

The virtual cell—a candidate co-ordinator for ‘middle-out’ modelling of biological systems

Dawn C. Walker and Jennifer Southgate

Submitted: 8th December 2008; Received (in revised form): 3rd February 2009

Abstract

Understanding the functioning of biological systems depends on tackling complexity spanning spatial scales from genome to organ to whole organism. The basic unit of life, the cell, acts to co-ordinate information received across these scales and processes the myriad of signals to produce an integrated cellular response. Cells interact with and respond to other cells through direct or indirect contact, resulting in emergent structure and function of tissues and organs. Systems biology has traditionally used either a ‘top-down’ or ‘bottom-up’ approach. However, neither approach takes account of heterogeneity or ‘noise’, which is an inherent feature of cellular behaviour and may have significant impact on system level behaviour. We review existing approaches to modelling that use cellular automata or agent-based methodologies, where individual cells are represented as equivalent virtual entities governed by simple rules. These paradigms allow a direct one-to-one mapping between real and virtual cells that can be exploited in terms of acquiring parameters from experimental systems, or for model validation. Such models are inherently extensible and can be integrated with other modelling modalities (e.g. partial or ordinary differential equations) to model multi-scale phenomena. Alternatively, hierarchical agent models may be used to explore the functions of biological systems across temporal and spatial scales. This review examines individual-based models and the application of the paradigm to explore multi-scale phenomena in biology. In so doing, it demonstrates how cellular-based models have begun to play an important role in the development of ‘middle-out’ models, but with considerable potential for future development.

Keywords: *cellular automata; cell biology; computational modelling; emergent behaviour; middle-out modelling; software agent; systems biology*

INTRODUCTION

‘I believe very strongly that the fundamental unit, the correct level of abstraction, is the cell and not the genome’

Sydney Brenner, Lecture at Columbia University, 2003 (as quoted in [1])

In contrast to physics, engineering or chemistry, biology is a data-driven science. In recent years, the development of high-throughput technologies for genomics, metabolomics, transcriptomics and proteomics has generated phenomenal amounts of data, such that in 2005, the rate of data generation was estimated to be of the order of 1 terabyte (10^{12} bytes)

Corresponding author. Prof. J. Southgate, Jack Birch Unit of Molecular Carcinogenesis, Department of Biology, University of York, York YO10 5YW, UK. Tel: +44-190-432-8705; Fax: +44-190-432-8704; E-mail: js35@york.ac.uk

Dawn Walker obtained a PhD in 2001 and is an RCUK Academic Fellow in the Department of Computer Science at the University of Sheffield. Her primary research interest is using computational models to understand the behaviour of complex biological systems. She is based in the Kroto Institute, part of the largest multidisciplinary research complex in the UK, comprising approximately 200 researchers from 20 disciplines.

Jennifer Southgate holds the Chair of Molecular Carcinogenesis and is Director of the Jack Birch Unit of Molecular Carcinogenesis in the Department of Biology at the University of York. Her research has involved developing human epithelial cell culture systems for functional tissue engineering and for investigating physiological and disease processes in the human urinary bladder. From this background, she has developed an interest in computational modelling as an interpretive and predictive tool of biological systems.

per day for proteomics alone, with a 5–10-fold increase expected year on year [2]. The field of bioinformatics has developed alongside the four 'omics' approaches as a means to capture multi-scale data inter-relationships and to facilitate the storage, access and analysis of this plethora of information. Predictive computational modelling is a field that has also grown in the past two decades. Driven by the need to make sense of the huge volumes of omics data and facilitated by the exponential growth in computational power, computational modelling holds the promise of revolutionizing biology as an interpretative and predictive science.

As discussed by Southern and colleagues [3], human biology bridges hierarchical scales of organization ranging from the gene, to proteins, individual biological cells, tissues, organs, and finally the organism, which directly interacts with its external environment. Associated with this hierarchical organization is a spectrum of spatial and temporal scales, the former ranging from $\sim 10^{-9}$ m at the gene/protein level to $\sim 10^1$ m for an individual (human) organism, and the latter ranging from milliseconds ($\sim 10^3$ s) for molecular interactions to 80 years ($\sim 10^9$ s) for the average human life expectancy.

The two traditional approaches adopted to interpret this complexity have been categorized as 'top-down' (driven by the observation of biological characteristics and attempting to construct theories that would explain the observed behaviour) and 'bottom-up' or reductionist (by studying the components of the system in isolation, then attempting to integrate the behaviour of each component in order to predict the behaviour of the entire system) [4].

Recently, there has been a growing interest in a 'middle out' approach, whereby the initial focus is on an intermediate scale that is gradually expanded to include both smaller and larger spatial scales. In his book 'The Music of Life—Biology Beyond the Genome' [1], Denis Noble attributes the term 'middle-out' to the renowned molecular biologist Sydney Brenner. The starting point for a 'middle-out' approach to modelling biological systems may be influenced by a number of factors, including the ready availability of relevant experimental data and the span of the biological scales of immediate interest to researchers working in a particular field. For instance, Cristofolino and colleagues proposed a middle-out approach to simulating bone remodeling under stress that started at the organ level [5].

The boundary loads were calculated from a whole-body perspective and dynamic biological structure was represented by constitutive equations, which also incorporated parameters relating to cellular turnover. Integration of the tissue and cellular levels of this multi-scale approach remain under development.

Taking a holistic view of the nature of biological systems, we support the argument that a natural starting point for the 'middle out' approach is the biological cell, which represents the basic unit of life. Cells are capable of integrating the 'hard-coded' information encompassed within their genetic and epigenetic makeup, with long- and short-range cell-generated and environmental cues, to regulate the profile of genes and proteins expressed and to modulate cellular phenotype and response. The summative effect of large numbers of individual cells communicating directly (cell–cell contact) and indirectly (cell–matrix contact or exogenous ligand–receptor interactions) with one another and their extracellular environment is the assembly of a homeostatic community or tissue. Disruption of tissue homeostasis resulting from dysregulation of cells within the community may be manifested on a macro-level as disease, with malignancy being the ultimate example.

