Separatum aus

Behring Institute Mitteilungen

Behring Institute Research Communications

No. 63 · June · 1979

Behring Inst. Mitt., No. 63, 169–210 (1979)

Molecular Anatomy*

N. G. ANDERSON, N. L. ANDERSON

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

In this presentation, we review the emerging field of Molecular Anatomy and how it makes possible radically new approaches to the understanding of human disease at the molecular level.

In *De Humani Corporis Fabrica*, published in 1543, a book which many consider to be the first book of modern science, Andreas Vesalius established the standards and some of the tools of the discipline of anatomy and stated the viewpoint to which anatomists have since held: to describe structure as completely as possible, to the limits of one's ability to separate and dissect. This basic concept has remained as the light microscope and then the electron microscope have made possible first microscopic anatomy and then ultrastructural analysis. Unfortunately, the systematic descent to the description of ever smaller constituents appears now to have found an impenetrable barrier in the laws of electron optics, and it has not been evident how the idea of "seeing all" could be brought down to the molecular level. Crystallography allows individual molecules to be viewed anatomically, and made possible the molecular biology revolution. However, if the anatomists' viewpoint is to be continued, and if one is to consider seriously establishing a science of molecular anatomy, then clearly some new approaches are required.

Vesalius made an additional contribution for which he is rarely given credit. The opening paragraph of his dedication to Charles V disposes of formalities very quickly and comes straight to the point, which is that the science of medicine had already been so subdivided into specialities that both knowledge and effectiveness were lost. There followed as ringing a defense of interdisciplinary research and communication as could then be managed. The problems of interdisciplinary research are not new.

If there is to be a science of molecular anatomy, it must clearly be an interdisciplinary one, borrowing techniques from biochemistry, biophysics, nuclear physics, chemistry, engineering, and mathematics, while retaining the essential viewpoints of anatomy, including the notion of completeness. The latter concept is an accepted one in nuclear particle physics where it is well understood that descriptions of nuclear forces and properties cannot be complete if particles have been overlooked.

As with the rest of science, the core problems are those of technique and of theory. First, how can the tens of thousands of human protein gene products be separated, quantitated, and identified? And secondly, are there important questions which can only be answered by a detailed study of the pattern of expression of human protein gene products? Present estimates of the number of human protein gene products (PGPs) range between 30,000 to 50,000, and to make matters worse, many of these are post-translationally modified.

The Program of Development

A complete list of PGPs, could it be obtained, would be of little interest in itself. Additional information about when and where they appear is required. Less than 10% of structural genes are thought to be expressed in any one human cell type, suggesting that some schedule or plan of genetic expression exists. It appears probable that the complex array of genes is organized into sets, and that the members of a set are turned on and off together. Hybridization

* This work was supported by the U. S. Department of Energy.

studies suggest that massive gene phasing occurs during development, and that different genes or gene sets are active at different times. One would like to be able to organize a master PGP list into sets, and ask how the expression of these is arranged during development. Does one set trigger the expression of another giving rise to sequencing rules? Does a set near the end of a developmental sequence turn off those temporally previous to it? Are there rules of exclusion such that certain sets cannot coexist, and are there rules of inclusion such that if one set is expresed, one or more other ones are then always turned on? Are sets inviolable, or can one gene be a member of more than one set? We do not know the answer to these questions; in fact, we do not know the members of a single human gene set for certain. What is probable is that if we had the list of PGPs from each human cell type at each stage in differentiation, these questions could be answered. It may be that none of our present ideas are correct, and that some totally new theory of how gene programming occurs will be required. Progress in the study of the temporal and functional organization of PGPs will be unsatisfactory in the absence of sufficient analytical data to support, eliminate, or suggest theories.

Further, if human cancer is a disease of the mechanisms underlying development, then it is unlikely that one will be understood without understanding the other, as has been frequently pointed out. The fundamental defect may be the reactivation of a gene set normal to early development, which puts in motion a chain of events also normal to that stage. Against the background of a normal differentiated cell, this reactivation could lead to problems which are difficult to imagine in detail. However, if one had reasonably complete lists of PGPs from tumor cells, these, by comparison with lists from normal cells at many developmental stages, may provide the necessary answers. The question of whether cancer involves the expression of only one or a few key genes could also be answered.

The purpose of this discussion is to make the point that if the objectives of the molecular anatomist could be reached, problems in human medicine that are not now solvable could be attacked.

Early History of the Molecular Anatomy Program

The molecular anatomy program described here had its inception in a series of seminars organized in 1959 at the request of *Dr. Alvin Weinberg*, which addressed the problem of diversifying the Oak Ridge National Laboratory in Oak Ridge, Tennessee. It was thought at that time that the problems associated with nuclear energy were nearing a practical solution, and it was not foreseen that there would be an energy crisis a decade and a half hence. One of us (NGA) presented a seminar and a proposal on the fractionation of human cells – how this might be gone about rationally, and what instrumental systems would be required to do it. It should be recalled that the Oak Ridge National Laboratory as a whole was largely dedicated to separations and analyses, that it had very large groups working on a wide range of analytical problems, and had been the key laboratory in the separation of the fission products and the isolation from them of medically useful radioisotopes.

The suggestion to attempt the complete analysis of human cells did not, therefore, appear to be at all out of place, and the basic concept appealed to nearly everyone except our fellow biologists. Quite obviously, it was then an impossible problem, and one that could never be carried to completion without major technical advances. In defense of the suggestion, it was noted that even attempting the project would define many interesting problems, and would give an umbrella under which work otherwise difficult to support could be carried out. But there were deeper questions. The first related to the indigenous obscurantism that has often pervaded biological research. It expresses itself as a semihope that the most basic bio-

logical problems will not be solved, indeed cannot be solved, but only worked on. Clearly the proposed program ran counter to this view.

The second problem was equally worrisome and relates to the question of how useful knowledge is actually acqired. Support for biomedical research appears to be largely based on the assumption that detailed knowledge of human cell constituents will arise as a vast accretion from unrelated, uncoordinated individual efforts. Cross-disciplinary work usually involves some coordination, organization, and direction, and was quite contrary to views then prevailing as to how biological research ought to be done. The contrast between what was being done in physics and in biology was quite evident.

We drew the conclusion that the basic concept of attempting to organize a molecular anatomy program making use of a very wide range of talents and technologies was sound, but that it should be approached in relatively slow stages. When the range of techniques available in 1959 was reviewed, it appeared that cell fractionation would have to be an essential ingredient in an orderly approach to cell disassembly and analysis. Further, if subcellular particles were to be the starting materials for further fractionation, then high-resolution, high-capacity separations systems would be required. The zonal centrifuge had been invented in 1955 by one of us, but it clearly needed much additional development. The history of the development of this instrument, or series of instruments, has been published [1,2], and a rather large literature concerned with the subject has evolved. The experience of that effort, carried out in collaboration with the Separations Systems Division of the Oak Ridge Gaseous Diffusion Plant, amply demonstrated what could be done when an interdisciplinary group grappled with an interesting interdisciplinary problem. At no time was the program allowed to be splintered so as to lose sight of the real biological problems. Every attempt was made to define aspects of the overall problem of cell dissection in terms intelligible to a specialist. It did not work to put two specialists (be they biologist, physicists, or engineers) together to solve a problem without someone in between who could talk to both. It might be thought that such communicating catalysts could be trained, possibly as bioengineers. This, in our experience, is simply not the case. (Where such people come from, what training, abilities, and motivations they possess is a subject for discussion elsewhere.) Some fifty experimental rotor systems were constructed, only a few of which became commercially available. The concept of continuous-sample-flow-with-banding grew out of this work [3-5] and led to the development of the large K series rotors now used throughout the world to purify influenza vaccines and to prepare quantities of virus for research purposes. The centrifuge program was phased out, not because all of the problems (especially those related to the separation of various types of cells from tissues) had been solved, but because it had been carried far enough for that day, and because other parts of the program needed attention. Parenthetically, we note that there is now need and reason to return to centrifuge development, if only for a brief period, to finish the development of several systems which we now very much need. (These will appear as the F, G, and H series rotors to be described elsewhere.)

During this period, the problem of analyzing intermediate molecular weight compounds in cells had been examined. The ion-exchange techniques developed at Oak Ridge for the separation of fission products and of plutonium had been adapted to biological analyses largely under the influence of Waldo Cohn, but no one at that time had pushed ion exchange separations toward their theoretical limits. *Moore* and *Spackman* at the Rockefeller Institute had obtained some of the information needed to build the amino acid analyzer from Waldo Cohn; however, comparable work had not been done with nucleic acid derivatives. As part of the Molecular Anatomy Program,

high-pressure, high-resolution chromatographic systems were developed for nucleotides and related compounds, and for sugars[6–9]. The sugar separations were based on earlier work of *Khym* and *Zill* at Oak Ridge[10]. By 1964 systems working up to 5,000 pounds per square inch were in operation, and their application to the study of urinary compounds had begun[8]. An accident robbed this work of much of its early impact. All of the initial work was done with one batch of ion exchange resin which had the unique property of separating bases, nucleosides, and nucleotides all the way out to the triphosphates. Only after the work was well along and published was it discovered that no other batch of resin could be found that would behave similarly. In the years that followed, considerable effort was put into setting this right, and in getting batches of resin synthesized that would make these separations. This portion of the program developed into the so called body fluids analysis program and made a number of interesting contributions in the area of inborn errors of metabolism[11].

High resolution separation of subcellular particles and of soluble or solubilized enzymes brought certain analytical problems into sharp focus. If many fractions are obtained during a separation, many analyses have to be run or the value of the resolution is lost. Further, the analyses must be done rather quickly, or many of the activities will have declined. This suggested that it might be necessary to originate a new analytical system, in which many analyses could be done rapidly and in parallel with the data fed directly into, and managed by a small computer[12]. The history of the centrifugal analyzer had been recorded elsewhere[13], and it need only be noted here that although the system has wide use in clinical chemistry, it is rarely employed in research and has not achieved the sophistication and versatility of which it is capable. For the purposes of molecular anatomy, however, it has provided a useful additional tool.

Protein Fractionation

The most difficult problem in molecular anatomy obviously lies in the area of protein and protein subunit separation and identification. Present estimates of the number of human structural genes range, as previously noted, from 30,000 to 50,000, and many of these protein gene products (PGPs) are posttranslationally modified, making the total number to be dealt with much larger. The question of whether methods can be developed for resolving such a very large number of PGPs depends on whether or not each is sufficiently unique to allow it to be isolated from all others. If each has some unique property, then in theory all can be resolved. The problem of isolating proteins rapidly and efficiently by running them through a series of systems, each of which separates on the basis of one property, ending with pure proteins, has not been solved. Rather, a series of techniques are available to the preparative biochemist who arranges and modifies them to suit each protein. It is interesting that the possibility of isolating any given protein has never been seriously questioned, but the posibility of resolving all of them from a given source has. The two problems are but different sides of the same coin. If there exist unresolvable pairs, then it is not possible to prove by any technique except crystallography, or possibly sequencing, that a given protein is not actually such a pair.

Methods based on selective absorption using either antibodies or immobilized ligands offered the possibility of isolating either individual proteins or groups. Using antibodies, it was possible to subtract groups of proteins from mixtures. For example, one could remove serum proteins from tissue extracts using a device for rapid recycling affinity chromatography which we termed a "cyclum"[14–15]. The cyclum and conventional separations systems could, in theory, fractionate a large percentage of tissue proteins, given sufficient starting material. However, enormous effort over a long period of time would be required. This

was a very discouraging point to reach, especially if completeness is important in practice, and if human cancer, for example, can not be explained in terms of already known constituents. Indeed, cancer might in the end be due to the very last protein gene product discovered. Given the complexity of the problem and the inadequacy of these methods, some new approach was clearly required.

Two-Dimensional Electrophoresis

The provisional answer to the problem of seeing analytically a very large fraction of all tissue proteins appeared quite dramatically in the 1975 publication of *Patrick O'Farell*[16] describing a two-dimensional electrophoretic method that used isoelectric focusing in one dimension and electrophoresis in the presence of sodium dodecyl sulfate (SDS) in the second. This method combined the two highest resolution separation techniques currently available. The method was applied to the separation of proteins and protein subunits that had been highly radiolabeled during synthesis, and detection was by autoradiography[16–24]. The technique was not an easy one to master, and it was difficult to do large number of analyses. However, with it one could "see" nearly all the PGPs of *E. coli*.

Two-dimensional electrophoresis has had a fairly long history, as is shown in Table 1 which lists the major events in the development of present techniques. Only with the work of *O'Farrell,* however, has truly high resolution been reached[16].

It seemed to us that several problems required solution if this method was to find general application in molecular anatomy. The first was to develop methods for doing many analyses in each dimension in parallel, and for using stains instead of autoradiography or fluorography for routine detection. The second concerned the problem of functional assignments. How can these be made in a systematic manner so that one can say with some assurance that a particular spot actually is, or is part of, a known enzyme? The third relates to positional standardization, and to assignment of isoelectric points and molecular weights, while the last problem is concerned with data reduction, that is, with how the separated PGPs may be quantitated, with how positional and quantitational data may be recorded, stored, and used, and with how large numbers of patterns may be intercompared to extract information.

The ISO-DALT System

The development of systems for multiple, parallel two-dimensional analytical electrophoresis is not complete, and we describe here status of the systems now in use. Published descriptions[45,46] are unfortunately already out of date at this writing.

The ISO System: The isoelectric focusing step is done in arrays of precision-bore glass tubes in the so called "ISO" apparatus shown in Figure 1. With this device, twenty isoelectric focusing gels may be quickly cast in parallel, loaded, run, and unloaded. The apparatus was devised with a view to complete automation when this may be required.

The DALT system: The recovered isoelectric focusing gels are briefly equilibrated with a SDS buffer, and then run in the second dimension on slab gels which are now semiautomatically cast in groups of twenty, using the system shown in Figure 2. The slab gels are usually gradient gels which allow better resolution of high molecular weight components, while still resolving low molecular weight ones. The slab gels are run between pyrex plates held together by a silicone rubber binding along one edge, and separated by spacers cemented to one of the plates. No clamps are used.

