Methods

A. Handling diffusion and reaction microscopically

Inputs to simulation programs are usually macroscopic measurable parameters such as the reaction rate constants and the diffusion coefficient of each biomolecular species. These macroscopic parameters describe the behaviors of large groups of molecules. However, Monte Carlo simulation requires the microscopic behavior of each individual molecule. We developed a theory to bridge macroscopic parameters with microscopic parameters.

First, the motion of molecules was modelled. Macroscopically, molecules diffuse in the direction of the concentration gradient with diffusion coefficient *D*. From a microscopic perspective, this is due to the random-walk motion of each molecule. The direction of each random walk step was isotropic and was defined by a spherical coordinate:

$$\begin{aligned} \theta &= rand,\ rand \in [0, 2\pi) \\ \phi &= \arccos(1 - rand),\ rand \in [0, 2] \end{aligned} \tag{1}$$

where *rand* was a random number in the above specified range. The random walk step length *r* had a radial Gaussian density profile and was described by:

$$\begin{aligned} r &= F^{-1}(rand),\ rand \in [0,1) \\ \text{where } F(r) &= erf\left(\frac{r}{\sqrt{4D\Delta t}}\right) - \frac{r}{\sqrt{\pi D\Delta t}}\exp\left(-\frac{r^2}{4D\Delta t}\right) \end{aligned} \tag{2}$$

where *erf* was the error function, *D* was the macroscopic diffusion coefficient, and Δt was the simulation time step duration. This model ensured that the simulated diffusion

coefficient always agreed with the input diffusion coefficient regardless of the Δt chosen for the simulation.

Second, in dealing with biomolecular reactions, they were divided into two categories: association and dissociation. In association reactions, two reagents reacted to form one product with rate constant k_a:

$$A + B \xrightarrow{k_a} C$$

It was not necessary to include more than two reagents in association reactions because based on the collision theory of reaction, it was physically impossible for three molecules to collide at the same moment. This was not a limitation of the program, as any biomolecular reaction can be rewritten as a set of association reactions involving two reagents, for instance:

$$A + B + C \rightarrow D \Rightarrow \begin{cases} A + B \rightarrow AB \\ AB + C \rightarrow D \end{cases}$$

This was instead an advantage because it provided flexibility to capture the kinetics of reaction intermediates (AB, in the above example). The probability of an association reaction between the two molecules depended on their collision probability and the reaction rate constant.

$$P_a = 1 - \exp\left(- \frac{k_a \left(1 - erf\left(\frac{l_0}{\sqrt{4(D_A + D_B)\Delta t}} \right) \right)}{4(D_A + D_B)\pi N_A l_0} \right) \tag{3}$$

The product molecule C was created on the line segment joining the two reagent molecules A and B. Its location on the line was such that the ratio of distance AC to CB was proportional to their average random walk step length.

In contrast, dissociation reactions took the form of one reagent and one or more products with a rate constant of k_d:

$$A \xrightarrow{k_d} B$$
$$A \xrightarrow{k_d} B_1 + B_2$$
$$A \xrightarrow{k_d} B_1 + B_2 + B_3 + \ldots$$

The probability of dissociation reactions was defined by:

$$P_d = 1 - e^{-k_d \Delta t} \quad (4)$$

In order to prevent instantaneous association reaction in the following time step, each of the product molecules took one step of random walk away from the reagent location where the reaction occurred.

B. Implementations

The MBS package was written in standard C++ and compiled using GCC3.4 (and above). It consisted of a program for simulating biomolecular reactions and a program for visualizing the results. For the simulation of reaction networks, the simulation program was controlled from command line by three setup scripts describing the reaction, the geometry and the experiment time course, respectively:

- Molecule and reaction script defines:

- Molecule information including diffusion coefficient and molecule size
- Reaction information including chemical formula and reaction rate constants
- Setup parameters such as pH, temperature, viscosity, total time duration, and time step duration

- Reaction geometry script defines the geometry of the reaction volume using a combination of simple geometries including spheres, spherical shells, boxes, box shells, cylinders, and cylindrical shells.

- Time course script allows one to:
  - Add or removing any molecule at arbitrary time step
  - Change environmental variables such as the temperature at arbitrary time step

The purpose of the scripts was to avoid modification and recompilation of the source code for different experiments. The simulation process was divided into equal duration time steps specified in the setup scripts (Figure M1). In each time step, all molecules were randomly moved according to equation (1) and (2). Molecules were selected randomly to ensure no molecules react statistically earlier than others to avoid biasing their probability of reaction. For each selected molecule, the type of reactions it can undergo was chosen in random order as well to avoid biasing toward certain types of reactions. Furthermore, if the chosen reaction was an association reaction, reaction partners were chosen at random to again avoid biasing their probability of reaction. The coordinates of each molecule and the concentration of each type of molecule were

recorded in data files for each time step. For the visualization of the results, the program used OpenGL to render a 3D movie of the molecular movements and reactions. The molecule concentrations over time were stored in a separate data file, which could be imported by standard data analysis software to analyze the reaction kinetics.

