Towards a Modeling Infrastructure for Studying Plant Cells[1]

Thomas Girke, Mihri Ozkan, David Carter, and Natasha V. Raikhel*

Center for Plant Cell Biology (T.G., M.O., D.C., N.V.R.), Department of Botany and Plant Sciences (T.G., D.C., N.V.R.), and Electrical Engineering Department (M.O.), University of California, Riverside, California 92521

The use of high-throughput technologies in recent years has generated extensive information on the various levels of cellular and developmental processes in plants. The major challenge, however, remains the integration of this information toward a broad understanding of how the different biological layers interact to form higher functional units like coordinated pathways, regulatory networks, or complex structures like cells or tissues (Fig. 1). Systems plant cell biology is the attempt to achieve a mechanistic understanding of the functional components of plant cells and of entire plants including their development by predicting their properties from numerical data that arise from interaction analyses of many systems elements. It will allow scientists to study and understand cellular dynamics and organismal function, to create a detailed model of cell regulation, and to provide system level knowledge for the network of signal transduction cascades that are essential for plant development and physiological function. To reach this goal, we must adopt mathematical and computational methods for modeling and simulating complex biological systems. Until now, much of biology has been descriptive and exploratory rather than focused on creating a quantitative simulation model. There are as yet no computer programs that can accurately model biological processes. New standards are required for designing and analyzing experiments that will allow us to implement tolerance levels for noise in large-scale data sets. A strong commitment to quantitative analysis of biological phenomena would have the long-term goal of being able to model biological processes, though attempts are being made to decipher the basis of biological patterning (Wolfram, 2002). Modeling approaches would, in turn, lead to an increased emphasis on hypothesis-driven research in plant biology. This approach has the potential to transform our traditional reductionist view of biological functions to a broader understanding of interrelated components that form a complex system. Central issues of modern biology, such as phenotypic expression, robustness, and adaptiveness of plant cells, may finally derive causal explanations from this attempt.

The paradigm of describing and analyzing biological systems on all levels constitutes a new research concept that utilizes the golden opportunities of modern genomics and proteomics techniques rather than representing a technology by itself (Hood, 1998; Kitano, 2002b). This complex enterprise is highly interdisciplinary in nature and demands tight interactions between biologists, mathematicians, computer scientists, engineers, and other specialists. Fundamental questions need to be addressed in these collaborations, for instance: (a) how modular and generic is the architecture of cell systems, (b) how similar are the cellular networks formed by evolution compared with their rationally designed analogs in engineering, (c) to what extent are network modules the true building blocks of evolution, and (d) how completely can we describe cell properties by combining all their modular activities? To address these questions, scientists from various disciplines need to initiate virtual cellome projects for mimicking the essential features of plant cells in silico (computational) in a similar fashion as physiome undertakings in medical sciences (Rudy, 2000; Noble, 2002). An even more challenging enterprise will be the engineering of completely artificial cells. These cellular machines could be advanced into extremely powerful model systems in the future which will open new opportunities to simulate and analyze cellular networks. With the establishment of these technologies, we will be able to obtain more subtle answers to the questions that plant biologists have been asking for decades: What are the relationships between plant cell structure and function? Is it possible to learn simple rules that operate in subcellular systems? What determines cell differentiation, and how does pattern formation occur? What regulates the size and number of the different cell types in a tissue or organ? What do we have to know to model simple cell types? How could simple cell models be used to simulate more complex cell structures and tissue assemblies? Can we identify common patterns that govern biological complexity based on complex network architecture of different cell types?

The answers to these questions require the availability of genome-wide expression patterns of all plant cells and tissues, access to global protein profiles in cells and tissues, novel methods for visualizing protein activities and their localizations in living

[1] This work was supported by the National Science Foundation (plant genome grant no. DBI–0211797 and Arabidopsis 2010 grant no. DBI–0210992 to N.V.R. as co-principal investigator).

* Corresponding author; e-mail natasha.raikhel@ucr.edu; fax 909–787–4437.

www.plantphysiol.org/cgi/doi/10.1104/pp.103.022103.

