

<i>Nereis. Revista Iberoamericana Interdisciplinar de Métodos, Modelización y Simulación</i>	14	73-88	Universidad Católica de Valencia San Vicente Mártir	Valencia (España)	ISSN 1888-8550
--	----	-------	---	-------------------	----------------

## Aplicación de la topología molecular a la predicción de la actividad repelente de un grupo de carboxiamidas frente a *Aedes aegypti*

Application of molecular topology to the prediction of the repellent activity of a group of carboxyamides against *Aedes aegypti*

Fecha de recepción y aceptación: 16 de febrero de 2022 y 20 de abril de 2022

DOI: 10.46583/nereis\_2022.14.1041

**B. Catalá-Sánchez<sup>1</sup>, A. Meca-Zapata<sup>1</sup>, M. Zúñiga-Orozco<sup>1</sup>, L. Bruno<sup>1</sup>, J. Gálvez<sup>1-2</sup> y R. García-Domenech<sup>1-2\*</sup>**

<sup>1</sup> Máster en Enfermedades Parasitarias Tropicales, Departamento de Farmacia y Tecnología Farmacéutica y Parasitología, Universitat de València.

<sup>2</sup> 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



### RESUMEN

El mosquito *Aedes aegypti* presenta características biológicas que lo convierten en un vector importante en el ciclo de transmisión de diferentes patógenos, en especial arbovirus. Durante la última década, la carga de dengue y otras enfermedades tales como la fiebre del zika o fiebre chikungunya se han incrementado. La expansión de dengue ocasionada por este vector, principalmente en América, representa actualmente uno de los problemas más grandes de salud pública, siendo endémico en más de 30 países. Frente a la falta de tratamientos específicos para tratar las infecciones transmitidas por este mosquito, los esfuerzos por hacerle frente se han focalizado en el control vectorial, aplicando diversas estrategias de combate, entre las que se incluye la aplicación de insecticidas y larvicidas para conseguir su eliminación, así como el uso de repelentes como estrategia de protección individual. La búsqueda de nuevos repelentes puede realizarse mediante diferentes herramientas, donde se incluye la topología molecular. En este contexto, computacionalmente a través de un modelo topológico-matemático, evaluamos la potencial actividad repelente de una serie de carboxiamidas preseleccionadas mediante el uso de análisis discriminante lineal y de regresión multilíneal. Tras este cribado virtual basado en el modelo seleccionado, proponemos nuevas estructuras químicas con una actividad potencialmente activa como repelentes contra *Aedes aegypti*.

**PALABRAS CLAVE:** *Aedes aegypti*, análisis de regresión multilíneal, análisis discriminante lineal, análisis QSAR, topología molecular, actividad repelente



## ABSTRACT

Due to its biological characteristics, the mosquito *Aedes aegypti* is an important vector in the transmission cycle of various pathogens, especially of arboviruses. The burden of dengue and other diseases such as the Zika virus infection or the chikungunya fever has increased over the last decade. The spread of dengue caused by this vector, mainly in America, represents one of the greatest public health problems, being endemic in more than 30 countries. Faced with the lack of specific treatments to treat the infections caused by this mosquito, efforts have focused on controlling the vector, applying various combat strategies, including the application of insecticides and larvicides to directly eliminate the mosquito population, as well as the use of repellents, as an individual protection strategy. The search for new repellents can be done using different tools, including molecular topology. In this context, using a topological-mathematical model, we evaluated the potential repellent activity of a series of preselected carboxyamides by using linear discriminant and multilinear regression analysis. After carrying out a virtual screening based on the selected model, new chemical structures with potential activity as repellents against *Aedes aegypti* are proposed.

**KEYWORDS:** *Aedes aegypti*, multilinear regression analysis, linear discriminant analysis, control, QSAR analysis, molecular topology, repellent activity

## INTRODUCCIÓN

El vector *Aedes aegypti*, también conocido como mosquito del dengue, mosquito momia o mosquito de la fiebre amarilla, está relacionado con la transmisión de enfermedades virales como dengue y fiebre amarilla, así como con otras enfermedades como chikungunya, fiebre de Zika y el Virus Mayaro [1].

Este mosquito pertenece al género *Aedes*, que incluye al estrechamente emparentado *Aedes albopictus*, vector también del dengue; y más concretamente al subgénero *Stegomyia*. Esta especie fue originalmente descrita como *Culex aegypti* por Carl Linnaeus y su discípulo Fredrik Hasselquist en 1762 pero fue posteriormente nombrada por otros autores como *Culex calopus* (Meigen en 1818) o *Stegomyia fasciata* (Taylor en 1903), entre otros. No fue hasta dos siglos después, cuando se solicitó en la Comisión Internacional de Nomenclatura Zoológica la evaluación de su nombre que concluyó con la denominación de la especie tal como actualmente la conocemos, *Aedes aegypti*; propuesta por primera vez por Dyar en la década de 1920 [2].

Morfológicamente, puede reconocerse por sus distintivas marcas blancas (en algunas regiones como Venezuela, se le conoce como ‘zancudo de pata blanca’), no obstante, sus diferencias en aspecto con respecto a otros mosquitos pueden ser ligeras [2]. Aunque machos y hembras pueden distinguirse por algunos caracteres como las antenas, uno de los rasgos de mayor relevancia en los que difieren es la alimentación. Mientras que el macho se alimenta de carbohidratos como el néctar obtenido de las plantas (fitófago), la hembra se alimenta de sangre (hematófaga) [3]. Ésta última, aunque puede hacerlo en cualquier momento del día, suele picar con más frecuencia al amanecer y al atardecer [2].

Debido a que en su ciclo biológico incluye diferentes estadios (holometábolos), algunos de ellos exclusivamente acuáticos como el larvario, los lugares propicios para que se dé la reproducción son



aquellos en los que existe agua estancada y limpia como en recipientes descubiertos y abandonados, tiestos, neumáticos desechados, agua de sumideros de los patios etc. [2,3].

En relación con su distribución mundial, el ancestro de la forma doméstica de *Aedes aegypti* se originó en África Subsahariana, pero actualmente se encuentra en regiones tanto tropicales como subtropicales de todo el mundo. Su distribución abarca las latitudes 40° norte y 40° sur y se encuentra comúnmente hasta los 1.700 metros sobre el nivel del mar e incluso, aunque en menos frecuencia, puede ir desde los 1.700 metros a los 2.200 m sobre el nivel del mar. Registros recientes han encontrado incluso su presencia a una altura de 2.302 metros sobre el nivel del mar.

Uno de los factores climáticos que más parece afectar a su distribución es la temperatura, ya que se trata de una especie altamente susceptible a temperaturas extremas y a climas cálidos-secos. Estudios recientes de modelización de su distribución potencial, teniendo en cuenta registros mundiales y diversas variables ambientales, demuestran que la temperatura es el predictor más importante, seguido por otros como la precipitación y los índices de vegetación; de modo que hechos como los nuevos asentamientos humanos en la montaña o el cambio climático podrían explicar el incremento en su rango de distribución, llegando a presentarse como endémico hasta en 30 países. Concretamente, aquellos países englobados en la región Neotropical son los que presentan una significativa abundancia del vector ya que éste se ve favorecido por las condiciones ambientales presentes en esta zona como la temperatura y la humedad.

