

## 1. FUNDAMENTOS TEÓRICOS

### 1.1. ENFERMEDAD DE PARKINSON.

La Enfermedad de Parkinson idiopática (EP) es un trastorno neurodegenerativo, de etiología desconocida, en el que el principal hallazgo neuropatológico es una degeneración de la sustancia negra y la consiguiente disfunción del sistema nigroestriatal. [4]

Es sabido que la pérdida de neuronas dopaminérgicas es un aspecto normal del envejecimiento; no obstante, cuando hablamos de EP nos referimos a la pérdida del 80 a 90% de estas neuronas, causando así manifestaciones obvias y limitantes. [5]

### 1.2. QUÉ ES LA ENFERMEDAD DE PARKINSON

El mal de Parkinson es una aflicción neurodegenerativa paulatina que ocurre cuando ciertas células nerviosas (neuronas) de la región cerebral conocida como sustancia negra (Esta estructura es una lámina de sustancia gris con neuronas intensamente pigmentadas (neuromelanina), localizada en el mesencéfalo, dorsal a los pedúnculos cerebrales. En ella se describen dos regiones: La parte dorsal es la llamada zona compacta, mientras que la parte ventral es la llamada zona reticulada) mueren o degeneran. [6], en esta muerte o degeneración se produce un depósito de material proteico rico en alfa sinucleína en el citoplasma de las neuronas, formando cuerpos de inclusión conocidos como cuerpos de Lewy [7], estos cuerpos son depósitos redondos que aparecen dentro de las células nerviosas dañadas. [8]

Normalmente cuando las neuronas están en buen estado producen una sustancia llamada dopamina. La dopamina es el mensajero químico encargado de transmitir señales de la sustancia negra al siguiente “puesto de retransmisión,” el cuerpo

estriado, lo que permite un funcionamiento suave y coordinado de los músculos y el movimiento del organismo. [9]

Generalmente los neurólogos no emiten un diagnóstico de mal de Parkinson a menos que se presenten dos de estos síntomas por un período considerable. La enfermedad es tanto crónica (que dura muy largo tiempo) como progresiva; o sea que sus síntomas se van empeorando con el tiempo. [9]

### **1.3. ¿CUÁL ES LA CAUSA DE LA ENFERMEDAD PARKINSON?**

La pérdida de las neuronas productoras de dopamina en la sustancia negra hace que las células nerviosas del cuerpo estriado emitan señales fuera de orden, lo que lleva a la incapacidad de controlar el movimiento normalmente. Se ha demostrado en estudios que el cerebro pierde cerca de 80 por ciento de las células productoras de dopamina en la sustancia negra antes de que se manifiesten los rasgos distintivos del EP.

Una de las teorías de la causa del EP es que los radicales libres (moléculas inestables y potencialmente dañinas generadas por reacciones químicas normales del organismo) tal vez contribuyen a la muerte celular, y conducen al mal de Parkinson. Los radicales libres son inestables porque les falta un electrón; para llenar la vacante, los radicales libres reaccionan con las moléculas vecinas (especialmente de metales como el hierro) en un proceso llamado oxidación. Se piensa que la oxidación causa daño a los tejidos, y también a las neuronas. Normalmente el daño causado por radicales libre es contrarrestado por los antioxidantes, sustancias químicas que protegen a las células de dicho daño [10]

### **1.4. LOS CUATRO RASGOS PRINCIPALES DEL EP**

- ✓ El temblor de reposo es la manifestación menos invalidante, tiene una frecuencia de 3-5 Hz y clásicamente se describe como un movimiento de «contar monedas». Es el primer síntoma en el 50-70% de los casos, de predominio asimétrico y distal en las extremidades superiores. Este síntoma

responde mal al tratamiento con levodopa (LD) y mejora con técnicas quirúrgicas.

- ✓ La rigidez predomina en los músculos flexores y se manifiesta como una hipertonía plástica o en rueda dentada cuando se superpone el temblor. Puede ser molesta o incluso dolorosa y es extremadamente sensible al tratamiento con LD.
- ✓ La bradicinesia es el síntoma más discapacitante. Es la lentitud del movimiento que impide o dificulta los movimientos sucesivos o simultáneos. Estas alteraciones pueden desaparecer bruscamente durante las cinesias paradójicas. La acinesia puede afectar a diferentes zonas corporales.
- ✓ La inestabilidad postural es la aparición gradual y tardía en la evolución de una dificultad del equilibrio. Es un síntoma muy incapacitante y es el que peor responde al tratamiento. Puede explorarse tirando del paciente hacia atrás para comprobar la recuperación del equilibrio (prueba del empujón).

[11]

#### **1.4.1. ALGUNOS SÍNTOMAS ADICIONALES MENOS PROBABLES**

- ✓ Depresión
- ✓ Cambios de personalidad
- ✓ Demencia
- ✓ Interrupción en el ciclo del sueño
- ✓ Problemas al hablar
- ✓ Dificultades en las relaciones sexuales [12]

#### **.1.5. CLASIFICACIÓN DE LOS PARKINSONISMOS**

El concepto de parkinsonismo como expresión clínica de la disfunción de las vías dopaminérgicas nigroestriadas, además de la mejor comprensión de las alteraciones neuropatológicas subyacentes en los pacientes con Parkinson.

✓ PARKINSONISMO PRIMARIO

La enfermedad de Parkinson (EP), Parkinson idiopático, o parálisis agitante, es reconocida como la forma prototipo de parkinsonismo.

✓ PARKINSONISMO POSTENCEFALICO

Representa la secuela clásica de la encefalitis letárgica, de la que muy pocos pacientes la sobreviven hasta hoy día.

✓ PARKINSONISMO IATROGENICO

Relacionado con la indicación de fármacos antipsicóticos antidopaminérgicos, que siendo de amplio uso, se presenta más relacionado con la administración de dosis altas e indicación prolongada en pacientes psiquiátricos.

✓ PARKINSONISMO PLUS

Los síntomas y signos de parkinsonismo pueden presentarse en asociación con otras afecciones degenerativas del sistema nervioso central de presentación relativamente infrecuente. Ejemplo de esto lo constituyen la *Parálisis supranuclear progresiva* o síndrome de Steele-Richardson-Olzewski, las *atrofias olivo-pontocerebelosas*, y el *síndrome de Shy-Drage*.

✓ PARKINSONISMO JUVENIL

Aunque la EP idiopática puede comenzar en pacientes menores de 45 años, es importante reconocer este grupo y considerar diagnósticos diferenciales distintos debe estudiarse la posibilidad de una enfermedad de Wilson, degeneración hepatolenticular que determina trastorno extrapiramidal en jóvenes y es susceptible de tratamiento médico, con resultados importantes para su curso y pronóstico.

✓ PARKINSONISMO SINTOMÁTICO O SECUNDARIO

Se presenta en asociación a trastornos que siendo frecuentes, por localización ocasional comprometen el mesencéfalo y la sustancia negra. Por ejemplo: lesiones traumáticas, isquémicas.

#### ✓ PSEUDOPARKINSONISMOS

Que incluye a un conjunto de afecciones con signos extrapiramidales, como el Temblor esencial benigno, el parkinsonismo aterioloesclerótico, el hidrocefalo normotensivo, y las alteraciones trémulo rígidas hipocinéticas producto del uso de sales de litio, asociadas a hipotiroidismo, a depresión, y trastornos de la marcha por distintas lesiones cerebrales. [13]

### **1.6. IMPORTANCIA DE LOS ANTAGONISTAS DE LA ADENOSINA A<sub>2A</sub>**

Durante las últimas cuatro décadas los científicos se han empeñado en encontrar una cura contra la enfermedad de Parkinson. Esta enfermedad ataca principalmente a personas de edad avanzada dejándolos incapacitado físicamente luego de sufrir este mal, durante los últimos años el número de personas que sufren de dicha enfermedad va en aumento generando una mayor preocupación en la comunidad mundial y a la vez un mayor interés por encontrar una posible cura o un tratamiento para las personas afectadas por esta enfermedad.

Adenosina es un nucleósido endógeno de purinas, presente en el espacio extracelular de los tejidos de mamíferos en el que modula una variedad de procesos fisiológicos. En el sistema nervioso central (SNC), la adenosina actúa como un modulador homeostático mediante el control de la liberación de neurotransmisores o post-sináptica neuronal respuestas. [14]

La triada clásica de las funciones cardinales motoras de la EP incluye la rigidez muscular y temblor en reposo, además de bradicinesia (es decir, la lentitud de movimientos). Así pues, las características adicionales de los antagonistas de los receptores A<sub>2A</sub> de especial relevancia clínica son la capacidad para aliviar la rigidez muscular y para contrarrestar el temblor en la enfermedad de Parkinson.

[15] En un ensayo clínico se indican que la combinación de los antagonistas de los receptores A<sub>2A</sub>, KW 6002 y un sud umbral de dosis de L-DOPA contrarrestado temblor de reposo de manera más eficaz de lo que hicieron otros síntomas cardinales de la EP [16]

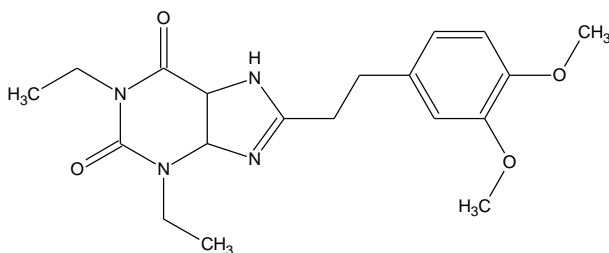


FIGURA 1. Estructura KW-6002 reportada por Gang Yao et al.

## 1.7. QUÍMICA COMPUTACIONAL

La química computacional es una rama de la química teórica cuyos principales objetivos son crear aproximaciones matemáticas eficientes y programas computacionales que calculen las propiedades de las moléculas (tales como la energía total, el momento dipolar, frecuencias vibracionales, densidad electrónica, entre otros) y aplicar estos programas a estudios químicos específicos. El término es también usado para cubrir las áreas de solapamiento entre la ciencia computacional y la química cuántica. [17]. Los programas usados en química computacional están basados en muchos métodos químico-cuánticos diferentes que resuelven la ecuación molecular de Schrödinger asociada con el Hamiltoniano molecular. Los métodos que no incluyen parámetros empíricos o semi-empíricos en sus ecuaciones son derivados directamente de principios teóricos son llamados métodos *ab initio*. La mayoría de las veces esto es referido como cálculos mecánicos cuánticos aproximados. Las aproximaciones hechas en estos casos son matemáticas por naturaleza, como usar una forma funcional más simple u obtener una solución aproximada para una ecuación diferencial complicada. [17]

### **1.7.1 SYBYL.**

Es un programa que está adquiriendo importancia en las investigaciones actuales de química computacional y en particular en la modelación de macromoléculas de interés biológico, ya que posee una serie de herramientas computacionales que nos permiten calcular diferentes propiedades moleculares fundamentales. La base que constituye el paquete de programas SYBYL permite el diseño de nuevos fármacos, nos permite editar y visualizar moléculas de diferentes tamaños y de diferente composición desde el punto de vista químico. Los datos de organización y análisis se distribuyen en una hoja de cálculo, con información química integrada en las herramientas computacionales del programa las cuales pueden de esa forma ser analizadas. [18]

### **1.7.2. TÉCNICAS DE SIMULACIÓN MOLECULAR**

Los métodos de simulación molecular que permiten obtener información microscópica del sistema objeto de estudio se pueden agrupar en dos grandes grupos: Los métodos cuánticos y los métodos clásicos. Los primeros representan el sistema molecular mediante un conjunto de núcleos y electrones, que siguen las leyes fundamentales de la mecánica cuántica. Los métodos clásicos describen el sistema como un conjunto de partículas elementales, localizadas sobre los núcleos, y cuyas interacciones se aproximan a una suma de términos energéticos representados por expresiones basadas en la mecánica clásica. El uso de potenciales clásicos implica una descripción menos rigurosa del sistema, pero abarata en gran medida los gastos computacionales. [19]

