Reprint from: Electrophoresis '79
Advanced Methods
Biochemical and Clinical Applications
Editor: Bertold J. Radola

HIGH-RESOLUTION TWO-DIMENSIONAL ELECTROPHORETIC MAPPING OF HUMAN PROTEINS

N. L. Anderson, J. J. Edwards, C. S. Giometti, K. E. Willard,
S. L. Tollaksen, S. L. Nance, B. J. Hickman, J. Taylor, B. Coulter,
A. Scandora, and N. G. Anderson

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

Introduction

The high resolution achievable with two-dimensional electrophoresis [isoelectric focusing in urea-NP40 followed by SDS-PAGE (1-4)] is sufficient to encourage us to contemplate mapping and cataloguing a large fraction of the estimated 30,000 to 50,000 human protein gene products (PGP's). The properties of only a few hundred human proteins (exclusive of immunoglobulins) have been described, and a very much smaller number characterized in detail. This leaves approximately 99% of the working parts of human cells not only unknown, but with functions largely unsuspected. We have therefore systematically investigated how a comprehensive catalogue might be made, and have devised methods for characterizing them en masse in the mapping process. Given automation of the analyses, direct gel or autoradiograph scanning, and the assistance of computers in quantitation, data storage, and analysis (described elsewhere in this volume), the project, usually referred to as molecular anatomy (4), is now technically feasible. List-based biology, which this project makes possible, will be a science in itself, and will be concerned with the details of the organization of gene expression (i.e., with the organization of genes into coregulated sets or batteries, with the sequencing of these sets during development), with misregulation in cancer, and with the details of genetic disease. The ultimate effects of this work on clinical diagnosis, on human genetics, and on molecular biology are difficult to overestimate.

This article reviews briefly the technology developed for running large numbers of high-resolution two-dimensional analyses, and considers some of the techniques that may be applied to the characterization of PGP's on gels. In view of the number of proteins involved, it is evident that methods requiring the isolation of each and every protein for systematic identification and description are too time consuming. The objective is to describe each spot by name and/or function, and to assign to each spot a sufficient number of properties to allow each to be distinguished from all other PGP's including the products of allelic genes. Methods are therefore required that provide information concerning many proteins in a single experiment. This approach uses the two-dimensional map as an organizational framework within which various kinds of data may be collated. Questions to be asked include: Is a given row of equi-spaced spots composed of variations (posttranslational modifications) of one polypeptide? If so, are the differences between spots due to phosphorylation,sialation, glycosylation, or something else? Which spots are of nuclear,cytoskeletal, mitochondrial, or plasma membrane origin? Which vary in abundance during the cell cycle, during differentiation, or from tissue to tissue? Which change during neoplastic transformation? The answers may provide clues to the identification of function, which is, of course, a central objective. While this approach to the discovery of function may seem indirect, we note that function is difficult to establish (given a protein or PGP as a starting point) except by systematic, unfortunately unenlightened trial and error. Starting with function (and therefore an assay) is much easier; however, one then finds only the proteins one expects to find, and this is a very limited subset of what exists.

Methods for Running Large Numbers of Analyses

First dimension isoelectric focusing. The ISO apparatus (2, 4) in its current form allows twenty 1.5-mm diameter focusing gels to be set up and run with minimal effort. The tubes are simultaneously filled with the acrylamide solution in one step by hydrostatic displacement. After polymerization the device is disassembled, excess gel is removed, and

the appropriate upper and lower buffers are added. Using multipipette loaders and gel extrusion devices (to be described elsewhere), 100 gels can be made and loaded with samples in 1-2 hours.

Second dimension SDS-PAGE. The DALT casting apparatus (3, 4) as used in our laboratory allows preparation of 1.5-mm thick gradient sodium dodecyl sulfate (SDS) slab gels (usually 10-20% acrylamide) in identical sets of 10, 20, or 80 gels. Following the method of J. I. Garrels (personal communication) the stacking gel is omitted and the 1.5-mm focusing gel is applied directly to the flat upper surface of the resolving slab, and sealed in place with hot 0.5% agarose in electrode buffer. Elimination of the stacking gel improves resolution (in the case of these small isoelectric focusing gels) and also allows the slabs to be stored indefinitely at 4°C before use. Each slab contains an integral serial number typed on a filter paper strip which is dropped between the glass plates before polymerization. The slab gel plates are square, without notches, and are constructed as books with a silicone rubber binding along one edge, and with the spacers attached to one plate. With ten DALT tanks, 100 slabs can be run in parallel.

