

## MODELACIÓN MULTIVARIADA DE LA ACTIVIDAD ANTICOAGULANTE EMPLEANDO LA METODOLOGÍA MODESLAB

Luis Torres<sup>1\*</sup>, Laura Machin<sup>1</sup>

<sup>1</sup>Instituto de Farmacia y Alimentos. Universidad de La Habana

\*email. [luistg@ifal.uh.cu](mailto:luistg@ifal.uh.cu)

### Resumen

En el trabajo se emplea la metodología de cálculo con el MODESLAB a la modelación de la actividad anticoagulante de diferentes fármacos. Para esto se realizó el cálculo de los momentos espectrales de la matriz de adyacencia entre aristas del grafo molecular con hidrógenos suprimidos, ponderada en la diagonal principal con diferentes parámetros a 985 compuestos entre activos e inactivos. Los descriptores calculados fueron usados en una serie de entrenamiento y otra de predicción para obtener y evaluar el modelo respectivamente. Con la serie de entrenamiento se desarrolló una función discriminante para la actividad anticoagulante obteniéndose una buena clasificación total de 92.29 %. El modelo fue validado mediante el uso de una serie de predicción externa con un total de 146 compuestos, obteniéndose una buena clasificación total de un 95.89 %. Se determinó la contribución de enlaces a la actividad (análisis subestructural), De esta forma se obtuvieron las zonas de iso-contribución positiva o farmacóforas, y las zonas de iso-contribución negativa, que pueden actuar como grupos transporte para las moléculas analizadas, lo cual nos da una idea de las regiones que pueden interactuar con un receptor determinado, así como aquellas que facilitan la llegada del fármaco al sitio de acción.

**Palabras clave:** QSAR, anticoagulante, iso-contribución, predicción

### Multi-varied modelation of the anticoagulant activity using the methodology modeslab

### Abstract

In the work the calculation methodology is used with the MODESLAB to the modeling of the anticoagulant activity of different drugs. For this was carried out the calculation of the spectral moments of the adjacency womb among edges of the molecular grapho with suppressed hydrogen, pondered in the main diagonal with different parameters to 985 compounds between active and inactive. The calculated describers were used in a series of training and another of prediction to obtain and to evaluate the pattern respectively. With the series of training a function discriminate was developed for the anticoagulant activity being obtained a good total classification of 92.29%.

The pattern was validated by means of the use of a series of external prediction with a total of 146 compounds, being obtained a good total classification of 95.89%. the contribution of connections was determined to the activity (analysis sub structural), this way the areas of positive iso-contribution, and the areas of negative iso-contribution were obtained that can act as groups transport for the analyzed molecules, that which gives us an idea of the regions that they can interaction with a certain receiver, as well as those that facilitate the arrival from the drug to the action place.

**Keywords:** QSAR, anticoagulant, iso-contribution, prediction

### Introducción

El desarrollo de la Química Orgánica ha permitido hasta nuestros días la síntesis e identificación de más de diez millones de moléculas diferentes. De especial interés son aquellas que manifiestan determinadas propiedades útiles para el hombre, como es el caso de los compuestos de aplicación en la Industria Farmacéutica. Especialmente en este tipo de industria, los métodos tradicionales de obtención de nuevos medicamentos son cada vez más caros, pues el conjunto de pruebas que debe superar una nueva entidad antes de ser aplicada en el hombre incluye la definición completa de sus propiedades químico-físicas, estudios de toxicidad, elaboración de métodos para determinar el estudio de su comportamiento en el organismo humano y el estudio de propiedades farmacéuticas generales. El método de prueba y error necesita ensayar sobre 10 mil compuestos de los cuales 10 superan todos los ensayos y solo 1 puede convertirse en una medicina de prescripción. La pérdida de una molécula en las etapas finales supone una gran frustración y pérdida de recursos. Debido al alto costo de desarrollo de nuevos principios activos para la Industria Farmacéutica, en los últimos años se ha desarrollado los métodos de diseño racional de fármaco. Estos métodos se basan en el estudio de la relación entre la estructura molecular de los fármacos y las propiedades farmacológicas (1-2). Entre las múltiples aproximaciones que existen para el diseño racional de fármaco están aquellas que se basan en el estudio de descriptores moleculares topológicos, los cuales son calculados a partir de conceptos de la teoría de grafos aplicados a la Química Orgánica.

La aproximación llamada Modeslab (3-5) basada en el cálculo de los momentos espectrales de la matriz de adyacencia entre los enlaces, la emplearemos como metodología de cálculo para los descriptores que emplearemos en nuestro trabajo encaminado a la predicción de la actividad anticoagulante, así como la interpretación químico física de la contribución de enlace a la misma.(5)

## Problema Científico

La obtención de fármacos anticoagulantes es un proceso largo y costoso llevado a cabo fundamentalmente basado en el método de prueba y error.

