

<i>Nereis. Revista Iberoamericana Interdisciplinar de Métodos, Modelización y Simulación</i>	13	25-39	Universidad Católica de Valencia San Vicente Mártir	Valencia (España)	ISSN 1888-8550
--	----	-------	---	-------------------	----------------

Aplicación de la topología molecular para la predicción de nuevos compuestos con actividad citotóxica antitripanosomal

Application of molecular topology in the prediction of new compounds with anti-trypanosomal cytotoxic activity

Fecha de recepción y aceptación: 17 de marzo de 2021 y 29 de abril de 2021

DOI: 10.46583/nereis_2021.13.823

C. T. Cortés Rodríguez¹, A. C. Onos Pérez¹, A. Moreno Murillo¹, C. Bonillo López¹, J. Gálvez¹⁻² y R. García-Domenech^{1-2*}

¹ Máster en Enfermedades Parasitarias Tropicales. Departamento de Farmacia y Tecnología Farmacéutica y Parasitología. Universitat de València.

² Departamento de Química Física, Facultad de Farmacia. Universitat de València.

* Correspondencia: Universitat de València. Facultad de Farmacia. Departamento de Química Física. Avenida V. A. Estellés, s/n. 46100 (Burjassot), Valencia. España. *E-mail*: ramon.garcia@uv.es



Universidad
Católica de Valencia
San Vicente Mártir

RESUMEN

La enfermedad de Chagas es una antropozoonosis parasitaria de transmisión mayoritariamente vectorial. Si bien afecta principalmente a América, constituye una de las mayores enfermedades desatendidas a nivel mundial. Solamente dos fármacos con un alto grado de toxicidad se encuentran disponibles para su tratamiento, por lo que es preciso aunar esfuerzos en pos del desarrollo de nuevas drogas antitripanosomales eficaces y con escasos efectos adversos. En este estudio, se aplicó la topología molecular para el desarrollo de un modelo QSAR capaz de predecir la actividad citotóxica antitripanosomal de un grupo de 39 compuestos. El modelo desarrollado mediante análisis lineal discriminante permitió clasificar correctamente al 93,3 % de los compuestos activos estudiados. Para la predicción de la potencia antitripanosomal se empleó un análisis de regresión multilíneal, que fue capaz de explicar el 90,8 % de la varianza. Ambos análisis fueron validados mediante un test interno de tipo “leave some out”. Se aplicó el modelo obtenido a nuevos compuestos, destacándose ocho moléculas potencialmente activas que podrán ser utilizadas en futuros ensayos preclínicos.

PALABRAS CLAVE: *Chagas, Trypanosoma cruzi, antiparasitarios, benznidazol, análisis QSAR.*

ABSTRACT

Chagas disease is a protozoan anthrozoosis primarily transmitted by vectors. America is the main affected continent, although it is considered a worldwide neglected tropical disease. Only two drugs with a high toxicity level are available for its treatment. Because of this, it is necessary to make an effort in order to develop new effective and safe therapies. In this study, molecular topology was used to create a QSAR model to predict anti-trypanosomal cytotoxic activity in a



group of 39 compounds. Linear discriminant analysis correctly classified 93.3 % of the active studied compounds. To predict the anti-trypanosomal potency, multilinear regression analysis was used; being this capable of explaining 90.8 % of the variance. Both analyses were validated by an internal “leave some out” test. The resulting model was applied to new compounds, selecting eight of them as potential active molecules that could be used in future preclinical studies.

KEYWORDS: *Chagas, Trypanosoma cruzi, antiparasitics, benznidazol, QSAR analysis.*

INTRODUCCIÓN

La enfermedad de Chagas o tripanosomiasis americana es una antropozoonosis parasitaria que afecta principalmente a 21 países latinoamericanos. Sin embargo, a raíz de la migración, ha traspasado esas barreras; siendo actualmente un problema de salud global. Está clasificada por la Organización Mundial de la Salud como una enfermedad tropical desatendida, existiendo alrededor de 7 millones de infectados y 70 millones de personas en peligro de contraerlo [1]. Los niños y las personas que residen en áreas rurales en condiciones de pobreza constituyen la población de mayor riesgo.

Es causada por el protozoo flagelado hemotisular *Trypanosoma cruzi* y es transmitida por insectos vectores de la familia de los triatomíneos [2]. Cuando el vector se alimenta de la sangre humana, excreta heces que contienen los tripomastigotes metacíclicos infectantes. Tras el rascado, esas heces entran en contacto con la herida de la picadura, por lo que los tripomastigotes penetran en el torrente sanguíneo. Los tripomastigotes entran en ciertas células y se transforman en amastigotes que se dividen por fisión binaria y producen más amastigotes y finalmente tripomastigotes, que vuelven al torrente sanguíneo manteniendo el ciclo reproductivo del parásito [3]. Aunque la transmisión vectorial es la forma de adquisición más frecuente, también ha de mencionarse la transmisión vertical [4], endovenosa [5] y oral [2].

La enfermedad tiene tres fases principales: la aguda, con presencia de parásitos en la circulación periférica; la fase crónica asintomática, que puede durar decenas de años o toda la vida; y la fase crónica sintomática. Aunque la primoinfección suele ser asintomática, puede manifestarse en ciertos casos como un chagoma o signo de Romaña. Otras presentaciones clínicas de la fase aguda, así como también de la reactivación chagásica, incluyen: fiebre, linfadenopatías, hepatoesplenomegalia y meningoencefalitis, entre otras. Dentro de las patologías asociadas a la fase crónica sintomática se hace mención de: miocardiopatías, arritmias, megacolon o megaesófago [2].

El benznidazol y el nifurtimox son dos fármacos disponibles frente a Chagas utilizados principalmente durante la fase aguda y la crónica asintomática. Su rol en la fase crónica sintomática es discutida y controversial, ya que a medida que avanza la enfermedad van perdiendo efectividad [3,6-10]. Por otro lado, aun siendo el benznidazol un fármaco de preferencia, presenta un alto grado de toxicidad para el ser humano [11]. Para encontrar nuevos fármacos efectivos y seguros, se han realizado numerosos estudios, algunos de los cuales involucran al posaconazol y otros inhibidores del ergosterol [12-15]. A pesar de ello, ninguno se encuentra validado para su uso frente a la enfermedad de Chagas.

