REVIEW

How microbial proteomics got started*

Frederick C. Neidhardt

Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI, USA

Received: December 13, 2010
Accepted: January 18, 2011

Publication in 1975 by Patrick O'Farrell of a procedure to separate the proteins of *Escherichia coli* in a two-dimensional array on polyacrylamide gels marked the birth of the field now called proteomics. Although O'Farrell's contribution was soon to have wide ranging effects on research in many fields, the initial impact was greatest in the arena of whole cell physiology. Refinements and amplification of the original procedure, including improved standardization and reproducibility of gel patterns, introduction of techniques to measure the quantity of protein in individual spots, and biochemical identification of the protein spots, afforded investigators the ability to explore for the first time the integrated working of control circuits in the living cell. From O'Farrell's contribution has grown the rich array of techniques currently employed and still being developed in the diverse field of microbial proteomics.

Keywords:
Integrative biology / Microbial physiology / Microbiology / Proteomics origin

It is an honor to be asked to write an introductory review on the origin of microbial proteomics for this special issue of *Proteomics*. Professor Michael Hecker, himself a pioneer in the field, has graciously invited me to say whatever I wish in this celebratory article. Such an invitation invites all sorts of mischief: personal recollections as opposed to reality… personal memory as opposed to historical fact.

Given the dangers inherent in such a loose charge, I shall adopt three measures to protect myself and you, the reader, from the most egregious errors. First, I shall confine myself to what I personally experienced, rather than guess what was in the minds of the dozens of sterling colleagues who were fellow progenitors of the field of microbial proteomics. Their stories are their stories, and as valid as mine.

Second, I shall confine myself to the origin of *microbial* proteomics, and not deign to trace how a single 1975 paper [1] by a single researcher blossomed into a major field currently occupying the skilled research efforts of countless individuals and international teams across the whole breadth of biology. Indeed, a full issue of *Proteomics* could not accommodate such a review.

Third, I shall not foolishly believe that I can validly give credit to all my colleagues, known and unknown, who learned as quickly as I did how the germinal event of microbial proteomics in 1975 would forever change the direction of bacterial physiology. Some colleagues are mentioned… those whose proximity to the author allowed us to share the excitement of the times. Many others are omitted, partly because of my ignorance, and partly through failing memory. To this latter group I feel especially apologetic. So, without further defense, here is my personal account of the origin of microbial proteomics.

My involvement in proteomics actually predates the origin of this field, in the sense that I had been subconsciously waiting for years for genomics to happen. I believe I was not alone in that regard; a number of us biologists experienced the same ''aha!'' moment in 1975 because we had been waiting for just such a revolutionary approach to hitherto intractable problems in cell physiology.

The explanation is simple. For most of the 20th century, the study of cell physiology was largely reductionistic; the living cell was taken apart and studied biochemically. In the fortunate case of bacterial studies the powerful marriage of biochemistry and genetic analysis led to notable triumphs, culminating in the field of molecular biology. Nevertheless, there were those of us who, consciously or subconsciously, were motivated by a desire to understand completely the workings of a cell, and for us, frustration seemed to grow in proportion to the success of the reductionistic approaches.

Correspondence: Professor Frederick C. Neidhardt, 7954 S. Galileo Lane, Tucson, AZ 85747-9609, USA
E-mail: fcneid@umich.edu
Fax: +1-520-305-4217

*Portions of this account appeared in: Neidhardt F. C., How Proteomics Got Started. *Small Things Considered* blog. December 21, 2009.

We benefited from the development of the central dogma; we cheered as the pathways of metabolism became clear; and many of us contributed to the growing knowledge of how bacteria regulate the molecular expression of their genes. Yet it was all too evident that these foundations failed to provide a means to answering our special questions about the living cell. Furthermore, we did not possess even the means to address our questions in experimentally useful terms.

