

Obtención de modelos de relación cuantitativa estructura actividad (QSAR) para la predicción de actividad antibacteriana en series heterogéneas de compuestos

Obtaining quantitative structure-activity relationship (QSAR) models for prediction to antibacterial activity on heterogeneous series of compounds

Lic. Argenis Soutelo-Jiménez, Dr. C. América García-López, Ms. C. Julio Rojas-Vargas, Ms. C. Yennys Hernández-Molina
asoutelo@cnt.uo.edu.cu, america@cnt.uo.edu.cu, jrojas@cnt.uo.edu.cu, yhmolina@cnt.uo.edu.cu

Facultad de Ciencias Naturales y Exactas. Universidad de Oriente, Santiago de Cuba, Cuba

Recibido: 8 de marzo de 2015

Aprobado: 15 de junio de 2015

Resumen

Se obtuvieron dos modelos discriminantes para la predicción de la actividad antibacteriana. El modelo 1 se obtuvo empleando descriptores TOPS-MODE y de fragmentos, el modelo 2 con descriptores 3D y de fragmentos, usando el Análisis Discriminante Lineal. El estudio se realizó con 402 compuestos reportados en la literatura. El modelo 1 clasificó el 90 % de casos activos y el 97 y 93 % de casos inactivos en las series de entrenamiento y predicción respectivamente, con una clasificación global de 93 y 91 %. El modelo 2 clasificó el 89 y 90 % de casos activos y el 95 y 91 % de casos inactivos en las series de entrenamiento y predicción respectivamente, con una clasificación global de 92 y 89 %. Estos resultados y los valores de los índices estadísticos de los modelos permitieron demostrar sus calidades. Además, fueron calculadas las contribuciones de los fragmentos a la actividad antibacteriana para ambos modelos.

Palabras clave: QSAR, antibacteriana, TOPS-MODE, descriptores 3D, LDA.

Abstract

Two discriminant models for prediction of antibacterial activity were obtained. Model 1 was obtained using descriptors TOPS-MODE and fragments, model 2 with 3D descriptors and fragments using Linear Discriminant Analysis. The study was performed with 402 compounds reported in the literature. Model 1 ranked 90 and 90 % of active cases and 97 and 93 % of inactive cases in training sets and prediction respectively, with an overall rating of 93 and 91 %. Model 2 ranked 89 and 90% of active cases and 95 and 91 % of inactive cases in training sets and prediction respectively, with an overall rating of 92 and 89 %. These results and the values of the statistical indices of the models allowed to show their qualities. In addition, the contributions were calculated fragments of antibacterial activity for both models.

Keywords: QSAR, antibacterial, TOPS-MODE, 3D descriptors, LDA.

Introducción

Las bacterias son células procariotas muy abundantes que se encuentran en cualquier medio, debido a la gran variedad de metabolismos que pueden presentar [1]. Aunque la gran mayoría de las bacterias son inofensivas existen algunas patógenas. Estas últimas son causantes de enfermedades infecciosas. Las infecciones bacterianas pueden ser tratadas con antibióticos. Estos medicamentos tienen acción bactericida si logran reducir considerablemente el desarrollo o eliminan totalmente al agente causante. A pesar de la gran cantidad de antibióticos existentes, es una necesidad insoslayable la búsqueda de nuevas entidades, atendiendo a que las bacterias crean resistencia a los mismos, y debido a la aparición de nuevas cepas y bacterias.

En los últimos años, la síntesis de nuevas entidades con actividad biológica de interés, es asistida por procedimientos *in silico*, que disminuyen notablemente el tiempo y los recursos necesarios para su introducción comercial. Una de ellas es la metodología que relaciona la estructura molecular con la actividad biológica denominada QSAR (siglas en inglés para la Relación Cuantitativa Estructura-Actividad). Esta es una hipótesis matemática que tiene su fundamento en el hecho de que la variación de la estructura de un grupo de moléculas es responsable de la variación de las propiedades químicas, fisicoquímicas, biológicas o farmacológicas [2-4].

En el grupo de investigación de Bioactivos y Química Sostenible de la Universidad de Oriente, se ha aplicado dicha metodología para obtener modelos teóricos que permiten predecir la actividad fungicida de agroquímicos y el diseño teórico de nuevas entidades, y para predecir la actividad insecticida [5, 6]. Posteriormente, atendiendo a la no existencia de una metodología disponible para predecir, de forma eficiente, si el nuevo fungicida tiene riesgo de resistencia debido a la especie fúngica, se desarrolló un modelo QSAR que considera el riesgo multirresistente, basado en descriptores sub-estructurales, desarrollados de una base de datos heterogénea de compuestos. De este modo, se obtuvo un modelo que permitió clasificar, diseñar y predecir a los fungicidas agroquímicos de acuerdo con las categorías del riesgo de resistencia [7].

En todos estos estudios se han confeccionado bases de datos de compuestos comerciales de una gran diversidad estructural, clasificando los mismos como activos o no, independientemente de la especie sobre la que actúen. Las mismas se han empleado para obtener modelos discriminantes de actividad, aplicando el modelo estadístico discriminante lineal basado en descriptores topológicos-sub-estructurales (como descriptores grafo-teóricos) y descriptores de fragmento, como punto de partida para el

cálculo de la actividad de fragmentos y el diseño de nuevas entidades. Esta metodología se pretende extender para el estudio de la actividad antibacteriana.

Materiales y métodos

Descriptores moleculares basados en fragmentos

En publicaciones anteriores, autores como Speck-Planche y col., emplearon descriptores moleculares basados en fragmentos y brindaron una buena descripción de ellos [8, 9]. De estos descriptores se obtiene información acerca de cuáles de los fragmentos que constituyen una molécula tienen influencia positiva o negativa en la actividad en estudio, lo que permite dirigir con más eficiencia el diseño de compuestos bioactivos en el sentido de minimizar el número de fragmentos con contribución negativa, y maximizar la presencia de fragmentos que contribuyan positivamente.

Suma de grupos funcionales

Los descriptores moleculares basados en la Suma de los Grupos Funcionales son descriptores moleculares simples definidos como el número de grupos funcionales específicos en una molécula. Ellos son calculados conociendo la composición molecular y la conectividad de los átomos. Se ha prestado mucha atención para distinguir si el mismo grupo funcional pertenece a un fragmento molecular alifático o aromático.

Fragmentos centrados en el átomo

Son descriptores moleculares simples definidos como el número de tipos de átomos específicos en una molécula. Ellos son calculados teniendo en cuenta la composición molecular, las hibridaciones y conectividades de los átomos.

Cada tipo de átomo es descrito en la molécula por sus átomos vecinos. Los átomos de hidrógeno y halógeno son clasificados por la hibridación y el estado de oxidación del átomo de carbono al cual ellos están enlazados y por los átomos de hidrógeno y heteroátomos que están unidos a un carbono en posición alfa (α).

Descriptores TOPS-MODE

El TOPS-MODE (Diseño Molecular Sub-Estructural Topológico) es un método de diseño molecular con una base grafo-teórica que necesita, para llegar a relaciones cuantitativas, el uso de datos muestrales, que son procesados estadísticamente. Esto lo convierte en un método general, que no necesita del conocimiento de los mecanismos que intervienen en un proceso dado para describirlo.

El enfoque TOPS-MODE en los últimos años ha sido intensamente estudiado y aplicado, tanto en el mundo académico como en la industria. Mediante este método de diseño molecular, se han desarrollado modelos para la predicción de compuestos antiinflamatorios [10], predecir coeficientes de permeabilidad [11], predecir actividad

fungicida [5] e insecticida [6], empleando en todos los casos como plataforma computacional el software MODESLAB [12].

Opera con la matriz de adyacencia entre enlaces del grafo molecular (E). Los descriptores usados por este método son el conjunto de los momentos espectrales de dicha matriz. El momento espectral de orden k (μ_k) se define como el trazo (Tr) o la suma de los valores de la diagonal principal de la matriz E^k . Esto se puede expresar matemáticamente según la ecuación (1) [13]:

$$\mu_k = Tr(E^k) = \sum_{i=1}^m k e_{ii} \quad (1)$$

donde e_{ii} son los elementos de la diagonal principal de la matriz E^k . En esta matriz los elementos no incluidos en la diagonal principal toman valores de 1 o 0, si los correspondientes enlaces son adyacentes o no entre sí.

