Proc. Nat. Acad. Sci. USA
Vol. 73, No. 1, pp. 238–242, January 1976
Microbiology

Chromatographic isolation of the hemagglutinin polypeptides from influenza virus vaccine and determination of their amino-terminal sequences

(sodium dodecyl sulfate gel filtration/hydrophobic proteins/disulfide linkage/primary structure/automatic Edman degradation)

D. J. BUCHER, S. S.-L. LI, J. M. KEHOE, AND E. D. KILBOURNE

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

Communicated by Igor Tamm, October 16, 1975

ABSTRACT **The influenza virus hemagglutinin polypeptides, HA_1 and HA_2, have been purified by gel filtration in the presence of sodium dodecyl sulfate from a vaccine preparation of the recombinant strain Heq1N2. Use of this technique for purification of the hemagglutinin polypeptides eliminated the need for proteolytic agents for removal of the hemagglutinin from the virus particles and 100–300 mg of virus yielded 10–30 mg of viral protein per chromatographic cycle. Because proteolysis is not required to remove the spikes from the viral envelope, the envelope-embedded HA_2 polypeptide was purified in its entirety for structural analysis.**

Amino-terminal sequence analysis of the smaller polypeptide, HA_2, revealed a cyclic repetition of glycyl residues through the first 24 residues at every third to fourth position. The sequence through the first 10 residues was identical to that presented by Skehel and Waterfield for other type A influenza viruses [(1975) *Proc. Nat. Acad. Sci. USA* 72, 93–97]. The HA_1 (Heq1) polypeptide, on the other hand, had different amino acids at three or four out of the first 10 residues of the amino-terminal sequence when compared to HA_1 from H0, H1, or H2 subtypes (Skehel and Waterfield). The present study has demonstrated the feasibility of the use of vaccine virus as a source of large quantities of viral protein for determination of primary structure.

The major structural protein of the envelope of influenza virus is the hemagglutinin, a glycoprotein which interacts with the neuraminic acid of cellular receptors. The hemagglutinin is synthesized as a glycoprotein of approximately 75,000 daltons. Proteolytic cleavage of the hemagglutinin by host proteases in the chick embryo *in ovo* is asymmetric and results in the production of two subunits which are held together by disulfide linkage (1). These subunits are HA_1 (50,000 daltons) and HA_2 (25,000 daltons). This prior cleavage of the hemagglutinin of virus grown in the chick embryo permits its separation by gel filtration under reducing and nonreducing conditions. Chromatography of the viral proteins without reduction separates the hemagglutinin (HA) from all viral proteins with the exception of the nucleoprotein (NP) (2). Reduction of the HA-NP fraction and chromatography results in the separation of NP, HA_1 and HA_2.

In the present study we report the isolation and amino-terminal sequence analysis of hemagglutinin from a production lot of an experimental formalin-inactivated influenza vaccine. The virus in the vaccine was X-38 (Heq1N2), a recombinant resulting from the triparental cross of A/Eng/42/72(MRC-11) (H3N2), A/PR/8/34 (HON1) (Ann Arbor variant), and A/equine/Prague/1/56 (Heq1Neq1) (E. D. Kilbourne, unpublished data). The hemagglutinin of the vaccine virus (Heq1) has little antigenic resemblance to the hemagglutinin of the human A subtype viruses (3), and accordingly has been utilized as an "irrelevant" hemagglutinin in neuraminidase-specific vaccines (4, 5).

Abbreviations: HA, hemagglutinin; NP, nucleoprotein; PTH, phenylthiohydantoin.

