Review

Twenty years of two-dimensional electrophoresis: Past, present and future

Norman G. Anderson[1, 2]
N. Leigh Anderson[1]

[1]**Large Scale Biology Corporation, Rockville, MD, USA**
[2]**Department of Pathology, George Washington University, Washington, DC, USA**

Contents

1 Introduction

It is not possible to do justice to the history of so interesting and complex a subject as two-dimensional electrophoresis in one lecture, or one short paper, or to credit fully the work of all of those who have contributed to it, many of whom are present. Hence, this is a review of how a fascinating area of research appears to two investigators. We meet on the 100th anniversary of the death of Louis Pasteur, a chemist interested initially in the basic question of what is unique about the chemistry of living systems, and then in how basic new concepts could be applied to the understanding and treatment of human diseases, *i.e.*, in both pure and applied research. We would therefore like to also mention how our subject today relates to the same two areas that interested Pasteur – namely, unique aspects of life itself, and the detection and treatment of disease.

Correspondence: Dr. N. G. Anderson, Large Scale Biology Corporation, Rockville, MD 20850-3338, USA (Tel: +301-424-5989; Fax: +301-762-4892)

Nonstandard abbreviation: HMG, 3-hydroxy-3-methylglutaryl

Keywords: Two-dimensional polyacrylamide gel electrophoresis / Epigenetics / Genome

1.1 History

One interesting aspect of the history of high resolution two-dimensional electrophoresis (2-DE) is shown in Fig. 1, which indicates the number of publications in this field by year. Since science is responsive to paradigms, or in the vernacular, to fashions, the question immediately arises: Is 2-DE going out of fashion? Are the paradigms which surrounded it being modified? For comparison, it is instructive to plot the rate of acquisition of genomic data with time. We can estimate by simple (and long) extrapolation that, if all of the data concerned man, and if none of it were overlapping, the complete sequence would be available approximately in the year 2018 (Fig. 2). The complete sequences of all expressed genes, which comprise an estimated 3% of the human genome, are expected to be completed very much earlier.

High throughput sequencing will continue after one complete genomic sequence is known, however. The reason for wanting to analyze many genomes completely is to find not just a few differences associated with a particular disease, but all of them. For example, we would like to know all of the differences between host and tumor cells, what differences are found in different metastases, the key differences between different ethnic groups, and of course, the sequence of the numerous genes which are believed to predispose us to different genetic diseases. Only complete genomic data will allow the detailed analysis of the problems of why genes are organized into chromosomes, and what the effects of changing gene order on chromosomes in man might be. Further, only careful analysis of large noncoding, non-

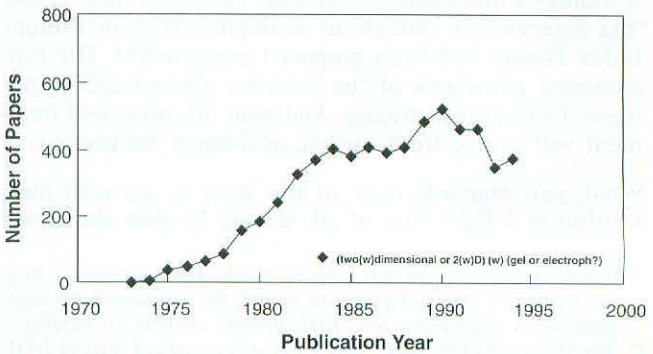

Figure 1. Results of a literature search conducted in Medline for published papers related to 2-DE by year, using as a query (two(W) dimensional or 2(w)D) (w) (gel or electroph?). The results probably underestimate the total number of 2-DE papers, but should be relatively accurate for comparing year to year.

0173-0835/96/0303-0443 $10.00+.25/0

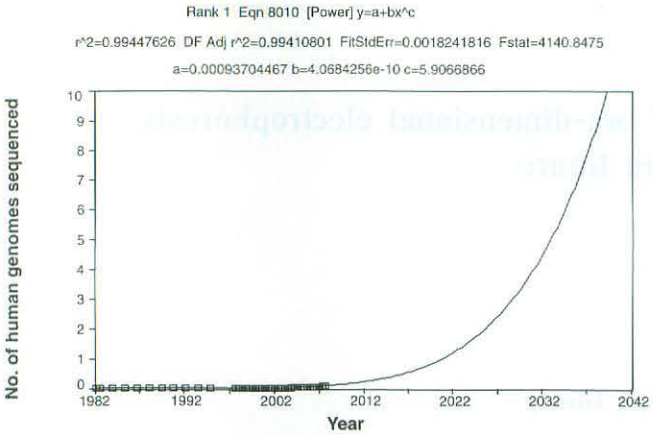

Figure 2. A prediction of the rate of genomic DNA sequencing as a function of time, based on a fit to the growth of sequence database contents to date and expressed as human complete genome equivalents.

promoter regions will allow the question of whether so called "junk" DNA has undiscovered functions, or can vary in a random manner. There have been optimistic predictions of a future in which massive, automated, possibly chip-based sequencing, combined with complete automation of all other aspects of the process, will come into use; in which clinical diagnostic procedures for variants of a large numbers of genes will be standard; and in which means will be found for almost casually mending genes. It is also proposed that it will be possible to modify or even remove individual genes, thus providing a cure for AIDS and other diseases in which viral genes are incorporated into human genomes. This is the projected view of the future for genomics.

Important byproducts of the current emphasis on genomics are enthusiasm, financial support, and the expectation of technical progress. The latter is important. In many fields of technology, especially those related to military objectives, the operational paradigm is that past progress can be extrapolated into the future. Hence systems are proposed in the expectation that, for example, strengths of materials, thrust per pound, or microprocessor speed will continue to increase in a quasi-predictable way. Such expectations and the financial support they generate provide driving forces for research and technology development. The Human Genome Project* is biology's first venture into what physicists have called "Big Science" [1], though an analogous Human Protein Index Project had been proposed previously**. The fundamental paradigms of the Genome Project are simply these: Genes cause disease. And both diagnosis and treatment will evolve from nucleic acid-based technology.

What, you may ask, does all this have to do with high resolution 2-DE? First of all, it may be that the down turn in the first graph showing 2-DE publications by year merely reflects the fact that many of the results we enthusiastically thought would result from 2-D studies are now coming out of genomics. For instance, we do not have to worry about how many genes (or proteins) there are in human cells, because the sequencers will shortly find out. Likewise, the question of where each gene is expressed is being determined by RNA or cDNA analysis of different cell types. The DNA-based efforts will continue to accelerate, and are even now generating vast on-line DNA and protein data bases to be followed by genes, vectors, expressed proteins, reagents, and kits, many of which, we believe, will ultimately be available by overnight delivery. The Genome Projects will thus make genes, reagents, and data available to individual investigators who are the ones capable of really exploring them in detail. We may expect every gene to yield a book, though whether such a book is ever printed on paper, as opposed to existing virtually as a collection of hyper-linked World Wide Web documents, is questionable.

The paradigm underlying most of the genomics boom is that knowing the parts will allow one to understand and ultimately fix the system. What is beyond argument is that current genome projects will provide the basic molecular anatomy of living cells. Given this fast track through the present and future, we are left with these questions: What comes after the Genome Project? Where, if anywhere, do high resolution 2-DE acrylamide gels fit in? Will there be anything left worth doing in biology? Often, in the history of science, just when everything seems to fall in place, some radical paradigm shift occurs. Could that be true here, and could some of the underlying concepts be in need of revision, or be actually wrong? Is there some large and fundamental area left to explore? And what technologies might be required to explore it?