Los elementos de la diagonal principal representan pesos de enlaces que describen propiedades estéricas, electrónicas e hidrofóbicas/polares de las moléculas, es decir son las contribuciones de los enlaces a las propiedades físico-químicas como son el coeficiente de partición, área de la superficie polar, polarizabilidad, cargas atómicas de Gasteiger-Marsilli, radios atómicos de Van der Waals, refracción molar, y descriptores moleculares de Abraham.

Los descriptores TOPS-MODE son considerados también como descriptores de fragmentos lo que es posible porque los momentos espectrales pueden expresarse como la combinación lineal de fragmentos moleculares, lo cual puede mostrarse en la ecuación (2):

$$\mu_k = \sum_{i=1}^s k c_i |F_i| \quad (2)$$

donde, al número $k c_i$ se le llama contribución del fragmento i al momento espectral de orden k ; $|F_i|$ es el número de veces que aparece el fragmento i en el grafo molecular, y s es el número de clases de fragmentos que contiene la molécula. A medida que aumenta el orden del μ_k , aumenta el número de fragmentos que contribuyen a él y la cantidad de información estructural que registran.

Descriptores 3D

Los descriptores TOPS-MODE han demostrado explicar, de forma razonable, una parte considerable de los fenómenos espaciales [14] correspondientes a los descriptores 3D. Sin embargo, Bath *et al.* indicaron que, muchas veces, esta información no es suficiente

y, por esta razón, los descriptores 3D son necesarios para obtener información más detallada de la molécula [15].

Ejemplo de lo anterior son los descriptores WHIM, los cuales, son contruidos de forma tal que capturan información relevante 3D de la molécula en relación con el tamaño, forma, simetría y distribución atómica respecto a una estructura invariante de referencia. El algoritmo consiste en la interpretación de los Principales Componentes de Análisis centrados en las coordenadas cartesianas de una molécula (matriz molecular de centrado), utilizando una matriz de covarianza ponderada, obtenida desde diferentes esquemas de ponderaciones para los átomos [16]. Los perfiles moleculares de Randic, como indica el nombre, busca la forma de transcribir la superficie molecular, el contorno y toda la forma de la estructura.

Selección de la muestra

La totalidad de moléculas analizadas en el estudio QSAR está formada por 402 estructuras, de ellas 202 son activas y 200 son inactivas. Se consideraron activas aquellas sustancias reportadas con una concentración mínima inhibitoria al 50 % (MIC₅₀) menor o igual que 10 µM, según el reporte de Prado-Prado [17] y referencias dentro del mismo. De las moléculas activas, 178 se tomaron de la National Center of Biotechnology Investigation (NCBI) [18], y 24 entre las reportadas en un estudio QSAR desarrollado por Prado-Prado [17]; las inactivas ante la actividad bajo estudio son fungicidas comerciales [19, 20].

Para el estudio, las mismas fueron agrupadas según semejanzas estructurales de diferentes clases de antibióticos, lo que se refleja en la tabla 1.

TABLA1. CLASES DE ANTIBIÓTICOS RECOPIRADOS, Y CANTIDAD DE ESPECIES POR FAMILIAS DE BACTERIAS

Clases de antibióticos	Especies de bacterias
Aminoglicósidos (15)	Estafilococos (15)
Cefalosporinas (30)	Estreptococos (7)
Cloramfenicol (2)	Salmonella (7)
Clorobiocin (15)	Pseudomonas (3)
Macrólidos (26)	Proteus (5)
Penicilinas (29)	Prevotella (6)

(continuación tabla 1)

Oxazolidinonas (7)	Porfiromonas (2)
Lincosamidas (1)	Mycobacterium (6)
Glicopéptidos (3)	Klebsiella (3)
Carbapenem (7)	Haemophilus (3)
Monobactámicos (1)	Fusobacterium (6)
Quinolonas (28)	Enterococos (3)
Sulfonamidas (20)	Enterobacter (4)
Tetraciclinas (9)	Corynebacterium (6)
Otros tipos (9)	Clostridium (11)
	Bacteroides (10)

Como se aprecia, la muestra garantiza una amplia diversidad estructural en este tipo de medicamentos, además estas moléculas son activas al menos a una especie de bacteria.

Generación de las estructuras y cálculo de los descriptores

Para este estudio se emplearon tres tipos de descriptores: de fragmentos, tridimensionales y TOPS-MODE. Se empleó el software ChemDraw Ultra 8.0 [21] para generar las estructuras y los códigos SMILES, según fuese el caso, y posteriormente se almacenaron estos en ficheros *.txt, los que se emplearon como entrada para el cálculo de los descriptores TOPS-MODE empleando el software ModesLab1.5 [12] (se calcularon todos los descriptores existentes en este programa informático). Las estructuras se optimizaron a nivel semiempírico AM1, empleando el software HyperChem 7.0 [22]; el fichero de extensión *.hingenerado, resultó la entrada para el software DRAGON 5.3 [23], donde se calcularon los descriptores de fragmentos y todos los tridimensionales para que el modelo fuera quien escogiera entre todas las variables.

Construcción de las series. Obtención y validación de las funciones discriminantes

Las moléculas se agruparon aleatoriamente en dos subgrupos: serie de entrenamiento (SE) y serie de predicción (SP), en proporciones de 75 % y 25 %, respectivamente del total de la muestra. La SE se utilizó en la obtención de las funciones discriminantes, en tanto la SP se empleó para la validación externa de los modelos encontrados.

La SE estuvo formada por 152 compuestos activos y 152 inactivos, para un total de 304 (75,62 % de la muestra). La SP la conformaron 50 compuestos activos, y 48 inactivos, para un total de 98 (24,38 % de la muestra). El Análisis Discriminante Lineal (LDA por

sus siglas en inglés) ha sido una de las técnicas más ampliamente utilizada en muchos estudios QSAR [7, 24].

Una expresión general para una función discriminante puede escribirse de la forma siguiente:

$$A = a_0 + a_1 D_1 + a_2 D_2 + \dots + a_k D_k \quad (3)$$

donde: A es la actividad antibacteriana, y tomará valores de 1 (compuesto antibacteriano) y de -1 (compuesto inactivo); a_1 , a_2 y a_k representan los coeficientes de la ecuación; a_0 es el término constante, y las variables D_1 , D_2 y D_k , representan las variables independientes métricas, en este caso los descriptores.

Se diseñó el estudio de dos funciones discriminantes (FD):

FD1 empleando descriptores TOPS-MODE, de suma de grupos funcionales y fragmentos centrados en el átomo.

FD2 empleando descriptores tridimensionales, de suma de grupos funcionales y fragmentos centrados en el átomo.

En cada caso, la función discriminante fue obtenida aplicando la técnica del Análisis Discriminante Lineal (LDA) implementada en el software STATISTICA 7.0 [25]. El método de selección de variables empleado fue el de regresión por etapas de todas las variables hacia delante (*forward stepwise*). La selección del mejor modelo estará sujeta al principio de la parsimonia, según el cual un fenómeno debe ser descrito con el número mínimo de elementos posibles.

Para la selección del modelo se tuvo en cuenta su calidad y capacidad predictiva, que viene definida por los valores de los índices estadísticos λ (lambda de Wilk), D^2 (distancia de Mahalanobis al cuadrado) y F (coeficiente de Fischer), el correspondiente P -valor y el porcentaje de clasificación dentro de cada grupo (para cada caso). El índice λ de Wilk evalúa la significación estadística de la capacidad de discriminación del modelo; puede tomar valores entre 0 (perfecta discriminación entre grupos) y 1 (ausencia de discriminación).

Mientras menor sea el valor, mayor es la diferencia entre las medias de las puntuaciones discriminantes de los grupos. La D^2 de Mahalanobis indica la separación de los respectivos grupos activos e inactivos, y puede ser tomado en cuenta mejor que F para un análisis de separación ideal de grupos; se reportan valores mayores que 1 para este índice estadístico. El coeficiente de Fischer (F) tiene que ver directamente con la varianza que logra explicar el modelo matemático por la introducción de un determinado número de variables. Cuanto más alto es el porcentaje de varianza explicada por el modelo, mayor será el valor de F .

