

Estudi del procés d'unió entre la benzamidina i la tripsina mitjançant docking molecular i simulacions de dinàmica molecular

Estudiant: Anna Maria Rustullet Farjas

Grau en Química

Correu electrònic: annarustullet@gmail.com

Tutor: Ferran Feixas Gerones

Empresa / institució Institut de Química Computacional i Catàlisi, departament de
Química de la Universitat de Girona.

Vistiplau tutor:

Nom del tutor: Ferran Feixas Gerones

Empresa / institució: Universitat de Girona

Correu(s) electrònic(s): ferran.feixas@udg.edu

Agraïments

Principalment vull agrair la realització d'aquest treball a l'equip de l'institut de Química Computacional i Catàlisi i especialment al meu tutor Ferran Feixas per la seva dedicació i suport ja que sense ell no hauria estat possible.

També vull fer especial menció a la coordinadora del Grau de Químic la Montse Rodrigues per la seva predisposició al resoldre dubtes i orientar-nos al llarg del grau.

Finalment agrair a la meva família el suport incondicional i per facilitar-me el camí.

Gràcies

Resum

Per dissenyar un nou fàrmac és fonamental entendre les interaccions entre la proteïna receptora i el fàrmac, tot i que el procés d'unió és difícil d'estudiar degut al caràcter altament dinàmic que presenten les proteïnes en el medi. En aquest estudi s'analitza amb mètodes computacionals com afecten els canvis conformacionals de la proteïna tripsina a l'hora de descriure les interaccions entre aquest enzim i el seu inhibidor benzamidina. A partir de combinar simulacions de dinàmica molecular (DM) amb càlculs de docking molecular es poden obtenir les energies d'interacció per diverses conformacions de la proteïna.

Per analitzar la dinàmica de la proteïna, s'han realitzat simulacions de dinàmica molecular en l'estat *apo* (tripsina) i *holo* (tripsina i benzamidina) de la proteïna que consten d'un total de 100,000,000 conformacions equivalents a un total de 200 ns. Les conformacions de la tripsina obtingudes a la simulació de dinàmica molecular de l'estat *apo* es poden classificar en tres grups: conformació tancada, semi-oberta i oberta, dependent de si el lligand benzamidina pot interaccionar amb el centre actiu (obert i semi-obert) o no pot interaccionar (tancat). A més, els canvis conformacionals del triptòfan 193 i de la glutamina 174 juguen un paper fonamental per determinar la conformació del centre actiu. Per poder obtenir les conformacions més representatives de cada simulació s'utilitzen mètodes de *clustering* que agrupen les conformacions obtingudes en la simulació segons les seves similituds estructurals. De les estructures seleccionades com a més representatives se'n va determinar l'orientació més favorable del lligand i l'energia d'interacció que presentava a través de tècniques de docking. Si l'orientació més favorable no era en el centre actiu es buscava l'energia d'aquesta interacció amb els aminoàcids del centre actiu en cas que el lligand hi pogués interaccionar, és a dir, que no presentés una conformació tancada. A més, es va realitzar un docking de l'estructura cristal·lografia de la tripsina (PDB:3PTB) amb el lligand benzamidina i 10 molècules que presentaven el grup benzamidina a l'estructura. Finalment, s'ha realitzat un estudi d'una simulació de dinàmica molecular posicionant el lligand fora del sistema proteic per intentar reproduir el camí d'interacció amb el centre actiu.

Resumen

Para el diseño de un nuevo fármaco es fundamental comprender las interacciones entre la proteína receptora y el fármaco, aunque el proceso de unión es difícil de estudiar debido al elevado carácter dinámico que presentan las proteínas. En este estudio se analiza con métodos computacionales como afectan los cambios conformacionales de la proteína tripsina a la hora de describir las interacciones entre esta enzima y su inhibidor benzamidina. La combinación de simulaciones de dinámica molecular (DM) con cálculos de docking molecular permite obtener las energías de interacción para varias conformaciones de la proteína.

Se han realizado simulaciones de dinámica molecular en el estado *apo* (tripsina) y *holo* (tripsina y benzamidina) de la proteína que constan de un total de 100,000,000 conformaciones equivalentes a un total de 200 ns. Las conformaciones de la tripsina obtenidas en la simulación de dinámico molecular del estado *apo* se pueden clasificar en tres grupos: conformación cerrada, semi-abierta y abierta, dependiendo de si el ligando puede interactuar con el centro activo (abierto y semi-abierto) o no puede interactuar (cerrado). Además, los cambios conformacionales del triptófano 193 y de la glutamina 174 juegan un papel fundamental para determinar la conformación del centro activo. Para poder obtener las conformaciones más representativas de cada simulación se utilizan métodos de *clustering* que agrupan las conformaciones obtenidas en la simulación según sus similitudes estructurales. De las estructuras seleccionadas como más representativas se determinó la orientación más favorable del ligando y la energía de interacción que presentaba a través de técnicas de docking. Si la orientación más favorable no era en el centro activo se buscaba la energía de esta interacción con los aminoácidos del centro activo en caso de que el ligando pudiera interactuar, es decir, que no presentase una conformación cerrada. Además, se realizó un docking de la estructura cristalografía de la tripsina (PDB:3PTB) con el ligando benzamidina y 10 moléculas que presentaban el grupo benzamidina a la estructura. Finalmente, se ha realizado un estudio de una simulación de dinámica molecular posicionado el ligando fuera del sistema proteico para intentar reproducir el camino de interacción con el centro activo.

Abstract

Understanding the process of protein-ligand binding is of capital importance to design a new drug. However, elucidating the nature of this process is complex because of the intrinsically dynamic character of proteins. In this study, we investigate the effect of protein conformational changes on the ligand-binding energies using computational methods. The system analysed is the trypsin-benzamidine enzyme-inhibitor complex. Molecular dynamics (MD) simulations have been successfully combined with molecular docking to obtain the interaction energy of each protein conformation.

Here, molecular dynamics simulations in the apo (trypsin) and holo (trypsin and benzamidina) protein states have been performed to obtain a total of 100,000,000 conformations which are equivalent to 200 ns of simulation time. The trypsin conformations obtained for the apo molecular dynamics simulations can be classified in three groups: open, semi-open and close structures, depending on the interaction of benzamidine with the amino acids of active site. In addition, the conformational changes of tryptophan 193 and glutamine 174 are fundamental to determine the different conformations of trypsin. To simplify the amount of information obtained, clustering methods are used to get the most representative structures of each simulation, by grouping the structures based on their similarity. The most representative structures were used to calculate the most favorable orientation of the benzamidine ligand and the interaction energy for the different protein conformations using molecular docking techniques. If the most favorable orientation predicted by docking does not target the active site, then less favorable orientations are checked to estimate the interaction energy with the active site. When the active site is found in the close conformation benzamidine does not fit in the active site and the binding event is not observed. Moreover, the crystallographic structure of trypsin (PDB: 3PTB) has been used as a benchmark for docking benzamidine and 10 molecules that contain the benzamidine group in their structure.

Finally, the ligand is placed in the solvent at a certain distance from the target protein to reproduce the binding pathway of benzamidine to trypsin.

Índex

1. Introducció	1
2. Objectius.....	5
3. Metodologia	7
3.1. <i>Docking</i>	7
3.2. Dinàmica molecular.....	8
3.3. Aspectes ètics i de sostenibilitat.....	9
4. Resultats i discussió	10
4.1. Determinació dels centres d'unió de la proteïna tripsina	10
4.2. <i>Docking</i> de l'estructura cristal·logràfica	11
4.3. Dinàmica molecular	16
4.4. Simulació amb el lligand fora la proteïna	26
5. Conclusions.....	27
6. Bibliografia.....	29

1. Introducció

El camí que porta al descobriment d'un nou fàrmac és un procés llarg i costós que involucra coneixements de diverses àrees com la química, la biologia, o la medicina. Generalment, per iniciar aquest procés cal, en primer lloc, identificar una proteïna involucrada en algun procés biològic relacionat amb la malaltia que es pretén tractar. En segon lloc, s'ha d'identificar una molècula que sigui capaç de desactivar o activar la proteïna escollida. Seguidament, aquesta molècula (anomenada *hit*) s'optimitza per tal d'augmentar-ne la selectivitat i potència. Si la molècula en qüestió mostra les propietats adequades avança cap a la fase d'assajos clínics. Una de les principals dificultats és trobar una molècula dins tot l'espai molecular que sigui activa per a la proteïna seleccionada.

En els darrers anys, la química computacional ha guanyat molt de pes en les primeres fases del descobriment i disseny de nous fàrmacs. L'augment del poder de càlcul dels ordinadors i la millora dels algorismes de càlcul ha permès desenvolupar tècniques computacionals que permeten identificar de forma ràpida i poc costosa inhibidors per a la proteïna diana. Una de les tècniques computacionals més utilitzades per al descobriment de molècules actives és el *docking* molecular.¹ Aquest mètode es basa en utilitzar una estructura cristal·logràfica de la proteïna diana per fer-hi *docking* molecular de llibreries de compostos fins a trobar algun compost que presenti activitat. Els programes de *docking* prediuen l'orientació més favorable de la molècula respecta a la proteïna i calculen de forma generalment poc acurada l'energia d'interacció. En principi, com més elevada és la interacció entre la proteïna i el lligand més activa serà la molècula. El principal problema del *docking* és que no té en compte la flexibilitat de la proteïna i això porta a una gran quantitat de falsos positius i negatius.² Tenir en compte la flexibilitat (dinàmica) de la proteïna diana és clau per a poder identificar molècules actives.³

La dinàmica molecular (DM) és una eina computacional que permet estudiar els canvis de conformació que tenen lloc a la proteïna.⁴ Comparat amb el *docking* molecular la DM és una tècnica més precisa però més cara des del punt de vista computacional, fet que dificulta la seva aplicació en el camp del disseny de fàrmacs.⁵ Tot i això, com es

veurà tot seguit, incorporar els efectes dinàmics de la proteïna en el procés de disseny de fàrmacs permetrà millorar la comprensió del mecanisme d'inhibició de la proteïna diana.

Comprendre el procés d'unió entre un lligand i una proteïna és de gran interès pel disseny de fàrmacs. Degut a la mobilitat de la proteïna, l'estudi complet d'aquest procediment comprèn la caracterització de la unió lligand-proteïna a través de les diverses conformacions que pot presentar la proteïna al interaccionar amb el lligand. En aquest estudi, la utilització de dinàmica molecular (DM) combinada amb el *docking* molecular permetrà caracteritzar les zones d'unió de l'inhibidor benzamidina (BEN) en les diverses conformacions que pot adoptar la proteïna tripsina, que és la proteïna estudiada en aquest treball.

La tripsina és un enzim format per 245 aminoàcids i produït pel pàncrees encarregat d'hidrolitzar proteïnes. Un dels inhibidors més coneguts de la tripsina és la benzamidina que interacciona específicament en la cavitat corresponent al centre actiu de la proteïna (Figura 1). Aquesta cavitat està formada per un conjunt d'aminoàcids que interaccionen per enllaços d'hidrogen amb la benzamidina. A la Figura 1 es poden observar els aminoàcids que interaccionen amb el lligand, entre els quals el que juga un paper més important és l'Asp 171 .