Allow me to explain how this situation applied to me. As a graduate student under Boris Magasanik at Harvard Medical School during the early 1950's I encountered questions for which there was no precedent in microbial cell biology. My project involved cell growth and the induced synthesis of enzymes in bacteria. I was delighted that bacterial growth could be monitored turbidimetrically with a Klett colorimeter, while the same instrument could provide colorimetric assays of enzyme activities. I appreciated my good fortune in having a mentor who did not require me to purify a protein. (Around me at that time in Harvard University's Department of Bacteriology and Immunology, now Microbiology and Molecular Genetics, were gifted individuals who on occasion were forced to purify proteins using laborious and personally onerous techniques. Not a life for me, I decided, even though H. Edwin Umbarger assured me that purifying an enzyme ''developed character.'')

Besides laziness, there was a second, more fundamental, reason I never purified a protein. Cell growth provided the *raison d'etre* for my interest in bacteria, and work that began by smashing cells into little bits seemed to me to be destroying the very object of interest.

For me, therefore, the limitations of reductionistic biology centered on two major aspects of microbial physiology involving proteins and their relation to cell growth rate: *catabolite repression* (or, more generally, how bacterial cells choose to utilize multiple carbon sources), and *growth rate modulation* (how bacterial cell size and composition are interrelated with growth rate). Catabolite repression dealt with the reality that bacterial (and yeast) cells faced with alternative sources of carbon and energy almost invariably used them sequentially in the order of the growth rate supporting capacity of these substrates [2]. Growth rate modulation included a set of fundamental laws of bacterial growth established in the early 1960's that related the size and composition of bacterial cells to the overall growth rate supporting property of their environment [3].

The vexing nature of these subjects was that the cellular outcomes of catabolite repression and of growth rate modulation were eminently easy to rationalize on the basis of their selective advantage to the cell, yet were bereft of molecular explanation (all too familiar a situation throughout biology). Nothing known about the *lac* operon allowed a biochemical explanation of the variation in β-galactosidase production under various growth conditions, and nothing in the central dogma explained the partitioning of macromolecular synthesis amongst proteins, RNA and DNA as a function of growth rate. Only now, 50 years later, are these processes approaching mechanistic solutions.

But many other questions could not be approached during the first three quarters of the 20th century by microbiologists motivated by the goal of understanding cellular life. They all emerged from observations that cellular components operate differently in the context of the whole cell than in isolation; or, to put it another way, that the behavior of the whole cannot be predicted by the properties of the individual parts. Here are some:

(i) What prevents bacteria of a given species from growing at the same rate on all carbon and energy sources?
(ii) Is there a growth rate-limiting step during balanced (steady-state) growth of a culture of bacteria? If so, what is it?
(iii) How many changes are made in a bacterium transitioning from growth to non-growth?
(iv) What adjustments enable most bacterial species to grow over a temperature range of 40 Celsius degrees?
(v) Given options, how do bacterial cells prioritize their choice of food?

This list reflects the author's interests, and can easily be expanded by those curious about other aspects of the integrated activity of a cell (as, for example, cell division).

By the mid 1970's, my mind, filled with such unanswered questions about growth physiology, was searching for a new way to approach the bacterial cell. The way was revealed, not by anyone in my laboratory, but by a graduate student named Patrick O'Farrell at the University of Colorado at Boulder. Steen Pedersen was a postdoctoral fellow in my laboratory at the University of Michigan at that time. He was one of the keenest of disciples of Ole Maaløe of the University Institute of Microbiology in Copenhagen (and one of his most honest critics). In 1974, Steen returned from a visit to Colorado and reported to our laboratory that a graduate student there had produced a two-dimensional polyacrylamide gel (2-D gel) process that could resolve many of the proteins of an *Escherichia coli* bacterial culture on an array that looked as cool as ''the sky on a starry night.''

Steen appreciated instantly the significance of O'Farrell's success, and his news electrified us. We realized that a fundamentally new approach to bacterial growth physiology had become possible. Instead of asking the cell for information about a protein of interest to us, *we could finally interrogate the cell about the proteins important to the cell* in any given situation. And *we could finally observe the integrated behavior of the entire panoply of gene regulatory devices*. This new power provided, we felt, the road to a global analysis of cell physiology. It is clear that the era of proteomics began in 1975 – the date of publication of Patrick O'Farrell's thesis research in *The Journal of Biological Chemistry*. His paper [1] was quickly recognized by a variety of molecular biologists

as a true technological breakthrough. Citations in the next 30 years numbered over 16 000 (in spite of the fact that the manuscript was initially rejected with two disparaging reviews which had to be overruled eventually by members of the journal's editorial board).

