A Multiscale Algorithm for Spatiotemporal Modeling of Multivalent Protein-Protein Interaction

MD Shahinuzzaman[1,2] **and Dipak Barua**[1,*]

[1]Missouri University of Science and Technology, Chemical and Biochemical Engineering, Rolla, 65409, USA.
Phone: 1-573-341-7560, Fax: 1-573-341-4377

[2]msh95@mst.edu

[*]baruad@mst.edu

ABSTRACT

This paper introduces a multiscale framework for spatiotemporal modeling of protein-protein interaction. Cellular protein molecules represent multivalent species that contain modular features, such as binding domains and phosphorylation motifs. The binding and transformations of these features occur at a small time and spatial scale. On the other hand, space and time involved in protein diffusion, colocalization, and formation of complexes could be relatively large. Here, we present an agent-based framework integrated with a multiscale Brownian Dynamics simulation algorithm. The framework employs spatial graphs to describe multivalent molecules and complexes with their site-specific details. By implementing a time-adaptive feature, the Brownian Dynamics algorithm enables efficient computation while capturing the site-specific interactions of the diffusing species at the sub-nanometer scale. We demonstrate these capabilities by modeling two multivalent molecules, one representing a ligand and the other a receptor, in a two-dimensional plane (cell membrane). Using the model, we show that the algorithm can accelerate computation by orders of magnitudes in both concentrated and dilute regimes. We also show that the algorithm enables robust model predictions against a wide range of selection of time step sizes.

1 Introduction

A key challenge to spatiotemporal modeling of multivalent species is to deal the multiscale nature of the problem (Ayton et al., 2007). Multivalent protein-protein interaction is a central theme of cell signaling and gene transcription. Cellular protein molecules represent multivalent species with modular features, such as binding domains and phosphorylation motifs (Bhattacharyya et al., 2006). These features at the submolecular scale play critical roles in their mutual recognition and binding. The molecular-scale domains and motifs of the cellular proteins mediate interactions at very small spatial and temporal scales. In contrast, the protein molecules are subject to long-range diffusion in and across the cellular compartments (Weng et al., 1999).

Several models have been developed in the past to study multivalent antigen-mediated IgE receptor clustering in the plasma membrane of mast cells (Goldstein and Perelson, 1984; Hlavacek et al., 1999; Mahajan et al., 2014; Monine et al., 2010). However, these models were either equation-based deterministic models or developed using a non-spatial Kinetic Monte Carlo approach called network-free simulation (Yang et al., 2008; Yang and Hlavacek, 2011; Sneddon et al., 2011). The explicit geometric features of molecules and complexes were missing in these models.

In this paper, we introduce a multiscale Brownian Dynamics (BD) simulation algorithm that enables accelerated computation while capturing the sub-nanometer scale site-specific interactions of the molecules and their complexes. We develop a model of a trivalent ligand and a trivalent receptor molecule and their diffusion, rotation, and site-specific binding in the plasma membrane of a cell. The multivalent molecules and their site-specific features are described by spatial graphs. The mobile ligand-receptor graphs form larger graphs (multimolecular complexes) that evolve dynamically during the course of a simulation. The lateral and rotational motion of the molecules and evolving complexes are simulated using the BD algorithm. Using the model, we provide analysis on the computational performance and accuracy of our algorithm. We show that the algorithm speeds up computation by orders of magnitude without compromising the accuracy in both concentrated and dilute regimes. Moreover, we show that the time-adaptive feature enables robust predictions against a wide range of selection of time steps.

2 Methods

Below, we first provide details of the model. We then explain the multiscale Brownian Dynamics algorithm.

2.1 Model

The model is a reaction-diffusion model of multivalent ligand-receptor interaction in the cell plasma membrane. The model is developed using the agent-based approach (Gilbert, 2008). More details are provided below.

2.1.1 Cell membrane, Molecules, and Complexes

We consider a 1 μm^2 two-dimensional plane, which represents a small part of the cell membrane ($\sim$ 0.3% of the surface area of a spherical cell of 5 μm radius). We consider a trivalent ligand and a trivalent receptor molecule that we describe by spatial graphs (Grünert and Dittrich, 2010) (Figure 1). A spatial graph is a collection of small circles (nodes). The spatial organization of these nodes creates a coarse-grained structure of a molecule. When two or more such molecule graphs combine in a prescribed manner, they form a complex graph. Each ligand or receptor molecule contains three equally-spaced binding arms. In each binding arm, one node is designated as the reaction center (Figure 1). The reaction centers serve as the mutual recognition sites of the ligand and receptor molecules. In a complex, the molecules remain linked via these reaction centers (Figure 1C).

