

————— *Artículo original* —————

Aplicación de la Topología Molecular en la búsqueda de nuevos agentes activos frente a Leishmania

Recibido el 28 de febrero de 2008

GARCÍA-DOMÉNECH, R. *; DOMINGO-PUIG, C.;
ESTEVE-MARTÍNEZ, M. A.; SCHMITT, J.; VERA-MARTÍNEZ, J.;
CHINDEMI, A. L.; GÁLVEZ, J.

*Departamento de Química Física. Facultad de Farmacia.
Universitat de Valencia*

RESUMEN

Se ha desarrollado un modelo topológico-matemático encaminado a buscar nuevos compuestos derivados del 3,5 dinitrobenceno sulfonamida y/o benzamida activos frente a Leishmania. Utilizando el análisis lineal discriminante y el análisis de regresión multilíneal se ha obtenido un modelo capaz de identificar la actividad del 90% de los compuestos estudiados. Tras realizar un barrido virtual se proponen nuevas estructuras potencialmente activas frente a *L. donovani*.

Palabras clave: Leishmania.—Análisis lineal discriminante.—Análisis de regresión multilíneal.—Índices topológicos.

* **Información de contacto:**

Ramón García Doménech.
Departamento de Química Física, Facultad de Farmacia, Universitat de Valencia.
Avda. V. A. Estellés, s/n. 46100 Burjassot (Valencia).
Telf.: 96 354 42 91. Fax: 96 354 48 92.
email: ramon.garcia@uv.es

ABSTRACT

Application of the Molecular Topology in the search of new agents actives against Leishmania

A topological-mathematical model has been arranged to search for new derivatives of 3,5-dinitrobenzene sulphonamide and/or benzamide, active against *L. donovani*. By using linear discriminant analysis as well as multilinear regression, a model capable to identify the activity of 90% of the analyzed compounds has been achieved. After carrying out a virtual screening based upon such a model, new structures potentially active against the parasite are proposed.

Key words: Leishmania.—Linear discriminant análisis.—Multilinear regresión análisis.—Topological descriptors.

INTRODUCCIÓN

La leishmaniasis es una enfermedad parasitaria causada por protozoos del género *Leishmania*. Ésta se transmite a través de la picadura de dípteros de la subfamilia *Phlebotominae*. Los principales reservorios son animales, tanto silvestres como domésticos, pero el ser humano también constituye un importante reservorio para algunas especies de *Leishmania* (*L. donovani*, *L. tropica*) (1).

Existen cuatro formas principales de leishmaniasis: cutánea, cutánea difusa, mucocutánea y la forma visceral. Esta última, llamada *kala azar*, es la más patógena y se caracteriza por una inflamación de hígado y bazo, acompañada de otros síntomas como fiebre, pérdida de peso y anemia. Sin tratamiento es fatal en el 100% de los casos (2).

La enfermedad es endémica en 88 países de cuatro continentes. En la actualidad, representa un auténtico problema de salud pública, con una prevalencia global de 12 millones de casos y una incidencia anual de 1,5-2 millones de casos (1-1,5 millones la forma cutánea y 500.000 la visceral) (3). Además la incidencia aumenta debido a los movimientos migratorios de las zonas rurales a las industrializadas y el consecuente hacinamiento en los suburbios de las grandes ciudades. La leishmaniasis está considerada por la OMS como una de las enfermedades olvidadas que ha entrado dentro del programa TDR (*Special Programme for Research and Training in Tropical Diseases*). Esta enfermedad afecta a la gente más pobre del planeta (4).

Los tratamientos utilizados hasta el momento están basados en los antimoniales pentavalentes como fármacos de primera línea y el isotianato de pentamidina y anfotericina B como fármacos de segunda elección. Pero ya se han descrito resistencias (tanto en la forma cutánea como en la visceral) a los antimoniales pentavalentes. Además el tratamiento se complica cuando la leishmaniasis aparece asociada al VIH o a inmunosupresión. También hay que remarcar que estos medicamentos son tóxicos y tienen muchos efectos secundarios (5). La miltefosina oral ha demostrado efectividad en el tratamiento de la leishmaniasis visceral, sin embargo, no reporta ninguna eficacia contra la forma cutánea (6).

Por todas estas razones, la necesidad de desarrollar nuevos medicamentos contra la lucha de esta enfermedad es urgente.

En la búsqueda de relaciones cuantitativas estructura-actividad, la topología molecular se ha convertido en la actualidad en una herramienta muy útil para caracterizar estructuralmente a una molécula. Basándose en el método de Conectividad Molecular (7), el cual fue iniciado por Randić (8) y posteriormente desarrollado y descrito por Kier y Hall (9), es posible identificar diferentes niveles estructurales de una molécula a través de unos valores numéricos denominados índices de conectividad. Para ello, cada molécula se asimila a un grafo en donde cada vértice representa a un átomo y cada eje a las conexiones existentes entre ellos. A partir del grafo, se construye la matriz topológica cuyos elementos adquieren valores nulos o no en función de las conexiones existentes entre dichos elementos en el grafo. El tratamiento posterior de la matriz topológica conduce a los índices de conectividad que caracterizan a ese grafo de una manera rápida y sencilla. Estos descriptores se han utilizado con éxito en la búsqueda, selección y diseño de nuevos compuestos activos en diferentes áreas terapéuticas y en el antiparasitario en particular (10-13).

En este trabajo nos centraremos en la actividad de varias dinitro-bencenosulfonamidas, compuestos derivados del herbicida orizalina, ya que su eficiencia contra leishmania ha sido demostrada previamente, inhibiendo la polimerización de tubulina en parásitos purificados y deteniendo su crecimiento en las fases G2/M del ciclo celular (14).