A esta problemática de expansión en su rango de distribución, se le suma el impacto en la salud de la población ocasionada por las enfermedades, mencionadas anteriormente, que son transmitidas por este vector. Según la OMS, se estima que esta especie de mosquito causa 50 millones de infecciones y 25000 muertes por año [2].

La falta de tratamientos específicos contra las infecciones transmitidas por este vector ha impulsado a la búsqueda de medidas de profilaxis, focalizadas principalmente en el control vectorial. Para ello, se están aplicando diversas estrategias de combate entre las que se incluyen el uso de insecticidas y larvicidas para conseguir su eliminación junto con estrategias de protección individual mediante el uso de repelentes [1].

Los repelentes disminuyen el contacto entre el mosquito y el humano, pudiendo disminuir la tasa de transmisión de enfermedades. Su uso ha sido empleado por la humanidad durante milenios, sobre todo los de origen natural como las hierbas o aceites y todavía es utilizado hoy en día entre un 50 y un 90% por los habitantes de los trópicos rurales [4].

Uno de los repelentes más conocido es el DEET (N, N-dietil-3-metilbenzamida), repelente sintético de amplio espectro y más eficaz, cuyo uso data desde 1950. Desde entonces, se han identificado muchos otros compuestos repelentes en pruebas de laboratorio y campo. Estos repelentes incluyen tanto compuestos químicos de origen sintético como de origen natural que se comercializan en diversas formulaciones (spray, vaporizadores, cremas, geles, lociones etc.) [4].

Los esfuerzos por encontrar repelentes superiores al DEET, mediante la búsqueda de análogos y otras carboxiamidas estructuralmente similares al mismo, se ha hecho con más frecuencia de manera experimental. Sin embargo, los estudios computacionales también son un método importante que puede ser empleado para diseñar nuevos fármacos, como es el caso de la topología molecular, particularmente la topología molecular [5], capaz de encontrar relaciones cuantitativas entre la estructura química de las moléculas y su actividad (QSAR) [4,6].



La topología molecular permite el cálculo sencillo de descriptores topológicos por medio de la representación de cada estructura molecular como un hidro-grafo, donde los átomos están representados por vértices y los enlaces por aristas. La conectividad entre cada átomo y los demás se incluye en las matrices topológicas (de distancia o de adyacencia), cuya manipulación matemática proporciona diferentes conjuntos de números llamados índices topológicos [5] con capacidad de caracterizar de manera sencilla y eficiente la estructura molecular. La selección adecuada de estos índices topológicos hace posible la obtención de una caracterización muy específica de cada compuesto químico, permitiendo su uso en modelos QSAR [7–11] y por tanto, con utilidad para la selección y diseño de nuevos fármacos [12].

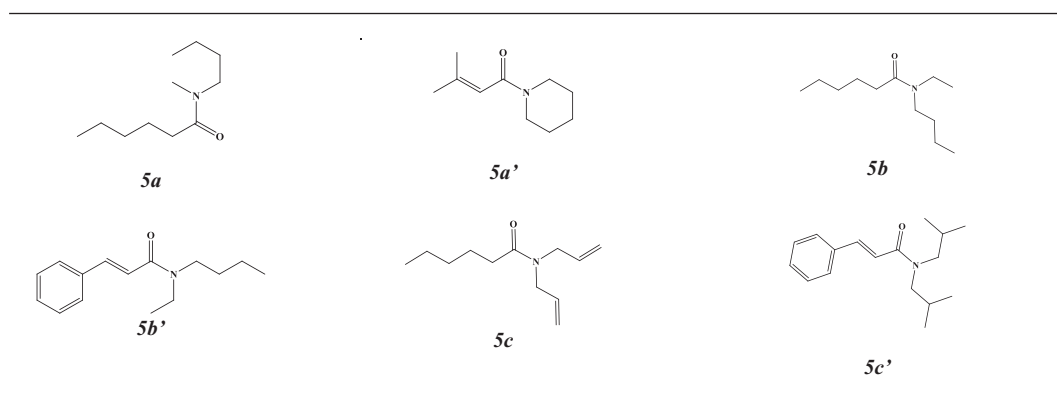
Bajo estos antecedentes, el presente estudio investigó la actividad potencial repelente de un grupo de carboxiamidas preseleccionadas con el objetivo de obtener un modelo topológico-matemático de predicción QSAR utilizando topología molecular, el análisis discriminante lineal y el de regresión multilíneal. Posteriormente, se realizó un cribado virtual de moléculas análogas a un subconjunto de las carboxiamidas iniciales mediante el uso del modelo QSAR preseleccionado a fin de recopilar nuevos compuestos con una actividad potencialmente activa como repelentes frente a *Aedes aegypti*.

## MATERIALES Y MÉTODOS

### Compuestos analizados

Para el presente estudio, se seleccionó un grupo de 43 moléculas carboxiamidas con actividad potencial como repelentes para *Aedes aegypti*. En las Tabla 1 y 2 se muestra la estructura química y la actividad repelente (valores de mínima dosis efectiva MED ( $\mu\text{mol}/\text{cm}^2$ ), respectivamente, para cada molécula, las cuales fueron recopilados a través de estudios previos [4].

