



UNIVERSIDAD DE JAÉN

Facultad de Ciencias Experimentales

Trabajo Fin de Grado

Diseño de fármacos mediante modelización molecular: aplicación a la leucemia mielocítica aguda

Alumno: Jaime Carazo Barrios

Junio 2015, Jaén

RESUMEN (ABSTRACT)

El diseño de fármacos por modelización molecular es un campo que va ganando importancia en los últimos años, no sólo por la cantidad creciente de literatura publicada, sino por el aumento de la potencia y la exactitud de los métodos. El objetivo de este estudio es demostrar la importancia de cierto dominio de BCL6 en la actividad biológica de esta proteína, utilizando la suite de programas de simulación molecular Amber12. Está demostrado que la sobreexpresión de BCL6 tiene significación clínica en varios tipos de linfoma: BCL6 interviene reprimiendo la transcripción de genes que se traducen en células B. El blanco del estudio es el dominio estructural BTB, situado en el extremo N-terminal de BCL6, en su forma complejada con el correpresor SMRT. La identificación de las interacciones más fuertes y estables entre proteína y correpresor es el primer paso para el desarrollo de fármacos contra el cáncer efectivos y específicos.

Drug design by molecular modeling is a field that is gaining importance during the last years, not only because of the increasing amount of literature published, but also because of the rise of power and accuracy in methods. The purpose of this study is to prove the importance of a certain domain of BCL6 in the biological activity of this protein, using the Amber12 suite of molecular simulation programs. It has been proven that overexpression of BCL6 has clinical significance in several types of lymphoma: BCL6 participates suppressing transcription of genes that translate into B cells. The target of the study is BTB structural domain, placed at N-terminal end of BCL6, in its form complexed with corepressor SMRT. Identification of the strongest and most stable interactions between protein and corepressor is the first step to develop effective and specific drugs against cancer.

ÍNDICE

1 Introducción	2
1.1 Antecedentes	2
1.2 Objetivos	12
2 Fundamento teórico	13
2.1 Linfoma difuso de células B grandes	13
2.2 Química física/métodos físico-químicos	16

3 Materiales y método	26
3.1 Materiales	26
3.2 Método	26
4 Resultados	30
4.1 Enlaces de hidrógeno	30
4.2 RMSD y RMSF	35
4.3 ΔG	36
5 Discusión	40
5.1 Enlaces de hidrógeno	40
5.2 RMSD y RMSF	41
5.3 ΔG	42
5.4 Conclusión	43
6 Bibliografía	43

1 INTRODUCCIÓN

1.1 Antecedentes

1.1.1 *BCL6*

1.1.1.1 *BCL6 y su estructura*

BCL6 (B-Cell Lymphoma 6) es una proteína relacionada con la activación y diferenciación de linfocitos, respuesta inmune, control del ciclo celular e inhibición de la apoptosis (1). Esta proteína puede encontrarse en tejidos diversos, pero se encuentra sobreexpresada en varias patologías relacionadas con los linfocitos B (2). Está codificada en el gen *BCL6*.

La proteína *BCL6* en su forma libre tiene unos 700 residuos (3), la isoforma difiere según la especie. En el extremo C-terminal, hay un dominio de 6 dedos de zinc (ZF). Tras estos, una región de residuos sin estructura terciaria. El extremo N-terminal se conforma en un dominio BTB/POZ (4). Este dominio BTB (Broad-Complex, tramtrak) (5), también llamado POZ (Pox Virus and Zinc finger), es el sitio activo de la proteína.

1.1.1.2 Correpressores de BCL6

Un correpresor es una sustancia que inhibe la expresión de un gen, sin unirse directamente al gen (los represores sí se unen). Desde el descubrimiento de BCL6 y su interés biológico, se fueron descubriendo también diversos correpressores, que son necesarios para que BCL6 pueda jugar su función biológica.

BCoR (BCL6 Corepresor) (9). *BCoR* es un gen que codifica (al menos) dos proteínas, BCoR (1721 residuos) y BCoR-S (1004 residuos). BCoR actúa como correpresor de la transcripción, sin embargo BCoR-S podría estar relacionada con el mecanismo de la represión de la transcripción. BCoR está expresado en tejidos humanos muy diversos. Interacciona específicamente con el dominio BTB/POZ de BCL6, impidiendo la interacción de NCOR y SMRT (9). La estructura del complejo BCoR-BCL6 está muy estudiada (8).

CtBP (C-terminal-Binding Protein) (10) es una proteína de unos 350 aminoácidos. Los demás correpressores que se citan están implicados en la represión de la transcripción génica, pero CtBP no lo está. Sin embargo, es clave para la autorregulación de BCL6 (10).

NCOR1 (Nuclear receptor Corepresor) (11) y NCOR2, también llamada SMRT (Silencing Mediator for Retinoid and Thyroid hormone receptors) (11), son correpressores de la transcripción, como BCoR. NCOR1 y NCOR2/SMRT tienen estructuras muy similares, hasta en un 45% (7). Aunque BCoR tiene una estructura muy diferente, los tres correpressores interaccionan con el dominio BTB/POZ de BCL6.

En este TFG, se estudia la estructura tridimensional del llamado complejo BCL6^{BTB}-SMRT^{BBD} (BBD significa BCL6 Binding Domain, dominio de unión con BCL6). Por lo que se conoce hasta el momento, es imprescindible que el dominio BTB de BCL6 dimerize para realizar cualquier actividad biológica. Tras la dimerización, el dominio BTB forma dos 'surcos laterales' con los cuales interaccionan los correpressores conocidos. En concreto, BCoR, NCOR y SMRT (6) son los correpressores conocidos que intervienen en la actividad de represión de la transcripción génica. Se pueden unir hasta dos moléculas, una por cada monómero de BTB, y la unión puede ser homogénea o heterogénea: se han descrito casos de la unión de un monómero de BCL6 con un correpresor y la unión de otro monómero de BCL6 con otro correpresor diferente, dando en cualquier caso un dímero activo como

resultado. La unión de uno de los correpresores con uno de los dos monómeros impide la unión de alguno de los otros a ese mismo monómero. Esta unión se produce de forma directa entre el dominio BBD de los correpresores y el BTB de BCL6 (7,8).

El archivo PDB utilizado en este TFG está conformado por el dímero del dominio BTB (cada monómero de 125 aminoácidos, llamados cadenas C y D) y dos oligopéptidos de SMRT (cada uno de 17 aminoácidos), con un total de 284 aminoácidos, y 18633 moléculas de disolvente (agua). Los oligopéptidos de SMRT se denominan cadenas A y B. (7)

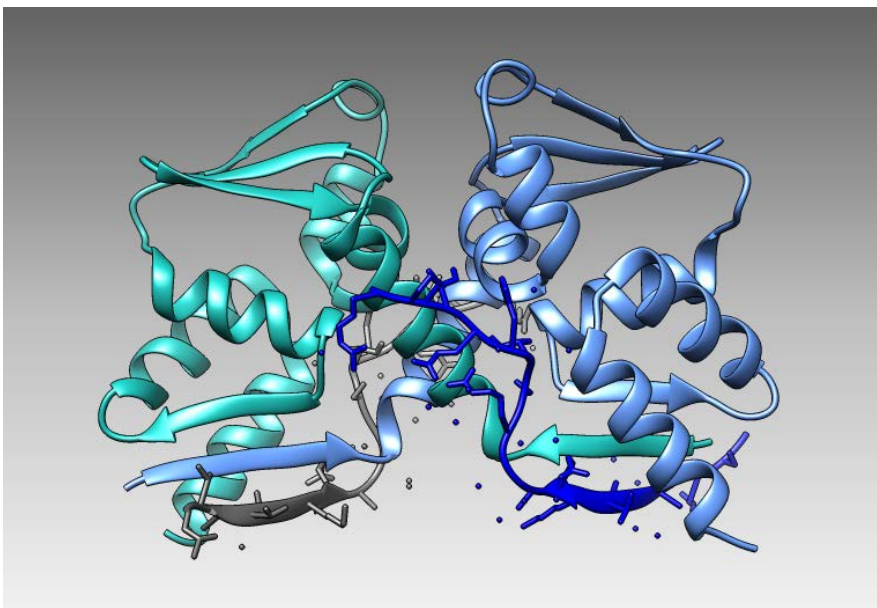


Figura 1: Complejo BCL6^{BTB}-SMRT^{BBD}. A: Azul oscuro B: Gris C: Azul claro D: Azul cyan

1.1.1.3 BCL6 y cáncer

Como ya se ha comentado (apartado 1.1.1.1), la proteína BCL6 está implicada en varias patologías relacionadas con los linfocitos B (linfoma folicular, linfoma MALT, leucemia mielocítica aguda) (2), pero también relacionado con el cáncer de mama (12) o con la aterosclerosis (la forma más común de arteriosclerosis) (13). La bibliografía sobre la relación entre BCL6 y linfomas, BCL6 y linfocitos B, y otras patologías relacionadas con linfocitos B es muy extensa y reciente (1,2,6,14-21). De todas estas patologías, la más frecuente es el linfoma difuso de células B grandes (DLBCL). En esta enfermedad, en el gen *BCL6* la translocación más observada es t(3;14), pero también se observa t(2;3) y t(3;22) (22).

Esta enfermedad se trata con inmunoquimioterapia, y en los casos más agresivos con radioterapia (22). Sin embargo, esta forma de tratamiento daña células sanas, por ello se busca un fármaco que ataque al mecanismo molecular de la enfermedad, cuyo principal exponente es BCL6 (23).

Se han ensayado, o están actualmente en período de prueba, algunos fármacos para el tratamiento del DLBCL. Por ejemplo, un péptido recombinado, conteniendo el dominio BBD de SMRT, un dominio TAT (y algunos otros dominios) bloquea la interacción de BCL6 con SMRT y NCoR (24). En otros artículos (25,4) encontramos estudios de diseño de fármacos por modelización molecular. En uno se encuentran varios compuestos pequeños que atenúan la represión de la transcripción más de un 50%, entre ellos uno (79-6) con mayor afinidad, que se une al dominio BTB de BCL6 matando las células que sufren DLBCL. Sin embargo, queda investigación que hacer, con objeto de mejorar este compuesto y retrasar el momento de su excreción (25).

También encontramos otro estudio de diseño de fármacos en el que se propone, de forma novedosa, el uso combinado de dos fármacos. Al suministrarlos, cada uno de los fármacos interacciona con una parte del dominio BTB de BCL6 (4).

1.1.2 Modelización molecular

La modelización molecular es el compendio de métodos teóricos y computacionales, utilizados para sistematizar e imitar el comportamiento de sistemas químicos. Se mantiene como una herramienta muy utilizada en ciencia hoy en día, en diversos campos como ciencia de materiales, diseño de fármacos, patogenicidad (descripción del proceso fisiológico) de una enfermedad o estudios metabólicos. Este TFG, como casi todos los estudios de diseño racional de fármacos, está basado en técnicas de modelización molecular (suelen aplicarse métodos experimentales posteriormente). Los programas utilizados se detallan en el apartado 3.1 Materiales. Puede señalarse sobre todo la suite de programas Amber (26). Los métodos utilizados se detallan en 2.2 Química física y en 3.2 Métodos.

1.1.2.1 Acoplamiento molecular

Hay varios métodos para la modelización molecular de fármacos, pero entre estos destaca el acoplamiento molecular (molecular docking) (27). Consiste en encontrar y optimizar la

orientación por la que se produce un proceso de interacción molecular. Esto es especialmente interesante para el diseño de nuevos fármacos, ya que estos generalmente interaccionan con el sitio activo de una proteína, y estas interacciones son muy dependientes de la orientación (modelo llave-cerradura). El acoplamiento molecular se basa en dos procesos: primero, muestrear las conformaciones del ligando en el sitio activo de la proteína (con algoritmos de muestreo), y luego ordenar las conformaciones por una función de puntuación (scoring function, SF) (27). Para reducir el número de falsos positivos, se pueden unir dos o más métodos de SF y más funciones de puntuación.

-Algoritmos de muestreo: Las posibles conformaciones (sin restringir) entre el ligando y el sitio activo de la proteína son muchas, y explorarlas todas supondría un gran gasto computacional. La misión de los algoritmos de muestreo es limitar el espacio conformacional en el cual se realiza la búsqueda, para intentar llegar a acierto con menos gasto computacional (de forma más eficiente). Pueden utilizarse tres enfoques para el acoplamiento: considerar tanto el ligando como el receptor rígidos, ligando flexible y receptor rígido, o ambos flexibles. El enfoque más utilizado es el de ligando flexible y receptor rígido, de modo que se pueda alcanzar un acuerdo entre exactitud y coste computacional (27).

-Funciones de puntuación: El papel de estas funciones es evaluar cada posible distribución relativa ligando-receptor, y calcular la intensidad de la interacción (27). Hay tres tipos de funciones: basadas en campos de fuerzas, en conocimiento o empíricas (28). Normalmente, se utiliza una combinación de varios tipos, conocida como puntuación por consenso (27).

En la puntuación basada en campos de fuerzas, se cuantifica la intensidad de la interacción ligando-receptor sumando las interacciones intermoleculares electrostáticas y de van der Waals. En la aproximación usada en este TFG, llamada protocolo de trayectoria única, se utiliza para el cálculo una misma trayectoria de dinámica molecular, y por ello los términos de interacción intramolecular se cancelan entre sí (energía interna e interacciones 1-4). Esta aproximación puede reducir significativamente el ruido, y cancelar errores (29). Se utiliza la ecuación de Coulomb para calcular los términos electrostáticos, y para los términos de van der Waals se utiliza la función de Lennard-Jones. Estos conceptos se explicarán con más detenimiento en el apartado 2.2.3 ΔG . La función de puntuación basada en campos de fuerzas es más lenta computacionalmente que otras (27).