Staining, destaining, photography, drying, and autoradiography. Staining and destaining, using Coomassie Blue in acetic acid-ethanol, is performed by shaking the gels in Rubbermaid 4-quart plastic boxes with sealing lids, 10 gels per box. Drying of gels is accomplished using a specially constructed 20-place gel dryer assembled on a lyophilizer modified to use a 500-liter/minute vacuum pump. Gels are pressed against film for autoradiography by placing the gel-film sandwich (assembled in a cardboard Kodak exposure holder) in one of 50 narrow reenforced shelves together with a plastic air bag which is continuously inflated to 2 psi. Wet gels are photographed through an interference filter (band pass at the Coomassie Blue absorption maximum) using high contrast film developed in a Kodak Versamat automatic processor. All photographs include a calibrated density wedge to enable accurate control of contrast, exposure, and development.

During the development of this system and for research in progress over 20,000 two-dimensional analyses have been performed to date.

Methods for Characterizing Proteins on Gels

The techniques that may be used for gel-based protein characterization fall into three general classes: (a) physico-chemical methods which exploit the properties of the molecules themselves; (b) subcellular fractionation which seeks to determine the locations of molecules in cells; and (c) whole cell biological experiments in which variations in gene expression are correlated with differentiation, malignancy, genetic disease, or experimental variables. Some of these techniques are illustrated in this paper, and as a reference in some of the examples, an ISO-DALT pattern of ^{35}S methionine-labeled GM 1500 human myeloma cells is used, as shown in Figure 1.

Physico-chemical methods for characterization. The most obvious protein characteristics employed in this work are isoelectric point and molecular weight. Unfortunately, neither can be measured directly from the gels with an accuracy close to the resolving power of the system unless standards are also incorporated. Therefore we have developed internal charge standard markers for pI measurement (5, and Hickman et al., this volume) and high-resolution SDS molecular weight markers (4, 6). While use of these procedures fails to provide exact physico-chemical data, the methods do provide highly reproducible coordinate information.

The amino acid composition of cellular proteins may be obtained by growing cells in a series of media, each containing a single ^{3}H or ^{14}C labeled amino acid as shown in Figure 2. By referring the relative incorporation to the amount of label in spots of known amino acid composition, such as actin, approximate compositions may be obtained for all spots. Proteins that completely lack a given amino acid are easily detected. Note the marked protein in Figure 2 which lacks tryptophan, as first noticed by J. I. Garrels.

A special case of amino acid labeling is the use of ^{14}C iodoacetamide (IAm) to label cysteine residues. This reaction, first suggested to us by T. Hunt,can be used to label proteins after synthesis and does not appreciably alter pI or SDS-molecular mass. The reaction has the advantage of proceeding essentially to completion. Hence, given the specific activity