METHODS

Preparation of Virus. Influenza vaccine prepared from X-38 virus (Heq1N2) was a gift from Lederle Laboratories, Pearl River, N.Y. The vaccine was a pool of three viral batches prepared separately. The virus was grown in the allantoic cavity of 10-day-old embryonated eggs. The allantoic fluids were harvested and the cellular debris was removed by low-speed centrifugation. The virus was pelleted by continuous flow at 20 liters/hr through a zonal rotor in a centrifuge operated at 35,000 rpm (75,000 × *g*). The viral pellet was resuspended in 1/20 the original volume of the allantoic fluid in 0.1 M sodium phosphate buffer, pH 7.0, and treated with formaldehyde at 1:4000 dilution for 2–4 days at 4°. The formaldehyde was lowered to a concentration of 0.00037% by dialysis. Contaminating protein in the viral preparation was removed by precipitating twice with 0.2 M $MgSO_4$ at pH 8.5 at room temperature and chilling overnight. The precipitated protein was removed by low-speed centrifugation at 1500 × *g* for 20 min. The viral preparation was concentrated to 1/60 the original volume of the allantoic fluid by pressure ultrafiltration in an Amicon unit. The viral preparation was filtered through the Diaflow unit versus 0.05 M sodium phosphate buffer and ethylene oxide was added to a final concentration of 0.01 mg/ml to sterilize the vaccine. The final vaccine preparation contained 0.019 mg/ml of thimerosal, 0.319 mg/ml of ethylene glycol, and 0.010 mg/ml of ethylene oxide in 0.05 M sodium phosphate buffer, pH 7.0.

The protein concentration of the vaccine was 1.0 mg/ml as assayed according to Lowry *et al.* (6). Virus was pelleted from the vaccine preparation by centrifugation at 48,200 × *g* for 90 min. A large amount of protein was not sedimentable: 0.47 mg/ml could be removed by pelleting, 0.53 mg/ml remained in the supernatant. Since the viral neuraminidase activity was distributed evenly between the pellet and the supernatant, the protein remaining in the supernatant was presumed to be derived from disrupted virus.

Purification of the Hemagglutinin. The pellet from the above procedure, consisting of 100–300 mg of protein, was disrupted with 10–15 ml of 10% sodium dodecyl sulfate and ultrasonicated with a Biosonik unit (Bronwill) until the solution was transparent, which generally required 2 min. The

