Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

A MULTI-SCALE, PHYSICS ENGINE-BASED SIMULATION OF CELLULAR MIGRATION

Terri Applewhite-Grosso

Department of Computer Science
The Graduate Center
The City University of New York
365 5th Avenue
New York, NY 10016, USA

Nancy Davis Griffeth
Elisa Lannon

Department of Mathematics and Computer Science
Lehman College - The City University of New York
250 Bedford Park Boulevard West
Bronx, NY 10468, USA

Uchenna Unachukwu
Stephen Redenti
Naralys Batista

Department of Biological Sciences
Lehman College - The City University of New York
250 Bedford Park Boulevard West
Bronx, NY 10468, USA

ABSTRACT

This research paper describes the design and prototyping of a simulation tool that provides a platform for studying how behavior of proteins in the cell membrane influences macro-level, emergent behaviors of cells. Whereas most current simulation tools model cells as homogeneous objects, this new tool is designed to modularly represent the cell's complex morphology and the varying distribution of proteins across the membrane. The simulation tool uses a physics engine to manage motion and collisions between objects. It also represents dynamic fluid environments, experimental surfaces, attachment bonds and interactions between the dynamically changing cell surface proteins. The prototype tool is described along with proposals for its use and further development.

1 INTRODUCTION

Modeling and simulation have become increasingly important recently in the field of systems biology, which aims to understand biological processes as whole systems instead of as isolated parts. This paper describes a new tool for simulating the migration of cells. The tool adapts existing technologies, including physics engines and rendering engines developed for computer graphics. It supports models that describe molecular bindings, the forces generated by the bindings, the composition of the membrane, and the shape of the cell, all of which influence the motion of a cell. Cell migration can then be studied as an emergent behavior arising from a model of activity on the membrane of the cell.

Experimental biologists require better tools to provide simulations based on experimentally-validated models in order to understand the relationship between low-level behaviors and emergent system properties such as cell migration. Our team is working with an experimental lab team at Lehman College to develop a simulation tool that addresses their needs. Challenges to developing this tool include:

- Bridging the gap between biologists and computer scientists.
- Finding collisions between objects.
- Modeling the forces produced by and after a collision, including changes in directions of objects and binding of molecules.
- Modeling large numbers of objects efficiently.
- Modeling complex cell shapes.
- Modeling variations in the composition of the cell membrane.
- Modeling the reactions of proteins to collisions with molecules (such as ligands) and other cells membranes.
- Visualizing the cells' behavior.
- Defining a model that can represent a large variety of proteins and other influences on the cells' behavior.

The contributions of this paper include:

1. Description of a simulation tool using both existing technologies and new models to address the above challenges.
2. A new model of the cell membrane providing for both changes in cell shape and variability in the composition of the membrane.
3. A new model of physical interaction of the cell with its environment, based on ligand gradients in the environment, receptors in the cell membrane, and the forces imposed by binding of membrane proteins to biological structures (such as other cells and the intercellular matrix).

2 THEORETICAL BACKGROUND

There are cellular simulations that model large numbers of cells as a continuous fluid which can be represented with a series of differential equations that explain how that fluid changes over time. Unlike these high level simulations, this project aims to model cells at a much smaller scale and represent them as composites of membranes, cytoskeleton and surface proteins. Therefore, the comparable research involves simulations that treat cells as individual units or composites of smaller units. Current simulation tools model cells as homogeneous objects. This new tool represents both the variable shape of the cell and the differences between how the disparate areas of the cell react to the external environment.

Using a lattice is an efficient approach to the computationally intensive process of determining collisions and distances between objects. A popular lattice-based model is the Cellular Potts Model (CPM). The CPM is built on a two- or three-dimensional lattice. Each lattice site is labeled with an integer identifier, called the spin, that defines what type of object it belongs to - extracellular matrix (ECM), cell or interstitial fluid, for example. An example is given in Figure 1. Spin values have also been used to represent different internal parts of the cell, such as the nucleus and the cytosol (Scianna, Preziosi, and Wolf 2013). The energy of the system is determined by a set of rules defined for all pairs of spin values. In each time step, a random selection of lattice sites have their spin changed to the spin of a randomly selected adjacent site. If this change reduces the overall energy of the system, the spin change is accepted. If it increases the energy, there is still a probability that the change will be accepted. Over time, the system tends towards lower energy. This model has been used to study, among other cellular processes, cellular migration, angiogenesis and tumor invasion (Bauer, Jackson, and Jiang 2007, Beltman, Marée, Lynch, Miller, and de Boer 2007, Scianna, Preziosi, and Wolf 2013). One advantage of this approach is that computation time is based solely on the size of the lattice, not on the number of cells or other objects represented. (Note that this is not an advantage when modeling a small number of objects.) The CPM can also represent varying cell shapes (within the confines of the grid). It cannot, however, represent each cell lattice site as having its own individual properties - that would require a different spin number for each individual lattice site in a single cell and would significantly increase the memory requirements for the model. Two different

