

# 10 Computational Cell Biology – The Stochastic Approach

**Thomas Simon Shimizu and Dennis Bray**

## 10.1 Introduction

In broad terms, the current interest in “computational cell biology” reflects the contemporary fascination with electronic networks of all kinds. There is a widespread feeling that the speed of computers and the sophistication of programmers can at last match the bewildering molecular complexity of living cells. This viewpoint is encouraged and fed by recent genomic studies, which have built up such an impetus that they are now pushing into areas outside the genome. Evidently, sequence information by itself cannot explain the functioning of a living cell or organism. It is also true that we have built up, over the past century, an enormous body of information about proteins and other molecules inside cells. Why should we not — the argument goes — store, collate and analyze these data by comprehensive, computer-intensive techniques similar to those currently employed to analyze genomes?

Unfortunately, no consensus presently exists as to how best to perform this analysis, or to what we can expect as a result. One clear-cut function of computers in cell biology is to store large amounts of information in a logical and accessible form. This role has seen its most public triumph in the generation of genomic databases, but also underpins the giant strides made in the determination of protein structure. Many databases containing integrated information on specific organisms have been developed and some of these allow the user to access specific molecular details on individual cells, such as WormBase<sup>1</sup> and EcoCyc<sup>2</sup>. There are also large programs containing information about specific cell types, such as the software developed by Denis Noble to analyze the behavior of heart muscle cells<sup>3</sup>. Several ambitious projects have been initiated that aim to simulate entire cells or parts of cells at a molecular level, such as E-CELL<sup>4</sup> and the Virtual Cell<sup>5</sup>.

Computers were of course used by biologists before the genomic era. The area of metabolic modeling, for example, with its roots in enzyme kinetics, was one of the earliest to be adapted to computer simulation. Many software packages have been written that allow the kinetic performance of enzyme pathways to be represented and evaluated quantitatively, such as GEPASI<sup>6</sup>, MIST<sup>7</sup> and

SCAMP<sup>8</sup>. This is also an area of commercial interest and biotechnology companies engaged in the production of food or drugs by fermentation or allied processes routinely evaluate their production by flux-analysis programs, often aided by metabolic control analysis.

Neurobiology is another computationally rich area. Whether because of their background in the physical science, or because of the computer-like nature of the brain, neurophysiologists have always been much more open to the use of computers than cell biologists. Several large computer packages have been developed and (what is far more significant) widely used as adjuncts to research neurophysiology. Packages such as GENESIS<sup>9</sup> and NEURON<sup>10</sup> provide integrated suites of routines for the recording and analysis of electrical data, the simulated performance of individual axons, and the investigation of networks of nerve cells and cortical activity.

In contrast to the above areas, topics that come under the rubric of core cell biology — those not directly concerned with DNA sequences, ions, or low molecular weight metabolites — are more difficult to handle computationally. Cell signalling, cell motility, organelle transport, gene transcription, morphogenesis and cellular differentiation cannot easily be accommodated into existing computational frameworks. Attempts to use computers in these areas are still at a stage of exploratory software development, usually in the hands of individual research groups. Conventional approaches using the numerical integration of continuous, deterministic rate equations sometimes provide a convenient route, especially when systems are very large or when molecular details are of little importance. But as the resolution of experimental techniques increases, so the limitations of conventional models become more evident. Difficulties include the combinatorial explosion of large numbers of different species, the importance of spatial location within the cell, and the instability associated with reactions between small numbers of molecular species.

## 10.2 Stochastic simulation of cellular processes

In recent years, a number of research groups have attempted to use a radically different approach to the processes occurring within cells<sup>11,12,13</sup>. The idea is to represent individual molecules rather than the concentrations of molecular species, and to apply Monte Carlo methods to predict their interactions. Motivation for this new approach comes from the realization that many crucial events in living cells depend on the interaction of small numbers of molecules

and hence are sensitive to the underlying stochasticity of the reaction processes. Under these conditions, the usual approach taken to biochemical reactions of analyzing their characteristic continuous, deterministic rate equations breaks down and fails to predict the behavior of the system accurately. Signalling pathways, for example, commonly operate close to points of instability and frequently employ feedback and oscillatory reaction networks that are sensitive to the operation of small numbers of molecules<sup>14,15</sup>. Only 200 K<sup>+</sup> and Na<sup>+</sup> channels responsive to changes in intracellular Ca<sup>2+</sup> are responsible for a key step in many neutrophil signalling pathways<sup>14</sup>. Gene transcription is controlled by small assemblies of proteins operating in an all-or-none fashion, so that whether a specific protein is expressed or not is, to some extent, a matter of chance<sup>16,17,18,19</sup>. The performance of sensory detectors such as retinal rod outer segments<sup>20,21</sup> and even the firing of individual nerve cells<sup>22,23</sup> are intrinsically stochastic.

In the stochastic modeling approach, rate equations are replaced by individual reaction probabilities and the output has a physically-realistic stochastic nature. Techniques are available by which large numbers of related species can be coded in an economical fashion and key concepts, such as signalling complexes and the thermally-driven flipping of protein conformations, can be embodied into the program. Stochastic modeling may help us to integrate biochemical and thermodynamic data in a coherent and manageable way.