El estudio consiste en la utilización conjunta del análisis lineal discriminante y la regresión multilínea para desarrollar un modelo topológico matemático capaz de identificar la actividad frente a

L. donovani y aplicarlo en la búsqueda de nuevos compuestos potencialmente activos.

MATERIAL Y MÉTODOS

Compuestos estudiados

Se eligió para este estudio un grupo de 57 compuestos [37 de ellos derivados de la 3,5 dinitrobenzoesulfonamida y 20 seleccionados de la base Maybridge Organics Compounds (15)] recogidos en el trabajo publicado por Karl A. Werbovetz (14). La Tabla 1 recoge la estructura química y la actividad antiparasitaria frente a *L. donovani* en términos de concentración inhibitoria 50, CI_{50} (mM) para cada compuesto (la numeración seguida es la misma que aparece en el trabajo).

Descriptores topológicos utilizados

Los índices topológicos empleados en el presente trabajo fueron los siguientes:

Índices de conectividad de Kier y Hall (7)

Los índices de conectividad ${}^m\chi_t$ corresponden en cada caso a la suma:

$${}^m\chi_t = \sum_{j=1}^{n_m} {}^mS_j$$

para todos los subgrafos de tipo t y orden m :

donde n_m es el número de subgrafos de tipo t y orden m . Los términos mS_j representan el inverso de la raíz cuadrada del producto de las valencias de los vértices que constituyen ese subgrafo.

$${}^mS_j = \left[\prod_{i=1}^{m+1} (\delta_i) \right]^{-1/2}$$

donde j define a un subgrafo en particular y d_i representa la valencia topológica de cada vértice (número de ejes que confluyen en ese vértice).

Calculamos índices de conectividad χ_i hasta orden 4 (subgrafos formados por cuatro ejes o aristas) y tipos path (subgrafos en los que las valencias de sus vértices son menores o iguales a 2), cluster (subgrafos que contienen sólo vértices con valencias 1, 3 y/o 4) y path-cluster (contienen vértices con valencias 1, 2, 3 y/o 4).

Los vértices representando a átomos distintos al carbono se identifican con una valencia $\delta^v = Z^v - H$, siendo Z^v el número de electrones de valencia del átomo objeto de estudio, y H el número de hidrógenos unidos a ese átomo, obteniéndose índices tipo ${}^m\chi_i^v$.

Otros descriptores que se han utilizado en este trabajo y que derivan de los anteriores son ${}^mD_i = {}^m\chi_i - {}^m\chi_i^v$ y ${}^mC_i = {}^m\chi_i / {}^m\chi_i^v$.

Índices de carga (16)

Desarrollados en nuestra Unidad de Investigación, evalúan la transferencia de carga entre pares de átomos así como la transferencia global de carga intramolecular. Se definen los índices G_k y J_k como:

$$G_K = \sum_{i=1}^{N-1} \sum_{j=i+1}^N |c_{ij}| \delta(k, d_{ij}) \quad \text{y} \quad J_K = G_K / (N-1)$$

donde:

N : Número de vértices (átomos diferentes al hidrógeno).

d_{ij} : Elementos de la matriz de distancia topológica.

δ : delta de Kronecker.

$c_{ij} = m_{ij} - m_{ji}$ y $M = A \times Q$, siendo:

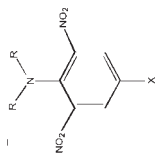
A : Matriz de adyacencia de vértices o de conectividad.

Q : Matriz coulombica: Para $i < j$, $q_{ij} = 1/d_{ij}^2$ y $q_{ii} = 0$.

Así, G_k representa la suma de todos los términos c_{ij} para cada par de vértices i y j separados a una distancia topológica k . En este trabajo $k < 6$.

Los índices de carga de valencia G_k^v y J_k^v se calculan siguiendo el procedimiento anterior usando una matriz modificada A' en cuya diagonal se introduce la electronegatividad de los heteroátomos.

TABLA 1. Estructura química de los compuestos estudiados y resultados de predicción de la actividad frente a *L. donovani* obtenidos para cada compuesto



| Compuesto | R | X | $CI_{50}^{exp^a}$ | FD^b | $\log CI_{50}^{exp^a}$ | $\log CI_{50}^{calc^c}$ | Clasif. |
|---|----------|---|-------------------|--------|------------------------|-------------------------|---------|
| Grupo de entrenamiento activo ($CI_{50} < 20 \mu M$) | | | | | | | |
| 46 | | | 0,5 | 0,87 | -0,301 | 0,776 | A |
| 48 | | | 2,3 | 5,23 | 0,362 | 0,751 | A |
| 33 | n-propil | | 2,5 | 3,09 | 0,398 | 0,837 | A |
| 11 | n-butil | | 2,6 | 3,66 | 0,415 | 0,913 | A |

TABLE 1. Estructura química de los compuestos estudiados y resultados de predicción de la actividad frente a *L. donovani* obtenidos para cada compuesto (cont.)