Tabla 1. Estructura química de los compuestos estudiados



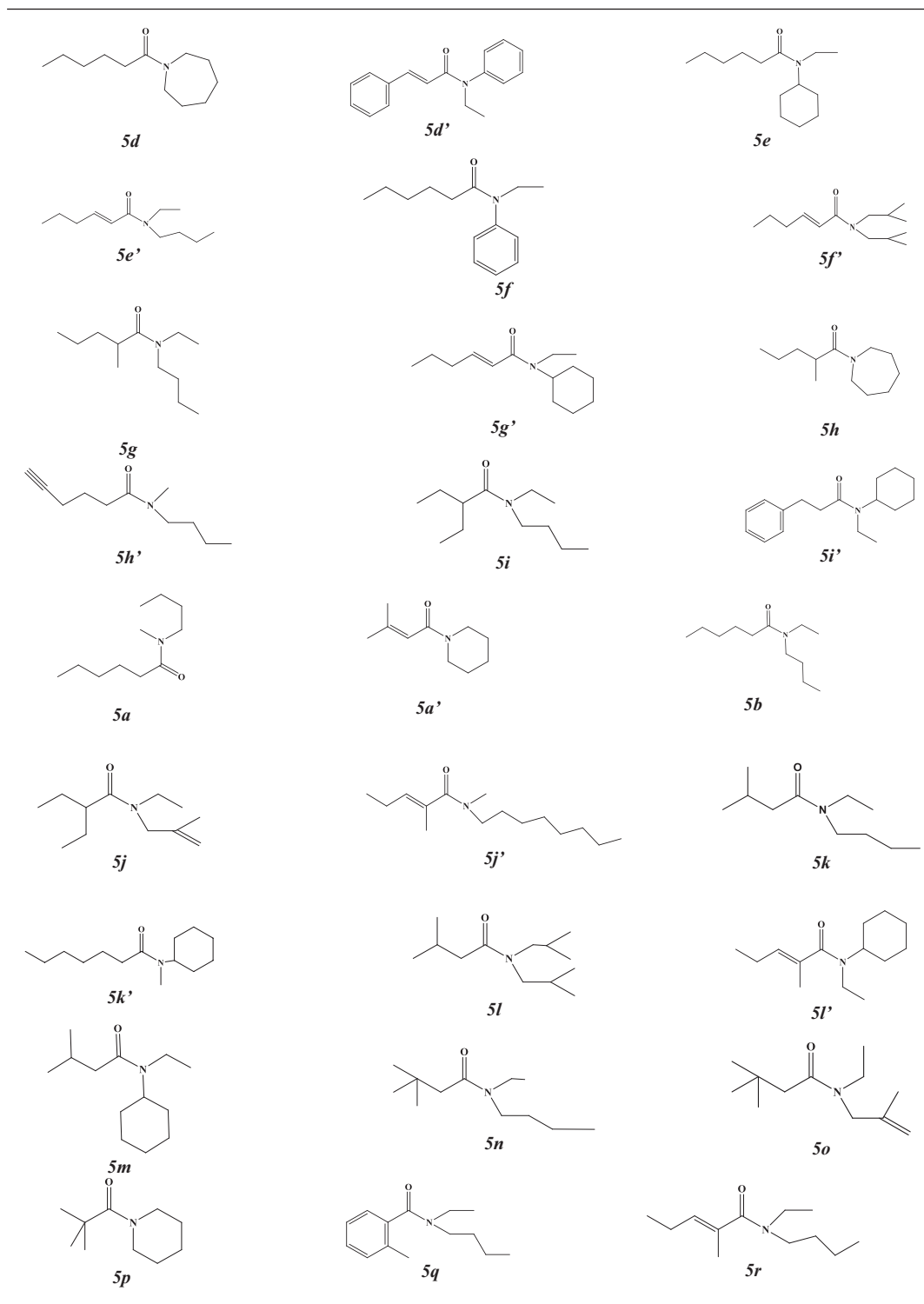
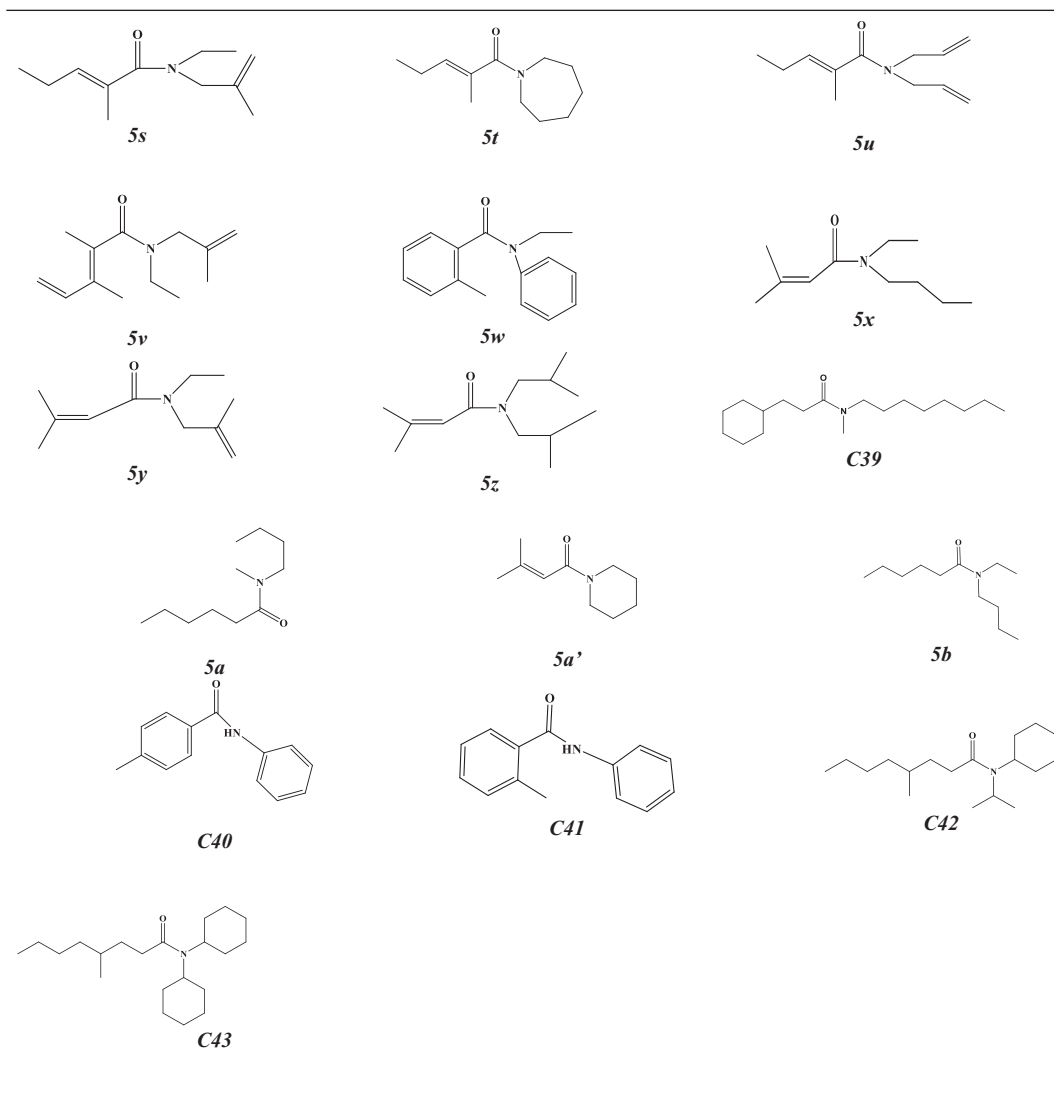


Tabla 1. Estructura química de los compuestos estudiados (Continued)

Tabla 2. Valores de la actividad repelente experimental, MED<sub>exp</sub>, frente a *Aedes aegypti* de 43 carboxiamidas [4]. MED como mínima dosis efectiva,  $\mu\text{mol}/\text{cm}^2$ , como medida de la concentración superficial mínima requerida para producir un efecto repelente.

