Cell, Vol. 121, 507–509, May 20, 2005, Copyright ©2005 by Elsevier Inc. DOI 10.1016/j.cell.2005.05.001

Towards Cellular Systems in 4D

Commentary

Peer Bork* and Luis Serrano*
EMBL
Meyerhofstr.1
69117 Heidelberg
Germany

Systems Biology: Hope versus Hype

The approaches to and therefore the definitions of systems biology have been varied. These range from collections of physiological data with quantified molecular parts lists (e.g., genes, expression levels, localizations) to abstract mathematical modeling of biological processes. The scale at which the discipline of systems biology focuses on a specific question or problem is also a matter of contention: a tiny protein can be a complicated biological system (e.g., we still do not fully understand how a protein folds) and, as is obvious, an entire ecosystem with thousands of species constitutes another type of complex biological system. The term "systems biology" will probably get further diffuse as it is now under the limelight and new funding opportunities will be available to very diverse scientific communities.

Irrespective of whether there is a consensus on the definition of the term or whether it remains fuzzy, "systems biology" aims at a quantitative understanding of biological systems to an extent that one is able to predict systemic features. It is expected that this understanding will allow for rational design and permit us to modify the behavior of biological systems (i.e., synthetic biology). This applies to any system comprised of biological components that is more than the mere sum of its components or, in other words, the addition of the individual components results in systemic properties that could not be predicted by considering the components individually.

What do we hope to accomplish with modern systems biology approaches? Implicitly, there are a number of steps involved, each of which contributes to greater biological understanding. A typical process pipeline starts with standardized quantitative and qualitative data collection, archiving, and management. It is followed by proper integration of the data allowing comparative evaluation (this can reveal biases in data collection and indicate generalizable or specific features of the system studied). A next step would be the idealized reconstruction of the experimental situation as close to reality as possible. This is the preposition for studying the internal consistency of concepts and would allow a more global interpretation of the experimental data. An important component is the ability to generalize from the experimental setup to the system under study which includes testing of the consequences. Even more, we expect to extrapolate to concepts that are inaccessible to current experimentation and hope to arrive at novel concepts that are not deducible from the details.

*Correspondence: bork@embl.com (P.B.); serrano@embl.de (L.S.)

Within this framework, the cell is an attractive biological system where the time has come not only to produce standardized quantitative data but also to integrate them to model, predict, and design spatial and temporal features of cellular processes.

Feedback Loops: From Generating Data, Their Integration and Interpretation, to More Data

One of the key components required for the quantitative understanding of biological systems and for a new era of holistic modeling is the availability of sufficient data. The various 'omics communities generate increasingly reliable and partially complementary collections of cellular parts lists covered in a plethora of data ranging from genes, expression patterns, protein-protein interactions, to cellular localizations. However, even with all these data there is still only limited access to spatial and temporal aspects of cellular processes and systems (concentrations, reaction rates, scaffolding requirements, etc.). We will need to develop experiments and concepts to overcome these limits. The parts lists not only need to be organized in two dimensions (e.g., protein interaction networks), but there is a need to generate a structural framework for such networks (see, e.g., Aloy et al., 2004 for first attempts in this direction). Since such a framework is context and time dependent, more data have to be generated to achieve a resolution that is sufficient for modeling and simulation.

As bioinformatics is moving from maintaining and crossreferencing data collections toward their proper integration, we are seeing now a fusion of data and tools with the emerging modeling and simulation platforms (ranging from simple Boolean logic to spatial stochastic simulations). This comes with the hope that a low-resolution knowledge (that is yet to be generated) of spatial and temporal data would be sufficient to understand, modify, or design a biological system as complicated as a living cell. If precise data are required for every single interaction or reaction that takes place in a cell, then it is not likely that systems biology will achieve its objectives in the near future. As the level of detail required is not known yet, iterations with experiments have to be planned and patience is required as the process of preparing sufficient data might take a while.

Generation of a Spatial Framework Integrated with Temporal Data

Even the simplest cell is quite heterogeneous and the importance of spatial localization in biological processes is indicated by many experiments. Compartmentalization and diffusion are used by living systems during diverse processes, from signal transduction to cell division. Thus space and dynamics are essential to clearly understanding biological processes. Our Cellular Systems Biology initiative at the Structural and Computational Biology unit is embedded in various complementary activities within European Molecular Biology Laboratories (EMBL). The initiative aims to generate a structural framework of the cell and to use this unique data to incorporate temporal aspects using a strong computational biology component and modeling

