Cell Signalling—The Proteomics of It All

Keith Ashman,[1] Michael F. Moran,[2] Frank Sicheri,[1] Tony Pawson,[1] and Mike Tyers[1]*

(Published 9 October 2001)

Cellular behavior arises from a labyrinthine network of protein-protein interactions that signal and integrate environmental and internal cues (*1*). Although the term "proteome" was coined to simply describe the full complement of proteins encoded by the genome (*2*), proteomics now usually connotes the global dynamics of protein function, which derives mainly from highly regulated protein-protein interactions. Numerous signal transduction pathways, often based on evolutionarily conserved cassettes of associated proteins, underlie the cellular response to specific stimuli. Such intracellular signaling proteins are often constructed in a modular fashion of independently folding domains that bind specific peptide motifs, phospholipids, nucleic acids, or other small molecules (*1*). These interaction domains serve to recruit proteins to an appropriate subcellular location and to direct their association with regulators and targets. In addition, separate proteins involved in common signaling events are frequently bound to the same scaffolding protein, thereby increasing the efficiency and specificity with which signals are conveyed. Although originally described in the context of signal transduction pathways, protein interactions now appear to organize a wide range of cellular activities, including protein and vesicle trafficking, the cell cycle, gene expression, protein degradation, and DNA repair, among others.

Increasingly, understanding the mechanics of individual pathways is within our grasp. However, the connections between pathways (often referred to as cross talk), the means through which they functionally compensate for one another, and the modulations that allow a single pathway to elicit quite different biological responses in distinct cells have largely eluded understanding. This is due in part to experimental difficulties in simultaneously monitoring responses in many pathways, particularly under physiological conditions, and in part to technical limitations in the analysis of protein networks. Thus, although we have, in principle, all of the pieces of the cellular puzzle in the form of genome sequence information, we have little idea as to how the cell functions as an integrated whole. It is accepted wisdom that the deluge of genomic and proteomic data will somehow lead to insights into such biological complexity. Here, we briefly review current and anticipated methods for proteomic analysis, and comment on the bioinformatics approaches needed to make sense of the bewildering flux of information through protein-protein interaction networks.

Proteomic Platforms

Signal transmission; metabolic processes, such as DNA, RNA, and protein synthesis; and the many specialized functions that occur within a given cell type are frequently orchestrated by multiprotein complexes. Such complexes are subject to cycles of regulated assembly and disassembly--or even destruction--and their localization within the cell is often tightly controlled and essential for their function (*1*). The events are often regulated by phosphorylation and by other covalent additions or processing that can affect protein interactions, stability, activity, and subcellular localization. To enable such dynamics, protein-protein interactions are frequently weak and thus often at the very edge of experimental detection. Several approaches have been devised to capture such interactions. The yeast two-hybrid system, pioneered by Stan Fields (*3*), has transformed the analysis of protein-protein interactions through its sensitivity and ease of use. This now-standard method is based on the ability of any given protein-protein interaction pair to couple a transcription activation domain to a DNA binding domain, and thereby to drive the expression of sensitive reporter genes in yeast. Various extensions of the two-hybrid system, including automated screens against highly complex libraries, have produced a wealth of new protein interaction data (*4-7*). However, there are some well-appreciated caveats to the method, including its tendency to generate a high level of false-positive and false-negative results, the fact that it detects only binary interactions, and the obvious absence of a physiologically relevant context (*8*). Indeed, there is surprisingly little overlap in the interactions detected by separate large-scale two-hybrid screens undertaken with the yeast proteome (*8*). Ideally, then, one would rather isolate native protein complexes from their true biological setting, followed by direct identification of each substituent. Until recently, this direct biochemical approach was limited to abundant or extremely stable complexes, or required months of arduous protein purification in the cold room, or both. But now, a veritable revolution is at hand with the advent of highly sensitive mass spectrometric identification methods (*9*).

Protein identification by mass spectrometry relies on the proteolytic cleavage of proteins into short peptides, ionization of the peptides either by electrospray (ESI) or matrix-assisted laser desorption (MALDI), and accurate peptide mass determination (*10*). In the ESI method, the liquid sample is physically sprayed and rapidly desolvated before entry into a mass analyzer, typically a quadrupole, which measures mass by the degree of deflection under a finely tuned electric field. In the MALDI method, samples are mixed with a chemical matrix, then ionized by brief laser pulses and mass analyzed in a time-of-flight (TOF) tube, which measures the time it takes ions to travel a given distance in a strong electric field (*9*). For single proteins, a comprehensive peptide mass map obtained by MALDI-TOF is often sufficient to unambiguously identify the protein by a database search of all possible tryptic peptides (*11*). However, a much more powerful method, called tandem mass spectrometry (MS/MS), introduces the additional step of peptide ion fragmentation at the amide bonds, which directly yields protein sequence information (*12, 13*). Numerous variations of these

[1]Programme in Molecular Biology and Cancer, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Canada, M5G 1X5. [2]MDS Proteomics, 251 Attwell Drive, Toronto, Canada, M9W 7H4.