Molécula	Nombre	MED <sub>exp</sub>	Molécula	Nombre	MED <sub>exp</sub>
5a	N-butil-N-metil-hexanamida	0,12	5l	N,N-diisobutil-3-metilbutanamida	0,41
5a'	Hexahidro-1-(3-metilcrotonil)-1H-azepina	0,14	5l'	(E)-N-ciclohexil-N-etil-2-metilpent-2-enamida	0,14
5b	N-butil-N-etilhexanamida	0,16	5m	N-ciclohexil-N-etil-3-metilbutanamida	0,17



Molécula	Nombre	MED exp	Molécula	Nombre	MED exp
5b'	N-butil-N-etil-cinamamida	10,75	5n	N-butil-N-etil-2,2-dimetilpropanamida	0,29
5c	N,N-dialhexanamida	0,20	5o	N-etil-2,2-dimetil-N-(2-metil-2-propenil)propanamida	0,47
5c'	N,N-bis(2-metilpropil)-3-fenil-2-propenamida	20,13	5p	1-(1-azepanil)-2,2-dimetil-1-propanona	0,31
5d	Hexahidro-1-(1-oxohexil)-1H-azepina	0,03	5q	N-butil-N-etil-2-metilbenzamida	0,16
5d'	N-etil-N,3-difenil-2-propenamida	20,25	5r	(E)-N-butil-N-etil-2-metil-2-pentenamida	0,12
5e	N-ciclohexil-N-etilhexanamida	0,27	5s	(E)-N-etil-2-metil-N-(2-metil-2-propenil)-2-pentenamida	0,18
5e'	(E)-N-n-butil-N-etil-2-hexanamida	0,27	5t	(E)-1-(1-azepanil)-2-metil-2-penten-1-ona	0,10
5f	N-etil-N-fenilhexanamida	0,63	5u	(E)-2-metil-N,N-di-2-propenil-2-pentenamida	0,42
5f'	(E)-N,N-di-(2-metilpropil)-2-hexenamida	0,63	5v	N-etil-2-metil-N-(2-metil-2-propenil)benzamida	0,15
5g	N-butil-N-etil-2-metilpentanamida	0,10	5w	N-etil-2-metil-N-fenilbenzamida	5,16
5g'	(E)-N-ciclohexil-N-etil-2-hexanamida	0,65	5x	N-butil-N-etil-3-metil-2-butanamida	0,19
5h	1-(1-azepanil)-2-metil-1-pentanona	0,10	5y	N-etil-3-metil-N-(2-metil-2-propenil)-2-butanamida	0,31
5h'	N-butil-N-metil-5-hexinamida	0,18	5z	N,N-diisobutil-3-metilcrotonamida	0,22
5i	N-butil-N,2-dietilbutanamida	0,13	C39	3-ciclohexil-N-metil-N-otilpropanamida	25,00
5i'	N,3-diciclohexil-N-etilpropanamida	20,50	C40	4-metil-N-fenilbenzamida	25,00
5j	N,2-dietil-N-(2-metil-2-propenil)butanamida	0,38	C41	2-metil-N-fenilbenzamida	25,00
5j'	(E)-N,2-dimetil-N-octilpent-2-enamida	0,13	C42	N-ciclohexil-N-isopropil-4-metiloctanamida	25,00
5k	N-butil-N-etil-3-metilbutanamida	0,13	C43	N,N-diciclohexil-4-metiloctanamida	25,00
5k'	N-ciclohexil-N-metilheptanamida	0,17			

## Descriptores moleculares

Este trabajo utilizó, con el fin de hacer una identificación estructural de cada compuesto, un conjunto de descriptores topológicos bien conocidos como los índices de Randić, Kier y Hall hasta el cuarto orden ( ${}^m\chi_t$ ,  ${}^m\chi^v$ ) [5,13], índices topológicos de carga (TCI), hasta quinto orden, (Ji, Gi), [14], cocientes y diferencias entre valencia e índices de conectividad sin valencia ( ${}^mC_t = {}^m\chi_t / {}^m\chi^v$  y  ${}^mD_t = {}^m\chi_t - {}^m\chi^v$ ) así como el índice de Wiener, W, longitud de la molécula, L, n° de pares de ramificaciones, PRi y valencia de las aristas, Vi. Cada compuesto se caracterizó por un conjunto de 62 descriptores



topológicos. Todos los índices utilizados en el presente estudio se obtuvieron con ayuda del programa Desmol2 accesible vía internet, <http://desmol2.uv.es>

## Algoritmos QSAR

### *Análisis lineal discriminante*

El objetivo del análisis lineal discriminante (ALD), considerado como un algoritmo heurístico capaz de diferenciar entre dos o más categorías u objetos, es hallar una función lineal capaz de discriminar entre las moléculas activas e inactivas mediante diferentes valores de sus descriptores. Para este análisis, consideramos 2 subconjuntos de compuestos: el primero, grupo A, con actividad repelente contra *Aedes aegypti* (compuestos con  $MED < 0,5 \mu\text{mol}/\text{cm}^2$ ) y el segundo, grupo I, sin actividad repelente contra este mismo vector (compuestos con  $MED > 0,5 \mu\text{mol}/\text{cm}^2$ ). La capacidad discriminante fue probada por el porcentaje de clasificaciones correctas en cada grupo (matriz de clasificación). La selección de los índices topológicos para la función discriminante se hizo bajo la evaluación del parámetro lambda de Wilk's,  $\lambda$ , (varía entre 0-1) el cual está relacionado con el análisis multivariante de la estadística de varianza que prueba la igualdad de las medias de grupo para la(s) variable (s) en la función discriminante. A partir de la función discriminante obtenida tras la selección de un número conveniente de índices topológicos (en función de N) que presentaban un menor valor de lambda de Wilk, se extrajo el diagrama de distribución de la actividad repelente, PDD. El diagrama se representó para establecer los intervalos de la función discriminante en los que la probabilidad/expectativa, E, de encontrar compuestos repelentes es máxima. Los PDD son gráficos similares a histogramas de funciones de conectividad donde las expectativas aparecen en el eje de ordenadas. Para un intervalo arbitrario de valores de una función dada, definimos la expectativa de actividad como:  $E_a = a / (I + 1)$ ; donde "a" es el número de compuestos activos en el intervalo dividido por el número total de compuestos activos y "I" es el número de compuestos inactivos en el intervalo dividido por el número total de compuestos inactivos. La expectativa de inactividad se define de manera simétrica, como  $E_i = I / (a + 1)$ . De modo que el diagrama de distribución proporciona una buena visualización de las regiones de mínima superposición permitiendo la selección de aquellas regiones en las que la probabilidad de encontrar compuestos activos alcanza un máximo [12].

Con el fin de validar la función de discriminación seleccionada, se llevó a cabo una validación interna de tipo leave-one-out para la cual, se elimina un caso del conjunto de datos y se repite el análisis con N-1 casos restantes junto con los descriptores topológicos que fueron seleccionados en el primer análisis discriminante lineal. El valor de actividad correspondiente al caso retirado será predicho y deberá ser el mismo que se obtuvo en el análisis discriminante lineal inicial. El procedimiento se repite hasta completar el número de casos presentes en los datos [15].

