

Títol del treball: Aplicació i validació d'eines bioinformàtiques per a la predicció d'interaccions proteïna-inhibidor en proteïnes quinasa

Nom estudiant: Bernat Alberch i Casas

Correu electrònic: alberchcasasbernat@gmail.com

Grau en Biotecnologia

Nom del tutor: Marc Garcia Borrás

Correu electrònic: marc.garcia@udg.edu

Data de dipòsit de la memòria a través de la plataforma de TFG: 07/06/2024

AGRAÏMENTS

En primer lloc, vull expressar el meu més sincer agraïment al meu tutor Marc Garcia i en especial a en Ferran Feixas per la seva inestimable ajuda, consells i dedicació al llarg de la realització d'aquest Treball de Fi de Grau. Les seves orientacions han estat fonamentals per al bon desenvolupament d'aquest treball.

També vull dedicar un agraïment a la meua família, parella, companys i amics per la seva companyia i suport emocional que han fet aquest camí més agradable i enriquidor.

I finalment, agrair a totes aquelles persones que d'alguna manera han contribuït a la realització d'aquest treball.

ÍNDEX

RESUM	i
RESUMEN	ii
ABSTRACT	iii
REFLEXIONS SOBRE ÈTICA, SOSTENIBILITAT I PERSPECTIVA DE GÈNERE	iv
1. INTRODUCCIÓ	1
2. OBJECTIVES	5
3. METODOLOGIA	6
3.1. Obtenció i preparació de les estructures de les proteïnes quinasa i dels lligands	6
3.2. Docking molecular	7
3.3. Reconstrucció de proteïnes	9
4. RESULTATS I DISCUSSIÓ	10
4.1. Validació protocols docking molecular per proteïnes quinasa a partir de PDBs coneguts	10
4.2. Validació docking molecular amb la proteïna quinasa completament reconstruïda	19
4.3. Reconstrucció completa de les proteïnes quinasa ancestrals de les quals es desconeix l'estructura cristal·logràfica	22
CONCLUSIONS	28
BIBLIOGRAFIA	29

RESUM

Les proteïnes quinasa participen en diverses vies de senyalització que afecten a la regulació de processos cel·lulars i estan involucrades en diverses malalties, fet que les ha convertit en dianes terapèutiques. Entendre la funció i estructura d'aquestes proteïnes i com interaccionen amb inhibidors pot ajudar al disseny de nous fàrmacs.

Concretament en aquest treball, s'utilitzen eines bioinformàtiques per estudiar la interacció entre les proteïnes quinasa Abl i Src i les seves proteïnes ancestrals amb l'Imatinib, un inhibidor de quinases que serveix per tractar leucèmia mieloide crònica. Alhora, s'analitza la precisió d'aquestes eines per fer càlculs de docking molecular (SwissDock) i per predir estructures 3D de proteïnes com el Swiss-Model i l'AlphaFold3.

Els resultats dels càlculs de docking molecular i la comparació dels modes d'unió de l'Imatinib amb com està unit en les estructures cristal·logràfiques (PDBs), han determinat que el SwissDock és un programa acurat per representar les interaccions entre proteïna quinasa i l'Imatinib en les proteïnes Abl, Src i l'ancestre comú (AS).

La reconstrucció d'elements de l'estructura no resolts en els PDBs d'estudi a través de models d'homologia, ha ajudat a entendre el paper clau de l'orientació del motiu DFG (DFG-out o DFG-in) i del bucle d'activació en les interaccions amb l'Imatinib d'aquestes proteïnes quinasa.

Per les proteïnes ancestrals A1, A2 i S1 de les quals es desconeix l'estructura s'ha fet una reconstrucció completa amb el servidor Swiss-Model utilitzant com a "templates" el PDB 4CSV (DFG-out) i amb el PDB 4UEU (DFG-in). Cap de les prediccions ha obtingut un mode d'unió de l'Imatinib correcte en els càlculs de docking molecular. Finalment, també s'han reconstruït les estructures de les proteïnes Src, AS, A1, A2 i S1 amb el servidor d'intel·ligència artificial AlphaFold3. Cap de les prediccions proposades ha obtingut millors resultats en els càlculs de docking molecular amb Imatinib en comparació amb les estructures cristal·logràfiques degut a una mala preorganització del motiu DFG.

En resum, s'observa que actualment és necessari disposar d'estructures cristal·logràfiques per a realitzar càlculs acurats de docking molecular en proteïnes quinasa. Ampliant el ventall de conformacions del DFG en les prediccions o podent afegir el lligand al fer la predicció de l'estructura, l'AlphaFold3 podria esdevenir una eina de gran utilitat en la predicció d'estructures moleculars.

RESUMEN

Las proteínas quinasa participan en varias vías de señalización que afectan a la regulación de procesos celulares y están involucradas en varias enfermedades, hecho que las ha convertido en dianas terapéuticas. Entender la función y estructura de estas proteínas y cómo interaccionan con inhibidores puede ayudar al diseño de nuevos fármacos.

Concretamente en este trabajo, se utilizan herramientas bioinformáticas para estudiar las interacciones entre las proteínas quinasa Abl y Src y sus proteínas ancestrales con Imatinib, un inhibidor de quinasas que se utiliza para tratar leucemia mieloide crónica. A su vez, se analiza la precisión de estas herramientas para hacer cálculos de docking molecular (SwissDock) y para predecir estructuras 3D de proteínas como el Swiss-Model y el AlphaFold3.

Los resultados de los cálculos de docking molecular y la comparación de los modos de unión del Imatinib con cómo está unido en las estructuras cristalográficas (PDBs), han determinado que el SwissDock predice bien las interacciones entre proteína quinasa y el Imatinib en las proteínas Abl, Src y el ancestro común AS.

La reconstrucción de elementos de la estructura no resueltos en los PDBs estudiados a partir de modelos de homología, ha ayudado a entender el papel de la orientación del motivo DFG (DFG-out o DFG-in) y del bucle de activación en las interacciones con Imatinib de estas proteínas quinasa.

Para las proteínas ancestrales A1, A2 y S1 de las cuáles se desconoce la estructura se ha hecho una reconstrucción completa con el servidor Swiss-Model utilizando como "templates" el PDB 4CSV (DFG-out) y PDB 4UEU (DFG-in). Ninguna de las predicciones ha logrado un modo de unión del Imatinib correcto en los cálculos de docking molecular. Finalmente, también se han reconstruido las estructuras de las proteínas Src, AS, A1, A2 i S1 con el servidor de inteligencia artificial AlphaFold3. Ninguna de las predicciones propuestas por el AlphaFold3 ha obtenido mejores resultados en los cálculos de docking molecular con Imatinib en comparación con las estructuras cristalográficas, debido a una mala preorganización del motivo DFG.

En resumen, se observa que actualmente es necesario disponer de estructuras cristalográficas para realizar cálculos certeros de docking molecular en proteínas quinasa. Ampliando la diversidad de conformaciones del DFG en las predicciones o pudiendo añadir el inhibidor para hacer la predicción de la estructura, el AlphaFold3 podría ser una herramienta de gran utilidad en la predicción de estructuras moleculares.

ABSTRACT

Kinase proteins participate in various signaling pathways that affect the regulation of cellular processes and are involved in numerous diseases, making them therapeutic targets. Understanding the function and structure of these proteins and their interactions with inhibitors can help in the design of new drugs.

In this study, bioinformatic tools are used to investigate the interaction between the kinase proteins Abl and Src and their ancestral proteins with Imatinib, a kinase inhibitor used to treat chronic myeloid leukemia. The accuracy of these tools is analysed for molecular docking calculations (SwissDock) and for predicting 3D protein structures, such as Swiss-Model and AlphaFold3.

The results of the molecular docking calculations and the comparison of Imatinib binding modes with its binding in crystallographic structures (PDBs) have determined that SwissDock is an accurate program for representing interactions between kinase proteins and Imatinib in Abl, Src, and common ancestor (AS) proteins.

The reconstruction of unresolved structural elements in the studied PDBs through homology models has helped understand the key role of the orientation of the DFG motif (DFG-out or DFG-in) and the activation loop in the interactions with Imatinib in these kinase proteins.

For the ancestral proteins A1, A2, and S1, whose structures are unknown, a complete reconstruction was performed with the Swiss-Model server using PDB 4CSV (DFG-out) and PDB 4UEU (DFG-in) as templates. None of the predictions obtained a correct binding mode for Imatinib in the molecular docking calculations. Finally, the structures of Src, AS, A1, A2, and S1 proteins were also reconstructed with the AlphaFold3 artificial intelligence server. None of the proposed predictions achieved better results in the molecular docking calculations with Imatinib compared to crystallographic structures due to poor preorganization of the DFG motif.

In summary, it is currently necessary to have crystallographic structures to perform accurate molecular docking calculations in kinase proteins. Expanding the range of DFG conformations in predictions or being able to add the ligand during structure prediction could make AlphaFold3 a very useful tool in molecular structure prediction.

REFLEXIONS SOBRE ÈTICA, SOSTENIBILITAT I PERSPECTIVA DE GÈNERE

REFLEXIÓ SOBRE ÈTICA

Les eines bioinformàtiques comporten moltes avantatges respecte l'experimentació tradicional però també algunes desavantatges. Fent càlculs de docking i veient com interacciona millor un medicament es poden evitar part dels experiments amb animals que es fan durant els assajos clínics de medicaments. Això pot millorar les pràctiques de recerca i evitar conflictes ètics als investigadors/es. En aquest treball s'analitza el potencial i la fiabilitat de l'AlphaFold3, una eina bioinformàtica que pot revolucionar les prediccions d'estructures tridimensionals. Aquest tipus d'eines també poden obrir portes a nous debats ètics com la falsificació de dades i el robatori o mal ús de dades personals dels usuaris d'aquestes, així com de dades genètiques usades com a base per eines com l'AlphaFold3 (Jumper et al., 2021).

Respecte als models d'homologia, hi ha un debat ètic sobre el biaix experimental que comporta afegir directament una proteïna model en la que eines com el Swiss-Model es basen per reconstruir una proteïna. Definitivament, les eines bioinformàtiques comporten millores a nivells ètics però també obren altres dilemes que han de ser resolts.

REFELXIÓ SOBRE SOSTENIBILITAT

A nivell de sostenibilitat, les eines bioinformàtiques clarament comporten millores respecte l'experimentació tradicional. Es gasten menys recursos, generant un menor impacte ambiental i contribuint al compliment dels objectius de desenvolupament sostenible i en la lluita contra el canvi climàtic. Alhora, tot i que utilitzant eines bioinformàtiques es generen menys residus que amb una experimentació al laboratori, s'ha de fer una bona gestió d'aquests residus com poden ser un gran ventall de dispositius electrònics, que costen molt de reciclar. També s'ha d'apostar per energies renovables per tal de reduir la petjada ecològica, ja que en càlculs computacionals a gran escala l'energia usada pot ser molta.

REFLEXIÓ SOBRE PERSPECTIVA DE GÈNERE

Històricament, les dones han estat infrarepresentades en camps com la ciència i la tecnologia, i la bioinformàtica i la intel·ligència artificial no en son una excepció (Buslón et al., 2023). Aquesta realitat sumada als fets de la bretxa salarial en el sector i el biaix de gènere que presenten els models d'intel·ligència artificial, han de suposar els principals eixos d'abordatge tant polítics com socials per revertir aquesta situació. Cal obrir un nou marc on els projectes de recerca amb un finançament igualitari siguin una realitat.

1. INTRODUCCIÓ

Les proteïnes quinasa son enzims que juguen un paper clau en molts aspectes de la biologia, participant en diverses vies de senyalització que afecten a la regulació de processos cel·lulars com la proliferació, apoptosi, diferenciació, el sistema nerviós, la resposta immunitària i fins i tot la transcripció o el cicle cel·lular (Manning et al., 2002). Aquests enzims funcionen transferint un grup fosfat de la molècula d'adenosinatrifosfat (ATP) a tres possibles aminoàcids específics d'una proteïna substrat (serina, treonina o tirosina), i ho fan mitjançant la següent reacció (Roskoski, 2016):



Aquesta fosforilació dona lloc a una classificació d'aquests enzims en dos grans grups, les proteïna-serina/treonina quinasa i les proteïna-tirosina quinasa. Aquest procés de fosforilació altera l'activitat, la localització o la interacció de la proteïna substrat, modulant així la seva activitat i per tant diverses vies de senyalització. És per això que les proteïnes quinasa s'han convertit en una diana pel desenvolupament de nous medicaments, ja que aquestes regulacions es donen en els àmbits de moltes malalties com càncer, diabetis, i problemes cardiovasculars, nerviosos i autoimmunes.

