*Proc. Natl. Acad. Sci. USA*
Vol. 73, No. 6, pp. 2142–2146, June 1976
Microbiology

# Mapping of the influenza virus genome: Identification of the hemagglutinin and the neuraminidase genes

(RNA/recombinant viruses/polyacrylamide gel electrophoresis/genetics)

PETER PALESE AND JEROME L. SCHULMAN

Department of Microbiology, Mount Sinai School of Medicine of the City University of New York, 100 St. and 5th Avenue, New York, N.Y. 10029

*Communicated by Robert M. Chanock, March 16, 1976*

**ABSTRACT** **Polyacrylamide gel electrophoresis of the RNA of influenza A/PR/8/34 (H0N1) and A/Hong Kong/8/68 (H3N2) viruses and recombinant viruses derived from them revealed that each contains eight RNA segments, the fourth of which codes for hemagglutinin. (The largest RNA of the segmented genome is counted as band 1.) The neuraminidase gene was identified as the sixth segment in the RNA pattern of influenza A/PR8 virus and as the fifth segment of A/Hong Kong virus. The molecular weights of the RNAs for the hemagglutinin and the neuraminidase genes lie in the range of 600,000–700,000.**

Analysis of RNA of influenza A viruses has revealed that the segmented genome contains at least seven distinct RNA pieces (review, ref. 1; refs. 2–6). These observations are in accord with the results of genetic experiments involving studies of recombinational frequencies among different strains of virus and characterization of temperature-sensitive mutants, which have indicated that transcripts of each of the virion RNA segments probably function as monocistronic messages coding for a single virus polypeptide (review, ref. 7).

The sizes of the different virion RNA segments range from a molecular weight of approximately $1 \times 10^6$ to 2 to $4 \times 10^5$ (5, 6). Presumably, complementary RNA (cRNA) transcripts of each are capable of coding for corresponding proteins with molecular weights ranging from $1 \times 10^5$ to approximately $2 \times 10^4$, a range which corresponds with that of isolated virus proteins (review, ref. 8; refs. 6 and 9). Although correlations in size have been used as a basis for speculation, it has not been possible until now to identify specifically which RNA segment is the gene for any specific virus protein.

In this communication we describe experiments that permit a preliminary mapping of the genomes of two different influenza A viruses. The method employed is based on our previous observations that the RNAs of several different strains have different patterns of migration on polyacrylamide gels and that analysis of the RNA pattern of a recombinant strain allows the identification of specific RNA segments which were derived from one or the other parent virus (6).

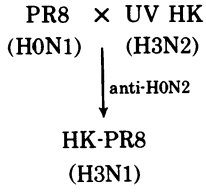
Unpublished results obtained with gels run at different temperatures or with different acrylamide concentrations indicated that the relative migration rates of various RNA bands were altered. Consequently, at present it is not clear whether differences in migration rates of equivalent RNA segments from different viruses are due to differences in molecular weight and/or to differences in secondary structure. We have improved this technique in the present experiments and, by analysis of a series of influenza A recombinant viruses, have succeeded in identifying the hemagglutinin and neuraminidase genes.

Abbreviations: PR8 virus, influenza A/PR/8/34 (H0N1) virus; HK virus, influenza A/Hong Kong/8/68 (H3N2) virus; HK-PR8 virus, influenza A/Hong Kong/8/68 (H3)–A/PR/8/34 (N1) virus; PR8-HK virus, influenza A/PR/8/34 (H0)–A/Hong Kong/8/68 (N2) virus; HK-HK virus, influenza A/Hong Kong/8/68 (H3)–A/Hong Kong/8/68 (N2) virus; PR8-PR8 virus, influenza A/PR/8/34 (H0)–A/PR/8/34 (N1) virus; Aichi-PR8 virus, influenza A/Aichi/68 (H3)–A/PR/8/34 (N1) virus.

