

Estudi computacional de la flexibilitat de la proteïna FPPS i mecanisme d'inhibició al·lostèric

Estudiant: Eduard Barcelona Garolera

Grau en Química

Correu electrònic: benedu7@hotmail.com

Tutor: Pedro Salvador Sedano

Cotutor: Ferran Feixas Geronès

Empresa / institució: Institut de Química Computacional i Catàlisi, Departament de Química, Universitat de Girona

Vistiplau tutor (i cotutor):

Nom del tutor: Pedro Salvador Sedano

Nom del cotutor: Ferran Feixas Geronès

Empresa / institució: Universitat de Girona

Correu(s) electrònic(s): pedro.salvador@udg.edu

ferran.feixas@udg.edu

Agraïments:

Abans de començar amb les memòries del treball de fi de grau, m'agradaria agrair especialment a en Ferran Feixas, cotutor del treball de fi de grau, per saber transmetre'm tota la seva experiència i coneixement. Per la seva dedicació i entrega en tot moment del projecte, ja que sense ell no hagués estat possible la confecció d'aquest.

També m'agradaria agrair a tots els companys de classe, per tot el recolzament que ens hem donat mútuament i per fer-me aquest camí de la recta final una mica més amè.

Finalment agrair, com no podia ser d'una altra manera, a tota la meua família que sempre ha estat al meu costat per recolzar-me i fer-me el camí més fàcil.

Resum

L'enzim farnesil pirofosfat sintasa (FPPS) és una proteïna involucrada en la ruta del mevalonat present en cèl·lules tumorals. Inhibir aquest enzim provoca l'apoptosi de cèl·lules canceroses. Els principals inhibidors d'aquest enzim són els fàrmacs bifosfonats. Durant els darrers anys s'ha comprovat que aquests fàrmacs han provocat una sèrie d'efectes secundaris. Per això és interessant dissenyar nous inhibidors per aquest enzim.

Pel disseny computacional de nous fàrmacs utilitzant tècniques de *docking* molecular és molt important disposar de l'estructura de la proteïna que es pretén inhibir. Tot i això, aquesta proteïna ha de tenir la conformació adequada. Per obtenir aquesta conformació determinada, és necessari conèixer els moviments que tenen lloc a la proteïna, les interaccions i el comportament del centre actiu o al·lostèrics. Per analitzar aquests paràmetres, en aquest estudi s'ha realitzat una simulació de dinàmica molecular de la proteïna FPPS per tal de conèixer els canvis que tenen lloc en la seva estructura. S'ha realitzat una simulació en els estats *apo* (FPPS) i *holo* (FPPS + inhibidor). Els resultats obtinguts mostren que la FPPS és una proteïna altament flexible en estat *apo* que pot alternar entre una conformació oberta i tancada. D'aquestes dues simulacions s'han escollit les estructures més rellevants i s'ha realitzat un *docking* amb dos lligands diferents. En primer lloc, s'ha realitzat el *docking* del substrat GPP amb les diferents estructures de la proteïna seleccionades. El GPP és un substrat de la FPPS que es col·loca al centre actiu de la proteïna. Mitjançant el *docking* i la posterior visualització dels resultats s'ha analitzat la posició del substrat al centre actiu de la proteïna i l'energia d'interacció d'aquest amb la proteïna per tal de saber quina és l'estructura més adequada per al disseny d'un nou fàrmac que es col·loqui al centre actiu. En segon lloc s'ha realitzat el *docking* amb el lligand BFH sobre les estructures seleccionades. El BFH és un inhibidor al·lostèric conegut de la proteïna que no conté el grup bifosfonat. Visualitzant els resultats s'ha analitzat la posició que aquest lligand pren en el centre al·lostèric i quins són els aminoàcids determinants per a aquesta interacció.

Mitjançant l'anàlisi de la disposició dels lligands GPP i BFH dins el centre actiu/al·lostèric respectivament, s'han escollit les millors conformacions del centre actiu i al·lostèric per al futur disseny de fàrmacs de la proteïna FPPS.

Resumen

La enzima farnesil pirofosfato sintasa (FPPS) es una proteína involucrada en la ruta del mevalonato presente en células tumorales. Inhibir esta enzima provoca la apoptosis de células cancerígenas. Los principales inhibidores de esta enzima son los fármacos bifosfonatos. Durante los últimos años se han encontrado muchos efectos secundarios relacionados con esta enzima. Por lo tanto, es interesante diseñar nuevos inhibidores por esta enzima.

Por el diseño computacional de nuevos fármacos utilizando técnicas de *docking* molecular es muy importante disponer de la estructura de la proteína que se pretende inhibir. Por eso, esta proteína tiene que tener la conformación adecuada. Para obtener esta conformación determinada, es necesario conocer los movimientos que tienen lugar a la proteína, las interacciones y el comportamiento del centro activo o alostérico. Para analizar estos parámetros, en este estudio se ha realizado una simulación de dinámica molecular de la proteína FPPS para conocer los cambios que tienen lugar en su estructura. Se ha realizado una simulación en los estados *apo* (FPPS) y *holo* (FPPS + inhibidor). Los resultados obtenidos muestran que la FPPS es una proteína altamente flexible en estado *apo*, que puede alternar entre una conformación abierta y cerrada. De estas dos simulaciones se han escogido las estructuras más relevantes y se ha realizado un *docking* con dos ligandos diferentes. En primer lugar, se ha realizado el *docking* del sustrato GPP con las diferentes estructuras de la proteína seleccionadas. El GPP es un sustrato de la FPPS que se coloca en el centro activo de la proteína. Mediante el *docking* y la posterior visualización de los resultados se ha analizado la posición del sustrato al centro activo de la proteína y la energía de interacción de este con la proteína para saber cuál es la estructura más adecuada para el diseño de un nuevo fármaco que se coloque al centro activo. En segundo lugar se ha realizado el *docking* con el ligando BFH sobre las estructuras seleccionadas. El BFH es un inhibidor alostérico de la proteína que no contiene el grupo bifosfonato. Visualizando los resultados se ha analizado la posición que este ligando toma en el centro alostérico y cuáles son los aminoácidos determinantes para esta interacción.

Mediante el análisis de la disposición de los ligandos GPP y BFH dentro del centro activo/alostérico respectivamente, se ha escogido las mejores conformaciones del centro activo y alostérico para el futuro diseño de fármacos de la proteína FPPS.

Abstract

The enzyme farnesyl pyrophosphate synthase (FPPS) is a protein involved in the route of the mevalonate present in tumour cells. Inhibiting this enzyme causes the apoptosis of carcinogenic cells. The most used drugs to inhibit this enzyme are the bisphosphonates. However, in the last years several secondary effects of these drugs have been identified. Therefore it is interesting to design new drugs for this enzyme. For the computational design of new drugs using molecular docking techniques it is very important to have the structure of the protein that one tries to inhibit. Because of it, this protein needs to be in the suitable conformation. To obtain this certain conformation, it is necessary to know the motions that take place into the protein, the interactions and the behaviour of the active or allosteric site. To analyse these parameters, in this study we have performed molecular dynamics simulations of FPPS to elucidate the dynamic changes that take place in its structure. We have carried out a simulation in the *apo* (FPPS) and *holo* (FPPS + inhibitor) states. The obtained results show that the FPPS is a highly flexible protein in the *apo* state and it can alternate between an open and closed conformation. From these two simulations the most relevant structures have been extracted and docking has been carried out with two different ligands. First, we have performed the docking of the GPP substrate into the selected conformations of the protein. The GPP is a substrate of the FPPS that is placed in the active centre of the protein. Using the docking results and visualization techniques, the position and the interaction energy of the substrate on the active site has been analysed, to know which is the most appropriate structure for the design of novel active site drugs. Secondly the docking was performed with the ligand BFH on the selected conformations. The BFH is an allosteric inhibitor of FPPS that does not contain the bisphosphonate group. By visualizing the results we were able to analyse the position that BFH adopts in the allosteric site and which are the most important amino acids to account for this interaction.

Through the analysis of the orientation and the interaction energy of the GPP and BFH ligands in the active/allosteric site respectively, we have been able to select the best conformations of the active and allosteric sites that can be used for the future design of drugs targeting the FPPS protein.

ÍNDEX:

| | |
|---|----|
| 1. INTRODUCCIÓ | 1 |
| 2. OBJECTIVES | 4 |
| 3. METODOLOGIA | 5 |
| 3.1 Dinàmica molecular..... | 5 |
| 3.2 <i>Docking</i> | 6 |
| 3.3 Visualització | 7 |
| 4 RESULTATS I DISCUSSIÓ | 8 |
| 4.1 Dinàmica molecular..... | 8 |
| 4.2 <i>Docking</i> | 10 |
| 4.2.1 <i>Docking</i> del substrat GPP..... | 11 |
| 4.2.2 <i>Docking</i> de l'inhibidor al·lostèric BFH | 22 |
| 5 CONCLUSIONS..... | 28 |
| 6. BIBLIOGRAFIA..... | 30 |

1. INTRODUCCIÓ

L'enzim farnesil pirofosfat sintasa (FPPS), és una proteïna encarregada de catalitzar una sèrie de reaccions de condensació seqüencials que porten de pirofosfat de dimetilalil (DMAPP) a geranil pirofosfat (GPP) i de GPP a farnesil pirofosfat (FPP). Aquest enzim forma part de la via del mevalonat¹, punt de partida de la formació de terpens, molècules fonamentals pel creixement i supervivència de la cèl·lula. Per exemple, la molècula de FPP, de fórmula $C_{15}H_{28}O_7P_2$, és un intermedi en la biosíntesi dels terpens, terpenoides i esterols. Aquests tres compostos, són sintetitzats dins la cèl·lula a partir de l'àcid mevalònic, provinent de l'acetil coenzim A.

Inhibir enzims involucrats en la via del mevalonat¹ és interessant des del punt de vista farmacològic. Actualment, existeixen un cert nombre de fàrmacs que inhibeixen proteïnes de la via del mevalonat; les estatines, utilitzades per reduir els nivells de colesterol, i els fàrmacs bifosfonats, utilitzats pel tractament de diverses malalties degeneratives dels ossos, com la osteoporosis, el càncer amb metàtesi òssia, etc. En aquest treball ens centrarem en l'estudi d'aquest últim tipus de compost, els compostos bifosfonats.

Els compostos bifosfonats presenten propietats farmacològiques degut a que són bon inhibidors de la proteïna FPPS, present en osteoclasts (cèl·lula òssia), fet que permet trencar la via del mevalonat bloquejant la prenilació de la proteïna GTPasa². La reacció de prenilació consisteix en la transferència d'un grup farnesil a la proteïna GTPasa. Fet que acaba provocant l'apoptosi de la cèl·lula, en aquest cas de l'osteoclast, aquesta és un tipus de cèl·lula òssia que degrada, reabsorbeix i remodela els ossos.

A la figura 1 es mostra la via del mevalonat¹, aquesta via parteix de dues molècules de acetil coenzim A que reaccionen mitjançant una condensació de Claisen per donar acetoacetil coenzim A. Després d'una sèrie de passos de reacció catalitzats per diferents enzims, els productes finals d'aquesta ruta són l'IPP i el DMAPP. L'IPP és el pirofosfat de 5-isopentil, mentre que el DMAPP és el pirofosfat de dimetilalil. Aquestes dues molècules, són precursors de la reacció catalitzada per a la proteïna FPPS que dóna com a substrat la molècula de geranil pirofosfat (GPP) i com a producte el farnesil pirofosfat (FPP).

A la Figura 2 es mostra la reacció que produeix GPP i FPP a partir de les molècules de IPP i DMAPP, on la proteïna FPPS actua com a catalitzador.

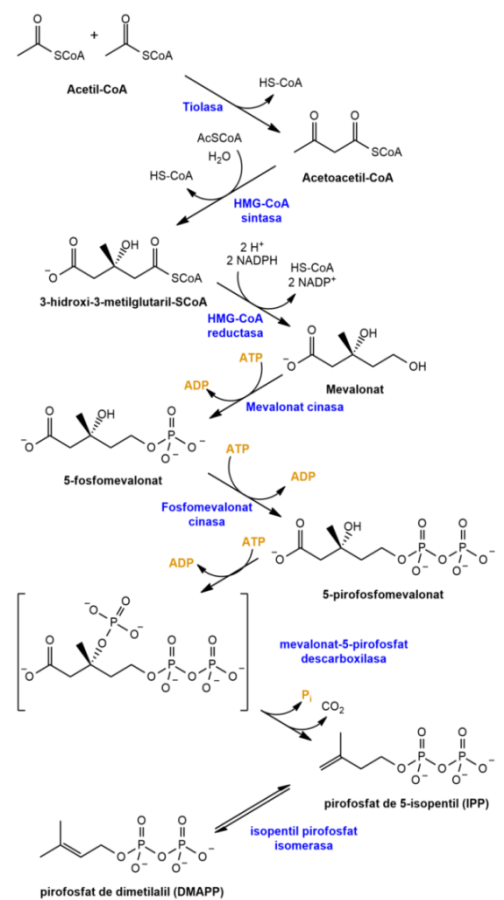
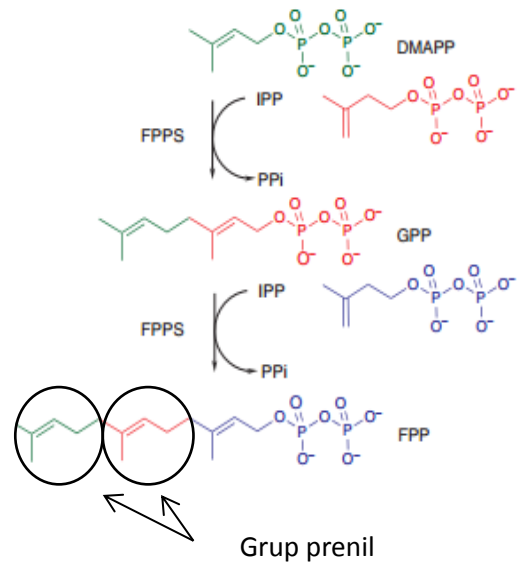


Figura 1. Ruta del mevalonat

Aquesta figura, mostra, en primer lloc i marcat en verd, l'estructura de la molècula de DMAPP, i en vermell, l'IPP. Aquestes dues molècules s'uneixen al centre actiu de la proteïna FPPS i a través d'una reacció de condensació produeixen la molècula de GPP.

