



**Universidad Michoacana de San
Nicolás de Hidalgo**



**Escuela de
Químico-Farmacobiología**

**“Análisis QSAR de difluoroquinolonas
asistido por algoritmos genéticos”**

Tesis profesional

Que para obtener el título de:
Químico-Farmacobiólogo

Presenta:
Otoniel Mojica Espíritu

Asesor de tesis:
D.C. Luis Chacón García

Morelia, Michoacán, Julio 2006

El presente trabajo de tesis fue desarrollado en el laboratorio de química orgánica del instituto de investigaciones Químico-Biológicas de la Universidad Michoacana de San Nicolás de Hidalgo, bajo la dirección del D.C. Luis Chacón García.

Con apoyo CONACYT (Proyecto 47907 “Síntesis de Inhibidores Potenciales de la Topoisomerasa”).

Con apoyo de la coordinación de investigación científica de la UMSNH (Proyecto 2.21 “Síntesis de Derivados de Fluoroquinolonas”).

Agradecimientos

Resulta imposible mencionar a todos aquello que de una u otra manera influenciaron el impulso que me a llevado a la culminación de esta etapa de mi vida, por ello quisiera antes que nada ofrecer un agradecimiento en general a todas las personas que han incursionado en mi vida. Aun así mencionare algunas de ellas:

Agradezco a mis padres que siempre me apoyaron en las buenas y en las malas, agradezco las enseñanzas que me proporcionaron ya que sin ellas esto no hubiese sido posible. Además les agradezco lo más importante para mi, gracias por la vida que me brindaron.

Gracias Hezart ya que has sido la inspiración y la fuerza que me ayuda a seguir adelante, pero sobre todo por que eres lo más hermoso que me pudo pasar en la vida, gracias hija.

Te agradezco Magdalena por compartir tantos momentos conmigo y ser un apoyo tan grande en mi vida, pero sobre todo por esa alegría inmensa de mi corazón “Hezart”.

Agradezco a mis hermanos por estar ahí en los momentos difíciles, por el apoyo que me han brindado y los hermosos momentos que hemos pasado juntos.

Al Dr. Luis Chacón García le agradezco por la oportunidad que me brindo y la confianza que deposito en mí, por el gran apoyo que me otorgó y sobre todo por su gran amistad.

Quisiera agradecer a todos mis amigos y amigas, a los cuales me gustaría mencionar pero tendría que agregar un libro anexo y aun si no habría espacio suficiente, así que les pido una disculpa y les agradezco por esa amistad tan especial que me brindaron y por el apoyo que recibí de ustedes, gracias.

Agradezco también a la Dra. Judit Araceli Aviña Verduzco, al Dr. Zurisaddai Hernández Gallegos y al M.C. Raúl Chavez Garibay, por ayudarme a culminar esta tesis con sus grandes aportaciones, gracias.

Índice

Índice.....	4
Introducción.....	5
Fundamento.....	8
1. Modelado molecular.....	9
2. Relación cuantitativa estructura-actividad (QSAR).....	12
3. Algoritmos genéticos.....	19
4. Quinolonas.....	25
Objetivos.....	29
Metodología.....	31
Discusión y resultados.....	33
1. Obtención de descriptores y Funciones de ajuste (Fitness functions).....	34
2. QSAR de difluoro quinolonas con actividad contra <i>Streptococcus pyogenes</i>	39
3. QSAR de difluoro quinolonas con actividad contra <i>Staphylococcus aureus</i>	49
4. QSAR de difluoro quinolonas con actividad contra <i>Escherichia coli</i>	54
5. QSAR de difluoro quinolonas con actividad contra <i>Pseudomona aeruginosa</i>	58
6. Predicción.....	62
Conclusiones.....	64
Referencias.....	66

Introducción

La afinidad de un fármaco por su receptor y el grado de actividad intrínseca que posee depende de su estructura química, y esta relación a menudo es muy precisa. Las modificaciones relativamente menores en la molécula pueden producir grandes cambios en las propiedades farmacológicas. En muchas ocasiones, el aprovechamiento de las relaciones entre la estructura y la actividad a permitido sintetizar agentes terapéuticos útiles.¹⁷

Por otro lado la necesidad de búsqueda de nuevos fármacos, es un proceso complejo e interdisciplinario que suele llevar 5 a 10 años de investigación y desarrollo, con un costo promedio de 231 millones de dólares por cada nuevo medicamento que sale al mercado. Lo más alarmante es que sólo una de cada 10 000 moléculas ensayadas pasa a la fase de desarrollo, una de cada 100 000 supera los ensayos clínicos y logra registrarse y sólo 3 de cada 10 nuevos medicamentos registrados recupera su inversión inicial.^{28,31}

En este sentido, el uso de ordenadores (computadoras) ha facilitado y disminuido costos en las diversas etapas de la investigación y desarrollo de nuevos fármacos. Así pues con el surgimiento de aplicaciones informáticas (tanto software como hardware) en el diseño de fármacos empleando métodos de predicción, se crea un filtro para obtener nuevos compuestos con alta probabilidad de presentar la actividad biológica deseada.³¹

Estas relaciones químicas pueden ser cuantificadas, diseñadas y predichas mediante técnicas de química computacional.³¹

En los años 60 Hansch y Fujita plantearon relaciones cuantitativas entre parámetros físico-químicos moleculares y actividad biológica mediante técnicas estadísticas (QSAR, de sus siglas en inglés Quantitative Structure Activity Relationship).⁹ En sus orígenes resultaba suficiente contar con herramientas matemáticas, sin embargo, el ordenador comenzó a ser necesario para resolver estas relaciones pues los datos dejaron de limitarse a lipofílicos, y otras propiedades fisicoquímicas medibles en el laboratorio, para enriquecerse con propiedades cuantificables mediante programas complejos de cómputo.^{9,28,31}

Los denominados métodos QSAR han agrupado históricamente a todas las técnicas que han intentado establecer modelos empíricos de comportamiento de las familias de compuestos biológicamente activos, con el fin de obtener, de la manera más eficaz posible, resultados óptimos de actividad, a partir de los datos de comportamiento de un número limitado de productos.⁹

Una técnica es precisamente el uso de inteligencia artificial, que agrupa Redes Neuronales y Algoritmos Genéticos (AGs), estos últimos son también conocidos como métodos evolutivos que se basan en la evolución en la naturaleza como medio para lograr una especie con características de sobrevivencia. Un proceso lento, con elevado costo pero altamente eficiente.²⁸

Los AGs son métodos adaptativos que pueden usarse para resolver problemas de búsqueda y optimización. Están basados en el proceso genético de los organismos vivos. A lo largo de las generaciones, las poblaciones evolucionan en la naturaleza de acuerdo con los principios de la selección natural y la supervivencia de los más fuertes, postulados por Darwin. Por imitación de este proceso, los Algoritmos Genéticos son capaces de ir creando soluciones para problemas del mundo real. La evolución de dichas soluciones hacia valores óptimos del problema depende, en buena medida, de una adecuada codificación de las mismas.^{24,25}

La presente tesis se centra en la utilización de métodos computacionales para un estudio QSAR de una serie de difluoroquinolonas, tomada de una base de datos descrita en la literatura utilizando AGs.⁷

Fundamento

Modelado molecular

El modelado molecular es una técnica relativamente nueva de estudio de la relación entre la estructura molecular y las propiedades químicas macroscópicas de la materia. La base del modelado molecular es el uso de métodos computacionales para simular, explicar o predecir la estructura tridimensional y las propiedades fisicoquímicas de las moléculas, con la ayuda esencial de métodos gráficos. Estas técnicas caben dentro de lo que recibe el nombre de Química Computacional.²²

Los *modelos cuantitativos* son aquellos que permiten asignar valores numéricos a ciertas magnitudes, asociadas a propiedades del sistema, y obtener como consecuencia del modelo una imagen del comportamiento de la sustancia o del sistema (tales modelos son, por ejemplo, las leyes que cubren aspectos parciales del comportamiento de las sustancias químicas). Un conjunto ordenado de modelos parciales constituye usualmente una teoría (o gran modelo) como lo son, por ejemplo, la termodinámica clásica o la cinética formal.^{22,31}

El nombre de *modelado molecular* se aplica a un conjunto de métodos que pretenden predecir las propiedades y el comportamiento de los sistemas químicos en forma numérica apoyándose en entornos gráficos para la manipulación de moléculas, y es mayormente indistinguible de lo que se conoce como *química computacional*, si bien este nombre implica más bien el conjunto de estudios que hace hincapié en lo numérico más que en lo gráfico. El objetivo fundamental de la química computacional (o, en general, el modelado molecular) es conseguir explicar los fenómenos químicos observados en el laboratorio y predecir otros fenómenos que aún no han sido observados, todo ello recurriendo a técnicas numéricas que permitan resolver las ecuaciones planteadas en los modelos surgidos de la aplicación de un determinado sistema de leyes físicas fundamentales a los sistemas moleculares.²²

La *química computacional* está compuesta de dos grandes áreas basadas en principios físicos distintos. Por una parte, se encuentra la *mecánica molecular* que se basa en la mecánica clásica, considerando a los átomos como partículas puntuales (dotadas de masa y carga) interaccionando unas con otras mediante enlaces que se asemejan a resortes, se

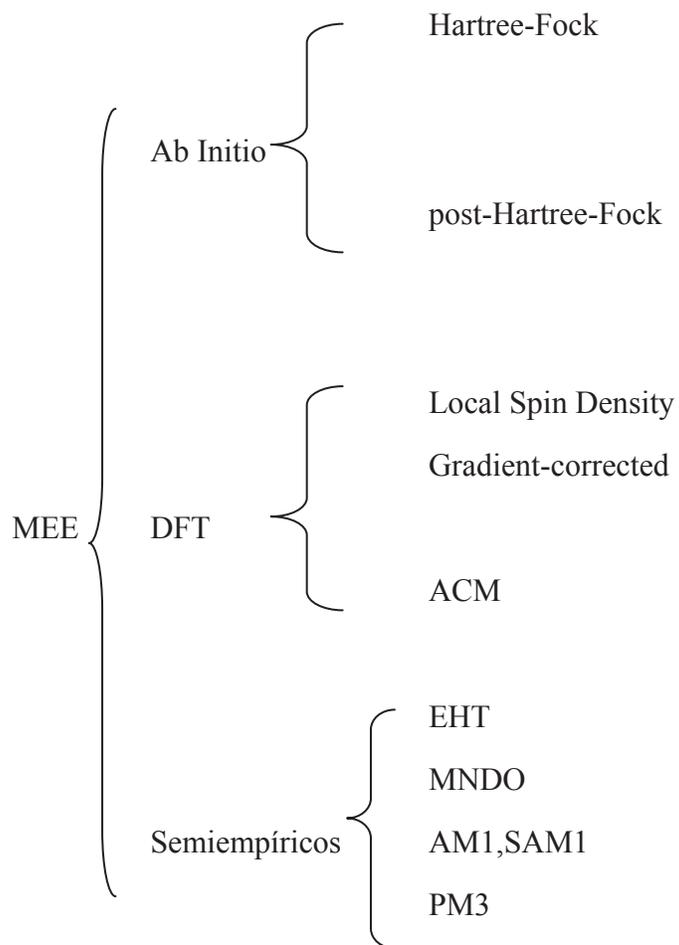
consideran como partículas puntuales con carga y masa fijas e invariables, interaccionando según la ley de Coulomb. Por otra parte, están los *métodos basados en estructura electrónica* que se utilizan en la aplicación de la Mecánica Cuántica a los sistemas atómicos y moleculares.

Tanto la Mecánica Molecular (MM) como los Métodos de Estructura Electrónica (MEE) permiten calcular un conjunto básico de propiedades, cuyos exponentes más destacados son la energía de un determinado arreglo de núcleos (conformación de una molécula), la geometría óptima de un sistema molecular (es decir, el arreglo geométrico de los núcleos considerados que brinde la energía más baja con el método de cálculo empleado) y otras propiedades moleculares (no todas ellas para todos los métodos) como pueden ser, por ejemplo, las frecuencias vibracionales y el momento dipolar.

Los distintos métodos dentro de la MM comparten el hecho general de basarse en las mismas leyes de la física clásica y se diferencian en (a) el tipo de ecuaciones que definen como varía la energía potencial de una molécula con la geometría de sus átomos componentes; (b) el tipo de los átomos, entendiéndose por ello las características que definen a un átomo de número atómico dado *en el entorno molecular en que se encuentre*; y/o (c) en el conjunto de parámetros que ajustan las ecuaciones aplicadas a los átomos definidos por su tipo, a los valores experimentales (o calculados por otros métodos) que se emplean para la parametrización. Estos tres componentes definen lo que se conoce con el nombre de campo de fuerza y definen completamente un método de MM.²²

Los MEE, como se mencionó anteriormente, emplean la mecánica cuántica en lugar de la mecánica clásica, para estudiar el comportamiento de núcleos y electrones (y no sólo núcleos como en MM). Los MEE se dividen en tres grupos principales (ab initio, funcionales de la densidad y semiempíricos) en función que no usen más que las constantes atómicas fundamentales o empleen datos experimentales para parametrizar parte de los cálculos. Todos los MEE se basan en el planteamiento y solución aproximada de la Ecuación de Schrödinger.

Los MEE son aplicables en la región de 10^3 átomos (para los métodos semiempíricos) a 10^1 (para los métodos ab initio mas sofisticados). Una clasificación somera de los métodos de cálculo mas frecuentemente empleados se muestra en el siguiente Esquema 1:



Esquema No. 1.- Clasificación de los métodos de cálculo mas frecuentemente empleados.

Relación cuantitativa estructura-actividad (QSAR)

El QSAR es un método que permite correlacionar la estructura de una serie de moléculas con la actividad biológica que presentan de manera cuantitativa.

La primera vez que se habla de un modelo "aditivo" de relación estructura actividad, tal como hoy lo entendemos, fue 1956 en un artículo de Bruice, Karasch y Winzler,⁹ donde los autores describían la utilización de un modelo *novo*. Pocos años después, Free y Wilson (1964) publicaron esencialmente el mismo método, sin que ellos ni los editores de la revista fueran capaces de detectar el trabajo anterior, y así el método ha sido conocido desde entonces como de Free-Wilson.^{9,16,28}

De manera simplificada, la técnica esta basada en el uso de series de compuestos derivados de una estructura en común, en los que se observan las alteraciones producidas en su actividad biológica en función de la presencia de sustituyentes diversos. A partir de estos datos, es posible calcular coeficientes aditivos de *novo* para cada sustituyente, con lo que es posible diseñar la mejor combinación de sustituyentes para alcanzar el óptimo de actividad de un fármaco.

Simultáneamente con el trabajo de Free y Wilson, Corwin Hansch en conjunto con Toshio Fujita pusieron a punto el denominado modelo paramétrico (también llamado modelo de Hansch) en el que se correlaciona la actividad biológica de series de productos con sus propiedades fisicoquímicas. Ambos métodos constituyen la base de la optimización de actividades con una serie de prototipos de serie, tal como se utiliza en la actualidad. El modelo extratermodinámico de Hansch brinda una explicación matemática del fenómeno recogida en la siguiente ecuación^{9,16}:

$$\text{Ecuación No. 1.- } \log A = f_h(X_h) + f_e(X_e) + f_s(X_s) + \text{cte}$$

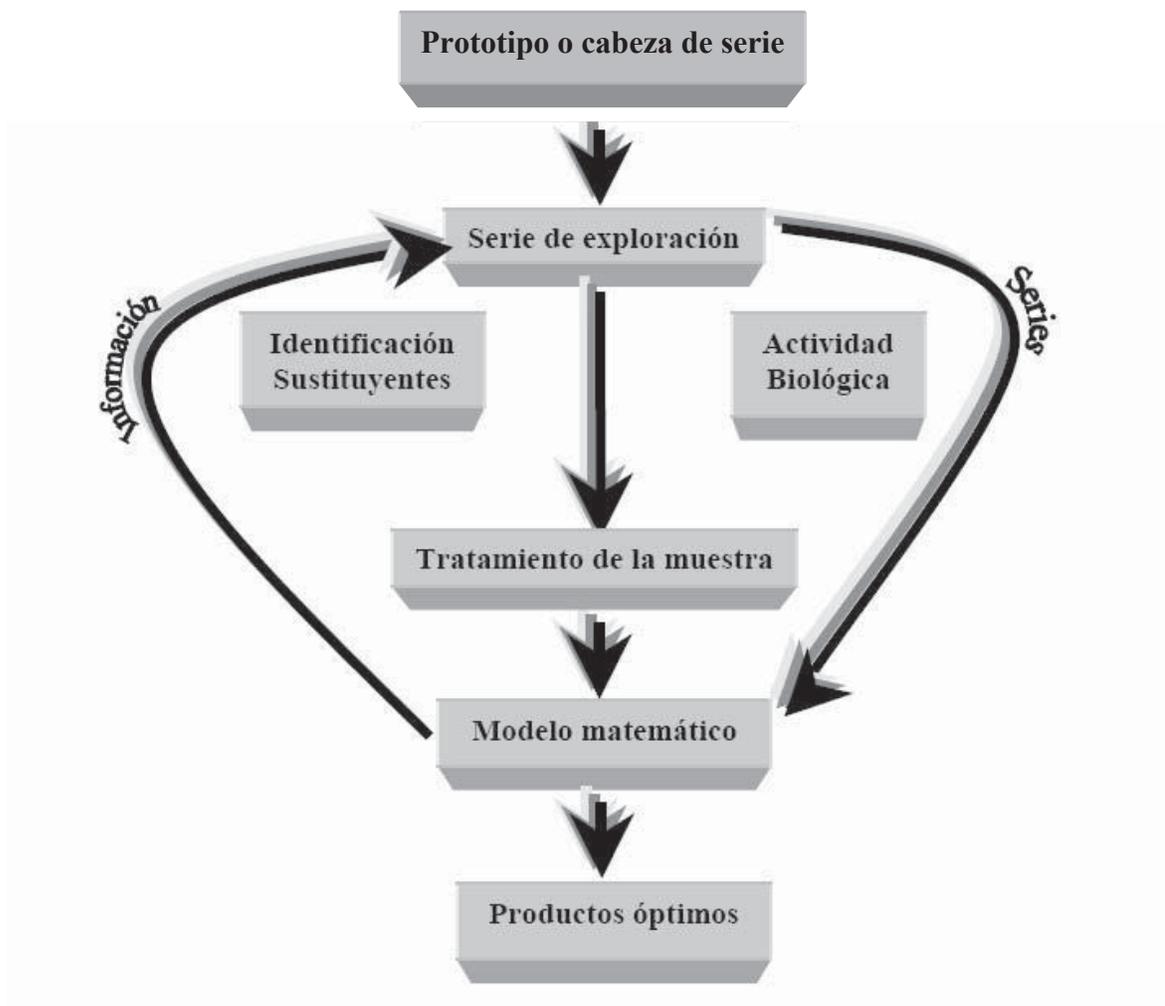
Donde A es la actividad y $f_h(X_h)$, $f_e(X_e)$, $f_s(X_s)$ son funciones de índices o parámetros hidrofóbicos, electrónicos o estéricos respectivamente. El término extratermodinámico

proviene que las relaciones se que describen en términos termodinámicos, aunque no se deducen de sus leyes. Considerando que la interacción de un fármaco con su blanco molecular se encuentran en equilibrio termodinámico y que el desplazamiento de este determina el grado de actividad como consecuencia de su interacción.

