



Departamento de Química Orgánica
Facultade de Farmacia

**Construcción QSAR de Redes
Complejas de Compuestos
de interés en Química Farmacéutica,
Microbiología y Parasitología**

Memoria presentada por:

ISELA GARCÍA PINTOS

para optar al grado de Doctor
por la Universidade de Santiago de Compostela

ISBN 978-84-9887-640-6 (Edición digital PDF)

**Construcción QSAR de Redes
Complejas de Compuestos
de interés en Química Farmacéutica,
Microbiología y Parasitología**

D. Xerardo García Mera, Prof. Titular y **D. Francisco Javier Prado Prado**, PDI Doctor Contratado por el Programa Ángeles Albariño, ambos del Departamento de Química Orgánica de la Universidad de Santiago de Compostela (USC). Así como, **D. Humberto González Díaz** PDI Doctor Contratado por el Programa Isidro Parga Pondal en el Departamento de Microbiología y Parasitología, Área de Parasitología, Facultad de Farmacia, USC.

INFORMAN:

Que la memoria titulada: **“CONSTRUCCIÓN QSAR DE REDES COMPLEJAS DE COMPUESTOS DE INTERÉS EN QUÍMICA FARMACÉUTICA, MICROBIOLOGÍA Y PARASITOLOGÍA”**, que para optar al grado de Doctor por la Universidade de Santiago de Compostela presenta D^a **ISELA GARCÍA PINTOS**, ha sido realizada bajo nuestra dirección, en el Departamento de Química Orgánica de la Facultad de Farmacia de la Universidad de Santiago de Compostela.

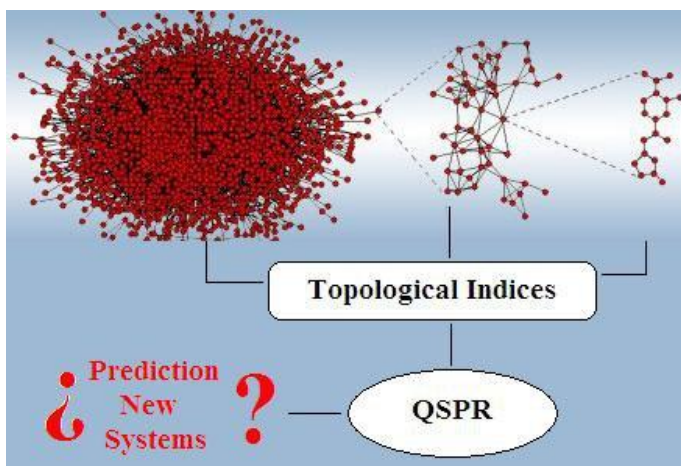
Y considerando que el trabajo constituye tema de Tesis Doctoral, autorizamos su presentación en la Universidade de Santiago de Compostela.

Y para que conste, expedimos el presente certificado en Santiago de Compostela a 20 de marzo de 2011.

Fdo. Dr. Francisco Prado Prado

Fdo. Prof. Dr. Xerardo García Mera

Fdo. Dr. Humberto González Díaz



"I think the next century will be the century of **Complexity**."
Stephen Hawking, January 23, 2000, SAN JOSE MERCURY NEWS

A mis padres y a mis hermanos

Este trabajo quisiera dedicárselo en primer lugar a mis padres y a mis hermanos, y expresarle mi agradecimiento a todas aquellas personas que lo han hecho posible, desde el primero al último:

a mis directores de tesis y compañeros de laboratorio
a la gente de otras universidades que han colaborado y me han ayudado
a mis amigos de la facultad y los de fuera de ella, por tantos buenos
momentos que hemos pasado juntos
a aquellas personas que de una manera especial me han ayudado a
realizar parte de este trabajo sin tener nada que ver con el
a quien ha estado ahí y me ha dado su apoyo para seguir adelante,
gracias por estar ahí,
a los que han confiado en mí desde el principio
a los que me han ayudado, a crecer cada día un poco más y ser quien soy

en general, mi agradecimiento va para todos aquellos que de una manera u otra forman parte de mi vida, saben como soy, y han estado siempre a mi lado, ya que dentro de todo este trabajo hay un poquito de cada uno de ellos, y porque sin ellos esto no sería posible.

ÍNDICE

ABREVIATURAS	V
1. INTRODUCCIÓN	1
1.1. Introducción al estudio del QSAR, esquema general de trabajo ..	3
1.2. Los descriptores moleculares y su interrelación	7
1.3. Cálculo de descriptores moleculares con Cadenas de Markov (CM)	13
1.4. Trabajos de Revisión Bibliográfica.....	20
1.4.1. Revisión de Estudios Teóricos de inhibidores de HGMR	21
1.4.2. Revisión de Estudios Teóricos de inhibidores de GSK-3, β - y γ -secretasas.....	23
1.4.3. Revisión de Estudios Teóricos de análogos de Vitamina D	24
1.5. Objetivos	25
2. RESULTADOS Y DISCUSIÓN	27
2.1. Estudio mt-QSAR y RC de inhibidores HGMR	30
2.2. Estudios QSAR de inhibidores de la GSK-3 α	32
2.3. Uso de Entropía en estudio mt-QSAR de inhibidores de la GSK-3.....	34
2.4. Uso de MARCH-INSIDE en estudio mt-QSAR de inhibidores de la GSK-3.....	36
2.5. Uso de ModesLab en estudio mt-QSAR de inhibidores de la GSK-3.....	38
2.6. Uso de QSAR y Docking en estudio de inhibidores de la GSK-3 β	39
2.7. Estudio mt-QSAR de interacción Diana-Proteína en antivirales	40
3. CONCLUSIONES	43

4. PUBLICACIONES (ANEXOS).....47

5. CURRICULUM

ABREVIATURAS

${}^{\text{SR}}\Pi_k$	Matriz estocástica
Θ_k	Distribución de probabilidades
ADL (LDA)	L inear D iscriminant A nalysis, término que proviene del inglés: Análisis Discriminante Lineal
ACP (PCA)	P rincipal C omponents A nalysis, término que proviene del inglés: Análisis de componentes principales
Actv	Actividad biológica
ANN	A rtificial N eural N etwork, término que proviene del inglés: Redes Neuronales Artificiales
ARN	Ácido ribonucleico
3D	Tridimensional
4D	Cuatro dimensiones
D	Descriptor Molecular
CM	C adenas de M arkov
CoMFA	C omparative M olecular F ield A nalysis
CoMSIA	C omparative M olecular S imilarity I ndices A nalysis
GSK-3	Enzima glicogen sintasa kinasa-3
HMGR	Enzima 3-hidroxi-3-metil-glutaril coenzima A reductasa
HMGRIs	Inhibidores de la enzima 3-hidroxi-3-metil-glutaril coenzima A reductasa

HTS	H igh- T hroughput- S creening, término que proviene del inglés: evaluación de alta eficacia
LNN	L inear N eural N etwork, término que proviene del inglés: Red Neuronal Lineal
m	Estado
[m] _k	Magnitudes medias
mt-	Multi-target, término que proviene del inglés: multi-diana
M	Matriz
MARCH-INSIDE	Markov Chain Invariants for Network Simulation and Design
p	Probabilidad
QSAR	Q uantitative- S tructure- A ctivity- R elationship, término que proviene del inglés: Relación-Cuantitativa-Estructura-Actividad
QSPR	Q uantitative- S tructure- P roperty- R elationship, término que proviene del inglés: Relación-Cuantitativa-Estructura-Propiedad
QSTR	Q uantitative- S tructure- T oxicity- R elationship, término que proviene del inglés: Relación-Cuantitativa-Estructura-Toxicidad
RC	R edes C omplejas
t	Tiempo
<i>T. cruzi</i>	<i>Trypanosoma cruzi</i>
TIs	T opological I ndex, término que proviene del inglés: Índices Topológicos

v	Vector
v^T	de v
$1^b, 2^a, 3^a \dots$	Son ejemplos del sistema usado para identificar los artículos de investigación en este trabajo. En el mismo todos los artículos científicos son identificados con un número que indica su orden de aparición en la Tesis y una letra con formato superíndice que indica el tipo de artículo. Los artículos de tipo (^a) son trabajos que usan descriptores moleculares basados en CM y los de tipo (^b) otros tipos de descriptores

INTRODUCCIÓN

1.1. Introducción al estudio del QSAR, esquema general de trabajo

Actualmente existen más de 15 millones de compuestos que han sido descubiertos o sintetizados en laboratorios químicos. Una gran cantidad de estos compuestos no ha encontrado aún aplicaciones farmacológicas, agroquímicas, industriales o de algún otro tipo. Esto es consecuencia directa de la diferencia existente entre la velocidad con que los nuevos compuestos son preparados y caracterizados, la cantidad de ellos que son sometidos a ensayos experimentales. La situación es más crítica si se tiene en cuenta que un gran número de los compuestos ensayados, de manera masiva por el método clásico de “prueba y error”, da resultados negativos. Este tipo de ensayos experimentales, especialmente los de corte farmacológico y toxicológico, son en general muy caros en términos de recursos materiales, humanos y de tiempo. También es de destacar el aspecto no sólo material, sino de tipo ético que conlleva la investigación con animales y su posterior sacrificio. En todo caso, nuevos paradigmas para el descubrimiento molecular han sido introducidos recientemente, basados en el uso de grandes librerías de compuestos químicos y sistemas robotizados para realizar ensayos biológicos. De tal modo los sistemas HTS (high-throughput screening), permiten la síntesis y ensayo de miles de compuestos cada día.¹

En este contexto, la industria farmacéutica ha reorientado las estrategias de búsqueda hacia métodos que permitan una selección o diseño racional de nuevos compuestos. En dicho sentido, los estudios QSAR (quantitative structure-activity-relationships) son usados como

¹ Kubinyi, H. *Rossiiskii Khimicheskii Zhurnal* **2006**, 50 (2), 5.

herramientas predicativas para el descubrimiento molecular. Estos métodos constituyen técnicas alternativas al ensayo masivo e indiscriminado de compuestos orgánicos, el cual debe ser realizado solamente a compuestos previamente seleccionados con modelos computacionales. Los métodos QSAR han emergido no sólo como una vía para el descubrimiento de compuestos con una actividad deseada sino además como un método con aplicaciones al estudio del mecanismo de acción de los mismos. El método QSAR se basa en la representación de la estructura molecular a través de ciertos números, denominados descriptores moleculares, los cuales son relacionados con la actividad biológica mediante técnicas de regresión.^{2,3,4}

Entre las técnicas de regresión usadas, son de destacar las técnicas lineares debido a su sencillez. En ellas se intenta modelar la actividad biológica como una función lineal multivariada de los descriptores moleculares. Por otra parte, en etapas tempranas del descubrimiento molecular así como del estudio del mecanismo de acción de los fármacos, es suficiente con tener una respuesta acerca de la probabilidad con que un fármaco tendrá la actividad o mecanismo bajo estudio, sin predecir el valor exacto. Particularmente, en este trabajo se utilizará el Análisis Discriminante Linear (ADL). El ADL posee la cualidad de ser simple permitiendo la clasificación de objetos en grupos predeterminados basándose en múltiples rasgos. En nuestro caso los objetos serán moléculas y los grupos, el grado de actividad o un mecanismo de acción

² Lutz, M.W.; Menius, J. A.; Laskody, R.G.; Domanico, P.L.; Goetz, A.G.; Saussy, D. L.; Rimele, T. *Network Science* **1996**, 2(9), September.

³ Loew, G.H.; Villar, H.O.; Alkorta, Y. *Pharm. Res.* **1993**, 10, 475.

⁴ Wess, G. *Drug Discovery Today*. **1996**, 1, 529.

determinado. La estrategia general de trabajo en QSAR con ADL se puede dividir en una serie de pasos que son ilustrados gráficamente en la **Figura 1**:^{5,6}

1. Recopilación de una serie de datos aleatoria, representativa y estratificada de moléculas con la actividad deseada y un grupo control que no posee la actividad bajo estudio.
2. Selección de los descriptores moleculares a utilizar.
3. Cálculo, mediante un programa computacional, de los descriptores moleculares seleccionados.
4. Utilización de los descriptores calculados a las moléculas recopiladas (serie de entrenamiento) para determinar modelos QSAR en un programa de cálculo estadístico.
5. Validación de los modelos QSAR contrastando la actividad predicha a las moléculas recopiladas (serie de predicción) con su actividad experimental.
6. Uso de los modelos encontrados para predecir la actividad a moléculas no ensayadas con anterioridad.

⁵ Kier, L.B.; Hall, L.H. *Topological indices and related descriptors in QSAR and QSPR*. Gordon and Breach, Amsterdam, **1999**.

⁶ Timmermann, H.; Todeschini, R; Consonni, V.; Mannhold, R; Kubinyi, H. *Handbook of molecular descriptors*. Ed., Wiley-VCH: Weinheim. **2002**.

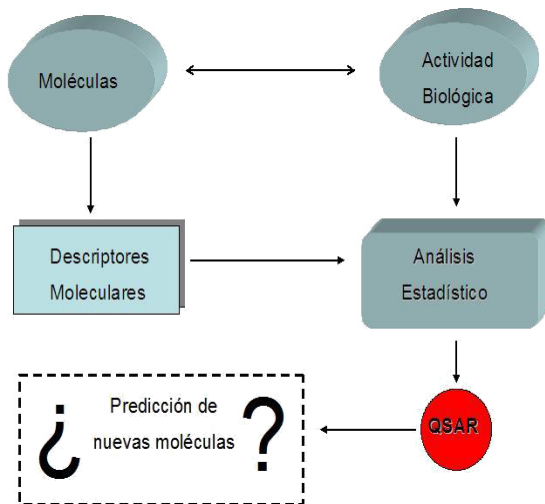


Figura1: Esquema general de trabajo con técnicas QSAR.

1.2. Los descriptores moleculares y su interrelación

El número de moléculas que puede ser obtenido por síntesis orgánica es tan vasto, que la probabilidad de seleccionar al azar una molécula que presente la actividad biológica deseada es prácticamente nula. Como consecuencia, ha surgido un gran interés en los métodos que relacionan la estructura molecular con la actividad biológica, especialmente el QSAR. Uno de los pilares para el desarrollo del QSAR, lo constituye el proceso de codificar la estructura química mediante descriptores moleculares.⁷ Se han definido en la actualidad más de 1600 descriptores moleculares diferentes, como podemos ver por ejemplo en el programa de cálculo DRAGON y el HANDBOOK de descriptores moleculares recopilados por sus autores, **Figura 2.**⁸

⁷ Devillers, J.; Balaban, A.T. Topological indices and related descriptors in QSAR and drug design, Amsterdam, **2000**.

⁸ Todeschini, R.; Consonni V. Handbook of Molecular Descriptors. Wiley VCH, Weinheim, Germany. **2000**.

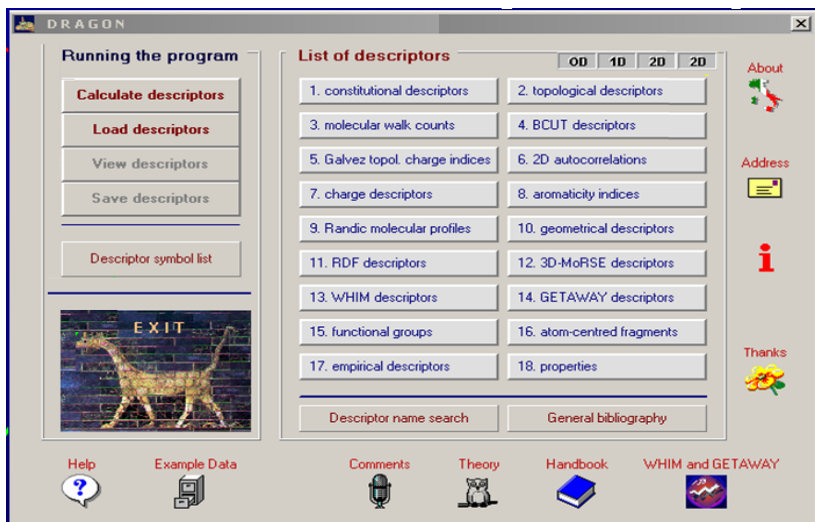


Figura 2. Portada principal del programa DRAGON que calcula más de 1600 descriptores moleculares agrupados en 18 familias diferentes.

En estos momentos no se detecta la “explosión” vista en el pasado, con respecto a la constante definición de nuevos índices estructurales o descriptores moleculares. Se observa en la actualidad cierta tendencia a una aplicación más diversificada e intensiva de los mismos.¹ No obstante, el inmenso número de propiedades a estudiar condiciona la continua introducción de nuevos índices basados en otros métodos, con la intención de que los químico-farmacéuticos posean un “arsenal” de descriptores moleculares lo más completo posible.²

Este contexto ha propiciado que algunos investigadores se hayan dedicado a la tarea de crear fórmulas matemáticas de los descriptores moleculares que ofrezcan un cuadro unificado de los mismos, para

facilitar su sistematización y estudio.⁹ Estas fórmulas pueden, además, indicar la dirección de búsqueda para nuevos descriptores moleculares. La gran mayoría de dichas ecuaciones usan representaciones matriciales de la estructura molecular, como es el caso de la matriz de adyacencia entre átomos, ilustradas en la **Figura 3**.¹⁰

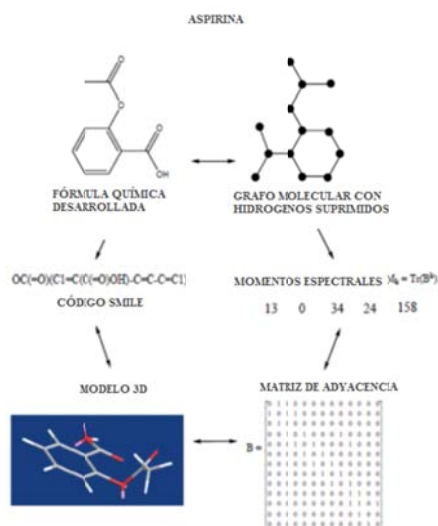


Figura 3. Distintas representaciones de la estructura molecular.

Muchos de los descriptores moleculares más conocidos en QSAR, sobre todo los índices topológicos, son susceptibles de una representación vector-Matriz-vector ($\mathbf{v} \cdot \mathbf{M} \cdot \mathbf{v}^T$). Por ejemplo, el primer descriptor molecular grafo-teórico, definido en un contexto químico, el índice de Wiener (W), es una forma cuadrática;¹¹ o sea, W es una forma $\mathbf{v} \cdot \mathbf{M} \cdot \mathbf{v}^T$ donde \mathbf{M} es una matriz simétrica y \mathbf{v}^T es el vector transpuesto de \mathbf{v} . Otros descriptores moleculares clásicos como los índices de Zagreb

⁹ Estrada, E. *Chem. Phys. Lett.* **2001**, 336, 248.

¹⁰ Kier, L.B.; Hall, L.H. *Topological Indices and Related Descriptors in QSAR and QSPR*. Gordon and Breach Sci. Pub.: Amsterdam, **1999**.

¹¹ Wiener, H. *J. Am. Chem. Soc.* **1947**, 69, 17.

M_1 y M_2 , el número de Harary (H), la invariante de Randic (χ), el índice de conectividad de valencia (χ^v), el índice de Balaban (J), el índice de topología molecular (MTI) y las auto-correlaciones de Moreau-Boroto (ATS_d), por citar sólo algunos, pueden ser expresados como transformaciones $\mathbf{v} \cdot \mathbf{M} \cdot \mathbf{v}^T$.⁸ También tiene cabida en este grupo los más recientes índices cuadráticos $q_k(X)$, lineares $f_k(X)$ y estocásticos $s_k(X)$, introducidos por Marrero-Ponce *et al.*^{12,13}

$$\begin{array}{lll}
 W = \frac{1}{2}(\mathbf{u} \cdot \mathbf{D} \cdot \mathbf{u}^T) & M_1 = \mathbf{v} \cdot \mathbf{A} \cdot \mathbf{u}^T & M_2 = \frac{1}{2}(\mathbf{v} \cdot \mathbf{A} \cdot \mathbf{v}^T) \\
 H = \frac{1}{2}(\mathbf{u} \cdot \mathbf{D}^k \cdot \mathbf{u}^T) & \chi = \mathbf{v}^l \cdot \mathbf{A} \cdot \mathbf{v}^{lT} & \chi^v = \mathbf{v}^{l1} \cdot \mathbf{A} \cdot \mathbf{v}^{l1T} \\
 J = \frac{1}{2} \cdot C \cdot (\mathbf{d}^l \cdot \mathbf{A} \cdot \mathbf{d}^{lT}) & MTI = \mathbf{v} \cdot (\mathbf{A} + \mathbf{D}) \cdot \mathbf{u}^T & ATS = \mathbf{w} \cdot \mathbf{m} \cdot \mathbf{B} \cdot \mathbf{w}^T \\
 q_k(X) = \mathbf{w} \cdot \mathbf{M} \cdot \mathbf{w}^T & f_k(X) = \mathbf{w} \cdot \mathbf{M} \cdot \mathbf{u}^T & s_k(X) = \mathbf{w} \cdot \mathbf{S}_k \cdot \mathbf{w}^T
 \end{array}$$

Como se ha podido ver, todos los símbolos de las matrices y vectores anteriores son de uso común en QSAR. Extensamente explicados en la literatura especializada, no lo serán aquí en detalle, queriéndose hacer hincapié, únicamente, en la idea del carácter unificador de las transformaciones $\mathbf{v} \cdot \mathbf{M} \cdot \mathbf{v}^T$.^{8,14}

Por otra parte, muchos estudios de química computacional hacen uso del concepto de momento espectral. Entre los índices basados en momentos espectrales los más conocidos son los momentos de energía $\mu(\mathbf{H})$, los conteos de caminos de auto-retorno $srwc^k$, los momentos espectrales de matrices de adyacencia entre enlaces $\mu(\mathbf{B})$ y $\mu(\mathbf{d}^l \mathbf{B})$, el índice I_3 de Estrada, para el grado de plegamiento de proteínas y el número de Kirchhoff (Kf). Todos estos índices pueden

¹² Marrero-Ponce, Y. J. *Chem. Inf. Comp. Sci.* **2004**, *44*, 2010.

¹³ Marrero-Ponce, Y.; González-Díaz, H.; Romero-Zaldívar, V.; Torrens, F.; Castro, E. A. *Bioorg. Med. Chem.* **2004**, *12*, 5331.

¹⁴ Estrada, E; Uriarte, E. *Curr. Med. Chem.* **2001**, *8*, 1573.

ser escritos en notación matemática mediante el operador traza de las matrices (Tr), que indica la suma de los valores en la diagonal principal de la matriz. Estos descriptores moleculares han sido clasificados tradicionalmente como un grupo aparte de los $\mathbf{v} \cdot \mathbf{M} \cdot \mathbf{v}^T$:⁸⁻
15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38

$$\begin{aligned} \text{srwc}_k &= \text{Tr}(\mathbf{A}^k) & \mu_k(\mathbf{B}) &= \text{Tr}(\mathbf{B}^k) & \mu_k({}^d\mathbf{B}) &= \text{Tr}[({}^d\mathbf{B} + \mathbf{W})^k] \\ \mu_k(\mathbf{H}) &= \text{Tr}(\mathbf{H}^k) & \text{Kf} &= a \cdot \text{Tr}(\mathbf{L}) & I_3 &= \frac{1}{k!} \sum_k \mu_k = \frac{1}{k!} \sum_k \text{Tr} \left[\mathbf{A} \left(\varphi, \Psi, \omega \right) \right] \end{aligned}$$

A pesar de los esfuerzos realizados con vistas a la unificación de los descriptores moleculares en el esquema $\mathbf{v} \cdot \mathbf{M} \cdot \mathbf{v}^T$, no se han realizado avances en la incorporación de índices como los momentos espectrales, a este prometedor esquema. En el presente trabajo se dará una

-
- ¹⁵ Gutman, I.; Rosenfield, V.R. *Theor. Chim. Acta* **1996**, *93*, 191.
¹⁶ Estrada, E. *Bioinformatics* **2002**, *18*, 1.
¹⁷ Estrada, E. *Chem. Phys. Lett.* **2000**, *319*, 713.
¹⁸ González, M.P.; Morales, A.H.; Molina R. *Polymer* **2004**, *45*, 2773.
¹⁹ González, M.P.; Morales, A.H.; González-Díaz H. *Polymer* **2004**, *45*, 2073.
²⁰ Morales, A.H.; González, M.P.; Rieumont J. B. *Polymer* **2004**, *45*, 2045.
²¹ Burdett, J.K.; Lee, S. *J. Am. Chem. Soc.* **1985**, *107*, 3063.
²² Burdett, J.K.; Lee, S. *J. Am. Chem. Soc.* **1985**, *107*, 3050.
²³ Lee, S. *Acc. Chem. Res.* **1991**, *24*, 249.
²⁴ Gutman, I. *Theor. Chim. Acta* **1992**, *83*, 313.
²⁵ Markovic, S.; Gutman, I. *J. Mol. Struct. Theochem* **1991**, *81*, 81
²⁶ Jiang, Y.; Tang, A.; Hoffmann, R. *Theor. Chim. Acta* **1984**, *66*, 183.
²⁷ Karwowski, J.; Bielinska-Waz, D.; Jurkowski, J. *Int. J. Quantum Chem.* **1996**, *60*, 185.
²⁸ Estrada, E.; González-Díaz H. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 75.
²⁹ González, M.P.; Terán, C. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3077.
³⁰ González, M.P.; Terán, C. *Bioorg. Med. Chem.* **2004**, *12*, 2985.
³¹ González, M.P.; Terán, C. *Bull. Math. Biol.* **2004**, *66*, 907.
³² González, M.P.; González-Díaz H.; Cabrera-Pérez, M.A.; Molina R. *Bioorg. Med. Chem.* **2004**, *12*, 735.
³³ González, M.P.; Morales, A.H. *J. Comput. Aid. Mol. Des.* **2003**, *10*, 665.
³⁴ González, M.P.; González-Díaz, H.; Molina, R.; Cabrera-Pérez, M.A.; Ramos de Armas, R. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1192.
³⁵ Cabrera-Pérez, M.A.; Bermejo, M. *Bioorg. Med. Chem.* **2004**, *22*, 5833.
³⁶ Cabrera-Pérez, M.A.; García, A.R.; Teruel C.F.; Álvarez, I.G.; Sanz, M.B. *Eur. J. Pharm. Biopharm.* **2003**, *56*, 197.
³⁷ Cabrera-Pérez, M.A.; González-Díaz, H.; Fernandez, T.C.; Pla-Delfina, J.M.; Bermejo, S.M. *Eur. J. Pharm. Biopharm.* **2002**, *53*, 317.
³⁸ Molina, E.; González-Díaz, H.; González, M.P.; Rodríguez, E.; Uriarte, E. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 515.

introducción

representación unificada de ambos grupos y esto permitirá simplificar el estudio sistematizado de los descriptores moleculares.

1.3. Cálculo de descriptores moleculares con Cadenas de Markov (CM)

Cadenas de Markov es el nombre de una teoría o tipo de modelo matemático definido por Markov.^{39,40} En nuestro trabajo se utilizará especialmente el método **MARCH-INSIDE** (del idioma Inglés: **Markov Chain Invariants for Network Simulation and Design**)^{41,42,43,44}; el cual emplea las CM para calcular descriptores moleculares mediante una aproximación sencilla a fenómenos tales como:

- a) distribución de electrones de valencia alrededor de los átomos de una molécula.
- b) propagación de una vibración en una cadena de ARN.
- c) propagación de interacciones electrostáticas superficiales en una proteína viral o en la estructura plegada 3D de una enzima.
- d) paso átomo por átomo de un fármaco desde el plasma a un tejido.
- e) interacción paso por paso de un fármaco con su receptor.

Aunque cabe destacar que las CM constituyen unos de los modelos de la teoría de probabilidades más usados, a continuación daremos algunas notas básicas sobre estas:

- Las CM estudian fenómenos estocásticos esto es: fenómenos que al ser estudiados mediante mediciones en el tiempo (t) el resultado obtenido no está determinado sino que puede obtenerse cada vez

³⁹ Markov, A.A. *Bull. Soc. Phys. Math. Kasan* **1906**, *15*, 155.

⁴⁰ Bharucha-Reid, A.T. Elements of Theory of Markov Process on the application, McGraw-Hill Series in Probability and Statistics, McGraw-Hill Book Company, New York. **1960**, 167.

⁴¹ González-Díaz, H.; Prado-Prado, F.; Ubeira, F.M. *Curr Top Med Chem.* **2008**, *8(18)*, 1676.

⁴² González-Díaz, H.; Duardo-Sanchez, A.; Ubeira, F.M.; Prado-Prado, F.; Pérez-Montoto, L.G.; Concu, R.; Podda, G.; Shen, B. *Curr Drug Metab.* **2010**, *11(4)*, 379.

⁴³ González-Díaz, H.; González-Díaz, Y.; Santana, L.; Ubeira, F.M.; Uriarte, E. *Proteomics.* **2008**, *8(4)*, 750.

⁴⁴ González-Díaz, H.; Vilar, S.; Santana, L.; Uriarte, E. *Curr Top Med Chem.* **2007**, *7(10)*, 1015.

con cierta probabilidad uno de varios resultados posibles. Es de destacar que todos los fenómenos antes mencionados (a-e) son por naturaleza estocásticos.

- En CM para referirse a un resultado determinado se habla de que el sistema ha ocupado un estado dado que cierto parámetro toma un valor determinado. Por ejemplo en la **Figura 4** se observa como el sistema pasa de un estado inicial al $t_0 = 0$ (en gris) a ocupar otros estados en tiempos sucesivos. En este trabajo hablaremos, además, de cómo cada estado está caracterizado por una magnitud fisicoquímica m_j , que depende del fenómeno en estudio. Con respecto a los ejemplos ofrecidos anteriormente en nuestro trabajo identificamos los siguientes pares sistema/estados en la **Tabla 1**.

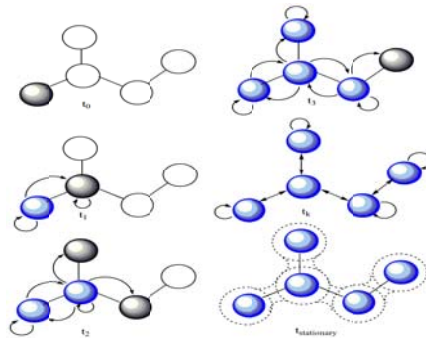


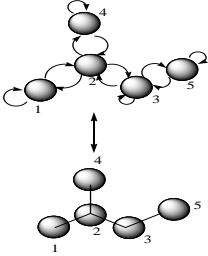
Figura 4. Ilustración gráfica de una cadena de Markov.

Tabla 1. Elementos o componentes de una CM para casos concretos estudiados en el presente trabajo.

Sistema	Estados (m_j)	Valores del Parámetro	Resultados esperados de la CM
Capa de electrones de valencia de la molécula.	Átomos (Electronegatividad) ó (electrones compartidos)	Presencia o ausencia de los electrones alrededor de cierto átomo j	Probs. absolutas con que los electrones se distribuyen alrededor de cada átomo
Vibración en una cadena de ARN	Nucleótidos (frecuencia de vibración)	Propagación permitida o prohibida de la vibración hasta un nucleótido j	Probs. absolutas con que un nucleótido participa en la vibración
Interacciones superficiales en una proteína viral	Aminoácidos Superficiales (Carga electrostática)	Participación o no de un aminoácido en Inter. Electrostaticas superficiales	Probs. absolutas de Inter. Electrostaticas superficiales
Propagación 3D de interacciones electrostáticas en proteínas	Aminoácidos (Carga electrostática)	Participación o no de un aminoácido en Inter. Electrostaticas 3D	Probs. absolutas de interacción electrostáticas 3D
Átomos de una molécula	Átomo (energía libre estándar)	Paso del átomo j a plasma, o tejido	Probs. absolutas de partición plasma/tejido

- En la CM se parte de una distribución inicial de probabilidades ${}^A p_0(j)$ que son las probabilidades absolutas iniciales ($t_0 = 0$) con que el sistema ocupa cada estado j .
- Dado un sistema con n estados estas probabilidades puede ordenarse en un vector de probabilidades iniciales ${}^0\pi = [{}^A p_0(1), {}^A p_0(2), {}^A p_0(3), \dots, {}^A p_0(n)]$.
- Las CM pueden ser representadas por un grafo dirigido, donde los vértices son los estados del sistema y los arcos representan la transición o paso del sistema de un estado a otro.
- Lo anterior se complementa con la representación matricial de las CM, convirtiéndolas en una herramienta muy versátil. De tal modo que las relaciones entre los estados del sistema expresadas por el grafo sobre el que se define la CM pueden ser resumidas a través de la matriz estocástica ${}^1\Pi$.⁴⁵
- Los elementos de esta matriz son las probabilidades ${}^1p_{ij}$ de que el sistema pase de un estado i en el tiempo $t_0 = 0$ a otro j en el tiempo $t_1 = 1$ o primer paso, ver **Tabla 2**:

Tabla 2. Representaciones grafo-teórica y matricial de una CM.

Grafo	Matriz ${}^1\Pi$
	${}^1\Pi = \begin{pmatrix} {}^1p_{11} & {}^1p_{12} & 0 & 0 & 0 \\ {}^1p_{21} & {}^1p_{22} & {}^1p_{23} & 0 & 0 \\ 0 & {}^1p_{32} & {}^1p_{33} & 0 & {}^1p_{35} \\ 0 & {}^1p_{42} & 0 & {}^1p_{44} & 0 \\ 0 & 0 & 0 & {}^1p_{53} & {}^1p_{55} \end{pmatrix}$

⁴⁵ Freund, J.A.; Poschel, T. Eds. Stochastic Processes in Physics, Chemistry, and Biology. In: Lect. Notes Phys. Springer-Verlag, Berlin, Germany 2000.

- Nótese que las transiciones del sistema de un estado i a otro j , que no esté directamente relacionado con él, están prohibidas en el t_i (sistemas con vértices no adyacentes en el grafo).
- Las CM cumplen la condición de que las probabilidades de transición ${}^k p_{ij}$ tanto para el primer paso (${}^1 p_{ij}$) como para tiempos mayores $t_k = k > 1$ depende solamente del estado que el sistema ocupaba en el tiempo inmediatamente anterior $t_{k-1} = k - 1$ pero no de tiempos anteriores.
- Las CM cumplen con las ecuaciones de Chapman-Kolmogorov, por lo que las probabilidades absolutas ${}^A p_k(j)$, con que el sistema ocupa determinado estado j en el tiempo k , se determinan como los elementos de los vectores ${}^k \pi = {}^0 \pi \cdot {}^k \Pi = {}^0 \pi \cdot ({}^1 \Pi)^k$. Así, las probabilidades absolutas de evolución del sistema a los tiempos $t_k = 0, 1, 2, 3, \dots, n$ quedan determinadas por:

$$\begin{aligned}
 {}^0 \pi &= {}^0 \pi \cdot {}^0 \Pi = {}^0 \pi \cdot ({}^1 \Pi) = {}^0 \pi \cdot I_n = [{}^A p_0(1) {}^A p_0(2) {}^A p_0(3) \dots {}^A p_0(n)] \\
 {}^1 \pi &= {}^0 \pi \cdot {}^1 \Pi = {}^1 \pi \cdot ({}^1 \Pi) = [{}^A p_0(1) {}^A p_0(2) {}^A p_0(3) \dots {}^A p_0(n)] \Pi = [{}^A p_1(1) {}^A p_1(2) {}^A p_1(3) \dots {}^A p_1(n)] \\
 {}^2 \pi &= {}^0 \pi \cdot {}^2 \Pi = {}^0 \pi \cdot ({}^1 \Pi)^2 = [{}^A p_0(1) {}^A p_0(2) {}^A p_0(3) \dots {}^A p_0(n)] \Pi \cdot \Pi = [{}^A p_2(1) {}^A p_2(2) {}^A p_2(3) \dots {}^A p_2(n)] \\
 {}^3 \pi &= {}^0 \pi \cdot {}^3 \Pi = {}^0 \pi \cdot ({}^1 \Pi)^3 = [{}^A p_0(1) {}^A p_0(2) {}^A p_0(3) \dots {}^A p_0(n)] \Pi \cdot \Pi \cdot \Pi = [{}^A p_3(1) {}^A p_3(2) {}^A p_3(3) \dots {}^A p_3(n)] \\
 &\vdots \\
 &\vdots \\
 &\vdots \\
 {}^k \pi &= {}^0 \pi \cdot {}^k \Pi = {}^0 \pi \cdot ({}^1 \Pi)^k = [{}^A p_0(1) {}^A p_0(2) {}^A p_0(3) \dots {}^A p_0(n)] ({}^1 \Pi)^k = [{}^A p_k(1) {}^A p_k(2) {}^A p_k(3) \dots {}^A p_k(n)]
 \end{aligned}$$

- Del mismo modo, las probabilidades de transición ${}^k p_{ij}$ pueden ser calculadas como los elementos de las matrices ${}^1\Pi = ({}^1\Pi)^k$.
- Las probabilidades ${}^k p_{ii}$ para $i = j$ se denominan probabilidades de auto-retorno, el sistema regresa al estado inicial, estas probabilidades se encuentran en la diagonal principal de las matrices estocásticas.
- Tanto las probabilidades absolutas iniciales ${}^A p_0(j)$ (elementos de ${}^0\pi$), como las probabilidades ${}^1 p_{ij}$ (elementos de ${}^1\Pi$) pueden ser determinadas a partir de las magnitudes fisicoquímicas m_j que caracterizan al estado:

$${}^A p_0(j) = \frac{m_j}{\sum_l m_l} \qquad {}^1 p_{ij} = \frac{\alpha_{ij} \cdot m_j}{\sum_l \alpha_{il} \cdot m_l}$$

- Donde, α_{ij} indica la adyacencia entre los dos estados (átomos, aminoácidos, nucleótidos).
- De lo anterior se desprende que es posible derivar de la CM ciertos números que la caracterizan, ya que dependen de: i) los estados presentes, ii) de su interconexión caracterizada por α_{ij} , y iii) de la tendencia del sistema a ocupar dichos estados caracterizados por su magnitud fisicoquímica m_j .
- Por tanto de ser aplicada la CM a un sistema molecular como los descritos anteriormente dichos números podrán ser usados como descriptores moleculares (D_h) para encontrar modelos QSAR de una actividad biológica ($Actv$) dada preferentemente de tipo lineal:

$$Actv = a_1 D_1 + a_2 D_2 + a_3 D_3 + a_4 D_4 \dots + a_h D_h + b$$

- Un ejemplo, en el sistema a) en el cual la CM representa a una molécula donde la posición de los electrones (parámetro) puede

estar alrededor de varios átomos (estados) el aspecto i) está ligado al tipo y cantidad de átomos en la molécula, el aspecto ii) a la presencia de enlaces específicos entre los átomos y el aspecto iii) a la electronegatividad con que cada átomo atrae los electrones.

- Entre los números introducidos en este trabajo con ese fin se encuentran:

- las probabilidades absolutas en sí (que pueden ser usadas como descriptores moleculares locales solamente).
- los momentos espectrales de la matriz estocástica (${}^{SR}\pi_k$), suma de las probabilidades de autoretorno:

$${}^{SR}\pi_k = Tr({}^k\Pi) = Tr(\Pi) = \sum_j^k p_{jj}$$

- las entropías de la distribución de probabilidades (Θ_k):

$$\Theta_k = -\sum_j^A p_k(j) \log^A p_k(j)$$

- y magnitudes medias ($[m]_k$) calculadas como sumas ponderadas de probabilidades absolutas:

$$[m]_k = -\sum_j^A p_k(j) \cdot m_j$$

1.4. Trabajos de Revisión Bibliográfica

En esta sección incluimos 3 trabajos de revisión bibliográfica sobre los antecedentes publicados en la literatura relacionados con los temas de esta tesis. En todos los trabajos revisamos los estudios QSAR con parámetros conceptuales que utilizan análisis de regresión ó ADL, incluyendo estudios teóricos para comprender los requisitos estructurales esenciales para la unión con el receptor. En estos trabajos de revisión también discutimos modelos 3D y 4D QSAR, CoMFA ó CoMSIA con diferentes compuestos. Los compuestos estudiados se centran en 3 tipos de dianas diferentes con acciones de importancia en química farmacéutica, microbiología y parasitología.

1.4.1. Revisión de Estudios Teóricos de inhibidores de HGMR

Medicamentos eficaces como las estatinas o los ácidos mevínicos son inhibidores de la enzima limitante de la biosíntesis del colesterol, conocida como 3-hidroxi-3-metil-glutaril coenzima A reductasa (HMGR, sigla en Inglés). Esta enzima es responsable de la doble reducción de la 3-hidroxi-3-metilglutaril coenzima A. A partir de estos compuestos se puede realizar la síntesis y evaluación de nuevos inhibidores de la HMGR (HMGRIs).

Otro uso potencial de este tipo de drogas puede ser el control de infecciones parasitarias. Se ha estudiado en el protozoo parásito *Trypanosoma cruzi* el efecto anti-proliferativo de la mevinolina (lovastatina), un fármaco de la familia de los HMGRIs, y su capacidad para potenciar la acción de inhibidores específicos de la biosíntesis de ergosterol, como el ketoconazol y terbinafina (tanto *in vitro* como *in vivo*). Estos resultados confirman la acción sinérgica contra las fases proliferativa de *T. cruzi* (tanto *in vitro* como *in vivo*) y la inhibición de la biosíntesis de ergosterol (*in vivo*). Además, la práctica médica sugiere que la mevinolina, combinada con azoles como el ketoconazol, puede ser utilizada en el tratamiento de la enfermedad de Chagas humana.

El alto número de posibles compuestos candidatos a ensayar crea la necesidad de desarrollar y aplicar modelos QSAR con el fin de orientar la síntesis de HMGRIs. En este trabajo, se revisan los diferentes estudios computacionales para distintas series de HMGRIs. En primer lugar, revisamos los estudios QSAR con parámetros conceptuales que utilizan el análisis de regresión, y los estudios QSAR para comprender los

requisitos estructurales esenciales para la unión con el receptor. En esta revisión también discutimos modelos 3D QSAR, CoMFA y CoMSIA con diferentes compuestos HGMRI.

1.4.2. Revisión de Estudios Teóricos de inhibidores de GSK-3, β -y γ -secretasas

Los inhibidores de las enzimas glucógeno sintasa quinasa 3 (GSK-3), así como la β y γ -secretasas son candidatos interesantes para desarrollar compuestos anti-Alzheimer. A su vez, los compuestos inhibidores de GSK-3 también son interesantes como antiparasitarios activos contra el *Plasmodium falciparum*, *Trypanosoma brucei* y *Leishmania donovani*; que son los agentes causantes del paludismo, la tripanosomiasis africana humana y la leishmaniosis. El alto número de posibles candidatos crea la necesidad de desarrollar y aplicar modelos QSAR con el fin de orientar la síntesis de estos inhibidores. En los dos trabajos presentados en esta sección, hemos revisado diferentes estudios computacionales para estos inhibidores. En primer lugar, revisamos los estudios QSAR con parámetros conceptuales. En estas revisiones también discutimos modelos 3D y/o 4D QSAR, CoMFA y CoMSIA.

1.4.3. Revisión de Estudios Teóricos de análogos de Vitamina D

1R,25-dihidroxitamina D₃, la forma hormonalmente activa de la vitamina D₃, además de regular la homeostasis del calcio y la mineralización ósea clásica, también promueve la diferenciación celular e induce algunas funciones biológicas relacionadas con el sistema inmunológico. Extensos estudios estructura-función han demostrado que es posible modificar la estructura del calcitriol para obtener análogos de la vitamina D₃. Estos compuestos son capaces de inducir, de una manera selectiva, las funciones biológicas relacionadas con la misma hormona. En este trabajo, hemos revisado diferentes estudios computacionales para inhibidores de Vitamina D. En primer lugar, revisamos los estudios QSAR con parámetros conceptuales. En esta revisión también discutimos modelos basados en funciones de distribución, 4D QSAR, CoMFA y CoMSIA con diferentes compuestos análogos de Vitamina D₃.

1.5. Objetivos

Objetivos Generales:

1. Desarrollar nuevos modelos QSAR aplicables a la predicción de la actividad biológica de compuestos contra una única diana o múltiples dianas, modelos QSAR multi-target (mt-QSAR), de interés en química farmacéutica, microbiología, y parasitología.
2. Desarrollar nuevas metodologías de construcción de RC de estos compuestos útiles en estudios de Bioinformática a partir de modelos QSAR o mt-QSAR.

Objetivos específicos:

- 1.1. Desarrollar modelos QSAR para la predicción de HGMRIs.
- 1.2. Desarrollar modelos mt-QSAR para la predicción de inhibidores de la GSK-3.
- 1.3. Desarrollar modelos mt-QSAR para la predicción de compuestos antivirales.
- 1.4. Realizar una revisión que permita valorar las perspectivas futuras de desarrollo de modelos QSAR, mt-QSAR, en análogos de Vitamina D.
- 2.1. Desarrollar una metodología de construcción de RC de compuestos HGMRIs a partir del modelo QSAR.
- 2.2. Desarrollar una metodología de construcción de RC de compuestos antivirales a partir del modelo mt-QSAR.

RESULTADOS Y DISCUSIÓN

En este acápite se presentarán todos los resultados obtenidos en forma de artículos de investigación ya publicados por el autor. Los 7 artículos presentados (6 artículos de revista y 1 capítulo de libro) están agrupados de acuerdo al objetivo específico que complimentan. Para cada artículo se presenta una breve sección explicativa en español de su importancia y los resultados alcanzados. En el apartado “4. Publicaciones” de esta Tesis se adjuntan las publicaciones correspondientes en el idioma en que fueron publicadas.

2.1. Estudio mt-QSAR y RC de inhibidores HGMR

Medicamentos eficaces como las estatinas o los ácidos mevínicos son HGMRI, inhibidores de la enzima HMGR; que limita la velocidad de biosíntesis del colesterol. Sin embargo, el elevado número de posibles compuestos candidatos a ensayar crea la necesidad de desarrollar modelos QSAR para guiar la síntesis de los HMGRI. Los modelos QSAR desarrollados con anterioridad en este sentido (ver trabajo de revisión bibliográfica presentado en la Introducción) presentan dos problemas principales: son aplicables únicamente a series homogéneas de compuestos y no tienen en cuenta la quiralidad de los compuestos.

En este trabajo, se propone por primera vez un modelo QSAR para una serie grande y heterogénea de HMGRI. El modelo se basa en TIs de las estructuras moleculares. Además proponemos la primera red compleja que describe las relaciones de similitud entre HGMRI usando como entrada las predicciones de este modelo. Ambos, el modelo QSAR y la red, fueron usados para predecir las diferencias en actividad debido a la quiralidad en más de 1600 isómeros quirales de HMGRI no explorados experimentalmente. También se presentó una versión reducida de esta red (Componente Gigante) que contiene el conjunto más representativo de los compuestos quirales candidatos a ser ensayados como HMGRI. El trabajo sugiere una nueva aplicación combinado el estudio QSAR y las RC.

Como conclusión podemos destacar que en este trabajo se propuso por primera vez un modelo QSAR basado en índices topológicos quirales y no quirales de una lista heterogénea de inhibidores de la HMGR. Este modelo se usó para la predicción de la actividad de la HMGR de nuevos inhibidores quirales. Además, la comparación entre isómeros quirales fue realizada mediante redes complejas quirales y no quirales. Estas redes se construyeron mediante previas predicciones QSAR. Una ventaja de este método es la posibilidad que ofrece para buscar nuevos compuestos quirales que no fueron caracterizados experimentalmente. Otra ventaja es que el empleo de este modelo QSAR es una guía importante para los experimentos sintéticos, para la búsqueda de nuevos candidatos a inhibidores de la HMGR y al mismo tiempo, la disminución de costes que eso supone.

2.2. Estudios QSAR de inhibidores de la GSK-3 α

En general, los inhibidores de distintas isoformas de la GSK-3 son candidatos interesantes para el desarrollo de compuestos anti-Alzheimer. Los inhibidores GSK-3 también son de interés como compuestos antiparasitarios activos contra *Plasmodium falciparum*, *Trypanosoma brucei* y *Leishmania donovani*, los agentes causantes de la malaria, tripanosomiasis africana humana y la leishmaniosis. Esto ha provocado una búsqueda activa de potentes y selectivos inhibidores de GSK-3. En este sentido, los estudios QSAR podrían desempeñar un papel importante en el descubrimiento de estos inhibidores de GSK-3. Por esta razón, en este trabajo hemos desarrollado modelos QSAR para los inhibidores de la GSK-3 α . En el estudio hemos usado ADL y ANN de casi 50.000 casos con más de 700 diferentes inhibidores de GSK-3 α . Los compuestos fueron obtenidos desde la base de datos ChEMBL, en total se utilizaron más de 20.000 moléculas diferentes para desarrollar los modelos QSAR. El modelo clasificó correctamente 237 de 275 compuestos activos (86,2%) y 14.870 de 15.970 compuestos inactivos (93,2%) en la serie de entrenamiento. El porcentaje general de buena clasificación fue de 93,0%. La validación del modelo se llevó a cabo mediante una serie de predicción externa. En esta serie el modelo clasifica correctamente 458 de 549 (83,4%) y 29.637 de 31.927 casos control (83,4%). El porcentaje general de buena clasificación fue del 92,7%. En este trabajo, proponemos tres tipos de ANN no lineales y se muestra otro modelo alternativo a los ya existentes en la literatura,

como ADL. El mejor modelo obtenido fue una ANN lineal (LNN): LNN: 236:236-1-1:1 que tuvo porcentaje de buena clasificación del 96%. Además, hicimos un estudio de los diferentes fragmentos que existen en las moléculas de la base de datos con el fin de ver cuales tenían más influencia en la actividad. Todo esto puede ayudar a diseñar nuevos inhibidores de GSK-3 α .

Como modelos no lineales calculamos las ANN utilizando los descriptores calculados con el DRAGON viendo que estos modelos eran otros métodos alternativos para estudiar la actividad de distintas familias de moléculas. Otra parte de este trabajo fue el estudio de las contribuciones de los fragmentos usando un modelo QSAR, lo que nos puede ayudar al diseño de los mejores inhibidores de GSK-3 α , y poder así luego sintetizarlos en el laboratorio, eliminando la síntesis de moléculas al azar ya que existe la posibilidad de que la mayoría sean inactivos.

2.3. Uso de Entropía en estudio mt-QSAR de inhibidores de la GSK-3

El desarrollo de modelos QSAR usando índices moleculares simples parece ser una prometedora técnica alternativa o complementaria al Docking fármacos-proteínas. Casi todas las técnicas QSAR se basan en el uso de descriptores moleculares, que son series numéricas que codifican la información química útil y que permiten correlacionar las propiedades estructurales y biológicas. Entropía de Shannon es uno de los parámetros más importantes con el fin de codificar la información estructural sobre los estudios QSAR. En este sentido, nuestro grupo de investigación ha introducido una nueva serie de índices estocásticos que pueden ser calculados con la técnica MARCH-INSIDE. El método MARCH-INSIDE se basa en el uso de CM para calcular probabilidades absolutas de la distribución de las diferentes propiedades atómicas dentro de la estructura molecular. Podemos aplicar la fórmula de Shannon a estas probabilidades absolutas para calcular los parámetros de la entropía de la distribución de las propiedades atómicas en la molécula. En este trabajo vamos a explorar el potencial de la técnica MARCH-INSIDE para buscar un modelo mt-QSAR para una serie heterogénea de compuestos inhibidores de la GSK-3. En el primer paso, los descriptores moleculares antes mencionados fueron calculados para una gran serie de compuestos activos/inactivos. Posteriormente se utilizó ADL para ajustar la función de clasificación. El modelo mt-QSAR fue validado después

con una serie de predicción externa mediante la técnica de re-sustitución.

En conclusión, podemos considerar la técnica MARCH-INSIDE como una buena alternativa para el desarrollo de nuevos inhibidores de la GSK-3 ya que clasifica correctamente a muchos inhibidores con diferente estructura molecular.

2.4. Uso de MARCH-INSIDE en estudio mt-QSAR de inhibidores de la GSK-3

En el trabajo aquí descrito, hemos desarrollado el primer modelo mt-QSAR capaz de predecir los resultados de 42 pruebas experimentales diferentes de inhibidores de GSK-3 con patrones estructurales heterogéneos. La técnica MARCH-INSIDE se utilizó para calcular rápidamente los valores promedio totales y locales de polarizabilidad, los coeficientes de partición n-octanol/agua (P), la refracción, el área de van der Waals y la electronegatividad de 4508 compuestos activos/inactivos, así como los valores medios de estos índices para los compuestos activos en 42 ensayos biológicos diferentes. Tanto los descriptores moleculares individuales como los valores promedio de cada prueba se utilizaron como entrada para un ADL. Encontramos una función de clasificación que clasifica correctamente 873 de 1218 casos de inhibidores de GSK-3 (97,4%) y 2140 de 2.163 casos de compuestos no activos (86,1%) en 42 ensayos farmacológicos diferentes. Además, el modelo clasifica correctamente 285 de 406 inhibidores de GSK-3 (96,3%) y 710 de 721 casos de compuestos no activos (85,4%) en la serie de validación externa. El resultado es importante porque, por primera vez podemos utilizar una sola ecuación para predecir los resultados de una serie heterogénea de compuestos orgánicos en 42 ensayos experimentales diferentes en lugar de desarrollar, validar y usar 42 modelos QSAR diferentes (uno para cada ensayo).

Podemos concluir que la técnica MARCH-INSIDE es una buena alternativa para el descubrimiento de nuevos inhibidores de la GSK-3 ya que proporciona una buena clasificación de compuestos con diversa estructura molecular.

2.5. Uso de ModesLab en estudio mt-QSAR de inhibidores de la GSK-3

En este trabajo se usó el software ModesLab para calcular descriptores topológicos de una gran serie de 3,370 compuestos activos e inactivos como inhibidores de la GSK-3. Se utilizó ADL para ajustar la función de clasificación y predecir esta serie heterogénea de compuestos. El estudio proporciona así un modelo mt-QSAR basado en el ModesLab para la evaluación general de este tipo de moléculas.

Podemos concluir que la metodología ModesLab con descriptores moleculares topológicos y compuestos con elevada diversidad estructural, es otra alternativa para el descubrimiento de inhibidores de la enzima GSK-3 comparable con los métodos descritos en la bibliografía.

2.6. Uso de QSAR y Docking en estudio de inhibidores de la GSK-3 β

Los descriptores VolSurf+ y GRIND pueden extraer la información presente en los Campos de Interacción Molecular calculados con la técnica GRID. Los descriptores VolSurf son más fáciles de interpretar y son de aplicación general a los temas ADME-Tox, mientras que los segundos son más sofisticados y por tanto más apropiados para estudios de farmacodinámica. En este trabajo se presenta un estudio de comparación de modelos QSAR para inhibidores de GSK-3 β no ATP competitivos obtenidos con los descriptores GRIND y VolSurf+. Los resultados indican no sólo que los descriptores Volsurf+ más simples son lo suficientemente buenos para predecir e interpretar químicamente el fenómeno investigado, sino que también se predijo una conformación bioactiva de la palinurina que puede orientar el diseño futuro de estos inhibidores.

2.7. Estudio mt-QSAR de interacción Diana-Proteína en antivirales

A día de hoy los modelos QSAR para compuestos antivirales tienen una limitación importante; sólo predicen la actividad biológica de los fármacos contra una única especie viral. Esto se determina por el hecho de que la mayoría de los descriptores moleculares actuales sólo pueden codificar información sobre la estructura molecular. Como resultado, la predicción de la probabilidad de que un fármaco sea activo frente a diferentes especies virales sólo con un modelo unificador es un objetivo de gran importancia. En este trabajo se utiliza la técnica MARCH-INSIDE para calcular parámetros de tipo entropía multi-objetivo para encontrar por primera vez un modelo mt-QSAR que predice la actividad biológica de 500 fármacos probados en la literatura contra 40 especies virales. Se utilizó un ADL para clasificar los fármacos como activos o no activos frente a las diferentes especies virales estudiadas. El modelo clasifica correctamente a 1424 de 1445 compuestos inactivos (98,55%) y 281 de 333 compuestos activos (84,38%). La validación del modelo se llevó a cabo por medio de una serie de predicción externa, clasificando correctamente 698 de 704 compuestos inactivos y 143 de 157 compuestos activos.

Este modelo, obtenido a partir de una base de datos amplia, mejora significativamente el anterior modelo QSAR, y el empleo de la entropía como descriptor molecular nos permite predecir la

actividad antiviral de muchos compuestos contra diversidad de patógenos virales.

CONCLUSIONES

Exponemos las conclusiones específicas, en correspondencia con los objetivos trazados, agrupadas en tres grupos, dada la naturaleza de los estudios realizados: 1) estudios QSAR/mt-QSAR de inhibidores de enzimas, 2) estudios de RC, 3) estudio mt-QSAR y de RC de múltiples enzimas:

Conclusiones específicas:

- 1.1. Se pudo desarrollar modelos QSAR para la predicción HGMRIs.
- 1.2. Desarrollamos modelos mt-QSAR para la predicción de inhibidores de la GSK-3.
- 1.3. Pudimos desarrollar modelos mt-QSAR para la predicción de compuestos antivirales.
- 1.4. Según la revisión realizada no existen modelos mt-QSAR en análogos a Vitamina D. Por lo que podemos concluir que el desarrollo de modelos mt-QSAR y RC para estos compuestos es un campo con perspectivas futuras.
 - 2.1. Desarrollamos una metodología QSAR de construcción de RC de compuestos HGMRIs.
 - 2.2. Desarrollamos una metodología de construcción de RC de compuestos anti-virales a partir del modelo mt-QSAR.

Conclusión general:

Podemos concluir que los nuevos modelos QSAR desarrollados son aplicables a la predicción de la actividad biológica de compuestos contra una única diana o múltiples dianas de interés en química farmacéutica, microbiología y parasitología. Además, podemos concluir que es posible

conclusiones

desarrollar nuevas metodologías de construcción de RC de estos compuestos en estudios de Bioinformática a partir de modelos QSAR o mt-QSAR.

PUBLICACIONES

A continuación se presenta un ANEXO con las publicaciones que se recogen en la Tesis siguiendo el orden establecido en la misma.

QSAR & Complex Network Study of the HMGR Inhibitors Structural Diversity

Isela García*, Yagamare Fall Diop and Generosa Gómez

Department of Organic Chemistry, University of Vigo, Spain

Abstract: Efficient drugs such as statins or mevinic acids are inhibitors of the rate-limiting enzyme of cholesterol biosynthesis, 3-hydroxy-3-methyl-glutaryl coenzyme A reductase (HMGR), an enzyme responsible for the double reduction of 3-hydroxy-3-methyl-glutaryl coenzyme A. These compounds promoted the synthesis and evaluation of new inhibitors for HMGR, named HMGRIs. The high number of possible candidates creates the necessity of Quantitative Structure-Activity Relationship models in order to guide the HMGR (3-hydroxy-3-methyl-glutaryl coenzyme A inhibitor) synthesis. In this work, we revised different computational studies for a very large and heterogeneous series of HMGRIs. First, we revised QSAR studies with conceptual parameters such as flexibility of rotation, probability of availability, etc; we then used the method of regression analysis; and QSAR studies in order to understand the essential structural requirement for binding with receptor. Next, we reviewed 3D QSAR, CoMFA and CoMSIA with different compounds to find out the structural requirements for 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) inhibitory activity

Keywords: QSAR; Complex network, Lipid-lowering agent, Cholesterol level, Atherosclerotic disease, Antiparasite drug, Trypanosoma cruzi, Chagas' disease, 3-hydroxy-3-methyl-glutaryl coenzyme A reductase, CoMSIA, CoMFA, topological indices.

1. INTRODUCTION

Hypercholesterolemia (the level of plasma cholesterol, and in particular LDL-associated cholesterol) is well-known as the main risks factor in atherosclerotic, the degenerative disease underlying myocardial infarction and stroke, and coronary heart diseases [1, 2].

In western countries, this is the leading cause of death, being more common than all cancers and leukemias combined. Briefly, atherosclerotic lesions develop as follows:

1. Small mechanical lesions in the vascular endothelium (the innermost layer of blood vessel walls) allow leakage of blood plasma into the muscular layers beneath. Formation of these small leakages is thought to be promoted by high blood pressure.
2. The lipoproteins that leaked into the tissue are degraded. Because of its low solubility, cholesterol released from degraded lipoproteins precipitates.
3. The cholesterol particles trigger invasion of phagocytic cells and in this way contribute to triggering inflammation, which in turn increases the tissue damage and turns the small, potentially repairable defects of the vessel wall into large lesions.

Clinical studies with lipid-lowering agents have established that the decrease of high serum cholesterol levels reduces the incidence of cardiovascular mortality.

Cholesterol occurs in animals but not in plants or fungi. These have similar sterols for similar purposes, which however cannot be converted into cholesterol. Therefore, it is essential for animals (particularly for plant-feeding ones, such as sheep, goat and vegetarians) to have a pathway for cholesterol biosynthesis. Statins and mevinic acids are two efficient drugs known as inhibitors of the rate-limiting enzyme of cholesterol biosynthesis, 3-hydroxy-3-methyl-glutaryl coenzyme A reductase (HMGR), enzyme responsible for the double reduction of 3-hydroxy-3-methyl-glutaryl coenzyme A into mevalonic acid. Synthesis starts with Acetyl-CoA in the mitochondrion, which is used to synthesize HMG-CoA. These reactions also occur in ketogenesis. However, while the entire process of ketogenesis occurs in the mitochondrion, the formation of HMG-CoA in sterol synthesis occurs in the cytosol. All subsequent steps occur in the smooth endoplasmic reticulum. HMG-CoA reductase reduces HMG-CoA to mevalonate, which in turn is con-

verted into various isoprene compounds. Several rounds of polymerization lead to the linear hydrocarbon molecule squalene, which is then converted into lanosterol. Subsequent modifications lead to cholesterol. The reactions of the synthetic pathway are shown in Fig. (1). The reaction catalyzed by HMG-CoA reductase is the first committed step, which means that from this point onwards the substrates have no other option than becoming a sterol. Therefore, HMG-CoA reductase is the main target of regulatory mechanisms, which in turn is being exploited in pharmacotherapy. The biosynthesis of cholesterol (the same as for many other metabolites) is regulated both by allosteric control of enzyme activity and by enzyme induction. The committed step is the reduction of HMG-CoA to mevalonate, and it therefore makes sense that HMG-CoA reductase is the primary target of regulation. The drug control of this enzyme is efficient in reducing the levels of cholesterol in plasma. [3, 4] The structure of statins and its derivatives is characterized by the desmethylmevalonic acid or by the lactone. The pharmacophore is connected to a lipophilic ring, such as hexahydronaphthalene, indole, pyrrole, pyrimidine or quinone, by a linking element (a two-carbon spacer). The biologically active form of mevinic acids is represented by the open chain hydroxyl-acid, which mimics the HMGR natural substrate [5].

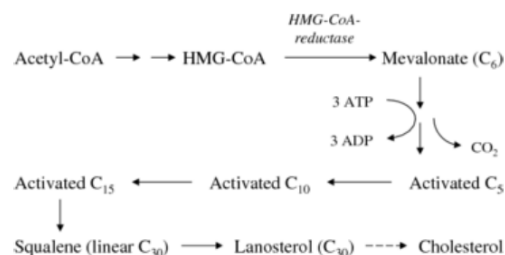


Fig. (1). Pathway.

The genetic regulation of cholesterol synthesis has been worked out only in the last decade, and it operates by quite a neat mechanism. Several membrane proteins participate in it (Fig. 2a): The Sterol Response Element Binding Protein (SREBP), the SREBP Cleavage Activating Protein (SCAP), and two SREBP-specific proteases (S1P and S2P). SCAP is the actual cholesterol sensor that changes conformation in response to changes of the cholesterol concentration within the ER membrane. If cholesterol is high,

*Address correspondence to this author at the Department of Organic Chemistry, University of Vigo, Spain; Tel: +34986813679; Fax: 986812262 E-mail: iselapintos@yahoo.es

SCAP does not bind SREBP, and SREBP is left behind when SCAP relocates to the Golgi (Fig. 2b). Nothing further happens in this case.

The action starts when the concentration of cholesterol in the ER membrane is low (Fig. 2c):

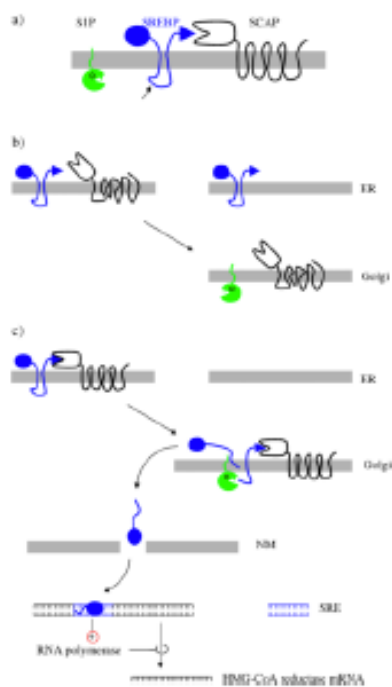


Fig. (2). Membrane based mechanism.

1. SCAP adopts a different conformation and now binds SREBP, taking it along for the ride to the Golgi apparatus.
2. In the Golgi, S1P and S2P ambush SREBP and cleave it.
3. The major fragment is free to leave the membrane and translocate to the nucleus.
4. In the nucleus, SREBP finds and binds to sterol response elements (SRE). SREs are short specific sequences that occur in various places of the genome.
5. One SRE is located upstream of the gene of HMG-CoA reductase. Binding of SREBP to this element increases transcription and therefore the overall activity of HMG-CoA reductase.

Another interesting potential use of this type of drugs could be the control of parasite infections [6-15]. Urbina *et al.* studied in the protozoan parasite *Trypanosoma (Schizotrypanum) cruzi* the anti-proliferative effects of mevinolin (lovastatin), a drug from the family of HMGR inhibitors (HMGRIs), and its ability to potentiate the action of specific ergosterol biosynthesis inhibitors, such as ketoconazole and terbinafine (both *in vitro* and *in vivo*). These results confirm the synergic action against the proliferative stages of *T. cruzi* (both *in vitro* and *in vivo*) and the ergosterol biosynthesis inhibition (*in vivo*) by acting in different points of the pathway. In addition, the medical practice suggests that mevinolin, combined with azoles, such as ketoconazole, can be used in the treatment of human Chagas' disease [16].

Alternatively, TI-QSAR models can be used to explore the relationships between the structural spaces of compounds as inhibitors for specific enzymes, such as MAO inhibitors [17], HIV-1 integrase inhibitors [18], and/or protease inhibitors [19] or tyrosinase inhibitors [20-22]. Unfortunately, the classic TIs do not consider important 3D features, such as chirality, and the biological activity of different stereo-isomers of the same molecule can be notably different [23-30].

The Complex Networks (CNs) [31] are similar to the graphs, composed by many nodes connected by edges [32]. In molecular sciences, structural units such as atoms [33-35], aminoacids [36], nucleotides [37-39], proteins [40, 41], RNAs [42], genes [43] or even organisms [44] play the role of nodes in the CNs. Complementarily, the chemical bonds (in case of classic molecular graph), chemical reactions, nucleotide-nucleotide hydrogen bonds [45, 46], amino acid-amino acid spatial contacts [47], metabolic pathways steps [48], protein-protein interactions [49], RNA-RNA co-expression [50], gene-gene regulation [51] or any other kind of structural or functional relationships play the role of edges in the CNs. These CNs can be used to characterize overall similarity/dissimilarity relationships [52] as well as properties such as bipartivity [53], community structure [54] or small world property [55], in order to unravel complex statistical properties of the phenomena under study. In a previous work, our group used drug-drug CNs to study the overall properties of the activity profiles of several anti-fungals against different species. The CNs were assembled based on the activity predictions made with a TI-QSAR model [56].

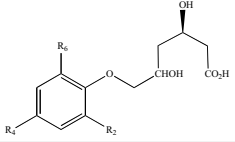
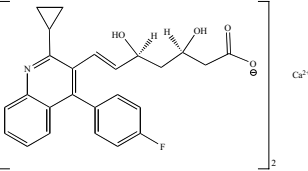
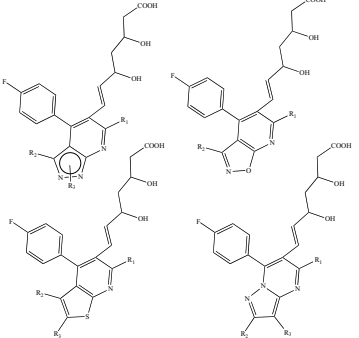
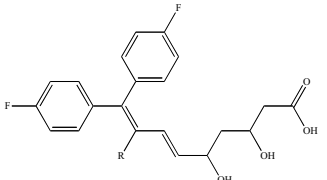
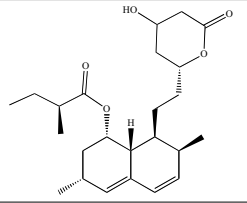
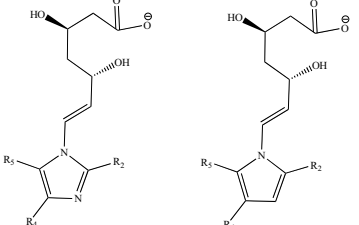
In the work described here, we review and comment different QSAR, CNs and other theoretic studies about HMGCoA reductase inhibitors.

1. Analysis of Tetrahedral Carbon in QSAR Studies.
2. A New HMG-CoA Reductase Inhibitor, NK-104: QSAR Studies.
3. Development of pharmacophoric model of pyridine and pyrimidine analogs as HMG-CoA inhibitors.
4. 3-Hydroxy-3-Methylglutaryl-Coenzyme A Reductase: Molecular Modeling, 3D QSAR, Inhibitor Design.
5. Potential Novel Inhibitors of Human HMG-CoA Reductase by Combining CoMFA 3D QSAR Modeling and Virtual Screening.
6. 3D QSAR and CoMSIA of HMGRIs by comparative molecular similarity indices analysis.
7. QSAR model with topological descriptors for predicting HMGRIs.
8. QSAR-based complex networks.

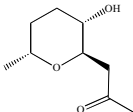
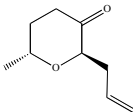
2. ANALYSIS OF TETRAHEDRAL CARBON IN QSAR STUDIES.

In one article by Prabhakar [57], an attempt has been made to outline a rule-based procedure to address the groups and atoms attached to the tetrahedral carbon in a quantitative manner. For this, computation of a few conceptual parameters, namely, flexibility of rotation, probability of availability, and net detachability, have been defined for the atoms/groups attached to the tetrahedral carbon. This has been used in a case study to explore structure-activity relations in some HMG-CoA reductase inhibitors that are showed in the Table 1. In this case, a new parameter was defined, net detachability (ND), to parameterize the ease of rotation and accessibility of R (R represents the substituent group on the main structural frame (M) of the molecule). The parameter ND is a ratio of rotational flexibility of R on the CM-CT bond and probability of availability of Ra, Rb, and/or Rc coupled with a constant phase difference between *R/S* isomeric centers, wherever necessary. Here, in a case study, the introduced concept was used to parameterize the flexibility of rotation with availability of the varying substituent group, R6,

Table 1.

Author	Compounds	Structures	Study	Ref.
Yenamandra <i>et al.</i>	Dihydroxy hexanoic acids		QSAR	[57]
Mutsukado <i>et al.</i>	NK-104		QSAR	[58]
Kashhedikar <i>et al.</i>	pyridine analogs and pyrimidine analogs		QSAR	[59]
Motoc	nonadienoic acids		3D-QSAR	[61]
Wan <i>et al.</i>	statins		3D-QSAR CoMFA	[64]
Chakraborti <i>et al.</i>	imidazolyls N-pyrrolyl heptenoates		3D-QSAR CoMSIA	[75]

(Table 1) contd....

Author	Compounds	Structures	Study	Ref.
García <i>et al.</i>	statins, decaresictins		QSAR	[62]
García <i>et al.</i>	statins, decaresictins		Complex Networks	[62]

of 6-aryloxy-3,5-dihydroxyhexanoic acids. This has been used as an independent parameter in the correlation analysis of HMGR inhibitory activity of these compounds.

3. A NEW HMG-CoA REDUCTASE INHIBITOR, NK-104: QSAR STUDIES

In this paper, Mutsukado Motoo and Suzuki Mikio, carried out searches for new HMG-CoA reductase inhibitors which led to the identification of NK-104 (Pitavastatin) [58] (Table 1), a type of statins, which are a class of drugs that lower cholesterol levels in people. They lower cholesterol by inhibiting the enzyme HMG-CoA reductase, which is the rate-limiting enzyme of the mevalonate pathway of cholesterol synthesis. Inhibition of this enzyme in the liver results in decreased cholesterol synthesis as well as increased synthesis of Low-Density Lipoprotein (LDL) receptors, resulting in an increased clearance of LDL from the bloodstream. Statins act by competitively inhibiting HMGCoA reductase, the first committed enzyme of the HMG-CoA reductase pathway. Because statins are similar to HMG-CoA on a molecular level they take the place of HMG-CoA in the enzyme and reduce the rate by which it is able to produce mevalonate, the next molecule in the cascade that eventually produces cholesterol, as well as a number of other compounds. This ultimately reduces cholesterol via several mechanisms. One of the most distinguishable structural features of NK-104 is possessing a cyclopropyl group on the central ring. QSAR analyses were performed by the method of regression analysis; a variety of physicochemical and steric parameters, including similarity indexes were used. The results showed that the activity was best correlated to the 7:1 ratio combined similarity index of shape and charge to a reference molecule (NK-104). This suggested the following: (1) The receptor-interaction is primarily steric and to a lesser electrostatic; (2) The interaction site should be a small pocket of limited size in both depth and width; (3) The cyclopropyl group would fit very well into that limited site with very good shape complementarity. Thus, NK-104 is indeed a fully optimized compound in this series.

4. DEVELOPMENT OF PHARMACOPHORIC MODEL OF PYRIDINE AND PYRIMIDINE ANALOGS AS HMG-CoA INHIBITORS

QSAR was established by Saxena, *et al.* [59] on a series of thirty-eight compounds of four different sets of condensed pyridine and pyrimidine analogs (see Table 1), for their hydroxymethyl glutaryl coenzyme (HMG-CoA) reductase inhibitor activity, in order to understand the essential structural requirement for binding with receptor, in terms of common biophoric and secondary sites employing APEX-3D software. Among several 3D pharmacophoric models with different sizes and arrangements, one model was selected based on $r^2 = 0.8$, $p < 0.001$, match equivalent to 0.38 and all the 38 compounds were considered. The generated 3D pharmacophore models can be useful for rational design and for identifying novel active compounds through a database search. In the past years

many successful applications in medicinal chemistry clearly demonstrate the usefulness of this ligand-based computational program [60]. A pharmacophore model or hypothesis, as defined within the Catalyst package, is a collection of chemical features placed in 3D space that represent the most important characteristics of a ligand to bind to its receptor or have a certain biological affinity. The results suggested that hydrophobicity, hydrogen acceptor and optimum steric refractivity play a dominant role in the inhibition of HMG-CoA reductase. The information obtained from the present study can be used to design and predict more potent molecules as HMG-CoA reductase inhibitors, prior to their synthesis.

