

Capítulo 4. METODOLOGÍA

4.1 HERRAMIENTAS *IN SILICO*

Se utilizó el banco de datos *Brookhaven Protein Data Bank* (PDB) para la obtención de la estructura tridimensional del receptor y su ligando.

Respecto al *software* utilizado, fueron cuatro los programas computacionales empleados:

1. Chem Office. Para representar esquemáticamente la estructura bidimensional de los ligandos.
2. SYBYL 6.7.1 Tripos Inc. Para la construcción tridimensional de los ligandos con una orientación espacial independiente del receptor.
3. ADT AutoDock Tools 4.0. Para la ubicación espacial adecuada de los ligandos dentro del sitio de unión del receptor.
4. MOE, Molecular Operating Environment de Chemical Computing Group Inc. Programa para el desarrollo del análisis al farmacóforo y la elaboración del estudio QSAR.

Se utilizó paquetería de Microsoft Office para la elaboración del proyecto, pero sin interés relevante para el análisis computacional.

4.2 OBTENCIÓN Y ELABORACIÓN DE LAS MOLÉCULAS

La estructura del receptor se obtuvo por medio del PDB, la cual es una fuente de datos a nivel mundial que contiene estructuras tridimensionales de biomoléculas determinadas por cristalografía. Cada estructura posee un ID asignado para identificarlas fácilmente. Del PDB se obtuvo la estructura tridimensional de la MAPk p38 α con su ligando SB203580 conocido por el ID 1A9U. Los 40 ligandos análogos de la SB203580 utilizados para el estudio, fueron sintetizados y probados por los investigadores de la Facultad de Farmacia de la Universidad de Tübingen, Alemania. De esa fuente se tomaron los siguientes datos:

- Estructura bidimensional.
- Nombre químico.
- pIC50 (-Log de la concentración inhibitoria 50).

La estructura de los ligandos se muestran en el **Anexo 1**.

Al contar con la estructura bidimensional de los ligandos, se pudo iniciar la construcción de las moléculas tridimensionalmente por medio del programa SYBYL 6.7.1 Tripos Inc.

Se tomó la decisión de construir los análogos de SB203580 a partir de la estructura que éste tiene en el modelo 1A9U para aprovechar su conformación. Este tipo de procedimiento provee de dos ventajas: uno es que se aprovecha la conformación

determinada en el modelo 1A9U que ya ha sido comprobada, y dos que la ramificación de los ligandos pueden someterse a cambios de orientación manual que teóricamente son favorecidos por las interacciones con el receptor. Como los análogos poseen un tronco común con el SB203580, la construcción de los análogos consistió en agregar aquellas ramificaciones que la SB203580 no poseía, o borrar aquellos grupos que la SB203580 tenía y los análogos no.

4.3 PREPARACIÓN DEL *DOCKING*

Con *docking* se refiere a la determinación de la conformación que los ligandos ocupan en el receptor. Teóricamente, la conformación que ocupa un ligando en su receptor será la de menor energía, ya que las interacciones que suceden en el interior receptor-ligando, son suficientes para mantener una conformación que sería inestable si el medio que rodeara al ligando fuera otro distinto al receptor [*Shattuck; 2004*]

Para esta determinación se utilizó el programa ADT AutoDock Tools 4.0.

Antes de iniciar el *docking*, AutoDock le proporciona cargas atómicas parciales a cada átomo. También distingue entre carbonos alifáticos y aromáticos. [*Huey & Morris; 2003*]
Antes de que el ligando se muestre en la pantalla de AutoDock, ADT la prepara para la iniciación. Este proceso involucra varios pasos:

- 1) ADT revisa y agrega los hidrógenos no polares, a menos que se le indique lo contrario.
- 2) ADT detecta si los ligandos ya poseen cargas o no. Si aún no las tienen ADT determina si el ligando es un péptido, revisando si los residuos aparecen en un conjunto estándar de los 20 aminoácidos más comunes. Si los residuos son aminoácidos, el programa les proporciona cargas Kollman a los ligandos. Si no, ejecuta las cargas Gasteiger, la cual requiere que los ligandos posean todos los hidrógenos, polares y no polares.
- 3) ADT determina los carbonos planares cíclicos al calcular el ángulo entre los carbonos adyacentes. Si el ángulo es menor de 7.5 grados, entonces se nombrará a esos carbonos como “A” de aromáticos.

