



Universidad de Jaén
Facultad de Ciencias Experimentales

TRABAJO FIN DE GRADO

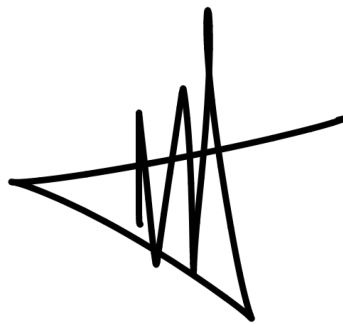
**ESTUDIO ASISTIDO
MEDIANTE SOFTWARE
DE LA INTERACCIÓN
ENTRE UN FÁRMACO
ANTICANCERÍGENO Y SU
DIANA BIOLÓGICA
(G6PDH)**

Alumno: Matilde Ansino Ortiz

Junio, 2020



ESTUDIO ASISTIDO MEDIANTE SOFTWARE DE LA INTERACCIÓN ENTRE UN FÁRMACO ANTICANCERÍGENO Y SU DIANA BIOLÓGICA (G6PDH)



Fdo. Matilde Ansino Ortiz

Jaén, junio 2020

ÍNDICE

1. RESUMEN DE LA PROPUESTA	7
2. ANTECEDENTES	9
2.1. La proteína G6PDH	9
2.1.1. <i>Estructura y función de la G6PDH</i>	11
2.1.2. <i>Deficiencia de la G6PDH</i>	13
2.2. Fundamento teórico	17
2.2.1. <i>Dinámica molecular</i>	17
2.2.2. <i>Campos de fuerza</i>	18
2.2.3. <i>Cálculo de la energía libre de unión</i>	20
3. OBJETIVOS DEL PROYECTO	23
4. METODOLOGÍA Y PLAN DE TRABAJO	25
4.1. Diseño de fármacos	25
4.2. Metodología	27
4.2.1. <i>Preparación de la caja del ligando solvatado (Tarea 1)</i>	28
4.2.2. <i>Preparación del homodímero G6PDH humano (Tarea 2)</i>	29
4.2.3. <i>Preparación del sistema (Tarea 3)</i>	29
4.2.4. <i>Producción o simulación de dinámica molecular (Tarea 4)</i>	31
4.2.5. <i>Estudio de las trayectorias reactivas (Tarea 5)</i>	32
5. BENEFICIOS DEL PROYECTO	35
5.1. Impacto científico-técnico social y/o económico	35
5.2. Plan de difusión	35
6. CRONOGRAMA DE TRABAJO	37
7. PRESUPUESTO	39
8. BIBLIOGRAFÍA	41

1. RESUMEN DE LA PROPUESTA

INVESTIGADOR PRINCIPAL: Matilde Ansino Ortiz

TÍTULO DEL PROYECTO: Estudio asistido mediante software de la interacción entre un fármaco anticancerígeno y su diana biológica (G6PDH)

RESUMEN

La glucosa-6-fosfato deshidrogenasa (G6PDH) es una enzima esencial implicada en la primera reacción de la rama oxidativa de la vía de las pentosas fosfato (PPP). La fase oxidativa de esta ruta proporciona los precursores para la biosíntesis de nucleótidos, lípidos y aminoácidos. En este caso, el proyecto propuesto se centra en la importancia del producto final, ribosa-5-fosfato, ya que éste es necesario para la síntesis de ácidos nucleicos y para la progresión del tumor. Dado que el producto sintetizado de la ruta PPP es dependiente de la enzima G6PDH, se sugirió dicha enzima como una nueva proteína diana para terapias contra el cáncer. La proteína objetivo se sobreexpresa en muchos tumores, produciendo un aumento de su actividad en una gran variedad de enfermedades cancerígenas. Esta relación tan estrecha con la enfermedad citada, convierte la proteína en una gran oportunidad de investigación. Hay dos posibles enfoques al estudio para inhibir su actividad, silenciando la expresión genética de la propia proteína, o bien, utilizando otros compuestos que inhiban su actividad enzimática.

En este proyecto se propone el estudio de un determinado compuesto orgánico del que experimentalmente se conoce su unión a la proteína, pero no se sabe en qué zona de la misma, ni la conformación adoptada por dicho compuesto al unirse. El estudio se realizaría por medio de simulaciones de dinámica molecular con el fin de obtener los posibles sitios de unión entre el ligando y la proteína, así como su modo de unión. Esto debería ayudar al diseño racional de inhibidores de G6PDH novedosos y potentes para fines terapéuticos

PROJECT TITLE: Software-assisted study of the interaction between an anticancer drug and its biological target (G6PDH)

SUMMARY

Glucose-6-phosphate dehydrogenase (G6PDH) is an essential enzyme involved in the first oxidative branch reaction of the pentose phosphate pathway (PPP). The oxidative phase of this route provides a precursor to the nucleotides, lipids and amino acids biosynthesis. In this case, the proposed project is focussed on the importance of the final product, ribose-5-phosphate, since it is necessary for the nucleic acid synthesis and tumour progression. Given that products synthesized by the PPP pathway are dependent on G6PDH, it has been suggested as a new target protein for anti-cancer therapies. The target protein is over expressed in many tumours, producing an increase in its activity in a great variety of cancerous diseases. The close relationship with the named disease, converts the protein into a great opportunity of investigation. There are two possible study focuses to inhibit its activity, silence the genetic expression of the own protein, better yet, utilizing other compounds that inhibit enzyme activity.

In this project, a proposed study of an organic compound that is experimentally known for binding to the protein, but not in which zone of the protein. The study will use molecular dynamic simulations, in order to obtain the possible binding sites between the protein and the ligand as well as it's form of binding. This should help the rational design of novel and potent inhibitors of G6PDH for therapeutic uses.

2. ANTECEDENTES

2.1. La proteína G6PDH

La glucosa-6-fosfato deshidrogenasa (G6PDH) es una enzima esencial citoplasmática que está presente en todas las células del organismo de los seres vivos, tanto animales como plantas. La función principal de la G6PDH humana es la catálisis en la primera reacción de la rama oxidativa de la ruta de las pentosas fosfato (PPP), vía de degradación del glucógeno que cumple con la finalidad de producir el coenzima NADPH (nicotinamida adenina dinucleótido fosfato), útil para donar electrones en rutas anabólicas, y ribosa 5-fosfato, metabolito esencial para la síntesis de nucleótidos y ácidos nucleicos.

La ruta se desarrolla en el citoplasma y se puede distinguir en dos fases: oxidativa y no oxidativa. En la fase oxidativa, una molécula de glucosa-6-fosfato (G6P) se transforma en ribulosa-5-fosfato (R5P), liberándose CO_2 y rindiendo dos coenzimas NADP reducidos a NADPH. En la fase no oxidativa, a partir de tres pentosas fosfato se forman dos fructosas-6-fosfato y un gliceraldehído-3-fosfato.

La función de la enzima G6PDH en la ruta se da en la primera etapa de la fase oxidativa donde cataliza la oxidación del carbono-1 de la G6P, formándose 6-fosfoglucono-1-lactona, y además, reduce el NADP A NADPH, interviniendo éste en un número importante de reacciones enzimáticas, como en la síntesis de glutatión reducido, ayudando así a la protección de las células ante el daño oxidativo. Esta enzima también interviene en procesos de hidrólisis de los ácidos nucleicos de la dieta, degradando glucosa o pentosas de los nucleótidos hasta formar agua y CO_2 .

