

SOFTWARE

Open Access



Cell4D: a general purpose spatial stochastic simulator for cellular pathways

Donny Chan^{1,2†}, Graham L. Cromar^{1†}, Billy Taj¹ and John Parkinson^{1,2,3*}

[†]Donny Chan and Graham L. Cromar contributed equally to this manuscript.

*Correspondence: jparkin@sickkids.ca

¹ Program in Molecular Medicine, Hospital for Sick Children, Toronto M5G 0A4, Canada

² Department of Molecular Genetics, University of Toronto, Toronto M5S 1A8, Canada

³ Department of Biochemistry, University of Toronto, Toronto M5S 1A8, Canada

Abstract

Background: With the generation of vast compendia of biological datasets, the challenge is how best to interpret 'omics data alongside biochemical and other small-scale experiments to gain meaningful biological insights. Key to this challenge are computational methods that enable domain-users to generate novel hypotheses that can be used to guide future experiments. Of particular interest are flexible modeling platforms, capable of simulating a diverse range of biological systems with low barriers of adoption to those with limited computational expertise.

Results: We introduce Cell4D, a spatial-temporal modeling platform combining a robust simulation engine with integrated graphics visualization, a model design editor, and an underlying XML data model capable of capturing a variety of cellular functions. Cell4D provides an interactive visualization mode, allowing intuitive feedback on model behavior and exploration of novel hypotheses, together with a non-graphics mode, compatible with high performance cloud compute solutions, to facilitate generation of statistical data. To demonstrate the flexibility and effectiveness of Cell4D, we investigate the dynamics of CEACAM1 localization in T-cell activation. We confirm the importance of Ca²⁺ microdomains in activating calmodulin and highlight a key role of activated calmodulin on the surface expression of CEACAM1. We further show how lymphocyte-specific protein tyrosine kinase can help regulate this cell surface expression and exploit spatial modeling features of Cell4D to test the hypothesis that lipid rafts regulate clustering of CEACAM1 to promote trans-binding to neighbouring cells.

Conclusions: Through demonstrating its ability to test and generate hypotheses, Cell4D represents an effective tool to help integrate knowledge across diverse, large and small-scale datasets.

Keywords: Computational modeling, Temporal-spatial models, Simulations, Meso-scale, Visualization

Background

The post-genome era is characterized by increasing availability of large, heterogeneous datasets detailing the molecules driving biological systems. These include genome-scale datasets encompassing the expression and dynamics of genes and their products [1–3], their localization [4, 5], interactions with other biomolecules [6, 7]



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

and organization within pathways [8, 9]. A major challenge is how best to unlock the full potential of these rich datasets to advance our understanding of complex biological processes. To address this challenge, a variety of computational modeling platforms have been developed, capable of exploiting these datasets to simulate the dynamics of the underlying systems at the meso-scale, bridging the nanometer scale of atoms and the micrometer scale of cellular structures [10–13].

Despite the availability of these platforms, widespread adoption has been limited, in part due to barriers concerning the level of computational expertise required for their operation. To help overcome these limitations, toolkits such as CellBlender [14], have been developed to help with the construction and visualization of models. At the same time there is a need for simulation platforms for cell biologists, without skills in computing, who nevertheless represent the domain experts and target audience for meso-scale simulators. To meet this need, we have developed Cell4D, a robust spatial stochastic simulation platform with integrated graphical visualization and a browser-based editing tool supporting the creation of sophisticated models written in XML format, compatible with Linux operating systems and allowing for further feature enhancements.

To demonstrate the ability of Cell4D to model a biological system and explore hypotheses, we applied it to examine the dynamics of carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1) in T-cell activation [15]. CEACAM proteins are glycosylated transmembrane adhesion molecules featuring an N-terminal IgV-like domain together with a variable number of IgC2-like domains, a transmembrane domain and a cytoplasmic tail [16–18]. They are widely expressed in many cell types and play a role in multiple functions including cell growth, metabolism and responding to infection. CEACAM1 is expressed by activated T cells and serves to transmit extracellular signals across the cell membrane through intercellular (trans) homophilic and heterophilic binding through its transmembrane domain to inhibit continued activation [19]. In resting T cells, CEACAM1 largely exists in the form of cis-homodimers, incapable of binding extracellular ligands. Clathrin-mediated internalization further ensures a relatively low concentration at the cell surface. Upon binding by active calmodulin, cis-homodimers disassociate facilitating trans-binding of the monomeric form of CEACAM1 to other monomers of CEACAM1 on adjacent cells [19]. Such binding appears higher in the presence of multiple IgC2 domains, suggesting a role for local accumulation (clustering) of CEACAM1 monomers [16, 20]. Regulation of CEACAM functionality is thought to operate through localized concentrations of Ca^{2+} [21] as well as phosphorylation of CEACAM1 by Src-family kinases such as Lck [22]. Furthermore, it has been hypothesized that regulation may also involve the preferential sequestration of CEACAM1 monomers within membrane microdomains [23]. Further details on CEACAM functionality are provided in the section *Clustering of CEACAM1 at the cell surface indicates a role for lipid rafts in regulating signaling and is dependent on calmodulin activation*.

Applying Cell4D we examined the dynamics of calmodulin activation through calcium binding and subsequently integrated this model to examine the role of calmodulin, Lck kinase and membrane microdomains to regulate the formation of local surface clusters of monomeric CEACAM1 to promote trans-binding.

Implementation

Cell4D is written in C++ and is supported to run under Linux (Ubuntu versions 20.04 and 22.04). To perform simulations, model files are loaded into the simulation environment which features a graphical interface that allows for real-time visualization of the running simulation (Fig. 1A). To help with the generation of model files, a Cell4D Model Editor—accessible at <https://compsysbio.org/cell4d> is provided (Fig. 1B). Simulations iterate over a user-defined number of timesteps. Within each timestep, molecular diffusion events are first simulated followed by reaction events (Fig. 1C). For small molecules represented by local concentrations within a lattice cell (c-voxel), the diffusion into neighboring c-voxels are calculated deterministically (Fig. 1D; [24]). C-voxels are typically used to represent spaces between 0.01 and 1 μm in length. For molecules and complexes represented as point particles, displacement is determined either stochastically (i.e., Brownian motion e.g. Fig. 1E) or deterministically (i.e. active transport). Full Details of the System and Methods are provided in Additional file 8.

