



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS QUÍMICAS.

ESTUDIOS DE RECONOCIMIENTO MOLECULAR DE AMIDAS DERIVADAS DEL
ÁCIDO CINÁMICO Y FENIL ACÉTICO Y SU POSIBLE BLANCO BIOLÓGICO
(APOPTOSIS)

INFORME DE INVESTIGACIÓN
PARA OPTAR POR EL GRADO DE
MAESTRO EN CIENCIAS

PRESENTA:

Q.I. ALDO YOSHIO ALARCÓN LÓPEZ

DR. ENRIQUE ÁNGELES ANGUIANO
FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLAN

CIUDAD DE MÉXICO, OCTUBRE 2020



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS QUÍMICAS.

ESTUDIOS DE RECONOCIMIENTO MOLECULAR DE AMIDAS DERIVADAS DEL
ÁCIDO CINÁMICO Y FENIL ACÉTICO Y SU POSIBLE BLANCO BIOLÓGICO
(APOPTOSIS)

INFORME DE INVESTIGACIÓN
PARA OPTAR POR EL GRADO DE
MAESTRO EN CIENCIAS

PRESENTA:

Q.I. ALDO YOSHIO ALARCÓN LÓPEZ

DR. ENRIQUE ÁNGELES ANGUIANO
FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLAN



CIUDAD DE MÉXICO, OCTUBRE 2020

Jurado Asignado

Presidente

Dr. Rodolfo Gómez Balderas

Vocal

Dra. Karina Martínez Mayorga

Vocal

Dr. Alfonso Sebastián Lira Rocha

Vocal

Dra. Catalina Soriano Correa

Secretario

Dra. Laura Domínguez Dueñas

El presente trabajo se realizó en el Laboratorio de Química Medicinal de la Facultad de Estudios Superiores Cuautitlán, Campo I, de la Universidad Nacional Autónoma de México, bajo tutela del Dr. Enrique Ramón Ángeles Anguiano, con el apoyo de la beca para maestría otorgada por el Consejo Nacional de Ciencia y Tecnología CONACYT (No. de registro becario: 878116)

Agradecimientos

A mis padres, que han sido mi mayor ejemplo de fortaleza, constancia y entrega, Amparo López y Miguel Alarcón †, quienes con todo el cariño y respeto me mostraron el mejor camino a seguir, que me enseñaron la importancia de todos los valores, que hoy me definen como persona, la importancia del trabajo para salir adelante, y lo imparables que se puede llegar a ser si estamos unidos, y que no hay batalla que te pueda vencer hasta que no puedas más, siempre hay algo por lo que seguir en pie.

A mis hermanas, que con tanta paciencia me ayudaron en muchos sentidos, dieron su ejemplo de virtud y responsabilidad, que consolaron mis tristezas, y que apoyaron cada una de las decisiones que finalmente me han traído aquí. Sigamos juntos y creciendo, como dice aquél poema: "Lo que el árbol tiene de florido, vive de lo que tiene sepultado".

A la Universidad Nacional Autónoma de México, que me ha brindado cada espacio para poder llevar a cabo mis proyectos, que no sólo es mi casa, si no la de muchos otros jóvenes que con anhelos se esfuerzan a diario por tener más oportunidades y crecer profesionalmente, y con ello, ayudan a mejorar al país.

A la Facultad de Estudios Superiores Cuautitlán, donde pude conocer a personas que valen mucho la pena, las alegrías y las tristezas, que son colegas, profesores y amigos, todo contacto deja trazas, y que sean éstas las que dejan marca sobre lo que seremos más adelante.

Al Consejo Nacional de Ciencia y Tecnología, por los fondos otorgados para poder concluir el presente trabajo (número de registro de becario: 878116). Que la ciencia siga siendo la herramienta fundamental para mejorar la calidad de vida de las personas.

A la Unidad de Posgrado de la universidad, que con tanta paciencia y amabilidad dieron seguimiento al proceso de graduación, y que siempre trabajan con total entrega a los alumnos.

Al Dr. Enrique Ángeles, por todo el apoyo durante el proyecto, que es un ser humano de calidad ejemplar, que sabe escuchar a los alumnos y que nos guía a todos dentro del laboratorio de química medicinal, sin limitar la capacidad e inquietudes de todos los que colaboramos en el mismo. Su inteligencia y constancia son un ejemplo a seguir

A los miembros del H. jurado. Dr. Alfonso Lira Rocha, Dr. Rodolfo Gómez Balderas, Dra. Karina Martínez Mayorga, Dra. Catalina Soriano Correa y Dra. Laura Domínguez Dueñas. Por todo el tiempo invertido en la revisión de este trabajo, cada una de las observaciones ayudó a dar un poco más de calidad a la presentación de los resultados del proyecto.

A todo el LQM. A la Dra. Ana María Velázquez, que con tanto cariño nos ha sabido acompañar y guiar, que tanto nos ayuda. A Alejandro Serda que siempre tiene la buena intención de ayudar, que es intrépido y atrevido, nunca dejes de experimentar cosas nuevas, eventualmente traen algo bueno. A Elizabeth Rul, que siempre escuchas y propones cosas con sensatez, que analizas todo y sabes aconsejarnos. Al Dr. Pablo Martínez Soriano, que tu inteligencia y amistad son algo que hay que atesorar, al Dr. Víctor Vázquez Valadez, sin duda tu determinación y franqueza hace que mantengamos los pies en la tierra, sigue creciendo mucho. A Pablo Aguirre, que con tu alegría siempre contagias el entusiasmo por el trabajo. Y al resto del equipo por estar pendiente siempre del trabajo, aún queda mucho por hacer para mejorar.

Índice

1. Objetivos	1
2. Introducción	2
3. Antecedentes	2
3.1. Cáncer	2
3.1.1. Características de las células tumorales.	2
3.1.2. Tipos de cáncer	3
3.1.3. Inducción de cáncer y factores de riesgo	4
3.1.4. Tratamientos contra el cáncer	5
3.1.5. Fármacos dirigidos a «dianas terapéuticas»	6
3.2. Modelado Molecular	7
3.2.1. Dinámica molecular.	7
3.2.2. Reconocimiento molecular (Docking) y desarrollo de fármacos	9
3.2.3. Minería de datos	11
3.2.4. Pesca de objetivos (Target fishing)	11
3.2.5. Servidores para target fishing por similitud.	12
3.2.6. La serie LQM-700.	13
4. Metodología.	15
4.1. Búsqueda de dianas biológicas	15
4.2. Estudios de reconocimiento molecular	15
4.3. Dinámica molecular.	17
5. Resultados y discusión	18
5.1. Blancos biológicos	18
5.2. Reconocimiento molecular	18
5.3. Dinámica Molecular	20
6. Conclusiones	23
7. Perspectivas	24
8. Bibliografía	25

Resumen

Diferentes estudios de reconocimiento molecular (docking) y de dinámica molecular fueron empleados para la propuesta de blancos biológicos, que pueden tener interacción con una serie de compuestos con actividad antineoplásica. Derivados del éster fenílico de los ácidos caféico y cinámico (serie LQM-700). Los blancos biológicos fueron hallados en una búsqueda por similitud de las estructuras de los compuestos LQM-700 utilizando diversos servidores de búsqueda comparativa como Target Hunter, Similarity Ensemble Approach (SEA), Drug Target Explorer, y Swiss Target Predictor.

Con base en los resultados de estudios previos dentro del laboratorio de química medicinal de la FESC-UNAM sobre la actividad antineoplásica de la serie LQM-700, es decir el porcentaje de inhibición del crecimiento celular en cultivos celulares PC-3, HeLa y MBA-Hb en conjunto con el uso de los servidores, se propusieron seis blancos biológicos como los más probables: MMP9, AIF, CB1, AKR1C3, AKR1B10, y MAP14K. Para cada uno de los blancos propuestos, los estudios de docking y dinámica molecular se llevaron a cabo para modelar las interacciones tridimensionales ligando-receptor, así como de la estabilidad de los complejos formados, con el fin de obtener aquél blanco biológico que con base en los resultados, se pueda proponer para estudios *in vitro* posteriores y validar así los experimentos computacionales.

1. Objetivos

Objetivo general.

Proponer blancos biológicos relacionados con la actividad antineoplásica de los compuestos de la serie LQM-700, mediante la utilización de herramientas computacionales disponibles en servidores de búsqueda por similitud, además de software de modelado molecular, como el reconocimiento molecular (docking) y dinámica molecular, que den índices suficientes para proponer al blanco biológico más probable de interaccionar.

Objetivos particulares

- Utilizar servidores de búsqueda por similitud para hallar blancos biológicos candidatos para la interacción con los compuestos LQM-700 y analizar los datos obtenidos para hallar aquéllos que estén relacionados con la biología molecular del cáncer.
- Obtener de bases de datos cristalográficos, Protein Data Bank (PDB), las estructuras que correspondan a los blancos biológicos encontrados en la búsqueda por similitud con resolución aceptable y preferentemente con un sitio activo ocupado con algún ligando.
- Realizar los estudios de reconocimiento molecular (Docking) en MOE para hallar las interacciones y el modo de acoplamiento entre los ligandos disponibles (LQM-700's) y los blancos propuestos derivados de la búsqueda por similitud, para determinar los mejores complejos ligando receptor basado en sus propiedades de energía calculadas.
- Preparar los mejores modelos obtenidos del reconocimiento para realizar una dinámica molecular y analizar los resultados de la misma con el fin de obtener los datos de Root Mean square deviation(RMSD) y radio de giro como principales variables de análisis y correlación con la estabilidad de los complejos.
- Proponer el blanco propuesto más favorecido en las simulaciones moleculares y el reconocimiento molecular para su posterior validación experimental.

2. Introducción

El uso de herramientas *in silico* en los últimos años ha servido para la exploración a nivel molecular de diversas teorías sobre el mecanismo de acción de los medicamentos, candidatos a fármacos o de compuestos con una actividad biológica específica, dando resultados que ayudan a la optimización estructural de los mismos para minimizar los efectos adversos que su administración provoca. El uso de estas herramientas tiene dos objetivos principales: optimizar la geometría de los compuestos con la actividad biológica y conocer el proceso de acoplamiento entre el ligando y el receptor correspondiente.

El presente trabajo pretende proponer un método para hallar a los blancos biológicos que se relacionan con el proceso de muerte celular (apoptosis) en células cancerosas. Esto se lograría al combinar técnicas de quimioinformática que realizan un algoritmo de comparación y búsqueda de moléculas en bases de datos disponibles y posteriormente dar una lista de blancos biológicos con una alta probabilidad de interacción con un ligando o lista de ligandos cuya actividad biológica esté reportada, pero que se desconozca su modo de acción.

Una vez obtenidos los blancos biológicos, la siguiente tarea es realizar estudios de reconocimiento molecular, para que con base en ésta herramienta computacional que tiene sus bases en la investigación progresiva sobre campos de fuerza y teorías de estructura electrónica, se obtengan datos comparables para poder discernir entre la más alta probabilidad de interacción de un ligando con su respectivo blanco biológico, dichos resultados nos darían una idea de la estabilidad de los complejos proteína-ligando formados dependiendo de su energía de afinidad.

Hallados los complejos con mayor afinidad, se realiza la preparación para una simulación de dinámica molecular, y se calculan el *Root Mean Square Deviation (RMSD)*, que es una medida de la desviación de la posición de los átomos con respecto a una inicial, *Root mean square fluctuation (RMSF)* que es la fluctuación sobre las posiciones promedio de cada aminoácido y Radio de giro que es una medida de la compactibilidad de una proteína, las cuales dan índices cuantitativos sobre la estabilidad y equilibrio del modelo simulado, para finalmente proponer a aquél blanco biológico que pueda ser evaluado experimentalmente, ahorrando así tiempo y recursos destinados al descubrimiento de nuevas dianas terapéuticas.

3. Antecedentes

3.1. Cáncer

El cáncer es una de las enfermedades que más afecta a la población anivel mundial, siendo la segunda causa de muerte en Europa y América [Kumar, 2015]. Lo vuelve a la búsqueda de tratamientos nuevos una de las preocupaciones principales en las áreas de investigación en materia de salud. El cáncer es una afección caracterizada principalmente por la alteración de uno o varios tipos de células durante algún proceso del ciclo celular.

Los cambios fisiológicos se presentan de forma progresiva, y se acompañan de modificaciones genéticas que le proporcionan a las células ventajas evolutivas y permiten su diseminación en el organismo [Gil-Bao et al., 2015].

3.1.1. Características de las células tumorales.

De acuerdo a Hanahan y Weinberg [Hanahan and Weinberg, 2011], las células tumorales poseen ciertas características que adquieren cuando dejan de ser células sanas en el proceso de neoplasia. Las características llamadas clásicas, esquematizadas en la Figura 1, se describen a continuación [Gil-Bao et al., 2015]:

1. Mantenimiento de las señales de proliferación

Las células tumorales consiguen mantener una replicación celular aumentada y autónoma mediante

la alteración de las señales de proliferación. En este sentido, son muchas las anomalías que pueden presentar: expresión alterada de factores de crecimiento, activación constitutiva de las vías de señalización intracelular, estimulación autocrina/paracrina que provoca la proliferación de las células del estroma y que redundan en una estimulación de las células tumorales y otros resultados.

