

# Estudio de la actividad inhibidora de acetilcolinesterasa de derivados de la *Huperzia* mediante técnicas QSAR y clasificación KNN

---

Piercosimo Tripaldi,<sup>1</sup> González Silvia<sup>2</sup> y Jiménez Vanessa<sup>2</sup>

## Introducción

El Alzheimer (AD) es un tipo de demencia que en la última época ha saltado a los medios de comunicación. Es una enfermedad neurodegenerativa progresiva que se caracteriza por una serie de rasgos clínicos y patológicos (Donoso, 2003).

El AD aumenta con la edad: se puede presentar del 1 al 2% en sujetos de 60 años, del 3 al 5% en sujetos de 70 años, del 15 al 20% en sujetos de 80 años, en un tercio y la mitad de los sujetos mayores a 85 años. Otros autores (Davidsson *et al.*, 2001; Shadlen y Larson, 1999; Benzi, y Moretti, 1998; Hasegawa, 1998) afirman que la pérdida de memoria, se debe a la pérdida de neuronas colinérgicas en diferentes áreas del sistema nervioso central, especialmente en la corteza y en el hipocampo. Según Sahakian (*et al.*, 1989) el tratamiento para esta afección es la utilización de medicamentos que bloquean el enzima acetilcolinesterasa (AChE) la cual es responsable de la hidrólisis de la acetilcolina. Existe un gran variedad de antagonistas colinérgicos que han demostrado aumentar la memoria de los hombres y de los animales. Sin embargo, los tratamiento contra el AD con antagonistas colinérgicos no han demostrado (Kozikowski *et al.*, 1996), en general, ser útil terapéuticamente en cuanto presentan muchas reacciones colaterales.

Un inhibidor reversible de la AchE de origen natural que es muy interesante por sus posibles aplicaciones en el tratamiento de la AD es la huperzina A. Esta molécula ha sido aislada por primera vez de la *Huperzia Serrata* por Liu y colaboradores (Liu, 1988). Este producto natural ha demostrado ser un potente y selectivo inhibidor natural de la AchE mientras que no ha presentado actividad sobre otros receptores y sistemas enzimáticos. Los estudios sobre esta clase de moléculas se han desarrollado mucho sea buscando sintetizar análogos de la huperzina A sea investigando otras posibles fuentes de esta familia de alcaloides. De manera que se han encontrado que también en licopodiáceas de otros continentes está presente, entre demás alcaloides, la huperzina A. Una de las más ubicua especie de *Huperzia* es la *Huperzia Saururus* (Ortega *et al.*, 2007) que se encuentra en la zona andina de Sudamérica en Madagascar y en las islas Mascareñas. En el Ecuador se encuentra además la *Huperzia Loxensis*:

---

1 Universidad del Azuay, Laboratorio de Análisis Ambientales, Cuenca-Ecuador.

2 Universidad técnica Particular de Loja.

Figura 1  
*Huperzia Loxensis*



Existen otras variedades de *Huperzia*, que son también difundidas, por ejemplo, la *Lechleri*, *Lindeni*, *Lilinfolia* (tenuifolia), etc. Todas las especies, presentan, (Benzi *et al.*, 1998) además de la Huperzina A también otros alcaloides.

La metodología QSAR (Quantitative Structure Activity Relationship) (Goldberg, 1989; Wehrens y Buydens, 1998; Todeschini y Consonni, 2000) es útil para predecir la actividad de nuevas moléculas mediante modelos matemáticos construidos utilizando características químicas estructurales, oportunamente digitalizadas y la actividad biológica de una base de datos de moléculas previamente estudiadas. El gran problema que se presenta en estos tipos de estudio es que difícilmente se alcanza un correcto balance entre moléculas y descriptores. El número de descriptores es casi siempre mucho más alto que el número de moléculas disponibles. Por esta razón es imperativo utilizar algoritmos que permitan la reducción de la dimensionalidad de la base de datos sin afectar la información útil. Otro problema que puede presentarse es que las actividades biológicas de la literatura pueden no ser reproducibles por lo tanto, a menudo, cuando se construyen bases de datos combinando resultados publicados, la calidad de los modelos continuos resulta escasa. Para superar esto se ha optado por utilizar modelos discretos, agrupando en clases de actividad las moléculas, de esta manera se ha intentado minimizar el efecto de la poca reproducibilidad.

