Proc. Nat. Acad. Sci. USA
Vol. 72, No. 1, pp. 93–97, January 1975

Studies on the Primary Structure of the Influenza Virus Hemagglutinin

(viral membrane protein/detergent and bromelain purification/hydrophobic tail/protein palindrome)

JOHN J. SKEHEL* AND MICHAEL D. WATERFIELD†

*National Institute for Medical Research, Mill Hill, London, NW7 1AA; and †Imperial Cancer Research Fund, Lincoln's Inn Fields, London, W.C. 2

Communicated by Jack L. Strominger, October 7, 1974

ABSTRACT **The amino-terminal sequence and composition of the subunits of the hemagglutinin (HA) of influenza virus has been determined. The hemagglutinin has been isolated by two techniques, (*1*) as the intact hemagglutinin after disruption of the virus in sodium dodecyl sulfate, giving 2 subunits of 58,000 daltons (HA1) and 26,000 daltons (HA2), and (*2*) after treatment of the virus with bromelain, giving 2 subunits of 58,000 daltons (BHA1) and 21,000 daltons (BHA2). In both preparations these subunits are linked by disulfide bonds. The amino-terminal sequences of HA1 and BHA1, and HA2 and BHA2 are the same. The composition of the 50 residue peptide associated with the membrane, which is removed from the C-terminus of HA2 by bromelain, is deduced and shown to be hydrophobic and contain 50% of the serine residues of HA2.**

The biosynthetic precursor of the hemagglutinin has been purified from the membranes of abortively infected chick fibroblasts and shown to have the same amino terminus as HA1. Thus the order of biosynthesis is NH_2-HA1-HA2-COOH. The amino-terminal sequence of BHA2—at the cleavage site of the precursor—is shown to be a palindrome:

NH_2-Gly-Leu-Phe-Gly-Ala-Ile-Ala-Gly-Phe-Ile-.

This sequence is conserved in representative viruses from each of the major pandemics. A region of homologous sequence is described between the hemagglutinins of influenza type A and B viruses.

The lipoprotein envelope of the influenza virion contains two virus-specified glycoproteins, a hemagglutinin and a neuraminidase. The former, which is the subject of this report, is responsible for the initial interaction between virus and cell during infection and it is the component of the virus against which neutralizing antibodies are directed. Because of these characteristics and since analyses of the properties of lipid-associated virus glycoproteins may be of value in elucidating the contribution of glycoproteins in general to membrane structure, studies of the amino-acid sequence of the hemagglutinin have been initiated.

The hemagglutinin is the major component of the virus envelope and accounts for approximately 30% of the total virus protein (1, 2). It is a rod-shaped molecule with an apparently triangular cross section, has a molecular weight of about 210,000, and is made up of two types of glycosylated subunits of apparent molecular weights 58,000 and 26,000 which are linked together in the molecule by disulfide bonds (3, 4). Analyses of protein synthesis in virus-infected cells have indicated that the two subunits are formed during virus assembly at the plasma membrane of the infected cell as the result of proteolytic cleavage of a glycosylated precursor polypeptide of molecular weight approximately 80,000 (5–8).

In this communication the amino-acid compositions and amino-terminal amino-acid sequences of the subunits and their biosynthetic precursor are reported. The results allow definition of both the site of cleavage of the precursor and the sequence of biosynthesis of the subunits. They also suggest that the carboxyl terminus of the smaller subunit is associated with the virus membrane and show that regions of sequence homology are present in the hemagglutinins of both type A and type B influenza viruses.

Abbreviations: HA, hemagglutinin isolated without proteolysis; BHA, hemagglutinin isolated after treatment with bromelain.

METHODS

Virus Strains. All viruses were grown in the allantonic cavity of 10-day-old embryonated eggs. Inocula for A0/Bel/42, A2/Singapore/57, X-31, MRC-11, and B/Lee were obtained from the World Influenza Centre, London and for A1/Weiss/57 from Dr. W. Dowdle, Communicable Disease Center, Atlanta. Viruses were concentrated and purified as described previously (3).

Hemagglutinin Isolation. (*1*) Following disruption of purified virus particles in 2% (w/v) sodium dodecyl sulfate at 20°, the dissociated proteins were separated by electrophoresis on cellulose acetate as described by Laver (9) except that Celogel (Chemetron, Milan) was used as supporting medium.

(*2*) Purified virus particles were incubated with stem bromelain (Sigma; EC 3.4.22.4) and the released hemagglutinin was purified by sucrose density gradient centrifugation as described previously (10).

