

Aplicación de la topología molecular para la predicción de la actividad anti-VIH-1 de un grupo de compuestos análogos del aciclovir y ganciclovir

Ramón García-Doménech^{*}, Mayra C. Montealegre, Edwige G. Nagham, Nancy Sandoval, Manuel Santana, Jorge Gálvez

^{*} Departamento de Química Física, Facultad de Farmacia, Universitat de Valencia

Avda. V. A. Estellés, s/n, 46100-Burjassot, Valencia, Spain.

Recibido el 27 de julio de 2009

RESUMEN

Usando el análisis de regresión multilíneal, se ha propuesto un modelo topológico-matemático de predicción de la actividad anti-VIH-1 para un grupo de compuestos análogos al aciclovir y ganciclovir. La caracterización estructural de cada compuesto se ha conseguido a través de índices topológicos. El modelo seleccionado para la predicción de la concentración eficaz 50 está formado por cuatro variables ($R^2 = 0.8762$ y $Q^2 = 0.8216$). La validación del modelo se ha hecho a través de los test de crovalidación, validación interna y aleatoriedad. Los resultados obtenidos confirman la capacidad de predicción de la propiedad estudiada.

Palabras clave: Topología molecular; Análisis de regresión multilíneal; Relaciones cuantitativas estructura-actividad; Actividad anti-VIH-1; Análogos del aciclovir y ganciclovir.

ABSTRACT**Application of the molecular topology to the prediction of anti-HIV-1 activity by acyclovir and ganciclovir analogs**

By using multilinear regression analysis, a topological-mathematical model has been built to predict the anti-HIV-1 activity for a group of acyclovir and ganciclovir analogs. The structural depiction was performed by using topological indices and one model with four variables for the prediction of EC_{50} ($R^2 = 0.8762$ and $Q^2 = 0.8216$) was selected. The model was checked by cross-validation, internal validation and randomization test. The results confirm its capability to predict the property analyzed.

Keywords: Molecular topology; Multilinear regression analysis; Quantitative structure-activity relationships; anti-HIV-1 activity; acyclovir and ganciclovir analogs.

1. INTRODUCCIÓN

El SIDA es la enfermedad que se desarrolla como consecuencia de la destrucción progresiva del sistema inmunitario (de las defensas del organismo), producida por un virus descubierto en 1983 y denominado Virus de la Inmunodeficiencia Humana (VIH).

La palabra SIDA proviene de las iniciales de Síndrome de Inmunodeficiencia Adquirida, que consiste en la incapacidad del sistema inmunitario para hacer frente a las infecciones y otros procesos patológicos. Existen dos subtipos de virus de inmunodeficiencia, el VIH-1 y el VIH-2.

En la actualidad el SIDA sigue siendo un gran problema de salud pública con una prevalencia global de 33,2 millones de casos y una incidencia de 2,5 millones de casos, de acuerdo con los datos más recientes en 2007 de ONUSIDA/OMS sobre la pandemia (1).

Existen alrededor de quince fármacos que se están utilizando en el tratamiento de la infección por el VIH. El tratamiento incluye la combinación de varios fármacos anti-retrovirales que evitan el deterioro inmunológico y suprimen la replicación viral (2).

La administración de estos fármacos anti-retrovirales ha permitido alargar considerablemente la supervivencia de los sujetos seropositivos, de manera que la enfermedad se ha convertido en un proceso crónico. La terapia anti-retroviral es compleja y supone la administración de al menos tres fármacos (triple terapia) con un elevado número de tomas y de comprimidos por día, lo que acarrea la existencia de efectos adversos e interacciones con otros fármacos. Todo ello hace que se siga investigando con el objetivo de encontrar nuevos fármacos más seguros y eficaces.

El aciclovir y el ganciclovir son dos antivíricos muy efectivos frente al virus del herpes simples y de la varicela zoster. Estos compuestos también han mostrado cierta actividad como agentes anti-VIH-1. Ante la posibilidad de poder constituir una nueva diana terapéutica, Almerico *et al.* han investigado recientemente la actividad anti-VIH-1 para un grupo de compuestos sintetizados análogos al aciclovir y ganciclovir, realizando posteriormente estudios encaminados a buscar las correspondientes relaciones estructura-actividad (3).

