

## Métricas de enriquecimiento para la identificación de estabilizadores del cuarteto G telomérico usando algoritmo genético

Correa Melissa<sup>1</sup>., Solórzano Santiago<sup>2</sup>

melissa.correa.osorio@gmail.com; solorzano.lescanao@gmail.com

<https://orcid.org/0000-0002-1022-2833><sup>1</sup>, <https://orcid.org/0000-0002-8154-5277><sup>2</sup>

<sup>1</sup>Universidad de las Américas, Facultad de Ingeniería y

Ciencias Aplicadas, Ingeniería en Biotecnología, Quito-Ecuador

<sup>2</sup>Consultor para el PNUD proyecto de las Naciones Unidas para el desarrollo  
Quito - Ecuador

Recibido (06/03/20), Aceptado (26/03/20)

**Resumen:** En este estudio se encuentra detallada una combinación de herramientas computacionales de acoplamiento y cribado virtual, en 108 moléculas activas y 3620 señuelos para encontrar estabilizadores del cuarteto G (G4). Para tener resultados más precisos se aplicaron combinaciones de programas de acoplamiento con quince funciones de puntuación energética. La validación y evaluación de las métricas se realizó con el algoritmo genético CompScore. Los resultados evidenciaron un aumento en BEDROC y EF del 50% en comparación a otras estrategias, además de reflejar un reconocimiento temprano de moléculas activas. A partir de estos resultados es posible trabajar con las moléculas que presentaron un buen reconocimiento temprano y evaluar su efecto como estabilizadores de G4. De esta manera se garantiza resultados más eficientes y precisos en la etapa preclínica para el desarrollo de anticancerígenos.

**Palabras Clave:** Métricas de enriquecimiento; telomero; cuarteto G (G4); CompScore.

## Enrichment metrics for the identification of stabilizers of the telomeric G quartet using genetic algorithm

**Abstract:** In this study a combination of computer tools for coupling and virtual screening is detailed, in 108 active molecules and 3620 decoys to find stabilizers for G quadruplex (G4). To have more precise results, combinations of coupling programs with fifteen energy scoring functions were applied. The validation and evaluation of the metrics was done with the CompScore genetic algorithm. The results showed an increase in BEDROC and EF of 50% compared to other strategies, as well as reflecting early recognition of active molecules. From these results it is possible to work with the molecules that showed a good early recognition and evaluate their effect as G4 stabilizers. This ensures more efficient and accurate results in the preclinical stage for the development of anticancer drugs.

**Keywords:** Enrichment metrics; telomere; G quadruplex (G4); CompScore.



## I. INTRODUCCIÓN

Los telómeros forman una parte fundamental para el desarrollo de la vida y el envejecimiento [1-3]. Existe evidencia que indica que en ciertos casos los telómeros son disfuncionales, como resultado de su acortamiento crítico en las células somáticas normales, lo cual provoca una respuesta al daño a nivel de ácido desoxirribonucleico ADN [4]. En las células tumorales, la telomerasa confiere la inmortalidad celular y también regula la generación de ácido ribonucleico mensajero ARNm. Estas características reflejan que la telomerasa relacionada con la aparición de tumores malignos y su difusión en diferentes órganos [5]. Algunos estudios sugieren que la estabilización del cuarteto G telomérico (G4), puede inhibir directamente la acción de la telomerasa mediante el bloqueo de la unión de la telomerasa al telómero [6]. Una gran variedad de estructuras G4 se han dilucidado a nivel atómico, generando una nueva vía para el diseño de ligandos dirigidos hacia la estructura del G4 de forma computacional [7]. Los últimos estudios realizados sobre estabilizadores de G4 mencionan que la interacción entre G4 y el ligando ejerce efectos antiproliferativos a través de mecanismos dependientes e independientes de los telómeros. No obstante, hasta el momento encontrar compuestos que estabilicen G4 ha resultado en un gran desafío. A pesar de que existen muchos datos disponibles sobre G4 y sus ligandos, solo un número limitado, se ha explorado para la detección virtual o en la detección de nuevas moléculas bioactivas. Aunque en la actualidad se usen múltiples técnicas para el descubrimiento de fármacos asistido por herramientas virtuales, existe una clara falta de comprensión en el mecanismo molecular primario que se encuentra tras la actividad de los aciertos que se consigue empleando estas técnicas [9]. Por otra parte, para la validación de los resultados requiere la combinación de distintos métodos o herramientas para superar las falencias que presenta un único protocolo, lo cual puede hacer que el proceso se considere más complejo [10].

A partir de esto, se ha planteado desarrollar una metodología de cribado virtual basada en estructura y como estrategia para la validación de datos se usará un algoritmo genético. La estructura de este trabajo inicia con los aspectos teóricos del tema en el apartado III, seguido de la metodología y programas aplicados en la sección III, a continuación se detallan los resultados obtenidos y se tiene como punto final las conclusiones.