En la fase oxidativa de la ruta de las pentosas fosfato se genera NADPH y R5P a partir de la glucosa-6-fosfato. Esta etapa consta de tres reacciones consecutivas:

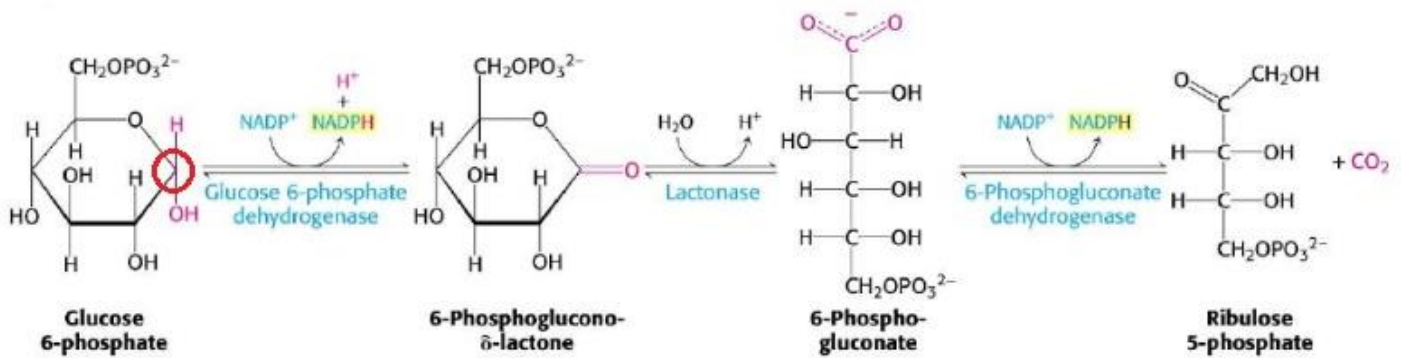


Ilustración 1. Etapa oxidativa de la ruta de las pentosas fosfato

1. Deshidrogenación de la G6P en el carbono 1 mediante la intervención de la enzima G6PDH. En esta primera reacción se obtiene 6-fosfogluconolactona y NADPH.
2. La enzima glucolactonasa cataliza la hidrólisis de la 6-fosfogluconolactona en 6-fosfogluconato (6FG).
3. Descarboxilación oxidativa de la 6FG produciendo R5P, catalizada por la 6-fosfogluconato desahidrogenasa, con la consecuente reducción de NADPH. En el paso final ocurre la isomerización de la R5P por la fosfopentosa isomerasa.

La fase no oxidativa es un poco más compleja que la oxidativa. La fosfopentosa isomerasa cataliza la isomerización de la R5P (cetosa) en R5P (aldosa). Simultáneamente, la fosfopentosa epimerasa sintetiza la xilulosa-5-fosfato (X5P) a partir de la R5P.

A partir de una molécula de R5P y dos de X5P transcurren una serie de reacciones de transferencia que culminan con la síntesis de dos moléculas de fructosa-6-fosfato (F6P) y una molécula de gliceraldehído-3-fosfato (GA3P). Estos productos entran a formar parte de la glucólisis/gluconeogénesis. Las reacciones son las siguientes:

1. $X5P + R5P \rightarrow$ sedoheptulosa-7-P + GA3P (transcetolasa)
2. Sedoheptulos-7-P + GA3P \rightarrow F6P + eritrosa-4-P (transaldolasa)
3. $X5P +$ eritrosa-4-P \rightarrow F6P + GA3P (transcetolasa)

Todo esto explica la gran importancia de la proteína diana del presente proyecto. Tiene un alto control sobre la ruta de las pentosas fosfato, y la hace indispensable en el organismo, ya que a partir de esta vía de degradación se sintetizan muchos metabolitos necesarios para la vida. Mantiene la producción, la reserva antioxidante (NADPH) y ofrece protección a las células ante posibles agentes oxidantes.

2.1.1. Estructura y función de la G6PDH

El monómero de la enzima G6PDH consta de 515 aminoácidos, con un peso molecular de 59 kDa aproximadamente [1]. Un modelo de estructura tridimensional de G6PDH fue publicada en 1996 [35] y posteriormente la estructura cristalina de la G6PDH humana se ha dilucidado con el paso del tiempo. La enzima es activa como tetrámero o dímero. Cada monómero consta de dos dominios: el dominio N-terminal (aminoácidos 27–200), con un sitio de unión a dinucleótidos $\beta - \alpha - \beta$, que lo llamaremos dominio 1 (aminoácidos 38–44); y un segundo dominio $\beta + \alpha$ más grande, consistente en una hoja antiparalela de nueve hilos, dominio 2.

Ambos dominios están unidos por una hélice α , que contiene el péptido de ocho residuos totalmente conservado actuando como el sitio de unión al sustrato (aminoácidos 198–206). La visualización de la estructura a una resolución de 3 Å muestra una molécula de coenzima NADP⁺ en cada subunidad del tetrámero, cerca del interfaz del dímero, aunque distante del sitio activo [2]. La estabilidad de la estructura cuaternaria de la enzima es crucial para la actividad normal de ésta.

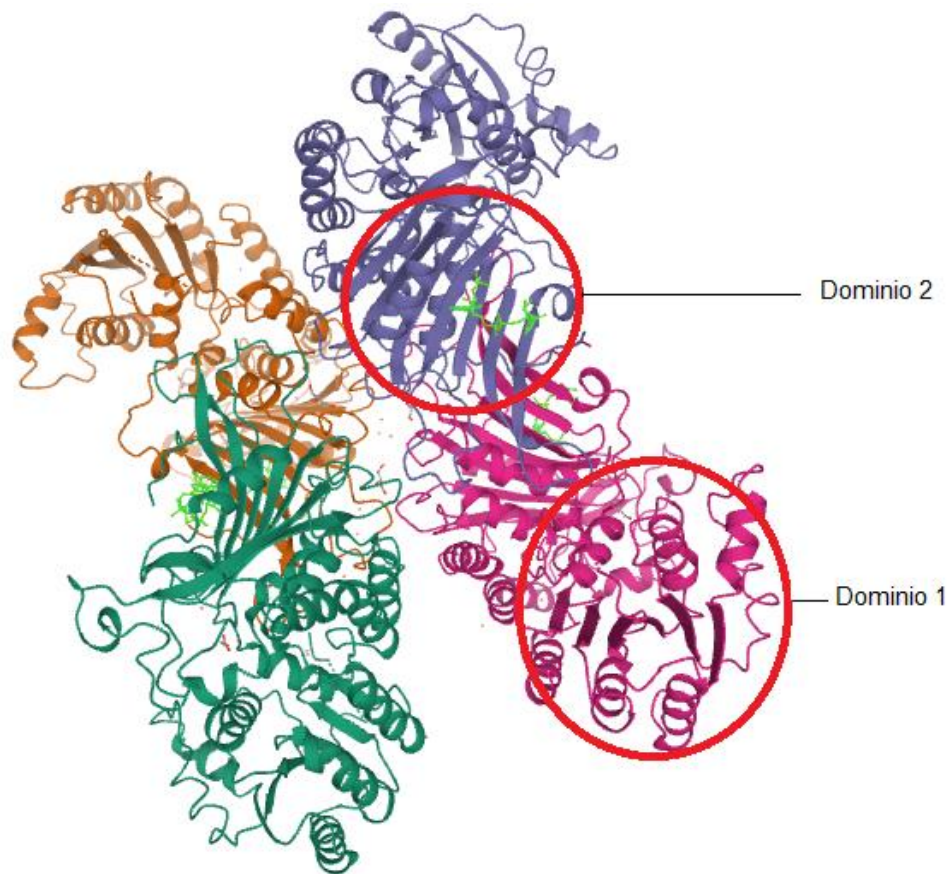


Ilustración 2. Tetrámero de la G6PDH

La Ilustración 2 presenta la estructura tetramérica de la G6PDH. Están expuestas las cuatro subunidades o monómeros del tetrámero en diferentes colores: amarillo, morado, verde y rosa. Cada monómero consta de dos dominios que denominaremos dominios 1 y 2 en la ilustración. Y las moléculas de NADPH se visualizan en color verde más claro.