Además se emplean los índices estadísticos que avalan la capacidad predictiva del modelo QSAR vienen dado por la sensibilidad, ecuación (4), especificidad, ecuación (5), valor predictivo positivo (VPP), ecuación (6), valor predictivo negativo (VPN), ecuación (7) y predicción total, ecuación (8). Usualmente se exige que el porcentaje de casos bien clasificados no sea inferior a un 75 %, para que el criterio de clasificación sea considerado como aceptable [16, 26, 27].

$$\text{Sensibilidad (\%)} = \frac{VP}{TA} \cdot 100 \quad (4)$$

$$\text{Especificidad (\%)} = \frac{VN}{TI} \cdot 100 \quad (5)$$

$$\text{VPP (\%)} = \frac{VP}{VP+FP} \cdot 100 \quad (6)$$

$$\text{VPN (\%)} = \frac{VN}{VN+FN} \cdot 100 \quad (7)$$

$$\text{Predicción total (\%)} = \frac{VP+VN}{TA+TI} \cdot 100 \quad (8)$$

donde:

TA - Total de casos activos

TI - Total de casos inactivos

VP - Casos activos clasificados correctamente

VN - Casos inactivos clasificados correctamente

FP - Casos inactivos clasificados incorrectamente

FN - Casos activos clasificados incorrectamente

La sensibilidad es la capacidad del modelo para clasificar correctamente los compuestos considerados como activos. La especificidad es la capacidad del modelo para clasificar correctamente los compuestos considerados como inactivos, y la predicción total es el porcentaje de compuestos totales clasificados correctamente. El VPP y VPN es la tendencia general del modelo para clasificar un compuesto como activo e inactivo respectivamente. Estos índices se calculan tanto para la serie de entrenamiento como para la de predicción.

Otro criterio para evaluar la calidad del modelo es la realización de la curva ROC (Receiver-Operating-Curve, por sus siglas en inglés) que se construye graficando los verdaderos positivos contra los falsos positivos, o sea, la sensibilidad contra (1-especificidad). Cuando la curva va a lo largo de la diagonal desde la esquina inferior izquierda a la superior derecha, representa los resultados de pura casualidad, o sea, que las variables bajo estudio no distinguen entre dos grupos, y entonces, el valor del área bajo la curva será de 0,5 (la curva ROC coincide con la diagonal); cuando el área bajo la curva es igual a 1, indica que no hay solapamiento entre las distribuciones [17].

Resultados y discusión

Modelo discriminante I

El modelo FD1 se obtuvo empleando descriptores TOPS-MODE, de suma de grupos funcionales y fragmentos centrados en el átomo.

En la ecuación (9) se presenta la mejor función discriminante con 11 variables que distingue entre los compuestos antibacterianos y los compuestos inactivos, junto con los parámetros estadísticos del LDA:

$$A = 1,578\ 40 + 0,280\ 70\ \mu(\text{Hyd})^1 - 0,007\ 67\ \mu(\text{Ab-sumB20})^2 - 0,159\ 95\ \text{nCp} - 2,239\ 10\ \text{nArCOOR} - 0,925\ 80\ \text{nArNH2} - 1,783\ 93\ \text{nArNR2} - 1,645\ 33\ \text{nSO2N} - 1,900\ 21\ \text{nOxirane} - 0,675\ 80\ \text{nBeta-La} - 0,675\ 80\ \text{N-067} - 1,046\ 058\ \text{N-068} \quad (9)$$

$$n = 304\ \lambda = 0,273\ 280\ \text{P-valor} = 0,000\ 0\ \text{D}^2 = 10,567\ 04\ \text{F} = 70,591\ 15$$

En la ecuación (9) existen dos descriptores de momento espectral: $\mu(\text{Hyd})^1$ representa el momento espectral de orden 1, ponderado por la hidrofobicidad, y el descriptor $\mu(\text{Ab-sumB20})^2$ codifica: el momento espectral de orden 2 ponderado por el coeficiente de partición agua-octanol. El modelo incluye, además, siete descriptores de suma de grupos funcionales que se describen a continuación: nCp, número de fragmentos tipo C (sp³) primario terminal; nArCOOR, número de fragmentos tipo ésteres aromáticos; nArNH₂, número de fragmentos tipo aminas primarias aromáticas; nArNR₂, número de fragmentos tipo aminas terciarias aromáticas; nSO₂N, número de fragmentos tipo sulfonamidas (thio-/dithio-); nOxirane, número de fragmentos tipo oxirane; nBeta-La, número de fragmentos tipo beta-lactama.

Por último, el modelo contempla dos descriptores de fragmentos centrados en el átomo: N-067, número de fragmentos tipo amina alifática secundaria; N-068, número de fragmentos tipo amina alifática terciaria.

De varios modelos analizados, sólo se seleccionó aquel donde el valor de λ estuviera cercano a 0 (0,27), lo que garantiza que los grupos provienen de poblaciones con medias significativamente diferentes; el valor D^2 es lo suficientemente grande para asumir que la distancia entre los centroides de los grupos activos e inactivos permite una buena discriminación, lo cual es corroborado por el alto valor del coeficiente de Fischer. Los estadígrafos indican que las variables escogidas explican en un alto por ciento la discriminación entre los grupos. Por tanto, podemos decir que el modelo escogido, que también presenta un pequeño *p-valor*, es de buena calidad estadística.

El modelo representado en la ecuación (9), como se muestra en la tabla 2, clasifica correctamente el 90,13 % de los compuestos activos, y el 97,36 % de los compuestos

inactivos en la SE, tiene un poder de discriminación apropiado, con un 93,75 % de buena clasificación global.

Los porcentajes de falsos activos e inactivos para la SE usada en el ajuste del modelo fueron de 1,32 % (4/304) y de 4,93 % (15/304), respectivamente. Los falsos activos son compuestos inactivos que el modelo los clasifica como antibacterianos, y los falsos inactivos son los compuestos activos clasificados como inactivos por el modelo. La tendencia general del modelo a clasificar un compuesto como activo e inactivo en la SE (VPP y VPN) fue de 97,16 % y 90,8 %, respectivamente. Esto significa que, si el modelo QSAR predice que un compuesto tiene actividad antibacteriana, la probabilidad de que el compuesto es activo es de un 97,16 %, y en el caso de que prediga que un compuesto no tiene actividad antibacteriana, la probabilidad de que sea cierto es de un 90,8 %.

TABLA 2. PORCENTAJES DE BUENA CLASIFICACIÓN EN LA SERIE DE ENTRENAMIENTO

Clasificación	% de casos correctos	Inactivos	Activos	Total
Inactivos	97,36	148	4	152
Activos	90,13	15	137	152
Total	93,75	163	141	304
VPP	VPN			
97,16	90,8			

Los criterios más importantes para aceptar o no un modelo discriminante se basan en las estadísticas para la serie de predicción tabla 3. En este sentido, el modelo desarrollado en este trabajo clasificó correctamente el 90 % de los compuestos activos, y el 93,75 % de los compuestos inactivos; el porcentaje de buena clasificación global fue de 91,84 %. Los porcentajes de falsos activos y falsos inactivos fueron de 3,06 % (3/98) y de 5,1 % (5/98). Los valores predictivos positivos y negativos (VPP y VPN) fueron de 93,75 % y 90 %, respectivamente.

TABLA 3. PORCENTAJES DE BUENA CLASIFICACIÓN EN LA SERIE DE PREDICCIÓN

Clasificación	% de casos correctos	Inactivos	Activos	Total
Inactivos	93,75	45	3	48
Activos	90	5	45	50
Total	91,84	50	48	98
VPP	VPN			
93,75	90			

Significación de las variables

En la tabla 4 se refleja la importancia relativa de las variables o descriptores, que viene dada por los valores de F , p -valor y coeficientes estandarizados (coef. Std); este último es el índice estadístico más importante para determinar la verdadera significación de las variables empleadas en el modelo, cuando estas tienen distintos órdenes de magnitud. Todas las variables usadas en el modelo son significativas, teniendo en cuenta que los p -valores son menores que 0,05 (nivel de significación). En este caso, el número de anillos β -lactámico (*nBeta-La*) es la variable de mayor importancia significativa, dado el mayor valor de F , menor p -valor y mayor valor modular de coeficiente estandarizado.

