

COMPARACIÓN DE MÉTODOS PARA LA PREDICCIÓN DE ESTRUCTURAS SECUNDARIAS DE ARN MINIMIZANDO LA ENERGÍA LIBRE

Comparing methods for RNA secondary structure prediction minimizing the free energy

Pedro Ayala Elizarraraz, Arturo González Gutiérrez¹, Rolando Tenoch Bárcenas Luna¹

¹ Universidad Autónoma de Querétaro

*Correo electrónico: aglez@uaq.mx

RESUMEN

El comportamiento de una molécula de ARN depende directamente de sus estructuras secundarias y terciarias; sin embargo, se ha demostrado que la predicción de estructuras secundarias con pseudonudos arbitrarios es un problema NP-completo. En este artículo se presenta un estudio comparativo de tres algoritmos basados en la técnica de programación dinámica. Los algoritmos de Nussinov y Zuker predicen estructuras secundarias sin pseudonudos, mientras que el algoritmo de Akutsu lo hace con pseudonudos simples. Asimismo, se muestra una metaheurística que utiliza un algoritmo genético para producir subestructuras cuasióptimas que a su vez permiten la predicción de estructuras secundarias con pseudonudos simples. Finalmente, se aborda un análisis experimental de los cuatro algoritmos utilizando instancias públicas de estructuras de ARN provistas por las bases de datos RNA STRAND y PseudoBase++.

PALABRAS CLAVE: Estructura Secundaria de ARN, Mínima Energía Libre, Pseudonudo, Programación Dinámica, Algoritmo Genético.

ABSTRACT

The behaviour of an RNA molecule is directly linked to its secondary and tertiary structures. However, it has been proved that the secondary structure prediction problem with arbitrary pseudoknots is an NP-complete problem. In this paper, a comparative study of three algorithms based on the dynamic programming technique is presented. The Nussinov and Zuker algorithms predict secondary structures without pseudoknots, whereas Akutsu algorithm does it with simple pseudoknots. A metaheuristic that uses a genetic algorithm to produce quasi-optimal substructures is shown which in turn allows the prediction of secondary structures with simple pseudoknots. Finally, an experimental analysis is proposed for the four algorithms using public instances of RNA structures provided by RNA STRAND and PseudoBase ++ databases.

KEYWORDS: RNA Secondary Structure, Minimum Free Energy, Pseudoknot, Dynamic Programming, Genetic Algorithm.

INTRODUCCIÓN

El ácido ribonucleico (ARN) es una molécula que consiste en una cadena formada por los nucleótidos: Adenina (A), Citosina (C), Guanina (G) y Uracilo (U). En comparación con el ADN, las moléculas de ARN son menos estables y muestran una mayor variabilidad en

su estructura tridimensional [1]. El ARN posee una gran relevancia, ya que está involucrado en varios procesos biológicos, tales como la codificación y decodificación de información genética; regulación, detección y comunicación de respuestas a señales celulares; así como la catalización de las reacciones biológicas [2].

Debido a que la funcionalidad de una molécula de ARN depende en gran medida de sus estructuras secundarias y terciarias, es importante su estudio y caracterización [3]. La caracterización y modelación de la estructura terciaria representa un mayor desafío científico y tecnológico que la estructura secundaria. Por ello, los esfuerzos se han enfocado en gran medida en la predicción de estructuras secundarias, ya que es experimentalmente accesible y contiene información para determinar la relación entre estructura y funcionalidad [4].

La estructura secundaria de una secuencia de ARN consiste en el plegamiento de la cadena de nucleótidos consigo misma mediante los enlaces de hidrógeno entre sus bases y en diferentes posiciones sobre la misma, donde cada nucleótido sólo puede formar parte de un par base [2].

DEFINICIONES

Secuencia de ARN: Sea $r=r_1\dots r_n$ una cadena y $\Sigma_{ARN}=\{A,C,G,U\}$ el conjunto que representa el alfabeto de los nucleótidos que conforman el ARN. Si r es una cadena conforme al alfabeto Σ_{ARN} , entonces, r es una secuencia de ARN.

La estructura secundaria se puede representar por el conjunto S_r de pares de índices, que corresponden con las posiciones de los pares base sobre la secuencia r ,

$$S_r \subseteq \{(i,j) \mid 1 \leq i < j \leq n\}$$

donde n es la longitud de la secuencia, i y j son posiciones en la cadena que representan bases complementarias sujetas a las siguientes restricciones [5], [6]:

1. Cada índice $k \in \{1, \dots, n\}$ ocurre a lo más una vez en el emparejamiento de S_r .
2. Para cada par (i,j) de S_r , el par de bases (i,j) es un par de tipo Watson-Crick: $(r_i,r_j) \in \{(A,U),(U,A),(-C,G),(G,C)\}$, o es un par tipo Wobble: $(r_i,r_j) \in \{(G,U),(U,G)\}$.
3. Para cada par (i,j) de S_r , $j-i \geq 4$ (donde 4 es el tamaño mínimo de una región de bases no complementarias llamada bucle en horquilla).