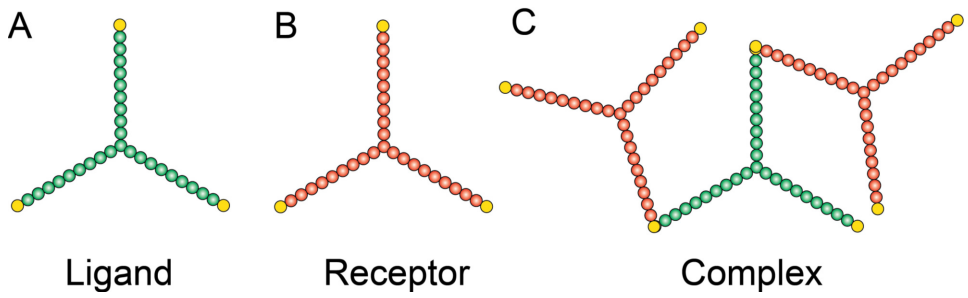

Figure 1: Illustration of the species (molecules and their complexes). (A) A trivalent ligand molecule. (B) A trivalent receptor molecule. (C) A ligand-receptor complex.

2.1.2 Motion of molecules and complexes

We consider both lateral diffusion (translation) and rotation of the molecule and complex graphs. The diffusion and rotation rates are determined by their effective size. To define the effective size for a graph, we consider a circle that completely contains the corresponding structure (Figure 2A). The circle is centered at the center of mass of the structure. The center of mass is determined by the mass of individual nodes and their spatial locations. We consider identical mass for all nodes because they have an identical size in the model (Table 1). The radius of the circle extends from this center to the farthest node. We define the circle area as the effective region (A), and its radius as the effective size (a). For a molecule graph, a remains fixed over time. However, for a complex graph, a is designated as $a(t)$ because it evolves with the complex size and geometry over time.

For a ligand or receptor molecule, we predefine their diffusion constants D_m (Table 1). For a complex graph, diffusion evolves based on the following rule: $D(t)/D_m = a_m/a(t)$, where a_m and $a(t)$ represent the effective size of a molecule and complex, respectively. This rule obeys the Stokes-Einstein formula according to which the diffusion coefficient is inversely related to particle radius. In a similar way, for a ligand or receptor molecule, we predefine their rotation speed (ω_m) (Table 1). For a complex, the rotation speed $\omega(t)$ obeys the following rule: $\omega(t)/\omega_m = a_m/a(t)$.

During the course of a simulation, we translate or rotate each graph as a rigid body. In each time step Δt, we first advance a graph by a distance $\sqrt{4D(t)\Delta t}\vec{e}$, where $\vec{e}$ is a unit vector in a random direction. We then rotate the graph around its center of mass by an angle $\omega(t)\Delta t$. We randomly select the rotation direction clockwise or anti-clockwise.

2.2 Time-adaptive Brownian Dynamics algorithm

The multiscale simulation algorithm is illustrated in Figure 2. The algorithm ensures collision-free motion of the structures defined by the molecule and complex graphs, as discussed below.

2.2.1 Calculation of adaptive step size

For each molecule or complex agent, the algorithm assigns a local search domain of radius S. In Figure 2B, the two concentric gray and blue circles illustrate a species agent and its search domain,