Segons Manning et al. (2002), es van identificar 518 gens de proteïnes quinasa humanes, el que representa aproximadament un 1.7% de tots els gens humans. D'aquests 518 gens, 385 eren del grup serina/treonina quinases, 90 eren tirosina quinases i 43 quinases que presenten funcionalitat similar a les tirosina quinases però que no es consideren en aquest grup.

Les proteïnes quinasa tenen una estructura relativament complexa que inclou diversos dominis essencials per a la seva funció. Tot i les particularitats de cada cinasa específica, trobem moltes regions i estructures comunes entre elles, tal i com la regió catalítica, que és on succeeix la fosforilació i que està dividida entre el domini N-terminal i el domini C-terminal. El domini N-terminal conté principalment cinc làmines- β (β_1 - β_5) i una α C-hèlix que conté un glutamat que fa un pont salí amb una lisina de la β_3 , que facilitien la catàlisi del procés de fosforilació. La presència d'aquest pont salí dona lloc a la conformació α C-hèlix-in. Aquesta posició de l'hèlix α C és necessària per tal que la proteïna quinasa pugui dur a terme la seva activitat completa, però no és suficient per si sola. Dins el domini C-terminal hi ha quatre làmines- β conservades (β_6 - β_9) de les quals en destaca la β_9 , ja que forma part d'una estructura anomenada bucle d'activació o "activation loop" (AS).

Aquest bucle d'activació és una seqüència d'aminoàcids que pot canviar la seva conformació provocant l'activació o inactivació de l'enzim. En la forma activa de l'enzim, aquest bucle d'activació forma una estructura oberta a la regió catalítica que permet la unió del substrat a la proteïna quinasa. En la forma inactiva, bloqueja l'accés del substrat i/o de l'ATP a la regió catalítica. Per tant, aquest bucle d'activació controla tant la unió del substrat com la catàlisi de la proteïna quinasa (Meharena et al., 2013). A la Figura 1 es mostra l'estructura de la quinasa EGFR, que serveix com a model per la gran majoria de les proteïnes quinasa.

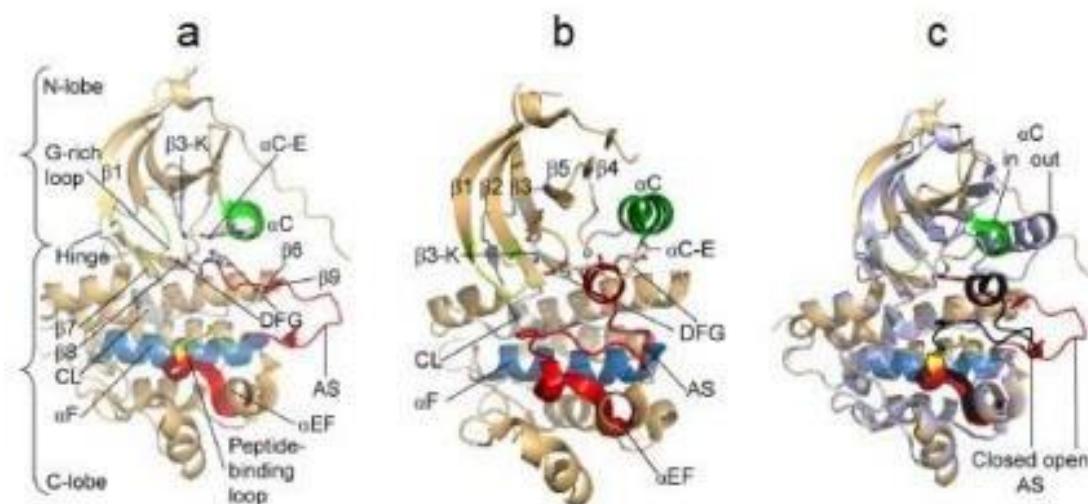


Figura 1: Estructura de la proteïna quinasa EGFR. (a) Conformació activa. (b) Conformació inactiva. (c) Conformacions activa i inactiva superposades (Roskoski, 2016).

La Figura 1 és un exemple gràfic de com s'organitzen aquestes dos conformacions activa i inactiva d'una quinasa, i en especial la Figura 1c, que permet veure-les superposades i poder entendre com varia la unió del substrat o dels inhibidors entre conformacions.

El bucle d'activació, per a la gran majoria de quinases, comença amb un motiu DFG (Asp-Phe-Gly) format per tres residus i acaba amb el motiu APE (Ala-Pro-Glu). Aquest motiu DFG es pot trobar en dues conformacions, "DFG-in" i "DFG-out". La conformació activa de les quinases ve donada generalment per la posició "DFG-in", i és quan l'aspartat d'aquest DFG es troba orientat cap a l'interior del lloc catalític, on pot unir-se al ió Mg^{2+} i coordinar la transferència del grup fosfat de la molècula d'ATP al substrat. En canvi, en la conformació "DFG-out", aquest aspartat es troba lluny del lloc catalític i a més la fenilalanina bloqueja l'entrada de la molècula l'ATP, donant lloc a l'estat inactiu de la quinasa. Per tant, la posició d'aquest motiu DFG és crucial per a l'activitat o inactivitat de moltes proteïnes quinasa, sent

un dels mecanismes més estudiats per a la regulació d'aquestes.

A part de la conformació "in" o "out" tant del bucle d'activació com del DFG-Asp, com també de la importància de la conformació "in" de la α C-hèlix per a la activitat de la quinasa, s'ha de considerar també la importància de la estructura de la espina reguladora, de la qual s'ha relacionat una estructura linear amb la forma activa de les quinases, i una estructura desordenada amb la forma inactiva. En aquest treball no s'entrarà en més detall sobre aquesta espina reguladora ja que no és rellevant pels resultats presentats més endavant.

Com s'ha esmentat abans, aquestes proteïnes quinasa han sigut una diana d'interès mèdic, ja que estan relacionades amb varies malalties. El primer inhibidor de proteïna quinasa aprovat per la FDA va ser l'Imatinib, també anomenat comercialment com a Gleevec, i és usat pel tractament de leucèmia mieloide crònica (LMC). Actualment hi ha 81 inhibidors de proteïnes quinasa aprovats per la FDA, la majoria dels quals per tractaments contra tumors malignes (Blue Ridge Institute for Medical Research, 2024). Moltes d'aquestes molècules petites inhibidores de quinases també són competidores de l'ATP a la fase catalítica inactiva, ja que interaccionen amb el lloc d'unió de l'ATP. Tot i això, actualment també hi ha molt d'interès en desenvolupar inhibidors al·lostèrics en proteïnes quinases que permetin regular l'activitat de la quinasa a través de canvis conformacionals en els motius principals descrits anteriorment.

Els inhibidors de quinases es poden dividir en tres categories. La primera (Tipus I) són les molècules que s'uneixen al lloc d'unió de l'ATP en la forma activa de la proteïna quinasa ("DFG-in" i " α C-hèlix-in"), els de tipus II són els inhibidors que s'uneixen a la forma inactiva de la quinasa (normalment "DFG-out" i " α C-hèlix-out"), i els de tipus III que són els inhibidors al·lostèrics, és a dir que no competeixen amb l'ATP per unir-s'hi, sinó que s'uneixen a un altre lloc que no és el lloc d'unió d'ATP de la quinasa. També es poden trobar altres classificacions on s'afegeixen els tipus I1/2, IV, V i fins i tot VI.

D'aquests tipus afegits amb posterioritat, l'únic que té rellevància per aquest treball és el I1/2, que correspon als inhibidors que s'uneixen a conformacions de les proteïnes quinasa que tenen la α C-hèlix en posició "out" però el DFG-Asp en posició "in".

L'Imatinib, l'inhibidor objecte d'estudi en aquest treball, s'uneix a la proteïna quinasa Abl en la seva forma inactiva, amb el "DFG-D-out", i per tant és classificat com a inhibidor de tipus II. Aquest Imatinib, del qual se'n mostra l'estructura molecular a la Figura 2b, el que fa és estabilitzar la conformació inactiva de l'Abl bloquejant la seva activitat oncogènica.

L'Imatinib s'uneix específicament a la proteïna quinasa Abl humana i també a Src humana

però amb significativament menys afinitat (constant d'inhibició per Abl és $K_i=13$ nM mentre que per Src és $K_i=2000$ nM). Aquestes dues proteïnes presenten diferent funcionalitat, però tenen la seqüència idèntica en un 54% i el bucle d'activació és gairebé idèntic en seqüència i estructura.

Per tant, l'Src té una afinitat entre 200-3000 cops inferior (en funció de com es realitza l'experiment) que l'Abl pel Gleevec (Agafonov et al., 2014). S'ha trobat que Abl i Src presenten dues conformacions del motiu DFG en equilibri (la forma DFG-in i la forma DFG-out). Aquest equilibri està desplaçat cap a la forma DFG-out en el cas d'Abl i això podria ser una explicació per la qual Abl podria presentar una elevada afinitat cap a l'Imatinib.

Wilson et al., 2015, a part de confirmar aquesta menor afinitat del Gleevec (Imatinib) per l'Src, també exploren l'origen de la pèrdua d'afinitat del fàrmac cap a Src i el guany d'afinitat per Abl. Per fer-ho proposen una reconstrucció bioinformàtica de les proteïnes ancestrals tant d'Abl com de Src, arribant a una proteïna ancestral comuna que anomenen AS (que va poder ser cristal·litzada i correspon al PDB 4CSV, aquesta proteïna presenta una constant d'inhibició intermitja de $K_i=910$ nM). Entre les proteïnes Src i Abl modernes i la proteïna quinasa ancestral, també arriben a dos proteïnes ancestrals de l'Abl (A1 i A2) i una per Src (S1) que es troben en una situació intermèdia entre AS i les modernes. A la Figura 2 es mostra l'arbre filogenètic de totes aquestes proteïnes quinasa, conjuntament amb l'estructura de l'Imatinib.

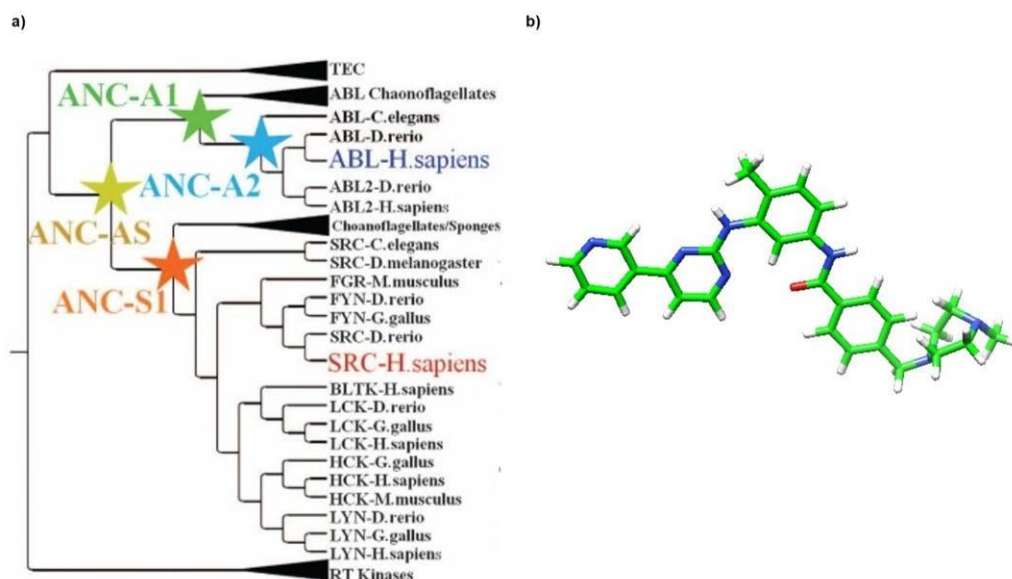


Figura 2: a) Arbre filogenètic de les proteïnes ancestrals d'Abl i Src, amb el seu ancestre comú, extret de Wilson et al., 2015. b) Estructura molecular de l'Imatinib (Gleevec) visualitzada amb el Chimera.

La conclusió principal és que Src ha perdut afinitat per l'Imatinib al llarg de la evolució i que Abl n'ha guanyat degut a canvis en la conformació preferent del motiu DFG. Tot i això, el mecanisme molecular a través del qual el Gleevec mostra una major selectivitat per Abl i AS respecte Src és encara desconegut. Les proteïnes quinasa són proteïnes molt dinàmiques i molt riques conformacionalment, fet que en dificulta la seva caracterització experimental a nivell molecular.