## Metodología

La moléculas consideradas en este estudio han sido 58 obtenidas de diferentes bases de datos; 15 moléculas se obtuvieron del trabajo de Bai (*et al.*, 2000) las demás 26 se obtu-

vieron de Kamalakaran Anand Solomon (*et al.*,2009), una del trabajo de Guang (*et al.*,2002), dos del trabajo de Haviv (*et al.*,2007) y diez del trabajo de Shi (*et al.*,2009). En el anexo 1 se presentan las estructuras de las moléculas. En la tabla 1 se reportan las actividades (IC50) de las moléculas, el Ln (1/IC50) y el cuartil de actividad de pertenencia (clase de actividad):

Tabla 1  
Actividades de las moléculas estudiadas

Molécula	IC50(micrM)	LN(1/IC50)	clase	Molécula	IC50(micrM)	LN(1/IC50)	clase
1	0.0631	2.7630	4	29	0.65	0.4308	3
2	158	-5.0626	1	30	0.453	0.7919	3
3	316	-5.7557	1	31	0.981	0.0192	2
4	0.2	1.6094	4	32	0.935	0.0672	2
5	3.16	-1.1506	1	33	0.906	0.0987	2
6	3160	-8.0583	1	34	0.921	0.0823	2
7	7.94	-2.0719	1	35	0.872	0.1370	2
8	25.1	-3.2229	1	36	1.262	-0.2327	2
9	1.58	-0.4574	1	37	0.965	0.0356	2
10	0.631	0.4604	3	38	0.84	0.1744	2
11	1.58	-0.4574	1	39	0.6	0.5108	3
12	0.891	0.1154	2	40	0.692	0.3682	3
13	1	0.0000	2	41	30.8	-3.4275	1
14	0.082	2.5010	4	42	16.6	-2.8094	1
15	0.285	1.2553	3	43	0.73	0.3147	3
16	0.437	0.8278	3	44	3.48	-1.2470	1
17	0.642	0.4432	3	45	0.455	0.7875	3
18	0.748	0.2904	3	46	1.995	-0.6906	1
19	0.818	0.2009	3	48	19.3	-2.9601	1
20	1.23	-0.2070	2	49	4.9	-1.5892	1
21	0.926	0.0769	2	50	14.2	-2.6532	1
22	0.91	0.0943	2	51	0.0752	2.5876	4
23	1.549	-0.4376	2	52	0.0297	3.5166	4
24	0.158	1.8452	4	53	0.21	1.5606	4
25	0.477	0.7402	3	54	0.0142	4.2545	4
26	0.217	1.5279	3	55	0.0168	4.0864	4
27	0.484	0.7257	3	56	0.189	1.6660	4
28	0.448	0.8030	3	57	0.0555	2.8914	4
				58	0.105	2.2538	4
				59	0.0985	2.3177	4

Se han digitalizado las moléculas utilizando el programa Hyperchem (Hyperchem.2000) y se han calculado mediante el programa Dragón (Talet,2006) 4000 descriptores de los cuales se han retenido solo los que presenten desviación estándar superior a 0.0001 para eliminar los descriptores constantes o “cuasi” constantes y entre pares de descriptores que presenten correlación superior a 0.95 para no considerar la información redundante. En total se han considerado 835 descriptores. La respuesta biológica de cada molécula (IC50 anti-ChE en micromoles determinada sobre la membrana de los eritrocitos de la ratas ha sido transformada en el correspondiente ln (1/IC50).

Se han subdivididos los valores de ln (1/IC50) en cuatro clases de actividad en función de los cuartiles como se muestra en la tabla 2:

Tabla 2  
Subdivisión en cuatro clases de actividad de los In (1/IC50)

<b>Cuartil</b>	<b>Clase de actividad</b>	<b>Resp.Maxima</b>
<b>1cuart</b>	<b>baja</b>	<b>-0.171</b>
<b>2cuart</b>	<b>medio-baja</b>	<b>0.140</b>
<b>3cuart</b>	<b>medio-alta</b>	<b>0.682</b>
<b>4cuart</b>	<b>alta</b>	<b>1.887</b>

El modelo ha sido construido utilizando el método de clasificación KNN (Ballabio, Todeschini, 2009) con selección contemporánea de los descriptores mejores mediante la técnica de los algoritmos genéticos en la versión de Leardi e colaboradores (Leardi, 1992). El data set se ha previamente autoescalado y subdividido casualmente en un training set (54 moléculas) y test set externo formado por 4 moléculas que aproximadamente representan el 10% de las moléculas totales. En la aplicación del algoritmo genético se han utilizado las opciones reportadas en la tabla 3.