Mediante la aplicación de la topología molecular, TM, se pueden crear modelos que permiten predecir la actividad de ciertos compuestos. La TM está relacionada con la teoría de grafo, siendo



este último una figura que cuenta con vértices y aristas. Se representa una molécula como grafo con sus átomos como sus vértices y los enlaces entre átomos como las aristas. De esta forma, se puede estudiar la conectividad de las moléculas y obtener así determinados índices topológicos sobre la base de las propiedades invariantes de los compuestos [16,17]. Los índices topológicos permiten predecir con precisión propiedades fisicoquímicas y biológicas de diversos grupos de compuestos. Es por esta razón que estas técnicas *in silico* basadas en los parámetros QSAR se emplean desde hace algunas décadas en el diseño y selección de nuevos compuestos activos. Contribuyen a la reducción de costos ya que permiten una adecuada y rápida selección de compuestos potencialmente candidatos a fármacos en ensayos clínicos. Se han utilizado en diferentes áreas terapéuticas, particularmente en la parasitaria [18-23]. Se describen, por ejemplo, estudios con aplicación de la topología molecular para la predicción de la actividad frente a *Trypanosoma cruzi* de compuestos derivados nitrotriazoles [22]. Melo-Filho *et al.* (2019), por otro lado, realizaron mediante parámetros de relación cuantitativa de la estructura-actividad o *quantitative structure-activity relationship* (QSAR) un filtrado de compuestos citotóxicos y con actividad antitripanosomal ante células de mamíferos; obteniendo 39 posibles compuestos activos de 1.063.926 evaluados [12].

En este trabajo, se describe la elaboración de un modelo topológico-matemático de predicción QSAR usando la topología molecular, el análisis lineal discriminante (ALD) y el análisis de regresión multilíneal. El objetivo de este estudio es aplicar la topología molecular con el fin de predecir la actividad antitripanosomal de un compuesto en dos etapas. En primer lugar, se intenta crear un modelo que pueda predecir si un compuesto desconocido es activo o no como citotóxico y antitripanosomal basándonos en el valor de sus descriptores topológicos. Para ello, se busca una función discriminante (FD) obtenida mediante un análisis lineal discriminante (ALD). En segundo lugar, se intenta buscar un modelo de predicción de la potencia de esa actividad mediante el valor de pEC50 predictivo, aplicando un modelo de regresión multilíneal. Por último, gracias a los dos modelos de predicción obtenidos, se realizará un screening de moléculas análogas a los tres compuestos más activos de Melo-Filho *et al.* (2019) [12]. De esta manera podemos proponer nuevos compuestos potencialmente activos frente a la actividad que buscamos a fin de poder analizarlas experimental y eficazmente a posteriori, confirmando o refutando la predicción.

MATERIALES Y MÉTODOS

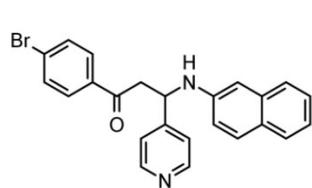
Compuestos analizados y ensayos realizados

Para este estudio se seleccionaron 39 compuestos considerando, como referencia, el fármaco benznidazol, que presenta una concentración media máxima eficaz (EC50) de 3,1 μM . De estos compuestos varios presentaron propiedades para combatir la enfermedad de Chagas, comprobadas experimentalmente en estudios previos, por lo que se dispone de sus respectivos valores de concentración EC50 (μM), así como de su estructura química (tabla 1).

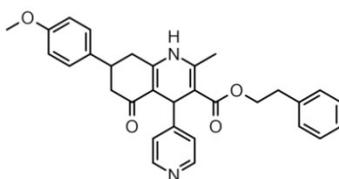
Cabe señalar que se ha mantenido la misma numeración de cada molécula con respecto al trabajo original de donde fueron tomados los compuestos que aparecen en la tabla 1.



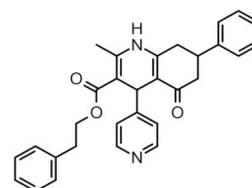
Tabla 1. Estructura química de los compuestos estudiados