La proteína G6PDH se encuentra en todas las células del organismo de los seres vivos, sin embargo, varía su concentración dependiendo del tejido en el que se encuentre. En los glóbulos rojos sanos, la G6PDH funciona solo entre el 1 y el 2% del máximo de su potencial. Existe una gran reserva de potencial reductor que se encuentra sustancialmente disminuido en sangre con células con deficiencia de G6PDH, y, en consecuencia, se llegan a características fisiopatológicas. Después de varios estudios y siendo evidente estas características clínicas, la deficiencia de la enzima G6PDH se estableció como un trastorno clínico.

De hecho, la OMS (Organización Mundial de la Salud) en 1967, hizo las primeras recomendaciones para la caracterización de la deficiencia citada que posteriormente fueron actualizadas. Inicialmente, este déficit se caracterizó bioquímicamente, midiendo la actividad enzimática y movilidad electroforética residual. Ya hay más de 400 variantes bioquímicas de la deficiencia de G6PDH desde que fue definida de acuerdo con otros criterios, en los cuales se incluyen; las propiedades fisicoquímicas (comportamiento cromatográfico y termoestabilidad) y variables cinéticas (variaciones en las concentraciones de sustratos, cambios de pH, uso de compuesto análogos del sustrato...).

2.1.2. Deficiencia de la G6PDH

El trastorno más frecuente del metabolismo de los glóbulos rojos es la deficiencia de glucosa-6-fosfato deshidrogenasa. Es el defecto humano más común en cuanto a reacciones enzimáticas se refiere. Existiendo más de 400 millones de individuos en el mundo con dicho trastorno, datos similares a la enfermedad de la malaria. Los casos detectados de manera más frecuente están en países del continente africano, en Asia, en regiones de la zona mediterránea y en Oriente Medio, aunque también se pueden dar casos en zonas de América del norte y del sur, y en algunos países del norte de Europa.

Se trata de una enfermedad hereditaria genética, debida a mutaciones en el gen G6PD, causando variantes funciones con muchos bioquímicos y fenotipos clínicos.

El déficit de enzimas G6PDH puede ser causado por una diferencia en la estructura de la enzima y que esto provoque un cambio cualitativo, también puede ser por una reducción en el número de moléculas enzimáticas, o por ambas causas.

Las variantes de la deficiencia de la G6PDH se clasifican en cinco grupos dependiendo de la actividad enzimática y las manifestaciones clínicas que presentan [3]:

- *Clase I:* Gravemente deficiente, asociado con anemia hemolítica crónica no esferocítica.
- *Clase II:* Gravemente deficiente (1–10% de actividad residual), asociado con anemia hemolítica aguda.
- *Clase III:* Moderadamente deficiente (10–60% de actividad residual).
- *Clase IV:* Actividad normal (60-150%).
- *Clase V:* Aumento de la actividad (> 150%).

Las variantes también se pueden clasificar como esporádicas polimórficas.

El examen de las variantes de G6PDH muestra que, en la mayoría de los casos, la deficiencia de G6PDH se debe a la inestabilidad enzimática, lo que implica que las sustituciones de aminoácidos en diferentes ubicaciones pueden desestabilizar la molécula enzimática.

Todas las mutaciones del gen G6PD que finalizan en deficiencia de glucosa-6-fosfato deshidrogenasa afecta a la secuencia de codificación [4]. En total se han descrito unas 140 mutaciones y en la mayoría, se han producido cambios de una sola base, que conducen a sustituciones de aminoácidos.

Por otro lado, dado que la enzima G6PDH interviene en la ruta metabólica de las pentosas fosfato, tiene un papel fundamental en la síntesis de los productos obtenidos en esta vía, y de esta enzima depende la cantidad de ribosa fosfato y NADPH sintetizados. La función principal de la coenzima en relación al déficit de G6PDH es primordial. La nicotinamida adenina dinucleótido fosfato dona electrones a la ruta del glutatión reducido utilizado para la protección del eritrocito contra el estrés oxidativo que pueda sufrir la célula por diferentes agentes. Ya que los glóbulos rojos no contienen mitocondrias, la única fuente electrónica que reciben éstos, es mediante la vía de las pentosas fosfato, es decir, del NADPH. Por lo tanto, la defensa del glóbulo rojo contra el daño oxidativo depende mayormente de la actividad de la enzima G6PDH. Hay una relación directa entre la supervivencia de la célula y la glucosa-6-fosfato deshidrogenada. La cantidad de enzima es proporcional a la cantidad de la coenzima NADPH, por lo que si hay menos cantidad de G6PDH lo habrá del NADPH y eso hace que disminuya la

defensa del glóbulo rojo, y por tanto, la rotura de éste. El proceso de destrucción del hematíe se conoce como hemólisis.

Entonces, una caída de los niveles de enzima se asocia invariablemente a episodios hemolíticos en individuos deficientes de G6PDH.

El diagnóstico definitivo de la deficiencia de G6PDH se basa en la estimación de la actividad enzimática, por análisis espectrofotométrico cuantitativo de la tasa de NADPH producción de NADP [5]. Para una determinación rápida de la población, se aplican varios métodos semicuantitativos, uno de ellos es la prueba desarrollada por Motulsky [6] de decoloración de tinte, y otras pruebas puntuales fluorescentes, que indican deficiencia de G6PD cuando no puede fluorecer la mancha de sangre bajo luz ultravioleta [7]. Por otro lado, se pueden ocasionar problemas en el diagnóstico para la medición de la actividad enzimática cuando el paciente presenta episodios de hemolisis aguda.

Aunque la mayoría de la población afectada por deficiencia de G6PDH no presenta síntomas, la enfermedad se manifiesta como una hemolisis aguda e ictericia neonatal. El daño oxidativo que sufren los glóbulos rojos en exceso puede ser provocado por la ingesta de algún alimento determinado (habas), por medicamentos o por diferentes infecciones. También, puede precipitarse la hemolisis en individuos que padezcan algún tipo de trastorno clínico, como diabetes o enfermedades cardiovasculares, con deficiencia de G6PDH.

A día de hoy, el mecanismo por el cual aumenta la sensibilidad de las células al daño oxidativo que conduce a la rotura de éstas no se conoce completamente.

Independientemente de la causa de la hemolisis, clínicamente se caracteriza por ictericia, fatiga, anemia y dolor de espalda.

Las manifestaciones clínicas de la deficiencia de la G6PDH más frecuentes son las siguientes:

- Diferentes anemias hemolíticas producidas por algún agente exógeno que desafíe el estrés oxidativo de las células. Estos trastornos serían anemia hemolítica inducida por infección, por fármacos y por determinados alimentos, lo que se conoce como favismo. Algunas variantes de G6PDH causan hemólisis crónica, lo que conduce a anemia hemolítica no esferocítica congénita.