2. **Evasión de la inhibición de la proliferación.**

Las células tumorales pierden la capacidad de inhibición de la proliferación dependiente de contacto, en las células normales existe una relación entre su número y la cantidad de factores de crecimiento presentes en el medio, de manera que dejan de dividirse al alcanzar una densidad celular determinada. Sin embargo, en las tumorales, dicha inhibición se encuentra alterada.

3. **Capacidad de invadir y metastatizar**

En tumores avanzados, la adhesión entre células cancerosas vecinas está disminuida debido a la adquisición de un fenotipo mesenquimal migratorio, lo que significa que hay una sobreexpresión de enzimas extracelulares y una menor expresión de moléculas de adhesión, como la cadherina E (lo que les da su estabilidad estructural). Esto permite que las células sean capaces de desprenderse de un conjunto y así invadir tejidos cercanos y/o metastatizar a órganos distantes a través del torrente sanguíneo.

4. **Inducción de angiogénesis.**

El crecimiento tumoral debe adquirir un fenotipo angiogénico que induzca la formación de nuevos vasos sanguíneos para cubrir el aporte de oxígeno y de nutrientes y la eliminación de dióxido de carbono y sustancias de desecho. La falta de neovascularización conduciría a la necrosis del tumor, por lo que los tumores necesitan inducir la angiogénesis para su desarrollo.

5. **Capacidad ilimitada de replicación.**

En las células no tumorales, las sucesivas mitosis provocan un acortamiento progresivo de los telómeros, de forma que, tras un número determinado de ciclos, la célula pierde su capacidad de replicarse. Sin embargo las células tumorales expresan la enzima telomerasa, que añade secuencias repetitivas de ADN telomérico y mantiene constante la longitud de los telómeros. La célula adquiere así una capacidad ilimitada de replicarse, sin entrar en senescencia.

6. **Resistencia a la muerte celular programada**

Las células tumorales son capaces de evadir la muerte por apoptosis, a pesar de acumular frecuentemente aberraciones genéticas y de estar sometidas a un ambiente adverso. Muchas células tumorales aumentan la expresión de factores antiapoptóticos, como Bcl-2, e inhiben la de factores proapoptóticos como Bax o Bim, por lo que alteran el balance normal que se requiere para inducir la muerte celular.

3.1.2. **Tipos de cáncer**

La clasificación de los tumores puede hacerse de diversas formas. Según su comportamiento o procedencia. Según éste, los tumores pueden ser benignos o malignos, los benignos tienen tasas de crecimiento controladas y no suelen invadir otros tejidos. Los tumores malignos presentan una mayor indiferenciación y un ritmo alto de crecimiento, pueden atravesar la membrana basal e invadir tejidos locales y diseminarse a distancia y provocar metástasis.

Como los tumores se pueden formar partiendo de distintos tipos celulares, se clasifican de acuerdo a su procedencia basados en el tipo histológico del que derivan. Así pues pueden derivar de tejido epitelial, mesenquimal, linfoide, nervioso o de células germinales. Para denominar a un tumor se usa la raíz del tejido de origen seguido del sufijo *-oma*. Los tumores benignos por células epiteliales se llaman papilomas o pólipos. Los que derivan de tejido glandular se denominan adenomas, los de origen epitelial se llaman carcinomas y los no epiteliales conocidos como sarcomas.



Figura 1: Para que una célula normal sufra el proceso de neoplasia para convertirse en una célula tumoral, ésta puede tener características bien definidas; partiendo de un daño genético se activan las mismas, como el mantenimiento de la proliferación, lo que lleva a un crecimiento incontrolado de tejido, mismo que necesita sangre para obtener nutrientes y es por eso que la angiogénesis es una pieza clave para regular el crecimiento de un tumor, así como la resistencia a la muerte celular por apoptosis y finalmente la capacidad de invadir tejido cercano o metastatizar a órganos lejanos, entre otras características.

3.1.3. Inducción de cáncer y factores de riesgo

La causa directa del cáncer no es solamente un factor determinante, la carcinogénesis es un proceso mediante el cual se desarrolla el cáncer y consta de varias etapas en las que se producen mutaciones y selección clonal de células que adquirirán una capacidad mayor de proliferación, supervivencia, invasión y metástasis.

La génesis del cáncer, suele comenzar con una o varias alteraciones genéticas que sin ser letales provocan cambios posteriores en las células. Éstas pueden ser heredadas, o bien ser producidas por agentes carcinógenos, los cuales pueden ser de tipo físico, químico o incluso biológico. Los carcinógenos físicos se refieren a un amplio rango de agentes: radiaciones electromagnéticas de diferentes tipos, corpusculares (alfa y beta), temperaturas extremas, traumas mecánicos y algunos tipos de materiales sólidos o en gel. Ambos tipos causan daños en el ADN que deben ser reparados en algún punto de control del ciclo celular antes de que la mitosis tenga lugar, de otro modo las células hijas heredarán las alteraciones genéticas que favorecerían la formación de un tumor. [Soffritti et al., 2003] Los carcinógenos de tipo químico pueden ser compuestos con potencial para inducir cáncer, aunque desafortunadamente dicho efecto no se descubre si no hasta después de varios años en contacto con dichas sustancias, ejemplo de lo anterior son los casos de tabaco y alcohol que inducen neoplasias en pulmón e hígado respectivamente. Asimismo existen compuestos carcinógenos que representan riesgos laborales en diversas áreas de la manufactura de productos diversos. [Gil-Bao et al., 2015] Además hay agentes biológicos que son capaces de inducir cáncer como virus, bacterias y parásitos. Probablemente *Helicobacter pylori* es el caso más destacable, ya que se estima que hasta el 30 % de la población tiene la mucosa gástrica colonizada por *H. pylori*, y no todos ellos tendrán manifestaciones clínicas, pero su presencia provoca gastritis crónica que conlleva a una disminución de las células D de la mucosa, que secretan somatostatina, hormona que inhibe la secreción de gastrina, por lo que hay una elevación en la producción de ésta, la cual estimula la secreción de ácido clorhídrico. La inflamación crónica en una célula hace que se produzcan radicales libres que pueden provocar daños en el ADN de las células y derivar en adenocarcinoma gástrico. [Gil-Bao et al., 2015]

3.1.4. Tratamientos contra el cáncer

Tratamientos clásicos.

Éstos incluyen a la cirugía para la extracción del tejido, la radioterapia (como resultado del desarrollo de la física en el siglo XX), y la quimioterapia, compuestos químicos encontrados en plantas en un principio, pero que con el paso del tiempo y el desarrollo de mejores técnicas para la investigación y evaluación de dichos compuestos ha cambiado el origen y eficacia de tales agentes.

Los compuestos usados en la quimioterapia convencional ejercen su acción fundamentalmente sobre las células del organismo que presentan un alto índice de proliferación, dichos fármacos no sólo poseen la actividad antitumoral, pues tienen la desventaja de no ser específicos contra las células tumorales, y por tanto son capaces de dañar a otras células en proceso de división activa y que resulta en efectos secundarios, como la disminución de glóbulos rojos, blancos y plaquetas en sangre, afectación de las células del intestino y folículos pilosos, etc. Los tratamientos clásicos se presentan en la Figura 2

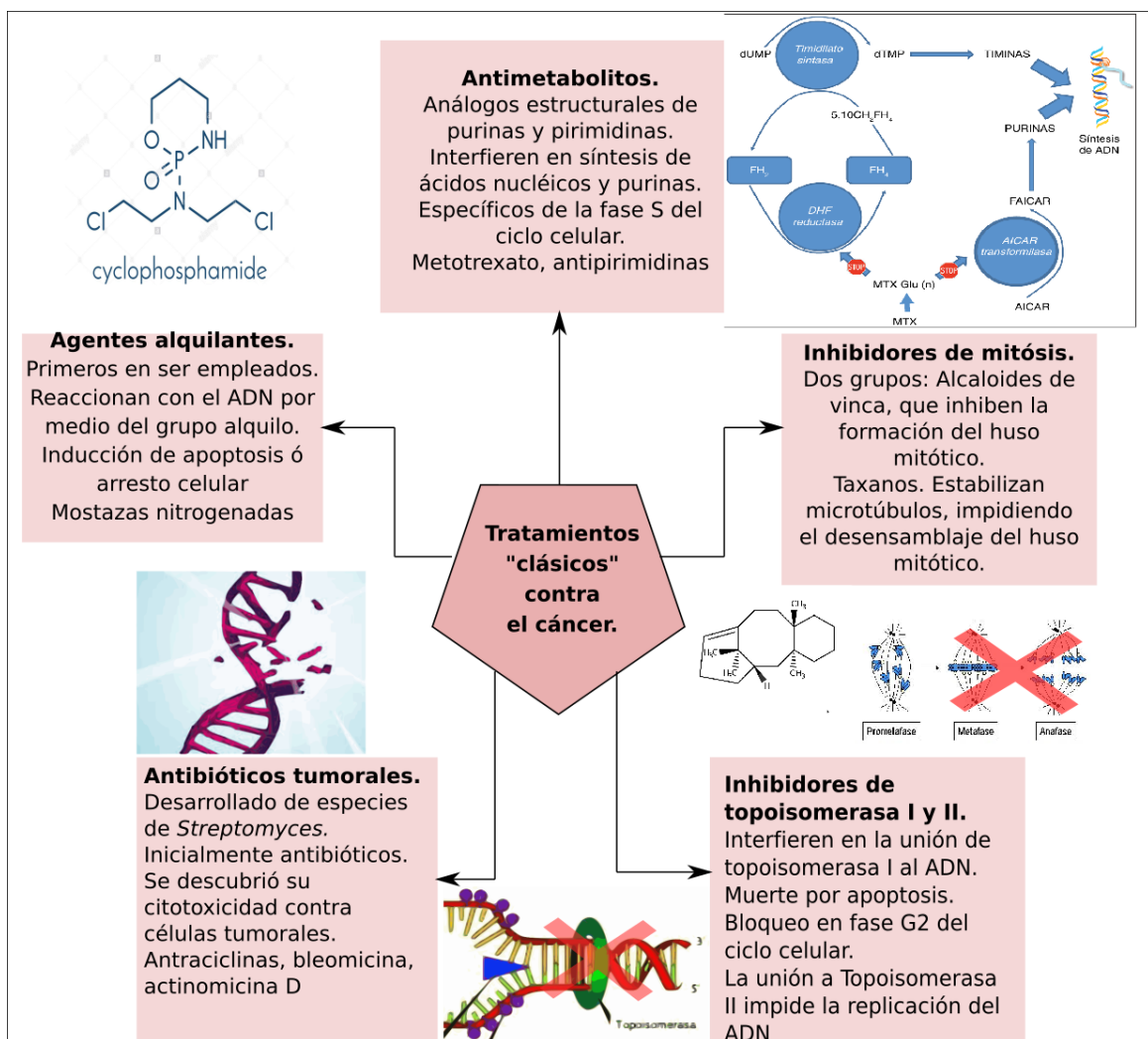


Figura 2: Tratamientos contra el cáncer: Se muestra la clasificación de los actuales tratamientos contra el cáncer, cuya acción depende del tipo de receptor al que se une y que produce diferentes expresiones fenotípicas.

Agentes alquilantes. Fueron los primeros fármacos empleados de manera eficaz contra el cáncer, siendo un grupo heterogéneo de medicamentos que reaccionan con el ADN por medio del grupo alquilo, generando

roturas en el ADN, resultando en la inducción de apoptosis o arresto en algún punto del ciclo celular. Los grupos de fármacos que constituyen esta familia son las mostazas nitrogenadas (melfalán, clorambucilo, ciclofosfamida, ifosfamida), las tetracinas, y derivados del cisplatino. Sus efectos secundarios característicos están relacionados fundamentalmente con el desarrollo de aplasia medular, y la aparición de mucositis, además de nefrotoxicidad y neurotoxicidad.

Antimetabolitos Son análogos estructurales de las purinas y pirimidinas, que intervienen en la síntesis de los ácidos nucleicos. Durante la replicación del ADN se incorporan al metabolismo celular y pasan a formar parte de las cadenas de ADN y ARN, generando una codificación incorrecta o dando lugar a roturas de sus cadenas. Inhiben enzimas necesarias para la síntesis de ADN, y por tanto son específicos de la fase S del ciclo celular. La fase S del ciclo celular se caracteriza precisamente por la síntesis (de ahí que sea S) de ácidos nucleicos para posteriormente pasar a una fase de espera llamada G₂. Los fármacos más relevantes son los antagonistas del folato como metotrexato, los arabinósidos y antipirimidinas.

Inhibidores de la mitosis. Existen dos grupos de fármacos antineoplásicos dentro de esta familia: los alcaloides de la vinca (vincristina, vinblastina y vinorelbina), empleados en el tratamiento de cáncer de mama y linfomas, y los taxanos (paclitaxel, docetaxel, abraxane) de uso contra tumores de próstata, mama, pulmón, ovario y estómago. Su mecanismo de acción difiere, los alcaloides de la vinca inhiben la formación del huso mitótico, mientras que los taxanos estabilizan los microtúbulos, impidiendo el desensamblaje del huso mitótico. Los efectos secundarios de este grupo farmacológico comprenden la mielosupresión y neurotoxicidad.