Para la búsqueda de modelos de predicción, la topología molecular ha demostrado ser una herramienta útil y eficaz. A través de ella, cada compuesto viene identificado por un conjunto de descriptores topológicos, los cuales caracterizan estructuralmente a esa molécula y pueden utilizarse en la búsqueda de modelos topológicos-matemáticos para la selección y diseño de nuevos compuestos activos en diferentes áreas terapéuticas (4-10).

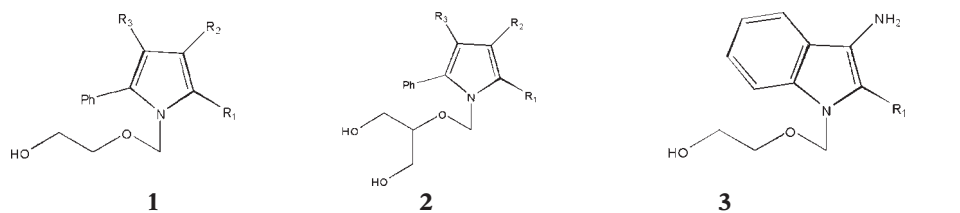
En este trabajo se estudia la actividad anti-VIH-1 de un grupo de compuestos análogos al aciclovir y ganciclovir para la búsqueda de un modelo de predicción de dicha actividad usando la topología molecular y el análisis de regresión multilínea.

2. MATERIALES Y MÉTODOS

2.1. Compuestos y actividad biológica estudiada

Para este estudio se seleccionaron un grupo de 30 compuestos derivados del aciclovir y ganciclovir, cuyas estructuras aparecen reflejadas en la Tabla 1. Los valores experimentales de actividad para cada compuesto vienen cuantificados a través de la concentración eficaz 50, CE_{50} expresada en unidades de concentración micromolar (3).

Tabla 1. Estructura química de cada uno de los compuestos estudiados, así como los valores de actividad [CE_{50} (μM)] experimental, su transformación logarítmica, $pCE_{50} = \text{Log}1/CE_{50}$, y valores calculados a partir del modelo topológico seleccionado, $pCE_{50}\text{calc}$, y a través de la crovalidación, $pCE_{50}(\text{cv})^c$



compuesto	R ₁	R ₂	R ₃	EC50 ^a (μM)	$pEC_{50}\text{exp}$	$pEC_{50}\text{calc}^b$	$pEC_{50}(\text{cv})^c$
1a	Me	COOEt	Cl	11.1	4.954	4.847	4.806
1b	Me	COOEt	NO ₂	33.3	4.477	4.657	4.705
1c	Ph	H	NO ₂	4.5	5.342	5.273	5.259
1d	Ph	Br	Br	5.3	5.279	5.297	5.300
1e	Ph	COOEt	Br	6.3	5.204	5.054	5.046
1f	Ph	COOEt	NO ₂	16.7	4.778	4.707	4.696
1g	Ph	CN	NO ₂	33.3	4.477	4.580	4.589
1h	Ph	Ph	Br	1.5	5.819	5.635	5.600
1i	Ph	Ph	NO ₂	3.7	5.431	5.472	5.479
1k	Ph	CN	NH ₂	33.3	4.477	4.573	4.589
1l	Ph	Ph	NH ₂	4.2	5.380	5.330	5.326
1m	Ph	COOH	NO ₂	33.3	4.477	4.506	4.510
1n	Ph	PH	N ₃	6.7	5.176	5.266	5.289
2a	Ph	Ph	Br	3.0	5.518	5.709	5.762
2b	Ph	Ph	NO ₂	4.5	5.342	5.526	5.559
2c	Ph	CN	Br	20.0	4.699	4.931	4.958
2d	Ph	CN	NO ₂	20.0	4.699	4.617	4.610
2e	Ph	COOEt	Br	12.5	4.903	5.107	5.129
2f	Ph	COOEt	NO ₂	16.7	4.778	4.751	4.746
2g	Me	COOEt	NO ₂	20.0	4.699	4.642	4.635
2h	Ph	CN	NH ₂	20.0	4.699	4.623	4.609
2i	Ph	Ph	NH ₂	1.5	5.819	5.404	5.344
2j	Ph	COOEt	NH ₂	20.0	4.699	4.797	4.809
2k	Ph	COOEt	NH ₂	20.0	4.699	4.700	4.701
2l	Ph	Ph	N ₃	5.0	5.301	5.310	5.312
2m	Ph	CN	N ₃	20.0	4.699	4.600	4.591
2n	Ph	COOEt	N ₃	20.0	4.699	4.644	4.626
2o	Me	COOEt	N ₃	20.0	4.699	4.602	4.577
3h	Ph			11.1	4.954	5.103	5.227
3i	2-Pyr			5.0	5.301	5.217	5.166

^a obtenidos de la referencia (3)

^b obtenidos a partir de la Eq. 1.