- La ictericia neonatal sucede cuando un organismo tiene niveles altos de bilirrubina en sangre. Esto se debe al reemplazo de un gran número de glóbulos rojos maduros.

Por otro lado, es esencial destacar la gran importancia que tiene la ruta de las pentosas fosfato. El producto coenzima NADPH es importante para lo descrito anteriormente, protección de glóbulos rojos frente a la oxidación. Pero, en los últimos años, el estudio del metabolismo de esta ruta relacionado con diferentes tumores está en auge. Ese aumento es debido a que uno de los productos obtenidos en la vía de las pentosas, como es la ribosa-5-fosfato, es clave para la síntesis de ácidos nucleicos y para el desarrollo de diferentes tumores. Por lo tanto, la ruta de las PPP juega un papel esencial en el crecimiento y proliferación de las células tumorales, lo que realza la importancia de la actividad de la enzima G6PDH.

Las células con una actividad proliferativa aumentada, como son las células tumorales, exhiben un flujo aumentado de la PPP mediado a través de la activación de la enzima G6PDH [8].

Dado que la replicación de ADN es más activa en las células tumorales que en las células normales, algunos fármacos presentan mejores resultados, haciéndose más efectivos en células tumorales. También, surge mayor efecto en células hematopoyéticas y otros tipos de células de replicación rápida, desgraciadamente. Todo esto hace pensar en el diseño de una alternativa más eficaz para el tratamiento de tumores haciendo uso de la inhibición de diferentes enzimas participativas de la ruta de las pentosas fosfato.

Esto relaciona directamente a nuestra enzima objeto, glucosa-6-fosfato deshidrogenasa, con el estudio del desarrollo de diferentes tratamientos para algunos tumores. De esta forma, la G6PDH se convierte en una proteína diana para estos tipos de terapias inhibiendo su actividad enzimática, así se elimina la expresión genética de la enzima, o haciendo uso de un compuesto que inhiba la actividad de ésta [9].

2.2. Fundamento teórico

La simulación de sistemas moleculares se puede realizar mediante métodos clásicos o cuánticos. Dependiendo de la complejidad del modelo se puede emplear la mecánica molecular (estudio de los átomos mediante las leyes de Newton), la mecánica cuántica que va más allá con el estudio de los electrones, y métodos híbridos denominados QM/MM (*Quantum Mechanics/Molecular Mechanics*) donde un centro reactivo se trata como cuántico mientras que el ambiente que rodea a éste se trata de modo clásico.

2.2.1. Dinámica molecular

La dinámica molecular (DM) es una técnica computacional que conlleva distintas ciencias como la Química, la Física y las Matemáticas para representar, describir y predecir el comportamiento de las moléculas. Tiene un amplio campo de aplicación, desde superficies catalíticas hasta grandes sistemas biológicos como las proteínas.

Por lo general, los sistemas moleculares suelen ser complejos y consistentes en gran número de moléculas, por lo que resulta imposible el estudio de sus propiedades de forma analítica. Por esta razón, la DM emplea métodos numéricos para su resolución.

En dinámica molecular hay que estudiar el costo computacional y la fiabilidad de los resultados. Por ejemplo, en dinámica molecular clásica se hace uso de las leyes de Newton, y su costo es menor que en mecánica cuántica. Por lo cual, hay muchas propiedades interesantes, como síntesis de enlaces, que no se pueden estudiar mediante este método ya que no contempla estados excitados. Los métodos basados en campo de fuerzas, mecánica/dinámica molecular, ignoran los movimientos de los electrones y calculan la energía del sistema como función de las posiciones nucleares.

Uno de los principios físicos de la dinámica molecular que cabe destacar por su uso en el proyecto propuesto es el ensamble canónico (NVT). En éste, el volumen (V) no variará y no se producirá intercambio de masa (N). La temperatura (T) se mantendrá alrededor de la deseada. Los tres parámetros quedarán constantes y utilizaremos un termostato para el control de la temperatura, en nuestro caso, Langevin [32].

En el método propuesto se aplicará un campo de fuerzas clásico a los átomos de la proteína y del disolvente. Los parámetros de éste ya han sido determinados mediante cálculos mecanocuánticos y con observables experimentales de otros sistemas modelo [31].

2.2.2. Campos de fuerza

Un campo de fuerzas se requiere para el cálculo de la energía de un sistema molecular. Un campo de fuerzas queda definido por la expresión matemática usada para determinar la energía del sistema más los parámetros de dicha expresión. Una ecuación generalizada de campo de fuerzas para la energía total de un sistema utilizada en mecánica molecular es la siguiente:

$$U = \sum_{\text{enlaces}} U^{\text{srt}} + \sum_{\text{ángulos}} U^{\text{bend}} + \sum_{\text{torsiones}} U^{\text{for}} + \sum_{i,j} U_{i,j}^{\text{LJ}} + \sum_{i,j} U_{i,j}^{\text{Coulomb}}$$

Ecuación 1. Ecuación generalizada de campo de fuerzas

En esta ecuación, los tres primeros términos corresponden a términos enlazantes y los dos últimos a términos de no enlace. Los términos enlazantes son los términos de energía de enlace de bandas, de enlace angular y de torsión. Los parámetros no enlazantes, correspondientes al término de Lennard-Jones y las contribuciones de Coulomb, describen las interacciones de van der Waals y electrostáticas respectivamente.

Se han desarrollado campos de fuerza específicos para modelización de biopolímeros, como Charmm [36] o Amber [13]. En nuestro caso, haríamos uso

del campo de fuerza Amber, que es muy popular para modelizar proteínas, ácidos nucleicos y carbohidratos.

Amber 18 [14] es un conjunto de programas para el estudio de sistemas mediante DM. Por otro lado, se usará el campo de fuerzas ff14SB [30] para la proteína.

La suite Amber nos proporcionará una serie de programas para aplicar el campo de fuerza ff14SB a biomoléculas que utilizaremos en una serie de etapas. El flujo de trabajo que emplearemos con dichos programas sería el siguiente:

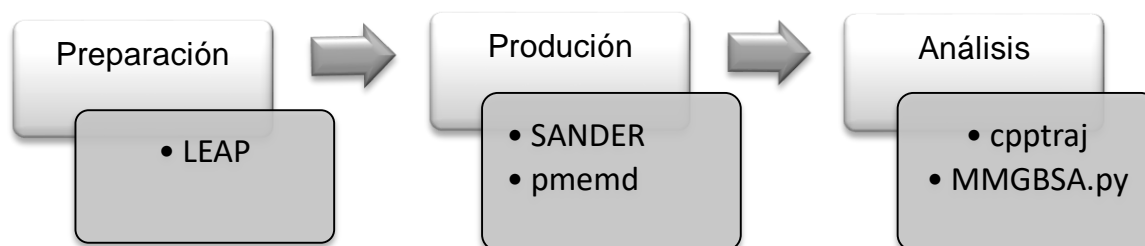


Ilustración 3. Flujo de trabajo de Amber

- El programa LEAP prepara los archivos de entrada necesarios para los programas de simulación.
- SANDER (*Simulated Annealing with NMR-Derived Energy Restraints*). Programa central de simulaciones que utilizaremos tanto para el proceso de minimización como para el de calentamiento que se detallará más adelante.
- Pmemd [21-23]. Una versión optimizada de sander que ejecuta las simulaciones en tarjetas gráficas (GPUs) en lugar de en CPUs.
- El programa cpptraj [17] de la suite Amertools 18 [18] se empleará para el análisis de los resultados de la simulación.
- MM-GBSA.py automatizará el cálculo de energía libre de unión según la metodología MMGBSA [26].