Los principios básicos de la metodología extratermodinámica pueden formalizarse en una serie de puntos, la mayoría de los cuales se extraen de la ecuación fundamental, y es lo que se conoce como el paradigma de Hansch,^{9,16,28} que establece que:

1. La actividad biológica es función de la estructura del fármaco.
2. La estructura del fármaco implica ciertas propiedades globales como la hidrofobicidad, carga neta, solubilidad etc. y ciertas propiedades locales como la distribución de la hidrofobicidad, carga y volumen en determinadas posiciones de la molécula.
3. Estas propiedades globales y locales pueden ser cuantificadas mediante ciertos parámetros.
4. Siempre existe una función que relaciona los cambios de actividad biológica con los cambios en las propiedades globales y locales, aunque puede ser que no sea sencilla ni evidente.

Las funciones estructura-actividad que propone el método extratermodinámico, constituye un método mediante el cual se pueden buscar los productos más activos entre un conjunto de candidatos. La metodología de investigación en el QSAR sigue, independientemente del modelo que se utilice, pasos comunes que se hallan sintetizados en el Esquema 2. El punto de partida implica en todos los casos la existencia de un cabeza de serie.



Esquema No. 2.- Los pasos de la metodología del análisis QSAR.

Cuando se dispone de un prototipo, es preciso diseñar una serie de exploración, que está constituida por un conjunto de análogos del prototipo, que permiten el establecimiento de las primeras relaciones estructura-actividad. Los miembros de la serie de exploración se sintetizan y evalúan en cuanto a actividad biológica respecta. Los miembros de la serie de exploración están constituidos por un núcleo común, y sustituyentes o fragmentos variables, que son característicos de cada producto de la serie. Dichos sustituyentes variables han de ser representados con descriptores, que serán utilizados como variables independientes (X_i) en el modelo. Los valores de actividad biológica (A) (expresados regularmente en forma logarítmica) se utilizan como variable dependiente.^{9,16,28}

Es de particular importancia la calidad de la serie de exploración a emplear, pues la calidad de las predicciones van a estar condicionadas en gran medida al diseño óptimo de esta serie, para así tratar de evitar las fallas en la predicción.

Elección de la serie

La calidad de una serie de exploración esta definida principalmente por dos aspectos:

1. *Disimilaridad de la serie*. Los productos de la serie deben ser, respecto a los parámetros elegidos, diferentes entre sí. No tiene sentido intentar estudiar cómo influye una característica sobre la actividad si sólo se estudian productos similares respecto a esta característica. Además, puesto que la serie debe ser una muestra del espacio experimental, el rango de variación de las características dentro de dicha serie debe ser equiparable al que existe en el espacio experimental.^{9,13,16,28,29}

2. *Ortogonalidad de la serie*. Los productos deben escogerse de tal forma que la variación en las características se produzca de manera independiente. Si la variación de una propiedad X_1 viene siempre acompañada de la variación en otra propiedad X_2 , no es posible conocer si el cambio de actividad de unos a otros productos se debe al cambio en la propiedad X_1 ó en la propiedad X_2 . En estas situaciones se dice que las propiedades X_1 y X_2 se encuentran *correlacionadas* en la serie. Una serie en la cual las correlaciones entre los parámetros son mínimas se dice que la serie es *ortogonal*. Una vez definida y evaluada la actividad biológica de la serie de experimentación se procede al tratamiento de la muestra de la cual se obtiene un modelo matemático que permitirá predecir uno o varios compuestos a los que les serán evaluadas sus actividades biológicas e incorporados, en caso de no presentar una actividad óptima, a la serie de exploración. Este proceso se repetirá el número de veces necesarias hasta lograr la eficacia esperada para el fármaco a obtener.^{9,13,16,28,29}

Tipos de estudios QSAR.

En la actualidad existen diferentes técnicas por las que se puede desarrollar un estudio QSAR en dependencia del tipo de tratamiento matemático que se le da a la muestra.

El método tradicional incluye el tratamiento estadístico de los datos por métodos multivariados, que incluye el análisis de regresión, en el que se valora la actividad biológica como variable dependiente del conjunto de descriptores moleculares que constituyen las variables independientes. Como resultado de esto se obtienen modelos que describen la actividad biológica como una determinada función matemática de los descriptores moleculares, bien sean estos estructurales o fisicoquímicos.^{16,28}

El método más familiar es la regresión, en la que se obtendrá la ecuación de una recta, un plano, o de un hiperplano, según sea el número de variables independientes incluida en la expresión. El análisis de regresión múltiple, es el más utilizado dentro del QSAR tradicional. En él, una vez establecidos los conjuntos de valores de las variables independientes X_i y la actividad biológica A , se obtiene un modelo en forma de ecuación de una recta (ecuación 1), la cual describe la dependencia de la actividad (A) en función del conjunto de descriptores (X_i) así como la magnitud de las contribuciones de cada uno de ellos.^{9,16,28}

Ecuación 2

$$\mathbf{A}=\mathbf{f}(\mathbf{X}_i) + \text{cte.}$$

El modelo así obtenido se analiza en función de su calidad estadística, para poder evaluar su capacidad de predicción. Cuanto mayor calidad estadística tenga el modelo, más confiables y exactas serán las predicciones a realizar, siempre y cuando se encuentran en el rango o cercanos a los intervalos de las propiedades incluidas en el modelo.²⁸ La calidad estadística de un modelo se evalúa por diferentes estadísticas. Los más comúnmente aceptados son el coeficiente de regresión r , el valor F de Fischer y la desviación estándar s . Cuanto más tienda a la unidad el valor de r , mejor ajuste tendrán los datos al modelo. El valor F correlaciona la varianza explicada (r^2) por el número de grados de libertad, con la

varianza no explicada ($1 - r^2$) por el número de variables del modelo. Cuanto más alto es el porcentaje de varianza explicada por el modelo mayor será el valor de **F**, mientras que la existencia de variables con baja contribución a la explicación de la varianza, tenderán a disminuir dicho valor, que tiende a infinito en los mejores modelos. La desviación estándar **s** depende de la varianza no explicada y de los grados de libertad del modelo y es una medida de cuanto se alejan los valores predichos por el modelo de la línea, plano o hiperplano. La tendencia a cero de este valor pudiera presuponer mayor calidad en la predicción. Un valor de **s** es válido cuando está en el mismo orden de magnitud del error experimental de las mediciones de la variable dependiente (la actividad biológica).^{9,16,28}

Consideraciones adicionales para el establecimiento de una ecuación con calidad se tienen en función del número de variables independientes a incluir en el modelo. Estas variables, deben estar limitadas por el número de compuestos que forman parte de la serie de exploración, de forma tal que idealmente, la relación existente entre ellos sea de una variable por cada 5 compuestos en la serie de exploración. La utilización excesiva de estas variables en una ecuación de regresión puede conducir a reproducciones de la muestra de entrenamiento eliminándole el valor predictivo al modelo. Es necesario señalar además que, mientras mayor número de variables que se incluyan en la ecuación, mayor será el número de parámetros a variar en el compuesto a obtener, hecho que se puede tornar difícil en la práctica sintética.^{9,16,28}

Validación de modelos.

El resultado final de un análisis QSAR será un modelo matemático. Para estar completamente seguros de que el modelo obtenido no resulta de un ordenamiento al azar de los datos es conveniente 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 varían las actividades. 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 actividad y a posibles efectos no incluidos en él (parámetros no considerados). Mientras mayor sea la diferencia entre la

varianza explicada sobre la no explicada mayor será la utilidad del modelo. Esto puede contrastarse con un análisis de varianza y un test de F; cuando el modelo no es satisfactorio se obtiene la *no significancia* en el test.^{16,18,19,28}

Otro problema es que el modelo no sea predictivo, es decir que la función propuesta no se cumpla para productos no incluidos en la serie de exploración. Esta problemática suele solucionarse realizando una validación cruzada (crossvalidation) con el procedimiento conocido como *leave-one-out* al modelo, generando los valores promedios de los coeficientes de correlación (r^2 ó r_{cross}) y de las desviaciones estándar obtenidos (S_{press}). Estos valores brindan una idea de la estabilidad estadística del modelo. Otra forma de medir o evaluar la calidad de la ecuación obtenida consiste en determinar la robustez (q_2) del mismo, y el mismo se calcula a través de la ecuación siguiente:

Ecuación 3
$$q_2 = (\text{SD-PRESS})/\text{SD}$$

Donde SD es la sumatoria al cuadrado de las desviaciones estándares de los residuales y PRESS es la sumatoria al cuadrado de los residuales. Este valor de q_2 derivado de la validación cruzada no es más que la habilidad del modelo para realizar la predicción. En un modelo donde r_{cross} no disminuya en más de un 20% con respecto al coeficiente de correlación, puede ser considerado como un modelo predictivo, lo mismo ocurre cuando se obtiene un valor de $q_2 > 0.5$.^{16,18,19,28}

Para que el estudio de QSAR se concluya con éxito debe obtenerse un modelo significativo y predictivo. Aún así, la validez del modelo estará siempre limitada al rango de los parámetros explorados por la serie de exploración, fuera del cual nunca debe considerarse válido. De hecho, esta es la principal limitación a la hora de buscar el óptimo de actividad predicho por un modelo.^{16,18,19,28}

Cuando el modelo contiene sólo términos lineales, basta con buscar sustituciones que maximicen los términos con coeficientes positivos y minimicen los términos con coeficientes negativos. Debe tenerse en cuenta, que si se rebasan los valores máximos y

mínimos de la serie de exploración, los resultados a obtener en los nuevos compuestos no tienen porqué responder al modelo, pues esos valores no fueron explorados y por tanto incluidos en el modelo. Un ejemplo de esto puede ser un modelo en el cual, con un incremento del coeficiente de partición se aumente la actividad; esto no implica que el con un incremento hasta el infinito de este coeficiente se aumentará ilimitadamente la actividad pues, a valores X de este parámetro puede resultar una incidencia negativa con respecto a la propia actividad, pues al cambiar lógicamente la distribución en el organismo del fármaco este ya no se encontrará en el compartimiento adecuado o sufrirá cambios en el metabolismo. Esto evidentemente, no pudo ser recogido por el modelo, pues en él este extremo de valores no fue considerado.^{16,18,19,28}

ALGORITMOS GENETICOS

La resolución de un estudio QSAR como fue hecho por Hansch en su momento requería solamente de herramientas matemáticas básicas, pues los parámetros utilizados se limitaban a estéricos, electrónicos y el coeficiente de partición. En la práctica es difícil obtener un buen modelo con solamente esos descriptores, además de que pueden ser otras variables las que definan la actividad y se encuentren fuera de ese grupo. Así, actualmente se puede contar, gracias a la quimioinformática, con cientos de descriptores moleculares, que llevarían a hacer cientos de regresiones para obtener un modelo adecuado. Es por esta razón que se ha acudido a estrategias computacionales como la Inteligencia Artificial (IA). Esta utiliza como herramientas para este tipo de estudios a las Redes Neuronales (RN) y Computación evolutiva. El principio de las RN se basa en el mecanismo de aprendizaje por las neuronas que aprenden a pensar; de la misma manera, al software especializado se le enseña a pensar y resolver un problema. Por otra parte los AGs se fundamentan en los mecanismos de evolución que han llevado a subsistir a las mejores especies de una población.^{21,24,25}

Los AGs nacieron con un investigador de la Universidad de Michigan llamado John Holland quien era consciente de la importancia de la selección natural y a fines de los 60s

desarrolló una técnica que permitió incorporarla a un programa de cómputo. Su objetivo era lograr que las computadoras aprendieran por sí mismas. ^{21,24,25}

Una definición bastante completa de un algoritmo genético es la propuesta por John Koza:

"Es un algoritmo matemático altamente paralelo que transforma un conjunto de objetos matemáticos individuales con respecto al tiempo usando operaciones modeladas de acuerdo al principio Darwiniano de reproducción y supervivencia del más apto, y tras haberse presentado de forma natural una serie de operaciones genéticas de entre las que destaca la recombinación sexual. Cada uno de estos objetos matemáticos suele ser una cadena de caracteres (letras o números) de longitud fija que se ajusta al modelo de las cadenas de cromosomas, y se les asocia con una cierta función matemática que refleja su aptitud. " ²⁴

Un algoritmo genético consiste en una función matemática o una rutina de software que toma como entradas a los ejemplares y retorna como salidas cuales de ellos deben generar descendencia para la nueva generación. ^{21,24,25}

Versiones más complejas de algoritmos genéticos generan un ciclo interactivo que directamente toma a la especie (el total de los ejemplares) y crea una nueva generación que reemplaza a la antigua una cantidad de veces determinada por su propio diseño. Una de sus características principales es la de ir perfeccionando su propia heurística en el proceso de ejecución, por lo que no requiere largos períodos de entrenamiento especializado por parte del ser humano, principal defecto de otros métodos para solucionar problemas, como los Sistemas Expertos. ^{21,24,25}

En la naturaleza los individuos de una población compiten entre sí en la búsqueda de recursos tales como comida, agua y refugio. Incluso los miembros de una misma especie compiten a menudo en la búsqueda de un compañero. Aquellos individuos que tienen más éxito en sobrevivir y en atraer compañeros tienen mayor probabilidad de generar un gran número de descendientes. Por el contrario individuos poco dotados producirán un menor número de descendientes. Esto significa que los genes de los individuos mejor adaptados se propagarán en sucesivas generaciones hacia un número de individuos creciente. La

combinación de buenas características provenientes de diferentes ancestros, puede a veces producir descendientes "superindividuos", cuya adaptación es mucho mayor que la de cualquiera de sus ancestros. De esta manera, las especies evolucionan logrando unas características cada vez mejor adaptadas al entorno en el que viven.²⁴

Los Algoritmos Genéticos usan una analogía directa con el comportamiento natural. Trabajan con una población de individuos, cada uno de los cuales representa una solución factible a un problema dado. A cada individuo se le asigna un valor ó puntuación, relacionado con la bondad de dicha solución. En la naturaleza esto equivaldría al grado de efectividad de un organismo para competir por determinados recursos. Cuanto mayor sea la adaptación de un individuo al problema, mayor será la probabilidad de que el mismo sea seleccionado para reproducirse, cruzando su material genético con otro individuo seleccionado de igual forma. Este cruce producirá nuevos individuos, descendientes de los anteriores, los cuales comparten algunas de las características de sus padres. Cuanto menor sea la adaptación de un individuo, menor será la probabilidad de que dicho individuo sea seleccionado para la reproducción, y por tanto de que su material genético se propague en sucesivas generaciones.^{21,24,25}

Ventajas de los AGs

- No necesitan conocimientos específicos sobre el problema que intentan resolver.
- Operan de forma simultánea con varias soluciones, en vez de trabajar de forma secuencial como las técnicas tradicionales.
- Cuando se usan para problemas de optimización maximizar una función objetivo- resultan menos afectados por los máximos locales (falsas soluciones) que las técnicas tradicionales.
- Resulta sumamente fácil ejecutarlos en las modernas arquitecturas masivamente paralelas.
- Usan operadores probabilísticos, en vez de los típicos operadores determinísticos de las otras técnicas.

Desventajas de los AGs

- Pueden tardar mucho en converger, o no converger en absoluto, dependiendo en cierta medida de los parámetros que se utilicen tamaño de la población, número de generaciones, etc.-.
- Pueden converger prematuramente debido a una serie de problemas de diversa índole.

