CytoComputational Systems – Perspectives and Tools of Thought

R. Paton[†]

Department of Computer Science, University of Liverpool, Chadwick Building, Peach Street, Liverpool L69 7ZF, United Kingdom

Cells are complex systems. For some people, they are like tiny chemical plants, or laboratories, or machines. For others they are more like computational devices or even computational systems. As data continue to be generated about them and their components and the systems that they make up, new perspectives and models are needed to deal with the complexity. Cells are more than bags of chemicals just as macromolecules are more than microscopic billiard balls or strings of beads. The challenges of an information processing view that complements the more commonly expressed chemical processing view needs to be taken into account.

The biomolecular, cellular and tissue levels of biological organisation have had considerable inspirational impact on the development of computational models and systems. Such innovations include neural computing, systolic arrays, genetic and immune algorithms, cellular automata, artificial tissues, molecular computing, and protein memories. With the rapid growth in biological knowledge there remains a vast source of ideas yet to be tapped. These include developments associated with biomolecular, genomic, enzymic, metabolic and signalling systems and the various impacts on distributed, adaptive, hybrid and emergent computation. Many biologists use language that is displaced from computer sciences not least, program (as in apoptosis), hardware–software, DNA – as data or program, code, gate, Boolean network, pattern recognition, and so forth. Indeed, many proteins (such as enzymes and transcription factors) carry out complex information processing tasks (as individual molecules) including pattern recognition, switching, logical decision-making, memory, and signal integration.

This book provides readers with a comprehensive exploration of this subject from a uniquely multidisciplinary point of view. Contributions from biologists, computer scientists, engineers and mathematicians are drawn together to provide a comprehensive picture of biological systems, both as sources for ideas about computation and as information processing systems. The varieties of perspectives that are presented provide an integrative view of a complex and evolving field of knowledge that needs new tools of thought to manage and mobilise the vast opportunities afforded to the postgenomic and biocomputational sciences.

1 Plan of the Book

The book begins with several chapters that represent a (bio)mimetic approach to engineering and computation. Tateson, a biologist by training who now works for BTexact Technologies, looks at the application of cellular analogies to telecommunication systems. He argues that in some cases there are sound reasons for believing that analogies will be helpful. We can identify biological systems, which resemble in some sense an artificial system, and then use our knowledge and understanding of the functioning of the biological system to improve or redesign the artificial system. In other cases there is little more than the assertion that 'nature knows best' and hence any artefact modelled on nature must be superior to an artefact for the same purpose devised by human reason alone. This is not a useful basis for redesigning our artificial systems, and in fact the analogy with nature is often 'bolted on' to the human–designed system to explain to non-engineers how it works rather than being genuinely useful at the design stage.

The chapter by Bull and Tomlinson shows how symbiogenetic mechanisms found at the cellular level can be successfully applied to computational learning. Symbiosis is the phenomenon in which organisms of different species live together in close association, potentially resulting in a raised level of fitness for one or more of the organisms. Symbiogenesis is the name given to the process by which symbiotic partners combine and unify – forming endosymbioses and then potentially transferring genetic material – giving rise to new morphologies and physiologies evolutionarily more advanced than their constituents. This process is known to occur at many levels, from intra-cellular to inter-organism. They use the abstract NKCS model of coevolution to examine endosymbiosis and its effect on the evolutionary performance of the entities involved. They suggest the conditions under which endosymbioses are more likely to occur and why; we find they emerge between organisms within a window of their respective "chaotic gas regimes" and hence that the association represents a more stable state for the partners. This general result is then exploited within a machine learning architecture to improve its performance in non-Markov problem domains.

Timmis, Knight, Castro and Hart discuss the growing field of Artificial Immune Systems (AIS) – that is using the natural immune system as a metaphor for solving computational problems. The field of AIS is relatively new and draws upon work done by theoretical immunologists such as Jerne, Perelson, and Varela. What is of interest to researchers developing AIS is not the modelling of the immune system, but extracting or gleaning useful mechanisms that can be used as a metaphor to help solve particular problems. It is quite common to see gross simplifications of the way the immune system works, but this is not a problem as it is inspiration computer scientists seek from nature rather than precise mechanisms. The review is organised in the following manner. First, reasons for why the immune system has generated such interest and is considered to be a good metaphor to employ are discussed. This is followed by a simple review of relevant immunology that creates many of the foundations for work reviewed in this contribution. Immunology is a vast topic and no effort has been made to cover the

whole area. Rather, only those ideas that have proved to be useful to the majority of research presented in this contribution are explained in some detail. Finally, a summary of the work presented in this contribution is provided, drawing main conclusions from the work presented and commenting on the perceived future of this emerging technology.

