

Análisis dinámico de un modelo 3D para el estudio de la leucemia mielogenia crónica

Sebastián Hernández¹, María Laura Ivanissevich² y Mirta Padín³

Unidad Académica Río Gallegos, Universidad Nacional de la Patagonia Austral

¹*shernandez@uarg.unpa.edu.ar*; ²*mivanissevich@uarg.unpa.edu.ar*;

³*mpadin@uarg.unpa.edu.ar*

Resumen

La leucemia mielogenia crónica es un tipo de cáncer donde el sistema inmune juega un rol muy importante, por lo que es interesante encontrar estrategias de intervención en la enfermedad. Aquí se analiza un modelo en tres variables de estado, correspondientes a los tipos de células intervinientes, el cual es altamente no lineal. Se pone énfasis en hallar los estados de equilibrio y analizar la estabilidad de los mismos. Estos estados se encuentran empleando algoritmos genéticos y el análisis de estabilidad se realiza por métodos habituales.

1. INTRODUCCIÓN

La leucemia mielogenia crónica (CML) es un cáncer que afecta células que circulan en el sistema sanguíneo. Esta enfermedad ocurre en más de 1 cada 100000 personas por año, y es la causa de aproximadamente el 15% de todas las leucemias en adultos. Se sabe además que el sistema inmune juega un rol muy importante en la dinámica de la CML (Sawyers, 1999). Esta enfermedad tiene tres fases distintas: crónica, acelerada, y de explosión. El modelo está dedicado al estudio de la fase crónica, que es la más larga (en promedio de 3 a 5 años) y cuando la cantidad de células crece en forma constante. Así, la población de células cancerosas es suficientemente grande durante la fase crónica de manera que la ley de acción de masas es válida para las interacciones de células en la sangre.

Se presenta el cálculo de los puntos de equilibrio en este sistema de tres dimensiones y con ecuaciones altamente no lineales y múltiples parámetros variantes a través de algoritmos genéticos (GAs). Los métodos clásicos de optimización y búsqueda, basados en técnicas analíticas, no se adaptan a estos problemas ya que exigen requisitos muy condicionantes sobre la función a optimizar y su complejidad aumenta considerablemente con la dimensionalidad (Holland, 1975). Los GAs, representan una técnica heurística de optimización (Goldberg, 1989), y se encuentra involucrada dentro de los esquemas evolutivos inspirados en los mecanismos biológicos naturales y conocidos originariamente como el Paradigma Neo-Darwiniano. Una población de individuos, que representan potenciales soluciones de algún problema, es sometida a transformaciones probabilísticas (reproducción, mutación, adaptación, selección, etc) y sólo algunos, los más aptos, sobreviven para regenerar una nueva población. Se espera que luego de algún número de generaciones el programa converja hacia los mejores individuos, los cuales deben estar muy cerca del óptimo.

La determinación de la estabilidad de estos puntos se realiza a través de métodos tradicionales incluyendo las simulaciones pertinentes que comprueban los resultados.

El trabajo se organiza como sigue: en la Sección 2 se describe el modelo, en la Sección 3 se describen las técnicas empleadas, en la siguiente se presentan los resultados obtenidos y finalmente se exponen las conclusiones.

2. EL MODELO BIOLÓGICO

El modelo matemático propuesto para la respuesta del sistema inmune humano al CML en un paciente hipotético pretende, a mediano plazo, identificar las direcciones promisorias para la investigación experimental en tratamientos de pacientes con CML colaborando con especialistas médicos.