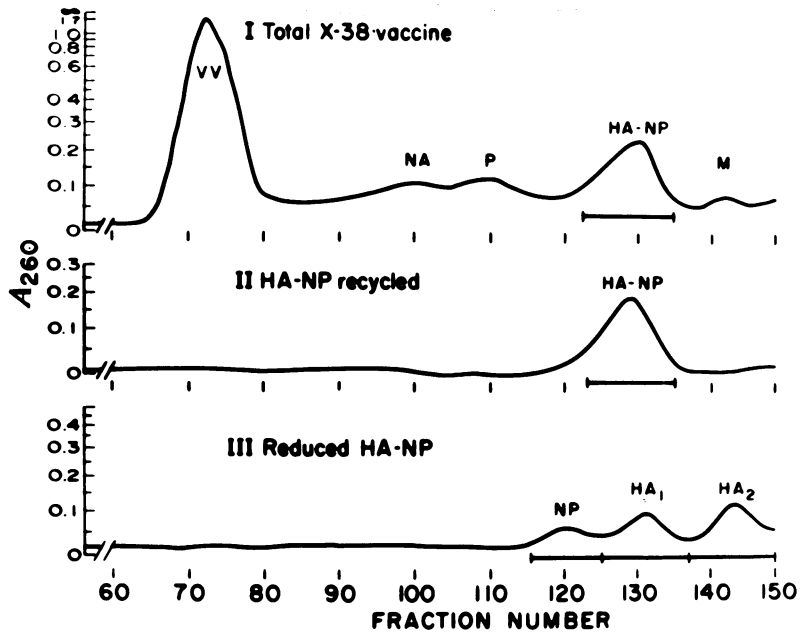

FIG. 1. Sodium dodecyl sulfate gel filtration on Bio-Gel A-5m of sodium dodecyl sulfate-disrupted viral vaccine. Chromatograms represent respectively, (I) chromatography of pelleted, disrupted influenza virus vaccine, (II) rechromatography of HA-NP fraction, (III) reduction with 0.1 M dithiothreitol and chromatography of HA-NP from II. Letters designating column fractions: VV, void volume; NA, neuraminidase; P, nonglycosylated polypeptide of molecular weight 122,000; HA-NP, hemagglutinin-nucleoprotein coeluted; M, nonglycosylated polypeptide of 27,000. Elution was performed with 0.02 M Tris·HCl buffer, pH 7.4, containing 0.1% sodium dodecyl sulfate, and 0.05% sodium azide at a flow rate of 12 ml/hr. The total quantity of protein applied was 140 mg. Fractions pooled for viral polypeptides are shown by bars under respective peaks.

viral preparation was then heated at 56° for 60 min before application to the column. The sample was pumped onto the base of a 2.5 × 90 cm column of Bio-Gel A-5m (Bio-Rad) connected in tandem to a second 2.5 × 90 cm column fitted with flow adapters (Pharmacia). The column material was previously equilibrated with 0.02 M Tris·HCl, pH 7.4, containing 0.1% sodium dodecyl sulfate and 0.05% sodium azide. All fractions were collected with an LKB 2000 fraction collector and the A_{260} was monitored with an LKB Uvicord I. Elution from the column was carried out with the same buffer at a flow rate of 12–15 ml/hr at room temperature. Elution of the HA-NP fraction required about 40 hr (see Fig. 1). The HA-NP fractions were pooled and concentrated by pressure ultrafiltration to 4.0 ml with the Amicon model 52 unit equipped with PM-10 membranes. The HA-NP pool was rechromatographed on the same pair of columns under the same conditions to remove residual P and M proteins, viral polypeptides of 122,000 and 27,000 daltons, respectively, as determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis (7) (see Fig. 1). The HA-NP pool was again concentrated to 4.0 ml. Dithiothreitol was added to 0.1 M and incubated at 37° for 30 min to reduce the hemagglutinin to HA_1 and HA_2 and thus permit their separation from NP. The reduced HA-NP preparation was chromatographed a third time on the same pair of columns and under the same conditions as for the first cycle (see Fig. 1). The results of polyacrylamide gel electrophoresis of the products of the third column cycle, pools of purified NP, HA_1, and HA_2 are shown in Fig. 2.

Removal of Sodium Dodecyl Sulfate and Preparation of Polypeptides for Sequencing. Sodium dodecyl sulfate was removed according to the procedure of Weber and Kuter using a 1.0 × 15 cm column of Dowex 1 (8). The viral proteins were incubated in a total of 10 ml of Tris-acetate buffer, pH 7.8, with 6 M urea at 37° for 30 min before being applied to the column. Viral protein samples were applied to the column in 5 ml aliquots and held for 15 min to improve removal of sodium dodecyl sulfate. An additional 50 ml of Tris-acetate buffer, pH 7.8, in 6 M urea was used to elute the viral protein. A total of 60 ml was collected and pooled from each column.

Following elution from Dowex 1, the hemagglutinin polypeptides, in approximately 60 ml of 6 M urea-buffer solution, were reduced with 0.01 M dithiothreitol in a sealed container for 2 hr at 37°. Sulfhydryl groups were blocked by the addition of 0.04 M iodoacetamide. The polypeptide solutions were dialyzed at 4° for 4–5 days with daily changes of distilled water. Following dialysis, HA_1 preparations remained in solution; HA_2 preparations formed a fine white

