Proceedings of the 2004 Winter Simulation Conference
R .G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, eds.

A SIMULATION METHODOLOGY IN MODELING CELL DIVISIONS WITH STOCHASTIC EFFECTS

Harsha K. Rajasimha
David C. Samuels

Virginia Bioinformatics Institute
Bioinformatics Facility I (0477)
Virginia Tech
Blacksburg, VA 24061, U.S.A.

Richard E. Nance

660 McBryde Hall (0106)
Department of Computer Science
Virginia Tech
Blacksburg, VA 24061, U.S.A.

ABSTRACT

We present a model to explain the effects of the long time between blood stem cell divisions and rapid cascades of progenitor cell divisions on the mitochondrial DNA drift. We allow four stochastic events in the system namely, mtDNA replication and degradation, cell division and death. To implement the conceptual model, we design two simulation models; one for a limited number of stem cells (20,000) over very long time scale (100 years) and another for the cell divisions of a progenitor cell resulting in a large number of blood cells (~10 million) over a shorter time span (25 days). Iterative enhancement with incremental builds constitutes the modeling methodology. We adopt the activity scanning conceptual framework for the model implementation. Initial transient and memory issues are resolved. By output data analysis, we conclude that the variation in mutation level occurs significantly due to time and less so due to cell divisions.

1 INTRODUCTION

For the reader unfamiliar with cell biology, we include some basic definitions.

Mitochondria: Cellular organelles responsible for energy production. Hence, they are called the power house of the cell. Mitochondria are unusual because they contain their own DNA molecules (mtDNA), separate from the DNA in the cell nucleus.

Stem cell: An unspecialized cell with self renewing capabilities that gives rise to a specific specialized cell.

Hematopoietic stem cell: A small population of undifferentiated cells in the bone marrow which give rise to the mature differentiated blood cells.

Blood progenitor cell: The ancestor cell for all types of blood cells originating directly from a stem cell division in the bone marrow.

Symmetric cell division: Always results in two identical daughter cells (Figure 1). For example, progenitor cells in our model are known to follow such divisions (Marley et al., 2003).

Asymmetric cell division: results in two daughter cells that are not be identical (Figure 1).

Mixed cell division: results in two daughter cells that may or may not be identical (Figure 1).

Synchronous cell division: all cells in a colony divide at the same time (Figure 2).

Asynchronous cell division: cells divide independently, at different times (Figure 2).

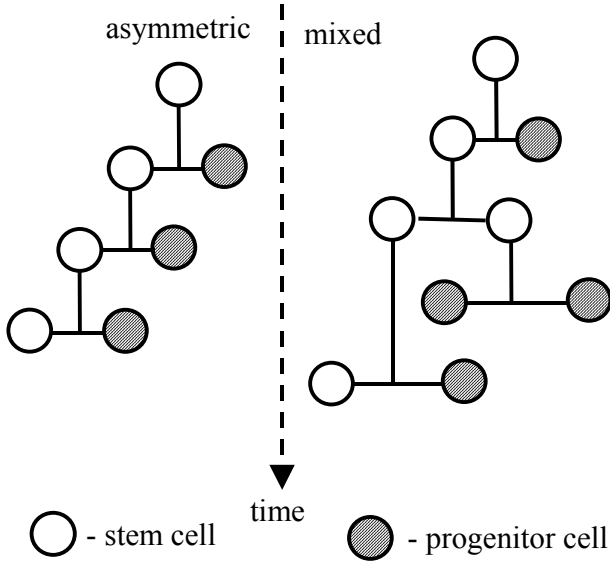

Figure 1: Symmetric, Asymmetric and Mixed Cell Divisions

A blood stem cell is an undifferentiated cell in the bone marrow that divides about once a year to yield two daughter cells. The divisions are asymmetric and result in the formation of two stem cells, two progenitor cells or one of each type.