### **1.7.3. TÉCNICAS CUÁNTICAS COMPUTACIONALES**

Las técnicas computacionales que ofrecen más confianza y rigurosidad son las basadas en la mecánica cuántica. Estos métodos permiten obtener de forma fiable propiedades tales como la geometría o la energía de sistemas químicos, y dado que consideran los electrones explícitamente, pueden cuantificar las propiedades relacionadas con estos (parámetros espectroscópicos, distribución de cargas,

momento dipolar, etc). Sin embargo, estas ventajas se obtienen a cambio de unos requerimientos computacionales muy elevados y por lo tanto, solo se pueden aplicar a sistemas constituidos por pocos átomos. [17]

La mecánica cuántica postula la existencia de una función de onda (o función de estado), simbolizada por  $\psi$ , que contiene toda la información posible del sistema. Para obtener esta función es necesario resolver la ecuación de Schrödinger:

$$H\psi = E\psi \quad (\text{Ecuación 1})$$

Donde **H** es el operador hamiltoniano que incluye la energía cinética y potencial de núcleos y electrones, y **E** la energía del sistema. A pesar de su aparente simplicidad, la resolución analítica de esta ecuación solo es posible para sistemas de tipo hidrogenoide. Con el objetivo de estudiar sistemas de mayor tamaño se han de introducir toda una serie de simplificaciones entre las que se destaca la aproximación de Born-Oppenheimer, la cual considera, de forma muy razonable, que el movimiento de los núcleos es mucho más lento que el de los electrones. Otras simplificaciones que se introducen son la representación de la función de onda en términos de orbitales moleculares (aproximación orbital), y la expresión de éstos orbitales como combinación lineal de orbitales atómicos (aproximación MO-LCAO). [17]. En la práctica existen tres metodologías básicas para la obtención de la función de onda: la *ab initio*, la semiempírica y la basada en la *teoría del funcional de la densidad (DFT)*. La filosofía de las tres metodologías es completamente diferente. Los métodos *ab initio*, emplean el formalismo químico cuántico de manera rigurosa, no empleando parámetros empíricos más allá de las magnitudes físicas fundamentales. El método del campo autoconsistente o de Hartree-Fock (HF), que es el más sencillo entre los *ab initio*, se basa en el teorema variacional y pretende encontrar la función de onda que minimiza la energía del sistema.

*Los métodos semiempíricos* nacieron con el propósito de proporcionar unas técnicas mecánico-cuánticas para el estudio de propiedades moleculares, de



forma que siendo suficientemente precisos y fiables como para tener un valor práctico en la investigación química, no presenten los inconvenientes de los procedimientos *ab initio*, esto es, que sean aplicables a grandes moléculas.

Este método hace uso de parámetros obtenidos a partir de datos experimentales con el fin de simplificar el cálculo. Son métodos relativamente baratos (en tiempo de computación), En general, resultan apropiados para: a) estudiar sistemas muy grandes para los cuales no se pueden utilizar otros métodos mecano-cuánticos; b) obtener estructuras de partida previas a una optimización *ab initio* o DFT (teoría del funcional de la densidad); c) caracterizar estados fundamentales de sistemas moleculares para los cuales el método semiempírico está bien parametrizado y bien calibrado; d) obtener información cualitativa sobre una molécula (orbitales moleculares, cargas atómicas o modos normales de vibración). [18]

Entre los métodos semiempíricos más populares en la actualidad cabría destacar el *Austin Model 1* (AM1) y el *Parametrization Method 3* (PM3), desarrollados por Dewar [19] y Stewart. [20]

#### **1.7.3.1. Optimización de Geometría.**

En el mundo de las ciencias computacionales, las técnicas aplicadas para conocer las geometrías de las moléculas, es bien conocida como “optimización de geometría”. En esta técnica se toman las aproximaciones geométricas y se hacen lo más exactas posibles. Éstas son una serie de interacciones realizadas sobre la molécula hasta que la energía de la molécula ha llegado a un mínimo. [21]

La geometría de equilibrio de una molécula corresponde a la disposición nuclear que minimiza  $U$ , la energía electrónica molecular incluyendo la repulsión internuclear. La función  $U$  da lo que se llama *superficie de energía potencial* (PES) de la molécula, llamada así debido a que  $U$  es la energía potencial en la ecuación de Schrödinger nuclear. [22]

Si  $U$  depende de dos variables, entonces una representación de  $U(q_1, q_2)$  en tres dimensiones da una superficie en el espacio tridimensional ordinario. Debido al gran número de variables,  $U$  es una “superficie” en un “espacio” abstracto de  $3N - 5$  dimensiones, siendo  $N$  el número de núcleos. Para obtener  $U$ , debemos resolver la ecuación de Schrödinger electrónica para muchas configuraciones nucleares, lo cual es una tarea formidable para una molécula grande. El cálculo de  $U$  para un agrupamiento particular de núcleos, se llama *cálculo de punto simple*, ya que da un punto de la PES molecular. [22]

Una PES está constituida por muchos puntos, cada punto representa una estructura molecular particular, la altura de la superficie para cada punto corresponde a la energía para esta estructura. [22]

Sabemos que una optimización de geometría ha finalizado de manera correcta cuando ésta ha convergido, es decir, cuando las fuerzas son cero, o la diferencia entre pequeñas variaciones de la configuración es muy pequeña. Este tipo de cálculo puede ser realizado mediante el uso de programas como Gaussian, el cual hace uso de los siguientes criterios de convergencia: [23]

- ✓ Las fuerzas deben ser esencialmente cero. Específicamente, el máximo componente de la fuerza, debe tener un valor por debajo de 0.00045 (interpretado como cero).
- ✓ La desviación cuadrática media de las fuerzas, debe ser esencialmente cero (o un valor por debajo de la tolerancia definido en 0.0003).
- ✓ El desplazamiento calculado para el próximo paso debe estar definido en un valor por debajo de 0.0018.
- ✓ La desviación cuadrática media del desplazamiento para el próximo paso debe estar por debajo de 0.0012.

### **1.7.3.2. MÉTODO SEMIEMPÍRICO AM1**

Es un método usado para cálculos cuánticos de la estructura electrónica molecular en química computacional. Está basado en la aproximación integral

NDDO (desprecia el solapamiento diferencial atómico). Específicamente, es una modificación del MNDO (NDDO modificado). Métodos relacionados a él son el PM3 y el antiguo MINDO. Desarrollado por Dewar y colaboradores publicado en 1985 [24], AM1 es un intento improvisado en el modelo MNDO reduciendo la repulsión de los átomos en distancias cercanas de separación. Los resultados de los cálculos en AM1 son algunas veces usados como punto de partida para la parametrización de campos de fuerza en mecánica molecular. [25]

## **1.8. DESCRIPTORES MOLECULARES**

La estructura de las moléculas puede ser descrita por muchas características moleculares las cuales pueden ser medidas o calculadas. Estas son llamadas descriptores moleculares. Los descriptores moleculares son el resultado final de un procedimiento lógico y matemático el cual transforma la información química codificada en una representación simbólica de una molécula en un número o resultado de un experimento. Existen diferentes tipos tales como: índice de información geométrica, electrónica y químico cuántico. [26]

## **1.9. RELACIONES ESTRUCTURA ACTIVIDAD Y MODELO DE PREDICCIÓN (QSAR)**

El campo que relaciona numéricamente la estructura molecular con sus propiedades químicas, fisicoquímica y biológica, es conocido como la relación actividad – estructura (QSAR).

Estas asociaciones son desarrolladas como modelos de predicción y permiten encontrar una ecuación que relacione las propiedades fisicoquímicas de las

moléculas, basándose en datos estructurales y propiedades físicas de las mismas. [27]

El estudio cuantitativo que relaciona la actividad biológica de un grupo de interés con una determinada similitud en su estructura es de gran interés e importancia para predecir cuales son las características estructurales que permiten el comportamiento de estas moléculas bajo determinadas condiciones. El proceso comienza cuando se concibe una hipótesis que relaciona los rasgos químicos de una serie de moléculas con su actividad biológica. [27]

Un modelo QSAR para una propiedad molecular permite determinar específicamente que factores son los directamente responsables en la misma, que tanto puede influir en ella y que parámetros puede afectar. [28]

## **2. RELACIONES ESTRUCTURA ACTIVIDAD Y MODELO DE PREDICCIÓN (HQSAR).**

El método HQSAR (holograma QSAR) ofrece la capacidad de generar rápida y fácilmente modelos de QSAR de alta calidad estadística. Esta técnica HQSAR emplea fragmentos de huellas digitales especializados como variables predictorias de la actividad biológica o de otro tipo de datos relacionados con la estructura. Estas herramientas de HQSAR se aplican para predecir el mejor modelo del sistema, usando la validación cruzada. Los resultados se ven como regiones coloreadas (codificación policromática) en las moléculas individuales, esta codificación policromática denota regiones de contribuciones a las actividades.

El método HQSAR solo requiere de estructuras 2D. En este modo, cada molécula en el conjunto se desglosa en varios fragmentos estructurales, que luego se organizan para formar un holograma molecular, una amplia forma de huella digital que codifica todos los posibles fragmentos moleculares (por ejemplo: lineal, ramificado, interacciones), y mantienen una cuenta del número de apariciones de cada fragmento. [29]

### 2.0.1. CONSTRUCCIÓN DEL MODELO HQSAR.

Los cálculos de Hogramas de un conjunto de moléculas para una base de datos del rendimiento de las estructuras en una matriz de datos de dimensiones N x L, donde N es el número de componentes en la base de datos y L es la longitud del Hograma Molecular, se usan para generar un modelo estadístico que muestre las variables descritas (número de ocupancias de los cubos en el hograma) para una propiedad observable, por ejemplo la actividad biológica expresada como

$-\log IC_{50}$ . [30]

La ecuación que describe el modelo HQSAR es:

$$X = C_0 + \sum_i C_i N_i \quad (\text{Ecuación 9})$$

donde:

- X= La propiedad objetivo (actividad)
- $C_0$ = El set de prueba
- $C_i$ = El coeficiente PLS por el cubo i en el hograma
- $N_i$ = El número de fragmentos incluidos en el cubo i

La implementación de un nuevo algoritmo del método parcial de los mínimos cuadrados (PLS) se usa cuando se asume que hay solo una variable dependiente. El resultado de la implementación de la técnica parcial de los mínimos cuadrados (PLS) está completamente de acuerdo con los resultados del SYBYL PLS. [30]

Algunos puntos concernientes a las opciones usadas en el análisis del PLS están resumidos a continuación:

1. El número máximo de componentes permitido (complejidad del modelo), para cada longitud del hograma analizado, es un set para un número máximo de componentes cualesquiera.

Por ejemplo, con un máximo de 10 componentes de un set determinado y una longitud del holograma de 53.

- En un set de entrenamiento de 15 moléculas el número de componentes que podría tener el set es 7 (15/2)
- En un set de entrenamiento de 100 moléculas el número de componentes que podría tener el set es 10 (desde 10 es más pequeño que 50(100/2) y 26 (53/2)).

2. el poder predictivo del modelo está determinado por el uso de una valoración cruzada estadística. Por convención, se usa el método "leave one out". Otras opciones son la validación no cruzada o validación cruzada usando un número específico de grupos.

3. Para el modelo final, el análisis HQSAR es regresado con el número de componentes del set para el número óptimo de componentes identificado por el análisis de validación cruzada, y el número de grupos de validación cruzada del set para cero.