### *Análisis de regresión multilineal*

La ecuación de regresión se obtuvo correlacionando la variable MED, transformada previamente de manera logarítmica (pMED), con los índices topológicos mencionados previamente mediante un





análisis de regresión multilínea (MLRA) por medio del software STATISTICA. Para ello, el programa buscó subconjuntos con 1, 2, 3, etc. variables independientes (índices topológicos) para predecir el comportamiento de la variable dependiente (pMED), aportando diferentes ecuaciones con sus correspondientes  $R^2$ , es decir, el porcentaje de la variación en la variable dependiente que es explicado por las variables independientes (modelo lineal). La selección del subconjunto de índices topológicos que mejor predecían la propiedad experimental, con el fin de obtener la ecuación de regresión, se hizo bajo el criterio de un  $R^2$  elevado ( $R^2 > 0,8$ ) y de un número de variables no muy elevado (en función de N), pero lo suficiente para predecir el comportamiento de la variable de estudio. Dicha relación, debería tener una alta significación estadística ( $p < 0,05$ ). Asimismo, se consideró, para el modelo seleccionado, un error estándar de estimación (EEE) inferior al 10% de variabilidad de la propiedad estudiada. Con el fin de validar la función de predicción del modelo seleccionado, se llevó a cabo una validación interna de tipo leave-one-out con la determinación del coeficiente de predicción  $Q^2$  ( $Q^2 > 0,5$  para considerar predictiva la función seleccionada) y un test de aleatoriedad, el cual consiste en intercambiar aleatoriamente el valor de la propiedad entre los compuestos y regresionar nuevamente utilizando el mismo modelo que había sido seleccionado previamente y calculando los valores de  $R^2$  y  $Q^2$ ; repitiendo el proceso un número determinado de veces.

#### *Aplicación del modelo QSAR obtenido a moléculas análogas y predicción de su actividad*

El modelo QSAR obtenido a partir del análisis discriminante lineal fue aplicado a nuevas moléculas, análogas a algunas de las carboxamidas utilizadas en los análisis anteriores, con el fin de evaluar su actividad potencial como repelentes contra *Aedes aegypti*. Para ello, se calcularon los índices topológicos de cada una de ellas con ayuda del programa Desmol2 y a continuación, se aplicó la función discriminante seleccionada anteriormente. Tras ello, para aquellas moléculas que fueron categorizadas como potencialmente activas, se les predijo su actividad biológica como repelentes mediante el uso del modelo obtenido del análisis de regresión multilínea seleccionado con anterioridad.

## RESULTADOS Y DISCUSIÓN

Varias de las estructuras presentes en los compuestos seleccionados para el estudio pueden encontrarse en compuestos naturales, como es el caso de los compuestos 5q-5w y C40-C41 (benzamidas), que pueden considerarse análogos de DEET (ingrediente activo en algunos productos extensamente usados para repeler insectos picadores como mosquitos y garrapatas).

La búsqueda de un modelo matemático-topológico útil para predecir la actividad repelente se realizó en dos pasos:

- I. En primer lugar, la selección de una función discriminante capaz de distinguir entre compuestos activos e inactivos en relación con su actividad repelente.
- II. En segundo lugar, la obtención de una función de regresión topológica capaz de medir la potencia de dicha actividad en términos de pMED.



Ambas funciones se comprenden en el marco del modelo matemático que permitiría la búsqueda y selección de nuevos compuestos potencialmente repelentes.

Para obtener la función discriminante, se aplicó el LDA a un conjunto de entrenamiento formado por 43 compuestos. El conjunto de entrenamiento incluía dos subconjuntos, calificados en compuestos activos (grupo A) e inactivos (grupo I), Tabla 3.

La función discriminante lineal seleccionada fue:

$$FD = -19,95 - 6,293 \times {}^3\chi_c^v + 157,59 \times D^1\chi - 3,15 \times L - 1,57 \times V3 \quad \text{Ec. 1}$$

N = 43 Lambda de Wilks = 0,21 F (4,38) = 36,27 p<0,00001

En la función discriminante FD intervienen los descriptores topológicos  ${}^3\chi_c^v$ ,  $D^1\chi$ , L y V3 que evalúan la ramificación, ( ${}^3\chi_c^v$  y V3) el volumen molecular,  $D^1\chi$  y la longitud molecular, L.

Los compuestos con  $FD > 0$  se corresponden con aquellos que tienen actividad como repelentes frente a *Aedes aegypti* (grupo A); mientras que los compuestos con  $FD < 0$ , son considerados con inactividad como repelentes frente al mismo vector (Grupo I). Los resultados obtenidos en el análisis discriminante lineal se muestran en la Tabla 3.

La matriz de clasificación es muy significativa para el conjunto de entrenamiento considerado, obteniendo unos resultados de 100% de sensibilidad (clasificación de los compuestos activos correctamente) y un 84,62% de especificidad (clasificación de los compuestos inactivos correctamente).

Tabla 3. Resultados obtenidos del análisis discriminante lineal y la regresión multilínea para el grupo de entrenamiento de repelentes frente a *Aedes aegypti*.

Comp.	MED exp <sup>a</sup> ( $\mu\text{mol}/\text{cm}^2$ )	pMED <sub>exp</sub>	Clas(exp)	Prob. Activ. <sup>b</sup>	FD <sup>c</sup>	Clas(Pred)	pMED calc <sup>d</sup>
5a	0,12	6,93	Grupo A	0,99	14,36	Grupo A	7,02
5a'	0,14	6,85	Grupo A	1,00	14,80	Grupo A	6,67
5b	0,16	6,81	Grupo A	0,99	8,64	Grupo A	7,10
5b'	10,75	4,97	Grupo I	0,00	-6,02	Grupo I	5,65
5c	0,20	6,71	Grupo A	0,99	8,66	Grupo A	6,56
5c'	20,13	4,70	Grupo I	0,00	-11,13	Grupo I	4,62
5d	0,03	7,48	Grupo A	0,99	11,80	Grupo A	6,88
5d'	20,25	4,69	Grupo I	0,00	-18,37	Grupo I	4,32
5e	0,27	6,57	Grupo A	0,97	3,73	Grupo A	6,74
5e'	0,27	6,56	Grupo A	0,99	5,58	Grupo A	6,81
5f	0,63	6,20	Grupo I	0,02	-3,71	Grupo I	6,13
5f'	0,63	6,20	Grupo I	0,61	0,47	Grupo I	6,13
5g	0,11	6,98	Grupo A	0,99	9,02	Grupo A	7,03
5g'	0,65	6,19	Grupo I	0,66	0,67	Grupo I	6,47
5h	0,10	6,99	Grupo A	0,99	12,17	Grupo A	6,88
5h'	0,18	6,74	Grupo A	0,99	12,79	Grupo A	6,62
5i	0,13	6,90	Grupo A	0,99	12,55	Grupo A	7,11
5i'	20,50	4,69	Grupo I	0,00	-8,00	Grupo I	5,28