## II. DESARROLLO

### A) Características de los telómeros y la telomerasa

En los mamíferos los extremos de los cromosomas se encuentran cubiertos por estructuras protectoras de ADN y proteínas de unión. A este complejo se le denomina telómero [11]. Los telómeros son estructuras conformadas de nucleoproteínas situadas en el extremo de cada brazo cromosómico. Cumple como principal función el mantenimiento de la estabilidad del genoma, al garantizar que los extremos naturales de los cromosomas no se confundan con sitios de daño en el ADN [12]. Además de presentar un papel fundamental frente a la señalización del daño del ADN, también evitan su reparación errónea. Esta función de activar la respuesta al daño e inhibir su reparación, podría tener una consecuencia importante en el proceso de envejecimiento celular dependiente de la longitud telomérica y la supresión tumoral [13].

La telomerasa es un heterodímero de transcriptasa inversa, formada por una subunidad enzimática y una catalítica de molde de ARN no codificante, para la síntesis de novo de secuencias de ADN no telomérico. Los componentes principales de la telomerasa son la transcriptasa inversa de la telomerasa (TERT) y el componente de ARN de la telomerasa (TERC). El molde de ARN, permite que TERT agregue repeticiones al terminal cromosómico. TERC se caracteriza por expresarse de forma constante, a diferencia de TERT que permanece inactiva en las células somáticas, generando el acortamiento constante en los telómeros [13]. La represión de la telomerasa en las células somáticas y el mecanismo de barrera de proliferación de los telómeros, generan una vía supresora que limita el crecimiento de las células tumorales [14]. Según resultados estadísticos, el 85% de células cancerígenas usan la telomerasa para mantener la longitud del telómero. En tumores que presentan un estado avanzado o metastásico, la actividad de esta ribonucleoproteína se ve incrementada. La telomerasa confiere inmortalidad a las células cancerígenas, adicionando repeticiones en tándem al extremo 3' del telómero, dando como consecuencia un mantenimiento o prolongación sobre la longitud del telómero [15].

### B) Definición y características del G4 telomérico

Las estructuras de ácido nucleico helicoidal de cuatro cadenas ricas en guanina reciben el nombre de G- cuádruplex o G4 [16]. Esta estructura es definida como una forma alternativa de ADN o ARN compuesta por varias capas

planas de guaninas denominadas tétradas, se forman por lo general por secuencias de tipo 5'-G3+N1-7G3+N1-7G3+N1-7G3+-3', donde N puede representar cualquier base nitrogenada (purinas y pirimidinas). El G4 se forma al ensamblarse las tétradas de guanina, de forma apilada hasta adaptar un bucle con interacciones que se consideren energéticamente favorables [16]. Estas se unen por enlaces de hidrogeno de tipo Hoogsteen, en donde cada tétrada rota aproximadamente 90° respecto a la adyacente. Para conseguir un buen ensamblaje es necesaria la presencia de cationes, con el fin de que se unan entre cada par de tétradas, coordinadas por los grupos carbonilo. La presencia de los cationes suele ser coordinada por las secuencias que conforman el bucle que brinda la estructura a G4 [17]. Los bucles que conectan cada tétrada en la topología paralela son de doble cadena conformados por pocos nucleótidos, mientras que en las formas antiparalela e híbrida su longitud se extiende con un mayor número de nucleótidos [18].

### C) Detección virtual de ligandos para G4

Hasta el momento existen más de 1.000 ligandos caracterizados, que ejercen efecto estabilizador sobre G4. Su descubrimiento ha partido de herramientas computacionales como cribado virtual (VS) y cribado tradicional de alto rendimiento (HTS) [17-18]. Estos ligandos se pueden dividir en dos clases químicas: compuestos heteroaromáticos planos y ligandos cíclicos o alicíclicos. Algunas de estas moléculas han sido cristalizadas con G4 y han sido usados para generar nuevos modelos de investigación [19]. Los métodos computacionales para encontrar nuevos estabilizadores son varios. Entre los más usados está el farmacóforo, un modelado basado en características específicas para garantizar las interacciones con un objetivo biológico definido y desencadenar su respuesta biológica. En conclusión este método se basa en unión del ligando a su objetivo [18]. Otra estrategia consiste en el modelamiento basado en formas. Se base en la estructura o geometría del compuesto para que este sea complementario al estar activo. Cuando la estructura de la molécula esta activa sirve como plantilla o molde para la generación de un modelo de forma. Para obtener mejores resultados usando esta estrategia se sugiere que se adicione información química. La característica común de estas dos estrategias es que filtran una gran cantidad de compuestos de bases de datos, pues es uno de los requisitos de ambos modelos. Por otra parte, las estrategias de acoplamiento o docking requieren específicamente datos estructurales del target. La eficiencia del método se estima mediante el cálculo de energía libre [15]. La combinación de múltiples métodos mejora el enriquecimiento de las moléculas activas en los resultados virtuales, además, ayuda a solucionar el problema de selectividad que suelen tener este tipo interacciones [14].

