Metrology in Life Sciences

White Paper for the Technology Innovation Program (TIP)
of the National Institute of Standards and Technology

Written by Zeev Smilansky

Anima Cell Metrology

December 2008

Contact: Dr. Zeev Smilansky

75 Claremont Road, Suite 102
Bernardsville, NJ 07924-2270
Voice: 908.766.6633
Fax: 908.766.6655
Mobile: 908.963.6200
Email: Zeev.Smilansky@Animapsl.com

Key words: Genomics, Proteomics, Measurement, Cell, Metrology, Technology, Systems Biology

Groping in the Darkness of a Cell

Novel biological research technologies, such as Genomics and Proteomics, raised high hopes of creating a paradigm shift- breakthroughs that will yield a new understanding of cellular processes and human disease, and pave the way to a bounty of new drugs and therapeutics. Unfortunately, first for Genomics and then for Proteomics, it became abundantly clear that the data produced, though of extreme interest, was insufficient for the anticipated breakthroughs [Miklos 01, 04]. So many puzzle pieces are still missing that the clear view of cellular machinery remains hidden from our eyes. As for the state of pharmaceutical technology, crisis is a word often used. In fact, despite huge increases in investments, the pace of drugs entering the market has not increased [http://www.fda.gov/cder/rdmt/default.htm].

Though there is more than one reason for this, a salient point is that while available genomic, proteomic and metabolic data is of a vast scale, the complexity of cellular machinery, not to mention tissue and whole organism, is on a much grander scale still. We are still groping our way in the darkness of a vast, complex system of which our understanding is scant and meager. It is not a coincidence that some novel pharmaceutical approaches are called "rational drug design": the name focuses our attention on the serendipity inherent in the business of pharmacology.

It is a fact that despite the wealth of targets discovered through current biological research tools, drug attrition rates remain extremely high. Often, information that can wipe out a compound is revealed only at late stages and at a high cost – sometimes only after spending

some unfortunate time on the market. Some recent examples are Merck's Cox-2 inhibitor Vioxx (developed with the "rational" paradigm), Bayer's cholesterol lowering Baycol, and American Home Products (now Wyeth) obesity drug Phen-fen.

In addition, etiology of numerous debilitating diseases remains obscure, and effective treatments remain as elusive as ever. Little progress has been made in the last decades on Autoimmune diseases, learning and memory impairment (e.g. fragile-x mental retardation, autism), cell/tissue proliferation impairments (e.g., Cardiac Hypertrophy, arteriosclerosis and Restenosis), multiple-factor diseases such as diabetes and metabolic syndrome ("syndrome X"), neurological disorders such as Muscular Dystrophy and public safety concerns such as antibiotic resistant bacteria.

We would like to argue that what is needed to make biology a real science and pharmacology a veritable engineering discipline is the development of new measurement tools and technologies, in particular for measuring dynamic data from within living cells, tissues and organisms. Data in biology and pharmacology is notoriously difficult to obtain; accurate, dependable data from living cells is almost unavailable. Only with such data can a quantitative – versus descriptive – scientific and engineering discipline be developed. There can be no exact science without exact data.

The Metrology Gap

Below we list some examples of available / desirable data and their methods of metrology.

1. Genes and genomes: our ability to measure and sequence genes and genomes goes back to the discovery of the polymerase chain reaction (PCR) in 1984 – barely 25 years ago. Today it is relatively easy to obtain sequence data of single genes and even entire genomes – quickly, accurately and cost effectively: it is expected that a complete human genome will be measurable within a few hours and for a few thousand dollars within the next few years [Mardis 2008].
2. Transcripts and transcriptomes: we can currently measure gene expression either for a specific gene product, using RT-PCR, or in large scale, using DNA microarrays (DNA chips) that allow the mRNA content of a sample to be measured. This technology is almost 20 years old, but the quality and reliability of the data it produces is still being actively debated [Draghici 2006].
3. Protein data: Analysis of proteins is much more complex than that of DNA or RNA, due to the distinct chemical properties of each protein and the wide dynamic range of protein expression. The two mainstay technologies of protein measurement and identification are 2D-Gels, invented in the 1970s, which are quantitative but lack sensitivity and accuracy; and mass spectrometry, developed about 10 years ago, which is sensitive and accurate but not quantitative. While massive efforts have been spent on the study of proteins, only a few thousand proteins of the 100,000 or so predicted in the human proteome have a function confidently assigned to them. In addition, critical attributes of proteins, such as correct folding, co- and post-translational modifications, subcellular localization, and interactions with other proteins are notoriously difficult to determine [Miklos 2001, Görg 2004, Domon 2006].

