*Proc. Natl. Acad. Sci. USA*
Vol. 74, No. 12, pp. 5421–5425, December 1977
Biochemistry

# High resolution two-dimensional electrophoresis of human plasma proteins

(genetic variants/polymorphism/genetic screening)

LEIGH ANDERSON AND NORMAN G. ANDERSON

Molecular Anatomy Program, Division of Biological and Medical Research, Argonne National Laboratory, Argonne, Illinois 60439

*Communicated by James V. Neel, September 22, 1977*

ABSTRACT **The two-dimensional electrophoretic technique of O'Farrell has been adapted to the analysis of human plasma proteins, and 30 polypeptides have been identified in the pattern produced. Genetic variants involving charge (isoelectric point) or size (molecular weight in the presence of sodium dodecyl sulfate) changes should be routinely detectable in at least 20 proteins at once, facilitating studies of human mutation rates.**

Various forms of electrophoresis have for some time been the best tools for analyzing complex protein mixtures, especially blood plasma. In 1937, Tiselius separated serum globulin into $\alpha,\beta$, and $\gamma$ components (1). Since then a steady increase in resolution has taken place due to the introduction of acrylamide gels (2), stacking systems (3, 4), isoelectric focusing (5), and more recently a variety of two-dimensional systems exploiting either a pair of electrophoretic methods (6–8) or electrophoresis and immunochemical specificity (9). The later methods can resolve perhaps 60 protein components in plasma (10) at the level of undissociated complexes, but the amount of detailed biophysical information available from such separations is limited by the fact that most undissociated plasma proteins are very large (40,000 to several million daltons). An alternative technique, recently developed by O'Farrell (11) and others (12–14) and adapted by us to allow convenient analysis of large numbers of samples in parallel (15, 16), separates proteins in two dimensions under dissociating conditions [isoelectric focusing in 9 M urea/2% Nonidet P-40 followed by electrophoresis in the presence of sodium dodecyl sulfate ($NaDodSO_4$)]. The system has three principal advantages over earlier two-dimensional methods: (*i*) it deals almost exclusively with individual gene products, simplifying genetic analysis of variants; (*ii*) it combines two different very high resolution separations yielding a combined resolution better than 0.1 charge unit and 1000 daltons for an average 50,000-dalton protein; and (*iii*) it may provide the best available way to investigate the *in vivo* characteristics of plasma proteins, since samples can be run that have been fully denatured in $NaDodSO_4$ immediately upon acquisition (17). By exploiting these advantages and using protein staining instead of autoradiography for detection, two areas of plasma protein study may be directly attacked: the search for genetic variants and the study of glycoprotein heterogeneity (N. L. Anderson and N. G. Anderson, unpublished observations). In this paper we concern ourselves primarily with the genetic aspect of the work. Here the potential advantages in monitoring 20–50 loci by one technique are considerable (18). Although two-dimensional electrophoresis is inherently more time-consuming than most one-dimensional techniques, we have constructed simple, inexpensive equipment (15, 16) that allows three people to run 80 or more samples per day routinely. By further mechanizing the running of large numbers of gels, it may be possible to examine plasma from some reasonable fraction of all newborns. Significant increases in mutation rate, due to environmental or other factors, could then be detected before major damage occurred, and important clinical information might be obtained as well.

Our program is aimed at cataloging human gene products for the purpose of exploring differentiation, carcinogenesis, and related questions. We have begun this effort by adapting the method of O'Farrell to the analysis of large numbers of human plasma samples in parallel (15, 16), by identifying a useful set of gene products in the pattern, and by undertaking initial characterizations of some known genetic polymorphisms and variants.

## MATERIALS AND METHODS

**Purified Proteins and Specific Immunoprecipitates Used for Identification.** As shown in Table 1, most of the 30 polypeptides identified in human plasma were obtained as purified proteins from others and also prepared by immunoprecipitation from fresh serum with commercially available specific antisera (Behring Diagnostics or Miles Laboratories). Immunoprecipitates were made by the following general procedure: 100 $\mu$l of antiserum in a 400-$\mu$l Beckman 152 microfuge tube was combined with 15, 30, and 60 $\mu$l of human serum on three successive days. The mixture was stored throughout at 4° and centrifuged in the microfuge for 1 min before each addition and before collection on the fourth day. The resulting precipitate was washed twice in cold 0.1 M phosphate buffer (pH 7). Immunoprecipitates made by reacting antisera with heparinized plasma contained may contaminants and were not used here. Hemoglobin and actin were identified on the basis of their abundance in erythrocyte lysates and platelets, respectively.

