Computational discrete models of tissue growth and regeneration

Francisco Azuaje

Abstract

Tissue growth and regeneration are fundamental processes underpinning crucial physiological and pathological conditions: ranging from normal blood vessel network development, response to stem cells therapy and cancers. Modelling of such biological phenomena has been addressed through mathematical and algorithmic approaches. The former implements continuous representations based on differential equations. The latter exploit operational descriptions in the form of computing programs to represent and execute the models. Within this area, models that define the cell as the fundamental unit of model development, as well as discrete representations of different model entities, are important to plan *in vitro* experiments and to generate new testable hypotheses. This article reviews the application of algorithmic discrete models, with a focus on tissue growth and regeneration phenomena in the context of health and disease. The review begins with an overview of basic concepts, problems and approaches of computational discrete models. This will include a discussion of basic assumptions and design principles. An overview of key cell-driven approaches and examples of applications in tissue growth and regeneration is provided. The specification, implementation and analysis of a model are illustrated with a hypothetical example, which mimics the branching and sprouting patterns observed in blood vessel network development. The article concludes with a discussion of current challenges and recommendations.

Keywords: cell-driven models; computational discrete models; cellular automata; agent-based models; angiogenesis; tissue growth; tissue regeneration

INTRODUCTION

A crucial aim of biology in the post-genome era is to understand how biological components interact in a dynamic, parallel or concurrent fashion. The spectrum of biological components ranges from molecules, cells, tissues and organs to entire organisms and populations. In complex biological systems, interactions at lower-scales of biological organization (e.g. molecules and cells) induce new, emergent properties that are observable at higher scales (e.g. tissue and organ) [1, 2]. Tissue growth and regeneration are among such complex phenomena. They are fundamental processes underpinning crucial physiological and pathological conditions: ranging from normal blood vessel network development, response to stem cells therapy and cancers.

Biological phenomena modelling may be categorized into two main research paradigms: mathematical and algorithmic approaches [3, 4]. The former relies on the simulation of systems using sets of differential equations, and encapsulate 'continuous' models of time and space. The latter, also referred to as computational models, use operational descriptions in the form of computing programs to represent and execute the models, i.e. an 'executable biology' [3]. As Priami [4] pointed out, mathematical models describe the numerical changes of the variables defining the system, while algorithmic models indicate how and why this dynamic occurs. Mathematical, continuum models have been traditionally applied to investigate sub-cellular behaviour, in which large amounts of molecules are

Corresponding author. Francisco Azuaje, Laboratory of Cardiovascular Research, Centre de Recherche Public-Santé, L-1150, 120 route d'Arlon, Luxembourg. E-mail: francisco.azuaje@crp-sante.lu

Francisco Azuaje is a Senior Researcher with the CRP-Santé in Luxembourg, where he leads research in translational bioinformatics, including disease biomarker discovery and algorithmic approaches to understanding biological systems. CRP-Santé is the leading Luxembourg public organization performing basic, pre-clinical and clinical research in biomedicine. Its Laboratory of Cardiovascular Research investigates translational applications to prognosis and treatment.

assumed to exist to enable their description with rate equations and average changes [5].

Within the area of algorithmic approaches, programming languages based on concurrency theory have become powerful tools to infer novel biological knowledge in complex cellular systems [4]. These methods represent biological entities as software programs, that are executed in parallel, and whose interactions are represented by the messages exchanged between them. Other examples of algorithmic approaches are those based on Petri nets, Boolean networks and other logic methodologies, software verification tools and cellular automata [3, 4, 6, 7]. Unlike methods based on differential equations, discrete algorithmic approaches represent distributed, non-sequential models of complex global behaviour, which cannot be inferred by specific numerical input–output relationships. Moreover, when the main objective is to model structures and behaviours at the tissue level, mathematical approaches can give qualitatively inconsistent or incorrect outputs [1, 5, 8]. This is because cells are usually represented as a single continuous variable. This also means that in traditional mathematical methods it is difficult to preserve the identity and properties of individual cells. Moreover, tissue-level visualization of the development of individual or groups of cells is not straightforward. This underscores not only the predictive power of algorithmic approaches, but also their visualization and qualitative interpretation capabilities.

Algorithmic approaches can determine system behaviour through the specification of *discrete* states, instead of 'average' responses of continuous variables. Thus, algorithmic discrete models require the researcher to design models on the basis of mechanisms and specific feasible cause–effect relationships to mimic natural phenomena [3]. In contrast to traditional mathematical models, algorithmic cell-based models aim to explicitly characterize spatial interactions and tissue-level architecture [5, 9]. In the area of tissue growth and regeneration, continuous models have shown to be unsuitable due to their limited capability to allow dynamic visualization of structures, such as those representing vascular networks [5].

In biological systems, the interaction of *entities* at one complexity scale, e.g. molecular and cellular scales, generates entities and interactions at a higher scale. Such a transition, from interactions to new entities and patterns, captures the idea of system emergence [1]. Algorithmic approaches are typically based on phenomenological descriptions of single cells, which individually may have little explanatory or predictive value. Nevertheless, their potential can be exploited when many of these single-entity models are combined and simulated in parallel to explain testable tissue-level responses [8].

Thus, discrete computational models comprise methods that simulate the development of autonomous entities and their mutual interactions. Such interactions also include relationships between the model entities and their environment. Examples of entities are proteins and cells, which are explicitly represented in space and time. Their dynamic, distributed and parallel simulation allows the visualization of higher-level, emergent behaviour at the tissue, organ or population levels.

These models can be useful to generate new hypotheses or insights, or to implement 'what-if question' engines. Such applications are possible when the behaviours of the computational model match those observed *in vitro* or *in vivo* experiments. Examples of questions that may be addressed are: can a physiological response be preserved by knocking-out a specific entity? What happens if a higher concentration of an external stimulant is added? These approaches are useful not only for testing hypotheses about mechanisms underlying biological phenomena, but also may be seen as useful tools for planning *in vivo* and *in vitro* experiments. Moreover, these methods are useful for exploring experimental settings and conditions that may be more difficult, or more expensive, to experimentally investigate in a high-throughput and systematic fashion.

This review focuses on algorithmic, cell-driven models of tissue growth and regeneration, with an emphasis on discrete computational models, i.e. those based on agent-based or cellular automata models. The reader is referred to [10] for a review focused on mathematical models in the specific context of tumour-induced angiognesis.

FUNDAMENTAL CONCEPTUAL ASSUMPTIONS AND PRINCIPLES

An important conceptual premise of computational discrete models is that, on one side, biological systems can generate complex patterns and responses in relatively simple developmental conditions. On the other, relatively simple laws can govern the

behaviour of these systems even in complex conditions [11].