Separation of the Hemagglutinin Subunits. (*1*) The two subunits of the hemagglutinin isolated by electrophoresis of the dissociated virus proteins on cellulose acetate were separated and purified in guanidine hydrochloride gradients as described by Laver (4).

(*2*) The two subunits of the bromelain-released hemagglutinin were reduced with 10 mM dithiothreitol in 6 M guanidine hydrochloride containing 0.2 M Tris·HCl, pH 8.5, at 37° for 2 hr. After cooling in ice the free sulfhydryl groups were alkylated with a 2-fold molar excess of iodoacetamide, and the protein was dialyzed against 0.05 M Tris in 1 M guanidine hydrochloride overnight and then against 6 M urea containing 0.2 M formic acid for 8 hr. The subunits were separated on a 2.5 × 100-cm column of Sephadex G-100 in 6 M urea containing 0.2 M formic acid, Fig. 1.

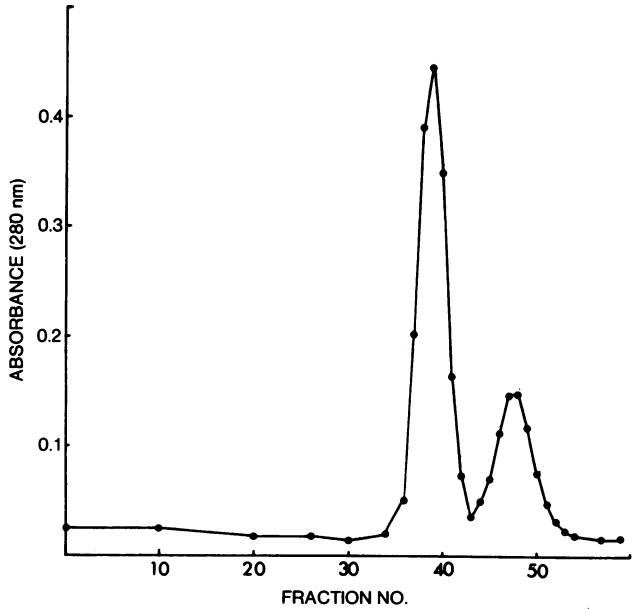

FIG. 1. Separation of the subunits BHA1 and BHA2 of the hemagglutinin released from virus particles by bromelain digestion. BHA preparations were reduced with dithiothreitol reagent in 6 M guanidine hydrochloride as described in *Methods* and after dialysis against 6 M urea in 0.2 M formic acid were subjected to gel filtration on a 2.5 × 100-cm column of Sephadex G-100 equilibrated with the same solvent. The separation was monitored by ultraviolet absorption at 280 nm.

Isolation of the Hemagglutinin Precursor Glycopolypeptide. Monolayer cultures of primary chick embryo fibroblasts (5×10^7 cells/petri dish) were infected at an added multiplicity of approximately 100 plaque-forming units/cell for 30 min at 20° and then washed twice, overlayed with 10 ml of Geys balanced salt solution, and incubated at 37° for 14 hr. The infected cells were then suspended in hypotonic buffer (10 mM Tris·HCl, 1 mM $MgCl_2$, pH 7.5) and homogenized as described previously (8). Nuclei were removed by centrifugation (1000 × *g*, 5 min) and a large particle fraction containing most of the cell membranes was obtained from the nuclear supernatant by centrifugation at 80,000 × *g* for 60 min. The pelleted material was resuspended in Tris–boric acid–EDTA buffer (0.66 M, 0.12 M, and 0.015 M, respectively, pH 8.9) containing 0.5% (w/v) sodium dodecyl sulfate and the hemagglutinin precursor isolated by electrophoresis on cellulose acetate in sodium dodecyl sulfate as for hemagglutinin purification procedure (*1*) above.

TABLE 1. *The amino-acid compositions of the hemagglutinin subunits*

Amino acid	HA1	BHA1	HA2	BHA2	HA2 − BHA2
Aspartic acid	67.8	66.3	36.0	31.9	4.1
Threonine	40.8	40.5	11.9	9.8	2.1
Serine	49.6	49.2	21.2	10.2	11.0
Glutamic acid	62.7	60.4	31.7	25.4	6.3
Proline	32.5	32.2	5.3	2.9	2.4
Glycine	44.7	43.2	24.0	19.4	4.6
Alanine	29.7	28.8	13.5	10.2	3.3
Valine	33.0	32.0	15.4	10.0	5.4
Cysteine*	12.6	12.5	5.2	3.4	1.8
Methionine†	4.0	4.0	5.8	4.4	1.4
Isoleucine	37.8	36.3	15.5	14.2	1.3
Leucine	47.4	48.3	21.7	13.3	8.4
Tyrosine	25.0	26.4	12.8	11.8	1.0
Phenylalanine	19.0	18.5	10.5	9.3	1.2
Histidine	15.3	15.4	4.6	4.6	0.0
Lysine	32.6	32.4	19.0	16.7	2.3
Arginine	25.8	25.8	6.1	4.8	1.3

* As cysteic acid.
† As methionine sulfone.

