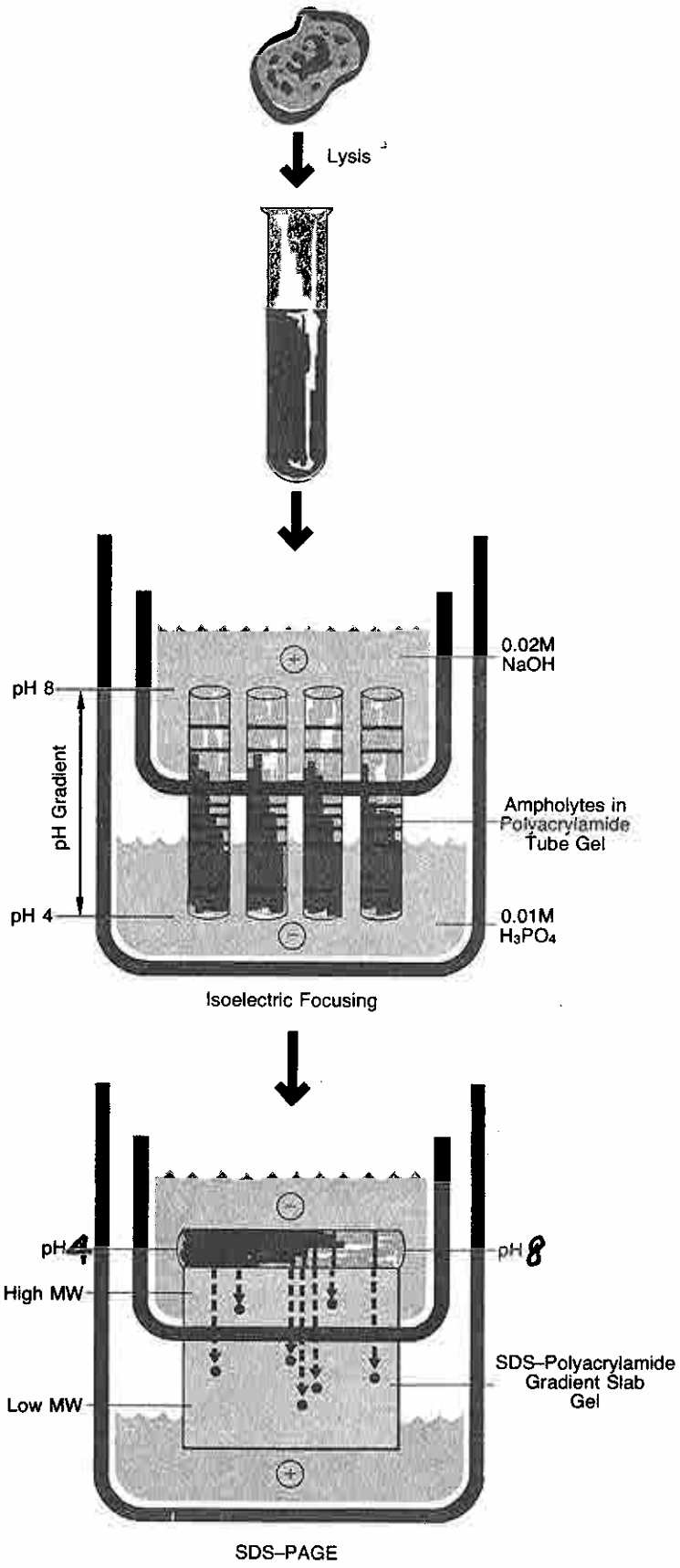



**Figure 1.** O'Farrell's 2-D gel technique
After lysis of specimen in a lysis buffer (8M urea, 2% NP-40, 5% mercaptoethanol), the sample undergoes separation on the basis of charge (isoelectric focusing). The gel is then transferred to the SDS–PAGE step, in which the polypeptides are further separated on the basis of size

# New Dimensions in Protein Analysis

## Two-Dimensional Gel Techniques Are Coming of Age as Image Processing Is Applied to Computer Analysis of the Data

Proteins occupy a central position in the expression of genetic information. They act as the major effectors of cellular activity, since genetic information is predominantly expressed in the structures and activities of the thousands of proteins coded for in DNA and synthesized during the life of a cell.

The individual mix of protein components in a given cell determines, to a very large extent, the state of differentiation and level of functional activity of that cell. These proteins may function as: 1) major structural elements in cells (e.g., microtubules); 2) enzymes controlling the rates of biochemical pathways; 3) receptors interacting with exogenous stimuli; and 4) extracellular products affecting other cells. If one could precisely identify and quantify the protein composition

Reprinted from Analytical Chemistry, 1981, *53*, 390A.

# Report

**Eric P. Lester**
Division of Biological Sciences
and Pritzker School of Medicine
Box 420
950 E. 59th Street
Chicago, Ill. 60637

**Peter F. Lemkin**
**Lewis E. Lipkin**
Image Processing Section
Laboratory of Pathology
National Cancer Institute
Bethesda, Md. 20205

of a particular cell or population of cells, then its state of differentiation and function could be determined.

Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) is an analytic biochemical technique that represents a powerful tool for the evaluation of the polypeptide components of complex mixtures of proteins, such as cells or body fluids. Computer processing of 2-D gel information will ultimately be capable of producing a complete catalog of the polypeptide components of virtually all cells, thus permitting the precise definition of cell types and activities. This article provides a brief summary of the state of the art of 2-D gels—the techniques involved, their applications, and the potential for future developments.

