

Estudio QSAR por algoritmos genéticos de reconocedores de surco del DNA

Anai Zavala Franco, Janett Betzabe González Campos y Luis Chacón García*

Laboratorio de Diseño Molecular, Instituto de Investigaciones Químico Biológicas, Edificio B-1, Ciudad Universitaria, Morelia, Michoacán, México. CP 58066

Resumen

Se describe un análisis de Relación Estructura Actividad Cuantitativa (QSAR) por Inteligencia Artificial mediante el uso de algoritmos genéticos de una serie de reconocedores de surco del DNA. Se encuentra una correlación entre las características estructurales y geométricas de reconocimiento intermolecular y la T_m del complejo ligando-DNA.

Palabras Clave: QSAR, Reconocedores de surco del DNA, descriptores moleculares.

Abstract

A quantitative structure activity relationship (QSAR) by the use of genetic algorithms of a series of DNA-groove binders is described. It was found a correlation between T_m of the DNA-recognizing molecule and its structural and geometric properties.

Keywords: QSAR, DNA-groove binders, molecular descriptors.

Introducción

Un algoritmo es por definición una serie de pasos organizados para resolver un problema específico. Durante la última década se ha incrementado el interés en el uso de algoritmos en las ciencias biológicas gracias a los avances en la computación. Una de las técnicas de computación evolucionaria, son los algoritmos genéticos [Estévez, 1997] que toman como inspiración la evolución genética y sus bases moleculares.

Los Algoritmos genéticos (AGs) son métodos adaptativos que pueden usarse para resolver problemas de búsqueda y optimización. Están basados en el proceso mediante el cual los organismos vivos, a lo largo de las generaciones, evolucionan de acuerdo a los principios de la selección natural y la supervivencia del “más fuerte”, postulados por Darwin. Los Algoritmos Genéticos, un área de las ciencias computacionales y las matemáticas imitan este proceso, lo que facilita buscar una solución a problemas reales. La evolución de dichas soluciones hacia valores óptimos del problema depende de una buena codificación de las mismas.

En la naturaleza los individuos de una población compiten entre sí en la búsqueda de recursos tales como comida, agua, refugio, incluso un compañero. Aquellos individuos que tienen más éxito en sobrevivir y en atraer compañeros tienen mayor probabilidad de generar un gran número de descendientes. Por el contrario, individuos poco dotados producirán un menor número de descendientes. Es posible pensar que los genes de los individuos mejor adaptados se propagarán en sucesivas generaciones hacia un número de individuos creciente. La combinación de buenas características provenientes de diferentes ancestros, puede a veces producir descendientes cuya adaptación es mucho mayor que la de cualquiera de sus ancestros.

Los AGs utilizan este razonamiento, de tal manera que se parte de una población inicial, siendo sus individuos una posible solución a un determinado problema. A cada individuo se le asigna un valor (en la naturaleza, es el grado de efectividad de un organismo para competir por determinados recursos). Mientras

más adaptado esté el individuo al problema, mayor será la probabilidad de que sea seleccionado para reproducirse, cruzando su material genético con otro seleccionado de la misma forma.

De esta manera se produce una nueva población de posibles soluciones, que contiene mejores características para una posible solución final satisfactoria. Así, a lo largo de las generaciones, las buenas características se propagan a través de la población. [Larrañaga *et al.*])

Una de las disciplinas que han aprovechado la inteligencia artificial y particularmente los AG's es la Química Farmacéutica dentro del diseño de fármacos a través de los métodos de relación estructura actividad cuantitativa o QSAR (por sus siglas en inglés Quantitative Structure-Activity Relationships) [Pastor, *et al.*] que han agrupado todas las técnicas intentando establecer modelos empíricos de comportamiento, sobre familias de compuestos biológicamente activos, como por ejemplo un grupo de no-nucleósidos inhibidores de VIH-1 [Hou T. J., *et al.*, 1999], para poder obtener óptimos de actividad, a partir de datos de comportamiento de un número limitado de compuestos. En resumen, un análisis QSAR busca una relación matemática entre las características fisicoquímicas de una serie de compuestos activos y su actividad biológica cuantificada, lo que permite explicar la actividad de dichos compuestos y más aún predecir la actividad de compuestos no sintetizados o aun no probados biológicamente. Así, es posible citar estudios realizados de QSAR en productos naturales (p. ejemplo flavonoides) [Stefanic-Petek, *et al.*, 2002], agentes antibacterianos [Koga, *et al.*, 1980], compuestos antifúngicos [Yalcin, *et al.*, 2000], estudios de toxicidad [Turabekova, *et al.*, 2004], compuestos citotóxicos [Suh, *et al.*, 2002], entre otros.