La anàlisi de les interaccions entre les proteïnes quinasa i un determinat fàrmac amb programes bioinformàtics, així com la predicció d'aquestes interaccions i afinitat, poden suposar un avenç molt important a l'hora de determinar sota quin mecanisme concret funciona la proteïna quinasa i com inhibir-la.

Actualment hi ha varis programes de docking molecular que serveixen per fer prediccions de com s'uneixen una proteïna i un lligand, i que permeten entendre millor aquestes interaccions. Tot i això, aquests protocols són encara poc acurats per al disseny de nous fàrmacs, de manera que és molt rellevant elaborar protocols de validació per proteïnes quinasa. La pregunta clau serà determinar quina conformació de la proteïna quinasa serà les que donen una millor predicció en càlculs de docking molecular tant a nivell de poder predir el mode d'enllaç i l'afinitat.

2. OBJECTIVES

Protein kinases are widely studied drug targets that are involved in a large number of biological processes. These proteins are highly dynamic and adopt multiple conformations to develop their function. Understanding the mechanism of action and selectivity of protein kinase inhibitors at the molecular level offer new possibilities for the development of therapeutic strategies for the treatment of diseases. However, the high conformational heterogeneity of protein kinases makes the structural characterization of key functional states a challenging task. Indeed, in the available crystal structures of protein kinases some key regions are unresolved.

This study is focused on exploring the accuracy of various bioinformatic tools to perform predictions of protein-ligand interactions and 3D structure predictions, on the sphere of protein kinases. The main goal is to elucidate the capacities and limitations of these bioinformatic tools, while studying the ancestral reconstruction of modern protein kinases Abl and Src (specified in Figure 2a) and their interaction with a known inhibitor Imatinib (also known as Gleevec). The analysis of these widely studied evolutionary related proteins will allow for a better assessment of the results obtained.

Thus, the following objectives will need to be accomplished in order to reach the main goal:

1. To analyse the main key structure particularities related to protein-ligand interaction of diverse protein kinases with Imatinib, mainly the ones related with the ancestral reconstruction of Abl and Src.
2. To validate the bioinformatic tool SwissDock as a reliable way to perform molecular docking calculations of protein kinases and predict the binding mode of Imatinib, while understanding the details of the interactions between these protein kinases and the drug Imatinib.
3. To bioinformatically reconstruct 3D parts of structures of these family of protein kinases which haven't been determined yet, to then analyse how these reconstructions affect the computational predictions of the binding affinity and binding mode of Imatinib.
4. To predict full protein structures by different methods, including AlphaFold3, a new artificial intelligence tool to reconstruct protein structures with only their sequence. Molecular docking calculations will be performed to determine the accuracy of these full protein structure predictions and their preorganization to interact with Imatinib.

3. METODOLOGIA

3.1. Obtenció i preparació de les estructures de les proteïnes quinasa i dels lligands

Per a l'objecte d'aquest treball ha sigut necessària l'obtenció de moltes estructures conegudes de proteïnes quinasa, així com dels dos lligands utilitzats. També s'han hagut de preparar i adequar aquests fitxers per a tal de poder ser acceptats pel programa SwissDock per poder fer el càlcul de docking molecular.

3.1.1. Protein Data Bank

Tant les estructures cristal·logràfiques en format .pdb com les seqüències FASTA utilitzades durant el treball s'han obtingut de la pàgina web del *Protein Data Bank* (<https://www.rcsb.org>). Es tracta d'una base de dades d'estructures tridimensionals obtingudes experimentalment de diferents molècules biològiques. En aquest treball s'han fet servir PDBs (fitxers obtinguts del *Protein Data Bank* en format .pdb) de diferents proteïnes quinasa, detallats més endavant a l'apartat de resultats i discussió.

3.1.2. PyMol

Un cop obtingudes les estructures de les proteïnes, aquestes s'han analitzat amb el visualitzador molecular PyMol, que permet analitzar les estructures tridimensionals

d'aquestes proteïnes d'interès. El PyMol també permet obrir directament una estructura només coneixent el PDB, sense necessitat de descarregar-la. Amb aquesta eina s'ha visualitzat l'orientació del DFG i dels aminoàcids del centre actiu, i també s'han preparat els PDBs dels quals s'ha fet docking molecular. S'ha eliminat qualsevol lligand que pogués estar unit a la proteïna i s'han afegit hidrògens. A través d'aquest programa també s'han eliminat cadenes repetides de la proteïna si ha fet falta.

3.1.3. Zinc Database

Per aquest treball s'ha utilitzat sobretot el lligand Imatinib (Gleevec), que s'ha obtingut en format SDF a través de la web Zinc Database (versió 15, <https://zinc15.docking.org>). Aquesta és una base de dades oberta de compostos que permet visualitzar la seva estructura amb programes informàtics (Sterling i Irwin, 2015). També s'ha fet servir el lligand derivat d'ATP que conté un metilè extra entre els dos grups fosfats terminals que s'anomena ACP, que s'ha extret també en format SDF del PDB.

Aquests fitxers es converteixen a format MOL2 per poder ser acceptats pel SwissDock, i aquesta conversió s'ha fet directament amb el PyMol.

3.2. Docking molecular

El docking molecular és una predicció computacional de com s'orienta una molècula quan interacciona amb el lloc d'unió d'una macromolècula i de quina afinitat presenta en aquesta orientació, en aquest cas una proteïna quinasa (Bugnon et al., 2024). El procés de docking proposa diverses posicions del lligand respecte la proteïna i les agrupa i ordena en clústers en base a la regió d'interacció, així com també ofereix informació de la energia lliure de Gibbs (ΔG) de la interacció de cada una de les posicions proposades. Aquesta ΔG es pot associar a l'afinitat que presenta una determinada molècula per una regió determinada de la macromolècula. Com més negatiu és el valor de la ΔG més afinitat presenta per la proteïna en qüestió.

3.2.1. SwissDock

Concretament, el programa SwissDock ofereix 256 posicions del lligand possibles per a cada càlcul de docking, i les ordena en clústers i segons la ΔG . El SwissDock ha ofert una nova versió del programa durant la realització d'aquest treball, amb millores respecte l'anterior versió. Han mantingut però la versió antiga en un nou link web (<http://old.swissdock.ch/docking>) i és la que s'ha seguit fent servir en aquest treball per mantenir la coherència dels resultats i no canviar paràmetres entre aquests càlculs de docking.

El SwissDock és gratuït i per fer-lo servir només s'ha d'adjuntar una proteïna "target" en format pdb preparada per al docking (una cadena, amb hidrògens i sense lligand) i un lligand en format mol2, així com un nom pel càlcul. El programa s'ha fet servir amb els paràmetres predeterminats per a tots els càlculs de docking al llarg del treball.

Per tal de validar el protocol de docking molecular, s'ha volgut comprovar l'eficàcia del SwissDock amb els següents PDBs de proteïnes quinasa: 4CSV, 2HYY, 1OPJ, 2OIQ i 4UEU dels quals es coneix l'orientació del lligand, que estan detallats a l'apartat de resultats i discussió.

3.2.2. Chimera

Els resultats del docking molecular s'han analitzat amb el Chimera, un programa que permet visualitzar estructures 3D i que permet fer una anàlisi molt completa de les orientacions del lligand proposades pel SwissDock. El primer pas és obrir el fitxer "target.pdb" que ofereix el càlcul de docking i amb l'eina "view dock" del Chimera obrir el fitxer "clusters.dock4.pdb", que permet veure les 256 orientacions del lligand.

Un dels passos és veure la similitud d'aquestes orientacions amb com està realment orientat el lligand en el PDB de l'estructura cristal·logràfica coneguda. Això s'ha fet de manera visual superposant aquestes dues estructures amb el Chimera, i també amb el càlcul de RMSD, una mesura quantitativa feta amb el mateix Chimera que calcula la diferència de posicions dels àtoms comparant les coordenades d'aquests. L'RMSD s'ha calculat amb la línia de comandes del Chimera escrivint el següent: `rmsd #1.x:lig #2:lig`, on x és l'orientació triada del lligand. Per a tots els resultats de docking l'RMSD s'ha calculat amb el mateix fitxer, que és el 4CSV però amb els àtoms canviats de nom per tal de que el Chimera ho pugui detectar. L'RMSD és un resultat que complementa a la ΔG , que com més negativa, indica una major afinitat entre proteïna i lligand.

En el procés de validació i posterior producció de resultats, s'ha triat sempre la orientació del docking que visualment més s'assembla a la del lligand en l'estructura cristal·logràfica, encara que tingués una ΔG una mica menys negativa que una altra orientació. Quan no hi ha hagut cap orientació que es pugui considerar correcta en base als PDBs coneguts, s'ha triat la orientació amb ΔG més negativa. Després s'han analitzat els ponts d'hidrogen d'aquesta orientació del lligand a través de l'eina "FindHbond" del Chimera, i s'han comparat amb els de l'estructura coneguda quan ha calgut per validar la fiabilitat del càlcul del SwissDock.

Amb el Chimera s'han generat totes les figures de l'apartat de resultats i discussió, i també s'ha fet servir per a alguna preparació de PDBs per al docking, de manera similar al PyMol.

3.3. Reconstrucció de proteïnes

3.3.1. Model d'homologia amb Swiss-Model

Als PDBs amb els que s'han fet càlculs de docking hi ha parts de l'estructura que hi falten, en concret solen faltar els residus corresponents al bucle d'activació que s'inicia amb el motiu DFG. A través dels models d'homologia es poden reconstruir aquestes parts que poden ser importants per a la unió del lligand a la proteïna. El Swiss-Model és un servidor gratuït que genera estructures proteiques en base a models d'homologia (<https://swissmodel.expasy.org>).

S'ha fet servir l'apartat de "user template", on s'afegeix la seqüència de la proteïna que es vol reconstruir en format FASTA i una proteïna model que li serveix al Swiss-Model com a base per construir aquesta nova proteïna, que han sigut els PDBs de les estructures conegudes. Després s'ha tornat a fer el càlcul de docking per veure si aquesta reconstrucció varia d'alguna forma la interacció entre quinasa i lligand.

També s'han reconstruït completament algunes proteïnes intermèdies en l'arbre de proteïnes ancestrals de Wilson et al., 2015 (veure Figura 2a), de les quals només se'n coneix la seqüència, i s'han utilitzat les proteïnes model 4CSV i 4UEU, una amb el DFG-out i l'altra amb DFG-in.

3.3.2. AlphaFold3

Recentment han sortit diverses eines per poder fer una predicció de com és una estructura 3D d'una proteïna (o altres molècules) sense necessitat de fer servir una proteïna coneguda com a model. Aquest és el cas del programa AlphaFold, desenvolupat, entre d'altres, per Google DeepMind, que permet realitzar prediccions suposadament molt acurades de les estructures proteiques a partir de la seva seqüència, obtinguda en format FASTA a través del PDB.

Al maig de 2024 va sortir la versió 3 de l'AlphaFold, que millora l'eficàcia de l'eina substancialment respecte la versió anterior (Abramson et al., 2024). S'ha fet servir la versió web de l'AlphaFold3 (<https://golgi.sandbox.google.com>), amb els paràmetres predeterminats. Aquesta versió permet no descarregar-se l'aplicació i, a data 30/05/2024, fer un màxim de 20 prediccions al dia.

Al fer una predicció, l'AlphaFold3 dona les 5 estructures amb més probabilitat que siguin correctes, i també un mapa amb valors de confiança de la predicció segons les zones de la proteïna (pLDDT).

Es parla de que eines com aquesta podrien suposar una revolució per al món de la bioinformàtica i la investigació, ja que per exemple permetria conèixer com s'uneixen fàrmacs a una proteïna determinada, i dissenyar-ne de nous sense necessitat d'obtenir experimentalment la estructura cristal·logràfica de la proteïna, com s'ha fet fins ara. A l'apartat de resultats s'intentarà comprovar l'eficàcia d'aquest programa veient com pre-organitza diverses proteïnes quinasa, entre elles proteïnes que se'n coneix l'estructura cristal·logràfica per poder fer una comparació, però també d'altres que només se'n coneix la seqüència. Serà important observar com orienta estructures clau com el DFG i veure'n també els valors de confiança de la predicció d'aquestes zones, així com la variabilitat de la posició del DFG ("in" o "out") entre les 5 estructures proposades.

4. RESULTATS I DISCUSSIÓ

4.1. Validació protocols docking molecular per proteïnes quinasa a partir de PDBs coneguts

A la primera fase dels resultats, el que s'ha volgut fer és comprovar com funciona el programa de docking SwissDock per a sistemes proteïna quinasa-ligand amb PDBs coneguts.