## Basic Technique

The analytic resolution of a single (one-dimensional) technique is vastly increased when a second independent analytic procedure is applied to components separated by the first technique. This principle has long been employed in areas such as thin-layer chromatography, and is confirmed again in 2-D PAGE. A major challenge for the future is the addition of still further dimensions to such techniques, as we shall see.

While a number of historical antecedents can be traced, O'Farrell's paper in the *Journal of Biological Chemistry* in 1975 was crucial to the development of 2-D PAGE. It demonstrated that two relatively high resolution one-dimensional techniques—isoelectric focusing in polyacrylamide gels and sodium dodecylsulfate (SDS) polyacrylamide gel electrophoresis—could be combined into a single two-dimensional technique. The combined method was capable of simultaneously resolving hundreds of individual polypeptide components of cells.

The essence of the technique is quite simple and is outlined in Figure 1. The specimen to be analyzed (cells, serum, etc.) is first treated with a denaturing and reducing buffer system such as 8M urea, 2% nonionic detergent, and 5% mercaptoethanol, thus freeing the individual polypeptide components. This sample is then subjected to isoelectric focusing, usually in a thin polyacrylamide tube gel also containing 8M urea and nonionic detergent in addition to carrier ampholytes (polyaminopolycarboxylic acids that migrate in the gel to form a pH gradient when subjected to an electrical field). Components of the sample are separated into discrete bands in the gels on the basis of their net intrinsic charges in the denaturing medium. Each component migrates in the pH gradient created in the gel by the action of the electrical field on the ampholytes until it reaches a pH at which its net charge is zero.

This focusing gel, containing the separated polypeptides, is then positioned atop an SDS polyacrylamide gradient slab gel and electrophoresed at right angles to the first separation dimension. SDS binds to the polypeptide components electrostatically, masking any charge differences between polypeptides, causing them to migrate through the second gel on the basis of their apparent molecular sizes. Thus two independent molecular parameters—charge and size—provide the basis for the two-dimensional separation. The final result is an array of spots distributed throughout the matrix of the slab gel (Figure 2). Although some proteins appear not to focus well in the first dimension, resulting in artifactual horizontal smears, the technique is capable of a high degree of precision and reproducibility.

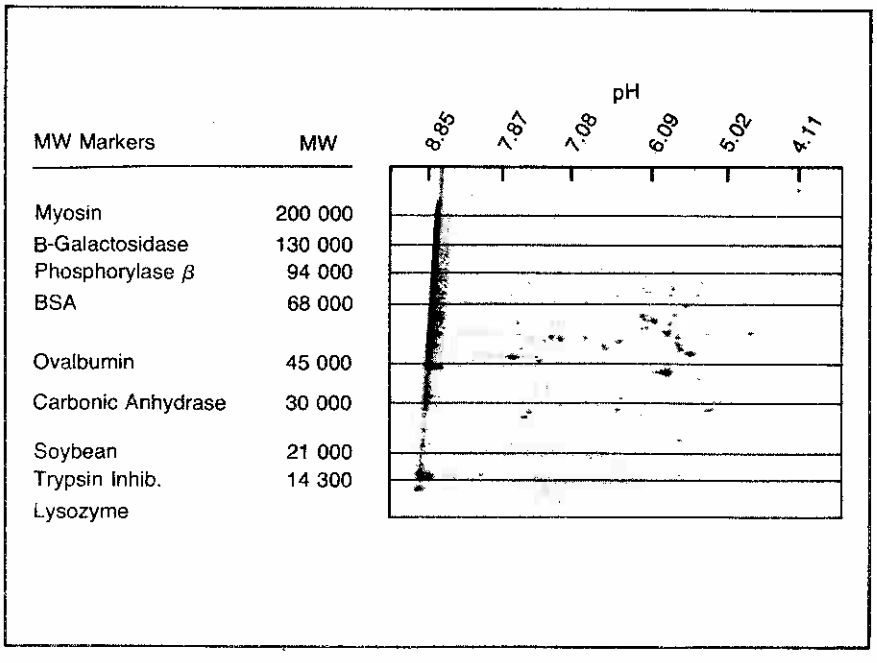



**Figure 2.** 2-D gel of human lymphocyte proteins. Coomassie Blue stained pattern

The development of devices such as the Iso-Dalt apparatus designed by Norman and Leigh Anderson at Argonne National Laboratory, which can focus 20 samples and run 10 slab gels simultaneously on identical gels, has permitted the analysis of large numbers of samples (See ANALYTICAL CHEMISTRY, **1979,** *51,* 707–8 A). Even with this high degree of reproducibility in the distribution of polypeptide spots, it is usually wise to run each sample in replicate gels to confirm polypeptide differences between samples.