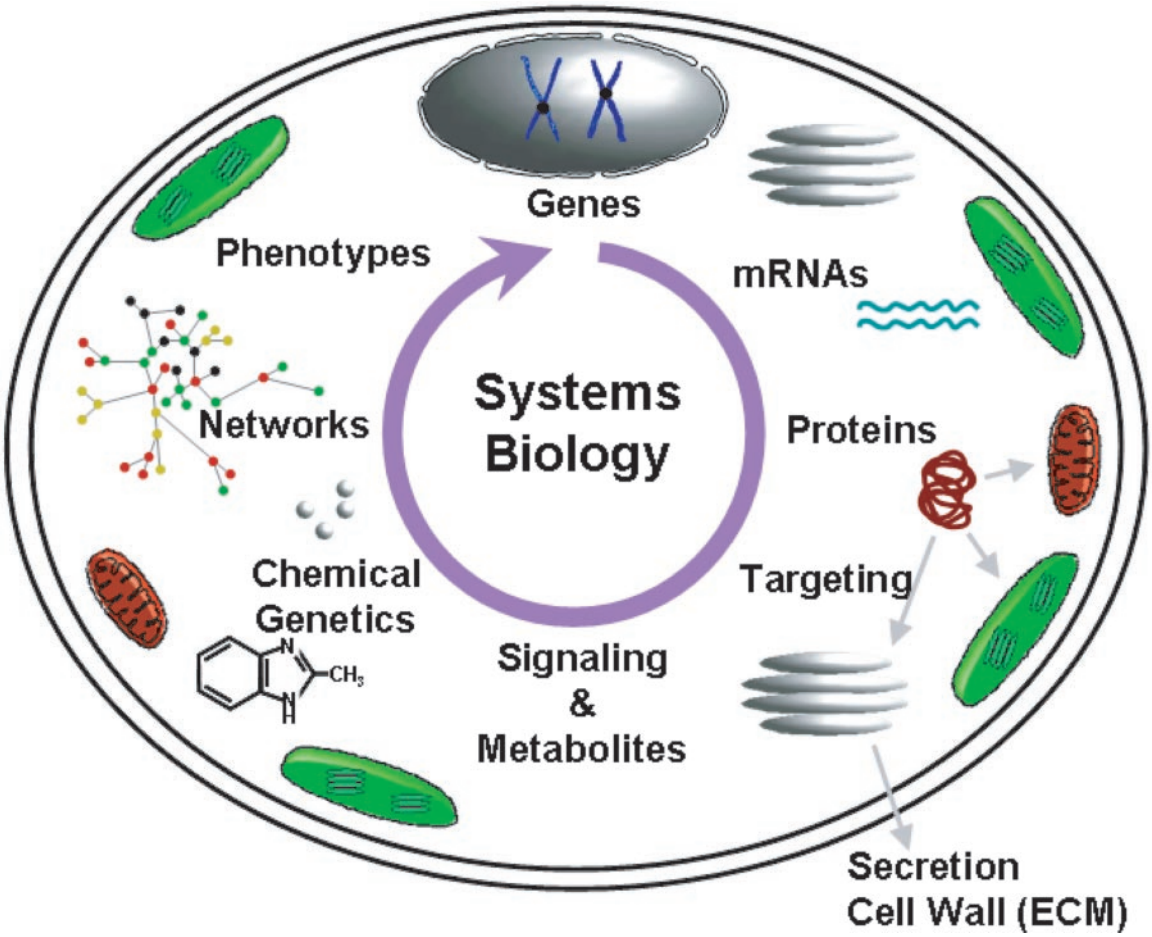

Figure 1. Holistic view of a simplified plant cell.

cells and organisms on a genome-wide scale (e.g. various florescence probes, fluorescence resonance energy transfer, and fluorescence redistribution after photobleaching; Chamberlain and Hahn, 2000; Houtsmuller and Vermeulen, 2001), approaches for finding complexes of macromolecules and methods that illuminate how plants use information, and the ability to integrate these "omics" profiles (see below). A system level understanding of the above-mentioned processes, for example, in Arabidopsis, will require a new combination of technologies and much closer interactions between system-minded plant biologists, computational biologists, engineers, chemists, and mathematicians. In silico biology will become a prominent feature in plant research. In addition, technical innovations in experimental devices that allow high-throughput and accurate measurements must be developed. This enterprise requires close integration of experiment, theory, and computation. Below, we present selected ideas that, in our view, could be of benefit to systems plant cell biology.