En el segon pas de condensació, el GPP reacciona amb una segona molècula d'IPP, donant lloc a la molècula de FPP.

Com es pot veure, la reacció bàsica d'aquesta ruta és la incorporació d'un grup prenil, marcat en verd i vermell, a la molècula GPP per donar FPP. De manera que la molècula es va allargant progressivament.



Tenint en compte que la proteïna FPPS és l'encarregada de catalitzar la reacció, el centre actiu de la proteïna, on es dona aquesta transferència del grup prenil, ha de ser prou gran per encabir totes aquestes molècules.

Figura 2. Reacció de prenilació catalitzada per la proteïna FPPS. Adaptada de Jahnke i col·laboradors⁵

Com ja s'ha dit, els bifosfonats són potents inhibidors de la proteïna FPPS, fent que totes les reaccions comentades anteriorment no es puguin portar a terme. Tot i que els bifosfonats són fàrmacs utilitzats en el tractament de nombroses malalties, entre les quals destaquen la osteoporosis, la malaltia de Paget³, malaltia que provoca l'engrandiment i deformació de l'os, o la metàtesi òssia associades al càncer de mama⁴ i de pròstata, durant els últims anys s'ha constatat que aquests medicaments tenen nombrosos efectes secundaris que són perjudicials en molts casos⁵. Per exemple, s'han donat casos clínics en que la presa de medicament de tipus bifosfonats es relaciona amb la osteonecrosis en l'os maxil·lar de la cara.

Alguns dels medicaments bifosfonats que han estat aprovats i usats durant anys per combatre aquest tipus de malalties han estat els següents:

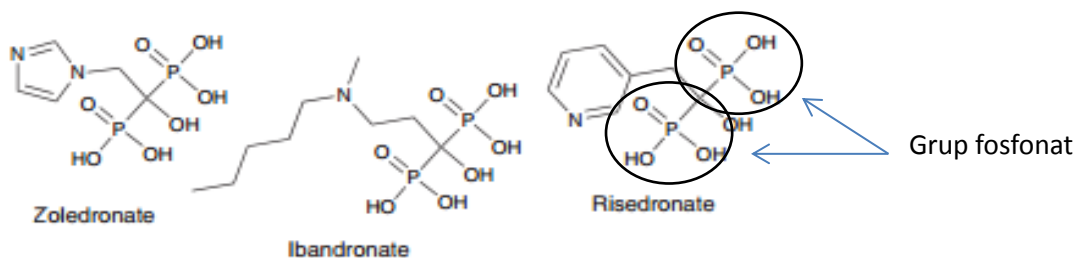


Figura 3. Grup de 3 fàrmacs bifosfonats aprovats

Com es pot observar, les tres molècules anteriors mostren dos grups fosfonats i presenten similitud amb els substrats de la proteïna FPPS com el DMAPP, IPP o GPP. A diferència de molts altres fàrmacs, els bifosfonats són molècules altament carregades i solubles en aigua, propietats que els fan poc selectius.

Degut als efectes secundaris i a les propietats especials que presenten aquests fàrmacs, nombrosos estudis s'han dedicat a la recerca d'un nou inhibidor de la proteïna FPPS que també bloquegi la ruta del mevalonat i per tant no permeti la formació de les molècules de GPP ni de FPP, però que no tingui els efectes secundaris que presenta els bifosfonats.

D'aquests nombrosos estudis, s'han obtingut un nombre destacable d'inhibidors no bifosfonats que presenten una bona interacció amb la proteïna FPPS⁵⁶. Per valorar aquesta interacció es sol analitzar el valor de IC₅₀, aquesta és una mesura de l'eficàcia d'una substància en la inhibició d'una funció biològica o bioquímica específica. Aquesta mesura quantitativa indica la quantitat d'un fàrmac que es necessita per inhibir un procés concret, per tant com més baix sigui el valor, més potent serà el fàrmac. Els valors s'expressen com a concentració molar. Fins l'any 2010, els valors de IC₅₀ per inhibidors no bifosfonats oscil·laven entre 5 i 20 µM (potència moderada), fins que l'any 2010 Jahnke i col·laboradors van dissenyar un inhibidor que mostrava un valor de IC₅₀=0.2 µM, aquest inhibidor és conegut com a BFH⁵.

El BFH es tracta d'un inhibidor al·lostèric, degut a que dins la proteïna no es col·loca en el centre actiu, sinó que ho fa en un centre al·lostèric de la proteïna molt proper al centre actiu. Aquest inhibidor, no té els efectes secundaris que presenten els medicaments bifosfonats degut a la diferència en la seva estructura.

Si observem la estructura del BFH observem que aquesta pot realitzar nombrosos tipus d'interaccions, una de les quals és l'enllaç d'hidrogen degut a la capacitat acceptora d'enllaç d'hidrogen que presenten els dos grups carboxilat de la molècula. Un altre tipus d'interacció que pot donar el BFH és a interacció π-stacking entre els 3 anells de la molècula i els aminoàcids aromàtics de la proteïna FPPS. És per aquestes

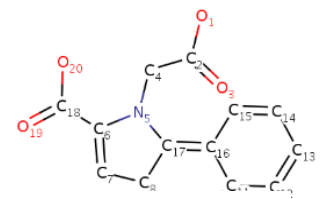


Figura 4. Estructura del BFH

interaccions que es vol aconseguir entendre com es col·loca el BFH dins el centre al·lostèric de la proteïna, quin és el mecanisme d'inhibició i perquè presenta unes millors característiques respecte als fàrmacs bifosfonats. Per a respondre aquestes preguntes s'ha realitzat l'estudi exhaustiu de la proteïna FPPS, degut a l'interès farmacològic i biològic que té aquesta, posant especial atenció amb les interaccions que estableix amb el substrat i l'inhibidor al·lostèric de la proteïna. A continuació es mostra una imatge on es veu la proteïna i la posició que pren l'inhibidor BFH, marcat en groc, al centre al·lostèric de la proteïna A més es pot veure la posició que pren el PO₄³⁻, marcat en vermell, el qual es col·loca al centre actiu de la molècula. També podem veure com es disposen els aminoàcids al voltant del grup fosfat i l'inhibidor BFH. Com veurem més endavant els aminoàcids més destacats són l'Asn50, la Lys48, la Phe230, i l'Arg104.

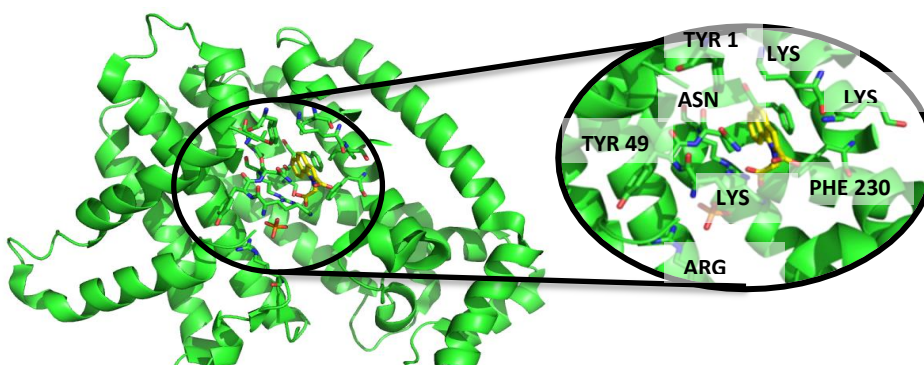


Figura 5. Estructura cristal·logràfica de la proteïna FPPS i representació del centre actiu i al·lostèric d'aquesta

2. OBJECTIVES

Farnesyl pyrophosphate synthase (FPPS) is an important drug target for the treatment of bone related diseases, but also as an antibacterial and antitumoral target. FPPS takes part in the biologically important mevalonate pathway catalysing the formation of farnesyl pyrophosphate, an important molecule for the synthesis of terpenes. Bisphosphonates are potent and highly effective inhibitors of FPPS. These inhibitors have been extensively used for the treatment of bone diseases such as osteoporosis, Paget's disease, or tumour-induced osteolysis. However, the high affinity of bisphosphonates for bone mineral seems suboptimal for the direct treatment of soft-tissue tumours and bacterial infections. Due to the secondary effects and the chemical properties of bisphosphonates drugs, a lot of effort has been made to design potent and effective non-bisphosphonate FPPS inhibitors. Recently, Jahnke and co-workers have suggested a number of non-bisphosphonate allosteric inhibitors of FPPS, the most potent being the BFH compound with an IC_{50} of $0.2 \mu M^5$. However, the mechanism by which this compound is able to inhibit FPPS is still unknown. Protein dynamics plays a key role on protein function. Understanding the role of protein dynamics is crucial to improve the drug design process.

The main objective of this study is to understand in detail the protein dynamics of FPPS and the mechanism of action of BFH in order to be able to improve the potency and selectivity of this compound.

First, we aim to evaluate the flexibility of FPPS by means of molecular dynamics simulations. Some proteins are more flexible than others and this have important consequences for understanding the mechanism of action of inhibitors. To this end, two molecular dynamics simulations will be performed, one with only the FPPS protein (*apo* state) and the other with the FPPS and the BFH inhibitor in the allosteric site (*holo* state). We will analyse the differences between the two states in order to understand the flexibility of the active and allosteric site regions. *Clustering* techniques will be used to extract the most relevant conformations of the molecular dynamics simulations. These selected protein conformations will be used in the second step.

Second, we will perform molecular docking on different structures extracted from the molecular dynamics simulations to understand the reaction mechanism and the mechanism of inhibition of FPPS. Docking with the substrate GPP and the inhibitor BFH will be carried out in a series of FPPS conformations. The docking of GPP will provide information about the size of the active cavity and help to understand which type of reactions could take place in it. For example, reactions such as the prenylation can take place in the active site of FPPS. Then, the same FPPS structures will be used to dock the inhibitor BFH. This will provide information of which amino acids are relevant for the interaction between the inhibitor and the protein and the dynamic properties of the allosteric cavity.

The final goal of this project is to identify the best protein conformations to design new active site and allosteric site drugs. These conformations must have the active or allosteric site open to make sure that the inhibitors fit correctly inside the cavity.

3. METODOLOGIA

En aquest apartat es parlarà sobre quines eines s'han emprat per a la realització d'aquest estudi. S'han utilitzat una sèrie de programes per a realitzar la dinàmica molecular de la proteïna FPPS com: el servidor H^{++} , o els programes *AMBER* i *Gromacs*. Per a realitzar el *docking* de substrat i inhibidor s'ha utilitzat el programa *Swissdock*, i finalment per a la visualització dels resultats i la creació d'imatges s'han utilitzat tres programes diferents; *Chimera*, *VMD* i *Pymol*. A continuació es farà una explicació més detallada de cada un.

3.1 Dinàmica molecular

La dinàmica molecular és una tècnica computacional que permet estudiar els processos dinàmics de tot tipus de molècules, entre les quals trobem les proteïnes⁷. Les proteïnes són sistemes amb molts àtoms fet que dificulta l'ús mètodes quàntics és per això que cal recórrer a conceptes de la física clàssica. L'energia de la proteïna es calcula per mitjà d'un camp de forces i la trajectòria que segueix la proteïna al llarg del temps ve definida per la llei de la dinàmica de Newton⁸⁹. Per a realitzar una dinàmica molecular cal una estructura de la proteïna de partida. Aquesta estructura pot ser obtinguda experimentalment o utilitzant models computacionals.

Preparació del PDB de la proteïna:

Per a fer una simulació de dinàmica molecular, és necessari un fitxer pdb de partida, en el nostre cas el pdb de la proteïna FPPS és el corresponent a l'estructura cristal·logràfica 3N6K (<http://www.rcsb.org/pdb/explore.do?structureid=3N6K>). Per a poder utilitzar aquest pdb cal preparar-lo per a realitzar la dinàmica molecular. L'estructura 3N6K no té hidrògens, per tant, el primer pas és el de protonar la proteïna mitjançant el programa H^{++} ¹⁰. Aquest programa es basa en calcular valors de pKa dels grups ionitzables de les macromolècules i afegir els àtoms d'hidrogen que falten d'acord amb els grups funcionals que envolten aquests grups ionitzables. Pel que fa a les histidines, la protonació es va fer de forma manual en base als aminoàcids del voltant. Un cop afegits els hidrògens, s'ha solvatat el sistema afegint 14635 molècules d'aigua en forma de caixa al voltant de la proteïna. Finalment, per tal d'aconseguir un sistema neutre, s'han afegit 12 ions de Na^+ distribuïts aleatòriament dins la caixa de simulació.

Producció de la simulació:

Un cop obtingut el pdb de la proteïna preparat amb els hidrògens i degudament solvat, s'ha fet servir el paquet de programes d'AMBER. El camp de forces utilitzat per descriure la proteïna és el AMBER ff14SB¹¹, aquest és el model físic de la proteïna, és l'expressió utilitzada per calcular l'energia.¹²

Per a poder realitzar una simulació de dinàmica molecular amb el paquet de programes Amber cal realitzar una sèrie de passos, el primer pas consisteix en una minimització, és a dir, s'optimitza el sistema fins a trobar un mínim local, el segon pas és l'anomenat *heating*, aquest consisteix en escalfar el sistema fins de 0 K a 300 K, sense permetre que el volum canviï, per donar energia cinètica al sistema per a que es moguin els

àtoms. En tercer lloc es fa el procés l'equilibració, procés en el qual pot canviar el volum del sistema permetent que l'aigua del sistema ocupi un major volum degut a l'increment de temperatura. Un cop el sistema s'ha equilibrat s'inicia pròpiament la trajectòria de dinàmica molecular. Cada una de les simulacions consta d'una trajectòria de 200 nanosegons, formats per un total de 100.000.000 de conformacions guardades cada 2 femtosegons. S'han realitzat dues simulacions, la primera on la proteïna té la forma *apo*, la proteïna sense lligand, i la segona on la proteïna en la forma *holo*, és a dir, amb l'inhibidor al·lostèric BFH. Totes les simulacions de dinàmica molecular s'han realitzat utilitzant la versió d'AMBER per unitats de processament gràfic (GPU)¹³ implementada a l'ordinador. Biocomp4 que pertany a l'Institut de Química Computacional i Catàlisi.