*To whom correspondence should be addressed: E-mail: tyers@mshri.on.ca

steps have been developed; for brevity, we focus only on the most capable current platforms.

A key factor for unambiguous protein identification is the quality and accuracy of the mass spectra, particularly when considering complex protein mixtures from higher organisms, such as humans. At present, the surest route to this end is through automated liquid chromatography (LC)-MS/MS, which separates peptide mixtures over a steep reverse phase high pressure liquid chromatography (HPLC) gradient. The peptides are delivered directly from the column outlet into an electrospray source-equipped mass spectrometer (*14*). The mass spectrometer enables the detection, selection, and fragmentation of individual peptides from a tryptic digest. The resultant mass spectra represent fragments generated from a selected precursor peptide ion (Fig. 1). Typically, the peptides break at peptide bonds and therefore the product fragment ions differ in mass by the particular amino acid residue liberated during the fragmentation. Algorithms that interpret the precursor-product relationship of the MS/MS spectra are then employed to interrogate DNA and protein sequence databases to identify the protein from which the peptide originated (Fig. 2) (*10*). With current technology, proteins

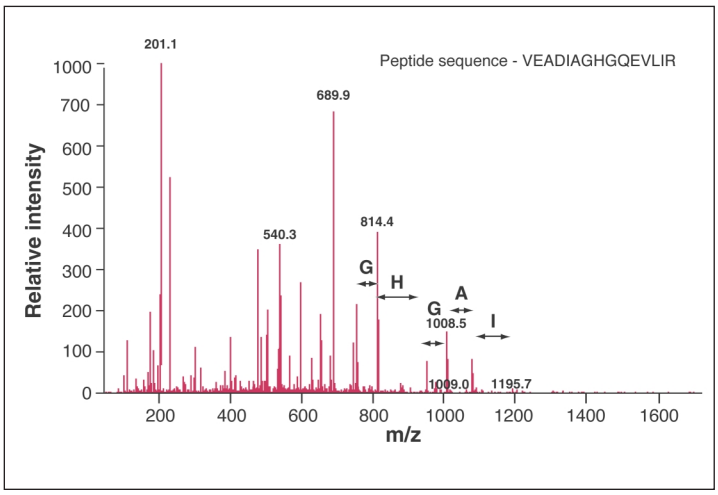

Fig. 1. A peptide fragmentation spectrum obtained by qTOF analysis. m/z is the ion mass to charge ratio. The amino acids are indicated in one-letter code.

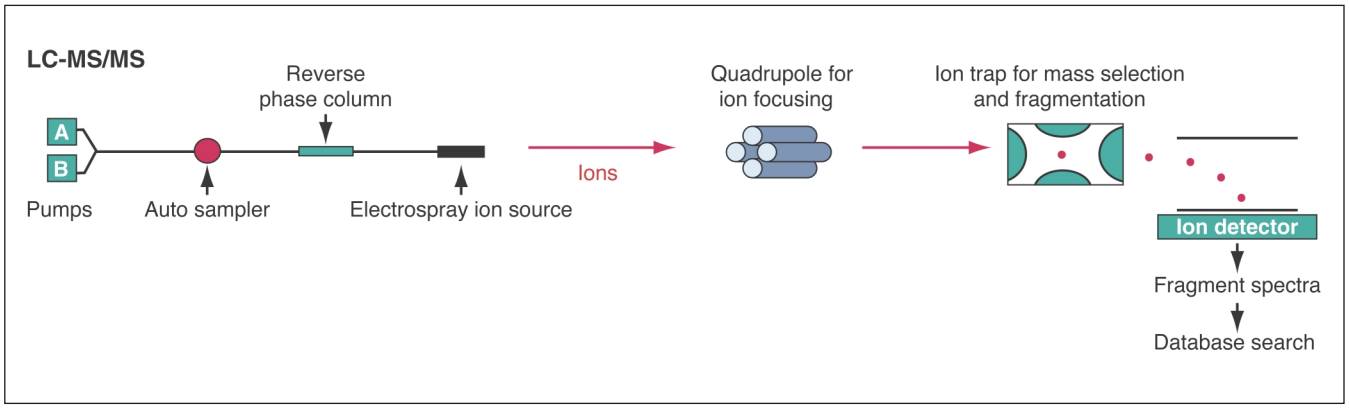

Fig. 2. A schematic of liquid chromatography-tandem mass spectrometry (LC-MS/MS).

can be readily identified down to a limit of approximately 100 fmol loaded onto a polyacrylamide gel by LC-MS/MS (*9*).