BIOINFORMATICS BRINGS IT ALL TOGETHER

The analysis of cellular systems requires extensive use of bioinformatics resources for data management, mining, modeling, and many other tasks. Because the bioinformatics requirements in plant cell biology are very similar to those in other areas, the following considerations are also relevant for many other fields in biology. The general systems analysis process can be divided into four stages of systems understanding following a classification from Kitano (2002a) that distinguishes between the identification of system structure, the analysis of its behavior, and the development of new system control and design strategies. It is an iterative process of analyzing and modulating cellular properties to continuously improve our models. An important question in this endeavor remains whether the current bioinformatics infrastructure is sufficiently prepared for the new requirement of systems-based approaches in plant biology because the data sets in this field are of unprecedented complexity and diversity. An additional dimension of complexity will be added by the incoming data from new technologies, like genome-wide studies of protein-protein interactions, subcellular protein localizations, single nucleotide polymorphisms, gene knockout effects, compound libraries, and associated phenotype data of chemical genetics screens. To manage these multidimensional

data sets, it will be necessary to develop a new generation of integrated databases to allow complex queries across diverse data types combined with new algorithms and flexible software pipelines for finding and simulating network architectures. Applications for analyzing single data types are even at the present time completely insufficient because bench and bioinformatics scientists have already started to identify correlations between sequence, RNA profiling, proteomics, and cellular imaging data (Ideker et al., 2002). Equally important are database interfaces for batch queries and scriptable access methods for large-scale comparisons to keep up with increased throughput requirements. Even for skilled bioinformaticians, organizing large data sets from different on-line sources is currently not an easy task largely due to the lack of standards for common data formats and limited interoperability of public databases (Stein, 2002). This issue is not limited to the plant community because it is common throughout the bioinformatics infrastructure across kingdoms and, in fact, in every application dealing with heterogeneous databases. Fortunately, there are several outstanding on-line resources available for the model plant Arabidopsis that have achieved a very high degree of data interconnectivity and multifunctionality: The Arabidopsis Information Resource (Garcia-Hernandez et al., 2002), The Institute for Genomic Research (http://www.tigr.org), and the Munich Information Center for Protein Sequences Arabidopsis Database (Schoof et al., 2002). Yet, there are few resources for comparing different plant species or plants with members from other kingdoms. They are of growing importance because comparisons of orthologous systems provide insight into important functional and evolutionary aspects of molecular network functions and organizations of cellular systems (Hartwell et al., 1999). For example, homology modeling, gene network comparisons, and discovery of regulatory DNA elements through phylogenetic footprinting are approaches that largely depend on the availability of extensive information from multiple species (Blanchette and Tompa, 2002). To facilitate these cross-species analyses, it will be crucial to initiate a plant meta-server project in the near future that mirrors the available information on all studied plant species and supports future plant systeome projects. In addition to organism-wide databases, it will be extremely important to incorporate the detailed information from more specialized sources like databases for gene families (e.g. PlantsP, http://plantsp.sdsc.edu/) or proteomics data (van Wijk, 2001). To reduce the isolated status of many data sources, numerous modifications need to be implemented in the current data infrastructure: consolidation of databases into larger warehouses, enhanced interoperability between databases, and increased support for open-source bio projects and web services by data providers (Stein, 2002). Considering the productivity and creativity of public database providers, strict standardization measurements appear to be less desirable than more pluralistic federation strategies. In addition, all these approaches will greatly benefit from the establishment of data annotation standards that provide more uniformity and the capability of organizing complicated data for the upcoming requirements of systems data exchanges (Hucka et al., 2002).