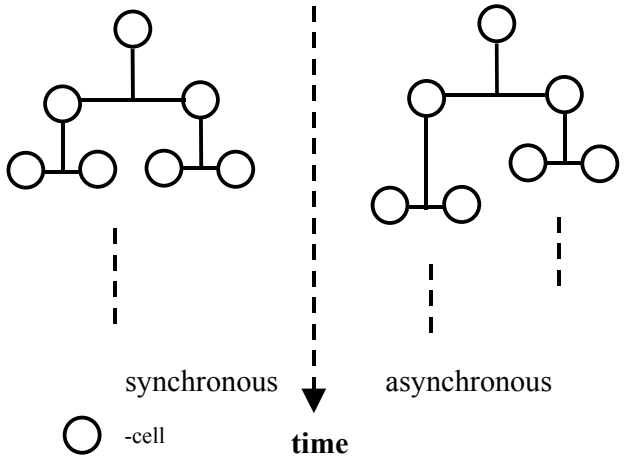

Figure 2: Synchronous and Asynchronous Cell Divisions

Progenitor cells undergo rapid cascades of cell divisions yielding millions of blood cells every day. In the following sections, we describe the system, the model, the model objectives and the requirements of the simulation study. The modeling methodology is described giving attention to the conceptual framework and the verification and validation (V&V) techniques employed. An initial transient problem is resolved and we comment on the indications of scalability and memory limitations from the preliminary results. We close with an observation regarding model insensitivity and offer several extensions that would incorporate more system features in future experimentation.

2 THE SYSTEM

The system to be modeled belongs in the domain of cell biology. The system consists of a number of human blood stem cells in the bone marrow. Each cell has a number of mitochondrial DNA (mtDNA) molecules (a few thousand). Each mtDNA molecule can be of either a wild-type mtDNA (W: The normal type we would expect in a cell), or a mutant-type mtDNA (M: abnormal due to insertion, deletion, or substitution of nucleotides). Each mtDNA molecule undergoes a process of replication wherein it creates a copy (replicate) of itself. Let N = W+M be the total number of mtDNA in a cell. Typically, a normal cell would contain all W mtDNA. The state where both W and M type mtDNA occur in the same cell is known as heteroplasmy. mtDNA are degraded with a half life ($T_{1/2}$). Under normal conditions, the cell replicates its mtDNA at a rate Ro to compensate for the loss of mtDNA by degradation. With both mtDNA replication and degradation events, and cell division and cell death events occurring independently in this stochastic system, the average mutation level (M/N) in the cell population drifts over time. The effect of this drift needs to be studied over a human life span. We observe the variation in the M/N level in a population of cells starting at different initial mutation levels over time. The effect of cell divisions are studied over a cascade of cell divisions leading to millions of cells. The literature supports the assumption that blood stem cells undergo cell division about once a year. These divisions are asymmetric and asynchronous, resulting in both blood progenitor cells and stem cells. A blood progenitor cell, once formed, undergoes a series of rapid cell divisions (approximately once a day) to yield mature blood cells after about 18-20 levels of cell divisions (Figure 3). In each cell division, random division determines the number of M or W type mtDNA that go into each of the daughter cells.

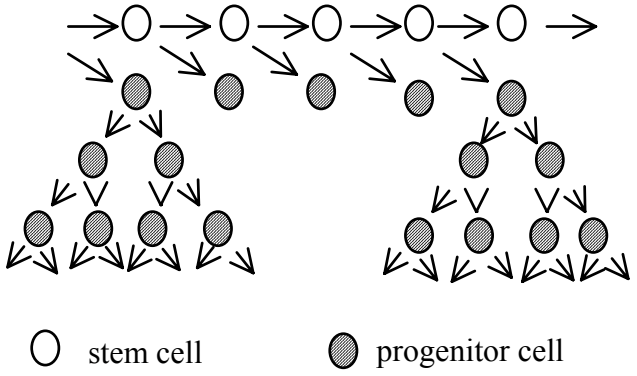

Figure 3: Cell Division Model for Blood Stem Cells

3 STUDY OBJECTIVES

Important in the understanding of the mtDNA mutation load on the blood cells are the relative contributions of two factors; the length of the interval separating divisions of the stem cell and the rate of the cascade of cell divisions among the progenitor cells. The objectives of this simulation study are:

1. How does the mutation level of mtDNA (M/(W+M)) drift over a time span of 100 years?
2. How does the mutation level of mtDNA drift over rapid cascades of cell division resulting in at least 2^{20} (a million) blood cells?
3. How does the drift in mutation level in the stem cells compare with that occurring in the progenitor cell division cascade?

