Stochasticity in signal transduction pathways

Vidal Rodriguez, J.

Publication date
2009

Chapter 3

Noise and Spatial Comparisons

In the previous chapter we have introduced the GMP method to simulate chemical reactions discretely, stochastically and spatially. We saw in the ubiquitous reversible reaction that the discretisation of space and the use of the operator split method affects the quality of the results in the short time scale. Other comparable methods, such as those described in section 1.4, model the reaction and diffusion processes in a similar manner and thus we expect them to display similar properties. Nonetheless, since the modelling for reaction and diffusion differs slightly among methods, it is important to understand the effects of each particular model onto these two fundamental aspects of biological reactions.

In this chapter we compare the results of several methods applied to more biological significant systems in order to elucidate the effects introduced by the different classes of methods. First we extend the analysis of the reversible reaction and place it in the context of gene expression regulation. Gene expression has been studied in numerous ways due to its central role in biology and stochastic effects. Second, we study the localisation of reactions on the surface of the inner-membrane for two networks: the PTS uptake pathway and the chemotaxis signalling network.

The contents of this chapter are based on the results published in

- Rodriguez, J. V., Kaandorp, J. A., Dobrzynski, M., and Blom, J. G. (2006). Spatial stochastic modelling of the phosphoenolpyruvate-dependent phosphotransferase (pts) pathway in escherichia coli. *Bioinformatics*, 22(15):1895–1901.

- Dobrzynski, M., Rodriguez, J., Kaandorp, J., and Blom, J. (2007). Computational methods for diffusion-influenced biochemical reactions. *Bioinformatics*

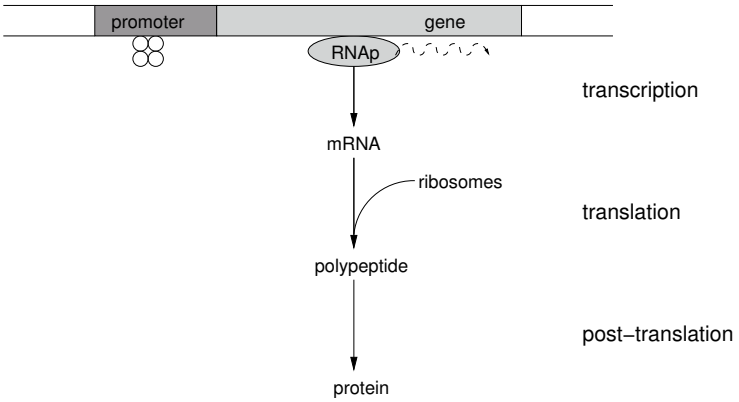


Figure 3.1: *Simple model of prokaryotic gene expression. A transcription factor is usually bound to the promoter site of a gene. An RNA polymerase can initiate transcription of the gene when the promoter is unoccupied. Then a single strand of mRNA is translated by multiple ribosomes producing multiple proteins before mRNA degrades. Proteins also degrade.*

3.1 Noise in Gene Expression

Gene expression is the mechanism by which new proteins are synthesized from the DNA. A widely used model for gene expression in prokaryotes is illustrated in Fig. 3.1. The synthesis of proteins is a multistep process: first is the *gene transcription* of genes into mRNA strands. The second is the *translation* of mRNA into a polypeptide chain that then folds into a minimum energy and stable configuration: the protein. The later translation step has been found to be the one with largest contribution to intrinsic noise, leading to population variation from isogenic populations (Ozbudak et al., 2002). In this line, Krishna et al. (2005) suggested that a more selective selection of subpopulations is caused by the intrisinc stochasticity, due to only small number of molecules, together with cooperative switches. Becskei et al. (2005) argues that the cell-to-cell variation is mainly due to the rare events of gene transcription rather than the low copy number of molecules.

It has also been observed that gene expression, under certain circumstances, is a bursty process (Ozbudak et al., 2002 ; Yu et al., 2006). If transcription initiation is rare, the synthesis of mRNA results in its translation by multiple ribosomes producing a burst of new proteins. When the transcription initiation occurs approximately one or two times per cell cycle, the number of proteins can differ significantly between cells, giving rise to a potential separation of populations.

It is common for prokaryote cells to negatively regulate gene expression, this is, a repressor is usually bound to the promoter site preventing the RNA polymerase molecule to initiate transcription (Hartl and Jones, 2005, chp. 11). Gene

transcription has been extensively studied for diffusion-limited effects with low copy number of transcription factors (< 50nM) (van Zon et al., 2006). Because of fast rebinding of the regulator (a transcription regulator or transcription factor) compared to that of RNA polymerase, the dissociation rate decreases inducing higher values of noise for transcription. Diffusion-limitation of the regulator gives rise to non-exponential distribution of activation, which produces higher levels of noise (van Zon et al., 2006 ; Dobrzynski et al., 2007).

3.1.1 The Model

The generic model for gene expression, as analysed in van Zon and ten Rein (2005), enables us to study noise caused by the combined effects of a low number of molecules and diffusion-limited reactions. Van Zon and ten Rein provide detailed continuous-in-space simulations of this system which we use to validate our GMP method. Additionally we use these accurate results to check our reasoning for choosing a proper lattice size.

The system we look at is a closed volume V of $\frac{4}{3}\pi\,\mu m^3$ with a DNA promoter binding site fixed in the centre and surrounded by freely diffusing RNA polymerase molecules (RNAp). Once the DNA-RNAp complex is formed with association rate k_a, it can either dissociate back to separate DNA and polymerase (with rate k_d) or produce a protein P with a production rate k_{prod} with subsequent complex dissociation. A protein can further decay at the rate k_{dec}. Obviously the single protein production step in this model encompasses both transcription and translation which in fact consist of many biochemical reactions. Protein degradation is also simplified and treated as a first-order decay reaction. The chemical reaction model and its parameters are shown in Table 3.1 and the initial conditions in Table 3.2.

Table 3.1: *Reaction scheme and parameters associated with the gene expression model*

Reaction			Rate
DNA + RNAp	$\xrightarrow{k_a}$	DNA·RNAp	$3 \cdot 10^9\,\mathrm{M}^{-1}s^{-1}$
DNA + RNAp	$\xleftarrow{k_d}$	DNA·RNAp	$21.5\,\mathrm{s}^{-1}$
DNA·RNAp	$\xrightarrow{k_{prod}}$	P+DNA+RNAp	$89.5\,\mathrm{s}^{-1}$
P	$\xrightarrow{k_{dec}}$	$\emptyset$	$0.04\,\mathrm{s}^{-1}$

When choosing our lattice size we use the reasoning developed in section 2.2.4. In van Zon et al. (2006) particles have a diameter of 5 nm (approximating the spherical size of an average real complex-shaped molecule). We therefore obtain for the minimum λ a value of $\lambda_o = 10\,\mathrm{nm}$ with $L = 100$. The λ_{rmfp} corresponds with $L \approx 18$.