## Hipótesis:

Obtener modelos predictivos mediante el ADL permite obtener nuevos fármacos anticoagulantes mediante un proceso mas corto y menos costoso.

## Objetivos:

- Lograr un modelo matemático mediante un Análisis Discriminante Lineal (ADL) que prediga la actividad anticoagulante.
- Validar el método obtenido.
- Determinar la contribución de enlace a la actividad.

## Materiales y Métodos

Para la búsqueda de modelos de clasificación se seleccionó una serie de entrenamiento, formada por dos componentes fundamentales: en primer lugar, un conjunto de compuestos químicos con una actividad farmacológica común a los cuales se les denomina *compuestos activos*, en segundo lugar un conjunto de compuestos con actividades farmacológicas diversas pero diferente a los activos, a este segundo grupo se ha dado el nombre de compuestos inactivos, esto debido a que no poseen la actividad farmacológica del primer grupo al cual se ha hecho referencia aquí (6-9). Debe destacarse que los compuestos pertenecientes tanto al primer conjunto como al segundo, se ha tenido en cuenta que tengan la mayor variabilidad posible desde el punto de vista estructural.(10-15). Para esta serie de entrenamiento se le calculó los momentos espectrales a cada uno de los compuestos que la componen ponderando los grafos moleculares con el momento dipolo estándar de enlace, se obtuvo una matriz de datos que contiene los momentos espectrales desde  $\mu_0$  hasta  $\mu_{15}$  para cada uno de los compuestos. Los momentos espectrales empleados en este trabajo fueron calculados, con el programa Modeslab (3) el cual genera los datos en ficheros de extensión .txt compatibles con el Microsoft Office. El procesamiento de los datos para crear nuevas variables se realizó con el tabulador electrónico *Microsoft Excel 2010*. (16). Los ficheros generados por Excel fueron procesados con el *software STATISTICA 8.0* (17) para *Windows* mediante un análisis de regresión lineal múltiple para buscar los modelos de predicción. Todos los *software* fueron corridos

sobre una computadora personal *Toshiba* modelo *Satellite M505 - S4972 Intel® Pentium®* con 3GB DDR2 RAM, CPU T4300 2.10GHz con *Windows 7* de 32-Bit de sistema operativo.

### Resultados

El primer paso para encontrar un modelo teórico que permita descubrir nuevos fármacos es diseñar una serie de entrenamiento representativa y aleatoria. En este trabajo se ha contado con una amplia data de 984 compuestos que comprenden compuestos activos e inactivos y estos últimos divididos en varios grupos farmacológicos; entre los que se encuentran antifúngicos, antihistamínicos, diuréticos, tuberculostáticos. Esta data fue dividida en dos subseries, una conteniendo 241 compuestos activos y 597 inactivos para formar la serie de entrenamiento y con un criterio similar se diseñó la serie de predicción conformada por 32 compuestos activos y 114 inactivos, a los cuales se le ha hecho un análisis de Cluster K-means.

El modelo de clasificación así como los parámetros estadísticos son mostrados a continuación

$$Act = 0.245 \mu_0 - (-0.3941) \mu(std)_3 - 0.591 \mu(std)_2 - 0.801 \mu(std)_1 \\ - 1.46 * 10 \mu(std)_5 - (-21.390)$$

Donde  $\lambda$  es la lambda de Wilks',  $D^2$  es la distancia de Mahalanobis y la F es la razón de Fisher

Para la discriminación de compuestos activos/inactivos estudiados en este trabajo el modelo clasifica correctamente el 91.29 % de los activos y el 92.81 de los inactivos en la serie de entrenamiento para una buena clasificación global del 92.37 %. El porcentaje de falsos activos y falsos inactivos en la serie de entrenamiento es 7.19 y 8.71 % respectivamente. Los falsos activos son aquellos compuestos inactivos que el modelo clasifica como activos y los falsos inactivos son los positivos que el modelo clasifica como inactivos

Uno de los criterios más importante para la aceptación o no de un modelo discriminante tal como se muestra en este trabajo está basado en los estadísticos de la serie de predicción externa. El modelo clasifica el 100 y el 94.74. % de los activos e inactivos respectivamente, para una clasificación global del 95.89 %.