### 10.3 Modeling bacterial chemotaxis

We have used both deterministic and individual-based stochastic programs to investigate the pathway of intracellular signals used by coliform bacteria in the detection of chemotactic stimuli<sup>24,25,261</sup>. The models are based on physiological data collected from single tethered bacteria of over 60 mutant genotypes. Quantitative discrepancies between computer simulations and experimental data throw a spotlight on areas of uncertainty in the signal transduction pathway, highlighting the importance of spatial organization to the logical operation of the pathway. In particular they emphasize the function of a specific, well-characterized, cluster of proteins associated with the chemotaxis receptors which acts like a self-contained computational cassette.

The individual-based stochastic simulation program STOCHSIM was writ-

---

<sup>1</sup> A resumé of this work together with a list of published references can be found at <http://www.zoo.cam.ac.uk/comp-cell>.

ten by Carl Firth as part of his PhD work at the University of Cambridge<sup>26</sup>. It was developed as part of a study of bacterial chemotaxis to be a more realistic way of representing the stochastic features of this signalling pathway and also as a means to handle the large numbers of individual reactions encountered<sup>26,27</sup>. The program provides a general-purpose biochemical simulator in which each molecule or molecular complex in the system is represented as an individual software object. Reactions between molecules occur stochastically, according to probabilities derived from known rate constants. An important feature of the program is its ability to represent multiple post-translational modifications and conformational states of protein molecules.

#### 10.4 Description of the STOCHSIM algorithm

In STOCHSIM, each molecule (not each population of molecular species) is represented as an individual software object, and a number of dummy molecules, or “pseudo-molecules”, are also included in the reaction system. Time is quantized into a series of discrete, independent time-slices, the size of which is determined by the most rapid reaction in the system. In each time-slice, STOCHSIM first chooses one molecule at random from the population of “real” molecules, and then makes another selection from the entire population including the pseudo-molecules. If two molecules are selected, they are tested for all possible bimolecular reactions for the particular reactant combination. If one molecule and one pseudo-molecule are chosen, the molecule is tested for all possible unimolecular reactions it can undergo.

Reaction probabilities are pre-computed at initialization time, and stored in a look-up table so that they need not be calculated during the execution of each time-slice. These probabilities scale linearly with the size of the time-slice (Eqs. 10.1 and 10.2 below), so that if the time-slices are sufficiently small, a single random number can be used to test for all possible reactions that a particular combination of reactants can undergo. Once the reactant molecules are chosen, the set of possible reactions are retrieved from the look-up table with their probabilities. STOCHSIM then iterates through these reactions in turn and computes a “cumulative probability” for each of the possible outcomes. The set of cumulative probabilities can then be compared with a single random number to choose which reaction, if any, occurs. If a reaction does occur, the system is updated accordingly and the next time-slice begins with another pair of molecules being selected.

The probabilities stored in the look-up table are calculated from the following five parameters: (i) the deterministic rate constant ( $k_1$  and  $k_2$  for uni- and bi-molecular reactions, respectively), (ii) the size of the time increment ( $\Delta t$ ), (iii) the number of molecules in the system ( $n$ ), (iv) the number of pseudo-molecules in the system ( $n_0$ ), and (v) the volume of the system ( $V$ ). Using these parameters, the probabilities for uni- and bi-molecular reactions ( $p_1$  and  $p_2$ , respectively) are obtained by:

$$p_1 = \frac{k_1 n(n + n_0) \Delta t}{n_0} \quad (10.1)$$

and

$$p_2 = \frac{k_2 n(n + n_0) \Delta t}{2N_A V}. \quad (10.2)$$

The previously published derivation of these expressions<sup>13,26</sup> is summarized in Appendix A for reference.

Whenever a molecular species in the system can exist in more than one state, then the program encodes it as a “multistate molecule” with a series of binary flags. Each flag represents a state or property of the molecule, such as a conformational state, the binding of ligand, or covalent modification (e.g. phosphorylation, methylation, etc.). The flags specify the instantaneous state of the molecule and may modify the reactions it can participate in. For instance, a multistate molecule may participate in a reaction at an increased rate as a result of phosphorylation, or fail to react because it is in an inactive conformation. The flags themselves can be modified in each time step as a result of a reaction, or they can be instantaneously equilibrated according to a fixed probability. The latter tactic is used with processes such as ligand binding or conformational change that occur several orders of magnitude faster than other chemical reactions in the system.

Let us say that in a particular time step, STOCHSIM has selected one or more multistate molecules. It then proceeds in the following manner. First any rapidly-equilibrated “fast flags” on the molecule are assigned to be on or off according to a weighted probability. A protein conformation flag, for example, can be set to be active or inactive, according to which other flags of the molecules are currently on. A ligand binding flag can, if desired, be set in a similar fashion, based on the concentration of ligand and the  $K_d$ . Once the fast flags have been set, then the program inspects the reactions available to the two species  $A$  and  $B$ . The chemical change associated with each type of

reaction (binding, phosphotransfer, methylation, etc.) is represented in the program together with “base values” of the reaction rate constants. The particular instantiation of the reaction, specified by the current state of the flags on *A* and *B*, is accessed from an array of values calculated at the beginning of the program, when the reaction system is being initialized. Values in the array modify the reaction probability according to the particular set of binary flags. In this manner, STOCHSIM calculates a set of probabilities, corresponding to the reactions available to the particular states of molecules *A* and *B*, and then uses a random number to select which reaction (if any) will be executed in the next step. The reaction will be performed, if appropriate, and the relevant slow flag flipped.