Table 3.2: *Intial concentrantions and number of molecules for the gene expression model*

	Concentration	Molecules	Dif. coef.
DNA		1	0
RNAp	30 nM	18	$1\,\mu\mathrm{m}^2\mathrm{s}^{-1}$
P		0	∞

3.1.2 Comparison of Noise

The parameters we use in the simulations are such that the protein synthesis process is limited by the diffusion of RNA polymerase molecules and that new proteins appear in bursts. We anticipate that stochastic effects will be significant under such circumstances. This idea is supported in Fig. 3.2, where the protein level behaviour in time is shown, as well as the distribution. While the averages remain the same, fluctuations levels for methods that explicitly account for space (i.e. GFRD and Smoldyn) are higher than those present in the widely used Gillespie's (or SSA) while the averages are the same (see standard deviation values in Table 3.3). The reader will also notice that GFRD yields significantly larger fluctuations than Smoldyn—in principle a method at the same level of detail. The reason for the differences between the two Brownian Dynamics (BD) methods are discussed later. For now it is sufficient to say that GFRD produces more trustworthy results since it is an exact method to solve diffusion-limited reversible reactions with only one reactive target.

Table 3.3: *Mean and standard deviation of the number of proteins in steady state for the three reference methods.*

	GFRD	Smoldyn	SSA
μ	1081	986	999
σ	147	61	28

The comparison of the RDME-class methods, MesoRD and Gillespie Multiparticle, reveals the behaviour for an increasing number of space divisions n_{sv} (the number of sub-volumes per unit length). We know that for $n_{sv} = 1$, the well-mixed case, RDME methods are equivalent to the CME which does not include space. In Fig. 3.3 we observe that, as the spatial detail increases, the predicted fluctuations reach the value obtained with the GFRD method, while the average number of proteins is the approximately the same (1032 ± 64), within statistical error, for all lattice sizes (see specific values for each method in Table 3.3.)

The fact that MesoRD and GMP are able to reproduce not only the average but also the correct variance of the solution, as compared to the reference result obtained from GFRD, shows their capability to give good results in the regime where spatial effects are important (the diffusion-limited regime with

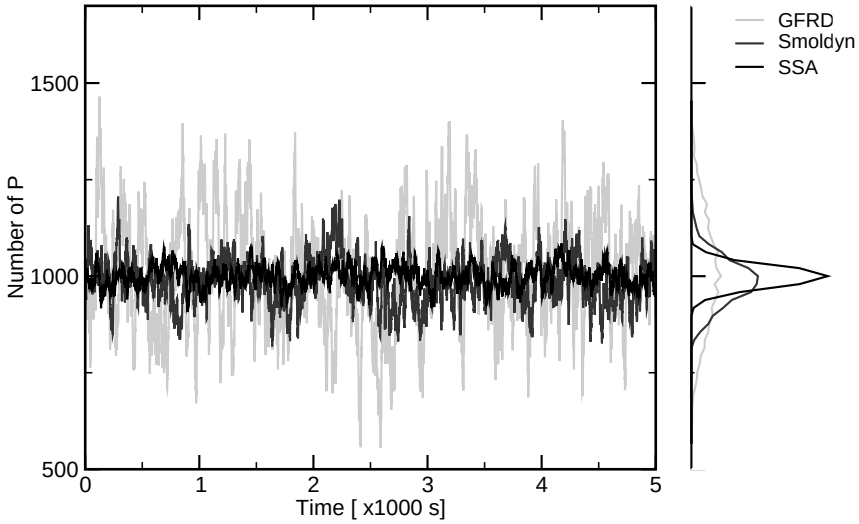


Figure 3.2: *Comparison of fluctuations in the number of proteins in steady state for the methods from literature (left panel shows the dynamics, and the right panel its distribution.) GFRD is the reference solution, since it models the diffusion-limited effects in an exact manner. On the other hand, SSA produces the smallest fluctuations because it is based on a well-mixed model. Smoldyn produces, unexpectedly, results in between, in spite of being a Brownian Dynamics method close to GFRD.*

small numbers of molecules).

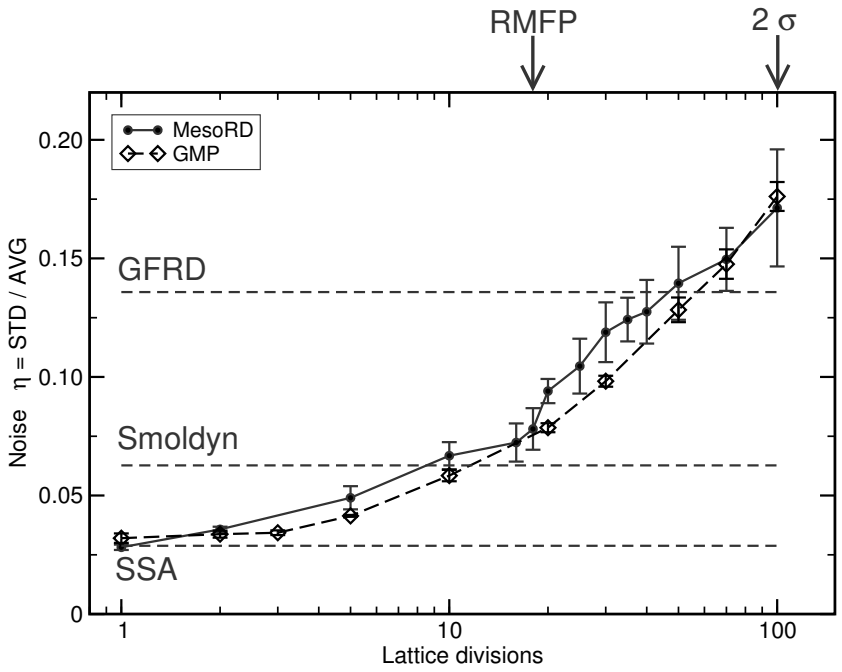


Figure 3.3: *Comparison of the noise levels for the different methods used to simulate gene expression. SSA is the standard Gillespie method result. Smoldyn and GFRD are two lattice-free methods. The bars indicate the standard error of the mean. RMFP is the reaction mean free path as defined in section 2.2.4. And the σ is the radius of the particle.*

3.1.3 Reversible Pair

In order to explain the differences in noise level predicted by various methods (Fig. 3.3) we look at a simple example of an isolated pair of molecules undergoing a reversible reaction, the same type of reaction as the first step in the gene expression model. This will allow us to examine the distribution of times between successive reactive events.

The target molecule is fixed in the center of the spherical domain, the second molecule diffuses freely, as in normal diffusion, with diffusion coefficient D. The molecules can undergo a reversible reaction with association and dissociation rates, κ_a and κ_d, respectively. We look at the time between consecutive bindings of the molecules. Simulations with GFRD, Smoldyn, GMP and SSA reveal that the distribution of inter-binding times is different for methods with and without space, and also that methods with spatial detail treat the diffusion-limited reactions differently.