**Treatment of Samples.** Denaturation in $NaDodSO_4$ and mercaptoethanol has been used here in hopes of abolishing any protein-modifying enzymatic activity at the outset and in order to separate tightly bound complexes. It is shown in this and other work (17) that the added $NaDodSO_4$ is removed from proteins during isoelectric focusing in urea/Nonidet P-40 (the $NaDodSO_4$ focuses alone at the extreme acid end of the first dimension gel producing a small swollen region which identifies that end). The following procedure was used to prepare samples of plasma and purified proteins in solution: a 10-$\mu$l sample was mixed with 20 $\mu$l of 2% $NaDodSO_4$/5% 2-mercaptoethanol/10% glycerol in a 400-$\mu$l microfuge tube, and the mixture was heated for 5 min in a 95° heating block. When cool, all or part of this sample could be applied to the isoelectric focusing gel; however, any more than the stated quantity of $NaDodSO_4$

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: $NaDodSO_4$, sodium dodecyl sulfate; Ig, immunoglobulin.

Table 1. Identification of plasma proteins

| Protein | Method of identification |
|---|---|
| $\alpha_2$ Macroglobulin | Purified protein[a], immunoprecipitate[b] |
| Ceruloplasmin | Immunoprecipitate[b] |
| Plasminogen[c] | Immunoprecipitate[b], removal by urokinase[d] |
| C3 activator | Purified protein[e], immunoprecipitate[b] |
| $\alpha_1$B glycoprotein | Purified protein[e], immunoprecipitate[b] |
| Hemopexin | Purified protein[e,f], immunoprecipitate[b] |
| Albumin | Purified protein[e,g] |
| Transferrin | Purified protein,[e,h,i] immunoprecipitate[b] |
| Fibrinogen[c] | Immunoprecipitate[b], removal by clotting |
| $\alpha_1$ Antichymotrypsin | Immunoprecipitate[b] |
| $\alpha_2$HS glycoprotein | Purified protein[e], immunoprecipitate[b] |
| $\alpha_1$ Antitrypsin | Purified protein[e], immunoprecipitate[b] |
| Antithrombin III | Purified protein[e], immunoprecipitate[b] |
| Gc-globulin | Purified protein[e], immunoprecipitate[b] |
| IgG | $\kappa$, $\lambda$, $\gamma$-Specific immunoprecipitates[b] |
| $\alpha_1$ Acid glycoprotein | Purified protein[e,j] |
| Haptoglobin[c] | Purified protein[e], immunoprecipitate[b,k], absence in cord blood |
| Actin | Molecular weight, presence in platelets as $\beta$, $\gamma$ forms |
| Arginine-rich lipoprotein | Purified protein from rabbit[l], presence in human low density lipoprotein[l,m] |
| Apo A-I lipoprotein | Purified protein[l], presence in high density lipoprotein[l,m] |
| Prealbumin | Purified protein[e], immunoprecipitate[b] |
| Hemoglobin[c] | Erythrocyte lysate, coimmunoprecipitation with haptoglobin |
| Apo A-II lipoprotein | Molecular weight, presence in high density lipoprotein[l,m] |
| Unidentified lipoproteins (perhaps Apo-Cs) | Presence in low density lipoprotein[l,m] |

[a] Red Cross, USA.
[b] Behring Diagnostics (American Hoechst Corp., Somerville, NJ) antiserum, commercially available.
[c] Chains assigned by known charge and molecular weight differences.
[d] G. Barlow, Abbott Labs, Chicago, IL.
[e] Behringwerke, Marburg, W. Germany.
[f] P. Alsen, Albert Einstein College of Medicine, New York, NY.
[g] Worthington Biochemical Corp., Freehold, NJ.
[h] C. Reimer, Center for Disease Control, Atlanta, GA.
[i] R. King, Letterman Army Institute for Research, San Francisco, CA.
[j] C. J. van Oss, State University of New York, Buffalo, NY.
[k] Miles-Yeda, Ltd. (Miles Laboratories, Inc., Kankakee, IL) antiserum, commercially available.
[l] V. and B. Shore, Lawrence Livermore Laboratory, Livermore, CA.
[m] B. A. Bonner, Argonne National Laboratory, Argonne, IL.