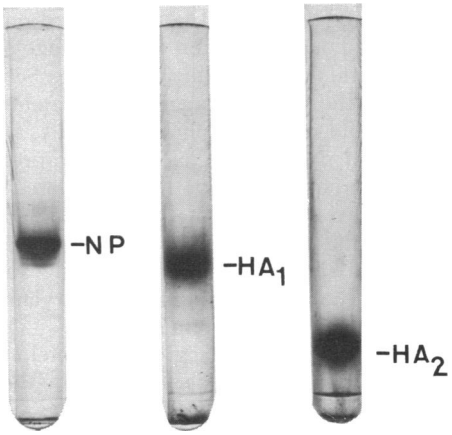

FIG. 2. Sodium dodecyl sulfate polyacrylamide gel electrophoresis of NP, HA_1, and HA_2 protein pools from Fig. 1, chromatographic cycle III. Approximately 10–15 μg of viral protein was applied per gel and stained with Coomassie brilliant blue. Viral proteins had been reduced and alkylated for sequence analysis.

Table 1. Sequential Edman degradation of HA_1

Step no.	Deduced residue	SP400*		TLC†
		$-SiH_3$	$+SiH_3$	
1	Asp	-0-	Asp	Asp
2	Lys	-0-	Lys‡	Lys
3	Ile	Leu/Ile	Ile	Leu/Ile
4	Ser	Ser	Ser‡	Ser
5	Leu	Leu/Ile	Leu	Leu/Ile
6	Gly	Gly	Gly	Gly
7	Tyr	Tyr‡	Tyr‡	Tyr‡
8	His	-0-	0	His
9	Ala	Ala	Ala	Ala
10	Val	Val	Val	Val

-0- indicates that the sample was analyzed but no phenylthiohydantoin (PTH) amino acid was detected. Leu/Ile indicates that PTH-amino acids of leucine and isoleucine cannot be differentiated unambiguously.

* Only those PTH-amino acids extracted with ethyl acetate following conversion were analyzed on the column of SP400 in gas–liquid chromatography. $+SiH_3$ and $-SiH_3$ indicate the analyses of silylated and unsilylated PTH-derivatives, respectively.

† TLC means the thin-layer chromatography of PTH-amino acids on polyamide sheets (5×5 cm) with a fluorescent indicator in the first solvent as described (13).

‡ Not identified unequivocally.

precipitate. All preparations were lyophilized for sequence analysis.

Automatic Edman Degradation. Amino-acid sequence was determined by Edman degradations on 0.1–0.2 μmol of hemagglutinin polypeptides, HA_1 and HA_2, performed automatically with a Beckman sequencer using *N,N'*-dimethylallylamine buffer and single acid cleavage (9–12). Phenylthiohydantoin-amino acids were identified by gas–liquid chromatography (11) and thin-layer chromatography (13). (For further details see footnotes of Table 1.)

RESULTS

The hemagglutinin polypeptides of influenza virus as HA_1 and HA_2 were obtained in large quantities from a batch of vaccine made from X-38 virus. When 300 mg of total viral vaccine protein was applied to the column at the first stage, 20–22 mg (7%) was recovered as NP, 32–38 mg (11–13%) as HA_1, and 12–20 mg (4–7%) as HA_2 at the third chromatographic cycle. Based on compilation of percentage composition of the polypeptides of the influenza virus by Schulze for those viruses with the total hemagglutinin present as HA_1 and HA_2, HA_1 should represent 19–21% and HA_2 should be 12–15% of the total protein (14). NP should represent 26–31% of total viral protein. Thus we have recovered as purified products approximately one-half of the available hemagglutinin polypeptides and approximately one-fourth of the NP, assuming all of the protein present in the vaccine to be viral protein. Sodium dodecyl sulfate polyacrylamide gels of the purified products are shown in Fig. 2.

A significant quantity of hemagglutinin protein was lost during the removal of sodium dodecyl sulfate by the Weber and Kuter technique (8). Recovery ranged from 40 to 80% of the protein applied. This technique was used to insure complete removal of sodium dodecyl sulfate prior to sequence analysis.