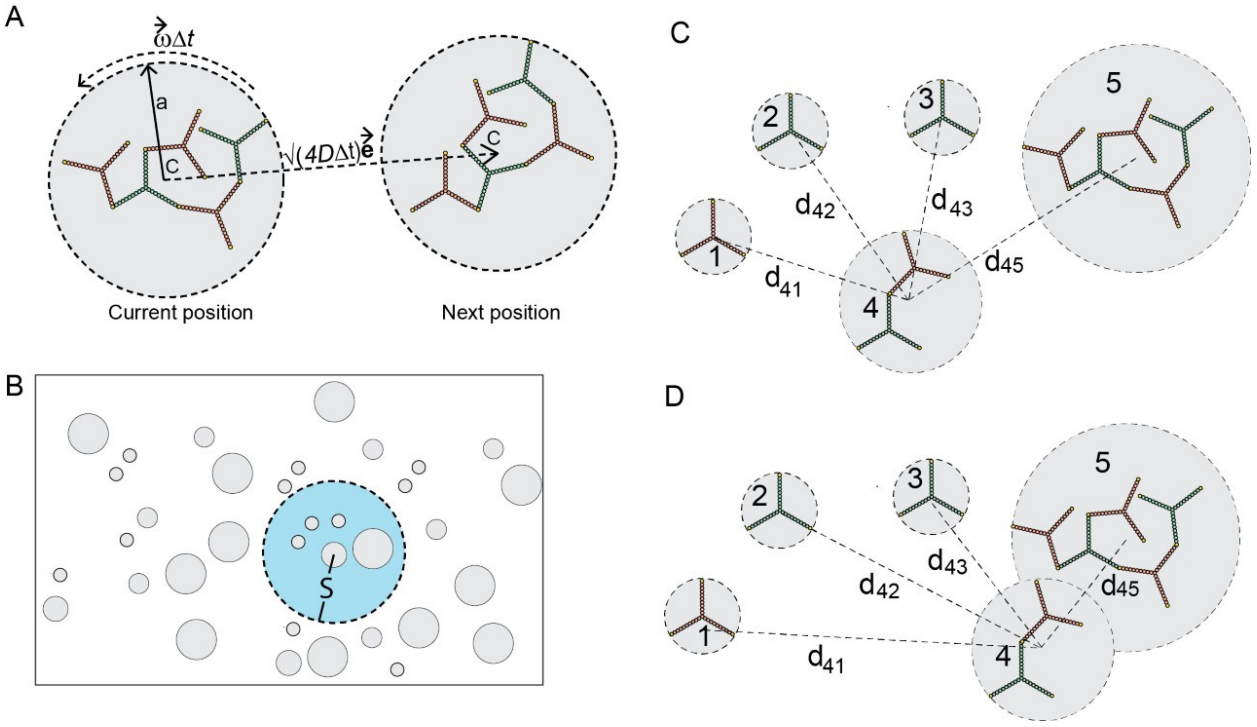

Figure 2: Illustration of the BD simulation algorithm. (A) A complex and its effective region (gray circle) and effective size (a). (B) A species (gray region at the center) and its search domain (blue region). The species has four collision partners (other gray regions with the search domain). (C) The center-to-center distance between a species (Agent 4) and its collision partners (Agents 1, 2, 3, and 5). (D) Overlap between the effective regions of agent 4 and 5. Such overlap is permitted ensuring the graphs within the gray circles never collide. When any such overlap happens in the entire system, the algorithm makes sure all species are advanced by less than 1 nm in each step.

respectively. The gray circle at the center represents the effective region of the species. The blue circle represents the search domain whose radius (S) is a tunable parameter.

In the algorithm, an upper bound l_{max} limits the particle jumps in a single step. l_{max} is the maximum distance any agent can travel over a single time step Δt. We set $l_{max} < S$. Thus, an agent can find all its potential (current) collision partners within its search domain. The agent in Figure 2B has four partners (other gray circles inside the blue circle).

In each BD step, the algorithm evaluates $d_{k,j}$, the center-to-center distance between each agent k and each of its collision partners $j \in \{1 \cdots, m_k(t)\}$ (Figure 2C,D). Here, the center of an agent implies the center of mass of the graph represented by the agent. The algorithm then computes the collision distance, $d^c_{k,j} = d_{k,j} - (a_k + a_j)$, where a_k and a_j are the effective size of the graphs

corresponding to agent k and j, respectively. Next, for the entire system, it identifies the smallest $d^c_{k,j}$: $d^c_{min} = min\{d^c_{k,j}\}$, where $k \in \{1, \cdots, n(t)\}$ (assuming the system has $n(t)$ species at time t), and $j \in \{1, \cdots, m_k(t)\}$. This distance, d^c_{min}, is then used to determine the time step size Δt, as discussed next.

If $d^c_{min} > 0$, all the grey circles representing the effective regions of the agents (graphs) in the system are at non-overlapping positions. Under this condition, the algorithm first evaluates the shortest possible time Δt^c that might lead to an overlap or collision between a pair of circles: $\Delta t^c = min\{(d^c_{k,j})^2/4(\sqrt{D_k(t)} + \sqrt{D_j(t)})^2\}$, where $k \in \{1, \cdots, n(t)\}$ and $j \in \{1, \cdots, m_k(t)\}$. $D_k(t)$ and $D_j(t)$ are the instantaneous diffusion coefficients of the corresponding agents. The algorithm then compares Δt^c with a prescribed lower-bound on the time step size, Δt_{min}. If $\Delta t^c > \Delta t_{min}$, the algorithm sets $\Delta t = \Delta t^c$, and advances all agents by $\sqrt{4D_k(t)\Delta t^c}$. This distance in a single step could represent a wide range of selections below the upper limit l_{max}, which has a default value of 100 nm in our simulations. However, if $\Delta t^c \leq \Delta t_{min}$, the algorithm sets $\Delta t = \Delta t_{min}$, and advances all agents by $\sqrt{4D_k(t)\Delta t_{min}}$. We consider $\Delta t_{min} = 10^{-5}$ seconds. At this time resolution, all graphs in our model are expected to advance by less than 1 nm in a single step.

