



UNIVERSIDAD DE BUENOS AIRES  
FACULTAD DE CIENCIAS EXACTAS Y NATURALES  
DEPARTAMENTO DE QUÍMICA BIOLÓGICA

**Determinantes moleculares de la interacción droga-proteína:  
uso de cosolventes como moléculas de prueba**

Tesis presentada para optar por el título de Doctor de la Universidad de Buenos Aires  
en el área Química Biológica

**Juan Pablo Arcón**

Director de tesis: Adrián Gustavo Turjanski

Consejero de estudios: Valeria Levi

Lugar de trabajo: Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires – Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN), UBA/CONICET.

Buenos Aires, 2018

# 1. INTRODUCCIÓN

*A medida que los métodos de la química estructural se apliquen de modo más profundo a problemas fisiológicos, se encontrará que la importancia del enlace de hidrógeno para la fisiología es mayor que la de cualquier otra característica estructural.*

Linus Pauling. The Nature of the Chemical Bond. 1939.

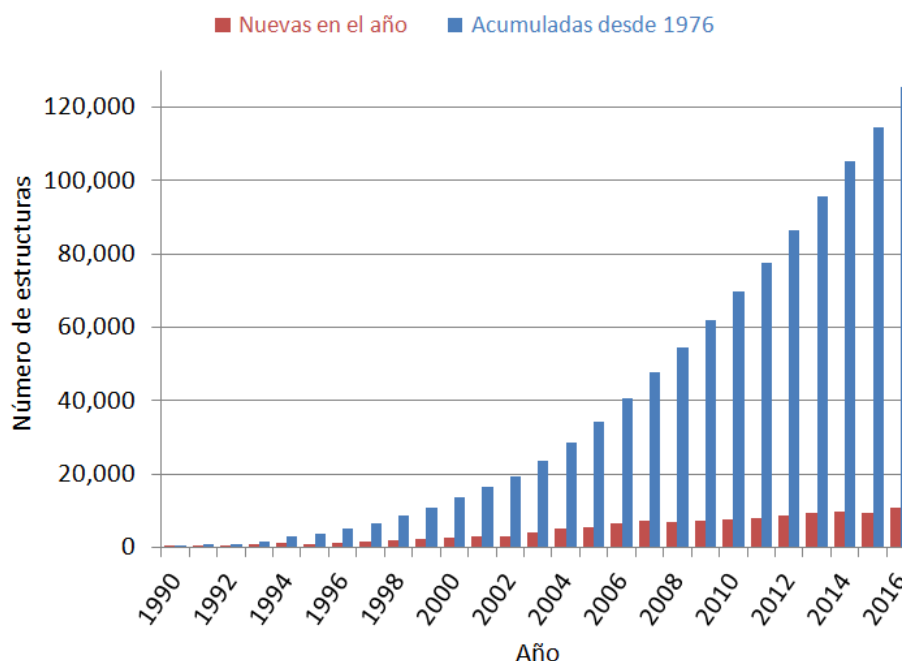
## 1.1. Descubrimiento de fármacos

### 1.1.1. Fármaco, droga, compuesto tipo droga y drogabilidad

En esta tesis utilizaremos los términos *fármaco* y *droga* de manera indistinta para referirnos a sustancias químicas aprobadas por una agencia de regulación oficial que sirven para el tratamiento, cura o prevención de una enfermedad.<sup>†</sup> En particular, nos centraremos exclusivamente en moléculas pequeñas, las cuales constituyen la gran mayoría de los fármacos y no trataremos con productos biológicos como anticuerpos, proteínas recombinantes o células madre, entre otros. El término *molécula pequeña* típicamente implica un compuesto orgánico, no polimérico, con un peso molecular inferior a 1.000 Da (Figura 1.1). En este contexto, se suele definir a un *compuesto tipo droga* como una molécula pequeña con ciertas características que le brindan una biodisponibilidad adecuada en humanos.<sup>1</sup> Existen diversos parámetros para evaluar dichas características y, por ende, la calidad de un compuesto como posible fármaco. Para un compuesto a ser administrado por vía oral en humanos, la regla de 5 de Lipinski<sup>2</sup> resultó ser un excelente punto de partida: no más de 5 grupos donores de enlace de hidrógeno, no más de 10 grupos aceptores de enlace de hidrógeno, peso molecular menor a 500 Da y logP (coeficiente de partición octanol-agua) menor a 5. Luego, se fueron agregando otras propiedades entre las que cabe destacar el límite de enlaces rotables<sup>3</sup> en 10, una alta solubilidad y la ausencia de grupos reactivos<sup>4</sup>. Por supuesto que existen excepciones a estas reglas, incluso entre drogas aprobadas y de venta masiva (Figura 1.1).

---

<sup>†</sup> Tomado de parte de la definición de droga según *Classification of Products as Drugs and Devices & Additional Product Classification Issues: Guidance for Industry and FDA Staff 2017*, Food and Drug Administration, U.S. Department of Health and Human Services.