Per tal d'extreure informació de la simulació, s'ha realitzat un *clustering* de la trajectòria, aquest consisteix en triar les conformacions més representatives del total de les 100.000.000 conformacions. Per fer-ho s'ha realitzat mitjançant el programa Gromacs¹⁴. El criteri de clusterització es basa en el càlcul de l'RMSD entre les diferents conformacions. L'RMSD és una mesura que permet agrupar les conformacions segons les seves similituds estructurals, com més gran és l'RMSD més gran és la diferència entre dues estructures. En aquest cas s'han agrupat estructures que presenten un RMSD màxim de 1.75 àngstroms entre elles.

3.2 Docking

El *docking* molecular és una tècnica computacional que permet calcular de forma ràpida l'orientació i l'energia d'interacció entre una proteïna i un lligand¹⁵. És una tècnica àmpliament utilitzada en el disseny de fàrmacs. En aquest treball s'ha realitzat el *docking*, mitjançant el programa *Swissdock*¹⁶. Normalment el *docking* es realitza sobre l'estructura cristal·logràfica de la proteïna que es pretén estudiar, però els resultats d'aquest *docking* no sempre són els més adequats degut a que no es considera la proteïna com un objecte rígid. En canvi, no és així ja que les proteïnes estan sotmeses a canvis constants de conformació. Per tal de millorar el poder predictiu del *docking*, és interessant considerar més d'una estructura de la mateixa proteïna¹⁷. Com s'ha explicat en l'apartat anterior s'ha clusteritzat la simulació de dinàmica molecular i s'han obtingut una sèrie d'estructures representatives. Com a lligands s'utilitza tan el GPP com el BFH.

Un cop obtingudes aquestes estructures, es procedeix a realitzar el *docking* amb el servidor *Swissdock*. Per fer-ho s'envia el pdb corresponent a l'estructura de la proteïna juntament amb l'estructura del lligand. El *Swissdock* proporciona una llista d'orientacions del lligand respecte la proteïna i uns valors d'energia d'interacció. Per fer-ho aquest programa simula la proteïna com a un objecte rígid, mentre que el lligand és mòbil. El *Swissdock* fa moure el lligand per moltes posicions de la proteïna i en cada una d'elles en calcula l'energia del sistema i seguidament les ordena de més a menys energia d'interacció. Per a realitzar aquest càlcul d'energia d'interacció entre el lligand i la proteïna, ho fa a partir d'un camp de forces de tipus CHARMM¹⁸ i mitjançant la següent equació:

$$\Delta G_{eff} = E_{intra}^{ligand} + E_{intra}^{recept} + E_{inter} + \Delta G_{elec.solv} + \sigma \cdot SASA \quad (1)$$

A continuació s'explica què vol dir cada paràmetre de la equació 1; el primer paràmetre correspon a

l'energia intramolecular del lligand, és a dir el BFH o el GPP. El segon paràmetre fa referència a l'energia del receptor, que en el nostre cas és la proteïna FPPS (aquesta energia serà diferent per a cada estructura). Pel que fa al tercer paràmetre, aquest es refereix a l'energia d'interacció entre el lligand i la proteïna, és a dir, l'energia d'atracció i repulsió de tipus no covalent que hi ha entre el lligand i la proteïna. Aquest valor és igual a la suma de l'energia d'interacció electrostàtica i l'energia de Van der Waals. El quart terme fa referència a l'energia de solvatació del sistema, té en compte la interacció entre el lligand i el solvent, que en el nostre estudi és l'aigua. Per últim, SASA (Solvent Accessible Surface Area) es refereix al solvent que hi cap en l'àrea de superfície on es col·loca el lligand. La mida d'aquesta cavitat dóna informació de lo bona que és la interacció. El símbol σ és un paràmetre constant que té com a valor $0.0072 \text{ kcal}/(\text{mol } \text{Å}^2)$.¹⁹

El mètode de *docking* té en compte aquesta fórmula al complet sempre i quan, s'utilitzi la opció *accurate*, en cas contrari, si s'utilitza *very fast* el *Swissdock* només té en compte els tres primers termes de l'equació 1, és a dir, no té en compte el solvent ni la interacció que aquest fa amb el lligand. En aquest treball es realitzen la majoria de càlculs amb l'opció *very fast* i es compraran en alguns casos amb el mode més acurat.

3.3 Visualització

Per últim, s'han utilitzat programes de visualització per tal d'analitzar els resultats que tant el programa de *docking* com la dinàmica molecular han proporcionat.

En primer lloc, s'ha utilitzat el *Chimera*, aquest és un programa de visualització de molècules i anàlisi d'aquestes. Aquest programa ens ha permès obrir els resultats del *Swissdock* i analitzar en quina posició s'ha col·locat el lligand respecte la proteïna així com l'energia d'interacció associada a cada orientació.²⁰ En segon lloc, per visualitzar les dinàmiques moleculars, s'ha utilitzat el VMD, aquest és un programa de visualització molecular, que serveix per analitzar simulacions de grans biomolècules. Aquest programa ens ha permès observar el moviment del lligand, en el cas de la simulació *holo*, i de tots els aminoàcids de la molècula al llarg de la simulació així com mesurar distàncies entre diferents àtoms de la proteïna.²¹ Per últim, s'ha utilitzat el *Pymol*, aquest igual que els dos programes anteriors es tracta d'un programa de visualització molecular, que en el nostre cas ens ha servit per visualitzar la proteïna, fer comparació de l'estructura entre diferents conformacions de la proteïna i per analitzar la posició relativa dels aminoàcids. També ens ha servit per a la realització de les imatges que apareixen en aquest treball.²²

3.4 Aspectes ètics i de sostenibilitat

Per a la realització d'aquest treball no s'ha utilitzat cap producte contaminant pel medi ambient, degut a que es tracta d'un treball computacional. És més, mitjançant aquest estudi computacional s'ha estalviat l'ús de productes tòxics, així com dissolvents, ja que si s'hagués realitzat aquest estudi al laboratori, s'haurien hagut de sintetitzar moltes molècules, inhibidor, lligand, etc. En canvi, les tècniques computacionals ens permeten realitzar proves d'activitat de compostos sense cap cost per al medi ambient. Tots els resultats d'aquest treball han estat extrets de dades experimentals i no s'han copiat de cap article.

4 RESULTATS I DISCUSSIÓ

La secció de resultats i discussió es divideix en els següents apartats. Primer s'analitza la flexibilitat de la proteïna amb dinàmica molecular en la seva forma *apo* i *holo*. En segon lloc, mitjançant conformacions extrems de la dinàmica molecular, s'avalua la interacció que mostra un substrat de la proteïna FPPS, en aquest cas el GPP, a través de tècniques de *docking*. Aquest segon pas es realitza, en primer lloc, amb conformacions extrems de la forma *apo* de la proteïna, i en segon lloc, per la forma *holo*, per tal de veure les diferències que presenta la dinàmica molecular en aquests dos casos. En tercer lloc, es realitza el mateix procediment amb l'inhibidor BFH. Finalment, es discuteix quines són les estructures més adequades per al disseny d'inhibidors que interaccionin o bé amb el centre actiu/centre al·lostèric de la proteïna FPPS.

4.1 Dinàmica molecular

La proteïna FPPS (PDB: 3N6K), és encarregada de processos biològics rellevants dins el cos humà com ja s'ha dit anteriorment, és molt interessant trobar un inhibidor potent i selectiu per aquesta. La interacció entre una proteïna i un inhibidor és un procés dinàmic on la proteïna canvia de conformació fins a trobar la que millor s'adapti a la forma i propietats de l'inhibidor. L'estructura cristal·logràfica representa la proteïna com una estructura rígida i dona poca informació sobre els moviments que tenen lloc a la proteïna, això pot fer que no sigui l'estructura més adequada per dissenyar nous inhibidors. Per fer-ho cal, en primer lloc, realitzar un estudi complet de la dinàmica d'aquesta proteïna, per tal de conèixer la seva mobilitat, així com, de les interaccions i característiques que tenen lloc al centre actiu i al·lostèric.

Per a realitzar aquest estudi s'han realitzat dues simulacions de dinàmica molecular, una a l'estat *apo* i l'altra a l'estat *holo*. Els conceptes de *apo* i de *holo* mostren en quin estat es troba la proteïna. Si parlem de *apo*, vol dir que la proteïna es troba sense lligands ni inhibidors, mentre que si parlem d'una proteïna *holo*, ens referim a que la proteïna no només està constituïda per aminoàcids, sinó que també pot contenir lligands, inhibidors o substrats. En aquest cas, la simulació *holo* es realitza amb l'inhibidor BFH unit al centre al·lostèric de la proteïna.

Per tal de determinar si la proteïna FPPS es tracta d'una proteïna altament flexible o d'una proteïna més rígida, s'ha estudiat la flexibilitat mitjançant dinàmiques moleculars. Per analitzar el grau de flexibilitat s'han realitzat mesures de distàncies entre aminoàcids seleccionats de la proteïna al llarg de la simulació.

El PDB 3N6K té 341 aminoàcids, el centre actiu d'aquesta proteïna es troba molt proper als dos extrems N- i C- terminal. En aquests extrems s'hi troba una tirosina (Tyr1) i una lisina (Lys341) (veure figura 5 a la introducció). S'ha mesurat la distància que hi ha entre els carboni alfa d'aquests dos aminoàcid (que és de 14.5Å a l'estructura cristal·logràfica). Per tal d'avaluar la flexibilitat de la proteïna s'ha comparat com canvia aquesta distància al llarg de la simulació de dinàmica molecular de la proteïna en els estats *apo* i *holo*.

A continuació es mostra la un gràfic (figura 6) en què es representa la distància en àngstroms entre a

tirosina i la lisina dels extrems de la proteïna, en blau es mostra la simulació *apo* i en vermell es mostra la simulació *holo*.

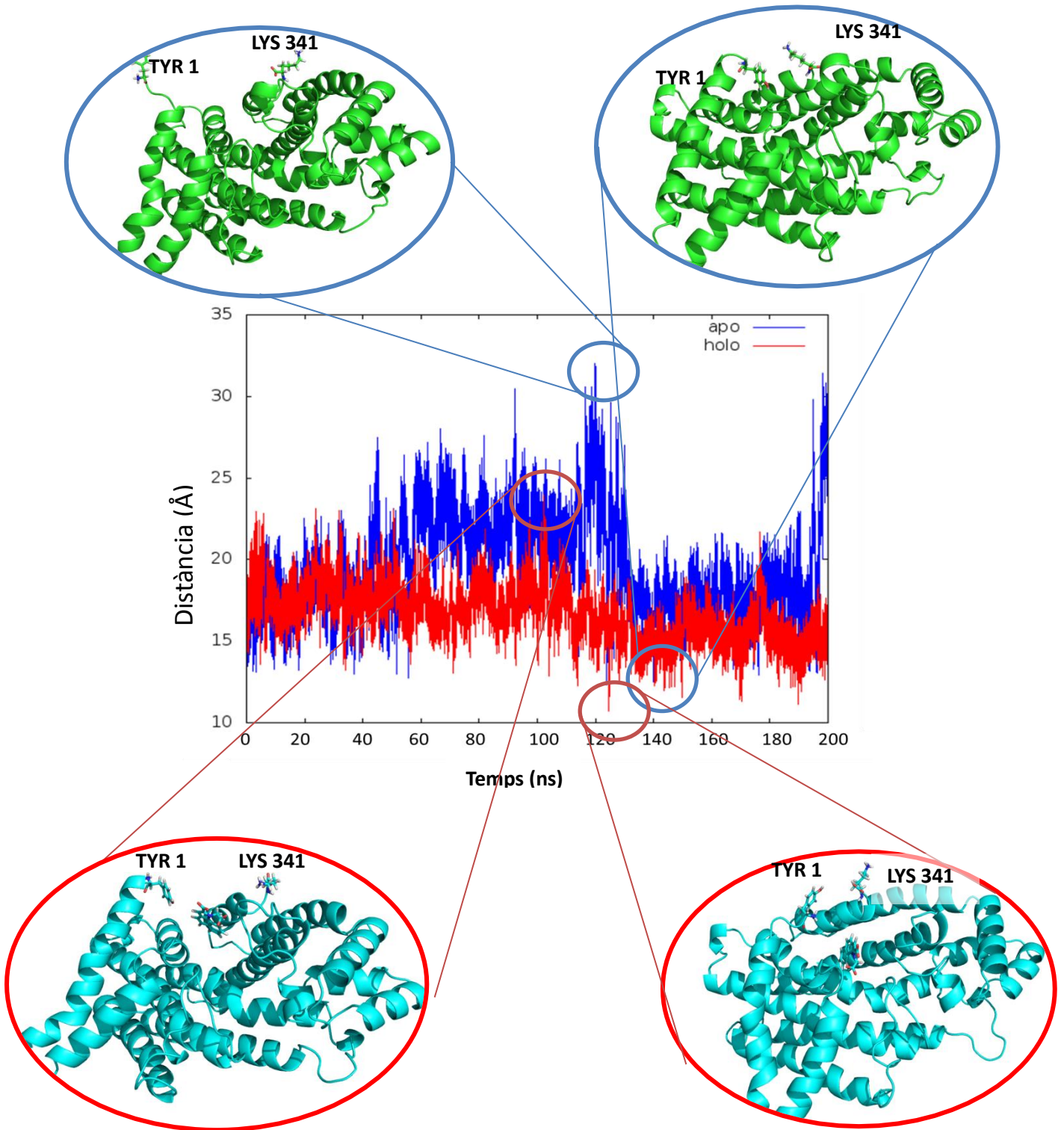


Figura 6. Gràfic que mostra l'evolució de la distància entre el C α de la Tyr1 i el C α de la Lys341 al llarg de la simulació en forma apo (blau) i la forma holo (vermell).

Si observem el gràfic de la figura 6 el que es pot veure és que la distància entre la tirosina i la lisina, en la forma *apo* de la proteïna, és a dir, sense lligand, oscil·la entre valors que van de 10 a 32 àngstroms. Aquesta

àmplia variació mostra la gran flexibilitat que té aquesta proteïna a la regió del centre actiu i centre al·lostèric. Les dues imatges marcades en blau, corresponen a la forma *apo*. A la primera imatge, que correspon al fotograma 599 (equivalent a 120 ns de la simulació), es pot veure que la lisina i la tirosina es troben a molta distància (31.9 Å), aquesta distància mostra l'alta flexibilitat de la proteïna, fent que el centre actiu de la proteïna es trobi en conformació oberta exposat al solvent. En canvi, la segona imatge, corresponen al fotograma 714 (145 ns de la simulació), es pot veure la lisina i la tirosina, que en aquest cas es troben molt a prop (10.8 Å), fent que el centre actiu de la molècula es trobi en conformació tancada.