^c obtenidos del análisis de crovalidación realizado con la Eq. 1.

2.2. Descriptores topológicos utilizados

La caracterización topológica de cada molécula se realiza a través del correspondiente grafo molecular, 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.

En este trabajo se han utilizado un grupo de descriptores bien conocidos tales como: índices de conectividad tipo Randić-Kier-Hall hasta orden cuarto (${}^m\chi_t$, ${}^m\chi_t^v$) (11, 12), índices de carga topológica, ICT, hasta orden quinto (J_m , G_m , J_m^v , G_m^v) (13), cocientes y diferencias entre índices de conectividad de valencia y de no valencia (${}^mC_t = {}^m\chi_t / {}^m\chi_t^v$ and ${}^mD_t = {}^m\chi_t - {}^m\chi_t^v$), así como el índice de Wiener (14) y otros índices tales como N (número de átomos en la molécula diferentes al hidrógeno), PRn (pares de ramificaciones separados por una distancia n), Vn (número de vértices con valencia n) y la longitud topológica de la molécula L (distancia topológica mínima que separan a los átomos más alejados entre sí). En total, cada compuesto fue caracterizado con una serie de 62 índices. Todos los descriptores usados en este trabajo han sido obtenidos a través del programa Desmol11 (disponible vía e-mail) (15).

2.3. Algoritmo QSAR: Análisis de regresión multilíneal

El estudio de regresión multilíneal se ha realizado con ayuda del programa BMDP 9R (16), utilizando los índices topológicos como variables independientes y la concentración eficaz 50, CE_{50} , como variable dependiente.

El estudio estadístico de la función seleccionada se hace atendiendo al número de casos utilizados N, el porcentaje de la varianza explicada R^2 , el error estándar de estimación, EEE, el valor F de Fisher-Snedecor y la significación estadística p.

La calidad predictiva y la robustez del modelo de predicción seleccionado debe de someterse a un test de validación. Las estrate-

gias que se suelen adoptar son (17): *a*) Crosvalidación; *b*) validación interna dividiendo la serie de compuestos estudiada en un grupo de entrenamiento y un grupo test; *c*) validación externa con un grupo test no usado en la búsqueda de la función de predicción; *d*) análisis de aleatoriedad y estabilidad. En este trabajo utilizaremos las estrategias (a, b y d) como criterio de validación.

3. RESULTADOS Y DISCUSIÓN

Partiendo del trabajo desarrollado recientemente por Almerico *et al.* (3) en el que realizaron un estudio QSAR con un grupo formado por 30 compuestos análogos al aciclovir y ganciclovir y sus correspondientes datos de actividad anti-VIH-1, se utilizó el análisis de regresión multilíneal para la búsqueda del modelo de predicción de actividad usando la concentración eficaz 50 en su transformación logarítmica, $pCE_{50} = \text{Log}1/CE_{50}$ como variable dependiente y el con-

Tabla 2. Análisis de regresión multilíneal realizado con pCE_{50} y diferentes números de variables

Variables	r^2	SEE
J_3^v	0.5558	0.2677
$^1\chi^v, G_2$	0.7818	0.1910
$2\chi^v, G_2, L$	0.8324	0.1704
$^1\chi^v, G_2, G_5^v, J_4^v$	0.8762	0.1495
$J_2, J_4^v, J_5^v, ^4D_{pc}, L$	0.8884	0.1449

Ecuación de predicción seleccionada

	Coefficiente	Std. Err.	p
Intercept	3.807	0.445	0.0000
$^1\chi^v$	0.437	0.044	0.0000
G_2	-0.316	0.042	0.0000
G_5^v	-0.433	0.105	0.0004
J_4^v	15.020	3.744	0.0005
N = 30	$R^2 = 0.8762$	$Q^2 = 0.8216$	
$F(4,25) = 44.3$	$p < 0.00000$	$EEE = 0.14493$	

junto de 62 descriptores calculados para cada compuesto como variables independientes.

