



Modelos Discretos Biológicos

Jorge Jaime Juárez Lucero¹, María del Rayo Graciela Guevara Villa², Adriana Hernández Beristain³,
Franco Rojas López¹

¹UPMP-CA-2, Universidad Politécnica Metropolitana de Puebla. Popocateptl S/N, tres cerritos, Puebla, Pue.

²Universidad Politécnica Metropolitana de Puebla. Popocateptl S/N, tres cerritos, Puebla, Pue.

³UPPue-CA-4, Universidad Politécnica de Puebla. Tercer carril del ejido "Serrano" S/N, San Mateo Cuanalá, Juan C. Bonilla, Puebla.

Resumen

Las matemáticas discretas han permitido un importante desarrollo en la biología molecular, han ayudado a crear nuevas áreas de investigación como la biología computacional y la bioinformática. El análisis y recorrido de los grafos permiten lograr la secuenciación del ADN. Los grafos etiquetados donde se asocian pesos a las aristas junto con la teoría de probabilidad permiten predecir la estructura del ARN y de proteínas. Por su parte, la biología molecular también ha aportado a encontrar soluciones a problemas generados dentro de las matemáticas discretas; la solución de Adleman permite encontrar el camino más corto entre dos nodos utilizando secuencias de ADN. Se han reportado bacterias que realizan funciones de lógica básica cuando se reproducen. Este es un breve recorrido de la relación existente entre las matemáticas discretas y la biología molecular.

Palabras clave: Matemáticas discretas, biología molecular, grafos, bacterias

Abstract

Discrete mathematics has an important development in molecular biology; it helped to create new research areas like computational biology and bioinformatics. The graph theory and search algorithms are used for DNA sequence analysis. Graph labeling with edge-labeled and probability theory are used for RNA structure and protein prediction. For other hand, molecular biology collaborates with discrete mathematics; the Adleman solution finds shortest way in graphs using little DNA sequences. It has been reported bacterial that can make logical functions when are growing. This is a brief review of the relation between discrete mathematics and molecular biology.

Keywords: Discrete mathematics, molecular biology, graphs, bacterial.

1 Introducción

Las matemáticas discretas es el área de las matemáticas que estudia los eventos no continuos, es decir, aquellos que sean discretos. Cuando pensamos en matemáticas discretas entonces estamos hablando de objetos contables, de relaciones entre diferentes elementos donde se tenga una relación de uno a uno. Forman parte de las Ciencias de la Computación y engloba el análisis de la lógica matemática, la teoría de conjuntos, la probabilidad, la combinatoria y la teoría de grafos. Por otro lado, la biología molecular se encarga de identificar secuencias genéticas donde se buscan pequeños segmentos que pertenezcan al ADN o al ARN y permitir identificar las subunidades que los conforman. También se puede predecir la estructura del ARN secundario o la forma tridimensional de las proteínas. Sin embargo, en los últimos años algunos estudios de biología molecular han aportado avances hacia problemas propuestos inicialmente en las matemáticas discretas haciendo que estas dos ciencias tengan una relación sinérgica.

2 Matemáticas discretas

Los pilares de las matemáticas discretas están constituidos por la lógica matemática (Patrick & Shirley, 1975); la teoría de conjuntos (Seymour, 1969); la combinatoria (Ralph, 1998); teoría de grafos (Richard, 2003); y Probabilidad (Javier, 2006).



3 Análisis biológico

La integración de las áreas que conforman las matemáticas discretas ha permitido la creación de dos nuevas líneas de investigación dentro de la biología molecular: La bioinformática y la biología computacional.

3.1 Bioinformática

Se refiere a la creación de algoritmos, técnicas estadísticas y teorías empleadas para resolver problemas prácticos derivados del análisis de datos biológicos, como puede ser el diseño de moléculas, el análisis del genoma, antropología evolutiva, interacciones proteicas, modelado tridimensional de genes o proteínas, estudio de micro arreglos, genética forense y el análisis de cualquier dato biomédico (Ellington Roni, 2010)

3.2 Biología computacional

Se dedica a construir hipótesis relacionadas con la investigación de algún problema biológico utilizando computadoras y empleando datos experimentales o simulados, por ejemplo en la predicción de la estructura de proteínas y seguimiento de propagación de enfermedades (Ellington Roni, 2010)

4 Aplicaciones de la matemática discreta para resolver problemas en biología

Conjuntando la lógica matemática, la teoría de conjuntos, la combinatoria, la probabilidad y la teoría de grafos se ha logrado un gran avance dentro de la biología molecular, algunas de sus aplicaciones se presentan a continuación.