5. 3-HYDROXY-3-METHYLGLUTARYL-COENZYME A REDUCTASE: MOLECULAR MODELING, 3D QSAR, INHIBITOR DESIGN

The 9,9-bis(4-fluorophenyl)-3,5-dihydroxy-8-(substituted)-6,8-nonadienoic acid analogues (see Table 1) represent a novel class of HMG-CoA reductase inhibitors developed in our laboratories. Motoc [61] delineated from inhibitory potency values the main topographical [60] and physicochemical features of the binding site probed by substituents attached to the C8 position of the analogues. Using a combination of receptor mapping and 3D QSAR techniques, it was possible to determine a logical candidate for the conformation of the inhibitor bound to the receptor and to derive a reliable 3D QSAR which related the HMG-CoA reductase inhibitory potency to the shape and size of both the binding site and C8-substituent of the inhibitor. They showed that the 3D QSAR derived here affords predictive utility.

6. POTENTIAL NOVEL INHIBITORS OF HUMAN HMG-CoA REDUCTASE BY COMBINING COMFA 3D QSAR MODELING AND VIRTUAL SCREENING

3-Hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR) catalyzes the formation of mevalonate. In many classes of organisms, this is the committed step leading to the synthesis of essential compounds, such as cholesterol. However, a high level of cholesterol is an important risk factor for coronary heart disease, for which an effective clinical treatment is to block HMGR using inhibitors like statins. The structures of catalytic portion of human HMGR complexed with six different statins [62] had been determined by a delicate crystallography study [63], which established a solid basis of structure and mechanism for the rational design, optimization, and development of even better HMGR inhibitors. In this study, Zhang [64] developed three-dimensional Quantitative Structure-Activity Relationship (3D QSAR) with Comparative Molecular Field Analysis (CoMFA) which was performed on a training set of up to 35 statins and statin-like compounds. Predictive models were established by using two different ways: (1) Models-fit, obtained by SYBYL conventional fit-atom molecular alignment rule, had cross-validated coefficients (q^2) up to 0.652 and regression coefficients (r^2) up to 0.977. Models-dock, obtained by FlexE by docking com-

pounds into the HMGR active site, had cross-validated coefficients (q^2) up to 0.731 and regression coefficients (r^2) up to 0.947. These models were further validated by an external testing set of 12 statins and statin-like compounds. Integrated with CoMFA + 3D-QSAR predictive models, molecular surface property (electrostatic and steric) mapping and structure-based (both ligand and receptor) virtual screening were employed to explore potential novel hits for the HMGR inhibitors. A representative set of eight new compounds of non-statin-like structures but with high pIC50 values were sorted out in the present study.

7. 3D QSAR AND COMSIA OF HMGRIS BY COMPARATIVE MOLECULAR SIMILARITY INDICES ANALYSIS

QSAR based on 2D or 3D structural information of the ligands alone, and involving three methods; Hologram QSAR (HQSAR), CoMFA, and Comparative Molecular Similarity Indices Analysis (CoMSIA) were developed in other work. CoMSIA of a set of 29 imidazolyl and N-pyrrolyl heptenoates (see Table 1) had been performed to find out the structural requirements for 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) inhibitory activity by Ramasamy Thilagavathi *et al.* [65] The HMG like side chain, a common moiety of statins, was used to align the molecules. The results allow the design of new chemical entities with high potency. For all those compounds, R6 is the major varying group and can be represented by M-R6. Here, these compounds were considered for the study of the spatial disposition of R6 vs HMGR inhibitory activity. The obtained equations show the correlations of the HMGR inhibitory activity of these analogues with the hydrophobicity, van der Waals volume, and/or ND of the varying substituent groups and structural units.

8. QSAR MODEL WITH TOPOLOGICAL DESCRIPTORS FOR PREDICTING HMGRIS

Efficient drugs such as statins or mevinic acids are inhibitors of the rate-limiting enzyme of cholesterol biosynthesis, HMGR, an enzyme responsible for the double reduction of 3-hydroxy-3-methyl-glutaryl coenzyme A into mevalonic acid. These compounds promoted the synthesis and evaluation of new inhibitors for HMGR, named HMGRIs. The high number of possible candidates creates the necessity of Quantitative Structure-Activity Relationship models in order to guide the HMGRi synthesis. There are two main problems with the reported QSAR models: the homogeneous series of the compounds and the chirality of many candidates. We proposed for the first time a QSAR model for a very large and heterogeneous series of HMGRIs [62]. The model was based on the Topological Indices (TIs) of molecular structures [66].

Some examples are the pyrimidine- and pyrrole-substituted compounds, that were evaluated for their ability to inhibit partially purified rat liver HMG-CoA reductase *in vitro* [67]. Inhibition of rat liver HMG-CoA reductase was measured by an *in vitro* assay procedure based on the direct conversion of or 3-hydroxy-3-methyl [3-14C]-glutaryl-CoA into [14C] mevalonolactone. The enzyme preparation and assay procedures used in this study were the same as those described in reference [67]. The HMG-CoA reductase inhibitory effect was examined through preparation of rat liver microsomes from Sprague-Dawley rats, which had been allowed free access to ordinary diets containing 2% cholestyramine and water for 2 weeks. The microsomes were purified as described in reference [67] too. We also performed an LDA in order to obtain a QSAR model that allowed us to classify or not a molecule as a potential HMGRi. The QSAR database is shown in the Supplement-

tary Materials of [62]. The equation of the QSAR model found was the following.

Where, N is the number of compounds used for training, Rc is the canonical regression coefficient, λ is the Wilks's statistic parameter, F is the Fisher ratio and p is the level of error. The high value of Rc = 0.7 (Rc may vary from 0 to 1) indicated a strong correlation and consequently supports the quality of the model. [68] On the other hand, the lower $\lambda = 0.54$ value found points to a better class or group separation (HMGRIs vs non-HMGRIs). Additionally, the F-test can reject the hypothesis of group overlapping with a level of error (p-level) < 0.001. Thus, we considered that the model performs a good separation of both groups. More direct evidence of the fitting quality are the high values of Sensitivity, Specificity and Accuracy [69] for both training and cv series (see also Table 2 in Ref. [62]). The *a priori* classification probability could be estimated by means of the receiver operating characteristics (ROC) curve [70] (Fig. 3 showed in Ref. [62]) that this is not a random model, but a statistically significant one, since the area under the ROC curve is significantly higher than the area under the random classifier curve equal to 0.5 [7]. The heteroscedasticity of a large set can be studied using a simple graphical method based on the examination of the variable residuals included in the model. (Fig. 5 showed in Ref. [62]). Due to the robustness of the LDA multivariate statistical techniques, the predictive ability and interference reached by using the proposed model should not be affected (see Fig. 3). The part including TIs of the QSAR can be used for the score prediction of a given potential HMGRi without taking into consideration chirality; however, if we also use the part including CTIs of the model, we can clearly differentiate between stereo-isomer derivatives and the same basic scaffold. In many of these cases, it is difficult to obtain all the possible isomers of the same compounds by organic synthesis; as a result, many of them have not been assayed biologically yet [71]. Consequently, the present QSAR model becomes an invaluable tool to provide the first virtual exploration of the molecular diversity of these compounds related to HMGRi. These predictions may be useful to guide the rational synthesis of the most promising candidates among the high number of potential candidates to be assayed.

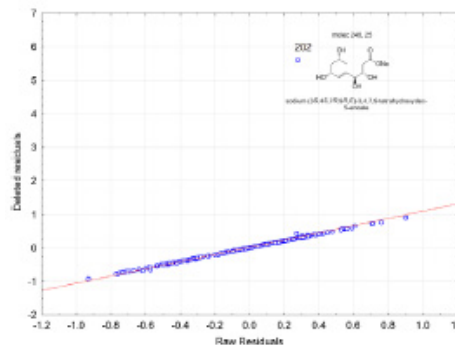


Fig. (3). Residuals versus deleted residuals plot for the LDA model.

9. QSAR-BASED COMPLEX NETWORKS

We constructed a drug-drug CNs and determined their complexity. The use of CNs for the investigation of data structure in this

$$HMGRi - score = 0.169 \cdot Cc - 0.029 \cdot Svd + 0.561 \cdot D - 0.002 \cdot W - 3.992 \cdot Nrc + 12.415 \cdot Bn + 12.167 \quad (1)$$

$$N = 183 \quad Rc = 0.7 \quad \lambda = 0.54 \quad F = 25.15 \quad p - level < 0.001$$

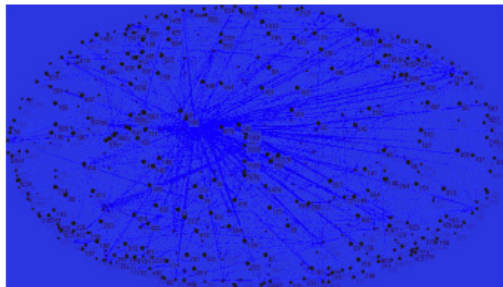


Fig. (4). Different layouts for cGC0.1.

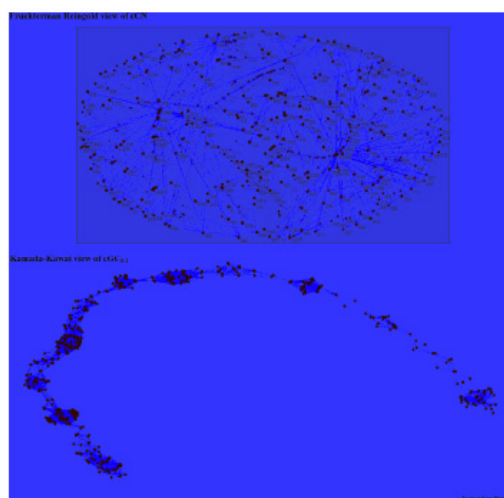


Fig. (5).

kind of problems by using large databases is nowadays an emerging and active field of research [72]. After calculating the two classes of DDD_{ij} distances ($cDDD_{ij}$ and $ncDDD_{ij}$) between all pairs of similar isomers (PSIs), we arranged them in two matrices (one that measured chirality and another which did not). Next, we explored different values of cut-off ranging, and we found that, for extremely low values of the cut-off, we obtained many unconnected nodes for very large values, all PSIs are linked to each other in the CN. Both extremes are useless because we need a CN that, for a given HMGR chiral compound (the input), could point out only few of the possible similar candidates to be assayed or obtained (the output). At the same time, in practical terms, the CN must be able to predict all possible input molecules (not many unconnected nodes). We calculated the probabilities of one node to be unconnected $p(u) = (\text{number of unconnected nodes})/1628$ and the probability $p(d) = (\text{average node degree})/1628 = z/1628$ in case all nodes have the largest degree (connecting all-with-all) by using different cut-off values (Fig. 8 in Ref. [62]). The parameter z (average node degree) represents the average number of HMGRIs, similar to one compound in the CN. A possible CN could be one with a 0.5 cutoff, having $p(u) = 0.007$ and $p(z) = 0.049$. It means that, of a total of 1628 chiral compounds for a given input candidate, we can select a series of $z = 79.5$ similar compounds and the CN is able to predict only 12 out of the 1628 compounds. However, for a more narrow search, we selected the CN determined by the cut-off equal to 0.1,

having $p(u)$ equal to 0.143 and $p(z)$ equal to 0.011. We also studied the effect of chirality on the CN construction with respect to CNs derived without considering the chirality. We can conclude that the omission of the chirality determines a transformation of the cCN into a ncCN, that leads us to node condensation or grouping, similar to other examples of complex networks. [73] The ncCN and cCN (cut-off = 0.1) are described in (Table SM4 and Table SM5 from the Supplementary Material of Ref. [62]) and plotted by Figs. (4 and 5).

The most important result is the reduction of the large universe of HMGRi isomers to a basic set of compounds in order to obtain the best candidates for the experimental assays. CentiBin has been used to reduce the CN by extracting the Giant Component (GC) [74] of the chiral compound CN with a cut-off equal to 0.1 (cGC0.1). The cGC0.1 contains 479 nodes (isomers) and 7074 edges (PSIs); the density of PSIs is only 0.62%. In cGC0.1, the average topological distance (D_c) for all PSIs is 14.96 and the diameter (D) is 49.5 (largest D). This CN is very important for practical use because it contains all the HMGRi chiral isomers with the most important degree of similarity in all their 1628 compounds. At first glance, cGC0.1 seems to be a very complex and highly interconnected network, but the visualization, after processing with Kamada-Kawai algorithms (Fig. 5), demonstrates that this component simplifies the information and could be of use as guide for future synthetic effort.

CONCLUSIONS

In the last years theoretical study of inhibiting of HMGR has increased; 2D-QSAR, 3D-QSAR (CoMSIA, CoMFA) and Complex Network. Different types of molecules were used, pyrimidines, pyridines, decarestrictins, statins, pyrrolidines, etc, chiral molecules and non-chiral molecules, etc; and in the QSAR studies was employed several molecular descriptors. All these methods allow to discover new inhibitory molecules, more active molecules and to reject new molecules before being synthesized.

ACKNOWLEDGEMENTS

We are grateful to the Xunta de Galicia (IN-CITE08PXIB314255PR) for partial financial support.

REFERENCES

- [1] Shimokata, K.; Yamada, Y.; Kondo, T.; Ichihara, S.; Izawa, H.; Nagata, K.; Murohara, T.; Ohno, M.; Yokota, M. Association of gene polymorphisms with coronary artery disease in individuals with or without nonfamilial hypercholesterolemia. *Atherosclerosis*, **2004**, *172* (1), 167-73.
- [2] Smilde, T. J.; van Wissen, S.; Wollersheim, H.; Kastelein, J. J.; Stalenhoef, A. F. Genetic and metabolic factors predicting risk of cardiovascular disease in familial hypercholesterolemia. *Neth. J. Med.*, **2001**, *59* (4), 184-95.
- [3] Leitersdorf, E.; Eisenberg, S.; Eliav, O.; Friedlander, Y.; Berkman, N.; Dann, E. J.; Landsberger, D.; Sehayek, E.; Meiner, V.; Wurm, M.; et al. Genetic determinants of responsiveness to the HMG-CoA reductase inhibitor fluvastatin in patients with molecularly defined heterozygous familial hypercholesterolemia. *Circulation*, **1993**, *87* (4 Suppl), III35-44.
- [4] Salazar, L. A.; Hirata, M. H.; Quintao, E. C.; Hirata, R. D. Lipid-lowering response of the HMG-CoA reductase inhibitor fluvastatin is influenced by polymorphisms in the low-density lipoprotein receptor gene in Brazilian patients with primary hypercholesterolemia. *J. Clin. Lab. Anal.*, **2000**, *14* (3), 125-31.
- [5] Suzuki, M.; Iwasaki, H.; Fujikawa, Y.; Sakashita, M.; Kitahara, M.; Sakoda, R. *Bioorg. Med. Chem.*, **2001**, *9*(2727).
- [6] Gonzalez-Diaz, H.; Prado-Prado, F. J.; Santana, L.; Uriarte, E. Unify QSAR approach to antimicrobials. Part 1: predicting antifungal activity against different species. *Bioorg. Med. Chem.*, **2006**, *14*(17), 5973-80.

- [7] Helguera, A. M.; Rodríguez-Borges, J. E.; García-Mera, X.; Fernandez, F.; Cordeiro, M. N. *J. Med. Chem.* **2007**, *50*(1537).
- [8] Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castanedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Sanchez, A. M.; Torrens, F.; Castro, E. A. Atom, atom-type and total molecular linear indices as a promising approach for bioorganic and medicinal chemistry: theoretical and experimental assessment of a novel method for virtual screening and rational design of new lead anthelmintic. *Bioorg. Med. Chem.*, **2005**, *13*(4), 1005-20.
- [9] González-Díaz, H.; Olazábal, E.; Santana, L.; Uriarte, E.; Castañedo, N. QSAR study of anticoccidial activity for diverse chemical compounds: prediction and experimental assay of trans-2-(2-Nitrovinyl)furan. *Bioorg. Med. Chem.*, **2007**, *15*(962-968).
- [10] Murcia-Soler, M.; Perez-Gimenez, F.; Garcia-March, F. J.; Salabert-Salvador, M. T.; Diaz-Villanueva, W.; Medina-Casamayor, P. Discrimination and selection of new potential antibacterial compounds using simple topological descriptors. *J. Mol. Graph Model*, **2003**, *21*(5), 375-90.
- [11] Prado-Prado, F. J.; Gonzalez-Diaz, H.; 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*(2), 897-902.
- [12] Mathur, K. C.; Gupta, S.; Khadikar, P. V. Topological modelling of analgesia. *Bioorg. Med. Chem.*, **2003**, *11*(8), 1915-28.
- [13] Calabuig, C.; Anton-Fos, G. M.; Galvez, J.; Garcia-Domenech, R. New hypoglycaemic agents selected by molecular topology. *Int. J. Pharm.*, **2004**, *278*(1), 111-8.
- [14] González, M. P.; González-Díaz, H.; Molina Ruiz, R.; Cabrera, M. A.; Ramos de Armas, R. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*(1192).
- [15] Votano, J. R.; Parham, M.; Hall, L. H.; Kier, L. B.; Oloff, S.; Tropsha, A.; Xie, Q.; Tong, W. Three new consensus QSAR models for the prediction of Ames genotoxicity. *Mutagenesis*, **2004**, *19*(5), 365-77.
- [16] Urbina, J. A.; Lazzardi, K.; Marchan, E.; Visbal, G.; Aguirre, T.; Piras, M. M.; Piras, R.; Maldonado, R. A.; Payares, G.; de Souza, W. *Antimicrob. Agents Chemother.*, **1993**, *37*(580).
- [17] Santana, L.; Uriarte, E.; González-Díaz, H.; Zagotto, G.; Soto-Otero, R.; Mendez-Alvarez, E. A QSAR model for in silico screening of MAO-A inhibitors. Prediction, synthesis, and biological assay of novel coumarins. *J. Med. Chem.*, **2006**, *49*(3), 1149-56.
- [18] Marrero-Ponce, Y. Linear indices of the "molecular pseudograph's atom adjacency matrix": definition, significance-interpretation, and application to QSAR analysis of flavone derivatives as HIV-1 integrase inhibitors. *J. Chem. Inf. Comput. Sci.*, **2004**, *44*(6), 2010-26.
- [19] Vilar, S.; Santana, L.; Uriarte, E. Probabilistic neural network model for the in silico evaluation of anti-HIV activity and mechanism of action. *J. Med. Chem.* **2006**, *49*(3), 1118-1124.
- [20] Marrero-Ponce, Y.; Khan, M. T.; Casanola Martin, G. M.; Ather, A.; Sultankhodzaev, M. N.; Torrens, F.; Rotondo, R. Prediction of Tyrosinase Inhibition Activity Using Atom-Based Bilinear Indices. *Chem. Med. Chem.*, **2007**, *2*(4), 449-478.
- [21] Casanola-Martin, G. M.; Marrero-Ponce, Y.; Khan, M. T.; Ather, A.; Sultan, S.; Torrens, F.; Rotondo, R. TOMOCOMD-CARDD descriptors-based virtual screening of tyrosinase inhibitors: evaluation of different classification model combinations using bond-based linear indices. *Bioorg. Med. Chem.*, **2007**, *15*(3), 1483-503.
- [22] Casanola-Martin, G. M.; Marrero-Ponce, Y.; Khan, M. T.; Ather, A.; Khan, K. M.; Torrens, F.; Rotondo, R. Dragon method for finding novel tyrosinase inhibitors: Biosilico identification and experimental *in vitro* assays. *Eur. J. Med. Chem.*, **2007**, *42*(11-12), 1370-81.
- [23] González-Díaz, H.; Sanchez, I. H.; Uriarte, E.; Santana, L. Symmetry considerations in Markovian chemicals 'in silico' design (MARCH-INSIDE) I: central chirality codification, classification of ACE inhibitors and prediction of sigma-receptor antagonist activities. *Comput. Biol. Chem.*, **2003**, *27*(3), 217-27.
- [24] Marrero-Ponce, Y.; Castillo-Garit, J. A. 3D-chiral Atom, Atom-type, and Total Non-stochastic and Stochastic Molecular Linear Indices and their Applications to Central Chirality Codification. *J. Comput. Aided Mol. Des.*, **2005**, *19*(6), 369-83.
- [25] Fabian, W. M.; Stampfer, W.; Mazur, M.; Uray, G. Modeling the chromatographic enantioseparation of aryl- and hetarylcarbinols on ULMO, a brush-type chiral stationary phase, by 3D-QSAR techniques. *Chirality*, **2003**, *15*(3), 271-5.
- [26] Golbraikh, A.; Tropsha, A. QSAR modeling using chirality descriptors derived from molecular topology. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*(1), 144-54.
- [27] Golbraikh, A.; Bonchev, D.; Tropsha, A. Novel chirality descriptors derived from molecular topology. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*(1), 147-58.
- [28] Castillo-Garit, J. A.; Marrero-Ponce, Y.; Torrens, F. *Bioorg. Med. Chem.*, **2006**, *14*(2398).
- [29] de Julián-Ortiz, J. V.; de Gregorio Alapont, C.; Ríos-Santamarina, I.; García-Doménech, R.; Gálvez, J. J. *Mol. Graph. Model.*, **1998**, *16*(14).
- [30] Castillo-Garit, J. A.; Marrero-Ponce, Y.; Torrens, F.; Rotondo, R. Atom-based stochastic and non-stochastic 3D-chiral bilinear indices and their applications to central chirality codification. *J. Mol. Graph Model*, **2007**, *26*(1), 32-47.
- [31] Zhang, Z.; Grigorov, M. G. Similarity networks of protein binding sites. *Proteins*, **2006**, *62*(2), 470-8.
- [32] Bonchev, D.; Buck, G. A. From molecular to biological structure and back. *J. Chem. Inf. Model*, **2007**, *47*(3), 909-17.
- [33] Ivanciuc, O.; Ivanciuc, T.; Klein, D. J. Quantitative structure-property relationships generated with optimizable even/odd Wiener polynomial descriptors. *SAR QSAR Environ. Res.*, **2001**, *12*(1-2), 1-16.
- [34] Ivanciuc, O.; Ivanciuc, T.; Klein, D. J.; Seitz, W. A.; Balaban, A. T. Wiener index extension by counting even/odd graph distances. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*(3), 536-49.
- [35] Ivanciuc, O. QSAR comparative study of Wiener descriptors for weighted molecular graphs. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*(6), 1412-22.
- [36] Gupta, N.; Mangal, N.; Biswas, S. Evolution and similarity evaluation of protein structures in contact map space. *Proteins*, **2005**, *59*(2), 196-204.
- [37] Gan, H. H.; Pasquali, S.; Schlick, T. Exploring the repertoire of RNA secondary motifs using graph theory: implications for RNA design. *Nucleic Acids Res.*, **2003**, *31*(11), 2926-2943.
- [38] Marrero-Ponce, Y.; Nodarse, D.; González-Díaz, H.; Ramos de Armas, R.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. A. Nucleic Acid Quadratic Indices of the "Macromolecular Graph's Nucleotides Adjacency Matrix". Modeling of Footprints after the Interaction of Paromomycin with the HIV-1 Psi-RNA Packaging Region. *Int. J. Mol. Sci.*, **2004**, *5*(276-293).
- [39] González-Díaz, H.; Agüero-Chapin, G.; Varona, J.; Molina, R.; Delogu, G.; Santana, L.; Uriarte, E.; Gianni, P. 2D-RNA-Coupling Numbers: A New Computational Chemistry Approach to Link Secondary Structure Topology with Biological Function. *J. Comput. Chem.*, **2007**, *28*(1049-1056).
- [40] Rose, A.; Schraegle, S. J.; Stahlberg, E. A.; Meier, I. Coiled-coil protein composition of 22 proteomes—differences and common themes in subcellular infrastructure and traffic control. *BMC Evol. Biol.*, **2005**, *5*(66).
- [41] Estrada, E.; Rodríguez-Velazquez, J. A. Subgraph centrality in complex networks. *Phys. Rev. E*, **2005**, *71*(5 Pt 2), 056103.
- [42] Yu, X.; Lin, J.; Masuda, T.; Esumi, N.; Zack, D. J.; Qian, J. Genome-wide prediction and characterization of interactions between transcription factors in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **2006**, *34*(3), 917-27.
- [43] Guido, N. J.; Wang, X.; Adalsteinsson, D.; McMillen, D.; Hasty, J.; Cantor, C. R.; Elston, T. C.; Collins, J. J. A bottom-up approach to gene regulation. *Nature*, **2006**, *439*(7078), 856-60.
- [44] Williams, R. J.; Berlow, E. L.; Dunne, J. A.; Barabasi, A. L.; Martinez, N. D. Two degrees of separation in complex food webs. *Proc. Natl. Acad. Sci. U.S.A.*, **2002**, *99*(20), 12913-6.
- [45] Marrero-Ponce, Y.; Castillo Garit, J. A.; Nodarse, D. Linear indices of the 'macromolecular graph's nucleotides adjacency matrix' as a promising approach for bioinformatics studies. Part 1: prediction of paromomycin's affinity constant with HIV-1 psi-RNA packaging region. *Bioorg. Med. Chem.*, **2005**, *13*(10), 3397-404.
- [46] González-Díaz, H.; de Armas, R. R.; Molina, R. Markovian negentropies in bioinformatics. I. A picture of footprints after the interaction of the HIV-1 Psi-RNA packaging region with drugs. *Bioinformatics*, **2003**, *19*(16), 2079-87.

- [47] Vullo, A.; Frascioni, P. Prediction of protein coarse contact maps. *J. Bioinform. Comput. Biol.*, **2003**, *1*(2), 411-31.
- [48] Lange, B. M.; Ghasseman, M. Comprehensive post-genomic data analysis approaches integrating biochemical pathway maps. *Phytochemistry*, **2005**, *66*(4), 413-51.
- [49] Chou, K. C.; Cai, Y. D. Predicting protein-protein interactions from sequences in a hybridization space. *J. Proteome Res.*, **2006**, *5*(2), 316-22.
- [50] Yu, X.; Lin, J.; Zack, D. J.; Qian, J. Computational analysis of tissue-specific combinatorial gene regulation: predicting interaction between transcription factors in human tissues. *Nucleic Acids Res.*, **2006**, *34*(17), 4925-36.
- [51] Margolin, A. A.; Nemenman, I.; Basso, K.; Wiggins, C.; Stolovitzky, G.; Dalla Favera, R.; Califano, A. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, **2006**, *7* Suppl 1(S7).
- [52] González-Díaz, H.; González-Díaz, Y.; Santana, L.; Ubeira, F. M.; Uriarte, E. Proteomics, networks and connectivity indices. *Proteomics*, **2008**, *8*(750-778).
- [53] Estrada, E. Protein bipartivity and essentiality in the yeast protein-protein interaction network. *J. Proteome Res.*, **2006**, *5*(9), 2177-84.
- [54] Newman, M. E. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E Stat Nonlin Soft Matter Phys.*, **2006**, *74*(3 Pt 2), 036104.
- [55] Kleinberg, J. M. Navigation in a small world. *Nature*, **2000**, *406*(6798), 845.
- [56] González-Díaz, H.; Prado-Prado, F. Unified QSAR and Network-Based Computational Chemistry Approach to Antimicrobials, Part 1: Multispecies Activity Models for Antifungals. *J. Comput. Chem.*, **2008**, *29*(656-657).
- [57] Prabhakar, Y. S. Analysis of Tetrahedral Carbon in QSAR Studies. *J. Chem. Inf. Comput. Sci.*, **1999**, *39*(650-653).
- [58] Mutsukado, M.; Suzuki, M. A New HMG-CoA Reductase Inhibitor, NK-104: QSAR Studies. *Kagaku Toronkai, Kozo Kassei Sokan Shinpojiumu Koen Yoshishu*, **2000**, *23-28*(192-193).
- [59] Saxena, M.; Soni, L. K.; Gupta, A. K.; Wakode, S. R.; Saxena, A. K.; Kaskhedikar, S. G. Development of pharmacophoric model of condensed pyridine and pyrimidine analogs as hydroxymethyl glutaryl coenzyme A reductase inhibitors. *Indian J. Biochem & Biophys. (IJB)*, **2006**, *43*(1), 32-36.
- [60] Web, C. S. o. t.
- [61] Motoc, I.; Sit, S. Y.; Harte, W. E.; Balasubramanian, N.; Wright, J. J. 3-Hydroxy-3-Methylglutaryl-Coenzyme A Reductase: Molecular Modeling, Three-Dimensional Structure-Activity Relationships, Inhibitor Design. *QSAR & Combinatorial Sci.*, **2006**, *10*(1), 30-35.
- [62] García, I.; Munteanu, C. R.; Fall, Y.; Gomez, G.; Uriarte, E.; González-Díaz, H. QSAR and complex network study of the chiral HMGR inhibitor structural diversity. *Bioorg. Med. Chem.*, **2009**, *17*(1), 165-75.
- [63] Istvan; Deisenhofer. *Science*, **2001**, *292*(1160-1164).
- [64] Qing, Y. Z.; Jian, W.; Xin, X.; Guang, F.; Yang, Y. L. R.; Jun, J. L.; Hui, W.; Yu, G. Structure-Based Rational Quest for Potential Novel Inhibitors of Human HMG-CoA Reductase by Combining CoMFA 3D QSAR Modeling and Virtual Screening. *J. Comb. Chem.*, **2007**, *9*(131-138).
- [65] Ramasamy Thilagavathi, R.; Kumar, R.; Aparna, V.; Sobhia, M. E.; Gopalakrishnan, B.; Chakraborti, A. K. Three-dimensional quantitative structure (3D QSAR) activity relationship studies on imidazolyl and N-pyrrolyl heptenoates as 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) inhibitors by comparative molecular similarity indices analysis (CoMSIA). *Bioorg. Med. Chem. Lett.*, **2005**, *15*(1027-1032).
- [66] González-Díaz, H.; Munteanu, C. R. Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and Social Networks, Transworld Research Network: Kerala, India **2010**.
- [67] Kuroda, M.; Endo, A. *Biochim. Biophys. Acta*, **1977**, *486*(70).
- [68] Hill, T.; Lewicki, P. STATISTICS Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining, StatSoft: Tulsa **2006**.
- [69] Lilien, R. H.; Farid, H.; Donald, B. R. Probabilistic disease classification of expression-dependent proteomic data from mass spectrometry of human serum. *J. Comput. Biol.*, **2003**, *10*(6), 925-46.
- [70] James, A.; Hanley, B. J. M. *Radiology*, **1982**, *143*(29).
- [71] Gómez, G.; Rivera, H.; García, I.; Estévez, L.; Fall, Y. The furan approach to carbocyclic systems. Synthesis of cyclohexane derivatives from butenolides through an intramolecular Michael addition. *Tetrahedron Lett.*, **2005**, *46*(35), 5819-5822.
- [72] Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M.; Hwang, D. U. Complex networks: Structure and dynamics. *Physics Reports*, **2006**, *424*(175-308).
- [73] Bianconi, G.; Barabasi, A. L. Bose-Einstein condensation in complex networks. *Phys. Rev. Lett.*, **2001**, *86*(24), 5632-5.
- [74] Bornholdt, S.; Schuster, H. G. *Handbook of Graphs and Complex Networks: From the Genome to the Internet*, WILEY-VCH GmbH & CO. KGaA: Weinheim **2003**.
- [75] Thilagavathi, R.; Kumar, R.; Aparna, V.; Sobhia, M. E.; Gopalakrishnan, B.; Chakraborti, A. K. Three-dimensional quantitative structure (3-D QSAR) activity relationship studies on imidazolyl and N-pyrrolyl heptenoates as 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) inhibitors by comparative molecular similarity indices analysis (CoMSIA). *Bioorg. Med. Chem. Lett.*, **2005**, *15*(4), 1027-32.

QSAR, Docking, and CoMFA Studies of GSK3 Inhibitors

Isela García*, Yagamare Fall and Generosa Gómez

Department of Organic Chemistry, University of Vigo, Spain

Abstract: GSK-3 inhibitors are interesting candidates to develop anti-Alzheimer compounds. GSK-3 β are also interesting as anti-parasitic compounds active against *Plasmodium falciparum*, *Trypanosoma brucei*, and *Leishmania donovani*; the causative agents for Malaria, African Trypanosomiasis and Leishmaniasis. The high number of possible candidates creates the necessity of Quantitative Structure-Activity Relationship models in order to guide the GSK3 (Glycogen Synthase Kinase 3 inhibitor) synthesis. In this work, we revised different computational studies for a very large and heterogeneous series of GSK-3Is. First, we revised QSAR studies with conceptual parameters such as flexibility of rotation, probability of availability, etc. We then used the method of regression analysis and QSAR studies in order to understand the essential structural requirement for binding with receptor. Next, we reviewed 3D-QSAR, CoMFA and CoMSIA with different compounds to find out the structural requirements for GSK-3 inhibitory activity.

Keywords: QSAR, Alzheimer, SAR, parasitic, fungi.

INTRODUCTION

The enzyme Glycogen Synthase Kinase 3 (GSK-3), so called because of its implication in the phosphorylation of glycogen synthase, is a serine/threonine kinase involved in many cellular processes. Glycogen Synthase Kinase-3 (GSK-3) is a serine-threonine kinase encoded by two isoforms in mammals, termed GSK-3 α and GSK-3 β [1]. Initially GSK-3 was implicated in muscle energy storage and metabolism, but since its cloning, a more generalized role in cellular regulation has emerged, highlighted by the wide array of substrates controlled by this enzyme that includes cytoplasmic proteins and nuclear transcription factors. GSK-3 targets encompass proteins implicated in Alzheimer's disease (AD), neurological disorders, in the *wnt* and insulin signaling pathway, glycogen and protein synthesis, regulation of transcription factors [2], embryonic development, cell proliferation and adhesion, tumorigenesis, apoptosis [3], circadian rhythm, etc. GSK-3 β knock-out mice die *in utero* [4], whereas GSK-3 α knock-out mice are viable and display improved glucose tolerance in response to glucose load and elevated hepatic glycogen storage and insulin sensitivity [5, 6].

Alzheimer's disease [7] is a serious and degenerative disorder that explains the gradual loss of neurons, and in spite of the efforts realized by the big pharmacists of the world, is still not a very clear reason of this pathology, because at present, it is the most recent reason of dementia in main elders. The fundamental characteristic of Alzheimer's disease is the presence in the brain of two injuries: the Neurofibrillary Tangles (NFTs) (Fig. 1), that are formed by paired helical filaments (PHF) whose main component is Tau Protein kinase (TPK), and insoluble β -amyloid ($A\beta$) plaques (Fig. 2) that are associated with active microglia [5, 8]. NFTs are composed of hyper-phosphorylated forms of the microtubule-associated protein tau, whereas $A\beta$ is derived from the proteolytic cleavage of β -amyloid precursor protein (APP). Active GSK-3 appears in neurons with pre-tangle changes [9] and there is an increased GSK3 activity in the frontal cortex in AD as evidenced by immunoblotting for GSK-3 phosphorylated at Tyr216 [10]. GSK3 expression is up-regulated in the hippocampus of AD patients and in post-synaptosomal supernatants derived from AD brain, although the latter study reports that there is no increase in GSK-3 enzymatic activity [5]. The functions of GSK-3 and its implication in various human diseases have triggered an active search for potent and selective GSK-3 inhibitors [11] in the last years, as shown in Fig. 3. Studies of

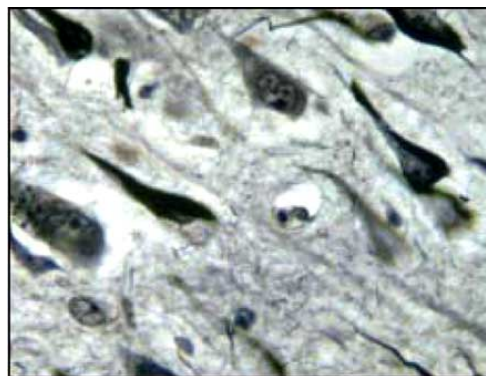


Fig. (1). Neurofibrillary tangles (NFTs).

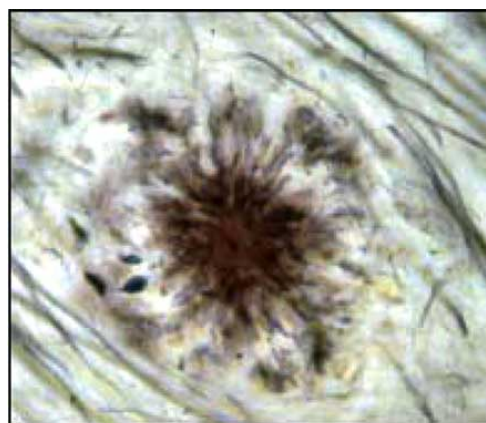


Fig. (2). β -amyloid ($A\beta$) plaques.

*Address correspondence to this author at the Department of Organic Chemistry, University of Vigo, Spain; Tel: +34986813679; Fax: +34986812262; E-mail: iselapintos@yahoo.es

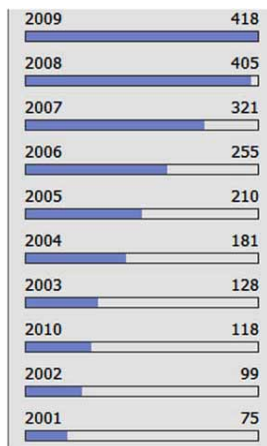


Fig. (3). Search of potent and selective GSK-3 inhibitors.

GSK-3 homologues in various organisms have revealed physiological roles for the enzyme in differentiation, cell fate determination, and spatial patterning to establish bilateral embryonic symmetry. Purified GSK-3 α and GSK-3 β exhibit similar biochemical and substrate properties [12, 13], and is known that in the phosphorylation of TPK/ Tau Protein Kinase takes part actively Glycogen Synthase Kinase 3 β (GSK-3 β), which not only plays a fundamental role in the synthesis of the glycogen (where it was identified by the first time), but it is very important in several processes as cellular signs, metabolic control, embryogenesis, cellular death and oncogenesis [14], and it is related to a wide range of neurodegenerative [15] diseases, bipolar mood disorders [16] and diabetes, hence the inhibition of this enzyme is accepted as a promising therapeutic strategy.

In 1988 Ishiguro and col. [17] isolated one enzyme when they were studying an extract of the brain and noticed the presence of paired helical filaments of Tau Protein Kinase, typical injury of Alzheimer's disease. TPKI and TPKII are the two kinases implied in this process and they found that TPKI has an identical structure to GSK-3 β . At this moment, there is an increasing interest in the evaluation of kinases from unicellular parasites as targets for potential new anti-parasitic drugs. The evolutionary difference between unicellular kinases and their human homologues might be sufficient to allow the design of parasite-specific inhibitors. The *Plasmodium falciparum* genome contains 65 genes that encode kinases, including three forms of GSK-3. An initial study showed that *P. falciparum* exports PfGSK-3 to the cytoplasm of host erythrocytes (which are devoid of GSK-3), where it colocalizes with parasite-generated membrane structures known as Maurer's clefts. The function of PfGSK-3 is unknown, but the presence of PfCK1, a CK1 homologue, in infected red blood cell supports the hypothesis that both kinases play a role in regulating the strong circadian rhythm of the parasite, which is responsible for the circadian fevers that are the characteristics of this infectious disease [18].

On the other hand, the vector-borne parasitic disease African trypanosomiasis, caused by members of the *Trypanosoma brucei* complex, is a serious health threat. It is estimated that 300,000 to 500,000 humans in sub-Saharan Africa are infected. If the disease is left inadequately treated, it often has a fatal outcome. Once infection is established, safe and effective therapy is critically important, yet it has been difficult to achieve. Despite the critical need, the available therapies are becoming less satisfactory due to the rising level of resistance to the available drugs, the long period of treat-

ment required to achieve a cure, and the unacceptable and sometimes severe adverse effects associated with current therapies [19]. An urgent priority is to identify and validate new targets for the development of safe, effective, and inexpensive therapeutic alternatives. Compounds that inhibit *T. brucei* GSK-3 activity and not host GSK-3 might be required for therapy for pregnant women and infants, in that GSK-3 regulates proteins critical in development, such as the *wnt* gene product. However, optimization of the selectivity of drug candidates for parasite kinases becomes an issue due to the highly conserved amino acids and protein conformation of the catalytic domains [20-23]. Understanding the differences in the substrate binding properties and the three-dimensional structures between mammalian and parasite GSK-3 enzymes is important for the optimization of selected target inhibitors for drug development [24, 25].

In this sense, the development of QSARs using simple molecular indices appears to be a promising alternative or complementary technique to drug-protein docking, high-throughput screening and combinatorial chemistry techniques. Almost all QSAR techniques are based on the use of molecular descriptors, which are numerical series that codify useful chemical information and enable correlations between statistical and biological properties [26-28]. Many recent works, for instance those published by González-Díaz *et al.* (to cite only one example), illustrated that the parameters used to seek QSAR-like predictive models of low-weight molecules may also be used to predict properties of proteins, RNAs, protein interaction networks (PINs), cerebral cortex, disease spreading, and other more complex systems [29-46]. In any case, a large number of examples have been published in which the use of molecular descriptors has become a rational alternative to massive synthesis and screening of compounds in medicinal chemistry [47, 48]; in the Fig. 4 we can see progress of GSK-3-QSAR studies in the last years. Indeed, experience has shown that the use of models fitted with large data sets of chemicals works as well as the use of models built from a series of homologous compounds and is also a more general method that can be applied in a broad spectrum of cases. The principal deficiency in the use of some molecular indices concerns their lack of physical meaning. In this respect, the introduction of novel molecular indices must obey physicochemical laws in order to ensure a theoretically rigorous interpretation of the results. In the work described here, we review and comment different theoretical studies about GSK-3 inhibitors in the last years.

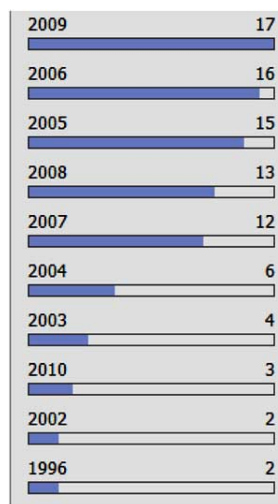


Fig. (4). Progress of GSK-3-QSAR studies in the last years.

1. 3D-QSAR studies on thiadiazolidinone derivatives as GSK-3 inhibitors.
2. 3D-QSAR and Docking studies of selective GSK-3 β inhibitors.
3. QSAR modeling of the inhibition of GSK-3.
4. Docking and 3D-QSAR for pyrimidin-2-amines as GSK-3 inhibitors.
5. 3D-QSAR for GSK-3 β inhibition by indirubin analogues.
6. 3D-QSAR and Docking for bisarylmaleimides as GSK-3, CDK-2 and CDK-4 inhibitors.
7. Construction of the pharmacophore model of GSK-3 inhibitors.
8. 3D-QSAR and Docking of pyrazolopyrimidines as GSK-3 β inhibitors.
9. 2D image-based approach for prediction of GSK-3 β inhibitors.
10. Free-Wilson QSAR Analysis to predict kinase selectivity profiles.
11. CoMFA and Docking for pyrazolo[3,4-*b*]pyrid[az]ines as GSK-3 inhibitors.
12. 2D and 3D-QSAR models for prediction of indirubin derivatives GSK-3 β inhibitors
13. Multi-target QSAR *in silico* screening for GSK-3 inhibitors.

REVIEW AND DISCUSSION

1. 3D-QSAR Studies on Thiadiazolidinone Derivatives as GSK-3 Inhibitors

In this article, Ana Martínez [49] reported a biological study following a SAR study about 2,4-disubstituted thiadiazolidinones (TDZD) (see Table 1), compounds that were described as the first ATP-noncompetitive GSK-3 inhibitors, where different structural modifications in the heterocyclic ring aimed to test the influence of each heteroatom. They synthesized various compounds such as hydantoins, dithiazolidindiones, rhodanines, maleimides, and triazoles and were screened as GSK-3 inhibitors. A CoMFA analysis was also performed highlighting the molecular electrostatic field connection in the interaction of TDZDs with GSK-3. CoMFA models were calculated for each molecular field considered alone or in combination, and the best PLS analysis was obtained by combining steric and electrostatic fields, leading to a good correlation between the IC₅₀ values predicted from the principal components extracted from these two fields and the experimental data ($r^2 = 0.922$, $q^2 = 0.654$, and $n = 5$). Since the relative contributions of steric and electrostatic molecular field were 42% and 58%, respectively. Favorable steric regions are found around the aromatic substituent attached to N4 as well as in the vicinity of N2. An increase of the steric field in the region between N2 and S is detrimental for the inhibitory activity, and a negatively charged region around the ring is favorable for activity. Moreover, first mapping studies indicate two binding modes which in turn might imply relevant differences in the mechanism that underlie the inhibitory activity of TDZDs.

2. 3D-QSAR and Docking Studies of Selective GSK-3 β Inhibitors

Bureau, R. [50] *et al.* in this paper carried out a 3D-QSAR (CoMFA) study with several GSK-3 β inhibitors. The cocrystallographic data of GSK-3 β vs 3-anilino-4-arylmaleimide were suitable to compare 3D-QSAR results with experimental intermolecular interactions. CoMFA analysis served to start the study of a new compound, a thieno[2,3-*b*]pyrrolizinone derivative (see Table 1) as GSK-3 inhibitor, because this study was not about the interactions registered in the active site. This comparison based on docking and simulation approaches allowed to confirm one preferential orientation of this ligand inside the active site, explaining the relationship

with the reference 3-anilino 4-arylmaleimide derivatives and its biological affinity.

3. QSAR Modeling of the Inhibition of GSK-3

Katritzky, A. R. [51] *et al.*, made a Quantitative Structure-Activity Relationship (QSAR) model of the *in vitro* biological activity (pIC₅₀) of 277 inhibitors of Glycogen Synthase Kinase-3 (GSK-3) (3-anilino-4-arylmaleimides), calculating different molecular descriptors by CODESSA PRO technique, geometrical, topological, quantum mechanical, and electronic descriptors. The linear (multilinear regression) and nonlinear (artificial neural network) models obtained, linked the structures to their reported activity pIC₅₀. Each multilinear model was verified by leave-one-out and internal validation methods confirmed the correct prediction of the inhibitory activity of 3-anilino-4-arylmaleimides (see Table 1), but was found that the Artificial Neural Network (ANN), which was built for all the data points, presented superior prediction over the multilinear models. These studies gave an insight into the dominant role played by the electrostatic, bonding, and steric interactions on the modulation of the inhibitory activity, hence the nature of GSK-3 inhibitor interaction was found to be electrostatic.

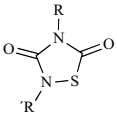
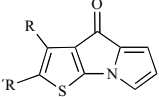
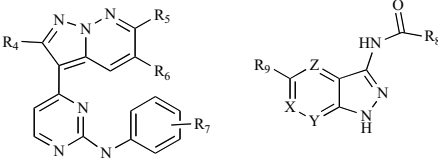
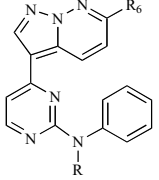
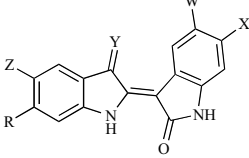
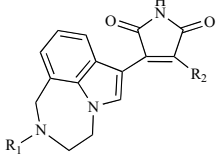
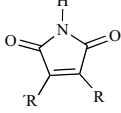
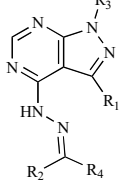
4. Docking and 3D-QSAR for Pyrimidin-2-amines as GSK-3 Inhibitors

In this paper, Guo [52] *et al.* carried out a study of Glycogen Synthase Kinase 3 (GSK-3) inhibition. Molecular docking and 3D-QSAR approaches were used for the study of the interaction mode of a series of N-phenyl-4-pyrazolo[1,5-*b*]pyridazin-3-ylpyrimidin-2-amine compounds (see Table 1) with human GSK-3. In the 3D-QSAR studies, the molecular alignment and conformation determination were of special importance. Flexible docking (AutoDock3.0.5) was used for the determination of 'active' conformation and molecular alignment. Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) were used to carry out 3D-QSAR models of 80 N-phenyl-4-pyrazolo[1,5-*b*]pyridazin-3-ylpyrimidin-2-amine compounds. The r^2 values of CoMFA were 0.870 and the r^2 values of CoMSIA were 0.861. The predictive ability of these models was validated by 10 compounds of the test set. Mapping these models back to the topology of the active site of GSK-3 led to a better understanding of the vital N-phenyl-4-pyrazolo[1,5-*b*]pyridazin-3-ylpyrimidin-2-amines-GSK-3 interactions. The results showed that combination of ligand-based and receptor-based modeling is a powerful approach to develop 3D-QSAR models.

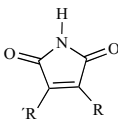
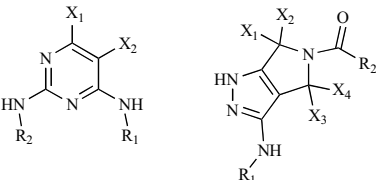
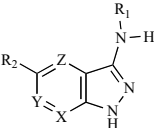
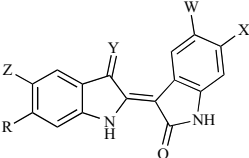
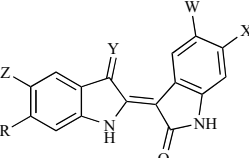
5. 3D-QSAR for GSK-3 β Inhibition by Indirubin Analogues

Yongjun Jiang [53] *et al.* studied Indirubin analogues that showed favorable inhibitory activity targeting GSK-3 β , which is closely related to the property and position of substituents. Two methods were used to build 3D-QSAR models for indirubin derivatives. The conventional 3D-QSAR (ligand-based) studies were performed based on the lower energy conformations employing atom fit alignment rule. The receptor-based 3D-QSAR models were also derived using bioactive conformations obtained by docking the compounds (see Table 1) to the active site of GSK-3. Conclusions of models based on two methods are similar and reliable. The ligand based CoMFA model had cross-validate coefficient q^2 of 0.790 with five components and conventional r^2 of 0.923, the CoMSIA model with q^2 of 0.780 for seven components and r^2 of 0.916 are obtained. These values indicate that the models were robust. The receptor-based CoMFA model had cross-validation coefficient (q^2) of 0.563 and regression coefficient (r^2) of 0.897 for four components. q^2 and r^2 for CoMSIA model were 0.766 and 0.908 with five components using the combination of steric + electric + hydrophobic fields, respectively. The results indicated that both ligand-based and receptor-based were feasible tools to build 3D-QSAR models.

Table 1. Summary of QSAR Studies of GSK-3 Inhibitors

Author	Compounds	Structure	Study	Refs.
Martinez, A.	TDZD derivatives		SAR 3D-QSAR	[49]
Bureau, R.	Pyrrolizinones		3D-QSAR Docking	[50]
Katritzky, A. R.	Aarylmaimides		QSAR	[51]
Guo, Z.	Pyrimidines		CoMSIA CoMFA	[52]
Jiang, Y.	Indirubin derivatives		3D-QSAR	[53]
Dessalew, N.	Maleimides		3D-QSAR	[54]
Feng-Chao, J.	Maleimides		CoMFA	[55]
Bharatam, P. V.	Pyrimidines		CoMFA CoMSIA	[56]

(Table 1) Contd....

Author	Compounds	Structure	Study	Refs.
Freitas, M. P.	Maleimides		QSAR ANN	[57, 58]
Sciabola, S.	Pyrimidines and Pyrrolopyrazoles		QSAR	[59]
Bharatam, P. V.	Pyrazoles		CoMFA Docking	[60]
Viney, L.	Indirubin derivatives		2D-QSAR 3D-QSAR	[61]
García, I.	Maleimides		QSAR	-

6. 3D-QSAR and Docking for Bisarylmaleimides as GSK-3, CDK-2 and CDK-4 Inhibitors

In this paper, Dessalew [54] developed three individual CoMFA models using 36 compounds of bisarylmaleimide series (see Table 1) to correlate with the GSK3, CDK2 and CDK4 inhibitory potencies. Selective Glycogen Synthase Kinase 3 (GSK3) inhibition over Cyclin Dependent Kinases such as Cyclin Dependent Kinase 2 (CDK2) and Cyclin Dependent Kinase 4 (CDK4) was an important requirement for improved therapeutic profile of GSK-3 inhibitors. The concepts of selectivity and additivity fields had been employed in developing selective CoMFA models for these related kinases. These models showed a satisfactory statistical significance: CoMFA-GSK3 (r^2_{con} , r^2_{cv} : 0.931, 0.519), CoMFA-CDK2 (0.937, 0.563), and CoMFA-CDK4 (0.892, 0.725). Three different selective CoMFA models were then developed using differences in pIC_{50} values. These three models showed a superior statistical significance: (i) CoMFA-Selective1 (r^2_{con} , r^2_{cv} : 0.969, 0.768), (ii) CoMFA-Selective 2 (0.974, 0.835) and (iii) CoMFA-Selective3 (0.963, 0.776). The selective models were found to outperform the individual models in terms of the quality of corre-

lation and were found to be more informative in pinpointing the structural basis for the observed quantitative differences of kinase inhibition. An in-depth comparative investigation was carried out between the individual and selective models to gain an insight into the selectivity criterion. To further validate this approach, a set of new compounds were designed which showed selectivity and were docked into the active site of GSK-3, using FlexX based incremental construction algorithm. The contour maps of the three different selective CoMFA models revealed a way that could be employed to bias the inhibitory activity towards either of the receptors and helped in explaining the structural basis for the experimentally observed differences in activity. The validity of the selective models had been checked using PLS statistical parameters and was found that it provided a better quality model as compared to the conventional individual models. On the other hand, molecules that showed a good fit in the contours of the selective models were found to exhibit a better selectivity in the inhibitory potencies used to develop the selective model. The final validation criterion designed new molecules using the individual and selective models by making a suitable modification to the basic skeleton in the favorable contour regions.

7. Construction of the Pharmacophore Model of GSK-3 Inhibitors

Feng-Chao, J. [55] generated a three dimensional pharmacophore model for glycogen synthase kinase-3 (GSK-3) inhibitors. A dataset consisting of 89 compounds (see Table 1) was selected on the basis of the information content of the structures and activity data as required by the CATALYST system. The better model was selected with a good correlation coefficient, $r^2 = 0.95$. This model was able to predict the activity of other known GSK-3 inhibitors not included in the model generation, and could be used further to identify structurally diverse compounds with desired biological activity by virtual screening. The analysis of the CoMFA study showed that a compound had electrostatic interactions with side chains of K85 and S66 via the NH group in region C around the 4-phenyl group.

8. 3D-QSAR and Docking of Pyrazolopyrimidines as GSK-3 β Inhibitors

Bharatam [56] *et al.* in this paper studied pyrazolopyrimidines derivatives (see Table 1) as GSK-3 β inhibitors. In the present work, they carried out three-dimensional Quantitative Structure Activity Relationship (3D-QSAR) studies on novel class of pyrazolopyrimidine derivatives as GSK-3 inhibitors reported to have improved cellular activity. Docked conformation of the most active molecule in the series, showed desirable interactions in the receptor, was taken as template for alignment of the molecules. CoMFA and CoMSIA models were generated using 49 molecules in training set. By applying leave-one-out (LOO) cross-validation study, r^2_{cv} values of 0.53 and 0.48 for CoMFA and CoMSIA, respectively and non-cross-validated (r^2_{ncv}) values of 0.98 and 0.92 were obtained for CoMFA and CoMSIA models, respectively, which indicates a good internal predictive ability of both models. The predictive ability of CoMFA and CoMSIA models was determined using a test set of 12 molecules, excluded from the model derivation was used, which gave predictive correlation coefficients (r^2_{pred}) of 0.47 and 0.48, respectively, indicating good external predictive ability of the model. This field higher $r^2_{bs} = 0.99$ and 0.96 for CoMFA and CoMSIA respectively, further supported the statistical validity of the developed models. Based upon the information derived from CoMFA and CoMSIA contour maps, we have identified some key features that explain the observed variance in the activity and have been used to design new pyrazolopyrimidine derivatives. The designed molecules showed better binding affinity in terms of estimated docking scores with respect to the already reported systems; hence suggesting that newly designed molecules can be more potent and selective towards GSK-3 β inhibition.

9. 2D Image-Based Approach for Prediction of GSK-3 β Inhibitors

Freitas, M. P. made different Quantitative Structure-Activity Relationship (QSAR) studies to evaluate GSK-3 inhibitors and their activity. A multivariate image analysis (MIA) [57] applied to QSAR analysis, whose descriptors were derived from pixels of two-dimensional (2D) chemical structures that may be built by using any appropriate drawing software, was used to model 17 Glycogen Synthase Kinase 3 (GSK-3) inhibitors. Calibration was carried out using Partial Least Squares (PLS) regression, and an r^2 value of 0.93 was obtained for four latent variables. Leave-one-out cross-validation and a robustness test were performed, and the model proved to be a suitable alternative QSAR approach for modeling this series of compounds.

In other paper, Freitas, M. P. [58] *et al.* calculated different descriptors with Dragon software through three different feature selection methods, namely Genetic Algorithm (GA), Successive Projections Algorithm (SPA), and fuzzy rough set Ant Colony Optimization (fuzzy rough set ACO). Each set of selected descriptors was regressed against the bioactivities of a series of glycogen

synthase kinase-3 β inhibitors, through linear and nonlinear regression methods, namely Multiple Linear Regression (MLR), Artificial Neural Network (ANN), and Support Vector Machines (SVM). While calibration with MLR yielded QSAR models only reasonably predictable, with r^2 ranging from 0.77 to 0.81 and r^2 test of 0.67 to 0.76, ANN and specially SVM were capable of estimating and predicting biological activities very accurately. According to PLS-based models, r^2 varied from 0.78 to 0.82 for the training set and from 0.70 to 0.80 for the test set. Leave-one-out (LOO) cross-validation experiments gave q^2 ranging from 0.76 to 0.79. These results were only comparable to MLR-based models, independent of the feature selection method used. Overall, Support Vector Machines (SVM) were suggested to be used as a regression method in QSAR studies where linear behavior is not expected or in those studies in which linear approaches do not work well.

10. Free-Wilson QSAR Analysis to Predict Kinase Selectivity Profiles

Sciabola, S. [59] *et al.* wrote this article about kinases that are involved in a variety of diseases such as cancer, diabetes, and arthritis. In recent years, many kinase small molecule inhibitors developed as potential disease treatments. Despite the recent advances, selectivity remains one of the most challenging aspects in kinase inhibitor design. To interrogate kinase selectivity, a panel of 45 kinase assays was developed in-house at Pfizer. The KSS is the Pfizer internal Kinase Selectivity Screening panel consisting of 45 different protein kinase bioassays selected based on bioinformatics and structural data to provide maximal coverage across subfamilies within the kinome. Here, they presented an application of *in silico* Quantitative Structure Activity relationship (QSAR) models to extract rules from this experimental screening data and make reliable selectivity profile predictions for all compounds enumerated from virtual libraries. They also proposed the construction of R-group selectivity profiles by deriving their activity contribution against each kinase using QSAR models. Such selectivity profiles can be used to provide better understanding of subtle structure selectivity relationships during kinase inhibitor design.

11. CoMFA and Docking for Pyrazolo[3,4-*b*]pyrid[az]ines as GSK-3 Inhibitors

Bharatam [60] *et al.* in this paper carried out a 3D-QSAR study of pyrazolo[3,4-*b*]pyrid[az]ine derivatives, which includes non-selective and selective GSK-3 inhibitors. The training set of CoMFA models were made from a 59 molecules (see Table 1). A test set containing 14 molecules was used to validate the CoMFA models. The CoMFA model generated by applying leave-one-out (LOO) cross-validation study gave a higher cross-validated correlation coefficient, $r^2_{cv} = 0.60$, which indicated a good internal prediction of the model; and conventional $r^2_{conv} = 0.97$; the predictive correlation coefficient, $r^2_{pred} = 0.55$, indicated good external predictive ability of CoMFA model. The 3D-QSAR study of pyrazolo[3,4-*b*]pyrid[az]ine derivatives led to the identification of regions of importance for steric and electronic interactions. Validation based on the molecular docking was also made to explain the structural differences between the selective and non-selective molecules in the given series of molecules.

12. 2D and 3D-QSAR Models for Prediction of Indirubin Derivatives GSK-3 β Inhibitors

The 3D-QSAR study made by Viney, L. *et al.* [61] of indirubin derivatives (see Table 1) was carried out with molecular descriptors calculated by CODESSA and Molconn-Z and a 3D-QSAR study based on the principle of the alignment of pharmacophoric features by PHASE module of Schrodinger. Statistically significant 2-D ($r^2 = 0.93$) and 3-D ($r^2 = 0.97$) QSAR models were generated using 36 molecules in the training set. The predictive correlation coefficient: $r^2_{pred} = 0.6$ for 2D- and $r^2_{pred} = 0.91$ for 3-D model. These studies

Table 2. Some Rules Used to Classify GSK-3 Inhibitors in Different Assay Conditions

Parameter	Isoform	Specie	Condition ^a
enzyme test (<i>in vitro</i>)			
%	β	-	0
cKi	α	-	>100
IC ₅₀ (nM)	α	-	>2000
IC ₅₀ (nM)	β	-	>2000
IC ₅₀ (μM)	β	-	>2
IC ₅₀ (μM)	nd	-	>2
IC ₅₀ (μM)	α/β	-	>2
IC ₅₀ E-9 (M)	nd	-	>2
pIC ₅₀	β	-	0
p IC ₅₀	nd	-	0
parasite			
EC ₅₀ (μM)	-	<i>T. brucei</i>	>2
IC ₅₀ (ng/mL)	-	<i>P. falciparum</i> ^b	NA
ED50 (μM)	-	<i>L. donovani</i>	<5
IC ₅₀ (ng/mL)	-	<i>P. falciparum</i> ^c	NA
IC ₅₀ (μg/mL)	-	<i>L. donovani</i>	NA
IC ₅₀ (μM)	-	<i>P. falciparum</i>	>20
IC ₅₀ (μM)	-	<i>L. mexicana</i>	>2
IC ₅₀ (μM)	-	<i>P. falciparum</i>	>2
IC ₅₀ (μM)	-	<i>P. falciparum</i> D6	NA
IC ₅₀ (μM)	-	<i>P. falciparum</i> W2	NA
IC ₉₀ (μg/mL)	-	<i>L. donovani</i>	NA
bacteria			
IC ₅₀ (μg/mL)	-	<i>M. intracellulare</i>	NA
IC ₅₀ (μg/mL)	-	MRS	NA
IC ₅₀ (μg/mL)	-	<i>S. aureus</i>	NA
IC ₅₀ (μM)	-	<i>M. intracellulare</i>	-
IC ₅₀ (μM)	-	MRSA	-
MIC (μg/mL)	-	<i>M. tuberculosis</i>	NA
cell line			
IC ₅₀ (μg/mL)	-	HVC	NC
IC ₅₀ (μM)	-	Hep2	NA
IC ₅₀ (μM)	-	HT29	NA
IC ₅₀ (μM)	-	LMM3	NA
IC ₅₀ (μM)	-	PTP	>2
IC ₅₀ (μM)	-	RD	>2
IC ₅₀ (μM)	-	U937	>2

^a Condition to be classified as non-active compound (LDA class 0); nd = not determined, NA = not active, NC = not cytotoxicity; ^b Chloroquine resistant W2 clone; ^c chloroquine sensitive D6 clone.

showed that 3D-QSAR model is a better predictive model for the indirubins compared to 2D-QSAR model. 2D-QSAR studies demonstrated that activity of these inhibitors can be increased by changing the polarity, electronegativity of the atoms that might be forming H-bonds with the receptor binding site. This study also correlates with the novel 3D-QSAR study generated volume occluded maps, which demonstrated that the activity can be increased by changing the electronegativity of the N-atoms involved in the H-bond interactions with the binding site of the receptor. Also 3D-QSAR results demonstrated that the biological activity of the indirubin-like molecules could be increased by placing the hydrophobic groups at particular positions on the phenyl ring of indirubins.

13. Multi-Target QSAR *In Silico* Screening for GSK-3 Inhibitors

In this paper, we studied the development of a discriminant function [62] that allowed the classification of organic compounds as active or non-active which is the key step in the present approach for the discovery of GSK-3 inhibitors. It was therefore necessary to select a training data set of GSK-3 inhibitors containing wide structural variability. The selection of discriminant techniques instead of regression techniques was determined by the lack of homogeneity in the conditions under which these values were measured. As reported in different sources, numerous IC₅₀ values lie within a range rather than a single value. In other cases, the activity was not scored in terms of IC₅₀ values but was quoted as inhibitory percentages at a given concentration. In Table 2, we summarize different rules used to classify compounds as active or inactive GSK-3 inhibitors in different assay conditions. Once the training series had been designed, forward stepwise Linear Discriminant Analysis (LDA) was carried out in order to derive the QSAR; this equation confirmed our intuitive hypothesis and we could conclude that the deviation of the parameters of one compound from the average values for active compounds tested in a given assay condition (CACq) was very important for the prediction of this compound as active. In particular, for i-GSK-3 small deviations in total polarizability and molecular refractivity indices as well as local deviations in refractivity and electronegativity for heteroatom-bound hydrogen atoms seem to be especially determinant. Specifically, the introduction of this last parameter in the model coincide with the GSK-3 activity observed in organic acid compounds with [63] this type of model based only on atomic physicochemical parameters and drug connectivity may become an interesting alternative for fast computational pre-screening of large series of compounds in order to rationalize synthetic efforts [52, 60, 64-68] complementary to more elaborated techniques 3D-QSAR, CoMFA, and CoMSIA studies that depend on a detailed knowledge of 3D structure. In any case, the present model is of more general application than the other known methods that apply only to compounds tested in only one CAC and/or belonging to only one homogeneous structural class of compounds. A confirmation of this stamen is that the present classification function has given rise to an efficient separation of all compounds with Accuracy = 89.1% (training series) and Accuracy = 88.3% (validation series). The names, observed classification, predicted classification and subsequent probabilities for all 4508 compounds (see examples in Table 1) in training and average validation are given as supplementary material. This level of total Accuracy, Sensitivity and Specificity is considered as excellent by other researchers that have used LDA for QSAR studies; see for instance the works of Garcia-Domenech, R., Prado-Prado, F. J.; Marrero-Ponce, Y., etc [69-84].

ACKNOWLEDGMENT

We are grateful to the Xunta de Galicia (INCITE08PXIB-314255PR) for partial financial support.