La Tabla 2 recoge los resultados estadísticos obtenidos en función del número de variables seleccionadas. La agrupación de variables que mejor correlaciona la actividad anti-VIH-1 es la formada por cuatro descriptores ($R^2 = 0.8762$ y $EEE = 0.145$). Todos ellos son estadísticamente significativos con valores de $p \leq 0.0005$. La incorporación de nuevas variables apenas modifican los parámetros anteriores.

La ecuación seleccionada fue:

$$pCE_{50} = 3.807 + 0.437 \ ^1\chi^v - 0.316G_2 - 0.433G_5^v + 15.020J_4^v \quad \text{Eq. 1}$$

$$N = 30 \quad R^2 = 0.8762 \quad EEE = 0.145 \quad Q^2 = 0.8216 \quad F(4,25) = 44.3 \\ p < 0.00001$$

Los índices topológicos que aparecen en la ecuación evalúan por una parte las transferencias de carga intramolecular, G_2 , G_5^v y J_4^v , así como la ramificación y el ensamblaje topológico, $^1\chi^v$.

El modelo seleccionado es capaz de explicar más del 87% de la varianza de la propiedad correlacionada ($R^2 = 0.8762$) con un error estándar de estimación inferior al 11% de la variabilidad en la que se mueve la propiedad ($EEE = 0.145$).

La Tabla 1, columna 7, y las Figuras 1 y 2 muestran los resultados de predicción obtenidos para cada compuesto. Todos ellos, a excepción del compuesto 2i (punto negro marcado en la Figura 2), presentan residuales inferiores a $\pm 2EEE$, lo cual es indicativo de la bondad y calidad de la ecuación seleccionada.

El primer estudio de validación de la Eq. 1 fue una crossvalidación interna. Para ello se elimina un caso del grupo y se realiza el análisis de regresión utilizando los N-1 restantes compuestos y prediciendo el valor de la propiedad del compuesto eliminado. El proceso se repite tantas veces como compuestos forme el grupo. El valor del coeficiente de predicción Q^2 nos informará sobre la calidad de la función seleccionada (Q^2 ha de tomar un valor superior a 0.5 para poder considerar predictiva la función obtenida) (18). El coeficiente de predicción obtenido fue de $Q^2 = 0.8216$, ligeramente

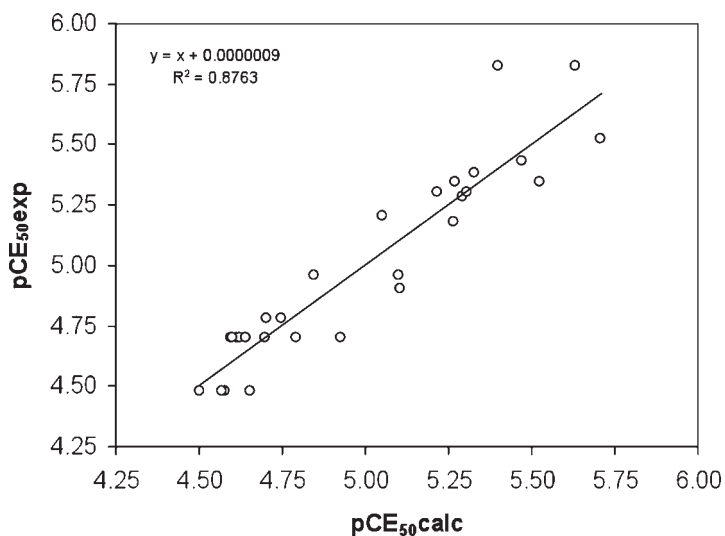


Figura 1. Representación gráfica de pCE_{50} experimental y pCE_{50} calculado para cada compuesto a partir de la función topológica de predicción seleccionada.