En la Figura 1, se muestran las subestructuras elementales que pueden contener la estructura secundaria mediante métodos de plegamiento por maximización



de pares base o por minimización de energía libre entre moléculas.

46

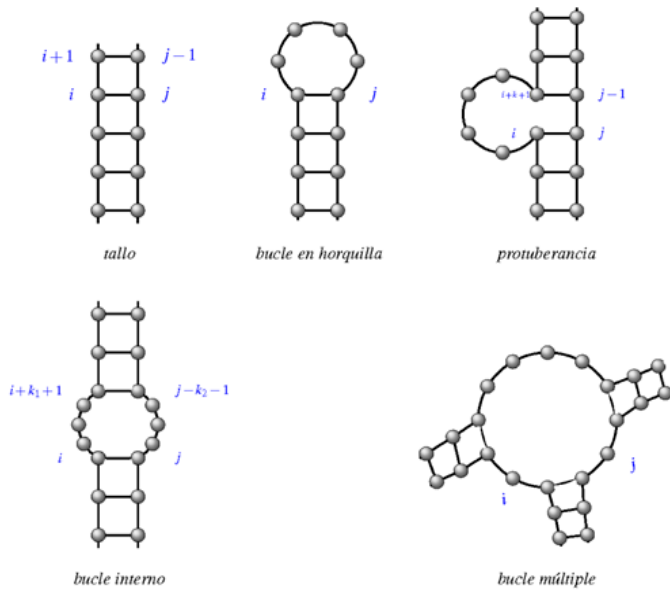


Figura 1. Representación de las posibles subestructuras que una cadena de arn puede adoptar. Las bases se muestran como puntos y las líneas representan las conexiones entre las bases [5]. Los índices i, j, k, k_1 y k_2 representan los pares base a partir de los cuales se forma la subestructura.

A continuación, se define el problema de la predicción de la estructura secundaria.

Predicción de la estructura secundaria: Dada una secuencia r conforme al alfabeto Σ_{ARN} , realizar el plegamiento de la secuencia utilizando las restricciones anteriormente listadas, generando las subestructuras permitidas [5], [7], [8], para obtener la estructura secundaria S_r con la mínima energía libre o la estructura que maximice el número de pares base.

Una topología o tipo de subestructura más compleja es el pseudonudo (Figura 2), la cual se caracteriza por el emparejamiento de bases en un bucle con residuos no apareados de la subestructura. Enseguida se describe formalmente el pseudonudo.

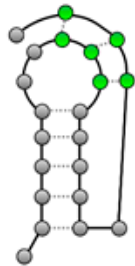


Figura 2. Representación esquemática de un pseudonudo. Las líneas punteadas representan las conexiones entre los pares base complementarios. Las conexiones entre pares base de los nodos resaltados conforman el pseudonudo.

Pseudonudo: Para todos los pares base de una estructura secundaria, si existen dos pares base (i,j) y (k,l) que no son continuos ($i < j < k < l$), ni anidados ($i < k < l < j$), pero se intersectan ($i < k < j < l$), entonces la estructura secundaria contiene un pseudonudo [9].

La intersección de los pares base (i,j) y (k,l) que obedece a la restricción ($i < k < j < l$), se muestra en la Figura 3. Si consideramos el par base (i,j) con las posiciones (1,13) y al par base (k,l) con las posiciones (6,21), claramente se observa esta intersección o superposición de pares base ($1 < 6 < 13 < 21$).

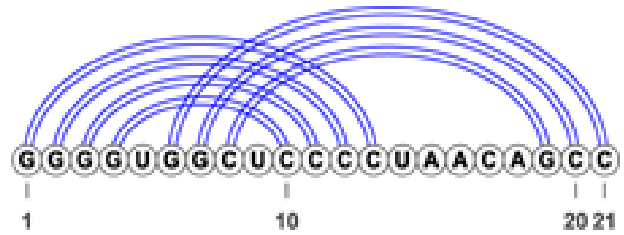


Figura 3. Representación planar de un pseudonudo.

Con base en las definiciones anteriores, el problema de la predicción de estructuras secundarias con pseudonudos se define como sigue:

Predicción de la estructura secundaria con pseudonudos: Dada una secuencia de ARN, la salida de la predicción con pseudonudos será la estructura secundaria que incluya pseudonudos que contienen la información de emparejamiento de bases.

