ALMA MATER STUDIORUM
UNIVERSITÀ DEGLI STUDI DI BOLOGNA

Seconda Facoltà di Ingegneria con Sede a Cesena
Corso di Laurea Specialistica in Ingegneria Biomedica

MODELLING AND SIMULATING IN SYSTEMS BIOLOGY: AN APPROACH BASED ON MULTI-AGENT SYSTEMS

Elaborata nel corso di:
Sistemi Intelligenti Distribuiti LS

Tesi di Laurea di:
SARA MONTAGNA

Relatore:
Prof. ANDREA OMICINI

Correlatori:
Prof. ALESSANDRO RICCI
Prof. ALESSANDRO SARTI

ANNO ACCADEMICO 2005–2006
SESSIONE III

KEYWORDS

Complex Systems
Systems Biology
Multi-agent Systems
Autopoiesis
Hematopoietic Stem Cell

Simple interactions can have consequences that are not predictable by intuition based on biological experience alone.

Lee A. Segel, (1980). ed.
Mathematical Models in Molecular and Cellular Biology,
Cambridge University Press,
Cambridge, England

Contents

Introduction

Systems Biology is an innovative way of doing biology recently raised in bio-informatics contexts, characterised by the study of biological systems as *complex systems* with a strong focus on the system level and on the interaction dimension. In other words, the objective is to understand biological systems as a whole, putting on the foreground not only the study of the individual parts as standalone parts, but also of their interaction and of the global properties that emerge at the system level by means of the interaction among the parts.

This thesis focuses on the adoption of *multi-agent systems* (MAS) as a suitable paradigm for Systems Biology, for developing models and simulation of complex biological systems. Multi-agent system have been recently introduced in informatics context as a suitabe paradigm for modelling and engineering complex systems. Roughly speaking, a MAS can be conceived as a set of autonomous and interacting entities, called *agents*, situated in some kind of *environment*, where they fruitfully interact and coordinate so as to obtain a coherent global system behaviour. The claim of this work is that the general properties of MAS make them an effective approach for modelling and building simulations of complex biological systems, following the methodological principles identified by Systems Biology. In particular, the thesis focuses on cell populations as biological systems.

In order to support the claim, the thesis introduces and describes *(i)* a MAS-based model conceived for modelling the dynamics of systems of cells interacting inside cell environment called niches. *(ii)* a computational tool, developed for implementing the models and executing the simulations. The tool is meant to work as a kind of *virtual laboratory*, on top of which kinds of *virtual experiments* can be performed, characterised by the definition and execution of specific

models implemented as MASs, so as to support the validation, falsification and improvement of the models through the observation and analysis of the simulations. A hematopoietic stem cell system is taken as reference case study for formulating a specific model and executing virtual experiments.

Thesis outline

The interest for the themes studied in this thesis began two years ago, by getting in touch with different but related works and studies on complex systems, in particular the theory on dissipative structure developed by Ilya Prigogine and the theory of autopoiesis developed by Humberto Maturana and Francisco Varela. What is really a complex system? How can we define and describe their dynamics? In particular, in the wide field of complex systems, my central interest was on biological complex systems.

The first chapter of the thesis has its focus in describing the properties of a biological system which make it complex, in particular hierarchical organization, non-linear dynamics that arise from the interaction between the components of the system, and emergent behaviours.

The second chapter focuses on traditional and state-of-the-art approaches currently used to study complex biological systems and their properties. Among them, one of the discipline which actually had a big success in the study of biological systems is molecular biology. Molecular biology is the study of biology at a molecular level and is the study of the structure, the biochemical composition and the functions of intracellular components. Despite the important scientific results obtained in this discipline, it adopts a reductionistic approach which does not permit the catch some important properties of biological systems, such the emergent behaviours arising from the interactions between components.

For this purpose, approaches recently emerged in the context of *Systems Biology* seem more appropriate. The main tools of investigation adopted by these approaches are *modelling* and *simulating* tools. They have been added to the experimental technologies used by molecular biology for data acquisition, and are the mean for interpreting,

analysing, elaborating such experimental data in order to build new knowledge on biological systems.

The third chapter provides a survey of the different formalisms used for modelling and simulating biological systems as found in state-of-the-art literature. The approaches can be subdivided in two main categories: mathematical models, based on differential equations (the most used); computational models, such as process algebra, petri-nets and agent-based and multi-agent systems models, as the most recent approaches.

The focus of this thesis is on multi-agent systems, as an approach for Systems Biology. Key concepts and abstractions of multi-agent systems are described in the fourth chapter, where also a brief survey of existing approaches applying MAS for modelling and simulating biological systems is reported.

MAS formalism gives to a modeller the opportunity of describing the behaviour of the system in terms of different, heterogeneous and organizationally closed entities, following a bottom-up approach. Specified the internal and interactive behaviour of such system's components, through the simulation process, we can observe the global emergent behaviour of the system. This approach clearly implies a change of view in comparison to the mathematical approaches based on differential equations, which exploit a top-down approach and describe the behaviour of the system in terms of global laws.

In the fifth chapter, a concrete MAS approach for modelling and simulating cellular populations is proposed. First, an abstract model is defined, useful to identify the components of the system and to describe their abstract behaviour. Then, a computational tool for doing simulations of cell systems modelled upon such an abstract model is introduced, as a kind of virtual laboratory for doing virtual experiments.

In the sixth and seventh chapter the abstract model and the computational tools are applied to a specific case study, hematopoietic stem cells (HSCs), making some concrete examples of MAS-based models and testing the effectiveness of the virtual laboratory tool. First, a review of the biochemical properties and behaviour of HSC systems is provided: how do they act, how do they interact with their micro-environment , how do they respond to external stimuli and how

do they change their local environment. Them, such a knowledge is exploited to build a MAS-based model of a HSC systems, and to perform some virtual experiments with the aim of validating the model and the approach, through the reproduction of known biological behaviours.

Chapter 1

Biological Systems as Complex Systems

1.1 Complex Systems

1.1.1 The paradigm shift

The history of science during the three century that followed the newtonian synthesis is a dramatic story indeed. At the beginning of the last century, when the program of classical science seemed near completion, physicists were almost unanimous in agreeing that the fundamental laws of the universe were deterministic and reversible. However at each such moment something invariably did not work out as anticipated: processes that did not fit this scheme were taken to be exceptions, merely trick due to complexity, which itself had to be accounted for by invoking our ignorance, or our lack of control of the variables involved. The scheme had to be enlarged, and the fundamental level remained elusive.

During the XX century, more and more scientists have come to think that many fundamental processes shaping nature are irreversible and stochastic; that the deterministic and reversible laws describing the elementary interaction may not be telling the whole story; that our physical world is no longer symbolized by the stable and periodic planetary motions that are at the heart of classical mechanics, but it is a world of instabilities and fluctuations, which are ultimately responsi-

ble for the amazing variety and richness of the forms and structures we see in nature all around us [35]. Wherever we look, we find evolution, diversification, and uncertainties.

We have long known that we are living in a pluralistic world, in which we find deterministic as well as stochastic phenomena, reversible as well as irreversible phenomena. We observe a great number of deterministic phenomena, such as the frictionless pendulum or the trajectory of the moon around the earth; moreover we know that many phenomena – the frictionless pendulum, for one – are also reversible, that means that future and past play the same role in the equations describing the motion or dynamics involved. But other processes such as diffusion or chemical reactions are irreversible, i.e. in such processes there is a privileged direction of time. Instead, to explain the variety of natural phenomena, we are forced to acknowledge the existence of stochastic processes, whose dynamics is nondeterministic, probabilistic, even completely random and unpredictable. Such processes appear in a vast number of natural phenomena – for example brain, ecology, sociology, pedestrian, and human immune cells behaviour – which are observed in the whole spectrum of scientific fields such as biology, medicine, social sciences, physics, mathematics and many others.

This critical knowledge point had so been achieved in numerous research fields simultaneously. At this moment scientists realized that classical approaches are not longer useful to formulate a wide range of problems. This leads to a new vision of matter, one no longer passive, as described in the mechanical world view, but associated with spontaneous activity.

This deep change is the result obtained in quite different areas of investigation. In particular, historically there are two disciplines that have dramatically modified the outlook of science. The first is *nonequilibrium physics*, with the discovery of fundamental new matter's properties of self-organization in far-from-equilibrium conditions. The second discipline is the modern *theory of dynamical system*: here the central discovery is the prevalence of instability, that means that small changes in initial conditions may lead to large amplifications of the effects of changes.

The new approach to understanding nature was pursued by outstanding researchers and their team around the world: Ilya Progogine

at the University of Brussels, Humberto Maturana at the University of Chile in Santiago, Francisco Varela at the École Polytechnique in Paris, Lynn Margulis at the University of Massachusetts, Benoit Mandelbrot at Yale University, and Stuart Kaufmann at the Santa Fe Institute. All they and others put the basis for the *science of complexity*, unifying and integrating new theories, ideas and methods from many disciplines which tried to explain natural phenomena, and identifying their properties and their behaviour. Several key discovery by these scientists have been hailed has revolutionary.

It is because off all these, that now we may consider applying new knowledge to situations for which the concepts of classical physics were insufficient or inappropriate, or even essentially meaningless. The new methods developed in this context lead to a better understanding of the environment in which we live, and in which we find both unexpected regularities as well as equally unexpected large-scale fluctuations.

To explain and better understand the variety of natural phenomena, the new science of complexity has been responsible of a paradigm shift from the mechanistic worldwide to a holistic view, arising a basic tension between the parts and the whole – the emphasis on the parts belong to a mechanistic or reductionist view, and the emphasis on the whole to a holistic or organismic view. Following this shift, the new paradigm of complex systems see the world as an integrated whole rather then a dissociated collection of parts.

1.1.2 The science of Complexity

The science of complexity has no precise definition but it refers at that science which studies how parts of a *complex system* give rise to the collective behaviours of the system, how the system interacts with its environment and how these interactions influence both the patterns of system's behaviour and of environmental conditions.

The challenge of the science of complexity is understanding the ways of describing complex systems and the process of formation of complex systems through pattern organization and evolution.

Main features of complex systems

Description of complex systems

There is no generally accepted formal definition of complex system. Informally, a complex system is an open system, composed by elements which interact through nonlinear dynamics and which constitute an organized entity, able to adapt itself to its environment, and to evolve in time. The collective behaviour that arises from the interactions between components is the emergent behaviour at the base of complex systems, which define the structure and the dynamic of the interconnections' network of the whole system.

Systems are typically defined to be complex systems if they exhibit the following properties.

1. Complex systems may be nested (*hierarchical organisation*)

   Complex systems are organized in the shape of a pyramid, with each row of elements linked to elements directly beneath it. In other words, they are systems that rank and organise their components, where each element of the systems (except for the top element) is subordinate to a single other element. The components of a complex system may themselves be complex systems.

   Such a kind of hierarchical organisation has an important and specific role, that is controlling a system composed by elements which have to act in a coordinate and harmonious way.

2. *Nonlinear Dynamics* and *Feedbacks*

   Starting from an initial state, the system evolves, following nonlinear dynamics, to singular states depending on the value of certain control parameters. These states, qualitatively different, are on the temporal level either stationary, periodic or chaotic oscillating without period. Moreover it means that small perturbation may cause a large effect (see butterfly effect), a proportional effect, or even no effect at all. In linear systems, effect is always directly proportional to cause.

   Such a kind of nonlinear dynamics often grow upon *network of interaction* between system's components that are feedback loops. In fact both negative (damping) and positive (amplifying)

feedback are often found in complex systems. In this way the effects of one or more elements' behaviour are feedback in the same way that the element themselves are altered.

3. Complex systems are *open*

   Complex systems in nature are usually open systems that is, they exist in a thermodynamic gradient or dissipate energy or continuous flux of matter or information. In other words, complex systems are usually far from equilibrium: but despite these fluxes, there may be pattern stability. According to Prigogine, this phenomenon arises *dissipative structures*. The term refers to the dynamics of non-equilibrium structures; that is, organized states that remain stable for long periods of time despite matter and energy continually flowing through them.

   Prigogine defines dissipative structures as islands of order in a sea of disorder, maintaining and even increasing their order at the expanse of greater disorder in their environment. They are stable non-equilibrium situations.

4. Complex systems have a *memory*

   The history of a complex system may be important. Because complex systems are dynamical systems they change over time, and prior states may have an influence on present states.

5. Complex systems may be *autopoietic*

   Autopoiesis literally means "self-reproduction" and expresses a fundamental complementarity between structure and function. Francisco Varela is the father of this theory.

6. Complex systems manifests *emergent phenomena*

   Complex systems, as collections of interacting elements, show characteristics that are properties of the collective behaviour of these elements. Such properties do not naturally arise out of the description of an individual component. To describe this spontaneous phenomena we use the term *emergence*. It refers to the process of complex pattern formation from more basic

constituent parts or behaviours, and manifests itself as an emergent property of the relationships between those elements. For a phenomenon to be termed emergent it should generally be unpredictable from a lower level description. At the very lowest level, the phenomenon usually does not exist at all or exists only in trace amounts: it is irreducible.

One form of emergence is *Self-Organization*, that is a process where the organization of a system spontaneously increases and the interactions between components become elaborate and orchestrated. With the phenomenon of self-organization emerge new structures and global properties which involve and organize, in a coordinate and harmonious way, a lot of system's elements. The crucial aspects for the system's behaviour, at this emergent hierarchical level, are no more the functions, the structure, and the behaviour of each element, but the relations and cooperation between these elements.

1.1.3 Dynamic of Complex Systems

After providing a descriptive feeling of the properties of complex systems by observing their behaviour and dynamic, we must take a more systematic, deeper look at these ideas in order to establish the vocabulary of complexity. Complex biological systems inherently behave between two broad regimes separated by a third-phase transition regime: the two broad regimes are *chaotic* and *ordered*, while the phase transition zone between them comprises a narrow third *complex* regime poised on the boundary of chaos: complex systems appear to lie in the ordered regime near the *edge of chaos*.

In this scenario it is useful explain what chaos is and which is the theory that define and describe it.

A brief history of Chaos Theory

During 1980s decade a fascinating theory changed the international scientific world: the *chaos theory*. It became in brief a formidable successful theory, that wants explain each natural phenomena through its general principle, and tried to explain the dynamics of a com-

plex system. Chaos theory, in fact, describes the behaviour of certain nonlinear dynamical systems that under certain conditions exhibit a phenomenon known as *chaos*, a state of extreme confusion and disorder. Kellert (1993) defines chaos theory as "the qualitative study of unstable aperiodic behaviour in deterministic nonlinear dynamical systems".

Chaos theory was formulated during the 1960s. Its story is one of many people – scientists who dared to think along new and unsuspected channels. We think that is not this thesis the true place where we can go into more deep of this theory. We'd like just talk about its beginning and cite the major scientists which work on it.

Mathematicians have known about nonlinearity since *Henri Poincaré* at the turn of this century. Most equations that attempt to predict the actions of nature or natural materials are close approximations rather than exact. They contain one or more factors of nonlinearity. Several courageous scientists were so intrigued with chaos, that they began to do research into both nonlinearity and turbulence.

In 1961, *Edward Lorentz* discovered the *butterfly effect*. He was trying to forecast the weather. He was running a long series of computations on a computer when he decided he needed another run. Rather than do the entire run again, he decided to save some time by typing in some numbers from a previous run. Later, when he looked over the printout, he found an entirely new set of results. The results should have been the same as before. After thinking about this unexpected result, he discovered that the numbers he typed in had been slightly rounded off. In principle, this tiny difference in initial conditions should not have made any difference in the result, but it did. From this, Lorentz determined that long-distant weather forecasts are impossible. Tiny differences in weather conditions on any one day will show dramatic differences after a few weeks, and these differences are entirely unpredictable. Although Lorentz's discovery was an accident, it planted the seed for the new theory of chaos.

A dynamical system to be classified as chaotic must so be sensitive to initial conditions. Sensitivity to initial conditions means that each point in such a system is arbitrarily closely approximated by other points with significantly different future trajectories. Thus, an arbitrarily small perturbation of the current trajectory may lead to

significantly different future behaviour. As a result of this sensitivity, the behaviour of systems that exhibit chaos appears to be random, exhibiting an exponential error dispersion, even though the system is deterministic in the sense that it is well defined and contains no random parameters.

One of the foremost contributors to the new science was *Benoit Mandelbrot*. Using a home computer, Mandelbrot pioneered the mathematics of *fractals*, a term which he coined in 1975. His fractals helped describe or picture the actions of chaos, rather than explain it. The striking principle he discovered was that many of the irregular shapes that make up the natural world, although seemingly random and chaotic in form, have a simple organizing principle. A new geometry of chaos was born.

In 1971, David Ruelle and Floris Takens described a phenomena they called a strange attractor. This strange phenomena was said to reside in what they called phase space and a whole new element of chaos theory was born.

Another pioneer of the new science was *Mitchell Feigenbaum*. His work in the late 1970s was so revolutionary that several of his first manuscripts were rejected for publication because they were so novel, they were considered irreverent. He discovered order in disorder. Feigenbaum showed that period doubling is the normal way that order breaks down into chaos. He calculated universal numbers which represent ratios in the scale of transition points that occur during the process of period doubling. These ratios are now called Feigenbaum numbers. Gleick (1987) mentions that Richard J. Cohen and his medical colleagues at MIT found that period doubling is associated with the onset of a heart attack. This finding brought chaos science into the domain of medical science.

By the mid 1970s, the movement toward chaos as a science was well underway and in 1977, the first conference on chaos theory was held in Italy. Perhaps the most startling finding to come out of this new scientific theory is that order exists within chaos. In fact, order comes from chaotic conditions.

Dynamic systems can show different kind of behaviour: from ordered until chaotic behaviour, depending on parameters that define their evolution. Every system could be found, in different circum-

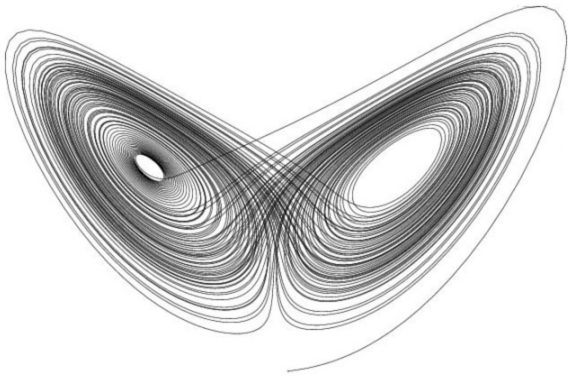


Figure 1.1: Lorentz strange attractor

stances, in a state of chaos, but this state could be reversible.

Attractors

The weave between order and chaos, that define complex phenomena, regularities in behaviour of systems which are extraordinarily complex, ask for formalisms and accepted concepts which allow their understanding.

A recurrent term in the description of complex systems and their dynamics is the term *attractor*, that is a region in phase space: since a large set of initial conditions the system will lead to orbits that converge to this region.

Phase space allows scientists to map information from complex systems, making a picture of their moving parts, and allowing insight into a dynamic system's possibilities. It is a mathematically constructed conceptual space where each dimension corresponds to one state variable of the system.

The simplest kind of attractor is a fixed point. An example of this is a pendulum subject to friction: no matter how it starts swinging, the pendulum always comes to rest at the same point. The next most complicated attractor is a limit cycle, forming a closed loop in phase space. A limit cycle describes stable oscillations, such as the motion of a pendulum clock or the beating of a heart.

Compound oscillations, or quasi-periodic behaviour are described by a torus, which resembles the surface of a doughnut. One oscillation is described around the larger perimeter of the doughnut, the other, perpendicular to it, around the smaller section. Higher dimensional tori can be used to describe combinations of more than two oscillations. (a multi-dimensionsal torus is called a hypertorus.) Despite the complexity of these latter examples, all these attractors describe predictable systems.

Chaotic, or strange attractors, on the other hand correspond to unpredictable motions and have a more complex geometric form. Strange attractors characterize chaotic phenomena, which show some regularities but in which the system's trajectory change continuously during time.

The dynamic of a complex system could be also defined by numerous attractors.

Bifurcation Theory

An other important concept for describing the dynamic behaviour of complex phenomena is the concept of *bifurcation*. A bifurcation is a qualitative change of the attractor of a dynamical system as the result of a moving system's parameter. It represents the sudden appearance of a qualitatively different solution for a nonlinear system as this parameter is varied. It may be accompanied by a change of the stability of an attractor. For example, a simple equilibrium, or fixed point attractor, might give way to a periodic oscillation as the stress on a system increases. Similarly, a periodic attractor might become unstable and be replaced by a chaotic attractor.

These types of mathematical constructs are the subject of *bifurcation theory*, that study how and when such bifurcations can occur.

1.1.4 Scientific Research Institute about Complex Systems

To complete this brief treatment we are going to cite some of the international research institute which have, as main focus of their researches, complex systems in the numerous fields of knowledge.

Santa Fe Institute

The Santa Fe Institute is devoted to creating a new kind of scientific research community, one emphasizing multidisciplinary collaboration in pursuit of understanding the common themes that arise in natural, artificial, and social systems. This unique scientific enterprise attempts to uncover the mechanisms that underlie the deep simplicity present in our complex world.

Since its founding in 1984, the Santa Fe Institute (SFI) is a private, independent research and education centre which has devoted itself to fostering a multidisciplinary scientific research community pursuing frontier science in the physical, biological, computational, and social sciences. SFI seeks to catalyse new research activities and serve as an institute without walls.

New England Complex Systems Institute (NECSI)

For over 10 years, The New England Complex Systems Institute (NECSI) has been instrumental in the development of complex systems science and its applications. NECSI conducts research, education, knowledge dissemination, and community development around the world for the promotion of the study of complex systems and its application for the betterment of society.

NECSI was founded by faculty of New England area academic institutions in 1996 to further international research and understanding of complex systems, that pervade all traditional fields of science.

NECSI research develops basic concepts and formal approaches as well as their applications to real world problems. To date, the contributions of NECSI researchers include studies of networks, agent-based modelling, multi-scale analysis and complexity, chaos and predictability, evolution, ecology, biodiversity, altruism, systems biology, cellular response, health care, systems engineering, negotiation, military conflict, ethnic violence, and international development.

Institute for the Study of Complex Systems (ISCS)

ISCS is a research organization in Palo Alto, California, USA, that specializes in evolutionary/functional approaches to complexity. The

director of the ISCS is Peter A. Corning, Ph.D., who is known especially for his work on the causal role of synergy in evolution. Current work at the Institute also includes a new approach to the relationship between thermodynamics and biology called "thermoeconomics", and a new, cybernetic approach to information theory called "control information".

1.2 Biological Complex Systems

Living beings are undoubtedly the most complex and organized objects found in nature, in view of their morphology and their functioning. In biological systems we can so recognize the cited properties of complex systems.

For example the simplest living system we know is a cell but, in order to give an even rough idea of cellular organization, the description of cell's components has to be quite elaborate; and the complexity increase dramatically when we try to picture how these cell's components are interlinked in a vast network, involving thousands of biochemical processes: large numbers of functionally diverse, and frequently multifuncional, *sets of biological elements interact selectively and nonlinearly* to produce complex behaviours. As a consequence biological systems are characterised by *network of elements*. Such organizational characteristics have both a geometric and functional component with adaptation to the surrounding environment being an essential ingredient for both development and long-term viability of the organism.

Moreover, living systems function under conditions *far from equilibrium*: an organism as a whole continuously receives fluxes of energy and of matter, which it transforms into quite different waste products evacuated to the environment [35]. These fluxes are indispensable so that biological system lives and functions.

At the biological level, complexity is also often associated at the *concept of emergence*: a biological system has a potential richness higher than that of the sum of its subsystems.

We are now looking for the details of complex features of biological systems.

1.2.1 Hierarchy

One of the key characteristic of the organization of living organisms is their hierarchical nature. Indeed, an outstanding property of all life is the tendency to form multi-levelled structures of systems within systems. Each of these forms a whole with respect to its parts while at the same time being a part of a larger whole. These multi-levelled structures have been called hierarchies.

Biological systems have different level of hierarchical organization: *(1)* sequences; *(2)* molecules; *(3)* pathways (such as metabolic or signalling); *(4)* networks, collections of cross-interacting pathways; *(5)* cells; *(6)* tissues; *(7)* organs.

In biological systems there is also a constant interplay between events at different levels. This interplay extends from the events that happen very slowly on a global scale right down to the most rapid events observed on a microscopic scale. A unique molecular event, like a mutation occurring in particularly fortuitous circumstances, can be amplified to the extent that it changes the course of evolution. In addition, all processes at the lower level of this hierarchy are restrained by and act in conformity to the laws of the higher level.

1.2.2 Interaction between components: Biological Networks

Complex systems are defined as systems of interacting parts where the state of one part is influenced by the state of one or more others. In biological systems we can find such a kind of interactions in a lot of biological networks, at different levels of the structural organisation, which are at the core of all biological functions.

In each of these networks we can find many elements with complex dynamics interacting with each other, evolving in time and changing the activities of other components. These processes are found from biochemical pathways, gene regulation mechanisms and metabolic reaction networks, inside a cell, to cell communication processes.