4.1 Ensamble de secuencias y teoría de grafos.

La secuenciación del código genético consiste en identificar cómo está estructurado el ADN de una bacteria, célula u organismo, comienza rompiendo físicamente el ADN en millones de fragmentos aleatorios (lo que se realiza en el laboratorio exponiendo la cadena de ADN con enzimas específicas). Supongamos que al terminar este análisis se obtienen los siguientes fragmentos: TGG, GGC, GGA, GAG, GCT, AGG (donde las siglas A, T, G, C corresponden a las bases nitrogenadas Adenina, Timina, Guanina y Citosina, respectivamente y son los pilares del ADN). Con ellas construyamos un grafo Hamiltoniano y un grafo Euleriano.

4.1.1 Grafo de superposición Hamiltoniano y Euleriano

Un camino Hamiltoniano consiste en encontrar una ruta que permita unir por sus aristas a todos los nodos sin repetir uno solo. Si aplicamos esta teoría a los fragmentos de ADN tenemos el grafo de la figura 1a. Si aplicamos el camino Hamiltoniano a este grafo obtenemos el mostrado en la figura 1b. La forma de construir este grafo dirigido fue colocar una arista entre dos nodos cuando se superpongan 2 nucleótidos de ADN, es decir, TGG tendrá una arista con los que tengan GGX donde X representa a cualquier nucleótido. Así TGG tiene una relación con GGC y con GGA, así sucesivamente se construye el grafo y se calcula su camino Hamiltoniano lo que nos proporciona la siguiente secuencia: TGGAGGCT conociendo el ADN del organismo que se estuviera analizando (Ogihara M., 1998). Se puede conseguir el mismo resultado si se realiza con un grafo de Euler. En éste se construyen los nodos tomando solamente dos nucleótidos de ADN por nodo, la arista corresponde a la unión de los nodos, por ejemplo el nodo TG junto con el nodo GG tendrá la arista TGG. Construyendo el camino Euleriano obtenemos la misma secuencia de ADN: TGGAGGCT (ver figura 1 C y D).

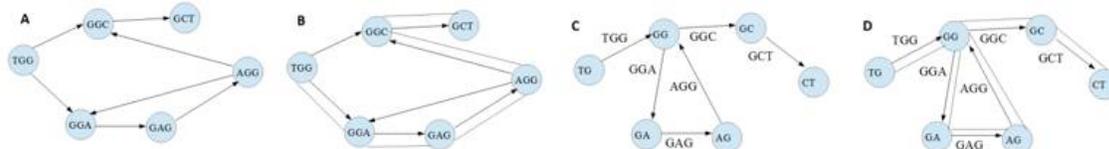


Figura 1. A) Grafo construido a partir de nucleótidos. B) Camino Hamiltoniano del grafo A. C) Grafo construido a partir de nucleótidos conocidos. D) Camino Euleriano del grafo C

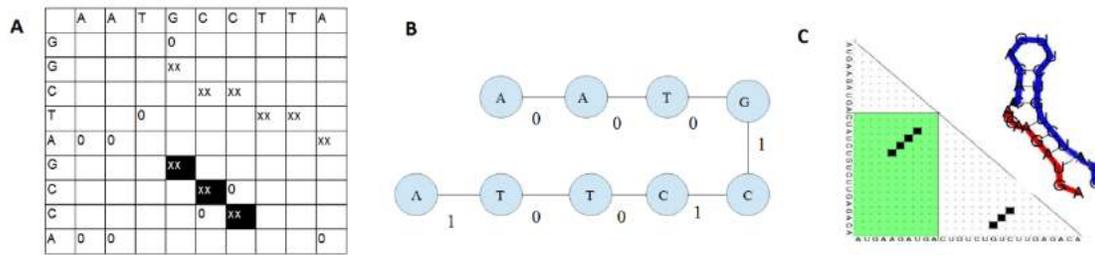


Figura 2. A) Grafo de rejilla. B) Pesos asignados al grafo de rejilla. C) ARN obtenido del análisis del grafo de rejilla.