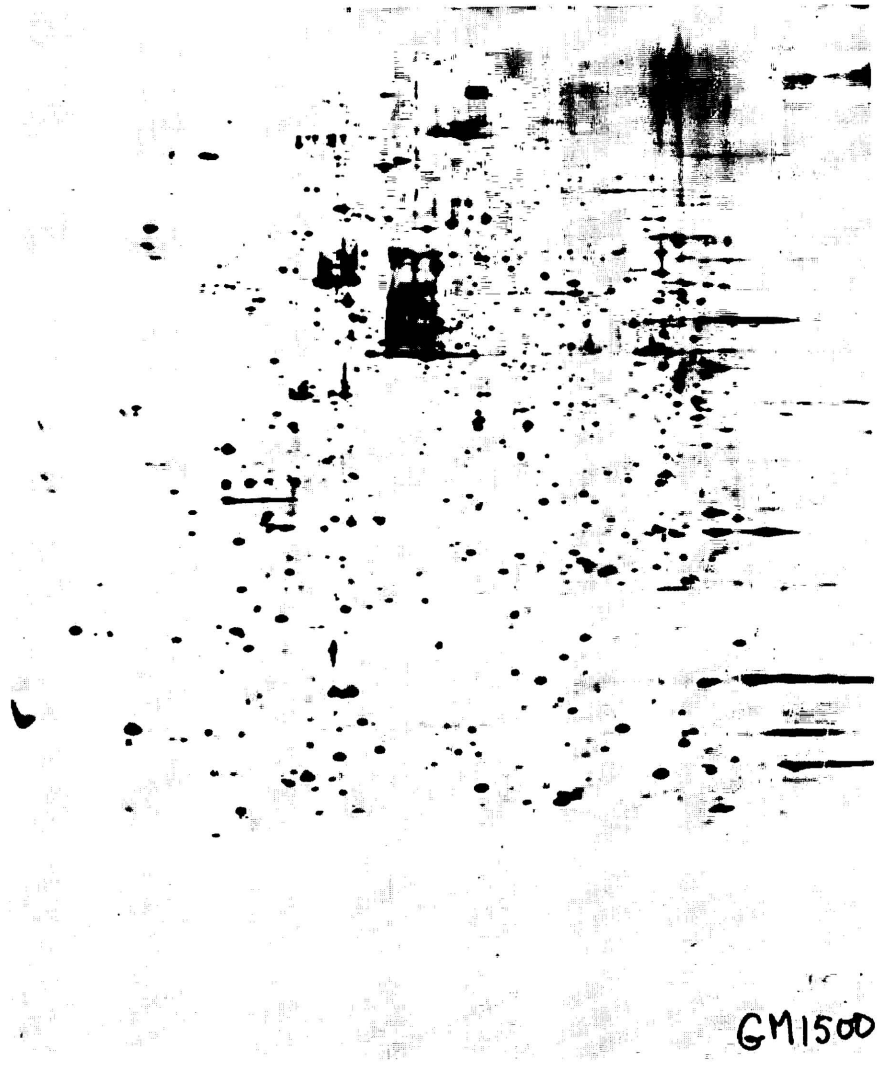

Fig. 1. Two-dimensional electrophoretic pattern of GM 1500 human myeloma cells (IgG, κ producers) labeled with ^{35}S methionine. LKB 3.5-10 Ampholines used with 10-20% gradient slab. Acid end of isoelectric focusing gel to left. (Cells kindly supplied by Dr. Carlo Croce, Wistar Institute).

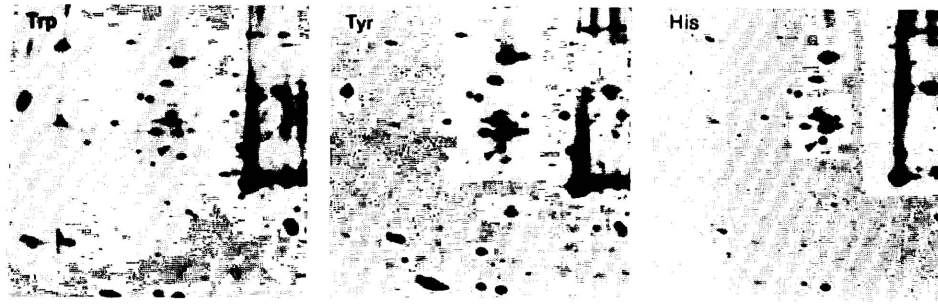

Fig. 2. Autoradiograph of human peripheral blood lymphocytes labeled with either tryptophan (Trp), tyrosine (Tyr), or histidine (His). Gels run as in Figure 1.