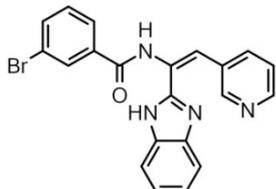
LabMol-96



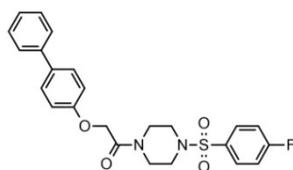
LabMol-97



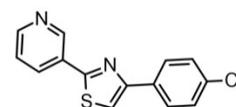
LabMol-98



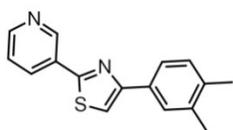
LabMol-100



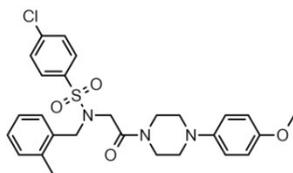
LabMol-101



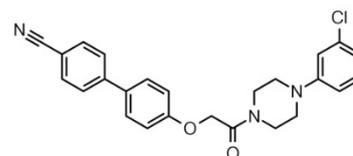
LabMol-102



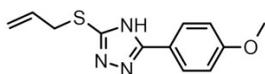
LabMol-103



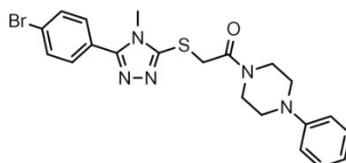
LabMol-104



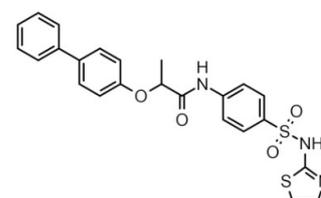
LabMol-105



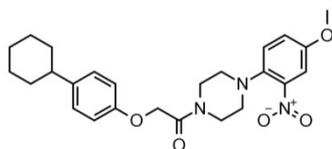
LabMol-106



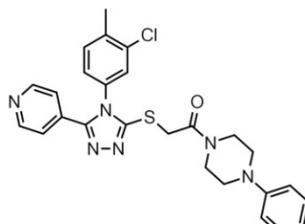
LabMol-107



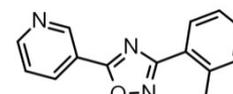
LabMol-108



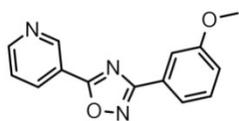
LabMol-109



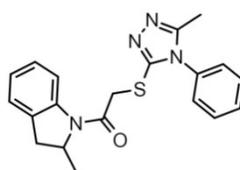
LabMol-110



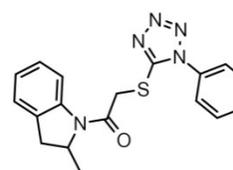
LabMol-111



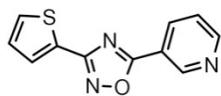
LabMol-112



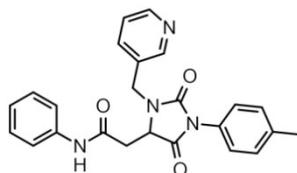
LabMol-113



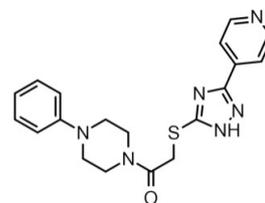
LabMol-114



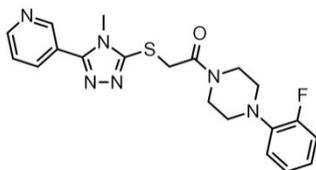
LabMol-115



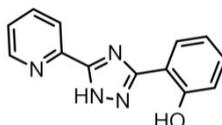
LabMol-116



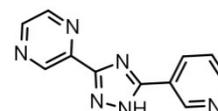
LabMol-117



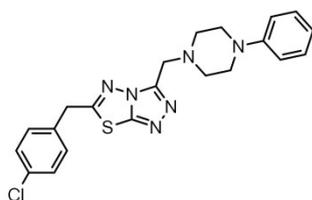
LabMol-118



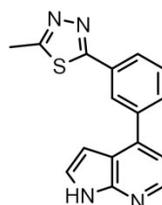
LabMol-119



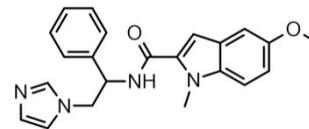
LabMol-120



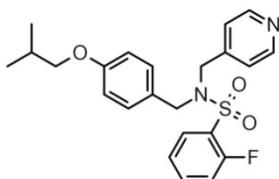
LabMol-121



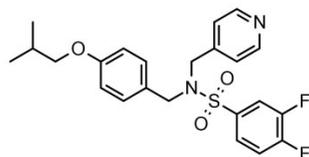
LabMol-122



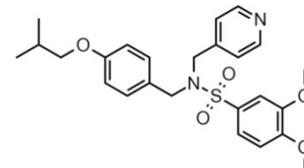
LabMol-123



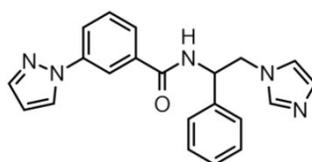
LabMol-124



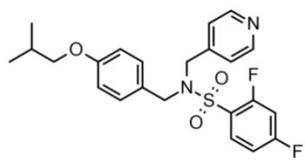
LabMol-125



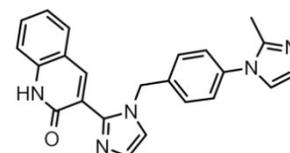
LabMol-126



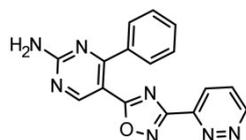
LabMol-127



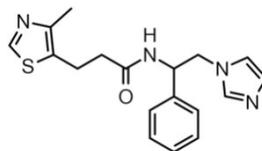
LabMol-128



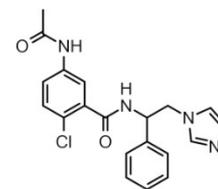
LabMol-129



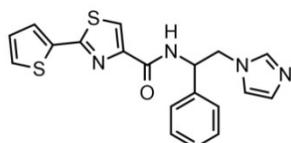
LabMol-130



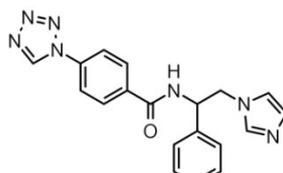
LabMol-131



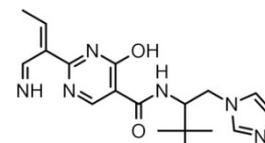
LabMol-132



LabMol-133



LabMol-134



LabMol-135



Descriptores moleculares

Para dibujar la estructura química de las moléculas se utilizó el software ChemDraw® Professional 15.1. Una vez obtenido el gráfico de la estructura química, se importaron las moléculas al software Dragon® [24] y se obtuvieron 583 índices topológicos (IT), tales como: índices de conectividad Randić, Kier y Hall (${}^m\chi_t$, ${}^m\chi^t$) [25] e índices topológicos de carga (ITC) (GGi, JGi) [26]. También se obtuvieron, cocientes y diferencias entre índices de conectividad de valencia y no valencia (${}^mCt = {}^m\chi_t / {}^m\chi^t$ y ${}^mDt = {}^m\chi_t - {}^m\chi^t$), el índice de Wiener (W) [27], y descriptores constitucionales sencillos como el número de vértices (N), longitud del grafo (L) y número de ramificaciones (R).

Análisis discriminante

El análisis lineal discriminante (ALD) es un algoritmo que nos permite distinguir entre dos o más categorías u objetos por medio de una función lineal. En nuestro caso, se trata de diferenciar o discriminar entre compuestos activos e inactivos según los valores de los descriptores de sus moléculas [28].

En este estudio, el grupo activo lo conforman compuestos con probada actividad antitripanosomal y con $EC_{50} < 10\mu M$. En cambio, el grupo inactivo se compone de los compuestos con un $EC_{50} > 10\mu M$. La capacidad discriminante se determinó en función del porcentaje de compuestos clasificados correctamente dentro de cada grupo.