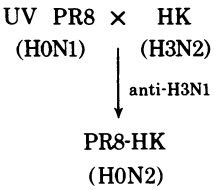
## MATERIALS AND METHODS

**Viruses.** The two parent viruses employed in these experiments were influenza A/PR/8/34 (H0N1) and A/Hong Kong/8/68 (H3N2) viruses. In the following, we will refer to these viruses as the PR8 and the HK viruses, respectively. One of these, PR8 virus, had previously been studied in the same system (5, 6). Both viruses were cloned by plaque-to-plaque passage in MDCK (canine kidney) cells (10) and then were passaged in 11-day-old embryonated eggs to prepare allantoic fluid seed viruses.

Recombinant viruses were derived from these two parents as follows: One or the other parent was diluted 10- to 100-fold and subjected to UV irradiation (General Electric, G.8.T5 lamp, 8 W) for 30 sec at a distance of 15 cm. Embryonated eggs were inoculated with mixtures containing equivalent concentrations of the two viruses and, after 24 hr, fluids were harvested, diluted serially 10-fold, mixed with an equal volume of a 1:10 dilution of the appropriate antiserum, and passaged in eggs. As illustrated schematically below, to select for an H3N1 recombinant, influenza A/Hong Kong/8/68 (H3)–A/PR/8/34 (N1) virus (HK-PR8 virus), eggs were inoculated with PR8 (HON1) virus and HK (H3N2) virus which had been irradiated with UV light and the yield was passaged in the presence of rabbit antiserum to H0N2 virus.




To select for the reciprocal recombinant influenza A/PR/8/34 (H0)–A/Hong Kong/8/68 (N2) virus, the PR8 parent was exposed to UV irradiation, and antiserum to H3N1 virus was employed as the selective antibody screen. This H0N2 recombinant is referred to as PR8-HK virus.




Back-recombinants, influenza A/Hong Kong/8/68 (H3)–A/Hong Kong/8/68 (N2) virus (HK-HK virus) and influenza A/PR/8/34 (H0)–A/PR/8/34 (N1) virus (PR8-PR8 virus),

antigenically identical to the two original H3N2 and H0N1 wild-type viruses, were isolated in a similar way employing HK-PR8 (H3N1) and PR8-HK (H0N2) recombinant viruses as parents. The back-recombinants, HK-HK and PR8-PR8, were selected using antisera against PR8 and HK viruses, respectively.

HK-PR8 × UV PR8-HK
(H3N1) (H0N2)
↓ anti-H0N1
HK-HK
(H3N2)
UV HK-PR8 × PR8-HK
(H3N1) (H0N2)
↓ anti-H3N2
PR8-PR8
(H0N1)

All newly isolated recombinant strains were cloned by plaque-to-plaque passage in MDCK cells in the presence of the appropriate antiserum screen and then inoculated into embryonated eggs to prepare seed virus. Serologic confirmation that each was a pure population containing the appropriate hemagglutinin and neuraminidase was obtained by methods described previously (11–13).

An additional recombinant, influenza A/Aichi/68 (H3)-A/PR/8/34 (N1) virus, used in this report was previously isolated for other purposes. It was derived in embryonated eggs following mixed infection of clone 1-5C-4 cells (14) with influenza PR8 (H0N1) and influenza A/Aichi/68 (H3N2) virus. The Aichi-PR8 (H3N1) recombinant was selected in the presence of antiserum to H0N2 virus.

