Modelling emergent order: from individual cell to tissue

R Smallwood[1*], M Holcombe[2], D Walker[3], D R Hose[4], Steven Wood[5], S Mac Neil[6] , J Southgate[7].

[1]Professor Rod Smallwood, Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield S1 4DP. *Corresponding author. r.smallwood@shef.ac.uk. Telephone +44 114 222 1840, fax +44 114 271 3403.
[2]Professor Mike Holcombe, Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield S1 4DP
[3]Dr Dawn Walker, Medical Physics & Engineering, University of Sheffield, Royal Hallamshire Hospital, Sheffield S10 2JF
[4]Dr D R Hose, Medical Physics & Engineering, University of Sheffield, Royal Hallamshire Hospital, Sheffield S10 2JF
[5]Dr Steven Wood, Medical Physics & Engineering, University of Sheffield, Royal Hallamshire Hospital, Sheffield S10 2JF
[6]Professor Sheila Mac Neil, Department of Engineering Materials, University of Sheffield, Sir Robert Hadfield Building, Mappin Street, Sheffield S1 3JD
[7]Professor Jenny Southgate, Jack Birch Unit of Molecular Carcinogenesis, Department of Biology, University of York, York YO10 5YW

Abstract

The emergence of order from highly complex systems without an over-arching explicit control structure is a fundamental feature of biological processes. In this paper, we start by developing a model of a single cell organism, the rice fungus *Magnaporthe grisea*, and show how gene activity can be introduced as part of the model. We then develop a modelling paradigm for describing the development of a multi-cellular epithelial tissue, from the interaction of individual cells, which combines rule-based software agents and a physical model. We examine how information from genomics, proteomics, and molecular and cell biology in general could be incorporated into the model, and the problems associated with relating cellular-scale and tissue-scale properties. Finally, we consider how the computational model can be closely coupled to *in vitro* biological models of epithelial tissue, providing a one-to-one mapping between *in virtuo* and *in vitro* worlds.

Modelling emergent order: from individual cell to tissue

R Smallwood, M Holcombe, D Walker, D R Hose, Steven Wood, S Mac Neil, J Southgate.

1 Introduction

Models of epithelial tissue either use postulated mechanisms in order to mimic cell growth (illustrative models) or build on known properties of the cells (explanatory models). The last three decades have seen the development of a number of models examining various aspects of cell culture and tissue behaviour. Models have varied considerably in terms of implementation and underlying concepts and assumptions, but have tended to increase in size and complexity in parallel with improvements in computer processing speed and capacity. Honda (1983) developed a graphical approach based on Dirichlet domains to model changes in 2D cell aggregates during morphogenesis. Ransom and Matela (1984) used Voronoi graphs to simulate colony growth. Odell *et al*. (1981) showed that the tightening of microfilament bundles at the apical surface of cells in response to mechanical stress was sufficient to explain the folding characteristic of many embryonic tissues. Computational modelling has been used to examine the validity of hypotheses proposed to explain morphogenic processes, most notably the Differential Adhesion Hypothesis proposed by Steinberg (1970). Glazier and Graner (1993) used a model based on the minimisation of surface energy to demonstrate that differential adhesion with fluctuations was sufficient to explain sorting in cell aggregates, without the need to consider active migration mechanisms.

Individual biological cells in a tissue or cell culture can be modelled as cellular automata, thus enabling the execution of rule sets according to the internal properties or parameters of each cell, and possibly its environment. Lim and Davies (1990) used a combination of cellular automata and Voronoi graphs to examine the growth rate and shape of cell clusters, with stochastic cell growth and division and cell death. Both Zygourakis *et al* (1991a,b) and Forestell *et al* (1992) used a purely automaton-based approach allowing stochastic selection of division. An important rule incorporated into these early automaton-type models is that of contact inhibition. Ruaan *et al.* (1993) introduced an extra element of complexity in considering the tendency of cells to spread if sufficient space is available, and regain a spherical shape and hence generate additional space for growth and division at higher cell densities. Cells were given the ability to migrate in order to find space for growth and division. The idea of random migration was further developed by Lee *et al.* (1995).

In addition to providing an insight into patterns observed in tissue culture growth, modelling has been used as a tool to simulate the formation of more complex tissue structures. Ryder et al (1999) simulated the development of the human cerebral cortex based on random migration, and the difference in cell cycle characteristics observed in cells of different ages. Stekel et al (1995) developed a model of the morphogenesis and homeostasis of the human epidermis. This is an illustrative model, with rules formulated to simulate observed behaviour, and not on the basis of well understood mechanistic behaviour of individual cells. For instance, one of the rules stipulates that stem cells emit a substance called ‘stem cell factor’, the concentration of which can be sensed by other cells in the model. Morel et al (2001) incorporated two distinct hierarchies in their model of epithelial tissue: a kinematic model of cell cycle regulation, incorporating both intracellular components (e.g. cyclins) and response to extracellular stimuli (e.g. growth factors); and a Voronoi graph-based tissue architecture model, with individual cells represented by polygons. In addition to models of development and behaviour associated with specific tissues such as epithelium or cortex, general embryonic morphogenesis has continued to be a field of active research. Hogeweg (2000) built on a more basic earlier model (Savill and Hogeweg, 1997) to construct a two level hierarchical simulation of the effect of morphogenic evolution on cell differentiation and differential adhesion.

One major factor absent in all the models discussed so far is the explicit consideration of forces acting on cells, and the resulting deformation and movement. In general, models that attempt to simulate the effect of physical forces are continuum, rather than automaton or agent based. Examples include the work of Brodland and Chen (2000) and Chen and Brodland (2000) who used finite element models to simulate various morphogenic process, such as stretching and engulfment, for confluent sheets consisting of a pre-determined number of cells. Interestingly, in contrast with the earlier energy-based models of Glazier and Graner (1993), the results of these

simulations suggested that differential adhesion alone is not sufficient to result in the sorting of a heterotypic cell population.

Palsson (2001) proposed a model which simulates three dimensional morphogenic processes using an automaton rather than continuum based approach. Cells in this model are considered as individual entities that can respond to the environment according to the values of their internal parameters, and physically interact via contact forces. Each cell has viscoelastic properties, and moves and deforms according to the equations of motions and deformation. The capacity exists to assign different properties or parameters according to designated cell type. Simulations produced using this model suggest that tissue structure and cell sorting can arise from differences in cell adhesion properties, and movement in response to a chemotactic gradient. Three dimensional embryonic models have been produced consisting of up to 10,000 'cells', each representing 4-16 actual biological cells. In common with many of the previous simulations of morphogenic behaviour, this model does not include the capacity for cell division and differentiation.

In this paper, we start by developing a model of a single cell organism, the rice fungus *Magnaporthe grisea*, and show how gene activity can be introduced as part of the model. We then develop a modelling paradigm for describing the development of a multi-cellular epithelial tissue, from the interaction of individual cells, which combines rule-based software agents and a physical model. We examine how information from genomics, proteomics, and molecular and cell biology in general could be incorporated into the model, and the problems associated with relating cellular-scale and tissue-scale properties. Finally, we consider how the computational model can be closely coupled to *in vitro* biological models of epithelial tissue, providing a one-to-one mapping between *in virtuo* and *in vitro* worlds.

2 Why is cellular interaction important

What is considered to be fundamental depends on perspective, but few would argue that the transition from single cell organisms to multi-cellular organisms is not fundamental to the development of complex life forms. Individual cells are autonomous entities containing the machinery for converting energy and substrates into more complex molecules and structures, for self-replication and differentiation, cell response and even self-destruction (apoptosis or programmed cell death). The structure and function of a multi-cellular organism is a result of the interaction of these autonomous cells. There is no external plan for the structure – no *deus ex machina.* There is a one-to-many mapping between genes and proteins, so the genome cannot of itself provide a plan for the structure of an organism. If we want to understand the mechanisms of formation of normal tissue, of wound healing, of progression to malignancy, we need to understand how the interaction of autonomous cells can produce a social entity, the tissue. Our current understanding is at a qualitative (descriptive) level, which differs between tissues, but nevertheless is very detailed even at the molecular level. For example, there is much known about adhesion molecules, such as the cadherins, that bind cells together to maintain tissue integrity and whose presence or absence can determine whether a tumour cell is benign or malignant.

The building of functional structures from autonomous cells is central to human biology, from embryology and development, to repair and even the development of cancer. We have started with epithelial tissue because it is (relatively) simple, consisting of only one cell type, is typically only a few cells (~500 μm) thick, there are good *in vitro* models available, and there are important clinical problems (e.g. epithelial cancers account for more than 90% of all adult malignancies and chronic non-healing skin wounds account for approximately 12% of the National Health Service budget in the UK. The *in vitro* biological models that are being used to develop our computational modelling encompass 2D and 3D models of uroepithelium (Southgate *et al*, 2002; Scriven *et al*, 1997), oral mucosa (Bhargarv *et al*, 2003) and skin (Chakrabarty *et al*, 1999; Ralston *et al*, 1999).