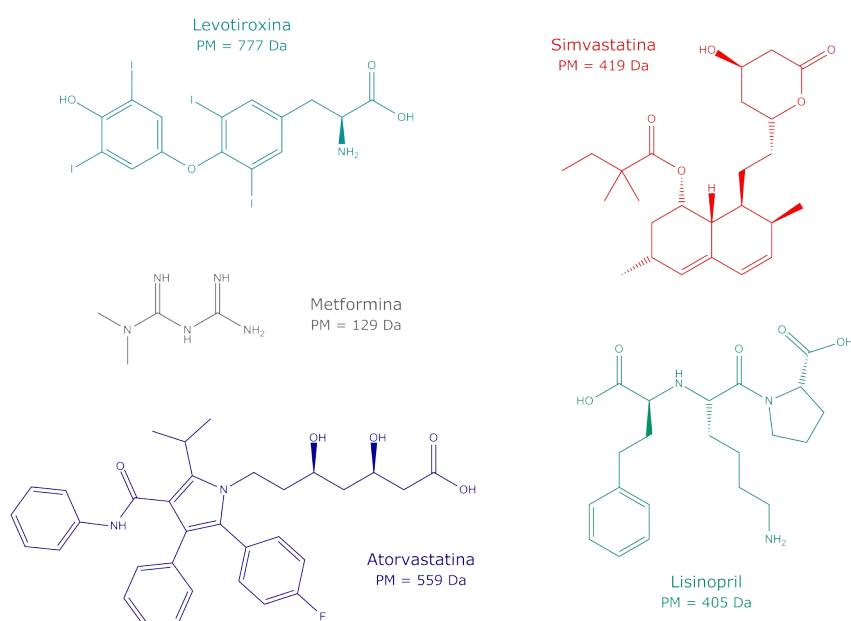


**Figura 1.4.** Cantidad de estructuras depositadas anualmente en la base de datos del PDB desde 1990 (barras rojas) y acumuladas desde 1976 (barras azules). Datos obtenidos del RCSB PDB<sup>35</sup> ([www.rcsb.org](http://www.rcsb.org)).

#### 1.1.4. Problemas actuales en el descubrimiento de fármacos

Los datos en la Figura 1.5 muestran que, si bien el número de nuevos medicamentos aprobados por la Administración de Alimentos y Fármacos de Estados Unidos (*Food and Drug Administration*, FDA) se mantuvo relativamente constante desde 1950, el número por cada 1.000 millones de dólares invertidos en Investigación y Desarrollo (I+D) en la industria farmacéutica se redujo notablemente a partir de la década de 1960.<sup>62</sup> A partir de 1962, las agencias de regulación elevaron los estándares para la introducción de nuevos fármacos<sup>63</sup> como resultado del caso mundialmente conocido de la droga talidomida, un fármaco que se vendió hacia finales de 1950 para aliviar náuseas durante el embarazo y que provocó que alrededor de 6.000 bebés nacieran con malformación de las extremidades en Alemania Federal, sobreviviendo sólo el 40%.<sup>64</sup> Se estima que hubo 10.000 casos en el mundo de los cuales sólo sobrevivió la mitad.

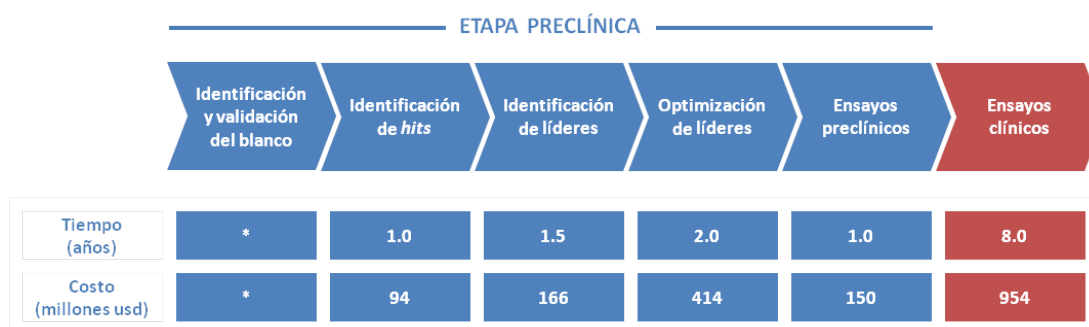
Es importante destacar que para que una droga ejerza su efecto, además de contar con las características mencionadas, debe unirse a un blanco molecular, el cual muchas veces es una proteína. La *drogabilidad* es un concepto que describe cuán factible es que la proteína blanco se una fuertemente a un compuesto tipo droga (típicamente  $K_d$  menor a  $1\ \mu\text{M}$ ) y que dicha unión sea biológicamente relevante, es decir, desencadene el efecto farmacológico deseado.<sup>5</sup> Más adelante veremos los fundamentos bioquímicos y termodinámicos de la unión droga-proteína.