2.2.3. Cálculo de la energía libre de unión

Para el cálculo de las energías libres de unión se utilizan dos métodos que combinan la mecánica molecular y modelos de disolvente implícito. Los métodos mencionados son Mecánica Molecular/Born Generalizado/Área Superficial (MMGBSA) y Mecánica Molecular/Poisson-Boltzman/Área Superficial (MMPBSA).

El método de Born Generalizado (GB), el que utilizaremos para el cálculo de las energías, es similar al método de Poisson-Boltzmann (PB) [27] (empleado para el cálculo de potenciales electrostáticos, es decir, cargas eléctricas estacionarias o cargas que tengan un movimiento lento no acelerado) considerando los iones como esferas con distinta constante dieléctrica.

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

Ecuación 2. Ecuación de Born Generalizada

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

Para el cálculo GB deberemos de controlar algunas variables como la concentración salina y el modelo GB específico que utilicemos. El programa MMGBSA.py agilizará los cálculos de energía libre de unión en estado final, posterior al procesamiento. MMGBSA.py utilizará el programa sander incluido en la suite Amber 18.

El área de superficie accesible al disolvente (SASA) es el área superficial de una biomolécula que puede interaccionar el disolvente. Muchas veces, durante el plegamiento de las proteínas quedan determinadas áreas escondidas, y por lo tanto, no podrá interaccionar con el disolvente.

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

Ecuación 3. Ecuación de la energía libre del SASA

El parámetro σ_i indica el nivel de solvatación del átomo i , y el término $SASA_i$ se refiere al área superficial accesible por ese átomo i .

Para el cálculo de dicha área podremos emplear el método LCPO [28] o mediante el algoritmo Shrake y Ruplay [29].

El cálculo de la energía libre total de unión se realizará a partir de las energías libres del complejo proteína-ligando (C) y los reactivos ligando (L) y receptor (R):

$$\Delta G = \langle G_C \rangle - \langle G_L \rangle - \langle G_R \rangle$$

Ecuación 4. Energía libre de unión estimada a partir de las energías libres de los estados

La energía libre G de cada uno de los estados se calculará mediante la suma:

$$G = E_{\text{int}} + E_{\text{vdW}} + E_{\text{el}} + E_{\text{polar solv}} + E_{\text{nopolar disol}} - TS$$

Ecuación 5. Suma de la energía libre de cada estado

El primer término, E_{int} , corresponde a la energía interna del estado, formada a su vez por las energías de longitud de enlace, ángulo de enlace y diedros.

Los dos términos siguientes, E_{vdW} y E_{el} , determinan las contribuciones de van der Waals y electrostáticas, respectivamente.

Los parámetros, $E_{\text{p solv}}$ y $E_{\text{np disol}}$, son las energías de solvatación. La energía de solvatación polar se obtendrá resolviendo la ecuación de GB expresada anteriormente.

El término $E_{\text{np disol}}$ está directamente relacionado con el área superficial accesible al disolvente, SASA, y se resolverá mediante la ecuación:

$$E_{\text{np disol}} = \alpha \text{SASA} + \beta$$

Ecuación 6. Energía libre de solvatación no polar

Y los dos últimos parámetros, T temperatura, multiplicada por la entropía S.

3. OBJETIVOS DEL PROYECTO

En este proyecto se plantea el diseño de una simulación de dinámica molecular, mediante la suite de programas Amber, utilizando la estructura cristalina de la enzima G6PDH humana, pudiendo analizar todo el proceso de unión de la proteína con un ligando, con el objetivo principal de obtener los posibles sitios de unión y modos de unión.

Dado que tenemos datos experimentales que confirman que determinado compuesto orgánico se une a la proteína y la inhibe, se propone el estudio mediante dinámica molecular para conocer el sitio de unión proteína-ligando y su modo de unión.

El análisis de la dinámica molecular permitirá identificar los diferentes sitios de unión proteína-ligando, además de ayudar a la elección del mejor sitio de unión. Esto permitirá, en un proyecto futuro, plantear mediante cálculos alquímicos, las posibles modificaciones sintéticas para mejorar la energía libre de unión del compuesto, y por tanto su actividad inhibidora.

4. METODOLOGÍA Y PLAN DE TRABAJO

4.1. Diseño de fármacos

Gran parte de los medicamentos que se utilizan a día de hoy son el resultado de un proceso de investigación muy complejo. Para la obtención de fármacos es necesario un gran esfuerzo de unión de diferentes disciplinas científicas para el descubrimiento y desarrollo de éstos, con la finalidad de conseguir efectos clínicos positivos y evitar los efectos secundarios en la medida de lo posible. El estudio y desarrollo de medicamentos durante muchos años se ha realizado mediante métodos experimentales. Pero en los últimos años, este proceso de desarrollo se ha conseguido acelerar gracias a nuevos métodos de cómputo, también conocidos como *in silico*. Estas técnicas computacionales permiten reunir modelos teóricos con mayor precisión, de esta manera procesan grandes cantidades de información sobre dichos fármacos que podrían ser utilizados en futuras investigaciones. Además, estos métodos *in silico* ayudan a entender algunos mecanismos de acción de principios activos de los medicamentos y, como consecuencia, se podría utilizar para mejorar determinadas propiedades de estos fármacos.

El diseño de fármacos mediante métodos computacionales tiene como principio el entendimiento de la relación estructura-actividad que presentan los compuestos y poder aplicarlo con fines terapéuticos deseados contra diferentes enfermedades [10]. Los objetivos principales de estos tipos de métodos son los siguientes:

- Identificar nuevos compuestos. Crear nuevas estructuras con efectos biológicos. Otro objetivo es la búsqueda de compuestos ya existentes que conlleven alguna actividad biológica específica.
- Selección de los candidatos. Estos métodos ayudan a diferenciar moléculas con el fin de conseguir aquella con las características más óptimas.
- Optimización. Mejora de las propiedades deseadas.

La detección de alto rendimiento (HTS, *high throughput screening*) ha sido un enfoque muy importante para el descubrimiento de fármacos desde la década de los años 90, gracias a la cual se desarrollaron fármacos para tratamientos de cáncer, VIH o la diabetes, entre otros [11]. Aun así, incluso utilizando enormes bibliotecas de compuestos, había un enorme espacio químico a explorar, en torno a 10^{63} posibles fármacos, permitiendo que el descubrimiento de fármacos basado en fragmentos (FBDD, de sus siglas en inglés) apareciera como un nuevo método para superar los inconvenientes que conllevaba la HTS. El FBDD se ha convertido hoy en día en una técnica que ha permitido el estudio y desarrollo de más de 30 candidatos a fármacos.

Muchos de los estudios del FBDD, desde una perspectiva teórica, se inspiran en el método MSCS (*multiple solvent crystal structures*) [12]. Este método hace uso de distintos codisolventes orgánicos al resolver la estructura cristalina de la proteína. De esta manera se relaciona las ubicaciones de estos compuestos orgánicos en la proteína con los sitios de unión, sitios alostéricos y puntos calientes. Con el paso del tiempo estos métodos han ido mejorando, empezando desde estructuras estáticas representativas de proteínas hasta llegar a simulaciones de dinámica molecular, permitiendo, con el uso de sondas que simulan la proteína en una solución de moléculas de codisolventes y agua, la adaptación conformacional de la proteína.