**RNA.** Monolayers of MDCK cells ($4 \times 10^6$ cells) were infected with different influenza viruses (2 to 3 dishes each) in the presence of approximately 15 mCi of $^{32}P$ (as orthophosphate, sterile in $H_2O$, New England Nuclear) at a multiplicity of 5–10 plaque-forming units/cell. After 18 hr virus was purified as described previously (6). All viruses examined replicated under these conditions, yielding hemagglutination titers that varied between 256 and 1024. RNA was extracted from purified virus (6). From three infected monolayers using 50 mCi of $^{32}P$, a total of 50,000–500,000 cpm of [$^{32}P$]RNA was obtained. No specific activity of the RNA was determined because of the small quantities of radioactive virus available. The RNAs were then analyzed on urea–polyacrylamide gels as originally described by Floyd *et al.* (15) and modified by us (6). In the present experiments, further modifications included the use of longer gels (23 cm) and, in some experiments, the use of other concentrations of acrylamide (2.6%) and *N,N'*-methylenebisacrylamide (0.15%). All gels were run for at least 16 hr at 120 V. Under these conditions, estimations of the molecular weights of segments 7 and 8 of PR8 virus varied slightly from those previously reported (5, 6). Band 7 was estimated to have a molecular weight of $0.42 \times 10^6$ and band 8 a molecular weight of $0.31 \times 10^6$, as compared to previously reported values of $0.47 \times 10^6$ and $0.39 \times 10^6$ (5, 6), respectively.

## RESULTS

The RNA patterns of PR8 (H0N1) and HK (H3N2) viruses and those of the two antigenic hybrid recombinants (H3N1 and H0N2 viruses) are compared in Fig. 1. In the case of PR8 virus, the RNA separates into eight distinct segments: a cluster of three