REFERENCES

- [1] Woodgett JR. Molecular cloning and expression of glycogen synthase kinase-3/factor A. *EMBO J* 1990; 9: 2431-8.
- [2] Troussard AA, Tan C, Yoganathan TN, Dedhar S. Cell-extracellular matrix interactions stimulate the AP-1 transcription factor in an integrin-linked kinase- and glycogen synthase kinase 3-dependent manner. *Mol Cell Biol* 1999; 19: 7420-7.
- [3] Turenne GA, Price BD. Glycogen synthase kinase3 beta phosphorylates serine 33 of p53 and activates p53's transcriptional activity. *BMC Cell Biol* 2001; 2: 12-21.
- [4] Hoefflich KP, Luo J, Rubie EA, Tasao MS, Jin O, Woodgett JR. Requirement for glycogen synthase kinase-3beta in cell survival and NF-kappaB activation. *Nature* 2000; 406: 86-90.
- [5] Hooper C, Killick R, Lovestone S. The GSK3 hypothesis of Alzheimer's disease. *J Neurochem* 2008; 104: 1433-9.
- [6] MacAulay K, Doble BW, Patel S, *et al.* Glycogen synthase kinase 3alpha-specific regulation of murine hepatic glycogen metabolism. *Cell Metab* 2007; 6: 329-37.
- [7] Olson RE. Secretase inhibitors as therapeutics for Alzheimer's disease. *Annu Rep Med Chem* 2000; 35: 31-40.
- [8] Vehmas AK, Kawas CH, Stewart WF, Troncoso JC. Immune reactive cells in senile plaques and cognitive decline in Alzheimer's disease. *Neurobiol Aging* 2003; 24: 321-31.
- [9] Pei JJ, Braak H, Grundke-Iqbal I, Iqbal K, Winblad B, Cowburn RF. Distribution of active glycogen synthase kinase 3beta (GSK-3beta) in brains staged for Alzheimer disease neurofibrillary changes. *J Neuropathol Exp Neurol* 1999; 58: 1010-9.
- [10] Leroy K, Yilmaz Z, Brion JP. Increased level of active GSK-3beta in Alzheimer' disease and accumulation in argyrophilic grains and in neurones at different stages of neurofibrillary degeneration. *Neuropathol Appl Neurobiol* 2007; 33: 43-55.
- [11] Droucheau E. Plasmodium falciparum glycogen synthase kinase-3: molecular model, expression, intracellular localisation and selective inhibitors. *Biochim Biophys Acta* 2004; 1697: 181-96.
- [12] Woodgett JR. cDNA cloning and properties of glycogen synthase kinase-3. *Methods Enzymol* 1991; 200: 564-77.
- [13] Ali A, Hoefflich KP, Woodgett JR. Glycogen Synthase kinase-3: properties, functions, and regulation. *Chem Rev* 2001; 101: 2527-40.
- [14] Grimes CA, Jope RS. The multifaceted roles of glycogen synthase kinase 3B in cellular signaling. *Prog Neurobiol* 2001; 65: 391-426.
- [15] Nadri C, Lipska B, Kozlovsky N, Weinberger DR, Belmaker RH, Agam G. GSK-3 β levels and activity in a neurodevelopmental rat model of schizophrenia. *Dev Brain Res* 2003; 141: 33-7.
- [16] Gould TD, Zarate CA, Manji HK. Glycogen synthase kinase-3: a target for novel bipolar disorder treatments. *J Clin Psychiatry* 2004; 65: 10-21.
- [17] Ishiguro K, Ihara Y, Uchida T, Imahori K. A novel tubulin-dependent protein kinase forming a paired helical filament epitope on tau. *J Bio Chem* 1988; 104: 319-21.
- [18] Meijer L, Flajolet M, Geengard P. Pharmacological inhibitors of glycogen synthase kinase 3. *Trends Pharmacological Sci* 2004; 25: 471-80.
- [19] Fairlamb AH. Chemotherapy on human African trypanosomiasis: current and future prospects. *Trends Parasitol* 2003; 19: 488-94.
- [20] Copeland RA, Pompliano DL, Meek TD. Drug-target residence time and its implications for lead optimization. *Nat Rev Drug Discov* 2006; 5: 730-2.
- [21] Liao JJ. Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J Med Chem* 2007; 50: 409-24.
- [22] Pink R, Hudson A, Mouries MA, Bending M. Opportunities and challenges in antiparasitic drug discovery. *Nat Rev Drug Discov* 2005; 4: 727-40.
- [23] Plyte SE, Hughes K, Nikkolakaki E, Pulverer BJ, Woodgett JR. Glycogen Synthase Kinase-3: functions in oncogenesis and development. *Biochim Biophys Acta* 1992; 1114: 147-62.
- [24] Dajani R, Fraser E, Roe SM, *et al.* Crystal structure of glycogen synthase kinase-3 beta: structural basis for phosphate-primed substrate specificity and autoinhibition. *Cell (Cambridge, Massachusetts)* 2001; 105: 721-32.
- [25] Ojo KK, Gillespie RG, Riechers A, *et al.* Glycogen Synthase Kinase 3 is a potential drug target for african trypanosomiasis therapy. *Antimicrob Agents Chemother* 2008; 52: 3710-7.
- [26] Nunez MB, Maguna FP, Okulik NB, Castro EA. QSAR modeling of the MAO inhibitory activity of xanthenes derivatives. *Bioorg Med Chem Lett* 2004; 14: 5611-7.
- [27] Todeschini R, Consonni V. *Handbook of molecular descriptors*. USA: Wiley VCH 2000.
- [28] Freund JA, Poschel T. Stochastic process in physics, chemistry and biology. In: *Lecture Notes on Physics*. Berlin, Germany: Springer-Verlag 2000.
- [29] González-Díaz H, Munteanu CR. Topological indices for medicinal chemistry, biology, parasitology, neurological and social networks. Kerala: Transworld Research Network 2010.
- [30] González-Díaz H, Duardo-Sánchez A, Ubeira FM, *et al.* Review of MARCH-INSIDE & complex networks prediction of drugs: AD-MET, anti-parasite activity, metabolizing enzymes and cardiotoxicity proteome biomarkers. *Curr Drug Metabol* 2010; 11: 379-406.
- [31] Vina D, Uriarte E, Orallo F, Gonzalez-Diaz H. Alignment-free prediction of a drug-target complex network based on parameters of drug connectivity and protein sequence of receptors. *Mol Pharm* 2009; 6: 825-35.
- [32] Pérez-Montoto LG, Prado-Prado F, Ubeira FM, González-Díaz H. Study of parasitic infections, cancer, and other diseases with mass-spectrometry and quantitative proteome-disease relationships. *Curr Proteomics* 2009; 6: 246-61.
- [33] González-Díaz H, Prado-Prado F, Pérez-Montoto LG, Duardo-Sánchez A, López-Díaz A. QSAR models for proteins of parasitic organisms, plants and human guests: theory, applications, legal protection, taxes, and regulatory issues. *Curr Proteomics* 2009; 6: 214-27.
- [34] Concu R, Dea-Ayuela MA, Perez-Montoto LG, *et al.* Prediction of enzyme classes from 3D structure: a general model and examples of experimental-theoretic scoring of peptide mass fingerprints of Leishmania proteins. *J Proteome Res* 2009; 8: 4372-82.
- [35] Agüero-Chapin G, Varona-Santos J, de la Riva GA, *et al.* Alignment-free prediction of polygalacturonases with pseudofolding topological indices: experimental isolation from coffea arabica and prediction of a new sequence. *J Proteome Res* 2009; 8: 2122-8.
- [36] Gonzalez-Diaz H, Prado-Prado F, Ubeira FM. Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach. *Curr Top Med Chem* 2008; 8: 1676-90.
- [37] González-Díaz H, González-Díaz Y, Santana L, Ubeira FM, Uriarte E. Proteomics, networks and connectivity indices. *Proteomics* 2008; 8: 750-78.
- [38] Rodríguez-Soca Y, Munteanu CR, Dorado J, Rabuñal J, Pazos A, González-Díaz. Plasmod-PPI: A web-server predicting complex biopolymer targets in plasmodium with entropy measures of protein-protein interactions. *Polymer* 2010; 51: 264-73.
- [39] Munteanu CR, Dorado J, Pazos Sierra A, *et al.* In information theory analysis of complex networks: statistical methods and applications. Emmert-streib F, Dehmer M, Mehler A, Eds. USA: Springer-Verlag 2010.
- [40] González-Díaz H, Agüero-Chapin G, Munteanu CR, *et al.* In advances in genetics Research. Osborne MA, Eds. New York: Nova Sciences 2010; Vol. 1.
- [41] Rodríguez-Soca Y, Munteanu CR, Prado-Prado FJ, Dorado J, Pazos Sierra A, Gonzalez-Diaz H. Trypano-PPI: A web server for prediction of unique targets in trypanosome proteome by using electrostatic parameters of protein-protein interactions. *J Proteome Res* 2010; 9(2): 1182-90.
- [42] Ferino G, Delogu G, Podda G, Uriarte E, González-Díaz H. Clinical chemistry research (ISBN: 978-1-60692-517-1). In: Shamham BH, Mitchem CHL, Eds. NY: Nova Science Publisher 2009.
- [43] Concu R, Podda G, Uriarte E, González-Díaz H. *Handbook of computational chemistry research*. In: Robson CT, Collett CD, Eds. USA: Nova Science Publishers 2009.
- [44] González-Díaz H, Vilar S, Santana L, Uriarte E. Medicinal chemistry and bioinformatics - current trends in drugs discovery with networks topological indices. *Curr Top Med Chem* 2007; 7: 1025-39.
- [45] Gonzalez-Diaz H, Saiz-Urria L, Molina R, Gonzalez-Diaz Y, Sanchez-Gonzalez A. Computational chemistry approach to protein

- kinase recognition using 3D stochastic van der Waals spectral moments. *J Comput Chem* 2007; 28: 1042-8.
- [46] González-Díaz H, Pérez-Castillo Y, Podda G, Uriarte E. Computational chemistry comparison of stable/nonstable protein mutants classification models based on 3D and topological indices. *J Comput Chem* 2007; 28: 1990-5.
- [47] Estrada E, Uriarte E. Recent advances on the role of topological indices in drug discovery research. *Curr Med Chem* 2001; 8: 1573-88.
- [48] Estrada E, Uriarte E, Montero A, Teijeira M, Santana L, De Clercq E. A novel approach for the virtual screening and rational design of anticancer compounds. *J Med Chem* 2001; 43: 1975-85.
- [49] Martınez A, Alonso M, Castro A, et al. SAR and 3D-QSAR Studies on thiazolidinone derivatives: exploration of structural requirements for glycogen synthase kinase 3 inhibitors. *J Med Chem* 2005; 48: 7103-12.
- [50] Lescot E, Bureau R, Sopkova-de Oliveira Santos J, et al. 3D-QSAR and docking studies of selective GSK-3 β inhibitors. Comparison with a thieno[2,3-b]pyrrolizidine derivative, a new potential lead for GSK-3 β ligands. *J Chem Inf Model* 2005; 45: 708-15.
- [51] Katritzky AR, Pacureanu LM, Dobchev DA, Fara DC, Duchowicz PR, Karelson M. QSAR modeling of the inhibition of Glycogen Synthase Kinase-3. *Bioorg Med Chem* 2006; 14: 4987-5002.
- [52] Xiao J, Guo Z, Guo Y, Chu F, Sun P. Inhibitory mode of N-phenyl-4-pyrazolo[1,5-b]pyridazin-3-ylpyrimidin-2-amine series derivatives against GSK-3: molecular docking and 3D-QSAR analyses. *Protein Eng Des Sel* 2006; 19: 47-54.
- [53] Jiang Y, Zhang N, Zou J, et al. 3D QSAR for GSK-3 β inhibition by indirubin analogues. *Eur J Med Chem* 2006; 41: 373-8.
- [54] Dessalew N, Bharatam PV. 3D-QSAR and molecular docking study on bisarylmaleimide series as glycogen synthase kinase 3, cyclin dependent kinase 2 and cyclin dependent kinase 4 inhibitors: An insight into the criteria for selectivity. *Eur J Med Chem* 2007; 42: 1014-27.
- [55] Ling L, Li-Na Z, Feng-Chao J. Construction of the pharmacophore model of glycogen synthase kinase-3 inhibitors. *Chinese J Chem* 2007; 25: 892-7.
- [56] Bharatam PV, Dessalew N, Patel DS. 3D-QSAR and molecular docking studies on pyrazolopyrimidine derivatives as glycogen synthase kinase-3B inhibitors. *J Mol Graph Model* 2007; 25: 885-95.
- [57] Freitas MP. A 2D image-based approach for modelling some glycogen synthase kinase 3 inhibitors. *Med Chem Res* 2007; 16: 461-7.
- [58] Freitas MP, Goodarzi M, Jensen R. Feature selection and linear/nonlinear regression methods for the accurate prediction of glycogen synthase kinase-3 β inhibitory activities. *J Chem Informat Model* 2009; 49: 824-32.
- [59] Sciabola S, Stanton RV, Wittkopp S, et al. Predicting kinase selectivity profiles using free-wilson QSAR analysis. *J Chem Inf Model* 2008; 48: 1851-67.
- [60] Patel DS, Bharatam PV. Selectivity criterion for pyrazolo[3,4-b]pyrid[az]ine derivatives as GSK-3 inhibitors: CoMFA and molecular docking studies. *Eur J Med Chem* 2008; 43: 949-57.
- [61] Viney L, Kristam R, Saini JS, Kristam R, Karthikeyan NA, Balaji VN. QSAR models for prediction of glycogen synthase kinase-3 β inhibitory activity of indirubin derivatives. *QSAR Comb Sci* 2008; 27: 718-28.
- [62] Van Waterbeemd H. In *Chemometric methods in molecular design*. Van Waterbeemd H, Ed. New York: Wiley-VCH 1995; Vol. 2, pp. 265-82.
- [63] Konda VR, Desai A, Darland G, Bland JS, Tripp ML. Rho iso-alpha acids from hops inhibit the GSK-3/NF-kappaB pathway and reduce inflammatory markers associated with bone and cartilage degradation. *J Inflamm (Lond)* 2009; 6: 26-34.
- [64] Rochais C, Duc NV, Lescot E, et al. Synthesis of new dipyrrolo- and furopyrrolopyrazinones related to triptenones and their biological evaluation as potential kinases (CDKs1-5, GSK-3) inhibitors. *Eur J Med Chem* 2009; 44: 708-16.
- [65] Simon D, Benitez MJ, Gimenez-Cassina A, et al. Pharmacological inhibition of GSK-3 is not strictly correlated with a decrease in tyrosine phosphorylation of residues 216/279. *J Neurosci Res* 2008; 86: 668-74.
- [66] Jacquemard U, Dias N, Lansiaux A, et al. Synthesis of 3,5-bis(2-indolyl)pyridine and 3-[(2-indolyl)-5-phenyl]pyridine derivatives as CDK inhibitors and cytotoxic agents. *Bioorg Med Chem* 2008; 16: 4932-53.
- [67] Tavares FX, Boucheron JA, Dickerson SH, et al. N-Phenyl-4-pyrazolo[1,5-b]pyridazin-3-ylpyrimidin-2-amine as potent and selective inhibitors of glycogen synthase kinase 3 with good cellular efficacy. *J Med Chem* 2004; 47: 4716-30.
- [68] Olesen PH, Sorensen AR, Urso B, et al. Synthesis and *in vitro* characterization of 1-(4-aminofurazan-3-yl)-5-dialkylaminomethyl-1H-[1,2,3]triazole-4-carboxylic acid derivatives. A new class of selective GSK-3 inhibitors. *J Med Chem* 2003; 46: 3333-41.
- [69] Calabuig C, Anton-Fos GM, Galvez J, Garcia-Domenech R. New hypoglycaemic agents selected by molecular topology. *Int J Pharm* 2004; 278: 111-8.
- [70] Garcia-Garcia A, Galvez J, de Julian-Ortiz JV, et al. New agents active against *Mycobacterium avium* complex selected by molecular topology: a virtual screening method. *J Antimicrob Chemother* 2004; 53: 65-73.
- [71] Prado-Prado FJ, Ubeira FM, Borges F, Gonzalez-Diaz H. Unified QSAR & network-based computational chemistry approach to antimicrobials. II. Multiple distance and triadic census analysis of antiparasitic drugs complex networks. *J Comput Chem* 2009; 31: 164-73.
- [72] Prado-Prado FJ, Martinez de la Vega O, Uriarte E, Ubeira FM, Chou KC, Gonzalez-Diaz H. Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modeling and comparative multi-distance study of the giant components of antiviral drug-drug complex networks. *Bioorg Med Chem* 2009; 17: 569-75.
- [73] Prado-Prado FJ, de la Vega OM, Uriarte E, Ubeira FM, Chou KC, Gonzalez-Diaz H. Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modeling and comparative multi-distance study of the giant components of antiviral drug-drug complex networks. *Bioorg Med Chem* 2009; 17: 569-75.
- [74] Prado-Prado FJ, Borges F, Perez-Montoto LG, Gonzalez-Diaz H. Multi-target spectral moment: QSAR for antifungal drugs vs. different fungi species. *Eur J Med Chem* 2009; 44: 4051-6.
- [75] Prado-Prado FJ, Gonzalez-Diaz H, de la Vega OM, Ubeira FM, Chou KC. 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: 5871-80.
- [76] Prado-Prado FJ, Gonzalez-Diaz H, 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.
- [77] Marrero-Ponce Y, Khan MT, Casanola Martin GM, et al. Prediction of Tyrosinase Inhibition Activity Using Atom-Based Bilinear Indices. *ChemMedChem* 2007; 2: 449-78.
- [78] Marrero-Ponce Y, Meneses-Marcel A, Castillo-Garrit JA, et al. Predicting antitrichomonal activity: a computational screening using atom-based bilinear indices and experimental proofs. *Bioorg Med Chem* 2006; 14: 6502-24.
- [79] Meneses-Marcel A, Marrero-Ponce Y, Machado-Tugores Y, et al. A linear discrimination analysis based virtual screening of trichomonocidal lead-like compounds: outcomes of *in silico* studies supported by experimental results. *Bioorg Med Chem Lett* 2005; 15: 3838-43.
- [80] Marrero-Ponce Y, Diaz HG, Zaldivar VR, Torrens F, Castro EA. 3D-chiral quadratic indices of the 'molecular pseudograph's atom adjacency matrix' and their application to central chirality codification: classification of ACE inhibitors and prediction of sigma-receptor antagonist activities. *Bioorg Med Chem* 2004; 12: 5331-42.
- [81] Murcia-Soler M, Perez-Gimenez F, Garcia-March FJ, Salabert-Salvador MT, Diaz-Villanueva W, Medina-Casamayor P. Discrimination and selection of new potential antibacterial compounds using simple topological descriptors. *J Mol Graph Model* 2003; 21: 375-90.

- [82] Cercos-del-Pozo RA, Perez-Gimenez F, Salabert-Salvador MT, Garcia-March FJ. Discrimination and molecular design of new theoretical hypolipaeamic agents using the molecular connectivity functions. *J Chem Inf Comput Sci* 2000; 40: 178-84.
- [83] Estrada E, Vilar S, Uriarte E, Gutierrez Y. *In silico* studies toward the discovery of new anti-HIV nucleoside compounds with the use of TOPS-MODE and 2D/3D connectivity indices. 1. Pyrimidyl derivatives. *J Chem Inf Comput Sci* 2002; 42: 1194-203.
- [84] Cronin MT, Aptula AO, Dearden JC, *et al.* Structure-based classification of antibacterial activity. *J Chem Inf Comput Sci* 2002; 42: 869-78.

Trends in Bioinformatics and Chemoinformatics of Vitamin D Analogs and Their Protein Targets

Isela García*, Yagamare Fall and Generosa Gómez

Department of Organic Chemistry, University of Vigo, Spain

Abstract: 1R,25-Dihydroxyvitamin D₃, the hormonally active form of vitamin D₃, besides regulating the homeostasis of calcium and classical bone mineralization, also promotes cellular differentiation and induces some biological functions related to the immunological system. Extensive structure-function studies have shown that it is possible to modify the calcitriol structure to obtain vitamin D₃ analogs that are capable of inducing, in a selective manner, the biological functions related to the same hormone. In this article, we revised QSAR studies with conceptual parameters such as flexibility of rotation, probability of availability, etc. we then used the method of regression analysis and QSAR studies in order to understand the essential structural requirement for binding with receptor. Next, we reviewed Radial Distribution Function, 4D-QSAR, CoMFA and Docking with different compounds to find out the structural requirements for GSK-3 inhibitory activity.

Keywords: QSAR, COMFA, docking, vitamin D₃, calcitriol.

1. INTRODUCTION

The vitamin D₃, across the 1 α ,25-Dihydroxyvitamin D₃ or calcitriol (see Fig. 1), her hormonally active form, regulates calcium homeostasis and this multifunctional hormone is also involved in other cellular processes, including cell differentiation, immune system regulation, and gene transcription [1-3]. These biological activities are mediated by the vitamin D receptor (VDR) [4] that is present in different tissues (see Fig. 2). However, the clinical utility of calcitriol is limited by its hypercalcemic effects (see Fig. 3) in treatment of cancers and skin disorders such as psoriasis, immune disorders, or malignant tumors [5]; hence there is accordingly much interest in the design and synthesis of vitamin D analogs with more selective biological effects. Thousands of vitamin D analogs have been synthesized and their biological activities evaluated and many of them have already been developed, and some are under development as clinical agents for treating metabolic bone diseases, skin diseases. A limited number of studies has been devoted to the search for new potent vitamin D analogs; the analogs of calcitriol synthesized so far show modifications at the side chain and conformational analysis has been carried out in order to predict a potentially active side-chain structure [6, 7].

In this picture, Bioinformatics and Chemoinformatics methods may play an important role in the study of Vitamin D analogs and their protein targets as well. Specifically, Quantitative Structure-Activity Relationships (QSAR) studies are used as predictive tools for the molecular development [8, 9]. Unfortunately, QSAR studies are generally based on databases considering only structurally parent compounds acting against one single microbial species. Up to today, there are nearly 1600 molecular descriptors that, in

principle, can be generalized and used to solve the former problem [10]. Many of these indices are known as molecular Topological Indices (TIs) or simply invariants of a molecular graph. In a recent review, our group has discussed recent advances in the field [11]. In addition to QSAR, Bioinformatics and Chemoinformatics methods which are useful to study vitamin D analogs and their targets may include techniques like Comparative Molecular Field Analysis (CoMFA), drug-target Docking, Sequence Alignment (SA) using BLAST or other methods. In a recent, preliminary review in the field published in Proteomics in 2008, discussed the use of these methods but only from the point of view of proteins [12]. After that, in Current Proteomics in December 2009, a collection of papers was published that reviewed many of these techniques [13-19]. In other recent issue guest-edited by González-Díaz *et al.* [20] presented a series of papers devoted to QSPR chemoinformatics and bioinformatics techniques; this issue was published in Current Topics in Medicinal Chemistry in 2008 [20-29]. González-Díaz *et al.* also guest-edited [30] an issue focused on graph TIs approach to Drug ADMET processes and Metabolomics, which was published in May 2010 in Current Drug Metabolism [31-39]. In a more recent issue published in 2010, a group of authors discussed many chemoinformatics and bioinformatics techniques in Current Pharmaceutical Design [40-49]. In this work, we review and comment different QSAR and other theoretic studies about Vitamin D and analogous (see Table 1):

2. QSAR STUDY OF SELECTIVE LIGANDS

In this paper [50] Huanxiang L. and col. developed a QSAR model of 87 selective ligands for the thyroid hormone receptor β 1 (TR β 1) using theoretical molecular descriptors to predict the binding affinity of compounds with receptor. The molecular descriptors were calculated by DRAGON software. Six most relevant structural descriptors to the studied activity were selected as the inputs of QSAR model by a robust optimization algorithm Genetic Algorithm (GA).

*Address correspondence to this autor at the Department of Organic Chemistry, University of Vigo, Spain; E-maol: iselapintos@yahoo.es

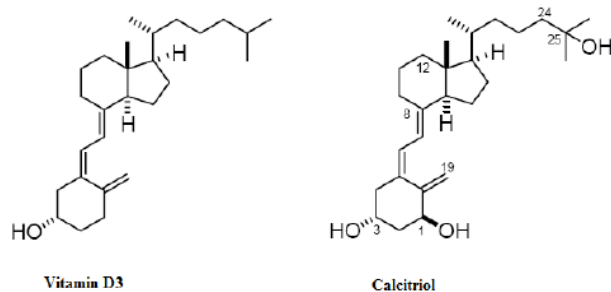


Fig. (1). Structure of Vitamin D3 and Calcitriol.

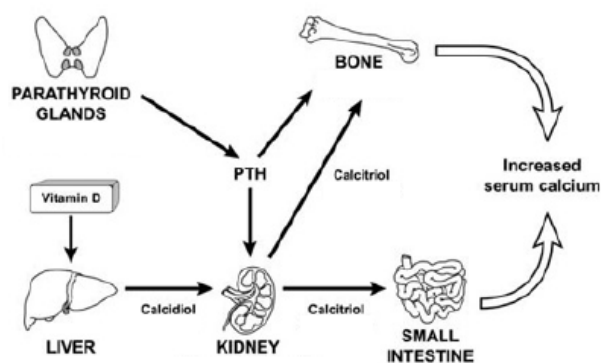


Fig. (2). Vitamin D activity

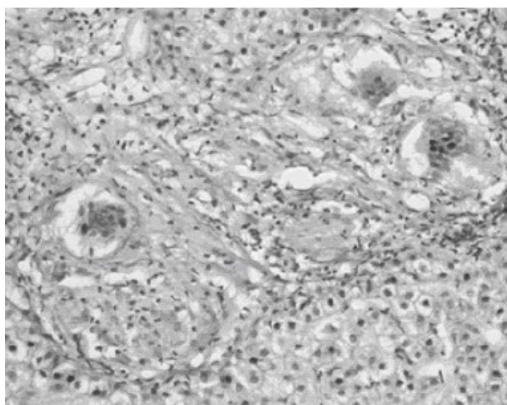


Fig. (3). Unusual case of severe hypercalcemia

The built QSAR model could be used to accurately predict the binding affinity of compounds (in the defined applicability domain) to TR β 1 and a simple correlation algorithm Multiple Linear Regression (MLR) was used to correlate the descriptors and $-\log IC_{50}$ values of the compounds. The built model was fully assessed by various validation methods, including internal and external validation, Y-randomization

test, chemical applicability domain, and all the validations which indicate that the QSAR model was robust and satisfactory. At the same time, the model proposed could also identify and provide some insight into what structural features were related to the biological activity of these compounds and provided some instructions for further designing the new selective ligands for TR β 1 with high activity.

Table 1.

Principal Author	Compounds	Studies	Ref.
Huanxiang L	TR β 1 ligands	QSAR	[50]
González MP	Vitamin D analogs	QSAR	[51]
González MP	Vitamin D analogs	QSAR	[52]
González MP	Vitamin D analogs	QSAR	[53]
Norgaard L	calcitriol analogs	QSAR	[54]
Schuster I	azols	CoMFA	[55]
Hormann RE	ecdysteroids	4D-QSAR CoMFA	[56]
Wurtz JM	dibenzoylhydrazines	Sequence alignment	[57]
Wurtz JM	dibenzoylhydrazines	Docking	[57]

3. QSAR STUDIES OF VITAMIN D RECEPTOR WITH WHIM DESCRIPTORS

In this paper Pérez González M. [51] studied the vitamin D receptor (VDR) affinity of 86 vitamin D analogs by means of the weighted holistic invariant molecular (WHIM) approach. A model being able to describe more than 71% of the variance in the experimental activity was developed with the use of the mentioned approach. In contrast, none of three different approaches, including the use of BCUT, Galvez topological charge indices, and 2D autocorrelations descriptors, was able to explain more than 38% of the variance in the mentioned property, even with more variables in the equation. The statistical information for the best regressions of affinity for VDR with these molecular descriptors showed that the WHIM descriptors explain the experimental variance of the data better than the other approximations. This behavior should be due to the fact that the 3D descriptors are better for modeling this biological property than the 2D descriptors because it is well known that the spatial configuration and the stereochemistry of the ligands are determinant factors for the affinity at the VDR.

4. RADIAL DISTRIBUTION FUNCTION DESCRIPTORS FOR VITAMIN D RECEPTOR

Pérez González M. [52] in his paper described the results of applying the Radial Distribution Function (RDF descriptors) approach for predicting the VDR affinity of 38 vitamin D analogs across QSAR methods, an alternative methodology for the research of new and better Vitamin D analogs with affinity for the VDR receptor. The model described 80% of the experimental variance, with a standard deviation of 0.35. Leave-one-out, bootstrapping and external set validation were carried out with the aim of evaluating the predictive power of the model. The model was able to establish a reliable linear dependence between descriptors just encoding the size and shape of the studied molecules and their VDR affinity. This fact strongly suggested that the main features controlling the VDR affinity are the molecular size and shape of the Vitamin D analogs of the whole molecule or a specific substituent or the interactions among different sub-structure

in the entire molecule. The values of their respective squared correlations coefficients were 0.72, 0.70 and 0.79. The RDF approach was compared with four other predictive models, but none of these could explain more than 71.0% of the variance with six variables in their respective models.

5. RADIAL DISTRIBUTION FUNCTION APPROACH FOR CALCITRIOL ANALOGS

In this article, Pérez González M. [53] applied the Radial Distribution Function (RDF) approach to the study of the chick intestinal VDR affinity of 49 Vitamin D analogs. Every QSAR investigation was based on an assumption of certain homogeneity. This implied similarity in the biological mode of action and measurement of biological activity of all the investigated compounds. In the last years, has been observed in development different protocols and biological test for the measurement of the affinity of Vitamin D analogs for the VDR. The principal tissues used in these protocols were from Bovine Thymus and Chick or Porcine intestinal VDR. A model being able to describe more than 77.5% of the variance in the experimental activity was developed with the use of the mentioned approach. In contrast, none of the four different approaches, including the use of Topological, BCUT, Randic molecular profiles and Geometrical descriptors were able to explain more than 55% of the variance in the mentioned property, with the same number of variables in the equation.

6. QSAR STUDY FOR PREDICTION OF METABOLIC STABILITY OF CALCITRIOL ANALOGS

Norgaard L. and col. in this paper [54] used a data set of 130 calcitriol analogs with their values of *in vitro* metabolic stability and they developed QSAR models. The metabolic stability of a drug is an important property for potential drug candidates. The analogs were encoded with molecular structure descriptors computed mainly with the commercial software QikProp and DiverseSolutions, using a very pragmatic approach to represent conformational diversity. Variable selection was carried out by five different variable selection techniques and Partial Least Squares Regression (PLS) mo-

dels were generated from the 130 analogs. The models were used for prediction of the metabolic stability of 244 virtual calcitriol analogs. Twenty of the 244 analogs were selected and the *in vitro* metabolic stability was determined experimentally. The PLS models were able to predict the correct metabolic stability for 17 of the 20 selected analogs, corresponding to a prediction performance of 85%. The results clearly demonstrate the utility of QSAR models in predicting the *in vitro* metabolic stability of calcitriol analogs. The final models can compete with previously published 3D-QSAR models and the results indicate that QSAR models indeed are useful in predicting the *in vitro* metabolic stability of calcitriol analogs.

7. QSAR STUDIES OF VITAMIN D HYDROXYLASE INHIBITORS

Schuster I. and col. in this article [55] designed some 400 different azole-type inhibitors and examined their capacity to selectively block vitamin D metabolism by CYP24 or synthesis by CYP27B, in human keratinocytes. Aiming at new drugs to efficiently treat diseases, in which either increased or decreased levels of active vitamin D were desirable. They built pharmacophore models of the active sites using commercial a software. The overlay of potent selective compounds indicated similar docking modes in the two-substrate pockets and allowed for identification of bioactive conformations. Superimposing these bioactive conformations with low energy conformers of 25(OH)D₃ suggested that the substrate-mimicked by strong inhibitors in size, shape and lipophilic character-binds to both enzymes in 6*s-trans* configuration. Pharmacophoric models implied a similar geometry of the substrate sites, nevertheless specific features of CYP24 and CYP27B could be defined. Bulky substituents in α -position to the azoles caused selectivity for CYP24, whereas bulky substituents in β -position could result in selectivity for CYP27B. Moreover, studies with small sterically restricted inhibitors revealed a probable location of the 3-OH-group of 25(OH)D₃ in CYP27B. For the sake of brevity, this report did not show further details on structural characteristics, exclusion, and selectivity criteria, which we obtained from our large sets of homogenous inhibition data. It was obvious that these data offered an excellent basis for pharmacophore-modeling and 3D-QSAR studies, especially using CoMFA. Since resulting information was continuously fed into the design of new inhibitors with improved selectivity, the "shape space" covered by the overall ensemble of compounds may be used to give a dynamic image of the active sites.

8. 4D-QSAR AND COMPARISON TO CoMFA MODELING OF ECDYSTEROIDS

Hormann R. E. and col. [56] used, to construct 4D-QSAR models, a training set of 71 ecdysteroids, for which the $-\log(EC_{50})$ potencies in the ecdysteroid-responsive BII cell line were measured. The ecdysteroid-responsive *Drosophila melanogaster* BII cell line was a prototypical homologous inducible gene expression system. Two modestly different alignments were identified ($Q_2 = 0.76-0.80$). These four models were used in consensus modeling to arrive at a three-dimensional pharmacophore. The C-2 and C-22 hydroxyls were identified as hydrogen-bond acceptor sites which enhance activity. A hydrophobic site near C-12 was consistent

with increasing activity. The side-chain substituents at C-17 were predicted to adopt semiextended "active" conformations which could fit into a cylinder-shaped binding pocket lined largely with non polar residues for enhanced activity. A test set of 20 ecdysteroids was used to evaluate the QSAR models. Two 4D-QSAR models for one alignment were identified to be superior to the others based on having the smallest average residuals of prediction for the prediction set (0.69 and 1.13 $-\log[EC_{50}]$ units). The correlation coefficients of the optimum 4D-QSAR models ($R_2 = 0.87$ and 0.88) were nearly the same as those of the best CoMFA model ($R_2 = 0.92$) determined for the same training set. However, the cross-validation correlation coefficient of the CoMFA model was less significant ($Q_2 = 0.59$) than those of the 4D-QSAR models ($Q_2 = 0.80$ and 0.80).

9. SEQUENCE ALIGNMENT OF EcRS WITH hRARg AND hVDR

Wurtz J-M. and col. [57] in this article saw that the EcR-LBD sequences exhibit a good conservation (54% residue identity) that is even higher within the diptera and lepidoptera subgroups (73 and 84%, respectively). The 11 helices (H1 to H12), of the alpha helical sandwich fold of nuclear receptors (NRs), were well identified (Fig. 4). The sequence alignment of the ecdysone receptors (EcRs) revealed strong amino acids conservation, (see Fig. 4), especially in the NR signature region encompassing helices H3 and H4 [58]. Furthermore, key residues in helix H1 were conserved among all ecdysone receptor members, especially Ile283, Leu286, Phe289, and Gln290 (in EcR *Chironomus tentans*, ctEcR). They anchor H1 to the core of the protein and the AH modified common to the retinoid acid (hRARg) (196-AH-197) and hVDR subgroup was replaced by the (FY)Q motif in EcRs (289-FQ290 in ctVDR) [59]. Helix H12 exhibited the typical AF-2AD motif with the conserved Glu508 (in ctEcR). A possibly important salt bridge interaction between this glutamate and Lys357 of H4 may be inferred from RAR and RXR data [60].

10. CONFORMATIONAL ANALYSIS AND DOCKING OF RH5849

In this paper, Wurtz J-M. and col. [57] know the crystal structures of the dibenzoylhydrazines RH5849 and RH5992 [61, 62]. Their most prominent features are the almost orthogonal twists of the central hydrazine bonds and the *cis*-amide bond to the *tert*-butyl group (Fig. 5). To explore their conformational flexibility, the rotational barrier of the central hydrazine bond (Θ) of RH5849 had been investigated by density functional theory. Therefore, two classes of minimum energy structures have been taken into account. The unsubstituted amide nitrogen of RH5849 remains planar, the *tert*-butyl nitrogen, though involved in an amide bond, was slightly pyramidal in each of the conformations investigated. Due to the influence of the NH-proton, the preferred orientation of the phenyl ring was 120 or 60° with respect to the carbonyl group of the *cis*-amide bond, while the other phenyl ring was almost coplanar ($\sim 20^\circ$) with respect to the carbonyl group of the *trans*-amide bond. The phenyl ring could almost be rotated freely when it is not substituted in the *ortho* position. The rotation of its counterpart was hindered to some extent by the effect of the NH-proton.

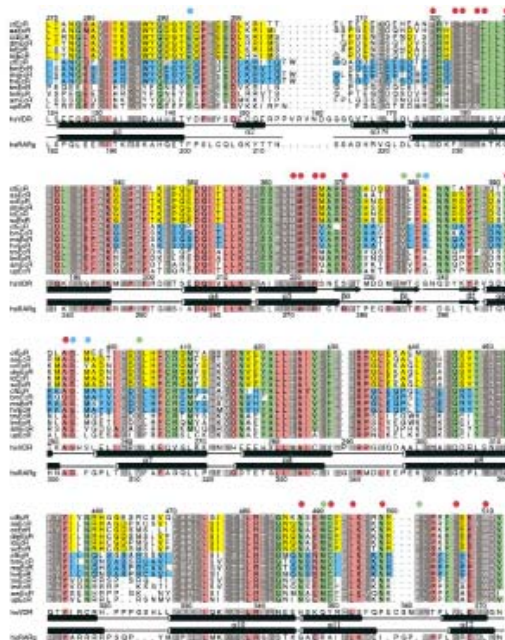


Fig. (4). Sequence alignment analysis of EcRs homology

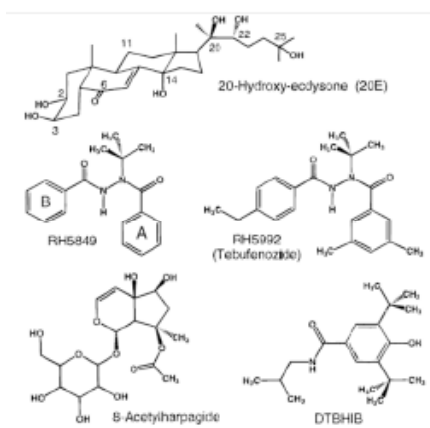


Fig. (5). Structures of ecdysone receptor ligands.

A crystal-like and a “rotated” conformation of RH5849 were then docked manually into the ctEcR-LBD models (Fig. 6). From the overall shape of the binding niches and that of the dibenzoylhydrazines, these were not being able to fully occupy the space in the binding niches. From geometrical reasons there existed only one principal orientation of dibenzoylhydrazines in the LBD. The RH58490 ctEcR complexes were then treated analogously to 20E/ctEcRs. As a result, in both models, the *tert*-butyl group of the dibenzoylhydrazines was located in the similar hydrophobic region occupied only partially by 20E. As for 20E, RH5849

occupied only partially by 20E. As for 20E, RH5849 adopted opposite orientations in both models and globally occupied the same volume. The bulky N-substituent can be favorably accommodated irrespective of which of the two principal conformations of the dibenzoylhydrazines we use. In the EcRra model, the synthetic ligand was tightly packed, due to the smaller size of the binding cavity. In the EcRvd, the ligand had more room to fit. The ligand carbonyl (A-ring) was hydrogen bonded to Asn488 in the EcRvd model but not in the EcRra model.

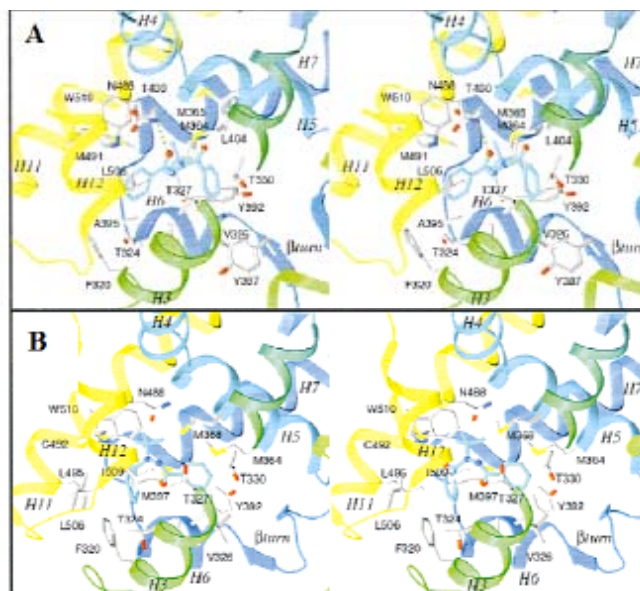


Fig. (6). **A.** Stereo view showing RH5849 EcR-LBD complex based on retinoid acid. **B.** Stereo view showing RH5849 EcR-LBD complex based on vitamin D.

Table 2.

PDB	Author	Year	Classification	Experiment	Resolution	Ref.
3H0A	Wang, Z.	2009	transcription	x-ray	2.10Å	[63]
2GL8	Min, J.R.	2006	Hormone/growth Factor Receptor	x-ray	2.40Å	-
1YNW	Shaffer, P.L.	2005	Transcription/dna	x-ray	3.00Å	[69]
1XAP	Germain, P.	2004	Transcription	x-ray	2.10Å	[70]
1EXA	Klaholz, B.P.	2000	Gene Regulation	x-ray	1.59Å	[71]
1EXX	Klaholz, B.P.	2000	Gene Regulation	x-ray	1.67Å	[71]
3LBD	Klaholz, B.P.	1999	Nuclear Receptor	x-ray	2.40Å	-
4LBD	Klaholz, B.P.	1999	Nuclear Receptor	x-ray	2.50Å	-
2LBD	Renaud, J.-P.	1997	Nuclear Receptor	x-ray	2.06Å	[72]
1HRA	Knegtel, R.M.A.	1994	DNA Binding Receptor	NMR	-	[73]

11. STUDY OF RETINOID ACID RECEPTOR

In this section, we show all structures of retinoid acid receptor found in RCSB PDB in the last years (Table 2). The structure of 3H0A [63] is the crystal structure of peroxisome proliferator-activated receptor gamma (PPARG) and retinoid acid receptor alpha (RXRa) in complex with 9-*cis* retinoid acid, co-activator peptide, and a partial agonist. This structure was studied by X-ray diffraction with a resolution of 2.10Å and it is formed by three polymers and two ligands (Fig. 7A). The structure of 2GL8 is the structure of human

retinoid acid receptor RXR-gamma ligand-binding domain. This structure was studied by X-ray diffraction with a resolution of 2.40Å and it is formed by one polymer (Fig. 7B). The structure of 1YNW [64] is the crystal structure vitamin D receptor and 9-*cis* retinoid acid receptor DNA-binding domain bound to a DR3 response element. This structure was studied by X-ray diffraction with a resolution of 3.00Å and it is formed by four polymers and one ligand. The 1XAP [65] is the structure of the ligand binding domain of the retinoid acid receptor beta. This structure was studied by X-ray diffraction with a resolution of 2.10Å and it is formed by

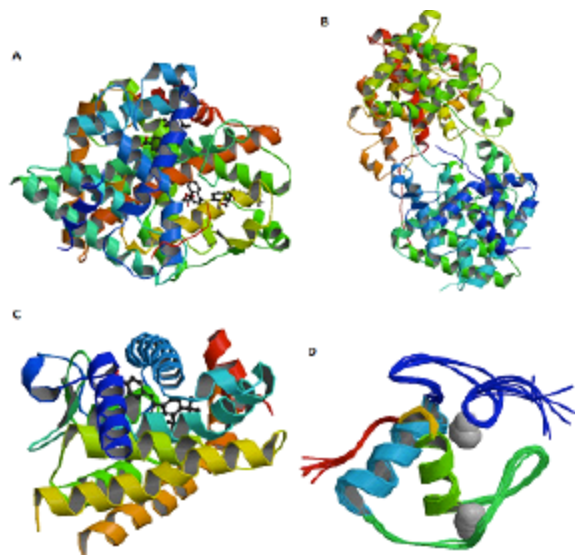


Fig. (7). View of some retinoic acid receptors: A. 3H0A; B. 2GL8; C. 1XAP; D. 1HRA.

one polymer and one ligand (Fig. 7C). The 1EXA is the enantiomer discrimination illustrated by crystal structures of the human retinoid acid receptor hRAR-gamma ligand binding domain; is the complex with the active *R*-enantiomer BMS270394 [66]. It is formed by one polymer and two ligands studied by X-ray diffraction with a resolution of 1.59Å. The 1EXX is the same that 1EXA but with the inactive *S*-enantiomer BMS270395 [66] and with a resolution of 1.67Å. The resolution of X-ray diffraction of the 3LBD is 2.40Å and is formed by one polymer and one ligand and 1EXA is the structure of the ligand binding domain of the human retinoid acid receptor gamma bound to 9-*cis* retinoid acid. The structure of the ligand binding domain of the human retinoid acid receptor gamma bound to the synthetic agonist BMS961 is named as 4LBD and her resolution of the X-ray diffraction is 2.50Å. It is formed by one polymer and one ligand. The 2LBD is the structure of the same receptor but to all-*trans* retinoid acid. Her resolution of the X-ray diffraction is 2.06Å [67]. The 1HRA is the solution structure of the human retinoid acid receptor-beta DNA-binding domain studied by solution NMR [68] (Fig. 7D).

ACKNOWLEDGMENT

We are grateful to the Xunta de Galicia (IN-CITE08PXIB314255PR) for partial financial support.

REFERENCES

- [1] Cernakova M, Kost'aloova D, Kettmann V, Plodova M, Toth J, Drimal J. Potential antimutagenic activity of berberine, a constituent of *Mahonia aquifolium*. *BMC Comp Altern Med* 2002; 2: 2.
- [2] Cao JW, Luo HS, Yu BP, Huang XD, Sheng ZX, Yu JP. Effects of berberine on intracellular free calcium in smooth muscle cells of Guinea pig colon. *Digestion* 2001; 64: 179-83.
- [3] Jang MJ, Jwa M, Kim JH, Song K. Selective inhibition of MAPKK Wis1 in the stress-activated MAPK cascade of *Schizosaccharomyces pombe* by novel berberine derivatives. *J Biol Chem* 2002; 277: 12388-95.
- [4] Li BX, Yang BF, Zhou J, Xu CQ, Li YR. Inhibitory effects of berberine on IK1, IK, and HERG channels of cardiac myocytes. *Acta Pharmacol Sin* 2001; 22: 125-31.
- [5] Choi DS, Kim SJ, Jung MY. Inhibitory activity of berberine on DNA strand cleavage induced by hydrogen peroxide and cytochrome c. *Biosci Biotechnol Biochem* 2001; 65: 452-5.
- [6] Iizuka N, Miyamoto K, Okita K, et al. Inhibitory effect of *Coptidis Rhizoma* and berberine on the proliferation of human esophageal cancer cell lines. *Cancer Lett* 2000; 148: 19-25.
- [7] Li H, Miyahara T, Tezuka Y, et al. The effect of kampo formulae on bone resorption *in vitro* and *in vivo*. II. Detailed study of berberine. *Biol Pharm Bull* 1999; 22: 391-6.
- [8] Chou KC. Structural bioinformatics and its impact to biomedical science. *Curr Med Chem* 2004; 11: 2105-34.
- [9] Chou KC, Wei DQ, Du QS, Sirois S, Zhong WZ. Progress in computational approach to drug development against SARS. *Curr Med Chem* 2006; 13: 3263-70.
- [10] Todeschini R, Consonni V. *Handbook of Molecular Descriptors*. Wiley-VCH: 2002.
- [11] Estrada E., Uriarte E. Recent advances on the role of topological indices in drug discovery research. *Curr Med Chem* 2001; 8: 1573-88.
- [12] González-Díaz H, González-Díaz Y, Santana L, Ubeira FM, Uriarte E. Proteomics, networks and connectivity indices. *Proteomics* 2008; 8: 750-78.
- [13] Torrens F, Castellano G. Topological Charge-Transfer Indices: From Small Molecules to Proteins. *Curr Proteomics* 2009; 204-13.
- [14] Concu R, Dea-Ayuela MA, Perez-Montoto LG, et al. 3D entropy and moments prediction of enzyme classes and experimental-theoretic study of peptide fingerprints in *Leishmania* parasites. *Biochim Biophys Acta* 2009; 1794: 1784-94.
- [15] Ivanciuc O. Machine learning Quantitative Structure-Activity Relationships (QSAR) for peptides binding to Human Amphiphysin-1 SH3 domain. *Curr Proteomics* 2009; 4: 289-302.
- [16] Vilar S, Gonzalez-Diaz H, Santana L, Uriarte E. A network-QSAR model for prediction of genetic-component biomarkers in human colorectal cancer. *J Theor Biol* 2009; 261: 449-58.
- [17] Giuliani A, Di Paola L, Setola R. Proteins as Networks: A Mesoscopic Approach Using Haemoglobin Molecule as Case Study. *Curr Proteomics* 2009; 6: 235-45.

- [18] Chou KC. Pseudo amino acid composition and its applications in bioinformatics, proteomics and system biology. *Curr Proteomics* 2009; 6: 262-74.
- [19] Chen J, Shen B. Computational Analysis of Amino Acid Mutation: a Proteome Wide Perspective. *Curr Proteomics* 2009; 6: 228-34.
- [20] Gonzalez-Diaz H. Quantitative studies on Structure-Activity and Structure-Property Relationships (QSAR/QSPR). *Curr Top Med Chem* 2008; 8: 1554.
- [21] Caballero J, Fernandez M. Artificial neural networks from MATLAB in medicinal chemistry. Bayesian-regularized genetic neural networks (BRGNN): application to the prediction of the antagonistic activity against human platelet thrombin receptor (PAR-1). *Curr Top Med Chem* 2008; 8: 1580-605.
- [22] Duardo-Sanchez A, Patlewicz G, Lopez-Diaz A. Current topics on software use in medicinal chemistry: intellectual property, taxes, and regulatory issues. *Curr Top Med Chem* 2008; 8: 1666-75.
- [23] Gonzalez MP, Teran C, Saiz-Urra L, Teixeira M. Variable selection methods in QSAR: an overview. *Curr Top Med Chem* 2008; 8: 1606-27.
- [24] Gonzalez-Diaz, H, Prado-Prado F, Ubeira FM. Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach. *Curr Top Med Chem* 2008; 8: 1676-90.
- [25] Helguera AM, Combes RD, Gonzalez MP, Cordeiro MN. Applications of 2D descriptors in drug design: a DRAGON tale. *Curr Top Med Chem* 2008; 8: 1628-55.
- [26] Ivanciu O. Weka machine learning for predicting the phospholipidosis inducing potential. *Curr Top Med Chem* 2008; 8: 1691-709.
- [27] Vilar S, Cozza G, Moro S. Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. *Curr Top Med Chem* 2008; 8: 1555-72.
- [28] Wang JF, Wei DQ, Chou KC. Drug candidates from traditional chinese medicines. *Curr Top Med Chem* 2008; 8: 1656-65.
- [29] Wang JF, Wei DQ, Chou KC. Pharmacogenomics and personalized use of drugs. *Curr Top Med Chem* 2008; 8: 1573-9.
- [30] Lin YY, Qi Y, Lu JY, *et al.* A comprehensive synthetic genetic interaction network governing yeast histone acetylation and deacetylation. *Genes & Dev* 2008; 22: 2062-74.
- [31] Zhong WZ, Zhan J, Kang P, Yamazaki S. Gender specific drug metabolism of PF-02341066 in rats-role of sulfoconjugation. *Curr Drug Metab* 2010; 11: 296-306.
- [32] Wang JF, Chou KC. Molecular modeling of cytochrome P450 and drug metabolism. *Curr Drug Metab* 2010; 11: 342-6.
- [33] Mrabet Y, Semmar N. Mathematical methods to analysis of topology, functional variability and evolution of metabolic systems based on different decomposition concepts. *Curr Drug Metab* 2010; 11: 315-41.
- [34] Martinez-Romero M, Vazquez-Naya JM, Rabunal JR, *et al.* Artificial intelligence techniques for colorectal cancer drug metabolism: ontology and complex network. *Curr Drug Metab* 2010; 11: 347-68.
- [35] Khan MT. Predictions of the ADMET properties of candidate drug molecules utilizing different QSAR/QSPR modelling approaches. *Curr Drug Metab* 2010; 11: 285-95.
- [36] Gonzalez-Diaz H, Duardo-Sanchez A, Ubeira FM, *et al.* Review of MARCH-INSIDE & complex networks prediction of drugs: ADMET, anti-parasite activity, metabolizing enzymes and cardiotoxicity proteome biomarkers. *Curr Drug Metab* 2010; 11: 379-406.
- [37] Gonzalez-Diaz H. Network topological indices, drug metabolism, and distribution. *Curr Drug Metab* 2010; 11: 283-4.
- [38] Garcia I, Diop YF, Gomez G. QSAR & complex network study of the HMGR inhibitors structural diversity. *Curr Drug Metab* 2010; 11: 307-14.
- [39] Chou KC. Graphic rule for drug metabolism systems. *Curr Drug Metab* 2010; 11: 369-78.
- [40] Concu R, Podda G, Ubeira FM, Gonzalez-Diaz H. Review of QSAR models for enzyme classes of drug targets: Theoretical background and applications in parasites, hosts, and other organisms. *Curr Pharm Design* 2010; 16: 2710-23.
- [41] Estrada E, Molina E, Nodarse D, Uriarte E. Structural contributions of substrates to their binding to P-Glycoprotein. A TOPS-MODE approach. *Curr Pharm Design* 2010; 16: 2676-709.
- [42] Garcia I, Fall Y, Gomez G. QSAR, docking, and CoMFA studies of GSK3 inhibitors. *Curr Pharm Design* 2010; 16: 2666-75.
- [43] Gonzalez-Diaz H. QSAR and complex networks in pharmaceutical design, microbiology, parasitology, toxicology, cancer, and neurosciences. *Curr Pharm Design* 2010; 16: 2598-600.
- [44] Gonzalez-Diaz H, Romaris F, Duardo-Sanchez A, *et al.* Predicting drugs and proteins in parasite infections with topological indices of complex networks: theoretical backgrounds, applications, and legal issues. *Curr Pharm Design* 2010; 16: 2737-64.
- [45] Marrero-Ponce Y, Casanola-Martin GM, Khan MT, Torrens F, Rescigno A, Abad C. Ligand-based computer-aided discovery of tyrosinase inhibitors. Applications of the TOMOCOMD-CARDD method to the elucidation of new compounds. *Curr Pharm Design* 2010; 16: 2601-24.
- [46] Munteanu CR, Fernandez-Blanco E, Seoane JA, *et al.* Drug discovery and design for complex diseases through QSAR computational methods. *Curr Pharm Design* 2010; 16: 2640-55.
- [47] Roy K, Ghosh G. Exploring QSARs with Extended Topochemical Atom (ETA) indices for modeling chemical and drug toxicity. *Curr Pharm Design* 2010; 16: 2625-39.
- [48] Speck-Planche A, Scotti MT, de Paulo-Emerenciano V. Current pharmaceutical design of antituberculosis drugs: future perspectives. *Curr Pharm Design* 2010; 16: 2656-65.
- [49] Vazquez-Naya JM, Martinez-Romero M, Porto-Pazos AB, *et al.* Ontologies of drug discovery and design for neurology, cardiology and oncology. *Curr Pharm Design* 2010; 16: 2724-36.
- [50] Tan YL, Goh D, Ong ES. Investigation of differentially expressed proteins due to the inhibitory effects of berberine in human liver cancer cell line HepG2. *Mol Biosyst* 2006; 2: 250-8.
- [51] Xu L, Liu Y, He X. Inhibitory effects of berberine on the activation and cell cycle progression of human peripheral lymphocytes. *Cell Mol Immunol* 2005; 2: 295-300.
- [52] Gonzalez MP, Gandara Z, Fall Y, Gomez G. Radial Distribution Function descriptors for predicting affinity for vitamin D receptor. *Eur J Med Chem* 2007.
- [53] Gonzalez MP, Puente M, Fall Y, Gomez G. In silico studies using Radial Distribution Function approach for predicting affinity of 1 alpha,25-dihydroxyvitamin D(3) analogues for Vitamin D receptor. *Steroids* 2006; 71: 510-27.
- [54] Jensen BF, Sorensen MD, Kissmeyer AM, *et al.* Prediction of *in vitro* metabolic stability of calcitriol analogs by QSAR. *J Comput Aided Mol Des* 2003; 17: 849-59.
- [55] Wang F, Zhou HY, Zhao G, *et al.* Inhibitory effects of berberine on ion channels of rat hepatocytes. *World J Gastroenterol* 2004; 10: 2842-5.
- [56] Ravi M, Hopfinger AJ, Hormann RE, Dinan L. 4D-QSAR analysis of a set of ecdysteroids and a comparison to CoMFA modeling. *J Chem Inf Comput Sci* 2001; 41: 1587-604.
- [57] Cernakova M, Kostalova D. Antimicrobial activity of berberine—a constituent of Mahonia aquifolium. *Folia Microbiologica* 2002; 47: 375-8.
- [58] Fink H. [On the problem of the minimal inhibitory concentration (MIC) of oxacillin against staphylococci]. *Arzneimittel-Forschung* 1965; 15: 630-2.
- [59] Tamura M, Takano S. [Influence of pH of media on the minimal inhibitory concentration of cycloserine to Mycobacterium tuberculosis]. *Kekkaku* 1965; 40: 213-8.
- [60] Miura T. Morphological observations of the influence of some chemical drugs on the growth of dermatophytes at concentrations directly below the minimal inhibitory concentration. *Tohoku J Exp Med* 1963; 80: 103-17.
- [61] Linsner H. [The mechanism of action of growth and inhibitory substances. IV. The concentration-activity curves of various synthetic cell-elongating growth substances in the presence of various quantities of synthetic inhibitors.]. *Biochim et Biophys Acta* 1954; 15: 25-30.
- [62] Gero E. [Inhibitory action of vitamin B1 on the oxidation of L-ascorbic acid. II. Influence of pH and of the vitamin B1 concentration; role of the various moieties of the vitamin B1 molecule.]. *Bull de la Soc de Chimie Biol* 1954; 36: 1335-42.
- [63] Raska, S.B. The Metabolism of the Kidney in Experimental Renal Hypertension: II. The Concentration of Cytochrome C and the Activities of the Cytochrome Oxidase and of the Succinic Dehydrogenase Systems in the Kidney of Dogs with Experimental Renal Hypertension. The Inhibitory Effect of Renin and of Kidney Tissue Preparations from Hypertensive Dogs on the Respiratory Enzymes. *J Exp Med* 1945; 82: 227-40.
- [64] Shaffer PL, Gewirth DT. Structural analysis of RXR-VDR interactions on DR3 DNA. *J Steroid Biochem Mol Biol* 2004; 89-90: 215-9.

- [65] Germain P, Kammerer S, Peluso-Iltis C, *et al.* Rational design of RAR-selective ligands revealed by RARbeta crystal structure. *Embo Rep* 2004; 5: 877-82.
- [66] Klaholz BP, Mitschler A, Belema M, Zusi C, Moras D. Enantiomer discrimination illustrated by high-resolution crystal structures of the human nuclear receptor hRARgamma. *Proc Natl Acad Sci USA* 2000; 97: 6322-7.
- [67] Renaud J.-P, Rochel N, Ruff M, Moras D. Crystal structure of the RAR-gamma ligand-binding domain bound to all-trans retinoic acid. *Nature* 1995; 378: 681-9.
- [68] Renaud J.-P, Rochel N, Ruff M, Moras D. The solution structure of the human retinoic acid receptor-beta DNA-binding domain. *J Biomol NMR* 1995; 3: 1-17.
- [69] Kong W, Li Z, Xiao X, Zhao Y. Microcalorimetric investigation of the toxic action of berberine on *Tetrahymena thermophila* BF(5). *J Basic Microbiol* 2010.
- [70] Zhang S, Zhang B, Xing K, Zhang X, Tian X, Dai W. Inhibitory effects of golden thread (*Coptis chinensis*) and berberine on *Microcystis aeruginosa*. *Water Sci Technol* 2010; 61: 763-9.
- [71] Liu L, Yu YL, Yang JS, *et al.* Berberine suppresses intestinal disaccharidases with beneficial metabolic effects in diabetic states, evidences from *in vivo* and *in vitro* study. *Naunyn-Schmiedeberg Arch Pharm* 2010; 381: 371-81.
- [72] Yu Y, Liu L, Wang X, Liu X, Xie L, Wang G. Modulation of glucagon-like peptide-1 release by berberine: *in vivo* and *in vitro* studies. *Biochem Pharmacol* 2010; 79: 1000-6.
- [73] Yang HZ, Zhou MM, Zhao AH, Xing SN, Fan ZQ, Jia W. [Study on effects of baicalin, berberine and Astragalus polysaccharides and their combinative effects on aldose reductase *in vitro*]. *Zhong Yao Cai* 2009; 32: 1259-61.

Review of Theoretical Studies for Prediction neurodegenerative inhibitors.

Francisco Prado-Prado^{1*}, Isela García²

¹ Department of Organic Chemistry, University of Santiago de Compostela, Spain.

² Department of Organic Chemistry, University of Vigo, Spain.

Abstract: Alzheimer's disease (AD) is characterized with several pathologies this disease, amyloid plaques, composed of the β -amyloid peptide and γ -amyloid peptide are hallmark neuropathological lesions in Alzheimer's disease brain. Indeed, a wealth of evidence suggests that β -amyloid is central to the pathophysiology of AD and is likely to play an early role in this intractable neurodegenerative disorder. AD is the most prevalent form of dementia, and current indications show that twenty-nine million people live with AD worldwide, a figure expected to rise exponentially over the coming decades. Clearly, blocking disease progression or, in the best-case scenario, preventing AD altogether would be of benefit in both social and economic terms. However, current AD therapies are merely palliative and only temporarily slow cognitive decline, and treatments that address the underlying pathologic mechanisms of AD are completely lacking. While familial AD (FAD) is caused by autosomal dominant mutations in either amyloid precursor protein (APP) or the presenilin (PS1, PS2) genes. First, we revised 2D QSAR, 3D QSAR, CoMFA, CoMSIA and Docking of β and γ -secretase inhibitors. Next, we review 2D QSAR, 3D QSAR, CoMFA, CoMSIA and Docking for GSK-3 α and GSK-3 β with different compound to find out the structural requirements.

Keywords: QSAR; CoMSIA; CoMFA; Docking; β -secretase inhibitors; γ -secretase inhibitors; Alzheimer's disease (AD).

Introduction

Pathologically, Alzheimer's disease (AD) is characterized by the accumulation of amyloid beta peptide ($A\beta$), as fibrillar plaques and soluble oligomers in high-order association brain regions. The presence of intracellular neurofibrillary tangles, neuroinflammation, neuronal dysfunction and death further characterizes this disease. Mounting evidence suggests that $A\beta$ plays a critical early role in AD pathogenesis, and the basic tenant of the amyloid (or $A\beta$ cascade) hypothesis is that $A\beta$ aggregates trigger a complex pathological cascade which leads to neurodegeneration [1]. A strong genetic correlation exists between FAD and the 42 amino acid $A\beta$ form ($A\beta_{42}$; reviewed in [2-4]). $A\beta$ is derived from APP and mutations in APP and PS increase $A\beta_{42}$ production and cause FAD with nearly 100% penetrance. Down's syndrome (DS) patients, who have an extra copy of the APP gene on chromosome 21, and FAD families with a duplicated APP gene locus [5], exhibit total $A\beta$ overproduction and all develop early-onset AD. In FAD, the $A\beta_{42}$ increase is present years before AD symptoms arise, suggesting that $A\beta_{42}$ is likely to initiate AD pathophysiology. The robust association of $A\beta_{42}$ overproduction with FAD argues strongly in favor of a critical role for $A\beta_{42}$ in the etiology of AD, including in SAD. Fibrillar and oligomeric forms of $A\beta$ appear neurotoxic in vitro and in vivo. Importantly, in specific transgenic (Tg) mouse models of AD the lack of

$A\beta$ correlates with the absence of neuronal loss and improved cognitive function [6-8]. Such data provides direct evidence for the amyloid hypothesis in vivo, and also indicates that $A\beta$ is directly responsible for neuronal death. Consequently, strategies to lower $A\beta_{42}$ levels in the brain are anticipated to be of therapeutic benefit in AD.

$A\beta$ peptide is generated following the sequential cleavage of APP by β - and γ -secretase in the amyloidogenic pathway [9, 10]. $A\beta$ genesis may be precluded if APP is cleaved by α -secretase within the $A\beta$ domain in the non-amyloidogenic pathway **Figure 1**. Recently, the secretases have been identified and the β -secretase is known to be β -site APP cleaving enzyme 1 (BACE1) [11-14], a novel aspartyl protease. BACE1 cleavage of APP is a prerequisite for $A\beta$ formation. $A\beta$ genesis is initiated by BACE1 cleavage of APP at the Asp+1 residue of the $A\beta$ sequence to form the N-terminus of the peptide. This scission liberates two cleavage fragments: a secreted APP ectodomain, APPs β and a membrane-bound carboxyl terminal fragment (CTF). In many instances, an increase in non-amyloidogenic APP metabolism is coupled to a reciprocal decrease in the amyloidogenic processing pathway, and vice-versa, as the α - and β -secretase moieties compete for APP substrate [10, 13]. In the case of γ -secretase is a multi-subunit protease complex, itself an integral membrane protein, those cleaves single-pass transmembrane proteins at residues within the

*Corresponding author: Prado-Prado, Francisco: francisco.prado@usc.es.

transmembrane domain. The most well-known substrate of gamma secretase is amyloid precursor protein, a large integral membrane protein that, when cleaved by both gamma and beta secretase, produces a short 39-42 amino acid peptide called amyloid beta whose abnormally folded fibrillar form is the primary component of amyloid plaques found in the brains of Alzheimer's disease patients. Gamma secretase is also critical in the related processing of the Notch protein [15, 16].

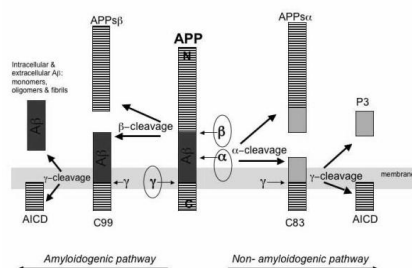


Figure 1. APP metabolism by the secretase enzymes.

Given that both secretase are the initiating enzyme in A β generation, and putatively rate-limiting, it is considered a prime drug target for lowering cerebral A β levels in the treatment and/or prevention of AD. Prior to its identification, numerous studies were undertaken to define the characteristics of β -secretase activity. Although the majority of body tissues exhibit β -secretase activity [17], highest activity levels were observed in neural tissue and neuronal cell lines [18]. Indeed, β -secretase appeared to predominate in neurons, with the level of β -secretase activity appearing lower in astrocytes [19]. Data showing that β -secretase efficiently cleaved only membrane-bound substrates [20] indicated that the enzyme was likely membrane-bound or closely associated with a [21] prevent the buildup of beta-amyloid and may help slow or stop the disease. However, current AD therapies are merely palliative and only temporarily slow cognitive decline, and treatments that address the underlying pathologic mechanisms of AD are completely lacking.

In the last years, a number of publications have appeared suggesting GSK-3 as a target for the treatment of AD. Two isoforms of GSK-3 exists, GSK-3 α and GSK-3 β , both share a high homology at their catalytic site but the α form possess an extended N-terminus with respect to the β form [22, 23]. The phosphorylation of proteins by GSK-3 is an important link in neural function [24-26]. Two are the characteristic neuropathological hallmarks of AD, Neurofibrillary Tangles (NFT's) and increase production of amyloid beta (A β) peptides,

where NFT's are composed of highly phosphorylated form of the microtubule associated protein tau [27] and studies have shown that GSK-3 is one of the main in vivo players of phosphorylation of tau protein [28]. It has been reported that Lithium, a GSK-3 inhibitor, block production of A β peptides by interfering with APP cleavage at γ -secretase step, where the target for Lithium is GSK-3 α [21, 22]. Phiel *et al.* [21] showed that selective reduction in concentration of the α isoform led to a decrease in the concentration of A β 40 and A β 42, primary constituents of amyloid plaques in AD. Thus inhibition of GSK-3 α could potentially provide dual therapy against AD, preventing the buildup of amyloid plaques and of neurofibrillary tangles [21, 29, 30].

GSK-3 β is a serine/threonine kinase and is thought to be a key factor for aberrant tau phosphorylation [31]. Activated GSK-3 β coexists with progression of NFT's and neurodegeneration in the AD brain [32-34]. A conditional GSK-3 β overexpressing transgenic mouse exhibits persistent tau hyperphosphorylation, pretangle-like somatodendritic localization of tau, neuronal death in hippocampus and cognitive deficits [35, 36]. These studies suggest that GSK-3 β is associated with AD progression, and GSK-3 β inhibition is expected to be a promising therapeutic approach for AD.

In this sense, quantitative structure-activity relationships (QSAR) could play an important role in studying these β and γ -secretase inhibitors. QSAR models are necessary in order to guide the β and γ -secretase inhibitors.

On the other hand, QSAR models can be used to explore the relationships between the structural spaces of compounds as inhibitors for specific enzymes, such as MAO inhibitors [37], HIV-1 integrase inhibitors [38], and/or protease inhibitors [39] or tyrosinase inhibitors [40-42]. In fact, Almost all QSAR techniques are based on the use of molecular descriptors, which are numerical series that codify useful chemical information and enable correlations between statistical and biological properties [43, 44]. Recently, the field has moved from small molecules to proteins and other systems. For instance, González-Díaz *et al.* discussed the use of these methods but only from the point of view of proteins [45]. Later, some groups published different papers in one special issue on QSAR but also restricted to the field of protein and proteomics [46-52]. In other recent issue, guest-edited by González-Díaz [53] appeared a series of papers devoted to QSAR/QSPR techniques for low-molecular-weight drugs [53-62]. Most recently, Prado-Prado *et al.* [63] published a mt-QSAR for anti-parasitic drugs. This year was

published other issue [64] focused on QSAR/QSPR models and graph theory used to approach Drug ADMET processes and Metabolomics [65-72]. Last, one of the most recent issues published is devoted to discuss the applications of QSAR in Pharmaceutical Design [73-82]. In the present work, we firstly revised the state-of-art on the design, synthesis, and biological assay of β and γ -secretase inhibitors. Next, we review previous works based on 2D-QSAR, 3D-QSAR, CoMFA, CoMSIA and Docking techniques, which studied different compounds to find out the structural requirements. The topics reviewed, discussed, and/or reported in this paper are:

- 1. Studies of γ -secretase inhibitors**
 - 1.1 Synthesis and Theoretical studies of γ -secretase inhibitors**
 - 1.2 Design and synthesis of pyridine derivatives as BACE-1 inhibitors**
 - 1.3 Discover non-peptide inhibitors of BACE-1 using VHTS**
 - 1.4 Distinct Pharmacological Effects of Inhibitors of γ -Secretase**
 - 1.5 3D-QSAR studies of γ -secretase inhibitors**
 - 1.6 MD simulations of A β fibril interactions**
- 2. Studies of β -secretase inhibitors**
 - 2.1 Synthesis, Theoretical studies and Biological Assay of β -secretase inhibitors**
 - 2.2 Models of novel pyridinium-based potent β -secretase inhibitory leads**
 - 2.3 CoMFA & CoMSIA of hydroxyethylamine derivatives as BACE-1 inhibitors**
 - 2.4 Virtual Screening and Protonation States at Asp32 and Asp228**
 - 2.5 Docking scoring function based on 2D-descriptors**
 - 2.6 Induced-Fit Docking of Peptidic and Pseudo-peptidic BACE-1 inhibitors**
- 3. Studies of GSK-3 α inhibitors**
 - 3.1. 2D-QSAR for 3-anilino-4-phenylmaleimides**
 - 3.2. 3D-QSAR and docking of 3-anilino-4-phenylmaleimides**
- 3.3. QSAR studies of Some GSK-3 α Inhibitory pyrimidines**
- 4. Studies of GSK-3 β inhibitors**
 - 4.1. Design, synthesis and SAR of oxadiazole derivatives**

4.2. Linear/Nonlinear Regression Methods for Prediction of Glycogen

4.3. Molecular modeling, docking and 3D-QSAR studies for maleimides

4.4. Molecular docking and biological testing of inhibitors of GSK-3 β

4.5. 3D-QSAR Modelling of Paullones

4.6. Modeling of Binding Mode of Benzo[e]isoindole-1,3-diones

Discussion

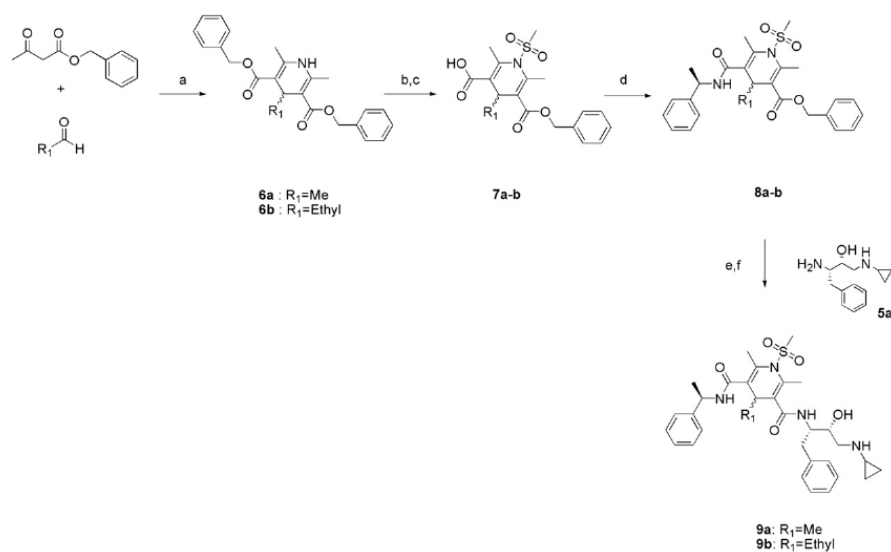
QSAR and Theoretical studies for neurodegene inhibitors

In this section we updated the contents presented in our recent review published in Current Drugs Metabolism [83]. The high number of possible candidates to β -secretase inhibitors creates the necessity of Quantitative Structure-Activity Relationship models in order to guide the β -secretase inhibitor synthesis. In this work, we revised different computational studies for a very large and heterogeneous series of β -secretase. First, we revised QSAR studies with conceptual parameters. Next, using method of regression analysis; and QSAR studies in order to understand the essential structural requirement for binding with receptor. Next, we review 3D QSAR, CoMFA and CoMSIA with different compound to find out the structural requirements for β -secretase inhibitors.

1. Studies of γ -secretase inhibitors

1.1. Design and synthesis of pyridine derivatives as BACE-1 inhibitors.

Soo-Jeong Choi, *et al.* [84], had designed and synthesized of 1,4-dihydropyridine derivatives as BACE-1 inhibitors using a 1,4-dihydropyridine (DHP) scaffold. They had synthesized new inhibitors of BACE-1 (the protein that has been shown to be an attractive therapeutic target in Alzheimer's disease) by modifying the known BACE inhibitor 2 containing a hydroxyethylamine (HEA) motif, see **Figure 2**. Using structure-based drug design based on computer-aided molecular docking, the isophthalamide ring was replaced with a 1,4-dihydropyridine ring as a brain-targeting strategy. After their synthesis, the dihydropyridine derivatives were evaluated their BACE-1-inhibitory activities using a cell-based, reporter gene assay system that measures the cleavage of alkaline phosphatase (AP)-APP fusion protein by BACE-1.



Reagent: (a) NH₄OAc, Ethanol, 90 °C, 24 h, 98%; (b) Methane sulfonylchloride, NaH, DMF, 0-60 °C, 4 h, 35%; (c) AlCl₃, Anisole, DCM, -50 °C to RT, 2 h, 21%; (d) R-methylbenzylamine, PyBOP, DIPEA, DCM, 1 h, 70%; (e) AlCl₃, Anisole, DCM, -50 °C to RT, 2 h, 30%; (f) Compound 5a, PyBOP, DIPEA, DCM, 15 min, 65%.

Figure 2. Synthesis of 1-methylsulfonamide-2,6-dimethyl-1,4-dihydropyridine derivatives.

Molecular modeling was performed using CDOCKER, a CHARMM based molecular dynamics docking algorithm Discovery Studio 2.0 (Accelrys). The BACE-1 structure co-crystallized was obtained from the PDBdata bank (PDB code: 2B8L). A protein clean process and a CHARMM-force field were sequentially applied. The area around 2 was chosen as the active site, with the radius set as at 8-Å. After removing 2 from the structure of the complex, a binding sphere in the three axis

directions was constructed around the active site. All default parameters were used in the docking process. CHARMM based molecular dynamics (1000 steps) were used to generate random ligand conformations and the position of any ligand was optimized in the binding site using rigid body rotation followed by simulated annealing at 700 K. Final energy minimization was set as the full potential mode. The final binding conformation was determined on the basis of energy (see **Figure 3**).

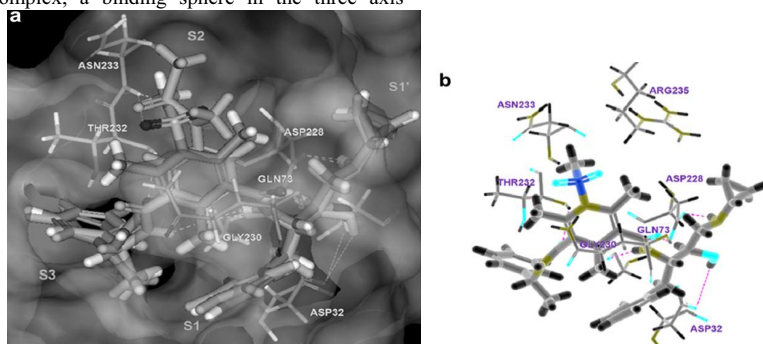


Figure 3. Overlay of inhibitor 2 (green) and inhibitor 9a in the BACE-1 active site, b. Interactions of 9a in the active site of BACE-1- Hydrogen bonds are shown with dotted lines (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Based on molecular docking results, we designed 1,4-DHP derivatives as BACE-1 inhibitors using five strategies. These were

replacement of the bulky *o*-methylbenzamide group in 2 with a benzyl ester or smaller acetyl group for binding in the S3 pocket; modification

of the sulfonamide group in the aromatic scaffold of **2** with alkyl ester or amide groups, maintaining the important hydrogen bonding with Asn233 in the S2 binding pocket; incorporation of additional hydrophobic interactions into the S1 binding pocket by introduction of alkyl or aryl groups, including methyl, ethyl, propyl, isopropyl, and phenyl groups; alteration of the cyclopropyl group at the R4 position to other aromatic groups, thus changing hydrophobic interactions in the S20 binding pocket by extension toward the prime-side of the enzyme; and, alterations at the 2 and 6 positions of the 1,4-DHP scaffold by synthesis of 2-monomethyl and 2,6-unsubstituted analogs.

Their results show that most of the 1,4-DHP analogs showed BACE-1-inhibitory activities with IC_{50} values in the range 8e30 mM, suggesting that the 1,4-DHP skeleton may be utilized to develop brain-targeting BACE-1 inhibitors.

1.2. Discover non-peptide inhibitors of BACE-1 using VHTS

A novel series of isatin-based inhibitors of b-secretase (BACE-1) using a virtual highthroughput screening approach have identified by Yi Moka *et al.* [85]. Structure-activity relationship studies revealed structural features important for inhibition. Docking studies suggest these inhibitors may bind within the BACE-1 active site through H-bonding interactions involving the catalytic aspartate residues.

They used AutoDock to separately dock the two proposed favored conformations 19 and 20 of compound **1** into the active site of BACE-1 (PDB code 1M4H). While the docking of conformer 19 did not give any solutions consistent with the observed biological activity, docking of conformer 20 revealed a binding pose which was consistent with the observed activities of compounds 1-8 (see **Figure 4**).

This figure shows the lowest-energy binding pose identified for conformer 20 within BACE-1. It is noteworthy that an analogous binding pose was also identified using eHiTS. The acetamide moiety is predicted to occupy the catalytic site, with the acetamide N-H acting as an H-bond donor to the catalytic residue Asp228 (H-bond length = 1.86 Å 0).

The phenol unit is predicted to make an H-bond contact with the backbone nitrogen of Thr232 (H-bond length = 2.20 Å 0) and to partly occupy the P2 substrate pocket. This feature implies the phenol might be involved in both the formation of the intramolecular H-bonding network, and also in intermolecular H-bonding interactions with the enzyme. The nitro group in **1** is predicted to extend into the P4 pocket, possibly participating in weak H-

bonding with the side chain of Arg307 (H-bond length = 2.13 Å 0), consistent with the slight decrease in the binding affinity exhibited by compound.

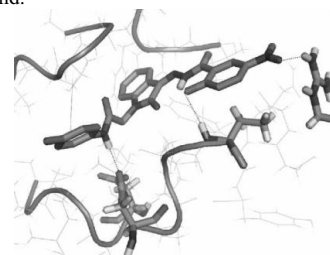


Figure 4. Binding pose of **1** (corresponding to conformer 20) in the BACE-1 active site generated using AutoDock.

In summary, the authors, using the virtual high-throughput screening software eHiTS, have discovered a novel non-peptidic inhibitor of BACE-1 based on an isatin motif. Studies of the biological activity of structural variants in combination with in silico docking suggest the inhibitor adopts a planar conformation, which is stabilized by intramolecular H-bonding from the phenolic moiety. Additionally, binding to BACE-1 appears to involve H-bonding interactions between the p-tolylamide of **1** and the catalytic residue Asp228.

A recent report detailing the discovery of a series of potent small molecule BACE-1 inhibitors compares the ligand efficiency (LE) of a range of reported inhibitors of BACE-1. In this study, the authors noted that despite the high potency of the previously reported peptide-based BACE-1 inhibitors such as OM99-2 (K_i = 1.6 nM), the relatively high molecular weights of these systems (e.g., OM99-2 has a molecular weight of 893) often result in them having relatively poor ligand efficiency (e.g., LE = 0.19 for OM99-2). For this study, although still somewhat below the preferred minimum value of LE = 0.3, compound **1** (molecular weight = 461) has LE = 0.22 and therefore, is closer to the preferred value than the potent but considerably larger peptidic inhibitors reported previously. They have demonstrated that eHiTS is a powerful screening tool to identify biologically active compounds quickly and efficiently.