The wealth of accessible experimental data available at the cellular level also makes it the natural starting point for developing 'middle out' predictive (simulation) models of biological phenomena [6]. The nature of the system behaviour under investigation in such a model will determine the level of cellular detail included. Where the phenomenon of interest occurs on a time scale that is significantly faster than cell turnover, it is not necessary to include representations of proliferation or apoptosis. For instance, Southern and colleagues describe multi-scale modelling approaches in cardiac physiology, with individual cells abstracted to collections of ion channels [3]; there is no concept of growth or proliferation in this model. Furthermore, a cellular-level approach does not necessarily imply that every individual cell is represented explicitly. In cases where the processes under consideration occur on longer time scales, and in which proliferation, apoptosis or migration may be of consequence but intercellular heterogeneity can be neglected, continuum approaches may be adopted. Here, the time-dependent changes in population density resulting from proliferation, apoptosis (cell death) and migration are represented by mathematical rate

equations (e.g. [7, 8]). Continuum models have been widely applied to subcellular components, where, providing that populations are large and well mixed, behaviour can be described by simple rate equations—for example, modifications to protein activity as a result of binding, dissociation, phosphorylation, synthesis and degradation. By contrast, cells can assume a variety of phenotypes and engage in asynchronous behaviour depending on their current genetic/proteomic state and the threshold of local extracellular cues. Whereas many protein-based simulations ignore spatial interactions, the relative positioning and interactions between cells is fundamental to developing emergent tissue architecture. Individual cell-based simulations thus have required a new type of modelling paradigm that is capable of including these characteristics.

CELLULAR AUTOMATA AND SOFTWARE AGENT APPROACHES

Cellular automata (CA)-based modelling has been applied to a range of systems, including molecular, bacterial, cellular and ecological (for reviews relating to biological systems see [9, 10] and for physical systems [11]). In this case, the ‘cellular’ does not necessarily refer specifically to biological cells, but to spatial elements in a 2D- or 3D-lattice. These elements can either represent the presence of a single entity, or a collection of similar entities. The state of each cellular automaton (e.g. whether it contains a live or dead biological cell or other entity of interest, or remains empty) is updated iteratively based on simple logical rules, which depend on its previous state, that of its neighbours, or some other local environmental variable. CA methods are useful for representing cellular behaviour in cases where cell shape and size can be ignored. By contrast, agent-based or ‘off lattice’ cellular models have more generic applications, as cells are not constrained to particular points in space and cell shape can be explicitly included as a model parameter. Unlike more traditional continuum-based modelling methods, both CA and software agent paradigms are highly suited to modelling emergent behaviour, where complex behaviour at the level of the biological system (e.g. tumour growth rate or wound healing) is generated as an outcome of direct and indirect interactions between large numbers of individual cells. For a recent review of

on- and off-lattice approaches, and a comparison with continuum models, see [12].

Agent-based modelling and CA modelling are terms that are often used interchangeably. There has been extensive debate about the precise definition of a software agent. Wooldridge and Jennings defined the four critical properties of an agent as ‘autonomy, social ability, responsiveness and proactiveness’ [13]. In the context of cellular biology, software agents are usually programmed to fulfill the first three of these characteristics, but as cells are not goal-driven, they are not usually considered to exhibit proactive behaviour. The main distinguishing features between CA and agent-based models are (i) CA models are lattice-based, whereas agent-based models may or may not be restricted in space (ii) generally agents exhibit more complex memory states and sets of behaviour and (iii) in agent models, the emphasis is on the explicit representation of discrete biological entities, whereas CA models focus on the state of elements or automata, which may contain more than one entity at any given time.

Over the last decade, both CA and software agent paradigms have been applied increasingly in the context of cell biology. A recent review discussed how the discrete nature of the agent model lends itself to validation by direct comparison with experimental systems—a process that is crucial in order to ensure that the adopted rules are representative of real biological behaviour [14]. Unfortunately, there is frequently a failure to validate computational models against the experimental counterpart, probably reflecting the hurdles of interdisciplinary collaboration and a lack of quantitative tools for comparing real and simulated systems.

A third example of a cellular based paradigm is the Cellular Potts models. Like CA models, this methodology is also lattice-based; with each lattice point defined either as inside or outside a biological cell. The locations of cell surfaces are iteratively determined by the minimization of a mathematical function representing the ‘effective energy’ of the cell, which comprises energy associated with surface interactions, cytoplasmic fluctuations and response to chemotactic stimuli [15]. This energy-based definition of cellular location and shape contrasts to the dependency of CA and software agents on simple logical rules that reflect directly observable behaviours (e.g. proliferation, apoptosis) governed by underlying biological mechanisms (e.g. subcellular

signalling). For this reason and due to space constraints we do not discuss Cellular-Potts models further in this review, but refer the interested reader to [16] for a recent discussion of how this type of model can be incorporated into a multi-scale simulation environment.

In order to represent multiple biological scales (e.g. cellular growth and migration that occur on a time scale of hours, or protein interactions involved in signal transduction that occur in seconds to minutes) a common approach is to implement computational models that are solved on multiple temporal scales, or contain hierarchical spatial scales (e.g. multiple subcellular compartments). For the remainder of this review, we will use the word *multi-scale* to refer to a model that includes more than one spatial or temporal component, whilst explicitly defining the biological scales, or components (cell, extracellular protein) encapsulated by the model.

The focus of this review is multi-scale CA and software agent models—i.e. those models that have begun to fulfill the criteria of being a 'middle out' exploration of biological systems and that have centred their approach on the concept of the individual cell as the mediator of higher order (tissue) structure and homeostasis. We have excluded multi-scale models that bridge the molecular to supra-cellular scale, but focus on physical, rather than biological behaviour (e.g. the electrophysiological/mechanical cardiac models of Noble and Hunter [17]). We have restricted our review to models that represent mammalian tissues (although multi-scale approaches have been employed in a number of non-mammalian systems including bacteria [18], dictyostelium [19], xenopus morphogenesis [20] and meristem development in plants [21]).

APPROACHES TO EXTENDING CELLULAR MODELS

Several multi-scale models that explicitly encapsulated cellular behaviour have appeared in the literature in recent years, but prior to this there was a general evolution in modelling approach. Single-scale cellular automata models were initially adopted for modelling cell behaviour, focussing on representing cell proliferation, migration and apoptosis [22]. Lattice-free or agent-based models were also developed, where the ability of cells to migrate or proliferate freely necessitated an explicit consideration of the concept of mechanical

interactions between individual agents. This problem was approached by both Drasdo *et al.* [23] and Walker *et al.* [24] by integrating agent models with numerical-based representations of intercellular forces. Although these examples considered only a single biological scale—the individual cell—as discussed above, the implementation of these models is, in fact multi-scale, as the temporal scales of the agent (biological) and mechanical model components are separated (specifically, a slow time scale represents cell behaviour, whilst a fast time scale is used to update cell location in response to intercellular forces, representing a process which is continuous in the real world).