### D) Métricas de enriquecimiento

Para el rendimiento del modelo se debe tomar en cuenta: área bajo la curva acumulativa (AUAC), Factor de enriquecimiento (EF) y discriminación de ROC mejorada por Boltzmann (BEDROC) [14].

$$UAC = 1 - \frac{1}{nN} \sum_{i=1}^n x_i \quad (1)$$

La ecuación 1, permite determinar la curva AUAC, en donde N es el número total de compuestos, n la cantidad de compuestos activos y  $X_i$  el ranking relativo del compuesto i.

La métrica EF calcula el número de veces que se enriquece una fracción con compuestos activos en relación con lo que se espera de una distribución uniforme de activos.

$$EF = \frac{\sum_{i=1}^n \delta_i}{Xn} \quad \text{cuando } \delta_i = \begin{cases} 1, & \text{si } r_i \leq xn \\ 0, & \text{si } r_i > xn \end{cases} \quad (2)$$

En la ecuación 2 se muestra la métrica EF, donde  $r_i$  es la clasificación relativa de  $X_i = r_i/N$  en toda la lista clasificada; X es la fracción de la lista ordenada ponderada entre 0 a 1.

$$BEDROC = \frac{RIE - RIE_{min}}{RIE_{max} - RIE_{min}} \quad (3)$$

La ecuación 3 muestra el cálculo para BEDROC considerando un valor de  $\alpha=160,9$ . Mientras que las ecuaciones 4 refleja cómo obtener BEDROC

$$RIE = \frac{\frac{1}{n} \sum_{i=1}^n e^{\alpha X_i}}{\frac{1}{N} \left( \frac{\alpha^{1-e^\alpha}}{e^N} \right)} \quad RIE_{min} = \frac{1 - e^{\alpha R_\alpha}}{R_\alpha (1 - e^\alpha)}$$

$$RIE_{max} = \frac{1 - e^{\alpha R_\alpha}}{R_\alpha (1 - e^{-\alpha})} \quad (4)$$

### III.METODOLOGÍA

Como punto inicial se buscó del banco de datos de proteínas RCSB PDB las estructuras de tipo paralela, anti paralela e híbrida del G4 telomérico [15]. A partir de la estructura se eligió como receptor la molécula con la molécula de acridina 3, 6,9 –trisustituida BRACO19 (3CE5). Para su preparación se usó de los paquetes de OEDocking el Software MakeReceptor [16]. En el proceso de preparación se eliminaron las moléculas de agua y el ligando. Además se agregaron átomos de hidrogeno y cargas. Para la Preparación del ligando Para generar las conformaciones 3D se empleó el software OMEGA. Se aprovecharon las herramientas MakeFraglib para la generación de fragmentos, Omega 2 para construcción 3D de los fragmentos guardando un total de 200 conformaciones [17]. Mientras que con el Software Quacpac se aplicó molcharge para adicionaron las cargas a las estructuras 3D [18]. Luego de obtenerse el ligando con un máximo de 200 conformaciones 3D mas las cargas y haber preparado el receptor, se realizó el acoplamiento con el software FRED [19]. Se puntuaron con Chemgauss4 500 moléculas para evaluar la calidad de las poses. Se guardaron aquellas con la puntuación más baja para cada conformación de ligando, dando un máximo de 20 poses. Para encontrar estabilizadores de G4 Se usó la base de datos G4LDB [20] y se buscaron compuestos en base a la estabilidad de FRET mayor a 10 grados y menor a 10 grados. De la información que se recopiló se filtró considerando la secuencia del telómero humano y las concentraciones experimentales de los compuestos, máxima de 20  $\mu$ M. Para generar moléculas señuelo Se usó la herramienta A Database of useful Decoys: Enhanced (DUD-E) [21]. En la herramienta se ingresó el código SMILES y luego de una hora se obtuvieron los resultados. Del archivo recibido se extrajo las líneas que contenían los ligandos y se agrupó en un solo documento. Los señuelos fueron separados en grupos se cien y se dejó solo con su código SMILES y su código de identificación.

#### A) Acoplamiento para las moléculas señuelo

Preparación del ligando y receptor: Para la preparación del ligando se usó una metodología similar a la aplicada anteriormente, usando los códigos SMILES de cada molécula señuelo y de los ligandos. La única diferencia que se efectuó tras generar los fragmentos, fue que inicialmente se crearon 200 conformaciones y en un siguiente resultado se corrió el mismo programa, pero generando una única conformación. La adición de cargas solo se efectuó sobre los resultados que generaron una única conformación. De igual modo se mantuvo el receptor preparado previamente. Para el acoplamiento se usó el software FRED [19]. Partiendo de las 200 conformaciones generadas. Únicamente se guardó una pose. También se efectuó el acoplamiento usando una alta resolución. A los resultados obtenidos se le adicionaron cargas usando del Software Quacpac molcharge [18].

#### B) Re-scoring

Preparación del receptor, esferas y caja: Se usó el software Dock 6.9 [22] y se mantuvieron los parámetros pre-establecidos por el programa para la obtención del receptor, las esferas y la caja. Para realizar la puntuación se usaron las herramientas de DOCK 6.8, OEDOCKING y GOLD. Los parámetros de cada programa se mantuvieron en sus valores por defecto. Además, se calcularon quince funciones de puntuación (tabla I) para las poses obtenidas durante el acoplamiento molecular [23].