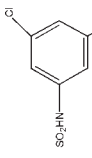
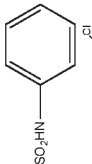
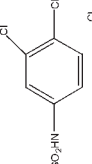
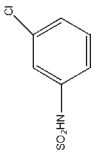
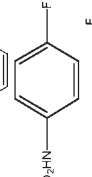
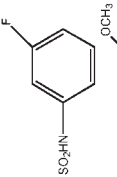
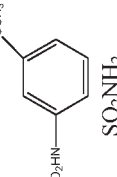

| Compuesto | R | X | $CI_{50}^{exp^a}$ | FD^b | $\log CI_{50}^{exp^a}$ | $\log CI_{50}^{calc^c}$ | Clasif. |
|-----------|----------|---|-------------------|--------|------------------------|-------------------------|---------|
| 02 | n-propil |  | 3,7 | 5,72 | 0,568 | 0,479 | A |
| 01 | n-propil |  | 5 | 1,42 | 0,699 | 1,069 | I |
| 32 | n-propil |  | 5 | 4,55 | 0,699 | 1,095 | I |
| 17 | n-propil |  | 5,5 | 3,53 | 0,740 | 0,837 | A |
| 12 | n-butil |  | 5,6 | 4,69 | 0,748 | 0,928 | A |
| 23 | n-butil |  | 5,7 | 5,29 | 0,756 | 0,719 | A |
| 30 | n-propil |  | 8,1 | 2,69 | 0,909 | 0,973 | A |
| 36 | n-pentil |  | 9 | 2,9 | 0,954 | 0,943 | A |

TABLA 1. Estructura química de los compuestos estudiados y resultados de predicción de la actividad frente a *L. donovani* obtenidos para cada compuesto (cont.)

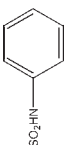

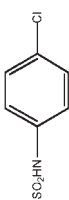

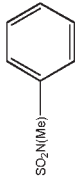
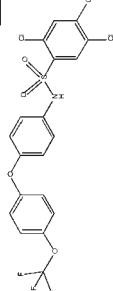

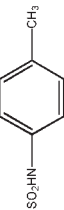
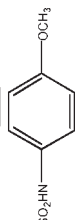
| Compuesto | R | X | CI ₅₀ exp ^a | FD ^b | logCI ₅₀ exp ^a | logCI ₅₀ calc ^c | Clasif. |
|---|----------|---|-----------------------------------|-----------------|--------------------------------------|---------------------------------------|---------|
| 35 | n-etil |  | 11 | 1,32 | 1,041 | 1,063 | I |
| 37 | n-hexil |  | 12 | 3,83 | 1,079 | 0,880 | A |
| 16 | n-propil |  | 13 | 3,34 | 1,114 | 1,012 | I |
| 26 | n-butil |  | 20 | 1,82 | 1,301 | 1,078 | I |
| Grupo de entrenamiento inactivo (CI₅₀ > 20 μM) | | | | | | | |
| 04 | n-propil |  | 21 | -6,38 | 1,322 | 1,375 | I |
| 44 | |  | 25 | -3,34 | 1,398 | 1,416 | I |
| 05 | n-propil |  | 27 | -7,46 | 1,431 | 1,569 | I |
| 15 | n-propil |  | 32 | 2,88 | 1,505 | 1,129 | I |
| 29 | n-propil |  | 32 | 1,83 | 1,505 | 1,262 | I |

TABLA 1. Estructura química de los compuestos estudiados y resultados de predicción de la actividad frente a *L. donovani* obtenidos para cada compuesto (cont.)

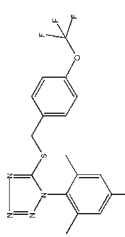
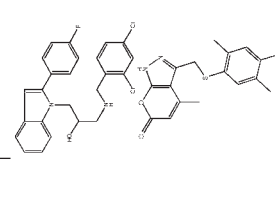
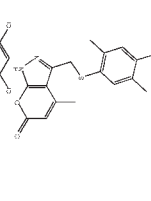
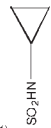
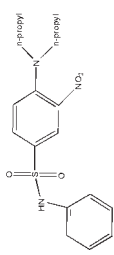

| Compuesto | R | X | $CI_{50}^{exp^a}$ | FD^b | $logCI_{50}^{exp^c}$ | $logCI_{50}^{calc^c}$ | Clasif. |
|-----------|---|---|-------------------|--------|----------------------|-----------------------|---------|
| 45 |  | | 32 | -5,88 | 1,505 | 2,004 | I |
| 54 |  | | 37 | -5,93 | 1,568 | 1,993 | I |
| 55 |  | | 39 | -4,74 | 1,591 | 1,729 | I |
| 09 | n-propil |  | 43 | -3,97 | 1,633 | 1,526 | I |
| 19 | n-propil | $SO_2NH(CH_2)_4CH_3$ | 43 | -1,21 | 1,633 | 1,442 | I |
| 22 |  | | 43 | -6,71 | 1,633 | 1,778 | I |
| 08 | n-propil | $SO_2NH(CH_2)_3CH_3$ | 50 | -1,89 | 1,699 | 1,396 | I |
| 10 | n-propil |  | 50 | -1,22 | 1,699 | 2,046 | I |

TABLA 1. Estructura química de los compuestos estudiados y resultados de predicción de la actividad frente a *L. donovani* obtenidos para cada compuesto (cont.)

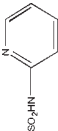
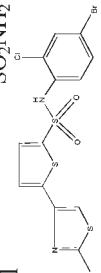
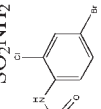
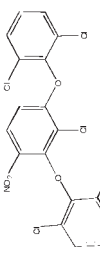
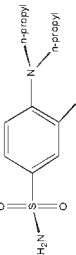
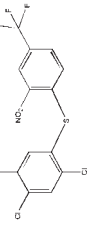
| Compuesto | R | X | $CI_{50,exp}^a$ | FD^b | $\log CI_{50,exp}^a$ | $\log CI_{50,calc}^c$ | Clasif. |
|-----------------|-------------|--|-----------------|--------|----------------------|-----------------------|---------|
| 06 | n-propil | $SO_2NH(CH_2)_2CH_3$ | 54 | -3,25 | 1,732 | 1,365 | I |
| 07 | n-propil | $SO_2N(CH_2CH_2CH_3)_2$ | 55 | -6,14 | 1,740 | 1,598 | I |
| 03 | n-propil |  | 60 | 1,35 | 1,778 | 1,098 | I |
| Oryzalin | n-propil |  | 65 | -0,7 | 1,813 | 1,425 | I |
| 43 | n-propil |  | 66 | -4,05 | 1,819 | 1,844 | I |
| 24 | H, n-propil | SO_2NH_2 | 67 | -1,28 | 1,826 | 2,100 | I |
| 25 | n-etil | SO_2NH_2 | 69 | -0,48 | 1,839 | 1,718 | I |
| 53 | |  | 80 | -3,34 | 1,903 | 1,847 | I |
| 27 | |  | 90 | -9,36 | 1,954 | 2,347 | I |
| 57 | |  | 98 | -3,17 | 1,991 | 1,460 | I |
| 13 | n-propil | $C \equiv N$ | >100 | -4,55 | >2,000 | 1,528 | I |