1.2 2-DE in perspective

Before delving more deeply into these matters, let us briefly return to our assigned subject: Twenty years of two-dimensional electrophoresis-past, present and future. First, let us place 2-DE in perspective. Analytical chemistry, as Arne Tiselius emphasized, progressed from single procedures for single elements, groups, compounds, or activities to methods based on physical separations linked to general detection methods. Hence, since the 1940s, great emphasis has been placed on electrophoresis, chromatography, and mass spectrometry, and more recently on capillary electrophoresis. Of these, only mass spectrometry has provided the resolution required to resolve really complex mixtures starting with elements, but until recently it could not adequately resolve macromolecules. To improve resolution, the concept evolved of combining, in a two-dimensional array, two methods which depended on different parameters. This was initially done using two different solvents sequentially in paper chromatography [3], or by combining electrophoresis in one dimension with chromatography in the second. The concept was also applied to centrifugal separations using sedimentation rate and isopycnic banding density as the two separation parameters [4].

* It is of interest that the first proposal for the Human Genome Project, written in 1983 and published in 1985 [2] suggested both large-scale DNA sequencing and 2-DE protein analysis in parallel.

** The Human Protein Index Task Force was organized around 2-DE technology by Senator Allen Cranston and members of the US Congress in August, 1980 with N. G. Anderson as Chairman. The election of that fall removed from office nearly every supporting elected member, and nearly all of the heads of agencies involved. The project was not funded.

More recently two-dimensional analysis has evolved as a general field with its own body of theory, as reviewed by Wankat [5]. The point is simply that if the two separation parameters are unrelated, then the final resolution should be the product of the resolution of the two methods separately – providing there was no resolution loss in mating the two techniques together. Obviously, more dimensions can be added, but multidimensionality poses problems in interfacing additional separation techniques without resolution loss, problems in data analysis, or problems in data display.

At least five attempts to develop 2-DE methods were described [6] before four papers appeared in 1975–76 which combined isoelectric focusing in urea with SDS-electrophoresis to make a separation based on the charge of a denatured protein in the first dimension (which is a direct function of amino acid composition), and molecular mass in the second [7–10]. Since several reviews and books have recorded the major events which followed the appearance of high resolution 2-DE [11–18], we will not review the development of this field in detail here. Enormous progress has been made in the development of databases for cells in culture [19, 20], plasma [17, 21, 22], rat and mouse liver [23, 24], heart tissue [25, 26], *E. coli* [27], and yeast [28], among others. Among the major technical advances has been the development of methods for identifying proteins from 2-DE gels by microsequencing and mass spectrometry [29–31], and of immobilized pH gradient electrophoresis as a practical tool to extend the pH range and improve reproducibility [32–34]. In addition, SWISS-PROT now provides a readily accessible data base of protein data, and SWISS-2-D-PAGE provides an annotated data base of patterns, both on the World Wide Web [35].

2-DE held the promise of providing means for writing a molecular anatomy of human and animal tissues, cells, and body fluids, and of bacterial cells and viruses, and for detecting changes occurring during development, aging, disease, and in response to experimental and environmental variables, for discovering both new disease markers, and targets for drug development. It also offered an opportunity to determine, when combined with cell fractionation, the intracellular location of proteins. To fulfill these promises, it was important that the method provide reproducible patterns, and that quantitative measurements of abundance could be made. Much progress has been made toward achieving these goals, and of providing data bases linked to interpretive computer programs. However, it must be admitted that progress has not been as fast as initially anticipated, and that some of the advances initially envisioned [15], including a map-based Human Protein Index, have been only partially completed. Complete indexed and annotated maps of all human tissues or those of experimental animals are not as yet available, and only a fraction of the human diseases listed in the nomenclatures of pathology have been explored using 2-DE. Unfortunately, resources for technical development have been limited, and the analytical process has been slow, labor intensive, and time consuming. No completely automated machines have appeared, such as exist for DNA synthesis and sequencing or in clinical chemistry.

As we have mentioned, much of the basic data for a molecular anatomy will derive from the Human Genome Project. How then will these two fields, nucleic acid analysis as represented by the Human Genome Project and analysis of what has been called the proteome (as presently embodied in 2-DE), evolve and interact?

2 Epigenesis

As emphasized by Strohman [36], the currently operational paradigm concerning the relationship of genes to human disease, and therefore some of the relevance of the Human Genome Project, requires revision*. There are, in fact, two informational systems in biology, one genetic, the other epigenetic. While the former is directly attacked by the Genome Projects, the latter reflects both environmental effects and complex interactions between genes and between gene products. Figure 3 illustrates

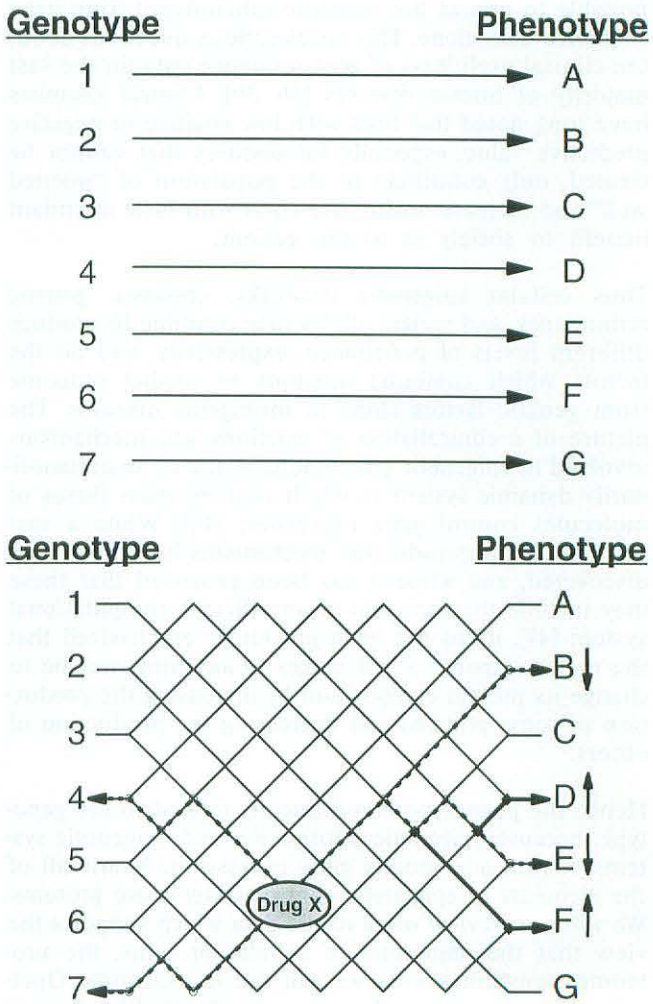

Figure 3. Schematic representations of systems in which there is a direct relationship between genotype and phenotype (a) and a preferred model in which genotype and phenotype are related by a complex network of interactions at the protein level (b). In the latter case, drug binding to a protein can lead to complex effects through the network, producing up or down regulation of specific proteins.

* Paradigms dictate the scope of normal research [37], even though most scientists may be unaware of them.

diagrammatically two alternative pictures of these relationships, which are first a one-to-one relationship between a gene and a disease, and second, a matrix of interactions in which a given genetic defect may or may not produce disease, or different levels of severity thereof in different individuals. Involved in this matrix are not only interactions between different gene products and different genes, but with the environment.