**Figura 1.1.** Drogas más prescritas en Estados Unidos durante el año 2015 según el número total de ventas. Tomado de *Prescribed Drug Estimates 2015, Medical Expenditure Panel Survey, Agency for Healthcare Research, U.S. Department of Health & Human Services*.

### 1.1.2. Etapas del descubrimiento y desarrollo de fármacos

Para llegar a obtener un fármaco hay que atravesar una serie de etapas consecutivas, las cuales se muestran esquemáticamente en la Figura 1.2. El descubrimiento y desarrollo de fármacos se divide principalmente en dos instancias: la fase preclínica, donde la droga se diseña y evalúa *in vitro*, en células y en animales, y los ensayos clínicos, donde el fármaco se prueba en humanos.



**Figura 1.2.** Modelo estándar actual del proceso de descubrimiento y desarrollo de fármacos. Se muestra el tiempo y el costo capitalizado que insumen cada una de las etapas para lograr la obtención de sólo un nuevo fármaco<sup>6</sup>. \*No hay datos para la etapa de elección y validación del blanco molecular porque resultan altamente variables.

Las estrategias abordadas en el presente trabajo de tesis se enmarcan en la fase preclínica, por lo que nos centraremos en la descripción de dichas etapas.

#### (i) *Identificación y validación del blanco*

Se trata principalmente de la obtención de un blanco para atacar, su caracterización y la puesta a punto de un ensayo experimental que permita reconocer el efecto de los compuestos candidatos a fármacos sobre el mismo. En general, por blanco se entiende una proteína involucrada en un proceso patológico que se desea tratar, aunque existan otros blancos moleculares como ARN mensajeros o regulatorios<sup>7</sup>. La validación del blanco elegido se extiende más allá en la cadena de descubrimiento, incluso hasta evaluar que su intervención terapéutica produce la eficacia esperada en pacientes. Nosotros trabajaremos sobre blancos moleculares proteicos ya validados para poner a prueba los métodos desarrollados.

#### (ii) *Identificación de compuestos hits*

Una vez identificado el blanco proteico, el camino general consiste en purificar ese blanco usando técnicas de ADN recombinante y desarrollar un método que permita medir la unión de compuestos, ya sea de manera directa o mediante la inhibición de la función proteica. Entonces, a partir de bibliotecas con gran cantidad de compuestos, se realizan ensayos experimentales de *screening* a gran escala (*high throughput screening*, HTS) que permiten seleccionar *hits* con una potencia de al menos 10  $\mu$ M contra el blanco.<sup>4,8</sup> Se suelen probar desde cientos de miles hasta millones de compuestos, lo que limita la aplicación del HTS a grandes empresas farmacéuticas o centros públicos de envergadura como el “NIH Chemical Genomics Center” o “The Scripps Research Institute” en Estados Unidos.<sup>9</sup> Además, no siempre resulta sencillo desarrollar una metodología que permita realizar este tipo de *screenings*. Por esta razón, muchas veces se utilizan técnicas de *screening* virtual, es decir el análogo *in silico* del HTS, para

acotar el número de compuestos que se ensayen en las pruebas experimentales correspondientes. Parte importante de la tesis se centra en técnicas de *screening* virtual por lo que éstas se introducen en las secciones 1.1.3.3 y 1.1.3.4.

#### (iii) *Identificación de compuestos líderes o “hit to lead”*

La definición precisa de una molécula *hit* o líder (*lead*) no es estricta. En general, un compuesto resultante del *screening* se considera *hit* si tiene potencia en el ensayo bioquímico dado (por ejemplo, afinidad de 10  $\mu\text{M}$  o menor) y factibilidad de compra o síntesis. A través de rondas iterativas de síntesis química y pruebas experimentales, un *hit* se transforma en un líder con mayor afinidad, buena comprensión de su relación estructura-actividad y un perfil ADMET (absorción, distribución, metabolismo, excreción y toxicidad) inicialmente favorable.<sup>4</sup>

#### (iv) *Optimización de compuestos líderes*

Los objetivos del paso de optimización son incrementar la afinidad por el blanco (farmacodinámica) hasta el rango 1-10 nM<sup>4</sup>, mejorar las propiedades fisicoquímicas de solubilidad, estabilidad y ADMET (farmacocinética) y aumentar la selectividad, es decir, reducir la afinidad por otros blancos (*off-targeting*).

#### (v) *Ensayos preclínicos*

Consisten en verificar la actividad de los compuestos sobre modelos celulares y mediante pruebas en animales. Se evalúan la seguridad, toxicidad, farmacocinética y farmacodinámica para determinar una dosis apta que sirva para empezar las pruebas en humanos.

#### (vi) *Ensayos clínicos*

Son las pruebas en humanos. Constan de cuatro fases que van desde pruebas de dosis subterapéuticas en sujetos sanos hasta la etapa final de farmacovigilancia una vez que el fármaco ya está aprobado. No entraremos en el detalle de la etapa clínica porque está fuera del alcance de la presente tesis.