TABLA 1. Estructura química de los compuestos estudiados y resultados de predicción de la actividad frente a *L. donovani* obtenidos para cada compuesto (cont.)

| Compuesto | R | X | $CI_{50}^{exp^a}$ | FD^b | $\log CI_{50}^{exp^a}$ | $\log CI_{50}^{calc^c}$ | Clasif. |
|-----------|----------|---|-------------------|--------|------------------------|-------------------------|---------|
| 14 | n-propil | | >100 | -1,45 | >2,000 | 1,711 | I |
| 34 | n-propil | | >100 | -0,25 | >2,000 | 1,383 | I |
| 38 | | | >100 | -7,15 | >2,000 | 1,929 | I |
| 39 | | | >100 | -1,28 | >2,000 | 1,805 | I |
| 40 | | | >100 | -4,54 | >2,000 | 2,114 | I |
| 49 | | | >100 | -3,31 | >2,000 | 1,800 | I |
| 50 | | | >100 | -3,05 | >2,000 | 1,696 | I |

TABLA 1. Estructura química de los compuestos estudiados y resultados de predicción de la actividad frente a *L. donovani* obtenidos para cada compuesto (cont.)

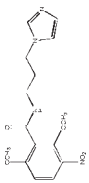
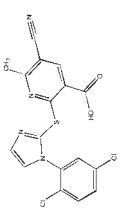
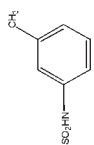
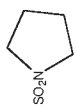
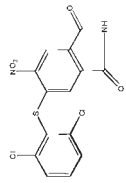
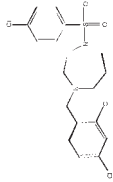
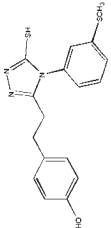
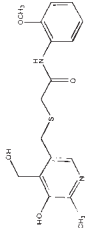
| Compuesto | R | X | CI ₅₀ exp ^a | FD ^b | logCI ₅₀ exp ^a | logCI ₅₀ calc ^c | Clasif. |
|--|--|---|-----------------------------------|-----------------|--------------------------------------|---------------------------------------|---------|
| 51 |  | | >100 | -8,57 | >2,000 | 2,215 | I |
| Grupo test inactivo (CI₅₀ > 20µM) | | | | | | | |
| 42 |  | | 21 | -9,01 | 1,322 | 1,442 | I |
| 31 | n-propil |  | 23 | 3,68 | 1,362 | 0,956 | A |
| 20 | n-propil | SO ₂ NH(CH ₂) ₅ CH ₃ | 26 | -0,58 | 1,415 | 1,526 | I |
| 21 | n-propil |  | 47 | -8,72 | 1,672 | 1,782 | I |
| 18 | n-propil | SO ₂ N((CH ₂) ₃ CH ₃) ₂ | 48 | -3,39 | 1,681 | 1,591 | I |
| 56 |  | | 60 | -9,48 | 1,778 | 1,900 | I |
| 41 |  | | 72 | -14,41 | 1,857 | 1,900 | I |

TABLA 1. Estructura química de los compuestos estudiados y resultados de predicción de la actividad frente a *L. donovani* obtenidos para cada compuesto (cont.)

| Compuesto | R | X | $CI_{50}^{exp^a}$ | FD ^b | $\log CI_{50}^{exp^c}$ | $\log CI_{50}^{calc^c}$ | Clasif. |
|-----------|---|-------------------|-------------------|-----------------|------------------------|-------------------------|---------|
| 28 | n-propil | CONH ₂ | 76 | -3,23 | 1,881 | 1,839 | I |
| 47 |  | | >100 | -9,33 | >2,000 | 1,606 | I |
| 52 |  | | >100 | -8,90 | >2,000 | 2,373 | I |

^a Valores de CI_{50} (μ M) obtenidos en la referencia (6).

^b Valores de la función discriminante, **Eq. 1.**

^c Valores de predicción de $\log CI_{50}$, **Eq. 2.**

Otros descriptores utilizados en este trabajo fueron:

N = número de vértices del grafo.

W = número de Wiener (17).

PR_i = pares de ramificaciones en el grafo separados por una distancia *i*.

En total se han utilizado 62 índices topológicos y para su cálculo el programa DESMOL11 (18).

Tratamiento informático

Análisis de regresión multilínea

El estudio de regresión multilínea lo hemos realizado con ayuda del programa BMDP 9R (19), utilizando los índices topológicos como variables independientes y la concentración inhibitoria-50 (CI₅₀) en su transformación logarítmica, logCI₅₀, como variable dependiente. El estudio estadístico de la función seleccionada se hace atendiendo al número de casos utilizados N, el coeficiente de correlación R, el error estándar de estimación, EEE, el valor F de Fischer y la significación estadística *p*.

La validación de la función de predicción seleccionada se ha realizado con la ayuda de un test externo.