D'aquesta manera es podrà validar o posar en dubte com es prediu aquesta interacció entre proteïna quinasa i l'Imatinib a través d'aquesta eina bioinformàtica.

Per això, abans cal analitzar les estructures cristal·logràfiques conegudes que ja estan pre-organitzades per unir-se a Imatinib, per poder comparar després els resultats dels dockings fets amb els PDBs d'aquestes proteïnes quinasa. Com a referència s'han agafat els PDBs coneguts de l'arbre filogenètic de les proteïnes ancestrals les quinases Abl i Src descrit a la Figura 2. Les quinases estudiades són Abl, l'ancestre comú entre Abl i Src (AS) i l'Src, dividides en els següents PDBs: 2HYY (Abl humana unida a Gleevec), 1OPJ (Abl de *Mus Musculus* unida a Gleevec), 4CSV (AS unida a Gleevec) i 2OIQ (Src de *Gallus Gallus* unida a Gleevec), totes aquestes presenten una conformació de DFG-out. A més, a mode de comparació, també s'ha inclòs el 4UEU (AS unida a ACP) que s'uneix a un lligand diferent, l'ACP, que és una variació de l'ATP i es troba amb la conformació DFG-in.

S'ha analitzat visualment a quina part de la proteïna s'uneix el lligand (normalment al centre actiu), i s'han analitzat també els diferents enllaços d'hidrogen entre el lligand i la proteïna al detall.

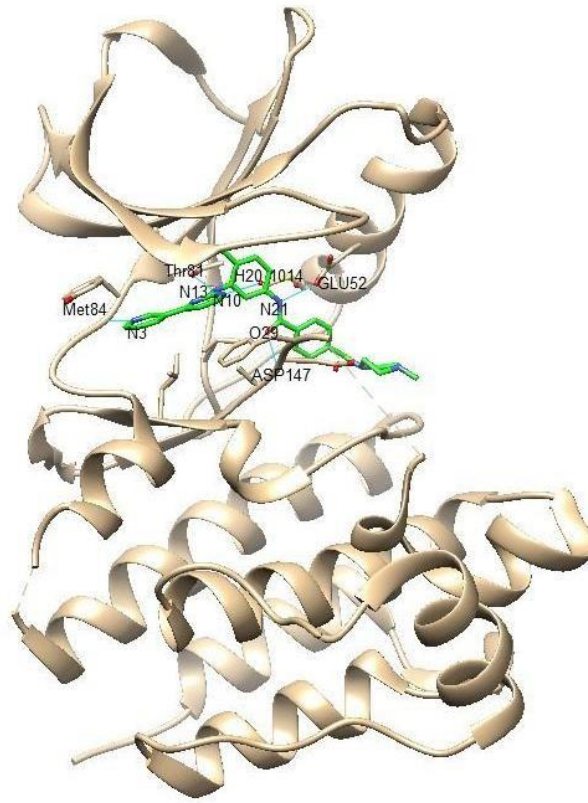


Figura 3: Estructura cristal·logràfica del PDB 4CSV al Chimera. En verd, l'Imatinib. En blau, els ponts d'hidrogen entre l'Imatinib i la proteïna AS. Es troben detallats els àtoms del lligand que conformen aquests ponts d'hidrogen, que també estan amb un color diferent al verd del lligand.

A la Figura 3 es mostra l'anàlisi de l'estructura cristal·logràfica de la proteïna ancestral (AS) entre Abl i Src, que és una de les proteïnes quinasa referència de l'objecte d'estudi, i que es farà servir per a futures comparacions de resultats.

Els enllaços d'hidrogen que es produeixen entre l'Imatinib i la proteïna quinasa són els següents (totes les distàncies que es donen són entre donador i acceptor): 1) entre el grup amino de l'esquelet de l'ASP 147 i l'oxigen del carbonil del lligand (O29) amb una distància entre donador i acceptor de 3.004 Å; 2) entre l'oxigen del grup carboxilat de la cadena lateral del GLU 52 i el grup amino del lligand (N21) amb una distància de 2.939 Å; 3) entre el grup amino de l'esquelet de la MET 84 i el nitrogen de l'anell piridina del lligand (N3) amb una distància de 2.929 Å; 4) entre el grup hidroxil de la cadena lateral de la THR 81 i el nitrogen de l'anell de pirazina del lligand (N13) amb una distància de 2.898 Å; i 5) entre una molècula d'aigua cristal·logràfica (H₂O) corresponent al residu 1014 del PDB 4CSV i el nitrogen de l'anell de pirazina del lligand (N10) amb una distància de 2.642 Å.

S'analitzen aquests enllaços d'hidrogen per tal de poder-los comparar amb els que assignarà la predicció de docking molecular a través del SwissDock i la seva futura anàlisi amb el Chimera. Això, juntament amb la ΔG i l'RMSD, permetrà validar el programa Swissdock per a la predicció de les interaccions entre quinases i Imatinib.

A la Figura 4 es visualitzen la resta d'estructures cristal·logràfiques del complex proteïna quinasa-ligand detallades anteriorment.

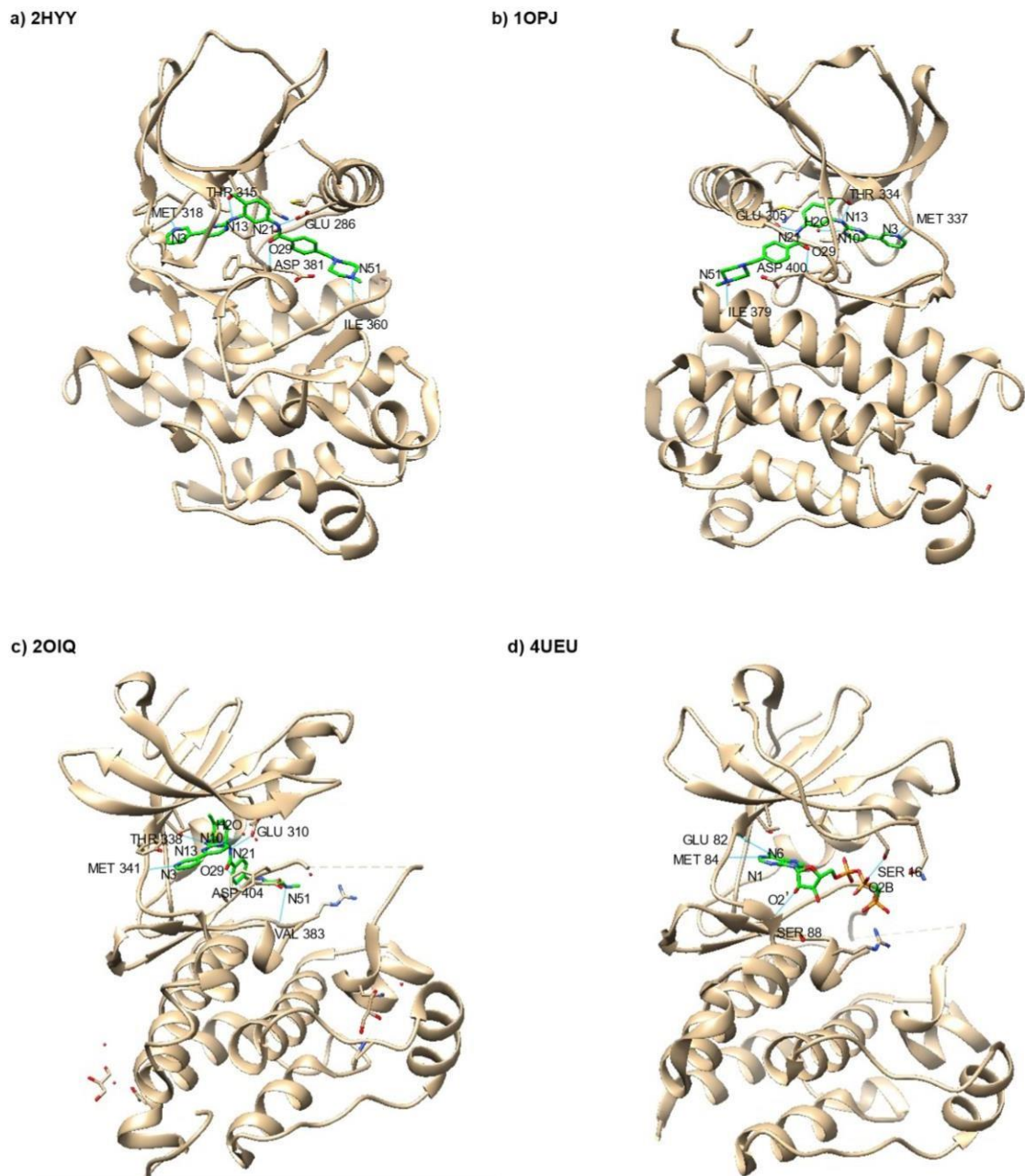


Figura 4: Estructures i ponts d'hidrogen de la resta de PDBs analitzades amb el Chimera. a) 2HYY. b) 1OPJ. c) 2OIQ. d) 4UEU. El 4UEU és l'únic que té com a lligand l'ACP, la resta es troben units a Imatinib.

Com es pot observar a la Figura 4, tots els PDBs que estan units a Imatinib (Fig. 4a,b i c) formen un elevat nombre d'enllaços d'hidrogen amb els mateixos residus, ja que les proteïnes Abl i Src son relativament properes en l'evolució, i també comparteixen la majoria d'enllaços d'hidrogen amb el 4CSV, l'ancestre comú entre ambdues. Tot i que els aminoàcids no corresponguin en quant a número degut a les diferències entre les seqüència de les proteïnes, si son el mateix aminoàcid que forma un enllaç d'hidrogen amb el mateix àtom del lligand. Per tant, els enllaços d'hidrogen són conservats entre les diverses proteïnes que estan relacionades evolutivament.

L'observació dels enllaços d'hidrogen del PDB 4UEU en presència d'ACP (Fig. 4d) i la comparació amb els enllaços d'hidrogen de la predicció del docking servirà de validació complementària a les estructures unides a Imatinib, per poder estudiar com d'acurat és el programa SwissDock per a altres lligands.

A partir d'aquí, s'han fet els càlculs de docking de totes aquestes estructures amb els PDBs preparats per tal de ser acceptats pel SwissDock (detallat a l'apartat de metodologia). La intenció és validar la precisió del programa a l'hora de predir aquesta interacció entre quinasa i lligand, concretament enfocat al Gleevec. El programa SwissDock ofereix un seguit d'orientacions possibles entre lligand i receptor que s'agrupen en diferents clústers. Per a cada orientació es proporciona un valor de ΔG .

A la Figura 5 es mostra la superposició del millor resultat del docking basat en el PDB 4CSV juntament amb el fitxer 4CSV original que conté el lligand posicionat tal i com està a l'estructura cristal·logràfica. S'han analitzat les orientacions del lligand que millor se superposaven visualment amb l'estructura cristal·logràfica del 4CSV referent a l'ancestre comú entre Src i Abl i també s'han considerat els valors de ΔG .