4.2 Predicción de ARN secundario.

El ARN secundario a menudo tiene funciones biológicas importantes como acarreador de aminoácidos o como parte de ribosomas para ayudar con la síntesis de proteínas. Para conseguirlo primero se construye una matriz de rejilla, posteriormente se analizan las energías existentes en los enlaces de las regiones propuestas a estar unidas y se aplica teoría de probabilidad para conseguir la estructura más favorable y estable. Cuando ya se conoce la secuencia genética y corresponde a una proteína con forma tridimensional desconocida la secuencia es introducida a servidores que permitan predecir su estructura lo cual se realiza con lo que se conocen como grafos de rejilla. El grafo de rejilla es una matriz de m filas por n columnas (figura 2 A). En la columna se coloca la secuencia genética de la proteína desconocida y en la fila la secuencia de la proteína con la que se esté comparando. Cuando exista una correspondencia entre sus nucleótidos (como por ejemplo AGGG en las filas y AGGG en las columnas) se le asigna el valor de 1, en caso de no haber relación el valor será de 0 como se muestra en la figura 2 B. Al final de este análisis el servidor devolverá la forma tridimensional (figura 2 C) de las correspondencias (Meidanis, 1997. Ellington Roni, 2010).

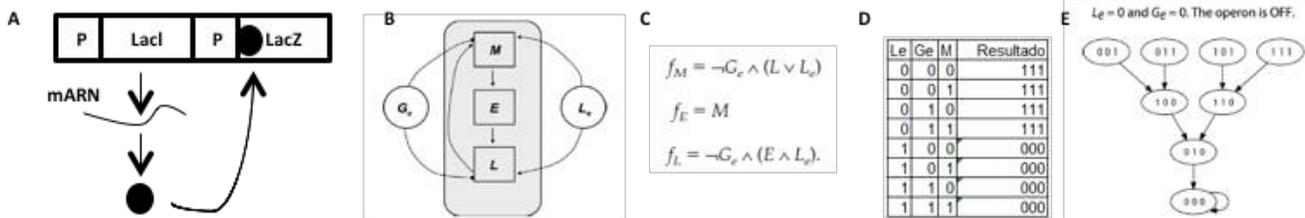


Figura 3. A) Modelo biológico del Operón. B) Modelo de bloques del Operón. C) Funciones de transferencia del Operón. D) Valores posibles para las funciones de transferencia. E) Grafo de combinaciones que tienen el Operón apagado.

4.3 Modelado del Operón

El operón es un complejo formado entre proteínas y ADN que dependiendo de su interacción con el exterior y ante la presencia de algún tipo de azúcar permite sintetizar la enzima que pueda degradar el alimento y ayudar a la bacteria a alimentarse (figura 3 A). En ausencia del azúcar, el gen se bloquea y deja de producir enzimas con el consiguiente ahorro energético. Para resolverlo, se realiza el grafo de la figura 3 B tomando como nodos los elementos involucrados como puede ser la presencia de lactosa interna, lactosa externa, la existencia de ARN mensajero y que contenga o no galactosa externa. Se colocan aristas que relacionen los factores internos y los factores externos tomando funciones de transición unidos por conectores que dándoles valores de 0 ó 1 se pueden calcular todos los estados posibles que permitan que el gen del operón se encuentre encendido o apagado como puede verse en la figura 3 C, D y E (Robeva Raina, 2010).

5 Biología molecular aplicada para resolver problemas en matemáticas discretas

Por el lado de la biología molecular se han hecho aportes hacia las matemáticas discretas y a continuación veremos algunos ejemplos.

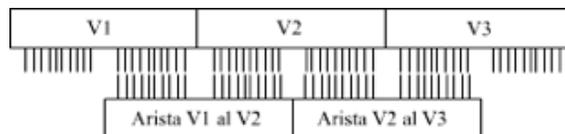


Figura 4. Grafo construido con nucleótidos de ADN. Las aristas son nucleótidos de una cadena doble de ADN.