3 Convergence of autonomous software and cellular behaviour

Our approach is based on the language of computational models. There are many different types of computational model with varying processing capabilities which differ, mainly, in the way that they are represented. We choose to use a powerful model based on three key components; internal state, internal memory and environmental interaction, since these seem to be the fundamental components of any reasonable model of a living system at a variety of levels. One difficulty with classical theories of computation is that they are essentially discrete so that events occur in discrete time steps and there is no concept of continuum, such as processes of metabolism which proceed over a time period in a continuous manner. We propose to exploit a

powerful computational approach based on communicating X-machines (Balanescu *et al*, 1999). X-machines are examples of discrete computational models that operate in finite environments (finite input sets, finite output sets and finite memory variables). Firstly we identify an X-machine (Holcombe and Ipate, 1998; Kefalas *et al*, 2003b) as a system which has internal states and an internal memory. The state transition functions will respond to events on the basis of both the environmental input as well as the current internal state. The system is in some state, an input **a** is received, the initial contents of the memory are **m** and, depending on both **a** and **m**, the system changes state and produces an output **x** and updates the memory to **m′**. This provides a general modelling mechanism which enables many of the problems associated with state explosion, which bedevil many efforts at modelling complex biological systems, to be dealt with sensibly. The memory can be used to abstract away detail in a way that does not prevent us from utilising it whenever necessary. These machine models are suitable for modelling many types of system. However, they can only model instantaneous processing and hence only finite discrete data is processed. Continuous functions and real valued data cannot be incorporated into traditional finite state machine models. Such systems are problematic when trying to deal with the complexities of some biological models and the hybrid X-machine (Holcombe *et al*, 2003) overcomes some of these problems.

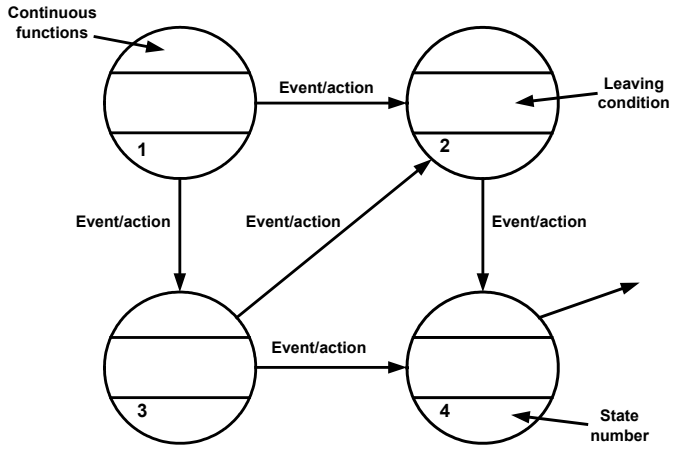

Figure 1. A hybrid machine

A hybrid machine has states and transitions as usual and responds to discrete events and performs discrete actions which are observable. The internal memory consists of a set of discrete variables and a set of continuous variables. The way the model works is that we identify a number of significant states of the system each of which describes some aspect of the system's behaviour. At certain moments the system changes state and moves to another state where different activities occur. The events and conditions that prompt state changes are explicitly identified, as are the sorts of processing that goes on in each state. All of these make use of the identified variables, both continuous and discrete and are expressed in terms of mathematical equations and properties. A key aspect of the hybrid machine is the memory, a collection of variables, continuous and discrete which help to model the cellular metabolism and structure. An important component of the memory is a description of the internal structure of the cells. When the system is in a given state there are sets of equations that apply to the system's continuous variables and all the while it is in that state, with time progressing, these variables change according to these equations. When either an appropriate external event occurs or a leaving condition is met (e.g. a set point) the system moves to its next state where a different set of equations take over. This sort of generic model can model many types of biological phenomena but needs refining to deal with some important issues.

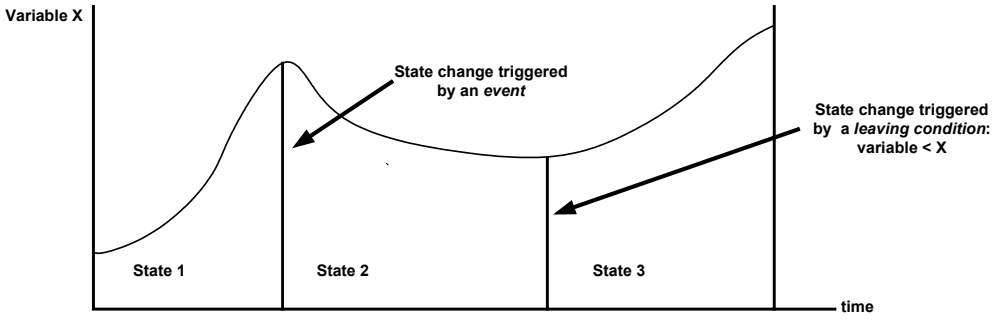

Figure 2: continuous variable in a hybrid machine

The overall system is then described by using a modified state transition diagram, see Figure 1, which describes the main states of the system; the transitions between these states; the equations pertaining to each state; the events which cause a state change - this could be either an external event or an internal leaving condition; and the results of the transition which will affect either internal discrete variables or external properties of the system. In some cases the transitions will be

prompted by signals derived either internally or by the communication between one part or component of the system, or by some external signal perceived in interaction with the system's environment, such as other cells in the organism or signals from outside the organism. Some simple examples that can be modelled this way include ion flow through voltage-gated channels and antigen-antibody interactions. The continuous variables can exhibit complex behaviour as shown in Figure 2.

The equations are often composed of relatively simple functions compared to the equations that try to describe the complete functions over all states. This is an advantage of a state-based approach and can be exploited in a variety of ways. Furthermore, these functions might be decomposable into lower level hybrid machines thus providing a mechanism for dealing with the undoubted complexity that such systems exhibit.

4 A single cell model – *Magnaporthe grisea*

Hybrid X-machines have been used to model simple organisms down to the level of gene activity (*Holcombe et al* 2003). *Magnaporthe grisea* exists in the atmosphere prior to infection as a spore which alights upon a leaf or other surface and undergoes a series of development processes which are determined by the genetic structure of the organism and interactions with its environment. The spore or *conidium* is a 3 celled structure, fig 3, which is present in the atmosphere in affected areas. These alight on the surface of rice leaves, normally contained within a dew drop, and attach themselves to the surface. This is possible despite the fact that the leaf surface is highly hydrophobic and is achieved by the conidium releasing from its tip a powerful adhesive stimulated by wetting.

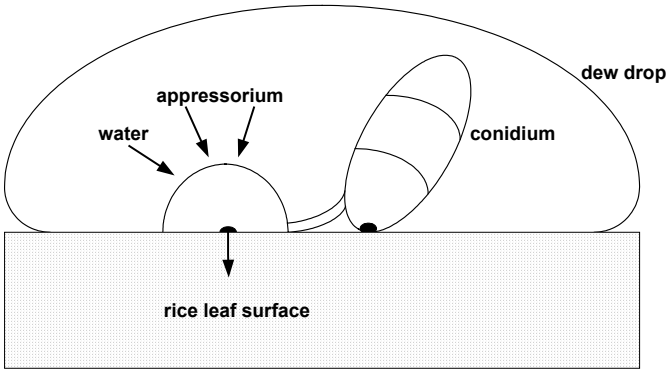

Figure 3: the rice blast fungus

The spore then germinates and produces a germ tube from one of the terminal cells of the conidium which then forms a hook and adheres to the leaf. An *appressorium* of a roughly hemispherical shape then develops at this point of contact. The penetration of the rice leaf surface is carried out by the build up of pressure within the appressorium. The pressures generated are as high as 8 MPa. As a result of this enormous pressure a penetration peg, formed at the part of the appressorium where it joins the leaf surface, is forced to penetrate the leaf surface. Once penetration has been achieved the fungus forms cylindrical cells called hyphae which initially spread into plant cells without causing damage or overt disease symptoms, but later produce toxic compounds and degradative enzymes which cause plant cell death. The fungus causes necrotic disease lesions on leaves which can be seen as dark oval spots. Each spot represents the point of a single appressorium-mediated infection and when they coalesce in heavy infection, whole leaves or entire seedlings can die (Talbot and Foster, 2001). The fungus produces conidia from disease lesions which propagate the fungus to new plants (Talbot, 1995). Figure 4 provides a top level hybrid machine model of the life history of the fungus. The internal variables (or memory) provide information about the status of various internal aspects of the fungus, for example the initial concentration of glycogen, glycerol etc. have to be modelled as variables. The build up of glycerol is thought to be responsible for the generation of appressorial turgor. One potential source of this glycerol is from the breakdown of glycogen. So we need to describe the internal turgor pressure of the appressorium generated by the concentration of glycerol. Each state has either a leaving condition, some condition that has to be satisfied by some internal parameter in order for a state change to occur, or there is some external event that triggers the state change.

