Chapter 2

Methods

2.1 Simulation

There exist various types of simulators for different biochemical reaction systems. Conventionally, biochemical reaction systems are described using concentrations as state variables and treated with ordinary differential equations (ODE). This approach is deterministic. It follows a well-mixed hypothesis, assuming either the molecules are in spatially homogeneous or their diffusions occur much faster than the reactions. Such an assumption is suitable for models without spatial concerns, for example, gene networks, and metabolic networks. COPASI is a software widely used for the ODE models.

Deterministic spatial models can be described in reaction-diffusion equations. These equations are essentially partial differential equations (PDE) that consist of the reaction terms and diffusion terms. The Virtual Cell is a software for simulating the reaction-diffusion systems. For stochastic models, in which state variables are discrete molecule numbers instead of continuous concentrations, the Gillespie algorithm, and its variations are used widely for simulations. The algorithm is exact regarding comparing to chemical master equations. It becomes computationally demanding when simulating large reaction systems.

When models are both stochastic and spatial, there are a few simulation approaches with varying spatial resolutions. Some divide the whole volume and perform Gillespie algorithm in

each subvolume. However, the way of dividing volume affects the simulation accuracy. Programs including MesoRD are for this type of simulations. A more accurate approach is to treat molecules individually as particles. In this case, space is not discretized but continuous. Molecules are particle-like objects. Their diffusions follow a statistics of Brownian motion and their reaction events depend on molecular collisions. Softwares such as Smoldyn and MCell are particle-based simulators.

2.1.1 Smoldyn

For the model described here, we have chosen to use Smoldyn. Smoldyn is a stochastic simulator for discrete time continuous space systems. It simulates the Brownian motion of molecules and diffusion-limited chemical reactions. The molecules are point-like and do not take up physical space. Details such as rotation, orientation, intermolecular forces are not considered. The particle-based approach in Smoldyn provides a microsecond level temporal resolution and a nanometer level spatial accuracy. In Smoldyn, only molecules of interest are explicitly included in the simulation system. Although solvent and other molecules are assumed to collide constantly with the molecules of interest, their effects are treated implicitly as affecting the diffusion of molecules in the system.

Smoldyn uses a fixed time step. An appropriate time step value is selected so that in one time step the probability of having the fastest first order reaction occur is at most 0.1. With this assumption, during a time step, whether a first order reaction occurs or does not occur can be viewed as a binary event, which follows a Poisson distribution. For three-dimensional isotropic Brownian diffusions, a mean square distance is calculated for each diffusible molecule for each spatial coordinate following the equation $MSD = \sqrt{2 \times r.v. \times dt}$, where r.v. is short for random variable and dt is time step. The MSD is calculated at each iteration and added to each diffusible molecule.

The reactions built into the Smoldyn simulator are mostly elementary reactions and can be at

most second order. First order reactions such as unbindings and (de)phosphorylations are modeled as Poisson processes with event rate equals the corresponding rate constant parameter measured from experiments. Probability is calculated for each reaction using equation $Pr = \exp(-rate \times dt)$. A random variable is generated and compared to the Pr value as a criterion to decide whether to proceed the particular reaction (r.v.<Pr). Second order binding reactions in Smoldyn are diffusion-limited, which means the binding step takes place fast once the reactants collide with each other. This concept is defined as below,
A + B $\xrightarrow{K_D}$ (AB)* $\xrightarrow{K_a}$ AB, $K_a >> K_D$.
Where K_a is the activation rate that allows bindings to complete, and K_D is the rate that two reactants diffuse together and collide. This condition is a reasonable assumption for reactions in solution phase such as intracellular environments, where lots of reactions involve enzymes can lower the activation energy and allow reactions to proceed fast.

Since molecules in Smoldyn do not possess physical volume, collisions are assumed to occur as long as two reactants are sufficiently close. The criterion for the closeness is a binding radius, which is calculated in Smoldyn based on simulation time step, reactants diffusion coefficients, the kinetic rate constant, and the possible reversal unbinding reactions. Reactants within the binding radius are allowed to bind.

The workflow of Smoldyn within a time step is as follow:
diffusion $\rightarrow$ surface interaction check $\rightarrow$ reactions $\rightarrow$ surface interaction check $\rightarrow$ imulation time update

Molecules are grouped into lists, which can be expanded by users. One default list, the diffusion list, contains all diffusible molecules. At each time step iteration, the simulator loops over the diffusion list and updates each molecule position by computing a random diffusion step. Then the simulator checks surface interactions for molecules with updated locations in case molecules move across surfaces that are reflective, absorptive, or jumping. Reactions simulated include zeroth order reactions, first order reactions, and second order reactions. In Smoldyn, it is arbitrary that lower order reactions are examined before higher ones. This arrangement

introduces bias. The simulator processes reactions by checking molecules in lists. If a molecule is involved in a zeroth or first order reaction, the reaction probability is used to decide whether the particular reaction occurs or not. For second order reactions, the criterion for a reaction to occur is the binding radius. Binding or unbinding may change a molecule's position. After the reaction stage, the simulator checks the surface interaction again for molecules that changed locations.

2.1.2 Modifications to Smoldyn

Our work requires modeling reactions between complex molecules, and this leads us to develop more sophisticated molecule and reaction features than available in the original Smoldyn. In our customized version, molecules are not just point-like objects. Complex molecules, such as CaMKII, consist of multiple subunits which can have independent reactions. Consequently, we have modified Smoldyn so that subunits take distinct physical locations. For diffusion, the position change is calculated for one subunit, and the other subunits shift their positions accordingly.

