6. Engineering approaches: what can we learn from it in Systems Biology?

Olaf Wolkenhauer
University of Rostock, Germany

The predefined title of this short chapter is misleading: Systems Biology *is* the merger of systems theory (engineering approaches) and molecular/cell biology. While bioinformatics has been closely associated with the field of computer science, Systems Biology is particularly attractive to researchers from the engineering and physical sciences. What this indicates is that an 'engineering' or 'systems-theoretic' approach is different from the way cell biological systems have been studied up to now. The emergence of Systems Biology is, in part, a consequence of the limitations we have reached in genomics and bioinformatics. While those areas are a different approach to investigating cellular processes, they are clearly complementary: an engineering approach relies on information about which macromolecules do matter in any particular cellular process and what their physical characteristics are. What this hints at is the fact that a signal and systems-oriented engineering approach is quite demanding in terms of the data it requires. As I shall argue below, the engineering approach is a necessity, not a choice, if we are to understand the functioning of the cell. The most important question we should therefore ask is 'What is necessary to ensure systems-theoretic approaches can work?'

While genomics and bioinformatics have focused on an effort to identify and catalogue the components that make up the cell, including their molecular characterisation and study of associations, the signal and systems-oriented perspective of Systems Biology focuses on functional activity, that is the dynamics of intra- and intercellular processes that determine cell function. An engineering approach is a 'way of thinking'. What this means and what we can or cannot learn from the engineering approach is discussed below. An engineering or systems-theoretic approach is characterised by the use of mathematical models. The important role mathematical models play is a consequence of the complexity of cellular processes, specifically the large number of variables, nonlinear interactions and temporal processes. Mathematical models are the extended arm of common sense; the only means we have to deal with non-intuitive complexity – no more but also no less.

In an article in *Current Biology* (Vol.15, No.21, 2005), Ronald Plasterk criticised the engineering approach and argued that:

'None of these modellers ever predicted that small microRNAs would play a role. One makes discoveries by watching, working, checking. They want to be Darwin, but do not want to waste years on the Beagle. They want sex but not love, icing but no cake. Scientific pornography'.

While a mathematical model (or more precisely the mismatch between a model and experimental data) can indicate whether additional variables or others than those selected, should be included in the model, knowledge of the components, and to some extent information about their molecular characteristics, must be available before we can establish a model of a dynamic system. A modeller could never predict that microRNAs would play a role; instead, the purpose of the model is to elucidate what role components have in the functioning of the cell. A mathematical model is used to characterise the function a component may have in the regulation and control of a processes, say gene expression. A model and computer simulation helps to validate hypotheses about the dynamic properties of a system and mechanisms (feedback interactions) that give rise to the behaviour observed in experiments. System biologists are interested in the consequences of dynamic interactions and perturbations, that is, how spatio-temporal changes in molecular concentrations determine cell function, including differentiation, apoptosis, proliferation etc. Plasterk apparently did not understand the role of models and modellers:

'One makes good models by watching gene expression, working on improved designs for experiments, checking hypotheses encoded by models. Modellers want quantitative data, but do not want to waste years in the lab (for which they are not trained). They want collaborations but not ignorance, support the experimentalist but not replace him. Interdisciplinary research'.

A cell, organ, or organism, understood as a 'system', is a network of components whose relationships and properties are largely determined by their function in the whole. The functionality is observed as the 'behaviour' of the system. The first and probably most important lesson of systems theory is that we can understand the behaviour of a system only if we systematically perturb it and record its response. A systems approach is thus characterised by input/output descriptions and from this, the most important role of the modeller in Systems Biology is to support the design of stimulus/response experiments. The role of nonlinear systems and control theory is then to provide methodologies to encode interactions of genes/proteins in the structure of the mathematical equations that form a model. The terms of these equations will reflect such processes as (de)activation, dimerisation, (de)phosphorylation, while the signs of these terms can indicate synthesis, degradation, positive or negative feedback relations. Parameter values emphasise terms and relate to the particular experimental set-up, cell type or cell line.