### Technical Variations

Two-dimensional PAGE is susceptible to innumerable variations, summarized in Table I. Some samples may be better solubilized by the addition of small amounts of the ionic detergent SDS (e.g., 0.5%) to the lysis buffer. Since the 8M urea in the focusing gel is capable of removing the SDS from the surface of the proteins, this does not interfere with their focusing. Ribosomal proteins have often been analyzed in a 2-D system involving acid/urea electrophoresis in the first dimension and SDS–PAGE in the second dimension, as described by Kaltschmidt and Wittman in 1970. Basic proteins, such as histones, are not resolved on standard O'Farrell gels because of their alkaline isoelectric points and the well-known degeneration of the cathodic region of isoelectric focusing gradients. To avoid this problem, one can reverse the positions of the acidic and basic buffer baths and perform a nonequilibrium electrophoresis in place of isoelectric focusing. This, coupled with the use of solubilizing agents such as salt, protamine, or phosphatidyl choline has permitted good resolution of basic nuclear proteins. Another solution to this problem has been to reverse the order of the two electrophoreses—SDS–PAGE first, followed by isoelectric focusing in a slab gel, since polypeptides that will not enter a first-dimension isoelectric focusing gel will nonetheless usually enter an SDS–PAGE gel. Even with these variations, some systems, such as plant proteins, have not been successfully analyzed on 2-D gels. A variety of marker protein mixtures can be electrophoresed as standards to determine both the molecular weight and relative charge of sample polypeptides. Carbamylation of a purified protein provides a mixture of charge variants differing by unit charges for use in standardizing the isoelectric (charge) dimension. Often the separation of polypeptides in a series of gels of similar samples (e.g., of mammalian cells) is so reproducible that there is no need to standardize each gel.

High resolution 2-D gels can be made even more informative by complementing them with a variety of other techniques, either preprocessing the samples or postprocessing the gels. Staining with protein-binding dyes, such as Coomassie Blue, represents the simplest form of the latter (Figure 2). The recent development of a silver stain capable of visualizing far smaller amounts of protein (well under 1 $\mu$g in an individual spot) has permitted analysis of trace protein constituents both in cells and in body fluids, such as unconcentrated cerebrospinal fluid.

Such staining procedures essentially reflect the steady-state concentration of the polypeptides visualized. Additional information about the synthesis rates of cellular polypeptides can be obtained simultaneously by incubating the viable cell population with a radioactive amino acid substrate for a brief period (e.g., 20 min) prior to disruption with lysis buffer. The radioactive polypeptides synthesized can be separated on a 2-D gel and detected and quantified autoradiographically, providing a fingerprint that is specific for the metabolic status of the cells under study. Experimental perturbations of cellular metabolism involving alterations in the synthesis rates of specific polypeptides may thus be studied with a previously impossible level of precision. The use of high specific activity radioisotopes coupled with autoradiography (or scintillant-enhanced autoradiography, termed fluorography, for $^{3}H$ labels) permits evaluation of up to approximately 1000 polypeptides simultaneously. The use of very large 2-D gels may ultimately provide resolution sufficient to detect the majority of cellular polypeptides.

Double labeling experiments with $^{3}H$- and either $^{14}C$- or $^{35}S$-labeled amino acids, which can be differentially visualized by performing fluorography followed by autoradiography on the same gel sequentially, can be designed to permit evaluation of individual polypeptide turnover rates. In addition, this type of double label visualization permits coelectrophoresis of control and experimental samples labeled with isotopes of differing energies (e.g., $^{3}H$ and $^{35}S$). Differential visualization of the two samples in the same gel removes the problem of artifactual variations in protein migration in the gels. Any difference between the $^{3}H$ and $^{35}S$ images must reflect a real difference between the two samples.

A wide variety of other isotopically labeled compounds may also be used to study protein metabolism, particularly post-translational protein modifications. Phosphoproteins are readily visualized autoradiographically in samples labeled with $^{32}P$. The $^{32}P$-labeled spots can then be further corre-

**Table I. Ancillary Techniques for 2-D Gels**

| |
|---|
| **Preprocessing cellular sample** |
| Biosynthetic pulse labeling with: |
| • radioactive amino acids |
| • radioactive phosphate |
| • other radioactive substrates (e.g., polyamines) |
| Plasma membrane labeling with $^{125}I$ or $NaB(^{3}H)_4$ |
| Cell fractionation |
| Immunoprecipitation with specific antisera |
| **Variations in 2-D technique** |
| Reverse SDS–PAGE and isoelectric focusing |
| Nonequilibrium pH gradient electrophoresis in place of isoelectric focusing |
| Acid/urea electrophoresis in place of isoelectric focusing (for ribosomes) |
| Use of other solubilizing agents—e.g., SDS, salt, protamine, phosphatidyl choline |
| **Postprocessing of gels** |
| Protein staining—Coomassie Blue or Silver staining |
| Transfer protein pattern to nitrocellulose or diazobenzyloxymethyl paper followed by: |
| • immunofixation |
| • enzymatic assay |
| • binding assay |
| Autoradiography (or fluorography), including: |
| • double-label fluoro/autoradiography |

lated with the patterns of amino acid-labeled proteins, permitting identification of charge variants of specific polypeptides created by phosphorylation. As another example, we have recently demonstrated the covalent incorporation of labeled polyamine into a few specific low molecular weight polypeptides in human lymphocytes. Surface membrane proteins can be labeled enzymatically with $^{125}I$ and displayed on 2-D gels (Figure 3).