TABLA 4. SIGNIFICACIÓN DE LAS VARIABLES PARA EL MODELO 1

Descriptores	F	p-valor	Coef. Std.
$\mu(\text{Ab-sumB20})^2$	16,08	0,000 08	-0,337 45
nBeta-La	149,03	0	-0,819 08
nArNR2	108,62	0	-0,784 97
$\mu(\text{Hyd})_1$	80,97	0	0,660 71
nSO2N	28,20	0	-0,397 77
nArCOOR	35,14	0	-0,431 14
N-068	20,16	0,000 01	-0,350 17
nArNH2	17,98	0,000 03	-0,310 33
nCp	12,06	0,000 59	-0,331 07
N-067	8,64	0,003 55	-0,209 01
nOxirane	4,92	0,027 39	-0,153 61

Correlación entre las variables

En la tabla 5 se muestran los valores de los coeficientes de correlación entre las variables incluidas en el modelo 1. Como se observa, el valor más alto es de **0,58** que es menor que 0,7 aceptado por la literatura como valor mínimo para asumir correlación entre variables. Además, las variables involucradas describen aspectos muy diferentes, una es de momento espectral y la otra es de número de fragmentos; por tanto se puede afirmar que no hay correlación significativa y se aceptan todas en el modelo discriminante.

TABLA 5. MATRIZ DE CORRELACIÓN ENTRE LAS VARIABLES PARA EL MODELO 1

Variables	$\mu(\text{Hyd})1$	$\mu(\text{Ab-sumB20})2$	nCp	nArCOO	nArNH2	nArNR2	nSO2N	nOxirane	nBeta-La	N-067	N-068
$\mu(\text{Hyd})1$	1										
$\mu(\text{Ab-sumB20})2$	-0,31	1									
nCp	-0,02	0,58	1								
nArCOOR	0,12	0,24	0,17	1							
nArNH2	-0,15	0,05	-0,15	0,03	1						
nArNR2	0,16	0,06	-0,13	-0,06	0,03	1					
nSO2N	-0,01	-0,12	-0,15	-0,05	0,34	-0,07	1				
nOxirane	-0,05	0,01	0,1	-0,02	-0,03	-0,02	-0,02	1			
nBeta-La	-0,26	0,11	-0,06	-0,04	0,2	-0,13	-0,08	-0,04	1		
N-067	-0,23	0,11	0,04	-0,05	-0,03	0,2	-0,07	-0,02	0,05	1	
N-068	-0,01	0,31	0,45	-0,02	-0,12	0,11	-0,09	0,09	-0,15	-0,03	1

Curva ROC (Receiver-Operating-Curve)

Las curvas ROC para ambas series de este modelo se muestran en la figura 1. El área bajo la curva para la serie de predicción es de 0,977, y de 0,961 para la serie de entrenamiento, lo que confirma la buena significación estadística del modelo. Esto significa para la serie de entrenamiento, que tiene un área de 0,96, que un compuesto seleccionado al azar de la serie activa tiene un valor de actividad mayor que el de un compuesto elegido al azar de la serie inactiva el 96 % de las veces. Una deducción similar puede realizarse para la serie de predicción.

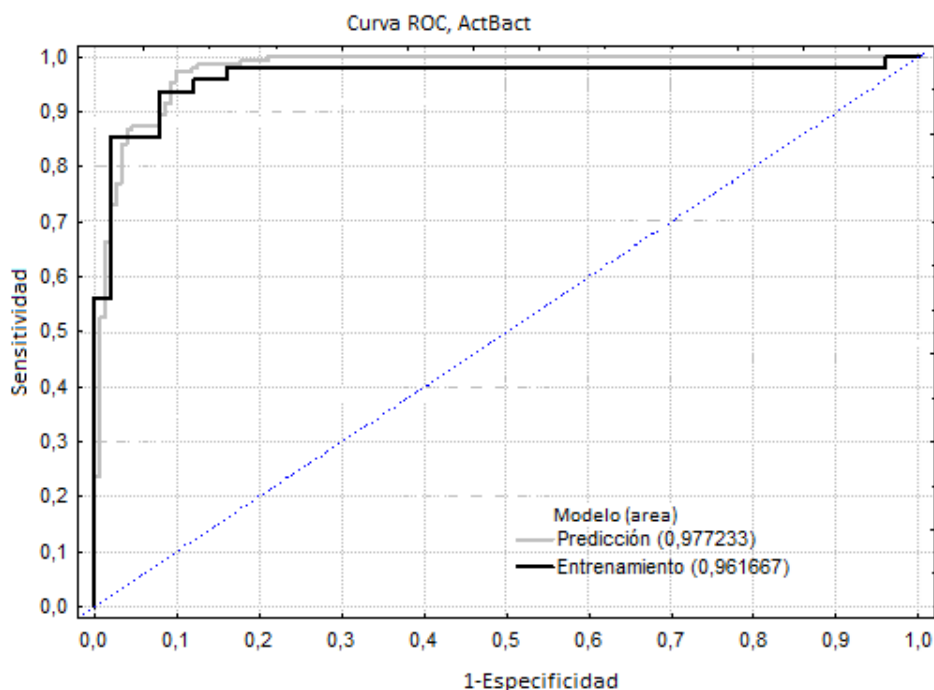


Figura 1. Curva roc para los descriptores del modelo 1

Interpretación de los descriptores más significativos del modelo 1

Un aspecto que es tan importante como la generación del modelo, es que los descriptores usados tengan una interpretación químico-física y/o estructural, de forma tal que brinden una información útil que permita un posterior diseño de nuevos compuestos.

En el caso de los descriptores basados en la suma de grupos funcionales y fragmentos centrados en el átomo se puede decir que estos expresan contribuciones a la actividad bactericida del fragmento en específico que ellos codifican. Esas contribuciones pueden ser consecuencias de propiedades químicas, o sea, pueden estar estrechamente relacionadas con su reactividad, pero al mismo tiempo pueden ser combinaciones de propiedades químicas con propiedades químico-físicas como la polarizabilidad u otros factores electrónicos y estéricos que no están contenidos dentro de la información que brindan los descriptores TOPS-MODE. No es correcto intentar dar un significado

específico a los descriptores basados en suma de grupos funcionales y fragmentos centrados en el átomo, pues esto puede suponer pérdida de información.

Estos descriptores siempre contendrán una contribución que es específica de cada fragmento, pero que puede ser debido a una combinación de diversos factores. En este sentido lo más importante es el signo del descriptor lo que dará una idea clara si la cantidad de ciertos fragmentos son favorables o no para el diseño de compuestos con actividad antibacteriana.

Los descriptores TOPS-MODE aparecen ponderados por distintas propiedades químico-físicas de las moléculas. El descriptor $\mu(Hyd)^1$ codifica la hidrofobicidad en regiones (fragmentos) de pequeño tamaño en la molécula; la que, debido al signo positivo en la ecuación, indica que un aumento de la hidrofobicidad puede resultar en un aumento de la actividad antibacteriana.

El descriptor $\mu(Ab-sumB20)^2$ codifica el coeficiente de partición [28], que podría ser agua-octanol en regiones (fragmentos) de pequeño tamaño en la molécula; la que por el signo negativo de su coeficiente, coincide con lo expresado para el descriptor anterior sobre el efecto que puede tener un aumento en la lipofilidad sobre la actividad antibacteriana.

Modelo discriminante II

El modelo FD2, empleando descriptores tridimensionales, de suma de grupos funcionales y fragmentos centrados en el átomo, se obtuvo con la misma distribución de datos empleada para generar el primer modelo.

En la ecuación (10) se presenta la mejor función discriminante, con diez variables que distingue entre los compuestos antibacterianos y los compuestos inactivos, junto con los parámetros estadísticos del LDA:

$$A = 4,378\ 77 - 3,926\ 74\ SHP2 + 0,003\ 10\ G(N...O) + 0,694\ 84\ G3e - 0,613\ 60\ nArNH2 - 1,043\ 79\ nArNR2 - 1,256\ 09\ nSO2N - 1,856\ 27\ nBeta-Lactams - 1,340\ 71\ nPyrroles - 0,373\ 67\ nHAcc - 0,789\ 64\ N-067 \quad (10)$$

$$n = 304 \quad \lambda = 0,306\ 598 \quad \mathbf{P-valor} = 0,000\ 0 \quad \mathbf{D}^2 = 8,986\ 876 \quad \mathbf{F} = 66,264\ 82$$

La ecuación (10) incluye, entre sus variables, tres descriptores del tipo tridimensional: **G(N...O)** representa la suma de distancias geométricas entre N-O, **G3e** tercer componente de simetría direccional del índice WHIM ponderado por las electronegatividades atómicas de Sanderson y **SHP2** codifica el promedio del índice forma de perfil de orden 2 de Randic. Aparecen, además, seis descriptores de suma de grupos funcionales: **nArNH2** representa el número de fragmentos tipo aminas primarias aromáticas, **nArNR2** número de fragmentos tipo aminas terciarias aromáticas, **nSO2N**

número de fragmentos tipo sulfonamidas (thio-/dithio-), **nBeta-Lactams** número de fragmentos tipo beta-lactama, **nPyrroles** número de fragmentos tipo pirrol y finalmente **nHAcc** número de átomos aceptores para enlaces por puente de H (N, O, F). Por último entró en el modelo un descriptor del tipo fragmento centrado en el átomo: **N-067** que codifica el número de fragmentos tipo amina alifática secundaria.