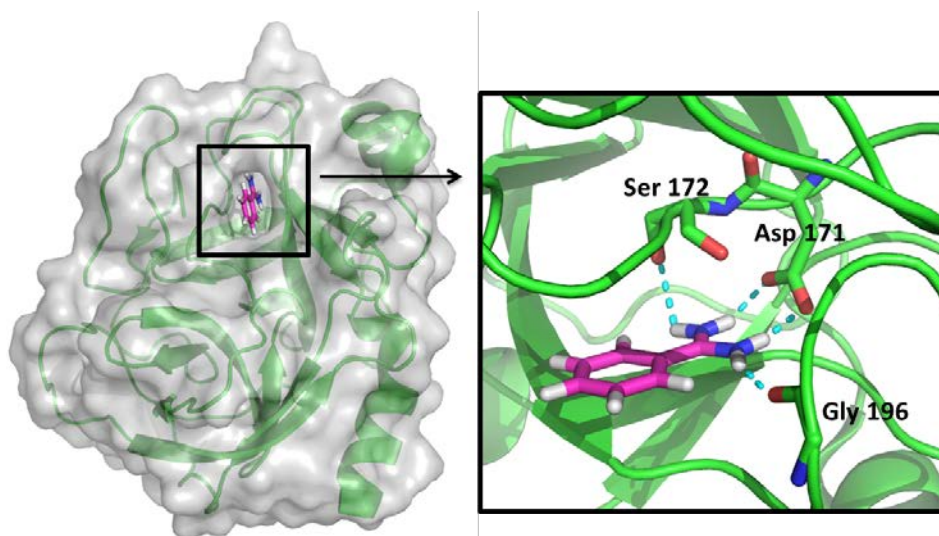


Figura 1. Estructura cristal·logràfica de la tripsina (PDB:3PTB) amb la benzamidina interaccionant al centre actiu.
<http://www.rcsb.org/pdb/explore/explore.do?structureId=3PTB>

Anteriors estudis de DM han determinat que la tripsina presenta diferents conformacions en l'estat *apo*, és a dir, sense cap lligand unit al centre actiu.⁶ Si la conformació del centre actiu permet la interacció entre la benzamidina i l'Asp 171 s'anomena conformació oberta o semi-oberta. En canvi, si aquesta interacció no és possible les conformacions es classifiquen com a tancades. Ambdues conformacions estan determinades pel moviment del *loop* marcat amb taronja a la Figura 2 que comprèn del residu 193 al 196.

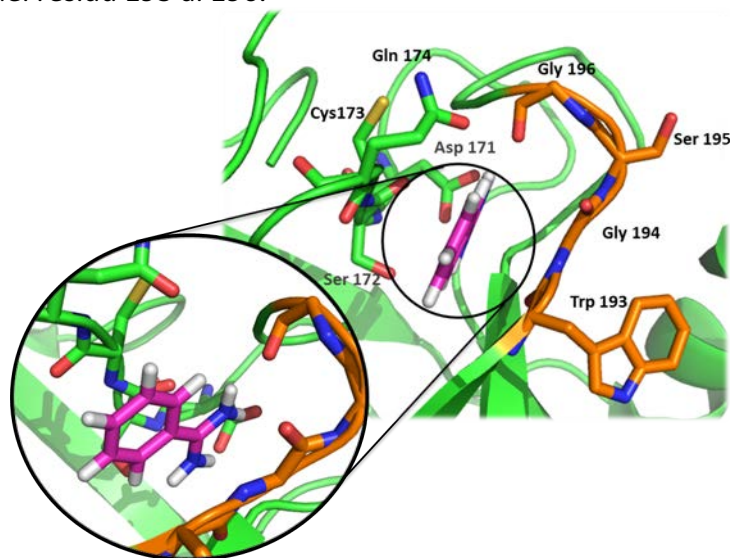


Figura 2. Els aminoàcids del centre actiu interaccionant amb la benzamidina. En taronja els aminoàcids 193-196 corresponents al *loop*.

A banda de la mobilitat conformacional del *loop* del centre actiu hi ha altres zones que també presenten variacions significatives al llarg de les simulacions de dinàmica molecular. A la Figura 3 s'observen diferents conformacions *apo* de la dinàmica molecular ressaltant les zones de major mobilitat extret de l'article d'en Frank Noé i col·laboradors.⁷ Aquestes zones corresponen als moviments que tenen lloc al centre actiu, al *loop* perifèric i al *loop* on s'uneix un ió calci.



Figura 3 Diferències entre les conformacions magenta i verda obtingudes a partir de

De Fabritiis et al.⁶ van realitzar una aproximació del procés d'unió de la benzamidina a la tripsina a partir de centenars de simulacions independents de DM. Van posicionar el lligand fora de la proteïna enmig del solvent i a partir de simulacions de dinàmica molecular es va intentar reproduir el procés d'unió a la proteïna. D'un total de 495 dinàmiques de 100 ns només el 37% resultaven favorables per la interacció del lligand al centre actiu. L'estudi d'aquestes simulacions van permetre obtenir un protocol dels possibles camins que podria seguir la benzamidina abans d'arribar a la zona d'interacció del centre actiu. La Figura 4 mostra la caracterització d'aquest protocol. Com es pot veure, per arribar al centre actiu (estat S4 a la Figura 4) la benzamidina pot interaccionar amb diferents parts de la proteïna: l'estat S0, envoltada només del solvent; l'estat S1, interaccionant amb els aminoàcids Y42, V72 i Y76 de la proteïna; l'estat S2, interaccionant amb els aminoàcids Y22, Y131 i P132; i l'estat S3, interaccionant amb l'aminoàcid T80, L81, Q155, T157 i M160. L'únic requisit és que per entrar al centre actiu (S4) cal passar abans per S3.

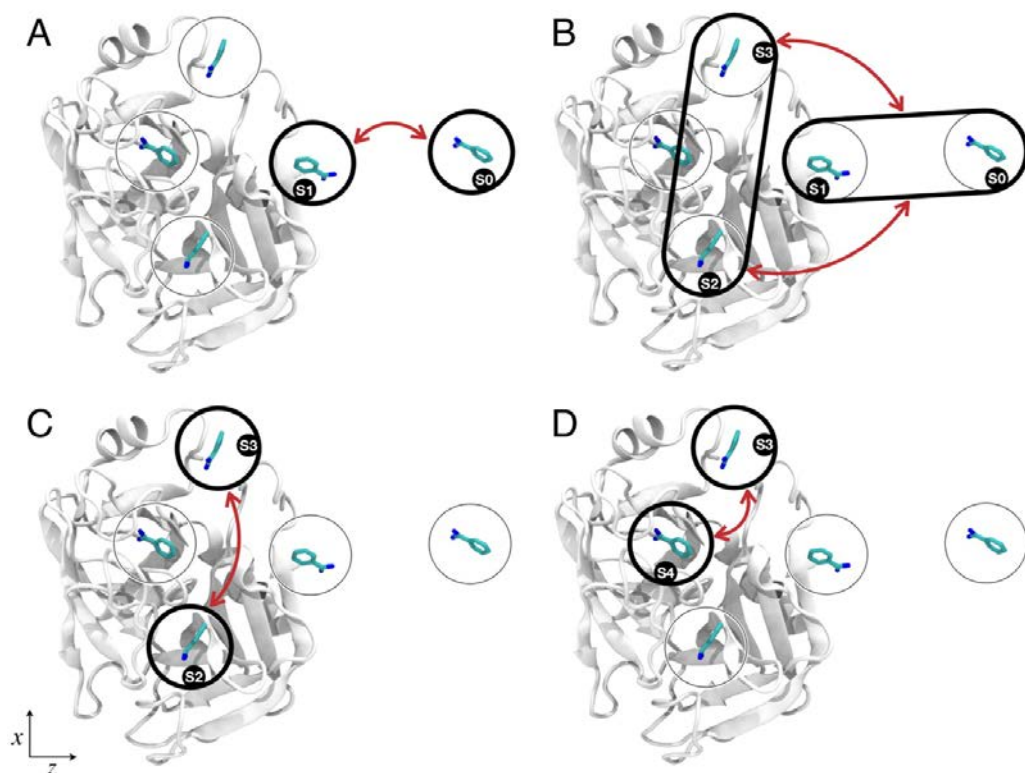


Figura 4. Possibles camins a seguir per la benzamidina abans d'interaccionar amb el centre actiu (S4). Imatge adaptada de la referència 6.

2. Objectius

In the last years, computational techniques have played a key role on the discovery and design of new drugs. In particular, molecular docking has been extensively used to identify compounds that present some activity for protein targets related to some particular diseases. Docking is a fast method because usually considers the protein as a rigid object. This fact significantly limits the efficacy of this computational technique. Proteins are dynamic objects that are able to interact with other molecules such as substrates, drugs, other proteins, or lipids. To establish interactions with other molecules, proteins usually alter their conformation to accommodate the binding partner. To properly study the process of protein-ligand binding we need to take into account both protein and ligand dynamics. Recently, Buch and co-workers reconstructed the complete protein-ligand binding process by means of molecular dynamics simulations for the enzyme-inhibitor complex trypsin-benzamidine.⁶ However, the computational cost associated with this technique limits its application.

The main aim of this work is to combine the advantages of molecular dynamics simulations with the advantages of molecular docking to design a protocol that allows us to study the protein-ligand binding process for the trypsin-benzamidine complex.

First, the protein dynamics of trypsin will be analysed by molecular dynamics simulations in both *apo* and complexed with benzamidine (*holo*) states. Molecular dynamics simulations will provide information about the most flexible parts of the protein and the importance of this flexibility on the active site cavity. The most relevant conformations of the simulation will be selected and the role of some particular aminoacids on the active site dynamics will be discussed.

Second, the binding process of benzamidine into the trypsin surface will be studied by means of molecular docking. To improve the efficacy and the information provided by molecular docking, several structures of trypsin extracted from MD simulations will be used to dock benzamidine. The results of docking will be used to rationalize which structures are in the closed or in the open conformation.

The ultimate goal of this project is to reconstruct the complete benzamidine binding process by means of molecular dynamics simulations. Benzamidine will be positioned

in the solvent and the regions of the trypsin surface visited prior to entering the active site will be discussed. Understanding how a ligand binds to a protein and how these interactions take place is of critical importance for the fields of biology and drug discovery.

3. Metodologia

Els programes de *docking* han esdevingut un eina molt important en el desenvolupament de fàrmacs.¹ Tot i això, una gran quantitat dels resultats obtinguts poden ser erronis ja que aquest càlcul no considera la flexibilitat de la proteïna. Per aquest motiu, l'estudi d'unió del lligand benzamidina a la proteïna tripsina s'ha dut a terme combinat simulacions de dinàmica molecular amb *docking* molecular. A continuació s'explicaran les eines utilitzades al realitzar els càlculs de *docking* i la simulació de dinàmica molecular.

3.1. Docking

Els càlculs de *docking* realitzats amb el programa *SwissDock*⁸ consideren l'estructura proteica rígida i orienten el lligand en diverses posicions de la superfície de la proteïna. Durant el *docking* el lligand pot patir rotacions per adaptar-se a l'estructura proteica i obtenir una interacció més favorable. Després d'orientar el lligand en les diverses zones de la proteïna es calculen les energies d'interacció utilitzant el camp de forces CHARMM⁹. El programa *SwissDock* presenta un tipus de *docking* més ràpid anomenat *very fast*, el qual calcula l'energia del sistema a partir de les energies intermoleculars de lligand i la proteïna i l'energia intermolecular entre el lligand i la proteïna (equació 1)¹⁰.

$$\Delta G = E_{intra}^{ligand} + E_{intra}^{recept} + E_{inter} \quad (1)$$

E_{intra}^{ligand} , energia intramolecular del lligand.

E_{intra}^{recept} , energia intramolecular de la proteïna.

E_{inter} , energia intermolecular entre el lligand i la proteïna.

Per calcular les energies del *docking* en mode *accurate* (equació 2)¹⁰, el programa considera l'efecte del solvent en les interaccions lligand-proteïna.

$$\Delta G = E_{intra}^{ligand} + E_{intra}^{recept} + E_{inter} + \Delta G_{elec.solv} + \sigma \times SASA \quad (2)$$

$\Delta G_{elec.solv}$, energia de solvatació.

σ , constant de valor 0.0072 kcal/(mol Å²).

SASA, Solvent accessible surface area.

Per visualitzar els resultats obtinguts dels càlculs de *dockings* s'han utilitzat els programes *Chimera*¹¹ i *PyMol*¹².

3.2. Dinàmica molecular

Una simulació de dinàmica molecular consisteix en l'estudi del moviment d'àtoms i molècules aplicant conceptes de la física clàssica com la segona llei de Newton.^{13, 14} Per a poder realitzar una simulació de dinàmica molecular d'una proteïna cal una estructura de partida guardada en format *pdb*. En el cas de la tripsina l'estructura cristal·logràfica escollida correspon al PDB 3PTB (<http://www.rcsb.org/pdb/explore/explore.do?structureId=3PTB>). Previ a iniciar la simulació de dinàmica molecular s'ha de realitzar un etapa de preparació del PDB.