En las SFs basadas en conocimiento, se hace análisis estadístico del complejo ligando-receptor para averiguar la frecuencia de contacto y la distancia entre ambos. Se hace la frecuencia de contacto asimilable a la fortaleza de la interacción: cuanto más fuerte sea una interacción, más probable es su ocurrencia. Este tipo de puntuación tiene poco coste de computación y este es su principal atractivo, sin embargo puede fallar en interacciones con metales o halógenos (27).

En las funciones de puntuación empíricas, la energía de interacción se descompone en sus componentes (enlaces de hidrógeno, interacción electrostática...). Cada componente se va ajustando por regresión lineal y se va sumando al valor final. El ajuste se realiza comparando con un grupo de complejos ligando-proteína, cuya afinidad conocemos de forma experimental (27).

1.1.2.2 Otros

Dada la creciente complejidad de los problemas científicos, los últimos estudios sobre estos temas van tendiendo (de forma lógica) hacia el uso combinado de varias técnicas, como acoplamiento molecular y dinámica molecular (30,31) o uniendo técnicas complejas de espectroscopia RMN (32). Se investigan temas fisiológicos diversos con intención de aplicarlos a la medicina, por ejemplo la forma que tiene la fosforilasa de Eschericia Coli (33) o para cuantificar las relaciones estructura-actividad de diversos inhibidores de la NADH deshidrogenasa (34). También la cascada de señalización de la proteína quinasa ErbB4, cuya mutación está relacionada con el fenotipo celular de varios cánceres (35). Finalmente, se cita el uso (con gran espacio para el avance) de macromoléculas con estructura de dendrímeros como vectores de medicamentos o agentes de diagnóstico (por ejemplo, radioisótopos para diagnosticar algunos cánceres, hipertiroidismo...) (36).

Resulta interesante hablar sobre el programa BOINC, de computación voluntaria en red (37). Fue desarrollado en 2002 en la Universidad de California, Berkeley. Básicamente, consiste en un programa que usa la capacidad de computación no utilizada de un ordenador para cálculos científicos en proyectos de todo el mundo. BOINC tiene más de 70 proyectos activos, algunos de ellos relacionados con la modelización molecular. Por ejemplo, CAS (38) se ocupa de la predicción de la estructura de proteínas, GPUGRID (39) investiga diversos temas en cáncer, VIH/SIDA y otros desde Barcelona, Ibercivis (40)

engloba proyectos tanto de diseño de fármacos como de astronomía o divulgación científica desde Madrid, World Community Grid (41) investiga ébola, malaria, cáncer infantil o lucha contra el hambre, o Rosetta (42), que investiga sobre la predicción de la estructura tridimensional de proteínas relacionadas con el cáncer, malaria, y muchas otras enfermedades.

1.1.3 Diseño de fármacos

El diseño de fármacos es la utilización de diversos métodos computacionales y técnicas experimentales para encontrar nuevos fármacos o mejorar fármacos existentes. Pueden mejorar los fármacos al tratar más síntomas de la enfermedad, conseguir los mismos efectos usando una dosis menor, llevar aparejados menos efectos secundarios... Hay dos tipos de enfoque, basado en ligando y basado en estructura (o receptor).

1.1.3.1 Enfoque basado en ligando

Cuando la estructura 3D de la molécula objetivo es desconocida, o no puede ser predicha por técnicas de modelización, una técnica muy usada es la QSAR (relaciones cuantitativas estructura-actividad). Como ya se ha comentado (1.1.2.2), la creciente complejidad de los problemas científicos deriva hacia un enfoque de técnicas integradas. Por tanto, aunque el enfoque basado en ligando se utilice menos, sí se utiliza integrado con otras técnicas (43).

El principio subyacente a estas técnicas es que la estructura molecular tiene características inherentes a sus propiedades (físicas, químicas y biológicas). Los modelos QSAR intentan encontrar una relación entre la estructura y las propiedades (químicas, físicas, actividad biológica...) de una molécula. Se basa en moléculas que se unen al receptor, es decir, que son activas frente a éste (43). A partir de la molécula activa que ya se conoce, se obtienen compuestos con propiedades mejoradas por modificaciones químicas (44).

Los modelos QSAR se diseñaron para encontrar la relación entre el valor calculado, relacionado con la propiedad, y el valor experimental. QSAR es también útil para el screening de fármacos efectivos y para anticipar posibles problemas en el desarrollo de nuevos agentes antes de realizar acoplamiento molecular o determinación experimental de toxicidad, tanto in vitro como in vivo. La necesidad de técnicas QSAR se hace patente cuando familias de compuestos bioactivos se enfrentan con poco conocimiento de la

estructura del blanco. Si se encuentra una relación entre algunas propiedades fisicoquímicas y la actividad biológica de ciertos compuestos, se hace posible predecir la actividad de compuestos modificados o relacionados (44).

Esta técnica ayuda y/o es precursora de otras de modelización molecular y diseño de fármacos basado en estructura.

1.1.3.2 Enfoque basado en estructura.

Se busca conocer e imitar los dominios y el centro activo para inducir o neutralizar alguna actividad biológica (45). Este es el enfoque más utilizado en diseño de fármacos. En base a la forma del sitio activo se proponen diferentes estructuras para el medicamento (44). La figura 1 muestra un proceso de diseño de fármacos, pero puede decirse que tiene 3 pasos generales:

1-Identificar la diana (proteína, ADN, etc) y conseguir estructura 3D

2-Diseño de novo o screening (random o docking)

3-Test preclínico y clínico

1-La primera etapa incluye lo relacionado con identificar la diana vinculada a la enfermedad, e identificar la estructura del sitio activo.

En esta etapa, se puede usar técnicas experimentales, como cristalografía de rayos X, espectroscopia (por ejemplo RMN o dicroísmo circular), microscopía electrónica... Y también técnicas computacionales, especialmente la modelización por homología.

En ésta, se toma la secuencia de aminoácidos de una proteína con estructura 3D conocida, y se compara con la secuencia de aminoácidos de una proteína con estructura 3D desconocida, para intentar deducirla. Se puede aplicar si las proteínas comparten al menos un 30% de la secuencia. Los motivos conservados y otros aspectos de la estructura secundaria se pueden deducir con bastante precisión en base a la estructura primaria. Varios aspectos de la estructura terciaria y cuaternaria suelen ser similares en proteínas con secuencia de aminoácidos similar, pero no siempre, ya que en ocasiones hay proteínas (chaperonas) que ayudan al plegamiento de otras proteínas. Por esto se utilizan varias secuencias de aminoácido de proteínas con estructura conocida, para asegurar que

el resultado sea más certero (44). También se puede intentar calcular la estructura terciaria de una proteína directamente a partir de la primaria. Este es el método *de novo*, de gran complejidad y coste computacional.

2-Se pueden llevar a cabo o el diseño de moléculas *de novo*, o la búsqueda por bibliotecas conocidas de compuestos y test iniciales para evaluar la viabilidad de compuestos (screening). Esta búsqueda se realiza en base a un farmacóforo, que es la unidad mínima con la actividad biológica que buscamos, por ejemplo en un fármaco. El screening puede ser aleatorio, o automatizado para buscar en base a cierta capacidad de los compuestos (inhibición o activación de alguna función proteica, mensajero celular...). La técnica para automatizar esta búsqueda se llama High Troughput Screening (HTS).

El screening también puede realizarse por técnicas génicas, como utilizando al promotor proveniente de ratones PEG-3 (46). PEG-3 se ha demostrado efectivo inhibiendo el crecimiento de células. Se prueban moléculas y sólo pasan al siguiente nivel las que presentan actividad contra ese inhibidor (46). Tras esta fase, se ordenan los compuestos prometedores en series, y los más prometedores se llaman cabezas de serie. Se suele volver a los ensayos para acotar los resultados aún más, antes de pasar a la fase 3.

3-Una vez superadas las fases ya comentadas, se pasaría a los test preclínicos y clínicos (del 0 al 4). En los test preclínicos se prueba el medicamento en cultivos celulares. Los test clínicos varían en dosis suministrada y número de participantes desde fase 0 (animales) o 1 (dosis muy pequeñas, pocos individuos) hasta fase 4 (mayores dosis, todos los pacientes).

A los ensayos preclínicos se les suele denominar test *in vitro*. A los ensayos clínicos, test *in vivo*, en contraste a los realizados en ordenador, a los que se les llama test *in silico*.

Un tipo de enfoque basado en estructura es el enfoque retrometabólico, que puede definirse como el diseño de fármacos enfocado al balance entre la actividad biológica necesaria y la toxicidad máxima tolerable. Este, a su vez, tiene varias estrategias, como añadir una fracción a los medicamentos que permita controlar la fase de excreción. Para esto pueden utilizarse análogos naturales (por ejemplo, progesterona). Una estrategia especialmente útil es la del metabolito inactivo, que comienza el diseño del medicamento a

partir de los metabolitos excretados que se detectan, de forma análoga al análisis retrosintético. Estas estrategias y otras pueden combinarse entre sí (46).

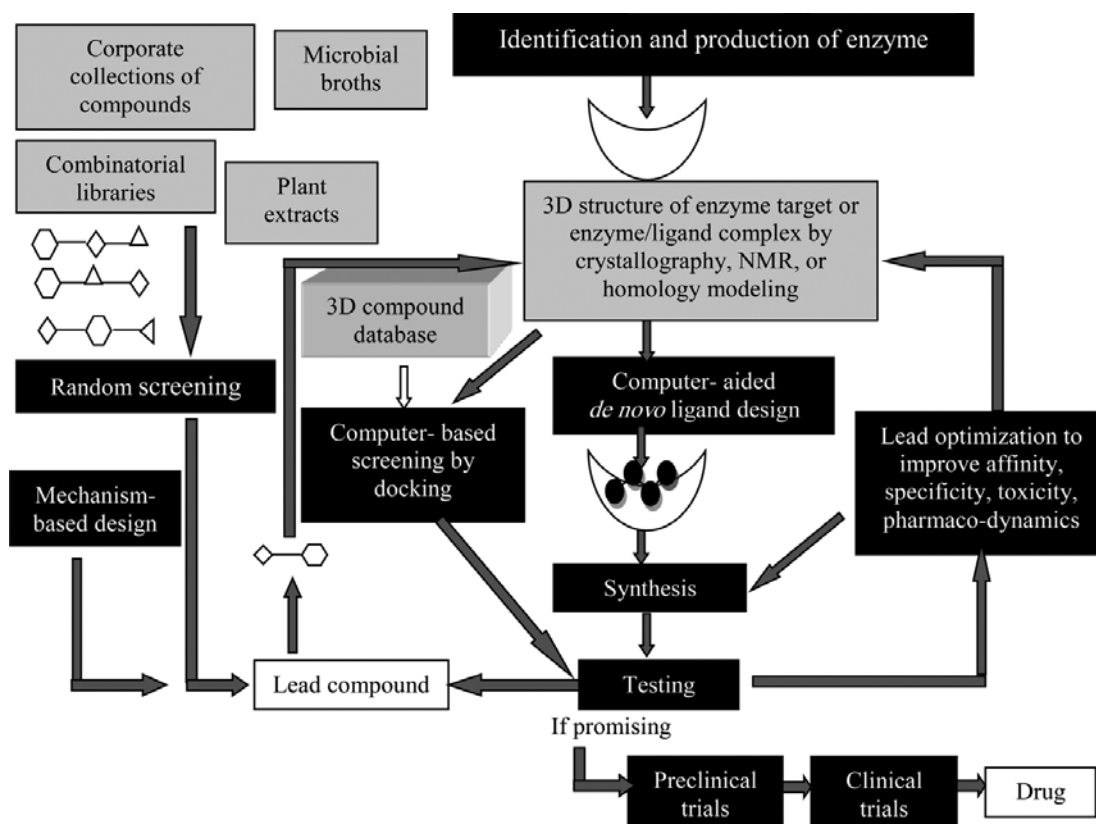


Figura 2: Esquema diseño de fármacos (44)

1.1.4 Proteómica

El proteoma es el conjunto de proteínas de un organismo, incluida toda modificación que hayan sufrido (47). Esto puede variar debido al estado del sistema (temperatura, estrés...).

La proteómica es el estudio de las proteínas, con el objeto de entender procesos biológicos (por ejemplo, el proceso de una enfermedad o la acción de un medicamento) y “descifrar los mecanismos de control de expresión génica”. (48)

Estos dos conceptos son interesantes en este TFG, porque los métodos utilizados en varios de los artículos consultados pueden clasificarse en el campo de la proteómica. (4,7,8,9,11,21,25).

La proteómica es una ciencia de importancia creciente en el estudio del cáncer, porque a menudo el proceso de esta enfermedad implica mecanismos proteicos complejos (49), y porque normalmente los biomarcadores y las dianas farmacológicas (en la fase de ensayo clínico) son proteínas, aparte de lo útil que es la proteómica para cribar candidatos a medicamentos (50).

De hecho, el número de publicaciones relacionadas con la proteómica ha aumentado mucho en los últimos diez años (49). Y a pesar de los prometedores resultados, queda aún mucho camino por recorrer, como por ejemplo ampliando estudios funcionales y validaciones clínicas (50). Se resumen tres usos terapéuticos de técnicas proteómicas (49):

-Identificación de dianas en población resistente al cáncer

El cáncer es una enfermedad con gran capacidad adaptativa frente a los medicamentos y a cualquier estímulo. El uso de técnicas como el knockdown de genes puede revertir esa sensibilización y hacer que el paciente vuelva a ser susceptible al tratamiento. (50)

-Identificación de dianas por la acción del fármaco (proteómica química)

Se utilizan pruebas basadas en la actividad, o pruebas basadas en técnicas tándem LC-MS (cromatografía líquida-espectrometría de masas). Se consigue identificar las proteínas implicadas en la actividad biológica, incluido su nivel de modificación, o conocer el modo de acción, por ejemplo el nivel de modulación de la expresión génica, o la inhibición de caminos de señalización. El objetivo último es conocer las interacciones específicas que definen la diana terapéutica, un gran paso para el desarrollo de nuevos fármacos. (49)

-Descubrimiento de dianas basada en el diferente estado de los tejidos (proteómica de tejidos).