Un grupo de compuestos con actividad farmacológica de interés son los que interactúan con el DNA, dado su potencial como agentes citotóxicos o antibacterianos. [Martínez y Chacón-García, 2005] Por su mecanismo de acción molecular, estos se pueden clasificar como agentes que interactúan

*Autor de correspondencia: Luis Chacón García, e-mail: lchacon@umich.mx

covalentemente con el material genético y aquellos que lo hacen de manera reversible mediante interacciones intermoleculares débiles tales como puentes de hidrógeno, fuerzas de van der Waals o interacciones electrostáticas. Estos últimos son de gran importancia pues, cuando se trata de fármacos, su reversibilidad en la interacción permite o facilita la destoxificación de la célula y por consiguiente del organismo. A su vez, se subdividen en intercaladores y en reconocedores de surco (RS) dependiendo del tipo de interacción con el DNA. **Figura 1.** [Martínez y Chacón-García, 2005]

Los RS, al ser compuestos que reconocen al material genético de manera reversible, están sujetos a los principios básicos de la termodinámica, de tal manera que la energía requerida para su interacción se define por la Ecuación 1:

$$\Delta G = \Delta H - T\Delta S \quad \text{Ecuación 1}$$

En donde G representa la energía libre de Gibbs, H la entalpía, S la entropía y T la temperatura del sistema que se considera constante.

Dada la ecuación 1, se deduce que el proceso será termodinámicamente favorecido (ΔG negativo) cuando el cambio en la entalpía se ve favorecido, por lo que a su vez un aumento en las interacciones intermoleculares estabilizantes (ΔH negativo) con el DNA producirá un proceso espontáneo y deseado para su óptimo reconocimiento intermolecular. Dentro de estas interacciones se encuentran los puentes de hidrógeno que son los que gobiernan la interacción de los RS con las bases púricas y pirimídicas.

Los RS pueden ser reconocedores, en el B-DNA, del surco mayor o del menor. [Neidle, 2001] El surco mayor es

principalmente reconocido por proteínas y aunque el surco menor es considerado menos específico es el responsable del reconocimiento de moléculas orgánicas pequeñas. [Bewley, *et al.*, 1998] Estas últimas adoptan, en su sitio de acción molecular, una conformación helicoidal.

En un estudio previo, se describió el efecto de la isohelicoicidad, en función de sus interacciones por puentes de hidrógeno, de una serie de N,N' -diarilalcanodiamidas con longitud variable y con actividad citotóxica moderada. Se propuso un descriptor molecular (descriptor D_{CL}) relacionado con el reconocimiento topológico y la distancia entre grupos capaces de formar puentes de hidrógeno con el DNA, obtenido a partir de la longitud de la molécula geoméricamente optimizada por métodos computacionales semiempíricos, teniendo en cuenta que se requiere una longitud específica para un reconocimiento geométrico óptimo con la doble hélice, considerando isohelicoicidad. [Chacón-García, 2001]

Un descriptor molecular es el resultado final de una lógica y de un procedimiento matemático que transforma la información química codificada dentro de una representación simbólica de una molécula en un número útil o el resultado de un cierto experimento estandarizado. [Todeschini, 2000] Los descriptores moleculares son números capaces de proveer datos teóricos y experimentales de la molécula. Aunque las propiedades moleculares no dependen solamente de la composición sino también de la conectividad, la molécula contiene implícita toda la información química, pero solamente una parte de ésta se puede extraer experimentalmente y el resto de manera teórica. El descriptor D_{CL} es un descriptor teórico que se obtiene a partir de la distancia de la molécula considerando los grupos extremos (NH_2-NH_2 , o los que forman puentes de hidrógeno con el

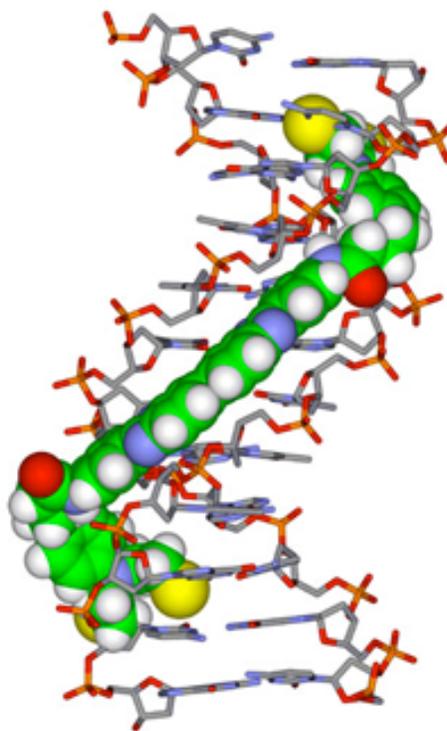
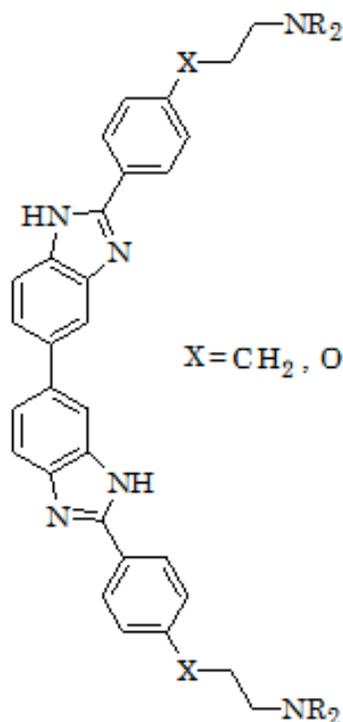


Figura 1. Representación de un reconocedor de surco y su interacción con el DNA

DNA), dividido por 4.202 o $3.4/\cos 36$. El resultado debe ser en principio el número de pares de bases más uno, que la molécula reconozca. Por lo tanto, si el número entero más cercano a este valor es restado al cociente de la división anterior, la diferencia que resulta será la separación del reconocimiento relativo, el cual es óptimo cuando es cero o inadecuado cuando es 0.5. [Chacón-García, 2001]

$$D_{CL} = |(Distancia\ entre\ NH_2-NH_2/4.202) - N| \quad \text{Ecuación 2.}$$

Donde N es el número entero más cercano.

