Proc. Natl. Acad. Sci. USA
Vol. 74, No. 4, pp. 1530–1534, April 1977
Biochemistry

Partial amino acid sequence of two major component polypeptides of hepatitis B surface antigen

(structure of HBsAg/hepatitis/synthetic vaccine)

DARRELL L. PETERSON, ISABEL M. ROBERTS, AND GIRISH N. VYAS

Department of Laboratory Medicine and the Liver Center, University of California School of Medicine, San Francisco, California 94143

Communicated by Rudi Schmid, February 9, 1977

ABSTRACT **Determination of the amino acid sequence of the immunogenic polypeptides of hepatitis B surface antigen may not only permit molecular localization of the distinct determinants *a*, *d*, and *y* but may also lead to the synthesis of a hapten useful in prophylactic immunization against hepatitis B virus infection. For this purpose, purified monotypic hepatitis B surface antigen of adw subtype was resolved into equal amounts of two major polypeptides (22,000 and 28,000 daltons) and up to six other minor polypeptides by polyacrylamide gel electrophoresis. With the periodate staining reaction, only the 28,000-dalton polypeptide stained as a glycoprotein. Guinea pigs immunized with the 22,000-dalton polypeptide produced potent antisera against determinants *a* and *d*, but the 28,000-dalton glycoprotein did not induce a response. Both polypeptides isolated by preparative polyacrylamide gel electrophoresis showed amino acid composition identical with that of the intact antigen. For both polypeptides, hydrazinolysis gave Ile as the carboxy-terminus, and carboxypeptidase A digestion gave the same terminal sequence, Val-Tyr-Ile. Both peptides also yielded an identical sequence of amino acids in nine steps of Edman degradation—Met-Glu-Asn-Ile-Thr-Ser(Cys)-Gly-Phe-Leu. Our data suggest that hepatitis B surface antigen contains a single major immunogenic 22,000-dalton polypeptide component, part of which is modified by the addition of carbohydrate to give rise to the glycopeptide of apparent molecular weight 28,000.**

Hepatitis B surface antigen (HBsAg) is a group of morphologically heterogeneous, complex, macromolecular structures found in the serum of patients with hepatitis B virus infection and also in the serum of apparently healthy chronic carriers of hepatitis B virus. Electron microscopy reveals three forms of HBsAg: 20-nm spherical particles, tubular structures of variable length and a diameter of 20 nm, and 40-nm spherical "Dane" particles (1–5). Of these, the 20-nm particles account for the bulk of the HBsAg present. All of these particles share common antigenic determinants. The *a* determinant is a group-specific determinant found in all HBsAg. In addition, there are two sets of mutually exclusive determinants, *d*/*y* and *w*/*r*. Thus, the four possible phenotypes are adw, ayw, adr, and ayr (6).

Purified 20-nm HBsAg has a molecular weight of 2 to 4 × 10^6 (7, 8). Previous reports have shown it to be composed of lipid, carbohydrate, and protein (9, 10). By sodium dodecyl sulfate ($NaDodSO_4$)/polyacrylamide gel electrophoresis the protein components have been resolved into two to seven bands. Some proteins stain as glycoproteins with the periodate-Schiff staining reaction and others appear to lack carbohydrate (11–13). Elucidation of the structure of the antigenic determinants of HBsAg has been the major goal of our studies (12, 13). Such knowledge might allow the synthesis of a peptide of the same sequence which could serve as a hapten and, when coupled to an appropriate carrier, may serve as a vaccine for immunization against hepatitis B virus infection. In order to elucidate this structure, it is essential to determine the number of proteins that make up HBsAg and to isolate these components in pure form. Because most of the HBsAg is present as 20-nm particles and because the adw subtype is the predominant form found in the United States, this form has been used exclusively for the biochemical characterization of major peptides of HBsAg.

We report here the results of analytic and preparative-scale polyacrylamide gel electrophoresis resolving HBsAg/adw into two major components (polypeptide I and polypeptide II), immunogenicity of these components in guinea pigs, and their amino acid compositions including their amino- and carboxy-terminal amino acid sequences.