La selección de los descriptores se basó en el parámetro de distribución Fisher-Snedecor (F), y el criterio de clasificación empleado fue la distancia de Mahalanobis. Se clasificó como activo o inactivo cualquier compuesto en función de su distancia más corta al centroide correspondiente [29]. El programa estadístico selecciona las variables para el cálculo de la función discriminante (FD) de manera sucesiva (*stepwise*), es decir, revisa todas las variables y la que contribuya más a la discriminación de los grupos será incluida en el modelo, mientras que la variable que menos contribuya a la predicción será eliminada. Por otra parte, para evaluar la calidad de la función discriminante se empleó el parámetro λ de Wilks, utilizando la prueba de igualdad de las medias de grupo para evaluar si las variables introducidas tienen poder discriminante [30]. De la misma manera, se evaluaron de forma sencilla la calidad de la función discriminante aplicándola a una prueba de validación interna del tipo *leave some out*.

Una vez seleccionada la función discriminante (FD), se realiza el diagrama de distribución de la actividad farmacológica (DDAF). Estos diagramas son descritos como gráficos de tipo histograma, que son útiles para establecer los intervalos de la función FD en donde la expectativa (E) o probabilidad de encontrar compuestos activos es máxima. En los DDAF, E se ubica en el eje de las ordenadas y los valores de la función discriminante, en el eje de las abscisas. Para cada intervalo arbitrario de cualquier función, se define la expectativa de actividad como: $Ea = a/(i+1)$, donde a es el cociente entre el número de compuestos activos en el intervalo dividido por el número total de compuestos activos, e i es el número de compuestos inactivos en el intervalo dividido por el número total de compuestos inactivos. La expectativa de inactividad también se obtiene de la misma manera, como $Ei = i/(a+1)$. Esta representación nos permite ver las zonas en las que la superposición o solapamiento es mínima,



así como determinar los intervalos de la función discriminante, donde la probabilidad de encontrar nuevos compuestos activos es máxima en relación con la elección de un falso activo [31].

Análisis de regresión multilineal (MLRA)

La regresión lineal múltiple permite generar un modelo lineal en el que el valor de la variable dependiente o respuesta, pEC_{50} , ($pEC_{50} = -\log[EC_{50}]$) se predice a partir de un conjunto de variables independientes llamadas predictores (IT). Para hacer el análisis de regresión multilineal (MLRA) se utilizaron los compuestos con valores cuantitativos de EC_{50} en unidades μM . Además, el compuesto LabMol-127 se eliminó al tratarse de un *outlier*. El número de compuestos para hacer el análisis quedó reducido a 24.

El juicio para la selección de variables consistió en usar la agrupación con menor número de variables para evitar un sobreajuste, que el valor del coeficiente de correlación múltiple, R^2 , sea alto ($R^2 > 0,8$) y que el error estándar de estimación fuera mínimo. Para aceptar la función de predicción se realizó un test de validación interna. La validación de la función de predicción se hace a través de una crovalidación interna del tipo *leave one out*; es decir, determinar el coeficiente de predicción (Q^2). Cada compuesto es eliminado del modelo y el valor de la actividad, pEC_{50} , se recalcula con los demás compuestos y descriptores de la ecuación seleccionada. El proceso se repite tantas veces como compuestos se estudien. Con los valores predichos se determina el valor del coeficiente de predicción (Q^2) y se compara con el valor del R^2 . Los valores de $Q^2 > 0,7$ nos muestran que la función obtenida es predictiva y, por tanto, que el modelo seleccionado es de buena calidad.

RESULTADOS Y DISCUSIÓN

Análisis lineal discriminante

La función discriminante obtenida mediante ALD fue aplicada a un total de 39 compuestos, utilizándose al benznidazol como fármaco de referencia. Los compuestos fueron clasificados como grupo activo si presentaban una EC_{50} menor a $10,0 \mu M$. La FD seleccionada fue:

$$FD = 8,073VEA1 - 4,540DECC - 12,33GATS4e - 7,734GATS8p - 6,292 \quad \text{Ec. 1}$$
$$N = 39 \quad \text{Wilks' Lambda, } \lambda = 0,376 \quad F(4, 34) = 14,11 \quad p < 0,00001$$

Los índices topológicos que intervienen en la Ec. 1 son: VEA1, índice obtenido a partir de los valores propios de la matriz de adyacencia; DECC, índice de excentricidad; GATS4e, índice de autocorrelación 2D de Geary $-\text{lag}4$ (que utiliza la electronegatividad atómica); y GATS8, índice de autocorrelación 2D de Geary $-\text{lag}8$ (que utiliza la polarizabilidad atómica) [24].

Las variables utilizadas en la Ec. 1 son estadísticamente significativas, con una $p < 0,05$.



A partir de la FD, un compuesto será considerado como potencial droga si la FD es mayor a 0. La matriz de clasificación ubicó correctamente en el grupo activo al 93,3 % de los compuestos y en el grupo inactivo al 91,7 % de los compuestos, correspondiendo estos valores a la sensibilidad y especificidad respectivamente. Clasificó a la totalidad de los compuestos con una exactitud del 92,3 %.

La clasificación de cada compuesto, el valor calculado de la función discriminante y la probabilidad de actividad se ilustran en la tabla 2.

Tabla 2. Resultados de la clasificación obtenida del análisis lineal discriminante y regresión multilínea para cada compuesto analizado

Comp.	EC ₅₀ exp ^a (μM)	pEC ₅₀	Clas(exp)	Prob. Activ. ^b	FDc	Clas(Pred)	pEC ₅₀ Pred ^d
LabMol-96	0,8	6,10	A	0,893	2,12	A	6,39
LabMol-97	0,4	6,40	A	0,656	0,65	A	5,95
LabMol-98	>200	<3,70	I	0,197	-1,41	I	NP
LabMol-100	18,9	4,72	I	0,001	-6,64	I	4,24
LabMol-101	>200	<3,70	I	0,003	-5,93	I	NP
LabMol-102	4,2	5,38	A	0,999	6,53	A	4,92
LabMol-103	2,7	5,57	A	0,998	6,49	A	5,64
LabMol-104	2,7	5,57	A	0,981	3,94	A	5,76
LabMol-105	>200	<3,70	I	0,131	-1,90	I	NP
LabMol-106	148,8	3,83	I	0,000	-8,04	I	3,57
LabMol-107	>200	<3,70	I	0,020	-3,88	I	NP
LabMol-108	>200	<3,70	I	0,024	-3,72	I	NP
LabMol-109	38,3	4,42	I	0,020	-3,90	I	4,37
*LabMol-110	7	5,15	A	0,075	-2,51	I	5,14
LabMol-111	18,6	4,73	I	0,011	-4,52	I	4,97
LabMol-112	86	4,07	I	0,013	-4,31	I	4,22
LabMol-113	36,9	4,43	I	0,065	-2,67	I	4,55
LabMol-114	108,9	3,96	I	0,000	-7,91	I	4,07
LabMol-115	>200	<3,70	I	0,002	-6,11	I	NP
LabMol-116	190,6	3,72	I	0,006	-5,07	I	4,12
LabMol-117	20,1	4,70	I	0,051	-2,92	I	5,08
LabMol-118	>200	<3,70	I	0,049	-2,97	I	NP
LabMol-119	>200	<3,70	I	0,056	-2,83	I	NP