When 2-D gels are used in conjunction with other techniques, a potentially infinite variety of $n$-dimensional analyses are possible. A typical example is the 2-D gel analysis of pulse-labeled polypeptides from specific subcellular fractions isolated by density gradient ultracentrifugation. The analysis of immunoprecipitates on 2-D gels provides a particularly powerful combination (Figure 4). With the advent of monoclonal antibody technology, it is now only a matter of time before we possess "libraries" with specific antibodies for most, if not all, proteins. It will then be a relatively simple matter to determine both the cellular location (by immunomicroscopy) and the metabolic status (by 2-D gels) of a large number of polypeptides.

duced by mitogens, substances that induce mitosis. Since we are examining only the 300–400 polypeptides present in highest concentration in lymphocytes, we are probably, for the most part, examining changes in the synthesis of major structural, or "housekeeping," polypeptides. Nonetheless, we can identify both an increase in the absolute synthesis rate of most polypeptides as well as alterations in the relative synthesis rates of many specific polypeptides. Similar results have been reported in association with altered growth rates in clonal rat cell lines. Alterations in the synthesis rates of specific proteins have been seen in hepatoma lines treated with glucocorticoids, a mouse lymphoma line treated with cyclic AMP, fibroblasts in the presence of interferon, and a variety of other systems.

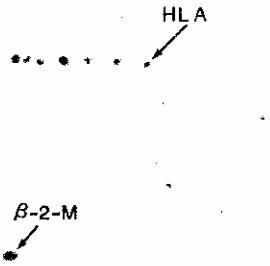



**Figure 4.** Autoradiograph of 2-D gel pattern

Immunoprecipitate of antihuman beta-2-microglobulin and a detergent lysate of radiolabeled human lymphocytes. The beta-2-microglobulin spot is at lower left, below a line of coprecipitated human leucocyte antigen (HLA) polypeptides

**Figure 3.** 2-D gel pattern

$^{125}I$-labeled surface proteins of a human T lymphoblastoid cell line, MOLT-4

## Current Applications

Most published studies utilizing 2-D gels have sought to identify differences in the autoradiographic patterns of pulse-labeled proteins from cells in tissue culture either produced by treatment with a hormone or similar agent or occurring during differentiation. Our studies are typical in that we have used 2-D gels to identify human lymphocyte polypeptides whose relative synthesis rates are altered during the course of growth activation induced by mitogens, substances that induce mitosis. Since we are examining only the 300–400 polypeptides present in highest concentration in lymphocytes, we are probably, for the most part, examining changes in the synthesis of major structural, or "housekeeping," polypeptides. Nonetheless, we can identify both an increase in the absolute synthesis rate of most polypeptides as well as alterations in the relative synthesis rates of many specific polypeptides. Similar results have been reported in association with altered growth rates in clonal rat cell lines. Alterations in the synthesis rates of specific proteins have been seen in hepatoma lines treated with glucocorticoids, a mouse lymphoma line treated with cyclic AMP, fibroblasts in the presence of interferon, and a variety of other systems.

Alterations in protein synthetic patterns during differentiation have been defined with 2-D gels during myogenesis in rabbits, in nerve cell lines treated with nerve growth factor, and in a mouse myeloid leukemia that differentiates in response to DMSO. Since protein synthetic patterns seem to reflect both the tissue of origin and the state of differentiation of the cell population studied, we and others have begun to examine the potential clinical applications of the technique. As an oncologist, I (E.P.L.) would find it particularly useful to be able to determine the precise cell type and state of differentiation of the malignancies I treat, and 2-D gels may make this possible in a single analysis of the tumor tissue. The very large data base implicit in a single 2-D gel pattern of tumor cell proteins may further contain a wealth of prognostic information, clues to expected biological behavior, and indications of likely responses to therapy, once we learn to interpret the data.

Yet another area in which 2-D gels are finding important applications is in genetics and the analysis of genetic diseases. Two-dimensional gels can readily detect shifts in the location of a single polypeptide spot in mutant *E. coli* strains known to differ from a control strain in a single genetic locus (protein). Studies are currently underway in our laboratory to identify mutant or deleted gene products in human genetic diseases such as cystic fibrosis.

At a more basic level, 2-D gel technology has shed light on one of the central concerns of genetics—the extent of genetic polymorphism in protein gene products. Whereas classical charge-dependent electrophoretic studies of cellular enzymes have shown a high frequency of polymorphism, 2-D gels, which visualize primarily structural proteins when applied in the usual fashion to whole cell extracts, show a very low frequency of polymorphism. This implies a high degree of evolutionary conservation of structural, as opposed to enzymatic, gene products.

## Computerized 2-D Gel Image Analysis

Simple inspection of a 2-D gel image may often be adequate to provide an important experimental result, such as detection of an alteration in the synthesis rate of a specific polypeptide of interest. Such a result, of course, largely ignores the often immense amount of data about other polypeptides visualized concurrently. Furthermore, simple inspection leads to essentially qualitative rather than quantitative conclusions.

We, and others, have been primarily concerned with the computerized analysis of 2-D gels of biosynthetically pulse-labeled polypeptides, since these provide information about the relative synthetic rates of a large number of polypeptides, potentially very useful data for a wide variety of biological problems. Here we shall confine our discussion to this problem, although analysis of other types of 2-D gel data (e.g., stained patterns) is quite analogous. We have called our system GELLAB—a software laboratory for gel analysis.

To realize the true potential of 2-D gels we must have an analysis of these images that is at once comprehensive, covering all polypeptides visualized,

**Table II. Computer Analysis of 2-D Gels of Radiolabeled Polypeptides**

| Problems | Solutions |
|---|---|
| **Data acquisition** | Scintillation counting of excised spots |
| | "Scintillation scanning" of intact gels (theoretical) |
| | Autoradiography/fluorography with image digitization by television or optical density scanning |
| **Spot detection and quantitation** | Image filtering |
| | Background subtraction |
| | Spot detection |
| | • Thresholding |
| | • Peak detection |
| | • Central core detection and propagation |
| | • Curve fitting |
| | Integration of spot density |
| **Gel comparison and spot mapping** | "Landmarking" relative to reference gel (R-gel) |
| | Coordinate transformation |
| | Spot pairing by empirical criteria |
| | Statistical measures of surety of pairing |
| **Data base construction** | "R-spot" sets |
| | Statistical analyses |

and quantitative. For this, a computerized analysis of these images is clearly required. The quantities of data are too vast for any other approach unless the number of gels or polypeptides analyzed is severely limited.