Limitaciones

El poder de los Algoritmos Genéticos proviene del hecho de que se trata de una técnica robusta, y pueden tratar con éxito una gran variedad de problemas provenientes de diferentes áreas, incluyendo aquellos en los que otros métodos encuentran dificultades. Si bien no se garantiza que el Algoritmo Genético encuentre la solución óptima del problema, existe evidencia empírica de que se encuentran soluciones de un nivel aceptable, en un tiempo competitivo con el resto de algoritmos de optimización combinatoria. En el caso de que existan técnicas especializadas para resolver un determinado problema, lo más probable es que superen al Algoritmo Genético, tanto en rapidez como en eficacia. El gran campo de aplicación de los Algoritmos Genéticos se relaciona con aquellos problemas para los cuales no existen técnicas especializadas. Incluso en el caso en que dichas técnicas existan, y funcionen bien, pueden efectuarse mejoras de las mismas hibridándolas con los Algoritmos Genéticos.²⁴

El Algoritmo Genético Simple

El Algoritmo Genético Simple, también denominado Canónico, necesita una codificación o representación del problema, que resulte adecuada al mismo. Además se requiere una función de ajuste ó adaptación al problema, la cual asigna un número real a cada posible solución codificada. Durante la ejecución del algoritmo, los padres deben ser seleccionados para la reproducción, a continuación dichos padres seleccionados se cruzarán generando dos hijos, sobre cada uno de los cuales actuará un operador de mutación. El resultado de la combinación de las anteriores funciones será un conjunto de individuos

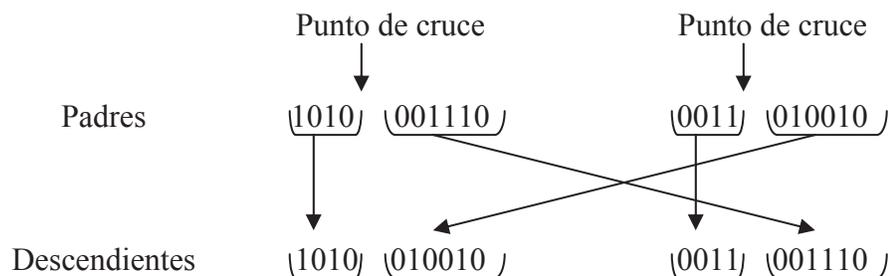
(posibles soluciones al problema), los cuales en la evolución del Algoritmo Genético formarán parte de la siguiente población.²⁴

Durante la fase reproductiva se seleccionan los individuos de la población para cruzarse y producir descendientes, que constituirán, una vez mutados, la siguiente generación de individuos. La selección de padres se efectúa al azar usando un procedimiento que favorezca a los individuos mejor adaptados, ya que a cada individuo se le asigna una probabilidad de ser seleccionado que es proporcional a su función de adaptación. Este procedimiento se dice que está basado en la ruleta sesgada. Según dicho esquema, los individuos bien adaptados se escogerán probablemente varias veces por generación, mientras que, los pobremente adaptados al problema, no se escogerán más que de vez en cuando.

Una vez seleccionados dos padres, sus cromosomas se combinan, utilizando habitualmente los operadores de cruce y mutación. Las formas básicas de dichos operadores se describen a continuación.

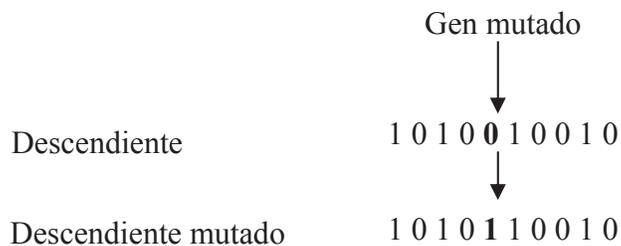
El *operador de cruce*, toma dos padres seleccionados y corta sus ristas de cromosomas en una posición escogida al azar, para producir dos subristras iniciales y dos subristras finales. Después se intercambian las subristras finales, produciéndose dos nuevos cromosomas completos (véase el Esquema 3). Ambos descendientes heredan genes de cada uno de los padres. Este operador se conoce como operador de cruce basado en un punto. Habitualmente el operador de cruce no se aplica a todos los pares de individuos que han sido seleccionados para emparejarse, sino que se aplica de manera aleatoria, normalmente con una probabilidad comprendida entre 0.5 y 1.0. En el caso en que el operador de cruce no se aplique, la descendencia se obtiene simplemente duplicando los padres.^{24,25}

Esquema No. 3.- Operador de cruce basado en un punto.



El operador de mutación se aplica a cada hijo de manera individual, y consiste en la alteración aleatoria (normalmente con probabilidad pequeña) de cada gen componente del cromosoma. El Esquema 4 muestra la mutación del quinto gen del cromosoma. Sí bien puede en principio pensarse que el operador de cruce es más importante que el operador de mutación, ya que proporciona una exploración rápida del espacio de búsqueda, éste último asegura que ningún punto del espacio de búsqueda tenga probabilidad cero de ser examinado, y es de capital importancia para asegurar la convergencia de los Algoritmos Genéticos.^{24,25}

Esquema No.4.- Operador de Mutación.



Si el AG ha sido correctamente implementado, la población evolucionará a lo largo de las generaciones sucesivas de tal manera que la adaptación media extendida a todos los individuos de la población, así como la adaptación del mejor individuo se irán incrementando hacia el óptimo global. El concepto de convergencia está relacionado con la progresión hacia la uniformidad: un gen ha convergido cuando al menos el 95 % de los individuos de la población comparten el mismo valor para dicho gen. Se dice que la población converge cuando todos los genes han convergido. Se puede generalizar dicha definición en el caso en que algunos de los individuos de la población hayan convergido.

^{24,25}

Algoritmos Genéticos Paralelos

En este apartado se introducirán tres maneras diferentes de explotar el paralelismo de los Algoritmos Genéticos, por medio de los denominados modelos de islas.

Modelos de islas.

La idea básica consiste en dividir la población total en varias subpoblaciones en cada una de las cuales se lleva a cabo un Algoritmo Genético. Cada cierto número de generaciones, se efectúa un intercambio de información entre las subpoblaciones, proceso que se denomina migración. La introducción de la migración hace que los modelos de islas sean capaces de explotar las diferencias entre las diversas subpoblaciones, obteniéndose de esta manera una fuente de diversidad genética. Cada subpoblación es una "isla", definiéndose un procedimiento por medio del cual se mueve el material genético de una "isla" a otra. La determinación de la tasa de migración, es un asunto de capital importancia, ya que de ella puede depender la convergencia prematura de la búsqueda. Se pueden distinguir diferentes modelos de islas en función de la comunicación entre las subpoblaciones.^{24,25}

Quinolonas

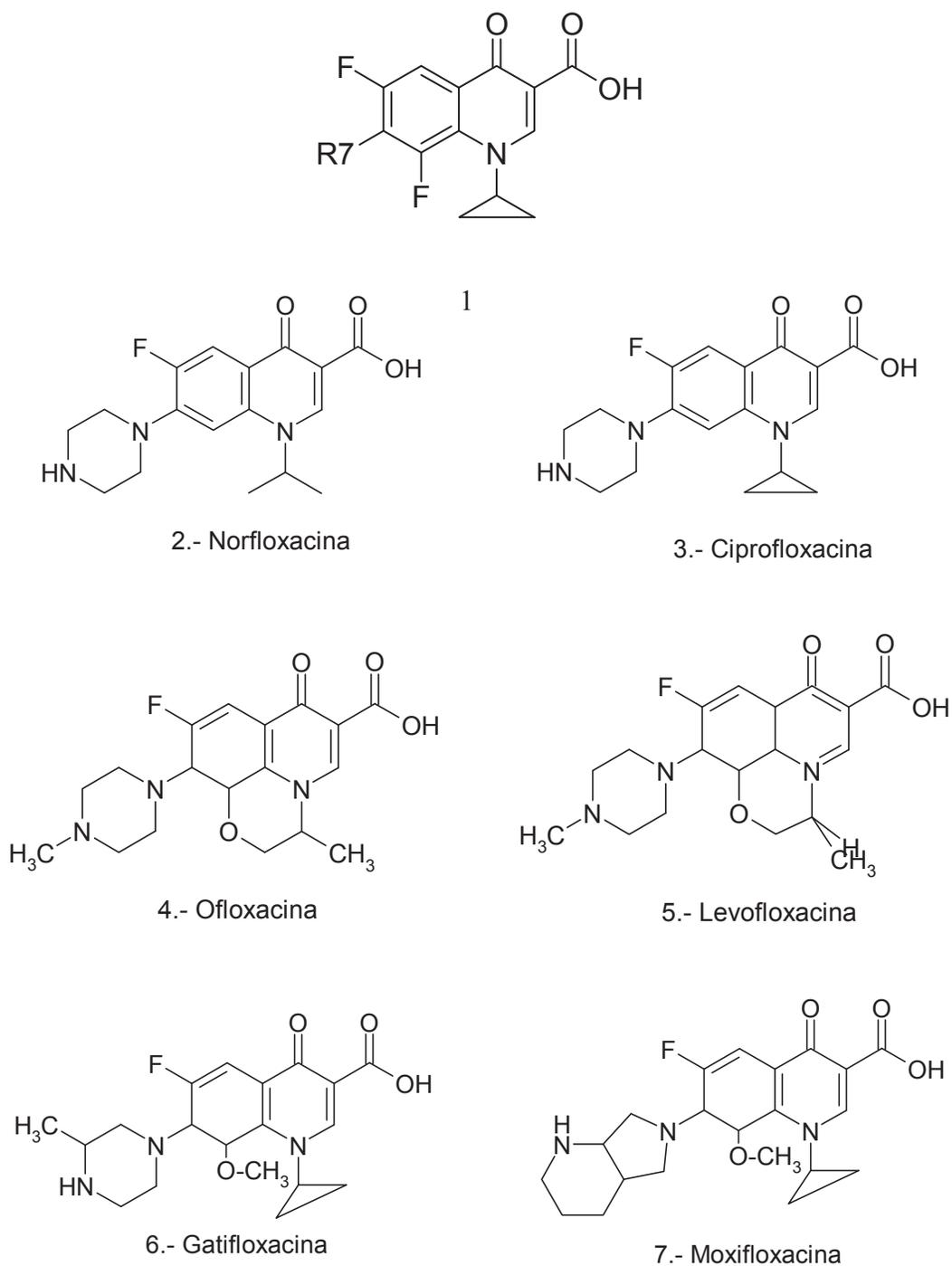
Puesto que en la presente tesis se aplica AGs para resolver un QSAR de una serie de difluoroquinolonas resulta conveniente explicar algunos aspectos importantes que permitan interpretar adecuadamente los modelos obtenidos.

Estructura química y clasificación

Las quinolonas de uso clínico tienen una estructura formada por dos anillos de 6 miembros, con un nitrógeno en la posición **1**, un grupo carbonilo en la posición **4** y un grupo carboxilo en la posición **3** (Figura 1.- **1**). La potencia y el espectro de actividad aumentan de manera significativa cuando hay átomos de flúor en la posición **6**. La actividad también se incrementa si en la posición **7** hay un grupo piperazínico (norfloxacin, ciprofloxacino presentada en la Figura 1.- **2** y **3**) o un grupo metilpiperazínico (ofloxacino, levofloxacino, gatifloxacino en la Figura 1.- **4**, **5** y **6**). Aquellos compuestos que llevan en la posición **7** un doble anillo derivado del anillo pirrolidónico

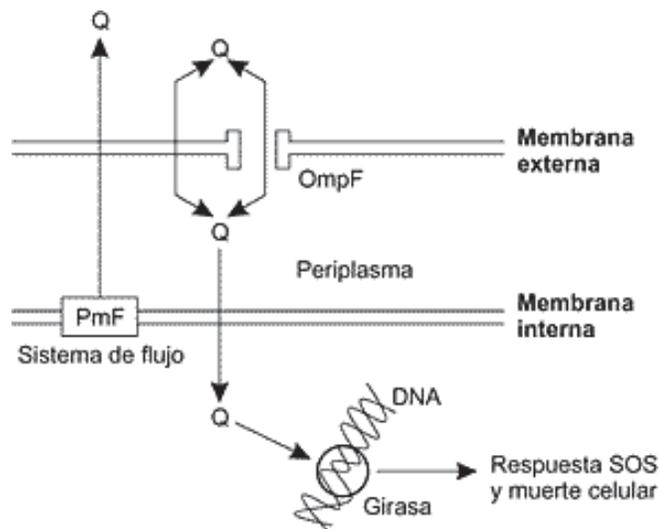
aumentan su actividad sobre bacterias grampositivas (moxifloxacino en la Figura 1.- 6).
10,26,27,30

Figura 1



Mecanismos de acción y resistencia

El mecanismo de acción de las Fluoro Quinolonas es complejo. Actúan en el ADN bacteriano, uniéndose a algunas de las topoisomerasas e inhibiendo su acción. Las topoisomerasas son enzimas que participan en el proceso de síntesis del ADN, mediante cambios en su topología. Se pueden clasificar en tipos I y II, dependiendo del mecanismo de acción. En las bacterias se encuentran las de tipo II que a su vez son subtipo II (girasa) y IV. En bacterias gram negativas, el blanco principal es la girasa. En bacterias gram positivas la principal diana es la topoisomerasa IV.^{10,23} El proceso de la actividad se resume en el Esquema 5:



Esquema No. 5.- El esquema muestra a grosso modo el mecanismo de acción de las quinolonas.²³

La presencia de una quinolona en el complejo ADN-girasa provoca una estabilización de la enzima inhibiendo su acción. La topoisomerasa IV forma complejos similares a los que se forman con la girasa. Los mecanismos más prevalentes de resistencia son alteraciones en alguna de las subunidades de la Girasa o de la Topoisomerasa.²³

Estudios QSAR en fluoroquinolonas

Los estudios QSAR para fluoro Quinolonas son escasos y en algunos casos inconclusos en su objetivo principal, por lo que se han propuesto algunas metodologías para su

desarrollo, como reducir las series de exploración, a un conjunto de moléculas con menos variaciones en sus sustituyentes,⁸ he incluso la creación de nuevos descriptores matemáticos específicos para la serie de exploración,^{1,8,11,15} estos generan conflictos a la hora de la interpretación del modelos y sobretodo al con gran numero de datos que se manejan en el estudio, ya que es requerido por lo menos 1 descriptor por cada molécula presente en el estudio, esto nos conduce a metodologías laboriosas que incluso no siempre son posibles de realizar.³¹ Estos factores han impedido un desarrollo pleno de los estudios QSAR para quinolonas, por ello se presenta una nueva metodología basada en algoritmos genético para el desarrollo de QSAR de fluoroquinolonas, permite el uso de un gran número de descriptores (a diferencia de los estudios QSAR encontrados donde los cálculos se complican con la presencia 15 a 20 descriptores). El presente estudio fue realizado con un promedio 1600 descriptores, cerca de 80 veces más descriptores que un QSAR tradicional, permitiendo así la creación de modelos más representativos y exactos.^{28,31}

El estudio QSAR permitirá identificar las variaciones físico-químicas en una serie de quinolonas y como afectan estas variaciones la actividad antibacteriana además en base a la cuantificación de estas variaciones, se logrará diseñar una base de datos para la creación de nuevas moléculas con un alto porcentaje de optimización para una actividad biológica específica.

La base de datos⁷ consultada para la elaboración de la presente tesis, considera solo 8 descriptores para la elaboración de un QSAR, entre los mas relevantes figuran el volumen total de van der Waals, el área total de la superficie y la carga electrónica del sustituyente en C7, y en ella se llega ala conclusión de que la carga electrónica y el área total de la superficie determinan la actividad contra las bacterias gram negativas. Esto es demostrado con los modelos obtenidos en el estudio pero no fue posible determinar una relación estructura-actividad para las bacterias gram positivas, declarando que la no obtención de buenos resultados para dicha actividad de las quinolonas se debe a los descriptores utilizados en el estudio. La presente tesis considera la ampliación del numero de descriptores presentes en el análisis y la reducción de tiempo en la elaboración de le análisis QSAR.

Objetivos

Objetivo general.-

- Elaborar un análisis cuantitativo estructura-actividad (QSAR) de difluoro quinolonas utilizando técnicas de inteligencia artificial.

Objetivos particulares.-

- Crear modelos QSAR con ayuda de algoritmos genéticos, de difluoroquinolonas, con alta capacidad predictiva considerando un número elevado de descriptores moleculares (1600).
- Utilizar los modelos obtenidos para la predicción de la actividad de nuevas moléculas que permitan validar externamente los modelos.

Metodología

Metodología.-

1. Selección de la serie de exploración.- Buscando antecedentes de QSAR en quinolonas se encontró una serie de exploración donde se realizan cambios estructurales en la posición C7 de las quinolonas⁷, teniendo interés en dichos cambios se retomó la serie de exploración, creando cambios en las condiciones de elaboración.
2. Obtención de la optimización geométrica.- Encontrada la serie de exploración se elaboraron las estructuras de las quinolonas en el software de HiperChem 6.0¹ Una vez elaboradas las moléculas son optimizadas geoméricamente con cálculos semiempíricos AM1.
3. Cálculo de los descriptores físico-químicos para la serie de exploración.- Una vez optimizadas las moléculas se calculó una serie de descriptores físico-químicos con ayuda del software Dragón².
4. Obtención de los modelos empleando algoritmos genéticos.- La elaboración del QSAR con una gran cantidad de descriptores sería muy complicado y laborioso, he incluso imposible, por lo que fue necesario apoyarse en el software Moby Digs³ el cual está basado en algoritmos genéticos.
5. Validación estadística de los modelos.- los modelos obtenidos con los algoritmos genéticos pueden tener errores en su selección y para evitar dicho error los modelos son validados en su carácter estadístico y predictivo.
6. Predicción de la actividad de compuestos ajenos a la serie de exploración.- Una vez realizada una predicción de moléculas dentro de la serie, se prosigue a predecir la actividad de moléculas ajenas a la serie y tengan una actividad ya conocida.

