Abstract
The developments in biochemistry and molecular biology over the past 30 years have produced an impressive parts list of cellular components. It has become increasingly clear that we need to understand how components come together to form systems. One area where this approach has been growing is cell signalling research. Here, instead of focusing on individual or small groups of signalling proteins, researchers are now using a more holistic perspective. This approach attempts to view how many components are working together in concert to process information and to orchestrate cellular phenotypic changes. Additionally, the advancements in experimental techniques to measure and visualize many cellular components at once gradually grow in diversity and accuracy. The multivariate data, produced by experiments, introduce new and exciting challenges for computational biologists, who develop models of cellular systems made up of interacting cellular components. The integration of high-throughput experimental results and information from legacy literature is expected to produce computational models that would rapidly enhance our understanding of the detail workings of mammalian cells.

Keywords: cell signalling; systems biology; network analysis; biochemical networks; graph theory
often the phenotype of cells. The integration of high-throughput experimental results and information from legacy literature is expected to produce computational models that would rapidly enhance our understanding of the detailed workings of mammalian cells (Figure 1).

Investigating many cellular components simultaneously, instead of few one at a time, has catalysed the formation of a new approach in biological research called systems biology. Systems biology has many definitions, but it surely involves the integration of experimental and computational approaches, and the systems view: the attempt to explain how the whole is made from its parts [2, 3]. This definition fits the current state of research in signal transduction in mammalian cells. Linear signal transduction pathways were first identified as a sequence of cascading events where extracellular ligands bind to receptors, evoking conformational changes in the receptors; this interaction then leads to the activation of molecules inside the cell. A series of activation events in the intracellular space results in propagating signals to cellular functional machines to alter cellular phenotypic behaviour. With time, it becomes more and more evident that cellular signalling pathways are connected to one another to form networks [4].

We are at a stage where there is a growing need for further understanding of the organizational and functional properties of combined pathways [5, 6]. Viewing protein–protein interactions as graphs (networks) where molecules are represented as vertices (nodes) and interactions are represented as edges (links) is useful for investigating cellular signalling networks [7–10]. Scientists have used the ‘balls and arrows’ diagrams for several decades, but the diagrams were never large enough for network/graph theory analysis. The information to construct a meso- or even large-scale ‘balls and arrows’ cellular connections map is contained within the legacy literature in biochemistry, molecular biology and cell physiology. We developed one such map by identifying interactions manually from primary research articles into a simple template. We ignored kinetics parameters and spatial–temporal relationships and focused on information flow to capture many of the known pathways in mammalian hippocampal CA1 neurons [11]. Data extraction of interactions from literature does pose challenges. One obstacle, for example, is the non-standardized nomenclature used to describe genes and proteins. We developed the CA1 neuron network model manually and the dataset underlying this network is highly reviewed by curators. There are alternative approaches to manually extract interactions from literature. Different automated and semi-automated approaches to build regulatory cellular networks from literature range from shallow parsers to complex natural language processing (NLP) implementations [12]. The major problem with those systems is accuracy in preserving biological information and constraints.

Regulatory networks, described as a system of nodes connected with directed links, are analysed using network analysis methods. The connectivity distribution scaling factor [13] and the clustering coefficients measurements [14] can be computed. Additionally, programs that identify the abundance of network motifs [15], small recurring circuits made of a few nodes, can be used to find and count such motifs. These measurements can be applied to subnetworks that are defined based on different biologically relevant criteria [11]. For example, we can generate and analyse subnetworks that include pathways starting at a specific receptor and converging into specific transcription factors. The comparison between many sub-networks created in a series based on similar criteria helped us identify some global properties of the architectural topology in the hippocampal neuronal regulatory cellular network [11]. Since, this process involves the measurement of

Figure 1: Integration of the high-throughput experimental results with knowledge gained in the past four decades is expected to breed computational models that would rapidly enhance our understanding of the detailed workings inside mammalian cells.
motifs and other features as a signal (i.e. functional connectivity) propagates through the system, we have termed this type of analysis pseudodynamics [6].