Several competing mass spectrometric technologies are on the horizon. In a method referred to as Orthogonal MALDI-quadrupole TOF (O-MALDI-qTOF), the MALDI ion source replaces the conventional electrospray ion source (*15*, *16*). This decouples the MALDI ion source from the TOF mass analyzer, so that peptides can be selected for MS/MS by a quadrupole and transmitted to a collision chamber for fragmentation, as in an electrospray instrument (Fig. 3). This approach yields good-quality MS/MS spectra and has the important advantage of trapping the samples in the solid phase matrix, which simplifies sample handling and is therefore amenable to automated high throughput analysis. Another nascent approach involves adaptations of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR-MS) and allows the analysis of mixtures containing thousands of peptides with extremely high accuracy (*17*). In FT-ICR-MS, peptide ions are constrained within an intense magnetic field and display a measurable resonance behavior at precise frequencies related to their mass-to-charge ratio (Fig. 4) [reviewed in (*18*)]. The amplitude of the signal generated is proportional to the number of ions in the sample and Syka *et al.* (*18*) recently described a hybrid instrument that accumulates ions upstream of an FT-ICR-MS analyzer that is able to characterize peptides present in low attomolar amounts (*19*). The highly accurate masses obtained by FT-ICR-MS allow very low tolerance thresholds during database searches and greatly improves the selectivity of protein identification with a minimum number of peptides identified per protein. When coupled with automated sample preparation, the O-MALDI or FT-ICR-MS methods are capable of making thousands of protein identifications per day per instrument, with little user intervention.

Isolation of Signaling Complexes

All mass spectrometric methods are necessarily limited by the initial biochemical step of protein complex purification, an area that lags far behind the technological sophistication of the mass spectrometers themselves. The balance between complex purity and

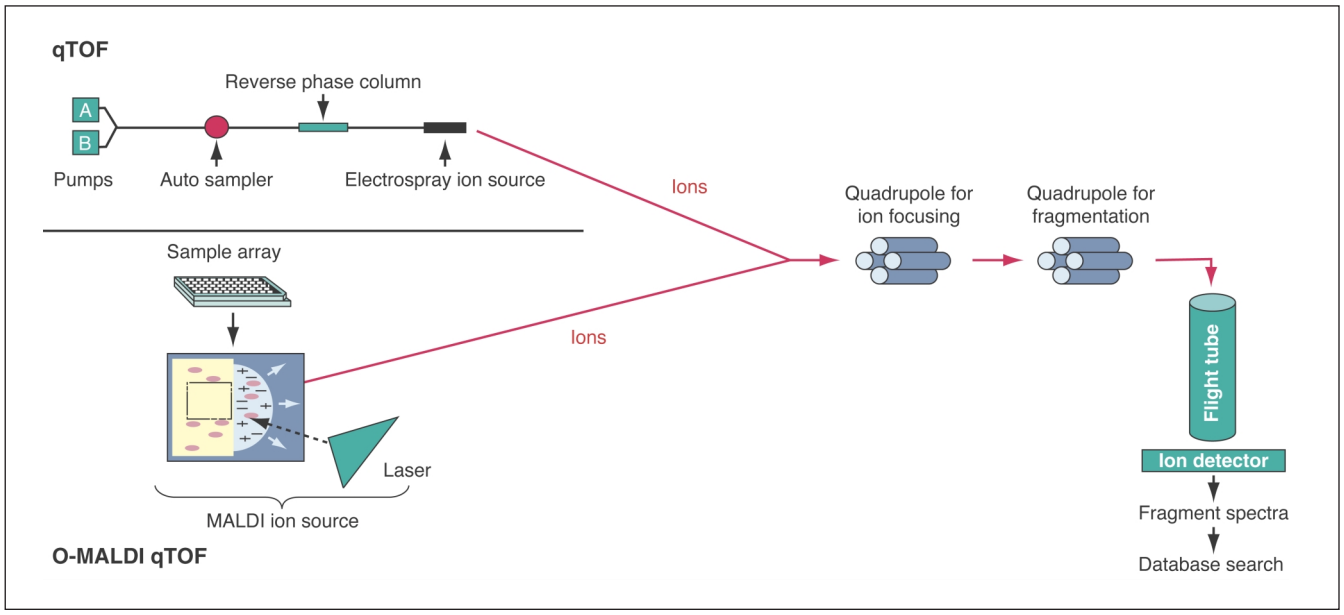

Fig. 3. A schematic of orthoganol matrix-assisted laser desorption quadrupole time-of-flight (O-MALDI-qTOF) analysis.