The ultimate goal of bioinformatics is not the management of systems data, but to utilize the data for the development of mathematical models to describe and predict the structure and behavior of plant cells and tissues. To formulate those models, there is a remarkable demand for new algorithms and software tools that allow more flexible and meaningful methods of analyzing and visualizing multifaceted systems data. First, cluster and network analysis tools, specifically designed for multidimensional data sets, will assist scientists in mining different data types simultaneously (Ideker et al., 2002). Second, new statistical methods are important to evaluate the significance of identified system structure and behavior patterns. Third, more modular software environments will allow scientists to bundle individual computational units into their own analysis pipelines in a user-friendly fashion. Fourth, new applications to simulate cellular systems will be extremely valuable for gaining additional insight into the architecture and performance of cellular networks (e.g. Virtual Biological Laboratory; Stelling et al., 2001).

Computational approaches require comprehensive and accurate data. In accordance, combined efforts of bioinformatics and experimental research appear to be the more promising strategy for biological discovery-oriented research, particularly at a time when data sets from various technologies are still incomplete and sometimes of improvable accuracy.

NEW TECHNOLOGIES FILL KNOWLEDGE GAPS

The need for highly accurate and comprehensive data at the cellular level demands a radical change in the way we design experiments toward more technology- and model-driven strategies. The introduction of nanotechnology, microfluidics, and highly integrated laboratory-on-a-chip systems will revolutionize our ability to collect, perturb, and measure cellular systems by providing new dimensions of automation, precision, and sensitivity down to single molecules. By using laboratory-on-a-chip systems, the time spent for a bioassay can be reduced, reagent cost can be minimized, and multidimensional assays on RNA, protein, and metabolite dynamics can be performed in parallel (Ozkan et al., 2001). This opens the question of how small should we go? The answer is entirely dependent upon the application and the size of the relevant biological elements. For example, devices for performing specific manipulations on cells

can be scaled down to cell scale dimensions, whereas devices based on molecular function (e.g. DNA analysis) can be made considerably smaller. Larger formats such as multiwell plates are more versatile, fit readily available equipment, and are amenable to automation, making them ideal for bridging the gap between sample-by-sample approaches and miniaturized high-throughput instrumentation.

The trend toward miniaturization came mainly from the human genome project and the related search for new therapeutic targets. With the number of genes in the human genome around 30,000 and the number of exclusive chemical targets about 100 million, up to 10^{12} assays would be required to completely map the structure activity space for all potential therapeutic targets. With the 3,456-well format in Aurora's Ultra-High-Throughput Screening System technology (Numann and Negulescu, 2001), the limit of "high-throughput" screening operates at the rates of 100,000 experiments per assay. This way, the time and cost required for each individual experiment is minimized appreciably. Many of these advances are already being adopted by mammalian cell researchers, but application to plant cell biology is in its infancy. This delay may be because of: (a) lower industrial support for the smaller market for green biotechnology products, and (b) the fact that many techniques have not been adapted to the structural characteristics of plant cells. The first concern can be addressed by developing lower cost high-throughput technologies, whereas the second issue requires an intelligent partnership between plant cell biologists and engineers.

Besides high-throughput screening, detection and imaging remain key elements for further developments in plant cell biology. With modern microscopy techniques, many physiological processes can be visualized directly or collected digitally and presented in an easily understood format. Fluorescent markers such as green fluorescent protein and fluorescence in situ hybridization can be used to locate proteins and DNA sequences, whereas fluorescence redistribution after photobleaching can measure motility of tagged molecules within and between cells (Chamberlain and Hahn, 2000; Houtsmuller and Vermeulen, 2001). In addition, fluorescent physiological markers can quantitatively indicate ion concentrations to test the reflexes of cells under different conditions. Whole cells and individual organelles can be gripped and moved by a laser beam (optical trapping) or selectively destroyed by a high-energy pulse of light (laser ablation; van den Berg et al., 1995).

All of these techniques reveal instant details of how cell components function within the living system. Combining them with high-throughput methodologies that we have discussed earlier, many parallel experiments can be performed to correlate hundreds of factors in different mutants and under different environmental conditions. For example, the new Atto Pathway HT microscope (www.atto.com), which is compatible with slide, dish, and multiwell plate samples, is well suited for plant material. It performs automated imaging by moving the optics below a stationary sample; therefore, whole plants and dishes of seedlings can be addressed and nonadherent cell suspensions can be viewed without agitation. For lower throughput, both the Pathway HT and Meridian InSight (Brakenhoff and Visscher, 1992) have real-time ocular viewing confocal visualization capabilities that are invaluable for rapid navigation around specimen space and subcellular screening.

