

## Diseño de fármacos con potencial actividad antitumoral

*Design of drugs with potential antitumoral activity*

Lic. Lisandra Rodríguez-Batista<sup>I</sup>, Dr. C. Julio César Escalona-Arranz<sup>II</sup>,  
MSc. Julio Alberto Rojas-Vargas<sup>II</sup>, Dr. C. Edgar Hernández-Sosa<sup>II</sup>

mbatista@hpuh.hlg.sld.cu; jcea1971a@gmail.com;  
jarojas@uo.edu.cu

<sup>I</sup>Instituto Nacional de Reservas Estatales, Holguín, Cuba; <sup>II</sup>Universidad de Oriente, Santiago de Cuba, Cuba.

Recibido: 25 de octubre de 2015

Aprobado: 30 de enero de 2016

---

### Resumen

El cáncer constituye un problema de salud, resultando la primera causa de muerte en el mundo. En el presente trabajo, se realiza un estudio de Relación Estructura Química-Actividad Farmacológica (QSAR) de 300 compuestos (200 activos y 100 inactivos) por Análisis Discriminante Lineal y Redes Neuronales, como instrumento para el diseño de nuevos fármacos antitumorales independientemente de su diversidad estructural. La muestra obtenida del Instituto del Cáncer de los Estados Unidos fue dividida por análisis *cluster* en una serie de entrenamiento y una de predicción. Se obtuvo una función discriminante y una red neuronal de 10 variables con un 84,33 % y 89,67 % de correcta clasificación. Se calcularon y cuantificaron las contribuciones de diferentes fragmentos estructurales a la actividad citostática, a partir de los cuales se diseñaron seis compuestos, a los que se les predice una buena actividad. La calidad predictiva del modelo obtenido por redes neuronales resultó superior al de discriminante lineal.

**Palabras clave:** cáncer, estudio QSAR, actividad antitumoral.

### Abstract

Cancer is a health problem, resulting in the first cause of death worldwide. In the present paper, a Quantitative Structure-Activity Relationship (QSAR) study was developed as a tool for the design of new drugs with potential antitumor activity. Discriminant Linear Analysis and Neural Network were the mathematics methods used to estimate the activity of in a data set consisting in 300 compounds. The biological activity, extracted from the US National Cancer Institute was divided by cluster analysis in a training and prediction series. A model with 10 variables and 84,33 % of correct classification was obtained by a discriminant function meanwhile, the neural network tested with the same number of variables resulted in a 89,67 % of accuracy. Also was calculated the contribution of different structural fragments on the cytostatic activity, and quantified their contribution. Six new compounds were designed predicting a good antitumor activity. In general, the predictive quality of the neural network model was higher than the linear discriminant.

**Keywords:** cancer, QSAR study, antitumor activity.

## Introducción

El descubrimiento de un nuevo medicamento y su desarrollo posterior es un proceso engorroso, comprendiendo en un sentido amplio, un gran conjunto de actos que culminan en la utilización terapéutica del mismo. Desde la obtención del nuevo fármaco hasta su registro comprende un total de 11,8 años de investigación, con un costo promedio de 231 millones de dólares por cada nuevo medicamento que salga al mercado. Lo más alarmante es que solo una de cada 10 000 moléculas ensayadas pasa a la fase de desarrollo, una de cada 100 000 supera los ensayos clínicos y logra registrarse, y solo 3 de cada 10 nuevos medicamentos registrados recupera su inversión inicial [1].

Por tal motivo, el diseño racional de fármacos constituye una útil herramienta en el desarrollo actual de nuevos medicamentos [2], contribuyendo a un aumento de las posibilidades de éxitos y a un decrecimiento de los costos. La teoría *QSAR*, del inglés *Quantitative Structure-Activity Relationships*, es una hipótesis matemática que tiene su fundamento en que de la estructura química de una molécula dependen sus propiedades farmacológicas [3]. Por ello, es usualmente empleada para la predicción de actividades de compuestos aún no sintetizados o ensayados farmacológicamente.

El cáncer es un problema de salud tanto para Cuba [4] como para el mundo [5, 6]. El mismo es caracterizado por un crecimiento anormal e incontrolado de células [7], manifestándose independientemente de la raza, cultura y nivel de desarrollo económico. Anualmente, se diagnostican millones de nuevos casos de cáncer en el mundo. Hasta el año 2014, en los Estados Unidos fueron diagnosticados 1 658 370 nuevos pacientes con cáncer y se estima que alrededor de 589 370 morirán a causa de esta enfermedad en el 2015 [6]. En este país, de acuerdo a un análisis de las Instituciones Nacionales de Salud, se proyecta que los gastos médicos por cáncer en el año 2020 ascenderán a 173 billones de dólares representando un 39 % de aumento con respecto al 2010 [8].

De acuerdo con informes realizados por la Organización Mundial de la Salud (OMS) la incidencia mundial del cáncer podría aumentar, pues según el desarrollo demográfico y social de nuestro planeta surgirán hasta el 2030 de 25 a 30 millones de nuevos casos [7]. Según datos de los Registros Estadísticos Médicos del Ministerio de Salud Pública, en Cuba fallecen más de 20 000 personas anualmente siendo las provincias más afectadas en diagnóstico y muertes por esta enfermedad La Habana, Holguín, Villa Clara y Santiago de Cuba [9].