Tyrrell considers the analogy between multi-cellular organisms and multi-processor computers as not too far-fetched, and well worth investigating, particularly when considering that nature has achieved levels of complexity that far surpass any man-made computing system. The aspect of biological organisms on which this chapter is centred is their phenomenal robustness: in the trillions of cells that make up a human being, faults are rare, and in the majority of cases, successfully detected and repaired. The Embryonics project (for embryonic electronics) is inspired by the basic processes of molecular biology and by the embryonic development of living beings. By adopting certain features of cellular organisation, and by transposing them to the two-dimensional world of integrated circuits in silicon, it will be shown that properties unique to the living world, such as self-replication and self-repair, can also be applied to artificial objects (integrated circuits). Self-repair allows partial reconstruction in case of a minor fault, while self-replication allows complete reconstruction of the original device in cases where a major fault occurs. These two properties are particularly desirable for complex artificial systems in situations that require improved reliability. To increase still further the potential reliability of these systems, inspiration has also been taken from biological immune systems – Immunotronics. The acquired immune system in humans (and most vertebrates) has a mechanism for error detection which is simple, effective and adaptable.

The chapter by McNeil and Snowdon serves to provide a further conceptual bridge between hardware and biology, this time by working with molecules at the nanoscale. The dawn of nanoscale science can be traced to a now classic talk that Richard Feynman gave on December 29th, 1959 to the annual meeting of the American Physical Society at the California Institute of Technology. In this lecture, Feynman suggested that there exists no fundamental reason to prevent the controlled manipulation of matter at the scale of individual atoms and molecules. Twenty one years later, Eigler and co-workers constructed the first man-made object atom-by-atom with the aid of a scanning tunnelling microscope. Given that there is "Plenty of room at the bottom" (the title of Feynman's talk), and biological systems have highly subtle and sophisticated meso- and micro-scale architectures, the exploitation of this level in medical and computational technologies will continue to challenge 21st century biocomputational science.

The kind of approach discussed in the previous chapters has an established record since the 1940s and the developments in cybernetics, digital electronics, and general models of computation. Thus we find the developments in digital models of neurones (the McCulloch–Pitts model), computational models of brains, and the origins of cellular automata. Many of the ideas that were spawned in the 1940s and 1950s, and many tools of thought for helping scientists and engineers to organise their knowledge of biological systems can be traced back to this time (as

can the revolution that took place in molecular biology). The viewpoint now moves along towards ways in which the languages of physical science and mathematics can enhance our appreciation of biological systems.

A common observation made by scientists who are working at multi-disciplinary interfaces (such as CytoComputational Systems) relates to the problems encountered by differences in vocabulary, emphasis, modelling approach, and attitudes to reduction and simplification. The next chapter looks at some ways displacements of ideas between the disciplines can take place. Nagl, Parish, Paton, and Warner consider ways of describing computational topics in molecular and cellular biology. Methods of classifying DNA, RNA and proteins are central to current methods for elucidating relationships between sequence, structure and function. The chapter looks at metaphors for the function of proteins and points to a unified view of proteins as computational devices capable of matching patterns as inputs and processing to result in alternative outputs. The requirement for a systems view of life is also pursued. As such this chapter provides an immediate bridge between the previous few and next few chapters. In addition it also anticipates a number of themes emerging in later chapters (such as Fisher *et al.* and Wolkenhauer *et al.*).

Following on from this computational stance, Bolouri and Schilstra provide a short review of the modelling of Genetic Regulatory Networks (GRNs). GRNs have a basic requirement to model (at least) some parts of a biological system using some kind of logical formalism. They represent the set of all interactions among genes and their products for determining the temporal and spatial patterns of expression of a set of genes. The origins of modelling the regulation of gene expression go back to the Nobel Prize winning work of Lwoff, Jacob and Monod on the mechanisms underlying the behaviour of bacterial viruses that switch between so-called lytic and lysogenic states. The authors briefly discuss some of the circuit-based approaches to GRNs such as the work of Kauffman, Thomas, and Shapiro and Adams.

The next two chapters address computational modelling of cells using very different approaches. The chapter by Gregory, Paton, Saunders and Wu reports on work concerned with Individual-Based Models (IBMs) of a 'virtual' bacterial cell. The goal of this project has been to explore the ecological and evolutionary trajectories of ‘artificial bacteria’. This is a great challenge both to understanding the cell in sufficient detail and to implementing a system on current computational architectures. Each bacterium is an independent agent with sufficient genomic and proteomic equivalents built into the model such that each individual cell can have up to 120 genes and 50,000 gene products. The chapter reports on the development of a model that has to incorporate multiple scales in both time and space.

Feng's chapter develops some mathematical models of stochastic computations in neurones and neural networks. It is a bridge between the cell-based work discussed previously, the tissue-based work we look at next, and the bioinspired approaches that started the book. Feng discusses the developments of his neuronal decision theoretic approach in relation to the role played by inhibitory inputs. The

cellular components execute a learning rule and the networks that are produced can be applied to statistical pattern recognition problems.

