Stochasticity in signal transduction pathways

Vidal Rodriguez, J.

Publication date
2009

Chapter 1

Introduction

One of the fascinating aspects of micro-organisms is their ability to endure and self-sustain under very diverse and changing environmental conditions. This applies not only to complex higher organisms made up of trillions of cells, but also unicellular micro-organisms, such as bacteria. Higher organisms may depend on a tight symbiotic relationship with many different bacteria present in their bodies; the human flora. This capacity to endure is attributed to be a response to evolutionary pressures that natural selection exerts uninterruptedly on every organism. Under this umbrella of evolution we can argue that all organisms are highly optimised to thrive in the environmental niche they usually inhabit, in so much as they are permitted to by the frequent short-term fluctuations, or slow long-term alterations, that threaten their life. As Kitano (2004) argues, in order to survive, life has developed extremely robust and reliable, yet flexible, complex systems that allow organisms to resist alterations in their ideal environmental conditions and to adapt to new ones.

The basic mechanisms that enable endurance and adaptability to changes seem to be universal (Szallasi et al., 2006, chap. 2). After all, we are very familiar with these type of mechanisms in higher organisms such as ourselves. For example, our visual organs tell us about changes in the physical conditions surrounding us. The response to a drop in the ambient light, results in the dilatation of the iris. Another example is that our body starts sweating in response to increases in the ambient temperature. These automatic mechanisms require essentially two abilities: sensing and responding. Organs rely on individual cells to perform sensory tasks, while the response mechanism may be a long chain of events initiated, however, at the sensory cell. In our light example, the signal is processed in the many sensory photoreceptor cells on the retina that react to the incoming photons and trigger electrical signals transported to the image processing part of the brain.

Unicellular micro-organisms have simple sensory-response mechanisms working at a molecular level. Recent growing genome-wide analyses of micro-organisms reveal an abundant number of genes involved in these sensory systems: around 1% of genes in eubacteria are for external signal sensing (West and Stock, 2001),

and the number of genes for internal signals is fourfold (Ulrich et al., 2005). Many of them still remain to be studied in detail and tied to the environmental conditions, or physical signals, they respond to. To implement these systems, a bacteria is constrained by the amount of resources and energy it devotes to these to build and sustain biochemical processes in a small, confined space. It seems there is little room for what is unnecessary, yet the presence of redundant systems may play an important role as long as they provide an enhanced robustness, endurance or adaptability capabilities (Kitano, 2004). This suggests that the regulation of the amount of proteins is optimised to support the endurance of the cells under the variable local conditions. In this line, evolutionary pressures favour minimisation of energy and resource usage, and optimisation of spatial organisation. It is as if organisms are limited by an energy budget, albeit with reserves. These limitations imply that the number of proteins ought to be kept at a low copy number as possible. Such resource economy apparently contradicts the view that robust and reliable systems require many molecules in order to reduce the stochasticity inherent to reactions. Under these conditions, the stochastic nature of chemical processes may have strong consequences in the cell's fate. A view has emerged in which stochastic phenomena at the single cell level are no longer a nuisance that interferes with other signals (Fedoroff and Fontana, 2002).

Taking a molecular-centric computational modelling approach we aim at unravelling general influential biochemical aspects in the functioning and organisation of sensory systems in bacteria in their spatial context, their stochastic and discrete nature as well as their genomic organisation. This shall enable us to study the global behaviour that arises from the interactions among the building blocks such as proteins and genes.

In the remainder of this introduction we describe the role that computers play in assisting cellular biology. It follows a description of the principal of general modelling principles and the computer methods used to solve complex models. Then we give a brief introduction to the concrete types of signalling systems and gene expression used in this thesis.

1.1 The Computer in Systems Biology

A major difficulty in experimental cell biology is the observation of detailed molecular mechanisms under, ideally, *in vivo* conditions in their spatio-temporal dimension and single molecule resolution, without disrupting the normal cell's functioning. To overcome the limitations of this direct approach, cell biologists have turned to computational modelling, or *in silico* modelling, as a complementary tool to aid to "build and test complex hypothesis" (Slepchenko et al., 2002). During the last three decades the use of computational methods have contributed substantially in furthering the understanding of physico-biochemical processes, from simple unicellular bacteria to multicellular complex organs (Noble, 2002 ; Mellman and Misteli, 2003). As Ridgway et al. (2006) highlights, "Interest in the possibility of dynamically simulating complex processes has escalated markedly

Table 1.1: *The publication of new journals on the boundary of biology and computational science attests to the importance of the co-operation of both fields as to forward the understanding of biology and especially of cellular phenomena.*

Journal	year 1st publication
BMC Systems Biology	2007
PLoS Computational Biology	2005
Molecular BioSystems	2005
in Silico Biology	1998
Bioinformatics	1985

in recent years". The range of computational approaches embraces not only the dynamic and often complex molecular aspects common to a computational cell biology approach, but also the vast static information, or bioinformatic data, gathered over the years such as DNA sequences, micro-array data or even the literature (data mining). Such computational-based approaches owe their success partly to the extensive elementary knowledge provided by the discipline of biochemistry and bioinformatics on the basic components such as DNA structure, protein functions and metabolic activity, and partly to the discipline of biophysics and mathematics to characterise and formulate in a rigourous mathematical framework the basic interactions and processes underlying the most elemental biological processes such as that of diffusion of particles in a solvent (Takahashi et al., 2002 ; Palsson, 2000). These come accompanied with an increase in the computational capabilities commonly accessed at a desktop level.