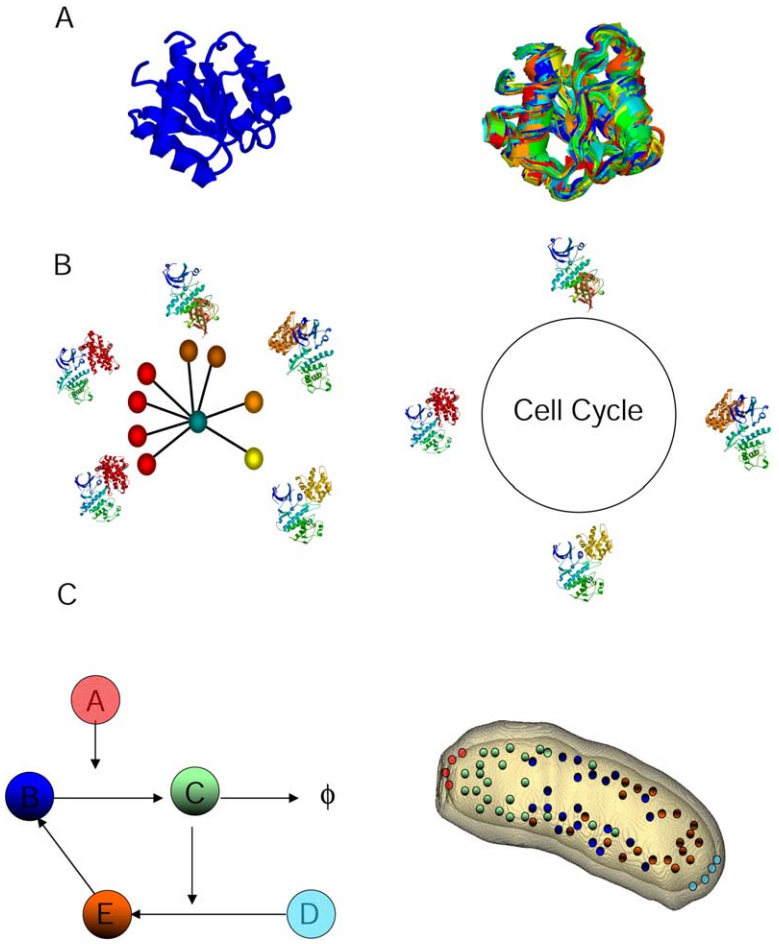

Figure 1. Spatial and Temporal Information from Proteins to Cells

(A) On the left a typical static structural representation of a biomolecule (in this case a protein) is shown. On the right a more realistic representation is shown in which an ensemble of conformations is recorded over time by NMR.

(B) On the left a typical representation for a network of interactions is given which is usually an average over several conditions or time steps. On the right this is considered as the interaction takes place only subsequently, at a given time point, for example, during cell cycle.

(C) On the left we show the typical graph representation for a protein network model without spatial constraints. On the right the same network is depicted in the context of a prokaryotic cell (EM tomogram), where there is an excluded volume effect in the center due to DNA.

techniques. While time is an integral part of most cellular models, incorporation of spatial information has only recently become feasible and, hence, the latter will be a crucial part of our efforts.

Starting with individual molecules or (partial) complexes determined by classical X-ray or NMR techniques and bridging them to subcellular structures obtained by lower-resolution techniques (e.g., single-particle EM) using computational biology tools, a large number of complexes and even some subcellular structures can now already be determined or modeled in an organism such as yeast (Figure 1). A large, EU-wide initiative is under way to collect three-dimensional structures of protein complexes and to further extend our first "bottom-up" attempts to model complexes around interaction networks (e.g., Aloy et al., 2004). In parallel, we have initiated "top-down" approaches to reconstruct high-resolution three-dimensional images of entire cells (Figure 2) using cryo-electron tomography (e.g., Kurner et al., 2005). These approaches are likely to soon cross a resolution barrier at which point the parts lists of complexes and subcellular structures can be readily identified in tomograms of a cell providing a new coordinate system onto which complexes and structures of decreasing size can be mapped. This spatial framework might serve as a bridge from individual molecules to a three-dimensional image of an entire cell (Figures 1 and 2). Despite its static nature, the framework will allow, for example, a more fine-tuned view of protein complex and interaction networks. In parallel to this initial phase, which will include a considerable amount of structural work for refinement, dynamic aspects are being considered at various scales. These range from comparing different states of a cell to individual molecular movements measured by NMR. In parallel to the measurements, data and tools are being customized to take on the spatial and temporal challenges at a cellular level (Figure 1). By integrating time series of gene expression to our quality-controlled networks of protein complexes and their interactions, the dynamics of complex formation can be revealed (e.g., de Lichtenberg et al., 2005). In order to model temporal network aspects in a spatial context, we are developing experimental (Isalan et al., 2005) and simulation techniques (Ander et al., 2005) that consider space and time and plan to integrate them with the spatial framework. These ambitious plans have to be seen in the context of various other systems biology activities spread across EMBL, ranging from simulations of spindle formation (see e.g., Surrey et al., 2001) to measuring and modeling cellular transport (e.g., Goerlich and Ellenberg, 2003) and signal transmission processes (Reynolds et al., 2003).