If $d^c_{min} \leq 0$, at least one pair of circles in the system is in overlapping positions (Figure 2D). This implies that at least one pair of graphs (contained by those overlapping circles) are in close proximity to collide or mediate site-specific binding. Under such a condition, a high-resolution step is preferred to account for the steric collision or site-specific binding. the algorithm identifies the time step Δt based on the following rule: $\Delta t = max\{min\{l^2_{min}/4D_k(t)\}, \Delta t_{min}\}$ for $k \in \{1, \cdots, n(t)\}$. Here, l_{min} is a defined small length (a tunable parameter in the algorithm). We set $l_{min} = 1$ nm. This ensures all graphs are advanced by less than 1 nm in each step whenever there is an overlap of the effective regions.

2.2.2 Collision rejection and site-specific binding

The algorithm allows overlap of two circles (effective graph regions) while ensuring that the graphs within those circles are advanced in a collision-free manner. If a move results in a collision between two graphs, the move is rejected for the two graphs only. We consider a thin reaction layer, $l_r = 0.1$

nm, around each reaction center. When the reaction center of a ligand molecule and that of a receptor molecule fall within l_r, an implicit bond is assumed that holds the molecules together. In this work, we consider high-affinity ligand-receptor interaction and assume irreversible bond formation. We prohibit intra-complex binding or ring formation (Monine et al., 2010; Mahajan et al., 2014). Only chain and branched structure are allowed.

2.2.3 Code implementation

The simulation algorithm and the model were specified in C++. The C++ code is provided in the Supplementary Information. All simulations were conducted on a dedicated node on AMAZON Elastic Compute Cloud (EC2): c3.4xlarge (16vCPU, 30 GiB, 2×160).

3 Results

3.1 Simulation of multivalent ligand-receptor systems

In Figure 3A and B, we show snapshots from our simulations. At time zero, the molecules were initiated at random non-overlapping positions in the membrane domain, and snapshots were taken after 500 seconds of simulation.

Figure 3C shows the time evolution of receptor clustering. The mean cluster size represents the average of the number of receptors incorporated in the ligand-receptor complexes at the indicated time points. Initially, all receptors were available as isolated molecules, and hence the mean cluster size was 1 at time zero. As time progressed, larger (and fewer) clusters populated in the system.