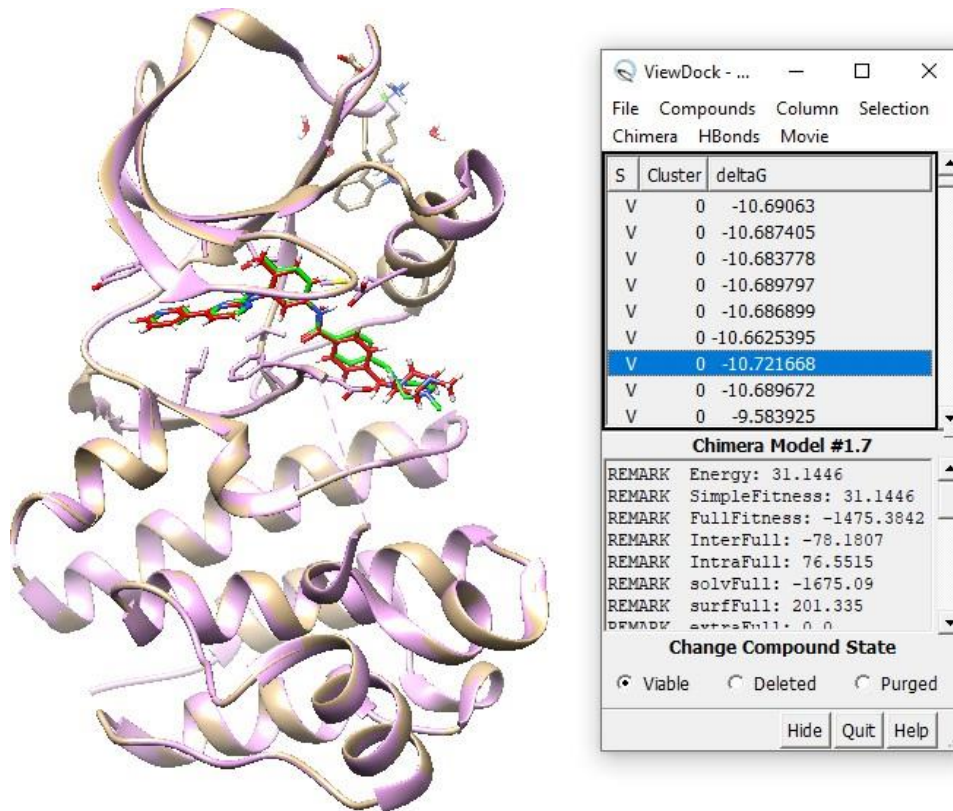


Figura 5: En ocre, estructura de la proteïna “target” oferta pel resultat de docking del 4CSV. En vermell, visualització del lligand de l’orientació 1.7 del clúster 0 del resultat del docking del 4CSV amb l’Imatinib. En lila, estructura cristal·logràfica del 4CSV. En verd, el lligand Imatinib unit a aquesta estructura del PDB 4CSV, a mode de comparació amb la predicció del docking. També es mostra la informació provinent del mòdul ViewDock del Chimera on s’observen diferents modes d’unió del lligand agrupats en diferents clústers i valors de ΔG .

Com es pot observar a la Figura 5, l’orientació 1.7 (clúster 0 i orientació 7 del total de 256) encaixa gairebé perfectament amb el lligand està unit a la proteïna cristal·logràfica ancestral entre Abl i Src, que correspon al fitxer 4CSV que està superposat. Tot i que la proteïna que s’ha agafat per fer el docking (4CSV sense el lligand) ja està preorganitzada per unir-se a Imatinib, aquest primer resultat visual és positiu per a la comprovació de la funcionalitat del SwissDock, tot i ser aviat per treure conclusions.

A la Figura 6 es mostren altres modes d’enllaç (“binding poses”) que tenen una ΔG significativament negativa i que, per tant, segons aquesta eina bioinformàtica prediu que hi ha una molt bona interacció, però realment la superposició amb l’estructura coneguda mostra que no és una orientació correcta del lligand, o com a mínim no tan bona com la posició 1.7.

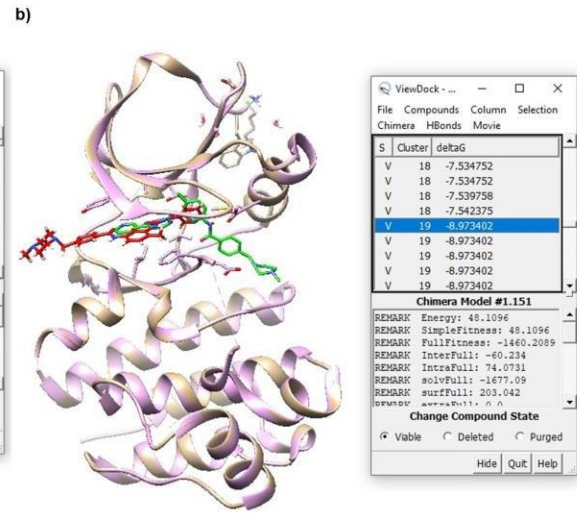
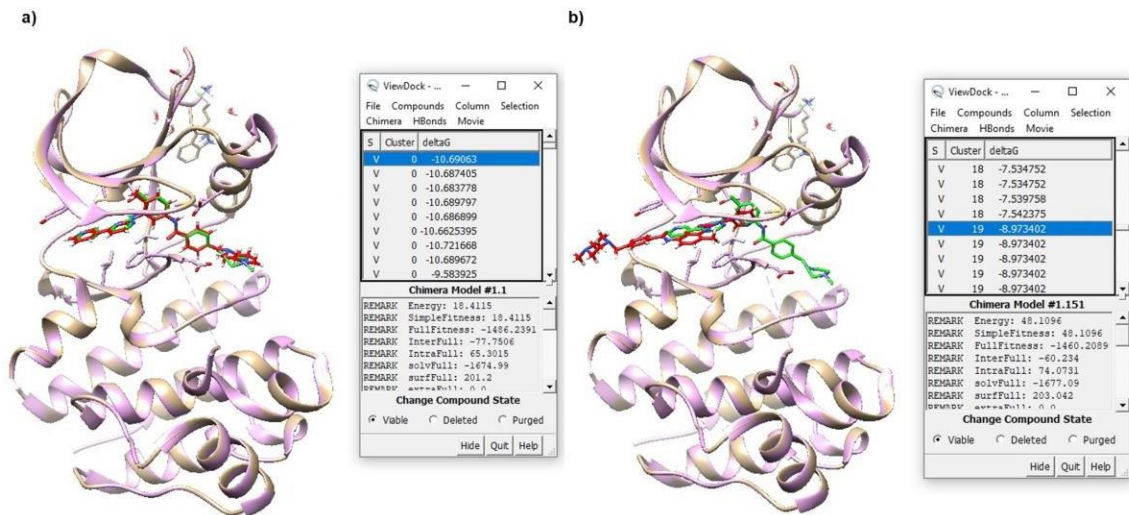


Figura 6: Visualització de les orientacions 1.1 del clúster 0 (a) i 1.151 del clúster 19 (b) del docking de la proteïna quinasa AS (4CSV) i comparació amb l'estructura cristal·logràfica del 4CSV. La llegenda de colors és la mateixa que la de la Figura 4, en ocre i vermell l'estructura i el lligand del docking i en lila i verd l'estructura i el lligand del 4CSV.

Concretament, en l'orientació 1.1 de l'Imatinib al docking (fig. 6a), tot i tenir una ΔG una mica menys negativa (-10.69 kcal/mol) que la de la posició 1.7 (-10.72 kcal/mol), s'observa una superposició molt semblant la de l'orientació 1.1. Per tant, es pot concloure que tots els elements del clúster 0 presenten una bona afinitat. La posició 1.151 referent al clúster 19 i orientació 151 del total 256 (fig. 6b), tot i presentar una de les ΔG més negatives fora del clúster 0 (-8.97 kcal/mol), s'observa clarament com difereix molt el lloc d'unió respecte la comparativa. Per tant, es conclou que és molt important analitzar els resultats visualment i no discriminar només en base al valor de ΔG .

A aquesta comparació visual s'hi suma el càlcul de l'RMSD, una mesura quantitativa que mesura la distància entre els àtoms del lligand del docking i els àtoms de la posició coneguda d'aquest mateix lligand a l'estructura 4CSV en aquest cas, i com més petit és l'RMSD, millor és la similitud entre la predicció del docking i l'orientació del lligand a l'estructura cristal·logràfica.

És una anàlisi útil ja que el càlcul de docking podria orientar l'Imatinib a la regió correcta de la cavitat definida pel bucle d'activació de la quinasa i predir una alta afinitat, però orientar-lo completament girat a com es coneix que s'hi uneix realment. L'RMSD i l'anàlisi d'interaccions permetrien discriminar entre aquestes situacions. És per això que a la Taula 1 hi ha la comparativa de l'RMSD d'aquestes 3 posicions estudiades pel docking del 4CSV.

Taula 1: Resultats de ΔG i RMSD de les “binding poses” 1.1, 1.7 i 1.151 del docking fet a partir de la proteïna del 4CSV i l’Imatinib.

“Binding pose”	ΔG (kcal/mol)	RMSD (Å)
1.1	-10.69	2.025
1.7	-10.72	0.609
1.151	-8.97	18.292

En el cas del docking del 4CSV i tal i com es mostra a la Taula 1, la posició 1.7 és la que té tan la ΔG més negativa com l’RMSD més petit. Per tant, és la posició que més s’ajusta a la orientació de l’Imatinib en l’estructura cristal·logràfica. La diferència principal entre la posició 1.1 i 1.7 fa referència a la rotació de l’anell de piridina que es troba a un dels extrems del lligand. En base a aquests resultats es pot concloure que amb el programa SwissDock s’ha pogut capturar l’orientació correcta de l’Imatinib unit a la proteïna quinasa referent a l’ancestre comú entre Src i Abl.

A partir d’aquest punt, s’ha triat sempre l’orientació del docking que s’ajusti més a la comparativa, sense necessàriament ser l’orientació que té una ΔG més negativa. L’RMSD ha ajudat a triar aquelles posicions que visualment eren molt semblants i que tenien una ΔG també molt similar. En cas que cap orientació oferta pel SwissDock s’hagi ajustat bé a la posició correcta, s’ha triat sempre la que ha tingut una ΔG més negativa. Se seguirà aquest criteri al llarg del treball i no es mostrarà una anàlisi tan detallada de les diferents prediccions dels dockings per la resta de PDBs per motius de concisió.

A la Taula 2 es mostren els resultats dels dockings moleculars de les altres quatre proteïnes quinasa que presenten PDBs coneguts esmentades amb anterioritat, juntament amb la posició 1.7 del docking del 4CSV. També s’afegeix una columna amb el número d’enllaços d’hidrogen i entre parèntesi els que conserva respecte l’estructura cristal·logràfica. L’enllaç d’hidrogen amb la molècula d’aigua cristal·logràfica no s’ha considerat en el docking molecular.

Es considera que un enllaç d’hidrogen és conservat si el forma el mateix aminoàcid de la proteïna quinasa que l’observat en el PDB.

Taula 2: Valors de ΔG , RMSD i el número d'enllaços d'hidrogen dels dockings realitzats per a les cinc mateixes proteïnes de les quals s'han estudiat abans els enllaços d'hidrogen: 4CSV, 2HYY, 1OPJ, 2OIQ, 4UEU. A la columna dels enllaços d'hidrogen, també s'informa entre parèntesi dels enllaços d'hidrogen que s'han conservat respecte el PDB original.

Docking PDB	"Binding pose"	ΔG (kcal/mol)	RMSD (Å)	Enllaços H (conservats)
4CSV	1.7	-10.72	0.609	3 (3)
2HYY	1.1	-11.45	1.648	3 (3)
1OPJ	1.1	-10.44	2.081	1 (1)
2OIQ	1.1	-11.41	0.913	4 (4)
4UEU	1.89	-8.73	-	2 (1)

Els resultats de la Taula 2 proporcionen diferents informacions. La primera, els dockings de les tres proteïnes quinasa principals (Abl, AS i Src) de l'estudi han donat molt bons resultats, amb unes ΔG molt negatives, uns RMSD petits, i 3 o 4 enllaços d'hidrogen conservats respecte l'estructura cristal·logràfica corresponent. Ara bé, l'1OPJ, que també és la proteïna Abl però de l'espècie *Mus Musculus*, ha donat un resultat pitjor del que s'esperava per la proteïna que segons bibliografia té millor interacció amb l'Imatinib, conservant només un enllaç d'hidrogen i amb un RMSD i ΔG que, tot i que dins d'uns valors acceptables, són pitjors que els tres resultats analitzats fins ara. El docking del 4UEU indica, a través de la ΔG i els enllaços d'hidrogen, que el programa SwissDock no ha aconseguit en cap de les 256 posicions acostar-se a una predicció prou correcta de com interacciona l'ACP amb la proteïna ancestral AS. Aquest fet podria ser degut a que la molècula d'ACP que presenta diversos grups fosfat, necessita un ió magnesi per estabilitzar-se al centre actiu. Aquest ió magnesi no s'ha pogut tenir en compte en els càlculs de docking ja que el SwissDock no permet afegir ions.

Una altra informació que es pot extreure de la Taula 2, és que no segueix la progressió d'afinitat entre Imatinib i proteïna esperada esperada segons els experiments de Wilson et al., 2015. S'esperaria que l'Abl humana (2HYY) donés la millor afinitat amb Imatinib, la qual cosa es compleix en clau ΔG però no en RMSD.