TABLA I. Funciones de puntuación.

Docking Program	Scoring Function
Dock 6.8	Grid
	Contact
	Continuous
	Hawkins
	PBSA
Gold	SASA
	PLP
	GoldScore
	ChemScore
OEdocking	ASP
	Shapegauss
	ChemScore
	ChemGauss 3
	ChemGauss 4
	PLP

### C) Modelos de cribado virtual

Los modelos de cribado virtual se basaron en la metodología CompScore, la cual se basa en combinar los componentes de distintas funciones de puntuación de acoplamiento que maximiza las métricas de detección virtual [23]. Los valores de los componentes de las funciones de puntuación energética fueron resumidos en una tabla empleando los códigos disponibles con CompScore. A continuación se separaron los ligandos de las moléculas señuelo. Se tomó el 80% de cada grupo y se formó un archivo para el entrenamiento de los modelos. El 20% restante de los datos se reservó para la validación del modelo obtenido. Durante el entrenamiento de los modelos se mantuvieron los parámetros preestablecidos para el método CompScore. El rendimiento del modelo se evaluó usando varias métricas: área bajo la curva acumulativa (AUAC), Factor de enriquecimiento (EF) y discriminación de ROC mejorada por Boltzmann (BEDROC) [23].

## IV.RESULTADOS

Se usó la estructura paralela con el código 1kfl. De la base de datos G4LDB se obtuvo un total de 291 estabilizadores iniciales con una actividad de FRET mayor a 10 grados, mientras que se obtuvieron 19 ligandos con actividad menor a 10 grados. Mientras que del programa A Database of useful Decoys, se obtuvo un total de 3728 moléculas de las cuales 3620 eran señuelo y los demás ligandos. Debido a que existían 3 pares de duplicados y 2 compuestos organometálicos. Se generaron 37 carpetas de señuelos y una de ligandos, además se agregó las moléculas que tenían actividad de FRET menor a diez grados. Normalmente se esperaría tener un mayor número de compuestos, sin embargo, en este caso algunos de los ligandos eran muy similares entre sí y por lo tanto comparten señuelos.

Los receptores, ligandos y señuelos para los cálculos de acoplamiento siguieron la metodología descrita anteriormente. A partir de los resultados para cada grupo tanto de señuelos como de ligandos se sometieron a diferentes estrategias de acoplamiento y puntuación. El usar el algoritmo de CompScore se proporciona un valor de enriquecimiento máximo, ya que combina 15 funciones de puntuación lo cual se considera una búsqueda exhaustiva [24].

Los resultados que se obtienen al usar el algoritmo CompScore en comparación con otras funciones de puntuación y sus componentes de puntuación resulta ser superior. El algoritmo genético (GA) ha proporcionado un aumento máximo de enriquecimiento de los resultados, de hasta un 45%. Los datos usados para entrenamiento, que corresponden al 80% de ligandos y 80% de señuelos. Reflejaron métricas más altas usando la metodología CompScore. Se obtuvo un valor de EF de 52,02. Entre las 15 funciones individuales de puntuación, el programa Gold con la función PLP presentó el valor más alto de EF = 20,43. Por otra parte entre los componentes de puntuación, el programa CompScore sugiere 15 componentes de puntuación entre las funciones incluidas en el modelo. De estos, se obtuvo un valor de EF= 18,58 para el componente OEPIPNonPola. Evidentemente estos valores resultan inferiores si se comparan con el algoritmo CompScore (figura 1).

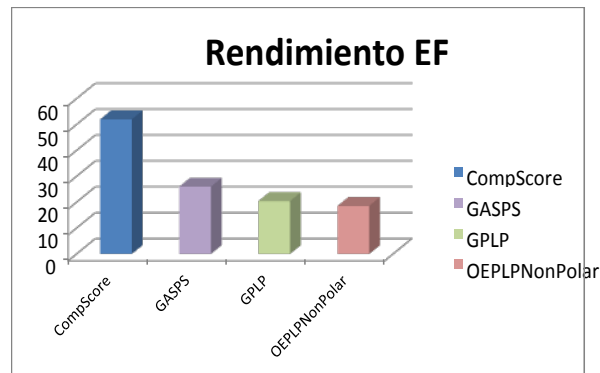


Fig 1. Mejores resultados obtenidos para EF evaluando el 1% de los datos.

En la figura 1 es Evidente que el mejor resultado se da usando el algoritmo CompsCore. Seguido del valor de GASPS el cual es reconocido por el algoritmo como la mejor puntuación individual en el 1% de los datos. El resultado de la mejor función puntuación y el mejor componente sugerido por el algoritmo se encuentran a continuación con resultados no muy altos en comparación con el resultado de CompScore. Está claro que el algoritmo efectivamente combino los componentes de puntuación para maximizar la métrica de enriquecimiento. Por otra parte, los datos de entrenamiento también maximizan el enriquecimiento de los activos usando

BEDROC con un valor de  $\alpha=160.9$ . En este caso CompScore superó significativamente a las funciones y componentes de puntuación arrojando un valor de BEDROC= 0,875. Siendo este valor más del doble del valor más alto obtenido entre los 15 componentes de puntuación comparados, cuyo valor fue de

BEDROC= 0,410. Aunque la función GPLP y los componentes GASPS y OEPLPNonpolar presenta valores altos en comparación a las demás (tabla II). No es relevante si se compara con los resultados de CompScore (figura 2).