caused streaking. Immunoprecipitate pellets were solubilized directly in 20 $\mu$l of $NaDodSO_4$/2-mercaptoethanol/glycerol, heated as for liquid samples.

**Two-Dimensional Electrophoresis.** The basic method has been described in great detail by O'Farrell (11). Our modifications, apart from the use of $NaDodSO_4$ denaturation (17), involve the use of labor-saving devices for simultaneously casting and running 20 isoelectric focusing gels (the ISO apparatus and procedures described in ref. 15) or 10 gradient $NaDodSO_4$ slab gels (the DALT apparatus and procedures described in ref. 16). Ampholines with wide pH ranges (LKB 3.5–10 or Brinkmann 2–11) were used throughout and focusing was allowed to proceed at 400–700 V for 20 hr, with or without prefocusing before sample loading. Slab gels were 10–20% gradients of acrylamide (including 2.7% bisacrylamide crosslinker). Proteins were visualized by staining slab gels overnight in 50% methanol/12% acetic acid/0.2% Coomassie Brilliant Blue R250, and then destaining in several changes of 10% ethanol/5% acetic acid (the last with a piece of wool felt to absorb residual stain). Photographs were taken through an interference filter (5500 Å narrow band; Baird-Atomic, Inc.) with 4 × 5 inch Kodak Contrast Process Pan film. Calibration with Behring Diagnostics standardized human transferrin indicates that a spot containing less than 50 ng of protein is visible. All gels are shown with acid end to the left and $NaDodSO_4$ electrophoresis proceeding downward (i.e., sample origin at upper right).

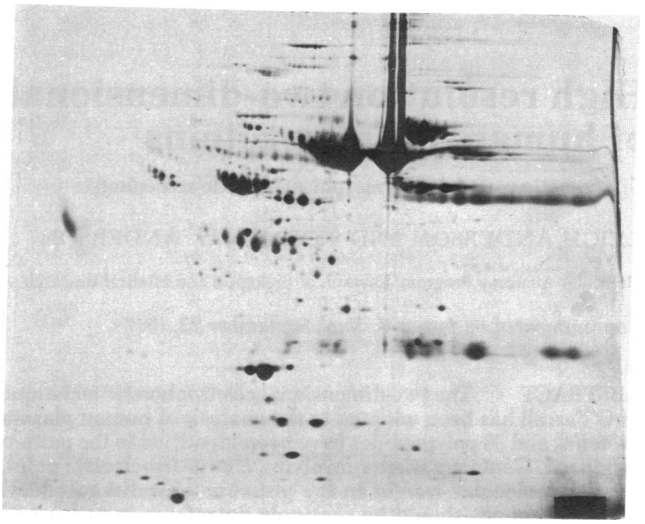

FIG. 1. Two-dimensional gel of human plasma proteins. The sample was 10 $\mu$l of fresh heparinized plasma denatured in $NaDodSO_4$/mercaptoethanol.

## RESULTS

### Identification of known plasma proteins in the two-dimensional pattern

A typical two-dimensional plasma protein pattern is shown in Fig. 1. Massive overloading of albumin is necessary to visualize the less abundant proteins, so that such gels often have vertical streaks near the site of albumin's isoelectric point. Most of the other proteins larger than immunoglobulin (Ig) light chains appear as groups of spots due to carbohydrate heterogeneity (principally neuraminidase-removable sialic acid). Thus the 300 or so spots visible in the pattern comprise perhaps 75–100 polypeptides.