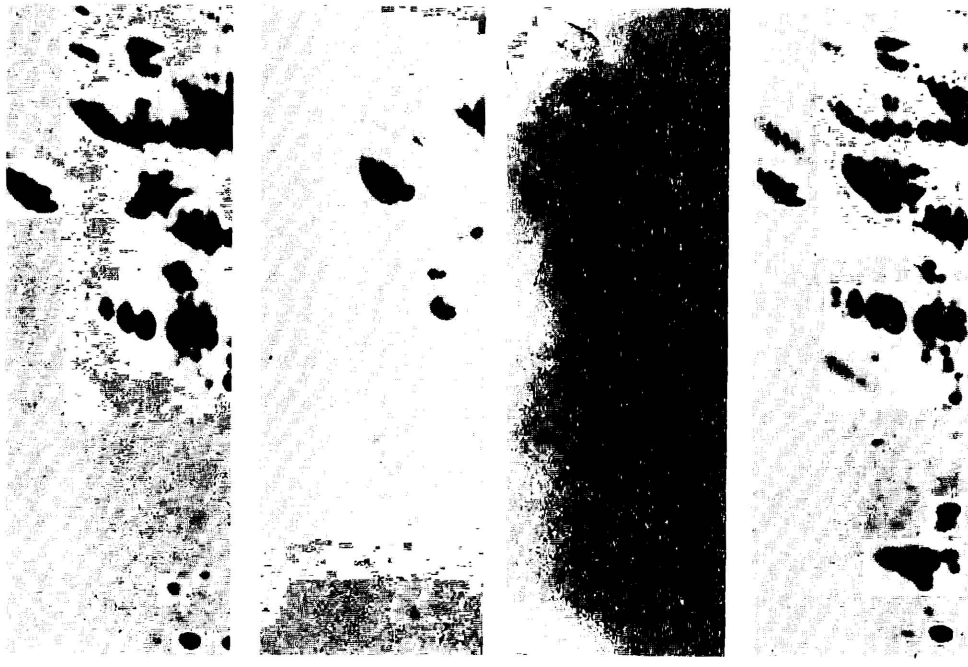

Fig. 3. Human plasma proteins labeled with ^{14}C iodoacetamide after disulfide reduction (A); labeled with ^{14}C IAm with no initial disulfide reduction, showing only "reactive" SH groups (B); stained with periodic acid-Schiff reagent (C); and stained with Coomassie Blue (D).

of the IAm, the number of cysteines in a protein, and the SDS molecular mass, one can calculate the absolute grams of protein in a spot as

$$g\ \text{protein} = \frac{(\text{cpm}\ ^{14}\text{C})\ (\text{protein mass} \cong \text{SDS-MW})}{(\text{No. of SH groups/molecule})(\text{cpm/mole}\ ^{14}\text{C IAm}).}$$

Further, if the number of SH groups for one protein in a pattern is known, then the number of SH groups in each and every other spot can be determined by measuring the relative spot intensities, and this information may be related to SDS molecular weight, determined from vertical spot positions.

Two types of SH groups may be distinguished using IAm (Figure 3). If the disulfide reduction is done first, all cysteines are labeled; if no initial reduction is done, only "reactive" or non-disulfide bonded cysteines will be labeled. This technique has been used on serum, muscle, and RBC lysates to identify proteins known to have many or no cysteines (G-6-PDH and Apo-Al lipoprotein, respectively) or those known to have only reactive (non-disulfide bonded) cysteines such as α_1-antitrypsin. Note that if ^{14}C iodoacetic acid is used, a single charge shift is produced for each SH substitution. By mapping mixtures of partially and completely substituted PGP's along with the untreated mixture, charge shift trains [analogous to those produced by charge shift standards (5)] are seen, enabling the number of SH groups to be directly counted for each spot.

It is also of interest to compare ^{35}S methionine labeling with ^{14}C IAm labeling as is done in Figure 4. ^{32}P labeling can be used to identify phosphorylated proteins in solubilized whole cell preparations as shown in Figure 5.

Additional physico-chemical methods include the use of group-specific stains, of lectin or of antibody binding to spots in gels, measurement of susceptibility to digestion with added enzymes, and the sequential use of two-dimensional non-denaturing and two-dimensional denaturing systems to enable enzymes to be identified histochemically in the first two-dimensional analysis, and the subunit composition to be determined in the second analysis.

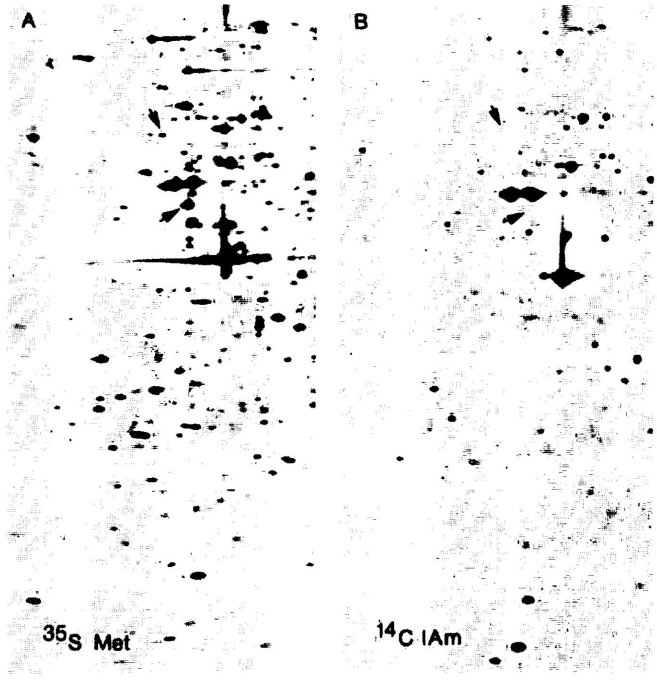

Fig. 4. Comparison of GM 1500 human myeloma cells labeled with ^{35}S methionine (^{35}S Met) or ^{14}C iodoacetamide (^{14}C IAm).