C. Systems Biology: scientific views about what, why and how

Taking cell differentiation as an example, the decision whether a cell differentiates or not will depend not only on the presence of a protein, whether a gene is 'on' or 'off', but on the history of various interacting proteins, a process referred to as a 'bistable system'. Even the simplest three-component model of such a process demonstrates that the observed behaviour can be understood only through experiments that vary not only initial conditions but also the duration and level (profile) of the stimulus. The system-theoretic concepts of 'identifiability', 'distinguishability' and 'observability' are important concepts in this context. The analysis of a model may reveal that there are multiple sets of parameter values that can all reproduce the same input-output behaviour. An improved design for the experiment may either remove this ambiguity or at least has the analysis alerting us to the uncertainty that can arise from such a situation. Closely related is the question of whether a given experiment would be capable of distinguishing between two hypothesised alternative mechanisms (model structures) that could generate the observed phenomena. What this discussion leads to is that experiments in Systems Biology tend to be more expensive and more time consuming. However, there is no alternative if we accept that in cells we are dealing with nonlinear dynamics. A consequence of this view is that research funding practices should appreciate the need for 'theoretical work', developing systems-theoretic methodologies, and that consumables budgets can increase if one generates quantitative time course data (including experiments to establish standards, normalise data and replicates to remove non-biological variability in measurements).

What the modeller describes as 'bistability', leads to switching-type behaviour; and an important task of Systems Biology is to identify functional units (subsystems) that realise such 'dynamic motifs', including for example 'oscillations', 'amplification', 'hysteresis' or

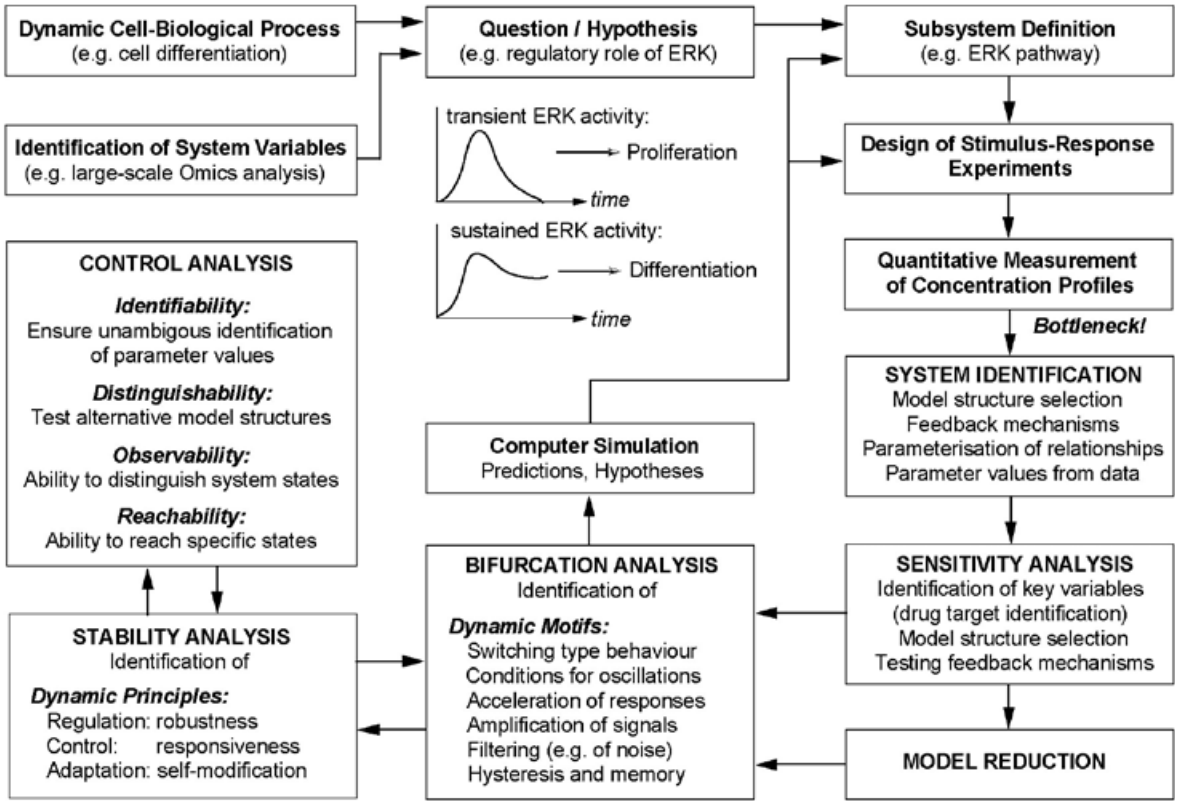

Figure 6.1: The (control) engineering approach as a Systems Biology workflow: merging cell biology with systems theory to study the functional organisation of cells, i.e., cell function understood as inter-and intracellular dynamic processes. The insert about the ERK signal transduction pathway is to provide an example in which the history of a signal (and not only the presence of a gene/protein) matters for the cellular process which decides upon the executed cell function. The role of an engineering approach is to elucidate the mechanisms (in particular feedback interactions) responsible for such observed phenomena: For nonlinear processes like the one above, these can be understood only with the help of mathematical modelling.