Inhibidores de la topoisomerasa. Estos interfieren en la unión de la topoisomerasa I al ADN, bloqueando su función y originando roturas irreversibles de doble cadena, por lo que la mayoría de las células mueren por apoptosis o serán bloqueadas en la fase G₂ del ciclo celular. Los inhibidores de la topoisomerasa II, empleados en el tratamiento de tumores microcíticos de pulmón y en tumores de testículo, impiden la replicación de ADN, manteniendo las roturas de doble cadena. Los efectos adversos más característicos son la diarrea y la mielosupresión en el caso de los inhibidores de la topoisomerasa I y la mielosupresión en los inhibidores de la Topoisomerasa II.

Antibióticos antitumorales. Este grupo de compuesto citotóxicos se desarrolló a partir de especies de *Streptomyces* en cultivo y concebidos como antibióticos, pero después de un tiempo se descubrió su citotoxicidad contra células tumorales. En este grupo se consideran a las antraciclinas y derivados, la bleomicina y la actinomicina D. Son empleados en el tratamiento de adenocarcinomas de mama, sarcomas o tumores testiculares. [Gil-Bao et al., 2015]

3.1.5. Fármacos dirigidos a «dianas terapéuticas»

Gracias a los recientes avances del conocimiento de las vías de señalización molecular que regulan el crecimiento de las células cancerosas, y el desarrollo biotecnológico ha permitido la síntesis de nuevos medicamentos que pueden ser anticuerpos o que pueden atravesar la membrana plasmática y bloquear proteínas específicas intracelulares. Esto hace que se elaboren tratamientos oncológicos específicos de una *diana molecular*. Dichos fármacos emplean como lugar de acción específico y selectivo aquéllos rasgos moleculares en los que las células tumorales difieren de las que no lo son. Esta estrategia busca la mejora en la eficacia de los tratamientos de los pacientes con cáncer y una disminución de la toxicidad.

3.2. Modelado Molecular

Las representaciones simbólicas de la estructura de los compuestos químicos es una de las primeras formas de modelado molecular, teniendo su origen a mediados de el siglo XIX, en el desarrollo de la teoría estructural de la química orgánica. A partir de entonces, con el paso de los años, se consideró al modelo de August Wilhelm Hofmann en 1865 como el modelo seminal de esferas y resortes para la representación de las moléculas. El esquema de colores que estableció Hofmann, sigue siendo utilizado hasta el día de hoy con ligeras modificaciones.

Los campos de fuerza o *forcefields*, se originaron en el siglo XX derivados de la espectroscopia vibracional que consideraba a las fuerzas que actuaban entre cada par de átomos relacionándolos con vibraciones que respondían básicamente a la ley de Hooke para un oscilador armónico. Estos *forcefield* describen la dependencia de la energía del sistema con respecto a las coordenadas de las partículas que lo conforman y cuya expresión matemática contiene al conjunto de ecuaciones y parámetros que intervienen para describir las contribuciones energéticas intra e inter moleculares [Cedillo-Cruz, 2020].

El concepto de un *forcefield* no fue de gran influencia si no hasta 1946, con la sugerencia de que usando un método de modelado molecular más cuantitativo, basado en la combinación de interacciones estéricas y en la mecánica Newtoniana y más específicamente en la combinación de los modos normales de vibración *bond stretching*, *angle bending*, y *modos vibracionales de torsión*, el tiempo de cómputo de cada sistema sería menor y por tanto volvería a estos algoritmos de cálculo mucho más eficientes. A esto se le conocería como **mecánica molecular** [Schlecht, 1998]

Hoy en día existe una variedad de campos de fuerza, cuya elección depende del sistema a simular, es decir, hay aquéllos aplicables a todos los átomos (como el llamado universal force field, UFF), o algunos que se aplican a sistemas específicos como proteínas o ácidos nucleicos. Un ejemplo de estos es CHARMM, que sirve para modelar principalmente proteínas usando mecánica molecular, aunque también puede referirse a un programa computacional homónimo utilizado para realizar simulaciones, y cuya forma es:

$$U(R) = \sum_{\text{Enlaces}} K_b(b - b_0)^2 + \sum_{K_{ub}} K_{ub}(s - s_0)^2 + \sum_{\text{Angulos}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{Torsiones}} K_\chi(1 + \cos(n\chi - \delta)) + \sum_{\text{impropios}} K_{\text{impropios}}(\rho - \rho_0)^2 + \sum_{\text{Noenlazados}} \epsilon \left[\frac{R_{\text{min}ij}^{12}}{r_{ij}} - \frac{R_{\text{min}ij}^6}{r_{ij}} \right] + \frac{q_i q_j}{\eta r(ij)}$$

Donde el primer término es el efecto enlazante, el segundo las interacciones Urey-Bradley, las cuales consideran el efecto de vibración de pares de átomos 1,3 no enlazados, el tercero son los efectos de ángulos de enlace, el cuarto término las torsiones, el quinto los diedros impropios, el sexto término los efectos no enlazantes (o de largo alcance) calculados con un potencial de tipo Lennard-Jones y el último son las interacciones electrostáticas.

3.2.1. Dinámica molecular.

El estudio de estructuras macromoleculares mediante el modelado molecular se ha logrado gracias al desarrollo de la dinámica, que resulta ser un punto clave en el entendimiento de las funciones biológicas basadas en las interacciones moleculares. Desde que las determinaciones estructurales fueron posibles, ha habido un incremento notable en el conocimiento de proteínas y ácidos nucleicos. Hasta ahora en la base de datos *Protein Data Bank (PDB)* se albergan más de cien mil entradas entre proteínas y ácidos nucleicos, ya sea en complejo o no con otra molécula.

Las biomoléculas mencionadas son entidades flexibles, y la dinámica ayuda a comprender sus cambios conformacionales que suceden mientras la proteína lleva a cabo su función. Como regla se puede decir que un complejo de cualquier proteína implica un rearrreglo conformacional, lo que es común en el ciclo

catalítico de las enzimas. [Hospital et al., 2015]

No sólo se obtiene el movimiento de los átomos y un resultado meramente visual, la resolución de las ecuaciones de movimiento de Newton, aunado a los efectos modelados por el campo de fuerza seleccionado nos darán tres medidas importantes, las cuales se ilustran en la Figura 3[Martínez, 2015]:

RMSD (Root Mean Square Deviation) Es una medida del cambio de posición relativo, ya sea global o localizado de los átomos, o bien un desplazamiento promedio de los átomos con respecto a una posición relativa, en un instante. Si se tiene un valor de RMSD muy grande, es decir mayor a 2 Å, significa que la dinámica está deformando mucho a la proteína y los datos no son confiables, por tanto es necesario estabilizar la fluctuación de la estructura durante el tiempo requerido.

RMSF (Root Mean Square Fluctuations) Es una medida promedio del desplazamiento de un átomo particular, o de un grupo de átomos, relativo a la estructura referencia. Cuando la estabilización del sistema está bien logrado, entonces el cálculo de RMSF tiene sentido.

Radio de giro. Es una medida de cuan compacta es una proteína, también nos da información relevante sobre su conformación y de su estabilidad a través del tiempo durante la dinámica. [Khan et al., 2016].

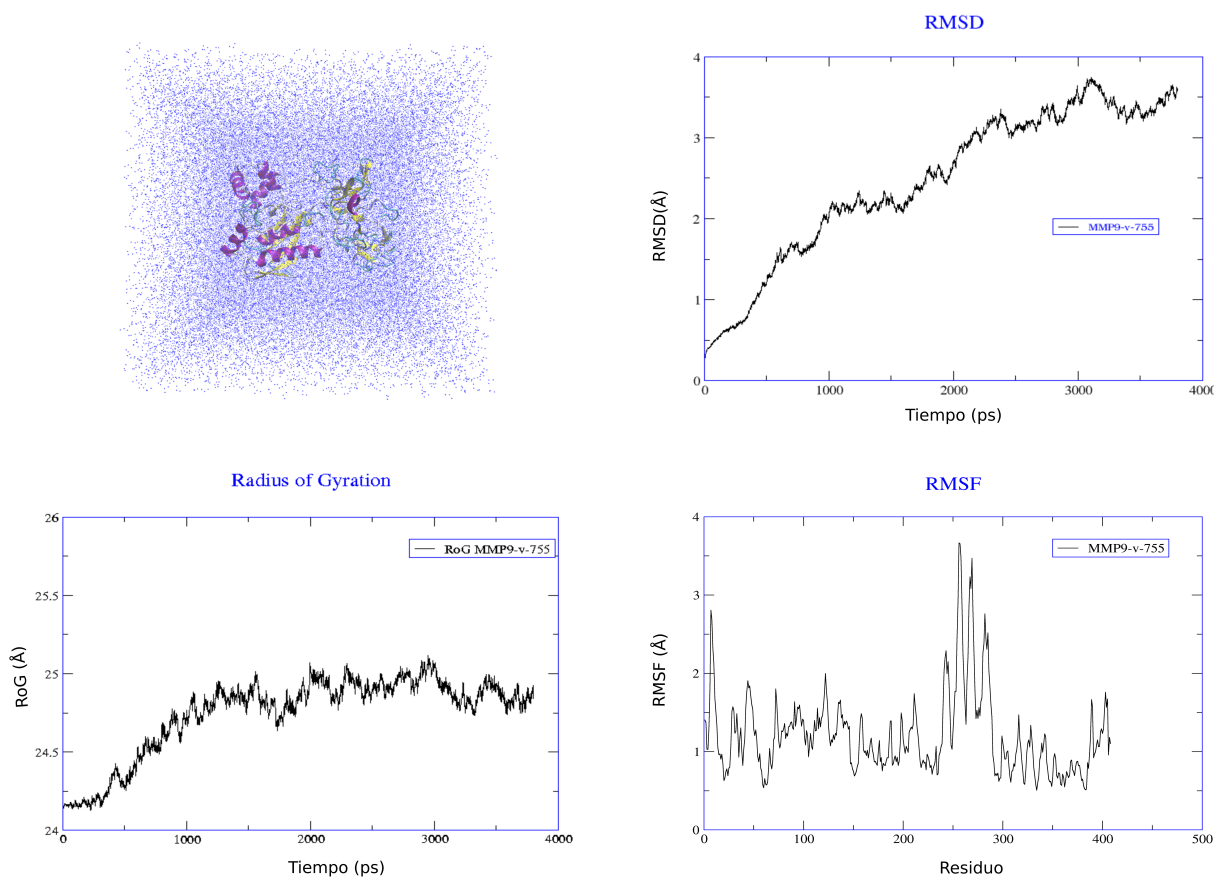


Figura 3: De una dinámica molecular, más que los resultados visuales del movimiento de los átomos, se obtienen tres medidas importantes. RMSD, RMSF y Radio de giro, que nos darán la información necesaria para interpretar la simulación y extrapolarla para la explicación de fenómenos visibles. (Ilustración adaptada de Khan et Al. (2016))

Simulaciones de dinámica molecular.

Fueron desarrolladas a finales de los 70's, y ha avanzado de simular algunos cientos de átomos con relevancia biológica hasta el modelado de proteínas completas con representaciones explícitas de disolvente, proteínas transmembrana, e incluso complejos macromoleculares como nucleosomas o ribosomas. Actualmente, sistemas de entre 50,000 y 100,000 átomos resultan rutinarios para muchas aplicaciones, siempre y cuando se cuente con los recursos computacionales.

El avance ha sido consecuencia del uso de computación de alta eficiencia (HPC) y la simplicidad del algoritmo básico de dinámica molecular. Es decir, una estructura experimental se obtiene o se realiza un modelo de homología, entonces el sistema puede ser representado en diferentes niveles de detalle, siendo atomístico, o grano grueso (*coarse-grained*), que se están volviendo populares en sistemas muy grandes. El disolvente explícito hace posible obtener los efectos de solvatación que reproduzcan la realidad incluyendo aquéllos de origen entrópico como el efecto hidrofóbico. Una vez que el modelo se construye, las fuerzas que actúan en cada átomo se obtienen de las ecuaciones de movimiento de Newton y el correspondiente campo de fuerza.

Los campos de fuerza son ecuaciones complicadas pero de cálculo sencillo, y se deben a la simplicidad de sus términos, es decir, tratando a los enlaces y ángulos de enlace como resortes, funciones periódicas para los diedros, potenciales de tipo Lennard-Jones para las interacciones de largo alcance, y la ley de coulomb para las interacciones electrostáticas hace que los cálculos de energía sean muy rápidos incluso para los sistemas grandes.

Existen distintos campos de fuerza, que difieren en el modo por el cual están parametrizados. Los parámetros no son necesariamente intercambiables y no todos los campos de fuerza permiten todos los tipos de moléculas, por ello es importante elegir el adecuado. Una vez que las fuerzas que actúan sobre los átomos del sistema se definen, las ecuaciones de Newton ayudan a definir las aceleraciones y velocidades de las partículas (átomos) para actualizar su posición. Como la integración del movimiento se realiza de modo numérico, para evitar la inestabilidad se utiliza el integrador de Verlet, en rangos que van de uno a dos femto segundos. Sin embargo, las simulaciones que son importantes en la escala de procesos biológicos están en el orden de micro segundos. Esto significa que se requiere de un costo computacional significativo para simular este tipo de sistemas, para solucionar esta desventaja, existen métodos de tipo *coarse grained* donde se disminuye el número de partículas, haciendo que el número de interacciones a calcular también sea menor.