The amino-acid sequence of the first 10 residues of vaccine-derived HA_1 was deduced from three runs of automatic Edman degradation on 0.1 μmol of the glycoprotein. The results are summarized in Table 1. It should be noted that the lysine at position no. 2 and tyrosine at position no. 7 were only tentatively identified. Difficulty with identification of the residues may reflect modification of amino acids during treatment of the viral preparation with formaldehyde or ethylene oxide for its use as a vaccine.

Table 2. Sequential Edman degradation of HA_2

Step no.	Deduced residue	SP400		TLC
		$-SiH_3$	$+SiH_3$	
1	Gly	Gly	Gly	Gly
2	Leu	Leu/Ile	Leu	Leu/Ile
3	Phe	Phe	Phe	Phe
4	Gly	Gly	Gly	Gly
5	Ala	Ala	Ala	Ala
6	Ile	Leu/Ile	Ile	Leu/Ile
7	Ala	Ala	Ala	Ala
8	Gly	Gly	Gly	Gly
9	Phe	Phe	Phe	Phe
10	Ile	Leu/Ile	Ile	Leu/Ile
11	Glu	-0-	Glu	Glu
12	Asn	-0-	Asn†	Asn
13	Gly	Gly	Gly	Gly
14	Trp	Trp†	Trp	Trp
15	Glu	-0-	Glu	Glu
16	Gly	Gly	Gly	Gly
17	Leu	Leu/Ile	Leu	Leu/Ile
18	Ile	Leu/Ile	Ile	Leu/Ile
19	Asp	-0-	Asp	Asp
20	Gly	Gly	Gly	Gly
21	?*	*	*	*
22	Tyr	Tyr	Tyr	Tyr
23	Gly	Gly†	Gly	Gly
24	Tyr	Tyr	Tyr	Tyr

For explanation of the abbreviations see the legend of Table 1.

* Amino acid was not positively identified.

† Not identified unequivocally.

Two runs of automatic Edman degradation on 0.2 μmol of HA_2 established the amino-acid sequence of 23 residues from the NH_2-terminus as shown in Table 2. The amino acid at position no. 21 was not positively identified.

DISCUSSION

Utilization of sodium dodecyl sulfate-gel filtration has permitted the separation of HA_1 and HA_2 from influenza virus inactivated by formaldehyde for use as a vaccine. Use of a detergent for separation of these polypeptides has circumvented the need for proteolytic agents and results in purification of both HA_1 and HA_2 in their entirety. Use of outdated vaccine as a source of viral hemagglutinin has resulted in tremendous savings of time and expense. From this source, viral proteins can be obtained in the large quantities necessary for elucidation of their primary structure.

Since separation of the hemagglutinin polypeptides by sodium dodecyl sulfate gel filtration is dependent on the subunit structure of hemagglutinin, this approach should be applicable to all strains of influenza viruses, providing purified HA_1 and HA_2 for structural analysis. Influenza viruses grown *in ovo* possess a hemagglutinin consisting of two polypeptides which have arisen from asymmetric cleavage of a precursor glycoprotein (1). The two polypeptides of the

HA_1		1 5 10	
X-38	(Heq1)	Asp-Lys- Ile- Ser- Leu-Gly-Tyr-His-Ala- Val	This report
Bel	(H0)	— Thr — — Ile — — — — Asx	Skehel and Waterfield (16)
Weiss	(H1)	Asn-Thr — — Ile — — — — Asx	(16)
Singapore	(H2)	Asn-Glu — — Ile — — — — Asx	(16)
HA_2		1 5 10 15 20 24	
X-38	(Heq1)	Gly-Leu-Phe-Gly-Ala- Ile- Ala-Gly-Phe- Ile- Glu-Asn-Gly-Trp-Glu-Gly-Leu-Ile-Asp-Gly-?-Tyr-Gly-Tyr	
Bel	(H0)	— — — — — — — — — — Glx-Gly —	Skehel and Waterfield (16)
Weiss	(H1)	— — — — — — — — — — Glx-Gly —	(16)
Singapore	(H2)	— — — — — — — — — — Glx-Gly	(16)
X-31	(H3)	— — — — — — — — — — Glx-Asx	(16)
MRC-11	(H3)	— — — — — — — — — — Glx-Asx —	(16)