Los métodos desarrollados en esta tesis se enmarcan principalmente en la obtención de *hits* de un modo más efectivo y eficaz mediante mejoras en las técnicas de *screening* virtual tradicionales. Asimismo, veremos que las estimaciones de energía libre de unión logradas en este trabajo sientan las bases para la utilización de dichos métodos en otras etapas como, por ejemplo, la optimización de líderes usando únicamente información del blanco proteico o la comparación rápida entre *hits* que difieren poco en su estructura.

### 1.1.3. Métodos computacionales para el descubrimiento de fármacos

#### 1.1.3.1. Métodos basados en propiedades de los ligandos

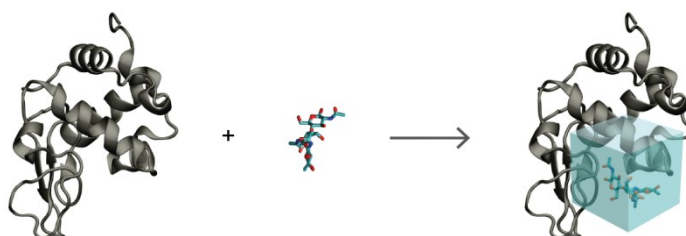
En el presente trabajo entenderemos por *ligando* a una molécula pequeña que se une a una proteína, si bien el término ligando tiene una definición amplia y suele usarse también para cualquier molécula que se une a otra molécula de interés. Los primeros métodos que se basaron en características de los ligandos para predecir la actividad de los mismos sobre blancos específicos fueron desarrollados en la década de 1960 y lograron establecer una relación entre la estructura química de los compuestos y su actividad biológica de manera cuantitativa (QSAR, *quantitative structure-activity relationship*). En sus comienzos se generaron modelos en 2D analizando el efecto de sustituyentes de anillos aromáticos en la actividad de ligandos sobre distintos blancos.<sup>10</sup> Actualmente se obtienen modelos 3D que provienen, por ejemplo, del cálculo de energías de interacción en grillas alrededor de un conjunto de ligandos de entrenamiento y su relación con la actividad conocida de estos mismos ligandos.<sup>11</sup> Otra estrategia se basa en las búsquedas por similitud de subestructuras que resulten interesantes para la unión al blanco. Así, los métodos de análisis de similitud química en 2D se utilizan para comparar los compuestos de una base de datos con ligandos activos del blanco de interés.<sup>12</sup> Finalmente, existen los métodos de farmacóforo derivado de ligandos. El farmacóforo representa un conjunto de rasgos funcionales distribuidos en el espacio 3D que pueden interactuar con un receptor específico para ganar actividad, es decir es un arreglo tridimensional en el espacio de funciones químicas que se creen responsables de la actividad biológica.<sup>13,14</sup> Se genera en base a propiedades estéricas y electrónicas del conjunto de ligandos de entrenamiento que se unen al mismo receptor y que resultan diversos estructuralmente. Entonces, se analizan las características que se repiten entre los ligandos y que presumiblemente son importantes para la unión y con ello se construye el modelo de farmacóforo. Los compuestos de una base de datos luego se comparan con el farmacóforo y aquellos que resulten compatibles con el mismo tendrán una alta probabilidad de unirse al blanco.<sup>15</sup> Tanto para el método de farmacóforo como para QSAR resulta muy importante el alineamiento de las moléculas de entrenamiento. Como veremos en esta tesis, un farmacóforo también puede derivarse a partir del blanco.<sup>14</sup>

Los métodos basados en propiedades de los ligandos se apoyan en la premisa de que compuestos con características similares tienen actividades similares. Su principal desventaja es que para establecer predicciones confiables requieren información de gran cantidad de compuestos (decenas a centenas) que a su vez muestren diversidad y cierta actividad sobre el blanco en estudio. Dicha información experimental no está nunca disponible para blancos relativamente novedosos. Al mismo tiempo, resulta más complicado descubrir compuestos con funcionalidades originales cuando el análisis se

centra en un gran número de ligandos ya conocidos. Por ello, en la presente tesis optamos por un enfoque basado en la estructura del blanco proteico, como se introduce en la siguiente sección.

### 1.1.3.2. Métodos basados en la estructura del blanco: *docking*

Los métodos de *docking* molecular permiten predecir la estructura de complejos proteína-ligando y estimar su afinidad de manera rápida (Figura 1.3). Usualmente, la proteína se mantiene en una dada conformación (rígida) y se identifican diferentes poses para el ligando (flexible), de las cuales se selecciona aquella que presente mayor afinidad. Se requiere conocer la estructura tridimensional del blanco, o al menos del sitio de unión, a nivel atómico y ésta puede provenir de experimentos de difracción de rayos X con resolución por debajo de 2,5 Å, de experimentos RMN o de modelos por homología de alta identidad. La estructura tridimensional de las moléculas pequeñas o posibles ligandos puede obtenerse también de estructuras cristalinas o de RMN, pero cuando es necesario *dockear* grandes cantidades de compuestos se modelan las conformaciones *de novo*. En general, el *docking* proteína-ligando no se aplica a toda la superficie proteica sino que, como se ilustra en la caja de la Figura 1.3, se restringe a un sitio de interés obtenido de información previa proveniente de cocristales o de programas de detección de bolsillos drogables, entre otros.