Dado lo anterior, resulta interesante explorar si el descriptor D_{CL} tiene aplicación a compuestos diferentes a los experimentados sobre todo con actividad biológica demostrada en DNA y así buscar compuestos novedosos con mayor actividad y menor toxicidad. Un parámetro de interacción de moléculas con el DNA es la T_m , es decir la temperatura a la cual un biopolímero se desnaturaliza en un 50%. Un incremento en la T_m representa mayor estabilidad del biopolímero. En el caso del DNA, la T_m se ve afectada al interactuar con intercaladores o RS, mismos que estabilizan la doble hélice.

En el presente trabajo se describe el estudio QSAR de una serie de reconocedores de surco con respecto a la T_m en DNA mediante Inteligencia Artificial utilizando AGs e incorporando en el método el descriptor D_{CL} diseñado ex profeso para el estudio de este tipo de compuestos.

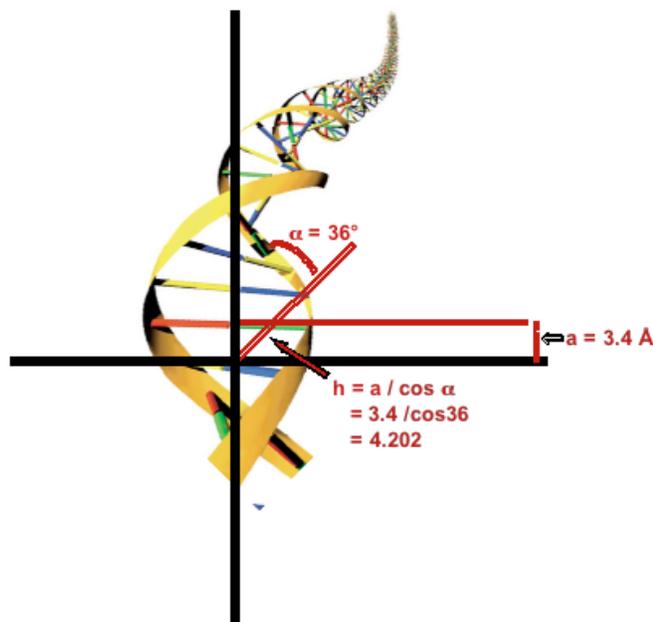


Figura 2. Esquema de las características del DNA que dan lugar al descriptor D_{CL} .

Materiales y métodos

La serie de 27 compuestos reconocedores de surco de DNA complementarios al dodecámero de Dickerson Drew (DDD) y la T_m que inducen en el DNA se obtuvieron de una base de

datos descrita en la literatura. [De Oliveira, *et al.*, 2003] La optimización geométrica de cada estructura se llevó a cabo con el programa Hyper Chem. [Hyper Chem, v. 6.03 Professional, 2000] Las estructuras primeramente fueron modeladas por mecánica molecular y posteriormente se optimizó su geometría mediante el método semiempírico AM1 que es el más comúnmente utilizado en moléculas orgánicas pequeñas con las características estructurales de las aquí estudiadas. Para la optimización geométrica se tomó como base el algoritmo Polak-Riviere [Press, *et al.*, 1986], con parámetros de 0.1 kcal/(\AA mol). Se procedió a obtener 1664 descriptores moleculares para cada uno de los compuestos, agrupados en 20 familias, con ayuda del programa DRAGON [Todeschini, *et al.*, 2005] y a estos se incorporó el descriptor D_{CL} que se obtuvo de acuerdo a su reporte en la literatura [Chacón-García, *et al.*, 2001]. Los valores de cada descriptor fueron importados al programa MobyDigs [Todeschini, *et al.*, 2004], por familias de descriptores y fueron “evolucionadas”, utilizando Algoritmos Genéticos. La evolución se detuvo tomando como criterio 700 generaciones o bien cuando los primeros 50 modelos fuesen estables. Debido a que el programa solamente respalda 10 familias, el análisis se realizó en dos partes y se unieron cuando ambas llegan al criterio de parada. Cuando ambas partes alcanzaban las generaciones apropiadas, se evolucionaron juntas, para combinar los 50 mejores modelos de ambas y con ello obtener resultados más confiables.

Los mejores 50 modelos finales, fueron validados estadísticamente tomando como base R^2 , Q^2 , F y s, que arroja el programa mismo y que sirvieron de base para hacer predicción de T_m .

Utilizando Excel™ se hace la validación externa de cada modelo, utilizando el coeficiente de correlación y descartando aquellos modelos que mostraban correlación entre sus descriptores.

En el caso del coeficiente de correlación entre descriptores, se consideró como criterio un máximo de 0.5, excluyendo aquellos conjuntos que presentaron un valor superior a éste entre sus descriptores.

Con la finalidad de validar la capacidad predictiva del modelo, se extrajo de la serie de datos el compuesto H y se le calculó la actividad con los modelos obtenidos.