The majority of cellular level-based models have evolved to incorporate phenomena occurring on a subcellular length scale (signalling pathways or gene networks), or extracellular processes involving the secretion or diffusion of proteins, synthesis or modification of extracellular matrix (ECM) or intercellular signalling, whereas relatively few models have been extended 'upwards' to include higher biological hierarchies, for instance, to simulate the formation of tissues or organs.

Biological multi-scale approaches may be classified under three main categories:

- (i) *Cellular-continuum approaches*. In this case, biological subscales (e.g. ECM proteins) are represented in the model as a field of values representing concentrations that are considered to be in steady-state. A single time scale is incorporated in order to represent cellular behaviour.
- (ii) *Spatially hierarchical approaches*. Subcellular components are explicitly represented in the model as a lower hierarchy of agent, but without separation of biological time scales.
- (iii) *Temporally separated approaches*. A second model modality (e.g. mathematical equations) is integrated to represent processes that occur on a faster time scale.

These three approaches are discussed in more detail below, and shown schematically in Figure 1.

Cellular-continuum approaches

The simplest approach to multi-scale modelling is to include the representation of a biochemical factor or protein that can influence the behaviour of the computational agents. Examples include the presence of a varying field that represents an exogenous

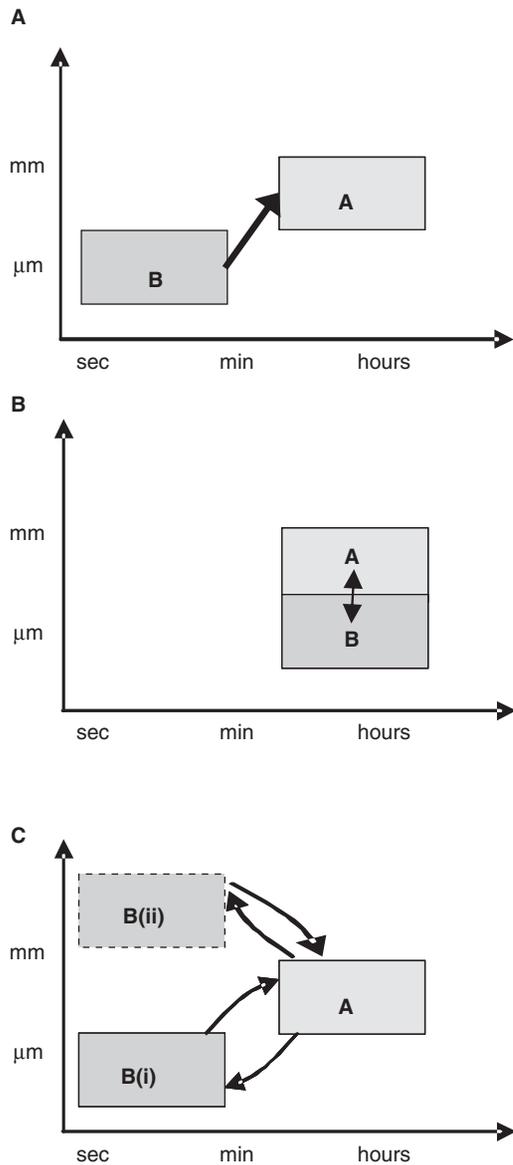


Figure 1: Placement on the SSM [55, 56] of alternative approaches to multi-scale modelling: **(A)** Individual based model of cellular behaviour (box A) is embedded in a continuum description of an extracellular field, which remains at steady state (box B); **(B)** cellular level model (box A) encapsulates an individual-based description of a subcellular component (box B), which is updated on a similar timescale and **(C)** subcellular process (box Bi) or biophysical process (box Bii) is represented by a second model component, which is updated on a faster timescale.

paracrine factor, nutrient or oxygen. In many cases, the concentration of such a factor can be assumed to be at steady-state with respect to the time scale of the cellular based model component.

This approach has been used in order to simulate spatial variations in solid tumour growth according

to the proximity of blood vessels [25] and microvessel assembly and remodelling [26]. In these examples, individual-based representations of tumour [25] or vascular components (pericyte, endothelial and smooth muscle cells) [26] were modelled in conjunction with a CA lattice, on which the relevant factors (oxygen and nutrients in [25]; transforming growth factor β (TGF- β), platelet-derived growth factor B (PDGF-B) and exogenous vascular endothelial growth factor (VEGF) in [26]) diffused on a similar time scale in accordance with simple rules.

As well as incorporating the concept of a continuum field that can influence the behaviour of cells in some way, a similar approach can be used to represent extracellular factors or structures that can be actively modified by the cells. Take, as example, ECM modification during the wound healing process. Dallon and colleagues extended an early continuum model that represented the interaction of fibroblast cells with the orientation of fibres in the surrounding collagen matrix [27] to incorporate an individual-based description of fibroblast migration on a pre-existing matrix (the ‘orientation’ model) [28]. Fibre alignment and density were represented by a vector field, which could be modified, deposited or degraded by fibroblast agents which were later permitted to proliferate [29]. A phenomenological representation of the temporal variation of TGF- β observed in wound healing experiments, and rules determining how fibroblast migration speed, proliferation and the deposition and degradation of matrix proteins were modified by the concentration of the growth factor, were later incorporated [30]. More recently, the role of chemoattractant produced within the wound bed was investigated [31]. The chemoattractant was assumed to have reached steady state prior to fibroblasts migrating into the wound, so remained fixed throughout the course of the simulation. From the results, it was predicted that the presence of the chemoattractant gradient reinforced collagen fibre alignment, as fibroblasts that initially migrated into the wound would lay down fibres to guide the following cells—behaviour with potential implications for tissue scarring.