Therefore, the basic assumption of discrete models is that the behaviour of individual biological entities can be encapsulated in a set of rules that combine both quantitative and qualitative information. Such rules allow each entity to respond to its environment or to make decisions. The rules are typically based on accepted models, assumptions or hypotheses published in the literature or under investigation. These rules typically describe cell–molecule, cell–cell or cell–environment interactions.

The rules are applied to change or update the *state* of the entities in a time- or space-specific context. A state may refer to a molecular concentration, cell activation state, or a specific cell type. The dynamic, integrated responses and behaviour of all autonomous entities induce a global, emergent behaviour, i.e. patterns and responses, that are not possible to predict based solely on the rules describing individual entity behaviour. Thus, local entity–entity and entity–environment interactions give rise to complex, global phenomena. This emergent behaviour is not specified in advance by the researcher and no global, centralized control of collective behaviour is considered. This means that the cells only respond to information in its local environment and the simulation exhibits self-organizing capability. Furthermore, all cells are fundamentally equal in terms of decision constraints, and no single cell is pre-defined as the 'pacemaker' of the simulation [12].

A rule explicitly describes the response of an individual entity to environmental stimuli, the states of its neighbouring entities, externally-defined conditions or its own state. Moreover, states are changed following stochastic or probabilistic conditions. For example, 'if *cell X* meets *cell Y* and *drug Z* is not present, then there is a 95% possibility that *cell X* will divide'.

The entities develop and update their states in an environment that is typically represented by geometric spaces in 2D or 3D, or grids, which aim to mimic the biological environment under study, e.g. cellular space. A 2D grid, for example, can be defined as a matrix of $m \times n$ sites. The localization of an entity on this space is typically described by Boolean variables indicating the presence (true or one) or absence (false or zero) of a specific entity in a grid site.

This digital space would be the equivalent of, for example, a Matrigel in an (*in vitro*) experiment of blood vessel growth. A Matrigel is a commercial protein mixture that mimics the extracellular environment in which blood vessels normally grow [8]. Thus, a computationally simulated space becomes a *digital Petri dish* or *in silico* cultured environment, with boundaries defined by the geometrical grid or surface defined prior to the simulation.

Time is defined by the cycles (or steps) that are required to execute the rules and update the states of all entities in the space simulated. Therefore, in a single simulation (time) step each cell senses its environment, applies rules and updates its state. The specific meaning of a simulation cycle and its relation to real biological time depends on the problem investigated. The separation between single simulation cycles may represent time scales ranging from nanosecond, milliseconds to hours or days.

In discrete models, cells and molecules are typically distributed randomly in the simulation grid. This would be equivalent to the random and isotropic distribution of cells and growth factors to initiate an *in vitro* experiment using, for example, a Matrigel [8]. Chemotaxis is a central concept in the computational modelling of tissue growth and regeneration. It is defined as the movement response of an entity to a specific signal, typically representing a substance produced by the cells or added externally as part of the experiments.

Discrete algorithmic models are non-deterministic or stochastic in the sense that they can allow the computation of different outcomes across different simulations. This is because state changes and rule execution can be controlled by probabilistic conditions, which is consistent with the operation of biological systems. These models are also useful in the presence of incomplete knowledge of the underlying mechanisms of the biological phenomena investigated. This advantage is derived from the different levels of abstraction that can be represented to obtain a given outcome, without the need to know all the specific conditions and relationships of the components in the natural system.

The validation of such models is done by *in vivo/ in vitro* analyses, which aim to reproduce the outcomes or response observed *in silico*. This can also be seen as an incremental, iterative process of experimental measurements and model adjustment steps. This allows the models (rules and parameters) to fit the observed biological data. This means that

different qualitative (visualization-based) and quantitative methodologies can be implemented to compare the computational and experimental measurements of interest, e.g. cell density or other entity concentrations. Examples of successful validation methodologies and case studies were reviewed by Thorne *et al.* [13].

Cells and molecules have been modelled and simulated as discrete entities or agents in several research domains, including cell proliferation and differentiation, tissue growth, death and regeneration. For example, models of epithelial cells and their responses to contact inhibition and calcium concentrations, in the context of wound healing, were investigated by Walker *et al.* [14]; and models of the propagation and fate of cancer cells were reported by Abbott *et al.* [15]. Discrete models of tissue growth and regeneration have been useful to answer questions related to the quantification of molecular variables that are difficult to measure *in vitro*, the robustness of the system against the perturbation of cell types and rules, the definition of the minimum sets of rules and parameters required for mimicking physiological responses and the identification of potential conflicting hypotheses [16].

KEY TYPES OF DISCRETE APPROACHES TO TISSUE GROWTH AND REGENERATION

In discrete computational models, the movement and state transformation of entities depend on the current state of the neighbouring entities. Rules may be specified as discrete 'if–then' representations or by the outcome of mathematical equations, which may also be constrained by either deterministic or probabilistic conditions [12].

Cellular automata are useful 'top–down' discrete modelling approaches, which have been applied to simulate a great variety of biological systems, including morphogenesis and tissue development [16]. As a typical discrete modelling approach, cellular automata capture systems-level mechanisms of complex biological phenomena by defining a series of decision rules that are implemented in a dynamic and parallel fashion. Most cellular automata models reduce the simulation process into a set of discrete time cycles and space coordinates, and small sets of rules to generate and update the state of different model entities, as described above and in ref. [6]. In a typical tissue growth model, at each simulation cycle a cell moves

in one of n directions with a certain probability. This is done iteratively through a number of simulation cycles. Advantages of these models include the relative simplicity of implementation and visualization, and its design flexibility and extensibility. The latter enables a relative straightforward reformulation of rules and entities using different types of programming languages.

The Cellular Potts model, an extension of cellular automata, is a discrete simulation approach that incorporates mathematical descriptions of cell connectivity and motility [17, 18]. Such mathematical descriptions represent growth constraints based on energy-minimization criteria and complementary dynamic conditions [19]. Cellular Potts models, unlike the traditional cellular automata approach, can specifically define cell structure in terms of shape and volume parameters [12, 19, 20]. In the context of tissue growth these models foster cell movements in those directions that minimize a local energy function [20]. The search and selection process is typically performed in a probabilistic fashion. In these models the interaction or aggregation between cells is defined by cell–cell contact energies. The central idea is that tissue localization, area and geometry are regulated by favouring stronger bonds, i.e. their contact energies, as well as larger cell boundaries [8]. Thus, the Cellular Potts model is a powerful approach to incorporating quantitative cellular information into discrete or cell-based models.