Pel que fa a la simulació de la forma *holo* de la proteïna, si observem el gràfic es pot veure que la distància entre els dos aminoàcids disminueix respecte a la simulació *apo*, i es pot observar que quan hi ha el lligand dins la proteïna, aquesta perd flexibilitat. Si observem les dues imatges, marcades en vermell, en la primera s'observa que la tirosina i la lisina es troben separades (23.5 Å), tot i que no tant com en el cas de la simulació *apo*, en canvi, en la segona imatge de la forma *holo*, la tirosina i la lisina es troben a poca distància, de manera que el centre actiu de la proteïna es troba més tancat (10.5 Å).

Si comparem els dos gràfics, es pot veure clarament que l'estructura de la forma *apo* és molt més flexible que no pas la forma *holo*. El valor mitjà d'aquesta distància i la seva desviació estàndard al llarg de la simulació *apo* és de 19.58 ± 3.00 Å mentre que per la simulació *holo* els valors són 16.68 ± 1.64 Å. Fet que demostra que la simulació *holo* és menys flexible que la *apo*. Aquesta afirmació s'explica pel fet que en el centre al·lostèric de la forma *holo* hi ha el lligand BFH, es pot observar en les imatges marcades en vermell, i aquest lligand interacciona amb el centre al·lostèric i fa que els aminoàcids no tinguin tanta llibertat de moviment degut a les interaccions que tenen lloc. En canvi, en la forma *apo* de la proteïna, ni el centre actiu ni el centre al·lostèric estan ocupats i presenta més variació, fet que demostra la gran flexibilitat de la proteïna. La dinàmica molecular dels estats *apo* i *holo* ha permès estudiar el canvi de conformació de la proteïna FPPS. Aquesta proteïna en estat *apo* mostra una alternança entre una conformació oberta i una conformació tancada. La pregunta que s'intentarà respondre a continuació és quina conformació és la més adequada per realitzar *docking* del substrat GPP i de l'inhibidor BFH.

4.2 Docking

El *docking* molecular és una tècnica que permet predir de forma ràpida la interacció entre una proteïna i un lligand. Un dels factors claus d'aquesta rapidesa és considerar la proteïna com un objecte rígid i no com un objecte dinàmic¹⁷. Aquesta aproximació fa que el *docking* no tingui un poder predictiu molt elevat comparat amb altres tècniques ja que els resultats obtinguts depenen de la conformació de la proteïna escollida. Les proteïnes molt flexibles adopten conformacions molt diferents al llarg del temps fet que complica l'elecció de la conformació més adequada.

Un cop analitzada la flexibilitat de la proteïna FPPS per mitjà de la dinàmica molecular queda clar que aquesta és altament flexible. Com s'ha mencionat a la introducció, aquesta proteïna té un centre actiu i un

centre al·lostèric conegut i s'han trobat inhibidors pels dos punts d'unió de la proteïna. Els inhibidors bifosfonats interaccionen amb el centre actiu mentre que els bifosfonats ho fan amb el centre al·lostèric. L'objectiu d'aquest apartat és trobar la millor conformació possible per a predir futurs inhibidors tan pel centre actiu com pel centre al·lostèric. El lligand escollit per buscar la millor conformació del centre actiu és el substrat GPP i pel centre al·lostèric s'utilitza l'inhibidor BFH.

Un cop realitzades les dinàmiques molecular s'han escollit conformacions de la simulació. El criteri de selecció ha estat: 1) a l'atzar (basat en la inspecció visual dels fotogrames de la simulació, i 2) amb criteri mitjançant tècniques de *clustering* (veure metodologia). A aquestes estructures se'ls ha realitzat un *docking* amb diferents molècules mitjançant el servidor web *Swissdock*. En primer lloc s'ha realitzat un *docking* de les estructures mitjançant un dels substrats de la proteïna FPPS, el geranil pirofosfat GPP, per tal d'identificar la conformació de la proteïna que millor descriu el centre actiu. En segon lloc s'ha realitzat un *docking* de les estructures obtingudes de les dinàmiques *apo* i *holo* amb un inhibidor del centre al·lostèric, en aquest cas s'ha utilitzat el BFH, que és un inhibidor conegut de la proteïna, per tal de conèixer les interaccions en aquest centre actiu de la proteïna i avaluar aquest inhibidor.

4.2.1 Docking del substrat GPP

En primer lloc s'ha realitzat el *docking* de l'estructura cristal·logràfica de la proteïna FPPS (PDB 3N6K) amb el substrat GPP, per tal de comprovar la interacció que es dona entre el substrat, GPP, i la proteïna. També es pretén comprovar si l'estructura 3N6K (cristal·litzada amb un inhibidor al·lostèric) serveix per predir correctament la interacció del substrat al centre actiu. El *Swissdock* assigna aquesta interacció un valor d'energia d'interacció de -15.62 kcal/mol, que aparentment mostra que hi ha una molt bona interacció entre la proteïna i el substrat. Per tal de veure com interacciona el substrat amb la proteïna, a continuació es mostra una imatge de la posició que ocupa el substrat a la proteïna i amb quins aminoàcids interacciona.

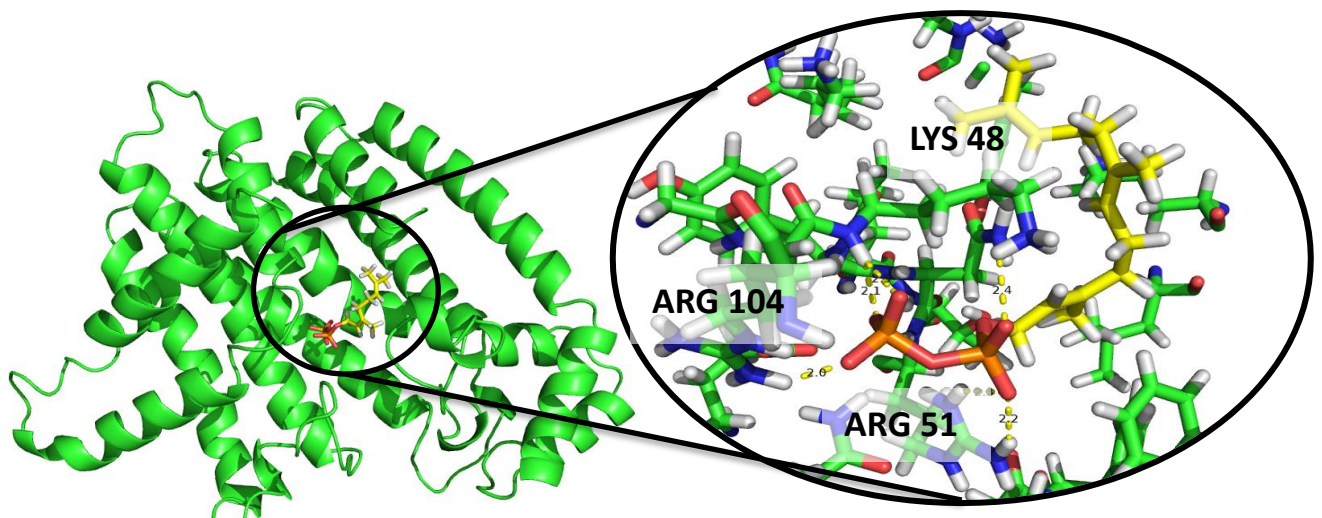


Figura 7. Posició del GPP en el *docking* de l'estructura cristal·logràfica 3N6K

En groc es pot observar la cadena carbonada de la molècula GPP, en taronja i vermell es mostra el grup pirofosfat (dos grups fosfonat), aquest últim grup ocupa la posició corresponent al centre actiu de la proteïna, mentre que la cadena carbonada ocupa el centre al·lostèric. Aquest centre al·lostèric és el mateix que ocupa l'inhibidor BFH, mentre que el centre actiu, és el mateix que ocupa el PO_4^{3-} a l'estructura cristal·logràfica.

Tal com es pot veure a la Figura 7, el grup pirofosfat del substrat interacciona amb la lisina 48 i les arginines 51 i 104 per mitjà d'enllaços d'hidrogen, mentre que la cadena carbonada fa interaccions hidrofòbiques amb els aminoàcids del centre al·lostèric.

A la Figura 7 també es pot observar que les distàncies entre aquests aminoàcids i el GPP són al voltant de 2 àngstroms, pròpia d'un enllaç d'hidrogen. Aquest enllaç d'hidrogen es dona entre un àtom electronegatiu i un hidrogen enllaçat covalentment a un altre àtom electronegatiu, en aquest cas, els oxígens de la molècula de GPP, actuen com a àtom electronegatiu acceptor d'hidrogen, mentre que els hidrògens dels diferents aminoàcids lisina i arginina actuen com a donador d'enllaç d'hidrogen. El fet de que la molècula de GPP tingui una càrrega de -3 situada sobre al grup pirofosfat i els aminoàcids arginina i lisina una càrrega positiva a la seva cadena lateral fa que s'estableixi una interacció electrostàtica forta de tipus pont salí que complementa els enllaços d'hidrogen formats. El dubte que sorgeix és si la orientació del substrat és l'adequada per a què la reacció que catalitza la proteïna FPPS es pugui portar a terme.

Com s'ha comentat a la introducció, per tal que es doni la reacció de prenilació, cal que en el centre actiu de la proteïna hi hagi suficient espai per què reaccionin l'IPP i el GPP per a donar el producte, que en aquest cas és l'FPP, on la proteïna FPPS actua com a catalitzador. Per a que hi hagi prou espai per a que es doni la interacció cal que el centre actiu de la proteïna sigui molt flexible ja que tan la molècula de GPP com FPP són significativament llargues i voluminoses. Tot i que el valor de ΔG és significativament negatiu i per tant la interacció predita pel *docking* és molt favorable des del punt de vista energètic, la orientació de la molècula de GPP respecte a la proteïna no és l'adequada ja que la cadena carbonada ocupa el lloc del centre al·lostèric i no del centre actiu, per tant la prenilació no es podria donar. Els resultats obtinguts amb el *Swissdock* (250 complexos lligand-proteïna analitzats) no mostren cap orientació de la molècula que situï la cadena carbonada dins el centre actiu.

El PDB 3N6K que s'ha utilitzat per a realitzar el *docking* ha estat cristal·litzat amb un inhibidor al centre al·lostèric. Això significa que la presència de l'inhibidor en el centre al·lostèric provoca el tancament del centre actiu, no deixant espai per què el GPP es pugui orientar correctament per a la catàlisi. Per tant, l'estructura 3N6K no és adequada per al disseny d'inhibidors que interaccionin amb el centre actiu. Aquests resultats són importants de cares a entendre el mecanisme d'inhibició de l'inhibidor al·lostèric.

Degut a que les proteïnes no són objectes rígids sinó que estan en constant moviment, s'ha realitzat una sèrie de dinàmiques moleculars per tal d'entendre els canvis conformacionals que tenen lloc i poder trobar

una estructura de la proteïna que permeti predir millor la interacció del substrat amb el centre actiu.

Docking GPP apo:

Primer de tot, com s'ha descrit anteriorment s'ha realitzat una simulació de la proteïna sense l'inhibidor (estat *apo*), un cop realitzada aquesta simulació s'han extret 1000 conformacions de la simulació de 200 ns, dels quals se n'han escollit 5 conformacions a l'atzar a partir de la visualització de la trajectòria de dinàmica molecular.

De cada una d'aquestes conformacions s'ha realitzat un *docking* mitjançant el servidor web *Swissdock* amb el GPP com a substrat.

A continuació es mostren els resultats obtinguts del *docking* sobre les conformacions extretes de la simulació *apo* de la proteïna amb el GPP com a substrat.

| | $\Delta G(\text{kcal/mol})$ | Superposició? |
|------------------------|-----------------------------|---------------|
| Conformació 182 | -11.12 | SÍ |
| Conformació 391 | -9.72 | NO |
| Conformació 611 | -11.25 | SÍ |
| Conformació 785 | -9.26 | SÍ |
| Conformació 978 | -9.93 | NO |

Taula 1. Resultats del *docking* del GPP respecte les conformacions extretes de la simulació de dinàmica molecular en la forma *apo*. Energia d'interacció en kcal/mol i existència de superposició amb la zona del centre actiu/al·lostèric

A la taula 1 es poden observar els diferents valor d'increment d'energia de Gibbs obtinguts, i també es pot observar si hi ha superposició amb el centre actiu/al·lostèric o no, aquest últim paràmetre s'ha estudiat mitjançant la comparació amb l'estructura cristal·logràfica 3N6K, per tant, existeix superposició si el GPP es col·loca a la mateixa posició que el BFH i PO_4^{3-} en l'estructura cristal·logràfica, és a dir, si es col·loca al centre actiu o al·lostèric de la proteïna.

Pel que fa als valors d' ΔG , observem que han disminuït respecte al *docking* de l'estructura cristal·logràfica, és a dir, la proteïna interacciona de manera més dèbil amb el GPP en aquestes estructures que no pas en l'estructura cristal·logràfica. Aquesta interacció dona més baixa que en l'estructura cristal·logràfica degut al moviment dels aminoàcids durant la simulació.

Això pot ser degut que les estructures que s'han triat de manera aleatòria no siguin representatives. Tot i això, com s'ha comentat abans un valor elevat de ΔG no implica que el GPP es posicioni correctament al centre actiu de la proteïna.

Per entendre millor com interacciona el substrat, GPP, amb la proteïna, s'han realitzat una sèrie d'il·lustracions que mostren com es col·loca aquest respecte a cada una de les conformacions utilitzades de la proteïna.