Se analizan tejidos tanto cancerosos como benignos por electroforesis diferencial 2D acoplada con espectrometría de masas. Se compara la expresión diferencial de las proteínas en tejido enfermo y tejido sano. Esto lleva a un mejor diagnóstico (tanto en efectividad como en anticipación) (49).

1.2 Objetivos

El objetivo de este TFG es realizar un estudio de las interacciones presentes entre el dominio BTB de BCL6 y la secuencia de 17 aminoácidos de SMRT conocida como BBD, cuando se encuentran en forma del complejo BCL6^{BTB}-SMRT^{BBD}. Este estudio se realizará mediante una simulación de dinámica molecular con la suite de programas Amber.

El análisis de la dinámica molecular permitirá identificar qué residuos están implicados en las interacciones más fuertes entre BCL6^{BTB} y SMRT^{BBD}. Conocer estos residuos permite establecer un farmacóforo (sección 1.1.3.2), que represente tales interacciones. Este farmacóforo es el primer paso para desarrollar un fármaco contra las enfermedades en las que BCL6 está implicada. Este fármaco actuaría uniéndose a la zona de interacción de BCL6^{BTB} e impidiendo su interacción con SMRT^{BBD}. Los resultados que se obtengan en este TFG se compararán con los obtenidos de varias fuentes bibliográficas, en el apartado de discusión.

2 FUNDAMENTO TEÓRICO

2.1 Linfoma difuso de células B grandes

El DLBCL es el tipo de cáncer que la desregulación de BCL6 induce en mayor medida. Esto se explica en el apartado 1.1.1.3 (página 9).

2.1.1 Definición cáncer

Pero ¿qué es el cáncer? “«Cáncer» es un término genérico que designa un amplio grupo de enfermedades (...). Una característica del cáncer es la multiplicación rápida de células anormales que se extienden más allá de sus límites habituales y pueden invadir partes adyacentes del cuerpo o propagarse a otros órganos (...)” (51)

Se puede entender el cáncer como la suma de dos procesos: la proliferación incontrolada de las células que forman un tejido (tumor) y la colonización de otros tejidos por migración a los tejidos adyacentes (invasión) o por propagación mediada por el aparato circulatorio y crecimiento en otro tejido (metástasis). Si un tumor no tiene capacidad de realizar metástasis, se le llama benigno, y maligno si tiene esa capacidad (52).

La gran mayoría (90-95%) de los cánceres ocurren por exposición a factores ambientales (tabaco, polución, infecciones...) que dañen el ADN y desemboken en la formación de un cáncer (no cualquier daño al ADN desemboca en cáncer). Una mínima parte (5-10%) están relacionados con la genética (siendo independientes de factores ambientales). A su vez, estos pueden estar relacionados con daño al ADN proveniente de procesos celulares en los que intervenga el ADN, o con la herencia genética (o con ambos).

El proceso de conversión de células sanas a células cancerígenas (o a tumores) y formación de cáncer se llama carcinogénesis. La diferencia entre que un tumor sea benigno o maligno radica en la carga genética. Los factores genéticos de cada tipo de cáncer son extremadamente variables, dando una naturaleza muy diferente para los más de 200 cánceres que afectan a las personas.

Sin embargo, podemos generalizar diciendo que para que se forme un cáncer tienen que alterarse los genes que regulan el crecimiento celular (tanto los inhibidores como los estimulantes). Antes de la aparición del tumor (neoplasia) en algunos cánceres aparecen lesiones precancerosas: displasia (aparición de células de forma aberrante) y/o la metaplasia (la célula cambia de tipo de tejido). El diagrama para la formación de tumores es el siguiente (52)

Daño al ADN --> Mutación genética --> Alteración de un ARN mensajero (o varios) --> Alteración de una proteína (o varias), o de uno (o varios) procesos metabólicos --> Alteración de la función celular --> Aparición del tumor

2.1.2 Tipos de cáncer

Se pueden hacer varias clasificaciones de los diferentes tipos de cáncer, y la OMS va actualizando su clasificación según se va avanzando en el campo y se va llegando a acuerdos sobre el tema. Aquí se resume la clasificación más utilizada, tipos de cáncer según el tejido afectado.

-Carcinoma: Cáncer de las células epiteliales. Puede ser, a su vez, de las células de la piel (melanoma) o de las células epiteliales que forman el revestimiento interno de los órganos (adenocarcinoma).

-Sarcoma: Cáncer del tejido conjuntivo (huesos, músculos, cartílagos, vasos sanguíneos...).

-Cáncer de células madre: Puede ser de células germinales (precursoras de espermatozoides y de óvulos) o de células parcialmente diferenciadas (blastoma).

-Hematopoyéticos: Cáncer de las células que forman la sangre. Puede ser, a su vez, de la médula ósea (leucemia) o del sistema linfático (linfoma):

Leucemia: Puede ser mieloide/linfoide, o aguda/crónica, dando 4 tipos generales.

Linfoma: Puede ser de linfocitos B, de linfocitos T y células NK, enfermedad de Hodgkin o desórdenes linfoproliferativos asociados con la inmunodeficiencia.

2.1.3 ¿Qué tipo de cáncer se estudia en este TFG?

DLBCL es un tipo de linfoma de linfocitos B. Los linfocitos son células del sistema inmunitario, que nos defiende contra enfermedades y otros agentes externos (53,54). Este sistema está formado mayoritariamente por leucocitos. Los linfocitos son un tipo de leucocitos (53,54). A su vez, los linfocitos se dividen en B, T y asesinos naturales (natural killers).

Los linfocitos B producen anticuerpos (también llamadas inmunoglobulinas) para reconocer antígenos específicos. Los linfocitos T reconocen antígenos presentados por el complejo mayor de histocompatibilidad. Los natural killers participan también reconociendo antígenos.

El proceso de formación de los componentes celulares de la sangre (glóbulos rojos, glóbulos blancos y plaquetas) se denomina hematopoyesis (la formación de linfocitos, linfopoyesis). La hematopoyesis se produce en la médula ósea, que se encuentra dentro de algunos huesos (vértebras, pelvis, cráneo...). Después, las células formadas se transportan hasta su destino. Los linfocitos van hasta el sistema linfático (la linfa tiene un alto contenido en linfocitos) y al aparato circulatorio (los linfocitos pueden representar entre un cuarto y un tercio de la sangre de una persona) (53,54).

La diferencia entre los diferentes linfomas reside en el momento exacto en el que sufren la transformación linfomatosas (de célula normal hacia célula cancerosa). Ésta puede ocurrir en la linfopoyesis, o en el transcurso de la respuesta inmunológica frente a un antígeno. El DLBCL surge de células B que están en el último estadio de la linfopoyesis, cuando están a punto de transformarse en células plasmáticas (24). La patogénesis (origen y desarrollo de una enfermedad) del DLBCL se encuentra revisada de forma extensa (6). Como el nombre de la enfermedad indica, las células que la sufren (denominadas linfomatosas) son de gran tamaño. Ya se ha comentado (apartado 1.1.1.3) que en las personas afectadas por el DLBCL se observan translocaciones en el gen *BCL6*. La translocación más observada es t(3;14), pero también se observan t(2;3) y t(3;22). El diagrama (simplificado) para la aparición de DLBCL sería el siguiente:

Daño a *BCL6* --> Translocaciones (entre ellas, t(3;14)) --> Alteración del ARN mensajero --> Sobreexpresión de la proteína BCL6 --> Alteración de la función de BCL6 --> Alteración de ciclo de vida celular --> Aparición del DLBCL

Encontramos las proteínas CD10, CD20, CD22, CD29, CD45 y CD79^a (63). A estas proteínas se les llama cúmulo de diferenciación (cluster of differentiation, CD), y funcionan como marcadores que nos permiten identificar la célula y algunas características (como el nivel de diferenciación). Las translocaciones y las CD encontradas definen el fenotipo de esta enfermedad.

2.2 Química física/métodos físico-químicos

2.2.1 Introducción

En modelización molecular pueden utilizarse tres enfoques, dependiendo del nivel de complejidad del modelo: la mecánica molecular (MM), que estudia los sistemas utilizando las leyes de Newton, acercándose a los átomos, la mecánica cuántica (QM), que da un paso más y también estudia los electrones, y los modelos híbridos cuántico-clásicos (QM/MM) (55)

Dentro de la química cuántica, se utilizan (entre otros) métodos *ab initio* y semiempíricos. En ambos métodos se utiliza el principio variacional para resolver la ecuación de Schrödinger de forma aproximada. El método variacional se basa en el hecho de que la función de ondas real de un sistema dará la mínima energía posible al aplicársele el

hamiltoniano del sistema; una función de ondas que no sea la real dará en cualquier caso una energía superior como resultado. Se utiliza una función de ondas de ensayo, que se minimiza mediante un método iterativo, hasta que entre dos iteraciones consecutivas la diferencia de energía calculada quede por debajo de un umbral predefinido.

Entre las diferentes metodologías que usan el principio variacional, la más simple es el método de Hartree-Fock, en el cual se evalúan las ecuaciones de Hartree-Fock/Roothaan-Hall en su totalidad. De forma matricial, son:

FC=SCE

F es la matriz de Fock. **C** es una matriz de coeficientes. **S** es la matriz de solapamiento. **E** es la matriz de energía orbital. A su vez éstas tienen otras componentes, de cierta complejidad, y la matriz **F** sólo se aplica a un electrón, por tanto en estos métodos se ocupa gran parte del tiempo calculando y manipulando integrales. Esto hace que tengan gran coste computacional y sólo puedan usarse para sistemas pequeños (57).

Los métodos *ab initio* están basados enteramente en mecánica cuántica y constantes físicas básicas. Actualmente hay disponible una amplia variedad de programas que utilizan métodos *ab initio*. Uno de los más conocidos es la serie de programas Gaussian (56).

Por otra parte, los métodos semiempíricos utilizan la misma metodología, pero aceleran el proceso haciendo más aproximaciones, ayudándose de datos experimentales. Para el ejemplo anterior de cálculo Hartree-Fock, la aproximación sería igualar la matriz de solapamiento a la matriz identidad, y por tanto la ecuación queda de esta forma:

FC=CE

Las aproximaciones permiten que estos métodos puedan utilizarse para estudiar sistemas de mayor tamaño (57).

2.2.2 Métodos mecánica molecular

En mecánica molecular se utiliza un conjunto de funciones potenciales, denominado campo de fuerzas, para expresar la energía de un sistema. Los campos de fuerzas realizan grandes aproximaciones, pero permiten estudiar sistemas de más átomos. Las funciones potenciales que conforman un campo de fuerzas dependen de una serie de parámetros,

como masa atómica, cargas parciales o términos energéticos asociados al enlace (stretching, bending...).

En este TFG, se utiliza dinámica molecular para simular el comportamiento del sistema BCL6^{BTB}-SMRT^{BBD} a lo largo de un periodo de tiempo de 5 ns, obteniéndose un fichero con los datos de la trayectoria (el objetivo se menciona en el apartado 1.2 Objetivos). Posteriormente, se utilizan otros métodos para extraer de la trayectoria los parámetros que nos interesan (enlaces de hidrógeno, energía libre...).

La dinámica clásica molecular es probablemente el método más conocido y extendido para el estudio de la flexibilidad proteica. En este método, se aplica un campo de fuerzas clásico a los átomos de la proteína y del disolvente. Los parámetros de este campo se calculan con cálculos mecanocuánticos, y/o con observables experimentales, tomados de sistemas modelo (55).

Durante la simulación, se permite que los átomos y las moléculas interactúen por un periodo de tiempo. El movimiento de cada átomo es calculado y dispuesto para analizar el comportamiento global, cada átomo tiene asociada su función potencial. El algoritmo de soporte para una simulación de dinámica molecular incluye: 1) la determinación de la posición y la velocidad iniciales de cada átomo; 2) el cálculo de las fuerzas que experimenta cada átomo (se utilizan potenciales interatómicos); 3) la evolución (integración) de la posición y la velocidad durante un periodo corto de tiempo, llamado timestep (2 femtosegundos, en este TFG). El paso 3) proporciona una nueva posición y velocidad, los cuales funcionan como posición y velocidad inicial para el paso 2) (43). Esto se repite hasta el final de la simulación. Suele ser complicado discernir si una trayectoria es lo suficientemente larga como para explorar de forma adecuada el espacio conformacional.

Hay dos técnicas que se pueden utilizar en las simulaciones para obtener a partir de una trayectoria de dinámica molecular datos de energías libres de unión mediante el uso del programa MMPBSA.py: la aproximación de trayectoria única (STP) y la aproximación de trayectoria múltiple (MTP). El primer método extrae el estado del complejo, receptor y ligando de una misma trayectoria, combinándose las trayectorias intermedias que cada procesador ha analizado (para el complejo, el receptor y el ligando) en una trayectoria única analizada por todos los procesadores. La aproximación de trayectoria múltiple utiliza trayectorias separadas para ligando, proteína y complejo (59).

El STP es un método más rápido, porque sólo se simula un sistema, y que puede minimizar el ruido, pero asume que la configuración espacial que ocupan el receptor y el ligando no varía entre el estado enlazado y no enlazado. Esto hace que la energía relativa a las distorsiones en el complejo no sea evaluada. Esto puede no inducir a errores de bulto, si no hay efectos derivados de cambios de conformación (típicos en proteínas y enzimas). Para el ligando, sin embargo, estos efectos pueden acumularse hasta un valor importante si el estado enlazado adopta una conformación con más tensión que la del ligando no enlazado en disolución. Estos errores pueden cancelarse parcialmente cuando se comparan ligandos similares con el mismo receptor. Pero evidentemente esto añade confusión a los resultados. Cuando se esté en duda, se debería ejecutar una trayectoria del ligando en solitario, para obtener la energía media del ligando en estado no enlazado.