lattice sites that are marked with the spin value for cytosol, will have the same characteristics whether they are in, for example, the front or the rear of the cell, while we know that these parts of the migrating cell behave in very different ways.

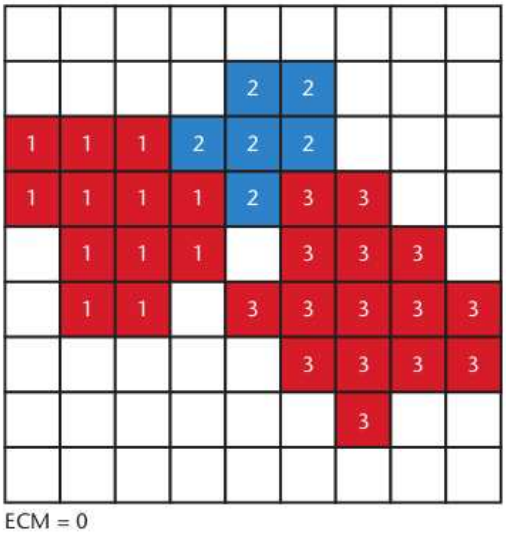

Figure 1: Example of lattice used for the Cellular Potts Model. (Cickovski, Aras, Alber, Izaguirre, Swat, Glazier, Merks, Glimm, Hentschel, and Newman 2007)

Simulations that do not use a lattice tend to represent cells as having simple shapes - spheres or ellipses. In some of these simulations, the shapes of these simulated cells can be changed in response to forces provided by other cells and the external environment. Each cell in the simulation can have its own levels of flexibility in different dimensions, so they can be treated as individuals. The calculations involved in determining collisions and distances between these simple shapes can be implemented efficiently and allow many cells to be represented. The responses of large numbers of cells to their environment can be determined if the shapes of the cells are restricted in this way. An example is given in Figure 2. The scale of these simulations go from the cellular level to that of tissues. (Palsson 2007)

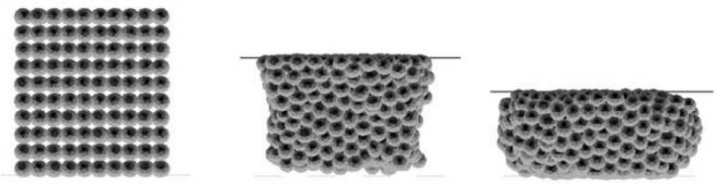

Figure 2: Example of ellipsoidal cells under pressure. (Palsson 2007)

While these versions of cells represent each individual cell as having its own particular characteristics, they still are not able to model varying characteristics in different parts of the same cell. This is problematic because the cell is a heterogeneous body. The shape can be extremely complex and the distribution of proteins varies across the surface and inside of the cell. Therefore, different parts of the cell will react differently to external stimulation. It is this kind of variable response to the environment that causes cells to migrate and that is what this research is hoping to show. The goal is to represent the level of the molecular bonds in order to elucidate how the dynamics of those bonds lead to cellular-level behavior.

3 THE SIMULATION TOOL

3.1 Using a Physics Engine to Model the Motion and Forces Affecting the Cells

Physics Engines were developed to apply physics to objects in a virtual world. A game engine is a physics engine that attempts to model the motion of objects in real time. These engines, once given a representation of the bodies in the virtual world, can quickly perform collision detection, use application-defined constraints to resolve the multiple forces that are acting on the bodies and generate data for visualizing the scenes in real time. Current graphics algorithms are used to make the movement in video games appear to be realistic and seamless. All of these abilities of physics engines will be used to make the simulation program fast and efficient. How the simulated cells respond to the collisions and forces that the physics engine detects must be defined by the application designer. How these responses are used in this simulation to represent the bonds that form between the cell and its environment and the changes in the surface proteins is elaborated in Sections 3.4 and 3.5.