El modelo determinístico propuesto por (Moore y Li, 2004) y que por tanto es exacto para poblaciones grandes, identifica dos poblaciones de células T definidas como sigue: las células T están en reposo si no son específicas a CML, pero pueden ser activadas; y las células T efectoras específicas para CML, las cuales están completamente “armadas” y capaces de acción inmediata. Las células T en reposo son activadas en el tejido linfático. Si una célula T en reposo es específica a CML, entonces puede comprometer un par MHC-péptido (complejo de mayor histocompatibilidad) en una célula antígeno-presente (APC). Si están presentes coestimuladores, las células T serán retenidas, activadas y proliferarán. Después de una semana de proliferación, la progenia diferencia entre células efectoras armadas (agresoras). Estas son entonces liberadas en la sangre y son capaces de montar un ataque inmune contra los antígenos CML encontrados, sin necesidad de coestimulación. Estas dos categorías (efectoras y en reposo) representan los comportamientos más dispares de células T que desea capturar el modelo.

El modelo consiste de un sistema de tres ecuaciones diferenciales ordinarias, donde dos de las variables son las descriptas y la tercera es la cantidad de células cancerosas de leucemia mielogénea crónica (CML). De análisis previos se sabe que cambios en la mayoría de los parámetros tienen pequeño efecto en los comportamientos a largo tiempo del sistema, mientras que el cambio en dos parámetros, la tasa de crecimiento de células cancerosas CML y la tasa natural de muerte (relacionados con ζ_6 y ζ_8 respectivamente), resulta en respuestas muy diferentes del sistema. En este modelo, las células linfocitos T en reposo toman contacto con el APC en el sistema linfático. Se asume que APC con antígeno a células CML están presentes en proporción al número de células de cáncer CML. Se usan las poblaciones de células CML con una constante de proporcionalidad para representar los números de APC que presenten antígeno CML. Las limitaciones de la respuesta inmune son representadas con un término de crecimiento de saturación, debido a los grandes números de APC que presentan antígeno CML el cual satura la linfa, pero relativamente pocas células en reposo T específicas de CML.

El sistema de ecuaciones diferenciales resultante utilizando técnicas de adimensionalización es el siguiente:

$$\begin{aligned}\frac{dT_n}{dt} &= 1 - T_n - \zeta_1 T_n \left(\frac{C}{C + \zeta_2} \right) \\ \frac{dT_e}{dt} &= \zeta_3 T_n \left(\frac{C}{C + \zeta_2} \right) + \zeta_4 T_e \left(\frac{C}{C + \zeta_2} \right) - \zeta_5 T_e - C T_e \\ \frac{dC}{dt} &= \zeta_6 C \ln \left(\frac{\zeta_7}{C} \right) - \zeta_8 C - C T_e\end{aligned}\tag{1}$$

T_n = células T en reposo (o no han sido aún expuestas a APC presentando antígeno CML, o no son específicas a CML),

T_e = células efectoras T específicas al CML, o sea con activación por un complejo péptido APC-MHC y coestimuladores,

C = células cancerosas de leucemia mielogenia crónica (CML).

El tiempo t se mide en días y las siguientes poblaciones de células en el sistema sanguíneo circulante son medidas como concentraciones de células por μl .

3. TECNICA APLICADA

Los Algoritmos Genéticos (GAs) operan sobre individuos codificados (Michalewicz, 1994; Wright, 1991),

emulando una cadena cromosómica conteniendo información genética a través de sus genes:

$$X \rightleftharpoons x = (g_1, g_2, \dots, g_l)$$

La variable X es transformada en cromosoma; un vector x formado por l genes g_i . La codificación del problema instala el proceso evolutivo en dos espacios simultáneamente: el espacio de búsqueda y el espacio problema (search-space vs problem-space) El primero contiene los genotipos mientras que el segundo los fenotipos, es decir la decodificación y manifestación como solución en el ambiente del problema. La medida de la aptitud de un cromosoma, o función de ajuste, es en general una función a valores reales positivos, $F(x)$, que mide la performance del individuo como solución del problema. En el ciclo generacional se persigue la maximización de $F(x)$ es decir el perfeccionamiento de los individuos a través de la acción de procesos y operadores genéticos como la selección, reproducción, cruzamiento, mutación, etc.