La generación actual de computadoras toma ventaja del paralelismo y aceleradores para optimizar el proceso. Los códigos de simulación más populares (AMBER, CHARMM, GROMACS, NAMD) han sido compatibles con lenguajes de multi procesadores (MPI) para disminuir el tiempo de cálculo. Pero más recientemente, las unidades de procesamiento gráfico (GPU's) son las que dan un tiempo de cálculo menor a las simulaciones. El uso combinado de MPI/GPU da robustez y mucho alcance al método en sí. [Hospital et al., 2015]

3.2.2. Reconocimiento molecular (Docking) y desarrollo de fármacos

Su aplicación principal, tanto para moléculas pequeñas o proteínas, es explorar el modo en el cual un ligando se une a su sustrato de manera típica, de donde se obtiene información importante y que es la base del diseño de fármacos basados en la estructura. El proceso de reconocimiento molecular es por naturaleza dinámico; las moléculas son entidades flexibles y por tanto el reconocimiento involucra re-arreglos estructurales, que afectan de modo observable las propiedades energéticas del complejo formado. La contribución de las técnicas experimentales para la determinación de estructuras y la purificación de muchas proteínas han permitido a las estrategias computacionales o *in silico* evolucionar en el aspecto del descubrimiento de fármaco. Para el diseño basado en la estructura, el reconocimiento molecular se convierte en una de las herramientas de uso común en los grupos de investigación desde los años ochenta.

El proceso involucra dos pasos esenciales: el primero es la predicción de la conformación del ligando, así

como de la posición y la orientación en los sitios disponibles de la proteína (llamados poses) y el segundo es el cálculo de la afinidad de acoplamiento. Ambos se realizan con base en métodos de muestreo y de registro (*scoring functions*) propios del algoritmo de reconocimiento. Y de los cuales se hablará en futuras secciones

Conocer el sitio de unión antes de realizar un estudio de reconocimiento molecular, mejora la eficiencia del proceso de modo significativo. Si no se conoce se pueden utilizar servicios en línea como GRID, POCKET, SurfNet, entre otros. Un docking ciego es aquél donde no se conoce el sitio de unión, lo que requiere de un tiempo computacional mayor, sin embargo puede dar indicios de sitios alostéricos para cierto tipo de ligando y es un auxiliar al proponer el modo de unión de un inhibidor o potenciador no competitivo. Debido a las limitaciones computacionales, usualmente el reconocimiento molecular se lleva a cabo considerando al ligando como una entidad flexible y al receptor como un cuerpo rígido, con el fin de disminuir los tiempos de cálculo. Basado en la teoría de llave-cerradura de Fischer, Koshland desarrolló el ajuste inducido, donde postula que el sitio de unión está en constante cambio debido a las interacciones que tiene el ligando con la proteína, y sugiere por tanto que el receptor debe ser tratado como flexible y consecuentemente describa mejor los eventos relacionados a la interacción.

Algoritmos de muestreo Debido a los grados de libertad traslacional, rotacional y conformacional tanto del ligando como del receptor, existe un gran número de posibles modos de unión entre dos moléculas. Realizar una evaluación de todos los modos resulta ineficiente y demandante para las capacidades computacionales. Por tanto se han desarrollado algoritmos de muestreo para tomar el número mínimo representativo de todas las posibilidades para agilizar el cálculo.

- **Algoritmos de correspondencia (Matching algorithms)** se basan en la forma de un ligando y hacen un mapeo en el sitio activo de una proteína en términos de la información química. El complejo ligando-receptor se representa como farmacóforo, las propiedades químicas como los donadores y aceptores de hidrógeno pueden tomarse en cuenta durante el proceso de ajuste, éstos algoritmos son relativamente rápidos.
- **Construcción progresiva (Incremental construction)** colocan al ligando en el sitio activo de manera fragmentada y progresiva. El ligando se divide en fragmentos realizando roturas en los enlaces rotatorios. El primero de estos fragmentos, usualmente aquél que podría tener la mayor interacción con la proteína, se posiciona en el sitio activo y los restantes se añaden de manera consecutiva hasta completar la forma total del ligando.
- **Métodos estocásticos.** Buscan en el espacio conformacional modificando un ligando o un grupo de estos, uno de los más conocidos es el método de *Monte Carlo*. El otro se auxilia de algoritmos genéticos, basados en la teoría de la evolución de Darwin.
- **Dinámica molecular.** Es el método que trata al movimiento del ligando y al receptor de mejor modo que el resto de los algoritmos. Sin embargo, su desventaja principal es que progresa en intervalos cortos y por tanto tiene dificultades cuando se encuentra con una barrera energética relativa, que es necesario superar para encontrar un mínimo global en la superficie de energía del sistema, con lo que se limita a optimizaciones locales. [Meng et al., 2011]

Funciones de registro (Scoring functions). Su propósito es identificar los arreglos correctos de los que no lo son, involucran una estimación más que el cálculo de la afinidad de acoplamiento entre la proteína y el ligando adoptando ciertas simplificaciones. Dichas funciones pueden estar basadas campo de

fuerza, ser empíricas, o estar basadas en el conocimiento.

Las funciones de tipo campo de fuerza calculan la suma de las interacciones no enlazantes. Los términos electrostáticos se calculan de acuerdo a la ley de Coulomb. Las interacciones de largo alcance se definen como potenciales de Lennard-Jones. Sus desventajas radican en la incapacidad que tienen de considerar el entorno real de la proteína con las interacciones de carga, y que tienen una velocidad de cálculo lenta. Para ello se utilizan las distancias de corte o *cut-off*, lo que reduce su exactitud.

En las funciones empíricas, la energía de enlace se conforma de componentes energéticos, como los puentes de hidrógeno, interacciones iónicas, efecto hidrofóbico y entropía de enlace. Cada componente se multiplica por un coeficiente y entonces es sumado para obtener una energía final. Los coeficientes se obtienen de un análisis de regresión con un conjunto de complejos proteína-ligando con afinidades de enlace conocidas. Poseen términos simples de energía, pero se limitan a que no pueden ser extrapoladas más allá del subconjunto de entrenamiento con el cual se llevó a cabo la regresión.

Las funciones basadas en el conocimiento utilizan el análisis estadístico de estructuras cristalinas de complejos proteína-ligando para obtener las frecuencias o distancias interatómicas de contacto. Se basan en la presunción de que entre más favorable sea una interacción, la frecuencia también será mayor, la cual es convertida en potenciales de pares. Así se favorecen los contactos y se penalizan las interacciones repulsivas en cada átomo del ligando con la proteína con un corte o truncamiento de cálculo (*cutoff*) determinado. Su ventaja es la simplicidad computacional que puede ser aprovechada en la exploración de grandes bases de datos. [Hospital et al., 2015]

3.2.3. Minería de datos

El reto principal en la era tecnológica es el análisis y la búsqueda en grandes bases de datos de la información requerida para cualquier fin. Buscar entre una colección de miles de compuestos, proteínas y genes es desafiante si no se cuenta con las herramientas que puedan cubrir la necesidad de una búsqueda óptima y bien orientada a obtener los datos que más se ajusten a el alcance y objetivos de un proyecto.

La minería de datos es un proceso de extracción significativa de grandes bases de datos, que revela patrones, tendencias y correlaciones. Que serían imposibles de visualizar sin la ayuda de sus técnicas. Usa tecnologías de reconocimiento de patrones, redes neuronales, lógica difusa, algoritmos genéticos y otras técnicas avanzadas [Cruz-Martínez, 2007].

3.2.4. Pesca de objetivos (Target fishing)

Los métodos aplicados a la búsqueda de dianas biológicas para su estudio se conocen como métodos de *target fishing*, y están diseñados con el fin de identificar el objetivo más probable para una molécula de entrada.

En años recientes, se han desarrollado herramientas para el llamado *in silico target fishing*. La mayoría de estos métodos están basados en el *principio de la similitud*, que establece que "*Las moléculas estructuralmente similares tienden a unirse a los mismos objetivos y poseen actividad biológica similar*". [Cereto-Massagué et al., 2015].

Considerando como fuentes a distintas bases de datos como **ChEMBL**, la cual contiene millones de moléculas y la información acerca de su actividad, ahora es posible combinar el *espacio químico* con el *espacio biológico* en modelos que permiten generar «espectros» biológicos que en principio son capaces de predecir la actividad fenotípica de nuevas moléculas basadas en su estructura y la información de otras similares.

Los métodos de búsqueda pueden ser clasificados en cuatro grupos: basados en la similitud, en minería de datos combinada con procesos de aprendizaje de máquina (*machine learning*) en ordenadores, reconocimiento molecular y el análisis de espectros de bioactividad. También están disponibles aquéllos

basados en la información de las interacciones proteína ligando (o huellas digitales de interacción PLIF).

Métodos de similitud molecular Son de los métodos más simples, se basan en encontrar moléculas similares con dianas moleculares reportadas, tienen la ventaja de que sólo requieren el cómputo de la similitud entre compuestos. Su entrada suele ser una representación molecular simple como SMILES, que son sencillos de calcular. Una de las limitaciones es que algunas veces, algunos compuestos inactivos pueden exhibir una buena similitud con moléculas activas, lo que da lugar a falsos positivos.

Usualmente se utilizan coeficientes como el de Tanimoto para establecer la similitud, con un valor mayor de 0.85 se considera aceptable, pero éste valor no siempre es el único determinante, pueden utilizarse otros valores como el valor E, que es un valor que ayuda a estimar la posibilidad de alineamiento en un comportamiento determinado, entre otros.

Uso de programación y aprendizaje de máquina El reto para las herramientas *in silico* es extender su capacidad predictiva a las consecuencias biológicas de un ligando unido a su receptor. Por tanto los métodos combinados de minería de datos y machine learning, han sido implementados para generar un enfoque llamado quimiogenómico, que busca la sistematización del genoma al identificar, analizar, expandir y predecir las interacciones de los ligandos con las proteínas por medio de pruebas *in vitro* e *in silico* [Arroyo-Rodríguez et al., 2020]. De tal modo que usando algún método previamente descrito, (clasificadores Bayesianos, redes neuronales, mapas auto-organizados) se desarrollan modelos predictivos, que deben estar basados en datos confiables y bien validados en el conjunto de entrenamiento y construcción de los modelos. Con este enfoque el problema de búsqueda de dianas se convierte en una clasificación a gran escala de compuestos y objetivos basados en los modelos.

Métodos basados en la estructura En éste tipo de métodos pueden utilizarse el llamado reconocimiento molecular inverso [Negi et al., 2018], la búsqueda por modelado de farmacóforos [Meshram et al., 2017] o por huellas digitales de interacción. Aunque están limitados a la disponibilidad de estructuras bien resueltas y modelos moleculares validados.

3.2.5. Servidores para target fishing por similitud.

Drug Target Explorer Es una herramienta que permite a los usuarios comparar moléculas con una base de datos curada basada en resultados experimentales, la cual hace una comparación con base en el índice de Tanimoto para encontrar moléculas similares a la estructura de entrada, las cuales ya poseen registros de tener interacciones con diversos blancos biológicos en bases de datos curadas y validadas, como ChEMBL, DrugBank, DGidb y Chemical Probes. El registro de salida muestra una lista de compuestos, una lista de objetivos, y además, la parte más amigable con el usuario es la elaboración de redes de similitud [Allaway et al., 2018]. El recurso está disponible en www.synapse.org/dtexplorer

Swiss Target Prediction Hace uso de las huellas digitales en 2D y 3D de las moléculas de entrada, y hace predicciones con base en el conocimiento de ligandos conocidos con su respectivo objetivo que tienen mucha similitud usando el argumento de que la biofísica del reconocimiento molecular sugiere que la similitud en la forma de distribución del potencial electrostático conllevará a efectos similares. El conjunto de proteínas es tomado de ChEMBL versión 16 y utiliza un criterio específico para remover casos ambiguos, luego evalúa interacciones específicas como acoplamiento directo y toma valores K_i , K_d , IC_{50} menor a $10\mu\text{M}$ en los ensayos con cinco organismos (humano, ratón, rata, res y equinos) como recursos de la información.

Este servidor posee un editor 2D basado en JAVA llamado MarvinSketch, que es de uso libre y que facilita la introducción de las moléculas, aunque también posee la facilidad de introducir la representación SMILES para iniciar la búsqueda [Gfeller et al., 2014]. El recurso está disponible en <http://www.swisstargetprediction.ch>

Target Hunter Hace uso de PubChem como base de datos para minar aquellos ensayos biológicos o bioquímicos que involucren la palabra clave "HTS", con actividad biológica y potencia en dosis menores a $1\mu\text{M}$. Hace uso de la Química Combinatoria para tener registrados más de 25.3 millones de compuestos que cumplen con la regla de los cinco de Lipinsky [Lipinski et al., 1997]. De los cuales 1.85 millones han sido probados en al menos un bioensayo, y 0.8 millones han sido reportados como activos. TargetHunter utiliza el algoritmo TAMOSIC (Targets Associated with its MOst Similar COunterparts) para la minería de datos de ChEMBL. El algoritmo tiene la ventaja de remover falsos positivos y calcular tasas de confianza. Además integra una herramienta llamada *BioassayGeoMap*, cuyo fin es el de encontrar colaboradores potenciales que podrían validar experimentalmente los objetivos biológicos predichos[Liu et al., 2013]. El recurso está disponible en <http://www.cbligand.org/TargetHunter>, para éste es necesario crear un usuario y contraseña para acceder al servicio de predicción.