1.1.1 The Research Cycle

A symbiotic relationship has been forged in the last half century between traditional wet experiment science and computer science. This symbiosis is acknowledged to be paramount to understanding complex biological phenomena (Kitano, 2002 ; Westerhoff and Palsson, 2004): computational methodologies form an integral part of the biological research cycle. The convergence of fields is well illustrated and supported by the recent publication of new journals with emphasis on the computational aspects of biological simulations (see Table 1.1).

Let us illustrate the role of computational modelling in a typical research cycle for computational (cell) biology (see Fig. 1.1 for a general overview). After a thorough study of the subject system, we produce a tentative model (step 1). This model, of course, already identifies and summarises the relevant biological aspects in the form of, for instance, a reaction model such as:

$$\begin{aligned} A + B &\overset{k_1}{\rightarrow} C + P \\ P &\rightarrow Z \end{aligned}$$

where each letter represents a biochemical element, and that upon reaction of, say, $A + B$ they produce two molecules with different properties that we label as C and D. Whenever possible, analytical tools are preferred to solve the model, because they provide a precise insight into the system's control and interactions among its components (step 2). For instance, the solution of the change in concentration of P is:

$$P(t) = C(P; t = 0) \exp(-kt).$$

For complex systems, however, obtaining algebraic solutions is not always feasible. In those cases where the complexity of the problem surpasses the mathematical capabilities —not uncommon in biology— computer numerical procedures, or simulations, are used (step 3). Numerical solutions yield instances of a solution, that is, one particular solution, given for each set of input parameters. However, in these explicit solutions the algebraic relationships between the variables, as expressed in the solution above, are lost; notwithstanding, we have a solution. Just as with mathematical models, computational models also need to rely on fundamental knowledge that guarantees the correctness of the procedure (step 4). The degree of formalism present in these bases of knowledge ultimately determines the reliability of the final solution. It is not uncommon to combine analytical and numerical methods to produce collaboratively a result. In these hybrid solutions a particular method is applied over a subset of the problem that is then fed as input to the other solver, and vice versa (steps 5 and 6). The final results gathered are interpreted by the modeller (step 7). If the results agree with the observed data our model is valid, if not, then the tentative model needs to be modified in order to overcome its limitations and repeat the cycle over (step 8).

In summary, extensive fundamental knowledge gathered by several traditional disciplines provides the necessary tools for computational biology to be used as an analytical tool with an acceptable degree of confidence (Takahashi et al., 2005).

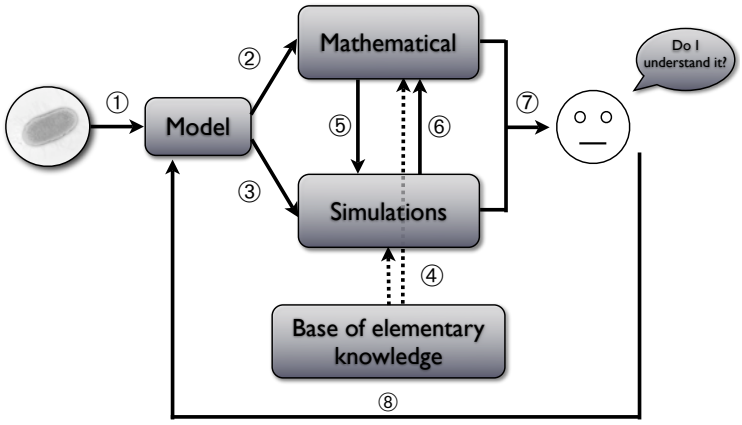

Figure 1.1: *Computational biology research cycle. Computer aided numerical solutions or simulations form an integral part in the path to understanding biological phenomena.*

1.2 Models: Purpose and Focus

Biological phenomena are in many cases complicated. The numerous reaction involved to understand some systems may be large. Additionally, the often nonlinear reactions contribute to the rise of complex pehnomena. For instance, *E. coli* has over 4190 translated proteins, 4377 genes[1], and 732 metabolites and over 932 reactions (Kim et al., 2007, and refs. therein). Complicate and complex phenomena are also observed in other temporal and spatial scales: from protein folding, to dynamics of whole ecosystems (Szallasi et al., 2006). In view of these overwhelming numbers, the role of modelling is to focus on the relevant parts necessary to understand a specific phenomenon, with the purpose being to elucidate the interactions among the components in a, preferably, quantitative way. In this manner, model focusing helps to cope with the complexity of the real system. As Kell and Welch say "our scientific heritage (in the Greek tradition) is to reduce in order to understand". The prevailing reductionist approach reduces the complexity of a system in two orthogonal directions: 1) scope, by focusing on smaller subsystems, and 2) detail, revealing the molecular or atomistic components as in a hierarchy (as depicted in Fig. 1.2(a)).

In this space the modeller is left to choose the right level of scope and detail necessary to describe the systems. However, modelling and computational limitations still pose some limitations to reach the whole. We find ourselves in a constrained position where there is a trade off between a molecular detail and a narrow scope system, or a coarse detail and a broad system scope (along the reductionist line in Fig. 1.2(a)).

[1] `http://redpoll.pharmacy.ualberta.ca/CCDB/cgi-bin/STAT_NEW.cgi`

1.2.1 Macro, Meso and Microscopic Scales

We can argue that the number of state variables increases with an increase in detail and scope. The complexity as the number of possible interactions among the components increases as well. By focusing on different levels of molecular detail we can identify a hierarchy of three layers. As illustrated in Fig. 1.2(b), at the top of this hierarchy a macroscopic paradigm is used to describe the fluxes among biochemical species, as these react with each other creating new, or modifying the properties of, molecules. Perhaps the archetypical example in this scale is that of metabolic networks, and the associated modelling method of differential equations (DE). Macroscopic models exclude molecular or atomistic details, allowing a broadening of the system's scope.