METODOLOGÍA Y MATERIALES

Las estructuras secundarias se determinan mediante técnicas de laboratorio. Las técnicas más utilizadas para determinar las estructuras de ARN son los rayos X, cristalografía y la resonancia magnética nuclear (NMR) [3], [10]. En los últimos años ha surgido un mayor interés por la predicción de dichas estructuras mediante la modelación computacional. Con los métodos computacionales es posible reducir el tiempo y costo de la predicción de estructuras secundarias para una molécula [3]; sin embargo, la modelación computacional no está exenta de limitantes. Un claro ejemplo de ello es la predicción de estructuras con pseudonudos, la cual incrementa la complejidad tanto en tiempo como en espacio. Se ha demostrado que la predicción de estructuras con pseudonudos utilizando el modelo termodinámico de minimización de la energía libre es un problema NP-completo [11], [12].

A lo largo de la historia se han implementado diferentes algoritmos utilizando técnicas como programación dinámica y algoritmos genéticos principalmente.

Entre los algoritmos que excluyen la predicción de pseudonudos y que se basan en la técnica de programación dinámica se encuentran:

Nussinov [13], que utiliza un enfoque de maximización de pares base bajo la hipótesis de que, entre más emparejamientos existan entre bases de la cadena de ARN, menor será la energía libre de la molécula.

Zuker [14], que emplea el concepto de minimización de energía libre mediante la asignación de energías termodinámicas a cada tipo de subestructura que conforman la estructura secundaria final; estas energías dependen del tamaño de las subestructuras, y la energía total de la secuencia se representa como la suma de todas las contribuciones individuales de estas subestructuras.

Ambos algoritmos exhiben un orden de complejidad $O(n^3)$.

Por otro lado, respecto a los algoritmos que incluyen la predicción con pseudonudos, destacan los propuestos por:

Rivas [15], que propone un algoritmo con enfoque de minimización de la energía libre y con restricción en el tipo de pseudonudo con un orden de complejidad $O(n^6)$.

Akutsu [11], que establece un algoritmo que maximiza el número de pares base e incluye la predicción de pseudonudos simples con un orden de complejidad $O(n^4)$.

Del algoritmo de Akutsu en [16] se propone una modificación que incluye el modelo termodinámico de minimización de energía libre y reduce la complejidad del espacio de $O(n^3)$ a $O(n^2)$, con un tiempo de ejecución de $O(n^4)$.

Con respecto a enfoques metaheurísticos para la predicción de estructuras secundarias, en los últimos años se ha propuesto una variedad de algoritmos genéticos, como GAKnot [9], el cual genera un conjunto de hélices mediante un enfoque similar al del algoritmo Nussinov. A partir de este conjunto de hélices se genera la población inicial, donde cada individuo se crea a través de las combinaciones de las hélices previamente encontradas. Posteriormente, se aplican los operadores genéticos de selección, cruza y mutación para obtener la solución.

El algoritmo propuesto en [17], GA-GRASP, combina un algoritmo genético y la metaheurística GRASP (Greedy Randomized Adaptive Search Procedure) en combinación del modelo termodinámico de energía libre. El algoritmo RNAknot [18] es una mejora del algoritmo GA-GRASP que incluye la predicción de todos los tipos de subestructuras básicas (Figura 1) y dos tipos de pseudonudos: simple (H-type) y Kissing-Hairpin, con un orden de complejidad $O(n^2) + O(k \times nbs)$, donde k es

el número de iteraciones y nbs es el número de individuos en la población.

En la Figura 4 se muestra el flujo básico de los algoritmos genéticos mencionados.



Figura 4. Diagrama de flujo del algoritmo genético.

En la Tabla 1 se presentan los enfoques de minimización de energía de cada modelo y su correspondiente algoritmo, así como sus restricciones en cuanto a si su predicción incluye o no los pseudonudos, indicando el tipo de pseudonudo en su caso.

Tabla 1. Caracterización de los algoritmos para modelación de estructuras secundarias.

ALGORITMO	MODELO DE MINIMIZACIÓN DE ENERGÍA	PREDICCIÓN CON PSEUDONUDOS	TIPO
Nussinov	Maximización de pares base	N	-
Zuker	Modelo Termodinámico	N	-
Akutsu	Modelo Termodinámico	S	Simple
RNAknot	Modelo Termodinámico	S	Kissing Hairpin

PARÁMETROS DE ENERGÍA

La base de datos del vecino más cercano (Nearest Neighbor Database, NNDB) [19] es un modelo termodinámico que asume que la estabilidad de un par base específico depende directamente de las bases vecinas. Esta base de datos proporciona los parámetros de energía que permiten la predicción de una estructura secundaria estable.

La estabilidad de una estructura secundaria se mide por el cambio de energía libre ΔG . Este cambio se aproxima como la suma de las contribuciones individuales

de cada subestructura que conforma la estructura secundaria [20]. A continuación, se enlistan los cambios de energía que pueden ocurrir durante la construcción de la estructura secundaria:

$\delta G = 0$ indica equilibrio en la estructura.

$\delta G > 0$ indica desestabilización en la estructura.