Although it sounds complicated, the above sequence of events within an individual iteration takes place very quickly and even a relatively slow computer can carry out hundreds of thousands of iterations every second. Moreover, the strategy has the advantage of being intuitively simple and close to physical reality. For example, it is easy, if required, to label selected molecules and to follow their changes with time. Lastly, the speed of the program depends not on the number of reactions but on the numbers of molecules *n* in the reaction system (with a time of execution proportional to  $n^2$ ). The STOCHSIM distribution<sup>2</sup> consists of a platform-independent core simulation engine encapsulating the algorithm just described, together with separate graphical and user interfaces.

### 10.5 Comparison with the Gillespie algorithm

Daniel Gillespie showed, in the 1970s, that it is possible to simulate chemical reactions by an efficient stochastic algorithm<sup>28</sup>. He showed that this algorithm gives the same results, on average, as conventional kinetic treatments<sup>29</sup>, and later provided a rigorous mathematical derivation for the procedure<sup>30</sup>. The Gillespie algorithm has since been used on numerous occasions to analyze biochemical kinetics, for example to simulate the stochastic events in lambda lysogeny<sup>11,31</sup>. In view of its evident success, the question therefore arises: Why in our work did we not use the Gillespie algorithm but chose to develop our own formulation? As shown in Appendix B, the Gillespie and STOCHSIM algorithms are based on equivalent fundamental physical assumptions. How-

<sup>2</sup> The latest version of STOCHSIM can be obtained via FTP from <ftp://ftp.cds.caltech.edu/pub/dbray/>.

ever, significant practical differences arise in applying the two algorithms to biochemical systems, as described below.

The Gillespie algorithm makes time steps of variable length, based on the reaction rate constants and population size of each chemical species. In each iteration, one random number is used to determine when the next reaction will occur, and another random number determines which reaction it will be. Both the time of the next reaction  $\tau$ , and the type of the next reaction  $\mu$  are determined by the rate constants of all reactions and the current numbers of their substrate molecules. Upon the execution of the selected reaction in each iteration, the chemical populations are altered according to the stoichiometry of the reaction, and the process is repeated. By avoiding the common simulation strategy of discretizing time into finite intervals, the Gillespie algorithm benefits from both efficiency and precision — no time is wasted on simulation iterations in which no reactions occur, and the treatment of time as a continuum allows the generation of an “exact” series of  $\tau$  values based on rigorously derived probability density functions.

However, the efficiency of the Gillespie algorithm comes at a cost, and its precision is guaranteed only for chemical systems with certain properties. The efficient algorithm that selects which reaction to execute next and what time interval to take, does not represent each molecule in the system separately. With regard to the reactions of a typical cell signalling pathway, for example, it cannot associate physical quantities with each molecule, nor trace the fate of particular molecules over a period of time. Similarly, without the ability to associate positional and velocity information with each particle, the algorithm cannot be easily adapted to simulate diffusion, localization or spatial heterogeneity. Indeed, the “exactness” of the Gillespie algorithm holds only for spatially homogeneous, thermodynamically equilibrated systems in which non-reactive molecular encounters occur much more frequently than reactive ones<sup>28,30</sup>.

A second limitation of the Gillespie algorithm (from a cell biological standpoint) is that it cannot easily handle the reactions of multistate molecules. Protein molecules are very frequently modified in the cell so as to alter their catalytic activity, binding affinity and so on. Cell signalling pathways, for example, carry information in the form of chemical changes such as phosphorylation or methylation, or as conformational states. A multi-protein complex may contain upwards of twenty sites, each of which can often be modified independently and each of which can, in principle, influence how the complex will participate in chemical reactions. With twenty sites, a complex can exist

in a total of  $2^{20}$ , or one million, unique states, each of which could react in a slightly different way. If our multi-protein complex interacts with only ten other chemical species, a detailed model may contain as many as ten million distinct chemical reactions, a combinatorial explosion. Any program in which the time taken increases in proportion to the number of reactions, as in a conventional, deterministic model, or in the Gillespie method, will come to a halt under these conditions.

We see therefore that STOCHSIM and the Gillespie algorithm take different approaches and are suited to different situations. STOCHSIM is likely to be slower than the Gillespie algorithm in calculating the eventual outcome of a small set of simple biochemical reactions, especially when the numbers of molecules is large. However, if the system contains molecules that can exist in a large number of states, then STOCHSIM will not only be faster but also closer to physical reality. It is easy, if required, to label selected molecules in this program and to follow their changes with time, including changes to their detailed post-translational modification and conformational states. Lastly, spatial structures can be incorporated into the STOCHSIM framework with relative ease, as one can directly define the spatial location of individual molecules — something that would be difficult to do with the Gillespie algorithm.