When large number of gels can be run at many different sample loadings, and spot intensities are determined with precision, additional protein characteristics may be examined. We have developed a heat denaturation system which allows many samples to be heated at a closely graded series of temperatures for the same length of time (7). The samples are then centrifuged and the supernatants are analyzed for enzymatic activity and by two-dimensional electrophoresis. The inflection points for the heat denaturation curve of spots may then be determined, and included as a fundamental property in the human protein index, and also related to enzymatic activity as described by Nance et al. in this volume.

Specific antibodies may also be used to identify spots by either mapping immunoprecipitated proteins, or by noting, by comparison, spots removed from mixtures by antibodies (8). Protein A sepharose (9) facilitates this work. Ultimately, monoclonal antibodies against each PGP will be required.

Fig. 5. Comparison of GM 607 human lymphoblastoid cells (IgM, κ producers) labeled with ^{35}S methionine (^{35}S Met) or ^{32}P.

Subcellular localization. We are interested in both the subcellular localization of PGP's and in the chromosomal localization of the genes coding for them. Conventional cell fractionation may be used for subcellular localization. However, the resolution of present two-dimensional electrophoresis is such as to give ultimately a better assessment of purity of cell fractions, once reliable markers have been identified, than conventional biochemistry or electron microscopy. The analysis of HeLa cells labeled with ^{35}S methionine and of nuclei isolated from them is illustrated in Figure 6. Nuclei have been isolated by several different methods, and actin and intermediate filament protein are found to be major nuclear proteins. One major nuclear phosphoprotein is common to HeLa and lymphoblastoid cells, but is minor or nonexistent in peripheral blood lymphocytes.

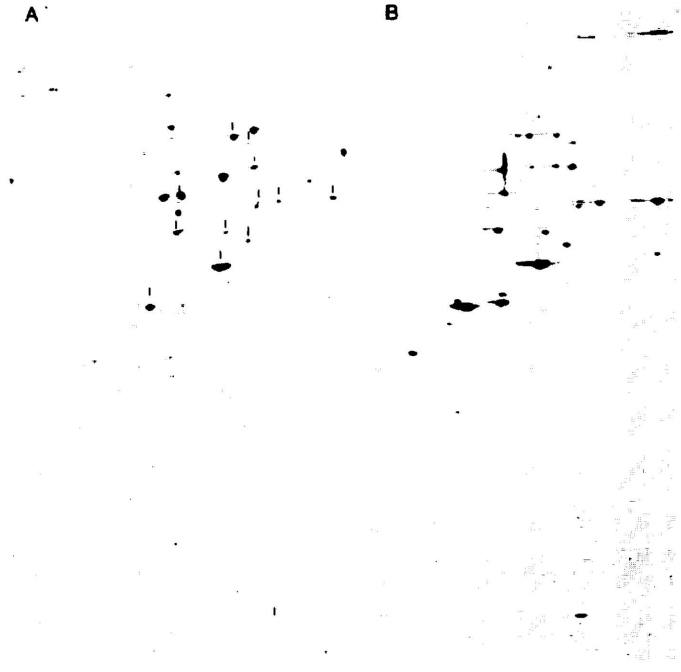

Fig. 6. Comparison of whole HeLa cells labeled with ^{35}S methionine (A) and nuclei isolated from them (B).

Other localization techniques include the isolation of so-called cytoskeletons by detergent treatment from labeled lymphocytes as shown in Figure 7. Note the protein released from cytoskeletons by RNase. The cell surface proteins of living cells may be labeled with ^{125}I by gentle methods to identify proteins containing externally exposed tyrosine or histidine. Unfortunately, ^{125}I autoradiographs give poor resolution, due to the high penetrating power of the radiation involved. Living cells may also be labeled by ^{3}H borohydride reduction of aldehydes formed either by sodium periodate or galactose oxidase oxidation of glycoprotein sugar residues. Some cell surface proteins may be digested with proteolytic enzymes while maintaining cell viability. These should be reduced in abundance or in spot density on 2-D analysis. Labeled proteins shed from cell surfaces may be readily identified by analysis of culture media.