In the standard Cellular Potts model, the effective energy, H, summarizes the effects of all cell behaviours [20]:

$$H = \sum_{\text{neighbouring sites}} J(\sigma(x),\sigma(x'))(1 - \delta(\sigma(x),\sigma(x'))) + \lambda_A \sum_{\sigma} (a(\sigma) - A_{\text{tar}}(\sigma))^2 + \text{other penalty terms}$$

where J represents the bond energy, $\sigma(x)$ and $\sigma(x')$ are neighbouring sites, $\delta(x,y) = 1$ when $x = y$ and 0 otherwise, $a(\sigma)$ is the area of cell σ, $A_{\text{tar}}(\sigma)$ is the target area, and λ_A encodes cell resistance to compression. Other penalty terms representing cell shape and chemotaxis, for example, can be added.

Cellular Potts models and others that go beyond the 'grid-'or 'lattice-based' space representation, and which explicitly aim to incorporate information on shape and size of cells, are often included in the category of 'agent-based' approaches [5]. Walker and Southgate [9] also asserted that a key distinguishing feature between traditional cellular automata and

agent-based models is that the latter commonly display more detailed memory states and behaviour rules. Moreover, agent theory is a generalization of cellular automata.

According to the categorization of biological growth patterning provided in ref. [21], computational cell-driven models mainly aim to mimic 'sighted' and 'self-regulatory' patterns. In the former, the generation of new tissue is determined by growth factors sensed in the immediate environment (e.g. grid neighbourhood), including specific molecules or cells. In the latter, growth is mainly regulated by cross-communication between entities in the simulation space regardless of the level of influence of environmental factors. In practice, as exemplified below, tissue growth and regeneration models may rely on the combination of both control mechanisms.

Regardless of the specific modelling approach implemented, discrete models are powerful tools to visualize system outcomes that can closely resemble both qualitative and qualitative aspects of *in vivo* and *in vitro* biological systems. At a minimum, discrete models may be used to formalize established biological theories or knowledge into a computational framework. Even in the absence of experimentally-validated novel predictions, such computational representations and simulations are useful for organizing and structuring relevant knowledge underlying necessary conditions for the emergence of a biological phenomenon [16].

EXAMPLES OF PROBLEMS AND APPLICATIONS IN TISSUE GROWTH AND REGENERATION
Blood vessel growth network models

Endothelial cells (ECs) are the fundamental components in the formation and evolution of new blood vessels from existing ones, i.e. angiogenesis. During angiogenesis, ECs divide, proliferate and migrate in the cellular matrix, leading to the development of new blood vessel networks. Hypoxia, the deprivation of oxygen supply, is a key trigger of vessel growth. Hypoxia-induced growth involves the signalling of different transcription factors that regulate different angiogenic genes, such as vascular endothelian growth factors (VEGF) genes [22]. Apart from VEGF genes, several other stimulants of growth have been studied, including chemokines, cytokines and hormones. Angiogenesis also depends on the

interplay between these and anti-angiogeneic molecules, as well as on their communication and sensing capability of local environmental signals, including blood flow and pressure and structural changes in their extra-cellular environment [22]. Such agonistic and antagonistic interplay also characterize angiogenesis in the context of tumour cell growth [23]. For example, placental growth factor (PlGF) may antagonize VEGF-A through the formation of heterodimers VEGF-A–PlGF, which can cause the reduction of pro-angiogenically active VEGF-A homodimers [23]. Molecules can often have a dual agonist-antagonist role in angiogenesis depending on the cell type and environmental conditions [24].

Angiogenesis may be categorized into angiogenic remodelling and sprouting [24]. In the former, an initial vascular network is modified through network enlargement or pruning. In the latter, cell sprouting is initiated in existing vessels and directed toward a vascular tissue. Examples of computational models for both categories are introduced below.

Discrete and continuous models of angiogenesis encoding relatively small sets of biologically-meaningful behaviours have been proposed [7, 24, 25], which have aided in our understanding of this complex process. These models can mimic the initiation, progression and development of ECs in different biomedically-relevant environments. In comparison to computational discrete models, traditional mathematical models do not incorporate causal mechanistic details at the cell or tissue level, which may represent an obstacle for systems visualization [26]. Discrete approaches typically represent the behaviour of individual ECs and their local cell–cell interactions according to logic-based or quantitative statements.

Although the computational simulation of individual cells or processes, such as adhesion and chemotaxis, cannot reproduce complex patterns associated with blood vessel growth, the combination and parallel execution of multiple processes and cells can indeed accurately generate biologically meaningful behaviours [8].

Examples of pioneering models of angiogenesis were published by Stokes and Lauffenburger [27], and Anderson and Chaplain [26]. These models used probability-based rules to govern the branching of vessels, e.g. a pre-defined probability of migration of a tip cell to a neighbouring lattice position. Older models included those based on random motility,

which were replaced by probabilistic models in which ECs are sensitive to chemoattractant gradients [28]. Indeed, such as probabilistic bias towards chemo-attractant signals, e.g. growth factors, has become a central element in different types of angiogenesis models [28]. Deterministic models of angiogenesis (i.e. non-probabilistic models) have also been reported using 2D space representations [25]. Such models can predict specific outcomes, which are defined by the sequence of previous system states, rather than providing probabilistic estimations [25]. Models of angiogenesis can also incorporate different entities representing antagonists or inhibitors of growth, such as angiostatins. These models allow the visualization of patterns of growth reduction and disorganization of the vessel sprouting process [28].

An example of experimentally-validated cellular automata models of microvascular growth was reported by Peirce *et al.* [29]. Their model included representations of different types of vascular-related cells: smooth muscle cells, precursor cells and ECs. The model rules were based on established knowledge of the individual behaviour of these cells, growth factor production and diffusion. A key simulation setting involved directed, focal application of a growth factor source to the simulated tissue space, which would be equivalent to a point delivery *in vivo*. Peirce *et al.*'s model [29] predicted increases in vascular network density after such simulated treatment. The resulting vessel network patterns resembled *in vivo* network re-modelling patterns, both graphically and quantitatively, i.e. vessel length density changes. Moreover, this *in silico* modelling enabled the identification of a functional patterning module that controlled tissue growth. Such a control module was characterized by a set of growth factor and cell–cell contact signalling, including VEGF and platelet-derived growth factor molecules.

The role of chemotaxis in angiogeneis has also been modelled using different types of cellular discrete models. Chemotaxis models are commonly based on the premise that ECs release substances that diffuse in the extracellular matrix (e.g. the 2D grid of a cellular automata model) and that attract other ECs [30]. Experimental research in different organisms indicate that ECs and other cell types exhibit limited maximum velocities and cannot accelerate in response to chemoattractant gradients [30, 31].

These models have shown how ECs display diverse response patterns to various chemoattractants, chemorepellants or cell-contact inhibitors. In the former category, VEGF and fibroblast factors can be represented. The latter may include semaphorins and Slit proteins [32]. Some of these models mathematically specify the velocity of the ECs according to the strength of the gradient of the chemoattractants, including linear proportional relations [30]. Using a Cellular Potts model, Merks *et al.* [32] later reported a chemotaxis-driven angiogenesis model in which both short- and long-range chemoattracting signals governed vessel network development. Short-range signals were simulated as the main growth control mechanism, while long-range signals played a network 'steering role'. Furthermore, their model allowed the investigation of chemotaxis inhibited or activated via cell contact. Resulting patterns resembled those observed in *in vitro* vessel networks.