Se emplearon los mismos criterios estadísticos para la selección del primer modelo en la obtención de la segunda función discriminante, la cual también presentó buena calidad estadística, teniendo en cuenta el análisis de los porcentajes de buena clasificación en la serie de entrenamiento y de predicción (tablas 6 y 7).

TABLA 6. PORCENTAJES DE BUENA CLASIFICACIÓN EN LA SERIE DE ENTRENAMIENTO

Clasificación	% de casos correctos	Inactivos	Activos	Total
Inactivos	95,39	145	7	152
Activos	89,47	16	136	152
Total	92,43	161	143	304
VPP (%)	VPN (%)			
95,1	90,06			

El modelo representado en la ecuación (10) clasifica correctamente el 89,47 % de los compuestos activos, y el 95,39 % de los compuestos inactivos en la SE, tuvo un poder de discriminación apropiado, con un 92,43 % de buena clasificación global. Los porcentajes de falsos activos e inactivos para la SE usada en el ajuste del modelo, fueron de 2,3 % (7/304) y de 5,26 % (16/304), respectivamente. La tendencia general del modelo a clasificar un compuesto como activo e inactivo en la SE (VPP y VPN) fue de 95,1 % y 90,06 %, respectivamente.

En la tabla 7 se muestran las estadísticas para la serie de predicción. El segundo modelo desarrollado en este trabajo clasificó correctamente el 88 % de los compuestos activos, y el 91,66 % de los compuestos inactivos; el porcentaje de buena clasificación global fue de 89,79 %. Los porcentajes de falsos activos y falsos inactivos fueron de 4,08 % (4/98) y de 6,12 % (6/98). Los valores predictivos positivos y negativos (VPP y VPN) fueron de 91,67 % y 88 %, respectivamente.

TABLA 7. PORCENTAJES DE BUENA CLASIFICACIÓN EN LA SERIE DE PREDICCIÓN

Clasificación	% de casos correctos	Inactivos	Activos	Total
Inactivos	91,66	44	4	48
Activos	88	6	44	50
Total	89,79	50	48	98
VPP (%)	VPN (%)			
91,67	88,00			

Significación de las variables

En la tabla 8 se refleja la importancia relativa de las variables. Todas las variables que forman el segundo modelo son significativas, teniendo en cuenta que los *p*-valores son menores que 0,05 (nivel de significación). En este caso, el número de átomos aceptores para enlaces por puente de H (N, O, F) [**nHAcc**] es la variable de mayor importancia significativa.

TABLA 8. SIGNIFICACIÓN DE LAS VARIABLES

Descriptor	F	p-valor	Coef. Std.
nHAcc	141,22	0,000 000	-1,767 43
G(N...O)	64,62	0,000 000	1,102 58
nBeta-La	77,23	0,000 000	-0,618 73
nArNR2	39,76	0,000 000	-0,459 29
nSO2N	15,54	0,000 101	-0,303 67
nPyrroles	12,31	0,000 521	-0,258 16
N-067	12,02	0,000 603	-0,244 22
nArNH2	7,58	0,006 246	-0,205 68
SHP2	9,43	0,002 335	-0,331 10
G3e	4,54	0,033 904	0,167 13

Correlación entre las variables

En la tabla 9 se muestran los valores de estos coeficientes para las variables incluidas en el modelo. Como se observa, el valor más alto es de 0,83 entre las variables **G(N...O)**, descriptor 3D y descriptor de suma de grupos funcionales, por lo que estadísticamente

una de ellas debía ser excluida, sin embargo, una variable complementa la información brindada por la otra, o sea, el descriptor **nHAcc**, atendiendo al signo de su coeficiente, sugiere la disminución, no eliminación, del número de átomos aceptores para enlaces por puente de H (N, O, F) en la estructura y el descriptor **G(N...O)**. Atendiendo al signo de su coeficiente en la ecuación discriminante, sugiere que dichos átomos aceptores deben estar lo más distante posible en la estructura química que se quiera luego diseñar, por eso se decidió mantenerlas en el modelo.

TABLA 9. MATRIZ DE CORRELACIÓN ENTRE LAS VARIABLES PARA EL MODELO 2

Variables	nHAcc	nBeta-La	"nSO2N"	G(N...O)	"nArNR2"	"nArNH2"	N-067	"SHP2"	nPyrroles	"G3e"
nHAcc	1,00									
nBeta-La	0,16	1,00								
nSO2N	-0,07	-0,07	1,00							
G(N...O)	0,83	0,03	-0,05	1,00						
nArNR2	0,04	-0,12	-0,07	-0,02	1,00					
nArNH2	0,15	0,20	0,33	0,12	0,03	1,00				
N-067	0,13	0,05	-0,06	0,06	0,20	-0,03	1,00			
SHP2	-0,71	-0,09	0,08	-0,49	0,03	-0,06	-0,07	1,00		
nPyrroles	0,24	-0,08	-0,05	0,21	-0,06	0,07	-0,05	-0,27	1,00	
G3e	-0,35	-0,19	-0,04	-0,15	-0,11	-0,11	-0,09	0,51	-0,10	1,00

Curva ROC (Receiver-Operating-Curve)

Las curvas ROC para ambas series de este modelo se muestran en la figura 2. El área bajo la curva para la serie de predicción es de 0,969, y de 0,972 para la serie de entrenamiento, lo que confirma la buena significación estadística del modelo.

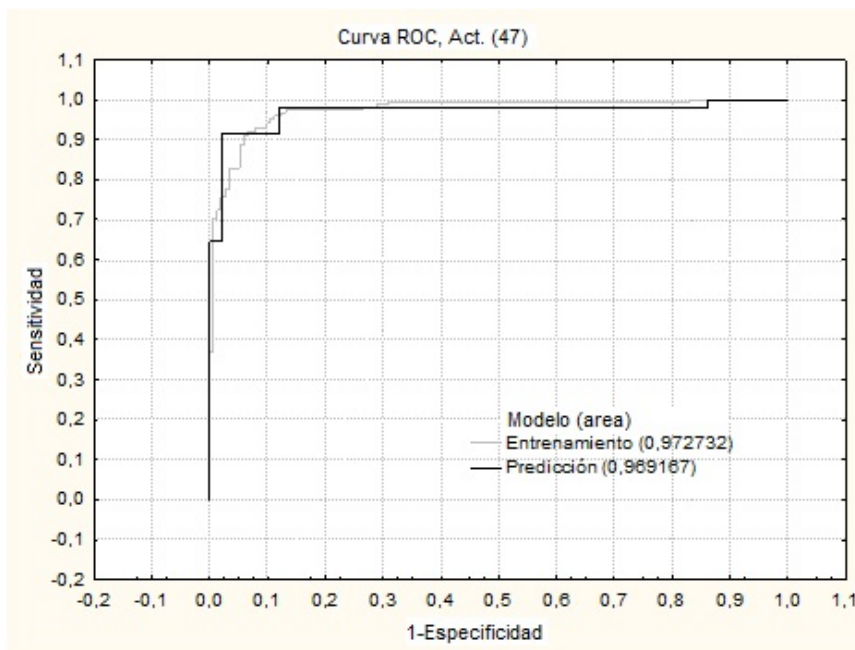


Figura 2. Curva ROC para los descriptores del modelo 2

Interpretación de los descriptores más significativos del modelo 2

La inclusión en este modelo de tres descriptores 3D indica la importancia de índices topográficos para describir la actividad antibacteriana.

De los descriptores de fragmento que entraron en el modelo1 sólo no se repiten, del modelo anterior, el número de átomos aceptores para enlaces por puente de H (N, O, F) (*nHAcc*) y el número de fragmentos tipo pirrol (*nPyrroles*), indicativo esto de que los cinco fragmentos comunes a ambos modelos deben tener un efecto significativo en la descripción de la actividad antibacteriana.