SEA (Similarity Ensemble Approach) Agrupa de modo cuantitativo y relaciona distintas proteínas basados en la similitud de sus respectivos ligandos. E incluso en el artículo original presentan casos en donde proteínas sin similitud secuencial ni geométrica participan en el modo de acción de algunos fármacos. Y no considera del todo cierto al principio de similitud. Los autores utilizaron MDL Drug Data Report para obtener sus receptores y crear su subconjunto de entrenamiento. Sus resultados requieren de un poco más de tiempo para ser interpretados de modo manual o visual pero tienen la ventaja de ayudar a crear redes de comparación entre ligandos y receptores y ayudar a encontrar efectos secundarios o incluso asociaciones fuera de un receptor específico, ya que es posible obtener archivos con extensión .csv que pueden ser después analizados con programación en *Python* [Keiser et al., 2007]. El recurso está disponible en sea.bkslab.org.

3.2.6. La serie LQM-700.

En el *Laboratorio de Química Medicinal* de la FES Cuautitlán, UNAM, se han desarrollado en los últimos años una serie de compuestos, análogos al ester fenetílico de ácido caféico (CAPE) y se han realizado estudios *in vitro* sobre su actividad y porcentaje de inhibición celular en las líneas celulares HeLa, MDA-MB-231 a $100\mu\text{M}$ durante 24h [Martínez Soriano, 2017] [Ruiz Sánchez, 2012] mostrando resultados prometedores para continuar con su investigación.

La estructura general de las amidas derivadas de CAPE se muestra en la Figura 4.

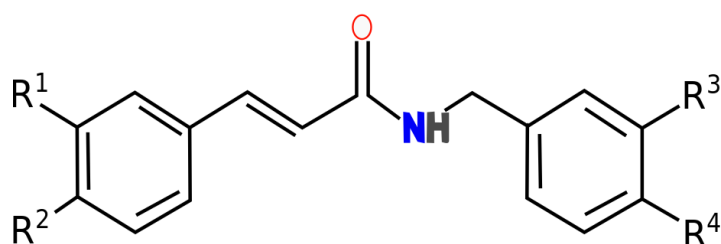


Figura 4: Para las amidas derivadas de CAPE, se realizaron ajustes en los sustituyentes de uno o ambos anillos aromáticos En alguna o más de una de las posiciones R1, R2, R3, y R4. Ver el trabajo de [Martínez Soriano, 2017]

La lista de los compuestos de la serie LQM-700 utilizados para obtener los blancos biológicos en el presente trabajo se muestra en la Tabla 1. Se observan los diversos tipos de sustituciones realizadas.

Tabla 1: Lista de los compuestos de la serie LQM-700 con los mayores porcentajes de inhibición de acuerdo a los reportes de Martínez-Soriano, (2017), utilizados como moléculas de entrada en el presente proyecto.

Compuesto	Estructura	Compuesto	Estructura
755		732	
CAPE		725	
752		721	
731		716	
745		709	
739		706	
738		753	
749			

4. Metodología.

4.1. Búsqueda de dianas biológicas

El primer paso es el de realizar búsquedas por índices de similitud. Considerando como entradas a las estructuras de los 10 compuestos más activos para cada línea celular, se encontraron un total de 15 compuestos más activos, por conveniencia en el tratamiento inicial de datos se omitieron temporalmente el resto de moléculas, concentrándose solamente sobre los que presentaron un porcentaje de viabilidad celular menor del 40 % [Martínez Soriano, 2017]. Y se crearon archivos de extensión *.mol para calcular sus descriptores químicos. El más importante para su uso posterior es el descriptor SMILES.

SMILES [Weininger, 1988] es una notación en una línea que describe la conectividad de las moléculas en un espacio 2D, lo que ayuda a la síntesis de información y ayuda a gestionar con mayor efectividad grandes bases de datos. Con base en este descriptor robusto es posible construir una estructura en 2D, e incluso con las mejoras recientes es posible calcular hasta algunos centros quirales y posiciones en 3D de los átomos. La ventaja de utilizar este tipo de descriptores en una línea en la era actual (que considera cada vez más a los recursos electrónicos) es que una entrada (query) puede ser comparada con bases de datos como ChEMBL que almacena 1.9 millones de compuestos reportados hasta la fecha, con constante actualización y que es una de las bases de datos más comúnmente utilizadas por los protocolos de cribado virtual.

El procedimiento consiste en utilizar cuatro servidores que ayudan a encontrar tanto moléculas similares como sus posibles blancos terapéuticos. En la figura 5 se muestra el diagrama de flujo utilizado para la búsqueda de blancos potenciales, utilizando una base de datos interna que liste los descriptores SMILES de los LQM-700 es posible entonces introducir la información necesaria en cada servidor para poder hallar blancos potenciales, a mayor probabilidad calculada se puede dar un mayor interés. Sin embargo, por delimitaciones en el caso de estudio es necesario hacer un tamizado de la respuesta biológica que tienen los blancos obtenidos, con el fin de enfocar la atención en inductores de apoptosis en células cancerosas o en casos particulares, hacia la respuesta real que se esté investigando. El proceso llevará a blancos potenciales que posteriormente serán estudiados mediante reconocimiento molecular.

4.2. Estudios de reconocimiento molecular

Una vez hallados los blancos potenciales se procede a realizar estudios de reconocimiento molecular, los cuales utilizan diversos parámetros para calcular las poses de menor energía y que maximicen la interacción entre ligando y receptor.

Antes de realizar los estudios, es necesario explorar la base de datos de proteínas del *Protein Data Bank (PDB)* [Burley et al., 2018] y hallar las estructuras derivadas de rayos X con la mejor resolución. Preferentemente éstas deben de tener un ligando natural para poder validar el protocolo de reconocimiento molecular [Ramírez and Caballero, 2018].

En la Figura 6 se presenta el diagrama a seguir después de encontrar un blanco potencial, se consideran las entradas con una resolución buena, de preferencia menor a 2Å para después realizar el tratamiento de la estructura y prepararla para el estudio de reconocimiento molecular, este enfoque ha sido una herramienta de uso muy frecuente durante el proceso de descubrimiento y desarrollo de fármacos, cuya metodología puede ser revisada en [Meng et al., 2011].

En general el proceso de reconocimiento molecular involucra los siguientes pasos:

- Eliminación de ligandos naturales co-cristalizados y moléculas de disolvente.
- Corrección de segmentos/residuos no resueltos por las técnicas cristalográficas. En caso de ser necesario se hará un modelado por homología para completar las secuencias incompletas y otorgar estabilidad y continuidad al sistema.

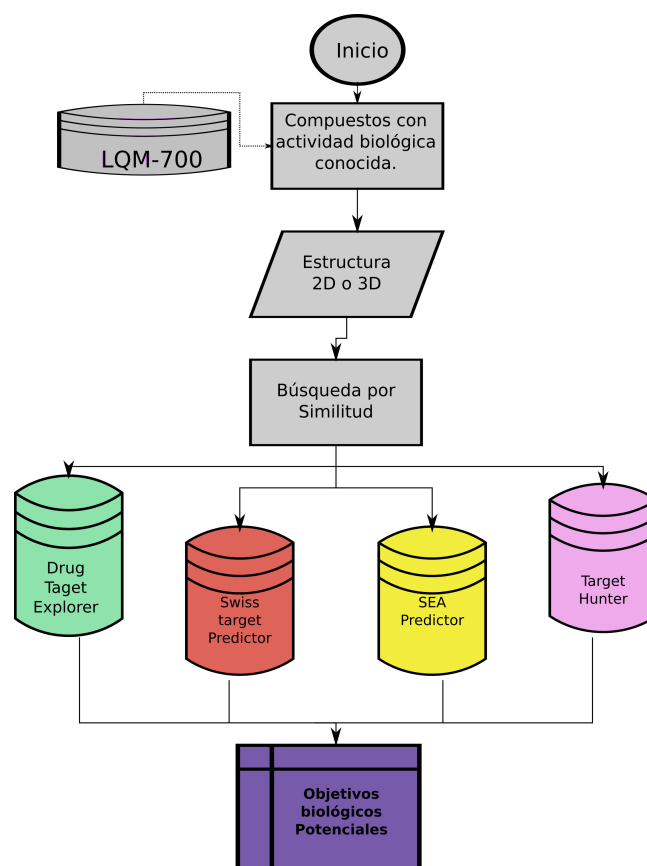


Figura 5: Procedimiento general para la búsqueda de los blancos biológicos con mayor probabilidad de ser aquéllos en los cuales los compuestos de la serie LQM-700 ejerzan su acción.

- Exploración para hallar el sitio de unión y generación de la cavidad a donde realizar el posicionamiento y análisis (SiteFinder).
- Minimización de la energía total del sistema, utilizando un campo de fuerza apropiado.
- Cálculo de poses mediante triangulación y posterior refinamiento con el uso de un "scoring function" para el cálculo de la energía de asociación proteína-ligando.
- Evaluación de la energía de asociación y datos de posición de cada resultado obtenido.

El programa para realizar el reconocimiento molecular fue Molecular Operating Environment (MOE), de Chemical Computing Group, el cual tiene una interfaz gráfica que facilita la introducción y manipulación de datos y procesos. El campo de fuerza fue MMFF94x [Halgren, 1996], el cual es aplicable tanto a macromoléculas y proteínas como a moléculas orgánicas pequeñas y ha demostrado tener un buen nivel de precisión. SiteFinder es una rutina de MOE que fue utilizado para crear un conjunto de átomos "dummy" dentro del sitio de unión más cercano al sitio del ligando natural, todo este procedimiento es similar al utilizado por [Negi et al., 2018].

En cuanto al Scoring Function para el refinamiento de los cálculos, de acuerdo al estudio realizado por Kalinowsky et al., (2018) se utilizó el posicionamiento por triangulación con LondondG, seguido de refinamiento con Affinity dG, que son dos de las funciones cuyas poses calculadas se asemejan mucho a las experimentales.

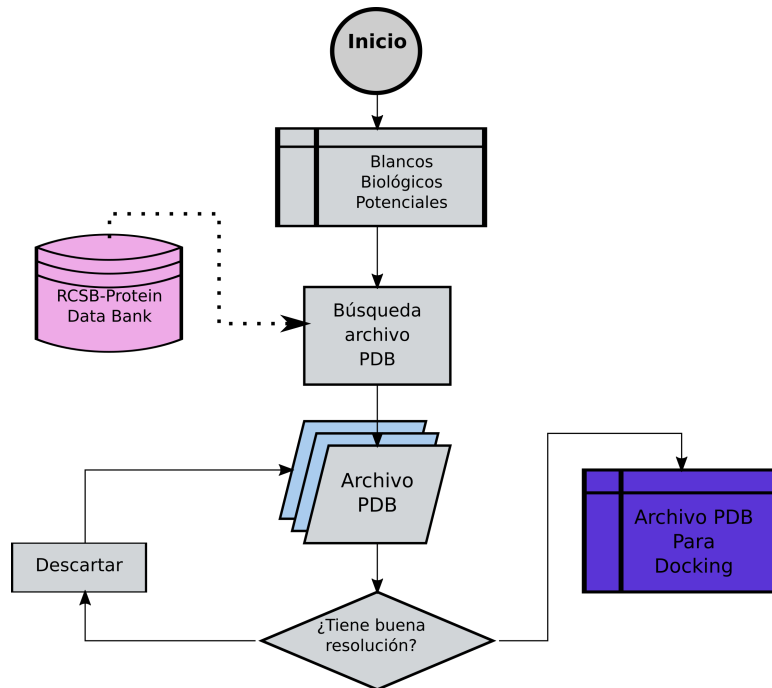


Figura 6: Una vez hallado un blanco, se busca una entrada PDB con buena resolución y validación para proseguir con el estudio de reconocimiento molecular.

4.3. Dinámica molecular.

La dinámica molecular se realizará una vez hallada la mejor pose de acoplamiento ligando-receptor. Usando motores de simulación como NAMD [Phillips et al., 2005] con el mismo campo de fuerza que para el estudio de reconocimiento molecular.

Como el protocolo de dinámica dicta, el primer paso para una simulación es contar con una entrada inicial, en este caso la mejor conformación hallada por el reconocimiento molecular. Después se seguirán los siguientes pasos.

- Solvatar en una caja determinada con un potencial, el disolvente será agua según el modelo TIP3P, con propiedades periódicas, es decir, que en cada eje tendrá una imagen igual para evitar la pérdida de materia, las dimensiones de la caja serán de 50 Å en cada eje (X,Y,Z).
- Para los cálculos de dinámica molecular fueron utilizados un barostáto y un termostáto de tipo langevin, usando los parámetros por defecto de NAMD.
- Es necesario realizar una minimización de energía durante algunos picosegundos, el objetivo final no es hallar un mínimo de energía global, si no más bien ajustar el sistema al campo de fuerza, a la distribución particular de las moléculas de disolvente y para relajar posibles efectos estéricos derivados de la predicción de las coordenadas de los átomos durante la generación del archivo de entrada.
- Después es necesario realizar un calentamiento lineal incrementando de 0 a 300 K durante 300 ps. A cada paso de integración las velocidades son reasignadas de una nueva distribución de Maxwell y la temperatura incrementará 0.001K.
- Después de esto se realizará un proceso de pre-equilibrio a la temperatura asignada (300 K) para equilibrar las energías cinética y potencial, de este modo se logra distribuir la energía proporcionada al sistema durante el calentamiento (En las simulaciones de dinámica molecular, la temperatura se

refleja como la cantidad de energía cinética que tiene cada partícula, átomo o molécula) en todos los grados de libertad involucrados.