In the specific cell context, proteins, genes and other molecules interact changing their activities: a gene can be inhibited or activated from transcription factors, an enzyme or other proteins can be acti-

vated through phosphorylation and so on. At a higher level we find interactions between cells such that of neural and immune networks. For example a neural network describes a population of physically interconnected neurons or a group of disparate neurons. Communication between neurons often involves an electrochemical process which change the depolarization state of a neuron. This process allow the transmission of signals and informations in the whole body. Other communication processes between cells of different tissues are at the basis of developmental processes and cell differentiation.

Nonlinearity and feedbacks

Often the interactions are *nonlinear* so it is not possible to reduce the system's behaviour to the sum of its parts and to consider the avarege effects. Common interactions in these systems are *feedback loops*, in which information from the output of a system transformation is sent back to the input of the system.

If the new input facilitates and accelerates the transformation in the same direction as the preceding output, they are positive feedbacks, whose effects are cumulative. If the new data produces an output in the opposite direction to previous outputs, they are negative feedbacks, whose effects stabilize the system. In the first case there is exponential growth or decline; in the second there is maintenance of the equilibrium [21].

Negative feedback loops are typically responsible for *regulation*, and they are obviously central in the *homeostasis* of biological systems. At the other hand, the last result of a positive feedback is often amplifying and "explosive": a small perturbation will result in big changes. This feedback, in turn, will drive the system even further away from its own original setpoint, thus amplifying the original perturbation signal, and eventually become explosive because the amplification often grows exponentially (with the first order positive feedback), or even hyperbolically (with the second order positive feedback).

Autocatalisis and Regulation

Nonlinearity in biology may arise intrinsically, independent from spatial inhomogeneities. Most of the chemical reactions are the results

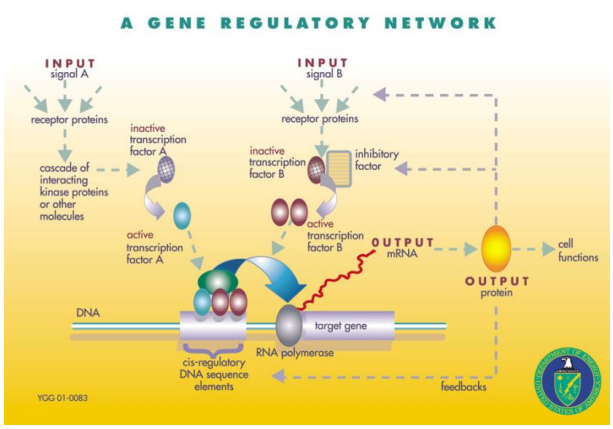


Figure 1.2: A wonderful example of regulatory network

of collisions. For instance $X + Y \rightarrow D$ implies that the molecules of species X and Y must interact and the function which represent this interaction could be nonlinear. Of particular interest are the nonlinearities related to *regulation*. Here a substance X may activate or inhibit the rate of its own production or of the production of another constituent, which in turns feeds back on the first substance.

In biology, regulation is intimately related to the peculiar structure and reactivity of the enzymes, but we find forms of regulations also at higher level. For example stem cells are a system with the specific aim of produce new cells, of a specific type, when the organism need them: these production are complexly regulate through signals which activated or disactivated stem cells in their reproduction and differentiations.

1.2.3 Interaction with the environment: Dissipative Structure

Biological systems also interact with their environment through a continual exchange of energy and matter. Through their interactions with the environment living systems continually maintain and renew themselves, using energy and resources from the environment for that

purpose. These continual flow of heat through the system maintain a form of non-equilibrium and generate a complex spatial pattern in which millions of molecules move coherently. Without these exchanges a living systems cannot exist.

The conditions of "far from equilibrium" don't necessarily imply form of instabilities; though it is opposite to an intuitive feeling, living organisms are able to maintain their life processes under conditions of non-equilibrium that may be stable.

This theory have been supported by Ilya Prigogine who describes the structure of living systems as a *dissipative structure*, putting a main emphasis on the openness and stability of biological systems. In the dissipative structure Prigogine see the coexistence of change (non-equilibrium) and stability [38]. *They maintain themselves in a stable state far from equilibrium.*

A cell is so seen as a stable structure with matter and energy continually flowing through it. The balancing force for these flows are chemical, in particular the catalytic loops in the cell's network that act as self-balancing feedback loops. At the cellular level the strong *inhomogeneities* that we can observe are responsible for some of these structures. For example the concentrations of potassium ions, K^+, inside a neurons, is higher then the outside environment, while the opposite is true for the sodium ions, Na^+. Such inequalities, which implies states of high non-equilibrium, are at the origin of processes such as the conduction of nerve impulse. Besides the cells are maintained by active transport and bio-energetic reactions like glycolysis or respiration.

Many of the characteristics of dissipative structure - the sensitivity to small changes in the environment, the relevance of previous history of critical points of choice, the uncertainty and unpredictability of the future - was revolutionary new concepts from the point of view of classical sciences.

1.2.4 Autopoiesis, Autonomy and Self-Organization

An other key characteristic of a living network is that it continually produces itself. According to the theory developed by Humberto Maturana and Francisco Varela, this property is defined by the term

autopoiesis. Autopoiesis literally means "self-reproduction" or "self-making", and in biological systems refer to a network pattern in which the function of each component is to participate in the production or transformation of other components in the network. In this way, the network is produced by its component and in turn produces those components [7] which: *(i)* through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and *(ii)* constitute it as a concrete unity in space in which the components exist by specifying the topological domain of its realization as such a network.

Since all components of an autopoietic network are produced by other components in the same network, the entire system is *organizationally closed*, even though it is open with regard to the flow of energy and matter. This organizational closure implies that a living system is *self-organizing* in the sense that its order and behaviour are not imposed by the environment but are established by the system itself: the continual self-making, in fact, also includes the ability to form new structure and new patterns of behaviour. Self-Organization refers, in particular, to a high degree of coordination among parts, in the sense of forming spatial and temporal patterns as a kind of spontaneous self-organization [44]. These patterns may emerge at levels far removed, from the level of single gene.

Moreover the organizational closure also implies that living systems are *autonomous*. As seen, this does not mean that their are isolated from their environment. On the contrary they interact with the environment, but this interaction does not determine their organization because they are *self*-organizing.

Therefore, autopoiesis is seen, in this context, as the pattern underlying the phenomenon of self-organization and autonomy, that is so characteristic of all living systems.

The canonical example of an autopoietic system, and one of the entities that motivated Maturana and Varela to define autopoiesis, is the biological cell. The enzymes alone form an intricate network of catalytic reactions, promoting all metabolic processes, and the energy carriers (as ATP) form a corresponding energy network to fuel them. Or let us think at genetic regulatory networks, that evolve along complex feedbacks loop made of interaction through which each

component help to produce or transform other components; thus these network are clearly autopoietic. At higher level every living organism continually renews itself, cells breaking down and building up structures, tissues and organs replacing their cells in continual cycles.

1.2.5 Structural Coupling

According to the Prigogine theory of dissipative structure a living system is an open system which recurrently interacts with its environment. According to Maturana and Varela and their theory of autopoiesis these living systems are autonomous, however, and through a process of self-organization reach a state of stability. When Ilya Prigogine describes the structure of a living system as a dissipative structure, his main emphasis is on the openness of that structure to the flow of energy and matter. When Maturana and Varela describe the pattern of life as an autopoietic network, by contrast, their main emphasis is on the organizational closure of that pattern. Thus a living system is both open and closed – it is organizationally closed but structurally open. Matter continually flows through it, but the system maintains a stable form, and it does so autonomously through self-organization.

These researchers use the word *organization* to connote the configuration of relations between components that define the class identity of a composite unity or system as a totality or singular entity; and the word *structure* to connote the physical embodiment of the system's pattern of organization, referring to the components and the relations between them that realize a system or a composite entity as a particular case of a particular class. The conservation of the organization of a system is a condition of existence: if the organization changes, the system disintegrates and something different appears in its place. The structure of a system is open to change, and can change in two ways: *(1)* structural changes through which the organization of the changing system is conserved: these are changes of state which conserve the class identity of the system; *(2)* structural changes through which the organization of the structurally changing system is lost, not conserved; these are disintegrative changes [28].

The structure of a living system changes both as a result of its

internal structural dynamics and as a result of its interactions. A realistic picture of autopoietic networks must so include a description of how living systems interact with their environment and how this interaction can preserve the autonomy of these systems. Indeed, such a description is an integral part of the theory of autopoiesis developed by Maturana e Varela. At the aim they introduce the concept of *structural coupling*.

Maturana and Varela claims that the interactions between system and environment happen through structural coupling, i.e. through interactions that trigger structural changes in the system. For example a cell membrane continually incorporates substances from its environment into the cell's metabolic processes. An organism's nervous systems changes its connectivity with every sense perception. These living systems are autonomous, however: the environment only triggers the structural changes of the system and it does not specify or direct them.

As a living system responds to environmental influences with structural changes, these changes will in turn alter its future behaviour. In other words, a structurally coupled system is a learning system. As long as it remains alive, a living organism will couple structurally to its environment. Its continual structural changes in response to the environment — and consequently its continuing adaptation, learning and development — are key characteristic of the behaviour of living beings.

In this process, the structure of the living system and the structure of the medium change together congruently as a matter of course, and the general result is that the history of interactions between two or more systems becomes a history of spontaneous recursive coherent structural changes in which all the participant systems change together congruently until they separate or disintegrate.

1.2.6 Stocasticity

Many studies have reported occurrence of stochastic fluctuations and noise in living systems. At intracellular level, observations of gene expression in individual cells clearly illustrates the stochastic nature of transcription and translation: similar initial conditions, such as con-

centrations of chemical species, temperature, pressure etc., have been shown to produce qualitatively different outcomes in the temporal evolution of a regulatory network. A classic example is the lysis/lysogenic switch of bacteriophage λ infecting *Escherichia coli* cells. It has two possible developmental pathways: due to noise the network may randomly evolve into one of the two bistable states [16].

Which are the origin of the observed stochasticity? At the microscopic level of functioning of cellular processes the interactions between the molecules - DNA, mRNA, protein, small molecules - follow the laws of physics. A fundamental result of theoretical statistical physics is the famous $\sqrt{n}$ law, which says that randomness or fluctuations in a system is inversely proportional to square root of the number of particles which can be considered as an index of the system size. As a result, low number of particles or low concentrations should result in high fluctuation. As the concentrations of the reacting species are increases they become less prominent tend to the determinstic solution.

Biochemical species participating in processes such as gene transcription, regulation and signalling transduction often occur in low copy numbers. As a result elementary reactions, such as polymerase binding or complex formation, take place with widely distributed reaction times. Such stochastic effects arising due to the inherent nature of biochemical interactions are often termed as *intrinsic noise*.

In the context of gene expression there exists an *extrinsic* component of *noise* too, arising from random fluctuations in other factors, e. g. the number of ribosomes, the stage of the cell cycle, mRNA degradation, and the cellular environment. These are due to the external environmental conditions. For example a transcription factor for a particular gene is mostly the protein product of another gene and thus its production is also probabilistic. In these situations, a protein product arising out of a stochastic activation of a gene, leads to a cascade of downstream stochastic events [30].

Chapter 2

Systems Biology: Modelling and Simulating Biological Systems

The ultimate goal of biology is to understand biological systems in sufficient detail to enable accurate, quantitative predictions about the behaviours of biological systems, including predictions on the effects of modifications of the systems. Because the properties of a biological system make it a complex system, as we have seen in the first chapter, reaching these challenges is really hard.

One of the main discipline which works at this aim is *Molecular Biology*: it had a great expansion and reached at really important discovery towards the middle of the twentieth century, when genetics began to explore the molecular structure of the gene. But during the last decades of that century it seemed no more enough and around the end of the century a new discipline born: *Systems Biology*, which proposes new tools and a different approach in studying biological systems.

2.1 Molecular Biology

The 1950s were the decade of the spectacular triumph of genetics: the elucidation of the physical structure of DNA has been hailed as the grater discovery in biology since Darwin's theory of evolution.

> On Feb. 28, 1953, Francis Crick walked into the Eagle pub in Cambridge, England, and, as James Watson later recalled, announced that "we had found the secret of life". Actually, they had. That morning, Watson and Crick had figured out the structure of deoxyribonucleic acid, DNA. And that structure a "double helix" that can "unzip" to make copies of itself confirmed suspicions that DNA carries life's hereditary information [James Watson & Francis Crick – by ROBERT WRIGHT].

For several decades, this triumphal success totally eclipsed the systems view of life.

The achievements of genetics brought about a significant shift in biological research, a new perspective: whereas cells were regarded as the basic building-blocks of living systems during the nineteenth century, the attention shifted from cells to molecules. Advancing in their explorations of the phenomena of life, biologists found that the characteristic of all living organisms were encoded in their chromosomes in the same chemical substance, using the same code script. After two decades the precise detailed of this code were unravelled. Biologist had discovered the alphabet of a truly universal language of life.

2.1.1 The Edges of Molecular Biology

This triumph of molecular biology resulted in the widespread belief that all biological functions can be explained in terms of molecular structures and mechanism. Thus most biologist have become fervent reductionists, concerned with molecular detailed. At the same time, the problems that resist the reductionistic approach of molecular biology become ever more apparent: while biologist know the precise structure of a few gene, they know very little about the ways in which genes communicate and cooperate. This means that, while molecular biology made enormous progress in understanding the structures and functions of many of the cell's subunits, it remained largely ignorant about the coordinating activities that integrate those operations into the functioning of the cell as a whole. Biological systems, in fact, have obvious both structure and organizational principles, and their behaviour cannot be understood either by "reading the DNA" (even

though in principle all the information is there) or by studying the biological components one by one or one level at a time. In few summarizing world: "*the whole is more then the sum of the parts*".

At now, even if the molecular biologists have been unravelling the functions of cellular components and networks, and the amount of molecular-level knowledge accumulated so far is absolutely amazing, yet we cannot say that we understand how a cell work: the process of understanding cellular components is far from finished, but it is becoming clear that simply obtaining a full part list will not tell us how a cell works. Rather, even for substructures that have been well characterized, there are significant difficulties in understanding how components *interact as systems* to produce the observed behaviour, both at static and dynamic level.

For example, at intracellular level, while an understanding of genes and proteins continues to be important, first of all the focus has to be on understanding a system structure. Because a biological system, as we have seen in the first chapter, is not just an assembling of genes and proteins, its properties cannot be fully understood merely by drawing diagrams of their interconnections. This diagram is an important first step but it is analogous to a static roadmap, whereas what we really seek to know are the traffic patterns, why such traffic patterns emerge, and how we can control them [22]. At these aims what we have to know is how the individual components of this diagram dynamically interact during operation. Similar problems occur also at each level of biological organization above the cellular level [8].

The complexity barrier between components and systems prevents us from predicting the behaviour of biological systems, and therefore from repairing them reliably. New concepts and new tools are clearly necessary to describe nature, tools for modelling and simulating biological complex systems. It enter in this context *Systems Biology*, an emergent discipline which goal is a predictive understanding of the whole.

2.2 Systems Biology

Around the year 2000, when the Institute of Systems Biology was established in Seattle, Systems Biology emerged as a movement in its own right, spurred on by the completion of various genome projects, by the large increase in "catalogue" from the *omics* (e.g. genomics and proteomics), by a growing understanding of how genes and their resulting proteins give rise to biological form and function, and by the accompanying advances in high-throughput experiments and bioinformatics. Systems Biology has grew up in recent years as an exciting new endeavour which aims at achieving a systems-level understanding of biological processes - and ultimatively whole cells and organisms: from the huge amounts of data that biologists collected, Systems Biology is building a science of the principles of operation of biological systems, based on the *integration and interaction between components*, i.e. on that interactions which are ultimately responsible for an organism's form and functions [1]. It is so a discipline that, instead of analysing individual components or aspects of an organism, focuses on all the components, and on the interactions among them, all as part of one system. To address the question, Systems Biology chooses modelling methods, which are implemented and then simulated through computational tools. Simulation is the process of using a developed model to analyse and predict the behaviour of the original system, doing experiment with this model. Because off the use of computational technique, in both modelling and simulating phase, Systems Biology is often called *in-silico* Biology.

Although Systems Biology believes that the essence of system lies in dynamics and it cannot be described merely by enumerating components of the system, at the same time it does not believe that only system's dynamic and structure is important without paying sufficient attention to diversities and functionalities of components structure. Both structure of the system and components plays indispensable role forming symbiotic state of the system as a whole. The effort provides a new approach for integrating quantitative data from a variety of sources, especially from genome-wide analyses, in conjunction with extensive use of a variety of different model: progress in Systems Biology is heavily dependent on a combination of experimental and

computational state-of-the-art techniques. Moreover, as the systems under study do not support an easy experimental access and analysis, models play an important role in gaining an insight into the systems' behaviour and structure [45]. In fact using knowledge from Molecular Biology, a systems biologist can causally model the biological system of interest and propose hypotheses that describe a system's behaviour.

Therefore Systems Biology is a new way of doing biology, starting with experimental knowledge, passing through *in-silico* modelling, and finally returning to biological experiments with the simulated results: it is so an approach that works if integrated with experimental biology. Figure 2.1 shows the combined application of experimental and computational tools.

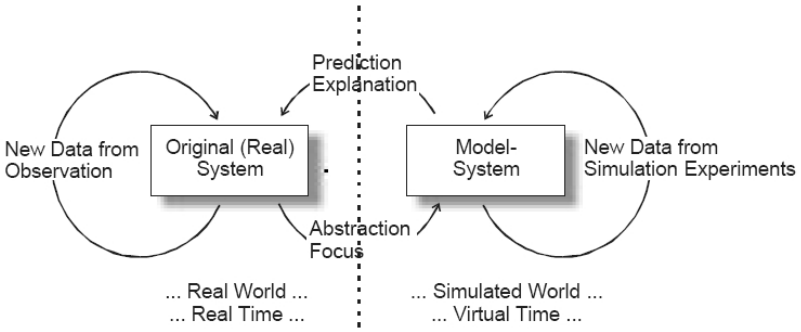


Figure 2.1: Systems Biology approach

Finally Systems biology is an interactive scientific approach in biological research that requires expert knowledge in several areas. Systems biology integrates and combines methods used in biology, mathematics, systems sciences and computer sciences. In order to be successful, close co-operation and the exchange of information between experimental research and theoretical computer-assisted simulation are necessary.

2.2.1 From Reductionism to Integration

Therefore, what Systems Biology propose is a shift from a reductionistic approach at understanding biological systems to an holistic approach. We want be more clear about the meaning of these two terms.

Reductionism

Reductionism can be defined as the belief that the behaviour of a whole or system is completely determined by the behaviour of the parts, elements or subsystems. In other words, if you know the laws governing the behaviour of the parts, you should be able to deduce the laws governing the behaviour of the whole.

From the first section we know that molecular biology strongly promoted the reductionistic approach, resulting in attribution of biological phenomena to the actions of genes.

The limitations of the reductionist model were shown even more dramatically by the problems of cell development and differentiation. In the very early stages of the development of higher organisms, the number of their cells increases from one to two, to four, etc., doubling at each step. Since the genetic information is identical in each cell, how can these cells specialize in different ways, becoming muscle cells, blood cells, bone cells, nerve cells and so on? This basic problem of development, which appears in many variations throughout biology, clearly flies in the face of the mechanistic view of life [7].

In summery reductionism has been highly successful in explaining some macroscopic phenomena, purely in term of the behaviour of constituent parts. However, this was predicted on the assumption that there were few parts and their interactions were simple, or that there were many parts but their interactions could be neglected. However the scope of reductionist approach is limited because these assumptions are not true in many systems of interest.

Holism

Systems theory, according to Complex Systems science, has always taken an anti-reductionist stance, noting that the whole is more than

the sum of the parts. In other words, the whole has "emergent properties" which cannot be merely reduced to properties of the parts.

Systems biology promote a holistic approach

Holism is the idea that all the properties of a biological system cannot be determined or explained by the sum of its component parts alone. Instead, the system as a whole determines in an important way how the parts behave.

The general principle of holism was concisely summarized by Aristotle in the Metaphysics: "The whole is more than the sum of its parts".

2.3 Modelling Methods

A model, as a tool for understanding, is an abstract representation, a schematic description of a system, theory, or phenomenon, that allows for investigation of the properties of the system and, in some cases, prediction of future outcome or studies of its characteristics. It is usually in the form of a set of objects and the relations between them.

Models represent aspects, a term that denotes a coherent set of properties or phenomena of biological interest. It is a skeleton, but not a replica of the real system, build with key components based on a mix of assumptions and known knowledge. It involves simplification, aggregation and omissions of details. The key to modelling is to identify the elements that can reflect chief global properties with incomplete information [12]. The *assumptions* that are under a model construction, condition or determine the relationship between models and the aspects they represent. These assumptions must be precisely documented and connected to the model for it to have meaning beyond the immediate use to which it has been put.

Modelling lies at the hart of Systems Biology. We can use experimental information to build a models at different biological scales, integrating them to create an orchestrated assemblage ranging from gross models of physiological functions through detailed models that build directly on molecular data. In this way these models should span from DNA and gene expressions to intracellular networks, to cell-to-cell and transmembrane signals, and through to the organ level.

2.3.1 A methodology for models construction

Usually, in order to understand a biological systems through modelling and simulation tools, we have to follow some advised steps, which are listed below and summarized in Figure 2.2.

1. Formulate the goal of the simulation study or research question, i.e. fixing the question addressed by the model.
2. Identify, explicate and justify the assumptions under the model.
3. Design the model in two consecutive steps:
   (a) coarse level → model concept;
   (b) detailed, formal level → model specification;
4. Select output values and measurements.
5. Select simulation software.
6. Implement the model.
7. Calibrate system and bug fixing;
8. Identify and do useful experiments.
9. Analyse simulation results [24].

To the purpose of building a model, the following methodology is used. Starting from an initial model, suggested by knowledge of regulatory mechanisms and available data, the behaviour of the system can be simulated for a variety of experimental conditions. Simulation attempts to predict the dynamics of systems so that the validity of underlying assumptions can be tested. Detailed behaviours of computer-executable models are first compared with experimental observation. Comparing the predictions with the observed experimental data gives an indication of the adequacy of the model. If the predicted and observed behaviour do not match, and the experimental data is considered reliable, the model must be revised. The activities of constructing and revising models of the regulatory network, simulating

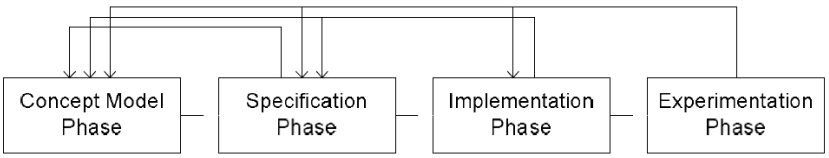


Figure 2.2: Main steps in a simulation study

the behaviour of the system, and testing the resulting predictions are repeated until an adequate model is obtained.

Models that survive initial validation can then be used *(1)* to make predictions to be tested by experiments; *(2)* to explore questions that are not amenable to experimental inquiry; *(3)* to analyse rapidly the effects of manipulating experimental conditions without having to perform complex experiments; *(4)* to verify hypothesis underlying the model that try to explain biological systems.

Diverse modelling methods are applied in the context of Systems Biology. They could be, fist of all, static or dynamic models. With a static model we have a photo of the system (e.g. the diagram of static interconnections, or the structure of an entity) at a specific instant of time, but we can't consider its evolution. In the follow we consider only dynamic models, which have the simulation phase with all the properties related to it. We can classify and characterise them taking into account three dimensions of modelling:

1. continuous and discrete;

2. quantitative and qualitative;

3. deterministic and stochastic;

These dimensions are not entirely independent nor are they exclusive. Many modelling approach are *hybrid* as they combine continuous and discrete, quantitative and qualitative, stochastic and deterministic aspects.

Another important aspect for the distinction of modelling approaches is at *which level* a model describe a system:

1. macro level;
2. micro level;
3. multiple levels of organization.

Explaining the properties of these different approaches in the follow, we may cite some specific tools or formalisms. We refer to the next chapter for the details.

2.3.2 Continuous and Discrete models

Continuous modelling approach

With a continuous model we assume that the variable of the systems change continuously during time. The continuous modelling approaches are perfectly suited for the reproduction of measured time-dependent trajectories and also easily allow the fitting of the model parameters. Continuous systems models are dominant type of model used in Systems Biology: they often use a set of differential equations.

The assumption underlying continuous models is still that the system behaves continuously with an infinite number of infinitely close state transitions in each time interval. The numerical integration merely serves to approximate this behaviour.

Discrete modelling approach

Often a cell's activity is perceived as being discrete rather then continuous, motivating the design of discrete systems models. In contrast to continuous systems models, discrete systems models assume only a finite number of state changes within a time interval. Depending on the time base that underlies the model, *discrete time stepped* approaches and *discrete event* approaches are distinguished. The former is based on time that advances in equidistant steps, whereas the latter allows to associate arbitrary time spans with each state of the system and thus is based on a continuous notion of time.