Las aplicaciones a problemas específicos no siempre reflejan la teoría, las principales desviaciones se deben por un lado a que hay dos presiones opuestas sobre la población que deben estar cuidadosamente balanceadas: la presión de selección de los mejores vs la diversidad de información genética. Por otro lado hay un límite finito sobre el tamaño de la población hipotéticamente ilimitada y sobre el número de generaciones (iteraciones).

El problema es hallar los ceros de una función f , $f : R^9 \rightarrow R$. El conocimiento específico de f permite una transformación tal que genera F como función a optimizar (fitness-form).

$$f \rightarrow F = -|f| + C.$$

La codificación del vector $x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9)$ es del tipo binaria, cada componente en un intervalo real que le es propio y con una precisión que varía entre 10^{-5} y 10^{-8} . Es así como se obtienen individuos (cromosomas) con un total de 199 bits de longitud. El algoritmo genético implementado finalmente tras una serie de ajustes se caracteriza por:

Método de Selección: Muestreo estocástico del resto. A cada individuo se le asigna un probabilidad de sobrevivencia proporcional a su medida de aptitud (fitness), de la cual se calcula el número esperado de copias. La parte entera de este número asegura determinísticamente una cantidad de clones y la parte fraccionaria es tratada como peso probabilístico en un sorteo. Se espera que los peores individuos tiendan a desaparecer. **Operador de Cruzamiento:** Binario de punto aleatorio simple. Esta instancia tiene como objetivo producir descendencia combinando la información genética por intercambio de subcadenas. Uno de los parámetros del GA es la probabilidad de cruzamiento, $pcross$, ésta indica la proporción esperada de individuos sobre los cuales actuará el operador. Tras el apareamiento de los individuos una posición de corte k , $1 \leq k \leq 198$, es elegida con distribución uniforme. Si bien la descendencia generada reemplaza a sus progenitores en la población se supone que permanecen clones de ellos intactos. **Operador de Mutación:** Inversión de un bit. Se intenta simular el fenómeno natural de la mutación genética y es por eso que actúa con muy baja probabilidad ($pmut$) sobre el valor de un gen (alelo). En una codificación binaria se invierten los valores 0 y 1.

La evolución del GA implementado se observa en las Figs. 1 y 2, siendo los parámetros óptimos: Tamaño de la población: $pop = 300$; Probabilidad de cruzamiento: $pcross = 0,3$; Probabilidad de mutación: $pmut = 0,02$

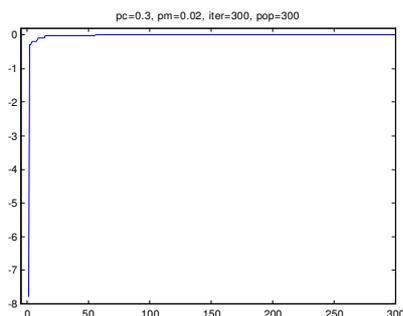


Figura 1: Solucion del orden 10^{-6}

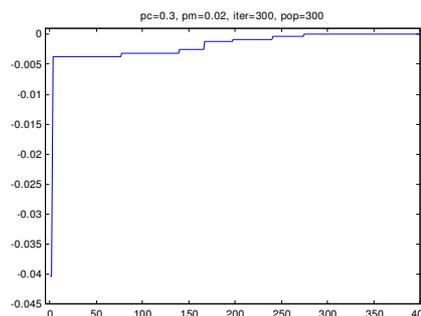


Figura 2: Solucion del orden 10^{-5}

4. RESULTADOS OBTENIDOS

Para hallar los puntos de equilibrio del sistema, y poder luego determinar la estabilidad de los mismos se empleó GA y se aplicarán técnicas convencionales para analizar la estabilidad de los mismos (Strogatz, 1994).