The first dimension gels are placed on top of the slab gels and are held in place using melted agarose to which molecular weight standards may be added (see below). Since the first dimension gels are thus sealed in place, and are not in liquid form as is often the case in other forms of electrophoresis, the slab gels

Table 1 Major Steps in the Development of High Resolution Two-Dimensional Electrophoresis

Proposed that higher resolution in electrophoretic separations may be achieved by "sieve" or "molecular filtration" effects using gels as supporting media	Smithies, 1955[25]
First two-dimensional electrophoretic separation.	Smithies and Poulik, 1956[26]
Use of acrylamide gel as an electrophoretic support medium	Raymond and Weintraub, 1959[27]
Development of stacking gel concept and suitable buffers	Ornstein, 1964[28]
Use of two unassociated parameters for separation-mobility and molecular weight	Margolis and Kendrick, 1969[29]
Isoelectric focusing followed by electrophoresis	Dale and Latner, 1969[30]
Mapping of tissue proteins for genetic studies-IEF followed by PAGE	Mako and Stegeman,1969[31]
Relationship of SDS electrophoretic mobility and molecular weight	Weber and Osborn, 1969[32]
Use of concentrated urea in gels and development of multiple-slab-gel system	Kaltschmidt and Wittman, 1970[33]
Introduction of SDS stacking gels	Laemmli, 1970[34]
Combination of IEF with SDS-PAGE	Stegeman, 1970[35]
Electrophoresis followed by SDS-PAGE	Martini and Gould, 1971[36]
IEF-SDS-PAGE of non-histone nuclear proteins	Barrett and Gould, 1973[37]
Acid urea electrophoresis-SDS PAGE of nuclear proteins	Orrick, Olson, and Busch, 1973[38]
Beginning of automation-system for casting centrifugally and simultaneously 500 tube gels	Neel, Tiffany, and Anderson, 1973[39]
Discovery that SDS reacts rapidly with proteins in urea without heating	Mets and Bogorad, 1974[40]
High resolution mapping: IEF followed by SDS-PAGE	O'Farrell, 1975[16]; Klose, 1975[41]; Scheele, 1975[42]; Iborra and Buhler, 1976[43]
Optimization of the system using very small samples and autoradiography	O'Farrell, 1975[16]
High resolution analysis of human serum	Anderson and Anderson, 1977[44]
Development of semi-automated ISO-DALT system	Anderson and Anderson, 1978a and b[45,46]

may be run on their sides. This capability allows a novel construction of the second dimension electrophoresis tank, which is divided into three compartments separated by silicone rubber septa with vertical slits. The slab gels are inserted in these slits (Figure 3). The arrangement provides electrical insulation between the ends of the slabs, and uniform cooling in the midle. In the present version of the system, ten gels may be run in parallel. This, however, is not the upper limit of the number that could be run in one tank. When the run is completed, the plates are removed, opened, and the gels are removed for fixing and staining. This is now done in groups of ten per plastic box, and the staining, destaining, and photography are all carried out by hand. Coomassie blue is the stain currently used. Additional developmental work is urgently needed to develop automatic processing systems for producing stained gels quickly, automatically, and reproducibly. There is also a need to develop so called active dyes, possibly based on umbelliferones, which

may be quantitatively attached to proteins and detected by fluorescence. Photography is done with a precision system, containing interference filters, with control of the shape of the Hurter-Driffield curve to give maximum contrast for faint spots, while still keeping the center of relatively dense spots on the curve. Prints are made using an automatic processing system, and a density step wedge is included in each photograph so that the entire system can be properly calibrated.

Mutation Rate Measurements

A basic objective of this work has been to develop the tools required for human genetic screening, both for medical purposes and to determine the background mutation rate in man. There is great concern throughout the world that this "normal" mutation rate may be changing due to exposure to ionizing radiation and to environmental chemical pollutants. As previously discussed[39], a very large number of measurements must be made to determine what the background mutation rate is and how it may be changing with time. This can be done either by making a very few measurements on a very large number of samples, or by making many measurements on a smaller number. About one third of amino acid substitutions give rise to alterations in isoelectric point (i.e., produce charge mutations) that can be observed with the ISO-DALT system. Maps of serum proteins, urinary proteins, red cell lysate proteins, white cell proteins, and proteins from hair follicles, together include over a thousand individual proteins or protein subunits (PGPs). It is important to produce such maps for several thousand individuals, to determine the level of variability normally present, and to discover whether cellular protein variants are as common as are variants of serum proteins. Preliminary evidence suggests that cell proteins are remarkably constant when compared to serum proteins, but much additional information is required.

Since a major objective of this work is human genetic research, it is necessary to develop systems and techniques that allow very large numbers of samples to be analyzed. At present, it is possible to run up to one hundred analyses per day; however, the present staff is not sufficiently large to do this routinely. One soon discovers that the major technical problems remaining lie in the area of standardization and data reduction, as will be discussed subsequently.

The resolution attainable by these system can be greatly improved by simply enlarging the gels. A 30 x 30 cm system has recently been constructed and shown to yield spots as small as the original system, yet distributed over four times the area. No fundamental limits to resolution have yet been approached.

Human Plasma Proteins

The analysis of human plasma proteins using the ISO-DALT system has been presented[44] and will be only briefly reviewed here. A typical pattern for human plasma is shown in Figure 4 and illustrates the effect of overloading due to the large amount of serum albumin present. A drawing of the major proteins is shown in Figure 5, which gives a somewhat better idea of the number of spots which are reproducibly present. The major protein (or protein subunits) are identified in Figure 6, where the error in assignments of α and γ fibrinogen subunits previously made[44] is corrected. When the initial study was published, we did not possess a series of molecular weight standards capable of covering the entire range of peptides present in plasma, and a tentative solution to the molecular weight standard problem will be described shortly. The heterogeneity of light and heavy chains is of great clinical interest. In studies on serum proteins from animal sources, even more discrete patterns of light chain spots are observed as is illustrated in Figure 7. It is interesting that some animal light chains have isoelectric points which are much more acid than those seen in human IgG. This is especially true of rabbit light chains. Recently, we have employed pre-

cipitation with immobilized streptococcal protein A to isolate human IgG for ISO-DALT mapping.

The heterogeneity of many of the serum proteins deserves comment. When the amount of sample used in the analyses is reduced, some of the proteins give very small spots, as little as a millimeter or less in diameter. O'Farrell[16] has previously noted that for a given homogeneous protein the size of a spot is concentration dependent. However, even at low concentrations, many serum proteins do not give single small spots, indicating heterogeneity that we classify as of three types. The first is charge heterogeneity which has been recently noted[16,19,20,44] and which gives rise to spots in rows along the horizontal axis. This heterogeneity may be due to the addition of charged groups such as sialic acid residues, or to deamidation. These alternatives may be distinguished by neuraminidase treatment as is shown in Figure 8 where the haptoglobin β-chain is shown before and after digestion. When the two forms are run as a mixture, it is possible to count the number of sialic acid residues present in the subunits natively.

The second type of heterogeneity appears to be due to the addition of variable numbers of neutral carbohydrate side chains. These are probably synthesized separately and transferred intact to plasma proteins, and could account for the vertical component of the stairstep configuration of many proteins in two-dimensional gels.

The third type of heterogeneity is apparently due to variations in the molecular weight to the transferred carbohydrate chains, and is thought to account for the elongation of some of the spots in a vertical direction, even at great dilution.

Much work remains to be done to explore the use of this analytical system to search for variants and to identify remaining spots. How information obtained may be used clinically is discussed under data reduction.

Salivary Proteins

Human salivary proteins constitute an interesting mixture which is easily obtainable, and which may prove very useful for genetic studies[47]. They present an interesting problem in the form of the double heterogeneity of amylase. In early one-dimensional electrophoresis, this enzyme was found to be present in the form of a series of bands. When a supporting gel was developed that provide some retardation due to sieving, a one-dimensional separation due to both charge and molecular weight resulted, and twice as many bands were seen. When the amylase was fractionated by gel filtration, two peaks differing slightly in molecular weight were obtained. These peaks, on electrophoresis in a sieving gel, proved to be the so called odd and even amylases, and the odd numbered ones are presumed to be higher molecular weight versions of the even amylases[47]. If this interpretation is correct, pairs of amylases exist with the same charge but differing in carbohydrate content and hence in molecular weight. Since in the two-dimensional system used here charge and molecular weight are dealt with separately, one sees two rows of spots directly above each other, confirming that "odd" and "even" amylases actually differ only by molecular weight, and not by charge (Figure 9). In addition, about 50 other spots are seen which now require identification.

Saliva presents a number of other very interesting problems. Its bacteriostatic powers have long been known[48], and are most evident in the pathologies that accompany loss of salivary gland function in man. Nerve growth factor[49] and an epithelial growth factor[50] are found in salivary glands, and from an examination of the pattern shown in Figure 9, it is evident that many other factors are also present which very probably have functions yet to be described. It is interesting to conjecture that the salivary glands, providing as yet unidentified factors important to the normal function of the alimentary tract, may be to the tract what endocrine glands are to

other organs. It is not unlikely also that saliva is important in wound healing, since animals apply it freely onto wounds. It will be of interest, therefore, to examine the effect of each of the many peptides present in saliva on a variety of cells and tissues. There may be good reasons why Mohammed's followers collected his saliva for curative purposes[51].

The fact that so few salivary proteins appear to be subject to the type of posttranslational modifications that give rise to rows of spots would appear to make saliva more interesting from a genetic viewpoint than is plasma.

Hair Follicles

Human hair follicles incubated in small volumes of sterile culture fluid containing ^{35}S methionine, extracted with SDS, and analyzed with the two-dimensional ISO-DALT system using autoradiography for detection, show a very large number of spots are seen as is shown in Figure 10. Over two hundred spots may be counted, only a very few of which have been thus far identified.

White Cells and Red Blood Cells

We are expending major efforts on mapping human white cells using both radiolabeling and staining. A map obtained using the whole white cell mixture is shown in Figure 11. The cells are being currently separated into T and B cells, and their maps will be compared with those of leukemic cells of various types. In addition, studies on the localization of known enzymes in red cell lysates are now rather far advanced[52].

Urinary Proteins

The concept that injured or diseased cells leak proteins into the circulation and that many of these proteins pass through the kidney into the urine is supported by both animal model studies[53-59] and by observations in man. Histuria following muscle damage was the first instance described[60]. Measurement of the level of liver or heart enzymes in serum as indicators of damage to these organs is widely employed; however, there has not been a systematic method for searching for indicators of tissue-specific damage that are of the right molecular weight to pass from plasma through the kidney into the urine. If such indicators could be found for each organ and possibly for tumors as well, an interesting spectrum of new diagnostic possibilities would be available. We have therefore asked whether urinary proteins could indeed be resolved by the ISO-DALT system, and whether interesting new proteins appear in disease. While many serum proteins are found in low concentration in urine in a relatively undamaged form, the view has prevailed that other proteins, if present, are probably degraded. It has been of interest, therefore, to examine normal urinary proteins and to see if incubation of them in concentrated form would give an indication of breakdown. As shown in Figure 12, incubation of concentrated urinary proteins at 37° for one hour produces no detectable effect. We have examined urine of human patients in a number of pathological states. An example of the type of change seen is given in Figure 13 which shows the wide spectrum of proteins present in the urine of a fatally burned patient. Our general conclusion from this preliminary survey is that many proteins are indeed found in urine that appear to have diagnostic possibilities. Rather than concentrate on their study, however, it has appeared to us more logical to map organs first, marking off all the spots common to all organs; then gradually and systematically to identify spots common to organ groups or germ layers; and finally, to identify those associated with one tissue or cell type. These tissue or cell specific spots then can be individually sought in the urine of patients having a disease of the organ in question. We have adopted this approach.

Soluble Phase Proteins of Rat Liver

Our intention is to map normal human tissue proteins to provide a baseline for the study of disease related change. For orienting studies,

however, it is necessary to use animal materials. Initially we were interested in seeing whether very minor difference between cells or cell fractions could be detected. Since it was known that there was at least one protein different between the soluble proteins of female and male rat liver[61], we have examined these with the results shown in Figures 14 and 15. Not one but two major proteins were discovered that were present in the male but almost totally absent in the female. In addition, one protein was found which was characteristic of female liver but not of male. In subsequent studies on subcellular fractions, spots characteristic of different organelles will be described. The patterns shown in Figures 14 and 15 are superimposable and give an indication of the reproducibility of the ISO-DALT system.

Pattern Comparison

Comparisons between patterns, such as those shown for soluble proteins of rat liver, are relatively simple when the gels are superimposable. However, even for such patterns, minor differences may be missed. A more convenient way to make comparisons between similar patterns is to use an optical comparator, such as is used for aerial mapping, which allows two images to be superimposed by the use of an optical train incorporating a variable magnification lens and a rotatable cylindrical lens to compensate for differences between gels. By alternately illuminating the two patterns (flickering), small differences may be noted. Two important problems are not solved by this procedure. The first is comparison between patterns where many or most spots are different, and the second is the problem of intercomparing large numbers of patterns as, for example, when one is asking which spots are common to a large number of different patterns from different organs and cell fractions.

The intercomparison of quite different patterns may be done by running each sample separately and then running the mixture as one additional analysis. An attempt is then made to find all of the spots in both patterns in the third gel pattern. In this manner, all spots can, in theory, be divided into shared and not shared. In practice, this is rather difficult to carry out, especially if a large number of tissues are to be examined. One way around this problem is to develop internal standards that can be added to any mixture. If these are for molecular weight, there is the added advantage of being able to make tentative identifications on the basis of this variable.

Molecular Weight Standards

It is customary in second dimension separations to add a mixture of known molecular weight proteins to run along one edge as a reference standard. However, the precision of this type of standard is not comparable to the precision inherent in the patterns obtained. In addition, many of the proteins used as standards do not give as sharp lines or spots as do freshly prepared tissue proteins, and it is difficult and expensive to prepare standards that adequately cover the molecular weight range of interest here (from approximately 8,000 to 250,000 daltons).