In the last one, i.e. in discrete event models, state transition functions define into which state to change triggered by external events, e.g. the collisions of species like enzymes and metabolites in a biological model, or triggered by the flow of time, e.g. after the time required for intra-molecular rearrangements.

Often systems can best be described by a combination of discrete and continuous models, e.g. if continuous processes exhibit discontinuities which require to switch from one continuous model to another one, or if leaving or entering a discrete phase depends on continuous processes that reach certain thresholds. Hybrid systems models combine continuous and discrete systems behaviour. Many modelling and simulating approaches for discrete and continuous systems have been extended to support hybrid systems models.

2.3.3 Deterministic and Stochastic models

Modelling is the process of structuring our knowledge about a given system. In this perspective, stochastic processes represent one means to express the uncertainty of our knowledge. A plethora of methods are dedicated to the problems of stochastic modelling, e.g. to estimate suitable distributions for random variates, and to interpret the results of the simulation runs.

From the view of the modelled system, integrating stochasticity into the models might also serve a slightly different purpose: randomness or "noise" arising from small numbers of molecules involved in processes like gene expression and regulation can directly be represented in the model. Although stochastic elements are often associated with discrete event models, they are also applicable to continuous system models. In Systems Biology, inclusions of stochastic elements for modelling continuous processes have gained ground recently. – e.g. chemical reaction equations are described by so called stochastic differential equations, which determine the probability with which a combination of molecules will react in a given time interval.

The stochastic discrete event models address specific constraints of continuous, deterministic models: concentrations do not necessarily change continuously, particularly if the dynamics of a small amount

of entities, like DNA molecules and plasmids, shall be modelled. In addition, sometimes, the dynamics of biological systems can be best approached in a stochastic manner, e.g. if the gene regulation is to be described, where stochastic fluctuations are abundant. The exact stochastic simulation approach is not practical for the simulation of metabolic processes, in which large numbers of molecules of the same kind are involved, due to the computational cost for the calculation of all individual molecular collisions.

Extensions of the approach overcome these difficulties and allow the stochastic simulation of systems composed of both intensive metabolic reactions and regulatory processes involving small numbers of molecules. The combination of stochastic discrete with continuous sub-models has stimulated the desire for an easy integration of stochastic aspects into continuous models. One common approach is to assume a normal distribution for key parameters of the differential equation system. The result is that stochasticity can now permeate the entire model [45].

2.3.4 Qualitative and Quantitative models

Quantitative

We define *quantitative models* that models whose variables are numerically scaled, in the case of differential equations the state space is given by real values vectors. This property is often associated with continuous models, even if continuous behaviour can also be described qualitatively.

Qualitative

We define *qualitative models* that models that allow predictions of qualitative properties of the dynamics of system's element interaction networks, that are invariant for a range of reaction mechanism and values of kinetic constants. The interest on qualitative models come from the following reasons. On the one hand, precise and quantitative information on reaction mechanisms and kinetic constants is not available for most networks of biological interest. In the other hand, in many situations predictions of qualitative rather then quantitative

dynamical properties are appropriate for gaining an understanding of the function of system's element interaction network [11].

2.3.5 Macro, Micro, Multi-level models

According to the complex systems properties of hierarchies, all biological systems are amenable to be represented as organised on different layers, ranging from genes and cells up to tissues, organs and organisms. Each level is essential to the general understanding of the system's wholeness, and it is autonomous with its own laws, pattern and behaviour. At the same time, no level can be understood in isolation independently of all the other levels, and the system as a whole can be understood only through the understanding and representation of all of its levels.

When we observe such a kind of complex organization, we may not, and in most cases we can't, analyse all hierarchical levels to understand the functioning of the biological system. We can so focus our attention at one level or more, depending on the problem.

Macro-level, continuous models – Macroscopic View

In a macro model and subsequent simulation, a complete system is tackled as one entity whose state variables are updated during simulation. Modelling, simulating and observation happens on one global level. The system is described by a set of state variables with their interdependencies, which can be expressed as rules, equations, constraints etc. All the simulations based on the macroscopic view are deterministic in nature. As a result, the system evolves along a fixed path from its initial state [42].

Typical representatives of this class are *differential equation* models which describe the time-dependent changes of the state variables, e.g. a biochemical system based on concentrations and reaction rates.

Focusing only on the population, we lose the representation of the individual and its locality, with the conditional and adaptive behaviour of each entity in its local environment [45].

Despite these limitations, macro simulation is used. Their advantages results from their relative simplicity and from their formal

aspect. First of all, in fact, differential equations are a really well understood and established framework, in which the complete model is documented concisely through formulas, and in which low number of parameters construction based on global input/output behaviour. With this approach, moreover, simulation experiments can be very fast (depending on the integration step).

Micro-level, discrete models – Microscopic View

Micro models are models that represent systems as comprising huge numbers of rather homogeneously structured entities. Only the behaviours of the individuals is explicitly modelled. The macro level of the system exists only as it aggregates results of the activities at micro level and is used for reflecting emergent phenomena, e.g. the development of specific spatial patterns. They do not have any behaviour of their own. Typical representative of this class are *cellular automata*.

The behaviour of the system is modelled by states changing at arbitrary points on a still continuous time scale. With discrete approaches models have emerged that integrate qualitative and stochastic aspects.

Multi-level models

Micro models often form only a transition to multi-level models, which describe a system at least at two different levels. Interactions are taking place within and between these levels: not only interdependencies at one organizational level but between different ones become of interest.

The importance of multi-level models has been emphasized for biological systems in particular, due to the great interplay that take place between different levels of hierarchical organization: "the whole is to some degree constrained by the parts (upward causation), but at the same time the parts are to some degree constrained by the whole (downward causation)". Moreover, the description of systems at different levels of abstraction and different time scales facilitates taking spatial and temporal structured processes into consideration.

Multi-level models are often defined as *hybrid* approaches: they are neither restricted to discrete models nor to continuous ones; they

can work deterministically or stochastically; they might be qualitative, quantitative or semi-quantitative.

2.3.6 Top-Down Versus Bottom-Up models

The studies of relational dynamics among many elements or parts of a system is different from both the top-down and the bottom-up approach.

Top-Down Approaches

In top down approaches we seek to analyse systems in comparatively general or high-level terms, lumping together subsystems in order to make the system easier to understand. With this approach an overview of the system is formulated, without going into detail for any part of it. The top-down model in fact, is often designed with the assistance of dark boxes that make it easier to bring to fulfilment but insufficient and irrelevant in understanding the elementary mechanisms.

To construct such approach is necessary a priory knowledge and attempt to disassemble it: we start with the intact system and we decompose it. In the typical top-down approach the behaviour at the bottom-up level is determined by the instruction from the top-level. The relations between the elements are fixed and each element preserve the same features.

Because with a top-down approach we must define behaviours of parts of the system so that they are consistent with the expected behaviour of the entire system, this approach not always works.

Bottom-up Approaches

In bottom-up approach, we study basic components and integrate the data to detect relevant patterns: the function of the components is well define (at least under a limited set of conditions), and is determined by detailed biochemical or molecular biological analysis.

In the bottom-up approach, the top level is generated by interactions among elements at the bottom level without instructions from the top level. Particularly, in this bottom-up approach the top level is no longer rigid, in contrast with the top-down approach, since the

level is not given in advance but is self-organized. Even when the top-level is self-organized, however, the relations between the elements and, after all, the state of the whole system, in general, is fixed.

The concept of bottom-up approach overlaps in some aspect to the concept of reductionism. This is why a bottom-up, data driven strategy will not work at all: we cannot build an understanding of biological systems from the understanding of the components alone; we must seek other approaches.

According to the principles of downward and upward causation, summarized in the "theory of two way causation", we cannot have organisms whose internal functioning flouts the rules of physics and chemistry (Up-ward causation). However, the laws of physics are completely insufficient to determine which shapes or organizations will evolve in the living world. Once a particular biological organization has emerged, it will strongly constrain the behaviour of its components (Down-ward causation). That means that high level phenomena are not reducible to physical laws but they must be consistent with them.

If this were true, then the modelling of some biological processes should not follow solely a bottom-up approach, hoping to go from simple laws to the desired phenomenon. At the same time it seems that to posit high-level of organizing principles and even downward causality is not enough because lost some physical and chemical laws governing the behaviour of the parts end the consequence of their interactions.

The top-down approach is to decompose the system to smaller parts. The bottom-up approach is to reconstitute elemental steps into larger parts. If the results of these approaches meet in the middle, and if they are consistent, we can be confident that we are on the right track. In other words, in studying biological complex systems, it was important to make the dynamics explicit for all levels. For example if genetic structure was changed in some way, it was important to know what happen at the cell level, but also at the level of the multicellular organism: there may be two or more levels of emergence that have to be explained because every level may influence the level below and above. In this way we have again underlay the importance of

multi-level models.

2.4 Simulation Methods

Simulation, as yet anticipated, is the process with which we can study the dynamic evolution of a model system, usually through computational tools. The Simulation phase is first of all used to validate the model, testing input/output behaviour; then simulation is used to do *virtual experiment* in a way that would be very hard to do if we had to actually do the experiment in real life. These experiments are useful to *(1)* produce quantitative correct *prediction* about the system's behaviour, depending on its input values; or *(2)* produce qualitative *explanations* about the system's behaviour, that is qualitative significant results that are sufficient for understanding the reaction of the system to input values [23].

2.4.1 Continuous or Discrete simulation

Although continuous systems models and subsequent simulations are the dominant type of models being used in Systems Biology, stochastic discrete event models are recently gaining ground as well.

In discrete event simulations, an event list is administrated by simulator, the time of the head is set on the event executed, produces other events that are inserted into the event list. Time is set to next event. They address specific constraints of continuous, deterministic models: concentrations do not necessarily change continuously, particularly if the dynamics of a small amount of entities.

Chapter 3

Modelling and Simulating Approaches in Systems Biology

In this chapter we are going to give an overview of formalisms proposed in the literature and to discuss modelling and simulation techniques appropriate for each of the formalisms. Formalisms to be discussed include directed graphs, Bayesian networks, Boolean networks, ordinary and partial differential equations, stochastic master equations, and some computational formalisms as Petri Nets, process algebra and multi-agent systems. It will come as no surprise that the review is not meant to be exhaustive: this is nor our main challenge in this thesis.

The diversity of modelling approaches, applied in Systems Biology, illustrates and suggests that, depending on the biological system, the available data and knowledge about the system, and the objective of the simulation study, modelling approaches are chosen deliberatively on demand and thus address the diverse needs of modelling and simulation in Systems Biology

Each of these models has different strengths and weakness. So the question is to be asked what do certain approaches offer in modelling biological systems when compared to others.

3.1 Structural model

The first step in modelling a biological system is often give a structural description of its components and of the interactions between these components, to understand how its parts are dynamically connecting to form a network. This is important for several reasons. For instance, if certain structural and organizational properties can be shown to imply specific dynamical properties, then the behaviour of the system could be inferred at least by verifying whether the network posses these structural and organizational properties.

3.1.1 Chemical Reactions

Chemical reactions are the lingua franca of biological modelling. They provide a unifying notation by which to express arbitrary complex chemical processes. Specifying chemical reactions is so fundamental, especially studying intracellular level, that the same set of chemical reactions can lead to different computational or mathematical models. In this sense, representing processes by chemical equations is more basic than using either differential equations, or stochastic processes, or something else, to run simulations to make predictions.

A general chemical reaction, such as

$$n_a A + n_b B \xrightarrow{k} n_c C + n_d D$$

states that some molecules of type A react with some of type B to form molecules of type C and D. The terms to the left of the arrow are called *reactants*; those on the right are called *products*. The n terms are called *stoichiometric coefficients* and are small integers. The value k on the reaction arrow is a *rate constant*. Chemical reactions do not occur instantaneously, but rather take some time to occur. The value k is a way of specifying the amount of time a reaction takes.

3.1.2 Directed Graphs

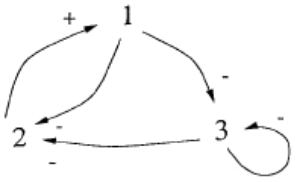
Most of biological systems, and often, most of levels of hierarchical organization of the system, involve entities which interact along complex networks. Probably the most straightforward way to model the

structure and organization of such a kind of networks is to view it as a *directed graph*.

A directed graph G is defined as a tuple $\langle V, E \rangle$, with V a set of vertices and E a set of edges. A directed edge is a tuple $\langle i, j \rangle$ of vertices, where i denotes the head and j the tail of the edge. The vertices of a directed graph correspond to elements of the network, while the edges denote interactions among these elements.

For instance, in a transcriptional regulatory network, nodes would represent genes with edges denoting the interactions between them. This would be a directed graph because, if gene A regulates gene B, then there is a natural direction associated with the edge between the corresponding nodes, starting at A and finishing at B. Directed graphs also arise in the study of neuronal networks, in which the nodes represent individual neurons and the edges represent synaptic connections between neurons.

The graph representation of a network can be generalized in several ways. The vertices and edges could be labelled, for instance, to allow information about the nature of interactions to be expressed. By defining a directed edge as a tuple $\langle i, j, w \rangle$ where w is a weigh which indicates how much strong is the interaction, and if it is a sign it can be indicated whether i is activated or inhibited by j [10].



(a)

$V = \{1, 2, 3\}$

$E = \{\langle 2, 1, - \rangle, \langle 3, 1, - \rangle, \langle 1, 2, + \rangle, \langle 2, 3, - \rangle, \langle 3, 3, - \rangle\}$

(b)

Figure 3.1: *(a)* Directed graph representing an interactions network and *(b)* its definition

The representation of a network as a graph allows the analysis of its structural properties by means of graph-theoretical techniques. The global connectivity properties of the network can, for instance, be described by the average degree and the degree distribution of the

vertices. The *degree* k of a vertex indicates the number of edges to which it is connected. $\langle k \rangle$ denotes the *average degree* and $P(k)$ the *degree distribution* of the graph, which measures the proportion of nodes in the network having degree k ($P(k) = \frac{N(k)}{N}$). The properties give an indication of the complexity of the graph and allow different types of graphs, and therefore of networks, to be distinguished. We cite in the following the most common ones:

1. *Random graph*: the vertices typically have $\langle k \rangle$ edges and the vertices having significantly more or less edges than $\langle k \rangle$ are extremely rare;

2. *Scale-free graph*: these types of graphs are inhomogeneous, in that most of the vertices have few edges, whereas some vertices, called (*hubs*), have many edges;

3. *Hierarchical graphs*: these types of graphs describes modular networks, i.e. they are formed by the repetition of nodes' cluster.

Observations

A number of operations on graphs can be carried out to make biologically relevant predictions about regulatory systems.

For instance, a search for paths between two elements, for instance, may reveal missing regulatory interactions or provide clues about redundancy in the network. Furthermore, cycles in the network point at feedback relations that are important for homeostasis and differentiation. Again, global connectivity characteristics of a network, such as the average and the distribution of the number of regulators per element, give an indication of the complexity of the network. Loosely connected subgraphs point at functional modules of the regulatory system of which the behaviour could be considered in isolation.

Alongside the potential benefits of applying graph theoretical methods in molecular biology, it should be emphasized that the complexity of the networks encountered in cellular biology and the mechanisms behind their emergence presents the network researcher with numerous challenges and difficulties. The inherent variability in biological data, the high likelihood of data inaccuracy and the need to incorporate dynamics and network topology in the analysis of biological

systems are just a sample of the obstacles to be overcome if we are to successfully understand the fundamental networks involved in the operation of living cells. The use of graphs alone is not always advised because it implements a static model which cannot successfully describe the dynamic of biological networks.

For further details of this modelling approach, interested readers can refer to a really exhaustive and well done report [27].

3.2 Boolean Networks

The simplest approach to characterizing the dynamics of biological networks is a Boolean model, which is often applied to studying molecular interaction networks inside a cell, as a method that allow predictions of *qualitative* properties of such systems [11], i.e. dynamical properties that are invariant for a range of reaction mechanism and values of kinetic constants. The qualitative properties express the intimate connections between the behaviour of the system and the structure of the network of molecular interactions, independently from the quantitative details of the latter. Also graphs are a qualitative approach.

To explain how a Boolean network can model a biological network, we give an example about again gene regulatory networks. As a first approximation, the state of a gene can be described by a Boolean variable expressing that it is active (**on, 1**) or inactive (**off, 0**) and hence that its products are present or absent. The change in gene expression can be described by making the assumption that the change in activation state of a gene is determined in a combinatorial fashion by the activation of other genes, in particular genes encoding for regulatory proteins. Interactions between elements can be represented by *Boolean functions* which calculate the state of a gene from the activation of other genes. The result is a Boolean network, an example of which is shown in Fig. 3.2. Recent comprehensive reviews of the use of Boolean network models can be found in Kauffman's book.

Let the n-vector $\hat{\mathbf{x}}$ of variables in a Boolean network represent the state of a regulatory system of n elements. Each $\hat{x}_i$ has the value 1 or 0, so that the state space of the system consists of 2^n states. The

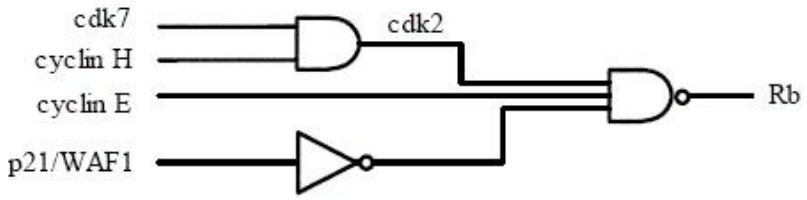


Figure 3.2: Boolean networks

state $\hat{x}_i$ of an element at time-point $t+1$ is computed by means of a Boolean function or rule $\hat{b}_i$ from the state of k of the n elements at the previous time-point t. (Notice that k may be different for each $\hat{x}_i$) The variable $\hat{x}_i$ is also referred to as the output of the element and the k variables from which it is calculated the inputs. In summary, the dynamics of a Boolean network describing a regulatory system are given by

$$\hat{x}_i(t+1) = \hat{b}_i(\hat{\mathbf{x}}(t)), 1 \leq i \leq n. \tag{3.1}$$

Transitions between states in a network are *deterministic*, with a single output state for a given input, and *synchronous*, in the sense that the outputs of the elements are updated simultaneously.

A sequence of states connected by transitions forms a trajectory of the system. Because the number of states in the state space is finite, the number of states in a trajectory will be finite as well. More specifically, all initial states of a trajectory will eventually reach a steady state or a state cycle, also referred to as point attractor or dynamic attractor, respectively. The states that are not part of an attractor are called transient states.

Probabilistic Boolean networks

Recently Boolean networks have been generalized to probabilistic Boolean networks (PBNs) to facilitate the incorporation of uncertainty in the

model and to represent cellular context changes in biological modelling. In essence, a PBN is composed of a family of Boolean networks between which the PBN switches in a stochastic fashion. In whatever framework Boolean networks are studied, their most important attribute is their attractors. Left to run, a Boolean network will settle into one of a collection of state cycles called attractors. The set of states from which the network will transition into a specific attractor forms the basin of the attractor. The attractors represent the essential long-run behaviour of the network. In a classical Boolean network, the network remains in an attractor once there; in a Boolean network with perturbation, the states form an ergodic Markov chain and the network can escape an attractor, but it will return to it or a different attractor unless interrupted by another perturbation; in a probabilistic Boolean network, so long as the PBN remains in one of its constituent Boolean networks it will behave as a Boolean network with perturbation, but upon a switch it will move to an attractor of the new constituent Boolean network. Given the ergodic nature of the model, the steady-state probabilities of the attractors are critical to network understanding. Heretofore they have been found by simulation.

Observation

This formalism is typically used to obtain a first representation of a complex system with many components until such time more detailed data become available. Its attractiveness is based on the intuitiveness of the representation of molecular regulation networks by Boolean functions. A Boolean model is advantageous in its simplicity and it does not require detailed data on how cellular components interact. Despite their apparent simplicity, Boolean models can provide many insights into the qualitative behaviour of the underlying system. For instance, Kauffman [20] has successfully employed Boolean models to explore self-organization phenomena and their implications in evolution. It is really well suited for modelling regulatory networks and signalling pathways. At now there is not a strong literature that experiment such a kind of approach in modelling biological system at higher levels. However, the classical formalisms make strong simpli-

fying assumptions, in particular the use of binary value for activation and synchronous transitions.

3.3 Bayesian Networks

Bayesian network is a form of probabilistic graphical model In the formalism of Bayesian networks, the structure of the system is modelled by a directed acyclic graph $G = \langle V, E \rangle$. The vertices $i \in V$, $1 \leq i \leq n$, represent elements and correspond to random variables X_i. For instance, if i is a gene, then X_i will describe the expression level of i. For each X_i, a conditional distribution $p(X_i|parents(X_i))$, is defined, where $parents(X_i)$ denotes the variables corresponding to the direct regulators of i in G. The graph G and the conditional distributions $p(X_i|parents(X_i))$, together defining the Bayesian network, uniquely specify a joint probability distribution $p(\mathbf{X})$, where

$$p(\mathbf{X}) = \prod_{i=1}^{n} p(X_i|parents(X_i)). \tag{3.2}$$

Observation

Statistical inference of graphical models has become an important tool in the reconstruction of biological networks of the type which model, i.e. gene regulatory interactions. In particular, the construction of a score-based Bayesian posterior density over the space of models provides an intuitive and computationally feasible method of assessing model uncertainty and of assigning statistical confidence to structural features [37]. Moreover the mathematical rules of probability theory are consistent rule for conducting plausible reasoning processes if:

1. we must reason with incomplete prior knowledge of and limited data on the biological system under study;

2. we must be able to update our inferences taking into account new data, without having to revisit the entire reasoning process;

as reasoning in biology imposes.

3.4 Nonlinear Ordinary Differential Equations (ODEs)

Being arguably the most widespread formalism to model dynamical systems in science and engineering, ordinary differential equations (ODEs) have been widely used to analyse biological systems, at each level of abstraction. At molecular level, we can model the dynamic of signalling and metabolic pathway taking concentrations of RNAs, proteins, and other molecules as time-dependent variables with values contained in the set of nonnegative real numbers. Regulatory interactions take the form of functional and differential relations between the concentration variables. At higher levels we can model cell-to-cell interactions, cell population, and cell system network, such as immune system, nervous system and endocrine system, considering always as variables concentrations of chemical molecule that are involved in these processes of communication and interaction.

In this way, at each level we are analysing, we have to identify some variables that can describe the state of the system, and that we call *state variables*.

To define a model with ODEs we can follow two methodologies:

1. *forward modelling*: structural model and parameters' value are known. We can directly define system of differential equations and then, through analytical or numerical methods, we can easily solve the problem;

2. *inverse modelling / reverse engineering*: if experimental data don't include parameters' value, a direct solution of the problem is essentially impossible, as they are normally hugely underdetermined and do not have an analytical solution. The normal approach is thus an iterative one in which a candidate set of parameters is proposed, the system run in the forward direction, and on the basis of some metric of closeness to the desired output – experimental data – a new set of parameters is tested. Eventually a satisfactory set of parameters, and hence solutions, will be found.

In the first case, to write a system of ODE, we start with a structural model, in which the reactions and the effectors are known. To get from a wiring diagram, to a set of ODEs, we must think about a network as a dynamical system whose state is changing from one moment of time to the next. We assign to each node in the diagram a single state variable, $X_i(t)$, that is a *macroscopic* collective variable of such a system and that are, in the most of cases, the concentration of species i. The collection of values of all these state variables $\{X_1, X_2, ..., X_n\}$ denote a complete set of variables to define the *instantaneous state* of the system $\mathbf{X}$. The time evolution of $X_i(t)$ will take the form, through a mathematical expression:

$$\frac{dX_i}{dt} = F_i(X_1, X_2, ..., X_n; \gamma_1, \gamma_2, ..., \gamma_m) \tag{3.3}$$

where F_i may be complicated functions of the state variables, and $\gamma_1, \gamma_2, ..., \gamma_m$, are some parameters present in the problem whose variation influence the evolution of the system and which can be modified by the external world. We call them *control parameters* [35].

In view of the diversity of biological phenomena, we expect that the structure of the function F_i will depend in a very specific way on the system considered and on the type of precess going on in this system. However, only certain basic features can be sorted out of this apparently bewildering variety, and the function will help us tackle complex phenomena in a systematic fashion.

The system of differential equations that we have defined, describes how the concentrations of the molecules involved in the process under study, changes over time due to its interactions with the other species in the network. The rate of each reaction must be represented by a *kinetic rate law*, which will have one or more *rate constants* associated with. By assigning specific values to these rate constants, we fine-tune general rate laws to particular reactions. The set of all rate constants needed to describe the reactions in a molecular interaction network is part of the *parameter set*, we define before. Because the kinetic rate laws on the right side of (3.3) equation are often nonlinear functions of the state variables, the function F_i is often *nonlinear*.

We can also extend the system including concentrations $\mathbf{u} \geq 0$ of

molecules which are inputs for the system as drugs, signal molecules, metabolites.

Notice that the ODEs tell us how each state variables is changing with respect to time; they do not tell us the value of X at any specific time t. Solving the differential equations is to find these functions, $X_i(t)$, for each variable i of the system. In order to solve equation (3.3) we must first prescribe a set of initial conditions $\{X_1(0), X_2(0), ..., X_n(0)\}$.