Una trayectoria única no produce diferencias en los términos energéticos de longitudes de enlace, ángulos o diedros entre las estructuras del complejo, receptor y ligando. Estos términos (longitud de enlace, ángulo de enlace, diedros e interacciones 1-4) se calculan, pero al restarse deberían cancelarse completamente.

El cálculo de las interacciones agua-agua es, en general, muy costoso computacionalmente, y puede hacerse con métodos de disolvente explícito o implícito. Los métodos de disolvente explícito (expresado directamente) tienen la ventaja de reflejar las interacciones entre soluto, agua e iones, por ejemplo en la formación de enlaces de hidrógeno. Todas estas interacciones pueden ser determinantes para la estructura molecular y la función. Sin embargo, el alto coste de calcular estas interacciones, y de explorar completamente el espacio de fase, puede hacer deseable el uso de métodos de disolvente implícito. Estos métodos representan el disolvente como un dieléctrico continuo, simplificando el cálculo de las interacciones entre pares disolvente-soluto.

Añadido a los efectos polares, ambos modelos deben reproducir los componentes apolares de las interacciones del disolvente (normalmente, agua), en su mayoría interacciones de van der Waals, y el coste entrópico de crear una cavidad e introducir el soluto en el disolvente (sin embargo, los términos entrópicos no suelen calcularse) (60).

Dos métodos que combinan mecánica molecular y modelos de disolvente implícito son los métodos de Mecánica Molecular/Poisson Boltzmann/Área Superficial (MM/PBSA) (61) y de

Mecánica Molecular/Born Generalizado/Área Superficial (MM/GBSA) (62). En este TFG se usan ambos métodos. Estos métodos han sido ampliamente utilizados para el cálculo de energías libres. Comparados con otros métodos más rigurosos, como los métodos de perturbaciones de energía libre (FEP) o los métodos de integración termodinámica (TI), MM/PBSA y MM/GBSA son más eficientes computacionalmente. La contribución polar a la energía de solvatación se calcula con el método PB o GB, y la contribución no polar con el método de área superficial accesible por el disolvente (29):

2.2.2.1 Método Poisson-Boltzmann

Es un método muy utilizado para calcular potenciales electrostáticos, es decir, cargas eléctricas estacionarias o con un movimiento lento no acelerado. El método se basa en la ecuación de Poisson-Boltzmann (PB), que es la siguiente:

$$\Delta[\epsilon(r)\Delta\Phi(r)] = -4\pi \rho(r) - 4\pi \lambda(r) \sum_i z_i c_i \exp(-z_i \Delta\Phi(r)/k_B T)$$

Donde $\epsilon(r)$ es la constante dieléctrica, $\Phi(r)$ el potencial electrostático, $\rho(r)$ la densidad de carga del soluto, $\lambda(r)$ la función de Stern, z_i la carga del ión i , c_i la densidad de ión i lejos del soluto, k_B la constante de Boltzmann, y T la temperatura.

Para usarla en MMPBSA.py se incluye &pb en el archivo de entrada. El usuario puede especificar variables termodinámicas externas para controlar el cálculo, basadas en condiciones experimentales, por ejemplo la fuerza iónica (istrng) y las constantes dieléctricas internas y externas (indi y exdi, respectivamente). Se recomienda utilizar el método por defecto para calcular la energía libre de solvatación no polar y sus variables asociadas (cavity_surften y cavity_offset), por optimización previa. El valor por defecto de la variable fillratio de 4.0 es recomendado, para evitar errores provenientes de que pequeñas moléculas queden fuera del cálculo, cuando el valor de fillratio se establece demasiado bajo. La escala es el número de puntos de red de PB por angstrom, y puede ajustarse para obtener mayor o menor resolución, siempre teniendo en cuenta que aumentar la resolución lleva aparejado un aumento en el tiempo de computación. Se establece un valor por defecto de 2.0, entendiendo éste como un compromiso entre velocidad y precisión. Se sugieren 1000 iteraciones para la ecuación de PB en su forma lineal (linit), éste valor es normalmente suficiente para que la simulación converja (63).

2.2.2.2 Método de Born Generalizado (GB)

El método GB es similar al método PB, tomando los iones como esferas con diferente constante dieléctrica:

$$G_s = (1/8\pi) (1/\epsilon_0 - 1/\epsilon) \sum_{A,B} (Q^A Q^B / f_{GB})$$

Donde ϵ_0 es la permitividad del vacío, y f_{GB} es una función de la distancia y la interacción disolvente-soluto.

El cálculo GB tiene sus propias variables a controlar, como concentración salina (saltcon) y el modelo GB específico utilizado (igb). Amber ofrece cinco modelos GB diferentes para usar con MMPBSA.py. Para el análisis de proteínas, se recomienda igb = 8 (para sander, que es un programa utilizado para simulaciones de dinámica molecular), basado en estudios recientes que muestra mayor estabilidad de los puentes salinos con este modelo.

Sin embargo, un valor de 2 para igb suele preferirse si no se tiene acceso a sander. MMPBSA.py puede usar sander para calcular parte del sistema del cálculo MMGBSA de forma cuántica, usando un hamiltoniano híbrido (QM/MM), con la variable ifqnt (59,63).

2.2.2.3 Método del área superficial accesible

El área superficial accesible por el disolvente (SASA) es el área superficial de una molécula que puede interaccionar con el disolvente. Por ejemplo, el plegamiento de una proteína sobre si misma esconde ciertas partes, que no estarían accesibles al disolvente mientras que la estructura tridimensional se conserve. ΔG_{SA} forma parte de la energía libre de solvatación no polar.

$$\Delta G_{SA} = \sum \sigma_i ASA_i$$

σ_i es un parámetro que indica el nivel de solvatación del átomo i, y ASA el área superficial accesible por el átomo i.

2.2.3 ΔG

En química computacional, uno de los objetivos es calcular la energía libre de Gibbs de los procesos de interacción (reacciones, solvatación...). En este TFG, se calcula la energía libre del proceso de interacción entre proteína (BCL6) y ligando (SMRT). ΔG se descompone en varias componentes:

$$\Delta G = \Delta G_{\text{gas}} + \Delta G_{\text{solv}} - TS$$

$$\Delta G_{\text{solv}} = \Delta G_{\text{polar}} + \Delta G_{\text{nopolar}}$$

$$\Delta G_{\text{gas}} = E_{\text{int}} + E_{\text{vdW}} + E_{\text{el}}$$

$$E_{\text{int}} = E_{\text{enlace}} + E_{\text{ang}} + E_{\text{dihed}}$$

ΔG_{solv} representa la energía de solvatación. Es la suma de la energía necesaria para quitar todas las cargas y volver a situarlas con presencia de un disolvente (esta es la energía de solvatación polar, representada por ΔG_{polar} o EPB/GB, según si se utiliza el método PB o GB) y de la energía necesaria para disolver una molécula cuyas cargas han sido retiradas (esta es la energía de solvatación no polar, representada por $\Delta G_{\text{nopolar}}$ o ESURF).

De forma general, $\Delta G_{\text{nopolar}}$ viene del efecto combinado de dos interacciones: la atracción de van der Waals entre el soluto y el disolvente, y el coste desfavorable de romper la estructura del disolvente alrededor del soluto (esta parte se entiende proporcional al área superficial accesible, SA, y se simboliza como ΔG_{SA} , explicada en el apartado 2.2.2.3 (26)).

ΔG_{gas} (o E_{gas}) es la energía total en fase gaseosa. Está formada por la energía interna (a su vez, formada por las energías de longitud de enlace, ángulo de enlace y diedros), la energía de van der Waals (E_{vdW}) y la energía electrostática (E_{el}).

Cada una de las componentes se computa de forma diferente. Computar absolutamente todas las interacciones es normalmente demasiado costoso para tamaños biomoleculares estándar. Para ahorrar coste computacional, se calculan sólo ciertas interacciones. La mayoría de los campos de fuerzas calculan la energía derivada de la tensión de enlace para átomos a dos enlaces de distancia, la flexión para átomos a 3 enlaces, torsión para átomos a 4 enlaces, y E_{el} y E_{vdW} para átomos a 4 enlaces o más (58). Ya que el cálculo de los términos energéticos está basado en parámetros y en conceptos como radio atómico y cargas parciales, es inevitable que surja un debate sobre la calidad de los parámetros (59). La mayor contribución a la ΔG total viene por las contribuciones de van der Waals y electrostática, que se explican a continuación:

2.2.3.1 Contribución electrostática

Es la energía debida a las cargas de los átomos, tanto permanentes como inducidas. La ecuación más básica que se utiliza es la de Coulomb:

$$E_{el}(R_{AB}) = Q^A Q^B / \epsilon R_{AB}$$

Donde R_{AB} es la distancia entre los átomos A y B, Q^A y Q^B las cargas de los átomos A y B, y ϵ la constante dieléctrica.

Las interacciones electrostáticas se han modelado tradicionalmente usando un modelo de cargas puntuales centradas en los átomos (cargas parciales) para representar la densidad de carga molecular. Los métodos más populares para extraer las cargas de la función de onda molecular están basados en el ajuste de las cargas atómicas al potencial electrostático molecular (MEP), calculado éste con métodos ab initio, o funciones de onda semiempíricas. El procedimiento de ajuste de las cargas consiste en minimizar la desviación al cuadrado entre el potencial producido por las cargas atómicas y el MEP (61). Estos potenciales no enlazantes son expresados como una suma de potenciales átomo-átomo simétricamente esféricos.

Se sabe que esta descripción representa una fuente importante de errores en los campos de fuerza actuales, sobre todo porque los monopolos o cargas parciales pueden variar enormemente con los cambios conformacionales. De hecho, una representación molecular realista necesita momentos dipolares (para incluir los pares no compartidos), y momentos cuadrupolares (para incluir los enlaces π). Una adecuada elección y modelización de las cargas parciales permite expresar los enlaces de hidrógeno de forma correcta (60)

Al asignar las cargas parciales, normalmente se obtienen cargas demasiado grandes para los átomos que no estén en la superficie de la molécula. Este problema se puede solucionar en parte añadiendo un término de penalización para las cargas parciales que no valgan cero, que asegure que sólo las cargas importantes mantienen un valor significativamente diferente a cero. Este ajuste restringido del potencial electrostático (RESP) se ha utilizado, por ejemplo, en Amber. También puede solucionarse representando la densidad de carga de forma explícita con una función continua, para proporcionar una descripción más exacta.

La transferencia de carga (CT) se refiere a la energía de interacción que surge con la deslocalización de los electrones de una molécula en los orbitales de otra molécula (60). En los sistemas de gran tamaño hay presentes múltiples cargas que interaccionan entre sí. Pero el cálculo de todas las posibles interacciones teóricas tendría demasiado coste computacional, por tanto a partir de cierta distancia las interacciones electrostáticas no se calculan, asumiendo el error derivado. La CT y las dificultades de cálculo de múltiples interacciones electrostáticas llevan a subestimar las cargas calculadas. Esto se compensa aumentando las cargas calculadas en un 10-15% y/o ajustando con datos experimentales (58).

La polarización se refiere a la redistribución de la densidad electrónica de una molécula, en presencia de un campo electromagnético externo. En biomoléculas, esto surge de forma natural de la distribución de carga de otras moléculas o átomos. La polarización genera interacciones no aditivas, atractivas, intra o intermoleculares, que pueden ser bastante significativas. En principio, la polarización puede ser descrita enteramente en términos de electrostática clásica, y se debe distinguir de la interacción dispersiva que surge de fluctuaciones instantáneas de la distribución de carga molecular que, en origen, es enteramente cuántica.

Los efectos de la polarización también deben ser incluidos en las simulaciones de dinámica molecular. Normalmente se evita el cálculo explícito de todos los efectos incluyendo una media de la polarización en forma de función del potencial de interacción (60).

2.2.3.2 Contribución de van der Waals

Es la energía que describe la repulsión o atracción entre átomos no directamente unidos no debida a las cargas de estos átomos. Puede ser interpretada como la parte no polar de la interacción entre átomos no enlazados. Se suele representar con una función de Lennard-Jones:

$$E_{\text{vdW}}(R_{\text{AB}}) = (C_1/R_{\text{AB}}^{12}) - (C_2/R_{\text{AB}}^6)$$

Donde R_{AB} es el radio de van der Waals, y C_1 y C_2 son constantes (ajustables de forma experimental). E_{vdW} es cero para distancias interatómicas largas, y es muy repulsiva en

distancias cortas. En términos cuánticos, esto último es debido al solapamiento de las nubes electrónicas de los dos átomos, ya que los electrones (de carga opuesta) se repelen entre sí. En distancias intermedias, hay una pequeña atracción de dipolo inducido entre estas nubes electrónicas.

Para moléculas grandes, el cálculo de la energía no enlazante es el cálculo que consume más capacidad computacional. La energía de van der Waals se calcula entre pares de átomos (pairwise), pero se corrige con datos experimentales.

El parámetro del radio de van der Waals de la ecuación de Lennard-Jones se calcula descomponiéndose como una suma de radios atómicos. Los átomos se asumen esféricos, pero esto puede suponer un problema en dos casos: Uno, con átomos de hidrógeno, ya que su distribución electrónica dista de ser esférica. El otro caso es con átomos con pares de electrones sin compartir (oxígeno, nitrógeno...). Es más difícil que los átomos (especialmente los más pequeños) con pares de electrones sin compartir sean esféricos. Sin embargo, los hidrógenos y los átomos con pares sin compartir son precisamente los átomos involucrados en enlaces de hidrógeno, y la tendencia es de representar los enlaces de hidrógeno sólo por interacciones electrostáticas (58).

2.2.4 Otros

También es una práctica habitual estudiar una parte de un sistema grande con un metodo superior de cálculo (por ejemplo, el centro activo de una enzima con métodos *ab initio*, y el resto de la enzima con métodos inferiores, por ejemplo QM/MM o mecánica molecular), todo esto para evitar un coste computacional excesivamente alto.