Molecular movement and reactions

Point particle movement is modeled through random walks, while bulk particle movement is modeled using Fick's laws of diffusion [24] and is simulated using the forward Euler method [25]. Reactions are implicitly defined based on the presence of substrates. Cell4D allows the definition of unimolecular and bimolecular reactions. For bimolecular reactions involving point particles, Cell4D uses the Andrews-Bray-adjusted

(See figure on next page.)

Fig. 1 Conceptual overview of Cell4D. **A** Screenshot of the Cell4D graphical interface. For further details on the interface see the project GitHub: <https://github.com/ParkinsonLab/cell4d>. **B** Model design interface. A web interface for creating and editing Cell4D model files. Custom XML model files can be loaded in by a user, or a preset example model can be selected. Once a model is loaded in, parameters of the model will automatically fill the text boxes in the interface. Users can edit the model by modifying the text in each textbox; real-time error messages will appear to prevent invalid inputs from being added. When the "Save" button on the bottom is pressed, if no textboxes have invalid inputs, the information in the text boxes of the current active tab will be saved to the loaded model. Users can switch between tabs that contain different model information such as modifying compartment spaces, molecular species, and reactions. Once all changes are saved on each tab, the user can click "Save model" on the top banner to download the loaded model onto their local device. **C** Simulation set up and flow time cycle logic. Parameters that describe system behavior such as how molecules behave within the simulation space as well as the way they interact with other molecules are described in an XML input file, which is then used to initialize the simulation space. The simulation then cycles through a series of steps until the end condition is met. Output occurs in two forms: tab-delimited files (.tsv) of molecule counts at each time step, and particle logs recording the position and state information of molecules in the simulation. **D** Diffusion of bulk molecules shown for a single voxel. At each time step, a portion of the bulk molecules for each c-voxel will diffuse into a neighboring c-voxel, based on the current concentration and the molecule's diffusion rate constant. This is calculated for all c-voxels at every timestep. (i) an initial setup where a c-voxel contains 100 molecules (orange) with three adjacent voxels that contain 5 molecules each (blue). Diffusion of molecules into grey voxels are disabled in this example. (ii) bulk molecule diffusion calculation is done for each voxel which depends on the system timestep length setting and the concentration of molecules in each voxel. (iii) image shows the concentration of molecules in each voxel after one timestep. **E** Implementation of off-lattice movement of point particles. (i) and (ii) show the diffusion path of the reactant (blue) after 1 μs for 0.2 μs and 1 μs timestep systems respectively. In (i), using 0.2 μs timesteps, the particle was able to enter the reaction radius of the first reactant (red) which would allow a reaction to occur. In (ii), using 1 μs timesteps, although the final diffusion path of the particle remains the same, there is no step where the second reactant has an opportunity to react with the first reactant. To avoid such cases Cell4D implements the Andrews-Bray-adjustment to artificially increase reaction radii of molecules according to the size of the time step (iii and iv) [26]

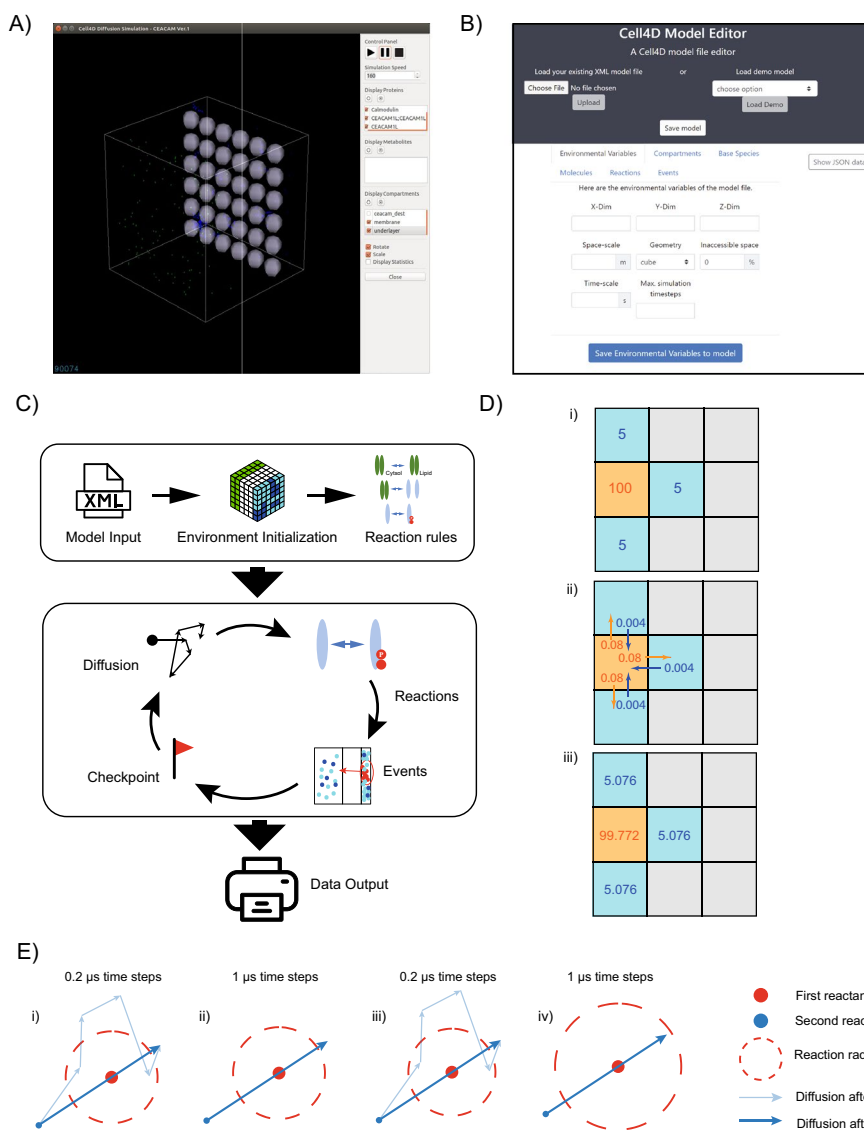


Fig. 1 (See legend on previous page.)