A pesar de las múltiples investigaciones que se desarrollan no todos los resultados han sido esperanzadores, pues diferentes tipos de neoplasias no responden a los tratamientos actuales; de ahí que sea tan importante el desarrollo de nuevos y más avanzados tratamientos contra los diferentes tipos de cáncer. Cuba resulta un país líder en el desarrollo de fármacos con actividad antitumoral, principalmente de aquellos que resultan antígenos específicos; sin embargo, estos fármacos resultan útiles exclusivamente en aquellas células donde se expresa dicho antígeno, limitando su efectividad en otros tipos de neoplasias [10].

En la actualidad, ningún fármaco resulta lo suficientemente general como para ser aplicable a los diferentes tipos de neoplasias conocidas; además, los existentes son compuestos que presentan variabilidad en su estructura y, a la vez, poseen diferentes mecanismos de acción. Esto provoca que el proceso de elaboración de nuevos citostáticos se vea dificultado por la carencia de un compuesto tipo sobre el cual desarrollar y expresar los caracteres químicos necesarios para evocar dicha actividad. Por ello, la obtención de modelos usando *QSAR* tradicional y por inteligencia artificial para la clasificación de compuestos con actividad antitumoral permitirá identificar las características estructurales necesarias para que un fármaco posea actividad antitumoral intrínseca, independientemente del tipo de neoplasia de que se trate, y a la vez, dará la posibilidad de diseñar nuevos compuestos con dicha actividad.

## **Materiales y métodos**

### **Materiales**

Computadora Intel Pentium Dual Core, con 2 GHz de aceleración y 2 GB de RAM sobre el sistema Windows XP.

### **Métodos**

#### *Selección de la muestra*

Un total de 300 compuestos de gran diversidad estructural evaluados en 11 tipos de neoplasias y 73 líneas celulares fueron tomados de la base de datos del Instituto Nacional del Cáncer de los Estados Unidos: *National Cancer Institute* (NCI) [11], actualización de mayo de 2012. La muestra consistió en 200 compuestos activos en al menos una de las 11 líneas celulares y 100 compuestos inactivos. La actividad medida (actividad antitumoral) se representó por un código binario en el que se le asignó el valor de 1 a las sustancias activas, y el de 0 a las inactivas.

### ***Generación de las estructuras y cálculo de los descriptores***

Las estructuras fueron representadas en el software ChemDraw Ultra 7.0 [12] para obtener los códigos *SMILES* y ser almacenados en ficheros de formatos *\*.txt* y *\*.smi*. Los *\*.txt* se emplearon para el cálculo de 226 descriptores TOPS-MODE con el software ModesLab 1.5, [13], y los *\*.smi* para calcular 157 descriptores de fragmentos empleando el software DRAGON 5.3 [14].

### ***Construcción de las series***

Se realizó un análisis *cluster* empleando el software STATISTICA 8.0 [15] para generar una serie de entrenamiento (para la obtención de los modelos) y una serie de predicción (para la validación externa del modelo), en proporciones de 75 % (150 compuestos activos y 75 inactivos) y 25% (50 compuestos activos y 25 inactivos), respectivamente.

Obtención y validación de los modelos

### ***Análisis Discriminante Lineal (ADL)***

El modelo QSAR tradicional fue obtenido por Análisis Discriminante Lineal definiendo como variable dependiente la condición de activo/inactivo y como variables independientes los 383 descriptores generados. Para la realización de este análisis discriminante se empleó el paquete de programas STATISTICA 8.0 y como método de selección de variables el de regresión por inclusión progresiva y por pasos de las variables (*forward stepwise*). La aceptación del modelo estuvo condicionada a su calidad estadística evaluando los parámetros lambda de Wilks ( $\lambda$ ), coeficiente de Fisher (F) y el cuadrado de la distancia de Mahalanobis ( $D^2$ ). Además, se tuvo en cuenta la correlación existente entre las variables obtenidas en el modelo no excediera el 90 %. Se determinaron las variables de mayor influencia o peso en la actividad citostática.

### ***Redes de neuronas (RN)***

Se obtuvo un modelo por redes neuronales empleando el programa Neuroshell 2 [16] definiendo una red neuronal simple de estructura consistente en (n) neuronas en la capa de entrada (según imponga el modelo discriminante), (y) neuronas en la capa oculta en dependencia del aprendizaje de la misma, y una neurona en la capa de salida consistente en la actividad antitumoral medida (actividad/inactividad). Como criterios de calidad del modelo se emplearon el error cuadrático medio, el máximo error absoluto y el valor del coeficiente de correlación. Se determinaron las variables de mayor influencia o peso en la actividad citostática.

***Determinación de la capacidad predictiva de los modelos***

Para evaluar la capacidad predictiva de ambos modelos (ADL y RN) se consideró la matriz de clasificación estimada. Aquellas predicciones mayores a 0,50 fueron consideradas como activas y las menores o iguales a 0,50 inactivas. Además, se calculó la matriz de calidad de clasificación determinándose sensibilidad, especificidad y predicción total para cada serie, empleando las ecuaciones (1), (2) y (3), definiéndolas como [17]:

$$\text{Sensibilidad (\%)} = \frac{\text{VP}}{\text{TA}} \times 100 \quad (1)$$

TA

$$\text{Especificidad (\%)} = \frac{\text{VN}}{\text{TI}} \times 100 \quad (2)$$

TI

$$\text{Predicción total (\%)} = \frac{\text{VP} + \text{VN}}{\text{TA} + \text{TI}} \times 100 \quad (3)$$

TA + TI

donde

**TA** - total de casos activos**TI** - total de casos inactivos**VP** - casos activos clasificados correctamente**VN** - casos inactivos clasificados correctamente**Sensibilidad:** capacidad que tiene el modelo para clasificar correctamente los casos activos.**Especificidad:** capacidad que tiene el modelo para clasificar correctamente casos inactivos.**Predicción total:** valor promedio de compuestos activos o inactivos clasificados correctamente.