To give each PGP (and each separable post-translational modification of it) a unique identification number, two methods are proposed. In the first, a standard reference gel is made for the PGPs of each cell type or tissue, and each spot is referred to a set of fixed coordinates. Images from subsequent gels from the same tissue are then modified by computer to superimpose the majority of spots. Remaining qualitative and quantitative differences are then recorded. The second approach assigns an isoelectric point value and molecular weight to each spot for identification and referencing. We delay routine use of the second method for cataloging purposes because, in the absence of precision standards, the values initially assigned will have large errors. However, for identification of peptides from literature data, and for comparison of patterns that have

very few spots in common, it is important to have good provisional intercomparison standards available now.

We have examined seed extracts, serum, saliva, urinary proteins, and a variety of tissues seeking a reference mixture that gives sharp lines with SDS electrophoresis, is reproducible, and may be easily obtained. Whole homogenates of rat heart or mixtures of whole homogenates and purified myofibrils were chosen because they yield approximately 86 bands, with the majority at relatively evenly spaced intervals. A few are present in considerable excess and serve as guides, dividing up the pattern in a convenient manner. In addition, heart tissue provides a source for several mixtures which may be used as standards of varying complexity. If the whole homogenate is used, the serum albumin and hemoglobin bands are readily apparent. By perfusing the heart, these may be eliminated from the whole homogenate pattern. The homogenate may be centrifuged at high speed to give a pattern containing all particulate proteins almost free of soluble proteins. The particulates in turn may be fractionated selectively. In the simplest case, myofibrils may be isolated and used alone[62], giving the proteins of the contractile system which are well known for skeletal muscle, and are being resolved for heart. While rabbit skeletal muscle proteins have been more completely characterized, heart proteins provide a better series of bands.

The tissue reference standards are added to the agarose gel used to seal the first dimension gel in place. They give reference bands which stretch all the way across the gel so that each spot may be accurately compared with them. Since both these "streak standards" and the isoelectric focusing gel are above the stacking gel for the second dimension SDS electrophoresis, they are compressed or "stacked" together and for all practical purposes appear to start electrophoresing from the same line. This technique solves one problem in reference standards which is lack of precision due to bowing or distortion of the gel pattern because both the experimental and the reference patterns are equally distorted.

Part of the standardization problem is illustrated by Figure 16 which shows a series of commercially available standards compared with rat heart muscle proteins prepared as described. Even at great dilution, these particular standards did not give sharp bands. To begin to calibrate the heart mixture, we made a series of single purified protein preparations and ran them as shown in Figure 17. Regardless of the dilution used, transferrin, catalase, and ovalbumin did not give sharp bands. For further comparison, our rat heart mixture was run against rabbit psoas muscle myofibrils and known proteins provisionally identified from *Porzio* and *Pearson*[63] (Figure 18). The identification of the latter needs additional confirmation beyond a positional one. Using the data thus obtained, we have numbered the rat heart pattern bands obtained using a mixture of 75% myofibrils and 25% whole homogenate (Figure 19) and then plotted these against molecular weight to give a calibration curve. Since a gradient gel was used, the curve would not be expected to be linear. The molecular weights for the major lettered bands are shown in Table 2. When

Table 2 Provisional SDS Molecular Weights of Major Bands in Rat Heart Standard Mixture (75% Myofibrils, 25% Whole Homogenate). Values Read Off of Figure 19

Band	Molecular Weight
B	15,400
C	17,200
D	21,400
E	26,200
F	29,200
G	35,700
H	40,100
I	42,600
J	52,100
K	64,200
L	102,000
M	147,000
N	200,000

the interpolated molecular weights of all bands are read off, it is discovered that at the lower molecular weight end of the spectrum, the molecular weight differences amount to about 2.5 amino acids. Hence it is important to examine this standard mixture much more thoroughly, and to calibrate it against as many different proteins as possible, with emphasis on those whose molecular weights are very accurately known.

Because there are a number of artifacts associated with the determination of molecular weights by SDS electrophoresis[64, 65], values determined using rat heart standards (or any others) cannot be considered to be accurate. The objective here, however, is not accurate molecular weight determination but rather accurate and reproducible positional location. Provided that the 86 bands obtained from rat heart homogenates are reproducible with animals from different sources and with the experimental procedures used in other laboratories, it should be possible to describe a protein as, for example, having a position 0.6 of the way between lines 31 and 32, in which case it would be described as 31.6.

The use of streak standards is illustrated in Figures 20 and 21, where human serum is run against two different versions of the streak standards described here. These figures illustrate the reproducibility that can be obtained with tissue reference standards.

It will be necessary to develop similar standards for isoelectric focusing, and some effort in this direction is now being made. When this effort is complete, it should be possible to do a larger fraction of identifications positionally.

Muscle Proteins

When a new high resolution system is used to examine even previously well studied tissues, one may expect to discover new things, and our study of muscle has been no exception. While the detailed identification of a number of muscle proteins, including enzymes, will be described elsewhere, we note here an unusual pattern of spots that has been observed with rat and rabbit heart, and with rat, rabbit, and human skeletal muscle. The pattern is shown in Figure 22 and in higher magnification in Figure 23. The rows of spots in the center of Figure 23 are at the limits of resolution of the system and provide a good test of it. The horizontal rows appear to represent incremental charge modifications, and the distance between the rows is indicative of a systematic molecular weight difference between the rows. If the charge differences are due to the systematic addition or subtraction of large carbohydrate complexes or peptides, then the rows should be tilted, as a few of them appear to be. For the most part, however, the charge and molecular weight modifications appear to be dissociated as if the fundamental molecule could have charge increments and molecular weight increments separately. These patterns are obtained with tissues removed with great speed, homogenized very rapidly, and then very quickly heated in SDS, and are very reproducible. It is interesting to speculate where in the contractile system a series of molecules graded both with respect to charge and molecular weight might be required. It is unlikely that this spectrum of proteins could be discovered by any other than the two-dimensional system described. In some respects, the proteins illustrated in Figure 23 resemble invertebrate paramyosin and may represent a hitherto unreported analogous protein in mammals.

Data Reduction

The studies described above demonstrate that it is possible to map tissue proteins systematically and to intercompare them in detail. We have not reviewed here the methods being developed to assist in the identification of spots and to assign enzyme or other functions to them based on the use of the centrifugal fast analyzer and on separations systems now under development. We have also not reviewed the problems of getting higher molecular weight components into these gels, and of

getting sharp isoelectric focusing at the basic end. These problems will doubtless find several solutions. What is evident, however, is that a truly enormous amount of data can now be obtained, and that the objective of molecular anatomy of "seeing" cells and subcellular components at the molecular level can now be reached. The central problem then becomes that of managing and reducing the data.

As described above, we have begun with photo-optical methods for visual intercomparison of gels and gel photographs, and find them very useful for the solution of simple problems. However, they do not solve the problem of the quantitative analysis and intercomparison of hundreds of spots on thousands of gels. This problem can only be solved with computerized image analysis systems.

Using a calibrated photographic system and an Optronics high-speed drum scanner, each pattern is scanned and reduced to 2–4 million pixel density values with density being expressed on a scale of 1–256. This information is obtained in the form of a raster, and each line is initially processed separately to determine where each spot begins, where its maximum or centroid is, where it ends, and whether the curve obtained represents one peak or can be resolved into two or more. All the maxima thus obtained are then scanned in the other direction and the location of the maximum for each spot determined. In addition, the absorbance (volume) of each spot (peak) is determined and three values are placed in memory: the x and y coordinates and the integrated absorbance. Using the standards added to the orginal sample and the calibrations built into the system, the absorbancies may be converted into micrograms or nanograms of protein.

For identification studies, the data are automatically plotted out on six foot square charts with each spot represented by a circle or elipse whose size is proportional to the amount of protein present, and with the coordinate values and quantitation printed out next to each spot. These charts are invaluable for recording experimental data as it is obtained. Examples of such charts are given in Figures 24 and 25.

The questions we wish to ask, however, concern the alteration of gene expression with cell line and with time, and the extent to which the function of different genes is coordinated, i.e., the extent to which they appear in sets. We would also like to know how gene expression (and all steps in RNA and protein synthesis including posttranslational modifications) are altered in response to drugs, mutagens, hormones, and pollutants. Obtaining the necessary information requires that small intergel distortions be corrected and that PGPs that are identical on two different gels be given the same coordinates. While completely computerized systems for making such corrections can be built, we have concluded that the initial matchings are best done by a human operator interacting with the computer. Our data reduction system is therefore being rebuilt to allow data to be processed more rapidly to the point where a spot location list is in memory. The pattern will then be displayed as a two-dimensional map in one color on a cathode ray tube while the reference standard will be displayed in a second color. The operator then manually identifies the major spots that should be superimposed, and the computer adjusts their coordinates so that they do. As this is done across the pattern, short range linear corrections are applied to all spots in a given area to superimpose them. Where there are difficulties, the original prints or negatives are available for intercomparison to see if errors of identification or errors due to extraneous marks on film have occurred. When correct superimposition is assured, and the distortional corrections made, the corrected data are placed in memory. PGP data lists may then be searched to answer specific questions such as, which are common to all tissues, which are unique to one tissue or cell, which always appear together, and which never do. We have previously examined in some detail how these questions apply to tumor

cells[106, 107]. In addition, a new description of tissue culture cell types will be available. From this may arise the possibility of identifying nuclear signals or controls involved in differentiation. For example, if acidic proteins of the nucleus are involved in the control of genetic expression, will it be possible to find cytoplasmic sets which always appear when a specific acidic nuclear protein is present?

Conclusions

Molecular anatomy (and later molecular pathology) can now be established as a science in its own right with a unique set of questions, objectives, and techniques. The development of this field requires a wide range of contributions from many scientific disciplines. Ultimately, a body of techniques will be developed which will be added to those of contemporary pathology and clinical chemistry, and one or more new specialties will arise.

We are now confronted with an interesting and instructive problem, which is, how should the systematic exploration of the molecular anatomy of human cells actually be done? The present model for biomedical research is largely built on the individual grant system which at present has two characteristics. The first is acceptance or rejection of a detailed plan, which requires that innovation already have occurred. The second is that anticipatory support for innovation is largely ruled out, as is quite evident from a detailed analysis of the support history of the work described here. Other forms of support suffer from "prior era" management difficulties – the work lies completely outside the personal experience of anyone who has left research to assume management functions, Thus, it is extraordinarily difficult to arrange to do work which directly relates new basic and applied research to human disease, even when the goals and the possibility of success are quite evident.

Molecular anatomy has a notable antecedent which deserves comment at this symposium. Identification of human blood proteins and the quantitation of many of them can be done in almost any major hospital throughout the world. One might think this capability was directly due to the vast amount of research on blood proteins done in various laboratories throughout the world, and that somehow the results of that work almost automatically reached the patient. The fact is that a disproportionately large amount of the credit for this accomplishment is due to work at the Behring Institute, as indicated from the lists in Table 3. The systematic analysis, isolations, and identification of human tissue proteins will require the same sort of dedication, organization, and effort previously devoted to blood proteins. In addition, it will be necessary to make monospecific antibodies against each tissue protein to allow a final check on localization in a tissue since each tissue contains more than one cell type, to make more certain identification of spots which appear in the same place in different tissues, and to allow the development of specific competition assays for diagnostic studies, for example, on serum or urine samples. It appears that the problem of producing monospecific antibodies will be best solved by using fused myeloma-plasma cells in culture, and the problem of preparing them, of sorting out producer cells, and of finding those producing the specific antibodies required will doubtless benefit from mechanization and automation.

It is now technically feasible to analyze human cells for the majority of protein gene products and to actually write a molecular anatomy based on the results. This revolutionizes both our concepts of research in human disease, especially cancer, and the way in which such research will actually be done.

This work has been made possible by the competency and dedication of the staff of the Molecular Anatomy Program which is organized as follows: erythrocyte proteins, *Dr. Jesse Edwards;* muscle protein, *Dr. Carol Smith;* tissue proteins, *Dr. Fred Giere;* ISO-DALT operations, *Sandra Tollaksen;* centrifugal fast analyzers, *Sharron Nance;* tissue culture studies,

Fig. 2: Multiple gradient-slab gel casting apparatus for casting 20 gels. This is a recent modification of the apparatus described in reference 46. The two chambered gradient apparatus (G) generates a gradient whose shape may be readily modified. The gradient flows through magnetic mixer (M) into a V-shaped input (I) where the velocity of flow is decreased in preparation for flow into slab gel casting chamber S which holds twenty pairs of hinged slab gel plates. The slab gel holder can be rotated so that the spaces between the slab gels form a V shaped compartment for additional flow control. When the holders are nearly full, the chamber is slowly rotated through the position shown to horizontal. Acrylamide is displaced out the input lines with displacing fluid P. Detailed operation is given in reference 46.