The difficulties involved in computer analysis of 2-D gel data can be subdivided into a number of problems (see Table II). These are: 1) raw data acquisition; 2) initial polypeptide spot detection and measurement, which we term "segmentation"; 3) comparisons of the same spots in replicate or similar gels, or spot mapping between gels; and 4) construction and analysis of a unified data base encompassing all spots in all gels in an experimental series. From this last level, we may further conceive of the construction of a theoretical "canonical" gel in which the location of each species of polypeptide from any and all sources is specified. Such a location, a function of isoelectric point and size, would be unique for each polypeptide in nature.

## Data (Image) Acquisition

While portions of gels containing individual polypeptides may be excised from gels and subjected to scintillation counting to accurately quantitate radioactivity (proportional to that polypeptide's synthesis rate), such a procedure is feasible only for a small proportion of the total data on a 2-D gel. It may, nonetheless, provide an important means for checking other methods of quantitation. In general, data acquisition must be accomplished by first producing an autoradiographic (or fluorographic) image on X-ray film, followed by digitization of this image. Digitization requires either a high-quality television system (capable of a minimum of 500–1000 noninterlaced line resolution) or a rotating drum optical density scanner (such as an Optronics P-1000 high-speed drum densitometer). We have utilized a Vidicon television camera at a resolution of 170–250 microns per picture point (pixel) to convert 2-D gel autoradiographs into 512 × 512 matrices of gray values. Of course, with a television system, an optical density scale must be included in each scanned image to permit conversion of gray scale values to optical density values. Furthermore, the autoradiographic image must itself be calibrated by exposure of the film to gel segments of known radioactivity in order to permit estimation of polypeptide radioactivity from optical density values. This second calibration applies to all images regardless of the method of digitization.

During each of these steps, care must be taken to remain within the dynamic range of both the film and the digitization hardware. For this purpose, a graded series of film exposures of a single gel may be necessary. While some workers have used "preflashed" film in an attempt to linearize the lower range of the film response curve, such an approach is not necessary if radioactive standards are exposed with each film. It may ultimately be possible to obviate these difficulties by the development of instrumentation capable of scanning for and quantitating radioactive emissions, or photons if fluorography is used, directly from intact 2-D gels. This would allow direct quantitation of radioactivity at each picture point.

## Spot Detection and Quantitation–Segmentation

The first crucial task in analyzing the digitized image (regardless of its method of acquisition) is the detection of its component spots. A variety of approaches have been employed by various groups. Most begin with one or more filter algorithms that are designed to reduce image noise. A typical 3 × 3 pixel convolution filter for image smoothing is shown below:

$$\begin{matrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{matrix}$$

In convolution filtering, the value of a "central" pixel is multiplied by, in this case, four. Its eight neighbor pixels are each multiplied × 1 or × 2, according to their positions relative to the central pixel. The nine products are summed and the total value is divided by 16, since (1 + 2 + 1 + 2 + 4 + 2 + 1 + 2 + 1) = 16. The original value of the central pixel is then replaced by this normalized value. This process is repeated for each pixel in the image.

Background values are then usually obtained by estimating the first peak of a density histogram of each of 64 subregions of the entire image, and subtracting an interpolated value from each spot measurement. Actual spot