### Análisis de contribución de enlace a la propiedad.

La evaluación de las contribuciones presupone la necesidad de superponer las contribuciones de todos los fragmentos en la molécula, particionándolas de forma invariante; para garantizar esto es necesario hacer dos tipos de particiones, una para los términos lineales y

otra para los de interacciones. Debido al carácter invariante necesario para la partición se escoge como mínima unidad estructural el enlace, teniendo en cuenta que este es la base de la matriz de adyacencia usada en el cálculo de los momentos espectrales. La partición de los términos lineales consiste simplemente en sumar todas las contribuciones de todos los fragmentos que pasan por un enlace premultiplicada por su peso estadístico

$$C(e) = \sum_i^m a_i \mu_i^L + \frac{a_{k,l}}{2} \sum_{k,l}^m (\mu_k^L \cdot \mu_l + \mu_l^L \cdot \mu_k) + \frac{a_0}{m}$$

Donde:  $C(e)$  es la contribución de enlace a la propiedad,  $a_i$  es el coeficiente estadístico del  $i$ -ésimo momento espectral total ( $\mu_i$ ) en el modelo de clasificación;  $a_{k,l}$  es el coeficiente estadístico de la interacción entre el  $k$ -ésimo y el  $l$ -ésimo momento espectrales totales ( $\mu_k, \mu_l$ );  $\mu_k^L$  y  $\mu_l^L$  son los momentos espectrales locales de orden  $k$  y  $l$  para el enlace ( $e$ ),  $a_0$  es el término independiente del modelo y  $m$  es el número de enlaces totales en la molécula.

### Discusión

La  $\lambda$  de Wilks' puede tomar valores entre 0 (discriminación perfecta) y 1 (no discrimina). La selección de los modelos se realizó en base a la calidad estadística de los mismos, los estadígrafos de comparación multivariada tenidos en cuenta para este fin fueron, en primer lugar la lambda ( $\lambda$ ) de Wilks. De un conjunto elevado de modelos obtenidos se seleccionaron sólo aquellos cuyo valor de  $\lambda$  es el más pequeño posible, ya que permite evaluar la hipótesis de que dos o más grupos provienen de poblaciones con medias significativamente diferentes para un conjunto de variables. Debido a que valores pequeños de ( $\lambda$ ) indican diferencias entre las medias de grupos, este se considera uno de los mejores criterios de comparación multivariada. Otro criterio estadístico importante a la hora de decidir qué modelo se debía seleccionar para realizar clasificaciones de calidad en el proceso de diseño de fármacos fue el cuadrado de la distancia de Mahalanobis ( $D^2$ ), la cual es una especie de distancia entre los centroides de cada uno de los supuestos grupos, por lo tanto su valor indicará, de una manera proporcional, la diferencia entre ambos, para grupos idealmente separados entre si la  $D^2$  como estimador insesgado de la F de Fisher debe ser mayor o igual a esta para

demostrar significación estadística en la prueba de hipótesis, ninguno de los modelos encontrados cumplió este requisito lo cual indica que no hay una separación total entre los grupos sino que existe cierta superposición en este caso  $D^2$  es menor que  $F$ , indicando que existe un solapamiento entre los grupos reales, este hecho puede ser explicado por la existencia de moléculas con un comportamiento dual pero no reportado en el grupo de los inactivos en la serie de entrenamiento. Un ejemplo puede ser los anticoagulantes. En la práctica médica son muy conocidos casos de compuestos con dicho comportamiento, como es el acetaminofeno, es por esa razón que se emplea el porcentaje de buena clasificación para evaluar cuando un modelo posee un alto poder discriminatorio entre dos grupos diferentes. Se hizo una valoración de la efectividad del proceso de clasificación. Para ello se tuvo en cuenta los resultados de la matriz de clasificación, para asegurar la alta calidad del modelo sólo se tomaron aquéllos en los cuales el porcentaje de casos bien clasificados no fuera inferior al 78%, en este sentido se tuvo en cuenta que el porcentaje de casos bien clasificados para el grupo de sustancias inactivas fuera alto, para evitar la aparición de "falsos activos" a la hora de la predicción, lo cual muestra la calidad de los modelos ya que evita la mala selección de un compuesto en el momento de realizar un diseño racional de fármacos. Además de esto, se prestó especial atención como criterio final de selección, a la capacidad predictiva de los modelos, caracterizado por el porcentaje de buena clasificación en la serie de predicción.

Para la discriminación de compuestos activos/inactivos estudiados en este trabajo el modelo clasifica correctamente el 91.29 % de los activos y el 92.81 de los inactivos en la serie de entrenamiento para una buena clasificación global del 92.37 %. El porcentaje de falsos activos y falsos inactivos en la serie de entrenamiento es 7.19 y 8.71 % respectivamente. Los falsos activos son aquellos compuestos inactivos que el modelo clasifica como activos y los falsos inactivos son los positivos que el modelo clasifica como inactivos.