Topa [33] has investigated blood vessel network growth in cancer tissue using a 'graph of cellular automata' model. This model combines the standard cellular automata approach with spatial descriptors based on graph theory [34, 35]. Such descriptors are vectors of statistical features representing the topology of the emergent blood vessel networks, which provide the basis for the automatic assessment and classification of the tissue patterns. This allows more accurate, quantitative comparisons between simulation results and *in vitro* experiments using pattern recognition techniques. Topa's model [33] of tumour-induced angiogenesis covered aspects of tumour cell activation, growth factor and nutrient distributions. It focused on the following process parameters: growth factor coefficients, oxygen gradient coefficient, Dll4/Notch signalling level in vessel cells and growth factor activation threshold. The analytical goal was to explore the influence of the Dll4/Notch ligand on the vessel sprouting process. Using the graph-theoretic feature description, quantitative comparisons of results from simulations implemented with different model parameters were conducted. Diverse simulated network topologies were compared using multi-dimensional scaling [33]. Based on this approach and adaptations, future analyses can involve the automatic matching of the outcomes of computational simulations to images of biological vascular networks.

Also in the area of tumour vasculature development, Gevertz and Torquato [36] proposed a 2D

cellular model that considered elements of miscro-vasculature remodelling together with tumour mass evolution. Their model explicitly addressed the problem of representing brain tumour growth dynamics, in the context of different types of cells: proliferating, non–proliferating (hypoxic) and necrotic cells. Moreover, they incorporated rules that mimicked the mechanical pressure caused by the skull on the brain. The interaction between tumour growth and microvasculature development was also modelled by considering reaction-diffusion equations that described the behaviour of proteins implicated in angiogenesis. Their model incorporated descriptions of VEGF and different receptors, as well as angiogenesis antagonists. One of the most interesting theoretical insights derived from this investigation was the illustration of how tumour expansion can occur in the absence of angiogenesis, which is compatible with evidence suggesting the limitations of anti-angiogenic therapies.

Milde *et al*. [5] reported a deterministic, 3D model of sprouting angiogenesis. Their model incorporated information about extracellular matrix and of different growth factors, and explored their effect on capillary growth. The model allowed the investigation of different roles of the extracellular matrix structure and the spatial distribution of the growth factors in the morphological development of the vascular networks. This model focused on the behaviour of ECs, including their departure from the vessel wall and entrance into the extracellular matrix. Key molecular entities such as VEGF, metalloproteinases and fibronectins were represented with numerical values. Reaction-diffusion equations controlled these molecular concentrations, which were then discretized on the grid. ECs were modelled as discrete, cellular automata. In this model, as in many angiogenesis models, ECs chemotactically responded to the presence of VEGF gradients. A key contribution of this model was the explicit representation of the extracellular matrix based on the random distribution of collagen fibers, which modulates EC migration.

Vasculogenesis is the *de novo* growth and development of primary blood vessel networks during embryonic development [24]. Examples of the application of computational cell-based approaches to vasculogenesis have been reported by Savill and Merks [18], Merks *et al*. [32], Merks and Koolwijk [20] and others. Monte Carlo models of chemotactically migrating cells have been proposed, for instance, to study both functional and morphological features of *de novo* vessel growth and angiogenesis [8].

Skin growth and tissue regeneration

Models of the *in vitro* development of skin epithelial cells (keratinocyte cell models) have been reviewed in ref. [37]. In these simulations different aspects of keratinocyte cell behaviour have been represented: cell location and bonds, differentiation, apoptosis and migration. Richmond *et al*. [37] also reported adaptations of such models to high-performance computing environments, which significantly increased the speed of execution of simulations and their 'real-time' visualizations.

Adra and co-workers have reported models of the human epidermis, including 2D and 3D agent-based models of keratinocyte colony formation [38]. They proposed to represent human epidermis as the integration of three functional layers: cellular behaviour level, molecular mechanisms level and a mechanical forces level. The integrated model mimicked epidermis growth and healing based on the exchange of information across these levels: from sub-cellular to cellular, and from cellular to tissue levels. The cellular level was implemented by an agent-based model that encoded fundamental rules of epidermis cell growth and injury. The molecular level model simulated the expression and signalling of the transforming growth factor (TGF–b1) in the cell, and was implemented with the complex pathway simulator (COPASI) system [39]. COPASI is a modelling tool that has been used to run time-course simulations of different types of biochemical pathways. Between–cell forces at the tissue level were modelled with a numerical physical layer. This integrated model allowed the authors to investigate hypotheses about the role of TGF–b1 at the cellular and molecular levels in the context of epidermal wound and healing, and based on the analyses of multiple keratinocyte populations [38]. The agent-based model was built with flexible large-scale agent modelling environment (FLAME) [40], which has also been applied to other biological modelling applications. FLAME allows the automatic generation of the simulation code for models specified in XML format [38]. Adra *et al*.'s agent-based model included different types of cells and rules relating to cell bonding and proliferation.

Grant *et al*. [41] investigated models that mimic epithelial growth behaviour in different cell culture conditions. The latter is particularly significant as

traditional models of cell growth tend to be based on spatial and biochemical assumptions that are incompatible with observations in different culture conditions [41]. Their models were based on the representation of cells with 'software agents'. Such cell-based agents were capable to assemble in stable cellular structures that were consistent with features observed in *in vitro* models. The four *in vitro* culture conditions simulated by their models were: (i) cell division, apoptosis and shape change, (ii) ECs plated on a layer of collagen, (iii) inversion of polarity relative to (i), and (iv) lumen formation in collagen. The behaviour of ECs was controlled by nine rules whose responses depended on local environmental stimuli. The rules determined cellular actions, such as cell division and death. The authors also argued that these types of models could not be implemented by applying traditional methodologies, such as Cellular Potts models. In this context, one of the primary assumptions of the Cellular Potts models could not have been satisfied: that morphogenesis is mainly the result of differential cellular adhesion [41].

