

Títol del treball:

Modelització de les diverses glicoformes de l'antigen prostàtic específic (PSA)

Estudiant: Aitor Palomanes Jimenez

Grau en Biotecnologia

Correu electrònic: aitor.palomanes@gmail.com

Tutor: Lluís Blancafort San Jose

Cotutor*:

Empresa / institució: Universitat de Girona

Vistiplau tutor (i cotutor*):

Nom del tutor: LLUÍS BLANCAFORT
Nom del cotutor*:
Empresa / institució: UdG
Correu(s) electrònic(s):
lluís.blancafort@udg.edu

*si hi ha un cotutor assignat

Data de dipòsit de la memòria a secretaria de coordinació:

RESUM

L'antigen prostàtic específic, conegut com PSA, és una glicoproteïna que es secreta en les cèl·lules de l'epiteli prostàtic i s'expressa en cèl·lules benignes i malignes. S'utilitza com a marcador del càncer de pròstata però té algunes limitacions com falsos positius ja que altres malalties poden augmentar els nivells de PSA. La solució a aquest problema de fiabilitat es pot trobar en la glicosilació de la PSA. Els estudis realitzats han demostrat que durant el càncer, la composició i la connectivitat de la cadena de sucres varia així com una reducció de la presència de fucosa en les cèl·lules tumorals. Així es podrien desenvolupar marcadors més eficients que permetessin distingir els casos de càncer de pròstata d'altres malalties com la hiperplàsia benigna de pròstata.

En el treball que es presenta s'ha realitzat una modelització de la proteïna PSA així com una simulació computacional de les diferents glicoformes que s'han generat. L'estructura original de la proteïna s'ha obtingut del PDB (3QUM) i s'ha modificat per obtenir tres glicoformes diferents. La primera glicoforma (G1) ha estat desglicosilada mentre que a les altres dos glicoformes (G2 i G3) se'ls ha afegit una glicosilació biantenària. La diferència entre les dos glicoformes biantenàries es troba en la presència o no de fucosa.

Per realitzar la simulació computacional de les proteïnes d'estudi s'ha utilitzat la mecànica molecular clàssica així com la dinàmica molecular. Després de finalitzar la dinàmica s'ha fet un anàlisi dels resultats calculant el rmsd, per observar la convergència de les simulacions, i els ponts d'hidrogen establerts entre la cadena de sucres i la proteïna, per observar quines interaccions són rellevants per l'estabilitat de la proteïna. També s'ha analitzat la distància dels ponts d'hidrogen al llarg del temps.

El temps de les tres simulacions que s'han realitzat es comprenen entre 38 i 57 ns. S'ha observat que en aquest temps, les simulacions no han convergit i les proteïnes no han arribat a la conformació més estable. Per altre banda en l'anàlisi dels ponts d'hidrogen, s'han trobat dos enllaços en la glicoforma amb fucosa que són més forts que els tres observats en la glicoforma sense fucosa. Pel que fa a l'anàlisi de les distàncies, s'ha observat que la G2 i la G3 pateixen un canvi conformacional cap al final de la simulació.

RESUMEN

El antígeno prostático específico, conocido como PSA, es una glicoproteína que se secreta en las células del epitelio prostático y se expresa en células benignas y malignas. Se utiliza como un marcador del cáncer de próstata pero tiene algunas limitaciones como falsos positivos debido a otras enfermedades pueden aumentar los niveles de PSA. La solución a este problema de fiabilidad se puede encontrar en la glicosilación de la PSA. Los estudios realizados han demostrado que durante el cáncer, la composición y la conectividad de la cadena de azúcares varían así como se da una reducción en la presencia de células tumorales en fucosa. Así se podría desarrollar marcadores más eficaces para distinguir los casos de cáncer de próstata de otras enfermedades como la hiperplasia prostática benigna.

En el trabajo que se presenta se ha realizado una modelización de la proteína PSA así como una simulación computacional de las diferentes glicoformas que se han generado. La estructura original de la proteína se ha obtenido del PDB (3QUM) y se ha modificado para obtener tres glicoformas diferentes. La primera glicoforma (G1) ha sido desglicosilada mientras que a las otras dos glicoformas (G2 y G3) se les ha añadido una glicosilación biantenaria. La diferencia entre las dos glicoformas biantenarias reside en la presencia o no de fucosa.