Smoluchowski method to reduce the possibility of missing reactions due to insufficient time-step resolution (Fig. 1E). For further details see Additional file 8.

Modeled systems

Calmodulin activation

Models used a simulation environment defined as a cube with side length of 8×10^{-7} m (0.08 μm). Calcium ions were defined as bulk molecules and calmodulin was defined as point particles with N- and C-terminal bindings sites. For some simulations, the model space was divided into a lattice comprising $10 \times 10 \times 10$ c-voxels (0.08 μm side length) which were subdivided into five $2 \times 10 \times 10$ cross-sectional compartments (quintiles Q1 thru Q5). Within Q1, four c-voxels were defined as sites of calcium-release.

CEACAM1 activation

Models used a simulation environment composed of $6 \times 6 \times 6$ c-voxels of length $0.2 \mu\text{m}$, subdivided into an outer 'membrane', a 'cytosolic interface', a 'cytosol' and an 'organelle'. For some simulations, within the membrane, microdomains were defined to represent lipid rafts. The model features CEACAM1 molecules capable of forming monomers or dimers, Lck molecules capable of phosphorylating CEACAM1, calmodulin which can dissociate CEACAM1 dimers to monomers and calcium, which can activate calmodulin.

Details of both systems are provided in Additional file 8.

Results

Cell4D accurately simulates diffusion and reaction events

Cell4D is a spatio-temporal simulation platform that performs simulations within a space defined by cubic lattice sites (c-voxels). The simulator features reaction-based rules governing molecular interactions, formation and dissolution of protein complexes, state changes (e.g., allowing post-translational modifications), enzymatic reactions and defined events (e.g., molecule trafficking between compartments). Using a hybrid on/off lattice approach, small molecules are represented as concentrations, diffusing via a grid pattern dividing the overall model space while larger molecules can be tracked as point particles that freely diffuse off-lattice. Within the lattice, multiple compartments can be defined (e.g., membranes or other organelles), providing boundaries and allowing the definition of compartment-specific rules governing molecule behavior. Compartments can contain other compartments and both compartments and the lattice environment itself can be trimmed to allow the representation of custom geometries. Molecules can be defined to occupy specific compartments (i.e. offering the option to exclude molecules from certain compartments). When a particle encounters the boundary of a disallowed compartment or the edge of the simulation space, its trajectory is deflected, such that it continues to move within its original lattice space. In the case of small molecule concentrations within a lattice, diffusion is disabled across the interface with restricted compartments. Rules may be combined within the same data model without the need for additional custom modifications and re-compilation of the C++ code.

In initial simulations we benchmarked the ability of Cell4D to accurately model molecular diffusion and reaction events (see Additional file 8). Focusing on diffusion, we found that Cell4D accurately models Brownian motion for both point particles and bulk molecules as predicted by Fick's laws (Additional file 1: Fig. S1 and Additional file 2: Fig. S2). In terms of reactions, simulated product formation for unimolecular reactions matched theoretical yields under all tested conditions (Additional file 3: Fig. S3). For bimolecular reactions involving both bulk molecules and point particles, we found that while timescale had negligible impact, accuracy increased for simulations with higher rates of diffusion (Additional file 4: Fig. S4). Bimolecular reactions involving only point particles were found to be sensitive to timescale and reaction rate (Additional file 3: Fig. S3B). To correct these errors identified in Additional file 3: Fig. S3B, we implemented the Andrews-Bray (AB) adjustment of the Smoluchowski method (Additional file 5: Fig. S5), resulting in high accuracy predictions under all conditions except those violating the assumption that the main constraint for reaction rates stems from particle collision

[26]. Having benchmarked the performance of Cell4D, we next demonstrate the ability of Cell4D to simulate a biological system.

Application of Cell4D to investigate mechanisms underlying CEACAM1 mediated T-cell activation

We chose to focus on CEACAM1 mediated T-cell activation to examine two aspects of Cell4D's capacity to model biochemical pathways. First, we use Cell4D to explore the dynamics of calcium:calmodulin binding dynamics, a critical step in the activation of CEACAM1 signaling. Second, we use Cell4D to examine the hypothesis that CEACAM1 signaling is regulated through its spatial organization at the cell surface and the underlying mechanisms that compete to determine its local concentration.

Calcium microdomains play a key role in calmodulin activation

To model calmodulin activation we applied a cooperative binding model [27] to confirm the importance of calcium microdomains in overcoming cellular buffering in signal transduction. Details on the cooperative binding kinetics involving the four calcium binding domains of calmodulin are provided in Additional file 8. First we examined a baseline model to establish that calmodulin binding occurs under physiological calcium concentrations. To compare simulations with theoretical predictions, we initially predicted the distribution of the four potential states of calmodulin: CaM_1, CaM_2, CaM_3, and CaM_4, representing calmodulin molecules with 1, 2, 3, or 4 binding sites occupied with calcium respectively, for 2 μM calmodulin exposed to a range of Ca^{2+} concentrations (0 to 16 μM) in a volume of 5×10^{-16} L (0.8 μm side length). Calculations were performed using the forward Euler method under the well-mixed assumption (see Additional file 8). The same conditions were then reproduced using Cell4D by placing the same concentration of molecules distributed randomly in a simulation space of $4 \times 4 \times 4$ c-voxels of length 0.2 μm . Results from simulations were similar to theoretical predictions and showed that under Ca^{2+} concentrations relevant to T cell activation (0.1 μM –1.2 μM), the proportion of saturated (activated) calmodulin (CaM_4) was negligible (< 1%; Additional file 6; Fig. S6). However, we found that modest activation of calmodulin (~20%) required Ca^{2+} concentrations an order of magnitude higher (> 10 μM) than observed experimentally [28, 29].