- **Preparació de la proteïna 3PTB**

Utilitzant el servidor *H++*¹⁵ és protonen els aminoàcids de la proteïna besant-se amb el càlculs dels seus pka en l'entorn proteic. La protonació de les histidines es realitza manualment degut a la seva complexitat. En els càlculs de *docking* el programa *SwissDock* addicionava els hidrògens automàticament, per això, aquest pas previ no era necessari. Un cop protonat el fitxer PDB s'ha solvatat el sistema afegint 8130 aigües en forma de caixa i neutralitzat amb 9 ions clor.

- **Dinàmica molecular**

Per realitzar la dinàmica molecular s'utilitza el paquet de programes *AMBER*.¹⁶ Abans d'iniciar la simulació el programa segueix un protocol per equilibrar el sistema a temperatura ambient (300K). El primer pas consisteix en una minimització del sistema complet a 0K fins a trobar el primer mínim local. Seguidament comença un procés d'escalfament (*heating*) a volum constant per aconseguir la temperatura de 300K, és a dir, donar energia cinètica al sistema per iniciar el moviment dels àtoms. Finalment, es procedeix al procés d'equilibració on es permet a les aigües ocupar el volum que els hi correspon a 300K, aquest procés es realitza a pressió constant. Un cop equilibrat el sistema es comença la simulació de la dinàmica molecular a volum i temperatura constant. El camp de forces utilitzat pel programa per calcular l'energia del sistema és l'*AMBER FF14SB*.¹⁷ La trajectòria generada consta d'un total de 100,000,000

conformacions on cada conformació correspon a un pas de 2 fs obtenint en total una simulació de 200 ns. S'han realitzat dues simulació una amb l'estat *apo* i una altre amb l'estat *holo* de la proteïna tripsina. Finalment, s'ha dut a terme una simulació de 400 ns amb el lligand benzamidina situat fora la proteïna. Totes les simulacions de dinàmica molecular s'han realitzat utilitzant la versió d'AMBER per unitats de processament gràfic (GPU)¹⁸ implementada a l'ordinador Biocomp4 que pertany a l'Institut de Química Computacional i Catàlisi.

- **Clustering**

El *clustering* obtingut pel programa *Gromacs*¹⁹ consisteix en l'agrupació de les diverses estructures recollides a la dinàmica molecular basant-se en el càlcul de RMSD. El RMSD és una mesura que permet agrupar les diverses estructures segons la seva similitud. D'aquesta manera podem obtenir conformacions representatives de la dinàmica molecular.

- **Visualització**

El programa utilitzat per la visualització de les simulacions de dinàmica molecular és el *VMD* que ens permet tant visualitzar el fotogrames com realitzar estudis comparatius de distàncies, angles, etc.

3.3. Aspectes ètics i de sostenibilitat

Els càlculs computacionals per la síntesis de nous fàrmacs permeten economitzar l'etapa de disseny del fàrmac ja que no gastem recursos. D'aquesta manera s'evita sintetitzar lligands i realitzar assajos bioquímics per lligands que no son actius per la proteïna.

Els resultats obtinguts en aquest estudi són originals i no són copiats d'altres articles.

4. Resultats i discussió

La secció de resultats es divideix en els següents apartats. En primer lloc es realitza el *docking* de l'estructura cristal·logràfica 3PTB per l'inhibidor benzamidina i 10 molècules que presenten el grup benzamidina a l'estructura. En segon lloc es realitzen dues simulacions de dinàmica molecular la primera amb la proteïna en estat *apo* i la segona en estat *holo*. La diferència entre les dues simulacions és que l'estat *holo* de la proteïna té el lligand a l'interior del centre actiu en canvi l'estat *apo* no. Seguidament s'extreuen les conformacions rellevants d'aquestes simulacions i es realitza un *docking* del lligand benzamidina sobre aquestes estructures. Finalment es fa una breu introducció dels resultats obtinguts en la primera simulació de dinàmica molecular amb el lligand posicionat for la proteïna, interaccionant amb el solvent.

4.1. Determinació dels centres d'unió de la proteïna tripsina

Previ als càlculs de *docking* i les simulacions de dinàmica moleculars s'utilitza el servidor *FTMap*²⁰ per predir les possibles zones d'interacció de la proteïna tripsina. La Figura 5 presenta els resultats obtinguts en aquest estudi on la zona de màxima interacció, el centre actiu, està encerclada.

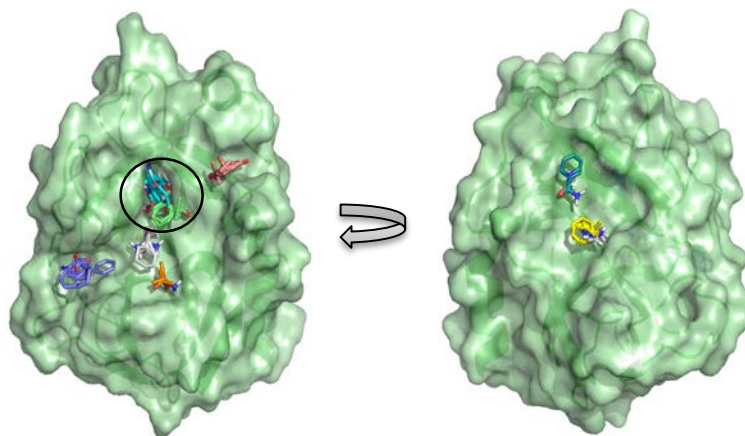


Figura 5. Possibles zones d'interacció de la proteïna tripsina predites pel servidor *FTMap*.

El centre actiu predit pel servidor és la zona on interacciona la benzamidina amb la proteïna tripsina. A més a més del centre actiu, el servidor *FTMap* també informa sobre la presència de possibles centres d'unió al·lostèrics. Com es pot veure a la Figura 5, alguns dels centres al·lostèrics predits pel servidor *FTMap* corresponen als estats S2 (lila) i S3 (rosa) predits per De Fabritiis⁶ mitjançant simulacions de dinàmica molecular (veure Figura 4).

4.2. Docking de l'estructura cristal·logràfica

Inicialment es realitza el *docking* sobre l'estructura cristal·logràfica (PDB:3PTB) per confirmar que el programa *SwissDock* prediu correctament la zona d'interacció entre el lligand benzamidina i la proteïna tripsina en la conformació òptima. Tal i com s'ha comentat a l'apartat de metodologia, es va efectuar un càlcul en mode acurat (*accurate*) i una altre en mode ràpid (*very fast*) per determinar si els dos tipus de càlcul que es poden realitzar amb el *SwissDock* presentaven diferències significatives.

A la Figura 6 és compara l'estructura cristal·logràfica (verda) amb l'estructura resultant del *docking* (blava), les quals presenten la mateixa conformació. El valor de RMSD entre les dues estructures és de 0.282 Å, fet que indica una gran similitud. A l'ampliació del centre actiu es pot corroborar que la benzamidina de l'estructura cristal·lografia (verda) està pràcticament solapada amb les benzamidina (BEN) dels *dockings* (magenta i lila).

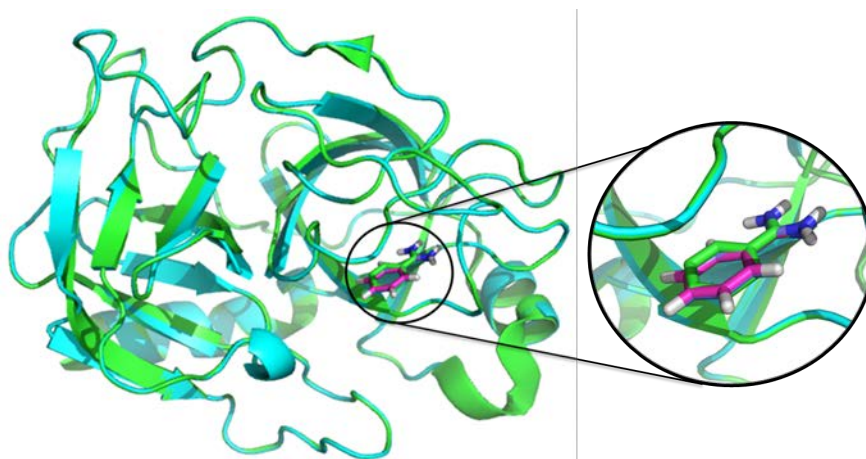


Figura 6. Comparativa de l'estructura cristal·logràfica (verd) amb la del Docking (blau). A l'ampliació del centre actiu: BEN estructura cristal·logràfica (verd), BEN docking very fast (lila), BEN docking accurate (magenta).

L'estructura de la benzamidina mostrada a la Figura 6 correspon a l'orientació de la molècula que presenta una energia d'interacció més favorable d'un total de 250 orientacions analitzades pel programa *SwissDock*. Aquests resultats validen l'ús del *docking* molecular per predir la interacció entre la tripsina i la benzamidina.

Al visualitzar un màxim solapament entre els càlculs acurat i ràpid es mesuren les distàncies d'enllaç de les diverses interaccions entre la benzamidina i la tripsina per comparar la seva similitud (Figura 7). Els aminoàcids responsables de la interacció del grup diamina amb l'estructura proteica són l'Asp 171, la Ser 172 i la Gly 196.

S'estableix un pont salí entre el grup carboxilat de l'Asp 171, carregat negativament, i la benzamidina, carregada positivament. Les distàncies es mesuren tant pel càlcul ràpid (Figura 7A) com en l'acurat (Figura 7B), respectivament. A més, l'anell aromàtic també presenta interaccions de tipus dipol-dipol induït i hidrofòbiques afavorint la interacció amb aquesta zona de la proteïna.

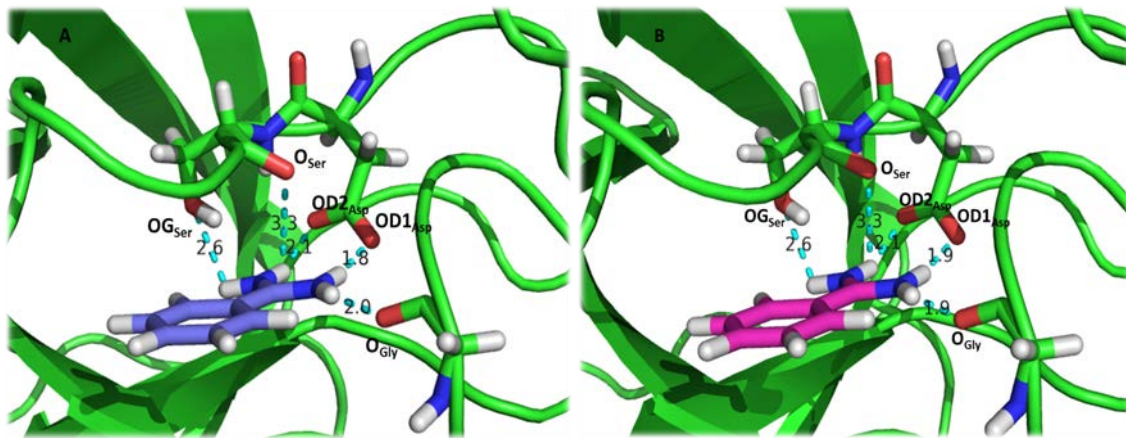


Figura 7 . Mesura de les distàncies entre el grup diamina del lligands i els aminoàcids del centre actiu. Imatge A càlcul ràpid imatge B càlcul acurate.

Comparant les distàncies dels càlculs en mode acurat i rapid s'observa que els valors difereixen en dues distàncies, però al fer la mitjana de tots els valors (Taula 1) aquesta diferència és imperceptible. Les distàncies de les diverses interaccions de la benzamidina amb els aminoàcids del centre actiu determinen l'energia del sistema, que en els dos casos s'obté un valor de -8,35 kcal/mol. A la següent taula es resumeixen els valors de l'energia d'interacció (ΔG en kcal/mol) i les distàncies intermoleculares obtinguts en el *docking*, que s'utilitzaran com a referència en propers estudis.

Taula 1. Valors d'energia d'interacció i distàncies del *docking* en mode very fast i accurate de l'estructura cristal·logràfica 3PTB.