El sistema tiene equilibrios asintóticamente estables $P_1 = (1, 0, 0)$ (estado de salud) y estados P_2 que existirán si se cumple $\frac{\zeta_8}{\zeta_6} < \ln \zeta_7$. Para hallar los equilibrios P_2 debe resolverse

$$0 = \zeta_8 - \zeta_6 \ln(\zeta_7) + \zeta_6 \ln C + \frac{\zeta_3 C(C + \zeta_2)}{(C + \zeta_2 + \zeta_1 C)[(C + \zeta_2)(C + \zeta_5) - \zeta_4]} \quad (2)$$

Esta ecuación fue resuelta empleando algoritmos genéticos para un amplio campo de variación de los parámetros ζ_i y C . En la Tabla 1 se indica los intervalos de variación de cada una de las variables. Ellos fueron seleccionados teniendo en cuenta los criterios expuestos en (Moore y Li, 2004).

Los puntos hallados que verifican 2 con un error del orden de 10^{-6} , son los exhibidos en la Tabla 1, donde se incluyen además de los valores de los 8 parámetros, el valor de las tres variables, dos de las cuales, T_e y T_n , fueron determinadas por resolución numérica de las ecuaciones 1. El umbral a partir del cual se considera que no hay remisión para el cáncer es $C = 22500$ (resultados P_i^*) y que existe remisión parcial para los restantes casos (indicados P_i).

Punto	P_1^*	P_2	P_3	P_4	P_5^*	Intervalo / Partición
ζ_1	0,6423	0,0778	0,0388	0,3645	0,4079	[0, 02; 1] / $2^{17} - 1$
ζ_2	88,4235	4,5828	8,5265	34,5613	92,0190	[0, 1; 100] / $2^{24} - 1$
ζ_3	0,004636	0,003900	0,003730	0,002497	0,005899	[0, 001; 0, 006] / $2^{16} - 1$
ζ_4	5,9756	2,6326	3,5761	5,5509	7,1484	[2, 10] / $2^{20} - 1$
ζ_5	3,6459	4,6362	2,3426	2,0050	3,2345	[0, 2; 5] / $2^{19} - 1$
ζ_6	5,4236	1,5516	3,4115	2,2448	4,8335	[1; 6] / $2^{19} - 1$
ζ_7	32382,7489	36481,2381	40176,9256	37944,3813	39990,3059	[10000; 50000] / $2^{32} - 1$
ζ_8	1,5420	1,1078	1,8921	1,8735	2,1196	[1; 16] / $2^{21} - 1$
T_n	0,609764	0,927830	0,962667	0,733289	0,711004	-
T_e	1,1559E-07	2,0249E-07	1,5560E-07	1,1098E-07	1,6205E-07	-
C	24369,0	17864,3	23072,8	16469,8	25793,3	[15000; 30000] / $2^{31} - 1$

Tabla 1: Valores obtenidos para cada uno de los parámetros por medio del GA. Valores solución de las variables originales. Intervalos de búsqueda y sus particiones.

A fin de conocer el comportamiento de la población de células cercano al punto de equilibrio se linealizó el sistema y se obtuvo la matriz jacobiana del sistema, que resultó ser:

$$J = \begin{pmatrix} -1 - \zeta_1 \frac{C}{C+\zeta_2} & 0 & -\zeta_1 T_n \frac{\zeta_2}{(C+\zeta_2)^2} \\ \zeta_3 \frac{C}{C+\zeta_2} & \zeta_4 \frac{C}{C+\zeta_2} - \zeta_5 - C & \zeta_3 C T_n \frac{\zeta_2}{(C+\zeta_2)^2} + \zeta_4 T_e \frac{\zeta_2}{(C+\zeta_2)^2} - T_e \\ 0 & -C & \zeta_6 \ln\left(\frac{\zeta_7}{C}\right) - \zeta_6 - \zeta_8 - T_e \end{pmatrix}$$

Cada uno de los puntos seleccionados resultó ser asintóticamente estable ya que los autovalores allí calculados, como se muestra en la Tabla 2, son todos negativos. La comprobación de los resultados se efectuó por simulación, tomando un horizonte de 750 días como marca la progresión de la enfermedad. Los valores iniciales de las poblaciones para $t = 0$, son $T_n(0) = 1510$, $T_e(0) = 20$ y $C(0) = 10000$ (Moore y Li, 2004). En las Figs. 3, 4 y 5 se observa la evolución en el tiempo para el punto P_1 .