Después de esos pasos ADT define el *Rigid Root* (conjunto rígido de átomos) y lo marca en una esfera verde. La porción rígida de una molécula está compuesta por un átomo raíz y todos los átomos conectados a él por enlaces no rotables.

Lo siguiente es la determinación de los enlaces rotables. Sólo los enlaces simples pueden rotarse. Los enlaces dobles, aromáticos o amido se consideran no rotables. El programa despliega la imagen de todos los enlaces del ligando y colorea de rojo los enlaces que no pueden rotarse, de púrpura los enlaces que son rotables pero que han sido inactivados para la disminución de cálculos o por falta de importancia, y de verde los rotables y activados. Esto es útil ya que permite establecer el número total de enlaces activos especificando si se quieren enlaces activos que muevan pocos átomos o que muevan

muchos. Después de estas variantes, el ligando editado se guarda en un archivo *pdbq* para proceder a los siguientes pasos del *docking*.

4.4 *AUTOGRID*

Lo que sigue es la preparación del *grid*, que representa las coordenadas donde se establecen los parámetros en los que el ligando puede moverse. El parámetro *grid* requiere que a cada elemento de los ligandos se le asigne un mapa, el cual establece el tipo de mapas que se va a computar, la localización y extensión de esos mapas y especifica los parámetros de energía potencial de los mismos. En general, un mapa se calcula por cada elemento en el ligando además de un mapa electrostático. Con esto, lo que se propone es darle a cada elemento un valor de energía, que más adelante cambiará conforme la conformación cambie. Este cambio de energía será positivo o negativo dependiendo de las interacciones que se formen entre el ligando y receptor.

Después de establecer los tipos de mapas, se procede a establecer el *grid* esto es una caja, *box*, que delimita la región que se va a analizar. La caja del *grid* debe ser suficientemente grande para permitir que el ligando rote libremente, aún cuando el ligando está en su conformación más extendida. Para establecer el *grid*, se escogió el ligando más voluminoso y se le colocó en el sitio de unión del receptor, el cuál ya está identificado en el 1A9U. De esa forma todos los ligandos, por ser más pequeños, podrán rotar en el *grid*. El *grid* marca una región límite en donde los cálculos de *docking* deben realizarse y no deben salirse de ella. El *grid* de cada ligando se guarda en un archivo con extensión *.gpf* y ese

archivo se corre en la ventana de AutoGrid, comando del programa ADT, para que se asignen los mapas, y el resultado se guarda en un archivo con extensión .glg que se utilizará en el *docking*.

4.5 DOCKING

Antes de correr la ventana AutoDock, comando propio de ADT, es necesario definir algunos parámetros como cuál es el centro y el número de torsiones, dónde empezar el ligando, qué algoritmo de *docking* emplear y cuántas series de *docking* realizar. AutoDock contiene cuatro diferentes algoritmos de *docking*: SA, que es el método original de Monte Carlo; GA, el algoritmo genético tradicional darwiniano; LS, de búsqueda local; y GA-LS que es un algoritmo genético híbrido con búsqueda local. El GA-LS también se le conoce como algoritmo genético lamarckiano o LGA. En el presente trabajo se utilizó el algoritmo GA-LS.

Cada método de búsqueda tiene sus propios conjuntos de parámetros. Y estos deben definirse antes de correr el experimento de *docking*. En el GA-LS el número de evaluaciones de energía y el número de generaciones afecta la duración del *docking*.

El *docking* necesita de 5 aspectos:

- a) La macromolécula: la estructura tridimensional de la MAPk p38 α obtenida del PDB.

- b) El ligando: que ya tiene establecidos los tipos de átomos, su centro, su número de torsiones, y su número de grados de torsión de libertad. Esta operación se hace con cada uno de los 40 ligandos.
- c) El parámetro del algoritmo genético: para éste se realizaron 100 corridas (*runs*), que son el número de conformaciones que se van a registrar, y tres sobrevivientes (*survive*).
- d) Correr la ventana de AutoDock para realizar el cálculo.
- e) Guardar el archivo con extensión .dpf.