Polyacrylamide Gel Electrophoresis in sodium dodecyl sulfate was as described before (3).

Amino-Acid Composition and Sequence Analysis. Proteins were hydrolyzed (11) in 1 ml of 6 N HCl containing 5 μl of butanedithiol in evacuated, sealed tubes for 12, 24, 48, and 72 hr except in the case of performic-acid-oxidized samples (12), which were only hydrolyzed for 24 hr. Values for amino acids were corrected for destruction during hydrolysis. Compositions were determined using a Durrum D500 amino-acid analyzer.

Automated Edman degradations were performed on 1- to 5-mg samples with a Beckman model 890 C sequencer using 1 M Quadrol buffer and two acid cleavages according to Edman and Begg (13). Phenylthiohydantoin derivatives were identified by gas chromatography (14) and, following back hydrolysis with hydroiodic acid (15), as amino acids on the Durrum amino-acid analyzer.

RESULTS

The Amino-Acid Composition and Amino-Terminal Sequences of the Glycopolypeptide Subunits. Two different procedures described in the previous section were used to isolate the hemagglutinin from purified virus particles of the strain A0/Bel/42. The first takes advantage of the peculiar electrophoretic mobility of the hemagglutinin of several influenza viruses on a cellulose acetate support in buffers containing sodium dodecyl sulfate (9). The protein isolated in this way (HA) is antigenically active, retains the ability to agglutinate chicken erythrocytes, and consists of two glycopolypeptide subunits of molecular weights 58,000 (HA1) and 26,000 (HA2) (Fig. 3). From such preparations of hemagglutinin, HA1 and HA2 subunits were separated following dissociation in guanidine hydrochloride (4).

The second hemagglutinin purification procedure involves digestion of purified virus particles with the protease bromelain. As reported previously (10) such preparations of the protein are antigenically active but do not aggregate and probably as a consequence do not agglutinate chicken erythrocytes. The protein purified by this procedure (BHA) consists of two types of subunit of molecular weights 58,000 (BHA1) and 21,000 (BHA2). These two subunits were separated by gel filtration in urea solutions (Fig. 1).

The results of amino-acid analyses and amino-terminal sequence analyses of these four subunits, HA1, HA2, BHA1, and BHA2 are shown in Tables 1 and 2, respectively. Their amino-acid compositions were calculated using values for their molecular weights of: HA1, 58,000; HA2, 26,000; BHA1, 58,000; and BHA2, 21,000. Since the subunits are glycopolypeptides these estimates, which are based on their mobility on polyacrylamide gel electrophoresis in sodium dodecyl sulfate (16), may not be accurate. However, for A0/Bel/42 most of the carbohydrate residues are associated with the larger subunits, HA1 and BHA1, and consequently the estimates of the molecular weights of the smaller subunits, HA2 and BHA2, are more

TABLE 2. *The amino-terminal amino-acid sequences of the hemagglutinin subunits*

BHA2	5 10
Bel	Gly-Leu-Phe-Gly-Ala-Ile-Ala-Gly-Phe- Ile -Glx-Gly-Gly
Weiss	Gly-Leu-Phe-Gly-Ala-Ile-Ala-Gly-Phe- Ile -Glx-Gly-Gly
Singapore	Gly-Leu-Phe-Gly-Ala-Ile-Ala-Gly-Phe- Ile -Glx-Gly
X-31	Gly-Leu-Phe-Gly-Ala-Ile-Ala-Gly-Phe- Ile -Gly-*Asx*
MRC-11	Gly-Leu-Phe-Gly-Ala-Ile-Ala-Gly-Phe- Ile -Glx-*Asx*-Gly
B/Lee	Gly-*Phe*-Phe-Gly-Ala-Ile-Ala-Gly-Phe-*Leu*-Glx-Gly-Gly
BHA1	10
Bel	Asp-Thr-Ile-Ser-Ile-Gly-Tyr-His-Ala-Asx
Weiss	Asn-Thr-Ile-Ser-Ile-Gly-Tyr-His-Ala-Asx
Singapore	Asn-*Glu*-Ile-Ser-Ile-Gly-Tyr-His-Ala-Asx
Precursor	
Bel	Asx-Thr-Ile
HA2	
Bel	Gly-Leu-Phe-Gly
HA1	
Bel	Asp-Thr-Ile-Ser