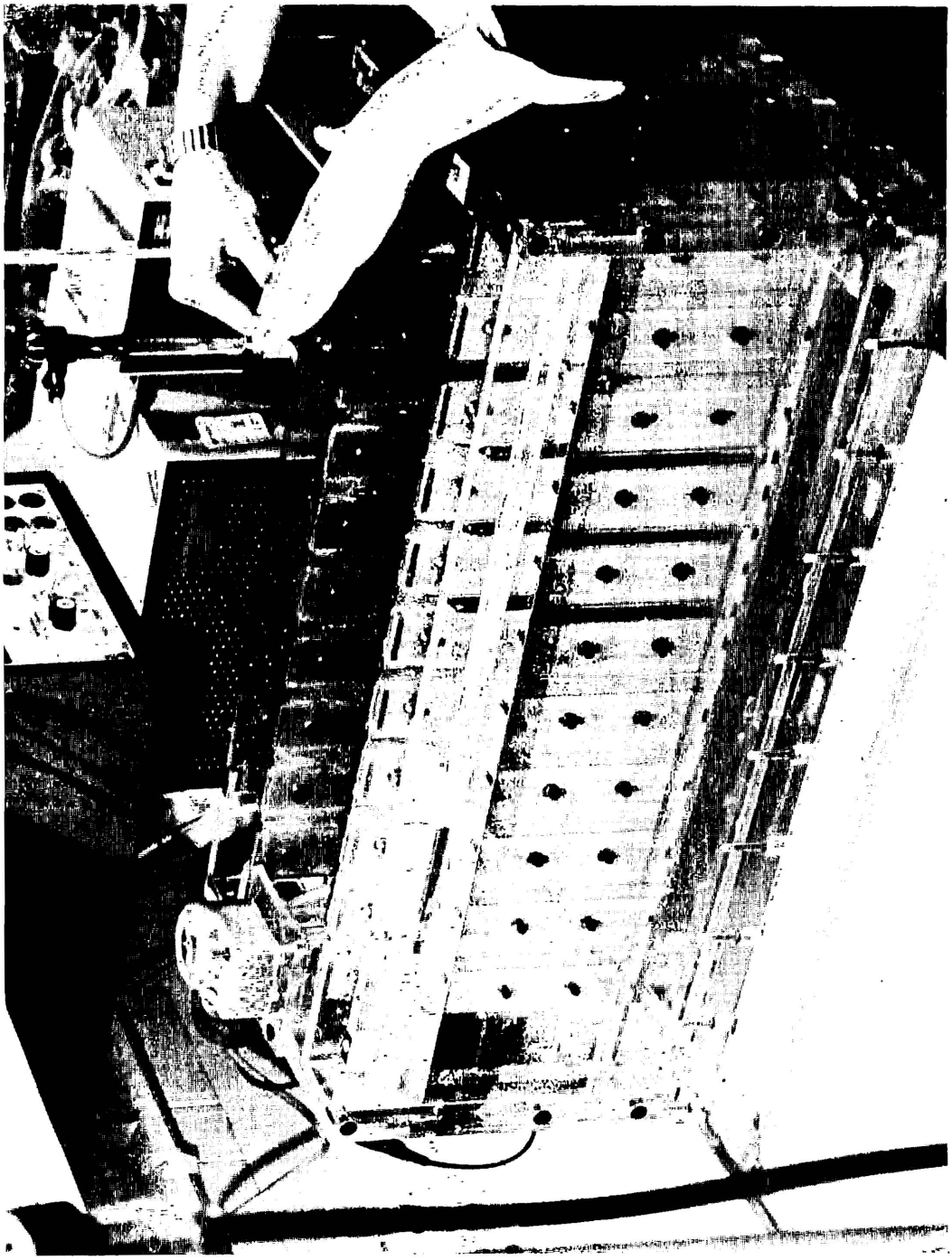


Fig. 3.: Dalt tank for electrophoresis of 10 slab gels simultaneously. The tank is divided into three compartments which are the electrode compartments on either side, and the cooling compartment in the center. The slabs are run horizontally, and are inserted into slits in silicone-rubber septa as shown to insure electrical isolation of the two ends of the gels. The buffer in the center chamber is rapidly circulated by a small pump and cooled by cooling coils. (Construction and operation described in detail in reference 46.)

Fig. 4: Two-dimensional separation of human plasma proteins using the ISO-DALT system. Note that the low pH (acid) end of the isoelectric focusing gel is the left. While some workers have placed the acid end on the right, we have adopted the convention of placing it on the left so that if actual pH values are plotted, the low numbers would start from the left in keeping with the universally adopted conventions of Cartesian coordinates. (From Proc. Nat. Acad. Sci. U.S. *74*, 5421, 1977 with permission.)

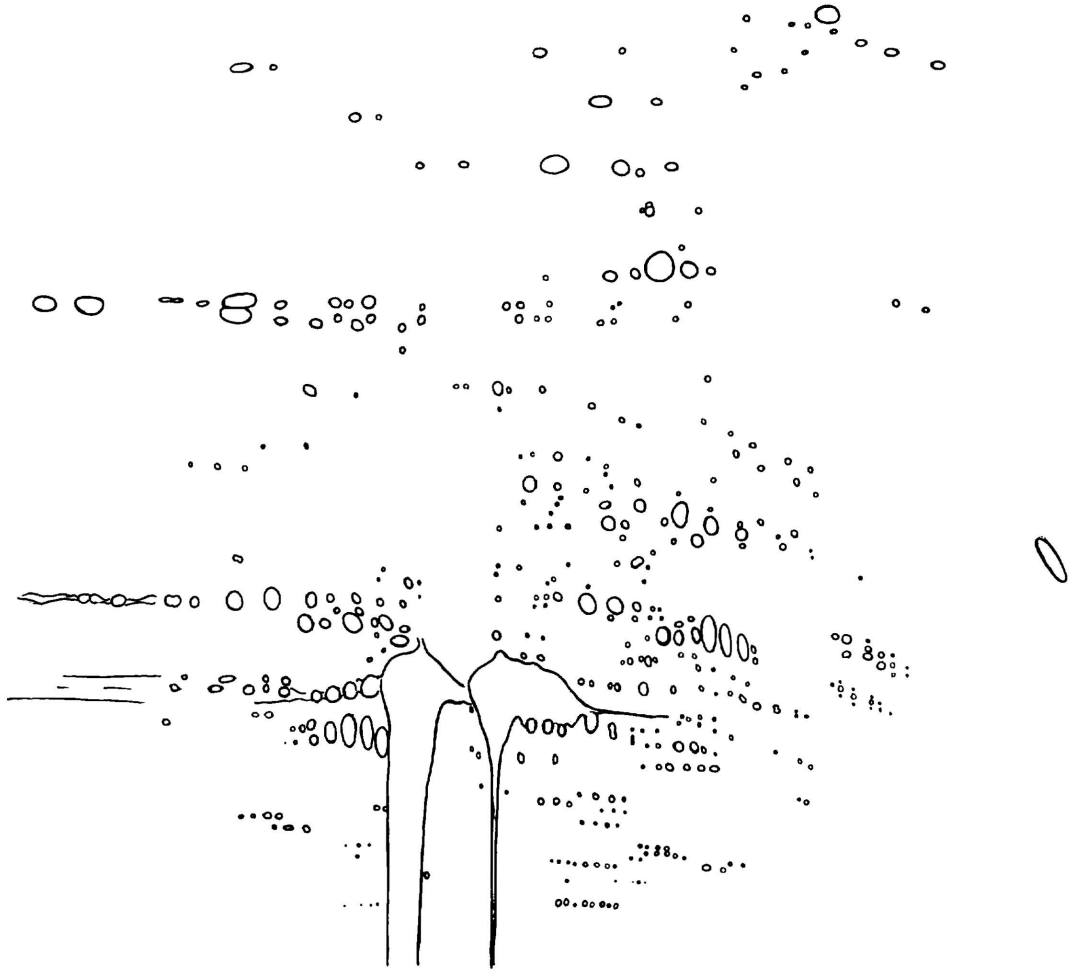


Fig. 5: Drawing of spots from Fig. 4 made using an aerial mapping plotting system.

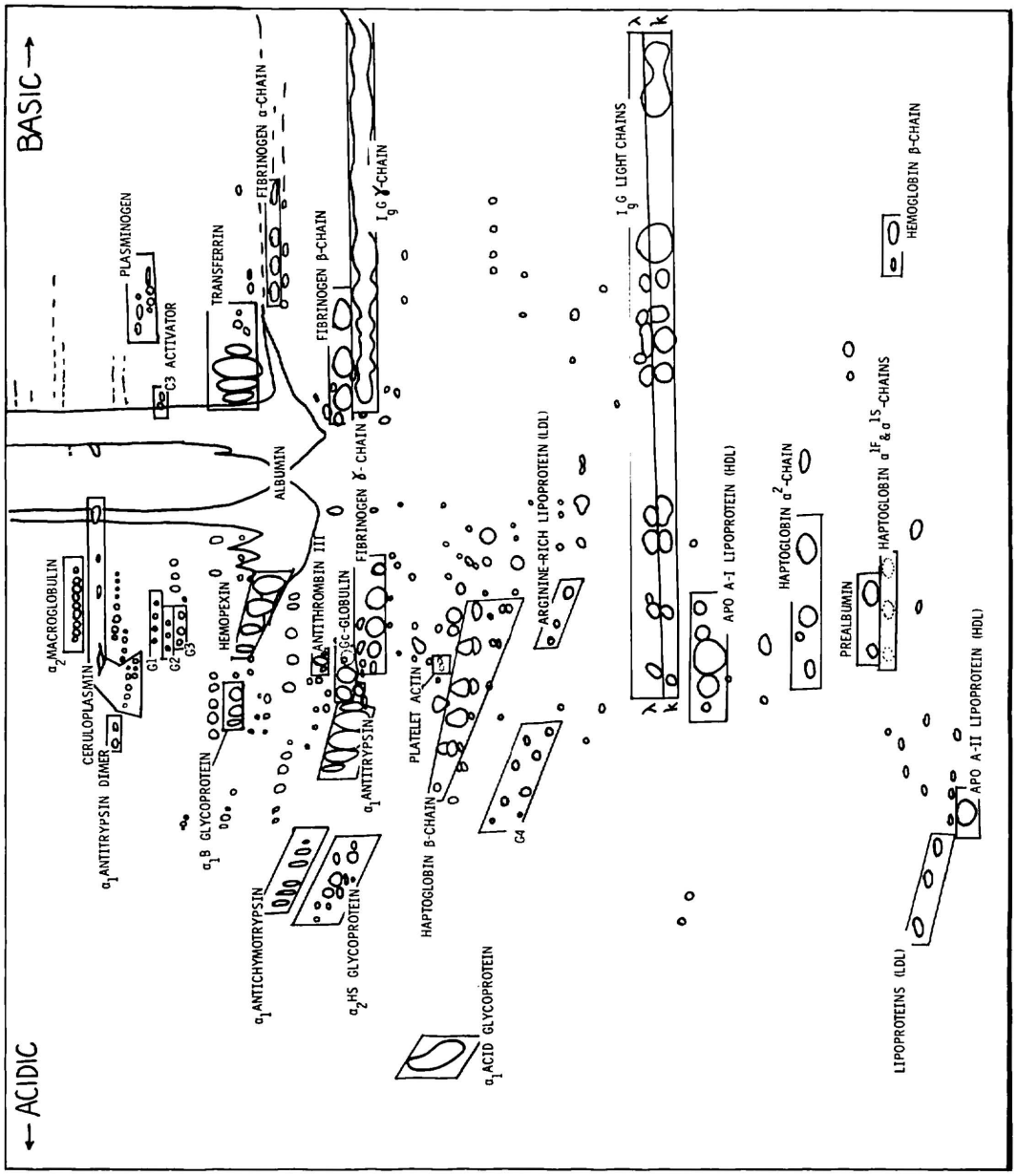


Fig. 6: Identification of spots in two-dimensional pattern of human plasma. (Revised from reference 44.)

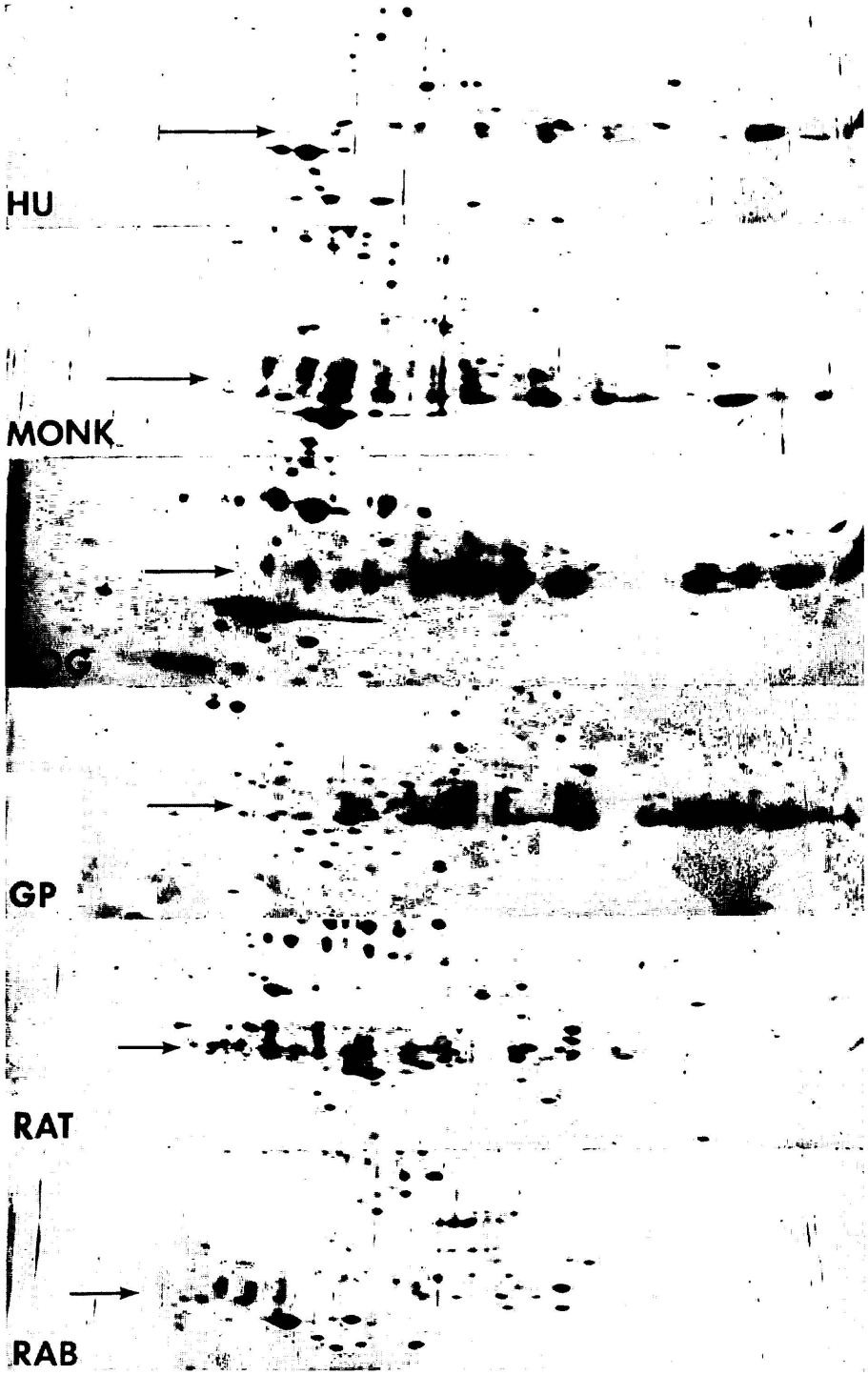


Fig. 7: Light chain patterns obtained from sera of different species. Note that rabbit light chains are much more acid (further to the left) than the light chains of most other species. Shown from top to bottom are human, monkey, dog, guinea pig, and rat light chain regions.