4. Live cell data: The cell is dynamic, ever changing, responding to its environment. However almost all our current knowledge about cells comes by analyzing cell extracts. Methods of measuring, recording, analyzing and monitoring the functioning of living cells, including their response to external stimuli and drug targets – are manifestly absent from the toolbox of current biology. Among the general processes most appropriate for being measured in living cells are transcription, translation, co- and post-translation modification, trafficking, and protein degradation.

   Systems biology is a new paradigm, proposing to study cells in ways similar to those use by engineers in studying machines, electronics or software. "*Identifying all the genes and proteins in an organism is like listing all the parts of an airplane. While such a list provides a catalog of the individual components, by itself it is not sufficient to understand the complexity underlying the engineered object. We need to know how these parts are assembled to form the structure of the airplane. … to understand how a particular system functions, we must first examine how the individual components dynamically interact during operation.*" [Kitano 2002]. Clearly, for Systems Biology to take off, it needs data that is of similar accuracy, availability and dependability as that available to aircraft, electronics or software engineers. In fact, we would need more data, since a single cell is orders of magnitude more complex than an airplane, a computer chip, or the most complex software program.

5. Tissue, organ, and organism: measuring live cells growing in a Petri dish is a great challenge; measuring cells in an intact tissue or organ is a challenge of an altogether different magnitude. The scope of what there is to be measured and how such measurement technology can be crafted is indeed breathtaking. A sampling of new concepts and brave projects are:

   a. Cardiology: in a recent NHLBI workshop on Systems Approach to Understanding Electromechanical Activity in the Human Heart, the first recommendation to NHLBI was "*Improve the quality and reliability of human cardiac electromechanical data…*" [http://www.nhlbi.nih.gov/meetings/workshops/electro.htm]

   b. Human Plasma: The Human Plasma Proteome Institute bravely undertakes to *"…study this complex fluid, whose thorough analysis holds the promise of a revolution in disease diagnosis and therapeutic monitoring…*", provided technical hurdles such as 10 orders of magnitude ratios in protein concentrations can be overcome [Anderson 2002]

   c. Metabolomics: this old-new discipline attempts to identify traces of chemical fingerprints that cellular processes leave behind. Metabolomics could yield a systems view of the operational organism, as well as important contributions to diagnostics and therapeutics. However, metabolomics will first *"…have to overcome some teething pains. Researchers have matched thousands of metabolites to cellular processes, but thousands more remain unidentified. And software's analytical powers are sinking fast beneath the rising tide of new information.*" [Daviss 2005]

   d. Cancer: a solid tumor becomes malignant and lethal when its cells begin to metastasize – migrate to new sites and grow there, eventually killing a critical or-

gan. Only a small fraction of the tumor's cells migrate; of these, a very small proportion sprout in a new location [Tarin 2006]. It is critical to develop ways to recognize these highly potent cells, initially ex-vivo and eventually in the blood stream; it is also critical to understand why such cells are able to sprout in specific tissues and not others. While the general process is fairly well understood and accepted, the technology should be developed that would be able to characterize cells ex-vivo, in-vivo and in the blood stream, as well as tissues, to enable early diagnosis and new therapies to be developed.

Economic and Social Impact

Healthcare is clearly recognized as an area of critical national importance. The United States faces increasing healthcare monetary demands, in part due to the increase in the aged population. Hospitalization and home care costs, as well as the cost of drugs, continue to rise. The latter has been, over the last few years, the fastest growing expense item in the total cost of health care (Reference: *The Factors Fueling Rising Healthcare Costs 2006: Prepared for America's Health Insurance Plans*, January 2006, Price Waterhouse Coopers, Pages 11, 12, 14 http://pwc.com/us/eng/about/ind/healthcare/pubfuel.html). At the same time, the challenge to find appropriate, cost effective treatments for numerous debilitating diseases is of prime importance. For example, the "war on cancer", while making important advances, still has a long way to go – 560,000 US citizens die of cancer each year, and malignant cancers almost invariably have no cure.

When considering productivity gains, current benchmarks in the pharmaceutical industry come to mind. According to the Tufts Center for the Study of Drug Development, current cost of drug development in the US is $1.2Bn *(source: Tufts CSDD press release, Nov 9, 2006: http://csdd.tufts.edu/NewsEvents/NewsArticle.asp?newsid=69)*, mostly spent on expensive trials of drug candidates that fail at the late stages of development due to the inability to predict efficacy or safety issues. Success rates are now very low: only one of almost 100 drug candidates, and only one of every ten drugs entering clinical trials will make it to the market. *(Sources: http://www.ftc.gov/be/workpapers/wp262.pdf; http://www.purdue.edu/dp/ptec/aicheV5.pdf)*.