Llama la atención en este segundo modelo que el descriptor 3D **G(N...O)**, suma de las distancias geométrica entre N-O, tiene un coeficiente positivo en la ecuación, indicando, como ya se explicaba anteriormente, que un aumento en la distancia entre estos átomos podría resultar en un aumento en la actividad antibacteriana; por otra parte, el signo negativo del coeficiente del descriptor número de átomos aceptores para enlaces por puente de hidrógeno, indica la tendencia a la disminución de este tipo de enlaces dentro de la molécula, como un factor favorecedor de la actividad antibacteriana.

Si la distancia entre los heteroátomos aumenta, la posibilidad de que aquellos heteroátomos (O o N) enlazados a átomos de hidrógeno puedan establecer puentes de hidrógeno intramolecular con el otro heteroátomo, disminuye; dejando, de este modo, lábiles los heteroátomos y los átomos de hidrógeno para formar interacciones por puente de hidrógeno con las moléculas diana.

Cálculo de las contribuciones

Una de las características más interesantes del enfoque *TOPS-MODE* en el diseño molecular es la posibilidad de obtener la contribución cuantitativa de cualquier clase de subestructura a la propiedad estudiada. El número de fragmentos estructurales que puede ser evaluado para determinar sus contribuciones a la actividad bactericida es, por supuesto, muy grande.

Con el objetivo de dar una idea del funcionamiento de la metodología *QSAR* usando descriptores *TOPS-MODE* y descriptores 3D, fueron calculadas las contribuciones de 24 fragmentos (figura 3) a la actividad antibacteriana, empleando los dos modelos obtenidos; los mismos se seleccionaron por dos razones: aparición reiterada y por interés del grupo de investigación. En la tabla 10 se muestran dichas contribuciones y a la vez se puede ir comparando el desempeño de cada uno.

Obsérvese que por el modelo 1 los fragmentos F1, 2, 4, 5, 6, 7, 9, 10, 11, 13, 14 y 15 son los que contribuyen positivamente con más intensidad a la actividad, así como el *F12* lo hace de manera negativa. Por el modelo 2 no se obtuvo ningún fragmento con contribuciones negativas, la más baja de todas la presenta el fragmento *F3*, y los fragmentos que contribuyen positivamente con mayor intensidad son: *F7* y *F8*, los cuales tienen en común al anillo bencénico, y esto está en concordancia con Zhou [29].

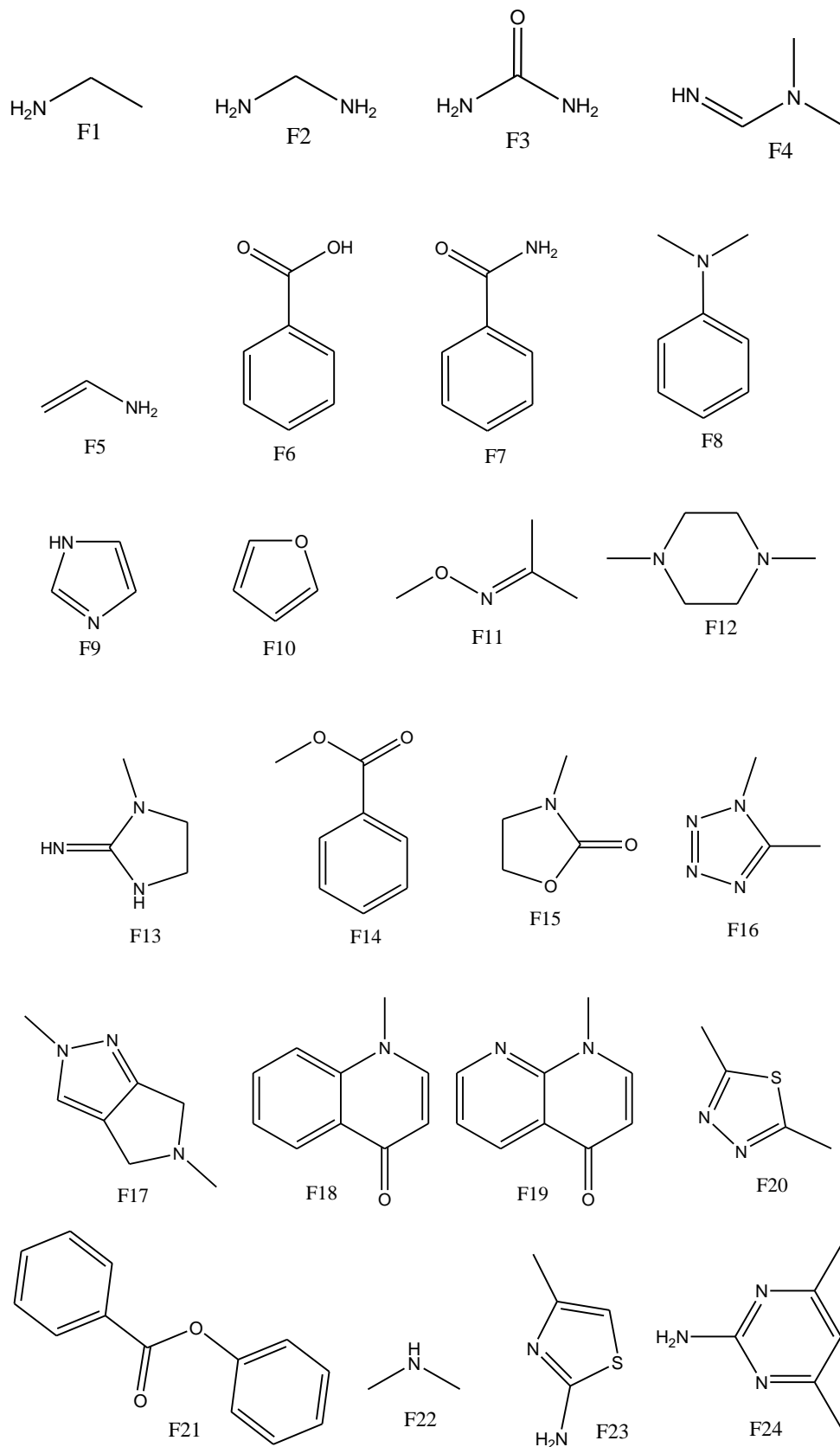


Figura 3. Estructura de los fragmentos utilizados para el cálculo de la contribución a la actividad antibacteriana

TABLA 10. CONTRIBUCIÓN DE LOS FRAGMENTOS A LA ACTIVIDAD ANTIBACTERIANA

Fragmento	Contribución		Fragmento	Contribución	
	Modelo 1	Modelo 2		Modelo 1	Modelo 2
F1	1,278	2,034	F13	1,769	1,000
F2	1,138	1,680	F14	1,943	2,410
F3	1,055	0,807	F15	1,540	1,251
F4	1,713	1,947	F16	0,862	1,703
F5	1,587	2,116	F17	0,103	1,445
F6	1,868	1,660	F18	0,162	1,352
F7	1,663	3,941	F19	0,015	1,001
F8	0,252	3,656	F20	1,064	1,955
F9	1,383	2,372	F21	0,021	2,164
F10	1,884	2,364	F22	0,837	2,089
F11	1,468	2,029	F23	0,353	1,361
F12	-0,444	1,767	F24	0,120	1,097

Desde el punto de vista práctico, uno de los principales significados del presente enfoque es detectar los fragmentos que contribuyen de forma negativa, para no ser considerados en un futuro diseño de compuestos con esta actividad, pero es un error pensar así, porque algunos de los fragmentos que contribuyen, de forma negativa, aparecen en moléculas activas, así como, algunos de los que contribuyen de forma positiva se encuentran en moléculas inactivas. Por lo que, la presencia o no de actividad biológica de un compuesto no está asociada a un único fragmento, sino que depende de la suma de todos los fragmentos presentes en la molécula.

En este caso, la suma de todos los fragmentos en los compuestos estudiados, determina que ellos sean clasificados como antibacterianos. Por otro lado, puede ser que estos fragmentos, teniendo contribución positiva en la actividad estudiada, no estén directamente relacionados con la interacción que pudiera existir entre el compuesto y el posible receptor biológico, y ello no sea determinante en el desarrollo de tal actividad.