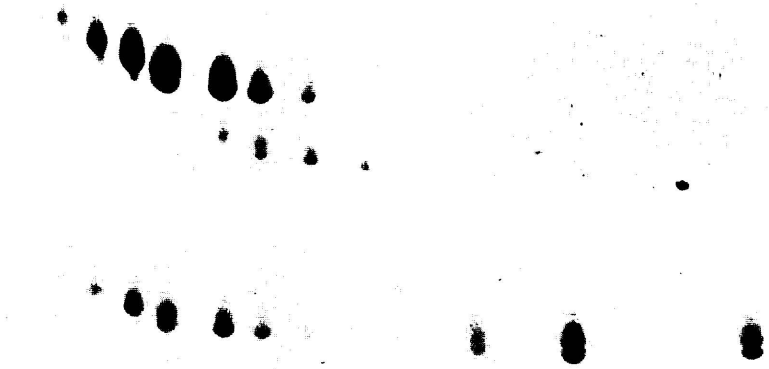


Fig. 8: Sections of two-dimensional gels showing the haptoglobin β-chain and the effect on it of neuraminidase digestion. The upper panel shows the Hp β-chain, and the lower panel a superimposition of the digested and undigested forms. The fact that the predominant desialated form (on the right) displays a residual stepwise molecular weight (vertical) heterogeneity suggests that neutral sugar structures are added (or not) as finished units.

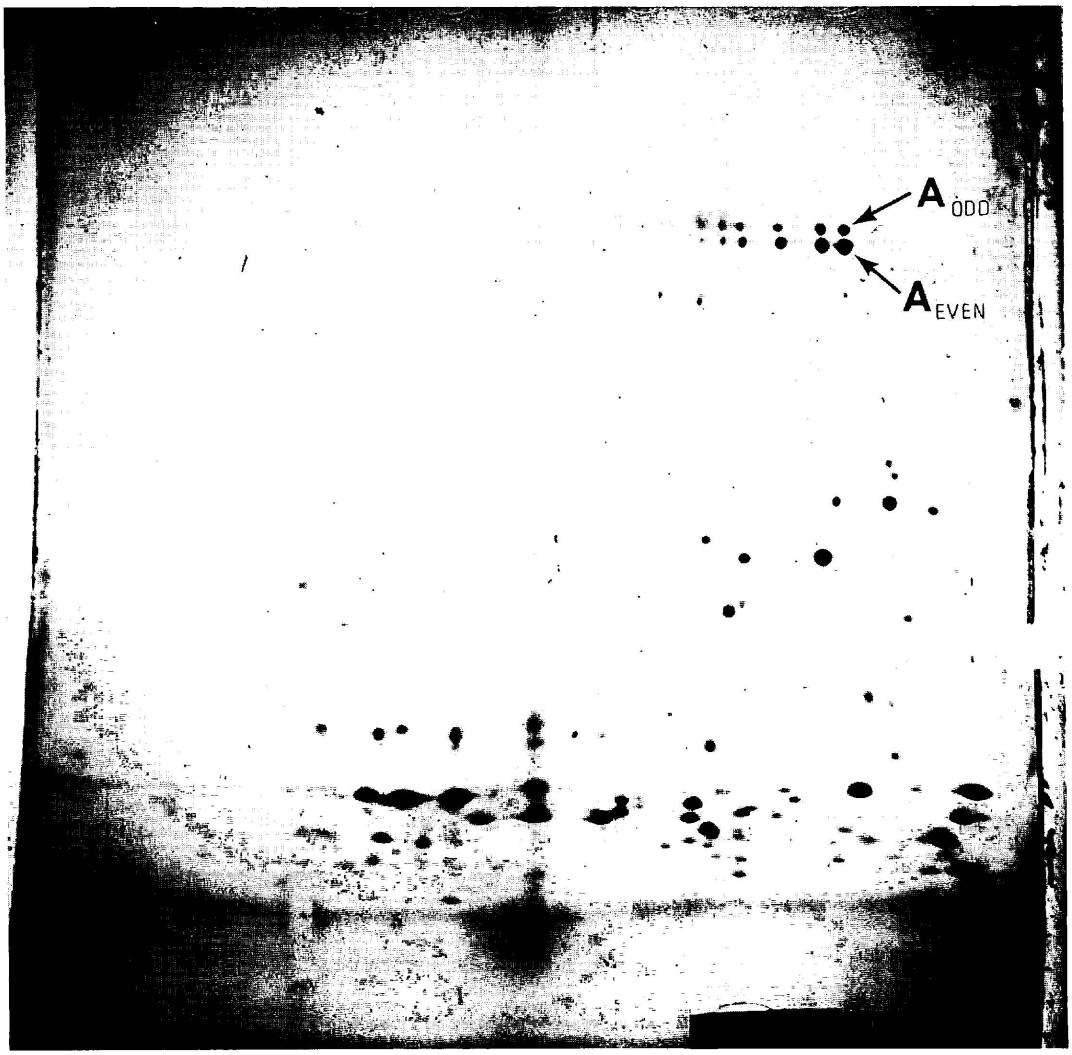


Fig. 9: Two-dimensional pattern of concentrated human salivary proteins. Labels indicate the "odd" and "even" amylases. Note the variety of low molecular weight PGPs along the bottom.

Fig. 10: Autoradiograph of a two-dimensional gel of proteins synthesized by a single human hair follicle during three days of *in vitro* labeling with ^{35}S methionine. Over two-hundred discreet spots are visible on the original autoradiograph (see also Figure 25).

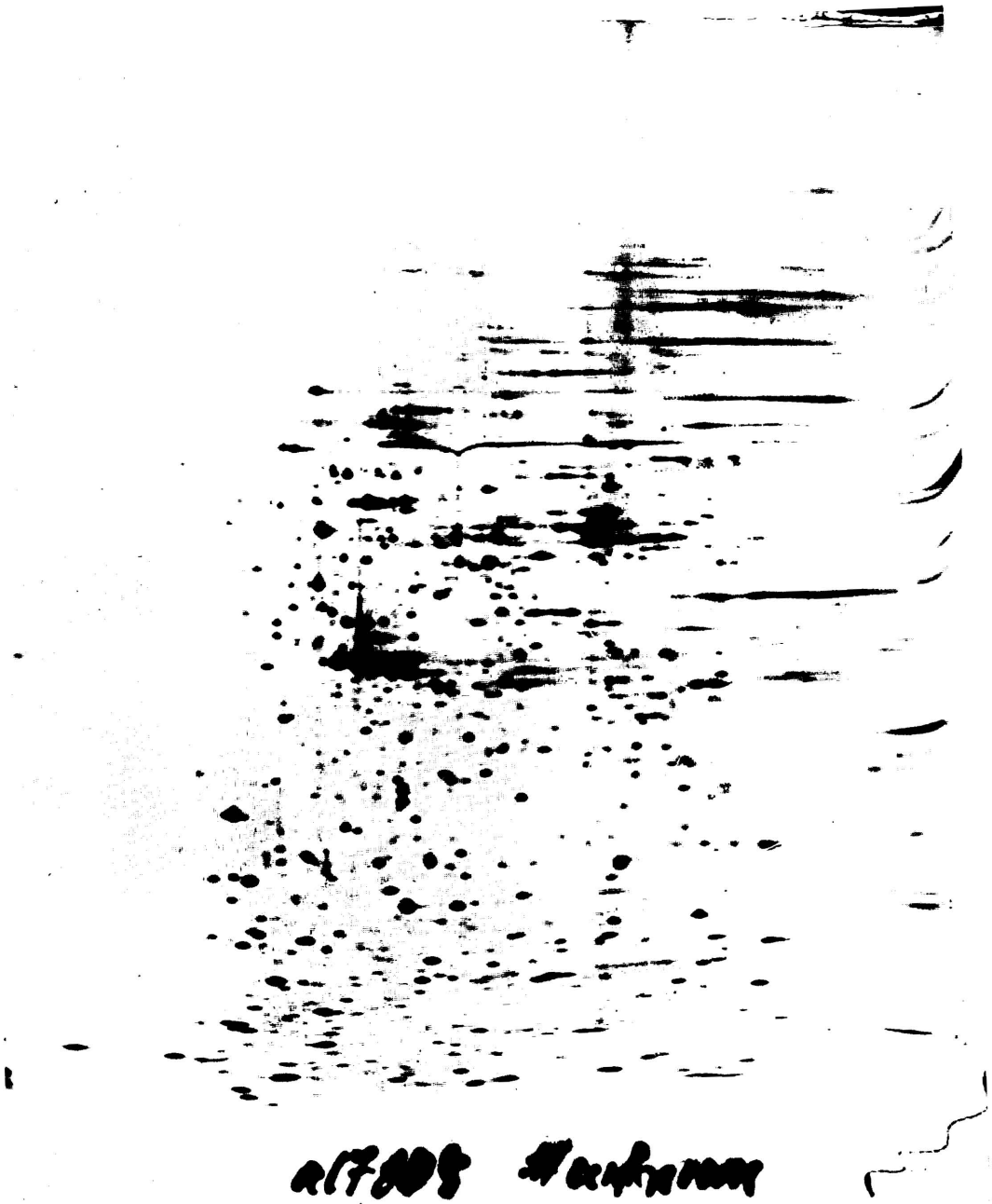
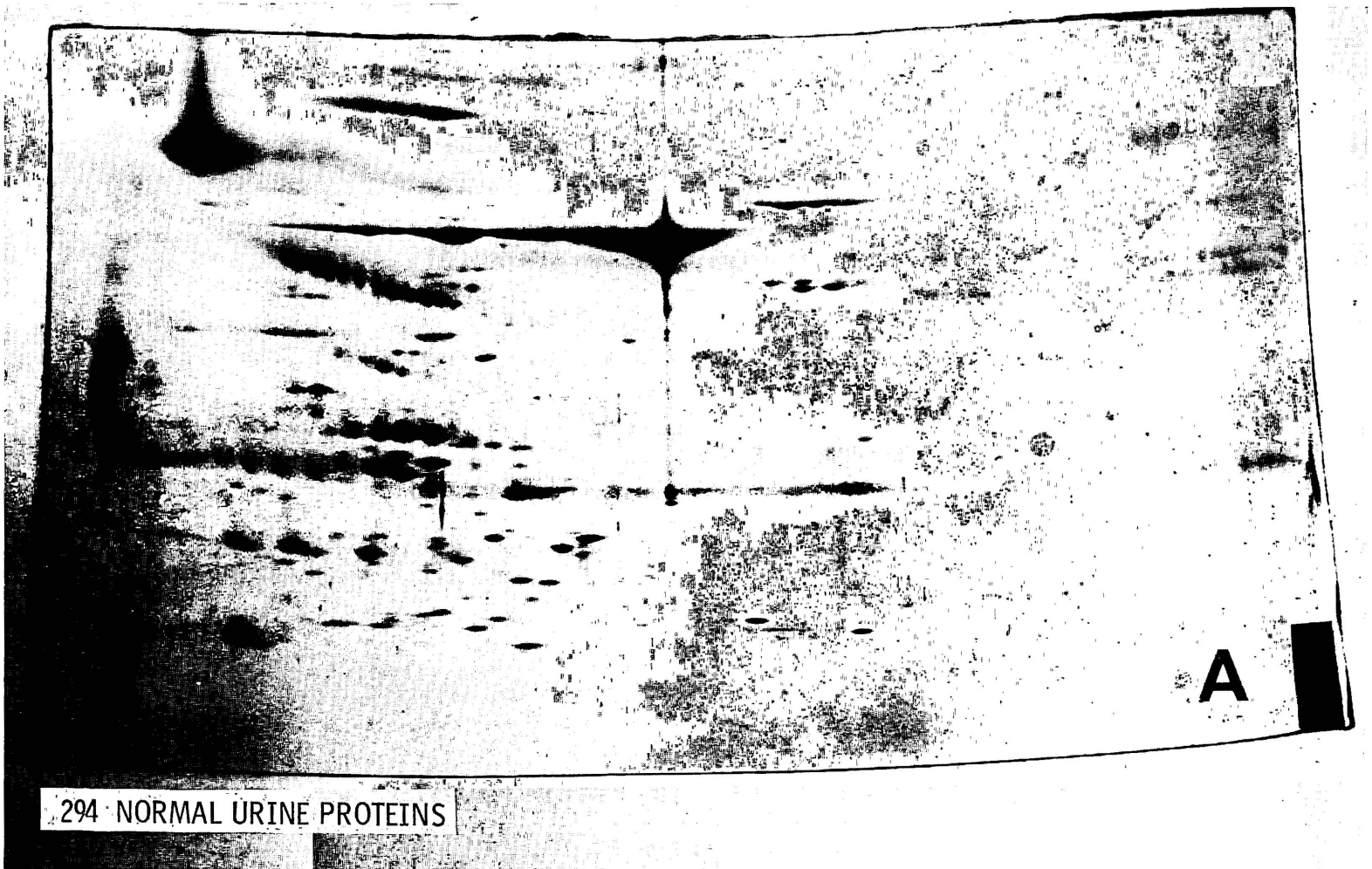


Fig. 11: Autoradiograph of a two-dimensional gel of labeled human PHA-stimulated lymphocytes.