$\delta G < 0$ indica que la estructura es estable.

Las reglas de cambio de energía se dividen por tipo de estructura: hélices, energías desestabilizadoras por tamaño de bucle (horquilla, protuberancia, interno), bucles internos generales (simétricos, asimétricos), extremos no apareados y bucles múltiples. Para cubrir la contribución energética de cada tipo de subestructura, la base de datos NNDB cuenta con un total de 12 883 parámetros de energía.

En la Tabla 2 se muestra la energía libre δG entre cada par base que conforma una hélice dentro de la estructura secundaria. Estos parámetros de energía son estándar y se aplican en la construcción de cualquier hélice.

Tabla 2. Energía (kcal/(mol)) de las hélices entre cada par base.

	AU	CG	GC	UA	GU	GU
AU	-0.9	-2.2	-2.1	-1.1	-0.6	-1.4
CG	-2.1	-3.3	-2.4	-2.1	-1.4	-2.1
GC	-2.4	-3.4	-3.3	-2.2	-1.5	-2.5
UA	-1.3	-2.4	-2.1	-0.9	-1.0	-1.3
GU	-1.3	-2.5	-2.1	-1.4	-0.5	+1.3
UG	-1.0	-1.5	-1.4	-0.6	+0.3	-0.5

SECUENCIAS DE ARN

En la Tabla 3 y Tabla 4, se muestran las secuencias de ARN utilizadas en el desarrollo del proyecto. De cada secuencia se cuenta con la estructura secundaria obtenida mediante experimentación (rayos X y/o resonancia magnética nuclear), la cual es utilizada como la estructura de referencia para la validación de los resultados.

Tabla 3. Secuencias de ARN sin pseudonudos en la estructura de referencia. Las secuencias fueron obtenidas de la base de datos RNA STRAND [21].

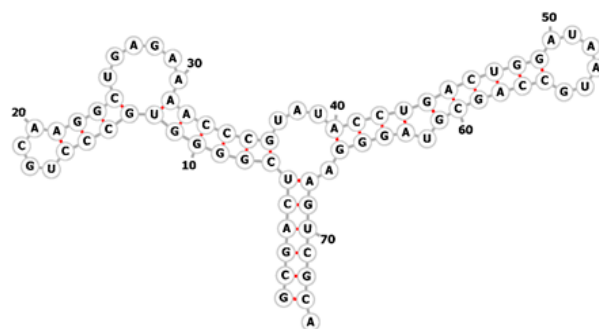
CLAVE	TIPO	LONGITUD NT	MÉTODO DE VALIDACIÓN
PDB_01130	ARN sintético	60	Rayos X
PDB_00481	ARN sintético	62	Rayos X
PDB_00136	Intron del Grupo II	70	Rayos X
PDB_01115	ARN sintético	74	Rayos X
PDB_01113	ARN sintético	75	Rayos X
PDB_01128	ARN sintético	75	NMR
PDB_01125	ARN sintético	86	NMR
PDB_00213	ARN sintético	101	NMR

CLAVE	TIPO	LONGITUD NT	MÉTODO DE VALIDACIÓN
PDB_00213	ARN sintético	126	Rayos X
PDB_01236	ARN sintético	153	Rayos X

Tabla 4. Secuencias de ARN con pseudonudos en la estructura de referencia. Las secuencias fueron obtenidas de la base de datos PseudoBase ++ [22].

CLAVE	TIPO	LONGITUD NT	MÉTODO DE VALIDACIÓN
PKB5	ARNt viral	41	NMR
PKB73	ARNm	33	NMR
PKB74	ARNm	28	NMR
PKB80	Cambio de marco Viral	49	NMR
PKB81	ARN sintético	26	NMR
PKB107	Cambio de marco Viral	52	NMR
PKB243	-	121	NMR
PKB273	Viral	48	NMR
PKB280	Cambio de marco Viral	68	NMR
PKB281	Cambio de marco Viral	43	NMR

En la Figura 5 se muestra la estructura correspondiente a la secuencia PDB_01115 en forma esquemática y notación Dot-Bracket.



GCGACUCGGGGUGCCUGCAAGGCGUGAGAAAACCCGUUACCUGA-CUGGAUAAUGCCAGCGUAGGGAAGUCGCA
((((((((((((((((.....)))))).....))))))...((((((((((((((((.....)))))).....)))))).....))))))

Figura 5. Estructura secundaria correspondiente a la secuencia pdb_01115.

SECUENCIAS DE ARN

A continuación, se enlistan los pasos que conforman la metodología para la predicción de estructuras secundarias de ARN (Figura 6).