Resultados

Los mejores 50 modelos obtenidos después de 700 evoluciones en la primer etapa y aprox. 1200 en la segunda, se muestran en la **Tabla 1** en la que se aprecia que el descriptor D_{CL} fue arrojado en los conjuntos 39 y 48. Del conjunto 39 se obtiene el Modelo 1. (vide supra)

La matriz de correlación de los descriptores involucrados en los conjuntos 39 y 48 se presenta en la **Tabla 2**.

En el caso del conjunto 39, existe una correlación mayor a 0.5, entre los descriptores RDF095 y E1e, por lo que se fragmentó a tres descriptores explorando dos conjuntos nuevos descartando RDF095 y E1e en cada caso dando lugar a los Modelos 2 y 3.

Modelo 1:

$$\log(\Delta T_m) = 0.0012 G(N..N) - 0.3368 DCL + 0.0371$$

$$n = 27, \quad r^2 = 0.8349, \quad F = 58.1512,$$

$$s = 0.0469, \quad Q^2 = 80.98$$

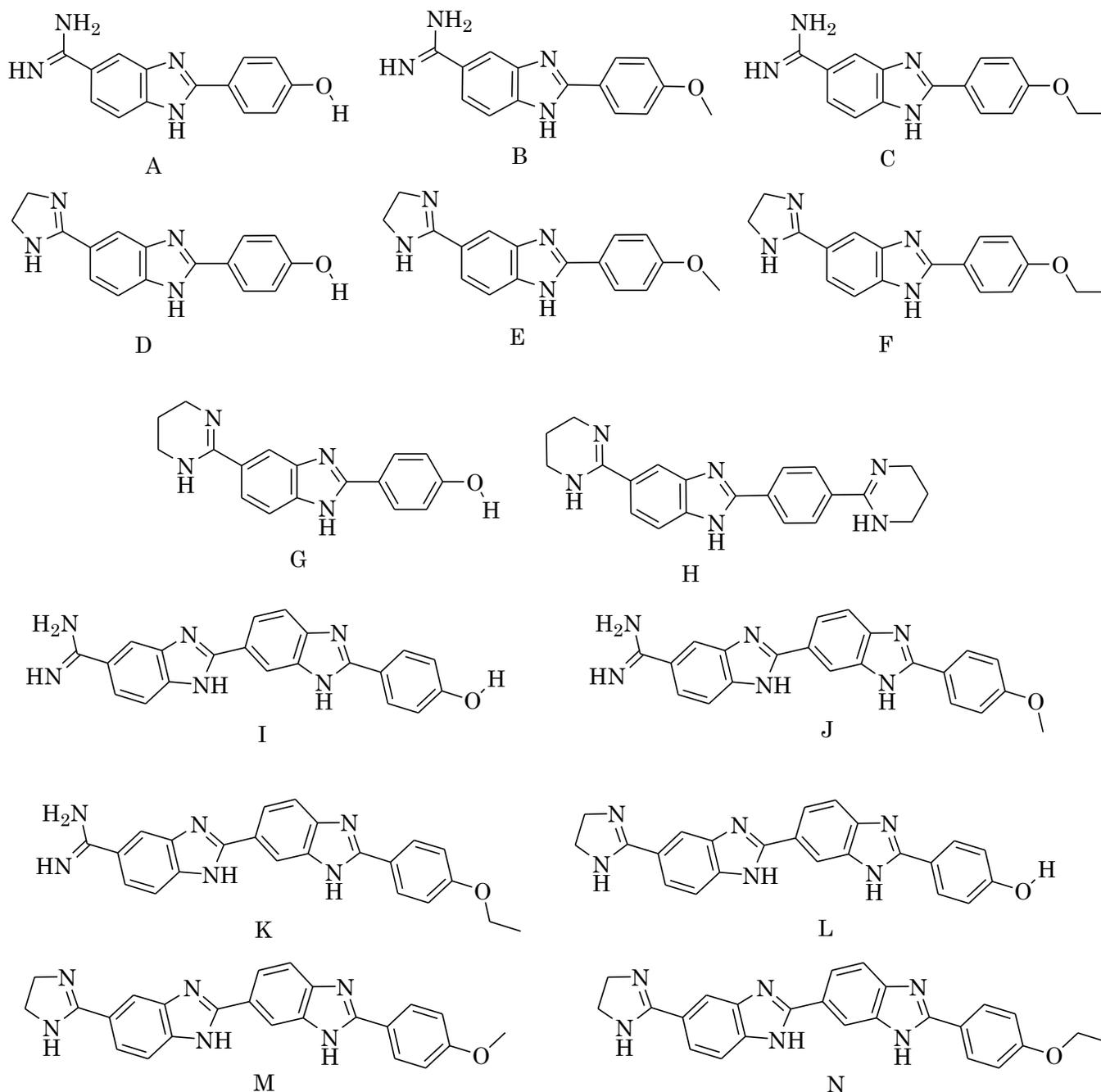


Figura 3. Estructura de los compuestos que conforman la serie de exploración.