'homeostasis'. Stability and bifurcation analysis are important tools for this task. Under the heading of 'system identification', the control engineer gathers tasks for parameter estimation and model structure selection. Once a (preliminary) model is found, parameter sensitivity analysis allows us to identify the influence variables have on the overall behaviour of a network. This enables an investigation into the robustness and/or responsiveness of a system and is also a natural starting point to discuss potential drug targets. Many of the existing systems-theoretic techniques are not well suited for short time series, uncertainty in data and for systems involving many variables: there is a need for basic research to develop new methodologies. Systems Biology is not the application of existing engineering tools to cell biology but a merger of both fields; both fields should co-evolve.

The aim of Systems Biology is to understand the relations between things such as molecules or cells, not the things in themselves. Cell function arises from interactions between molecules and is not a property of any one molecule. The engineering perspective of Systems Biology is thus characterised by a shift towards an understanding of functional activity, away from the identification, molecular characterisation and cataloguing of the components that make up the cell. The complexity and limitations of Systems Biology are primarily a consequence of a large number of variables, interacting in space and time in a nonlinear fashion. Because of limited time frames for projects, funding constraints and also technological limitations that prevent us from quantifying large numbers of gene/proteins in time course experiments (at different levels of scale), a dynamic model of a pathway is necessarily 'wrong' – a phenomenological representation of a hypothesised principle that governs observed phenomena. Mathematical modelling is therefore the *art* of making appropriate assumptions; a process by which we represent one thing by another because understanding consists of reducing one type of reality to another. The purpose of modelling is therefore abstraction: the reduction of a complex reality to essential features. But even if inaccurate in this sense, a model can be useful by guiding the experimentalist in the design of his experiments, helping in the decision as to which variables to measure and how.

An important role of the modeller is therefore his/her involvement in the design of experiments. An advantage engineers and physicists have in this is that in addition to their analytical skills, they are not afraid of getting their hands dirty with experimental data. The sceptical wet-lab scientist may find that even if a mathematical model is a long way off, engineers and physicist can be helpful allies in understanding the physical properties (specifications) and limitations of measurement devices (e.g. its linear range, reproducibility, accuracy, etc.). Being able to quantify the accuracy and variability of instruments is an important step in interpreting experimental data. The real bottleneck for a success of engineering approaches in Systems Biology is advances in the generation of quantitative and sufficiently rich time series data sets. Progress in Systems Biology will depend on improved technologies that can quantify temporal changes in stimulus-response experiments. This can be done only in close collaboration with the engineering and physical sciences. What we can learn from engineering approaches is that measurement technologies to generate data and methodologies to analyse data cannot be separated.

7. The role of information technology for Systems Biology

Heikki Mannila
Helsinki Institute for Information Technology (HIIT), Helsinki University of Technology and University of Helsinki, Finland

Systems Biology aims at understanding biological entities at the systems level. To do this, we have to be able to observe many parts of the structure and dynamics of the entity, store and transform data, link it with many other types of observations, and model the data. Consider, for example, the task of modelling the behaviour of a single cell. We can obtain direct and indirect information about the genome of the cell, gene expression under experimental conditions, the metabolites and pathways in the cell, etc. Managing this data and using it to build a useful model of the cell will require huge advances also in information technology. Information technology is vital for Systems Biology: it is needed in measurement, in management and curation of the data, and in data analysis. Existing methods are not going to be enough, as Systems Biology poses unprecedented challenges to all these areas.

Measurement, storage and retrieval, and analysis

One key factor in the rise of Systems Biology is the rapid development of *measurement* technologies. We can measure many aspects of the operation of biological systems with high accuracy and in tremendous volume. The advances in high-throughput measurement techniques such as microarray methods have required many innovations from information technology.

Data management and curation are crucial for the accurate analysis of any larger mass of observational data. Especially in Systems Biology we have to understand well the conditions under which the data have been collected, otherwise the prediction of complex cellular functions cannot be achieved.