4 THE MODEL

The model we develop will represent the above system with the following assumptions:

1. During cell division, roughly 50% of total mtDNA is distributed to each daughter cell.
2. A cell dies and is removed from the population only if its mutation level exceeds the set threshold level (~90%).
3. mtDNA replication, mtDNA degradation, cell division and cell death are the only events that affect the mtDNA mutation load in the cell.

4. The progenitor cells in Figure 3 divide to produce several specialized cell types. However, the model treats all these cells as the same type.

4.1 MODEL EQUATIONS

Table 1: Table of Model Parameters and Equations

Description	Parameter	Set to value
The target total mtDNA population in a cell	Nt	~5000
The time between cell divisions	D	1 year (stem cells) 1 day (progenitor cells)
The number of wild-type mtDNA in a cell	W	W = pp*Nt
The number of wild-type mtDNA in a cell	M	M = (1-pp)*Nt
Half life of mtDNA	$T_{1/2}$	10 days (from literature)
The wild-type mtDNA level in a cell	pp	set to a value between 10% to 90%
The simulation time step	Δt	1 hour (for both stem and progenitor cell models)
The probability of a cell division occurring in time Δt		$\ln(2) * \Delta t / D$
The mtDNA destruction coefficient	τ	$\tau = T_{1/2} / \ln(2)$
The mtDNA replication rate	Ro	$Ro=Nt/\tau+\ln(2)*Nt/D$
The mean value parameter for the poisson subroutine	xm	$xm = Ro * \Delta t$
The probability of a cell death event occurring in a time interval $(t, t+\Delta t)$		= (number of living cells at time t)*Δt/(D*Nt)

4.2 MODEL PARAMETERS TO BE SET

MAXCELLS: Maximum number of cells that could exist at anytime

MAXDIV: Maximum number of cell division cascades in the synchronous model

MAXHOUR: Maximum number of hours that the simulation should run

THRESHOLD: A cell dies and is removed from the population if its mutation level exceeds the set threshold level (~90%)

5 MODELING METHODOLOGY

Iterative enhancement with incremental builds constitutes the methodology of model development. Different perspectives are offered in the literature (Balci 2001, Sargent 2001). While we do not use a specific methodology, the simulation model is developed with careful consideration to the methodology and CF. To achieve the objectives and meet the requirements, we design two simulation models; one for a limited number of stem cells (20,000) over very long time scale (100 years) and another for the cascaded cell divisions of a progenitor cell resulting in a large number of blood cells (10 million) over a shorter time span (25 days). Separating the programmed stem cell model from the progenitor cell model is the strategy adopted to overcome memory and time limitations of the PC. The RNG and probability distribution subroutines are verified, first, the stem cell model with synchronous cell divisions is implemented and results verified. The next step is to extend stem cell model to asynchronous divisions. We then build the synchronous and asynchronous cell division versions of the blood progenitor cell model. This is followed by design and execution of simulation experiments and output data analysis.

The algorithm used to generate a uniform random number taken from the book "Numerical Recipes in C" is originally proposed by Park and Miller (1988). The subroutines for generating a Poisson distribution and binomial distribution are also taken from the same source. All these subroutines are implemented as 'C' functions and included as a header file. A poisson distribution is used to set the number of W and M mtDNA that are degraded during each time step. A binomial distribution is used to set the number of W and M type mtDNA to be copied during a time step based on the current numbers in the cell.