**Figura 1.3.** Esquema del *docking* de un ligando al sitio de unión de una proteína.

Para lograr sus objetivos, los programas de *docking* necesitan básicamente dos componentes: (i) un algoritmo de búsqueda eficiente que halle diferentes poses del ligando en relación a la estructura del blanco y (ii) una función de puntaje para estimar la afinidad de cada pose e ir seleccionando las poses energéticamente más favorables.

**Algoritmo de búsqueda.** El algoritmo de búsqueda es el encargado de explorar de manera efectiva el espacio conformacional disponible para el ligando en el sitio de interés. Dada la enorme cantidad de posibilidades, se usan técnicas heurísticas que permiten hallar poses aceptables en tiempos de simulación acotados. En la gran mayoría de los casos los programas tienen implementados métodos estocásticos como



*simulated annealing* (Monte Carlo)<sup>16,17</sup> o algoritmos genéticos<sup>18–20</sup> que operan haciendo cambios aleatorios sobre los grados de libertad conformacionales del ligando. Las poses del ligando que se van obteniendo se evalúan mediante una función de probabilidad predefinida que define si se mantienen como opción o se eliminan. Los grados de libertad evaluados suelen restringirse a la traslación, rotación y las torsiones de enlaces. Las distancias y ángulos de enlace permanecen fijas, incluso en anillos alifáticos, de modo de simplificar la búsqueda.

**Funciones de puntaje.** Las funciones de puntaje de los programas de *docking* están diseñadas para estimar la afinidad de manera rápida. Si bien existen enfoques rigurosos basados en la fisicoquímica del proceso de unión, las funciones de puntaje están preparadas para predecir modos de unión y para discriminar entre compuestos que se unen al blanco y aquellos que no, más que para brindar un valor preciso de energía libre de unión. Existen tres grandes categorías de funciones de puntaje usadas en *docking*: (i) basadas en campos de fuerza, (ii) empíricas y (iii) potenciales estadísticos. Las funciones basadas en los campos de fuerzas diseñados para dinámica molecular como AMBER<sup>21</sup> y CHARMM<sup>22</sup> calculan las interacciones intermoleculares de van der Waals y electrostáticas de a pares entre los átomos del ligando y la proteína y, en algunos casos, agregan términos de enlaces de hidrógeno, energía intramolecular del ligando y desolvatación<sup>23</sup>. Por su parte, las funciones de puntaje empíricas consisten en la suma de términos de interacción considerados importantes para la unión, cada uno de ellos multiplicado por un coeficiente cuyo valor se obtiene de un ajuste por regresión lineal múltiple a energías de unión experimentales para un conjunto de ligandos conocidos.<sup>24</sup> Finalmente, los potenciales estadísticos (funciones *knowledge-based*) son el resultado de observaciones estadísticas de contactos cercanos intermoleculares entre proteínas y ligandos en bases de datos estructurales como Protein Data Bank (PDB)<sup>25</sup> o Cambridge Structural Database (CSD)<sup>26</sup>, los cuales son utilizados para derivar un potencial de fuerza media.<sup>27</sup> Cabe realizar una mención final a funciones de puntaje alternativas que no entran en la clasificación anterior y que se basan en aprendizaje automático<sup>28</sup> o en descriptores de la forma estructural y contactos complementarios<sup>29</sup>.

La Tabla 1.1 lista los programas de *docking* más utilizados y las opciones que ofrecen respecto de la búsqueda y puntaje de ligandos. La elección de uno u otro programa se basará en la aplicación deseada (búsqueda de modos de unión, *screening* virtual de bases de datos) y la permeabilidad de los códigos para implementar modificaciones. De la Tabla 1.1 vemos que las funciones de puntaje generalmente están ajustadas a conjuntos de compuestos conocidos que idealmente cubren un amplio espectro de blancos. La versatilidad de aplicación que se gana con estos conjuntos redundará en un desempeño lejos del óptimo para cada blanco particular. En nuestro trabajo combinaremos las funciones de puntaje generales con información específica de interacciones de cada blanco de interés para demostrar que los métodos de *docking*

pueden lograr resultados satisfactorios en una proporción de blancos mucho mayor a la actual.