When the spore lands on the leaf, the release of its glue to attach to the surface is a passive process caused simply by the presence of water. This triggers germination and germ tube development. The break of dormancy is a genetic signal, but derives from a passive external effect (presence of water). The dormancy breaking signal would trigger gene expression involved with polar growth and further adhesion. Some state changes will be triggered by the recognition of some intracellular signal and the switching on and/or off of some specific genes

which then become active in the new state and control the metabolism until the next state transition. In the diagram we indicate this thus: *signal4 / MAGB, MAC1, PMK1,* where *signal4* is some internal signal indicating that the hook formation process has completed and the transcription factors encoded by the genes *MAGB, MAC1, PMK1* are activated. These genes are probably always expressed at low levels, but activation of the transcription factors results in de novo induction of genes required for movement to stages 3 through to 6. A state transition involving a leaving condition might be state 7 → state 8 where the leaving condition is the internal appressorium pressure reaching a set point sufficient to cause the puncturing of the host leaf surface. Whilst in state 7 this pressure is undergoing change mediated by some suitable set of equations determined by the inflow of water through the appressorium surface. If this surface is regarded as a hemisphere and the porosity of the surface is constant throughout then the rate of increase in mass, the pressure is given by the linear equation: $p(t) = \rho t/r + initial\ pressure$ [equation_7]. During state 7, then, the turgor pressure is increasing linearly, assuming that there is sufficient water surrounding the appressorium, until the set point is reached at which time the transition to state 8 occurs, thus triggering further gene activation for the next phase of development. In state 7 there are a number of important processes which determine the production of glycerol. Genes involved in the breakdown of lipids, glycogen and trehalose (a fungal storage carbohydrate) are all thought likely to be involved in the production of glycerol. For example, the genes *GPH1* which encodes glycogen phosphorylase and *AGL1* which encodes an amyloglucosidase are involved in breaking down glycogen into glucose units which can then be used to make glycerol. The equations operating during state 7 (equations_7) will describe the metabolic activity relating to glycerol production. Stage 7 may require a different signalling pathway, perhaps also involving cAMP. GPH1 and AGL1 are targets of this pathway, rather than effectors. They bring about the change in metabolic state however, acting in concert with a number of other enzymes that leads to state 7.

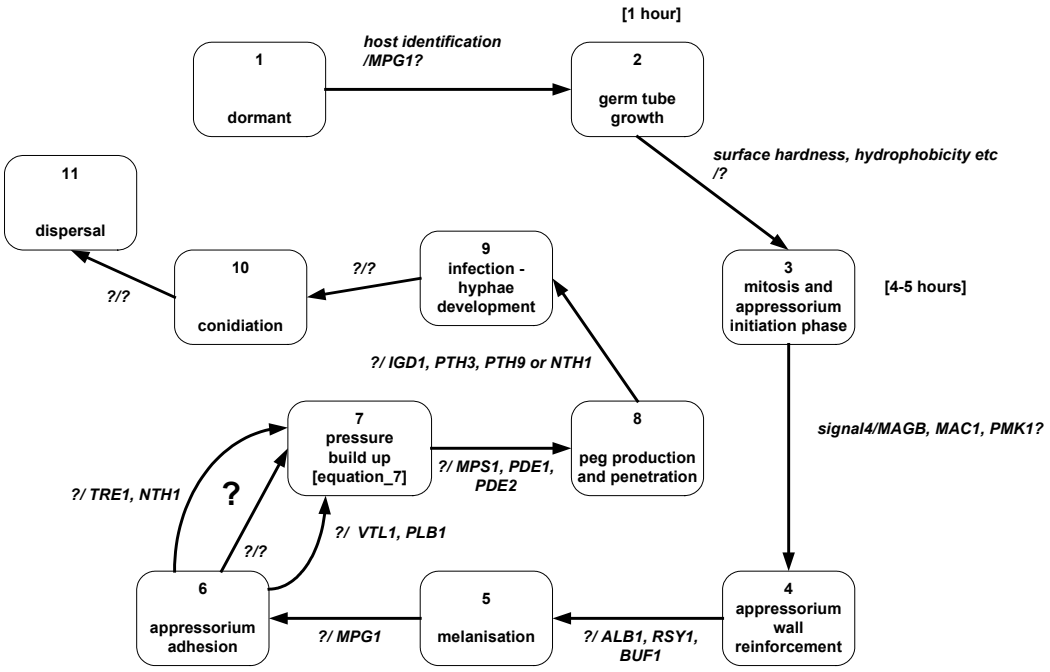

Figure 4: a hybrid machine model of *Magnaporthe grisea*

5 Communication in multicellular structures.

A simple organism such as *M. grisea* engages in communication principally with its environment and the internal communication is fairly rudimentary. Higher organisms, however, exhibit highly sophisticated communication capabilities and to examine these we need to extend the basic model to include communication. In particular, in the context of the self-assembly of tissues, we need to include explicit chemical signalling between neighbouring cells, chemical signalling within cells, and transduction of mechanical forces exerted by neighbouring cells into chemical signals. Self-assembly in biological systems appears to be based on local signalling and local processing capabilities from which the overall system behaviour emerges, rather than being organised as some high level control system which determines what each component, a cell in this case, is to do under the specific circumstances pertaining at that moment. The concept of an "autonomous" agent is one which

is useful in this respect. A number of types of biological systems have been modelled in this way (Georghe *et al*, 2001), including communities of various social insects such as species of ants and bees. Each insect is considered as one of these agents and can behave independently of any explicit external instructions. If societies of individual insects can be modelled in this way then it is worth considering how societies of cells might fare under a similar modelling paradigm. The metaphor that we are investigating is based around the concept of a cell as an agent. In order to do this in a way that can be exploited both in terms of simulation but also using promising approaches to the automated analysis of complex models we need to conceptualise the agent model suitably. In the context of self-assembly of tissue, we also need to introduce links to both the underlying biological mechanisms and the physical forces which determine cell motility and organisation. For example, oral mucosa and skin epithelial cells in 3D structures rapidly cease horizontal migration and assemble into a multi-layered epithelium with varying degrees of differentiation. Experimental investigation shows that the nature of the underlying substratum (e.g. Ralston *et al* 1999) and the introduction of a physical stimulus in the form of an air-liquid interface act as strong inducers of differentiation, but the degree of differentiation achieved is dictated by the genes expressed in the cells themselves (Bhargarv *et al* 2003). The X-machine makes a natural candidate for modelling an agent (Kefalas *et al* 2003a). We start with a simple set of rules which describe what the agent must/could do under various different circumstances. The set of rules may be defined with an explicit prioritisation that determines which rule is to be used under which environmental and internal conditions. Thus, perhaps, the top rule provides a general metabolic processing activity typical of the cell's normal state of activity but on the receipt of some event in the cell's immediate environment such as a signal from a neighbouring cell or some external process the cell undergoes a change in activity which is reflected in a new set of metabolic processing activities captured in a new set of rules.

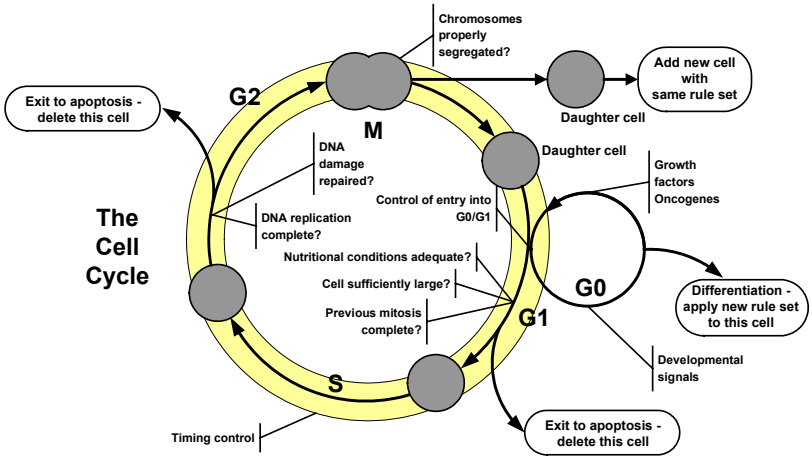

Figure 5: a simplified representation of the cell cycle

The starting point for modelling a single cell as an agent is the cell cycle, a simplified version of which is shown in figure 5. DNA is synthesised in the S phase, and other cellular macromolecules are synthesised throughout the G1, S, and G2 phases, leading to the cell roughly doubling in size before mitosis (M, cell division). The preparation for mitosis takes place during the G2 phase. Non-dividing cells enter the G0 or quiescent phase. Within the cell cycle there are checkpoints (denoted by questions in the diagram), which have to be successfully passed for the cell cycle to proceed – otherwise the cell proceeds to apoptosis (programmed cell death). For a single cell, provided with adequate nutrients, a rule set can be developed from the diagram and combined with typical times for each phase, to give the top level in a hierarchical model of the cell. If the cells can leave the stem cell pool – e.g. by differentiation (e.g. in skin and oral mucosa stem cells produce transit amplifying cells), then a differentiation rule is also required – differentiation will change the rule set for the cell in some way. A basic rule set has been presented by Walker *et al* (2003), which does not include either metabolic events or prioritisation of rules at this stage of development. The detailed rule set may well prove different for epithelial cells of different origin – thus, under conditions of physiological calcium, keratinocytes can become relatively differentiated in both 2D and 3D models *in vitro* to a much greater extent than oral mucosal epithelial cells or uroepithelial cells. This can, in principle, be extended to include every process within the cell.