## 10.6 Spatial extensions to STOCHSIM

The original version of STOCHSIM (1.0) treated the entire reaction system as a uniformly mixed solution. Although this is clearly not how molecules are arranged within living cells, the omission of spatial heterogeneity has been a norm in biochemical simulations because it greatly facilitates modeling and reduces the computational load of simulation. However, as the resolution of our understanding of biochemical processes increases, it is becoming clear that even in bacteria, the spatial organization of molecules often play an important role<sup>32,33</sup>.

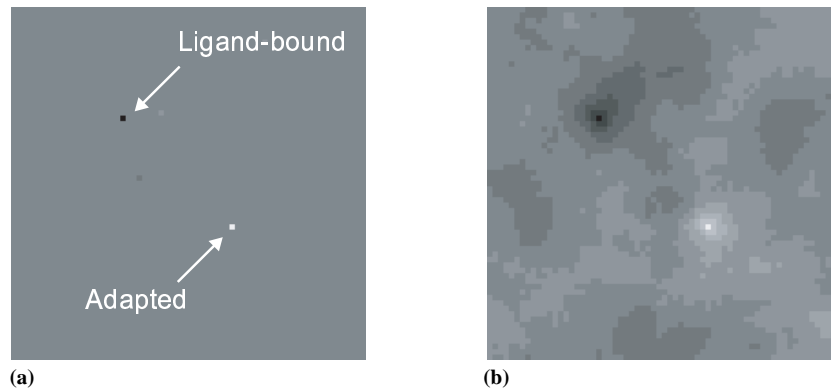
In the chemotaxis pathway, the membrane receptors are not only associated with the signalling molecules CheW and CheA in the cytoplasm, but also clustered together, usually at one pole of the cell<sup>34</sup>. The density of packing of molecules in the cluster implies a regular arrangement, and recent model building led to the proposal of a hexagonal lattice built from CheA and CheW into which receptor dimers are inserted in sets of three<sup>35</sup>. An arrangement of this kind would create a “microenvironment” within the cytoplasm which

could sequester certain molecules and exclude others simply through binding affinities (and without an internal membrane). A regular lattice would allow neighboring receptors to influence each other's activity by what has been termed "conformational spread"<sup>36,37</sup>. We must also consider the time taken for diffusible components of the signal pathway, notably CheY and its phosphorylated derivative, which have to shuttle repeatedly between the receptor complex on the plasma membrane and the flagellar motors. Although this time is short, consistent with distances of less than a micrometer and a diffusion coefficient of around  $5 \times 10^{-8} \text{ cm}^2\text{s}^{-1}$ , it is in principle measurable by recent techniques<sup>38,39</sup>. A fully realistic model would have to deal with not only time delays but also the possibilities that diffusing species might have a non-uniform distribution and move within privileged channels in the interior of the cell.

Considerations such as these encouraged us to extend STOCHSIM to incorporate spatial representation, and the positions of the important molecular species within the cell. As a first step, we have so far introduced modifications that allow us to represent the two dimensional arrangement of receptors in the plane of the plasma membrane. These changes, embodied in version 1.2 of STOCHSIM, assign two-dimensional coordinates to the array of receptors and permit such phenomena as the spread of activity from one receptor to its neighbor (Figure 10.1), and the diffusive movement of individual molecules bound to the surface array, to be represented.

### Model of bacterial chemotaxis with "conformational spread"

In the *Escherichia coli* chemotaxis pathway, one of the major discrepancies between simulation and experiment thus far has been in the sensitivity of the system to very small changes in attractant concentration<sup>36</sup>. For example, computer-based estimates of the minimal detectable concentration of aspartate is on the order of 100 nM<sup>36</sup> whereas responses to concentration jumps as small as 5 nM have been detected experimentally<sup>40</sup>. In order to test the possibility that the aforementioned "conformational spread" mechanism could account for this exquisite sensitivity observed in real bacteria, we have incorporated a two-dimensional representation of the receptor clusters into our previous STOCHSIM model<sup>27</sup> of the *E. coli* chemotaxis pathway. In that model, chemotactic receptors were modeled as multistate complexes with 11 binary flags to represent their various states. One of these flags represents the conformational state of the receptor, and was controlled by a rapid equilibrium. The probability of this flag being on or off, and hence the receptor being active or inactive, depended on the binding of ligand and the receptor's methylation