The chosen physics engine is JBullet, a Java port of the Bullet physics engine. Because it is built using Java, it is a cross-platform solution. JBullet supports rigid shapes, static and moving triangular meshes and multiple types of constraints. JBullet is used to determine when and where cells are colliding with each other and with objects in the environment. It models the buoyancy supplied by the fluid in which the cells are moving. It is used to determine the amount of force that is being applied to a set of molecular bonds in order to find out when they should break. And finally, its binding to OpenGL (a cross-platform 3D rendering API) is used to create images and video of the simulation to assist with experimental analysis.

3.2 Representing the Fluid Environment

Ligand assays are generally conducted in a water-based fluid. This fluid can contain specific concentrations or gradients of diffusible ligands which can then bind to the cell's receptors. As ligands diffuse, their concentrations change in both time and space. Therefore, the development of concentration gradients is modeled using partial differential equations (Kong, Able, Dudu, and Vazquez 2010). The Finite Element Method (FEM) is a numerical approach that can provide an approximate solution. As there is no open-source Java library for solving partial differential equations, they can be solved using Matlab's FEM libraries.

3.3 Representing Surfaces

A goal of this simulation is to represent the interactions of surfaces - the cell's membranes, the extra-cellular matrix (ECM), and the experimental containers. In the laboratory, the cell's environment can include plates or dishes coated with adhesion molecules, artificially created scaffolds with varying biophysical properties, and various ligands, possibly dissolved in solution, maintained in a concentration gradient or attached in various ways to different surfaces. *In vivo*, they are often surrounded with a heterogeneous system of ECM fibers. All of these different components must be able to be represented by the program.

Because the surfaces of these different components can vary in shape, the best way to represent them for a physics engine is to use a triangular mesh. This kind of mesh is frequently used in games to represent dynamic objects with complex shapes. An additional advantage this data structure provides is the ability to define different characteristics for each triangle of the surface. It allows heterogeneous distributions of proteins and differing rates of change to be represented over a surface. The size of the triangles can be adjusted to get an appropriate level of granularity. The color of the triangles can also be used in the visual representation of the simulation. Each triangle can be colored to show the surface concentrations of different proteins as they change over time. In addition, by using different numbers of triangles for different levels of granularity, as in Figure 3, the triangular mesh can represent all of the varied shapes that an object can take on.

Take the cell itself as an example. The simulation represents the cell's shape with a mesh of triangles. If we are looking at several surface membrane proteins, each triangle can keep track of the number of those

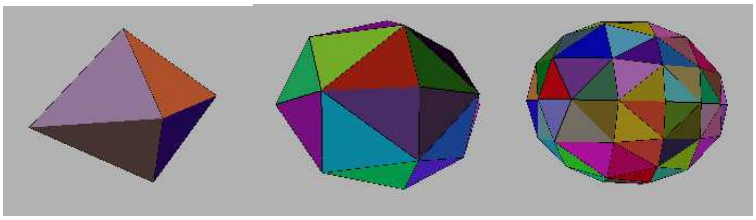

Figure 3: Levels of granularity for the cell's triangular mesh.

proteins which are on the surface and how those numbers are changing over time. This mesh can also represent the multitude of shapes that the cell can take on. In some circumstances the cells migrate using a spherical shape, called amoeboid migration, whereas at other times, they move by generating pseudopodia and lamellipodia, called mesenchymal migration, as can be seen in Figure 4. As the forces on the cell faces change over time, the number and shape of the triangles in the mesh can adjust to reflect the changes in the cell's shape. This allows the cell's shape to also be an emergent property of the model.