The identities of various proteins are established by running pure preparations or immunoprecipitates (made by reacting specific antisera with whole serum) alone and combined with a small amount of plasma to facilitate position identification relative to other proteins (Fig. 2). Thirty identifications have so far been made, and these are shown diagrammatically in Fig. 3. In almost every case, the molecular weight in $NaDodSO_4$, determined by mobility in the vertical direction, agrees closely with published values; exceptions are ceruloplasmin (which exists in two main forms of about 90,000 daltons, but for which no generally accepted value is available) and apo A-I lipoprotein, which, according to sequence and other studies (21), should appear at 28,000 daltons instead of less than 23,500 (level of Ig light chains), as observed. Although a number of known proteins

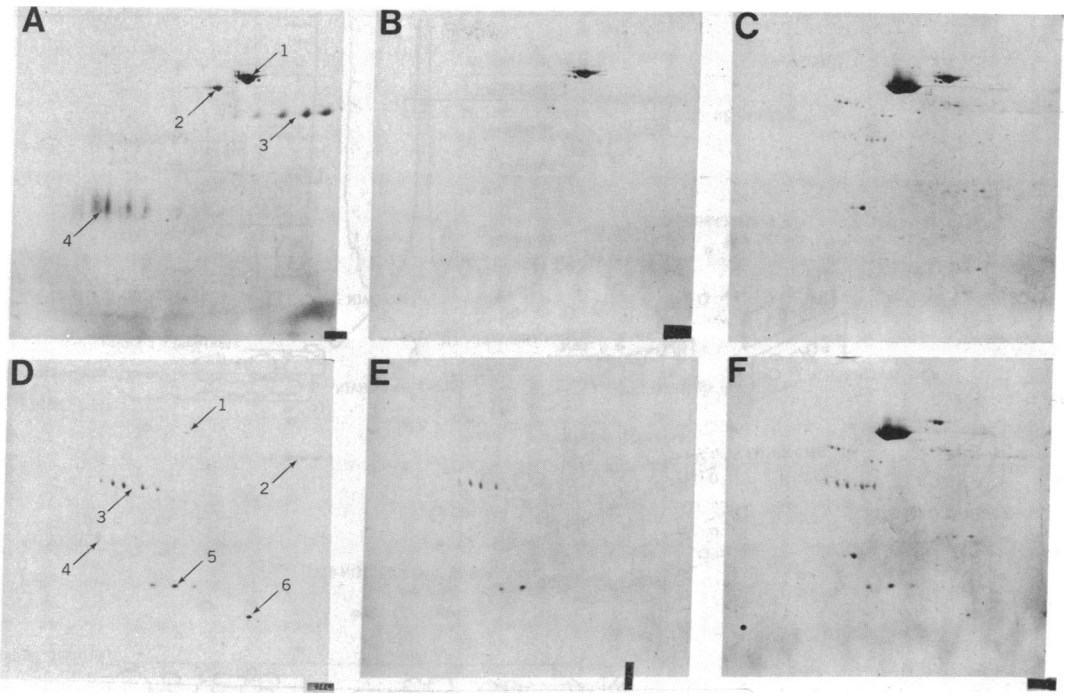


FIG. 2. Two-dimensional gels illustrating methods used to identify individual proteins. (*A*) Immunoprecipitated transferrin; (*B*) Behring purified transferrin; (*C*) purified transferrin mixed with 0.7 μl of plasma; (*D*) immunoprecipitated haptoglobin; (*E*) Behring purified haptoglobin; (*F*) purified haptoglobin mixed with 0.7 μl of plasma. Arrows in *A* indicate: (*1*) transferrin, (*2*) albumin contaminating the immunoprecipitate, (*3*) heavy chains of the precipitating IgG, and (*4*) IgG light chains. Arrows in *D* indicate: (*1*) contaminating albumin, (*2*) heavy chains of the precipitating IgG, (*3*) haptoglobin β-chain, (*4*) light chains of the precipitating IgG, (*5*) haptoglobin $\alpha^2$-chain, and (*6*) hemoglobin β-chain which coprecipitates with the haptoglobin to which it is bound in plasma. The haptoglobin β-chain (*D3*) consists of six or seven main forms (the arc indicated by the arrow) differing by single charges; slightly below and to the right is a faint arc of three spots corresponding to a form of the β-chain cleaved *in vivo* (a few percent of the total).

remain to be identified, it seems likely that there are in the pattern many hitherto undetected species.