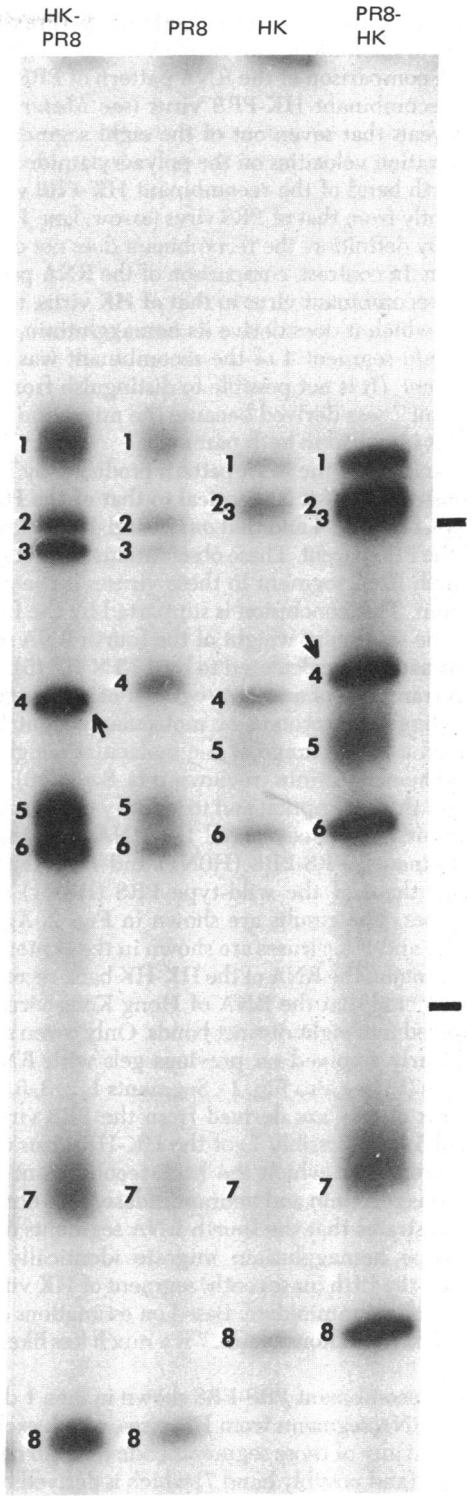



FIG. 1. Analysis of [$^{32}P$]RNAs of influenza HK (H3N2), influenza PR8 (H0N1), and their reciprocal recombinant viruses, HK-PR8 (H3N1) and PR8-HK (H0N2). Lane 1 (from left): influenza HK-PR8 (H3N1) virus; arrow indicates the HK (H3) hemagglutinin. Lane 2: influenza PR8 (H0N1) virus. Lane 3: influenza HK (H3N2) virus; bands 2 and 3 do not separate well on this gel (compare Fig. 2, lane 3). Lane 4: influenza PR8-HK (H0N2) virus; arrow indicates the PR8 (H0) hemagglutinin. Migration is from top to bottom. Bars to the right indicate the position of 23S (molecular weight: $1.07 \times 10^6$) and 16S (molecular weight $0.55 \times 10^6$) *Escherichia coli* ribosomal RNA. Samples were run on a 2.8% polyacrylamide gel (23 cm long) containing 0.14% *N,N'*-methylenebisacrylamide.

slow-moving RNA bands, a second cluster of three faster-moving bands in the middle, and two fast-moving segments at the bottom. A comparison of the RNA pattern of PR8 virus to that of the recombinant HK-PR8 virus (see *Materials and Methods*) reveals that seven out of the eight segments have identical migration velocities on the polyacrylamide slab gel. Only the fourth band of the recombinant HK-PR8 virus migrates differently from that of PR8 virus (arrow, lane 1), a virus from which, by definition, the recombinant does not derive its hemagglutinin. In contrast, comparison of the RNA pattern of the HK-PR8 recombinant virus to that of HK virus, the other parent, from which it does derive its hemagglutinin, demonstrates that *only* segment 4 of the recombinant was derived from HK parent. (It is not possible to distinguish from which parent segment 7 was derived because the migration velocity of this band is identical in both parents.)

Similarly, analysis of the RNA pattern produced by PR8-HK virus demonstrates that it is identical to that of the HK virus parent *except* for band 4 and that *only* band 4 was clearly derived from the PR8 parent. These observations suggest strongly that the fourth RNA segment in these viruses is the gene for hemagglutinin. This conclusion is supported by the fact that, previously, the molecular weight of the fourth RNA segment of PR8 virus had been calculated to be $0.73 \times 10^6$ (5). A complete cRNA transcript of an RNA segment of this length is capable of coding for a protein of molecular weight 70,000–80,000, which is in the range of the molecular weight of the glycosylated hemagglutinin (reviews, refs. 8 and 16).

To confirm this assumption and to identify the genes coding for neuraminidase, we compared the RNA patterns of the back-recombinants, PR8-PR8 (H0N1) and HK-HK (H3N2) viruses, with those of the wild-type PR8 (H0N1) and HK (H3N2) viruses. The results are shown in Fig. 2. Again, the RNAs of PR8 and HK viruses are shown in the center. Lane 4 to the right contains the RNA of the HK-HK back-recombinant. It should be noted that the RNA of Hong Kong virus can be clearly resolved into eight distinct bands. Only seven segments could be clearly resolved on previous gels with RNA from H3N2 viruses (6) (see also Fig. 1). Segments 1, 2, 3, 6, and 8 of HK-HK virus clearly are derived from the PR8 virus. Only bands 4 and 5 (and possibly 7) of the HK-HK virus originate from HK virus, with which the back-recombinant shares a common hemagglutinin and neuraminidase. This observation again demonstrates that the fourth RNA segments of viruses with the same hemagglutinin migrate identically. Consequently, only the fifth (or seventh) segment of HK virus could be the gene for neuraminidase. Based on estimations of its size (0.42 to $0.47 \times 10^6$ daltons), band 7 is a much less likely candidate.

The back-recombinant PR8-PR8 shown in lane 1 derives its largest three RNA segments from HK virus, which excludes the possibility that any of those segments code for PR8 neuraminidase. Band 8 (and possibly band 7) which is derived from PR8 virus is too small to code for neuraminidase, leaving bands 5 and 6 as the only candidates for the neuraminidase gene. Band 5 may be excluded, because a band of identical velocity appears in the RNA pattern of the HK-HK back-recombinant virus, which derives its neuraminidase from HK virus and not from PR8 virus. This leaves band 6 as the most likely candidate for the PR8 neuraminidase gene. As previously described, band 6 of PR8 virus has an apparent molecular weight of $0.62 \times 10^6$ (6). The size of this RNA would therefore be sufficient to contain the information for the neuraminidase molecule (molecular weight, approximately 60,000) (17).