It is known that for a particle diffusing in an infinite three-dimensional space the probability that it reaches a specific target at a specified time (the first-

passage probability) has a power-law distribution (Redner, 2001). This is depicted in the log-log plot in Fig. 3.4. All spatial methods reproduce the power-law behavior (straight line) for times shorter than the average time needed for a single molecule to reach the boundary; a classical result for diffusion in an infinite 3D space (Weiss, 1994 ; Redner, 2001). On such a short time scale the molecule is not influenced by the finite boundary because it simply did not have time to travel the distance to reach it. On the other hand, diffusion in a closed finite volume results in an exponential decay of the first-passage probability for long times. Exponential behavior is characteristic to mean-field approaches like the chemical master equation, where the time of a next reaction is independent of the molecules' position (the well-mixed assumption). Note that, the exponential part of the result obtained with GFRD can be reproduced by changing the forward reaction rate in SSA from the intrinsic rate κ_a to the overall *on* rate coefficient $k_{\mathtt{ON}}$ (Agmon and Szabo, 1990 ; van Zon and ten Rein, 2005), which "includes" the time needed to reach the target by diffusion and the time to undergo a chemical reaction.[‡] It is important to note that the average of the first-passage distribution SSA($k_{\mathtt{ON}}$) in Fig. 3.4 differs from the rest of the experiments. This fact indicates that the simple change of the reaction rates from the intrinsic to the *overall* rates which include the effects of diffusion will not preserve the average time between bindings. The plot also shows that an increase of spatial resolution of an RDME-class method like GMP results in the distribution which can be as close as desired to the exact solution given by GFRD. For clarity we draw only intermediate distributions with small (5) and average (30) number of sub-volumes per unit volume. The simulation with $n_{sv} = 50$ overlaps with the result from GFRD.

[‡]The overall *on* reaction rate equals $\frac{1}{k_{\mathtt{ON}}} = \frac{1}{\kappa_a} + \frac{1}{K_D}$, where κ_a is the intrinsic association rate, and K_D is the diffusion-limited reaction rate given by $4\pi\sigma D$, dependent on the reaction distance σ and the relative diffusion coefficient D of two reacting molecules. Note that inverses of rates are equivalent to quantities with a dimension of time.

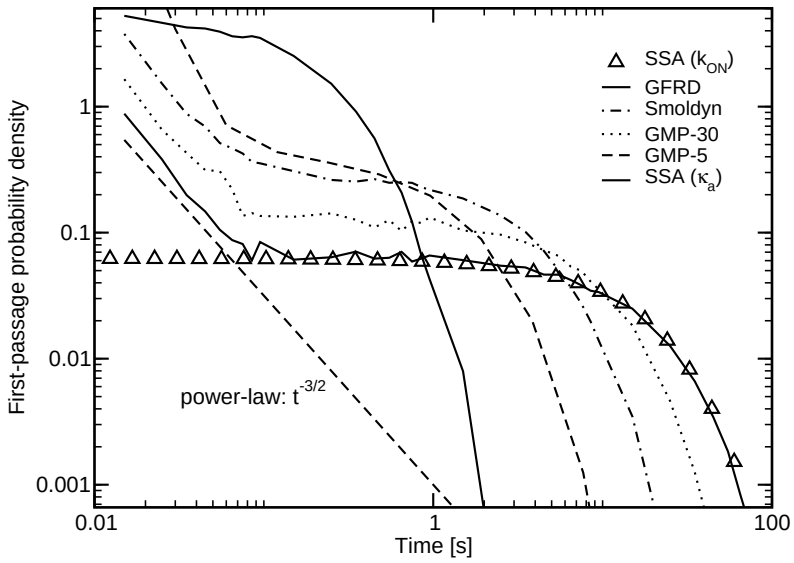


Figure 3.4: *Probability density function of time between subsequent bindings of an isolated pair of particles, also known as the first-passage probability. Methods including spatial effects, GFRD, Smoldyn and GMP, reproduce the power-law behavior for short times. For times larger than* $\approx 0.1s$*, which is the average time needed to reach the boundary, the first-passage probability exhibits an exponential decay. GMP-5 and GMP-30 denote simulations where the whole volume's side is divided into 5 and 30 sub-volumes. The left- and rightmost curves are computed with SSA with the forward reaction rate equal to the* intrinsic κ_a *and to the* overall k_{ON} *reaction rate, respectively. The averages of all distributions are the same, equal* $1/\kappa_a$*, except for the* $\mathtt{SSA}(k_{ON})$*, where the average is larger and amounts to* $1/k_{ON}$*. Note, that the position of the power-law line is chosen arbitrarily; it only compares the slope of data for short times obtained from different methods.*

3.1.4 Discussion

A theoretical study of gene expression shows that fluctuations arise in diffusion-limited processes not only due to the small number of reactants but also may result from spatial effects (van Zon et al., 2006). In the comparison we show that not all computational methods predict fluctuations correctly, although the average is the same as those given by the mean-field models (CME, ODE, PDE). Smoldyn yields much smaller fluctuations comparing to GFRD, even though both methods are based on Brownian dynamics. The reason for an incorrect prediction of the second moment by Smoldyn lies in the way it deals with diffusion-limited reversible reactions. The assumption that every collision is reactive leads to the introduction of the unbinding radius such that it reconstructs the macroscopic geminate recombination probability. By doing so a part of the spatial fluctuations is averaged and the resulting first-passage probability lies between the exact solution of the Smoluchowski diffusion equation obtained with GFRD and the mean-field result from SSA for the well-mixed system (Fig. 3.3).

The methods based on the reaction-diffusion master equation, MesoRD and GMP, are able to correctly predict the fluctuations. The key issue here is to choose such a space discretisation (division into sub-volumes) that the size of a single sub-volume is at least of the order of the systems correlation length. Otherwise the assumption about the local independence of the reaction probability from the inter-particle distance does not hold. If the requirement of well-mixed sub-volumes is not satisfied, spatial fluctuations are averaged which is clearly visible in Fig. 3.2. For a small number of volume sub-divisions both the noise in the protein level and the distribution of times between bindings approach the prediction of the SSA model. On the other hand, if the number of sub-volumes is increased up to the limit where two molecules completely fill one sub-volume, the first-passage probability gradually recovers the desired characteristics typical for the diffusion process in a closed volume: power-law behavior for short times and exponential decay for long times (Fig. 3.2).

A word of caution about the notion of exact fluctuations needs to be added here. Although we treat noise computed with GFRD as a reference value, one should bear in mind that this is a result of fluctuations with a rather simple Brownian dynamics model for chemical reactions. We are ignoring here other, possibly important, microscopic effects, like hydrodynamic interactions, electrostatic forces or molecular crowding. These certainly affect the diffusion process (Echevera et al., 2007) but their significance for enhancing noise in biological systems is an open issue. Still, methods like GFRD which solve numerically the Smoluchowski model for diffusion-limited chemical reactions will provide an upper bound for the magnitude of fluctuations if compared to mesoscopic methods based on the master equation. The latter contains additional physical assumptions in order to simplify computations at the cost of averaging microscopic phenomena. In section 3.1.3 we argue that for a reversible reaction of a pair of particles the methods reproduce the first-passage probability differently, which is the cause for the variation in noise. The power-law region constricts also with increasing number of molecules or with accelerating the diffusion process (not

shown in this study). Then the system abandons the low-molecule-number and the *short-correlation-length* regime and the distribution converges towards the mean-field exponential behaviour. This can be properly approximated either by RDME-based methods with a coarse discretisation or simply by the SSA algorithm. If the overall forward reaction rate κ_{ON} is taken instead of the intrinsic κ_a, the SSA is also able to reproduce the exponential decay of the first-passage probability. This case equals to the one obtained from GFRD or RDME methods with the large number of sub-volumes. However, for obvious reasons, the power-law part is not recovered (in SSA the next event is drawn from the Poisson distribution), and thus the average inter-binding time is larger than that of GFRD.