Spatially hierarchical approaches

A second approach to multi-scale modelling is the explicit inclusion of additional hierarchies of agents representing subcellular or supracellular entities that interact on a similar time scale. This approach allows

subcellular molecular and protein interactions that might impact upon cellular behaviour to be abstracted to stochastically driven events (e.g. bonding or dissociation of particular cell surface receptors and ligands). These events are determined by simple rules, which, if desired, may be 'tuned' to represent the continuous kinetics of biochemical interactions. The explicit inclusion of spatially defined subagents allows the interactions of cells with their micro-environment (e.g. diffusing biochemical factors or extracellular structures) to be resolved on a finer spatial scale. Conversely, the inclusion of supra-cellular agent hierarchies can allow the investigation of cellular or subcellular interactions on tissue, organ, or even systemic behaviour.

An example of this approach is a multi-scale agent model developed to explore the mechanism of initiation of angiogenesis in response to VEGF [32]. In this case, the model, representing a short segment of a single vessel, incorporated two distinct hierarchies of agents: memAgents, representing distinct segments of endothelial cell membrane, and cell agents representing endothelial cells with either a tip or stalk phenotype. MemAgents, which were encapsulated within individual endothelial cell agents contained receptors for the growth factor VEGF. Ligation of VEGF receptors resulted in activation of delta ligand, which in turn, activated notch receptors and down-regulated VEGF receptors on neighbouring cells (i.e. a negative feedback loop). Activation and regulation processes were modelled by equation-based rules. At each time step, proteins associated with MemAgents were summed over each of the endothelial cell agents, with the number of activated VEGF receptors determining the allocation of a tip or stalk phenotype. MemAgents associated with tip cells could extend or retract filopodia by the transfer of 'actin tokens'. Simulations run in uniform and gradient distributions of VEGF produced distinct emergent behaviours, with random extension and retraction of filopodia in uniform VEGF fields, and filopodia directed towards the VEGF source in linear gradients. The adoption of an alternating 'salt and pepper' pattern of tip and stalk cell phenotypes was an emergent behaviour in both cases, with the model predicting an oscillating, rather than stable pattern in high VEGF concentrations.

A related biological phenomenon is the interaction of immune cells with the luminal surface of blood vessels. Tang and colleagues developed a multi-scale agent-based model of an *in vitro* flow

chamber experiment, comprising encapsulated software agents (representing virtual leukocytes) that could interact with flow chamber surface agents (representing sections of the vessel wall *in vivo*) via a set of membrane units [33]. The interaction of membrane and surface units was determined by the rule-based interaction of various cytokines and adhesion molecules, allowing the formation of contact zones. The presence or absence of bonds within contact zones caused the leukocyte agents to ratchet forwards on the surface, leading to intermittent rolling behaviour followed by firm adhesion. This emergent behaviour closely resembled the behaviour observed experimentally. As well as illustrating the applicability of a multi-scale agent approach, where agents representing successively smaller scales can be encapsulated within one another, this article also set out to model an *in vitro* rather than an *in vivo* system, where the behaviour of agents at different levels could be validated more easily.

As well as encapsulating subcellular agents, cellular level models can also be encapsulated within supracellular agent hierarchies. The process of systemic inflammation underlying multiple organ failure as a consequence of acute respiratory distress syndrome (ARDS) was explored by combining agent representations of organ and vascular surfaces [34]. Specifically, an agent model of endothelial/inflammatory cell interactions [35, 36] was combined with a second, hierarchical agent model of the organ luminal surface comprising gut or pulmonary epithelial cells interconnected by tight junctions. The integrity of the epithelial junctions could be disrupted by the presence of pro-inflammatory factors, including intracellular NF- κ B and nitric oxide. In the combined model, endothelial and epithelial surface models were represented as parallel layers, with the connecting space representing blood or lymphatic vessels across which inflammatory cell agents were free to move. Model predictions were validated against *in vivo* observations reported in the literature. To our knowledge, this is the only hierarchical agent approach that to date has been applied to simulate a systemic disease process.

As an alternative to explicitly representing subcellular structures using a hierarchical agent approach, subcellular biological scales may be incorporated phenomenologically into a cellular level model. For instance, the role of random genetic mutations in determining the relevant 'fitness' of

clonal subpopulations in tumour biology was explored by allowing random changes in proliferation rate to simulate tumour cells in a CA-based model of tumour expansion [37]. More recently, a phenomenological representation of the loss of the metastasis-suppressor E-cadherin protein has been used in an agent-based model to explore the interactions between normal and mutated cells within mixed populations (D.C. Walker *et al.*, submitted for publication).

Temporally separated approaches

This category of model explicitly incorporates representations of processes which change on a faster time scale than cellular scale phenomena (most notably proliferation). Examples include (i) the intercellular diffusion of autocrine or paracrine growth factors, which can typically diffuse a distance roughly equivalent to a cell diameter in less than a second, (ii) receptor-binding or intracellular protein phosphorylation events, which occur within seconds to minutes at typical concentrations or (iii) biophysical phenomena such as vascular flow, that may result in the transport of biochemical factors, or even cells. These approaches are more powerful in explicitly capturing the complex interactions between cells and their biochemical and biophysical microenvironments, as well as the detailed dynamics of intercellular interactions.

The modelling of integrated multiplex biological processes with inherent time scale differences (e.g. biochemical regulation of cell proliferation) can be achieved by applying time-splitting techniques, whereby a larger time step is used to represent the slower, cellular level behaviour, and a faster time scale represents the signalling processes. By contrast to the models described in sections 'Cellular-continuum approaches' and 'Spatially hierarchical approaches' above, where subcellular processes are represented by a lower hierarchy of agents whose states are also determined by simple rules, in this case the processes are usually represented by mathematical differential equations that are solved using standard numerical techniques. The two model components are usually run consecutively, with an update of the agent or CA model followed by an update of the submodel representing the faster process. In the case of intercellular signalling via the diffusion of biochemical factors, changes in cell number, location or factor secretion will influence the boundary conditions or sources of molecular species in the

signalling model, whereas the spatial distribution or intracellular concentration of key molecular or protein species will influence cell fate decisions. This scheme is thus well suited for exploration of the concept of intercellular interactions by diffusive or membrane bound signalling, or cellular interaction with the molecular environment, for instance, tumour cell behaviour in response to diffusing oxygen or nutrients. In one example, the role of phospholipase C (PLC γ) in determining cell phenotype in gliomas was investigated by extending previously developed single scale CA models (e.g. [25, 38]) to include the concept of autocrine TGF α signalling [39, 40]. These extended models explicitly encapsulated the processes of ligand release and diffusion, binding of receptors and activation of an intracellular-signalling cascade that controlled the decision to adopt a migratory or proliferative phenotype, and provided a feedback mechanism by modulating the expression of genes encoding receptor and ligand. The previously developed 2D CA model framework was used to simulate tumour cell behaviour, whereas diffusion of TGF α , nutrients and oxygen from a blood vessel were represented by partial differential equations (PDEs), and growth factor binding, intracellular signalling pathways and gene interactions were represented by ordinary differential equations (ODEs). Simulations generated using this multi-scale model suggested that the density of receptors on the cell surface could influence the rate of tumour expansion up to a maximum saturation level, and that both migratory and proliferative cell phenotypes were necessary for a high rate of tumour expansion. A 3D implementation of this model was also applied to investigate the effect of epidermal growth factor (EGF) signalling in non-small cell lung cancer [41]. In a recent paper, the ODE-based Tyson-Novak model of the cell cycle [42] was incorporated within modelled cells, allowing a more in-depth exploration of proliferation control [43].