Uno de los criterios más importante para la aceptación o no de un modelo discriminante tal como se muestra en este trabajo está basado en los estadísticos de la serie de predicción externa. El modelo clasifica el 100 y el 94.74. % de los activos e inactivos respectivamente, para una clasificación global del 95.89 %. El porcentaje de falsos activos e inactivos en la serie de predicción es: 5.26% y 0 %, respectivamente.

El porcentaje de falsos activos e inactivos como se describió anteriormente es el

porcentaje de compuestos mal clasificados en cada una de la respectivas series esto puede estar motivado por diversos factores; uno de los cuales puede estar asociado con la metodología de cálculo MODESLAB, la cual opera con momentos espectrales los que codifican información subestructural, y como es bien conocido de los estudios de REA en la molécula de un fármaco se puede encontrar una parte farmacófora y otra correspondiente a los grupos transportes, los cuales no son esenciales para la actividad pero al estar presentes en la molécula se tiene en cuenta su influencia sobre el cálculo del momento espectral, entonces puede darse el caso que exista una mayor contribución de los grupos transportes al momento espectral que no del grupo farmacóforo; esto evidentemente puede llevar a clasificar un fármaco inactivo como activo y viceversa.

### **Análisis de contribución de enlace a la propiedad**

Se evaluará los enlaces como positivos o negativos según su contribución a la actividad, donde la suma total de estas contribuciones será la contribución total de la molécula. Para ilustrar lo que será discutido en detalle, en los siguientes ejemplos

Entiéndase como zona isoactiva aquella agrupación atómica cuyos enlaces tiene grado mayor o igual a uno y contribución positiva a la propiedad, esto garantiza que al tener los enlaces grado mayor o igual que uno en la zona isoactiva no existan elementos aislados y al ser todos positivos no quede encubierto un efecto por otro mayor y de signo contrario, de igual manera la zona isonegativa será aquella en la cual todos sus enlaces tengan grado mayor igual a uno y de contribución negativa. La partición en zonas de isoactividad permite sumar las contribuciones de enlaces para obtener contribuciones de grupos específico en una molécula que tenga un carácter invariante. Las zonas de isocontribución (isoactivas o isonegativas) pueden ser subdividas en grupos atómicos más pequeños siguiendo diferentes criterios como pueden ser criterios sintéticos, farmacológicos electrónicamente, farmacodinámica, u otros según los fines del investigador. La subdivisión de una zona de isocontribución en grupos sintéticos se realizará utilizando las reglas de desconexión y permitirá diseñar una molécula activa siguiendo el criterio de accesibilidad sintética. las zonas que aparecen sombreadas se corresponden con las zonas activas de la molécula aún sin especificar su valor de contribución.

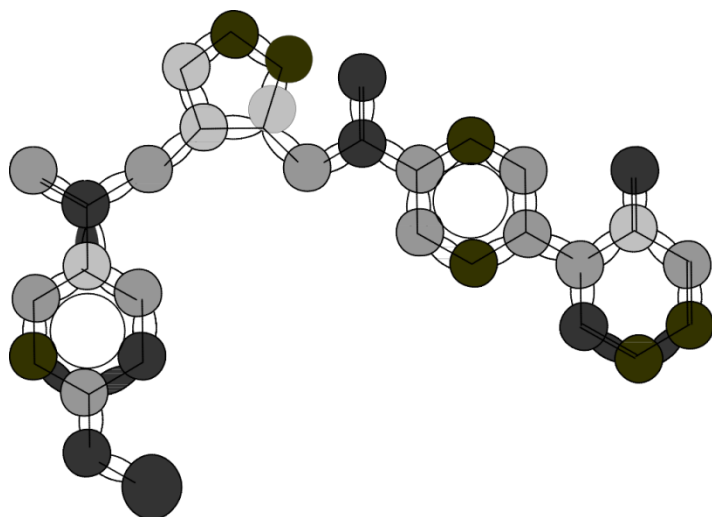


Figura 1. Estructura de la molécula 023A20

Se puede apreciar en la molécula anterior que existe un predominio de las zonas de contribución positivas en los anillos aromáticos y átomos de carbonos con deficiencia electrónica o exceso de esta, lo que nos lleva a poder plantear que la principal forma de interacción electrostática con el receptor es de tipo dipolo dipolo, esto está en plena correspondencia con lo planteado en la literatura en lo que se refiere a los estudios de REA para este tipo de compuesto donde la práctica demuestra el incremento de la actividad al introducir sustituyentes electroaceptores pero en este caso o podemos particularizar un poco más y afirmar que se trata de sustituyentes que extiendan conjugación, tal como es el caso del carbonilo. También por otra parte es importante destacar que desde el punto de vista sintético esto tiene su implicación puesto que estos grupos de mayor contribución a la actividad pueden ser introducidos o generados en pasos específicos de reacción, con el fin de aumentar el rendimiento y disminuir productos colaterales.