6 Rules and mechanisms

The rules provide a qualitative description of events within the cell, but the rules are the result of the operation of a mechanism. For instance, if we are considering metabolic activity within the cell, we may have a rule of the

form {if nutritional conditions adequate, then …., else ….}. Underlying this qualitative rule are the biochemical pathways which produce the required proteins; diffusion of nutrients through the surrounding medium; transport across the cell membrane; etc. So, in principle, it is possible to model the mechanisms which determine the output state of the rules. Mechanisms will be known to a greater or lesser extent, and can replace qualitative rules as more knowledge of the system becomes available. For instance, establishing adequate nutritional conditions may require a rule {if [substrate x] > y, then …., else …..}, where the mechanism of manufacturing x, or the method by which its concentration y is measured, are both unknown. If this proves to be a critical path in the model, then the mechanism will have to be determined experimentally. The model thus acts as a driver for experiment.

When there is more than one cell present, the interaction between cells has to be introduced. The cells 'communicate' (intercellular signalling), and the tissue as a whole (the continuum) 'communicates' with individual cells through both chemical diffusion through the tissue and through mechano-transduction (mechanical strain eliciting chemical signals within the cell), and, possibly more importantly, integrin-mediated bonding. The cells adhere to each other and can also be actively motile (in contrast to passive movement due to the physical forces between adhesion molecules). In the agent-based model, the agents access a common memory area. In order to model this type of system we need to use some form of communication broker that reflects how different agents can send and receive their messages. This is done by using a sparse matrix. Suppose that we have a collection of agents, in this case represented as individual X-machines. We now have to try to identify the communication channels and how these might work. Suppose that there are N agents and each is potentially able to communicate with any other. We thus have an N x N communication matrix. Agent i can send a message to agent j by placing the message in the (i, j)th place in the matrix during a write communication process. When agent j is in a state to read the message from agent i it reads the message from this position in the matrix. Since most agents will only communicate with their immediate neighbours the matrix will be sparse with only a small number of valid slots and with no diagonal elements.

Suppose that an agent can only communicate with other agents within a specific distance of it. There is a global memory that maintains the current position of each agent (an N-vector of co-ordinates). At any communication state an agent can interrogate this memory to ascertain which other agents are within communication distance. A number of strategies can be used to determine which and how many attempts at a communication can be made. The agent then puts data into the appropriate slots of the communication matrix and continues processing, moving or whatever.

The key thing is to define the rules of the agent including the inter-agent communication mechanism, to examine the complete model to see what emergent behaviour is produced, and to verify by comparison with the 'real thing' – in this case, the cell culture systems.

7 The cell as software agent

The proof-of-concept model consists of a number of cells interacting with their environment, which comprises a two-dimensional, square substrate with user-defined dimensions and modifiable exogenous calcium ion concentration. The purpose of this stage of the model is to test the basic concept of cell-as-software agent, and the link between the agent model and a physical model of the world. The user can select the number of cells seeded and whether to place them randomly on the substrate, or in specified locations. The seed cell radii and epithelial type can also be selected by the user. All seeded cells are designated to be members of the stem or transit amplifying (TA) cell classes. Model execution is based on an iterative process, with each tick or agent model iteration representing approximately 30 minutes in real time. This was achieved by selecting mean cell cycle duration and migration rates, so that the cell cycle step, or distance migrated by cells at each time step is, on average, similar to that undergone by a real cell during this time period. These data were based on a mean cell cycle duration of 60 hours for keratinocytes (Dover, 1983) and 15 hours for urothelial cells (Southgate, 1994) and a migration speed of approximately 1μm/ minute, as measured from time lapse sequences of urothelial cell culture. These parameters would be expected to vary as a result of culture conditions as well as cell type.

At each tick of the model clock, each cell is interrogated in turn, and depending on the state of its internal parameters (e.g. current position in cell cycle, flag indicating bond to substrate) and its environment (number and proximity of neighbouring cells, concentration of calcium ions) a number of rules are executed that may change the state of the cell's internal parameters. During this process, cells can receive messages from other cells in their immediate vicinity, or send messages to other cells by reading from and writing to a structure designated as a

communication matrix. An additional global data structure, which can be accessed by all cells in the model, contains information relating to the exact position and dimensions of every cell. Each cell knows the location (distance and azimuth) of all neighbours which could come into contact with it (by growing or migrating) within one time step. This distance is approximately 10*(initial cell radius) = 100 μm (cells can grow to 30 μm radius and migrate up to 30 μm per time step). This information is updated every time a cell moves or changes shape. If a cell divides during a model iteration, one daughter cell overwrites the data structure of the parent, whilst the second daughter is added to end of the string of existing cell structures. Further details of rule sets pertaining to particular cell behaviours are given below.

Each cell in our model is represented as assigned to a particular object class according to its designated cell type. The current permitted types are stem cells (specifically epithelial stem cells); transit amplifying cells; mitotic cells; post-mitotic cells; dead cells. Stem cells are predicted to exist in many types of epithelial tissue, although a definitive biochemical/antigenic marker has yet to be found. The fundamental property of such cells is that they have an unlimited capacity for self-renewal, and thus are critical in maintaining a constant cell number in normal homeostatic tissue. There is some evidence that in certain types of epithelium, stem cells may have a particular spatial distribution, may cycle more slowly or may express different integrins and thus exhibit distinct contact properties to transit amplifying cells (Jensen, 1999; Potten, 2002). We have currently ignored such characteristics, some of which remain speculative, and restrict our stem cells to differ from TA cells in terms of their self-renewal properties only. Ten percent of the initial seed cells are designated to be stem cells (Potten, 1988). The permitted behaviour set for this cell type are: bonding (to substrate and other cells); spreading (once bonded to substrate); lateral migration (once bonded to substrate); apoptosis (if failure to bond to substrate); cell growth and division. Transit Amplifying (TA) cells are designated to form approximately 90% of the basal compartment of epithelial tissue. Unlike stem cells, they are capable of only a limited number of division cycles before becoming committed to terminal differentiation. The number of rounds of division allowed appears to vary between different epithelial tissue types. For instance, as previously stated, it is not unusual for keratinocytes (Chakrabarty *et al*, 1999) or oral mucosa cells (Bhargarv *et al*, 2003) in 3D culture to express some cytokeratin isotypes associated with differentiation, but this is not observed in the culture of urothelial cells (Southgate *et al*, 1994). It is, of course, quite possible that this disparity can be accounted for by the different conditions used to culture these cell types, in contrast to being hard-coded for each cell type at an epi-genetic level. However, whilst this point remains in doubt, the number of TA cell divisions allowed are limited to 3 for keratinocyte models and 30 for urothelial cell models. The permitted behaviour set for TA cells is identical to stem cells in all other respects. Mitotic cells are a subset of both stem and TA cells that are progressing through the M phase of the cell cycle. In cell culture, mitotic cells are observed to round-up, thus losing adhesion with their neighbours, and to some extent with the underlying substrate. For this reason, model cells assume a rounded morphology, breaking all intercellular bonds on entering the M phase (defined as the final 20% of the cell cycle) and are not permitted to form new bonds or actively migrate until mitosis is complete. Post-mitotic cells are produced by the division of TA cells that have undergone the maximum permitted number of mitotic cycles. They are rarely observed in monolayer culture of urothelial cells, but are relatively common in keratinocyte cultures which rapidly become enucleated keratin envelopes reflecting their barrier function *in vivo*. The extension of our model to be truly three dimensional will allow these cells to assume the characteristics of cells committed to terminal differentiation, and migrate upwards to reside in the supra-basal layer. Currently, post-mitotic cells are permitted access to the same rule sets as stem and TA cells, with the obvious exception of cell growth and division. Failure of normal cells in culture to adhere to the substrate results in the programmed cell death (apoptosis) of that cell. The dead cells are then fragmented and phagocytosed by their neighbours. Model cells that have failed to bond to the substrate before the end of their G1 phase become dead cells, shrink and are removed from the model. In reality, most cells adhere very quickly after seeding, and hence dead cells are a rare occurrence.

In addition to the internal parameters unique to the designated class of a cell, every cell structure in the model contains a set of parameters that defines the exact location of its centre in x, y and z Cartesian co-ordinates, and the length of its axes in each direction. Hence, although the model remains essentially two dimensional at this stage, each cell is defined as a three dimensional ellipsoid. This allows three dimensional changes in cell morphology, such as cell spreading and rounding, to take place.

The central hinge of the model is the cell cycle. All cells capable of cycling (i.e. stem and TA cells) have an internal clock, and provided that they are not inhibited, progress one tick through the cycle at each model iteration. At model initiation, or when mitosis occurs to create new daughter cells, the new TA cells are

designated a fixed S-G2-M phase length comprising approximately 50% of the total cell cycle duration according to the model cell type (keratinocytes = 60 hours or 120 model iterations, urothelial cells = 15 hours or 30 iterations) and a G1 phase length that is selected randomly from a normal distribution with mean equal to half the total cycle length and standard deviation equal to a tenth of the mean. This reflects the situation that is observed in real cells where the variation in the cell cycle duration between cells of the same lineage is derived almost entirely from variation in G1 duration.

The G1 phase of the cell cycle is the growth phase during which cells effectively double their volume by synthesis of proteins within the cytoplasm. At the first tick of the G1 phase of a model cell, the cell's volume is calculated along with the volume increase required at each subsequent tick in order that the volume is exactly doubled at the end of G1. This volume increment is added to the cell at each G1 tick.