We can imagine three types of solutions of a system of ODEs.

1. Analytical solution:

   under very special circumstances, i.e. when the function F_i is linear, it is possible to write the solution of a set of ODEs in terms of exponential functions, $\exp(\lambda_i t)$, and harmonic functions, $\sin(\omega_i t + \phi_i)$.

2. Numerical solution:

   Alternatively, one can take recourse to numerical techniques. In numerical simulation, the exact solution of the equations is approximated by calculating approximate values $\{X_1, X_2, ..., X_n\}$ for $\mathbf{X}$ at consecutive time-points $t_0, t_1, ..., t_m$. A variety of computer tools specifically adapted to the simulation and analysis of biochemical reaction systems are available, such as GEPASI [29].

3. Qualitative solution:

   Whereas numerical integration of the ODEs gives us quantitative information about solution, sometimes we are more interested in answer qualitative questions, like – what will the network do for $t \to \infty$, that is characterize the stable attractor of the system – or – how will the long-term behaviour of the network change if I double the rate of synthesis of a specific protein, that is characterize the dependence of the stable attractors on any parameter in the ODEs [9].

Once we have the *time course* of each variables, we can design *state space* of the system of ODEs. At each point of this space, differential equations define a vector that tell us which direction and how far dynamical system will move over the next small increment of time Δt.

The collection of vectors describing the behaviour of the system at each time point, is called *vector field*. A solution of ODEs is just a curve that stars at some initial point and follows the vector field. On the state space we can identify one or more attractors.

Simulation of the functioning of a system is often complemented by *bifurcation analysis* tools, to investigate the sensitivity of steady states and limit cycles to parameter values. From a practical perspective, well-polished computational techniques (e.g., bifurcation analysis - analysis of qualitative changes in the dynamics of a system caused by the variation of some system parameters) and software tools (e.g., Xppaut at
`http://www.math.pitt.edu/ bard/xpp/xpp.html`) are available for high-level analysis of system dynamics using ODEs.

For further details and examples of this modelling approach, we suggest reading [9], a chapter of an exhaustive book on modelling tools used for biochemical network.

Observations

Although the great importance and usefulness of ODEs approach to biological models cannot be denied, we should not lose sight of the fact that the physical basis for this approach leaves something to be desired. Not all systems can be modelled with differential equations. Specifically differential equations assume that: *(i)* concentrations are well-defined quantities, *(ii)* rate constants are well-defined quantities, *(iii)* the system is spatially homogeneous, *(iv)* concentrations vary deterministically over time and *(v)* concentrations vary continuously and continually. In the following we are going to explain better these points.

3.4.1 Assumption of spatially homogeneous system

A system of ordinary differential equations is a "well-stirred" chemical reactor, so that component concentrations don't vary with respect to space. This assumption hardly seems appropriate for an intact cell. Whether it is a good approximation or not depends on the time and

space scales involved. To bring an example, we know that molecular diffusion is sufficiently fast to mix proteins throughout a yeast-size cell in less then a minute. If we are interested in cell cycle precesses, which have a time scale of hours, then the "well-stirred" assumption is justified.

However, in many cases, one may assume that the processes considered are synchronous in all parts of biological objects, and therefore dependence on space coordinates is absent. In these cases we usually deal with ODE.

When spatial information is required, then partial differential equations would be indicated.

3.4.2 Assumption of continuous and deterministic system

The second basic assumption under a system of ordinary differential equations, is that variables are continuous function of time and vary deterministically. These assumptions may be valid only if the number of entities, such as molecules, of each species in the reaction volume is sufficiently large (say, thousands of entities each, at least).

In a biological cell, these underlying assumptions for the ODEs are often violated: a realistic model must take the inherent randomness into account and therefore need to be of stochastic nature. Stochasticity becomes more pronounced with the decrease in the number of molecules. As the concentrations of the reacting species increase the fluctuations become less prominent and tend towards the deterministic solution. This fact is important from the computational complexity point of view as stochastic simulations are more expensive to run compared to deterministic Ordinary Differential Equations (ODE)-based simulations.

When stochastic behaviour has to be modelled, then stochastic differential equations would be indicated.

Granted these two simplifying assumptions, the ODEs are a very useful language in which to express mathematically the dynamical consequences of a biological network. If then the mathematical consequences of the mechanism do not agree with observations, we must

search for the problems in out hypothesis. If the consequences agree with the observations, then we can have some confidence that we are on the right track to understanding the mechanism.

Finally, remembering the classification that we introduced in the previous chapter we can say, without doubts, that continuous systems models can easily be translated into a set of differential equations, independently of being defined as graphs, or as a set of chemical reactions.

3.5 Partial Differential Equations (PDEs)

Differential equations of the form (3.3) describe processes while abstracting from spatial dimensions. As mentioned before, the systems of interest are assumed, implicitly, to be spatially homogeneous.

There are situations in which these assumptions are not appropriate. From cells to tissues and organisms, biological systems display spatially inhomogeneous structures. All processes, in fact, develop in time and space. It might be necessary, for instance, to distinguish between different compartments of a cell, say the nucleus and the cytoplasm, and to take into account the diffusion of regulatory proteins or metabolites from one compartment to another. Again, gradients of protein concentrations across cell tissues are a critical feature in development processes.

The introduction of time delays for diffusion effects allows some aspects of spatial inhomogeneities to be dealt with, while preserving the basic form of the rate equations. In the case that multiple compartments of a cell, or multiple cells, need to be explicitly modelled, a more drastic extension of (3.3) becomes necessary.

Given a continuous variable $l \in [0, \lambda]$, where λ represents the size of the system, let's define the state of the system X as functions of both t and l. The time variation of the concentration of each substance X_i, is computed trough a partial differential equation, in the form:

$$\frac{\partial X_i}{\partial t} = F_i(\mathbf{x}) + \delta_i \frac{\partial^2 X_i}{\partial l^2} \tag{3.4}$$

where δ_i is the diffusion constant for the species i.

In order to solve equation (3.4) we must prescribe other then a set of initial conditions $\{X_1(0), X_2(0), ..., X_n(0)\}$, also a set of boundary conditions in $l = 0$ and $l = \lambda$. If it is assumed that no diffusion occurs across the boundaries $l = 0$ and $l = \lambda$, the boundary conditions become:

$$\frac{\partial^2 X_i}{\partial l^2}(0,t) = 0 \; and \; \frac{\partial^2 X_i}{\partial l^2}(\lambda,t) = 0 \tag{3.5}$$

Excellent introduction into PDEs and their applications in biology can be found in [19].

Observation

Continuum description of a system gave by partial differential equations is appropriate if one is interested in the dynamics on scales that are large compared to molecular or cell, in general unit of the system under study, length scales. In such a description, the discrete nature of the single entity forming the system is neglected. Instead, the state of the system is given in terms of continuous functions of space and time. On smaller scales, the discrete nature of the components cannot be neglected and a stochastic and discrete description is required, where we consider both inhomogeneous particle distribution and the stochastic aspects of biological events (next section) [26].

3.6 Stochastic Modelling

Despite their broad applications, ODE-based kinetic models are criticized because they implicit assume that the time evolution of the concentrations of interacting cellular species is both continuous and deterministic, particularly for intracellular processes.

In the first place, the time evolution of a chemical reacting system is not a continuous process, because molecular population levels obviously can change only by discrete integer.

In the second place, cellular events, at both intra- and extracellular level, are triggered by random collisions between molecules. If each type of event occurred numerous times per generation, this

randomness could possibly average out and the cells or cellular population could behave deterministically. But many proteins are expressed at nanomolar levels, which correspond to only tens or hundreds of molecules per cell. As a consequence of the small numbers of the interacting species, many central cellular reactions, occur so infrequently that substantial relative fluctuations arise spontaneously. By affecting the rates of other reactions, these fluctuations can propagate through networks and spread to any cellular process, and then also to higher level of organism [17].

Stochasticity in the dynamics arises in one of the two following ways: *intrinsic stochasticity* is inherent to the system, arising due to the relatively small number of reactant molecules, whereas *extrinsic stochasticity* originates due to random variation of one or more environmental factors, e. g. temperature and concentrations of the reactant species.

Many aspects of life in an individual cell are therefore best understood probabilistically. Deterministic in nature, ODE models will fail to predict such fluctuations. For this reason, some researchers have questioned the use of deterministic simulations in characterizing the behaviours of biological systems, and suggested using *stochastic simulations* instead. Several algorithms are available for carrying out stochastic simulations [36].

For stochastic modelling, it is feasible to keep track of every molecule in the system. *Concentrations changes of discrete number of molecules* corresponding to certain reaction events. Also, in contrast to the deterministic system, *the change is random or stochastic in nature.* In some literature the approaches that have these properties implements a *mesoscopic level* model/view.

The following paragraphs of this section outline some of the representative algorithms along with a comparative study of strengths and limitations of each algorithm [30].

3.6.1 Chemical Master Equation

The temporal behaviour of a spatially homogeneous mixture of molecular species can be described by a Chemical Master Equation [18]. Chemical Master Equation is a form of mathematical formalism that

describes the transition of the system from one state to another state using probabilistic methods. Before introducing the Master Equation, we first define the following notations:

- $\mathbf{M}$ = number of reactions
- $\mathbf{N}$ = number of species
- $\mathbf{X} = [X_1, X_2, ..., X_i, ...X_N]$ = number of molecules of species i in the system – $i = [1, 2, ..., N]$
- $p(\mathbf{X}, t)$ = probability of the system in state X at time t
- c_j = stochastic kinetic constant for reaction j – $j = [1, 2, ..., M]$
- R_j = reaction j – $j = [1, 2, ..., M]$
- $\alpha_j \Delta t$ = probability of R_j happening in time (t, Δt) given that the system is in the state $\mathbf{X}$ at time t.
- $\beta_j \Delta t$ = probability that the system is one R_j reaction removed from the $\mathbf{X}$ and undergoes the R_j reaction in time (t, Δt).

Given the notations, we can describe the evolution of $p(\mathbf{X}, t)$ in terms of the rates α and β as follows:

$$p(\mathbf{X}, t + \Delta t) = p(\mathbf{X}, t) \left(1 - \sum_{j=1}^{M} \alpha_j \Delta t \right) + \sum_{j=1}^{M} \beta_j \Delta t. \tag{3.6}$$

The first term on the right hand side of (3.4) represents the probability at which $\mathbf{X}$ remains its state, whereas the second term is the probability at which $\mathbf{X}$ undergoes one reaction in time (t, $t + \Delta t$). Reorganizing (3.4), and taking the limit as $\Delta t \to 0$, gives the final form of Master Equation (3.5). Notice that the transition of the state of the system is described through changes of the probability of the system being in a certain state, $p(\mathbf{X}, t)$. Hence, the inherent stochasticity of the system is mathematically formalized in this context:

$$\frac{\partial p(\mathbf{X}, t)}{\partial t} = \sum_{j=1}^{M} (\beta_j - \alpha_j p(\mathbf{X}, t)). \tag{3.7}$$

The Master Equation approach tries to write a system of equations for every possible transition state and solve them simultaneously. Solving that equation gives the complete probability distribution at any point in time. Even better, it is linear differential equations with constant coefficients, so one can actually solve it.

There is of course a catch: Chemical Master Equation needs one variable for each possible state of the system. For all but the simplest systems, this number of variables becomes huge, and so one cannot even write out the full master equation, let alone solve it. Generating a single trajectory is significantly easier: ones need to generate a sequence of state transitions and the times of which they occur. However, when the dimensionality of problem increases, the possible trajectories explode combinatorially and the problem becomes intractable.

In view of this limitation, Gillespie devised a better way of generating such trajectories [18]. Instead of writing the whole Master Equation explicitly, the Gillespie Algorithm generates trajectories by picking reactions and times according to the correct probability distributions so that the probability of generating a given trajectory with the simulation algorithm is exactly the same as the solution of the Master Equation [15].

Two different classes of stochastic simulation algorithms exist for chemically reacting system, namely the Gillespie Algorithm and StochSim Algorithm. The following sections will describe these algorithms in detail, stating the strengths and limitations of each. Various improvement and enhancement schemes proposed by

3.6.2 Gillespie algorithms

In 1976 Gillespie developed a discrete stochastic simulator algorithm to solve the Chemical Master Equation based on the assumptions that the system is homogeneous and well mixed.

The Gillespie algorithm makes time steps of variable length; in each time step, based on the rate constants and population size of each chemical species, one random number is choose which reaction will occur, and another random number determines how long the time step will last.

At each time step, the chemical system is exactly in one state.

The idea is to directly simulate the time evolution of the system. Basically, the algorithm determines the nature and occurrence of the next reaction, and consequently of the next transition, given that the system is in state α at time t. Given a system with total number of reaction channels **N** and total number of species **M**, there are at most **N** possible transitions from a given state. The key is to choose random number using a computer random number generator, and use those to pick transitions.

Gillespie proposed two methods for accomplishing the simulation.

Consider a system of **N** reaction. Gillespie propose two *exact stochastic simulation* algorithms

The Gillespie algorithm is by far the most popular. Following a Monte Carlo procedure, the Gillespie algorithm predicts the time evolution of the system by determining when and in what order the next reaction is going to occur. This algorithm has a rigorous theoretical foundation, and is shown to give exact solution for a network of elementary reactions occurring in a well-stirred environment. It often generates dynamics drastically different from the prediction by deterministic simulations, particularly when some reactions have nonlinear terms in their rate expressions.

Although the Gillespie algorithm reveals stochastic fluctuations resulting from small molecular numbers, several outstanding questions make it unclear whether or to what extent it is more appropriate than a deterministic approach in modelling cellular reaction networks. Also, stochastic simulations by the Gillespie algorithm are often much more time consuming than deterministic simulations. In fact, the computation time of this algorithm approximately scales with the frequency of the reaction events: the more reactions there are, or the more molecules there are, the longer the computation will take for a given simulated time span.

One way to model coupled chemical reactions stochastically at a mesoscopic level is to the use the Stochastic Simulation Algorithm (SSA) proposed by Gillespie. This Monte Carlo algorithm yields a correct realization of the process, but the computing time required to approximate the probability distribution of the species in the system is often dictated by the reactions involving the molecules with the largest copy numbers or the fastest reaction rates. They may well

be the components where the stochastic description is the least important. The convergence rate is also slow for this method and it can be computationally cumbersome to obtain detailed information of the probability distributions when the number of different reacting molecules is large.

3.6.3 StochSim algorithm

The computer program StochSim was written by Carl Morton-Firth as part of his PhD work at the University of Cambridge [34]. The program provides a general purpose biochemical simulator in which individual molecules or molecular complexes are represented as individual software objects: the algorithm treats the biological components, for example, enzymes and proteins, as individual objects interacting according to probability distribution derived from experimental data.

Reactions between molecules occur stochastically, according to probabilities derived from known rate constants. In every iteration, a pair of molecules is tested for reaction. Due to the probabilistic treatment of the interactions between the molecules, StochSim is capable of reproducing realistic stochastic phenomena in the biological system.

Simulation time is quantised into a series of discrete, independent time-slices, the size of which is determined by the most rapid reaction in the system. At the start of the simulation, the user assigns the maximum number of molecules, **N**, the system will use. In each time-slice, one molecule is selected at random from N possibilities (the probability of selection of each molecule is 1/N). Then, another object, either a molecule or dummy molecule, is selected at random from N possibilities. Another random number is then generated and used to see if a reaction occurs. The probability of a reaction is retrieved from a look-up table and if the probability exceeds the random number, the particles do not react. On the other hand, if the probability is less than the random number, the particles react, and the system is updated accordingly. The next time-slice then begins with another pair of molecules being selected.

Molecules that can exist in more than one state can be encoded in the program as a "multistate molecule" with a series of binary flags. Each flag represents a state or property of the molecule, such as a

conformational state, ligand binding, phosphorylation, methylation, or other covalent modification. The flags specify the instantaneous state of the molecule and may modify the reactions it can perform. For instance, a multistate molecule may participate in a reaction at an increased rate as a result of phosphorylation, or fail to react because it is in an inactive conformation. The flags themselves can be modified in each time-slice as a result of a reaction, or they can be instantaneously equilibrated according to a fixed probability. The latter tactic is used with processes such as ligand binding or conformational change that occur several orders of magnitude faster than other chemical reactions in the system.

Under special circumstances when the number of reactions is small and the number of molecules is large, the Gillespie algorithm is more efficient than the StochSim algorithm [33].

3.7 Computational Models

Also computer science techniques and formalisms are used to build, analyse and interpret biological models. We are now going to briefly survey these computational approaches that exploit the similarities between networks of biochemical components (molecules, cells or other) and networks of computing processes.

One of the fundamental dimension which characterize a computational model, is the possibility to have a large number of components that work independently or concurrently, and also interact with each other from time to time through different form of communication methods. Biological molecular systems, as we know, may be viewed as complex concurrent processes. Cellular mechanisms are complex and dynamic: multiple genes are transcribed, multiple types and copies of transcripts are translated to proteins, and those proteins partake in multiple signalling processes and metabolic reactions in a concurrent fashion. On higher level, huge amounts of different cells work concurrently inside a tissue to achieve complex goals. Conceptually, because off these properties of biological systems, it seems interesting to treat them and their components as *concurrent processes* and use techniques from the global computing field to study their behaviour.

For example, a network of cells can be seen as a computing machinery, made of processing components which interact and cooperate to achieve a common goal.

Petri Nets, *process calculi* and *agent-based* and *Multi-agent systems* are some of these computational models, which are best known in the context of computer and engineering. The most comprehensive works used Petri nets for representation, simulation and analysis of cellular pathways. For other reasons particularly promising is the use of process calculi, which provide the basis to study in a more systematic way hypotheses on properties of complex systems of biochemical reactions.

Instead the adoption of agent technologies and multi-agent systems constitutes an emerging area in Systems Biology. The use of agents in computational and systems biology suggest the design of agent-based systems, tools and languages for modelling the biological processes themselves. More specifically, an agent can be considered a high-level software abstraction that provides a convenient and powerful way to describe a complex software entity in terms of its behaviour within a contextual computational environment. Agents provide designers and developers with a way of structuring an application around autonomous, communicative elements [31]. Next chapter is all dedicate at this last formalism.

3.7.1 Petri Nets

One approach to the representation of concurrent, distributed, asynchronous, parallel, deterministic and non deterministic systems is based on a mathematical concept called a Petri Net. This powerful formal specification tool is mainly used for dealing with performability issues in systems with concurrent processes with local behaviour. Petri nets are so used to describe processes as concurrent and interacting machines which engage in internal actions and communications with their environment or user.

The simplest kind of Petri Net is a bipartite digraph, i.e. a graph with two types of node and directed arcs which connect nodes of different types. The two types of node are called *place nodes*, represented as circles, and *transition nodes*, represented as boxes; places may hold

tokens, transition represent time events; hence this type of net is also known as a place-transition net. The arcs may be labelled with an integer weight, but if unlabelled are assumed to have a weight equal to 1. Places may be marked by an integer number of tokens.

The overall state of a system of n places is represented by a vector of size n consisting of the markings on each place. The arcs connected to a transition node define sets of input places and output places for that transition. In a simple Petri Net, a transition is enabled if all of its input places have a marking equal to or greater than the weight of the arc connecting that place to the transition. When a transition is enabled, it may be fired to remove a number of tokens from each input place equal to the weight of the connecting input arc, and create a number of new tokens at each output place equal to the weight of the connecting output arc.

Many extensions to the simple Petri Net model have been developed for various modelling and simulation purposes. These high-level nets include the following: *(i)* Hierarchical Petri Nets, which allow composition relations where a previously defined net is represented by a single place or transition in a new net; *(ii)* Hybrid Petri Nets, which incorporate places which may take continuous values instead of integer numbers of tokens; *(iii)* Timed Petri Nets, in which places and/or transitions may be assigned deterministic time delays; *(iv)* Stochastic Petri Nets, in which places and/or transitions may be assigned delays which are given by a probability distribution; *(v)* Coloured Petri Nets, which allow tokens to have internal structure and transitions to have more complex firing rules.

In Figure 3.3 continuous and discrete transitions of hybrid functional Petri net are represented. *(a)* An example of continuous transition. Four input arcs are attached to continuous transition T_C: two continuous input arcs from continuous places P_1 and P_4, and two test input arcs from continuous place P_2 and discrete place P_3. a_i is the weight of arc from place P_i for $i = [1, 2, 3, 4]$. Two continuous arcs are headed from the transition T_C to continuous places Q_1 and Q_2, respectively. Variables b_1 and b_2 are assigned to these arcs as weights. *(b)* An example of discrete transition. Four input arcs are attached to discrete transition T_D: two discrete input arcs from discrete place P_1 and continuous place P_3, and two test input arcs from discrete place

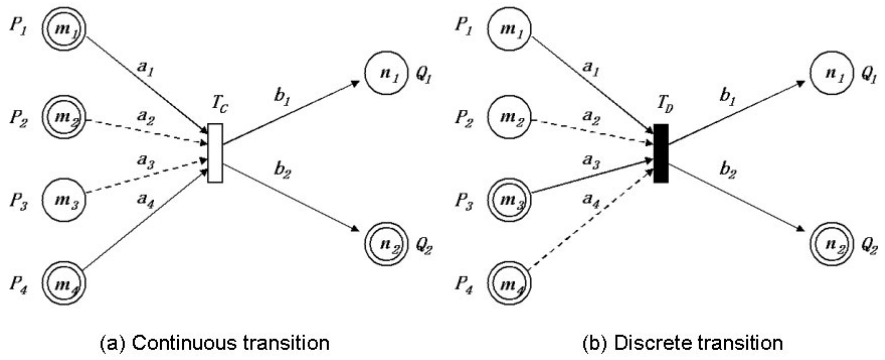


Figure 3.3: Petri Net

P_2 and continuous place P_4. a_i is the weight of arc from place P_i for $i = [1, 2, 3, 4]$. Two output arcs are headed from the transition T_D to discrete place Q_1 and continuous place Q_2. Variables b_1 and b_1 are assigned to these arcs as weights.

Petri Nets in Systems Biology

Petri nets have recently been proposed as a potential tool for modelling, composing and analysing biological systems. Petri Nets of various different kinds have been used in several studies in Systems Biology, both as structural network models for qualitative analyses and as quantitative models using high level nets.

There are two aspects of biological networks that can be represented and analysed using Petri net based models:

1. *Network structure*

   Network structural properties can be analysed using methods already devised for the structural analysis of Petri nets. Some progress has been made in this area already, for example the use of Place and Transition invariants, although there are many interesting net properties such as boundedness, liveness and reachability for which a biological context still remains to be investigated.

2. *Network behaviour*

   Individual cellular processes can be viewed as discrete event systems in which sets of single molecules act as species, products and mediators in these reactions and processes. Hence, the concentrations of these entities can ultimately be expressed as integer units. The behaviour of biological networks over time has been modelled with stochastic and hybrid Petri nets. When simulating a network over time the rates of the reaction must be captured. Currently these rates are obtained from a combination of experimental results, expert knowledge and experimentally manipulating the network representation.

An example has been shown in Figure 3.4: glycolysis pathway (GP) is a sequence of reactions that converts glucose into pyruvate with the concomitant production of a relatively small amount of ATP.

3.7.2 Π-calculus

The π-calculus is one of the process algebra, which was developed by Robin Milner as a formal language for concurrent computational processes. It provides a framework for the representation, simulation, analysis and verification of mobile communication systems. A typical system in the π-calculus consists of multiple concurrent processes. Pairs of processes interact with each other by sending and receiving messages in a synchronized way. This communication is done on complementary (input and output) channels. The content of messages can be also channel's name. As a result of such a communication event, the recipient process may now use the received channel for further communication. This feature, called *mobility*, allows the network "wiring" and structure to change with interaction. Mobile systems are made up of few operators to compose elementary actions (say α) over distributed channels (denoted hereafter by their names, given in lower-case letters).

These operators are: sequentialisation (α.P), parallel composition (P—Q), name declaration (ν x), and recursion ($recx.P$).

Note, that the calculus can be non-deterministic: stochastic *pi*-calculus. Thus, when several options are available, the actual interac-

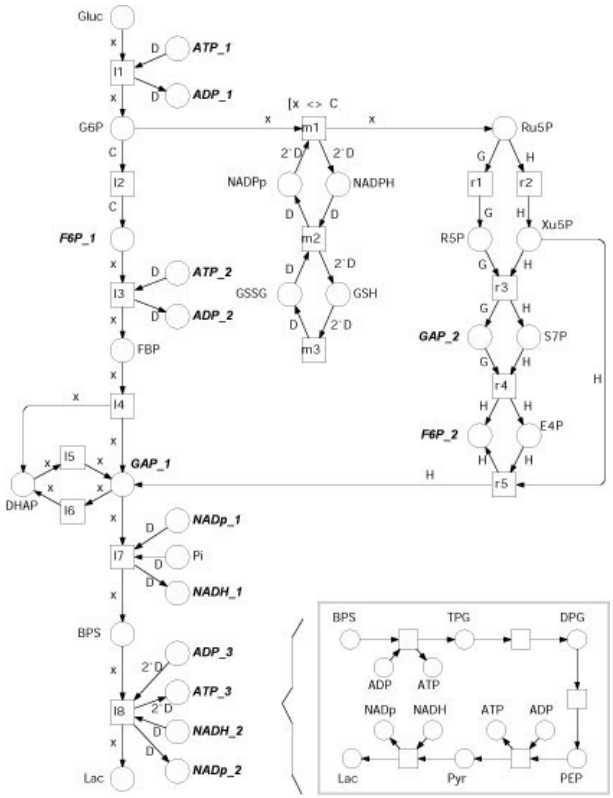


Figure 3.4: Petri Net

tion that occurs is chosen on a completely random basis.