Para realizar la simulación computacional de las proteínas del estudio se ha utilizado la mecánica molecular clásica así como la dinámica molecular. Después de finalizar la dinámica se ha realizado un análisis de los resultados calculando el rmsd, para observar la convergencia de las simulaciones, y los puentes de hidrógeno entre la cadena de azúcares y la proteína, para observar que interacciones son relevantes para la estabilidad de la proteína. También se ha analizado la distancia de los puentes de hidrógeno en función del tiempo.

El tiempo de las tres simulaciones que se han realizado se comprende entre 38 y 57 ns. Se ha observado que en este tiempo, las simulaciones no han convergido y las proteínas no han llegado a la conformación más estable. Por otro lado, en el análisis de los puentes de hidrógeno se han encontrado dos enlaces en la glicoforma con fucosa que son más fuertes que los tres observados en la glicoforma sin fucosa. Respecto al análisis de las distancias, se ha observado que la G2 y la G3 sufren un cambio conformacional hacia el final de la simulación.

ABSTRACT

The prostate-specific antigen, known as PSA, is a glycoprotein that is secreted in the prostatic epithelial cells and is expressed both in benign and malignant cells. It is used as a marker of the prostate cancer but has some limitations as false positives since other diseases can increase PSA levels too. The solution to this reliability problem can be found in the PSA glycosylation. Studies have shown that in cancer, the composition and connectivity of the sugar chain change as well as the presence of fucose is reduced. Studying this changes can lead to the development of more efficient markers to distinguish cases of prostate cancer from other diseases such as benign prostatic hyperplasia.

In the work presented a PSA modeling has been made as well as a computational simulation of different glycoforms generated. The original structure of the protein has been obtained from the PDB (3QUM) and has been modified to obtain three different glycoforms. The first glycoform (G1) has been deglycosylated while the other two glycoforms (G2 and G3) have been glycosylated with a biantennary glycosylation. The difference between the two biantennary glycoforms is in the presence or absence of fucose.

To perform the computational simulation of the proteins, classical molecular mechanics has been used as well as the molecular dynamics. After completing the dynamics, an analysis of the rmsd, to study the convergence of the simulations, and the hydrogen bonds, to observe the relevant interactions for protein stability, has been made. The hydrogen bonds distances have also been analyzed.

The time of the three simulations that have been made is comprised between 38 and 57 ns. In the elapsed time, the simulations have not converged and proteins have not achieved the most stable conformation. Concerning the hydrogen bonds analysis, two interactions have been found in the glycoform with fucose that are stronger than the observed in the glycoform without fucose. Finally, the analysis of the hydrogen bonds distances has showed that the G2 and G3 suffer a conformational change toward the end of the simulation.

ÍNDIX

1. INTRODUCCIÓ	1
1.1. Antigen prostàtic específic	1
1.2. Relació amb el càncer de pròstata	1
1.3. Glicosilació	2
1.3.1. N-glicosilació	3
1.4. Sistemes d'estudi	4
2. OBJECTIVES	5
3. METODOLOGIA	6
3.1. Simulació computacional	6
3.2. Mecànica Molecular	7
3.2.1. Camp de forces	7
3.3. Dinàmica molecular	8
3.4. Simulació dels sistemes d'estudi	9
3.4.1. Preparació dels sistemes d'estudi	9
3.4.2. Minimització del sistema d'estudi	10
3.4.3. Dinàmica molecular del sistema d'estudi	10
3.4.4. Rmsd	11
3.4.5. Anàlisi dels ponts d'hidrogen i les distàncies	12
4. RESULTATS	13
4.1. Rmsd i convergència de les trajectòries	13
4.2. Anàlisi dels ponts d'hidrogen	14
4.2.1. Evolució de les distàncies	14
5. DISCUSSIÓ	17
6. CONCLUSION	18
7. BIBLIOGRAFIA	19

1. INTRODUCCIÓ

1.1. Antigen prostàtic específic

L'antigen prostàtic específic o PSA és una proteasa serina que es sintetitza a les cèl·lules de l'epiteli prostàtic. La PSA és una glicoproteïna de 28.400 Da que compren 237 residus i la seva funció és liquar el semen coagulat, així els espermatozoides poden escapar i progressar per fertilitzar l'òvul en el tracte reproductor femení [1]. Es secreta en el plasma seminal on es troba a una concentració entre 0,5 – 3 mg/ml. La pròstata està recoberta del lumen prostàtic, que si es danya o es trenca per alguna malaltia pot permetre el pas de la PSA als capil·lars sanguinis [2].