Resultados

Y

Discusión

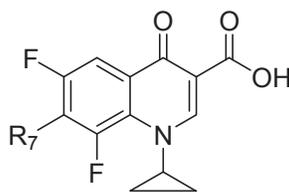
Discusión y Resultados

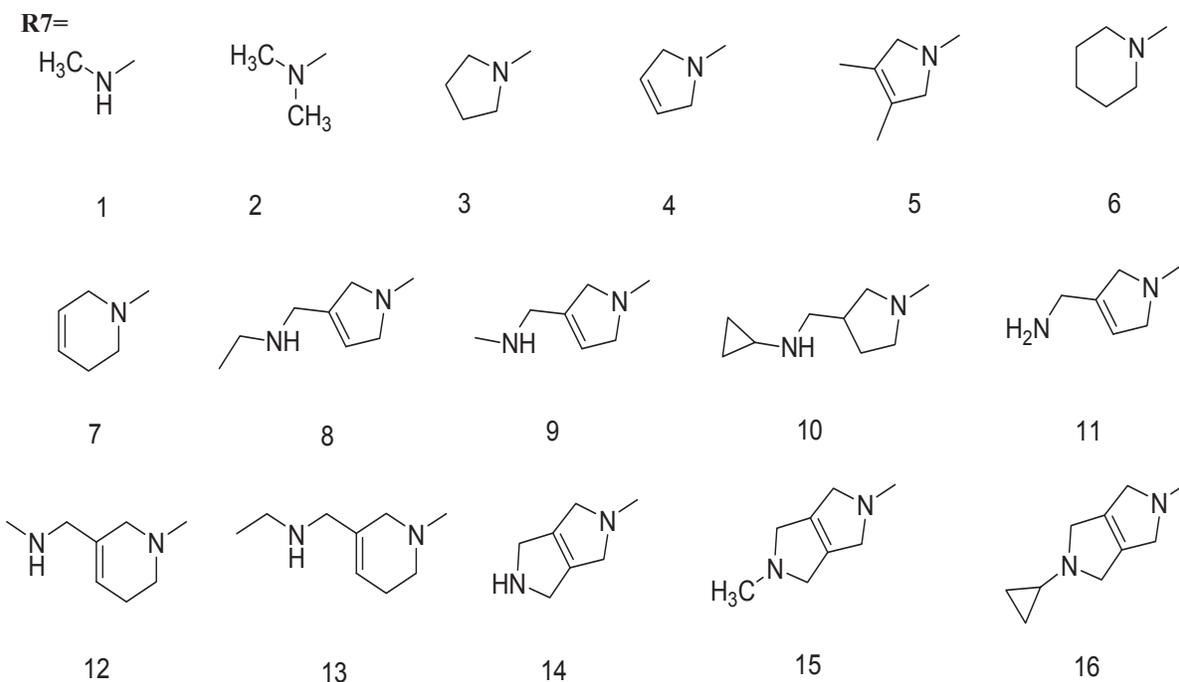
1. Obtención de descriptores y Funciones de ajuste (Fitnes functions)

Para realizar el QSAR de un conjunto de compuestos es necesario conocer su actividad biológica cuantitativamente. Así, se sabe que las fluoroquinolonas presentan actividad antibacteriana de amplio espectro. Puesto que la actividad sobre cada tipo de bacteria está marcada por las diferencias en respuesta al antibiótico, que pueden ir desde metabólicas hasta de transporte, no es del todo factible realizar un estudio QSAR que generalice una relación entre la estructura y la actividad responsable de causar la muerte de cualquier microorganismo. Por lo anterior resulta necesario realizar un estudio QSAR para cada bacteria de interés. Aun, entre bacterias del mismo genero y especie se presenta este efecto pues entre dos congéneres una puede ser resistencia al grupo de compuestos y por lo tanto comportarse de manera diferente a su pariente silvestre.

La actividad antibacteriana es registrada convencionalmente como la dosis inhibitoria 50 (DI₅₀) o también llamada concentración mínima inhibitoria (MIC), la cual es expresada en µg/ml e indica la concentración a la cual se inhibe el crecimiento del 50% de las bacterias expuestas al antibiótico. Para el desarrollo de esta tesis, el análisis QSAR se realizó, tomando una base de datos⁷, con una serie de exploración de 16 difluoro quinolonas, que tiene como característica la variación de sustituyentes en C7, ilustrados en la Figura 2,⁷ probados en cepas de *Streptococcus pyogenes*, *Staphilococcus aureus*, *Pseudomona aeruginosa* y *Escherichia coli*.

Figura 2





Cada una de los compuestos fue modelado con métodos semi-empíricos AM1; posteriormente se calcularon 1664 descriptores para cada una de las moléculas. Los descriptores utilizados pueden agruparse por bloques de la siguiente manera:

Tabla 1. Clasificación de los descriptores utilizados para el QSAR.²⁰

<i>Bloque No.</i>	<i>Tipo de Descriptor</i>	<i>No. de desc.</i>
1	Constitucionales	48
2	Topológicos	119
3	Walk and path counts	47
4	Indicadores de conectividad	33
5	Indicadores de información	47
6	Autocorrelaciones 2D	96
7	Edge adjacency indices	107
8	Burden eigenvalue	64

9	Indicadores topológicos de carga	21
10	Indicadores basados en Eigenvalores	44
11	Perfiles Randico-Moleculares	41
12	Geométricos	74
13	RDF	150
14	3D-MoRSE	160
15	WHIM	99
16	GETAWAY	197
17	Conteo de grupos funcionales	154
18	Fragmentos Atomo-centrado	120
19	Carga	14
20	Propiedades Moleculares	29

Como se mencionó en los antecedentes de esta tesis, un análisis QSAR se lleva a cabo tradicionalmente por regresión lineal múltiple, lo cual resultaría en un caos si se intentase con cientos de descriptores como es este caso. De ahí que fue necesario el uso de algoritmos genéticos.

Primeramente se sometieron los descriptores a una selección, lo cual se obtuvo agrupándolos en familias, que se muestran en la Tabla 1. Debido a que los descriptores están agrupados conforme a las características de dichos parámetros, una vez especificadas las familias, se especificó la actividad biológica en común de la serie de exploración. El programa utilizado presenta la ventaja de permitir variaciones dentro del criterio de selección (funciones de ajuste), así como el tamaño de la ecuación y el número de evoluciones deseadas.

Las funciones de ajuste seleccionadas y una breve descripción de las mismas se expone a continuación:

Q^2 .- Capacidad predictoria: La asintota Q^2 tiene su origen en la capacidad o “bondad” del modelo en el ajuste de recta y la capacidad predictiva. Este valor derivado de la validación cruzada no es más que la habilidad del modelo para realizar la predicción. En un modelo donde r_{cross} no disminuya en más de un 20% con respecto al coeficiente de correlación, puede ser considerado como un modelo predictivo.^{18,19,21}

R^2 .- Coeficiente de determinación: Compara los valores Y estimados y reales, y los rangos con valor de 0 a 1. Si es 1, hay una correlación perfecta en la muestra, es decir, no hay diferencia entre el valor Y estimado y el valor Y real. En el otro extremo, si el coeficiente de determinación es 0, la ecuación de regresión no es útil para predecir un valor Y.^{18,19,21}

K_x .- Coeficiente total de correlación: Muestra una evaluación de la relación existente entre las variables.²¹

Cada una de las funciones de ajuste mencionadas anteriormente son expresadas porcentualmente en tabulaciones creadas por el software, y permiten seleccionar los modelos que posean los valores más altos de cada uno de estos parámetros. Una vez especificadas las funciones de ajuste se regula el número máximo y el mínimo de descriptores permitidos para un adecuado análisis. Si se desea hacer una selección de descriptores, el número de estos deberá abarcar una muestra representativa de cada familia y si ya se tienen seleccionados los descriptores o es un grupo reducido es preferible especificar que se requiere un modelo con pocos componentes.

Posteriormente se especifica el número de ciclos al que equivaldrá una evolución o generación, así como el momento en el cual será adicionado un nuevo descriptor a los modelos ya creados y comenzar con una nueva selección de modelos, dicho proceso será consecutivo y el software comenzará con el número de descriptores que se haya especificado como mínimo y creará todas las combinaciones existentes hasta que ingrese el

siguiente descriptor, hasta llegar al límite mayor, es necesario tomar en cuenta que si el nuevo descriptor se agrega en un rango corto de ciclos se estarán limitando los modelos debido a que las combinaciones entre los descriptores no llegaran a su fin y se corre el riesgo de pasar por alto descriptores que puedan ser importantes. En el caso contrario, si las combinaciones entre los descriptores llega a su fin y no se adiciona el nuevo descriptor a combinar, el software generara un bloqueo de cálculos: debido a que ya no existen mas combinaciones entre los descriptores existentes en ese momento. Por lo tanto, estas especificaciones tendrán que ajustarse de acuerdo a lo observado en las evoluciones y al grado de exactitud que se requiera en el momento (si es selección de variables o elaboración de un modelo).

Para la obtención de los modelos en este trabajo se tomaron las siguientes funciones de la Tabla 2:

Tabla 2. Funciones de ajuste y parámetros considerados para la obtención de los modelos.

Función de ajuste		Opciones de evolución			
Q ²	0-100	N. de modelos	50	N. de ciclos para ingreso de variables	20 000 iter
R ²	0-100	Mínimo de descriptores	2		
K _x	0-100	Máximo de descriptores	7 en selec. 4 en ec.	*Las otras opciones no se modificaron	

Una vez especificadas esta variantes en el software se comienzan los cálculos, mismos que serán tan grandes como el número de variables (o descriptores) que se estén usando y del número de descriptores permitidos en los modelos. Hay que considerar que las combinaciones entre descriptores se incrementa logarítmicamente conforme se adicionan nuevas variables al modelo, por lo que se requieren calculos demasiado grandes que incluso pueden ser infinitos, debido a lo cual el proceso puede tardar horas o días. Observando las población de una forma individual se encontró que después de un determinado número de evoluciones las primeras ecuaciones permanecen invariables y conforme avanzan los cálculos, el número de ecuaciones sin variación incrementa por lo que se tomo como

criterio para detener el proceso de evolución cuando 20 o más modelos permanecen sin variaciones de ningún tipo (aprox. o los 700 evoluciones); aun así, en el momento de seleccionar modelos finales se permitió que estos cálculos continuaran para observar posibles modificaciones en los modelos, cosa que no se observó al menos en los primeros 20 modelos en más de 5000 evoluciones.

Una vez obtenidos los modelos, fue valorada su capacidad de predicción recalculando Q^2 . Una vez hecho esto se llevó a cabo la predicción de la actividad para cada uno de los compuestos de la serie utilizando el mismo software, lo cual se corroboró manualmente empleando un programa de cálculo convencional (Excel de Microsoft).⁴ Se encontraron algunas variaciones en r^2 , F (F de Fisher) y s (desviación estándar), debido a la diferencia en el número de cifras después del punto que utiliza cada programa, pero no fueron significativas aunque se puede decir que la hoja de cálculo fue mas precisa además de permitir realizar un análisis minucioso de los modelos, siendo así, una forma alternativa de validación externa.^{18,19,21}

2. QSAR de difluoro quinolonas con actividad contra *Streptococcus pyogenes* (cepa.- 308)

Los modelos seleccionados para la actividad contra *Streptococcus pyogenes*, son presentados en la Tabla 7, la cual se presenta al final del análisis, el modelo principal de esta serie es mostrado a continuación (**Hoja de resultados 1**) así como los parámetros registrados en los cálculos de regresión lineal y sus respectivas matrices:

Este modelo está constituido por 4 descriptores entre los cuales se encuentran parámetros topológicos y de carga, estos descriptores son los siguientes:

D/Dr06.- Índice de distancia/desvió de los ciclos.²⁰

Hoja de resultados 1: Modelo 1.- *Streptococcus pyogenes* (cepa.- 308)

Ecuación. 4

$$\text{Log (1/MIC)} = (-0.00719) \text{ D/Dr06} + (3.68273) \text{ ATS7m} \\ + (-6.10207) \text{ ESpm10d} + (2.50817) \text{ BELp6} + 57.02181$$

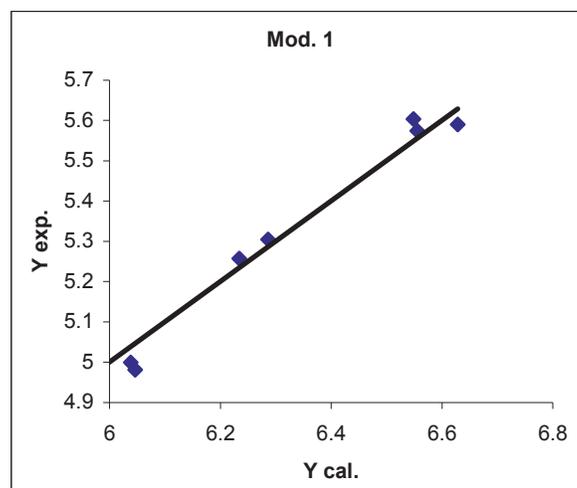
$$N= 16 \quad R^2= 0,992 \quad F= 368,05 \quad s= 0,055 \quad Q^2= 97.31$$

Tabla N. 3

ID	Log de la Actividad Experimental	Log de la Actividad calculada	Descriptores			
			D/Dr06	ATS7m	ESpm10d	BELp6
1	3.672	4.645	97.217	8.141	13.991	1.083
2	3.993	4.986	103.757	8.327	14.04	1.084
3	4.933	5.866	112.386	8.524	14.067	1.236
4	4.629	5.744	112.386	8.514	14.081	1.236
5	4.664	5.698	125.869	8.692	14.18	1.236
6	4.649	5.633	189.565	8.61	14.066	1.236
7	4.346	5.355	189.565	8.593	14.073	1.167
8	5.59	6.628	141.528	8.835	14.132	1.325
9	5.574	6.555	133.751	8.799	14.131	1.324
10	5.603	6.548	148.272	8.861	14.152	1.323
11	5.256	6.234	126.269	8.752	14.128	1.236
12	4.999	6.038	221.747	8.869	14.118	1.236
13	5.304	6.286	233.618	8.92	14.118	1.294
14	4.981	6.046	126.023	8.813	14.23	1.319
15	4.997	5.95	133.505	8.857	14.265	1.323
16	4.724	5.702	148.025	8.904	14.317	1.323

Tabla. 4

Matriz de correlación				
	D/Dr06	ATS7m	ESpm10d	BELp6
D/Dr06	1			
ATS7m	0,52990444	1		
ESpm10d	0,05510756	0,7505404	1	
BELp6	0,24292012	0,89135695	0,74834895	1



Hoja resultados del modelo No. 1. En esta hoja se encuentran recabada toda la información del modelo, la hoja de resultados cuenta con la ecuación y sus respectivos parámetros estadísticos, la matriz de correlación de los descriptores (tabla 3), la matriz de las actividades con el valor de los descriptores y el gráfico de la regresión lineal.

- **ATS7m**.- Descriptor de autocorrelación espacial calculado en función de una gráfica desarrollada por las propiedades físico-químicas de la molécula, siendo en este caso el peso molecular y la masa atómica.²⁰
- **ESpm10d**.- Momento espectral de la molécula (momento dipolar).²⁰
- **BELp6**.- Propiedad específica del descriptor; polarizabilidad atómica.²⁰

El valor de los descriptores para cada molécula y los datos estadísticos resultantes de la exploración se muestran en la Tabla 3 de la Hoja de resultados 1, al igual que los resultados de la actividad experimental y la calculada, mostrando además el gráfico 1, de la regresión lineal.

Aunque se puede apreciar una alta predicción en el modelo, el número de descriptores rebasa el número ideal, que sería un descriptor por cada cinco compuestos a explorar y lo que es más, hasta este punto faltaría corroborar la correlación que pudiese existir entre los descriptores, por lo que se llevó a cabo el análisis matricial, que se muestra en la tabla 4.

En la tabla 3 se aprecia que existen descriptores altamente correlacionados (valores superiores a 0.5), lo que indica redundancia. La correlación entre los descriptores **ATS7m** y **BELp6**, que es la más alta en la matriz. Por lo anterior se consideró conveniente separar los descriptores relacionados y crear dos ecuaciones para entonces seleccionar el mejor modelo y después fragmentarse nuevamente y realizar las combinaciones existentes entre ellos. Así, se encontró que solo dos de los descriptores estaban determinando los resultados y que los otros influían en el resultado al interactuar con los descriptores determinantes. El resultado se muestra en la Ecuación 5.

Ecuación 5.-

$$\text{Log (1/Act.)} = (0.0091124) \text{ D/Dr06} + (-4.58010035) \text{ ATS7m} \\ + (6.42299726) \text{ ESpm10d} + (-52.6137627)$$

N= 16	R ² = 0.973	F= 149.03	s= 0.094	Q ² = 95.55
-------	------------------------	-----------	----------	------------------------

En la Ecuación 5 se puede observar que fue eliminado el descriptor **BELp6** el cual se encontraba altamente relacionado con el descriptor **ATS7m** y el **ESpm10d**, aunque este último en menor proporción. Aunque se puede observar un decremento en los valores r^2 y F , y una elevación en la s , aún así el modelo sigue siendo estadísticamente aceptable.