likely to be accurate. From this data it is apparent that the amino-acid compositions of HA1 and BHA1 are essentially identical but BHA2 contains about 50 amino-acid residues less than HA2. The amino-terminal sequence analyses further indicate that HA1 and BHAl, and HA2 and BHA2 are identical. Further, they imply that the modifications that result from the bromelain digestion procedure occur at the carboxyl terminal portion of the glycopolypeptides.

The Direction of Synthesis and the Site of Proteolytic Cleavage of the Hemagglutinin Precursor. As stated above, the hemagglutinin appears to be synthesized during influenza virus replication as the product of a single influenza virus gene and to be cleaved subsequently, presumably by a host cell protease, to generate the two glycopolypeptide subunits HA1 and HA2. However, in chicken cells this sequence of events does not occur during certain abortive infections, which do not result in the production of infectious progeny virus. For example in this study of chick embryo fibroblasts infected with A0/Bel/42, the hemagglutinin precursor is not cleaved and neither HA1 nor HA2 are detected. Instead, the precursor protein accumulates in the plasma membrane of infected cells (Fig. 2) and can be isolated from cell homogenates as described in *Methods*. The amino-terminal sequence of the precursor glycopolypeptide is shown in Table 2. The results indicate that the amino terminals of the precursor and HA1 are identical and together with the data obtained for the HA1, HA2, BHA1, and BHA2 subunits allow the following deductions: (*1*) The direction of synthesis of the two glycopolypeptides is from the amino terminus of HA1 to the carboxyl terminus of HA2 and (*2*) the cleavages of the precursor molecule generate the amino terminus of HA2 and the carboxyl terminus of HA1.

The Amino-Terminal Sequence of BHA1 and BHA2 from Different Strains of Influenza. A comparison of the amino-terminal sequences of representative viruses from the major subgroups of human type A influenza viruses shows remarkably little variation of sequence within the small region examined for both of the hemagglutinin subunits. Amino-acid substitutions were found at position 2 in the BHA1 subunit of A2/Singapore/57 and at positions 12 and 15 in the BHA2 subunits of X-31 and MRC-11. The amino-terminal sequence of the BHA2 subunit of a single type B influenza virus, B/Lee, was also determined and as shown in Table 2 this region of the type A and B virus proteins is clearly homologous.

DISCUSSION

Biosynthesis and Cleavage of the Hemagglutinin Precursor. The results presented here form the first part of an analysis of amino-acid sequence of the influenza virus hemagglutinin and have allowed a number of interesting conclusions to be made

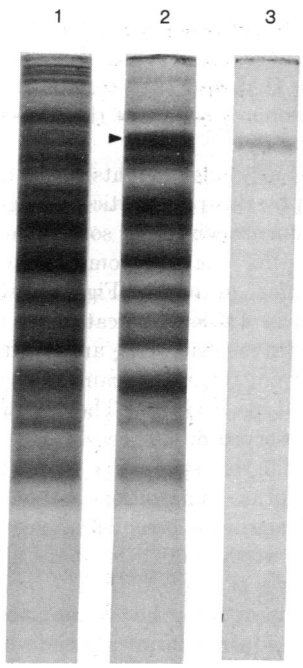

FIG. 2. Isolation of the hemagglutinin precursor from homogenates of A0/Bel/42-infected chick embryo fibroblasts. Monolayer cultures of chick embryo fibroblasts were infected and processed and the hemagglutinin precursor was isolated by electrophoresis on cellulose acetate as described in *Methods*. Purification of the protein was assayed by electrophoresis in polyacrylamide gels containing sodium dodecyl sulfate. (*1*) Uninfected cell large particle fraction. (*2*) Infected cell large particle fraction which was applied to the cellulose acetate for isolation of the precursor. (*3*) Hemagglutinin precursor.