Després, l'hauria de seguir amb una afinitat més baixa l'AS (4CSV), que es compleix amb ΔG però l'RMSD és més petit que pel 2HYY, i per tant s'ha acostat més a la posició de la seva estructura cristal·logràfica. Per l'Src (2OIQ), que hauria de ser la que mostra una pitjor afinitat, trobem que té la segona ΔG més negativa (pitjor que l'Imatinib en el docking de 2HYY però millor que al de 4CSV) i que té la segona millor superposició respecte la posició correcta de l'Imatinib (pitjor que l'Imatinib en el docking de 4CSV però millor que al de 2HYY). El resultat que trenca més la tendència esperada és el del docking de 2OIQ. Això es podria explicar conjuntament amb per què l'1OPJ no ha donat el resultat que s'esperava comparat amb l'Abl humana, i és que tant 2OIQ com l'1OPJ son proteïnes quinasa d'espècies que no son la humana, i això podria afectar a aquesta afinitat esperada. Per altra banda, també es pot considerar que el docking molecular prediu correctament l'orientació però no l'afinitat del lligand amb la proteïna. Per poder determinar l'afinitat d'una manera més precisa caldria tenir en compte els efectes dinàmics a través d'una dinàmica molecular.

A continuació, a la Figura 7, hi ha detallats els enllaços d'hidrogen que formen les orientacions seleccionades de l'Imatinib dels dockings de les proteïnes 4CSV i 2HYY, a mode d'exemple.

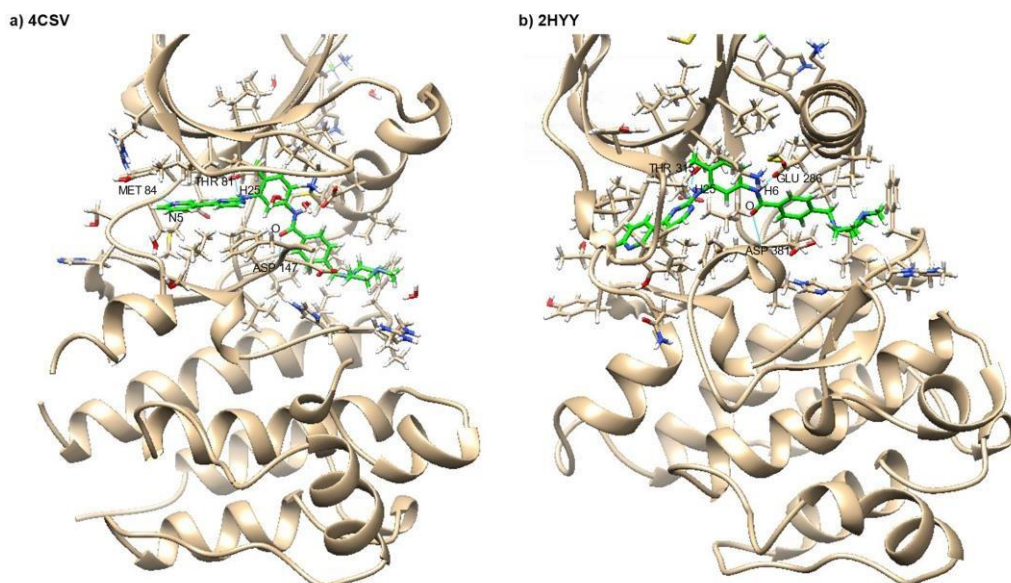


Figura 7: Visualització dels resultats del docking de les estructures 4CSV (a) i 2HYY (b), amb els ponts d'hidrogen que formen les orientacions del lligand triades detallats. Visualitzat amb el Chimera

Un cop analitzats els resultats de docking es pot concloure que el SwissDock prediu amb molt bona capacitat la interacció entre l'Imatinib i les proteïnes quinasa tot i que no permet discriminar entre l'afinitat de les diferents proteïnes. En canvi, amb el lligand ACP, cap de les orientacions del lligand proposades pel programa ha sigut bona.

4.2. Validació docking molecular amb la proteïna quinasa completament reconstruïda

Observant amb detall els fitxers PDB utilitzats, es pot veure que hi ha una part de l'estructura que no està resolta, ja que les regions molt flexibles no solen aparèixer a les estructures cristal·logràfiques. Mirant la seqüència amb el PyMol, tal i com es mostra a la Figura 8, s'observa que justament aquesta part de l'estructura que falta correspon exactament els següents aminoàcids després del motiu DFG i que corresponen al bucle d'activació. Aquest motiu estructural és molt important per explicar l'activitat de les quinases i, per tant, no tenir-lo en compte pot afectar a les prediccions realitzades amb el docking molecular.

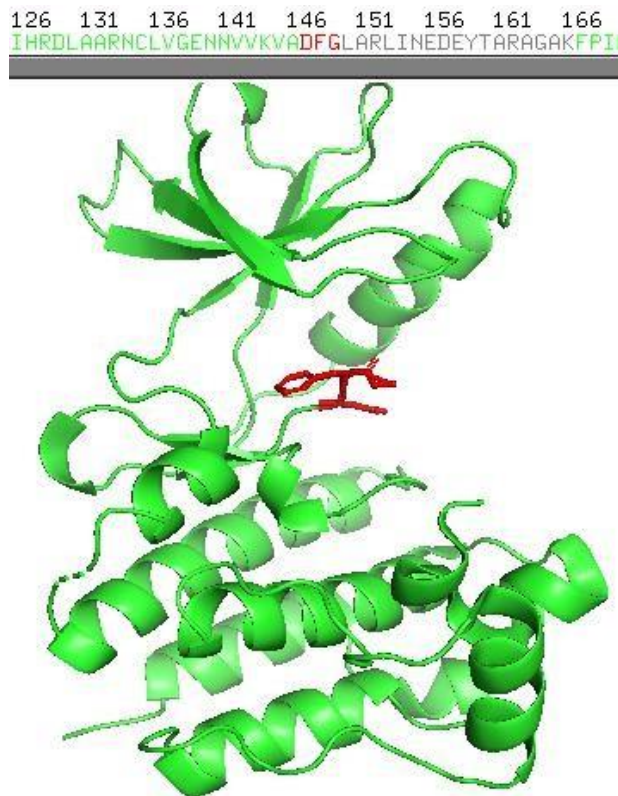


Figura 8: Fitxer PDB 4CSV sense el Gleevec visualitzat al PyMol. A dalt de la imatge, una part de la seqüència, amb el DFG seleccionat. En gris, la part de la seqüència que no està representada a l'estructura 3D. A la imatge també surt seleccionat el DFG, de color vermell i amb la opció "Sticks" del PyMol.

A la Figura 9 hi apareixen les reconstruccions del 4CSV i 2HYY a mode d'exemple, així com la superposició entre elles.

A la Figura 8 es pot veure clarament com just després del motiu DFG hi ha una part de l'estructura que està sense reconstruir, que correspon a la part grisa de la seqüència aminoacídica mostrada a la Figura referent al bucle d'activació. Com que se'n coneix la seqüència i que no hi sigui aquesta part de l'estructura clau de les quinases podria ser important pels resultats de docking, el següent pas és reconstruir aquesta part de l'estructura mitjançant models d'homologia.

La reconstrucció s'ha fet amb el servidor d'homologia Swiss-Model i només s'ha realitzat per les quatre proteïnes que estan preorganitzades per interaccionar amb Imatinib, el PDB 4UEU s'ha descartat pel mal resultat del docking.

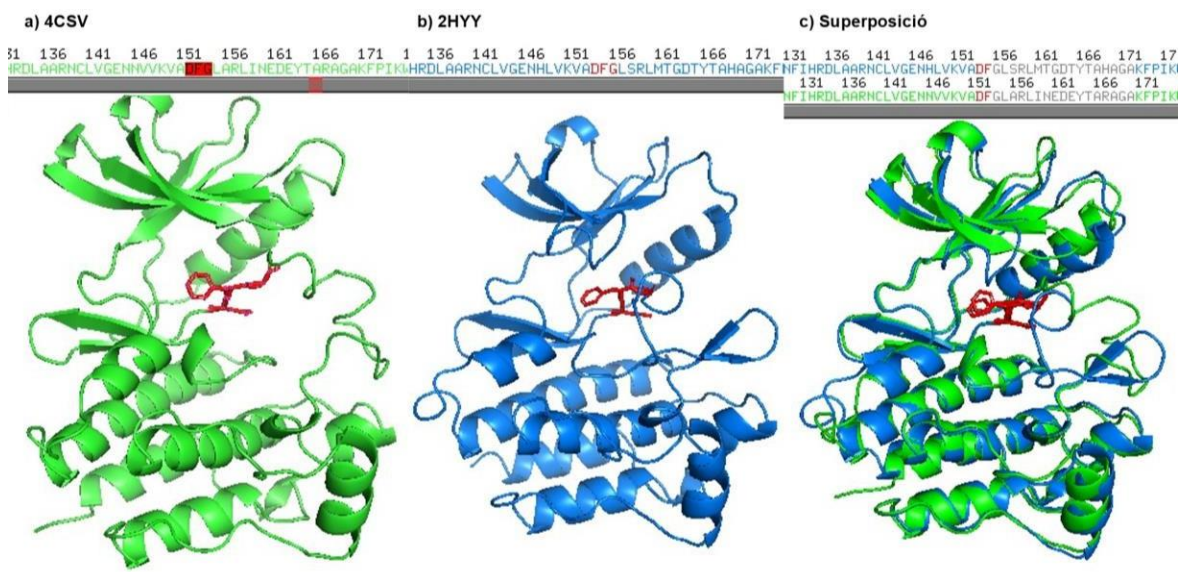


Figura 9: Visualització amb el PyMol de les reconstruccions parcials dels fitxers 4CSV (a, en verd) i 2HYY (b, en blau). A l'apartat c s'hi troba la superposició entre aquestes dues estructures amb els mateixos colors que a les dues anteriors. A les tres imatges hi ha el DFG seleccionat, en vermell i amb la opció "Sticks" del PyMol.

Es pot observar a la Figura 9 que la reconstrucció ha sigut diferent pel 4CSV que pel 2HYY, sobretot si ens fixem en la superposició de la fig. 9c i que per tant això pot afectar a com s'uneix l'Imatinib, ja que aquesta part reconstruïda forma part del bucle d'activació.

Un cop obtingudes les quatre reconstruccions, s'han introduït els fitxers amb la proteïna completa al SwissDock i s'han tornat a calcular els dockings utilitzant l'Imatinib com a lligand. A la Taula 3 es mostren els resultats obtinguts per aquests dockings moleculars.

Taula 3: Resultats dels dockings de les quatre proteïnes reconstruïdes: 4CSV, 2HYY, 1OPJ i 2OIQ, amb l'Imatinib com a lligand. Hi apareix l'orientació del lligand triada, la ΔG , l'RMSD i els ponts d'hidrogen que forma aquesta orientació de l'Imatinib amb la proteïna, juntament amb els que son conservats en parèntesi.

Docking PDB	"Binding pose"	ΔG (kcal/mol)	RMSD (Å)	Enllaços H (conservats)
4CSV	1.1	-9.75	1.834	1 (1)
2HYY	1.1	-11.02	1.971	2(2)
1OPJ	1.2	-10.56	2.122	1(1)
2OIQ	1.6	-10.62	1.721	2 (2)

Els resultats de docking mostrats a la Taula 3 indiquen que de manera generalitzada les reconstruccions han empitjorat una mica les interaccions i afinitat entre proteïna quinasa i el Gleevec, però que tot i així segueixen essent unes prediccions bones. Totes quatre proteïnes presenten un RMSD més gran que les orientacions dels dockings amb els PDBs sense reconstruir, i per tant una pitjor superposició que aquests respecte l'Imatinib de les estructures cristal·logràfiques. Només el lligand l'1OPJ reconstruït presenta una lleugera millora en la ΔG que la seva estructura sense reconstruir, la resta han obtingut una ΔG menys negativa que per les seves estructures sense reconstruir, amb una diferència de 0.5 kcal/mol per 2HYY i de gairebé 1 kcal/mol per 4CSV i 2OIQ.

Respecte als enllaços d'hidrogen, exceptuant l'1OPJ que ha mantingut l'únic enllaç d'hidrogen que havia pogut conservar el docking de la seva estructura sense reconstruir, la resta han perdut un o varis enllaços d'hidrogen, deixant patent la interacció lleugerament més feble d'aquests dockings amb les estructures reconstruïdes. A la Figura 10 es mostra el resultat del docking de la proteïna 1OPJ reconstruïda.

