

CAPITULO I: ACOPLAMIENTO MOLECULAR

1.1 Generalidades

El acoplamiento molecular es una técnica rápida y económica que en las últimas décadas se ha empleado ampliamente, tanto a nivel académico, como industrial. Esta disciplina se ha consolidado en los últimos años, aunque muchos aspectos todavía son un desafío y no existen rutas directas y precisas para la identificación de los verdaderos ligandos entre un conjunto de moléculas, ni para identificar con precisión la conformación correcta del ligando dentro de una diana biológica determinada.

Este método se ha incorporado a la investigación farmacéutica con éxito, debido a una gran variedad de programas para el estudio de sistemas biológicos y químicos complejos. Hoy en día, debido a la disponibilidad de una variedad de algoritmos de acoplamiento, la comprensión de las ventajas y limitaciones de cada método es de fundamental importancia en el desarrollo de estrategias efectivas y la generación de resultados relevantes ^{4,5}.

1.2 Definición

El acoplamiento molecular se define como una técnica computacional, que permite predecir la estructura de complejos intermoleculares que se establecen entre dos o más moléculas. Uno de los casos más estudiados con esta metodología es la interacción proteínas-ligandos, seguido de las interacciones proteína-proteína. Su principal objetivo es predecir energías y modos de enlace, y los estudios suelen hacerse con antelación a los ensayos experimentales.

Posteriormente, después del estudio de acoplamiento se consideró la mejor pose obtenida, que casi siempre es la de menor energía de unión. Alternativamente, también podemos considerar el menor valor de RMSD por sus siglas en inglés (Root Mean Square Deviation) que se refiere a la desviación de la raíz cuadrática media con respecto a una estructura de referencia⁶.

Es innegable, la versatilidad de esta herramienta permitiendo probar con casi todo tipo de estructura como ligando, con la mayoría de macromoléculas registradas en las bases de datos como diana biológica. Esto abre una ventana colosal a nuevas investigaciones las cuales, inician el camino al descubrimiento de innumerables compuestos candidatos a fármacos. En la **figura 1** se resaltan los elementos necesarios para el estudio de acoplamiento molecular.

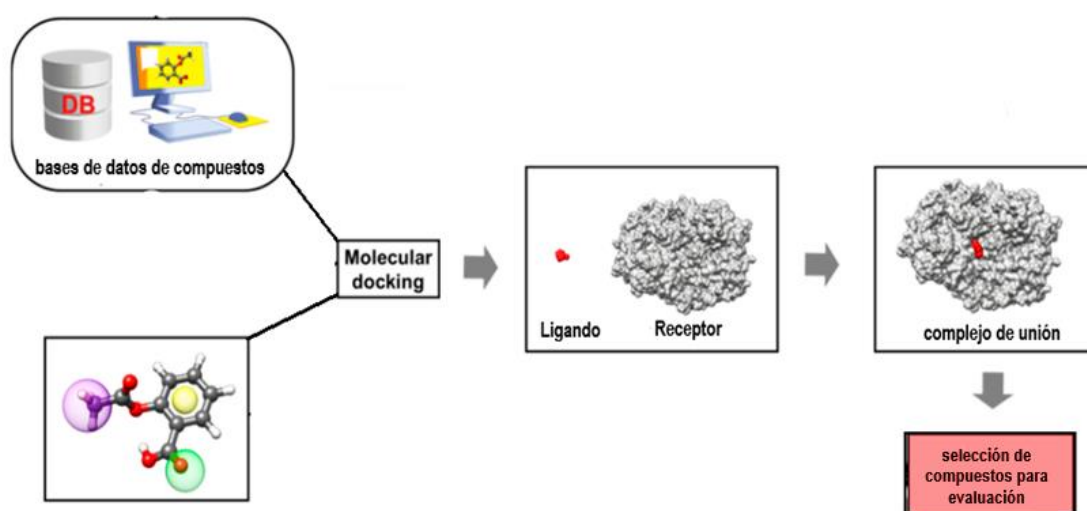


Figura 1. Elementos necesarios para el acoplamiento molecular¹⁸

1.3 Tipos de Acoplamiento Molecular

Se desarrollaron dos metodologías generales para el acoplamiento. En primer lugar, el enfoque de cuerpo rígido que está estrechamente relacionado con el modelo clásico de Emil Fischer. En este modelo, el ligando y el receptor se consideran dos cuerpos independientes que se reconocen entre sí en función de la forma y el volumen. El segundo enfoque es el acoplamiento flexible, este enfoque considera un efecto recíproco del reconocimiento de proteínas y ligandos en la conformación de cada parte⁷. La **Figura 2** muestra una esquematización de los dos enfoques principales para el acoplamiento molecular.

