The Biology of Interacting Things: The Intuitive Power of Agent-Based Models

By Alexander Gelfand
In the early 1990s, when James A. Glazier, PhD, first became interested in using agent-based modeling to simulate biological phenomena, the field was so new that he had to borrow ideas from the study of metal and soap.

**Times have changed:** Over the last 10 years, agent-based models (ABMs) have become an important component of the biomedical researcher’s toolkit.

By their nature, ABMs would seem to be a perfect fit for biology. First developed in the 1940s, they simulate complex systems by having autonomous virtual agents (cells, anyone?) interact with each other and their environment according to preprogrammed rules, often with a degree of built-in randomness. Yet when agent-based programming languages and modeling software came along in the 1990s, they were slow to gain traction in biomedical circles. Representing the many cells in a biological system using ABMs can be expensive compared to running biological simulations based on differential equations, and the mathematical techniques used to analyze and optimize equation-based models do not necessarily work on ABMs.

Yet ABMs do carry advantages for biomedical research. Among other things, they are intuitive, work well in three dimensions, and can reproduce complex behaviors with just a few simple (even incomplete) rules. Moreover, progress toward hybridizing ABMs with other approaches, such as differential equations, is making them more powerful than ever. These plusses, along with increased computing power, are helping biomedical applications of ABMs take off, as scientists use them to investigate everything from tumor formation to bacterial growth.

**Soap to Cells**

As a doctoral candidate in physics in the late 1980s, Glazier, who now directs the Biocomplexity Institute at Indiana University Bloomington, studied the evolution of bubbles in soap froth. The surprisingly broad implications of that work led him to collaborate with a group of researchers at Exxon who were using computational models derived from statistical physics to investigate the related phenomenon of grain formation and growth in metals. A few years later, while working as a post-doctoral fellow in Sendai, Japan, in the laboratory of Yasuji Sawada, PhD, Glazier met François Graner, PhD, who was studying the microscopic fresh-water creatures known as hydra. Hydra are renowned for their regenerative capabilities—chop them into hamburger, and the cells will rearrange themselves to form a whole...
new organism—and Graner wanted to test the hypothesis that cell adhesion allowed a hydra’s two main cell types to sort themselves into larger structures during regeneration. Glazier realized that a modified form of the model that he and the Exxon researchers had been using could also simulate cell-sorting by treating each cell as an individual unit subject to basic physical forces and constrained by a few rules. Paulien Hogeweg, PhD, a Dutch theoretical biologist at the University of Utrecht who helped coin the term “bioinformatics” in the 1950s, later elaborated on Glazier and Graner’s initial modeling efforts, adding biological mechanisms like cell differentiation and chemotaxis to create what became known as the Glazier-Graner-Hogeweg (GGH) model.

The GGH model was one of the first agent-based models designed specifically for biological purposes. Over the past two decades, scientists have used increasingly sophisticated forms of it to simulate multicell phenomena as diverse as wound healing, stem-cell differentiation, and skin pigmentation. Often, they have relied on CompuCell3D, an open-source modeling environment that Glazier and his collaborators Mark Alber, PhD, and Jesus Izaguirre, PhD, at the University of Notre Dame began developing in 2000 and whose development is currently led by Maciej Swat, PhD, at Indiana University. CompuCell3D is meant to help researchers concentrate on the biology behind their simulations rather than on the nuts and bolts of model building. To that end, the software allows users to select the cell types and behaviors they want from a series of drop-down menus. It also lets them add modules that use partial differential equations to describe the chemical fields that influence cell migration and differentiation, or ones that use ordinary differential equations to describe the dynamics of biochemical networks inside cells and the distribution of chemicals at the whole-body level. (Such hybrid models, which combine agent-based and equation-based methods for greater efficiency and multi-scale capability, are becoming increasingly popular.) Once users have made their selections, the software generates draft code that can be manually edited.