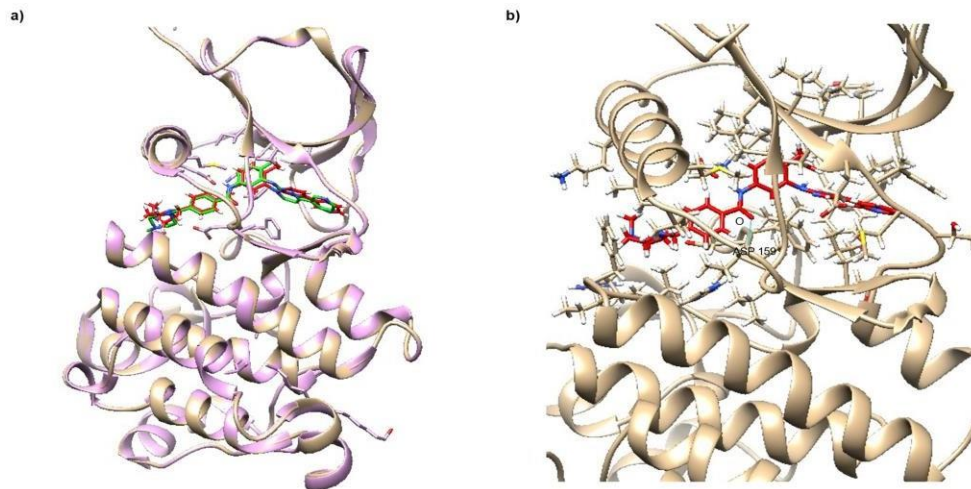


Figura 10: Resultat del docking de la reconstrucció d'1OPJ. A la imatge a) En ocre i vermell l'estructura i orientació 1.2 del lligand de la reconstrucció 1OPJ, respectivament. En lila i verd l'estructura i lligand de l'estructura cristal·logràfica 1OPJ, respectivament. b) Enllaç d'hidrogen que forma l'orientació 1.2 del lligand de la reconstrucció de l'1OPJ.

Val la pena centrar-se sobretot en el resultat de l'1OPJ (fig. 10) ja que, tot i que la resta de resultats de docking de les proteïnes reconstruïdes han sigut força bons, és l'única reconstrucció que ha obtingut una interacció pràcticament idèntica a la del docking sense reconstruir, o fins i tot una mica més negativa en termes de ΔG . Aquests resultats indiquen que aquesta reconstrucció del bucle d'activació amb el Swiss-Model ha sigut acurada, ja que la part reconstruïda no ha interferit geomètricament amb com s'uneix l'Imatinib a l'Abl de *Mus Musculus*. Aquests resultats permeten validar el Swiss-Model com una bona eina per a fer reconstruccions parcials de proteïnes quinasa a través d'una proteïna "template". Tot i això, s'ha de tenir en compte que la reconstrucció del bucle d'activació es realitza sense la presència de lligand, pel que pot adoptar conformacions que bloquegin total o parcialment el lloc d'unió dels inhibidors de quinasa de tipus II que són els que ocupen aquesta regió. Aquest fet és difícil de controlar quan es realitza el model d'homologia i, per tant, es recomana realitzar proves amb diferents conformacions del bucle d'activació.

4.3. Reconstrucció completa de les proteïnes quinasa ancestrals de les quals es desconeix l'estructura cristal·logràfica

Com s'ha exposat amb anterioritat a la introducció, a part de les proteïnes les quals s'ha obtingut l'estructura cristal·logràfica, Wilson i col·laboradors van reconstruir les seqüències d'altres proteïnes ancestrals d'Abl i Src però no se'n van obtenir les estructures cristal·logràfiques, concretament A1, A2 i S1. A1 i A2 es troben en una situació intermèdia entre AS i Abl i S1 entre AS i Src. D'aquestes proteïnes experimentalment s'ha determinat que totes presenten afinitat per l'inhibidor Imatinib.

A partir d'aquestes tres seqüències i sense tenir informació de l'estructura, en aquest treball s'ha volgut construir-les de tres maneres diferents. Les dues primeres amb models d'homologia a través del programa Swiss-Model (el mateix que les reconstruccions parcials), amb el 4CSV com a model per una banda i amb el 4UEU per l'altra, com a comparació. S'han escollit aquests dos models perquè en el PDB 4CSV la proteïna quinasa es troba en conformació DFG-out i en el cas de 4UEU en conformació DFG-in. La tercera s'ha determinat utilitzant intel·ligència artificial mitjançant el programa AlphaFold3. Un cop obtingudes les estructures del Swiss-Model A1, A2 i S1, s'ha tornat a fer el docking amb l'Imatinib, tant per les que han estat construïdes a partir del 4CSV com el 4UEU.

A la Taula 4 i 5 hi ha els resultats dels dockings d'aquestes tres construccions amb el "template" 4CSV i 4UEU, respectivament.

Taula 4: Resultats dels dockings de les prediccions estructurals de seqüències ancestrals intermèdies amb el 4CSV "template". Hi apareix l'orientació de l'Imatinib triada, la ΔG , l'RMSD respecte el fitxer 4CSV i els enllaços d'hidrogen formats pel lligand.

Docking PDB	"Binding pose"	ΔG (kcal/mol)	RMSD (Å)	Enllaços H
A1	1.3	-8.94	15.701	0
A2	1.16	-9,42	6.782	0
S1	1.1	-10,20	2.503	1

Taula 5: Resultats dels dockings de les prediccions estructurals de seqüències ancestrals intermèdies amb el 4UEU "template". Hi apareix l'orientació de l'Imatinib triada, la ΔG , l'RMSD respecte el fitxer 4CSV i els enllaços d'hidrogen formats pel lligand.

Docking PDB	"Binding pose"	ΔG (kcal/mol)	RMSD (Å)	Enllaços H
A1	1.9	-8,92	12.055	0
A2	1.2	-8,97	15.269	1
S1	1.1	-9,06	15.920	1

Els resultats de la Taula 4 i 5 mostren de manera generalitzada que el Swiss-Model no ha sigut capaç de construir aquestes tres proteïnes ancestrals de manera acurada a partir de la seva seqüència, amb cap dels dos fitxers model utilitzats. En comparació amb les proteïnes amb estructura cristal·logràfica coneguda, aquesta informació es pot extreure observant els valors poc negatius de ΔG , els valors de RMSD generalitzadament molt alts, i la poca capacitat de formar ponts d'hidrogen de les orientacions del lligand triades, a excepció de la proteïna S1 reconstruïda a partir del model 4CSV.

Si tornem a fer un cop d'ull a la Figura 2a, ja que segons Wilson et al., 2015 la major afinitat de l'Imatinib és amb Abl i la pitjor amb Src, passant per l'ancestre comú (AS). Així doncs, l'ordre de major a menor afinitat que esperariem d'aquestes tres proteïnes seria: $A2 > A1 > S1$. Això no es compleix en base als resultats de les reconstruccions fetes amb el Swiss-Model, però com que els resultats son generalitzadament dolents tampoc es pot pretendre fer aquesta comparació.

L'únic resultat que ha sortit amb uns paràmetres de ΔG i RMSD molt acceptables ha sigut la S1 construïda amb el 4CSV. S'esperaria que aquesta S1 tingués la menor afinitat de les tres, però té la millor de les tres amb diferència. Als resultats de la Taula 4 sí que s'observa la tendència de que A2 ha tingut millor interacció que A1, però la mala interacció dels dos tampoc dona peu a poder treure'n una conclusió clara. El que sí que es pot determinar és que de manera generalitzada, les reconstruccions fetes amb el 4CSV com a model han sigut millors que les realitzades amb el 4UEU com a model, tal i com ho demostren els resultats dels dockings. Aquestes observacions van en la línia dels resultats experimentals que mostren que l'Imatinib presenta més afinitat per la conformació DFG-out (PDB 4CSV) que per la DFG-in (PDB 4UEU). Al fer la reconstrucció completa, el bucle d'activació ocupa part de la zona d'unió del Imatinib de manera que el docking no surt correctament. Es pot concloure doncs que la reconstrucció de proteïnes quinasa a partir de models d'homologia no és del tot acurada per determinar si un inhibidor té o no afinitat.

Buscant una alternativa per construir aquestes estructures de manera acurada per predir la interacció amb Imatinib s'ha acabat provant el servidor d'intel·ligència artificial AlphaFold3. Aquest servidor web, per cada seqüència proporcionada ofereix cinc estructures diferents de cada reconstrucció ordenats del model 0 al 4, essent el model 0 teòricament més acurat que la resta. També ofereix un mapa de la proteïna amb uns valors de confiança per cada regió d'aquesta, que indiquen com de fiable ha sigut la reconstrucció de cada part. Amb l'AlphaFold3 s'ha construït l'estructura d'aquestes tres proteïnes (A1, A2, S1), però també s'ha fet amb la seqüència de Abl (PDB 2HYY) i de l'ancestre comú (PDB 4CSV), a mode de validació d'aquesta eina.

Abans de passar cap als resultats dels dockings posteriors cal fer una ullada a la Figura 11, on es mostren el model 0 i 1 del 4CSV proposats per l'AlphaFold, com també la superposició del model 0 amb l'estructura cristal·logràfica i el mapa amb valors de confiança ofert per l'AlphaFold.

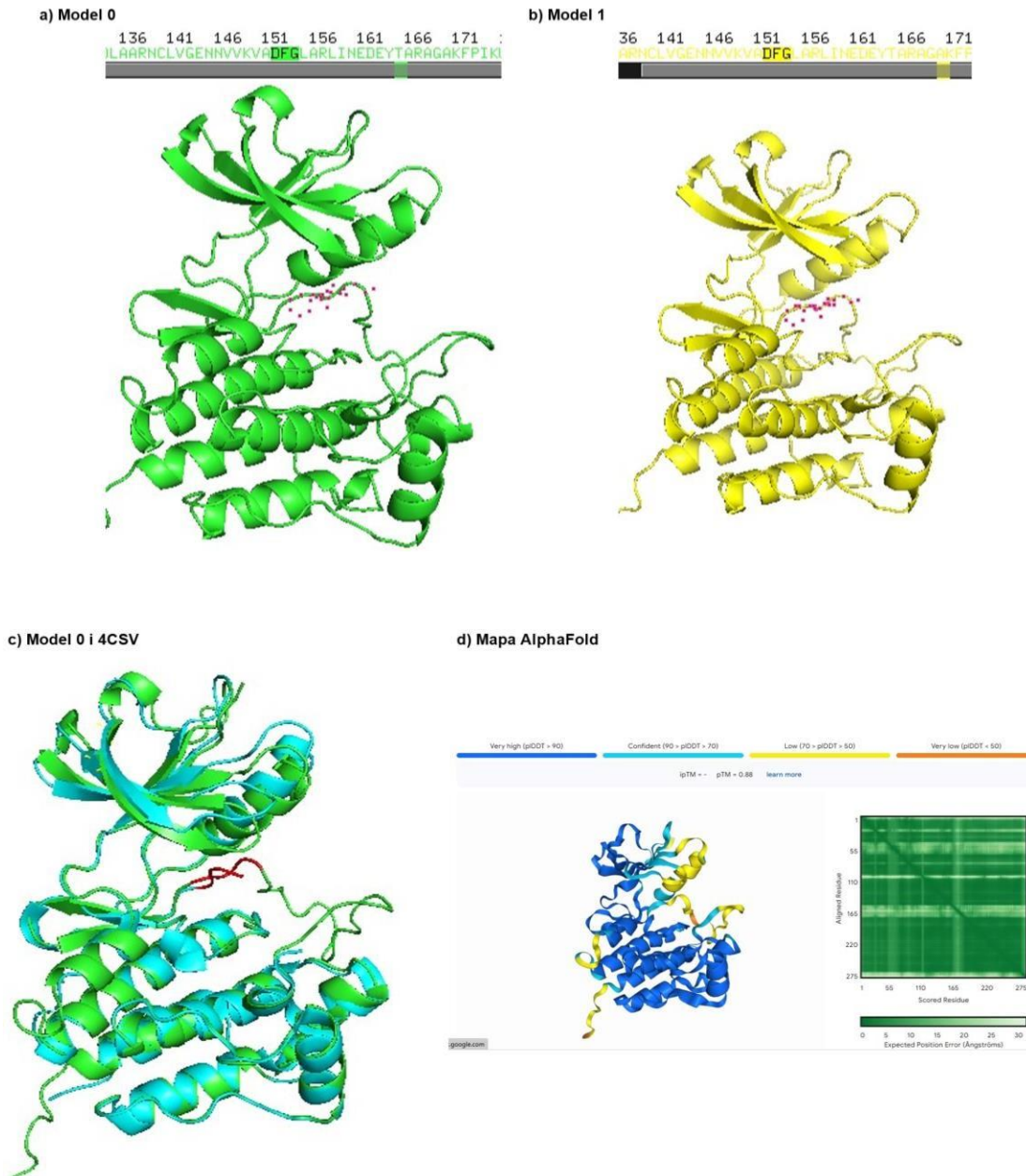


Figura 11: a) Visualització del resultat “model 0” de la reconstrucció a partir de la seqüència del 4CSV de l'AlphaFold, amb el PyMol. b) Visualització del resultat “model 1” de la reconstrucció a partir de la seqüència del 4CSV de l'AlphaFold, amb el PyMol. c) Estructura “model 0” (en verd) i l'estructura cristal·logràfica 4CSV (en blau) superposades. En les dues molècules es troba seleccionat el motiu DFG, en vermell. d) Mapa dels valors de confiança que indiquen la fiabilitat de la predicció, ofert per l'AlphaFold.