Uno de los parámetros importantes entre los métodos teóricos que usan la dinámica molecular es la concentración de codisolvente. Por un lado, una alta concentración de este compuesto conduce a la desnaturalización de la proteína, y por otro lado, se necesitan términos de repulsión para evitar la agregación de codisolventes hidrofóbicos. Cada método ha empleado estrategias diferentes para minimizar dichos inconvenientes.

Otro enfoque teórico relacionado con el FBDD, es la predicción del sitio de unión y modo de unión de un conjunto de fragmentos, en el que un fragmento se coloca aleatoriamente en la caja solvatada y se realizan simulaciones cortas de dinámica molecular (DM). Otra forma de encontrar dicho sitio de unión es mediante el uso

de un gran número de ligandos en la caja, pero en este caso es necesario el uso de un término de repulsión de Lennard-Jones.

4.2. Metodología

En este proyecto proponemos el estudio de un determinado compuesto orgánico que hemos comprobado experimentalmente que se une a la enzima G6PDH e inhibe su actividad enzimática. Para ello, se utilizará un método sistemático y semiautomático basado en dinámica molecular. Este método nos permitirá identificar los sitios de unión proteína-ligando y su modo de unión. El análisis de dinámica molecular nos ayudará a elegir el mejor sitio de unión, y en próximos proyectos, modificar algunas propiedades de ese ligando para mejorar su actividad mediante el uso de cálculos alquímicos.

Este nuevo método podría ser útil para iniciar los primeros pasos de un proyecto de FBDD, en el cuál se obtienen las afinidades y los sitios calientes de una base de datos de fragmentos a una proteína objetivo, para hacer crecer los fragmentos más prometedores y que finalmente se obtenga un medicamento.

El método propuesto, que hemos llamado fdMD (dinámica molecular basada en fragmentos), emplea cajas de simulación de pequeños fragmentos solvatados (copia del ligando) con moléculas de agua usando un término potencial repulsivo de Lennard-Jones para evitar su agregación. De este modo, la solvatación de diferentes proteínas con esta caja preequilibrada hace que contenga un menor número de moléculas de ligandos en estos sistemas de simulación, evitando así el problema de la desnaturalización de la proteína durante la simulación.

El método tiene varias ventajas claras; por una parte, permite solvatar con un número bajo de ligandos la proteína diana, de modo que evita su desnaturalización, y, por otra parte, se obtiene una base de datos de cajas solvatadas con ligandos, muy útil para diferentes estudios que puedan surgir en futuros proyectos de FBDD.

Además, permite clasificar los sitios de unión encontrados mediante el uso de un conjunto de descriptores eligiendo el punto de unión mejor posicionado.

De esta manera, descartaremos con confianza los sitios de unión espurios.

Todos los cálculos descritos en este proyecto se realizarán empleando el campo de fuerzas Amber con el conjunto de programas Amber 18. El solvente se considerará explícitamente haciendo uso de moléculas de agua TIP3P [15] y la distancia de corte para el cálculo de interacciones intermoleculares se establecerá a 11 Å. Las interacciones electrostáticas de largo alcance se tratarán mediante el método de Ewald de malla de partículas [16]. Para el análisis de las trayectorias se utilizará el programa cptraj de la suite Amertools 18. Para este análisis será necesario definir un átomo de carbono al que se le asignará el nombre C99. Dicho átomo de carbono se elegirá de entre los átomos de carbono centrales del ligando, y les adicionaremos un potencial de repulsión de Lennard-Jones para evitar la aglomeración de las copias del ligando presentes en la simulación. Las energías libres de unión se calcularán utilizando la metodología MMGBSA y el programa KDEEP [19]. Este último se encargará de predecir las afinidades de unión mediante redes convolucionales 3D de última generación.

4.2.1. Preparación de la caja del ligando solvatado (Tarea 1)

Previo a la preparación del homodímero protéico es necesario preparar el ligando para obtener la caja de solvatación equilibrada. Para el ligando se utilizará el campo de fuerzas GAFF (*generalized amber force field*) [37]. El ligando se solvatará con moléculas de agua TIP3P mediante el programa Leap, generando así una caja de ligando solvatado. A continuación, la caja se someterá a una minimización y ejecución de DM. La instantánea final se exportará para utilizarla como un cubo de solvatación.

4.2.2. Preparación del homodímero G6PDH humano (Tarea 2)

La estructura 3D inicial de la G6PDH humana se obtendrá del *Protein Data Bank* con el código de entrada 1QKI [20]. Para este trabajo, se requiere obtener el homodímero G6PDH y la estructura de monómero sin NADP⁺. Para conseguir el homodímero se seleccionarán las cadenas A y B de la estructura tetramérica de la enzima, debido a que esta estructura es dependiente del pH y presenta un equilibrio rápido de dímero-tetrámero. La estructura cristalina de la enzima se presenta con la mutación Canton (1376G→T). Esta mutación se cambiará para poder obtener la proteína natural a través del programa Leap de la suite Amber. A continuación, se eliminarán las moléculas de glicerol y ácido glicólico, mientras que se mantendrán las aguas estructurales.

A través del programa Leap de la suite Amber, se agregarán átomos de hidrógeno, se neutralizará el sistema con un número apropiado de contraiones de sodio y se solvatará el sistema con un número apropiado de copias de la caja de ligando solvatado con agua TIP3P. Además, con este programa se asignarán los parámetros del campo de fuerzas ff14sb de todos los átomos de la simulación. El sistema final consistirá en las subunidades A y B de G6PDH, contraiones de sodio, NADP⁺ y una caja cúbica de moléculas de agua y un número variable de moléculas del ligando.

4.2.3. Preparación del sistema (Tarea 3)

Para que el sistema esté en equilibrio es necesario realizar las etapas de preparación de éste. Para ello, acondicionamos el sistema en una serie de pasos.

- *Minimización*

Para eliminar el posible estrés conformacional de la proteína es necesario el paso de la minimización. Para aplicar el proceso de minimización se necesitará: fichero de topología (esto es, los parámetros del campo de fuerzas a utilizar para cada átomo), fichero con las coordenadas de todos los átomos (ambos obtenidos con el

programa Leap en la Tarea 1), y por último, fichero de entrada (input), que incluirá todas las órdenes necesarias para la ejecución del programa utilizado para minimizar, que en este caso será sander.

La minimización se realizará en 5 etapas de 10000 ciclos (maxcyc=10000) cada una. El comando que se debe de utilizar es imin=1 y se refleja como ntmín=1, que indica el método de minimización. Para los 10000 ciclos (ncyc=10000) se usará el método de descenso más pronunciado (steepest descent).

```
Minim_1

&cntrl
  imin=1,
  maxcyc=10000, ncyc=10000, ntmín=1, ntp = 1000,
  cut=10., irest=0, ntx=1,
  ntb=1, ntr=0,
  restraintmask=':1-15,25-59,68-296,301-419,432-506,518-530',
  restraint_wt=1.0,
&end
```

Ilustración 4. Ejemplo de fichero de entrada para la minimización

- *Calentamiento*

La estructura ya minimizada se someterá a un protocolo de calentamiento y equilibrado con el programa pmemd.

Se aumentaría la temperatura gradualmente para evitar la destrucción del sistema debido a un cambio brusco de temperatura y haga que el sistema adquiriera, dentro del mismo, velocidades demasiado altas. El aumento gradual de la temperatura quedaría reflejado con el comando ntt=3, siendo éste el valor máximo deseado y que es controlado por el termostato Langevin. En esta etapa de calentamiento se utilizará el comando ntb=1, indicando el proceso a volumen constante.