Docking	ΔG						Mitjana
	(kcal/mol)	OD1 _{Asp} -H9 _{BEN}	OD2 _{Asp} -H8 _{BEN}	OG _{Ser} -H7 _{BEN}	O _{Ser} -H8 _{BEN}	O _{Gly} -H6 _{BEN}	
Very fast	-8.35	1.8	2.1	2.6	3.3	2	2.4
Accurate	-8.35	1.9	2.1	2.6	3.3	1.9	2.4

Al comprovar que el *docking* preveu correctament la zona on la benzamidina interacciona amb la proteïna es realitza un estudi complementari amb altres lligands que presenten una estructura similar a la benzamidina.

Les proteïnes són objectes dinàmics que canvien constantment de conformació, adaptant la seva forma a la naturalesa del lligand. El *docking* molecular considera a la proteïna com un objecte rígid i una estructura cristal·logràfica determinada no té perquè ser l'estructura adequada per predir la interacció amb tot tipus de lligands. Per tal de comprovar, si l'estructura cristal·logràfica corresponent al PDB 3PTB (cristal·litzada amb la benzamidina al centre actiu) és representativa no només per la benzamidina sinó també per molècules similars, s'escullen 10 molècules del Protein Data Bank²¹ que presentin el grup benzamidina dins la seva estructura per determinar-ne l'energia i la distància de la interacció. A la Figura 8 es recullen les 10 molècules escollides.

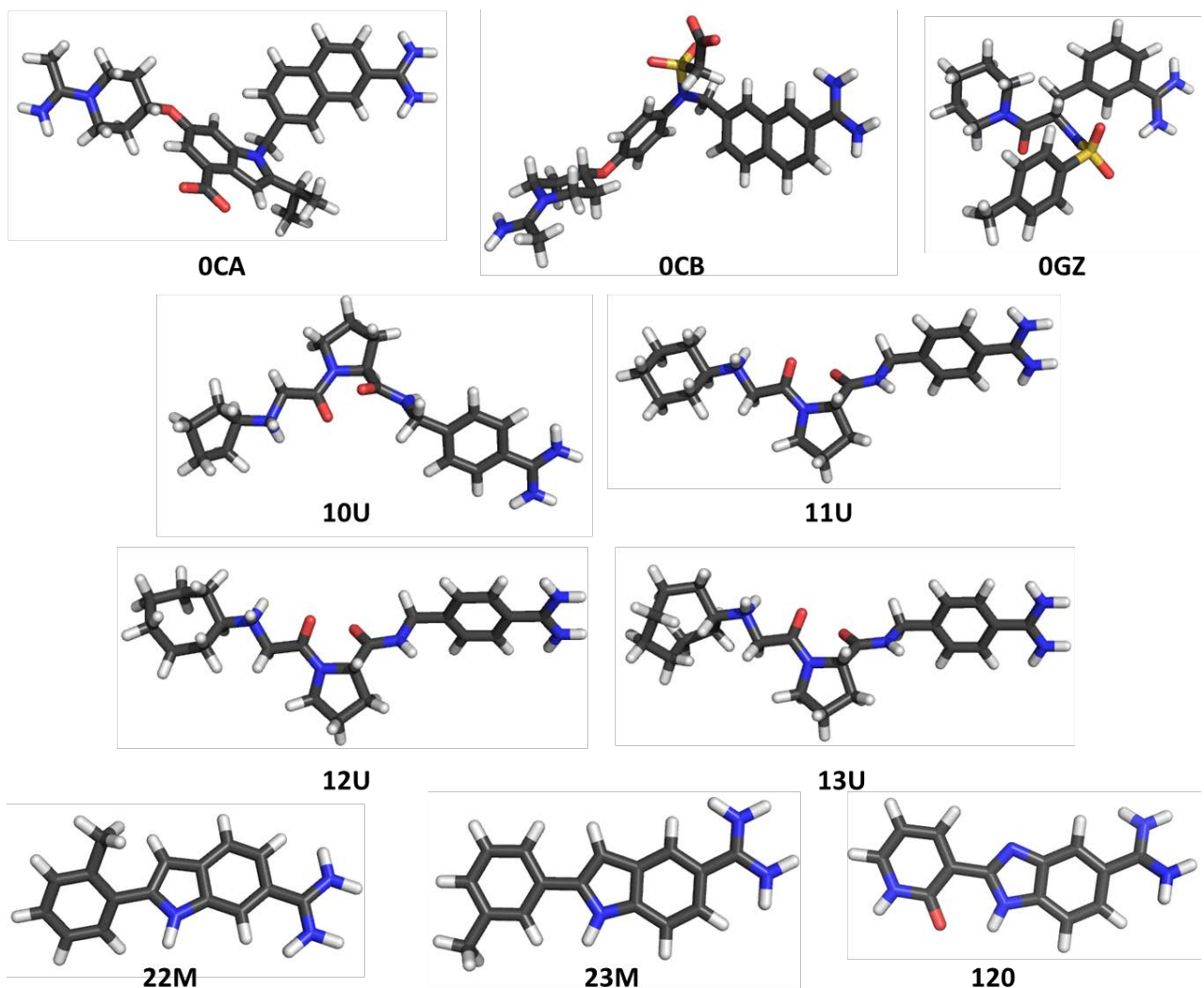


Figura 8. Estructura de les 10 molècules del Protein Data Bank que presenten el grup benzamidina.

Els valors energètics obtinguts amb el programa *SwissDock* per cada molècula es recullen a la Figura 9, juntament amb els valors de referència de la benzamidina.

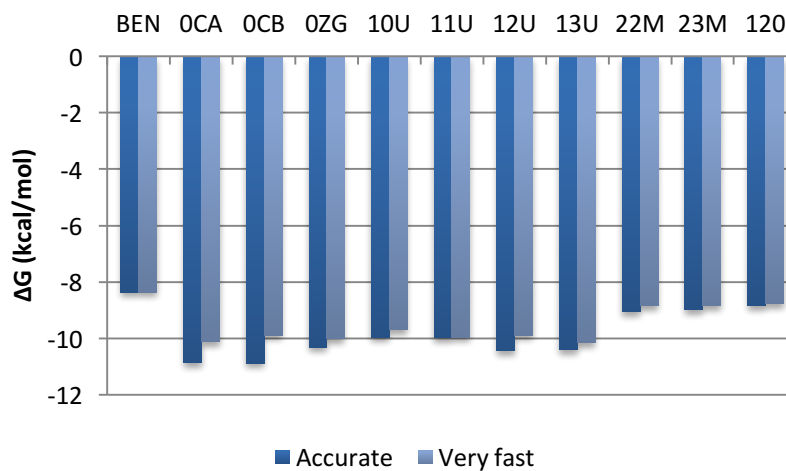


Figura 9. Valors d'energia d'interacció de les 10 molècules i la benzamidina calculats amb el programa *SwissDock*.

Si es comparen els valors obtinguts en els dos tipus de càlculs s'observa clarament que els càlculs en mode acurat són lleugerament superiors als realitzats en mode ràpid. Aquest fet és degut a que l'expressió utilitzada per a calcular l'energia d'interacció en el mode acurat té en compte d'una manera més completa les interaccions que tenen lloc entre el lligand i la proteïna (veure metodologia). Tot i això, els dos valors d'energia obtinguts en les 10 molècules són superiors als de la benzamidina en tots els casos.

Per poder trobar una explicació a aquest valors d'energia es fa una comparativa de les distàncies del grup benzamidina de les de les 10 molècules als aminoàcids del centre actiu amb què interaccionen (Figura 10). Els valors de la distància representats són la mitjana dels 5 enllaços d'hidrogen que presenta el lligand amb els aminoàcids descrits a la Taula 1. A excepció de la molècula OCA en el càlcul en mode ràpid, la resta de molècules interaccionen amb els aminoàcids del centre actiu amb els hidrògens que formen part del grup benzamidina.

Al comparar els dos tipus de càlculs (Figura 10) s'observa que els valors obtinguts en el càlcul en mode ràpid són molt variats i difereixen significativament dels càlculs més acurats i dels valors d'energia obtinguts per a la benzamidina. Aquest fet és degut a que el model ràpid del *docking* no permet predir de forma precisa la orientació i

l'energia d'interacció de la molècula respecte la tripsina. En canvi, si comparem els valors dels càlculs acurats la majoria presenten una distància d'interacció similar als de la benzamidina. Aleshores, si la interacció és similar l'elevada energia d'interacció es deguda a altres interaccions amb la resta de l'estructura molecular. Aquestes interaccions són majoritàriament de tipus hidrofòbic i tenen lloc amb les cadenes laterals dels aminoàcids que envolten el centre actiu tals com la molècula 10U que presenta interaccions hidrofòbiques entre l'anell aromàtic de l'estructura i la cadena lateral de l'aminoàcid Asn 97.

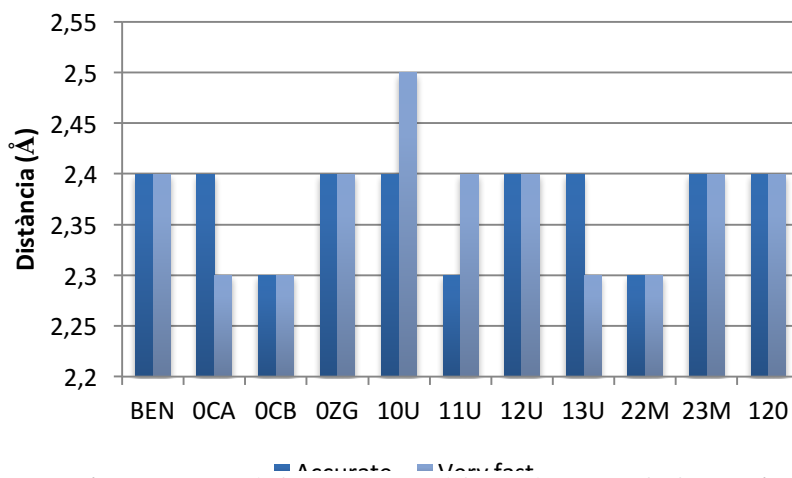


Figura 10. Distàncies mitjanes de les interaccions del grup diamina amb els aminoàcids del centre actiu. Excepció de la molècula OCA en el docking very fast.

Anteriorment s'ha mencionat que en tots els casos la interacció de les molècules amb el centre actiu és amb els hidrògens del grup diamina de la benzamidina a excepció de la molècula OCA. La Figura 11 mostra que el càlcul en mode acurat del *docking* realitzat amb el programa *SwissDock* ha predit correctament la zona d'interacció. És a dir, el grup benzamidina de la molècula OCA s'orienta correctament respecte al centre actiu. En canvi, en el càlcul amb mode ràpid la interacció amb el centre actiu té lloc per l'extrem oposat, o sigui, la benzamidina no s'orienta de forma adequada per afavorir les interaccions per enllaç d'hidrogen amb l'Asp 171. Aquest és un clar exemple que a vegades el càlcul en mode ràpid fa una predicció errònia tan de l'energia d'interacció com de l'orientació de la molècula i que per obtenir resultats més fiables s'ha de fer un *docking* acurat. A partir d'aquí, la resta de càlculs presentats seran tots realitzats en mode acurat.

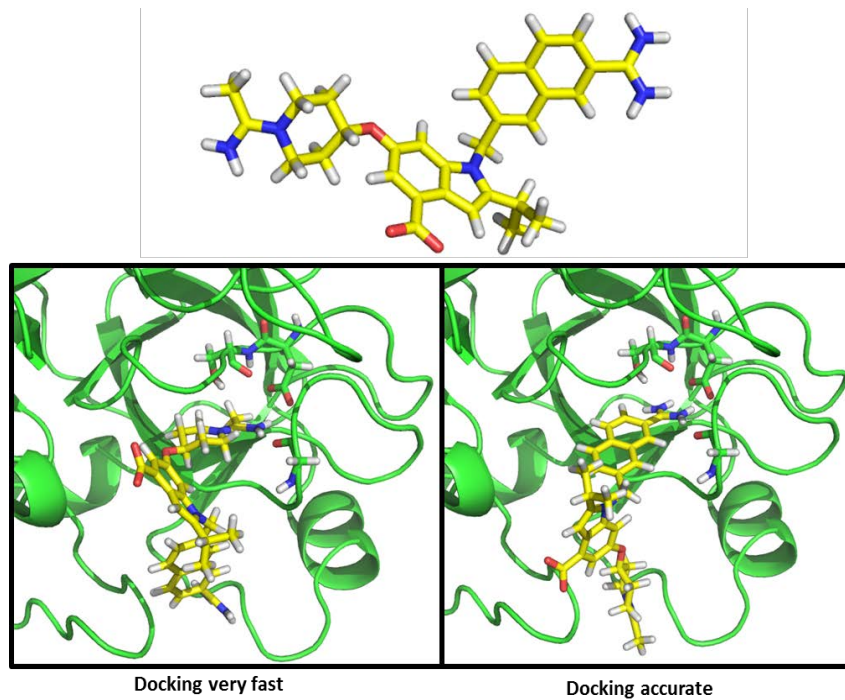


Figura 11. Comparativa de l'orientació de la molècula OCA pels dos modes de càlcul.