contamination by nonspecific background proteins forms the yin-yang of protein complex isolation. Delicate complexes tend to fall apart under all but the very mildest conditions. Inevitably, though, such conditions lead to high levels of background caused by nonspecific adsorption to the affinity resins and to tube and tip surfaces encountered during complex isolation. Currently, protein complexes are typically captured in simple one- or two-step purification procedures. Expression of a recombinant, tagged protein of interest in transfected cells allows recovery of the bait protein and any associated proteins on a solid resin that binds the tag with high affinity. Similarly, a high-affinity antibody, preferably a monoclonal that recognizes a single exposed epitope, can be used to capture native complexes from untransfected cells or tissues. Alternatively, for proteins that can be expressed in sufficient yield, cell or tissue lysates can be passed over a protein affinity column that selectively retains interacting proteins before stringent elution from the column (*20*).

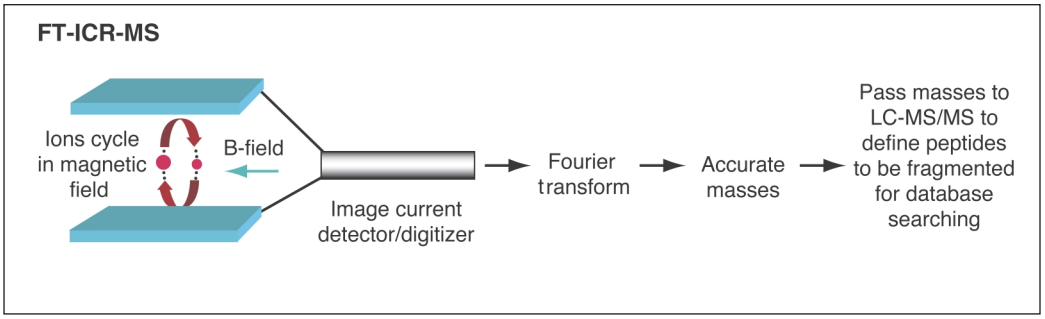

Fig. 4. A schematic of Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS).

Each approach has its strengths and weakness. Affinity tagging has the advantage of allowing complex assembly to occur in the living cell, but suffers from the constraint of transfection in cultured cells. At present, the direct antibody approach is hampered by the lack of high-affinity antibodies to most proteins, especially antibodies that efficiently precipitate protein complexes. The protein affinity column approach can detect weak interactions directly from primary cell culture or tissues, but many proteins of biomedical interest are large and difficult to express. The latter problem can be circumvented to some extent by using columns made of isolated protein domains, which are often more amenable to high-level expression, but these in turn may be susceptible to nonphysiological interactions. In sum, there is no perfect method for complex isolation, so an obvious need exists for new capture methods and surface chemistries to minimize losses during complex isolation.

At present, reliable identification of individual constituents usually requires polyacrylamide gel separation of the protein complex. This step serves not only to resolve multiprotein complexes into isolated subunits, but also yields qualitative information on subunit stoichiometry and the degree of the inevitable background contamination. A cost of the gel step is incomplete recovery of peptide fragments after digestion and an overall sample loss due to liquid handling. Soon, though, gel separation will be replaced by "gel-free" methods that directly digest the complex on the capture resin. The limitation in gel-free analysis has been the ability to deconvolve the complex spectra generated by multiprotein mixtures. At least for abundant complexes, such as the 26S proteasome (*21*), the gel-free approach is now feasible. One simple potential solution to the

background contamination problem is high-throughput analysis of many different protein complexes under uniform isolation conditions, which empirically defines background binding, thereby allowing specific signals to be discerned.