Before we could now learn what the cell had to teach us about its complement of proteins and about adjustments to different environmental conditions, this new ability to listen to the cell required the addition of several features to the O'Farrell technique.

First, we recognized that we had to standardize the 2-D gel system of O'Farrell in order to compare the protein arrays from different samples. This required extreme attention to details of procedures and quality of regents. The genius of O'Farrell's system was that it employed two independent properties of proteins to separate them: their molecular weight and their isoelectric point. IEF in a gel tube containing ampholines to establish a pH gradient produced the first dimension: proteins lined up by their charge. Placing the resulting tubular gel on an electrophoretic gel slab containing sodium dodecyl sulfate, allowed the polypeptides previously resolved by charge now to be segregated by their size. The resulting 2-D gel was then stained and dried for subsequent inspection. A beautiful picture – but to be useful, 2-D gels had to be reproducible, and this was not an easy task for a number of reasons. In the end it took years of perfecting sample preparation and gel casting (not to mention improvements in ampholines) to get to the stage where computer-driven pattern matching could align a whole series of ''starry patterns'' from the multiple samples of an experiment.

Second, once the pattern-matching problem was in hand (no small feat), the issue became one of accurate measurement of the quantity of protein in the individual spots across the gel set. Clever uses, first of isotopes, then of differentially colored samples, were devised to obtain reasonable quantification. As a result, it became possible for the cell to display much of the array of changes made in its proteome (the totality of its several thousand proteins) as the cell adapted to its environment.

Fortunately, these tasks of standardizing and quantifying O'Farrell gels were approached by many individuals skilled in scientific technology. James Garrels then at Cold Spring Harbor Laboratory, Norman G. and N. Leigh Anderson at Argonne National Laboratory, and Julio Celis at the University of Aarhus, Denmark, were some of the people who early on used their considerable skills to refine and expand the usefulness of 2-D gel technology.

But still a third attribute had to be added to 2-D gels for maximum usefulness: the identities of the ''starry'' spots on the gels had to be determined. For the bacterium *E. coli* and its close cousins, my laboratory in Ann Arbor mounted a full-scale effort to identify spots on the 2-D gels with known proteins. Hundreds of protein spots were identified through the use of purified proteins (naturally donated by others – you would not catch me purifying a protein) and mutants in known genes [4]. Everyone in my laboratory contributed to this effort; unfair as this is, I'll single out only two because of their germinal work in identifying spots and because of their tireless energy in teaching the 2-D gel process to all the others: Dr. Ruth A. VanBogelen and Ms. Teresa Phillips.

Needless to say, the identification of spots might be regarded as tedious drudgery – and it was – save for the thrill that we were simultaneously making discovery after discovery using the 2-D gels: heat-shock and cold-shock proteins, proteins under stringent control, proteins that vary monotonically with growth temperature, proteins that vary with growth rate – and we were not simply learning which proteins exhibit a certain behavior, but what *fraction* of the cell's proteome was involved in different physiological responses to stress or starvation. These discoveries led Dr. VanBogelen and her colleagues to the concept of *protein signatures* [5]. A *protein signature* is the set of proteins that by their amplification or suppression, signal a particular physiological stress state of the cells. One learned how to recognize when a cell was in a state of energy starvation, or oxidative stress, or membrane damage, or… the list goes on. One can imagine the gigantic usefulness of this approach when a pharmaceutical company is exploring how a potentially useful therapeutic agent acts.

But we should bring this story to a close quickly, because from the mid 1990's onward the explosion of cell protein technology transformed the field from what Pat O'Farrell had created to one with a formidable arsenal of techniques for protein resolution and measurement. The term *proteome* was introduced in 1996 [6] to refer to the totality of proteins in a cell, and this quickly gave rise to the noun, *proteomics*, to designate studies of the proteome [7]. The 2-D gel technique introduced by Pat O'Farrell has inspired others to develop improved techniques for monitoring the global pattern of a cell's total protein complement. The availability of DNA sequences with reasonably accurate annotations for the genomes of hundreds (thousands?) of species has made it possible to develop separation techniques that enable tandem mass spectrometry to provide the ''second dimension'' to primary fractionation procedures, and as a result, enable protein identifications an order of magnitude beyond that which was achieved in the first two decades of the 2-D era.