Multi-scale methods have also been used in order to study the role of biochemical signalling in the context of growing epithelial cell populations. In one example, differences between predicted and measured growth kinetics obtained respectively from a lattice-free agent model simulation and epithelial cells grown in culture [24] indicated the existence of a contact-mediated pro-proliferative mechanism that was not present in early versions of the model. These observations prompted the development and

integration of model components representing the production, diffusion and binding of autocrine growth factors in growing epithelial cell populations [44], and juxtacrine-mediated intracellular signalling via EGF receptor (EGFR) activation [45]. In the latter case, the size and frequency of intercellular contacts, which provided sites of intercellular signalling, were closely based on observations derived from time-lapse microscopy studies of real cells grown in different culture environments. Again, diffusive intercellular signalling processes were simulated using PDEs, and juxtacrine and intracellular signalling was represented using ODEs, with data passed at each iteration from the slower agent model to the faster signalling model. New data relating to receptor occupancy [44] or activation of downstream intracellular signalling molecules [45] was passed back to the agent model, where it influenced the decision of individual agents to progress through the G1/G0 cell cycle checkpoint. Simulations generated by this model indicated that the response of individual cells was influenced by the local microenvironment, leading to population heterogeneity. This heterogeneity was masked when data was averaged over the entire cell population (as is the case for many cell biological-based assays, such as western blotting) and highlighted the danger of extrapolating population-derived data to individual cells. Similar results were reported in [46], where a generic model consisting of a fixed ring configuration of cells, each encapsulating a generic three-step activator-inhibitor signalling pathway with feedback properties and coupled by diffusion of the end product of the pathway, was used to demonstrate the loss of resolution associated with averaging measurements over space and time.

The role of direct cell-cell contact-mediated signalling in the context of tumour invasion has also been investigated using multi-scale models [47]. In this case, the adhesive and migratory behaviour of individual cell agents was determined by the intracellular distribution of E-cadherin and cytoplasmic β -catenin, with the latter represented by ODEs. Loss of intercellular contact resulted in an increase of free cytoplasmic β -catenin, active 'random' migration, and reduced intercellular adhesive force. The results suggested that tumour growth could be controlled directly by intercellular interactions. The down-regulation of adhesion molecules by one cell was shown to result in a 'wave' of loss of adhesion and onset of active migration, leading to the emergence of invasive behaviour.

Finally, multi-scale approaches have also been used to integrate individual based descriptions of cells with representations of vascular flow—a biophysical process which, like signalling processes, occurs on a faster time scale than cellular growth and proliferation. In [48], a CA-based model of individual tumour cell migration and proliferation was integrated with an explicit PDE-based representation of oxygen diffusion from blood vessels. The topology of the blood vessel network was explicitly represented and was itself the emergent outcome of a separate model that was run prior to the tumour cell CA model. A pressure gradient was applied to an initial configuration of vessels, with the resultant flow rate through the network calculated using Kirchoff's Laws. Vessel radii were updated to reflect remodelling in response to changes in flow, and the new flow rate calculated and fed back into the first model until a steady state was reached. The final network topology, and hence haematocrit distribution, provided the oxygen source for the tumour CA model described above (which, again, was extended to incorporate the Tyson–Novak cell cycle model [42]). Hence, as well as incorporating subcellular scale phenomena into a cellular level CA model, this work demonstrated the applicability of physics-based models in generating emergent tissue structures (in this case, vessel networks) that may in turn influence the behaviour of the surrounding cells. Although the vessel and tumour growth are separated in time, with the final vessel network geometry forming static nutrient sources in the tumour growth model, in reality, the two systems would evolve together. This suggests that full integration of the two models could provide a more realistic representation of the system.

A fully integrated approach to combining vascular flow and cellular interactions has been developed in order to study inflammatory processes [49]. In this case, endothelial cells lining the vessel surface and circulating leukocytes were represented as individual agents, whereas bulk blood transport was modelled separately using a network flow model. Although the transport of leukocytes by flow was not explicitly modelled, the prevalence of cells available for interaction at each time-step was determined by the computed haemodynamic parameters. Transport of various soluble cytokines [interleukins, tumour necrosis factor α (TNF α) and nitric oxide (NO)] secreted by endothelial agents was included. Rules for adhesion and migration depended on the concentrations of these factors, as well as wall shear

stress computed by the flow model and the density of adhesion molecules on the leukocyte and endothelial agent surfaces. Hence, unlike [33], the effect of haemodynamics on the leukocyte/endothelial interactions were explicitly included. In simulations, this model successfully represented the sequence of leukocyte–endothelial interactions observed *in vivo* (initial contact, rolling, firm adhesion and extravasation). Simulated molecular ‘knock out’ studies also matched independently published experimental observations.

FUTURE CHALLENGES

A potential criticism of agent- and CA-based models of cellular behaviour is that they are intrinsically phenomenological and not capable of encapsulating the complexity of a biological system. As argued above and in [14], careful comparison of virtual and experimental models provides a key to allaying these criticisms. The development of multi-scale agent models, and in particular those that aim to capture the protein or gene level mechanisms that determine cell phenotype, offers a further opportunity to ensure that simulations capture the real system processes as accurately as possible. Many of the examples discussed above consist of a cellular and subcellular (usually signalling-related) component and fit a category of bi-scale models. Relatively few models have encompassed supra-cellular level scales, with the exception of An’s work on modelling systemic inflammatory disease [34]. One explanation for the prominence of the cell/pathway model is the accessibility of signal transduction pathways and also the relative simplicity of representing multiple time scales using time-splitting approaches, which have been applied widely to the modelling of physical systems, particularly those utilizing lattice-based methodologies (e.g. [50]).