Al hacer un análisis minucioso de los 50 modelos arrojados por algoritmos genéticos se encontró que el modelo correspondiente a la Ecuación 5 figuraba en la posición 34 con factor de predicción (Q^2) de 95.55.

Como consecuencia de la omisión del descriptor **ATS7m** en el Modelo 1 se obtiene la Ecuación 6:

Ecuación. 6

$$\text{Log (1/Act.)} = (0.00011585) \text{ D/Dr06} + (3.39021764) \text{ ESpm10d} \\ + (-8.13185716) \text{ BELp6} + (-38.0787835)$$

$$N= 16 \quad R^2= 0.844 \quad F= 21.73 \quad s= 0.229$$

Se observa que el descriptor eliminado es importante en el modelo, pero no lo suficiente para invalidar estadísticamente el resultado en su ausencia, lo que indica que el descriptor en cuestión no es determinante.

La fragmentación del modelo se llevó hasta dos variables cuyas ecuaciones con los descriptores **D/Dr06** y **ATS7m** presentan valores de r^2 menores a 0.7 y las s superiores a 0.25, por lo que fueron descartados.

Los modelos obtenidos expresan, en común, que la actividad está en función de tres diferentes factores fisicoquímicos. El descriptor **D/Dr06** es una propiedad topológica que implica la presencia de heterociclos, que en la serie de exploración se pueden observar que son de cinco y seis miembros, ambos nitrogenados como se observa en las Figuras 2 y 3 (la

figura 3 se presenta en la pag. 43), estos a su vez poseen sustituyentes lineales, cíclicos e incluso se aprecian biciclos, todos cuentan con nitrógeno en su estructura. El descriptor **D/Dr06** indica que la actividad será afectada positivamente con la presencia de dichos heterociclos, con un particular incremento en la actividad cuando hay pirroles sustituidos en la posición 3.

Otro factor que cabe resaltar es la autocorrelación espacial **ATS7m**, la cual también es de índole topológica, sin embargo esta indica las modificaciones presentes en los heterociclos, es decir los sustituyentes que estos posean. Un análisis minucioso muestra que la actividad será beneficiada por los sustituyentes lineales en la posición C-3 de los heterociclos, como se observa con los sustituyentes **10** y **13** de las Figuras 2 y 3.

El tercer factor que influye es la polarizabilidad de la molécula que puede ser atribuida a los nitrógenos encontrados en los heterociclos y en las aminas secundarias encontradas como sustituyentes. En la ecuación se aprecia que a mayor polarizabilidad de la molécula está presenta una mejor actividad, lo cual posiblemente esté involucrado con el reconocimiento a la girasa, enzima responsable de la actividad, pues cabe recordar que las interacciones entre especies deficientes en electrones (electrófilo) se dan con especies ricas en electrones (nucleófilo) y que entre más polarizable sea la nube electrónica mas permisible, geométricamente, será la interacción.

Modelo 2

El **Modelo 1** fue producto de la selección de cuatro descriptores cuya fragmentación llevó a ecuaciones de tres de ellos. Lo anterior motivó a realizar una nueva búsqueda de ecuaciones con solo dos descriptores partiendo el análisis desde un principio con ese objetivo. El producto de esta nueva búsqueda fue la Ecuación 7 que se encuentra en la **Hoja de resultados 2**.

Los descriptores obtenidos en este modelo son los siguientes:

- **SHP2**.- Interacciones geométricas en la estructura 3D de las moléculas; analiza la similitud y/o diversidad en las características de la molécula.²⁰
- **G(O..F)**.- Parámetro obtenido de coordenadas X y Y refiriéndose a dos heteroátomos.²⁰

La ecuación 8, que involucra dos descriptores, presenta una $Q^2 = 90.45$ la cual se puede considerar como un valor alto de predicción. Presenta alta correlación entre la actividad calculada y la observada experimentalmente (Gráfica 2). El gráfico, realizado con el logaritmo del inverso de la concentración, muestra una clara pero ligera dispersión de las actividades que conserva la tendencia a la línea recta.

El modelo presenta dos características singulares; la primera es que uno de los descriptores implica la distancia entre los fluoros y los oxígenos [**G(O..F)**], cambio que conforme a la adición de los diversos sustituyentes son aparentemente muy pequeños pero que matemáticamente son significativos en la actividad de las quinolonas. En la Tabla 5 se puede apreciar que entre menor sea la distancia, entre los átomos de fluor y oxígeno, la actividad será optimizada. Los datos de actividad y los valores de los descriptores se presentan en la Tabla 5 que se encuentra en la **Hoja de resultados 2**.

El segundo aspecto que se toma en cuenta en el **Modelo 2** es el cambio en la estructura 3D de los sustituyentes. Al analizar la estructura de las quinolona, la actividad y el valor del descriptor **SHP2** se observa que las moléculas presentan una mejor actividad al poseer sustituyentes constituidos por heterociclos de cinco átomos que poseen a su vez sustituyentes lineales en la posiciones **C3** con inclinación hacia la parte anterior a los anillos de la quinolona como se muestra en la Figura 3.

Hoja de resultados No.2: Modelo No.2.- *Streptococcus pyogenes* (cepa.- 308)

$$\text{Log (1/Act.)} = (-12.35847) \text{ SHP2} + (-6.612466) \text{ G(O..F)} + (284.73295)$$

Ecuación 8

$$N= 16 \quad R^2= 0.946 \quad F= 116.117 \quad s= 0.136 \quad Q^2 = 90.45$$

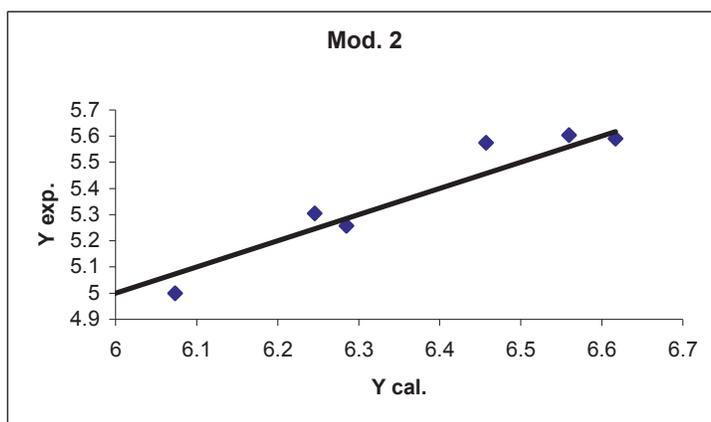
Tabla N. 5

ID	Log de la Actividad Experimental	Log de la Actividad calculada	Descriptores	
			SHP2	G(O..F)
1	3.672	4.529	0,464	41,659
2	3.993	5.106	0,466	41,568
3	4.933	5.886	0,42	41,536
4	4.629	5.781	0,42	41,552
5	4.664	5.722	0,429	41,544
6	4.649	5.564	0,399	41,624
7	4.346	5.564	0,399	41,624
8	5.59	6.616	0,377	41,506
9	5.574	6.456	0,391	41,504
10	5.603	6.559	0,372	41,524
11	5.256	6.284	0,406	41,502
12	4.999	6.073	0,368	41,605
13	5.304	6.245	0,353	41,607
14	4.981	5.892	0,411	41,552
15	4.997	5.743	0,415	41,567
16	4.724	5.892	0,403	41,567

Tabla. 6

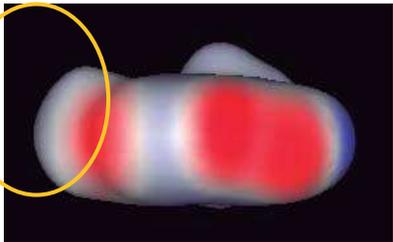
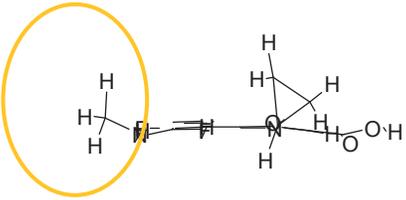
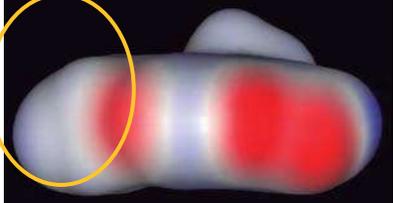
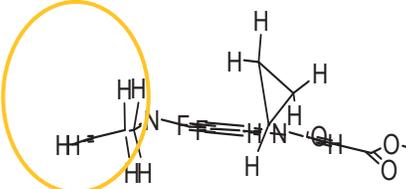
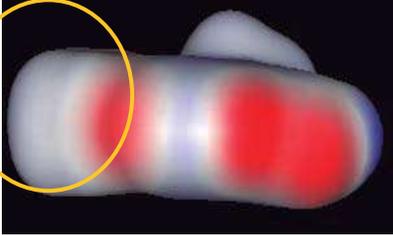
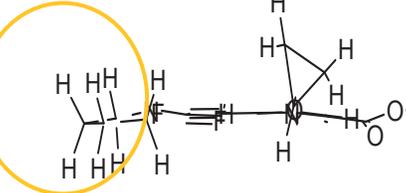
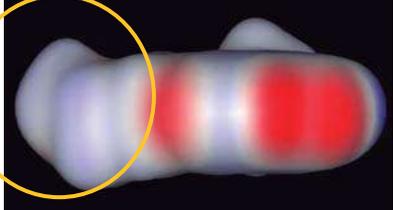
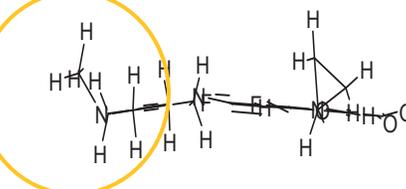
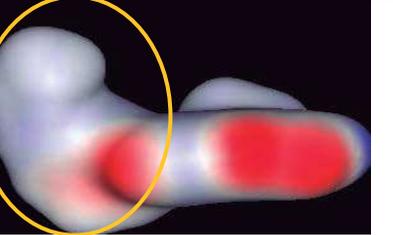
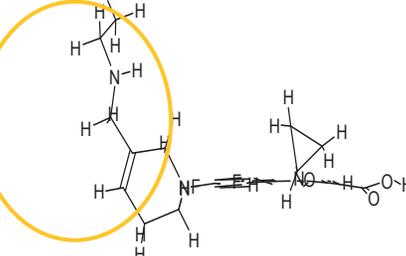
Matriz de correlación		
	SHP2	G(O..F)
SHP2	1	
G(O..F)	0.17865377	1

Grafico. 2



Hoja resultados del modelo No. 2. En esta hoja se encuentran recabada toda la información del modelo, la hoja de resultados cuenta con la ecuación y sus respectivos parámetros estadísticos, la matriz de correlación de los descriptores (tabla. 5), la matriz de las actividades con el valor de los descriptores y el grafico de la regresión lineal.

Figura 3

ID	SHP2	Estructura en 3D	Formula desarrollada
1	0,464		
4	0,42		
3	0,42		
15	0,415		
13	0,353		

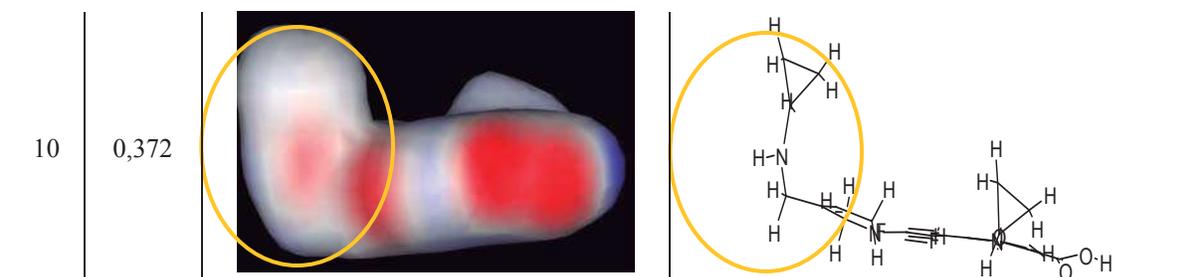


Figura 3.- Ilustración de las moléculas **1, 3, 4, 10, 13 y 15** en donde el círculo indica la distorsión que refiere el descriptor **SHP2**. Las quinolonas ilustradas se observan de forma horizontal de los anillos en estructura base de las moléculas, solo se presentan 6 moléculas de la serie por ser las mas representativas y se muestran 2 tipos de estructuras: en la segunda columna tendremos el valor del descriptor **SHP2**, en la tercer columna se muestra la forma en 3D del área de superficie molecular, en la cuarta columna se presenta la estructura desarrollada. Las moléculas se encuentran ordenadas conforme a su actividad de menor a mayor.^{5,6}

A partir de los datos de la tabla 5 y la ecuación 8, se puede deducir que una disminución en los valores del descriptor **SHP2** incrementará la actividad antibacteriana; aunque se trata de valores numéricos, una interpretación física resulta más fácil al analizar la Figura 3.

Los modelos 1 y 2 para *Streptococcus pyogenes* (cepa.- 308), presentan diferencias en el tipo de descriptores de las cuales, es relevante que, en el modelo 1, 1.1, y 1.2 predomina un descriptor que involucra el momento dipolar **ESpm10d**, mientras que en el Modelo 2 predomina un descriptor de tipo topológico SHP2.

Pese a la diferencia en el tipo de descriptores, en la interpretación de los modelos ambos coinciden en que las sustituciones en **C7** presentan una mayor actividad al poseer heterociclos que confieran una zona plana que incremente la superficie de la molécula. A su vez, de acuerdo al Modelo 2, los heterociclos al tener sustituciones en el **C3'**, independientemente de la estructura del anillo, presentan un incremento en la actividad, que esta relacionada con la forma que la molécula adopta (una inclinación que se desvía del plano de la quinolona) (Figura 3).

Al realizar la búsqueda de nuevos modelos con un numero menor de descriptores y tratar de fragmentar la ecuación del modelo 2, no se encontraron mejores ecuaciones estadísticamente validadas ni con capacidad de predicción relevante. A continuación se presentan los 50 modelos calculados para *Streptococcus pyogenes*, del cual fue seleccionado el modelo anteriormente descrito

Tabla. 7.- Modelos para *Streptococcus pyogenes*

ID	Tamaño	Modelos	R2	Q2	Q2boot	Kx	F	s
1	4	D/Dr06 ATS7m ESpm10d BELp6	99.21	98.58	97.31	57.83	344.97	0.054
2	4	D/Dr06 ATS7m ESpm11d BELp6	99.2	98.58	97.24	57.27	342.86	0.055
3	4	D/Dr06 X5v ATS7m ESpm10d	99.29	98.36	97.89	62.57	385.6	0.051
4	4	D/Dr06 X5v ATS7m ESpm11d	99.16	98.07	97.08	62.03	325.68	0.056
5	4	D/Dr05 EEig15d BEHm2 BELp6	98.72	97.81	96.53	78.7	211.88	0.069
6	4	EEig15d ESpm11d BELp6 G(O..F)	98.76	97.77	96.3	65.69	219.66	0.068
7	4	EEig15d ESpm10d BELp6 G(O..F)	98.72	97.73	96.6	66.28	212.87	0.069
8	4	D/Dr06 EEig15d BEHm2 BELp6	98.54	97.41	95.94	59.19	186.06	0.074
9	4	D/Dr05 EEig15d BEHm2 Mor03e	98.49	97.36	96.03	58.72	179.37	0.075
10	4	RDF030u RDF115p HATS5e TE1	98.5	97.33	96.33	61.56	180.42	0.075
11	4	D/Dr05 ATS7m ESpm11d GGI3	98.62	97.3	94.43	72.57	196.67	0.072
12	4	D/Dr06 X5sol ATS7m ESpm10d	98.68	97.3	94.2	63.78	205.41	0.07
13	4	D/Dr05 EEig15d BEHm2 Mor03u	98.54	97.25	96.23	59.31	185.06	0.074
14	4	RDF030u RDF115v HATS5u AMR	98.58	97.23	96.3	51.07	191.48	0.073
15	4	D/Dr06 ATS7m ESpm11d Mor32p	98.42	97.21	94.77	48.37	171.4	0.077
16	4	D/Dr05 X5sol EEig15d BEHm2	98.39	97.21	95.99	78.38	168.13	0.078
17	4	D/Dr06 ATS7m ESpm11d Mor16v	98.52	97.2	95.01	53.49	183.48	0.074
18	4	RDF030u RDF115v HATS5e TE1	98.77	97.19	96.32	61.39	220.04	0.068
19	4	D/Dr06 ATS7m ESpm10d Mor16v	98.58	97.18	93.73	53.76	191.59	0.073
20	4	D/Dr05 D/Dr06 ATS7m ESpm10d	98.52	97.18	91.58	62.12	182.63	0.074
21	4	D/Dr05 D/Dr06 ATS7m ESpm11d	98.48	97.12	92.92	61.86	178.61	0.075
22	4	D/Dr06 ATS7m ESpm10d BELv6	98.46	96.42	93.77	58.29	176.32	0.076
23	4	D/Dr05 X5sol ATS7m ESpm11d	97.97	96.06	92.2	74.68	132.92	0.087
24	4	D/Dr06 ATS7m ESpm10d VRp2	98.09	95.93	92.08	63.02	141.55	0.084
25	4	D/Dr05 EEig15d ESpm11d BEHm2	97.37	95.72	92.95	82.51	101.91	0.099
26	4	D/Dr06 ATS7m ESpm10d PCWTe	97.56	95.68	87.94	56.44	110.07	0.095
27	4	D/Dr06 EEig15d ESpm10d BELp6	97.81	95.59	90.05	58.02	122.55	0.091
28	4	D/Dr06 ATS7m BEHm2 BELp6	98.05	95.57	93.75	58.84	138.22	0.085
29	3	D/Dr05 ATS7m ESpm11d	97.37	95.57	94.43	69.92	148.1	0.095
30	3	D/Dr05 EEig15d BEHm2	97.09	95.57	94.77	77.11	133.51	0.1
31	3	D/Dr06 ATS7m ESpm10d	97.41	95.55	92.69	47.26	150.73	0.094
32	3	D/Dr06 ATS7m ESpm11d	97.34	95.51	92.4	46.34	146.17	0.096
33	3	D/Dr05 ATS7m ESpm10d	97.26	95.47	94.21	70.35	142.04	0.097
34	4	D/Dr05 ATS7m ESpm10d ESpm11d	97.38	95.33	92.64	78.62	102.21	0.099
35	4	D/Dr05 ATS7m EEig15d ESpm11d	97.46	95.31	92.52	75	105.38	0.098
36	4	D/Dr06 ATS7m ESpm10d Ku	97.8	95.28	90.27	58.9	122.04	0.091
37	4	D/Dr05 ATS7m ESpm11d QXXv	97.4	94.94	91.6	60.51	103.06	0.099
38	4	D/Dr05 ATS7m ESpm11d BELp6	97.39	94.93	91.25	73.83	102.6	0.099
39	4	D/Dr05 ATS7m ESpm10d BELp6	97.27	94.84	91.3	74.26	97.99	0.101
40	4	D/Dr05 EEig15d BEHm2 QXXv	97.18	94.7	91.87	61.95	94.86	0.103
41	4	D/Dr05 EEig15d BEHm2 Mor25m	97.14	94.61	92.46	76.86	93.25	0.103
42	4	D/Dr05 EEig15d BEHm2 RDF030u	97.09	94.51	92.41	64.38	91.87	0.104
43	4	D/Dr06 ATS7m ESpm11d QXXv	97.35	94.31	85.97	56.96	101.17	0.099
44	4	D/Dr05 ATS7m EEig15d BEHm2	97.77	94.25	92.91	76.38	120.5	0.091
45	4	D/Dr05 EEig15d ESpm11d BELv6	97.02	94.23	90.51	75.07	89.49	0.106
46	4	D/Dr06 ATS7m EEig15d ESpm11d	97.38	94.2	90.11	61.79	102.33	0.099