Abbreviations: HBsAg, hepatitis B surface antigen; $NaDodSO_4$, sodium dodecyl sulfate.

MATERIALS AND METHODS

Purification of HBsAg. HBsAg was purified from the plasma of an apparently healthy chronic carrier of HBsAg by a combination of isopycnic banding and differential rate sedimentation through cesium chloride gradients according to the method described by Bond and Hall (14). The material was checked for purity by Ouchterlony double-diffusion analysis with antiserum against normal human serum.

Analytical Electrophoresis. $NaDodSO_4$/polyacrylamide gel electrophoresis was performed with the Biophore gel system (Bio-Rad Laboratories, Richmond, CA). Twelve percent acrylamide gels were pre-electrophoresed with the running buffer [0.025 M Tris/0.025 M acetic acid/0.1% (wt/vol) $NaDodSO_4$], pH 6.6. Pre-electrophoresis was performed at room temperature with a current of 3 mA per gel. The samples were made 1% (wt/vol) in $NaDodSO_4$ and 1% (wt/vol) in 2-mercaptoethanol and heated for 1 hr at 60°; then they were mixed with bromophenol blue as a tracking dye and a 50-μl aliquot was applied to each gel. Electrophoresis was performed at 8 mA per gel until the tracking dye was approximately 1 cm from the end. The gels were removed and stained for protein with Coomassie blue or for carbohydrate with the periodate-Schiff stain according to the methods of Fairbanks *et al.* (15).

Preparative Electrophoresis. Large-scale separation of HBsAg into its component protein was performed by preparative electrophoresis with a custom-built apparatus (Fig. 1). The apparatus was adapted from the method of Ryan *et al.* (16) with the following improvisations. The gel column was 2.5 cm in diameter and 15 cm in length. The acrylamide gel and the electrophoresis buffer were formulated as described by Stephens (17). The gel was cast by pouring approximately 40 ml of 7.5% (wt/vol) acrylamide into the column; then 40% (wt/vol) sucrose was injected into the elution tubing to create a space (1–2 mm) between the gel and the dialysis membrane (Spectrapor-3000 membrane). A 3% (wt/vol) stacking gel was then carefully layered over the 7.5% gel, followed by a buffer layer. The gel cured in about 30 min. However, it was allowed to stand overnight before use because this apparently prevented its being

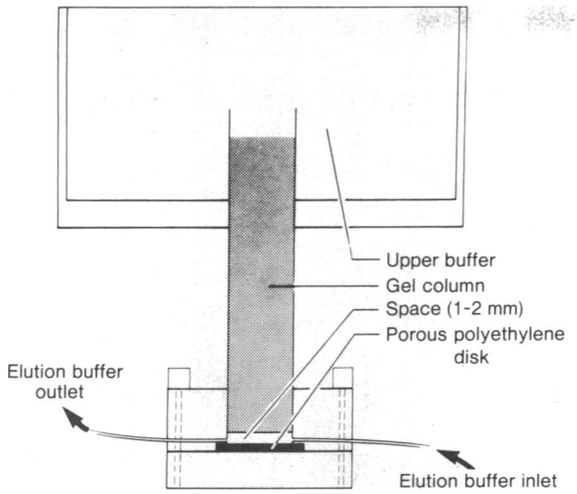

FIG. 1. Diagram of custom-made preparative polyacrylamide gel electrophoresis system used for resolution and elution of two major component polypeptides of HBsAg. Elution buffer was pumped in with ISCO Dialagrad pump, and fractions from the outlet were collected after the effluent had passed through a recording spectrophotometer.

displaced later by the pumping of the elution buffer. With bromophenol blue as a tracking dye, the gel was pre-electrophoresed before use. Approximately 10 mg of purified HBsAg in 2 ml of water was dialyzed extensively against 0.025 M Tris·glycine, pH 9.5, containing 1% (wt/vol) $NaDodSO_4$. The solution was then made 1% (wt/vol) in 2-mercaptoethanol and heated at 60° for 1 hr. The sample in 1–2 ml of 0.025 M Tris·glycine, pH 8.4, containing 20% (wt/vol) sucrose was loaded on the gel and electrophoresis was performed at room temperature at 25 mA. Buffer was pumped through the elution chamber at the rate of 16 ml/hr, and fractions were collected at 6-min intervals. The elution was monitored with a Uvicord spectrophotometer at 280 nm. Peaks were pooled and concentrated by vacuum dialysis. $NaDodSO_4$ was removed by the method of Tuszynski and Warren (18).