Global Protein Interaction Networks

An ambitious and now attainable goal is a comprehensive protein interaction map of the human proteome. If, as now seems likely, virtually all facets of cellular function are controlled through the assembly of proteins into specific complexes, the global analysis of protein interaction networks provides an approach to discerning the cell's functional organization. Aside from needing more mass spectrometry firepower, a current bottleneck to achieving this goal is the lack of a unified, complete set of human complementary DNAs (cDNAs) in appropriate expression vectors. This problem is not trivial, given the difficulty in obtaining full-length cDNA clones for many genes. A number of academic groups and companies have assembled large collections of mammalian cDNAs, both mouse and human, but none of these clones are in expression-ready format, nor are the collections anywhere near complete. To rectify this situation, an industry-academia consortium called FLEXgene (for Flexible Expression) has been formed under the leadership of Ed Harlow and the Harvard Institute for Proteomics (http://www.hip.harvard.edu/). The primary aim of FLEXgene is to assemble a fully sequence-validated set of all human cDNAs in a recombinational cloning system, such that large clone sets can be moved into any desired expression system with ease. The cost of this project, which is already well under way, is estimated at over $100 million over a period of three years. In the meantime, systematic proteomics in mammalian systems will have to rely on a ragtag collection of expression constructs in various formats.

A tremendous amount of information stands to be gained by assembling protein interaction networks from more tractable model eukaryotes, such as yeast, for which complete sets of affinity tagged proteins already exist [for example, see (*22*)]. The evolutionary conservation of many cellular functions will allow human interaction networks to be deduced from those of model organisms. Conversely, much of the increased complexity in cells of higher eukaryotes apparently results not from a wholesale increase in numbers of genes and proteins, but the incorporation of new interaction domains and motifs into preexisting proteins, thereby endowing them with additional connections and functions.

Equally important, lessons regarding the structure and dynamics of complex protein networks will be learned first from analysis of model proteomes (*23*). One issue concerns the very definition of a protein complex. In a sense, proteins form "valence" shells of secondary interactions, tertiary interactions, and so on. Presumably, these higher order interactions will connect and coordinate all aspects of cellular behavior. Just how to define these boundaries and discriminate them from nonspecific background has not yet been tackled. Here again, the blanket approach of one-pass, high-throughput analysis of all complexes in the cell may reveal compelling connections that may otherwise have been overlooked.

Finally, there is the issue of combinatorial protein function and, in metazoans, the critical problem of tissue specificity. The number of proteins encoded by the human genome seems inadequate to the task of building a complex organism, despite the fact that alternative splicing and posttranslational modifications greatly increase the number of protein isoforms. One resolution to this conundrum may be that proteins function in combination with one another (that is to say, in protein complexes) and that the same protein may associate with different combinations of partners in distinct cells, or in response to distinct signals, yielding a variable biological output. This promiscuity could greatly expand the information content of the proteome, but will also complicate its analysis. As one example, a series of modular proteins, each with multiple protein-protein interaction domains, has been implicated in the control of epithelial cell polarity and in asymmetric cell division, two processes that appear crucial for the organization of tissues and organs in animals. In *Drosophila*, these proteins assemble into complexes that control cell fate, but the nature of the interactions and thus the effect on cell division differ from one cell type to another (*24*). Most of these proteins (for example, PAR-3, PAR-6, Numb, and Dlg) are conserved in vertebrates, which are likely to exhibit more complex ways of regulating cell polarity and asymmetry.

Signaling Dynamics

Perhaps the most important aspect of proteomics to understand is how the protein circuitry is switched from state to state. Protein complexes are often continuously remodeled at numerous levels, including subunit composition, subcellular localization, and associated enzymatic activities. More often than not, complex formation is controlled by numerous phosphorylation events, although other forms of posttranslational modification, including proteolytic processing, acetylation, methylation, glycosylation, prenylation, sulfatation, polyadenosine diphosphate (poly-ADP) ribosylation, and covalent attachment of small protein modifiers, such as ubiquitin, can dramatically affect protein interactions. The systematic detection of protein phosphorylation remains a challenge for current mass spectrometric methods, in part because phosphoserine and phosphothreonine residues are very labile and often lost during the peptide fragmentation step. However, advances in MS intrumentation and approaches involving the chemical modification of peptide phosphoamino acids suggest that the global analysis of cellular phosphoproteins may be attainable (*25*-*27*). Applications of these techniques on a proteome-wide scale, and having the bioinformatics tools to facilitate the interpretation of such complex data, may provide measures of cellular physiology, which would be of enormous benefit in assessing and comparing cellular states. For example, an ability to compare such phosphoprofiles between diseased and normal tissues or those treated or not with drug candidates may be a powerful method to diagnose and treat disease in a patient-specific fashion.

Protein microarrays are another promising emergent method for systematic detection of protein interactions (*28*). This approach is analogous to DNA microarrays in that individually purified proteins are spotted at high density on a solid support, which is then probed with any other protein or compound of interest. Proof-of-concept results for peptide and small molecule binding have been obtained for protein arrays corresponding to the entire yeast proteome (*22*). Similarly, enzymatic assays have been carried out with systematic arrays of all yeast kinases (*29*, *30*). Obstacles to these types of array-based approaches include the ever-present background issue for complex protein mixtures, the limited amounts of protein that can be captured for detection, and the enormous range of physicochemical properties displayed by proteins.