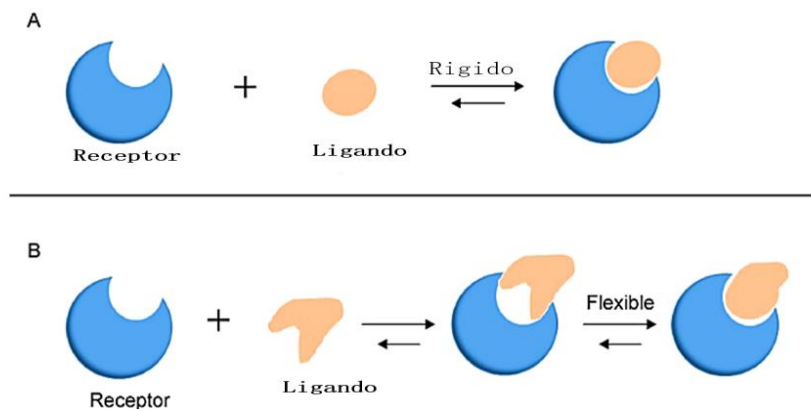


Figura 2. Enfoques principales para el acoplamiento molecular: **A)** cuerpo rígido y **B)** ajuste inducido o flexible²⁰.

1.3.1 Acoplamiento rígido

El acoplamiento rígido considera que la conformación del sistema de acoplamiento (receptor y ligando) no cambia. De acuerdo con esto, es adecuado para sistemas grandes, como proteína-proteína⁸, y ácido nucleico-proteínas. Además, el acoplamiento rígido no requiere múltiples cálculos, por lo que es el método más simple⁹. También, produce una gran cantidad de conformaciones acopladas con una complementariedad de superficie favorable, seguido de la reorganización de las conformaciones utilizando la energía libre de aproximación¹⁰.

Es importante tener en cuenta que cuando el ligando y el receptor se tratan como cuerpos rígidos, el espacio de búsqueda es muy limitado, considerando solo tres grados de libertad de traslación y tres de rotación.

1.3.2 Acoplamiento flexible

En los estudios de acoplamiento estándar, los ligandos se acoplan libremente en un receptor rígido. Sin embargo, se ha visto que la flexibilidad de la cadena laterales juega un papel crucial en los complejos ligando-proteína. Estos cambios permiten que el receptor altere su sitio de unión de acuerdo con la orientación del ligando.

Durante este proceso, el procedimiento de acoplamiento permite modificar fácilmente la conformación del sistema de acoplamiento (receptor y ligando). Dado que las variables del receptor y el ligando aumentan junto con el número de átomos, es necesario considerar varios factores adicionales. Estos incluyen el hecho de que el cálculo es sustancial y el proceso de acoplamiento es demasiado complicado⁹. La **figura 3**, resume las principales características de los tipos de acoplamiento molecular.

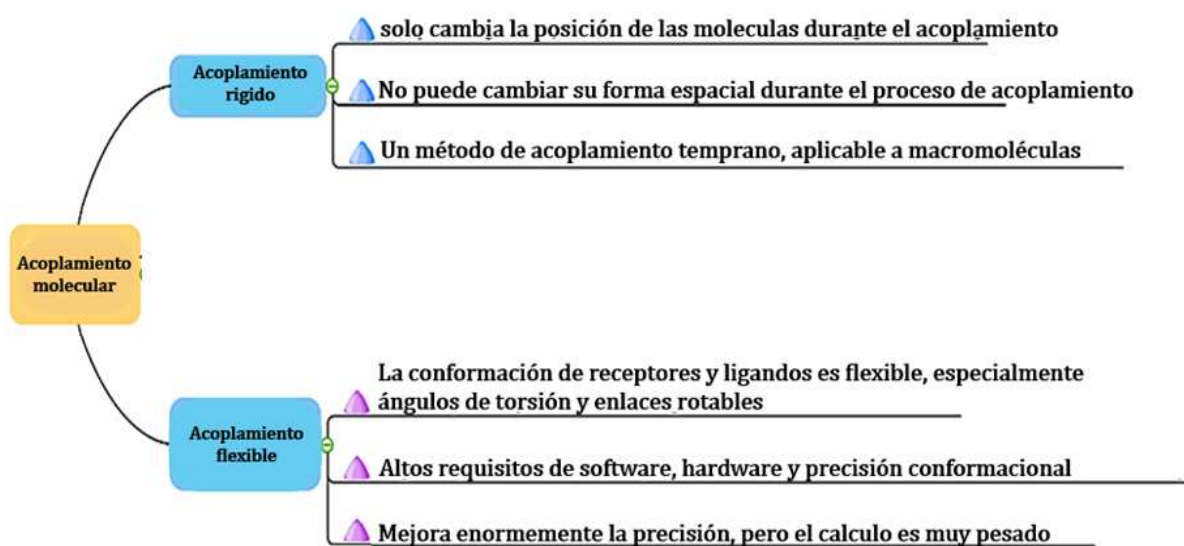


Figura 3. Resumen de las principales características para cada tipo de acoplamiento⁵⁸.