It is estimated that 2% of human diseases may be caused by direct, one-to-one Mendelian expression of a single trait [36]. The remaining 98% of human diseases are polygenic and/or epigenetic in origin. And the message of genetics is that a gene cannot act by itself [38]. It acts only in conjunction with other genes and with the environment. And the presence of one gene, or even of several, does not guarantee that a given trait or disease will be expressed. Genes exhibit variable penetrance, or expressivity, and may inhibit each other (epistasis). Hence for most polygenic diseases it will not soon be possible to predict the outcome (phenotype) from gene sequence data alone. This raises serious questions about the clinical usefulness of gene sequence data for the vast majority of human diseases [36, 39]. Clinical chemists have long noted that tests with low positive or negative predictive value, especially for diseases that cannot be treated, only contribute to the population of "worried well" and increase health care costs with little attendant benefit to society or to the patient.

Thus cellular epigenetic networks, epistasis, genetic redundancy, and variant alleles may combine to produce different levels of penetrance, expressivity, and all the factors which confound attempts to predict outcome from genetic factors alone in multigenic diseases. The picture of a constellation of reactions and mechanisms involved in epigenetic phenomena is one of an extraordinarily dynamic system in which concentration fluxes of molecules control gene expression [40]. While a vast array of signal transduction mechanisms have now been discovered, and while it has been proposed that these may include the elements of an efficient computational system [41], it has not been previously emphasized that the major response a cell makes to anything may be to change its protein composition by increasing the production of some proteins and decreasing the production of others.

Hence the phenotype is not linearly related to the genotype, because epigenetic information and epigenetic systems control and modify gene expression. Nearly all of the elements of epigenetic control systems are proteins. We will now review some of the data which supports the view that the sum total of cellular proteins, the proteome, constitutes what we will call the Genome Operating System, shown diagrammatically in Fig. 4.

3 The Genome Operating System

The molecular phenotype of living cells is extraordinarily flexible. There does not appear to be a fixed set point for the rate of synthesis or breakdown, and therefore steady-state concentration, of any cellular protein. Instead a "kan ban" or "just in time" system exists to make what is needed as needed. The implications of this concept, in which thousands of genes and thousands of proteins are interactively involved, are profound indeed. As recently pointed out by Bray [41], a large number of cellular proteins have as their primary function the transfer and processing of information. These have generally been described as almost static, fixed elements in an information processing system. However, the large literature on regulation (reviewed in the series *Regulation*, edited by Weber) provides numerous instances in which quantitative changes are produced by exposure to drugs, hormones, and toxic agents. Both up- and down-regulation occur. The conclusion is that cells are reactive systems in which information flows not only from genes to proteins, but in the reverse direction as well. And if a cell includes mechanisms for sensing both the amount and/or function of each of large numbers of individual proteins, and the binding to them of drugs or signal substances, then we may postulate that all cellular proteins are, in some sense, sensors, receptors, or information transfer units, and constitute what may be termed in the aggregate a cytosensorium. If the afferent arm of this system is called the cytosensorium, then the efferent arm may be called the cytoeffectorium. Together they make up the genome operating system.

We propose that the genome operating system translates almost any alteration in the quantity or configuration of a cellular protein into a signal which causes up- or down-regulation of either individual proteins or batteries of proteins, in many cases including the protein initially affected. If this concept or paradigm is correct, elimina-

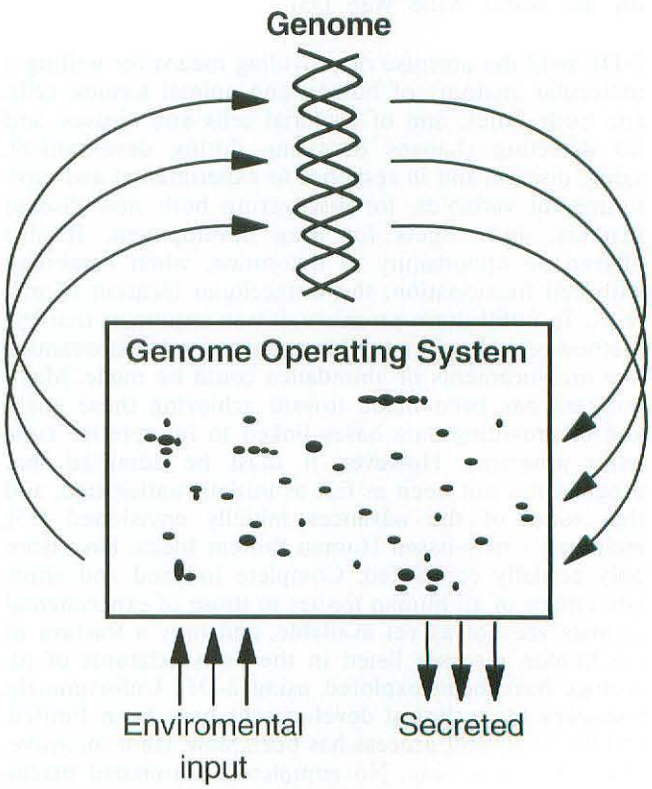

Figure 4. A simplified schematic view of the Genome Operating System.

tion of any single cellular protein, for example in a knockout mouse, would change the remaining molecular phenotype in some way. Further, any drug, except those which produce their effects through a fast transient mechanism, for example an antidote to a poison, should affect gene expression and protein abundance because, to be a drug, it must bind somewhere. These concepts are based on the following findings and postulates.

4 Pharmacology of molecular phenotype effects

In drug development the objective is most often to increase or decrease a protein-mediated activity, hence one would like some general method for discovering or finding drugs which would either up-regulate or down-regulate a specific activity. With only a very few exceptions, all drugs which have been carefully examined by 2-DE produce quantitative pattern changes in target organs such as liver. (The exceptions may be due to incomplete coverage of the entire IEF pH range, or choice of the wrong organ to analyze). Examples are now too numerous to review in detail here. Several studies conducted in our lab will suffice for illustration. Merck's Mevacor® (lovastatin) is a half-billion-dollar-a-year drug which lowers plasma cholesterol levels. It was developed to inhibit 3-hydroxy-3-methylglutaryl (HMG) CoA reductase, which it indeed does. When one examines the effect of this drug alone or in combination with another cholesterol lowering agent, it is found that the physical amount of HMG CoA synthase (coregulated with HMG CoA reductase) increases significantly in amount (Fig. 5), while other proteins are found to decrease (Fig. 6) [23]. However, it now appears that the major factor in cholesterol lowering produced this drug is the co-upregulation of low density lipoprotein receptor [42] resulting in the removal of low density lipoprotein (LDL) cholesterol from plasma, and not simple inhibition of the cholesterol synthesis pathway. This suggests a new concept in drug discovery and development which is to up- (or down-)regulate one desired protein by constructing a drug to target a second protein positively or negatively coregulated with it. Exploiting this idea depends on knowing the details of the genome operating system and of coregulation.