Comp.	EC ₅₀ exp ^a (μM)	pEC ₅₀	Clas(exp)	Prob. Activ. ^b	FDc	Clas(Pred)	pEC ₅₀ Pred ^d
LabMol-120	3,4	5,47	A	0,809	1,44	A	5,23
LabMol-121	>200	<3,70	I	0,001	-6,55	I	NP
LabMol-122	>200	<3,70	I	0,003	-5,70	I	NP
LabMol-123	13,3	4,88	I	0,020	-3,91	I	5,05
*LabMol-124	>200	<3,70	I	0,795	1,36	A	NP
LabMol-125	0,4	6,40	A	0,952	3,00	A	6,29
LabMol-126	0,5	6,30	A	0,857	1,79	A	5,98
LabMol-127	0,3	6,52	A	0,854	1,76	A	<i>Outlier</i>
LabMol-128	0,7	6,15	A	0,973	3,58	A	6,31
LabMol-129	1,2	5,92	A	0,986	4,27	A	5,74
*LabMol-130	>200	<3,70	I	0,986	4,27	A	NP
LabMol-131	>200	<3,70	I	0,057	-2,80	I	NP
LabMol-132	2,7	5,57	A	0,991	4,66	A	5,76
LabMol-133	>200	<3,70	I	0,045	-3,05	I	NP
LabMol-134	0,3	6,52	A	0,863	1,84	A	6,39
LabMol-135	5,8	5,24	A	0,963	3,26	A	5,43

^a Valores de EC₅₀ experimental tomados del trabajo de Melo-Filho *et al.* 2019 [12]; ^b probabilidad de que el compuesto sea activo; ^c valores de FD obtenidos con la Ec. 1; ^d valores de la actividad predichos con la Ec. 2; * Compuestos erróneamente clasificados por la función discriminante, Ec. 1. A: activo; I: inactivo; NP: No predicho al no clasificarse como activo por la función discriminante. El compuesto clasificado como outlier no se ha tenido en cuenta para obtener la función de predicción.

La validación de la FD se realizó mediante un test interno de tipo *leave some out*. Para ello, los compuestos fueron divididos en 4 subgrupos: A, B, C y D, cada uno con aproximadamente el 25 % de los compuestos. A continuación, se realizó el ALD con los mismos índices topológicos de la FD y se usó como grupo de entrenamiento tres de los cuatro subgrupos. El cuarto se utilizó como test de predicción. El proceso se repitió cuatro veces, intercambiando los subgrupos de entrenamiento y de predicción. Los valores promedio de sensibilidad y especificidad del grupo de entrenamiento y grupo test fueron 90,0 y 96,8 %; 94,4 y 83,3 %, respectivamente. Los resultados obtenidos de discriminación fueron, entonces, similares a los observados por la FD seleccionada.

En la figura 1 se muestra el diagrama de actividad obtenida con la FD. Mediante observación del diagrama, el dominio de aplicación de la FD para la búsqueda de potenciales compuestos activos se da para valores de FD comprendidos entre 0 y 7.



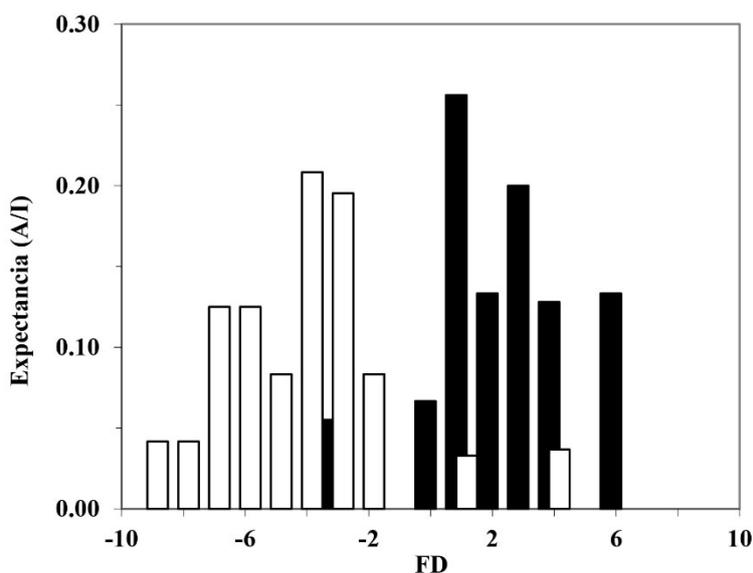


Fig. 1. Diagrama de distribución de actividad antitripanosomal obtenida al graficar la expectancia (E) frente a la distribución discriminante (FD). Las barras de color negro representan al grupo activo y las blancas, al inactivo

Regresión multilíneal

Para hacer el análisis de regresión multilíneal (MLRA) se utilizaron los compuestos con valores cuantitativos de EC_{50} . Se eliminó el compuesto LabMol-127 por tratarse de un *outlier*, al igual que los compuestos cuyos valores de concentración estaban indicados en la tabla 2 como >200 . De este modo, el tamaño de la muestra quedó reducido a 24 compuestos. La función de predicción seleccionada fue:

$$pEC_{50} = -27,93 - 0,30nO + 1,76J + 2,62MATS8p + 15,658BELv1 \quad \text{Ec. 2}$$

N=24 $R^2=0,909$ $Q^2=0,841$ $EEE=0,290$ $F(4,19)=47,2$ $p<0,00001$

Puesto que el p obtenido es menor a 0,05, podemos afirmar que todas las variables del modelo son estadísticamente significativas. En la función de predicción intervienen cuatro índices topológicos: el número de átomos de oxígeno (nO), un descriptor constitucional; un descriptor topológico, el índice de Balaban (J), un índice de autocorrelación 2D que utiliza la polarizabilidad atómica, el índice de autocorrelación de Moran $-\text{lag}8$; y uno de los valores propios de Burden, el valor propio más bajo obtenido con la matriz de Burden, ponderada por los volúmenes atómicos de Van der Waals [24].