El *docking* proporciona los resultados en forma de gráficos de energía (en el eje X) contra conformaciones posibles (en el eje Y) llamados *clusters*. Se registran en forma de columnas donde se acoplan en una misma columna conformaciones con semejante posición y energía, si tienen una posición diferente o energía diferente se acomodan en otra columna. Una columna con muchas conformaciones representa la columna *most*, donde se encuentra la conformación que más veces se repitió. La columna con menor energía es la columna *best*. En ocasiones la columna *best* era la misma que la *most*. Después de correr *docking*, se analizaron los clusters y de las gráficas de los mismos se escogió el ligando con el de menor energía para realizar la prueba de QSAR.

4.6 FARMACÓFORO

El análisis del farmacóforo consiste en encontrar grupos consenso entre todas las moléculas, es decir, características que todos los ligandos compartan, se realizó en el

programa MOE. Esto con la finalidad de descubrir las características a considerar para elegir los descriptores de QSAR y del estudio SAR.

Lo primero es buscar la opción MOE | *File* | *New* | *Pharmacophore Query*. En la ventana de MOE, aparecen esferas coloreadas (puntos de anotación) en cierta posición en el ligando que denotan las propiedades potenciales de interés como se muestra en la **Figura 4.1**. Los puntos de anotación son códigos coloreados dependiendo del tipo de propiedad (ejemplo: verde brillante es aromático, púrpura es donador).

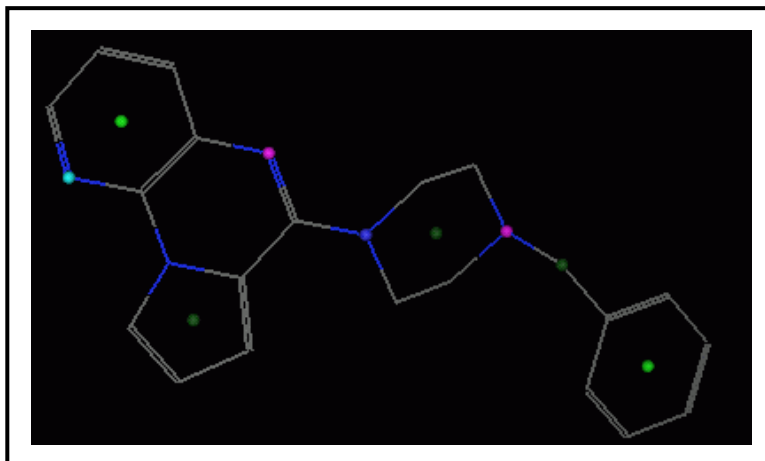


Figura 4.1. Imagen ilustrativa de los grupos identificados en el farmacóforo. Obtenido del programa MOE.

Un esquema de farmacóforo define el conjunto de atributos usados para construir puntos de anotación de ligandos etiquetados que pueden utilizarse en la búsqueda, ver **Figura 4.2** y **Figura 4.1**. El esquema utilizado es PCH (Polaridad-Carga-Hidrofobicidad). El esquema define como cada ligando de las bases de datos es clasificado.

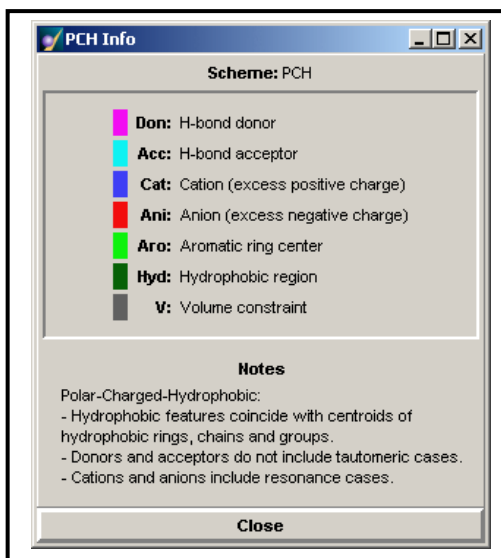


Figura 4.2. Esquema PCH con códigos por colores de MOE. Obtenido del programa MOE.

Bajo este tipo de opciones se pueden deducir las propiedades principales de cada ligando y tener una idea del tipo de descriptores que se deben utilizar.

4.7 QSAR

Esta parte del trabajo se realizó en el programa MOE.