La inclusión o no de fragmentos que tienen contribución negativa en el diseño de nuevas moléculas, suponiendo que contribuyan a la actividad biológica deseada, debe dirigirse con cuidado. En primer lugar, estos fragmentos pueden ser efectivamente inhibidores de la actividad biológica investigada, y pueden ser eliminados de la

estructura de los nuevos compuestos propuestos. Sin embargo, estos compuestos pueden contener otras subestructuras, las cuales pueden ser realmente responsables de la actividad y contribuir, de manera decisiva, a la distribución, metabolismo y farmacocinética de tales compuestos.

Evaluación de la calidad de predicción del modelo ante compuestos no comerciales con actividad experimentalmente reportada

Como un criterio adicional de validación del modelo se calcularon compuestos reportados como activos ante diferentes bacterias, los que no son medicamentos formulados comercialmente. De los once compuestos estudiados, nueve son predichos satisfactoriamente por el modelo 1, para un 81,8 %, y 8 por el modelo 2, para un 72,7 %.

En la tabla 11 se muestra un primer grupo de los compuestos estructuralmente formado por nucleósidos, unidos a D-glucosa (figura 4) que fueron sintetizados y determinada su dosis mínima inhibitoria contra: *Staphylococcus aureus*, *Listeria inovanii*, *Klebsiellapneumoniae*, *Salmonella* spp. y *Escherichiacoli*, por Amara y col. [30]. De los valores de actividad predicha lo único que se tomó para análisis en esta parte fue el signo (-) inactivo y (+) activo, o sea, no se compara las magnitudes de MIC₅₀ con los valores de actividad predicha.

TABLA 11. ACTIVIDAD EXPERIMENTAL DE UN GRUPO DE NUCLESÓSIDOS Y PREDICCIÓN POR LOS MODELOS

Compuestos	Gram positiva MIC ₅₀ (µg/mL)		Actividad Predicha	Gram negativa MIC ₅₀ (µg/mL)		
	S. aureus	L. inovanii		K. pneumoniae	Salmonella spp.	E. coli
1	0,94	1,875	-0,425	1,875	1,875	1,875
			-0,707*			
2	-	-	0,736	1,875	-	-
			0,192*			
3	1,875	1,875	0,835	-	-	-
			-0,582*			
4	0,94	-	-0,694	-	-	1,875
			-0,446*			

*Actividad predicha por el modelo 2, el resto predicha por el modelo 1.

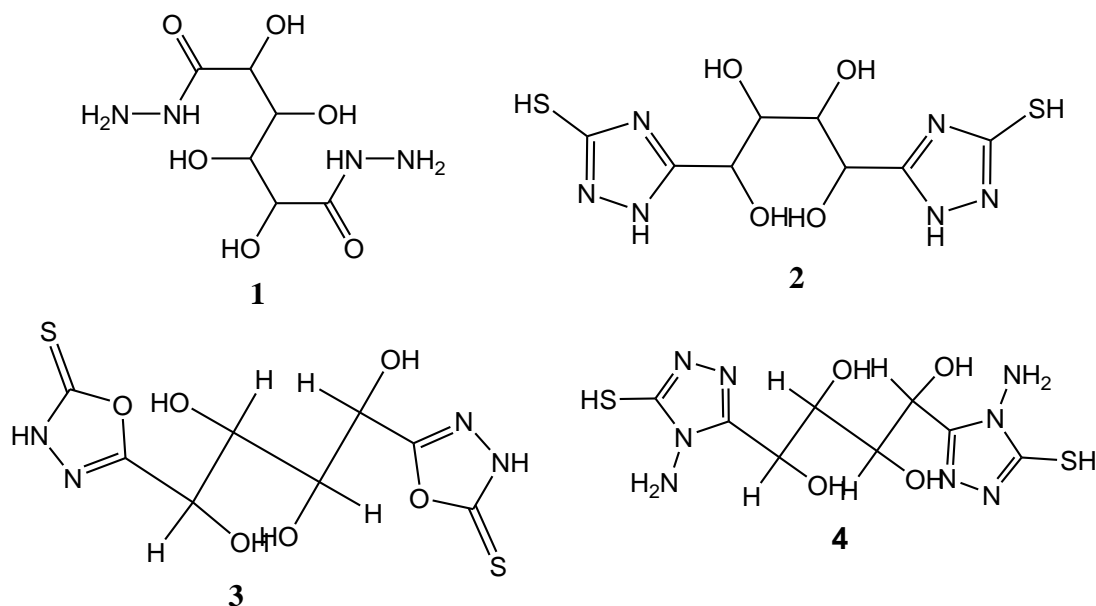


Figura 4. Grupo de compuestos formados por nucleósidos y d-glucosa

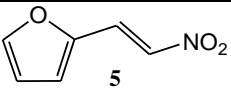
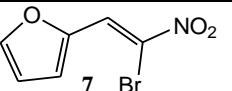
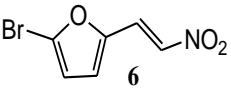
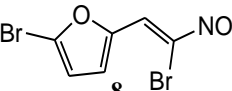
Se pudo apreciar que los compuestos mal predichos por ambos modelos, fueron los que tenían menor valor calculado del descriptor que explica la hidrofobicidad en el modelo 1, $\mu(\text{Hyd})^1$ y mayor valor de la variable $n\text{HAcc}$ en el modelo 2 que describe la cantidad de heteroátomos que pueden formar puentes de hidrógeno y por tanto son más hidrofílicas.

Atendiendo a los signos de los coeficientes de dichos descriptores en ambos modelos, la actividad se favorece con un incremento de $\mu(\text{Hyd})^1$ y una disminución de $n\text{HAcc}$ en su modelo respectivo. Según algunos autores [29], el diseño de nuevos fármacos debe seguir el camino de la obtención de moléculas más lipofílicas para favorecer su acceso a los puntos dianas y hacer más probable el efecto deseado.

Seguidamente, se estudió la capacidad predictiva de los modelos para otro grupo de antibacterianos, formado por cuatro compuestos furánicos sintetizados en el Centro de Bioactivos Químicos de la Universidad Central de las Villas (UCLV) “Martha Abreu” [31].

En la tabla 12 se compara la actividad ante la bacteria *E. coli* con la actividad predicha por los modelos. Como se puede apreciar, ambos modelos predicen como activos a todos los compuestos analizados.

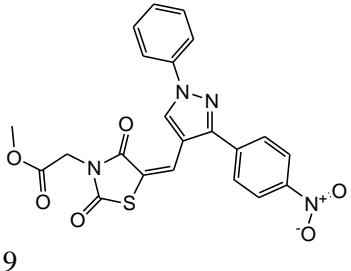
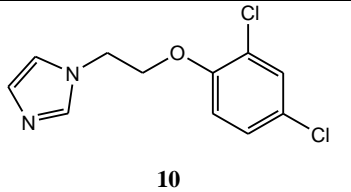
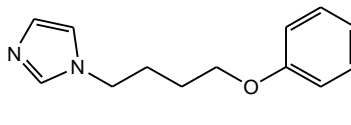
TABLA 12. COMPARACIÓN DE LA ACTIVIDAD EXPERIMENTAL DE UN GRUPO DE COMPUESTOS FURÁNICOS Y PREDICCIÓN POR LOS MODELOS

Compuestos	MIC (µg/mL)	Act. pred.	Compuestos	MIC (µg/mL)	Act. pred.
	50	1,847		12,5	2,131
		1,987*			1,940*
	25	2,065		3,125	2,349
		1,897*			1,862*

*Actividad predicha por el modelo 2, el resto predicha por el modelo 1.

En la tabla 13 se comparan las actividades de tres compuestos recientemente sintetizados y determinada su actividad contra: *Staphylococcus aureus*, *Bacillus subtilis* y *Salmonella typhimurium*, según reportan sus autores, Deepak y col. [32] y Khalafi-Nezhad, A. y col. [33]. Como puede apreciarse en la misma, ambos modelos predicen acertadamente a los mismos como antibacterianos. Además, se puede comprobar que cuando se aumenta la distancia entre los átomos aceptores para puente de hidrógeno, aumenta considerablemente también la actividad antibacteriana.