Amino Acid Analysis. Aliquots (150 μg) of protein were hydrolyzed in sealed, evacuated tubes containing 1.0 ml of constant-boiling HCl for 24–96 hr at 110°. Amino acid analyses were performed with the single-column system of a Durrum D-500 amino acid analyzer.

Edman Degradation. Amino-terminal sequence determinations were performed with a Beckman 890C automatic sequencer. The sample in 0.5 ml of water was applied with the standard application technique, followed by use of the fast protein Quadrol program 972172-C (Beckman, Palo Alto, CA). The released thiazolinone was identified by one or a combination of the following: (*i*) conversion to the phenylthiohydantoin by treatment with 1.0 M HCl for 10 min at 80° (21) followed by gas chromatography on 10% SP-400; (*ii*) amino acid analysis after hydrolysis by either HI or NaOH/dithionite (22); (*iii*) thin-layer chromatography of the phenylthiohydantoins on silica gel, with chloroform/methanol, 80:20 (vol/vol) as the developing solvent (23) for distinguishing amides from corresponding acids.

Hydrazinolysis. For determination of the carboxy-terminal amino acid, hydrazinolysis was performed essentially according to the method of Fraenkel-Conrat and Tsung (24). Samples of protein (350 μg) were freeze-dried in Pyrex tubes. The tubes were constricted, 0.2 ml of anhydrous hydrazine was added, and the tube was sealed under vacuum. Hydrazinolysis was allowed to proceed for 24 hr at 80°. The tubes were opened and the hydrazine was removed under vacuum. The dried material was dissolved in 50 μl of 0.2 M sodium citrate, pH 2.2, and centrifuged at 14,000 × *g* for 5 min to remove any particulate material. The entire sample was applied to the Durrum amino acid analyzer. Prolonged regeneration for 1 hr with 0.2 M NaOH between samples was necessary to prevent large base-line shifts on consecutive samples.

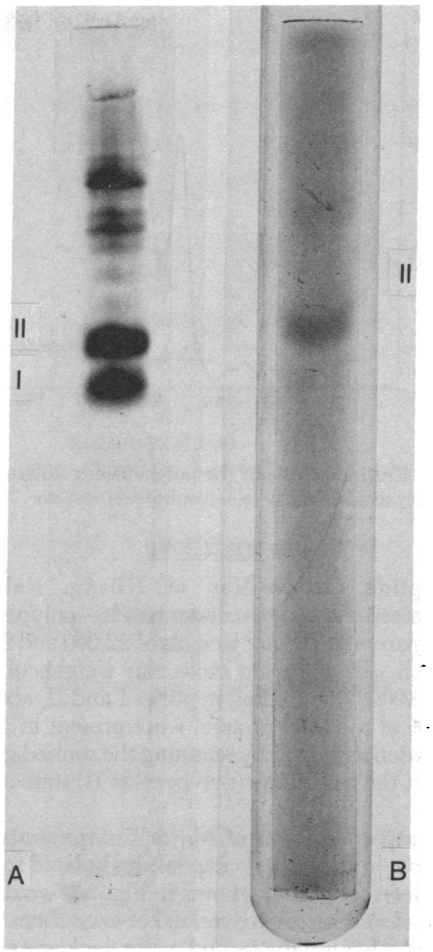

FIG. 2. HBsAg resolved by analytical $NaDodSO_4$/polyacrylamide gel electrophoresis and stained by Coomasie blue showed two major components, polypeptides I and II (*A*). Identical gel stained with periodate-Schiff stain (*B*) showed that polypeptide II is a glycoprotein.

Carboxypeptidase Digestion. Carboxypeptidase A digestion was performed according to the method of Ambler (25). Aliquots (150 μg) of peptides in 50 μl of 0.2 M *N*-ethylmorpholine acetate, pH 8.5, were treated with 1 μg of carboxypeptidase A for various times at room temperature. The pH was lowered to pH 2.2 with 1.0 M HCl and the sample was applied to the amino acid analyzer.