A mesoscopic description level emphasises the discrete nature of chemical systems and focuses on the individual molecules, such as proteins. Compared to a macroscopic level, there is an increase in the spatial and temporal resolution. Global behaviour emerges from the interactions between particles, which increments the number of variables needed to characterise the state, making the system more complicated, but not necessarily more complex. Nevertheless, the rules that define the interactions between components are simplified because each molecule is abstracted as a single particle, instead of multiple constituting atoms. We would associate such a level of detail with a microscopic scale, suitable to perform detailed studies with virtual representations of the biomolecules and their atomic interactions. The most salient example of these types of modelling is that of protein folding.

1.2.2 Regimes

As emphasised previously, biological processes vary largely in terms of three important factors: the spatial scale, the temporal scale and the number of molecules involved in the process. For example, the collision of two molecules in a high concentration medium is on the order of milliseconds, while in a low concentration medium these may take seconds to minutes. The size of a typical protein is $5\,nm$, while the host cell, for instance *E. coli* has a typical volume of $1-3\,\mu m^3$, which means that a protein is $\approx 10^9$ times smaller. These broad ranges of spatial, temporal and molecular scales might present a challenge for models and, especially, for the methods used to solve them, either analytical or computational. Adding more detail to a model increases the number of variables and, consequently, its complexity. So as models aim at narrowing the scope of parameters in order to reduce the complexity of the model, specialised methods have been developed to tackle a particular scale of the parameters in an efficient manner.

As illustrated in Fig. 1.3, we can distinguish four regimes depending on the abundance of molecules and the degree of heterogeneity of the system. In contrast to the previous scale hierarchy, this classification emphasises the spatial dimension and discreteness of the chemical systems.

The regime that is probably the most widely used in cell biology is characterised by a large number of molecules, a spatially homogeneous system and

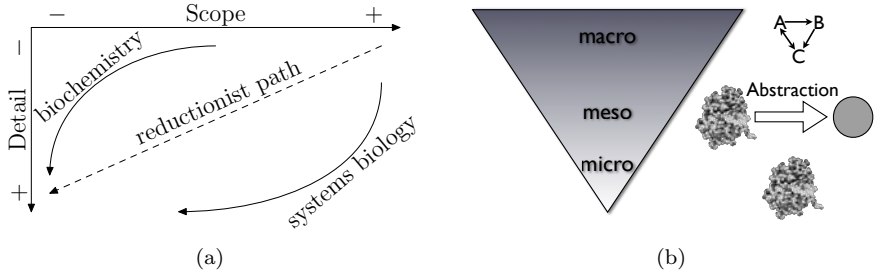

Figure 1.2: *Panel (a) a reductionist path reduces the scope of the system while increasing the level of detail. Panel (b) shows the three most common abstraction levels and typical models in each level: a reaction network which may be solved with differential equations, a particle-base model where the complexity of a proteins is abstracted into a spherical object with equivalent mesoscopic properties, and microscale simulations for protein-folding.*

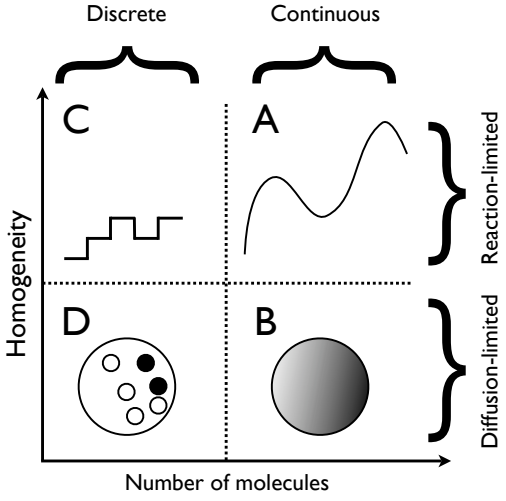

Figure 1.3: *Classification of biochemical regimes according to two fundamental criteria: number of particles and systems' homogeneity. The dynamics of the system may be described deterministically or stochastically. Transition between regimes is gradual and large systems may span over several regimes*

reaction-limited kinetics(A). As a result, the dynamics can be described in a temporally continuous-manner and using differential equations. Typical examples in this regime include metabolic networks.

When the kinetics of the systems become limited by diffusion, spatial heterogeneity arises (B). This may arise due to a combination of high affinity and a relatively slow diffusion, or a localisation of reactions, for instance, on the membrane. In this case, molecules perform few non-reactive collisions before reaction occurs. Because of the high number of particles these systems can also be described in a continuous manner, in both time and space.

In the two regimes described so far, on the right half of Fig. 1.3,the dynamics of the systems can be described in a deterministic way. This is possible in the presence of a high number of particles, as the size of the ensemble averages out the individual fluctuations in reaction and diffusion. In other words, the stochastic component of each individual reaction is negligible compared to the dynamics as a whole.

Regime C and D are discrete versions of their continuous counterparts. Their main difference resides in that small changes in the number of molecules present have a larger impact in the overall population. Under these circumstances, opposed to the previous A and B regimes, the stochastic nature of reactions is significant. In this scenario deterministic methods may fail to capture essential dynamics. We shall see some examples of this situation in section 1.3.

The clear division between regimes is only for illustrative purposes. Instead, transitions between regimes are gradual and in some cases a phenomenon may fall into more than one regime. A system may also span over several of these regimes. Multi-regime systems shall be more common as more detail for more systems is required. This regime classification, together with the scale hierarchy, are a reflection of the different types of computational methods available to model the rich diversity of biological phenomena.