Each complex has multiple subunits or molecules, and each molecule has at least one binding sites (Figure 2.1). The states of binding sites are implemented as bit vectors. Value 1 means a site is bound or phosphorylated; whereas a value 0 means a site is unbound or unphosphorylated. Hence, each molecule can have $2^{sites\,num}$ states. Depending on the binding situation of the reactant(s), there can be a set of possible reaction choices, which are chained in an arbitrary sequence. The customized simulator picks one reaction from the sequence randomly and checks the reactant(s) against the probability (for first order reactions) or the binding radius (for second order reactions). If the picked reaction gets rejected, the simulator continues to the next reaction in the sequence until all reactions are examined or one passes the check. Using random picking, we avoid the potential bias caused by a fixed sequence of choices.

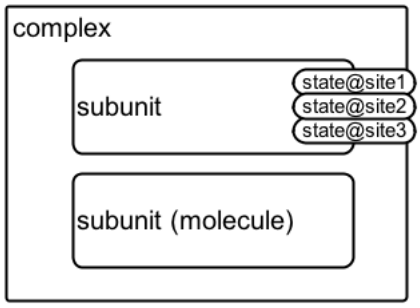

Figure 2.1: A complex that consists of two subunits, each of which contains three sites

To illustrate, suppose we have a calmodulin molecule with no calcium ions bound (apoCaM) and one nearby calcium ion can interact with either an N lobe site or a C lobe site of calmodulin. The two reaction choices can be expressed:

```
cam{N1==0, N2==0, C1==0, C2==0, Ng==0, K==0, Kp==0} + ca{cam==0}
->cam[N1=1]~ca[cam=1]
cam{N1==0, N2==0, C1==0, C2==0, Ng==0, K==0, Kp==0} + ca{cam==0}
->cam[C1=1]~ca[cam=1]
```

In the equations, {} represents the states at relevant binding sites and also can be considered as the precondition for the reaction to occur; + represents binding; ~ is a bond formed between molecules; [] codes for sites that are updated in products. The bit vector of the binding states of the calmodulin molecules is '0000000', which equals 0. This binding state value is paired with the identity of the calmodulin molecule, resulting in the first key. Likewise, a second key value can be calculated for the calcium ion. The simulator uses the key pair to look up a reaction choice in a pre-loaded reaction table, which is a hashtable with a key pair input. The lookup result is the sequence of two reactions which are listed above. Then one reaction is picked randomly. Since this reaction is second order, if the distance between the reactants is less than the binding radius, the two reactants bind accordingly, with sites involved in the binding updated to new values by updating the bit vectors. Otherwise, the simulator continues to the next reaction in the sequence. To speed up simulations, the model space is pre-divided into unit boxes during model initiation. Only molecule pairs located in the same box or neighbor

boxes are considered as nearby for reaction checking.

2.1.3 Categories of reactions

In the original Smoldyn, two reactants bind and collapse into one product. The reactants are removed from the system, and a product molecule is generated anew. As a result, the product carries a different serial number, making it difficult to log the reaction history of a particular molecule. In the customized version, all reactions and occur at the binding site level, defined for an individual molecule or subunit. When two reactants bind, we use pointers to connect the molecules at the involved binding sites. Their physical positions overlap, point to each other and become synchronized. Therefore, they are bound without being replaced by newly generated molecules, making the simulator possible to track the reaction history of each molecule. When two molecules are bound together, one is considered as a master molecule, and the other is an attached "slave." The master molecule determines what diffusion coefficient of the complex. For example, when Ca2+ binds to CaM, CaM is the master molecule, and the Ca2+/CaM diffuses at the same speed as CaM.

Reactions are categorized by the number of reactants involved and products generated. Below are the reaction types included in the model.

1) `A->B`

An example reaction is that a phosphorylated CaMKII subunit without CaM bound becomes dephosphorylated, e.g.,

```
camkii{P==1,cam==0} -> camkii[P=0].
```

Where P stands for the phosphate binding site and cam stands for the CaM binding site of bK. Assuming abundant phosphate molecules are around, so phosphate are treated implicitly, and the reaction is unimolecular and first order.

2) `A~C->B~C`

An example is a CaM-bound CaMKII subunit becomes dephosphorylated. Because a

CaM-bound CaMKII subunit dephosphorylates at a different rate from an unbound one, this reaction is listed seperately, e.g.,

```
camkii{cam==1,P==1}~cam{K==1,Kp==1} -> camkii[cam=1,P=0]~cam[K=1,Kp=0].
```

In contrast to the first example, the reaction here is still first order but involves two molecules. The bound CaM is essential because the site named as Kp on CaM is needed to update the state. This site reflects the phosphorylation state of the CaMKII subunit bound to the cam. It is important to have this artificial Kp site because whether the bound CaMKII subunit is phosphorylated or not affects other reactions that involve the CaM. For example, a CaM bound with a phosphorylated CaMKII subunit interacts differently with a free ca2+ than a CaM attached to an unphosphorylated CaMKII subunit.

Another particular case belonging to this category is the autophosphorylation reaction between CaMKII subunits, in which a subunit bound with a CaM may become phosphorylated if its neighbor subunit gets bound or phosphorylated. The reaction rules are described as below:

```
camkii{cam==1, P==0, n->P==1}~cam{K==1, Kp==0} -> camkii[P=0]~cam[Kp=1]
camkii{cam==1, P==0, n->cam==1}~cam{K==1, Kp==0} -> camkii[P=0]~cam[Kp=1].
```

The `n->P` site and `n->cam` site of CaMKII are defined to reflect the binding and phosphorylation status of the neighbor subunit. The Kp site of cam is also updated in the reaction to track the phosphorylation status of the camkii.