Production of Antisera. A solution of 25–50 μg of protein in 0.2 ml of saline was mixed with an equal volume of Freund's complete adjuvant and injected subcutaneously into several guinea pigs. After 3 weeks the blood was collected by heart puncture. The serum was separated and stored frozen.

Serological Assays. HBsAg and antisera to it were tested with the commercially available AUSRIA and AUSAB radioimmunoassay kits (Abbott Laboratories, Chicago). Serologic specificity was determined by neutralization of antisera with equal amounts of plasma containing high-titer HBsAg/adw or HBsAg/ayw (26).

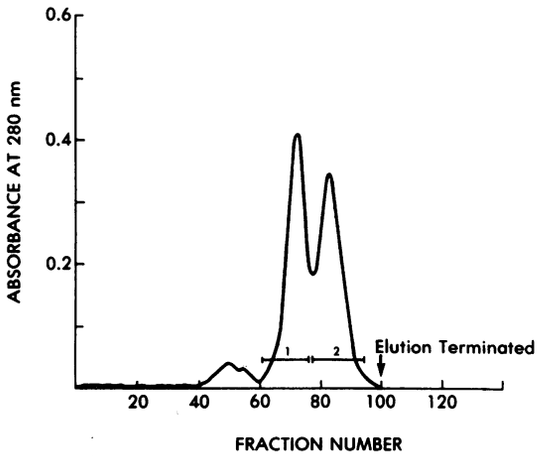

FIG. 3. Elution profile of the polypeptides isolated in the preparative polyacrylamide gel electrophoresis system.

RESULTS

Polypeptide Composition of HBsAg. $NaDodSO_4$/gel electrophoresis showed two major bands—polypeptides I and II with apparent molecular weights of 22,000 and 28,000—and other bands with apparent molecular weights of 16,000 and 40,000–90,000 (Fig. 2). Polypeptides I and II, accounting for about 75% of the total protein, were present in nearly equal concentration as judged by scanning the stained gels. Of these two bands, the larger one (polypeptide II) stained as a glycoprotein.

Preparative Isolation of Major Components of HBsAg. The elution profile of the polypeptides isolated from preparative gel electrophoresis is shown in Fig. 3. Two distinct peaks were obtained with some overlap between them. Elution was continued only until the second major peak was eluted. No attempt was made to recover the other components. The fractions were pooled as shown, concentrated by vacuum dialysis to approximately 2 ml, and then dialyzed extensively against water. By analytical gel electrophoresis, polypeptide I was shown to be essentially pure; only minor contamination with polypeptide II was detected. The polypeptide II fraction was contaminated with polypeptide I to an extent of about 20%. Therefore, polypeptide II was further purified by a second passage through the preparative electrophoresis system and was obtained in essentially pure form (Fig. 4).

Table 1. Amino acid composition of HBsAg and its major protein components

Amino acid	Mole %		
	HBsAg	Pop I*	Pop II*
Aspartic acid	5.0	5.6	5.7
Threonine	7.8	8.4	8.6
Serine	11.7	11.9	13.0
Glutamic acid	5.0	5.8	5.7
Proline	12.8	12.5	10.6
Glycine	7.2	6.1	6.6
Alanine	3.4	3.5	3.3
Half cystine†	6.8	7.2	7.0
Valine	4.7	4.1	3.9
Methionine	3.2	3.0	3.1
Isoleucine	5.6	5.7	5.1
Leucine	11.9	11.2	10.9
Tyrosine	2.2	2.4	2.3
Phenylalanine	5.1	4.9	5.1
Lysine	2.2	2.2	2.0
Histidine	0.6	0.7	0.6
Arginine	2.4	2.2	2.1

* Polypeptides I and II.
† As cysteic acid.