La PSA forma part de les cal·licreïnes, que són un subgrup de proteases serines que tenen diferents funcions fisiològiques. El gen que codifica per aquesta proteïna, també coneguda com KLK3, es troba en el cromosoma 19 en la localització 19q13.41. Es troba a prop dels gens que codifiquen per altres cal·licreïnes com KLK2, KLK4 o KLK15 [3],[4].

1.2. Relació amb el càncer de pròstata

El càncer de pròstata és un dels càncers més mortals per homes en tot el món i la segona causa de mort relacionada amb càncer en els Estats Units. Tot i així, és una malaltia curable si es detecta a temps. Per aquest motiu els homes de més de 50 anys es sotmeten a revisions de càncer de pròstata. L'examinació rectal digital (ERD) i l'anàlisi dels nivells de PSA són mètodes ben establerts per fer un cribratge i poder diagnosticar aquest tipus de càncer. Però el ERD és un mètode poc sensible quan el càncer es troba en una etapa primerenca i més sensible en etapes tardanes [5].

La PSA s'expressa en cèl·lules malignes com les tumorals i s'allibera al sèrum incrementant el seu nivell, permetent la detecció i diagnosi del càncer de pròstata. La PSA total ha permès incrementar l'habilitat de detectar i tractar els casos de càncer de pròstata en etapes primerenques. Però l'ús de la PSA com a marcador tumoral és un mètode que presenta algunes limitacions ja que l'elevació dels seus nivells també es pot donar en pacients amb hiperplàsia benigna de pròstata (HBP) [6]. Així doncs, es dona una falta d'especificitat sobretot en el rang de 4 a 10 ng/ml, on el càncer de pròstata es troba únicament en el 25% del casos [7]. En aquest rang els mètodes no poden diferenciar clarament entre càncer de pròstata i altres malalties de pròstata benignes.

D'aquesta forma els cribratges del nivell de PSA redueixen la mortalitat dels pacients amb càncer de pròstata però donen un elevat nombre de falsos positius que impliquen biòpsies de pròstata innecessàries. A més, els homes grans tendeixen a tindre uns nivells de PSA més elevats que no corresponen a cap malaltia de la pròstata [5].

Per aquestes limitacions, s'estan desenvolupant altres mètodes per tal d'augmentar la sensibilitat i tenir un poder de resolució major alhora de decidir si un pacient pateix o no càncer de pròstata. Algunes variacions del mètode poden ser l'anàlisi de PSA associat a l'edat, la densitat de la PSA, el percentatge de PSA lliure i la PSA complexa [7].

Un altre variant del mètode és l'estudi de la glicosilació de la PSA, que és l'alternativa en la que ens centrarem en aquest treball. La PSA és una glicoproteïna amb una única N-glicosilació en l'asparagina 61 (Asn 61). El procés de glicosilació cel·lular, especialment la sialilació, es veu pertorbat en els tumors i aquests canvis poden permetre la distinció entre la PSA d'origen normal i la d'origen tumoral. L'espectrometria de masses mostra que la PSA del càncer de pròstata presenta un contingut major d'àcid siàlic α -2,3 respecte àcid siàlic α -2,6 i una disminució del contingut de fucosa core [8]. A més, la PSA del plasma seminal presenta estructures d'oligosacàrids de tipus complex biantenàries amb àcid siàlic mentre que la PSA del càncer de pròstata presenta una barreja d'oligosacàrids biantenaris i triantenaris amb àcid siàlic [2].

1.3. Glicosilació

Les modificacions post-transcripcionals, com la glicosilació, juguen un paper molt important en el control de l'activitat enzimàtica. A més, la majoria de membranes plasmàtiques i proteïnes secretores són glicosilades [9]. La glicosilació es basa en el procés d'addició de sucres a una proteïna i té una influència important en el plegament, estabilitat, reconeixement molecular i activitat específica de les proteïnes.