Análisis lineal discriminante

La búsqueda de un modelo topológico-matemático capaz de identificar a un compuesto con actividad frente a leishmania se ha realizado con la ayuda del análisis lineal discriminante, ALD. Mediante esta técnica es posible obtener funciones lineales capaces de clasificar los compuestos estudiados en categorías o grupos (20). En nuestro caso, el ALD se ha aplicado a un amplio grupo de compuestos con actividad frente a *L. donovani*, siendo la CI₅₀, la propiedad utilizada para distribuir a cada compuesto en el grupo activo o inactivo. Con ayuda del paquete estadístico BMDP 7M el programa selecciona los

índices a incorporar en la función discriminante sobre la base del valor de F de Fisher-Snedecor, y el criterio de clasificación seguido es la distancia de Mahalanobis. La calidad de la función discriminante obtenida se evalúa a través del parámetro de Wilks, λ , y con la ayuda de un test externo.

Diagramas de distribución de la actividad farmacológica, DDAF

Una vez obtenidas las funciones discriminantes, es interesante realizar el correspondiente diagrama DDAF para cada una de ellas. Estos gráficos son útiles para determinar el intervalo de la función discriminante en el que la expectancia, E, o probabilidad de encontrar compuestos activos es máxima (21). Los DDAF son histogramas en los que en ordenadas aparece la expectancia E y en abcisas el valor de la función discriminante, FD. Para un intervalo arbitrario de FD, se puede definir la expectancia de actividad, E_a , como: $E_a = a/(i+1)$ en donde a e i representan el número de compuestos activos (o inactivos) en dicho intervalo dividido por el número total de compuestos activos (o inactivos). De forma similar se puede definir la expectancia de inactividad, E_i , como: $E_i = i/(a+1)$.

RESULTADOS Y DISCUSIÓN

La búsqueda de un modelo topológico-matemático útil para predecir la actividad frente a *L. donovani* del grupo de compuestos, objeto de estudio, se realizó en dos etapas:

- Etapa 1. Seleccionar una función discriminante con el objetivo de clasificar y diferenciar los compuestos activos de los claramente inactivos.
- Etapa 2. Seleccionar una función topológica de predicción de la actividad en términos de CI_{50} .

Ambas funciones formarían el modelo matemático capaz de identificar y predecir la actividad de cada uno de los compuestos estudiados y servir de plataforma para la búsqueda de nuevos compuestos activos.

Para obtener la función discriminante, aplicamos el ALD a un grupo de entrenamiento formado por todos los compuestos activos (con valores de $CI_{50} < 20 \mu\text{M}$) y, aproximadamente el 80% del resto de los compuestos catalogados como inactivos ($CI_{50} > 20 \mu\text{M}$). El 20% restante de compuestos inactivos se utilizará como grupo test externo para validar la función discriminante seleccionada.

La función discriminante elegida, FD, fue:

$$\text{FD} = -7,74 - 5,32 \text{}^4\chi_{\text{pc}} - 0,542 \text{G}_1^{\text{v}} + 5,11 \text{G}_3 + 5,16 \text{G}_4^{\text{v}} + 18,52 \text{J}_2^{\text{v}} - 141,58 \text{J}_3^{\text{v}} + 2,51 \text{}^3\text{D}_c \quad \text{Eq. 1}$$

$$\text{N} = 48 \quad \text{F} = 8,97 \quad \lambda(\text{Wilks' lambda}) = 0,389$$

En base a esta función, un compuesto será clasificado con potencial actividad frente a leishmania si $\text{FD} > 0$. En cualquier otro caso se consideraría inactivo. La matriz de clasificación obtenida es muy significativa: el 100% de los compuestos activos son clasificados correctamente, así como el 90,6% del grupo inactivo (29 de 32 compuestos) lo que representa en su conjunto un 93,8% de clasificación correcta.

Los índices topológicos que aparecen en la **Eq. 1** evalúan por una parte el grado de ramificación de la molécula ($\text{}^4\chi_{\text{pc}}$, $\text{}^3\text{D}_c$) y por otra la distribución intramolecular de carga (G_1^{v} , G_3 , G_4^{v} , J_2^{v} , J_3^{v}).

La Tabla 1, columna 5, recoge los valores de FD obtenidos para cada compuesto de la serie de entrenamiento.

Una forma sencilla de evaluar la calidad de la función discriminante seleccionada consiste en aplicarla a un grupo test externo. En nuestro caso disponemos de diez compuestos no utilizados en el análisis discriminante, seleccionados de forma aleatoria y todos ellos con valores de $CI_{50} > 20 \mu\text{M}$. La Tabla 1, columna 5, recoge los resultados obtenidos para cada compuesto. Como puede observarse, todos los compuestos, excepto el **31** son clasificados correctamente como inactivos ($\text{FD} < 0$), lo que representa un porcentaje de acierto del 90%.

La Figura 1 muestra el diagrama de distribución de la actividad antileishmanial obtenido con la función discriminante FD. Las barras negras y blancas corresponden con el grupo activo e inactivo, respectivamente. Como puede observarse, los compuestos activos se reagru-

pan preferentemente dentro del intervalo de FD comprendido entre 0,5 y 7. De tal manera que la máxima probabilidad de encontrar un compuesto activo debe mostrar un valor de FD dentro de ese intervalo.

Con la finalidad de poder predecir el valor de la actividad frente a *L. donovani* para cada compuesto estudiado, realizamos un análisis de regresión multilíneal usando los descriptores topológicos como variables independientes y la CI_{50} en su transformación logarítmica $\text{Log}CI_{50}$ como variable dependiente. Utilizamos la misma serie de entrenamiento que para el ALD. Los compuestos con valores de $CI_{50} > 100 \mu\text{M}$ se les ha asignado un valor de $\text{log}CI_{50} = 2.000$. Lógicamente es una aproximación que habrá que considerar al evaluar la función de predicción que se obtenga.