3) `A~B -> A + B`

An example is a CaM with only the N1 site bound with a Ca2+ and the Ca2+ unbinds from this site, e.g.,

```
cam{N1==1, N2==0, C1==0, C2==0, Ng==0, K==0, Kp==0}~ca{cam==1}
-> cam[N1=0] + ca[cam=0].
```

After unbinding, the physical locations of the two reactants separate from each other. The binding site that previously pointed to the other molecule becomes a null pointer.

4) `A+B -> A~B`

An example is that one used previously describing an apoCaM binding to a free Ca2+ at the N1 site, e.g.,

```
cam{N1==0, N2==0, C1==0, C2==0, Ng==0, K==0, Kp==0} + ca{cam==1}
-> cam[N1=1]~ca[cam=0].
```

After binding, the two reactants' physical locations overlap and become synchronized.

Among the four reaction types, type I-III are first order and type IV is second order. Except for type I that involves one reactant, all the rest involve two reactants. Other types of reactions shown in the original Smoldyn but not listed above include A+B-¿C+D, which is suitable for gas phase reactions. For solution phase as in the intracellular environment, the solvents tend to cage reactants and allow them staying bound instead of quickly split up. Michaelis-Menten reactions in the form `A+B<->A~B->A+C` are divided into three separate reactions: `A+B->A~B, A~B->A+B, A~B->A+C`.

2.1.4 Workflow of the customized Smoldyn

The general simulation workflow in the customized Smoldyn is the same as the original version. The difference is that we shuffle the molecule lists at every iteration. Since reactions are not categorized according to the number of reactants, it is not necessarily true that first order reactions should get simulated before second order reactions. In our approach, the simulator loops through the molecule lists and for each molecule, Type I reactions proceed first. When it comes to type II-IV reactions which require two molecules, the simulator examines the unit space (a.k.a. molecule box in Smoldyn) that contains the molecule and select another molecule, which is either the same box space or a neighboring one, to guarantees the two molecules are sufficiently close. The selected two molecules may already be bound. Then the simulator looks up possible reactions depending on the binding states of the molecule pair.

Also, at most one reaction is allowed during a time step between one pair of binding sites. For example, the simulator checks every molecule in lists in turn. If a CaM gets examined first and its N1 site binds an unexamined Ca2+ located nearby, both binding sites involved in the reaction become no longer reactive in the time step. This restriction prevents a reverse unbinding between the two molecules later when the bound Ca2+ gets examined. However, for a multi-site molecule, more than one reaction can occur within a time step. After every reaction, the reactant molecules update their binding states, so that repeat binding events are impossible. The arbitrary sequence in which molecules get examined may seem to introduce bias since molecules located first get examined earlier and then affect the available reaction choices for other molecules. For example, if a CaM binds a CaMKII first, then the molecule will interact with Ca2+ with higher affinity. However, we tried to minimize the bias by shuffling the molecule list at every iteration. So the arbitrary "advantage" for the precedent molecules does not accumulate.

Because binding sites are defined for molecules, a molecule can have more than one state. When searching possible reactions for the reactant(s), it is necessary to know not only the molecule identity but also the binding state. To illustrate, in the original Smoldyn, camN1C0 is considered as a molecules species with identity camN1C0; whereas in the customized Smoldyn, it consists of two molecules: a CaM with identity cam and binding state value 1, and a Ca2+ a with identity ca and binding state 1. The reaction network is implemented as a hash table using a key pair with molecule identity and binding state.

The hash table implementation followed a `g_hash_table` package included in the GLib library, an open source library. The hash table approach also enables fast loading and expanding the reaction network. Reaction rules are read in by lines. Depending on how specifically the rules are described, each line of a reaction can be expanded into multiple elementary reactions. Since some binding sites can work independently from each other, rules can be described for necessarily relevant binding sites, without all possible combinations being specified. Instead, all possible combinations are expanded during loading the rules into the hash table. For example, for a bare cam binding with ca at an N1 site, it is not necessary to specify the states of C1 and

C2 sites. However, when parsing the reaction rule, the simulator automatically calculates all possible states of the involved reactant states. In this case, cam can be:

```
N1==0, N2==0, C1==0, C2==0, K==0, Kp==0, Ng==0
N1==0, N2==0, C1==0, C2==0, K==0, Kp==0, Ng==0
N1==0, N2==0, C1==1, C2==0, K==0, Kp==0, Ng==0
N1==0, N2==0, C1==1, C2==1, K==0, Kp==0, Ng==0
```

The syntax here allows a "fuzzy" yet compact way to describe a reaction network, therefore avoids an issue of "combinatorial explosion", which is a common problem in systems biology when modeling and simulating macromolecules. In eukaryotic cells, macromolecules contain multiple binding sites and exhibit various binding states. These molecules are often modeled as multi-state systems. Depending on how the binding sites cooperate with each other, the total combinations of states can grow exponentially, resulting in problems to represent the states of a macromolecule ("the specification problem") and to simulate the time evolution of the states ("the computation problem"). A common way to handle this problem is to follow a rule-based approach, which allows reaction rules to be described implicitly and avoid enumerating all possible combinations.