Aparte de la energía y los enlaces de hidrógeno, también se calcula la desviación cuadrática media (RMSD, Root-Mean-Square Deviation) de las posiciones atómicas, y el factor de desviación cuadrática media (RMSF, Root-Mean-Square factor). La RMSD mide la distancia media entre átomos de estructuras superpuestas de proteínas a lo largo de la simulación, y la RMSF mide la desviación entre la posición de una partícula y una posición de referencia. En este TFG, se comparan las coordenadas iniciales de la proteína con las coordenadas finales, con ello se sabe cuánto se han movido los átomos estudiados durante el plegamiento.

La hipótesis ergódica dice que todos los microestados tienen la misma probabilidad de ser accedidos cuando el tiempo tiende a infinito. Esto significa que el espacio conformacional sólo se explorará en su totalidad con un tiempo muy prolongado. Teniendo en cuenta esta hipótesis, se admite que existirán errores cuanto el tiempo de cálculo no sea lo suficientemente grande. Es una hipótesis estadística, y se supone que se ha cumplido si al final de la simulación el parámetro (por ejemplo, energía libre o RMSD) es estable. Esto no siempre es así, y es relativamente fácil caer en errores por varias cuestiones, como lo pequeña que es la barrera de energía entre las conformaciones de mínima energía de macromoléculas.

3 MATERIALES Y MÉTODO

3.1 Materiales

3.1.1 Ordenador de sobremesa Pentium Dual-Core CPU E5300 @2.60GHz x 2, memoria RAM 2.0 GB, tarjeta gráfica Gallium 0.4 sobre AMD RV620. Volumen de disco 294 GB. Sistema operativo Ubuntu 14.04 LTS de 32 bits (64).

3.1.2 Suite de programas de simulación molecular Amber12, que incluyen código fuente y demos (26). Campo de fuerzas Amber ff12SB (65). Programas de apoyo, no incluidos en Amber12: Chimera (66), VMD (67) y Xmgrace.

3.1.3 Fichero .pdb del complejo BCL6^{BTB}-SMRT^{BBD}, con código 1R2B (7,68).

3.2 Método

El proceso de computación es el siguiente:

Construcción de los complejos--> Minimización--> Preparación del sistema--> Producción --> PTRAJ y CPPTRAJ--> MMPBSA.py

3.2.1 Construcción de los complejos

Se prepara un archivo .top (de topología) de BCL6. Éste archivo define los parámetros del campo de fuerzas ensayado para cada uno de los átomos del sistema. Se prepara a partir del archivo .pdb, con el programa LeaP.

Son necesarias ciertas modificaciones sobre el fichero .pdb de cara a la correcta asignación de los parámetros del campo de fuerzas ff12SB (4). Primero, en el dominio D faltaban dos aminoácidos (ALA5 y GLU129), que tuvieron que ser añadidos utilizando Chimera, y copiando de forma manual, los residuos del dominio C al D sobre el fichero pdb. Además, había varias mutaciones experimentales en SMRT, necesarias para solubilizar la proteína. Estas mutaciones se revirtieron eliminando los átomos de la sidechain (cadena lateral) de los aminoácidos, y cambiando los átomos para pertenecer al backbone (esqueleto) de los originales (LEU1, VAL2, ALA3, VAL5 y ALA8). Se utiliza LeaP para añadir los átomos que falten, y Reduce para añadir los hidrógenos ausentes, orientar y asignar el estado de las cadenas laterales.

Para MMPBSA.py (63), son necesarios tres archivos de formato PRM (otro tipo de archivo de topología): complejo ($BCL6^{BTB} + SMRT$), receptor ($BCL6^{BTB}$) y ligando (SMRT). Estos se preparan a partir del fichero .top de BCL6, con la utilidad ante-MMPBSA.py.

3.2.2 Minimización

El paso de minimización energética es necesario para eliminar posible estrés conformacional de la proteína. Se aplica un protocolo de minimización en varias etapas ya ensayado (4). Para ejecutar la minimización (y cualquier proceso en Amber) son necesarios: 1) Fichero de topología; 2) Fichero con las coordenadas de todos los átomos; 3) Fichero de entrada (input), que incluye las órdenes que tiene que ejecutar el programa. La minimización se realiza en 5 etapas de 5000 ciclos (maxcyc=5000) cada una. El comando utilizado es imin=1, y se indica ntmin=1 y ncyc=5000 para utilizar el algoritmo de máxima pendiente durante toda la minimización, que constará de 5000 ciclos (a no ser que se alcance el objetivo de gradiente establecido anteriormente).

En la primera etapa, se mantienen fijos los átomos de proteína y receptor, y se minimiza el disolvente y contraiones. En la segunda, tercera y cuarta etapa se mantienen fijos los átomos del backbone de los aminoácidos, con restricción energética de 10, 1 y 0.1 kcal/mol en cada sucesiva etapa. En la quinta etapa, se permite el movimiento de todos los átomos.

3.2.3 Preparación del sistema

Las etapas de preparación son necesarias para que el sistema se encuentre en equilibrio. También podría hacerse una simulación sin que el sistema se encuentre en equilibrio, aunque es menos común. La frecuencia de colisión es de 5 colisiones por ps, establecida por el comando `gamma_In`. Durante los pasos de preparación y de producción, se utiliza el algoritmo de restricción SHAKE, con el comando `ntc=2`, para que los enlaces que impliquen hidrógeno se encuentren siempre en distancias de equilibrio.

La utilización de SHAKE elimina la libertad de movimiento de estiramiento de los enlaces (estiramiento es el tipo de movimiento más veloz), y por lo tanto permite que la integración se haga en pasos de más tiempo (2 fs en vez de 1 fs o menos) (26).

3.2.3.1 Calentamiento

Aumento gradual de la temperatura hasta el valor deseado, con el comando `ntt=3`. Es necesario hacerlo de forma gradual, para evitar que el cambio súbito de temperatura destruya el sistema al adquirir partes del mismo velocidades extremadamente altas. Se utiliza el comando `ntb=1`, para que el volumen sea constante. En ésta simulación, la etapa de calentamiento dura 200 ps, con un aumento de temperatura de 1.5 K/s.

3.2.3.2 Equilibrio

Al alcanzar la temperatura deseada (300 K), se mantiene la presión constante (comando `ntb=2`), y se pasa a aumentar la densidad del sistema, hasta que se encuentre en equilibrio. Éste proceso ocupa, aproximadamente, 200 ps más.

3.2.4 Producción

Tras estas etapas, el sistema está preparado para que se pueda producir la simulación de dinámica molecular. Se utiliza volumen constante (`ntb=1`) y sin restricciones para la presión. El comando `iwrap=1` hace que las coordenadas mostradas en los archivos estén sujetas en una “caja primaria”. Esto significa que el programa se queda con la imagen periódica del átomo que se encuentre más cercana a la mitad de la “caja primaria”. Es recomendable utilizar esta opción, para evitar que se produzca un exceso de difusión a alta

temperatura y las coordenadas del resultado sobrepasen los formatos de archivo (de la trayectoria, por ejemplo). Si no se utiliza `iwrap=1`, puede usarse `ptraj` para reubicar las moléculas en la “caja primaria” (26,59).

Durante la simulación, con el comando `dt=0.002` se indica que el timestep (unidad mínima de paso temporal) es de 0.002 ps (2fs), y con el comando `ntwx=5000` se indica que cada 5000 timestep se escribe el estado del sistema en el archivo `mdcrd` (de la trayectoria). Como $0.002 \times 5000 = 10$ ps, cada 10 ps se va a reflejar el estado del sistema en la trayectoria. A ésta descripción del estado instantáneo del sistema se le llama 'snapshot' (foto instantánea). Como 5 ns son 5000 ps, y se toma una snapshot cada 10 ps, quedan 500 unidades de tiempo con las cuales se constituye la trayectoria, o 'frames'. De éstas frames se obtendrá el estado del sistema, y por tanto todas las variables (ΔG , RMSD...).

3.2.5 PTRAJ y CPPTRAJ

Se ejecutó con el programa PTRAJ un script cuyo resultado fue un estudio temporal de los enlaces de hidrógeno que se establecen en la proteína, tanto intra como intermoleculares. Sólo se toman en consideración los que ocurren con una distancia máxima de 3.5 Å, una pervivencia mínima del 15% de la trayectoria, y un ángulo de enlace de entre 180° y 120° (por tanto, el rango total es 240°). Entre estos enlaces, se seleccionaron los más importantes, y los enlaces de hidrógeno dinámicos. Estos son los enlaces que se establecen durante toda o gran parte de la trayectoria, pero conforme la proteína se va plegando van intercambiando el átomo con el cual se establecen, normalmente dentro de un mismo par de residuos de aminoácido. La tabla 5 muestra algunos ejemplos de enlaces de hidrógeno dinámicos.

Con CPPTRAJ, se averiguó la distancia de estos enlaces de hidrógeno en cada foto de la trayectoria, con el comando `distance`, seguido por los átomos a calcular, por ejemplo “`distance d3 :9@O :13@H`”. Para esto, es necesario centrar la trayectoria respecto a las condiciones que hemos impuesto (72). Esto se indica con el comando `autoimage` (también pueden usarse otros como `image`, `origin`, `center`). También con CPPTRAJ se calculó la RMSD total frente al tiempo, y el RMSF.

3.2.6 MMPBSA.py

3.2.6.1 Ya con MMPBSA.py, se averigua el ΔG por frame, con sus diferentes contribuciones. En el cálculo, se utilizan el método PB y el método GB. Las 500 frames se leen de una en una, y se utiliza la aproximación de trayectoria única (sección 2.2.3).

Para PB, el valor de la tensión superficial es 0.00542 kcal/mol Å². Los valores de la constante dieléctrica externa e interna son, respectivamente, 80 y 1. Los valores del radio utilizados serán los mismos que los del archivo .top. El valor de la constante es 0.92 kcal/mol. Los parámetros 'scale', 'fillratio' e 'istrng' se mantuvieron en su valor por defecto (28).

Para GB, el valor de la tensión superficial es 0.0072 kcal/mol Å². Se utilizó el método de Onufriev (69). Los demás parámetros se mantuvieron en su valor por defecto (28). Para el cálculo por residuo se utilizan los mismos valores.

3.2.6.2 También con MMPBSA.py, se averigua la contribución al ΔG total que efectúa cada uno de los residuos de aminoácido con el método GB. El método PB no permite efectuar descomposición por residuo. Esto ocurre por motivos inherentes al método PB: mientras que en el método GB los términos energéticos pueden ser descompuestos por medio del radio de los átomos, en el método PB esto no ocurre. La descomposición por residuo se hizo con la energía electrostática 1-4 añadida a la EEL total, y la van der Waals 1-4 añadida a los términos VDWAALS potenciales. Sin embargo, ya que usamos una trayectoria única, estos términos se calculan de forma que se cancelan entre sí al sumar complejo, receptor y ligando.

4 RESULTADOS

4.1 Enlaces de hidrógeno

4.1.1 Ocupación

Para el total de 500 frames, resultaron 446 donantes (donante significa base de Lewis, es decir dador de electrones) y 494 aceptores (aceptor significa ácido de Lewis, es decir aceptor de electrones). 'Específico' significa que participa el sidechain del aminoácido. 'No específico' significa que participa el backbone del aminoácido, por tanto si cambiamos este residuo por otro residuo el enlace es propenso a establecerse de todas formas. Teniendo

en consideración sólo los enlaces intermoleculares con una ocupación mayor al 15%, tenemos 303 enlaces de hidrógeno. En la siguiente tabla se muestran algunos datos de ocupación:

Porcentaje de ocupación	Número de enlaces	Porcentaje del total de enlaces
>15	303	100
>50	242	79.87
>75	189	62.38
>90	148	48.85

Tabla 1: Enlaces de hidrógeno totales

Porcentaje de ocupación	Número de enlaces	Porcentaje del total de enlaces de SMRT
>15	42	100
>50	24	57.14
>75	18	42.86
>90	17	40.48

Tabla 2: Enlaces de hidrógeno en los que participa SMRT.

	Donante de enlaces		Aceptor de enlaces
Residuo	No específico	Específico	No específico
ALA3A	1		1
VAL5A	1		1
GLU7A	1	2	1
GLY9A			2
ILE12A	1		
HIS13A	1		
GLU14A		2	
ILE15A	1		1
PRO16A	1		
ARG17A	1	2	
LEU1B			3

ALA3B	1		
VAL5B	1		
GLU7B	1	8	
ILE12B	1		
HIS13B	1		
GLU14B		3	
ILE15B	1		
PRO16B	1		

Tabla 3: Descripción enlaces de hidrógeno en los que participan las cadenas de SMRT.