To be sure, the current armamentarium of proteomics is being used in highly targeted ways to explore previously identified sets of ''proteins of interest,'' but I want to emphasize that Pat O'Farrell's development of the first method of spreading out the proteins of a cell was at the start, and particularly for me, the initiation of an exciting new grammar of scientific questioning. This new grammar is essential to the goal of solving (i.e. modeling) a living cell.

The papers in this special issue on *Microbial Proteomics* point to the richness of current research in microbial proteomics. Stress responses, posttranslational modifications, the proteomics of special environmental communities, molecular pathogenesis, and exciting propects for

industrial and pharmaceutical applications are all well represented. It will be exciting to track what is learned in the coming years.

What does the future hold for the special interest of those of us who recognize the contribution of proteomics to solving a cell? There are many roads being followed. I draw attention to one path as being a logical extension of the original use of proteomics for an integrative and total systems analysis of a cell. The protein–protein interactions within the cell are being explored by techniques that permit one to determine protein–protein contacts in situ. The use of tandem affinity purification combined with mass spectrometry (TAP-MS) [8] allows one to begin defining the functional organization of the proteome within the cell [this volume, Chapter by R. Herrmann et al. ''The proteome of a minimal organism''] and [this volume, Chapter by P. Noirot ''A cluster of hubs within a bacterial protein–protein interaction network: functional exploration by an integrative approach'']. This work should provide one more layer of the information needed before we can claim to understand life at the microbial cell level. Other layers will surely be needed as well.

Finally, permit me a personal note. Were a historian of science to tell a story that illustrates the defining features of scientific exploration in our era, he or she would do well to choose the history of proteomics. Within such a narrative would appear such themes as:

(i) the frustrating dependence of thoughtful investigation on key technical advances;
(ii) the role of the single, inspired (usually young) scientist in introducing an astounding breakthrough;
(iii) the collaboration of individuals from assorted scientific fields to improve upon and expand on an initial technical advance;
(iv) the excitement of scientists whose questions about the nature of things depended on the new discovery;
(v) the manner in which chemists, physicists, statisticians, image analysts, and systems analysts can collaborate to bring a new field to maturity;
(vi) the collaboration of basic and applied scientists fostered through the new technique; and finally,
(vii) the demonstration of the international aspects of science in this era.

To have played a role in this great story is one of the pleasures of my life.

The author has declared no conflict of interest.

References

[1] O'Farrell, P. H., High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 1975, *250*, 4007–4021.

[2] Neidhardt, F. C. and Magasanik, B., Effect of mixtures of substrates on the biosynthesis of inducible enzymes in *Aerobacter aerogenes*. *J. Bacteriol.* 1957, *73*, 260–263.

[3] Schaechter, M., Maaløe, O., and Kjeldgaard, N. O., Dependence on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. *J. Gen. Microbiol.* 1958, *19*, 592–606.

[4] Van Bogelen, R. A., Abshire, K. Z., Pertsemlidis, A., Clark, R. L., Neidhardt, F. C., Gene-protein database of *Escherichia coli* K-12, in: Neidhardt, F. C. (Ed. in Chief), Curtiss III, R., Ingraham, J. L., Lin, E. C. C. et al. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 6th Edn, Chapter 115, ASM Press, Washington, DC 1996, pp. 2067–2117.

[5] VanBogelen, R. A., Schiller, E. E., Thomas, J. D. and Neidhardt, F. C., Diagnosis of cellular states of microbial organisms using proteomics. *Electrophoresis* 1999, *20*, 2149–2159.

[6] Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K. et al., From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Nat. Biotechnol.* 1996, *14*, 61–65.

[7] James, P., Protein identification in the post-genome era: the rapid rise of proteomics. *Q. Rev. Biophys.* 1997 *30*, 279–331.

[8] Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M. et al., A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 1999 *17*, 1030–1032.