The next two chapters provide very interesting insights into the workings of mathematical modellers as their work addresses very specific biological problems. Monk's chapter looks at his work dealing with spatial patterning in explicitly cellular environments. Pattern formation in multicellular organisms generally occurs within populations of cells that are in close contact. It is thus natural and important to consider models of pattern formation that are constructed using a spatially discrete cellular structure. Here, the particular case of pattern formation in cellular systems that depends on contact-dependent (juxtacrine) signalling between cells is discussed.

At another scale of biological organisation, MacGregor, Leng and Brown describe a model of the hypothalamic and pituitary components involved in controlling growth hormone release. Their model has been developed by gathering and attempting to formalise the experimental data about the system but has been kept as simple as possible, focusing on the functional rather than mechanical properties of its components. They show that a relatively simple model can be capable of producing complex behaviour and accurately reproducing the behaviour and output of a real brain system.

We now address a collection of modelling approaches that build on a computational perspective (that is, the underlying metaphor is focused on computation or information processing rather than the dynamics expressed in models based on continuous mathematics). Holcombe notes how computational models have been of interest in biology for many years and have represented a particular approach to trying to understand biological processes and phenomena from a systems point of view. One of the most natural and accessible computational models is the state machine. These come in a variety of types and possess a variety of properties. This chapter discusses some useful ones and looks at how machines involving simpler machines can be used to build plausible models of dynamic, reactive and developing biological systems that exhibit hierarchical structures and behaviours.

Another computational stance that has been related to coding/decoding issues concerns the sequence structures and scrambling of genes. Sant and Amos examine some recent work on models of recombination in *Ciliates*. They describe how these single-celled organisms ‘compute’ by unscrambling their genetic material. They present a number of models of this process, and discuss their implications. This could have useful implications for the development of cellular computers.

In contrast to the previous chapter, which looks at genes and DNA, the chapter by Fisher, Malcolm and Paton looks at proteins and, specifically, the modelling of signalling proteins as algebraic agents. Harking back to the previous chapter by Nagl *et al.,* the authors begin by looking at proteins as computational agents. Protein information processing networks are discussed, notably secondary messenger signaling, signalling kinases and phosphatases, scaffold proteins and

protein–protein interactions. The final section of the paper develops an algebraic model of protein interactions based on rewrite rules.

The final part of the previous chapter points to an area of mathematics called Category Theory that has variously been applied to problems in theoretical biology since the 1960s. However, it is inaccessible to many non-mathematicians and yet is very useful at helping to integrate knowledge. The short piece by Brown, Paton and Porter seeks to give non-specialists, and especially biologists, an accessible introduction to the subject especially in relation to hierarchical structures.

Wolkenhauer and Kolch present an approach that can be used to investigate genome expression and regulation through mathematical systems theory. The principal idea is to treat gene expression and regulatory mechanisms of the cell cycle, morphological development, cell differentiation and environmental responses as controlled dynamic systems. Although it is common knowledge that cellular systems are dynamic and regulated processes, to date they have not been investigated and represented as such. The kinds of experimental techniques that have been available in molecular biology largely determined the material reductionism, which describes gene expression by means of molecular characterisation. Instead of trying to identify genes as causal agents for some function, role or change in phenotype they relate these observations to sequences of events.

The final chapter by Johnson can be viewed as one summary approach to some of the issues the CytoCom Network had to address, namely the kinds of natural processes that can be regarded as computations. In recent years the idea of using computational concepts as a way of understanding biological systems has become of increasing importance; this conceptual use of computational ideas should be contrasted with the equally valuable activity of using computers as tools for interpreting biological data and simulating biological systems. He suggests that this computational attitude towards biological systems has been valuable in computer science itself, too; by observing how biological systems solve problems, new algorithms for problem solving on computers can be developed.

2 History of the CytoComputational Systems Project

The CytoCom project was one of a number of networks of scientists funded by the UK Engineering and Physical Sciences Research Council looking at Emerging Computing Paradigms. In our case we were using non-neural cells and tissues as the area of focused study and discussion. We held five workshops between 1999 and 2001, in Liverpool, Leeds, Hertford, Sheffield and London. This book, together with a number of other publications, further networks and funded research, were some of the achievements. A less quantifiable achievement was the general increase in awareness of this level of biological organisation as a source of computational ideas. CytoCom grew from an existing, though loose community of scientists interested in and attending an international series of workshops called IPCAT (Information Processing in Cells and Tissues). The first IPCAT was held

in Liverpool in 1995 and since then we have had workshops every other year in Sheffield (1997), Indianapolis (1999) and Leuven (2001). The fifth workshop was held in Lausanne (2003) and the sixth is planned for York in 2005.

Acknowledgement. CytoCom was made possible through support from the UK Engineering and Physical Sciences Research Council (EPSRC) and we are especially grateful to Jim Fleming and Mark Hylton for all their help and encouragement. We are also very grateful to Helen Forster for all her help and administrative support.