1.3 Distinct Pharmacological Effects of Inhibitors of γ -Secretase

Toru Sato, *et al.* [86], have report that helical peptide inhibitors designed to mimic SPP substrates and interact with the SPP initial substrate-binding site (the “docking site”) inhibit both SPP and γ -secretase, but with submicromolar potency for SPP. SPP was labeled by helical peptide and transition-state analogue affinity probes but at distinct sites.

Nonsteroidal anti-inflammatory drugs, which shift the site of proteolysis by SPP and γ -secretase, did not affect the labeling of SPP or γ -secretase by the helical peptide or transition-state analogue probes. On the other hand, another class of previously reported γ -secretase modulators, naphthyl ketones, inhibited SPP activity as well as selective proteolysis by γ -secretase. These naphthyl ketones significantly disrupted labeling of SPP by the helical peptide probe but did not block labeling of SPP by the transition-state analogue probe. With respect to γ -secretase, the naphthyl ketone modulators allowed labeling by the transition-state analogue probe but not the helical peptide probe. Thus, the naphthyl ketones appear to alter the docking sites of both SPP and γ -secretase. These results indicate that pharmacological effects of the four different classes of inhibitors (transition-state analogues, helical peptides, nonsteroidal anti-inflammatory drugs, and naphthyl ketones) are distinct from each other, and they reveal similarities and differences with how they affect SPP and γ -secretase (see **Figure 5**).

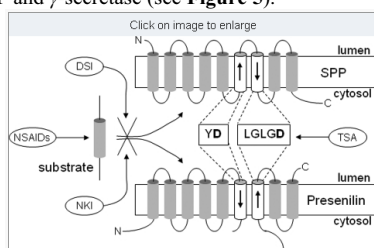


Figure 5. Schematic mechanisms of inhibitor. Transition-state analogue (TSA) inhibitor targets the active site, and helical peptide docking site inhibitor (DSI) prevents initial substrate interaction with the protease. NSAIDs target the substrate, and naphthyl ketone inhibitors (NKI) disrupt the initial interaction between substrate and protease (selectively for APP in the case of γ -secretase).

1.3. 3D-QSAR studies of γ -secretase inhibitors

A 3D-QSAR analysis on a series of 67 benzodiazepine analogues reported as γ -secretase inhibitors using molecular field analysis (MFA), with G/PLS to predict steric and electrostatic molecular field interaction for the activity have performed by Tarnvir Sammi *et al.* [87].

The MFA study was carried out using a training set of 54 compounds. The predictive ability of model developed was assessed using a test set of 13 compounds (r^2_{pred} as high as 0.729). The analyzed MFA model has demonstrated a good fit, having r^2 value of 0.858 and cross validated coefficient, r^2_{cv} value as 0.790. The analysis of the best MFA model provided insight into possible modification of the molecules for better activity.

To obtain effective 3D-QSAR models the method that they used for performing the

alignment was the maximum common subgroup (MCSG). This method looks at molecules as points and lines, and uses the techniques of graph theory to identify patterns. It finds the largest subset of atoms in the shape reference compound that is shared by all the structures in the data set and uses this subset for alignment. A rigid fit of atom pairing was performed to superimpose each structure so that it overlays the shape reference compound. The most active bold-faced portion of molecule 1, was used as the template for the superposition (see **Figure 6**).

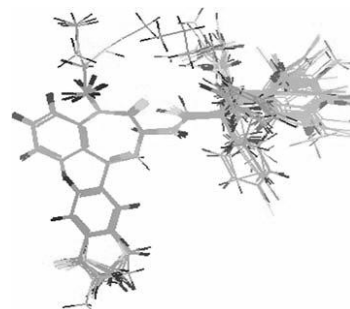


Figure 6. Stereoview of all the aligned molecules.

G/PLS technique available in QSAR environment of Cerius2 software was used to perform regression analysis of data. As there were large numbers of points used as independent variables, genetic partial least squares (G/PLS) were used to derive QSAR models. G/PLS is derived from two QSAR calculation methods: Genetic function approximation (GFA) and partial least squares (PLS). The GFA algorithm approach builds multiple models rather than single model; it automatically selects which features are to be used in model. Further it is better at discovering combinations of features that take advantage of correlations between multiple features. In PLS, variables might be overlooked during interpretation or in designing the next experiment even though cumulatively they are important. It gives a reduced solution, which is statistically more robust than multiple linear regressions (MLA). The linear PLS model finds “new variables” (latent variables or X scores) which are linear combinations of original variables. To avoid over fitting, a strict test for the significance of each consecutive PLS component is necessary and then stopping when the components are non-significant. Cross validation a practical and reliable way of testing this significance. G/PLS combines the best features of GFA and PLS. In GFA; equation models have a randomly chosen proper subset of independent variables. As a result of multiple

linear regressions (MLA) on each model, the best ones become the next generation and two of them produce an offspring. This was repeated 50,000 (default, 5000 times). For other settings, all defaults were used. Application of G/PLS thus allows the construction of large QSAR equations while still avoiding over fitting and eliminating most variables. The best model was

selected on statistical measures such as data points (n), square correlation coefficient (r^2), cross-validated correlation coefficient (r_{cv}^2), predicted correlation coefficient (r_{pred}^2), predicted sum of squares (PRESS), bootstrap correlation coefficient (r_{obs}^2) (see **Table 1**).

Table 1. Various statistical parameters along with their numerical value obtained for the best model

	Parameter	Value
1.	Data points (n)	54
2.	Square of correlation coefficient (r^2) for training set	0.858
3.	Leave one out cross validated correlation coefficient (r_{cv}^2)	0.790
4.	Predicted sum of squares (PRESS)	16.086
5.	Number of PLS components (C)	5
6.	Simple correlation coefficient (r_{pred}^2) for test set	0.729
7.	Predicted correlation coefficient (r_{pred}^2)	0.685
8.	Bootstrap correlation coefficient (r_{bs}^2)	0.843
9.	Least Square error (LSE)	0.208
10.	Predicted root mean square error ($RMSE_{pred}$)	0.579
11.	Slope of regression line of observed vs predicted activity passing through origin(k)	1.004
12.	Slope of regression line of predicted vs observed activity passing through origin(k')	0.987
13.	Correlation coefficient for regression line of observed vs predicted activity passing through origin(R20)	0.999
14.	Correlation coefficient for regression line of predicted vs observed activity passing through origin	0.998

(* Correlation coefficient calculated using Eq.3)

1.4. MD simulations of A β fibril interactions

Neil J. Brucea, *et al.*, use molecular dynamics simulations to compare the model of interaction of an active (LPFFD) and inactive (LHFFD) β -sheet breaker peptide with an A β fibril structure from solid-state NMR studies. This study is based in the matter that have accumulation and aggregation of the 42-residue amyloid- β (A β) protein fragment, which originates from the cleavage of amyloid precursor protein by β and γ secretase, correlates with the pathology of Alzheimer's disease (AD). Possible therapies for AD include peptides based on the A β sequence, and recently identified small molecular weight compounds designed to mimic these, that interfere with the aggregation of A β and prevent its toxic effects on neuronal cells in culture.

Here, they found that LHFFD had a weaker interaction with the fibril than the active peptide, LPFFD, from geometric and energetic considerations, as estimated by the MM/PBSA approach. Cluster analysis and computational alanine scanning identified important ligand-fibril contacts, including a possible difference in the effect of histidine on ligand-fibril π -stacking interactions, and the role of the proline residue in establishing contacts that compete with those essential for maintenance of the inter-monomer β -sheet structure of the fibril. Their results show that molecular dynamics simulations can be a useful way to classify the stability of docking sites. These mechanistic insights into the ability of LPFFD to reverse aggregation of toxic A β will guide the redesign of lead compounds, and aid in developing realistic therapies for AD and other diseases of protein aggregation (see **Figure 7**).

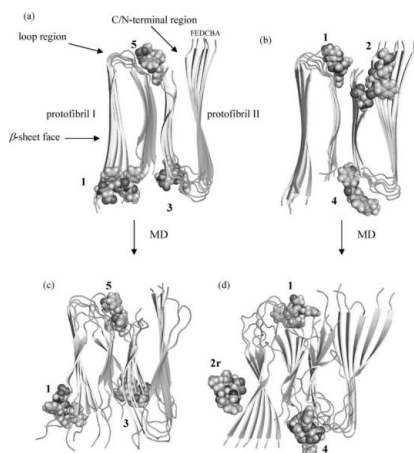


Figure 7. Comparison of docked poses of (a) active peptide LPFFD (b) inactive peptide LHFFD; and comparison of MD-refined poses of (c) active peptide LPFFD and (d) inactive peptide LHFFD.

2. Studies of β -secretase inhibitors

2.1. Models of novel pyridinium-based potent β -secretase inhibitory leads

In one article by Afaf Al-Nadaf *et al.* [88] explore the pharmacophoric space of 129 known BACE inhibitors have potential as anti-Alzheimer's disease treatments. The QSAR analysis employed to select optimal combination of pharmacophoric models and 2D physicochemical descriptors capable of explaining bioactivity variation ($r^2 = 0.88$, $F = 60.48$, $r^2_{\text{LOO}} = 0.85$, r^2_{PRESS} against 25 external test inhibitors = 0.71). They were obliged to use ligand efficiency as the response variable because the logarithmic transformation of bioactivities failed to access self-consistent QSAR models. The authors constructed three pharmacophoric models emerged in the successful QSAR equation suggesting at least three binding modes accessible to ligands within BACE binding pocket. The QSAR equation and pharmacophoric models were validated through ROC curves (see **Table 2**), and were employed to guide synthesis of novel pyridinium-based BACE inhibitors.

Table 2. ROC_s performances of QSAR-selected pharmacophores as 3D search queries.

Pharmacophore	ROC _s		A			FNR _t
	AUC _s	CC _s	PC _s	PR _s	TPR _s	
Hypo10/10	0.982	0.961	0.988	0.28	0.345	0.011
Hypo6/18	0.981	0.961	0.975	0.06	0.311	0.024
Hypo1/21	0.738	0.961	0.961	0.96	0.898	0.038

^a ROC: receiver operating characteristic, ^b AUC: area under the curve, ^c ACC: overall accuracy,

^d SPC: overall specificity, ^e TPR: overall true positive rate, ^f FNR: overall false negative rate.

2.2. CoMFA & CoMSIA of hydroxyethylamine derivatives as BACE-1 inhibitors

Ashish Pandey *et al.*[89] were developed three-dimensional quantitative structure-activity relationship (3D-QSAR) models based on comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA), on a series of 43 hydroxyethylamine derivatives, acting as potent inhibitors of β -site amyloid precursor protein (APP) cleavage enzyme (BACE-1). They used a crystal structure of the BACE-1 enzyme (PDB ID: 2HM1) with one of the most active compound presented in this paper was available, and we assumed it to be the bioactive conformation of the studied series, for 3D-QSAR analysis. Statistically significant 3D-QSAR model was established on a training set of 34 compounds, which were validated by a test set of 9 compounds. For the best CoMFA model, the statistics are, $r^2 = 0.998$, $r^2_{\text{cv}} = 0.810$, $n = 34$ for the training set and $r^2_{\text{pred}} = 0.934$, $n = 9$ for the test set. For the best CoMSIA model (combined steric, electrostatic, hydrophobic, and hydrogen bond donor fields), the statistics are $r^2 = 0.978$, $r^2_{\text{cv}} = 0.754$, $n = 34$ for the training set and $r^2_{\text{pred}} = 0.750$, $n = 9$ for the test set, see **Table 3**. The resulting contour maps, produced by the best CoMFA and CoMSIA models, were used to identify the structural features relevant to the biological activity in series of analogs. The data generated from the present study will further help to design novel, potent, and selective BACE-1 inhibitors.

Table 3. PLS summary of CoMFA and CoMSIA results.

Statistical parameters	results.	
	CoMFA (S E)	CoMSIA (S EHD)
Number of molecules in training set	34	34
Number of molecules in test set	9	9
r^2_{cv}	0.810	0.754
NOC	7	4
SEE	0.063	0.204
r^2	0.998	0.978
F -test	2009.08	324.673
r^2_{bs}	0.999	0.989
SD _{bs}	0.001	0.006
r^2_{pred}	0.934	0.750
Percentage of field contributions		
S	47.4	24.8
E	52.6	34.0
H	–	26.3
D	–	14.9

Abbreviations: S (steric field), E (electrostatic field), H (hydrophobic field), D (hydrogen bond donor field) r^2_{cv} = Cross-validated correlation coefficient by PLS LOO

method, NOC=Optimum number of components as determined by PLS LOO cross-validation study, SEE=Standard error of estimate, r^2 =Conventional correlation coefficient, r_{bs}^2 =Correlation coefficient after 100 runs of bootstrapping, SD_{bs} =Standard deviation from 100 runs of bootstrapping, r_{pred}^2 =Predictive correlation coefficient.

2.3. Virtual Screening and Protonation

States at Asp32 and Asp228

György M. Keseru *et al.*[90] performed a comparative virtual screen for β -secretase (BACE1) inhibitors using different docking methods (FlexX and FlexX-Pharm), scoring functions (Dock, Gold, Chem, PMF, FlexX), protonation states (default and calculated), and protein conformations (apo and ligand bound). Apo and ligand bound conformations of BACE1 were both found to be suitable for virtual screening. Assigning calculated protonation states to catalytic Asp32 and Asp228 residues resulted in significant improvement of enrichment factors as calculated at 1% of the ranked database. The authors used 1FKN to obtain no enrichment by FlexX/D-Score that was improved to ligand when considering calculated protonation states. They also show that combining calculated protonation states with pharmacophore constraints using FlexX-Pharm/D-Score improved enrichment further to ligand. Enrichments reported in this study suggest our screening protocol will be effective in the virtual screening of large compound libraries for BACE1 inhibitors.

2.4 Docking scoring function based on 2D-Descriptors

In this paper Csaba Hetényi [91] showed a key step in the molecular engineering of such potent lead compounds is the prediction of the energetics of their binding to the macromolecular targets. Although sophisticated experimental and in silico methods are available to help this issue, the structure-based calculation of the binding free energies of large, flexible ligands to proteins is problematic. In this study, a fast and accurate calculation strategy is presented; following modification of the scoring function of the popular docking program package AutoDock and the involvement of ligand based two-dimensional descriptors. Quantitative structure-activity relationships with good predictive power were developed. The best results of this paper were shown in **Table 4**. Thorough cross-validation tests and verifications were performed on the basis of experimental binding data of biologically important systems. The capabilities and limitations of the ligand based descriptors were analyzed. According to the authors the application of these results in the early phase of lead design will contribute to precise predictions, correct selections, and consequently a higher success rate of rational drug discovery.

Table 4. The results produced by the best CoMFA and CoMSIA models.

QSAR	descriptor (Di)						
	coefficient (Ri)	error of coeff.	t-value	R ²	R _{cv} ²	R _s ²	F-value
A	gGTH	3.1216 × 10 ⁻¹	2.4686 × 10 ⁻²	0.799	0.774	0.105	93.36
	RPCGEN	3.2582 × 10 ¹	6.9963				
	constant	-4.1980	6.9930 × 10 ⁻¹				
B	gGTH	2.7077 × 10 ⁻¹	2.2926 × 10 ⁻²	0.859	0.838	0.076	93.17
	RPCGEN	5.7129 × 10 ¹	8.1307				
	J	-6.2410 × 10 ⁻¹	1.4148 × 10 ⁻¹				
	constant	-4.6864	6.0281 × 10 ⁻¹				

Standard deviations (s^2), squares of the correlation coefficients (R^2), and leave-one-out cross-validated correlation coefficients (R_{cv}^2) of the regressions are tabulated.

2.5 Induced-Fit Docking of Peptidic and Pseudo-peptidic BACE-1 inhibitors

Inhibition of β -secretase (BACE 1) has recently been investigated as a promising therapeutic approach in the treatment of Alzheimer's disease, and a growing number of BACE 1 inhibitors and crystal structures of BACE 1/inhibitors complexes have been reported. Nicolas Moitessier *et al.*[92] report herein a predictive computational method and its application to potential BACE 1 inhibitors. Using a training set of 50 known highly flexible inhibitors, they developed a docking method that accounts for the flexibility of both the protein and the inhibitors. Protein flexibility is accounted for using a specifically designed genetic algorithm. In this paper developed a scoring function consisting

of force field evaluation of the inhibitor/protein interactions and two additional terms for hydrogen bonding and entropy change upon binding. Discarding three outliers from the training set, the protocol was found to perform well with an rmsd of 1.19 kcal/mol and r^2 value of 0.789. Evaluation of the predictive power was carried out by virtual screening of 80 synthetic compounds. The significant enrichment at the top of the ranking list in active compounds demonstrated the ability of the docking and scoring protocol to rank the compounds relative to their activities.

3. Studies of GSK-3 α inhibitors

3.1. 2D-QSAR for 3-anilino-4-phenylmaleimides

In this paper, Sivaprakasam, P. *et al.* [30] reported a 2D-QSAR exploration of the physicochemical (hydrophobic, electronic, and steric) and structural requirements among 3-anilino-4-phenylmaleimides toward GSK-3 α binding. Using Fujita-Ban and Hansch QSAR analysis, electronic and steric interactions at the 4-phenyl ring and hydrophobic interactions at the 3-anilino ring were shown to be crucial. Hansch type QSAR was still widely used in the lead optimization stage of synthetic and other projects.

Fujita-Ban analysis of 3-anilino-4-phenylmaleimides revealed that certain structural features such as Cl, OCH₃, and NO₂ mono substitution at any position around the 4-phenyl ring were favorable for GSK-3 α inhibition. Substituents at the 3-anilino ring such as 3-Cl, 4-Cl, 5-Cl, 3-COOH, 4-OH, and 4-SCH₃ were positively and 3-OH was negatively correlated with GSK-3 α inhibitory activity.

Through Hansch QSAR analyses, they found that the GSK-3 α inhibitory activity was enhanced by: 1. Electron-withdrawing, bulky *ortho* substituents at 4-phenyl ring; 2. 4-chloro substitution around anilino ring; 3. 3-anilino rather than 3-N-methylanilino derivatives; 4. Hydrophobic *meta* substituents on the anilino ring. Overall, QSAR models 13a and 14a suggested electronic and steric effects at the 4-phenyl ring and hydrophobic effects at the 3-anilino or 3-N-methylanilino ring were crucial. Their 2D-model (**Figure 8**) illustrated these effects which are essential for binding of the maleimides to the GSK-3 α enzyme. Their analysis had provided key information regarding ligand–target interactions which they believed will help medicinal chemists to design more potent GSK-3 α inhibitors.

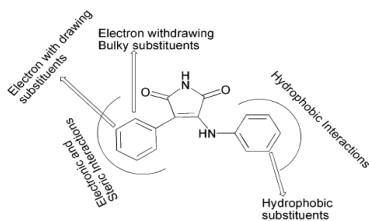


Figure 8. Proposed model based on 2D-QSAR analyses showing the nature of interactions and substitution requirements for effective binding of 3-anilino-4-phenylmaleimides with the GSK-3 α isoform.

3.2. 3D-QSAR and docking of 3-anilino-4-phenylmaleimides

In this article [93] was reported 3D-QSAR analyses using CoMFA and CoMSIA and

molecular docking studies on 3-anilino-4-phenylmaleimides as GSK-3 α inhibitors, in order to better understand the mechanism of action and structure-activity relationship of these compounds. Comparison of the active site residues of GSK-3 α showed that all the key amino acids involved in polar interactions with the maleimides for the β isoform were the same in the α isoform, except that Asp133 in the β isoform was replaced by Glu196 in the α isoform. They prepared a homology model for GSK-3 α , and showed that the change from Asp to Glu should not affect maleimide binding significantly. Our best CoMFA model contained steric and electrostatic fields and had $n = 56$, $q^2 = 0.844$, $r^2 = 0.942$, $SEE = 0.104$, $F = 162.49$ and $r^2_{pred} = 0.779$ for five components. CoMFA electrostatic contours revealed that increased negative charge at the *meta* position of the 4-phenyl ring was favorable for the activity. They found that electron withdrawing groups at the *meta* and *para* positions around the anilino ring were important for enhancing activity. Electron-withdrawing bulky *ortho* substituents on the 4-phenyl ring were conducive to GSK-3 α inhibition. CoMSIA model showed the importance of hydrogen bond donor groups on these ligands for enhanced activity. The best CoMSIA model (S + E + D) had $n = 56$, $q^2 = 0.833$, $r^2 = 0.932$, $SEE = 0.113$, $F = 111.67$ and $r^2_{pred} = 0.803$ for six components. Comparatively, 3-N-methylanilino derivatives were less active than 3-anilino derivatives.

Docking studies revealed the binding poses of three subclasses of these ligands, namely anilino, N-methylanilino and indoline derivatives, within the active site of the β isoform, and helped to explain the difference in their inhibitory activity.

3.3. QSAR studies of Some GSK-3 α Inhibitory pyrimidines

Jamloky, A. *et al.* in this paper [22] studied a series of pyrimidines which was performed to gain structural insight into the binding mode of the molecules to the GSK-3 α . The molecular modeling studies were performed using CS Chem. Office 2001 molecular modeling software version 6.0. MOPAC module was used to minimize the energy and calculate of the descriptors. The thermodynamic and steric features of the pyrimidines were highly correlated with GSK-3 α inhibitory activity. The positive coefficient of PMI-Y in the model suggested the presence of bulky substituents oriented towards Y-axis of the molecule will enhance the GSK-3 α inhibitory activity. The observation supports the hypothesis that the presence of the bulky substituents like bromine with inherent hydrophobic character may involve in nonspecific interaction with the ATP

binding site. The results of the study suggested that introduction of bulky groups at C-5 position of the hydrophobic interaction with the ATP binding site of the enzyme. This may be attributed to the strain exerted by the two adjacent phenyl rings on the planar pyrazolo (3,4-*b*) pyridine ring thereby partly disrupting the hydrogen bonding interaction between nitrogen in the pyrazolo group and the complementary group in the enzyme.

Studies of GSK-3 β inhibitors

4.1. Design, synthesis and structure-activity relationships of 1,3,4-oxadiazole derivatives

Saitoh, M. *et al.* [94] reported design, synthesis and structure-activity relationships of a novel series of oxadiazole derivatives as GSK-3 β inhibitors. Among these inhibitors, compound 20x (see Figure) showed highly selective and potent GSK-3 β inhibitory activity in vitro and its binding mode was determined by obtaining the X-ray co-crystal structure of 20x and GSK-3 β (see Figure 9,10). The hydrogen bonding interaction of the benzimidazole core with the hinge region and the oxadiazole with Asp200 were observed. Additionally, interaction of 4-methoxyphenyl group with Arg141 was observed.

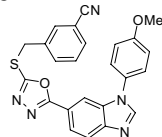


Figure 9. Structure of 20x

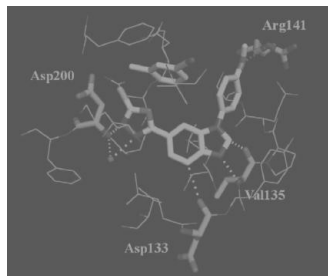


Figure 10. X-ray co-crystal structure of 1 in complex with GSK-3 β

4.2. Linear/Nonlinear Regression Methods for Prediction of Glycogen Synthase Kinase-3 β Inhibitory Activities

Matheus P. Freitas *et al.* [95] realized linear/nonlinear regression methods as multiple linear regression (MLR), artificial neural network (ANN), and support vector machines (SVM) with a series of glycogen synthase kinase-3 β (GSK-3 β) inhibitors using calculated Dragon descriptors. Few variables were selected from a pool of calculated Dragon descriptors through three different feature selection

methods, namely genetic algorithm (GA), successive projections algorithm (SPA), and fuzzy rough set ant colony optimization (fuzzy rough set ACO). The fuzzy rough set ACO/SVM-based model gave the best estimation/prediction results, demonstrating the nonlinear nature of this analysis and suggesting fuzzy rough set ACO, first introduced in chemistry here, as an improved variable selection method in QSAR for the class of GSK-3 β inhibitors. MLR yielded QSAR models only reasonably predictable, with r^2 ranging from 0.77 to 0.81 and r^2_{test} of 0.67 to 0.76, ANN and specially SVM were capable of estimating and predicting biological activities very accurately.

4.3. Molecular modeling, docking and 3D-QSAR studies for maleimides

Ki Hwan Kim *et al.* [96] in this article carried out molecular modeling and docking studies with three-dimensional quantitative structure relationships (3D-QSAR) to determinate the correct binding mode of glycogen synthase kinase 3 β (GSK-3 β) inhibitors. For the 3D-QSAR (CoMFA and CoMSIA), they used 51 substituted benzofuran-3-yl-(indol-3-yl)maleimides. Two binding modes of the inhibitors to the binding site of GSK-3 β are investigated. The binding mode 1 yielded better 3D-QSAR correlations using both CoMFA and CoMSIA methodologies. The three-component CoMFA model from the steric and electrostatic fields for the experimentally determined pIC₅₀ values has the following statistics: $R^2(\text{cv}) = 0.386$ and $\text{SE}(\text{cv}) = 0.854$ for the cross-validation, and $R^2 = 0.811$ and $\text{SE} = 0.474$ for the fitted correlation. $F(3,47) = 67.034$, and probability of $R^2 = 0(3,47) = 0.000$. The binding mode suggested by the results of this study was consistent with the preliminary results of X-ray crystal structures of inhibitor-bound GSK-3 β . The 3D-QSAR models were used for the estimation of the inhibitory potency of two additional compounds.

4.4. Molecular docking and biological testing of new inhibitors of GSK-3 β

Ya. V. Lavrovskii *et al.* [97] used in this paper a series of new heteroaryl-substituted oxadiazole-5-carboxamide inhibitors of GSK-3 β . Molecular docking was used for the rational selection of synthesized compounds for the subsequent biological testing. It was established that the inhibitory activity of the synthesized compounds strongly depends on the character of substituents in the phenyl ring and the nature of terminal heterocyclic fragments. The most active compounds inhibit GSK-3 β at IC₅₀ in the micromolar range and could be considered as potential drug candidates.

4.5. 3D-QSAR Modelling of Paullones

D. I. Osolodkin *et al.* [98] realized 3D-QSAR study allows one to suggest ways of modification of the molecule to increase its physiological activity. Comparative molecular field analysis (CoMFA) [7] and comparative molecular similarity indices analysis (CoMSIA) [8] are among the most widely used 3D-QSAR methods. The energy of van der Waals and electrostatic interactions of a probe atom (with the charge +1) with molecules of the training set (CoMFA) or the electrostatic, van der Waals, hydrophobic, and donor/acceptor similarity indices (CoMSIA) were used as descriptors. The equation for activity prediction was derived using the partial least squares (PLS) method. The ability of graphic representation of PLS model coefficients was the advantage of the methods and allowed the user to suggest substitutions affecting activity and/or selectivity of the molecules. They had built a new 3D-QSAR model for GSK-3 β inhibition by paullones by means of CoMFA method. This model can be used as a guide for design of new paullone GSK-3 β inhibitors.

4.6. Modeling of Binding Mode of Benzo[e]isoindole-1,3-diones

Zhen Yang *et al.* [99] synthesized benzo[e]isoindole-1,3-dione derivatives, and the effects on GSK-3 β activity and zebrafish embryo growth were evaluated. A series of derivatives showed obvious inhibitory activity against GSK-3 β . The most potent inhibitor, 7,8-dimethoxy-5-methylbenzo[e]isoindole-1,3-dione, showed nanomolar IC₅₀ and obvious phenotype on zebrafish embryo growth associated with the inhibition of GSK-3 β at low micromolar concentration. The interaction mode between this compound and GSK-3 β was characterized by computational modeling. To rationalize the structure-activity relationships of these compounds, the binding modes of the most potent inhibitors 8a and 8b (see **Figure 11**) were modeled using docking simulations. Compounds 8a and 8b were docked into the ATP binding site of GSK-3 β , and the binding modes of lowest energy were analyzed. Compounds 8a and 8b fit the ATP pocket of GSK-3 β well. The maleimide motif of type II formed a pair of hydrogen bonds with the hinge region (Glu133 and Val135) of GSK-3 β , similar to the binding mode of other known maleimides GSK-3 β inhibitors. The two methoxy oxygen atoms formed another two hydrogen bonds with the positively charged Lys85. The methyl group of the methoxy at C-8 position docked to the small back cleft of GSK-3 β . This binding mode explicitly explained the important role of the two methoxy groups at C-7 and C-8 positions. Other result was the 4-ethyl group of 8b docks

to the minor hydrophobic pocket formed by Ile62 and Val70 in the front of the ATP binding site of GSK-3 β (**Figure 12**), which contributed to its higher binding affinity compared to 8a. The docking results also provided a template to understand the structure-activity relationships of other compounds.

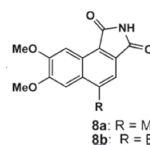


Figure 11. Structure of 8a and 8b

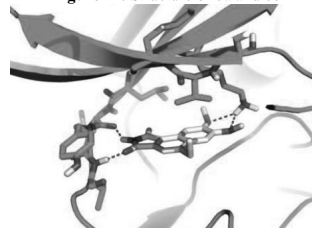


Figure 12. Docked binding modes of compounds 8b in the ATP binding site of GSK-3 β .

CONCLUSIONS

Theoretical studies such as QSAR models have become a very useful tool in this context to substantially reduce time and resources consuming experiments. The functions of β and γ -secretase and its implication in Alzheimer's disease have triggered an active search for potent and selective β and γ -secretase inhibitors. In this paper we can see that the development of theoretical and QSAR models to study γ -secretase inhibitors are usually not many achieved so far, and most of these works present docking studies. Watching this situation we need to develop QSAR models with γ -secretase inhibitors. In this sense, QSAR could play an important role in studying these γ -secretase inhibitors. QSARs can be used as predictive tools for the development of molecules. In this work we developed a new ANN RBF model using the ModesLab descriptors, based on a large database using about 10,000 different drugs obtained from the ChEMBL server.

Acknowledgements

Prado-Prado F. thanks sponsorships for research position at the University of Santiago de Compostela from *Angeles Alvariño*, Xunta de Galicia. All authors acknowledge the Project 07CSA008203PR.

References

- [1] Golde, T. E., Dickson, D. and Hutton, M., Filling the gaps in the abeta cascade hypothesis of Alzheimer's

- disease. *Curr Alzheimer Res*, **2006**, 3, 421-30
- [2] Hutton, M., Perez-Tur, J. and Hardy, J., Genetics of Alzheimer's disease. *Essays Biochem*, **1998**, 33, 117-31
- [3] Younkin, S. G., The role of A beta 42 in Alzheimer's disease. *J Physiol Paris*, **1998**, 92, 289-92
- [4] Sisodia, S. S., Alzheimer's disease: perspectives for the new millennium. *J Clin Invest*, **1999**, 104, 1169-70
- [5] Rovelet-Lecrux, A., Hannequin, D., Raux, G., Le Meur, N., Laquerriere, A., Vital, A., Dumanchin, C., Feuillet, S., Brice, A., Vercelletto, M., Dubas, F., Frebourg, T. and Campion, D., APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat Genet*, **2006**, 38, 24-6
- [6] Ohno, M., Sametsky, E. A., Younkin, L. H., Oakley, H., Younkin, S. G., Citron, M., Vassar, R. and Disterhoft, J. F., BACE1 deficiency rescues memory deficits and cholinergic dysfunction in a mouse model of Alzheimer's disease. *Neuron*, **2004**, 41, 27-33
- [7] Ohno, M., Cole, S. L., Yasvoina, M., Zhao, J., Citron, M., Berry, R., Disterhoft, J. F. and Vassar, R., BACE1 gene deletion prevents neuron loss and memory deficits in 5XFAD APP/PS1 transgenic mice. *Neurobiol Dis*, **2007**, 26, 134-45
- [8] Laird, F. M., Cai, H., Savonenko, A. V., Farah, M. H., He, K., Melnikova, T., Wen, H., Chiang, H. C., Xu, G., Koliatsos, V. E., Borchelt, D. R., Price, D. L., Lee, H. K. and Wong, P. C., BACE1, a major determinant of selective vulnerability of the brain to amyloid-beta amyloidogenesis, is essential for cognitive, emotional, and synaptic functions. *J Neurosci*, **2005**, 25, 11693-709
- [9] Selkoe, D. J., Alzheimer's disease: genes, proteins, and therapy. *Physiol Rev*, **2001**, 81, 741-66
- [10] Vassar, R., BACE1: the beta-secretase enzyme in Alzheimer's disease. *J Mol Neurosci*, **2004**, 23, 105-14
- [11] Hussain, I., Powell, D., Howlett, D. R., Tew, D. G., Meek, T. D., Chapman, C., Gloger, I. S., Murphy, K. E., Southan, C. D., Ryan, D. M., Smith, T. S., Simmons, D. L., Walsh, F. S., Dingwall, C. and Christie, G., Identification of a novel aspartic protease (Asp 2) as beta-secretase. *Mol Cell Neurosci*, **1999**, 14, 419-27
- [12] Sinha, S., Anderson, J. P., Barbour, R., Basi, G. S., Caccavello, R., Davis, D., Doan, M., Dovey, H. F., Frigon, N., Hong, J., Jacobson-Croak, K., Jewett, N., Keim, P., Knops, J., Lieberburg, I., Power, M., Tan, H., Tatsuno, G., Tung, J., Schenk, D., Seubert, P., Suomensaaari, S. M., Wang, S., Walker, D., Zhao, J., McConlogue, L. and John, V., Purification and cloning of amyloid precursor protein beta-secretase from human brain. *Nature*, **1999**, 402, 537-40
- [13] Vassar, R., Bennett, B. D., Babu-Khan, S., Kahn, S., Mendiaz, E. A., Denis, P., Teplow, D. B., Ross, S., Amarante, P., Loeloff, R., Luo, Y., Fisher, S., Fuller, J., Edenson, S., Lile, J., Jarosinski, M. A., Biere, A. L., Curran, E., Burgess, T., Louis, J. C., Collins, F., Treanor, J., Rogers, G. and Citron, M., Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science*, **1999**, 286, 735-41
- [14] Yan, R., Bienkowski, M. J., Shuck, M. E., Miao, H., Tory, M. C., Pauley, A. M., Brashier, J. R., Stratman, N. C., Mathews, W. R., Buhl, A. E., Carter, D. B., Tomasselli, A. G., Parodi, L. A., Heinrichson, R. L. and Gurney, M. E., Membrane-anchored aspartyl protease with Alzheimer's disease beta-secretase activity. *Nature*, **1999**, 402, 533-7
- [15] Goate, A., Chartier-Harlin, M. C., Mullan, M., Brown, J., Crawford, F., Fidani, L., Giuffra, L., Haynes, A., Irving, N., James, L. and et al., Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature*, **1991**, 349, 704-6
- [16] Schellenberg, G. D., Bird, T. D., Wijsman, E. M., Orr, H. T., Anderson, L., Nemens, E., White, J. A., Bonnycastle, L., Weber, J. L., Alonso, M. E. and et al., Genetic linkage evidence for a familial Alzheimer's disease locus on chromosome 14. *Science*, **1992**, 258, 668-71
- [17] Haass, C., Schlossmacher, M. G., Hung, A. Y., Vigo-Pelfrey, C., Mellon, A., Ostaszewski, B. L., Lieberburg, I., Koo, E. H., Schenk, D., Teplow, D. B. and et al., Amyloid beta-peptide is produced by cultured cells during normal metabolism. *Nature*, **1992**, 359, 322-5

- [18] Seubert, P., Oltersdorf, T., Lee, M. G., Barbour, R., Blomquist, C., Davis, D. L., Bryant, K., Fritz, L. C., Galasko, D., Thal, L. J. and et al., Secretion of beta-amyloid precursor protein cleaved at the amino terminus of the beta-amyloid peptide. *Nature*, **1993**, 361, 260-3
- [19] Zhao, J., Paganini, L., Mucke, L., Gordon, M., Refolo, L., Carman, M., Sinha, S., Oltersdorf, T., Lieberburg, I. and McConlogue, L., Beta-secretase processing of the beta-amyloid precursor protein in transgenic mice is efficient in neurons but inefficient in astrocytes. *J Biol Chem*, **1996**, 271, 31407-11
- [20] Citron, M., Teplow, D. B. and Selkoe, D. J., Generation of amyloid beta protein from its precursor is sequence specific. *Neuron*, **1995**, 14, 661-70
- [21] Phiel, C. J., Wilson, C. A., Lee, V. M.-Y. and Klein, P. S., GSK-3 α regulates production of Alzheimer's disease amyloid- β peptides. *Nature*, **2003**, 423, 435-439
- [22] Jamloki, A., Karthikeyan, C. and Sharma, S. K., QSAR Studies on Some GSK-3 α Inhibitory 6-aryl-pyrazolo-(3,4-b)pyrimidines. *Asian Journal of Biochemistry*, **2006**, 1, 236-243
- [23] Ali, A., Hoefflich, K. P. and Woodgett, J. R., Glycogen Synthase Kinase-3: Properties, Functions, and Regulation. *Chem Rev*, **2001**, 101, 2527-2540
- [24] Martinez, A., Castro, A., Dorransoro, I. and Alonso, M., Glycogen synthase kinase 3 (GSK-3) inhibitors as new promising drugs for diabetes, neurodegeneration, cancer, and inflammation. *Med Res Rev*, **2002**, 22, 373-84
- [25] Martinez, A., Alonso, M., Castro, A., Perez, C. and Moreno, F. J., First Non-ATP Competitive Glycogen Synthase Kinase 3B (GSK-3B) Inhibitors: Thiazolidinones (TDZD) as Potential Drugs for the Treatment of Alzheimer's Disease. *J Med Chem*, **2002**, 45, 1292-1299
- [26] Hagit, E.-F., Glycogen synthase kinase 3: an emerging therapeutic target. *Trends in Molecular Medicine*, **2002**, 8, 126-132
- [27] Lee, V. M., Goedert, M. and Trojanowski, J. Q., Neurodegenerative tauopathies. *Ann Rev Neurosci*, **2001**, 24, 1121-1159
- [28] Flaherty, D. B., Sorrea, P. J., Tomasienicz, G. H. and Wood, G. J., Phosphorylation of human tau protein by microtubule-associated kinases: GSK3beta and cdk5 are key participants. *J Neuro Sci Res*, **2000**, 62, 463-472
- [29] Bhat, R. V., Budd Haerberlein, S. L. and Avila, J. J., Glycogen synthase kinase 3: a drug target for CNS therapies. *J Neurochem*, **2004**, 89, 1313-1317
- [30] Sivaprakasam, P., Xiea, A. and Doerksen, R. J., Probing the physicochemical and structural requirements for glycogen synthase kinase-3a inhibition: 2D-QSAR for 3-anilino-4-phenylmaleimides. *Bioorganic & Medicinal Chemistry Letters*, **2006**, 14, 8210-8218
- [31] Ishiguro, K., Takamatsu, M., Tomizawa, K., Omori, A., Takahashi, M., Arioka, M., Uchida, T. and Imahori, K., Tau protein kinase I converts normal tau protein into A68-like component of paired helical filaments. *J. Biol. Chem.*, **1992**, 267, 10897-10901
- [32] Pei, J. J., Tanaka, T., Tung, Y. C., Braak, E., Iqbal, K. and Grundke-Iqbal, I., Distribution, Levels, and Activity of Glycogen Synthase Kinase-3 in the Alzheimer Disease Brain. *J. Neuropathol. Exp. Neurol.*, **1997**, 56, 70-78
- [33] Pei, J. J., Braak, H., Grundke-Iqbal, I., Iqbal, K., Winblad, B. and Cowburn, R. F., Distribution of active glycogen synthase kinase 3beta (GSK-3beta) in brains staged for Alzheimer disease neurofibrillary changes. *J. Neuropathol. Exp. Neurol.*, **1999**, 58, 1010-1019
- [34] Baum, L., Hansen, L., Masliah, E. and Saitoh, T., Glycogen synthase kinase 3 alteration in Alzheimer disease is related to neurofibrillary tangle formation. *Mol. Chem. Neuropathol.*, **1996**, 29, 253-261
- [35] Lucas, J. J., Hernandez, F., Gomez-Ramos, P., Moran, M. A., Hen, R. and Avila, J., Decreased nuclear -catenin, tau hyperphosphorylation and neurodegeneration in GSK-3 conditional transgenic mice. *EMBO J.*, **2001**, 20, 27-39
- [36] Hernandez, F., Borrell, J., Guaza, C., Avila, J. and Lucas, J. J., Spatial learning deficit in transgenic mice that conditionally over-express GSK-3beta in the brain but do not form tau

- filaments. *J. Neurochem.*, **2002**, 83, 1529-1533
- [37] Santana, L., Uriarte, E., González-Díaz, H., Zagotto, G., Soto-Otero, R. and Mendez-Alvarez, E., A QSAR model for in silico screening of MAO-A inhibitors. Prediction, synthesis, and biological assay of novel coumarins. *J Med Chem*, **2006**, 49, 1149-56
- [38] Marrero-Ponce, Y., Linear indices of the "molecular pseudograph's atom adjacency matrix": definition, significance-interpretation, and application to QSAR analysis of flavone derivatives as HIV-1 integrase inhibitors. *J Chem Inf Comput Sci*, **2004**, 44, 2010-26
- [39] Vilar, S., Santana, L. and Uriarte, E., Probabilistic neural network model for the in silico evaluation of anti-HIV activity and mechanism of action. *J Med Chem* **2006**, 49, 1118-1124
- [40] Marrero-Ponce, Y., Khan, M. T., Casanola Martin, G. M., Ather, A., Sultankhodzhaev, M. N., Torrens, F. and Rotondo, R., Prediction of tyrosinase inhibition activity using atom-based bilinear indices. *ChemMedChem*, **2007**, 2, 449-78
- [41] Casanola-Martin, G. M., Marrero-Ponce, Y., Khan, M. T., Ather, A., Sultan, S., Torrens, F. and Rotondo, R., TOMOCOMD-CARDD descriptors-based virtual screening of tyrosinase inhibitors: evaluation of different classification model combinations using bond-based linear indices. *Bioorg Med Chem*, **2007**, 15, 1483-503
- [42] Casanola-Martin, G. M., Marrero-Ponce, Y., Khan, M. T., Ather, A., Khan, K. M., Torrens, F. and Rotondo, R., Dragon method for finding novel tyrosinase inhibitors: Biosilico identification and experimental in vitro assays. *Eur J Med Chem*, **2007**, 42, 1370-81
- [43] Nunez, M. B., Maguna, F. P., Okulik, N. B. and Castro, E. A., QSAR modeling of the MAO inhibitory activity of xanthenes derivatives. *Bioorg Med Chem Lett*, **2004**, 14, 5611-5617
- [44] Todeschini, R. and Consonni, V., Handbook of Molecular Descriptors. *Wiley VCH*, **2000**,
- [45] González-Díaz, H., González-Díaz, Y., Santana, L., Ubeira, F. M. and Uriarte, E., Proteomics, networks and connectivity indices. *Proteomics*, **2008**, 8, 750-778
- [46] Zhao, C. J. and Dai, Q. Y., [Recent advances in study of antinociceptive conotoxins]. *Yao Xue Xue Bao*, **2009**, 44, 561-5
- [47] Jacob, R. B. and McDougal, O. M., The M-superfamily of conotoxins: a review. *Cellular and Molecular Life Sciences*, **2010**, 67, 17-27
- [48] Giuliani, A., Di Paola, L. and Setola, R., Proteins as Networks: A Mesoscopic Approach Using Haemoglobin Molecule as Case Study. *Curr Proteomics*, **2009**, 6, 235-245
- [49] Vilar, S., Gonzalez-Diaz, H., Santana, L. and Uriarte, E., A network-QSAR model for prediction of genetic-component biomarkers in human colorectal cancer. *Journal of Theoretical Biology*, **2009**, 261, 449-58
- [50] Concu, R., Dea-Ayuela, M. A., Perez-Montoto, L. G., Prado-Prado, F. J., Uriarte, E., Bolas-Fernandez, F., Podda, G., Pazos, A., Munteanu, C. R., Ubeira, F. M. and Gonzalez-Diaz, H., 3D entropy and moments prediction of enzyme classes and experimental-theoretic study of peptide fingerprints in Leishmania parasites. *Biochimica et Biophysica Acta*, **2009**, 1794, 1784-94
- [51] Torrens, F. and Castellano, G., Topological Charge-Transfer Indices: From Small Molecules to Proteins. *Curr Proteomics*, **2009**, 204-213
- [52] Vázquez, J. M., Aguiar, V., Seoane, J. A., Freire, A., Serantes, J. A., Dorado, J., Pazos, A. and Munteanu, C. R., Star Graphs of Protein Sequences and Proteome Mass Spectra in Cancer Prediction. *Curr Proteomics*, **2009**, 6, 275-288
- [53] Gonzalez-Diaz, H., Quantitative studies on Structure-Activity and Structure-Property Relationships (QSAR/QSPR). *Curr Top Med Chem*, **2008**, 8, 1554
- [54] Ivanciuc, O., Weka machine learning for predicting the phospholipidosis inducing potential. *Curr Top Med Chem*, **2008**, 8, 1691-709
- [55] Gonzalez-Díaz, H., Prado-Prado, F. and Ubeira, F. M., Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach. *Curr Top Med Chem*, **2008**, 8, 1676-90
- [56] Duardo-Sanchez, A., Patlewicz, G. and Lopez-Diaz, A., Current topics on software use in medicinal chemistry: intellectual property, taxes, and regulatory issues. *Curr Top Med Chem*, **2008**, 8, 1666-75

- [57] Wang, J. F., Wei, D. Q. and Chou, K. C., Drug candidates from traditional chinese medicines. *Curr Top Med Chem*, **2008**, 8, 1656-65
- [58] Helguera, A. M., Combes, R. D., Gonzalez, M. P. and Cordeiro, M. N., Applications of 2D descriptors in drug design: a DRAGON tale. *Curr Top Med Chem*, **2008**, 8, 1628-55
- [59] Gonzalez, M. P., Teran, C., Saiz-Urra, L. and Teijeira, M., Variable selection methods in QSAR: an overview. *Curr Top Med Chem*, **2008**, 8, 1606-27
- [60] Caballero, J. and Fernandez, M., Artificial neural networks from MATLAB in medicinal chemistry. Bayesian-regularized genetic neural networks (BRGNN): application to the prediction of the antagonistic activity against human platelet thrombin receptor (PAR-1). *Curr Top Med Chem*, **2008**, 8, 1580-605
- [61] Wang, J. F., Wei, D. Q. and Chou, K. C., Pharmacogenomics and personalized use of drugs. *Curr Top Med Chem*, **2008**, 8, 1573-9
- [62] Vilar, S., Cozza, G. and Moro, S., Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. *Curr Top Med Chem*, **2008**, 8, 1555-72
- [63] Prado-Prado, F. J., Garcia-Mera, X. and Gonzalez-Diaz, H., Multi-target spectral moment QSAR versus ANN for antiparasitic drugs against different parasite species. *Bioorg Med Chem*, **2010**, 18, 2225-31
- [64] Gonzalez-Diaz, H., Network topological indices, drug metabolism, and distribution. *Curr Drug Metab*, 11, 283-4
- [65] Khan, M. T., Predictions of the ADMET properties of candidate drug molecules utilizing different QSAR/QSPR modelling approaches. *Curr Drug Metab*, 11, 285-95
- [66] Mrabet, Y. and Semmar, N., Mathematical methods to analysis of topology, functional variability and evolution of metabolic systems based on different decomposition concepts. *Curr Drug Metab*, 11, 315-41
- [67] Martinez-Romero, M., Vazquez-Naya, J. M., Rabunal, J. R., Pita-Fernandez, S., Macenlle, R., Castro-Alvarino, J., Lopez-Roses, L., Ulla, J. L., Martinez-Calvo, A. V., Vazquez, S., Pereira, J., Porto-Pazos, A. B., Dorado, J., Pazos, A. and Munteanu, C. R., Artificial intelligence techniques for colorectal cancer drug metabolism: ontology and complex network. *Curr Drug Metab*, 11, 347-68
- [68] Zhong, W. Z., Zhan, J., Kang, P. and Yamazaki, S., Gender specific drug metabolism of PF-02341066 in rats--role of sulfoconjugation. *Curr Drug Metab*, 11, 296-306
- [69] Wang, J. F. and Chou, K. C., Molecular modeling of cytochrome P450 and drug metabolism. *Curr Drug Metab*, 11, 342-6
- [70] Gonzalez-Diaz, H., Duardo-Sanchez, A., Ubeira, F. M., Prado-Prado, F., Perez-Montoto, L. G., Concu, R., Podda, G. and Shen, B., Review of MARCH-INSIDE & complex networks prediction of drugs: ADMET, anti-parasite activity, metabolizing enzymes and cardiotoxicity proteome biomarkers. *Curr Drug Metab*, 11, 379-406
- [71] Garcia, I., Diop, Y. F. and Gomez, G., QSAR & complex network study of the HMGR inhibitors structural diversity. *Curr Drug Metab*, 11, 307-14
- [72] Chou, K. C., Graphic rule for drug metabolism systems. *Curr Drug Metab*, 11, 369-78
- [73] Concu, R., Podda, G., Ubeira, F. M. and Gonzalez-Diaz, H., Review of QSAR Models for Enzyme Classes of Drug Targets: Theoretical Background and Applications in Parasites, Hosts, and other Organisms. *Current Pharmaceutical Design*, **2010**, 16, 2710-23
- [74] Estrada, E., Molina, E., Nodarse, D. and Uriarte, E., Structural Contributions of Substrates to their Binding to P-Glycoprotein. A TOPS-MODE Approach. *Current Pharmaceutical Design*, **2010**, 16, 2676-709
- [75] Garcia, I., Fall, Y. and Gomez, G., QSAR, Docking, and CoMFA Studies of GSK3 Inhibitors. *Current Pharmaceutical Design*, **2010**, 16, 2666-75
- [76] González-Díaz, H., QSAR and Complex Networks in Pharmaceutical Design, Microbiology, Parasitology, Toxicology, Cancer, and Neurosciences. *Current Pharmaceutical Design*, **2010**, 16, 2598-600
- [77] Gonzalez-Diaz, H., Romaris, F., Duardo-Sanchez, A., Perez-Mototo, L. G., Prado-Prado, F., Patlewicz, G. and

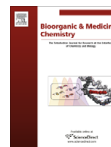
- Ubeira, F. M., Predicting drugs and proteins in parasite infections with topological indices of complex networks: theoretical backgrounds, applications, and legal issues. *Current Pharmaceutical Design*, **2010**, 16, 2737-64
- [78] Marrero-Ponce, Y., Casanola-Martin, G. M., Khan, M. T., Torrens, F., Rescigno, A. and Abad, C., Ligand-Based Computer-Aided Discovery of Tyrosinase Inhibitors. Applications of the TOMOCOMD-CARDD Method to the Elucidation of New Compounds. *Current Pharmaceutical Design*, **2010**, 16, 2601-24
- [79] Munteanu, C. R., Fernandez-Blanco, E., Seoane, J. A., Izquierdo-Novo, P., Rodriguez-Fernandez, J. A., Prieto-Gonzalez, J. M., Rabunal, J. R. and Pazos, A., Drug Discovery and Design for Complex Diseases through QSAR Computational Methods. *Current Pharmaceutical Design*, **2010**, 16, 2640-55
- [80] Roy, K. and Ghosh, G., Exploring QSARs with Extended Topochemical Atom (ETA) Indices for Modeling Chemical and Drug Toxicity. *Current Pharmaceutical Design*, **2010**, 16, 2625-39
- [81] Speck-Planche, A., Scotti, M. T. and de Paulo-Emerenciano, V., Current pharmaceutical design of antituberculosis drugs: future perspectives. *Current Pharmaceutical Design*, **2010**, 16, 2656-65
- [82] Vazquez-Naya, J. M., Martinez-Romero, M., Porto-Pazos, A. B., Novoa, F., Valladares-Ayerbes, M., Pereira, J., Munteanu, C. R. and Dorado, J., Ontologies of drug discovery and design for neurology, cardiology and oncology. *Current Pharmaceutical Design*, **2010**, 16, 2724-36
- [83] Garcia, I., Diop, Y. F. and Gomez, G., QSAR & complex network study of the HMGR inhibitors structural diversity. *Curr Drug Metab*, **2010**, 11, 307-14
- [84] Choi, S. J., Cho, J. H., Im, I., Lee, S. D., Jang, J. Y., Oh, Y. M., Jung, Y. K., Jeon, E. S. and Kim, Y. C., Design and synthesis of 1,4-dihydropyridine derivatives as BACE-1 inhibitors. *Eur J Med Chem*, **2010**, 45, 2578-90
- [85] Yi Mok, N., Chadwick, J., Kellett, K. A., Hooper, N. M., Johnson, A. P. and Fishwick, C. W., Discovery of novel non-peptide inhibitors of BACE-1 using virtual high-throughput screening. *Bioorg Med Chem Lett*, **2009**, 19, 6770-4
- [86] Sato, T., Ananda, K., Cheng, C. I., Suh, E. J., Narayanan, S. and Wolfe, M. S., Distinct pharmacological effects of inhibitors of signal peptide peptidase and gamma-secretase. *J Biol Chem*, **2008**, 283, 33287-95
- [87] Sammi, T., Silakari, O. and Ravikumar, M., Three-dimensional quantitative structure-activity relationship (3D-QSAR) studies of various benzodiazepine analogues of gamma-secretase inhibitors. *J Mol Model*, **2009**, 15, 343-8
- [88] Al-Nadaf, A., Abu Sheikha, G. and Taha, M. O., Elaborate ligand-based pharmacophore exploration and QSAR analysis guide the synthesis of novel pyridinium-based potent beta-secretase inhibitory leads. *Bioorg Med Chem*, **2010**, 18, 3088-115
- [89] Pandey, A., Mungalpara, J. and Mohan, C. G., Comparative molecular field analysis and comparative molecular similarity indices analysis of hydroxyethylamine derivatives as selective human BACE-1 inhibitor. *Mol Divers*, **2010**, 14, 39-49
- [90] Polgar, T. and Keseru, G. M., Virtual screening for beta-secretase (BACE1) inhibitors reveals the importance of protonation states at Asp32 and Asp228. *J Med Chem*, **2005**, 48, 3749-55
- [91] Hetenyi, C., Paragi, G., Maran, U., Timar, Z., Karelson, M. and Penke, B., Combination of a modified scoring function with two-dimensional descriptors for calculation of binding affinities of bulky, flexible ligands to proteins. *J Am Chem Soc*, **2006**, 128, 1233-9
- [92] Moitessier, N., Therrien, E. and Hanessian, S., A method for induced-fit docking, scoring, and ranking of flexible ligands. Application to peptidic and pseudopeptidic beta-secretase (BACE 1) inhibitors. *J Med Chem*, **2006**, 49, 5885-94
- [93] Sivaprakasam, P., Daga, P. R., Xie, A. and Doerksen, R. J., Glycogen synthase kinase-3 inhibition by 3-anilino-4-phenylmaleimides: insights from 3D-QSAR and docking. *J Comput Aided Mol Des*, **2009**, 23, 113-127
- [94] Saitoh, M., Kunitomo, J., Kimura, E., Hayase, Y., Kobayashi, H., Uchiyama,

- N., Kawamoto, T., Tanaka, T., Mol, C. D., Dougan, D. R., Textor, G. S., Snell, G. P. and Itoh, F., Design, synthesis and structure–activity relationships of 1,3,4-oxadiazole derivatives as novel inhibitors of glycogen synthase kinase-3beta. *Bioorganic & Medicinal Chemistry*, **2009**, 17, 2017-2029
- [95] Freitas, M. P., Goodarzi, M. and Jensen, R., Feature Selection and Linear/Nonlinear Regression Methods for the Accurate Prediction of Glycogen Synthase Kinase-3 β Inhibitory Activities *Journal of Chemical Information and Modeling* **2009**, 49, 824-832
- [96] Kim, K. H., Gaisina, I., Gallier, F., Holzle, D., Blond, S. Y., Mesecar, A. and Kozikowski, A. P., Use of molecular modeling, docking, and 3D-QSAR studies for the determination of the binding mode of benzofuran-3-yl-(indol-3-yl)maleimides as GSK-3 β inhibitors. *J Mol Model*, **2009**, 15, 1463-1479
- [97] Ryzhova, E. A., Koryakova, A. G., Bulanova, E. A., Mikitas, O. V., Karapetyan, R. N., Lavrovskii, Y. V. and Ivashchenko, A. V., Synthesis, Molecular Docking, and Biological Testing of New Selective Inhibitors of Glycogen Synthase Kinase 3beta. *Pharmaceutical Chemistry Journal*, **2009**, 43, 148-153
- [98] Osolodkin, D. I., Shulga, D. A., Tsareva, D. A., Oliferenko, A. A., Palyulin, V. A. and Zefirov, N. S., The Choice of Atomic Charges Calculation Scheme in 3D-QSAR Modelling of GSK-3 β Inhibition by Paullones. *Biochemistry, Biophysics and Molecular Biology*, **2010**, 434, 274-278
- [99] Zou, H., Zhou, L., Li, Y., Cui, Y., Zhong, H., Pan, Z., Yang, Z. and Quan, J., Benzo[e]isoindole-1,3-diones as Potential Inhibitors of Glycogen Synthase Kinase-3 (GSK-3). Synthesis, Kinase Inhibitory Activity, Zebrafish Phenotype, and Modeling of Binding Mode. *J. Med. Chem.*, **2010**, 53, 994-1003



Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

QSAR and complex network study of the chiral HMGR inhibitor structural diversity

Isla García^a, Cristian Robert Munteanu^{b,c}, Yagamare Fall^a, Generosa Gómez^a, Eugenio Uriarte^d, Humberto González-Díaz^{c,*}

^a Department of Organic Chemistry, University of Vigo, Spain

^b Department of Chemistry, REQUIMTE/Faculty of Science, University of Porto, 4169-007 Porto, Portugal

^c Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain

^d UBICA, Institute of Industrial Pharmacy, Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain

ARTICLE INFO

Article history:

Received 23 May 2008

Revised 31 October 2008

Accepted 6 November 2008

Available online 9 November 2008

Keywords:

QSAR

Topological indices

Complex network

Chiral compound

Lipid-lowering agent

Cholesterol level

Atherosclerotic disease

Anti-parasite drug

Trypanosoma cruzi

Chagas' disease

3-Hydroxy-3-methyl-glutaryl coenzyme A reductase

ABSTRACT

Efficient drugs such as statins or mevinic acids are inhibitors of the rate-limiting enzyme of cholesterol biosynthesis, 3-hydroxy-3-methyl-glutaryl coenzyme A reductase (HMGR), an enzyme responsible for the double reduction of 3-hydroxy-3-methyl-glutaryl coenzyme A into mevalonic acid. These compounds promoted the synthesis and evaluation of new inhibitors for HMGR, named HMGRIs. The high number of possible candidates creates the necessity of Quantitative Structure–Activity Relationship models in order to guide the HMGR synthesis. There are two main problems of the reported QSAR models: the homogeneous series of the compounds and the chirality of many candidates. In this work, we propose for the first time a QSAR model for a very large and heterogeneous series of HMGRIs. The model is based on the Topological Indices (TIs) of molecular structures. Using the predictions of this model as input, we construct the first complex network that describes the drug–drug similarity relationships for more than 1600 experimentally non-explored chiral HMGRIs isomers. We also presented a reduced version of this network (Giant Component) that contains the most representative set of chiral HMGR candidates. The work suggests a new mixed application in the QSAR study of relevant aspects of structural diversity by using chiral/non-chiral TIs, combined with complex networks.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Hypercholesterolemia is well-known as the primary risk factor in atherosclerotic and coronary heart diseases.^{1,2} Clinical studies with lipid-lowering agents have established that the decrease of high serum cholesterol levels reduces the incidence of cardiovascular mortality. Statins and mevinic acids are two efficient drugs known as inhibitors of the rate-limiting enzyme of cholesterol biosynthesis, 3-hydroxy-3-methyl-glutaryl coenzyme A reductase (HMGR), enzyme responsible for the double reduction of 3-hydroxy-3-methyl-glutaryl coenzyme A into mevalonic acid. The drug control of this enzyme is efficient in reducing the levels of cholesterol in plasma.^{3,4} The structure of statins and its derivatives is characterized by the desmethylmevalonic acid or by the lactone. The pharmacophore is connected to a lipophilic ring, such as hexahydronaphthalene, indole, pyrrole, pyrimidine or quinine, by a

linking element (a two-carbon spacer). The biologically active form of mevinic acids is represented by the open chain hydroxyl-acid, which mimics the HMGR natural substrate.⁵

Another interesting potential use of this type of drugs could be the control of parasite infections. Urbina et al. studied in the protozoan parasite *Trypanosoma* (*Schizotrypanum*) *cruzi* the anti-proliferative effects of mevinolin (lovastatin), a drug from the family of HMGR inhibitors (HMGRIs), and its ability to potentiate the action of specific ergosterol biosynthesis inhibitors, such as ketoconazole and terbinafine (both in vitro and in vivo). These results confirm the synergic action against the proliferative stages of *T. cruzi* (both in vitro and in vivo) and the ergosterol biosynthesis inhibition (in vivo) by acting in different points of the pathway. In addition, the medical practice suggests that mevinolin, combined with azoles, such as ketoconazole, can be used in the treatment of human Chagas' disease.⁶

The success of HMGR drugs encourages many researchers to look for new derivatives. The high asymmetry of many interesting compounds (they are commonly present various chiral centres) makes difficult to obtain new drug candidates by low-cost efficient

* Corresponding author. Tel.: +34 981 563100; fax: +34 981 594912.

E-mail addresses: humberto.gonzalez@usc.es, gonzalezdiaz@yahoo.es (H. González-Díaz).

asymmetric organic synthesis or synthesis followed by asymmetric separation.⁷ Thus, the development of computational methods becomes an important step for the exploration of the large space of potential chiral isomers and for the selection of a higher probability of HMGR activity in the asymmetric synthesis. In Figure 1 we illustrate the asymmetric structures of some well-known HMGRs, including those above-mentioned. In order to solve this problem, we use the Quantitative Structure–Activity Relationships (QSAR), a mathematical relationship that links in a quantitative manner the chemical structure and the pharmacological activity of compounds.^{8–11} The reported models are efficient and they are based mainly on 3D descriptors (3D-QSAR) and/or based on specific homologous compound series.^{12–14} One of the most efficient and fast QSAR techniques (relative to 3D-QSAR) makes use of molecular graph Topological Indices (TIs)^{15–18} in order to describe the molecular structure of drugs. The TI-based QSAR (TI-QSAR) models can be used to explore *in silico* different biological activities in a large series of compounds. In this sense, we can use TI-QSAR to predict general activities such as: anti-fungal,¹⁹ anti-cancer,²⁰ anti-parasite,^{21,22} sedative/hypnotic, anti-bacterial,^{23,24} analgesia,²⁵ hypoglycaemic activity,²⁶ herbicide action²⁷ or drug toxicity.²⁸ Alternatively, we can use TI-QSAR models to explore the relationships between the structural spaces of compounds as inhibitors for specific enzymes, such as MAO inhibitors,²⁹ HIV-1 integrase inhibitors,³⁰ and/or protease inhibitors³¹ or tyrosinase inhibi-

tors.^{32–34} Unfortunately, the classic TIs do not consider important 3D features, such as chirality, and the biological activity of different stereo-isomers of the same molecule can be notably different. Therefore, several authors have recently defined novel Chiral TIs (CTIs), which opens a fast investigation gateway of large spaces of one/many chiral center stereo-isomers. CTI-based QSARs (CTI-QSARs) are as fast as TI-QSAR and chiral sensitive because they do not assume the knowledge of 3D-structure.^{35–42} However, a TI-QSAR or CTI-QSAR models for a large and heterogeneous series of HMGRs have not been reported up to date.

In general, HMGRs, such as statins, are molecules with chiral centers. In many cases, only a low percent of the possible stereo-isomers of one molecular skeleton was obtained by synthesis, characterized and assayed to know their activity profiles *in vitro* and *in vivo*. Using a CTI-QSAR, it is possible to predict the activity profiles of many derivatives of different compounds in a computer simulation. This allows us to select for synthesis the derivatives with a high predicted activity and a similar profile to other successful drugs. Once predicted the most significant derivatives, it may be highly difficult to experimentally access by organic chemical synthesis all selected derivatives (as in the case of many isomers in the entire space of stereo-isomers). Thus, it is highly important to draw some conclusions about the overall degree of similarity in activity profiles of whole spaces of compounds and their derivatives. We can achieve this goal by means of similarity/dissimilar-

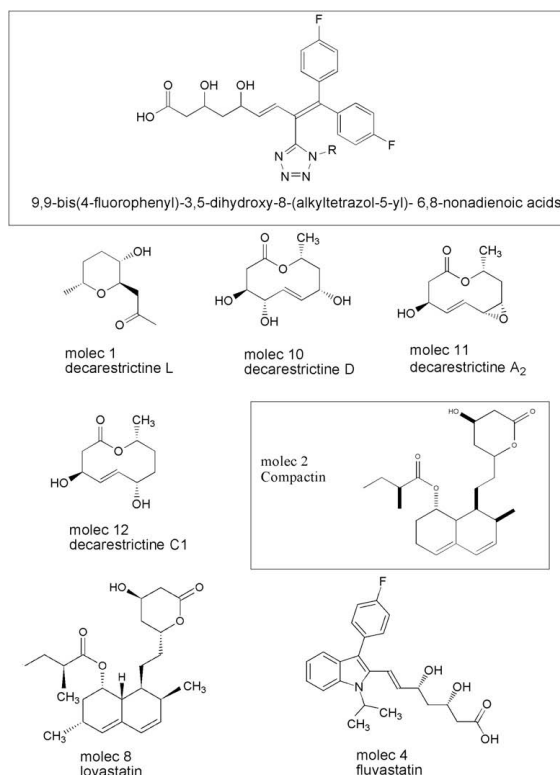


Figure 1. Different types of chiral HMGRs.

ity Complex Networks (CNs).⁴³ These CNs are similar to the graphs, composed by many nodes connected by edges.⁴⁴ In molecular sciences, structural units such as atoms,^{45–47} aminoacids,⁴⁸ nucleotides,^{49–51} proteins,^{52,53} RNAs,⁵⁴ genes⁵⁵ or even organisms⁵⁶ often play the role of nodes in the CNs. Complementarily, the chemical bonds (in case of classic molecular graph), chemical reactions, nucleotide–nucleotide hydrogen bonds,^{57,58} amino acid–amino acid spatial contacts,⁵⁹ metabolic pathways steps,⁶⁰ protein–protein interactions,⁶¹ RNA–RNA co-expression,⁶² gene–gene regulation⁶³ or any other kind of structural or functional relationships play the role of edges in the CNs. These CNs can be used to characterize overall similarity/dissimilarity relationships¹⁸ as well as properties such as bipartivity,⁶⁴ community structure⁶⁵ or small world property,⁶⁶ in order to unravel complex statistical properties of the phenomena under study. In a previous work, our group used drug–drug CNs to study the overall properties of the activity profiles of several anti-fungals against different species. The CNs were assembled based on the activity predictions made with a TI-QSAR model.⁶⁷

There have been reported no CNs for the space of stereo-isomers of potential HMGRIs. In this work, we compiled the biological activity and molecular structure of several well-known HMGRIs. Next, we calculated several TIs and CTIs for these molecules and constructed the first CTI-QSAR models for HMGRIs. Using these models, we predicted the in vitro and in vivo activity profile of all the possible chiral isomers of the studied molecules. With these values we assembled CNs for a large space composed by chiral HMGRIs isomers. We were able to design the CNs for HMGRIs and compared them with the CN resulted when omitted the chirality information. In addition, we proposed a reduced CN that contains the most relevant isomers to be experimentally assayed. The present work may become a useful tool to guide the synthesis of new HMGRi chiral isomers.

2. Results and discussion

2.1. QSAR model

In the first step, we performed an LDA in order to obtain a QSAR model that allowed us to classify or not a molecule as a potential HMGRi. The QSAR database is shown in Table SM1 and the correspondent chemical structures and names in Table SM2 (from the Supplementary Materials). The equation of the QSAR model found is the following:

$$\begin{aligned} \text{HMGRi-score} &= 0.169 \cdot \text{Cc} - 0.029 \cdot \text{Svd} + 0.561 \cdot \text{D} - 0.002 \cdot \text{W} \\ &\quad - 3.992 \cdot \text{Nrc} + 12.415 \cdot \text{Bn} + 12.167 \\ N &= 183 \quad \text{Rc} = 0.7 \quad \lambda = 0.54 \quad F = 25.15 \\ p\text{-level} &< 0.001 \end{aligned} \quad (1)$$

where N is the number of compounds used for training, Rc is the canonical regression coefficient, λ is the Wilks's statistic parameter, F is the Fisher ratio and p is the level of error. This model is based on the following parameters presented in Table 1: cluster count (Cc), sum of valence degree (Svd), Wiener index (W), number of R centers (Nrc) and normal chiral/non-chiral balance (Bn). In this work, the high value of $\text{Rc} = 0.7$ (Rc may vary from 0 to 1) indicates a strong correlation and consequently supports the quality of the model.⁶⁸ On the other hand, the lower $\lambda = 0.54$ value found points to a better class or group separation (HMGRIs vs non-HMGRIs). Additionally, the F -test can reject the hypothesis of group overlapping with a level of error (p -level) < 0.001 . Thus, we consider that the model performs a good separation of both groups. More direct evidence of the fitting quality are the high values of Sensitivity, Specificity and Accuracy for both training and cv series (see Table 2).⁶⁹

Table 1
Symbols and names of the molecular descriptors used in this work

Symbol	Name	TI	Chiral
W	Wiener index	Yes	No
J	Balaban index	Yes	No
Cc	Cluster count	Yes	No
Svd	Sum of valence degree	Yes	No
D	Diameter	Yes	No
R	Radius	Yes	No
Sha	Shape attribute	Yes	No
Shc	Shape coefficient	Yes	No
Sod	Sum of degree	Yes	No
Mti	Molecular topological index	Yes	No
Nrb	Number of Rotable Bonds	Yes	No
Tc	Total connectivity	Yes	No
Tvc	Total valence connectivity	Yes	No
Ntc	Number of total centers	Yes	Yes
Ncc	Number of chiral centers	Yes	Yes
Nncc	Number of non-chiral centers	Yes	Yes
Nsc	Number of S centers	Yes	Yes
Nrc	Number of R centers	Yes	Yes
Bg = Ncc – Ncc	General chiral/non-chiral balance	Yes	Yes
Bn = (Ncc – Nncc)/Ntc	Normal chiral/non-chiral balance	Yes	Yes
Brs = (Nrc – Nsc)/(Ncc + 1)	Relative RS balance	Yes	Yes
Inv	In vivo test indicator	No	No
Psa	Polar surface area	No	No

Table 2
QSAR model results

Parameter	Statistics for classes		Predicted output	
	Value (%)	Observed input	Non-HMGRIs	HMGRIs
Specificity	81.3	Train		
		Non-HMGRIs	39	9
		HMGRIs	14	121
Sensitivity	89.6	Total	53	130
		Accuracy	87.4	
Specificity	71.4	Cv		
		Non-HMGRIs	10	4
		HMGRIs	7	39
Sensitivity	84.8	Total	17	43
		Accuracy	81.7	

All the parameters included in the equation have demonstrated to present statistically a significant contribution to the classification using the Fisher test, see Table 3. The *Forward stepwise* analysis selects both TIs (Cc, Svd, D, and W) and CTIs (Nrc and Bn), but does not include the In vivo test indicator parameter (Inv). This could indicate that there are no important differences for in vitro versus in vivo results in this database and confirms the use of HMGRi in vitro test results. The results stimulate, in general, the in vivo experiments and, in particular, the search for QSAR models. Cc is related to structure cycles and has a positive contribution of 0.619 to the HMGRi-score. Thus, it could be rationalized, considering that most of these molecules are cycle-containing structures. The diameter of the graph (D) has also a positive contribution of 0.561, possibly due to the fact that drugs need to have certain distance between the separated atoms, in order to guarantee the interaction with the receptor. The negative contribution of -0.002 for W indicates that groups with high level of molecular branching could decrease the HMGRi-score. Considering the relative character of the R-S notation, the general interpretation of the CTI contribution should be avoided and reduced only to specific molecules. In contrast with other QSAR models, with easier structural explanation,⁷⁰ we prefer to consider the structural interpretation of the present (TI + CTI)-QSAR models with caution and showing it simply as a prediction instead of an explicative tool.

The parametrical assumptions such as the normality, homoscedasticity (homogeneity of variances) and non-colinearity have the same importance in the application of multivariate statistic

Table 3
Summary of the Forward stepwise analysis

TIs	F ^a	p-level ^b	State ^c
W	35.48685	0.0001	In
Cc	5.38917	0.0214	In
Nrc	76.91498	0.0001	In
Sovd	25.07916	0.0001	In
D	12.11721	0.0006	In
(Nrc – Nncc)/Ntc	34.76868	0.0001	In
R	0.01094	0.9168	Out
Sha	0.40046	0.5277	Out
Shc	0.09681	0.7561	Out
Sod	0.98024	0.3235	Out
Mti	1.15043	0.2849	Out
Nrb	0.13234	0.7164	Out
Tc	1.27195	0.2609	Out
Tvc	1.47973	0.2254	Out
Inv	0.17810	0.6735	Out
Ncc	0.27370	0.6015	Out
Nncc	0.36751	0.5451	Out
(Nrc-Nsc)/(Ncc + 1)	0.92655	0.3370	Out
Nsc	0.27370	0.6015	Out
Ntc	0.20965	0.6476	Out
Nncc-Ncc	0.53599	0.4650	Out
J	1.54234	0.2159	Out
PsA	0.03872	0.8442	Out

^a Fisher ration.

^b Probability level of error for F.

^c In/out denote the included/not-included variables in the model.

techniques to QSAR^{71,72} as the correct specification of the mathematical form has. The validity and statistical signification of any model is conditioned by the above-mentioned factors.