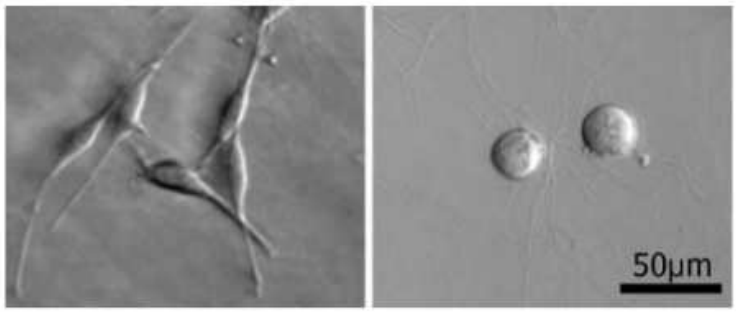

Figure 4: Mesenchymal and Amoeboid migration. (Panková, Rösel, Novotný, and Brábek 2010)

3.4 Representing Bonds

When the surfaces of cells and walls approach each other within a minimum distance, the proteins and receptors are able to bind to each other. (See Figure 5) Constraints, as defined in JBullet, are used to represent these bonds. The constraints can also be displayed in the images. A constraint's parameters allow its ability to rotate to be adjusted and/or restricted.

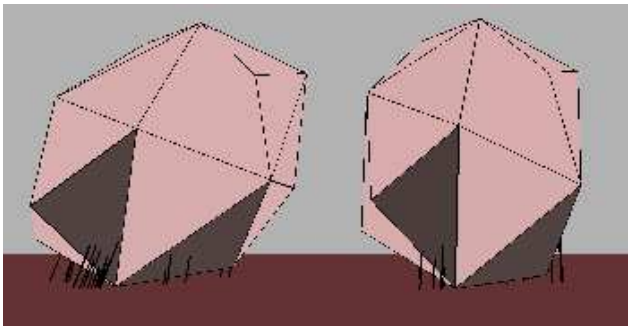

Figure 5: Example of cells bonded to a surface.

Since using one constraint for each molecular bond would be prohibitively expensive in terms of computational efficiency, each constraint represents a user-defined number of molecular bonds. When two surfaces with compatible proteins come within a minimal distance, the simulation tool creates the resulting

constraints. The algorithm to define the number of bonds/constraints generated is stochastic as molecular bonding is inherently probabilistic. The position of the bonds across the surfaces is randomly decided.

The constraints also break probabilistically. The probability of breakage is determined by the age of the constraint and the forces applied to it. When cells bind to surfaces coated with ECM proteins, if the bonds last long enough, they develop into focal adhesions (FA) which are large, stable complexes of proteins that anchor the cell to the ECM and whose forces on the cell's cytoskeleton transmit information to the cell. These complexes are strong, but also degrade over time. These changes in the flexibility and strength of the bonds are reflected in the constraint's parameters and in the probability of breakage.

3.5 Representing Protein Interactions

In order to model the interactions between the molecules on the cell membrane and those in the environment, the simulation needs to accurately represent the types and numbers of molecules on the triangular segments of the cell's surfaces as well as how the numbers of those molecules change over time. These changes can be represented by a system of differential equations that describe the binding of ligands to receptors and the internalization and secretion of receptors to the surface of the cell. Thus treating the triangular segments as hybrid automata is the most natural model for this behavior: each triangular segment has a state, determined by the number of proteins (receptors and integrins) in the membrane represented by that segment, and transitions are determined by the differential equations controlling the changes in the numbers of proteins (Lynch, Segala, and Vaandrager 2003). Shankaran, Resat, and Wiley (2007) present a system of equations that describe how the number of surface receptors changes over time using the number of unbound receptors (along with their internalization and secretion rates), the number of bound receptors (with their internalization rates), and the rates at which the receptors bind and release ligand. This system of equations is used in the simulation to determine the surface density of the membrane proteins and it is primarily the change in these internalization and secretion rates that creates the changes in the molecular interactions that happen at the cell's surface - as the number of receptors that bind to the ECM (such as integrins) changes over time, the number and strength of available bonds will also change.

The relationships between surface proteins can be quite complex. When a surface receptor binds to a ligand, it changes the cell's behavior by initiating localized protein cascades inside the cell that then change the trafficking rates of other surface proteins. For example, binding of receptor A can cause receptor B to be internalized at an increased rate, thereby reducing the number of receptor B proteins at that location of the cell's membrane. Alternatively, it could decrease the internalization rate. Or it could change the rate at which receptor B is excreted to the surface. In addition, a ligand binding to receptor B could cause some change in rates for receptor A. Finally, enough of a change in a given protein inside the cell can cause the cell to produce (or to stop producing) other proteins. All of these interactions have to be represented in the simulation. Note that while the cascade of proteins that are generated by the binding of a receptor are important, this simulation focuses on the resulting changes in the surface density of the proteins found in the same vicinity (specifically on the same triangular segment).