It is clear that the extension of single-level CA/software agent models to incorporate mechanistic components is only a first step in fulfilling the potential of the cell as the starting point for a middle-out approach that spans the entire biological continuum from gene to organism. However, methods of representing multiple spatial scales are non-trivial. Approaches adopted within the materials modelling community include classical multi-grid or domain decomposition methods, or more recently, gap-tooth or Heterogenous Multiscale Methods (HMM). The details of these methodologies are

beyond the scope of this review. However, we refer interested readers to [51, 52] for a discussion of these techniques. It is worth noting that these methodologies have been developed for continuum type (e.g. finite element) methodologies and their application to models representing dynamic biological processes, such as tissue growth and regeneration, may not be straightforward. More appropriate may be the continuum-atomistic coupling methodologies discussed in [53].

As yet, there is no universally adopted theoretical or computational framework for the assembly of multi-scale biological models. A recent description of a process for designing multi-scale, multi-science (i.e. encapsulating physical, chemical and biological phenomena) computational modelling frameworks based on the construction of a Scale Separation Map (SSM) has been proposed [54, 55]. This involves dissecting a complex system into constituent processes and representing each process within logarithmic axes representing time and space. The relationship(s) between any pair of processes—whether overlapping, in contact, or separated, and in the latter case, the relative direction of the offset, determines the optimal computational methodology for representing the processes. Those with sufficient spatial and/or temporal separation to justify separate models can be connected via *conduits* that can pass the necessary data and also perform any required operations (e.g. interpolation, summation or averaging in time and space), as determined by the nature of the individual models. The relative positions on the SSM of the three categories of model identified in this review are shown in Figure 1. In terms of the categories identified in [54], it can be seen that cellular steady-state approaches described above correspond to a special uni-directional case of micro-macro coupling (case 3.1 in [54]) where the cellular level behaviour does not (usually) influence the subcellular process, whereas temporally-separated processes involving subcellular signalling correspond to bi-directional micro-macro coupling. Spatially hierarchical agents represent coarse and fine structures operating on similar temporal scales (case 2) and temporally separated processes coupling cellular models with blood flow correspond to case 3.2 in [54], where slower processes on a smaller scale are coupled to faster processes operating on a larger scale.

There are other theoretical considerations associated with the development of multi-scale,

multi-paradigm models. For instance, within equation-based models of signalling mechanisms, the process of quantifying uncertainty with approaches such as sensitivity analysis is well established. How errors and uncertainties propagate from one biological scale (e.g. subcellular signalling events) to higher levels (e.g. tissue growth or assembly) is an area that has undergone relatively little investigation, though a recent publication reports a cross-scale analysis in which components of a particular modelled pathway were identified as critical in determining cellular behaviour [56].

In addition to the scarcity of theoretical frameworks for the development of multi-scale models, there is also a lack of software frameworks (the reader is referred to [57] for a general review of single-scale agent-based modelling platforms). With the exception of [34], which was developed in the NetLogo environment (<http://ccl.northwestern.edu/netlogo/>), all models discussed in this review have involved custom-written code, usually in languages with object oriented capability to represent the agent or individual based component, in combination with libraries of standard numerical solvers for the PDE or ODE component. In some cases, external solvers have been integrated with generalized agent platforms (e.g. [39, 40]) where customized Java code was integrated with Repast (<http://repast.sourceforge.net/>), or [49] where Matlab code was integrated with NetLogo. A recent publication has proposed an open source C++/Python based computational framework for modelling immunological interactions [58]. The Multiscale Systems Immunology (MSI) framework incorporates a catalogue module—a predefined database of biological and physical data for specified cell/protein types, as well as modules representing cell motility, chemokine diffusion and reaction (these are solved sequentially using a time splitting technique). The option exists to substitute model modules (e.g. diffusion) with alternative solution techniques. The developers have placed emphasis on usability for non-expert programmers. Finally, a multi-scale computational framework based on the SSM conceptual framework is currently under development – MUSCLE or Multi-Scale Coupling Library and Environment [59]. It remains to be seen whether the frameworks are adopted for the design of future multi-scale models and if there is uptake by the biological/medical communities.

A potential drawback of multi-scale, as opposed to single scale, modelling approaches is the additional

computational overhead required to solve these models. Particular agent-based frameworks (e.g. FLAME— <http://www.flame.ac.uk/>) have been optimized for parallel computation, but the protocols used for the parallelization of agent models are not necessarily applicable to, for example, the numerical solution of PDEs representing more rapid intercellular diffusion processes. As far as we are aware, none of the models discussed in this review have been implemented on a parallel architecture, and hence all are ultimately limited in the number of agents and the complexity of the subcellular processes represented. These issues raise particular challenges for computer scientists and software engineers involved in the design of such computational frameworks. The vision for CA and software agent-based models is that they will provide a solution for systems biology to interpret dynamic, biological relationships, most importantly at the cellular level, that can then be integrated into more general multi-scale models. However, fulfilment of this vision will depend on future development and amalgamation of bioinformatics and modelling tools, including automated methods to process and encapsulate information from large-omic datasets. As argued by An [34], the strength of the software agent paradigm is that it offers a conceptually intuitive, non-mathematical framework within which researchers in the biomedical community can represent, exchange and revise dynamic biological knowledge. Individual based models can serve as a substantiation of conceptual models, and provide a means for simulation to test hypotheses. The inherent capacity of the paradigm to represent non-deterministic and heterogeneous behaviour offers the opportunity to investigate the role of diversity in cellular populations, and highlight potential implications for experimental methodologies [45].

There are many challenges, both computational and theoretical, facing those involved with the development of multi-scale models. These are in addition to the challenges relating to the engagement of biologists and extraction of appropriate biological data to both inform and validate models, which are central to all computational biological investigations of any scale or paradigm. However, by embracing such challenges, life and physical/computational scientists can work together towards the goal of developing tools to aid understanding and prediction of complex biological systems, and ultimately, to guide intervention in human disease.