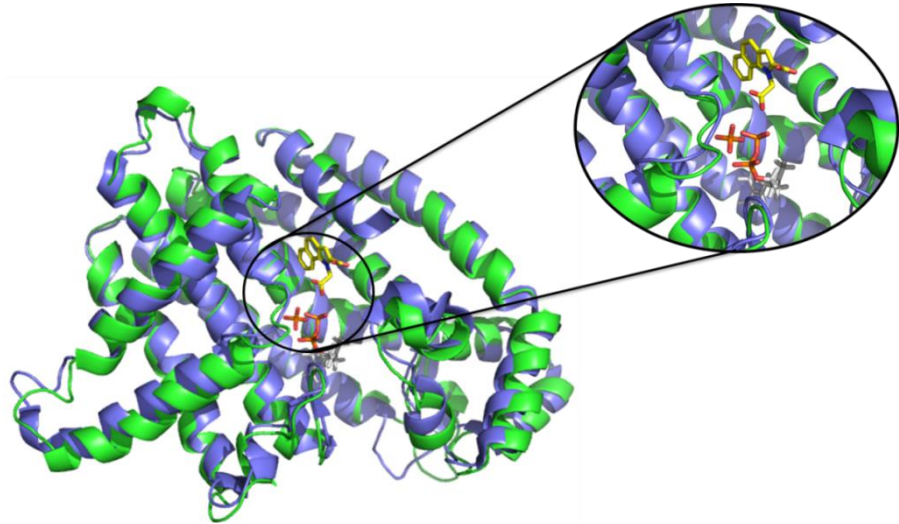


Figura 8. Comparació de l'estructura cristal·logràfica 3N6K (lila) i l'estructura de la conformació 182 (verd)

En aquesta il·lustració (Figura 8) es pot observar com es disposa el GPP dins la proteïna FPPS. En blau es mostra l'estructura cristal·logràfica, mentre que en verd es mostra la conformació 182. Es pot observar que el grup pirofosfat es solapa amb el PO₄³⁻, per tant es col·loca en el centre actiu de la proteïna, mentre que la cadena carbonada, marcada en blanc en la imatge, es disposa de diferent manera que en la estructura cristal·logràfica (veure Figura 8). En el cas de l'estructura cristal·logràfica, la cadena carbonada es col·loca cap a la part de dalt de la proteïna, que correspon al centre al·lostèric, en canvi en la conformació 182, la cadena es disposa cap a dins la proteïna, com s'ha comentat abans això pot ser degut a que a l'estructura cristal·logràfica la part inferior del centre actiu de la proteïna estigui parcialment tancat i per tant la cadena carbonada s'ha de disposar cap al centre al·lostèric. Tot i que la ΔG és menys negativa que en el *docking* utilitzant l'estructura cristal·logràfica, la orientació del GPP en aquest cas és la correcta ja que la cadena carbonada ocupa el centre actiu.

Pel que fa a les conformacions de la simulació en què el substrat es posiciona correctament al centre actiu, la 611 i 785, la cadena carbonada també es disposa diferent que en el cas de l'estructura cristal·logràfica.

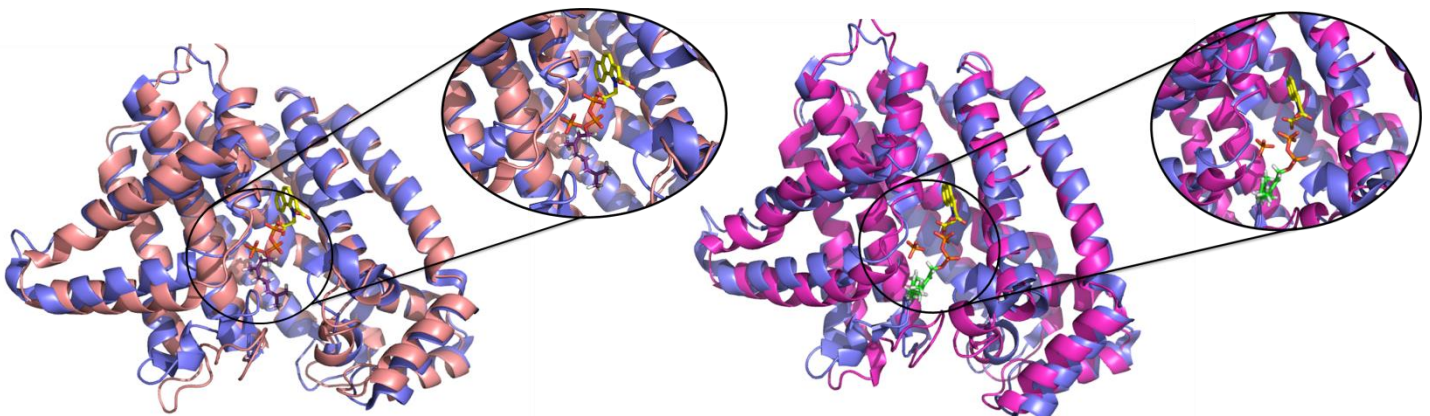


Figura 9. Comparació de les estructures de les conformacions 611 (rosa) i 785 (fúcsia) amb l'estructura cristal·logràfica 3N6K (lila)

Per altre banda, en les conformacions de la proteïna corresponents 978 i 391, el GPP no es col·loca al centre actiu, a continuació es mostra una imatge de la proteïna en la conformació 978 per veure la zona d'interacció més favorable del GPP segons el *Swissdock*.

En la Figura 10 es pot observar en blau l'estructura cristal·logràfica, mentre que en blau més clar, la conformació 978, encerclat es troba el GPP, com es pot observar no es col·loca en el centre actiu com en els casos anteriors. Aquest fet és degut a que en aquesta conformació de la proteïna el centre actiu estigui tancat o parcialment tancat, fet que impedeix que el GPP entri al centre actiu de la molècula..

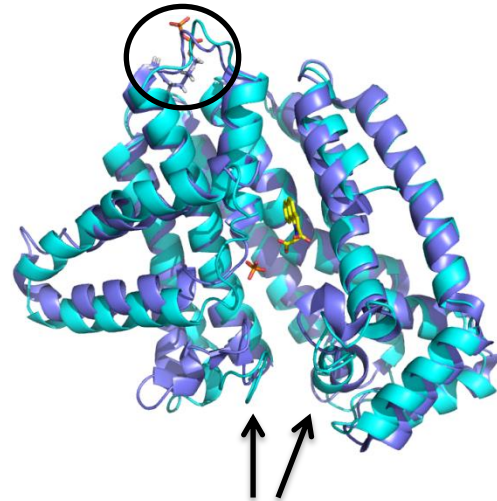


Figura 10. Estructura de la conformació 978 (blau) en comparació amb l'estructura 3N6K (lila)

Aquest tancament de la proteïna és degut al desplaçament que han patit els *loops* marcats a la Figura 10, que fan l'efecte de porta d'entrada pel substrat, tot impedit que el substrat entri i es col·loqui al centre actiu de la proteïna. En l'estructura corresponent a la conformació 978 els dos *loops* marcats es troben a 8.2 Å (distància entre Tyr1 i Lys341), mentre que a l'estructura cristal·logràfica els mateixos *loops* es troben a 14.4 Å. Queda clar que amb aquesta diferència de distàncies, és normal que en el cas de l'estructura de la conformació 978 el GPP no aconsegueixi entrar al centre actiu o al·lostèric.

Per altre banda, cal dir que aquestes estructures han donat valors de ΔG baixos, i per tant, pot ser que les estructures escollides aleatòriament no siguin les millors conformacions per a realitzar estudis de *docking*. Tot i no obtenir resultats energèticament molt favorables, es pot comprovar que a partir de la simulació *apo* s'han obtingut estructures adequades per dissenyar un fàrmac que es col·loqui en el centre actiu. Ja que en els casos de les conformacions 182, 611 i 785, el substrat s'orienta correctament en el centre actiu. Tot i això, cal estudiar si les estructures que s'han són representatives de la simulació o no.

Seleccionar conformacions a l'atzar pot portar a escollir estructures poc representatives o poc habituals al llarg de la simulació. Existeixen tècniques que permeten extreure conformacions biològicament rellevants d'una simulació de dinàmica molecular. És per això que s'ha realitzat un *clustering* (veure metodologia) de la simulació de dinàmica molecular per tal d'obtenir estructures més representatives de totes les conformacions visitades durant la simulació. El *clustering* és un mètode de agrupació basat en un criteri concret, en el nostre cas, el criteri ha estat la semblança basada en l'RMSD que hi havia entre les diferents estructures que pren la proteïna durant la simulació (veure metodologia).

Un cop obtingudes les 5 estructures més representatives de tota la simulació, s'ha realitzat un *docking* de cada una d'aquestes amb la molècula de GPP per tal de veure com interacciona la proteïna amb aquest.

A continuació es mostra una taula dels resultats obtinguts pel *docking* de les 5 estructures, nomenades

clústers. El clúster 1 és el més representatiu de tota la simulació mentre que la resta de clústers corresponen a conformacions de la proteïna cada cop menys freqüents al llarg de la simulació:

| | $\Delta G(\text{kcal/mol})$ | Superposició? |
|--------------------|-----------------------------|---------------|
| Clúster 001 | -11,46 | SÍ |
| Clúster 002 | -13,72 | SÍ |
| Clúster 003 | -9,62 | NO |
| Clúster 004 | -12,73 | SÍ |
| Clúster 005 | -11,58 | NO |

Taula 2. Resultats del *docking* del GPP a les diferents conformacions obtingudes del *clustering* de la simulació de la proteïna *apo* amb la opció *very fast* del *Swissdock*. S'indica l'energia d'interacció en kcal/mol i si existeix superposició amb la regió del centre actiu/al·lostèric.

Observant la taula, podem veure que dels 5 clústers, tant el clúster 005 com el clúster 003, el GPP no es col·loca al centre actiu de la molècula, ja que s'ha comparat amb la posició de l'inhibidor BFH i PO_4^{3-} i no es col·loca en la mateixa posició. En canvi, la resta d'estructures sí que prenen lloc al centre actiu o al·lostèric.

També podem veure que els valors d'energia de Gibbs han incrementat lleugerament, cosa que mostra que la interacció entre la molècula GPP i les diferents estructures ha millorat respecte a les conformacions escollides a l'atzar. Aquests diferents clústers són més representatius i permeten una millor predicció de la interacció lligand-proteïna.

El servidor *Swissdock* té dues expressions per a calcular la ΔG , la *very fast* i la *accurate* (veure metodologia). L'opció *very fast* permet calcular la ΔG i predir l'orientació del lligand d'una forma ràpida tot i que en moltes ocasions es subestimen algunes interaccions importants. Fins a aquest moment, els resultats de *docking* presentats han estat realitzats mitjançant l'opció *very fast*. Per tal d'avaluar si aquesta opció és correcta s'ha realitzat el mateix *docking* però en aquest cas, s'ha realitzat mitjançant la opció *accurate*.

Els resultats del *docking* mitjançant la opció *accurate* són els que es mostren a continuació:

| | $\Delta G(\text{kcal/mol})$ | Superposició? |
|--------------------|-----------------------------|---------------|
| Clúster 001 | -12,01 | SÍ |
| Clúster 002 | -13,97 | SÍ |
| Clúster 003 | -10,39 | NO |
| Clúster 004 | -13,33 | SÍ |
| Clúster 005 | -11,79 | NO |

Taula 3. Resultats del *docking* del GPP a les diferents conformacions obtingudes del *clustering* de la simulació de la proteïna *apo* amb la opció "*accurate*" del *Swissdock*. S'indica l'energia d'interacció en kcal/mol i si existeix superposició amb la regió del centre actiu/al·lostèric.

Igual que en el cas anterior, dels 5 clústers tant el clúster 003 com el clúster 005, el GPP no es col·loca al centre actiu, i la resta sí que es col·loquen al centre actiu de la proteïna, igual que en el cas anterior amb

l'opció *very fast*. Pel que fa a les energies de Gibbs, les 5 ΔG han disminuït en energia, s'han fet més negatius, això vol dir que la interacció, en tots els casos, ha augmentat. Per tant, l'opció *accurate* permet descriure millor les interaccions que tenen lloc tot i que el cost computacional és més elevat.

Un altre dels desavantatges del *docking* és que considera a la proteïna rígida. La proteïna no es pot adaptar al lligand. El *Swissdock* té una opció que permet considerar la flexibilitat de les cadenes laterals dels aminoàcids de la proteïna per tal d'adaptar-se millor al lligand. Aquesta opció augmenta també el temps de càlcul de la ΔG i predicció de l'orientació del lligand. Per tal d'avaluar els efectes d'aquesta opció, s'han realitzat els mateixos *dockings* mitjançant la opció del *Swissdock* de flexibilitat.

Els resultats d'aquest estudi, mitjançant la opció flexibilitat es mostren a continuació:

| | $\Delta G(\text{kcal/mol})$ | Superposició? |
|--------------------|-----------------------------|---------------|
| Clúster 001 | -11,37 | SÍ |
| Clúster 002 | -13,62 | SÍ |
| Clúster 003 | -10,92 | NO |
| Clúster 004 | -12,82 | SÍ |
| Clúster 005 | -11,82 | NO |

Taula 4. Resultats del *docking* del GPP a les diferents conformacions obtingudes del *clustering* de la simulació de la proteïna apo amb la opció "flexibility" del *Swissdock*. S'indica l'energia d'interacció en kcal/mol i si existeix superposició amb la regió del centre actiu/al·lostèric.

Pel que fa als resultats obtinguts mitjançant el *docking* dels 5 clústers realitzats amb flexibilitat, es pot observar que els valors d' ΔG s'han mantingut en la majoria de casos en comparació a la Taula X. Mentre que la superposició amb el centre actiu és exactament igual en els dos casos anteriors. Per tal de comprovar tots els valors a partir dels resultats obtinguts per les tres diferents opcions; *very fast*, *accurate*, i *flexibility*, s'ha realitzat un diagrama amb els 3 clústers en que sí hi ha superposició amb el centre actiu/al·lostèric, és a dir, el clúster 001, el clúster 002 i el clúster 004.



Figura 11 Diagrama de barres que mostra la comparació dels resultats del *docking* en "very fast", "accurate" i "flexibility". Valors d'energia d'interacció en Kcal/mol.

Mitjançant aquest diagrama de barres es pot observar que la distribució segueix un patró igual en tots els casos, és a dir, tant per els models *very fast*, *accurate* i *flexibility*, el clúster 002 és el que té una energia més negativa, i per tant que la interacció amb el substrat és la més favorable, en canvi, en tots els casos, el clúster 001 és el menys negatiu i per tant el que interacciona més dèbilment amb el substrat. A partir d'aquest punt tots els resultats presentats són obtinguts amb l'opció *very fast*.

Per tal d'entendre millor com interacciona el substrat dins la proteïna en cada un dels clústers, a continuació, es mostra la conformació que pren la proteïna i com es col·loca el substrat dins aquesta.