Para la validación de la función de predicción, se realizó también una validación cruzada interna del tipo leave one out. Se obtuvo un valor del coeficiente de predicción, $Q^2=0,841$.



En la figura 2 se muestra la correlación obtenida entre los valores experimentales ($pEC_{50 \text{ exp}}$), y los predichos ($pEC_{50 \text{ pred}}$), con la función de predicción seleccionada. Se observa un R^2 de 0,9086, por lo que el 90,86 % de la variabilidad de pEC_{50} queda explicada por el modelo.

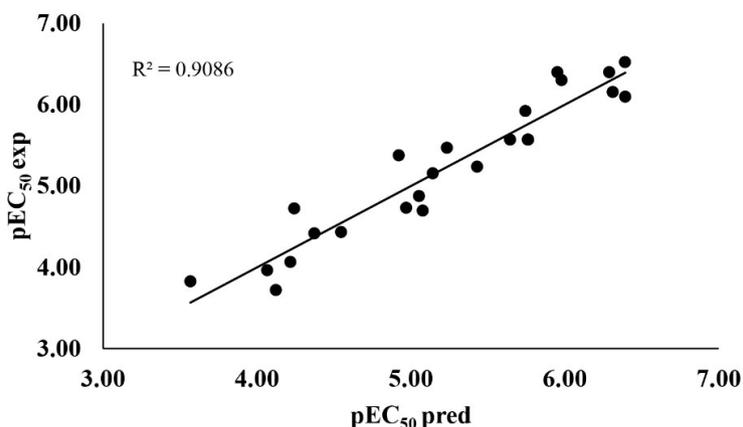


Fig. 2. Representación gráfica de pEC_{50} experimental frente a pEC_{50} predicha obtenida a partir de la función de predicción seleccionada, Ec. 2

La presencia de posibles *outliers* se visualiza en la figura 3. Puesto que el error estándar de estimación (EEE) es igual a 0,290, podemos decir que no se observan valores fuera del rango de $\pm 2*EEE$.

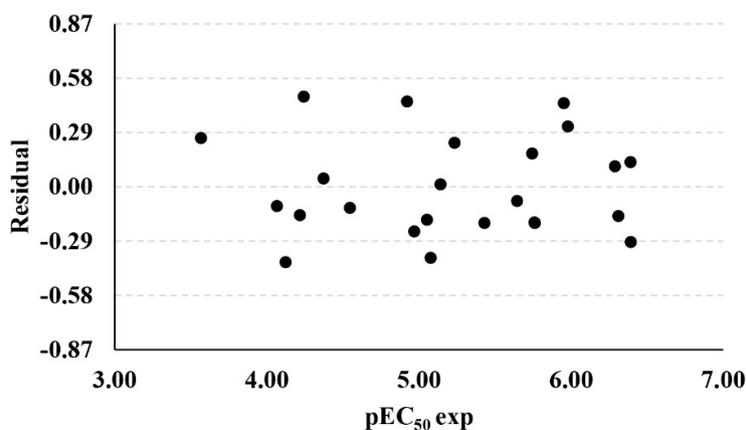


Fig. 3. Correlación entre los valores residuales resultantes de la resta del pEC_{50} calculado y el experimental o conocido con los valores del pEC_{50} experimentales. El intervalo en que se presentan corresponde con $\pm 2*EEE$

En conclusión, puesto que la bondad de ajuste es correcta ($R^2 > 0,9$), el coeficiente de predicción obtenido de la crosvalidación interna, $Q^2 > 0,8$ y además todos los valores se encuentran dentro del rango de $\pm 2*EEE$, se acepta el modelo de regresión lineal múltiple como apto para la predicción

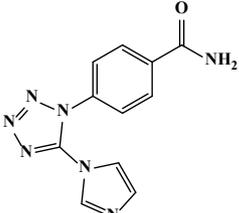
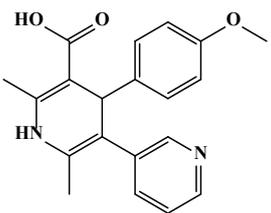
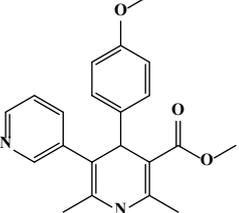
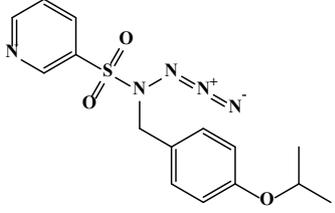
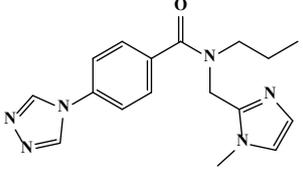
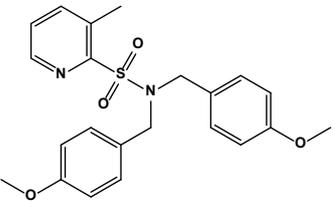
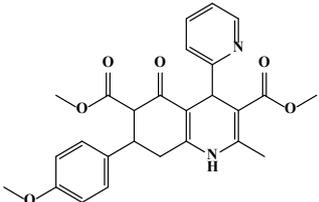
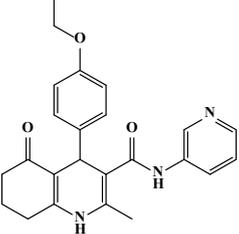


de pEC_{50} . Así, este modelo podrá ser usado en la predicción de la actividad antitripanosomal de los nuevos compuestos.