While the average concentration of calcium in a cell is approximately 1–2 μM , local concentrations can be significantly higher (of the order 100 μM –1 mM) in regions proximal to active calcium channels, which we term *microdomains* [27]. To explore the range and impact of such channels on calmodulin activation, we constructed a model composed of five compartments (defined as quintiles, Q1–Q5) each composed of $2 \times 10 \times 10$ c-voxels of 0.08 μm . 1 mM of Ca^{2+} was then introduced in Q1 at each timestep and allowed to freely diffuse and exit from Q5 (Fig. 2A). To assess the potential impact of the spatial organization of Ca^{2+} release, three arrangements were investigated (see Additional file 8). As expected, Ca^{2+} concentrations decreased with increasing distance from the source, ranging from 1–10 μM across the five quintiles (Fig. 2B). Further, the relative proportion of activated calmodulin (CaM_4) reflects the distance from the source of Ca^{2+} (from 5–10% to 20–30% for Q5 and Q1 respectively; Fig. 2C). Configuration of calcium channels had minimal impact on either

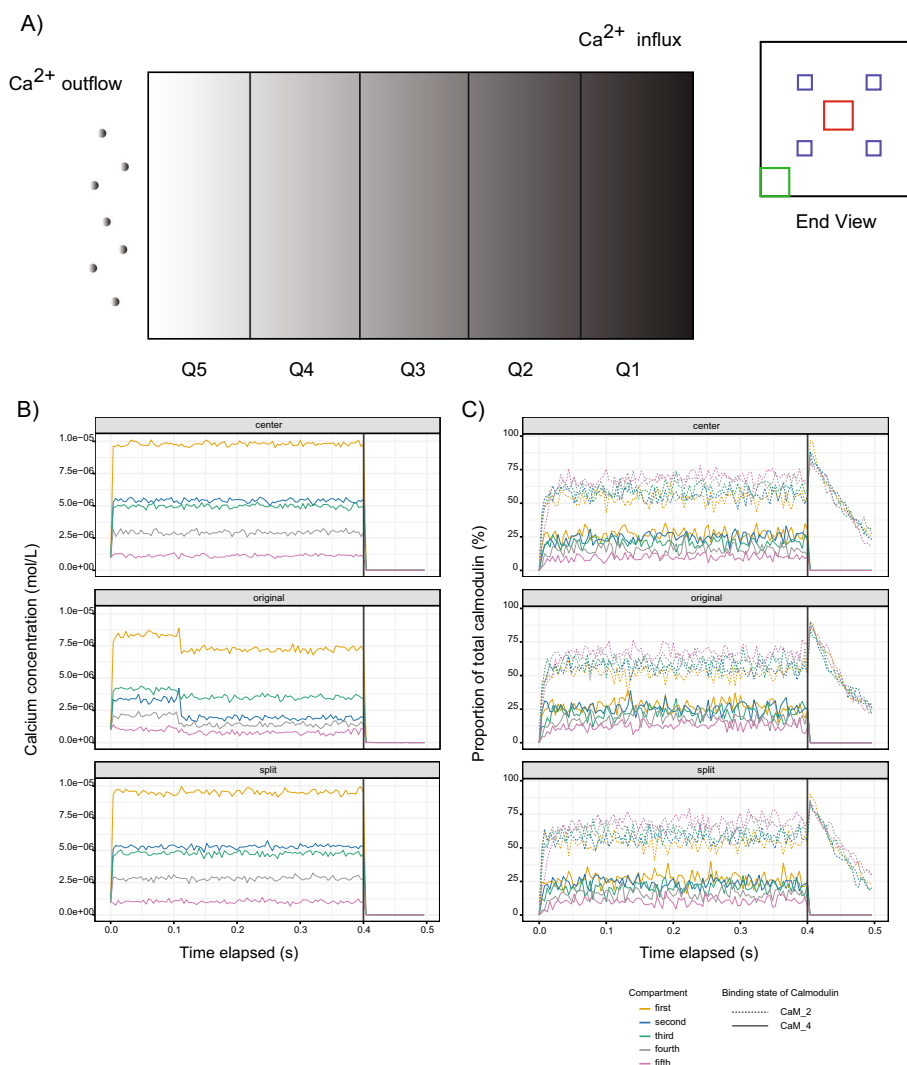


Fig. 2 Activation of calmodulin in Ca^{2+} microdomains. **A** Schematic representation of simulations of calmodulin activation involving five compartments (Q1–Q5). End view shows the spatial orientation of the center, original and split types (indicated by red, green, and blue squares respectively) of microdomain-layout used to release Ca^{2+} into the system. Calcium is removed at the Q5 region to maintain the expected overall Ca^{2+} concentration of approximately 1–2 μM . The top and bottom of the environment represent hard boundaries that constrain Ca^{2+} diffusion. **B** Effective average concentration (mol/L) of Ca^{2+} in each of the compartments over the course of the simulations. **C** Average saturation of two states of calmodulin (CaM_2 and CaM_4) in each compartment over the course of the simulations

calcium concentrations or activated calmodulin. Interestingly, in Q5, despite the proportion of primed calmodulin (CaM_2) being ~75%, activated calmodulin remained relatively low (5–10%). This suggests that while elevated background concentrations of Ca^{2+} are sufficient to maintain a primed population of calmodulin, its activation appears transient and likely requires proximity to sources of Ca^{2+} , a local effect that is not well captured in well-mixed models.

Having established the conditions under which elevated Ca^{2+} levels lead to calmodulin activation, we next consider factors affecting the spatial organization of CEACAM1 in response to T-cell activation.

Clustering of CEACAM1 at the cell surface indicates a role for lipid rafts in regulating signaling and is dependent on calmodulin activation

The primary mechanism of CEACAM1 function is to transduce extracellular signals to the cytosol through intracellular *trans*-binding. In resting T cells, CEACAM1 concentration at the cell membrane remains relatively low due to clathrin-mediated internalization, involving the adaptor protein complexes AP1 and AP2. CEACAM1 in resting T cells is primarily found in the cis-dimeric form, predicted to be inactive due to steric hinderance that prevents access to the ITIM sites in the cytoplasmic tail, blocking SHP1 recruitment which leads to downstream CEACAM1 inhibitory function (not considered here). Upon binding with calmodulin, CEACAM1 dimers are converted to their monomeric form. Phosphorylation of tyrosine residues in the ITIM sites by Lck, both blocks binding to AP1 and AP2 (preventing internalization and resulting in retention of CEACAM1 at the cell surface) as well as promote binding to their counterparts on adjacent cells in *trans* (resulting in the suppression of host immune responses). With the observation of CEACAM1 clustering at the cell surface, it has further been suggested that the CEACAM1 binding dynamics may also be driven, at least in part, by partitioning through lipid rafts [16, 21, 23]. Here we are interested in using Cell4D to examine potential mechanisms, including interactions involving calcium, calmodulin, Lck kinase as well as the involvement of lipid rafts, that drive the accumulation of CEACAM1 at the cell surface (Fig. 3A).