Many other instances of regulational surprises exist. Oltipraz, a compound being studied as a cancer chemopreventive, upregulates a number of proteins including aflatoxin B_1 aldehyde reductase, which metabolizes one of the most potent natural carcinogens [43]. One could not wish for a more relevant result. In studies of enzyme induction by the chlorinated hydrocarbon mixture Araclor 1254 in liver, the endoplasmic reticulum is usually isolated for analysis because the doctrine is that the interesting effects occur in this subcellular fraction. However, if one looks at the 2-DE protein pattern of whole liver from animals treated with this compound, most of the changes found (and not previously suspected) occur in the soluble and mitochondrial fractions [44]. Similarly,

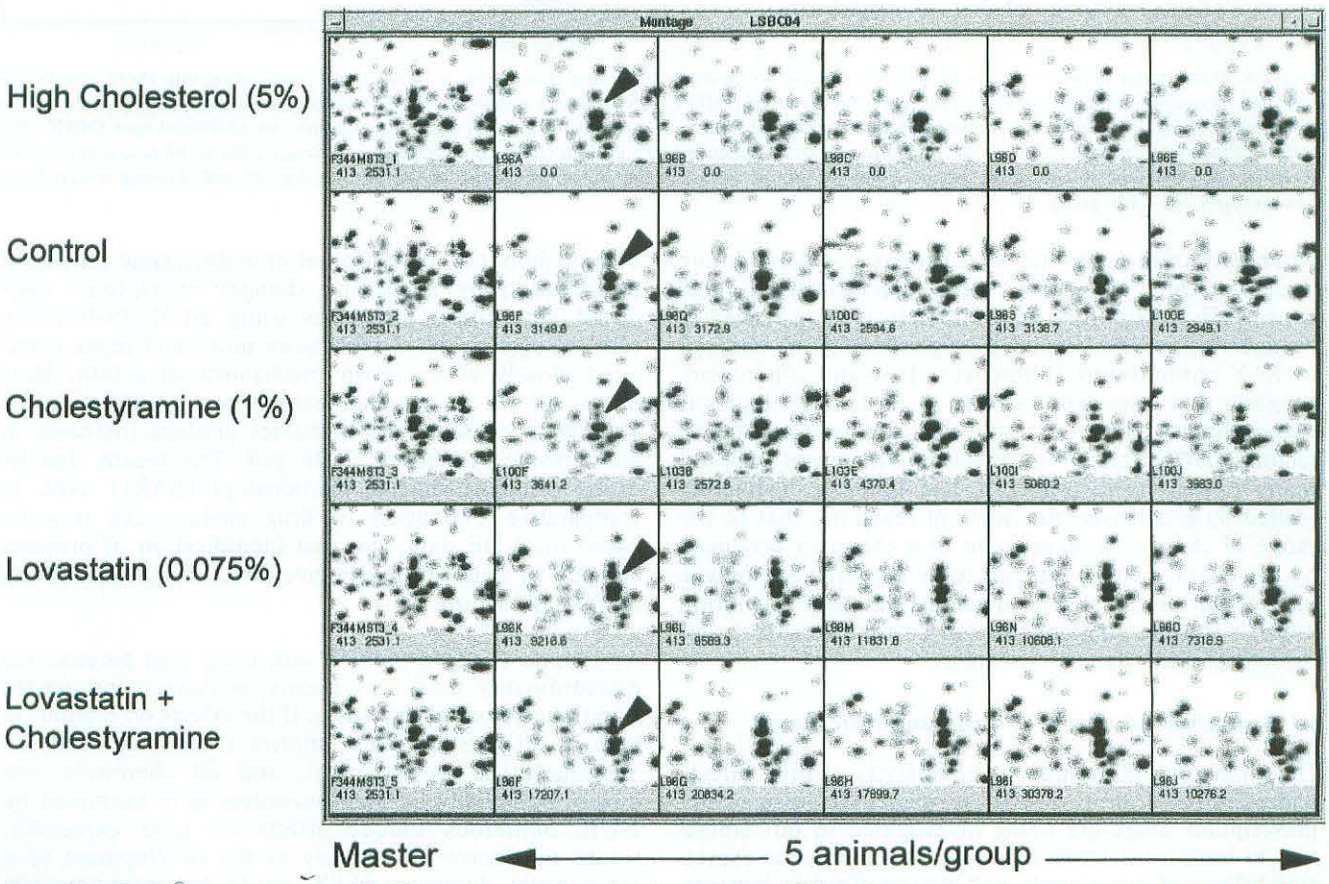

Figure 5. A Kepler® system montage of 2-DE patterns of rat liver, showing the effects of four cholesterol-altering treatments versus controls (one row of images per group). All treatment compounds were administered to five male F344 rats in diet at the concentrations shown for 7 days. An arrow indicates the principal spot comprising cytosolic HMG-CoA synthase.