1.3 Stochastic Phenomena in Bacteria

Much of the interest in stochasticity of biological processes arises because traditional deterministic models fail to capture the intrinsic random phenomena of reactions. In systems such as signalling and gene expression, dynamic behaviour and spatial organisation are of particular interest (Kholodenko et al., 2000 ; Morishita et al., 2006 ; Kholodenko, 2003; 2006), as these are susceptible to show qualitatively different results from their deterministic counterparts. Arkin et al. (1998) carries out stochastic simulations of the phage λ lysis-lysogeny bifurcation pathway to characterise the statistical properties of the two distinct subpopulations. The switch (or bistable) mechanism is affected by the thermal fluctuations in the stochastic rate of rate-limitingreaction!rate-limiting reactions within gene expression giving rise to a distinct phenotypes. This phenomena is illustrated in Fig. 1.4 (extracted from Arkin et al. (1998)), a conventional approach fails to characterise the differences between both populations. Heterogeneous populations may be advantageous in evolutionary terms, as the rich phenotypic diver-

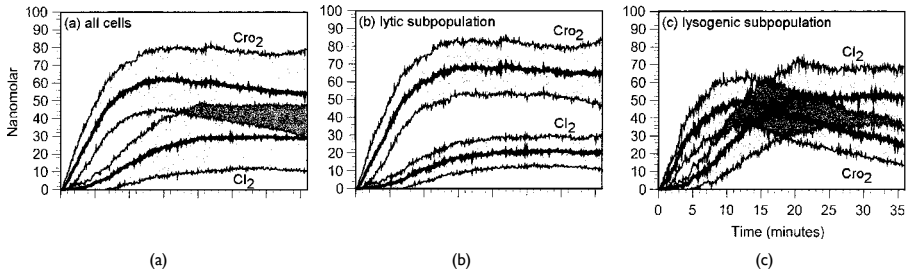

Figure 1.4: *The stochastic simulations for the phage λ lysys-lysogeny show clearly a phenotypic differentiation (b and c), which is masked by common experimental methods (a). Figures taken from Arkin et al. (1998).*

sity may provide some individuals with advantageous adaptability mechanisms against adverse, unforeseen environmental fluctuations. Acar et al. (2008) suggest that this diversity arises from the intrinsic stochasticity of gene expression in response to environmental pressures. Once a calm environment emerges, the pressure for phenotypic differentiation decreases, which implies a stable gene expression. In mechanistic terms, such phenotypic differentiation arises when certain events occur, or certain protein-concentration levels are reached. This happens in some cells before a differentiating process starts, such as a switch or cell division, while in other cells it does not. In these examples the gene expression lies at the core of the mechanism that shapes this stochastic behaviour. van Zon et al. (2006) showed that positive regulation of genes with a low number of regulators gives rise to a burst-type synthesis of mRNA and proteins. The synthesis of proteins in bursts may also be enhanced by the transcription and translation processes, as was shown by Dobrzynski and Bruggeman (2009). As a result, there can be large differences in the number of proteins expressed in each cell of an isogenic population (McAdams and Arkin, 1997).

Stochastic effects may also have a spatial dimension. Although the spatial effects are more relevant in larger cells, computer models of small prokaryotes predict spatial heterogeneity, mostly in the form of gradients (Francke et al., 2003). In spite of the small volume of bacterial cells (approx. $1\,\mu m^3$) Fange and Elf (2006) show that Min oscillations, which oscillate between the poles, help to locate the middle point before division. Moreover, a stochastic model reveals the emergence of small clusters on the membrane, whereas the deterministic shows a continuous concentration gradient. The stochastic model also allows to predict that oscillations may become aperiodic when the concentration of MinD and MinE are reduced below certain threshold levels. In addition, the emergence of MinD clusters on the membrane can only be explained by the stochastic kinetics of low copy numbers of molecules (Howard and Rutenberg, 2003). Similarly, (Doubrovinski and Howard, 2005) find that the stochastic model of Soj oscillations in the nucleoids of *Bascillus subtilis*, a remnant of *E. coli* Min oscillations, yields irregular oscillations as observed in experiments,

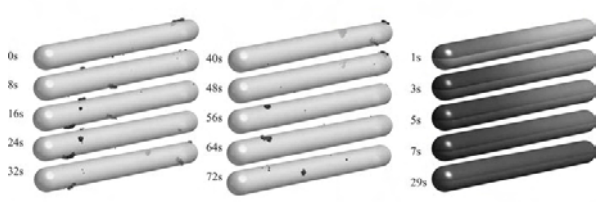

Figure 1.5: *Clusters of MinD appear* only *in the stochastic simulation (left and middle column), while a deterministic simulation shows a continuous gradient. Figure taken from Fange and Elf (2006).*

while a deterministic version of the model often stalls in a stable state. It is then hypothesised that the small stochastic fluctuations due to small numbers of particles are essential in some oscillatory systems as a means to escape from stable states. Spatial phenomena extend also to static systems such as the bacterial chemotaxis receptor, where the lateral interaction between individual sensory receptors is essential to amplify the signal (Shimizu et al., 2003 ; Shimizu, 2002)

Stochastic phenomena are however not always beneficial. Weak signals may be lost during their transport, circadian clocks corrupted (Barkai and Leibber, 2000), and memory reduction due to noise Acar et al. (2005). In fact, nature has also evolved network architectures, such as negative feedback loops, to provide resistance against disruptive stochastic events. In this way these systems may work well even at a low copy number of molecules (Gonze et al., 2002).