Antes de correr QSAR, a los ligandos, que ya habían pasado por ADT, se les asignaron las cargas parciales para una correcta descripción de los parámetros. Para la

asignación de cargas de los elementos de la molécula es necesario establecer antes que nada el campo de fuerza (*force field*) por lo que se utilizó MMFF94s, un comando de MOE para moléculas pequeñas (*Molecular Force Field* desarrollado en 1994), y después se calcularon las cargas con la opción de *potential charges*. A la estructura de la proteína obtenida de PDB también fue procesada para determinar su campo de fuerza, para esto se utilizó el comando AMBER especializado para proteínas.

MOE posee una amplia lista de descriptores que se pueden utilizar, ver **Figura 4.3**:

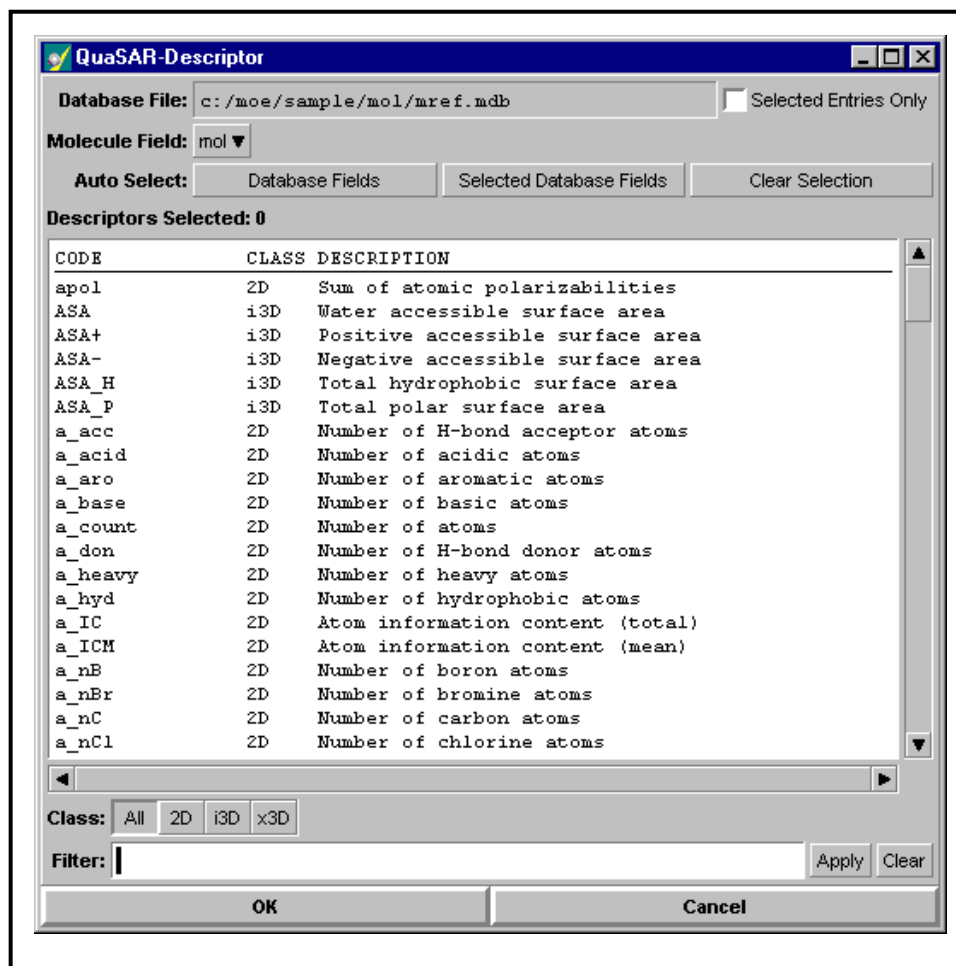


Figura 4.3. Pantalla de MOE con los descriptores que pueden utilizarse para realizar estudio QSAR. Obtenido del programa MOE.

Antes de realizar el modelo, los ligandos son divididos en dos conjuntos, uno contiene el 90% de los ligandos y se utiliza para elaborar el modelo QSAR (*training set*), mientras que el otro 10% es utilizado para su validación (*test set*).