TABLA 13. COMPARACIÓN DE LA ACTIVIDAD EXPERIMENTAL DE UN GRUPO DE COMPUESTOS Y PREDICCIÓN POR LOS MODELOS

Compuestos	MIC ₅₀ (µg/mL)		Modelo 1	Modelo 2
	S. aureus	B. Subtilis		
Deepak y col. [32] 	64	32	1,704	0,844
Khalafi-Nezhad, A. y col. [33]	S. aureus	S. Tiphy		
	7,5	8,5	2,241	2,202
	7,5	7,5	2,268	2,302

Conclusiones

Tanto el modelo discriminante basado en descriptores TOPS-MODE y suma de grupos funcionales y fragmentos centrados en el átomo, como el modelo basado en descriptores tridimensionales y suma de grupos funcionales y fragmentos centrados en el átomo, con 11 y 10 variables, respectivamente, tienen calidades estadísticas análogas, por lo que podrán ser aplicados en el diseño de nuevos fármacos. Se seleccionaron 24 fragmentos, a los cuales se les determinó su contribución a la actividad antibacteriana, empleando los dos modelos discriminantes. Ambos modelos en algunos de sus descriptores indican la conveniencia de que en la molécula a diseñar predomine el carácter hidrofóbico.

Referencias bibliográficas

1. CONSORTIUM, H. M. P. "A framework for human microbiome research", *Nature*. 2012, **486** 215–221.
2. KING, R. B. *Chemical Applications of Topology and Graph Theory*. Amsterdam: Elsevier, 1983.
3. SEXTON, W. A. *Chemical Constitution and Biological Activity*. New York: D. Van Nostrand, 1950.
4. HANSCH, C. "A quantitative approach to biochemical structure-activity relationships". *Acc Chem Res*. 1969, **2** 232--239.
5. SPECK PLANCHE, A.; GUILARTE MONTERO, L.; YERA BUENO, R.; ROJAS VARGAS, J. A.; GARCÍA LÓPEZ, A.; URIARTE, E.; MOLINA PÉREZ, E. "Rational design of new agrochemical fungicides using substructural descriptors". *Pest Manag Sci*. 2011, **67** (4) 438-445.
6. SPECK PLANCHE, A.; CORDEIRO, M. N.; GUILARTE MONTERO, L.; YERA-BUENO, R. "Current computational approaches towards the rational design of new insecticidal agents". *Curr Comput Aided Drug Des*. 2011, **7** (4) 304-314.
7. SPECK PLANCHE, A.; KLEANDROVA, V. V.; ROJAS-VARGAS, J. A. "QSAR model toward the rational design of new agrochemical fungicides with a defined resistance risk using substructural descriptors". *Mol Divers*. 2011, **15** (4) 901-909.
8. 9. SPECK PLANCHE A; SCOTTI, M. T; EMERENCIANO, V. P; GARCÍA LÓPEZ, A; M. P. E.; URIARTE, E. "Design of novel antituberculosis compounds using graph-theoretical and substructural approaches.". *Mol Divers*. 2009, **13** 445–458.

9. _____. "Designing novel antitrypanosomal agents from a mixed graph-theoretical substructural approach". *J. Comput Chem.* 2010, **31** 882–894.
10. PÉREZ GONZÁLEZ, M.; DIAS, L. C.; HELGUERA, A. M.; RODRÍGUEZ, Y. M.; DE OLIVEIRA, L. G.; GÓMEZ, L. T.; DÍAZ, H. G. "TOPS MODE based QSARs derived from heterogeneous series of compounds. Applications to the design of new anti-inflammatory compounds". *Bioorganic & Medicinal Chemistry.* 2004, **12** (16) 4467-4475.
11. PÉREZ GONZÁLEZ, M.; MORALES HELGUERA, A.; GONZÁLEZ DÍAZ, H. "A TOPS-MODE approach to predict permeability coefficients.". *Polymer.* 2004, **45** (6) 2073-2079.
12. ESTRADA, E.; GUTIERREZ, Y. *ModesLab.* version 1.5 edn, Copyright 2002-2004.
13. ESTRADA, E.; RAMÍREZ, A. "Edge Adjacency Relationships and Molecular Topographic Descriptors. Definition and QSAR Applications." *J. Chem Inf and Comput Sci.* 1996, **36** 837-843.
14. ESTRADA, E; MOLINA, E; I., P.-L. "Can 3D structural parameters be predicted from 2D (topological) molecular descriptors?". *J. Chem Inf Comput Sci.* 2001, **41** 1015–1021.
15. BATH, P. A; POIRRETE, A. R; P., W. "The extent of the relationship between the graph-theoretical and the geometrical shape coefficients of chemical compounds". *J. Chem Inf Comput Sci.* 1995, **35** 714–716.
16. TODESCHINI, R.; CONSONNI, V. *Handbook of Molecular Descriptor.* Germany: VCH, 2000.
17. PRADO-PRADO, F. J.; GONZÁLEZ-DÍAZ, H.; SANTANA, L.; UNIARTE, E. "Unify QSAR approach to antimicrobials. Part 2: Predicting activity against more than 90 different species in order to halt antibacterial resistance". *Bioorg Med Chem.* 2007, **15** 897-902.
18. DATABASE, P. 2013, <http://pubchem.ncbi.nlm.nih.gov/>.
19. MERCK AND CO. INC. (2000) *The Merck Index on CD-ROM.* Version 12:3. Published on CD-ROM by Chapman & Hall/CRC. Whitehouse Station NJ. USA. Copyright 2000. [www document]
20. PESTICIDESDATABASE. 2013, http://www.alanwood.net/pesticides/class_pesticides.html.

21. CAMBRIDGE SOFT CORPORATION. ChemDraw Ultra (Chemical Structure Drawing Standard). version 8.0 edn, Copyright 1985-2003.
22. HYPERCUBE INC. HyperChem Release for Windows (Molecular Modeling System) version 7.1 edn. Copyright 2002.
23. TODESCHINI, R.; CONSONNI, V.; MAURI, A.; PAVAN, M. *Dragon Professional version*. Software version 5.3 edn. Milano. Italy: TALETE Inc., Copyright 2005.
24. SPECK PLANCHE, A.; KLEANDROVA, V. V.; CORDEIRO, M.; S., N. D. "New insights toward the discovery of antibacterial agents: Multi-tasking QSBER model for the simultaneous prediction of anti-tuberculosis activity and toxicological profiles of drugs". *European Journal of Pharmaceutical Sciences*. 2012, **48** 812-818.
25. STATSOFT INC. STATISTICA (Data Analysis Software System). Version 7 edn, Copyright 1984-2004.
26. GONZÁLEZ, H.; PRADO, F.; SANTANA, L.; URIARTE, E. "Unified QSAR approach to antimicrobials. Part 2: Predicting activity against more than 90 different species in order to halt antibacterial resistance". *Bioorg Med Chem*. 2007, **15** 897- 902.
27. GONZÁLEZ, H.; PRADO, F.; SANTANA, L.; URIARTE, E. "Unified QSAR approach to antimicrobials. Part 3: First multi-tasking QSAR model for Input-Coded prediction, structural back-projection, and complex networks clustering of antiprotozoal compounds". *Bioorg Med Chem*. 2008, **16** 5871-5880.
28. ABRAHAM, M. H. "Scales of Solute Hydrogen-bonding: Their Construction and Application to Physicochemical and Biochemical Processes ". *Chem. Soc. Rev.* 1993, 73-83.
29. ZHOU, Y.; SUN, Z.; FROELICH, J. M.; HERMANN, T.; WALL, D. "Structure activity relationships of novel antibacterial translation inhibitors: 3,5-Diamino-piperidinyl triazines". *Bioorganic & Medicinal Chemistry Letters*. 2006, **16** (20) 5451-5456.
30. AMARA, S.; OTHMAN, A. A. "A convenient new synthesis, characterization and antibacterial activity of double headed acyclo-C-nucleosides from unprotected d-glucose". *Arabian Journal of Chemistry*. 2012, **Article in press**.
31. MOLINA, E. Centro de Bioactivos Químicos. Villa Clara: Universidad Central de Las Villas, 2001.
32. DEEPAK K., Aneja; POONAM, Lohan; SANJIV, Arora; CHETAN, Sharma; K. R. A.; PRAKASH, O. "Synthesis of new pyrazolyl-2, 4-thiazolidinediones as antibacterial and antifungal agents". *Organic and Medicinal Chemistry Letters*. 2011, **1** 15.

33. KHALAFI NEZHAD, A.; SOLTANI RAD, M. N.; MOHABATKAR, H.; ASRARI, Z.; HEMMATEENEJAD, B. "Design, synthesis, antibacterial and QSAR studies of benzimidazole and imidazole chloroaryloxyalkyl derivatives". *Bioorganic & Medicinal Chemistry*. 2005, **13** (6) 1931-1938.