Comp.	MED exp <sup>a</sup> ( $\mu\text{mol}/\text{cm}^2$ )	pMED <sub>exp</sub>	Clas(exp)	Prob. Activ. <sup>b</sup>	FD <sup>c</sup>	Clas(Pred)	pMED calc <sup>d</sup>
5j	0,38	6,43	Grupo A	0,99	14,13	Grupo A	6,94
5j'	0,13	6,90	Grupo A	0,92	2,54	Grupo A	6,20
5k	0,13	6,90	Grupo A	0,99	10,82	Grupo A	6,88
5k'	0,17	6,76	Grupo A	0,99	6,35	Grupo A	6,56
5l	0,41	6,39	Grupo A	0,99	5,70	Grupo A	6,31
5l'	0,14	6,85	Grupo A	0,98	4,53	Grupo A	6,67
5m	0,17	6,76	Grupo A	0,99	5,90	Grupo A	6,63
5n	0,29	6,54	Grupo A	0,99	10,37	Grupo A	6,69
5o	0,47	6,33	Grupo A	0,99	11,95	Grupo A	6,53
5p	0,31	6,50	Grupo A	0,99	13,52	Grupo A	6,49
5q	0,16	6,81	Grupo A	0,98	4,28	Grupo A	6,32
5r	0,12	6,93	Grupo A	0,99	9,44	Grupo A	7,00
5s	0,18	6,74	Grupo A	0,99	11,02	Grupo A	6,83
5t	0,10	7,01	Grupo A	0,99	12,60	Grupo A	6,84
5u	0,42	6,38	Grupo A	0,99	9,46	Grupo A	6,50
5v	0,15	6,84	Grupo A	0,99	5,87	Grupo A	6,16
5w	5,16	5,29	Grupo I	0,00	-8,07	Grupo I	5,32
5x	0,19	6,72	Grupo A	0,99	11,64	Grupo A	6,89
5y	0,31	6,50	Grupo A	0,99	13,22	Grupo A	6,69
5z	0,22	6,66	Grupo A	0,99	6,53	Grupo A	6,31
C39	25,00	4,60	Grupo I	0,01	-4,26	Grupo I	5,14
C40	25,00	4,60	Grupo I	0,00	-11,82	Grupo I	4,61
C41	25,00	4,60	Grupo I	0,00	-8,46	Grupo I	4,74
C42	25,00	4,60	Grupo I	0,00	-11,70	Grupo I	5,32
C43	25,00	4,60	Grupo I	0,00	-10,90	Grupo I	3,67

<sup>a</sup> Valores de MED experimental tomados del trabajo de Oliferenko et al., 2013 [4]; <sup>b</sup> probabilidad de que el compuesto sea activo; <sup>c</sup> valores de FD obtenidos con la Ec. 1; <sup>d</sup> valores de la actividad predichos con la Ec. 2.

La Figura 1 muestra el diagrama de distribución de la actividad repelente obtenida con la función FD (las barras blancas y negras representan conjuntos inactivos y activos, respectivamente). Mediante observación del diagrama, el dominio de aplicación de la FD para la búsqueda de potenciales compuestos repelentes activos frente a *Aedes aegypti* se da para valores de FD comprendidos entre 0 y 14.

Con la finalidad de validar la función FD seleccionada, se llevó a cabo una crovalidación interna tipo “leave-one-out”, para la cual, se eliminó un caso del conjunto de datos y se repitió el análisis discriminante lineal con los casos restantes (N-1). Los resultados obtenidos confirmaron la validez de la prueba al clasificar correctamente a los compuestos de estudio, observando que el valor de actividad correspondiente al caso retirado se correspondía con el predicho. El proceso se repitió tantas veces como casos presenta la data.



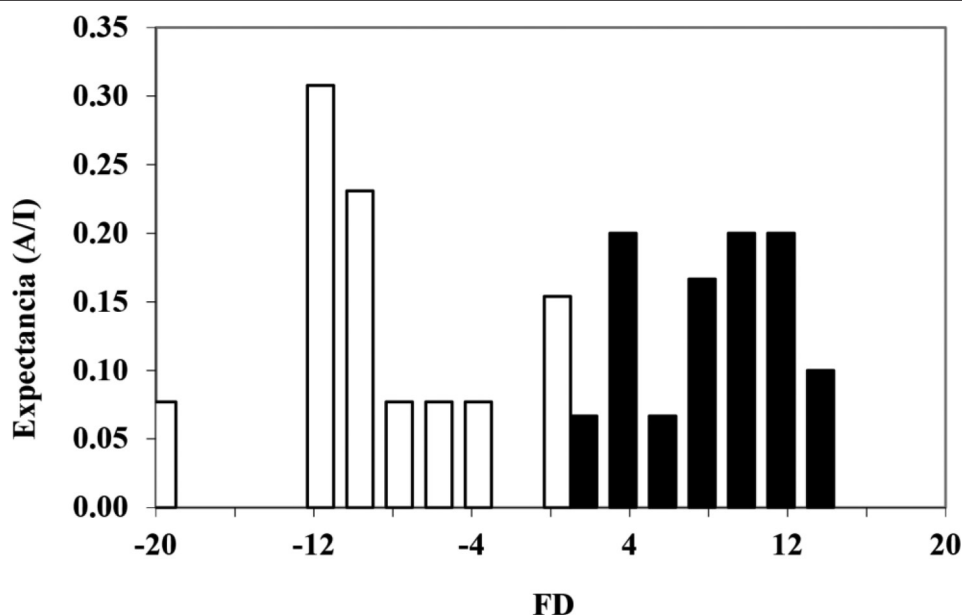


Figura 1. Diagrama de distribución de actividad repelente frente a *Aedes aegypti* obtenida al graficar la expectancia, E, frente a la función discriminante, FD. Las barras de color negro representan al grupo activo y las blancas al inactivo.

Para predecir el valor teórico de pMED de cada compuesto, se realizó un análisis de regresión multilínea utilizando los descriptores topológicos como variables independientes y pMED como dependiente. En este análisis se utilizó el mismo conjunto de entrenamiento que en LDA. Los resultados obtenidos se muestran en la Tabla 3, Fig.2. y Fig. 3.

La función de regresión multilínea seleccionada fue:

$$pMED = 96,11 - 0,58 \times {}^3\chi_c^v - 83,68 \times C^0\chi - 0,004 \times W - 0,103 \times V3 \quad \text{Ec. 2}$$

$$N=43 \quad R^2 = 0,88 \quad Q^2 = 0,84 \quad EEE= 0,32 \quad F(4,38) = 71,64 \quad p < 0,0001$$

En la función de regression, Ec. 2, intervienen los descriptores topológicos  ${}^3\chi_c^v$ ,  $C^0\chi$ , W (índice de Wiener) y V3.

La función seleccionada presenta un valor de  $R^2$  superior a 0,80 (0, 88) lo que explica más del 88% de la varianza.

Para validar la función de regresión multilínea seleccionada, se llevó a cabo una crovalidación interna tipo “leave-one-out” con la determinación del coeficiente de predicción  $Q^2$ . El valor de dicho coeficiente fué de  $Q^2=0,84$ , lo que indica una alta capacidad predictiva de la ecuación.