Π-calculus in Systems Biology

The use of π-calculus as a means to analyse biological networks and to represent pathways is suggested in a lot of work developed in recent years.

The abstract characteristics of biological systems are the same as those of distributed and mobile systems. Many processes are active simultaneously over a set of physical resources for which they compete while cooperating to accomplish a common goal. Acquisition of a resource from a process or reception of a message upon which choices have to be taken can surely affect the future behaviour of the

whole system and even change the logical interconnection structure among processes. Trust barriers and administrative domains work as membranes that can be passed only by those processes that possess the right keys - hence the concept of localisation of processes is an important one.

Mimicking the description of mobile and distributed systems in a biological domain, it can be stated that processes are the biological components. Sharing of channels establishes the interconnection topology of the system and represents the interaction potentials of components together with their affinity. Scopes of channels or explicit binders represent the boundaries within which interactions through such channels may occur. Since channel names can be sent as data along channels, the interconnection topology varies dynamically, so modelling the impact of an interaction on the future behaviour of the whole system. The above interpretation immediately provides a dynamic description of the temporal as well as causal evolution of the system in hand: we only need to run the program.

The π-calculus is suitable for modelling various intracellular molecular systems, including transcriptional circuits, metabolic pathways, and signal transduction networks.

We suggest the reading of [40] to have an example on how the π-calculus can be used to model biochemical networks as mobile communication systems. In this paper authors treat molecules and their individual domains as computational processes, where their complementary structural and chemical determinants correspond to communication channels. Chemical interaction and subsequent modification coincide with communication and channel transmission.

Based on its formal semantics, the model is amenable to computer simulation, analysis and formal verification, using a combination of existing and self-developed tools. The compositional nature of the calculus allows incremental modelling of complex networks and alternation between different levels of complexity. This is instrumental for studying the modular design of biological systems.

An extension of π-Calculus in form of the stochastic π-Calculus, supports the definition of discrete event models and their execution by discrete event simulation. In the stochastic π-Calculus actions are assigned rates according to the rates of the corresponding biochemical re-

actions and a probabilistic distribution function is defined, driving the selection of the action to fire among all the ones enabled. Openly available simulation systems like BiosPI also push the application of the stochastic π-Calculus. Recent developments like BioAmbients which is based on the stochastic π-Calculus, allow the description of spatial cell compartments, and entities moving from one compartment to the next and thus increase the expressiveness of the language [39].

Chapter 4

Multi-agent Systems Approaches in Systems Biology

In this chapter we explain the basic concepts and principles concerning agents and multi-agent systems, considering — in particular — the Agents and Artifacts meta-model. Such concepts and principles provide a background to understand how a multi-agent system can be used in modelling and simulating biological systems.

4.1 Basic Concept of MASs

The area of Multi-agent systems (MASs) brings together and draws on results, concepts and ideas of many disciplines including artificial intelligence (AI), distributed artificial intelligence (DAI), Parallel & Distributed Systems P&D, Mobile Computing, Programming Languages and Paradigms (PL), Software Engineering (SE), Robotics.

Generally speaking, MAS is an effective paradigm for modelling, understanding, and engineering *complex systems*, providing a basic set of high level abstractions that make is possible to directly capture and represent the main aspects of such complex systems, such as interaction, multiplicity and decentralization of control, openness, dynamism to cite few.

A MAS can be characterized by three key abstractions in a MAS:

agents, *societies* and *the environment*. Agents are the basic active components of the systems, executing pro-actively and autonomously some kind of work. Agent societies are formed by set of agents that suitably interact and communicate so as to do some kind of collective work, which requires the contribution — including knowledge, skills, activities — of multiple agents. In order to do their work agents typically need to exploit and affect the environment where they are situated.

4.1.1 The notion of Agent

The first key concept of a MAS is that of "agent". To give an agent definition is not easy because, actually now, there is no accepted notion of the term "agent", and indeed there is a good deal of ongoing debate and controversy on this subject; we can find a lot of agent definitions, more or less convergent, coming from the different area cited above.

Essentially there is a general consensus and agreement that *autonomy* is central to the notion of agent. Part of the difficulty is that various attributes associated with agent are of different importance for different domains.

Nevertheless, some sort of definition is important. The most accepted and cited is that of [47] where: *an agent is a computer system that is situated in some environment, and that is capable of autonomous action in this environment in order to meet its design objectives.*

There are some agent properties to note, because they represent the fundamental keys for the proposal that we intend to develop in the next chapters:

1. agents are clearly identifiable entities with well defined boundaries;

2. agents are situated in a particular environment over which they have partial control and observability – they receive/sense inputs related to the state of the environment through some form of sensors and they act on the environment through effectors;

3. agents are autonomous, i.e. they have both control of their internal state and over their own behaviour;

Besides these properties which are central for this work, in particular for the MAS model that will be described in next chapters, many other features can be found in the models of agents found in the literature, which makes such a notion a powerful modelling and engineering abstraction, useful for several heterogeneous contexts.

Different kind of computational and programming models / architectures have been devised for defining agent structure and behaviour, depending on the specific application domains. Generalising the approaches, some points can be identified:

1. agents are designed to fulfil a specified role, as a set of tasks to achieve. As such they are entities with some objective;

2. agents are entities with inferential capabilities – it dynamically can compute new data representing a new solution to a given problem; new knowledge inferred from old data; new methods to solve a given problem; new laws describing a portion of the world; on these new data, new solution, new knowledge, new methods, new laws it can autonomously choose the following action to perform;

3. agents are capable of exhibiting flexible problem-solving behaviour in pursuit of their design objectives - being both *reactive* (sensing environment changes and behaving accordingly or able to respond in a timely fashion to changes that occur in the environment) and *proactive* (deliberating upon its own course of actions based on its mental representation of the world or able to opportunistically adopt goal and take the initiative).

As we said, various definitions and classifications besides this one have been given in literature for the agent concept, with different characterisations coming from different fields. From the different contexts, different acceptation of the agent abstraction have emerged, still sharing the basic issues of autonomy and situatedness. A synthesis is currently ongoing in the MAS community.

4.1.2 Agent Societies

When adopting an agent-oriented view, it soon becomes apparent that most application domain contexts require or involve multiple agents: to represent the decentralised nature of the problem, the multiple loci of control, the multiple perspectives or the competing interests. As obvious, MAS emphasise the multiplicity of the agents composing a system as a "society" of agents.

Each agent of this system, a *MAS agent*, is an autonomous entity pursuing its goal / task by interacting with other agents. In this view a MAS agent does not lives in isolation: it lives within an *agent society*. Its main features are: *(i)* autonomy / proactivity, *(ii)* interactivity / reactivity / situatedness....[SPIEGARE MEGLIO,MA FORSE NN QUI]

Interaction is therefore a main dimension of multi-agent systems. Agents need to interact either to fulfil their individual tasks or to take part to the collective processes which characterise the MAS as a whole. Interactions can vary from simple semantic interoperation through traditional client/server type of interactions, to rich social interactions (the ability to cooperate, coordinate, and negotiate in the same social context). More generally, we can say that different kind of interaction models can be adopted for enabling interaction and communication inside a MAS.

A first main distinction is between direct and mediated interaction. In the former case agents interact by directly exchanging informations; in the latter case, the media enabling agent interaction and communication, are explicitly modelled, and actually become fundamental in the engineering of the whole MAS. In this case the interaction and communication is not considered to take place directly between agents, but through these media as first class entities of the system. Such media could be the environment itself, or a different object.

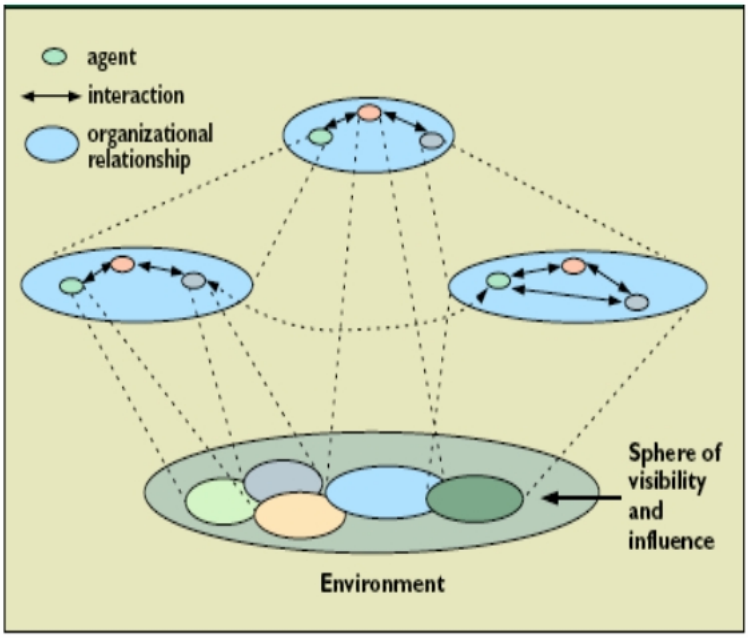


Figure 4.1: MAS

4.1.3 The notion of environment and the A&A meta-model

First of all, the notion of environment is what makes it possible to define agents' situated-ness, defining a notion of locality, a model for the stimuli that can be perceived by the agents and also a model for the actions available to agents to affect their environment. Accordingly, the environment plays a key role of *enabler* and *mediator* of agents' interaction: such a mediation role is crucial for realising forms of communication among agents than are complimentary to direct communication (which takes place directly between agents through some kind of agent communication language). By generalising this point, the A&A meta-model [41] a notion of "working environment" is introduced, modelling those parts of the environment which are explicitly conceived and designed to be fruitfully exploited by agents in their working activities.

The notion of artifact

Following A&A, such working environments can be generally modelled as a dynamic set of entities called *artifacts*, organised in *workspaces*. The notion of artifact is the core *abstraction* of A&A: it is meant to represent any entity belonging to the working environment—hence existing outside the agent mind—that is created, shared & used (and eventually disposed) by agents to carry on their activities, in particular social ones. So, an artifact (type) is typically meant to be explicitly designed by MAS engineers so as to encapsulate some kind of *function*, here synonym of "intended purpose".

Artifacts work then as the basic building blocks to compose complex working environments: MAS designers can define different types of artifacts, according to the need of the application at hand or to devise a library of reusable artifacts. Analogously to artifacts as studied in human science disciplines, two basic categories of artifacts could be identified: *resources* and *tools*. While resources are primary source and target result of the agent activities, tools are artifacts used as "instrument" to achieve some objective or execute some task.

The artifact abstraction leads to a notion of *use* that is the basic kind of relationship among agents and artifacts, besides creation and disposal. Accordingly, the notion of *usage interface* is defined, as the basic set of *operations* and *observable states* and *events* that an artifact expose so as to be usable by agents. Informally, we can think about an agent interacting with an artifact through its usage interface as follows: an agent executes actions that result in the triggering of some artifact operations, which then leads to the observation of events or the evolution of the artifact state. Such an abstraction strictly mimics the way in which humans use their artifacts: a simple example is the coffee machine, whose usage interface includes suitable controls—such as the buttons—and means to make (part of) the machine behaviour observable—such as displays—and to collect the results produced by the machine—such as the coffee can.

Artifacts then are meant to model those parts of the MAS which are not autonomous or proactive: they are meant to represent passive entities that are useful if and only if properly (created and) used by agents.

4.2 Abstract Architectures for MAS

4.2.1 Abstract architectures for Agents

We can easy formalize the abstract view of agents presented so far, that does not help us to construct them, since it gives us no clues about how to design the decision function *action*. For this reason, we will now begin to refine our abstract model of agents, by breaking it down into subsystem. As we refine our view of agents, we find ourselves making design choices that mostly relate to the subsystems that go to make up an agent – what data and control structure are present.

An *agent architecture* is essentially a map of the internals of an agent – its data structure, the operations that maybe performed on these data structures, and the control flow between these data structures. There are a huge number of different types of agent architecture, with very different views on data structures and algorithms. In the remainder of this thesis, however, we will survey some fairly high-level design decisions.

First, let us assume that the environment may be in any of a finite set E of discrete, instantaneous states $E = \{e_1, e_2...\}$. At any given instant, the environment is assumed to be in one of these states. The effectoric capability of an agent is assumed to be represented by a set $A = \{a_1, a_2...\}$ of *actions*, through which it transforms the state of the environment. Then abstractly, an agent can be viewed as a function

$$action : S^* \rightarrow A$$

which maps sequence of environment states to actions: an agent decides what action to perform on the basis of its history – its experiences to date. These experience are represented as a sequence of environmental states – those that the agent has so far encountered.

The non deterministic behaviour of an environment can be modelled as a function

$$env : S \times A \rightarrow \rho(S)$$

in which the environment starts in some state, and the agent begins by choosing an action to perform on that state. As a result of this action, the environment can respond with a number of possible states. However, of course, only one state will actually result – though the agent does not know in advantage which it will be.

We can represent the interaction of agents and environment as a *history*. A history h is a sequence:

$h : s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} s_2 \xrightarrow{a_2} s_3 \xrightarrow{a_3} ...$

which is thus a sequence of interleaved environment states and actions.

Purely reactive agent

Certain type of agents decides what to do without reference to their history. They base their decision making entirely on the present, with no reference at all to the past. We will call such agents *purely reactive*, since they simply respond directly to their environment. Formally the behaviour of a purely reactive agent can be represented by a function

$action : S \rightarrow A.$

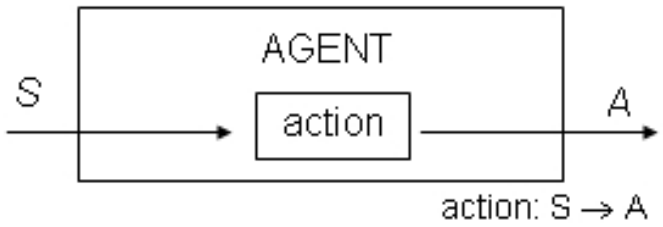


Figure 4.2: Purely reactive agent

Agent with perceptions

In this agent an agent's decision function is now separated into *perception* and *action* subsystems [Fig. 4.3]. The idea is that the function

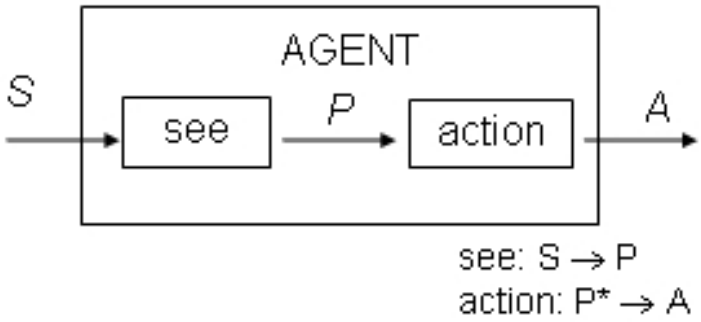


Figure 4.3: Agent with perceptions

see captures the agent's ability to observe its environment through sensors, whereas the *action* function represents the agent's decision making process. Let P be a (non-empty) set of percepts. Then *see* is a function

$$see : S \rightarrow P$$

which maps environment states to percepts, and *action* is now a function

$$action : P^* \rightarrow A$$

which maps sequences of percepts to actions.

State-based agent

We have so far been modelling an agent's decision function action as from sequences of environment states or percepts to actions. This allows us to represent agents whose decision making is influenced by environment history. We want now consider something more, i.e. agents

that *maintain state*.

These agents have some internal data structure, which is typically used to record information about the environment state and history. Let I be the set of all internal states of the agent. An agent's decision making process is the based, at least in part, on these information. The perception function *see* for a state-based agent is unchanged, mapping environment states to percepts as before:

$see : S \rightarrow P.$

The action-selection function *action* is now defined a mapping

$action : I \rightarrow A$

from internal state to action. An additional function *next* is introduced, which maps an internal state and percept to an internal state:

$next : I \times P \rightarrow I$

The behaviour of a state-based agent can be summarised as follows. The agent starts in some initial internal state i_0. It then observes its environment state s, and generates a percept *see(s)*. The internal state of the agent is then updated via the *next* function, becoming set to *next(i_0, see(s))*. The action selected by the agent is then *action(next(i_0, see(s)))*. This action is then performed, and the agent enter another cycle, perceiving the world via *see*, updating its state via *next*, and choosing an action to perform via *action*.

4.2.2 Abstract architectures for the environment

Following the A&A metamodel, an environment can be designed in terms of sets of artifacts, distributed and collected in workspaces. The abstract architecture of an artifact can be characterised here by:

- a usage interface, defining which operations can be invoked by agents to exploit artifacts functionalities. Such an interface includes also the set of observable events that are generated by the artifact, as a consequence of operation execution.

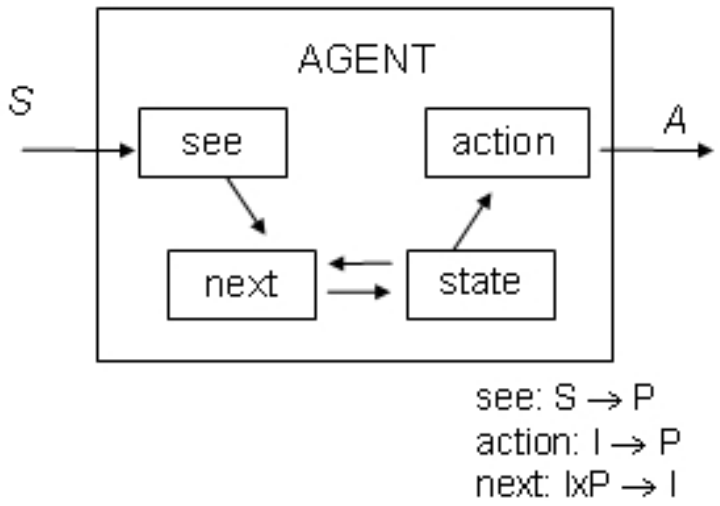


Figure 4.4: Agent with state

- artifact functions or operations, which defines the functionalities that an artifact encapsulates, which are exploited by the agents through the interface.

4.3 Modelling and Simulating Biological Systems as MASs

MASs have been widely recognised as a good approach for modelling and simulating complex systems. As such, they can be seen as a promising choice also for modelling and simulating biological complex systems, in particular those scenarios in which traditional approaches, such as the one most used based on Ordinary Differential Equations (ODE), fail.

MASs dimension of *interaction* is one of the main reasons that makes MASs a rather new modelling and simulating method available in Systems Biology.

As we know, in biological systems, interactions inside the system

and between the system and its environment plays a fundamental role and must be explicitly taken into account in order to understand the overall system behaviour, typically emergent, along with related effects concerning non-linearity, stochastic phenomena, feedbacks, and so on [32].

With MASs, biological systems are so modelled as a set of interacting autonomous components , i.e. a set of agents, and the chemical environment is modelled through an environment abstraction. With MASs, we have methods to: *(i)* model individual structures and behaviours of different entities of the biological system as different agents (*heterogeneity*); *(ii)* model local interactions between biological entities/agents (*locality*); *(iii)* model the environment structures and dynamics; *(iv)* define biological entities/agents-environment interactions.

It is so clear that multi-agent systems seems to promote a natural form of modelling, where active entities in the original systems are interpreted as actors in the model [25].

On the other hand, simulating a MAS model means execute it as a program and study its time evolution: *(i)* observing individual and environment evolution; *(ii)* observing global system properties as emergent properties from agent-environment and inter-agent local interaction; *(iii)* making in-silico experiments. From local interactions a multi-agent system arises a global coherent behaviours and emergent structures, with the same processes of self-organization that we can observe in real biological systems. The systemic, emergent properties that characterize a biological system are so reproducible and analysable in the virtual system.

4.3.1 A methodology

We propose now a possible methodology that can be followed in designing, developing and simulating a multi-agent model.

Step 1: Identify general objectives of the simulation study.

Step 2: Identify the scope of the model – what aspects should be contained, which not – and the assumptions and parameters - with justifications of their actual setting.

Step 3: Do a coarse plan of experiments – which are possible configuration tests; listing of simulation experiments.

Step 4: Do a coarse characterisation of the system (agents, interactions, environment):

1. Identify agents, observing and defining:

   – which are the actors involved in the process under study: active entities of the process are agents in the model;

   – which is their role;

   – which are the interactions between them;

   – which is the result of the interactions;

   – which new entity are produced by the composition of simpler entities.

2. Characterize the environment defining spatial aspects (build map) and its possible functions.

3. Define a coarse agent behaviour:

   – taking the agents perspective and formulate what you would do for performing your agent-goal;

   – information that the agent might receive, actions that it might perform - what effect should these actions have?

Step 5: Specification of Multi-Agent Model.
This step correspond to "classical" specification phases in Software Engineering. Its aim is making the coarse concept more concrete: detailed, almost executable model, at least guideline. We suggest in the follow which are the main components of this specification, to take into account:

1. definition of agent classes, their internal structure, i.e. definition of agent architecture. We can identify, from a modellers point of view, two main forms of models:

(a) *Behaviour-oriented* agent models:
The modeller describes agent behaviour, status and dynamics with some kind of formalisms such as rules, activity, graphs. Reactions to perceptions/status changes are defined by the modeller. Agent goal is implicit treated.

(b) Goal-oriented agent models:
The modeller identifies and associates goals of the agents, which possess some forms of cognitive capabilities, which make it possible to explicitly represent their goals, and which possess a set of actions that they may execute after selecting that goal. Reactions are not predefined, but goal dependent. Agent goals are explicit treated in the agent behaviour.

2. definition of the environmental elements and their relation to agents;

3. concrete definition of interactions and organisational structures;

4. definition of the computation procedures for performance values.

4.3.2 What can we Model with MASs

Based on the attempts to represent the same model in different frameworks, [25] identifies some basic characteristics in systems to be modelled that advise for the use of a multi-agent modelling and then simulation:

(a) When feedback loops in the agents behaviour are important, but the conditional behaviour they are based on is not purely locally determined. When the ability of the entity to decide is not only based on its local surroundings but also must relate to more or less global properties or values, then a multi-agent simulation is well suited.

(b) When the feedback loops are not fixed in the sense that the number of affected entities is not predefinable or the existence of the feedback loop is depending on additional factors, then the formulation of flexible agent behaviour has advantages to network structures with fixed

connections.

(c) When inhomogeneous space is relevant, then abstractions using stochastic terms for flight times, etc. might not always be sufficient to reproduce the effects of space onto different behavioural patterns, or at least make a valid representation of behaviour unnecessarily complex compared to an multi-agent model. The problem seems to be still more sophisticated, when the configuration of the relevant spatial patterns has to undergo dynamic changes.

(d) When flexible conditional or even adaptive individual behaviour has to be formulated, then it could be easier to concentrate on the behaviour of an agent than to describe a network that is passed through by a token, even if it carries an internal structure. Adaptivity of behaviour was not a central point in this study, but is also an important feature that is rather directly representable in an agent-based model, whereas it causes problems in other modelling frameworks.

(e) When interactions with flexible individual participants have to be represented, then it might be hard to formulate it using either uniform entities or predefined sequences of processes. When it is not irrelevant who the interaction partner of a particular agent is, or the agent may flexibly decide not to interact at all, then a focus on the agent, its behaviour and reasoning may be advisable. Another advantage of multi-agent simulation consists in its un-fixed interaction participants. That facilitate the modelling of variable agent numbers. When an agent has to be erased or a new agent enters the scenario, it may start interacting with the other agents without complex reconfigurations of the system. However, this advantage is relative to the application domain.

Developing, designing and finally implementing a multi-agent simulation is not trivial. The situation is even worse, as there exists neither an unified formal framework for multi-agent models nor a widely accepted methodology for developing multi-agent simulations. The modeller must carefully think about whether the instrument of multi-

agent simulation is necessary and explain why there is no "simpler" method for modelling and simulating the system. This even leads to the basic question: what the advantages of multi-agent simulation are?

In the following we want to tackle the question, comparing different standard methods with MASs method.

4.4 A Comparison with Standard Approaches

For what we have said until now some of the differences between MAS approach and traditional ones are clear. For argumenting why multi-agent simulation should be used, we have to know weak points of other paradigm and modelling and simulating approaches, and then we can talk about benefits and drawbacks of a MAS approach.

Weak point of mathematical approaches

Some of the weak points of standard approaches are yet underlined in the third chapter, where we made a review of the modelling techniques applied in Systems Biology. Now we recall and deepen some of these points.