Cellular automata models have also been reported to investigate tissue regeneration in the context of stem cell therapy. Galvão *et al.* [42] addressed this problem in damaged tissue in the mouse heart as a consequence of cardiomyopathy caused by Chagas disease [43]. Their 2D computational tissue model allowed them to quantify the effect of bone marrow cell transplantation on the amount of fibrosis reduced, i.e. good regeneration outcome. It specified six types of entities: inflammatory cell, fibrosis area, cardiomyocyte, proinflammatory cytokine, bone marrow stem cell, and the *Trypanosoma cruzi* parasite (the etiological agent of Chagas disease). In their model the entities randomly move on a grid, with each entity having eight neighbouring sites to move. State transition rules for entity movement and transformation were incorporated. For example, 'if the number of stem cells is different from zero in the neighbourood of an inflammatory cell, then the apoptosis of this cell occurs, i.e. the site of the cell is emptied'. The effects of stem cell therapy on apotosis of inflammatory cells and fibrosis reduction were consistent with those obtained in *in vitro* studies involving experimental measurements at different times after treatment. A key conclusion was that tissue regeneration is mainly driven by the concentration pattern and distribution of the stem cells transplanted [42].

Table 1 summarizes some of the models introduced above in terms of biomedical area, modelling approach and main contributions.

BASIC DESIGN PRINCIPLES AND ILLUSTRATIVE EXAMPLE

Use the right level of description to catch the phenomena of interest. Don't model bulldozers with quarks. Goldenfeld and Kadanoff [11].

Any modelling task begins with a question: the definition of the natural problem to be investigated and simulated [11]. This problem formulation step will subsequently guide the specification of the level of abstraction or detail of the model, as well as the types of measurements to be obtained during the model development and validation.

The next crucial point of any discrete modelling design task is the identification and isolation of the fundamental processes and sub-processes to be simulated [4, 20]. This will lead to the computational specification of a biologically-feasible prototype model, which can be seen as the foundation for subsequent model expansions or re-definitions. Moreover, such an initial specification would represent the core building platform to iteratively add new entities and rules that better fit the behaviours observed in experimental validations. As pointed out above, the challenge is to indentify the most adequate level of abstraction, which is both necessary and sufficient to reproduce cellular and tissue-level behaviour. Thus, this is a problem domain-specific challenge. These computational representations are based on biologically plausible descriptions of individual cell behaviour, which typically capture qualitative relationships or responses.

This will be followed by the specification of entity types and rules, using for example 'if-then' statements. This is necessarily coupled with the formulation of space- and time-related constraints, control parameters and the identification of their *in vitro/in vivo* counterparts. Figures 1 and 2 illustrate the design, simulation and outcomes of a cellular automata model that mimics the branching and sprouting patterns observed in a blood vessel network growth experiment.

Figure 1A shows the entities, rules, input parameters and outputs of this model. Three model entities were considered: Cells, growth factors (GFs) and

Table I: Overview of key contributions

| Biomedical problem	Approach	Key contributions	*In vitro* or *in vivo* validation?	References
Microvascular remodelling	Cellular automata	Characterization of functional module controlling vessel network remodelling	Yes	[25]
Angiogenesis	Hybrid model (discrete and equation-based) of endothelial cells	Assessment of the influence of the extra-cellular matrix structure and VEGF concentrations on vessel structure	No	[5]
Epidermis growth and healing	Integrated model of sub-cellular, cellular and tissue levels. Cellular level implemented with discrete, agent-based models	Insights into the regulation of epithelial formation and healing of wounds of different sizes	No	[35]
Angiogenesis	Cellular Potts model	Investigation of vascular network patterns driven by cell contact inhibition of chemotaxis	Yes	[24]
Skin epithelial cells	Template-driven framework for agent-based modelling, computing parallel architectures	Real-time simulation visualization, significant increase in simulation speed	Yes	[34]
Vasculature evolution in brain tumour growth	Hybrid approach: cellular automata and differential equations	Connection of quantitative aspects of tumour mass and microvasculature evolution. Prediction of tumour expansion in the absence of angiogenesis	No	[33]
Epithelial cell growth	Agent-based models	Simulation of *in vitro* cell morphogenesis observed in different cell culture conditions	Yes	[38]
Tumour-induced angiogenesis	Cellular automata	Analysis of effect of the Dll4/ Notch ligand on the vessel sprouting process. Quantitative description of simulation-derived vessel networks using graph features	No	[31]
Tissue regeneration after stem cell treatment	Cellular automata	Tissue regeneration depends on the concentration pattern of the stem cells transplanted	Yes	[39]

Discrete models of tissue growth and regeneration. *In vitro* or *in vivo* validation refers to experiments specifically designed to assess the computational model investigated.

antagonists (Ants). Cells represent the units of tissue growth, mimicking the behaviour of ECs, for example. The GFs and Ants promote and inhibit cell proliferation, respectively. In each simulation iteration, i.e. a cycle, the states of all entities are updated following a single rule, whose outcome (cell state) depends on the presence or absence of GFs and Ants in a Cell's neighbourhood. Thus, an existing Cell will grow and migrate to a neighbouring site only if a GF is present in that site, and if no Ant is present. When this rule is satisfied, the Cell moves to the new site with a probability (P).

A Cell's neighbourhood is defined as the nearest, single-site space surrounding the Cell, i.e. six sites including the current position (Figure 1B). Note that in this model GFs and Ants are designed to diffuse from the bottom to the top of the grid only. In

this model a squared grid of 300×300 sites defines the simulation space (Figure 1C). The simulation outputs are quantified by measuring the Cell density observed at the end of a simulation (i.e. total number of Cells divided by the grid area). Figure 1D summarizes the simulation algorithm. The initialization phase comprises the definition of the initial 'vascular' tissue (bottom of Figure 1C) and the random distribution of GFs and Ants on the grid. After initializing the environment and model parameters, a cycle consists of the application of the rule to each grid site, followed by the random diffusion of GFs and Ants on the grid. A simulation is completed after a user-defined number of cycles.