1.4 Algoritmos de Búsqueda.

En la etapa de búsqueda conformacional, los parámetros estructurales de los ligandos, como los grados de libertad de torsión (diedro), de traslación y de rotación, se modifican de forma incremental (**Figura 4A**). Los algoritmos de búsqueda conformacional realizan esta tarea mediante la aplicación de métodos de búsqueda sistemáticos y estocásticos¹¹. Los métodos de búsqueda sistemática promueven ligeras variaciones en los parámetros estructurales, cambiando gradualmente la conformación de los ligandos¹². El algoritmo sondea el panorama energético del espacio conformacional y, después de numerosos ciclos de búsqueda y evaluación, converge a la solución de energía mínima correspondiente al modo de unión más probable (**Figura 4B**). Aunque el método es eficaz para explorar el espacio conformacional, puede converger a un mínimo local en lugar de un mínimo global. Este inconveniente se puede superar realizando búsquedas simultáneas a partir de diferentes puntos del panorama energético.

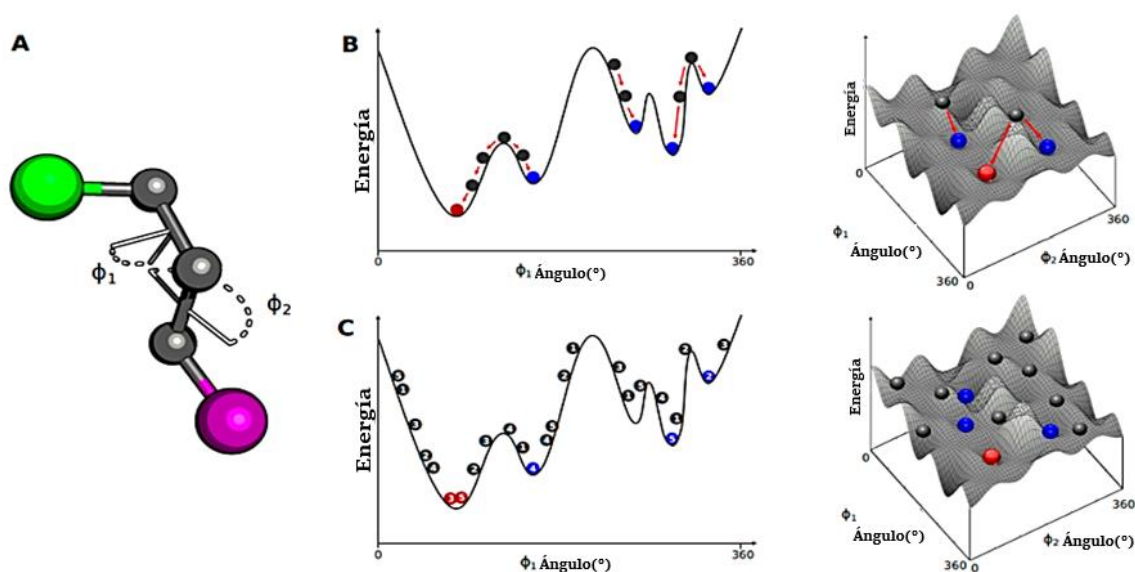


Figura 4. Métodos de búsqueda conformacional de moléculas pequeñas¹³.

Los métodos estocásticos llevan a cabo la búsqueda conformacional modificando aleatoriamente los parámetros estructurales de los ligandos. Para ello, el algoritmo genera conjuntos de conformaciones moleculares y llena una amplia gama del panorama energético (**Figura 4C**). Esta estrategia evita hallar la solución final en un mínimo de energía local y aumenta la probabilidad de encontrar un mínimo global. Como el algoritmo promueve una amplia cobertura del panorama energético, el costo computacional de este procedimiento será considerable y por tanto una limitación importante¹⁴.

Existe una amplia gama de algoritmos los cuales se derivan de lo tratado anteriormente a continuación, se describen brevemente los más importantes:

1.4.1 Emparejamiento rápido de forma (SM)

Los algoritmos de coincidencia de formas son enfoques que tienen en cuenta la superposición geométrica entre dos moléculas. Se emplean diferentes algoritmos para realizar varios alineamientos entre ligando y receptor. Este enfoque puede identificar los posibles sitios de unión de una proteína mediante una búsqueda de superficie macromolecular. Además, los algoritmos específicos de SM establecen posibles conformaciones de los sitios de unión predichos¹⁵.