Modelo 2:

$$\log(\Delta T_m) = 0.0012 G(N..N) - 0.0004 RDF095u - 0.3374 DCL + 0.0403$$

$$n = 27, \quad r^2 = 0.8352, \quad F = 37.1585, \\ s = 0.0479, \quad Q^2 = 80.79$$

Modelo 3:

$$\log(\Delta T_m) = 0.0011 G(N..N) + 0.8935 E1e - 0.3567 DCL - 0.4731$$

$$n = 27, \quad r^2 = 0.8970, \quad F = 63.8463, \\ s = 0.0379, \quad Q^2 = 87.57$$

La T_m calculada para cada compuesto se comparó con la experimental. Los valores de los descriptores, su respectiva T_m calculada y experimental se resume en la **Tabla 3**.

Discusión

La T_m está ligada de manera directamente proporcional a la estabilidad de la doble hélice del DNA y la interacción estabilizante de un fármaco con su blanco molecular incrementa sus posibilidades de coexistir en el ambiente celular. De ahí que entre mayor sea la T_m en los compuestos que interactúan con el material genético estos tendrán potencialmente actividad superior

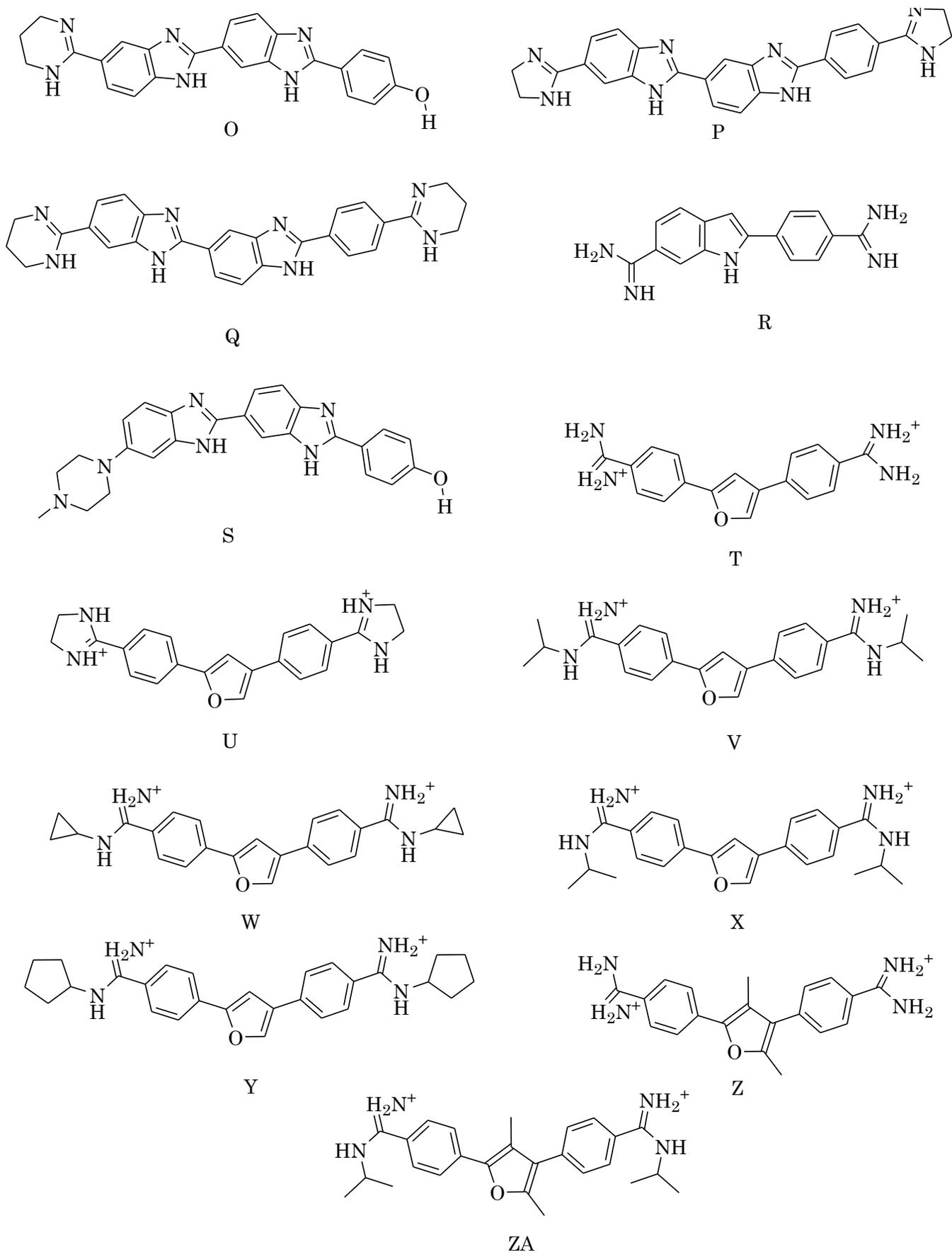


Figura 3. Estructura de los compuestos que conforman la serie de exploración (continuación).