3.2 PTS in *Escherichia coli*

In the previous section we saw the effects that diffusion-limited reactions and a low copy number of molecules have on the noise levels of proteins. It is also instructive to study the performance of the GMP method, and by extension of the RDME-based methods, applied on a problem that belongs to another regime. In this section we shift our study to a system in a regime where the number of particles is moderately large (in the order of thousands) and therefore a deterministic modelling approach seems more suitable (recall regimes Fig. 1.3). In this regime the principal point of interest is the spatial properties rather than the fluctuations. Therefore, results of the stochastic simulations can be compared with results obtained from a partial-differential-equation (PDE) model of the same process.

3.2.1 The Model

The phosphoenolpyruvate:glucose phosphotransferase system (or PTS), as illustrated in Fig. 3.5, is a moderate complex network, involving 17 species with fast and slow reactions, and of particular interest it has reactions between cytosolic and membrane-bound molecules. The PTS is a group-transfer pathway situated at the beginning of the glycolysis pathway (Stryer, 1987). It is present in many bacteria, such as our model organism *E. coli*, and it is involved in the uptake and concomitant phosphorylation of external glucose that feeds the glycolysis. A phosphoryl group derived from PEP is carried along a series of bimolecular reactions to the membrane-bound IICB, which is responsible for taking the external glucose and transferring it into the cytoplasm. Table 3.4 summarizes the complete list of reactions of the model as shown in Fig. 3.5 and as analysed by Francke et al. (2003). In that paper the kinetic parameters were obtained from Rohwer et al. (2000) and the diffusion coefficients were based on measurements by Elowitz et al. (1999).

Cytosolic concentrations have been converted into numbers of molecules using 1/8th of a sphere's volume of 1 μm radius, discretised in a lattice of $L = 20$. All molecules are initially uniformly distributed in their respective compartment, be it the cytosolic or the membrane sites as described in section 2.2. The membrane

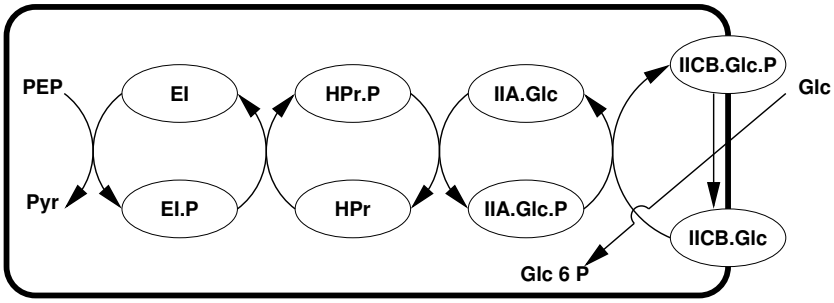


Figure 3.5: *Model of PTS in* Escherichia coli. *A phosphoryl group derived from PEP is carried along a series of bimolecular reactions to the membrane-bound IICBwhich is responsible for the uptake of glucose (Glc).*

is a crowded environment and more viscous than the cytosol. Consequently, diffusion coefficients are usually one order of magnitude slower than for molecules of similar size but free to diffuse in the cytosol. Thus, the mobility of membrane molecules has little impact on the global dynamics of the system. Additionally, due to the irregular neighbourhood of the membrane sites on the lattice, which would add additional computational operations in the diffusion step, membrane-bound particles remain immobile.

We argue that the much slower diffusion of membrane-bound molecules should not have a significant impact on the overall system's dynamics (Kholodenko et al., 2000).

Surface concentrations of membrane-bound molecules require special consideration. We assume that the concentration reported in Rohwer et al. (2000) for IICB was calculated as if the membrane-bound molecules were diluted in the cytosol and its concentration was measured to be $15\,\mu mM$. According to this, the number of IICB molecules should be 3100. Since the procedure of fitting IICB bulk concentration in the original model by Rohwer does not explicitly indicate the number of membrane molecules, we have adjusted the number of IICB molecules in such a way as to approximate the steady state of IIA·P·IICB and IICB·P· Glc. Additionally, since we have to compare surface concentrations and numbers of molecules, we scale the surface concentrations of the PDE solution in such a way that the initial IICB concentration matches the number of particles, as is used in Fig. 3.6(c). By trying different numbers, we found that using 2100 particles for IICB, corresponding to an average of 4.3 particles per site, yields a good approximation for IIA · P · IICB and IICB · P · Glc but a less accurate approximation of IICB and IICB · P. Finally, external Glc molecules are confined to membrane sites in the reported concentration without interacting with the cytosolic molecules.

For the PTS system at steady state the average value for λ_{rmfp} for cytosolic reactions is 0.06 ± 0.05. Since all species must share the same lattice, the optimal lattice size L is different for each species, making even more difficult the selection

Table 3.4: *Reaction list and parameters associated to the PTS model illustrated below. These parameters were taken from the spatial PDE model discussed in Francke et al. (2003), converting concentrations to number of molecules. Special consideration is put on the conversion of the membrane-bound concentration*. This parameter is adjusted to the value 2100 to obtain a better approximation of the PDE solution and is used for the results reported here.*
† *Bimolecular reaction rates are given in* $\mu M^{-1} min^{-1}$*, and monomolecular in* min^{-1}*.*
♯ *Diffusion coefficients are given in* $\mu m^2 min^{-1}$*.*
‡ *Initial concentrations for cytosolic species are given in* μM*, the surface concentration for IICB is given in* $\mu M \mu m$

Reaction			Rate†
EI + PEP	$\overset{k_1}{\rightarrow}$	EI·PEP	1960
EI + PEP	$\overset{k_{-1}}{\leftarrow}$	EI·PEP	480000
EI·PEP	$\overset{k_2}{\rightarrow}$	EI·P + Pyr	108000
EI·PEP	$\overset{k_{-2}}{\leftarrow}$	EI·P + Pyr	294
HPr + EI·P	$\overset{k_3}{\rightarrow}$	HPr· P·EI	14000
HPr + EI·P	$\overset{k_{-3}}{\leftarrow}$	HPr· P·EI	14000
HPr· P·EI	$\overset{k_4}{\rightarrow}$	HPr·P + EI	84000
HPr· P·EI	$\overset{k_{-4}}{\leftarrow}$	HPr·P + EI	3360
IIA + HPr·P	$\overset{k_5}{\rightarrow}$	IIA·P·HPr	21960
IIA + HPr·P	$\overset{k_{-5}}{\leftarrow}$	IIA·P·HPr	21960
IIA·P·HPr	$\overset{k_6}{\rightarrow}$	IIA·P + HPr	4392
IIA·P·HPr	$\overset{k_{-6}}{\leftarrow}$	IIA·P + HPr	3384
IICB + IIA·P	$\overset{k_7}{\rightarrow}$	IICB·P·IIA	880
IICB + IIA·P	$\overset{k_{-7}}{\leftarrow}$	IICB·P·IIA	880
IICB·P·IIA	$\overset{k_8}{\rightarrow}$	IICB·P + IIA	2640
IICB·P·IIA	$\overset{k_{-8}}{\leftarrow}$	IICB·P + IIA	960
Glc + IICB·P	$\overset{k_9}{\rightarrow}$	Glc·P·IICB	260
Glc + IICB·P	$\overset{k_{-9}}{\leftarrow}$	Glc·P·IICB	389
Glc·P·IICB	$\overset{k_{10}}{\rightarrow}$	Glc·P + IICB	4800
Glc·P·IICB	$\overset{k_{-10}}{\leftarrow}$	Glc·P + IICB	$5.4 \cdot 10^{-3}$

Species	Initial Conc. ‡	Molecules
EI	5	1577
HPr	50	15766
IIA	40	12613
IICB	3	3100*
PEP	2800	882890
Pyr	900	283780
Glc	500	70767
Glc·P	50	15766

Species	Dif. coef.♯
EI, EI·P, EI·P·Pyr	197.8
EI·P·HPr	189.1
HPr, HPr·P	378.0
HPr·P·IIA	262.1
IIA, IIA·P	300
PEP, Pyr, Glc	18000
all IICB	0

of the most adequate lattice discretisation to minimise artefacts. Thus we are bound to introduce some error in the species whose optimal L differs from the chosen. In this case, and based upon previous tests, we have chosen a lattice size of $L = 20$, which yields reasonable average results. Since the discretisation is most likely to affect noise levels rather than average values (tests not shown), in this section we do not consider whether the noise levels are accurate or not.