Key Points

- Biological systems are suited to 'middle out' modelling approaches.
- The virtual cell is a promising, paradigm for middle-out modelling because, as the basic unit of life, the cell is responsible for processing and integrating information from other cells and from the environment to regulate homeostasis and emergent tissue-level behaviour.
- Software agents and cellular automata provide useful paradigms for modelling biological systems as they can be used as direct representations of cells, facilitating model validation against experimental systems.
- The modelling of complex biological phenomena that extend across temporal and spatial scales may be achieved by combining different modalities (e.g. software agent with ordinary differential equations) or by using hierarchical agent models.
- Different approaches to combine model modalities give different degrees of temporal/spatial resolution.
- Existing models are primarily bi-scale, consisting of a cellular level component integrated with a single supra- or subcellular model. In the majority of cases, the second component represents an inter- or intra-cellular signalling process.
- Remaining challenges for computational modellers include the development and adoption of a generalized framework and software for multi-scale modelling, alongside the continuing challenges of engaging with biologists and extracting appropriate biological data to inform and validate models.
- The future vision for computational modelling is to provide predictive models of complex multi-scale, dynamic biological systems, but to reach this full potential is a grand challenge that will require automated methods to integrate omic datasets.

FUNDING

JS holds a research chair funded by York Against Cancer. DW holds an RCUK Fellowship.

References

1. Noble D. *The Music of Life: Biology Beyond the Genome*. Oxford: Oxford University Press, 2006.
2. US Department of Energy Office of Science. *Genomics: GTL Roadmap*, 2005. <http://genomicsgtl.energy.gov/roadmap/index.shtml> (8 December 2008, date last accessed).
3. Southern J, Pitt-Francis J, Whiteley J, *et al.* Multi-scale computational modelling in biology and physiology. *Prog Biophys Mol Biol* 2008;**96**:60–89.
4. Noble D. The rise of computational biology. *Nature reviews. Mol Cell Biol* 2002;**3**:459–63.
5. Cristofolini L, Taddei F, Baleani M, *et al.* Multiscale investigation of the functional properties of the human femur. *Philos Transact A Math Phys Eng Sci* 2008;**366**:3319–41.
6. Noble D. Modeling the heart – from genes to cells to the whole organ. *Science* 2002;**295**:1678–82.
7. Gaffney E, Maini P, Sherratt J, Tuft S. The mathematical modelling of cell kinetics in corneal epithelial wound healing. *J Theor Biol* 1999;**197**:15–40.
8. Johnston M, Edwards C, Bodmer W, *et al.* Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer. *Proc Natl Acad Sci USA* 2007;**104**:4008–13.
9. Ermentrout GB, Edelstein-Keshet L. Cellular automata approaches to biological modeling. *J Theor Biol* 1993;**160**:97–133.
10. Deutsch A, Dormann S. *Cellular Automaton Modeling of Biological Pattern Formation*. Boston: Birkhauser, 2004.
11. Chopard B, Droz M. *Cellular Automata Modeling of Physical Systems*. Cambridge: Cambridge University Press, 1998.
12. Byrne H, Drasdo D. Individual-based and continuum models of growing cell populations: a comparison. *J Math Biol* 2008;**58**(4–5):657–87.
13. Wooldridge M, Jennings NR. Intelligent agents – theory and practice. *Knowl Eng Rev* 1995;**10**:115–52.
14. Thorne BC, Bailey AM, Peirce SM. Combining experiments with multi-cell agent-based modeling to study biological tissue patterning. *Brief Bioinformatics* 2007;**8**:245–57.
15. Graner F, Glazier JA. Simulation of biological cell sorting using a two-dimensional extended Potts model. *Phys Rev Lett* 1992;**69**:2013–6.
16. Cickovski T, Aras K, Alber MS, *et al.* From genes to organisms via the cell – a problem-solving environment for multicellular development. *Comput Sci Eng* 2007;**9**:50–60.
17. Noble D. From the Hodgkin-Huxley axon to the virtual heart. *J Physiol* 2007;**580**(Pt 1):15–22.
18. Lees M, Logan B, King J. Multiscale models of bacterial populations. *Proc 2007 Winter Simulat Conf* 2007;**1–5**:860–9.
19. Dallon J, Jang W, Gomer RH. Mathematically modelling the effects of counting factor in *Dictyostelium discoideum*. *Math Med Biol* 2006;**23**:45–62.
20. Robertson S, Smith C, Langhans A, *et al.* Multiscale computational analysis of *Xenopus laevis* morphogenesis reveals key insights of systems-level behavior. *BMC Syst Biol* 2007;**1**. <http://www.biomedcentral.com/1752-0509/1/46> (8 December 2008, date last accessed).
21. Stoma S, Lucas M, Chopard J, *et al.* Flux-based transport enhancement as a plausible unifying mechanism for auxin transport in meristem development. *PLoS Comput Biol* 2008;**4**:e1000207.
22. Zygorakis K, Bizios R, Markenscoff P. Proliferation of anchorage-dependent contact-inhibited cells: I. Development of theoretical models based on cellular automata. *Biotechnol Bioeng* 1991;**38**:459–70.
23. Drasdo D, Hohme S. Individual-based approaches to birth and death in avascular tumors. *Math Comput Modell* 2003;**37**:1163–75.
24. Walker D, Southgate J, Hill G, *et al.* The Epitheliome: modelling the social behaviour of cells. *Biosystems* 2004;**76**:89–100.
25. Mansury Y, Kimura M, Lobo J, Deisboeck TS. Emerging patterns in tumor systems: simulating the dynamics of multicellular clusters with an agent-based spatial agglomeration model. *J Theor Biol* 2002;**219**:343–70.
26. Peirce SM, Van Gieson EJ, Skalak TC. Multicellular simulation predicts microvascular patterning and in silico tissue assembly. *FASEB J* 2004;**18**:731–3.
27. Dallon JC, Sherratt JA. A mathematical model for fibroblast and collagen orientation. *Bull Math Biol* 1998;**60**:101–29.
28. Dallon JC, Sherratt JA, Maini PK. Mathematical modelling of extracellular matrix dynamics using discrete cells: fiber

