

# Autómatas Celulares en la Simulación de la Dinámica de Dominios Proteicos

Juan C. Castrejón E.

Universidad Nacional Autónoma de México

6 de Junio de 2023



# Tabla de Contenidos

- 1 Introducción
- 2 Pregunta de Investigación
- 3 Objetivo
- 4 Método
- 5 Definición del automáta
- 6 Resultados
- 7 Futuro



# Importancia de la evolución de dominios de proteínas en la biología y medicina.

- Pueden proporcionar información sobre la historia de la vida y la función de los organismos actuales.
- Se podría prever cómo progresarían las enfermedades o cómo podrían evolucionar los patógenos, lo que puede informar el desarrollo de nuevos tratamientos o medicamentos.



## Uso de métodos computacionales para modelar y predecir la evolución de dominios de proteínas.

- Aprovechar el modelo de sustitución de aminoácidos propuesto por Margaret Dayhoff.

[illegible]

Figure: Matriz de Dayhoff.



# Autómatas celulares como herramienta poderosa para modelar sistemas dinámicos complejos.

- Los autómatas celulares son modelos matemáticos que simulan sistemas complejos mediante reglas simples. Un autómatá celular tiene una cuadrícula de celdas, cada una de las cuales puede estar en un número finito de estados. El estado de una celda en un momento dado depende de su propio estado y del estado de sus vecinos en el paso de tiempo anterior, según un conjunto de reglas de transición.



# Autómatas celulares como herramienta poderosa para modelar sistemas dinámicos complejos.

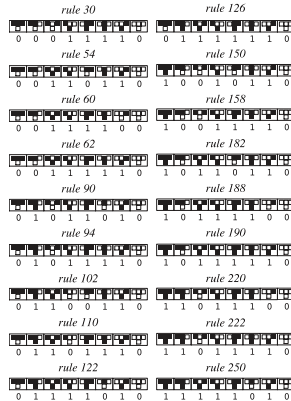


Figure: Ejemplos de reglas



# Autómatas celulares como herramienta poderosa para modelar sistemas dinámicos complejos.

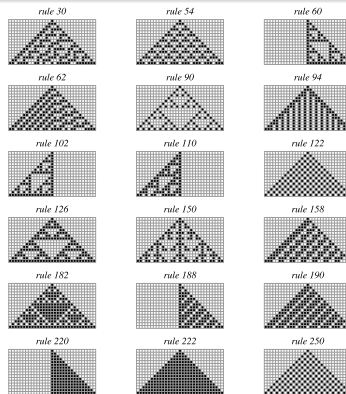


Figure: Ejemplos de reglas propagadas.



# Pregunta de Investigación

- ¿Cómo se pueden utilizar los autómatas celulares para simular la evolución de dominios de proteínas?





# Objetivo

- Desarrollar un modelo basado en autómatas celulares para simular la evolución de dominios de proteínas.



# Método

- Crear un autómatá celular para representar la estructura y función de una proteína.
- Definir reglas de transición basadas en la dinámica de evolución de dominios de proteínas.
- Simular la evolución de dominios de proteínas utilizando el autómatá celular propuesto.



## Definición del autómatá

- Se utilizará un autómatá celular 1D (CA) para simular la evolución del dominio de las proteínas.
- El CA es un arreglo unidimensional de celdas, cada una con uno de  $(g+3)$  posibles estados, que representan clases de dominios, dos símbolos de terminación y un estado vacío.
- Al inicializar, la celda central se establece en un dominio ancestral, mientras que las demás se configuran como vacías.
- El CA evoluciona durante una cantidad de pasos de tiempo igual al número de dominios en la proteína con la mayor cantidad de dominios en el conjunto de datos de entrenamiento.



## Reglas de transición

- Regla A (Herencia): Si una celda ha evolucionado hacia un dominio (es decir, su estado no es "0/"), heredará este dominio en el próximo paso de tiempo.
- Regla B (Evolución de adelante hacia atrás): Si una celda y su vecino de la derecha están ambos en el estado vacío, y su vecino de la izquierda está en un estado de dominio, la celda evolucionará hacia un dominio en el próximo paso de tiempo. El dominio se selecciona utilizando el algoritmo de selección de ruleta basado en la matriz F2B (probabilidad previa de evolución de adelante hacia atrás).



## Reglas de transición

- Regla C (Evolución de atrás hacia adelante): Si una celda y su vecino de la izquierda están ambos en el estado vacío, y su vecino de la derecha está en un estado de dominio, la celda evolucionará hacia un dominio en el próximo paso de tiempo. El dominio se selecciona utilizando el algoritmo de selección de ruleta basado en la matriz B2F (probabilidad previa de evolución de atrás hacia adelante).
- Regla D (Retorno a la inanimación): Si una celda y sus vecinos están todos en el estado vacío, la celda permanecerá en el estado vacío en el próximo paso de tiempo.



## Reglas de transición

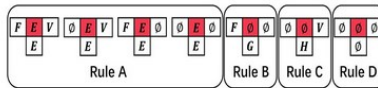


Figure: Ayuda visual para reglas.



# Resultados

- Se ha mostrado una implementación novedosa para representar la evolución de los dominios proteicos, tomando como ventaja la naturaleza compleja de los autómatas celulares para reflejar con simplicidad un fenómeno difícil de estudiar que se puede aprovechar en múltiples áreas científicas.



## Futuro

- La implementación usa dominios proteicos tomados de una base de datos establecida para generar estructuras reales.
- Un proceso llamado MDAT (herramienta de alineación multi dominio) puede ser usada con estructuras reales y las generadas por este método para producir valores útiles que miden la similitud en la evolución de las proteínas.
- ¿Qué tan factibles son los métodos de generación aleatoria para la búsqueda de estructuras?





## Referencias

- Qiu, Y., Zhang, X. (2020). Using Cellular Automata to Simulate Domain Evolution in Proteins. *Frontiers in Genetics*, 11, 515. <https://doi.org/10.3389/fgene.2020.00515>
- Schwartz, J. T., Neumann, J. V., and Burks, A. W. (1967). Theory of self-reproducing automata. *Q. Rev. Biol.* 21:745. doi: 10.2307/2005041
- Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978) A model of evolutionary change in proteins, in Dayhoff, M.O. Edition, *Atlas of Protein Sequence and Structure*. Natl. Biomed. Res. Found., Washington DC, 5(3), 345- 352.



## Recursos

- Repositorio de Github <https://github.com/JCastrejonE/proyecto-computacion-genomica/tree/main/>
- Código de esta presentación <https://github.com/JCastrejonE/proyecto-computacion-genomica/blob/main/presentacion/main.tex>