3.2.2 Results

A first check of the validity of our method consists of looking at the System's Mass Evolution in Time (SMET) which shows the total number of particles present in the system over time regardless of their position. In Fig. 3.6 the SMET for all species shows a good agreement with the reference solution, the PDE model as described in Blom and Peletier (2000). These results are a direct consequence of the initial decision to adjust the number of molecules of IICB to

match the IIA·P·IICB steady state. Nevertheless we note in Fig. 3.6(a) a very good agreement of the EI-related molecules. In Fig. 3.6(b), despite a good initial agreement ($t < 0.0005\,\text{min}$), the total number of molecules starts to diverge, and settling at levels around 6% for HPr species, and 20% for IIA·P. Finally, the level of the phosphorylated IICB·P, is 4 times as high as the PDE solution. However, in absolute numbers, the extra ≈100 molecules in IICB·P, seem to be lacking in IICB.

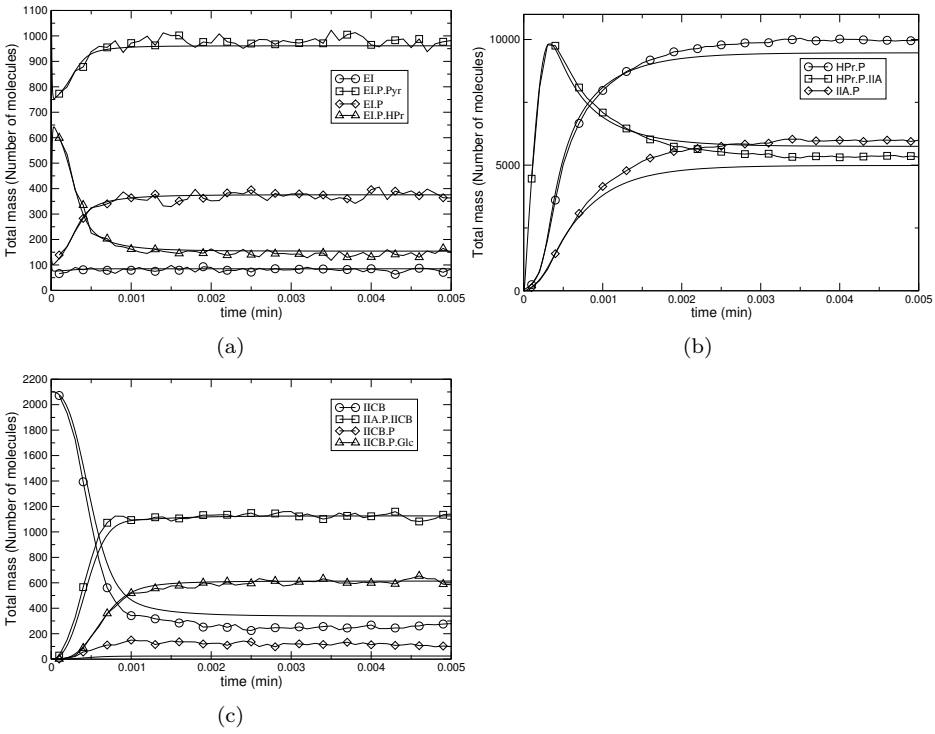
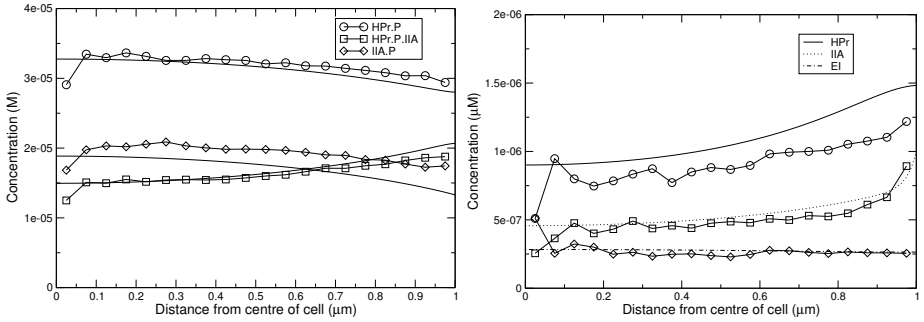


Figure 3.6: *System Mass Evolution in Time (SMET) graph for the PTS pathway with parameters shown in Table 3.4 and a cell's radius of* $1\,\mu m$*. Only one trial is shown here for the GMP result. EI-related enzymes show a very good agreement between the GMP result and the PDE solution. The membrane concentration of IICB enzyme in subfigure (c) was scaled to match the number of molecules at* $t = 10^{-5}$ *s.*

In Fig. 3.7 we show the radial gradients, at steady state ($t_{sim} = 0.005\,\text{min}$) for some enzymes. To obtain the gradient along the radius we average the number of particles over all sites with the center in the range $[r - 0.5/L, r + 0.5/L)$. We then average again over 50 trials of the same systems with uniform random distributed particles. Therefore, results near the membrane will always contain far more points than those near the center, so the average near the center may show a larger deviation from the real mean. In Fig. 3.7(a), we observe two

different behaviors of the obtained gradients: the phosphorylated enzymes HPr show only a difference in the gradient higher near the membrane (10%), whereas for IIA·P its gradient is upward shifted along the whole radius (30% near the membrane, and 10% near the centre of the cell). These differences are linked with those observed in the SMET in Fig. 3.6. Note, that due to the spherical geometry, a slight divergence near the membrane has a larger contribution to the SMET difference than a divergence near the center. Turning to the three basic enzymes shown in Fig. 3.7(b), we find for HPr a relative concentration difference between the center and the membrane of the cell of 66% in both the PDE solution and the GMP averaged solution. However, in absolute value, GMP's solution for the radial gradient is consistently 16% lower than the PDE solution. Both the gradients for IIA and EI show an excellent agreement. Nevertheless, some of the molecules involved in their reactions, such as IIA·P, IICB·P and HPr·P, show a noticeable difference as seen in Fig. 3.6.