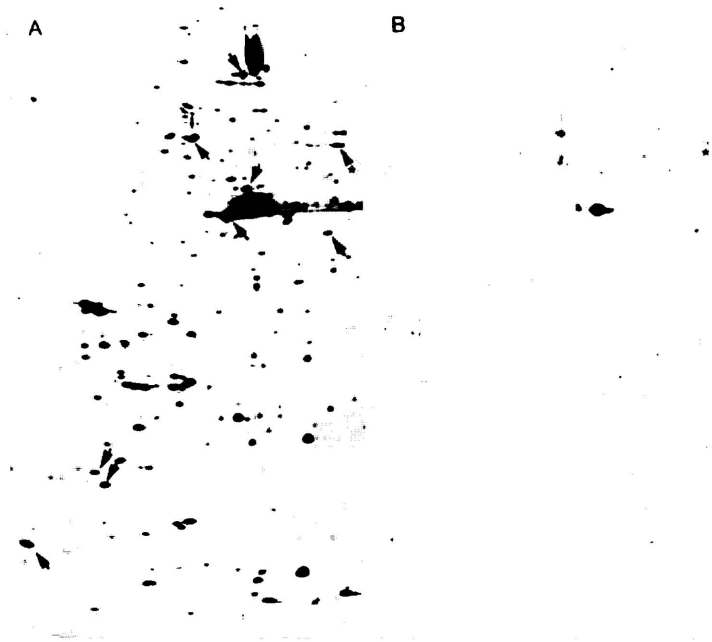

Fig. 7. Comparison of whole human peripheral blood lymphocytes (A) and detergent-prepared cytoskeletons (B). ^{35}S methionine label. The star indicates protein released from the cytoskeleton by RNase.

Protein characterization based on "whole cell" biology. As different human cell types are analyzed, PGP's will be characterized as belonging to the so-called basic set found in all nucleated cells, or as being present in less than all, i.e., to be expressed selectively during differentiation, or during different stages in the cell cycle. In addition to knowing when each structural gene is expressed, it is important to know the chromosomal location of each gene. This problem may be approached by examining fused human-mouse cells which retain only selected human chromosomes, as illustrated in Figure 8, to see which human proteins are still expressed. When marker proteins have been found for all human chromosomes, and variants of the markers (and of many other cell proteins) discovered, then massive chromosomal attributions will follow from the analysis of families bearing these variants (Figure 9). The study of genetic polymorphisms in man is also basic to attempts to determine both the background