Se consideró el modelo como bueno cuando los parámetros antes mencionados fueran superiores al 75 % [17].

***Determinación de la contribución de los fragmentos***

Los fragmentos fueron seleccionados a partir de los 200 compuestos activos de la data y su procesamiento siguió igual curso que las moléculas ensayadas, sustituyendo los mismos en la ecuación del modelo por DL obteniendo así la contribución real de cada fragmento.

### *Diseño de las nuevas estructuras*

Para el diseño de las nuevas estructuras con potencial acción citotóxica se consideraron aquellos fragmentos de mayor aporte (positivo) a la actividad medida y la factibilidad de síntesis química de las nuevas entidades propuestas. Para cada uno de los compuestos diseñados fueron calculadas (empleando la misma metodología explicada anteriormente) la actividad teórica de cada uno de ellos. Se consideró como buena actividad cuando el valor superara 90.

## **Resultados y discusión**

### *Construcción de las series. Análisis Cluster*

Se generaron 5 *clusters*, extrayéndose de cada uno tres compuestos para la serie de entrenamiento y uno para la serie de predicción. Para aquellos *clusters* donde no se pudo mantener esta proporción, se favoreció la inclusión de los compuestos a la serie de entrenamiento, obteniendo así una mayor diversidad estructural en aquella serie sobre la cual se construyó el modelo.

Obtención de los modelos

### *Análisis Discriminante Lineal*

El modelo discriminante obtenido constó de 10 variables, ocho de las cuales corresponden a descriptores fragmentales. Los valores del coeficiente de Fisher (F), del p-valor ( $p$ ), de la distancia de Mahalanobis ( $D^2$ ) y de la lambda ( $\lambda$ ) de Wilks, confirman la calidad del modelo, el cual se describe en la ecuación 4:

$$\begin{aligned} \text{Act} = & -1,24 + 0,31 \text{ nR=Cs} + 1,11 \text{ nRNH}_2 - 0,98 \text{ nOxolanes} - 1,11 \text{ nSO}_2\text{OH} + 0,87 \text{ N-} \\ & \text{078} \\ & - 8,70 \cdot 10^{-7} \mu(\text{PS})^4 + 2,08 \text{ nCHRX}_2 - 0,16 \text{ C-024} + 1,44 \mu(\text{Ab-R2})^1 - 0,72 \text{ Cl-089} \\ n= & 300 \quad \lambda= 0,61 \quad D^2= 2,83 \quad p= 0,00 \quad F_{(10,289)}= 18,31 \quad (4) \end{aligned}$$

donde

**Act :** Actividad antitumoral.

**nR=Cs:** Número total de carbonos alifáticos secundarios.

**nRNH<sub>2</sub>:** Número de aminas primarias.

**nOxolanes:** Número de oxolanos.

**nSO<sub>2</sub>OH:** Número de ácidos sulfónicos.

**N-078:** Agrupamientos Ar-N=X o X-N=X

**$\mu(\text{PS})^4$ :** Momento espectral de orden 4 ponderado por la Superficie polar.

**nCHRX<sub>2</sub>:** Número de CHRX<sub>2</sub>

**C-024:** Agrupamientos R--CH--R

**$\mu(\text{Ab-R}_2)^1$ :** Momento espectral orden 1 ponderado por la refractividad molar.

**Cl-089:** Número de elementos Cloro enlazados a un C (sp<sup>2</sup>).

### *Redes neuronales*

El modelo obtenido resultó una red neuronal de estructura simple tipo 10:6:1. Las diez variables de la capa de entrada fueron las mismas variables señaladas por el análisis discriminante. Este modelo resultó en un error cuadrático medio de 0,089, un coeficiente de correlación  $r = 0,79$  y uno de determinación de 0,60, considerados aceptables para este tipo de estudio. Los valores predichos por este modelo y por el de Discriminante lineal (ADL) para cada uno de los 300 compuestos se muestran en la tabla 1 en la que se observa que varios de los compuestos mal predichos coinciden en ambos modelos, denotando que las mayores limitaciones del proceso de predicción no radican en la metodología o algoritmo para encontrar los mismos.