In the simulation, an interaction is defined as having a Receptor, the protein that binds to the ligand, and a Target, the protein whose rate of internalization or secretion is being affected. There is evidence that the response of Target proteins is associated with the number of Receptor molecules that are bound to ligand molecules (Knauer, Wiley, and Cunningham 1984). This number is used to determine the internalization and secretion rates of the target surface proteins.

Many biological reactions, including the response to receptor binding, can be represented by a logistic curve as in Figure 6. Below an activation point, there is no response to binding of the receptor. Then there is an area of the curve where increasing the number of bound receptors increases the effect on the Target. Finally, at or above saturation levels, the rate change has reached its maximum value and increasing the number of bound receptors does not change the rate. The function for this curve,

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}},$$

is computationally intensive. Since this response occurs on multiple segments of membrane, across multiple cells, using a more efficient approximation is advantageous. A simpler approximation of this curve has a straight line representing the curve at the activation stage and can be represented with four parameters - the baseline rate, the maximal response rate, the activation point and the saturation point as shown in Figure 7 (Ye, Entcheva, Grosu, and Smolka 2005). Note that the maximal response rate could be either greater than or less than the baseline rate, depending on whether binding of the receptor increases or decreases the target protein's rate of change.

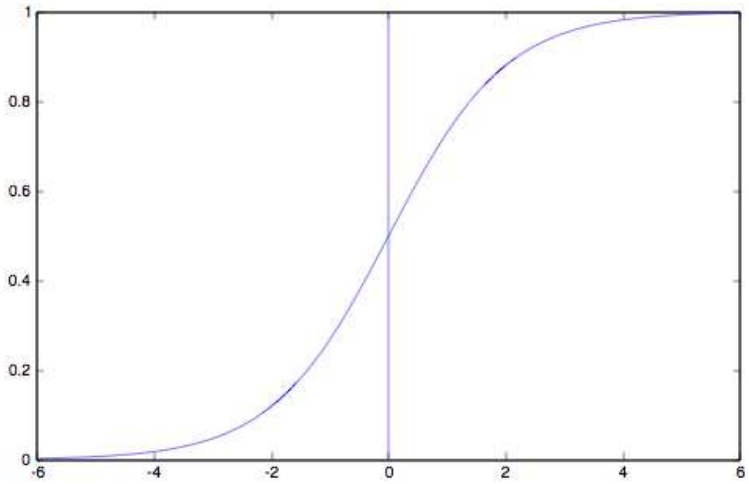

Figure 6: Graph of basic logistic function.

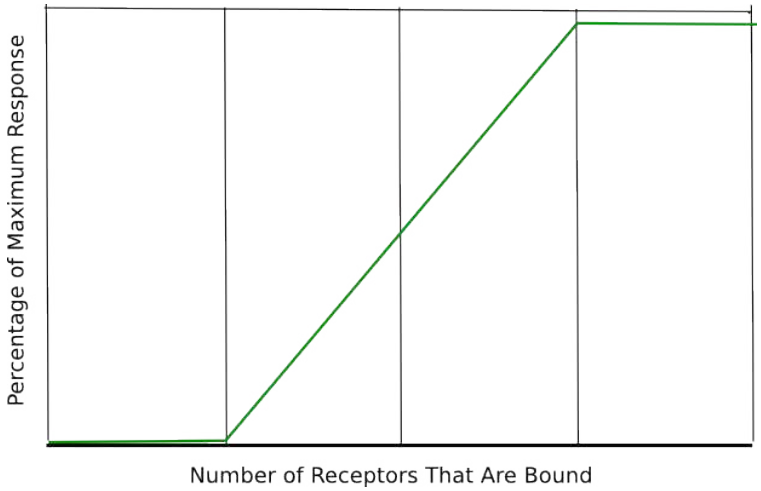

Figure 7: General response curve of internalization or secretion rates.

4 EFFECTS ON TOOL PERFORMANCE

The most time-consuming aspect of the model comes from displaying the objects on the screen. While the physics engine is optimized to perform the coloring, shading and rendering, these processes take time to generate an image. The tool can be used with or without visualization because users can have different requirements at different times. If the goal is simply to collect numerical data, visualizing the cell's motion may not be very important and the simulation can be set to run quickly without generating images. If,

however, the user wants to view how the number of surface molecules change over time and how the cells actually move, the simulation can be run with images.