Commercial imaging systems generally lack the beam intensity for routine use of nonimaging interactive applications such as photobleaching, photoactivation, trapping, and ablation, but most can be coupled to a Photonic Systems MicroPoint Nitrogen pumped dye laser (www.photonic-instruments.com), which delivers nanosecond pulses of any color from UV to deep red. For collecting large images at any optical resolution, an automated microscope can be used to seamlessly stitch together multiple extended-focus fields of view. And, for greater depth penetration, one can combine optical and physical sectioning for digitizing three-dimensional volumes of large blocks of tissue or even whole seedlings (Carter, 1994).

With a diverse selection of imaging methodologies available to us, the bottleneck of data collection is reduced, and access to the sample is improved. However, there remains the difficulty of preparing material for large experimental runs and of digesting the gigabytes of data generated by these high-performance imaging systems. New imaging systems are even more forgiving of sample type. For example, screening whole plants at subcellular resolution can be performed on a single instrument by adding high-power objective lenses to dissection microscopes or very low-power lenses to compound microscopes (M2 FL S, Zeiss, Jena, Germany; CFI Macro Plan UW 0.5× objective, Nikon, Tokyo). On-the-fly compression and analysis can be utilized to keep the data as condensed as possible, perhaps reducing the contents of each well to a few relevant statistics, then enabling the operator to drill down to study in detail those wells that produce the most interesting scores. This combination of hands-off screening and easy interactive review will allow the investigator to contemplate larger multiparameter experiments, to more readily recognize unexpected trends, and to manage a larger body of interrelated data.

Besides the current progress in imaging plant cells, technological developments in the genomics and proteomics domain are already being used for screening large knockout and transgenic populations to develop novel traits in economically important crop plants. These efforts have facilitated the generation of an unprecedented base of genetic and phenotypic diversity. These developments occurred along with the sequencing of complete genomes and the rapid development of multiparallel (high-throughput) technologies capa-

ble of sorting, accessing, and detecting the properties of biological systems (Somerville and Somerville, 1999). Most prominent among these technologies has been the establishment of mass hybridization protocols for the determination of the expression levels of many thousands of genes. In parallel, the science of proteomics—globally cataloging cellular protein content and state—is advancing at a phenomenal pace, and there are many excellent examples of its application in plant science (van Wijk, 2001). At the cellular level, cellular and tissue arrays are already in use for nonclinical mammalian cell research. In general, the use of cultured cells in vitro for nonclinical research raises questions about possible changes in cell behavior and function relative to the changes in its native environment. However, due to convenience and easy access, in vitro cell-based studies remain the first choice. In plant cell biology research, cell cultures that maintain tissue identity need to be developed for this purpose.

Similarly, technical innovations in experimental devices are essential for further advancing systems biology research. Raman spectroscopy, near-field scanning optical microscopy, and femto-second laser analysis are promising new approaches that permit direct visualization of molecular interactions (Lewis et al., 1999; Peleg et al., 1999), whereas surface plasmon resonance offers high-sensitivity kinetics analysis of native macromolecules in a microfluidic environment (Schuck, 1997; http://www.biacore.com).

For plant cell research to immediately benefit from these recent micro- and nanotechnologies, there needs to be close communication between scientists and engineers to seriously address complex biological problems on a systems level. Rather than waiting until the new technology is mature enough to adopt, fruitful collaborations between plant cell biologists and engineers are needed to ensure proper feedback so that newly developed instrumentation is amenable to work with plant material. This is one of the goals of the Center for Plant Cell Biology at the University of California (Riverside), where scientists with diverse backgrounds including biology, chemistry, computer science, and engineering adopt a team approach toward studying plant systems.

CONCLUSION

Systems biology will be the dominant concept and driver for future research in plant cell biology. If recognized and understood as a coordinated and multidisciplinary effort for developing the required infrastructure to model complex processes in plants, it will not just be the next buzzword of the postgenomics era.

ACKNOWLEDGMENTS

The authors thank Timothy Galitski for valuable scientific comments and Kathy Barton for editing the manuscript.