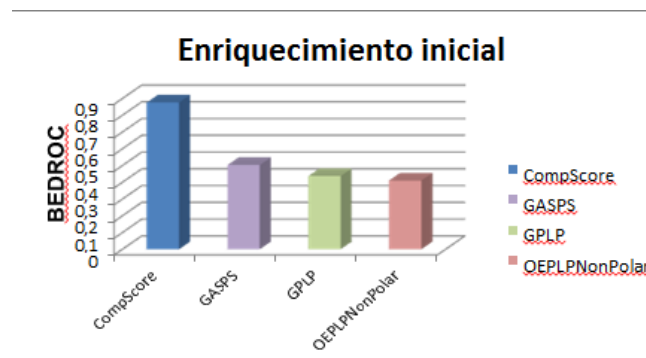


Fig 2. Resultado de la métrica BEDROC.

En la figura 2 CompScore es claramente el mejor resultado. Mientras los otros componentes y funciones presentan valores similares entre sí, pero poco relevantes comparados con el resultado de CompScore. La similitud entre los resultados, que presentan las funciones y componentes de puntuación, sugiere que el componente GASPS puede lograr un rendimiento similar a la mejor función de puntuación completa GPLP. Este tipo de resultados es explicado por la metodología de CompScore. Según su hipótesis los componentes de las funciones de puntuación son de suma importancia e incluso pueden ser más relevantes que las funciones de puntuación completas.

TABLA II. Resultados obtenidos de los componentes de las funciones individuales de puntuación sugeridas por el algoritmo.

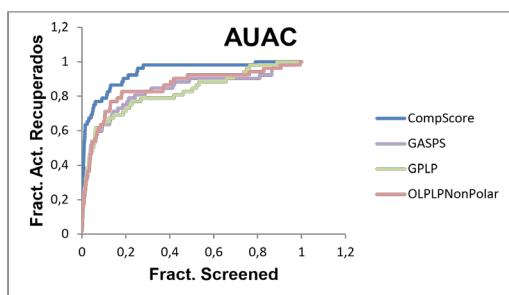
Componentes	AUAC	EF	BEDROC
D6HawkinsGB	0,583	0	1,165E-05
D6PbsaPB	0,529	7,432	1,17E-01
D6PbsaSA	0,754	7,432	0,147
D6SASAComOther	0,549	7,432	0,084
D6SASAOther	0,563	7,432	0,091
D6SASAPercOtherExpo	0,549	7,432	0,084
D6SASAPhilic	0,691	1,858	0,042
GASPASP	0,466	0	0,004
GCSLipo	0,462	0	5,205E-07
GPLPDECLash	0,634	1,858	0,042
GPLPDETors	0,734	0	0,017
OECCG3Don	0,427	1,858	0,040
OECCG4ProDesolv	0,503	0	1,135E-06
OEPLPNonPolar	0,850	18,580	0,410
OEShapegauss	0,811	14,864	0,314
GASPS	0,832	26,013	0,502

Para las métricas EF y BEDROC el programa Gold presentó los componentes de puntuación con valores más altos pero también contiene el componente con menores valores, tanto para EF (0) como para BEDROC (0,004). Por lo tanto no se considera a los componentes y funciones de puntuación de forma individual como estrategias significativas frente a la metodología de CompScore (tabla III).

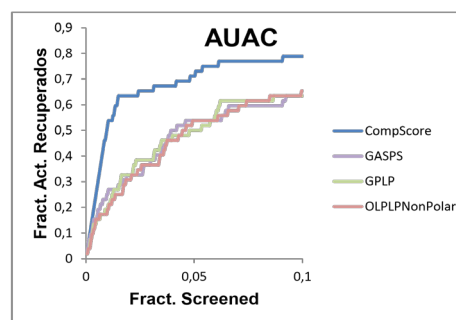
TABLA III. Resultados de las 15 funciones de puntuación analizadas por el algoritmo CompScore

Funciones	AUAC	EF	BEDROC
DContact	0,744	14,864	0,315
DContinuous	0,512	0	0,005
DGrid	0,498	0	0,0006
DHawkins	0,516	1,858	0,024
DPBSA	0,513	0	0,004
DSASA	0,316	0	0,0001
GASP	0,546	1,858	0,034
GChemScore	0,665	11,148	0,235
GGoldScore	0,581	1,858	0,021
GPLP	0,821	20,438	0,436
OEChemGauss 3	0,825	9,290	0,166
OEChemGauss 4	0,772	11,148	0,171
OEChemScore	0,616	5,574	0,100
OEPLP	0,817	14,864	0,332
OEShapegauss	0,811	14,864	0,314

Los resultados del AUAC no fueron tan altos en los componentes de puntuación analizados (tabla 2) ni en las funciones de puntuación evaluadas (tabla 3). Sin embargo los valores de GASPS 0,823 GPLP 0,821 y OEPLP-NonPolar 0,850 si presentaron resultados significativos, muy cercanos al resultado obtenido por la metodología del GA que presentó un valor de 0,937 (figura 3).