**TABLA 1 VALORES DE ACTIVIDAD PREDICHOS POR AMBOS MODELOS PARA LOS 300 COMPUESTOS ENSAYADOS**

Co.	Se.	Act.	Probabilidad		Co.	Se.	Act.	Probabilidad		Co.	Se.	Acti.	Probabilidad	
			ADL	RN				ADL	RN				ADL	RN
1	E	1	72,16	100,00	101	E	1	99,92	100,00	201	E	0	22,50	0,00
2	E	1	69,39	100,00	102	E	1	95,51	100,00	202	E	0	80,11	88,00
3	E	1	83,56	100,00	103	E	1	99,86	100,00	203	E	0	67,30	88,00
4	E	1	6,95	100,00	104	E	1	26,00	12,00	204	E	0	91,49	100,00
5	E	1	98,80	100,00	105	E	1	77,82	100,00	205	E	0	53,22	88,00
6	E	1	81,10	100,00	106	E	1	98,98	100,00	206	E	0	35,24	84,00
7	E	1	98,77	100,00	107	E	1	24,43	12,00	207	E	0	70,91	88,00
8	E	1	92,60	100,00	108	E	1	99,95	100,00	208	E	0	88,95	48,00
9	E	1	8,54	12,00	109	E	1	90,50	100,00	209	E	0	60,50	100,00
10	E	1	79,16	100,00	110	E	1	91,52	100,00	210	E	0	64,25	100,00
11	E	1	59,51	100,00	111	E	1	96,97	100,00	211	E	0	87,85	100,00
12	E	1	71,49	100,00	112	E	1	85,58	100,00	212	E	0	59,79	100,00
13	E	1	63,80	99,00	113	E	1	97,51	100,00	213	E	0	34,25	90,00
14	E	1	8,97	100,00	114	E	1	97,96	100,00	214	E	0	70,80	88,00
15	E	1	87,69	100,00	115	E	1	95,03	100,00	215	E	0	72,14	88,00
16	E	1	77,70	100,00	116	E	1	96,88	100,00	216	E	0	80,68	100,00
17	E	1	2,42	100,00	117	E	1	89,98	100,00	217	E	0	81,85	99,00
18	E	1	1,60	11,00	118	E	1	8,85	100,00	218	E	0	39,48	95,00
19	E	1	6,47	12,00	119	E	1	98,20	100,00	219	E	0	86,28	89,00

20	E	1	97,45	100,00	120	E	1	99,09	100,00	220	E	0	66,72	88,00
21	E	1	78,68	100,00	121	E	1	72,05	72,00	221	E	0	62,98	88,00
22	E	1	4,37	100,00	122	E	1	97,26	100,00	222	E	0	59,00	88,00
23	E	1	97,16	100,00	123	E	1	81,34	98,00	223	E	0	35,78	1,00
24	E	1	90,23	100,00	124	E	1	99,49	100,00	224	E	0	82,59	90,00
25	E	1	98,74	100,00	125	E	1	96,63	100,00	225	E	0	15,12	0,00
26	E	1	97,99	100,00	126	E	1	92,67	100,00	226	P	1	22,74	1,00
27	E	1	71,89	100,00	127	E	1	98,17	100,00	227	P	1	48,14	12,00
28	E	1	96,13	100,00	128	E	1	98,81	100,00	228	P	1	52,13	96,00
29	E	1	35,14	95,00	129	E	1	99,90	100,00	229	P	1	98,74	100,00
30	E	1	98,18	100,00	130	E	1	96,28	100,00	230	P	1	61,56	100,00
31	E	1	85,05	100,00	131	E	1	98,31	100,00	231	P	1	91,91	100,00
32	E	1	32,57	14,00	132	E	1	77,46	100,00	232	P	1	81,00	100,00
33	E	1	53,17	100,00	133	E	1	93,58	100,00	233	P	1	77,08	100,00
34	E	1	72,91	100,00	134	E	1	95,06	100,00	234	P	1	97,91	100,00
35	E	1	79,82	100,00	135	E	1	94,52	100,00	235	P	1	98,17	100,00
36	E	1	86,45	100,00	136	E	1	94,43	99,00	236	P	1	98,10	100,00
37	E	1	71,30	100,00	137	E	1	60,63	100,00	237	P	1	95,02	100,00
38	E	1	90,53	100,00	138	E	1	29,11	100,00	238	P	1	90,96	100,00
39	E	1	86,51	100,00	139	E	1	70,48	100,00	239	P	1	70,57	100,00
40	E	1	90,53	100,00	140	E	1	51,83	92,00	240	P	1	98,52	100,00
41	E	1	96,83	100,00	141	E	1	80,62	100,00	241	P	1	96,56	100,00
42	E	1	94,76	100,00	142	E	1	33,87	12,00	242	P	1	92,39	100,00
43	E	1	53,11	12,00	143	E	1	90,73	100,00	243	P	1	95,18	100,00
44	E	1	99,96	100,00	144	E	1	94,02	100,00	244	P	1	98,48	100,00
45	E	1	95,25	100,00	145	E	1	97,15	100,00	245	P	1	99,96	100,00
46	E	1	95,25	100,00	146	E	1	87,48	100,00	246	P	1	92,40	100,00
47	E	1	94,59	100,00	147	E	1	74,96	100,00	247	P	1	52,64	100,00
48	E	1	13,49	15,00	148	E	1	96,41	100,00	248	P	1	63,43	100,00
49	E	1	53,66	100,00	149	E	1	72,05	72,00	249	P	1	70,90	100,00
50	E	1	91,75	100,00	150	E	1	97,52	100,00	250	P	1	52,30	72,00
51	E	1	27,44	12,00	151	E	0	71,14	87,00	251	P	1	78,45	100,00
52	E	1	87,53	100,00	152	E	0	89,40	92,00	252	P	1	82,86	100,00
53	E	1	78,95	100,00	153	E	0	30,47	0,00	253	P	1	19,95	12,00
54	E	1	85,31	100,00	154	E	0	72,64	88,00	254	P	1	96,43	100,00
55	E	1	99,74	100,00	155	E	0	67,65	100,00	255	P	1	88,22	100,00
56	E	1	90,86	100,00	156	E	0	64,71	100,00	256	P	1	83,28	100,00
57	E	1	86,52	100,00	157	E	0	94,73	100,00	257	P	1	93,49	100,00
58	E	1	89,84	100,00	158	E	0	79,77	100,00	258	P	1	94,55	100,00
59	E	1	74,19	100,00	159	E	0	71,08	88,00	259	P	1	94,43	100,00
60	E	1	70,14	100,00	160	E	0	73,97	98,00	260	P	1	46,33	16,00
61	E	1	32,16	100,00	161	E	0	75,88	88,00	261	P	1	93,46	100,00
62	E	1	97,27	97,00	162	E	0	88,68	100,00	262	P	1	95,22	100,00
63	E	1	84,19	100,00	163	E	0	84,07	90,00	263	P	1	87,81	100,00
64	E	1	98,88	100,00	164	E	0	57,32	82,00	264	P	1	98,94	100,00