También se utilizará el algoritmo de restricción SHAKE [24] y la simulación se realizará en el ensamble canónico NVT, con el comando `ntc=2`, durante los pasos previos de preparación y de producción. Esto implicará que todos los enlaces que conlleven hidrógeno se limitarán a su valor de equilibrio utilizando dicho algoritmo. Al utilizar SHAKE eliminaremos la libertad de movimiento de los enlaces y permitirá que la integración se realice en pasos de mayor duración.

```
1-paso Calentamiento (c)
&cntrl
  imin=0, irest=0, ntx=1, nmropt=1,
  ntc=2, ntf=2, dt=0.002, tol=0.000001,
  cut=9.0,
  nstlim=100000,
  ntp=10000,
  ntb=1, ntp=0,
  ntt=3, gamma_ln=3.0, ig = -1,
  ntwv=0, ntwe=1000, ntwx=1000, ntwr=1000,
  ntr=0, ibelly=0,
  restraintmask=':1-15,25-59,68-296,301-419,432-506,518-530',
  restraint_wt=1.0
```

Ilustración 5. Ejemplo fichero de entrada para el proceso de calentamiento

- *Equilibrado*

Una vez alcanzada la temperatura deseada, 300 K, se mantendrá el sistema a presión constante (`ntb=2`), en el ensamble NPT, es decir, manteniendo el número de partículas, presión y temperatura constantes. De esta forma, al variar el volumen y mantener una presión constante, nuestro sistema aumentaría su densidad hasta que se encuentre en equilibrio.

4.2.4. *Producción o simulación de dinámica molecular (Tarea 4)*

Tras las etapas descritas el sistema ya estaría preparado para someterse a la simulación de dinámica molecular. En este proceso, lanzaremos dos dinámicas moleculares usando `gaMD` [33-34].

La dinámica molecular acelerada de Gauss (gaMD) es un método de muestreo biomolecular mejorado que utiliza un potencial armónico de refuerzo para suavizar la superficie de energía potencial del sistema. gaMD permite el muestreo sin restricciones de estas biomoléculas sin tener que establecer coordenadas de reacción predefinidas. gaMD se encuentra implementado en el programa pmemd.

Se realizará a volumen constante, $n_{tb}=1$, y sin restricciones para la presión.

Durante el proceso de simulación, se indicará el tiempo de duración de cada paso (timestep) de la dinámica molecular, con el comando $dt=0.002$, tendrá un tiempo de 0.002 ps por paso. Con el comando $ntwx=1000$, indicaremos cada cuántos pasos se realizará un archivo mcdrd de la trayectoria, en este caso será cada mil pasos. En nuestra simulación, el estado del sistema en la trayectoria se reflejará cada 2 ps, ya que $0.002 \times 1000 = 2$. A esta lectura del estado instantáneo del sistema se le conoce como 'snapshot'.

4.2.5. Estudio de las trayectorias reactivas (Tarea 5)

Para el análisis de las trayectorias se utilizarán 4 secuencias de comandos conforme con el flujo de trabajo propuesto para nuestro método fdMD.

La primera secuencia de comandos (fdMD-OneTraj.cmd), eliminará las moléculas de agua y los contraiones de la trayectoria, dividiendo además la trayectoria global en N trayectorias proteína-ligando, siendo N el número de ligandos en la simulación, y escribirá la última instantánea de cada una de las N trayectorias en formato PDB.

En el segundo script (fdMD_ReactiveTraj.py), decidirá cuáles de las N trayectorias son reactivas, analizando si el ligando está interactuando con un posible sitio de unión durante un tiempo de simulación establecido al final de la trayectoria. El script considera reactivas aquellas trayectorias en las que la distancia para el átomo C99 entre la última instantánea y las anteriores es inferior a 5 Å durante al menos el 90% de los últimos 20 ns de la simulación.

La tercera (fdMD_MMGBSA_Send.cmd) y cuarta (fdMD_MMGBSA_Analyze.py) secuencia de comandos junto con las trayectorias reactivas generarán entradas

necesarias para la ejecución del siguiente paso, los cálculos de MMGBSA [25] y sus gráficos correspondientes. El cuarto script calculará la energía de enlace promedio para todas las trayectorias reactivas.

En consecuencia, obtendremos: los archivos de salida que permitirán visualizar los posibles sitios de unión, el tiempo de residencia (tiempo que permanece el ligando en cada posible sitio de unión) y la energía de unión de MMGBSA. Además, se podrá calcular la energía de unión con KDEEP a partir de la última instantánea de cada trayectoria, para poder comparar ambos valores y proponer los posibles sitios de unión.

5. BENEFICIOS DEL PROYECTO

5.1. Impacto científico-técnico social y/o económico

Como consecuencia de los experimentos realizados previamente y la gran influencia que suponen las técnicas computacionales para la investigación del diseño y conocimiento de nuevos fármacos, surgió el objetivo principal del proyecto propuesto, la obtención del sitio de unión proteína-ligando y su adaptación conformacional mediante técnicas *in silico*.

En nuestro caso, con estas ayudas de técnicas de modelización molecular, se propone la simulación de la proteína al unirse con un determinado compuesto orgánico y el estudio posterior a la dinámica molecular. El hecho de que la proteína diana sea la G6PDH ampliará el rango de beneficios ya que se trata de una de las enzimas humanas más importantes del organismo.

Los resultados obtenidos en el proyecto podrían ser la base de futuros proyectos de investigación para nuevas terapias u ofrecer otras aplicaciones haciendo uso del análisis que se obtenga de este estudio.

5.2. Plan de difusión

Una vez realizado el proyecto se seguirá con un plan de estudio con los resultados obtenidos y poder aplicarlo en diferentes ámbitos. El beneficio primordial y necesario sería poder obtener un tratamiento adecuado para multitud de enfermedades que a día de hoy no tienen ninguna terapia. Por supuesto, para ello, sería necesario un estudio farmacológico completo en el que se comprobaría realmente su beneficio y los efectos secundarios que conlleve.

Para conseguir una mayor difusión del trabajo se baraja la idea de la asistencia a congresos una vez obtenidos los resultados, y acceder a aquellos que estén a nuestro alcance.

Beneficios del proyecto

Todo esto se llevará a cabo considerando que los resultados obtenidos fuesen de interés. Aun así, se espera la publicación de los resultados en revistas científicas de ámbito internacional.

6. CRONOGRAMA DE TRABAJO

A continuación, se indican las tareas a realizar en el proyecto con sus correspondientes responsables y el tiempo empleado para cada tarea:

<i>Tareas/actividades</i>	<i>Personas involucradas</i>	<i>Primer año*</i>
Tarea 1. Preparación de la caja del ligando solvatado	J.M. Granadino, M. Montejo	x
Tarea 2. Preparación del homodímero G6PDH humano	J.M. Granadino , M. Ansino	x x
Tarea 3. Preparación del sistema	M. Ansino , M. Montejo	x x x
Tarea 4. Producción o simulación de dinámica molecular	J.M. Granadino , M. Ansino	x x
Tarea 5. Estudio de las trayectorias reactivas	J.M. Granadino, M. Ansino	x x x x

*Las x corresponden a los meses de trabajo necesarios para realizar cada tarea a lo largo de un año.

7. PRESUPUESTO

Todo el trabajo se realizaría a través de software. Por ello, será necesario un servidor GPU, puesto que utilizaremos la versión optimizada para GPU, pmemd, para la dinámica. Se estima un importe de 1000€ para el servidor.