La función seleccionada, **Eq. 2**, fue:

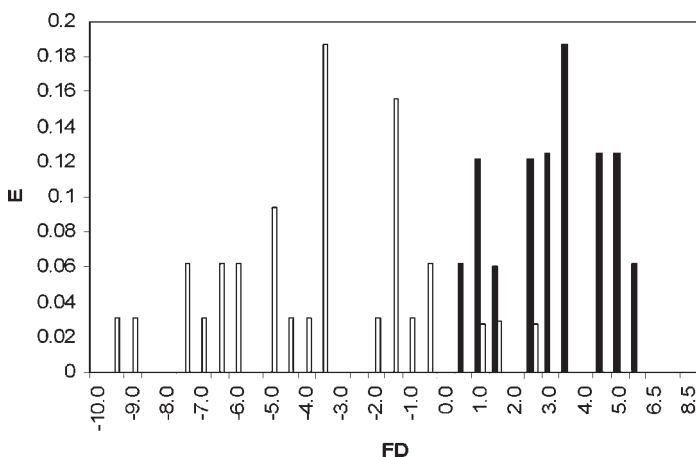


FIGURA 1. *Diagrama de distribución de la actividad frente a L. donovani obtenido a partir de la función de discriminación seleccionada FD, Eq. 1 (las barras negras representan al grupo activo y las blancas al grupo inactivo).*

$$\text{Log } CI_{50} = 13,32 + 0,814 \text{ } ^4\chi_p + 1,381 \text{ } G_3^v - 32,16 \text{ } J_3^v + 0,0018W - 0,717N + 0,332PR1 - 0,263PR2$$

$$N = 48 \quad R = 0,806 \quad EEE = 0,366 \quad F = 10,6 \quad p < 0,0001$$

La Tabla 1 y la Figura 2 recogen los valores de predicción obtenidos a partir de la **Eq. 2** para cada compuesto. Si nos fijamos en la serie de entrenamiento (compuestos representados con puntos blancos en

la Figura 2), aproximadamente el 75% de los compuestos presentan residuales inferiores a $\pm 1\text{EEE}$. El compuesto con un valor residual más alto, es precisamente el más activo de la serie estudiada, **46**. El ALD clasifica a este compuesto como activo. Teniendo en cuenta la aproximación apuntada anteriormente en relación con los compuestos con $\text{CI}_{50} > 100 \mu\text{M}$ podemos considerar aceptable la predicción de la propiedad objeto de estudio. La mayoría de los compuestos más activos, $\text{CI}_{50} < 10 \mu\text{M}$, presentan $\text{LogCI}_{50\text{calc}} < 1.000$. Por otra parte, todos los compuestos tomados como inactivos, $\text{CI}_{50} > 20 \mu\text{M}$, muestran valores de $\text{LogCI}_{50\text{calc}} > 1.000$. En base a estos datos, la función de predicción seleccionada, **Eq. 2**, nos permitirá seleccionar un compuesto con probabilidad de ser activo siempre que el $\log\text{CI}_{50\text{calc}} < 1.000$.

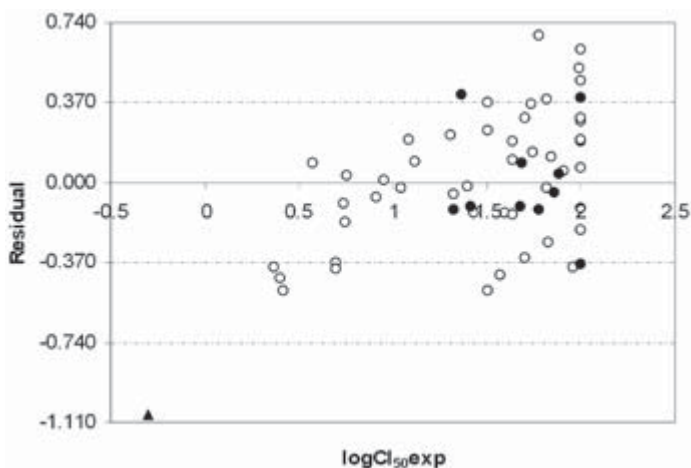


FIGURA 2. *Representación gráfica del residual obtenido para cada compuesto a partir de la Eq. 2 frente al valor experimental, $\log\text{CI}_{50\text{exp}}$.*

La validación de la **Eq. 2** se hizo a través del test externo. Los resultados de predicción del test están reflejados al final de la Tabla 1. Al tratarse de un grupo de compuestos inactivos, todos, excepto el compuesto **31**, presentan valores de $\log\text{CI}_{50\text{calc}} < 1.000$. Los residuales de los compuestos del grupo test son inferiores a $\pm 1\text{EEE}$ o próximos a $\pm 1\text{EEE}$ (ver Figura 2, compuestos representados por puntos negros).

En base a los resultados obtenidos del análisis lineal discriminante y de la regresión multilínea, se puede establecer un modelo topológico para la búsqueda y selección de nuevos agentes activos frente

a *L. donovani* el formado por dos funciones, FD y LogCI_{50} (**Eqs. 1 y 2**) y que cumpla la siguiente condición:

Si $\text{FD} > 0,5$ y $\text{FD} < 7$ y $\text{logCI}_{50} < 1.000$, el compuesto será seleccionado con teórica actividad frente a *L. donovani*. En cualquier otro caso el compuesto será clasificado como inactivo.