1. Secuencia de ARN: obtención y almacenamiento de las secuencias de ARN. Dichas secuencias son almacenadas localmente en archivos de texto plano en formato FASTA.
2. Cálculo de la energía libre: esta es una etapa en conjunto con el algoritmo de predicción; para ello se utiliza el modelo termodinámico del vecino más

- cercano, en el que se resume la contribución energética de cada subestructura que conforma la estructura secundaria final.
3. Algoritmo de predicción: se producen cuatro soluciones conforme a los algoritmos de Nussinov, Zuker, Akutsu y RNAknot.
 4. Notación Dot-Bracket: estructura secundaria predicha en un formato específico para permitir su posterior visualización.
 5. Estructura secundaria: es el resultado de la predicción (estructura secundaria), en el formato establecido y energía libre o número máximo de pares base, con la cual el prototipo desplegará el resultado.

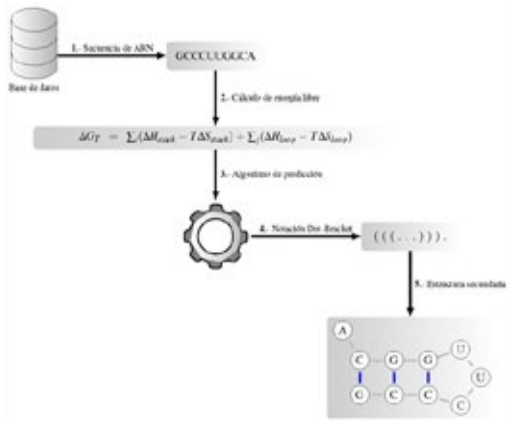


Figura 6. Diagrama de flujo del proceso de la predicción de estructuras secundarias.

RESULTADOS Y DISCUSIÓN

Para evaluar la precisión de la predicción de los métodos, se utiliza la técnica descrita en [23], que utiliza los conceptos de sensibilidad y valor predictivo positivo.

La sensibilidad (S) representa el porcentaje de pares base presentes en la estructura de referencia que ocurren en la estructura secundaria predicha, este valor es calculado con la siguiente ecuación:

$$S = \frac{VP}{VP + FN} \quad (1)$$

El valor predictivo positivo (VPP) describe qué porcentaje de pares base predichos ocurren en la estructura secundaria de referencia, se obtiene mediante la ecuación:

$$VPP = \frac{VP}{VP + FP} \quad (2)$$

Donde VP es el número de pares base predichos correctamente, FN el número de pares base en la estructura de referencia que no están en la estructura predicha y FP es el número de pares base predichos que no están en la estructura de referencia.

Adicionalmente, se incluye la medida de precisión Valor-F, que varía entre 0 y 1, donde 0 indica que no hay pares base en común entre ambas estructuras y 1 indica una precisión perfecta. Se obtiene con la siguiente ecuación:

$$Valor - F = \frac{2 \times S \times VPP}{S + VPP} \quad (3)$$

ANÁLISIS DE RESULTADOS

El conjunto de secuencias de prueba utilizado para evaluar el rendimiento, a través de los conceptos de sensibilidad (Ec. (1)) y de valor predictivo positivo (Ec. (2)), consta de 20 secuencias; 10 son libres de pseudonudos y las otras 10 incluyen un pseudonudo de tipo simple en la estructura de referencia. Se obtuvo la eficiencia con respecto a las métricas definidas mediante la ejecución de cada algoritmo.

En la Tabla 5 se muestra el promedio del Valor-F (Ec. (3)) por algoritmo, agrupado por estructura con o sin pseudonudo en la estructura de referencia. La estructura obtenida de manera experimental se asume que tiene un Valor-F de 1.

Tabla 5. Valor-F promedio en estructuras sin y con pseudonudos.

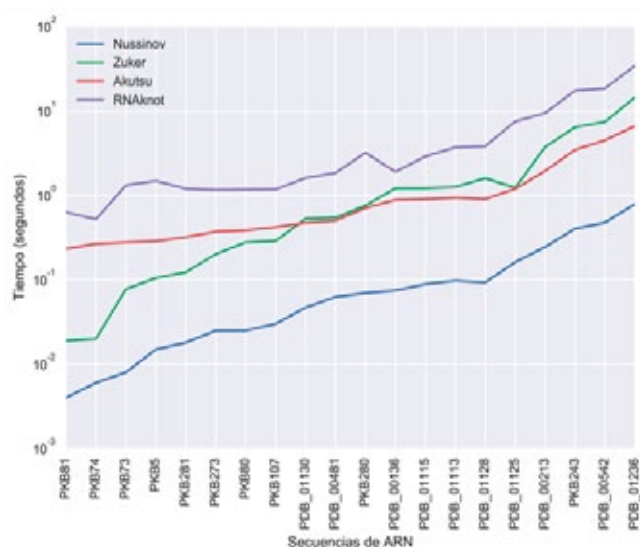
ALGORITMO	VALOR-F	
	ESTRUCTURAS SIN PSEUDONUDOS	ESTRUCTURAS CON PSEUDONUDOS
Nussinov	0.2655	0.2329
Zuker	0.7816	0.4817
Akutsu	0.4766	0.7349
RNAknot	0.4609	0.8039