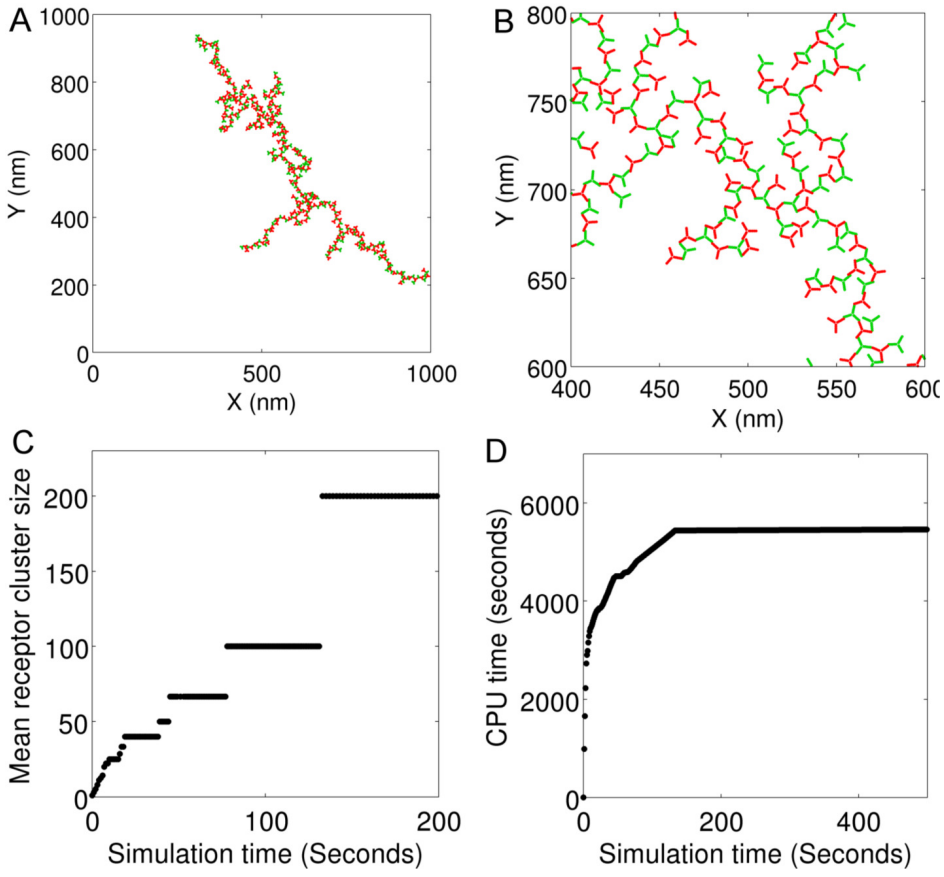

Figure 3: Simulation of a trivalent ligand-trivalent receptor system. The ligand and receptor molecules are indicated in green and red, respectively. (A) A branched complex structure formed in the simulation. (B) A small part of the structure in Panel A. (C) Time evolution of receptor aggregation. (D) Computation (CPU) time versus simulation time.

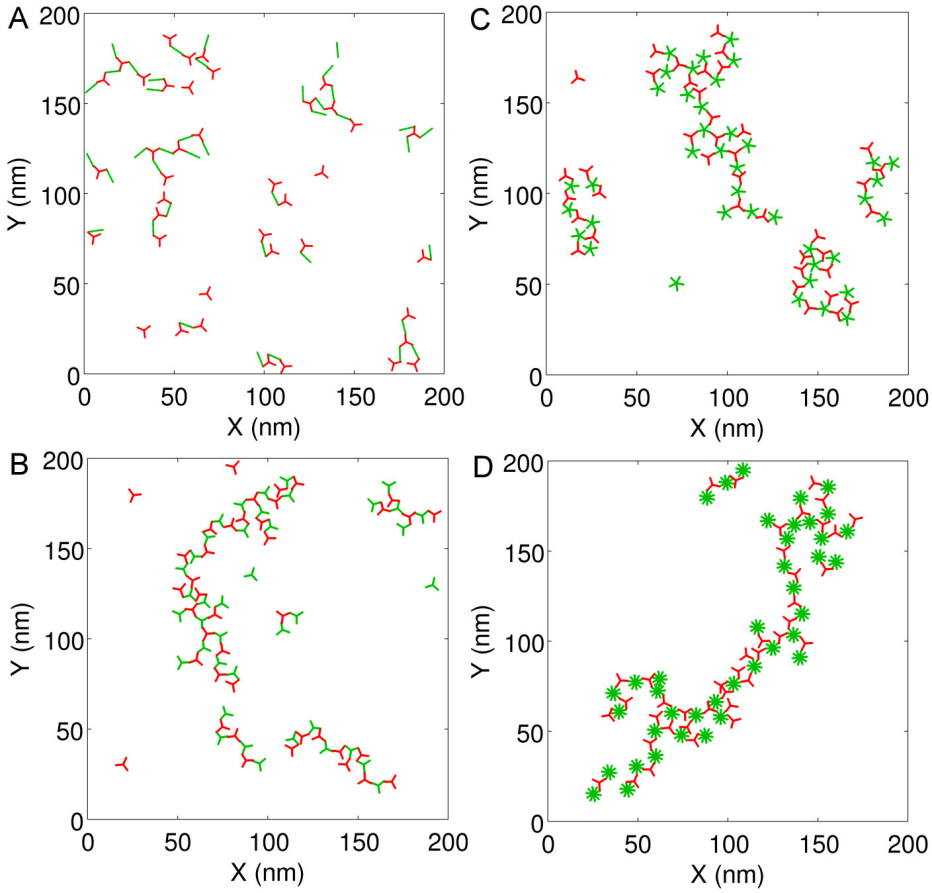

Figure 4: Simulation snapshots of multivalent ligand (green)-receptor (red) systems. In all cases the receptor is trivalent while the ligand is bivalent (A), trivalent (B), pentavalent (C), and decavalent (D).

Figure 3D shows how the computation time (CPU time) evolved as the simulation progressed. Most of the computation was spent in the first few seconds because the system was in a concentrated regime (species were either single molecules or small complexes). As the simulation progressed, larger (and fewer) species formed and the system approached a dilute regime. In the dilute regime, the adaptive algorithm took larger steps, thus accelerating computation.