The most commonly used models of biological systems are based on ODE and PDE. Differential equations are a really well understood and established framework: their advantages results from their relative simplicity and from their formal aspect. They promote a macro-model approach, since the complete system is tackled as one object whose state variables are updated during simulation. Modelling, simulating and observation happens on one level, the global level. The characteristics of a population are averaged together and the model attempts to simulate changes in the averaged characteristics of the whole population.

Focusing only on the population, we lose the representation of the individual and its locality, with the conditional and adaptive behaviour of each entity in its environment/locality, and we ignore the local processes performed by low-level components [42]. A particular entity or individual groups are no longer accessible [5].

Besides, in most biological systems the number of different types of entities which compound them has an important role and makes the classical approach through differential equations hard and often without any simple analytical solution.

To give an example, if traditional mathematical approaches have proved to be quite effective in modelling and simulating metabolic pathways, the same does not hold for signalling pathways. The mechanisms underlying these complex behaviours involve many interacting components and cannot be understood by experiments alone. No adequate mathematical models are known for analysing such pathways.

4.4.1 Advantages of Multi-agent systems simulation

MAS is a powerful paradigm. Generally speaking, they allow, of course, to stay quite close to biological reality: each real entity is a simulated agent, that means that it provides a really intuitive way of modelling, enabling more researcher to use simulation. In addition, we can design a lot of different entities/agents, encapsulating heterogeneous structures and behaviours, and a inhomogeneous environment, dealing with biological heterogeneity in space and population.

The virtual model so designed has amazing dynamic properties: each agent could be viewed, modified, removed from the model or added to the model very easily, introducing a peculiar property of the Multi-agent simulation.

Multi-level approach

Agent and MAS dimensions, as individual and social dimensions that coexist in the same model, suggest that a simulations based on the agent and multi-agent paradigm give, at the same time, at least two hierarchical level of description of biological system. They integrate aspects that can be found both in micro and macro techniques to simulation. On the one side, in the same way as micro techniques, agent-based approaches model specific behaviour of individual entities or components. On the other side, in the same way as in macro techniques, agent-based approaches promote the investigation of sys-

temic properties that cannot be understood at the individual component level, but require the introduction of new categories for their description. In other words, agent-based approaches make it possible to simulate and analyse emergent properties, which can be understood as properties of the ensemble of the components in the overall [6].

This is not all: what can be described at one level as an individual agent, at a more detailed level can be described as a society of agents (zooming in) and vice-versa (zooming out), according to hierarchy principle that characterize biological systems. The hierarchy principle suggests that agent-oriented processes and methods should support some forms of MAS layering, allowing modellers to design and develop MAS along different levels of abstractions - a number of independent, but strictly related, MAS layers.

The same biological system is so described at different levels of abstraction: MAS models are, undoubtedly, multi-level models.

Top-down & bottom-up approach

An agent-based or multi-agent simulation can follows both "top-down" and "bottom-up" approaches. The former approach starts with the phenomena (described on a global level) and systematically derive agents/behaviours, constructing an organisational structure. The latter approach starts with the observation of the real entity, from which it derives simulated agents, hoping that the global behaviour can be produced; it follows a try and error strategy and the parameter calibration is costly.

But with an agent formalism we can also combine these two approaches, using global knowledge for constraining/guiding bottom-up development,that means that the modeller has to start by determining the objective, then use objective to determine coarse dependences, use this knowledge for constraining the bottom-up approach.

Hybrid approach

An agent-based or multi-agent simulation can implement an hybrid method. We can in fact model different aspects or components of the system in different way, obtaining an hybrid model, discrete and continuous, deterministic and stochastic, qualitative and quantitative.

We can also integrate existing and new approaches to model individual and social behaviours, so that reusing existing algorithm. For instance for defining agent behaviour, in terms of reactive and proactive rules, we can adopt existing models, already implemented and validated, such as models of gene regulation networks which describe their dynamical evolution.

Modular Approach

A multi-agent approach is a modular approach because it allows consecutive specifications of the components of the system, of its parameters or variables; or it allows the addition of new elements or variables formerly not considered without change the existing MAS model, simply add them to the existing MAS model.

4.4.2 Disadvantages of Multi-agent systems simulation

As mentioned above multi-agent simulation currently has a lot of drawbacks. The huge parameter space that has to be searched for valid system behaviour causes an immense effort on justification, modelling and simulation. The lack of a formal framework makes the un-ambiguous presentation of a model a rather hard task. Thus there a some properties of the original system that on the other hand make the method of multi-agent simulation not advisable, although it seems to become a rather popular method.

(a) It is not clear, what parts of the system can be identified as agents, then multi-agent simulation is not apt. Components with simple non-autonomous behaviour or systems with fixed direct connections between components with well defined input-output behaviour can be tackled with better developed methods.

(b) If the considered space has a large extension or the agent numbers are huge, then an abstraction of homogeneous space and homogeneous societies may still be satisfying. A macro simulation approaches might to be sufficient. One has to regard that a simulation of millions of

agents takes very much time, especially compared to the computations necessary for simulating a set of differential equations.

*(c)*If a formal analysis of the model without simulating is necessary, e.g. for detecting deadlocks, etc., then a modelling method resulting in an exact and explicit model is necessary. Such a modelling method does not yet exist for multi-agent models. This restriction might change as there is a lot of ongoing work about formal specification in distributed artificial intelligence aiming at tools for software specification and verification.

Chapter 5

An Abstract Model of a Cellular Population based on MASs

In this chapter we propose a Multi-Agent based model of a *cellular population* embedded in its niche. Our abstract model is founded on A&A conceptual framework as specific reference for modelling and designing the simulation framework. Such a meta-model makes it possible to model a set of interacting cells in a niche as a multi-agent system, with societies of agents interacting in the same environment.

The abstract model proposed tries to be faithful and to reproduce the complex properties of biological systems, which are listed and explained in the first chapter. We have developed such model and subsequent simulation following the methodology proposed in the second chapter. At the beginning of each section we explicitly identify the step of the methodology in which we are. In this chapter we face the first points of the methodology. The last points are faced in the last seventh chapter. We intentionally do not consider some points of this methodology that we consider less important for our goal.

5.1 Model and Simulation Goal

> *STEP 1.* Formulate the goal of the simulation study or research question, i.e. fixing the question addressed by the

model.

The first goal of our model is to propose a concrete example of a multi-agent based model engineered with A&A conceptual framework. We model a cellular population. For this purpose, we experiment a bottom-up approach, in which the single entities of the system are modelled specifying their behaviour and how they interact with their environment. Following this approach, we aim at capturing the emergent properties which originate from the local interactions between components.

5.2 Model Assumptions

STEP 2. Identify, explicate and justify the assumptions under the model.

Here is a list of some strong assumptions on the background of the model:

- the first fundamental assumption of the approach is that every global properties and phenomenon emerges from the local interaction of the individual components; so the approach is heavily based on a *principle of locality*.
- we assume that molecules have not spatial extension, and that they have a brownian movement, as the result of the collisions. Cells instead are supposed to have a spatial extension.
- the membrane proteins are supposed to have a uniform distribution on the cell membrane, and to be continuously moving. So, if a signalling molecule is found in the perceptive region of the cell and there is a membrane protein with a free binding site complimentary to the molecule, the binding is established.

5.3 An Abstract Model

STEP 3. Design the model in two consecutive steps:

1. coarse level → model concept;
2. detailed, formal level → model specification;

We face in this chapter the point *(1)*, defining the concept of the abstract model. The model specification will be described in the seventh chapter.

5.3.1 The Cell System as A&A System

Modelling a biological complex system in terms of a MAS accounts for directly identify and represent the individual entities of the system as agents, the environment where such entities are situated and the basic type of interaction, both in terms of agent-environment interaction and inter-agent communication. The environment defines the topology of the system.

In this way, through a process of direct mapping we can easily obtain a model of the static structure of our cellular population which can be a multi-agent system, where each cell is modelled as an "agent" — with a heterogeneous type and behaviour—, and the chemical environment is modelled through an "environment" abstraction — an artifact or set of artifacts, according to the A&A metamodel described in previous chapter. With this computational metaphor we can also model the interaction between the cells as some kind of interaction between agents, which could be direct, modelling for example neural signal transmission, or indirect, mediated by environment, as in the case of the release and consumption of signalling molecules.

Going from agent-based systems to *multi-agent systems* account for considering the collective or *social* aspect of the system as a primary aspects, beyond the behaviour of the individual autonomous entities. The MAS dimension leads to focus then on the behaviour of the system as an ensemble of interacting components, in terms of both *emergent* global behaviours —including self-organising behaviours—, and *designed* collective behaviours, that can be fully characterized and explained (or *reduced*) in terms of the individual entities behaviour. Accordingly, we aim at capturing the global properties of cell systems by integrating the properties that emerge from the individual interaction of the (cell) agents with (possibly) the global laws and constrains enforced by the environment (artifacts).

5.3.2 Modelling a Cell as an Agent

The agent architecture that we have chosen for modelling a generic cell is the architecture of state-based agent. We have already defined it in the previous chapter.

We are now opportunely integrating and adapting this architecture for modelling a cell and its interaction with other cells and with the biochemical environment where it is suited.

Given such notions from molecular and cellular biology, we think that this is the best architecture to capture the cell functions and behaviour. Cells are, in fact, entities with a state, through which past events/history are memorized and a specific behaviour is identified. This behaviour consist of a sets of general cellular functions which can be modelled by a sets of reactive and proactive rules, that define agent's interactive behaviour with environment and inner agent's modifications. These rules are triggered on the basis of cell agent state and stimuli coming from the environment.

5.3.3 The Cell Agent State

Each gene of a cell contains the information for thousands of proteins: it is first transcribed into a specific mRNA molecule, which in turn guide the synthesis of protein molecules by the more complex machinery of translation. Proteins are the principal catalysts for almost all the chemical reaction in the cell, and have a host of other functions as well – maintaining structures, generating movements, sensing signals and so on. Proteins, above all, are the molecules that put the cell's genetic information into action.

We identify the state of the cells in their gene expression profiling, that measure the concentrations of different mRNA, i.e. in the transcription levels of each gene, which could be active (transcripted) or inactive (not transcripted). The gene expression level reflect in fact the type of proteins, and in which number, we find inside a cell in a specific time instant, and consequently most of the actions and functions that the cell has to and can perform instant after instant.

This is the reason way we choose to model the cell agent state through a set of state variables which represent the expression level of each gene.

Eg. `gene-x(0.0)`

Variables varies continuously in the interval [0, 1], where the minimum and maximum value mean that the gene is completely active or completely inactive. In the example above the gene is not transcripted.

The cell agent state at time t is so identified by the set of state variable's values. Depending on the state, the cell agent has particular functions and behaviours: at each state in fact correspond a particular set of proteins which are expressed.

The state modification is then done as a consequence of the stimuli coming from the environment, opportunely trasducted, or as an effect of particular internal modification. Both of these modifications are defined by the reactive and proactive rules characterizing agent's behaviour.

5.3.4 Interaction with the Environment: The Cell Membrane

Cells membrane are crucial to the life of the cell. The plasma membrane encloses the cell, defines its boundaries, and maintains the essential differences between the cytosol and the extracellular environment. In all cells, the plasma membrane also contains proteins that act as sensors of the external signals, allowing the cell to change its behaviour in response to environmental cues; these protein sensors, or receptors, transfer information – rather then ions ecc. These proteins in fact can bind in particular sites, specific molecules and consequently activate a signal-transduction pathway.

A *signal-transduction pathway* is a series of molecular changes that converts a signal on target cell's surface into a specific response inside the cell. The main elements of a signal-transduction pathway in which the target cell's response is the transcription (turning on) of a gene are the following:

1. The signal cell first secretes the signal molecules.

2. The secreted molecules binds to specific receptor protein embedded in the target cell's plasma membrane.

3. The binding activates the first in a series of relay proteins within the target cell. Each relay protein activates another.

4. The last relay molecule in the series activates a transcription factor.

5. The factor triggers transcription of a single gene.

6. Subsequently, translation produces the corresponding protein.

Therefore, as a consequence of this transduction, the cell changes its state and also its functions/activities, depending on the new proteins which are produced.

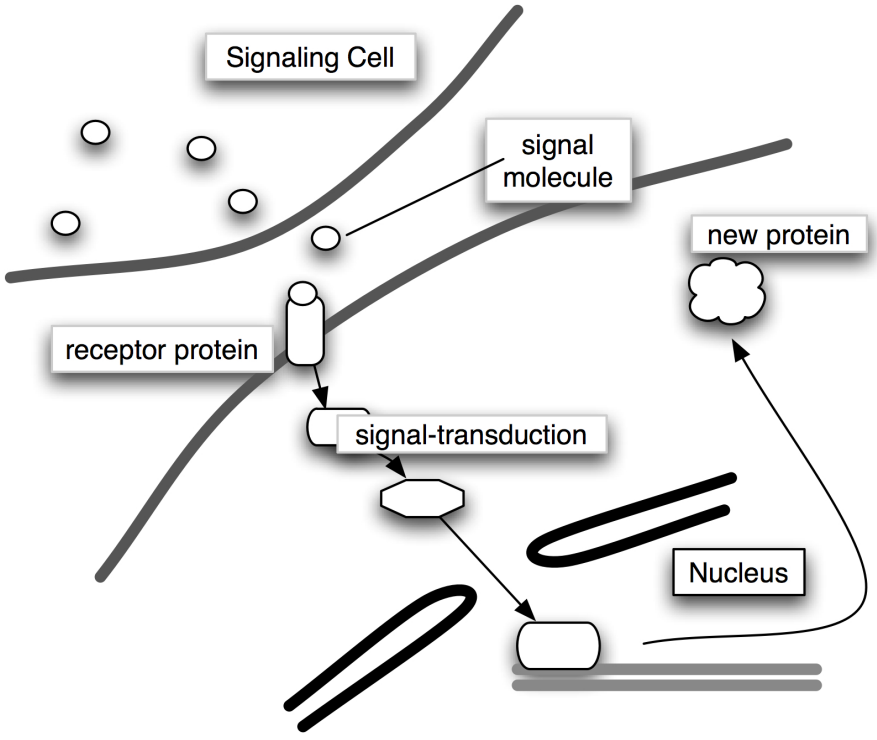


Figure 5.1: Signal-transduction pathway

see function: the source of interaction

The membrane proteins are modelled as the agent's sensor. Each agent has a particular set of protein sensors, which are defined by the specific binding sites that they expose. These binding sites can interact with specific binding sites of other proteins or with specific molecules of extracellular environment.

The state of a protein sensor is characterized by the set of free and/or bounded binding sites which it has in that moment. The state of a protein sensor represent the tridimensional structure of a protein identified in molecular biology. At each state is associated the eventual generation of an internal *perception*.

The sensors has then a set of transition rules for moving between different state, as a consequence of the creation of a new binding or of the liberation of a site.

The model is also provided of a mechanism for the bindings' break. At a biological level this break happens because *(i)* there are other molecules in the environment which casually collide with the bounded molecules or protein, *(ii)* the cell digest the bounded molecule at the aim of interrupting the signal-tranduction *(iii)* the bounded molecule is naturally reduced (it has a limited life-time).

For capturing all these aspects we have defined a probability distribution function, linearly time dependent. This function has the follow time evolution:

$$P(t) = \begin{cases} kt & \text{per } 0 \leq t \leq T \\ 1 & \text{per } t > T \end{cases} \tag{5.1}$$

where $k = 1/T$ and T is the maximum possible value for the binding's duration.

All the membrane proteins are encapsulated in the plasma membrane, whose function is collecting all the perception generated as a consequence of an external stimulus.

Finally, depending on the type of membrane proteins which are encapsulated in the plasma membrane, on the state of each of that, on the molecules' type or membrane proteins of other cells which are in the perception radius of the cell agent, the plasma membrane determines the satte transition of each protein and collects the consequent perceptions.

We have so modelled the perception function *see* for a state-based agent, mapping environment states to perceptions:

$see : Environment\ State \rightarrow Perception.$

5.3.5 The Cell Agent behaviour

Each agent of MAS encapsulates a dynamic state, a dynamic behaviour and the full control of such a behaviour, establishing time by time which kind of *actions* to take upon the environment—according to some kind of criterion, which can be defined in terms of agent *goal* or *task*—and how to *react* to stimuli *perceived* by the environment.

***next* Function: Agent Reactive Behaviour**

The reactive behaviour of a cell agent is defined by a set of *reactive rules*, which can be used to define the cell reaction to the perceptions generated by membrane proteins . Once triggered, the effect of a reactive rule is changing one or multiple state variables inside the agent, updating their value with a quantity that is specified and controlled by the rule. The reactive rules are meant to play the same role and function of the signalling-transduction pathway inside the cell, which are fundamental to transmit an external signal to the inner part of the cell and then to modify expression of one or multiple genes.

Then, a reaction rule maps an internal state and perception into a new internal state I':

$next : I \times Perception \rightarrow I'$ (REACTIVE RULE)

***action* Function: Agent Proactive Behaviour**

The pro-active behaviour of an agent is described by a set of rules — called proactive rules — which define the set of actions that the cell agent perform depending on its inner state. Such actions include also updates of the inner state itself. Each cell, as anticipated, has certain

particular functions on the basis of the number and type of proteins which compose it. These numbers are controlled inside the cell by the gene regulation networks. For this reason, pro-active rules are triggered depending on the dynamic value of the state variables, which represent the gene expression level. The actions and functions that a cell agent can do can be of different kinds, including also the update of the gene expression level (explicitly modelling the gene regulation networks). As a result, some basic cell processes can be started, such as cell division, differentiation, apoptosys and so on.

The action-selection function *action* can be defined by the following mapping:

$action : I' \rightarrow A \times I''$ (PROACTIVE RULE)

5.3.6 Modelling Cell Niches as Agent Environments

Finally, we explicitly model cell *niches* as MAS environments where cell agents are situated. We keep the term "niche" to identify such MAS environments. An individual niche encapsulates different kind of functions:

- it defines the topology of the space where cells and molecules are immersed;
- it functions as a container and controller for the overall set of molecules, including those playing the role of cues such as the cytochines, and proteins.
- it functions as an artifact enabling the interaction between the cell agents and the molecules, or between the cell agents.

It's worth noting that a niche controls the movement of molecules, but not the movement of the cell agents, which are supposed to autonomous.

A complete biosystems can be modelled then by a network of niches, linked together so as to make it possible the migration of cell

agents and molecules from one niche to another, according to the laws specified to rule such migrations. A cell agent then can be situated in a niche and then migrates to another niche; the reactive and proactive behaviour of the cell agent can change then according to the niche where the cell agent is currently situated, and the different kind of signalling molecules that are part of the niche.

5.4 The Simulator

STEP 6. Implement the model and a simulation software.

The model is implemented and executed on top of a simulator called CPSIM, developed to function as a *virtual laboratory* to execute *virtual experiments*.

The simulator can be conceived as an abstract (virtual) machine, and an experiment as a kind of a program on top of such a machine, defining the multi-agent system representing the specific cell population. The execution of the multi-agent system program corresponds to the execution of the virtual experiments, with the possibility for the users to observe and control the experiments by means of a suitable user interface.

The simulator has been implemented in Java, an object-oriented language with a runtime architecture — based on the Java Virtual Machine — that makes it possible to execute the application on almost all the main operating systems and machines (on every system where the Java platform is available, actually). Besides the portability, the choice of such an Object-Oriented programming language with a runtime architecture — despite of the performance penalty with respect to other languages without a runtime architecture, such as the C or C++ languages — has been effective for rapid prototyping such a quite complex application, based on a multi-agent system architecture.

The main parts of such an architecture are the following:

- `BioSystem` class, which represents the overall multi-agent system modelling the cell population. It keeps track of the overall cell agents and niches that are part of the system. Mainly, it provides a services to initialize and then step on the state of the multi-agent system by a certain delta time (a cycle), step by step,

computing the sequence of states that defines MAS execution. An updating cycle of the overall MAS accounts for the following main steps:

1. update niches states
2. let cell agents step on, providing them the information about their current locality, that is the set of the molecules and cells that are currently part of their neighbourhood
3. collect and execute cell agent actions, which possibly change their local environment

- `Cell`, which represents a cell inside the system, as an agent of the MAS. A cell agent is composed by three main parts: the membrane represented by the class `CellMembrane` — containing the set of reactive rules that define agent reactive behaviour, represented by the class `ReactiveRule` —, the inner state represented by the class `CellState` — and the set of proactive rules, represented by the class `ProactiveRule`. The `CellMembrane` is then composed by a set of specific proteins represented by classes extending the base class `MembraneProtein`. A membrain protein then is described in terms of a set of binding sites, represented by the `BindingSite` class, and its behaviour is defined in terms of a set of states, represented by the class `ProteinState`, along with state transitions, represented by the `ProtStateTransition` class.

  Besides these components, the class `CellPhase` is used to represent a cell phase of a cell: as defined in the abstract model, a cell can have one or more phases, which define different configurations of cell structure and behaviour. A single `CellPhase` groups a specific configuration for the membrane, the state, the set of the proactive rules and also a specific cell shape, represented by the class `CellShape`.

  Then, a specific cell agent can be defined by extending the class `Cell` — which functions as a base class for defining new cell types —, and initializing its structure so as to contains a specific set of cell phases.

  An individual simulation step for a `Cell` agent accounts for:

1. detecting the structural changes occurred in the membrane, given the current neighbourhood of the cell
2. updating accordingly the part of the inner state which is affected by the execution of the triggered reactive rules
3. collecting the perceptions generated by such updates
4. checking all the proactive rules and executing the ones that are triggered, according the current set of perceptions and the inner state of the cell

- `Niche` class, which represents a niche inside the system, that is an environment part (artifact) inside the MAS. A `Niche` is responsible to manage the set of molecules, represented by the class `Molecule`, which enable the interactions between the Cell agents and the niche environment. Also, the `Niche` defines and control the spatial property and topology of the space of the environment where the cells (and the molecules) are immersed, enabling and constraining cell (agents) and molecules movements.

Besides these classes which represent the most part of the *model* of the simulator, the controller part is represented by the `SimControllerAgent`, which is responsible to initialize and control the simulation execution, moving on the simulation step by step and reacting to user commands, triggered through the *view* part, that is the graphical user interface (GUI).

The view part is composed by three main panels: the *view panel* — implemented by the `ViewPanel` class — is used to provide a graphical representation of the simulation state and evolution; the *control panel* — implemented by the `ControlPanel` class — provides the GUI controls that the user can trigger to interact and control the simulation; and finally the *inspection panel* — implemented by the `InspectionPanel` class — makes it possible to inspect the value of parameters concerning the structure and dynamics of the simulation, such as the current number of a certain type of a cell or of a molecule.

Chapter 6

Modelling and Simulating Hematopoietic Stem Cells with Multi-agent Systems

Cells evolved originally as free living individuals, but the cells that matter most to us, as human beings, are specialized members of a multicellular community. They have lost features needed for independent survival and acquired peculiarities that serve the needs of the body as a whole. Although they share the same genome, they are spectacularly diverse: more than 200 different types are traditionally recognized in the human body. These collaborate with one others to form a multitude of different tissues, arranged into organs performing widely varied functions [3].

In this chapter we first describe the functions and lifestyle of a specialized type of cells in the adult body of a vertebrate, blood cells. In particular we are going to describe how new specialized cells born, how they live and die.

Then we try to explain why a model of stem cell is useful for understand their behaviour.

6.1 Blood cell

Blood contains many types of cells with different functions , ranging from the transport of oxygen to the production of antibodies. Some

TYPE OF CELL	MAIN FUNCTIONS
Red blood cells (erythrocytes)	Transport O_2 and CO_2
White blood cells (leucocytes)	
Granulocytes	
Neutrophils	Phagocytose and destroy invading bacteria
Eosinophils	Destroy larger parasites and modulate allergic inflammatory responses
Basophils	Release istamine (and in some species serotonin) in certain immune reactions
Monocytes	Become tissue macrophages, which phagocytose and digest invading microorganisms and foreign bodies as well as damaged senescent cells
Linfocytes	
B cells	Make antibodies
T cells	Kill virus-infected cells and regulate activities of other leucocytes
Natural killer (NK cells)	Kill virus-infected cells and some tumor cells
Platelets	Initiate blood clotting

Table 6.1: The various type of blood cells and their functions, as reported in [3]

of these cells function entirely with vascular system, while others use the vascular system as a means of transport and perform their function elsewhere. All blood cells, however, have certain similarities in

their life history. They all have limited life-span and are produced throughout the life of the animal. Most remarkably, they are all generated ultimately from a common stem cell in the bone marrow. This *hempoietic* (blood forming) *stem cell* is thus multipotent, giving rise to all the types of terminally differentiated blood cells as well as some other types of cells, such as osteoclasts in bone.

The classification of blood cells is shown in Table 5.1, with the main functions of each of them.

6.2 Hematopoietic Stem Cell (HSC)

Before going in the details of hematopoietic stem cells, we are now doing a brief introduction at the characteristics of a general stem cell.

6.2.1 Stem Cell Definition

A cellular population has to be self-renewing. It must therefore contain some cells that generate a mixture of progeny, including daughters that remain undifferentiate like their parent, as well as daughters that differentiate. Cells with this property are called *stem cells*. They have so a important role in such a variety of tissues that it is useful to have a formal definition.