Un ejemplo de este tipo de métodos lo encontramos en las aplicaciones del acoplamiento rígido, donde el software ZDOCK¹⁶, adopta el concepto básico de SM como su recurso principal, combinando la desolvatación con los parámetro electrostáticos a través de un algoritmo de transformada rápida de Fourier. Logrando una alta precisión en simulaciones de acoplamiento proteína-proteína reproduciendo datos con un RMSD inferior a 2,5 Å¹⁵.

1.4.2 Construcción Incremental (IC)

Este método divide el ligando en fragmentos que se acoplan por separado en el sitio del receptor. Una vez acoplados los fragmentos, las piezas se fusionan. Esta fragmentación

permite que el algoritmo considere la flexibilidad del ligando. Los fragmentos rígidos que se acoplan inicialmente funcionan como "anclajes" que están unidos secundariamente por partes flexibles de ligando que tienen enlaces rotativos. De esta manera, el ligando se "construye" gradualmente dentro del sitio de unión del receptor¹⁵.

También conocido como métodos de anclaje y crecimiento, el algoritmo IC se ha empleado en programas de acoplamiento como DOCK¹⁷, y FLEXX⁵.

1.4.3 Simulación de Monte Carlo (MC)

En este enfoque, los ligandos se modifican gradualmente mediante la rotación del enlace y la traslación o rotación de todo el ligando. También se puede cambiar más de un parámetro a la vez para obtener una conformación particular. Luego, esa conformación se evalúa en el sitio de unión basándose en el cálculo de la energía utilizando la mecánica molecular y luego se rechaza o acepta para la siguiente iteración en función de la constante de probabilidad de Boltzmann. La aceptación o rechazo de la conformación es función del cambio de energía con respecto a un parámetro T , que se puede interpretar físicamente como temperatura¹⁸. Este criterio de aceptación o rechazo hace que este método sea diferente a los demás. Mientras que el otro algoritmo favorece la disminución de energía, en el método MC también son posibles aumentos. Para valores más altos de T es probable que se produzcan aumentos. Si se comienza con un valor alto de T , entonces se pueden saltar pequeñas barreras de energía y la configuración puede moverse más allá de los mínimos locales, esto es, particularmente útil en situaciones donde se busca un mínimo global entre muchos mínimos locales. Un resultado interesante del enfoque MC es la búsqueda Tabú, que mantiene un registro del espacio de búsqueda del sitio de enlace que ya han sido estudiados. por lo tanto, garantiza que el sitio de enlace se explore al máximo¹⁵. Una de las principales preocupaciones con el enfoque MC es la incertidumbre de la convergencia, que puede mejorarse realizando múltiples ejecuciones independientes. Los ejemplos de aplicación del método MC incluyen una versión anterior de AutoDock, Affinity Dock Visión 1.0.3, Glam Dock, MCDOCK, PRODOCK, Ribo Dock, etc¹⁹.

1.4.4 algoritmo genético (GA)

Los algoritmos genéticos (GA), hacen parte de los métodos estocásticos, al igual que el enfoque Monte Carlo (MC). La idea del GA proviene de la teoría de evolución de Darwin. Donde se codifican los grados de libertad como cadenas binarias denominadas genes. Estos genes forman el "cromosoma" que en realidad representa la pose del ligando. Luego, ambos elementos se relacionan por medio de mutación y cruce que hacen parte de los operadores genéticos de GA. En contraste, la mutación produce cambios aleatorios en los genes; el cruce intercambia genes entre dos cromosomas. Cuando los operadores genéticos afectan a los genes, el resultado es una nueva estructura de ligando. Las nuevas estructuras se evaluarán mediante la función de puntuación, y las que sobreviven (es decir, superan el umbral) se pueden usar para la próxima generación. Los algoritmos genéticos se han utilizado en programas como AutoDock, GOLD, DIVALI y DARWIN²⁰.

1.5 Funciones de Puntuación

Las funciones de puntuación son métodos matemáticos de aproximación para estimar la afinidad de unión, que es una herramienta principal para la optimización de los resultados de acoplamiento, encontrando el ligando de mayor afinidad en el objetivo macromolecular. Su precisión puede validarse en pruebas de refinamiento de acoplamiento, donde se mira si el método de puntuación es capaz de predecir datos experimentales¹⁵.