Tabla 1. Modelos obtenidos con mejor ajuste estadístico

Conj. No.	Descriptores involucrados.	R ²	Conj. No.	Descriptores involucrados.	R ²
1	D/D ATS1m EEig02r GGI9	95.24	26	IDDE EEig02r GGI5 GGI8	94.94
2	Xu ATS1m EEig02r GGI9	95.51	27	EEig02r GGI5 GGI8	94.41
3	ATS2m MATS8e MATS8p GATS1p	95.35	28	EEig02r GGI5 GGI8 JGI3	94.8
4	EEig02r GGI4 GGI5 GGI8	95.32	29	piPC01 MATS8v MATS8e GATS1p	94.43
5	EEig02r GGI5 GGI8 LP1	95.27	30	EEig02r GGI5 GGI8 AEigZ	94.68
6	piPC01 MATS8e MATS8p GATS1p	95.13	31	EEig02r GGI5 GGI8 AEigm	94.68
7	ATS1m EEig02r GGI9 AEigm	95	32	EEig02r GGI5 GGI8 Eig1Z	94.68
8	ATS1m EEig02r GGI9 AEigZ	95	33	MATS8p EEig02r GGI5 GGI8	94.56
9	EEig02r GGI5 GGI8 H-048	95.14	34	EEig13x EEig02r GGI5 GGI8	94.51
10	EEig02r GGI5 GGI8 C-033	95.14	35	X0Av EEig02r GGI5 GGI8	94.68
11	ATS1m EEig02r GGI9 Eig1Z	94.99	36	VDA EEig02r GGI5 GGI8	94.67
12	EEig02r ESpm09u GGI5 GGI8	95.03	37	nC EEig02r GGI5 GGI8	94.56
13	EEig02r ESpm10u GGI5 GGI8	95.05	38	D/D EEig02r GGI5 GGI8	94.63
14	IVDM MATS8v MATS8e GATS1p	94.74	39	G(N..N) RDF095u E1e DCL1	93.26
15	EEig02r ESpm07u GGI5 GGI8	94.86	40	RBN EEig02r GGI5 GGI8	94.56
16	EEig02r GGI5 GGI8 E1e	95.12	41	GATS1p EEig02r GGI5 GGI8	94.55
17	EEig02r GGI5 GGI8 VEA1	94.98	42	X1v EEig02r GGI5 GGI8	94.53
18	EEig02r ESpm08u GGI5 GGI8	94.84	43	IDDM TIC0 GATS7v	92.78
19	EEig02r GGI3 GGI5 GGI8	94.97	44	IAC IDDM GATS7v	92.78
20	ATS1m MATS8v MATS8e GATS1p	94.56	45	GATS7v G(N..N) Ui	87.02
21	ATS1m MATS8e MATS8p GATS1p	94.72	46	GATS7v EEig02r	86.42
22	EEig02r GGI5 GGI8 C-026	94.8	47	E1u nCbH C-002 Ui	89.15
23	EEig02r GGI5 GGI8 C-040	94.8	48	G(N..N) DCL1	83.49
24	EEig02r GGI5 GGI8 E1u	95.04	49	GATS7v Ui	84.58
25	EEig02r GGI5 GGI8 nConj	94.81	50	nCar C-033 Ui	84.11

lo cual, claro está, sin considerar otros factores como absorción o metabolismo del fármaco. La importancia del descriptor D_{CL} radica justamente en este punto.

A partir de un análisis de regresión lineal simple, la relación entre D_{CL} y la T_m no resulta estadísticamente significativa por lo que se infiere que la geometría del reconocimiento por puentes de hidrógeno en función de su helicoicidad, como único factor a considerar, no es suficiente para explicar este parámetro termodinámico. Sin embargo, en los Modelos 1 a 3 se aprecia que tanto R^2 , F , s , y Q^2 son estadísticamente significativos y que al menos considerando los 1665 descriptores involucrados en la

búsqueda de modelos, que permitan explicar y predecir la T_m de los reconocedores de surco considerados en este estudio, hay otros factores que pueden afectar el reconocimiento y la estabilidad de los compuestos de la serie explorada y el DNA.

En el caso del Modelo 1, con base al coeficiente de determinación se establece que el 83.5% de la variabilidad de T_m es explicada tanto por el descriptor D_{CL} como por el descriptor geométrico $G(N..N)$ que considera la distancia entre nitrógenos, de existir dos, lo cual resulta razonable pues el adecuado reconocimiento geométrico es esencial en la formación de la supramolécula. A diferencia del Descriptor D_{CL} , el descriptor $G(N..N)$ no involucra la helicoicidad del DNA y la distancia entre nitrógenos se obtiene a partir de la estructura geoméricamente optimizada al vacío que no necesariamente corresponde a la conformación en su sitio de acción.

El Modelo 2 involucra, además de los descriptores que participan en el Modelo 1, el descriptor teórico RDF095u; que forma parte de los descriptores de función de distribución radial (del cual derivan sus siglas: Radial Distribution Function).

El término exponencial incluido en la ecuación que define la función RDF es $(R - r_{ij})^2$, donde r_{ij} es la distancia entre el

Tabla 2. Correlación entre los descriptores involucrados en los modelos 1- 3.

	G(N..N)	RDF095u	E1e	DCL
G(N..N)	1	0.44	0.09	-0.23
RDF095u		1	0.53	-0.15
E1e			1	0.1
DCL				1

Tabla 3. Comparación de Tm calculada y experimental y valores de los descriptores moleculares utilizados en los modelos 1-3.