(a) (b)

Figure 3.7: *Profile gradients along the radius of a 1/8th of an sphere for the PTS system. Lines without glyphs are the solution of the PDE model in Blom and Peletier (2000). Each point is the average of all sites with center in* $[r-0.5/L, r+0.5/L]$*, and then averaged again over 50 trials. Note the good agreement of IIA, a consequence of the tuning of IICB molecules to match* $IIA \cdot P \cdot IICB$ *as seen in figure 3.6(c). EI has no significant gradient, while HPr differs by about 15% from the value predicted by the PDE model. The first point near the center shows a large deviation because it is only the average of one site and* 50 *trials.*

3.2.3 Discussion

For the PTS system, the main focus was on the reactions between membrane and cytosolic species, as well as on the gradients that may arise due to the spatial localisation of the membrane-bound reactions and the diffusion of cytosolic molecules. Due to the moderately high concentrations of all species, and the architecture of the network, fluctuations seem not to be of much significance. As shown in Fig. 3.7, only the species related to HPr and IIA show a significant gradient, in contrast to a flat gradient by EI which has the lowest concentra-

tion. This case illustrates that spatial inhomogeneities do not depend on the concentration levels, but on the whole pathway architecture, depending on the combination of reaction rates, diffusion coefficients, and most importantly, spatial localisation of reactions.

Having tuned the initial number of membrane-bound molecules of IICB, we obtained a good approximation of IIA·P·IICB and IIA, but IICB, IICB·P and IIA·P differ from the PDE solution (see Fig. 3.6 and 3.7). This difference is propagated down the pathway irregularly. For instance, it relatively strongly affects the gradient of HPr, though not in absolute sense ($< 2\,10^{-7}\,\mu M$) (see Fig. 3.7(b)) and not of IIA. The reaction procedure between bulk (cytosol) and surface (membrane) molecules, which is approximated by a bulk to bulk reaction in a small volume, remains a topic for further research. Note, that in the GMP method we have to use a constant number of particles rather than a constant concentration for the membrane-bound molecules as used in PDE models. Although this is a more realistic representation of the underlying biological process and avoids the arbitrary use of concentrations of membrane-bound molecules, this introduces a practical difficulty, since this number is often unknown. In PDE models this concentration is estimated from the bulk concentration, which does not necessarily provide a correct estimation.

3.3 Chemotaxis: Activation of Flagellar Motors

In the PTS example, membrane-bound molecules were not arranged in any particular configuration. However, in other systems they can be found forming clusters and also pairing with other molecules forming multi-units to carry out a more complex function. Using methods such as GMP, such microscopic configuration may escape the capabilities of the mesoscopic tools. Highly localised reactions might present some problems for the GMP and other lattice-based methods. Our next system, the bacterial chemotaxis, has two membrane-bound subsystems with a particular spatial micro-organisation of molecules into a cluster and flagellar motors. This spatial arrangement of flagellar-motor molecules is not exactly reproduced by mesoscopic tools, but it is instead approximated using the lattice discretisation. We compare the the results between GMP and MesoRD against the stochastic lattice-free methods Smoldyn. Smoldyn uses a continuous spatial description which together with the hard-ball molecules can achieve high spatial resolution of the diffusion and reaction processes.

Chemotaxis is a particular case of the generic class of two-component signal transduction pathways. This two-component signal transduction pathway does not belong to the majority of signalling networks that affect gene expression. Instead, the receiver of the signal are membrane-bound molecular motors that control the rotation of the flagella (see Fig. 1.6.) These motors are responsible for the motion of the cell.

E. coli is too small to sense spatial gradients to aid steer the swimming direction. Instead it uses temporal sensing to decide whether to keep swimming straight (the cell thinks it is swimming in the right direction), or tumble (when

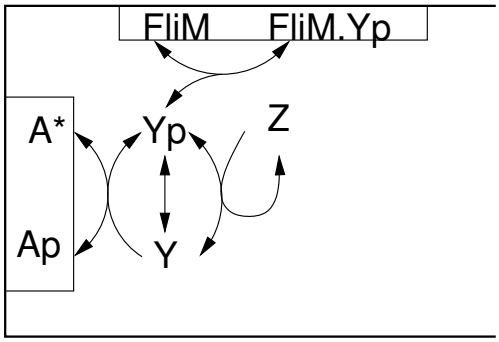


Figure 3.8: *CheY diffusion model for* E. coli.*CheA* kinase dimers form an array of 1260 chemotaxis receptors on the interior side of the membrane wall. CheAundergoes autophosphorylation. The phosphryl group is transfered to CheY. Then CheYpdiffuses throughout the cytosol and up to 34 of them can bind to a FliMreceptor. A total of 8200 CheYsignalling molecules (both non- and phosphorylated) and 1600 CheZdimers diffuse freely in the cytoplasm.*

swimming in the wrong direction.) Nutrients are sensed by a transmembrane cluster of receptors located at one end of the elongated cell and the signal is transferred to the flagellar motors. We might wonder why is there a cluster of receptors if the cell is too small to sense external gradients. In this section we will investigate some consequences of receptor clustering. However, in the chemotaxis case, it is known that the clustering of the chemotactic sensors function as a mean of obtaining a strong cooperative effect between adjacent sensors which increases the sensitivity of the whole sensor (Shimizu et al., 2003 ; Kim et al., 2002). Transmission of this signal is achieved by diffusion of one chemical species (CheYp).

Despite the small size of *E. coli*, spatial and stochastic simulations predict that gradients of the transmission signal in the cytosol exist, but they are shallow and are difficult to detect in vivo Lipkow et al. (2005). These gradients arise due to the spatial localization of the receptor on one end of the cell, together with a certain degree of diffusion limitation of the receptor related reactions.

3.3.1 The Model

We use a signal transduction system, also found in *E. coli*, which has been studied using Brownian Dynamic simulations. Thus now both tools are discrete, although Smoldyn is lattice-free (Andrews and Bray, 2004). Such feature gives some advantage over GMP when there are particles that need to be placed at particular points in the domain. The model presented by Lipkow et al. (2005), is the diffusion of phosphorylated CheY, a subset of the chemotaxis pathway.

Chemotaxis is probably one of the most well-studied *taxi* systems in bacteria. The model we use in this thesis is based upon that described by Lipkow et al. (2005), which we also use as a reference solution since it was studied with the

Table 3.5: *Diffusion coefficient for the Chemotaxis model.*

Species	Location	Diffusion coefficient
CheA,CheA*	Membrane wall	static
CheY,CheYp	Cytoplasm	$10\,\mu\mathrm{m}^2$
CheZ	Cytoplasm	$6\,\mu\mathrm{m}^2$
FliM	Membrane wall	static

Table 3.6: *Reaction rate parameters for the Chemotaxis model.*

	Reaction	Rate constant
Unimolecular	$\mathrm{CheA}^* \rightarrow \mathrm{CheAp}$	$3.4 \cdot 10^1\,\mathrm{s}^{-1}$
	$\mathrm{CheY} \rightarrow \mathrm{CheYp}$	$5.0 \cdot 10^{-5}\,\mathrm{s}^{-1}$
	$\mathrm{CheYp} \rightarrow \mathrm{CheY}$	$8.5 \cdot 10^{-2}\,\mathrm{s}^{-1}$
	$\mathrm{CheYp} \rightarrow \mathrm{FliM} + \mathrm{CheYp}$	$2.0 \cdot 10^1\,\mathrm{s}^{-1}$
Bimolecular	$\mathrm{CheAp} + \mathrm{CheY} \rightarrow \mathrm{CheA}^* + \mathrm{CheYp}$	$1.0 \cdot 10^8\,\mathrm{M}^{-1}\mathrm{s}^{-1}$
	$\mathrm{CheZ} + \mathrm{CheYp} \rightarrow \mathrm{CheZ} + \mathrm{CheY}$	$1.6 \cdot 10^6\,\mathrm{M}^{-1}\mathrm{s}^{-1}$
	$\mathrm{FliM} + \mathrm{CheYp} \rightarrow \mathrm{FliM} \cdot \mathrm{CheYp}$	$5.0 \cdot 10^6\,\mathrm{M}^{-1}\mathrm{s}^{-1}$

particle-based and lattice-free method Smoldyn (Andrews and Bray, 2004). An illustration of the model as listed in Table 3.6 is shown in Fig. 3.8. Table 3.5 gives the additional parameters needed for the spatial simulation tools such as GMP and MesoRD.