The other simulation parameters, as defined in Figure 1A, and applied to generate the results shown in Figure 2 were: #iniCells = 300
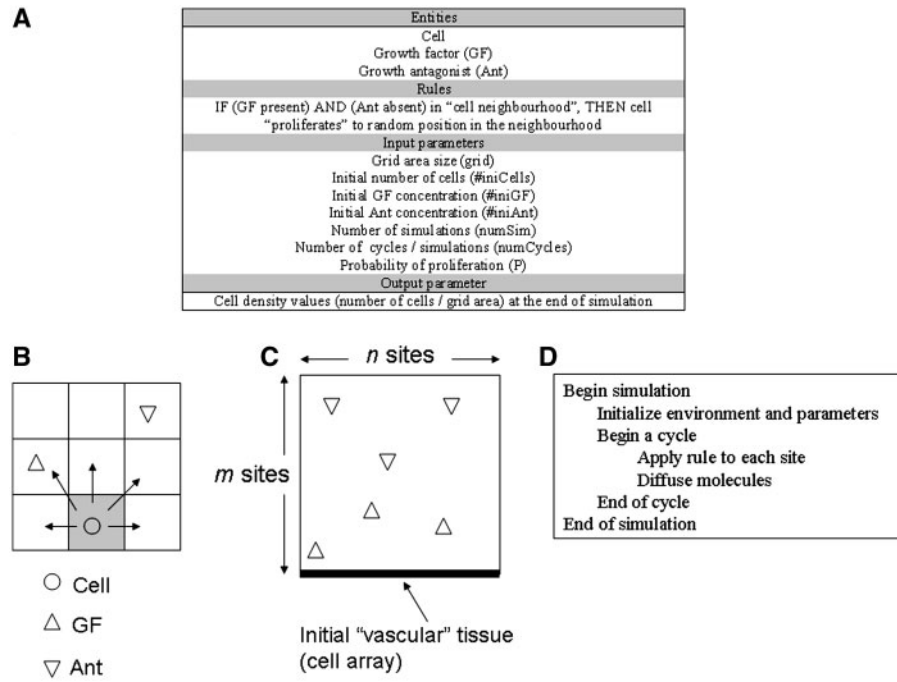
Figure 1: Specification of an example model. A 'vascular network growth' model based on cellular automata: (**A**) definition of entities, rules, input and output parameters. (**B**) Graphical illustration of a cell's neighbourhood. Arrows indicate possible cell proliferation directions. (**C**) Schematic representation of a simulation space. Thick line at the bottom represents the initial multi-cellular tissue, from which sprouting and branching subsequently originate. Symbols are used to represent different molecules randomly dispersed on a grid with $m \times n$ sites. (**D**) Algorithmic description of a simulation.

(initial cell array), numCycles = 300, numSim = 1000, P = 0.05. Different simulations were implemented with these parameters and different #iniGF and #iniAnt values. Row A in Figure 2 shows results from simulations in which #iniAnt = 10 000 for a range of #iniGF values. This panel shows examples of graphical outputs obtained from the simulations: #iniGF = 1000 (A1) and #iniGF = 10 000 (A2). Row B shows results from simulations with #iniGF = 10 000 for different #iniAnt values. The third column of Figure 2 shows these dynamic responses in terms of the observed (mean) Cell density values. As expected, increasing #iniGF values, while keeping #iniAnt constant, will induce larger and denser tissue networks (A3). Conversely, increases in #iniAnt, with #iniGF kept fixed, limits network development.

CHALLENGES AND FINAL REMARKS

Some of the most critical challenges in the development of computational discrete models of tissue growth and regeneration are: the determination of computational model parameters, the definition of the optimum level of abstraction, limitations related to computing power, their interfacing with continuous computational or mathematical models based on differential equations, the documentation and sharing of models, and their validation using *in vivo* or *in vitro* experiments [13]. The latter is essential to demonstrate both the biological relevance and applicability of models.

The definition of qualitative or quantitative parameters for the formulation of rules or states has been traditionally based on validated knowledge obtained from the literature or from human experts. Depending on the level of abstraction of the model, i.e. how many details or aspects are to be mimicked by the model, rule definition may become a complex process that may be assisted by mathematical estimation procedures or by small–scale experimental studies. However, the required level of qualitative understanding may be compromised if too many processes and variables are included [11]. Merks and Glazier [8] recommended starting investigations with a minimal model of entity types and rules. Such models can then be incrementally
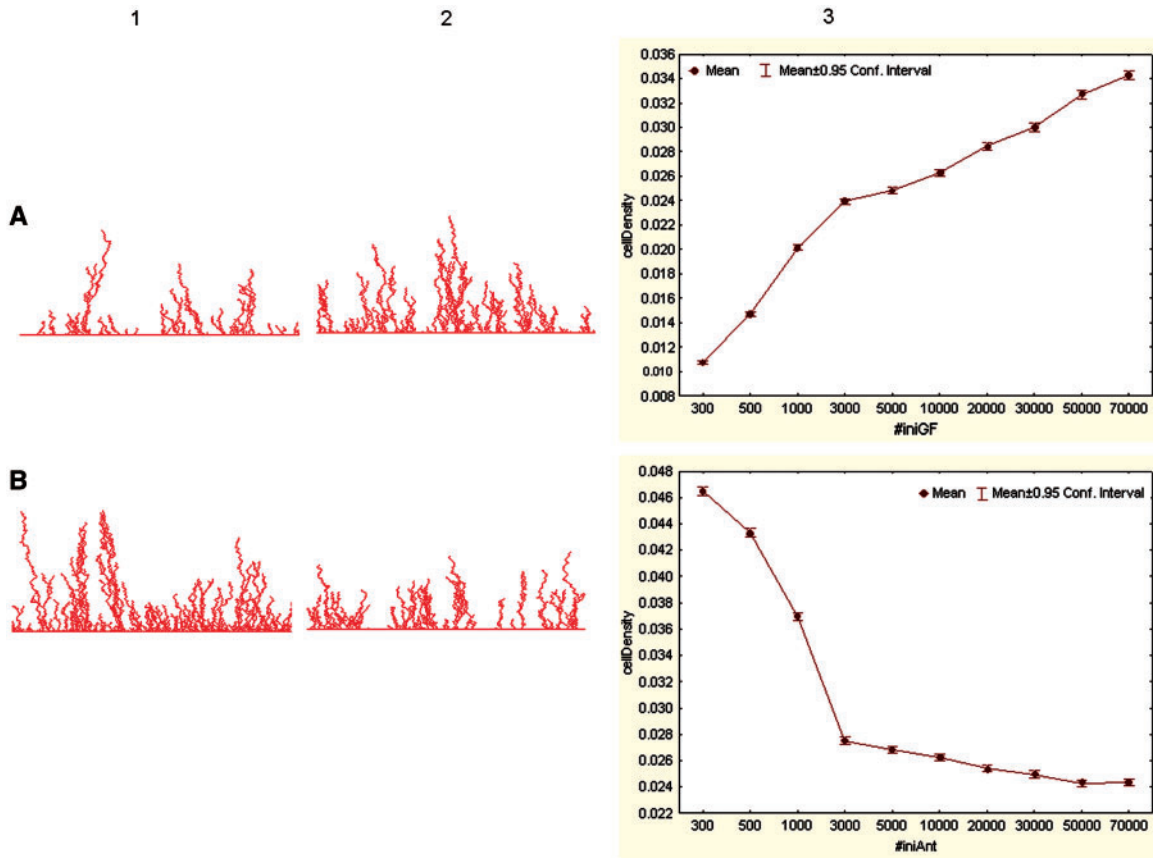
Figure 2: Example model outcomes and patterns. (**A**) and (**B**) show examples of the simulation results with constant GF and Ant concentration values (initial number of molecules) respectively. Row A shows patterns observed with #iniAnt = 10 000, and #iniGF = 1000 (A1) and #iniGF = 10 000 (A2). B1 and B2 show, respectively, output examples with #iniAnt = 500 and #iniAnt = 5000, with #iniGF = 10 000. Plots A3 and B3 summarize simulation results for different ranges of #iniAnt and #iniGF, and 1000 simulations. Cell density mean and (0.95) confidence intervals are shown at each concentration value.

expanded to accommodate additional mechanisms or entities, instead of trying to simulate all known behaviours from the beginning.