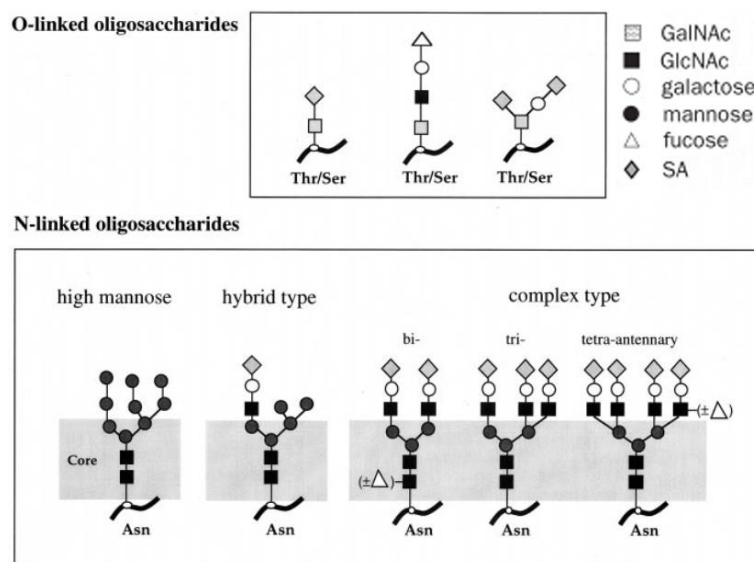


Figura 1. Diferents estructures de O-glicosilació i N-glicosilació [9].

es troben en dominis fortament glicosilats, només s'uneixen a serina o treonina i únicament es sintetitzen quan la cadena polipeptídica ha finalitzat. Per altre banda en la N-glicosilació, els sucres s'uneixen a la proteïna en blocs de 14 sucres durant la traducció[10].

Aquest treball s'ha centrat en la N-glicosilació ja que és la que es dona en la PSA.

La glicosilació es dona en el citosol, el reticle endoplasmàtic i l'aparell de Golgi i hi són involucrats diversos passos. Afecta en gran mesura les propietats de les proteïnes i les seves funcions biològiques

Hi ha dos tipus de glicosilacions: les O-glicosilacions i les N-glicosilacions. Les primeres contenen entre 1 i 20 sucres,

1.4. Sistemes d'estudi

Aquest treball s'ha desenvolupat a partir de l'estructura de PSA obtinguda del *Protein Data Bank* (PDB) on s'ha resolt l'estructura cristal·lina de la PSA formada per dos unitats de PSA i dos anticossos monoclonals [14]. La referència de la proteïna amb la que es treballa en el PDB és 3QUM, i així es referenciarà al llarg del treball.

Per realitzar l'estudi s'han utilitzat diferents glicoformes de la proteïna 3QUM. La primera glicoforma (G1) correspon a la proteïna desglicosilada, és a dir, s'ha eliminat la cadena glicosídica que portava la proteïna 3QUM en l'Asn 61. La segona glicoforma (G2) correspon a la 3QUM amb una N-glicosilació biantenària, amb àcid siàlic α -2,6 en els dos extrems de les branques d'oligosacàrids i fucosa unida al primer sucre de la cadena. La tercera glicoforma (G3) correspon a la 3QUM amb una N-glicosilació biantenària, àcid siàlic α -2,6 en els dos extrems de les branques d'oligosacàrids i sense fucosa. La G2 i la G3 comparteixen la cadena de sucres i l'estructura de la proteïna, i únicament es diferencien per la presència o no de fucosa.

Per realitzar aquest treball s'ha escollit la glicosilació biantenària amb una seqüència de sucres determinada ja que és la forma més habitual en mamífers. A més, s'ha escollit la isomeria i connectivitat que es troba en pacient amb càncer de pròstata. En la *figura 3* es mostren les dues cadenes d'oligosacàrids que presenten les dos glicoformes.

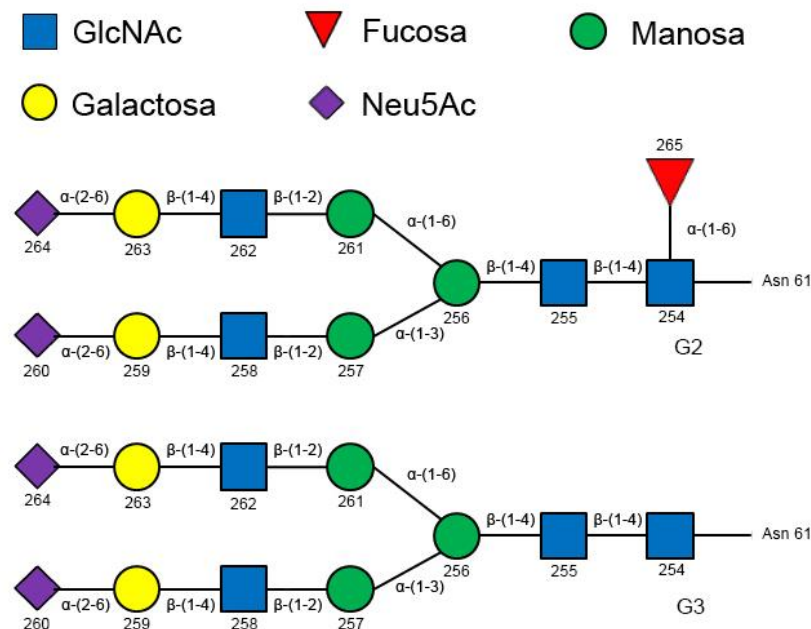


Figura 3. Estructura dels oligosacàrids que s'han utilitzat en les glicoformes 1 i 2 de la proteïna 3QUM. Les estructures es diferencien per la presència de fucosa en la G1. Neu5Ac és una forma d'anomenar l'àcid siàlic.