If one accepts sample preparation by quick $NaDodSO_4$/mercaptoethanol denaturation as a valid method of preserving the nondisulfide covalent structure of proteins for subsequent analysis, then this technique should provide an accurate record of the molecular forms present *in vivo*. We believe that this may indeed be the only way to obtain such a record, since fractionation of undenatured plasma must upset the equilibrium of modifying enzymes, macromolecular substrates, and inhibitors present in the circulation, and lead to production of nonphysiological protein forms. In our experience, immunoprecipitated proteins almost always correspond better than classically purified proteins to spots in the fresh plasma pattern. The instances in which molecules appear to be present in more than one form (aside from carbohydrate heterogeneity), despite quick denaturation, thus appear to represent true *in vivo* modifications. Ceruloplasmin, for example, exists as two forms of roughly equal abundance differing by 5,000–10,000 daltons; the larger form interacts strongly with precipitated albumin in the first dimension causing horizontal streaking, while the smaller form does not. The haptoglobin β-chain also exhibits two components, the smaller (lower molecular weight) one comprising only a few percent of the total (Fig. 2). Last, we confirm the presence in plasma of two forms of plasminogen differing by about 2000 daltons, the larger of which is more acidic (19). In each of these examples, the ability to analyze the immunoprecipitate under dissociating conditions has been crucial. Both the existence of *in vivo* modification and some biophysical data concerning the proteins themselves can be deduced directly from the gels.

A similar pattern results when urea instead of $NaDodSO_4$ is used to make up the sample, except that antithrombin III is not seen. The only probable artifacts discovered so far concern anomalous fluctuations in the apparent molecular weight of $\alpha_1$-antitrypsin in $NaDodSO_4$. Freezing and thawing of isoelectric focusing gels before $NaDodSO_4$ electrophoresis can cause such alterations, but the effect appears not to be due to insufficient $NaDodSO_4$ equilibration before the second dimension, disulfide formation (alkylation has no effect), or proteolysis (fast and slow migrating forms run together when re-electrophoresed in $NaDodSO_4$).

## Detection of polymorphisms and genetic variants

Several polymorphisms are immediately evident from comparisons of plasma from a few individuals: haptoglobin light chains $\alpha^2$, $\alpha^{1F}$, and $\alpha^{1S}$, Gc-globulin's three forms (22), several alleles of $\alpha_1$ antitrypsin, and two common forms of $\alpha_2$-HS glycoprotein (Fig. 4). Preliminary results from a series of several hundred plasma samples indicate the existence of rarer allelic forms of arginine-rich lipoprotein and unidentified protein G4 as well (data not shown). In studies of a series of transferrin variants provided by J. V. Neel and coworkers, isoelectric point shifts consistent with the known electrophoretic differences are observed (Fig. 5). All of this evidence is consistent with the accepted view (11, 23) that single charge differences (due to amino acid substitutions or to chemical modifications) are easily visible in this system. For those proteins (such as transferrin or G1, G2, and G3 in Fig. 3) that appear as a series of spots (due to sialic acid or other carbohydrate heterogeneity), a charge change mutation shifts the series as a whole (Fig. 5). How it is that many charge changes due to mutation can be detected in electrophoretic systems that do not reveal charge heterogeneity due to varying sialic acid content is not clear.