La Tabla 1, última columna, muestra la clasificación de actividad obtenida para cada compuesto estudiado siguiendo del modelo topológico seleccionado. Para el grupo de entrenamiento, el porcentaje de acierto de predicción de actividad antileishmanial es del 90% (43 de 48 compuestos correctamente clasificados). Todos los compuestos inactivos son clasificados como inactivos. Para el grupo test externo, la predicción correcta es del 90% (9 de 10 compuestos correctamente clasificados).

Por último realizamos una búsqueda de nuevos compuestos activos con la ayuda de la base Scifinder Scholar (22) y el modelo topológico seleccionado. Se trata de buscar derivados de 3,5 dinitro benceno sulfonamidas y/o benzamidas. La Tabla 2 recoge algunos de los resultados obtenidos. De los compuestos cuya actividad frente a *L. donovani* se describe en la bibliografía (23), los inactivos son reconocidos por el modelo (compuestos 908843-65-2, 908843-73-2, 908843-57-2, 908843-76-5 y 908843-77-6) y buena parte de los activos (compuestos 908843-67-4, 908843-72-1, 902778-99-8). Otros compuestos no evaluados frente a *L. donovani* son reconocidos como teóricamente activos por el modelo topológico. Algunos de ellos poseen actividad antifúngica (compuestos 62745-75-9, 71112-25-9) (24, 25).

Los resultados obtenidos en esta investigación nos animan a seguir trabajando en este campo al abrirse una vía para la búsqueda de nuevos compuestos activos frente a *L. donovani* con la ayuda de la topología molecular.

AGRADECIMIENTOS

Esta investigación ha sido financiada por el Fondo de Investigación Sanitaria, Ministerio de Sanidad (proyecto SAF2005-PI052128) y el Master Internacional en Enfermedades Parasitarias Tropicales (Universitat de Valencia).

TABLE 2. Screening molecular realizado con la base Scifinder Scholar para la selección de nuevos compuestos teóricamente activos frente a *L. donovani* derivados de 3,5 dinitro benceno sulfonamidas y/o benzamidas

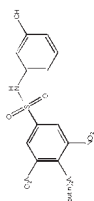
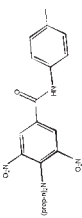
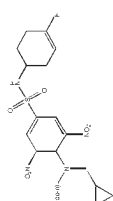
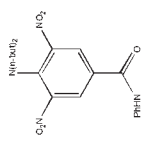
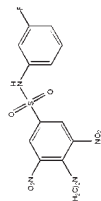
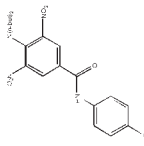
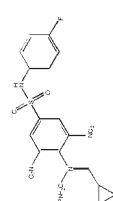
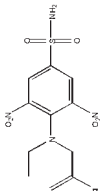
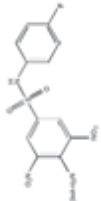
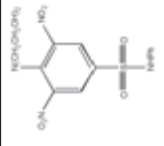
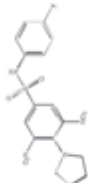
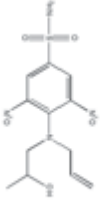
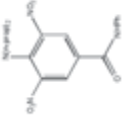
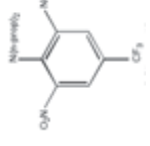
| Compuesto ^a | CI ₅₀ ^b | FD ^c | log CI ₅₀ ^d | Clasif. | Compuesto ^a | CI ₅₀ ^b | FD ^c | log CI ₅₀ ^d | Clasif. |
|--|-------------------------------|-----------------|-----------------------------------|---------|---|-------------------------------|-----------------|-----------------------------------|---------|
|  | 2,8 | 5,76 | 0,72 | A |  | >100 | -0,11 | 1,15 | I |
| 908843-67-4 | | | | | 908843-57-2 | | | | |
|  | 4,5 | -0,56 | 0,82 | I |  | >100 | 1,12 | 1,02 | I |
| 90-8843-70-9 | | | | | 908843-76-5 | | | | |
|  | 6,9 | 8,96 | 0,29 | I |  | >100 | 2,20 | 1,05 | I |
| 908843-63-0 | | | | | 908843-77-6 | | | | |
|  | 8,4 | 2,44 | 0,50 | A |  | — | 0,78 | 0,98 | A |
| 908843-72-1 | | | | | 62745-75-9 | | | | |

TABLA 2. Screening molecular realizado con la base Scifinder Scholar para la selección de nuevos compuestos teóricamente activos frente a *L. donovani* derivados de 3,5 dinitro benceno sulfonamidas y/o benzamidas (cont.)

| Compuesto ^a | CI ₅₀ ^b | FD ^c | log CI ₅₀ ^d | Clasif. | Compuesto ^a | CI ₅₀ ^b | FD ^c | log CI ₅₀ ^d | Clasif. |
|---|-------------------------------|-----------------|-----------------------------------|---------|---|-------------------------------|-----------------|-----------------------------------|---------|
|  | 9,3 | 4,69 | 0,93 | A |  | — | 1,98 | 0,96 | A |
| 902778-99-8 | | | | | 908610-42-1 | | | | |
|  | 73 | 1,57 | 1,34 | I |  | — | 0,32 | 0,85 | A |
| 908843-65-2 | | | | | 71112-25-9 | | | | |
|  | >100 | -1,26 | 1,16 | I |  | — | -2,54 | 1,41 | I |
| 908843-73-2 | | | | | Trifluralin | | | | |

^a Estructura y número de referencia.

^b Valores de CI₅₀ (μM) obtenidos en la referencia (23).

^c Valores de la función discriminante, Eq. 1.

^d Valores de la función discriminante, Eq. 2.