FIG. 3. Amino-terminal sequences of influenza virus hemagglutinin subunits. The identification of respective residues for HA_1 and HA_2 (Heq1) is shown in Tables 1 or 2. A — indicates the amino acid is identical to that in X-38 protein.

hemagglutinin are held together by disulfide bonds; reduction of the disulfide bonds permits their separation from other viral proteins by sodium dodecyl sulfate gel filtration.

In addition to providing pure HA_1 and HA_2, the sodium dodecyl sulfate gel filtration technique simultaneously purifies the NA, P, NP, and M polypeptides for analysis of these components. If desired, the neuraminidase can be removed in active form by an affinity isolation technique prior to sodium dodecyl sulfate gel filtration (2).

Formaldehyde has long been used as an agent for the inactivation of toxin and viral preparations for vaccines. Formaldehyde interacts with nucleic acids, rendering them noninfective, but does not seriously affect the stability or antigenicity of the external viral proteins by the substitution of available lysine, amide, and cysteine side chains (15). Evidently the conditions of formaldehyde inactivation of the virus for vaccine production did not introduce modifications of amino acids of the hemagglutinin polypeptides that interfere significantly either with the purification or automatic sequencing of these polypeptides. The amino-terminal sequences of hemagglutinin HA_1 and HA_2 subunits of equine type 1 Heq1 (X-38) are clearly homologous to the partial sequences of type A subtypes (H0, H1, H2 and H3) (16) (Fig. 3).

The amino-acid sequence of the first 10 residues of HA_1 of equine type 1 has three or four differences from those of human type A subtypes (H0, H1, and H2), and these differences can be explained by single nucleotide substitutions. The amino-acid sequence of the first 10 residues of HA_2 of equine type 1 is identical to that of human type A subtypes (H0, H1, H2, and H3) of influenza virus.

On the basis of these limited preliminary studies, it appears that the amino-acid sequences of HA_2 subunits are more conserved than those of HA_1 subunits. This is consistent with earlier observations that the antigenicity of influenza viral hemagglutinin appears to reside with HA_1 rather than HA_2 (17, 18) and the finding of greater variation in peptide maps of HA_1 than of HA_2 (19). Recent findings of Webster and Laver, however, show significant differences in peptide maps of HA_2 from H2 and H3 subtypes of influenza virus (20). The partial sequence of HA_2 of equine type 1 (X-38) is identical to other influenza virus A subtypes through the first 10 residues. The asparagine-glycine difference at residue no. 12 can only be explained by two nucleotide replacements.

The amino-acid sequence of the first 24 residues of HA_2 hemagglutinin shows several features of interest: (*1*) a pseudo-palindrome of seven amino acids centered on the isoleucine at position 6, observed by Skehel and Waterfield, is confirmed (16); (*2*) Almost one-third of 23 residues identified are glycine, although only about 10% of entire HA_2 are glycine, based on the data of Laver and Baker (21), and Skehel and Waterfield (16); (*3*) These glycine residues are regularly distributed, that is, every three or four residues are glycine residues. If this portion of HA_2 is an α-helix, this finding suggests a spatial constraint for one side of the helix, as the glycyl residues would have no side chain.

Although the limited information available shows a larger number of differences in amino acid sequence among HA_1 polypeptides than among HA_2 polypeptides sequenced from the amino-termini, and the principal antigenicity to hemagglutinin is thought to reside with HA_1; the sequences of the HA_1 and HA_2 polypeptides must be determined in their entirety before antigenicity can be related to amino-acid sequence. Given the present technique for purification of the polypeptides, which provides large quantities of viral polypeptides unmodified by proteolytic enzymes, it should be possible to accomplish the complete sequence analysis of both polypeptides of influenza viral hemagglutinin.