Utilizando los modelos anteriormente descritos, se ha realizado un *screening* virtual para proponer nuevos compuestos potencialmente activos. Sobre la base de los resultados obtenidos con el ALD y el MRLA, un compuesto se seleccionará como activo frente a *Trypanosoma cruzi* si la función discriminante toma valores comprendidos entre 0 y 6 y la función de predicción, valores de $pEC_{50} > 6,0$ ($EC_{50} < 1\mu M$).

A partir de los tres compuestos más activos LabMol-97, LabMol-125 y LabMol-134, se han seleccionado 150 análogos de la base de datos molecular PubChem [32]. Después de obtener sus índices topológicos se seleccionaron los ocho compuestos encontrados en la tabla 3, ya que son activos en cuanto a la función discriminante y presentan un valor de predicción $pEC_{50} > 7,00$, lo que se corresponde a una actividad $EC_{50} < 0,1\mu M$.

Tabla 3. Identificación y estructura química de algunos de los compuestos seleccionados por el modelo topológico desarrollado con posible actividad frente a *Trypanosoma cruzi* [32]

 <p>CID-1512007 FD=1,53 $EC_{50}=0,004\mu M$</p>	 <p>CID-67850921 FD= 2,26 $EC_{50}=0,01\mu M$</p>	 <p>CID- 19107552 FD=3,25 $EC_{50}=0,02\mu M$</p>
 <p>CID- 151534349 FD=1,58 $EC_{50}=0,06\mu M$</p>	 <p>CID- 50971427 FD= 0,79 $EC_{50}=0,074\mu M$</p>	 <p>CID- 153936258 FD= 1,92 $EC_{50}=0,079\mu M$</p>
 <p>CID- 2939950 FD=3,60 $EC_{50}=0,08\mu M$</p>	 <p>CID-2939380 FD=7,08 $EC_{50}=0,08\mu M$</p>	

En la última década se han realizado numerosos ensayos clínicos con nuevas drogas candidatas para el tratamiento de la enfermedad de Chagas. Aunque el benznidazol sea uno de los fármacos de



referencia, su rol en aquellos pacientes que padecen patologías propias de la fase crónica es controversial [33]; sin mencionar que además presenta un alto grado de toxicidad. El posaconazol y otros derivados azólicos han sido utilizados en numerosos ensayos, con tasas elevadas de fallo al tratamiento [34].

El desarrollo de nuevas drogas antitripanosomales constituye un desafío. Mediante un modelo QSAR, hemos predicho la actividad antitripanosomal de ocho moléculas usando al benznidazol como referencia. Se ha realizado una búsqueda bibliográfica de dichos compuestos a través de PubChem [32]. Ninguno ha sido estudiado previamente como antiparasitario. CID-1512007 ha sido clasificada como inactiva frente a *Mycobacterium tuberculosis*, *Staphylococcus aureus* y *Vibrio cholerae*, entre otros microorganismos. CID-151534349 está categorizado como agonista de prostaglandinas, sin evaluación previa respecto a la actividad antimicrobiana. CID-2939950 ha sido estudiada como potencial droga frente a *Mycobacterium tuberculosis*, y ha demostrado ser inactiva. Hemos de mencionar que varias moléculas se encuentran patentadas [32].

CONCLUSIONES

Se ha conseguido realizar un cribado eficaz de moléculas potencialmente activas al aplicar la topología molecular en la búsqueda de ecuaciones de predicción para compuestos citotóxicos antitripanosomales. Los modelos QSAR basados en descriptores topológicos grafo-teóricos permiten optimizar en tiempo y coste la búsqueda de nuevos compuestos y su posterior caracterización experimental. Se describen ocho moléculas potencialmente activas que han de ser estudiadas a posteriori a fin de corroborar o refutar su papel como fármacos contra la enfermedad de Chagas.

AGRADECIMIENTOS

C. T. C. R., A. C. O. P., A. M. M. y C. B. L. agradecen al máster en Enfermedades Parasitarias Tropicales de la Universitat de València el haber permitido realizar este trabajo. J. G y R. G. D agradecen al proyecto: PID2019-107464RB-C22, “Diseño racional de nuevas herramientas de fitoprotección (Newcroproctools)” concedido por el Ministerio de Ciencia e Innovación de España por el material utilizado para la realización de este trabajo.

REFERENCIAS BIBLIOGRÁFICAS

- [1] World Health Organization. *Integrating neglected tropical diseases into global health and development: fourth WHO report on neglected tropical diseases*. World Health Organization; 2017.
- [2] Pérez-Molina JA, Molina I. Chagas disease. *Lancet*. 2018;391(10115):82-94.
- [3] Moretti NS, Mortara RA, Schenkman S. Trypanosoma cruzi. *Trends Parasitol*. 2020;36(4):404-5.