Los descriptores se dividen en varias categorías dependiendo del parámetro que van a calcular. Hay descriptores de propiedades fisicoquímicas, de cuenta del número de átomos y enlaces, de conectividad (Kier & Hall), de adyacencia y distancia, de propiedades del farmacóforo y descriptores de cargas parciales. [*Shattuck; 2004*]

Para elegir los correctos se realizaron los cálculos de regresión lineal simple de cada uno de ellos, seleccionando los que poseían las R^2 más altas. No obstante, este tipo de procedimientos puede ser engañoso, ya que muchos parámetros pueden no mostrar una relación lineal con la actividad, pero podrían mostrar una relación más lineal si se realiza la regresión con más parámetros, ya sea bidimensional, tridimensional o n-dimensional, por lo que se prefirió realizar el análisis de farmacóforo para tener una idea más amplia de los descriptores que se desean utilizar.

Una vez elegidos los descriptores se procedió a realizar el modelo lineal. Para eso se utilizó la opción de MOE de *QuaSAR Model*, el cual calcula la desviación estándar y el coeficiente de regresión de la regresión lineal múltiple, **ver Figura 4.4**.

4.8 VALIDACIÓN

Cuando el modelo ya está listo, hay que validarlo. Para eso se necesitan los valores de actividad calculados por el modelo (\$PRED) que son los valores que la ecuación predice, la diferencia entre el valor de la actividad del modelo y la experimental (\$RES) o residuales, y la diferencia absoluta entre el valor del modelo y la actividad, dividida por la raíz cuadrada del error cuadrado de la media del conjunto de datos (\$Z-SCORE) valor z, ver Figura 4.5.

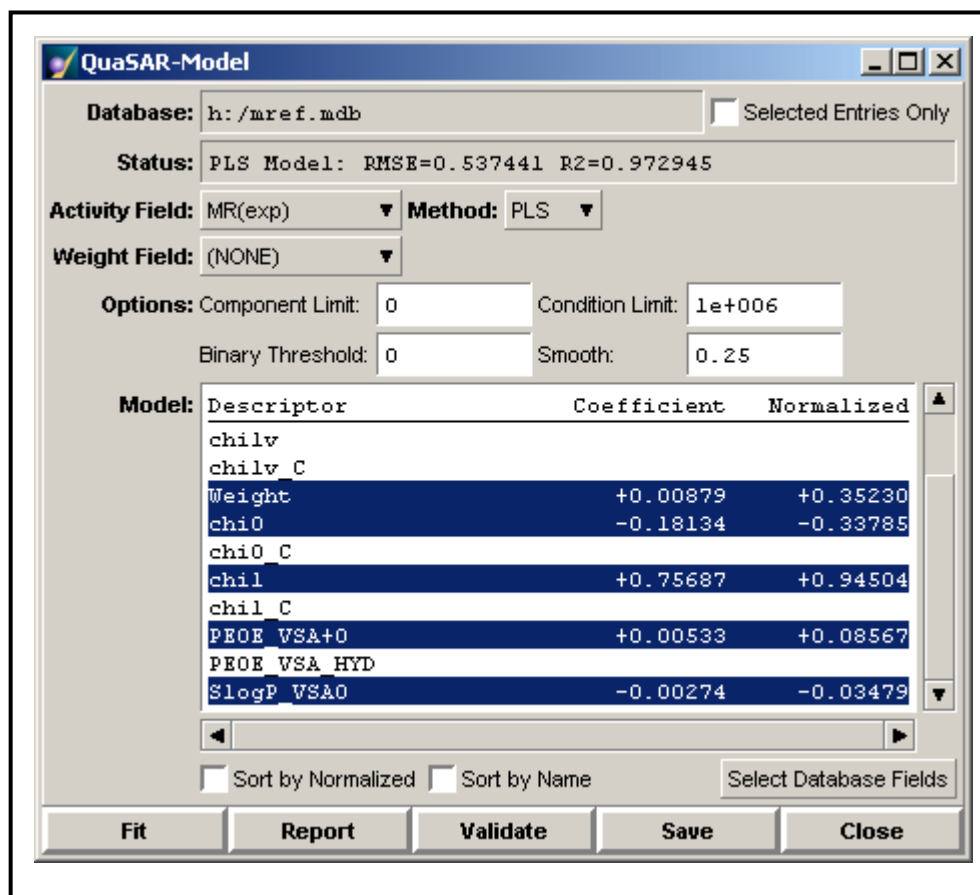


Figura 4.4. Ventana de selección de descriptores para la elaboración del modelo QSAR mediante el programa MOE. Obtenido del programa MOE.