To examine the impact of lipid rafts, we constructed two models (*raft* and *no-raft* models). Both models involve an environment composed of $6 \times 6 \times 6$ voxels of length $0.2 \mu\text{m}$ sub-divided into four compartments representing the cell membrane, a region of cytosol immediately adjacent to the membrane, the rest of the cytosol and an organelle (see Additional file 8). Particles in the membrane were restricted to diffusion in 2-dimensions only. Calmodulin was confined to the cytosolic interface with the

(See figure on next page.)

Fig. 3 Modeling CEACAM1 signaling. **A** Schematic of reactions used in the model. CEACAM1 dimers are transported between the membrane and cytosol compartments. Within the membrane, CEACAM1 dimers disassociate to monomers based on interactions with activated calmodulin. Src-family kinases (Lck) phosphorylate the ITIM regions of the CEACAM1 cytoplasmic tail, preventing its transport back to the cytosol and shifting the equilibrium of CEACAM1 localization to the membrane. The membrane region can be defined into lipid-ordered (lipid rafts) and lipid-disordered regions. CEACAM1 preferentially associates with these regions based on its oligomeric state, as shown by the solid and dashed orange arrows between the two membrane regions that indicate transport. The end state of the activated T cell consists of the clustering of CEACAM1 monomers within lipid rafts. **B** Representation of a 2D membrane compartment with lipid raft sub-compartments. CEACAM1 dimers (green) are generally localized outside of lipid raft regions (indicated in dark yellow), while CEACAM1 monomers (blue) preferentially localize within lipid raft compartments. **C** The lipid raft CEACAM1 model was tested using 0, 10, 20 molecules of Lck and 0, 2, 5, 10, 20 molecules of active calmodulin to examine the effects of both proteins on CEACAM1 surface expression. Error bars represent standard deviation for 6 replicates. Results show that CEACAM1 surface concentration is dependent on the concentration of activated calmodulin, but not Lck. **D** Impact of *trans*-binding rate constants (i.e. unbound monomer to *trans*-bound (immobilized) monomer) on CEACAM clustering. Simulations were performed for both the lipid raft model (left) and the no-raft model (right). The binding rate constant shows a positive correlation with the total count of *trans*-bound (clustered) CEACAM1 monomers. For low binding constant conditions in the presence of Lck and CaM in the no-raft model, there appears to be a threshold effect where the rate of cluster formation is slow in the beginning of the simulation, but accelerates after a certain point. For the no-raft Lck-absent models, low calmodulin levels did not lead to a significant clustered CEACAM1 population, while high calmodulin only produced an increased surface CEACAM1 concentration at the highest binding rate constant

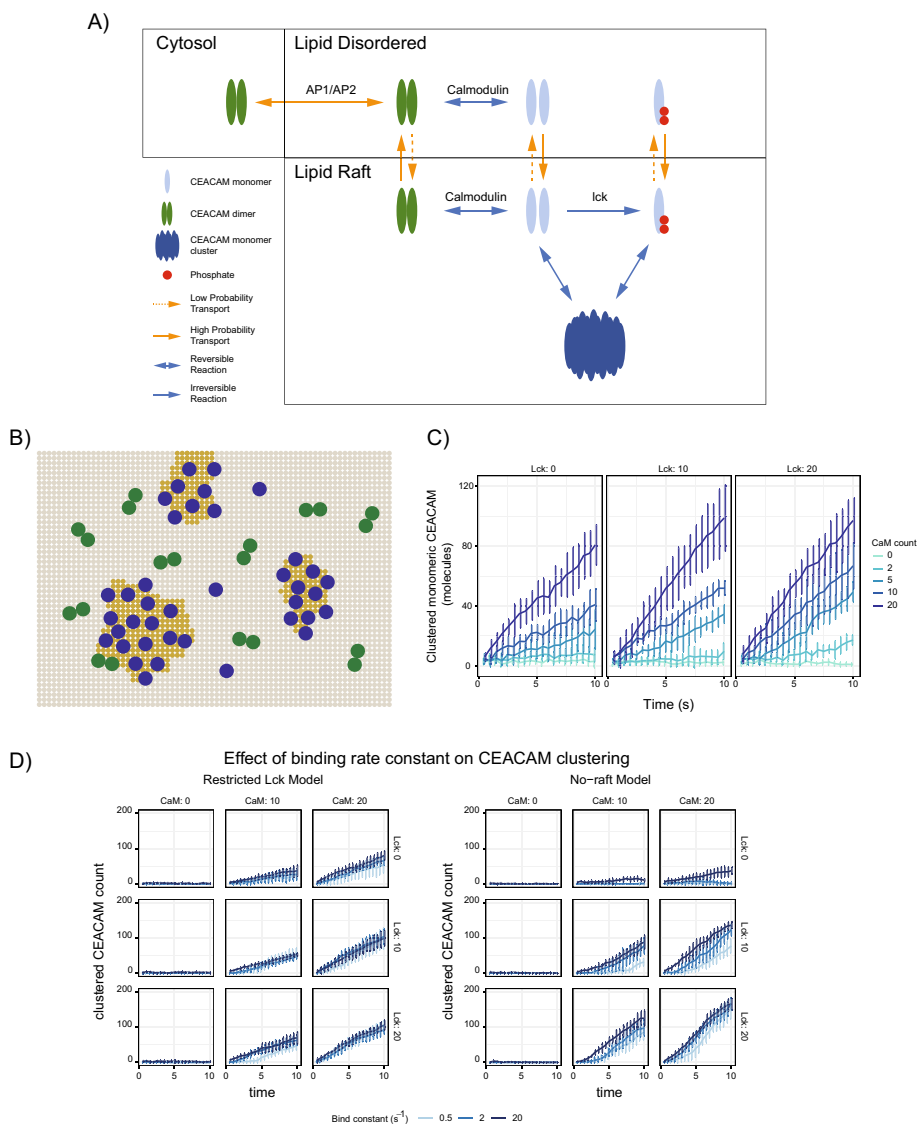


Fig. 3 (See legend on previous page.)