The simple linear mathematical form of the model in our work was chosen in the absence of prior information. Figure 2 shows that the training cases against the residuals did not present any characteristic pattern.⁷³

A better threshold for the a priori classification probability can be estimated by means of the receiver operating characteristics (ROC) curve.⁷⁴ Figure 3 shows that this is not a random model, but a statistically significant one, since the area under the ROC curve (both for training = 0.93 and validation = 0.83) is significantly higher than the area under the random classifier curve equal to 0.5 (diagonal line).⁷⁵

The validity of the LDA models depends on the normal distribution of the sample used as well as the homogeneity of their variances. Thus, we carry out two significant tests of normality:

Chi-Square and Kolmogorov–Smirnov tests. We found statistically significant differences ($p < 0.01$) on the respective values (Chi-Square, d). This result rejects the hypothesis of normal distribution of the sample in study (Fig. 4).⁷²

The heteroscedasticity of a large set can be studied using a simple graphical method based on the examination of the variable residuals included in the model. Figure 5 shows that the LDA model variables (against the residuals) do not present any pattern. This indicates that homocedasticity assumption is fulfilled.⁷²

Due to the robustness of the LDA multivariate statistical techniques, the predictive ability and interference reached by using the proposed model should not be affected (see Fig. 6).

The part including TIs of the QSAR can be used for the score prediction of a given potential HMGRI without taking into consideration chirality; however, if we also use the part including CTIs of the model, we can clearly differentiate between stereo-isomer derivatives and the same basic scaffold. In many of these cases, it is difficult to obtain all the possible isomers of the same compounds by organic synthesis; as a result, many of them have not been assayed biologically yet.⁷⁶ Consequently, the present QSAR model becomes an invaluable tool to provide the first virtual exploration of the molecular diversity of these compounds related to HMGRI. These predictions may be useful to guide the rational synthesis of the most promising candidates among the high number of potential candidates to be assayed. Specifically, if we used the model, we should select those compounds with a higher TIs-based value and (TI + CTI)-based HMGRI-scores of the isomer of interest, but a low value should be selected for the other isomers. In Figure 7, we illustrate this fact for a large series of 1628 chiral derivatives of known HMGRIs. The details on the configuration of the compound and the chiral and non-chiral QSAR results are given in Table SM3 from the Supplementary Material.

2.2. QSAR-based complex networks

Supposing we have experimentally demonstrated that one chiral isomer of a compound has a good HMGRI activity, but it is faced with problems of toxicity or relatively difficult synthesis. In this kind of situation, it is highly interesting the prediction of other similar isomers of the same compound (or other close to it). This can help to lead the synthesis to those compounds with similar activity but probably different toxicity and/or synthesis accessibility. In more theoretical terms, it may be also useful to answer the question about how diverse the universe of chiral isomers of

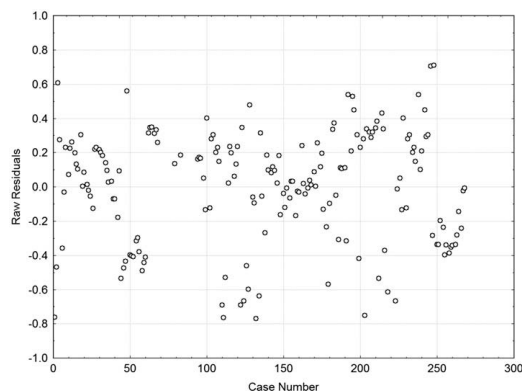


Figure 2. Training cases against the residuals.

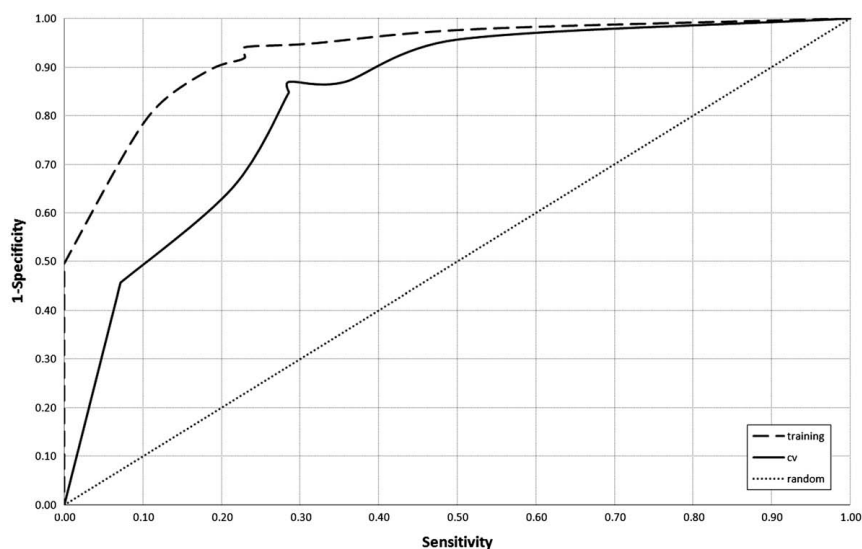


Figure 3. ROC curve for the HMGRl/non-HMGRl model.

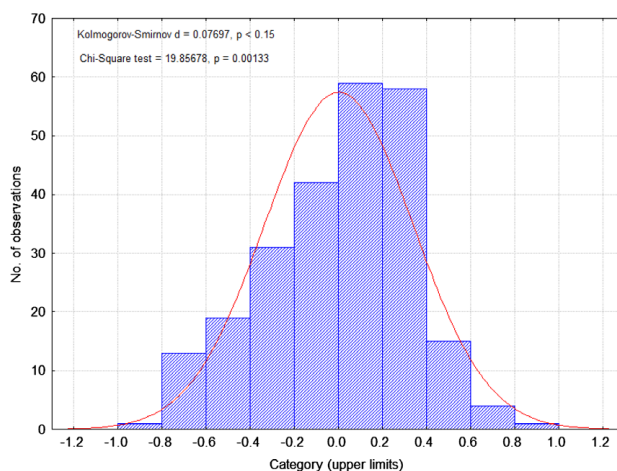


Figure 4. Distribution for LDA model residuals, Chi-Square and Kolmogorov-Smirnov tests.

HMGRIs is, in terms of both structure and activity. In order to solve this problem, we constructed a drug–drug CNs and determined their complexity. The use of CNs for the investigation of data structure in this kind of problems by using large databases is nowadays an emerging and active field of research.⁷⁷ This method will help to decide whether we should dedicate or not material resources to investigate specific families of chiral isomers. After calculating the two classes of DDD_{ij} distances ($cDDD_{ij}$ and $ncDDD_{ij}$) between all pairs of similar isomers (PSIs), we arranged them in two 1628×1628 matrices (one that measured chirality and another which did not). Next, in order to calculate the Boolean matrices

and to construct two drug–drug interactions CNs for HMGRIs, we explored different values of cut-off ranging, from a cut-off DDD_{ij} of 0.0001 to 35 dimensionless units of distance (Table 4).

We found that, for extremely low values of the cut-off, we obtain many unconnected nodes (HMGRl chiral isomers not similar to any other); for very large values, all PSIs are linked to each other in the CN. Both extremes are useless because we need a CN that, for a given HMGRl chiral compound (the input), could point out only few of the possible similar candidates to be assayed or obtained (the output). At the same time, in practical terms, the CN must be able to predict all possible input molecules (not many uncon-

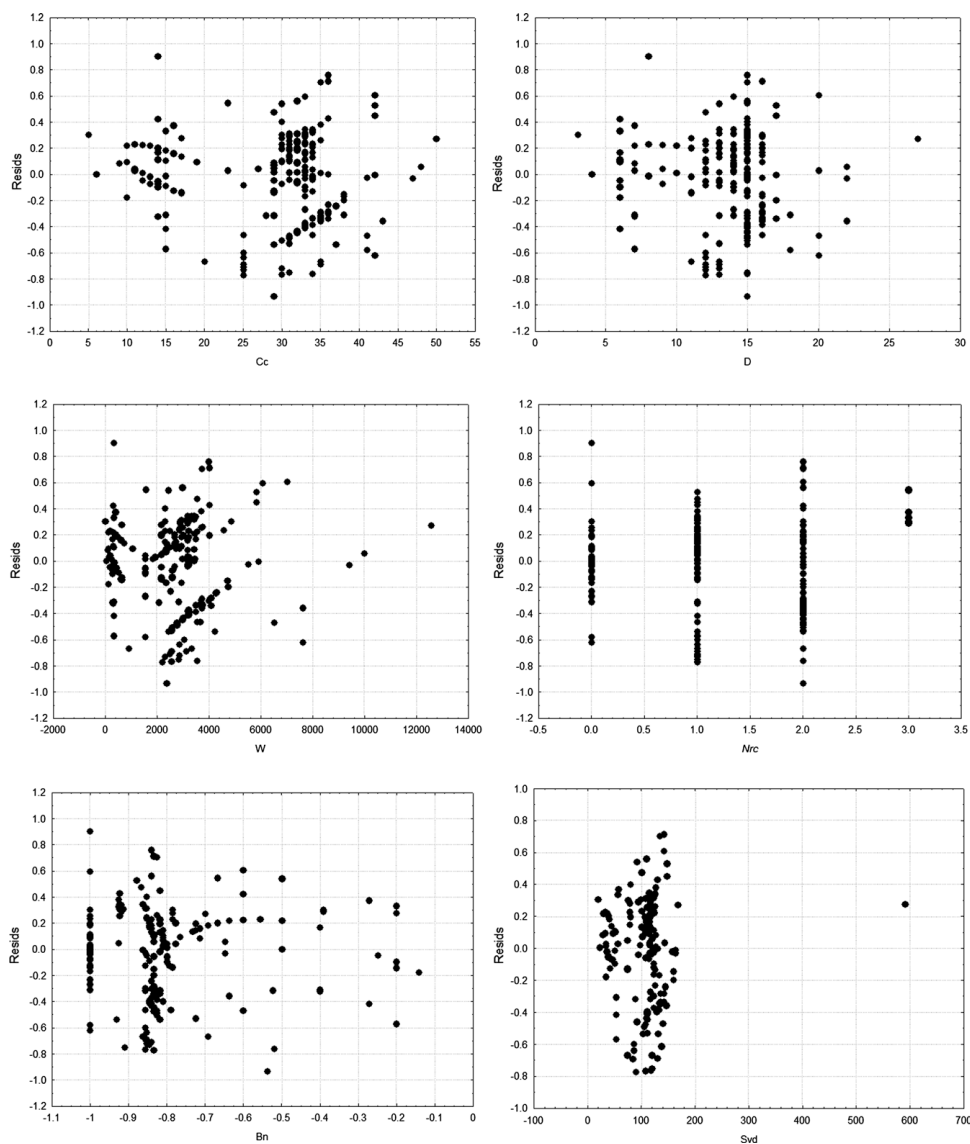


Figure 5. Graphical analysis of homogeneity of variances (variables vs residuals).

nected nodes). We calculated the probabilities of one node to be unconnected $p(u) = (\text{number of unconnected nodes})/1628$ and the probability $p(d) = (\text{average node degree})/1628 = z/1628$ in case all nodes have the largest degree (connecting all-with-all) by using different cut-off values (Fig. 8). The parameter z (average node degree) represents the average number of HMGRIs, similar to one compound in the CN. A possible CN could be one with a 0.5 cut-off, having $p(u) = 0.007$ and $p(z) = 0.049$. It means that, of a total

of 1628 chiral compounds for a given input candidate, we can select a series of $z = 79.5$ similar compounds and the CN is able to predict only 12 out of the 1628 compounds. However, for a more narrow search, we selected the CN determined by the cut-off equal to 0.1, having $p(u)$ equal to 0.143 and $p(z)$ equal to 0.011.

We also studied the effect of chirality on the CN construction with respect to CNs derived without considering the chirality. The first CN was derived using both the chiral and non-chiral parts

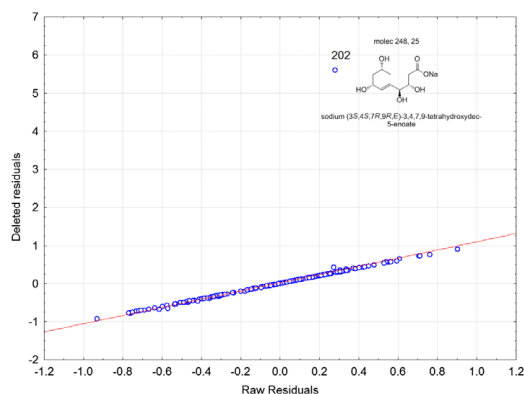


Figure 6. Residuals versus deleted residuals plot for the LDA model.

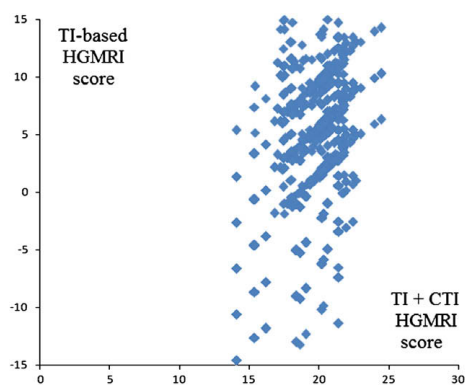


Figure 7. Chiral sensitive versus non-chiral in silico screening of chiral HMGRI.

of the QSAR model in the distance function ($cDDD_{ij}$), whereas the second one ($ncDDD_{ij}$) was based only on the non-chiral part (see Section 4). Figure 8 illustrates also the behavior of the ncCN in terms of $p(z)$, $p(u)$ and cut-off. In general, we observed that the ncCN reaches higher values of $p(z)$ for lower cut-off values compared with the cCN. This can be explained by the fact that all common chiral isomers have different structural parameters in the cCN and consequently different $p(z)$; these isomers do not always connect at the same cut-off. In contrast, all chiral isomers of the same compound in ncCN become degenerated nodes, due to the fact they have identical $p(z)$ values. Thus, it is not necessary to extend the cut-off scanning. In practice, we could condense all this nodes with identical behavior in only one node. We can conclude that the omission of the chirality determines a transformation of the cCN into a ncCN, that leads us to node condensation or grouping, similar to other examples of complex networks.⁷⁸ The ncCN and cCN (cut-off = 0.1) are described in Table SM4 and Table SM5 (from the Supplementary Material) and plotted by Figures 9 and 10.

The most important result is the reduction of the large universe of HMGRI isomers (1628 compounds) to a basic set of compounds in order to obtain the best candidates for the experimental assays. CentiBin has been used to reduce the CN by extracting the Giant Component (GC)⁷⁹ of the chiral compound CN with a cut-off equal

Table 4
Cut-off scanning

TIs and CTIs QSAR		Cut-off		TIs QSAR				
$p(z)$	$p(u)$	z	u	u	z	$p(u)$	$p(z)$	
0.000	1.000	0.0	1628	0.0001	1628	0.0	1.000	0.000
0.000	0.851	0.8	1385	0.001	1533	0.3	0.942	0.000
0.001	0.654	1.5	1064	0.005	976	3.2	0.600	0.002
0.001	0.520	2.4	847	0.01	808	5.6	0.496	0.003
0.006	0.210	9.2	342	0.05	448	25.4	0.275	0.016
0.011	0.143	17.5	233	0.1	340	46.9	0.209	0.029
0.049	0.007	79.1	12	0.5	128	232.7	0.079	0.143
0.059	0.007	95.5	12	0.6	128	288.3	0.079	0.177
0.070	0.007	114.1	12	0.7	64	322.8	0.039	0.198
0.080	0.007	130.8	12	0.8	64	375.8	0.039	0.231
0.093	0.006	152.0	9	0.9	64	420.8	0.039	0.258
0.103	0.006	167.9	9	1	64	448.5	0.039	0.276
0.111	0.006	180.0	9	1.1	64	510.1	0.039	0.313
0.120	0.006	195.7	9	1.2	64	546.7	0.039	0.336
0.132	0.001	214.5	2	1.3	64	615.7	0.039	0.378
0.139	0.001	226.2	2	1.4	0	653.3	0.000	0.401
0.149	0.001	242.1	2	1.5	0	677.3	0.000	0.416
0.196	0.001	319.5	1	2	0	826.4	0.000	0.508
0.287	0.001	467.7	1	3	0	1084.2	0.000	0.666
0.387	0.001	630.6	1	4	0	1268.5	0.000	0.779
0.470	0.001	764.5	1	5	0	1382.2	0.000	0.849
0.762	0.001	1241.2	1	10	0	1561.9	0.000	0.959
0.913	0.001	1485.9	1	15	0	1562.2	0.000	0.960
0.973	0.001	1584.4	1	20	0	1562.2	0.000	0.960
1.000	0.000	1628.0	0	25	0	1562.2	0.000	0.960
0.994	0.000	1617.6	0	30	0	1562.2	0.000	0.960
0.994	0.000	1618	0	35	0	1562.2	0.000	0.960

to 0.1 ($cGC_{0.1}$). The $cGC_{0.1}$ contains 479 nodes (isomers) and 7074 edges (PSIs); the density of PSIs is only 0.62%. In $cGC_{0.1}$, the average topological distance (Dt) for all PSIs is 14.96 and the diameter (Di) is 49.5 (largest Dt). This CN is very important for practical use because it contains all the HMGRI chiral isomers with the most important degree of similarity in all their 1628 compounds. A component is a subset of nodes in the CN where each node is reachable from the others by some path through the CN. At first glance, $cGC_{0.1}$ seems to be a very complex and highly interconnected network, but the visualization, after processing with Kamada–Kawai algorithms (Fig. 10), demonstrates that this component simplifies the information and could be of use as guide into the future synthesis effort. In fact, $cGC_{0.1}$ contains a more representative set of compounds from which synthesis candidates can be selected. In Table 5 we compare the results obtained in this work with other CNs⁷⁹, previously reported.

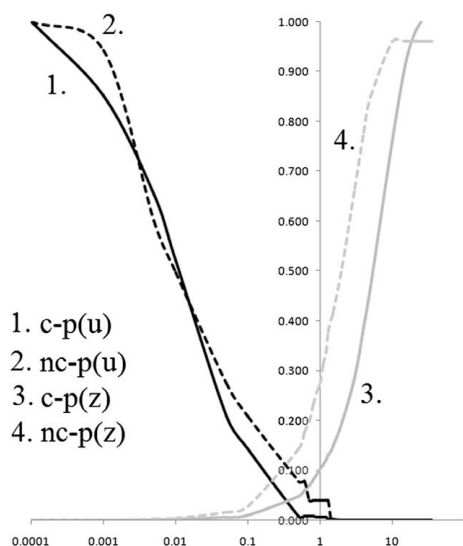


Figure 8. Cut-off scanning for TIs and CTIs CN (curves 2 and 3) or TIs only CN (2 and 4).

3. Conclusions

This work proposes, for the first time a QSAR model based on topological indices and chiral topological indices of a large heterogeneous series of HMGRIs. This model is used for the prediction of the HMGR activity of new chiral isomers. In addition, the comparison between the chiral isomers was carried out using the chiral and non-chiral complex networks. These networks were constructed based on the previous QSAR predictions. An advantage of this method is the possibility to explore the activity profile of a large number of chiral compounds that are not experimentally characterized. Another advantage is the use of this QSAR model

as an important guide of synthetic experiments by providing the best candidates and, consequently, lowering the costs.

4. Methods

4.1. Biochemical assays for evaluation of compound activity

The biological activities of the components used in the QSAR model were known from previous experimental results. Some examples are the pyrimidine- and pyrrole-substituted, that were evaluated for their ability to inhibit partially purified rat liver HMG-CoA reductase in vitro.⁸⁰ Inhibition of rat liver HMG-CoA reductase was measured by an in vitro assay procedure based on the direct conversion of *or*-3-hydroxy-3-methyl [3-¹⁴C]-glutaryl-CoA to [¹⁴C]mevalonolactone. The enzyme preparation and assay procedures used in this study were the same as those described in Ref. 80. The HMG-CoA reductase inhibitory effect was examined through preparation of rat liver microsomes from Sprague-Dawley rats, which had been allowed free access to ordinary diets containing 2% cholestyramine and water for 2 weeks. The microsomes were purified as described in Ref. 80. Details about the assays of each compound can be found in the references included in the QSAR database (see Table SM1 from the Supplementary Material).

4.2. QSAR model

A database from literature^{12,81–94} containing in vitro and/or in vivo assayed HMGRIs was used (Table SM1 from the Supplementary Material). ChemDraw Ultra 10.0 was employed to draw the compounds (Table SM2, Supplementary Material) and Chem3D Ultra 10.0 to calculate the TIs (Table SM1, Supplementary Material). The resulted TIs are used as numerical parameters that describe the molecular structure in the statistical analysis. In Table 1 we give the complete list of symbols, names and some features of these indices. The QSAR model was constructed with the multivariate regression technique, the Linear Discriminant Analysis (LDA), employing the *Forward stepwise* method for the selection of variables. All statistical analyses and data exploration were carried out in STATISTICA 6.0.⁹⁵ In the actual work, the independent data test is used by splitting the data randomly in a training series used for a model construction and a cross-validation (cv) one. The calculation of the CTIs was implemented in a Microsoft Excel sheet.⁹⁶

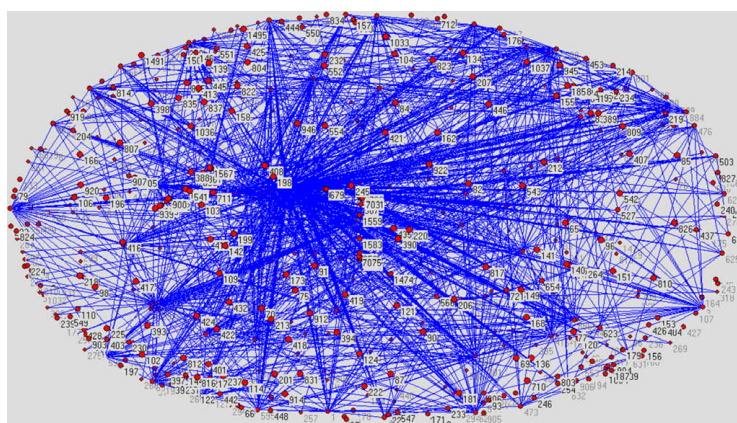
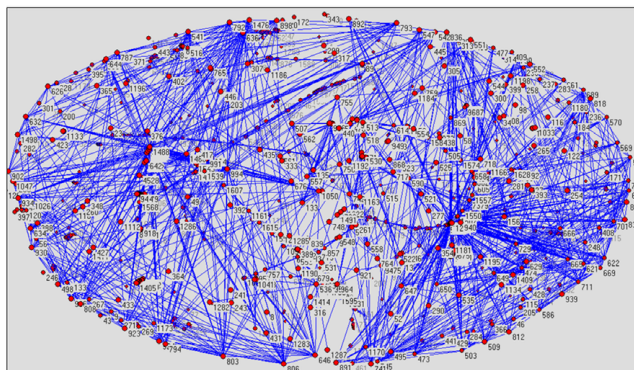
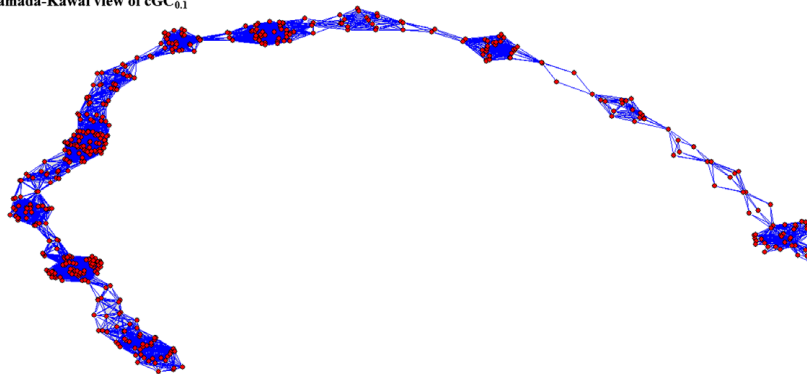


Figure 9. Different layouts for cGC₀₁.

Fruchterman Reingold view of cCN

Kamada-Kawai view of cGC_{0.1}

*output from Pajek

Figure 10.

The general formula of the QSAR classification function is the following:

$$\text{HMGRI-score} = \sum_m w_m \cdot {}^m\text{TI} + \sum_n w_n \cdot {}^n\text{CTI} + w_0 \quad (2)$$

where HMGRI-score is the continue and dimensionless score value for the HMGRI/non-HMGRI classification that gives relatively higher values to molecules with more probability to act as HMGRIs. ${}^m\text{TI}_i$ and ${}^n\text{CTI}_i$ are the TIs and the CTIs of type n and m , w_m and w_n are the coefficient (weights) of these indices in the QSAR model and w_0 is the independent term.

The reported statistical parameters of the QSAR model are the following: N , R_c , λ , F , and p -level as well as Sensitivity, Specificity, and Accuracy for both training and cv.⁹⁵ N is the number of molecules used to train the model, R_c is the canonical regression coefficient, λ is Wilks statistic parameter, F is Fisher ratio and p -level is the probability of error. The R_c , similarly to other regression coefficients, indicates the strength of the correlation between the inputs (TIs) and the output of the model (HMGRI-score). The predictions made with the QSAR model were used to construct a complex network.

4.3. Complex network construction

We generated a series with all the possible stereo-isomers of the above-mentioned molecules. This series of compound was used to explore the molecular diversity of chiral HMGRIs with the CN. We did not use these compounds for the QSAR model fitting or validation because they were generated in the computer, but almost none of them have been experimentally explored yet. The construction of the CNs is given by the following steps:

- First, we predicted the value of Drug–Drug Distance (DDD_{ij}) for all possible pairs of compounds of this third database by substituting the TIs values in the QSAR model. We ordered these values and formed a distance matrix of n rows by n column. ${}^m\text{TI}_i$ and ${}^m\text{TI}_j$ are the TIs of type m for the drugs i and j . ${}^n\text{CTI}_i$ and ${}^n\text{CTI}_j$ represent the CTIs of type n for the drugs i and j . w_m and w_n are the coefficient (weights) of these indices in the QSAR model, then we calculated two DDD_{ij} as a Manhattan distance⁹⁷ with QSAR-weighted components. One type of DDD_{ij} considers only TIs and the other includes both TI and CTI components. We named the first as the non-chiral distance (ncDDD_{ij}) and the other one as the chiral distance (cDDD_{ij}):

Table 5
Behavior of z and C for different real versus random CNS

CNS ^a	n	z	C_{measured} ^b	C_{random} ^b
<i>Drug–drug activity similarity networks</i>				
1 HMGRI _s cGC _{0.1}	479	—	—	—
2 HMGRI _s nGC	—	—	—	—
3 Anti-fungal	59	22	0.37	0.37
4 Anti-parasites	380	3.34	0.01	0.01
5 <i>S. cerevisiae</i> mutants	24	—	—	—
6 ATC classification	1014	—	—	—
<i>Biological networks</i>				
7 Metabolic network	315	28.3	0.59	0.09
8 Neural network	282	14.0	0.28	0.049
9 Food web	134	8.7	0.22	0.065
<i>Social networks</i>				
10 Biology collaborations	1 520 251	15.5	0.081	0.00001
11 Mathematics collaborations	253 339	3.9	0.15	0.00002
12 Company directors	7 673	14.4	0.59	0.002
13 Word co-occurrence	460 902	70.1	0.44	0.0002
14 Film-actor collaborations	449 913	113.4	0.2	0.0003
<i>Technological networks</i>				
15 WWW (sites)	153 127	35.2	0.11	0.0002
16 Internet	6 374	3.8	0.24	0.0006
17 Power grid	4 941	2.7	0.08	0.0005

^a CNS 1 and 2 have been reported here by the first time; CN3—see González-Díaz, H., et al. *J. Comput. Chem.* **2008**, *29*, 656; CN4—Prado-Prado, F. J., et al. *Bioorg. Med. Chem.* **2008**, *16*(11), 5871.

^b These are clustering coefficients determined as the ratio $C = z/n$ for the network measured versus random network with the same number of nodes.

$$ncDDD_{ij} = \left| \sum_m w_m \cdot ({}^m T_{i_j} - {}^m T_{i_j}) \right| \quad (3a)$$

$$cDDD_{ij} = \left| \sum_m w_m \cdot ({}^m T_{i_j} - {}^m T_{i_j}) + \sum_n w_n \cdot ({}^n CT_{i_j} - {}^n CT_{i_j}) \right| \quad (3b)$$

- Next, we transformed the distance matrix into a Boolean matrix, assigning a value of 1 to all very similar pairs of drugs and 0 otherwise. The drugs are similar if the DDD_{ij} is lower than a pre-determined cut-off value equal to 0.1.
- We saved the Boolean matrix in a *.mat* format and used it as input for the Pajek application, (a free software for CNS analysis). The same software was used to visualize the CN.
- In the last step, we used the module for analysis of distribution function fit from STATISTICA 6.0⁹⁵ in order to investigate the normality of the node degrees (number of similar molecules for each molecule) within the space of chiral derivatives of HMGRI_s.

Acknowledgements

We acknowledge the useful comments and kind attention of the editor Prof. Herbert Waldmann and two unknown reviewers. The authors thank the Portuguese Fundação para a Ciência e a Tecnologia (FCT) (SFRH/BPD/24997/2005). The corresponding author González-Díaz H. acknowledges contract/grant sponsorship for a research position funded by the Program Isidro Parga Pondal of the 'Xunta de Galicia'.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2008.11.007.

References and notes

- Shimokata, K.; Yamada, Y.; Kondo, T.; Ichihara, S.; Izawa, H.; Nagata, K.; Murohara, T.; Ohno, M.; Yokota, M. *Atherosclerosis* **2004**, *172*, 167.
- Smilde, T. J.; van Wissen, S.; Wollersheim, H.; Kastelein, J. J.; Stalenhoef, A. F. *Neth. J. Med.* **2001**, *59*, 184.
- Leitersdorf, E.; Eisenberg, S.; Eliav, O.; Friedlander, Y.; Berkman, N.; Dann, E. J.; Landsberger, D.; Sehayek, E.; Meiner, V.; Wurm, M., et al. *Circulation* **1993**, *87*, 1135.
- Salazar, L. A.; Hirata, M. H.; Quintao, E. C.; Hirata, R. D. *J. Clin. Lab. Anal.* **2000**, *14*, 125.
- Suzuki, M.; Iwasaki, H.; Fujikawa, Y.; Sakashita, M.; Kitahara, M.; Sakoda, R. *Bioorg. Med. Chem.* **2001**, *9*, 2727.
- Urbina, J. A.; Lazard, K.; Marchan, E.; Visbal, G.; Aguirre, T.; Piras, M. M.; Piras, R.; Maldonado, R. A.; Payares, G.; de Souza, W. *Antimicrob. Agents Chemother.* **1993**, *37*, 580.
- Sit, S. Y.; Parker, R. A.; Motoc, I.; Han, W.; Balasubramanian, N.; Catt, J. D.; Brown, P. J.; Harte, W. E.; Thompson, M. D.; Wright, J. J. *J. Med. Chem.* **1990**, *33*, 2982.
- Choulier, L.; Andersson, K.; Hamalainen, M. D.; van Regenmortel, M. H.; Malmqvist, M.; Altschuh, D. *Protein Eng.* **2002**, *15*, 373.
- Du, Q. S.; Huang, R. B.; Wei, Y. T.; Du, L. Q.; Chou, K. C. *J. Comput. Chem.* **2008**, *29*, 211.
- Li, Y.; Wei, D. Q.; Gao, W. N.; Gao, H.; Liu, B. N.; Huang, C. J.; Xu, W. R.; Liu, D. K.; Chen, H. F.; Chou, K. C. *Med. Chem. (Shariqah, United Arab Emirates)* **2007**, *3*, 576.
- Sirois, S.; Tsoukas, C. M.; Chou, K. C.; Wei, D.; Boucher, C.; Hatzakis, G. E. *Med. Chem. (Shariqah, United Arab Emirates)* **2005**, *1*, 173.
- Thilagavathi, R.; Kumar, R.; Aparna, V.; Sobhia, M. E.; Gopalakrishnan, B.; Chakraborti, A. K. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1027.
- Prabhakar, Y. S. *Drug Des. Discov.* **1992**, *9*, 145.
- Prabhakar, Y. S.; Saxena, A. K.; Doss, M. J. *Drug Des. Deliv.* **1989**, *4*, 97.
- González-Díaz, H.; Vilar, S.; Santana, L.; Uriarte, E. *Curr. Top. Med. Chem.* **2007**, *7*, 1015.
- Gozalbes, R.; Doucet, J. P.; Derouin, F. *Curr. Drug Targets Infect. Disord.* **2002**, *2*, 93.
- Estrada, E.; Uriarte, E. *Curr. Med. Chem.* **2001**, *8*, 1573.
- González-Díaz, H.; González-Díaz, Y.; Santana, L.; Ubeira, F. M.; Uriarte, E. *Proteomics* **2008**, *8*, 750.
- González-Díaz, H.; Prado-Prado, F. J.; Santana, L.; Uriarte, E. *Bioorg. Med. Chem.* **2006**, *14*, 5973.
- Helguera, A. M.; Rodriguez-Borges, J. E.; Garcia-Mera, X.; Fernandez, F.; Cordeiro, M. N. *J. Med. Chem.* **2007**, *50*, 1537.
- Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castanedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Sanchez, A. M.; Torrens, F.; Castro, E. A. *Bioorg. Med. Chem.* **2005**, *13*, 1005.
- González-Díaz, H.; Olazabal, E.; Santana, L.; Uriarte, E.; Castañedo, N. *Bioorg. Med. Chem.* **2007**, *15*, 962.
- Murcia-Soler, M.; Perez-Gimenez, F.; Garcia-March, F. J.; Salabert-Salvador, M. T.; Diaz-Villanueva, W.; Medina-Casamayor, P. *J. Mol. Graph. Model.* **2003**, *21*, 375.
- Prado-Prado, F.; González-Díaz, H.; Santana, L.; Uriarte, E. *Bioorg. Med. Chem.* **2007**, *15*, 897.
- Mathur, K. C.; Gupta, S.; Khadikar, P. V. *Bioorg. Med. Chem.* **2003**, *11*, 1915.
- Calabuig, C.; Anton-Fos, G. M.; Galvez, J.; Garcia-Domenech, R. *Int. J. Pharm.* **2004**, *278*, 111.
- González, M. P.; González-Díaz, H.; Molina Ruiz, R.; Cabrera, M. A.; Ramos de Armas, R. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1192.
- Votano, J. R.; Parham, M.; Hall, L. H.; Kier, L. B.; Oloff, S.; Tropsha, A.; Xie, Q.; Tong, W. *Mutagenesis* **2004**, *19*, 365.
- Santana, L.; Uriarte, E.; González-Díaz, H.; Zagotto, G.; Soto-Otero, R.; Mendez-Alvarez, E. *J. Med. Chem.* **2006**, *49*, 1149.
- Marrero-Ponce, Y. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2010.
- Vilar, S.; Santana, L.; Uriarte, E. *J. Med. Chem.* **2006**, *49*, 1118.
- Marrero-Ponce, Y.; Khan, M. T.; Casanola-Martin, G. M.; Ather, A.; Sultankhodzhaev, M. N.; Torrens, F.; Rotondo, R. *ChemMedChem* **2007**, *2*, 449.
- Casanola-Martin, G. M.; Marrero-Ponce, Y.; Khan, M. T.; Ather, A.; Sultan, S.; Torrens, F.; Rotondo, R. *Bioorg. Med. Chem.* **2007**, *15*, 1483.
- Casanola-Martin, G. M.; Marrero-Ponce, Y.; Khan, M. T.; Ather, A.; Khan, K. M.; Torrens, F.; Rotondo, R. *Eur. J. Med. Chem.* **2007**, *42*, 1370.
- González-Díaz, H.; Sanchez, I. H.; Uriarte, E.; Santana, L. *Comput. Biol. Chem.* **2003**, *27*, 217.
- Marrero-Ponce, Y.; Castillo-Garit, J. A. *J. Comput. Aided Mol. Des.* **2005**, *19*, 369.
- Fabian, W. M.; Stampfer, W.; Mazur, M.; Uray, G. *Chirality* **2003**, *15*, 271.
- Golbraikh, A.; Tropsha, A. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 144.
- Golbraikh, A.; Bonchev, D.; Tropsha, A. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 147.
- Castillo-Garit, J. A.; Marrero-Ponce, Y.; Torrens, F. *Bioorg. Med. Chem.* **2006**, *14*, 2398.
- de Julián-Ortiz, J. V.; de Gregorio Alapont, C.; Ríos-Santamarina, I.; García-Doménech, R.; Gálvez, J. *J. Mol. Graph. Model.* **1998**, *16*, 14.
- Castillo-Garit, J. A.; Marrero-Ponce, Y.; Torrens, F.; Rotondo, R. *J. Mol. Graph. Model.* **2007**, *26*, 32.
- Zhang, Z.; Grigorov, M. G. *Proteins* **2006**, *62*, 470.
- Bonchev, D.; Buck, G. A. *J. Chem. Inf. Model.* **2007**, *47*, 909.
- Ivanciuc, O.; Ivanciuc, T.; Klein, D. J. *SAR QSAR Environ. Res.* **2001**, *12*, 1.

46. Ivanciuc, O.; Ivanciuc, T.; Klein, D. J.; Seitz, W. A.; Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 536.
47. Ivanciuc, O. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1412.
48. Gupta, N.; Mangal, N.; Biswas, S. *Proteins* **2005**, *59*, 196.
49. Gan, H. H.; Pasquali, S.; Schlick, T. *Nucleic Acids Res.* **2003**, *31*, 2926.
50. Marrero-Ponce, Y.; Nodarse, D.; González-Díaz, H.; Ramos de Armas, R.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. A. *Int. J. Mol. Sci.* **2004**, *5*, 276.
51. González-Díaz, H.; Agüero-Chapin, G.; Varona, J.; Molina, R.; Delogu, G.; Santana, L.; Uriarte, E.; Gianni, P. *J. Comput. Chem.* **2007**, *28*, 1049.
52. Rose, A.; Schraegle, S. J.; Stahlberg, E. A.; Meier, I. *BMC Evol. Biol.* **2005**, *5*, 66.
53. Estrada, E.; Rodríguez-Velazquez, J. A. *Phys. Rev. E* **2005**, *71*, 056103.
54. Yu, X.; Lin, J.; Masuda, T.; Esumi, N.; Zack, D. J.; Qian, J. *Nucleic Acids Res.* **2006**, *34*, 917.
55. Guido, N. J.; Wang, X.; Adalsteinsson, D.; McMillen, D.; Hasty, J.; Cantor, C. R.; Elston, T. C.; Collins, J. *J. Nature* **2006**, *439*, 856.
56. Williams, R. J.; Berlow, E. L.; Dunne, J. A.; Barabasi, A. L.; Martinez, N. D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12913.
57. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Nodarse, D. *Bioorg. Med. Chem.* **2005**, *13*, 3397.
58. González-Díaz, H.; de Armas, R. R.; Molina, R. *Bioinformatics* **2003**, *19*, 2079.
59. Vullo, A.; Frasconi, P. *J. Bioinform. Comput. Biol.* **2003**, *1*, 411.
60. Lange, B. M.; Ghassemian, M. *Phytochemistry* **2005**, *66*, 413.
61. Chou, K. C.; Cai, Y. D. *J. Proteome Res.* **2006**, *5*, 316.
62. Yu, X.; Lin, J.; Zack, D. J.; Qian, J. *Nucleic Acids Res.* **2006**, *34*, 4925.
63. Margolin, A. A.; Nemenman, I.; Basso, K.; Wiggins, C.; Stolovitzky, G.; Dalla Favera, R.; Califano, A. *BMC Bioinformatics* **2006**, *7*, 57.
64. Estrada, E. *J. Proteome Res.* **2006**, *5*, 2177.
65. Newman, M. E. *Phys. Rev. E: Stat. Nonlin. Soft Matter Phys.* **2006**, *74*, 036104.
66. Kleinberg, J. M. *Nature* **2000**, *406*, 845.
67. González-Díaz, H.; Prado-Prado, F. *J. Comput. Chem.* **2008**, *29*, 656.
68. Hill, T.; Lewicki, P. *STATISTICS Methods and Applications*; Tulsa: StatSoft, 2006.
69. Lilien, R. H.; Farid, H.; Donald, B. R. *J. Comput. Biol.* **2003**, *10*, 925.
70. Vilar, S.; Estrada, E.; Uriarte, E.; Santana, L.; Gutierrez, Y. *J. Chem. Inf. Model.* **2005**, *45*, 502.
71. Bisquerra Alzina, R. *Introducción conceptual al análisis multivariante: Un enfoque informático con los paquetes SPSS-X, BMDP, LISREL y SPAD*; Barcelona: PPU, 1989.
72. Stewart, J.; Gill, L. *Econometrics*; Prentice Hall: London, 1998.
73. Dillon, W. R.; Goldstein, M. *Multivariate Analysis: Methods and Applications*; Wiley: N.Y., 1984.
74. James, A.; Hanley, B. *J. M. Radiology* **1982**, *143*, 29.
75. Helguera, A. M.; Rodríguez-Borges, J. E.; García-Mera, X.; Fernández, F.; Cordeiro, M. N. *J. med. chem.* **2007**, *50*, 1537.
76. Gómez, G.; Rivera, H.; García, I.; Estévez, L.; Fall, Y. *Tetrahedron Lett.* **2005**, *46*, 5819.
77. Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M.; Hwang, D. U. *Phys. Rep.* **2006**, *424*, 175.
78. Bianconi, G.; Barabasi, A. L. *Phys. Rev. Lett.* **2001**, *86*, 5632.
79. Bornholdt, S.; Schuster, H. G. *Handbook of Graphs and Complex Networks*; Wiley-VCH GmbH & Co. KGa: Weinheim, 2003.
80. Kuroda, M.; Endo, A. *Biochim. Biophys. Acta* **1977**, *486*, 70.
81. Turabi, N.; DiPietro, R. A.; Mantha, S.; Ciosek, C.; Rich, L.; Tu, J. I. *Bioorg. Med. Chem.* **1995**, *3*, 1479.
82. Watanabe, M.; Koike, H.; Ishiba, T.; Okada, T.; Seo, S.; Hirai, K. *Bioorg. Med. Chem.* **1997**, *5*, 437.
83. Romo, D.; Harrison, P. H.; Jenkins, S. I.; Riddoch, R. W.; Park, K.; Yang, H. W.; Zhao, C.; Wright, G. D. *Bioorg. Med. Chem.* **1998**, *6*, 1255.
84. Colle, S.; Taillefumier, C.; Chapleur, Y.; Liebl, R.; Schmidt, A. *Bioorg. Med. Chem.* **1999**, *7*, 1049.
85. Suzuki, M.; Iwasaki, H.; Fujikawa, Y.; Kitahara, M.; Sakashita, M.; Sakoda, R. *Bioorg. Med. Chem.* **2001**, *9*, 2727.
86. Taillefumiera, C.; Fornelb, D.; Chapleur, Y. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 615.
87. Alali, F.; Zeng, L.; Zhang, Y.; Ye, Q.; Hopp, D. C.; Schwedler, J. T.; McLaughlin, J. L. *Bioorg. Med. Chem.* **1997**, *5*, 549.
88. Schroepfer, G. J., Jr.; Parish, E. J.; Chen, H. W.; Kandutsch, A. A. *J. Biol. Chem.* **1977**, *252*, 8975.
89. Beck, G.; Kessler, K.; Baader, E.; Bartmann, W.; Bergmann, A.; Granzer, E.; Jendralla, H.; von Kerekjarto, B.; Krause, R.; Paulus, E., et al. *J. Med. Chem.* **1990**, *33*, 52.
90. Zhang, Q. Y.; Wan, J.; Xu, X.; Yang, G. F.; Ren, Y. L.; Liu, J. J.; Wang, H.; Guo, Y. *J. Comb. Chem.* **2007**, *9*, 131.
91. Grabley, S.; Granzer, E.; Hutter, K.; Ludwig, D.; Mayer, M.; Thiericke, R.; Till, G.; Wink, J.; Philipps, S.; Zeeck, A. *J. Antibiot. (Tokyo)* **1992**, *45*, 56.
92. Gohrt, A.; Zeeck, A.; Hutter, K.; Kirsch, R.; Kluge, H.; Thiericke, R. *J. Antibiot. (Tokyo)* **1992**, *45*, 66.
93. Grabley, S.; Hammann, P.; Hutter, K.; Kirsch, R.; Kluge, H.; Thiericke, R.; Mayer, M.; Zeeck, A. *J. Antibiot. (Tokyo)* **1992**, *45*, 1176.
94. Chan, C.; Bailey, E. J.; Hartley, C. D.; Hayman, D. F.; Hutson, J. L.; Inglis, G. G. A.; Jones, P. S.; Keeling, S. E.; Kirk, B. E.; Lamont, R. B.; Lester, M. G.; Pritchard, J. M.; Ross, B. C.; Scicinski, J. J.; Spooner, S. J.; Smith, G.; Steeples, I. P.; Watson, N. S. *J. Med. Chem.* **1993**, *36*, 3646.
95. StatSoft, Inc., 2002.
96. Microsoft Corp., Microsoft Excel, 2002.
97. Zhang, W. *Environ. Monit. Assess.* **2007**, *124*, 253.

Theoretical study of GSK-3 α : Neural Networks QSAR studies for the design of new inhibitors using 2D-descriptors

Isela García ^{*}, Yagamare Fall, Xerardo García-Mera
Francisco Prado-Prado

Abstract GSK-3 targets encompass proteins implicated in AD, neurological disorders. The functions of GSK-3 and its implication in various human diseases have triggered an active search for potent and selective GSK-3 inhibitors. In this sense, QSAR could play an important role in studying these GSK-3 inhibitors. For this reason we developed QSAR models for GSK-3 α , LDA and ANNs from nearly 50000 cases with more than 700 different GSK-3 α inhibitors obtained from ChEMBL database server; in total we used more than 20000 different molecules to develop the QSAR models. The model correctly classified 237 out of 275 active compounds (86.2%) and 14870 out of 15970 non-active compounds (93.2%) in the training series. The overall training performance was 93.0%. Validation of the model was carried out using an external predicting series. In these series the model classified correctly 458 out of 549 (83.4%) compounds and 29637 out of 31927 non-active compounds (83.4%). The overall predictability performance was 92.7%. In this work, we propose three types of non Linear ANN and we show that it is another alternative model to the already existing ones in the literature, such as LDA. The best model obtained was Linear Neural Network: LNN: 236:236-1-1:1 which had an overall training performance of 96%. In addition, we did a study of different fragments that exist in the molecules of the database in order to see which fragments had more influence in the activity. All this can help to design new inhibitors of GSK-3 α . The present work reports the attempts to calculate within a unified framework probabilities of GSK-3 α inhibitors against different molecules found in the literature.

Keywords: GSK-3 α , QSAR, Artificial Neural Network, Linear Neural Network, Fragment contribution, Linear Discriminant Analysis.

Introduction

Alzheimer's disease (AD) [1] is a serious and degenerative disorder that causes a gradual loss of neurons, and in spite of the efforts realized by the big pharmaceutical companies of the world, the origin of this pathology is still not very clear. Glycogen synthase kinase-3 (GSK-3) is a serine-threonine kinase encoded by two isoforms in mammals, termed GSK-3 α and GSK-3 β [2]. GSK-3 targets encompass proteins implicated in AD, neurological

disorders, in the *wnt* and insulin signaling pathway, glycogen and protein synthesis, regulation of transcription factors [3], embryonic development, cell proliferation and adhesion, tumorigenesis, apoptosis [4], circadian rhythm,... etc. GSK-3 β knock-out mice die *in utero* [5], whereas GSK-3 α knock-out mice are viable and display improved glucose tolerance in response to glucose load and elevated hepatic glycogen storage and insulin sensitivity [6, 7]. The functions of GSK-3 and its implication in various human diseases have triggered an active search for potent and selective GSK-3 inhibitors [8] in the last years. In this sense, quantitative structure-activity relationships (QSAR) could play an important role in studying these GSK-3 inhibitors (GSKI-3 α); QSARs can be used as predictive tools for the development of molecules [9, 10]. Computer-aided drug design techniques based on QSAR could play an important role in drug discovery programs. The QSAR approach involves the development of models that relate the structure of drugs with to their biological activity against different targets [11, 12]. In principle, there are currently more than 1600 molecular descriptors that may be generalized and used to solve the problem outlined above [13, 14]. Many of these indices are known as Dragon Descriptors (DDs). DRAGON provides 1664 molecular descriptors divided into several families: 0D (constitutional descriptors), 1D (e.g., functional group counts), 2D (e.g., topological descriptors and connectivity indices), and 3D (e.g., GETAWAY, WHIM, RDF and 3D-MoRSE descriptors) [15-17]. Numerous different molecular descriptors have been reported to encode chemical structures in QSAR studies. Furthermore, there are multiple chemometric approaches that can, in principle, be selected for this step. Multiple linear regression (MLR), linear discriminant analysis (LDA) [18], partial least squares (PLS) and different kinds of artificial neural networks (ANN) can be used to relate molecular structure (represented by molecular descriptors) with to biological properties. The ANNs are particularly useful in QSAR studies in which the linear models fit poorly due to high data complexity^[15, 18], an example was the work of Prado-Prado *et al.* in which four types of non Linear Artificial neural networks (ANN) were developed to calculate within a unified framework probabilities of antiparasitic action of drugs against different parasite species [19-21].

There are several different kinds of ANN and these include multilayer perceptron (MLP), radial basis functions (RBF) and PNNs; the latter ANN is a variant of RBF systems. In particular, PNN is a type of neural network that uses a kernel-based approximation to form an estimate of the probability density functions of classes in a classification problem [22].

In this article, we developed QSAR models for GSK-3 α , LDA and ANNs from more than 48000. This database contains 700

I. García Y. Fall G. Gómez
Department of Organic Chemistry, University of Vigo, Vigo, Spain
e-mail: iselapintos@yahoo.es
X. García-Mera F. Prado-Prado
Department of Organic Chemistry, Faculty of Pharmacy,
USC, 15782 Santiago de Compostela, Spain

different molecules inhibitors of GSK-3 α (active against GSK-3 α); in total we used more than 20000 different molecules non-active or not have interaction with GSK-3 α to develop a different QSAR models. All the tested compounds were compiled from ChEMBL database <http://www.ebi.ac.uk/chembl/db/index.php/target/browser/classification> [23, 24].

First we used 237 molecular descriptors calculated with DRAGON software [16]. Next, we developed LDA and ANNs models in order to find new compounds that present inhibitor action against GSK-3 α and hence might be used as new treatments for neurological pathologies such as Parkinson's disease (PD) or AD. Last, to probe if our QSAR models are consistent, we established a Molecular Fragments study; we used different fragments from the known molecules of the database in order to see which fragments had more influence in the activity, and which fragments interacted more with the GSK-3 α protein. The design of new inhibitors of this enzyme is very important for the study of neurodegenerative diseases [25, 26].

The present is the first work reports the attempts to calculate within an unified framework probabilities of GSK-3 α inhibitors against different molecules using a tool server ChEMBL database.

Methods

Linear classifier

A database from ChEMBL database [23] containing assayed GSK-3 α inhibitors was used (**Table SM** from the Supplementary Material). The DRAGON software 4.0 [16] was utilized here and provides 1664 descriptors classified as zero- (0D) one- (1D), two- (2D) and three-dimensional (3D) descriptors depending on whether they are computed from the chemical formula, substructure list representation, molecular graph or geometrical representation of the molecule, respectively [27]. In this work, we calculated the following descriptors: 2D autocorrelations, Burden eigenvalues, topological charge indices, eigenvalue-based indices, functional group counts, atoms-centred fragments, charge descriptors and molecular properties. The QSAR model was constructed with the multivariate regression technique, the LDA, employing the Forward stepwise method for the selection of variables. All statistical analyses and data exploration were carried out in STATISTICA 6.0 [28]. In the actual work, the independent data test is used by splitting the data randomly in a training series used for a model construction and a cross-validation (CV) one. The general formula of the QSAR classification function is the following:

$$GSKI-3\alpha_{score} = \sum W_m \cdot {}^m 2D_i + W_0 \quad (1)$$

where GSKI-3 α_{score} is the continuous and dimensionless score value for the GSKI-3 α /non-GSKI-3 α classification that gives relatively higher values to molecules with more probability to act as GSKI-3 α , ${}^m 2D_i$ are the 2Ds of type m , W_m is the coefficient (weights) of these indices in the QSAR model and W_0 is the independent term.

The reported statistical parameters of the QSAR model are the following: N, χ^2 and p-level as well as Sensitivity, Specificity, and Accuracy for both training and CV [28]. N is the number of molecules used to train the model, λ is Wilks statistic parameter, is Chi-square and p-level is the probability of error.

Nonlinear classifier

We processed our data with different ANNs using the STATISTICA 6.0 software [28] looking for a better model to predict activity against GSK-3 α . Three types of ANNs were used, namely, Radial Basis Function (RBF) [29], Multi Layers Perceptron (MLP) and Linear Neural Network (LNN). The profile of a ANN is: Ni:I-H1-H2-O:No. It means that we have inputs variables (Ni), neurons in the input layer (I), neurons in the first hidden layer (H1), in the second hidden layer (H2), neuron in the output layer (O) and output variable (No).

We can used a very simple type of ANN called Linear Neural Network (LNN) to fit this discriminant function. The model deals with the classification of a compound set with or without affinity on different receptors. A dummy variable Affinity Class (AC) was used as input to codify the affinity. This variable indicates either high (AC = 1) or low (AC = 0) affinity of the drug by the receptor. S(DTP)_{pred} or DTP affinity predicted score is the output of the model and it is a continuous dimensionless score that sorts compounds from low to high affinity to the target coinciding DTPs with higher values of S(DTP)_{pred} and nDTPs with lowest values. In **Equation 2**, b represents the coefficients of the LNN classification function, determined by the ANN module of the STATISTICA 6.0 software package [28]. We used Forward Stepwise algorithm for a variable selection.

Let be ${}^k \chi(G)$ drugs molecular descriptors and ${}^k \xi(R)$ receptor or drug target descriptors for different drugs (d) with different receptor; we can attempt to develop a simple linear classifier of mt-QSAR type with the general formula:

$$S(DTP)_{pred} = \sum_{k=0}^5 b(G_k) \cdot \chi(G) + \sum_{k=0}^5 b(R_k) \cdot \xi(R) + b \quad (2)$$

We assessed the quality of models with different statistical parameters like Specificity (see **Equation 2**), Sensitivity (see **Equation 3**), Accuracy (see **Equation 4**) and ROC curve (Receiver Operating Characteristic curve) which is a [graphical](#) plot of the [sensitivity](#), or true positives, vs. (1-[specificity](#)), or false positives,

$$specificity = \frac{NTN}{NTN + NFP} \quad (3)$$

$$sensitivity = \frac{NTP}{NTP + NFN} \quad (4)$$

$$accuracy = \frac{NTP + NTN}{NTP + FN + FP + TN} \quad (5)$$

where NTN means number of true negatives, NFP is number of false positives, NTP is number of true positives, NFN is number of false negatives, FN is false negatives, FP is false positives and TN is true negatives.

Study of molecular fragments

In this work we calculated contributions of different molecular fragments for activity against 15 different molecules inhibitors of *GSK-3 α* . In so doing we gave the following steps [11, 21]:

- First, we calculated the specie-dependent atomic descriptors included in the QSAR equation for selected molecular fragments using DRAGON software [16].
- Second, we calculated the contribution scores of each fragment against 15 species of molecules studied by substituting the atomic descriptors into the QSAR equation using the Microsoft Excell application.
- Third, the contributions of each molecular fragment were standardized dividing each value by the sum of all contributions for each molecule. These molecular fragment contributions can indicate the potential relation between molecular fragments with the activity against *GSK-3 α* , each separated fragment or each fragment inside a molecule.
- Fourth, contributions for each atom of the drug were scaled into a percentage value.

- Fifth, scaled atom contributions were grouped into different molecular fragments.
- Sixth, molecular fragment contributions to the biological activity were back-projected onto the molecular structure for obtaining a colour-scaled biological structure-activity.

Data set

We developed QSAR models for *GSK-3 α* , LDA and ANNs from more than 48000. This database contains 700 different molecules inhibitors of *GSK-3 α* (active against *GSK-3 α*); in total we used more than 20000 different molecules non-active or not have interaction with *GSK-3 α* to develop a different QSAR models. All the compounds were compiled from ChEMBL database

<http://www.ebi.ac.uk/chembl/db/index.php/target/browser/classification> [23, 24]. This is a database of bioactive drug-like small molecules, which contains 2-D structures, calculated properties (e.g. logP, Molecular Weight, Lipinski Parameters, etc.) and abstracted bioactivities (e.g. binding constants, pharmacology and ADMET data). ChEMBL normalises the bioactivities into a uniform set of end-points and units where possible, and also tags the links between a molecular target and a published assay with a set of varying confidence levels. The data is abstracted and curated from the primary scientific literature, and covers a significant fraction of the structure activity relationship (SAR) and discovery of modern drugs. The codes and activity for all compounds as well as the references used to collect them are depicted in **Table SM** of the supplementary material file.

Results and Discussion

LDA

In this paper we obtained a LDA study, **Equation 6**, and we can observe that eighteen variables entry inside equation:

$$\begin{aligned} GSK-3\alpha_{act} = & -34.3 \cdot AT51m + 19.3 \cdot AT53m - 14.3 \cdot AT54e - 2.9 \cdot AT56e + 37.0 \cdot AT51p + 39.1 \cdot MAT53e \\ & - 43.3 \cdot MAT53p - 3.7 \cdot GAT53e - 4.0 \cdot GAT52e - 46.5 \cdot BELm3 - 34.6 \cdot BELm4 + 6.8 \cdot BELe8 + 41.0 \cdot BELp3 \\ & + 36.0 \cdot BELp4 + 1.1 \cdot GGI3 - 227.2 \cdot JGI4 + 163.4 \cdot VEm2 + 40.7 \end{aligned} \quad (6)$$

$N = 48721 \quad \chi^2 = 3127.3 \quad p\text{-level} < 0.001$

The nomenclature used in the descriptors of the equation is the same as establishing the Dragon software, where N is the number of cases, χ^2 is the Chi-square and p is the level of error. The model correctly classified 237 out of 275 active compounds (86.2%) and 14870 out of 15970 non-active

compounds (93.2%) in the training series. The overall training performance was 93.0%. Validation of the model was carried out using an external predicting series. In this series the model classified correctly 458 out of 549 (83.4%) antiparasitic compounds and 29637 out of 31927 non-active compounds (92.8%). The overall predictability performance was 92.7% (see **Table 1**).

Table 1 comes about here

ANN models

The ANN models are non-linear models useful to predict the biological activity of a large datasets of molecules. This technique is an alternative to linear methods such as LDA [30, 31]. **Figure 1** depicts the networks maps for some of the ANN models. In general, at least one ANN of every types tested was statically significant. However, one must note that the profiles of each network indicate that these are highly nonlinear and complicated models [32-34].

Figure 1 comes about here

In **Figure 2**, we depict the ROC-curve [35, 36] for LNN tested. Notably, the ROC curve can also be represented equivalently by plotting the fraction of true positives (sensitivity) vs. the fraction of false positives (1-specificity). The vitality of this type of procedures developing ANN-QSAR models has been demonstrated before [37]; see, for instance, the work of Fernandez and Caballero [38]. The same is true about the ANNs tested in this work, we illustrated a ROC-curve for the best ANN model, in this case was a LNN (236:236-1:1) which results of ROC curve values (AUC) were with an area higher than 0.98. It indicates that the present model gives statistically significant results and clearly different from those obtained with a random classifier (area = 0.5) [39]. To shows how important is this result, we compared the present model with other model used to address the same problem. We processed our data with ANNs looking for a better model [31].

The network found was LNN and it showed training performance higher than 96%. The summary of results is showed in **Table 1**. After direct inspection of the results reported in **Table 1** for ANN methods, we can conclude that a complex ANN method is a good method to predict the activity. We compare different types of networks to obtain a better model; **Table 1** shows the classification matrix of the different

networks. LNN 236:236-1:1 was taken as the main network because it presented a wider range of variables, 236 inputs in the first layer and 236 neurons in second layer, and two sets of cases (Training and Validation). Another tested networks found were MLP 22:22-27-1:1, RBF 98:98-740-1:1 presented low accuracy and PNN 237:237-16760-2-2:1 had a very low percentage of DTPs leading to possible errors in the model although its accuracy was very good, see **Table 1**. We depict the ROC-curve for LNN 236:236-1:1 to show how reliable was the network model developed, see **Figure 2**.

Figure 2 comes about here

Study of molecular fragments (F)

One application of QSAR is the calculation of the contribution of different molecular fragments to the desired activity [40, 41]. In this sense, one important application of QSAR models is the calculation of molecular fragments contribution to activities or action against different drug targets [11]. Before, to get the different QSAR models we calculated the contribution of different molecular fragments from the better model obtained in this work. The LDA model was better than ANN models obtained, because the LDA model presented 18 variables to obtained one result (active or non-active GSK-3 α), while the best LNN model was presented 236 variables to obtained the same results. In spite of percentage of good classification the LNN model presents the highest classified percentage, but LNN model needs more variables to predict one result than LDA model. This LDA model developed in this work is a single equation that uses few parameters to predict the inhibitory action of a new compound against GSK-3 α . For this reason, to obtain the best and the most reliable results, we calculated fragments of the molecules using the LDA method. As a result, we selected different molecular fragments against 15 molecules of database; we selected these molecules at random. In **Figure 3**, we observed contribution of all fragments for the 15 molecules selected at random whose results in our model are as follow: green colour indicates major contribution to the activity, yellow colour indicates medium contribution and red colour indicates minor contribution to the biological activity against GSK-3 α . An example is the Dasatinib see Louise N. Johnson [42], where a pair of hydrogen bonds is formed in the hinge region of the ATP-binding site (i.e. between the 3-nitrogen of the aminothiazole ring of dasatinib and the amide nitrogen

of Met318 and between the 2-amino hydrogen of dasatinib with the carbonyl oxygen of Met318). A hydrogen bond is also formed between the side chain hydroxyl oxygen of Thr315 and the amide nitrogen of dasatinib.

Figure 3 comes about here

Conclusion

The functions of GSK-3 and its implication in various human diseases have triggered an active search for potent and selective GSK-3 inhibitors. Nowadays theoretical studies such as QSAR models have become a very useful tool in this context to substantially reduce time and resources consuming experiments. In this work we developed a new LDA model using the Dragon descriptors, using a large data base using about 20000 different drugs obtained from the ChemBL server. We conclude that a large database gives a much more precise model; the use of tools such as ChemBL database enables us to develop models with large data bases, and this helps us make the results more reliable. To improve the model we developed non-linear models and compared them to LDA. We proposed non-linear models, and for the first time, we proposed ANN models based on Dragon Descriptors series of GSK-3 α , and we concluded that they are alternative methods to study the activity of different families of molecules compared with other methods found in the literature. The use of tools such as fragment contributions can help us design the best GSK-3 α inhibitors for their later synthesis in the laboratory, eliminating the synthesis of molecules at random which may have few possibilities of being active.

Acknowledgments

Prado-Prado F. thanks sponsorships for research position at the University of Santiago de Compostela from Angeles Alvarino, Xunta de Galicia. All authors acknowledge the Project 07CSA008203PR. We are grateful to the Xunta de Galicia (INCITE08PXIB314255PR) for partial financial support.

References

- [1] Hübscher U, Maga G, Spadari S. Eukaryotic DNA polymerases. *Annu Rev Biochem.* 2002;71:133-63.
- [2] Benek K, Kunkel TA. Functions of DNA polymerases. *Adv Protein Chem.* 2004;69:137-65.
- [3] Kornberg A, Baker TA. DNA replication. New York 1992.
- [4] Takahashi S, Yonezawa Y, Kubota K, Ogawa N, Maeda K, Koshino H, et al. Pyranicin, a non-classical annonaceous acetogenin, is a potent inhibitor of DNA polymerase, topoisomerase and human cancer cell growth. *International Journal of Oncology.* 2008;32:451-8.
- [5] Loeb LA, Agarwal SS, Dube DK, Gopinathan KP, Travaglini EC, Seal G, et al. Inhibitors of mammalian DNA polymerases: possible chemotherapeutic approaches. *Pharmacol Ther, Part A.* 1977;2:171-93.
- [6] Miura S, Izuta S. DNA polymerases as targets of anti-cancer nucleotides. *Curr Drug Targets.* 2004;5:191-5.
- [7] Mizushima Y, Kasai N, Iijima H, Sugawara F, Yoshida H, Sakaguchi TK. Sulfo-quinovosyl-acyl-glycerol (SQAG), a eukaryotic DNA polymerase inhibitor and anticancer agent. *Curr Med Chem: Anti-Cancer Agents.* 2005;5:613-25.
- [8] Albertella MR, Lau A, O'Connor MJ. The overexpression of specialized DNA polymerases in cancer. *DNA Repair.* 2005;4:583-92.
- [9] Nakamura R, Takeuchi R, Kuramochi K, Mizushima Y, Ishimaru C, Takakusagi Y, et al. Chemical properties of fatty acid derivatives as inhibitors of DNA polymerases. *Organic & Biomolecular Chemistry.* 2007;5:3912-21.
- [10] So AG, Downey KM. Eukaryotic DNA replication. *Crit Rev Biochem Mol Biol.* 1992; 27:129-55.
- [11] Mizushima Y, Saito A, Tanaka A, Nakajima N, Kuriyama I, Takemura M, et al. Structural analysis of catechin derivatives as mammalian DNA polymerase inhibitors. *Biochem Biophys Res Commun.* 2005;333:101-9.
- [12] Colot V, Rossignol JL. Eukaryotic DNA methylation as an evolutionary device. *BioEssays.* 21:402-11.
- [13] Jean B, Margot M, Cristina C, Heinrich L. Mammalian DNA Methyltransferases Show Different Subnuclear Distributions. *J Cell Biochem.* 2001;83:373-9.
- [14] Sun NJ, Ho Woo S, Cassady JM, Snapka RM. DNA Polymerase and Topoisomerase II Inhibitors from *Psoralea corylifolia*. *J Nat Prod.* 1998;61:362-6.
- [15] Nunez MB, Maguna FP, Okulik NB, Castro EA. QSAR modeling of the MAO inhibitory activity of xanthenes derivatives. *Bioorg Med Chem Lett.* 2004 Nov 15;14(22):5611-7.
- [16] Todeschini R, Consonni V. *Handbook of Molecular Descriptors.* 2000.
- [17] Freund JA, Poschel T. Stochastic processes in physics, chemistry, and biology. *Lect Notes Phys.* Berlin, Germany: Springer-Verlag 2000.
- [18] Estrada E, Uriarte E. Recent advances on the role of topological indices in drug discovery research. *Curr Med Chem.* 2001;8:1573-88.
- [19] Estrada E, Uriarte E, Montero A, Teijeira M, Santana L, De Clercq E. A Novel Approach for the Virtual Screening and Rational Design of Anticancer Compounds. *J Med Chem.* 2001;43:1975-85.
- [20] Verma RP, Kurup A, Hansch C. On the role of polarizability in QSAR. *Bioorg Med Chem.* 2005 Jan 3;13(1):237-55.
- [21] WU RS, Wolpert-DeFilippes MK, Quinn FR. Quantitative Structure-Activity Correlations of Rifamycins as Inhibitors of Viral RNA-Directed DNA Polymerase and Mammalian α and β DNA Polymerases. *J Med Chem.* 1980;23(3):256-61.
- [22] Quinn FR, Driscoll JS, Hansch C. Structure-Activity Correlations among Rifamycin B Amides and Hydrazides. *J Med Chem.* 1975;18(4):332-9.

- [23] Wright GE, Gambino JJ. Quantitative Structure-Activity Relationships of 6-Anilinoouracils as Inhibitors of *Bacillus subtilis* DNA Polymerase III. *J Med Chem.* 1984;27:181-5.
- [24] Gass KB, Cozzarelli NR. Further Genetic and Enzymological Characterization of the Three *Bacillus subtilis* Deoxyribonucleic Acid Polymerases. *J Biol Chem.* 1973;248:7688-700.
- [25] Clements J, D'Ambrosio J, Brown NC. Inhibition of *Bacillus subtilis* deoxyribonucleic acid polymerase III by phenylhydrazinopyrimidines. Demonstration of a drug-induced deoxyribonucleic acid-enzyme complex. *J Biol Chem.* 1975;250:522-6.
- [26] Chakraborty AK, Majumder HK. Mode of action of pentavalent antimonials: specific inhibition of type I DNA topoisomerase of *Leishmania donovani*. *Biochem Biophys Res Commun.* 1988;152:605-11.
- [27] Liu LF. DNA topoisomerase poisons as antitumor drugs. *Annu Rev Biomed.* 1989;58:351-75.
- [28] Ray S, Hazra B, Mitra B, Das A, Majumder HK. Diospyrin, a bisnaphthoquinone: a novel inhibitor of type I DNA topoisomerase of *Leishmania donovani*. *Mol Pharmacol.* 1998;54:994-9.
- [29] Sakaguchi K, Sugawara F, Mizushima Y. Inhibitors of eukaryotic DNA polymerase. *Seikagaku.* 2002;74:244-51.
- [30] Talete srl, ed. DRAGON for Windows (Software for Molecular Descriptor Calculations). Version 5.3 ed 2005.
- [31] StatSoft.Inc. STATISTICA (data analysis software system), version 6.0, www.statsoft.com. Statsoft, Inc. 6.0 ed 2002.

Received:

Revised:

Accepted:

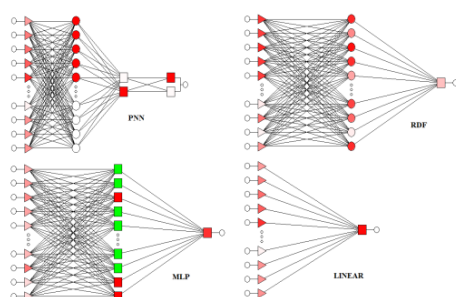


Figure 1

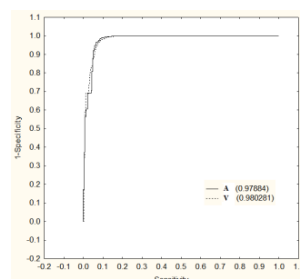


Figure 2

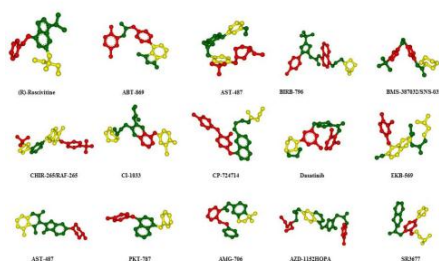


Figure 3

Table 1

Model profile	Train			Stat. Par.	Validation		
	Active	Non-Active	%		%	Active	Non-Active
LDA	237	38	86.2	Sn	83.4	458	91
	1100	14870	93.2	Sp	92.8	2290	29637
			93.0	Ac	92.7		
LNN 236:236-1:1	258	17	93.8	Sn	93.3	513	37
	860	15625	94.8	Sp	94.4	1834	31146
			96.2	Ac	95.6		
MLP 22:22-27-1:1	97	178	35.3	Sn	34.6	190	360
	11038	5447	33.0	Sp	33.3	22004	10976
			33.1	Ac	33.3		
RBF 98:98-740-1:1	210	65	76.4	Sn	72.4	398	152
	4713	11772	71.4	Sp	71.4	9441	23539
			71.5	Ac	71.4		

First computational chemistry multi-target model for anti-Alzheimer, anti-parasitic, anti-fungi, and anti-bacterial activity of GSK-3 inhibitors in vitro, in vivo, and in different cellular lines

Isela García · Yagamare Fall · Generosa Gómez · Humberto González-Díaz

Received: 5 March 2010 / Accepted: 13 September 2010
© Springer Science+Business Media B.V. 2010

Abstract In the work described here, we developed the first multi-target quantitative structure–activity relationship (QSAR) model able to predict the results of 42 different experimental tests for GSK-3 inhibitors with heterogeneous structural patterns. GSK-3 β inhibitors are interesting candidates for developing anti-Alzheimer compounds. GSK-3 β are also of interest as anti-parasitic compounds active against *Plasmodium falciparum*, *Trypanosoma brucei*, and *Leishmania donovani*; the causative agents for Malaria, African Trypanosomiasis and Leishmaniasis. The MARCH-INSIDE technique was used to quickly calculate total and local polarizability, *n*-octanol/water partition coefficients, refractivity, van der Waals area and electronegativity values to 4,508 active/non-active compounds as well as the average values of these indexes for active compounds in 42 different biological assays. Both the individual molecular descriptors and the average values for each test were used as input for a linear discriminant analysis (LDA). We discovered a classification function which used in training series correctly classifies 873 out of 1,218 GSK-3 cases of inhibitors (97.4%) and 2,140 out of 2,163 cases of non-active compounds (86.1%) in the 42 different tests. In addition, the model correctly classifies 285 out of 406 GSK-3 inhibitors (96.3%) and 710 out of 721 cases of non-active compounds (85.4%) in external validation series. The result is important because, for the

Electronic supplementary material The online version of this article (doi:10.1007/s11030-010-9280-3) contains supplementary material, which is available to authorized users.

I. García (✉) · Y. Fall · G. Gómez
Department of Organic Chemistry, University of Vigo, Vigo, Spain
e-mail: iselapintos@yahoo.es

H. González-Díaz
Department of Microbiology and Parasitology, Faculty of Pharmacy, USC, 15782 Santiago de Compostela, Spain

first time, we can use a single equation to predict the results of heterogeneous series of organic compounds in 42 different experimental tests instead of developing, validating, and using 42 different QSAR models. Lastly, a double ordinate Cartesian plot of cross-validated residuals (first ordinate), standard residuals (second ordinate), and leverages (abscissa) defined the domain of applicability of the model as a squared area within ± 2 band for residuals and a leverage threshold of $h = 0.0044$.

Keywords Glycogen synthase kinase-3 (GSK-3) inhibitors · Alzheimer's disease · Malaria · *Plasmodium falciparum* · Trypanosomiasis · *Trypanosoma brucei* · QSAR · Markov Model · Linear discriminant analysis

Introduction

The most common cause of dementia in elderly people is Alzheimer's disease (AD) [1]. This serious and degenerative disorder results in the gradual loss of neurons. A more paradigmatic fact is, in spite of the efforts realized by part of the big pharmaceutical companies and of the world governments, there is no credible explanation for Alzheimer's pathology. The fundamental characteristic of the AD is the presence of two principal types of brain pathology: the neurofibrillary tangles, formed by the hyperphosphorylated Tau protein and the senile plaques formed by the aggregation of the β -amiloide peptide.

In the mammals, the enzyme glycogen synthase kinase-3 (GSK-3) has two isoforms GSK-3 α and GSK-3 β [2] and studies reveal their physiological role in the cellular differentiation. The two isoforms show similar biochemistry and properties. According to some studies, this enzyme controls a wide number of substrates including cytoplasmic

proteins and nuclear transcription factors. The GSK-3 enzyme also involved in neurological disorders (e.g., *wnt* and insulin signaling pathway, glycogen and protein synthesis, regulation of transcription factors, embryonic development, cell proliferation and adhesion, tumorigenesis, apoptosis, circadian rhythm, etc.), hence the importance of finding powerful inhibitors for this enzyme [3,4].