La selección de una longitud de holograma que conduzcan al "mejor" HQSAR se basa en el análisis del PLS que da cualquiera de las más altas validaciones cruzadas  $r^2$  o el más bajo error estándar asociado con el análisis de validación cruzada. Los resultados impresos de la validación cruzada después de incluir el número óptimo de componentes, corresponden al valor máximo de  $r^2$  o al menor error estándar dependiendo del criterio elegido. [30]

### **2.0.2. INTERPRETACIÓN DE LOS MODELOS HQSAR.**

Una importante función de un modelo HQSAR, además de predecir las actividades de moléculas, es proporcionar consejos acerca de que posibles fragmentos moleculares son importantes contribuyentes a la actividad. El químico sintético puede combinar esta información con el conocimiento de la máxima estructura común para obtener nuevas moléculas que puedan conducir a la síntesis de

nuevas drogas. Esta sección proporciona una breve descripción de la técnica utilizada para abordar esta cuestión. **[30]**

La contribución a la actividad de cada átomo en una molécula en el conjunto de datos se calculará de la siguiente manera:

1. Generar cada fragmento en el holograma, a su vez.
2. Identificar el papel del fragmento y el coeficiente correspondiente de PLS que se le asignara en el holograma. Este valor se considera la contribución del fragmento a la actividad.
3. La contribución a la actividad de cada átomo en el fragmento se toma como el coeficiente PLS de que fragmentan (en el paso anterior) dividido por el número de átomos en el fragmento. De este modo, todos los átomos se supone que contribuyen por igual a la actividad de un determinado fragmento. Si un fragmento se encuentra dos veces se cuenta dos veces.
4. La contribución total a la actividad de un determinado átomo se obtiene mediante la suma de las contribuciones individuales atómicas de cada uno de los fragmentos que contienen el átomo. **[30]**

### **2.0.3. CÓDIGO DE COLORES HQSAR.**

Los cálculos muestran la contribución atómica a la actividad, las moléculas reflejan la contribución atómica individual a la actividad por medio de un código de colores. Los colores como rojo, rojo naranja y naranja reflejan pobre o negativa contribución a la actividad, mientras colores como verde, azul verdoso y amarillo, reflejan favorable o positiva contribución. Los átomos con contribución intermedia o neutra son coloreados de blanco. **[30]**

## 2.1. ESTADÍSTICA INFERENCIAL

La estadística inferencial nos permite trabajar con una variable a nivel de intervalo o razón, así también se puede comprender la relación de dos o más variables y nos permitirá relacionar mediante ecuaciones, una variable en relación de la otra variable llamándose regresión lineal y una variable en relación a otras variables llamándose regresión lineal múltiple.

Casi constantemente en la práctica de la investigación estadística, se encuentran variables que de alguna manera están relacionados entre sí, por lo que es posible que una de las variables puedan relacionarse matemáticamente en función de otra u otras variables. [31]

### Procedimiento para hallar el Coeficiente de Regresión

Para determinar el valor del coeficiente de regresión de una manera fácil y exacta es utilizado el método de los mínimos cuadrados de dos maneras:

#### 1.- Forma Directa

De la ecuación de la recta:

$$Y = a_0 + a_1x \quad (\text{Ecuación 10})$$

Si  $a_0$  y  $a_1$ , se obtienen a partir de las ecuaciones normales:

$$\sum y = a_0N + a_1\sum x \quad (\text{Ecuación 11})$$

$$\sum xy = a_0\sum x + a_1\sum x^2 \quad (\text{Ecuación 12})$$

Aplicando normales Y sobre X tenemos:



$$a_1 = \frac{n \sum xy - \sum x \sum y}{N \sum x^2 - (\sum x)^2} \quad (\text{Ecuación 13})$$

El coeficiente de regresión es =  $a_1$

De la misma manera la recta de regresión de “X” sobre “Y” será dada de la siguiente manera:

$$X = b_0 + b_1 y \quad (\text{Ecuación 14})$$

Donde:  $b_0$  y  $b_1$  se obtienen a partir de las ecuaciones normales:

$$\sum x = b_0 N + b_1 \sum y \quad (\text{Ecuación 15})$$

$$\sum xy = b_0 \sum y + b_1 \sum y^2 \quad (\text{Ecuación 16})$$

Aplicando normales X sobre Y tenemos:

$$b_0 = \frac{\sum x \sum y^2 - \sum y \sum xy}{N \sum y^2 - (\sum y)^2} \quad (\text{Ecuación 17})$$

$$b_1 = \frac{N \sum xy - \sum x \sum y}{N \sum y^2 - (\sum y)^2} \quad (\text{Ecuación 18})$$

El coeficiente de Regresión es =  $b_1$

## 2. Forma Indirecta del método de los mínimos cuadrados.

El fundamento de este método es de las desviaciones de X respecto a su media aritmética X.

$$y = \left[ \frac{\sum xy}{\sum x^2} \right] x \quad (\text{Ecuación 19}) \qquad x = \left[ \frac{\sum xy}{\sum y^2} \right] y \quad (\text{Ecuación 20})$$

Ecuación de **y** sobre **x** y ecuación de **x** sobre **y**, donde:

$$x = X - X^- \quad (\text{Ecuación 21})$$

$$y = Y - Y^- \quad (\text{Ecuación 22})$$

x, y = desviaciones, X = media aritmética, Y = media aritmética. **[54]**

### 2.1.1. REGRESIÓN LINEAL MÚLTIPLE

La regresión lineal múltiple es uno de los métodos estadísticos más ampliamente utilizados. En química este método es el indicado para evaluar la relación entre un conjunto de variables (X) llamadas independientes y una variable (Y) denominada dependiente. **[32]**

### 2.2.2. ANÁLISIS DE REGRESIÓN MÚLTIPLE

En el caso general, el modelo de regresión lineal múltiple con k variables responde a la ecuación:

$$Y_i = -\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i \quad (\text{Ecuación 23})$$

Donde:

i : indica uno de los valores de la población para cada variable.

$\alpha$  : es la ordenada en el origen

$\beta_{ki}$  : es la pendiente de la relación lineal entre  $X_{ki}$  y  $Y_i$

$\varepsilon_i$  : Término de error o residual

Los coeficientes  $\beta_{ki}$  son estimados siguiendo el criterio de mínimos cuadrados.

Una de la principales dificultades a la hora de ajustar un modelo de regresión múltiple surge cuando es necesario identificar entre el conjunto de variable disponible, aquellas que están correlacionadas con la respuesta y que la predicen de la mejor forma posible. Una alternativa es construir un modelo por inclusión o hacia adelante (forward) , considerando en primer lugar la relación de cada variable con la respuesta e ignorando las demás variables, valorándola por medio de las medidas de ajuste del modelo como la significancia, coeficiente de correlación, entre otros parámetros estadísticos. Aquella que muestre un mejor ajuste será introducida en un modelo inicial.

El segundo paso consiste en seleccionar entre las variables restantes la que al introducirla en el modelo permita explicar una mayor parte de la variabilidad residual. La comparación entre distintos modelos debe hacerse en términos del valor relativo de los coeficientes de determinación y el contraste F parcial. Este esquema será repetido hasta que ninguna otra variable entre a formar parte del modelo. [32]

### **2.2.3. COEFICIENTE DE CORRELACIÓN MÚLTIPLE**

En el contexto del análisis de la regresión lineal simple el coeficiente de correlación múltiple mide el grado de asociación entre dos variables. El coeficiente de correlación R, varía entre -1 y 1. Si R es cercano a cero se concluye que no existe una correlación lineal significativa entre las variables  $X_1, X_2, \dots, X_k$  y Y pero si R está cerca de -1 o +1, decimos que existe una buena correlación lineal significativa entre las variables. [33] Se define, a partir de los n pares de observaciones, mediante: [34]

$$1 - R = \frac{SSE}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (\text{Ecuación 25})$$

Verificándose que  $-1 \leq R \leq 1$

Su cuadrado,  $R^2$ , denominado *coeficiente de determinación múltiple*, puede interpretarse como el porcentaje de variabilidad de Y explicada o debida a la recta de regresión, en tanto que puede comprobarse que:

$$1 - R = \frac{SSE}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (\text{Ecuación 26})$$

Cuando todos los puntos se encuentran sobre la recta de regresión estimada, es decir, "el ajuste es perfecto", la suma de cuadrados de residuos, SSE, toma el valor cero y, por tanto,  $R^2 = 1$ . El denominador de la última expresión es una medida de la variabilidad total de las n observaciones de la variable respuesta.

**[34]**

### 2.3. VALIDACIÓN CRUZADA

En un análisis HQSAR el resultado final es un modelo matemático. Para estar completamente seguros de que el modelo obtenido no resulta de un ordenamiento al azar de los datos es común verificar la calidad del mismo.

El método de regresión lineal múltiple proporciona los valores para los coeficientes que hacen que la función matemática propuesta explique al máximo como varía la retención. Independientemente de esto siempre quedará una varianza no explicada por el modelo -los residuales-, que corresponden a los errores en las medidas de los datos de retención y a posibles efectos no incluidos en él. **[35]**

Por lo cual se utiliza la técnica de "Validación Cruzada" (Cross validation) con el procedimiento conocido como *leave-one-out* que consiste en excluir una molécula de la base de datos y se recalcula de nuevo el modelo con las N-1 moléculas

restantes, el dato excluido es calculado con el nuevo modelo obtenido. Esto es realizado para cada una de las N moléculas usadas para desarrollar el modelo. Luego se correlacionan los datos calculados para cada una de las moléculas obtenidas por el método anterior con los datos experimentales. En este caso el coeficiente de correlación debe ser igual o similar al encontrado en el método. **[20]**

### 3. METODOLOGÍA

La metodología del presente trabajo fue llevada a cabo en varias etapas, todas ellas desarrolladas en las instalaciones del Grupo de Química Cuántica y Teórica de la Universidad de Cartagena.

El conjunto de datos y de moléculas utilizadas en el análisis se obtuvieron de la literatura [1], para mayor conveniencia, la lista de datos y compuestos se muestran en la Tabla 1a–1f, las cuales presentan la estructura mostrada y la nomenclatura usada para describir las posiciones de los sustituyentes en la Figura 2-7.

**Figura 2. Compuesto C 1-4v referenciado en la Tabla 1a.**

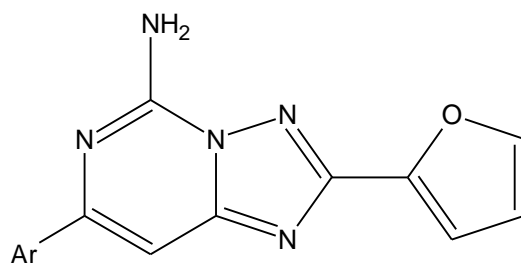


Tabla 1a. Estructura química y valores pK<sub>i</sub> (nM) de las moléculas C1-4v antagonistas del receptor de adenosina A2A

Compuesto	Sustitúyete	pK <sub>i</sub> Experimental
	Ar	
C1	Ph	8.15
C4a	2-CH <sub>3</sub> -Ph	7.66
C4b <sup>MP</sup>	3-CH <sub>3</sub> -Ph	8.43
C4c	4-CH <sub>3</sub> -Ph	8.08
C4d	2-CH <sub>3</sub> O-Ph	8.37
C4e	3-CH <sub>3</sub> O-Ph	8.30
C4f	4-CH <sub>3</sub> O-Ph	8.14
C4g	2,4-(CH <sub>3</sub> O) <sub>2</sub> -Ph	8.54
C4h	3,4-(CH <sub>3</sub> O) <sub>2</sub> -Ph	8.14
C4i	2,5-(CH <sub>3</sub> O) <sub>2</sub> -Ph	8.35
C4j	2,6-(CH <sub>3</sub> O) <sub>2</sub> -Ph	5.77
C4k	2,3-(CH <sub>3</sub> O) <sub>2</sub> -Ph	7.66
C4l	3,4,5-(CH <sub>3</sub> O) <sub>3</sub> -Ph	8.35
C4m	2,3,4-(CH <sub>3</sub> O) <sub>3</sub> -Ph	8.35
C4n	2-Tiofenil	8.36
C4o	4-Piridina	8.72
C4p	3-Piridina	8.18
C4q	2-Piridina	8.85
C4r	3-( <sup>i</sup> Pr)-Ph	8.33
C4s	3-(Ph)-Ph	8.26
C4t <sup>MP</sup>	3-(NH <sub>2</sub> )-Ph	8.30
C4u	3-(CN)-Ph	8.27
C4v	3-(CH <sub>2</sub> OH)-Ph	8.55

MP. Moléculas de prueba

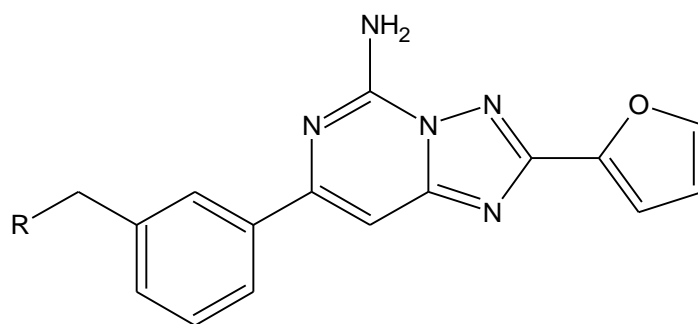
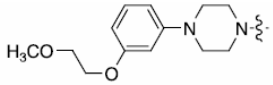
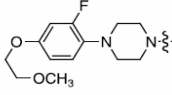
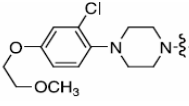
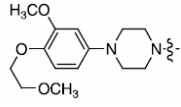
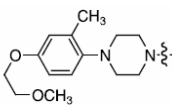
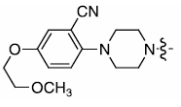
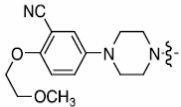
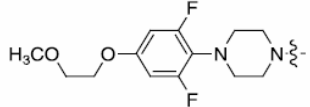
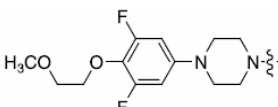
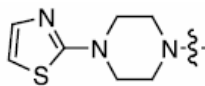


Figura 3. Compuesto C 8a-8q referenciado en la Tabla 1b.