Support for the hypothesis that it is the sixth segment of PR8

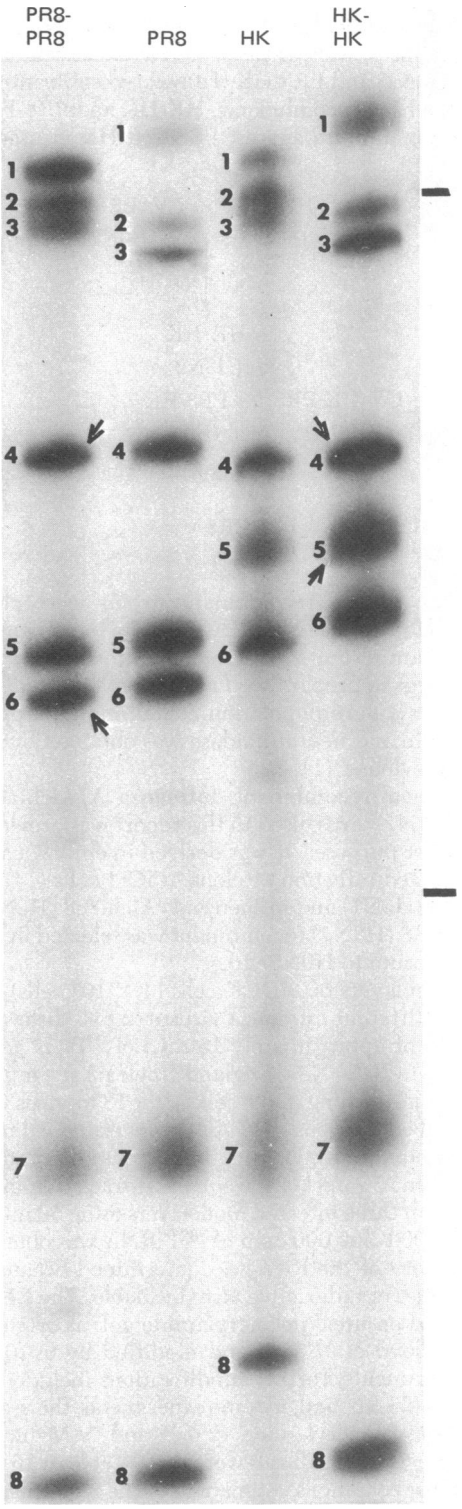



FIG. 2. Analysis of [$^{32}$P]RNAs of influenza A/HK (H3N2) and influenza A/PR/8/34 (H0N1) virus and of the back-recombinant viruses, HK-HK (H3N2) and PR8-PR8 (H0N1). Lane 1 (from left): influenza PR8-PR8 (H0N1) virus; arrows indicate the PR8 (H0) hemagglutinin and the PR8 (N1) neuraminidase (bands 4 and 6). Lane 2: influenza PR8 (H0N1) virus. Lane 3: influenza HK (H3N2) virus. Lane 4: influenza HK-HK (H3N2) virus; arrows indicate the HK (H3) hemagglutinin and the HK (N2) neuraminidase (bands 4 and 5, respectively). Migration is from top to bottom. Bars to the right indicate the position of 23S and 16S *E. coli* ribosomal RNA. Samples were run on a 2.6% polyacrylamide gel (23 cm long) containing 0.15% *N,N'*-methylenebisacrylamide. The origin at the top of the gel is not shown (compare with Fig. 1).