Con respecto a las instancias sin pseudonudos, el algoritmo Zuker exhibe una mayor precisión con un 78.1 %, seguido por Akutsu con 47.6 %, luego RNAknot con 46 % y finalmente Nussinov con 26.5 %. El de Zuker es buen resultado, ya que la mayoría de algoritmos de predicción logran un porcentaje de efectividad de entre 60 % y 80 %, como se muestra en las comparaciones realizadas en [9] y [23]. En tanto para Akutsu y RNAknot, el porcentaje es superior a 45 %, lo cual es entendible debido a que la estructura de referencia no incluye pares base que formen un pseudonudo, pero la predicción de estos algoritmos incluye posibles pseudonudos en la mayoría de las instancias. Sin embargo, hay instancias como PDB_00136, PDB_B01125 y PDB_00213, donde la métrica VPP supera el 70 % (Apéndice A, Tabla 7). Finalmente, el porcentaje es muy bajo respecto a la efectividad del algoritmo Nussinov; la principal razón parece estar ligada a que este método se basa en la

tiempo de ejecución; sin embargo, como se mostró anteriormente, el algoritmo RNAknot tiene un mejor desempeño en cuanto a precisión con respecto a la estructura de referencia. Lo mismo sucede para los algoritmos Nussinov y Zuker: mientras que el algoritmo Nussinov tiene un mejor desempeño en tiempo de ejecución, el algoritmo de Zuker tiene por mucho un mejor desempeño en cuanto a la precisión de la estructura secundaria predicha.

Figura 9. Tiempos de ejecución.

Los algoritmos Nussinov, Zuker y Akutsu se implementaron construyendo un sistema en el lenguaje de programación Python versión 3.7.3, mientras que para el



algoritmo RNAknot se utilizó el software implementado en [18]. Las pruebas se realizaron en una computadora portátil Dell G3 con procesador Intel(R) Core(TM) i5-8300H CPU@ 2.30GHZ, RAM 16GB.

CONCLUSIONES

En este artículo se presenta un análisis sobre los principales algoritmos basados en programación dinámica y algoritmos genéticos para la predicción de estructuras secundarias. Con los resultados obtenidos a través de las métricas planteadas observamos que los algoritmos RNAknot, Akutsu y Zuker, que incluyen un modelo termodinámico para la minimización de energía, tienen un porcentaje de precisión mayor que el algoritmo Nussinov, que utiliza un enfoque de maximización de pares base. También se observa que existe una variabilidad en la precisión y sensibilidad con respecto a las diferentes instancias. Por ello, es importante incluir varios algoritmos en el desarrollo de un sistema robusto para la predicción de estructuras secundarias de ARN, para obtener una mejor propuesta a partir de los

diferentes resultados.

Como trabajo futuro se planea el diseño e implementación de un algoritmo genético que incluya el modelo termodinámico de minimización de energía, el cual puede ser alimentado con poblaciones iniciales de estructuras secundarias subóptimas desde el punto de vista energético.

Con respecto a la validación de resultados, un trabajo a futuro interesante consiste en incluir una validación desde la perspectiva topológica de la estructura que complemente los parámetros de sensibilidad y precisión.

REFERENCIAS

- [1] A. Polanski and M. Kimmel, *Bioinformatics*. Berlin: Springer Verlag, 2007.
- [2] A. Punetha, P. Sarkar, S. Nimkar, H. Sharma, Y. KNR, and S. Nagaraj, "Structural bioinformatics: Life through the 3d glasses," in *Bioinformatics: Sequences, Structures, Phylogeny* (A. Shanker, ed.), pp. 191–253, Springer Verlag, 2018.
- [3] A. E. Fatmi, M. A. Bekri, and S. Benhlima, "Rna secondary structure prediction based on genetic algorithm and comparative approach," in *2018 4th International Conference on Optimization and Applications (ICOA)*, pp. 1–7, April 2018.
- [4] A. Churkin, L. Weinbrand, and D. Barash, *Free Energy Minimization to Predict rna Secondary Structures and Computational rna Design*, pp. 3–16. New York, NY: Springer New York, 2015.
- [5] H.-J. Böckenhauer and D. Bongartz, *Algorithmic Aspects of Bioinformatics*. Berlin: Springer Verlag, 2007.
- [6] Z. Liu, D. Zhu, W. Cui, and N. Liu, "An approximation scheme for rna folding structure prediction including pseudoknots," in *2013 Ninth International Conference on Computational Intelligence and Security*, pp. 6–10, Dec 2013.
- [7] C. Reidys, *Combinatorial Computational Biology of rna: Pseudoknots and Neural Networks*. Berlin: Springer Verlag, 2011.
- [8] R. Lorenz, M. T. Wolfinger, A. Tanzer, and I. L. Hofacker, "Predicting rna secondary structures from sequence and probing data," *Elsevier: Advances in rna Structure Determination*, vol. 103, pp. 86–98, 2016.
- [9] K. Tong, K. Cheung, K. Lee, and K. Leung, "Gaknot: Rna secondary structures prediction with pseudoknots using genetic algorithm," in *2013 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pp. 136–142, 2013.
- [10] Y. Zhao, Q. Ni, and Z. Wang, "Computational features evaluation for rna secondary structure predic-



tion," in *2009 2nd International Conference on Biomedical Engineering and Informatics*, pp. 1–5, Oct 2009.