47	2	SHP2 G(O..F)	94.03	90.45	90.73	17.87	102.32	0.137
48	2	BELv6 RDF030u	92.18	89.62	89.1	71.47	76.63	0.157
49	1	RDF030u	84.33	79.21	80.58	0	75.34	0.215
50	1	BELv6	72.64	66.78	67.33	0	37.17	0.283

En la tabla 7 , se presentan los 50 modelos obtenidos para el QSAR de *Streptococcus pyogenes* , en conjunto con los parámetros mas relevantes calculados por el Moby Digs

3. QSAR de difluoro quinolonas con actividad contra *Staphylococcus aureus* (cepa.- SG-511)

Staphylococcus aureus es un patógeno común, en el ser humano, que se localiza principalmente en las mucosas y la piel. Puede originar abscesos y forúnculos en la dermis además de provocar osteomielitis, endocarditis y otro gran número de infecciones. *Staphylococcus aureus* es un coco que crece agrupado en racimos (de ahí su raíz "Staphylo"), que responde positivamente a la tinción de Gram, es aerobio y anaerobio facultativo. Es importante cuando de enfermedades hospitalarias se trata y de ahí la importancia de tener un control terapéutico del mismo.

El análisis QSAR para la cepa SG-511 arrojó un modelo de interés que se describe a continuación:

El “mejor modelo” encontrado está constituido por cuatro descriptores, comprendidos entre topológicos, atómicos y electrónicos, mismos que expresan en la Ecuación 9. Los valores correspondientes a cada descriptor y la actividad experimental de cada compuesto se presenta en se encuentra en la Hoja de Resultados 3.

En donde:

- **X5sol** (solvation connectivity index chi-5).- Descriptor topológico resultante de un vértice gradual, indica los índices de solvatación de la molécula con base en la conectividad de la de los átomos (omite hidrógenos).²⁰

Hoja de resultados 3. *Staphylococcus aureus* (cepa.- SG-511)

$$\text{Log (1/Act.)} = (-0.161802683697) X5\text{sol} + (-30.8714618905691) \text{Gu} +$$

Ecuación.9

$$(31.3945446175892) \text{H8m} + (12.2265963234019)$$

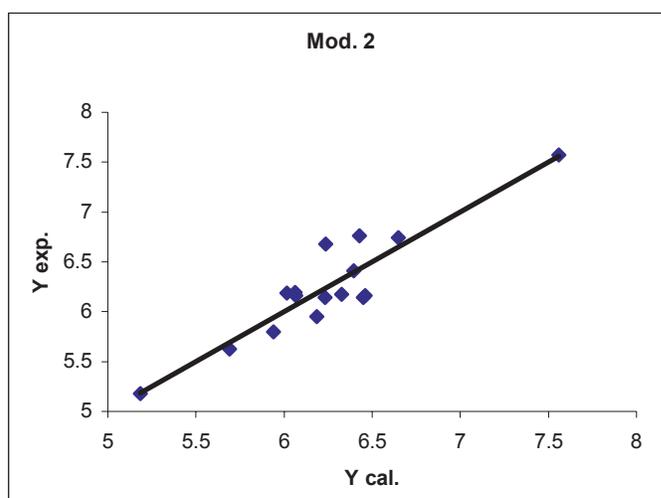
$$N= 16 \quad R^2= 0.847 \quad F= 22.204 \quad s= 0.23 \quad Q^2 = 84.55$$

Tabla N. 8

ID	Log de la Actividad Experimental	Log de la Actividad calculada	Descriptores		
			X5sol	Gu	H8m
1	5.178	5.185	4.629	0.214	0.01
2	5.798	5.940	4.755	0.197	0.018
3	6.679	6.237	5.41	0.186	0.02
4	6.407	6.394	5.41	0.186	0.025
5	6.158	6.067	6.114	0.199	0.031
6	6.144	6.231	6.022	0.184	0.021
7	6.141	6.450	6.022	0.183	0.027
8	6.745	6.648	6.37	0.19	0.042
9	7.574	7.559	6.158	0.184	0.064
10	6.758	6.426	6.547	0.18	0.026
11	6.160	6.459	5.993	0.191	0.035
12	6.192	6.064	6.639	0.176	0.011
13	5.953	6.183	6.851	0.17	0.01
14	6.174	6.327	6.782	0.184	0.028
15	6.190	6.017	7.187	0.195	0.031
16	5.625	5.690	7.709	0.212	0.04

Tabla No. 9

Matriz de correlación			
	X5sol	Gu	H8m
X5sol	1		
Gu	-0.1926955	1	
H8m	0.30857628	0.12287317	1



Hoja resultados del modelo No. 3 En esta hoja se encuentran recabada toda la información del modelo, la hoja de resultados cuenta con la ecuación y sus respectivos parámetros estadísticos, la matriz de correlación de los descriptores (tabla. 8), la matriz de las actividades (log (1 / [mmol/ml]) con el valor de los descriptores y el grafico de la regresión lineal.

- **Gu** (G total symmetry index / unweighted).- Descriptor global que evalúa la simetría total de la molécula, es calculado a partir de el grafico y los parámetros estadísticos de la molécula en 3D²⁰
- **H8m** (H autocorrelation of lag 8 / weighted by atomic masses).- Descriptor proveniente de la representación matricial de la estructura molecular (matriz x, y, z), describe la magnitud geométrica de la molécula representando las cadenas lineales y ramificaciones en la molécula, el calculo esta basado en la masa atómica.²⁰

La matriz de correlación en la Hoja de resultados 3 (Tabla 9), no muestra relación entre los descriptores. **X5sol** describe la solvatación de la molécula, **Gu** representara la simetría del sustituyente **H8m** la magnitud de la molécula.

La solvatación (**X5sol**) en los compuestos es una de las características más importantes para la actividad biológicas de las moléculas, ya que la absorción esta en función directa con la capacidad de solvatación de las moléculas, como se muestra en la tabla 8, donde se observa que los compuestos **8**, **9** y **10**, las cuales presentan la mayor actividad, poseen un índice de solvatación elevado y el compuesto **1** que presenta la menor actividad tiene baja solvatación, sin embargo se puede observar que en los compuestos **13**, **15** y **16** los índices de solvatación son los más elevados lo que estaría implicando un efecto parabólico en el efecto de la solvatación y la actividad.

La índice de simetría (**Gu**) en los sustituyentes se presenta inversamente proporcional a la actividad, como se observa en la tabla 8, en los compuestos **9** y **10**, así mismo es posible observar que en los compuestos **1** y **16** la actividad es menor y el índice de simetría es elevado.

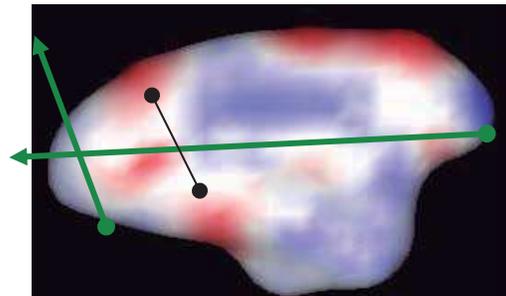
El tercer descriptor (**H8m**) representa cadenas o ramificaciones en los sustituyentes, midiendo así la magnitud de la molécula, este índice afecta la actividad proporcionalmente, como se observa en los compuestos **8** y **9**. El efecto de los descriptores **Gu** y **H8m** puede observarse en la estructura de la serie de exploración como se muestra en la Figura 4:

Figura 4

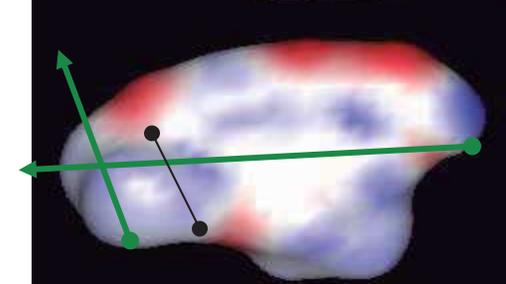
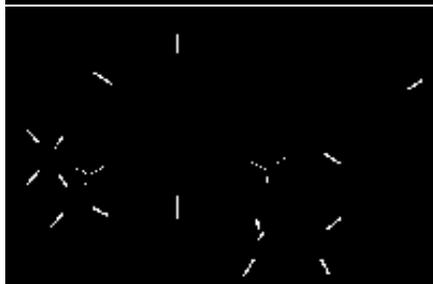
ID Estructura en 3D, forma linear

Estructura en 3D, volumen de la molécula

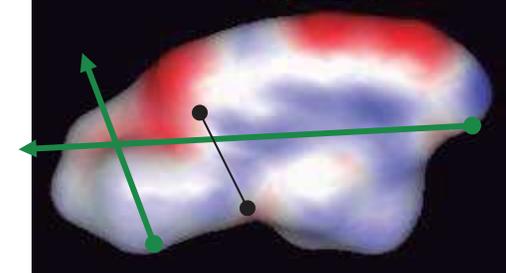
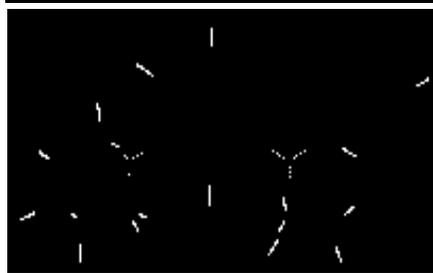
1



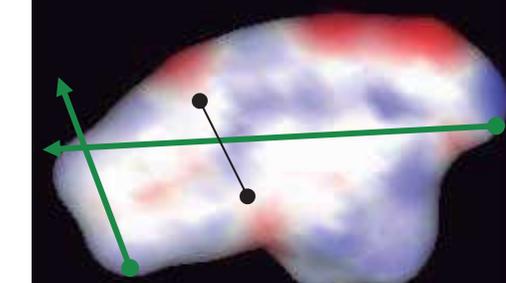
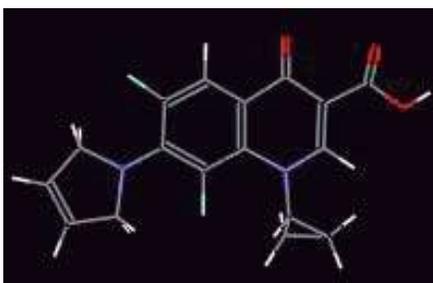
2



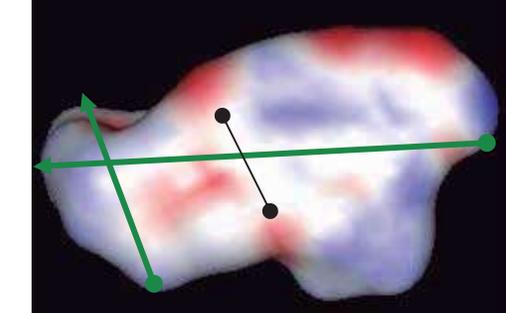
7



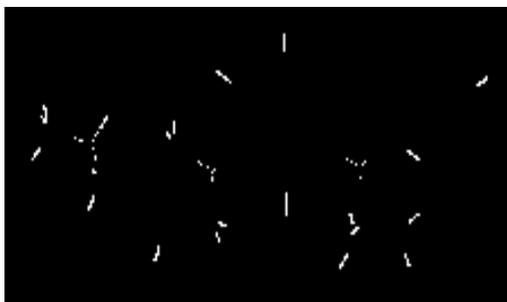
4



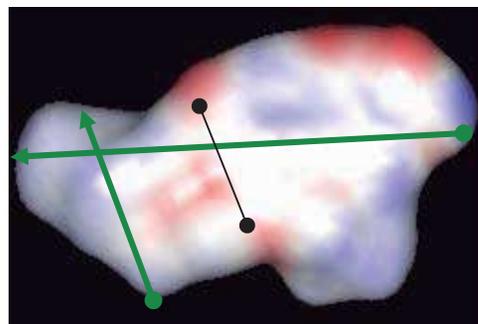
8



9



Simetría y magnitud
de la molécula



División del sustituyentes
y la estructura base

Figura 4.- Representación de las moléculas: 1, 2, 7, 4, 8, 9 en donde se señala con flechas verdes la magnitud en la estructura de la molécula, así como el simetría de esta. La flecha negra indica la división entre el sustituyente y el núcleo base de la quinolona. Los compuestos se ordenan de menor a mayor actividad. En la parte izquierda se observa la estructura en barras y en la parte derecha el volumen de la molécula.^{5,6}

Al realizar la búsqueda de modelos con un numero menor de descriptores y tratar de fragmentar la ecuación del Modelo 3, no se encontraron mejores ecuaciones estadísticamente validadas ni con capacidad de predicción relevante, por ello solo se presenta un modelo. A continuación se presentan los 20 modelos calculados para *Staphylococcus aureus*, del cual fue seleccionado el modelo anteriormente descrito:

Tabla 10 .- Modelos para *Staphylococcus aureus*

ID	tamaño	Modelos	R2	Q2	Q2boot	Kx	F	s
1	3	SEig H8m H7e	87.84	82.41	75.5	58.57	28.9	0.205
2	3	BELe4 BELp5 Gu	87.73	79.5	76.74	43.4	28.61	0.206
3	3	HIC H8m H7e	85.89	79.15	71.78	53.6	24.36	0.221
4	3	BELp4 BELp5 Gu	88.44	78.87	76.7	42.5	30.6	0.2
5	3	HTu H8m H7e	86.59	78.01	71.12	54.06	25.82	0.215
6	3	BELv5 BELe4 Gu	85.09	77.18	69	40.4	22.83	0.227
7	3	X5sol Gu H8m	84.55	76.94	68.9	21.31	21.89	0.231
8	3	RDF150v HATS5u H8m	85.73	76.38	48.55	27.96	24.02	0.222
9	3	X5 Gu H8m	84.14	76.27	68.07	23.46	21.22	0.234
10	3	VEA2 H8m H7e	83.74	76.1	65.1	51.58	20.6	0.237
11	3	ATS3m Gu H8m	83.86	75.54	68.05	24.87	20.79	0.236
12	3	E1s Gu H8m	84.27	75.43	62.35	36.88	21.42	0.233
13	3	RDF150v H8m HATS5e	84.27	75.26	31.82	35.46	21.43	0.233
14	3	G2e HATS5u H8v	85.98	74.56	68.6	3.42	24.53	0.22
15	3	BELv5 BELp4 Gu	85.55	74.32	71.06	39.26	23.67	0.223
16	3	Gu H8m H7e	82.64	72.99	63.18	24.96	19.04	0.245
17	3	BELp4 BELp5 HATS5e	86.73	72.85	72.66	70.44	26.15	0.214
18	3	SRW10 Gu H8m	82.16	72.37	62.48	19.19	18.42	0.248

19	3	BELp4 G(O..F) HATS5e	92.57	85.43	72.83	51.56	34.24	0.167
20	3	Gu H6m H8m	87.84	82.41	76.22	58.57	28.9	0.205

En la tabla 10, se presentan los modelos obtenidas en el QSAR para el *Staphylococcus aureus*, y los parámetros calculados para cada uno de los modelos.