6 CONCEPTUAL FRAMEWORK

For implementing the simulation model, we adopt the Activity Scanning (AS) Conceptual Framework (CF) with fixed increment time flow mechanism (TFM) (Balci 1988). Our objectives require that we perform statistical analysis on the state of the system at specific intervals of time in order to make valid predictions. For example, consider the following questions: Given an initial number of stem cells to start with, what are the number of cells living after 50 years? What are the number of blood cells produced per day? Starting from a given initial mutation level in given number of cells, how does the mutation level decay over a 5, 10, 20 year span? To answer such questions, AS with fixed time increments allows a simple and logical design of the program flow. Second, the simulation termination condition is time based and statistical routines are to be called at regular intervals of time in order to save the system state at those times. Third, since there are only four different events that can occur at each time point namely, cell division, cell death, mtDNA copy, and mtDNA degradation, it is easy to scan them. Lastly, we deal with a large number of cells (millions), and an ES approach would mean scheduling (or queuing) millions of events whenever an event occurs (since several events are expected to at each time step). This would be memory inef-

ficient (Derrick, Balci, and Nance 1989). AS overcomes this problem by assigning a probability of occurrence of an event in each time step. Also, we do not save all the attributes of the system at every time step, but only those averaged over intervals of time.

7 MODEL V&V TECHNIQUES

We build the model with particular attention to model Verification and Validation. We adopt the following V&V techniques:

Analysis by plot: Individual subroutines such as RNG, Poisson, binomial, cell division, and statistics subroutines are verified using short programs and plotting the output data. Subroutines derived from published sources such as Park and Miller (1998) are subject to one time verification to eliminate any implementation errors. For the blood cell model, one of the V&V procedure is to analyze the counts of the number of cells living and dying due to mutation threshold (Figure 4). We note that there is no single cell death in the first twelve days and thereafter the death count raises exponentially as the living cell count does. In addition, the death events are small in number indicating that the cascaded cell divisions do not cause many cells to exceed the mutation threshold.

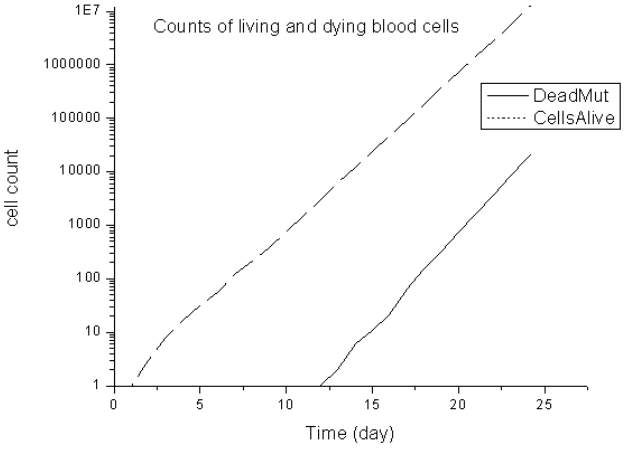

Figure 4: Blood Cell Counts Dying Due to Mutation Threshold Compared with Living Cell Counts over Time

Desk checking: Subroutines that are derived from reliable sources are subjected to a one time desk checking before utilizing them in the simulation. However, the simulation programs are subjected to routine desk checking whenever additions or modifications are made.

Figure 5 is used for desk checking purposes to ensure that the data structures for the critical parameters are storing the right data. The validation criteria set for this model are based on the standard expectations or the biological scenario derived from literature. The total number of mtDNA in a cell (N = W+M) should be fairly constant (very little initial variation and then it stabilizes). If the model is implemented right, the average number of W and M type mtDNA in all living stem cells should behave as shown in Figure 5.

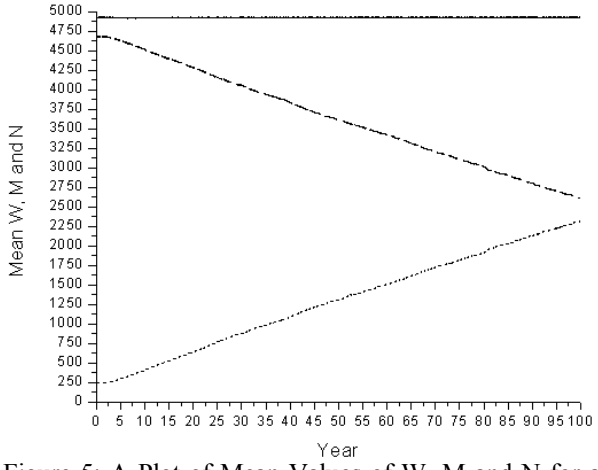

Figure 5: A Plot of Mean Values of W, M and N for a Stem Cell Simulation Run Starting at 95% Mutation Level with the Threshold for Cell Death Set to 100%. Nt is Set to 5000 and W and M are Set to 5% and 95% of Nt Respectively

Simulation runs performed for exact parameter set with different random number seeds (idum) serve as a verification procedure.