Approximately half way through G1, a checkpoint has been introduced that, at present, governs whether a cell proceeds through the remainder of the cell cycle based on the number of contacts the cell has formed with neighbouring cells, and its morphology. In the culture of several epithelial cell types, cells situated at the centre of cell colonies grown in physiological or high calcium media withdraw from the cell cycle and become quiescent. It is believed that this phenomenon, termed 'contact inhibition' is mediated via intercellular signalling – cells are aware of the existence of their nearest neighbours via the cadherin-mediated, and hence calcium-dependent physical bonds (St Croix *et al* 1998; Orford and Byers 1999). The rule for a model cell to become contact-inhibited is that it has four or more bonds with neighbouring cells (or, alternatively three bonds and be adjacent to the 'edge' of the substrate). Cells require a spread morphology in order to progress further in the cell cycle. For this reason, model cells also check that they are bonded to the substrate and sufficiently spread at this checkpoint. Cells which have not attached at this stage undergo apoptosis.

Cells which do not fulfil the criteria to pass through the G1 checkpoint enter a quiescent, or G0 phase of the cell cycle. The cell will remain in this state until enough of its intercellular bonds are broken to allow it to re-enter the cycle (this may occur if a neighbouring bonded cell reaches the mitotic phase of it cycle, rounds up, and breaks its intercellular bonds), and the cell has spread sufficiently. A corollary of these rules is that they allow the concept of wound repair to be introduced in a confluent monolayer model.

If a cell passes the checkpoint, or fulfils the requirements to leave the G0 state, it will then progress to the S and G2 phases. Biological cells synthesise new DNA during the S phase, and prepare for mitosis during the second lag phase, or G2. However, as our model does not currently consider sub-cellular phenomena explicitly, during these phases model cells simply increment their cycle clock by a single tick at each model iteration. Further checks relating to phenomena such as occupation of growth factor receptors (G1 phase) and successful DNA replication (G2 phase) will be incorporated at a later stage of model development.

On approaching the final 20% of the cycle, defined to constitute the mitotic or M phase, the cell alters its morphology to become rounded, breaks any bonds it may have with neighbouring cells and is redefined as a member of the mitotic cell class. Cells in this class are defined as two distinct but overlapping spheres, which move apart over the duration of the M phase to produce two adjacent volumes that are eventually designated as two separate daughter cells. The first daughter cell structure overwrites the position of the parent in the model, and the second is added to the end of the existing list of cells. Stem cells always undergo asymmetric division, i.e. produce one daughter that is also a stem cell, and a second that is a transit amplifying cell, whereas transit amplifying cells always produce transit amplifying daughters. New cells, whether seeded onto the substrate, or the product of mitosis, have a rounded morphology and are not bonded to the substrate – i.e. they are not polarised. At each subsequent model iteration, there is a probability-based calculation of the cell binding to the substrate. This calculation is biased to reflect the fact that it is extremely unusual for cells in culture not to adhere within the first hour or so after seeding. Once adherent, a cell is then permitted to engage in a number of cell behaviours, including spreading and migration. The location of all cells in the model is continually updated and stored in a global data structure, which allows every cell to know the relative positions of all its neighbours. For each cell, at every model iteration, a binding probability is calculated for each neighbour within a set distance (defined to be half the specified radius of the original seeded cells, which in the case of the models discussed in this paper is 10μm). The probability of bonding for a pair of cells is inversely proportional to the separation between the cell edges and related to the environmental calcium concentration via a sigmoid function with the inflection point at 1.0mM. This relationship is based on published data (Baumgartner *et al* 2000) on the relationship between the binding activity of E-Cadherin-mediated bonds and exogenous calcium ion concentration. This study used atomic force microscopy techniques to show that an extracellular calcium of

greater than 1.0 mM is required for dimerisation of the E-Cadherin receptor. A random number between 0 and 1 is generated and bonds will be formed with cells whose probability constant is greater than this number. All cells that form intercellular bonds automatically become bound to the substrate, if they are not already attached.

The first cell records the identity flags of the successfully bonded neighbouring cells within its data structure and sends a signal to these neighbours via the global communication matrix. On receiving the signal, a cell adds the flag of the sender to its own data structure. Hence every cell in the model that is able to form bonds will contain a list of flags indicating the cells to which it is linked. Bonds between cells will be recognised when the data is assembled for the next physical correction and non-adjacent linked cells will feel an attractive force proportional to their separation, resulting in them moving together until their edges are adjacent.

Cells seeded onto the substrate and new daughter cells produced by mitosis are not attached to the substrate and are spherical in shape. In order to be able to progress through the cell cycle, real biological cells spread by forming numerous attachments to the substrate, thus assuming a flattened morphology (Huang and Ingber 1999). This process is emulated in the model by increasing the surface area of attached cells by a small amount each model iteration, whilst maintaining a constant volume. Cells continue to spread until their radius has increased by a factor of approximately 1.5. Cells that are attached to the substrate but have no intercellular bonds are free to migrate laterally on the substrate surface. The allowed migration distance per model iteration is fixed at approximately 2.5 times its radius. This figure was obtained from measurement of urothelial cell movement from successive frames of a time-lapse recording of a scrape wound closure assay. Migration directions are randomly assigned to new cells. At each subsequent iteration, a cell will tend to carry on migrating in the same direction, but may randomly choose another direction within 60° of its original trajectory. If the migration path is blocked by a neighbouring cell or the edge of the substrate, the cell will move by a smaller distance so that its edge is adjacent to that of its neighbour. If no movement at all is possible in the chosen direction, the cell will alter its trajectory by 30 degrees and try again, continuing to do so up to 10 times, after which it will remain stationary. Such checks, and similar ones that are executed when selecting the direction of the separation in mitotic cells, for instance, are essential to ensure that cells do not fall off the edge of the model, or overlap one another by more than a single cell radius, as this could lead to errors during the physical correction. Stem and TA cells that have failed to attach to the substrate by the end of their G1 phase are redefined as dead cells. These cells shrink, and are removed from the model.

After each time step of the agent model, the position and size of all the cells is passed to the physical model, which calculates the changes in position which occur as a result of growth, division, and physical interaction, and returns the new cell positions to the agent model for the next time step (figure 6).

8 The physical world

At the current stage of development, the physical realisation of the cells and their interactions is very simplistic. The cells are rigid spheroids which can change the ratio of height to radius depending on the cell cycle stage (they are less disc-like in mitosis). They are confined to a plane i.e. only a monolayer of cells is modelled. Cell growth, movement and mitosis events during each agent-based iteration will result in forces between neighbouring cells. In addition, the edges of neighbouring cells that are bonded to one another may not be contiguous.

The acceleration induced in the *ith* cell, by the force exerted by its *n* neighbours is given by Newton's second law, with damping:

$$m_i \ddot{u}_i + c_i \dot{u}_i = \sum_{j=1}^{n} F \qquad (1.1)$$

where m_i is the mass of the cell, c_i is its damping constant and u is its displacement. The mass of the cell is calculated from the length of its two primary axes *a* and *b*, which is information retained in the computational data structure representing each cell. As the cells only interact in two dimensions, this is based on the planar area of the cell, assuming a unit density:

$$m_i = \pi a^2 \qquad (1.2)$$

The damping constant, c_i is proportional to the mass of the cell:

$$c_i = \alpha m_i \tag{1.3}$$

where the proportionality constant α is assigned according to the global location of the cell. Cells close to the edge of the model are given higher damping factors to inhibit their movement and prevent them being pushed off the edge of the model. Damping factors in the *x* and *y* direction can be manipulated independently, thus a cell may be more easily pushed in one direction than the other.

During each agent iteration, a cell determines the position of all other cells within approximately 10 cell radii. In addition, each cell contains information regarding its own position and radius. When the physical model is assembled, a check is made to identify whether the growth of these nearby cells would cause them to attempt to occupy the same physical space. In addition, each data structure contains a list of the flags of other agents to which it has formed an intercellular bond. Forces are only exerted by neighbours that are bonded or attempt to occupy the same space. The force exerted the cell *i* by neighbour *j* is defined to be proportional to the 'overlap' or edge separation O_{ij} between the cells.

$$F_{ij} = k_{ij} O_{ij} \tag{1.4}$$

where k_{ij} is an arbitrary 'stiffness' defined to be equal to the inverse of the separation of the centres of cells *i* and *j*. If cells *i* and *j* have radii r_i and r_j and are located at $\left(x_i, y_i\right)$ and $\left(x_j, y_j\right)$ respectively, then:

$$O_{ij} = \sqrt{\left(x_j - x_i\right)^2 + \left(y_j - y_i\right)^2} - \left(r_i + r_j\right) \tag{1.5}$$

Note that the signs of the constants are constructed in such a way that cells experience a repulsive force from neighbours attempting to occupy the same space, and an attractive force from bonded neighbours that may be separated by a small distance.