(Continuación tabla 1)

65	E	1	10,75	97,00	165	E	0	33,08	88,00	265	P	1	94,39	100,00
66	E	1	94,66	100,00	166	E	0	81,33	89,00	266	P	1	40,17	12,00
67	E	1	97,89	100,00	167	E	0	67,89	100,00	267	P	1	56,79	98,00
68	E	1	99,53	100,00	168	E	0	81,88	89,00	268	P	1	96,88	100,00
69	E	1	79,42	100,00	169	E	0	18,06	0,00	269	P	1	72,38	100,00
70	E	1	87,74	100,00	170	E	0	65,48	84,00	270	P	1	89,36	100,00
71	E	1	99,00	100,00	171	E	0	60,81	100,00	271	P	1	91,61	100,00
72	E	1	97,29	100,00	172	E	0	71,45	99,00	272	P	1	37,32	12,00
73	E	1	94,20	100,00	173	E	0	42,16	0,00	273	P	1	88,61	100,00
74	E	1	90,88	100,00	174	E	0	61,23	100,00	274	P	1	79,26	100,00
75	E	1	98,30	100,00	175	E	0	50,69	54,00	275	P	1	43,23	100,00
76	E	1	66,40	100,00	176	E	0	10,24	0,00	276	P	0	71,66	100,00
77	E	1	52,70	83,00	177	E	0	33,06	37,00	277	P	0	80,05	82,00
78	E	1	84,07	100,00	178	E	0	21,99	0,00	278	P	0	82,44	88,00
79	E	1	80,62	100,00	179	E	0	77,19	88,00	279	P	0	51,74	98,00
80	E	1	56,96	97,00	180	E	0	75,88	88,00	280	P	0	72,91	100,00
81	E	1	84,90	100,00	181	E	0	97,59	100,00	281	P	0	77,03	100,00
82	E	1	68,66	100,00	182	E	0	43,65	92,00	282	P	0	60,05	86,00
83	E	1	65,46	26,00	183	E	0	61,94	76,00	283	P	0	54,96	100,00
84	E	1	83,68	100,00	184	E	0	97,22	100,00	284	P	0	71,96	88,00
85	E	1	98,62	100,00	185	E	0	73,55	88,00	285	P	0	91,49	100,00
86	E	1	92,99	100,00	186	E	0	86,05	87,00	286	P	0	62,83	88,00
87	E	1	92,75	100,00	187	E	0	64,99	93,00	287	P	0	35,50	87,00
88	E	1	99,23	100,00	188	E	0	78,81	89,00	288	P	0	58,78	88,00
89	E	1	97,03	100,00	189	E	0	78,25	88,00	289	P	0	15,20	0,00
90	E	1	99,52	100,00	190	E	0	84,55	90,00	290	P	0	65,94	88
91	E	1	94,03	100,00	191	E	0	45,40	86,00	291	P	0	75,87	88,00
92	E	1	91,47	100,00	192	E	0	63,50	88,00	292	P	0	32,07	0,00
93	E	1	73,71	100,00	193	E	0	86,30	93,00	293	P	0	57,17	85,00
94	E	1	99,28	100,00	194	E	0	79,41	89,00	294	P	0	59,00	88,00
95	E	1	96,39	100,00	195	E	0	78,25	88,00	295	P	0	41,10	76,00
96	E	1	99,51	100,00	196	E	0	98,06	100,00	296	P	0	97,22	100,00
97	E	1	93,22	100,00	197	E	0	87,75	100,00	297	P	0	64,29	88,00
98	E	1	91,74	100,00	198	E	0	55,59	81,00	298	P	0	47,74	30,00
99	E	1	92,69	100,00	199	E	0	30,89	98,00	299	P	0	63,82	88,00
100	E	1	99,77	100,00	200	E	0	79,59	100,00	300	P	0	35,62	0,00

Co. Compuestos Se. Serie Act. Actividad E: entrenamiento P: predicción 1: Activo 0: Inactivo

En el modelo bidimensional de correlación de Pearson, ninguna de las variables independientes presentó correlaciones superiores a 0,6 entre sí por lo que cada una de las variables aporta información importante y no redundante a la actividad modelada.

Independientemente de la forma de obtención de los modelos (ADL y RN), las variables de mayor influencia o peso en la predicción de la actividad antitumoral fueron  $\mu(\text{Ab-R2})^1$ , C-024,  $\mu(\text{PS})^4$  y  $n\text{R}=\text{Cs}$ . La variable  $n\text{R}=\text{Cs}$ , describe el número de carbonos

alifáticos secundarios. El valor positivo en la ecuación de este descriptor, indica que los compuestos con potencial actividad citostática deben presentar un elevado número de fragmentos estructurales que cumplan con esta premisa.