Simulation/theoretical comparison: The simulation output data is compared to the experimental data sets on blood cells provided by Rahman et al., (2001) that spans over a 19-year time frame and Howell et al., (2000) data spans just about 7 years. We observed a good overlap between the W and M levels in the cells with the experimental values. This served as the major verification procedure for this research. However, not enough data is available on longer time scales; as such experiments are both costly and become infeasible to conduct in one's lifespan. Despite all this, some experiments are ongoing in this direction using the blood samples at birth and present. If blood samples are saved at birth, comparing the values from those with current values in the same person can give us the data across the kind of time-scales we are looking at. With that, we should be able to provide a standard behavior curve (for each set of initial conditions) using this simulation model for experimentalists to compare with.

Subject matter expert: Several of the plots and desk checking procedures are designed based on the experience and intuition of a domain expert.

8 RESULTS

Simulation output data analysis is performed using Origin package from OriginLab.

Stem cell model: From experiments, it is known that cells can survive and even function normally with mutation

level of up to 85%. Since it is not exactly known what the threshold level for cell death is, we need to test if the exact threshold value affects the model behavior significantly. A simulation experiment is conducted with THRESHOLD set at 100% so that a cell would die only if the mutation level reached 100% in that cell. The result for the two simulation runs starting at 75% and 95% initial M level is shown in Figure 6. Another simulation experiment is setup with THRESHOLD set to 90%. Interestingly, the behavior with 90% THRESHOLD for cell death does not deviate much from the behavior observed at 100% threshold even over a time scale of 100 years (plot not shown).

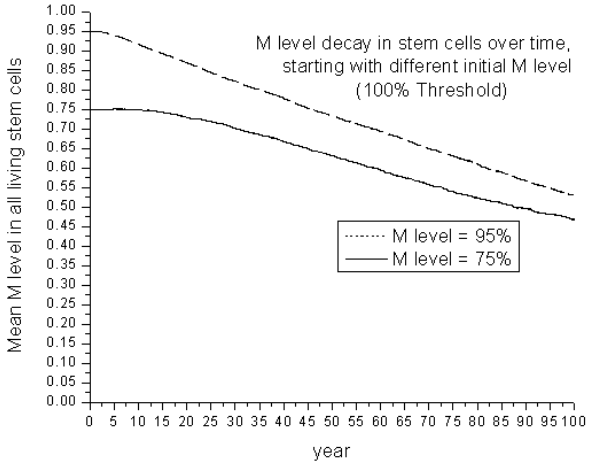

Figure 6: A Plot of Mean M Level (M/N) versus Time

Progenitor cell model: The results from the cascaded cell division model of progenitor cells shows that the average M level of all cells resulting from a cascade of cell divisions did not differ from that of the progenitor cell (Figure 7). As shown in Figure 7, the variation in the mean M level in about 10 million cells resulting from a single progenitor cell is in the order of 10^{-3} which is negligible.