Figure 4 shows some representative snapshots from simulations that involved different other multivalent ligand-receptor combinations. A movie file showing the time-evolution of a trivalent ligand-trivalent receptor system is included in the Supplementary Information.

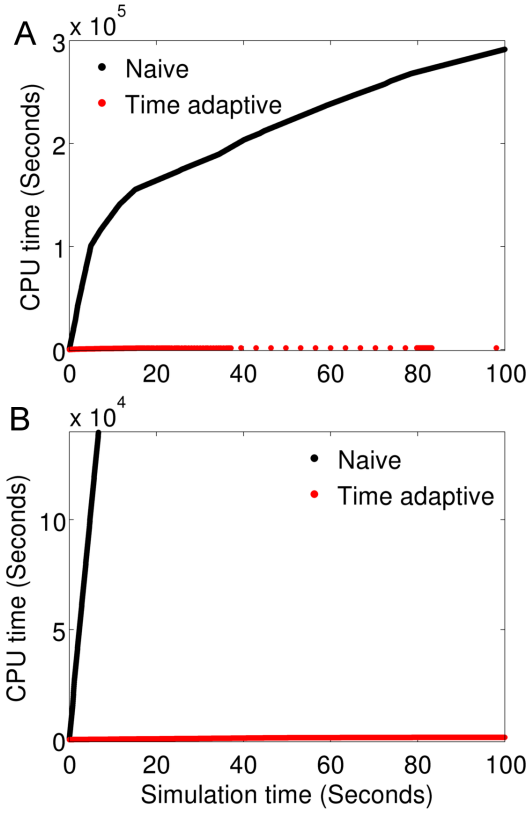

Figure 5: Computational speed gain in the time-adaptive algorithm. All simulations involved 100 trivalent receptor and 100 trivalent ligand molecules. (A) Concentrated regime: $1,000$ nm $\times$ $1,000$ nm membrane domain. (B) Dilute regime: $10,000$ nm $\times$ $10,000$ nm membrane domain.

3.2 Gain in computation speed

We evaluated the computational performance of our algorithm against a naive implementation of the model (Figure 5). In the naive implementation, the time step Δt was held fixed and the search domain (S) was infinite. The fixed Δt was chosen so that a single ligand or receptor molecule in each step could advance only by 1 nm, which is the smallest step size for the adaptive algorithm. The trivalent ligand-receptor system was simulated using the naive and adaptive approach for a dilute and a concentrated regime (Figure 5A and B, respectively), and corresponding computation (CPU) times were compared.

The figure shows a dramatic reduction in computation time in the adaptive algorithm. The difference between the naive and adaptive algorithm is even more pronounced in the dilute regime, where the naive algorithm spent most of the computation in particle advancement.

Robustness and accuracy

The results we have discussed so far represents $l_{max} = 100$ nm. A larger l_{max} accelerates computation but a smaller l_{max} is desired for higher spatial and temporal resolution (accuracy). However, as discussed in Methods, the algorithm requires $l_{max} \leq S$, where S is the local search radius associated with each species in the system. A larger S may facilitate computation by permitting a larger l_{max} but involves more computation in checking possible collision among the species.

We investigated three different combinations of l_{max} and S (Figure 6). The black lines in both panels of Figure 6 represents the case where $S = 100$ nm and $l_{max} = 1$ nm ($S : l_{max}$ = 100:1). We expected this combination to yield more accurate simulations because it enforced a high spatiotemporal resolution by constraining all particle jumps below 1 nm. The green lines represent the case where we considered the default scenario (Table 1): $S = 100$ nm and $l_{max} = 100$ nm ($S : l_{max}$ = 100:100). The red lines represent the case where we set $S = 10$ nm, and $l_{max} = 10$ nm ($S : l_{max}$ = 10:10). We compared the prediction accuracy (Figure 6A) and computational efficiency (Figure 6B) among these three cases.

Figure 6A shows the predicted mean receptor cluster size as a function of time under the above three conditions. Except for the stochastic variations, the predictions are closely similar in all three cases, indicating the robust behavior of the time adaptive feature. Despite the distinct upper bounds (l_{max}), the algorithm adapted the step sizes as necessary to capture the site-specific interactions with similar accuracies.

Figure 6B compares computation time among the three conditions above. First, it is interesting to note that the 100:1 case took only $\sim$ 3,500 CPU seconds despite the adaptive feature was compromised with $l_{max} = 1$ nm. Corresponding naive simulation, where we also restricted the step size below 1 nm, took $\sim 300,000$ seconds (Figure 6A). Therefore, the accelerated computation in the 100:1 case is not due to the adaptive feature. Rather it is due to the local search feature of the algorithm. The 100:1 case involves search over a domain of 100 nm radius, whereas the naive implementation involves a search over the entire membrane domain to avoid species collision.