[11] T. Akutsu, "Dynamic programming algorithms for rna secondary structure prediction with pseudoknots* 1,* 2," *Discrete Applied Mathematics*, vol. 104, pp. 45–62, 08 2000.

[12] S. Sheikh, R. Backofen, and Y. Ponty, "Impact of the energy model on the complexity of rna folding with pseudoknots," in *Combinatorial Pattern Matching*, pp. 321–333, Springer Berlin Heidelberg, 07 2012.

[13] R. Nussinov and A. B. Jacobson, "Fast algorithm for predicting the secondary structure of single-stranded rna" *Proceedings of the National Academy of Sciences of the United States of America*, vol. 77 11, pp. 6309–13, 1980.

[14] M. Zuker and P. Stiegler, "Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information," *Nucleic Acids Research*, vol. 9, pp. 133–148, 01 1981.

[15] E. Rivas and S. R. Eddy, "A dynamic programming algorithm for RNA structure prediction including pseudoknots," *Journal of Molecular Biology*, vol. 285, no. 5, pp. 2053–2068, 1999.

[16] A. Kravchenko, "Predicting RNA secondary structures including pseudoknots," 2009.

[17] A. E. Fatmi, A. Chentoufi, M. A. Bekri, S. Benhlma, and M. Sabbane, "A heuristic algorithm for rna secondary structure based on genetic algorithm," in *2017 Intelligent Systems and Computer Vision (ISCV)*, pp. 1–7, April 2017.

[18] A. El Fatmi, M. A. Bekri, and S. Benhlma, "Rnaknot: A new algorithm for RNA secondary struc-

ture prediction based on genetic algorithm and grasp method," *Journal of Bioinformatics and Computational Biology*, vol. 17, no. 05, p. 1950031, 2019.

[19] D. H. Turner and D. H. Mathews, "Nndb: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure," *Nucleic Acids Research*, vol. 38, pp. D280–D282, 2009.

[20] D. Huson, "Lecture notes Algorithms in Bioinformatics I," December 2006.

[21] M. Andronescu, V. Bereg, H. Hoos, and A. Condon, "RNA strand: the RNA secondary structure and statistical analysis database," *BMC bioinformatics*, vol. 9, p. 340, 09 2008.

[22] M. Taufer, A. Licon, R. Araiza, D. Mireles, F. H. D. van Batenburg, A. P. Gulyaev, and M.-Y. Leung, "Pseudobase++: an extension of pseudobase for easy searching, formatting and visualization of pseudoknots," *Nucleic acids research*, vol. 37, p. D127–35, January 2009.

[23] Y. Zhao, J. Wang, C. Zeng, and Y. Xiao, "Evaluation of rna secondary structure prediction for both base-pairing and topology," *Biophysics Reports*, vol. 4, 06 2018.

[24] P. Kerpedjiev, S. Hammer, and I. L. Hofacker, "Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams," *Bioinformatics*, vol. 31, pp. 3377–3379, 06 2015.

[25] K. Darty, A. Denise, and Y. Ponty, "VARNA: Interactive drawing and editing of the RNA secondary structure," *Bioinformatics*, vol. 25, pp. 1974–1975, 04 2009.

Tabla 7. Métricas de precisión por instancia y algoritmo.

CLAVE	ALGORITMO	VP	FN	FP	S	VPP	VALOR-F
PDB.00136	Nussinov	18	8	10	0.6428	0.6428	0.6666
	Zuker	18	8	10	0.6923	0.8571	0.7659
	Akutsu	19	7	9	0.7307	0.6785	0.7037
	RNAknot	18	8	1	0.6923	0.9473	0.80
PDB.01130	Nussinov	0	20	24	0.0	0.0	0.0
	Zuker	20	0	0	1.0	1.0	1.0
	Akutsu	16	4	10	0.80	0.6153	0.6956
	RNAknot	0	20	19	0.0	0.0	0.0
PDB.00481	Nussinov	0	26	24	0.0	0.0	0.0
	Zuker	14	12	0	0.5384	1.0	0.70
	Akutsu	0	26	22	0.0	0.0	0.0
	RNAknot	7	19	7	0.2692	0.50	0.35