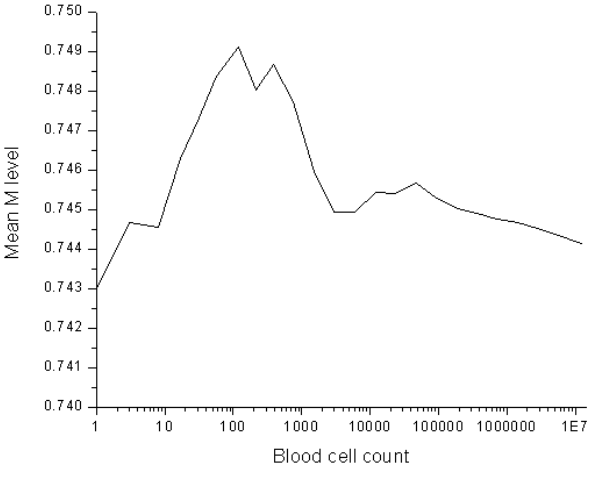

Figure 7: Mean M Level of Blood Cells as a Function of Their Cell Count

9 THE PROBLEM OF INITIAL TRANSIENT

Figure 6 also shows the problem of initial transient for two simulation runs starting at 75% and 95% mutation levels. The run at 75% mutation level takes longer for individual cells to exceed the set 100% threshold for cell death. While the M level starts to decay after about 10 years in the 75% case, the same starts in less than 5 years for the 95% case. One reason is that in a single simulation run, all cells are initialized to the same initial state, which is not very close to the natural situation. Hence, we allow the first few years of simulation to allow randomization of the initial state of different cells. Figure 8 shows the stem cell population distribution for a simulation run starting at year 0 and 70% mutation level, after 6 years and after 60 years. As we can see, the distribution after 6 years is spread around the mean 70% level from 40% to 90%. However, after a span of 60 years, the population has spread the entire spectrum from 0 to 90% with more than 1500 cells fixed at all W type cells. The current procedure to determine the end of transient period is more ad hoc and expert heuristic based. A much cleaner way of handling this issue of initial transient is still of interest.

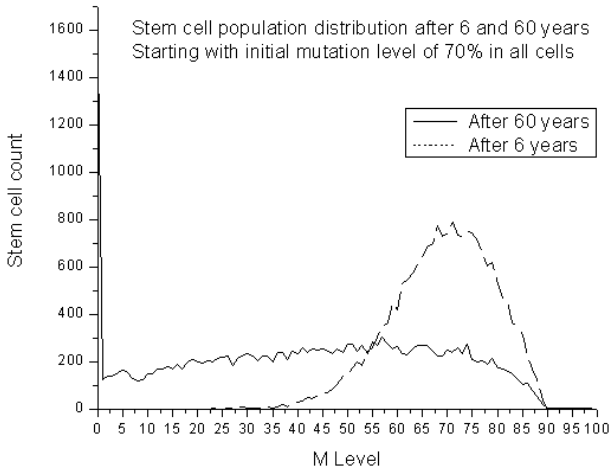

Figure 8: Stem Cell Population Distribution Based on Their M Level

To check the effects of starting simulation runs on cells with different starting conditions, the initial mutation level is set to a range of values from 20% to 75%. Figure 6 shows the decay in M levels in cell populations starting at different initial M levels (75% and 95%). From the plot, both lines are nearly parallel, i.e., initial condition does not noticeably impact the slopes. The explanation for this is that the amount of time needed for the cells to exceed the M THRESHOLD for cell death is proportional to the initial M level at which each cell starts. In our model, all cells start with the same M level in a particular simulation run. But, the rate at which the M level decays itself does not depend on the initial M level of the cell.

10 MEMORY AND SCALABILITY ISSUES

One point to mention here is the memory limitation posed by data structure design. With the initial cell class definition, each instance of the cell object has 4 float variables (W and M values for private and public access) and 1boolean variable (cell dead or alive). The maximum number of cells that can be statically declared at the beginning of the program is about half a million because of the limited memory available on the stack. Since the requirement is to simulate at least a million cell objects (~ 2^{20}), we consider the following improvements:

1. Shift to dynamic memory allocation in which case, we would have to create a cell object every time a cell division event occurred and destroy one, every time a cell death event occurred. Traversing a linked list of more than a million cells at every time step is highly inefficient. This shift would be quite costly in terms of resources and effort.

   A way around this problem is to declare the array of objects in the heap instead of on the stack.

   i.e., instead of declaring:

   ```
   // This allocates memory on the stack
   Cell cell[MAXCELLS];
   ```

   declare this way:

   ```
   // This allocates memory on the heap
   Cell *cell;
   cell = new Cell[MAXCELLS];
   ```

   In this case, the number of cell objects (MAXCELLS) is limited only by the amount of available secondary memory on the computer. This imposes an additional overhead on the program, as it now has to handle more of segmentation issues than before to get cell objects between main and virtual memory to perform operations.