La evaluación y clasificación de las conformaciones de ligando es un aspecto crucial en el cribado virtual. Cuando el interés es saber cómo se une un solo ligando a una biomolécula, la función de puntuación debe predecir la orientación acoplada que represente con mayor precisión la estructura del complejo intermolecular. No obstante, si el interés es evaluar múltiples ligandos, la función de puntuación debe ser capaz de clasificar a un ligando en relación con otro¹⁹.

Las funciones de puntuación generalmente estiman la energía de enlace de un complejo evaluando una amplia gama de propiedades que incluyen; interacciones intermoleculares, de solvatación, efectos electrostáticos y entrópicos¹³, para llegar lo más cerca posible de la energía de enlace real en un tiempo mínimo. Las funciones de puntuación populares tienen

un equilibrio adecuado entre la estimación precisa de la energía de unión y el costo computacional en términos de tiempo²¹.

Investigaciones previas clasificaron las funciones de puntuación en tres clases principales: las de campo de fuerza, empíricas y basadas en el conocimiento. Sin embargo, dado el avance en investigación actual, en estudios para las funciones de puntuación, especialmente en las interacciones proteína-ligando, fue necesario revisar las funciones de puntuación recientes obteniendo un nuevo esquema de clasificación, que clasifica las funciones de puntuación en basadas en la física, empíricas, basadas en el conocimiento y en el aprendizaje automático²². En la **figura 5** podemos observar algunos ejemplos de estas funciones.

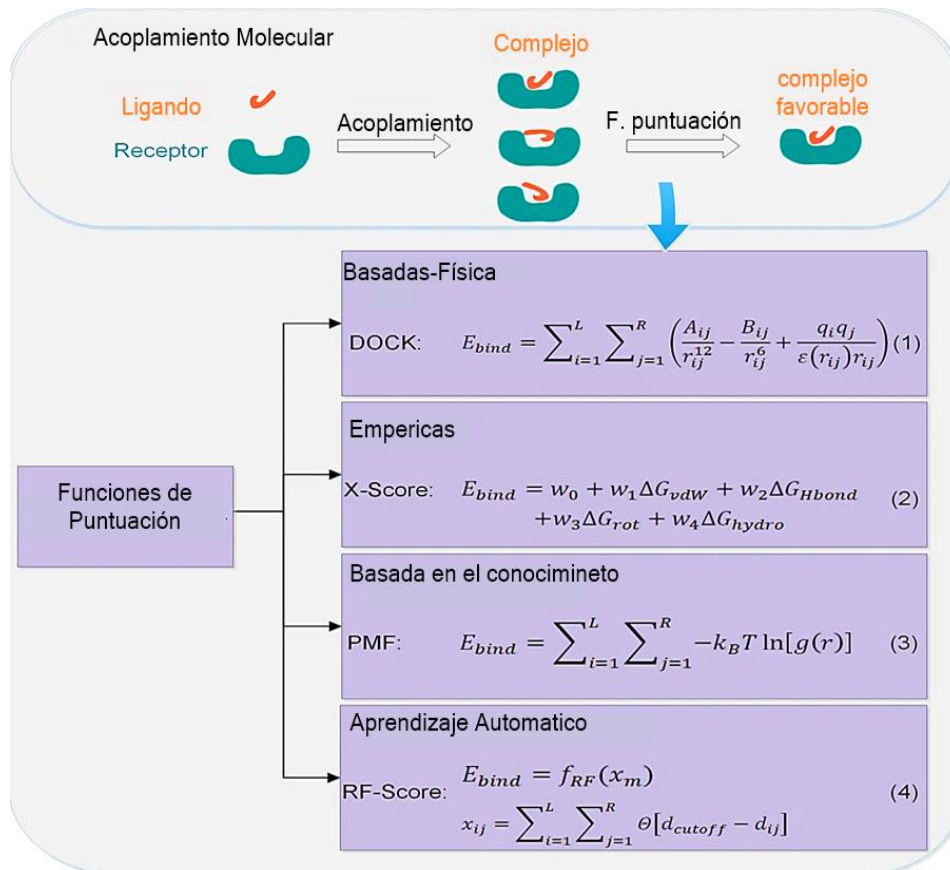


Figura 5. clasificación de las funciones de puntuación para el acoplamiento de proteínas y ligandos²².