(A) Curva AUAC del 100% de datos



(B) curva AUAC del 1% de los datos

Fig 3.a Curvas completas AUAC de la metodología CompScore, de las funciones y componentes de puntuación. B. curvas para el 10% de los datos

La validación externa se realizó siguiendo la metodología antes descrita con el 20% de ligandos y 20% de señuelos. De los resultados obtenidos anteriormente del algoritmo CompScore se efectuó la recalificación de los datos externos y se midió el enriquecimiento sobre los datos restaurados. Es por tal motivo que únicamente se analizaron las 15 funciones completas de puntuación y el algoritmo CompScore.

Para los datos de validación la metodología de CompScore presentó un resultado para la métrica EF= 40,44 un valor muy alto en comparación con el mejor de los resultados de las funciones de puntuación completa GPL que obtuvo un valor EF= 20,22 (figura 4). Considerando que este mismo valor se obtuvo con la función completa GASPS, se podría creer que ambas funciones de forma individual presentan un buen enriquecimiento. Sin embargo el resultado para la métrica de BEDROC y el AUAC es inferior (tabla 4). El único fenómeno que se podría considerar relevante es que la función completa GPLP del programa Gold, presenta resultados relevantes tanto para los datos de entrenamiento como de validación. En promedio las funciones de los programas Gold y OpenEye presentan un valor de EF=13,480 en ambos casos, mientras que las funciones del programa Dock 6.8 fueron inferiores, con un promedio EF=5,616. Por tanto, sería más recomendable usar estos programas para calcular la métrica EF, ya que presentan valores más altos de enriquecimiento para el 1% de los datos.

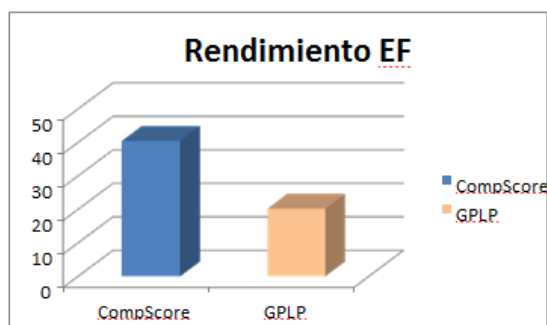


Fig 4. Resultados de EF usando la metodología CompScore en comparación con el mejor resultado de las funciones completa de puntuación GPLP.

Para los resultados de BEDROC el programa Gold presentó entre sus funciones completas en promedio un valor de 0,762 siendo este el valor más alto, seguido de las funciones de OpenEye que presentaron un valor de BEDROC=0,231 y DOCK6.8 con el promedio más bajo 0,105 (tabla 4). Entre estos resultados, el puntaje más alto fue para la función PLP del programa Gold, con un valor de 0,386. Sin embargo, este valor es bajo si se compara con el de la metodología CompScore donde se obtuvo un valor de BEDROC=0,848 (Figura 5).



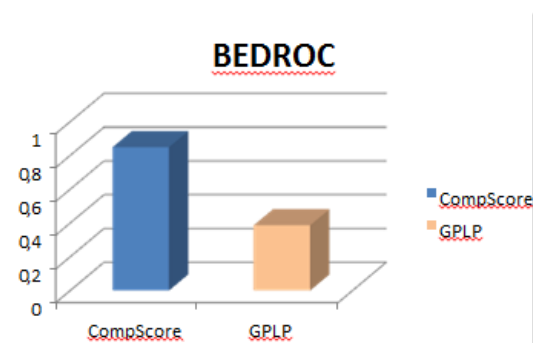


Fig 5. Resultado de BEDROC en la validación externa.

Finalmente el resultado AUAC para CompScore fue de 0,950 y para GPLP fue de 0,836 (Figura 6). En este como en los otros resultados analizados la metodología CompScore fue la mejor estrategia para el enriquecimiento de las métricas.

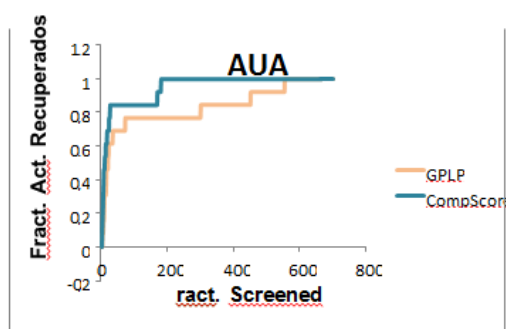


Fig 6. AUAC para los resultados de validación

Tabla IV. Resultados obtenidos de las funciones individuales de puntuación para los resultados de validación externa.