BIBLIOGRAFÍA

- (1) CHENG, C. (1986): *General Parasitology*, pp. 121-127. Academic Press College Division. Orlando.
- (2) LEISHMANIASIS. <http://www.who.int/topics/leishmaniasis/en/>
- (3) WEEKLY EPIDEMIOLOGICAL RECORD. WHO 2002 (44); 77: 365-372.
- (4) WEEKLY EPIDEMIOLOGICAL RECORD. WHO 2002 (25); 77: 205-212.
- (5) HIDE, G.; MOTRAM, J. C.; COOMBS, G. H. and HOLMES, P. H. (1997): *Trypanosomiasis and leishmaniasis, biology and control*. Chemotherapy of Human Leishmaniasis and Trypanosomiasis (Simon L. Croft, J. A. Urbina and Reto Brun), pp. 245-257. CAB International. Wallingford, UK.
- (6) BRENDLE, J. J.; OUTLAW, A.; KUMAR, A.; BOYKIN, D. W.; PATRICK, D. A.; TIDWELL, R. R. and WERBOVETZ, K. A. (2002): Antileishmanial activities of several classes of aromatic dications. *Antimicrobial Agents and Chemotherapy*. 46(3): 797-807.
- (7) KIER, L. B. and HALL, L. H. (1986): *Molecular Connectivity in Structure-Activity Analysis*. Research Studies Press, Letchworth, UK.
- (8) RANDIC, M. (1975): On characterization of molecular branching. *J. Am. Chem. Soc.* 97: 6609-6615.
- (9) KIER, L. B.; MURRAY, W. J.; RANDIC, M. and HALL, L. H. (1976): Molecular Connectivity V: Connectivity Series concept applied to density. *J. Pharm. Sci.* 65: 1226-1230.
- (10) GALVEZ, J.; GARCÍA-DOMÉNECH, R.; DE JULIÁN-ORTIZ, J. V. and SOLER, R. (1995): Topological Approach to Drug Design. *J. Chem. Inf. Comput. Sci.* 35(5): 938.
- (11) MAHMOUDI, N.; DE JULIÁN-ORTIZ, J. V.; CICERÓN, L.; GALVEZ, J.; MAZIER, D.; DANIS, M.; DEROUIN, F. and GARCÍA-DOMÉNECH, R. (2006): Identification of new anti-malarial drugs by linear discriminant analysis and topological virtual screening. *J. Antimicrob. Chemother.* 57(3): 489-497.
- (12) GOZALBES, R.; BRUN-PASCAUD, M.; GARCÍA-DOMÉNECH, R.; GALVEZ, J.; GIRARD, P. M.; DOUCET, J. P. and DEROUIN, F. (2000): Anti-Toxoplasma activities of 24 quinolones and fluoroquinolones in vitro: prediction of activity by molecular topology and virtual computational techniques. *Antimicrob. Agents Chemother.* 44(10): 2771-2776.
- (13) DE JULIÁN-ORTIZ, J. V.; GALVEZ, J.; MUNOZ-COLLADO, C.; GARCÍA-DOMÉNECH, R. and GIMENO-CARDONA, C. (1999): Virtual Combinatorial Syntheses and Computational Screening of New Potential Anti-Herpes Compounds. *J. Med. Chem.* 42(17): 3308-3314.
- (14) DELFIN, D. A.; BHATTACHARJEE, A. K.; YAKOVICH, A. J. and WERBOVETZ, K. A. (2006): Activity of and Initial Mechanistic Studies on a Novel Antileishmanial Agent Identified through in Silico Pharmacophore Development and Database Searching. *J. Med. Chem.* 49: 4196-4207.
- (15) DATA BASE: Maybridge Organics Compounds, <http://www.maybridge.com>
- (16) GALVEZ, J.; GARCÍA-DOMÉNECH, R.; SALABERT, M. T. and SOLER, R. (1994): Charge Indices. New topological descriptors. *J. Chem. Inf. Comput. Sci.* 34: 520-525.

- (17) WIENER, H. (1947): Structural determination of Paraffin boiling points. *J. Am. Soc.* 69: 17-20.
- (18) DESMOL11 SOFTWARE; Unidad de Investigación de Diseño de Fármacos y Conectividad Molecular. Facultad de Farmacia, Universitat de Valencia.
- (19) DIXON, W. J. (1990): BMDP Statistical Software. University of California, Berkeley.
- (20) FURNIVAL, G. M., WILSON, R. W. (1974): Regressions by leaps and bounds. *Technometrics* 16: 499-511.
- (21) GÁLVEZ, J.; GARCÍA-DOMÉNECH, R.; DE GREGORIO-ALAPONT, C.; DE JULIÁN-ORTIZ, J. V. and POPA, L. (1996): Pharmacological distribution diagrams: a tool for the new drug design. *J. Mol. Graph. Model.* 14: 272-276.
- (22) DATA BASE: SCIFINDER SCHOLAR. <http://www.cas.org/support/academic/sf/index.html>
- (23) GEORGE, T. G.; JOHNSAMUEL, J.; DELFÍN, D. A.; YAKOVICH, A.; MUKHERJEE, M.; PHELPS, M. A.; DALTON, J. T.; SACKETT, D. L.; KAISER, M.; BRUND, R. and WERBOVETZA, K. A. (2006): Antikinetoplastid antimitotic activity and metabolic stability of dinitroaniline sulfonamides and benzamides. *Bioorg. Med. Chem.* 14: 5699-5710.
- (24) BECK, J. R. and YAHNER, J. A. (1977): *Fungicides containing 2,6-dinitroaniline derivatives*. Ger. Offen. 22 pp.
- (25) BECK, J. R. and GAJEWSKI, R. P. (1979): *Fungicidal and herbicidal methods*. U.S. 11 pp.