1.5.1 Funciones de puntuación basadas en la física

Las funciones de puntuación basadas en la física, donde se incluyen las funciones basadas en campos de fuerza, los modelos de solvatación y los métodos de mecánica cuántica se ilustran en la **figura 6**. La función basada en el campo de fuerza calcula la energía de unión coincidiendo la interacción de Van Der Waals y la interacción electrostática entre los pares de átomos de proteína-ligando (**Ec.1 de la Figura 6**), que también consideran la contribución de la entalpía a la energía. Dado que no se tiene en cuenta a la entropía y el efecto solvente, el rendimiento de la función de campo de fuerza no es lo suficientemente buena. Por lo tanto, la función de campo de fuerza debe mejorarse incorporando la entropía de torsión de los ligandos y el efecto de solvatación / desolvatación descrito por modelos de solventes explícitos e implícitos²³ (**Ec.2 de la Figura 6**).

Estudios recientes han desarrollado de funciones de puntuación basadas en la mecánica cuántica (QM) para abordar los desafíos de las interacciones covalentes, la polarización y la transferencia de carga en el acoplamiento²⁴. Sin embargo, estas funciones basadas en QM tienen mayor precisión y costo computacional que las basadas en campo de fuerza. Por esta razón, se desarrolló un enfoque híbrido de mecánica cuántica/mecánica molecular (QM / MM) (**Ec.3 en la Figura 6**)²⁵.

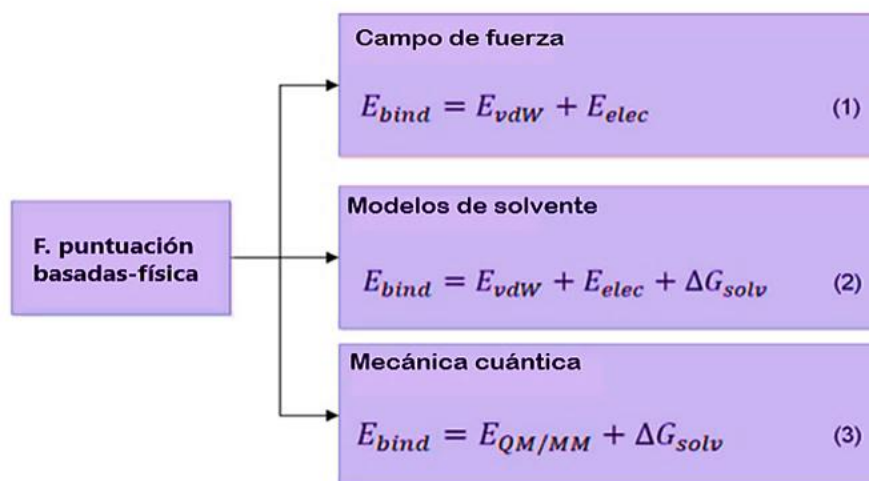


Figura 6. descripción de las funciones de puntuación basada en la física²²

En general, las funciones basadas en la física pueden calcular directamente las interacciones entre los átomos de la proteína y el ligando y su energía libre de unión, con una precisión predictiva relativamente mayor que otros tipos de funciones debido a la consideración de la entalpía, la solvatación y la entropía. Sin embargo, los modelos híbridos MM/QM en estas funciones físicas son computacionalmente caros²².

1.5.2 Funciones de puntuación empíricas

Las funciones de puntuación empíricas, estiman la afinidad de unión de un complejo al resumir los factores energéticos importantes para la unión de proteínas y ligandos, tales como enlaces de hidrógeno, efectos hidrófobos, choques estéricos, etc. Usualmente se emplea un conjunto de entrenamiento con afinidades de unión conocidas para optimizar las cargas de los factores energéticos para las funciones empíricas mediante análisis de regresión lineal^{26,27}. Un ejemplo de funciones empíricas, es la función X-score²⁸, (**Ec.2 en la figura 5**).

Aunque las funciones empíricas descomponen las afinidades de unión proteína-ligando en varios términos de energía individuales, similares a las funciones basadas en la física, generalmente emplean una forma funcional flexible e intuitiva distinta de los modelos establecidos que utilizan las funciones basadas en la física. Debido a sus términos de energía simples, estas funciones son buenas para predecir la afinidad de unión, la pose del ligando y el cribado virtual con un bajo costo informático, pero no es adecuado usarlas para describir la relación entre la afinidad de unión y las estructuras cristalinas²².