Tabla 1b. Estructura química y valores pK<sub>i</sub> (nM) de las moléculas C8a-8q antagonistas del receptor de adenosina A<sub>2A</sub>

Compuesto	Sustituyente	pK <sub>i</sub> Experimental
	R	
C8a		8.44
C8b		8.34
C8c		8.13
C8d <sup>MP</sup>		8.05
C8e		8.34
C8f <sup>MP</sup>		7.98
C8g		8.55



C8h		8.40
C8i		8.57
C8j		8.54
C8k		8.48
C8l		8.57
C8m		8.44
C8n		8.96
C8o <sup>MP</sup>		8.19
C8p		8.66
C8q		8.62

**MP. Moléculas de prueba**

Figura 4. Compuesto E7-14referenciado en la Tabla 1c.

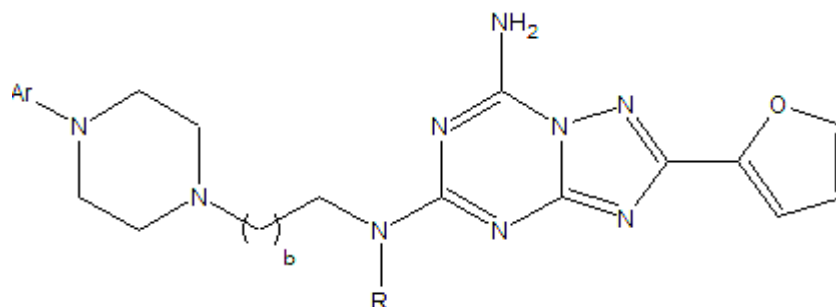
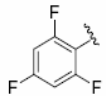
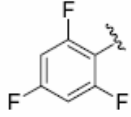
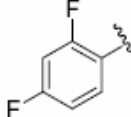
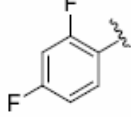
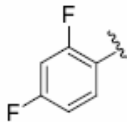


Tabla 1c. Estructura química y valores pKi (nM) de las moléculas E7-14 antagonistas del receptor de adenosina A2A

Compuesto	Sustitúyete	Valor	Sustitúyete	pKi Experimental
	Ar	b	R	
E7	Ph	1	H	7.66
E8	Ph	2	H	6.72
E9		1	H	7.62
E10		2	H	7.00
E11		1	H	8.15
E12		2	H	7.03

E13 <sup>MP</sup>	Ph	1	CH <sub>3</sub>	7.96
E14		1	CH <sub>3</sub>	8.40

MP. Moléculas de prueba

Figura 5. Compuesto E16-19 referenciado en la Tabla 1d.

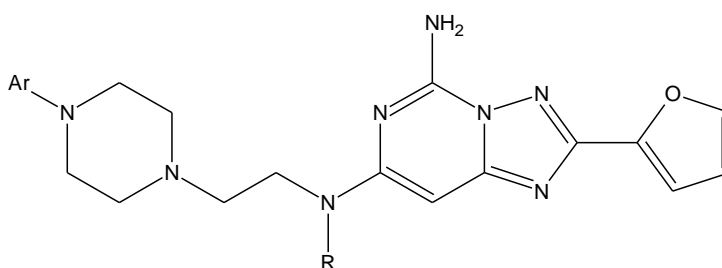
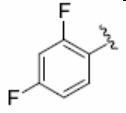
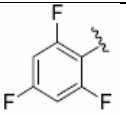
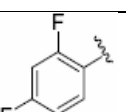
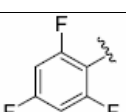


Tabla 1d. Estructura química y valores pK<sub>i</sub> (nM) de las moléculas E16-19 antagonistas del receptor de adenosina A<sub>2A</sub>

Compuesto	Sustituyente	Sustituyente	pK <sub>i</sub> Experimental
	Ar	R	
E16		H	7.26
E17		H	7.04
E18		CH <sub>3</sub>	8.19
E19 <sup>MP</sup>		CH <sub>3</sub>	7.62

MP. Moléculas de prueba

Figura 6. Compuesto H8-25referenciado en la Tabla 1b.

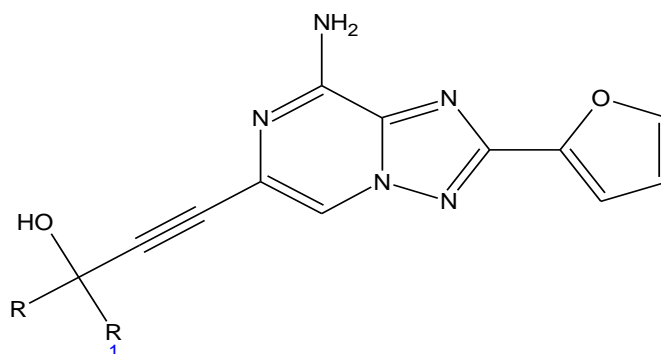


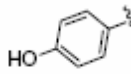
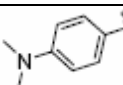
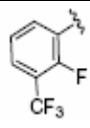





Tabla 1e. Estructura química y valores pKi (nM) de las moléculas H8-25 antagonistas del receptor de adenosina A2A.

Compuesto	Sustituyente		pKi Experimental
	R	R <sub>1</sub>	
H8		Me	8.43
H9			8.15
H10			8.05
H12		Me	7.92
H13 <sup>MP</sup>		H	6.47
H14		CF <sub>3</sub>	7.13
H15		Me	8.96
H16		Me	7.60
H17		Me	7.48

H18		Me	7.48
H19		H	6.59
H20		Me	7.38
H21		CF <sub>3</sub>	5.96
H22		ME	5.80
H23			6.52
H24			8.14
H25			7.92

MP. Moléculas de prueba

Figura 7. Compuesto H26-33referenciado en la Tabla 1b.

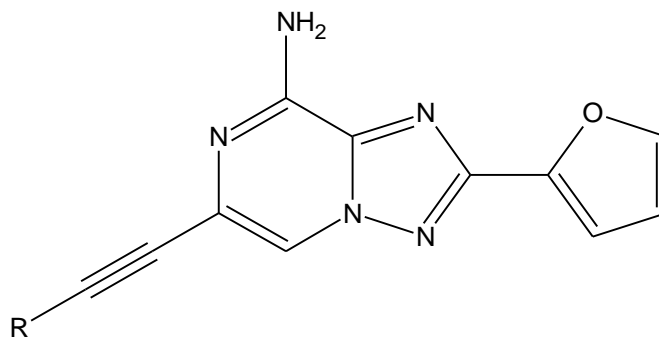
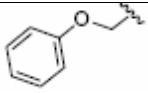
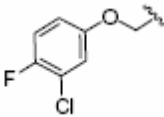
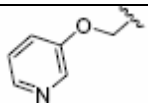
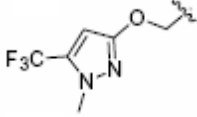
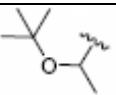
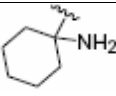
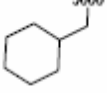


Tabla 1f. Estructura química y valores pKi (nM) de las moléculas H26-33 antagonistas del receptor de adenosina A2A.

Compuesto	Sustituyente	pKi Experimental
	R	
H26 <sup>MP</sup>		6.41
H27 <sup>MP</sup>		6.51
H28		7.04
H29		7.05
H30		7.43
H31		6.68
H33		6.55

**MP. Moléculas de prueba**

La metodología utilizada consta de las siguientes etapas:

- ❖ Selección y generación de las estructuras 3D de las moléculas de estudio.
- ❖ Optimización de las estructuras
- ❖ Análisis estadístico para la obtención del modelo
- ❖ Validación del modelo

A continuación se describe con más detalle cada una de las diferentes etapas.

### **3.1. SELECCIÓN Y GENERACION DE LAS ESTRUCTURAS 3D DE LAS MOLECULAS DE ESTUDIO.**

En el presente trabajo se estudiaron 78 moléculas reportadas por (Yao G. et al 2005), (Vu B. et al 2004) y (Matasi Julios et al 2005), se seleccionaron 76 moléculas antagonistas del receptor de adenosina A2A ya que 2 moléculas no presentaban valores de Ki definidos. Estas 76 moléculas fueron dibujadas en el programa CORINA\_3D que es un software que sirve para generar estructuras tridimensionales. ([http://www.molecularnetworks.com/online\\_demos/corina\\_demo.html](http://www.molecularnetworks.com/online_demos/corina_demo.html)).

CORINA genera estructuras 3D a partir de estructuras dadas en 2D. El rendimiento, velocidad y fiabilidad de CORINA es una perfecta aplicación para la conversión de grandes conjuntos de datos. Posteriormente estas estructuras tridimensionales generadas en CORINA fueron revisadas y corregidas en el programa AMPAC GUI8.

### **3.2 OPTIMIZACION DE GEOMETRIAS**

La optimización de geometrías de las moléculas fueron ejecutadas con el conjunto de programas gaussian03, estos cálculos se realizaron mediante el método semi-empírico AM1 obteniendo así un mínimo de energía para cada uno de los compuestos en estudio, lo que permitió así obtener moléculas con una conformación estable y adecuada.

### **3.3. ANÁLISIS ESTADÍSTICO Y OBTENCIÓN DEL MODELO**

El análisis, los cálculos y la visualización del modelo HQSAR obtenido que relaciona la actividad de las moléculas de estudio con la estructura molecular se realizaron con el paquete SYBYL 7.3. El HQSAR utiliza diferentes parámetros para la generación del modelo. Los parámetros como: tamaño de fragmento, longitud del holograma y distinción de fragmentos son al mismo tiempo importante para la generación de hologramas moleculares.

El modelado HQSAR se llevó a cabo como ya se describió, mediante el mismo tipo de parámetros estadísticos utilizando el paquete SYBYL 7.3 para el desarrollo del modelo y la generación de los fragmentos se utilizó la siguiente serie de distinción de fragmentos:

- Los átomos (A)
- Los enlaces (B)
- Las conexiones (C)
- Los átomos de Hidrogeno (H)
- La quiralidad (Ch)
- Donantes y Aceptores (DA)

Posteriormente con el fin de evaluar el proceso de generación del holograma, se utilizaron varias combinaciones de los parámetros de estudio (distinción de fragmentos), como sigue: A/B, A/B/C, A/B/C/H, A/B/C/H/Ch, A/B/C/H/Ch/DA, A/B/H, A/B/C/Ch, A/B/DA, A/B/C/DA, A/B/H/DA, A/B/C/Ch/DA, A/B/C/H/DA, y A/B/H/Ch/DA. Los análisis fueron ejecutados sobre las 12 series de longitudes del holograma establecidas de 53, 59, 61, 72, 83, 97, 151, 199, 257, 307, 353 y 401. usando el valor por defecto de tamaño de fragmento de (4-7) ; con el fin de obtener mejores valores estadísticos, se aplica el segundo criterio estadístico utilizando distintos tamaños de fragmentos :(2-5), (3-6), (4-7), (5-8), (6-9), (9-10) el cual se refiere a la mínima y máxima longitud de los fragmentos moleculares, para luego por medio de estos distintos tamaños de fragmentos escoger el mejor modelo.

La calidad de los modelos se analizó estadísticamente teniendo en cuenta el coeficiente de correlación ( $r^2$ ) y se evaluó por los valores de ( $q^2$ ).