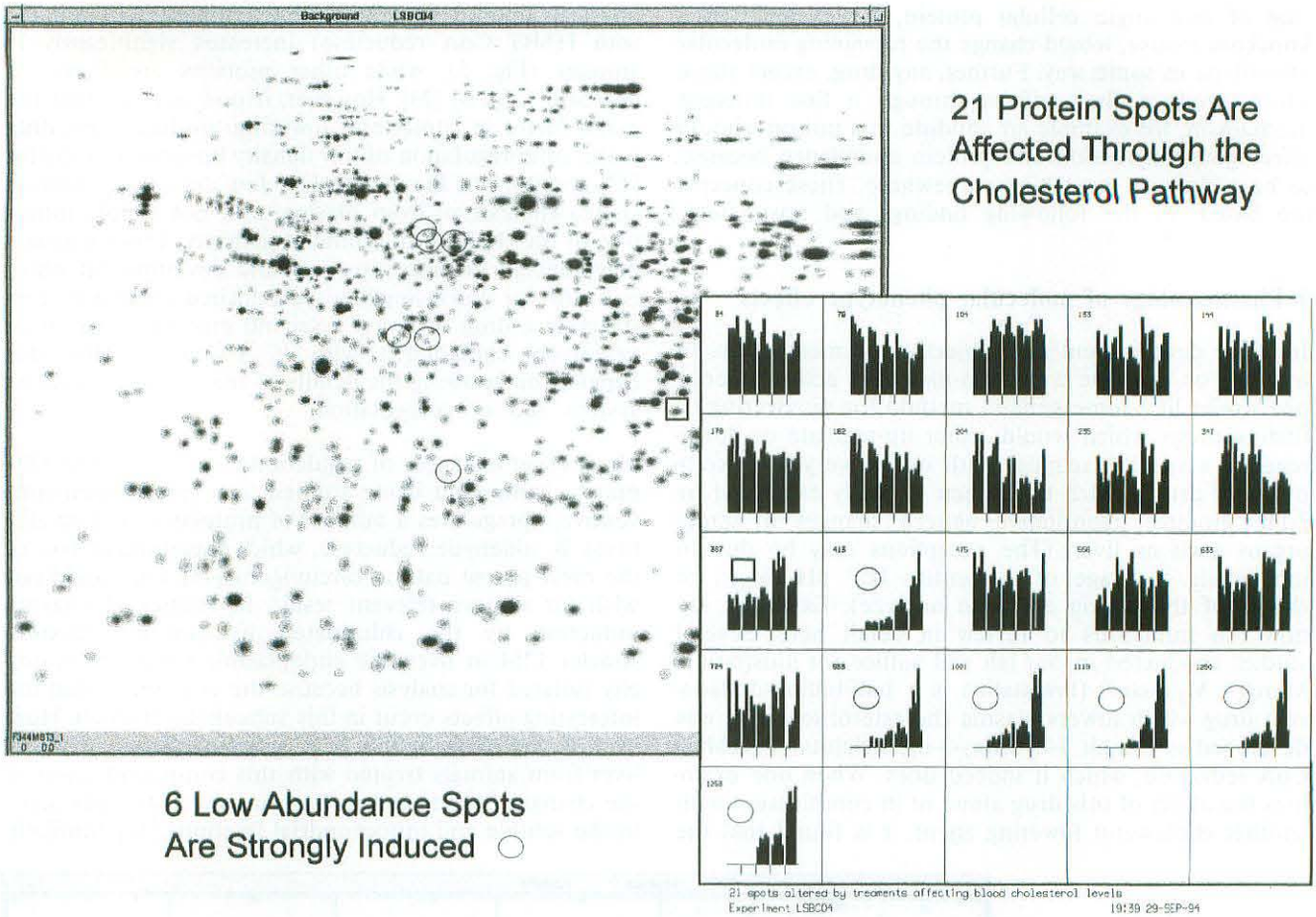

Figure 6. Quantitative data on a series of proteins affected by cholesterol-altering treatments. A series of six spots falling into three groups surrounded by ovals (a top group of three that represent forms of HMG CoA synthase, a middle group of two presumed charge isomers, and a lone lower spot) show consistent quantitative alterations across all treatment groups indicative of tight coregulation. An additional spot (MSN: 367, square) shows anti-synergistic effects of lovastatin and cholestyramine: this protein (as yet unidentified) is strongly induced by peroxisome proliferators. Each bargraph square shows results from five groups shown in Fig. 5 beginning with cholesterol on the left, and showing results from five animals for each group.

when liver of male and female mice are compared, more than one third of proteins differ significantly in abundance [45]. Nearly all of these differences are believed to be due to hormone-mediated epigenetic factors and not to X–Y chromosome differences. This and other work suggests that gene expression in at least some tissues is extraordinarily plastic, is responsive to a large number of factors, and almost always involves not one or two, but many protein gene products. The limits of variability (called by geneticists "the norm of reaction", that is, the range of changes in expression of a character produced by epigenetic factors) remains to be explored for individual cellular proteins. It is apparent that high resolution 2-DE is the pre-eminent method currently available for exploring such dynamic changes.

4.1 A database of effects of drugs and toxic agents

The expanding literature on the analysis of drug effects using 2-DE will be reviewed elsewhere. At present, fifty prescription drugs are being investigated in our laboratory to build a reference data base of rodent gene expression effects of compounds well-characterized in humans. This work constitutes, in part, an extension to the molecular level of classical methods for histological detection of mitochondrial, peroxisomal or endoplasmic reticulum proliferation by measuring changes in proteins associated with those organelles using 2-DE. Preliminary results suggest that 2-DE shows protein changes correlated closely with known mechanism of action. Thus drugs shown to produce peroxisome proliferation by electron microscopy (EM) studies produce increases in peroxosome proteins in 2-DE gels. This means that so-called structure-activity relationships (SARs) used in comparative evaluation of drug analogs can now be based on 2-DE data, and that identification of proteins affected by a new drug can give direct insight into mechanisms of action.

As Claude Bernard pointed out, drugs and poisons are extraordinarily useful as means to dissect and understand physiological processes. If the effects on animal tissues of all new chemical entities (NCEs) currently in pharmaceutical development, and all chemicals produced industrially in large quantities were examined by 2-DE, numerous unique effects on gene expression would be discovered, leading to the development of a mechanistic database which would be extraordinarily useful in both pharmacology and toxicology. This would recreate the golden era of discovery and exploration of

early toxicology and would stand in marked contrast to the present situation where the sacrifice of the many millions of rats and mice used in regulatory studies per year worldwide results in little new data of general use in biological or medical research, and in remarkably few (published) interesting discoveries. If, as proposed, the effects of a wide variety of compounds on experimental animals were studied by 2-DE as part of routine toxicological and pharmacological studies, an extremely valuable database of compounds affecting specific regulational pathways in the genome operating system would be developed. It is therefore our view that addition of 2-DE to routine toxicology would enormously and profitably invigorate both pharmacology and toxicology, providing that the data was properly analyzed and made available. The results would allow the identification of the major toxic agent in complex mixtures, provide new lead compounds for drug discovery projects, and provide reagents for dissecting regulational pathways.

4.2 Global protein analysis and drug development

We have chosen to stress global protein analysis by 2-DE over many other important applications that could be mentioned because we believe it will provide the major driving force for much further research and development. A historical perspective of paradigms in drug development over time is illustrated diagrammatically in Fig. 7. The general direction has been toward the development of automated molecular-level screening techniques, the production of large sets of compounds for screening, and computerized drug design. Selected compounds are subjected to safety assessment in relatively standardized animal studies including histological examination of tissues and long-term carcinogenesis studies. Many of these procedures are prescribed by regulatory agencies. Thus the trend has been toward more and more complex studies at ever increasing expense. Estimates for the current cost of taking one new chemical entity all the way to use extrapolate to approximately $400 million for 1995 [46]. Clearly anything that would reduce or eliminate whole stages in this costly and time-consuming process would be welcome.

The most costly outcomes are failure at a late stage of development, or withdrawal of a drug after introduction. These failures are largely due to negative effects seen either in animals or in man. An experimental animal is the most complex detector available, and 2-DE provides a method for reading multiple outputs using small tissue samples. This suggests the almost heretical idea of doing 2-DE-based animal studies very early in drug development. Positive results would indicate the drug has been absorbed and distributed, and pattern changes would indicate mechanisms of action and toxicity. The options appear to be either many animals with few and often subjective end points, or quantitative multifactorial 2-DE analyses involving a thousand or more proteins on a few animals. Clearly an improvement in use of animals as detectors can lead to substantial decreases in overall animal use. As the 2-DE database of drug effects becomes larger, as associations are made between pharmaceutical effects and individual spot changes, as known and trusted histological markers (peroxisome, mitochondrial, or endoplasmic reticulum proliferation) are shown to correlate closely with changes in 2-DE patterns, and as 2-DE analyses are automated, the cost of drug discovery and development may be expected to drop while also becoming more efficient.

Note that 2-DE mapping is also applicable to the evaluation and validation of alternatives to animal testing. Either the alternative exhibits the same molecular responses to a drug as does the model and is therefore a valid alternative, or it does not.