Funciones	AUAC	EF	BEDROC
GASP	0,812	20,221	0,234
GChemScore	0,715	13,480	0,207
GGoldScore	0,683	0	0,011
GPLP	0,835	20,221	0,386
DGrid	0,516	0	0,021
DContact	0,845	13,480	0,231
DContinuous	0,538	6,740	0,054
DHawkins	0,541	6,740	0,108
DPBSA	0,534	6,740	0,216
DSASA	0,181	0	3,261E-26
OEShapegauss	0,836	13,480	0,180
OEChemScore	0,682	13,480	0,243
OEChemGauss 3	0,787	13,480	0,209
OEChemGauss 4	0,709	13,480	0,194
OEPLP	0,847	13,480	0,324

Por otra parte, aunque la metodología CompScore proporciona un alto enriquecimiento para las métricas EF y BEDROC, no se evidencia una distinción clara entre compuestos con baja o nula estabilización. Esto se comprobó al buscar la posición en la que el algoritmo situó los compuestos de FRET menor a 10 grados. Muchos de ellos se encontraban en posiciones intermedias, mientras dos de ellas se encontraron entre los primeros 100 datos.

Es recomendable que la métrica EF sea enriquecida cuando se quiere evaluar la clasificación de los métodos en el cribado virtual. Dado que su principal propósito es clasificar al principio de una lista ordenada los compuestos activos (ligandos) en una fracción específica. Mientras que BEDROC cumple la función de representar el enriquecimiento temprano. Si se desea maximizar su valor es necesario proporcionar el parámetro  $\alpha$ .

## V.CONCLUSIONES

El uso de moléculas activas e inactivas permite un mejor resultado de acoplamiento para su posterior validación. El uso de las herramientas de CompScore a partir de su algoritmo genético genera un aumento significativo en las métricas de enriquecimiento. El valor EF de 40,44 en los datos de validación es muy superior al obtenido en las funciones de puntuación individuales, lo mismo ocurre en los datos de entrenamiento. Esta característica demuestra que el modelo tiene una calidad superior en comparación a otros. Se ve enriquecidas ambas métricas EF y BEDROC. De forma menos evidente aumenta el valor de AUAC. La capacidad de enriquecimiento de EF considerando el 1% de las moléculas activas es más acertada que considerar fracciones de 5% 10% o 20% ya que las funciones de puntuación generaron una lista ordenada de compuestos activos en orden ascendente. En conclusión la estrategia de VS mostro características relevantes de enriquecimiento y capacidades adecuadas para el reconocimiento temprano. Sin embargo también es evidente que el modelo no distingue entre compuestos con buena o mala estabilización.

## REFERENCIAS

- [1]M. Porru, P. Zizza, M. Franceschin, C. Leonetti y A. Biroccio. «EMICORON: A multi-targeting G4 ligand with a promising preclinical profile» 2017. *Biochimica et Biophysica Acta - General Subjects*, 1861(5), 1362–1370. [En línea]. Available: <https://doi.org/10.1007/s00294-018-0836-6>
- [2]K. Tomita. «How long does telomerase extend telomeres ? Regulation of telomerase release and telomere length homeostasis». *Current Genetics*, 64(6), 1177–1181. 2018. [En línea]. Available: <https://doi.org/10.1016/j.bbagen.2016.11.010>
- [3]M. Jafri, S. Ansari, M. Alqahtani y J. Shay. «Roles of telomeres and telomerase in cancer , and advances in telomerase- targeted therapies. *Genome Medicine.*, 2016. [En línea]. Available: <https://doi.org/10.3390/ijms19020482>
- [4]J. Huppert y S. Balasubramanian. «G-quadruplexes in promoters throughout the human genome». 35(2), 406–413. 2007. [En línea]. Available: <https://doi.org/10.1016/j.bbapap.2017.06.012>
- [5]S. Joy, Vijayakumar, Y. M., & Sunhye, G. (2015). «Role of computer-aided drug design in modern drug discovery». *Archives of Pharmacal Research*. 2015. [En línea]. Available: <https://doi.org/10.1007/s12272-015-0640-5>
- [6]S. Asamitsu, S. Obata, Z. Yu, T. Bando y H. Sugiyama. «Recent Progress of Targeted G-Quadruplex-Preferred Ligands Toward Cancer Therapy». *Molecules*, 24(3), 429. 2019. [En línea]. Available: <https://doi.org/10.3390/molecules24030429>
- [7]R. Monsen y J. Trent. «Biochimie G-quadruplex virtual drug screening : A review». *Biochimie*, 152, 134–148. 2018. [En línea]. Available: <https://doi.org/10.1039/c9cc06748e>
- [8]J. Beauvarlet, P. Bensadoun, E. Darbo, G. Labrunie, E. Richard, I. Draskovic, M. Djavaheri-mergny. «Modulation of the ATM / autophagy pathway by a G-quadruplex ligand tips the balance between senescence and apoptosis in cancer cells». 1–18. 2019. [En línea]. Available: <https://doi.org/10.1093/nar/gkz095>
- [9].Z. Crees, J. Girard, Z. Rios, G. Botting, K. Harrington y C. Shearrow. « Oligonucleotides and G-quadruplex stabilizers: targeting telomeres and telomerase in cancer therapy».2014. [En línea]. Available: <https://doi.org/10.2174/1381612820666140630100702>
- [10].M. Meier, A. Moya-torres, N. Krahn, M. Mcdougall, L. Orriss, E. Mcrae, T. Patel. «Structure and hydrodynamics of a DNA Gquadruplex with a cytosine bulge». (May), 1–13. 2018. [En línea]. Available: <https://doi.org/10.1093/nar/gky307>
- [11].M. Ivancich, Z. Schrank, L. Wojdyla, B. Leviskas y A. Kuckovic. «Treating Cancer by Targeting Telomeres and Telomerase». 2017. [En línea]. Available: <https://doi.org/10.3390/antiox6010015>
- [12].J. Maciejowski y T. Lange. De. «Telomeres in cancer : tumour». Nature Publishing Group. 2017. [En línea].