- orientation and tissue regeneration. *J Theor Biol* 1999;**199**:449–71.
29. Dallon J, Sherratt J, Maini P, Ferguson M. Biological implications of a discrete mathematical model for collagen deposition and alignment in dermal wound repair. *IMA J Math Appl Med Biol* 2000;**17**:379–93.
 30. Dallon JC, Sherratt JA, Maini PK. Modeling the effects of transforming growth factor-beta on extracellular matrix alignment in dermal wound repair. *Wound Repair Regen* 2001;**9**:278–86.
 31. McDougall S, Dallon J, Sherratt J, Maini P. Fibroblast migration and collagen deposition during dermal wound healing: mathematical modelling and clinical implications. *Philos Trans A Math Phys Eng Sci* 2006;**364**:1385–405.
 32. Bentley K, Gerhardt H, Bates PA. Agent-based simulation of notch-mediated tip cell selection in angiogenic sprout initialisation. *J Theor Biol* 2007;**250**:25–36.
 33. Tang J, Ley KF, Hunt CA. Dynamics of in silico leukocyte rolling, activation, and adhesion. *BMC Syst Biol* 2007;**1**:14.
 34. An G. Introduction of an agent-based multi-scale modular architecture for dynamic knowledge representation of acute inflammation. *Theor Biol Med Model* 2008;**5**:11. <http://www.tbiomed.com/content/5/1/11> (8 December 2008, date last accessed).
 35. An G. Agent-based computer simulation and SIRS: building a bridge between basic science and clinical trials. *Shock* 2001;**16**:266–73.
 36. An G. In silico experiments of existing and hypothetical cytokine-directed clinical trials using agent-based modeling. *Crit Care Med* 2004;**32**:2050–60.
 37. Kansal AR, Torquato S, Chiocca EA, Deisboeck TS. Emergence of a subpopulation in a computational model of tumor growth. *J Theor Biol* 2000;**207**:431–41.
 38. Deisboeck TS, Berens ME, Kansal AR, *et al.* Pattern of self-organization in tumour systems: complex growth dynamics in a novel brain tumour spheroid model. *Cell Prolif* 2001;**34**:115–34.
 39. Athale C, Mansury Y, Deisboeck T. Simulating the impact of a molecular 'decision-process' on cellular phenotype and multicellular patterns in brain tumors. *J Theoret Biol* 2005;**233**:469–81.
 40. Athale CA, Deisboeck TS. The effects of EGF-receptor density on multiscale tumor growth patterns. *J Theoret Biol* 2006;**238**:771–9.
 41. Wang Z, Zhang L, Sagotsky J, Deisboeck T. Simulating non-small cell lung cancer with a multiscale agent-based model. *Theoretical Biol Medical Modelling* 2007;**4**. <http://www.tbiomed.com/content/4/1/50> (8 December 2008, date last accessed).
 42. Tyson JJ, Novak B. Regulation of the eukaryotic cell cycle: molecular antagonism, hysteresis, and irreversible transitions. *J Theoret Biol* 2001;**210**:249–63.
 43. Zhang L, Athale C, Deisboeck T. Development of a three-dimensional multiscale agent-based tumor model: simulating gene-protein interaction profiles, cell phenotypes and multicellular patterns in brain cancer. *J Theoret Biol* 2006;**244**:96–107.
 44. Walker D, Wood S, Southgate J, *et al.* An integrated agent-mathematical model of the effect of intercellular signalling via the epidermal growth factor receptor on cell proliferation. *J Theoret Biol* 2006;**242**:774–89.
 45. Walker DC, Georgopoulos N, Southgate J. From pathway to population – a multiscale model of juxtacrine EGFR-MAPK signalling. *BMC Syst Biol* 2008;**2**. <http://www.biomedcentral.com/1752-0509/2/102> (8 December 2008, date last accessed).
 46. Rajesh S, Sinha S. Measuring collective behaviour of multicellular ensembles: role of space-time scales. *J Biosci* 2008;**33**:289–301.
 47. Ramis-Conde I, Drasdo D, Anderson AR, Chaplain MA. Modeling the influence of the E-cadherin-beta-catenin pathway in cancer cell invasion: a multiscale approach. *Biophys J* 2008;**95**:155–65.
 48. Alarcon T, Byrne HM, Maini PK. A cellular automaton model for tumour growth in inhomogeneous environment. *J Theoret Biol* 2003;**225**:257–74.
 49. Bailey A, Thorne B, Peirce S. Multi-cell agent-based simulation of the microvasculature to study the dynamics of circulating inflammatory cell trafficking. *J Biomed Eng* 2007;**35**:916–36.
 50. Alemani D, Chopard B, Galceran J, Buffle J. LBGK method coupled to time splitting technique for solving reaction-diffusion processes in complex systems. *Phys Chem Chem Phys* 2005;**7**:3331–41.
 51. Weinan E, Engquist B. Multiscale modeling and computation. *Notices AMS* 2003;**50**:1062–70.
 52. Panchal JH, Choi HJ, Allen JK, *et al.* Robust multiscale simulation-based design of multifunctional materials. *International Conference on Advanced Materials Design & Development. Notices of the AMS Goa, India, 2005.*
 53. Miller RE, Tadmor EB. Hybrid continuum mechanics and atomistic methods for simulating materials deformation and failure. *MRS Bull* 2007;**32**:920–6. Available at <http://www.srl.gatech.edu/Members/jpanchal/Publications/JP.HC.goa.dlmrev%20april06.ver2.pdf>.
 54. Hoekstra AG, Lorenz E, Falcone JL, Chopard B. Toward a complex automata formalism for multiscale modeling. *Int J Multiscale Comput Eng* 2007;**5**:491–502.
 55. Evans D, Lawford P, Gunn J, *et al.* The application of multiscale modelling to the process of development and prevention of stenosis in a stented coronary artery. *Philos Trans A Math Phys Eng Sci* 2008;**366**:3343–60.
 56. Wang Z, Birch CM, Deisboeck TS. Cross-scale sensitivity analysis of a non-small cell lung cancer model: linking molecular signaling properties to cellular behavior. *Biosystems* 2008;**92**:249–58.
 57. Railsback SF, Lytinen SL, Jackson SK. Agent-based simulation platforms: Review and development recommendations. *Simulat Trans Soc Model Simulat Int* 2006;**82**:609–23.
 58. Mitha F, Lucas T, Feng F, *et al.* The multiscale systems immunology project: software for cell-based immunological simulation. *Source Code Biol Med* 2008;**3**. Available at <http://www.scfbm.org/content/3/1/6> (8 December 2008, date last accessed).
 59. Hegewald J, Krafczyk M, Tolke J, *et al.* An agent-based coupling platform for complex automata. *Comput Sci – ICCS* 2008;**5102**:227–33.