Por otro lado, también se llevó a cabo un test de aleatoriedad, que consiste en intercambiar aleatoriamente los valores de la propiedad entre los compuestos y regresionar nuevamente utilizando los mismos descriptores topológicos que fueron seleccionados. El proceso se repite varias veces. Los resultados obtenidos se muestran en la Fig. 4. Como puede observarse, la función de predicción seleccionada es altamente estable. En todos los casos, los valores de  $R^2$  y  $Q^2$  son muy bajos, excepto cuando a cada compuesto se le asigna el valor real que le corresponde de la propiedad pMED.



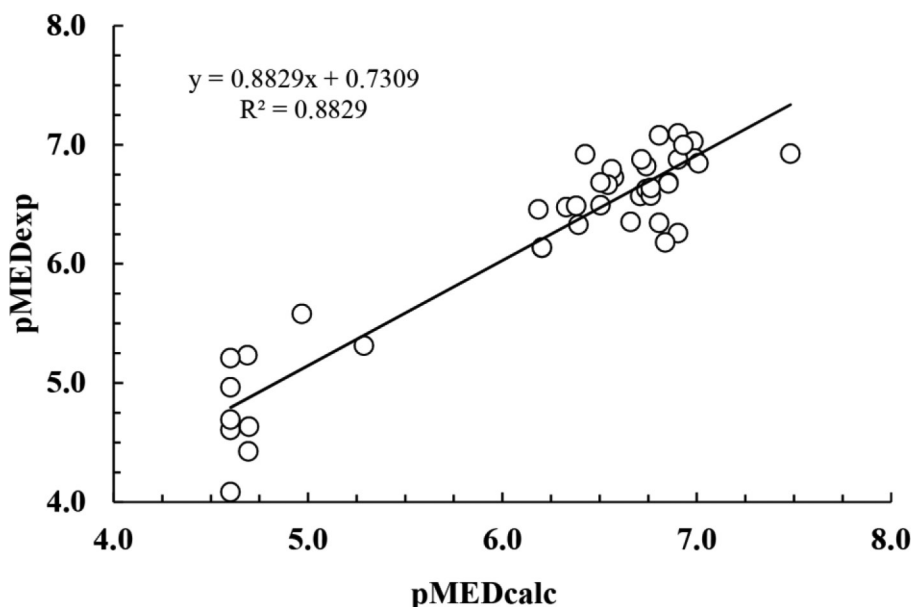


Figura 2. Representación gráfica de pMED experimental frente a pMED calculada obtenida a partir de la función de predicción seleccionada, Ec. 2.

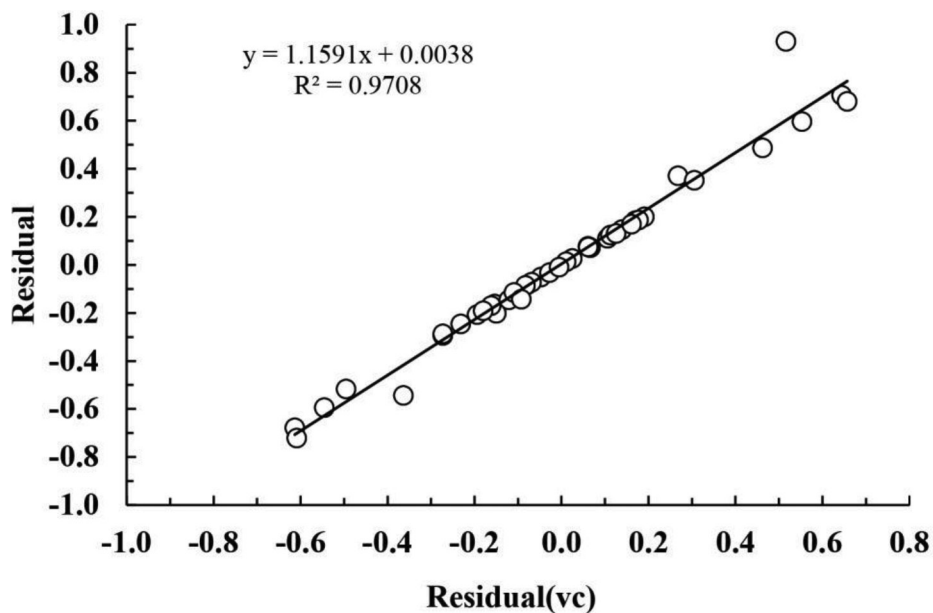


Figura 3. Representación gráfica de los valores residuales resultantes de la resta del pMED experimental y el calculado con los residuales de la validación *leave one out*.



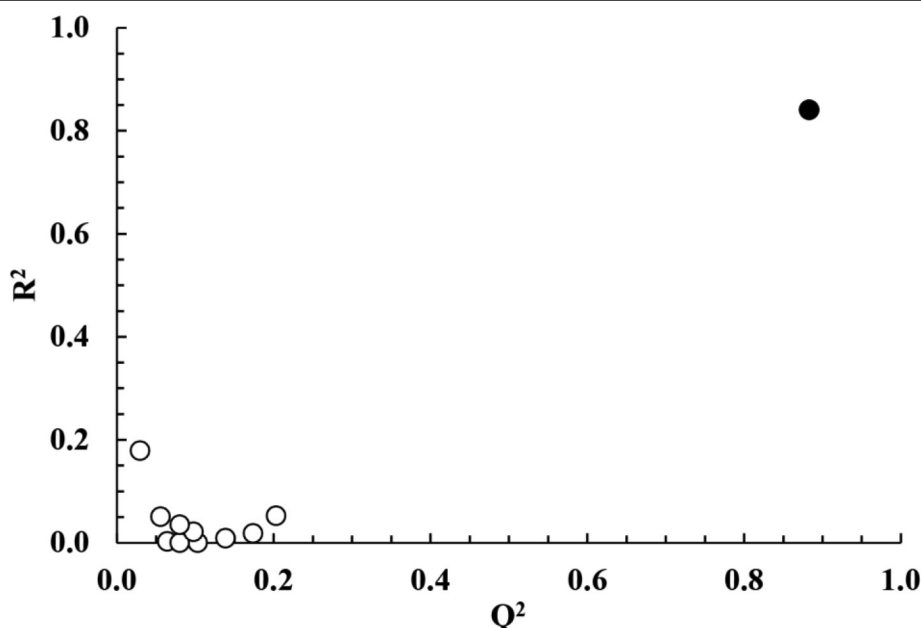


Figura 4. Representación gráfica del test de aleatoriedad realizado a la función de predicción, Ec. 2

Sobre los resultados obtenidos del LDA y del análisis de regresión multilínea (Tabla 3) se puede configurar un modelo topológico para la búsqueda de nuevos agentes repelentes. La búsqueda se realiza sobre moléculas carboxamidas de acuerdo con los siguientes requisitos: si  $14 > FD > 0$  entonces el compuesto se etiqueta como potencial activo (grupo A); de lo contrario, el compuesto se clasifica como inactivo (grupo I).