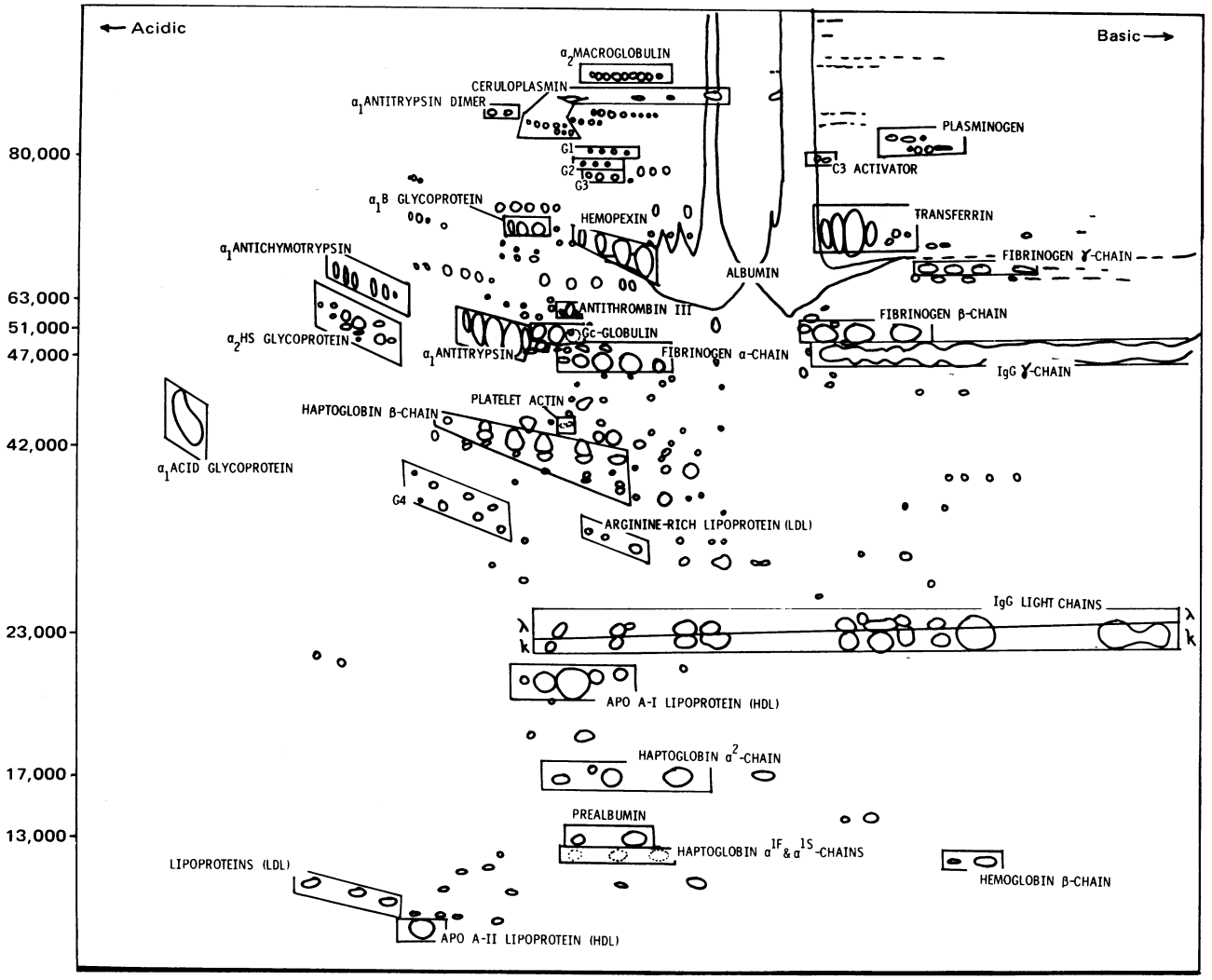


FIG. 3. Diagram drawn from the gel shown in Fig. 1, and labeled to indicate positions of known plasma proteins. Hemopexin and the C3-activator are somewhat obscured by albumin overloading. Ceruloplasmin appears to be present in two major and two minor forms (all between 80,000 and 90,000 daltons), each present as a row of four or more dots due to sialic acid heterogeneity. The highest molecular weight form interacts strongly with the albumin precipitate, while the others do not. Plasminogen exists in two forms: the Glu-form (upper horizontal row of dots) and the Lys-form (lower row, more basic) (19). Gc-globulin can be present as three spots; the left-hand pair appears to correspond to type 1, and the right-hand spot to the type 2 allele. The immunoglobulin light chains ($\kappa$ and $\lambda$) are partially resolved (20) and show similar isoelectric distributions. Identification of the lipoproteins is based on the presence of spots in certain of the low (LDL) and high (HDL) density lipoprotein fractions, as well as similarity to isolated materials for the arginine-rich and apo A-I lipoproteins. Platelet actin, Gc-globulin spot 3, and the haptoglobin $\alpha^{1F}$ and $\alpha^{1S}$ chains are shown although they were not present in the sample run in Fig. 1. As yet unrecognized glycoproteins G1, 2, 3, and 4 are labeled for use in the *text*. The hemoglobin $\alpha$-chain is too basic to appear in a separation with these ampholytes.