The defining properties of a stem cell are as follows:

1. it is not itself terminally differentiated (that is, it is not at the end of a pathway of differentiation);

2. it can divide without limit (or at least for the lifetime of the animal);

3. when it divides, each daughter has a choice: it can either remain a stem cell, or it can embark on a course that commits it to terminal differentiation.

Although it is a part of the definition of a stem cell that it should be able to divide, it is no part of the definition that it should divide rapidly; in fact, stem cells usually divide at a relative law rate. They are required wherever there is a recurring need to replace differentiated

cells that cannot themselves divide, and this includes a great variety of tissues. Thus stem cells are of many types, specialized for the genesis of different classes of terminally differentiated cells –epidermal stem cells for epidermis, intestinal stem cell for intestinal epithelium, hemopoietic stem cell for blood, and so on.

Each stem cell system nevertheless raises similar fundamental questions. What factors determine whether the stem cell divides or stays quiescent? What decides whether a given daughter cell differentiates or remains a stem cell? And where the stem cell can give rise to more then one kind of differentiated cell –as in very often case– what determines which differentiation pathway is followed?

Trying to respond, in part, at these questions, we are going to explore in the next section the hematopoietic stem cell (HSC), which is the best-characterized adult stem cell is.

6.2.2 Role of Hematopoietic Stem Cell

Each day the human body produces billions of new white blood cells, red blood cells, and plates to replace blood cells lost to normal turnover precesses as well as to illness or trauma.

All of the mature blood cells in the body are generated from a relatively small number of hematopoietic stem cells and progenitors. HSCs are able to generate every lineage found in the hematopoietic system including red blood cells, platelets and a variety of lymphoid and myeloid cells.

HSCs generate the multiple hematopoietic lineages through a successive series of intermediate progenitors. These include common lymphoid progenitors (CLPs), which can generate only B, T, NK cells, and common myeloid progenitors (CMPs) which can generate only red cells, platelets, granulocytes, and monocytes. Downstream of the CLPs and CMPs are more mature progenitors that are further restricted in the number and type of lineages that they can generate. Ultimately, terminally differentiated cells are produced that cannot divide and undergo apoptosis after a period of time ranging from hours (for neutrophils) to decades (for some lymphocytes). A summary of the process of blood development is presented in Fig 5.1.

In this way, during homoeostasis, a proportion of stem cells are

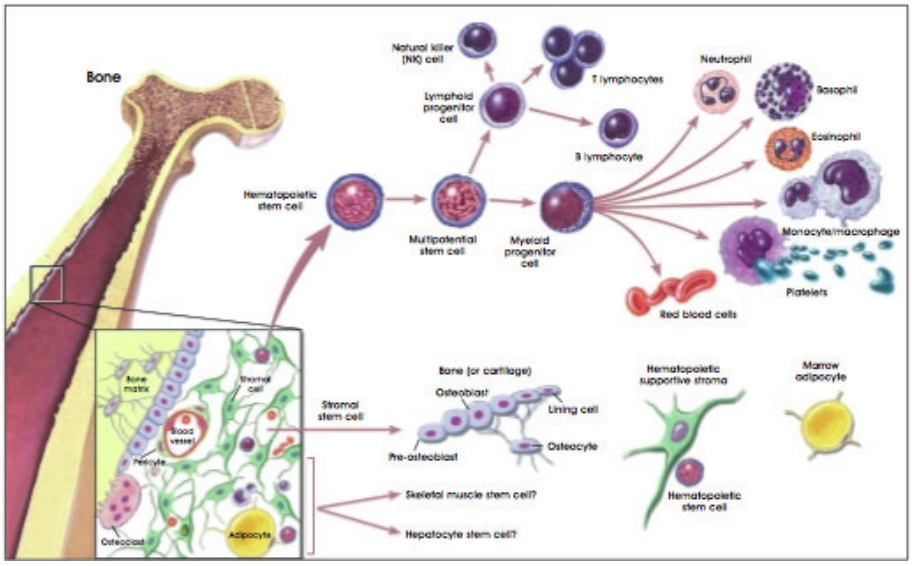


Figure 6.1: A summary of the process of blood development

expected to divide at least occasionally, to maintain a constant flow of short-lived progenitors that can generate enough cells to replace those that are constantly lost during normal turnover. This result is obtain balancing asymmetric division results in sufficient self-renewal – primitive cells– to sustain hematopoiesis throughout life and sufficient differentiation to produce mature cells with specialized capabilities necessary for blood. In this way, under homoeostatic conditions in the adult, the number of tissue stem cells remains relatively constant, despite the fact that they proliferate, because they not only self-renew but also produce differentiated progeny.

During times of physiologic stress such as haemorrhage, it is hypothesized that division may shift to more symmetric division favouring differentiation to replenish necessary mature cell pools [4].

A variety of homoeostatic mechanisms allow blood cell production to respond quickly to stress such as bleeding or infection and then return to normal levels when the stress is resolved, as we are going to explain in the next paragraph.

6.2.3 Hematopoiesis

The process of hematopoiesis involves a complex interplay between the intrinsic genetic processes of blood cells and their microenvironment. This interplay determines whether HSCs, progenitors, and mature blood cells remain quiescent, proliferate, differentiate, self-renew, or undergo apoptosi. Intrinsic genetic processes are cell autonomous, mechanisms possibly determined by developmental state. The specific microenvironment of stem cells has been historically called the haematopoietic-inductive microenvironment or "stem-cell niche" [46].

All of the genetic and environmental mechanisms that govern blood production operate by affecting the relative balance of these fundamental cellular processes. Under normal conditions, the majority of HSCs and many progenitors are quiescent in the G0 phase of the cell cycle; however, many of the more mature progenitors are proliferating and producing mature offspring.

In the absence of any stresses, this is balanced by the rate of apoptosis in progenitors and mature cells. In the event of a stress such as bleeding or infection, several processes occur. Stored pools of cells in the marrow or adherent to the endothelium are quickly released into the circulation in order to localize to the site of injury. Fewer progenitors and mature cells undergo apoptosis. In addition, quiescent progenitors and HSCs are stimulated by a variety of *growth factors* to proliferate and differentiate into mature white cells, red blood cells, and platelets. When the bleeding, infection, or other underlying stress ceases and the demand for blood cells returns to normal, the anti-apoptotic and proliferative processes wind down, blood cells are redistributed back to their storage sites, and the kinetics of hematopoiesis return to baseline levels [43].

6.2.4 Regulation of hematopoietic stem cell growth

Many of the different types of signals that are exchanged between stem cells and niche cells, as well as some of the signalling pathways that control stem cells maintenance, self-renewal and differentiation, have recently been identified. Between them:

1. Colony Stimulating Factor 1 (Macrophage)

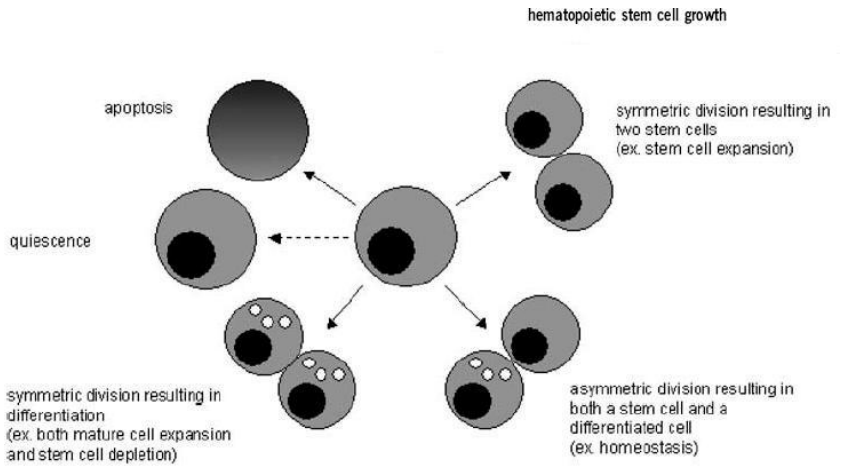


Figure 6.2: Possible HSC fates

2. Colony Stimulating Factor 2 (Granulocyte-Macrophage)
3. Colony Stimulating Factor 3 (Granulocyte)
4. Erythropoietin
5. Interleukin 1
6. Interleukin 3
7. Interleukin 4
8. Interleukin 5
9. Interleukin 6
10. Interleukin 7
11. Interleukin 11
12. OPN
13. Transforming Growth Factor-beta

6.2.5 Symmetric and asymmetric division in HSCs

We talk about both symmetric and asymmetric division. Let's now explain better their meaning.

Symmetric division

The vast majority of cell divisions are symmetrical, producing identical daughter cells and leading (in the absence of apoptosis) to increased numbers of cells, both generating two identical daughter cells with stem-cell function or two differentiated daughter cells.

Asymmetric division

A single division of a HSC can result in the formation of both an identical stem cell and a more highly mature cell. An individual stem cell can give rise to two non-identical daughter cells, one maintaining stem-cell identity and the other becoming a differentiated cell. There are two mechanisms by which this asymmetry can be achieved, depending on whether it occurs pre- (divisional asymmetry), or post- (environmental asymmetry) cell division.

- *Divisional asymmetry.*

  In divisional asymmetry, specific cell-fate determinants in the cytoplasm (mRNA and/or proteins) redistribute unequally before the onset of cell division. During mitosis, the cleavage plane is oriented such that only one daughter cell receives the determinants. Therefore, two non-identical daughter cells are produced, one retaining the stem-cell fate while the other initiates differentiation.

- *Environmental asymmetry.*

  An alternative way to achieve asymmetry is by exposure of the two daughter stem cells to different extrinsic signals provided by distinct local micro environments. Therefore, a stem cell would first undergo a symmetric self-renewing division, producing two identical daughter cells. While one daughter cell would remain in the niche microenvironment, conserving its stem-cell fate, the

other would contact (passively or actively) a different microenvironment that would no longer preserve its stem-cell phenotype but would instead produce signals initiating differentiation.

6.3 Need of a model of stem cell

Stem cell research has been arguably the biggest growth area of medical science in recent years. However, experiments are very limited because it is not possible to track stem cells in the adult human body. Even if you remove cells from the human body and look at them in the laboratory you can only tell if you had some stem cells in your original sample some weeks later when you see what the cells have actually done. However, we do know what stem cells do at the system level. They maintain the population of the various functional cells in our bodies, they can maintain there own number, and can also recover populations of stem and functional cells after disease, injury or radiation theory [13].

So we know what the system of stem cells does functionally but we do not have much idea of what happens at the level of individual cells and this gives us a clear reason as to why we might want to model and simulate systems of stem cells. Specifically we are interested in relating the behaviour of the system (macro) to what happens at the local cell (micro).

We don't know whether the fate of stem cells is pre-determined or stochastic, and whether the fate cells relies on their internal state, or on extra-cellular micro-environmental factors.

Until quite recently it seems that many researchers that stem cell fate was essentially pre-determined and that it was simply a matter of time before a stem cell did what it was pre-programmed to do. However, it is now reasonably clear that stem cell fate is a function of the local interaction with its environment.

We summarize what we see are the key reasons for the systematic development of formal models and simulations.

1. It is not possible to investigate how stem cells react by looking at

dead tissue, and much stem cell research is based on observation of dead, 2-D slides. Building simulations allows researchers to test possible cell behaviours that can then be related back to observable laboratory results.

2. In the adult stem cells cannot be distinguished morphologically from other primitive non-differentiated cell types. It is therefore hard, if not impossible, to observe their behaviour in the dynamic system of which they are a part.

3. The size and complexity of stem cell systems mean that without simulation, it is not possible to consider the whole system. Simulations provide an important tool for understanding the global behaviour of complex systems reacting agents.

4. There is no way to determine whether any individual isolated cell is a stem cell, or, to be able to model what its potential behaviour might be. It is not possible to make any definite statements about this cell. At best it can be tracked and its behaviour observed though clearly any particular behaviour is simply one of many possibly paths. The notion of a stem cell refers to the wide-ranging set of potential behaviours that it might have, and these are influenced by internal, environmental, and stochastic processes. Simulations provide a way of determining which behaviours are essential to stem cells and which are incidental in systems that have been studied in the laboratory.

5. the number of possible interactions and behaviours of a large number of stem cells makes the system an extremely complex (in all the senses described above) one. Theoretical simplifications are therefore key to understanding fundamental properties.

6. When you consider experimental evidence you have seen only one behaviour. This behaviour may have been one of many, and it is the potential for cells to behave in certain ways that might be key to defining them. Modelling and simulation is a much more effective device for understanding "behavioural potential" than looking at completed chains of events in the lab.

7. Though our work has been explicitly concerned with modelling the adult human body, it is clear that simulation does not involve any ethical difficulties such as extracting stem cells from an embryo in such a way that it is sacrificed.

8. And of course, simulation is cheap [14].

There have been several attempts to build formal models of these theories, so that predictions can be made about how and why stem cells behave either individually or collectively.

Chapter 7

A MAS-based Model for A Hematopoietic Stem Cells System

Following the abstract model proposed in the fifth chapter, in this chapter a first model of a multiagent system representing a population of hematopoietic stem cells is described and first results obtained by running experiments on top of the simulator described in previous chapter briefly discussed.

7.1 A MAS-based Model for Hematopoietic Stem Cells

Recalling the methodology described in chapter two:

> *STEP 3.* Design the model in two consecutive steps:
>
> 1. coarse level → model concept;
> 2. detailed, formal level → model specification;

Based on the abstract model described in the fifth chapter, in this section we apply the model by defining a concrete model the a hematopoietic stem cells system. The cellular population is composed by cells

which are specified with a set of different phases that corresponds to the differentiation steps of the stem cells.

The description of the hematopoietic stem cell — in the following — is given in a bottom-up style: first we describe some individual parts that are partially shared by cell phases, i.e. membrane proteins, state variables, reactive rules, proactive rules; then, we describe the the set of cell phase that characterise an homopoietic cell. Finally, we specify how the hematopoietic cell is composed, by referring all the parts.

7.1.1 Cell plasma membrane and membrane protein

The cell plasma membrane is the model collector of all the membrane proteins which are responsible of the interaction with the cell environment (as everything out to the cell). The cell membrane is also the collector of the stimuli which arise from the structural change of the membrane proteins.

Identified the molecules or proteins which are in the perception region of each cell, each protein of plasma membrane, depending on its own state and on the free binding sites associated to this state, can bind one ore more molecules and consequently make a transition in an other state. At each state of the protein is associated the eventual generation of one stimulus.

The following is the brief description of the membrane proteins considered in our case study. For each protein, the name and some basic information are reported: the number of binding sites, the type of molecule or protein that can be bound by the binding sites (ligand), the possible states of the protein, and the perceptions generated (P.g.) by the protein when being in the specified states.

1. *CD44*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
1 BS	OPN	s_0 = OPN bound → stimulus fired (P.g.)
		s_1 = OPN not bound

2. *Colony Stimulating Factor 1 receptor (Macrophage)*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
CSF1 BS	M-CSF	s_0 = M-CSF bound → stimulus fired (P.g.) s_1 = M-CSF not bound

3. *Colony Stimulating Factor 2 receptor(Granulocyte-Macrophage)*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
CSF2 BS	GM-CSF	s_0 = GM-CSF bound → stimulus fired (P.g.) s_1 = GM-CSF not bound

4. *Colony Stimulating Factor 3 receptor (Granulocyte)*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
CSF3 BS	G-CSF	s_0 = G-CSF bound → stimulus fired (P.g.) s_1 = G-CSF not bound

5. *Erythropoietin receptor*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
EPO Rec. BS	Erythropoietin	s_0 = EPO bound → stimulus fired (P.g.) s_1 = EPO not bound

6. *Interleukin 1 receptor*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
IL1 Rec. BS	Interleukin 1	s_0 = IL1 bound → stimulus fired (P.g.) s_1 = IL1 not bound

7. *Interleukin 3 receptor*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
IL3 Rec. BS	Interleukin 3	s_0 = IL3 bound → stimulus fired (P.g.) s_1 = Il3 not bound

8. *Interleukin 4 receptor*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
IL4 Rec. BS	Interleukin 4	s_0 = IL4 bound → stimulus fired (P.g.) s_1 = IL4 not bound

9. *Interleukin 5 receptor*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
IL5 Rec. BS	Interleukin 5	s_0 = IL5 bound → stimulus fired (P.g.) s_1 = IL5 not bound

10. *Interleukin 6 receptor*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
IL6 Rec. BS	Interleukin 6	s_0 = IL6 bound → stimulus fired (P.g.) s_1 = IL6 not bound

11. *Interleukin 7 receptor*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
IL7 Rec. BS	Interleukin 7	s_0 = IL7 bound → stimulus fired (P.g.) s_1 = IL7 not bound

12. *Interleukin 11 receptor*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
IL11 Rec. BS	Interleukin 11	s_0 = IL11 bound → stimulus fired (P.g.) s_1 = IL11 not bound

13. *Jagged*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
Jagged BS	Notch BS	s_0 = Notch bound → stimulus fired (P.g.) s_1 = Notch not bound

14. *Kit*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
Kit BS	Kit Ligand	s_0 = SCF bound → stimulus fired (P.g.) s_1 = SCF not bound

15. *N-cadherin*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
N-cadherin BS	N-cadherin BS	s_0 = N-cadherin bound → adhesion s_1 = N-cadherin not bound

16. *Notch*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
Notch BS	Jagged BS	s_0 = Jagged bound → stimulus fired (P.g.) s_1 = Jagged not bound

17. *Kit Ligand*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
Kit Ligand BS	Kit BS	s_0 = Kit bound → stimulus fired (P.g.) s_1 = Kit not bound

18. *TGFβ Receptor*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
TGF Rec BS	TGFβ	s_0 = TGFβ bound → stimulus fired (P.g.) s_1 = TGFβ not bound

19. *VLA4*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
VLA4 BS	VCAM1 BS	s_0 = VCAM1 bound → stimulus fired (P.g.) s_1 = VCAM1 not bound

20. *VCAM1*

Binding Sites (BS)	Ligand	State & Stimuli (Perceptions)
1 BS	TGFβ	s_0 = VLA4 bound → stimulus fired (P.g.) s_1 = VLA4 not bound

7.1.2 Cell State Variables

Following the abstract model, we define a set of state variables which represent the expression level of different genes. Because the number of gene in a human cell is about of 30.000 genes, and because we don't have the complete knowledge of which of them are involved in the processes we are interested in, we abstract by the specific gene involved introducing a set of "virtual genes". Each one of these virtual genes represents the set of genes (or, analogously the translated protein) which are responsible of a specific process. The value of a state variable ranges between 0 and 1. When the value reaches the maximum value, the corresponding process is triggered and can start. Currently, each process is realised as a set of proactive rules.

- APOPTOSIS State Variable:

  This variable represents all the genes involved in the apoptosis process, that is a process of deliberate life relinquishment by a cell in a multicellular organism.

- DIFFERENTIATION State Variables:

  The great differences among cells in an organism result from the selective expression of genes. This is the reason why, during the differentiation processes, different genes must be activated in different cells at different times. Each cell type has a different pattern of turned genes: particular cells express particular genes for specialized proteins. This suggests that in order for cells to differentiate – to become different from one another – certain genes must somehow be activated, while others remain inactive.

  With this variable we identify all the genes which are specific for the development of a cell in a specific pattern. We have in this way a differentiation state variable for each of the different phases of the cell (e.g. DIFFe for Erythrocyte, or DIFFgemm fro CFU-GEMM phase ans so on).

- QUIESCENCE State Variable: This variable is representative for those genes which leavethe cell in the G_0 phase of the cell cycle; quiescence is the state of a cell when it is not dividing.

- SENESCENCE State Variable: This variable is representative for that genes which are responsible for the cellular senescence. Cellular senescence is the phenomenon where cells lose the ability to divide. Cellular senescence is a state that occurs in response to DNA damage or degradation that would make a cell's progeny nonviable; it is often a biochemical alternative to the self-destruction of such a damaged cell by apoptosis.

- Symmetric DIVISION State Variable:

  This variable is representative for that genes which are responsible for the beginning of the cell cycle whose last small segment causes the cell division. Cell division is the process by which a cell, called the parent cell, divides into two cells, called daughter cells.

- Asymmetric DIVISION State Variables:

  Some kinds of stem cells are thought to undertake asymmetric cell division, generating one daughter cell that remains a stem

cell and one daughter cell that differentiates. For Hematopoietic Stem Cells, however, whether asymmetric cell division occurs during self-renewal is not known with certainty. It is instead possible that hematopoiesis occurs via symmetrical divisions, that sometimes give rise to two daughter HSC, and that at other times give rise to progeny that are committed to differentiate. The balance between self-renewal versus differentiation would therefore be regulated by the control of these two kinds of symmetrical cell division.

7.1.3 Reactive Rules

A reactive rule is easily defined specifying: *(i)* one or more stimuli (perceptions) that trigger it, *(ii)*the target state variable (SV) whose value has to be updated (representing the gene expression level, which is changed because of the signal-transduction pathway), *(iii)* the "updating force", that is amount of the update. Here it follows a set of reactive rules. Some rules are omitted because off their strong analogy and similarity with other cited. Some other

1. Stimulus N-cadherin BOUND → QUIESCENCE SV ((+)value)

2. Stimulus VCAM1 BOUND → QUIESCENCE SV ((+)value)

3. Stimulus OPN BOUND → QUIESCENCE SV ((+)value)
   APOPTOSIS SV ((+)value)

4. Stimulus IL1 BOUND → Sym DIVISION SV ((+)value)

5. Stimulus IL3 BOUND → Sym/Asym DIVISION SV ((+)value)

6. Stimulus IL7 BOUND → Sym/Asym DIVISION SV ((+)value)

7. Stimulus EPO BOUND → Sym/Asym DIVISION SV ((+)value), Ery

8. Stimulus GM-CSF BOUND → Sym/Asym DIVISION SV ((+)value), (Neut,Macr)

7.1.4 Proactive Rules

As for the abstract model proposed we defined a set of proactive rule which are triggered depending on the cell agent state, i.e. when one or more state variables reach a specific value.

Apoptosis Proactive Rule

This proactive rule is triggered when the APOPTOSIS state variable reach the one value. When fired, it execute the action of `die()`.

Differentiation Proactive Rule

This proactive rule is triggered when a DIFFERENTIATION variable versus a specific phase reach the one value. Once it is triggered the differentiation process starts. It ends with the transformation of the cell from one phase to the target phase associated to the specific DIFFERENTIATION state Variable.

Produce Molecule Proactive Rule

This rule implements the production of the molecules of the desired type, in the desired number.

Symmetric Division Proactive Rule

This proactive rule models the cellular behaviour during the cell-division cycle. The cell cycle consists of several phases. In the first *phases* G_1, the cell growth and become larger. When it was reached a certain size it enters in the next *phase* S, in which DNA synthesis takes place. The cell duplicates its hereditary material and a copy of each chromosome is formed. During the next *phase* G_2, the cell check that DNA-replication is complete and prepares for cell division. Chromosomes are separated in the next *phase* M and the cell divides in two daughter cells. After division the cells are back in G1.

It is essential that the different phases are precisely coordinated. The phases must follow in correct order, and one phase must be completed before the next phases can begin.

Our modelled rule of division is so composed by 5 different proactive rules, each for each different phase of the cell cycle. We have:

1. Symmetric Division Proactive Rule phase G_1
2. Symmetric Division Proactive Rule phase S
3. Symmetric Division Proactive Rule phase G_2
4. Symmetric Division Proactive Rule phase Mitosis
5. Symmetric Division Proactive Rule phase Cytokinesis (cytoplasm physical division)

The first rule is triggered every time that the SYMMETRIC DIVISION state variable exceeds the value 1. It launches the division process. The next rules are triggered every time that the control's conditions are verified, so as to model the precise coordination and control among the different phases of the cell cycle. The control's conditions are also modelled through state variables which represent that gene/proteins necessarily express in the different phases.

Cells that have temporarily or reversibly stopped dividing are said to have entered a state of quiescence called G_0 phase, while cells that have permanently stopped dividing due to age or accumulated DNA damage are said to be senescent. We model the first cell state with the state variable of QUIESCENCE, which has value 1 in phase G_0. To enter in the division process through the phase G_1, this variable has to have a value minor or equal to 0, representing that the cell is no more in a state of quiescence. And finally we model with an other state variable of SENESCENCE equal to one the concept that the cell can no more divide.

Asymmetric Division Proactive Rule

We have also modelled the process of asymmetric division. It is really similar to the rule of symmetric division, and it develops along four different steps:

1. Asymmetric Division Proactive Rule step 1

2. Asymmetric Division Proactive Rule step 2
3. Asymmetric Division Proactive Rule step 3
4. Asymmetric Division Proactive Rule step 4

The last step has a consequence to produce two cells: one becomes differentiated in the target phase and the other becomes equal to the original cell.

Change Direction Proactive Rule

This rule implements the movement of the cell, individuating step by step a casual new directions of the velocity vector.