Analicemos ahora el caso de una molécula considerada el grupo de las inactivas pero que tiene un comportamiento peculiar

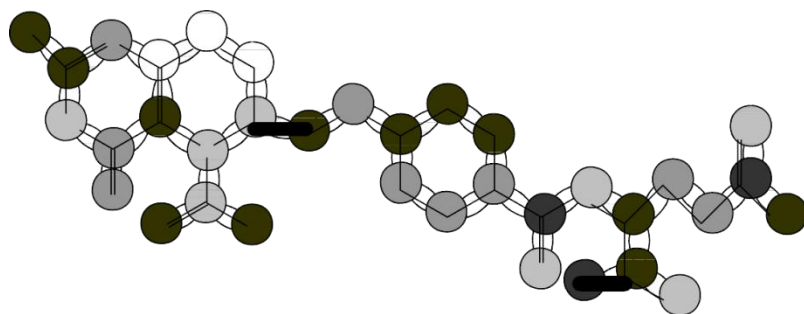


Figura 2. Estructura del ácido fólico

En este caso podemos ver como existe una mayor cantidad de enlaces que contribuyen negativamente a la actividad ( los de menos intensidad en el color) pero llama la atención que esta molécula a pesar de ser inactiva se clasifica como un falso activo por el modelo, con anterioridad en este trabajo se da una explicación a las posibles causas de aparición de los falsos activos y una de ellas esta dada por la posibilidad real de que la molécula manifieste dicha actividad y que no se haya probado hasta ese momento, ya que el criterio de inactividad que se sigue en su selección es el hecho de no tenerla reportada, pues bien esta puede ser la posible causa de su clasificación como tal por el modelo, si observamos en su estructura vemos la existencia de grupos atómicos que reúnen la características necesarias para establecer la interacción con el receptor pues coinciden con los presentes en las moléculas del grupo de los activos para nuestro análisis, esto puede ser de importancia significativa en la búsqueda de nuevos agentes anticoagulantes, ya que se tendría la ventaja de conocer las vías de obtención de estos compuestos, así como se dispondría de estudios toxicológicos que acortarán el tiempo de desarrollo de nuevas aplicaciones terapéuticas.

### Referencias

1. Lajiness, M.S., *Molecular similarity-Based Methods for Selecting Compounds for Screenig. In Computacional Chemical Graph Theory.* 1990, new york.
2. Gálves, J., *Diseño de medicamentos.* 1995, España.
3. Estrada, E., *Aplication of aproximations Toss Mode.* J. Chem. Inf. Comput, 1995. **35**.

4. E, E., *Applications of Aproximations in adyacence Matrix of edge*. J. Chem Inf. Comput, 1998.
5. E, E., *Spectral Moment of edge adyacency matrix*. J. Chem. Inf. Comput, 1998.
6. Goodman, *Bases Farmacológicas de la terapéutica*. 2000, Los Angeles.
7. Malgor, L.A., *Farmacología*. 2002: España.
8. Foye, W.O., *Principios de Química Farmacéutica*. 1984, España.
9. Españoles, C.d.F., ed. *Base de Datos del Medicamento*. 2002: España.
10. R, K., *A priori molecular descriptors in QSAR: a case of HIV-1 protease inhibitors. I. The chemometric approach*. J Mol Graph Model, 2003.
11. H, Y., *QSAR studies of HIV-1 integrase inhibition*. Bioorg Med Chem., 2002. **12**.
12. MG, F., *QSAR studies of the pyrethroid insecticides. Part 3. A putative pharmacophore derived using methodology based on molecular dynamics and hierarchical cluster analysis*. J Mol Graph Model., 2003.
13. Y, F., *Quantitative structure-antitumor activity relationships of camptothecin analogues: cluster analysis and genetic algorithm-based studies*. J Med Chem, 2004.
14. T, S., *Classification of environmental estrogens by physicochemical*
15. Masart, D., *The Interpretation of Chemical Analytical Data by the use of Cluster Analysis*. 2003: New York.
16. *Microsoft Excel for Windows XP Ver. 7.0 Copyright 1985-2010*. 2010.
17. STATISTICA 8.0 for WINDOWS (2008), Statsoft, Inc