- [4] Messenger LA, Bern C. Congenital Chagas disease: current diagnostics, limitations and future perspectives. *Curr Opin Infect Dis.* 2018;31(5):415-21.
- [5] Cura CI, Lattes R, Nagel C, Gimenez MJ, Blanes M, Calabuig E, *et al.* Early Molecular Diagnosis of Acute Chagas Disease After Transplantation With Organs From Trypanosoma cruzi-Infected Donors. *Am J Transplant.* 2013;13(12):3253-61.
- [6] Viotti R, Vigliano C, Lococo B, Bertocchi G, Petti M, Alvarez MG, *et al.* Long-term cardiac outcomes of treating chronic Chagas disease with benznidazole versus no treatment: a nonrandomized trial. *Ann Intern Med.* 2006;144(10):724-34.
- [7] Lana M de, Lopes LA, Martins HR, Bahia MT, Machado-de-Assis GF, Wendling AP, *et al.* Clinical and laboratory status of patients with chronic Chagas disease living in a vector-controlled area in Minas Gerais, Brazil, before and nine years after aetiological treatment. *Mem Inst Oswaldo Cruz.* 2009;104(8):1139-47.
- [8] Andrade ALSS, Martelli CMT, Oliveira RM, Silva SA, Aires AIS, Soussumi LMT, *et al.* benznidazole efficacy among Trypanosoma cruzi-infected adolescents after a six-year follow-up. *Am J Trop Med Hyg.* 2004;71(5):594-7.
- [9] Estani SS, Segura EL, Ruiz AM, Velazquez E, Porcel BM, Yampotis C. Efficacy of chemotherapy with benznidazole in children in the indeterminate phase of Chagas' disease. *Am J Trop Med Hyg.* 1998;59(4):526-9.
- [10] Morillo CA, Marin-Neto JA, Avezum A, Sosa-Estani S, Rassi Jr A, Rosas F, *et al.* Randomized trial of benznidazole for chronic Chagas' cardiomyopathy. *N Engl J Med.* 2015;373(14):1295-306.
- [11] Yun O, Lima MA, Ellman T, Chambi W, Castillo S, Flevaud L, *et al.* Feasibility, drug safety, and effectiveness of etiological treatment programs for Chagas disease in Honduras, Guatemala, and Bolivia: 10-year experience of Médecins Sans Frontières. *PLoS Negl Trop Dis.* 2009;3(7):e488.
- [12] Melo-Filho CC, Braga RC, Muratov EN, Franco CH, Moraes CB, Freitas-Junior LH, *et al.* Discovery of new potent hits against intracellular Trypanosoma cruzi by QSAR-based virtual screening. *Eur J Med Chem.* 2019;163:649-59.
- [13] Paucar R, Moreno-Viguri E, Pérez-Silanes S. Challenges in Chagas disease drug discovery: a review. *Curr Med Chem.* 2016;23(28):3154-70.
- [14] Urbina JA. Ergosterol biosynthesis and drug development for Chagas disease. *Mem Inst Oswaldo Cruz.* 2009;104:311-8.
- [15] Maguire JH. Treatment of Chagas' Disease —time is running out. *Mass Medical Soc;* 2015.
- [16] García-Doménech R, Gálvez J, de Julián-Ortiz J V, Pogliani L. Some new trends in chemical graph theory. *Chem Rev.* 2008;108(3):1127-69.
- [17] Amigó JM, Villar Amigó V, Falcó Montesinos A, Gálvez J. Topología molecular. *Boletín la Soc Española Matemática Apl N.* 2007;39.
- [18] Zanni R, Gálvez-Llompert M, García-Doménech R, Gálvez J. Latest advances in molecular topology applications for drug discovery. *Expert Opin Drug Discov.* 2015;10(9):945-57.
- [19] Martínez Rodríguez MÁ, Seguí López-Peñalver RJ, Alcácer Tomás G, Gálvez Álvarez J, García-Doménech R. Aplicación de la topología molecular a la predicción y optimización de la actividad repelente de mosquitos de derivados de la N-acylpiperidina. *Nereis.* 2014;(6):19-26.



- [20] Carrillo JS, Rizza C, Álvarez BE, Hernández D, JG, García-Doménech R. Aplicación de la topología molecular en la búsqueda de nuevos compuestos basados en Azaauronas derivados de las auronas naturales como potenciales antimaláricos. Application of Molecular Topology in the search for new composites based on Azaaurones derived. *Nereis Interdiscip Ibero-American J Methods, Model Simul.* 2017;9:49-62.
- [21] Mateo AJ, García RS, Sarmiento AM, Guidet LG, Gálvez J, García-Doménech R. Application of molecular topology for predicting the leishmanicidal activity of a group of compounds derived from pyrrolo [1,2- α] quinoxaline. *An la Real Acad Nac Farm.* 2016;82(3):317-23.
- [22] Rodríguez Grande C, Villalba de Gregorio B, Bort Carbonell M, Giordanelly Mendicoa C, Gálvez Álvarez J, García Doménech R. Aplicación de topología molecular a la predicción de la actividad frente a *Trypanosoma cruzi* de compuestos derivados nitrotriazoles. *Nereis Rev Iberoam Interdiscip métodos, Model y simulación.* 2016;(8):11-22.
- [23] Souard F, Okombi S, Beney C, Chevalley S, Valentin A, Boumendjel A. 1-Azaaurones derived from the naturally occurring auronas as potential antimalarial drugs. *Bioorg Med Chem.* 2010;18(15):5724-31.
- [24] Dragon for windows (software for molecular descriptor calculations), version 5.4., Talete srl.; Milan, Italy, 2006.
- [25] Kier LB, Hall LH. Molecular connectivity in structure-activity analysis. *Research studies*; 1986.
- [26] Gálvez J, García R, Salabert MT, Soler R. Charge indexes. New topological descriptors. *J Chem Inf Comput Sci.* 1994;34(3):520-5.
- [27] Wiener H. Structural determination of paraffin boiling points. *J Am Chem Soc.* 1947;69(1):17-20.
- [28] García-Doménech R, López-Peña W, Sanchez-Perdomo Y, Sanders JR, Sierra-Araujo MM, Zapata C, *et al.* Application of molecular topology to the prediction of the antimalarial activity of a group of uracil-based acyclic and deoxyuridine compounds. *Int J Pharm.* 2008;363(1-2):78-84.
- [29] Gálvez J, Gálvez-Llompant M, García-Domenech R. Introduction to molecular topology: basic concepts and application to drug design. *Curr Comput Aided Drug Des.* 2012;8(3):196-223.
- [30] Gramatica P, Sangion A. A historical excursus on the statistical validation parameters for QSAR models: a clarification concerning metrics and terminology. *J Chem Inf Model.* 2016;56(6):1127-31.
- [31] Gálvez J, García-Doménech R, de Gregorio Alapont C, de Julián-Ortiz J V, Popa L. Pharmacological distribution diagrams: a tool for de novo drug design. *J Mol Graph.* 1996;14(5):272-6.
- [32] Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, *et al.* PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* 2021;49(D1):D1388-95.
- [33] Marin-Neto JA, Rassi Jr A, Morillo CA, Avezum A, Connolly SJ, Sosa-Estani S, *et al.* Rationale and design of a randomized placebo-controlled trial assessing the effects of etiologic treatment in Chagas' cardiomyopathy: the BENznidazole Evaluation For Interrupting Trypanosomiasis (BENEFIT). *Am Heart J.* 2008;156(1):37-43.
- [34] Morillo CA, Waskin H, Sosa-Estani S, del Carmen Bangher M, Cuneo C, Milesi R, *et al.* Benznidazole and posaconazole in eliminating parasites in asymptomatic *T. cruzi* carriers: the STOP-CHAGAS trial. *J Am Coll Cardiol.* 2017;69(8):939-47.