Mol	Valores de Descriptores				Modelo 1		Modelo 2		Modelo 3	
	G(N..N)	E1e	RDF095u	DCL	Tm experimental	Tm calculada	Tm experimental	Tm calculada	Tm experimental	Tm calculada
A	24.316	0.575	4.758	0.488	0.845	0.81	0.845	0.7994	0.845	0.783
B	23.889	0.562	7.728	0.406	0.792	0.813	0.792	0.8488	0.792	0.814
C	24.317	0.554	7.057	0.499	0.69	0.812	0.69	0.791	0.69	0.743
D	24.437	0.574	4.632	0.476	0.748	1.399	0.748	0.8072	0.748	0.789
E	24.514	0.557	6.465	0.467	0.875	1.263	0.875	0.8118	0.875	0.768
F	24.571	0.556	2.24	0.462	0.845	1.399	0.845	0.8181	0.845	0.77
G	24.118	0.58	6.573	0.462	0.845	1.382	0.845	0.814	0.845	0.808
I	95.084	0.578	11.616	0.009	1.362	1.391	1.362	1.3996	1.362	1.4
J	95.192	0.582	8.696	0.141	1.362	1.339	1.362	1.2668	1.362	1.267
K	95.169	0.567	13.052	0.0088	1.362	1.25	1.362	1.3984	1.362	1.369
L	95.612	0.582	11.051	0.026	1.378	1.54	1.378	1.3839	1.378	1.394
M	95.454	0.584	16.137	0.018	1.387	1.487	1.387	1.3857	1.387	1.408
N	95.604	0.573	11.548	0.067	1.412	1.118	1.412	1.3399	1.412	1.323
O	94.905	0.587	12.857	0.153	1.405	1.346	1.405	1.2495	1.405	1.266
P	250.12	0.579	15.269	0.426	1.479	0.976	1.479	1.5434	1.479	1.482
Q	248.159	0.595	17.576	0.464	1.487	0.994	1.487	1.4874	1.487	1.477
R	78.287	0.566	5.165	0.239	1.233	0.969	1.233	1.1243	1.233	1.083
S	98.504	0.583	11.24	0.071	1.253	0.884	1.253	1.3468	1.253	1.356
T	59.421	0.584	7.935	0.349	0.959	0.868	0.959	0.9778	0.959	0.978
U	59.678	0.584	8.019	0.326	0.857	0.951	0.857	0.996	0.857	0.998
V	59.322	0.621	12.935	0.358	1.037	0.974	1.037	0.9664	1.037	1.048
W	59.06	0.657	15.291	0.475	1.037	0.946	1.037	0.88	1.037	1.024
X	58.745	0.611	12.924	0.498	0.968	1.352	0.968	0.8655	0.968	0.913
Y	59.085	0.679	22.585	0.381	1.113	0.81	1.113	0.9406	1.113	1.157
Z	59.719	0.524	8.178	0.353	0.813	0.813	0.813	0.9753	0.813	0.862
ZA	59.311	0.565	25.011	0.388	0.732	0.812	0.732	0.9341	0.732	0.911
H	105.089	0.599	12.542	0.088	1.233	1.399	1.233	1.3518	1.233	1.405

par de átomos *i, j* y *R* es la distancia de referencia en intervalos de 1.0 a 15.5 Ang. Esto significa que todos los pares de átomos con distancias geométricas cerca de *R* juegan un papel significativo en el descriptor. Este descriptor puede estar ponderado por masas atómicas, electronegatividad, polarizabilidad, etc. En este caso, RDF095u, establece a *R* como 9.5 Ang y la sigla *u* indica que no está ponderado por ningún término en particular (unweighted). Esta información se resume a que las distancias geométricas dentro de la estructura en exactamente 9.5 Ang juegan un papel importante en el reconocimiento molecular y por lo tanto en la estabilidad del complejo DNA-Reconocedor de surco formado. Del Modelo 2, se aprecia que RDF095u tiene una contribución negativa y por lo tanto entre mayor es este componente numéricamente, contribuye disminuyendo a la *Tm*. Sin embargo, el coeficiente de RDF095u en el modelo es de 0.0004, lo cual se traduce a una contribución matemáticamente insignificante y los datos calculados a partir del modelo con

respecto a los experimentales son prácticamente iguales a los que respecta al Modelo 1.

Por su parte, el Modelo 3, que al igual que el Modelo 2 tiene el mismo origen en su conjunto de datos obtenido por el proceso de evolución, incluye al descriptor E1e que forma parte de los descriptores conocidos como WHIM (weighted holistic invariant molecular descriptors). Estos son índices moleculares 3D que representan diferentes fuentes de información química en términos de tamaño, forma, simetría y distribución atómica y son obtenidos a partir de sus coordenadas de la estructura tridimensional con diferentes esquemas de ponderación. El descriptor E1e está ponderado en las electronegatividades de Sanderson (que se fundamenta en el recíproco del volumen atómico). Es interesante este resultado pues la *Tm* está definida, como se mencionó anteriormente, por interacciones estabilizantes dentro de las cuales se encuentran las interacciones dipolo-dipolo y estas a su vez están relacionadas directamente con la diferencia

de electronegatividad de las especies que interactúan. En este caso la variabilidad del modelo del 89.7 % es explicada por los descriptores D_{CL} , $G(N..N)$ y $E1e$, es decir por aspectos de helicidad, geométricos y de electronegatividad.