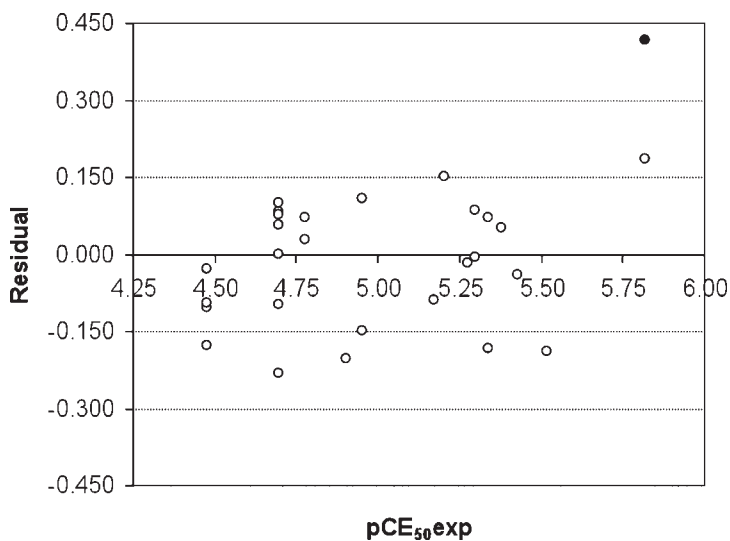


Figura 2. Representación gráfica de los residuales ($pCE_{50}exp - pCE_{50}calc$) frente a pCE_{50} experimental obtenidos para cada compuesto a partir de la función topológica de predicción seleccionada, Eq. 1.

inferior a la varianza de la ecuación seleccionada ($R^2 = 0.8762$) y muy por encima del valor mínimo exigido ($Q^2 > 0.5000$).

Los valores de pCE50 predichos en la crosvalidación para cada compuesto aparecen en la columna 8 de la Tabla 1. Los resultados son similares a los mostrados en la columna 7.

En una segunda etapa se realizó una validación interna dividiendo el grupo de compuestos estudiado en cinco subgrupos (A-E) conteniendo cada subgrupo aproximadamente el 20% de los compuestos analizados: serie A (1d, 1l, 2a, 2f, 2k, 3h); serie B (1e, 1k, 2b, 2g, 2l, 3i); serie C (1c, 1h, 1n, 2e, 2j, 2o); serie D (1b, 1g, 1m, 2d, 2i, 2n) y serie E (1a, 1f, 1l, 2c, 2h, 2m). Cuatro de los cinco subgrupos (A, B, C y D), (A, B, C y E), (A, B, D y E), (A, C, D y E) y (B, C, D y E) se utilizarán como grupo de entrenamiento, mientras que el subgrupo restante hará de grupo test. La ecuación de regresión obtenida con cada grupo de entrenamiento, usando los mismos descriptores, se utiliza para predecir la actividad de los compuestos del correspondiente grupo test.

La Tabla 3 muestra los valores de R^2 y R^2_{pred} para cada grupo de entrenamiento y test analizado. La variabilidad de R^2 y R^2_{pred} es pequeña para cada análisis y el valor promedio obtenido tanto para R^2 como para R^2_{pred} ($R^2 = 0.8813$ y $R^2_{\text{pred}} = 0.8676$) es similar al de la función seleccionada, $R^2 = 0.8762$.

En una tercera etapa se realizó un análisis de aleatoriedad de la función topológica seleccionada. En este caso es aconsejable hacer-

Tabla 3. **Resultados del estudio de validación interna realizado al modelo de predicción seleccionado.**

Series entrenamiento	Serie test	R^2 (entrenamiento)	R^2_{pred} (test)
A, B, C, D y E	—	0,8762	—
B, C, D y E	A	0,8770	0,9386
A, C, D y E	B	0,8740	0,8814
A, B, D y E	C	0,8739	0,8569
A, B, C, y E	D	0,9016	0,8892
A, B, C, y D	E	0,8853	0,7720
Promedio		0,8813	0,8676

lo, ya que la propiedad correlacionada se mueve en un estrecho margen de valores ($pCE_{50\max} = 5.819$ y $pCE_{50\min} = 4.477$). El análisis consiste en intercambiar aleatoriamente el valor de pCE_{50} entre los compuestos y regresionar utilizando los mismos descriptores de la Eq. 1. El proceso se repite varias veces (diez en nuestro caso), calculándose para cada uno los valores de R^2 y Q^2 .

La Figura 3 recoge los resultados obtenidos. Como era de esperar, la regresión obtenida en todos los casos, excepto cuando cada compuesto tiene asignado su valor real de pEC_{50} , es muy baja.