2.2 Modeling

Only the soma part of the neuron was included in the model, assuming that dendrites and axons do not affect the signaling events in the cell body. By ignoring the dendritic processes and the axon, we considered the voltage elevation have already propagated to the soma via dendrites, and therefore the model focuses on the downstream events.

2.2.1 Compartments

A pyramid geometry (Figure 2.2) was used to represent a randomly selected part of a spherical-shaped cell. The base of the pyramid represents the cell membrane, and the tip represents the center of the nucleus. The model consists of three compartments: the nanodomain, the cytosol, and the nucleus. Each compartment contains a distinct set of

molecule species and reaction rules. The concentrations and diffusion properties of these molecules also vary with compartments.

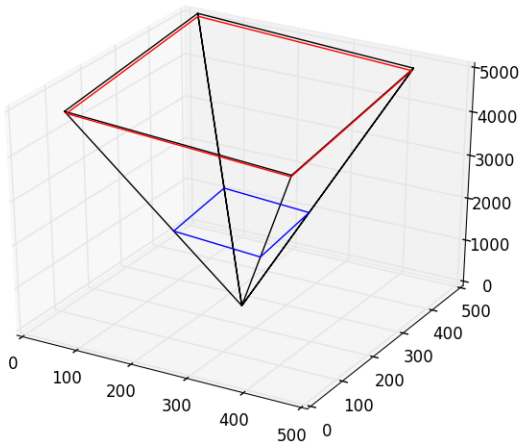

Figure 2.2: The pyramid geomeotry of the model. The red square represents the intracellular nanodomain boundary and the blue square represents the nucleus membrane. Scale units are nm.

2.2.2 Molecules

Molecules involved in the model are as follows: calcium ions (Ca2+), calmodulin (CaM), Ca2+/CaM kinase II (CaMKII), neurogranin (Ng), Ca2+/CaM kinase kinase (CaMKK), Ca2+/CaM kinase IV (CaMKIV), protein kinase A (PKA), protein phosphatase 1 (PP1), protein phosphatase 2A (PP2A), and cAMP response element-binding protein (CREB). Ca2+, CaM, and CaMKII are located in all three compartments; Ng molecules attach to the cell membrane and are present only in the nanodomain; all the other molecules exist exclusively in the nucleus. Table 2.1 lists the detailed distribution of these molecules is listed regarding molecule numbers and concentrations at the initial condition.

We focused on how Ca2+ signals affect the nuclear CREB phosphorylation. The process can be divided into two stages: Ca2+ signal transportation and nuclear phosphorylation of CREB by Ca2+/CaM. In the transportation step, according to Ma et al.(2014), Ca2+ ions and CaM are

the cargo molecules; CaMKII, in particular, the gamma subtype, are suggested as the shuttle molecules. We included a high concentration of immobilized Ng in the nanodomain, assuming that Ng plays a modulatory role in the signaling cascade by capturing apoCaM in the nanodomain close to the Ca2+ channels and gradually release them when the Ca2+ influx starts, therefore providing a reservoir of Ca2+ buffer near Ca2+ entrance. To make the model necessarily simple, we did not differentiate the subtypes of CaMKII. All CaMKII are assumed to be able to bind Ca2+/CaM, autophosphorylate and translocate to the nucleus.

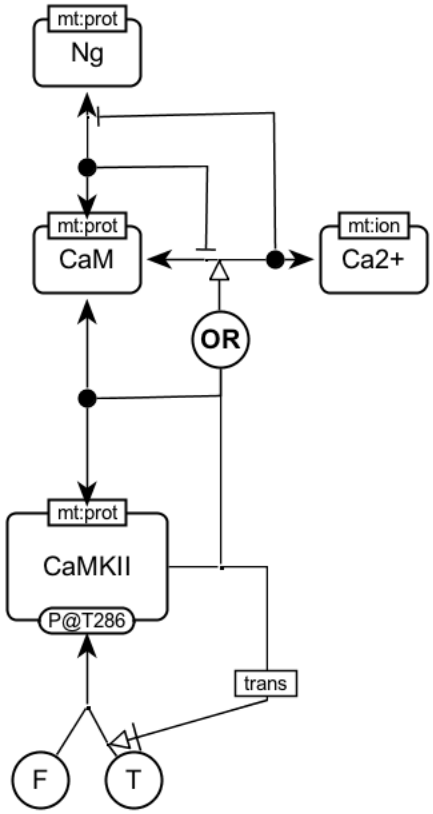

Figure 2.3: SBGN Entity Relation diagram shows the interactions between Ca2+, CaM, CaMKII and Ng.

Except for CaMKII, all other molecule species are monomers. We modeled CaMKIIs as holoenzymes in a two-ring structure, similar to their natural composition. Both the top and the bottom ring consists of 6 subunits. Based on the estimation from a study by Gaertner et al.(2004), we used 8 nm for the inter-subunits distance and 10 nm for the space between two

rings.