### 3.4. VALIDACIÓN DEL MODELO

El modelo fue validado mediante el cálculo de los valores de  $q^2$ , los valores de ( $q^2$ ) se calculan a partir de “leave-one-out” (LOO), también conocido como la técnica de validación cruzada, el cual se basa en que uno de los compuesto se omite de la serie de datos y se vuelve a calcular la regresión, el valor determinado se compara con el valor real y esto se repite hasta que cada dato sea omitido una vez, la suma de los cuadrados de esta residual omisión se utiliza para calcular  $q^2$ . Los valores de ( $q^2$ ) pueden considerarse una medida de la potencia de predicción. Para tal fin separamos en dos grupos el conjunto de las 76 moléculas seleccionadas, de las cuales 10 moléculas se excluyeron como grupo de prueba y por lo tanto las 66 moléculas restantes conformaron el grupo de entrenamiento.

Todos los valores de  $K_i$  (nM) fueron transformados en valores de  $pK_i$  (nM), del grupo de entrenamiento se obtuvo un nuevo modelo con el cual se predijo el valor de  $pK_i$  de las 10 moléculas pertenecientes al grupo de prueba.

Una vez generados los datos mediante esta técnica se realizó una correlación con los  $pK_i$  calculados y los experimentales reportados en la literatura, hallando así un nuevo coeficiente de correlación múltiple el cual fue comparado con el que aparece en el modelo, teniendo en cuenta que estos deben ser iguales o similares para que el modelo propuesto sea aceptado.

## 4. RESULTADOS Y DISCUSIONES

Inicialmente para el desarrollo del modelo HQSAR utilizamos los diversos tipos de fragmentos es decir, Átomos (A), Enlaces (B) y Conexiones (C), Átomos de Hidrógenos (H), Quiralidad (Ch) y Donantes y Aceptores (DA) y se desarrollaron numeroso modelos con las combinaciones de estos parámetros como A/B, A/C, A/B/C, A/H/Ch, A/B/Ch, A/B/DA, A/B/C/Ch, A/B/C/DA, A/B/H/Ch, A/B/H/DA, A/B/Ch/DA, y A/B/C/H/Ch/DA.

Todos estos modelos fueron construidos con el valor por defecto del tamaño de fragmento de (4-7) a través de las 12 diversas longitudes del holograma, los cuales son un conjunto de números primos que van desde 53 a 401. La idea de fondo es ajustar todos los parámetros existentes en la técnica HQSAR para la generación del modelo más robusto posible. La Tabla 2 resume los resultados de diferentes tipos de fragmentos y diferentes longitudes del holograma. La combinación teniendo como tipo de fragmentos, Átomos (A), Enlace (B), Conexiones (C), y Donantes y Aceptores (DA) arrojó el mejor modelo con valores de ( $q^2= 0.402$  y  $r^2= 0.817$ ). (Modelo 8)

Los resultados también indican que el uso de otras distinciones de los fragmentos moleculares disminuye la calidad de los modelos según la medición estadística. Cabe señalar que nuestros resultados confirman estudios previos, que indican que cuando donante y aceptor están siendo utilizados en la generación del fragmento, los átomos de hidrogeno no deben ser considerados. Esto puede explicarse por el aumento drástico en el número de fragmentos generados cuando ambos se consideran como opciones en el módulo de construcción (30).

**Tabla 2. Resultados de los análisis HQSAR de la influencia de diversas distinciones de fragmentos sobre los principales parámetros estadísticos utilizando los fragmentos de tamaño por defecto (4-7)**

Modelo	Distinción de fragmento	$r^2$	SEE	$q^2$ (LOO)	SEP	HL	N
1	A/B	0.552	0.529	0.337	0.644	97	3
2	A/C	0.787	0.378	0.403	0.632	53	7
3	A/B/C	0.444	0.585	0.306	0.654	59	2
4	A/H/Ch	0.508	0.555	0.417	0.604	401	3
5	A/B/H	0.572	0.522	0.339	0.648	151	4
6	A/B/DA	0.758	0.395	0.364	0.641	353	5
7	A/B/C/Ch	0.553	0.529	0.328	0.649	61	3
<b>8</b>	<b>A/B/C/DA</b>	<b>0.817</b>	<b>0.350</b>	<b>0.402</b>	<b>0.632</b>	<b>53</b>	<b>7</b>
9	A/B/H/Ch	0.655	0.472	0.400	0.623	59	5
10	A/B/H/DA	0.688	0.449	0.450	0.596	59	5
11	A/B/Ch/DA	0.755	0.398	0.372	0.637	257	5
12	A/B/C/H/Ch/DA	0.724	0.426	0.371	0.643	97	6

$r^2$ , coeficiente de correlación; SEE, error estándar de  $r^2$ ;  $q^2$ , coeficiente de correlación de validación cruzada; SEP, error estándar de  $q^2$ ; HL, longitud del holograma; N, número de componentes.

A los mejores modelos HQSAR generados se les realizó un análisis PLS para investigar si con diferentes tamaños de fragmentos, referente a la máxima y mínima longitud de los fragmentos moleculares, se podría mejorar los parámetros estadísticos. Los resultados HQSAR de los distintos tamaños de fragmentos se resumen en la Tabla 3. El mejor modelo se obtuvo mediante un holograma de

longitud de 199, número óptimo de componentes de 10, un tamaño de fragmento de (7-10) y usando Átomos (A), Enlace (B), Conexiones (C), y Donantes y Aceptores (DA) como tipos de fragmentos. Una diferencia significativa se observó en los valores estadísticos cuando al modelo se le modificó el tamaño fragmento de (1-4) a (7-10), al aumentar el tamaño de los fragmentos de (7-10) se vió favorecido por la mejora de los datos estadísticos, expresada como ( $r^2 = 0.903$  y  $q^2 = 0.476$ ).

**Tabla 3. Parámetros estadísticos obtenidos del modelo 8 (A/B/C/DA) con los diferentes tamaños de fragmentos**

Tamaño del fragmento	$r^2$	SEE	$q^2$ (LOO)	SEP	HL	N
1-4	0.651	0.471	0.417	0.609	199	4
2-5	0.710	0.433	0.418	0.614	199	5
3-6	0.699	0.441	0.422	0.611	71	5
4-7	0.817	0.350	0.402	0.632	53	7
5-8	0.662	0.464	0.347	0.644	97	4
6-9	0.703	0.434	0.389	0.623	97	4
<b>7-10</b>	<b>0.903</b>	<b>0.262</b>	<b>0.476</b>	<b>0.608</b>	<b>199</b>	<b>10</b>

$r^2$ , coeficiente de correlación; SEE, error estándar de  $r^2$ ;  $q^2$ , coeficiente de correlación de validación cruzada; SEP, error estándar de  $q^2$ ; HL, longitud del holograma; N, número de componentes.

El mejor modelo HQSAR se obtuvo con la combinación de parámetros A/B/C/DA presentando valores estadísticos de ( $r^2= 0.903$  y  $q^2= 0.476$ ) con un cambio de tamaño de fragmento de (7-10), longitud del holograma de (HL=199), estándar de

error de la correlación de (Std. Error =0.262) y un error estándar de la validación cruzada de (Std.Error – Cross Val = 0.608). A partir de estos datos, después de la aplicación de varias técnicas estadísticas para la detección de valores atípicos tales como el diagrama de caja y bigotes para la diferencia de los valores observados menos predichos arrojan puntos por fuera del intervalo comprendido entre  $(\bar{x} \pm 3s)$  donde  $\bar{x}$  es la media de los datos en cuestión y  $s$ , la desviación estándar de los mismos, se encontró que las moléculas H23, H24 y H25 presentaron grandes valores residuales y deben ser consideradas como outliers. Es considerable que los compuestos (H23, H24 y H25) Figura 6-Tabla 1e, presentan una naturaleza estructural diferente, con la presencia de alcoholes propargil sustituidos directamente al grupo acetileno en comparación de otras moléculas del conjunto de datos que cuentan con la presencia de un segundo sustituyente ya sea arilo o alquilo dándole un carácter quiral, característica que no presentan las moléculas H23,H24 y H25, probablemente por estas razones se encuentran fuera del ámbito de la predicción del modelo. La eliminación de estos compuestos (H23, H24, H25) y el posterior re-análisis del conjunto de datos (73 moléculas). Da como resultado un HQSAR con una mejor calidad, favorecido por la mejora de los datos estadísticos, expresada como ( $r^2 = 0.947$  y  $q^2 = 0.534$ ). Con la combinación de parámetros A/B/C/DA presentando un cambio de tamaño de fragmento de (7-10), longitud del holograma de (HL=199), estándar de error de la correlación de (Std. Error =0.193) y un estándar de error de la validación cruzada de (Std.Error – Cross Val = 0.575).

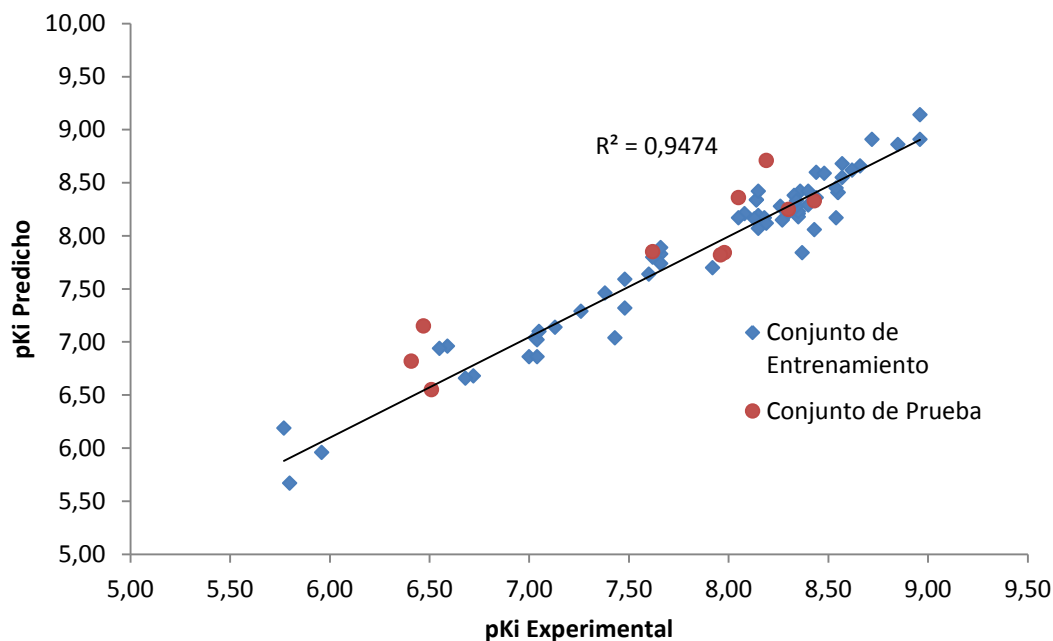
#### **4.1. VALIDACIÓN DEL MODELO**

En cuanto a la validación de un modelo HQSAR, una medida de consistencia interna está disponible en forma de  $q^2$ . Sin embargo, la prueba más importante en un modelo HQSAR es su capacidad para predecir el valor de la actividad biológica para los nuevos compuestos. Como la estructura codificada en una huella digital

en 2D está directamente relacionada con la actividad biológica de las moléculas en el conjunto de entrenamiento, El modelo HQSAR debe ser capaz de predecir la actividad biológica de nuevos compuestos a partir de su huella dactilar. Por lo tanto, el poder de predicción del modelo HQSAR, utilizando 63 moléculas del conjunto de entrenamiento se evaluó mediante la predicción de los valores de  $pK_i$  para 10 moléculas de prueba, que no se incluyeron en el conjunto de entrenamiento para la generación del modelo, el proceso de validación externa puede ser considerado el más valioso método de validación. Los resultados obtenidos sobre la evaluación del modelo se muestran en la Tabla 4, mientras que la representación grafica de los datos experimentales frente a los predichos de ambos conjuntos de entrenamiento (generación del modelo) y de prueba (evaluación externa) se muestran en la Figura 8.

Los resultados muestran que en conjunto de las moléculas de prueba, las cuales representan características estructurales diferentes incorporadas dentro del conjunto de entrenamiento, son razonablemente bien predichas. El buen acuerdo entre los valores de actividad experimental y predicha indica la alta confiabilidad de nuestro modelo HQSAR. El poder predictivo del modelo generado presenta una buena correlación ( $r^2_{\text{pred}} = 0.931$ ). Lo que es notable dado el universo químico de las estructuras de las cuales se derivó el modelo.