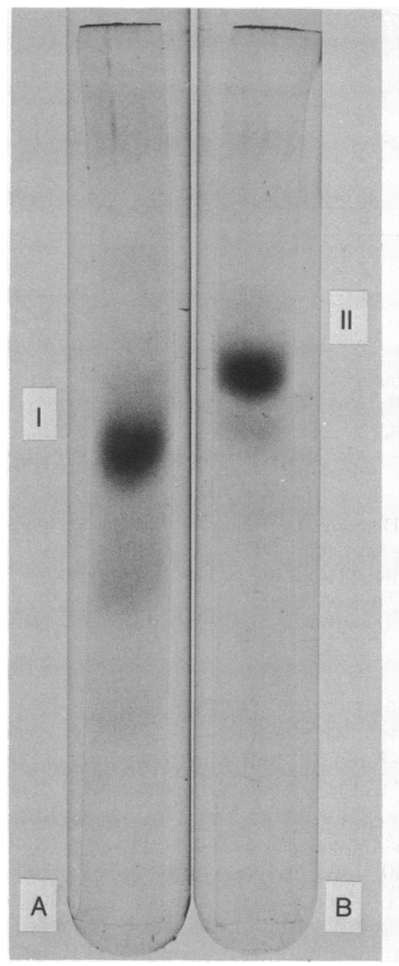

FIG. 4. Electrophorograms showing purity of polypeptides I and II used in all biochemical and immunological work.

Amino Acid Analysis of Purified Polypeptides I and II. Amino acid analysis of the purified polypeptides I and II along with that of intact HBsAg is shown in Table 1. They had essentially identical amino acid composition. The most noticeable features of the amino acid composition are a large amount of proline and half cystine and a large percentage of hydrophobic residues.

Amino-Terminal Sequence of HBsAg Peptides. The results of automatic Edman degradation performed on 1–2 mg of each of the major polypeptides are shown in Table 2. They were identical at every step. Only step 6 was uncertain: HI hydrolysis gave alanine and NaOH/dithionite gave nothing, consistent with either serine or half-cystine.

Hydrazinolysis. Hydrazinolysis of either polypeptide I or II liberated isoleucine as a free amino acid in a yield of 30–40%. Only trace amounts of other amino acid were released. This is in the expected range for the yield of carboxy-terminal isoleucine (27).

Carboxypeptidase A Digestion. Carboxypeptidase A digestion of either polypeptide I or II released isoleucine, tyrosine, and valine, with only traces of other residues. Based on the kinetics of the release of the amino acids, the likely carboxy-terminal sequence is Val-Tyr-Ile.

Table 2. Amino acid sequence of major polypeptide components of HBsAg

Step	Amino acid	Yield* Pop I	Yield* Pop II	Method of identification†
1	Met	100	100	NaOH
2	Glu	100	100	HI; TLC
3	Asn	83	88	HI; TLC
4	Ile	75	79	HI
5	Thr	62	65	HI
6	Ser or Cys	50	53	NaOH; HI
7	Gly	72	76	HI
8	Phe	61	65	HI
9	Leu	70	71	HI

* Yield is defined as the amount of amino acid obtained relative to the amount of amino acid obtained in step 1. Pop I = polypeptide I; Pop II = polypeptide II.

† Amino acids were identified by amino acid analysis after hydrolysis of the phenylthiazolinone by HI (indicated by HI) or by NaOH (indicated by NaOH) or they were identified by thin-layer chromatography (TLC).

Production of Antibodies to Components of HBsAg. When injected into guinea pigs, only polypeptide I elicited an antibody response. The antibodies were completely neutralized by a single absorption with HBsAg/adw. However, neutralization with HBsAg/ayw only decreased the measured radioactivity in the AUSAB test by about half relative to the control neutralization with normal human serum (20,000 cpm). Additional neutralization with HBsAg/ayw did not influence the residual antibody to HBsAg/*d*. Hence, we concluded that polypeptide I produced antibodies against HBsAg/*a* and HBsAg/*d* whereas polypeptide II did not produce any antibody response. Two additional experiments with booster immunization gave identical results.

DISCUSSION

The amino acid composition of HBsAg reported here is essentially the same as previously reported (7, 12). The high content of proline, cysteine, and hydrophobic amino acids may be largely responsible for the high degree of resistance to physical denaturation and proteolytic digestion.