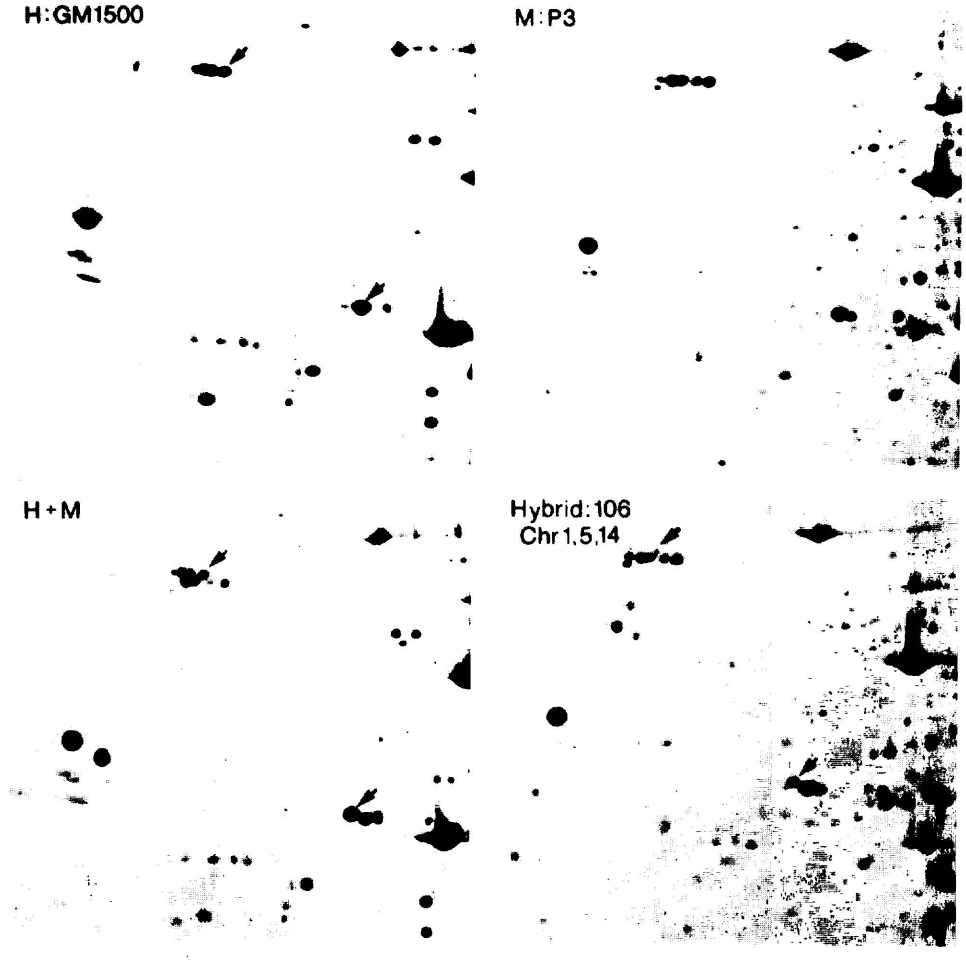

Fig. 8. Demonstration of technique used to relate specific spots to human chromosomes. H: human GM 1500 cells. M: mouse P3 cells. H + M: mixture of extracts prepared from both cells, i.e. human and mouse. Hybrid 106:, containing only human chromosomes 1, 5, and 14 (provided by C. Croce and M. Shander). Human proteins common to hybrid and unhybridized human GM 1500 cells are indicated by arrow.

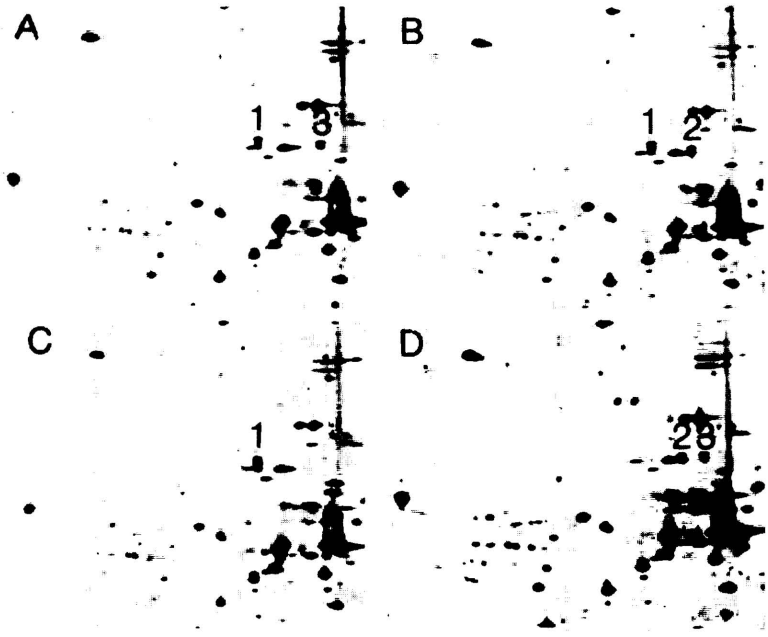

Fig. 9. Autoradiographs of peripheral blood lymphocytes from one family labeled with ^{35}S methionine. (A) father, (B) mother, (C) son, (D) daughter. Spot labeled 1 is common to both parents and son. Spot 2 is found only in mother and daughter, while spot 3 is in father and daughter only.

mutation rate in man and variations in it which may be produced by radiation, or environmental mutagens. This would be done by examining maps of offspring in search of mutant PGP's not present in either parent. Thus far, we have examined over 100 lymphocyte samples from unrelated individuals and have identified a number of polymorphic systems. Preliminary data suggest that the average heterozygosity for 200 peripheral blood lymphocyte loci (1-3%) is between 1/3 and 1/5 that expected from serum and red blood cell lysate protein studies.

Experiments of nature - i.e., human genetic disease - will also provide important clues to function, as the presence or absence of specific spots is correlated with specific symptoms.