En primer lloc es mostra el clúster 001:

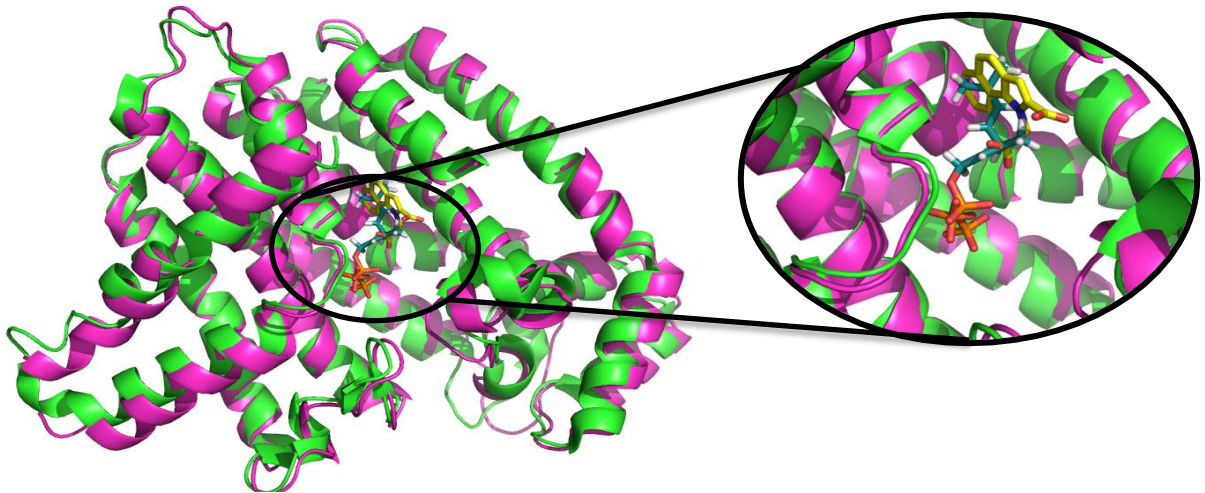


Figura 12. Estructura del clúster 001 (rosa) comparada amb l'estructura cristal·logràfica 3N6K (verd)

A la Figura 12 es pot veure l'estructura cristal·logràfica representada en verd, mentre que sobreposada, en lila, es mostra l'estructura corresponent al clúster 001. Es pot observar que en aquest cas, a diferència de l'apartat anterior la cadena carbonada del GPP es col·loca en el centre al·lostèric on es troba l'inhibidor BFH, en canvi, el centre actiu, on es troba el PO_4^{3-} , només està ocupat pel grup pirofosfat del GPP.

Pel que fa a la resta d'estructures en què el substrat, GPP, es superposa amb el BFH, a continuació es mostra una imatge que mostra com es col·loca aquest en la proteïna (Figura 13).

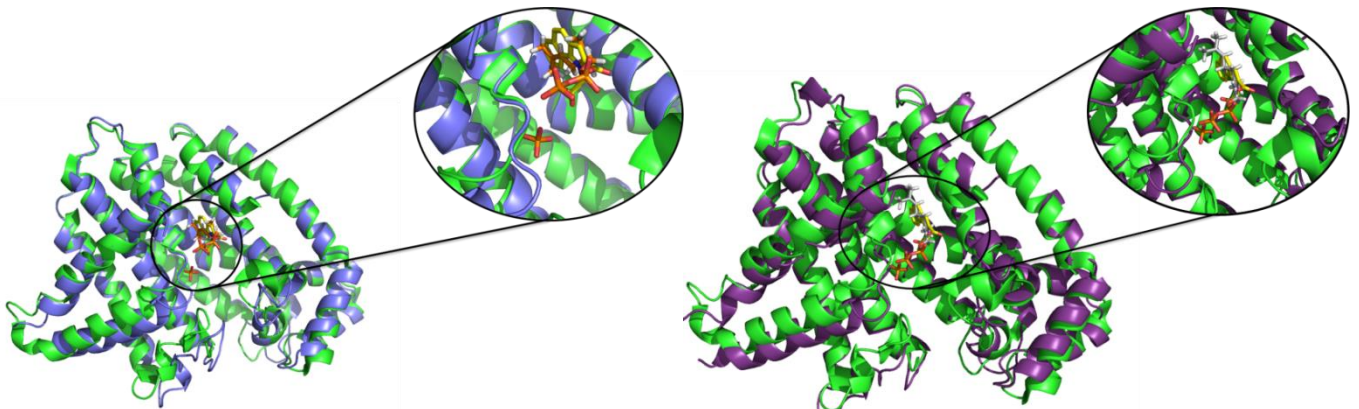


Figura 13. Comparativa de l'estructura cristal·logràfica (verd) amb l'estructura del clúster 002 (lila) i 004 (violeta)

Es pot observar que en els dos casos, clúster 002 i clúster 004, la cadena carbonada del substrat, marcada en groc, es col·loca en el centre al·lostèric de la proteïna.

Per tant, a diferència de les estructures triades aleatòriament en què la cadena del GPP es disposava correctament dins el centre actiu, en aquest cas, en les estructures escollides mitjançant el *clustering* el substrat s'ha col·locat ocupant el centre al·lostèric de la proteïna. Això fa que, igual que per l'estructura cristal·logràfica, aquestes estructures no siguin les més adequades per a predir la interacció entre el substrat i la proteïna. Per tant, el *clustering* de la simulació *apo* en conformacions representatives no ens ha apartat conformacions amb el centre actiu obert. Per tal de millorar aquest fet, caldria generar més clústers per veure si s'obtenen conformacions representatives de la simulació amb el centre actiu obert.

A continuació, en la Figura 14, es mostra una il·lustració pel clúster 003, en que el substrat no es solapa amb el BFH, i per tant, el GPP no entra al centre actiu de la molècula. Igual passa amb el clúster 005.

Si observem la Figura 14, es pot veure com el substrat no ha entrat al centre actiu, això és degut a la flexibilitat de la proteïna i que en aquest cas, els *loops* marcats en la imatge estan desplaçats respecte la estructura cristal·logràfica, fent que el substrat no tingui espai per entrar al centre actiu de la proteïna.

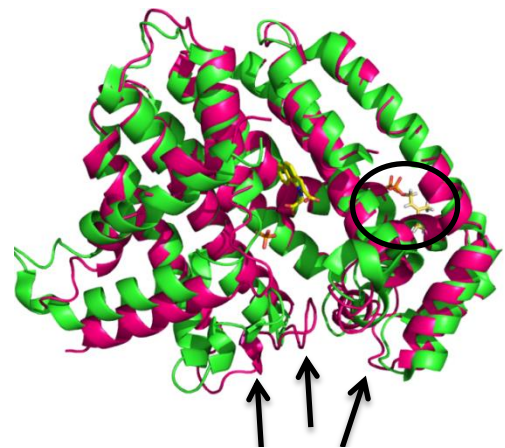


Figura 14. Comparativa entre l'estructura del clúster 003 (vermell) i l'estructura cristal·logràfica (verd)

Docking GPP holo:

Anteriorment s'ha demostrat que la proteïna és més flexible en estat *apo* que en estat *holo*. La selecció d'estructures al llarg de la simulació *apo* ha permès identificar conformacions poc freqüents on el centre actiu està obert per a què el GPP s'hi col·loqui. Tot seguit, s'analitza quin efecte té pel centre actiu realitzar la simulació amb el lligand BFH ocupant el centre al·lostèric. Si l'inhibidor BFH és capaç de mantenir tancat el centre actiu, el *docking* del GPP no podrà situar la cadena carbonada dins el centre actiu.

Un cop realitzada la simulació en l'estat *holo*, igual que en el cas anterior es tracte d'una simulació de 200 ns, de la qual se n'han extret 1000 conformacions. Analitzar 1000 conformacions suposaria molt de temps, per tant, cal realitzar una tria de les conformacions més rellevants de la simulació. Mitjançant el programa Gromacs (veure metodologia) s'obtenen 4 estructures, aquestes representen bastant bé les conformacions més representatives de la simulació.

A cada una d'aquestes estructures se li ha realitzat un *docking* amb el substrat de la proteïna, el GPP, per saber com interacciona i en quina posició es col·loca dins la molècula.

A continuació es mostren els resultats del *docking* de les quatre estructures obtingudes del programa Gromacs:

| | $\Delta G(\text{kcal/mol})$ | Superposició? |
|------------------|-----------------------------|---------------|
| Clúster 1 | -13,93 | NO |
| Clúster 2 | -15,97 | SÍ |
| Clúster 3 | -13,18 | NO |
| Clúster 4 | -14,15 | SÍ |

Taula 5. Resultats del *docking* de les diferents conformacions obtingudes del *clustering* de la simulació de la proteïna *holo* amb la opció "very fast" del *Swissdock*. S'indica l'energia d'interacció en kcal/mol i si existeix superposició amb la regió del centre actiu/al·lostèric

Com es pot veure en la taula anterior, només en dos casos (clúster 2 i clúster 4) hi ha una superposició del substrat amb la zona del centre actiu/al·lostèric, això vol dir que només en aquests dos casos, el GPP es col·loca en el centre actiu o al·lostèric de la molècula. Cal remarcar, que en el cas del clúster 2, on el GPP es col·loca en el centre actiu o al·lostèric, l'increment d'energia de Gibbs pren un valor molt negatiu, per tant mostra que existeix una gran interacció entre el GPP i la proteïna.

A continuació es mostra una imatge on es mostra el centre actiu del clúster 2 per tal d'analitzar amb quins aminoàcids interacciona, per tant, quins són els aminoàcids importants per explicar aquesta interacció tant forta.

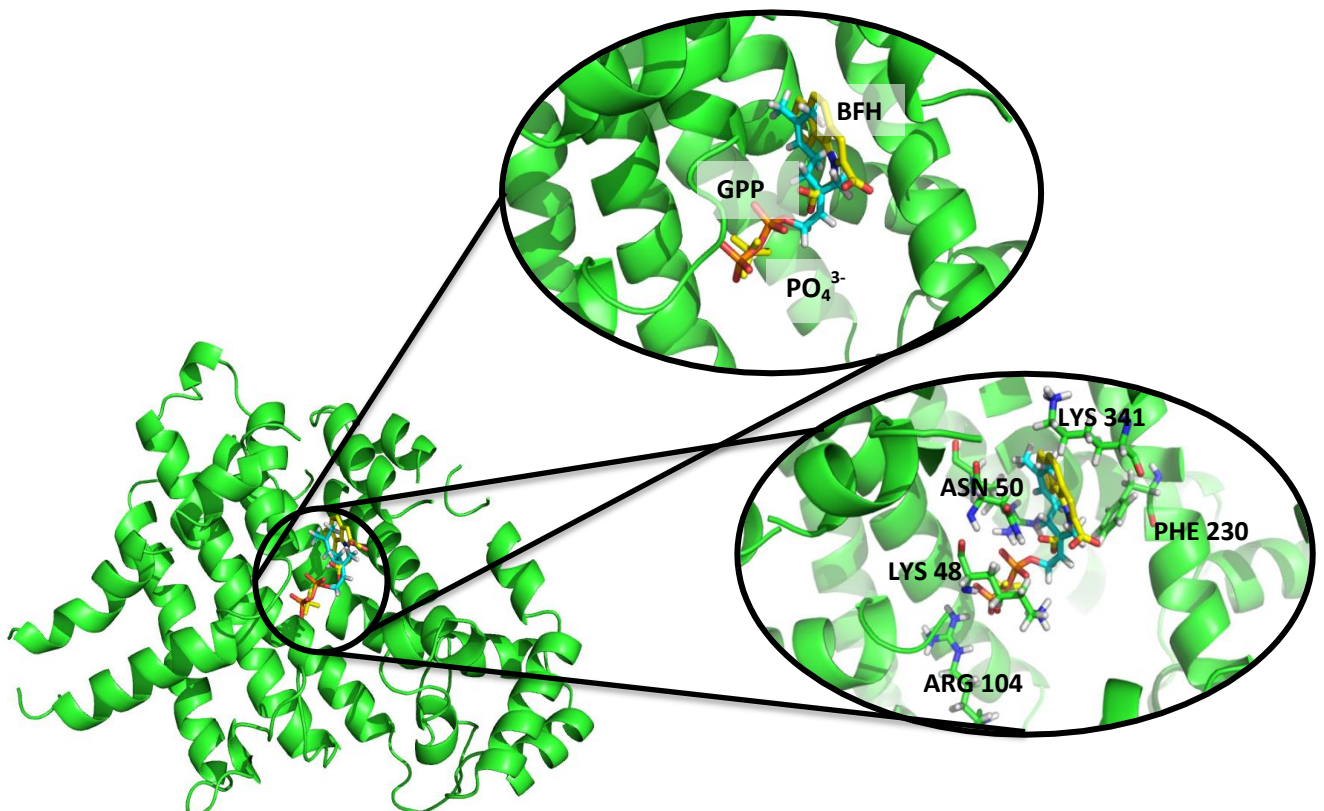


Figura 15. Representació de la posició del GPP (blau) en el *docking* de la proteïna en forma *holo* el clúster 2. Els aminoàcids més rellevants per descriure la interacció estan representats en verd i el BFH i el PO₄³⁻ en groc.

A la Figura 15 es mostra, en primer cas una ampliació del centre actiu de la molècula, mentre que la segona ampliació es mostra el centre actiu amb tots els aminoàcids importants que formen aquest centre actiu.

Si observem la primera ampliació corresponent al clúster 2 es pot veure en groc el lligand BFH i el PO_4^{3-} , mentre que en blau i taronja es mostra el GPP, la part blava correspon a la cadena carbonada mentre que la part taronja correspon al pirofosfat. Com es pot observar el pirofosfat ocupa el mateix espai que el fosfat, és a dir, el centre actiu de la proteïna. Pel que fa a la cadena alifàtica del substrat, es col·loca ocupant el centre al·lostèric de la proteïna, és a dir, el mateix espai que el BFH en l'estructura cristal·logràfica. La posició del GPP és clau per explicar aquest valor tant negatiu d'energia d'interacció, i per tant, el motiu pel qual el *docking* prediu una interacció tant forta. Per explicar aquesta interacció forta cal mirar la segona ampliació on es mostren els aminoàcids presents al centre actiu de la molècula i per tant amb els que interacciona el substrat. En la segona ampliació es pot observar que els aminoàcids són pràcticament els mateixos que en l'estructura cristal·logràfica. Primer de tot cal observar la posició de l'arginina 104, aquest aminoàcid presenta una càrrega positiva a la part de la cadena nitrogenada, marcada en blau. Aquesta càrrega positiva interacciona amb les càrregues negatives del grup pirofosfat del substrat mitjançant una interacció electrostàtica. Aquest enllaç és molt fort degut a l'atracció entre les càrregues positiva i negativa de les dues molècules.