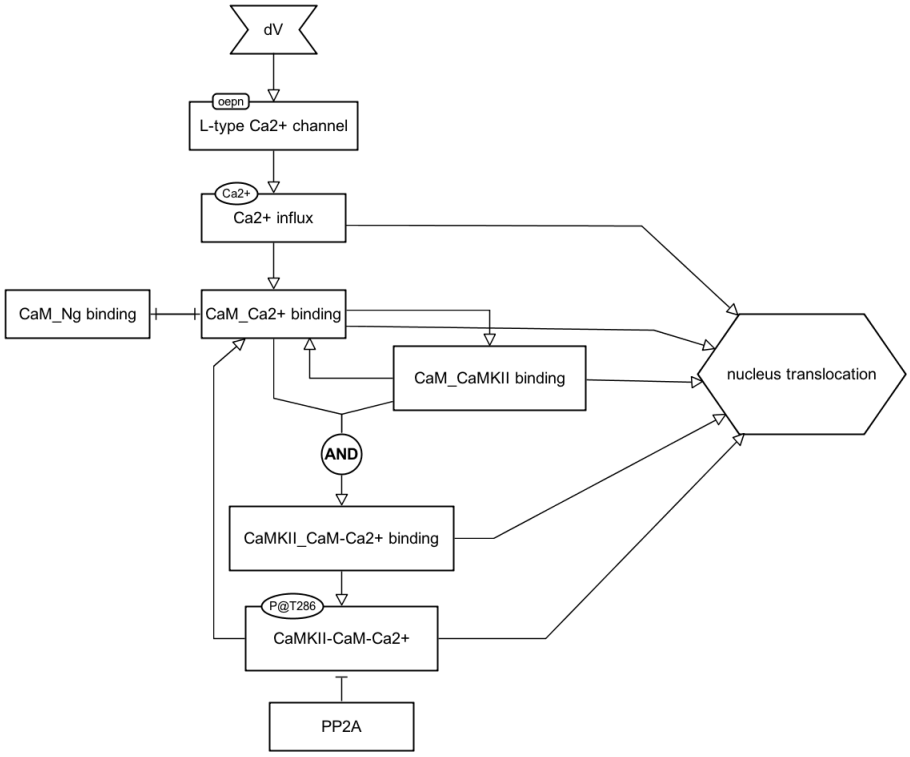

Figure 2.4: SBGN Activity Flow diagram shows the events following Ca2+ influx in the cytosol.

Since all molecule species in the model participate in chemical reactions, each molecule possesses at least one binding site. In particular, CaM can interact with Ca2+, Ng, CaMKII, CaMKK and CaMKIV, therefore playing an essential role in the cascade. Each CaM consists of four Ca2+ binding sites (two on the N-lobe and two on the C-lobe), one Ng binding site, one binding site to interact Ca2+/CaM kinases such as CaMKII, CaMKK, and CaMKIV. It also contains a functional site to reflect the phosphorylation state of the bound CaMKII, in case bound to a CaMKII. Each CaMKII subunit consists of CaM binding site, a phosphorylation site, and two other functional sites to reflect the binding and phosphorylation states of its next neighbor subunit. Bindings sites can be either where real binding reactions happen, or functional domains that reflect the state change of a molecule. There are multiple ways to define the binding sites for a molecule species. However, the more binding sites included, the

more a total number of states the simulator has to be handle. In general, the number of binding sites depends on the specificity of the corresponding reaction system.

In the second stage, Ca2+/CaM molecules that entered the nucleus carry on the downstream events to activate CaMKK and CaMKIV. The activated CaMKIV catalyzes the phosphorylation of CREB. Neither free Ca2+ ions nor apoCaM alone can initiate the cascade. Therefore, it is essential for Ca2+/CaM to enter in a bound complex form. Hypothetically, using CaMKII as shuttle proteins is advantageous in escorting Ca2+ and CaM traveling through the cytoplasm, where Ca2+ buffers and CaM binding proteins are potentially everywhere. Also when dropping Ca2+ and CaM in the nucleus, the two molecules are more likely to stay bound, therefore ready to be picked up by CaMKK and CaMKIV.

2.2.3 Ca2+ influx

Ca2+ influx files were generated from a multi-compartment neuron model, which was previously built using the NEURON simulator. The neuron model has detailed morphology (Figure 2.5), consisting of separate compartments for dendrites (898 apical and 228 basal compartments), 190 spines, 18 somatodendritic compartments, an axon hillock, an initial segment(AIS) and an axon with reduced morphology. Passive parameters such as the membrane resistance, the axial resistance, and the membrane capacitance are automatically fit with a parameter searching algorithm (Holmes et al., 2006). Active conductance was distributed depending on compartments. For example, I_h(hyperpolarization-activated K+ channels) are distributed only on dendrites. Various ion channels are involved, such as the delayed rectifier K+ current, A-type K+ current, fast Na+ current and persistent Na+ current. The distributions and conductance parameters of these ion channels were fit manually to recorded action potentials.

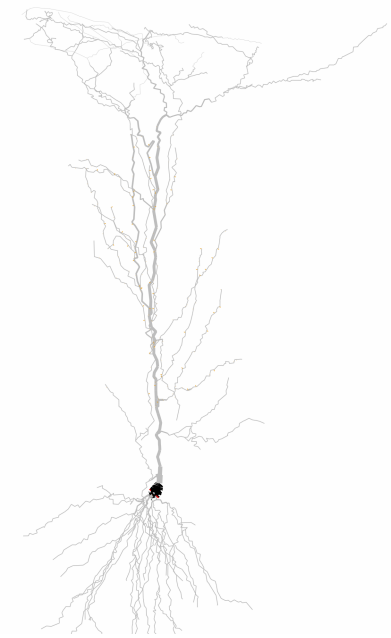

Figure 2.5: Detailed morphology of the modeled CA1 neuron. The grey part represents dendrites and the black represents soma.