Assuming the vector $i \rightarrow j$ is orientated at angle θ with respect to the *x* axis, resolving into two orthogonal directions and substituting into (1.4) gives:

$$F_{ij(x)} = k_{ij}\left\{\left(x_j - x_i\right) - \left(r_i + r_j\right)\cos\theta\right\} \tag{1.6a}$$

$$F_{ij(y)} = k_{ij}\left\{\left(y_j - y_i\right) - \left(r_i + r_j\right)\sin\theta\right\} \tag{1.6b}$$

The co-ordinates of cell *i* at time *t* are the co-ordinates $\left(x_{Oi}, y_{Oi}\right)$ at the previous time step plus the displacement u_{xi}, u_{yi} during the current time step, so equation 1.6a can be written:

$$F_{ij(x)} = k_{1ij}\left\{\left(x_{Oi} - x_{Oj}\right) + \left(u_{xi} - u_{xj}\right)\right\} - k_{2xij}\left(r_i + r_j\right) \tag{1.7}$$

where k_1 and k_{2x} are constants depending on the location of the two cells:

$$k_{1ij} = \frac{1}{\sqrt{\left(x_j - x_i\right)^2 + \left(y_j - y_i\right)^2}} \tag{1.8a}$$

$$k_{2xij} = \frac{\cos\theta}{\sqrt{\left(x_j - x_i\right)^2 + \left(y_j - y_i\right)^2}} \tag{1.8b}$$

Equation 1.6b can be similarly rewritten, but in this case:

$$k_{2yij} = \frac{\sin\theta}{\sqrt{\left(x_j - x_i\right)^2 + \left(y_j - y_i\right)^2}} \tag{1.8c}$$

The value of the differential terms in equation (1.1) at time *t*, can be written as difference equations:

$$\ddot{u}_{i,t} = \frac{u_{i,t} - 2u_{i,t-\Delta t} + u_{i,t-2\Delta t}}{\Delta t^2} \tag{1.9a}$$

$$\dot{u}_{i,t} = \frac{u_{i,t} - u_{i,t-\Delta t}}{\Delta t} \tag{1.9b}$$

where $u_{i,t-\Delta t}$ and $u_{i,t-2\Delta t}$ denote the displacement of the *ith* cell evaluated at the two previous time points. Substituting equations 1.7 and 1.9 a and b into 1.1 gives the complete equation of motion for cell *i* in the x direction:

$$m_i\left\{\frac{u_{i,t} - 2u_{i,t-\Delta t} + u_{i,t-2\Delta t}}{\Delta t^2}\right\} + \alpha_i m_i\left\{\frac{u_{i,t} - u_{i,t-\Delta t}}{\Delta t}\right\} = \sum_{j=1}^{n}\left[k_{1ij}\left\{\left(x_{Oi} - x_{Oj}\right) + \left(u_{xi} - u_{xj}\right)\right\} - k_{2xij}\left(r_i + r_j\right)\right]$$

Similar equations can be assembled to describe the x and y displacements of every particle in the model. Equation 1.1 can then be written in matrix form, with every row representing the displacement of one cell, and rearranged:

$$\left[\frac{[M]}{\Delta t^2} + \frac{\alpha[M]}{\Delta t} - [K1]\right]\{u_{x,t}\}$$
$$= [K1]\{x_O\} - [K2]\{r\} + \frac{2[M]\{u_{t-\Delta t}\}}{\Delta t^2} + \frac{\alpha[M]\{u_{t-\Delta t}\}}{\Delta t} - \frac{[M]\{u_{t-2\Delta t}\}}{\Delta t^2}$$

This equation is in the standard matrix equation form $Ax = B$ and hence can be solved by commonly used numerical analysis techniques such as Gaussian Elimination or conjugate gradients, to obtain the vector containing the cell displacements at the current iteration. Displacements in *x* and *y* are solved independently and the new positions of each cell obtained from its previous location and calculated displacements. The choice of time step Δt and damping constants α are critical in determining the time taken for a stable solution to be reached and were optimised empirically.

The solution at time *t* depends on the solutions at the previous two time points, so at the first two iterations of each solve an all zero vector is substituted for both $u_{i,t-\Delta t}$ and $u_{i,t-2\Delta t}$. The solution takes several iterations to stabilise as a result. The constants in the $[K1]$ and $[K2]$ matrices depend on the current position of the cells and are updated between every iteration. Also, it is possible that as the cells are displaced, they will lose contact with old neighbours, or exert a force on other cells that were not previously in contact. This necessity to update positional information at each iteration means that this process of resolving forces is the limiting factor in optimising the solution time of the whole model. Once a solution has been found the new position information is inserted into the individual cell structures, and the next iteration of the agent model commences. Hence this physical correction is an iterative process contained within the main iterative structure of the primary agent model. In high density culture conditions, it may be impossible to find a solution where there is no force between adjacent cells. In order to circumvent this problem, an error value, corresponding to the sum of any 'overlaps' with adjacent cells is recorded for every cell at every physical iteration. This list of values is termed the *error vector*. When the difference in the norm of this error vector between two consecutive iterations falls below a threshold value, the physical correction is terminated and the current solution is returned. Any remaining force is

interpreted as squashing of cells, and during the next agent based iteration, cells will round up, reducing their radius to minimise the force acting. Thus, at high cell density, cells will adopt an increasingly round morphology and will eventually be prevented from passing the mid G1 checkpoint.

9 From individual cells to tissue

A more sophisticated model of the physical environment has to recognise that tissue is three dimensional, contains cells which are not rigid, elastic bonds whose strength and density varies, and active movement of cells. All cell movement, and consequent sorting and self-assembly, is driven by physical forces, but these can be both passive – i.e. the result of cell growth and the formation of bonds between adjacent cells – and active – i.e. the result of cells altering their size and shape in order to move. Forces acting on the cells will result in mechano-transduction through both integrin-mediated bonding and the cytoskeleton, which may alter cell behaviour. Of particular importance is the need to link mechano-transduction at the individual cell level to continuum descriptions of tissue on a millimetre length scale. For example, an important feature of skin epithelial cells is their ability to contract the underlying dermis (Chakrabarty *et al*, 2001) which almost certainly reflects part of the wound healing programme for skin which is seen to a much lesser extent with oral epithelial cells. Three types of models have been used to describe the mechanical properties of individual cells: continuum models (a continuous elastic shell surrounding a continuous viscous or visco-elastic core); orthogonal non-linear elements with a constant volume constraint; and a tensegrity structure (6 rigid struts plus 24 elastic cables in tension). Continuum models are, by their nature, unable to provide links to the internal cytoskeletal mechanics. Stamenović *et al* (1996) developed a tensegrity-based approach to model cytoskeletal mechanics in a general sense i.e. to provide an adequate model of the mechanical properties of an individual cell without explicitly modelling the complex architecture of actin filaments and microtubules within the cell. The mechanical properties of a 6 strut plus 24 cable tensegrity structure, when subjected to uniaxial stress, were similar to those found for shear stress applied to endothelial cells. Without further study, it is not clear how the cable tension could be related to mechano-transduction. Palsson (2001) modelled *Dictyostelium discoideum* as a viscoelastic ellipsoid, with the three axes represented by a non-linear spring in parallel with a series spring and dashpot, under the constraint of constant volume. Chemotaxis and cell bonding were included in the model, and sorting due to differential adhesion was demonstrated. Tensile and compressive forces applied to the cells will result in a change in length of the three orthogonal spring-dashpot elements, which could provide a mechanism for incorporating mechano-transduction.

In addition to compressive forces between contiguous cells resulting from cell growth, the distribution of bonds between cells will give rise to translational and rotational forces about all three axes, as will cell motility (which is a result of bonds between an active extension of the cell and another cell or substrate). The physical model has to resolve these forces. One possible approach to solving the physical problem is to make use of techniques developed for Distinct Element Methods (DEM). The initial physical model is conceptually related to DEM, which are widely used for modelling powders and grain structures in chemical engineering and geology. DEM have been used for both linear and non-linear visco-elastic collisions. In particular, the techniques used to handle the elastic response of cohesive aggregates (Jefferson *et al* 2002) are closely related could be used for simple elastic cells with rigid bonds, but would require substantial development to handle visco-elastic cells which grow and divide. Efficient proximity and contact detection algorithms are also needed (binding, and hence tension, between cells is a function of proximity, physical contact generates compressive forces). Williams *et al* (1999) have shown that altering the object representation scheme can reduce the collision detection problem in DEM from $O(n)$ to $O(\sqrt{n})$.

It could be argued that developing a global model of the physical world is negating the original concept, which was to use agents to provide local processing because there is no over-arching explicit control structure. If the same argument is applied to the physical forces – each individual cell moves a result of the forces applied by the contiguous cells – then the resolution of the forces ought to take place at an individual cell level i.e. be part of the structure of the agents. We are currently exploring this option.