Independently of the level of detail, models will always entail simplifications of specific individual or collective behaviours, which can be useful enough to facilitate the visualization of biologically relevant, high-level responses. The speed of the simulations will be constrained by the number and complexity of the rules, space and level of abstraction. For example, computational speed will be reduced in proportion to the number of rules, entities and simulation space size.

Qualitative or quantitative inconsistencies of observed tissue-level patterns between discrete models can also assist researchers in detecting potential gaps in their knowledge. In addition, by performing large-scale perturbation analyses of these models, researchers can gain insights into potential

causal mechanisms and their resulting phenotypes, which can be linked to specific environmental or molecular alterations [4, 20]. Furthermore, the detection of qualitative and quantitative differences between computational and *in vitro* models provides the basis for an incremental, iterative process of adjustment and refinement of the computational model. The successful and rigorous experimental validation of computational models also depends on the capacity of computational scientists, biologists and clinical researchers to establish effective communication. There is a need for cooperation informed not only by expected mutual benefits, but also by an understanding of the limitations encountered in the computational modelling and 'wet lab' investigation of complex biological phenomena. Challenges and hurdles posed by such cross-disciplinary research environments may also be overcome with the availability of user-friendly tools to visualize and manipulate

models, and of more quantitative methods to compare computational and *in vitro* modelling observations.

One should also expect that different models or sets of cell behaviour rules can lead to similar system- or tissue-level output patterns. Furthermore, it is possible that observed discrepancies between computational models and experimental observations may be caused by incorrect assumptions or partly inaccurate rule descriptions [20]. This motivates researchers to enhance the levels of quantitative information included in their discrete models or to compare them to mathematical models, such as those based on differential equations. This strategy may, on the one hand, highlight potential inconsistencies or differences between *in silico* and *in vitro* models. On the other hand, it may aid in the reduction of the uncertainty associated with the patterns and quantitative observations generated by computational models. Current and future progress in molecular and cellular measurement instrumentation will contribute to a more accurate definition of model parameters, including motility rates, entity concentrations and tissue-level outputs. Examples of such technologies include cell-specific gene/protein expression analysis, microfluidics and microscopy [20].

Previous research has suggested that successful, experimentally-validated discrete models, such as those based on variations of the cellular automata approach, should be followed by more detailed mathematical modelling and simulations [44]. Conversely, researchers have also applied cellular automata models to recreate quantitative responses and patterns observed in models based on differential equations. The quantitative resolution of cellular automata patterns and measurements may be augmented by increasing the number of states and rules, the re-definition of cell neighbourhoods and averaging over large numbers of simulations [44]. Cell-based, discrete models can also be adapted to specify sub-cellular behaviours, such as regulatory networks, and to represent multi-scale spatial and temporal hierarchies [5].

The connection of computational discrete models, such as cellular automata, with mathematical models based on differential equations or alternative algorithmic approaches, such as those based on concurrent computing languages, should be decided according to specific goals and resource constraints. Advances have been reported to exploit the complementarities of discrete and continuous approaches [13].

The challenge of sharing and disseminating models goes beyond the documentation of software programs. This also requires a high-level, but detailed, description of entities, assumptions, rules, space and time scales to allow the re-implementation and reproduction of simulations and results [13].

There is a need to continue developing tools for aiding model specification, design and implementation. For instance, many computational discrete models of tissue growth and regeneration have been based on customized software prototypes and oriented to serialized computing implementations. Models suitable for high-performance computing simulations, including multi-core processing architectures, can benefit both the efficiency and complexity coverage of simulations in this area [37]. FLAME is an example of a computer-aided specification framework that can be used for specifying, storing and simulating discrete models on parallel computing architectures [37]. In FLAME, the models are defined and stored in an XML-format file. These files are automatically transformed into simulation code based on templates, which then represent the input to the simulation engine.

The development of techniques to understand biological complexity has traditionally relied on the application of 'top-down' and 'bottom-up' approaches. In the former, the analysis of system-level patterns and features guides the design of models to explain the observed, global-scale behaviours. The latter, also known as the reductionist paradigm, focuses on the study of the system components in relative isolation, and aims to merge the resulting observations into models capable to predict system-level behaviour. 'Middle-out' approaches would represent a third category. 'Middle-out' approaches aim to develop models in which an intermediate scale between lower- and higher-level scales would represent the starting point of an investigation [45]. In such middle-out approach, the level of detail or scale is then incrementally adapted according to research requirements. Computational cell-driven models can be seen as the natural drivers of middle-out approaches to modelling biological systems. According to Noble [45], Walker and Southgate [9] and other researchers, this should be a logical consequence of the recognition of the cell as the fundamental unit of life. Middle-out modelling may also facilitate the development of extensible,

foundational models that can be easily re-used, adapted or enhanced according to domain-specific requirements [41].

Computational discrete models offer qualitative and quantitative analytical approaches to investigating tissue growth and regeneration phenomena based on structured, visually-driven dynamic representations. This makes them suitable as relatively inexpensive tools for exploring a diverse range of testable hypotheses, model parameters and 'what-if' query scenarios.

Key Points

- Tissue growth and regeneration are fundamental processes underpinning crucial physiological and pathological conditions.
- Computational discrete models of tissue growth and regeneration can aid in planning *in vitro* experiments and to generate new testable hypotheses.
- Discrete computational models comprise methods that simulate the development of autonomous entities, including cells and environmental signals, and their mutual interactions.
- This makes them suitable as inexpensive tools for visually and quantitatively exploring a diverse range of testable hypotheses and 'what-if' query scenarios.

References

1. Cohen IR, Harel D. Explaining a complex living system: dynamics, multi-scaling and emergence. *J R Soc Interface* 2007;**4**:175–82.

2. Nuzzo R. On Simulating growth and form. *Biomed Comp Rev* 2008;17–23.

3. Fisher J, Henzinger TA. Executable cell biology. *Nat Biotechnol* 2007;**25**:1239–49.

4. Priami C. Algorithmic systems biology. *Commun ACM* 2009;**52**:80–88.

5. Milde F, Bergdorf M, Koumoutsakos P. A hybrid model for three-dimensional simulations of sprouting angiogenesis. *Biophys J* 2008;**95**:3146–60.

6. Schiff JL. *Cellular Automata: A Discrete View of the World 2008*. Hoboken, NJ, USA: Wiley-Interscience, 2008.

7. Azuaje F, Devaux Y, Wagner DR. Identification of potential targets in biological signalling systems through network perturbation analysis. *Biosystems* 2010;**100**:55–64.