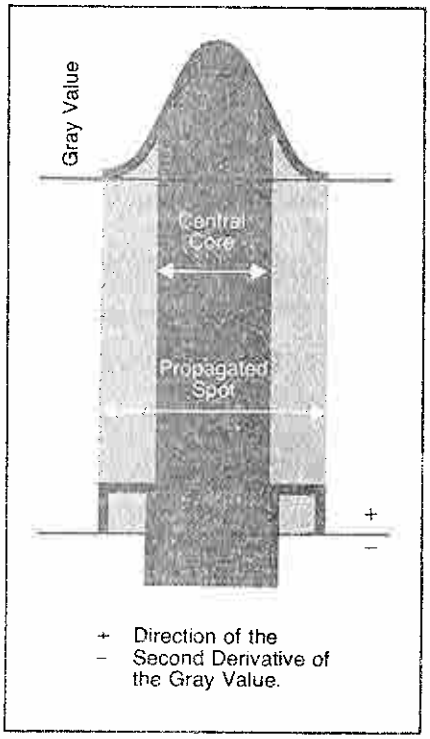



**Figure 5.** Central core and its propagation for a theoretical spot

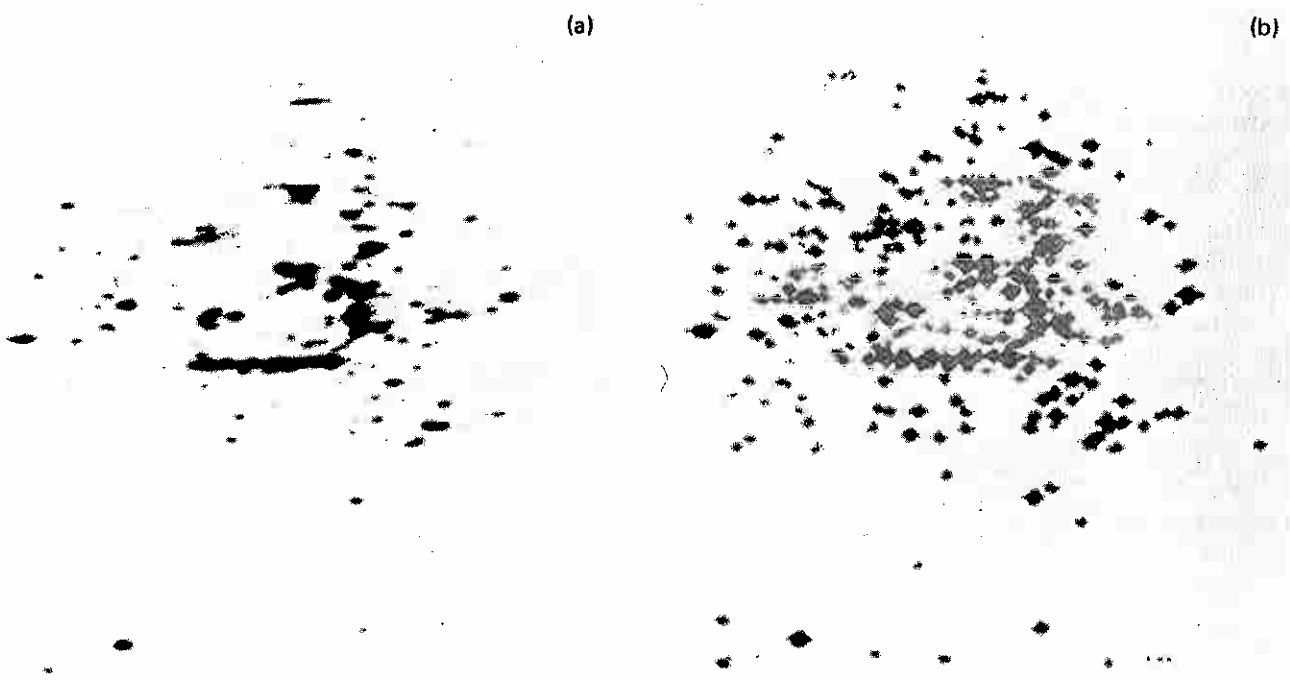



**Figure 6.** Computer processing a 2-D gel image. (a) Native image; (b) Computer-generated image

detection may be performed by thresholding, i.e., setting an empirical value to determine a spot boundary—but this is often unsatisfactory because of the marked variation in spot intensity and background. An alternative approach involves detection of density peaks followed by evaluation of adjacent inflection points (points of maximum rate of density change) to determine spot boundaries. Our approach has utilized detection of the central core of each spot, which is defined as those pixels having negative second derivatives of the density. This is because the rate of increase of density as the spot is scanned from its edge to its center is less in the central core than at the preceding inflection point (Figure 5). Once the central core has been found, adjacent pixels may be added to it (propagation of the central core) on the basis of ancillary algorithms (e.g., monotonically decreasing values moving out from the central core are included until background values or an adjacent spot is reached). An example of an original and a "segmented" 2-D gel is shown in Figure 6. The superimposition of two such images is represented schematically on the cover of this issue.

Yet another approach involves the fitting of Gaussian or skewed Gaussian curves to the detected peaks. Theoretically, it should be possible to "deconvolve" the entire image into its component overlapping density curves, thus permitting precise measurement of each spot in spite of overlapping. Such a possibility, however, would potentially require large increases in computation, and even current methods are relatively costly since each image has about 250 000 pixels that must be evaluated by successive algorithms. No single set of algorithms has proven ideal for all areas of all images, and direct comparisons of the various programs now available are needed. It may ultimately be best to combine these various approaches, using one set of algorithms for certain areas of a gel and applying a second set to other areas not well segmented by the first method.

## Gel Comparison

Having achieved, by whatever means, an adequate detection and measurement estimate of the spots in a gel, it is next necessary to compare that gel with its control gel to determine experimentally induced variations. Segmentation algorithms output a list of spots, with associated densities, X and Y coordinates, and perhaps other data. Such lists must be matched with lists for similar gels, mapping each spot from one gel onto its counterpart in another gel.

Our approach here is typical. We landmark each gel by comparing it with a reference gel from the same series and noting the X, Y coordinates of a set of 15–24 easily identifiable spots matched between both gels. The same set of spots is used for comparisons between the reference gel and all gels that are paired to it. The X, Y coordinates of these landmark spots are used by a set of gel comparison algorithms to map all of the spots in one landmark region in one gel onto the spots in the same region of the reference gel. While some groups have used these coordinates to transform the coordinates of all the spots in the gel, we have simply used empirical criteria based on spot distance from the landmark spot and distance between members of a pair to accomplish pairing (Figure 7). Our final output consists of a series of gel comparison files, each of which consists of a list of spots paired with the spots present on the reference gel.

We have found it particularly useful to develop software that allows us to rapidly compare (flicker) two gel images on a video display and interactively landmark them. The use of video display or plotter output to visually evaluate the gel analysis process at each step is essential.