Despite its user-friendly interface, some very sophisticated computation is taking place under the hood. Most cell properties, behaviors, and interactions are bundled together in a single function, called the “effective energy” (the terminology harks back to the model’s roots in physics), which incorporates all of the forces acting on whatever agents are being simulated—cells, parts of cells, environmental features—and the rules that govern how they will respond. The cells live in regular 2-D or 3-D lattices, like pixels in a digital microscope image—a feature that links the GGH model to simpler cellular automata that represent cells as points on grids. Unlike automata, however, cells in the GGH model have volume, are deformable, and are affected not only by their immediate neighbors but by a host of other factors. They can also move about in three dimensions, providing a degree of spatial realism that is extremely valuable for tissue simulations.

The GGH model is also inherently stochastic: cells move about by randomly exploring their environment, responding to whatever forces have been programmed into the simulation, and moving on average towards a state of least energy. That randomness, says Glazier, is what gives cells the freedom to reorganize themselves. It also gives rise to very complex and even unexpected aggregate behaviors—behaviors that could not necessarily be predicted from the underlying rules. This quality, known as emergence, is both the hallmark of agent-based modeling and the secret to its success. “That complexity is the reason that this kind of modeling works,” says Glazier, who adds that modelers can simplify many of the rules governing the individual agents in an ABM and still generate realistic global behaviors, so long as they include the key biological mechanisms—

**Emergence is both the hallmark of agent-based modeling and the secret to its success.**
Emergent Rules

According to Gary An, MD, the ability to generate complex emergent behaviors in the absence of comprehensively detailed knowledge makes agent-based modeling ideally suited to testing hypotheses and conducting in silico trials.

An, associate professor of surgery at the University of Chicago, first came to agent-based modeling while working as a trauma surgeon at Cook County Hospital in the 1990s. Frustrated by the lack of medications he and his colleagues had for treating sepsis, a potentially fatal condition that occurs when the immune system's own response to injury or infection triggers inflammation throughout the body, An began building agent-based models of sepsis using SWARM, a software platform for multi-agent simulations of complex systems developed by the Santa Fe Institute. Since then, he has continued to use ABMs to investigate acute inflammation, often in collaboration with his friend and colleague Yoram Vodovotz, PhD, an immunologist and professor of surgery at the University of Pittsburgh. He has also helped others apply similar models in their own research.

An was first attracted to ABMs because he found them to be more intuitive than equation-based models. “I wasn’t a math guy,” he says. “I didn’t think in terms of differential equations and calculus. I thought in terms of things doing things”—i.e., cells interacting with other cells—and things doing things is agent-based modeling.” But he has come to appreciate ABMs as tools for dynamically embodying what we know (or think we know) about biological systems and processes, and as platforms for testing hypotheses that can yield unexpected insights.

CompuCell3D can be used to simulate cell sorting using various rules for cell adhesion. These snapshots show the dynamics of cluster formation during a 5000-cell aggregate simulation with five different levels of cadherins (represented by the five different colored cells). Reprinted from Zhang Y, et al., Computer Simulations of Cell Sorting Due to Differential Adhesion. PLoS One 2011; 6(10):e24999. doi: 10.1371/journal.pone.0024999.

“I didn’t think in terms of differential equations and calculus. I thought in terms of things doing things,” An says. “And things doing things is agent-based modeling.”
Three virtual populations of 500 individuals and ran for 15,000 time steps, corresponding to approximately 40 years. Genetic mutations were allowed to accumulate over time, ultimately impairing cell function and resulting in cancer.

The first version of the model did a good job of accurately simulating normal cell population dynamics and breast physiology. But it could not generate estrogen receptor-positive (ER+) tumors, which are in fact the most common kind. This “huge fail,” An says, not only indicated a serious flaw in the model (because the rules governing the agents were based on the best available knowledge concerning breast cancer), but also pointed to a serious gap in researchers’ understanding of the pathogenesis of ER+ tumors.