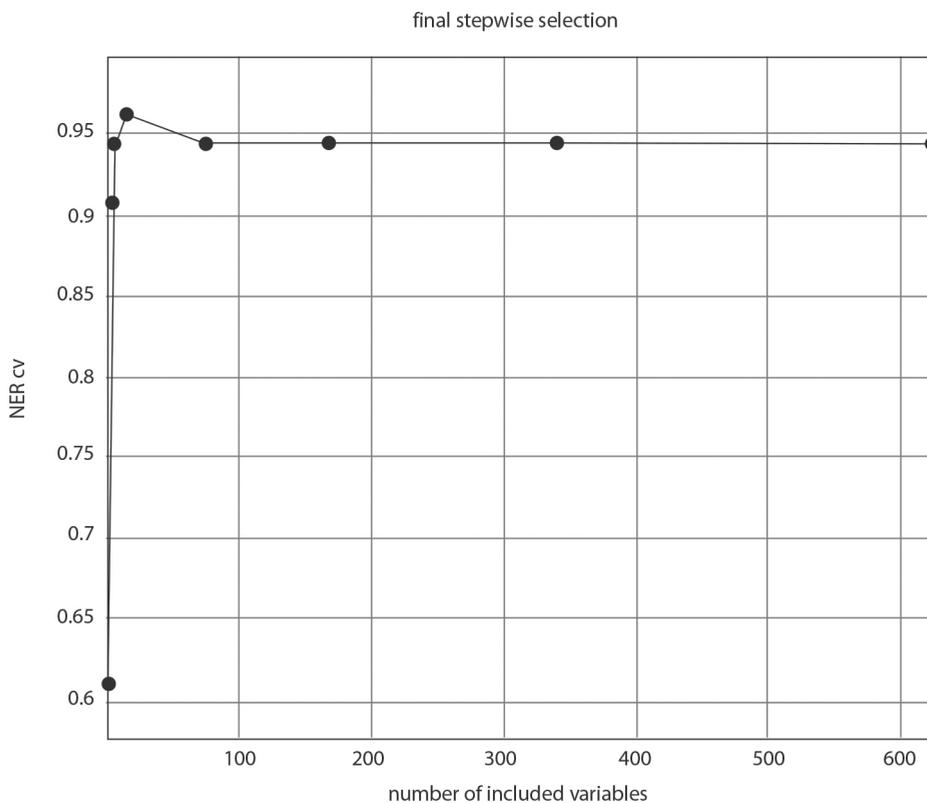
Tabla 3  
Condiciones de aplicación del algoritmo genético de selección de los descriptores moleculares

<b>Metodo:</b>	<b>K.N.N.</b>
<b>Escalado</b>	<b>Autoscaling</b>
<b>Grupos de cross-validación</b>	<b>5</b>
<b>Tipo de cross-validación</b>	<b>Ventanas venecianas</b>
<b>Número de cromosomas</b>	<b>30</b>
<b>Probabilidad de mutación</b>	<b>0.01</b>
<b>Probabilidad de cruce</b>	<b>0.5</b>
<b>Repeticiones</b>	<b>100</b>
<b>Número de Evaluaciones</b>	<b>250</b>
<b>Kernel</b>	<b>Lineal</b>

## Resultados

Una vez aplicado el método de selección de los descriptores al modelo de clasificación se han obtenido los siguientes resultados: el número óptimo de descriptores utilizados ha sido 16, en la figura 2 Se presenta el gráfico del número de descriptores en el modelo vs. el No Error Rate Cross, validado internamente con cinco grupos de extracciones.

Figura 2  
Variación del NER al aumentar del número de descriptores considerados



Se puede notar que luego de los primeros siete descriptores el NER no mejora en forma apreciable, en efecto con los primeros siete se obtiene un valor de NER de 0.944 mientras que con 16 descriptores se mejora solo de dos centésimos, es decir 0.963.

