Horizons

Nature **454**, 424-426 (24 July 2008) | doi:10.1038/454424a; Published online 23 July 2008

Life, logic and information

Paul Nurse[1]

Top of page

Abstract

Focusing on information flow will help us to understand better how cells and organisms work.

Biology stands at an interesting juncture. The past decades have seen remarkable advances in our understanding of how living organisms work. These advances have been built mostly on molecular biology: applying the ideas that the gene is the fundamental unit of biological information and that chemistry provides effective mechanistic explanations of biological processes. These approaches, combined with an increasing ability to analyse highly complex biomolecular mixtures both qualitatively and quantitatively, have led to our present good understanding of cells and organisms and to significant improvements in our knowledge of human disease.

But comprehensive understanding of many higher-level biological phenomena remains elusive. Even at the level of the cell, phenomena such as general cellular homeostasis and the maintenance of cell integrity, the generation of spatial and temporal order, inter- and intracellular signalling, cell 'memory' and reproduction are not fully understood.

This is also true for the levels of organization seen in tissues, organs and organisms, which feature more complex phenomena such as embryonic development and operation of the immune and nervous systems. These gaps in our knowledge are accompanied by a sense of unease in the biomedical community that understanding of human disease and improvements in disease management are progressing too slowly.

One reason for this is that our past successes have led us to underestimate the complexity of living organisms. We need to focus more on how information is managed in living systems and how this brings about higher-level biological phenomena. There should be a concerted programme to investigate this, which will require both the development of the appropriate languages to describe information processing in biological systems and the generation of more effective methods to translate biochemical descriptions into the functioning of the logic circuits that underpin biological phenomena.

Living organisms are complex systems made up of many interacting components, the behaviour of which is often difficult to predict and so is prone to unexpected outcomes. Systems analyses of living organisms have used a variety of biochemical and genetic interaction traps with the emphasis on identifying the components and describing how these interact with each other. These approaches are essential but need to be supplemented by more investigation into how living systems gather, process, store and use information, as was emphasized at the birth of molecular biology.

Two iconic examples of this early thinking are the structure of DNA and the transcriptional regulation of the *lac* operon. The DNA double helix is beautiful not only because it is an elegant structure but because that structure reveals that DNA can act as a digital information storage device that can be precisely copied. Similarly, the mechanism of the *lac* operon (a set of nucleotides that regulates the metabolism of lactose) can be described in terms of molecular interactions between DNA, protein and metabolites. But these interactions make sense only when they are translated into a negative feedback loop that processes information about the level of lactose in the environment to regulate the rate of *lac* operon transcription.

This type of thinking needs to be embraced more comprehensively in all studies of living processes. We need to describe the molecular interactions and biochemical transformations that take place in living organisms, and then translate these descriptions into the logic circuits that reveal how information is managed. This analysis should not be confined to the flow of information from gene to protein, but should also be applied to all functions operating in cells and organisms, including chemical interactions and transformations as well as physical phenomena, such as electrical signalling and mechanical processes.

Information management

The study of cells is likely to be particularly effective for this programme because the cell is the simplest entity that shows complex biological phenomena. Furthermore, model cellular systems, such as bacteria and yeasts, developing eggs of worms and flies, frog-egg extracts and mammalian cells, provide a range of powerful complementary genetic, genomic and biochemical experimental approaches.

Given the conservation of many processes, the model eukaryotic systems have the added advantage of being relevant to human cells. The aim should be to analyse cells more effectively with the intention of then applying those approaches to more difficult organismal problems and to human disease. Two phases of work are required for such a programme: to describe and catalogue the logic circuits that manage information in cells, and to simplify analysis of cellular biochemistry so that it can be linked to the logic circuits.

For the first phase, the logic circuits that operate within cells need to be broken down into the individual segments that carry out specific computational functions. I shall call these segments 'logic modules'. One example of such a module is the negative feedback loop,

which often operates in a homeostatic manner. Another example is the positive feedback loop, which can generate irreversible switch behaviour from one state to another. Combinations of modules will produce more sophisticated outcomes: for example, reversible toggle switches, timers and oscillators.

The behaviour of the outputs from modules will be influenced by the shapes of the response curves embedded within them, with the outputs generated depending on whether, for example, the curves are linear, hyperbolic or sigmoidal. Modules could act as a short-term memory device, as seen in a G protein locked in a GTP-bound state, or as a long-term digital memory device as in the case of DNA. The identification of the logic modules used in cellular systems will allow a catalogue to be generated that defines the logic 'tool-kit' that is available to cells.

A useful analogy is an electronic circuit. Representations of such circuits use symbols to define the nature and function of the electronic components used. They also describe the logic relationships between the components, making it clear how information flows through the circuit. A similar conceptualization is required of the logic modules that make up the circuits that manage information in cells.