Under normal conditions, the Tau protein stabilizes the microtubules which form the cytoskeleton, but in pathological conditions this protein is hyperphosphorylated at many points by means of the GSK-3 enzyme, provoking the microtubules loss of form; in addition, it is important in the cellular processes, metabolic control, embryogenesis, and oncogenesis, and is related to a wide range of neurodegenerative diseases, bipolar mood disorders and diabetes, hence is nowadays one of the main research targets.

In 1988, Ishiguro et al. [5] while studying brain extract, isolated one enzyme displaying the generation of paired helical filaments of Tau protein kinase, an injury typical of Alzheimer's disease. A more in depth study reveal the existence of two kinases, TPKI and TPKII. The discovery that the structures were identical to GSK-3 β enzyme came as a surprise.

In addition to the study of new GSK-3 inhibitors, researching other alternatives is also important. One promising area is the study of kinases that come from unicellular parasites. One of the most studied is the *Plasmodium falciparum*, which genome contains 65 genes that encode kinases as the GSK-3 α and β ; *P. falciparum* exports PfGSK-3 and though its function is unknown, it is believed that PfCK1 and PfGSK-3 influence the regulation of the strong circadian rhythm of the parasites causing circadian fevers, when both are in the infected red blood cell.

Trypanosomiasis, caused by the *Trypanosoma brucei*, is another critical illness which affects about 300,000–500,000 humans in sub-Saharan Africa. If left untreated, the disease is fatal; however, recovery is possible, though difficult, due to the diseases resistance to available drugs, lengthy treatment periods, and the adverse effects of conventional therapies [6].

One of the problems is the therapy of pregnant women and infants, because the GSK-3 regulates the critical protein of the development such as the *wnt* gene product [7], hence the importance of selecting well the objective having the three-dimensional structure in mind between mammalian and parasite GSK-3 enzyme understanding well the differences of the substrate binding properties, because is more representative of the reality [8].

It is necessary to bear in mind that in all these reported cases (e.g. parasites, fungi, etc.), it is not taken into account the influence of the compounds over the GSK-3 enzyme, which is one of the main objectives of this work.

The development of new quantitative structure–activity relationship (QSAR) methods, with simple molecular indexes, is a promising alternative to drug–protein docking,

high-throughput screening, and combinatorial chemistry techniques, because all of them employ molecular descriptors that allows the relation between the biological and statistical properties [9].

Literature is full of cases where QSAR are used as alternative methods to random synthesis, and where molecular descriptors are used [10,11]. Recently multi-target studies have been published which include microbes or studies of the enzyme against the receptor, though none of them used enzymes, as in our case [12]. Indeed, experience demonstrates the use of models fitted with large data sets of chemicals works as well as the use of models built from a series of homologous compounds, and can be applied in a broad spectrum of cases. The principal problem of some molecular descriptors is that they lack of physical meaning. For this reason, the introduction of novel molecular indices must obey physicochemical laws in order to ensure a theoretically rigorous interpretation of the results.

On the other hand, our research group has introduced a novel series of stochastic indices in the so called MARCH-INSIDE approach. The method is based on the use of Markov Models (MM) to codify useful chemical structure information in terms of molecular electron delocalization, polarizability, refractivity, and *n*-octanol/water partition coefficient [13]. In this work, we explore the potential of MARCH-INSIDE to seek a QSAR for GSK-3 inhibitors from a heterogeneous series of compounds. In the first step, the aforementioned molecular descriptors were calculated for a large series of active/non-active compounds. LDA was subsequently used to obtain a classification function. The QSAR developed was then validated with an external predicting series.

Materials and methods

Computational methods

The MARCH-INSIDE approach [14] is based on the calculation of the averages of the different physicochemical molecular properties (ap). For instance, it is possible to derive average estimations of molecular descriptors or group indices [15]: electro-negativities ${}^k\chi(G)$, refractivities ${}^kMR(G)$, polarizabilities ${}^k\alpha(G)$, van der waals area ${}^kA_{vdW}(G)$ and logarithms of water/*n*-octanol partition coefficients $\log{}^kP$. In a compact notation, we write ${}^kD_t(G)$ where D_t is the type of descriptor. $D_t = \chi, MR, \alpha, A_{vdW}$ and $\log{}^kP$; $G = \text{total}, C_{\text{sat}}, C_{\text{inst}}, \text{Hetero}$ and H_x (see details in Supplementary material 1).

Multi-target linear discriminant analysis

Linear discriminant analysis was used to construct the classifiers. One of the most important steps in this work was the

organization of the spreadsheet containing the raw data used as input for the LDA because this is not a classic classifier. Herein, the diagram of the paper is peculiar. We expected to use a two-group discriminant function to assign compounds into two possible groups: compounds which belong to a particular group and compounds which do not belong to a particular group. With this in mind, we needed to indicate what group we want to predict. To accomplish this, we made the following steps, which are essentially the same given by Concu et al. [16] for the QSAR study of six classes of enzymes:

- (1) We produced raw data representing each compound input as a vector made up of 1 output variable, 450 structural variables (inputs) divided in values (refer to the first term of the Eq. 9), averages (refer to the second term of the Eq. 9) and differences between values and averages (refer to the third term of the Eq. 9); and the compound assay conditions query (CAC_q) variable. CAC_q is an auxiliary not used to construct the model.
- (2) The first element (output) is a dummy variable (Boolean) called observed group (OG); OG = 0 if the compound belongs to the class to which we refer in CAC_q and 1 otherwise (OG = 1). We could repeat each compound more than once in the raw data. In fact, we could repeat each compound 46 times corresponding to 42 CAC_q assay conditions (see Table 1). The first time we used the CAC_q = CAC number. It means that we used the real CAC class of the compound that it is the CAC_q of this compound. In this case, the LDA model had to give the highest probability to the group OG = 0 because it had to predict the real class of the compound. The remnant 46 times we use an CAC class number different to the real in CAC_q and then the LDA model had to predict the highest probability for the group OG = 1. This indicated that the compound did not belong to this group.
- (3) The χ , MR, α , A_{vdW} and log^kP values cannot be constants because the model fail when we change OG values. This can create an inconvenience when we wanted to use the model for a real enzyme, since we have only one unspecific prediction and we need 42 specific probabilities, 1 confirming the real class and 41 giving low probabilities for the other CAC_q. This problem is solved by introducing variables characteristic of each CAC class referred on the CAC_q but without giving information in the input about the real CAC class of the protein. To this end, we used the average value of each χ , MR, α , A_{vdW} and log^kP for all enzymes that belonged to the same CAC class. We also calculated the deviation of the χ , MR, α , A_{vdW} and log^kP from the respective group indicated in CAC_q. Altogether, we have then (150 χ , MR, α , A_{vdW} and log^kP values) + (150 χ , MR,

α , A_{vdW} and log^kP average values for CAC class) + (150 χ , MR, α , A_{vdW} and log^kP deviations values for CAC class average) = 450 input variables. It is of major importance to understand that we never used as input CAC_q, so the model only includes as input the χ , MR, α , A_{vdW} and log^kP values for the protein entry and the average and deviations of these values from the CAC_q, which is not necessarily the real CAC class. The general formula for this class of LDA model is shown below, where $S(E)$ is not the probability, but a real valued score that predicts the propensity of a compound to act as an inhibitor of a given class:

$$\begin{aligned}
 S(E) &= \sum_{k,G,D_i} b_k \cdot {}^k D_i(G) \\
 &\quad + \sum_{k,G,D} c_k \cdot {}^k \bar{D}_i(G) \\
 &\quad + \sum_{k,G,D} d_k \cdot \left({}^k D_i(G) - {}^k \bar{D}_i(G) \right) + a_0 \\
 &= \sum_{k,G,D_i} b_k \cdot {}^k D_i(G) + \sum_{k,G,D} c_k \cdot {}^k \bar{D}_i(G) \\
 &\quad + \sum_{k,G,D} d_k \cdot \Delta^k D_i(G) + a_0 \quad (9)
 \end{aligned}$$

LDA forward stepwise analysis was carried out for variable selection to build up the models [15]. All the variables included in the model were standardized. Subsequently, a standardized linear discriminant equation that allows comparison of their coefficients was obtained [17]. The square of Canonical regression coefficient (Rc) and Wilk's statistics (U) were examined in order to assess the discriminatory power of the model ($U = 0$ perfect discrimination, being $0 < U < 1$); the separation of the two groups of proteins was statistically verified by the Fisher ratio (F) test with an error level $P < 0.05$.

Data set

The data set was conformed to 1,221 different molecules, a set of marketed and/or reported drugs/receptor pairs where affinity/non-affinity of drugs with the receptors was established taking into consideration the IC₅₀, ki , pki , and similar values. In consequence, we managed to collect 1,191 cases active compounds in different CAC_q. In addition, we used a negative control series of 3,316 cases of non-active compounds in different CAC_q. In the two data sets used, there were the following training series: 896 cases of active compounds plus 2,485 cases of non-active compounds (3,381 in total). Predicting series: 296 + 831 = 1127 in total. The names or codes for all compounds are depicted in Supplementary material 2, due to space constraints, as well as the

Table 1 Compound Assay Conditions query (CACq) characteristic of different tests

Test	Parameter	Isoform/	Specie	Condition ^a
Enzyme test (in vitro)				
1	%	β	–	0
2	cKi	α	–	>100
3	IC ₅₀ (nM)	α	–	>2000
4	IC ₅₀ (nM)	β	–	>2000
5	IC ₅₀ (nM)	nd	–	>2000
6	IC ₅₀ (μ M)	β	–	>2
7	IC ₅₀ (μ M)	nd	–	>2
8	IC ₅₀ (μ M)	α/β	–	>2
9	IC ₅₀ E-9 (M)	nd	–	>2
10	pIC ₅₀	β	–	0
11	p IC ₅₀	nd	–	0
Parasite				
12	EC ₅₀ (μ M)	–	<i>T. brucei</i>	>2
13	IC ₅₀ (ng/ml)	–	<i>P. falciparum</i> ^b	NA
14	ED50 (μ M)	–	<i>L. donovani</i>	<5
15	IC ₅₀ (ng/ml)	–	<i>P. falciparum</i> ^c	NA
16	IC ₅₀ (μ g/ml)	–	<i>L. donovani</i>	NA
17	IC ₅₀ (μ M)	–	<i>P. falciparum</i>	>20
18	IC ₅₀ (μ M)	–	<i>L. mexicana</i>	>2
19	IC ₅₀ (μ M)	–	<i>P. falciparum</i>	>2
20	IC ₅₀ (μ M)	–	<i>P. falciparum</i> D6	NA
21	IC ₅₀ (μ M)	–	<i>P. falciparum</i> W2	NA
22	IC ₉₀ (μ g/ml)	–	<i>L. donovani</i>	NA
Bacteria				
23	IC ₅₀ (μ g/ml)	–	<i>M. intracellulare</i>	NA
24	IC ₅₀ (μ g/ml)	–	MRS	NA
25	IC ₅₀ (μ g/ml)	–	<i>S. aureus</i>	NA
26	IC ₅₀ (μ M)	–	<i>M. intracellulare</i>	–
27	IC ₅₀ (μ M)	–	MRSA	–
28	MIC (μ g/ml)	–	<i>M. tuberculosis</i>	NA
Cell line				
29	IC ₅₀ (μ g/ml)	–	HVC	NC
30	IC ₅₀ (μ M)	–	Hep2	NA
31	IC ₅₀ (μ M)	–	HT29	NA
32	IC ₅₀ (μ M)	–	HVC	NA
33	IC ₅₀ (μ M)	–	HVC	NC
34	IC ₅₀ (μ M)	–	LMM3	NA
35	IC ₅₀ (μ M)	–	PTP	>2
36	IC ₅₀ (μ M)	–	RD	>2
37	IC ₅₀ (μ M)	–	U937	>2
Fungi				
38	IC ₅₀ (μ g/ml)	–	<i>C. neoformans</i>	NA
39	IC ₅₀ (μ M)	–	<i>C. albicans</i>	–
40	IC ₅₀ (μ M)	–	<i>C. neoformans</i>	–
41	IC ₅₀ (μ M)	–	No	>2
42	EC ₅₀ (μ M)	–	HIV-1	NA

nd not determined, NA not

active, NC not cytotoxicity

^a Condition to be classified as

non-active compound (LDA

class 0)

^b Chloroquine resistant W2

clone

^c Chloroquine sensitive D6 clone

references consulted to compile the data in this table. This series is composed at random by the most representative families of GSK-3 inhibitors taken from the literature (see Supplementary material 3). The remaining compounds were a heterogeneous series of inactive compounds including members of the aforementioned families and compounds including in the Merck index [18].

Results and discussion

The development of a discriminant function [19] that allows the classification of organic compounds as active or non-active is the key step in the present approach for the discovery of GSK-3 inhibitors. It was therefore necessary to select a training data set of GSK-3 inhibitors containing wide structural variability. To describe all the compounds a series of conditions were defined that are indicated in the supplementary material gathered from the bibliography. The selection of discriminant techniques, rather than regression techniques, was determined by the lack of homogeneity in the conditions under which these values were measured. As reported in different sources, numerous IC_{50} values lie within a range rather than a single value. In other cases, the activity is not scored in terms of IC_{50} values but are quoted as inhibitory percentages at a given concentration. Once the training series had been designed, forward stepwise LDA was carried out in order to derive the QSAR:

$$\begin{aligned}
 i\text{-GSK-3} = & 0.0359 \cdot \Delta\alpha_4 (\text{Total}) - 0.0217 \cdot \Delta MR_0 (\text{Total}) \\
 & - 0.0057 \cdot \Delta MR_1 (C_{\text{inst}}) \\
 & - 0.0115 \cdot \Delta MR_0 (\text{Hetero}) \\
 & - 0.0152 \cdot \Delta\chi_0 (H_x) + 10.7644 \\
 \lambda = & 0.42, \quad F = 918.4, \quad p < 0.001 \quad (10)
 \end{aligned}$$

The statistical significance of this model was determined by examining Wilk's λ statistic, Fisher ratio (F), and the p level (p). This equation confirmed our intuitive hypothesis leading us to conclude the deviation of the parameters of one compound, from the average values for active compounds tested in a given assay condition (CAC_q), is very important for the prediction of this compound as active. Note that Eq. 10 only contains deviation measures, third type of terms in Eq. 9. In particular, i -GSK-3 seemed to be important small deviations in total polarizability and molecular refractivity indexes, as well as local deviations in refractivity and electronegativity, for heteroatom-bound hydrogen atoms. Specifically, the introduction of this last parameter in the model coincides with the i -GSK activity observed in organic acid compounds [20]. This type of model based only on atomic physicochemical parameters and drug connectivity may become an interesting alternative for fast computational pre-screening of large series of compounds in order to rationalize synthetic efforts

Table 2 Training and validation results

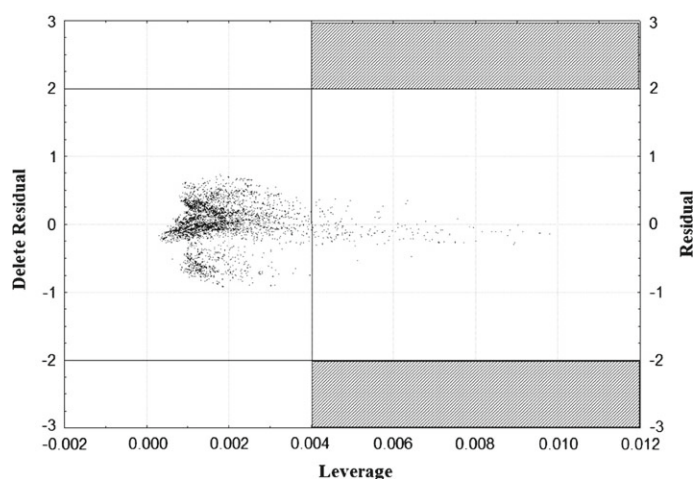
Group	Parameter	%	GSK-3 inhibitors	Non-active
Training				
GSK-3 inhibitors	Sensitivity	97.4	873	23
Non-active	Specificity	86.1	345	2140
Total	Accuracy	89.1		
Validation				
GSK-3 inhibitors	Sensitivity	96.3	285	11
Non-active	Specificity	85.4	121	710
Total	Accuracy	88.3		

[21, 22] complementary to more elaborated techniques such as 3D-QSAR, CoMFA, and CoMSIA studies that depend on a detailed knowledge of 3D structure. In any case, the present model is of more general application than the other known methods that apply only to compounds tested in only one CAC and/or belonging to only one homogeneous structural class of compounds. A confirmation of this statement is that the present classification function has given rise to an efficient separation of all compounds with accuracy = 89.1% (training series) and accuracy = 88.3% (validation series), see Table 2 for details. The names, observed classification, predicted classification and subsequent probabilities for all 4,508 cases of compounds in training and average validation are given as supplementary material. This level of total accuracy, sensitivity and specificity is considered as excellent by other researches that have used LDA for QSAR studies [12, 23–29].

Domain of applicability of the model

The interest in QSAR has steadily increased in recent decades. It is generally acknowledged that these empirical relationships are valid only within the same domain for which they were developed. However, model validation is occasionally neglected and the application domain poorly defined [30]. The purpose of this section is to outline how validation and domain definition determines in which situation it is correct to use the model [31]. The domain of applicability can be characterized in various ways as it is defined by the descriptors used in the model and the studied response. Here, a leverage approach was employed to verify the prediction reliability [31–33]. The leverage (h) is calculated by $h_i = x_i(X_T X)^{-1} \times T_i$ ($i = 1, \dots, m$), where x_i is the descriptor row-vector of the query compound, i, m is the number of query compounds, and X is the $n \times k$ matrix of the training set (k is the number of model descriptors, and n is the number of training set samples). The limit of normal values for X outliers (h^*) is set as $3(k+1)/n$, and a leverage

Fig. 1 Domain of applicability



greater than h^* . For the training set this means the chemical is highly influential in determining the model, while for the test set it means the prediction is the result of substantial extrapolation of the model and as such could be unreliable [34, 35]. To examine this in further detail, a double ordinate Cartesian plot of deleted-residuals (first ordinate), standard residuals (second ordinate), and leverages (abscissa) defined the domain of applicability of the model as a squared area within ± 2 band for residuals and a leverage threshold of $h = 0.0044$. As can be noted in Fig. 1, almost all cases used in training and validation lie within this area. Some sequences have leverage higher than the threshold, but show leave-one out (LOO) residuals, deleted-residuals, and standard residuals within the limits. In closing, no apparent outliers were detected and the model can be used with high accuracy in this applicability domain [36–38].

Conclusions

In this work, we have shown that the MARCH-INSIDE methodology can be considered as a good alternative for developing GSK-3 inhibitors in a fast and efficient way. This approach is able to correctly classify the GSK-3 inhibitory activity of compounds with different structural patterns.

Acknowledgments We are grateful to the Xunta de Galicia (INCITE08PXIB314255PR) for partial financial support. H. González-Díaz acknowledges partial financial support from Program Isidro Parga Pondal, Xunta de Galicia.

References

- Olson RE (2000) Secretase inhibitors as therapeutics for Alzheimer's disease. *Ann Rep Med Chem* 35:31–40. doi:10.1016/S0065-7743(00)35005-9
- Woodgett JR (1990) Molecular cloning and expression of glycogen synthase kinase-3/factor A. *EMBO J* 9:2431–2438
- Woodgett JR (1991) cDNA cloning and properties of glycogen synthase kinase-3 methods. *Enzymol* 200:564–577. doi:10.1016/0076-6879(91)00172-S
- Ali A, Hoeflich KP, Woodgett JR (2001) Glycogen synthase kinase-3: properties, functions, and regulation. *Chem Rev* 101:2527–2540. doi:10.1021/cr000110o
- Ishiguro K, Ihara Y, Uchida T, Imahori K (1988) A novel tubulin-dependent protein kinase forming a paired helical filament epitope on tau. *J BioChem* 104(3):319–321
- Fairlamb AH (2003) Chemotherapy on human African trypanosomiasis: current and future prospects. *Trends Parasitol* 19:488–494. doi:10.1016/j.pt.2003.09.002
- Plyte SE, Hughes K, Nilkolakaki E, Pulverer BJ, Woodgett JR (1992) Glycogen synthase kinase-3: functions in oncogenesis and development. *Biochim Biophys Acta* 1114:147–162. doi:10.1016/0304-419X(92)90012-N
- Ojo KK, Gillespie RG, Riechers A, Napuli AJ, Verlinde CL, Buckner FS et al (2008) Glycogen synthase kinase 3 is a potential drug target for african trypanosomiasis therapy. *Antimicrob Agents Chemother* 52: 3710–3717
- Freund JA, Poschel T (2000) Stochastic processes in physics, chemistry, and biology (lecture notes in physics). Springer-Verlag, Berlin
- Estrada E, Uriarte E (2001) Recent advances on the role of topological indices in drug discovery research. *Curr Med Chem* 8:1573–1588. doi:10.2174/0929867013371923
- Estrada E, Uriarte E, Montero A, Teijeira M, Santana L, De Clercq E (2001) A novel approach for the virtual screening and rational design of anticancer compounds. *J Med Chem* 43:1975–1985
- Prado-Prado FJ, Borges F, Perez-Montoto LG, Gonzalez-Diaz H (2009) Multi-target spectral moment: QSAR for antifungal

- drugs vs. different fungi species. *Eur J Med Chem* 44:4051–4056. doi:10.1016/j.ejmech.2009.04.040
13. González-Díaz H, Torres-Gomez LA, Guevara Y, Almeida MS, Molina R, Castanedo N et al (2005) Markovian chemicals “in silico” design (MARCH-INSIDE), a promising approach for computer-aided molecular design III: 2.5D indices for the discovery of antibacterials. *J Mol Model* 11:116–123. doi:10.1007/s00894-004-0228-3
 14. Gonzalez-Díaz H, Prado-Prado F, Ubeira FM (2008) Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach. *Curr Top Med Chem* 8:1676–1690. doi:10.2174/156802608786786543
 15. Santana L, Uriarte E, González-Díaz H, Zagotto G, Soto-Otero R, Mendez-Alvarez E (2006) A QSAR model for in silico screening of MAO-A inhibitors. Prediction, synthesis, and biological assay of novel coumarins. *J Med Chem* 49:1149–1156. doi:10.1021/jm0509849
 16. Concu R, Dea-Ayuela MA, Perez-Montoto LG, Bolas-Fernandez F, Prado-Prado FJ, Podda G et al (2009) Prediction of enzyme classes from 3D structure: a general model and examples of experimental-theoretic scoring of peptide mass fingerprints of Leishmania proteins. *J Proteome Res* 8:4372–4382. doi:10.1021/pr9003163
 17. Kutner MH, Nachtsheim CJ, Neter J, Li W (2005) Standardized multiple regression model. *Applied linear statistical models*, 5th edn. McGraw Hill, New York, pp. 271–277
 18. Hall Ca (1996) *The Merck Index*, 12th ed. Merck & Co, New Jersey
 19. Van Waterbeemd H (1995) Discriminant analysis for activity prediction. In: Van Waterbeemd H (ed) *Chemometric methods in molecular design*. Wiley-VCH, New York pp 265–282
 20. Konda VR, Desai A, Darland G, Bland JS, Tripp ML (2009) Rho iso-alpha acids from hops inhibit the GSK-3/NF-kappaB pathway and reduce inflammatory markers associated with bone and cartilage degradation. *J Inflamm (Lond)* 6:26–34. doi:10.1186/1476-9255-6-26
 21. Jacquemard U, Dias N, Lansiaux A, Bailly C, Loge C, Robert JM et al (2008) Synthesis of 3,5-bis(2-indolyl)pyridine and 3-[(2-indolyl)-5-phenyl]pyridine derivatives as CDK inhibitors and cytotoxic agents. *Bioorg Med Chem* 16:4932–4953
 22. Olesen PH, Sorensen AR, Urso B, Kurtzhals P, Bowler AN, Ehrbar U et al (2003) Synthesis and in vitro characterization of 1-(4-aminofurazan-3-yl)-5-dialkylaminomethyl-1H-[1,2,3]triazole-4-carboxylic acid derivatives. A new class of selective GSK-3 inhibitors. *J Med Chem* 46:3333–3341. doi:10.1021/jm021095d
 23. Calabuig C, Anton-Fos GM, Galvez J, Garcia-Domenech R (2004) New hypoglycaemic agents selected by molecular topology. *Int J Pharm* 278:111–118. doi:10.1016/j.ijpharm.2004.03.012
 24. Cercos-del-Pozo RA, Perez-Gimenez F, Salabert-Salvador MT, Garcia-March FJ (2000) Discrimination and molecular design of new theoretical hypolipaeic agents using the molecular connectivity functions. *J Chem Inf Comput Sci* 40:178–184. doi:10.1021/ci9900480
 25. Murcia-Soler M, Perez-Gimenez F, Garcia-March FJ, Salabert-Salvador MT, Diaz-Villanueva W, Medina-Casamayor P (2003) Discrimination and selection of new potential antibacterial compounds using simple topological descriptors. *J Mol Graph Model* 21:375–390. doi:10.1016/S1093-3263(02)00184-5
 26. Estrada E, Vilar S, Uriarte E, Gutierrez Y (2002) In silico studies toward the discovery of new anti-HIV nucleoside compounds with the use of TOPS-MODE and 2D/3D connectivity indices. 1. Pyrimidyl derivatives. *J Chem Inf Comput Sci* 42:1194–1203. doi:10.1021/ci0255331
 27. Cronin MT, Aptula AO, Dearden JC, Duffy JC, Netzeva TI, Patel H et al (2002) Structure-based classification of antibacterial activity. *J Chem Inf Comput Sci* 42:869–878. doi:10.1021/ci025501d
 28. Prado-Prado FJ, Ubeira FM, Borges F, Gonzalez-Diaz H (2010) Unified QSAR & network-based computational chemistry approach to antimicrobials. II. Multiple distance and triadic census analysis of antiparasitic drugs complex networks. *J Comput Chem* 31:164–173. doi:10.1002/jcc.21292
 29. Prado-Prado FJ, Martínez de la Vega O, Uriarte E, Ubeira FM, Chou KC, Gonzalez-Diaz H (2009) Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modeling and comparative multi-distance study of the giant components of antiviral drug-drug complex networks. *Bioorg Med Chem* 17:569–575. doi:10.1016/j.bmc.2008.11.075
 30. Oberg T (2004) A QSAR for baseline toxicity: validation, domain of application, and prediction. *Chem Res Toxicol* 17:1630–1637. doi:10.1021/tx0498253
 31. Gramatica P (2007) Principles of QSAR models validation: internal and external. *QSAR Comb Sci* 26:694–701. doi:10.1002/qsar.200610151
 32. Eriksson L, Jaworska J, Worth AP, Cronin MT, McDowell RM, Gramatica P (2003) Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ Health Perspect* 111:1361–1375. doi:10.1289/ehp.5758
 33. Tropsha A, Gramatica P, Gombar VK (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb Sci* 22:69–77. doi:10.1002/qsar.200390007
 34. Melagraki G, Afantitis A, Sarimveis H, Koutentis PA, Kollias G, Igglessi-Markopoulou O (2009) Predictive QSAR workflow for the in silico identification and screening of novel HDAC inhibitors. *Mol Divers*. 13:301–311. doi:10.1007/s11030-009-9115-2
 35. Li J, Gramatica P (2009) The importance of molecular structures, endpoints’ values, and predictivity parameters in QSAR research: QSAR analysis of a series of estrogen receptor binders. *Mol Divers*. doi:10.1007/s11030-009-9212-2
 36. Papa E, Villa F, Gramatica P (2005) Statistically validated QSARs, based on theoretical descriptors, for modeling aquatic toxicity of organic chemicals in *Pimephales promelas* (fathead minnow). *J Chem Inf Model* 45:1256–1266. doi:10.1021/ci050212i
 37. Liu H, Papa E, Gramatica P (2006) QSAR prediction of estrogen activity for a large set of diverse chemicals under the guidance of OECD principles. *Chem Res Toxicol* 19:1540–1548. doi:10.1021/tx0601509
 38. Gramatica P, Giani E, Papa E (2006) Statistical external validation and consensus modeling: a QSPR case study for K(oc) prediction. *J Mol Graph Model* 25:755–766. doi:10.1016/j.jmgm.2006.06.005



Transworld Research Network
37/661 (2), Fort P.O.
Trivandrum-695 023
Kerala, India

Complex Network Entropy: From Molecules to Biology, Parasitology, Technology, Social,
Legal, and Neurosciences, 2011: 17-29 ISBN: 978-81-7895-507-0
Editors: Humberto González-Díaz, Francisco J. Prado-Prado and Xerardo García-Mera

2. Entropy Multi-target QSAR model for Anti-Parasitic and Anti-Alzheimer GSK-3 inhibitors

Isela García¹, Yagamare Fall¹, Generosa Gómez¹
and Humberto González-Díaz²

¹Department of Organic Chemistry, University of Vigo, Spain; ²Department of Microbiology and
Parasitology, Faculty of Pharmacy, USC, Santiago de Compostela, 15782, Spain

1. Introduction

In this moment, there is an increasing interest in the evaluation of kinases from unicellular parasites as targets for potential new anti-parasitic drugs. The evolutionary difference between unicellular kinases and their human homologues might be sufficient to allow the design of parasite-specific inhibitors. The *Plasmodium falciparum* genome contains 65 genes that encode kinases, including three forms of Glycogen synthase kinase-3 (GSK-3). An initial study showed that *P. falciparum* exports PfGSK-3 to the cytoplasm of host erythrocytes (which are devoid of GSK-3), where it colocalizes with parasite-generated membrane structures known as Maurer's clefts.³ The function of PfGSK-3 is unknown, but the presence PfCK1, a CK1 homolog, in infected red blood cell supports the hypothesis that both kinases play a role in regulating the strong circadian rhythm of the parasite, which is responsible for the circadian fevers that are characteristic of this infectious disease [1].

Correspondence/Reprint request: Dr. Humberto González-Díaz, Department of Microbiology and Parasitology
Faculty of Pharmacy, USC, Santiago de Compostela, 15782, Spain. E-mail: gonzalezdzah@yahoo.es

The vector-borne parasitic disease African trypanosomiasis, caused by members of the *Trypanosoma brucei* complex, is a serious health threat. It is estimated that 300,000 to 500,000 humans in sub-Saharan African are infected. If the disease is left inadequately treated, it often has a fatal outcome. Once infection is established, safe and effective therapy is critically important, yet it has been difficult to achieve. Despite the critical need, the available therapies are becoming less satisfactory due to the rising level of resistance to the available drugs, the long period of treatment required to achieve a cure, and the unacceptable and sometimes severe adverse effects associated with current therapies [2]. An urgent priority is to identify and validate new targets for the development of safe, effective, and inexpensive therapeutic alternatives. Compounds that inhibit *T. brucei* GSK-3 activity and not host GSK-3 might be required for therapy for pregnant women and infants, in that GSK-3 regulates proteins critical in development, such as the *wnt* gene product. However, optimization of the selectivity of drug candidates for parasite kinases becomes an issue due to the highly conserved amino acids and protein conformation of the catalytic domains [3-6]. Understanding the differences in the substrate binding properties and the three-dimensional structures between mammalian and parasite GSK-3 enzymes is important for the optimization of selected target inhibitors for drug development [7, 8]. To report of all these cases, more parasites, fungi, etc. exist that are keep out for compounds that also disable the enzyme GSK-3, and in it consists our aim objective of this work.

On the other hand, Alzheimer's disease (AD) is the most recent reason of dementia in the elders at present [9]. This serious and degenerative disorder explains the gradual loss of neurons, and in spite of the efforts realized by the big pharmacists of the world, still is not very clear the reason of this pathology. The fundamental characteristic of Alzheimer's disease is the presence in the brain of two injuries: the neurofibrillary tangles that are formed by paired helical filaments (PHF) whose main component is Tau protein kinase (TPK) and the senile plaques formed by the aggregation of the β -amiloide peptide. In addition, GSK-3 is a serine-threonine kinase encoded by two isoforms in mammals, termed GSK-3 α and GSK-3 β [10]. Initially GSK-3 was implicated in muscle energy storage and metabolism, but since its cloning, a more generalized role in cellular regulation has emerged, highlighted by the wide array of substrates controlled by this enzyme that includes cytoplasmic proteins and nuclear transcription factors. GSK-3 targets encompass proteins implicated in Alzheimer's disease, neurological disorders, in the *wnt* and insulin signaling pathway, glycogen and protein synthesis, regulation of transcription factors, embryonic development, cell proliferation and adhesion, tumorigenesis, apoptosis, circadian rhythm,... etc.

The functions of GSK-3 and its implication in various human diseases have stimulated and active search for potent and selective GSK-3 inhibitors [11]. Studies of GSK-3 homologues in various organisms have revealed physiological roles for the enzyme in differentiation, cell fate determination, and spatial patterning to establish bilateral embryonic symmetry [12]. Purified GSK-3 α and GSK-3 β exhibit similar biochemical and substrate properties [12, 13], and it is known that in the phosphorylation of Tau protein kinase takes part actively glycogen synthase kinase 3 β (GSK-3 β), which not only plays a fundamental role in the synthesis of the glycogen (where it was identified by the first time), but it is very important in several processes as cellular signs, metabolic control, embryogenesis, cellular death and oncogenesis [14], and it is related to a wide range of neurodegenerative [15] diseases, bipolar mood disorders [16] and diabetes, by this the inhibition of this enzyme is one of the therapeutic aims more promoters of those which are fighting at present. In 1988 Ishiguro and col. [17] isolated one enzyme when they were studying an extract of the brain that there was showing the generation of paired helical filaments of Tau protein kinase, typical injury of Alzheimer's disease. TPKI and TPKII are the two kinases implied in this process and they found that TPKI has an identical structure to GSK-3 β .

In parallel, the development of QSARs using simple molecular indices appears to be a promising alternative or complementary technique to drug-protein docking, high-throughput screening and combinatorial chemistry techniques. Almost all QSAR techniques are based on the use of molecular descriptors, which are numerical series that codify useful chemical information and enable correlations between statistical and biological properties [18-20]. Shannon entropy and other entropy-like measures are one of the more prominent parameters in order to codify structural information on QSAR studies [21, 22]. In this direction, our research group has introduced a novel series of stochastic indices in the so called MARCH-INSIDE approach. The method is based on the use of Markov Models (MM) to calculate absolute probabilities of the distribution of different atomic properties within the molecular skeleton with specific bonding patterns. Using these absolute probabilities we can apply the Shannon formula to calculate entropy parameters of the distribution of atomic properties in the molecule [23, 24]. In this work we will explore the potential of MARCH-INSIDE to seek a QSAR for GSK-3 inhibitors from a heterogeneous series of compounds. In the first step, the aforementioned molecular descriptors were calculated for a large series of active/non-active compounds. LDA was subsequently used to fit a classification function. The QSAR developed was then validated with an external predicting series by the re-substitution technique.

2. Materials and methods

Computational methods. The MARCH-INSIDE approach [25-27] is based on the calculation of the different physicochemical molecular properties as an average of atomic properties (ap). For instance, it is possible to derive average estimations of molecular descriptors or group indices [28, 29].

Multi-target Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) was used to construct the classifiers. One of the most important steps in this work was the organization of the spreadsheet containing the raw data used as input for the LDA because this is not a classic classifier. Herein, the schematisation of the paper is peculiar. Our expectation is to use a two-group Discriminant function to classify compounds into two possible groups: compounds that belong to a particular group and compounds that do not belong to this group. To this end, we have to indicate somehow what group we pretend to predict in each case. In this regard, we made the following steps, these steps are essentially the same given by Concu *et al.* [30, 31] for the QSAR study of six classes of enzymes:

1. We created a raw data representing each compound input as a vector made up of 1 output variable, 108 structural variables (inputs) divided in values (see the first term of the equation 1), averages (see the second term of the equation 1) and differences between values and averages (see the third term of the equation 1); and the Compound Assay Conditions query (CAC_q) variable. CAC_q is an auxiliary not used to construct the model.
2. The first element (output) is a dummy variable (Boolean) called Observed Group (OG); OG = 0 if the compound belongs to the class to which we refer in CAC_q and 1 otherwise (OG = 1). We could repeat each compound more than once in the raw data. In fact, we could repeat each compound 38 times corresponding to 38 CAC_q Assay Conditions (see Table 1). The first time we used the CAC_q = CAC number. It means that we used the real CAC class of the compound in CAC_q. In this case, the LDA model had to give the highest probability to the group OG = 0 because it had to predict the real class of the compound. The remnant 38 times we use an CAC class number different to the real in CAC_q and then the LDA model had to predict the highest probability for the group OG = 1. This indicated that the compound did not belong to this group.

Table 1. Compound Assay Conditions query (CACq).

q	Param.	Enz.	Iso.	Target	Type	Class	Cond.	Obs.
12	IC ₅₀ (µg/mL)	no	no	MRS	bacterium	0	NA	=
13	IC ₅₀ (µg/mL)	no	no	<i>S. aureus</i>	bacterium	0	NA	=
23	IC ₅₀ (µM)	no	no	MRSA	bacterium	0	—	=
24	IC ₅₀ (µM)	no	no	Hep2	bacterium	0	—	=
36	MIC (µg/mL)	no	no	no	bacterium	0	NA	=
14	IC ₅₀ (µg/mL)	no	no	HVC	CL	0	NA	=
26	IC ₅₀ (µM)	no	no	HVC	CL	0	NA	=
27	IC ₅₀ (µM)	no	no	RD	CL	0	NC	<
28	IC ₅₀ (µM)	no	no	U937	CL	0	2	>
25	IC ₅₀ (µM)	no	no	HT29	CL	0	NA	=
1	%	GSK-3	β	no	enzyme	0	0	=
2	cKi	GSK-3	α	no	enzyme	0	100	=
9	IC ₅₀ (nM)	GSK-3	α	no	enzyme	0	2000	>
10	IC ₅₀ (nM)	GSK-3	β	no	enzyme	0	2000	>
11	IC ₅₀ (nM)	GSK-3	nd	<i>M. intracellulare</i>	enzyme	0	2000	>
18	IC ₅₀ (µM)	GSK-3	β	no	enzyme	0	2	>
19	IC ₅₀ (µM)	GSK-3	nd	<i>T. brucei</i>	enzyme	0	2	>
20	IC ₅₀ (µM)	GSK-3	nd	<i>P. falciparum</i>	enzyme	0	2	>
22	IC ₅₀ (µM)	GSK-3	α/β	<i>M. intracellulare</i>	enzyme	0	2	>
34	IC ₅₀ E-9 (M)	GSK-3	nd	<i>L. donovani</i>	enzyme	0	2	>
37	pIC ₅₀	GSK-3	β	no	enzyme	0	0	=
38	pIC ₅₀	GSK-3	nd		enzyme	0	0	=
29	IC ₅₀ (µM)	no	no	<i>C. albicans</i>	fungi	0	2	>
15	IC ₅₀ (µg/mL)	no	no	<i>C. neoformans</i>	fungi	0	NC	<

Table 1. Continued

16	IC ₅₀ (µg/mL)	no	no	<i>L. donovani</i>	parasite	0	NA	=
30	IC ₅₀ (µM)	no	no	<i>P. falciparum</i>	parasite	0	—	=
6	ED ₅₀ (µM)	no	no	<i>P. falciparum</i> W2	parasite	0	5	<
7	IC ₅₀ (ng/mL)	no	no	<i>P. falciparum</i> D6	parasite	0	NA	=
8	IC ₅₀ (ng/mL)	no	no	no	parasite	0	NA	=
17	IC ₅₀ (µg/mL)	no	no	no	parasite	0	NA	=
21	IC ₅₀ (µM)	GSK-3	no	no	parasite	0	20	>
31	IC ₅₀ (µM)	no	no	<i>P. falciparum</i> D6	parasite	0	2	>
32	IC ₅₀ (µM)	no	no	<i>P. falciparum</i> W2	parasite	0	NA	=
33	IC ₅₀ (µM)	no	no	no	parasite	0	NA	=
5	EC ₅₀ (µM)	no	no	<i>L. donovani</i>	parasite	0	NT	=
35	IC ₉₀ (µg/mL)	no	no	<i>M. tuberculosis</i> h	parasite	0	NA	=
3	EC ₅₀ (µM)	no	no	CE	no	0	2	>
4	EC ₅₀ (µM)	no	no	HIV-1	virus	0	inactive	=

CE = Cell Efficacy glycogen synthase stimulation, HVC = Human Vero cells, W2 = choroquine resistant W2 clone, D6 = choroquine sensitive D6 clone, CL = Cellular Line, *M. tuberculosis* h = *M. tuberculosis* (H37Rv), nd = not determined, NA = not active, NC = not cytotoxicity

3. The problem in this type of organization of raw data is that $\theta_k(G)$ values are global or local compound constants that depend only on structure. Consequently, if these latter and LDA are based only on these values, they will necessarily fail when we change OG values. An inconvenient in this regard occurs if we pretend to use the model for a real enzyme, since we have only one unspecific prediction and we need 38 specific probabilities, 1 confirming the real class and 37 giving low probabilities for the other CAC_q . We can solve this problem introducing variables characteristic of each CAC class referred on the CAC_q but without giving information in the input about the real CAC class of the protein. To this end, we used the average value of each $\theta_k(G)$ for all enzymes that belonged to the same CAC class. We also calculated the deviation of the $\theta_k(G)$ from the respective group indicated in CAC_q . Altogether, we have then 36 $\theta_k(G)$ values + 36 $\theta_k(G)_{avg}$ average values for CAC class + 36 $\theta_k(G)_{dev}$ deviation values from CAC class average = 108 input variables. It is of major importance to understand that we never used as input CAC_q , so the model only includes as input the $\theta_k(G)$ values for the protein entry and the average and deviations of these values from the CAC_q , which is not necessarily the real CAC class. The general formula for this class of LDA model is shown below, where $S(CAC_q)$ is not the

probability but a real valued score that predicts the propensity of a compound to act as an inhibitor of a given class:

$$\begin{aligned}
 S(E) &= \sum_{k,G,Dt} b_k \cdot \theta_k(G) + \sum_{k,G,D} c_k \cdot \theta_k(G)_{avg} + \sum_{k,G,D} d_k \cdot (\theta_k(G) - \theta_k(G)_{avg}) + a_0 \quad (1) \\
 &= \sum_{k,G,Dt} b_k \cdot \theta_k(G) + \sum_{k,G,D} c_k \cdot \theta_k(G)_{avg} + \sum_{k,G,D} d_k \cdot \theta_k(G)_{dif} + a_0
 \end{aligned}$$

LDA forward stepwise analysis was carried out for variable selection to build up the models [29]. All the variables included in the model were standardized in order to bring them onto the same scale. Subsequently, a standardized linear discriminant equation that allows comparison of their coefficients was obtained [32]. The square of Canonical regression coefficient (Rc) and Wilk's statistics (U) were examined in order to assess the discriminatory power of the model (U = 0 perfect discrimination, being $0 < U < 1$); the separation of the two groups of proteins was statistically verified by the Fisher ratio (F) test with an error level $p < 0.05$.

Data Set. The data set was conformed to a set of marketed and/or reported drugs/receptor pairs where affinity/non-affinity of drugs with the receptors was established taking into consideration the IC_{50} , ki , pki ,... values. In consequence, we managed to collect 1012 cases active compounds in different CAC_q . In addition, we used a negative control series of 2536 cases of non-active results for compounds evaluated at different CAC_q . The two data sets used were: training series with 249 active + 638 non-active (887 in total) and validation series with 763 + 1898 = 2661 cases in total. The names or codes for all compounds are depicted in the Supporting Information, due to space constraints, as well as the references consulted to compile the data in this table. This series is composed at random by the most representative families of GSK-3 inhibitors taken from the literature (supplementary material). The remaining compounds were a heterogeneous series of inactive compounds including members of the aforementioned families and compounds including in the Merck index [33].

3. Results and Discussion

General QSAR for GSK-3 inhibitors. The development of a discriminant function [34] that allows the classification of organic compounds as active or non-active is the key step in the present approach for the discovery of GSK-3 inhibitors. It was therefore necessary to select a

training data set of GSK-3 inhibitors containing wide structural variability. To define all the compounds there have been defined a series of conditions that are indicated in the supplementary material gathered from the bibliography. The selection here of discriminant techniques instead of regression techniques was determined by the lack of homogeneity in the conditions under which these values were measured. As reported in different sources, numerous IC₅₀ values lie within a range rather than a single value. In other cases, the activity is not scored in terms of IC₅₀ values but is quoted as inhibitory percentages at a given concentration. Once the training series had been designed, forward stepwise Linear Discriminant Analysis (LDA) was carried out in order to derive the QSAR, see the full equation as well as the compact notation of the model:

$$s(\text{CACq}) = -1.75 \cdot \theta_2(C_{\text{inst}})_{\text{avg}} + 8.39 \cdot \theta_0(X)_{\text{avg}} + 2.33 \cdot \theta_5(\text{Het})_{\text{avg}} \\ + 1.66 \cdot (\theta_0(\text{Total}) - \theta_0(\text{Total})_{\text{avg}}) + 0.49 \cdot (\theta_1(C_{\text{inst}}) - \theta_1(C_{\text{inst}})_{\text{avg}}) + 0.62 \quad (2a) \\ \lambda = 0.54 \quad F = 147.8 \quad p < 0.001$$

$$s(\text{CACq}) = -1.75 \cdot \theta_2(C_{\text{inst}})_{\text{avg}} + 8.39 \cdot \theta_0(X)_{\text{avg}} + 2.33 \cdot \theta_5(\text{Het})_{\text{avg}} \\ + 1.66 \cdot \theta_0(\text{Total})_{\text{dif}} + 0.49 \cdot \theta_1(C_{\text{inst}})_{\text{dif}} + 0.62 \quad (2b) \\ \lambda = 0.54 \quad F = 147.8 \quad p < 0.001$$

The statistical significance of this model was determined by examining Wilk's λ statistic, Fisher ratio (F), and the p-level (p). This equation confirm our intuitive hypothesis and we can conclude that the deviation of the parameters of one compound from the average values for active compounds tested in a given assay condition (CACq) is very important for the prediction of this compound as active. In any case, the present model is of more general application than the other known methods that apply only to compounds tested in only one CAC and/or belonging to only one homogeneous structural class of compounds. A confirmation of this stamen is that the present classification function have given rise to an efficient separation of all compounds with Accuracy = 84.0% (training series) and Accuracy = 84.4% (validation series), see **Table 2** for details. The names, observed classification, predicted classification and subsequent probabilities for all 3548 compounds in training and average validation are given as supplementary material. This level of total Accuracy, Sensitivity and Specificity is considered as excellent by other researches that have used LDA for QSAR studies; see for instance the works of Garcia-Domenech, R., Prado-Prado, F. J.; Marrero-Ponce, Y., etc [35-50].

Table 2. Training and validation results.

Group	Parameter	%	GSK-3 inhibitors	Non-active
Training				
GSK-3 inhibitor	Sensitivity	96.4	240	9
Non-active	Specificity	79.2	133	505
Total	Accuracy	84.0		
Validation				
GSK-3 inhibitor	Sensitivity	96.3	735	28
Non-active	Specificity	79.6	387	1511
Total	Accuracy	84.4		

4. Conclusions

In this work we have shown that the MARCH-INSIDE methodology can be considered a good alternative for developing GSK-3 inhibitors in a fast and efficient way. This approach is able to correctly classify the GSK-3 inhibitory activity of compounds with different structural patterns.

Acknowledgment

We are grateful to the Xunta de Galicia (INCITE08PXIB314255PR) for partial financial support. González-Díaz, H. acknowledges financial support from Program Isidro Parga Pondal, Xunta de Galicia.

5. References

1. Meijer L, Flajolet M, Geengard P. Pharmacological inhibitors of glycogen synthase kinase 3. *Trends in Pharmacological Sciences*. 2004 September 2004;25(9):471-80.
2. Fairlamb AH. Chemotherapy on human African trypanosomiasis: current and future prospects. *Trends Parasitol*. 2003;19:488-94.
3. Copeland RA, Pompliano DL, Meek TD. Drug-target residence time and its implications for lead optimization. *Nat Rev Drug Discov*. 2006;5:730-2.
4. Liao JJ. Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J Med Chem*. 2007;50:409-24.
5. Pink R, Hudson A, Mouries MA, Bending M. Opportunities and challenges in antiparasitic drug discovery. *Nat Rev Drug Discov*. 2005;4:727-40.
6. Plyte SE, Hughes K, Nilkolakaki E, Pulverer BJ, Woodgett JR. Glycogen Synthase Kinase-3: functions in oncogenesis and development. *Biochim Biophys Acta*. 1992;1114:147-62.

7. Dajani R, Fraser E, Roe SM, Young N, Good V, Dale TC, et al. Crystal structure of glycogen synthase kinase-3 beta: structural basis for phosphate-primed substrate specificity and autoinhibition. *Cell (Cambridge, Mass)*. 2001;105:721-32.
8. Ojo KK, Gillespie RG, Riechers A, Napuli AJ, Verlinde CL, Buckner FS, et al. Glycogen Synthase Kinase 3 is a potential drug target for african trypanosomiasis therapy. *Antimicrob Agents and Chemother*. 2008 October 2008;52(10):3710-7.
9. Olson RE. Secretase inhibitors as therapeutics for Alzheimer's disease. *Annu Rep Med Chem*. 2000;35:31-40.
10. Woodgett JR. Molecular cloning and expression of glycogen synthase kinase-3/factor A. *EMBO J*. 1990 Aug;9(8):2431-8.
11. Doucheau E. *Plasmodium falciparum* glycogen synthase kinase-3: molecular model, expression, intracellular localisation and selective inhibitors *Biochim Biophys Acta*. 2004;1697:181-96.
12. Ali A, Hoeflich KP, Woodgett JR. Glycogen Synthase Kinase-3: Properties, Functions, and Regulation. *Chem Rev*. 2001;101:2527-40.
13. Woodgett JR. cDNA cloning and properties of glycogen synthase kinase-3 *Methods Enzymol*. 1991;200:564-77.
14. Grimes CA, Jope RS. The Multifaceted roles of glycogen synthase kinase 3B in cellular signaling. *Prog Neurobiol*. 2001;65:391-426.
15. Nadri C, Lipska B, Kozlovsky N, Weinberger DR, Belmaker RH, Agam G. *Dev Brain Res*. 2003;141(1,2):33-7.
16. Gould TD, Zarate CA, Manji HK. Glycogen Synthase Kinase-3: A Target for Novel Bipolar Disorder Treatments. 2004;65(1):10-21.
17. Ishiguro K, Ihara Y, Uchida T, Imahori K. A Novel Tubulin-Dependent Protein Kinase Forming a Paired Helical Filament Epitope on Tau. *J Bio Chem*. 1988;104(3):319-21.
18. Nunez MB, Maguna FP, Okulik NB, Castro EA. QSAR modeling of the MAO inhibitory activity of xanthenes derivatives. *Bioorg Med Chem Lett*. 2004 Nov 15;14(22):5611-7.
19. Todeschini R, Consonni V. *Handbook of Molecular Descriptors*. Wiley VCH. 2000.
20. Freund JA, Poschel T. *Stochastic processes in physics, chemistry, and biology. Lect Notes Phys*. Berlin, Germany: Springer-Verlag 2000.
21. Stahura FL, Godden JW, Bajorath J. Differential Shannon entropy analysis identifies molecular property descriptors that predict aqueous solubility of synthetic compounds with high accuracy in binary QSAR calculations. *J Chem Inf Comput Sci*. 2002 May-Jun;42(3):550-8.
22. Stahura FL, Godden JW, Xue L, Bajorath J. Distinguishing between natural products and synthetic molecules by descriptor Shannon entropy analysis and binary QSAR calculations. *J Chem Inf Comput Sci*. 2000 Sep-Oct;40(5):1245-52.
23. González-Díaz H, Marrero Y, Hernandez I, Bastida I, Tenorio E, Nasco O, et al. 3D-MEDNEs: an alternative "in silico" technique for chemical research in toxicology. 1. prediction of chemically induced agranulocytosis. *Chem Res Toxicol*. 2003 Oct;16(10):1318-27.
24. González-Díaz H, Aguero G, Cabrera MA, Molina R, Santana L, Uriarte E, et al. Unified Markov thermodynamics based on stochastic forms to classify drugs

- considering molecular structure, partition system, and biological species: distribution of the antimicrobial G1 on rat tissues. *Bioorg Med Chem Lett*. 2005 Feb 1;15(3):551-7.
25. Gonzalez-Díaz H, Prado-Prado F, Ubeira FM. Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach. *Curr Top Med Chem*. 2008;8(18):1676-90.
 26. González-Díaz H, González-Díaz Y, Santana L, Ubeira FM, Uriarte E. Proteomics, networks and connectivity indices. *Proteomics*. 2008;8:750-78.
 27. González-Díaz H, Vilar S, Santana L, Uriarte E. Medicinal Chemistry and Bioinformatics – Current Trends in Drugs Discovery with Networks Topological Indices. *Curr Top Med Chem*. 2007;7(10):1025-39.
 28. Santana L, Gonzalez-Diaz H, Quezada E, Uriarte E, Yanez M, Vina D, et al. Quantitative structure-activity relationship and complex network approach to monoamine oxidase a and B inhibitors. *J Med Chem*. 2008 Nov 13;51(21):6740-51.
 29. Santana L, Uriarte E, González-Díaz H, Zagotto G, Soto-Otero R, Mendez-Alvarez E. A QSAR model for in silico screening of MAO-A inhibitors. Prediction, synthesis, and biological assay of novel coumarins. *J Med Chem*. 2006 Feb 9;49(3):1149-56.
 30. Concu R, Dea-Ayuela MA, Perez-Montoto LG, Prado-Prado FJ, Uriarte E, Bolas-Fernandez F, et al. 3D entropy and moments prediction of enzyme classes and experimental-theoretic study of peptide fingerprints in Leishmania parasites. *Biochim Biophys Acta*. 2009 Aug 28;1794(12):1784-94.
 31. Concu R, Dea-Ayuela MA, Perez-Montoto LG, Bolas-Fernandez F, Prado-Prado FJ, Podda G, et al. Prediction of Enzyme Classes from 3D Structure: A General Model and Examples of Experimental-Theoretic Scoring of Peptide Mass Fingerprints of Leishmania Proteins. *Journal of proteome research*. 2009 Sep 4;8(9):4372-82.
 32. Kutner MH, Nachtsheim CJ, Neter J, Li W. Standardized Multiple Regression Model. *Applied Linear Statistical Models*. Fifth ed. New York: McGraw Hill 2005:271-7.
 33. Hall Ca. The Merck Index, twelfth ed. 1996.
 34. Van Waterbeemd H. Discriminant Analysis for Activity Prediction. In: Van Waterbeemd H, ed. *Chemometric methods in molecular design*. New York: Wiley-VCH 1995:265-82.
 35. Calabuig C, Anton-Fos GM, Galvez J, Garcia-Domenech R. New hypoglycaemic agents selected by molecular topology. *Int J Pharm*. 2004 Jun 18;278(1):111-8.
 36. Garcia-Garcia A, Galvez J, de Julian-Ortiz JV, Garcia-Domenech R, Munoz C, Guna R, et al. New agents active against Mycobacterium avium complex selected by molecular topology: a virtual screening method. *J Antimicrob Chemother*. 2004 Jan;53(1):65-73.
 37. Prado-Prado FJ, Ubeira FM, Borges F, Gonzalez-Diaz H. Unified QSAR & network-based computational chemistry approach to antimicrobials. II. Multiple distance and triadic census analysis of antiparasitic drugs complex networks. *J Comput Chem*. 2009 May 6.
 38. Prado-Prado FJ, Martinez de la Vega O, Uriarte E, Ubeira FM, Chou KC, Gonzalez-Diaz H. Unified QSAR approach to antimicrobials. 4. Multi-target

- QSAR modeling and comparative multi-distance study of the giant components of antiviral drug-drug complex networks. *Bioorg Med Chem.* 2009 Jan 15;17(2):569-75.
39. Prado-Prado FJ, de la Vega OM, Uriarte E, Ubeira FM, Chou KC, Gonzalez-Diaz H. Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modeling and comparative multi-distance study of the giant components of antiviral drug-drug complex networks. *Bioorg Med Chem.* 2009;17:569-75.
 40. Prado-Prado FJ, Borges F, Perez-Montoto LG, Gonzalez-Diaz H. Multi-target spectral moment: QSAR for antifungal drugs vs. different fungi species. *Eur J Med Chem.* 2009 May 5;44(10):4051-6.
 41. Prado-Prado FJ, Gonzalez-Diaz H, de la Vega OM, Ubeira FM, Chou KC. 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 Jun 1;16(11):5871-80.
 42. Prado-Prado FJ, Gonzalez-Diaz H, 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 Jan 15;15(2):897-902.
 43. Marrero-Ponce Y, Khan MT, Casanola Martin GM, Ather A, Sultankhodzhaev MN, Torrens F, et al. Prediction of Tyrosinase Inhibition Activity Using Atom-Based Bilinear Indices. *ChemMedChem.* 2007 Apr 16;2(4):449-78.
 44. Marrero-Ponce Y, Meneses-Marcel A, Castillo-Garit JA, Machado-Tugores Y, Escario JA, Barrio AG, et al. Predicting antitrichomonal activity: a computational screening using atom-based bilinear indices and experimental proofs. *Bioorg Med Chem.* 2006 Oct 1;14(19):6502-24.
 45. Meneses-Marcel A, Marrero-Ponce Y, Machado-Tugores Y, Montero-Torres A, Pereira DM, Escario JA, et al. A linear discrimination analysis based virtual screening of trichomonacidal lead-like compounds: outcomes of in silico studies supported by experimental results. *Bioorg Med Chem Lett.* 2005 Sep 1;15(17):3838-43.
 46. Marrero-Ponce Y, Diaz HG, Zaldivar VR, Torrens F, Castro EA. 3D-chiral quadratic indices of the 'molecular pseudograph's atom adjacency matrix' and their application to central chirality codification: classification of ACE inhibitors and prediction of sigma-receptor antagonist activities. *Bioorg Med Chem.* 2004 Oct 15;12(20):5331-42.
 47. Murcia-Soler M, Perez-Gimenez F, Garcia-March FJ, Salabert-Salvador MT, Diaz-Villanueva W, Medina-Casamayor P. Discrimination and selection of new potential antibacterial compounds using simple topological descriptors. *J Mol Graph Model.* 2003 Mar;21(5):375-90.
 48. Cercos-del-Pozo RA, Perez-Gimenez F, Salabert-Salvador MT, Garcia-March FJ. Discrimination and molecular design of new theoretical hypolipemic agents using the molecular connectivity functions. *J Chem Inf Comput Sci.* 2000 Jan;40(1):178-84.
 49. Estrada E, Vilar S, Uriarte E, Gutierrez Y. In silico studies toward the discovery of new anti-HIV nucleoside compounds with the use of TOPS-MODE and

- 2D/3D connectivity indices. 1. Pyrimidyl derivatives. *J Chem Inf Comput Sci.* 2002 Sep-Oct;42(5):1194-203.
50. Cronin MT, Aptula AO, Dearden JC, Duffy JC, Netzeva TI, Patel H, et al. Structure-based classification of antibacterial activity. *J Chem Inf Comput Sci.* 2002 Jul-Aug;42(4):869-78.

Article

Using Topological Indices to Predict Anti-Alzheimer and Anti-Parasitic GSK-3 Inhibitors by Multi-Target QSAR *in Silico* Screening

Isela García *, Yagamare Fall and Generosa Gómez

Department of Organic Chemistry, Faculty of Chemistry, University of Vigo, Spain

* Author to whom correspondence should be addressed; E-Mail: iselapintos@yahoo.es;
Tel.: +34-986-813-679; Fax: +34-986-812-262.

Received: 8 June 2010; in revised form: 27 July 2010 / Accepted: 2 August 2010 /

Published: 9 August 2010

Abstract: *Plasmodium falciparum*, *Leishmania*, *Trypanosomes*, are the causers of diseases such as malaria, leishmaniasis and African trypanosomiasis that nowadays are the most serious parasitic health problems worldwide. The great number of deaths and the few drugs available against these parasites, make necessary the search for new drugs. Some of these antiparasitic drugs also are GSK-3 inhibitors. GSKI-3 are candidates to develop drugs for the treatment of Alzheimer's disease. In this work topological descriptors for a large series of 3,370 active/non-active compounds were initially calculated with the ModesLab software. Linear Discriminant Analysis was used to fit the classification function and it predicts heterogeneous series of compounds like paullones, indirubins, meridians, *etc.* This study thus provided a general evaluation of these types of molecules.

Keywords: glycogen synthase kinase-3 (GSK-3) inhibitors; Alzheimer's disease; *Plasmodium falciparum*; *Trypanosoma brucei*; *Leishmania*; QSAR

1. Introduction

Neurofibrillary tangles (NFTs) are one of the characteristic neuropathological lesions of Alzheimer's disease (AD) [1] and other neurodegenerative processes such as frontotemporal dementia, Pick's disease, progressive supranuclear palsy, and corticobasal degeneration [2].

Alzheimer's disease affects 5–10% of the population over 65 years of age. The dementia associated with AD results from the selective death of neurons, which is associated with several anatomopathological hallmarks such as senile neuritic plaques and neurofibrillary tangles [3,4]. Three molecular actors clearly play a role in the development of AD: the amyloid β peptide (A β), presenilins-1 and -2 and the microtubule-associated protein tau.

Glycogen Synthase Kinase-3 (GSK-3) is a senile-threonine kinase ubiquitously expressed and involved in the regulation of many cell functions [5]. GSK-3 was originally identified as one of the five protein kinases that phosphorylate glycogen synthase [6], being implicated in type-2 diabetes [7]. GSK-3 is also known to phosphorylate the microtubule-associated protein tau in mammalian cell [8]. This hyperphosphorylation is an early event in neurodegenerative conditions, such as Alzheimer's disease [9], also involved is a second kinase, called CDK-5 [10].

Two GSK-3 genes (α and β) have been cloned from vertebrates. The intrinsic biochemical properties of GSK-3 are also conserved, with greater differences between GSK-3 α and GSK-3 β than between species [11].

Interest in GSK-3 has grown far beyond glycogen metabolism during the past decade and GSK-3 is known to occupy a central stage in many cellular and physiological events, including Wnt and Hedgehog signaling, transcription, insulin action, cell-division cycle, response to DNA damage, cell death, cell survival, patterning and axial orientation during development, differentiation, neural functions, circadian rhythm and others. In 1988 Ishiguro *et al.* [12] isolated one enzyme when they were studying an extract of the brain, which showed the generation of paired helical filaments of Tau protein kinase, typical injury of Alzheimer's disease.

In this moment, there is an increasing interest in the evaluation of kinases from unicellular parasites and fungi as targets for potential new anti-parasitic drugs. The evolutionary difference between unicellular kinases and their human homologues might be sufficient to allow the design of parasite-specific inhibitors. The *Plasmodium falciparum* genome contains 65 genes that encode kinases, including three forms of GSK-3. *P. falciparum* is one of the *Plasmodium* species that cause malaria in humans. It is transmitted by the female *Anopheles* mosquito. Malaria has the highest rates of complications and mortality; for example in 2006 it accounted for 91% of all 247 million human malaria infections (98% in Africa) and 90% of the deaths.

Leishmania is a genus of trypanosome protozoa, and is the parasite responsible for the disease leishmaniasis. *Leishmania* commonly infects hyraxes, canids, rodents and humans; and currently affects 12 million people in the world. There are many drawbacks to current chemotherapy for leishmaniasis, including problems of low efficacy, severe toxic side effects, and emerging drug resistance [13-15].

Leishmania parasites possess a complex life cycle in which the parasite passes between the sandfly vector and the mammalian host, during which time the parasite oscillates between rapidly dividing and cell cycle-arrested forms. The cell cycle of *Leishmania* is closely regulated, as in other eukaryotes, and integrated with its differentiation between the various life cycle stages.

Trypanosomes are a group of kinetoplastic protozoa distinguished by having only a single flagellum. All members are exclusively parasitic, found primarily in insects. A few genera have life-cycles involving a secondary host, which may be a vertebrate or a plant. These include several species that cause major diseases in humans. African Trypanosomiasis disease is caused by members of the

Trypanosma brucei complex, is a serious health threat. It is estimated that 300,000 to 500,000 humans in sub-Saharan African are infected. Despite the critical need, the available therapies are becoming less satisfactory due to the rising level of resistance to the available drugs, the long period of treatment required to achieve a cure, and the unacceptable and sometimes severe adverse effects associated with current therapies [16]. However, optimization of the selectivity of drug candidates for parasite kinases becomes an issue due to the highly conserved amino acids and protein conformation of the catalytic domains [17-20]. Our main objective in this work was to identify numerous examples of anti-parasitic, anti-fungi, *etc.* and GSK-3 inhibitors in order to obtain a model for the optimization of selected target inhibitors for drug development [21,22].

There are more than one thousand theoretical descriptors available in the literature to represent molecular structures, and one usually faces the problem of selecting those which are the most representative for any particular property under consideration. Topological indices [23-33], the most commonly used molecular descriptors, have been widely used in the correlation of physicochemical properties of organic compounds. In chemical graph theory, molecular structures are normally represented as hydrogen depleted graphs, whose vertices and edges act as atoms and covalent bonds, respectively. Chemical structural formulas can be then assimilated to undirected and finite multigraphs with labeled vertices, commonly known as molecular graphs. Topological indices, also known as graph theoretical indices, are descriptors that characterize molecular graphs and contain a large amount of information about the molecule, including the numbers of hydrogen and non-hydrogen atoms bonded to each non-hydrogen atom, the details of the electronic structure of each atom, and the molecular structural features [34]. In this article, we use ModesLab software [35] for calculating Topological indices (see Table 1), which is very important in the area of development of molecular descriptors and its applications to quantitative structure-property (QSPR), quantitative structure-activity (QSAR) relationship and drug design. ModesLab also provides a very useful way to define the properties of atoms, bonds and fragments by an extension of SMILES language and use these properties in molecular descriptors calculations.

Table 1. Topological Indices used in the present study.

Index	Description
$\chi(P), \chi(C), \chi(PC), \chi(Ch)$	Randic branching index
$\chi_v(P), \chi_v(C), \chi_v(PC), \chi_v(Ch)$	Valence connectivity
$e(P), e(C), e(PC), e(Ch)$	Epsilon index
${}^1\kappa, {}^2\kappa, {}^3\kappa$	Kappa index
${}^1\kappa(\alpha), {}^2\kappa(\alpha), {}^3\kappa(\alpha)$	Kappa (alpha) index
\emptyset	Flexibility index
M_1	Zagreb M_1 index
M_2	Zagreb M_2 index
H	Harary number
J	Balaban index

The development of QSARs using simple molecular indices appears to be a promising alternative or complementary technique to drug-protein docking, high-throughput screening and combinatorial chemistry techniques. Almost all QSAR techniques are based on the use of molecular descriptors,

which are numerical series that codify useful chemical information and enable correlations between statistical and biological properties [36-38].

A large number of examples have been published in which the use of molecular descriptors has become in a rational alternative to massive synthesis and screening of compounds in medicinal chemistry [39,40]. The principal deficiency in the use of some molecular indices concerns their lack of physical meaning.

2. Results and Discussion

2.1. General QSAR for GSK-3 Inhibitors

The development of a discriminant function [41] that allows the classification of organic compounds as active or non-active is the key step in the present approach for the discovery of GSK-3 inhibitors. It was therefore necessary to select a training data set of GSK-3 inhibitors containing wide structural variability. Linear Discriminant Analysis (LDA) was used to construct the classifiers. One of the most important steps in this work was the organization of the spreadsheet containing the raw data used as input for the LDA because this is not a classic classifier. Herein, the schematisation of the paper is peculiar. Our expectation is to use a two-group discriminant function to classify compounds into two possible groups: compounds that belong to a particular group and compounds that do not belong to this group. To this end, we have to indicate somehow what group we pretend to predict in each case. In this regard, we made the following steps, these steps are essentially the same given by Concu *et al.* [42,43] for their QSAR study of six classes of enzymes. In this study, each compound may be assayed on q^{th} different sets of conditions in the pharmacological tests, which are defined as Compound Assay Conditions query (CAC_q). These conditions are indicated in the supplementary material. The selection here of discriminant techniques instead of regression techniques was determined by the lack of homogeneity in the conditions under which these values were measured. As reported in different sources, numerous IC₅₀ values lie within a range rather than a single value. In other cases, the activity is not scored in terms of IC₅₀ values but is quoted as inhibitory percentages at a given concentration. Once the training series had been designed, forward-stepwise Linear Discriminant Analysis (LDA) was carried out in order to derive the QSAR (see **Equation 1**):

$$\begin{aligned}
 i - \text{GSK} - 3_{\text{TI}} = & 12.2755 \cdot {}^6\chi(\text{Ch}) + 5.9786 \cdot {}^5e(\text{Ch}) + 0.243 \cdot {}^1\kappa - 0.8553 \cdot \phi \\
 & + 1.9549 \cdot {}^3\chi(\text{C})_{\text{dif}} - 1.2747 \cdot {}^5\chi_v(\text{P})_{\text{dif}} - 43.1819 \cdot {}^4e(\text{C})_{\text{dif}} \quad (1) \\
 & - 1.2034 \cdot {}^5e(\text{P})_{\text{dif}} + 3.705 \cdot {}^6e(\text{P})_{\text{dif}} - 0.5713 \cdot {}^6e(\text{pC})_{\text{dif}} - 0.2127 \\
 \lambda = & 1.00 \quad F = 163.45 \quad p < 1.00
 \end{aligned}$$

The statistical significance of this model was determined by examining Wilk's λ statistic, Fisher ratio (F), and the p-level (p). The model is based on two types of parameters, the first type are parameters for single molecules. The type one includes first fourth parameters in the model. The following parameters presented in Table 1: randic branching index ($\chi(\text{Ch})$, $\chi(\text{C})$), epsilon index ($e(\text{Ch})$, $e(\text{C})$, $e(\text{P})$, $e(\text{pC})$), kappa index (κ), flexibility index (ϕ) and valence connectivity ($\chi_v(\text{P})$). In addition, we can see in the model parameters that quantify the difference between the structure of the drug and the structure of the drugs active for a given set of conditions CAC_q (see the last six parameters in the

Equation 1a-1f). We quantify this information in terms of the difference between the descriptors (Ds) of the drug and the average of Ds of active drugs for a given condition (see Methods section). The exact formulas for these terms present in the model are:

$${}^3\chi(C)_{dif} = {}^3\chi(C) - {}^3\bar{\chi}(C) \quad (1a)$$

$${}^5\chi_v(P)_{dif} = {}^5\chi_v(P) - {}^5\bar{\chi}_v(P) \quad (1b)$$

$${}^4e(C)_{dif} = {}^4e(C) - {}^4\bar{e}(C) \quad (1c)$$

$${}^5e(P)_{dif} = {}^5e(P) - {}^5\bar{e}(P) \quad (1d)$$

$${}^6e(P)_{dif} = {}^6e(P) - {}^6\bar{e}(P) \quad (1e)$$

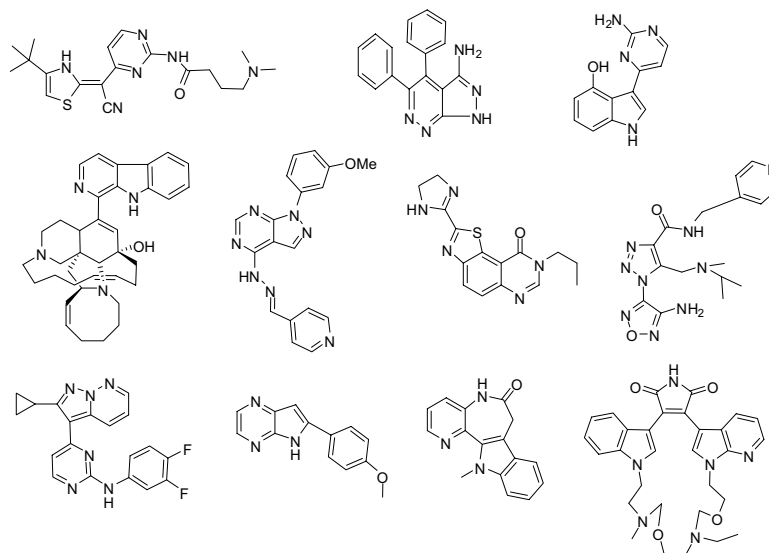
$${}^6e(pC)_{dif} = {}^6e(pC) - {}^6\bar{e}(pC) \quad (1f)$$

The introduction of this last parameter in the model (1) coincides with the i-GSK-3 activity observed in organic acid compounds [44] and drug connectivity may become an interesting alternative for fast computational pre-screening of large series of compounds in order to rationalize synthetic efforts [45-51] complementary to more elaborated techniques 3D-QSAR, CoMFA, and CoMSIA studies that depend on a detailed knowledge of 3D structure. In any case, these present models are of more general application than the other known methods that apply only to compounds tested in only one CAC and/or belonging to only one homogeneous structural class of compounds. Confirmation of this statement comes from the fact that the present classification function has given rise to an efficient separation of all compounds: with Accuracy = 99.1% in training series and Accuracy = 86.8% in validation series for the topological function, see Table 2 for details. The names, observed classification, predicted classification and subsequent probabilities for all 3,370 compounds in training and average validation are given as supplementary material. This level of total Accuracy, Sensitivity and Specificity is considered as excellent by other researches that have used LDA for QSAR studies and taking into account the great variety of compounds (see Figure 1), due to the fact that their structures are very different; see for instance the works of Garcia-Domenech, Prado-Prado and Marrero-Ponce *et al.* [52-67].

Table 2. Training and validation results.

Group	Parameter	%	GSKI-3	Non-active
<i>Training</i>				
GSKI-3	Sensitivity	95.3	854	42
Non-active	Specificity	82.8	77	371
Total	<i>Accuracy</i>	91.1		
<i>Validation</i>				
GSKI-3	Sensitivity	95.3	282	14
Non-active	Specificity	84.6	179	985
Total	<i>Accuracy</i>	86.8		

Figure 1. Some compounds studied in this article.