Para hacer uso de técnicas de modelización molecular es necesario obtener una licencia de Amber 18 lo que supondría 500€.

En el caso de que sea concedido este proyecto y poder llevar a cabo el trabajo descrito, con la finalidad de conseguir mejores resultados se intentaría una mayor difusión del proyecto mediante publicaciones y asistencia a congresos que estén a nuestro alcance. Se esperan tres publicaciones de los resultados del proyecto y la asistencia de, al menos, dos congresos. El precio estimado por publicación rondaría 1500 € y la asistencia a congresos 2500€ por persona aproximadamente.

Importe total solicitado: 11.000€

8. BIBLIOGRAFÍA

- [1] Luzatto, L. (1995). Glucose-6-phosphate dehydrogenase deficiency. *The metabolic and molecular bases of inherited disease*, 3367-3398.
- [2] Au, S. W., Gover, S., Lam, V. M., & Adams, M. J. (2000). Human glucose-6-phosphate dehydrogenase: the crystal structure reveals a structural NADP⁺ molecule and provides insights into enzyme deficiency. *Structure*, 8(3), 293-303
- [3] World Health Organization. (1989). Working group glucose-6-phosphate dehydrogenase deficiency. *Bull WHO*, 67, 601-611.
- [4] Vulliamy, T., Luzzatto, L., Hirono, A., & Beutler, E. (1997). Hematologically important mutations: glucose-6-phosphate dehydrogenase. *Blood Cells, Molecules, and Diseases*, 23(2), 302-313.
- [5] Beutler, E. (1984). *Red cell metabolism: a manual of biochemical methods*.
- [6] Motulsky, A. G., & Campbell-Kraut, J. M. (1961). *Population Genetics of Glucose-6-phosphate Dehydrogenase Deficiency of the Red Cell*. Bobbs-Merrill.
- [7] BEUTLER, E., & Halasz, A. (1966). A series of new screening procedures for pyruvate kinase deficiency, glucose-6-phosphate dehydrogenase deficiency, and glutathione reductase deficiency. *Blood*, 28(4), 553-562.
- [8] Wise, D. R., DeBerardinis, R. J., Mancuso, A., Sayed, N., Zhang, X. Y., Pfeiffer, H. K., ... & Thompson, C. B. (2008). Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. *Proceedings of the National Academy of Sciences*, 105(48), 18782-18787.

- [9] Zhao, Z.-B., Liu, Y., & Yao, Y. (2014). Computational determination of binding structures and free energies of glucose 6-phosphate dehydrogenase with novel steroid inhibitors. *Journal of Molecular Graphics and Modelling*, 51, 168-172.
- [10] Saldívar-González, F., Prieto-Martínez, F. D., & Medina-Franco, J. L. (2017). Descubrimiento y desarrollo de fármacos: un enfoque computacional. *Educación química*, 28(1), 51-58.
- [11] Macarron, R., Banks, M. N., Bojanic, D., Burns, D. J., Cirovic, D. A., Garyantes, T., ... & Schopfer, U. (2011). Impact of high-throughput screening in biomedical research. *Nature reviews Drug discovery*, 10(3), 188-195.
- [12] Allen, K. N., Bellamacina, C. R., Ding, X., Jeffery, C. J., Mattos, C., Petsko, G. A., & Ringe, D. (1996). An experimental approach to mapping the binding surfaces of crystalline proteins. *The Journal of Physical Chemistry*, 100(7), 2605-2611.
- [13] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., ... & Kollman, P. A. (1996). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 1995, 117, 5179– 5197. *Journal of the American Chemical Society*, 118(9), 2309-2309.
- [14] Case, D. A., Darden, T. A., Cheatham III, T. E., Simmerling, C. L., Wang, J., Duke, R. E., ... & Merz, K. M. (2008). AMBER, version 10. University of California: San Francisco, CA.
- [15] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2), 926-935.
- [16] Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of chemical physics*, 98(12), 10089-10092.

- [17] Roe, D. R., & Cheatham III, T. E. (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation*, 9(7), 3084-3095.
- [18] Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham III, T. E., Cruzeiro, V. W. D., ... & Gohlke, H. AMBER 2018; 2018. University of California, San Francisco.
- [19] Jiménez, J., Skalic, M., Martínez-Rosell, G., & De Fabritiis, G. (2018). K deep: Protein–ligand absolute binding affinity prediction via 3d-convolutional neural networks. *Journal of chemical information and modeling*, 58(2), 287-296.
- [20] Au, S. W., Gover, S., Lam, V. M., & Adams, M. J. (2000). Human glucose-6-phosphate dehydrogenase: the crystal structure reveals a structural NADP+ molecule and provides insights into enzyme deficiency. *Structure*, 8(3), 293-303.
- [21] Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., & Walker, R. C. (2012). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *Journal of chemical theory and computation*, 8(5), 1542-1555.
- [22] Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., & Walker, R. C. (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *Journal of chemical theory and computation*, 9(9), 3878-3888.
- [23] Le Grand, S., Götz, A. W., & Walker, R. C. (2013). SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Computer Physics Communications*, 184(2), 374-380.
- [24] Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3), 327-341.

[25] Miller III, B. R., McGee Jr, T. D., Swails, J. M., Homeyer, N., Gohlke, H., & Roitberg, A. E. (2012). MMPBSA.py: an efficient program for end-state free energy calculations. *Journal of chemical theory and computation*, 8(9), 3314-3321.

[26] Atkins, P. W., & De Paula, J. (2008). *Atkins química física* (No. 544 ATK). Editorial Médica Panamericana.

[27] Fogolari, F., Brigo, A., & Molinari, H. (2002). The Poisson–Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *Journal of Molecular Recognition*, 15(6), 377-392.

[28] Weiser, J., Shenkin, P. S., & Still, W. C. (1999). Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *Journal of Computational Chemistry*, 20(2), 217-230.

[29] Rey, E. Cálculo multiprocesado de SASA.

[30] Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation*, 11(8), 3696-3713.

[31] Orozco, M. (2014). A theoretical view of protein dynamics. *Chemical Society Reviews*, 43(14), 5051-5066.

[32] Wu, X., & Brooks, B. R. (2011). Force-momentum-based self-guided Langevin dynamics: A rapid sampling method that approaches the canonical ensemble. *The Journal of chemical physics*, 135(20), 204101.

[33] Miao, Y., Feher, V. A., & McCammon, J. A. (2015). Gaussian accelerated molecular dynamics: Unconstrained enhanced sampling and free energy calculation. *Journal of chemical theory and computation*, 11(8), 3584-3595.

[34] Miao, Y., Sinko, W., Pierce, L., Bucher, D., Walker, R. C., & McCammon, J. A. (2014). Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *Journal of Chemical Theory and Computation*, 10(7), 2677-2689.

[35] Naylor, C. E., Rowland, P., Basak, A. K., Gover, S., Mason, P. J., Bautista, J. M., ... & Adams, M. J. (1996). Glucose 6-phosphate dehydrogenase mutations causing enzyme deficiency in a model of the tertiary structure of the human enzyme.

[36] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. A., & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, 4(2), 187-217.

[37] Aduri, R., Psciuk, B. T., Saro, P., Taniga, H., Schlegel, H. B., & SantaLucia, J. (2007). AMBER force field parameters for the naturally occurring modified nucleosides in RNA. *Journal of Chemical Theory and Computation*, 3(4), 1464-1475.