Note Added in Proof. In the absence of radioactive labeling of the sulfhydryl group of cysteine and half-cystine with iodoacetamide, the PTH-serine and PTH-*S*-carboxymethylcysteine are not readily distinguishable. Thus the serine at residue no. 4 of HA_1 must be considered presumptive.

We gratefully acknowledge the assistance of Dr. Ruth Kaplan in the identification of PTH amino acids on thin-layer chromatography and M. Diane Forde for excellent technical assistance. This work was monitored by the Commission on Influenza of the Armed Forces Epidemiological Board and was supported (in part) by the U.S. Army Medical Research and Development Command under Research Contract no. DADA17-69-C-9137, by NIH Research Grant no. AI09304 and AI 09810-06 awarded by the National Institute of Allergy and Infectious Diseases, by NIH Research Grant no. CA 18621-01 awarded by the National Cancer Institute, and by Contract no. U-2076E from the Health Research Council of the City of New York. J.M.K. is an Established Investigator of the American Heart Association.

1. Lazarowitz, S. G., Compans, R. W. & Choppin, P. W. (1971) *Virology* **46**, 830–843.
2. Bucher, D. J. (1975) in *Negative Strand Viruses*, eds. Mahy, B. W. J. & Barry, R. D. (Academic Press, London), Vol. 1, pp. 133–143.
3. Kilbourne, E. D. (1968) *Science* **160**, 74–76.
4. Kilbourne, E. D., Schulman, J. L., Couch, R. B. & Kasel, J. A. (1972) in *International Virology 2, Proceedings of the Second International Congress of Virology*, ed. Melnick, J. L. (S. Karger, Basel), pp. 118–119.
5. Couch, R. B., Kasel, J. A., Gerin, J. L., Schulman, J. L. & Kil-

bourne, E. D. (1974) *J. Infect. Dis.* **129,** 411–420.
6. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. (1951) *J. Biol. Chem.* **193,** 265–275.
7. Bucher, D. J. & Kilbourne, E. D. (1972) *J. Virol.* **10,** 60–66.
8. Weber, K. & Kuter, D. J. (1971) *J. Biol. Chem.* **246,** 4504–4509.
9. Edman, P. & Begg, G. (1967) *Eur. J. Biochem.* **1,** 80–91.
10. Li, S.-L., Hanlon, J. & Yanofsky, C. (1974) *Biochemistry* **13,** 1736–1744.
11. Pisano, J., Bronzert, T. J. & Brewer, H. B., Jr. (1972) *Anal. Biochem.* **45,** 43–59.
12. Kehoe, J. M. & Capra, J. D. (1971) *Proc. Nat. Acad. Sci. USA* **68,** 2019–2021.
13. Summers, M. R., Smythers, G. W. & Oroszlan, S. (1973) *Anal. Biochem.* **53,** 624–628.
14. Schulze, I. T. (1973) *Adv. Virus Res.* **18,** 1–55.
15. Fraenkel-Conrat, H. (1968) in *Molecular Basis of Virology,* ed. Fraenkel-Conrat, H. (Reinhold Book Corp., New York), p. 152.
16. Skehel, J. J. & Waterfield, M. D. (1975) *Proc. Nat. Acad. Sci. USA* **72,** 93–97.
17. Brand, C. M. & Skehel, J. J. (1972) *Nature New Biol.* **238,** 145–147.
18. Eckert, E. A. (1973) *J. Virol.* **11,** 183–192.
19. Laver, W. G. & Webster, R. G. (1973) *Virology* **51,** 383–391.
20. Webster, R. G. & Laver, W. G. (1975) in *The Influenza Viruses and Influenza,* ed. Kilbourne, E. D. (Academic Press, New York), in press.
21. Laver, W. G. & Baker, N. (1972) *J. Gen. Virol.* **17,** 61–67.