**Tabla 1.1.** Enfoques de los programas de *docking* más utilizados.

Programa	Algoritmo de búsqueda	Función de puntaje	Observaciones
AutoDock <sup>30</sup>	Algoritmo genético + minimización local. <i>Simulated annealing</i> (opcional).	Basada en campo de fuerzas y ajustada por afinidad de complejos conocidos.	Ligando flexible.
DOCK <sup>31</sup>	Algoritmo genético.	Basada en campo de fuerzas (pocos términos).	Método simple y rápido de coincidencia geométrica entre el ligando rígido y la imagen invertida del sitio de unión ( <i>hot spots</i> ).
GLIDE <sup>32</sup>	Generación previa de conformaciones + búsqueda exhaustiva + minimización.	Empírica. Basada en fisicoquímica.	Ligando flexible. <i>Software</i> pago.
GOLD <sup>19</sup>	Algoritmo genético.	Empírica. Basada en fisicoquímica. Potencial estadístico.	Ligando flexible. <i>Software</i> pago.
rDock <sup>33</sup>	Algoritmo genético + <i>Simulated annealing</i> + minimización.	Empírica. Basada en fisicoquímica.	Ligando flexible.
Vina <sup>34</sup>	Mutación + Optimización local + Metropolis	<i>Knowledge based</i> + empírica.	Ligando flexible.

### 1.1.3.3. *Screening* virtual

El *screening* virtual consiste en la evaluación automática de grandes bases de datos de compuestos químicos mediante programas informáticos con el fin de hallar aquellas moléculas que se unan a un blanco de interés<sup>35</sup>. Su objetivo central es filtrar *grosso modo* el enorme espacio químico de compuestos posibles y obtener un número manejable de estos compuestos que se puedan sintetizar o comprar para probar experimentalmente. En una campaña de *screening* virtual se debe producir, a partir de una biblioteca de miles a millones de compuestos virtuales, un subconjunto enriquecido donde la proporción de compuestos que realmente se unen al blanco de interés es mucho mayor que en la base de datos original<sup>36</sup>. En la práctica es importante que la biblioteca de partida sea diversa, esté curada y que los compuestos se puedan adquirir comercialmente o sean sintetizables de modo más o menos directo. Algunos ejemplos de estas bibliotecas incluyen las de empresas vendedoras de compuestos como Asinex Gold & Platinum Collections (Asinex Co.) o Enamine Screening Collection (Enamine Ltd.) y las bases de datos públicas como ZINC<sup>37</sup> o PubChem<sup>38</sup>. Como dijimos, para los

resultados de *screening* es aceptable hallar compuestos con actividad del orden de 10  $\mu\text{M}$ <sup>4,8</sup>, de modo que luego puedan ser optimizados por química medicinal. Es importante aclarar que el uso de *screening* virtual no pretende identificar candidatos a fármacos listos para los ensayos clínicos, sino más bien de encontrar *hits* iniciales con funciones químicas que no se hayan asociado previamente al blanco de estudio, a un ritmo mayor y a un costo mucho menor que el *screening* experimental a gran escala.

Las técnicas de *screening* virtual abarcan ambos tipos de métodos computacionales descritos en las secciones 1.1.3.1 y 1.1.3.2, los que derivan de propiedades de los ligandos y aquellos que se basan en la estructura de los blancos. La posibilidad de realizar *screening* virtual se originó en la década de 1970 cuando se introdujeron metodologías para realizar búsquedas en bases de datos de compuestos utilizando fragmentos estructurales bidimensionales (2D).<sup>39-41</sup> A partir de allí se fueron introduciendo una amplia variedad de metodologías entre las que se destacan las búsquedas por similitud química de ligandos en 2D<sup>42</sup> y 3D<sup>43,44</sup>, el mapeo de farmacóforos<sup>45</sup>, y el *docking* molecular de compuestos contra el blanco de interés<sup>46</sup>.

Los programas de *docking* utilizados para realizar *screening* virtual de bases de datos basado en estructura son considerablemente más lentos que los métodos basados en ligando. Sin embargo, vimos que estos últimos tienen la desventaja de necesitar un conjunto de ligandos conocidos que se unan al blanco y su capacidad de encontrar nuevos compuestos activos está limitada por la diversidad o exhaustividad del conjunto de ligandos utilizados para construir el modelo. Por esta razón, los métodos de *docking* tienen cierta preponderancia en el descubrimiento de fármacos novedosos y nos centraremos en este tipo de *screening* en la presente tesis. La factibilidad del *screening* virtual basado en estructura usando *docking*, también conocido como *high throughput docking*, se logró gracias a la simplificación de los algoritmos de búsqueda y las funciones de puntuación. Veremos que hoy en día, en nuestro grupo podemos probar computacionalmente un millón de compuestos en 5 días o menos usando un *cluster* de tamaño moderado (1.000 procesadores).