The ability to detect mutations in many loci at once is obviously useful for human genetic screening. We estimate that at least 30 gene products could be monitored routinely with 5- to 10-$\mu$l samples of plasma. The estimate is quite conservative in the sense that photographs of the gels could be stored for re-examination at a time when more is understood about variations in the minor spots. With improvements in technique, particularly the removal of nearly all the albumin from samples by immunoadsorption, it should be possible to monitor the products of 60–100 genetic loci by using plasma and simple optical techniques for comparing gels. More sophisticated methods using a computerized densitometer (to be described elsewhere) will hopefully contribute further to sensitivity and speed.

Concerns regarding genetic damage due to pollutants are best met by a human genetic monitoring system followed by appropriate action if increases in mutation rate are observed. Assuming a baseline mutation rate of $0.5 \times 10^{-5}$/locus per generation, it will require two samples of approximately 6,000,000 observations each to detect a 50% increase in mutation rate (24). An effort of this magnitude can only succeed given a variety of technical advances (25). When the present approach is extended to other easily obtained protein mixtures such as erythrocyte, platelet, leukocyte, salivary, and urinary proteins, it is likely that 300 gene products could be observed routinely in adults. A more practical approach to monitoring could make use of perhaps 200 proteins (from plasma, erythrocytes, leukocytes, and platelets) present in placental cord blood, and would entail two successive samples of 30,000 each. This would involve running 120,000 gels, if each mixture is treated separately, and is currently a feasible project with our apparatus. Note that the spots do not have to be identified with known proteins to be used in such a genetic screen, although as variants are accumulated, the motivation to make identifi-

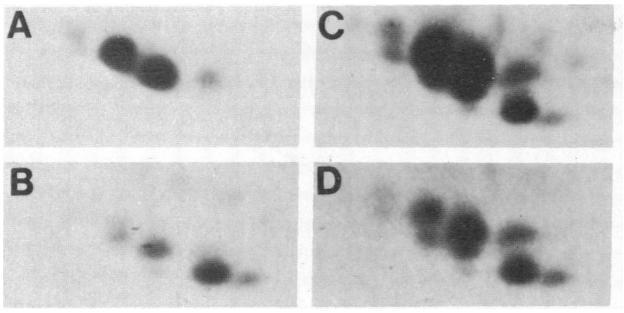


FIG. 4. Sections of the two-dimensional gels showing two common $\alpha_2$-HS glycoprotein types, which we call N and L. The patterns are portions of whole plasma separations: (*A*) type L pattern, (*B*) type N pattern, (*C*) mixture of N and L plasmas, (*D*) pattern from a single plasma showing both types. Since the N and L patterns would overlap extensively in either single dimension, they would be difficult to distinguish by other techniques, and to our knowledge have not been previously resolved.

cations will progressively increase. Instead, it will suffice to treat each spot or lattice as a unit of empirical genetic data which can be further characterized if the need arises.

## CONCLUSIONS

The introduction of high resolution protein mapping techniques will eventually lead to the enumeration of most, if not all, human gene products, and to an understanding of differentiation at the molecular level. First, however, it is necessary to demonstrate that a relatively well understood collection of proteins can be separated and that some of the major characteristics of each can be determined correctly. We believe this has been done in the present analysis of the plasma proteins, and that a powerful system for human genetic analysis has been generated. The next stage, a survey of human cell types, will require more sophisticated data reduction and new techniques for identifying known components.