En la tabla 4 se presenta el set de los siete descriptores seleccionados, la clase de actividad experimental y la clase de actividad predicha para las moléculas del *training set*.

Tabla 4  
 Descriptores seleccionados para el training set,  
 clase de actividad experimental y clase predicha por KNN

Molec	'Eta_beta_A'	'PW3'	'JGI3'	'HATS4s'	'B02[C-O]'	'Eig10_EA(dm)'	'R7e+'	KNN_Class	KNN_CALC
6	-0.640	0.936	0.946	-0.429	0.267	-0.365	-0.417	1	1
3	0.161	1.197	0.638	-0.630	0.267	-0.365	-0.633	1	1
2	0.505	1.563	0.561	-0.588	0.267	-0.255	-0.633	1	1
38	-0.240	1.354	0.792	-0.435	0.267	-0.365	-0.633	1	1
7	-0.240	1.354	0.792	-0.441	0.267	-0.365	-0.676	1	1
39	0.161	0.779	0.407	-0.606	0.267	-0.365	-0.633	1	1
45	0.161	0.700	0.253	-0.661	0.267	0.756	-0.719	1	1
44	-0.125	0.596	0.330	-0.687	0.267	-0.177	-0.719	1	1
41	-1.156	0.413	0.176	-0.620	0.267	-0.242	-0.676	1	1
5	0.734	0.622	-0.055	-0.661	0.267	-0.197	-0.633	1	1
43	0.447	0.857	0.022	-0.618	0.267	-0.184	-0.848	1	1
8	0.447	0.491	-0.440	-0.658	0.267	-0.184	-0.762	1	1
10	-0.640	0.936	0.946	-0.429	0.267	-0.365	-0.417	1	1
21	-1.270	-1.548	-1.286	0.047	0.267	-0.365	-0.115	2	2
34	-1.614	-1.260	-0.517	0.147	0.267	-0.365	0.186	2	2
12	-1.270	-1.208	-0.901	-0.008	0.267	-0.365	-0.029	2	2
29	-1.270	-1.208	-0.901	0.131	0.267	-0.365	-0.029	2	2
30	-0.755	-1.260	-0.517	-0.135	0.267	-0.365	-0.158	2	2
19	-1.270	-1.548	-1.286	0.102	0.267	-0.365	0.272	2	2
32	-1.098	-1.208	-0.901	0.279	0.267	-0.365	-0.245	2	2
20	-0.240	-1.208	-0.901	0.265	0.267	-0.365	1.695	2	2
31	-1.098	-1.208	-0.901	0.114	0.267	-0.365	1.522	2	2
11	-1.098	-1.234	-0.440	0.026	0.267	-0.365	1.565	2	2
33	-1.270	-1.548	-1.286	0.027	0.267	-0.365	1.781	2	2
35	-1.270	-1.574	-0.055	0.054	0.267	-0.365	1.824	2	2
18	-1.270	-1.208	-0.901	0.092	0.267	-0.365	1.522	2	2
17	-1.270	-1.574	-0.055	0.006	0.267	-0.365	1.651	2	2
40	-1.270	-1.208	-0.901	-0.065	0.267	-0.365	1.307	2	2
57	-0.755	-0.973	-1.286	-0.212	0.267	-0.365	0.746	3	2
27	-1.270	-1.548	-1.286	0.167	0.267	-0.365	0.703	3	2
16	0.734	0.282	1.561	1.848	0.267	-0.365	0.186	3	3
9	1.077	0.361	1.023	1.883	0.267	-0.365	1.695	3	3
36	1.077	0.361	1.023	2.902	0.267	-0.365	2.169	3	3
25	1.592	0.700	0.869	1.805	0.267	-0.365	1.565	3	3
23	1.191	0.256	1.638	2.833	0.267	-0.365	1.608	3	3
42	0.619	0.518	1.484	1.673	0.267	-0.365	0.272	3	3
28	1.191	0.596	1.407	2.622	0.267	-0.365	0.402	3	3
26	0.962	0.936	0.946	-0.343	-3.677	-0.910	-0.288	3	3
15	1.306	1.093	2.100	-0.321	-3.677	-0.806	-0.417	3	3
14	1.306	1.093	2.100	-0.320	-3.677	-0.962	-0.417	3	3
24	-0.240	1.406	2.331	-0.310	-3.677	-0.365	-0.460	3	3
48	-0.354	0.622	-0.055	-0.769	0.267	3.649	-1.020	4	4
4	1.191	1.040	0.484	-0.122	0.267	-0.365	-0.374	4	1
51	1.420	0.596	-0.286	-0.783	0.267	5.516	-1.193	4	4
22	0.962	0.073	-0.670	-0.718	0.267	0.886	-0.805	4	4
53	0.562	-0.058	-0.747	-0.736	0.267	0.886	-0.805	4	4
54	0.161	-0.162	-0.824	-0.756	0.267	0.886	-0.891	4	4
13	-0.182	-0.241	-0.978	-0.765	0.267	0.886	-0.891	4	4
46	0.734	0.152	-0.517	-0.772	0.267	0.886	-1.020	4	4
1	0.333	0.021	-0.593	-0.786	0.267	0.886	-1.150	4	4
52	-0.011	-0.084	-0.670	-0.796	0.267	0.886	-1.107	4	4
47	1.134	-0.058	-0.747	-0.774	0.267	0.439	-1.063	4	4
50	0.734	-0.162	-0.824	-0.787	0.267	0.439	-1.063	4	4
49	0.333	-0.267	-0.901	-0.797	0.267	0.439	-1.063	4	4