Aunque la capacidad predictiva de los modelos 1 a 3 es alta ($Q^2 = 80.98, 80.79, 91.31$, respectivamente) y la validación externa con el compuesto **H** así lo demuestra, el análisis de regresión mostró ser ligeramente superior en el Modelo 3, mismo que explica de manera más amplia la interacción con el DNA no solo considerando aspectos geométricos sino además electrónicos. La contribución del descriptor D_{CL} es importante en los tres modelos y abre la posibilidad de utilizarse, en conjunto con el descriptor $E1e$ y $G(N..N)$ para predecir la T_m de compuestos potencialmente reconocedores de surco que no han sido explorados en QSAR, por ejemplo compuestos peptídicos.

Agradecimientos

Con apoyo económico de CIC proyecto 2.18

Referencias

- Bewley, Carole A.**, Gronenborn Angela M. y Clore Marius G. (1998). *Minor Groove – Binding. Architectural proteins: Structure, Function, and DNA Recognition. Annual Review of Biophysics and Biomolecular Structure* 27: 105-31.
- Chacón-García Luis, Martínez Roberto.**(2001) Cytotoxic activity and QSAR of N,N' -diaryllkanediamides. *European Journal of Medicinal Chemistry* 36: 731-736.
- De Oliveira, A. M.**; Custodio, F. B.; Donnici, C. L.; Montanari, C. A. (2003) *European Journal of Medical Chemistry* 38: 141-155.
- Estévez Valencia Pablo.** (1997). Optimización mediante algoritmos genéticos. *Anales del Instituto de Ingenieros de Chile. Agosto. pp. 83-92.* [http://www.inele.ufro.cl/apuntes/Tutores Inteligentes/Pablo%20Estevez%20-%20Optimizacion%20Mediante%20AGs.pdf](http://www.inele.ufro.cl/apuntes/Tutores%20Inteligentes/Pablo%20Estevez%20-%20Optimizacion%20Mediante%20AGs.pdf)
- Hou, T.J.**, Wang J.M. y Xu X.J. (1999). *Application of genetic algorithms on the structure – activity correlation study of a group of non-nucleoside HIV – 1 inhibitors. Chemometrics and intelligent laboratory systems* 45: 303-310.
- Hyper Chem 6.03** for Windows; HyperCube Inc., 2000.
- Koga Hiroshi, Itoh Akira, Murayama Satoshi, Suzue Seigo y Irikura Tsutomu.** (1980). Structure- Activity Relationships of Antibacterial 6,7-and 7,8- Disubstituted 1-Alkyl-1,4-dihydro-4-oxoquinoline-3-carboxylic Acids. *European Journal of Medical Chemistry* 23: 1358-1363.
- Larrañaga Pedro e Inza Iñaki.** Tema 2. Algoritmos genéticos. *Departamento de Ciencias de la computación e Inteligencia Artificial. Universidad del País Vasco – Euskal Herriko Unibertsitatea.* <http://www.redesr.com/~talos/t2geneticos.pdf>
- Martínez R.** y Chacón-García L. (2005) *Current Opinion in Medicinal Chemistry.* 12: 127-151.
- Neidle Stephen.** (2001). DNA minor-groove recognition by small molecules. *Natural Product Reports* 18: 291-309.
- Pastor M.** y Alvarez – Builla J. Técnicas QSAR en diseño de fármacos. *Departamento de Química Orgánica. Universidad de Alcalá. E-28871 Alcalá de Henares. Madrid.*
- Press W, H.**, et. al., (1968). *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, Chapter 10.
- Software Microsoft Office Excel 2007.** Para Windows. Microsoft Corporation. 2006.
- Stefanic-Petek Alenka, Krbavcic Ales, Solmajer Tom.** (2002). QSAR of flavonoids: differential inhibition of aldose reductase and p56lck protein tyrosine kinase. *Croatica Chemica Acta. CCACAA* 75 (2) 517-529.
- Suh Myung-Eun, Park So-Young y Lee Hyun-Jung.** (2002). Comparison of QSAR Methods of Anticancer 1-N-Substituted Imidazoquinoline-4,9-dione Derivatives. *Bull. Korean Chemical Society. Vol. 23, No. 3,* 417
- Todeschini Roberto y Consonni V.** (2000). Manual de descriptores moleculares, Wiley-VCH. <http://www.moleculardescriptors.eu/books/handbook.htm>
- Todeschini R.**; Ballabio, D.; Consonni, V.; Mauri, A.; Pavan, M. (2004) Mobydigs Computer Software, 1.0; TALETE srl: Milano.
- Todeschini R.**, Consonni, V.; Mauri, A.; Pavan, M. (2005). DRAGON—Software for the calculation of molecular descriptors. Ver. 5.3, Talete srl, Milano, Italy.
- Turabekova Malakhat A.** y Rasulev Bakhtiyor F. (2004). *A QSAR Toxicity of a Series of Alkaloids with the Lycoctonine Skeleton. Molecules* 9: 1194-1207.
- Yalcin Ismail, Ören Ilkay, Temiz Özlem y Aki Sener Esin.** (2000). QSARs of some novel isosteric heterocyclics with antifungal activity. *Acta Biochimica Polonica.* 2: 481-486.