Residuo	Tipo de residuo	Enlaza con	Tipo de residuo
ALA3A	Hidrofóbico	CYS129	Polar, sin carga
ALA3B	Hidrofóbico	CYS4	Polar, sin carga
VAL5A	Hidrofóbico	GLN131	Polar, sin carga
VAL5B	Hidrofóbico	GLN6	Polar, sin carga
GLU7A	Ácido	THR133	Polar, sin carga
GLU7B	Ácido	THR8	Polar, sin carga
ARG10A	Básico	ILE12A	Hidrofóbico
ARG10B	Básico	ILE12B	Hidrofóbico
SER11A	Polar, sin carga	GLN135	Polar, sin carga
SER11B	Polar, sin carga	ASP10	Ácido
ILE12A	Hidrofóbico	ARG10A	Básico
ILE12B	Hidrofóbico	ARG10B	Básico
HIS13A	Básico	ASN142	Polar, sin carga
HIS13B	Básico	ASN17	Polar, sin carga
GLU14A	Ácido	ARG145	Básico
GLU14B	Ácido	ARG20	Básico
ILE15A	Hidrofóbico	ARG145	Básico
ILE15B	Hidrofóbico	ARG20	Básico
PRO16A	Hidrofóbico	ARG149	Básico
PRO16B	Hidrofóbico	ARG24	Básico

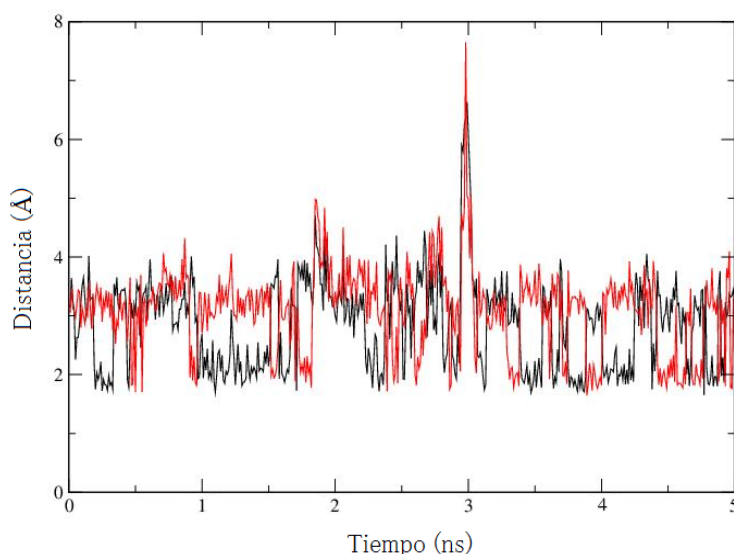
Tabla 4: Descripción enlaces de hidrógeno en cadenas A y B de SMRT

A continuación se presentan algunos datos de enlaces de hidrógeno dinámicos, y luego se muestra la representación de la distancia frente a tiempo en el apartado 4.1.2. Se eligieron porque son enlaces relacionados entre sí: cuando uno de ellos aumenta la distancia (es decir, pierde fuerza) el otro enlace la recupera, y viceversa.

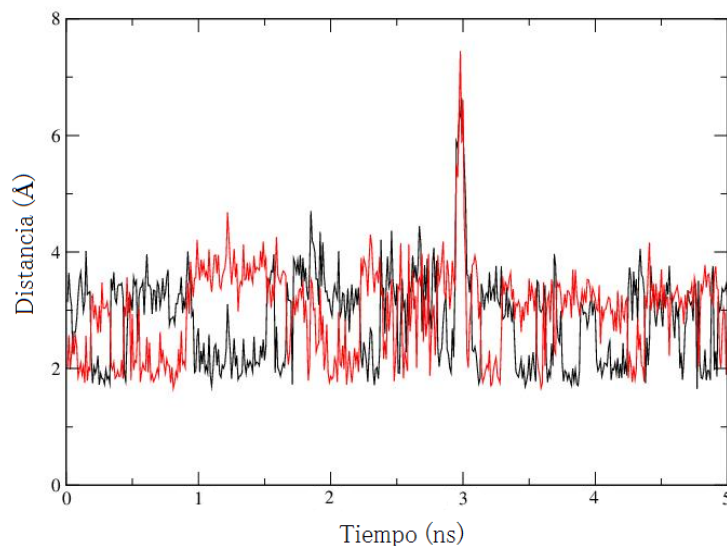
Enlace	Residuo donante	% de ocupación	Átomo donante	Átomo aceptor	Residuo aceptor
1	GLU7B	36.4	OE2	NHZ2	LYS244
2	"	36.2	OE1	NHZ2	"
3	"	32.0	OE1	NHZ3	"
4	"	29.4	OE2	NHZ3	"
5	"	20.8	OE2	NHZ1	"
6	"	20.0	OE1	NHZ1	"
	GLU14A	58.2	OE2	NHH12	ARG145
	"	37.6	OE1	NHH12	"

Tabla 5: Datos de enlaces dinámicos

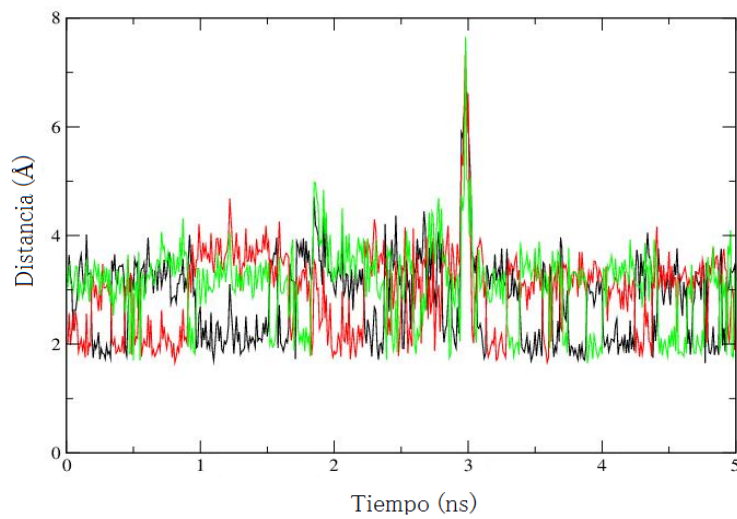
4.1.2 Distancia del enlace frente al tiempo



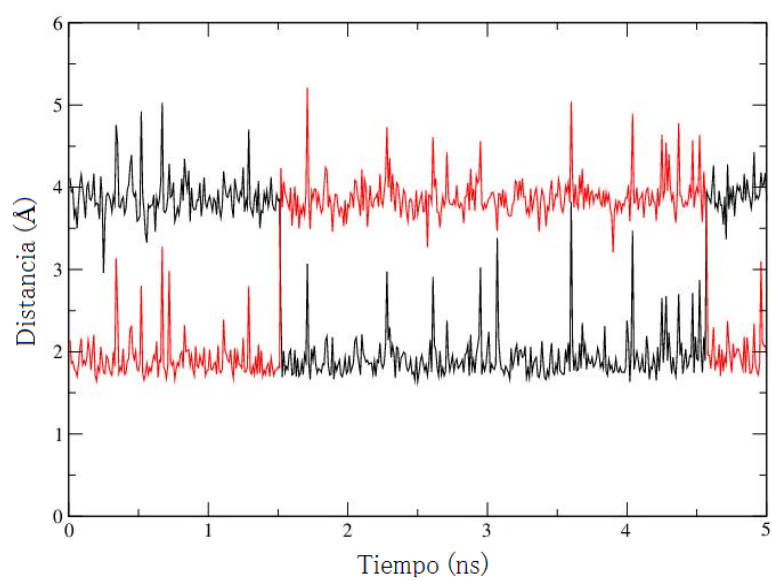
Gráfica 1: Enlaces de hidrógeno 1 (negro) y 4 (rojo) entre GLU7B y LYS244



Gráfica 2: Enlaces de hidrógeno 1 (negro) y 5 (rojo) entre GLU7B y LYS244

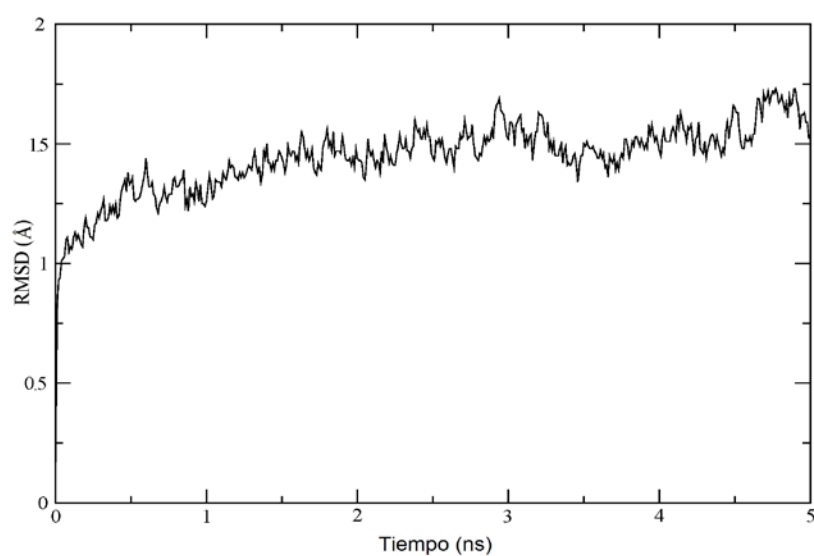


Gráfica 3: Enlaces de hidrógeno 1 (negro), 4 (rojo) y 5 (verde) entre GLU7B y LYS244

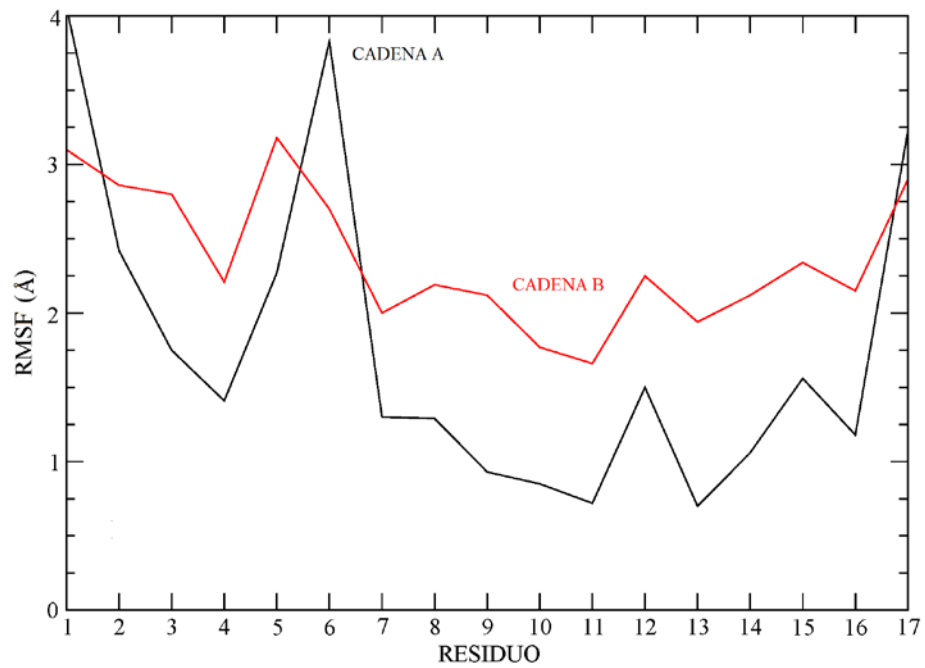


Gráfica 4: Enlaces de hidrógeno entre GLU14A y ARG145

4.2 RMSD y RMSF



Gráfica 5: RMSD durante la simulación

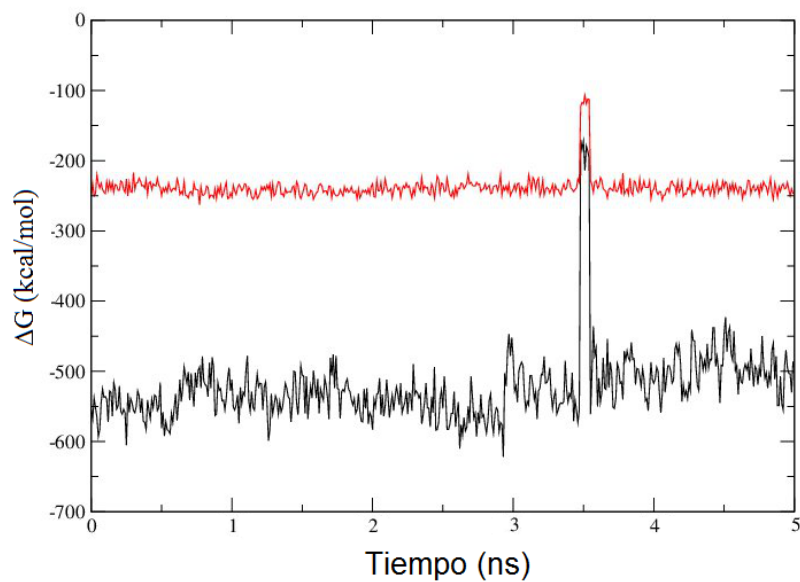


Gráfica 6: RMSF por residuo

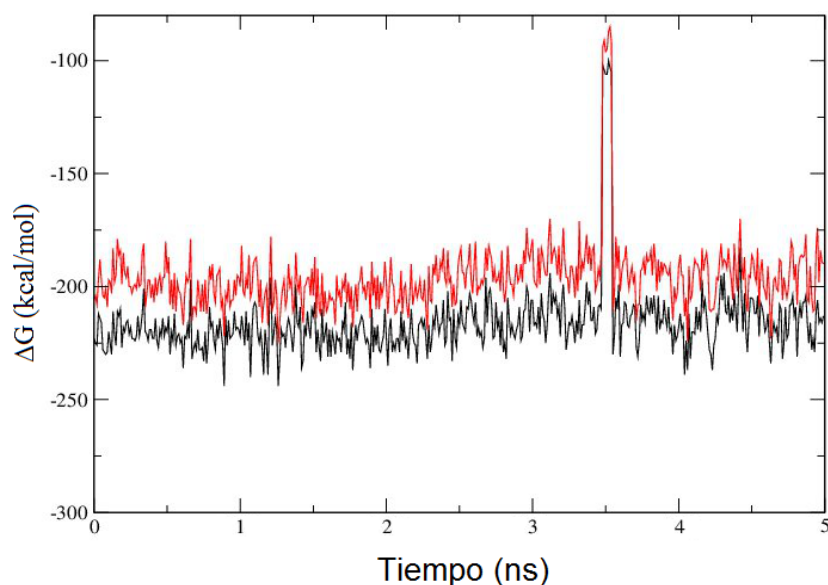
4.3 ΔG

En el apartado 4.3 todas las unidades son kcal/mol.

4.3.1 ΔG frente a tiempo



Gráfica 7: ΔG electrostática (negro) vs van der Waals (rojo) (método GB)



Gráfica 8: ΔG del método Poisson-Boltzmann (rojo) vs método Born generalizado (negro)

4.3.2 ΔG total

Pr significa que forma parte del cálculo por residuo y pf por frame.