An additional feature needed in gel comparisons is an estimate of the reliability of each spot pair detected. This estimate is a function of the distance of the pair from the relevant landmark spot (dL, Figure 7) as well as the distance between the spots in the pair (dP, Figure 7), and is essentially a statistical measurement. It must be determined experimentally by comparing replicate gels of the same sample and is, of course, profoundly affected by the quality and reproducibility of spot detection.

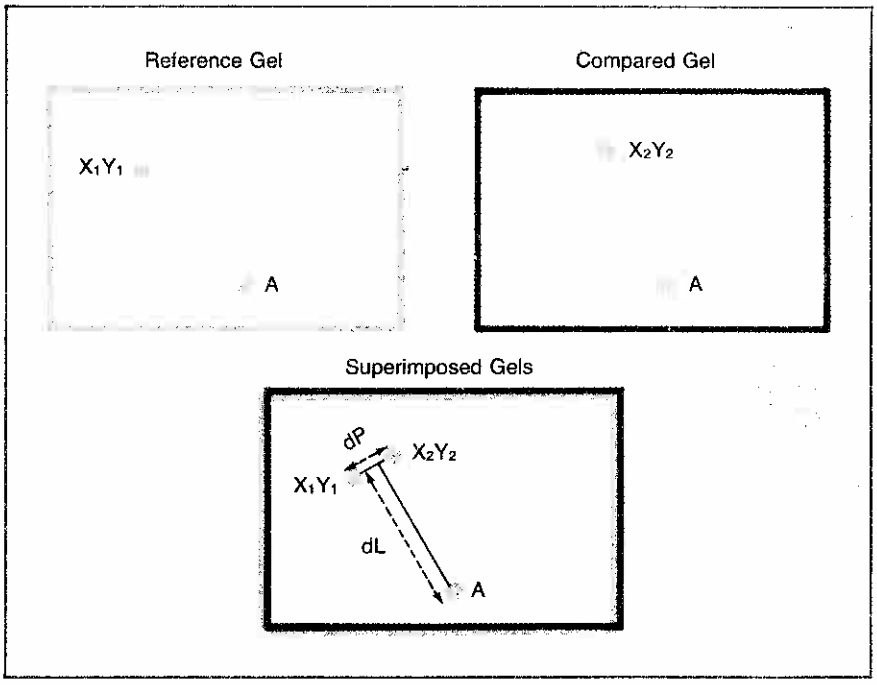



**Figure 7.** Spot pairing parameters. A = landmark spot; dL, dP—see text

### Data Base Construction and the Canonical Gel

Once a series of gel comparison files has been constructed, the final stage of gel analysis can begin with the construction of a unified data base encompassing all of the valid data on all of the gels in an experimental series. The essential element in such a series is what we call an R-spot set. This consists of a given spot on the reference gel (the gel to which all of the other gels were compared) and the spot paired with it in each of the other gels in the series. Ideally, if this polypeptide was present in each sample, correctly electrophoresed on each gel, visualized on each autoradiograph, correctly segmented by the spot detection algorithms, and correctly paired to the R-spot by the gel comparison algorithms, then there will be one member of the R-spot set for each gel in the series (Figure 8). Such an ideal is currently achievable in only a fraction of the R-spot sets in our current data bases. However, the current ability to include over 100 gels, with thousands of spots from each gel in a single data base, allows statistical determination of the level of error inherent in the system. Such statistical analysis may define spots that change significantly in response to the initial experimental manipulation of the cellular samples. To evaluate such a result, it is very useful to have a 2-D gel image displaying these spots with markings (Figure 9).

Many computerized 2-D gel analysis systems currently perform in the optical density mode, rather than in a mode representing actual radioactivity in each polypeptide. In the density mode, the value of each spot must be normalized and may be expressed either as a percent of the total density of all the spots or as a ratio to the density of one or more selected polypeptides. Our current system flexibly allows reexamination of the entire data base after normalizing it in a variety of ways.

A large data base such as this permits detailed and sophisticated statistical dissection to evaluate unsuspected interrelations between polypeptides and alterations in rates of polypeptide synthesis induced by a variety of experimental manipulations. We have been particularly careful to build a software system that will output data into files that may be independently analyzed by the powerful statistical software packages currently available in most computer centers. Even in its current developmental state, our system, GELLAB, appears significantly more effective than visual analysis in detecting experimentally induced alterations in polypeptide synthesis rates. Since each of the four steps in 2-D gel analysis is accomplished separately (data acquisition, image segmentation, image comparison, and data base construction) we can substitute and compare various hardware and software components at each step until the system is optimized.

By appropriately combining such data bases, it will someday be possible to create a true canonical gel, by which we mean an idealized gel image in which the location of each polypeptide found in any and all experimental series is specified within statistical limits. Such a theoretical image may then be used recursively to guide both the spot detection and gel comparison algorithms. Ultimately, it may provide an intellectual framework and nomenclature by which the state of differentiation and activity of any cell population may be defined.