L'asparagina 50 també és un aminoàcid clau per explicar la gran interacció que dona el GPP en aquest clúster, aquest és important per la proximitat que té aquest amb el substrat (1.2 Å). Si mirem l'estructura de l'asparagina, podem veure que conté acceptors d'enllaç d'hidrogen, com els carbonils, però també té donadors d'enllaç d'hidrogen com les amines. En aquest cas la interacció principal és un enllaç d'hidrogen entre el grup amina de la cadena lateral de l'asparagina i els oxígens del grup pirofosfat

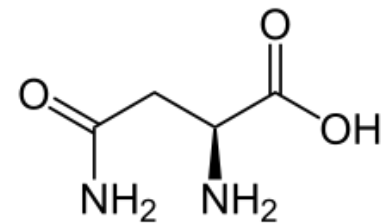


Figura 16. Estructura de l'aminoàcid asparagina

del GPP. Aquest enllaç es dona entre l'oxigen 7 del pirofosfat i el grup NH_2 de l'asparagina.

Per altre banda, les lisines 341 i 48 realitzen interaccions hidrofòbiques amb la cadena alifàtica del GPP.

Un altre tipus d'interacció, és la que fa amb l'aminoàcid, fenilalanina 230, en aquest cas es produeix una interacció hidrofòbica, que estabilitza la cadena carbonada del substrat.

A partir de la dinàmica molecular de la forma *holo* de la proteïna, s'ha pogut obtenir una estructura que mostra una interacció similar a la que presenta l'estructura cristal·logràfica, i per tant aquesta estructura podria ser la òptima si es volgués dissenyar un nou inhibidor.

Cal remarcar, que en aquest cas, el clúster 2 es troba en la forma *holo* i per tant, amb un inhibidor, BFH al centre al·lostèric, això fa que el centre actiu de la proteïna quedi parcialment tancat i per tant que la cadena del substrat ocupi el centre al·lostèric, igual que en el cas de la estructura cristal·logràfica, per tant tot i que

el valor d'energia de Gibbs és favorable, no es tracta d'una bona estructura per a dissenyar fàrmacs que es col·loquin al centre actiu de la molècula, però sí que seria una bona estructura per a dissenyar un fàrmac que es col·loqui al centre al·lostèric.

En resum, el clúster 2 obtingut a partir de la simulació *holo* és l'estructura més adequada per a realitzar *docking* amb inhibidors al·lostèrics mentre que la conformació 611 de la simulació *apo* és la més adequada per a predir molècules que interaccionin amb el centre actiu de la proteïna FPPS.

4.2.2 *Docking* de l'inhibidor al·lostèric BFH

Per tal d'entendre millor el mecanisme d'acció de l'inhibidor al·lostèric BFH i de cares a predir una estructura que sigui bona per a predir futurs fàrmacs per al centre al·lostèric, s'ha realitzat un estudi de *docking* amb l'inhibidor que se sap que es col·loca al centre al·lostèric de la proteïna, en aquest cas s'ha triat l'inhibidor al·lostèric BFH, que com ja s'ha comentat a la introducció, es tracta d'un inhibidor de la proteïna FPPS i que no presenta els efectes secundaris associats als fàrmacs bifosfonats.

***Docking* BFH *apo*:**

Per a la realització dels *dockings* s'han escollit les mateixes estructures que en l'apartat anterior; primer s'ha realitzat una dinàmica molecular amb la proteïna en forma *apo*, i s'han triat 5 estructures de dues maneres diferents: a l'atzar a partir de la visualització de la trajectòria de dinàmica molecular i mitjançant *clustering* de la simulació.

A més s'ha realitzat una segona dinàmica molecular de la proteïna FPPS en la forma *holo*, on el BFH es troba dins la proteïna, en el centre al·lostèric i s'han extret conformacions rellevants. Un cop obtingudes aquestes estructures s'ha realitzat el *docking* mitjançant el programa *Swissdock*, però en aquest cas amb l'inhibidor al·lostèric BFH com a lligand.

Abans de d'analitzar les dades extretes del *docking* en *apo* i *holo*, cal comentar que també s'ha realitzat el *docking* amb el BFH amb l'estructura cristal·logràfica de la proteïna FPPS 3N6K.

Com era d'esperar, el *docking* ha predit que el lligand es col·loca al centre al·lostèric, molt proper al centre actiu, mentre que el valor d' ΔG ha estat de -10.36 kcal/mol, aquest valor ens servirà per comparar amb els resultats de les conformacions *apo* i *holo*.

A continuació es mostren els resultats obtinguts del *docking* de les estructures escollides a l'atzar i escollides mitjançant el *clustering* a partir de la simulació *apo*:

| | $\Delta G(\text{kcal/mol})$ | Superposició? |
|------------------------|-----------------------------|---------------|
| Conformació 182 | -9.93 | NO |
| Conformació 391 | -10.53 | NO |
| Conformació 611 | -9.39 | NO |
| Conformació 785 | -8.98 | NO |
| Conformació 978 | -10.3 | NO |
| Clúster 001 | -10.29 | SÍ |
| Clúster 002 | -10.57 | SÍ |
| Clúster 003 | -8.36 | NO |
| Clúster 004 | -9.5 | NO |
| Clúster 005 | -9.35 | NO |

Taula 6. Resultats del *docking* del BFH a les diferents conformacions obtingudes de la simulació de la proteïna *apo* amb la opció "very fast" del *Swissdock*. S'indica l'energia d'interacció en kcal/mol i si existeix superposició amb la regió del centre actiu/al·lostèric

Com es pot observar en la Taula 6, a les conformacions escollides a l'atzar de la simulació, en cap cas existeix superposició amb el centre actiu o al·lostèric. Per a valorar si hi havia superposició o no, s'ha comparat la posició del BFH que prediu el programa de *docking* respecte la posició inicial de l'estructura cristal·logràfica, és a dir, el centre al·lostèric de la proteïna. A continuació es mostra l'exemple d'una de les conformacions on es pot veure que el BFH no entra al centre al·lostèric de la proteïna.

A la Figura 17 s'hi ha afegit el BFH, marcat en color magenta, en la posició inicial de l'estructura cristal·logràfica, per tal de poder comparar respecte la posició que prediu el *docking*, la posició més favorable del BFH per la conformació 182 que prediu el *docking* està marcada en blau.

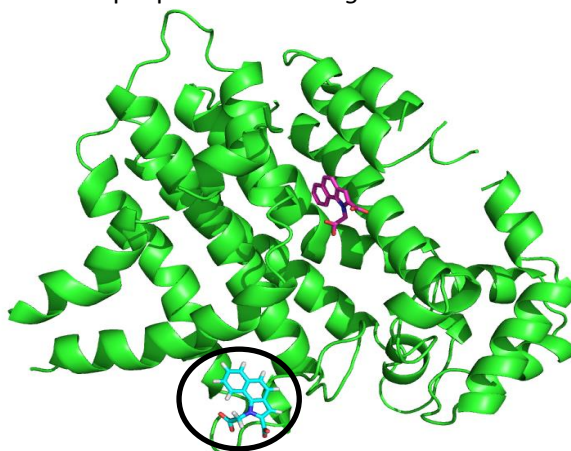


Figura 17. Estructura de la conformació 182 de la forma *apo* de la proteïna FPPS, BFH corresponent a l'estructura cristal·logràfica en magenta i l'obtingut a partir del *docking* en blau

Com es pot veure en la Figura 17, la posició que prediu el *docking*, encerclada en negre, no correspon al centre al·lostèric ni actiu de la proteïna. Aquest fet, determina que quan la proteïna pren la forma *apo*, durant la dinàmica molecular, aquest centre al·lostèric també s'ha tancat parcialment, i que les conformacions triades a l'atzar, en tots els casos, el centre al·lostèric de la proteïna, ha pres una conformació tancada. Aquest fet lliga amb l'obertura del centre actiu observada per aquestes mateixes conformacions al fer el *docking* amb el substrat GPP.

Mitjançant el *docking* de les estructures escollides mitjançant el mètode de *clustering* que s'ha realitzat de la proteïna, en que s'han escollit les 5 estructures més representatives, podem veure si les estructures escollides a l'atzar són representatives o no pel centre al·lostèric. Anteriorment hem vist que aquestes estructures no són representatives per observar l'obertura del centre actiu.

A simple vista, ja podem veure que milloren les conformacions analitzades a l'atzar, ja que mitjançant el *docking* dels 5 clústers, en dos dels cinc casos, el lligand BFH es col·loca correctament al centre al·lostèric de la molècula. A continuació es pot observar una imatge del clúster 001, que mostra la posició que prediu el programa de *docking* del lligand a la proteïna.

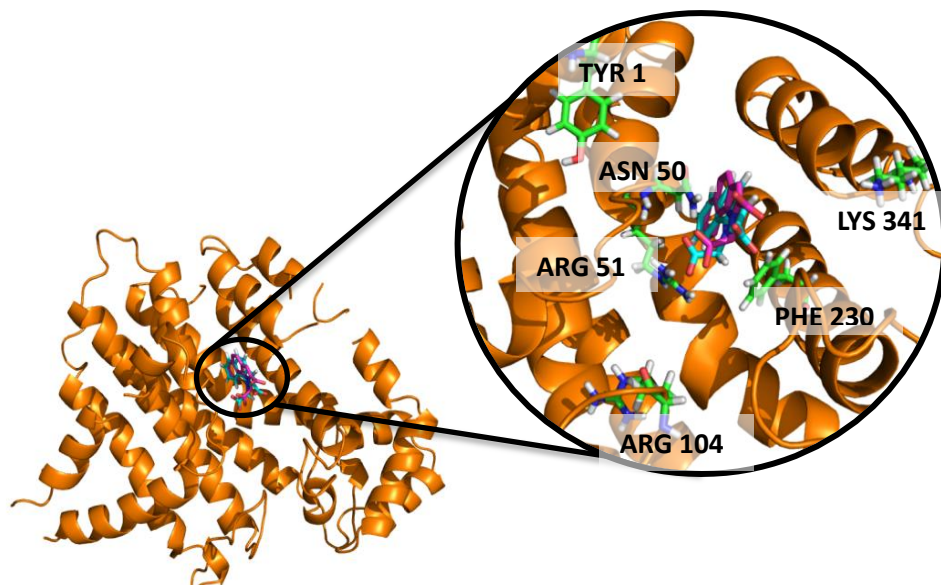


Figura 18. Estructura del clúster 001 i representació dels aminoàcids del centre al·lostèric. El BFH corresponent a l'estructura cristal·logràfica en magenta i el l'obtingut amb el *docking* en blau

Igual que en el cas anterior, el BFH de l'estructura cristal·logràfica, marcat en magenta, només s'hi mostra per tal de comparar la posició respecte la que prediu el *docking* del lligand. Per tant, si observem el BFH, marcat en blau, podem observar que la superposició és pràcticament completa i per tant es pot assegurar que aquest es col·loca en el centre al·lostèric de la proteïna. Així doncs el clúster 1 seria una bona estructura pel disseny de fàrmacs al·lostèrics ja que dóna uns valors similars als obtinguts pel *docking* del BFH a l'estructura cristal·logràfica.

En verd es troben marcats els aminoàcids amb els que interacciona el BFH al centre al·lostèric.

Principalment interacciona amb l'Asn50 formant un enllaç d'hidrogen entre el carboxilat acceptor d'enllaç d'hidrogen del lligand i l'amina de l'asparagina que és donadora d'enllaç d'hidrogen. Una altra interacció que es dona en aquesta estructura és la interacció hidrofòbica de tipus π -stacking que té lloc entre l'anell de l'aminoàcid Phe230, fenilalanina, i els anells del lligand BFH. Una altra interacció molt important és la que fa amb l'aminoàcid Arg51, aquesta arginina, fa enllaç d'hidrogen entre l'amina del seu extrem amb el carboxilat del lligand. Cal destacar la posició de l'arginina 104 que es troba molt lluny i que no interacciona amb el lligand, més endavant es parlarà de la importància d'aquest aminoàcid.

El mateix passa amb el clúster 002, que a més són en els que s'han obtingut uns resultats d'energia més negatius, per tant, on la interacció és més forta. En canvi, els altres tres clústers (003, 004 i 005), la posició del BFH no es correspon amb el centre al·lostèric de la proteïna igual que en el cas en què s'han pres les estructures aleatòriament.

Com ja hem dit, mitjançant els resultats obtinguts de la dinàmica molecular de la proteïna en la forma *apo*, la conformació del centre al·lostèric de la proteïna es troba tancat, excepte en les conformacions representades pels clústers 001 i 002. Per tant, es pot dir que en global les estructures extremes de la simulació en la forma *apo* de la proteïna no són prou bones per si es volgués dissenyar un inhibidor al·lostèric.

Docking BFH holo:

Per tal de buscar una estructura de la proteïna en que tingui el centre al·lostèric prou obert com per dissenyar un inhibidor al·lostèric, s'ha realitzat la dinàmica molecular en la forma *holo* de la proteïna. La simulació amb la forma *holo* s'ha realitzat col·locant al centre al·lostèric. Un cop obtingudes les 4 estructures més representatives de la dinàmica molecular mitjançant el programa Gromacs, s'ha realitzat el *docking* amb el BFH, per tal d'avaluar on predir el programa de *docking* que es col·locarà aquest lligand.

Els resultats del *docking* es mostren a continuació en forma de taula:

| | $\Delta G(\text{kcal/mol})$ | Superposició? |
|------------------|-----------------------------|---------------|
| Clúster 1 | -11,47 | SÍ |
| Clúster 2 | -11,76 | SÍ |
| Clúster 3 | -10,75 | SÍ |
| Clúster 4 | -10,44 | SÍ |

Taula 7. . Resultats del *docking* del BFH a les diferents conformacions obtingudes de la simulació de la proteïna *holo* amb la opció "very fast" del *Swissdock*. S'indica l'energia d'interacció en kcal/mol i si existeix superposició amb la regió del centre actiu/al·lostèric

Com es pot veure en la taula anterior, en la forma *holo* de la proteïna, en tots els clústers existeix una superposició, és a dir, en tots els casos el lligand es col·loca en el centre al·lostèric de la proteïna, a més si observem els valors de l'energia d'interacció, observem que són valors més alts que els obtinguts per a l'estructura cristal·logràfica i per a les conformacions extremes de la simulació *apo*.