membrane while particles representing CEACAM1 could move between compartments through defined transport events (representing clathrin-mediated internalization and export to the membrane). In the raft model, CEACAM1 monomers are preferentially enriched in membrane microdomains, while CEACAM1 dimers are preferentially excluded. Further, Lck kinases, previously shown to be enriched and functionally segregated in lipid rafts [30], were constrained within microdomains, representing rafts, defined within the membrane (Fig. 3B). Thus, in this model, Lck is only able to phosphorylate CEACAM1 (through its ITIM domain) located within the lipid rafts. Simulations were performed for 200,000 timesteps of 50 μ s (10 s total) and explored the impact of different concentrations of Lck and Ca²⁺-activated calmodulin, on the concentration and distribution of CEACAM1. Models initiated with no calmodulin or Lck represent the T cell model at rest, while the T cell activated state

was modeled through the inclusion of 20 molecules of activated calmodulin and Lck (0.1 μM).

For both raft and no-raft models, surface clustering of CEACAM1, representing *trans* bound CEACAM1, was found to depend on the availability of activated calmodulin (CaM_4; Fig. 3C). However, counter-intuitively, in the absence of rafts, the total amount of CEACAM1 localized to the membrane exceeded that of the raft model (Additional file 7: Fig. S7). One possible explanation is that the preferential accumulation of monomeric CEACAM1 in the rafts may increase the relative concentration and promote self-association into homodimers, thus reducing the amount of monomeric CEACAM1 available for phosphorylation by Lck. Where such unexpected behaviors emerge, there is an opportunity to question the model and its assumptions. Here, possible refinements could include changes in the relative rates of CEACAM transport, *trans*-binding or, the incorporation of additional components modeling the activation and partitioning of Lck into lipid rafts.

To further explore this behavior, we examined the impact of CEACAM1 dimerization rate constants in both the raft and no raft models, on CEACAM1 clustering (Fig. 3D). For most simulations, the number of clustered CEACAM1 molecules (defined as monomers of CEACAM1 that have been transiently and reversibly immobilized on the membrane—mimicking *trans*-binding) increased over the course of the simulation. Unlike the no raft model in which clustering of CEACAM1 was sensitive to the dimerization rate constants, such constants had minimal impact in the raft model. This suggests that for CEACAM1 signaling, in addition to regulating local concentrations of Lck, lipid rafts may also play an important role in modulating the ability of CEACAM1 to form clusters, reducing its sensitivity to reaction kinetics.

Discussion

We present Cell4D, a novel tool for the spatiotemporal simulation of biological processes and pathways. Combining deterministic, cellular automata for the efficient diffusion of small molecules (as concentrations within a virtual lattice), with probabilistic Brownian motion of discrete molecules, Cell4D is a feature-rich and flexible program that can accurately simulate a variety of biological pathways. Cell4D is among the few cell simulation programs that can produce biologically accurate simulations while providing a graphical output of the model which we consider to be an essential feature in a tool designed to be used by non-computational, biological domain experts for hypothesis generation. Alternative, popular simulators in this space include Smoldyn [10] and VCell [31]. Smoldyn is more established, currently offering a wider variety of modeling options than Cell4D, but this comes at the expense of usability. Compared to Cell4D's graphic model building tool and underlying user readable XML model files, Smoldyn data models are formatted in plain text using a syntax that may be unintuitive for users inexperienced with programming. Also, Smoldyn simulation outputs are not enabled by default. Rather, it is up to the user to define the data they are interested in analyzing, then create an output format so that it can be processed and analyzed downstream. VCell is also feature-rich, offering a variety of model types and a graphic interface for model building making it a friendly choice for beginning modelers. However, spatial-stochastic models do not produce simulation data output, a crucial feature for users who

are ultimately interested in quantitative analysis of their model systems. Cell4D bridges this gap by providing both graphical and non-graphical modes, the latter capable of generating quantitative results in the absence of graphical overhead. Documentation is also provided for advanced users who wish to use Cell4D in parallel on high performance computing clusters.

In our initial simulations, we validated the ability of Cell4D to accurately model molecular diffusion and reactions, the latter exploiting the Andrews-Bray (AB) adjustment of the Smoluchowski method [26]. Next, we applied Cell4D to model two complementary aspects of CEACAM1-mediated signaling: a calcium-calmodulin interaction model and a CEACAM1 localization model. Previously it has been suggested that monomeric CEACAM1 cluster at the cell surface to help amplify CEACAM1-mediated signaling, potentially through the formation of a lattice-like arrangement of trans-dimers involving CEACAM1 monomers on neighbouring cells [16, 21, 23]. Our simulations support the requirement for local and transient spikes in $[Ca^{2+}]$ to activate primed calmodulin, and subsequently predict a dependence of CEACAM1 cluster size and surface concentration on active calmodulin. Furthermore, we showed that competing mechanisms have the potential to influence CEACAM1 clustering including the sequestration of Lck and CEACAM1 within lipid rafts, which at the same time may require a threshold amount of CEACAM1 to maintain the formation of clusters. In addition to representing an important biological process, with the involvement of molecular diffusion, spatial compartmentalization, active transport, reactions and state changes, CEACAM1 signaling represents a suitably complex system to test the functional capabilities of Cell4D.

It has been previously shown that CEACAM1 clusters in lipid raft regions are predominantly monomeric, while CEACAM1 outside of lipid rafts primarily exists in a dimeric state [32]. To replicate this behavior in Cell4D, we defined microdomains, representing lipid rafts, to preferentially accumulate monomeric CEACAM1 (and Lck) and exclude dimeric CEACAM1. Furthermore, in our model CEACAM1 monomers have a low probability of spontaneously becoming immobilized (representing a trans-binding event). This combination of accumulation and immobilization amplifies the localization of monomers within lipid-rafts. Furthermore, through increasing the local concentration of CEACAM1 monomers, clustering is promoted. Thus, the presence of lipid raft regions in the Cell4D models creates a spatial dynamic that matches the known observation of CEACAM1 clusters being primarily monomeric within lipid rafts, and dimeric CEACAM1 being localized in lipid-disordered regions outside of rafts [33].