La necesidad del incremento de átomos de carbono en una estructura química implica un aumento de las propiedades lipófilas del fármaco, característica necesaria para atravesar las membranas biológicas del organismo humano. El origen del cáncer es atribuido a alteraciones que ocurren en diferentes partes de la célula e incluso, muchas veces, en la expresión del ADN [7], el cual se encuentra en el interior del núcleo, constituyendo este una barrera membranosa adicional que deberán traspasar los fármacos que exhiben esta actividad farmacológica.

Ratificando lo anteriormente expuesto, se encuentra en el modelo la variable  $\mu(\text{Ab-R2})^1$ , un índice asociado a la refractividad molar, parámetro químico que correlaciona de manera inversa el volumen molar con la polarizabilidad de la molécula. De tal forma, un incremento del volumen molar o un decrecimiento de la polarizabilidad de la molécula incrementarán el valor de la refractividad molar, la cual es directamente proporcional a la liposolubilidad de la misma. El valor positivo de este índice en la ecuación, confirma la necesidad de que las moléculas con actividad citostática presenten una elevada lipofilia.

La variable que representa la superficie polar  $\mu(\text{PS})^4$  tiene una influencia negativa sobre la actividad antitumoral. Este descriptor es indicativo del carácter hidrófilo de la molécula, por lo que su influencia inversa resulta coincidente con la de los dos índices anteriormente explicados. A pesar de que se describe la influencia por separado de cada uno de los factores, ello no debe ser interpretado como algo absoluto. La estructura química en sí es única, y en ella confluyen o se presentan grupos químicos que podrán favorecer o desfavorecer la actividad. Por ello, la actividad debe ser interpretada como una sumatoria de influencias (positivas o negativas) de cada uno de los descriptores que sobre ella influyen.

#### *Validación de los modelos*

Se estimó a partir de los modelos obtenidos la matriz de clasificación general (tabla 2) y la matriz de calidad de clasificación de las dos series las cuales expresan los parámetros de sensibilidad, especificidad y predicción total (tabla 3).

**TABLA 2 MATRIZ DE CLASIFICACIÓN GENERAL ESTIMADA POR LOS MODELOS DISCRIMINANTE LINEAL Y REDES DE NEURONAS**

MATRIZ						
Clasificación de los compuestos	Modelo ADL			Modelo RN		
	Predicciones acertadas (%)	Clasificados		Predicciones acertadas (%)	Clasificados	
		Inactivos	Activos		Inactivos	Activos
Inactivos	77,00	77	23	86,00	86	14
Activos	88,00	24	176	91,5	17	183
Total	84,33	101	199	89,67	103	197

**TABLA 3 MATRIZ DE CALIDAD DE CLASIFICACIÓN ESTIMADA POR EL MODELO EN EL ANÁLISIS DISCRIMINANTE LINEAL PARA LAS SERIES DE ENTRENAMIENTO Y PREDICCIÓN**

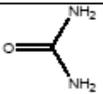
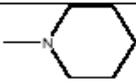
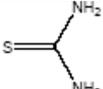
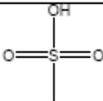
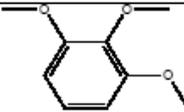
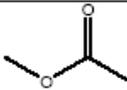
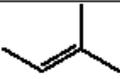
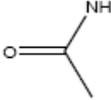
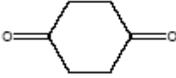
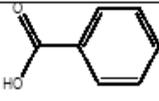
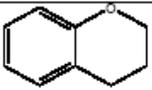
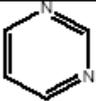
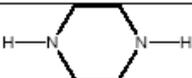
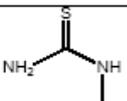
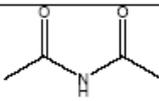
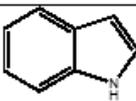
CLASIFICACIÓN												
Parámetros	Modelo ADL						Modelo RN					
	Entrenamiento			Predicción			Entrenamiento			Predicción		
	(%)	Bien	Mal	(%)	Bien	Mal	(%)	Bien	Mal	(%)	Bien	Mal
Sensitividad	88,00	132	18	88,00	44	6	92,67	139	11	88,00	44	6
Especificidad	77,33	58	17	76,00	19	6	86,67	65	10	84,00	22	3
Predicción total	84,44	190	35	84,00	63	12	90,67	204	21	86,67	66	9

Como se observa en las tablas 2 y 3 todos los valores obtenidos superaron el 75 % de aciertos considerándose como buenos según la bibliografía [17]; estos resultados permiten afirmar que ambos modelos son capaces de predecir la actividad antitumoral de nuevas entidades químicas. El modelo por Redes neuronales resultó mejor para cada uno de los parámetros considerados. Para el porcentaje de predicciones acertadas, sensibilidad, especificidad, y predicción total resultó superior en 5,34 %, 2,33 %, 8,67 % y 4,45 %, respectivamente.