C. Diffusion and basic reaction kinetics

To test our diffusion model, 10^4 non-interacting molecules were initially placed at the location of one point and allowed to diffuse over time. Then, the molecule distribution in our simulated diffusion was verified to be correct by comparing the result to the deterministic solution. Next, the diffusion process was verified to be independent of Δt by modeling identical diffusion processes using $D=10^{-10}$ m^2s^{-1} under different Δt ranging from 10^{-5} s to 10^{-3} s.

To test simple reaction kinetics, association and dissociation reactions were simulated using parameters that resemble conditions inside cells: biomolecules were homogeneously distributed in a spherical volume with a radius of 1 μm; the concentration of biomolecules was varied in the range of 100 nM to 10 μM; association rate k_a was varied from 10^4 to 10^6 $M^{-1}s^{-1}$; dissociation rate k_d was varied from 1 to 100 s^{-1}. Lastly, the resulting kinetics curves were compared with the deterministic solutions.

D. The predator-prey model

The oscillator was simulated using the following biomolecular reactions with association and dissociation parameters that ensured an oscillation would occur:

$$A + X \xrightarrow{k_{a1}} 2X \qquad k_{a1}=2\times10^{5}M^{-1}s^{-1}$$
$$X + Y \xrightarrow{k_{a2}} 2Y \qquad k_{a2}=2\times10^{6}M^{-1}s^{-1}$$
$$Y \xrightarrow{k_{d}} D \qquad k_{d}=200s^{-1}$$

The total simulation duration was 1 s with a Δt of 0.1 ms. The diffusion coefficients D of all molecular species were varied in the range of 10^{-10} to 10^{-12} m^2s^{-1}.

E. Prokaryotic and eukaryotic genetic oscillator

Genetic circuits for both prokaryotic and eukaryotic cell were constructed using the following parameters: inhibitor-DNA complex dissociation constant: 10^{-7} M, mRNA synthesis rate: 500 s^{-1}, protein synthesis rate: 500 s^{-1}, mRNA degradation rate: 50 s^{-1}, protein degradation rate: 50 s^{-1}, copy number of plasmid DNA: 50. The total simulation duration was 3 s with Δt of 1 ms. Both the prokaryotic and eukaryotic cell had a spherical volume with a radius of 2 μm. In the case of the eukaryotic cell, it had a spherical nucleus with a radius of 0.75 μm, which contained the DNA.

F. Ca^{2+} wave

A cylindrical compartment with a radius of 0.1 μm and a length of 10 μm was constructed. The membrane of the compartment was embedded with Ca^{2+}-dependent Ca^{2+} channels (CDCC) at a range of concentrations. These CDCCs bound to outside Ca^{2+} which opened the channel, releasing Ca^{2+} inside the compartment. The flux of Ca^{2+} depended on the concentration gradient across the membrane, similar to the IP_3 receptor proteins on the surface of endoplasmic reticulum (ER). In our simulation, the Ca^{2+} concentration inside the compartment was much higher than outside, hence, when Ca^{2+}

induced the initial release of Ca^{2+}, these Ca^{2+} ions bound to more CDCCs and released more Ca^{2+} in a positive feedback loop. The Ca^{2+} concentration inside the compartment was similar to the physiological values inside ER ~500 μM.

G. Availability and requirements

Name: Monte Carlo Biomolecular Reaction Simulator

Web address: http://individual.utoronto.ca/ktruong/software.htm

Operating system requirement: Windows 2000/XP, can be recompiled in Linux

Programming language: C/C++

License: Open-source and free for academic and non-profit use only

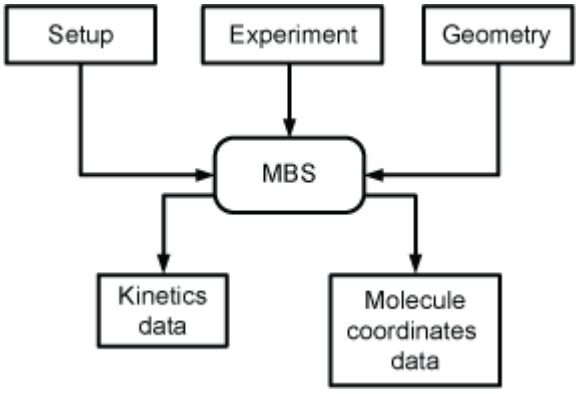

Figure M1 - The input and output files of MBS.

MBS required three separate input scripts (i.e. setup, experiment and geometry scripts) and generated two data files (i.e. kinetics data and molecular coordinates data).