The initial identification of the logic modules operating in cells requires detailed biochemical descriptions of the interactions between different molecular components. Knowledge of the rate constants and strengths of interactions allows models to be built and differential equations to be generated and solved. If constraints exist as to what sorts of modules and linkages can generate effective and robust behaviours, then fewer possibilities will need to be considered. The tool-kit of modules and of the linkages between them that operate in cells may thus be limited, reducing the complexity of the problem that has to be solved.

“Studies at higher system levels are likely to inform those at the simpler level of the cell and vice versa.”

Knowledge of which modules are operational and how these are linked into circuits will help us to understand the flow of information. We need to know how information is gathered from various sources, from the environment, from other cells and from the short- and long-term memories in the cell; how that information is integrated and processed; and how it is then either used, rejected or stored for later use. The aim is to describe how information flows through the modules and brings about higher-level cellular phenomena, investigations that may well require the development of new methods and languages to describe the processes involved.

The next phase will be to simplify the analysis of the cellular biochemistry and link it with the logic modules. Key to this is determining which molecules interact with each other. This analysis is well under way with the application of various interaction-trapping approaches, such as two-hybrid methods, protein purification followed by mass spectrometry, and genetic screens for synthetic lethality. A further approach will be the systematic cataloguing of the position of fluorescently tagged proteins in living cells to

identify which proteins are near to each other and how that proximity may change over time. These spatial and temporal descriptions of molecules within living cells should simplify the analysis by defining a limited set of cellular spatial and temporal 'domains' that need to be considered. All these data will then need to be organized into databases, relating different cell types and model systems.

The next step is difficult, as it involves the mapping of molecular interactions and biochemical functions onto the logic modules, in effect linking the cellular chemistry tool-kit with the logic tool-kit. The success of this mapping will depend on whether there are sufficient regularities between specific logic modules and specific interacting molecules, at least at some level of probability.

Such regularities may not exist if natural selection has recruited many different components from the chemical tool-kit to generate specific examples of the logic tool-kit. However, there may be sufficient regularities to make this mapping possible. The fact that life on Earth generally uses nucleic acids as digital information-storage devices, gives some cause for optimism. Another example may be protein kinases and phosphatases that act antagonistically, which behave like switches.

As we learn more about how molecules interact to generate logic modules it may become less necessary to know the details of the rate constants and the molecular concentrations and to solve the differential equations that they generate. If detailed modelling reveals that certain molecules wired together in particular ways are often associated with specific modules, then it might become possible to predict some behaviours without having precise measurements of the variables involved. Simply knowing which molecular components are present and how they are linked together might be sufficient to speculate about which logic module is in operation. If this is the case, then the module can be considered as a black box and it may be necessary to concentrate only on *in vivo* measurements of key inputs into and outputs from the black box to confirm that the logic module is behaving in the expected manner.

Analysis in practice

How could such a programme work in practice? First the higher-level cellular phenomenon of interest has to be identified. Examples of such processes include chemotaxis, mating, signalling and aspects of cellular reproduction. One approach would then be to mutationally saturate the phenomenon by use of forward genetics and genome-wide deletion collections to identify as many of the genes involved as possible. Application of standard bioinformatic procedures would link the genes identified with specific biochemical and molecular functions. Identifying which molecules interact with each other, and how, can be established by use of the interaction trap, and by spatial and temporal cellular domain databases.

So far this approach is relatively conventional. The next steps will be to use the databases described above to determine the probability that specific components of the chemical tool-kit are associated with a particular logic module. Finally, the modules will be linked

together into a complete circuit, allowing outputs to be predicted so that the functioning of the circuit can be translated into a narrative of information flow to describe how the cellular phenomenon works.

What issues might we expect to encounter if this programme is adopted? One important consideration is that because the logic modules and circuits are combined into networks, an understanding of how such networks operate in cells will be crucial. Complex networks have been well analysed in other spheres of human activity. For example, transportation networks such as flight routes and connections are often found to have diverse numbers of linkages between hubs in the network such that some hubs become crucial because they are highly connected to many other hubs. Network analysts call such networks 'scale-free'. It seems that biological networks derived from genetic, protein–protein and transcriptional interaction studies are also often scale-free. So far, analysis has suggested that these hubs are likely to be ancient in origin and so arose early in evolution.

It is important to realize that unlike simpler networks such as those seen in transportation systems, linkages between hubs in cellular networks will not all be of a similar physical and logic type. Some will represent stable physical interactions and others will reflect more transient biochemical reactions. Furthermore, the logic consequences will vary, either negative or positive in action, for example. In the future it will be necessary to use representations that capture more effectively the different linkages connecting hubs in biological systems. Biological networks are also more flexible and fluid, and can reconnect and reassemble in different ways to generate alternative networks with changed outcomes. The language used to properly represent biological networks will need to accommodate these variations in logic structures.