Per tant, l'estructura cristal·logràfica corresponent al PDB 3PTB és una estructura adequada per a realitzar estudis de *docking* amb lligands com la benzamidina o que contenen el grup benzamidina dins la seva estructura.

4.3. Dinàmica molecular

Les proteïnes són molècules altament flexibles capaces de canviar de conformació per permetre la interacció amb altres molècules com per exemple substrats, inhibidors o altres proteïnes. Les estructures cristal·logràfiques de les proteïnes només donen informació sobre l'estructura però no sobre la dinàmica de la proteïna. Entendre els processos dinàmics que tenen lloc a la proteïna és clau per entendre el mecanisme a través del qual es poden inhibir. Per estudiar la flexibilitat de la tripsina i la rellevància d'aquests moviments que tenen lloc per l'enllaç de la benzamidina es realitzaran simulacions de dinàmica molecular.

En total es realitzen dues dinàmiques moleculars la primera amb la proteïna en estat *apo* (proteïna sense lligands) i la segona amb la proteïna i el lligand benzamidina unit al centre actiu (estat *holo*), les dues simulacions corresponen a 200 ns. La trajectòria obtinguda conté un total de 100,000,000 conformacions que s'han guardat cada 2 fs. Aquesta gran quantitat d'informació s'ha de reduir seleccionant les conformacions més

rellevants. Per l'estudi de les diverses conformacions que presenta la proteïna al llarg de la dinàmica s'utilitza l'eina de *clustering* del programa *Gromacs* que selecciona les estructures més representatives de la simulació, anomenades clústers. Les diferents conformacions de la trajectòria s'agrupen segons la seva semblança estructural basat en el càlcul de l'RMSD.

A la Figura 12 s'observen les 5 conformacions més representatives de la proteïna en la dinàmica molecular *apo*. El clúster 1 correspon a la conformació més representativa de la simulació, el clúster 2 és la segona estructura més representativa i així successivament. Aquesta tècnica permet trobar conformacions rellevants de la simulació sense necessitat d'analitzar tota la trajectòria. Al comparar-les amb l'estructura cristal·logràfica (verda) s'observa que hi ha diversos *loops* que presenten mobilitat la llarg de la dinàmica.

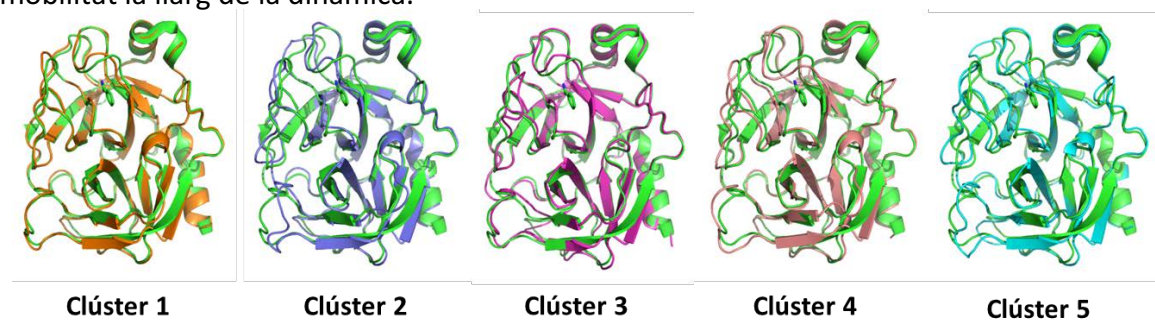


Figura 12. Cinc clústers obtinguts a partir de les simulacions de dinàmica molecular que representen els canvis conformationals de tota la proteïna comparats amb l'estructura cristal·logràfica.

Si comparem les 5 estructures conjuntament (Figura 13) s'observa que la major mobilitat de la proteïna es centra en dues zones principalment; en el *loop* de la zona inferior esquerra (corresponent al *loop* on s'uneix el calci mencionat a la Figura 3) i el *loop* del centre actiu. Ja que la benzamidina ha d'interaccionar preferentment amb el centre actiu realitzarem un estudi més detallat de la mobilitat d'aquest *loop*. Al ampliar la zona del centre actiu podem observar que aquest *loop* es troba en tres estats, un estat obert (conformacions verd (3PTB), taronja (clúster 1) i fúcsia (clúster 3)), un estat semi obert (rosa (clúster 4)), i en un estat tancat (lila (clúster 2) i blau (clúster 5)).

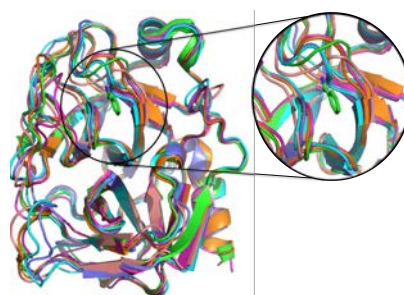


Figura 13. Comparativa dels 5 clústers i l'estructura cristal·logràfica (verda).

Aquest primer *clustering* ens informa sobre els canvis conformacionals que tenen lloc globalment a la proteïna. Però, per poder estudiar acuradament el moviment d'aquest *loop* en la dinàmica sense lligand es realitza un segon *clustering* basat en l'agrupació de les conformacions de la proteïna centrada només en les variacions que tenen lloc al centre actiu. Els aminoàcids seleccionats per a realitzar aquest *clustering* del centre actiu són els residus 171, 173, 173, 174, 194, 195, 196 i 197. Aquest *clustering* genera un total de 5 noves estructures representatives dels moviments que tenen lloc en els aminoàcids del centre actiu (Figura 14).

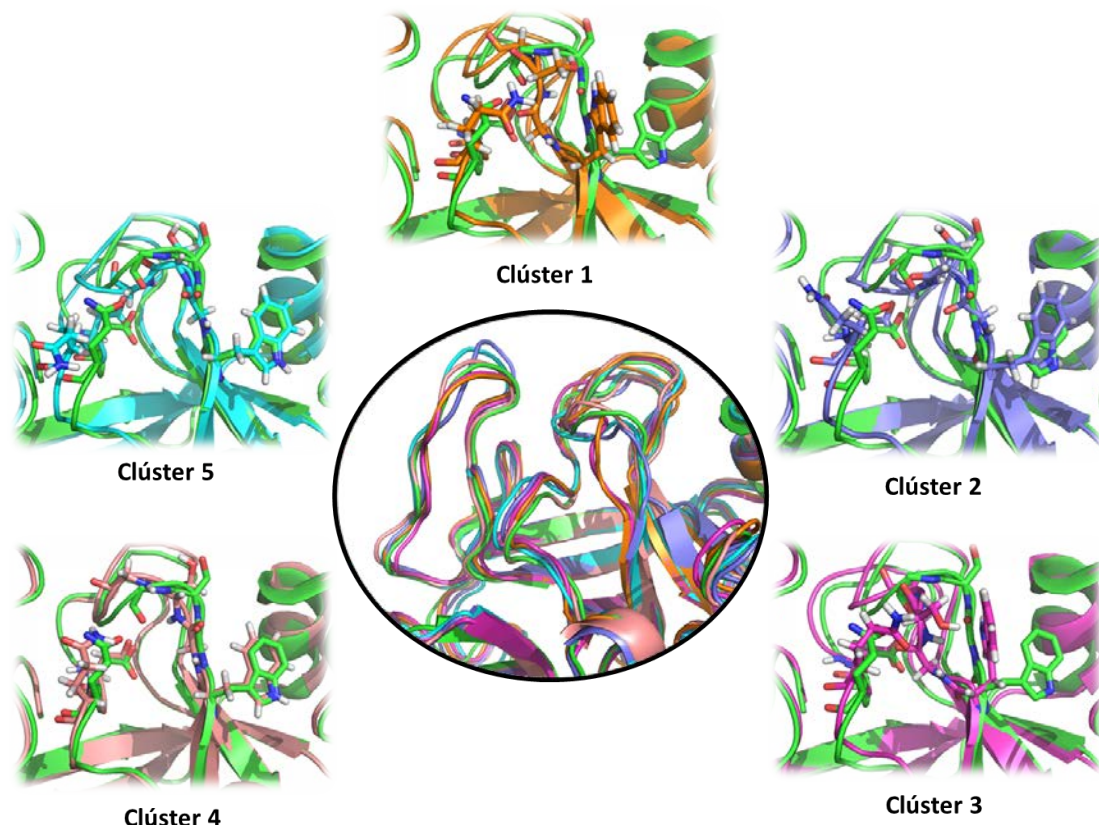


Figura 14. Comparativa conjunta dels clústers i l'estructura cristal·logràfica i comparativa individual de cada clúster amb l'estructura cristal·logràfica 3PTB (verd).

Al centre de la Figura 14 es comparen els 5 clústers juntament amb l'estructura cristal·logràfica (verda). En aquest cas, el *loop* del centre actiu només presenta dos estats: estat obert (clúster 1 i 3), estat tancat (clúster 2, 4 i 5). A les imatges individuals de cada clúster comparat amb la cristal·logràfica es podem observar diferents posicions dels aminoàcids en cada un dels estats. De les 5 estructures la que presenta una estructura més similar a la cristal·logràfica és el clúster 4. Dels aminoàcids que formen el centre actiu la Gln 174 i el Trp 193 són els que presenten una major mobilitat la llarg de la simulació de dinàmica molecular. Durant la simulació s'observa

que la Gln 174 interacciona per enllaços d'hidrogen amb els residus 194, 195 i 196 del *loop* provocant el seu tancament i observant la conformació tancada de la tripsina. A més, la Gln 174 interacciona amb els anells del Trp 193 provocant la modificació de l'angle díedre d'aquest aminoàcid (Figura 15).

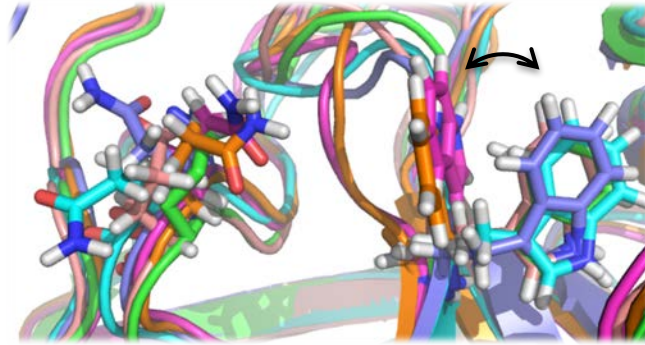
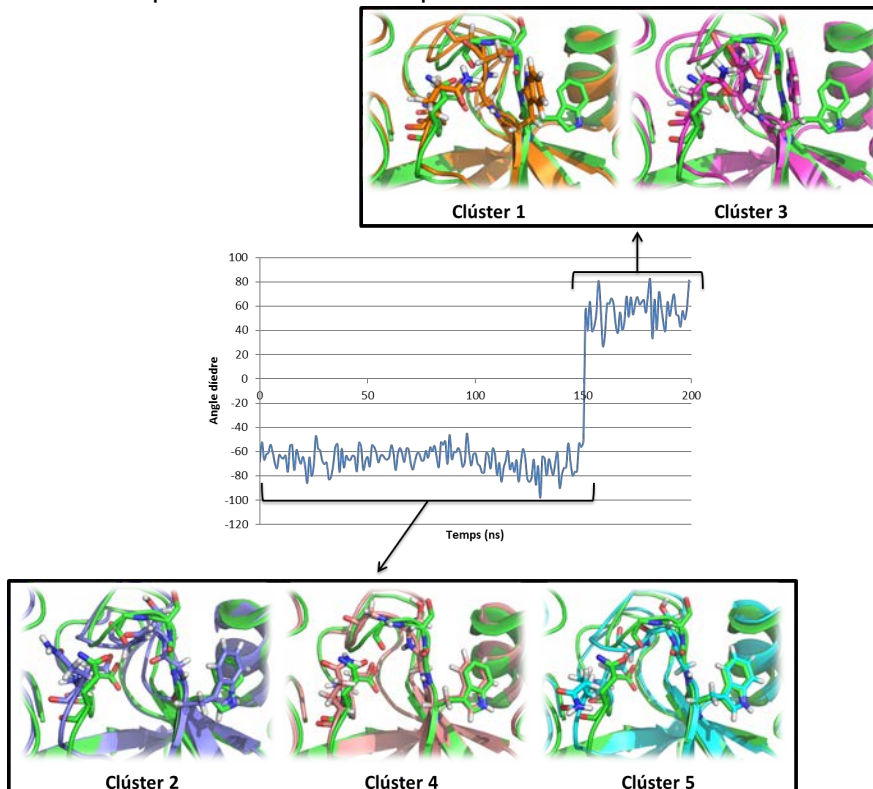


Figura 15. Comparativa dels 5 clústers i l'estructura cristal·logràfica (verda) de la interacció hidrofòbica entre la Gln 174 i el trp 193.