294 NORMAL URINE PROTEINS

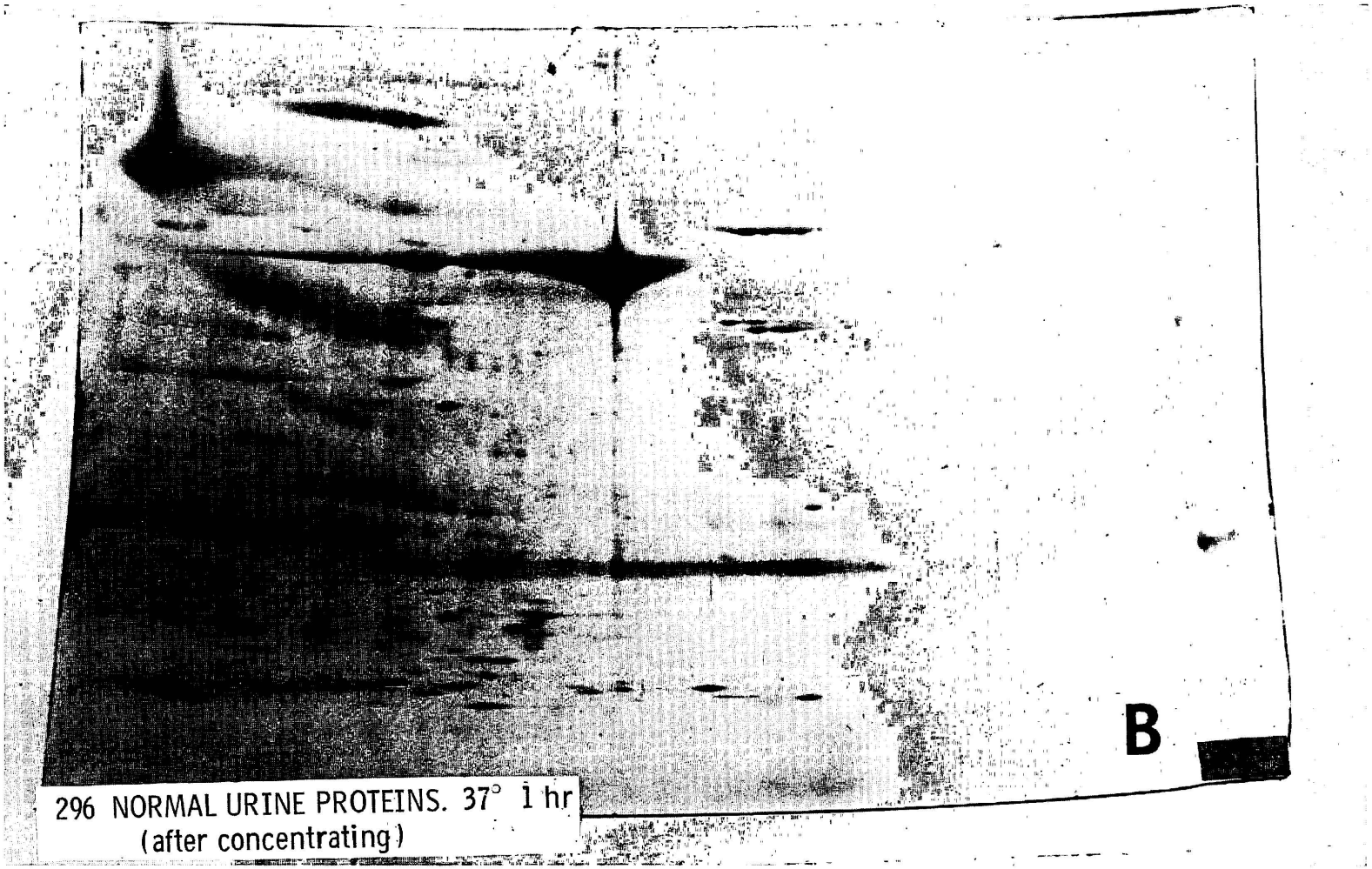


Fig. 12: Two-dimensional gels of normal human male urinary proteins. The proteins were concentrated ~ 1000-fold by dialysis against Carbowax 20M. Sample A was run directly while B was incubated first for 1 hr at 37° C after concentration to detect proteolysis. None is observed.

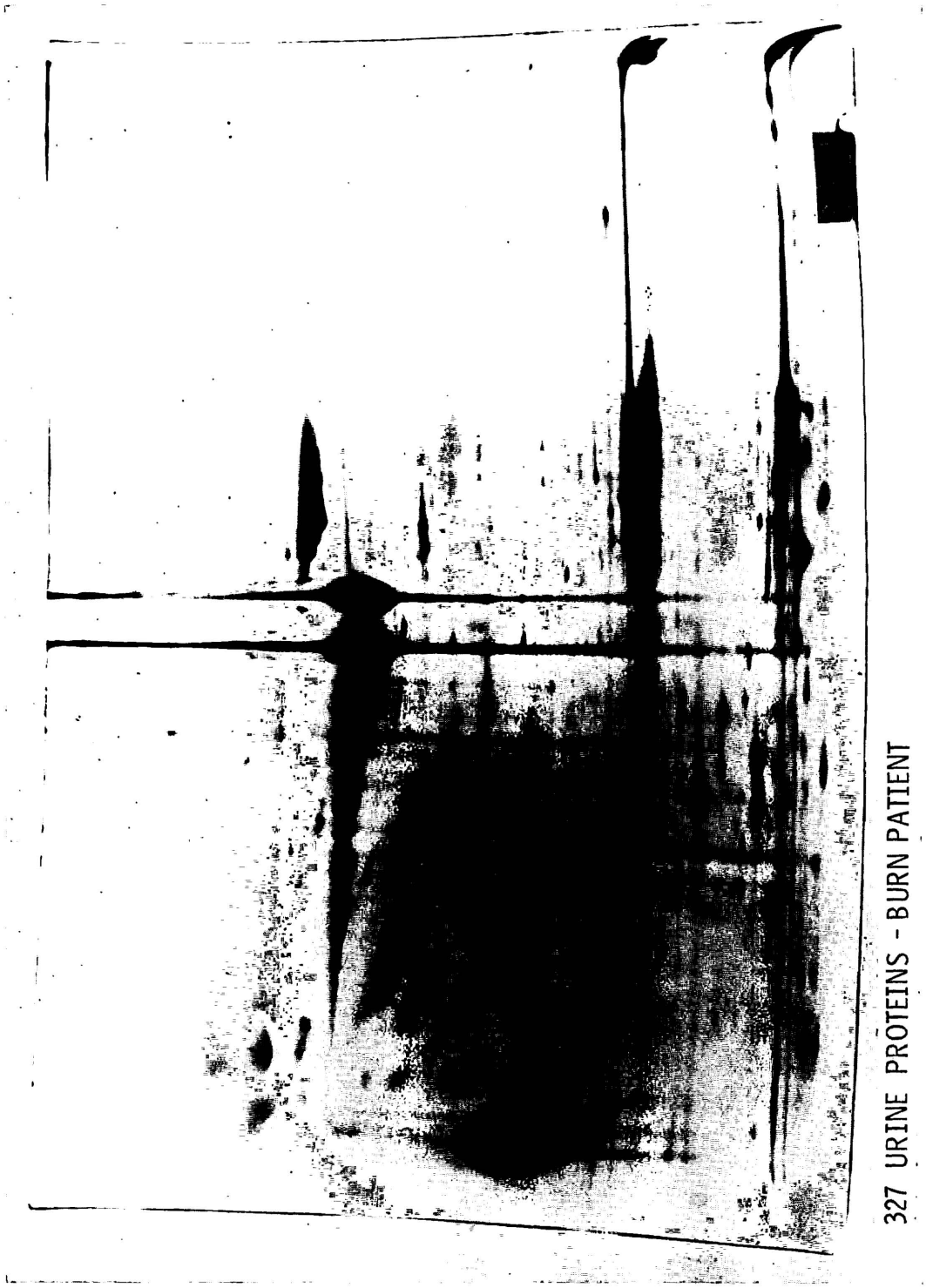


Fig. 13: Two-dimensional gel of urinary proteins excreted by a patient with severe burns.

Fig. 14: Two-dimensional gel of female rat liver soluble proteins. The arrow indicates the female-specific protein.

Fig. 15: Two-dimensional gel of male rat liver soluble proteins. Arrows indicate the two male specific proteins. All proteins not indicated by arrows (several hundred spots) appear always to be common to both male and female.

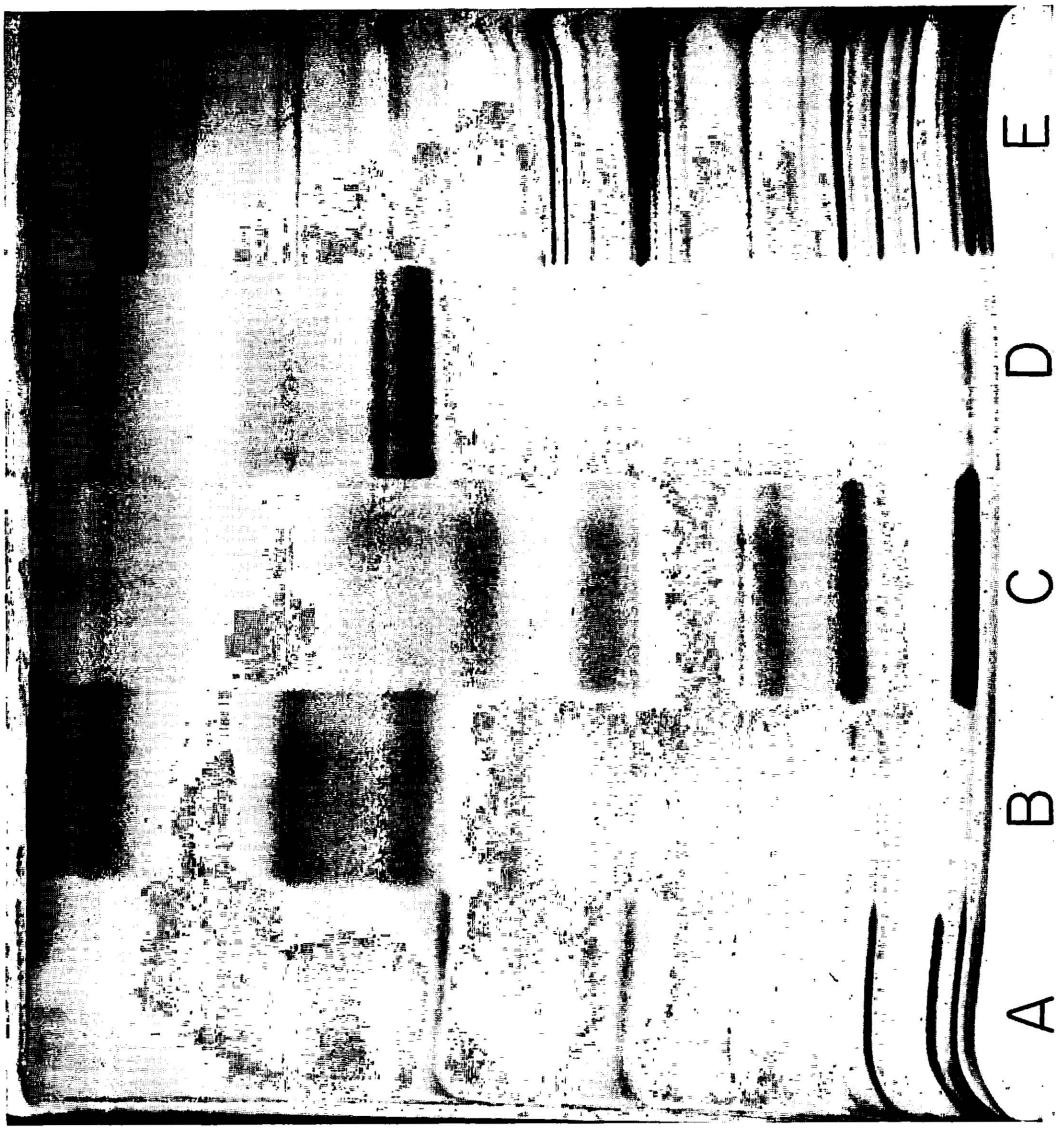


Fig. 16: One-dimensional SDS electrophoresis of four commercially available molecular weight standards compared to rat heart homogenate preparation. A. Dalton IV standards containing (top to bottom) bovine serum albumin (BSA), ovalbumin, pepsin, trypsinogen, β-lactoglobulin, and egg white lysozyme. B. Cross-linked BSA. C. Cross-linked hemoglobin. D. Hemocyanin. E. Our rat heart homogenate standard. (A-D obtained from Sigma Chemical Co.).