2. Reduce the memory requirements of each cell object so as to hold many more cells. We achieved this by eliminating two private float variables to save W and M values for each cell. The model is still intact since we have the two public variables that undergo changes as events occur at every time step.

11 CONCLUSIONS AND FUTURE WORK

Results: We present a simplified model of mtDNA drift in blood stem cells. The main result of the model is that long times between divisions allows the mtDNA mutation levels to increase beyond the threshold for cell death. Hence, such population of cells will show a decrease in mutation level as more and more cells die. On the other hand, the drift in mutation level introduced by rapidly dividing progenitor cells yielding millions of blood cells is negligible. All cells resulting from a single progenitor cell would have nearly identical mutation level i.e., the variation in the M level occurs mainly due to time and not due to rapid cell divisions.

Simulation run times: The current stem and progenitor cell simulation models take approximately about 4.5 hours each to complete on a Intel P4 2.4 GHz processor with 256 MB RAM. The stem cell model is set to run for 100 years and the progenitor cell is set to expand to about 12 million cells. We note that the model design is highly amenable to parallelization to achieve faster simulation for even larger numbers of cells for even longer durations.

Model complexities: The cell object can be made more complex to include other cellular compartments and/or the model can be applied to a growing cell culture or an embryo. The blood stem cell model can be extended to include the various types of specialized blood cells that could be produced such as erythrocytes, platelets, etc. The model is easily adaptable to other stem cell types.

Model extensions and applications: Incorporating the actual mtDNA sequence data in the model and applying a particular mutation could be one of the extensions of the model. One question that such a model would help us examine concerns the idea that mtDNA with large deletion mutations would take less time to replicate than wild-type mtDNA since the number of base pairs replicated would be much less. **Preimplantation Genetic Diagnostics (PGD)** for advising gravid couples regarding genetically inherited mitochondrial diseases (Dean et el., 2003). Since **cancer** involves uncontrolled proliferation of cells (Byrne and Preziosi 2003), an extension of our current model can be used to examine specific questions about the characteristics of cancer cells and the role of mitochondrial mutations in cancer (Breward, Byrne, and Lewis 2003). **Apoptosis or programmed cell death**, which has sparked (or renewed) significant interest in the recent past can be studied by extending the current model. Cancer can be thought of as a condition where cell death fails to occur at a reasonable rate with cell proliferation occurring at normal pace. Understanding the two in tandem will give valuable insights into the basic science of apoptosis in cancer cells.

ACKNOWLEDGMENTS

D.C. Samuels and H.K. Rajasimha thank the commonwealth of Virginia for financial support and their mitochondria research team at VBI for their valuable inputs. Rajasimha especially thanks Dr. Bradshaw for providing useful suggestions.

REFERENCES

Balci, O. (1988) The Implementation of Four Conceptual Frameworks for Simulation Modeling in High-Level Languages, Proceedings of the Winter Simulation

Conference, M. Abrams, P. Haigh and J. Comfort (eds), 287-295. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Balci, O. (2001), "A Methodology for Certification of Modeling and Simulation Applications," *ACM Transactions on Modeling and Computer Simulation*, 11, (4): 352-377.

Breward, Christopher J. W., Helen M. Byrne, and Vlaire Lewis (2003) A Multiphase Model Describing Vascular Tumor Growth, Bulletin of Mathematical Biology 65, 609–640

Byrne, H., and L Preziosi (2003). Modeling solid tumor growth using the theory of mixtures *JOURNAL OF THE IMA* 20 (4): 341-366

Dean, N.L., Brendan J. Battersby, Asangla Ao, Roger G. Gosden, Seang LinTan, and Eric A. Shoubridge (2003) Prospect of preimplantation genetic diagnosis for heritable mitochondrial DNA diseases. Molecular Human Reproduction 9(10): 631-638

Derrick, E.J., O. Balci and R.E. Nance (1989) A Comparison of Selected Conceptual Frameworks for Simulation Modeling, Proceedings of the Winter Simulation Conference, E.A. MacNair, K.J. Musselman and P. Heidelberger (eds), 711-718. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Howell, N., Soumitra S. Ghosh, Eoin Fahy, and Laurence A. Bindoff (2000) Journal of the Neurological Sciences 172:1-6

Park, S.K., and K.W. Miller, (1988), *Communications of the ACM*, 31, :1192–1201.