4. QSAR de difluoro quinolonas con actividad contra *Escherichia coli* (cepa.- 308)

Hasta este punto, se ha descrito el análisis QSAR tomando en cuenta datos de MIC en bacterias Gram positivas. En esta sección se describe un estudio sobre *Escherichia coli*, una de las bacterias Gram negativas más representativa y usada con fines experimentales. Es un bacilo anaerobio facultativo, móvil por flagelos peritricos (que rodean su cuerpo), no forma esporas, es capaz de fermentar la glucosa y la lactosa. Esta se encuentra en la flora normal del intestino y suele asociarse con algunas infecciones intestinales. Cabe recordar que el ácido nalidíxico fue la primer quinolona utilizada como antibiótico y se empleaba precisamente en este tipo de infecciones. Actualmente el ciprofloxacino es una de las primeras opciones para el tratamiento de infecciones urinarias y del tracto gastrointestinal generadas por bacterias gram negativas. El QSAR se describe a continuación:

Modelo No.4

La “mejor ecuación” encontrada para *E. coli* se encuentra constituida por descriptores de tres grupos: topológicos, atómicos y de volumen y se describe en la ecuación 10 que se presenta en la Hoja de resultados 4.

En donde:

- **GAST8e** (Geary autocorrelation - lag 8 / weighted by atomic Sanderson electronegativities).- Descriptor de correlación basado en la electronegatividad atómica de Sanderson.²⁰
- **GAST4p** (Geary autocorrelation - lag 4 / weighted by atomic polarizabilities).- Descriptor de correlación basado en la Polarizabilidad de la molécula.²⁰

Hoja de resultados 4 *Escherichia coli* (cepa.- 308)

$$\text{Log (1/Act.)} = (4.62945) \text{ GAST8e} + (-13.80315) \text{ GAST4p} + (-0.86086) \text{ VRm2} + (-0.86086)$$

Ecuación. 10

$$N= 16 \quad R^2= 0.962 \quad F= 101.59 \quad s= 0.153 \quad Q^2= 93.4$$

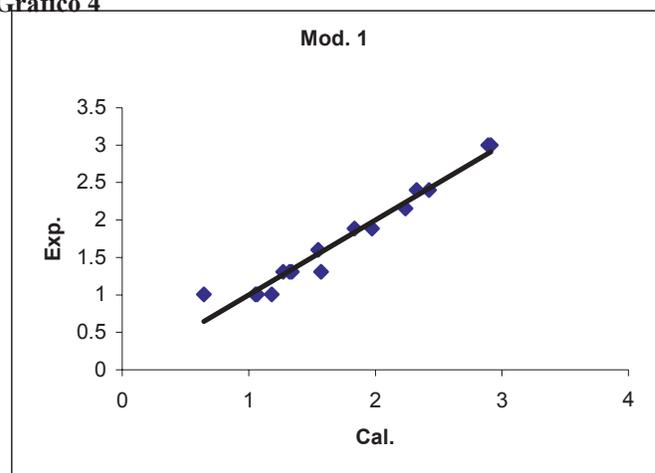
Tabla N. 11

ID	Log de la Actividad Experimental	Log de la Actividad calculada	Descriptores		
			GATS8e	GATS4p	VRm2
1	5.778	5.796	1.098	1.139	5.529
2	5.798	5.758	0.996	1.1	5.65
3	6.126	6.078	0.882	1.015	6.028
4	5.831	6.099	0.927	1.029	6.021
5	5.565	5.736	0.899	1.032	6.244
6	5.55	5.606	0.859	1.035	6.132
7	5.548	5.183	0.881	1.074	6.116
8	6.745	6.820	1.065	0.995	6.471
9	7.574	7.461	1.118	0.973	6.364
10	6.489	6.439	1.053	1.006	6.673
11	6.955	6.981	1.173	1.033	6.255
12	6.476	6.55	1.018	1.001	6.436
13	5.915	5.928	0.96	1.02	6.541
14	6.970	6.908	1.034	0.978	6.474
15	7.588	7.505	0.983	0.911	6.581
16	5.625	5.685	1.005	1.03	6.906

Tabla 12

Matriz de correlación			
	GATS8e	GATS4p	VRm2
GATS8e	1		
GATS4p	-0.05327981	1	
VRm2	0.11477975	-0.72148456	1

Grafico 4



Hoja resultados del modelo No. 4 En esta hoja se encuentran recabada toda la información del modelo, la hoja de resultados cuenta con la ecuación y sus respectivos parámetros estadísticos, la matriz de correlación de los descriptores (tabla 11), la matriz de las actividades con el valor de los descriptores y el grafico de la regresión lineal.

- **VRm2** (average Randic-type eigenvector-based index from mass weighted distance matrix).- Descriptor calculado mediante la matriz del gráfico molecular; Se omiten los hidrógenos. La matriz es calculada con los índices de conectividad que se basan en la propiedad de la masa atómica.²⁰

Los datos de actividad calculada vs experimental se pueden apreciar en la Tabla 11 y en la Gráfica 4 que se encuentran Hoja de resultados 4, en donde se puede observar alta capacidad de predicción.

Para llevar a cabo adecuadamente la interpretación, resulta conveniente analizar los descriptores **GATS8e**, **GATS4p** y **VRm2**, con respecto a la actividad. Los valores para cada compuesto se muestran en la Tabla 11. Donde se observa que el descriptor de **GAST8e** esta basado en la electronegatividad atómica, misma que afecta positivamente a la actividad como se puede comparar en los compuestos **9** y **5**. A menor electronegatividad, menor la actividad. Es importante señalar que el descriptor fue calculado con omisión de hidrógenos y que la electronegatividad calculada estará en función de los átomos de nitrógeno afectados por el número de átomos de carbono en el sustituyente.

La polaridad en la molécula se encuentra asociada al descriptor **GAST4p** el cual afecta negativamente a la actividad. Al observar el valor de la polaridad en conjunto con la actividad es notable que entre más alta sea polaridad, la actividad es menor como en el compuesto **7**; en el caso contrario, el compuesto **15** presenta una actividad máxima y una polaridad baja. Es decir, existe una relación inversamente proporcional.

VRm2 (índice de conectividad) afecta positivamente la actividad en la serie. Como ejemplo más notable se tienen los compuestos **9** y **15** como las más activas y en el caso contrario compuestos **1** y **2** en donde la conectividad es muy baja al igual que su actividad. El índice de conectividad al no considerar hidrógenos solo se ve incrementado por la presencia de ciclos en los sustituyentes.

En resumen, para *Escherichia coli*, la presencia de ciclos así como la presencia de átomos electronegativos, como el nitrógeno en los heterociclos y en algunas cadenas lineales, afectan positivamente la actividad de los compuestos explorados. La polaridad por el contrario desfavorece la actividad. Al realizar la búsqueda de nuevos modelos con un número menor de descriptores y tratar de fragmentar la ecuación del Modelo 4, no se encontraron modelos representativos estadísticamente y los factores de predicción fueron sumamente bajos, por ello únicamente se discute un modelo. A continuación se presentan los 50 modelos calculados para *Escherichia coli*, del cual fue seleccionado el modelo anteriormente descrito:

Tabla 13.- Modelos para *Escherichia coli*.

ID	Tamaño	Modelos	R2	Q2	Q2boot	Kx	F	s
1	3	GATS8e GATS4p VRm2	96.03	93.49	91.32	37.03	96.7	0.154
2	3	GATS8e GATS4p VRZ2	96.03	93.49	91.57	37.02	96.64	0.154
3	3	HVcpx BEHm8 QYYv	96.17	93.25	92.1	89.87	100.31	0.151
4	3	HVcpx BEHm8 RDF135u	96.71	93.01	92.42	80.24	117.47	0.14
5	3	GATS4p G1m BLTA96	95.54	92.82	91.46	32.71	85.68	0.163
6	3	GATS4p G1m MLOGP	95.53	92.79	90.39	32.71	85.41	0.163
7	3	GATS4p G1m BLTF96	95.5	92.74	89.78	32.65	84.82	0.164
8	3	GATS4p G1m BLTD48	95.5	92.73	90.24	32.72	84.8	0.164
9	3	nSK GATS8e GATS4p	95.64	92.42	91.1	36.16	87.66	0.161
10	3	X1 GATS8e GATS4p	95.52	92.28	90.69	35.95	85.24	0.163
11	3	X1sol GATS8e GATS4p	95.52	92.28	90.58	35.95	85.21	0.163
12	3	GATS4p G1m MLOGP2	95.12	92.24	90.11	32.31	78.01	0.17
13	3	XMOD GATS8e GATS4p	95.43	92.12	90.37	36.4	83.52	0.165
14	3	GATS8e GATS4p BEHm8	95.13	91.96	89.78	21.09	78.22	0.17
15	3	GATS8e GATS4p QYYv	95.3	91.75	90.16	34.26	81.15	0.167
16	3	GATS4p BELv4 BLTD48	95.15	91.69	90.1	41.22	78.45	0.17
17	3	GATS4p BELv4 MLOGP	95.11	91.64	89.9	41.36	77.78	0.171
18	3	GATS8e GATS4p BLTA96	94.9	91.56	89.51	39.28	74.49	0.174
19	3	GATS4p BELv4 BLTF96	95.07	91.56	89.93	41.11	77.18	0.171
20	3	GATS4p BELv4 BLTA96	95.08	91.56	89.2	41.4	77.28	0.171
21	3	GATS4p BELe4 BLTD48	94.87	91.55	89.21	40.14	74.02	0.175
22	3	GATS4p BELe4 MLOGP	94.84	91.51	89.31	40.29	73.46	0.175
23	3	GATS8e GATS4p MLOGP	94.86	91.48	88.92	39.36	73.8	0.175
24	3	GATS4p BELe4 BLTA96	94.81	91.45	88.99	40.33	73.1	0.176
25	3	GATS8e GATS4p BLTD48	94.84	91.43	89.05	39.43	73.52	0.175
26	3	GATS4p BELe4 BLTF96	94.79	91.42	89.44	40.02	72.84	0.176
27	3	GATS4p H8p BLTA96	95.09	91.39	89.02	22.32	77.46	0.171
28	3	GATS8e GATS4p BLTF96	94.81	91.39	88.3	39.37	73.12	0.176
29	2	GATS4p BLTA96	93.75	91.35	88.93	19.11	97.55	0.185
30	2	GATS4p BLTD48	93.75	91.32	90.32	18.78	97.51	0.185

31	2	GATS4p MLOGP	93.73	91.31	89.5	19.13	97.25	0.186
32	3	GATS4p H8p BLTF96	95.02	91.26	89.51	22.03	76.29	0.172
33	2	GATS4p BLTF96	93.7	91.25	89.18	18.38	96.66	0.186
34	3	GATS4p E1s BLTA96	94.37	91.21	88.57	33.53	67.07	0.183
35	3	GATS4p MLOGP2 BLTA96	94.18	91.2	86.96	52.63	64.78	0.186
36	3	GATS8e GATS4p MLOGP2	94.68	91.16	88.87	38.52	71.14	0.178
37	3	GATS4p MLOGP MLOGP2	94.1	90.98	85.42	52.63	63.8	0.187
38	3	GATS4p RDF140p BLTA96	93.89	90.86	87.95	32.75	61.5	0.191
39	3	GATS4p R5u MLOGP	94.12	90.79	87.71	16.44	63.97	0.187
40	3	GATS4p RDF140p BLTF96	93.86	90.78	87.61	32.42	61.16	0.191
41	3	GATS4p MLOGP BLTF96	93.86	90.71	88.4	53.3	61.09	0.191
42	3	GATS4p MLOGP2 BLTF96	93.97	90.67	86.39	52.53	62.35	0.19
43	3	GATS4p RDF035m BLTD48	93.88	90.65	88.13	29.31	61.41	0.191
44	3	GATS4p BLTF96 BLTD48	93.93	90.61	86.16	53.24	61.9	0.19
45	3	GATS4p RDF135u BLTA96	93.75	90.59	87.2	28.21	60.04	0.193
46	3	GATS8e GATS4p Tu	94.63	90.56	88.49	37.44	70.43	0.179
47	3	GATS4p RDF135u BLTD48	93.76	90.54	86.95	27.93	60.05	0.193
48	3	GATS4p BELv4 MLOGP2	94.52	90.53	88.71	40.03	69	0.181
49	2	GATS4p MLOGP2	93.19	90.52	89.18	14.86	89	0.193
50	3	HVcpx GATS4p BLTD48	93.95	90.51	87.52	42.87	62.14	0.19

En la tabla 13, se presentan los modelos obtenidas en el QSAR para el *Escherichia coli*, y los parámetros calculados para cada uno de los modelos.

5. QSAR de difluoro quinolonas con actividad contra *Pseudomona aeruginosa* (cepa.- 308)

Las Pseudomonas son bacilos rectos o ligeramente curvados, Gram-negativos, oxidasa positivos, aeróbicos estrictos aunque en algunos casos pueden utilizar el nitrato como aceptor de electrones. El catabolismo de los glúcidos se realiza por la ruta de Etner-Doudoroff y el ciclo de los ácidos tricarbónicos. Algunos miembros del género son psicrófilos. Otros sintetizan sideróforos fluorescentes de color amarillo-verdoso con gran valor taxonómico. Es común la presencia de plásmidos. No forman esporas. La importancia de estudiar compuestos activos contra este tipo de bacterias radica en que es una bacteria que provoca infecciones intrahospitalarias y presenta resistencia o baja sensibilidad a la mayoría de los antibióticos incluso a las mismas quinolonas. El estudio QSAR se describe a continuación:

El modelo seleccionado para *Pseudomona aeruginosa* está constituido por 3 descriptores, expresados en la Ecuación 11, de la Hoja de resultados 4.

En donde:

- **BELv6** (lowest eigenvalue n. 6 of Burden matrix / weighted by atomic van der Waals volumes).- Descriptor que evalúa la similaridad o diversidad de una molécula, el cálculo proviene de las aproximaciones de Burden y está basado en el volumen de las fuerzas de van der Waals.²⁰
- **BEHp6** (highest eigenvalue n. 6 of Burden matrix / weighted by atomic polarizabilities).- Descriptor que evalúa la similaridad o diversidad de una molécula, el cálculo proviene de las aproximaciones de Burden y está basado en la polarizabilidad de los átomos.²⁰
- **Mor25e** (3D-MoRSE - signal 25 / weighted by atomic Sanderson electronegativities).- Los descriptores MoRse usan las coordenadas espaciales de la estructura en 3D, esta coordenadas son transformadas en un curva de difracción electrónica, esta curva es basada en propiedades moleculares en este caso es utilizada la electronegatividad de Sanderson.²⁰

Por su definición, es de esperarse que los descriptores **BELv6** y **BEHp6** estén, pues ambos fueron calculados con una matriz de Burden aunque están basados en propiedades diferentes.

Las actividad calculada y experimental en la Hoja de resultados 5 (Tabla 14 Y Gráfica 5), poseen una pequeña discrepancia entre si, lo cual concuerda con la $R^2 = 0.936$ y $Q^2 = 89.27$. En la misma tabulación se observan los valores de los descriptores para cada una de las moléculas así como la actividad que presenta cada una de ellas:

Hoja de resultados 5 *Pseudomonas aeruginosa* (cepa.- 308)

Ecuación 11 **Log (1/Act.) = (4.56538) BELv6 + (-4.58520) BEHp6 + (0.36394) Mor25e + (10.74877)**

N= 16 R²= 0.936 F= 59.41 s= 0.08 Q² = 89.27

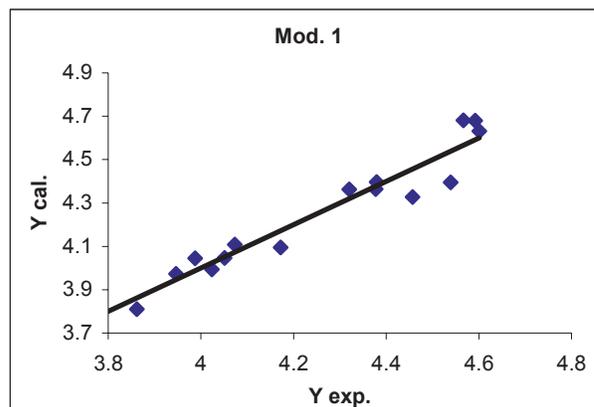
Tabla 14

ID	Log de la Actividad Experimental	Log de la Actividad calculada	Descriptores		
			BELv6	BEHp6	Mor25e
1	3.973	3.946	1.074	2.616	0.794
2	3.994	4.023	1.117	2.644	0.821
3	4.631	4.6	1.272	2.708	1.268
4	4.327	4.456	1.272	2.708	0.873
5	4.362	4.32	1.272	2.748	1.001
6	4.047	4.051	1.272	2.817	1.131
7	4.044	3.987	1.202	2.757	1.079
8	4.396	4.379	1.325	2.814	1.331
9	4.681	4.566	1.325	2.763	1.202
10	4.108	4.073	1.324	2.844	0.88
11	4.364	4.377	1.272	2.718	0.78
12	4.095	4.171	1.272	2.809	1.362
13	3.809	3.861	1.32	2.912	1.205
14	4.679	4.591	1.322	2.716	0.717
15	4.394	4.538	1.325	2.744	0.887
16	3.82	3.785	1.324	2.895	0.733

Tabla No. 15

Matriz de correlación			
	BELv6	BEHp6	Mor25e
BELv6	1		
BEHp6	0.70310691	1	
Mor25e	0.24203238	0.34879067	1

Grafico 5



Hoja resultados del modelo 5. En esta hoja se encuentran recabada toda la información del modelo, la hoja de resultados cuenta con la ecuación y sus respectivos parámetros estadísticos, la matriz de correlación de los descriptores (tabla No.14), la matriz de las actividades con el valor de los descriptores y el grafico de la regresión lineal.