A continuació es mostra una imatge del clúster 1 en què es pot veure com es col·loca el lligand BFH al centre actiu de la proteïna a més dels aminoàcids més importants amb els quals interacciona.

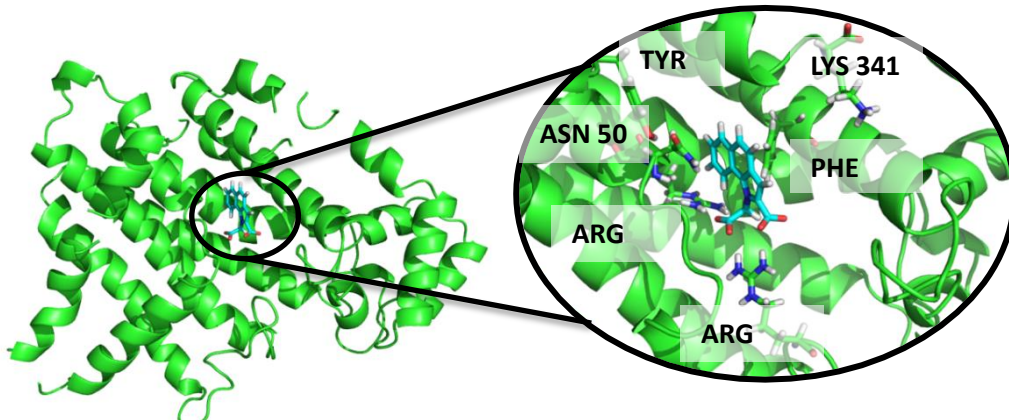


Figura 19. Estructura del clúster 2 i representació dels aminoàcids del centre al·lostèric de la proteïna FPPS

En la imatge es pot veure, marcat en blau, el lligand BFH que es col·loca en el centre al·lostèric de la proteïna, al voltant té una sèrie d'aminoàcids amb els quals interacciona. Igual que en l'apartat anterior, fa enllaç d'hidrogen amb l'arginina 51, amb la asparagina 50 i π -stacking amb la fenilalanina 230. Però en aquest cas, a diferència de l'estructura cristal·logràfica el grup guanidina de la cadena lateral de l'arginina 104 es troba molt propera al lligand, fent que es doni interacció per enllaç d'hidrogen. És important aquesta interacció ja que com es pot veure clarament a la imatge, l'Arg104 s'orienta cap al lligand, en canvi en l'apartat anterior, s'ha vist com l'arginina no s'orientava cap al lligand sinó que quedava allunyada. Per tant, es pot concloure que és aquesta interacció que fa que disminueixi el valor d'energia de Gibbs fent així augmentar la seva interacció. Aquest fet passa en tots els 4 clústers de la simulació *holo*.

A continuació es mostra un gràfic on es pot veure el canvi de posició de l'arginina 104 en la simulació *apo* i *holo*. La distància s'ha mesurat respecte l'aminoàcid Asn50, ja que és un aminoàcid molt proper al centre al·lostèric i és molt poc mòbil, així la diferència de distància que es pot veure a la Figura 21 és degut al desplaçament de l'Arg104:

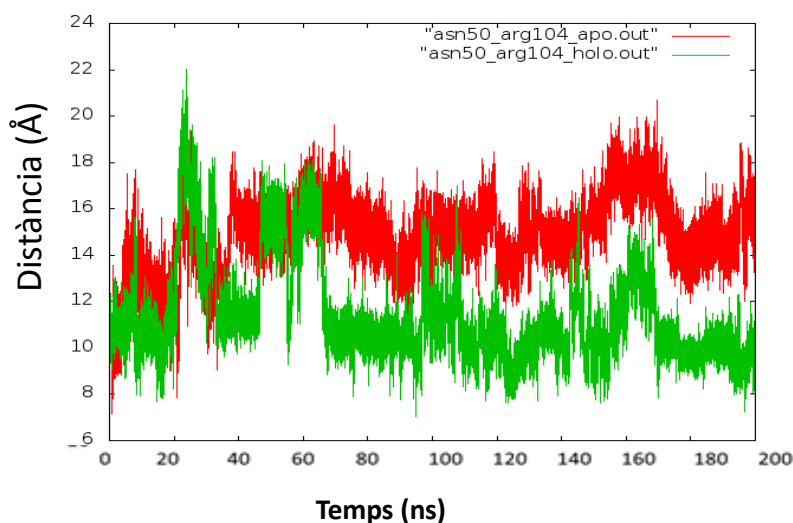


Figura 20. Gràfic que mostra la distància entre el carboni alfa de l'arginina 104 i el carboni alfa de la asparagina 50 en les simulacions apo i holo

En vermell es pot veure les distàncies al llarg de la dinàmica molecular de la proteïna en la forma *apo*, a l'inici de la simulació, la distància oscil·la entre els 10 i 16 àngstroms de distància entre el C α de l'arginina 104 i el C α de l'asparagina 50, en canvi a mesura que avança la simulació, la distància es manté més o menys igual i fins i tot en algun punt la distància s'allarga. Per tant, es pot concloure que l'arginina 104 en cap moment de la simulació *apo* s'apropa al centre al·lostèric de la proteïna.

Per altre banda en verd es pot veure les distàncies entre l'arginina 104 i l'asparagina 50 al llarg del recorregut de la simulació holo. A l'inici de la simulació, les distàncies són més o menys igual que les de la simulació *apo*, però a mesura que avança la simulació es pot veure que la distància disminueix fins a valors propers a 6 àngstroms. Si analitzem la mitjana i desviació estàndard d'aquesta distància al llarg de la simulació, s'obté un valor de 14.88 ± 1.65 Å per a la forma *apo* i 11.43 ± 2.17 Å per a la *holo*. Per tant, es pot concloure que l'arginina al llarg de la simulació en *holo* s'apropa al centre al·lostèric.

Cal destacar que la distància en holo de l'arginina respecte el lligand BFH és molt més petita que la distància amb l'asparagina, en canvi, la comparació entre les simulacions *apo* i *holo* s'ha fet amb l'asparagina ja que en la forma *apo* de la proteïna, el BFH no hi és i per tant no s'ha pogut fer la comparació.

Aquest fet, ens mostra que en la simulació *holo*, com era d'esperar, el centre al·lostèric es troba obert ja que totes les simulacions extremes de la dinàmica molecular, el BFH es col·loca al centre al·lostèric, és per això que es pot assegurar que qualsevol d'aquestes 4 estructures, sobretot les dues primeres degut a que tenen l'energia de Gibbs més negativa, serien bones estructures per predir o buscar un nou inhibidor al·lostèric per a la proteïna FPPS.

5 CONCLUSIONS

FPPS is an interesting anticancer and antibacterial target that participates in the mevalonate pathway. Inhibiting this enzyme can lead to the apoptosis of the cancerous cell or bacteria. Bisphosphonate compounds were the only known drugs active for this protein but recently novel non-bisphosphonate inhibitors that target the allosteric FPPS have been proposed. However, the mechanism of action of these allosteric drugs was still poorly understood. Understanding the structure and flexibility of proteins is of capital importance for improving the design of new inhibitors.

In this work, we have first studied the flexibility of the FPPS protein in the *apo* and *holo* states by means of molecular dynamics simulations. From the simulations in the *apo* state we can conclude that FPPS is a very flexible protein that alters the conformation from an open to a closed state. Molecular dynamics simulations have shown that the most mobile parts are the N- and C-terminal domains of the protein. We arrived to this conclusion, by analysing the distance between the alpha carbons of the amino acids Tyr1 and Lys341 at the beginning and at the end of the protein sequence. This distance goes from 10 to 35 angstroms during the simulation in the *apo* state. However, when BFH is bound to the allosteric site, the protein is significantly less flexible. We also have evaluated the flexibility by analysing the motions in the loops of the protein, to do this we have compared the molecular dynamics of the *apo* and *holo* protein by comparing the position of these loops in the simulations with the position of the loops of the crystallographic structure. This particular flexibility of FPPS and the flexibility of the active/allosteric sites allow the protein to admit molecules of different nature and size, as the substrate GPP and the inhibitor BFH that bind the active site and the allosteric site respectively.

Then, to analyse the effect of FPPS flexibility on the substrate and inhibitor binding, we have extracted selected protein conformations from both molecular dynamics simulations. *Docking* with one substrate (GPP) and one inhibitor (BFH) has been carried out to describe the interaction between the ligand and the protein in different conformations. The first conclusion that we have reached in this part of the work is that the crystallographic structure may not represent the best protein conformation for molecular *docking*. This fact is due to the huge flexibility that FPPS has. The crystallographic structure is a rigid and sometimes poor approach of the protein, and that's why it is important to analyse conformations obtained from molecular dynamic simulations. In general we have observed that when the allosteric site is in the open conformation, the active site is closed, and the ligands occupy the allosteric site. While when the active site is open the ligands prefer to occupy the active site region. For the conformations extracted from the *holo* simulation with the BFH in the active site the active site remains closed and the substrate cannot be positioned in the proper orientation. Second we have performed the *docking* with different options of the *Swissdock* program: very fast, accurate, and flexibility. The results of these studies are very similar, and the differences between them are not significant. The conclusion is that with the very fast option of the *Swissdock* we can

perform the complete study.

Finally, our ultimate goal was to identify the best protein conformation to design new inhibitors for the FPPS protein. If we want to design an active site drug that inhibits the FPPS protein we should select the conformation 611 from the *apo* simulation of the protein. This fact can be explained because the orientation of the GPP in the active site of this particular structure presents the correct orientation. On the other hand, if we want to design an allosteric site inhibitor for the FPPS protein, we should select the conformation corresponding to the cluster 2 from the *holo* simulation. In this conformation the ligand BFH fits in the allosteric site correctly and the amino acids around this allosteric centre interact strongly with it. These important amino acids that interact with the ligand are; Arg104, Arg51, Asn50 and Phe230. The role of Arg104 is particularly important to properly describe the interaction between the allosteric inhibitor and FPPS.

Future work.

The role of Arg104 seems to play a critical role in the allosteric site inhibitor binding. In the crystal structure the interaction between Arg104 and BFH is not observed because Arg104 is interacting with a phosphate ion. Phosphate molecules are sometimes used to stabilize protein structures for crystallization. In this case it is not clear whether this ion plays a critical role for the protein or it is just used in the crystallization process. To understand this fact, we are currently running molecular dynamics simulations with the BFH in the allosteric site and the phosphate ion in the active site. These results will give us information about the role of Arg104 into the inhibitor binding process.

6. BIBLIOGRAFIA

1. Goldstein JL, Brown MS. Regulation of the mevalonate pathway. *Nature*. 1990;343:425-430. doi:10.1038/343425a0.
2. Kavanagh KL, Guo K, Dunford JE, et al. The molecular mechanism of nitrogen-containing bisphosphonates as antiosteoporosis drugs. *Proc Natl Acad Sci U S A*. 2006;103(20):7829-7834. doi:10.1073/pnas.0601643103.
3. Reid IR, Hosking DJ. Bisphosphonates in Paget's disease. *Bone*. 2011;49(1):89-94. doi:10.1016/j.bone.2010.09.002.
4. Stresing V, Daubiné F, Benzaid I, Mönkkönen H, Clézardin P. Bisphosphonates in cancer therapy. *Cancer Lett*. 2007;257(1):16-35. doi:10.1016/j.canlet.2007.07.007.
5. Jahnke W, Rondeau J-M, Cotesta S, et al. Allosteric non-bisphosphonate FPPS inhibitors identified by fragment-based discovery. *Nat Chem Biol*. 2010;6(9):660-666. doi:10.1038/nchembio.421.
6. De Schutter JW, Zaretsky S, Welbourn S, Pause A, Tsantrizos YS. Novel bisphosphonate inhibitors of the human farnesyl pyrophosphate synthase. *Bioorganic Med Chem Lett*. 2010;20(19):5781-5786. doi:10.1016/j.bmcl.2010.07.133.
7. Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE. Biomolecular simulation: a computational microscope for molecular biology. *Annu Rev Biophys*. 2012;41:429-452. doi:10.1146/annurev-biophys-042910-155245.
8. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. *Nature*. 1977;267(5612):585-590. doi:10.1038/267585a0.
9. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nat Struct Biol*. 2002;9(9):646-652. doi:10.1038/nsb0902-646.
10. Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A. H++: A server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res*. 2005;33(SUPPL. 2). doi:10.1093/nar/gki464.
11. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput*. 2015;11:3696-3713. doi:10.1021/acs.jctc.5b00255.
12. D.A. Case, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K.M. VB. The Amber Molecular Dynamics Package. *AMBER*. 2014;14. <http://ambermd.org/>.
13. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald. *J Chem Theory Comput*. 2013;9(9):3878-3888. doi:10.1021/ct400314y.
14. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun*. 1995;91(1-3):43-56. doi:10.1016/0010-4655(95)00042-E.
15. Meng X-Y, Zhang H-X, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des*. 2011;7(2):146-157. doi:10.2174/157340911795677602.
16. Grosdidier A, Zoete V, Michielin O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res*. 2011;39(SUPPL. 2).
17. Kokh DB, Wade RC, Wenzel W. Receptor flexibility in small-molecule docking calculations. *Wiley Interdiscip Rev Comput Mol Sci*. 2011;1(2):298-314. doi:10.1002/wcms.29.
18. Grosdidier A, Zoete V, Michielin O. Fast docking using the CHARMM force field with EADock DSS. *J Comput Chem*. 2011;32(10):2149-2159.
19. Grosdidier A, Zoete V, Michielin O. EADock: Docking of small molecules into protein active sites with a multiobjective evolutionary optimization. *Proteins Struct Funct Genet*. 2007;67(4):1010-1025.
20. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605-1612. doi:10.1002/jcc.20084.
21. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph*. 1996;14(1):33-38. doi:10.1016/0263-7855(96)00018-5.
22. DeLano WL. The PyMOL Molecular Graphics System, Version 1.1. *Schrödinger LLC*. 2002:<http://www.pymol.org>. doi:10.1038/hr.2014.17.