Método	GB (pr)	GB (pf)	PB (pf)
VDWAALS	-2319.2115	-2319.2115	-2319.2115
EEL	-19081.9650	-19081.9650	-19081.9650
EGB	-3130.4948	-3129.6344	-3198.4286
ESURF	105.7843	105.7843	79.3396
G gas	-21401.1765	-21401.1765	-21401.1765
G solv	-3024.7105	-3023.8501	-3119.0890
TOTAL	-24425.8869	-24425.0266	-24520.2654

Tabla 5: Complejo (residuos 1-284)

Método	GB (pr)	GB (pf)	PB (pf)
VDWAALS	-1978.9839	-1978.9839	-1978.9839

EEL	-16666.3743	-16666.3743	-16666.3743
EGB	-2815.1108	-2814.3372	-2886.4798
ESURF	106.8213	106.8213	76.9343
G gas	-18645.3582	-18645.3582	-18645.3582
G solv	-2708.2895	-2707.5158	-2809.5455
TOTAL	-21353.6477	-21352.8741	-21454.9038

Tabla 6: Receptor (residuos 1-250)

Método	GB (pr)	GB (pf)	PB (pf)
VDWAALS	-101.4308	-101.4308	-101.4308
EEL	-1891.9696	-1891.9696	-1891.9696
EGB	-894.0270	-893.7813	-904.2642
ESURF	31.7070	31.7070	28.5310
G gas	-1993.4004	-1993.4004	-1993.4004
G solv	-862.3200	-862.0743	-875.7333
TOTAL	-2855.7204	-2855.4747	-2869.1336

Tabla 7: Ligando (residuos 251-284)

Método	GB (pr)	GB (pf)	PB (pf)
VDWAALS	-238.7968	-238.7968	-238.7968
EEL	-523.6211	-523.6211	-523.6211
EGB	578.6431	578.4841	592.3155
ESURF	-32.7440	-32.7440	-26.1256
G gas	-762.4179	-762.4179	-762.4179
G solv	545.8991	545.7401	566.1898
TOTAL	-216.5188	-216.6778	-196.2280

Tabla 8: Diferencia (complejo - receptor - ligando)

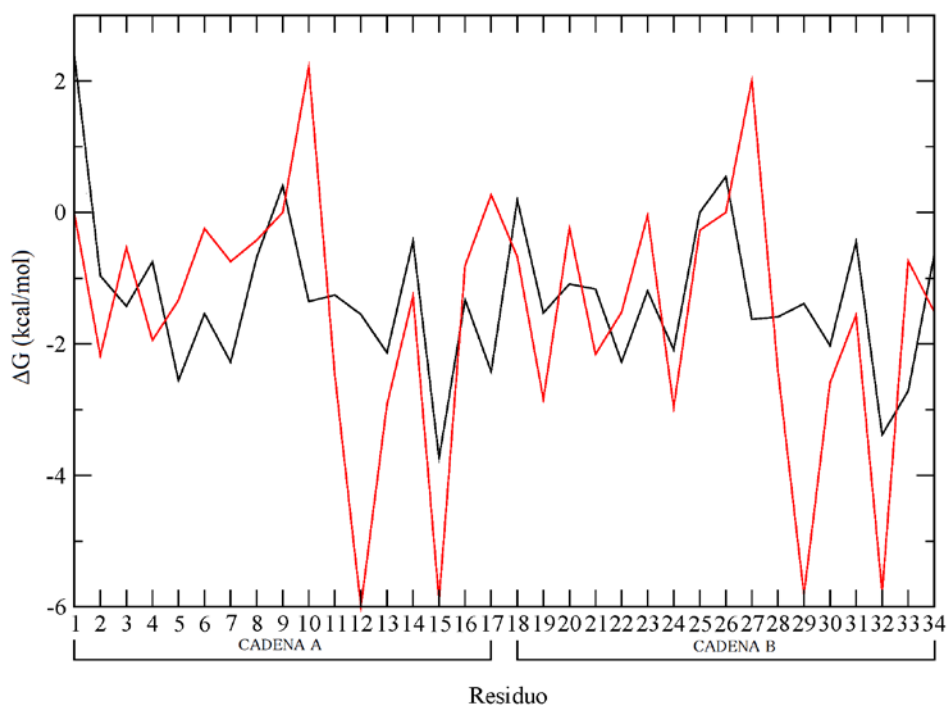
4.3.3 ΔG por residuo

Residuo	VDWAALS	EEL	ΔG_{solv}	TOTAL
---------	---------	-----	-------------------	-------

LEU1A	-0.6890	2.8344	0.2713	2.4167
VAL2A	-3.2293	-1.9993	2.1920	-3.1367
ALA3A	-1.7000	-2.7480	2.5882	-1.9598
THR4A	-3.5046	-1.8159	2.6249	-2.6957
VAL5A	-2.7053	-5.5251	4.3400	-3.8904
LYS6A	-2.6928	11.8356	-10.9290	-1.7862
GLU7A	-4.1813	-13.7733	14.9328	-3.0218
ALA8A	-1.7688	-2.4530	3.1172	-1.1045
GLY9A	-1.5065	-1.5109	3.4239	0.4064
ARG10A	-6.1788	1.8831	5.1588	0.8631
SER11A	-4.4682	-3.7739	4.5107	-3.7314
ILE12A	-7.8919	-2.7201	3.0615	-7.5506
HIS13A	-6.4474	-6.5433	7.9300	-5.0608
GLU14A	-1.5367	-26.4868	26.3149	-1.7087
ILE15A	-6.2317	-10.2250	6.8175	-9.6392
PRO16A	-2.2057	-4.2242	4.2960	-2.1339
ARG17A	-2.2445	-21.5707	21.6626	-2.1527
LEU1B	-0.5229	-47.8450	47.8777	-0.4902
VAL2B	-4.3913	-7.4176	7.4419	-4.3670
ALA3B	-1.6561	-1.4365	1.7663	-1.3263
THR4B	-2.7848	-2.5210	1.9805	-3.3253
VAL5B	-2.5143	-4.2639	2.9853	-3.7929
LYS6B	-2.7982	10.1750	-8.6185	-1.2417
GLU7B	-3.7070	-51.3633	50.0151	-5.0552
ALA8B	-1.2575	0.2178	0.7679	-0.2718
GLY9B	-1.7034	-0.2931	2.5418	0.5453
ARG10B	-6.3644	6.7838	-0.0342	0.3853
SER11B	-4.5761	-4.5111	5.0953	-3.9919
ILE12B	-7.5612	-1.9141	2.2741	-7.2012
HIS13B	-5.6497	-6.2036	7.2376	-4.6156
GLU14B	-1.6354	-34.8661	34.4941	-2.0074
ILE15B	-6.4914	-8.7511	6.0898	-9.1527

PRO16B	-2.636	-4.8544	3.4655	-3.4525
ARG17B	-4.4389	-13.9298	16.2190	-2.1497

Tabla 9: Diferentes contribuciones al ΔG de binding y ΔG de binding de cada residuo de ambas cadenas de SMRT



Gráfica 9: ΔG sidechain (rojo) frente a ΔG backbone (negro)

5 DISCUSIÓN

En esta sección se indican los residuos que se han detectado más importantes. Como se comentó antes, sólo se tendrán en cuenta los residuos de SMRT. Esto es porque todo este proceso va orientado a explorar la naturaleza de las interacciones del dominio BTB de BCL6 con dos moléculas del correpressor SMRT, para así encontrar un ligando equivalente a éste, que pueda devolver a BCL6 a niveles normales de actividad.

5.1 Enlaces de hidrógeno

Aquí se comentan los tipos de residuos, la ocurrencia de enlaces, y la distribución. Varios de los enlaces específicos son, además, enlaces dinámicos. Esto significa que los átomos implicados no se mantienen durante toda la trayectoria, los átomos van cambiando pero siempre son del mismo residuo. Esto ya se mostró en la tabla 5 y gráfica 1 a 4.

Los donantes son átomos de oxígeno de doble enlace. Algunos pertenecen al backbone, y otros al sidechain. Los enlaces de hidrógeno no específicos se establecen de igual manera en ambas cadenas de SMRT, como se muestra en la tabla 4.

Los donantes específicos son los residuos de GLU7A, GLU14A, GLU7B y GLU14B, especialmente el 7B, que participa en 8 enlaces de hidrógeno usando los dos oxígenos de su cadena lateral. Estos enlaces son entre un grupo carboxilato y (la mayoría) con la cadena lateral de la arginina. En condiciones fisiológicas ambos grupos se encuentran cargados, y la atracción electrostática hace que sean más fuertes que los enlaces no específicos.

Los aceptores son hidrógenos que están unidos a átomos de nitrógeno, tanto en el backbone de los aminoácidos como a la cadena lateral de arginina, pero también a un grupo hidroxilo. Los enlaces de hidrógeno más significativos son entre la ARG10A y ARG10B e ILE12A y ILE12B, y los que se dan entre la SER11A y SER11B y la cadena lateral de dos residuos de HIS (135 y 10, respectivamente). Otra observación notable es que las interacciones de SMRT se producen con una zona bastante localizada del receptor, desde ASP127 a ARG149.

5.2 RMSD y RMSF

La RMSD es una medida de la distancia entre los átomos, de cuánto se han movido a lo largo del tiempo. La gráfica 5 muestra la RMSD de la trayectoria completa. Esta RMSD se va estabilizando con el tiempo, esto es un indicador de que la trayectoria converge.

La gráfica 6 muestra el RMSF por residuo de ambas cadenas de SMRT. En la gráfica, sobresalen 5 residuos, con un RMSF aproximadamente mayor de 3. Estos son LEU1A, LYS6A, ARG17A, LEU1B y VAL5B. Esto nos indica que son los residuos que más se han desplazado respecto a su posición inicial.

5.3 ΔG

5.3.1 ΔG frente a tiempo

Normalmente, la ΔG por frame se utiliza para comprobar si la dinámica molecular ha convergido o no. En la gráfica 8, se muestra una comparación entre la ΔG calculada por el método PB y por el método GB. En esa gráfica se aprecia que el valor de ΔG permanece estable con el tiempo, y con esto se podría decir que la trayectoria ha convergido.

5.3.2 ΔG total

El cálculo que sale más cercano al experimental es el mayor, del método PB, -196.228 kcal/mol. Es, sin embargo, bastante diferente al experimental, de -6.78 kcal/mol (7). Esto es por limitaciones inherentes al método. Se asume que el valor absoluto (el experimental) de ΔG no nos interesa, sino el relativo con fines comparativos: $\Delta G = \Delta G_{\text{complejo}} - \Delta G_{\text{receptor}} - \Delta G_{\text{ligando}}$. Al efectuarse la diferencia de ΔG , el término entrópico se cancela.

5.3.3 ΔG por residuo

En la tabla 9 se observan los valores de la contribución de van der Waals, electrostática y total a la energía de interacción, por cada residuo. Por parte de van der Waals, los residuos que se encuentran más interesantes son VAL2, THR4, GLU7, ARG10, SER11, ILE12, HIS13, GLU14 e ILE15 (de ambas cadenas).

Los residuos que incluyen enlaces de hidrógeno numerosos o notables son GLU7, ARG10, SER11, GLU14 e ILE15 (de ambas cadenas).

En el sentido de la contribución electrostática, VAL5, GLU7, SER11, HIS13, GLU14, ILE15, PRO16, ARG17 (de ambas cadenas) y LEU1B y VAL2B presentan un valor atractivo de magnitud importante. LYS6 y ARG10 presentan un valor repulsivo alto.

Para el total, VAL5, SER11, ILE12, HIS13, ILE15 presentan un valor alto.

Enlaces H	Vdwaals	Eel	Total
	VAL2		

	THR4		
VAL5		VAL5	VAL5
GLU7	GLU7	GLU7	
ARG10	ARG10		
SER11	SER11	SER11	SER11
	ILE12		ILE12
	HIS13	HIS13	HIS13
GLU14	GLU14	GLU14	
ILE15	ILE15	ILE15	ILE15
		PRO16	
		ARG17	

Tabla 10: Residuos considerados más importantes en SMRT, según el parámetro considerado

Teniendo estos datos en cuenta, se puede apreciar la importancia relativa del sidechain y el backbone. Puede verse que en aminoácidos apolares la importancia es similar, o algo mayor el peso del backbone, mientras que en ácidos y básicos el sidechain cobra mayor importancia.

5.4 Conclusión

Teniendo los datos anteriores en cuenta, se puede concluir que los residuos que más importancia tienen en el proceso de interacción entre BCL6 y SMRT son VAL5, GLU7, SER11, HIS13, GLU14 e ILE15.

Estos resultados concuerdan, en parte o en su totalidad, con los de la bibliografía utilizada. Estos artículos establecen que la zona clave de interacción es desde GLY9A hasta ARG17A (7,24), otro que los residuos importantes son desde GLU14A hasta ARG17A (25), o THR4A, VAL5A, GLU7A, HIS13A e ILE15A (4).