One thing all these approaches share is the use of stochastic computational methods to investigate the cause of apparently random phenomena at a molecular scale that was often beyond the capabilities of wet experiments. The validation of the computational results could not be done at a microscopic molecular interaction level, but often by observing a macroscopic phenomenon. Not until recent technological developments in microscopy, starting in the early 90s, was the direct visualisation of a single membrane molecule of single cell possible (Sako et al., 2000). Recent developments allowed for direct *in vivo* observations of the elusive events in gene expression, enabling for the tracking in real-time of the protein synthesis at the single molecule resolution (Cai et al., 2006 ; Yu et al., 2006), or the kinetics of binding and dissociation of transcription factors (Elf et al., 2007).

1.4 Computer Methods for Stochastic Chemical Reactions

Stochastic models are not a recent development, however their application to biology has undergone a recent acceleration (Chandrasekhar, 1943 ; jachimowski et al., 1964 ; McQuarrie et al., 1964 ; Ishida, 1964). In spite of the existence of stochastic mathematical frameworks such as stochastic differential equations

for dynamical systems, their analytical treatment is often unfeasible for complex systems (Gardiner, 1983 ; Baras and Mansour, 1996). In simple systems, however, the analytical solutions of dynamical systems often shed valuable insight on the fundamental processes of stochastic reactions. And it is precisely this dynamic component, the one that makes analytical treatment (McQuarrie et al., 1964). Paulsson (2004) uses, Ω-expansion theory (linear noise approximation) to analyse the steady state properties of the fluctuations (or noise) in gene networks.

Numerical or simulation methods represent a pragmatic alternative to study complex stochastic systems with single-molecule detail. These methods share two common features. The first is the discrete treatment of molecules, as opposed to a continuous concentration representation. (We do not refer to continuous time or space, but only to the state characterised by the number of molecules of each species.) The second is that a single simulation run produces only one stochastic trajectory of the system state. Hence, there is a need for an often large number of independent runs to obtain reliable statistical distribution of the dynamics. It is precisely this, computationally expensive, distribution which is the most valuable information provided by stochastic systems. With it, it is possible to characterise, for instance, multiple steady states, oscillatory behaviour or extinction of species not properly captured by a deterministic models (see section 1.3). Thus, computing explicit stochastic models often has a noticeable demand for computing power. One of the principal reasons is that every single reaction needs to be simulated explicitly. A wide range of computational methods and tools has been developed, and still continues to be, tailored to the specific needs of several types of systems (as roughly classified in section 1.2). In this thesis we emphasise the use of spatial and stochastic methods. These methods are a natural evolution of the homogeneous stochastic methods, just as PDEs are to ODEs. We should note that there are software packages that incorporate multiple techniques, deterministic, stochastic, spatial and homogeneous, in order to tackle complex multi-regime systems (Takahashi et al., 2004 ; Loew and Schaff, 2001).

Spatially homogeneous The Stochastic Simulation Method (SSA) (or direct method, or Gillespie's method Gillespie (1977)) stands as an exact solver of the underlying temporally homogenous birth-death Markov process (Gillespie, 1992a). The model describing this dynamic process is also known as the Chemical Master Equation (CME), which has been argued to be founded on microphysical foundations (Gillespie, 1992b). The discrete state of the system $X(t)_n$, with n different chemical species, is advanced one reaction at a time by calculating probabilistically the time at which the first reaction will occur in relation to the current state at a given time. Thus, simulation time advances in a variable manner and in each step one action (reaction) is executed, in contrast to other methods with fixed time step, such as StochSim (Morton-Firth and Bray, 1998). Gillespie's method also belongs to the *discrete event* simulators category. This reduces the possibility of introducing errors in the presence of events occurring in a shorter time than the time step. Additionally, a lack of a

prefixed time step simplifies the use and reduces the possibility of errors due to the difficulty in determining the number of stochastic reactive events that will be shorter than the time step. Another differentiating feature is that StochSim treats each particle as an individual software entity, in contrast to a population number of Gillespie's method. Storing each particle's state individually allows the tracking of its multiple state changes (such as phosphorylation or methylation) instead of the population of each possible state. As we shall see, Gillespie's approach is more widely adopted. Later enhancement introduced by Gibson and Bruck (2000) to the original SSA method focused on reducing the computational cost. This improved method reuses random numbers, reduces the overhead in the mechanism that computes the next reaction time by tracking the dependencies among reactions, and uses an absolute time instead of the relative approach used in the original SSA. Another further enhancement to the SSA allows for the execution of multiple reactions in a single step. The τ-leaping methods encompasses in one step the result of multiple reactions, thus the increase, or decrease, in the number of molecules might be larger than one (Gillespie and Petzold, 2003). This method reduces the computational cost at the expense of a loss in the SSA exactness status (the precision can be tuned by a parameter). The original τ-leaping method has also been reviewed and enhanced in order to further reduce errors (Rathinam et al., 2003 ; Chatterjee et al., 2005 ; Xu and Cai, 2008) All these methods assume that the reactions are not influenced by the diffusion of the particles. That is, the kinetics of the system can be well described by reaction-limited reaction rates, which are equivalent to the macroscopic reaction rates often used in classic ODE modelling.