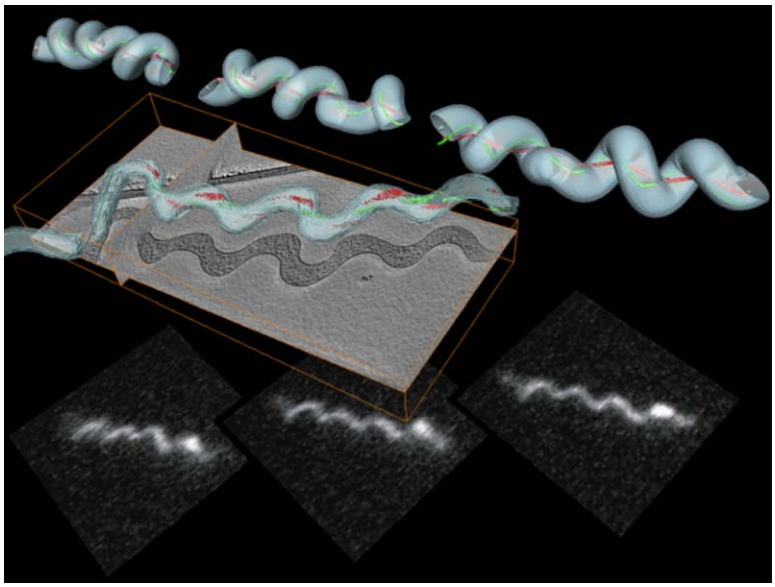

Figure 2. Modeling of the Conformational Changes and Underlying Cellular Motility of the Filament Bundles Attached at the Cell Wall of *Spiroplasma melliferum*

Cryo-electron tomography in conjunction with confocal microscopy reveals the motility modes of this organism at a molecular resolution. The center inset represents a three-dimensional rendering of a cryo-electron tomogram of a *Spiroplasma melliferum* cell combined with 2 nm thick slices visualized at different planes. At the bottom, three time-lapse confocal images visualize the cell movement at a resolution level of apprximately 200 nm. Above is the computational merge of data from single-particle electron microscopy, cryo-electron tomography, and confocal micrcopy, which resemble the observation from confocal microscopy very closely and explain cellular motility. Image provided by A. Frangakis, EMBL.

The goal is to bring all these activities together to explore spatial and temporal properties at the cellular level to the extent that we can predict and modify cell-phenotypic features.

Systems Biology, an Interdisciplinary Approach

There will be many elegant approaches toward systemic understanding, but it is likely that concerted efforts will have a higher chance to achieve an impact on science, and hence a number of diverse groups from our unit already contribute to the initiative. Together with the various interdisciplinary activities within EMBL and with collaborators world-wide, the analysis, prediction, and design of spatial and temporal features of entire cells based on molecular data will become feasible and have the potential of moving our understanding of living systems into four dimensions.

Acknowledgments

The authors are very grateful to Achilleas Frangakis from EMBL for providing the images used in Figure 2 as well as to Jens Reich and Iain Mattaj for critical reading of the manuscript and helpful discussions. L.S. is partly supported by EC grant LSHG-CT-2003-505520. P.B. is partially supported by EC grant LSHG-CT-2004-503568.

References

Aloy, P., Bottcher, B., Ceulemans, H., Leutwein, C., Mellwig, C., Fischer, S., Gavin, A.C., Bork, P., Superti-Furga, G., Serrano, L., and Russell, R.B. (2004). Structure-based assembly of protein complexes in yeast. Science *303*, 2026–2029.

Ander, M., Beltrao, P., Di Ventura, B., Ferkinghoff-Brog, J., Foglierini, M., and Kaplan, A. (2005). Iemerle, C., Tomas_Oliveira, I & Serrano, L. SmartCell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks. Syst. Biol. *1*, 129–138.

de Lichtenberg, U., Jensen, L.J., Brunak, S., and Bork, P. (2005). Dynamic complex formation during the yeast cell cycle. Science *307*, 724–727.

Goerlich, D., and Ellenberg, J. (2003). 4D imaging to assay complex dynamics in live specimens. Nat. Cell Biol. *307* (Suppl), S14–S19.

Kurner, J., Frangakis, A.S., and Baumeister, W. (2005). Cryo-electron tomography reveals the cytoskeletal structure of *Spiroplasma melliferum*. Science *307*, 436–438.

Isalan, M., LeMerle, C., and Serrano, L. (2005). Engineering gene networks to emulate Drosophila embryonic pattern formation. PLoS Biol. *3*, e64. 10.1371/journal.pbio.0030064

Reynolds, A.R., Tischer, C., Verveer, P.J., Rocks, O., and Bastiaens, P.I. (2003). EGFR activation coupled to inhibition of tyrosine phosphatases causes lateral signal propagation. Nat. Cell Biol. *5*, 447–453.

Surrey, T., Nedelec, F., Leibler, S., and Karsenti, E. (2001). Physical properties determining self-organization of motors and microtubules. Science *292*, 1167–1171.