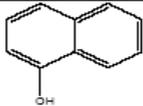
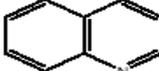
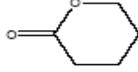
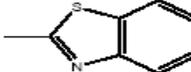
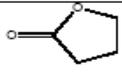
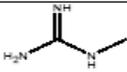
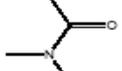
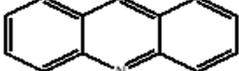
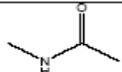
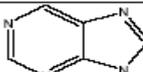
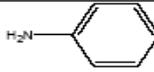
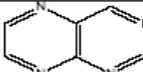
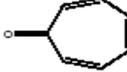
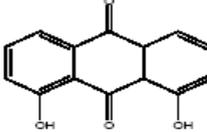
#### *Cálculo de la contribución de los fragmentos*

Se calculó la contribución de 62 fragmentos estructurales. Entre los fragmentos que influyen negativamente en la actividad evaluada, se encuentran aquellos que presentan carbonilos y sulfonilos, pero particularmente los fragmentos que mayor influencia negativa presentan son los agrupamientos ureídicos y tioureídicos. De otro lado, los que mayor influencia positiva tienen sobre la actividad son policiclos principalmente los de naturaleza heteronuclear (tabla 4). Estos sistemas anulares se caracterizan por ser altamente hidrófobos, aspecto que concuerda con la interpretación dada en el modelo, a la variable  $\mu(\text{Ab-R2})^1$ .

**TABLA 4 ESTRUCTURA Y CONTRIBUCIÓN DE LOS FRAGMENTOS QUE MÁS APORTAN (POSITIVA Y NEGATIVAMENTE) A LA ACTIVIDAD ANTITUMORAL MODELADA**

No	Fragmento	Contribución	Nro	Fragmento	Contribución
50		-0,861 50	9		0,166 88
35		-0,717 57	14		0,171 30
56	$\text{H}_2\text{N}-\text{N}=\text{CH}_2$	-0,714 95	21		0,189 01
29		-0,649 83	31		0,206 78
47		-0,578 52	44		0,213 46
25		-0,567 64	52		0,248 13
2		-0,511 58	34		0,272 70
53	$\text{S}=\text{CH}_2$	-0,463 31	40		0,299 40
20	$\text{N}\equiv\text{C}-$	-0,430 37	36		0,359 10
51	$\text{N}=\text{C}=\text{S}$	-0,387 15	58		0,391 51
55		-0,386 22	54		0,427 73
39		-0,374 88	60		0,592 34

(Continuación tabla 4)

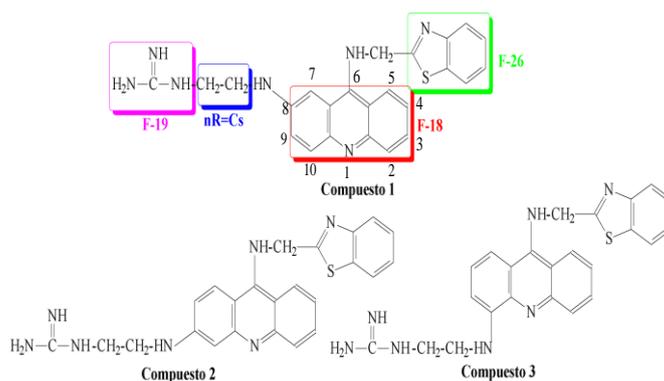
1		-0,356 54	30		0,646 01
45		-0,329 76	15		0,654 55
42		-0,321 06	61		0,667 41
41		-0,299 46	22		0,683 18
49		-0,279 16	26		0,890 02
48		-0,279 15	19		1,015 11
46		-0,267 68	18		1,254 80
33		-0,239 15	59		1,374 70
5		-0,207 02	62		1,504 45
12		-0,175 37	27		2,545 20
23		-0,132 53	38		2,756 95

Una de las ventajas que ofrece el análisis de las contribuciones de fragmentos resulta en la interpretación del por qué el modelo incurre en errores de predicción. La estructura química de un compuesto puede dividirse y representarse en diversos fragmentos, cada uno de los cuales contribuirá de manera positiva o negativa a la actividad medida. La sumatoria de estas contribuciones resultará en la actividad predicha a la molécula. Los compuestos activos 9 y 104 son clasificados incorrectamente pues cuentan en su estructura con fragmentos que según el modelo influyen negativamente en la actividad. En el caso de la sustancia 9 influyen negativamente los fragmentos 46 y 23 (tabla 4). El compuesto 104 también cuenta con el fragmento 23, además de los fragmentos 2, 12 y 42 influyendo negativamente a la actividad.

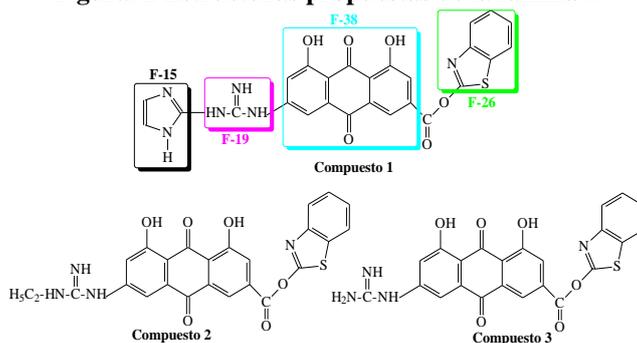
También se obtuvieron compuestos inactivos que son mal predichos por ambos modelos, tal es el caso del 169 y el 289 presentando ambos una base púrica, la cual como fragmento (fragmento 59) es identificada por su aporte significativo a la actividad antitumoral.

### *Diseño de las nuevas estructuras*

Utilizando los fragmentos con mayores aportes a la actividad antitumoral y algunas variables obtenidas en el modelo, se diseñaron un total de seis compuestos pertenecientes a dos familias diferentes (fig. 1 y 2). Para el diseño de la familia 1 se utilizaron los fragmentos 19, 18 y 26 y se cumplió con una de las variables que mayor importancia muestra en el modelo obtenido que es:  $nR=Cs$ . (fig. 1), mientras que para el diseño de la familia 2 se utilizaron los fragmentos 19, 38 y 26, y en el caso particular del compuesto 1 se utilizó el fragmento 15 (fig. 2).