El No Error Rate del modelo con siete descriptores ha sido de 0.944; la matriz de confusión del modelo se presenta a continuación en la tabla 4.

Tabla 5  
Matriz de confusión del modelo de clasificación KNN

	clase1	clase2	clase3	clase4
clase1	13	0	0	0
clase2	0	15	0	0
clase3	0	2	11	0
clase4	1	0	0	12

Otros parámetros de calidad del modelo se presentan a continuación:

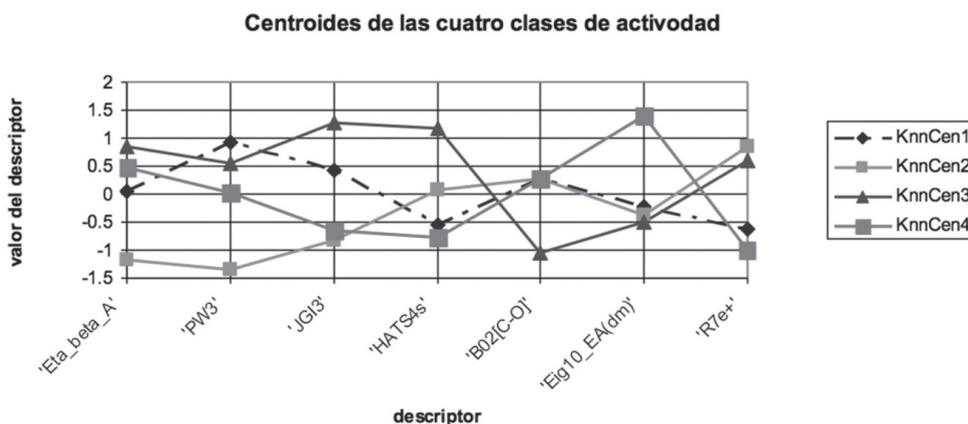
Tabla 6  
Parámetros adicionales de la calidad del modelo de clasificación KNN

Parámetro	Clase 1	Clase 2	Clase3	Clase 4
precisión	0.928571429	0.882353	1	1
sensibilidad	1	1	0.846154	0.923076923
especificidad	0.975609756	0.948718	1	1

Los parámetros de calidad adicionales indican un modelo generalmente bueno frente con algunas deficiencias leves en la precisión con la cual asigna las moléculas de la clase dos, presenta una leve falla en la sensibilidad para la clase tres mientras que la especificidad de las clase es bastante homogénea.

En la figura 3 se presenta el gráfico de los centroides de las cuatro clases de actividad. Este gráfico permite visualizar cuales son los valores promedio de los descriptores en cada clase de actividad y por lo tanto inferir su influencia en la actividad de una nueva molécula.