The topology of biochemical regulatory networks is often compared with artificially generated networks. These efforts attempt to capture the evolutionary processes that shape regulatory networks to further advance our understanding about the network's functional design. The duplication–divergence model [16] is an attractive network-growth algorithm that is relevant in understanding regulatory cellular networks structure. Artificial networks can also be simulated as Boolean networks to assess their dynamical stability. Gene regulatory networks have long been viewed as Boolean networks [17]. A resurgence of interest in Boolean networks modelling is now occurring [18]. Simplification of complex chemical networks with variable rates and concentrations to Boolean networks appears to be a useful coarse-grain representation of biochemical regulatory systems. Thus, some of the dynamics of cellular signalling regulatory networks may be captured using Boolean networks modelling. However, it is important to keep in mind that simplifying cellular regulatory networks to Boolean networks is only a first step. Prediction capabilities of Boolean networks are qualitative. To understand fully the dynamics of regulatory cellular networks quantitative modelling approaches are required [19]. These range from simplest ordinary differential equations representations of mass-action and enzyme kinetics [20] to stochastic [21] and hybrid modelling platforms [22]. Most complex modelling platforms consider spatial boundaries in 2D [23] or 3D [24] with both deterministic and stochastic reactions.

Computational modelling of networks from literature without considering the results from high-throughput experiments is limited. The integration of the two types of datasets (networks from literature and networks from high-throughput experiments) is likely to produce models that would have the best predictive capabilities. The high dimensionality of such dynamic systems raises challenges for computational biologists. The current quantitative modelling methods can only handle a handful of variables. New formalisms would have to be developed. Going back to the concepts of information transfer, new modelling approaches would directly deal with information content embedded in bio-molecular interactions. For example, one could propose that concepts such as Shannon entropy [25] and information gain [26] would be computed instead of flux balance in mass-action kinetics [27] for signalling networks. Such approaches have been already successfully implemented for the analysis of signalling networks in mammalian cells [28].

In conclusion, the process of discovering how molecules work together to perform the functions of a cell, is analogous to solving a jigsaw puzzle. The pieces of this puzzle were created from a 3D VRML image. The image was automatically created from interactions extracted from literature that describe the signalling network from norepinephrine to reach intracellular components in five steps of coupled interactions.

Figure 2: The process of discovering how bio-molecules work together to perform the functions of a cell, is analogous to solving a jigsaw puzzle. The pieces of this puzzle were created from a 3D VRML image. The image was automatically created from interactions extracted from literature that describe the signalling network from norepinephrine to reach intracellular components in five steps of coupled interactions.
The analogy to a conventional jigsaw puzzle ends at this step, a cell is dynamic, puzzle pieces move, there are many pieces of the same type, and many pieces of different types, the pieces are so tiny that we cannot see them, even with the best microscopes, and the puzzle is in 4D instead of 2D. Despite all of these complexities, by integrating networks from literature with high-throughput experimental results, and developing computational models from such data, we may be able to solve most of the 4D jigsaw puzzle that is the functional cell.

Key Points
- Integration of knowledge accumulated in the past 30–40 years about cellular components and their interactions could be combined with high-throughput emerging technologies that allow the measurements of binary interactions between many cellular components simultaneously. These data can be used to build computational models of mammalian cells with predictive power.
- We have developed a meso-scale mammalian neuronal cellular regulatory network from biomedical research literature to represent many of the known signalling pathways and their effectors in CA1 neurons; using graph theory methods we have uncovered some of the global topological properties of this network.
- Analysis of subnetworks (within the meso-scale network) formed by expanding connectivity in steps from source nodes such as extracellular ligands is useful in identifying network motifs (circuits made of few interacting nodes), that are capable of information processing, and the juxtaposition of these motifs within the structure of the network.
- Understanding the dynamics of mammalian cell signalling networks is analogous to solving a jigsaw puzzle except that the puzzle is in four dimensions and the pieces are very small.

Acknowledgements
We thank Sherry L. Jenkins for her assistance in developing Figure 2. A.M. is supported by the NIH training grant GM-62754. Research in the Iyengar Laboratory is supported by grants GM-072853 and GM-54508 from NIH.

References