#### 1.1.3.4. Comparación entre *screening* virtual y experimental

A principios de los años 2000 se registraron múltiples casos en la industria farmacéutica en los cuales el *screening* virtual arrojó resultados superadores a los ensayos de *screening* experimental a gran escala (HTS). Investigadores de la empresa Pharmacia realizaron *screening* virtual basado en *docking* para detectar inhibidores de la tirosina fosfatasa-1B, una enzima blanco clave de la diabetes tipo 2, obteniendo como resultado 365 *hits*, de los cuales 127 mostraron inhibición efectiva (tasa de acierto del 35%).<sup>47</sup> Al mismo tiempo, realizaron un HTS tradicional contra el mismo blanco, para 400.000 compuestos de la biblioteca de la compañía. Sólo 85 mostraron inhibición, de modo que la tasa de acierto fue de solo 0,02%. En 2003, mientras se

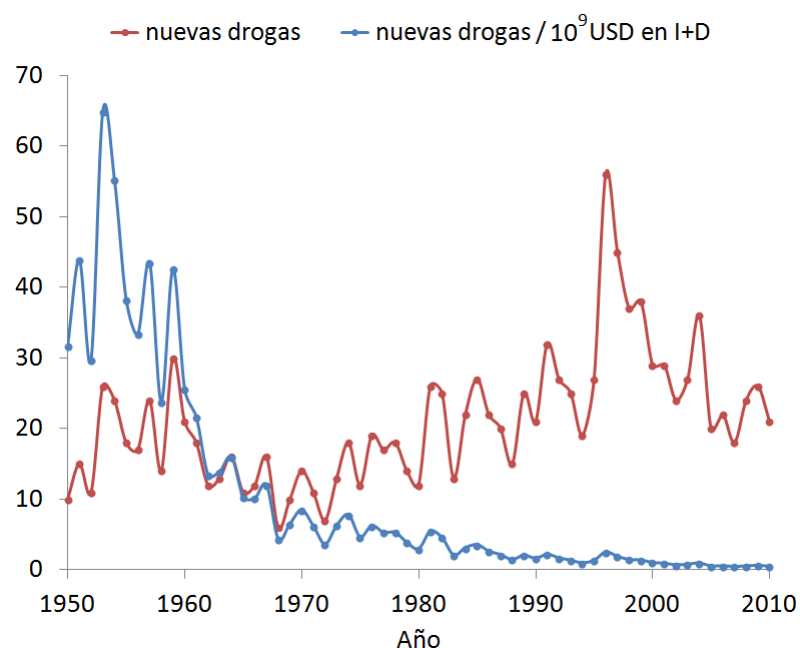
buscaban nuevos inhibidores del receptor quinasa de TGF- $\beta$ 1, un grupo en Eli Lilly usó HTS tradicional, mientras que en Biogen Idec utilizaron *screening* virtual basado en las interacciones estructurales entre un inhibidor débil y el receptor.<sup>8</sup> El grupo de Biogen Idec identificó 87 *hits*, siendo el mejor de ellos idéntico en estructura al compuesto descubierto en Eli Lilly, pero a un costo y una carga de trabajo muy reducidos. Incluso más rotundo resulta el caso de búsqueda de inhibidores de la girasa de ADN en Roche, donde los ensayos de HTS no fueron capaces de producir ningún compuesto líder apropiado más allá de los conocidos, mientras que por *screening* virtual obtuvieron una serie de *hits* que pudieron ser luego optimizados.<sup>48</sup>

Como se observa en estos y otros casos<sup>49</sup>, la tasa de acierto, definida como la cantidad de compuestos que se unen al blanco en una concentración particular dividida por la cantidad de compuestos probados experimentalmente, es de 100 a 1.500 veces más alta en *screening* virtual que en HTS. Por ello, el descubrimiento de fármacos asistido por computadora ha surgido como una forma de disminuir significativamente la cantidad de compuestos necesarios para ensayar experimentalmente, al tiempo que se mantiene el nivel de descubrimiento de compuestos líderes. Esto redunda en un costo mucho menor para la alternativa virtual. Además, el número de compuestos disponibles para realizar *screening* virtual excede el número que se pueden ensayar por HTS. Si bien los ejemplos anteriores muestran que la química computacional y el modelado molecular proporcionan herramientas para dirigir y aumentar la eficiencia del *screening*, tanto la metodología virtual como la experimental son herramientas que pueden ser utilizadas en paralelo, ya que los *hits* de una y otra suelen ser disímiles y complementarios.<sup>47</sup>

A pesar de los casos recién ejemplificados, al día de hoy el éxito en campañas de *screening* virtual sigue siendo muy dependiente del tipo de blanco bajo estudio y la función de puntaje elegida para *rankear* los diferentes compuestos a ensayar.<sup>50-53</sup> Esto representa un claro problema a la hora de embarcarse en la elección de un blanco y un método de *docking* apropiado. Nuestra estrategia buscará ampliar las capacidades de los métodos de *docking* en relación con este punto. Se desarrollará entonces una función de puntaje diferenciada para cada blanco proteico, la cual se aplicará en la búsqueda conformacional y en la estimación final de la energía libre de unión. El objetivo es aprovechar tanto las capacidades de los algoritmos de *docking* ya establecidos como la información específica de las interacciones relevantes que se encuentren en cada proteína particular.