- Finalmente se realizará el estado de producción de una trayectoria de dinámica molecular en un ensamble NPT, que nos dará la idea de las fluctuaciones de energía presentes en el sistema.

Una vez realizada la dinámica se evalúa su comportamiento con base en los resultados de RMSD, RMSF, y radio de giro.

5. Resultados y discusión

5.1. Blancos biológicos

Derivado de la búsqueda por similitud, se encontraron seis blancos biológicos que resultaron ser potenciales candidatos para su evaluación posterior. Se resumen en la tabla 2.

Tabla 2: Blancos objetivos mejor catalogados derivados de la búsqueda por similitud usando los servidores descritos con anterioridad.

Blanco	Expresión y actividad	Entrada PDB utilizada	Resolución (Å)	Referencia
MMP9	Es una metaloproteasa dependiente de calcio y zinc sobreexpresada en distintos tipos de cáncer. Regula morfogénesis y metástasis	6ESM	1.1	[Kotipatruni et al., 2012] [Kalva et al., 2016] [Nutti et al., 2018]
AIF	Factor de inducción de apoptosis. Induce la muerte celular programada independiente de caspasas.	1M6I	1.8	[Srivastava, 2007] [Shen et al., 2015] [Ye et al., 2002]
CB1	Receptor de canabinoides, inhibe la proliferación y que causa apoptosis en ciertos tipos de melanoma.	5XR8	2.98	[Brand et al., 2018] [Lisovyy et al., 2009] [Hua et al., 2017]
AKR1C3	Aldo ceto reductasa 1C3. Se encuentra sobre expresada en cáncer prostático resistente a la castración.	5HNT	2.0	[Tian et al., 2014] [Li et al.,] [Verma et al., 2018]
AKR1B10	Aldo ceto reductasa 1B10. Está sobre expresada en cáncer de mama, promueve el crecimiento tumoral y metástasis	4GQ0	2.1	[Ma et al., 2012] [Liping et al.,]
MAP14K	Proteín quinasa 14 activada por mitógeno. Se relaciona con distintos procesos celulares incluidos la proliferación, diferenciación, transcripción, regulación y desarrollo.	5LAR	1.5	[Cheng et al., 2017] [Petersen et al., 2016]

Cada uno de los blancos biológicos reportados en la Tabla 2 se eligió con base en el criterio de su relación con la actividad antineoplásica reportada por [Martínez Soriano, 2017] y [Ruiz Sánchez, 2012], descartándose otros resultados que si bien son interesantes desde el punto de vista de la polifarmacología, no son de interés ni del alcance de las simulaciones a realizar, además de que no se tienen datos internos que avalen dicha actividad, como es el caso de genes relacionados con asma, e isquemia como ALOX5 [Lisovyy et al., 2009] y otros más relacionados a afecciones cardiacas y neurológicas, sin embargo no fue necesario realizar más estudios con ellos *in silico*, ya que sin la experimentación previa necesaria, no es posible extrapolar información de los resultados computacionales.

5.2. Reconocimiento molecular

Una vez hallados los blancos biológicos más importantes se continúa con el estudio realizando el modelado computacional de reconocimiento molecular, calculando así las energías de asociación relacionadas con cada uno de los receptores. La Tabla 3 muestra las energías de asociación calculadas en el paso de refinamiento usando Affinity dG como función de registro.

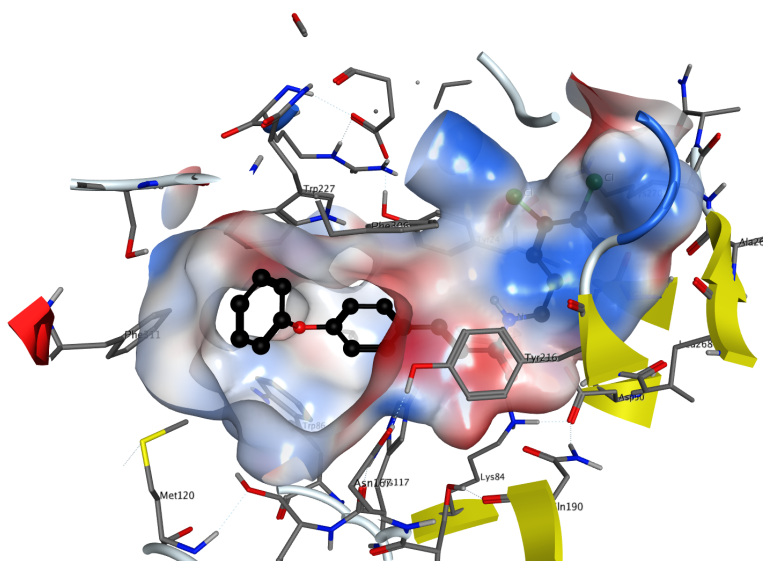


Figura 7: Representación de la superficie molecular de la cavidad del receptor AKR1C3 con LQM-755 como ligando. La superficie está coloreada con base en el potencial electrostático calculado por MOE, donde el azul son zonas positivas, el blanco son zonas neutras y el rojo son zonas negativas. Se observa que los átomos de cloro se acoplan en la region positiva.

Para el compuesto LQM-755, que fue el más activo en las tres líneas celulares con las que se evaluó su inhibición de crecimiento, se realizó un análisis más exhaustivo, por lo que se continuó con el proceso de dinámica molecular. En la Tabla 3 se puede observar que el complejo receptor-ligando con menor energía corresponde al receptor AKR1C3, probablemente a este se le deba prestar más atención en el proceso de análisis de las trayectorias de dinámica molecular, así como a los ligandos líderes en energía contra cada blanco biológico.

Tabla 3: Resultados de energía de acoplamiento de cada uno de los ligandos con el respectivo blanco molecular encontrado

Objetivo	755	CAPE	752	731	745	739	738	732	725	721	716	709	706	753	749
MMP9	-6.4005	-5.1278	-5.9547	-4.7873	-4.9970	-4.4192	-4.733	-4.5218	-4.4965	-5.9587	-4.8355	-4.778	-4.0117	-5.0972	-4.2721
AIF	-6.4370	-5.8135	-6.5905	-5.3781	-5.1146	-6.1393	-5.8824	-5.3661	-6.2991	-6.5248	-5.4493	-5.1299	-5.7576	-6.6758	-6.4049
AKR1B10	-9.2966	-7.8296	-8.8965	-9.0578	-7.2273	-8.0844	-7.5031	-6.9811	-8.7773	-9.2795	-8.7967	-7.8560	-7.6807	-8.6822	-8.7323
AKR1C3	-9.6287	-7.6384	-8.4701	-6.6977	-6.4741	-7.6377	-7.2001	-6.7814	-7.8185	-8.8355	-6.4989	-6.3179	-6.4432	-9.1831	-7.7893
CB1	-4.8294	-4.3026	-4.7507	-3.3927	-3.5412	-3.5267	-4.1338	-3.7676	-4.2714	-4.8446	-4.0772	-3.8696	-3.6025	-4.0476	-3.4279
MAP14K	-7.5805	-6.5046	-6.6809	-4.1342	-6.7999	-7.5278	-7.0830	-6.8134	-7.5612	-8.2958	-6.4949	-6.7662	-5.9301	-7.3787	-7.6377

Se puede representar tal conformación en dos dimensiones (Figura 8), aunque se pierde la representación espacial de los ángulos diedros en el compuesto y la proteína, es más sencillo visualizar que aparentemente no existen interacciones inter moleculares entre el ligando y el receptor de tipo puentes de hidrógeno o apilamiento π . Sin embargo, es posible notar que el contorno de interacción rodea a casi toda la molécula, y que además la exposición del ligando es mínima. Si analizamos el acomodamiento de los anillos aromáticos como se muestra en la Figura 8, con respecto a los residuos aromáticos más cercanos en la zona del cafeato, hay cuatro aminoácidos aromáticos

En adición, creando una superficie molecular del receptor como potencial electrostático mostrado en la Figura 7 se observa que la conformación del ligando dentro de la cavidad de la proteína se orienta de tal modo que los átomos de cloro tienen una probabilidad de estabilización electrostática al estar en una

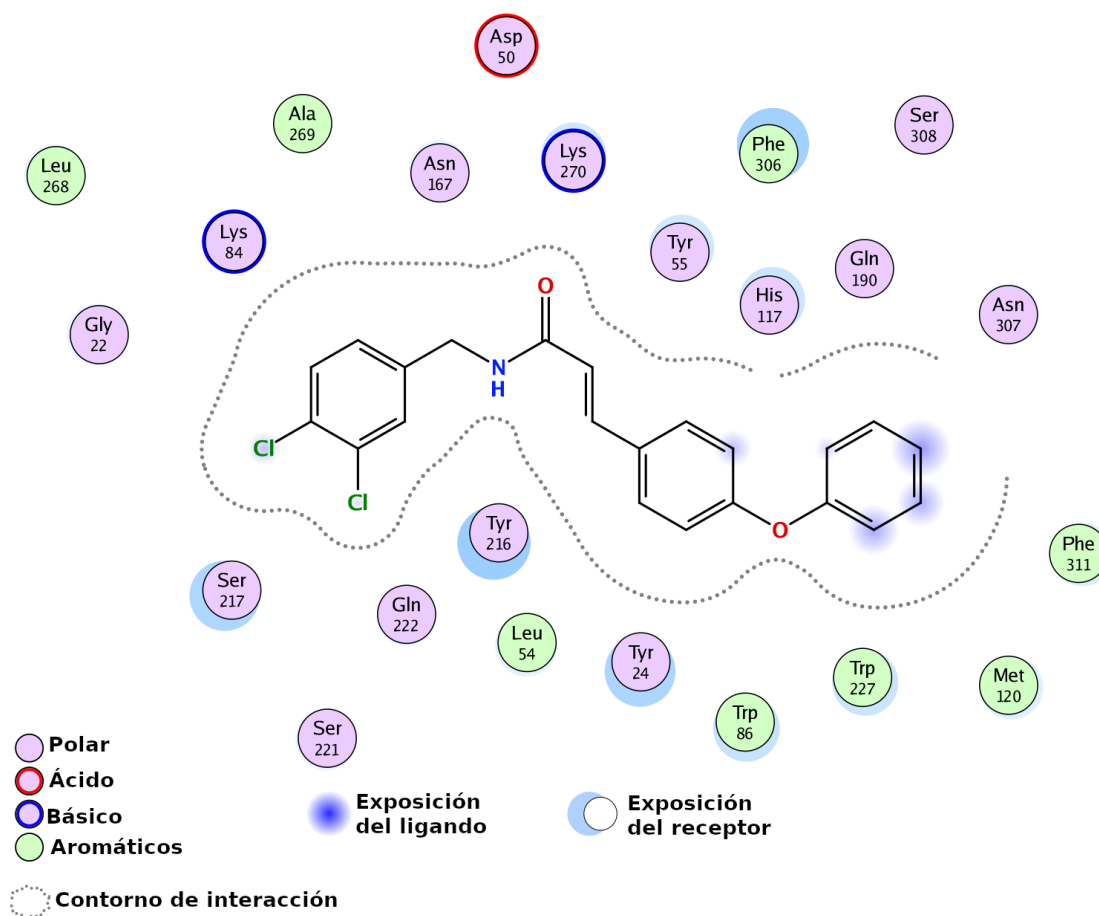


Figura 8: Se observan a los residuos que están más cerca del ligando y su naturaleza ácida, básica o aromática, así como el contorno de interacción y la exposición del ligando dentro de la cavidad.

región positiva (en azul). Esto podría explicar la diferencia entre los compuestos LQM-755 y LQM-752, cuya disimilitud radica en los átomos de cloro, ya que la actividad de inhibición del primero es prácticamente el doble (Porcentaje de viabilidad de crecimiento celular: LQM-755 18%; LQM-752 40%). Conclusion similar se podría plantear para el compuesto LQM-753, que posee un carbono extra entre el grupo amida y el anillo aromático aumentando sus enlaces rotatorios, disminuyendo así la probabilidad de inclusión del ligando dentro de la cavidad del receptor, explicando así porqué la actividad antineoplásica es tres veces menor que la de LQM-755 (60% de viabilidad de crecimiento celular).

En la Tabla 3 se muestran los resultados promedio de la energía de acoplamiento derivado del reconocimiento molecular. Se nota que el complejo con una energía de unión más favorable es el formado por LQM-755 y AKR1C3, proteína que se sobre expresa en cáncer de próstata resistente a la castración, [Tian et al., 2014]. Asimismo resulta interesante observar el comportamiento tendiente del complejo AKR1B10-LQM 755 a ser más estable, probablemente por la naturaleza desigual de sus cavidades. Este resultado podría corroborarse con los resultados de dinámica molecular.