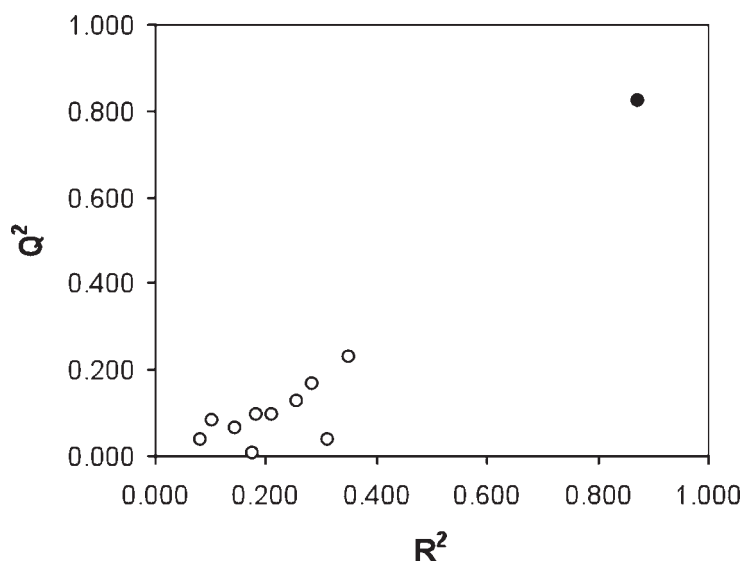


Figura 3. Estudio de aleatoriedad realizado con la función de predicción seleccionada.

Los resultados obtenidos en los tres test de validación realizados nos lleva a concluir que la función topológica de predicción de la actividad anti-VIH-1 para el grupo de compuestos estudiado es robusta y estable, pudiéndose utilizar en la predicción de esta propiedad para otros compuestos estructuralmente análogos a los estudiados.

Por último indicar que si comparamos los resultados obtenidos por nosotros con los de Almerico *et al.* (utilizando parámetros fisi-

coquímicos y topológicos obtenidos con los programas TSAR 3.2, VAMP 6.0 y ASP 3.2) (3) se observa una significativa diferencia: La mejor regresión obtenida por Almerico *et al.* fue usando seis variables ($R^2 = 0.782$, $EEE = 0.20$ y $Q^2 = 0.667$), mientras que la nuestra, con cuatro variables y utilizando sólo descriptores topológicos, fue superior ($R^2 = 0.8762$, $EEE = 0.145$ y $Q^2 = 0.8216$). Si a ello unimos la sencillez y rapidez de cálculo de los descriptores utilizados nos lleva a concluir que la topología es una herramienta de gran utilidad en la búsqueda de modelos de predicción QSAR.

4. CONCLUSIONES

El principal objetivo de este trabajo ha sido la selección de un modelo matemático de predicción de la actividad anti-VIH-1 para un grupo de compuestos análogos al aciclovir y ganciclovir. Para ello se ha utilizado la topología molecular como herramienta para caracterizar la estructura de cada molécula de una forma rápida y sencilla. A partir del análisis de regresión multilíneal y utilizando sólo descriptores topológicos, se ha propuesto un modelo topológico-matemático que permite la predicción de la actividad trabajando con la concentración eficaz 50. El modelo contiene cuatro variables capaces de explicar el 87,6% de la varianza de la propiedad estudiada.

La calidad de la ecuación de predicción seleccionada ha sido validada a través de una crovalidación, test interno y test de aleatoriedad. Los resultados indican una alta robustez.

5. AGRADECIMIENTOS

Este trabajo ha sido realizado con la ayuda del Master Internacional en Enfermedades Parasitarias Tropicales de la Universitat de Valencia (curso 2008/2009).