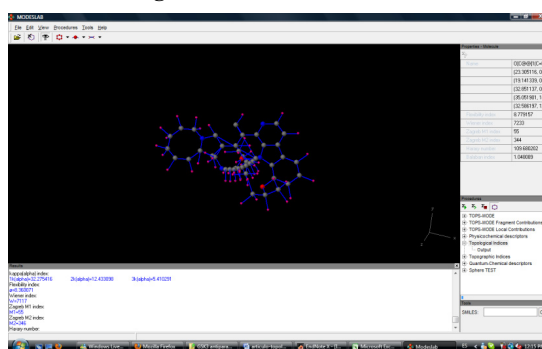


3. Materials and Methods

3.1. Computational Methods

The dataset formed by 3,370 cases was divided into 43 groups depending of their GSK-3 inhibitory activity and antiparasitic, antifungi, *etc.* activity. The model was obtained with topological descriptors. ModesLab (Molecular Descriptor Laboratory) version 1.5 software [35] was used to calculate all descriptors (see Figure 2).

Figure 2. ModesLab software.



The total analyzed variables were 189. The variables ${}^3\chi(\text{Ch})$, ${}^4\chi(\text{Ch})$, ${}^3\chi_v(\text{Ch})$, ${}^4\chi_v(\text{Ch})$, ${}^3e(\text{Ch})$, ${}^4e(\text{Ch})$, ${}^3\chi(\text{Ch})_{\text{avg}}$, ${}^4\chi(\text{Ch})_{\text{avg}}$, ${}^3\chi_v(\text{Ch})_{\text{avg}}$, ${}^4\chi_v(\text{Ch})_{\text{avg}}$, ${}^3e(\text{Ch})_{\text{avg}}$, ${}^4e(\text{Ch})_{\text{avg}}$, ${}^3\chi(\text{Ch})_{\text{dif}}$, ${}^4\chi(\text{Ch})_{\text{dif}}$, ${}^3\chi_v(\text{Ch})_{\text{dif}}$, ${}^4\chi_v(\text{Ch})_{\text{dif}}$, ${}^3e(\text{Ch})_{\text{dif}}$, ${}^4e(\text{Ch})_{\text{dif}}$ was eliminated of the database because all of them were equal to 0 and

constants. The quality of the model was determined by examining the Wilk's statistic, the square of Mahalanobis distance (D^2), the Fisher ratio (F) and the number of variables in the equation. Discrimination functions were obtained by using the forward-stepwise linear discriminant analysis as implemented in Statistica 6.0.

3.2. Multi-target Linear Discriminant Analysis (LDA)

In this regard, we performed the following steps, which are essentially the same given by Concu *et al.* [42,43] for their QSAR study of six classes of enzymes:

- (1). We created a raw data representing each compound input as a vector made up of 1 output variable, 189 structural variables (inputs) divided in values (see the first term of the Equation (1)), averages (see the second term of the Equation (1)) and differences between values and averages (see the third term of the Equation (1)); and the CAC_q variable. CAC_q is an auxiliary not used to construct the model.
- (2). The first element (output) is a dummy variable (Boolean) called Observed Group (OG); OG = 0 if the compound belongs to the class to which we refer in CAC_q and 1 otherwise (OG = 1). We could repeat each compound more than once in the raw data. In fact, we could repeat each compound 43 times corresponding to 43 CAC_q Assay Conditions (see **Table 3**). The first time we used the $CAC_q = CAC$ number. It means that we used the real CAC class of the compound in CAC_q . In this case, the LDA model had to give the highest probability to the group OG = 0 because it had to predict the real class of the compound. The remnant 43 times we use an CAC class number different to the real in CAC_q and then the LDA model had to predict the highest probability for the group OG = 1. This indicated that the compound did not belong to this group.

Table 3. Compound Assay Conditions query (CAC_q).

Group	Parameter	Enzyme	Isoform	Enzyme/ Organism	Species	Activity (1,0)	Condition	Observ.
1	%	GSK-3	beta	enzyme	no	0	0	=
2	cKi	GSK-3	alfa	enzyme	no	0	100	>
3	EC ₅₀ (μM)	no	no	no	Cell Efficacy	0	2	>
4	EC ₅₀ (μM)	no	no	no	glycogen synthesis stimulation	0	inactive	=
5	EC ₅₀ (μM)	no	no	no	β-catenin synthesis	0	2	>
6	EC ₅₀ (μM)	no	no	parasite	<i>T. brucei</i>	0	2	>
7	EC ₅₀ (μM)	no	no	virus	HIV-1	0	NA	=
8	ED ₅₀ (μM)	no	no	parasite	<i>L. donovani</i>	0	5	<
9	IC ₅₀ (ng/mL)	no	no	parasite	<i>P. falciparum</i> (chloroquine resistant W2 clone)	0	NA	=
10	IC ₅₀ (ng/mL)	no	no	parasite	<i>P. falciparum</i> (chloroquine sensitive D6 clone)	0	NA	=

Table 3. Cont.

11	IC ₅₀ (nM)	GSK-3	alfa	enzyme	no	0	2000	>
12	IC ₅₀ (nM)	GSK-3	beta	enzyme	no	0	2000	>
13	IC ₅₀ (nM)	GSK-3	nd	enzyme	no	0	2000	>
14	IC ₅₀ (µg/mL)	no	no	bacterium	<i>M. intracellulare</i>	0	NA	=
15	IC ₅₀ (µg/mL)	no	no	bacterium	MRS	0	NA	=
16	IC ₅₀ (µg/mL)	no	no	bacterium	<i>S. aureus</i>	0	NA	=
17	IC ₅₀ (µg/mL)	no	no	cell line	Human Vero cells	0	NC	<
18	IC ₅₀ (µg/mL)	no	no	fungus	<i>C. neoformans</i>	0	NA	=
19	IC ₅₀ (µg/mL)	no	no	parasite	<i>L. donovani</i>	0	NA	=
20	IC ₅₀ (µM)	GSK-3	beta	enzyme	no	0	2	>
21	IC ₅₀ (µM)	GSK-3	nd	enzyme	no	0	2	>
22	IC ₅₀ (µM)	GSK-3	no	parasite	<i>P. falciparum</i>	0	20	>
23	IC ₅₀ (µM)	GSK-3	α/β	enzyme	no	0	2	>
24	IC ₅₀ (µM)	no	no	bacterium	<i>M. intracellulare</i>	0	—	=
25	IC ₅₀ (µM)	no	no	bacterium	MRSA	0	—	=
26	IC ₅₀ (µM)	no	no	cell line	Hep2	0	NA	=
27	IC ₅₀ (µM)	no	no	cell line	HT29	0	NA	=
28	IC ₅₀ (µM)	no	no	cell line	Human Vero cells	0	NA	=
29	IC ₅₀ (µM)	no	no	cell line	Human Vero cells	0	NC	<
30	IC ₅₀ (µM)	no	no	cell line	LMM3	0	NA	=
31	IC ₅₀ (µM)	no	no	cell line	PTP	0	2	>
32	IC ₅₀ (µM)	no	no	fungus	<i>C. albicans</i>	0	—	=
33	IC ₅₀ (µM)	no	no	fungus	<i>C. neoformans</i>	0	—	=
34	IC ₅₀ (µM)	no	no	parasite	<i>L. mexicana</i>	0	2	>
35	IC ₅₀ (µM)	no	no	parasite	<i>P. falciparum</i>	0	2	>
36	IC ₅₀ (µM)	no	no	parasite	<i>P. falciparum</i> D6	0	NA	=
37	IC ₅₀ (µM)	no	no	parasite	<i>P. falciparum</i> W2	0	NA	=
38	IC ₅₀ E-9 (M)	GSK-3	nd	enzyme	no	0	2	>
39	IC ₉₀ (µg/mL)	no	no	parasite	<i>L. donovani</i>	0	NA	=
40	MIC (µg/mL)	no	no	bacterium	<i>M. tuberculosis</i> (H37Rv)	0	NA	=
41	pIC ₅₀	GSK-3	beta	enzyme	no	0	0	=
42	pIC ₅₀	GSK-3	nd	enzyme	no	0	0	=
43	IC ₅₀ (µM)	no	no	no	Cell Efficacy	0	2	>

The problem in this type of organization of raw data is that the descriptor values are compound constants. Consequently, if these latter and LDA are based only on these values, they will necessarily fail when we change OG values. An inconvenient in this regard occurs if we pretend to use the model

for a real enzyme, since we have only one unspecific prediction and we need 43 specific probabilities, one confirming the real class and 42 giving low probabilities for the other CAC_q. We can solve this problem introducing variables characteristic of each CAC class referred on the CAC_q but without giving information in the input about the real CAC class of the protein. To this end, we used the average value of each descriptor for all enzymes that belonged to the same CAC class. We also calculated the deviation of all the descriptors from the respective group indicated in CAC_q. Altogether, we have then (63 topological descriptor values) + (63 topological descriptor values average values for CAC class) + (63 topological descriptor values deviations values for CAC class average) = 189 input variables. It is of major importance to understand that we never used as input CAC_q, so the model only includes as input the physicochemical-topological descriptor values for the protein entry and the average and deviations of these values from the CAC_q, which is not necessarily the real CAC class. The general formula for this class of LDA model is shown below (see **Equation 2**), where S(E) is not the probability but a real valued score that predicts the propensity of a compound to act as an inhibitor of a given class:

$$\begin{aligned} S(E) &= \sum_{k,G,D_i} b_k \cdot {}^k D_i(G) + \sum_{k,G,D} c_k \cdot {}^k \bar{D}_i(G) + \sum_{k,G,D} d_k \cdot ({}^k D_i(G) - {}^k \bar{D}_i(G)) + a_0 \\ &= \sum_{k,G,D_i} b_k \cdot {}^k D_i(G) + \sum_{k,G,D} c_k \cdot {}^k \bar{D}_i(G) + \sum_{k,G,D} d_k \cdot \Delta^k D_i(G) + a_0 \end{aligned} \quad (2)$$

In a compact notation we write ${}^k D_i(G)$, where D_i is the type of descriptor; G is the types of subgraphs studied in the molecular connectivity $G = \text{path } (p)$, clusters (C), path-clusters (pC) and chains (rings) (Ch).

LDA forward stepwise analysis was carried out for variable selection to build up the models [68]. All the variables included in the model were standardized in order to bring them onto the same scale. Subsequently, a standardized linear discriminant equation that allows comparison of their coefficients was obtained [69]. The square of Canonical regression coefficient (Rc) and Wilk's statistics (U) were examined in order to assess the discriminatory power of the model ($U = 0$ perfect discrimination, being $0 < U < 1$); the separation of the two groups was statistically verified by the Fisher ratio (F) test with an error level $p < 0.05$.

3.3. Data Set

The data set was conformed to a set of marketed and/or reported drugs/receptor pairs where affinity/non-affinity of drugs with the receptors was established taking into consideration the IC50, k_i , pki , values. In consequence, we managed to collect 1,192 examples of active compounds in different CAC_q. In addition, we used a negative control series of 2,178 cases of non-active compounds in different CAC_q. In the two data sets used, there were the following training series: 474 active compounds plus 896 non-active compounds (1,370 in total), predicting series: 296 + 1,164 = 1,460 in total. Due to space constraints the names or codes for all compounds are listed in supplementary material **SM1** in the Supporting Information, as well as the references consulted to compile the data in this table. This series is composed at random by the most representative families of GSK-3 inhibitors taken from the literature (supplementary material **SM2**). The remaining compounds were a

heterogeneous series of inactive compounds, including members of the aforementioned families and compounds included in the Merck index [70].

4. Conclusions

In this work we have shown that the ModesLab methodology using topological indices can be considered a good alternative for developing GSK-3 inhibitors in a fast and efficient way with respect to other methods of the literature. This approach is able to correctly classify the GSK-3 inhibitory activity of compounds with different structural patterns.

Acknowledgements

We are grateful to the Xunta de Galicia (INCITE08PXIB314255PR) for partial financial support and to Estrada, E. for donation of this valuable tool (ModesLab software) for the realization of this work.

References

1. Olson, R.E. Secretase inhibitors as therapeutics for Alzheimer's disease. *Annu. Rep. Med. Chem.* **2000**, *35*, 31-40.
2. Brion, J.P.; Anderton, B.H.; Authélet, M.; Dayanandan, R.; Leroy, K.; Lovestone, S.; Ocatve, J.N.; Pradier, L.; Touchet, N.; Tremp, G. Neurofibrillary Tangles and Tau Phosphorylation. *Biochem. Soc. Symp.* **2001**, *67*, 81-88.
3. Braak, H.; Braak, E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* **1991**, *82*, 239-259
4. Iqbal, K.; Alonso, A.; Chen, S.; Chohan, M.O.; El-Akkad, E.; Gong, C.; Khatoon, S.; Li, B.; Liu, F.; Rahman, A.; TAnimukai, H.; Grundke-Iqbal, I. Tau pathology in Alzheimer disease and other tauopathies. *BBA-Mol. Basis Dis.* **2005**, *1739*, 198-210.
5. Cohen, P.; Frame, S. The renaissance of GSK3. *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 769-776.
6. Woodgett, J.R.; Cohen, P. Multisite phosphorylation of glycogen synthase. Molecular basis for the substrate specificity of glycogen synthase kinase-3 and casein kinase-II (glycogen synthase kinase-5). *Biochim. Biophys. Acta* **1984**, *788*, 339-347.
7. Nikoulina, S.E.; Ciaraldi, T.P.; Mudailar, S.; Mohideen, P.; Carter, L.; Henry, R.R. Potential role of glycogen synthase kinase-3 in skeletal muscle insulin resistance of type 2 diabetes. *Diabetes* **2000**, *49*, 263-271.
8. Lovestone, S.; Reynolds, C.H.; Latimer, D.; Davis, D.R.; Anderton, B.H.; Gallo, J.M.; Hanger, D.; Mulot, S.; Marquardt, B. Alzheimer's disease-like phosphorylation of the microtubule-associated protein tau by glycogen synthase kinase-3 in transfected mammalian cells. *Curr. Biol.* **1994**, *4*, 1077-1086.
9. Imahori, K.; Uchida, T. Physiology and pathology of tau protein kinases in relation to Alzheimer's disease. *J. Biochem.* **1997**, *121*, 179-188.

10. Takashima, A.; Murayama, M.; Yasutake, K.; Takahashi, H.; Yokoyama, M.; Ishiguro, K. Involvement of cyclin dependent kinase5 activator p25 on tau phosphorylation in mouse brain. *Neurosci. Lett.* **2001**, *306*, 37-40.
11. Ryves, W.J.; Harwood, A.J. Lithium Inhibits Glycogen Synthase Kinase-3 by Competition for Magnesium. *Biochem. Biophys. Res. Commun.* **2001**, *280*, 720-725.
12. Ishiguro, K.; Ihara, Y.; Uchida, T.; Imahori, K.A. Novel Tubulin-Dependent Protein Kinase Forming a Paired Helical Filament Epitope on Tau. *J. Bio. Chem.* **1988**, *104*, 319-321.
13. Arana, B.; Rizzo, N.; Diaz, A. Chemotherapy of cutaneous leishmaniasis: A review. *Med. Microbiol. Immunol.* **2001**, *190*, 93-95.
14. Bryceson, A. Current issues in the treatment of visceral leishmaniasis. *Med. Microbiol. Immunol.* **2001**, *190*, 85-87.
15. Sundar, S. Treatment of visceral leishmaniasis. *Med. Microbiol. Immunol.* **2001**, *190*, 89-92.
16. Fairlamb, A.H. Chemotherapy on human African trypanosomiasis: Current and future prospects. *Trends Parasitol.* **2003**, *19*, 488-494.
17. Copeland, R.A.; Pompliano, D.L.; Meek, T.D. Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discov.* **2006**, *5*, 730-732.
18. Liao, J.J. Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J. Med. Chem.* **2007**, *50*, 409-424.
19. Pink, R.; Hudson, A.; Mouries, M.A.; Bending, M. Opportunities and challenges in antiparasitic drug discovery. *Nat. Rev. Drug Discov.* **2005**, *4*, 727-740.
20. Plyte, S.E.; Hughes, K.; Nilkolakaki, E.; Pulverer, B.J.; Woodgett, J.R. Glycogen Synthase Kinase-3: Functions in oncogenesis and development. *Biochim. Biophys. Acta* **1992**, *1114*, 147-162.
21. Dajani, R.; Fraser, E.; Roe, S.M.; Young, N.; Good, V.; Dale, T.C.; Pearl, L.H. Crystal structure of glycogen synthase kinase-3 beta: Structural basis for phosphate-primed substrate specificity and autoinhibition. *Cell (Cambridge, Mass.)* **2001**, *105*, 721-732.
22. Ojo, K.K.; Gillespie, R.G.; Riechers, A.; Napuli, A.J.; Verlinde, C.L.; Buckner, F.S.; Gelb, M.H.; Domostoj, M.M.; Wells, S.J.; Scheer, A.; Wells, T.N.C.; Voorhis, C.V. Glycogen Synthase Kinase 3 is a potential drug target for african trypanosomiasis therapy. *Antimicrob. Agents Chemother.* **2008**, *52*, 3710-3717.
23. González-Díaz, H.; Munteanu, C.R. *Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and Social Networks*; Transworld Research Network: Kerala, India, 2010.
24. Gonzalez-Díaz, H.; Prado-Prado, F.; Ubeira, F.M. Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach. *Curr. Top Med. Chem.* **2008**, *8*, 1676-1690.
25. González-Díaz, H.; González-Díaz, Y.; Santana, L.; Ubeira, F.M.; Uriarte, E. Proteomics, networks and connectivity indices. *Proteomics* **2008**, *8*, 750-778.
26. González-Díaz, H.; Vilar, S.; Santana, L.; Uriarte, E. Medicinal chemistry and bioinformatics – current trends in drugs discovery with networks topological indices. *Curr. Top Med. Chem.* **2007**, *7*, 1015-1029.

27. Gonzalez-Diaz, H.; Duardo-Sanchez, A.; Ubeira, F.M.; Prado-Prado, F.; Perez-Montoto, L.G.; Concu, R.; Podda, G.; Shen, B. Review of MARCH-INSIDE & complex networks prediction of drugs: ADMET, anti-parasite activity, metabolizing enzymes and cardiotoxicity proteome biomarkers. *Curr. Drug Metab.* **2010**, *11*, 379-406.
28. Helguera, A.M.; Combes, R.D.; Gonzalez, M.P.; Cordeiro, M.N. Applications of 2D descriptors in drug design: A DRAGON tale. *Curr. Top Med. Chem.* **2008**, *8*, 1628-1655.
29. Caballero, J.; Fernandez, M. Artificial neural networks from MATLAB in medicinal chemistry. Bayesian-regularized genetic neural networks (BRGNN): Application to the prediction of the antagonistic activity against human platelet thrombin receptor (PAR-1). *Curr. Top Med. Chem.* **2008**, *8*, 1580-1605.
30. Vilar, S.; Cozza, G.; Moro, S. Medicinal chemistry and the molecular operating environment (MOE): Application of QSAR and molecular docking to drug discovery. *Curr. Top Med. Chem.* **2008**, *8*, 1555-1572.
31. Khan, M. T. Predictions of the ADMET properties of candidate drug molecules utilizing different QSAR/QSPR modelling approaches. *Curr. Drug Metab.* **2010**, *11*, 285-295.
32. Garcia, I.; Diop, Y.F.; Gomez, G. QSAR & complex network study of the HMGR inhibitors structural diversity. *Curr. Drug Metab.* **2010**, *11*, 307-314.
33. Martinez-Romero, M.; Vazquez-Naya, J.M.; Rabunal, J.R.; Pita-Fernandez, S.; Macenlle, R.; Castro-Alvarino, J.; Lopez-Roses, L.; Ulla, J.L.; Martinez-Calvo, A.V.; Vazquez, S.; *et al.* Artificial intelligence techniques for colorectal cancer drug metabolism: Ontology and complex network. *Curr. Drug Metab.* **2010**, *11*, 347-368.
34. Mao, B.Y.; Chou, K.C.; Maggiora, G.M. Topological analysis of hydrogen bonding in protein structure. *Eur. J. Biochem.* **1990**, *188*, 361-365.
35. Estrada, E.; Gutiérrez, Y. *ModesLab*, versión 1.5, 2002-2004.
36. Nunez, M.B.; Maguna, F.P.; Okulik, N.B.; Castro, E.A. QSAR modeling of the MAO inhibitory activity of xanthones derivatives. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5611-5617.
37. Todeschini, R.; Consonni, V. *Handbook of molecular descriptors*; Wiley-VCH: Weinheim, Germany, 2000.
38. Freund, J.A.; Poschel, T. Stochastic processes in physics, chemistry, and biology. In *Lect. Notes Phys.*; Springer-Verlag: Berlin, Germany, 2000.
39. Estrada, E.; Uriarte, E. Recent advances on the role of topological indices in drug discovery research. *Curr. Med. Chem.* **2001**, *8*, 1573-1588.
40. Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. A Novel Approach for the Virtual Screening and Rational Design of Anticancer Compounds. *J. Med. Chem.* **2001**, *43*, 1975-1985.
41. Van Waterbeemd, H. Discriminant Analysis for Activity Prediction. In *Chemometric Methods in Molecular Design*; Van Waterbeemd, H., Ed.; Wiley-VCH: New York, NY, USA, 1995; Volume 2, pp. 265-282.
42. Concu, R.; Dea-Ayuela, M.A.; Perez-Montoto, L.G.; Prado-Prado, F.J.; Uriarte, E.; Bolas-Fernandez, F.; Podda, G.; Pazos, A.; Munteanu, C.R.; Ubeira, F.M.; Gonzalez-Diaz, H. 3D entropy and moments prediction of enzyme classes and experimental-theoretic study of peptide fingerprints in Leishmania parasites. *Biochim. Biophys. Acta.* **2009**, *1794*, 1784-1794.

43. Concu, R.; Dea-Ayuela, M.A.; Perez-Montoto, L.G.; Bolas-Fernandez, F.; Prado-Prado, F.J.; Podda, G.; Uriarte, E.; Ubeira, F.M.; Gonzalez-Diaz, H. Prediction of Enzyme Classes from 3D Structure: A General Model and Examples of Experimental-Theoretic Scoring of Peptide Mass Fingerprints of Leishmania Proteins. *J. Proteome Res.* **2009**, *8*, 4372-4382.
44. Konda, V.R.; Desai, A.; Darland, G.; Bland, J.S.; Tripp, M.L. Rho iso-alpha acids from hops inhibit the GSK-3/NF-kappaB pathway and reduce inflammatory markers associated with bone and cartilage degradation. *J. Inflamm.* **2009**, *6*, 26-34.
45. Rochais, C.; Duc, N.V.; Lescot, E.; Sopkova-de Oliveira Santos, J.; Bureau, R.; Meijer, L.; Dallemagne, P.; Rault, S. Synthesis of new dipyrrolo- and furopyrrolopyrazinones related to tripentones and their biological evaluation as potential kinases (CDKs1-5, GSK-3) inhibitors. *Eur. J. Med. Chem.* **2009**, *44*, 708-716.
46. Simon, D.; Benitez, M.J.; Gimenez-Cassina, A.; Garrido, J.J.; Bhat, R.V.; Diaz-Nido, J.; Wandosell, F. Pharmacological inhibition of GSK-3 is not strictly correlated with a decrease in tyrosine phosphorylation of residues 216/279. *J. Neurosci. Res.* **2008**, *86*, 668-674.
47. Patel, D.S.; Bharatam, P.V. Selectivity criterion for pyrazolo[3,4-b]pyrid[az]ine derivatives as GSK-3 inhibitors: CoMFA and molecular docking studies. *Eur. J. Med. Chem.* **2008**, *43*, 949-957.
48. Jacquemard, U.; Dias, N.; Lansiaux, A.; Bailly, C.; Loge, C.; Robert, J.M.; Lozach, O.; Meijer, L.; Merour, J.Y.; Routier, S. Synthesis of 3,5-bis(2-indolyl)pyridine and 3-[(2-indolyl)-5-phenyl]pyridine derivatives as CDK inhibitors and cytotoxic agents. *Bioorg. Med. Chem.* **2008**, *16*, 4932-4953.
49. Xiao, J.; Guo, Z.; Guo, Y.; Chu, F.; Sun, P. Inhibitory mode of N-phenyl-4-pyrazolo[1,5-b]pyridazin-3-ylpyrimidin-2-amine series derivatives against GSK-3: Molecular docking and 3D-QSAR analyses. *Protein Eng. Des. Sel.* **2006**, *19*, 47-54.
50. Tavares, F.X.; Boucheron, J.A.; Dickerson, S.H.; Griffin, R.J.; Preugschat, F.; Thomson, S.A.; Wang, T.Y.; Zhou, H.Q. N-Phenyl-4-pyrazolo[1,5-b]pyridazin-3-ylpyrimidin-2-amines as potent and selective inhibitors of glycogen synthase kinase 3 with good cellular efficacy. *J. Med. Chem.* **2004**, *47*, 4716-4730.
51. Olesen, P.H.; Sorensen, A.R.; Urso, B.; Kurtzhals, P.; Bowler, A.N.; Ehrbar, U.; Hansen, B.F. Synthesis and *in vitro* characterization of 1-(4-aminofurazan-3-yl)-5-dialkylaminomethyl-1H-[1,2,3]triazole-4-carboxylic acid derivatives. A new class of selective GSK-3 inhibitors. *J. Med. Chem.* **2003**, *46*, 3333-3341.
52. Calabuig, C.; Anton-Fos, G.M.; Galvez, J.; Garcia-Domenech, R. New hypoglycaemic agents selected by molecular topology. *Int. J. Pharm.* **2004**, *278*, 111-118.
53. Garcia-Garcia, A.; Galvez, J.; de Julian-Ortiz, J.V.; Garcia-Domenech, R.; Munoz, C.; Guna, R.; Borrás, R. New agents active against Mycobacterium avium complex selected by molecular topology: A virtual screening method. *J. Antimicrob. Chemother.* **2004**, *53*, 65-73.
54. Prado-Prado, F.J.; Ubeira, F.M.; Borges, F.; Gonzalez-Diaz, H. Unified QSAR & network-based computational chemistry approach to antimicrobials. II. Multiple distance and triadic census analysis of antiparasitic drugs complex networks. *J. Comput. Chem.* **2009**, *31*, 164-173.

55. Prado-Prado, F.J.; Martinez de la Vega, O.; Uriarte, E.; Ubeira, F.M.; Chou, K.C.; Gonzalez-Diaz, H. Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modeling and comparative multi-distance study of the giant components of antiviral drug-drug complex networks. *Bioorg. Med. Chem.* **2009**, *17*, 569-575.
56. Prado-Prado, F.J.; de la Vega, O.M.; Uriarte, E.; Ubeira, F.M.; Chou, K.C.; Gonzalez-Diaz, H. Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modeling and comparative multi-distance study of the giant components of antiviral drug-drug complex networks. *Bioorg. Med. Chem.* **2009**, *17*, 569-575.
57. Prado-Prado, F.J.; Borges, F.; Perez-Montoto, L.G.; Gonzalez-Diaz, H. Multi-target spectral moment: QSAR for antifungal drugs vs. different fungi species. *Eur. J. Med. Chem.* **2009**, *44*, 4051-4056.
58. Prado-Prado, F.J.; Gonzalez-Diaz, H.; de la Vega, O.M.; Ubeira, F.M.; Chou, K.C. 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.
59. Prado-Prado, F.J.; Gonzalez-Diaz, H.; 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.
60. Marrero-Ponce, Y.; Khan, M.T.; Casanola Martin, G.M.; Ather, A.; Sultankhodzhaev, M.N.; Torrens, F.; Rotondo, R. Prediction of Tyrosinase Inhibition Activity Using Atom-Based Bilinear Indices. *Chem. Med. Chem.* **2007**, *2*, 449-478.
61. Marrero-Ponce, Y.; Meneses-Marcel, A.; Castillo-Garit, J.A.; Machado-Tugores, Y.; Escario, J.A.; Barrio, A.G.; Pereira, D.M.; Nogal-Ruiz, J.J.; Aran, V.J.; Martinez-Fernandez, A.R.; Torrens, F.; Rotondo, R.; Ibarra-Velarde, F.; Alvarado, Y.J. Predicting antitrichomonal activity: A computational screening using atom-based bilinear indices and experimental proofs. *Bioorg. Med. Chem.* **2006**, *14*, 6502-6524.
62. Meneses-Marcel, A.; Marrero-Ponce, Y.; Machado-Tugores, Y.; Montero-Torres, A.; Pereira, D.M.; Escario, J.A.; Nogal-Ruiz, J.J.; Ochoa, C.; Aran, V.J.; Martinez-Fernandez, A.R.; *et al.* A linear discrimination analysis based virtual screening of trichomonacidal lead-like compounds: Outcomes of in silico studies supported by experimental results. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3838-3843.
63. Marrero-Ponce, Y.; Diaz, H.G.; Zaldivar, V.R.; Torrens, F.; Castro, E.A. 3D-chiral quadratic indices of the 'molecular pseudograph's atom adjacency matrix' and their application to central chirality codification: Classification of ACE inhibitors and prediction of sigma-receptor antagonist activities. *Bioorg. Med. Chem.* **2004**, *12*, 5331-5342.
64. Murcia-Soler, M.; Perez-Gimenez, F.; Garcia-March, F.J.; Salabert-Salvador, M.T.; Diaz-Villanueva, W.; Medina-Casamayor, P. Discrimination and selection of new potential antibacterial compounds using simple topological descriptors. *J. Mol. Graph. Model.* **2003**, *21*, 375-390.
65. Cercos-del-Pozo, R.A.; Perez-Gimenez, F.; Salabert-Salvador, M.T.; Garcia-March, F.J. Discrimination and molecular design of new theoretical hypolipaemic agents using the molecular connectivity functions. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 178-184.

66. Estrada, E.; Vilar, S.; Uriarte, E.; Gutierrez, Y. In silico studies toward the discovery of new anti-HIV nucleoside compounds with the use of TOPS-MODE and 2D/3D connectivity indices. 1. Pyrimidyl derivatives. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1194-1203.
67. Cronin, M.T.; Aptula, A.O.; Dearden, J.C.; Duffy, J.C.; Netzeva, T.I.; Patel, H.; Rowe, P.H.; Schultz, T.W.; Worth, A.P.; Voutzoulidis, K.; Schuurmann, G. Structure-based classification of antibacterial activity. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 869-878.
68. Santana, L.; Uriarte, E.; González-Díaz, H.; Zagotto, G.; Soto-Otero, R.; Mendez-Alvarez, E.A. QSAR model for in silico screening of MAO-A inhibitors. Prediction, synthesis, and biological assay of novel coumarins. *J. Med. Chem.* **2006**, *49*, 1149-1156.
69. Kutner, M.H.; Nachtsheim, C.J.; Neter, J.; Li, W. Standardized Multiple Regression Model. In *Applied Linear Statistical Models*, 5th ed.; McGraw Hill: New York, NY, USA, 2005; pp. 271-277.
70. Budavari, S. *The Merck Index*, 12th ed.; Merck & Co, Inc: Whitehouse Station, NJ, USA, 1996.

Sample Availability: Not available.

© 2010 by the authors; licensee MDPI, Basel, Switzerland. This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).



Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: <http://www.elsevier.com/locate/ejmech>

Original article

An application of two MIFs-based tools (Volsurf+ and Pentacle) to binary QSAR: The case of a palinurin-related data set of non-ATP competitive Glycogen Synthase Kinase 3 β (GSK-3 β) inhibitors

Giuseppe Ermondi^a, Giulia Caron^{a,*}, Isela Garcia Pintos^b, Michela Gerbaldo^a, Manuel Pérez^b, Daniel I. Pérez^c, Zoila Gándara^b, Ana Martínez^c, Generosa Gómez^b, Yagamare Fall^b

^aCASSMedChem Laboratory, DSTF at the Centre for Innovation, Università di Torino, Via Quarello 11, 10135 Torino, Italy

^bDepartamento de Química Orgánica, Facultad de Química, Universidade de Vigo, 36200 Vigo, Spain

^cInstituto de Química Médica, CSIC, Juan de la Cierva 3, 28006 Madrid, Spain

ARTICLE INFO

Article history:
Received 19 August 2010
Received in revised form
18 November 2010
Accepted 21 December 2010
Available online 9 January 2011

Keywords:
Alzheimer's disease
GRIND
MIFs
Palinurin
Binary QSAR
VolSurf descriptors

ABSTRACT

VolSurf+ and GRIND descriptors extract the information present in MIFs calculated by GRID: the first are simpler to interpret and generally applied to ADME-Tox topics, whereas the latter are more sophisticated and thus more suited for pharmacodynamics events. Here we present a study which compares binary QSAR models obtained with VolSurf+ descriptors and GRIND for a data set of non-ATP competitive GSK-3 β inhibitors chemically related to palinurin for which the biological activity is expressed in binary format. Results suggest not only that the simpler Volsurf+ descriptors are good enough to predict and chemically interpret the investigated phenomenon but also a bioactive conformation of palinurin which may guide future design of ATP non-competitive GSK-3 inhibitors.

© 2011 Elsevier Masson SAS. All rights reserved.

1. Introduction

Quantitative Structure–Activity Relationships (QSARs) strategies are widely used in medicinal chemistry [1]. Briefly, QSAR are methods for estimating a given biological activity of a chemical from its molecular structure [2]. Among others, two of the major concerns about QSAR are the sources of biological activities which should provide a high number of quality data and the choice of adequate molecular descriptors (MD) among the plethora of molecular determinants reported in the literature [1].

The automation of experiments through robotics to effectively perform hundreds of thousands of experiments in a short time is generally called High Throughput Screening (HTS). In recent years, the use of HTS has produced a large amount of biological data that classifies compounds as active or inactive (some HTS also report a discrete measure; e.g., activity on a scale from 1 to 10). As a consequence HTS technologies provide a source of semi-quantitative biological activities of good quality that could potentially be used to

build QSAR models that facilitate the identification of drug candidates [3].

Among the overabundance of *in silico* methodologies, the Molecular Interaction Fields (MIFs) based tools represent well suited methods to handle QSAR strategies [4]. MIFs derived descriptors are of particular interest in molecular discovery since they are conceived to describe molecular interactions of pharmaceutical nature. A MIF in fact describes the spatial variation of the interaction energy between a molecular target and a chosen probe and its calculation is mediated by the GRID software [5–7] which uses a potential based on the total interaction energy (Eq. (1)) between a target molecule and a probe (which may be an atom or a group)

$$E_{XYZ} = \text{Sum}[ELJ] + \text{Sum}[EHB] + \text{Sum}[EQ] + [S] \quad (1)$$

in which E_{XYZ} is the total energy of interaction of the selected target/probe couple, $\text{Sum}[]$ indicates pair wise energy summation (ELJ, EHB, and EQ are the Lennard–Jones, the hydrogen bond and the electrostatic terms, respectively) between the probe at its grid point and each appropriate atom of the target, and $[S]$ is the appropriate entropic term at the grid point.

* Corresponding author. Fax: +39 0112367282.
E-mail address: giulia.caron@unito.it (G. Caron).

MIFs are extremely rich in information [8,9] but consist of a large number of variables that are generated to describe the non-bonded interaction energies between one or more probes and each drug molecule. Much effort has been devoted to developing methods that optimize the selection of only those variables of importance [10]. In early attempts, MIFs value at the grind points were directly correlated with activities using Partial Least Squares (PLS) statistical techniques (GRID/GOLPE method [11]). GRID/GOLPE method is a 3D-QSAR approach similar to CoMFA [12] and requires the alignment of the molecules. The alignment is often the bottleneck in the whole computational study [4] and to avoid it two excellent commercially available software packages were recently designed and released to automatically extract the information present in MIFs in the form of numerical descriptors: Volsurf+ [13,14] and Pentacle [15,16] (a recent improvement upon the original ALMOND software) both by Molecular Discovery.

In general terms, Volsurf+ descriptors are obtained from MIFs by calculating the volume or the surface of the interaction contours [4,13,14] at predefined energy values whereas Pentacle descriptors (called GRIND) are the results of a filtering procedure based on energetic and distribution criteria and relative position of points.[15] As a result Volsurf+ descriptors represent an intermediate level of descriptors between the whole MIF and localised Pentacle GRIND which are expected to represent every single region of the MIF with a potential relevance in the interaction between the small molecule and the receptor. As a consequence of their definition that depends on the global characteristics of the molecules, Volsurf+ descriptors are well suited to describe some ADME properties, whereas modelling of pharmacological target-based interaction requires the use of descriptors able to catch the specificity of the interaction as GRIND.[4] The state of art of the applications of the two software seems to confirm this classification [4,13–17], but the comparison of

these two tools [18] remains a crucial step to shed light on the identification of the best method to describe a given biological event.

Currently, there is a significant evidence, both *in vitro* and *in vivo*, that Glycogen Synthase Kinase 3 β (GSK-3 β) plays a crucial role in neurodegeneration in general and AD in particular.[19] Because of these data, inhibition of GSK-3 β is accepted as a promising strategy for the treatment of AD and other neurodegenerative diseases.[20] Recently, Palinurin (**1** in Fig. 1) has emerged as a potent ATP non-competitive inhibitor of GSK-3 β [21]. Palinurin is a linear furanosesterterpene previously reported from the Mediterranean sponge *Ircinia variabilis* and was reisolated from a Red Sea sponge *Ircinia dendroides* [22]. The total synthesis of palinurin was recently carried out by some of us [23] and all the intermediates were systematically tested as inhibitors of GSK-3 β and some of them resulted active [24].

In this study we compare the QSAR models obtained with Volsurf+ and Pentacle for the aforementioned data set of non-ATP competitive GSK-3 β inhibitors chemically related to palinurin to check whether Volsurf+ descriptors could replace GRIND in the interpretation of pharmacodynamic events when these latter are expressed in binary format. The predictive skill of both models and their chemical interpretability are also discussed as well.

2. Results and discussion

2.1. General considerations

Palinurin (**1** in Fig. 1) is a linear sesterterpene characterized by a furan ring at one end and by a conjugated tetroneic acid at the other. Some intermediates investigated in this study bear a furan, others a tetroneic acid and many compounds out of the 58 bear one or two *tert*-butyldimethylsilyl (TBS, TBDMS) ethers as well as *tert*-

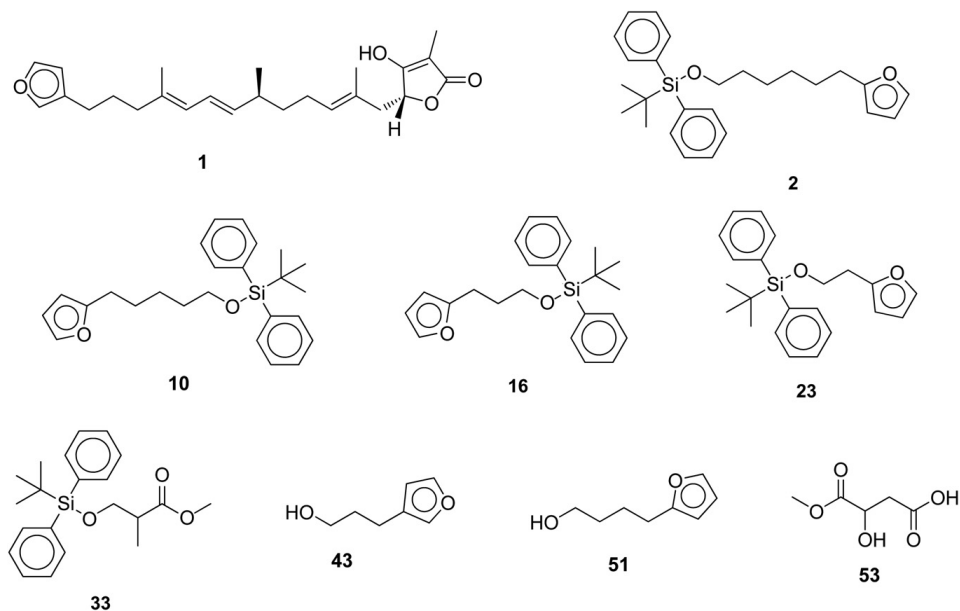


Fig. 1. Chemical structures of compounds discussed in the text (the SMILES codes for the whole data set are given in the Supplementary material).

Table 1
Experimental activity (values taken from [22,24,31]).

Compounds	Activity ^a	Binary activity ^b
1	7.19	1
2	5.08	1
3	7.89 ^c	1
4	8.44 ^c	1
5	9.00	1
6	11.00	1
7	11.00	1
8	14.00	1
9	15.72 ^c	1
10	16.13	1
11	17.39	1
12	18.00	1
13	22.00	1
14	23.00	1
15	24.19 ^c	1
16	33.48	1
17	35.00	1
18	42.00	1
19	43.00	1
20	44.00	1
21	45.20	1
22	47.00	1
23	47.64	1
24	48.00	1
25	48.62	1
26	50.00	-1
27	52.00	-1
28	55.60	-1
29	58.00	-1
30	62.98 ^c	-1
31	65.30	-1
32	67.00	-1
33_S	67.00	-1
33_R	77.00	-1
34	67.40	-1
35	67.65 ^c	-1
36	68.00	-1
37	68.00	-1
38	70.00	-1
39	72.00	-1
40	72.20	-1
41	74.90	-1
42	76.00	-1
43	80.00	-1
44	82.81 ^c	-1
45	84.80	-1
46	84.90	-1
47	87.20 ^c	-1
48	88.00	-1
49	90.00	-1
50	91.00	-1
51	92.05	-1
52	92.44	-1
53	94.00	-1
54	94.90	-1
55	97.70	-1
56	98.20	-1
57	100.00	-1
58	100.00	-1
59	100.00	-1

^a Inhibitor potency of investigated compounds expressed in percentage at 50 μ M of the respective enzyme activity.

^b A binary value equal to 1 was assigned to compounds with percentage at 50 μ M of the GSK-3 β activity less than 50% and vice versa.

^c Activity referred to the racemic mixture.

butyldiphenylsilyl (TBDPS) ethers (protecting groups widely used in organic chemistry) [23].

This collection of compounds together with their GSK-3 β percentage of inhibition response (Table 1) constitutes a valuable data set to submit to a QSAR study for at least two reasons: from a QSAR point of view the quality of the activity data is good (the activity measurements were carried out in the same laboratory) and

the experiments were carried out with a sort of HTS approach [3,13,25,26], whereas from a pharmacological point of view palinurin and related compounds are a class of compounds of potential interest in the research of new strategies for the treatment of the AD. Moreover, the data set includes Si derivatives, compounds of potential interest in medicinal chemistry.

Since the percentage of inhibition is not an appropriate biological endpoint because of the nonlinear characteristic of dose–response relationships [1], we decided to adopt a simple binary classification (Table 1): compounds with percentage at 50 μ M of the respective enzyme activity less than 50% were set as active (value = 1) whereas compounds with percentage greater or equal to 50% were set inactive (value = -1). Evidently, this classification was quite drastic and should be kept in mind in the evaluation of the final models.

Standard statistical tools present in Volsurf+ and Pentacle were used to build binary QSAR models as already discussed elsewhere [14] and the quality of the models were assigned using tailored parameters (see Section 4 for details) instead of classical r^2 and q^2 as suggested in the literature [25,27].

Moreover, assumed the OECD principles of validation [2] as an essential requirement for QSAR studies, to test the quality of the final model we used the flow chart shown in Fig. 2. In particular the original data set was split as described below in a training set (TR00) and a test set (TS00) to carry out the external validation. To establish the robustness of the final models M00-V and M00-P and thus the dependence of the results from the selected test set, internal validation runs were also performed. To do that, TR00 was split in two training (TR01 and TR02) and test sets (TS01 and TS02), respectively (Fig. 2), by a simple random selection [28] to be equally representative for two sets of descriptors (VolSurf+ and GRIND). The composition of training and test sets is reported in Table 2. Partial PLS models were sought for TR01 and TR02 both with Volsurf+ (M01-V and M02-V, respectively) and GRIND (M01-P and M02-P, respectively).

At first sight some additional insights could be obtained by individuating series of compounds showing particular chemical features. All compounds containing a silyl-ether moiety show good GSK-3 β inhibitor properties. From a comparison between the structure of compound 43 and 51, inactive compounds, and 23, 16, 10 and 2, the active ones (which differ one from the other for the number of carbon atoms which separate the silyl ether from the furan ring, 2, 3, 5 and 6, respectively), it is possible to deduce that the presence of a hydrophobic moiety was mandatory for biological activity.

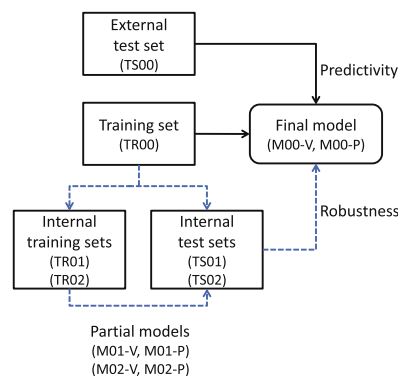


Fig. 2. Flow chart of the QSAR strategy.

Table 2
Training and test sets composition.

Model	Data set	Compounds	Binary activity
Final M00	Training set (TR00)	1, 2, 5, 6, 7, 8, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 26, 27, 28, 29, 31, 32, 33_R, 33_S, 34, 36, 37, 38, 39, 40, 41, 42, 43, 45, 46, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59	1 -1
	Test set (TS00)	3_R, 3_S, 4_R, 4_S, 9_R, 9_S, 15_R, 15_S 30_R, 30_S, 35_R, 35_S, 44_R, 44_S, 47_R, 47_S	1 -1
Partial M01	Training set (TR01)	1, 2, 6, 7, 8, 11, 12, 16, 17, 18, 19, 20, 22, 23, 24, 25 28, 29, 32, 33_R, 36, 37, 38, 39, 40, 41, 42, 43, 46, 48, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59	1 -1
	Test set (TS01)	5, 10, 13, 14, 21 26, 27, 31, 33_S, 34, 45, 49	1 -1
Partial M02	Training set (TR02)	1, 5, 6, 7, 10, 11, 12, 13, 14, 16, 17, 18, 20, 21, 23, 24 26, 27, 28, 31, 32, 33_R, 33_S, 34, 38, 39, 40, 43, 45, 48, 49, 50, 51, 52, 53, 54, 55, 57, 58, 59	1 -1
	Test set (TS02)	2, 8, 19, 22, 25 29, 36, 37, 41, 42, 46, 56	1 -1

2.2. Molecular chirality

The presence of chiral compounds should be considered in QSAR models and several strategies are reported in the literature. Some authors include for any molecule all possible stereoisomers in the data set as separate structures and assign to all of them the same activity. [14,29,30] Following this procedure the model suffers from a greater noise without loss of stability and reliability.

Here the use of a binary activity suggested to avoid the introduction of a further noise and thus we decided to exclude experimental data related to racemic mixtures from the training set (TR00) but to use them as an external data set (TS00) [28].

Forty-two compounds have a chiral center: for 32 of them the activity was measured for a single stereoisomer, for 8 compounds the racemic mixture was tested and for **33** the activity of both stereoisomers (**33_R** and **33_S**) was measured. The small difference in percentage of inhibition at 50 μ M between **33_R** and **33_S** (77% and 67%, respectively) indicates that both stereoisomers belong to the same active/inactive category. This experimental evidence suggests that for the considered data set the activity of the racemic mixtures should assign the compounds to the right active/inactive category. Final PLS models were sought for TR00 both with Volsurf+ (M00-V) and GRIND descriptors (M00-P) and their predictivity was assessed with TS00.

2.3. Volsurf+ model

The 3D structures built as explained in Section 4 were imported in Volsurf+ as mol2 format. For each molecule 96 descriptors were calculated: 28 were from water (OH2) probe, 28 from hydrophobic (DRY) probe, 5 from the hydrogen bond acceptor probe (O), 5 from the hydrogen bond donor probe (N1), 10 represent 3D

pharmacophoric descriptors related to MIFs, 2 (Log P oct/wat and Log P n-hex/wat) were obtained from internal models based on MIFs and finally 18 descriptors were independent of GRID maps.

The binary activity was then imported in Volsurf+ as dependent variable (Y) and a relation between Y and Volsurf+ descriptors (X) was looked for using standard PLS tool implemented in the software. PLS algorithm combined with Volsurf+ descriptors was already used elsewhere [9] to build models with discrete response Y, nevertheless in this study data were also submitted to a Discriminant Analysis (DA) procedure to check the reliability of the method (data not shown). Since DA confirmed PLS results with no substantial improvement, the use of the statistical tools implemented in Volsurf+ was preferred to avoid long and complex export procedures.

A model with 1 latent variable (LV) was obtained using all the 96 descriptors with an internal overall accuracy of 0.83. VIPs plot (reported in Supplementary material) was checked to reduce the number of variables, thus the final Volsurf+ model (M00-V) was obtained using only 21 descriptors (Table 3) with an improvement in precision, recall and overall accuracy (see Section 4 for parameters definitions). Ten out of the 21 selected descriptors were derived from the water probe (4 of them are related to the size of the molecules whereas 6 (W3-W8) reflect the hydrophilic regions present in the molecules), 9 from the DRY probe and the remaining two were Log P oct/wat and Log P n-hex/wat.

M00-V shows (Table 3) both internal and external accuracies equal to 0.88. Six compounds in TR00 (**20**, **25**, **33_S**, **36**, **40**, **42**) were incorrectly predicted. At a deeper insight we remarked that the percentage of inhibition spanned by the two false negatives **20** and **25** ranges from 44% to 50%, which is close to the threshold that splits active from inactive compounds. It is thus possible that the model could fail in predicting activity that differs from the threshold of

Table 3
Statistical results for Volsurf+ models. All the models was obtained using PLS with 1 latent variable and 21 Volsurf+ descriptors (see text for details).

Model	Data set	Number of compounds	Binary activity	Number of true positives		False negatives		Precision	Recall	Accuracy
				Number of true negatives	False positives					
Final M00-V	TR00	21	1	19	20, 25	0.83	0.90	0.88	0.88	0.88
		31	-1	27	33_S, 36, 40, 42	0.93	0.87			
	TS00	8	1	6	15_R, 15_S	1.00	0.75	1.00	0.75	0.88
		8	-1	8	none	0.80	1.00			
Partial M01-V	TR01	14	1	12	20, 25	0.86	0.86	0.89	0.86	0.89
		22	-1	20	36, 42	0.91	0.91			
	TS01	7	1	7	none	0.88	1.00	0.94	0.88	0.94
		9	-1	8	33_S	1.00	0.89			
Partial M02-V	TR02	16	1	15	20	0.79	0.94	0.88	0.79	0.88
		24	-1	20	26, 33_R, 33_S, 40	0.95	0.83			
	TS02	5	1	4	25	0.57	0.80	0.67	0.57	0.67
		7	-1	4	29, 36, 42	0.80	0.57			

about 5–10%. The reasons for which the four false positives are incorrectly predicted are not clear, apart from **42** which is the only compound bearing an aliphatic cyano (CN) moiety and thus it could be either not well considered by the model or not well parametrised in the GRID force field. Finally M00-V predicts successfully most compounds belonging to the test set apart from **15_R** and **15_S** which are false negatives.

Besides the external validation the model was submitted to an internal validation and two partial models, M01-V and M02-V, were obtained (Table 3). M01-V and M02-V confirms M00-V results (i.e. similar statistics and similar outliers, **26** and **29** show activity close the threshold values and could explain the differences with M00-V).

With the help of VIPs plot the most relevant descriptors (VIP > 1) were identified (Fig. 3a). The hydrophobic descriptors D1–D8 (hydrophobic volumes calculated at eight different energy from –0.2 to –1.6 kcal/mol) and the Hydrophobic Surface Area HSA (calculated via the sum of hydrophobic region contributions to molecular surface area) were positively related to the activity and confirm the fundamental role played by hydrophobicity in governing the inhibitory properties of the compounds as expected from the preliminary analysis. Three additional relevant descriptors (molecular volume V, molecular surface S and the molecular globularity G, a descriptor of the shape of the molecule) reflecting molecular shape and volume were also positively related to activity.

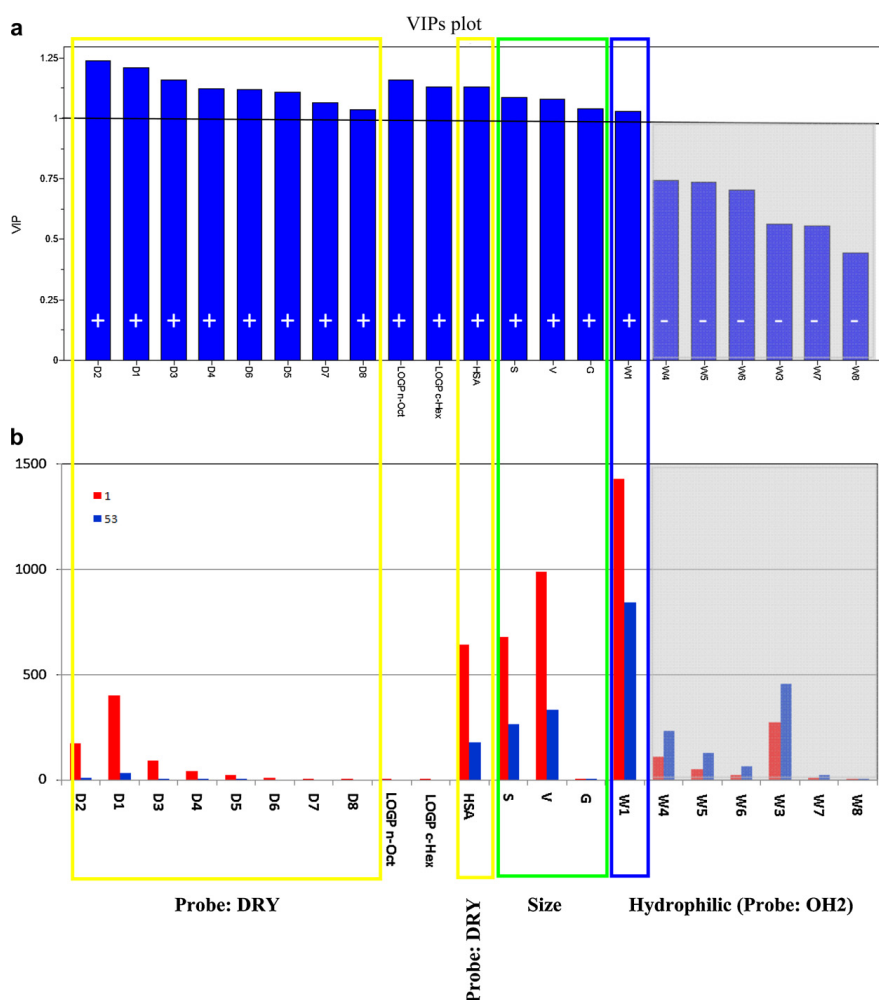


Fig. 3. Volsurf+ results. Rectangles have the following codes: yellow for the DRY probe descriptors, green for the size-related descriptors and blue for the descriptors due to the OH2 probe. The descriptors not relevant in the analysis are opaque. (a) VIPs plot for PLS the final model M00-V. The sign of the PLS coefficient (+/–) is in white. (b) Volsurf+ descriptors in raw format for **1** (active) and **53** (inactive). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Finally, the model also suggests that an increase in LogP_{oct}/wat and LogP_{hex}/wat is expected to improve the inhibitor potency of compounds. Fig. 3b compares raw values of Volsurf+ descriptors for **1**, active molecule, and **53**, inactive molecule. Descriptors increasing GSK-3 β inhibitor potency (D1–D8, W1, HSA, S, V, G) show larger values for **1** than **53** whereas the reverse is true for descriptors detrimental for activity (W3–W8). Summing up, Fig. 3b reflects the presence of hydrophobic moieties in **1** and the polar character of **53**.

2.4. GRIND-based model

The 3D structures were imported in Pentacle as mol2 format and GRIND (see Section 4 for details) were calculated.

The binary activity was then imported in Pentacle as dependent variable (Y) and a relation between Y and GRIND descriptors (X) was looked for as described above for Volsurf+ model. Results are summarized in Table 4. A final model, M00-P (Fig. 2), with three LVs was obtained with an internal accuracy of 0.94 and an external accuracy of 0.88. Three compounds in TR00 were incorrectly predicted (**12**, **20** and **42**), **12** and **20** are false negatives whereas **42** is a false positive. The test set TS00 was rightly predicted apart from **15_R** and **15_S**. Interestingly, **12**, **42**, **15_R** and **15_S** were already incorrectly predicted in M00-V, in particular the bad prediction of **42** confirms the presence of problems associated with an aliphatic CN moiety. The internal validation procedure confirmed results obtained for M00-P which is a statistical model of good quality and sufficiently robust.

The chemical interpretation of the model was assessed by selecting the ten most relevant descriptors (Table 5): six are related to DRY and TIP probes, three to N1 probe and one to TIP-N1 probes. Fig. 4 shows the main GRIND for palinurin (4a-d) and **53** (4e), active and inactive as GSK-3 β inhibitor, respectively. The favourable distances (i.e. distances that increase inhibition potency, in red) between DRY nodes (yellow, Fig. 4a) suggest that the presence of hydrophobic moieties separated from about 8 to 12 Å is mandatory for activity, whereas favourable distances between TIPs nodes (green, Fig. 4c) indicate that potent non-ATP competitive GSK-3 β inhibitors have some steric constrictions. Finally, GRIND due to the N1 probe (light blue nodes) indicate that the presence of a hydrogen bond acceptor group is important (Fig. 4d) but it should be properly localised (three N1-N1 GRIND have negative coefficients, Fig. 4e in blue).

2.5. Models comparison

Results reported above demonstrate that GRIND-based final model shows slightly better statistics than M00-V, being the same the external accuracy in the prediction. Moreover compounds incorrectly predicted were largely shared by the two models and in most

Table 5
The most relevant GRIND descriptors.

Descriptor	Fields	Distance range (Å)		Coefficient sign
11	DRY-DRY	4.4	4.8	+
18	DRY-DRY	7.2	7.6	+
30	DRY-DRY	12	12.4	+
35	DRY-DRY	14	14.4	+
404	DRY-TIP	15.2	15.6	+
222	TIP-TIP	15.6	16	+
576	N1-TIP	10.8	11.2	+
141	N1-N1	7.6	8	-
136	N1-N1	5.6	6	-
146	N1-N1	9.6	10	-

situations false negatives have percentage of inhibition near the threshold separating active from inactive molecules.

Chemical interpretation was roughly superposable: Figs. 3 and 4 show the final results obtained using Volsurf+ and Pentacle for an active molecule, Palinurin **1**, and an inactive molecule, **53**. Summing up the two models highlight the same basic information: the size of the molecule and the presence of more hydrophobic regions play a central role to increase GSK-3 β inhibitor potency whereas the presence of hydrophilic regions is less important and potentially detrimental for activity.

As expected GRIND-based model gave more detailed information about the potential interaction regions with the enzyme active site than Volsurf+. Fig. 5 compares graphical results given by the two models for palinurin. Fig. 5a shows the filtered nodes due to the DRY probe in yellow. Among them, eight nodes are more important than others because they define the most relevant DRY-DRY GRIND (red distances). Fig. 5b shows the MIF due to the DRY probe at the energy level of -0.8 kcal/mol, the corresponding volume is D4, one of the most relevant descriptors found for M00-V model. DRY-DRY GRIND are clearly more localised than D4 but evidently this feature is not fundamental to obtain significantly better statistical results. This finding could be ascribed to the binary format of the biological activity.

3. Conclusions

This study compares the quality of the QSAR models obtained using two well-known series of molecular descriptors (Volsurf+ and GRIND) to describe the non-ATP competitive GSK-3 β inhibitor activity expressed in a binary format of a series of palinurin-related compounds of potential application in AD.

Taken together results suggest that the binary quality of the biological activity does not require the use of a sophisticated 3D-QSAR approach as GRIND-based is, being the use of the simplest Volsurf+ method sufficient to extract the most relevant information.

Table 4

Statistical results for Pentacle models. All the models was obtained using PLS with 3 latent variables and 609 AMANDA/CLACC descriptors (see text for details).

Model	Data set	Exp Bin activity	Binary activity	Number of true positives	False negatives	Precision	Recall	Accuracy
				Number of true negatives	False positives			
Final M00-P	TR00	21	1	19	12 , 20	0.95	0.90	0.94
		31	-1	30	42	0.94	0.97	
	TS00	8	1	6	15_R , 15_S	1.00	0.75	0.88
		8	-1	8	None	0.80	1.00	
Partial M01-P	TR01	14	1	14	None	0.93	1.00	0.97
		22	-1	21	42	1.00	0.95	
	TS01	7	1	7	None	1.00	1.00	1.00
		9	-1	9	None	1.00	1.00	
Partial M02-P	TR02	16	1	15	20	1.00	0.94	0.98
		24	-1	24	None	0.96	1.00	
	TS02	5	1	4	25	0.67	0.80	0.75
		7	-1	5	36 , 42	0.83	0.71	

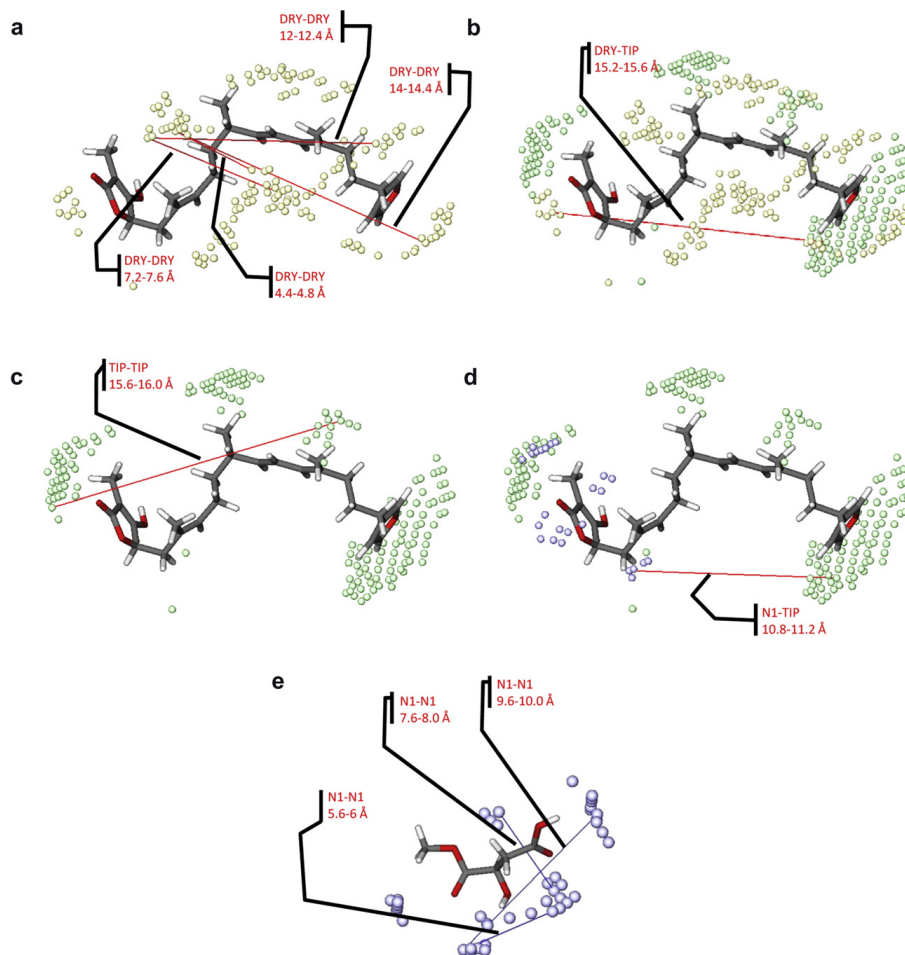


Fig. 4. Relevant GRIND descriptors for **1**(a–d) and **53**(d). Nodes selected using ALMOND filtering for DRY, TIP and N1 probes, yellow, green and blue respectively. Distances in red are favourable whereas distances in blue are not favourable to inhibitor potency. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

The level of information encoded in these pharmacodynamic data could in some way be assimilated to some pharmacokinetic measured values (e.g. distribution volume, clearance, etc.) which could be seen as “intrinsically rough” since containing the contribution of a combination of events largely dominated by hydrophobic/hydrophilic properties. This hypothesis is also supported by the successful application of VolSurf descriptors in ADME-Tox topics.

Nowadays a number of biological data of HTS quality are reported in patents describing synthetic procedures of a number of proprietary compounds. A more and more major concern in modern medicinal chemistry is thus the recovery of these data and their usage in building QSAR models. In particular of great relevance is the selection of the best strategy to perform a useful QSAR study where useful means that the final model should give reliable

information in line with the quality of the available biological data, i.e. no mis- and over-interpretation should be given. This study demonstrates that to achieve this aim the *ad hoc* molecular descriptors combined with the principle of QSAR models validation should be used.

4. Material and methods

4.1. Activity

Inhibitor potency of investigated compounds (expressed in percentage at 50 μM of the respective enzyme activity) are listed in Table 1 [22,24,31] and their binary value as well (compounds with percentage less than 50% were set as active, i.e. value = 1 whereas

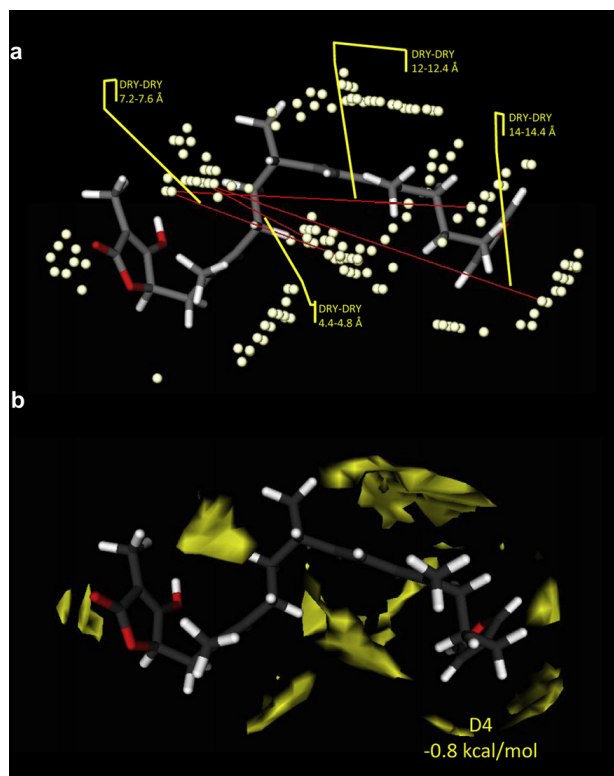


Fig. 5. Comparison of graphical results, palinurin (1) taken as an example: (a) the most relevant GRIND obtained with Pentacle; (b) D4 descriptor obtained with VolSurf+.

compounds with percentage greater or equal to 50% are set inactive, i.e. value = -1).

4.2. Data set preparation

The 2D structures were drawn in ChemDraw (version 8.0, CambridgeSoft Corporation, Cambridge, UK, 2007) using chiral notation. The ChemDraw conversion tool was then used to generate SMILES codes (the list of SMILES codes of investigated compounds is reported in the Supplementary material) that were then submitted to Omega (version 2.3.2, OpenEye Scientific Software, Inc., Santa Fe, USA, 2008). Briefly, Omega uses fragment templates along bonds to assemble initial models, after which it begins a torsion search with an assessment of freely rotatable bonds. Finally, Omega furnishes a list of conformations ranked by their MMFF energy. For any compound the conformer with the lowest energy was selected.

4.3. GRID-based descriptors calculation

Descriptors used in the paper are based on GRID (version 22c, Molecular Discovery Ltd, Pinner, Middlesex, UK, 2009, <http://www.moldiscovery.com>) force field [4–7].

4.3.1. VolSurf+

VolSurf+ (version 1.0.4, Molecular Discovery Ltd, Pinner, Middlesex, UK, 2009, <http://www.moldiscovery.com>) firstly calculates the GRID molecular interaction fields (MIFs) between the molecules in the data set and four different probes (OH2, DRY N1 and O probes that mimic respectively water, hydrophobic, hydrogen bond acceptor and hydrogen bond donor interaction of the compounds with the environment) using default settings and secondly converts MIFs into 128 molecular descriptors using an automatic procedure which has been presented in detail elsewhere [4,9,13,14]. Ninety-four out of 128 descriptors are obtained mainly by elaboration of MIFs and describe molecular properties whereas 34 ADME descriptors and the partition coefficients of the neutral species in the alkane/water and in octanol/water systems were predicted using models implemented in VolSurf+.

In this study we decided to exclude all ADME descriptors from the final model except for the partition coefficients of the neutral species in the alkane/water and in octanol/water systems and thus only 96 descriptors were used to generate the QSAR models.

4.3.2. Pentacle

Pentacle, (version 1.0.5, Molecular Discovery Ltd, Pinner, Middlesex, UK, 2010, <http://www.moldiscovery.com>), the new software that recently replaced Almond, adopts a standard procedure already

described by some of us in other papers [17,32]. Briefly, the Pentacle methodology involves three steps [15,33]: computing a set of Molecular Interaction Fields (MIFs) for molecules in the data set (as in VolSurf+), filtering the MIFs to extract the most relevant regions, and encoding the filtered MIFs into molecular descriptors named GRIND (GRid INdependent Descriptors). In this study we applied both ALMOND [33] and AMANDA [15] for filtering, whereas both traditional MACC2 and the new CLACC [34] procedure were used as encoding algorithms. As a result, filtering and encoding were combined in four methods to produce GRIND descriptors: ALMOND/MACC2, ALMOND/CLACC, AMANDA/MACC2 and AMANDA/CLACC. We built models with all the combinations (see Supplementary material) and select the model with the best statistical performance, (AMANDA/CLACC). Four probes (DRY, O, N1 and TIP) were used and default values for all Pentacle parameters were retained.

All calculations were performed on a Linux based server and on standard PCs operating with Microsoft Windows 7.

4.4. Statistical analysis

QSAR models were obtained using standard PLS methods implemented in VolSurf+ and Pentacle.

In VolSurf+ the correct interpretation of the statistical model was carried out with the help of the variable importance in the projection (VIP), VIP plot was combined with the PLS coefficient table. The VIP values reflect the importance of terms in the model both with respect to Y and with respect to X, but they do not consider the sign of the coefficients. Conversely, the PLS coefficients represents the contribution of each single descriptor to the model only with respect to Y. Coefficients with positive values increase Y and the reverse is true for coefficients with negative values. In practice, the VIP plot shows which are the most important descriptors, while the PLS coefficients table indicates which of them increases or decreases Y.

In Pentacle a similar procedure is automatically available and permits to find the most relevant variables of the model.

The Discriminant Analysis was performed with SIMCA-P (Version 11.0.0, Umetrics, Umea, SE, 2005, <http://www.umetrics.com/>).

Performance of the binary QSAR models was measured using the main standard parameters for classification models as described by Jacobsson et al. [27], in particular:

- Accuracy, is the overall classification accuracy of a prediction model, including both active and inactive compounds. It is defined by:

$$\text{Accuracy} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{fp} + \text{tn} + \text{fn}} \quad (2)$$

where tp is the number of true positives, tn is the number of true negatives, fp is the number of false positives and fn is the number of false negatives.

- Precision is a measure of the accuracy of predicting a specific class. In (2) the precision of the active class is defined by:

$$\text{Precision} = \frac{\text{tp}}{\text{tp} + \text{fp}} \quad (3)$$

Analogous equation could be written to define the precision of the inactive class

- Recall is a measure of the ability of a prediction model to select instances of a certain class from a data set. For active class, it is defined by the formula:

$$\text{Recall} = \frac{\text{tp}}{\text{tp} + \text{fn}} \quad (4)$$

Acknowledgement

We thank Molecular Discovery for the complimentary copy of Pentacle software sent us before the official release. We are grateful to the Xunta de Galicia (INCITE08PXIB314255PR) for financial support.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.ejmech.2010.12.024](https://doi.org/10.1016/j.ejmech.2010.12.024).

References

- [1] C.D. Selassie, Burger's medicinal chemistry and drug discovery, In: Drug Discovery, vol. 1. D.J. Abraham, Hoboken, John Wiley & Sons, New York, 2003, 1–48.
- [2] Guidance document on the validation of (quantitative) structure–activity relationship ((Q)SAR) models, OECD Series on Testing and Assessment 69 (OECD Document ENV/JM/MONO). Organization for Economic Co-operation and Development, 2007, 55–65.
- [3] Z. Zhou, Y. Wang, S.H. Bryant, QSAR models for predicting cathepsin B inhibition by small molecules – continuous and binary QSAR models to classify cathepsin B inhibition activities of small molecule, *J. Mol. Graphics Model* 28 (8) (2010) 714–727.
- [4] G. Cruciani, Molecular Interaction Fields – Applications in Drug Discovery and ADME Prediction, Wiley-CH, Zurich, 2005.
- [5] P.J. Goodford, A computational procedure for determining energetically favorable binding sites on biological important macromolecules, *J. Med. Chem.* 28 (1985) 849–857.
- [6] R.C. Wade, K.J. Clark, P.J. Goodford, Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds, *J. Med. Chem.* 36 (1993) 140–147.
- [7] R.C. Wade, P.J. Goodford, Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds, *J. Med. Chem.* 36 (1993) 148–156.
- [8] G. Caron, A. Nurisso, G. Ermondi, How to extend the use of grid-based interaction energy maps from chemistry to biotopics, *Chem. Med. Chem.* 4 (1) (2009) 29–36.
- [9] G. Caron, G. Ermondi, Calculating virtual logP in the alkane/water system ($\log P_{\text{alk}}^{\text{alk}}$) and its derived parameters $\Delta \log P_{\text{alk}}^{\text{alk}}$ and $\log D_{\text{alk}}^{\text{alk}}$, *J. Med. Chem.* 48 (2005) 3269–3279.
- [10] G. Cruciani, Comparative molecular field analysis using GRID Force-field and GOLPE variable selection methods in a study of inhibitors of glycogen, *J. Med. Chem.* 37 (1994) 2589–2601.
- [11] J. Nilsson, H. Wikström, A. Smilde, S. Glase, T. Pugsley, G. Cruciani, M. Pastor, S. Clementi, GRID/GOLPE 3D quantitative structure–activity relationship study on a set of benzamides and naphthamides, with affinity for the dopamine D3 receptor subtype, *J. Med. Chem.* 40 (6) (1997) 833–840.
- [12] R.D. Cramer, D.E. Patterson, J.D. Bunce, Comparative Molecular Field Analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, *J. Med. Chem.* 110 (1988) 5959–5967.
- [13] G. Cruciani, P. Crivori, P. Carrupt, B. Testa, Molecular fields in quantitative structure–permeation relationships: the Volsurf+ approach, *J. Mol. Struct. (THEOCHEM)* 503 (2000) 17–30.
- [14] P. Crivori, G. Cruciani, P. Carrupt, B. Testa, predicting blood–brain barrier permeation from three-dimensional molecular structure, *J. Med. Chem.* 110 (2000) 2204–2216.
- [15] M. Pastor, Development and validation of AMANDA, a new algorithm for selecting highly relevant regions in molecular interaction fields, *J. Chem. Inf. Mod.* 48 (9) (2008) 1813–1823.
- [16] A. Durán, I. Zamora, M. Pastor, Suitability of GRIND-based principal properties for the description of molecular similarity and ligand-based virtual screening, *J. Chem. Inf. Mod.* 49 (9) (2009) 2129–2138.
- [17] G. Caron, G. Ermondi, Influence of conformation on GRIND-based three-dimensional quantitative structure–activity relationship (3D-QSAR), *J. Med. Chem.* 50 (2007) 5039–5042.
- [18] C.G. Fortuna, V. Barresi, G. Musumarra, Design, synthesis and biological evaluation of trans 2- (thiophen-2-yl) vinyl heteroaromatic iodides, *Bioorg. Med. Chem.* 18 (12) (2010) 4516–4523.
- [19] C. Hooper, R. Killick, S. Lovestone, The GSK3 hypothesis of Alzheimer's disease, *J. Neurochem.* 104 (6) (2008) 1433–1439.
- [20] A. Martinez, D.I. Peréz, GSK-3 inhibitors: a ray of hope for the treatment of Alzheimer's disease? *J. Alzheimer Dis.* 15 (2) (2008) 181–191.