The clue to solving this mystery lay in the model itself. Since ER+ cells are normally prevented from proliferating by the suppression of the receptor c-Met, the agent rules specified that ER+ cells were not allowed to divide. That, in turn, meant that mutations to the cells couldn’t accumulate to be passed on to future generations and lead to cancer. So An and his colleagues began looking for something that would allow ER+ cells to proliferate—something that would ordinarily be responsible for suppressing c-Met, but that could be impaired.

A literature search identified the gene RUNX3 as a possible candidate; and once it was incorporated into the model and permitted to mutate, ER+ cells acquired the capacity to replicate and accumulate damage, resulting in the appearance of ER+ tumors.

The discovery that RUNX3 might play a role in breast cancer by regulating ER+ cell proliferation could be clinically useful. For example, An raises the prospect of one day screening for decreased expression of RUNX3 as a warning sign of increased risk for ER+ tumors. But the discovery process also highlights one of the advantages of agent-based modeling. An equation-based model, An says, might simply have been designed to reproduce the rates of ER+ tumor occurrence seen in the real world, and would therefore have masked the underlying mechanism. The agents in the DEABM, however, could not reproduce those rates without having the proper mechanism written into their rules in the first place—making the absence of that mechanism painfully clear. As Glazier says, “An agent-based model is constructive—it includes only what you put in. If you leave out a key mechanism, you will never replicate the biology.”

“An agent-based model is constructive—it includes only what you put in. If you leave out a key mechanism, you will never replicate the biology,” Glazier says.

---

**If It Grows Like Skin, and It Looks Like Skin...**

Even when the underlying rules for a model are incomplete, researchers can use ABMs to test hypotheses “before killing rats or growing cells,” An says. Robert Isfort, PhD, and his colleagues at Procter & Gamble have made the most of this capability. Working together with researchers at the University of Sheffield in England, the Procter & Gamble group employed agent-based modeling to test no less than three competing theories of how epidermal tissue maintains and renews itself over time. In the process, they resolved a central question in skin biology and helped advance the field of stem cell research.

According to the oldest hypothesis, known as asymmetric division, stem epithelial cells drive epidermal regeneration by dividing either to form new stem cells, or to form progenitor cells that go on to produce the differentiated progeny that make up the outer layers of the skin. Another, more recent hypothesis, known as population asymmetry, holds that progenitor cells are primarily responsible for skin renewal through stochastic differentiation, with stem cells playing only a secondary role. The third and latest hypothesis, population asymmetry with stem cells (PAS), contends that stochastic differentiation of both stem cells and progenitor cells is required to maintain and regenerate skin tissue. With experimental data to support all three, the question remained: which hypothesis was correct?

Using a modified form of a human skin model that the Sheffield group had developed with an agent-based modeling platform called FLAME, the international team...
of researchers translated all three hypotheses into separate, stochastic ABMs, each with slightly different rules for cell division and differentiation; the probability of stem cell division, for instance, changed from model to model. They then ran each simulation for the equivalent of three years. In addition, the virtual skin in the PAS model was wounded at the three-year mark, and the simulation was run for the equivalent of an additional year to see how it would respond.

Although the physical forces acting on the cells as they adhered to the lower level of the epidermis or migrated to the upper surface of the skin remained the same in all three models, variations in cell division and differentiation produced strikingly different outcomes. Most surprisingly, says Isfort, the models derived from the first two hypotheses were unable to produce colonies of mother and daughter cells that behaved realistically over the long term. Consequently, while all three models yielded mature epidermal layers with similar cellular organization after three years, only the model instantiating the PAS hypothesis, according to which both stem and progenitor cells divide and differentiate stochastically, was able to generate tissue that acted “like the real stuff.” In addition to fueling future experimental research on stem and progenitor cells, this work could lead to new therapies for repairing skin that has been wounded, or suffered damage through the normal aging process.