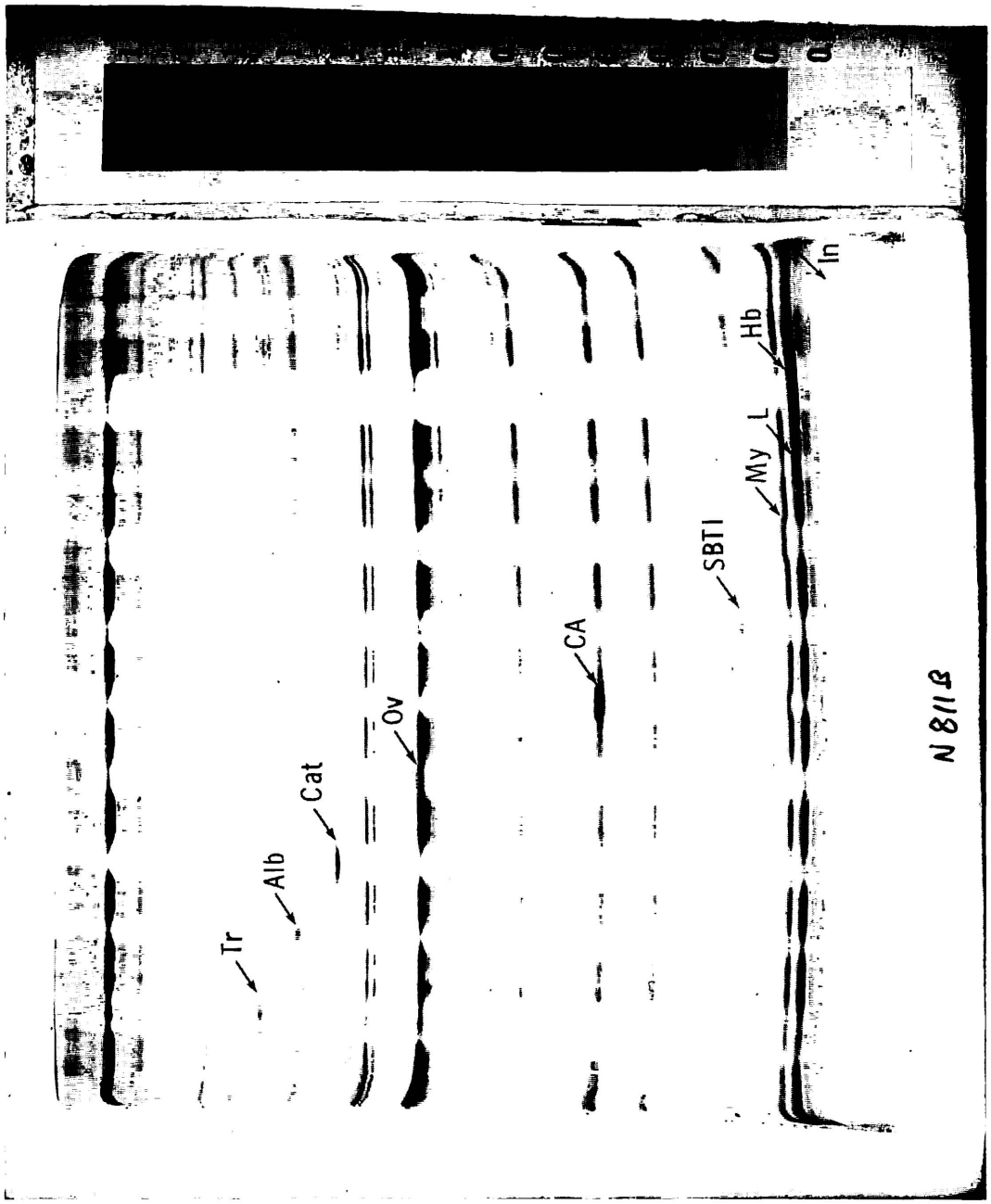


Fig. 17: Various purified proteins run against our rat heart protein standard (alternating background lanes). Tr, transferrin (Cal Biochem); Alb, human serum albumin (Behring); Cat, bovine liver catalase (Sigma); Ov, ovalbumin (Worthington); CA, bovine erythrocyte carbonic anhydrase (Worthington); SBTI, soybean trypsin inhibitor (Worthington); My, whale myoglobin (Cal Biochem); L, egg white lysozyme (Sigma); and In, insulin (Lilly).

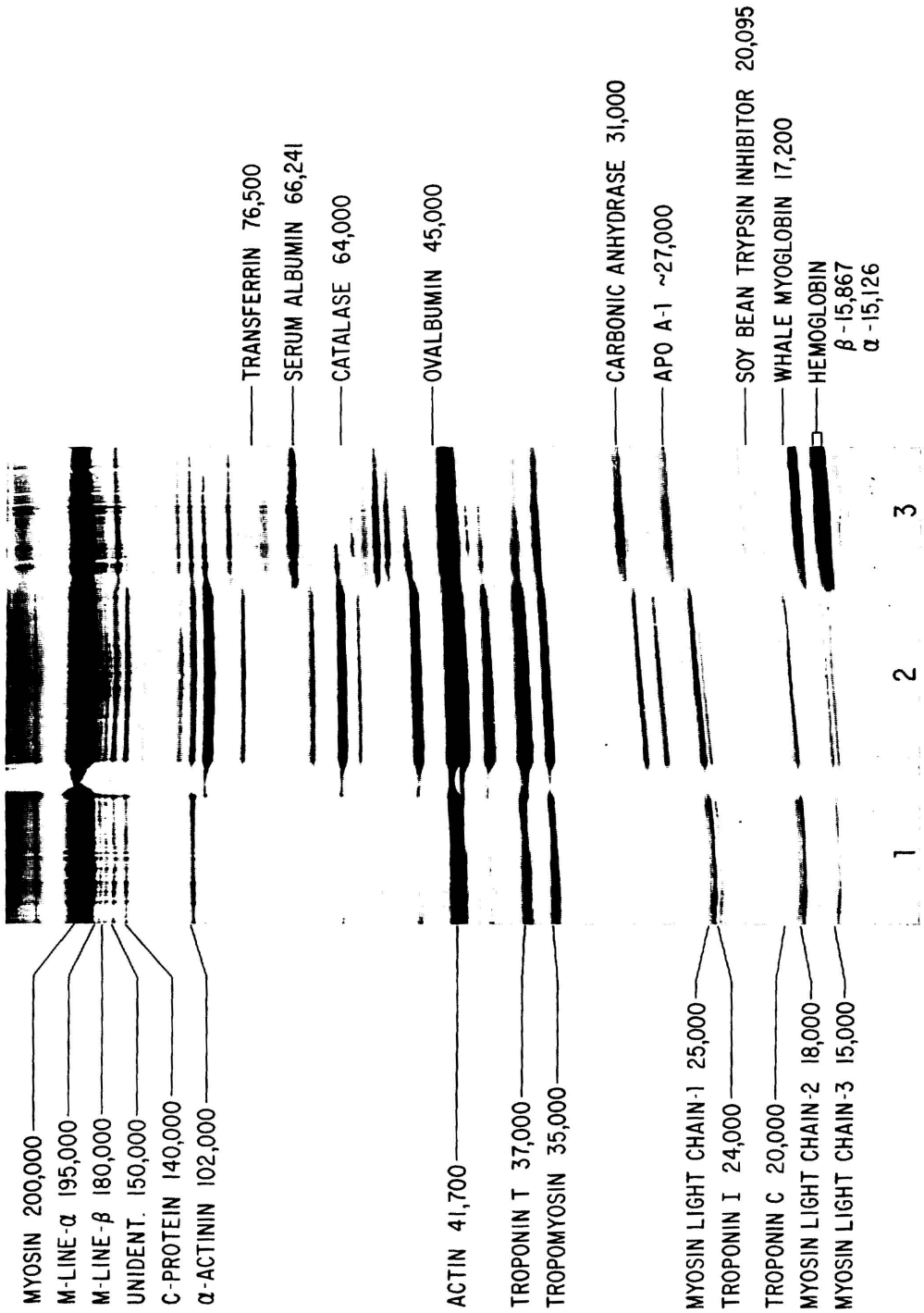


Fig. 18: Comparisons of one-dimensional SDS gel patterns of 1. Rabbit psoas muscle myofibrils. 2. Rabbit psoas muscle homogenate. 3. Rat heart standard preparation. Tentative rabbit muscle myofibrillar protein identifications (on the left) are from Porzio and Pearson[9], and numbers are approximate molecular weights.

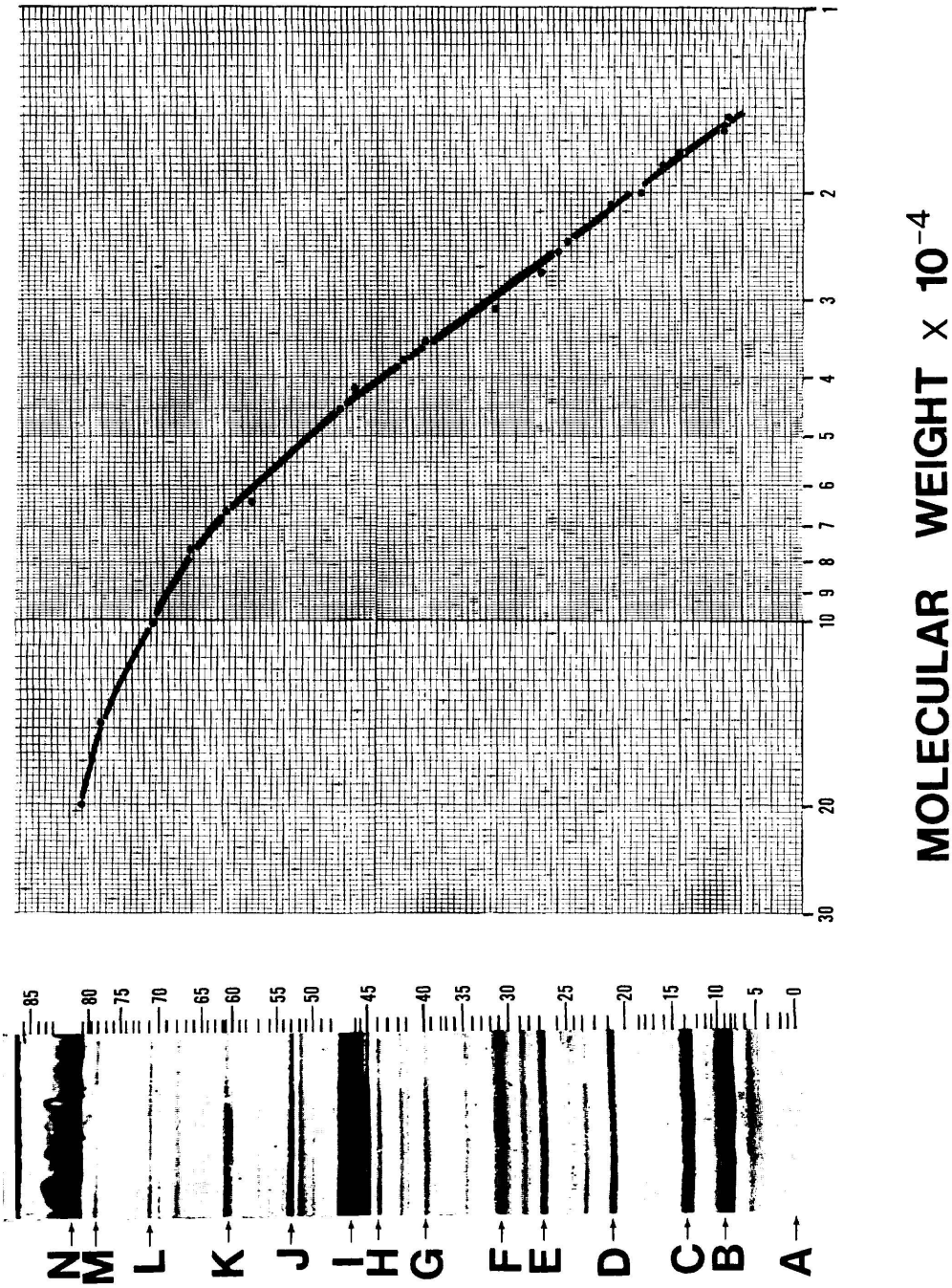


Fig. 19: Calibration curve of molecular weight versus SDS electrophoretic mobility for proteins of the rat heart standard mixture. Experimentally, the hearts were very rapidly excised and homogenized in a relaxing buffer plus pyrophosphate, and the myofibrils isolated by a slight modification of the method of Zak, Etlinger, and Fischman[62]. The homogenates or myofibrils were diluted in melted agarose (0.7% agarose, 1% SDS, 0.5% dithiothreitol, 0.125 M tris, to pH 6.8 with HCl) to contain 5 or 10 mg tissue per ml, or the myofibrils from that amount of tissue. The solutions were heated briefly to 100° to denature the proteins and stop proteolysis, and were then stored frozen at –20° C until used. For use, the agarose was melted using microwave radiation and added as a layer 3–5 mm thick above the slab gels which were prepared, run, stained, and photographed essentially as described[45, 46].

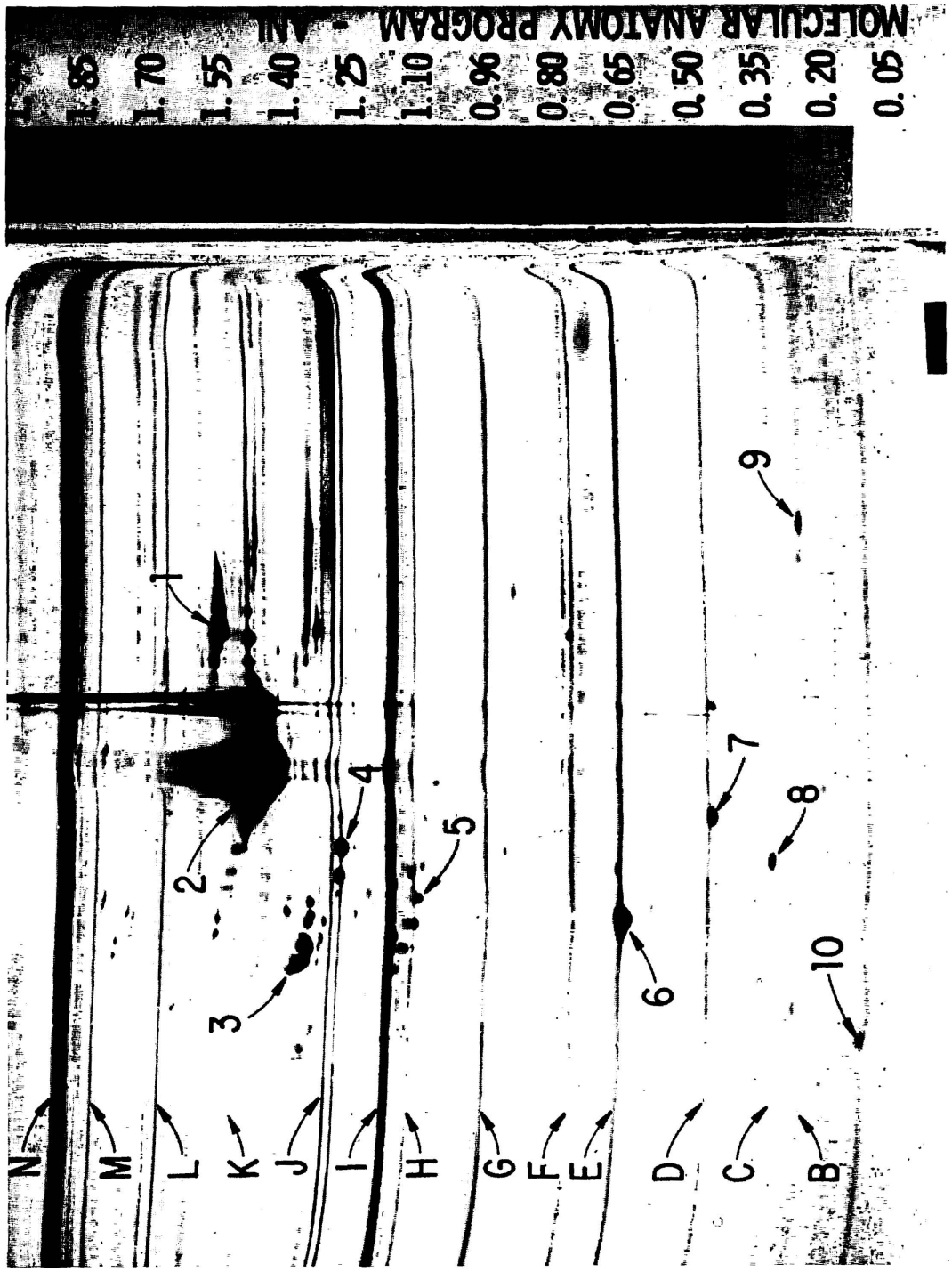


Fig. 20: Two-dimensional patterns of human serum proteins with rat heart streak standards added to agarose used to seal first-dimension gel in place. Lines are lettered to match lines in Figure 19. Proteins are: 1. transferrin, 2. albumin, 3. α_1-antitrypsin, 4. fibrinogen γ-chain, 5. haptoglobin β-chain, 6. Apo A-I lipoprotein, 7. haptoglobin α^2-chain, 8. prealbumin, 9. hemoglobin β-chain, 10. Apo A-II lipoprotein. Identifications are from Anderson and Anderson[7].