Per tal d'avaluar més extensivament aquest canvi de conformació del triptòfan s'ha representat la variació de l'angle díedre (corresponent a la rotació de l'enllaç format pel carboni alfa i carboni beta de Trp 193) al llarg de la dinàmica molecular *apo* (veure Figura 16). Es pot observar que el triptòfan es manté en conformació oberta durant els primers 150 ns de la simulació. A partir del nanosegon 150, el Trp 193 canvia de conformació fet que provoca el desplaçament del *loop* i el tancament del centre actiu. Aleshores, el clústers 2, 4 i 5 són representatius d'estructures anteriors als 150 ns i els clústers 1 i 3 corresponen a estructures posteriors.



Després d'estudiar detalladament el moviment dels aminoàcids a la dinàmica molecular es realitza el *docking* de la benzamidina sobre els 5 clústers obtinguts. L'objectiu és determinar sí la benzamidina es pot unir correctament al centre actiu tot i els canvis de conformació que aquest pateix. El *SwissDock* posiciona la benzamidina a 250 orientacions diferents sobre la superfície de la tripsina i per a cada una d'aquestes calcula l'energia d'interacció de manera que es pot establir un rànquing d'orientacions. La Figura 17 mostra la orientació més favorable predita pel *SwissDock* de la benzamidina de color blau i en color vermell la posició de la benzamidina més propera al centre actiu en el clústers que s'ha pogut observar que es produïa la interacció.

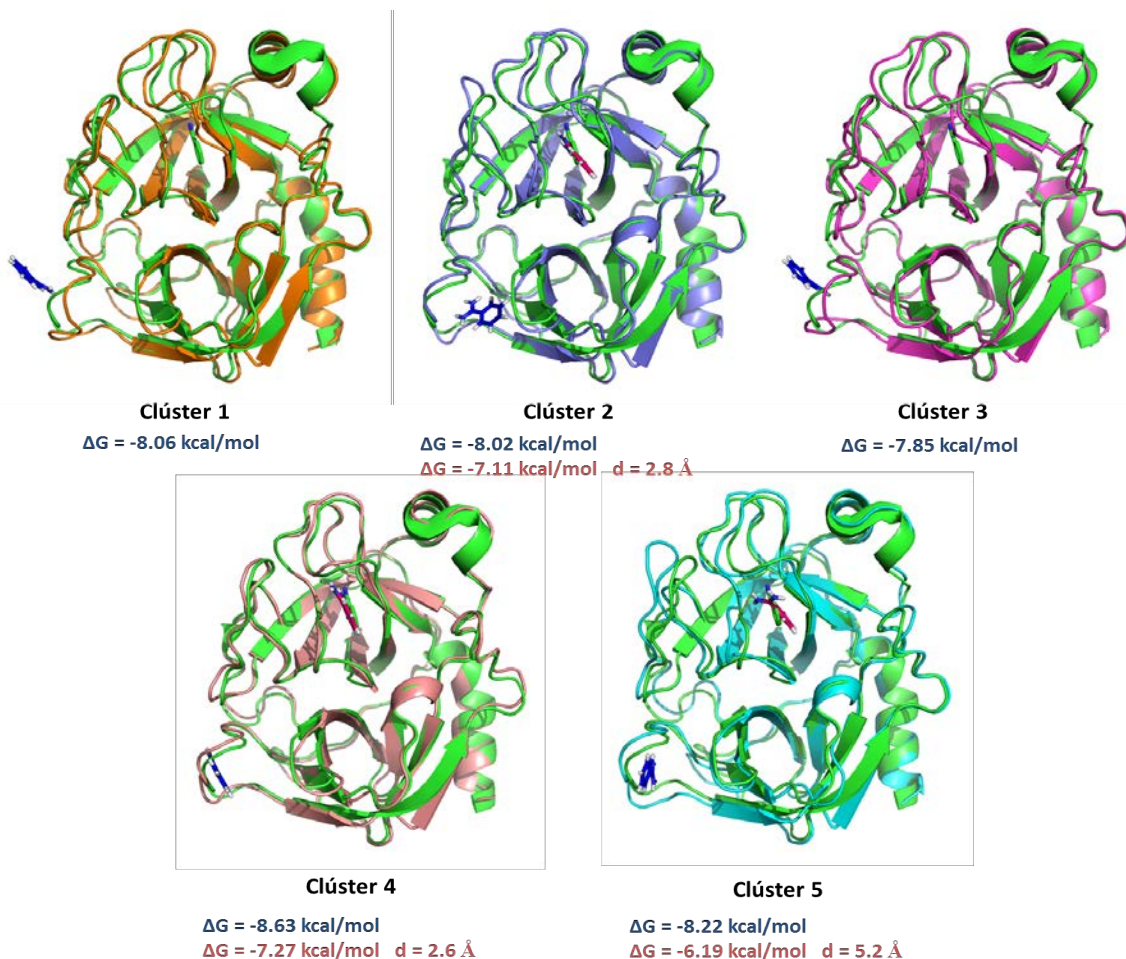
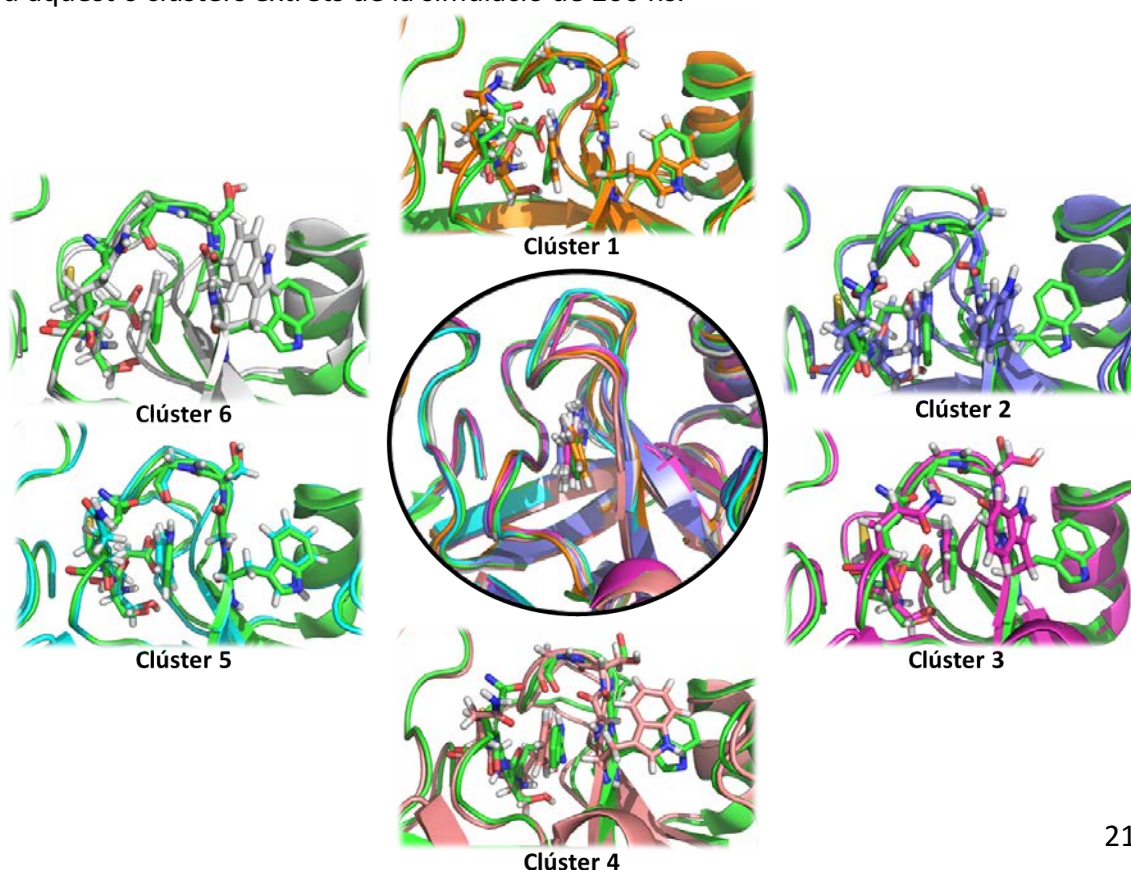


Figura 17. Orientació de la benzamidina en els dockings dels 5 clústers amb les corresponents energies d'interacció en kcal/mol en cas que la benzamidina s'uneixi al centre actiu també es presenta les distàncies mitjanes amb els aminoàcids del centre actiu. .

Aquests resultats corroboren els dos tipus de conformació que hem observat a la dinàmica. Els clústers 1 i 3 de conformació tancada el *docking* preveu que la benzamidina no interaccionarà amb el centre actiu ja que la posició del loop no permet l'entrada al lligand. En canvi, en els altres clústers la benzamidina interacciona amb el

centre actiu però amb una energia inferior a la obtinguda amb l'estructura cristal·logràfica (-8.35 kcal/mol). El clúster 4 de conformació oberta és el que presenta una energia inferior (-7.27 kcal/mol) degut a les interaccions hidrofòbiques de l'anell de la benzamidina amb l'estructura de la proteïna. Cal remarcar que en tots els sistemes la orientació més favorable, és a dir, l'orientació de la benzamidina que dóna una energia més negativa predita pel *docking* no es correspon a la zona del centre actiu. És interessant comprovar que en tots els casos la interacció més favorable es produeix amb el loop on s'enllaça el catió calci que és una zona rica en grups aspartats carregats negativament. Per tant, l'estructura més representativa per preveure la unió de la benzamidina amb el centre actiu és de moment la cristal·logràfica. Quan el loop del centre actiu es tanca la benzamidina no s'hi pot unir.

Per tal de comprendre amb més detall els moviments que tenen lloc al centre actiu i la importància que tenen aquests per a l'enllaç de la benzamidina, s'ha realitzat una segona dinàmica molecular amb la proteïna i el lligand benzamidina unit al centre actiu (simulació *holo*). Aquesta DM es porta a terme pel mateix procediment que en l'anterior cas. Al centrar-nos amb el moviment del centre actiu agrupem les conformacions en 6 clústers representatius. La Figura 18 mostra la comparativa d'aquest 6 clústers extrets de la simulació de 200 ns.



Al comparar conjuntament les 6 estructures d'aquesta s'observa que el *backbone* dels residus 194, 195 i 196 que formen el loop no ha canviat la conformació. Per tant, totes les estructures proteiques són obertes degut a la presència de la benzamidina al centre actiu. Al realitzar el *docking* sobre aquests sis clústers la benzamidina hauria de posicionar-se correctament al centre actiu. En tots el clústers s'observa una similitud amb l'estructura cristal·logràfica tot i que els aminoàcids Gln 174 i Trp 193 continuen presentant certa mobilitat. En aquesta simulació la Gln 174 només presenta interaccions hidrofòbiques amb el Trp 193 degut a la presència de la benzamidina a l'interior del centre actiu. Aquestes interaccions hidrofòbiques provoquen el canvi de conformació del Trp 193, tot i que no és una conformació gaire favorable degut a la presència de la benzamidina i el Trp 193 retorna a la posició original tal i com mostra la Figura 19, fent que el centre actiu de la proteïna es mantingui sempre obert.