Lattice-based Apatially Heterogeneous When reactions are diffusion-limited, or diffusion-influenced, the spatial dimension of the system needs to be incorporated. A common approach is to discretise space into small sub-volumes, small enough to regard it as a locally homogeneous sub-system. In this manner the position of particles needs not to be explicit: It is held by the lattice — which is mostly regular, but may also be unstructured Bernstein (2005) ; Engblom et al. (2009). The condition of local homogeneity is essential to allow existing homogeneous methods, mostly based on Gillespie's method, to be executed locally. Particle diffusion through space occurs by allowing particles to jump to neighbouring sites in a stochastic manner. Stundzia and Lumsden (1996) transform mesoscopic diffusion coefficients into corresponding transition probability rates in a similar way to Gillespie for reaction rates. These probabilistic rates are then used directly by Gillepie's methods as a sort of special reaction: one that changes the state of the neighbouring sites as particles jump to them. In the same line as the enhancements for homogenous methods, Elf et al. (2003) proposed recently the more efficient next-subvolume method, implemented in MesoRD tool (Hattne et al., 2005), which only updates those sites that change after each reaction. These two methods are exact solvers of the Reaction-Diffusion Master Equation (RDME), the spatial equivalent to the CME, on regular lattices. With the aim to further enhance the performance of this class of method Marquez-Lago and Burrage (2007) apply a particular case of the τ-leaping methods to substantially

reduce the computational time at the cost, again, of a minor degradation in accuracy. All these methods fall into the category of mesoscopic models since the solvent is ignored and the motion of particles is assumed to be normal diffusion. The principal distinguishing feature of RDME methods to CME methods is the inclusions of a diffusion mechanism for particles.

Prior to RDME methods, Cellular Automata (CA) were simple rule-driven spatial methods, discrete in space, time and state, used to primarily study pattern formation of physical and chemical systems, as opposed to intracellular (Ermentrout and Edelstein-Keshet, 1993). In the most basic form of CA, each particle occupied a lattice site, roughly equivalent in size (in principle RDME methods assume no particle size, although, as we shall see in section 3.1, there are some limitation in choosing the lattice size). The rules for the movement of particles developed initially as a combination of deterministic ballistic motion and direction randomisation upon collision (a physical model for gas molecules in lattice gas automata, LGA) giving rise to the mesoscopic diffusion phenomena from microdynamic interactions (Boon et al., 1996). Later on, a general acceptance of random walks on a lattice as a diffusion model for larger molecules suspended in a solvent (for instance proteins in the cytosol) would be the principal paradigm adopted (as we have seen above in RDME-based methods). Often the reactive rules in CAs were more qualitative than quantitative, and were tailored for a particular dynamical system (Weimar and Boon, 1994 ; Weimar, 1997). Moreover, the different diffusion coefficients of multi-chemical-species systems limited their use. Growing interest in the potential of CA to study fluctuating phenomena in multi-species biochemical systems resulted in the development of more rigourous kinetic rules to match the dynamics of the mean field approaches (Chopard et al., 1994 ; Weimar, 2002).

Lattice-free spatially heterogenoeous Diffusion in RDME and CA methods is a discrete-in-space mesoscopic process, albeit continuous and discrete in time. In its original conception, however, Brownian Dynamics (BD) simulations are a continuous process in space yet discrete in time process. In other words, it is lattice-free. Although there is consensus with the diffusive motion of particles, the reacting scheme has been implemented in various forms. GFRD (van Zon and ten Wolde, 2005) uses the Smoluchowski analytical solution for the reaction of two particles close to each other (Rice, 1985). The fact that particles do not need to collide allows the diffusion mechanism to use a larger time step (in the Smoldyn case collisions occur by using a larger *binding* radius, larger than the real size of the particle).

1.5 Signalling Systems in Bacteria

In this thesis we aim to shed light on the spatial and stochastic role in bacterial signaling systems. Bacteria are unicellular organisms (prokaryotic) that thrive in a broad range of environments, including very inhospitable sea bottom vents, the intestines and desktops in every office. They can be found almost

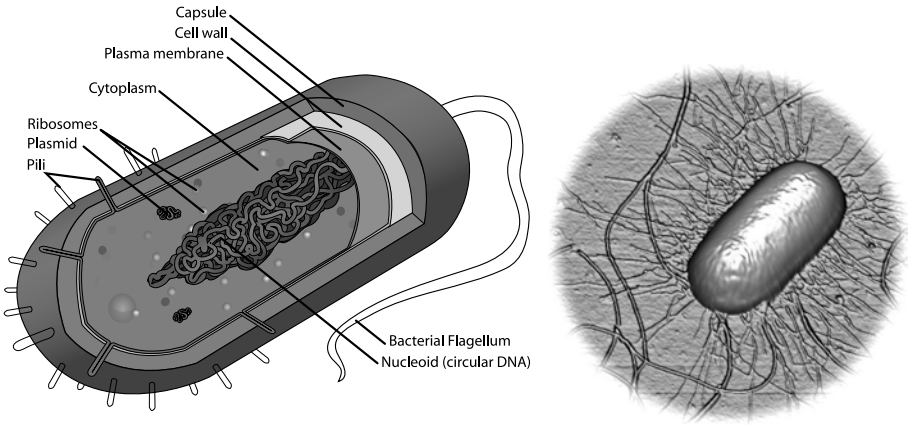

Figure 1.6: *In these illustrations we observe the rod-like shape of* E. coli *and the principal salient structures such as pili and flagella. (left) cartoon representation of the main structures. In this thesis we will pay attention to phenomena related to the flagella, cytoplasm, plasma membrane and nucleoid. (right) 3D image obtained with a scanning probe miscroscopy.* `http://oleeichhorn.com/images/nanoart/13_e_coli.jpg`