10 Validation

We take the view that computational models of biological systems that do not satisfy two conditions (the ability to predict function at a systems biology level, and the existence of robust validation against biological models) are only of academic interest. In the model discussed in this paper there is a one-to-one mapping between the *in vitro* epithelial models and the computational model, with each cell in the biological model having an identical corresponding cell in the computational modelling at the initial seeding stage. Cell growth and division is a stochastic process, so the two models would not be expected to yield identical outcomes after a period of growth under identical conditions – indeed, this would not be expected for a pair of *in vitro* models. Nevertheless, it is reasonable to assume strong similarity, and suitable metrics to compare *in vitro* and *in virtuo* development are being developed. To date, validation of the simple *in virtuo* model has been confined to comparison of real and simulated growth rates for urothelial tissue in normal and low Ca^{2+} environments (Walker *et al* 2003) and wound healing in a urothelial model (Walker et al 2004 submitted)

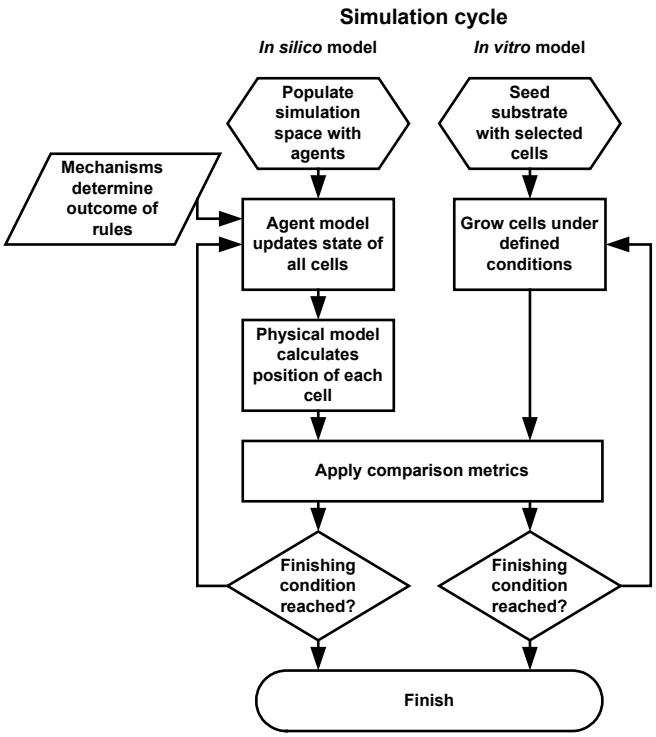

Figure 6: the simulation cycle

11 Computational issues

Building models and carrying out simulation can provide a great deal of information and inspiration for further experimental investigations. It can also offer a platform within which *in virtuo* analysis and reasoning can take place. The first question to answer is whether the model accurately and usefully represents the biological reality. This can not be answered simply, we have to integrate the model building and the experimental validation into as seamless and constructive a framework as possible with the model posing questions about the biology and the biology posing questions about the model. We have certainly found that a systems view of the phenomena helps to formulate hypothesis, questions and experiments that might not arise otherwise. Similarly, the models are challenged by the experimental data that is collected during this process and in our case this leads to the refinement of the underlying rules and processing functions, thus improving the model. It is an iterative process and so there is no real concept of a *correct* model. We aim to produce useful models and how this is judged is determined by the use to which they are put.

One aspect of any computational model, whether of a biological system or of any other type of system is that it needs to be both consistent and as complete as it is practical to aspire to. Methods exist for testing models by developing scenarios (e.g. changing from physiologically relevant levels of calcium to low calcium media which markedly affects epithelial cell biology) - essentially sets of events and environmental conditions which can be applied to the model to see if the resultant behaviour fits with what we expect. It is quite common for models to exhibit unforeseen behaviour, perhaps under some conditions the model either behaves in unpredictable ways or simply fails in some sense. To build a robust model therefore requires an extensive period of validation. Techniques (e.g. Holcombe & Ipate, 1998) exist which can guide this process, although no theory exists for a complex communicating hybrid model as yet.

What can we do with a model apart from simulation? One potential approach is to apply ideas from model checking (Clarke *et al* 1999). This process involves coding up the model in a suitable logic and using powerful software to analyse the model. The models are usually based on state machines and the process involves posing a question which is written in a logic style language and then using the model checker to determine whether it is possible to answer the question. Typically the question might involve statements along the lines of: is it always the case that when the quantity of a critical substance exceeds a particular critical value then a given action

occurs or a given state is entered, is there a state in which a particular property of the model holds, is there a path of behaviour in the model such that every state in that path has a certain property and so on. Model checkers have successfully explored models with very large numbers (billions) of states. Model checking for communicating X-machines has been developed by Eleftherakis using the XmCTL logic (Eleftherakis *et al*, 2001). This provides an important basis for future analysis of large, complex biological models. It is not enough that we only rely on simulation for our understanding of the system, it may be that some highly critical sequence of behaviour only occurs under conditions that we never get round to simulate, yet this knowledge might be important. Rather than trying things out to see what happens we try to identify interesting or undesirable phenomena and see whether the model can ever exhibit them (e.g. predicting what would happen if tissue engineered epithelia were constructed with a sub-optimal percentage of stem cells - this has been a long-standing and almost unanswerable concern for the clinical use of tissue engineered skin. Would one expect the skin to break down over a patient's life-time? A computational model could be used to predict the life-time wear characteristics of the tissue–engineered skin under both normal and wounded (regenerative) conditions. This kind of backwards reasoning is relatively novel and is still at an early stage but looking at the development of computational models in the long term the ability to do this will be extremely powerful. Questions such as: whether there are wound healing advantages to enriching the skin stem cell population or lowering the wound bed calcium for patients with chronic wounds could be assessed initially *in virtuo*. The use of the model could inform the design of *in vitro* (and possibly *in vivo*) experiments to then test these questions

12 The larger scheme: the Physiome Project

The long-range goal of the Physiome Project is to understand and describe the human organism, its physiology and pathophysiology, and to use this understanding to improve human health. A major aim is to develop computer models to integrate the observations from many laboratories into quantitative, self-consistent and comprehensive descriptions (Hunter *et al* 2002). The models that have been hitherto been seen as central to this endeavour fall into two classes – supra-cellular models (whole body system models; whole body continuum models; and tissue and whole organ continuum models) and sub-cellular models (sub-cellular ODE models; sub-cellular Markov models; and molecular models). The development of a model which is specifically based on the individual cell and its interaction with neighbouring cells provides an essential link between supra-cellular and sub-cellular models. In principle, the sub-cellular models provide the mechanisms to replace the rules in the agent model. An important goal for the cellular model is to provide a means for describing the parameters in constitutive equations of tissue (continuum models of the macro world) in terms of events at a sub-cellular level.

13 Conclusion

Modelling biological systems such as epithelial organisation will immediately necessitate a more quantitative approach to describing cell/cell social behaviour. When used in parallel with well characterised 2D and 3D epithelial culture systems then we anticipate that descriptive rule sets will be the first " deliverable" and that these will inform our understanding of some of the biological rules implicit in tissue organisation. It seems a safe prediction that the next "deliverable" will be the ability to run experiments *in virtuo* to test concepts and understanding. In the long term, we propose that the *in virtuo* model will function as a predictive tool of cell behaviour *in vitro* and *in vivo*.

14 References

Balanescu, T., Cowling, A.J., Gheorgescu, H., Gheorghe, M., Holcombe, M., & Vertan, C. 1999. Communicating stream X-machines systems are no more than X-machines. Journal of Universal Computer Science, 5, 494-507

Bhargava S, Chapple C R, Layton C, Bullock A J and MacNeil S 2003 Tissue-engineered buccal mucosa for substitution urethroplasty. British Journal of Urology. In press

Bassingthwaighte, J.B. 1995 Towards modelling the human physiome. Adv. Expt. Med. Biol. 382, 331-339.

Bassingthwaighte, J.B. 2000a Symposium on Integrative Biology of the Heart. Ann. Biomed Eng 28, 835-1058.

Bassingthwaighte, J.B. 2000b Strategies for the Physiome Project. Ann. Biomed. Eng. 28, 1043-1058.

Baumgarter W, Hinterdorfer P, Ness W, Raab A, Vestweber D, Schindler H and Drenckhahn D. 2000.

Cadherin interaction probed by atomic force microscopy. PNAS 97 4005-4010.

Booth C, Harnden P, Trejdosiewicz LK, Scriven S, Selby PJ, Southgate J 1997 Stromal and vascular invasion in an human in vitro bladder cancer model. Lab Invest 76, 843-857

Brodland G W, Chen H. H. 2000. The Mechanics of Heterotypic Cell Aggregates:Insights from Computer Simulations. Journal of Biomechanical Engineering 122, 402-407.

Chakrabarty K H, Dawson R A, Harris P, Layton C, Babu M, Gould L, Phillips J, Leigh I, Green C, Freedlander E, Mac Neil S 1999 Development of autologous human dermal-epidermal composites based on sterilised human allodermis for clinical use. British Journal of Dermatology 141, 811-823

Chakrabarty, K.H., Heaton, J., Dalley, A.J., Dawson, R.A., Freedlander, E., Khaw, P.T. and Mac Neil, S 2001. Keratinocyte-driven contraction of reconstructed human skin. Wound Repair and Regeneration, 9, 95-106

Chen H H, Brodland G. W. 2000. Cell-level Finite Element Studies of Viscous Cells in Planar Aggregates. Journal of Biomechanical Engineering 122, 394-401.

Clarke E M, Grumberg O, Peled D, 1999 *Model checking*, MIT Press.

Dover R and Potten C S 1983. Cell Cycle Kinetics of Cultured Human Epidermal Keratinocytes. Journal of Investigative Dermatology 80 423-429

Eleftherakis, G., & Kefalas P. 2001. Towards model checking of finite state machines extended with memory through refinement. In G.Antoniou, N.Mastorakis, & O.Panfilov (Eds.). (pp.321-326). Advances in Signal Processing and Computer Technologies, World Scientific and Engineering Society Press.