4.3 Screening for chemotherapeutic and anti-human immmuodeficiency virus (HIV) drugs

A more speculative, but no less important role for 2-DE lies in drug discovery screening itself. Screening for chemotherapeutic or anti-viral (including anti-HIV) drugs with cells in culture assumes that the same or analogous molecular targets are expressed in these cells as in patients. For example, in the current US National Cancer Institute (NCI) *in vitro* screen, 60 cell lines derived from human tumors are used for the initial screen, and positive compounds are then tested against some of the same cells in nude mice [47]. An initial investigation of these cell lines, carried out in collaboration with John

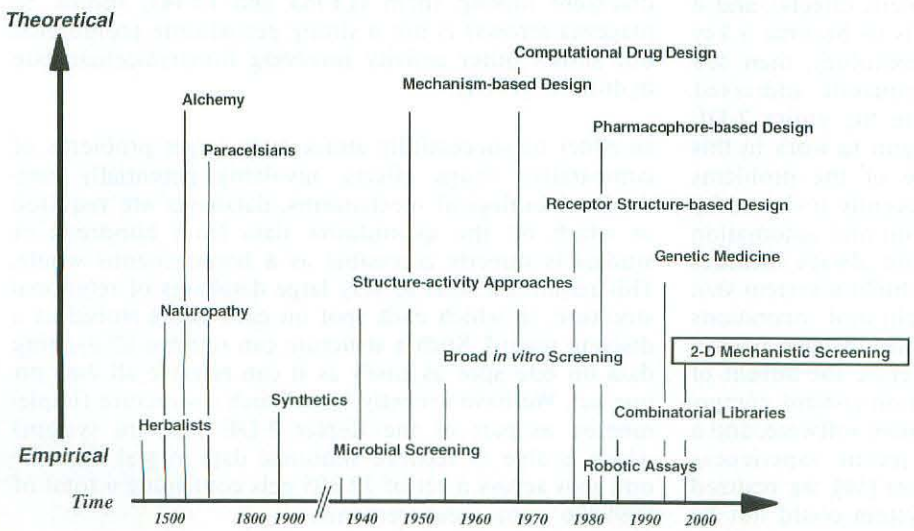

Figure 7. A schematic representation of trends in drug discovery paradigms.

Weinstein of NCI, suggests that 2-DE can provide fundamental new data from such *in vitro* screens (data to be published elsewhere) that may be directly relevant both to the selection of cell lines and to the choice compounds for further study.

With regard to screening for anti-HIV drugs, the current doctrine is that there is more chance of therapeutic success with a combination of drugs than with a single agent or "magic bullet". Ideally the combination should include drugs which have different targets. The HIV genome and its gene products interact with host cell components in ways that have not been fully explored, but which produce both up- and down-regulation of host proteins as well as the expression of viral ones [48]. These interactions themselves provide a large number of potential targets for new drugs, and the challenge is to develop a screening system which will identify compounds which hit these interaction targets.

In the present NCI HIV screening system, cytotoxicity assays are used, which turn up large numbers of marginally effective compounds, and thus far, no magic bullets. In practice, effective drugs are often developed from lead compounds which are not fully effective (are not magic bullets), but which uncover a useful mechanism. The core problem is how to distinguish those few compounds which work by some new unique mechanisms, and which are potential lead compounds for more effective drugs. We and others have therefore proposed to develop a screening system in which the effects of candidate agents on protein changes occurring in infected target cells are measured using quantitative 2-DE. This would require detailed quantitative analyses of the time course of changes in gene expression after infection, the development of automated systems for doing large numbers of analyses, and programs for analyzing and intercomparing very large numbers of patterns.

5 Technical barriers: gel automation, database construction and data visualization

If 2-DE is to achieve many of the goals that may be allocated to it in a revised paradigm of molecular biology (*i.e.*, analysis of regulation and epigenetic effects), and if global quantitative protein mapping is to become a key element in drug discovery and in toxicology, then several technological issues must be squarely addressed. First, it is now essential to automate the entire 2-DE process. Several laboratories have begun to work in this direction, including our own. Some of the problems involved in automation have been recently reviewed by Peccaud [49]. Note than mechanization and automation are not synonymous since automation always includes feedback control. The challenge is to build a system that performs an extended series of physical operations under tight control, something that would have been a difficult and expensive proposition before the advent of fast, small computers, advanced motion-control components, object-oriented instrument control software, and a variety of sensors. Based on our recent experiences building a large scale DNA synthesizer [50], we realized early on that an automated 2-DE system could not be assembled from commercial bench-top robotic instruments, but rather had to be designed and built using regular industrial components. As expected, the prototype 2-DE system that is emerging is large and complex: it is 35 feet long, run by three computers, incorporates a total of 24 axes of rotation and 47 separate parameter controllers. We have elected to accommodate gel formats up to 50 cm (IEF) × 25 cm (SDS). The design objective is continuous operation without operator involvement, from autoinjection of samples from barcoded vials, to quantitative data deposited in a relational database. Much of the hardware and software is currently on hand, and we expect full operation on a developmental basis in 1996.

A second major requirement is emerging in data analysis and visualization. Most 2-DE work to date has been organized in discrete experiments, comparing, for instance, a set of treated samples with appropriate controls. Such experiments often involve hundreds of gels, and hundreds of thousands of protein abundance measurements. Hence the major requirement has been for means of 'summarizing' complex quantitative effects. The difficulty of developing, analyzing and presenting such summaries is one of the main factors accounting for the relative paucity of large-scale quantitative 2-DE studies. Historically 2-DE data has been presented in conventional forms such as labeled maps, as bargraphs showing changes occurring in the integrated densities of individual spots during an experiment (as in Fig. 6), or as numerical tables. None of these provide an intuitive means for comparing two or more complex quantitative patterns of gene expression change in the manner required for routine use. To address this problem, we have recently developed 'arrowplots' (Fig. 8) that show simultaneously the magnitude, polarity and significance of changes in many spots with multiple treatments. Such displays are generated in real-time on a PC screen and allow comparative perusal of large data sets. In the case shown [51], the overall similarity of effect of six compounds, which represent five significantly different structural classes of compounds producing peroxisome proliferation in rat liver, is quite striking, and accurately reflects the likelihood that they operate through binding to a single DNA-binding receptor (PPAR). The most divergent among them (LY163 and LY443, shown by magenta arrows) is not a strong peroxisome proliferator, but shows other activity involving fumarylacetoacetate hydrolase (FAH).

In order to successfully attack even larger problems of comparative drugs effects, involving potentially hundreds of biological mechanisms, databases are required in which all the quantitative data from hundreds of studies is directly accessible as a homogeneous whole. This requires a shift to very large databases of relational structure, in which each spot on each gel is stored as a discrete record. Such a structure can retrieve all existing data on one spot as easily as it can retrieve all data on one gel. We have recently tested such a structure (implemented as part of the Kepler 2-DE software system) which is able to retrieve statistical data in real time for one spot across a set of 12 305 gels containing a total of 8099806 spot measurements.