- [21] D. Alonso, A. Martínez, Marine compounds as a new source for glycogen synthase kinase 3 inhibitors. in: A. Martínez, A. Castro, M. Medina (Eds.), *Glycogen Synthase Kinase 3 (GSK-3) and Its Inhibitors: Drug Discovery and Developments*. Wiley-Interscience, New Jersey, 2006, pp. 257–280.
- [22] D. Alonso, I. Dorronsoro, A. Martínez, G. Panizo, A. Fuertes, M.J. Pérez, E. Martín, D.I. Pérez, M. Medina (2005) WO/2005/054221.
- [23] M. Pérez, D.I. Pérez, A. Martínez, A. Castro, G. Gómez, Y. Fall, The first enantioselective synthesis of palinurin, *Chem. Comm.* (2009) 3252–3254.
- [24] A. Martínez Gil, M. Medina Padilla, Y. Fall Diop, G. Gómez Pácios, M. Pérez Vázquez, E. Martín Aparicio, A. Fuertes Huerta, A.: M. Del Monte Millán, M.L. Navarro Rico, M.J. Pérez Puerto (2008) WO/2008/080986.
- [25] K.M. Thai, G.F. Ecker, A binary QSAR model for classification of hERG potassium channel blockers, *Bioorg. Med. Chem.* 16 (2008) 4107–4119.
- [26] J.L. Fells, R. Tsukahara, J. Liu, G. Tigyí, A.L. Parrill, 2D binary QSAR modeling of LPA 3 receptor antagonism, *J. Mol. Graphics Modell.* 28 (8) (2010) 828–833.
- [27] M. Jacobsson, P. Lide, Improving structure-based virtual screening by multivariate analysis of scoring data, *J. Med. Chem.* 46 (2003) 5781–5789.
- [28] V. Consonni, D. Ballabio, R. Todeschini, Evaluation of model predictive ability by external validation techniques, *J. Chemometrics* (2010) 194–201.
- [29] C. Dezi, M. Alvarado, E. Ravin, C.F. Masaguer, Multistructure 3D-QSAR studies on a series of conformationally constrained butyrophenones docked into a new homology model of the 5-HT 2A receptor, *J. Med. Chem.* 50 (2007) 3242–3255.
- [30] G. Hervé, G. Caron, J. Duché, P. Gaillard, R. Noorsaadah, A. Tsantili-Kakoulidou, P. Carrupt, P. D'Athis, J. Tillement, B. Testa, Ligand specificity of the genetic variants of human $\alpha 1$ -acid glycoprotein: generation of a three-dimensional quantitative structure–activity relationship model for drug binding to the A variant, *Mol. Pharmacol.* 138 (1998) 129–138.
- [31] A. Martínez Gil, M. Medina Padilla, A. Castro Morera, M. Alonso Cascón, G. Gómez Pácios, M. Pérez Vázquez, M. Pérez Fernández, Y. Fall Diop, S.: Herrero Santos, E. Martín Aparicio, A. Fuertes Huerta, M. Del Monte Millán, M.L. Navarro Rico, M.J. Pérez Puerto (2008) WO/2008/080988.
- [32] G. Ermondi, G. Caron, GRIND-based 3D-QSAR to predict inhibitory activity for similar enzymes, OSC and SHC, *Eur. J. Med. Chem.* 43 (2008) 1462–1468.
- [33] M. Pastor, G. Cruciani, I. Mclay, S. Pickett, S. Clementi, GRIND-Independent Descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors, *J. Med. Chem.* 43 (2000) 3233–3243.
- [34] M. Pastor, A. Duran, *Manual of Pentacle*. Molecular Discovery Ltd., Barcelona, 2010.

Accepted Manuscript

Entropy multi-target QSAR model for prediction of antiviral drugs complex networks

Francisco J. Prado-Prado, Isela García-Pintos, Xerardo García-Mera, Humberto González-Díaz

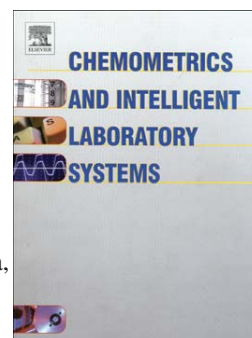
PII: S0169-7439(11)00031-1
DOI: doi: [10.1016/j.chemolab.2011.02.003](https://doi.org/10.1016/j.chemolab.2011.02.003)
Reference: CHEMOM 2331

To appear in: *Chemometrics and Intelligent Laboratory Systems*

Received date: 27 May 2009
Revised date: 9 February 2011
Accepted date: 11 February 2011

Please cite this article as: Francisco J. Prado-Prado, Isela García-Pintos, Xerardo García-Mera, Humberto González-Díaz, Entropy multi-target QSAR model for prediction of antiviral drugs complex networks, *Chemometrics and Intelligent Laboratory Systems* (2011), doi: [10.1016/j.chemolab.2011.02.003](https://doi.org/10.1016/j.chemolab.2011.02.003)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Entropy Multi-Target QSAR model for Prediction of Antiviral drugs Complex Networks

Francisco J. Prado-Prado^{1,*}, Isela García-Pintos¹,
Xerardo García-Mera¹ and Humberto González-Díaz^{2,*}.

¹ *Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, 15782, Santiago de Compostela, Spain.*

² *Department of Microbiology and Parasitology, Faculty of Pharmacy, USC, 15782, Spain*

Abstract: The antiviral QSAR models today have an important limitation. Only they predict the biological activity of drugs against only one viral species. This is determined due the fact that most of the current reported molecular descriptors encode only information about the molecular structure. As a result, predicting the probability with which a drug is active against different viral species only with a single unifying model is a goal of major importance. In this paper we use the Markov Chain theory to calculate new multi-target entropy to fit a QSAR model that predict by the first time a mt-QSAR model for 500 drugs tested in the literature against 40 viral species. We used Linear Discriminant Analysis (LDA) to classify drugs into two classes as active or non-active against the different tested viral species whose data we processed. The model correctly classifies 1424 out of 1445 non-active compounds (98.55%) and 281 out of 333 active compounds (84.38%). Overall training predictability was 95.89%. Validation of the model was carried out by means of external predicting series, the model classifying, thus, 698 out of 704 non-active compounds and 143 out of 157 active compounds. Overall validation predictability was 97.68%. The present work report the first attempts to calculate within a unify framework probabilities of antiviral drugs against different virus species based on entropy analysis. We assembled for the first time a drug-virus complex network, for observed possible mechanism of action for the different drugs against viruses.

Keywords: Antiviral drugs; QSAR; Entropy; Markov Chain model; Linear Discriminant Analysis.

***corresponding authors:** FJ Prado-Prado (fenol1@hotmail.com) and H González-Díaz (humberto.gonzalez@usc.es)

1. Introduction

Examples of diseases caused by viruses include the common cold (produced by any one of a variety of related viruses), AIDS (caused by HIV) and cold sores (caused by herpes simplex) which produced some of the major health problems in the last 30 years. Other relationships are being studied such as the connection of Human Herpesvirus 6 (HHV6), one of the eight known members of the human herpes virus family, with organic neurological diseases such as multiple sclerosis and chronic fatigue syndrome. Recently, it has been shown that cervical cancer is caused, at least partially, by papillomavirus, representing the first significant evidence in humans for a link between cancer and an infective agent. The relative ability of viruses to cause disease is described in terms of virulence.

Consequently, there is an increasing interest on the development of rational approaches for discovery of antiviral drugs. In this sense, a very important role may be played by computer-added drug discovery techniques based on Quantitative-Structure-Activity-Relationship (QSAR) models [1]. Unfortunately, almost QSAR studies, including those for antiviral activity and others, use limited databases of structurally parent compounds acting against one single fungus species [2]. One important step in the evolution of this field was the introduction of QSAR models for heterogeneous series of antimicrobial compounds; see for instance the works of Cronin, de Julián-Ortiz, Galvéz, Domenech, Gosalbez, Marrero-Ponce, Torrens, *et al.* and others [3-15]. As a result, researchers may predict very heterogeneous series of compounds but often need to use/develop as many QSAR equations as microbial species are necessary to be predicted. In any case, if you aim to predict activity against different targets you still need to use one different QSAR model for each target.

An interesting alternative, is the prediction of structurally diverse series of antimicrobial compounds (antiviral in this case) against different targets (mechanisms) using complicated non-linear Artificial Neural Networks with multi-class prediction, *e.g.* the work of Vilar *et al.* [16]. We can understand strategies developed in this sense as Multi-Objective Optimization (MOOP) techniques; in this case we pretend to optimize the activity of antiviral drugs against many different objectives or targets (viral species). A very useful strategy related to the MOOP problem use Derringer's desirability function and many QSAR models for different objectives[17]. In this sense, it is of major importance the development of unified but simple linear equations explaining the antimicrobial activity, in the present work antiviral activity, of structurally-heterogeneous series of compounds active against as many targets (viral species) as possible. We call this class of QSAR problem the multi-target QSAR (mt-QSAR) [18, 19].

There are near to 2000 chemical molecular descriptors that may be in principle generalized and used to solve the mt-QSAR problem. Many of these indices are known as Topological Indices (TIs) or simply invariants of a molecular graph G . We can rationalize G as a draw composed of vertices (atoms) weighted with physicochemical properties (mass, polarity, electronegativity or charge) and edges (chemical bonds) [20]. In any case, many of these indices have not been extended yet to encode additional information to chemical structure. One alternative to mt-QSAR is the substitution of classic atomic weights by target specific weights. For instance, we introduced and/or reviewed TIs that use atomic weights for the propensity of the atom to interact with different microbial targets [21] or undergoes partition in a biphasic systems or distribution to biological tissues [22-24]. The method, called MARCH-INSIDE approach, MARKovian CHEMicals IN SILICO Design, calculates TIs using Markov Chain theory. In fact, MARCH-INSIDE define a Markov matrix to derive matrix invariants such as stochastic spectral moments, mean values, absolute probabilities, or entropy measures, for the study of molecular properties. Applications to macromolecules have extended to RNA, proteins and blood proteome[25-30]. In particular, one of the classes of MARCH-INSIDE descriptors is defined in terms of entropy measures; which have demonstrated flexibility in many bioorganic and medicinal chemistry problems such as: estimation of anticoccidial activity, modelling the interaction between drugs and HIV-packaging-region RNA, and predicting proteins and antiviral activity [24, 31-33]. We give high importance to entropy measures due to it have been largely demonstrate as an excellent function to codify information in molecular systems, see for instance the important works of Graham [34-39]. However, have not been studied the proficiency of entropy indices (of MARCH-INSIDE type or not) to solve the mt-QSAR problems in antiviral compounds.

The present study develops the first mt-QSAR model based on entropy indices to predict antiviral activity of drugs against different viral species. The model fits one of the largest datasets used up-to-date in QSAR studies, number of entries 2 600+ cases; which is the result of forming different (antiviral compounds/viral target) pairs. One application of mt-QSAR is develop of Complex-Network (CN). In the previous work, we constructed a similar type of CNs for antiprotozoal compounds [40] and species of parasites based on an mt-QSAR. For this reason, we assembled drug-virus complex network using the model, for observed possible mechanism of action for the different drugs against viruses.

2. Methods

2.1. Markov entropy (θ_k) for drug-target k -th step-by-step interaction

One can consider a hypothetical situation in which a drug molecule is free in the space at an arbitrary initial time (t_0). It is then interesting to develop a simple stochastic model for a step-by-step interaction between the atoms of a drug molecule and a molecular receptor in the time of desencadenation of the pharmacological effect. For the sake of simplicity, we are going to consider from now on a general structure less receptor. Understanding as structure-less molecular receptor a model of receptor which chemical structure and position it is not taken into consideration. Specifically, the molecular descriptors used in the present work are called stochastic entropies θ_k , which are entropies describing th connectivity and the distribution of electrons for each

atom in the molecule [41]. The initial entropy of interaction a j-th atom of the drug with the target ${}^0\theta_j(s)$ is considered as a state function so a reversible process of interaction may be came apart on several elemental interactions between the j-th atom and the receptor. The 0 indicates that we refer to the initial interaction, and the argument (s) indicates that this energy depends on the specific viral species. Afterwards, interaction continues and we have to define the interaction probability ${}^k\theta_{ij}(s)$ between the j-th atom and the receptor for specific viral specie (s) given that i-th atom has been interacted at previous time t_k . In particular, immediately after of the first interaction ($t_0 = 0$) takes place an interaction ${}^1p_{ij}(s)$ at time $t_1 = 1$ and so on. So, one can suppose that, atoms begin its interaction whit the structure-less molecular receptor binding to this receptor in discrete intervals of time t_k . However, there several alternative ways in which such step-by-step binding process may occur[24, 42, 43]. The **Figure 1** illustrates this idea.

Figure 1 comes about here

The entropy ${}^0\theta_j(s)$ will be considered here as a function of the absolute temperature of the system and the equilibrium local constant of interaction between the j-th atom and the receptor ${}^0\gamma_j(s)$ for a give microbial species. Additionally, the energy ${}^1\theta_{ij}(s)$ can be defined by analogy as $\gamma_{ij}(s)$ [24, 42, 44]:

$${}^0\theta_j(s) = -R \cdot T \cdot \log {}^0\Gamma_j(s) \quad (1) \quad {}^1\theta_{ij}(s) = -R \cdot T \cdot \log {}^1\Gamma_{ij}(s) \quad (1)$$

The present approach to antimicrobial-species-specific-drug-receptor interaction has two main drawbacks. The first is the difficulty on the definition of the constants. In this work, we solve the first question estimating ${}^0\gamma_j(s)$ as the rate of occurrence $n_j(s)$ of the j-th atom on active molecules against a given specie with respect to the number of atoms of the j-th class in the molecules tested against the same specie $n_t(s)$. With respect to ${}^1\gamma_{ij}(s)$ we must taking into consideration that once the j-th atom have interacted the preferred candidates for the next interaction are such i-th atoms bound to j by a chemical bond. Both constants can be then written down as[24, 42, 44]:

$${}^0\Gamma_j(s) = \left(\frac{n_j(s)}{n_t(s)} + 1 \right) = e^{\frac{{}^0\theta_j(s)}{RT}} \quad (2) \quad {}^1\Gamma_{ij}(s) = \left(\alpha_{ij} \cdot \frac{n_j(s)}{n_t(s)} + 1 \right) = e^{\frac{{}^1\theta_{ij}(s)}{RT}} \quad (3)$$

Where, α_{ij} are the elements of the atom adjacency matrix, $n_j(s)$, $n_t(s)$, ${}^0\theta_j(s)$, and ${}^1\theta_{ij}(s)$ have been defined in the paragraph above, r is the universal gases constant, and t the absolute temperature. The number 1 is added to avoid scale and logarithmic function's definition problems. The second problem relates to the description of the interaction process at higher times $t_k > t_1$. Therefore, mm theory enables a simple calculation of the probabilities with which the drug-receptor interaction takes place in the time until the studied effect is achieved. In this work we are going to focus on drugs-microbial structure less target interaction. As depicted in figure 1, this model deals with the calculation of the probabilities (${}^k p_{ij}$) with which any arbitrary molecular atom j-th bind to the structure less molecular receptor given that other atom i-th has been bound before; along discrete time periods t_k ($k = 1, 2, 3, \dots$); ($k = 1$ in grey), ($k = 2$ in blue) and ($k = 3$ in red) throughout the chemical bonding system. The procedure described here considers as states of the mm the atoms of the molecule. The method arranges all the ${}^0\theta_j(s)$ values in a vector $\theta(s)$ and all the ${}^1\theta_{ij}(s)$ entropies of interaction as a squared table of $n \times n$ dimension. After normalization of both the vector and the matrix we can built up the corresponding absolute initial probability vector $\phi(s)$ and the stochastic matrix ${}^1\Pi(s)$, which has the elements ${}^0p_j(s)$ and ${}^1p_{ij}(s)$ respectively. The elements ${}^0p_j(s)$ of the above mentioned vector $\phi(s)$ constitutes the absolute probabilities with which the j-th atom interact with the molecular target or receptor in the species s at the initial time with respect to any atom in the molecule [24, 42, 44]:

$${}^0p_j(s) = \frac{{}^0\theta_j(s)}{\sum_{a=1}^m {}^0\theta_a(s)} = \frac{-RT \cdot \log \left(\frac{n_j(s)}{n_t(s)} + 1 \right)}{\sum_{a=1}^m -RT \cdot \log \left(\frac{n_a(s)}{n_t(s)} + 1 \right)} = \frac{\log \left(\frac{n_j(s)}{n_t(s)} + 1 \right)}{\sum_{a=1}^m \log \left(\frac{n_a(s)}{n_t(s)} + 1 \right)} \quad (4)$$

Where, m represents all the atoms in the molecule including the j-th, n_a is the rate of occurrence of any atom a including the j-th with value n_j . On the other hand, the matrix is called the 1-step drug-target interaction

stochastic matrix. ${}^1\Pi(s)$ is built too as a squared table of order n , where n represents the number of atoms in the molecule. The elements ${}^1p_{ij}(s)$ of the 1-step drug-target interaction stochastic matrix are the binding probabilities with which a j -th atom bind to a structure less molecular receptor given that other i -th atoms have been interacted before at time $t_1 = 1$ (considering $t_0 = 0$) [18, 24, 42, 44]:

$${}^1p_{ij}(s) = \frac{{}^1\theta_{ij}(s)}{\sum_{a=1}^n {}^1\theta_{ia}(s)} = \frac{\alpha_{ij} \cdot (-RT) \cdot \log\left(\frac{n_j(s)}{n(s)} + 1\right)}{\sum_{a=1}^n \alpha_{ia} \cdot (-RT) \cdot \log\left(\frac{n_a(s)}{n_T(s)} + 1\right)} = \frac{\alpha_{ij} \cdot \log\left(\frac{n_j(s)}{n_T} + 1\right)}{\sum_{a=1}^n \alpha_{ia} \cdot \log\left(\frac{n_j(s)}{n_T(s)} + 1\right)} \quad (5)$$

By using, $\varphi(s)$, ${}^1\Pi(s)$ and chapman-kolgomorov equations one can describe the further evolution of the system.¹⁰⁻¹⁷ summing up all the atomic free energies of interaction ${}^0\theta_j(s)$ pre-multiplied by the absolute probabilities of drug-target interaction ${}^ap_k(j,s)$ one can derive the average changes in entropies ${}^k\theta_s$ of the gradual interaction between the drug and the receptor at a specific time k in a given microbial species (s) [24]:

$${}^k\theta_s = \varphi(s) \cdot {}^k\Pi(s) \cdot {}^0\theta(s) = \varphi(s) \cdot [{}^1\Pi(s)]^k \cdot \theta^1\Pi(s) = \sum_{j=1}^n {}^k\theta_j(s) = \sum_{j=1}^n {}^ap_k(j,s) \cdot {}^0\theta_j(s) \quad (7)$$

Such a model is stochastic *per se* (probabilistic step-by-step atom-receptor interaction in time) but also considers molecular connectivity (the step-by-step atom union in space throughout the chemical bonding system).

2.2. Statistical analysis

As a continuation of the previous sections, we can attempt to develop a simple linear QSAR using the MARCH-INSIDE methodology, as defined previously, with the general formula:

$$Actv = a_0 \cdot {}^0\theta_s + a_1 \cdot {}^1\theta_s + a_2 \cdot {}^2\theta_s + a_3 \cdot {}^3\theta_s + \dots + a_k \cdot {}^k\theta_s + b_0 \quad (7)$$

Here, ${}^k\theta_s$ act as the microbial species specific molecule-target interaction descriptors. The calculation of these indices has been explained in supplementary material by space reasons. We selected Linear Discriminant Analysis (LDA) to fit the classification functions. The model deals with the classification of a set of compounds as active or not against different microbial species[44]. A dummy variable (Actv) was used to codify the antimicrobial activity. This variable indicates either the presence (Actv = 1) or absence (Actv = -1) of antimicrobial activity of the drug against the specific species. In equation (1), a_k represents the coefficients of the classification function and b_0 the independent term, determined by the least square method as implemented in the LDA module of the STATISTICA 6.0 software package[45]. Forward stepwise was fixed as the strategy for variable selection[44]. The quality of LDA models was determined by examining Wilk's U statistic, Fisher ratio (F), and the p-level (p). We also inspected the percentage of good classification and the ratios between the cases and variables in the equation and variables to be explored in order to avoid over-fitting or chance correlation. Validation of the model was corroborated by re-substitution of cases in four predicting series [44, 45].

2.3. Data set

The data set was formed by a set of marketed and/or very recently reported antiviral drugs which low reported $MIC_{50} < 10 \mu M$ against different virus. The data set was conformed to more of 500 different drugs experimentally tested against some species of a list of 40 viruses. Not all drugs were tested in the literature against all listed species so we were able to collect 2639 cases (drug/species pairs) instead of 500×40 cases. The names or codes and activity for all compounds as well as the references used to collect it are depicted in supplementary material files, see **Table 1SM**.

3. Results and discussion

3.1. mt-QSAR model.

One of the main advantages of the present stochastic approach is the possibility of deriving average thermodynamic parameters depending on the probability of the states of the MM. The generalized parameters fit on more clearly physicochemical sense with respect to our previous ones [24, 42, 43]. In specific, this work introduces by the first time a linear mt-QSAR equation model useful for prediction and MOOP [46] of the antiviral activity of drugs against different viral target species or objectives. The best model found was:

$$\begin{aligned} actv = & -1.75 \cdot \theta_2(s)_{het} + 0.99 \cdot \theta_1(s)_{total} + 0.66 \cdot \theta_4(s)_{total} - 3.19 \cdot \theta_3(s)_{C_{sat}} - 12.39 \cdot \theta_0(s)_{C_{unst}} \\ & - 2 - 23 \cdot \theta_1(s)_{Het} + 11.79 \cdot \theta_4(s)_{H-Het} - 2.86 \end{aligned} \quad (8)$$

$$N = 2639 \quad \lambda = 0.248239 \quad \chi^2 = 2463.469 \quad p < 0.001$$

In the model the coefficient λ is the Wilk's statistics, statistic for the overall discrimination, χ^2 is the Chi-square, and p the error level. In this equation, ${}^k\theta_s$ where calculated for the totality (T) of the atoms in the molecule or for specific collections of atoms. These collections are atoms with a common characteristic as for instance are: heteroatom (*Het*), unsaturated Carbon atoms (C_{unst}), saturated Carbon atoms (C_{sat}) and hydrogen bound to heteroatom (*H-Het*). The model correctly classifies 1 424 out of 1 445 non-active compounds (98.55%) and 281 out of 333 active compounds (84.38%). Overall training predictability was 95.89%. Validation of the model was carried out by means of external predicting series, the model classifying, thus, 698 out of 704 non-active compounds and 143 out of 157 active compounds. Overall validation predictability was 97.68% see **Table 1**.

Table 1 comes about here

The more interesting fact is that ${}^k\theta_s$ have the skill of discerning the active/no-active classification of compounds among a large number of viral species. This property is related to the definition of the ${}^k\theta_s$ using species-specific atomic weights (see supplementary material file for method). It allows us to model by the first time a very heterogeneous a diverse data with more than 2 600 cases (one of the largest in QSAR). Another interesting characteristic of the model is that the ${}^k\theta_s$ used as molecular descriptors depend both on the molecular structure of the drug and the viral species against which the drug must act. The codification of the molecular structure is basically due to the use of the adjacent factor α_{ij} to encode atom-atom bonding, molecular connectivity. The other aspect that allows encoding molecular structural changes is that the entropy ${}^k\theta_s$ are atom-class specific. This property is related to the definition of the ${}^k\theta_s$. The values of these species and specific atomic standard free energies reported herein for the first time are given in **Table 2** for some atoms and more than 40 species. For example, one change in the molecular structure of, e.g. S by O, necessarily implies a change in the moments of interaction. Moreover, the most interesting fact is that ${}^k\mu_s$ are the molecular descriptors reported for antimicrobial mt-QSAR studies able to distinguish among a large number of viral species. The present work is the first reported mt-QSAR model using entropy ${}^k\theta_s$ as a molecular descriptor that allow one predicting antiviral activity of any organic compound against a very large diversity of viral pathogens.

3.2 Plot of number cases vs. Total accuracy analysis.

To see the behavior of our model with respect to the number of cases, we have developed, a plot of accuracy against number of cases, using our model, see **Figure 2**. This graph displays the behavior of our model; we increase significantly the number of cases to see how it affects to the accuracy of the model. For this, we use our model and increase the cases for 5 000 to 5 000 for a total of 45 000 and observed its accuracy. We see in the **Figure 2** 15 000 cases the accuracy is 95%, but is unrealistic, because most cases are negative and the

model predicted well, but the positive value them wrong and cause this percentage. As a result, we see that the accuracy of the model is near to 95.89% (2 639 cases), but in the end the graph shows that the accuracy is 67.23%. The number of cases used in this last experiment is about 45 000 (negative) for the first class and about 500 for the second class (positive). According to the **Figure 2** we can conclude that our training dataset with 2639 cases have a correct balance with respect to the ratio of positive/negative cases. However, the data can become seriously unbalanced if we increase the number of negative cases, and it is observed by decreasing the accuracy of the model.

Figure 2 comes about here

3.3 Drug-Virus Complex networks (DV-CN) assembly.

Haggarty, Clemons, and Schreiber published a very interesting works related to the chemical profiling of different fungi strains to construct CNs [47]. The network derived was based on only 23 drugs and 10 strains of one single species *S. cerevisiae* (budding yeast). In any case, the method reported uses experimental values and is unable to add new drugs or strains to the network at least you measured it. In the previous work of this series we constructed a similar type of CNs for antiprotozoal compounds [40] and species of parasites based on an mt-QSAR. It allowed us to add growth the CN using the mt-QSAR to predict new nodes without to measure the results experimentally. In addition, we have reported very recently an mt-QSAR and CN for antifungal compounds [19]. Consequently, one possible application for this type of mt-QSAR model is the construction of multi-virus affinity profile network for antiviral drugs as well. In fact, different classes of viral drugs very possibly cover several different mechanism of action (MOA) [48]. Precisely, this type of CN may help to infer the MOA for different drugs if they link in the CN to the same/different nodes representing viruses. In this sense, we assembled drug-virus complex network using the model, for observed possible mechanism of action for the different drugs against viruses. In so doing, we used the MARCH-INSIDE 2.0 software to calculate the θ_k values. We substituted these values in the model and predicted the probability with the 40 viruses interact with dataset drugs. In order to recall the capacity of the mt-QSAR to predict new CNs we selected a database of recently assayed drugs with their respectively virus. We constructed a new observed Drug-Virus DV-CN see **Figure 3** (A), obtaining a CN with 204 vertices (virus or drugs) and 377 DP (edges, virus-drug pairs), an average distance equal to 3.67, see **Figure 3** (B).

Figure 3 comes about here

The same as before, we constructed a new predicted DV-CN obtaining an average distance equal to 3.24 with 302 DP (edges, virus-drug pairs) and 134 vertices (virus or drugs). In this, we illustrate visually both observed DV-CN and predicted DV-CN. Here we show how our DV-CN, which was assembled using our mt-QSAR model, can help us recognize a possible MOA of drugs against viruses, see **Figure 4**. Complex networks can be useful to identify possible MOAs of new drugs that have antiviral activity. An example of this, in the **Figure 4** we can see how the Lamivudine and Adefovir are active drugs against Hepatitis B, both are nucleoside analog reverse transcriptase inhibitor[49, 50]. According to the **Figure 4** we can see that the drug clevudine (L-FMAU) [51] is also active against Hepatitis B, so it may be possible to have the same mechanism of action than adefovir and lamivudine.

Figure 4 comes about here

4. Conclusions

Entropy based mt-QSAR equation is able to predict the biological activity of antiviral drugs in more general situations than the traditional QSAR models; which predict the biological activity of drugs against only one viral species. The present model with a very large data set improves significantly the previous QSAR models and may help to perform MOOP of drug activity against different viral species. This mt-QSAR methodology improves models using entropy as a molecular descriptor that allow predicting antiviral activity of any organic compound against a very large diversity of viral pathogens. One application for the present model is the construction of Complex networks; they can be useful to identify possible mechanisms of action of new drugs that have antiviral activity.

Acknowledgments

H. González-Díaz thanks sponsorships for a research position at the University of Santiago de Compostela from the Isidro Parga Pondal Program, Xunta de Galicia. F. Prado-Prado thanks sponsorships for research position at the University of Santiago de Compostela from Angeles Alvariño, Xunta de Galicia.

References

- [1] J. Prado-Prado, O. Martínez de la Vega, E. Uriarte, F.M. Ubeira, K.-C. Chou, H. González-Díaz, Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modeling and comparative multi-distance study of the giant components of antiviral drug–drug complex networks., *Bioorg. Med. Chem.*, doi:10.1016/j.bmc.2008.11.075 (2008).
- [2] F. Fratev, E. Benfenati, 3D-QSAR and molecular mechanics study for the differences in the azole activity against yeastlike and filamentous fungi and their relation to P450DM inhibition. 1. 3-substituted-4(3H)-quinazolinones, *J Chem Inf Model*, 45 (2005) 634-644.
- [3] M.T. Cronin, A.O. Aptula, J.C. Dearden, J.C. Duffy, T.I. Netzeva, H. Patel, P.H. Rowe, T.W. Schultz, A.P. Worth, K. Voutzoulidis, G. Schuurmann, Structure-based classification of antibacterial activity, *J. Chem. Inf. Comput. Sci.*, 42 (2002) 869-878.
- [4] Y. Marrero-Ponce, J.A. Castillo-Garit, E. Olazabal, H.S. Serrano, A. Morales, N. Castanedo, F. Ibarra-Velarde, A. Huesca-Guillen, A.M. Sanchez, F. Torrens, E.A. Castro, Atom, atom-type and total molecular linear indices as a promising approach for bioorganic and medicinal chemistry: theoretical and experimental assessment of a novel method for virtual screening and rational design of new lead anthelmintic, *Bioorg. Med. Chem.*, 13 (2005) 1005-1020.
- [5] Y. Marrero-Ponce, R. Medina-Marrero, F. Torrens, Y. Martínez, V. Romero-Zaldivar, E.A. Castro, Atom, atom-type, and total nonstochastic and stochastic quadratic fingerprints: a promising approach for modeling of antibacterial activity, *Bioorg. Med. Chem.*, 13 (2005) 2881-2899.
- [6] Y. Marrero-Ponce, A. Meneses-Marcel, J.A. Castillo-Garit, Y. Machado-Tugores, J.A. Escario, A.G. Barrio, D.M. Pereira, J.J. Nogal-Ruiz, V.J. Aran, A.R. Martínez-Fernandez, F. Torrens, R. Rotondo, F. Ibarra-Velarde, Y.J. Alvarado, Predicting antitrichomonal activity: a computational screening using atom-based bilinear indices and experimental proofs, *Bioorg. Med. Chem.*, 14 (2006) 6502-6524.
- [7] A. Montero-Torres, M.C. Vega, Y. Marrero-Ponce, M. Rolon, A. Gomez-Barrio, J.A. Escario, V.J. Aran, A.R. Martínez-Fernandez, A. Meneses-Marcel, A novel non-stochastic quadratic fingerprints-based approach for the 'in silico' discovery of new antitrypanosomal compounds, *Bioorg. Med. Chem.*, 13 (2005) 6264-6275.
- [8] A. Meneses-Marcel, Y. Marrero-Ponce, Y. Machado-Tugores, A. Montero-Torres, D.M. Pereira, J.A. Escario, J.J. Nogal-Ruiz, C. Ochoa, V.J. Aran, A.R. Martínez-Fernandez, R.N. Garcia Sanchez, A linear discrimination analysis based virtual screening of trichomonacidal lead-like compounds: outcomes of in silico studies supported by experimental results, *Bioorg. Med. Chem. Lett.*, 15 (2005) 3838-3843.
- [9] M.C. Vega, A. Montero-Torres, Y. Marrero-Ponce, M. Rolon, A. Gomez-Barrio, J.A. Escario, V.J. Aran, J.J. Nogal, A. Meneses-Marcel, F. Torrens, New ligand-based approach for the discovery of antitrypanosomal compounds, *Bioorg. Med. Chem. Lett.*, 16 (2006) 1898-1904.
- [10] Y. Marrero-Ponce, A. Meneses-Marcel, O.M. Rivera-Borroto, R. Garcia-Domenech, J.V. De Julian-Ortiz, A. Montero, J.A. Escario, A.G. Barrio, D.M. Pereira, J.J. Nogal, R. Grau, F. Torrens, C. Vogel, V.J. Aran,

- Bond-based linear indices in QSAR: computational discovery of novel anti-trichomonal compounds, *J. Comput. Aided Mol. Des.*, 22 (2008) 523-540.
- [11] R. Garcia-Domenech, J. Galvez, J.V. de Julian-Ortiz, L. Pogliani, Some new trends in chemical graph theory, *Chem. Rev.*, 108 (2008) 1127-1169.
- [12] Y. Marrero-Ponce, M.T. Khan, G.M. Casanola-Martin, A. Ather, M.N. Sultankhodzhaev, R. Garcia-Domenech, F. Torrens, R. Rotondo, Bond-based 2D TOMOCOMD-CARDD approach for drug discovery: aiding decision-making in 'in silico' selection of new lead tyrosinase inhibitors, *J. Comput. Aided Mol. Des.*, 21 (2007) 167-188.
- [13] A. Garcia-Garcia, J. Galvez, J.V. de Julian-Ortiz, R. Garcia-Domenech, C. Munoz, R. Guna, R. Borrás, Search of chemical scaffolds for novel antituberculosis agents, *J Biomol Screen*, 10 (2005) 206-214.
- [14] A. Garcia-Garcia, J. Galvez, J.V. de Julian-Ortiz, R. Garcia-Domenech, C. Munoz, R. Guna, R. Borrás, New agents active against *Mycobacterium avium* complex selected by molecular topology: a virtual screening method, *J. Antimicrob. Chemother.*, 53 (2004) 65-73.
- [15] A. Meneses-Marcel, O.M. Rivera-Borroto, Y. Marrero-Ponce, A. Montero, Y. Machado Tugores, J.A. Escario, A. Gomez Barrio, D. Montero Pereira, J.J. Nogal, V.V. Kouznetsov, C. Ochoa Puentes, A.R. Bohorquez, R. Grau, F. Torrens, F. Ibarra-Velarde, V.J. Aran, New antitrichomonal drug-like chemicals selected by bond (edge)-based TOMOCOMD-CARDD descriptors, *J Biomol Screen*, 13 (2008) 785-794.
- [16] S. Vilar, L. Santana, E. Uriarte, Probabilistic neural network model for the in silico evaluation of anti-HIV activity and mechanism of action., *J. Med. Chem.*, 49 (2006) 1118-1124.
- [17] M. Cruz-Monteagudo, F. Borges, M.N. Cordeiro, J.L. Cagide Fajin, C. Morell, R.M. Ruiz, Y. Canizares-Carmenate, E.R. Dominguez, Desirability-based methods of multiobjective optimization and ranking for global QSAR studies. Filtering safe and potent drug candidates from combinatorial libraries, *J Comb Chem*, 10 (2008) 897-913.
- [18] H. González-Díaz, F.J. Prado-Prado, L. Santana, E. Uriarte, Unify QSAR approach to antimicrobials. Part 1: Predicting antifungal activity against different species, *Bioorg. Med. Chem.*, 14 (2006) 5973-5980.
- [19] H. González-Díaz, F. Prado-Prado, Unified QSAR and Network-Based Computational Chemistry Approach to Antimicrobials, Part 1: Multispecies Activity Models for Antifungals, *J. Comput. Chem.*, 29 (2008) 656-657.
- [20] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, 2002.
- [21] H. Gonzalez-Diaz, F. Prado-Prado, F.M. Ubeira, Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach, *Curr Top Med Chem*, 8 (2008) 1676-1690.
- [22] H. González-Díaz, M.A. Cabrera-Pérez, G. Agüero-Chapín, M. Cruz-Monteagudo, N. Castañedo-Cancio, M.A. del Río, E. Uriarte, Multi-target QSPR assemble of a Complex Network for the distribution of chemicals to biphasic systems and biological tissues., *Chemometrics Intellig. Lab. Syst.*, 94 (2008) 160-165.
- [23] M. Cruz-Monteagudo, H. González-Díaz, G. Agüero-Chapin, L. Santana, F. Borges, R.E. Domínguez, G. Podda, E. Uriarte, Computational Chemistry Development of a Unified Free Energy Markov Model for the Distribution of 1300 Chemicals to 38 Different Environmental or Biological Systems, *J. Comput. Chem.*, 28 (2007) 1909-1922.
- [24] H. González-Díaz, G. Agüero, M.A. Cabrera, R. Molina, L. Santana, E. Uriarte, G. Delogu, N. Castanedo, Unified Markov thermodynamics based on stochastic forms to classify drugs considering molecular structure, partition system, and biological species: distribution of the antimicrobial G1 on rat tissues, *Bioorg. Med. Chem. Lett.*, 15 (2005) 551-557.
- [25] H. González-Díaz, E. Uriarte, Proteins QSAR with Markov average electrostatic potentials, *Bioorg. Med. Chem. Lett.*, 15 (2005) 5088-5094.
- [26] L. Saiz-Urra, H. González-Díaz, E. Uriarte, Proteins Markovian 3D-QSAR with spherically-truncated average electrostatic potentials, *Bioorg. Med. Chem.*, 13 (2005) 3641-3647.
- [27] G. Ferino, G. Delogu, G. Podda, E. Uriarte, H. González-Díaz, Quantitative Proteome-Disease Relationships (QPDRs) in Clinical Chemistry: Prediction of Prostate Cancer with Spectral Moments of PSA/MS Star Networks, in: B.H.a.S. Mitchem, Ch.L. (Ed.) *Clinical Chemistry Research* (ISBN: 978-1-60692-517-1), Nova Science Publisher, NY, 2009.

- [28] R. Concu, G. Podda, E. Uriarte, H. González-Díaz, A New Computational Chemistry & Complex Networks approach to Structure-Function and Similarity Relationships in Protein Enzymes, in: C.T.a.R. Collett, C.D. (Ed.) Handbook of Computational Chemistry Research, Nova Science Publishers 2009.
- [29] H. González-Díaz, Y. González-Díaz, L. Santana, F.M. Ubeira, E. Uriarte, Proteomics, networks and connectivity indices, *Proteomics*, 8 (2008) 750-778.
- [30] H. González-Díaz, S. Vilar, L. Santana, E. Uriarte, Medicinal Chemistry and Bioinformatics – Current Trends in Drugs Discovery with Networks Topological Indices, *Curr Top Med Chem*, 7 (2007) 1025-1039.
- [31] H. Gonzalez-Diaz, L. Saiz-Urra, R. Molina, L. Santana, E. Uriarte, A model for the recognition of protein kinases based on the entropy of 3D van der Waals interactions, *J Proteome Res*, 6 (2007) 904-908.
- [32] H. González-Díaz, Y. Marrero, I. Hernandez, I. Bastida, E. Tenorio, O. Nasco, E. Uriarte, N. Castanedo, M.A. Cabrera, E. Aguila, O. Marrero, A. Morales, M. Perez, 3D-MEDNEs: an alternative "in silico" technique for chemical research in toxicology. 1. prediction of chemically induced agranulocytosis, *Chem. Res. Toxicol.*, 16 (2003) 1318-1327.
- [33] H. González-Díaz, R. Molina, E. Uriarte, Markov entropy backbone electrostatic descriptors for predicting proteins biological activity, *Bioorg. Med. Chem. Lett.*, 14 (2004) 4691-4695.
- [34] D.J. Graham, Information Content in Organic Molecules: Brownian Processing at Low Levels, *Journal of chemical information and modeling*, 47 (2007) 376-389.
- [35] D.J. Graham, D. Schacht, Base Information Content in Organic Molecular Formulae, *J. Chem. Inf. Comput. Sci.*, 40 (2000) 942.
- [36] D.J. Graham, Information Content in Organic Molecules: Structure Considerations Based on Integer Statistics, *J. Chem. Inf. Comput. Sci.*, 42 (2002) 215.
- [37] D.J. Graham, C. Malarkey, M.V. Schulmerich, Information Content in Organic Molecules: Quantification and Statistical Structure via Brownian Processing, *J. Chem. Inf. Comput. Sci.*, 44 (2004).
- [38] D.J. Graham, M.V. Schulmerich, Information Content in Organic Molecules: Reaction Pathway Analysis via Brownian Processing, *J Chem Inf Comput Sci*, 44 (2004).
- [39] D.J. Graham, Information Content and Organic Molecules: Aggregation States and Solvent Effects, *Journal of chemical information and modeling*, 45 (2005).
- [40] F.J. Prado-Prado, H. González-Díaz, O. Martínez de la Vega, F.M. Ubeira, K.C. Chou, 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.*, 16 (2008) 5871-5880.
- [41] H. Gonzalez-Diaz, E. Tenorio, N. Castanedo, L. Santana, E. Uriarte, 3D QSAR Markov model for drug-induced eosinophilia--theoretical prediction and preliminary experimental assay of the antimicrobial drug G1, *Bioorg. Med. Chem.*, 13 (2005) 1523-1530.
- [42] H. González-Díaz, M. Cruz-Monteagudo, R. Molina, E. Tenorio, E. Uriarte, Predicting multiple drugs side effects with a general drug-target interaction thermodynamic Markov model, *Bioorg. Med. Chem.*, 13 (2005) 1119-1129.
- [43] M. Cruz-Monteagudo, H. González-Díaz, Unified drug-target interaction thermodynamic Markov model using stochastic entropies to predict multiple drugs side effects, *Eur. J. Med. Chem.*, 40 (2005) 1030-1041.
- [44] H. Van Waterbeemd, Discriminant Analysis for Activity Prediction, in: H. Van Waterbeemd (Ed.) *Chemometric methods in molecular design*, Wiley-VCH, New York, 1995, pp. 265-282.
- [45] StatSoft.Inc., STATISTICA (data analysis software system), version 6.0, www.statsoft.com. Statsoft, Inc., in, 2002, pp. STATISTICA (data analysis software system), version 6.0, www.statsoft.com. Statsoft.
- [46] M. Cruz-Monteagudo, F. Borges, M.N. Cordeiro, Desirability-based multiobjective optimization for global QSAR studies: application to the design of novel NSAIDs with improved analgesic, antiinflammatory, and ulcerogenic profiles, *J Comput Chem*, 29 (2008) 2445-2459.
- [47] S.J. Haggarty, P.A. Clemons, S.L. Schreiber, Chemical genomic profiling of biological networks using graph theory and combinations of small molecule perturbations, *J Am Chem Soc*, 125 (2003) 10543-10545.
- [48] J. Kerkvliet, L. Papke, M. Rodriguez, Antiviral effects of a transgenic RNA-dependent RNA polymerase, *J Virol*, 85 (2011) 621-625.

- [49] J. Balzarini, L. Naesens, E. De Clercq, New antivirals - mechanism of action and resistance development, *Curr Opin Microbiol*, 1 (1998) 535-546.
- [50] J. Neyts, E. De Clercq, Mechanism of action of acyclic nucleoside phosphonates against herpes virus replication, *Biochem Pharmacol*, 47 (1994) 39-41.
- [51] B.E. Korba, P.A. Furman, M.J. Otto, Clevudine: a potent inhibitor of hepatitis B virus in vitro and in vivo, *Expert Rev Anti Infect Ther*, 4 (2006) 549-561.

LEGENDS FOR FIGURES AND TABLES TO BE INSERTED IN THE TEXT

Figure 1. Alternative routes to step-by-step drug-target Markov interaction.

Figure 2. Plot of number cases VS Total accuracy analysis.

Figure 3. Observed Complex Network and Predicted Complex Network.

Figure 4. Drug-Virus Complex Network

Table 1. Results of the model, analysis, validation.

Table 2. Standard atomic free energy values for atom-receptor interactions.

Table 1. Results of the model, analysis, validation.

Parameter	%	Classes	Non-active	Antiviral
Analysis				
Sensitivity	98.55	Non-active	1424	21
Specificity	84.38	Antiviral	52	281
Accuracy	95.89			
Validation				
Sensitivity	99.15	Non-active	698	6
Specificity	91.08	Antiviral	14	143
Accuracy	97.68			

The positive cases are in black

Table 2. Standard atomic free energy values for atom-receptor interactions.

VIRUS	C	H	N	O	S	Cl	Br	I	F	P
ADENOVIRUS-2	0,30	0,30	0,27	0,28	0,00	0,00	0,00	0,00	0,00	0,60
BIOSTAR COLD VIRUS	0,48	0,48	0,52	0,42	0,00	0,48	0,00	0,00	0,00	0,00
CAMELPOX VIRUS	0,24	0,27	0,30	0,21	0,00	0,00	0,00	0,00	0,00	0,43
COWPOX VIRUS	0,25	0,26	0,19	0,20	0,00	0,00	0,00	0,00	0,12	0,25
COXSACKIEVIRUS	0,32	0,32	0,34	0,29	0,00	0,37	0,00	0,00	0,00	0,00
ECHOVIRUS	0,27	0,27	0,28	0,24	0,00	0,12	0,00	0,00	0,00	0,00
ENTEROVIRUS	0,47	0,47	0,56	0,37	0,00	0,00	0,00	0,00	0,00	0,00
EPSTEIN-BARR VIRUS	0,41	0,36	0,34	0,30	0,00	0,00	0,00	0,00	0,30	0,00
HEPATITIS B VIRUS	0,49	0,49	0,47	0,55	0,65	0,10	0,60	0,00	1,07	0,75
HEPATITIS C VIRUS	0,29	0,29	0,39	0,29	0,27	0,12	0,12	0,00	0,18	0,00
HERPES SIMPLEX VIRUS	0,09	0,15	0,07	0,12	0,00	0,00	0,00	0,00	0,02	0,35
HERPES SIMPLEX VIRUS 1	0,58	0,65	0,71	0,81	0,17	0,00	0,22	0,00	0,00	0,00
HERPES SIMPLEX VIRUS 2	0,35	0,44	0,43	0,60	0,12	0,00	0,30	0,00	0,00	0,00
HIV-1	0,30	0,31	0,29	0,28	0,29	0,27	0,37	0,00	0,41	0,31
HIV-2	0,23	0,26	0,26	0,29	0,60	0,00	0,00	0,00	0,00	0,30
HUMAN CYTOMEGALOVIRUS	0,29	0,30	0,32	0,34	0,10	0,35	0,40	0,00	0,14	0,32
HUMAN HERPESVIRUS 6A	0,21	0,23	0,31	0,21	0,00	0,00	0,00	0,00	0,00	0,00
HUMAN HERPESVIRUS-6B	0,32	0,35	0,44	0,33	0,00	0,00	0,00	0,00	0,00	0,00
HUMAN PAPILLOMAVIRUS	0,51	0,51	0,49	0,46	0,37	0,00	0,00	0,00	0,48	0,00
HUMAN PARAINFLUENZA VIRUS 1	0,59	0,58	0,78	0,55	0,30	0,30	0,00	0,00	0,00	0,00
HUMAN PARAINFLUENZA VIRUS 2	0,59	0,58	0,78	0,55	0,30	0,30	0,00	0,00	0,00	0,00
HUMAN PARAINFLUENZA VIRUS 3	0,59	0,58	0,78	0,55	0,30	0,30	0,00	0,00	0,00	0,00
HUMAN PICORNAVIRUS	1,32	0,45	1,10	0,08	0,08	0,00	0,00	0,00	0,24	0,00
HUMAN RHINOVIRUS	0,62	0,63	0,66	0,57	0,00	0,73	0,18	0,00	0,60	0,00

HUMANHERPES VIRUS 8	0,25	0,25	0,22	0,30	0,00	0,00	0,00	0,00	0,00	0,00
INFLUENZA B	0,26	0,24	0,33	0,24	0,00	0,00	0,00	0,00	0,00	0,00
INFLUENZA C	0,15	0,10	0,18	0,30	0,00	0,00	0,00	0,00	0,00	0,00
INFLUENZA H1N1	0,26	0,26	0,27	0,29	0,00	0,00	0,00	0,00	0,26	0,00
INFLUENZA H2N2	0,42	0,42	0,36	0,23	0,00	0,00	0,00	0,00	0,00	0,00
INFLUENZA H3N2.	0,34	0,36	0,34	0,25	0,00	0,00	0,00	0,00	0,12	0,00
INFLUENZA H5N1	0,46	0,50	0,48	0,29	0,00	0,00	0,00	0,00	0,00	0,00
INFLUENZA H6N1	0,66	0,69	0,54	0,50	0,00	0,00	0,00	0,00	0,00	0,00
INFLUENZA H9N2	0,17	0,16	0,24	0,28	0,00	0,00	0,00	0,00	0,00	0,00
INFLUENZA VIRUS	0,60	0,57	0,54	0,57	0,00	1,00	0,00	0,00	0,30	0,00
MOLONEY SARCOMA VIRUS	0,67	0,91	1,27	0,56	0,00	0,12	0,00	0,00	0,00	0,16
MONKEYPOX VIRUS	0,22	0,24	0,29	0,18	0,00	0,00	0,00	0,00	0,00	0,38
POLIOVIRUS	0,30	0,32	0,30	0,28	0,00	0,18	0,00	0,00	0,00	0,00
RESPIRATORY SYNCYTAL VIRUS	0,44	0,44	0,44	0,44	0,00	0,00	0,00	0,00	0,00	0,00
SENDAI VIRUS	0,33	0,38	0,39	0,49	0,30	0,30	0,00	0,00	0,00	0,00
VACCINIA VIRUS	0,48	0,53	0,36	0,40	0,15	0,00	0,30	0,00	0,03	0,42
VARICELLA-ZOSTER VIRUS	0,40	0,48	0,45	0,59	0,18	0,24	0,48	0,00	0,18	0,00

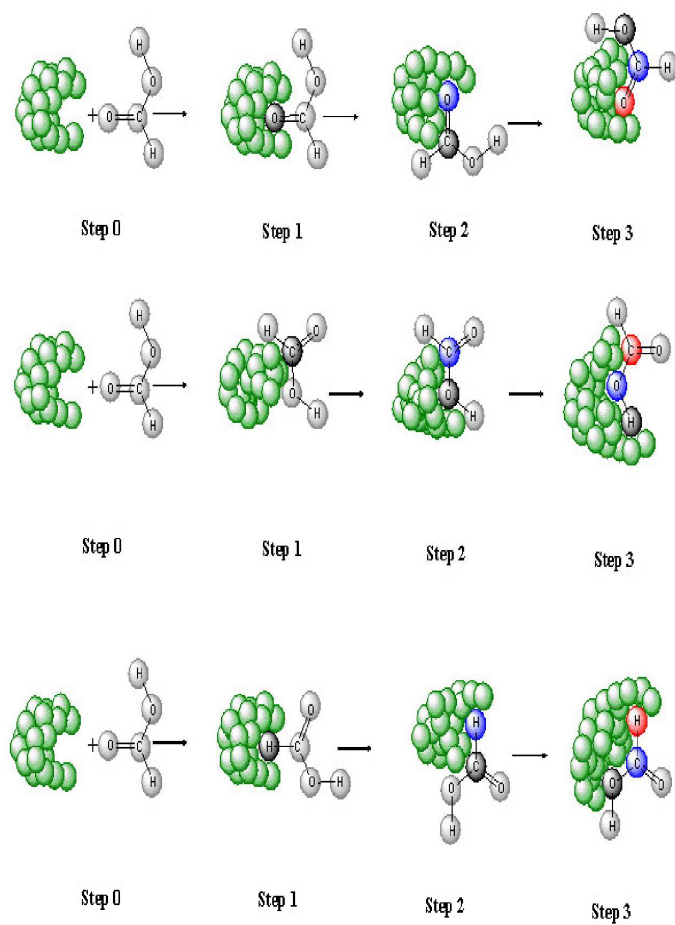
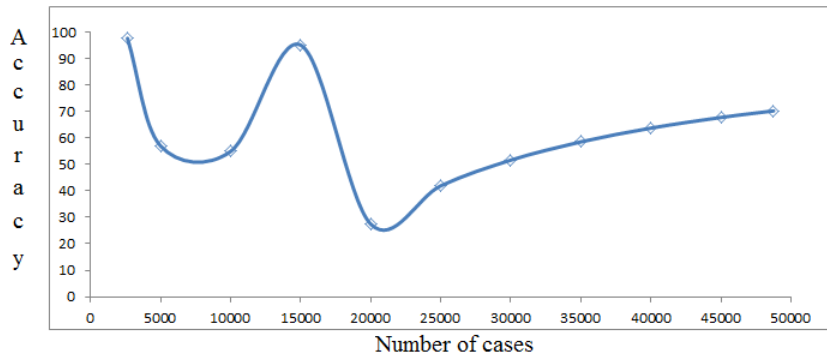


Fig. 1



ACC

Fig. 2

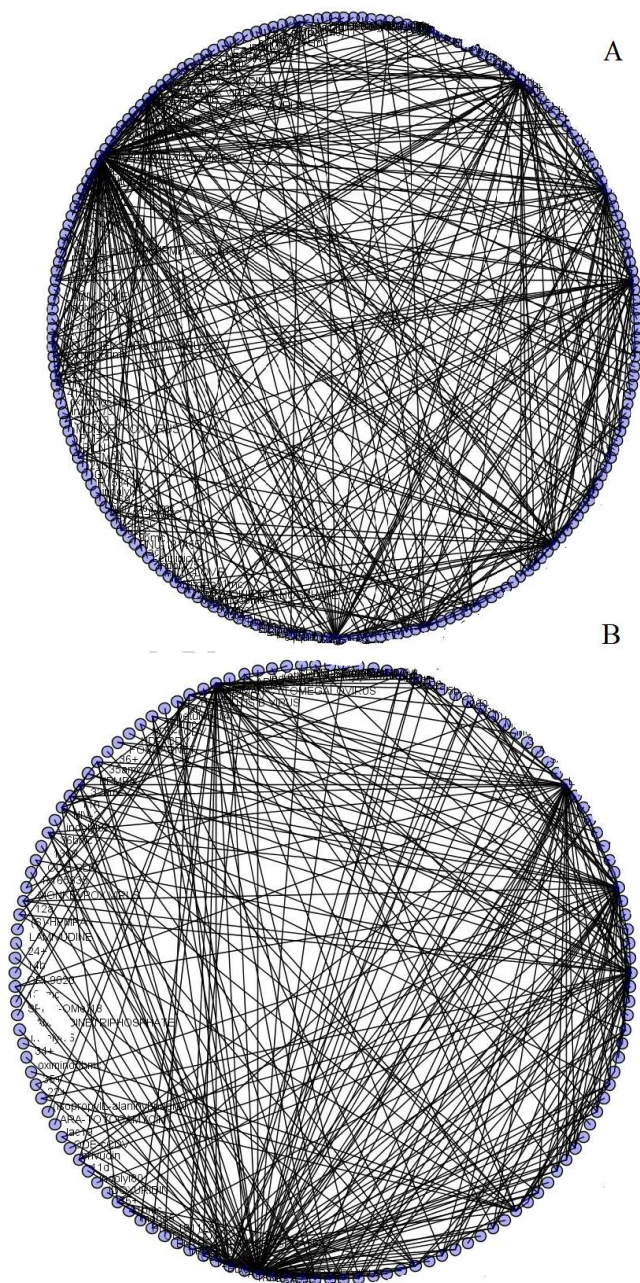


Fig. 3

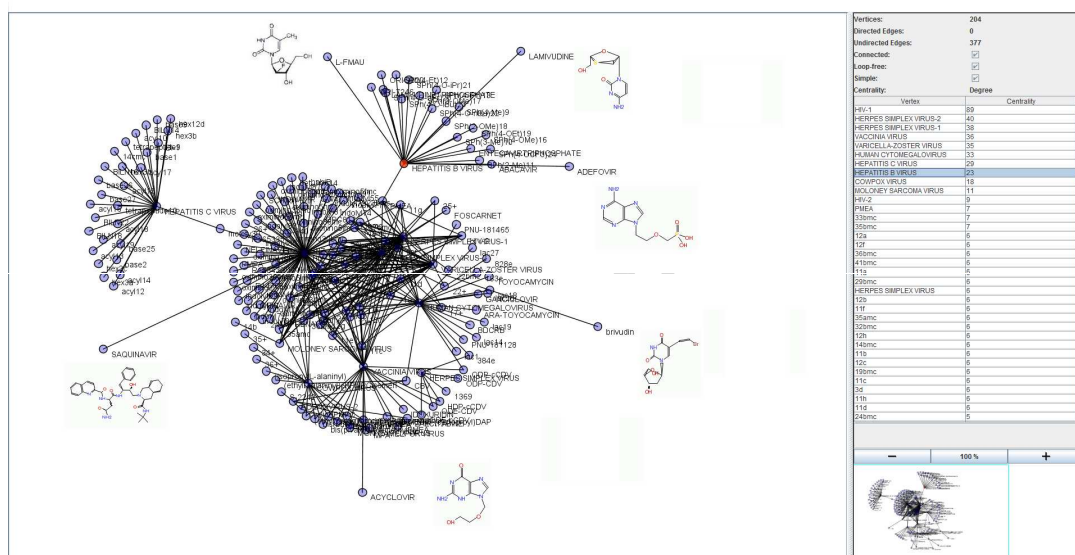


Fig. 4

CURRICULUM

C. V. científico

ISELA GARCÍA PINTOS

iselapintos@yahoo.es

isela.garcia.pintos@rai.usc.es

DATOS PERSONALES

Apellidos: García Pintos

Nombre: Isela

DNL/Pasaporte:774023L Fecha de nacimiento: 05/10/1980

Sexo: Mujer

Nacionalidad: Española

FORMACIÓN ACADÉMICA

<i>Titulación Superior</i>	<i>Centro</i>	<i>Fecha</i>
DEA	Universidad de Santiago de Compostela	Julio 2006
Tesina en Químicas	Universidade de Vigo	Julio 2005
Licenciado en Químicas	Universidade de Vigo	Diciembre 2003

<i>Doctorado</i>	<i>Centro</i>	<i>Director/a tesis</i>	<i>Fecha</i>
Doctor en Química	Universidade de Vigo	Yagamare Fall Diop Generosa Gómez Pacios	Julio 2008

IDIOMAS (R = REGULAR, B = BIEN, C = CORRECTAMENTE)

<i>Idioma</i>	<i>Habla</i>	<i>Lee</i>	<i>Escribe</i>
Gallego	C	C	C
Inglés	R	R	R

PARTICIPACIÓN EN PROYECTOS DE INVESTIGACIÓN

TÍTULO DEL PROYECTO: Síntese de moléculas con interese farmacolóxico

ENTIDAD FINANCIADORA: Universidade de Vigo (Contrato-Programa con Grupos de Investigación de Referencia)

DURACIÓN DESDE: 2006 *HASTA:* 2009

INVESTIGADOR/A PRINCIPAL: *Yagamare Fall Diop*

TÍTULO DEL PROYECTO: *Red Gallega de Investigación y Descubrimiento de Medicamentos*

ENTIDAD FINANCIADORA: *Xunta de Galicia (expediente 2007/118)*

DURACIÓN DESDE: 2007 *HASTA:* 2009

INVESTIGADOR/A PRINCIPAL: *María Isabel Loza García*

TÍTULO DEL PROYECTO: *Síntese de moléculas con interese farmacolóxico*

ENTIDAD FINANCIADORA: Xunta de Galicia (Consolidación y Estructuración de Unidades de Investigación Competitivas)

INCITE08ENA314019ES

DURACIÓN DESDE: 2008 *HASTA:* 2008

INVESTIGADOR/A PRINCIPAL: *Emilia Tojo Suárez*

TÍTULO DEL PROYECTO: *Diseño, síntese e avaliación de novos inhibidores de GSK-3beta e outros compostos con potencial actividade fronte ao cancro e a enfermidade de Alzheimer*

ENTIDAD FINANCIADORA: Xunta de Galicia

(INCITE08PXIB314255PR)

DURACIÓN DESDE: 2008 *HASTA:* 2011

INVESTIGADOR/A PRINCIPAL: *M^a Generosa Gómez Pacios*

TÍTULO DEL PROYECTO: *Síntese de moléculas con interese farmacolóxico*

ENTIDAD FINANCIADORA: Xunta de Galicia (Consolidación y Estructuración de Unidades de Investigación Competitivas)

INCITE09E1R314094ES

DURACIÓN DESDE: 2009 *HASTA:* 2009

INVESTIGADOR/A PRINCIPAL: *Emilia Tojo Suárez*

TÍTULO DEL PROYECTO: *Desarrollo de una Química verde por medio de la utilización de líquidos iónicos*

ENTIDAD FINANCIADORA: MAE (A/023040/09)
DURACIÓN DESDE: 2010 HASTA: 2010
INVESTIGADOR/A PRINCIPAL: Yagamare Fall Diop

TÍTULO DEL PROYECTO: *Red Gallega de Investigación y Descubrimiento de Medicamentos*
ENTIDAD FINANCIADORA: Xunta de Galicia (expediente 2010/43)
DURACIÓN DESDE: 2010 HASTA: 2011
INVESTIGADOR/A PRINCIPAL: María Isabel Loza García

TÍTULO DEL PROYECTO: *Síntese de moléculas con interese farmacolóxico*
ENTIDAD FINANCIADORA: Xunta de Galicia (Consolidación y Estructuración de Unidades de Investigación Competitivas) INCITE845B-2010/020
DURACIÓN DESDE: 2010 HASTA: 2010
INVESTIGADOR/A PRINCIPAL: Emilia Tojo Suárez

TÍTULO DEL PROYECTO: *Desarrollo de una Química verde por medio de la utilización de líquidos iónicos*
ENTIDAD FINANCIADORA: MAE (A/030052/10)
DURACIÓN DESDE: 2011 HASTA: 2011
INVESTIGADOR/A PRINCIPAL: Yagamare Fall Diop

TÍTULO DEL PROYECTO: *Síntese de moléculas con interese farmacolóxico*
ENTIDAD FINANCIADORA: Universidade de Vigo (Contrato-Programa con Grupos de Investigación de Referencia)
DURACIÓN DESDE: 2009 HASTA: 2011
INVESTIGADOR/A PRINCIPAL: M^a Carmen Terán Moldes

PUBLICACIONES

CLAVE: L= libro completo, CL.= capítulo de libro, A= artículo, R= revisión/"review", E= editor/a

(*) En el caso de aquellas publicaciones que estén en tramitación y aún no hayan sido publicadas, indicar únicamente la situación en la que se encuentra la publicación. (**) Esta información no deberá ser rellenada para los artículos publicados en áreas de conocimiento donde esta clasificación no sea de aplicación.

AUTORES/AS (p.o. de firma): Generosa Gómez, Hilda Rivera, Isela García, Laura Estévez, Yagamare Fall

TÍTULO: The furan approach to carbocyclic systems. Synthesis of cyclohexane derivatives from butenolides through an intramolecular Michael addition

REF. REVISTA/LIBRO: Tetrahedron Letters *CLAVE: A*
VOLUMEN: 46 (35) *PÁGINAS INICIAL Y FINAL:* 5819-5822
FECHA PUBLICACIÓN ():* 2005

AUTORES/AS (p.o. de firma): Isela García, Generosa Gómez, Marta Teijeira, Carmen Terán, Yagamare Fall

TÍTULO: The furan approach to oxacycles. Part 4: A synthesis of (+)-Decarestrictine L

REF. REVISTA/LIBRO: Tetrahedron Letters *CLAVE: A*
VOLUMEN: 47 *PÁGINAS INICIAL Y FINAL:* 1333-1335
FECHA PUBLICACIÓN ():* 2006

AUTORES/AS (p.o. de firma): Isela García, Manuel Pérez, Pedro Besada, Generosa Gómez, Yagamare Fall

TÍTULO: Synthetic studies toward Zoapatanol

REF. REVISTA/LIBRO: Tetrahedron Letters *CLAVE: A*
VOLUMEN: 49 *PÁGINAS INICIAL Y FINAL:* 1344-1347
FECHA PUBLICACIÓN ():* 2008

AUTORES/AS (p.o. de firma): Isela García, Manuel Pérez, Zoila Gándara, Generosa Gómez, Yagamare Fall

TÍTULO: The furan approach to azacyclic compounds

REF. REVISTA/LIBRO: Tetrahedron Letters *CLAVE: A*
VOLUMEN: 49 *PÁGINAS INICIAL Y FINAL:* 3609-3612

FECHA PUBLICACIÓN ()*: 2008

AUTORES/AS (p.o. de firma): Garcia, Isela; Munteanu, Cristian Robert; Fall, Yagamare; Gomez, Generosa; Uriarte, Eugenio; Gonzalez-Diaz, Humberto

TÍTULO: QSAR and complex network study of the chiral HMGR inhibitor structural diversity

REF. REVISTA/LIBRO: Bioorganic & Medicinal Chemistry

CLAVE: A

VOLUMEN: 17

PÁGINAS INICIAL Y FINAL: 165-175

FECHA PUBLICACIÓN ()*: 2009

AUTORES/AS (p.o. de firma): Isela Garcia, Yagamare Fall, Generosa Gómez

TÍTULO: QSAR & complex network study of the HMGR inhibitors structural diversity

REF. REVISTA/LIBRO: Current Drug Metabolism *CLAVE*: R

VOLUMEN: 11

PÁGINAS INICIAL Y FINAL: 307-314

FECHA PUBLICACIÓN ()*: 2010

AUTORES/AS (p.o. de firma): Isela Garcia, Yagamare Fall, Generosa Gómez

TÍTULO: QSAR, Docking, and CoMFA Studies of GSK3 Inhibitors

REF. REVISTA/LIBRO: Current Pharmaceutical Design *CLAVE*: R

VOLUMEN: 16

PÁGINAS INICIAL Y FINAL: 2666-2675

FECHA PUBLICACIÓN ()*: 2010

AUTORES/AS (p.o. de firma): Isela Garcia, Yagamare Fall, Generosa Gómez

TÍTULO: Using Topological Indices to Predict Anti-Alzheimer and Anti-Parasitic GSK-3 Inhibitors by Multi-Target QSAR *in Silico* Screening

REF. REVISTA/LIBRO: Molecules

CLAVE: A

VOLUMEN: 15

PÁGINAS INICIAL Y FINAL: 5408-5422

FECHA PUBLICACIÓN ()*: 2010

AUTORES/AS (p.o. de firma): Isela García, Yagamare Fall, Generosa Gómez, Humberto González-Díaz

TÍTULO: First computational chemistry multi-target model for anti-alzheimer, anti-parastic, anti-fingi, and anti-bacterial activity of GSK-3 inhibitors in vitro, in vivo, and in different celular lines.

REF. REVISTA/LIBRO: Molecular Diversity *CLAVE:* A
VOLUMEN: *PÁGINAS INICIAL Y FINAL:*
FECHA PUBLICACIÓN ():* 2010 (DOI 10.1007/s11030-010-9280-3)

AUTORES/AS (p.o. de firma): Giuseppe Ermondi; Giulia Caron; Isela Garcia Pintos ; Michela Gerbaldo; Manuel Pérez; Daniel I. Pérez; Zoila Gándara; Ana Martínez; Generosa Gómez; Yagamare Fall

TÍTULO: Comparing VolSurf+ and GRIND-based approach to build QSAR models with biological activity in binary format: the case of a palinurin-related data set of non-ATP competitive Glycogen Synthase Kinase 3 β (GSK-3 β) inhibitors

REF. REVISTA/LIBRO: European of Medicinal Chemistry
CLAVE: A

VOLUMEN: *PÁGINAS INICIAL Y FINAL:*
FECHA PUBLICACIÓN ():* 2011 (en prensa)

AUTORES/AS (p.o. de firma): Isela García; Yagamare Fall; Generosa Gómez, Humberto González-Díaz

TÍTULO: Entropy Multi-target QSAR model for Anti-Parasitic and Anti-Alzheimer GSK-3 inhibitors

REF. REVISTA/LIBRO: Complex Network Entropy: From Molecules to Biology, Parasitology, Technology, Social, Legal, and Neurosciences, ISBN: 978-81-7895-507-0, Editors: Humberto González-Díaz, Francisco J. Prado-Prado and Xerardo García-Mera *CLAVE:* CL

VOLUMEN: *PÁGINAS INICIAL Y FINAL:* 17-29
FECHA PUBLICACIÓN ():* 2011

AUTORES/AS (p.o. de firma): Isela García; Yagamare Fall; Generosa Gómez

TÍTULO: Trends in Bioinformatics and Chemoinformatics of Vitamin D analogues and their protein targets

REF. REVISTA/LIBRO: Current Bioinformatics *CLAVE:* R

VOLUMEN: 6 *PÁGINAS INICIAL Y FINAL:*
FECHA PUBLICACIÓN ():* 2011 (en prensa)

AUTORES/AS (p.o. de firma): Francisco Prado-Prado, Isela García

TÍTULO: Review of Theoretical Studies for Prediction
neurodegenerative inhibitors

REF. REVISTA/LIBRO: Mini-Reviews in Medicinal Chemistry

CLAVE: MR

VOLUMEN:

PÁGINAS INICIAL Y FINAL:

FECHA PUBLICACIÓN (*): 2011 (*en prensa*)

AUTORES/AS (p.o. de firma): Francisco J. Prado-Prado, Isela García,
Xerardo García-

Mera, Humberto González-Díaz

TÍTULO: Entropy multi-target QSAR model for prediction of antiviral
drugs complex networks

REF. REVISTA/LIBRO: Chemometrics and Intelligent Laboratory
Systems

CLAVE: A

VOLUMEN:

PÁGINAS INICIAL Y FINAL:

FECHA PUBLICACIÓN (*): 2011 (DOI:

10.1016/j.chemolab.2011.02.003)

**PARTICIPACIÓN EN CONTRATOS DE INVESTIGACIÓN
DE ESPECIAL
RELEVANCIA CON EMPRESAS Y/O ADMINISTRACIONES**

TÍTULO DEL CONTRATO: *Síntesis de la Palinurina y de sus
análogos*

EMPRESA/ADMINISTRACIÓN

FINANCIADORA:

NEUROPHARMA (España)

DURACIÓN DESDE: 2005 **HASTA:** 2006

INVESTIGADOR/A RESPONSABLE: Yagamare Fall Diop

ESTANCIAS EN CENTROS DE INVESTIGACIÓN

CLAVE: D=doctorado, P=postdoctoral. Y= invitado/a,
C=contratado/a, O=otras (especificar)

CENTRO: Facoltà di Farmacia, Università degli studi di Torino

LOCALIDAD: Torino *PAÍS:* Italia *AÑO:* 2009

DURACIÓN: 3 meses

TEMA: Diseño de modelos QSAR para la síntesis de los nuevos inhibidores de GSK-3beta, implicados en la enfermedad de Alzheimer, y para el estudio de la Ecotoxicidad de los Líquidos Iónicos

CLAVE: P

CONGRESOS

AUTORES/AS: Marta Teijeira, Carmen Terán, Emilia Tojo, Generosa Gómez, Concepción Saa, Isela García

TÍTULO: The Furan Approach to Oxacycles. A Synthesis of Decarestrictine L

TIPO DE PARTICIPACIÓN: Comunicación

CONGRESO: 5th Spanish Italian Symposium on Organic Chemistry

PUBLICACIÓN: Póster PO-62

LUGAR DE CELEBRACIÓN: Santiago de Compostela

AÑO: 2004

AUTORES/AS: Laura Estévez, Pilar Canoa, Generosa Gómez, Maikel Pérez, Isela Pintos, Marta Teijeira, Carmen Terán

TÍTULO: Síntesis de análogos de isonucleósidos derivados de tetrahidropirano

TIPO DE PARTICIPACIÓN: Comunicación

CONGRESO: XIV Congreso Nacional de la Sociedad Española de Química Terapéutica

PUBLICACIÓN: Póster C-87

LUGAR DE CELEBRACIÓN: Bilbao y Leioa

AÑO: 2005

AUTORES/AS: Isela García, Generosa Gómez, Yagamare Fall

TÍTULO: Síntesis del Ácido (+)-(S,S)-(cis-6-metiltetrahidropiran-2-il)acético

TIPO DE PARTICIPACIÓN: Comunicación

CONGRESO: XXXI Reunión Bienal de la Real Sociedad Española de Química

PUBLICACIÓN: Póster G1-P56

LUGAR DE CELEBRACIÓN: Toledo

AÑO: 2007

AUTORES/AS: Yagamare Fall; Isela García; Generosa Gomez; Emilia Tojo; Seila Boullosa; Gonzalo Pazos; Hilda Rivera; Alioune Fall

TÍTULO: Obtención de polioxaciclos por oxidación del furano

TIPO DE PARTICIPACIÓN: Comunicación

CONGRESO: XV Congreso Nacional de la Sociedad Española de Química Terapéutica

PUBLICACIÓN: Póster P-90

LUGAR DE CELEBRACIÓN: San Lorenzo de El Escorial

AÑO: 2007

AUTORES/AS: Generosa Gómez, Andrea Zúñiga, Emilia Tojo, Isela García

TÍTULO: Enantioselective synthesis of the civet constituent cis-(6-methyltetrahydropyran-2-yl)acetic acid

TIPO DE PARTICIPACIÓN: Comunicación

CONGRESO: XV Tenth Tetrahedron Symposium

PUBLICACIÓN: Póster B213

LUGAR DE CELEBRACIÓN: París

AÑO: 2009

OTROS MÉRITOS

-Prácticas en el laboratorio central de ensayos físico-químicos de la empresa E.N.C.E. en Agosto-Septiembre de 2004 en Pontevedra.

-Veedor en la Empresa Consello Regulador Rías Baixas en Septiembre de 2008 en Pontevedra.