**Figura 8.** Valores experimentales y predichos de  $pK_i$  a partir del modelo HQSAR



Los valores residuales moderadamente bajos, representan un modelo HQSAR confiable y que puede ser usado para predecir la actividad biológica de nuevos compuestos estructurales. Se predijeron valores  $pK_i$  que son muy parecidos a los valores  $pK_i$  experimentales, estos valores se apartan en promedio entre sí a  $\pm 0.3$  unidades.

La única excepción son los compuestos H13 Y C8O, en la prueba con desviaciones de  $\pm 0.7$  y  $0.5$ , respectivamente, de donde se puede concluir que no existe diferencia significativa entre los valores reales y los predichos por el modelo, esto debido a que los resultados obtenidos proporcionaron valores reproducibles para predecir las actividades de los compuestos en estudio.

**Tabla 4. Actividades observadas versus las predichas utilizando el Modelo 8 HQSAR.**

<b>Compuestos</b>	<b>pKi Observado</b>	<b>pKi predicho</b>	<b>Valor residual</b>
<b>Conjunto de entrenamiento</b>			
C1	8.15	8.19	-0.04
C4A	7.66	7.83	-0.17
C4C	8.08	8.21	-0.13
C4D	8.37	7.84	0.53
C4E	8.3	8.22	0.08
C4F	8.14	8.34	-0.2
C4G	8.54	8.17	0.37
C4H	8.14	8.34	-0.2
C4I	8.35	8.21	0.14
C4J	5.77	6.19	-0.42
C4K	7.66	7.89	-0.23
C4L	8.35	8.18	0.17
C4M	8.35	8.23	0.12
C4N	8.36	8.42	-0.06
C4O	8.72	8.91	-0.19
C4P	8.18	8.17	0.01
C4Q	8.85	8.86	-0.01
C4R	8.33	8.38	-0.05



C4S	8.26	8.28	-0.02
C4U	8.27	8.15	0.12
C4V	8.55	8.41	0.14
C8A	8.44	8.60	-0.16
C8B	8.34	8.28	0.06
C8C	8.13	8.16	-0.03
C8E	8.34	8.33	0.01
C8G	8.55	8.41	0.14
C8H	8.4	8.29	0.11
C8I	8.57	8.68	-0.11
C8J	8.54	8.45	0.09
C8K	8.48	8.59	-0.11
C8L	8.57	8.55	0.02
C8M	8.44	8.36	0.08
C8N	8.96	9.14	-0.18
C8P	8.66	8.66	0.00
C8Q	8.62	8.62	0.00
E10	7	6.86	0.14
E11	8.15	8.07	0.08
E12	7.03	7.04	-0.01
E14	8.4	8.42	-0.02
E16	7.26	7.29	-0.03

E17	7.04	7.02	0.02
E18	8.19	8.12	0.07
E7	7.66	7.74	-0.08
E8	6.72	6.68	0.04
E9	7.62	7.80	-0.18
H10	8.05	8.17	-0.12
H12	7.92	7.70	0.22
H14	7.13	7.14	-0.01
H15	8.96	8.91	0.05
H16	7.6	7.64	-0.04
H17	7.48	7.59	-0.11
H18	7.48	7.32	0.16
H19	6.59	6.96	-0.37
H20	7.38	7.46	-0.08
H21	5.96	5.96	0.00
H22	5.8	5.67	0.13
H28	7.04	6.86	0.18
H29	7.05	7.10	-0.05
H30	7.43	7.04	0.39
H31	6.68	6.66	0.02
H33	6.55	6.94	-0.39
H8	8.43	8.06	0.37

H9	8.15	8.42	-0.27
<b>Compuestos</b>	<b>pKi Observado</b>	<b>pKi predicho</b>	<b>Valor residual</b>
<b>Conjunto de prueba</b>			
C4B	8.43	8.33	0.1
C4T	8.3	8.25	0.05
C8D	8.05	8.36	-0.31
C8F	7.98	7.84	0.14
C8O	8.19	8.71	-0.52
E13	7.96	7.82	0.14
E19	7.62	7.85	-0.23
H13	6.47	7.15	-0.68
H26	6.41	6.82	-0.41
H27	6.51	6.55	-0.04

Una importante propiedad de la técnica HQSAR, además de predecir las actividades de las moléculas de prueba, es proporcionar consejos acerca de los fragmentos moleculares que pueden ser importantes contribuyentes a la actividad biológica. Se utilizó un código de colores con el fin de mostrar las principales contribuciones atómicas individuales de cada molécula a la actividad. Los colores que reflejan contribuciones pobres (negativas) se encuentran en el extremo rojo del espectro (rojo, rojo-naranja y naranja), mientras que los colores que reflejan contribuciones favorables (positivas) se encuentran en el extremo verde (amarillo, verde-azul y verde). Los átomos coloreados de blanco representan contribuciones intermedias.

Las regiones coloreadas de cian son consideradas como núcleo común en todas las moléculas de estudio presentes para el análisis de las contribuciones atómicas.

Las moléculas en el conjunto de datos se pueden clasificar en cuatro grupos basados en la sustitución de los grupos funcionales, el primer grupo consta de 23 moléculas (C1-4v en la Tabla 1a.). Los análogos sustituidos con los grupos toliilo y metoxi-fenil presentaron contribuciones positivas a la actividad con la presencia de color verde y amarillo como se puede ver en la Figura 9, una de las más activas moléculas C4v de esta categoría, presenta grupos meta-sustituidos, además de contar con la presencia de un grupo -OH que probablemente contribuye a su alta actividad, mientras que las regiones blancas se consideran como contribuyentes intermedias a la actividad de la molécula. (Tabla 1a.). Con excepción del compuesto C4j que posee una muy baja actividad lo que sugiere que la ortogonalidad entre los árenos y el sistema de anillos triazolopirimidina no era tolerado, por esta razón se puede observar en la Figura 10. Las contribuciones negativas en color rojo de la molécula C4j (Tabla1).

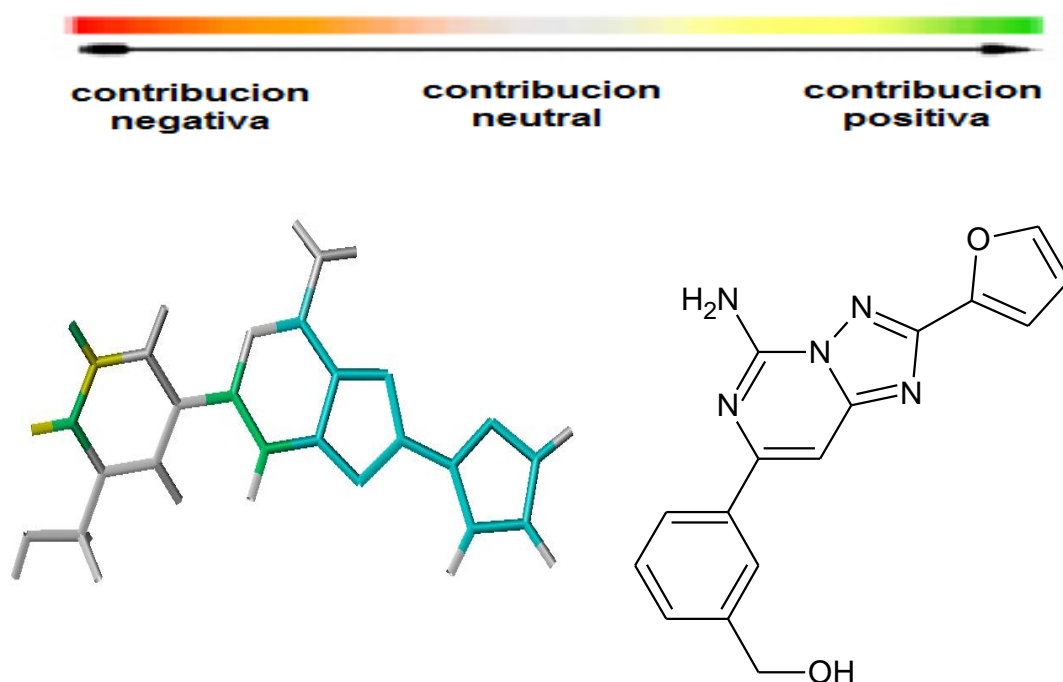


Figura 9. Contribución atómica individual (compuesto C4v de las moléculas de estudio) a la actividad.

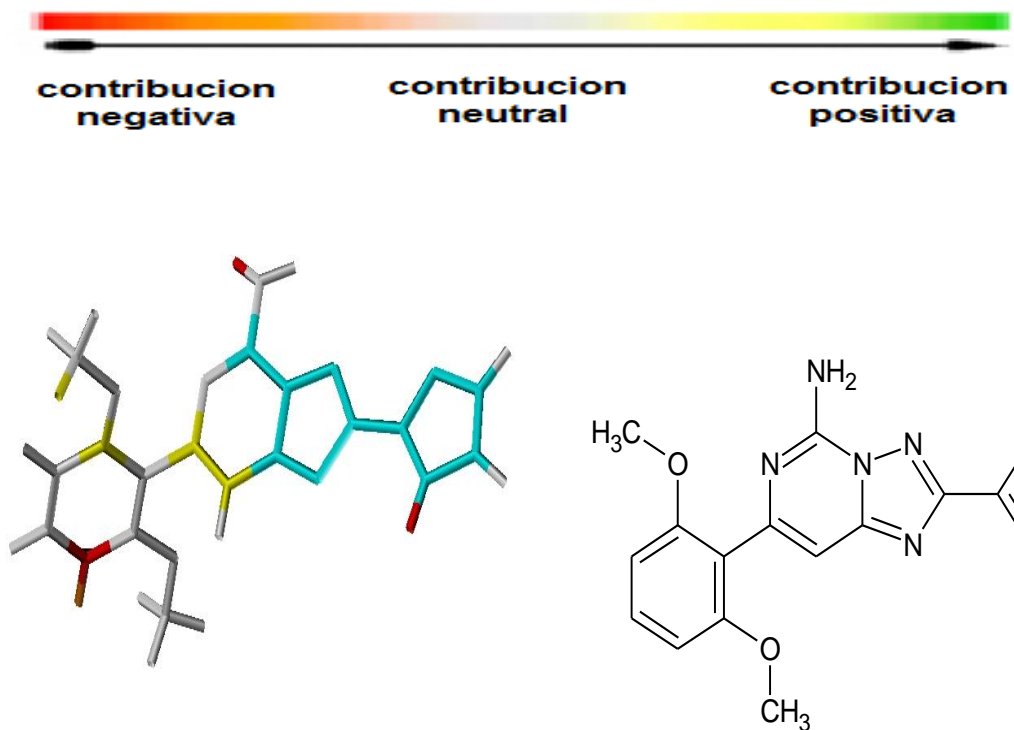


Figura 10. Contribución atómica individual (compuesto C4j de las moléculas de estudio) a la actividad

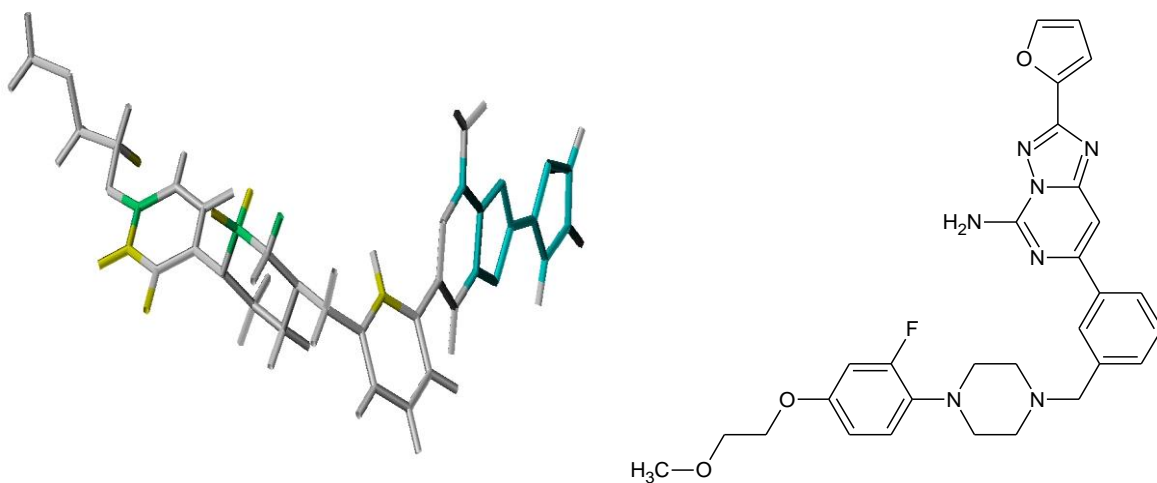


Figura 11. Contribución atómica individual (compuesto C8i de las moléculas de estudio) a la actividad.