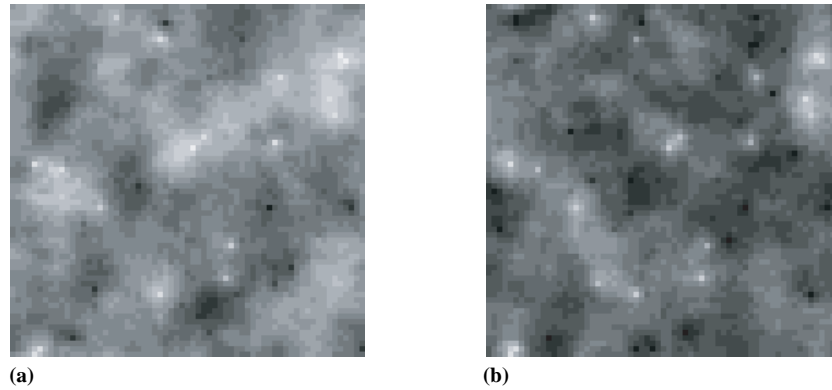


**Figure 10.1**

Graphical representations of the spatial patterns of receptor activity in a STOCHSIM simulation of the *E. coli* chemotaxis pathway with clustered receptors. The two arrays shown above are both square lattices of  $50 \times 50$  closely packed receptors, represented by one pixel each. These views are averaged over 0.1 seconds of simulated time, and the activities are represented by sixteen grey levels, with white corresponding to active and black to inactive receptors. In this simulation, ligand binding and methylation reactions were disabled in order to reveal the patterns due to activity spread alone. One receptor was permanently assigned to the ligand-bound (ligated and zero-methylated) state and another to the adapted (unligated and four-methylated) state. All other receptors were in the two-methylated state. In (a), no coupling reactions are defined and no spread of conformation is observed, but in (b), nearest-neighbor interactions allow the activity of certain receptors to “spread” over a wide range.

state. In our new spatially extended model, an additional “coupling” factor that depends on the number of neighbors in the active state, has been defined. The more active neighbors a receptor has, the higher the probability of being active. The spread of conformations that results from this can be visualized in time-averaged views of the receptor cluster (Figures 10.1 and 10.2).

Preliminary results of our simulations indicate that the conformational spread mechanism can indeed serve to enhance the chemotactic response at the cost of higher steady-state noise. In simulations where receptor clusters were first adapted to various background concentrations of attractant and tested for their response to a subsequent doubling in stimulus, significant amplification of the signal is observed (Figure 10.3). The level of amplification is not as high as that reported previously<sup>36,37</sup>, but the performance of the receptor cluster is less dependent on the precise value of the coupling strength. These differences arise because the previous models did not consider the multiple methylation states of the receptors, which can be covalently modified with up to four methyl groups. We are now investigating the effect of spatial patterns of methylation,

**Figure 10.2**

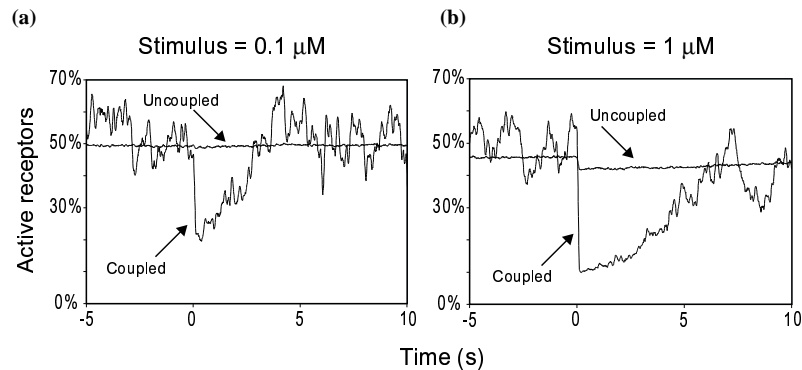
Patterns of receptor activity in an unconstrained simulation, in which ligand binding and methylation reactions were enabled in a coupled receptor array. These patterns are averaged over 0.01 seconds of simulated time, equivalent to roughly 16.7 million simulation iterations, in which each receptor flipped about 783 times on average. Discrete white patches correspond to the vicinities of adapted (highly methylated), and therefore highly active, receptors and black patches to centers of inactivity near bound ligand. In this simulation, the background attractant concentration was set to  $10^{-8}$  M ( $\sim \frac{1}{100} K_d$ ), and then doubled. The two patterns represent the average activity of intervals at 10 ms (a) prior to, and (b) after the doubling of stimulus. Rapid suppression of activity is observed, despite the very low concentration of stimulus.

an example of which is shown in Figure 10.4.

### 10.7 Future directions

The obvious next steps for development of STOCHSIM are the implementation of other geometries (e.g. triangular and hexagonal) for the two-dimensional arrays, further extending the spatial representation to a third dimension, and the development of a more generally accessible interface. Recent models of the neuromuscular junction include a realistic representation of the folds of the muscle membrane surface, the position and state of individual synaptic vesicles, and even the location of individual calcium ions<sup>41</sup>. Three-dimensional representation of a cell may require some form of compartmentalization of its contents, whether into regularly spaced volume elements (voxels) or more biologically relevant compartments, such as nucleus, and membrane cortex. The interface development is now focused around a cross-platform GUI<sup>42</sup>, and a simple command-line interface is also provided for ease of scripting.

On the more general question of the future of modeling in cell biology, it



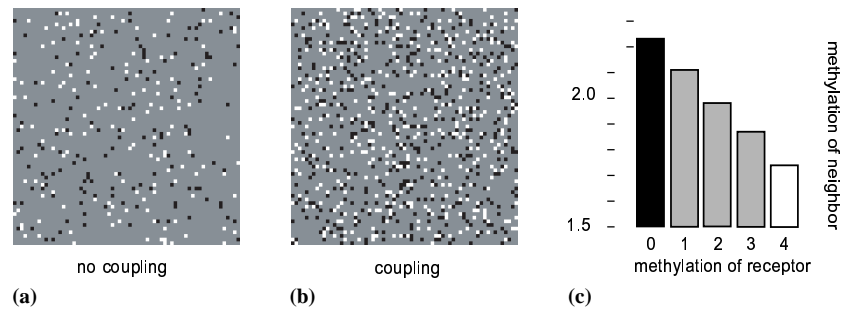
**Figure 10.3**

The enhancement of response achieved by coupling interactions in the receptor cluster model. Changes in total receptor activity during a doubling of stimulus at two background concentrations, (a)  $0.1 \mu\text{M}$  and (b)  $1 \mu\text{M}$  are shown. The concentration of ligand was doubled at time 0 in both (a) and (b). Significant enhancement is observed at both concentrations; the coupled array shows clear amplification of the ligand signal in (a), and in (b) only the coupled array shows a significant response to the doubling of attractant, demonstrating that the coupling could also act to increase the range of concentrations to which the system can respond.