**Hybrid Vigor**

Despite its strengths, agent-based modeling can, at times, be slower and less efficient than equation-based models. That is why some researchers are creating hybrid approaches that combine the two methods. Yoram Vodovotz, who has co-authored a number of papers on agent-based modeling with Gary An, says that while ABMs can often be assembled more quickly than equation-based models, their stochastic and emergent properties sometimes make it difficult to relate outcomes to specific causal factors. In 2004, for example, Vodovotz and An both published papers in *Critical Care Medicine* on sepsis, with Vodovotz simulating a population of patients using a deterministic model based on ordinary differential equations, and An using an ABM. In that case, says Vodovotz, the mathematical model allowed him to trace individual patient outcomes to particular configurations—to say that patient X, for example, died because of a specific pathogen load, or a genetic predisposition to acute inflammatory response—whereas the ABM could only indicate that a certain percentage of virtual patients hadn’t responded to treatment, without revealing precisely why.

Emergence also makes it harder to set and optimize parameters in ABMs than in mathematical models. In an equation-based model, for example, the modeler can simply program parameters like rate constants, which characterize the rates of biochemical reactions in a system. In an ABM, however, rate constants must emerge from the individual interactions of the agents; one must run the simulation first, then measure the rate constants and tweak the model if the numbers don’t match experimental data. Vodovotz says that this makes parameter optimization in ABMs “very nontrivial,” adding: “It’s one of those grand challenge types of problems.”

Moreover, while equation-based models can be analyzed using well-established mathematical techniques, the complex patterns that emerge from ABMs can be difficult to quantify and analyze with the same degree of rigor. And while agent-based methods are very good at simulating local interactions between heterogeneous populations of cells at multiple scales and with a high degree of spatial realism—for example, simulating how different kinds of cells migrate from place to place, adhere to one another, and arrange themselves in the macroscopic patterns found in real-life tissues—differential equations provide a
more efficient and less expensive way of modeling well-mixed systems and other phenomena that can be adequately represented at the continuum level, such as blood flow.

Just as Glazier and his team at Indiana University have gradually expanded CompuCell3D to incorporate equation-based models of the biochemical pathways inside cells and the chemical flows between organs and tissues, Vodovotz has also been building hybrid models that offer the best of both worlds. He and his colleagues at the University of Pittsburgh have developed an open-source software package called SPARK (Simple Platform for Agent-based Representation of Knowledge) that can integrate mathematical and agent-based modeling techniques, and have used it to demonstrate how basic inflammatory mechanisms can lead to both positive and negative outcomes in various kinds of tissue. In a paper published in May 2013 in *PLoS Computational Biology*, for example, Vodovotz employed a hybrid model to simulate the formation of pressure ulcers, or bedsores, on the skin of patients with spinal cord injuries, a common and potentially life-threatening occurrence.

The model used ordinary differential equations to simulate blood flow in the skin based on non-invasive measurements taken from injured individuals and an uninjured control group; and a stochastic ABM to simulate the blood vessels, cells, and signaling molecules (epithelial cells and macrophages; pro- and anti-inflammatory cytokines) that are involved in the formation of pressure ulcers. According to the rules governing the agents in the ABM, damaged epithelial cells released inflammatory cytokines that caused further damage; but they could also be healed by anti-inflammatory cytokines at a rate that depended on the amount of oxygen delivered by the blood. The simulation produced realistic-looking pressure ulcers at rates suggesting that people with spinal cord injuries are more likely to form them than people without such injuries, perhaps due to changes in vascularity—a finding that could lead to tools for predicting the risk of ulcer formation based on non-invasive measurements of blood flow.

Christian Jacob, PhD, a computer scientist at the University of Calgary who has used ABMs to simulate everything from ant colonies to traffic congestion, has also built a platform for constructing hybrid models, albeit one that takes the whole body as its canvas. Developed by PhD student Tim Davison and other graduate students in Jacob’s Evolutionary & Swarm Design Lab, the software suite, which goes by the name LINDSAY Composer, can be used to create interactive 3-D simulations and visualizations of human physiological processes across multiple scales, from systems and organs to cells and sub-cellular structures. Users can drag and drop objects into their simulations from a component library that contains templates for various agents (e.g., cells, pathogens), all of which come with their own customizable sets of properties and interaction rules. Like SPARK and CompuCell3D, LINDSAY Composer can also combine mathematical and agent-based models—feeding data, for instance, from a mathematical model of molecular concentration gradients to an agent-based model of cell development. Jacob’s ultimate goal is to create a comprehensive 3-D interactive model of human anatomy and physiology, called LINDSAY Virtual Human, which will enable users to zoom seamlessly from the whole-body scale right down to the molecular level for both medical education and research purposes.