1.5.3 Funciones de puntuación basadas en conocimientos.

También denominadas funciones de puntuación basadas en el potencial estadístico, estas funciones emplean potenciales de energía que se derivan de la información estructural en el interior de las estructuras atómicas determinadas experimentalmente. El principio detrás de las funciones de puntuación basadas en el conocimiento es simple: los potenciales por pares se obtienen directamente de la frecuencia de aparición de pares de átomos en una

base de datos utilizando la relación de Boltzmann inversa. Para los estudios de ligandos en proteínas, los potenciales se calculan mediante la ecuación (1):

$$A = \sum_i^{lig} \sum_j^{prot} \omega_{ij}(r) \quad (1)$$

Esto es el potencial dependiente de la distancia entre el par de átomos i-j, es decir, $\omega_{ij}(r)$, se deriva de un análisis de Boltzmann inverso, dado por la ecuación (2):

$$\omega_{ij} = -k_B T \ln[g_{ij}(r)] = -k_B \ln \left[\frac{\rho_{ij}(r)}{\rho_{ij}^*} \right] \quad (2)$$

donde, $\rho_{ij}(r)$ es la densidad numérica del par de átomos i-j a la distancia r, y ρ_{ij}^* es la densidad numérica del mismo par de átomos en un estado de referencia donde se asume que las interacciones interatómicas son cero^{22,29}.

Por otro lado, en comparación con las funciones de campo de fuerza, y las funciones empíricas, las funciones de puntuación basadas en el conocimiento ofrecen un buen equilibrio entre precisión y velocidad. Debido a que los potenciales se extraen de las estructuras en lugar de intentar reproducir las afinidades conocidas por ajuste, y debido a que la base de datos estructural de entrenamiento puede ser grande y diversa, las funciones de puntuación basadas en el conocimiento son bastante robustas y relativamente insensibles a el conjunto de entrenamiento. Su característica por pares también permite que el proceso de puntuación sea tan rápido como las funciones de puntuación empírica.

1.6 Software para Acoplamiento Molecular

El acoplamiento molecular In Silico se realiza mediante softwares específicamente formulados. Desde que se desarrolló el primer programa de acoplamiento en 1982, ha surgido un creciente número de software para acoplamiento molecular que se actualizan constantemente. Sin embargo, existen diferencias de precisión y funcionalidad entre las diferentes versiones de programas disponibles actualmente. Se han informado más de 60 herramientas de acoplamiento molecular en la literatura. Actualmente, los más utilizados incluyen principalmente AutoDock, DOCK, GOLD, FlexX, AutoDock Vina, ICM, Glide,

Surflex y Affinity. Como se muestra en la **Tabla 1**, donde se comparan varios tipos de software de acoplamiento, algoritmos, métodos de evaluación, precisión de acoplamiento, tipos de acoplamiento y más^{10,30,31}.

Tabla 1. Comparación de software de acoplamiento molecular⁹.

Software	Algoritmo	Método de evaluación	Exactitud (%)	Tipo	Velocidad	Costo
AutoDock	Algoritmo genético lamarckiano (LGA)	Energía libre semiempírica	49	Semi-flexible	Promedio	gratis
DOCK	Coincidencia Geométrica	Campo de fuerza molecular	--	Semi-flexible	Rápido	gratis
GOLD	Genético	Energía libre semiempírica	78	flexible	Rápido	pago
FlexX	Construcción Incremental	Energía libre semiempírica	58	Semi-flexible	Rápido	pago
AutoDock Vina	Broyden–Fletcher– Goldfarb–Shanno	Energía libre semiempírica	78	Semi-flexible	Rápido	gratis
ICM	Optimización Global Estocástica	Energía libre semiempírica	--	Semi-flexible	Rápido	pago
Glide	Búsqueda sistemática	Energía libre semiempírica	82	Semi-flexible	Promedio	pago
Surflex	Similitud molecular basada en superficie	Energía libre semiempírica/ Campo de fuerza molecular	75	flexible	Rápido	pago
Affinity	Monte Carlo (MC)	Campo de fuerza molecular	--	flexible	Lento	pago
LigandFit	Monte Carlo (MC)	Campo de fuerza molecular	46	Semi-flexible	Rápido	gratis
Discovery Studio	Mecánica Molecular (MD)	Campo de fuerza molecular	--	flexible	Lento	gratis

El software de acoplamiento semiflexible AutoDock es uno de los más utilizados por ser de código abierto y por su facilidad de uso, aunque no es el más preciso según los datos de la **Tabla 1**, este consta de dos partes: AutoGrid, que se utiliza para calcular el nivel de energía contenido en la red, y Auto-Dock, se emplea principalmente como herramienta de búsqueda para determinar la conformación y las puntuaciones óptimas. Luego, tenemos al software AutoDock Vina que, aunque es basado en el programa anterior, exhibe una mayor precisión y velocidad de acoplamiento, ya que, deja de lado el uso de los algoritmos genéticos y utiliza la optimización de gradiente para establecer el punto mínimo. Este hecho pospone al AutoDock vina como uno de los mejores software de código abierto disponible en la web, puesto que, su exactitud está por encima no solo de AutoDock, sino del resto software de código abierto⁹.