A la Figura 11a i 11b es poden veure els models 0 i 1 de la construcció de l'estructura del 4CSV que ha fet l'AlphaFold3 a partir de la seva seqüència. Son dues estructures molt similars, igual que amb les altres tres proposades per aquesta eina, i sobretot son pràcticament idèntiques pel que fa a l'orientació del DFG (es troba en conformació DFG-in), el centre actiu i el bucle d'activació (regió que presenta els valors més baixos de confiança). Aquesta poca diversitat a l'hora d'orientar el DFG dona poques possibilitats de seleccionar una altra de les estructures proposades si la que s'analitza dona un mal resultat de docking, ja que si totes tenen el DFG orientat de la mateixa manera, no pot haver-hi diversitat de resultats. Això ha passat no només amb el 4CSV, sinó amb totes les cinc estructures que s'han predit amb l'AlphaFold3.

Respecte la Figura 11c, la superposició entre el model 0 de la predicció i el 4CSV cristal·logràfic permet valorar d'una manera visual com de bona és aquesta predicció. El més rellevant és fixar-se amb el motiu DFG, marcat en vermell a la imatge, i veure com la predicció adopta una orientació diferent (DFG-in) d'aquest que la que té en l'estructura cristal·logràfica (DFG-out). Tenint en compte que la orientació del DFG és un element clau en la interacció de les proteïnes quinasa amb els lligands de tipus I i II, això pot afectar severament al resultat de docking d'aquesta estructura i al poder predictiu de les estructures generades amb AlphaFold3.

La Figura 11d mostra un mapa amb uns valors de confiança, que exposen quines parts de l'estructura ha pogut determinar amb més o menys certesa l'AlphaFold. Es pot observar que una de les parts amb la classificació de baixa i fins i tot molt baixa resolució és el DFG, part del centre actiu i sobretot la zona corresponent al bucle d'activació. Això clarament significa que l'AlphaFold no ha pogut predir de manera acurada aquestes estructures, com s'ha comprovat a la Figura 11c, i com ha passat també amb la resta d'estructures predites. Això és en part degut a que les proteïna quinasa són altament flexibles i el programa AlphaFold3 només captura una determinada conformació. Fet que pot limitar l'ús d'aquestes estructures pel descobriment de nous fàrmacs.

Per reafirmar que amb aquesta mala orientació del DFG hi ha una mala interacció amb el lligand, s'han realitzat els càlculs de docking d'aquestes dues molècules de validació (4CSV i 2HYY) i també amb les tres proteïnes ancestrals intermèdies amb les que s'ha treballat fins ara (A1, A2 i S1), amb l'Imatinib igual que al llarg del treball. Els resultats d'aquests dockings es mostren a la Taula 6.

Taula 6: Resultats dels dockings de les reconstruccions a partir de la seqüència amb l'AlphaFold següents: 4CSV, 2HYY, A1, A2 i S1, totes amb l'Imatinib com a lligand. Hi apareix l'orientació del lligand triada, la ΔG , l'RMSD respecte el fitxer 4CSV i els enllaços d'hidrogen formats pel lligand.

Docking PDB	"Binding pose"	ΔG (kcal/mol)	RMSD (Å)	Enllaços H
4CSV	1.13	-8.92	12.285	0
2HYY	1.14	-9.23	11.331	0
A1	1.9	-8,93	12.102	0
A2	1.3	-9,21	12.309	0
S1	1.9	-8,17	17.550	0

Els resultats de la Taula 6 mostren uns resultats de docking força dolents de manera generalitzada, confirmant així la mala predicció d'estructures que ha fet l'AlphaFold en aquestes 5 proteïnes quinasa. Aquests mals resultats de docking molt probablement son deguts a la mala orientació que ha ofert l'AlphaFold del DFG en aquestes reconstruccions, tal i com s'ha comentat amb anterioritat. Cap orientació del lligand dels cinc dockings ofereix un valor de RMSD per sota dels 11 Å, indicant la mala superposició amb l'Imatinib a l'estructura 4CSV de totes les prediccions. Cap de les estructures ha format cap enllaç d'hidrogen amb les orientacions de l'Imatinib triades.

El que també es pot extreure d'aquests resultats és la lleugera tendència esperada dels resultats de les tres proteïnes ancestrals intermitges (A1, A2 i S1) sobretot a nivell de ΔG , que ordena les interaccions de la forma esperada segons la bibliografia: $A2 > A1 > S1$. Tot i això, no es pot confirmar aquesta tendència degut al mal resultat generalitzat de les interaccions i degut a que l'RMSD tampoc confirma que el mode d'enllaç sigui el correcte.

En definitiva, l'AlphaFold3 no ha predit de forma correcta les estructures tridimensionals que s'han estudiat en aquest treball i encara li queda molt camí per recórrer a aquesta eina per poder esdevenir una alternativa fiable per a les reconstruccions d'estructures de les proteïnes quinasa, o com a mínim per tal de que aquestes estructures puguin ser usades per a càlculs de docking amb Imatinib i puguin assimilar-se a les interaccions reals d'aquestes.

De totes maneres, afinant la diversitat de resultats que dona l'AlphaFold i potser incloent estructures que a priori donin un pitjor "Score" als valors de confiança de la predicció però que tinguin el DFG orientat de forma diferent, es podria arribar a algunes estructures que podrien ser molt acurades pels càlculs de docking molecular, esdevenint així de manera indubtable una eina molt útil per a tots els processos que podria estalviar amb un sol click. Una altra possibilitat de millora consistiria en poder incloure el lligand en la predicció de l'estructura proteica. De manera que es podria intentar estabilitzar una conformació en concret i així millor la precisió dels càlculs de docking.

CONCLUSIONS

In this study, we explored the performance of different bioinformatic tools to predict the binding mode and binding affinity of known inhibitor Imatinib with a number of evolutionary related protein kinases including Abl, Src and common protein ancestors (AS, A1, A2, S1).

First, molecular docking results of the known structures of protein kinases determined the good accuracy of SwissDock in terms of predicting protein kinase-Imatinib binding mode in Abl, Src and AS. The wideness of possible positioning of the ligand helped choosing the right binding pose, based on the crystallographic PDBs comparisons. Although the ligand orientation with the most negative ΔG was not always the one with better RMSD with respect to the X-ray binding mode, ΔG was generally a good indicator of which binding poses were the most accurate.

Second, partial reconstruction of the activation loop of the protein kinases studied helped understanding why the DFG motif and the activation loop are very relevant structures for the kinase in order to interact with a ligand like Imatinib. Homology models based Swiss-Model made generally good partial reconstructions of the activation loop of these kinases, leading to slightly worse interactions of these predictions with Imatinib than the know crystallographic ones but keeping a plausible binding mode.

Third, however, docking results performed on the Swiss-Model full structure predictions of common ancestors A1,A2 and S1 protein kinases that present unknown X-ray structures were in general bad, in both PDB 4CSV template predictions and PDB 2HYY template ones. Comparing to the orientation of Imatinib between the predicted structures and the PDB 4CSV, all binding poses selected presented an RMSD greater than 11 Å, which demonstrate the poor binding mode these predicted structures have made with Imatinib. Although directing Swiss-Model to construct these structures with 4CSV template, which has a good preorganization to interact with Imatinib, this bioinformatic tool could not make structures that reproduced this good interactions through molecular docking.

Fourth, the full models of the proteins were also predicted with AlphaFold3. This tool could not create 3D structures in any of their models that had better interactions with Imatinib in molecular docking calculations than the ones created by Swiss-Model. The lack of diversity in the orientation of the DFG motif in AlphaFold3 results led to poor docking results of these proposed structures with Imatinib. This was assessed by predicting the structure of 4CSV (AS ancestor) and 2HYY (Abl) PDBs, which also led to poor docking results compared to the original 4CSV and 2HYY PDBs.

Even though AlphaFold3 has not made protein kinase 3D structures that could recompose the interaction with Imatinib, enhancing the diversity of the DFG orientation in the 5 models that the tool propose could help finding a prediction that would be able to be used for molecular docking calculations in a reliable way. In this sense, if AlphaFold3 can solve these accuracy problems (involving preorganization of their predicted structures), it could without any doubt become a very useful tool in molecule structure predictions.

In general, we found that molecular docking predictions in protein kinases are strongly dependent on having an available X-ray structure. Computational protein structure predictions of protein kinase structures are not representative of the inhibitor-bound state leading to inaccurate binding mode predictions. Even with evolutionary related proteins with similar sequences, these computational structure predictions lead to inaccurate results. Therefore, these results suggest that molecular docking calculations in protein kinases should be analyzed with care, especially if they are used for the discovery of new drugs. The high dynamicity of protein kinases challenge current experimental and computational methods. Including molecular dynamics simulations in these protocols can help to identify the most relevant conformations of protein kinases prior molecular docking calculations.

BIBLIOGRAFIA

- Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., ... Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 1-3. <https://doi.org/10.1038/s41586-024-07487-w>
- Agafonov, R. V., Wilson, C., Otten, R., Buosi, V., & Kern, D. (2014). Energetic dissection of Gleevec's selectivity towards human tyrosine kinases. *Nature Structural & Molecular Biology*, 21(10), 848-853. <https://doi.org/10.1038/nsmb.2891>

- Bugnon, M., Röhrig, U. F., Goullieux, M., Perez, M. A. S., Daina, A., Michielin, O., & Zoete, V. (2024). SwissDock 2024: Major enhancements for small-molecule docking with Attracting Cavities and AutoDock Vina. *Nucleic Acids Research*, gkae300. <https://doi.org/10.1093/nar/gkae300>
- Burley, S. K., Berman, H. M., Bhikadiya, C., Bi, C., Chen, L., Di Costanzo, L., Christie, C., Dalenberg, K., Duarte, J. M., Dutta, S., Feng, Z., Ghosh, S., Goodsell, D. S., Green, R. K., Guranovic, V., Guzenko, D., Hudson, B. P., Kalro, T., Liang, Y., ... Zardecki, C. (2019). RCSB Protein Data Bank: Biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Research*, 47(D1), D464-D474. <https://doi.org/10.1093/nar/gky1004>
- Buslón, N., Cortés, A., Catuara-Solarz, S., Cirillo, D., & Rementeria, M. J. (2023). Raising awareness of sex and gender bias in artificial intelligence and health. *Frontiers in Global Women's Health*, 4, 970312. <https://doi.org/10.3389/fgwh.2023.970312>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589. <https://doi.org/10.1038/s41586-021-03819-2>
- Manning, G., Whyte, D. B., Martinez, R., Hunter, T., & Sudarsanam, S. (2002). The Protein Kinase Complement of the Human Genome. *Science*, 298(5600), 1912-1934. <https://doi.org/10.1126/science.1075762>
- Meharena, H. S., Chang, P., Keshwani, M. M., Oruganty, K., Nene, A. K., Kannan, N., Taylor, S. S., & Kornev, A. P. (2013). Deciphering the Structural Basis of Eukaryotic Protein Kinase Regulation. *PLOS Biology*, 11(10), e1001680. <https://doi.org/10.1371/journal.pbio.1001680>
- PKI | Blue Ridge Institute for Medical Research. (consultat el 15 de maig de 2024). <https://brimr.org/protein-kinase-inhibitors/>
- Roskoski, R. (2016). Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. *Pharmacological Research*, 103, 26-48. <https://doi.org/10.1016/j.phrs.2015.10.021>
- Sterling, T., & Irwin, J. J. (2015). ZINC 15 – Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling*, 55(11), 2324-2337. <https://doi.org/10.1021/acs.jcim.5b00559>
- Wilson, C., Agafonov, R. V., Hoemberger, M., Kutter, S., Zorba, A., Halpin, J., Buosi, V., Otten, R., Waterman, D., Theobald, D. L., & Kern, D. (2015). Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science*, 347(6224), 882-886. <https://doi.org/10.1126/science.aaa1823>