**Figura. 1 Estructuras propuestas de la familia 1**



**Figura.2 Estructuras propuestas de la familia 2**

A las seis estructuras diseñadas se les predice una buena actividad antitumoral, según los modelos obtenidos por ADL y RN. Para la familia 1 la actividad predicha para los tres compuestos fue la misma; por ADL 99,65 y por redes neuronales 100. En el caso de la familia 2 se comportó similar en el modelo obtenido por ADL oscilando la actividad predicha entre 99,64 a 99,92, mientras que por redes neuronales se obtuvo una predicción para la actividad de los tres compuestos de 100. Aunque las diferencias

numéricas son muy pequeñas en apariencia los compuestos de la familia 2 son mejor predichos que los de la familia 1.

No obstante, es de señalar que ninguno de los dos modelos es discriminatorio para los compuestos pertenecientes a esta familia 1, debido a que en todos los casos el valor de la actividad predicha es el mismo, no diferenciando los patrones de sustitución en las posiciones 8, 9 y 10 del anillo de la acridina o fragmento 18. Esto puede estar dado por el hecho de que en el modelo se incluyen preferentemente variables de tipo fragmento y no variables de naturaleza topográfica, que son las que usualmente resultan de mayor capacidad discriminante de isómeros de posición.

## Conclusiones

*Se logró la obtención de modelos QSAR por Análisis Discriminante Lineal (ADL) y por Redes de Neuronas (RN), con la calidad estadística adecuada para la predicción de la actividad antitumoral de nuevos compuestos y/o sustancias no ensayadas, resultando el modelo obtenido por redes neuronales superior al de Análisis Discriminante Lineal. Además, el análisis de la influencia de los fragmentos estructurales, permitió diseñar seis nuevos compuestos químicos a las cuales se les predice una potencial acción antitumoral.*

## Referencias bibliográficas

1. ESCALONA, J. C.; CARRASCO, R.; PADRÓN, J. A., *Introducción al diseño racional de fármacos*, primera edición, Ciudad de La Habana: Universitaria, 2008, 45 p. ISBN 978-959-16-0647-1
2. MANLY, C. J.; LOUISE-MAY, S.; HAMMER, J. D., "The impact of informatics and computational chemistry on synthesis and screening", *Drug Discov. Tod.*, 2001, 6(21), 1101-1110.
3. BROWN, N.; LEWIS, R. A., "Exploiting QSAR methods in lead optimization", *Curr. Opin. Drug Discov. Devel.*, 2006, 9(4), 419-424.
4. SANSÓ, F. J.; SOBERATS, P.; GALBÁN, A.; TORRES, R. M., "Mortalidad por cáncer en Cuba", *Cub. Sal. Púb.*, 2010, 36(1), 78-94.
5. MINISTERIO DE SANIDAD Y POLÍTICA SOCIAL, *Estrategia en cáncer del Sistema Nacional de Salud*, Madrid, 2010, 166 p. ISSN 840-10-055-2
6. SIEGEL, R.; MILLER, K.; JEMAL, A., "Cancer statistics, 2015", *Can Jour. Clin.*, 2015, 65(1), 5-29.

7. CAVALLI, F., *Cáncer El gran desafío*, La Habana: Ciencias Médicas, 2012. 157p. ISB: 978-959-212-809-5
8. MARIOTTO, A. B.; YABROFF, K. R.; SHAO, Y.; FEUER, E. J.; BROWN, M. L., "Projections of the Cost of Cancer Care in the United States: 2010-2020", *J. Natl. Can. Inst.*, 2011, 103(2), 117-128.
9. MINISTERIO DE SALUD PÚBLICA, *Anuario Estadístico de Salud 2014*, 43ra. edición, La Habana, 2015. 205 p., ISSN: 1561-4425
10. MAURIZ, J. L.; ORDOÑEZ, R.; PRIEGO, N.; GONZÁLEZ, J., "La biotecnología en la salud humana: el hito de los anticuerpos monoclonales", *Biotec. Sal. Hum.*, 2014, 12, 12-33.
11. NCI DATABASE, Disponible en: [http://cactus.nci.nih.gov/download/nci/NCI-Open\\_2012-05-01.sdf.gz](http://cactus.nci.nih.gov/download/nci/NCI-Open_2012-05-01.sdf.gz) Consultado el 25de enero de 2013.
12. CHEMDRAW Ultra 7.0 Corp., C. ChemDraw Ultra (Chemical Structure Drawing Standard), Version 8.0, EE.UU., 2004.
13. GUTIÉRREZ, Y.; ESTRADA, E., "MODESLAB 1.5 (MOlecular DEScriptors LABoratory) for Windows", Universidad de Santiago de Compostela, España, 2002.
14. TODESCHINI, R.; CONSONNI, V.; MAURI, A.; PAVAN, M. DRAGON, Professional, Software versión 5.3; TALETE Inc.: Milano, Italy, 2005.
15. STATISTICA 8.0 STATISTICA (Data Analysis Software System), Version 8. StatSoft, Inc.: 2004.
16. NEUROSHELL 2.0 © Inc versión 2 Ward Systems Group: 2008.
17. TODESCHINI, R.; CONSONNI, V., *Handboock of molecular descriptor*, primera edición, República Federal Alemana, 2000, 522 p. ISBN 3-52-29913-0