Several parameters can affect the frame rate of the simulation. As cells are added to the simulation, it takes time to process the interactions between each other and the walls. Because the triangular mesh is treated as a single object, using cells with more segments does not have a large effect on calculating collisions and movement. But the number of molecules and the trafficking rates of each segment must be updated in each time step and that also takes time. So using more detailed cells will make each frame take longer to process.

These parameters will also affect the amount of required memory usage. The effects of increasing the number of cells, segments or proteins on memory are relatively straightforward. Each segment requires memory to keep track of the number of bound and unbound molecules of each surface proteins along with the rates of change for each protein interaction. Adding more cells, increasing the number of segments of each cell and increasing the number of surface proteins and protein interactions will require more memory usage.

5 CONCLUSION

A prototype of the simulation that uses cells whose membranes are represented as near-spherical triangular meshes has been developed. The current version simulates cells in a microfluidic channel responding to gradients of Epidermal Growth Factor (EGF), modeling the experiment found in Unachukwu, Sauane, Vazquez, and Redenti (2013). The changes in number of integrins and EGF Receptors found on the cells' surfaces have also been modeled. These virtual cells form bonds between their membranes and the walls of the channel and migrate by rolling through the channel. This movement in an environment with no external ligand is similar to that of cells in the laboratory.

The next step is to match the virtual cells' behavior to that of the cells in the laboratory by using parameter estimation (Korsunsky, McGovern, LaGatta, Olde Loohuis, Grosso-Applewhite, Griffeth, and Mishra 2014). The value of parameters for particular proteins and their interactions will be estimated by comparing the simulation results to experimental data collected on Retinal Progenitor Cells by John Uchenna in the Redenti Lab at Lehman College. While some of the parameters used in the simulation are taken from experimentally determined data, some will be found by determining the range of values that produce behavior similar to that found in the lab. Once these parameters have been determined for the experiment using EGF, the simulation will model a similar experiment using stromal cell-derived factor 1 (SDF-1), another protein being tested in the Redenti Laboratory. In addition, because the external environment and gradients are user-defined, this simulation tool can also be used to model cell reactions to non-chemical gradients, like gravitational, electrical and magnetic fields.

Finally, a user interface that will not require significant programming on the part of the user is being developed. The target audience for the tool is researchers in the biological fields who are not necessarily familiar with programming. The ability of these researchers to use this simulation tool to model novel assays and experiments will be a significant contribution to their work and to the research community.

REFERENCES

Bauer, A. L., T. L. Jackson, and Y. Jiang. 2007. "A Cell-Based Model Exhibiting Branching and Anastomosis During Tumor-Induced Angiogenesis". *Biophysical Journal* 92 (9): 3105–21.

Beltman, J. B., A. F. Marée, J. N. Lynch, M. J. Miller, and R. J. de Boer. 2007. "Lymph Node Topology Dictates T Cell Migration Behavior". *The Journal of Experimental Medicine* 204 (4): 771–780.

Cickovski, T., K. Aras, M. S. Alber, J. A. Izaguirre, M. Swat, J. A. Glazier, R. M. Merks, T. Glimm, H. G. E. Hentschel, and S. A. Newman. 2007. "From Genes to Organisms Via the Cell: A Problem-Solving Environment for Multicellular Development". *Computing in Science & Engineering* 9 (4): 50–60.

Knauer, D. J., H. S. Wiley, and D. D. Cunningham. 1984. "Relationship Between Epidermal Growth Factor Receptor Occupancy and Mitogenic Response. Quantitative Analysis Using a Steady State Model System.". *Journal of Biological Chemistry* 259 (9): 5623–5631.

Kong, Q., R. A. Able, V. Dudu, and M. Vazquez. 2010. "A Microfluidic Device to Establish Concentration Gradients Using Reagent Density Differences". *Journal of Biomechanical Engineering* 132 (12): 121012.1–9.

Korsunsky, I., K. McGovern, T. LaGatta, L. Olde Loohuis, T. Grosso-Applewhite, N. Griffeth, and B. Mishra. 2014. "Systems Biology of Cancer: A Challenging Expedition for Clinical and Quantitative Biologists". *Frontiers in Bioengineering and Biotechnology* 2:1–17.