El segundo grupo consta de 17 moléculas (C8a-q en la Tabla 1b) las sustituciones de grupos fenilos o bien 4- (fenil)-piperidina o derivados de la 4-(fenil)-piperizina en la posición meta fue bien tolerada, presentando así contribuciones positivas observadas con la presencia de color verde, la incorporación del grupo metoxetoxi parece particularmente beneficioso a las contribuciones favorables, cualquier molécula, que presente regiones coloreadas de blanco, verde o amarillo, indican contribuciones intermedias o contribuciones favorables a la actividad, ver Figura11, (molécula C8i –Tabla1b).

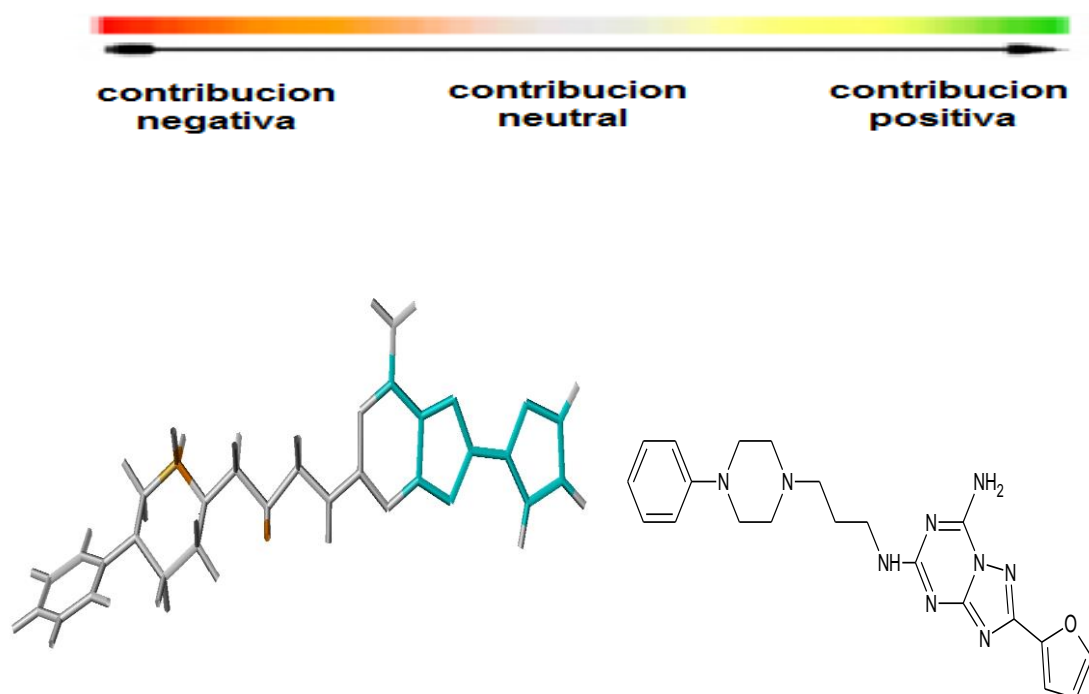


Figura 12. Contribución atómica individual (compuesto E8 de las moléculas de estudio) a la actividad.

La tercera clases de molécula consta de 12 moléculas (E7-E19 en las Tablas 1c-d) estos derivados de la piperazina con variaciones en la longitud del espaciado entre la fracción de piperazina, nos muestra que las contribuciones son pobres al momento de aumentar el espaciado entre la fracción de piperazina como se

muestra en la molécula E8, con contribuciones de color naranja (Tabla 1c, ver Figura 12) que presenta una baja actividad debido al aumento de espacios de moléculas de carbono. Por el contrario la molécula E14 (Tabla 1d) representada en la Figura 12 presenta contribuciones favorables a la actividad, debido a al corto espaciado de carbono entre la fracción de piperazina y a la metilación del grupo NH resultando así una molécula con una alta actividad biológica, representada con la presencia de color verde y amarillo. Tanto la molécula E8 como la E14 (Figura 12 y 13) presentan contribuciones intermedias (color blanco), presentando una mayor región de contribución intermedia a la actividad la molécula E8.

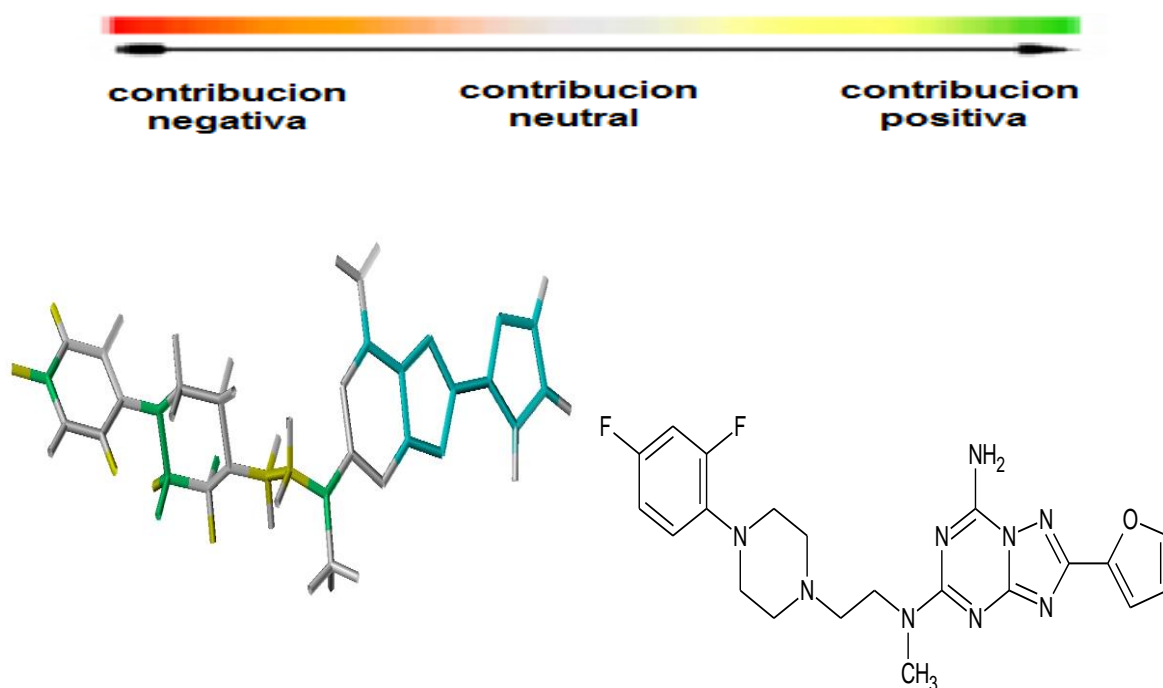


Figura 13. Contribución atómica individual (compuesto E14 de las moléculas de estudio) a la actividad.

El último grupo consta de 21 moléculas (H8-33 en la Tabla 1e-f) la presencia del grupo arilo o alquilo en la cadena lateral del acetileno es claramente beneficiosa en la actividad, como lo demuestra la más activa molécula H15 (Tabla 1e) de esta

categoría con una alta presencia de color verde y amarillo que nos indica una contribución favorable o positiva a la actividad y la coloración blanca una contribución intermedia a la actividad, ver Figura 14. A excepción de las moléculas H21 y H22 (Tabla 1e) que presentan contribuciones extremadamente pobres, es probable que cuando R es igual a p-(dimetilamino)-fenil, un grupo básico o el voluminoso 2-fluoro-3-(trifluorometil)-fenilo, contribuyen a la gran pérdida de la actividad. Por esta razón se observa (ver Figura 15.) coloraciones rojas y rojo-naranja en estas moléculas en particular.

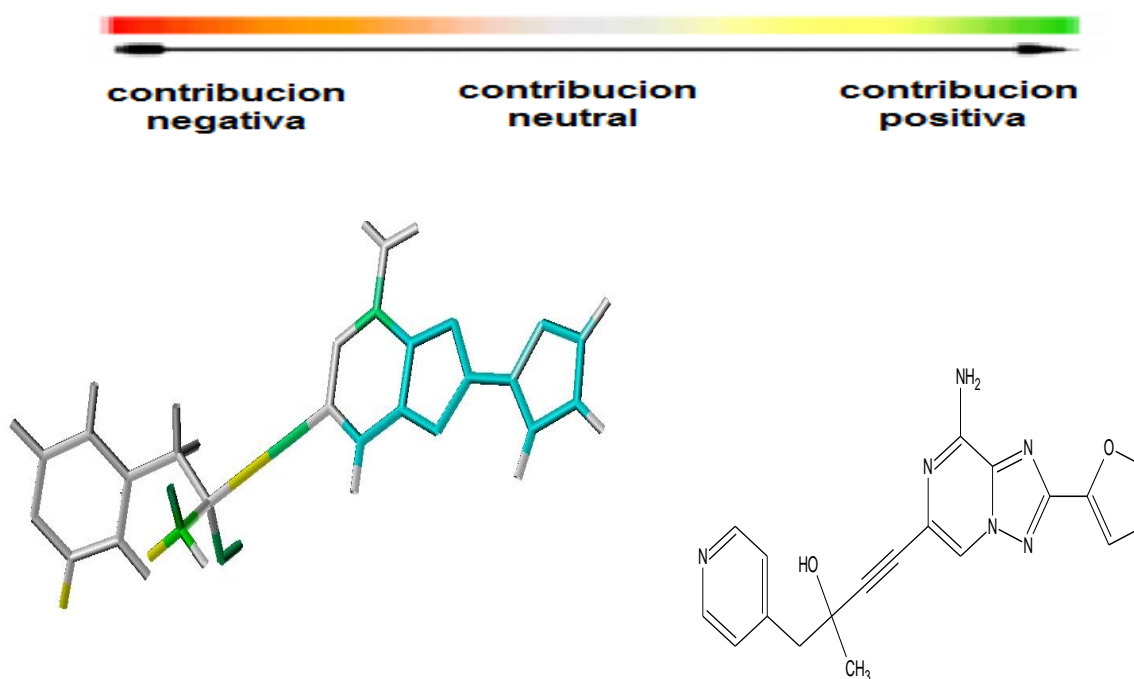


Figura 14. Contribución atómica individual (compuesto H15 de las moléculas de estudio) a la actividad.



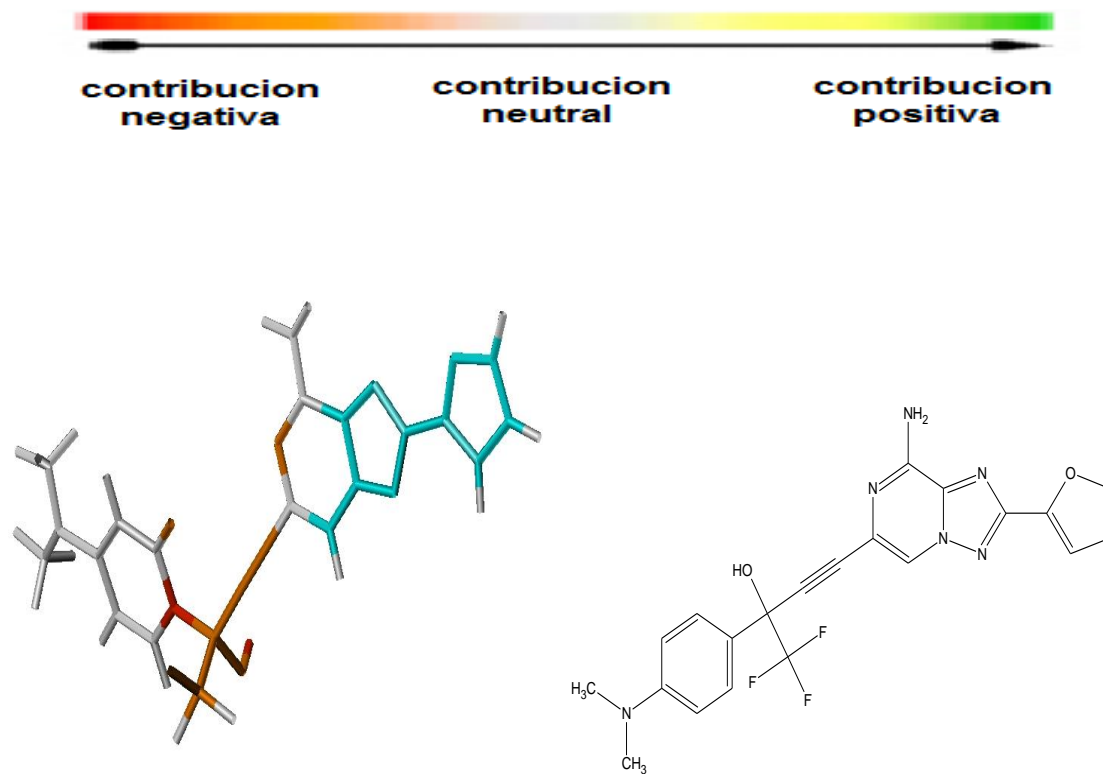


Figura15. Contribución atómica individual (compuesto H21 de las moléculas de estudio) a la actividad.