7.1.5 Cell Phases

The phase that we have considered in our model are indicated in Figure 7.1. Note that some complexity is omitted from the diagram. The complete diagram can be found in [2]. Lymphocytes come from Lymphoid line, while granulocytes, monocytes, megakaryocytes, and erythrocytes come from Myeloid line. Among myeloid cells, granulocytes and monocytes/macrophage have a common precursor, CFU-GM.

Each phase is identified by:

1. a name
2. a shape (which include also dimension and colour) [`shape(name,width,height,colour)`]

and is composed by a set of:

1. proteins [`protein(name, number)`];
2. state variables [`state_variable(name, initial_value)`];
3. of reactive rules [`reactive_rule(triggering_STIMULUS,(target_variable(Δ Value)))`];
4. of proactive rules [`proactive_rule(name,(parameters))`].

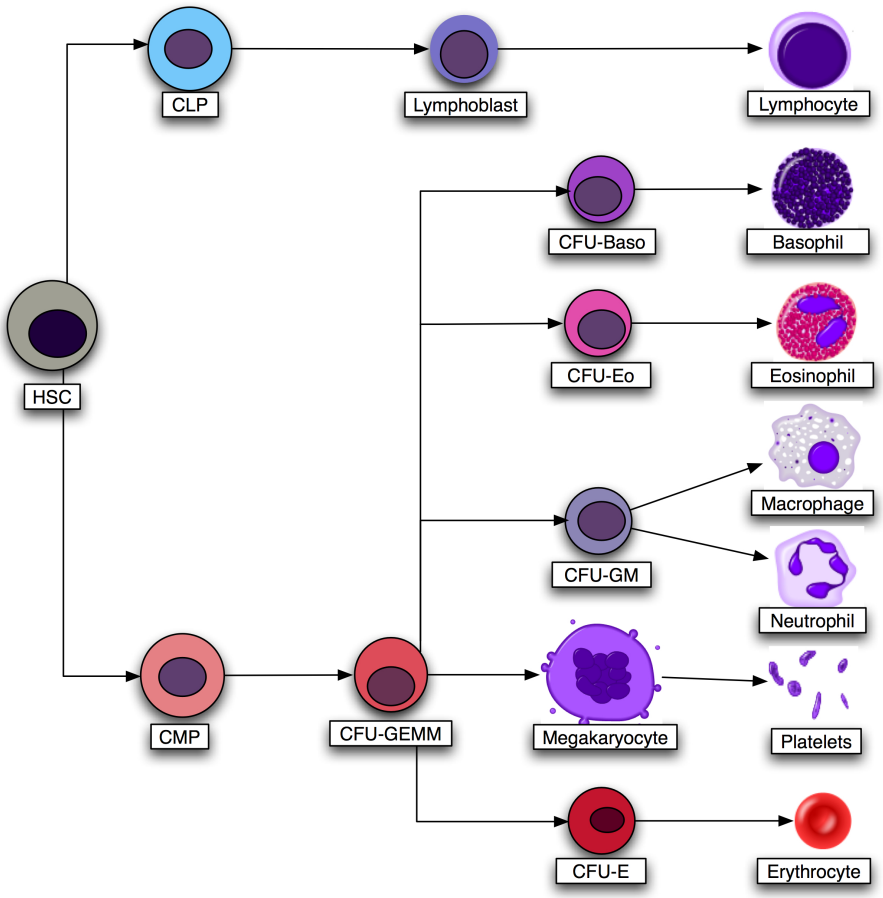


Figure 7.1: Diagram of Hematopoietic Stem Cell Lineage

Hematopoietic Stem Cell Phase

We define here the specific components of the HSC are the precursor cells which give rise to all the types of both the myeloid and lymphoid lineages. The others are obviously different in the type and number of proteins, in the description of reactive and proactive rules. We omitted these particulars for not making this model presentation too heavy.

```
shape(round,0.1,0.1,grey)

protein(CD44,x)
protein(IL1receptor,x)
protein(IL3receptor,x)
protein(IL6receptor,x)
protein(IL7receptor,x)
protein(Kit,x)
protein(N_cadherin,x)
protein(Notch,x)
protein(VLA4,x)

state_variable(DIFFERENTIATION,0.0)
state_variable(symDIVISION,0.0)
state_variable(asymDIVISION,0.0)
state_variable(APOPTOSIS, 0.0)
state_variable(ADHESION, 1.0)
state_variable(QUIESCENCE, 1.0)

reactive_rule(STIM_Jagged_BOUND,
 (symDIVISION(0.005),QUIESCENCE(-0.005),DIFFERENTIATION(-0.05)))
reactive_rule(STIM_N_cadherin_BOUND,(QUIESCENCE(0.005)))
reactive_rule(STIM_SCF_BOUND,
 (symDIVISION(0.005),QUIESCENCE(-0.005),APOPTOSIS(-0.005)))
reactive_rule(STIM_VCAM1_BOUND,(QUIESCENCE(0.05)))
reactive_rule(STIM_OPN_BOUND,(QUIESCENCE(0.05),APOPTOSIS(0.005)))
reactive_rule(STIM_IL1_BOUND,
 (symDIVISION(0.005),QUIESCENCE(-0.005),DIFFERENTIATION(0.001)))
reactive_rule(STIM_IL3_BOUND,(asymDIVISION(0.05),QUIESCENCE(-0.005)))
reactive_rule(STIM_IL6_BOUND,
 (symDIVISION(0.005),QUIESCENCE(-0.005),DIFFERENTIATION(0.0005)))
reactive_rule(STIM_IL7_BOUND,(asymDIVISION(0.05),QUIESCENCE(-0.05)))


proactive_rule(ApoptosisProactiveRule)
proactive_rule(SymDivisionProactiveRule)
proactive_rule(AsymDivisionProactiveRule,PH_CMP)
```

```
proactive_rule(AsymDivisionProactiveRule,PH_CLP)
proactive_rule(DiffProactiveRule,PH_CMP)
proactive_rule(ChangeDirProactiveRule,0.001)
proactive_rule(QuiescenceProactiveRule);
```

7.2 Preparing and Executing Virtual Experiments

STEP 4. Select output values and measurements.
STEP 5. Select simulation software.
STEP 8. Identify and do useful experiments.

First Experiment

It is now generally accepted that all blood cells are made from a relatively few uncommitted cells which are capable of mitosis and of differentiation into committed precursors of each of the main types of blood cell. In this first experiment, we start with a single stem cell and we verify that the cell properly reacts to all the different kind of stimuli, executing the right actions.

From the execution of the experiment we verified — as expected — that an individual cell stem, even if equipped with all the protein membranes, keeps quiescent, starts the symmetric / asymmetric division and follows a specific path of the lineage depending on the molecules that are found in the niche.

In particular we show here what happen if we compose the environment with molecules of Interleukin 3 and Erythropoietin [Figure 7.2].

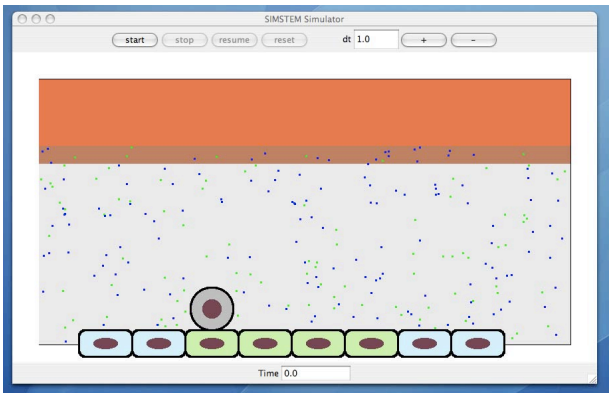


Figure 7.2: Simulation Snapshot 1

The expected behaviour is the final production of a set of Erythrocytes passing through the intermediate phases of Colony forming units (pink), Colony forming Unit-GEMM (light red) and Colony Forming Unit-E (dark red) [Figure 7.3].

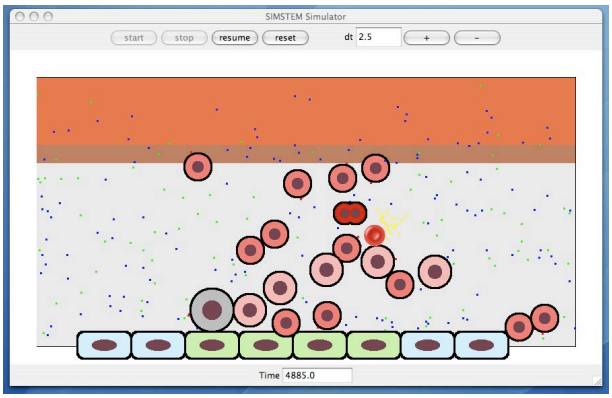


Figure 7.3: Simulation Snapshot 2

When the mature cells are produced they pass in the blood niche and starts to produce TGFβ (the yellow molecules) which inhibit the division of CMP while accelerate the differentiation of CFU-GEMM and CFU-E, so that the number of Erythrocytes does not become too high [Figure 7.4].

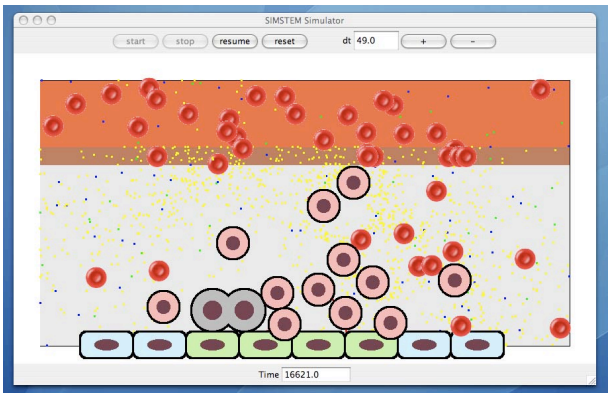


Figure 7.4: Simulation Snapshot 3

Finally, when the Erythrocytes die, they produce IL3 and EPO so that reactivate the progenitor and maintain the set of red blood cells essential for the organism life. With these consecutive steps we have

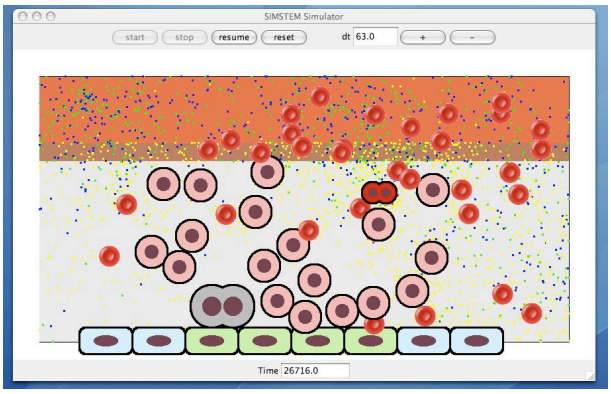


Figure 7.5: Simulation Snapshot 4

modelled the mechanisms of feedback control and regulation.

Second Experiment

In this second experiment, again we start with a single HSC, situated in a niche containing all the types of molecules as in the real case. In

this case we can observe that – as expected – different kind of mature cells are finally generated, since the choice of the lineage path to follow is not deterministic, since the same signalling molecule can trigger the division along multiple paths.

Third Experiment – normal turnover

In this third experiment, we introduce a second niche, representing the blood. The blood niche functions as collector of all the mature cells, which are meant to migrate from the niche to the blood and then be spread over other niches (tissues). For each cell a maximum lifetime is specified: when a mature cell dies, it produces some signalling molecules that trigger its production in the bone marrows, which functions as a pool containing (not mature) hematopoietic stem cell. By executing the experiment, we can observe the *turnover* of the cells.

Fourth Experiment – hypoxia

This experiment explores the behaviour of the cell system reacting to a global condition mimicking some kind of physiological abnormality. We can observe an emergent behaviour of the cell population in the overall that finally results in removing the abnormality, bringing back the system to a normal situation.

As a specific example, we consider Erythropoiesis, which is is stimulated by hypoxia (lack of oxygen). The lack of oxygen does not act directly on the hemopoietic tissues, but instead stimulates the production of a hormone, erythropoietin. This hormone then stimulates hemopoietic tissues to produce red cells. Erythropoietin is a glycoprotein. It is inactivated by the liver and excreted in the urine. It is now established that erythropoietin is formed within the kidney by the action of a renal erythropoietic factor erythrogenin on plasma protein, erythropoietinogen.

Fifth Experiment – disease response

The last experiment investigates the system dynamics in the case of a physiological condition of disease, represented in particular by the

presence of an infection. The hematopoietic cell system is triggered so as to activate a large number of white blood cells.

7.3 Calibrating and Analysing the Results

Following the methodology, last points of the overall process consist in:

STEP 7. Calibrate system and bug fixing;
STEP 9. Analyse simulation results.

For the former point, the model adopted makes it possible to calibrate the system by:

- acting on the *local* properties and parameters of the individual parts, in particular by tuning the structure and behaviour of the cell agents. Examples of this kind of calibration includes tuning the number of membrane proteins, changing the behaviour of the proactive rules, tuning the timings taken by a cell to realise some kind of process.

- acting on the environment, by tuning the number of molecules — and possibly of cells — contained in the niches.

For the latter point, the virtual laboratory application — that is, the simulator — makes it possible to dynamically observe and analyse the evolution of the systems dynamics both from a qualitative point of view, through visualisation, and from a quantitative point of view, making it dynamically inspectable the numerical value of some basic parameter, such as the number of cells of a certain kind contained in the niche in a specific moment.

Chapter 8

Conclusions

The objective of this master thesis was twofold: on the one side, to investigate the adoption of multi-agent systems as an approach for modelling and simulating complex biological systems, and as a mean to realise the methodological principles suggested by Systems Biology; on the other side, to experiment concretely the approach, by introducing a first model based on a multi-agent system to model the dynamics of systems of cells interacting inside some niches, taking as a reference case study the hematopoietic cell stems system, and by developing a basic computational tool for supporting the execution of proper simulations. The tool is meant to work as a kind of *virtual laboratory*, on top of which kinds of *virtual experiments* can be performed, characterised by the definition and execution of specific models implemented as MASs, so as to support the validation, falsification and improvement of the models through the observation and analysis of the simulations.

First of all, the results achieved have confirmed the effectiveness of multi-agent systems as a tool for modelling and simulating such complex systems as the ones focussed by this thesis, that is biological systems, cell systems in particular. Differently from traditional modelling approaches, which account for the *a-priori* formulation of the global laws that govern systems' evolution, the approach presented in this thesis starts from modelling the behaviour of the individual parts of the system and *their interaction*, obtaining the global properties as emergent properties of system execution, observable *a-posteriori*. In

the specific case study considered in chapter seven, the global properties of the behaviour of the hemapoietic stem cell system emerge from the behaviour of the individual stem cells and their interaction within the bone marrow and blood niches, and can be observed by executing the multi-agent systems modelling the cell population, executing the simulation. This makes the approach an effective complimentary tool to be used with traditional approaches, useful to investigate those aspects that are not directly captured by those approaches, such as individual behaviours, interactions, strong non-linearities, emergent behaviours and properties.

Then, the results achieved show that the specific MAS model introduced for modelling cell populations and the related computational tool for executing the simulations — the virtual laboratory, implemented on top of the Java language and platform —have been quite effective and flexible in supporting the implementation and execution of virtual experiments. In particular, the tools, in spite of their simplicity, provide a first concrete support for scientist / biologist / engineer users to *(i)* describing the behaviour of the individual parts and their interaction quite intuitively, adopting a basic set of quite powerful high-level abstractions, reducing the gap between the real system to be modelled and the virtual counterpart; *(ii)* formulating models — by integrating hypothesis and experimental data — using different levels of detail, which can be refined incrementally, as soon as new knowledge and hypothesis are available; *(iii)* exploiting a simulation environment that gives full control on the simulation, making it possible not only to observe but also interact with the simulated system in execution, changing — for instance — the structure of the system by introducing or removing molecules in niches, dynamically.

Bibliography

[1] A portal site for systems biology
`http://www.systems-biology.org/`.

[2] A portal site of GenomeNet Database Service
`http://www.genome.jp/dbget-bin/show_pathway?hsa04640+3562`.

[3] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular biology of the cell - 4 ed.* Garland Science, New York, 2002.

[4] E. C. Attar and D. T. Scadden. Regulation of hematopoietic stem cell growth. *Leukemia*, 18(11):1760–1768, November 2004.

[5] P. Ballet, J. Tisseau, and F. Harrouet. A multiagent system to model an human humoral response. *Systems, Man, and Cybernetics, 1997. 'Computational Cybernetics and Simulation'., 1997 IEEE International Conference*, 1:357–362, Oct 1997.

[6] N. Cannata, F. Corradini, E. Merelli, A. Omicini, and A. Ricci. An agent-oriented conceptual framework for systems biology. In C. Priami, E. Merelli, P. P. Gonzalez, and A. Omicini, editors, *Transactions on Computational Systems Biology III*, volume 3737 of *Lecture Notes in Computer Science*, pages 105–122. Springer, 2005.

[7] F. Capra. *The Web of Life. A New Synthesis of Mind and Matter.* Flamingo, London, 1997.

[8] L. Cardelli. Abstract machines of systems biology. *Transactions on Computational Systems Biology. III*, 3737:145–168, September 2005.

[9] E. D. Conrad and J. J. Tyson. Modeling Molecular Interaction Networks with Nonlinear Ordinary Differential Equations. In Z. Szallasi, J. Stelling, and V. Periwal, editors, *System Modeling in Cell Biology - From Concepts to Nuts and Bolts*, chapter 6, pages 97–123. MIT Press, 2006.

[10] H. De Jong. Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. *Journal of Computational Biology*, 9(1):67–103, 2002.

[11] H. de Jong and D. Ropers. Qualitative approaches to the analysis of genetic regulatory networks. In Z. Szallasi, J. Stelling, and V. Periwal, editors, *System Modeling in Cell Biology - From Concepts to Nuts and Bolts*, chapter 7, pages 125–147. MIT Press, 2006.

[12] P. Dhar, H. Zhu, and S. Mishra. Computational approach to systems biology: From fraction to integration and beyond. *IEEE Transactions on Nanobioscience*, III(3):144–152, Sept. 2004.

[13] M. d'Inverno and J. Prophet. Multidisciplinary investigation into adult stem cell behavior. In C. Priami, E. Merelli, P. P. Gonzalez, and A. Omicini, editors, *Transactions on Computational Systems Biology III*, volume 3737, pages 49–64, 2005.

[14] M. d'Inverno and R. Saunders. Agent-based modelling of stem cell self-organisation in a niche. In *Engineering Self-Organising Systems*, pages 52–68, 2004.

[15] M. A. Gibson. *Computational methods for stochastic biological systems*. PhD thesis, 2000. Adviser-Jehoshua Bruck.

[16] M. A. Gibson and J. Bruck. A Probabilistic Model of a Prokaryotic Gene and Its Regulation. In J. M. Bower and H. Bolouri, editors, *Computational Modelling of Genetic and Biochemical Network*, chapter 2, pages 49–71. MIT Press, 2001.

[17] M. A. Gibson and E. Mjolsness. Modeling the Activity of Single Genes. In J. M. Bower and H. Bolouri, editors, *Computational*

Modelling of Genetic and Biochemical Network, chapter 1, pages 3–48. MIT Press, 2001.

[18] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977.

[19] V. V. Ivanov and N. V. Ivanova. *Mathematical Models of the Cell and Cell Associated Objects*. Elsevier Science Ltd, 2006.

[20] S. A. Kauffman. *Origins of order: Self-organization and selection in evolution*. Oxford University Press, Oxford, 1993.

[21] D. B. Kell and J. D. Knowles. The Role of Modeling in Systems Biology. In Z. Szallasi, J. Stelling, and V. Periwal, editors, *System Modeling in Cell Biology - From Concepts to Nuts and Bolts*, chapter 1, pages 3–18. MIT Press, 2006.

[22] H. Kitano. Systems biology: A brief overview. *Science*, 295:1662–1664, March 2000.

[23] H. Kitano. Computational systems biology. *Nature*, 420:206–210, November 2002.

[24] F. Klügl and E. Norling. Agent-based Simulation: Social Science Simulation and Beyond. Technical report, The Eighth European Agent Systems Summer School (EASSS 2006), Annecy, July 2006.

[25] F. Klügl, C. Oechslein, F. Puppe, and A. Dornhaus. Multi-agent modelling in comparison to standard modelling. *Artificial Intelligence, Simulation and Planning in High Autonomy Systems*, pages 105–110, 2002.

[26] K. Kruse and J. Elf. Kinetics in spatially extended systems. In Z. Szallasi, J. Stelling, and V. Periwal, editors, *System Modeling in Cell Biology - From Concepts to Nuts and Bolts*, chapter 9, pages 177–198. MIT Press, 2006.

[27] O. Mason and M. Verwoerd. Graph theory and networks in biology, March 2006.

[28] H. Maturana. Autopoiesis, Structural Coupling and Cognition: A history of these and other notions in the biology of cognition. Technical report, Biologia University of Chile, August 2002.

[29] P. Mendes. Gepasi: a software package for modelling the dynamics, steady states and control of biochemical and other systems. *Comput. Appl. Biosci.*, 9(5):563–571, October 1993.

[30] T. C. Meng, S. Somani, and P. Dhar. Modelling and simulation of biological systems with stochasticity. *In Silico Biology*, 4(3):293–309, 2004.

[31] E. Merelli, G. Armano, N. Cannata, F. Corradini, M. d'Inverno, A. Doms, P. Lord, A. Martin, L. Milanesi, S. Mller, M. Schroeder, , and M. Luck. Agents in bioinformatics, computational and systems biology. *Briefings in Bioinformatics*, pages 1–15, 2006.

[32] S. Montagna, A. Ricci, and A. Omicini. Agents & Artifacts for Systems Biology: Toward a framework based on TuCSoN. In A. Genco, A. Gentile, and S. Sorce, editors, *Industrial Simulation Conference 2006 (ISC 2006)*, pages 25–32, Palermo, Italy, 5–7 June 2006. EUROSIS (The European Simulation Society) & ETI (The European Technology Institute).

[33] C. A. J. Morton-Firth and D. Bray. Stochastic simulation of cell signalling pathways. In J. M. Bower and H. Bolouri, editors, *Computational Modelling of Genetic and Biochemical Network*, chapter 9, pages 263–286. MIT Press, 2001.

[34] C. J. Morton-Firth. *Stochastic Simulation Of Cell Signalling Pathways*. PhD thesis, University of Cambridge, 1998.

[35] G. Nicolis and I. Prigogine. *Exploring Complexity: An Introduction*. W. H. Freeman and Company, New York, October 1989.

[36] J. Paulsson and J. Elf. Stochastic Modeling of Intracellular Kinetics. In Z. Szallasi, J. Stelling, and V. Periwal, editors, *System Modeling in Cell Biology - From Concepts to Nuts and Bolts*, chapter 8, pages 149–175. MIT Press, 2006.

[37] V. Periwal. Bayesian inference of biological systems: The logic of biology. In Z. Szallasi, J. Stelling, and V. Periwal, editors, *System Modeling in Cell Biology - From Concepts to Nuts and Bolts*, chapter 4, pages 53–71. MIT Press, 2006.

[38] I. Prigogine and I. Stengers. *La Nouvelle Alliance. Les Métamorphoses de la Science*. Gallimard, Paris, February 1986.

[39] A. Regev, E. M. Panina, W. Silverman, L. Cardelli, and E. Shapiro. BioAmbients: An abstraction for biological compartments. *Theoretical Computer Science*, 325(1):141–167, 2004.

[40] A. Regev, W. Silverman, and E. Y. Shapiro. Representation and simulation of biochemical processes using the pi-calculus process algebra. In *Pacific Symposium on Biocomputing*, pages 459–470, 2001.

[41] A. Ricci, M. Viroli, and A. Omicini. Programming MAS with artifacts. In R. P. Bordini, M. Dastani, J. Dix, and A. El Fallah Seghrouchni, editors, *3rd International Workshop "Programming Multi-Agent Systems" (PROMAS 2005)*, pages 163–178, AAMAS 2005, Utrecht, The Netherlands, 26 July 2005.

[42] E. E. Santos, D. Guo, E. S. Jr., and W. Onesty. A multi-agent system environment for modelling cell and tissue biology. In *PDPTA*, pages 3–9, 2004.

[43] C. M. Smith. Hematopoietic stem cells and hematopoiesis. *Cancer Control*, 10(1):9–16, Jenuary/February 2003.

[44] A. M. Uhrmacher, D. Degenring, and B. Zeigler. Discrete Event Multi-level Models for Systems Biology, June 1990.

[45] A. M. Uhrmacher, D. Degenring, and B. Zeigler. Discrete Event Multi-level Models for Systems Biology. In C. Priami, editor, *Transactions on Computational Systems Biology I*, volume 3380 of *Lecture Notes in Computer Science*, pages 66–89. Springer, 2005.

[46] A. Wilson and A. Trumpp. Bone-marrow haematopoietic stem-cell niches. *Nature Reviews Immunology*, 6(2):93–106, February 2006.

[47] M. Wooldridge. Intelligent Agents. In G. Weiß, editor, *Multiagent Systems - A Modern Approach to Distributed Artificial Intelligence*, pages 27–77. MIT Press, Cambridge, Massachusetts, USA, April 1999.