Lynch, N., R. Segala, and F. Vaandrager. 2003. "Hybrid I/O Automata". *Information and Computation* 185 (1): 105–157.

Palsson, E. 2007. "A 3-D Deformable Ellipsoidal Cell Model with Cell Adhesion and Signaling". In *Single-Cell-Based Models in Biology and Medicine*, 271–299. Springer.

Panková, K., D. Rösel, M. Novotný, and J. Brábek. 2010. "The Molecular Mechanisms of Transition Between Mesenchymal and Amoeboid Invasiveness in Tumor Cells". *Cellular and Molecular Life Sciences* 67 (1): 63–71.

Scianna, M., L. Preziosi, and K. Wolf. 2013. "A Cellular Potts Model Simulating Cell Migration on and in Matrix Environments". *Mathematical Biosciences and Engineering* 10 (1): 235–61.

Shankaran, H., H. Resat, and H. S. Wiley. 2007. "Cell Surface Receptors for Signal Transduction and Ligand Transport: A Design Principles Study". *PLoS Computational Biology* 3 (6): e101.

Unachukwu, U. J., M. Sauane, M. Vazquez, and S. Redenti. 2013. "Microfluidic Generated EGF-Gradients Induce Chemokinesis of Transplantable Retinal Progenitor Cells Via the JAK/STAT and PI3kinase Signaling Pathways". *PLoS ONE* 8 (12): e83906.

Ye, P., E. Entcheva, R. Grosu, and S. A. Smolka. 2005. "Efficient Modeling of Excitable Cells Using Hybrid Automata". In *Proceedings of Computational Methods in Systems Biology*, Volume 5, 216–227.

AUTHOR BIOGRAPHIES

TERRI APPLEWHITE-GROSSO is a doctoral candidate in the Computer Science Department at the Graduate Center, City University of New York. She holds a B.S. in Biology from Temple University and an M.S. in Computer Science from The City College of the City University of New York. Her research interests lie in joining her background in molecular and cellular biology with simulation modeling. She is also a high school teacher with an M.A. in Secondary Science Education from Teachers College, Columbia University. Her email address is tgrosso@gradcenter.cuny.edu.

NANCY DAVIS GRIFFETH is a Professor of the Department of Mathematics and Computer Science at Lehman College and The Graduate Center of the City University of New York, New York, NY. She conducted research in databases, distributed systems, and computer networking prior to her work in computational biology. She was on the faculty at Georgia Tech from 1979 to 1988 and at Northwestern University Kellogg School of Management from 1976 to 1979. She received her Ph.D. degree from the University of Chicago in 1978 and her B.A. from Harvard in 1968. Her email address is nancy.griffeth@lehman.cuny.edu.

UCHENNA UNACHUKWU is currently a doctoral candidate in the Biochemistry program at the Graduate Center, City University of New York. His current research delineates molecular mechanisms governing photoreceptor progenitor cell migration following transplantation in diseased mammalian retinal models and he conducts his work in the laboratory of Dr. Stephen Redenti. Prior to his doctoral study, he received an M.A. in biology from Lehman College. An avid educator, Uchenna also holds an M.S. in teaching and adolescent education from Pace University and is currently an adjunct lecturer at various colleges in the City University of New York educational system. His email address is ucheuna@yahoo.com.

STEPHEN REDENTI is an Associate Professor of Molecular, Cell and Developemental Biology and Biochemistry at Lehman College of the City University of New York, Bronx, NY. He holds a Ph.D. in Neuroscience from Hunter College of the City University of New York and conducted postdoctoral work in Retinal Tissue Engineering at the Schepens Eye Research Institute at Harvard Medical School. His lab is currently focused on researching cellular development and regeneration of retinal tissue. His email address is stephen.redenti@lehman.cuny.edu.

NARALYS BATISTA received her Bachelors of Arts in Chemistry at Lehman College of the University of New York where her research interests included in-silico modeling of biological systems. She will be attending Columbia University College of Physicians and Surgeons. Her email address is naralys-batista@gmail.com.

ELISA LANNON is a Graduate Student of the Department of Mathematics and Computer Science at Lehman College of the City University of New York, Bronx, NY. She holds a B.A. in Mathematics and an M.A. in Mathematics from Lehman College. Her research interests lie in cultivating the field of Mathematics for its applications to the medical sciences, specifically nervous system disorders and injuries. Her email address is elisa.lannon@gmail.com.