The sensing system of chemotaxis consist of a cluster of receptors located mainly at one end of the cell. When the cell detect external food supply, CheA dimers activate (CheA*) and pass their phosphoryl group the CheY. Active CheA* autophosphorylate. Phosphorylated CheYp diffuse freely through the cytosol until it either binds reversibly to the motors FliM, or is dephosphorylated by scavengers CheZ or it autodephosphorylates.

In line with Lipkow et al. (2005) set of parameters, *E. coli* has dimensions of $2.52 \times 0.88 \times 0.88\mu$m. For GMP to reach the same level of spatial detail as Smoldyn, a very fine discretisation is required. This constrain is imposed by the particular arrangement of the FliM motors. Lattice sites with length L_{sv} equal to 40 nm will be used. For this lattice size the FliM cluster occupies 4 sites (2×2). The receptor is a one-layer array of 13×13 and 7×7 sub-volumes for $L_{sv} = 40$ and 80 nm, respectively.

Table 3.7: *Average time (in seconds) to reach motor occupancy of 10 CheYpmolecules. Results are averaged over 10 runs for every method.*

Method	Comments	T_{M1}	T_{M2}	T_{M3}	T_{M4}
Smoldyn	$\Delta t = 0.2\,ms$	0.11	0.19	0.22	0.29
MesoRD	$L_{sv} = 20\,nm$	0.06	0.10	0.15	0.21
MesoRD	$L_{sv} = 40\,nm$	0.06	0.11	0.15	0.22
MesoRD	$L_{sv} = 80\,nm$	0.06	0.10	0.14	0.19
GMP	$L_{sv} = 40\,nm$	0.06	0.08	0.17	0.23

3.3.2 Results

To compare GMP's solution quality we use as a reference a solution that is produced with Smoldyn. We also use MesoRD to compare solutions, which despite being an exact solver for the RDME has the same spatial resolution limitations as GMP.

The first measurement to compare is the time required to reach a given level of flagellas's motor occupancy. Initially all CheA dimers in the receptor are in phosphorylated form, while CheY and CheZ are freely diffusing in the cytoplasm. The results in Fig.3.9 show averages over 10 runs of 1 second for every method. The growth of the number of motorbound CheYp is visibly slower for motors placed farther along the cell. GMP produces slightly higher averages which can be attributed to the splitting error between reaction and diffusion (Table 3.7). Molecules diffuse in *jumps* due to the fixed diffusion time-step which affects the first-passage properties of the diffusing front while the macroscopic mean square displacement is reproduced correctly. Finally, the average time to reach the given threshold is noticeably higher for Smoldyn. This difference cannot be explained by a wrong treatment of the nonlinear reactions in RDME due to sub-volumes larger than the correlation length. Simulations for different sizes of the sub-volumes yielded the same results within statistical error.

In MesoRD and GMP a reaction may occur when reactants are in the same sub-volume. For BD-based methods like Smoldyn two reacting molecules need to be within the binding radius in order for an event to occur. This is a stricter constraint because it is possible for two molecules to simply pass each other despite being in a very close vicinity which would result in a reaction in RDME-type methods (unless the discretisation is such that each sub-volume contains just one motor molecule).

Another property of the CheYp diffusion model we have studied is the average and the noise in the motor occupancy in steady-state (Table 3.8).

As to compare MesoRD and GMP results, we picked the simulations with 40 nm sub-volumes. Both RDME-type methods yield very similar results, although again, averages from GMP are slightly higher than those obtained from

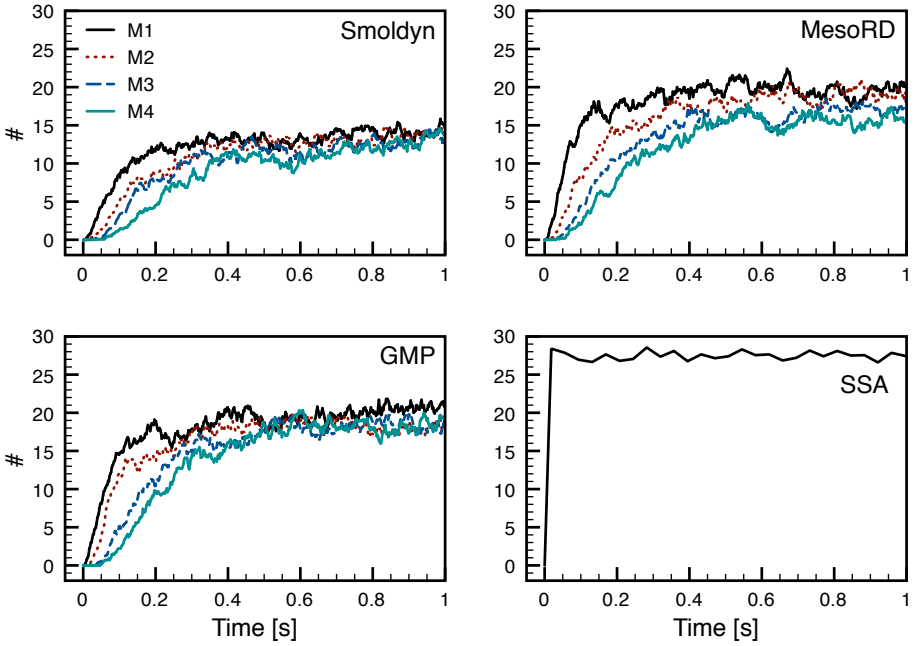


Figure 3.9: *Change in flagellar motor occupancy in time. Time-step used in Smoldyn is* 0.2 *ms. The side of the sub-volume* L_{sv} *in MesoRD and GMP is* 40 *nm. results do not change significantly when* $L_{sv} = 20$ *or* 80 *nm are used. Note that the larger difference is between the Smoldyn and the two RDME-type methods. The SSA (Gillespie's) case is shown in order to show the strong effect of diffusion in the system.*

MesoRD; Smoldyn computes approximately 20% lower averages. Note the interesting effect regarding noise in the motor occupancy, which increases for motors placed further from the receptor. This behavior of noise can be attributed to the concentration gradient of CheYp (high at the receptor and low at the posterior end). A smaller CheYp concentration at motor four compared to motor one results in a drop of the average motor occupation, and causes the fluctuations in binding to the FliM cluster to increase.