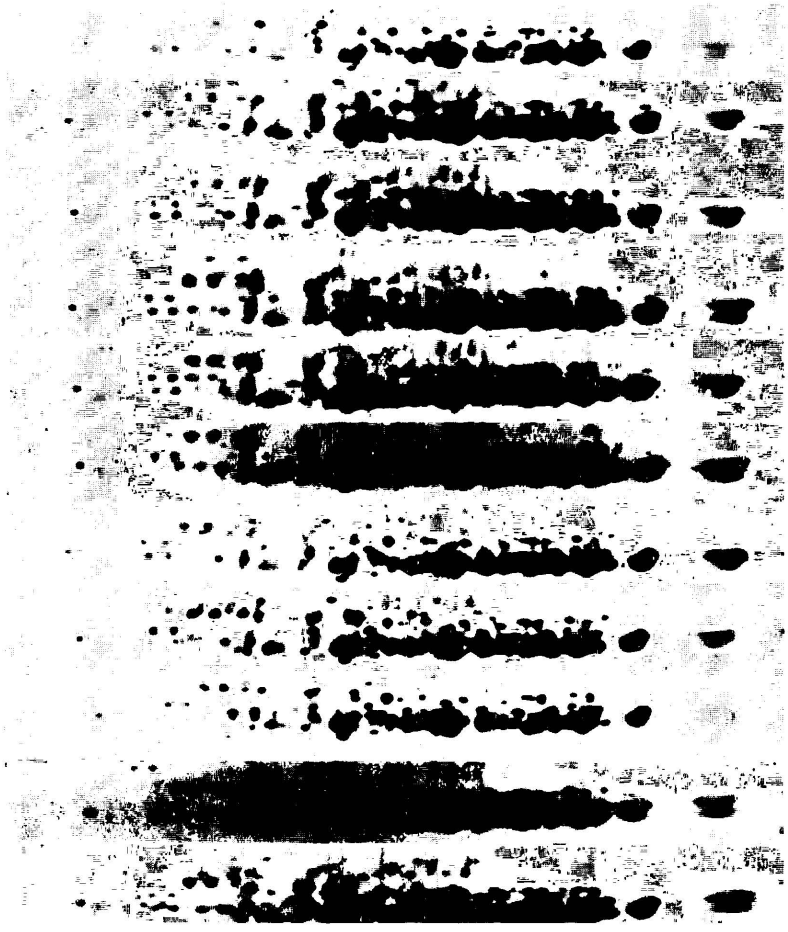

Fig. 10. Two-dimensional maps of IgG light chains prepared from eleven individual BALB/c mice.

The response of cells or organisms to controlled experimental variables will also provide important clues. It is probable that 2-D mapping of alterations in PGP maps of cultured cells will be an important part of drug and environmental chemical monitoring to determine which substances alter gene expression, or alter the posttranslational modification of PGP's. More subtle changes may also be monitored. For example, Figure 10 shows 2-D maps of IgG light chains from eleven individual Balb/c mice;

70-80% of the spots are superimposable. Common spots may indicate many genes in common and/or a common antigenic history. Quite obviously, high resolution two-dimensional mapping allows these problems to be approached experimentally.

Summary and Conclusions

It is now technically feasible to resolve and map two dimensionally a very large fraction of all human protein gene products (monomeric proteins or protein subunits), and to characterize each resolved component in relation to physico-chemical properties, intracellular location, genetics, and function. The organization of the protein portion of the human molecular index is now a matter of active concern. It will ultimately involve organization and classification of proteins into coregulated sets (plus the nonregulated basic set if such exists), into linkage groups depending on the chromosomal localization of the individual genes, and by differentiational groups that characterize differentiating or differentiated cells. A requirement of the index is a sufficient body of descriptive data to distinguish any one protein gene product from all others.

The implications of this work are many; one, however, stands out. Early cancer detection, proper tumor classification, evaluation of therapy, elucidation of mechanisms of action of antitumor agents, and a detailed understanding of how a tumor cell differs from a normal one, all require a complete index. This alone justifies the effort.

Acknowledgment

This work was supported by the U. S. Department of Energy under contract No. W-31-109-ENG-38.

References

1. O'Farrell, P. H.: J. Biol. Chem. 250, 4007-4021 (1975).
2. Anderson, N. G., Anderson, N. L.: Anal. Biochem. 85, 331-340 (1978).
3. Anderson, N. L., Anderson, N. G.: Anal. Biochem. 85, 341-354 (1978).
4. Anderson, N. G., Anderson, N. L.: Behring Inst. Mitt. 63, 169-210 (1979).
5. Anderson, N. L., Hickman, B. J.: Anal. Biochem. 93, 312-320 (1979).
6. Giometti, C. S., Anderson, N. G., Tollaksen, S. L., Edwards, J. J., and Anderson, N. L.: Anal. Biochem., in press, (1979).
7. Anderson, N. L., Eisler, W. J., and Anderson, N. G.: Anal. Biochem. 91, 441-445 (1978).
8. Anderson, N. L., Anderson, N. G.: Proc. Nat. Acad. Sci. USA. 74, 5421-5425 (1977).
9. Pearson, T., Anderson, N. L. in preparation.