Each spine contains NMDA receptors and AMPA receptors. The spines receive synaptic stimulations, which were generated using a probabilistic release file based an experiment recorded action potential failure rates on the postsynaptic neuron (Kim et al., 2012). Once synapses receive a stimulation, the AMPARs open, leading to excitatory postsynaptic potentials (EPSP) and then NMDARs open, letting inward Ca2+ currents to further elevate the dendritic membrane potential. The local dendritic depolarizations join with other dendritic potentials, propagate toward the soma and activate the low threshold Na+ channel on the AIS. Subsequently, action potentials start in the AIS, propagate backward to the soma and facilitate the voltage-dependent Ca2+ channels open, letting extracellular Ca2+ flow into the cell. We recorded the inward Ca2+ current density and converted the recording to the number of Ca2+ ions using the equation

$Ca^{2+}ions = I_{Ca^{2+}current} \times N_{avgadro} \times A_{surface} \times \frac{1}{Ca^{2+}cation} \times dt.$

$A_{surface}$ is the area of the membrane patch of the pyramid geometry, and dt is the simulation time step of the Smoldyn Ca2+ signaling pathway model. The converted Ca2+ influx file was fed into the Smoldyn model as an input stimulus.

We adjusted the parameters of the neuron model and generated various types of Ca2+ influx. One is a long-lasting tetanus stimulation which consists of uninterrupted trains of spikes, following the probabilistic presynaptic release. The other is a theta-burst stimulation, the same total number of spikes that were grouped into sets of 5 spikes with an interval roughly 200 ms between spike sets. The 200 ms interval gives rise of 5 Hz of stimulation frequency. The plasma membrane Ca2+ ATPase pump (PMCA) is included in the model to extrude some Ca2+ influx implicitly as an absorbing membrane. The extrusion rate was calculated by $pump\,rate = rate_{single\,pump} \times pump\,density \times dt$.

2.2.4 Reactions

In cells, there are numerous Ca2+ buffer proteins, such as calbindins (CB), parvalbumins (PV) and calretinins (CR). These buffers are present in varying concentrations depending on cell types (Schwaller 2010). For example, PV is prevalent in interneurons but not in primary neurons. Since we are interested in excitatory neurons, PV can be ignored in the model. Considering a minimal set of reactions, we did not include CB and CR either. Given the small volume of the model and the relatively short time scale (seconds level) that we are interested in, we assume that the absence of CB and CR would not affect the transportation process of Ca2+ ions. Thus in the model, only CaM is included as a Ca2+ binding protein. In general, CaM alone has a Ca2+ binding affinity within 5×10^7 M to 5×10^6 M, which falls in the range of intracellular Ca2+ concentration and is lower than regular Ca2+ buffers, and is thus usually considered as a Ca2+ sensor (Chin et al., 2000). However, multiple CaM binding proteins can modify the binding affinity of CaM, some of which can enhance the affinity to allow CaM stably sequester Ca2+ ions.

The N-lobe sites and the C-lobe sites of CaM bind with Ca2+ independently. However, within each lobe, the bindings are cooperative. CaM have multiple downstream substrates, including Ng, CaMKII, CaMKK, and CaMKIV. These molecules can interact with CaM loaded with varying number of Ca2+ ions. Table 2.2 shows the dynamic effects of Ca2+ loading on the

interaction between CaM and its binding proteins. Ng has the highest affinity with apoCaM. As a CaM picks up more Ca2+ ions, the binding between the Ng and the CaM decreases. In contrast, the interactions between CaM kinases and CaM become stronger as CaM gets loaded with more Ca2+. Also, the modifying effect of CaM binding proteins on CaM and Ca2+ affinity also varies. Attached with a CaMKII enhances the affinity of a CaM for Ca2+ binding. For example, a CaMKII bound CaM increases its second C-lobe site from a K_D of 400 nM to 120 nM. In contrast, attaching to a Ng decreases the K_D to 500 μM.

In the model, CaMKII subunits can undergo four reaction stages: activation through Ca2+/CaM binding, autophosphorylation at Thr286 or Thr287 catalyzed by trans-subunit, Ca2+/CaM-independent activation (trapping) and deactivation through dephosphorylation. More detailed transition states such as autoinhibition at the resting level, or phosphorylation at Thr305 or Thr306 to block CaM binding, are not considered in the model.

Autophosphorylation between subunits is an essential step for CaMKII holoenzymes to carry Ca2+/CaM as shuttle proteins. The intra-holoenzyme interactions allow phosphorylation to propagate and cooperative activation among subunits. In brief, what happens during autophosphorylation is that when a subunit N binds a CaM, its phosphorylation site at Thr286 becomes exposed, and its kinase domain unfolds. With an open kinase domain, the subunit N is ready to phosphorylate the subunit N-1; in the meantime, if the other neighbor subunit N+1 is activated, the subunit N can be phosphorylated at Thr286. Phosphorylated CaMKII subunit gains higher affinity in binding CaM/Ca2+. Also, a CaM bound with a phosphorylated CaMKII subunit becomes more affinitive with Ca2+. After phosphorylation, even when Ca2+/CaM drops from the CaMKII subunit, the CaMKII subunit is still activated, leading to Ca2+/CaM independent trapping stage.