virus that codes for neuraminidase was obtained from RNA gels shown in Fig. 3, in which the RNA patterns of three viruses, all of which contain PR8 neuraminidase, were compared. The patterns of PR8 virus (lane 1) and the recombinant HK-PR8 (lane 3) are identical to those which were demonstrated in Fig. 1 (although different RNA preparations were employed). It is evident that the Aichi-PR8 recombinant does not contain a fifth segment in common with the other two viruses, although it too contains PR8 neuraminidase. In contrast, the sixth segment migrates identically in all three viruses, again suggesting that it is the sixth and not the fifth segment of PR8 virus that is the gene for neuraminidase. Although we have not examined the RNA pattern of the Aichi virus parent of the Aichi-PR8 recombinant, we assume that the fifth segment of the recombinant probably is derived from Aichi virus. It should be noted that the latter had not been cloned and that the Aichi-PR8 recombinant was isolated in clone 1-5C-4 cells, which could have resulted in the selection of a variant fifth RNA segment from the heterogenous Aichi virus population. In any case, the Aichi-PR8 recombinant virus was included in these studies merely to confirm the association of a common sixth RNA segment in viruses containing PR8 neuraminidase.

## DISCUSSION

Analysis of the RNAs of two strains of influenza A virus and of five recombinant strains demonstrated that each contains eight distinct RNA segments. The fortunate circumstance that seven of the eight segments from the two parent strains migrate differently in this gel system has provided a unique opportunity to analyze recombinant viruses to determine from which parent they derive their genes. As a result of such analysis, we propose that the fourth segment in both influenza A/Hong Kong/68 (H3N2) and A/PR/8/34 (H0N1) viruses is the gene for hemagglutinin. The fact that the fourth RNA segments of viruses which contain a common hemagglutinin *always* migrate at the same velocity, whereas the fourth segments of viruses with different hemagglutinins *never* migrate in the identical position, is unlikely to be due to chance. For similar reasons, we propose that the fifth segment of Hong Kong virus and the sixth of PR8 virus are the genes for the respective neuraminidases. It should be noted that, in each instance, the molecular weight of the RNA closely approximates that which would be required to code for the corresponding gene product, lending further support to the large body of genetic evidence indicating that transcripts of each virion RNA segment function as monocistronic messages (review, ref. 7).

We believe that extension of this approach will permit mapping of other genes and may make it possible to correlate biological properties of specific strains with particular genes or groups of genes. The latter could be accomplished by comparisons of antigenically identical recombinant viruses that contain different combinations of genes derived from the two parent viruses. For example, the PR8-PR8 back-recombinant derives its three largest genes from HK virus. Another antigenically identical back-recombinant has been identified in which only two of these three genes are of HK virus origin (not shown). Analysis of the biological characteristics of similar recombinants (obtained from viruses that have not been exposed to UV light irradiation) may ultimately permit the identification of genes that are associated with strain-related differences in growth rate, host range, plaque morphology, and virulence for man and animals. If this speculation is correct, this approach could add a new dimension to the understanding of the genetics of influenza viruses.