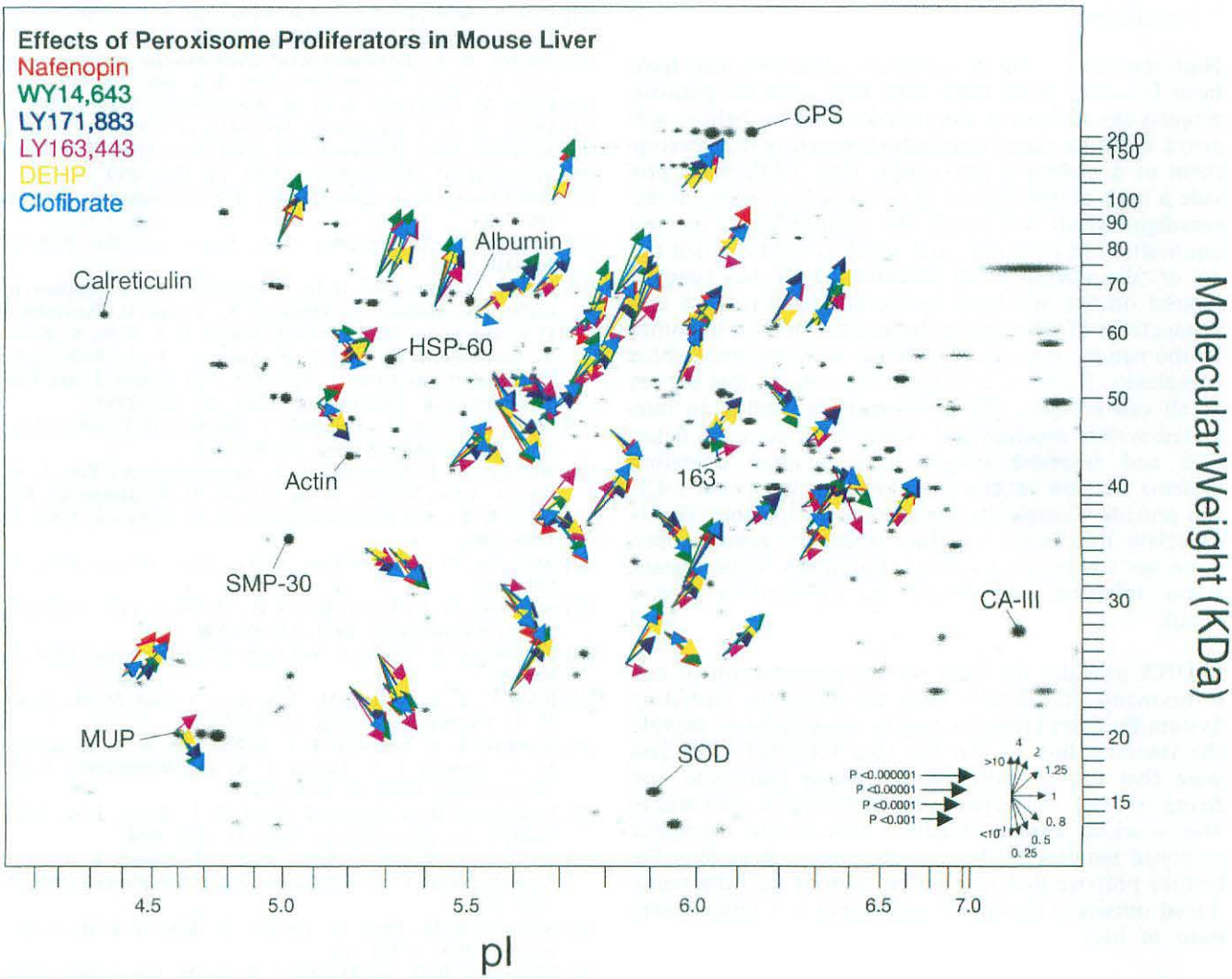

Figure 8. An 'arrowplot' of comparative protein changes within a series of six compounds, five of which are known peroxisome proliferators (and hence hepatocarcinogens) in rodents [51]. The background picture is the master 2-DE gel pattern of proteins from B6C3F1 mouse liver. Approximately 1500 spots are shown, ranging from high molecular mass (300 kDa at the top) to low (13 kDa at the bottom) and an isoelectric point of 4.0 at the left to 7.0 at the right. Short arrows project from selected protein spots showing quantitative changes in treated animals versus controls: each arrow expresses the effect of one agent on one spot by summarizing the statistical difference between measurements on liver samples from treated animals (5 per group) versus measurements on a set of parallel control animals. Arrow angle indicates relative magnitude and direction of change (abundance ratio treated/control), and arrow length indicates t-test statistical significance of the difference (as shown in the legend at lower right). Red arrows show result for nafenopin, green for WY 14,643, blue for LY171,883, magenta for LY163,443, yellow for diethylhexylphthalate (DEHP) and cyan for clofibrate. Some proteins are labeled for reference: CPS is carbamoyl phosphate synthase, SMP-30 is the senescence marker protein-30, MUP is the major mouse urinary protein, CA-III is carbonic anhydrase III, HSP-60 is the 60 kDa heat shock protein of mitochondria, SOD is superoxide dismutase, and 163 is a protein strongly induced by peroxisome proliferators but absent from current sequence databases.

6 Nucleic acid *vs.* protein-based monitoring of gene expression patterns

Differential display of eukaryotic messengers [52], and quantitative versions of it, provide means for identifying genes expressed in different cell types, and a measure of quantitative differences in gene expression [53, 54]. For studies on the effects of drugs and other agents on gene expression, quantitation in the range of ±10–15% is required. At present this can be obtained with 2-DE, but not with nucleic acid-based techniques. Only protein-based methods allow post-translational protein processing (including proteolytic cleavage, glycosylation, sialation, or phosphorylation) to be detected. There is no necessary close quantitative relationship between mRNA abundance and the steady-state abundance of its protein product, since the rate of turnover of individual proteins varies widely. Further, there is at present no method for determining the intracellular location of a protein from DNA sequence or mRNA-based studies. Lastly, while the function of a gene can be guessed by comparison with the sequence of genes coding for proteins of known function, proof of function still requires biochemical studies using the proteins themselves. High resolution multicomponent protein analysis will therefore continue to occupy an important place in biological research.

7 Conclusions

High resolution 2-DE, first introduced in 1975, may have been far ahead of its time; only now, after the genome projects are well along, can it make what we believe will prove to be its main contribution, which is the development of a molecular physiology. Thus 2-DE maps provide a body of information that makes most sense in the paradigm which will follow the current phase of "the enunciation of all genes", and which could be called the era of "the analysis of all functions". 2-DE has been inhibited on the one hand by the comparative ease and productivity of DNA-based approaches* and on the other by the natural aversion of scientists towards unasked-for complexity. However, to explore the concept that the set of all cell proteins (the proteome) constitutes an integrated system incorporating within itself complex detection and response systems (the genome operating system) we now require the global analyses that 2-DE can provide. Complexity has become inescapable. This is especially true as the responses which the genome operating system makes to drugs, hormones, toxic agents, stress, infection, and disease are explored in greater detail.

If DNA provides the rules for the construction of cell components (the LAW), then the Genome Operating System Program Elements may be considered to provide the interpretation of that law (the GOSPEL). We propose that the key difference between inanimate and living systems, and between biochemistry *in vitro* and *in vivo* – which Pasteur sought – lies in the organized response network of the genome operating system. We further propose that this system cannot be fully reproduced outside a living cell, and hence is a unique attribute of life.