Recognition of the Challenge

In the FY 2009 Administration Research and Development Budget Priorities report, John H Marburger and Stephen S. McMillin write, in the section entitled "Understanding Complex Biological Systems": "*Access to new biotechnological tools and increasing amounts of genetic sequence data opens new avenues for research into the functional implications of gene expression… Agencies should focus research at the:*

- *cellular/sub-cellular and the organism/population/community levels; and*
- *interface of the life, physical and computational sciences.*"

In a recent NIH program announcement (PA-07-452, New Technology for Proteomics and Glycomics), the authors write: " Proteomics technologies and methods remain largely inadequate, particularly with respect to quantitative and real time measurements… "

In 1998, the National Institute of General Medical Studies (NIGMS) held a workshop on complex biological systems (interestingly, this had no recent follow up). At the same year, the NIGMS issued a program announcement entitled "Quantitative Approaches to the Analysis of Complex Biological Systems (PA-98-024)". In the opening paragraph, the authors write: "*The purpose of this initiative is to support new quantitative approaches to the study of complex, fundamental biological processes by encouraging non-traditional collaborations across disciplinary lines. The National Institute of General Medical Sciences (NIGMS) will … support … physicists, engineers, mathematicians, and other experts with quantitative skills relevant to the analysis of complex systems.*"

The recent emergence of the discipline of "systems biology" points attention to some of the same issues that we address in this document. In 2005, The World Technology Evaluation Center (WTEC - a technology assessment organization supported by several US government agencies, including NIH, NIST, NFS and DARPA) published a report entitled "Assessment of International Research and Development in Systems Biology" (http://www.wtec.org/ sysbio/). In the executive summary, the authors of the report write "*… For the past 40 years the paradigm for predicting phenotype has focused on single gene defects. This extraordinarily powerful approach has been the major contributor to an understanding of the function of individual genes and proteins. It seems less likely that it will yield an understanding of complex biological behavior, from individual cellular activities such as motility to the operation and integration of organ systems. Indeed, the underlying assumption for all the excitement surrounding systems biology is that phenotype is governed by the behavior of networks, rather than simply the consequence of individual gene action. In its essence systems biology is the development of approaches to the understanding of biological networks and consequently to the determination of biological phenotype… Understanding input-output behavior of the network… is more effectively reached through the systems approach, since network behavior is more complex than can be understood intuitively.*"

Clearly, the success of systems biology will depend on the availability, accuracy and overall quality of the data it proposes to use for developing its models. Systems biology emphasizes the modeling and analysis of biological data, while we focus here on the foundations prerequisite for this task - measuring data that is unmeasurable today, in particular dynamic, live data. Clearly, modeling could only be as useful and as accurate as the input data it uses to produce its models.

Finally, in "Modeling the heart: from genes to cells to the whole organ" [Noble 2002], Prof. Denis Noble (Oxford) says: "*successful physiological analysis requires an understanding of the functional interactions between the key components of cells, organs, and systems, as well as how these interactions change in disease states. This information resides neither in the genome nor even in the individual proteins that genes code for. It lies at the level of (extensive) protein interactions within the context of subcellular, cellular, tissue, organ, and system structures. … The rapid growth in biological databases; models of cells, tissues, and organs; and the development of powerful computing hardware and algorithms have made it possible to explore functionality in a quantitative manner all the way from the level of*

genes to the physiological function of whole organs and regulatory systems. Systems physiology of the 21st century is set to become highly quantitative and, therefore, one of the most computer-intensive disciplines."

We beg to add: provided these computer farms can be fed with quantitative, accurate, sensitive data, that someone found a way to measure.

Why TIP Funding is Essential

The reason why a program such as TIP is a necessary enabler is simple: almost no-one funds development of measurement tools for biology and pharmacology.

First, the US investment community has been badly disappointment by technology companies in the 2001 bust, most notably by momentous data gathering flops such as Celera and Millenium. Ever since, the term "platform technology" has become a forbidden term; companies developing technology do their best to hide this fact and portray themselves as a "drug discovery" or "drug development" companies, in some cases even rushing to buy a couple of molecules to solidify their new persona. Venture capital, on its part, will quickly mark off any company resembling a "technology platform" and not even bother to consider what the company is proposing to achieve.

Second, government funding agencies, most notably the NIH, with their highly esteemed peer-review system, usually favor projects that can show impressive "initial results" and have good chances of success. The peer review process rightly respects taxpayers' money and places its funds on relatively safe bets. In addition, development of a technology in the life-science discipline is often considered to be too "technological" and insufficiently "scientific" – unless it has the context of a well established discipline. Thus, determination of a new of a new protein structure would be considered properly scientific in the established science of X-Ray crystallography; a project suggesting development of an entirely new microscopy technique would be hard-pressed to find a program that will fund it.