“Living machines are not intelligently designed and will often be redundant and overly complex.”

Dynamic signals

Another interesting feature of logic circuits in biological systems is the roles that temporal organization or dynamics may have. Signalling pathways within or between cells have generally been thought of as linear sequences that lead to on/off switches. An analogy for such a sequence is a railway signal that results in only one of two outcomes, a stop or a go signal. If dynamics is introduced into signalling pathways, richer behaviours can emerge. For example, if signals are pulsed down a pathway and the changing outputs are monitored, much more complex information can be transmitted.

A metaphor here would be the use of the Morse code and the telegraph to communicate messages. Pulses of information sent along the telegraph generate a code for letters and as a consequence sentences can be communicated. This converts the same signalling pathway from a simple on/off switch to a device that can transfer, for example, the works of Shakespeare. It is likely that dynamics has been exploited more generally in the

evolution of biological systems for signalling purposes, allowing the communication of more complex information.

Spatial organization of signalling pathways within cells will also enrich behaviours, with different outcomes being possible in different regions of the cell depending on the spatial context of the input and output signals. Logic circuits can also give rise to behaviours that generate spatial organization, as in the case of Alan Turing's reaction-diffusion equations. Because cells are extended in space, the spatial organization generated by logic circuits will contribute to spatial order within the cell, for example by acting as position-locating mechanisms during the generation of cellular form.

Finally, we need to take account of the biological origins of the logic circuits and networks that operate in cells. Because natural selection operates on pre-existing living organisms, novelties will initially arise as add-ons to systems already in existence, almost guaranteeing some redundancy. Thus, man-made machines, which are generally intelligently designed, will differ from the logic machines found in life. Living machines are not intelligently designed and will often be redundant and overly complex.

We should anticipate these differences and be prepared for the additional complexity to be found in the logic circuits that manage information in cells. Lessons will also be learned from the higher levels of biological organization seen in communities of individuals, in ecological systems and during evolutionary change. The principles and rules that underpin how information is managed may share similarities at these different levels even though their elements are completely different. Studies at higher system levels are thus likely to inform those at the simpler level of the cell and vice versa.

I have suggested that cells and experimentally amenable model systems should form the major part of this programme at this point in time, but ultimately what we learn with these simpler biological systems needs to be applied to more complex multicellular organisms and to humans if we are to fully understand organismal biology and improve treatment of human disease. Part of the problem of shifting these approaches to organisms will be one of scale, of having to deal with more genes, more involved structures and more complex phenomena. It will also be necessary to take full account of ecological and environmental interactions as well as the evolutionary context of the organism under study. In addition, we will have to develop methodologies to properly investigate intact living organisms, including humans in both the healthy and the diseased state. Particularly important for this work will be the development of high-resolution sensitive imaging procedures to monitor biomolecules in real time and in space. This is the return to whole-organism and human physiology that many have argued is long overdue, but with a renewed emphasis on the logic of life and the management of information.

Programme requirements

What is required that is not already generally in place to pursue this programme effectively? Perhaps the most pressing need is to develop the appropriate theoretical

approaches to analyse the management of information flow and to investigate the logic systems that are responsible for that flow.

I see this best being developed not as a 'big science' project but by individual scientists working alone and together in small interactive workshop groups meeting on a regular basis. The groups will need to be multidisciplinary, including information theorists, mathematicians, physicists, chemists and computer scientists working closely with experimental biologists who have good biological intuition and who can communicate with members of the other disciplines. Different workshop groups could interact with each other through digital conversations to share ideas.

The training of advanced undergraduate and graduate biologists also needs to shift in its emphasis. The separation of molecular and cell biologists from those that study organism biology, ecology and evolution has weakened biomedical research, and the emphasis on learning large numbers of facts in molecular- and cell-biology courses and during medical training has reduced the necessary exposure to the ideas central to biology.

Time needs to be made during education to expose biomedical scientists to other scientific disciplines to ensure good communication between biologists and other disciplines so that theory is always well embedded in biological facts and experiments. Placing a greater emphasis on ideas during teaching and training will have the added advantage of attracting excellent students to the whole biological and biomedical research endeavour.

Success in the programme will require sophisticated databases that can manage different types of data from a range of experimental systems that can be used to generate connections and handle probabilities of outcomes. New experimental techniques are required to allow better *in vivo* analysis of living systems with sophisticated imaging for real-time experiments. The analyses will also need to develop beyond single-organism studies in closely defined unchanging laboratory conditions, and move towards more complex ecological circumstances working with societies of organisms in changing environments.