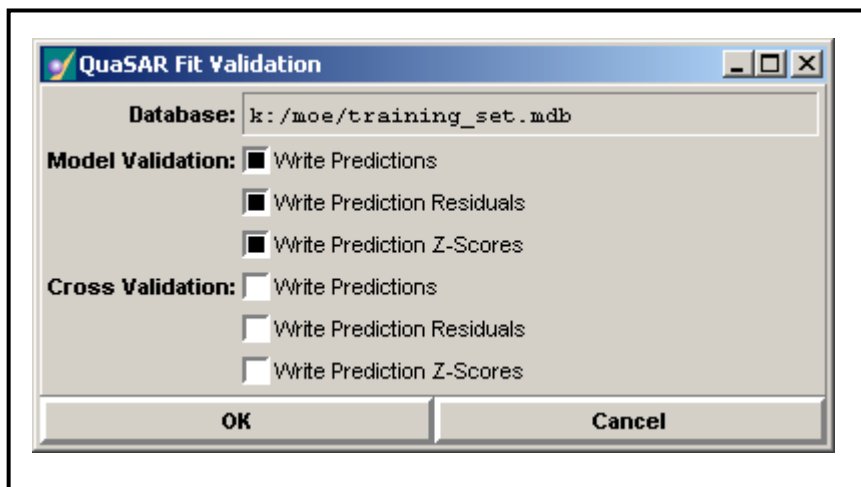


Figura 4.5. Selección de valores Z para la evaluación del modelo. Obtenido del programa MOE.

Se utilizaron 4 métodos de validación. *Leave One Out*, *Cross Validation*, Valor Z, χ^2 . Los primeros 3 pueden ser calculados en el programa MOE, la evaluación por el método chi cuadrada (χ^2) fue diseñado por nosotros mismos.

4.8.1 *Leave One Out* (LOO)

El método LOO se refiere a sacar cada uno de los 39 ligandos (*training set*) usados en el modelo de la ecuación lineal para calcular sus actividades por medio de la ecuación lineal. Una vez calculada la actividad, el ligando se vuelve a incluir en el modelo y se saca al ligando posterior, se calcula su actividad con el modelo, y el proceso se repite con cada uno de los 39 ligandos que participan en la ecuación. Con esto se pretende calcular el valor de los residuales entre la actividad experimental determinada en la Universidad de Tübingen, Alemania, y el valor de la actividad predicho de la ecuación lineal.

Los datos de la actividad obtenidos experimentalmente por investigadores de la Facultad de Farmacia de la Universidad de Tübingen y los datos de la actividad obtenidos por la ecuación se representan en un gráfico donde se pone el ID de las moléculas en el eje X y los valores de pIC_{50} en el Y para apreciar gráficamente las diferencias.

Los valores que se obtengan de la diferencia del pIC_{50} experimental y el pIC_{50} teórico calculado con la ecuación, se utilizarán para realizar el estudio de *Z Score*.

4.8.2 *Z Score*

Este es un método de análisis que realiza MOE calculando los valores *Z* de cada ligando para revisar su confiabilidad. El programa muestra que valores de *Z* mayores a 2.5 indican moléculas que se salen del rango permitido por lo que disminuye la confiabilidad (*outlier*). Los resultados se representan en un gráfico donde en el eje X se ponen los nombres de las moléculas y en eje Y los valores de *Z*. Este es un método que confirma si los descriptores utilizados son los correctos, ya que un valor de más de 2.5 representa un ligando que no está siendo evaluado correctamente con los descriptores utilizados.

4.8.3 *Cross Validation*

Los cuatro ligandos (*test set*) que se separaron servirán para evaluar al modelo como moléculas que no pertenecen a la ecuación predictiva, pero de los cuales si se conoce la pIC_{50} . Los valores no deben estar muy alejados y se reportan en un gráfico que compara los valores obtenidos experimentalmente con los calculados teóricamente.

4.8.4 Chi cuadrada χ^2

Este es un estudio de confiabilidad estadístico que se base en el análisis de los residuales para valores predecibles, esto es la diferencia entre los valores obtenidos experimentalmente y los obtenidos con el modelo predictivo. Esta diferencia se eleva al cuadrado para eliminar signos, se suman y se observa si el valor cae dentro del rango establecido para valores de χ^2 con determinado nivel de confianza. Los niveles de confianza van desde un máximo de 90% de área bajo la curva a un mínimo de 99% de área bajo la curva. Los rangos se encuentran establecidos en tablas estadísticas.