El descriptor **BELv6** representa el volumen del sustituyente. Se aprecia que la actividad es beneficiada con el incremento del volumen. En algunos compuestos como en **3**, **9** y **14** el volumen es el más alto y las actividades son las más altas, mientras que en los compuestos **1** y **2** el volumen es muy bajo y la actividad también lo es.

La polarizabilidad en las moléculas es parametrizada por **BEHp6**. El análisis indica que el compuesto es más activo cuando presenta una polarizabilidad baja como en el caso de **3**, **9** y **14**. Al incrementar la polarizabilidad, la actividad es afectada negativamente como se puede apreciar en los compuestos **13** y **16**.

El descriptor **Mor25e** mide la electronegatividad atómica y afecta positivamente la actividad de las quinolonas de la serie, el incrementarse la electronegatividad en la molécula, la actividad se ve beneficiada como en el caso de compuestos **9** y **3**. Al disminuir la electronegatividad, la actividad disminuye como en el caso de las moléculas **1** y **16**.

A continuación se presentan los 50 modelos calculados para *Pseudomona aeruginosa*, del cual fue seleccionado el modelo anteriormente descrito:

Tabla 16.- Modelos para *Pseudomona aeruginosa*

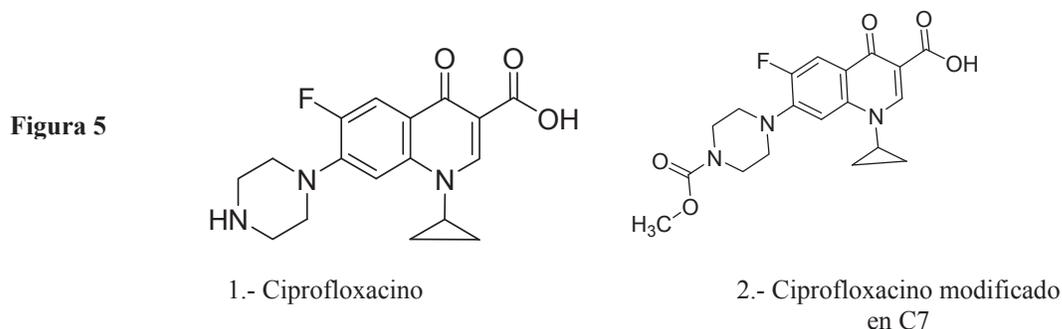
ID	Size	Models	R2	Q2	Q2boot	Kx	F	s
1	4	RDF035u Gu HATS7v RPCG	97.62	95.1	93.31	49.77	112.62	0.052
2	4	RDF035u Gu HATS7v Qneg	97.64	95.02	93.07	49.06	113.63	0.052
3	4	RDF035e Gu HATS7v RPCG	97.34	94.54	92.64	49.92	100.58	0.055
4	4	RDF035u Mor01p Gu HATS7v	97.01	94.53	92.24	50.5	89.16	0.059
5	4	RDF035e Gu HATS7v Qneg	97.28	94.32	92.18	49.22	98.51	0.056
6	4	RDF035u RDF110m RDF155v RDF100e	97.18	94.11	0	52.28	94.61	0.057
7	4	RDF035u RDF100u RDF110m RDF155v	97.22	94	0	52.46	96.17	0.057
8	4	RDF035u RDF110m RDF100e RDF155e	97.27	93.96	0	52.55	98.11	0.056
9	4	RDF035u Mor01u Gu HATS7v	96.75	93.91	91.27	50.64	81.91	0.061
10	4	BEHp6 RDF035e Gu HATS7v	96.5	93.82	90.64	50.48	75.75	0.063
11	4	BEHp6 RDF035u Gu HATS7v	96.42	93.74	90.94	50.61	74.15	0.064
12	4	RDF035u RDF100u RDF110m RDF155e	97.3	93.69	0	52.72	99.09	0.056
13	4	RDF035e Mor01p Gu HATS7v	96.63	93.67	90.75	50.31	78.81	0.062
14	4	RDF035e Gu Du HATS7v	96.91	93.55	89.96	36.92	86.12	0.06
15	4	W3D RDF035u Gu HATS7v	96.51	93.47	90.41	49.29	75.98	0.063
16	4	AGDD RDF035u Gu HATS7v	96.3	93.04	90.13	49.75	71.48	0.065
17	4	QYYm RDF035u Gu HATS7v	96.08	92.84	89.15	49.7	67.4	0.067
18	4	QYYv RDF035u Gu HATS7v	96.05	92.81	89.62	49.61	66.81	0.067
19	4	RDF035e Mor01u Gu HATS7v	96.26	92.81	89.75	50.53	70.77	0.066
20	4	RDF035u Gu Du HATS7v	96.69	92.79	89.81	36.67	80.42	0.062
21	3	BELv6 BEHp6 Mor25e	93.65	89.27	87.6	44.91	59	0.082

22	4	RDF035u RDF110m RDF155e R8m	93.35	87.03	0	51.95	38.62	0.087
23	3	RDF035u RDF110m RDF155e	92.26	86.36	0	46.64	47.69	0.09
24	3	RDF035u RDF110m RDF155v	91.98	85.97	0	46.38	45.9	0.092
25	4	G(O..F) RDF035e Gu HATS7v	93.43	85.88	81.75	35.81	39.11	0.087
26	3	RDF035e Gu HATS7v	92.48	85.86	84.84	33.65	49.22	0.089
27	4	RDF035e Gu HATS7v R8m	92.52	84.67	78.31	53.1	34.03	0.093
28	3	RDF035u Gu HATS7v	91.8	84.66	83.14	33.73	44.78	0.093
29	4	RDF035u RDF110m RDF155v Gu	92.82	84.27	0	50.14	35.54	0.091
30	4	RDF035u RDF035e Gu HATS7v	92.75	83.69	79.8	51.3	35.17	0.091
31	4	BELp6 RDF035e Gu Du	90.81	81.76	68.53	42.14	27.16	0.103
32	2	BELv6 BEHp6	86.14	80.47	79.02	70.31	40.41	0.116
33	2	RDF035e R8m	84.31	77.43	75.87	72.17	34.92	0.123
34	3	RDF035e HATS7v R8m	88.12	76.09	73.64	56.13	29.68	0.112
35	3	BELv6 RDF035u RDF155e	85.63	74.6	0	30.26	23.83	0.123
36	3	G(O..F) RDF035e R8m	84.31	74.5	69.77	43.8	21.49	0.129
37	4	RDF035u RDF035e HATS7v R8m	88.58	73.03	60.39	64.92	21.33	0.115
38	3	BELv6 BEHp6 RDF155e	86.65	71.06	0	49.39	25.95	0.119
39	3	SNar RDF035u R8m	83.25	70.35	62.5	67.45	19.88	0.133
40	3	BEHp6 BELp6 DP04	82.31	68.5	63.96	79.94	18.61	0.136
41	4	BELv6 BEHp6 RDF110m RDF155e	86.77	66.59	0	62.38	18.03	0.123
42	4	BELv6 QYYm RDF035u HATS7v	84.95	65.56	52.2	68.03	15.52	0.131
43	4	SNar Dz RDF035u R8m	83.51	64.09	47.85	73.53	13.93	0.138
44	3	BELv6 RDF035u Mor25e	80.44	62.77	56.72	32.72	16.45	0.143
45	3	DP03 DP04 RDF155e	73.52	62.41	0	69.88	11.1	0.167
46	2	BELv6 RDF035e	73.94	60.56	55.55	39.97	18.44	0.159
47	2	BELv6 RDF035u	74.19	60.55	55.45	40.93	18.68	0.158
48	3	BELv6 G2s GVWAI-50	74.7	59.79	49.66	34.65	11.81	0.163
49	1	Mor32p	56.14	43.31	45.99	0	17.92	0.199
50	1	Mor32v	53.32	39.4	42.89	0	15.99	0.205

En la tabla 16, se presentan los modelos obtenidos en el QSAR para el *Escherichia coli*, y los parámetros calculados para cada uno de los modelos.

6.-Predicción

Una vez obtenidos los modelo y valorarlos se prosiguió a predecir la actividad ya determinada para el ciprofloxacino y para compuestos previamente sintetizados (Compuesto 1 y 2, Figura.5):



Esta molécula fue sintetizada como proyecto de investigación de tesis, por uno de los colaboradores del laboratorio ³², quien realizó las pruebas cualitativas de actividad.

Los datos de actividad calculada se encuentran en la tabla 17:

Tabla 17

Molécula	Descriptores			Actividad calculada
1	SHP2	G(O..F)		<i>S. pyogenes</i>
	0.399	41.435		6.814
2	0.365	35.513		46.3
1	X5sol	Gu	H8m	<i>S. aureus</i>
	6.793	0.187	0.04	6.61
2	7.412	0.186	0.09	8.11
1	GATS8e	GATS4p	VRm2	<i>E. coli</i>
	0.871	1.118	6.126	4.52
2	0.982	1.083	6.376	5.3
1	BELv6	BEHp6	Mor25e	<i>P. aeruginosa</i>
	1.272	2.779	0.639	4.04
2	1.272	2.817	0.939	3.98

Experimentalmente solo fueron utilizadas 2 cepas: *S. aureus* y *E. coli*, ambas quinolonas mostraron halo de inhibición pero en el caso de la Molécula 2 el halo fue más grande que en el ciprofloxacino, si se comparan las actividades experimentales con las calculadas por los modelos, la predicción es correcta, caso que se podrá constatar en el momento en que sean realizadas las pruebas cuantitativas de la actividad.

Conclusiones

Conclusiones.-

- La técnica basada en algoritmos genéticos para la elaboración del análisis QSAR, reduce tiempo en la elaboración de dicha metodología. Esta técnica permite el uso de un gran número de descriptores físico-químicos, evitando así la limitación del análisis QSAR e incrementando la exactitud de los modelos obtenidos.
- Los modelos predictivos para las bacterias *Streptococcus pyogenes*, *Staphilococcus aureus*, *Pseudomona aeruginosa* y *Escherichia coli*. Fueron comprobados estadísticamente con muy buenos resultados.
- Las propiedades fisicoquímicas que afectan la actividad bactericida de las quinolonas, pueden utilizarse para proponer difluoroquinolonas con un óptimo de actividad biológica.

Referencias

Referencias.-

1. Software **Hiper Chem 6.03** for Windows; *HiperCube Inc., 2000*
2. Software **Dragón versión 5.3 profesional** para Windows; *Copyright © 2003-2005 by Talete srl, Milano, Italy*
3. Software **Moby Digs Versión 1.0 profesional**; *Copyright © 2004-2005 by Talete srl, Milano, Italy*
4. Software **Microsoft Excel 2000**; *Copyright © Microsoft Corporation 1985-2001.*
5. Software **Viewer pro; Version 5.0**; *Copyright © 2000 by Accelrys Inc.*
6. Software **ChemDraw Pro Version 8.0**; April 23, 2003; © 1985-2003 *Cambridgesoft Corporation.*
7. Lee, K. W.; Kwon, S. Y.; Hwang, S.; Lee, J.; and Kim, H.; **Quantitative Structure-Activity Relationships (QSAR) Study on C-7 Substituted**; *Bull. Korean Chem. Soc. 1996, Vol. 17, pags.- 147-152.*
8. Potemkin, V. A.; Grishina, M. A; Belik, A. V. ; and Chupakhin, O. N.; **Quantitative Relationship Between Structure and Antibacterial Activity of Quinolone Derivatives**; *Pharmaceutical Chemistry Journal 2002, Vol. 36, pags.- 22-25.*
9. Pastor, M. y Alvarez-Builla, J.; **Técnicas QSAR en Diseños de Fármacos**; *Departamento de química Orgánica. Universidad de Alcalá; E-28871 Alcalá de Henares, Madrid.*
10. Bambeke, F. V.; Michot, J. M.; Eldere, J. Van; and Tulkens, P. M.; **Quinolones in 2005 an update**; *Clinical Microbiology and Infection, April 2005, Vol. 11, pags.- 256-280.*
11. Gonzalbes, R.; Brun-Pascaud, M.; García-domenech, R.; Gálvez, J.; Girard, P. M.; Doucet, J. P.; and Derouin, F.; **Prediction of Quinolone Activity Against Mycobacterium avium by Molecular Topology and Virtual Computational Screening**; *Antimicrobial Agents and Chemotherapy, Oct. 2000, Vol. 44, pags.- 2764-2770.*
12. Gonzalbes, R.; Brun-Pascaud, M.; García-domenech, R.; Gálvez, J.; Girard, P. M.; Doucet, J. P.; and Derouin, F.; **Anti-Toxoplasma Activities of 24 Quinolones and**

- fluoroquinolones In Vitro: Prediction of Activity by Molecular Topology and Virtual Computacional Tecniqúe; *Antimicrobial Agents and Chemotherapy, Oct 2000, Vol. 44*, pags.- 2771-2776.**
13. Topliss, J. G.; **Some observations on classical QSAR; *Perspectives in Drug Discovery and Design, 1 (1993)*, pags.- 253-268**
 14. Stanton, D. T.; **On the Physical Interpretation of QSAR Models; *J. Chem. Inf. Comput. Sci. 2003, Vol. 43*, pags.- 1423-1433.**
 15. Renau, T. E.; Sanchez, J. P.; Gage, J. W.; Devar, J. A. ; Shapiro, M. A.; Gracheck, S. J.; and Domagala, J. M.; **Structure-Activity Relationships of the Quinolone Antibacterials against Mycobacteria: Effect of Structural Changes at N-1 and C-7; *J. Med. Chem. 1996, Vol. 39*, pags.- 729-735.**
 16. Selassie, C. D.; **Hitory of Quanttitative Structure-Activity Relationships; *Burger's Medicinal Chemistry and Drug Discovery: January 15, 2003*.**
 17. Hardman, J. G.; Limbird, L. E.; Gooman Gilman, A.; **Las bases farmacológicas de la terapéutica; *Editorial McGraw-Hill; 10° Edición 2003, Vol. I y II*, pags.- 35-38, 1197-1203.**
 18. Box, G. E. P.; Huter, W. G.; Hunter, J. S.; **Estadística para investigadores; *Editorial Reverté S.A.; 1° Edición 1999 Mexico*; pags.- 463-524.**
 19. Ramírez, Y.; Rodríguez, M.; Herrera J. L.; **Estadística II; *Universidad Tecnológica de Mexico; Colección EJE Matemáticas FACS; 1° Edición Abril 2003*; pags.- 235-304.**
 20. Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M.; **Manual de Dragón para Windows versión 5.3 prefecional, 2005**
 21. Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M.; **User Manual Moby Digs Versión 1.0, 2004-2005.**
 22. Herrera Busio R.; **Curso de modelado molecular; 2003, *Universidad Michoacana de San Nicolás de Hidalgo; Instituto de investigaciones Químico Biológicas*.**
 23. Agrawal, V. K.; Bano, S.; Mathur, K. C.; and Khadikar, P.V.; **Novel application of Wiener vis-á-vis Szeged indices: Antitubercular activities of quinolones; *Proc. Indiana acad. Sci. (Chem, Sci.), April 2000, Vol. 112*, pags.- 137-146.**
 24. **Algoritmos genéticos; <http://eddyalfaro.galeon.com/geneticos.html>**

25. Arámbula C., F.; Pérez M., M.; **Algoritmos Genéticos**;
<http://tigre.aragon.unam.mx/geneticos/>
26. Taléns-Visconti, R.; Garrigues, T.M.; y Cantón, E.; **Mecanismos de resistencia bacteriana a las quinolonas**; *Revista Española de Quimioterapia: Marzo 2002; Vol. 15*
27. Gutiérrez-Zufiaurre, N.; **Relación entre estructura, actividad y efectos adversos de las Quinolonas**; *Rev. Esp. Quimioterap, septiembre 2004, Vol. 17*, pags.-234-243.
28. Escalona, J.C.; Carrasco, R.; Padrón, J. A.; **Introducción al diseño de Fármacos**;
Folleto para la docencia de la asignatura de Farmacia, Universidad de Oriente.
29. Elguero, J.; Gutthann, S.; **Bases de datos en farmacología y terapéutica**;
Fundación Dr. Antonio Esteve.
30. Alós, J. I.; **Quinolonas**; *Enferm Infecc Microbiol Clin 2003, 7-02-2003*, pags.-261-267.
31. Gutiérrez de Terán C., H.; **Modelización molecular de los receptores de adenosina y sus ligandos en el marco de diseño de Fármacos asistido por ordenador**; *Tesis doctoral, Febrero del 2004, Barcelona, España*, pags.- 9-10, 21-23 y 25-38.
32. Aguirre Serrano M. A.; **Relación estructura-actividad de derivados del ciprofloxacino**; 2006; *Tesis Químico Fármaco Biólogo; Universidad Michoacana de San Nicolás de Hidalgo.*