Finally, what organization would be better placed to promote the science and technology of measurement than the National Institute of Standards and Technology? What better flag for NIST to be raising in the 21st century than that of bringing metrology to the life-sciences?

Conclusion

The two sister disciplines, biology and pharmacology, have made enormous strides forward over the last three decades. Still, it is commonly agreed that our understanding of cell, tissue and organism is meager. It is a fact that a majority of drug candidates fail phase three trials – after throwing away hundreds of millions of dollars. This would be totally unacceptable in any other scientific or engineering discipline pursuing a multi-billion dollar project – for example in designing a new passenger airline, a new wafer fab, or even a new space vehicle. Can anyone imagine the majority of space shuttle missions being lost? But this is the situation in biology and drug development: an undertaking worth $1.2B fails, more often than not, when its efficacy is tested. Such failures can lead to devastating effects such as Pfizer exiting the cardiology field after failure with its CTEP inhibitor Torcetrapib, and Bayer dramatically downsiz-

ing cardiology after failure with Baycol. In fact, as a result of these failures, cardiology drug development is facing very sparse times.

We contend that a large part of today's shiny new biotech relies on a small number of key advances in metrology: the discovery of the polymerase chain reaction (PCR) in 1984 by Kary Mullis at Cetus (a development which had nothing to do with Cetus' "business model", took months to convince the management to devote some attention to, and is the only asset remaining from that company), the development of the sequencing machine, discovery of fluorescent proteins, and the development of protein identification by mass spectrometry.

Let us conclude with this story by Lee Hood, one of the developers of today's most celebrated measurement instrument – the sequencing machine: "*In the late 1970s, a friend approached me and said it was certainly unfortunate that only my group had the highly sensitive protein sequencer. Why didn't I commercialize it and make it available to the scientific community…. I went to 19 companies with the fully developed protein sequencer and a vision of the three other instruments (the DNA and protein synthesizers and the DNA sequencer) and how collectively they would transform biology. Not one of the 19 companies I visited was interested and after three visits to the one company I thought would be an ideal partner, Beckman Instruments, I was told not to come back. To say I was discouraged is an understatement…. Shortly thereafter, I gave a lecture to the Caltech trustees on the vision of how our four instruments would change the world of biology. One of the Trustees, Arnold Beckman, came up to me afterwards and said, "This is fascinating. It is just what my company needs." I pointed out that his company had already turned me down three times. After some additional hesitation, Murph Goldberger finally agreed… to start the company that became Applied Biosystems. Today, Applied Biosystems is world leader in molecular instrumentation*" [Hood 2002].

Today, it seems, developing and commercializing measurement instruments for biology is as challenging as it was 30 years ago.

References

[Miklos 2001] Miklos GL and Maleszka R, *Protein functions and biological contexts*, Proteomics 2001 1(2):169-78

[Anderson 2002] Anderson, NL and Anderson, NG, *The Human Plasma Proteome, History, Character, And Diagnostic Prospects,* 2002, Molecular & Cellular Proteomics 1:845–867.

[Hood 2002] Hood L., *My Life and Adventures Integrating Biology and Technology,* A Commemorative Lecture for the 2002 Kyoto Prize in Advanced Technologies

[Kitano 2002] Kitano H, *Systems Biology: A Brief Overview*, 2002, Science 295:1 1662-1664

[Noble 2002] Noble, D. *Modeling the Heart--from Genes to Cells to the Whole Organ,* 2002, Science 295:1, 1678-1682

[Görg 2004] Görg A, Weiss W, Dunn MJ., *Current two-dimensional electrophoresis technology for proteomics.* Proteomics. 2004 Dec;4(12):3665-85

[Miklos 2004] Miklos GL and Maleszka R, *Microarray reality checks in the context of a complex disease*, Nat Biotechnol. 2004 22(5):615-21

[Daviss 2005] Daviss B, *Growing Pains for Metabolomics*, 2005, The Scientist 19(8):25

[Domon 2006] Domon B and Aebersold R, Mass Spectrometry and Protein Analysis, 2006, Science 312:5771, 212 - 217

[Draghici 2006] Draghici S., Khatri P., Eklund A.C., Szallasi Z. *Reliability and reproducibility issues in DNA microarray measurements*, 2006 Trends in Genetics, 22 (2), pp. 101-109.

[Tarin 2006] Tarin, D. *Breast Disease* 26, 13-25 (2006-2007).

[Mardis 2008] Mardis ER, *Next-generation DNA sequencing methods*, Annu Rev Genomics Hum Genet. 2008;9:387-402.