The two other cases (100:100 and 10:10) revealed modest improvements in computation speed

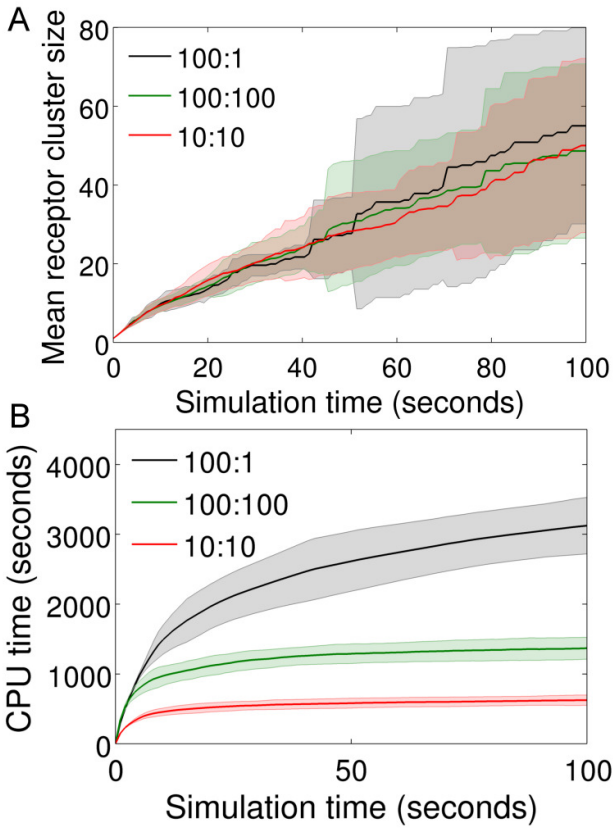

Figure 6: Sensitivity and computational performance under different combinations of S and l_{max}. Three different combinations were chosen and compared. The lines in different color correspond to different ratio of S and l_{max}: 100 nm:1 nm (black), 100 nm:100 nm (green), and 10 nm:10 nm (red). Each line is the average of 30 realizations. The shaded regions represent corresponding standard deviations.

compared to the 100:1 case. Nevertheless, these two cases plateaued rapidly as the system approached from the concentrated to dilute regime. This indicates that the adaptive feature could be of greater advantage in the dilute regimes. On the other hand, the local search feature alone can improve computation speed significantly in a concentrated regime.

4 Discussion

In this work, we have demonstrated a computationally efficient algorithm for modeling reaction-diffusion systems involving multivalent molecules. Here, we limited our focus to modeling cell-surface ligand-receptor assembly, a key early step of immunoreceptor and receptor tyrosine kinase (RTK) signaling (Puffer et al., 2007; Schlessinger, 2000; Deak et al., 2016). The approach and analysis could be adapted for modeling and simulation of more complex multivalent species or particle interactions in molecular biology or other fields.

Among the existing spatiotemporal modeling tools, SRSim (Grünert and Dittrich, 2010) provides a unique capability for spatiotemporal modeling of macromolecular assembly. SRSim also employs spatial graphs to represent protein molecules and complexes. However, SRSim simulations more comparable could be expensive for the time and spatial scales of signaling protein interactions. MCell (Bartol et al., 2015) is a spatiotemporal modeling tool with the special capability to consider complex geometries for cellular compartments. However, MCell treats reaction species as featureless points (Kerr et al., 2008). Therefore, it has a limited ability to incorporate site-specific attributes of the biomolecules. VirtualCell (Loew and Schaff, 2001) is another spatiotemporal modeling tool. However, it is primarily limited to partial differential equation (PDE)-based deterministic modeling. It might be useful to extend our multiscale approach to incorporate some of the unique features of the above tools. In particular, our algorithm might be coupled to features creating complex geometries for the reaction compartments, as implemented in MCell.

In this work, we have limited the scope to modeling protein-protein interactions in the plasma membrane, which is often treated as a two-dimensional space. However, a three-dimensional extension of the approach might enable modeling the cytoplasmic and extracellular compartments. We expect that the computational benefits of our algorithm will be significant in a three-dimensional space where simulation of particle advancement could be more expensive than that in a two-dimensional system.