There have been discussions regarding whether βCaMKII (bK) and γCaMKII (gK) play different roles in the signaling cascade. Both subtypes can interact with Ca2+/CaM, but the only the gK can enter the nucleus due to its NLS tail. It has been observed that the bK subtype is faster at autophosphorylation, which leads to a hypothesis that bK phosphorylate gK and facilitate the

latter to carry Ca2+/CaMs to the nucleus. However, a few problems arise when modeling the hypothesized inter-holoenzyme interaction. Consider modeling the process using a Michaelis-Menten way with three separate steps: binding, unbinding and phosphorylation. Based on the kinetic parameters from the Pepke et al.(2010), the phosphorylation step is much slower than the unbinding rate `0.5 to 1.25` s^1 vs. `13 to 4500` s^1, implying that inter-holoenzyme phosphorylation does not often occur right after binding. Also in our trial simulations, the reaction rule leads to illegitimate bindings between CaMKII complexes and forming interconnected macromolecule clusters, simply due to a slower unbinding than binding. Though experiment studies have shown that CaMKII in dendritic spines can self-associate and form scaffold structures (Hudmon et al., 2005, Ashpole et al., 2010), such a process takes longer than the seconds time scale adopted in the simulation. Also, allowing the self-association add the complexity of the simulator. Currently, the simulator treats molecules as point-like objects without taking real volume. Self-association requires CaMKII complexes to be organized in a specific manner because volume enclosed by the two layered rings are excluded for other CaMKII subunits. Without simulating real volume, it is unlikely to prohibit CaMKII subunits diffusing across each other. Since it is arbitrary to add extra rules to exclude such questionable reactions, inter-holoenzyme phosphorylation is not included in the reaction network.

Reactions in the nucleus were from the model by Bhalla(ref). We used the same reaction kinetics but expanded the reaction rules. Since CaM contains four Ca2+ binding sites, we allow CaMKK and CaMKIV to interact with partially loaded CaM. Also, CREB can be phosphorylated by CaMKIV with CaM with at least 2 Ca2+ bound.

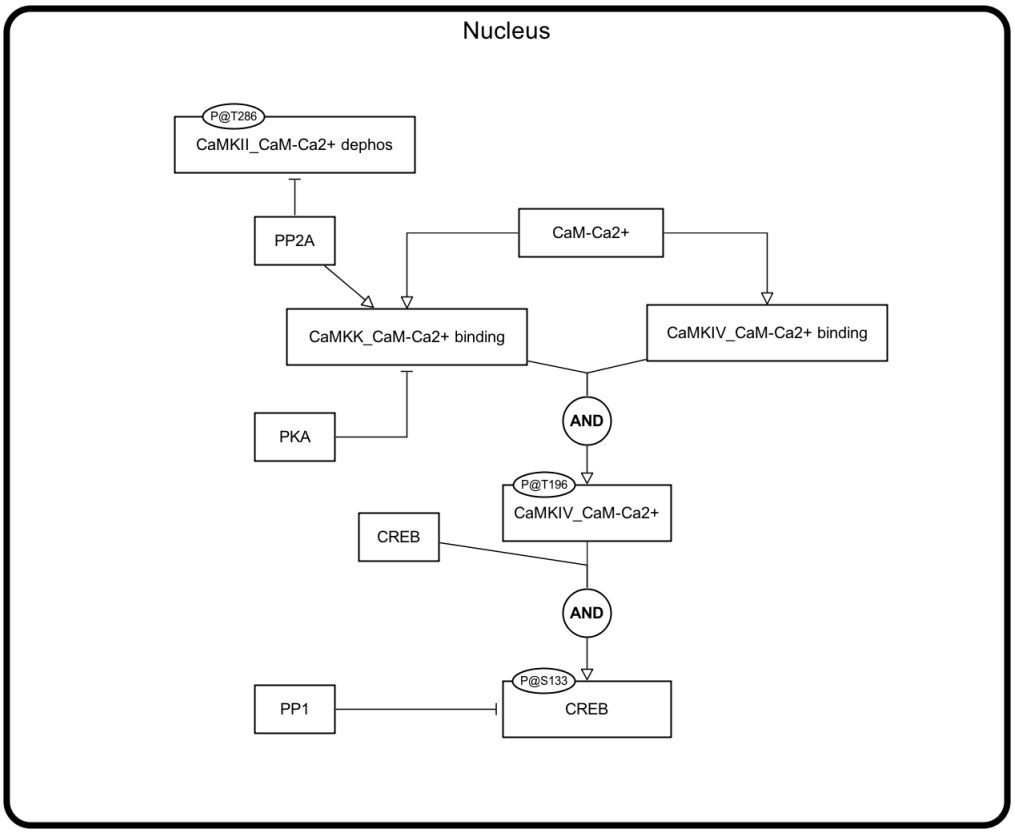

Figure 2.6: SBGN Activity Flow diagram showing the nucleus signaling events after molecules translocate from the cytosol.

2.2.5 Boundary conditions

All surfaces of the pyramid are reflective. We attempted to construct periodic boundaries for the four side surfaces but failed since these surfaces are not in parallel. We further tried to approximate the pyramid geometry by layering cube-shaped compartments with decreasing diameters. However, the different jumping distance at two consecutive compartments results in a problem for multi-subunits macromolecules. When two subunits of a macromolecule jump over the surfaces and change positions according to different jumping distances, the inter-distance between the subunits fails to keep constant. Although the reflective side surfaces may impose artifacts on lateral diffusions, we focus on analyzing the diffusion of the radial

direction toward the nucleus.

Surfaces between compartments are selectively transparent to allow some molecules to diffuse across. For example, the surface between the cytosol and the nucleus compartments are transparent to Ca2+, CaM, and CaMKII subunits but reflective all the nucleus resident molecules.