Para el conjunto de prueba, el grado de éxito promedio es del 95,4% de exactitud (precisión). La alta calidad estadística del modelo QSAR dado para este conjunto de datos prueba su capacidad discriminante.

Por otro lado, se llevó a cabo el estudio de la actividad repelente de algunos compuestos análogos a algunas de las carboxamidas estudiadas. Para ello, se procedió a la clasificación de su actividad potencial como repelentes mediante el empleo de la función discriminante lineal obtenida con anterioridad.

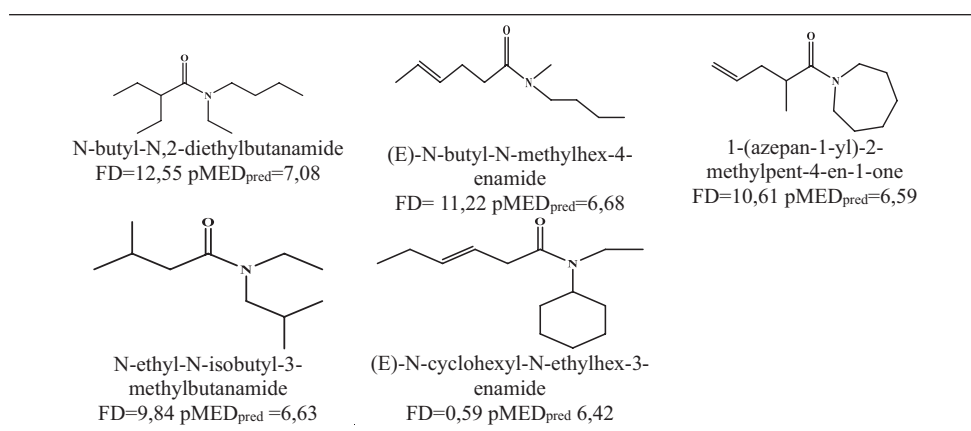
Posteriormente, empleando la función de predicción del modelo de regresión multilínea, se procedió también al cálculo de la pMED de los compuestos activos para observar su nivel de efectividad. Los resultados obtenidos se muestran en la Tabla 4.

Todos los compuestos presentan valores de  $pMED_{pred} > 6,30$  lo que corresponde con una concentración  $MED < 0,5 \mu\text{mol}/\text{cm}^2$ . Destaca el compuesto N-butyl-N,2-diethylbutanamide con una potencial actividad repelente de  $MED=0.08 \mu\text{mol}/\text{cm}^2$ .

Estos sugerentes resultados deben ser corroborados con los correspondientes ensayos de actividad repelente, los cuales, deben permitir la validación o evaluación del modelo propuesto y servir como herramienta útil para la búsqueda de nuevos compuestos con mayor actividad frente al vector *Aedes aegypti*.



Tabla 4. Identificación química de algunos de los compuestos análogos a las carboxamidas, seleccionados por el modelo topológico desarrollado, con potencial actividad repelente frente a *Aedes aegypti*



## CONCLUSIONES

La topología molecular se ha utilizado con éxito para encontrar un modelo QSAR para predecir la actividad repelente de un grupo de 43 moléculas carboxamidas con una actividad potencial como repelentes para el vector *Aedes aegypti*. Todos los descriptores moleculares utilizados son grafo-téoricos. El modelo matemático empleado en este trabajo conserva las principales características estructurales que involucran a la propiedad correlacionada y, por tanto, puede ser aplicable, tras las correspondientes comprobaciones empíricas, a la búsqueda de nuevos compuestos activos mediante exploración virtual en bases de datos con el fin de seleccionar compuestos con actividad potencial como repelentes que podrían ser utilizados de forma efectiva frente a este vector.

## AGRADECIMIENTOS

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

## REFERENCIAS BIBLIOGRÁFICAS

- [1] Número V, Agosto M, El EN, San D, Molinos JDELOS. LEVEL OF KNOWLEDGE IN THE PREVENTION OF TRANSMISSIBLE DISEASES BY AEGYPTI. 2021;(December 2018).
- [2] Gómez Garcia GF. *Aedes (Stegomyia) aegypti* (diptera: Culicidae) and its importance for human health. Rev Cubana Med Trop. 2018;70(1):55–70.



- [3] Montero G, Fca B. Biología de *Aedes aegypti*. *WwwProduccion-AnimalComAr* [Internet]. 2009;1–4. Available from: [http://www.produccion-animal.com.ar/fauna/79-Aedes\\_aegypti.pdf](http://www.produccion-animal.com.ar/fauna/79-Aedes_aegypti.pdf)
- [4] Oliferenko P V., Oliferenko AA, Poda GI, Osolodkin DI, Pillai GG, Bernier UR, et al. Promising *Aedes aegypti* Repellent Chemotypes Identified through Integrated QSAR, Virtual Screening, Synthesis, and Bioassay. *PLoS One*. 2013;8(9):1–13.
- [5] Kier LB, Hall LH. Molecular connectivity VII: Specific treatment of heteroatoms. *J Pharm Sci*. 1976;65(12):1806–9.
- [6] García-Domenech 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.
- [7] De Gregorio C, Kier LB, Hall LH. QSAR modeling with the electrotopological state indices: Corticosteroids. *J Comput Aided Mol Des*. 1998;12(6):557–61.
- [8] Duart MJ, Antón-Fos GM, De Julian-Ortiz J V., Gozalbes R, Gálvez J, García-Domenech R. Use of molecular topology for the prediction of physicochemical, pharmacokinetic and toxicological properties of a group of antihistaminic drugs. *Int J Pharm*. 2002;246(1–2):111–9.
- [9] García-Domenech R, Gálvez J, de Julián-Ortiz J V., Pogliani L. Some new trends in chemical graph theory. *Chem Rev*. 2008;108(3):1127–69.
- [10] Hosoya H, Gotoh M, Murakami M, Ikeda S. Topological index and thermodynamic properties. 5. How can we explain the topological dependency of thermodynamic properties of alkanes with the topology of graphs? *J Chem Inf Comput Sci*. 1999;39(2):192–6.
- [11] Ivanciuc O, Ivanciuc T, Balaban AT. Quantitative structure-property relationship study of normal boiling points for halogen-/oxygen-/sulfur-containing organic compounds using the CODESSA program. *Tetrahedron*. 1998;54(31):9129–42.
- [12] Gálvez J, García-Domenech 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.
- [13] Kier LB, Hall LH. General definition of valence delta-values for molecular connectivity. *J Pharm Sci*. 1983;72(10):1170–3.
- [14] Gálvez J, Garcia R, Salabert MT, Soler R. Charge Indexes. New Topological Descriptors. *J Chem Inf Comput Sci*. 1994;34(3):520–5.
- [15] Besalú E. Fast computation of cross-validated properties in full linear leave-many-out procedures. *J Math Chem*. 2001;29(3):191–204.