5.3. Dinámica Molecular

Para los estudios de dinámica, se tomaron las mejores (top scoring) poses derivadas del reconocimiento molecular del complejo Receptor-LQM755, se prepararon y solvataron en cajas de 50 Å. La primera

simulación se llevó a cabo durante un tiempo corto de simulación para analizar la posible estabilización de los complejos durante 3 nano segundos. Se descartó la entrada de CB1, debido a su baja afinidad de interacción receptor-ligando. Los resultados de RMSD y Radio de giro (RoG) se presentan en la figura 9.

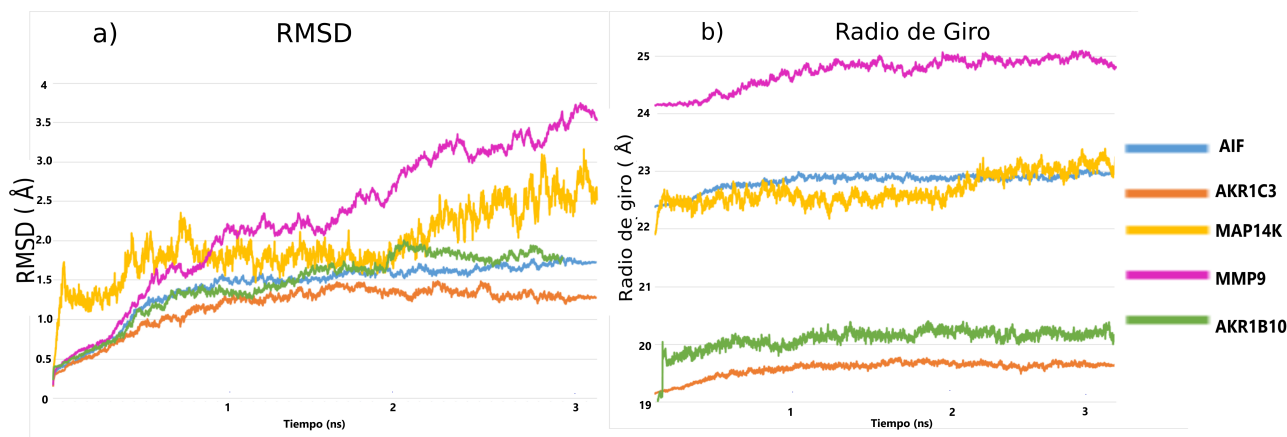


Figura 9: RMSD (a) y Radio de giro (b) para los complejos de cada blanco con LQM-755 durante una simulación corta para evaluar el comportamiento inicial de los complejos y así usar menos tiempo de cómputo para los complejos menos estables.

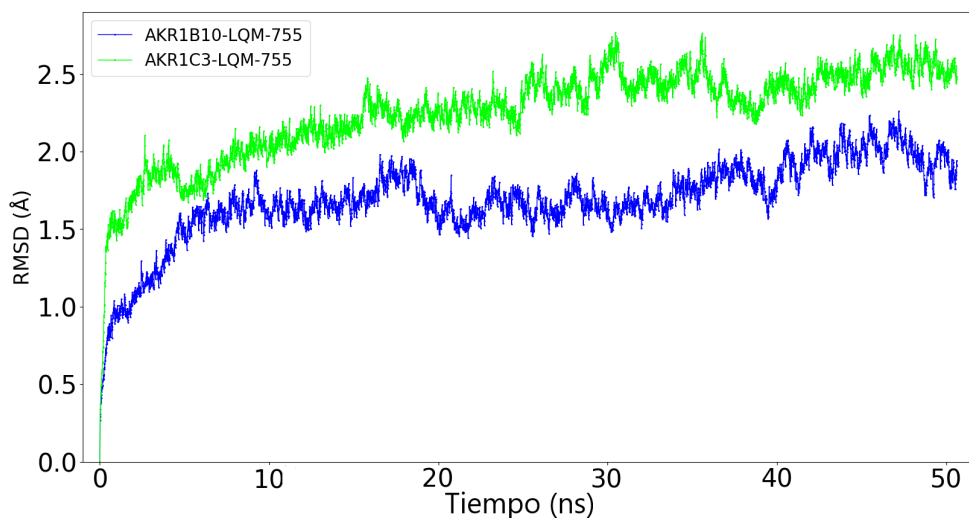


Figura 10: RMSD de los sistemas AKR1C3 (Verde) y AKR1B10 (azul) en complejo con el ligando LQM-755 se muestran, puede notarse que para el complejo de LQM-755 con AKR1CB10, el valor es menor y oscila alrededor de 1.5 Å, .

Se observa que los comportamientos de menor variación con respecto a sus estructuras originales son AKR1C3 y AKR1B10, seguidos de AIF y MAP14K, el que presenta mayor variación tanto en RMSD como con el radio de giro es MMP9, lo que puede significar que éste sea el complejo menos estable de los analizados. Lo anterior es de interés si se considera que al principio de la búsqueda, MMP9 era uno de los blancos con mayor probabilidad de acuerdo a los servidores, y aunque se han realizado estudios de dinámica molecular con probables agentes cancerígenos basados en la búsqueda por farmacóforo [Kalva et al., 2016], considerando a MMP9 como punto de partida, no podría considerarse como el blanco biológico con mayor

afinidad hacia los compuesto de la serie LQM-700, al menos desde el enfoque del presente trabajo.

La siguiente tarea fue extender el tiempo de cálculo para los sistemas con una menor variación de distancias con respecto a la original, y se continuó, después del calentamiento de 300 ps y pre equilibrio de 300 ps, de 50 000 ps de fase de producción con presión constante de 1 atm, con un tiempo de muestreo igual a 1 ps para los sistemas AKR1C3 y AKR1B10 con el compuesto líder (LQM-755). Las gráficas de RMSD y Radio de Giro, se muestran en las Figuras 10 y 11.

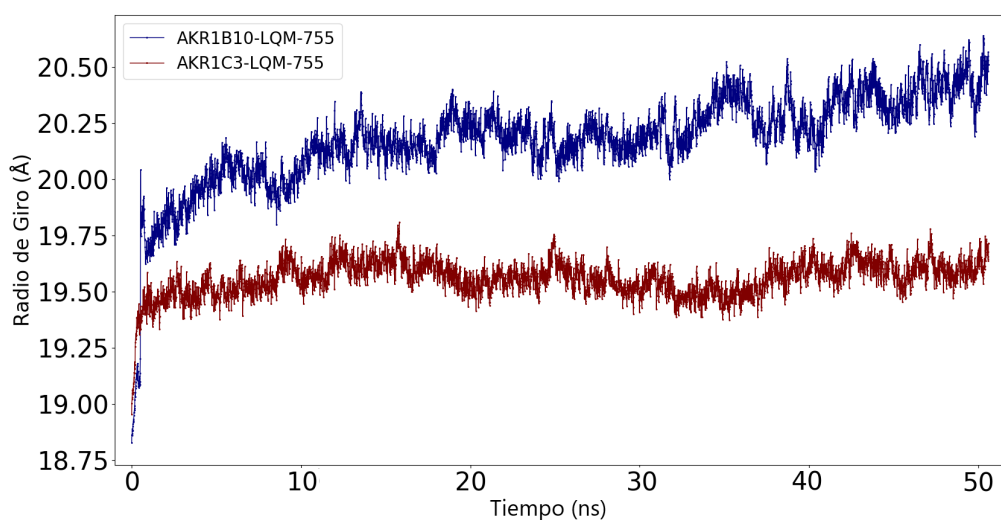


Figura 11: Radio de giro para los complejos receptor-ligando de AKR1C3 (Rojo) y AKR1B10 (Azul) con LQM-755. Se observa que el radio de giro varía menos en el caso de AKR1C3, lo que es un indicativo de mayor compacticidad y estabilidad.

Puede observarse que para este caso, la fluctuación de RMSD del complejo AKR1B10 con LQM-755 corrobora los datos obtenidos durante los estudios de reconocimiento molecular, y que la tendencia a obtener mejores energías de afinidad puede deberse a una mejor interacción del ligando con el sitio receptor de AKR1B10. Esto tiene validez si se toma en cuenta lo mencionado por [Ramírez and Caballero, 2018], donde no siempre la mejor energía de afinidad representa al complejo receptor-ligando más estable.

Por otra parte, también es visible que para el caso de AKR1C3, la proteína se comporta de modo más compacto, es decir, el radio de giro es menor, y esto bien puede deberse al tamaño y conformación de la proteína al inicio de la simulación, y que el valor numérico se debe a su naturaleza de extensión. Resulta interesante observar que en comparación con el complejo de AKR1B10, hay un menor ruido en los valores y que su tendencia se ve más centralizada al valor de 19.5 Å.

6. Conclusiones

Se ha logrado la recopilación y propuesta de objetivos biológicos para una serie de amidas derivadas del ester fenetílico de ácido caféico mediante el uso de servidores de búsqueda por similitud, logrando encontrar una lista de cientos de objetivos que se relacionan a distintos tipos de afecciones cardiacas, enfermedades vasculares, cáncer o bien que controlan procesos biológicos normales como la inflamación. Algunos de los blancos biológicos encontrados en los servidores coinciden con otros que han sido reportados en la literatura [Tolba et al., 2013] [Murtaza et al., 2015] como objetivos para el diseño racional de fármacos, lo que indica una potencial confiabilidad del uso de estos servidores.

Se analizó de modo manual comparando con los datos de Gene Cards (www.genecards.org) y Uniprot (www.uniprot.org) la información de los blancos biológicos obtenidos en los servidores de búsqueda por similitud, eliminando aquéllos blancos que tenían poca relación con el proceso de inducción de apoptosis o un posible efecto importante en el ciclo celular, prefiriéndose aquéllas proteínas sobre expresadas en los tipos de cáncer correspondientes a los estudiados por Martínez Soriano [Martínez Soriano, 2017], con lo cual la lista de objetivos potenciales aptos para continuar con la investigación se redujo a seis. Sin embargo es en este paso donde el procedimiento sufre un efecto de cuello de botella, ya que analizar cada gen y proteína de modo individual es la parte que más tiempo consume y dado que el método pretende minimizar los tiempos de búsqueda, es necesario buscar o desarrollar una herramienta más automatizada para depurar los datos con baja correlación a los fenómenos observables de inhibición de crecimiento celular.

Se obtuvieron las estructuras de las proteínas de la lista de objetivos candidatos: AKR1C3, AKR1B10, CB1, MMP9, AIF y MAP14K, cada proteína con una sobre expresión en tejido canceroso de distintos tipos, enfocándose en los tejidos de las líneas celulares con las que experimentó Martínez Soriano, de las cuales se seleccionaron las estructuras mejor resueltas por cristalografía de rayos X. Cada una fue preparada usando el campo de fuerza MMFF94x, el cual es usado para el modelado de proteínas y algunos ácidos nucleicos y que además ha permitido el modelado de moléculas pequeñas aplicadas a la Química Medicinal. Con lo anterior se logró hacer el primer acercamiento a los estudios de reconocimiento molecular, además de que también puede ser un punto de partida para analizar los conjuntos de proteínas en función de sus familias y llevar a cabo estudios conformacionales como Modelos de Redes Anisotrópicas (ANM por sus siglas en inglés), que son modelos elásticos que ayudarían a agrupar cada proteína en función de los ligandos asociados o su ausencia.

Los estudios de reconocimiento molecular usando el campo de fuerza MMFF94x aunado a las funciones de registro Affinity dG y London dG sugieren que las mejores energías de acoplamiento corresponden al complejo receptor-ligando más estable formado entre LQM-755 y AKR1C3, aunque por tendencia los valores relativos más bajos se observan para el complejo con AKR1B10. Tal diferencia podría explicarse por el análisis de superficies de las cavidades de la proteína y sus cargas electrostáticas, es probable que la cavidad en AKR1C3 difiera ligeramente de la de AKR1B10 y por tanto exista el comportamiento observado. Para ello resulta necesario replicar los resultados del reconocimiento molecular y analizar con mayor detenimiento cada uno de los complejos receptor-ligando formados y de sus cavidades e interacciones intermoleculares.

Finalmente, las estructuras de los complejos con mayor afinidad derivadas del reconocimiento molecular se llevaron a estudios de dinámica molecular para evaluar la estabilidad de los mismos, así como de posibles cambios conformacionales de la proteína correspondiente, la dinámica molecular indica que los complejos que se estabilizan más con un RMSD menor y una mayor compacidad, son los de AKR1B10 y AKR1C3 con el compuesto LQM-755. Esto puede ser un indicador para realizar estudios experimentales sobre el acoplamiento de las entidades mencionadas, para esto habría que plantear el protocolo experimental, no sin antes replicar los experimentos *in silico* y con ello tener una perspectiva más explorada para realizarlos. Aunque en el presente trabajo se muestran los gráficos de simulaciones individuales, cada proceso se repitió en al menos tres ocasiones y se obtuvieron resultados comparables.

7. Perspectivas

Validar la confiabilidad del método al confirmar las interacciones entre los blancos biológicos propuestos y las moléculas de la serie LQM-700. Para ello se requiere proponer un método analítico de determinación de constantes de afinidad, o alguna técnica que dé indicios sobre el comportamiento de las moléculas y proteínas estudiadas, como la resonancia magnética nuclear de las proteínas en solución.

Realizar una comparación entre la efectividad del método aplicado a otro tipo de compuestos con actividades biológicas diversas, con el fin de evaluar el comportamiento de las búsquedas, o bien repetir el proceso de búsqueda pero con moléculas cuyo objetivo biológico esté bien identificado, todo esto como parte de validación de la efectividad del proceso.