2.2.6 Initial Conditions

Simulations all start with a resting state, which is calculated using COPASI, software for modeling ODE-based chemical reactions. Initial conditions are calculated for each compartment separately. The compartment volume is entered, as well as the total fixed concentrations of Ca2+, CaM, and the reactions between them. For the nanodomain compartment, the concentration and reactions of Ng participated are also entered. Complex molecules such as bK and gK, as well as nucleus located molecules, are not entered for input, considering that their small numbers of copies or relatively slow kinetics have minor effects on the resting state.

2.2.7 Simulation time step

The simulation uses a fixed time step. A timestep should be small enough to resolve the fastest first order reaction without overrunning the whole simulation. The fastest first-order reaction the unbinding between CaM N1 site and Ca2+. The reaction has a rate constant 4.0 ms^{-1}, which on average equals a 0.25 ms interval. We use the $\frac{1}{10}$ of the length as the time step, allowing each interval approximately a probability of $1 - \exp \frac{1}{10}$.

2.2.8 Diffusion

Intracellular environment is highly crowded and packed with molecules of various sizes and geometries, leading to molecules fail to follow conventional Brownian diffusion either because of bumping into barriers or being unable to access soma excluded space. Theoretically, molecular crowding leads to volume exclusion, which distorts the space a molecule can effectively explore

for free diffusion. When measuring the mean squared displacement (MSD) of a single molecule vs. time, in Brownian diffusion, MSDt; whereas in anomalous diffusion, this linear relation does not hold and the diffusion coefficient is not a constant as in Brownian diffusion (Figure 2.2.6). Anomalous diffusion can be categorized into subdiffusion and superdiffusion. In cells, subdiffusion is much more common than superdiffusion.

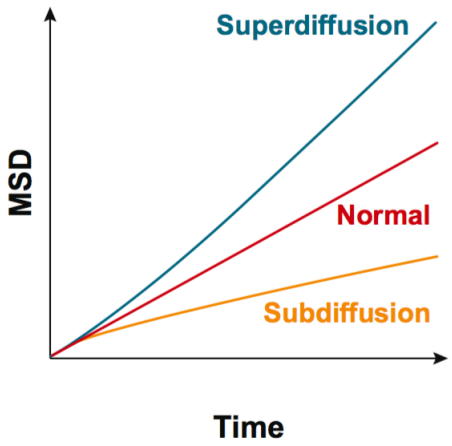

Figure 2.7: Three types of diffusions categorized by the relation between MSD and diffusion time. Figure from Dix and Verkman (2008) Figure 1a.

Molecular crowding-related subdiffusion is relevant to our model of Ca2+ signaling through the nanodomain. Tardross et al. (2013) indicated that Ca2+ ions might slow down by ten times in the nanodomain, either due to crowding or Ca2+ buffers. Large macromolecules such as CaMKII may also exhibit slow diffusion because of associating with F-actin when CaMKII are not activated. Li et al.(2016) showed that CaMKII become mobilized when Ca2+ ions enter from L-type Ca2+ channels, indicating a dissociation from F-actins.

However, it is not straightforward to simulate volume exclusion in the original Smoldyn. Smoldyn allows to add immobilized crowders, but extra details are necessary about the number and sizes of crowders. Besides, to simulate volume exclusion for each molecule by adding extra interacting rules is too computationally demanding. Some simulators are based on Green's Function Reaction Dynamics (GFRD), implement hard sphere for molecules and allow to simulate volume exclusion in an exact manner. Other relatively simplified simulators discretize

space into lattices and molecules are allowed to move on if its neighbor sites on the lattice are not completely taken. Also, there are phenomenological approaches to model subdiffusion by using for example continuous time random walk, which is not equivalent to the previously mentioned simulation methods.

In general, intracellular subdiffusion due to molecular crowding is still a debating topic. Dix and Verkman (2008) pointed out that a universal subdiffusion in the whole intracellular space is unlikely to be true. Instead, it is possible that molecules are slowed down in particular compartments, depending on the spatial heterogeneity of the compartment environment, and overall their MSD vs. stime exhibit a transient phase of diffusion coefficient shift (Di Rienzo et al., 2014).

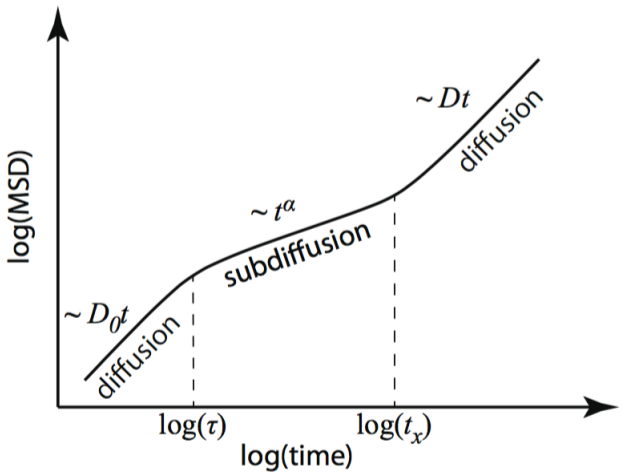

Figure 2.8: Schematic log(MSD) as a function of log(time) shows the idea of transient or intermediate subdiffusion phase. Figure from Holfling et al. (2013) Figure 1.

Based on the idea of compartment-dependent diffusion, we implemented this feature into our customized Smoldyn. We then varied the diffusion coefficients of interested molecules in different compartments. For example, inspired by Tardross et al.,(2013), we had a few simulations with Ca2+ ions slowed down in the nanodomain by ten times.