Figura 3  
Valores de los descriptores seleccionados en los centroides de las clases de actividad



De la precedente figura se desprende que la clase cuatro, que agrupa las moléculas con mayor actividad anti ACE, es caracterizada por valores altos del descriptores: 'Eig10\_EA (dm)', mientras tiene valores medio bajos de los demás descriptores. El descriptor Eig10\_AEA (dm) pertenece al grupo de los índices topológicos calculados a partir de la matriz de conectividad que resume la información sobre el número de enlaces entre pares de átomos (hidrógenos excluidos). En este caso la matriz de conectividad es la Matriz Aumentada que presenta en la diagonal parámetros específicos de cada átomo. El descriptor en estudio es el décimo autovalor de la matriz aumentada de conectividad. Finalmente el modelo con siete descriptores se ha aplicado a las moléculas del test set para predecir su actividad y contrastarla con la experimental. En la tabla cinco se presentan los resultados de la predicción externa

Tabla 6  
Predicción de la clase de actividad del test set externo

<b>Test _Externo</b>	<b>clase verd</b>	<b>clase asig</b>
<b>7</b>	<b>1</b>	<b>1</b>
<b>20</b>	<b>2</b>	<b>2</b>
<b>37</b>	<b>2</b>	<b>3</b>
<b>48</b>	<b>1</b>	<b>1</b>

La molécula 37 del test set externo ha sido mol clasificada por lo que el Ner en predicción con este esquema de training/test sets ha sido de 0.75.

## Discusión

El modelo obtenido presenta un muy buen poder predictivo, el relativamente bajo valor del Ner externo se debe al pequeño número de moléculas utilizadas. A diferencia de muchos modelos QSAR continuos encontrados en literatura por ejemplo: J.Kaur (2000); que propone modelos de regresión no validados, o Kalamalakaran (2009) que propone modelos validados internamente con una sola molécula a la vez, este modelo discreto presenta la ventaja de ser *cross* validado internamente con cinco grupos de cancelación para cada iteración por lo que su estabilidad es mucho más significativa que la de modelos continuos poco validados y de abarcar moléculas cuyas actividades han sido obtenidas por diferentes grupos de investigadores por lo que resulta ser más general.

## Conclusiones

El trabajo demuestra la utilidad de aplicar la metodología Qusra para predecir valores de actividad biológica sin aplicar técnicas complejas de crianza de animales y su sacrificio.

Su utilidad inmediata se centra en las ramas farmacéutica y fotoquímica para obtener informaciones rápidas sobre las propiedades curativas de las moléculas contenidas en las *Huperzias*, especilamente la *Huperzia Loxensis*.