Eleftherakis G, Kefalas P, Sotiriadou A, 2002 Formal verification of agent models, in Proc. 2nd Hellenic Conf. On AI (SETN02), 25-435

Eleftherakis G, Kefalas P, Sotiriadou A, 2001 XmCTL: extending temporal logic to facilitate formal verification of X-machine models. Analel Uni. Bucuresti, Math.-Inf. 50, 79-95

Eves P, Layton C, Hedley S, Dawson R A, Wagner M, Morandini R, Ghanem G, Mac Neil S 2000 Characterisation of an in vitro model of human melanoma invasion based on reconstructed human skin. British Journal of Dermatology 142, 210-222

Forestell S P, M. B. J., Kalogerakis N, Behie L A 1992. A Cellular Automaton Model for the Growth of Anchorage-dependent Mammalian Cells used in Vaccine Production. Chemical Engineering Science 47, 2381-2386.

Georghe, M., Holcombe, M., & Kefalas, P. 2001. Computational models for collective foraging, 4th International Workshop on Information Processing in Cells and Tissues, Leuven, Belgium

Glazier J A and Graner F. 1993. Simulation of the Differential Adhesion driven Rearrangement of Biological Cells. Physical Review E 47, 2128-2153.

Hill G, Varley C, PellegrinS, Trejdosiewicz L K and Southgate J 2003 Human urothelial cell proliferation andmigration isdifferentially regulated by autocrine EGFR signalling during re-epithelialisation in vitro. In preparation

Hogeweg, P. 2000. Evolving Mechanisms of Morphogenesis: on the Interplay between Differential Adhesion and Cell Differentiation. Journal of Theoretical Biology 203, 317-333.

Holcombe M, Holcombe L, Gheorghe M, Talbot N 2003 A hybrid machine model of rice blast fungus, *Magnaporthe grisea*. BioSystems, 68, 223-228

Holcombe, M., & Ipate, F. 1998. Correct systems: Building a business process Solution. Springer Verlag, London

Honda, H. 1983. Geometrical Models for Cells in Tissues. International Review of Cytology 81, 191-248.

Huang S and Ingber D E 1999. The structural and mechanical complexity of cell-growth control. Nature Cell Biology 1 E131-E138.

Hunter P, Robbins Pand Noble D, 2002, The IUPS human physiome project. Pflugers Arch - Eur J Physiol 445:1–9

Jefferson G, Haritos G K and McMeeking R M 2002 The elastic response of a cohesive aggregate – a discrete element model with coupled particle interaction. Journal of the Physics and Mechanics of Solids 50 2539-2575

Jensen U B, Lowell S and Watt F M 1999. The spatial relationship between stem cells and their progeny in the basal layer of human epidermis: a new view based on whole-mount labelling and lineage analysis. Development 126 2409-2418.

International Human Genome Mapping Consortium 2001 A physical map of the human genome. Nature 409, 934-941

Kefalas P, Holcombe M & Gheorghe M, 2003a, "A formal method for the development of agent-based systems", in Intelligent Agent Software Engineering, V. Plekhanova,,ed. Idea Group, 68-98

Kefalas, P., Eleftherakis, G., & Kehris, E. 2003b. Modular modeling of large-scale systems using communicating X-machines, Proceedings of the 8th Panhellenic Conference in Informatics, Greek Computer Society. In press

Lee Y, K. S., McIntire L V, Zygourakis K 1995. A Cellular Automaton Model for the Proliferation of Migrating Contact-Inhibited Cells. Biophysical Journal 69, 1284-1298.

Lim J H F and Davies G. A. 1990. A Stochastic Model to Simulate the Growth of Anchorage Dependent Cells on Flat Surfaces. Biotechnology and Bioengineering 36, 547-562.

Loew L M and Schaff J C 2001 The virtual cell: a software environment for computational cell biology. Trends in Biotechnology 19, 401-406

Morel D, Marcelpoil R., Brugal G 2001. A proliferation control network model:the simulation of two-dimensional epithelial homeostasis. Acta Biotheoretica 49, 219-234.

Morton-Firth C J and Bray D 1998 Predicting temporal fluctuations in an intracellular signalling pathway. J Theor Biol 192, 117-128

Odell G M, O. G., Alberch P, Burnside B 1981. The Mechanical Basis of Morphogenesis. Developmental Biology 85, 446-462.

Orford K, Orford C C and Byers S W 1999. Exogenous expresion of Beta-Catenin regulates contact inhibition, anchorage-independent growth, anoikis and radiation-induced cell cycle arrest. J Cell Biol 146 855-867.

Palsson, E. 2001. A three-dimensional model of cell movement in multi-cellular systems. Future Generation Computer Systems 17, 835-852.

Potten C S and Morris R J 1988. Epithelial stem cells in vivo. J Cell Science Suppl. 10 45-62.

Potten C S and Booth C 2002. Keratinocyte Stem Cells: a Commentary. J Investigative Dermatology 119 888-899.

Ralston D R, Layton C, Dalley A J, Boyce S G, Freedlander E, Mac Neil S 1999 The requirement for basement membrane antigens in the production of human epidermal/dermal composites in vitro. British Journal of Dermatology 140, 605-615

Ransom R and Matela R. J. 1984. Computer Modelling of Cell Division during Development using a Topological Approach. Journal of Embryology and Experimental Morphology 83 Suppl.: 233-259.

Ruaan R C, Tsai. G. J., Tsao G T 1993. Monitoring and Modelling of Density-Dependent Growth of Anchorage-Dependent Cells. Biotechnology and Bioengineering 41, 380-389.

Ryder E F, Bullard L, Hone J, Olmstead J, Ward M O 1999. Graphical simulation of early development of the cerebral cortex. Computer methods and Progress in Biomedicine 59, 107-114.

Savill N J, Hogeweg P. 1997. Modelling Morphogenesis: From Single Cells to Crawling Slugs. Journal of Theoretical Biology 184, 229-235.

Scriven SD, Booth C, Thomas DF, Trejdosiewicz LK, Southgate J 1997 Reconstitution of human urothelium from monolayer cultures. J Urol 158, 1147-1152

Southgate J, Hutton KA, Thomas D F and Trejdosiewicz L K 1994 Normal human urothelial cells in vitro: proliferation and induction of stratification. Lab Invest 71, 583-594

Southgate J, Masters JR, Trejdosiewicz LK 2002 Culture of Human Urothelium. In: Freshney RI, Freshney MG, eds. Culture of Epithelial Cells. 2nd Ed. New York: J Wiley and Sons, Inc.

St Croix B, Sheehan C, Rak J W, Florenes V A, Slingerland J M and Kerbel R S 1998. E-Cadherin dependent growth suppression is mediated by the cyclin-dependent kinase inhibitor p27 KIPI. J Cell Biol 142 557-571.

Stamenović D, Fredberg J J, Wang N, Butler J P and Ingber D E 1996 A Microstructural Approach to Cytoskeletal Mechanics based on Tensegrity. J Theor Biol 181, 125-136

Steinberg, M. S. 1970. Does differential adhesion govern cell self-assembly processes in histogenesis? Equilibrium configurations and the emergence of a hierarchy among populations of embryonice cells. Journal of Experimental Zoology 173, 395-433.

Stekel D, Rashbass. J., Williams E D 1995. A computer graphic simulation of squamous epithelium. Journal of theoretical biology 175, 283-293.

Talbot N J, 1995 Having a blast: exploring the pathogenicity of Magnaportha grisea. *Trends in Microbiology*, 3, 9-16.

Talbot N J & Foster A J, 2001 Genetics and genomics of the rice blast fungus Magnaporthe grisea: developing an experimental model for understanding fungal diseases of cereals, Adv. Bot. Res., 34 , 263- 287.

Tomita M, Hashimoto K, Takahashi K, Shimuzu T S, Matsuzaki Y, Miyoshi F, Saito K, Tanida S, Yugi K, Venter J C and Hutchison C A 1999 E-CELL: software environment for whole-cell simulation. Bioinformatics 15, 72-84

Venter C et al 2001 The sequence of the human genome. Science 291, 1304-1351

Walker D C, Southgate J, Hill G, Holcombe M, Hose D R, Wood S M, Mac Neil S, Smallwood R H 2003 Agent-based modelling of the social behaviour of cells. Biosystems. In press.

Williams J R and O'Connor R (1999) Discrete element simulation and the contact problem. Archives of Computational Methods in Engineering

Zygourakis K, Bizios R., Markenscoff P 1991a. Proliferation of Anchorage-Dependent Contact-Inhibited Cells: I. Development of Theoretical Models based on Cellular Automata. Biotechnology and Bioengineering 38, 459-470.

Zygourakis K, Markenscoff P., Bizios R, 1991b. Proliferation of Anchorage-Dependent Contact-Inhibited Cells: II. Experimental Results and Validation of the Theoretical Models. Biotechnology and Bioengineering 38, 471-479.

Walker D C, Hill G, Wood S M, Smallwood R, Southgate J 2004. Agent-based computational modeling of epithelial cell monolayers: predicting the effect of exogenous calcium concentration on the rate of wound closure. Submitted

.