Databases and Modeling: Assembling the Whole from the Parts

Suppose for the moment that all relevant protein interaction data were in hand--how would one visualize such a vast network, much less use the data set to predict cellular behavior? Before

modeling efforts can begin, a unified database that captures all aspects of interaction data is needed, essentially as a Genbank equivalent for functional interactions. To this end, several database systems are under construction, as reviewed recently (*31*). The Biomolecular Interaction Network Database (BIND) is based on a flexible ASN.1 architecture designed in anticipation of the coming flood of proteomics data (http://www.bind.ca/). BIND is currently supported through joint government and industry programs and will be accessible as a fully open public database through an initiative called Blueprint (http://137.82.44.24/). Another advanced collection, the Database of Interacting Proteins (DIP) (http://dip.doe-mbi.ucla.edu/) now holds over 2600 protein interaction records parsed from the biomedical literature (*32*). A number of commercial databases with varying degrees of free access have also been constructed (*31*). In the long term, it will be critical that a uniform standard, or at least cross-platform compatibility, is implemented.

As mentioned above, most of the proteins involved in signal transduction and other forms of information transfer in the cell are constructed in a modular fashion from interaction domains and catalytic domains (*1*). It is possible to identify such interaction domains based on their primary sequence (*33*) and to assess their ligand-binding properties through the use of degenerate peptide libraries [for example, see (*34*)]. This approach allows the prediction of protein interactions based primarily on primary sequence information (*35*) [see Scansite (http://cansite.bidmc.harvard.edu/cantley85.html)]. Such predicted networks can then be compared with experimentally determined interactions using tools incorporated into BIND and other databases (*31*).

Given the complexity of biological systems, it is not surprising that most efforts at mathematical modeling of cellular behavior are fraught with faulty assumptions and amount to little more than a mathematical description of empirical observations. Obviously, sound data is the first requirement for meaningful modeling. This in itself is a huge hurdle, particularly since biological systems exhibit a range of behaviors, from straightforward Michaelian responses to highly complex nonlinear responses, including ultrasensitivity and phenotypic buffering, or the resistance to changes in phenotype arising from cellular perturbations (*36-37*). Many crucial decisions made by the cell may integrate signals that vary from state-to-state by twofold or less, yet at the same time such states may also be highly resistant to large changes in a single parameter. Despite the obvious impediments to successful modeling, some sound efforts have already been made to model the dynamics of carbon source utilization and cell cycle control in yeast (*40, 41*). It goes without saying that any attempt to model cellular behavior as a whole must incorporate all other available large scale datasets, including microarray-based expression profiles, models of protein structure from structural genomic analysis (*42*), genomewide mutational analysis (*43-45*), and systematic genetics (*46*).

The Immediate Future

Emergent technological breakthroughs in mass spectrometry will soon permit relevant protein interactions and posttranslational modifications in any given biological response to be comprehensively identified. In combination with large-scale mapping of protein-protein interactions, we can expect assembly of the dynamic networks that control cellular behavior to emerge over the next decade. In combination with systematic efforts to define gene-gene interactions and transcriptional profiles, a complete map of cellular wiring will emerge. The use of a standardized database for data archiving and modeling, by both academia and industry, will be crucial to maximize returns on large-scale experiments. Creative application of this knowledge base is expected to herald a new era in biology, the transition from a blind process of random archiving to engineering and systems analysis of the cellular machine.