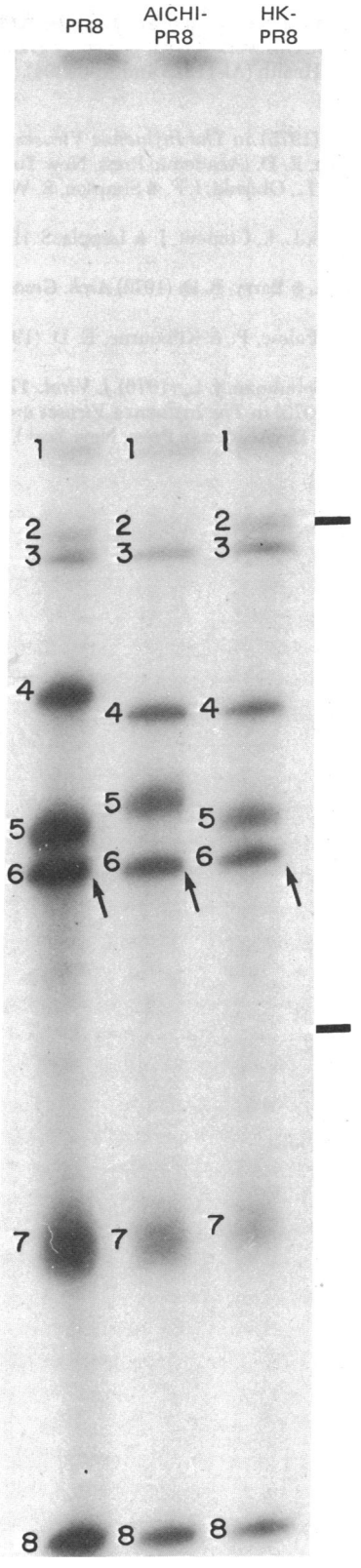



FIG. 3. Analysis of [$^{32}$P]RNAs of influenza PR8 (H0N1) virus and two recombinant viruses, Aichi-PR8 (H3N1) and HK-PR8 (H3N1). Lane 1 (from left): influenza PR8 (H0N1) virus. Lane 2: influenza Aichi-PR8 (H3N1) virus. Lane 3: influenza HK-PR8 (H3N1) virus. Arrows indicate the PR8 (N1) neuraminidase common to all three viruses. Migration is from top to bottom. Bars to the right indicate the position of 23S and 16S *E. coli* ribosomal RNA. Electrophoresis conditions are as in Fig. 1.

We thank Ms. Sidna Rachid and Ms. Kaye Leitzinger for expert technical assistance. This work was supported by grants from the National Institutes of Health (Al-11823 and Al-09394).

1. Pons, M. W. (1975) in *The Influenza Viruses and Influenza*, ed. Kilbourne, E. D. (Academic Press, New York), pp. 145–170.
2. Bishop, D. H. L., Obijeski, J. F. & Simpson, R. W. (1971) *J. Virol.* **8**, 74–80.
3. Lewandowski, L. J., Content, J. & Leppla, S. H. (1971) *J. Virol.* **8**, 701–707.
4. Bromley, P. A. & Barry, R. D. (1973) *Arch. Gesamte Virusforsch.* **42**, 182–196.
5. Ritchey, M., Palese, P. & Kilbourne, E. D. (1976) *J. Virol.* **18**, 738–744.
6. Palese, P. & Schulman, J. L, (1976) *J. Virol.* **17**, 876–884.
7. Sugiura, A. (1975) in *The Influenza Viruses and Influenza*, ed. Kilbourne, E. D. (Academic Press, New York), pp. 171–213.
8. Compans, R. W. & Choppin, P. W. (1975) "Reproduction of large RNA viruses," in *Comprehensive Virology*, eds. Fraenkel-Conrat, H. & Wagner, R. R. (Plenum Press, New York), Vol. 4, pp. 179–252.
9. Skehel, J. J. (1971) *J. Gen. Virol.* **11**, 103–109.
10. Krug, R. M. (1971) *Virology* **44**, 125–136.
11. Wong, S. C. & Kilbourne, E. D. (1961) *J. Exp. Med.* **113**, 95–110.
12. Aminoff, D. (1961) *Biochem. J.* **81**, 384–392.
13. Palese, P. & Schulman, J. L. (1974) *Virology* **57**, 227–237.
14. Kilbourne, E. D. (1969) in *Fundamental Techniques in Virology*, eds. Habel, K. & Salzman, N. (Academic Press, New York), pp. 146–160.
15. Floyd, R. W., Stone, M. P. & Joklik, W. K. (1974) *Anal. Biochem.* **59**, 599–609.
16. Schulze, I. (1975) in *The Influenza Viruses and Influenza*, ed. Kilbourne, E. D. (Academic Press, New York), pp. 53–83.
17. Wrigley, N. G., Skehel, J. J., Charlwood, P. A. & Brand, C. M. (1973) *Virology* **51**, 525–529.