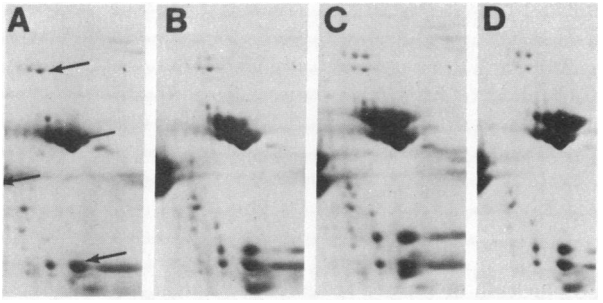


FIG. 5. Sections of two-dimensional gels showing the appearance of transferrin variants compared to the normal form. (*A*) Photograph of a normal plasma gel; (*B–D*) gel of a variant transferrin-containing plasma superimposed on the normal and displaced slightly upward to allow comparison. (*A*) Arrows indicate (from top to bottom) the C3-activator glycoprotein, transferrin, albumin, and the fibrinogen $\beta$-chain; (*B*) homozygous transferrin variant (type B) with one extra negative charge; (*C*) heterozygous variant (type C-B) with one extra negative charge (the variant pattern is a combination of the normal and the variant in *B*); (*D*) heterozygous variant (type C-D, D type undetermined) with one extra positive charge. Samples were 1.3 $\mu$l of plasma prepared as described in *Materials and Methods*.

We thank the following for gifts of purified proteins: G. Barlow, C. Reimer, P. Alsen, C. J. van Oss, V. and B. Shore, J. V. Neel, B. A. Bonner, the Red Cross, and the Behringwerke. In addition, we thank Dr. J. V. Neel for comments on the manuscript and Sandra Tollaksen and Barbara Hickman for excellent technical assistance. This work was supported by the U.S. Energy Research and Development Administration, and in the early stages by the Medical University of South Carolina.

1. Tiselius, A. (1937) *Biochem. J.* **31,** 313–317.
2. Raymond, S. & Weintraub, L. (1959) *Science* **130,** 711.
3. Davis, B. J. (1964) *Ann. N.Y. Acad. Sci.* **121,** 404–427.
4. Laemmli, U. K. (1970) *Nature* **227,** 680–685.
5. Righetti, P. & Drysdale, J. W. (1971) *Biochim. Biophys. Acta* **236,** 17–28.
6. Margolis, J. & Kenrick, K. G. (1969) *Nature* **221,** 1056–1057.
7. Dale, G. & Latner, A. L. (1969) *Clin. Chem. Acta* **24,** 61–68.
8. Kaltschmidt, E. & Wittman, H. G. (1970) *Anal. Biochem.* **36,** 401–412.
9. Laurell, C.-B. (1965) *Anal. Biochem.* **10,** 358–361.
10. Freeman, T. & Smith, J. (1970) *Biochem. J.* **118,** 869–873.
11. O'Farrell, P. H. (1975) *J. Biol. Chem.* **250,** 4007–4021.
12. Barrett, T. & Gould, H. J. (1973) *Biochim. Biophys. Acta* **294,** 165–170.
13. Scheele, G. A. (1975) *J. Biol. Chem.* **250,** 5375–5385.
14. Iborra, F. & Buhler, J.-M. (1975) *Anal. Biochem.* **74,** 503–511.
15. Anderson, N. G. & Anderson, N. L. (1977) *Anal. Biochem.*, in press.
16. Anderson, N. L. & Anderson, N. G. (1977) *Anal. Biochem.*, in press.
17. Ames, G. F.-L. & Nikaido, K. (1976) *Biochemistry* **15,** 616–623.
18. Klose, J. (1975) *Humangenetik* **26,** 231–243.
19. Robbins, K. C., Boreisha, I. G., Arzadon, L., Summaria, L. & Barlow, G. H. (1975) *J. Biol. Chem.* **250,** 4044–4047.
20. Virella, G. & Coelho, I. M. (1974) *Immunochemistry* **11,** 157–160.
21. Morrisett, J. D., Jackson, R. L. & Gotto, A. M., Jr. (1975) *Annu. Rev. Biochem.* **44,** 183–207.
22. Ruoslahti, E., Simons, K. & Ehnholm, C. (1971) *Eur. J. Immunol.* **1,** 45–48.
23. Steinberg, R. A., O'Farrell, P. H., Friedrich, U. & Coffino, P. (1977) *Cell* **10,** 381–391.
24. Neel, J. V. (1971) *Perspect. Biol. Med.* **13,** 522.
25. Neel, J. V., Tiffany, T. O. & Anderson, N. G. (1973) in *Chemical Mutagens*, ed. Hollaender, A. (Plenum Press, New York), Vol. 3, p. 117.