5 Acknowledgements

This work was supported by the National Science Foundation [1609642]. No additional external funding was received to support this work. The funders had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

6 Author Disclosure Statement

No competing financial interests exist.

References

Ayton, G. S., Noid, W. G., and Voth, G. A. 2007. Multiscale modeling of biomolecular systems: in serial and in parallel *Current opinion in structural biology* 17, 192–198.

Bartol, T. M., Dittrich, M., and Faeder, J. R. 2015. MCell *in Encyclopedia of Computational Neuroscience* 1673–1676 Springer.

Bhattacharyya, R. P., Reményi, A., Yeh, B. J., et al. 2006. Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits *Annu. Rev. Biochem.* 75, 655–680.

Deak, P. E., Vrabel, M. R., Pizzuti, V. J., et al. 2016. Nanoallergens: A multivalent platform for studying and evaluating potency of allergen epitopes in cellular degranulation *Experimental Biology and Medicine* 1535370216644533.

Gilbert, N. 2008. *Agent-based models* 153 Sage.

Goldstein, B. and Perelson, A. S. 1984. Equilibrium theory for the clustering of bivalent cell surface receptors by trivalent ligands. Application to histamine release from basophils *Biophysical journal* 45, 1109–1123.

Grünert, G. and Dittrich, P. 2010. Using the SRSim software for spatial and rule-based modeling of combinatorially complex biochemical reaction systems *in International Conference on Membrane Computing* 240–256 Springer.

Hlavacek, W. S., Posner, R. G., and Perelson, A. S. 1999. Steric effects on multivalent ligand-receptor binding: exclusion of ligand sites by bound cell surface receptors *Biophysical journal* 76, 3031–3043.

Kerr, R. A., Bartol, T. M., Kaminsky, B., et al. 2008. Fast Monte Carlo simulation methods for biological reaction-diffusion systems in solution and on surfaces *SIAM journal on scientific computing* 30, 3126–3149.

Loew, L. M. and Schaff, J. C. 2001. The Virtual Cell: a software environment for computational cell biology *TRENDS in Biotechnology* 19, 401–406.

Mahajan, A., Barua, D., Cutler, P., et al. 2014. Optimal aggregation of FcεRI with a structurally

defined trivalent ligand overrides negative regulation driven by phosphatases *ACS chemical biology* 9, 1508.

Monine, M. I., Posner, R. G., Savage, P. B., et al. 2010. Modeling multivalent ligand-receptor interactions with steric constraints on configurations of cell-surface receptor aggregates *Biophysical journal* 98, 48–56.

Puffer, E. B., Pontrello, J. K., Hollenbeck, J. J., et al. 2007. Activating B cell signaling with defined multivalent ligands *ACS chemical biology* 2, 252–262.

Schlessinger, J. 2000. Cell signaling by receptor tyrosine kinases *Cell* 103, 211–225.

Sneddon, M. W., Faeder, J. R., and Emonet, T. 2011. Efficient modeling, simulation and coarse-graining of biological complexity with NFsim *Nature methods* 8, 177–183.

Weng, G., Bhalla, U. S., and Iyengar, R. 1999. Complexity in biological signaling systems *Science* 284, 92–96.

Yang, J. and Hlavacek, W. S. 2011. The efficiency of reactant site sampling in network-free simulation of rule-based models for biochemical systems *Physical biology* 8, 055,009.

Yang, J., Monine, M. I., Faeder, J. R., et al. 2008. Kinetic Monte Carlo method for rule-based modeling of biochemical networks *Physical Review E* 78, 031,910.

Parameter	Comment
L = 1,000 nm	Membrane length
W = 1,000 nm	Membrane width
$n_L = 100$	Ligand copy number.
$n_R = 100$	Receptor copy number
$n_{siteL} = 3$	Number of binding arms in a ligand
$n_{siteR} = 3$	Number of binding arms in a receptor
$n_{eL} = 5$	Number of nodes per ligand arm
$n_{eR} = 5$	Number of nodes per receptor arm
$r_e = 0.5$ nm	Radius of each node
$D_m = 10^{-10}$ cm^2/s	Diffusion coefficient of a ligand or receptor molecule
$\omega_m = 10$ radian/s	Rotation speed of a ligand or receptor molecule.
$l_r = 0.1$ nm	Reaction layer thickness.
$S = 100$ nm	Radius of the local search domain.
$l_{max} = 100$ nm	Upper limit on particle jump size.
$l_{min} = 1$ nm	Limit on jump size when effective regions overlap.
$\Delta t_{min} = 10^{-5}$ s	Lower bound on time step size.

Table 1: Default parameter values of the model.