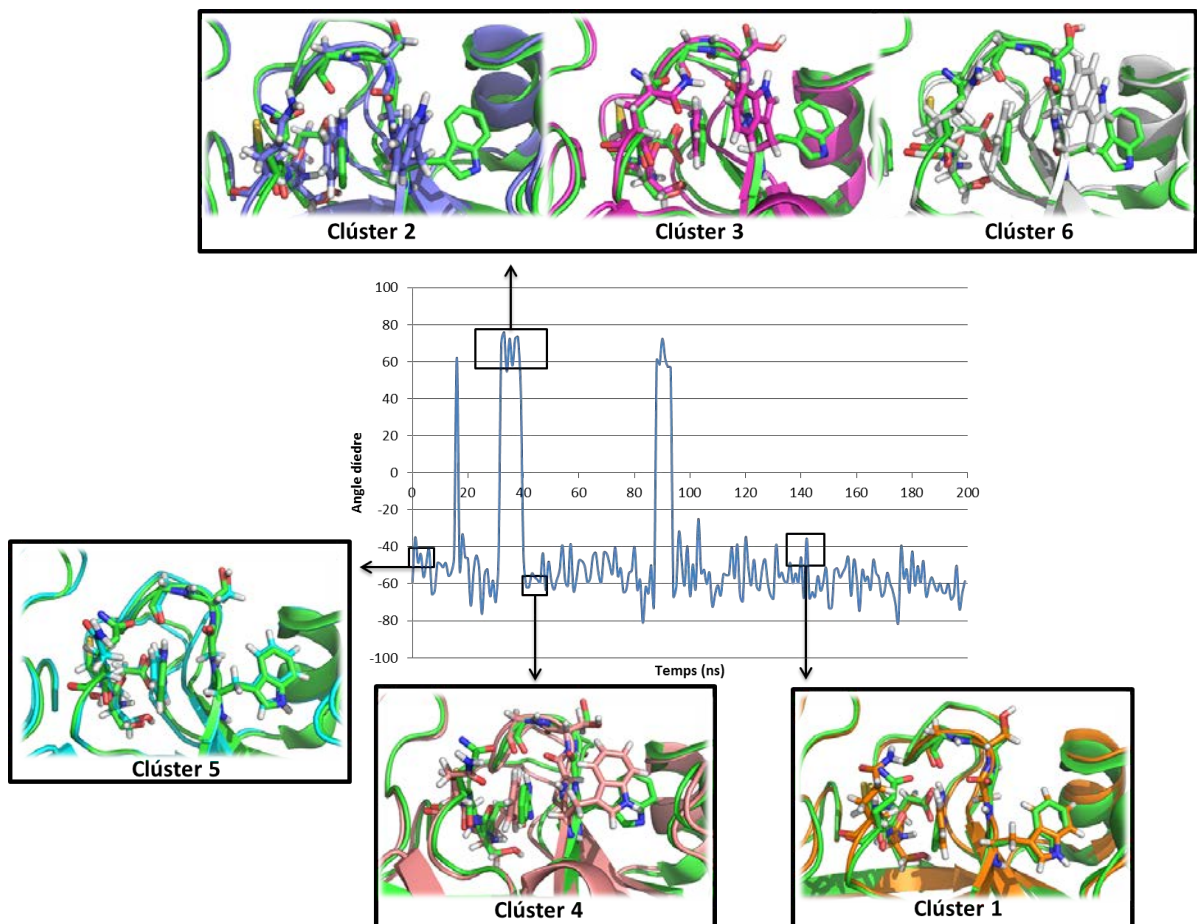


Figura 19. Variació de l'angle dièdre del Trp 193 (corresponent a la rotació de l'enllaç que formen el carboni alfa i carboni beta del Trp 193) al llarg de la simulació de la configuració en holo.

Anteriorment s'ha comentat que les 6 estructures són obertes ja que les estructures s'han extret de la simulació *holo* amb la benzamidina al interior del centre actiu. Per tant, es realitza el *docking* per comprovar si aquestes estructures són igual de vàlides que la cristal·logràfica alhora de predir la interacció. La Figura 20 mostra els resultats obtinguts d'aquest *docking* amb el programa *SwissDock*. Tot i que en tots els càlculs la benzamidina ha interaccionat amb els aminoàcids del centre actiu, no en tots els casos la interacció amb el centre actiu era la d'energia més favorable. Tot i això, els valors energètics en la majoria de les estructures són bastant pròxims als valors de la benzamidina a l'estructura cristal·logràfica.

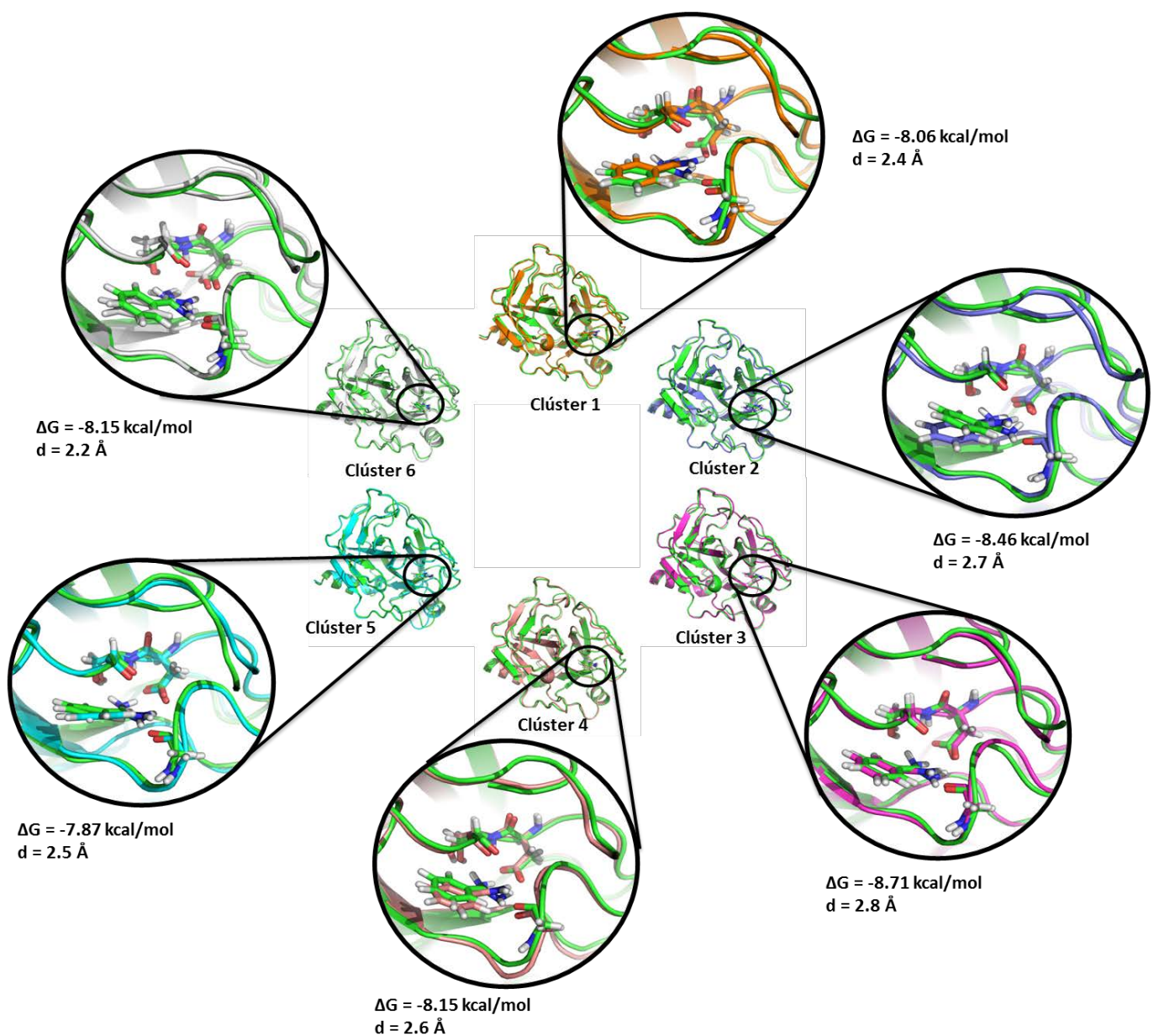


Figura 20. Predicció del programa *SwissDock* de la orientació de la benzamidina al centre actiu amb les corresponents energies d'interacció i distàncies mitjanes.

Per determinar si la distància de les interaccions per enllaç d'hidrogen són un factor important al determinar l'energia de cada estructura es realitza un gràfic (Figura 21). En el gràfic s'hi representen els valors obtinguts pels clústers que permetien la interacció de la benzamidina al centre actiu tant de la simulació de la proteïna en estat *apo* com *holo* (clúster 3 de la Figura 12, clúster 2, 4 i 5 de la Figura 14, els 6 clústers de la Figura 18, entre d'altres). Per tots aquests clústers s'ha calculat la mitjana de les distàncies tal i com s'ha fet a la Taula 1. En principi, s'espera que una menor distància mitjana amb els aminoàcids del centre actiu haurà de donar una energia d'interacció més favorable. S'observen dos grups, un format per distàncies mitjanes inferiors a 3.5Å que presenten energies més negatives i un altre grup format per distàncies mitjanes superiors a 3.5Å amb energies més altes. Per tant, si la interacció és menor a 3.5Å les energies són més negatives i més similars a les de referència (punt vermell). Tot i que s'observen els dos grups la diferència no es òbvia ja que apart d'aquestes interaccions també hi ha les interaccions hidrofòbiques amb l'anell aromàtic de la benzamidina que són les que permeten determinar el valor energètic final. Tot i això si que existeix certa correlació entre la distància mitjana i l'energia d'interacció.

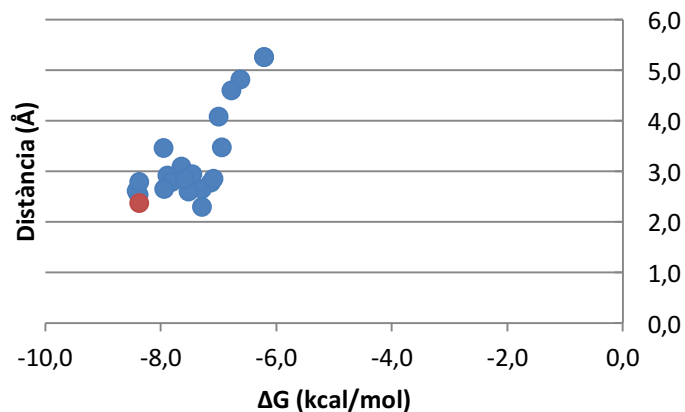


Figura 21. Energia d'interacció i distàncies mitjana dels dockings que presenten la interacció de la benzamidina amb el centre actiu. En vermell el valor de referència de la benzamidina.

Les energies obtingudes pel programa *SwissDock* per a les estructures extretes de la simulació de la proteïna en estat *holo* són bastant similars a les del *docking* de referència. Per aquest motiu, es va voler comprovar si alguna d'aquestes estructures podria ser millor que l'estructura cristal·logràfica per les 10 molècules d'estructura similar a la benzamidina (Figura 9). Per realitzar aquest estudi es va utilitzar el clúster 4 de la simulació *holo* ja que dels les 6 estructures es la que presenta un conformació i

energia més similar a la cristal·logràfica, amb el Trp193 en conformació oberta. Primer es van compara les energies obtingudes anteriorment per l'estructura cristal·logràfica amb les calculades pel clúster 4 (Figura 22).

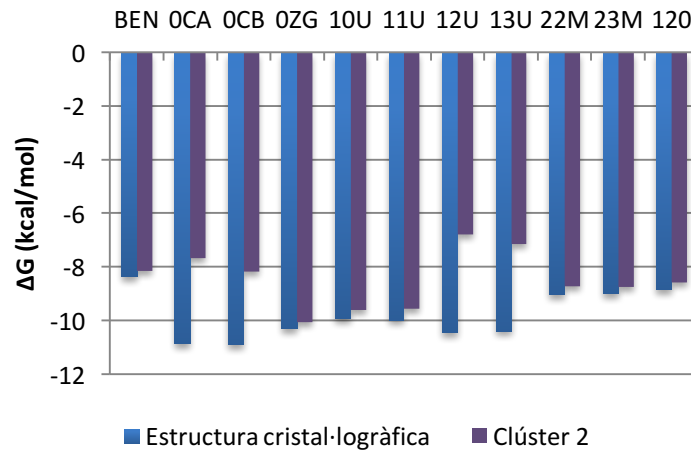


Figura 22. Valors d'energia d'interacció de les 10 molècules i la benzamidina calculats amb el programa SwissDock pel clúster 4.