The identical amino-terminal and carboxy-terminal sequences of amino acids in polypeptides I and II indicate that HBsAg/adw consists of a single major polypeptide chain or two homologous polypeptide chains that differ only in limited areas of their structure rather than being two completely different proteins as has been generally assumed (10–13). The apparent difference in the molecular weights of the two components reflects the contribution of the carbohydrate moiety of the glycoprotein (polypeptide II). This is consistent with the general observation that glycoproteins migrate at a slower rate on Na$DodSO_4$/gel electrophoresis than would be expected based only on their protein molecular weight (28). Therefore, it is likely that HBsAg consists of a single major polypeptide chain, part of which has been modified by the addition of carbohydrate. This likelihood is particularly attractive in view of the small size of the hepatitis B virus genome which is reported to be a 1.6×10^6-dalton double-stranded DNA (29). Such a genome can code for only 700–800 amino acids and could not make all of the apparently different proteins of HBsAg resolved by gel electrophoresis. These results are analogous to those reported for the major envelope glycoproteins of Rauscher murine leukemia virus, which have apparent molecular weights of 48,000 and 62,000 but apparently identical amino acid sequences (30).

The nature of the minor components of HBsAg has not been determined. However, it is possible that these higher molecular weight components are aggregates of the major peptide and glycopeptide. Such aggregates resistant to treatment with Na$DodSO_4$, mercaptethanol, and heating have been reported for other proteins and glycoproteins (31, 32).

Consistent with the findings of other investigators (10, 33), the nonglycosylated polypeptide I from preparative gel electrophoresis was found to be immunogenic in guinea pigs, giving rise to antibodies against HBsAg/*a* and HBsAg/*d*. Therefore, this peptide contains antigenic determinants found on the intact HBsAg. However, guinea pigs immunized by us with glycosylated polypeptide II, obtained by purification with two successive passages through the preparative gel electrophoresis system, showed no antibody response. The discrepancies in immunogenicity of glycosylated peptides reported by two groups (10, 33) could be due to contaminating polypeptide I in polypeptide II cut out of gels and used in immunization experiments. This is a more likely explanation of differences among the three observations (including our findings reported here).

The carbohydrate composition of intact HBsAg is 45.8 μg of *N*-acetylglucosamine, 12.1 μg of mannose, 8.3 μg of galactose, and 8.5 μg of sialic acid per mg of protein (H. Shirachi and N. Ishida, personal communication). Because these carbohydrates commonly bind to asparagine and threonine residues, it is conceivable that steric hindrance is operative in the failure of the glycopeptide (polypeptide II) to elicit an antibody response in immunized guinea pigs. Further, it has been shown that, after reduction and alkylation, otherwise immunogenic HBsAg undergoes conformational changes and fails to elicit antibodies in immunized guinea pigs (34). If, indeed, the sixth step in the sequence is a cysteine residue, it is conceivable that the amino-terminal sequence is conformationally rendered in an immunogenic form. Such a hypothesis can be tested with synthesis of the amino-terminal sequence and determination of its immunogenicity in guinea pigs.

Complete information about the amino acid composition of HBsAg and its components, amino acid sequence in the haptenic determinant *a* of HBsAg, and synthesis of such a haptenic peptide could achieve the goal of a synthetic vaccine against hepatitis B virus infection.

We thank Jennifer Harford and Rivka Zisling for technical assistance and Warren Eveleth for editorial assistance. This work was supported in part by Contract N01-3-3000 from the National Heart, Lung, and Blood Institute and Liver Center Grant P-50-AM18520.

1. Bayer, M. E., Blumberg, B. S. & Werner, B. (1968), *Nature* **218**, 1057–1059.
2. Sutnick, A. I., London, W. T. & Blumberg, B. S. (1969) *Am. J. Dig. Dis.* **14**, 189–194.
3. Almeida, J. D., Zuckerman, A. J., Taylor, P. E. Waterson, A. P. (1969) *Microbios* **1**, 117–123.
4. Hirschman, R. J., Shulman, N. R., Barker, L. F. & Smith, K. O. (1969) *J. Am. Med. Assoc.* **208**, 1667–1670.
5. Dane, D. S., Cameron, C. H. & Briggs, M. (1970) *Lancet* **i**, 695–698.
6. Le Bouvier, G. L. (1973) *Ann. Int. Med.* **79**, 894–896.
7. Dreesman, G. R., Hollinger, F. B., Suriano, J. R., Fujioka, R. S., Brunschwig, J. P. & Melnick, J. L. (1972) *J. Virol.* **10**, 469–476.
8. Kim, C. Y. & Tilles, J. G. (1973) *J. Clin. Invest.* **52**, 1176–1186.