Available: <https://doi.org/10.1038/nrm.2016.171>

[13].M. Zurek, J. Altschmied, S. Kohlgrüber, N. Ale-agma y J. Haendeler. «Role of Telomerase in the Cardiovascular System». (Cvd), 1–11. 2016. [En línea]. Available: <https://doi.org/10.3390/genes7060029>

[14].G. Zhou, X. Liu, Y. Li, S. Xu y C. Ma. «Telomere targeting with a novel G-quadruplex-interactive ligand BRA-CO-19 induces T-loop disassembly and telomerase displacement in human glioblastoma cells». 7(12). 2016. [En línea]. Available: <https://doi.org/10.18632/oncotarget.7808>

[15]H. Berman. (2000). The Protein Data Bank. Nucleic Acids Research, 28(1), 235–242. [En línea]. Available: <https://doi.org/10.1093/nar/28.1.235>.

[16]Cheminformatics Software, Molecular Modeling Software | OpenEye Scientific. Retrieved June 11, 2019, [En línea]. Available:<https://www.eyesopen.com/>.

[17]OMEGA 3.1.1.2: OpenEye Scientific Software, Santa Fe, NM. [En línea]. Available: <http://www.eyesopen.com>

[18]QUACPAC 2.0.1.2: OpenEye Scientific Software, Santa Fe, NM. [En línea]. Available: <http://www.eyesopen.com>

[19]OEDOCKING 3.3.1.2: Software científico OpenEye, Santa Fe, NM. [En línea]. Available: <http://www.eyesopen.com>

[20]Q. Li, J. Xiang, Q. Yang, H. Sun, A. Guan y Y. Tang. (2013). G4LDB: a database for discovering and studying G-quadruplex ligands. Nucleic Acids Research, 41(Database issue), D1115–23. [En línea]. Available: <https://doi.org/10.1093/nar/gks1101>

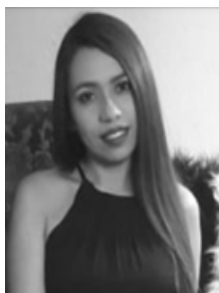
[21]M. Mysinger, M. Carchia, J. Irwin y B. Shoichet. (2012). Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. Journal of Medicinal Chemistry, 55(14), 6582–6594. [En línea]. Available:<https://doi.org/10.1021/jm300687e>

[22]P. Lang, S. Brozell, S. Mukherjee, E. Pettersen, E. Meng, V. Thomas e I. Kuntz. (2009). DOCK 6 : Combining techniques to model RNA – small molecule complexes. 1219–1230. [En línea]. Available: <https://doi.org/10.1261/rna.1563609.HIV>

[23]J. Truchon y C.Bayly. (2007). Evaluating Virtual Screening Methods : Good and Bad Metrics for the “ Early Recognition ” Problem. 488–508. [En línea]. Available: <https://doi.org/10.1021/ci600426e>

[24]D. Castillo-González, J. Mergny, A. De Rache, G. Pérez-Machado, M. Cabrera- Pérez, O. Nicolotti, M. Cruz-Monteaugudo. (2015). Harmonization of QSAR Best Practices and Molecular Docking Provides an Efficient Virtual Screening Tool for Discovering New G-Quadruplex Ligands. Journal of Chemical Information and Modeling, 55(10), 2094–2110. <https://doi.org/10.1021/acs.jcim.5b00415>

## RESUMEN CURRICULAR



**Melissa Correa**, Egresada de Ingeniería en Biotecnología en la Universidad de las Américas en Quito-Ecuador. Con participación destacada en el primer foro de CEBIO, congreso internacional de Biotecnología UDLA y final continental Solacyt 2020 en México. Desarrollo de proyectos en Biorreactores de inmersión temporal y fotobiorreactores presentados a nivel internacional.



**Santiago Solórzano** Máster en Inteligencia Artificial cuenta con una Ingeniería en Redes y Telecomunicaciones, Tecnología en Electrónica Instrumentación & Aviónica; consultor para el PNUD. Su formación le ha permitido manejar y desarrollar proyectos con tecnología de punta impulsando soluciones de múltiples ingenierías, en el ámbito de la Robótica, IA, IoT, WSN, Electrónica.