seems unavoidable that stochastic representations will be increasingly useful. As the resolution of experimental techniques improves, they will generate large quantities of data relating to the behavior of individual cells and molecules. There will be an increasing emphasis on situations in which cell behavior depends on small numbers of molecules and analysis of such situations will naturally invoke individual-based simulations with a stochastic basis. Presently available simulation programs such as *STOCHSIM* will doubtless become integrated into larger software packages that allow non-specialist users to quickly identify their requirements and obtain results. The remorseless increase in the power and speed of computers available to the modeling community will accompany, and empower, these developments.

### Acknowledgments

We wish to thank Matthew Levin for proofreading and providing critical comments on this manuscript, as well as Nicolas Le Novère, Tom Duke and Gavin Brown for helpful discussions.

**Figure 10.4**

Spatial patterns of methylation. As in the activity snapshots (Figures 10.1 and 10.2), each pixel represents a single receptor in (a) an array without coupling and (b) an array with coupling. Four- and zero-methylated receptors (the extreme receptors) are highlighted in black and white, respectively, while receptors in all other methylation states are shown in gray. A notable feature of the changes in methylation state distribution between the uncoupled and coupled arrays is that the extreme receptors are more abundant and tend to be closer together in the latter. This tendency is due to the relationship between receptor activity and methylation reactions, and is also reflected in (c) the average number of methyl groups per neighbor for each methylation state.

## References

- Stein, L., Sternberg, P., Durbin, R., Thierry-Mieg, J., and J., S. *Nucleic Acids Res.* **29**, 82–86 (2001).
- Karp, P. D., Riley, M., Saier, M., Paulsen, I. T., Paley, S. M., and Pellegrini-Toole, A. *Nucleic Acids Res.* **28**, 56–59 (2000).
- Noble, D. *Oxford Cardiac Electrophysiology Group Website*. <http://noble.physiol.ox.ac.uk/>, (2001).
- Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T. S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J. C., and Hutchison III, C. A. *Bioinformatics* **15**, 72–84 (1999).
- Schaff, J., Fink, C. C., Slepchenko, B., Carson, J. H., and Loew, L. M. *Biophys. J.* **73**, 1135–1146 (1997).
- Mendes, P. *Comput. Appl. Biosci.* **9**, 563–571 (1993).
- Ehlde, M. and Zacchi, G. *Comput. Appl. Biosci.* **11**, 201–207 (1995).
- Sauro, H. *Comput. Appl. Biosci.* **9**, 441–450 (1993).
- Wilson, M. A., Bhalla, U. S., Uhley, J. D., and Bower, J. M. In *Advances in Neural Information Processing Systems*, Touretzky, D., editor, 485–492 (Morgan Kaufmann, San Mateo, CA, 1989).
- Hines, M. In *Neural Systems: Analysis and Modeling*, Eeckman, F. and Norwell, M. A., editors, 127–136. Kluwer, (1993).
- McAdams, H. H. and Arkin, A. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 814–819 (1997).
- Stiles, J. R., Bartol, T. M., Salpeter, E. E., and Salpeter, M. M. In *Computational Neuroscience*, Bower, J. M., editor, 279 (Plenum, New York, 1998).
- Morton-Firth, C. J. and Bray, D. *J. Theor. Biol.* **192**, 117–128 (1998).
- Hallett, M. B. *Perspect. Biol. Med.* **33**, 110–119 (1989).
- Goldbeter, A. *Biochemical Oscillations and Cellular Rhythms*. Cambridge University Press, Cambridge, (1996).