Fig. 21: Two-dimensional pattern of human serum proteins with unperfused rat heart standards. B and C are hemoglobulin and myoglobin.

Fig. 22: Two-dimensional gel of whole rat heart homogenate.

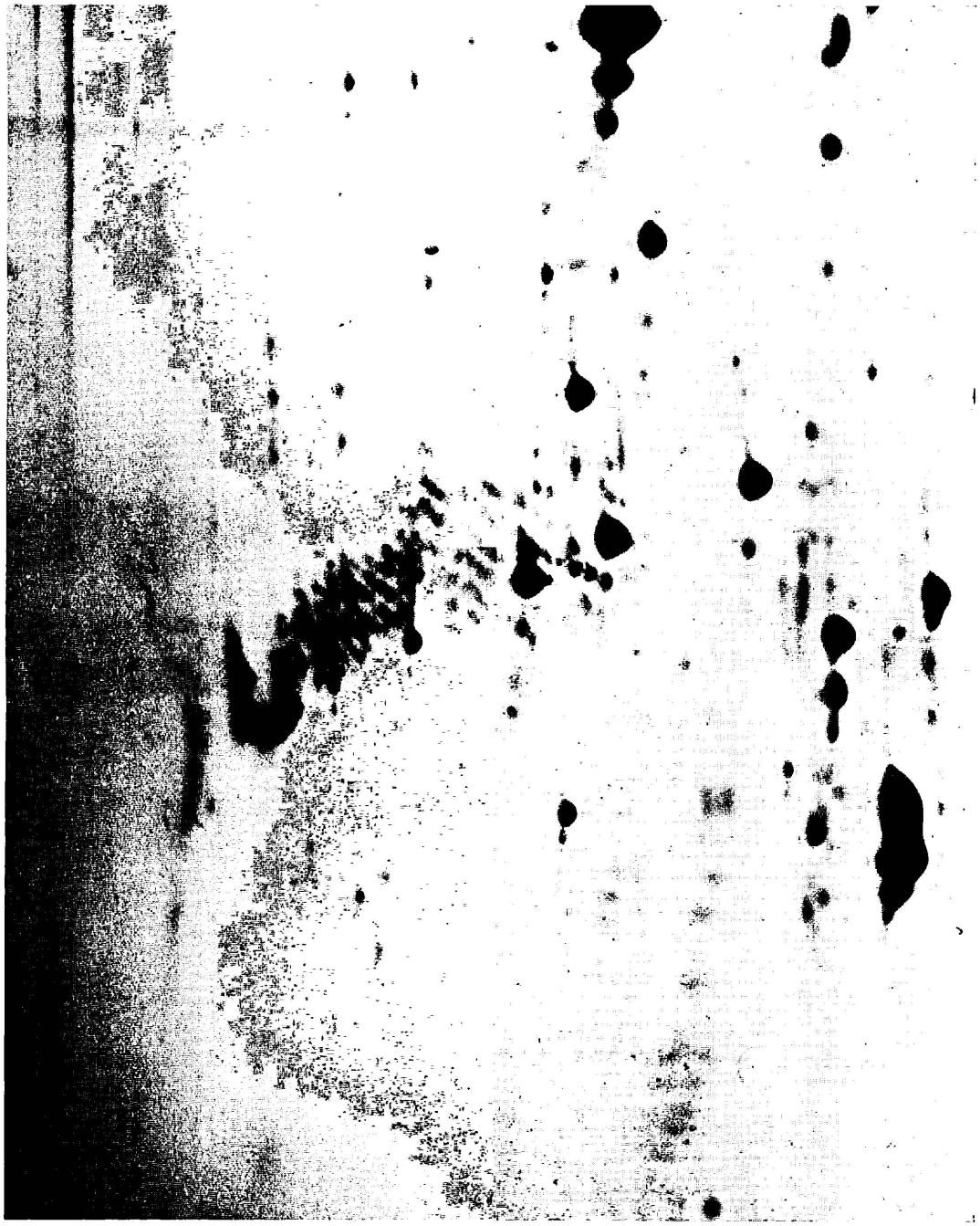


Fig. 23: Enlarged portion of the gel in Figure 22 showing a novel pattern of charge and molecular weight heterogeneity thought to exist *in vivo*. This peculiar protein has properties not previously described to our knowledge; it resembles the pattern produced by invertebrate paramyosin and may, in fact, be a mammalian equivalent of that thick-filament protein.

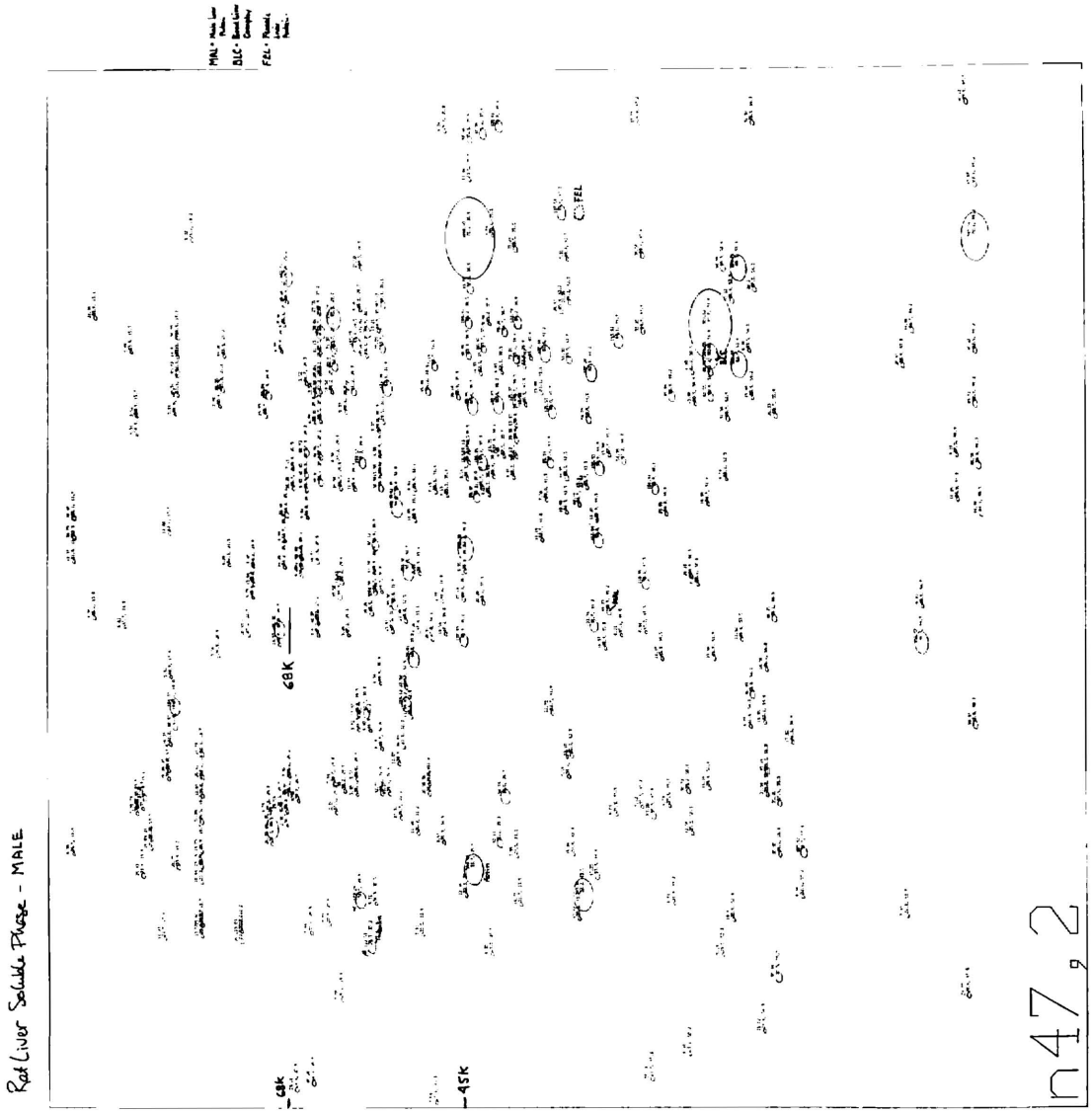


Fig. 24: Computer generated plot of the data extracted from the gel in Figure 15. A computer program produced a list of spot positions and integrated densities from the 2-D scan of Figure 15, and this information was then plotted (either 3 x 3 or 6 x 6 feet) with ellipses representing spots (size of ellipse related to integrated spot density). A comparison of Figure 24 with Figure 15 shows that even this prototype system correctly detects and quantitates most spots. Above the center of each ellipse is written the density and x-y coordinates of the spot.

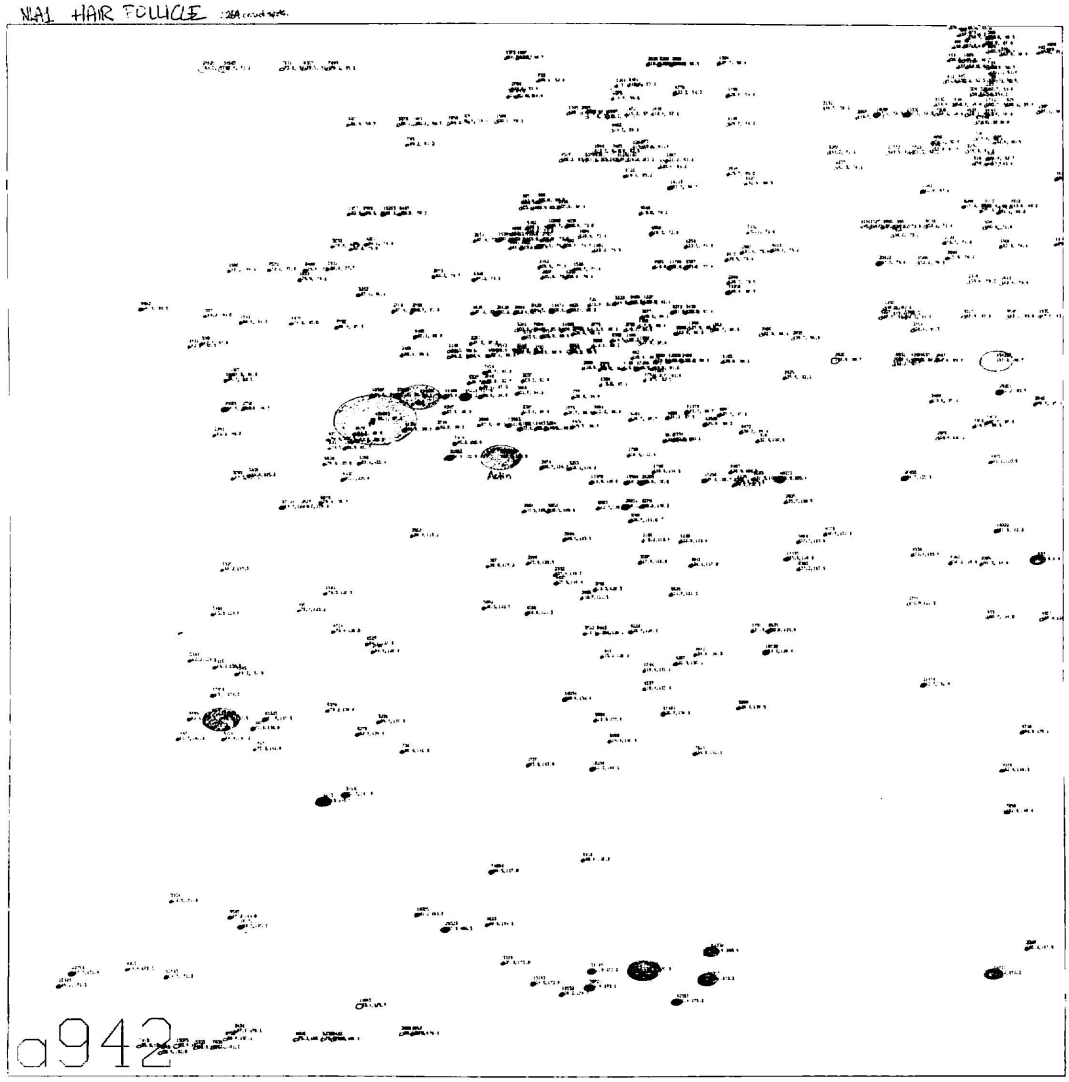


Fig. 25: Computer plot of the data extracted from the gel in Figure 10 (human hair follicle). Solid spots are those correctly detected, while open ellipses are artifacts arising from the mistaken processing of streaks. In the computerized analytical systems being built provision will be made for visual comparison of computer drawn pattern with original photographic image.