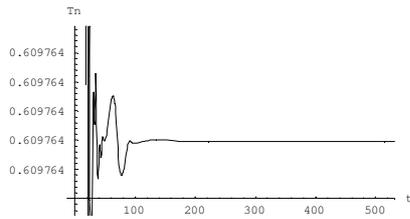


Figura 3: T_n en función de t .

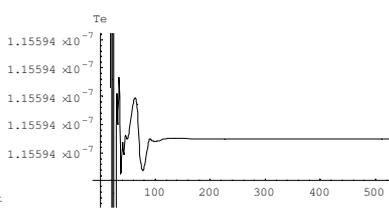


Figura 4: T_e en función de t .

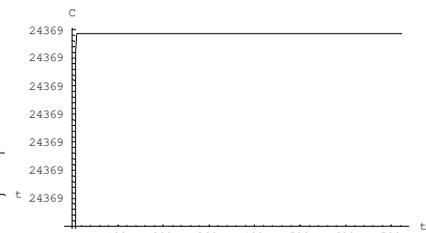


Figura 5: C en función de t .

Punto	Valor de la función optimizada	λ_1	λ_2	λ_3
P_1	4,0120E-06	-24366,7	-5,42361	-1,63996
P_2	4,2578E-06	-17866,3	-1,5516	-1.07778
P_3	-5,4420E-06	-23071,6	-3,41147	-1,03878
P_4	-6,4668E-06	-16466,2	-2,24478	-1,36372
P_5	1,1649E-05	-22305,6	-3,607	-1,42202
P_6	-2,2073E-05	-25789,4	-4,83354	-1,40646

Tabla 2: Parámetros. Rango de variación y partición del espacio de búsqueda.

5. CONCLUSIONES

Se presentó un modelo dinámico en tres dimensiones y altamente no lineal para el estudio de leucemia mielogenia crónica. Se hallaron puntos de equilibrio del mismo a través del empleo de algoritmos genéticos. Se determinó la estabilidad de los mismos a través del cálculo de los autovalores del sistema linealizado y obtuvieron simulaciones que comprobaron los resultados numéricos hallados.

Este trabajo se encuentra en su fase preliminar, restanto aún describir varios comportamientos dinámicos. Resulta de interés práctico incrementar el espacio de muestreo de los parámetros y además conocer la sensibilidad de los parámetros, si bien se sabe que los parámetros de mayor interés son ζ_6 y ζ_8 .

Agradecimientos:

Los autores agradecen el financiamiento otorgado por la Universidad Nacional de la Patagonia Austral.

6. REFERENCIAS

Goldberg, D.E. Genetic Algorithms in Search, Optimization and Machine Learning, Addison Wesley (1989).

Hale, J. Oscillations in Nonlinear Systems, Dover Pub. Inc., New York (1963).

Holland, J.H. Adaptation in natural and artificial systems, The University of Michigan Press (1975).

Michalewicz, Z. Genetic Algorithms + Data Structures = Evolution Programs, Springer-Verlag (1994).

Moore, H. y N. Li "A mathematical model for chronic myelogenous leukemia (CML) and T cell interaction," Journal of Theoretical Biology, 227, 513-523 (2004).

Strogatz, S. H. Nonlinear Dynamics and Chaos - With Applications to Physics, Biology, Chemistry and Engineering, Perseus Books Publishing, Cambridge (1994).

Wright, A.H. Foundations of Genetic Algorithms- Genetic Algorithms for Real Parameter Optimization, Morgan Kaufmann Publishers, (1991).