We acknowledge that our models are based on hypothesized mechanisms as well as being subject to inevitable modeling constraints. For example, since calcium ions diffuse quickly and interact with calmodulin at fast kinetic rates (10^7 to $10^{10} \text{ M}^{-1} \text{ s}^{-1}$), the system resolves quickly requiring less simulation time. However, these kinetic rates necessitate a high-resolution time scale to avoid missing reactions between timesteps. In comparison, the localization model involving CEACAM1 transport, with molecular species that have slower diffusion rates and slower reaction rates (10^4 to $10^7 \text{ M}^{-1} \text{ s}^{-1}$), require more simulation time to reach equilibrium. The solution was the creation of separate, complimentary models, which apart from demonstrating Cell4D's modeling flexibility, allowed each part of the pathway to be simulated with appropriate simulation time and space scales.

We further note that the simulation of vesicle formation and transport of CEACAM1 between compartments was simplified using a single, probabilistic transport event. Future modifications of Cell4D are anticipated to support dynamic compartments (i.e., compartments which move, merge with, or bud from other compartments). In initial experiments, we examined the inclusion of clathrin adaptor protein complexes, AP1 and AP2, as discrete point particles in the cytoplasmic and cytosolic interface regions. However, the approach of modeling clathrin adaptor binding as bimolecular reactions had two major limitations. First, binding of an adaptor to CEACAM1 does not directly transport the molecule to its destination. Instead, transport requires a host of additional proteins to recruit clathrin and promote the formation of vesicles containing multiple copies of CEACAM1 [34]. Second, following assumptions for Smoluchowski reactions, reactions in Cell4D occur at rates that are diffusion-limited. Given that vesicles form over a timescale of the order of seconds, their formation violates this assumption. Although Cell4D could simulate sufficient CCV transport events to reach system equilibrium, an implementation of these events would require simulation lengths of hundreds of millions of timesteps, requiring extended computational run-times. As an example, the CEACAM1 lipid raft model described above were performed for 200,000 timesteps, taking approximately 5 h on a single Intel 80-thread CPU running at 2.4 GHz.

Computational models can be a powerful tool for understanding biological systems. Ideally, models maintain a balance where details that have a low impact on system behavior are ignored or abstracted away, while more critical aspects are preserved, producing models robust to input parameters and capable of emergent behaviors that provide novel insights and testable hypotheses. Such tools, we argue, are best constructed and interpreted by cell biologists who are the domain experts. For this we built Cell4D to be a user-friendly, feature-rich, and flexible framework for users to develop complex pathway models and generate large amounts of simulation data. Further, we developed a graphic interface to allow visualization of Cell4D simulations in real time. Such visualizations enhance a user's understanding of the system being modeled which is especially useful when observing the impact of changing parameters. For example, visualization of our CEACAM1 system allows users to see how increasing concentrations of calmodulin result in increased surface accumulation of CEACAM1. Examples of simulation visualizations are provided as movies on the project GitHub site: <https://github.com/ParkinsonLab/cell4d>.

The current Cell4D modeling environment is implemented through collections of *c*-voxels allowing the definition of arbitrary shapes representing different e.g. compartments. In the future, we hope to expand on this capability by providing a sculpting tool to make the process of defining biologically shaped compartments more practical. Beyond this, we suggest that the most significant improvements to be made in terms of simulating a 'real biology environment' within this modeling paradigm will involve improving efficient use of memory and compute resource, enabling the simulation of models that are larger and/or higher resolution. Nevertheless, we stress that simulations need only be sufficiently detailed to observe a given phenomenon in order to have empirical benefit (Additional file 8).

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-024-05739-0>.

Additional file 1: Fig. S1. RMSD values of Cell4D particle and bulk diffusion across multiple timescales.

Additional file 2: Fig. S2. RMSD of Cell4D bulk molecule diffusion and particle diffusion across space scales.

Additional file 3: Fig. S3. Reaction products generated from unimolecular and two-particle bimolecular reactions over time.

Additional file 4: Fig. S4. Summary of bimolecular products over time of bulk-particle reactions using a Smoluchowski-based method to calculate reaction probability.

Additional file 5: Fig. S5. Summary of two-particle bimolecular products over time using the Andrews-Bray adjusted Smoluchowski method to calculate reaction radii.

Additional file 6: Fig. S6. Saturation of calmodulin under varying calcium concentrations in well-mixed or microdomain conditions.

Additional file 7: Fig. S7. Comparison of surface CEACAM1 clustering in both Cell4D model variants.

Additional file 8: Text. Detailed Methods.

Acknowledgements

Thanks to D. Han and K. Keerthisingham for additional coding support and N. Liu for software testing. We are grateful to A. Funahashi for advice on XML schema.

Author contributions

D.C. wrote code, designed and performed experiments. G.L.C. helped design experiments and oversaw code development. B.T. wrote code and supported software development. G.L.C. and J.P. supervised the project. D.C., G.L.C. and J.P. wrote the manuscript. All authors reviewed the manuscript. J.P. conceived the project.

Funding

This work was supported by the University of Toronto's Medicine by Design initiative, funded by the Canada First Research Excellence Fund (CFREF) and the Natural Sciences and Engineering Research Council (RGPIN-2019-06852).