Received February 11, 2003; returned for revision February 12, 2003; accepted February 12, 2003.

LITERATURE CITED

Blanchette M, Tompa M (2002) Discovery of regulatory elements by a computational method for phylogenetic footprinting. Genome Res **12:** 739–748

Brakenhoff GJ, Visscher K (1992) Confocal imaging with bilateral scanning and array detectors. J Microsc **165:** 139–146

Carter D (1994) What do you get when you cross a confocal microscope with a microtome? Instant 3D reconstruction of really thick or opaque specimens. Microsc Soc Can Bull **22:** 9–12

Chamberlain C, Hahn KM (2000) Watching proteins in the wild: fluorescence methods to study protein dynamics in living cells. Traffic **1:** 755–762

Garcia-Hernandez M, Berardini TZ, Chen G, Crist D, Doyle A, Huala E, Knee E, Lambrecht M, Miller N, Mueller LA et al. (2002) TAIR: a resource for integrated Arabidopsis data. Funct Integr Genomics **2:** 239–253

Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. Nature **402:** C47–C52

Hood L (1998) Systems biology: new opportunities arising from genomics, proteomics and beyond. Exp Hematol **26:** 681

Houtsmuller AB, Vermeulen W (2001) Macromolecular dynamics in living cell nuclei revealed by fluorescence redistribution after photobleaching. Histochem Cell Biol **115:** 13–21

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle J, Kitano H (2002) The ERATO Systems Biology Workbench: enabling interaction and exchange between software tools for computational biology. Pac Symp Biocomput 450–461

Ideker T, Ozier O, Schwikowski B, Siegel AF (2002) Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics Suppl **18:** S233–S240

Kitano H (2002a) Looking beyond the details: a rise in system-oriented approaches in genetics and molecular biology. Curr Genet **41:** 1–10

Kitano H (2002b) Systems biology: a brief overview. Science **295:** 1662–1664

Lewis A, Radko A, Ben Ami N, Palanker D, Lieberman K (1999) Near-field scanning optical microscopy in cell biology. Trends Cell Biol **9:** 70–73

Noble D (2002) Modeling the heart–from genes to cells to the whole organ. Science **295:** 1678–1682

Numann R, Negulescu PA (2001) High-throughput screening strategies for cardiac ion channels. Trends Cardiovasc Med **11:** 54–59

Ozkan M, Ozkan CS, Kibar O, Wang MM, Bhatia S, Esener SC (2001) Heterogeneous integration through electrokinetic migration. IEEE Eng Med Biol Mag **20:** 144–151

Peleg G, Lewis A, Linial M, Loew LM (1999) Nonlinear optical measurement of membrane potential around single molecules at selected cellular sites. Proc Natl Acad Sci USA **96:** 6700–6704

Rudy Y (2000) From genome to physiome: integrative models of cardiac excitation. Ann Biomed Eng **28:** 945–950

Schoof H, Zaccaria P, Gundlach H, Lemcke K, Rudd S, Kolesov G, Arnold R, Mewes HW, Mayer KF (2002) MIPS *Arabidopsis thaliana* Database (MAtDB): an integrated biological knowledge resource based on the first complete plant genome. Nucleic Acids Res **30:** 91–93

Schuck P (1997) Use of surface plasmon resonance to probe the equilibrium and dynamic aspects of interactions between biological macromolecules. Annu Rev Biophys Biomol Struct **26:** 541–566

Somerville C, Somerville S (1999) Plant functional genomics. Science **285:** 380–383

Stein L (2002) Creating a bioinformatics nation. Nature **417:** 119–120

Stelling J, Kremling A, Ginkel M, Bettenbrock K, Gilles ED (2001) Towards a virtual biological laboratory. *In* H Kitano, ed, Foundations of Systems Biology. The MIT Press, Cambridge, MA, pp 189–212

van den Berg C, Willemsen V, Hage W, Weisbeek P, Scheres B (1995) Cell fate in the Arabidopsis root meristem determined by directional signalling. Nature **378:** 62–65

van Wijk KJ (2001) Challenges and prospects of plant proteomics. Plant Physiol **126:** 501–508

Wolfram S (2002) A New Kind of Science. Wolfram Media, Inc., Champaign, IL