En tots els casos les energies obtingudes en aquest darrer càlcul són inferiors a les de l'estructura cristal·logràfica. Per raonar aquests valors a més és realitza una gràfica comparant les distàncies d'enllaç entre les 10 molècules i el aminoàcids del centre actiu (Figura 23).

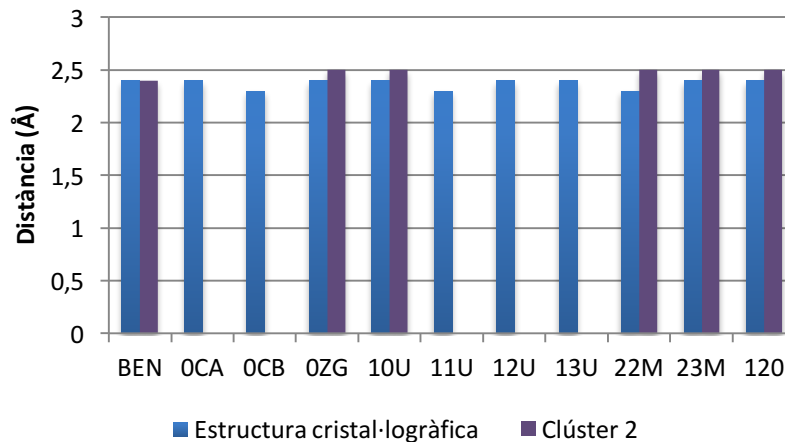


Figura 23. Distàncies mitjana de les interaccions del grup diamina amb els aminoàcids del centre actiu del clúster 4.

La Figura 23 mostra clarament que en alguns casos la molècula no s'ha unit al centre actiu per tant algunes energies no corresponen a la interacció de la molècula al centre actiu. A més a més, la distància mitjana és superior a la de l'estructura cristal·logràfica. Aleshores podem afirmar que aquesta estructura no és la més adequada per realitzar estudis de *docking* amb altres molècules que no siguin la benzamidina.

4.4. Simulació amb el lligand fora la proteïna

Per obtenir més informació sobre el procés d'unió del lligand a la proteïna es va realitzar una simulació col·locant el lligand fora del sistema proteic (enmig del solent) esperant que al cap de cert temps es produís la interacció amb el centre actiu. El protocol seguit va ser el mateix que el proposat per De Fabritiis i col·laboradors.⁶ En un total de 400 ns de simulació no s'ha observat interacció entre els aminoàcids del centre actiu i el lligand benzamidina, tot i que la benzamidina sí ha interaccionat momentàniament amb alguns dels estats que va predir De Fabritiis i col·laboradors.

La Figura 24 mostra les interaccions de la benzamidina amb les diversos estats predits a l'article.⁶ Cal remarcar que per interaccionar amb els estats S1, S2 i S3 sempre ha partit de l'estat S0. A més, cal fer especial menció a l'aminoàcid W193 el qual varia la seva conformació depenent de la interacció amb el lligand. Quan el lligand interacciona a l'estat S3 l'aminoàcid presenta un conformació tancada i al canviar-la el lligand marxa d'aquest estat. Futurs estudis de simulacions de dinàmica molecular, possiblement dinàmica accelerada, seran realitzat amb l'esperança d'observar la interacció de la benzamidina amb els aminoàcids del centre actiu.

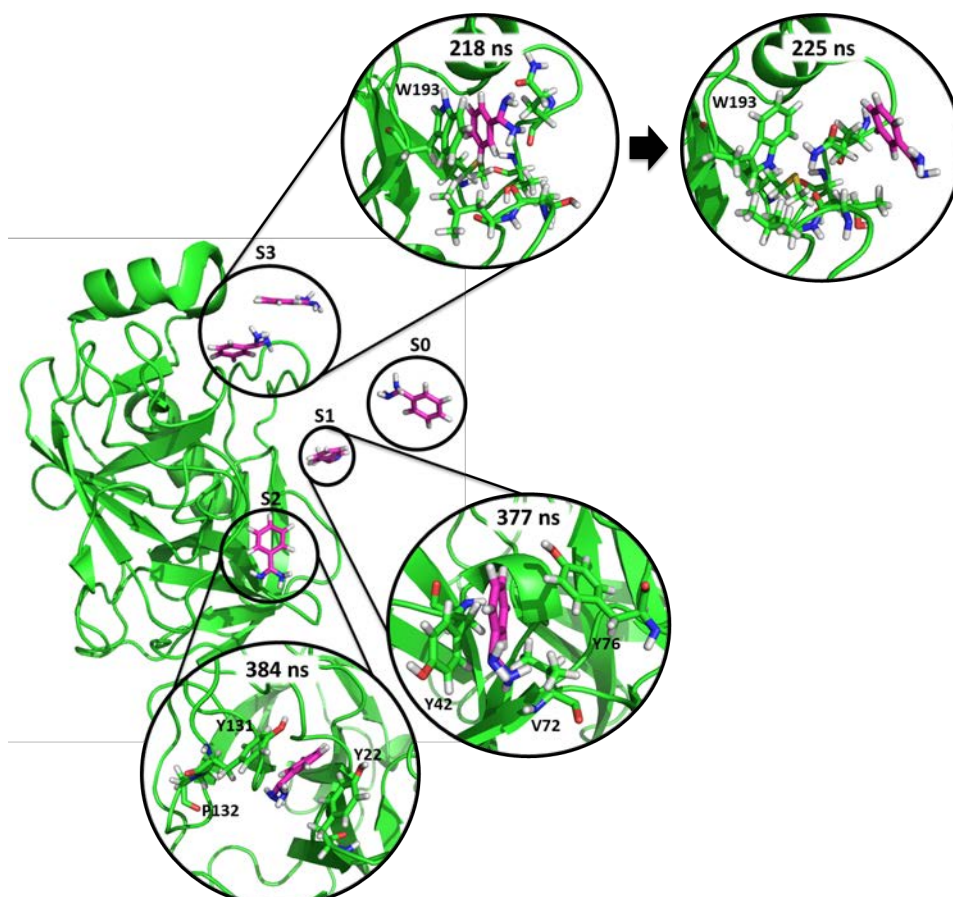


Figura 24. Caracterització dels estats S0, S1, S2 i S3 de la dinàmica molecular amb el corresponent nano segon de la simulació 26

5. Conclusions

Understanding how a ligand binds to a protein receptor and how these interactions take place is of critical importance for the field of drug discovery. The complexity of this association process is due to the highly dynamic nature of proteins. In last years, computational methods have played a key role on improving the understanding of the protein-ligand binding. Techniques such as molecular docking are commonly used in the pharmaceutical industry to discover and design novel drugs. However, it is difficult to account for protein dynamics when using docking methods, which clearly limits the efficacy of this kind of methods. On the other hand, molecular dynamics simulations are commonly used to study protein dynamics but at a high computational cost compared. This study combines the advantages of molecular dynamics simulations with the advantages of molecular docking to design a protocol for describing the protein-ligand binding process for the trypsin-benzamidine complex.

First, we have assessed the performance of molecular docking as implemented in the *SwissDock* server. To this end, we have docked benzamidine into the PDB 3PTB corresponding to the trypsin crystallographic structure co-crystallized with benzamidine in the active site. *SwissDock* has been able to properly describe the binding mode and the interaction energy of benzamidine for this particular structure in comparison to the crystal structure. The process has been repeated for 10 molecules that contain a benzamidine group in their structure. All of them bind to the active site according to docking results. This validates the use of molecular docking for the trypsin-benzamidine system. These docking results are the reference for the next steps.

Second, we have explored trypsin flexibility by means of molecular dynamics simulations. The molecular dynamics of trypsin have been performed in the *apo* state and in the *holo state* (with benzamidine). Then the most representative conformations of the simulation has been selected and analysed with quantitative and qualitative methods with special focus on the active site dynamics.

Third, to evaluate the implications of these conformational changes we have performed docking calculations on the most representative structures from molecular

dynamics simulations. On the one hand, the *apo* molecular dynamics simulation show three relevant conformations: open, semi-open and closed active site pocket. These three conformations are determined by the motion of Trp 193 and Gln 174. The open conformations (corresponding to cluster 1 and 3, Figura 12) and semi-open conformations (cluster 4, Figura 12) allow the binding of benzamidine in the active site, where benzamidine is able to interact with Asp 171. In the closed conformations (cluster 2 and 5, Figura 12), benzamidine can not interact with the active site, which confirms that the closed conformations of trypsin cannot bind benzamidine. However, in the three open conformations the interaction energies are lower than the one obtained from the crystallographic structure (8.35 kcal/mol). On the other hand, the selected conformations from the *holo* simulation are all found in the open conformation. That is, the loop of the active found is always found in the open conformation and benzamidine was able to bind the active site in all cases. The docking results for these structures show similar interaction energy values as compared to the crystallographic structure, which confirms that the protein is always in the open conformation in the *holo* state. However, the crystal structure is still the most representative for describing the interaction with other molecules with the benzamidine group.

Finally, we tried to reconstruct the complete benzamidine binding process by molecular dynamics simulations. We positioned benzamidine in the solvent region (S0) as a starting point and we started to simulation. The simulation shows that benzamidine interacts with states S1, S2 and S3 described previously by De Fabritiis but we have not observed the binding in the active site after 400 ns of simulation.

Future work

First, we are going to extend the length of molecular dynamics simulations to try to observe the binding of benzamidine in the active site. Then, accelerated molecular dynamics will be used to accelerate this binding process in order to be able to extend this methodology to more complex systems. Finally, we are going to repeat the process for the 10 molecules that contain benzamidine in their structure.

6. Bibliografia

1. 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.
2. 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.
3. Waszkowycz B, Clark DE, Gancia E. Outstanding challenges in protein – ligand *docking* and structure-based virtual screening. *WIREs Comput Mol Sci.* 2011;1:229-259. doi:10.1002/wcms.18.
4. 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.
5. Borhani DW, Shaw DE. The future of molecular dynamics simulations in drug discovery. *J Comput Aided Mol Des.* 2012;26(1):15-26. doi:10.1007/s10822-011-9517-y.
6. Buch I, Giorgino T, De Fabritiis G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc Natl Acad Sci U S A.* 2011;108(25):10184-10189. doi:10.1073/pnas.1103547108.
7. Plattner N, Noé F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat Commun.* 2015;6(May):7653. doi:10.1038/ncomms8653.
8. 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). doi:10.1093/nar/gkr366.
9. Grosdidier A, Zoete V, Michielin O. Fast *docking* using the CHARMM force field with EADock DSS. *J Comput Chem.* 2011;32(10):2149-2159. doi:10.1002/jcc.21797.
10. 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. doi:10.1002/prot.21367.
11. 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.

12. DeLano WL. The PyMOL Molecular Graphics System, Version 1.1. *Schrödinger LLC*. 2002:<http://www.pymol.org>. doi:10.1038/hr.2014.17.
13. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. *Nature*. 1977;267(5612):585-590. doi:10.1038/267585a0.
14. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nat Struct Biol*. 2002;9(9):646-652. doi:10.1038/nsb0902-646.
15. 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.
16. Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip Rev Comput Mol Sci*. 2013;3(2):198-210. doi:10.1002/wcms.1121.
17. 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.
18. 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.
19. Abraham MJ, Murtola T, Schulz R, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015;1-2:19-25. doi:10.1016/j.softx.2015.06.001.
20. Bohnuud T, Beglov D, Ngan CH, et al. Computational mapping reveals dramatic effect of Hoogsteen breathing on duplex DNA reactivity with formaldehyde. *Nucleic Acids Res*. 2012;40(16):7644-7652. doi:10.1093/nar/gks519.
21. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235-242. doi:10.1093/nar/28.1.235.