References

1. T. Pawson, J. D. Scott, Signaling through scaffold, anchoring, and adaptor proteins. *Science* **278**, 2075-2080 (1997).
2. M. R. Wilkins, C. Pasquali, R. D. Appel, K. Ou, O. Golaz, J. C. Sanchez, J. X. Yan, A. A. Gooley, G. Hughes, I. Humphery-Smith, K. L. Williams, D. F. Hochstrasser, From proteins to proteomes: large scale protein identification by two- dimensional electrophoresis and amino acid analysis. *Biotechnology* (N.Y.) **14**, 61-65 (1996).
3. S. Fields, O. Song, A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-246 (1989).
4. T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, Y. Sakaki, A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4569-4574 (2001).
5. J. C. Rain, L. Selig, H. De Reuse, V. Battaglia, C. Reverdy, S. Simon, G. Lenzen, F. Petel, J. Wojcik, V. Schachter, Y. Chemama, A. Labigne, P. Legrain, The protein-protein interaction map of *Helicobacter pylori*. *Nature* **409**, 211-215 (2001).
6. B. Schwikowski, P. Uetz, S. Fields, A network of protein-protein interactions in yeast. *Nature Biotechnol.* **18,** 1257-1261 (2000).
7. P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnson, S. Fields, J. M. Rothberg, A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623-627 (2000).
8. P. Uetz, R. E. Hughes, Systematic and large-scale two-hybrid screens. *Curr. Opin. Microbiol.* **3**, 303-308 (2000).
9. M. Mann, R. Hendrickson, A. Pandey, Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **10**, 437-473 (2001).
10. D. Figeys, L. D. McBroom, M. F. Moran, Mass spectrometry for the study of protein-protein interactions. *Methods* **24**, 230-239 (2001).
11. A. Pandey, J. S. Andersen, M. Mann, Use of mass spectrometry to study signaling pathways. *Science*'s STKE (2000), http://stke.sciencemag.org/cgi/content/abstract/OC_sigtrans;2000/37/pl1
12. T. R. Covey, E. C. Huang, J. D. Henion, Structural characterization of protein tryptic peptides via liquid chromatography/mass spectrometry and collision-induced dissociation of their doubly charged molecular ions. *Anal. Chem.* **63**, 1193-200. (1991).
13. D. F. Hunt, J. R. Yates, 3rd., J. Shabanowitz, S. Winston, C. R. Hauer, Protein sequencing by tandem mass spectometry. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6233-6237 (1986).
14. H. R. Morris, T. Paxton, M. Panico, R. McDowell, A. Dell, A novel geometry mass spectrometer, the Q-TOF, for low-femtomole/attomole-range biopolymer sequencing. *J. Protein Chem.* **16**, 469-479 (1997).
15. A. V. Loboda, A. N. Krutchinsky, M. Bromirski, W. Ens, K. G. Standing, A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: design and performance. *Rapid Commun. Mass Spectrom.* **14**, 1047-1057 (2000).
16. A. Shevchenko, A. Loboda, A. Schevchenko, W. Ens, K. G. Standing, MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research. *Anal. Chem.* **72**, 2132-2141 (2000).
17. S. E. Martin, J. Shabanowitz, D. F. Hunt, J. A. Marto, Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **72**, 4266-4274. (2000).
18. A. G. Marshall, C. L. Hendrickson, G. S. Jackson, Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* **17**, 1-35 (1998).
19. J. E. P. Skya, G. C. J. Stafford, S. Horning, J. Shabanowitz, D. F. Hunt, J. A. Marto, A linear quadrupole ion trap fourier transform mass spectrometer: a new tool for proteomics. Abstract, 49th American Society for Mass Spectrometry conference, Chicago, IL (2001).
20. T. Formosa, J. Barry, B. M. Alberts, J. Greenblatt, Using protein affinity chromatography to probe structure of protein machines. *Methods Enzymol.* **208**, 24-45 (1991).
21. R. Verma, S. Chen, R. Feldman, D. Schieltz, J. Yates, J. Dohmen, R. J. Deshaies, Proteasomal proteomics: identification of nucleotide-sensitive pro-

teasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes. *Mol. Biol. Cell* **11**, 3425-3439 (2000).
22. H. Zhu, M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R. A. Dean, M. Gerstein, M. Snyder, Global analysis of protein activities using proteome chips. *Science* 26 July 2001 (10.1126/science.1062191)
23. A. Wagner, Robustness against mutations in genetic networks of yeast. *Nature Genet.* **24**, 355-361 (2000).
24. Y. Bellaiche, A. Radovic, D. F. Woods, C. D. Hough, M.-L. Parmentier, C. J. O'Kane, P. J. Bryant, F. Schweisguth, The partner of Inscuteable/Discs-Large complex is required to establish planar polarity during asymmetric cell division in *Drosophila*. *Cell* **106**, 355-366 (2001).
25. H. Steen, B. Kuster, M. Fernandez, A. Pandey, M. Mann, Detection of tyrosine phophorylated peptides by precursor ion scanning quadrupole TOF mass spectrometry in positive ion mode. *Anal. Chem.* **73**, 1440-1448 (2001).
26. Y. Oda, T. Nagasu, B. Chait, Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nature Biotechnol.* **19**, 379-382 (2001).
27. H. Zhou, J. Watts, R. Aebersold, A systematic approach to the analysis of protein phosphorylation. *Nature Biotechnol.* **19**, 375-378 (2001).
28. H. Zhu, M. Snyder, Protein arrays and microarrays. *Curr. Opin. Chem. Biol.* **5**, 40-45 (2001).
29. G. MacBeath, S. L. Schreiber, Printing proteins as microarrays for high-throughput function determination. *Science* **289**, 1760-1763 (2000).
30. H. Zhu, J. F. Klemic, S. Chang, P. Bertone, A. Casamayor, K. G. Klemic, D. Smith, M. Gerstein, M. A. Reed, M. Snyder, Analysis of yeast protein kinases using protein chips. *Nature Genet.* **26**, 283-289 (2000).
31. G. D. Bader, W. V. C. Hogue, in *Genomics and Bioinformatics*, C. W. Sensen, Ed. (Wiley-VCH, Weinheim, Germany, 2001), vol. 5b, pp. 399-413.
32. I. Xenarios, E. Fernandez, L. Salwinski, X. J. Duan, M. J. Thompson, E. M. Marcotte, D. Eisenberg, DIP: The database of interacting proteins: 2001 update. *Nucleic Acids Res.* **29**, 239-241 (2001).
33. J. Schultz, F. Milpetz, P. Bork, C. P. Ponting, SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5857-5864 (1998).
34. Z. Songyang, S. E. Shoelson, J. McGlade, P. Olivier, T. Pawson, X. R. Bustelo, M. Barbacid, H. Sabe, H. Hanafusa, T. Yi, Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. *Mol. Cell Biol.* **14**, 2777-2785 (1994).
35. M. B. Yaffe, G. C. Leparc, J. Lai, T. Obata, S. Volinia, L. C. Cantley, A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nature Biotechnol.* **19**, 348-353 (2001).
36 J. E. Ferrell, Jr., E. M. Machleder, The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes. *Science* **280**, 895-898 (1998).
37 J. E. Ferrell, Jr., Building a cellular switch: more lessons from a good egg. *Bioessays* **21**, 866-870 (1999).
38. D. C. LaPorte, D. E. Koshland, Jr., Phosphorylation of isocitrate dehydrogenase as a demonstration of enhanced sensitivity in covalent regulation. *Nature* **305**, 286-290 (1983).
39. J. L. Hartman, B. Garvik, L. Hartwell, Principles for the buffering of genetic variation. *Science* **291**, 1001-1004 (2001).
40. K. C. Chen, A. Csikasz-Nagy, B. Gyorffy, J. Val, B. Novak, J. J. Tyson, Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol. Biol. Cell* **11**, 369-391 (2000).
41. T. Ideker, V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, R. Aebersold, L. Hood, Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929-934 (2001).
42. D. Baker, A. Sali, Protein structure prediction and structural genomics. *Science* 293, 93-96 (2001).
43. E. A. Winzeler, D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J. D. Boeke, H. Bussey, A. M. Chu, C. Connelly, K. Davis, F. Dietrich, S. W. Dow, M. El Bakkoury, F. Foury, S. H. Friend, E. Gentalen, G. Giaever, J. H. Hegemann, T. Jones, M. Laub, H. Liao, N. Liebundguth, D. J. Lockhart, A. Lucau-Danila, M. Lussier, N. M'Rabet, P. Menard, M. Mittmann, C. Pai, C. Rebischung, J. L. Revuelta, L. Riles, C. J. Roberts, P. Ross-MacDonald, B. Scherens, M. Snyder, S. Sookhai-Mahadeo, R. K. Storms, S. Véronneau, M. Voet, G. Volckaert, T. R. Ward, R. Wysocki, G. S. Yen, K. Yu, K. Zimmermann, P. Philippsen, M. Johnston, R. W. Davis, Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901-906 (1999).
44. P. Gonczy, G. Echeverri, K. Oegema, A. Coulson, S. J. Jones, R. R. Copley, J. Duperon, J. Oegema, M. Brehm, E. Cassin, E. Hannak, M. Kirkham, S. Pichler, K. Flohrs, A. Goessen, S. Leidel, A. M. Alleaume, C. Martin, N. Ozlu, P. Bork, A. A. Hyman, Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**, 331-336 (2000).
45. A. G. Fraser, R. S. Kamath, P. Zipperlen, M. Martinez-Campos, M. Sohrmann, J. Ahringer, Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**, 325-330 (2000).
46. A. H. Y. Tong, M. Evangelista, A. B. Parsons, H. Xu, G. D. Bader, N. Page, M. Robinson, S. Raghibizadeh, C. W. V. Hogue, H. Bussey, B. Andrews, M. Tyers, C. Boone. Systematic genetic analysis with ordered arrays of yeast deletion mutants (submitted).