Received August 31, 1995

8 References

[1] Weinberg, A. M., *Reflections on Big Science*, MIT Press, Cambridge, USA 1967, pp. 182.
[2] Anderson, N. G., Anderson, N. L., *Am. Biotechnol. Lab.* Sept./Oct. 1985.
[3] Consden, R., Gordon, A. M., Martin, A. J. P., *Biochem. J.* 1944, *38*, 224–232.
[4] Anderson, N. G., *The Development of Zonal Centrifuges and Ancillary Systems for Tissue Fractionation and Analysis*, National Cancer Institute Monograph No. 21, NIH, Bethesda, MD 1966.
[5] Wankat, P. C., *Sep. Sci. Technol.* 1984–5, *19*, 801–829.
[6] Gianazza, E., Righetti, P. G. R., in: Righetti, P. G., Van Oss, C. J., Vanderhoff, J. W., (Eds.), *Electrokinetic Separation Methods* Elsevier/North Holland Biomedical Press, Amsterdam 1979, pp. 293–311.
[7] Klose, J., *Humangenetik* 1975, *26*, 231–243.
[8] Scheele, G. A., *J. Biol. Chem.* 1975, *250*, 5375–5385.
[9] Iborra, G., Buhler, J. M., *Anal. Biochem.* 1976, *74*, 503–511.
[10] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.
[11] Celis, J., Bravo, R. (Eds.), *Two-dimensional Gel Electrophoresis of Proteins*, Academic Press, Orlando 1984.
[12] Dunbar, B. S., *Two-Dimensional Electrophoresis and Immunological Techniques, Plenum Press*, New York 1987, pp. 372.
[13] Dunn, M. J., Burghes, A. H. M., *Electrophoresis* 1983, *4*, 97–116.
[14] Dunn, M. J., *J. Chromatogr.* 1987, *418*, 145–185.
[15] Anderson, N. G., Anderson, N. L., *Clin. Chem.* 1982, *28*, 739–748.
[16] Vesterberg, O., *Electrophoresis* 1993, *14*, 1243–1249.
[17] Hochstrasser, D. F., Tissot, J.-D., *Adv. Electrophoresis* 1993, *6*, 268–382.
[18] Anderson, N. G., Anderson, N. L., *Behring Inst. Mitt.* 1979, *63*, 169–210.
[19] Celis, J. E., Rasmussen, H. H., Gromov, P., Olsen, E., Madsen, P., Leffers, H., Honoré, B., Dejgaard, K., Vorum, H., Kristensen, D. B., Østergaard, M., Haunso, A., Jensen, N. A., Celis, A., Basse, B., Lauridsen, J. B., Ratz, G. P., Andersen, A. H., Walbum, E., Kjaergaard, I., Andersen, I., Puype, M., Van Damme, J., and Vandekerckhove, J., *Electrophoresis* 1995, *16*, 2177–2240.
[20] Wirth, P. J., Luo, L., Fujimoto, Y., Bisgaard, H. C., and Olson, A. D., *Electrophoresis* 1991, *12*, 931–954.
[21] Wasinger, V. C., Cardwell, S. J., Serpa-Paljak, A., Yan, J. X., Gooley, A. A., Wilkins, M. R., Cuncan, M. W., Harris, R., Williams, K. L., and Humphery-Smith, I., *Electrophoresis* 1995, *16*, 1090–1094.
[22] Anderson, N. L., Anderson, N. G., *Electrophoresis* 1991, *12*, 883–906.
[23] Anderson, N. L., Esquer-Blasco, R., Hofmann, J.-P., Anderson, N. G., *Electrophoresis* 1991, *12*, 907–930.
[24] Giometti, C. S., Taylor, J., Tollaksen, S., *Electrophoresis* 1992, *13*, 970–991.
[25] Baker, C. S., Corbett, J. M., May, A. J., Yacoub, M. H., Dunn, M. J., *Electrophoresis* 1992, *13*, 723–726.
[26] Kovalyov, L. I., Shishkin, S. S., Effimochkin, A. S., Kovalyova, M. A., Ershova, E. S., Egorov, T. A., and Musalyamov, A. K., *Electrophoresis* 1985, *16*, 1160–1169.
[27] VanBogelen, R. A., Sankar, P., Clark, R. L., Bogan, J. A., Neidhardt, F. C., *Electrophoresis* 1992, *13*, 1014–1054.
[28] Latter, G. I., Boutell, T., Mondaro, P. J., Kobayashi, R., Futcher, B., McLaughlin, C. S., and Garrels, J. I., *Electrophoresis* 1995, *16*, 1170–1174.
[29] Abersold, R. H., Pipes, G., Hood, L. E., Kent, S. B. H., *Electrophoresis* 1988, *9*, 520–530.
[30] Rasmussen, H. H., Van Damme, J., Puype, M., Gesser, B., Celis, J. E., and Vanderkerckhove, J., *Electrophoresis* 1991, *12*, 873–882.
[31] Patterson, S. D., *Electrophoresis* 1995, *16*, 1104–1114.
[32] Bjellqvist, B., Pasquali, C., Ravier, F., Sanchez, J.-C., and Hochstrasser, D. F., *Electrophoresis* 1993, *14*, 1357–1365.
[33] Görg, A., Postel, W., Günther, S., *Electrophoresis* 1988, *9*, 531–546.
[34] Görg, A., Boguth, G.,. Obermairer, C., Posch, A., Weiss, W., *Electrophoresis* 1995, *16*, 1079–1086.
[35] Sanchez, J.-C., Appel, R. D., Golaz, O., Pasquali, C., Ravier, F., Bairoch, A., and Hochstrasser, D. F., *Electrophoresis* 1995, *16*, 1131–1151.
[36] Strohman, R., *Bio/Technology* 1994, *12*, 156–164.
[37] Kuhn, T. S., *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago 1970, 210p.
[38] Suzuki, D. T., Griffith, A. J. F., Miller, J. H., Lewontin, R. C., *An Introduction to Genetic Analysis*, W. H. Freeman and Co., New York 1986, 612pp.
[39] Hubbard, R., *Amer. Sci.* 1995, *83*, 8–10.
[40] Maddox, J., *Nature* 1992, *355*, 201.
[41] Bray, D., *Nature* 1995, *376*, 307–312.
[42] Bilheimer, D. W., Grundy, S. M., Brown, M. S., Goldstein, J. L., *Proc. Natl. Acad. Sci. USA* 1983, *80*, 4124–4128.
[43] Anderson, L., Steel, V. K., Kelloff, G. J., Sharma, S., *J. Cellular Biochem.* 1995, *22* (Suppl.), 108–116.
[44] Anderson, N. L., Swanson, M., Giere, F. A., Tollaksen, S., Gemmell, A., Nance, S., and Anderson, N. G., *Electrophoresis* 1986, *7*, 44–48.
[45] Anderson, N. L., Giere, F. A., Nance, S. L., in: Galteau, M.-M., Siest, G. (Eds.), *Progres Recents en Electrophorèse Bidimensionelle*, Presses Universitaires de Nancy, Nancy 1986, pp. 253–260.
[46] DiMasi, J. A., Hansen, R. W., Grabowski, H. G., Lasagna, L., *J. Health Econ.* 1991, *10*, 107–142.

* Ivan Lefkovits has pointed out that if one simple fact of biology had been different (if restriction enzymes had not existed), then the protein technologies would have had to fill the role currently occupied by the DNA technologies, and the history of 2-DE would have been vastly different.

[47] Boyd, M. R., Shoemaker, R. H., McLemore, T. L., Johnston, M. R., *Thoracic Oncol.* Part 7, Chapter 51, W. B. Saunders, Philadelphia 1986.
[48] Kettman, J. R., Robinson, R. A., Kuhn, L., Lefkovitz, I., *Electrophoresis* 1991, *12*, 554–569.
[49] Peccoud, J., *Bio/Technology* 1995, *13*, 741–745.
[50] Anderson, N. G., Anderson, N. L., Taylor, J., Goodman, J., *Appl. Biochem. Biotechnol.* 1995, *54*, 19–42.
[51] Anderson, N. L., Esquer-Blasco, R., Richardson, F., Foxworthy, P., Eacho, P., *Toxicol. Appl. Pharmacol.* 1996, in press.
[52] Liang, P., Pardee, A. B., *Science* 1992, *257*, 967–971.
[53] Schena, M., Shalon, D., Davis, R. W., Brown, P. O., *Science* 1995, *270*, 467–470.
[54] Velculescu, V. E., Zhang, L., Voglestein, B., Kinzler, K. W., *Science* 1995, *270*, 484–487.