Anexos

Figura 5  
Las moléculas obtenidas del trabajo de D. L. Bai han sido

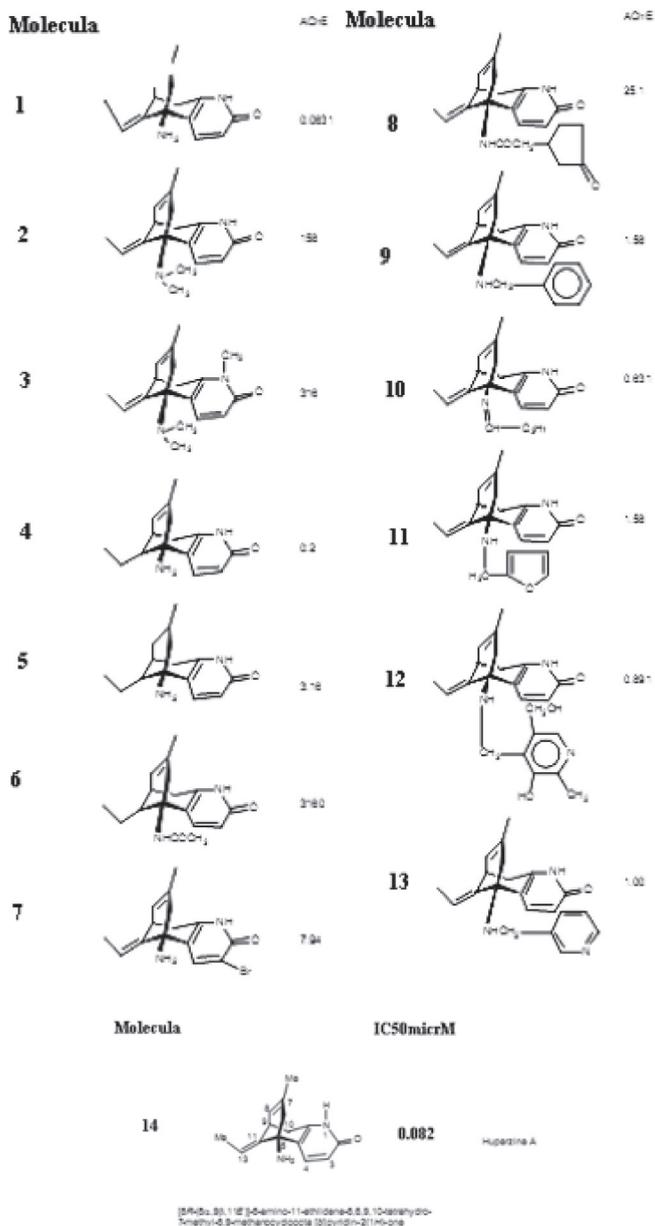


Figura 6  
Las moléculas obtenidas del trabajo de Kalamalakran han sido

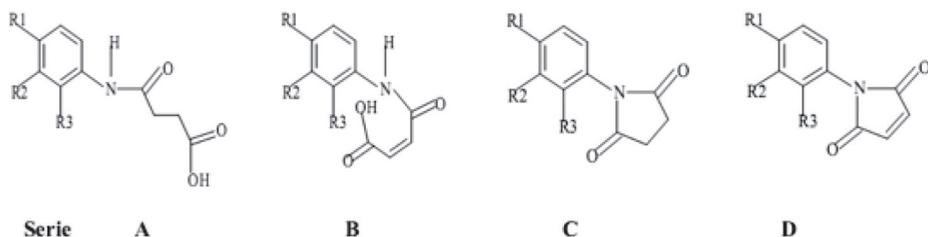


Figura 7  
La molécula obtenida del trabajo de Guang ha sido

Compd. No.	Substituents			IC50 micrMol
	R1	R2	R3	
15	NH <sub>2</sub>	H	H	0.235
16	H	CO	H	0.437
17	H	H	NH <sub>2</sub>	0.642
18	H	H	OH	0.748
19	H	OCH <sub>3</sub>	H	0.813
20	H	H	H	1.23
21	OH	H	H	0.925
22	CO <sub>2</sub> H	H	H	0.91
23	NO <sub>2</sub>	H	H	1.549
24	CO <sub>2</sub> H	H	H	0.158
25	H	CO <sub>2</sub> H	H	0.477
26	Cl	H	H	0.217
27	H	Cl	H	0.444
28	H	H	Cl	0.448
29	H	OH	H	0.65
30	H	H	NH <sub>2</sub>	0.453
31	H	H	OCH <sub>3</sub>	0.931
32	F	H	H	0.935
33	H	H	OH	0.906
34	H	CO <sub>2</sub> H	H	0.921
35	H	CO <sub>2</sub> H	H	0.872
36	H	H	OCH <sub>3</sub>	1.252
37	H	OH	H	0.955
38	H	NH <sub>2</sub>	H	0.84
39	NH <sub>2</sub>	H	CH	0.6
40	H	H	Cl	0.622

Figura 8  
Las moléculas obtenidas del trabajo de Haviv han sido

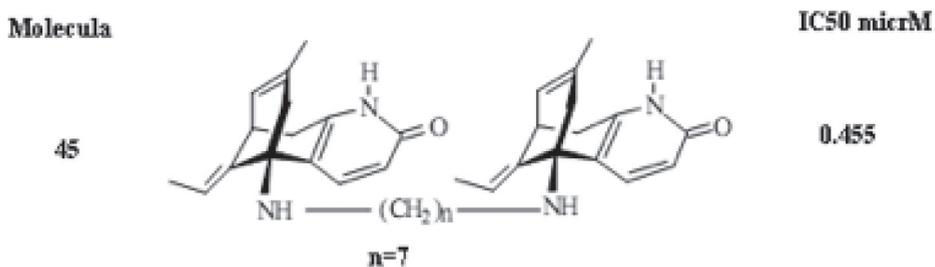


Figura 9  
Del trabajo de Shi se obtuvieron las siguientes moléculas