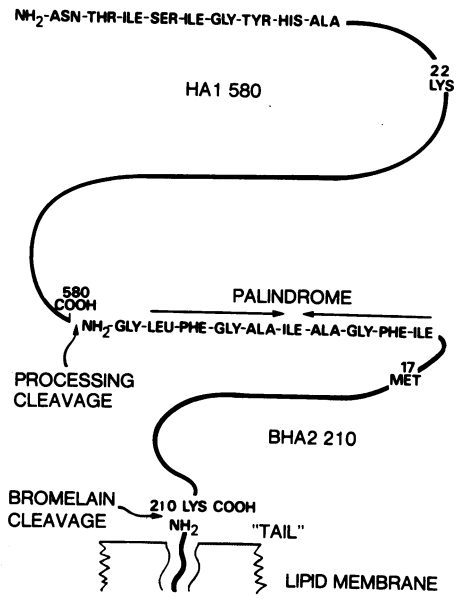

FIG. 3. Schematic representation of the hemagglutinin.

concerning the structure and biosynthesis of the molecule, which are summarized in Fig. 3. It has been suggested previously from the results of pulse-chase experiments in influenza virus-infected cells that the two hemagglutinin glycopolypeptide subunits, HA1 and HA2, are derived from a single glycosylated precursor as a result of proteolytic cleavage(s) (5–8). The present results, which indicate the identity of the amino-terminal sequences of the putative precursor and HA1, support this conclusion. They also indicate the sequential synthesis of the HA1 and HA2 regions of the precursor and imply that the amino terminus of HA2 is derived as a result of limited proteolysis.

The apparent molecular weights of the hemagglutinin subunits generated by this proteolytic cleavage vary from strain to strain of influenza virus. In some cases this variation is quite extensive—for example compare A0/Bel/42 and A2/Singapore/57 hemagglutinins in Fig. 4. However, the sequence analyses shown in Table 2 indicate that the amino-acid sequence at the cleavage site—the amino terminus of BHA2—does not vary from strain to strain and varies only slightly from type A to type B viruses. These findings suggest, therefore, that the observed differences in the molecular weights of the subunits of different strains are probably due to variation in their carbohydrate compositions and not to variation in the position of cleavage of the precursor molecules (17).

The sequence at the amino terminus of BHA2 shows two additional features of interest. (*1*) The first 10 residues are all uncharged and apart from the glycine residues are hydrophobic and (*2*) the sequence forms a pseudo-palindrome of seven residues centered on the isoleucine residue at position 6. This structure can be extended to nine residues if the conservative substitution of leucine for isoleucine at positions 2 and 10 is included. In the influenza B type virus BHA2 has two conservative amino-acid substitutions at positions 2 (phenylalanine for leucine) and position 10 (isoleucine for leucine).

The structural significance of palindromes in proteins is not clear, although the importance of palindromes as structural features of nucleic acids is currently being established (18, 19).

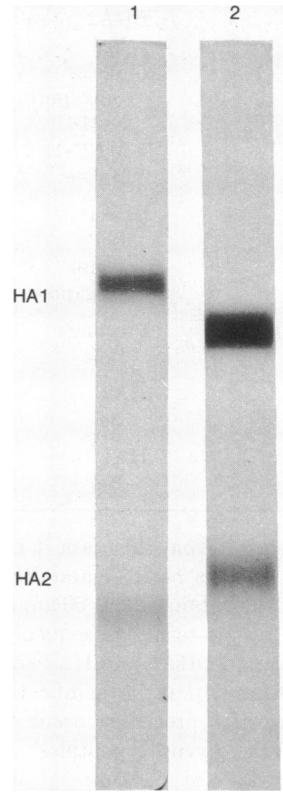

FIG. 4. Polyacrylamide gel electropherograms of the polypeptide components of different influenza viruses which demonstrate the apparent variation in molecular weight of the hemagglutinin subunits. Purified HA preparations were disrupted and their polypeptide composition was analyzed by polyacrylamide gel electrophoresis in sodium dodecyl sulfate as described in *Methods*. The hemagglutinin subunits are indicated as HA1 and HA2. (*1*) The HA of A0/Bel/42 (Ho N_1). (*2*) The HA of A2/Singapore/57 (H_2N_2). The nomenclature in parentheses is that proposed by the World Health Organization (21), for classification of these virusus.

It is possible in this case that this amino-acid sequence represents a specific recognition sequence for the enzyme that cleaves the precursor to generate the two subunits. It is also possible to construct a palindromic nucleic-acid sequence, which could code for the amino-acid palindrome.

Comparison of the Amino-Acid Sequences of the Hemagglutinins from Different Strains of Influenza. It is clear from the data presented in Table 2 that a high degree of sequence homology between the BHA2 subunits of the different A type viruses exists at the amino terminus. Similarly, conservation of sequence in this region is also seen between type A and type B virus hemagglutinins, in which the sequences of the first 20 residues of the BHA2 subunits are clearly homologous (unpublished).