Los cambios en la conformación de las proteínas estudiadas podrían evaluarse con métodos no descritos en el presente trabajo, como lo son el análisis de componentes principales (PCA), el análisis de dinámica esencial (*Essential dynamics*) y modelos elástico que brindan un enfoque distinto al del presente trabajo.

En cuanto al tiempo de procesamiento de la información, dado que las listas son complicadas de interpretar y comprender las funciones de cada blanco derivado de la búsqueda por similitud, se propone escribir un programa en un lenguaje de alto nivel como Python, que nos permita aminorar los tiempos de análisis y encontrar de manera rápida a un mayor número de candidatos mediante filtros específicos. Por lo que se propone la continuación de la investigación en la rama de la Química Computacional dentro del grupo de investigación.

8. Bibliografía

Referencias

- [Allaway et al., 2018] Allaway, R. J., La Rosa, S., Guinney, J., and Gosline, S. J. (2018). Probing the chemical–biological relationship space with the Drug Target Explorer. *Journal of Cheminformatics*, 10(1):41.
- [Arroyo-Rodríguez et al., 2020] Arroyo-Rodríguez, J., González-Young, F., Hernández-Vega, R., Lizano-Quirós, C., Quesada-Murillo, J., and Mora-Roman, J. J. (2020). Quimiogenómica y Quimiogenética: Definiciones, técnicas y aplicaciones en el ámbito farmacéutico. *Revista Tecnología en Marcha*, 33(1):Pág. 3–16.
- [Brand et al., 2018] Brand, S., Roy, S., Schröder, P., Rathmer, B., Roos, J., Kapoor, S., Patil, S., Pommerenke, C., Maier, T., Janning, P., Eberth, S., Steinhilber, D., Schade, D., Schneider, G., Kumar, K., Ziegler, S., and Waldmann, H. (2018). Combined Proteomic and In Silico Target Identification Reveal a Role for 5-Lipoxygenase in Developmental Signaling Pathways. *Cell Chemical Biology*, 25(9):1095–1106.e23.
- [Burley et al., 2018] Burley, S. K., Berman, H. M., Christie, C., Duarte, J. M., Feng, Z., Westbrook, J., Young, J., and Zardecki, C. (2018). RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. *Protein Science*.
- [Cedillo-Cruz, 2020] Cedillo-Cruz, E. (2020). *Simulaciones computacionales para el estudio de la retención de plomo en agua usando moléculas tensoactiva*. PhD thesis, Universidad Nacional Autónoma de México.
- [Cereto-Massagué et al., 2015] Cereto-Massagué, A., Ojeda, M. J., Valls, C., Mulero, M., Pujadas, G., and Garcia-Vallve, S. (2015). Tools for in silico target fishing. *Methods*, 71(C):98–103.
- [Cheng et al., 2017] Cheng, F., Twardowski, L., Fehr, S., Aner, C., Schaeffeler, E., Joos, T., Knorpp, T., Dorweiler, B., Laufer, S., Schwab, M., and Torzewski, M. (2017). Selective p38 α MAP kinase/MAPK14 inhibition in enzymatically modified LDL-stimulated human monocytes: Implications for atherosclerosis. *FASEB Journal*, 31(2):674–686.
- [Cruz-Martínez, 2007] Cruz-Martínez, M. d. R. (2007). *Minería de Datos Multiperspectiva*. PhD thesis, Universidad Nacional Autónoma de México.
- [Gfeller et al., 2014] Gfeller, D., Grosdidier, A., Wirth, M., Daina, A., Michielin, O., and Zoete, V. (2014). SwissTargetPrediction: a web server for target prediction of bioactive small molecules. *Nucleic Acids Research*, 42(W1):W32–W38.
- [Gil-Bao et al., 2015] Gil-Bao, I., Baraibar, I., Castañón, E., Nistal, E., Bandrés, E., Agirre, X., and Prósper, F. (2015). Biología celular del cáncer. In *Biología celular biomédica*, chapter 16, pages 365–395.
- [Halgren, 1996] Halgren, T. A. (1996). Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry*, 17(5-6):490–519.
- [Hanahan and Weinberg, 2011] Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144:646–674.
- [Hospital et al., 2015] Hospital, A., Goñi, J. R., Orozco, M., and Gelpí, J. L. (2015). Molecular dynamics simulations: advances and applications. *Advances and applications in bioinformatics and chemistry : AABC*, 8:37–47.
- [Hua et al., 2017] Hua, T., Vemuri, K., Nikas, S., Laprairie, R., Wu, Y., Qu, L., Pu, M., Korde, A., Jiang, S., Ho, J., Han, G., Ding, K., Li, X., Liu, H., Hanson, M., Zhao, S., Bohn, L., Makriyannis, A., Stevens, R., and Liu, Z. (2017). Crystal structures of agonist-bound human cannabinoid receptor CB1. *Nature*, 547:468–471.
- [Kalva et al., 2016] Kalva, S., Agrawal, N., Skelton, A. A., and Saleena, L. M. (2016). Identification of novel selective MMP-9 inhibitors as potential anti-metastatic lead using structure-based hierarchical virtual screening and molecular dynamics simulation. *Molecular BioSystems*, 12(8):2519–2531.
- [Keiser et al., 2007] Keiser, M. J., Roth, B. L., Armbruster, B. N., Ernsberger, P., Irwin, J. J., and Shoichet, B. K. (2007). Relating protein pharmacology by ligand chemistry. *Nature Biotechnology*, 25(2):197–206.
- [Khan et al., 2016] Khan, F. I., Wei, D. Q., Gu, K. R., Hassan, M. I., and Tabrez, S. (2016). Current updates on computer aided protein modeling and designing.
- [Kotipatruni et al., 2012] Kotipatruni, R. R., Nalla, A. K., Asuthkar, S., Gondi, C. S., Dinh, D. H., and Rao, J. S. (2012). Apoptosis Induced by Knockdown of uPAR and MMP-9 is Mediated by Inactivation of EGFR/STAT3 Signaling in Medulloblastoma. *PLoS ONE*, 7(9).
- [Kumar, 2015] Kumar, S. (2015). Drug Targets for Cancer Treatment: An Overview. *Medicinal Chemistry*, 5(3):115–123.

- [Li et al.,] Li, C., Zhao, Y., Zhang, H., and Hu, X. Crystal Structure of AKR1C3 complexed with octyl gallate. *TO BE PUBLISHED*.
- [Liping et al.,] Liping, Z., Xuehua, Z., Shangke, C., and Jing, Z. Crystal structure of AKR1B10 complexed with NADP⁺ and Caffeic acid phenethyl ester. *TO BE PUBLISHED*.
- [Lipinski et al., 1997] Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 23(1-3):3–25.
- [Lisovyy et al., 2009] Lisovyy, O. O., Dosenko, V. E., Nagibin, V. S., Tumanovska, L. V., Korol, M. O., Surova, O. V., and Moibenko, O. O. (2009). Cardioprotective effect of 5-lipoxygenase gene (ALOX5) silencing in ischemia-reperfusion. *Acta Biochimica Polonica*, 56(4):687–694.
- [Liu et al., 2013] Liu, H., Su, W., Wipf, P., Ma, C., Wang, L., and Xie, X.-Q. (2013). TargetHunter: An In Silico Target Identification Tool for Predicting Therapeutic Potential of Small Organic Molecules Based on Chemogenomic Database. *The AAPS Journal*, 15(2):395–406.
- [Ma et al., 2012] Ma, J., Luo, D.-X., Huang, C., Shen, Y., Bu, Y., Markwell, S., Gao, J., Liu, J., Zu, X., Cao, Z., Gao, Z., Lu, F., Liao, D.-F., and Cao, D. (2012). AKR1B10 overexpression in breast cancer: Association with tumor size, lymph node metastasis and patient survival and its potential as a novel serum marker. *International Journal of Cancer*, 131(6):E862–E871.
- [Martínez, 2015] Martínez, L. (2015). Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. *PloS one*, 10(3):e0119264.
- [Martínez Soriano, 2017] Martínez Soriano, P. A. (2017). *Síntesis y actividad antineoplásica de derivados y/o análogos del ácido cinámico*. PhD thesis, UNAM-FESC.
- [Meng et al., 2011] Meng, X.-Y., Zhang, H.-X., Mezei, M., and Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design*, 7(2):146–57.
- [Meshram et al., 2017] Meshram, R. J., Baladhye, V. B., Gacche, R. N., Karale, B. K., and Gaikar, R. B. (2017). Pharmacophore Mapping Approach for Drug Target Identification: A Chemical Synthesis and in Silico Study on Novel Thiadiazole Compounds. *Journal of clinical and diagnostic research : JCDR*, 11(5):KF01–KF08.
- [Murtaza et al., 2015] Murtaza, G., Sajjad, A., Mehmood, Z., Shah, S. H., and Siddiqi, A. R. (2015). Possible molecular targets for therapeutic applications of caffeic acid phenethyl ester in inflammation and cancer. *Journal of Food and Drug Analysis*, 23(1):11–18.
- [Negi et al., 2018] Negi, A., Bhandari, N., Shyamlal, B. R. K., and Chaudhary, S. (2018). Inverse docking based screening and identification of protein targets for Cassiarin alkaloids against Plasmodium falciparum. *Saudi Pharmaceutical Journal*, 26(4):546–567.
- [Nutti et al., 2018] Nutti, E., Cuffaro, D., Bernardini, E., Camodeca, C., Panelli, L., Chaves, S., Ciccone, L., Tepshi, L., Vera, L., Orlandini, E., Nencetti, S., Stura, E., Santos, M., Dive, V., and Rossello, A. (2018). Development of Thioaryl-Based Matrix Metalloproteinase-12 Inhibitors with Alternative Zinc-Binding Groups: Synthesis, Potentiometric, NMR, and Crystallographic Studies. *J. Med. Chem.*, 61:4421–4435.
- [Petersen et al., 2016] Petersen, L., Blakskær, P., Chaikuad, A., Christensen, A., Dietvorst, J., Holmkvist, J., Knapp, S., Korinek, M., Larsen, L., Pedersen, A., Rohm, S., Slok, F., and Hansen, N. (2016). Novel p38alpha MAP kinase inhibitors identified from yeast DNA-encoded small molecule library. *Medchemcomm*, 7:1332–1339.
- [Phillips et al., 2005] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of computational chemistry*, 26(16):1781–802.
- [Ramírez and Caballero, 2018] Ramírez, D. and Caballero, J. (2018). Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data? *Molecules*, 23(5):1–17.
- [Ruiz Sánchez, 2012] Ruiz Sánchez, M. B. (2012). *Evaluación in vitro de la citotoxicidad de compuestos derivados del éster fenilico del ácido caféico (CAPE), en diferentes líneas celulares de cáncer*. PhD thesis, UNAM.
- [Schlecht, 1998] Schlecht, M. F. (1998). *Molecular modeling on the PC*. John Wiley & Sons, Inc., New York.
- [Shen et al., 2015] Shen, S.-M., Guo, M., Xiong, Z., Yu, Y., Zhao, X.-Y., Zhang, F.-F., and Chen, G.-Q. (2015). AIF inhibits tumor metastasis by protecting PTEN from oxidation. *EMBO reports*, 16(11):1563–1580.

- [Soffritti et al., 2003] Soffritti, M., Minardi, F., and Maltoni, C. (2003). Physical Carcinogens. In Kufe, D., RE, P., RR, W., and Al., E., editors, *Holland-Frei Cancer Medicine*, chapter 21. Hamilton (ON), 6th edition.
- [Srivastava, 2007] Srivastava, R. (2007). *Apoptosis, Cell Signaling, and Human Diseases, molecular mechanisms, volume 2*.
- [Tian et al., 2014] Tian, Y., Zhao, L., Zhang, H., Liu, X., Zhao, L., Zhao, X., Li, Y., and Li, J. (2014). AKR1C3 overexpression may serve as a promising biomarker for prostate cancer progression. *Diagnostic pathology*, 9:42.
- [Tolba et al., 2013] Tolba, M. F., Azab, S. S., Khalifa, A. E., Abdel-Rahman, S. Z., and Abdel-Naim, A. B. (2013). Caffeic acid phenethyl ester, a promising component of propolis with a plethora of biological activities: A review on its anti-inflammatory, neuroprotective, hepatoprotective, and cardioprotective effects. *IUBMB Life*, 65(8):699–709.
- [Verma et al., 2018] Verma, K., Gupta, N., Zang, T., Wangtrakuldee, P., Srivastava, S. K., Penning, T. M., and Trippier, P. C. (2018). AKR1C3 Inhibitor KV-37 Exhibits Antineoplastic Effects and Potentiates Enzalutamide in Combination Therapy in Prostate Adenocarcinoma Cells. *Molecular Cancer Therapeutics*, 17(9):1833–1845.
- [Weininger, 1988] Weininger, D. (1988). SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.*, 28:31–36.
- [Ye et al., 2002] Ye, H., Cande, C., Stephanou, N., Jiang, S., Gurbuxani, S., Larochette, N., Daugas, E., Garrido, C., Kroemer, G., and Wu, H. (2002). DNA binding is required for the apoptogenic action of apoptosis inducing factor. *Nat. Struct. Mol. Biol.*, 9:680–684.