Additionally, we provide the SSA result for the motor occupancy (lower-right plot in Fig. 3.9 and Table 3.8) with the same reaction rates as in the other simulations. The occupancy for only one motor cluster is shown because all of them are equivalent if space is not included. This is clearly a wrong approach to model CheY diffusion, nevertheless it gives an indication of the error one can make when spatial information is omitted either by not accounting for geometry or by a lack of correction in the diffusion-limited reaction rates. The average occupation, which is higher than in the other methods, is a direct consequence of a lack of delay due to the diffusion of CheYp towards the motors. Lower noise,

Table 3.8: *Average and noise in the level of motor occupancy in steady-state. Averages were computed from simulations of length 21 s after the equilibration period of 1 s. Parameters of the simulations are the same as in Figure 3.9. MesoRD and GMP used* $L_{sv} = 40\,nm$.

	M1		M2		M3		M4	
Method	$\langle N \rangle$	η	$\langle N \rangle$	η	$\langle N \rangle$	η	$\langle N \rangle$	η
Smoldyn	13.9	0.20	13.2	0.21	12.5	0.23	12.4	0.23
MesoRD	19.1	0.17	18.0	0.18	16.5	0.20	16.0	0.22
GMP	20.3	0.15	19.0	0.16	18.6	0.16	18.1	0.16
SSA	27.5	0.08						

on the other hand, results from the constant and immediate supply of reactants while in the case of simulations accounting for space, the supply is considerably lower due to the appearance of the CheYp concentration gradient.

3.3.3 Conclusions

In this example of chemotaxis we are able to compare the mean and noise results across different simulation methods. The principal point of discussion is whether a more detailed method (such as lattice-free Smoldyn) can yield a significantly more accurate and trustworthy result than good approximate methods (MesoRD and GMP). It is is not surprising to find such similarity in the results of MesoRD and GMP since as we discussed in section 3.1.2 the levels of noise are expected to be lower due to the operator-split mechanism. However, the 20% lower average levels highlight the differences between the two methods. The fact that Smoldyn has a higher spatial resolution for complex structures such as the flagellar motor arrangement gives it an advantage over the lattice-based methods. Nonetheless we should be cautious to take Smoldyn's solutions as an accurate and trustworthy one due to the differences in noise levels obtained in gene expression as studied in section 3.1.2. After all, the physics underlying Smoldyn are not as robust as those underlying GFRD (see further discussion in supplement material in Dobrzynski et al. (2007)). Given the uncertainties discussed above it is difficult to be confident that the levels of noise correspond to actual ones, since there is a lack of *in vivo* measurements. In spite of such difficulties, we should not get discouraged to use these computational methods for simulating processes which, in fact, are unfeasible to measure in the laboratory. Only recent technology has been developed to visualise individual reaction in single cells (Ishijima and Yanagida, 2001 ; Cai et al., 2006 ; Elf et al., 2007). Nevertheless, we could expect in the near future to have the right technology to perform measurements that can be directly compared to these explicit simulations.

Regarding the proper lattice-site size we seem to have a contradiction with

with the gene expression results. For Chemotaxis the systems has a lower dependence on the lattice-site size, as evidenced by the results shown in Table 3.7. Such decrease could be explained if this system is less diffusion-limited than the gene expression, which in fact it has only one binding molecule instead of 34 for each flagellar motor. We find that this argument is also in line with the idea that more homogeneous systems require less spatial information. However, the we have tried a limited range of spatial resolutions, and as we can note in the fourth motor, spatial resolution seems to have a barely noticeable (3%) effect. We find ourselves again in a situation where it is difficult to estimate beforehand the optimal lattice discretisation size without prior knowledge of the real solution.

In summary, the results obtained with the operator-split method GMP are consistent, within a certain margin of variability, to other available methods.

3.4 Conclusions

We have applied the Gillespie Multi-Particle (GMP) method to three representative models for stochastic and spatial modelling of intracellular biochemical processes. In spite of the existence of tailored methods for a specific usage, such as GFRD, the generality of the GMP method and its class of RDME-based methods performs well in a variety of scenarios. Choosing a proper unique lattice size can be tricky when dealing with complex systems with multiple, diverse types of molecules, specially when the mean reaction distance or their effective sizes vary significantly. The recommendations of Baras and Mansour (1996) and as explained and followed in section 2.2.4 seem to point to a reasonable lattice size. Lattice-free methods, such as Smoldyn, circumvent this problem by dealing explicitly with particle sizes. However, in the gene expression model, we have to admit that we were surprised to find that Smoldyn's results were in close agreement to those of the reference solution provided by GFRD. Andrews and Bray claim that the method is rigourously derived to fulfil the detailed balance principle and thus it is expected to produce results matching those of GFRD. After consulting the author, and using a broad range of parameters, we still do not have a clear picture of why this methods yields such low noise values.

We have left out complex systems with feedback leading to amplification of signal and with multiple attractor points. In this area, though, we could expect to observe larger differences between the different methods since the small differences, obtained in the study cases of this chapter, can be amplified and drive the systems state into different steady states, should these exist.

Although the use of explicit simulations offers direct insights into the final solution of the (complex) chemical system, an analytical understanding of it, or at least of some key parts, may offer a greater, more valuable understanding of the dependencies among the reactants as well as their spatial location. In the next chapter we focus on the study of a simple signalling system consisting of only two chemical reactions in which one of them is spatially restricted to the membrane, adding thus a potential spatial dimension to the problem. The sim-

Table 3.9: *Scaling of the computational cost for the spatial discrete methods presented in this comparison. Abbreviations: N_S – number of molecules of a given species, N_R – number of reaction channels, N_{sv} – the total number of sub-volumes, $\langle \tau_R \rangle$ – average time between reactions, $\langle \tau_D \rangle$ – average time between diffusive movements, D – diffusion coefficient, k_R – rate of reaction R, L_{sv} – length of the sub-volume, L – length of the total volume.*

Method	Computational cost	Comment
GFRD	$\sim \sum_S N_S$	Diffusive movements.
(event-driven)	$\sim \sum_{N_R} \prod_{S \in R} N_S$	Reactive distances.
Smoldyn	As GFRD	As GFRD.
(fixed time step)		
MesoRD	$\sim \log N_R$	Gibson and Bruck (2000).
(event-driven)	$\sim \log N_{sv}$	A sub-volume adds
		a diffusive reaction.
	$\sim \langle \tau_D \rangle^{-1}$	
	$\sim \langle \tau_R \rangle^{-1}$	
GMP[a]	$\sim \sum_S N_S$ [b]	Diffusive movements.
	$\sim N_R$	As in the SSA.
	$\sim \tau_D^{-1}$	Fixed diffusion time step.
	$\sim \langle \tau_R \rangle^{-1}$	

Where: $\langle \tau_D \rangle \propto L_{sv}^2 / D \cdot N_{sv}/N_S$, $\langle \tau_R \rangle \propto L_{sv}^3 / k_R \cdot \prod_{S \in R} N_{sv}/N_S$, $\tau_D \propto L^2 / D \cdot N_{sv}^{-2/3}$.

[a] The scheme is event-driven for reactions but the diffusion time step τ_D is fixed. The diffusion time step is assigned for every diffusing species.

[b] If $N_S / N_{sv} > 90$ molecules are moved in bulk, otherwise one-by-one in τ_D. Note that for event-driven schemes the cost of diffusive movements or of computing reactive distances is given per iteration time step.

plicity of such system enables us to derive analytical expressions and to validate them with computer simulations of similar nature, this is explicit simulations in space, than the GMP method used in this chapter and introduced in chapter 2. We shed light into the design properties of the two-component signalling systems and how a few particles may be organised to achieve a robust and efficient degree of functionality.