 24). The RNA chain length estimates are from Table 1 and are based on electrophoretic mobilities, except for RNAs 7 and 8, where the values from oligonucleotide mapping are used.

structural protein NS_2. There are several possible explanations of this apparent discrepancy: NS_2 may not in fact be virus coded, or it may be encoded by the same RNA species as one of the other proteins, and then arise either from separate translation or by proteolytic cleavage of another protein.

We have obtained evidence to support these gene–polypeptide assignments, which is summarized here and will be presented in detail elsewhere. Influenza virus is considered to be a negative strand virus, that is, the virion RNA is complementary to the RNA with messenger activity (14, 21, 25). We have used the separated virion RNA species to determine by RNA·RNA hybridization the levels of each of the complementary RNA species in appropriate extracts of virus-infected cells. We have found that, with RNA from membrane-bound polyribosomes, the highest levels of hybridization are to RNA of Band 4, followed by Band 6. With RNA from free polyribosomes, the highest levels of hybridization are to RNAs 5 and 7. These results are wholly in accord with our assignments: the glycoproteins hemagglutinin and neuraminidase (encoded by RNAs 4 and 6, respectively) are known to be synthesized by membrane-bound polysomes (26–28), while the nucleocapsid protein (Band 5) is a major species synthesized on free polysomes (28). The matrix protein (Band 7) is also a major protein in infected cells, but its site of synthesis is obscure. In addition, in all polysomal RNA preparations, RNA complementary to Bands 1–3 is present at a low level, consistent with the low relative rates of synthesis of P_1, P_2, and P_3 (23).

The ability to isolate the individual RNA species of influenza virus opens up several possible fields of study. First, it should provide new approaches to the problems of the molecular basis of antigenic changes in human pandemic influenza virus strains. Second, since the RNA species of influenza virus probably correspond to individual genes, with their various regulatory sequences, they present an unusual opportunity for study of animal virus gene structure and regulation. Finally, as outlined above, the separated RNA species can be used to examine the process of virus infection by measuring individual virus mRNA species from infected cells.

We thank R. D. Barry, N. H. Carey, A. R. Carroll, and B. W. J. Mahy for discussion and criticism. This work was supported in part by Medical Research Council Grants. P.F. and C.N. thank A. J. Hale for provision of research facilities.

1. Burnet, F. M. (1956) *Science* **123**, 1101–1104.
2. Barry, R. D. (1961) *Virology* **14**, 398–405.
3. Hirst, G. K. (1962) *Cold Spring Harbor Symp. Quant. Biol.* **27**, 303–309.
4. Davies, P. & Barry, R. D. (1966) *Nature* **211**, 384–387.
5. Duesberg, P. H. (1968) *Proc. Natl. Acad. Sci. USA* **59**, 930–937.
6. Pons, M. W. & Hirst, G. K. (1968) *Virology* **34**, 385–388.
7. Skehel, J. J. (1971) *J. Gen. Virol.* **11**, 103–109.
8. Bishop, D. H. L., Obijeski, J. F. & Simpson, R. W. (1971) *J. Virol.* **8**, 74–80.
9. Horst, J., Content, J., Mandeles, S., Fraenkel-Conrat, H. & Duesberg, P. (1972) *J. Mol. Biol.* **69**, 209–216.
10. White, D. O. (1974) *Curr. Top. Microbiol. Immunol.* **63**, 1–48.
11. Laver, W. G. & Webster, R. G. (1973) *Virology* **51**, 383–391.
12. Skehel, J. J. (1974) in *Evolution in the Microbial World*, eds. Carlile, M. J. & Skehel, J. J. (Cambridge University Press, London), pp. 321–342.
13. Floyd, R. W., Stone, M. P. & Joklik, W. K. (1974) *Anal. Biochem.* **59**, 599–609.
14. Glass, S. E., McGeoch, D. & Barry, R. D. (1975) *J. Virol.* **16**, 1435–1443.
15. Jeppesen, P. G. N., Barrell, B. G., Sanger, F. & Coulson, A. R. (1972) *Biochem. J.* **128**, 993–1006.
16. DeWachter, R. & Fiers, W. (1972) *Anal. Biochem.* **49**, 184–197.
17. Frisby, D. P., Newton, C., Carey, N. H., Fellner, P., Newman, J. F. E., Harris, T. J. R. & Brown, F. (1976) *Virology*, in press.
18. Peacock, A. C. & Dingman, C. W. (1968) *Biochemistry* **7**, 668–674.
19. Loening, U. E. (1968) *J. Mol. Biol.* **38**, 355–365.
20. Stanley, W. M. & Bock, R. M. (1965) *Biochemistry* **4**, 1302–1311.
21. Etkind, P. R. & Krug, R. M. (1974) *Virology* **62**, 38–45.
22. Hirst, G. K. (1973) *Virology* **55**, 81–93.
23. Inglis, S. C., Carroll, A. R., Lamb, R. A. & Mahy, B. W. J. (1976) *Virology*, in press.
24. Skehel, J. J. (1972) *Virology* **49**, 23–36.
25. Mahy, B. W. J. & Barry, R. D., eds. (1975) *Negative Strand Viruses* (Academic Press, London).
26. Compans, R. W. (1973) *Virology* **51**, 56–70.
27. Klenk, H. D., Wollert, W., Rott, R. & Scholtissek, C. (1973) *Virology* **57**, 28–41.
28. Hay, A. J. (1974) *Virology* **60**, 398–418.