Jacob’s introduction to agent-based modeling came during the 1980s when he first encountered computer simulations of flocking birds. He was immediately impressed with the method’s capacity to handle mixed populations of agents in three-dimensional space—a capacity that proved crucial to a project that Jacob and

---

**ABMs for Biomedicine**

Many possible ABM software programs exist (including NetLogo, which might be best for the ABM novice; http://ccl.northwestern.edu/netlogo/), but the five listed below are featured in this story. Although they all accomplish somewhat the same thing, they were developed using different programming languages, possess varying levels of support and documentation, and have been used to build different models.

Curious investigators are encouraged to visit their respective websites to learn more about them.

**CompuCell3D**: www.CompuCell3D.org

**SWARM**: www.swarm.org

**FLAME**: http://www.flame.ac.uk

**SPARK**: http://www.pitt.edu/~cirm/spark/

**LINDSAY Composer**:

http://lindsayvirtualhuman.org/?q=node/59
a former student, Vladimir Sarpe, MSc, recently undertook using LINDSAY Composer.

In a paper published in BMC Bioinformatics, Jacob and Sarpe describe how they used a three-dimensional ABM of the human immune system to simulate and visualize the body's response to influenza A virus, from the initial infection of epithelial cells in the lungs to the destruction of the virus by lymphocytes. The model included such agents as T cells, B cells, viruses, and antibodies that were programmed to interact according to various rules in two distinct 3-D environments: within the lung tissue, and inside a lymph node. From a computational perspective, each environment was treated separately—the lymph node and lung tissue simulations were in fact executed on different computing nodes—but they communicated with one another via "controllers" that shared information as necessary. A dendritic cell in the lung that encountered a virus, for example, would engulf the pathogen and transport it to the lymph node to activate the T and B cells. They in turn would produce killer T cells and antibodies that would travel back to the lung tissue in order to neutralize the virus and destroy the infected epithelial cells. The simulation even generated "memory" T and B cells that stuck around after the initial infection to enable a faster response upon subsequent exposure to the virus.

**Getting ABMs Into More Hands**

The high computational overhead incurred by ABMs remains a challenge. In the case of their immune-system simulation, Jacob and Sarpe sidestepped the issue by relying on a relatively small number of agents—a few thousand, far less than the actual number of cells and viruses that would really be involved—and used probabilities (of becoming infected, of releasing antibodies, of reproducing) to generate the kinds of emergent behaviors that would arise with more realistic numbers of moving parts. As a result, the model was able to produce outcomes that accorded both with clinical data, and with the results of a robust equation-based model.

That approach might not always be ideal, however; so for Jacob, driving down the computational expense of agent-based modeling has become an area of research unto itself. In a paper published this year in the journal Simulation, he and his colleagues reported that they were able to reduce the number of agents in a simulation by creating so-called "observers" that recognized patterns in the behaviors of groups of agents, and replaced those groups with single meta-agents that subsumed their behaviors. When applied to a blood-clotting simulation in which 12 different blood factors were represented as agents, average run-time was cut almost in half.

By making agent-based modeling more affordable, such advances could also help put ABMs into the hands of more scientists. And that would be good news for biomedical researchers who do not necessarily know much about machine-learning algorithms or parameter optimization, but who do find it easy to grasp a modeling technique that so faithfully reproduces the kinds of objects, interactions, and behaviors that they observe in nature.

"They actually think of these agents," Jacob says, "without knowing it."