### 1.1.3.5. Uso de métodos computacionales en el descubrimiento de fármacos

A principios de la década de 1980 ya existía un fuerte interés en el potencial del diseño de fármacos asistido por computadora. De hecho en 1981 la revista *Fortune* publicó en su portada un artículo titulado "La próxima revolución industrial: diseño de drogas por computadora en Merck".<sup>54</sup> En lo que refiere al desarrollo basado en estructura, el crecimiento exponencial de estructuras depositadas en el Protein Data Bank (PDB)<sup>25</sup> tuvo enormes consecuencias (Figura 1.4). El PDB es una base de datos de estructura tridimensional de macromoléculas biológicas que cuenta actualmente con más de 130.000 estructuras de proteínas y ácidos nucleicos, así como también de sus complejos con lípidos, azúcares y diversos ligandos entre los que se incluyen los compuestos tipo droga. Todas las estructuras depositadas en el PDB fueron obtenidas experimentalmente, en gran medida por difracción de rayos X, seguido por modelos de RMN. En los últimos años también fueron apareciendo más estructuras obtenidas por microscopía electrónica. Hoy se requieren al menos tres órdenes de magnitud menos horas de trabajo para obtener la estructura tridimensional de una proteína mediante cristalografía de rayos X<sup>55</sup> y las bases de datos de estructura tienen 300 veces más entradas que hace 25 años (ver Figura 1.4), lo que facilita en gran medida la identificación de compuestos líderes a través de estrategias basadas en estructura. A lo largo de todo este período hubo incontables casos en que el diseño de drogas asistido por computadora fue determinante en el descubrimiento de compuestos que superaron ensayos clínicos y se convirtieron en terapias novedosas para el tratamiento de enfermedades. El captopril, un inhibidor de la enzima convertidora de angiotensina aprobado en 1981 como un fármaco antihipertensivo, fue el primer fármaco derivado de un modelo estructural de la proteína blanco, siguiendo estudios de relación estructura actividad (SAR) sobre los ligandos.<sup>56,57</sup> La dorzolamida, inhibidor de la anhidrasa carbónica aprobado en 1995 para el tratamiento del glaucoma, fue el primer fármaco obtenido por diseño basado en estructura a partir de estructuras de rayos X del blanco y sus complejos.<sup>56,58</sup> Tal vez los casos más conocidos de éxito de las metodologías computacionales para encontrar nuevos compuestos tipo droga se relacionan con el VIH.<sup>59,60</sup> En 1995 y 1996 se aprobaron tres inhibidores de la proteasa del VIH, saquinavir, ritonavir e indinavir. Los tres fueron obtenidos por diseño basado en estructura y cumplieron un rol fundamental en revertir las muertes crecientes por SIDA.<sup>54</sup> En particular, el *screening* virtual de bases de datos de compuestos químicos basado en *docking*, central en esta tesis, ha identificado nuevos ligandos para más de 50 receptores con estructura conocida o incluso, en algunos casos, modelada por computadora.<sup>61</sup>



**Figura 1.5.** En rojo se muestra el número de nuevas drogas aprobada por la FDA cada año y en azul se muestra el mismo número por cada 1.000 millones de dólares invertidos en investigación y desarrollo (I+D) en los Estados Unidos.<sup>62</sup>

A las regulaciones estrictas se suma la existencia de un catálogo cada vez más amplio y mejorado de medicamentos aprobados, lo cual aumenta la complejidad del proceso de desarrollo de nuevos medicamentos. Es por ello que aún queda mucho trabajo por hacer y sobre el cual el diseño de drogas asistido por computadora puede aportar de manera significativa. En particular, las técnicas basadas en estructura resultan promisorias porque las estimaciones del número total de posibles blancos farmacológicos en el genoma humano<sup>65</sup> es alrededor de 3.000, mientras que el número de blancos ya atacados por drogas<sup>66</sup> ronda los 550, lo que demuestra que todavía quedan muchos blancos por explotar.

**PREGUNTAS:**

- 1) ¿Qué es el acoplamiento molecular y cómo se usa en el descubrimiento de fármacos?
- 2) ¿Cómo predicen los algoritmos de acoplamiento molecular la afinidad de unión entre un ligando y una proteína?
- 3) ¿Cuáles son las ventajas y limitaciones del acoplamiento molecular como herramienta para el descubrimiento de fármacos?
- 4) ¿Qué papel juega la flexibilidad de los ligandos y de las proteínas en el acoplamiento molecular?
- 5) ¿Qué desafíos quedan en el campo de acoplamiento molecular, y ¿cuáles son algunas soluciones potenciales?
- 6) ¿Cómo podría combinarse el acoplamiento molecular con otros métodos computacionales y experimentales para mejorar la descubrimiento?