Availability of data and materials

Project name: Cell4D. Project home page: <https://github.com/ParkinsonLab/cell4d>. Operating system(s): Linux. Programming language: C++. Other requirements: freeglut3, libsbml5, libwxgtk3.0. License: Custom—see Project home page. Any restrictions to use by non-academics: Not for commercial use.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 13 September 2023 Accepted: 11 March 2024

Published online: 21 March 2024

References

1. Passi A, Tibocha-Bonilla JD, Kumar M, Tec-Campos D, Zengler K, Zuniga C. Genome-scale metabolic modeling enables in-depth understanding of big data. *Metabolites*. 2021;12(1):14.
2. Seaver SMD, Liu F, Zhang Q, Jeffryes J, Faria JP, Edirisinghe JN, Mundy M, Chia N, Noor E, Beber ME, et al. The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Res*. 2021;49(D1):D575-d588.
3. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41(Database issue):D991-995.
4. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419.
5. Karlsson M, Zhang C, Méar L, Zhong W, Digre A, Katona B, Sjöstedt E, Butler L, Odeberg J, Dusart P, et al. A single-cell type transcriptomics map of human tissues. *Sci Adv*. 2021;7(31):eabh2169.
6. Oughtred R, Rust J, Chang C, Breitkreutz BJ, Stark C, Willems A, Boucher L, Leung G, Kolas N, Zhang F, et al. The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci*. 2021;30(1):187–200.

7. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirschs R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 2021;49(D1):D605–d612.
8. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, Griss J, Sevilla C, Matthews L, Gong C, et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* 2022;50(D1):D687–d692.
9. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 2016;44(D1):D457–462.
10. Andrews SS. Smoldyn: particle-based simulation with rule-based modeling, improved molecular interaction and a library interface. *Bioinformatics.* 2017;33(5):710–7.
11. Kerr RA, Bartol TM, Kaminsky B, Dittrich M, Chang JC, Baden SB, Sejnowski TJ, Stiles JR. Fast Monte Carlo simulation methods for biological reaction–diffusion systems in solution and on surfaces. *SIAM J Sci Comput.* 2008;30(6):3126.
12. Resasco DC, Gao F, Morgan F, Novak IL, Schaff JC, Slepchenko BM. Virtual Cell: computational tools for modeling in cell biology. *Wiley Interdiscip Rev Syst Biol Med.* 2012;4(2):129–40.
13. Sanford C, Yip ML, White C, Parkinson J. Cell++—simulating biochemical pathways. *Bioinformatics.* 2006;22(23):2918–25.
14. Czech J, Dittrich M, Stiles JR. Rapid creation, Monte Carlo simulation, and visualization of realistic 3D cell models. *Methods Mol Biol.* 2009;500:237–87.
15. Morales VM, Christ A, Watt SM, Kim HS, Johnson KW, Utku N, Texeira AM, Mizoguchi A, Mizoguchi E, Russell GJ, et al. Regulation of human intestinal intraepithelial lymphocyte cytolytic function by biliary glycoprotein (CD66a). *J Immunol.* 1999;163(3):1363–70.
16. Gray-Owen SD, Blumberg RS. CEACAM1: contact-dependent control of immunity. *Nat Rev Immunol.* 2006;6(6):433–46.
17. Boulton IC, Gray-Owen SD. Neisserial binding to CEACAM1 arrests the activation and proliferation of CD4⁺ T lymphocytes. *Nat Immunol.* 2002;3(3):229–36.
18. Sadarangani M, Pollard AJ, Gray-Owen SD. Opa proteins and CEACAMs: pathways of immune engagement for pathogenic Neisseria. *FEMS Microbiol Rev.* 2011;35(3):498–514.
19. Kim WM, Huang YH, Gandhi A, Blumberg RS. CEACAM1 structure and function in immunity and its therapeutic implications. *Semin Immunol.* 2019;42:101296.
20. Tchoupa AK, Schuhmacher T, Hauck CR. Signaling by epithelial members of the CEACAM family—mucosal docking sites for pathogenic bacteria. *Cell Commun Signal.* 2014;12(1):27.
21. Patel PC, Lee HS, Ming AY, Rath A, Deber CM, Yip CM, Rocheleau JV, Gray-Owen SD. Inside-out signaling promotes dynamic changes in the carcinoembryonic antigen-related cellular adhesion molecule 1 (CEACAM1) oligomeric state to control its cell adhesion properties. *J Biol Chem.* 2013;288(41):29654–69.
22. Chen Z, Chen L, Qiao SW, Nagaishi T, Blumberg RS. Carcinoembryonic antigen-related cell adhesion molecule 1 inhibits proximal TCR signaling by targeting ZAP-70. *J Immunol.* 2008;180(9):6085–93.
23. Driouchi A, Gray-Owen SD, Yip CM. Correlated STORM-homoFRET imaging reveals highly heterogeneous membrane receptor structures. *J Biol Chem.* 2022;298(10):102448.
24. Fick A. On liquid diffusion. *Lond Edinb Dublin Philos Mag J Sci.* 1855;10(63):30–9.
25. Pletcher RH, Tannehill JC, Anderson D. *Computational fluid mechanics and heat transfer.* 2nd ed. Boca Raton: CRC Press; 1997.
26. Andrews SS, Bray D. Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Phys Biol.* 2004;1(3–4):137–51.
27. Faas GC, Raghavachari S, Lisman JE, Mody I. Calmodulin as a direct detector of Ca²⁺ signals. *Nat Neurosci.* 2011;14(3):301–4.
28. Joseph N, Reicher B, Barda-Saad M. The calcium feedback loop and T cell activation: how cytoskeleton networks control intracellular calcium flux. *Biochim Biophys Acta.* 2014;1838(2):557–68.
29. Lewis RS. Calcium signaling mechanisms in T lymphocytes. *Annu Rev Immunol.* 2001;19(1):497–521.
30. Ventimiglia LN, Alonso MA. The role of membrane rafts in Lck transport, regulation and signalling in T-cells. *Biochem J.* 2013;454(2):169–79.
31. Loew LM, Schaff JC. The virtual cell: a software environment for computational cell biology. *Trends Biotechnol.* 2001;19(10):401–6.
32. Driouchi, A. Characterization of CEACAM1 spatial distribution, self-association and dynamics using high-resolution microscopy techniques. Ph.D. Thesis. Toronto, ON, Canada: University of Toronto; 2019.
33. Driouchi A, Giuliani M, Gray-Owen S, Yip CM. Characterization of CEACAM1 and lipid raft nanoclustering, association and structure by dSTORM and homo-FRET imaging. *Biophys J.* 2016;110(3):205a.
34. Kononenko NL, Haucke V. Molecular mechanisms of presynaptic membrane retrieval and synaptic vesicle reformation. *Neuron.* 2015;85(3):484–96.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.