Así mismo, dentro de los programas de código cerrado o pagos tenemos a GOLD, desarrollado por el Centro de datos cristalográficos de Cambridge. El cual, utiliza un algoritmo genético (GA) para el acoplamiento de receptores y ligandos. Este programa tiene un gran porcentaje de exactitud, por lo que es particularmente atractivo, para los grupos de investigación dedicados al desarrollo de moléculas bioactivas. De igual forma, los programas Glide y Surflex, gozan de gran popularidad en la comunidad científica, debido a su óptimo desempeño en la simulación de moléculas, el gran número de publicaciones donde aparecen estos programas lo ratifica. Aunque, a pesar de que el Glide tiene un porcentaje de exactitud más alto que el Surflex y el GOLD, estos últimos permiten acoplamientos flexibles, lo que se asemejaría un poco más al proceso de acoplamiento de forma biológica.

1.7 Limitaciones

La principal limitación del acoplamiento molecular se debe a la carencia de confiabilidad en la capacidad de las funciones de puntuación para proporcionar energías de enlace precisas. Debido, al hecho de que algunos términos de interacción intermolecular son difíciles predecir con exactitud, tal es el caso, para el efecto de solvatación y el cambio de entropía. Además, algunas interacciones intermoleculares pocas veces se consideran en las funciones de puntuación, y han demostrado ser importantes. Por ejemplo, se sabe que la unión de halógeno contribuye a la afinidad de unión proteína-ligando, y también lo hacen las interacciones guanidina-arginina, pero estas no son consideradas dentro de los resultados del acoplamiento. Es decir, que no están parametrizadas dentro del archivo de ejecución de los programas^{19,32}.

Así mismo, todavía es un problema sin resolver la precisión de las moléculas de agua en un bolsillo de unión durante el acoplamiento, este problema es de gran relevancia. En primer lugar, porque las estructuras cristalinas de rayos X carecen de la información de coordenadas del hidrógeno, debido a la dispersión ineficaz de los átomos más pequeños. Si no se conoce la posición exacta del hidrogeno esto nos puede llevar a inexactitudes para encontrar moléculas de agua que podrían estar formando puentes entre el ligando y el receptor. En segundo lugar, se dispone de poca información teórica con suficiente

veracidad para poder predecir cuantas moléculas de agua son afectadas por los ligandos y que tan fuerte es el efecto³³.

Otro ejemplo donde falla el acoplamiento molecular es en la simulación del complejo transtirretina-tiroxina (**figura 7**), donde resultado arroja dos modos de unión posibles, uno similar al modo de unión nativo y otro perteneciente a un dominio de unión alternativo, con una desviación cuadrática media (RMSD) de 8,97 Å. Posteriormente, en la simulación energética la solución de menor energía elegida por el software de acoplamiento fue la de mayor RMSD, por lo que es claro, que el acoplamiento molecular no pudo realizar la predicción correcta del modo de unión con lo que se concluye que podríamos obtener muchos falsos negativos durante el proceso de cribado virtual³³.

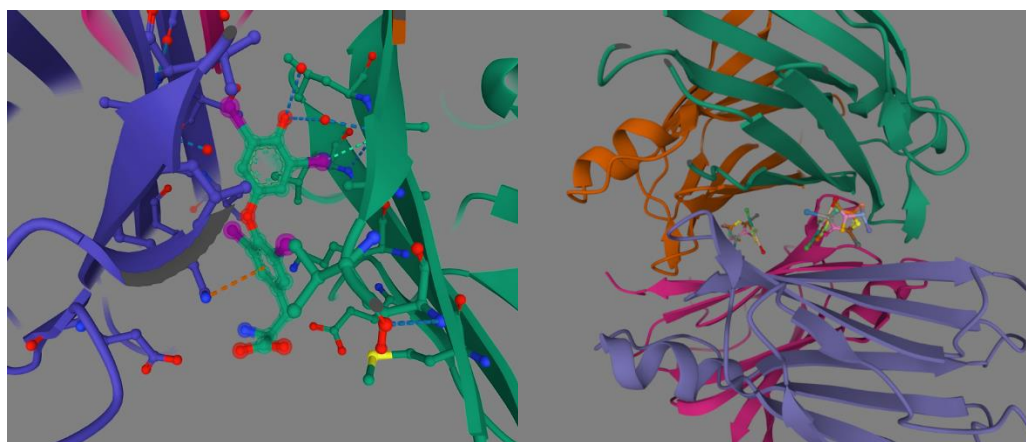


Figura 7. Ejemplo del complejo de transtirretina con tiroxina (T4), perteneciente a una rata PDB(1IE4)³⁴