6 BIBLIOGRAFÍA

- 1: Margarita Sánchez-Beato, Abel Sánchez-Aguilera, Miguel A. Piris, Cell cycle deregulation in B-cell lymphomas, *Blood Journal* 101(4) (2003) 1220-1235.
- 2: Daniela Capello, Gianluca Gaidano, Molecular pathophysiology of indolent lymphoma, *Haematologica* 85 (2000) 195-201.
- 3: <http://www.ncbi.nlm.nih.gov/protein/?term=bcl6>
- 4: J.M. Granadino-Roldán, C. Obiol-Pardo, M. Pinto, A. Garzón, J. Rubio-Martínez, Molecular dynamics analysis of the interaction between the human BCL6 BTB domain and its SMRT, NcoR and BCOR corepressors: The quest for a consensus dynamic pharmacophore, *Journal of Molecular Graphics and Modelling* 50 (2014) 142–151.
- 5: Susan Zollman, Dorothea Godt, Gilbert G. Privé, Jean-Louis Couderc, Frank A. Laski, The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in *Drosophila*, *Proceedings of the National Academy of Sciences of the United States of America* 91 (1994) 10717-10721.
- 6: Weimin Ci, Jose M Polo, Ari Melnick, B-cell lymphoma 6 and the molecular pathogenesis of diffuse large B-cell lymphoma, *Current Opinion in Hematology* 15(4) (2008) 381-390.
- 7: K.F. Ahmad, A. Melnick, S. Lax, D. Bouchard, J. Liu, C.L. Kiang, et al., Mechanism of SMRT corepressor recruitment by the BCL6 BTB domain, *Molec. Cell* 12 (2003) 1551-1564.
- 8: A.F. Ghetu, C.M. Corcoran, L. Cerchietti, V.J. Bardwell, A. Melnick, G.G. Prive, Structure of a BCOR corepressor peptide in complex with the BCL6 BTB domain dimer, *Molec Cell* 29 (2008) 384–391.
- 9: Khanh D. Huynh, Wolfgang Fischle, Eric Verdin, Vivian J. Bardwell, BCoR, a novel corepressor involved in BCL-6 repression, *Genes & Development* 14 (2000) 1810-1823.
- 10: Lourdes M. Mendez, Jose M. Polo, J. Jessica Yu, Melissa Krupski, B. Belinda Ding, Ari Melnick, et al., CtBP Is an Essential Corepressor for BCL6 Autoregulation, *Molecular and Cellular Biology* 28(7) (2008) 2175-2186.

- 11: Khanh Huynh, Vivian J. Bardwell, The BCL-6 POZ domain and other POZ domains interact with the co-repressors N-CoR and SMRT, *Oncogene* 17 (1998) 2473-2484.
- 12: S. R. Walker, S. Liu, M. Xiang, M. Nicolais, K. Hatzi, E. Giannopoulou, et al., The transcriptional modulator BCL6 as a molecular target for breast cancer therapy, *Oncogene* 34(9) (2015) 1073-1082.
- 13: Grant D. Barish, Ruth T. Yu, Malith S. Karunasiri, Diana Becerra, Jason Kim, Tiffany W. Tseng, et al., The Bcl6-SMRT/NCoR Cistrome Represses Inflammation to Attenuate Atherosclerosis, *Cell Metabolism* 15 (2012) 554-562.
- 14: Jose M. Polo, Tania Dell'Oso, Stella Maris Ranuncolo, Leandro Cerchietti, David Beck, Gustavo F. Da Silva, et al., Specific peptide interference reveals BCL6 transcriptional and oncogenic mechanisms in B-cell lymphoma cells, *Nat. Med.* 10(12) (2004) 1329-1335.
- 15: Leandro Cerchietti, Jose M. Polo, Gustavo F. Da Silva, et al., Sequential Transcription Factor Targeting for Diffuse Large B-Cell Lymphomas, *Cancer Res.* 68 (2008) 3361-3369.
- 16: Weimin Ci, Jose M. Polo, Leandro Cerchietti, Rita Shaknovich, Ling Wang, Shao Ning Yang, et al., The BCL6 transcriptional program features repression of multiple oncogenes in primary B cells and is deregulated in DLBCL, *Blood Journal* 113 (2009) 5536-5548.
- 17: Michael R. Green, Carolina Vicente-Dueñas, Isabel Romero-Camarero, Chih Long Liu, Bo Dai, Inés González-Herrero, et al., Transient expression of Bcl6 is sufficient for oncogenic function and induction of mature B-cell lymphoma, *Nature Comm.* 5, (2014).
- 18: Katerina Hatzi, Yanwen Jiang, Chuanxin Huang, Francine Garrett-Bakelman, Micah D. Gearhart, Eugenia G. Giannopoulou, et al., A Hybrid Mechanism of Action for BCL6 in B Cells Defined by Formation of Functionally Distinct Complexes at Enhancers and Promoters, *Cell Reports* 4 (2013) 578-588.
- 19: Katerina Hatzi, Ari Melnick, Breaking bad in the germinal center: how deregulation of BCL6 contributes to lymphomagenesis, *Trends in Molecular Med.* 20(6) (2014) 343-352.

- 20: Katia Basso, Riccardo Dalla-Favera, Roles of BCL6 in normal and transformed germinal center B cells, *Immunological Reviews* 247 (2012) 172-183.
- 21: Ribal Bassil, William Orent, Marta Olah, Ahmed T. Kurdi, Michael Frangieh, Thomas Buttrick, et al., BCL6 Controls Th9 Cell Development by Repressing *IL9* Transcription, *Journal Immunology* 193 (2014) 198-207.
- 22: J. Sans-Sabrafen, C. Besses Raebel, J. L. Vives Corrons, *Hematología clínica* (5^a edición (2006) Elsevier.
- 23: Samir Parekh, Gilbert Privé, Ari Melnick, Therapeutic targeting of the BCL6 oncogene for diffuse large B-cell lymphomas, *Leukemia & Lymphoma* 49(5) (2008) 874-882.
- 24: Leandro Cerchietti, Shao Ning Yang, Rita Shaknovich, Katerina Hatzi, Jose M. Polo, Amy Chadburn, et al., A peptomimetic inhibitor of BCL6 with potent antilymphoma effects in vitro and in vivo, *Blood Journal* 113(15) (2009) 3397-3405.
- 25: Leandro Cerchietti, Alexandru F. Ghetu, Xiao Zhu, Gustavo F. Da Silva, Shijun Zhong, Marilyn Matthews, et al., A small-molecule inhibitor of BCL6 kills DLBCL cells in vitro and in vivo, *Cancer Cell* 17 (2010) 400-411.
- 26: D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, et al., (2012), *Amber12*, University of California, San Francisco.
- 27: Xuan-Yu Meng, Hong-Xing Zhang, Mihaly Mezei, Meng Cui, Molecular Docking: A powerful approach for structure-based drug discovery, *Current Computer-Aided Drug Design* 7(2) (2011) 146–157.
- 28: Kitchen DB, Decornez H, Furr JR, Bajorath J, Docking and scoring in virtual screening for drug discovery: methods and applications, *Nat Rev Drug Discov* 3(11) (2004) 935–949.
- 29: Tingjun Hou, Junmei Wang, Youyong Li, Wei Wang, Assessing the performance of the MM/PBSA and MM/GBSA methods: I. The accuracy of binding free energy calculations based on molecular dynamics simulations, *Journal Chem. Inf. Model.* 51(1) (2011) 69-82.

- 30: Haijing Yu, Yu Fang, Xia Lu, Yongjuan Liu, Huabei Zhang, Combined 3D-QSAR, molecular docking, molecular dynamics simulation, and binding free energy calculation studies on the 5-hydroxy-2H-pyridazin-3-one derivatives as HCV NS5B polymerase inhibitors, *Chemical Biology & Drug Design* 83(1) (2014) 89-105.
- 31: Fangfang Wang, Wei Yang, Yong. Shi, Guowei Le, 3D-QSAR, molecular docking & molecular dynamics studies of a series of ROR γ t inhibitors, *J Biom Struct Dyn* (2014) 1-12.
- 32: Vahid Jahed, Ali Zarrabi, Abdol-khalegh Bordbar, Mohammad Sadegh Hafezi, NMR (^1H , ROESY) spectroscopic and molecular modelling investigations of supramolecular complex of β -cyclodextrin and curcumin, *Food Chemistry* 165 (2014) 241-246.
- 33: Branimir Bertoša, Goran Mikleušević, Beata Wielgus-Kutrowska, Marta Narczyk, Matea Hajnić, Ivana Lešćić Ašler, et al., Homooligomerization is needed for stability: a molecular modelling and solution study of Escherichia coli purine nucleoside phosphorylase, *FEBS Journal* 281(7) (2014) 1860-1871.
- 34: Yun-Jie Miao, Xiao-Fang Xu, Fei Xu, Yong Chen, Jian-Wei Chen, Xiang Li, The structure-activity relationships of mono-THF ACGs on mitochondrial complex I with a molecular modelling study, *Natural Products Research* 28(21) (2014) 1929-1935.
- 35: Shannon E. Telesco, Rajanikanth Vadigepalli, Ravi Radhakrishnan, Molecular Modeling of the ErbB4/HER4 Kinase in the Context of the HER4 Signaling Network Helps Rationalize the Effects of Clinically Identified HER4 Somatic Mutations on the Cell Phenotype, *Biotechnology Journal* 8(12) (2013) 1452-1464.
- 36: Nuno Martinho, Helena Florindo, Liana Silva, Steve Brocchini, Mire Zloh, Teresa Barata, Molecular Modeling to Study Dendrimers for Biomedical Applications, *Molecules* 19 (2014) 20424-20467.
- 37: <http://boinc.berkeley.edu/> (cita BOINC)
- 38: <http://casathome.ihep.ac.cn/> (cita CAS)
- 39: <http://www.gpugrid.net/> (cita GPUGRID)

- 40: <http://www.ibercivis.es/> (cita ibercivis)
- 41: <http://www.worldcommunitygrid.org/> (cita WCG)
- 42: <http://boinc.bakerlab.org/rosetta/> (cita rosetta)
- 43: Hung-Jin Huang, Hsin Wei Yu, Chien-Yu Chen, Chih-Ho Hsu, Hsin-Yi Chen, Kuei-Jen Lee, et al., Current developments of computer-aided drug design, Journal of the Taiwan Institute of Chemical Engineers 41 (2010) 623-625.
- 44: Bahareh Honarparvar, Thavendran Govender, Glenn E. M. Maguirre, Mahmoud E. S. Soliman, Hendrik G. Kruger, Integrated Approach to Structure-Based Enzymatic Drug Design, Chemical Reviews 114 (2014) 493-537.
- 45: Michael Hedvat, Luni Emdad, Swadesh K. Das, Keetae Kim, Santanu Dasgupta, Shibu Thomas, Selected Approaches for Rational Drug Design and High Throughput Screening to Identify Anti-Cancer Molecules, Anti-Cancer Agents in Med Chem 12(9) (2012) 1143-1155.
- 46: Yashumati Ratan Bhardwaj, Ashutosh Pareek, Vivek Jain, Dharma Kishore, Chemical delivery systems and soft drugs: Retrometabolic approaches of drug design, Saudi Pharmaceutical Journal 22 (2014) 290-302.
- 47: M. R. Wilkins, C. Pasquali, R. D. Appel, K. Ou, O. Golaz, J. C. Sanchez, et al., From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis, Biotechnology 14(1) (1996) 61-65.
- 48: N. Leigh Anderson, Norman G. Anderson, Proteome and proteomics: new technologies, new concepts and new words, Electrophoresis 19(11) (1998) 1853-1861.
- 49: Shanchun Guo, Jin Zou, Guangdi Wang, Advances in the proteomic discovery of novel therapeutic targets in cancer, Drug Design, Development and Therapy 7 (2013) 1259-1271.
- 50: Haruhiko Kamada, Proteomic analysis for the development of biopharmaceuticals, Yakugaku Zasshi Journal of the Pharmaceutical Society of Japan 133(9) (2013) 925-930.

- 51: OMS cáncer <http://www.who.int/mediacentre/factsheets/fs297/es/>
- 52: John Mendelsohn, Peter M. Howley, Mark A. Israel, Joe W. Gray, Craig B. Thompson, Molecular Basis of Cancer, 3ª edición (2008) Saunders.
- 53: José Ramón Regueiro Gonzalez, Carlos López Larrea, Segundo González Rodríguez, Eduardo Martínez Naves, Biología y patología del sistema inmunitario, 4ª edición (2011) Médica Panamericana.
- 54: Thomas J. Kindt, Richard A. Goldsby, Barbara A. Osborne, Inmunología de Kubi, 6ª Edición (2007) McGrawHill Interamericana.
- 55: Modesto Orozco, Theoretical view of protein dynamics, Chem Soc Rev 43 (2014) 5051.
- 56: <http://www.gaussian.com/index.htm>
- 57: Alan Hinchliffe, Modelling Molecular Structures, 2ª Edición (2000) John Wiley and Sons.
- 58: Frank Jensen, Introduction to computational chem., 2ª Ed. (2007) John Wiley & Sons.
- 59: D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, et al., (2012), AmberTools13, University of California, San Francisco.
- 60: G. Andrés Cisneros, Mikko Karttunen, Pengyu Ren, Celeste Sagui, Classical electrostatic for biomolecular simulations, Chemical Reviews 114 (2014) 779-814.
- 61: F. Fogolari, A. Brigo, H Molinari, The Poisson–Boltzmann Equation for Biomolecular Electrostatics: a Tool for Structural Biology, J. of Molec. Recognit. 15 (2002) 377-392.
- 62: Peter Atkins, Julio De Paula, Química Física, 8ª edición (2008) Médica Panamericana.
- 63: Bill R. Miller, T. Dwight McGee, Jason M. Swails, Nadine Homeyer, Holger Gohlke, Adrian E. Roitberg, MMPBSA.py: An Efficient Program for End-State Free Energy Calculations, Journal of Chemical Theory and Computation 8(9) (2012) 3314–3321.

64: <http://releases.ubuntu.com/14.04/>

65: Victor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, Carlos Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins: Struct., Funct. and Bioinf.* 65 (2006) 712–725.

66: Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, Thomas E. Ferrin, UCSF Chimera--a visualization system for exploratory research and analysis, *J. Comp. Chem.* 25(13) (2004) 1605-1612.

67: William Humphrey, Andrew Dalke, Klaus Schulten, VMD: Visual Molecular Dynamics, *Journal of Molecular Graphics* 14(1) (1996) 33-38.

68: Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T.N. Bhat, Helge Weissig, et al., The Protein Data Bank, *Nucleic Acids Research* 28(1) 235-242.

69: Alexey Onufriev, Donald Bashford, David A. Case, Exploring protein native states and large-scale conformational changes with a modified GB model, *Proteins: Struct., Funct. Bioinf.*, 55 (2) (2004) 383–394.