8. Merks RM, Glazier JA. Dynamic mechanisms of blood vessel growth. *Nonlinearity* 2006;**19**:C1–10.

9. Walker DC, Southgate J. The virtual cell – a candidate co-ordinator for 'middle-out' modelling of biological systems. *Brief Bioinform* 2009;**10**:450–61.

10. Mantzaris NV, Webb S, Othmer HG. Mathematical modeling of tumor-induced angiogenesis. *J Math Biol* 2004;**49**:111–87.

11. Goldenfeld N, Kadanoff LP. Simple lessons from complexity. *Science* 1999;**284**:87–9.

12. Alber MS, Kiskowski MA, Glazier JA, *et al*. On cellular automaton approaches to modeling biological cells. In:

13. Rosenthal J, Gilliam DS, (eds). *Mathematical Systems Theory in Biology Communication and Finance*. New York: Springer, 2002;1–40.

13. Thorne BC, Bailey AM, Peirce SM. Combining experiments with multi-cell agent-based modeling to study biological tissue patterning. *Brief Bioinform* 2007;**8**:245–57.

14. Walker DC, Hill G, Wood SM, *et al*. Agent-based computational modeling of wounded epithelial cell monolayers. *IEEE Trans Nanobiosci* 2004;**3**:153–63.

15. Abbott RG, Forrest S, Pienta KJ. Simulating the hallmarks of cancer. *Artif Life* 2006;**12**:617–34.

16. Peirce SM. Computational and mathematical modeling of angiogenesis. *Microcirculation* 2008;**15**:739–51.

17. Glazier JA, Graner F. Simulation of the differential adhesion driven rearrangement of biological cells. *Phys Rev* 1993;**E47**:2128–54.

18. Savill NJ, Merks RMH. The cellular Potts model in biomedicine. In: Anderson ARA, *et al*., (eds). *Single-Cell-Based Models in Biology and Medicine*. Basel, Switzerland: Birkhaüser, 2007;137–150.

19. Merks RMH, Newman SA, Glazier JA. Cell-Oriented modeling of in vitro capillary development. *Lecture Notes Comp Sci* 2004;**3305**(2004):425–34.

20. Merks RMH, Koolwijk P. Modeling Morphogenesis in silico and in vitro: towards quantitative, predictive, cell-based modeling. *Math Model Nat Phenom* 2009;**4**:149–71.

21. Prusinkiewicz P. Paradigms of pattern formation: towards a computational theory of morphogenesis. In: Carbone A, *et al*., (eds). *Pattern Formation in Biology, Vision, and Dynamics*. Singapore: World Scientific Publishing Company, 2002;91–5.

22. Carmeliet P. Angiogenesis in health and disease. *Nat Med* 2003;**9**:653–60.

23. Cao Y. Positive and negative modulation of angiogenesis by VEGFR1 ligands. *Sci Signal* 2009;**2**:re1.

24. Yancopoulos GD, Davis S, Gale NW, *et al*. Vascular-specific growth factors and blood vessel formation. *Nature* 2000;**407**:242–8.

25. Sun S, Wheeler MF, Obeyesekere M, *et al*. A deterministic model of growth factor–induced angiogenesis. *Bull Math Biol* 2005;**67**:313–37.

26. Anderson ARA, Chaplain MAJ. Continuous and discrete mathematical models of tumor-induced angiogenesis. *Bull Math Biol* 1998;**60**:857–99.

27. Stokes CL, Lauffenburger DA. Analysis of the roles of microvessel endothelial cell random motility and chemotaxis in angiogenesis. *J Theor Biol* 1991;**152**:377–403.

28. Jones PF, Sleeman BD. Angiogenesis - understanding the mathematical challenge. *Angiogenesis* 2006;**9**:127–38.

29. Peirce SM, Van Gieson EJ, Skalak TC. Multicellular simulation predicts microvascular patterning and in silico tissue assembly. *FASEB J* 2004;**18**:731–3.

30. Merks RMH, Brodsky SV, Goligorksy MS, *et al*. Cell elongation is key to in silico replication of in vitro vasculogenesis and subsequent remodeling. *Dev Biol* 2006;**289**:44–54.

31. Walker GM, Sai J, Richmond A, *et al*. Effects of flow and diffusion on chemotaxis studies in a microfabricated gradient generator. *Lab Chip* 2005;**5**:611–8.

32. Merks RM, Perryn ED, Shirinifard A, *et al*. Contact-inhibited chemotaxis in de novo and sprouting blood-vessel growth. *PLoS Comput Biol* 2008;**4**:e1000163.

33. Topa P. Dynamically reorganising vascular networks modelled using cellular automata approach. *Lecture Notes Comp Sci* 2008;**5191**(2009):494–9.

34. Aittokallio T, Schwikowski B. Graph-based methods for analysing networks in cell biology. *Brief Bioinform* 2006;**7**: 243–55.

35. Goryczka S, Arodz T. Complex-network-based methodology for analysis of biomedical data. *Bio-algorithms Med-systems* 2006;**2**:19–26.

36. Gevertz JL, Torquato S. Modeling the effects of vasculature evolution on early brain tumor growth. *J Theor Biol* 2006; **243**:517–31.

37. Richmond P, Walker D, Coakley S, *et al*. High performance cellular level agent-based simulation with FLAME for the GPU. *Brief Bioinform* 2010;**11**:334–347.

38. Adra S, Sun T, MacNeil S, *et al*. Development of a three dimensional multiscale computational model of the human epidermis. *PLoS One* 2010;**5**:e8511.

39. Hoops S, Sahle S, Gauges R, *et al*. Copasi- a complex pathway simulator. *Bioinformatics* 2006;**22**:3067–74.

40. Coakley S. Formal software architecture for agent-based modelling in biology. PhD thesis, Department of Computer Science, University of Sheffield, Sheffield, UK, 2007.

41. Grant MR, Mostov KE, Tlsty TD, *et al*. Simulating properties of in vitro epithelial cell morphogenesis. *PLoS Comput Biol* 2006;**2**:e129.

42. Galvão V, Miranda JG, Ribeiro-dos-Santos R. Development of a two-dimensional agent-based model for chronic chagasic cardiomyopathy after stem cell transplantation. *Bioinformatics* 2008;**24**:2051–6.

43. Higuchi ML, Benvenuti LA, Martins-Reis M, *et al*. Pathophysiology of the heart in Chagas' disease: current status and new developments. *Cardiovasc Res* 2003;**60**: 96–107.

44. Ermentrout GB, Edelstein-Keshet L. Cellular automata approaches to biological modeling. *J Theor Biol* 1993;**160**: 97–133.

45. Noble D. *The Music of Life: Biology Beyond the Genome*. Oxford: Oxford University Press, 2006.