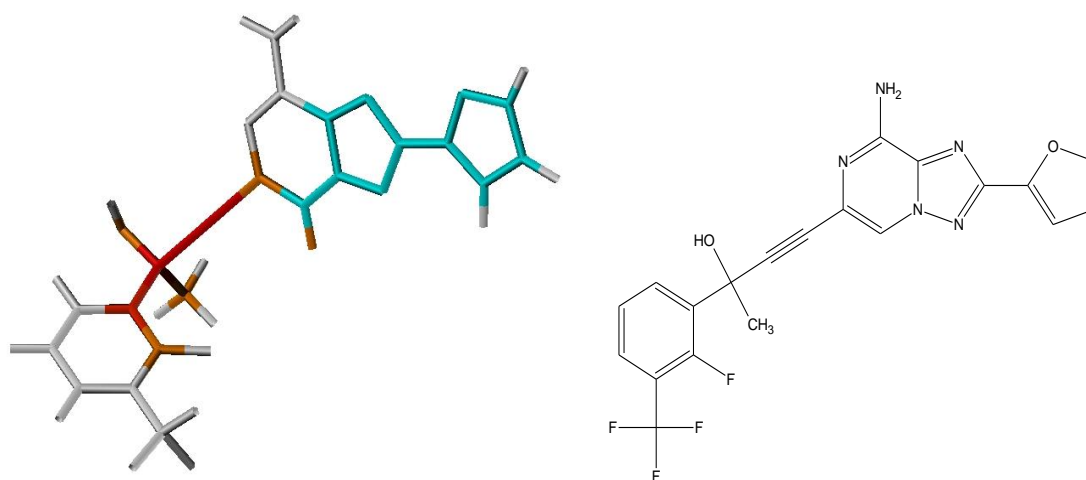


Figura16. Contribución atómica individual (compuestos H22 de las moléculas de estudio) a la actividad.

## CONCLUSIONES

A partir del estudio del holograma de la relación cuantitativa estructura–actividad (**HQSAR**), de derivados de la triazol-triazina, triazol-pirimidina y triazol-pirazina se puede concluir lo siguiente:

El análisis 2D-QSAR fue desarrollado para construir un modelo estadísticamente bueno con un importante poder de correlación para un conjunto de antagonistas del receptor de adenosina  $A_{2A}$ , el mejor modelo se obtuvo utilizando la combinación de parámetros de Átomos, Enlaces, Conectividad, Donador y aceptor (A/B/C/D) y con el tamaño de fragmento de 7-10, dio valores estadísticamente significativos de  $r^2$  y  $q^2$  de 0.947 y de 0.534, respectivamente.

No existe diferencia significativa entre los valores reales y los predichos por el modelo, esto debido a que los resultados obtenidos proporcionaron valores reproducibles para predecir las actividades de los compuestos en estudio.

Los resultados de este estudio demuestran que el **HQSAR**, usando el método de Partial Least Square (PLS), puede generar modelos confiables para la predicción de propiedades tales como la actividad biológica que en éste caso depende básicamente de los parámetros moleculares: tamaño del fragmento, longitud del holograma y tipo de fragmento, (H, DA, C, Ch, B, A.).

El HQSAR demostró contar con amplias aplicaciones en conjunto de datos grandes debido al corto tiempo de cálculo y potente capacidad ejecutiva. Muchos factores como el de la no exigencia de la estructura 3D, simple reproducción y los códigos visuales de color hacen a la técnica HQSAR poseer ventajas significativas en el estudio de la relación cuantitativa estructura-actividad.

Con el uso de los métodos de validación tanto interna como externa se demuestra nuevamente que el modelo obtenido es confiable y se observó que las

predicciones de las actividades biológicas de algunas moléculas usando este modelo son muy cercanas a los valores experimentales.

En general, se concluye que la información obtenida en nuestro análisis acerca de que fragmentos moleculares están directamente relacionados con la actividad biológica, a través de la codificación poli cromática, además de predecir las actividades de moléculas. Será útil en la orientación química para la obtención de nuevos y más activos antagonistas del receptor de adenosina  $A_{2A}$ .

## BIBLIOGRAFIA

1. Chi B. Vu, et al., Triamino derivatives of triazolotriazine and triazolopyrimidine as adenosine A2a receptor antagonists. *Bioorganic & Medicinal Chemistry Letters* 14 (2004) 4835–4838.
2. Gang Yao, et al., Synthesis of alkyne derivatives of a novel triazolopyrazine as A2A adenosine receptor antagonists. *Bioorganic & Medicinal Chemistry Letters* 15 (2005) 511–515.
3. Geoffrey B. Varty and Deen B. Tulshian, et al., 2-(2-Furanyl)-7-phenyl[1,2,4]triazolo[1,5-c]pyrimidin-5-amine analogs: Highly potent, orally active, adenosine A2A antagonists. Part 1. *Bioorganic & Medicinal Chemistry Letters* 15 (2005) 3670–3674.
4. Jesús Norelis Lorenzo, et al., Deterioro del aprendizaje visoespacial en la enfermedad de Parkinson. *Psicothema* 2009. Vol. 21, nº 1, pp. 21-26
5. Lilian María Vargas Barahona, Enfermedad de Parkinson y la Dopamina. *BUN Synapsis* Vol. 2 Abril - Junio 2007
6. Magda Osman, et al., Patients with Parkinson's disease learn to control complex systems via procedural as well as non-procedural learning. *Neuropsychologia*, Volume 46, Issue 9, July 2008, Pages 2355-2363.

7. Alex Iranzo. Parkinson's Disease and Sleepiness. *Sleep Medicine Clinics*, Volume 1, Issue 1, March 2006, Pages 127-137.
8. A.C. Williams and D.B. Ramsden. Nicotinamide homeostasis: A xenobiotic pathway that is key to development and degenerative diseases. *Medical Hypotheses*, Volume 65, Issue 2, 2005, Pages 353-362
9. Amanda L. Price. Explicit Category Learning in Parkinson's Disease: Deficits Related to Impaired Rule Generation and Selection Processes. *Neuropsychology*, Volume 20, Issue 2, March 2006, Pages 249-257.
10. Guy Sawle. Parkinson's disease. *Medicine*, Volume 32, Issue 11, 1 November 2004, Pages 93-96
11. Anders Björklund, et al., Neural transplantation for the treatment of Parkinson's disease. *The Lancet Neurology*, Volume 2, Issue 7, July 2003, Pages 437-445.
12. Takashi Kato, et al., Pathophysiological brain mapping in Parkinson's disease. *International Congress Series*, Volume 1232, April 2002, Pages 571-577.
13. Oscar Jiménez Leighton, Enfermedad De Parkinson Y Parkinsonismos. <http://www.med.ufro.cl/Recursos/neurologia/doc/c17.pdf>

14. Elisabetta Tronci, et al., Characterization of The Antiparkinsonian effects of the new Adenosine A<sub>2a</sub> receptor antagonist ST1535: Acute and subchronic studies in rats. *European Journal of Pharmacology* 566 (2007) 94–102 .
15. Correa, M., et al., The adenosine A<sub>2A</sub> antagonist KF17837 reverses the locomotor suppression and tremulous jaw movements induced by haloperidol in rats. Possible relevance to parkinsonism. *Behav. Brain Res.* 148, 47–54.
16. Bara-Jimenez, W., et al., Adenosine A<sub>2A</sub> receptor antagonist treatment of Parkinson's disease. *Neurology* 61, 293–296
17. G. Cuevas, F. Cortés, INTRODUCCIÓN A LA QUIMICA COMPUTACIONAL. 1<sup>ra</sup> Edición en español. Fondo de Cultura Económica. México (México) (2003). Cap. 5
18. Muegge, I.; Martin, Y. C.; Hajduk, P. J.; Fesik, S. W. *J. Med. Chem.* Evaluation of PMF Scoring in Docking Weak Ligands to the FK506 Binding Protein.; (Expedited Article); 1999;
19. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P., AM1: A New General Purpose Quantum Mechanical Molecular Model, *J. Am. Chem. Soc.* 107, 3902 (1985).
20. J. J. P. Stewart, Optimization of parameters for semiempirical methods I. Method, *J. Comput. Chem.* 10, 209, (1989).

21. Combet C, Jambon M, Deléage G, Geourjon C. Geno3D: automatic comparative molecular modelling of protein. *Bioinformatics*. 2002 Jan;18(1):213-4
22. Levine, N. *Química Cuántica 5<sup>ta</sup> edición*, Prentice Hall, Madrid (2001). Pag 638-639
23. Foresman, J. B. *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems* By David Young (Cytoclonal Pharmaceuticals Inc.). Wiley-Interscience: New York. 2001.
24. M. J. S. Dewar, E. G. Zoebish, E. F. Healy, J. J. P. Stewart, AM1: A new purpose quantum mechanical molecular model, *J. Am. Chem. Soc.* 107, (1985)
25. AM1, <http://en.wikipedia.org/wiki/AM1> (22/02/08)
26. Tutorial 1. What is a molecular descriptor, by V. Consonni, and R. Todeschini; Milano Chemometrics and QSAR Research Group. Website: <http://www.taletе.mi.it>. (19/01/2008)
27. Nirmalakhandan, N. Y Speece, R. E. (1988). Structure – Activity Relationships. *Environ. Sci. Technol.* **22** : 606 – 615.
28. C. (1969). A Quantitative Approach the Biochemical Structure – Activity Relationships. *Acc. Chem. Res.* **2**: 232 – 239.

29. Tiago L. Moda, Carlos A. Montanari y Adriano D. Andricopulo. (2007). Hologram QSAR model for the prediction of human oral bioavailability. 7738-7745.
30. David Lewis INC Tripos notas técnicas 1997. Manual HQSAR
31. Técnicas de regresión: Regresión lineal múltiple, ( 15/10/2007)  
[http://www.fisterra.com/mbe/investiga/regre\\_lineal\\_multi/regre\\_lineal\\_mu-lti.asp](http://www.fisterra.com/mbe/investiga/regre_lineal_multi/regre_lineal_mu-lti.asp)
32. Técnicas de regresión: Regresión lineal múltiple, (15/10/2007)  
[http://www.fisterra.com/mbe/investiga/regre\\_lineal\\_multi/regre\\_lineal\\_multi.a-sp](http://www.fisterra.com/mbe/investiga/regre_lineal_multi/regre_lineal_multi.a-sp)
33. Miller, J.N. y Miller, J.C. Estadística y Quimiometría para Química Analítica, Cuarta Edición, Editorial Prentice Hall, España. 2002.
34. Katritzky, A.R., y Gordeeva, E.V. (1993). *J. Chem. Inf. Comput. Sci.* **33** : 835 – 857.
35. [http://e-stadistica.bio.ucm.es/glosario/coef\\_corre.html](http://e-stadistica.bio.ucm.es/glosario/coef_corre.html) (27/02/2008)
36. J. C. Escalona, R. Carrasco, J. A. Padrón, Introducción al diseño de fármacos, Folleto para la docencia de la asignatura de farmacia, Universidad de Oriente, [www.fq.uh.cu/investig/lqct/imagenes2/disen.pdf](http://www.fq.uh.cu/investig/lqct/imagenes2/disen.pdf), (22/02/08)