everywhere. Surrounding environmental conditions, a wide array of chemical substances and physical conditions, may change, or fluctuate. These changes are not always beneficial for the micro-organism, as they may pose a serious threat to its survival. It is important that they have evolved with quick, efficient and economical mechanisms to be positioned with a survival advantage (Stocker et al., 2008 ; Korobkova et al., 2004 ; Pernestig et al., 2003 ; Stephenson et al., 2000). When speaking of changes, or fluctuations, we should consider the temporal scale of these phenomena in relation to the temporal scale of the process subject to study. Signaling systems are specialised in detecting temporal changes; notwithstanding, these have evolved over evolutionary time scales to be adapted to these fluctuating conditions. For example, in *Escherichia coli* (*E. coli*, as illustrated in Fig. 1.6) chemotaxis has evolved an efficient sensory system to detect nutrients, and a response mechanism that enables it to move in order to reach for more nutrients (Kollmann et al., 2005). The sensor consists of an cluster of chemical sensors to the presence of nutrients. While the response mechanism consists of a number of flagella that rotate in order to propel and turn themselves. Thus, this mechanism can be viewed as an information flow system: capturing, transporting and processing signals.

To perform the sensory-response task, we identify in, a generic system, three main components that allow the information flow. Fig. 1.7 shows the specialised function of each component:

Figure 1.7: *Architecture and information flow in a sensory pathway. In some cases the sensor is also responsible for the transport of the signal to the receiver (dashed box).*

- A *sensor* specialised in detecting a particular signal such as a chemical, temperature or light. A sensor often detects the signal by a specialised input domain . For chemical signals this means that a molecule or chemical element binds the sensor's input. Upon input activation, the sensor activates an output domain, which can then pass the information down the path.
- The *receiver* of a signal understands the message and is the gate to activate other intracellular processes. It is possible that a receiver responds to more than one signal. Since this system is separated from the sensor, these two need a communication channel or transport.
- The *transport* of a signal can be done by another intermediate transport molecule, from the sensor to the receiver, or by the sensor molecule itself (dashed box in Fig. 1.7)(Ulrich et al., 2005 ; Hoch, 2000). It is also possible that the transport of the signal is effectuated not only by one molecule but by transferring the signal through a metabolic network.

The actual number of implementations of this architecture is by no means small. Sensors can take multiple forms and have different degrees of complexity. They can interact with one or multiple molecules. In other instances, the sensor and the transport can be the same molecule to detect internal signals. And the complexity of the transport chain may range from a single molecule composition to a cascade, or other more complex networks of molecules. A particular instance, central to this thesis, is the two-component signaling systems (TCS).

1.5.1 Two-Components Signalling Systems

Two-Components Signalling Systems (TCS) sensory systems were first characterised in the late 1980s by Nixon et al. (1986) and named *two-component signal regulatory systems* (see (Hoch, 2000 ; Rodrigue et al., 2000 ; Stock et al., 2000 ; Wolanin et al., 2002 ; Galperin, 2004 ; Cashin et al., 2006) for reviews, or (Simon et al., 2007 ; Inouye and Dutta, 2003) for monographs). Even though genomic analysis can reveal most of the genes encoding for the sensors and transporter molecules, the signal they respond to is not always known. Of the known conditions that TCSs regulate, nutrient acquisition, energy metabolism and virulence are just a few (Throup et al., 2000). Besides the omnipresence of

TCSs in bacteria (with few exceptions as seen Table 1.2). Evidences of TCSs in other organisms such as plant cells (Chang and Stewart, 1998) and single-cell eukaryotes (Chang and Stewart, 1998) have been reported too.

The first component of a TCS functions as a sensor and is usually a histidine-kinase; the second is a transporter or response regulator. Wide genomic studies on bacteria and archaea found that a large fraction (87%) of the two-component regulatory systems have as a target the modulation of gene expression (Ulrich et al., 2005). In this case, the response regulator is also called a transcription factor (or a regulator). Other receiver systems are: the motility system, composed of flagella (see Fig. 1.6), and metabolic networks. This simple two-step architecture requires only one or two types of molecules making the whole mechanism rather economical, in cell energy terms, to operate efficiently. This two-step architecture can be contrasted with a multi-step (or more commonly known as signaling) cascade, which is actually more common in larger and more complex eukaryotic cells. Despite this fact, two-component regulatory systems are also found in the more complex eukaryotic cells, although they are not the dominant type (Stock et al., 2000). Why eukaryotic cells have adopted a more efficient signalling architecture. Although this topic falls out of the scope if this thesis, some analysis is presented in a forthcoming publication (Dobrzynski et al., 2009).

Returning to bacteria, we should also mention that there is what is believed to be an even more primitive, extensive and economical sensory-response system where theses functions are carried out by only one single molecule. Following the same logic, these systems are know as one-component signalling systems. Contrary to one-component systems that are mostly used for intracellular sensing, two-component systems are the dominant type for extra-cellular sensing in prokaryotic cells (Ulrich et al., 2005). The strategic anchoring of sensors across the inner membrane, with access to both the exterior (the cell wall is permeable) and the cytosol, provides the right spatial placement for such signalling.

Spatial Localisation of Molecules is Relevant

The spatial localisation of sensors and receivers can play an important role for the signalling functioning (Hoch, 2000, and references therein). For example, in chemotaxis sensors form a cluster in order to co-operatively increase signal sensitivity. A consequence of confining sensors to one pole of *E. coli* is the rise of a spatial gradient of the response regulator across the cytosol (Lipkow et al., 2005). The spatial gradient contains, in a way, a memory of where the signal originated, that is, where the concentration is higher. If there was no gradient, then without knowledge of the position of origin of the sensor we could not infer the origin of the signal. For these gradients to arise the diffusion transport needs to be relatively slow relative to the production rate of the particles. However, the transporter molecule can lose the signal before it has reached the receptor as a result, for instance, of auto-dephosphorylation or cleavaging. In this situation, more particles are needed in order to guarantee the delivery of the signal, avoiding false negatives (Morishita et al., 2006).