9. Steiner, S., Huebner, M. T. & Dreesman, G. R. (1974) *J. Virol.* **14,** 572–577.
10. Dreesman, G. R., Hollinger, F. B. & Melnick, J. L. (1975) *Am. J. Med. Sci.* **270,** 123–129.
11. Gerin, J. L., Holland, P. V. & Purcell, R. H. (1971) *J. Virol.* **7,** 569–576.
12. Vyas, G. N., Williams, E. W., Klaus, G. G. B. & Bond, H. E. (1972) *J. Immunol.* **108,** 1114–1118.
13. Rao, K. R. & Vyas, G. N. (1976) *J. Biol. Stand.* **4,** 295–304.
14. Bond, H. E. & Hall, W. T. (1972) *J. Infect. Dis.* **125,** 263–268.
15. Fairbanks, G., Steck, T. L. & Wallach, D. F. H. (1971) *Biochemistry* **10,** 2606–2617.
16. Ryan, T. E., Woods, G. M., Kirkpatrick, F. H. & Shamoo, A. E. (1976) *Anal. Biochem.* **72,** 353–358.
17. Stephens, R. E. (1975) *Anal. Biochem.* **65,** 369–379.
18. Tuszynski, G. P. & Warren, L. (1975) *Anal. Biochem.* **67,** 55–65.
19. Hugli, T. E. & Moore, S. (1972) *J. Biol. Chem.* **247,** 2828–2834.
20. Hirs, C. H. W. (1967) in *Methods in Enzymology,* ed. Hirs, C. H. W. (Academic Press, New York), Vol. 11, pp. 59–62.
21. Pisano, J. J. & Bonzert, T. J. (1969) *J. Biol. Chem.* **244,** 5597–5607.
22. Smithies, O., Gibson, D., Fanning, E. M., Goodfleish, R. M., Ballantyne, D. L. & Gilman, J. G. (1971) *Biochemistry* **10,** 4912–4921.
23. Pataki, G. (1966) *Chromatography* **21,** 133–135.
24. Fraenkel-Conrat, H. & Tsung, C. M. (1967) in *Methods in Enzymology,* ed. Hirs, C. H. W. (Academic Press, New York), Vol. 11, pp. 151–155.
25. Ambler, R. P. (1967) in *Methods in Enzymology,* ed. Hirs, C. H. W. (Academic Press, New York), Vol. 11, pp. 155–166.
26. Vyas, G. N., Adelberg, S. G. & Perkins, H. A. (1973) *Science* **182,** 1368–1369.
27. Locker, R. H. (1954) *Biochim. Biophys. Acta* **14,** 533.
28. Weber, K., Pringle, J. R. & Osborn, M. (1972) in *Methods in Enzymology,* ed. Hirs, C. H. W. (Academic Press, New York), Vol. 26, pp. 3–27.
29. Robinson, W. S., Clayton, D. A. & Greenman, R. L. (1974) *J. Virol.* **14,** 384–391.
30. Marquardt, H. & Oroszlan, S. (1974) *Fed. Proc.* **35,** 1610.
31. Bensodoun, A., Ehnholm, C., Steinberg, D. & Brown, W. V. (1974) *J. Biol. Chem.* **249,** 2220–2227.
32. Slutzky, G. M. & Ji, T. H. (1974) *Biochim. Biophys. Acta* **373,** 337–346.
33. Shih, J. W. & Gerin, J. L. (1975) *J. Immunol.* **115,** 634–639.
34. Vyas, G. N., Rao, K. R. & Ibrahim, A. B. (1972) *Science* **178,** 1300–1301.