Thus the coalescence of a variety of techniques from such disparate disciplines as tissue culture, biochemical analysis, image processing, and statistical computer analysis of large data bases has produced a powerful new system for the evaluation of questions that arise in fields such as human ge-

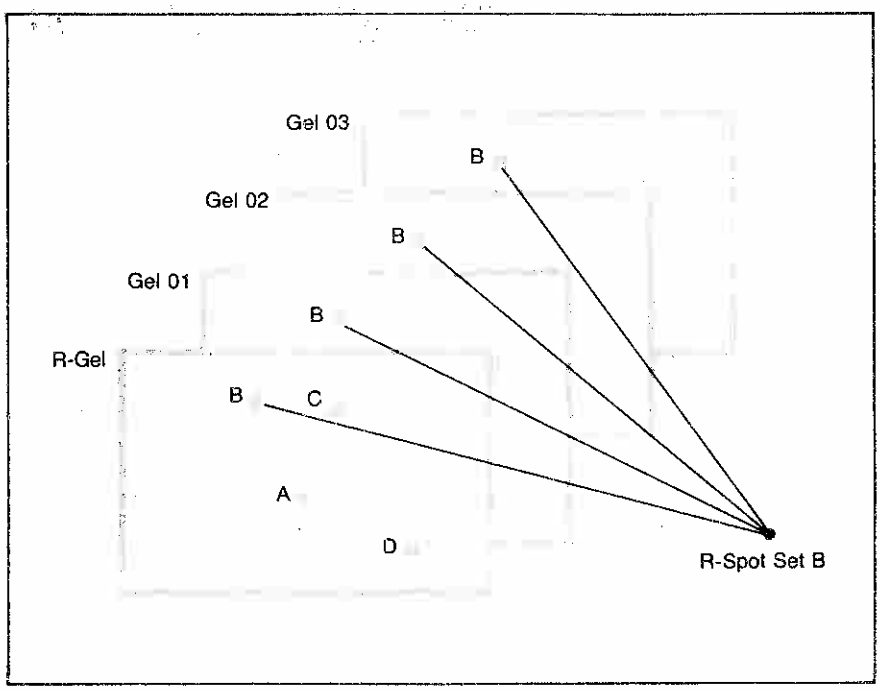



**Figure 8.** The definition of an R-spot set

**Figure 9.** A 2-D gel image with labeling of spots found by statistical analysis of a data base

netics and developmental biology, and in the study of disease.

## Suggested Reading


(1) P. H. O'Farrell, High resolution two-dimensional electrophoresis of proteins, *J. Biol. Chem.*, **250**, 4007–21 (1975).

(2) E. P. Lester, P. Lemkin, L. Lipkin, and H. L. Cooper, Computer-assisted analysis of two-dimensional electrophoreses of human lymphoid cells, *Clin. Chem.*, **26**, 1392–1402 (1980).

(3) L. Lipkin and P. Lemkin, Data-base techniques for multiple two-dimensional polyacrylamide gel electrophoresis analyses, *Clin. Chem.*, **26**, 1403–12 (1980).

(4) N. L. Anderson and N. G. Anderson, High resolution two-dimensional electrophoresis of human plasma proteins, *Proc. Natl. Acad. Sci.* (USA), **74**, 5421–5 (1977).

(5) J. I. Garrels, Two-dimensional gel electrophoresis and computer analysis of proteins synthesized by clonal cell lines, *J. Biol. Chem.*, **254**, 7961–77 (1979).

(6) C. R. Merril, R. C. Switzer, and M. L. Van Keuren, Trace polypeptides in cellular extracts and human body fluids detected by two-dimensional electrophoresis and a highly sensitive silver stain, *Proc. Natl. Acad. Sci.* (USA), **76**, 4335–9 (1979).

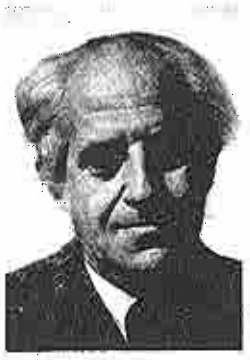
(7) E. H. McConkey, B. J. Taylor, and D. Phan, Human heterozygosity: A new estimate, *Proc. Natl. Acad. Sci.* (USA), **76**, 6500–4 (1979).

(8) J. Bossinger, M. J. Miller, K-P. Vo, E. P. Geiduschek, and N-H. Xuong, Quantitative analysis of two-dimensional electropheretograms, *J. Biol. Chem.*, *254*, 7986–98 (1979).

(9) P. Lemkin and L. Lipkin, GELLAB: A Computer System for 2-D Gel Electrophoresis Analysis. I: Segmentation and System Preliminaries. II: Spot Pairing. III: Multiple Gel Analysis, *Comput. Biomed. Res.*, in press, 1981.


*E. P. Lester*

*P. F. Lemkin*



*L. E. Lipkin*

*Eric P. Lester received a BS in biochemistry and an MD (1972) at the University of Chicago. He is presently an assistant professor in the Hematology/Oncology Section of the University of Chicago's Pritzker School of Medicine. His research involves the study of biochemical aspects of growth control in normal and malignant lymphocytes.*

*Peter F. Lemkin received a BS and an MS in electrical engineering from Cornell and the University of Rochester, respectively. In 1978 he received a PhD in computer science from the University of Maryland. His research projects at the National Institutes of Health (NIH), where he has been employed since 1968, have included the computer analysis of optical microscope blood cell images, the application of artificial intelligence techniques to biological image processing, and the development of a multiple gel data base for the analysis of 2-D gel images.*

*Lewis E. Lipkin is chief of the Image Processing Section in the Laboratory of Pathology of the National Cancer Institute. Following an undergraduate major in mathematics and graduate work in physical chemistry at NYU, he received the MD degree from what is now Downstate Medical Center, Brooklyn, N.Y., in 1949. His research interests have included quantitative microscopy and computer-aided image processing. He has most recently been active in 2-D gel electrophoresis, algorithms for prediction of nucleic acid secondary structure, and quantitation of cell movement and phagocytosis.*