Table 1.2: *Number of potential two-component systems in different organisms of different kingdoms.*

Kingdom	Organism	Approx. number of 2CSTN
Bacteria	*Escherichia coli*	30
	Mycoplasma genitalium	0
	Synechocystis sp	80
	Bacillus subtilis	70
	Haemophilus influenzae	9
	Helicobacter pylori	11
	Thermotoga maritima	19
Fungi	*Saccharomyces cerevisiae*	1
	Candida albicans	2
Animalia	*Caenorhabditis elegans*	0

Additionally, the organization of these sensors on the two-dimensional inner-membrane can be exploited in order to increase or decrease the signal response efficiency. On the one hand, Kentner and Sourjik (2006) have shown that a high degree of clustering of chemotactic sensors have a signal amplification effect (increasing the sensor sensitivity)i.e. the activation of one sensor facilitates the activation of adjacent sensors. On the other hand, il Lim and Yin (2005) proposed that scattering receptors over the membrane, even into small clusters or rafts, might have a beneficial effect on reducing the signalling time.

Another type of spatial organization, which is usually not incorporated in the spatial stochastic methods presented earlier, concerns the genome. Genes in bacteria are organised in what is believed to be functional groups or operons. Thus, multiple genes are contained in one operon, or transcription of an operon results in a polycistronic mRNA. This ensures that expression levels in these gene families are correlated. This is viewed as an economical yet robust mechanism to guarantee the proper functioning of systems (Iber, 2006).

1.6 Objectives of this Thesis

Interest in how living organisms benefit from the stochastic processes occurring at the cellular level has been developing rapidly. Because the origins of stochasticity are well understood from a mechanistic point of view. Advances have come either in the form of computational simulations of complex biochemical processes or as analytical derivations usually of more elementary processes. Recently spatial effects, such as localisation of reactions on the membrane walls or on the DNA, together with their associated stochastic effects challenge current computational tools and open new questions about the implications of the nature random events—stochasticity.

From the computational perspective we investigate how different simulation methods cope with some small networks in which stochastic and spatial effects are treated explicitly. Concretely,

- how an operator-split reaction-diffusion mechanism affects noise levels and the distribution times of reaction between molecules.

To this end we developed a general simulation method and contrasted the results against other methods with similar capabilities.

This computational framework provides us with a base to analyse in more detail the mechanisms involved in two-component signalling transduction pathways in bacteria. In this work we question *whether the genome has evolved in a particular order providing an evolutive advantage to organisms, and in particular to the environmental sensory systems in bacteria.* Because of their relationship, a generic signalling system is studied alongside with its associated gene expression system, responsible of the protein synthesis that enables the signalling function. The objective is to elucidate

- the connection between the spatial localisation of reactions and the stochastic effects in gene expression for an effective signalling.

Fulfilling this objective requires the use and analysis of bioinformatics data on two-component signalling-related genes, together with analytical work of the fundamental processes.

1.7 Overview of this Thesis

In this thesis we aim to understand the influences of spatial localisation of reactions under stochastic conditions. In Chapter 2 we introduce a method used for simulation of reaction-diffusion chemical processes. Our method uses a combination of multi-particle diffusion developed by Chopard et al. (1994). This approach is taken so as to reduce the number of random events, as it allows for multiple particles to jump to a neigbouring site. For reactions we use the Gillespie mechanism, much in line with other RDME methods as described in section 1.4. This combination implies the method is only approximate, yet, we shall see that the decline in accuracy is not significant. With the aim to keep the two processes independent, we study some computational enhancements that reduce the diffusion costs in the regimes where the number of particles per site is low. The comparison of our method against other available methods is discussed in Chapter 3. We use the models of biochemical networks available in the literature to compare the quality of the solutions. First, we compare the fluctuations that arise in gene expression as a result of diffusion-limited response regulators as done by van Zon et al. (2006). We find that our method performs similarly to an exact solver of the RDME, and that with a carefully chosen lattice discretisation the fluctuations levels produced by GFRD can be achieved. Surprisingly, Smoldyn produces significantly lower levels of fluctuations. Second, we contrast the results of the PTS system against a partial differential model by Francke et al.

(2002) ; Blom and Peletier (2000), which yields shallow gradients due to the membrane-localised reactions. The last case is also related to spatial gradients in the cytosol, but this time in the chemotaxis system and compared to Smoldyn results provided by Lipkow et al. (2005). The last two chapters are concerned with two-component signalling systems in a generic form. In chapter 4 we use simulations to analyse the response time as a a function of time between reactive events. We explore in detail the consequences that of the spatial organisations of sensors on the membrane and the ratio between sensors and response regulators have on the response time under restricted conditions in the total number molecules involved. These analyses required detailed spatial discretisation similar to those provided by Brownian dynamics. We used, however, a random walk on a lattice methodology and a first-passage approach for the kinetics. The results obtained suggested that for an efficient signalling response, the expression of both sensors and response regulators needed to be strongly correlated and in ratios close to 1:1. In Chapter 5 we study the genomic organisation of two-component signalling genes in bacterial organisms. We discuss the advantages of operon organisation, and why in other instances TCS genes are in separate operons. The analysis considers known types of random effects occurring during transcription and translation, which seemingly have been overlooked in many recent studies, but which, with the recent technological developments of *in vivo* single molecule visualisation, may contribute to furthering our understanding of stochastic cellular processes.