6. BIBLIOGRAFÍA

1. <http://www.who.int/mediacentre/news/releases/2007/pr61/es/index.html>.
2. Llibre, J. M.; Falco, V.; Tural, C.; Negro, E.; Pineda, J. A.; Muñoz, J.; Ortega, E.; Videla, S.; Sirera, G.; Martínez, E.; Miralles, C.; Iribarren, J.; Galindo, M. J.; Domingo, P.; D'Arminio-Monforte, A.; Miró, J. M. & Clotet, B. (2009) The changing face of HIV/AIDS in treated patients. *Current HIV Research*. 7(4): 365-77.
3. Almerico, A. M.; Tutone, M. & Lauria, A. (2009) A QSAR study investigating the potential anti-HIV-1 effect of some Acyclovir and Ganciclovir analogs. *ARKIVOC*. (viii): 85-94.
4. Gálvez, J.; García-Doménech, R.; de Julián-Ortiz, J. V. & Soler, R. (1995) Topological Approach to Drug Design. *Journal of Chemical Information and Computer Sciences*. 35(5): 272-284.
5. Gozalbes, R.; Brun-Pascuad, M.; García-Doménech, R.; Gálvez, J.; Girard, P. M.; Ddoucet, J. P. & Derouin, F. (2000) Anti-Toxoplasma activities of 24 uinolones and Fluoroquinolones *in vitro*: Prediction of activity by molecular topology and virtual computational techniques. *Antimicrobial Agents and Chemotherapy*. 44(10): 2771-2776.
6. Julián-Ortiz, J. V.; Gálvez, J.; Muñoz-Collado, C.; García-Doménech, R. & Gimeno-Cardona, C. (1999) Virtual combinatorial syntheses and computational screening of new potential anti-herpes compounds. *Journal of Medicinal Chemistry*. 42(17): 3308-3314.
7. Duart, M. J.; Antón-Fos, G. M.; Alemán, P. A.; Gay-Roig, J. B.; González-Rosende, M. E.; Gálvez, J. & García-Doménech, R. (2005) New potential antihistaminic compounds. Virtual combinatorial chemistry, computational screening, real synthesis, and pharmacological evaluation. *Journal of Medicinal Chemistry*. 48(4): 1260-1264.
8. García-Doménech, R.; Gálvez, J.; Julián-Ortiz, J. V. & Pogliani, L. (2008) Some new trends in chemical graph theory. *Chemical Reviews*. 108(3): 1127-1169.
9. Mahmoudi, N.; García-Doménech, R.; Gálvez, J.; Farhati, M.; Franetich, J.-F.; Sauerwein, R.; Hannoun, L.; Derouin, F.; Danis, M. & Mazier, D. (2008) New active drugs against liver stages of Plasmodium predicted by molecular topology. *Antimicrobial Agents and Chemotherapy*. 52(4): 1215-1220.
10. García-Doménech, R.; Domingo-Puig, C. & Esteve-Martínez, M. A. (2008) Aplicación de la topología molecular en la búsqueda de nuevos agentes activos frente a Leishmania. *Anales de la Real Academia de Farmacia*. 74(3): 345-367.
11. Kier, L. B.; Murray, W. J.; Randic, M. & Hall, L. H. (1975) Molecular connectivity V: connectivity series concept applied to density. *Journal of Pharmaceutical Sciences*. 65(8): 1226-1230.
12. Kier, L. B. & Hall, L. H. (1982) General definition of valence delta-values for molecular connectivity. *Journal of Pharmaceutical Sciences*. 72(10): 1170-1173.
13. Gálvez, J.; García-Doménech, R.; Salabert, M. T. & Soler, R. (1994) New Topological Descriptors. *Journal of Chemical Information and Computer Sciences*. 34(3): 520-525.

14. Wiener, H. J. (1947) Structural determination of paraffin boiling points. *Journal of the American Chemical Society*. 69(1): 17-20.
15. DESMOL11 Software. Unidad de Investigación de Diseño de Fármacos y Conectividad Molecular. Facultad de Farmacia Universitat de València, España.
16. Dixon, W. J. (1990) BMDP Statistical software manual. Vols. I, II. Dixon, W. J. Eds. University of California Press, Berkeley, California.
17. Roy, P. P.; Leonard, J. T. & Roy, K. (2008) Exploring the impact of size of training sets for the development of predictive QSAR models. *Chemometrics and Intelligent Laboratory Systems*. 90(1): 31-42.
18. Allen, D. M. (1974) The relationship between variable selection and data augmentation and a method for prediction. *Technometrics*. 16(1): 125-127.

*** Información de contacto:**

Dr. Ramón García-Doménech.

Dept. Química Física, Facultad de Farmacia, Universitat de Valencia.

Avda. V. A. Estellés, s/n.

46100-Burjassot, Valencia, Spain.

Telf.: +34 963 544 291.

Email: ramon.garcia@uv.es