CLAVE	ALGORITMO	VP	FN	FP	S	VPP	VALOR-F
PDB.01115	Nussinov	9	17	18	0.3461	0.3333	0.3396
	Zuker	20	6	1	0.7692	0.9523	0.8510
	Akutsu	6	20	22	0.2307	0.2142	0.2222
	RNAknot	10	16	10	0.3846	0.50	0.4347
PDB.01113	Nussinov	8	19	19	0.2962	0.2962	0.2962
	Zuker	18	9	1	0.6666	0.9473	0.7826
	Akutsu	6	21	22	0.2222	0.2142	0.2181
	RNAknot	6	21	11	0.2222	0.3529	0.2727
PDB.01128	Nussinov	0	27	27	0.0	0.0	0.0
	Zuker	12	15	6	0.4444	0.6666	0.5333
	Akutsu	14	13	19	0.5185	0.4242	0.4666
	RNAknot	8	19	9	0.2962	0.4705	0.3636
PDB.01125	Nussinov	18	18	18	0.50	0.50	0.50
	Zuker	33	3	0	0.9166	1.0	0.9565
	Akutsu	27	9	6	0.75	0.8181	0.7826
	RNAknot	24	12	0	0.6666	1.0	0.80
PDB.00213	Nussinov	9	32	30	0.2195	0.2307	0.2250
	Zuker	30	11	3	0.7317	0.9090	0.8108
	Akutsu	32	9	9	0.7804	0.7804	0.7804
	RNAknot	23	18	9	0.5609	0.7187	0.6301
PDB.00542	Nussinov	8	42	40	0.16	0.1666	0.1632
	Zuker	32	18	8	0.64	0.80	0.7111
	Akutsu	19	31	30	0.38	0.3877	0.3838
	RNAknot	19	31	14	0.38	0.5757	0.4578
PDB.01236	Nussinov	26	28	32	0.4814	0.4482	0.4642
	Zuker	37	17	14	0.6851	0.7254	0.7047
	Akutsu	29	25	30	0.5370	0.4915	0.5132
	RNAknot	24	30	18	0.4444	0.5714	0.50
PKB5	Nussinov	0	8	13	0.0	0.0	0.0
	Zuker	5	3	6	0.6250	0.4545	0.5263
	Akutsu	8	0	6	1.0	0.5714	0.7272
	RNAknot	4	4	4	0.50	0.50	0.50
PKB73	Nussinov	7	5	4	0.5833	0.6363	0.6086
	Zuker	7	5	2	0.5833	0.7777	0.6666
	Akutsu	12	0	0	1.0	1.0	1.0
	RNAknot	11	1	0	0.9166	1.0	0.9565



PKB74	Nussinov	5	6	5	0.4545	0.50	0.4761
	Zuker	7	4	0	0.6363	1.0	0.7777
	Akutsu	11	0	0	1.0	1.0	1.0
	RNAknot	8	3	0	0.7272	1.0	0.8421
PKB80	Nussinov	0	12	15	0.0	0.0	0.0
	Zuker	0	12	10	0.0	0.0	0.0
	Akutsu	12	0	5	1.0	0.7058	0.8275
	RNAknot	12	0	0	1.0	1.0	1.0
PKB81	Nussinov	2	6	6	0.25	0.25	0.25
	Zuker	5	3	0	0.6250	1.0	0.7692
	Akutsu	8	0	0	1.0	1.0	1.0
	RNAknot	6	2	0	0.75	1.0	0.8571
PKB107	Nussinov	0	12	17	0.0	0.0	0.0
	Zuker	0	12	11	0.0	0.0	0.0
	Akutsu	12	0	5	1.0	0.7058	0.8275
	RNAknot	12	0	0	1.0	1.0	1.0
PKB243	Nussinov	0	31	45	0.0	0.0	0.0
	Zuker	20	11	7	0.6451	0.7407	0.6896
	Akutsu	22	9	19	0.7096	0.5365	0.6111
	RNAknot	11	20	16	0.3548	0.4074	0.3793
PKB273	Nussinov	5	6	11	0.4545	0.3225	0.3703
	Zuker	6	5	6	0.5454	0.50	0.5217
	Akutsu	10	1	6	0.9090	0.6250	0.7407
	RNAknot	6	5	3	0.5454	0.6666	0.60
PKB280	Nussinov	6	8	17	0.4285	0.2608	0.3243
	Zuker	7	7	11	0.50	0.3888	0.4375
	Akutsu	12	2	13	0.8571	0.48	0.6153
	RNAknot	13	1	0	0.9285	1.0	0.9629
PKB281	Nussinov	3	5	9	0.3750	0.50	0.4385
	Zuker	3	5	3	0.3750	0.50	0.4285
	Akutsu	0	8	10	0.0	0.0	0.0
	RNAknot	8	0	1	1.0	0.8888	0.9411