In addition, inspection of the BHA1 amino-terminal sequences of the three type A viruses also reveals a high degree of homology (Table 2) with differences at positions 1 and 2.

The extent of sequence variations between strains will only be revealed by further sequence analysis.

Release of the Hemagglutinin from Virus Particles by Bromelain. Bromelain treatment of influenza viruses releases the protein BHA, which is made up of two subunits with molecular weights 58,000 (BHA1) and 21,000 (BHA2). The amino-terminal sequences of these subunits are identical to those of the corresponding subunits of the hemagglutinin isolated by detergent disruption of the virus—HA1 and HA2. In addition the amino-acid compositions of HA1 and BHA1 are almost identical, while HA2 and BHA2 show clear differences. This suggests that bromelain removes a carboxyl terminal peptide from the HA2 subunit and that this peptide is associated with the lipid membrane of the virus. The amino-acid composition of this peptide can be deduced from the differences in composition between the HA2 and BHA2 subunits and is shown in Table 1. Because of its association with the membrane lipid this "tail" peptide might be expected to be largely composed of hydrophobic amino acids. The results indicate that of the 50 residues, 14 may be charged, and 21 of the remainder can be regarded as hydrophobic (20). It is particularly striking that half of the serine residues of the HA2 subunit are located in the tail peptide.

The carboxyl terminus of BHA2 was found by carboxypeptidase digestion to be lysine. The carboxyl terminus of HA2 could not be determined by this technique, probably because of its insolubility in nondenaturing solvents—a property presumably related to the hydrophobicity of the subunit, which aggregates in 6 M guanidine. Comparative studies of the cyanogen bromide fragments of BHA2 and HA2 that will allow further characterization of the "tail" fragment obviously represent an important part of future analyses of the hemagglutinin molecule.

We thank Ann Gurnett and David Stevens for excellent assistance. M.D.W. is an Established Investigator of the American Heart Association.

1. Schulze, I. T. (1973) *Advan. Virus Res.* **18,** 1–55.
2. Laver, W. G. (1973) *Advan. Virus Res.* **18,** 57–103.
3. Skehel, J. J. & Schild, G. C. (1971) *Virology* **44,** 396–408.
4. Laver, W. G. (1971) *Virology* **45,** 275–288.
5. Lazarowitz, S. G., Compans, R. W. & Choppin, P. W. (1971) *Virology* **46,** 828–843.
6. Skehel, J. J. (1972) *Virology* **49,** 23–36.
7. Klenk, H. D., Scholtissek, C. & Rott, R. (1972) *Virology* **49,** 723–734.
8. Hay, A. J. (1974) *Virology,* **60,** 398–418.
9. Laver, W. G. (1964) *J. Mol. Biol.* **9,** 109–124.
10. Brand, C. M. & Skehel, J. J. (1972) *Nature New Biol.* **238,** 145–147.
11. Smyth, D. G., Stein, W. H. & Moore, S. (1963) *J. Biol. Chem.* **238,** 227–234.
12. Hirs, C. H. W. (1967) in *Methods in Enzymology,* ed. Hirs, C. H. W. (Academic Press, New York), Vol. XI, p. 197.
13. Edman, P. & Begg, G. (1967) *Eur. J. Biochem.* **1,** 80–91.
14. Pisano, J. J., Bronzert, T. J. & Brewer, H. B. (1972) *Anal. Biochem.* **45,** 43–59.
15. Smithies, O., Gibson, D., Fanning, E. M., Goodfliesch, R. M., Gilman, J. G. & Ballantyne, D. L. (1971) *Biochemistry* **10,** 4912–4921.
16. Shapiro, A. L., Vinuela, E. & Maizel, J. V. (1967) *Biochem. Biophys. Res. Commun.* **28,** 815–820.
17. Skehel, J. J. (1974) *Symp. Soc. Gen. Microbiol.* **XXIV,** 321–342.
18. Maniatis, T., Ptashne, M., Barrell, B. G. & Donelson, J. (1974) *Nature* **250,** 394–397.
19. Wilson, D. A. & Thomas, C. A., Jr. (1974) *J. Mol. Biol.* **84,** 115–144.
20. Tanford, C. J. (1962) *Amer. Chem. Soc.* **84,** 4240–4247.
21. *Bull. W.H.O.* (1971) **45,** 119–126.