Press W.H., Brian P. Flannery, Saul A. Teukolsky, William T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*.

Rahman S., J. Poulton, D. Marchington, and A. Suomalainen, (2001) Decrease of 3243 A->G mtDNA Mutation from Blood in MELAS Syndrome: A Longitudinal Study. *American Journal of Human Genetics* 68:238-240.

Sargent, R.G. (2001) Some Approaches and Paradigms for Verifying and Validating Simulation Models, *Proceedings of the Winter Simulation Conference*, B.A. Peters, J.S. Smith, D.J. Medeiros, and M.W. Rohrer (eds), 106-114. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Marley, S.B, J.L. Lewis and M.Y. Gordon (2003) Progenitor cells divide symmetrically to generate new colony-forming cells and clonal heterogeneity, *British Journal of Haematology* 121: 643-648.

AUTHOR BIOGRAPHIES

HARSHA K. RAJASIMHA is a research associate at the Virginia Bioinformatics Institute and a masters candidate in the computer science department at Virginia Polytechnic Institute and State University. He received a B.E. (2000) in computer science & engineering from Bangalore University, India and briefly worked as a lecturer in computer science in the same college. He gained industry experience interning as a software design engineer in test at Microsoft Corporation, Redmond, WA (Summer 2002). He worked as a graduate research assistant at the Virginia Bioinformatics Institute since December 2002 in the mitochondria research group. He intends to pursue doctoral research in bioinformatics. His research interests include biological data integration and modeling & simulation of life systems. He served as the secretary of the computer science graduate student council during the year 2003. He can be reached by email at <hrajasim@.vbi.vt.edu> and his web address is <http://bioinformatics.cs.vt.edu/~hrajasim>.

DAVID C. SAMUELS, Ph.D., is a research assistant professor at the Virginia Bioinformatics Institute. He received a B.Sc. degree in physics from Washington University in St. Louis in 1983 and A Ph.D. in physics from the University of Oregon in 1990. He was a faculty member in the Department of Mathematics at the University of Newcastle-upon-Tyne from 1996-2002. His principal research interests are in computational cell biology. Specifically, his group at the Virginia Bioinformatics Institute is currently focused on models of mitochondria, the cellular organelles responsible for the generation of energy. He can be reached by email at <dsamuels@vbi.vt.edu> and his web address is <https://www.vbi.vt.edu/article/articleview/220>

RICHARD E. NANCE, Ph.D., is a professor emeritus of computer science and director of the Systems Research Center at Virginia Polytechnic Institute and State University. He received B.S. and M.S. degrees from N.C. State University in 1962 and 1996, and Ph.D. degree from Purdue University in 1968. He has served on the faculties of Southern Methodist University and Virginia Tech, where he was Department Head of Computer Science, 1973-1979. Professor Nance has held research appointments at the Naval Surface Weapons Center and at the Imperial College of Science and Technology (UK). Within ACM, he has chaired two special interest groups: Information Retrieval (SIDIR), 1970-71 and Simulation Activities Board, the Outstanding Service Awards Subcommittee, the ad hoc Conference Procedures Committee and the ad hoc Film Committee that produced "Computers in Life." He served on the Editorial Panels of Communications of the ACM for research contributions in simulation and statistical computing, and Journal of Operations Research and Computer Science for contributions in simulation. The author of papers on discrete event simulation, performance modeling and evaluation, and computational structures and techniques of Operations Research 1978-82, and as Department Editor for Simulation, Automation and Information Systems of IIE Transaction, 1976-81. He can be reached by email at <nance@vt.edu> and his web address is <http://www.cs.vt.edu/info/people/vitae/Nance.html>.