5.1 Computadoras de ADN y solución de Adleman

Uno de los problemas que más ha generado gasto computacional es la búsqueda del camino más corto entre dos nodos, mientras más nodos aparezcan en el sistema más se complica el análisis de los datos. Adleman encontró la forma de hallar este camino por medio del uso de secuencias genéticas. Asigna nucleótidos a cada nodo y toma los segmentos de ADN construidos para formar una hebra de ADN que incluya todos los nucleótidos obteniendo una cadena doble de ADN completa que al secuenciarla contiene la ruta más corta que incluye todos los nodos como lo muestra la figura 4. (Leonard, 1994).

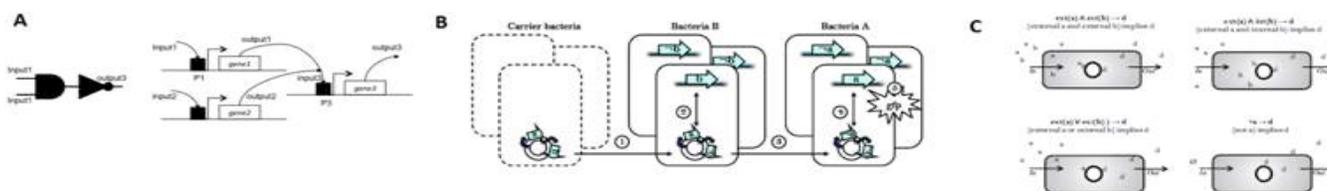


Figura 5. A) Esquema genético de una compuerta AND y NOT. B) Bacterias modificadas para expresar la compuerta del esquema A. C) Bacterias genéticamente modificadas para expresar diferentes compuertas lógicas.

5.2 Compuertas genéticas, lógica bacteriana y compuertas bacterianas

Con el desarrollo de la biología sintética (modificar bacterias para que realicen funciones útiles al hombre), se han desarrollado compuertas lógicas empleando secuencias genéticas interactuando con proteínas, la figura 5 A muestra el equivalente genético de una compuerta AND negada. (Bode, 2001). La figura 5 B y C presentan bacterias modificadas para que durante su crecimiento puedan resolver problemas simples de lógica matemática (Anderson Christopher, 2007).

6 Conclusiones

Los avances más importantes en bioinformática y en biología computacional se han dado gracias a las matemáticas discretas. Éstas son un instrumento básico para las ciencias computacionales con una relación muy estrecha con la biología molecular. Por otro lado, la biología molecular comienza a aportar soluciones a problemas complejos dentro de las matemáticas discretas que posiblemente no puedan realizarse con los equipos de cómputo con que disponemos actualmente. Con el avance que está teniendo la biología sintética se busca la construcción de una computadora formada a base de bacterias o a base de secuencias de ADN.

7 Referencias

- Anderson Christopher, V. C. (2007). Environmental signal integration by a modular AND gate. *Mol Syst Biol*, 1-8.
- Bode, B. (2001). Coping with kinetic and thermodynamic barriers: RMCE, an efficient strategy for the targeted integration of transgenes. *Current opinion in biotechnology*, 473-480.
- Ellington Roni, W. J. (2010). RNA Secondary structure prediction by using discrete mathematics: An interdisciplinary research experience for undergraduate students. *Life Science Education*, 348-356.
- Javier, M. P. (2006). *Fundamentos de probabilidad*. México: Ediciones Paraninfo.
- Leonard, A. (1994). Molecular computational of solutions to combinatorial problems. *Science*, 1021-1024.
- Meidanis, S. J. (1997). *Introduction to computational molecular biology*. USA: PWS Publishing company.
- Ogihara M., R. A. (1998). Circuit evaluation: thoughts on a killer application in DNA computing. *Computing with Bio-molecules. Theory and Experiments*, 111-126.
- Patrick, S., & Shirley, H. (1975). *Primer curso de lógica matemática*. España: Reverte.
- Ralph, G. (1998). *Matemáticas discretas y combinatoria*. México: Pearson Educación.
- Richard, T. (2003). *Introduction to graph theory*. USA: Dover Books on Mathematics.
- Robeva Raina, D. R. (2010). Mathematical biology modules based on modern molecular biology and modern discrete mathematics. *CBE Life Sciences Education*, 227-240.
- Seymour, L. (1969). *Teoría de conjuntos y temas afines*. Colombia: Mc Graw Hill.