16. Ko, M. S. H. *J. Theor. Biol.* **153**, 181–194 (1991).
17. Kingston, R. and Green, M. R. *Curr. Biol.* **4**, 325–332 (1994).
18. Tjian, R. and Maniatis, T. *Cell* **77**, 5–8 (1994).
19. McAdams, H. H. and Arkin, A. *Trends. Genet.* **15**, 65–69 (1999).
20. Lamb, T. D. *Biophys. J.* **67**, 1439–1454 (1994).
21. Van Steveninck, R. D. R. and Laughlin, S. B. *Int. J. Neural Systems* **7**, 437–444 (1996).
22. Smetters, D. K. and Zador, A. *Curr. Biol.* **6**, 1217–1218 (1996).
23. White, J. A., Rubinstein, J. T., and Kay, A. R. *Trends Neurosci.* **23**, 131–137 (1999).
24. Bray, D., Bourret, R. B., and Simon, M. I. *Mol. Biol. Cell* **4**, 469–482 (1993).
25. Bray, D. and Bourret, R. B. *Mol. Biol. Cell* **6**, 1367–1380 (1995).
26. Morton-Firth, C. J. *Stochastic Simulation of Cell Signalling Pathways*. PhD thesis, University of Cambridge, Cambridge, UK CB2 3EJ, (1998).
27. Morton-Firth, C., Shimizu, T., and Bray, D. *J. Mol. Biol.* **286**, 1059–1074 (1999).
28. Gillespie, D. T. *J. Comput. Phys.* **22**, 403–434 (1976).
29. Gillespie, D. T. *J. Phys. Chem.* **81**, 2340–2361 (1977).
30. Gillespie, D. T. *Physica A* **188**, 404–425 (1992).
31. Arkin, A., Ross, J., and McAdams, H. H. *Genetics* **149**, 1633–1648 (1998).
32. RayChaudhuri, D., Gordon, G. S., and Wright, A. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 1332–1334 (2001).
33. Norris, V., Turnock, G., and Sigeo, D. *Mol. Microbiol.* **19**, 197–204 (1996).
34. Maddock, J. R. and Shapiro, L. *Science* **259**, 1717–1723 (1993).
35. Shimizu, T. S., Le Novère, N., Levin, M. D., Beavil, A. J., Sutton, B. J., and Bray, D. *Nat. Cell Biol.* **2**, 792–796 (2000).
36. Bray, D., Levin, M. D., and Morton-Firth, C. J. *Nature* **393**, 85–88 (1998).
37. Duke, T. A. J., and Bray, D. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10104–10108 (1999).
38. Elowitz, M. B., Surette, M. G., Wolf, P.-E., Stock, J. B., and Leibler, S. *J. Bacteriol.* **181**, 197–203 (1999).
39. Cluzel, P., Surette, M., and Leibler, S. *Science* **287**, 1652–1655 (2000).
40. Segall, J. E., Block, S. M., and Berg, H. C. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 8987–8991 (1986).
41. Stiles, J. R. and Bartol, T. M. In *Computational Neuroscience*, De Schutter, E., editor, 87–127 (CRC Press, Boca Raton, 2000).
42. Le Novère, N. and Shimizu, T. S. *Bioinformatics (in press)* (2001).

### Appendix A: Derivation of reaction probabilities in STOCHSIM

First consider the following unimolecular reaction with substrate  $A$ :

$$\frac{d[A]}{dt} = -k_1[A] \quad (\text{A.1})$$

If the size of the STOCHSIM time-slice  $\Delta t$  is sufficiently small, the change in the number of reactant molecules,  $\Delta n_A$ , within this interval will be between 0

and 1, and is given by

$$\Delta n_A = -k_1 n_A \Delta t \quad (\text{A.2})$$

where  $k_1$  is the deterministic rate constant.

In the STOCHSIM algorithm, the expected value of  $\Delta n_A$  within a single time-slice is

$$\begin{aligned} -\Delta n_A &= \text{Pr}(\text{molecule of A is selected in the first selection}) \\ &\quad \times \text{Pr}(\text{pseudo-molecule is selected in the second selection}) \\ &\quad \times p_1 \end{aligned} \quad (\text{A.3})$$

$$-\Delta n_A = \frac{n_A}{n} \times \frac{n_0}{n + n_0} \times p_1 \quad (\text{A.4})$$

Equating Eqs. (A.2) and (A.4) gives

$$p_1 = \frac{k_1 n(n + n_0) \Delta t}{n_0}. \quad (\text{A.5})$$

The probability for the bimolecular reaction can be derived similarly. Consider the following reaction with substrates  $B$  and  $C$ :

$$\frac{d[B]}{dt} = -k_2[B][C] \quad (\text{A.6})$$

In a very small  $\Delta t$ ,

$$\Delta n_B = -\frac{k_2 n_B n_C \Delta t}{2N_A V} \quad (\text{A.7})$$

where  $V$  is the volume of the reaction system, and  $N_A$  is Avogadro's constant.

In STOCHSIM, the expected value of  $\Delta n_B$  within a single time-slice is

$$\begin{aligned} -\Delta n_B &= \{\text{Pr}(\text{molecule of B is selected in the first selection}) \\ &\quad \times \text{Pr}(\text{molecule of C is selected in the second selection}) \\ &\quad \times p_2\} \\ &\quad + \{\text{Pr}(\text{molecule of C is selected in the first selection}) \\ &\quad \times \text{Pr}(\text{molecule of B is selected in the second selection}) \\ &\quad \times p_2\} \end{aligned} \quad (\text{A.8})$$

$$-\Delta n_B = 2 \times \frac{n_B}{n} \times \frac{n_C}{n + n_0} \times p_2 \quad (\text{A.9})$$

Equating Eqs. (A.7) and (A.9) gives

$$p_1 = \frac{k_2 n(n + n_0) \Delta t}{2N_A V}. \quad (\text{A.10})$$

### Appendix B: Equivalence of physical assumptions in the Gillespie and STOCHSIM algorithms

Gillespie rigorously derived his algorithm from what he called the *fundamental hypothesis* of stochastic chemical kinetics. To show that the STOCHSIM algorithm can also be derived from this same hypothesis, we can translate the expressions for the STOCHSIM probabilities (Eqs. 10.1 and 10.2) into the Gillespie formalism and show that in the limit  $\Delta t \rightarrow 0$ , it reduces to probability expressions derived directly from the *fundamental hypothesis*.

Gillespie's *fundamental hypothesis* states that the probability  $\pi$  of an elementary reaction  $R$ , occurring within the infinitesimal time interval  $\delta t$ , can be expressed as

$$\pi = hc\delta t \quad (\text{B.1})$$

where  $h$  is the number of distinct molecular reactant combinations for the  $R$  reaction, and  $c$  is its stochastic rate constant. As with the deterministic rate constant  $k$ , the stochastic rate constant  $c$  can be interpreted to account for the mean rate at which reactant molecules collide, and the "activation energy" required for the reaction to occur. The relationship between the two constants are such that if the effects of fluctuations and correlations in reactant concentrations can be considered negligible, the following holds true for uni- and bi-molecular reactions<sup>28</sup>:

$$k_1 \doteq c_1 \quad (\text{B.2})$$

for the unimolecular rate constant  $k_1$  and the unimolecular stochastic rate constant  $c_1$  (both with dimensionality  $\text{s}^{-1}$ ), and

$$k_2 \doteq N_A V c_2 \quad (\text{B.3})$$

for the bimolecular rate constant  $k_2$  (with dimensionality  $\text{M}^{-1}\text{s}^{-1}$ ) and the bimolecular stochastic rate constant  $c_2$  (with dimensionality  $\text{s}^{-1}$ ). Here,  $V$  is the volume of the reaction system and  $N_A$  is Avogadro's constant.

For a specific unimolecular reaction  $R_1$  with species  $A$  as reactant,  $h$  in Eq. (B.1) is simply the number of molecules of the reactant species ( $h = n_A$ ). For

a specific bimolecular reaction  $R_2$  with species  $B$  and  $C$  as reactants,  $h$  is the product of the number of each species ( $h = n_B n_C$ ). Therefore, the probability of an  $R_1$  reaction occurring within  $\delta t$  is

$$\pi_1 = n_A c_1 \delta t, \quad (\text{B.4})$$

and the probability of an  $R_2$  reaction occurring within  $\delta t$  is

$$\pi_2 = n_B n_C c_2 \delta t. \quad (\text{B.5})$$

We now proceed to inspect how Gillespie's reaction probabilities ( $\pi_1$  and  $\pi_2$ ) are related to the reaction probabilities in STOCHSIM. In doing so, it is important to note that the uni- and bi-molecular reaction probabilities stored in STOCHSIM's look-up tables ( $p_1$  and  $p_2$  in Eqs. 10.1 and 10.2) are conditional probabilities, i.e. they are the probability of a certain reaction occurring given that its reactant molecules have been chosen by STOCHSIM in the current time-slice. However, Gillespie's reaction probability  $\pi$  is the probability of a given reaction occurring within *any* given time interval, so we must first obtain the equivalent quantities for STOCHSIM reactions.

For a unimolecular reaction with a look-up table probability of  $p_1$  in STOCHSIM, the probability  $\varpi_1$  of this reaction occurring in any given time-slice can be written as

$$\varpi_1 = p_{uni} \cdot \frac{n_A}{n} \cdot p_1 \quad (\text{B.6})$$

where  $p_{uni} = n_0/(n + n_0)$  is the probability that a unimolecular reaction is tested for in each time-slice. Similarly for a bimolecular reaction whose look-up table probability is  $p_2$ ,

$$\varpi_2 = p_{bi} \cdot 2 \cdot \frac{n_B}{n} \cdot \frac{n_C}{n} \cdot p_2 \quad (\text{B.7})$$

where  $p_{bi} = 1 - p_{uni} = n/(n + n_0)$  is the probability that a bimolecular reaction is tested for in each time-slice. Using  $p_{uni}$  and  $p_{bi}$ , Eqs. (10.1) and (10.2) can be rewritten as

$$p_1 = \frac{nk_1 \Delta t}{p_{uni}} \quad (\text{B.8})$$

and

$$p_2 = \frac{n^2 k_2 \Delta t}{2N_A V \cdot p_{bi}}. \quad (\text{B.9})$$

Substituting (B.8) into (B.6) and (B.9) into (B.7) yields the following expres-

sions for  $\varpi_1$  and  $\varpi_2$ :

$$\varpi_1 = n_A k_1 \Delta t \quad (\text{B.10})$$

$$\varpi_2 = n_B n_C \frac{k_2}{N_A V} \Delta t \quad (\text{B.11})$$

If the assumptions made in obtaining Eqs. (B.2–B.3) are valid, we may further substitute (B.2) into (B.10) and (B.3) into (B.11), and take the limit  $\Delta t \rightarrow 0$  to obtain

$$\varpi_1 = n_A c_1 \delta t = \pi_1 \quad (\text{B.12})$$

and

$$\varpi_2 = n_B n_C c_2 \delta t = \pi_2. \quad (\text{B.13})$$

We see, therefore, that the reaction probabilities for uni- and bi-molecular reactions calculated by the Gillespie and STOCHSIM algorithms are equivalent.