2. OBJECTIVES

- Protein modeling and molecular dynamics of the protein 3QUM and its glycoforms.
 - Build the computational model.
 - Carry out the simulation of the glycoforms.
 - Rmsd analysis to study the convergence of the dynamics runs.

- Check whether glycosylation and changes in the sugar chain affects its stability.
 - Analyze the hydrogen bonds of the glycoforms between the protein and the sugar chain.

3. METODOLOGIA

3.1. Simulació computacional

Des del desenvolupament dels ordinadors digitals a meitats del segle passat, la simulació computacional ha permès avançar en el nostre coneixement dels processos fonamentals de la natura com l'estructura molecular i la reactivitat en la química, la interacció fàrmac-receptor en la biologia o la mecànica de fluids en la física. Això s'ha utilitzat per intentar explicar i predir el comportament de sistemes complexos [15].

Les simulacions computacionals s'utilitzen per intentar entendre les propietats de les molècules en termes de la seva estructura i les seves interaccions microscòpiques. Serveix com a complement dels experiments convencionals i ens permet fer nous descobriments que no es poden fer per altres mètodes [16].

Amb la simulació computacional podem realitzar una aproximació teòrica del que passaria si ho portéssim a la pràctica. Però per acostar-nos al model real existeixen molts paràmetres que s'han d'establir i existeixen diverses formes de fer-ho. Els principals paràmetres que s'han de predir en una simulació són l'energia de les partícules i la seva trajectòria, que és la posició de les partícules del sistema i la seva evolució al llarg del temps.

Hi ha diversos mètodes per aproximar-nos a l'energia del sistema, però els més utilitzats són la Mecànica Quàntica (MQ) i la Mecànica Molecular (MM). La MQ es basa en la resolució aproximada de l'equació de Schrödinger segons el moviment dels electrons, que com a resultat dona l'estructura molecular i l'energia així com informació sobre els enllaços de les molècules. A més, en aquesta mecànica es consideren les molècules com grups de nuclis i electrons. Aquest mètode és precís però presenta un cost i un temps de càlcul elevat. Per altra banda, la MM no té en compte els electrons, considera cada àtom com una partícula i aplica les lleis de la Física Clàssica per modelar els sistemes moleculars. És aplicable per sistemes de gran mida i per això és la mecànica que s'ha escollit per a desenvolupar aquest treball [17].

Pel que fa a la trajectòria de les partícules del sistema, existeixen dos aproximacions: Dinàmica Molecular (DM) i Monte Carlo (MC). La DM calcula l'evolució real del sistema on les posicions atòmiques i les velocitats s'obtenen aplicant les lleis de Newton. Per altra banda, en la simulació MC, les posicions que poden ocupar les partícules es donen a l'atzar donant com a resultat algunes conformacions que no serien possible d'obtenir per DM. Però la DM presenta alguns avantatges respecte la simulació MC ja que permet establir les propietats dinàmiques del sistema com els coeficients de transport o la resposta a les pertorbacions en funció del temps. En aquest treball s'ha utilitzat la DM per realitzar les simulacions [16].

3.2. Mecànica Molecular

La MM és el mètode que s'escull per realitzar simulacions de proteïnes ja que és fonamental en l'estudi de la flexibilitat conformacional. És aplicable per sistemes de gran mida i per tractar sistemes biològics. A cada àtom se li assigna un radi i una carrega neta parcial, els enllaços es tracten com si fossin molles amb constants de força i distància d'equilibri determinades i no es poden trencar els enllaços covalents [18].

Es tracta d'un mètode clàssic que utilitza models per predir l'energia d'una molècula en funció de la seva conformació. Això permet prediccions de l'equilibri geomètric, els estats de transició i l'energia relativa entre conformacions o entre diferents molècules. La MM es pot utilitzar per obtenir l'energia potencial en simulacions de DM de molècules de gran mida. No és apropiada per reaccions de trencament d'enllaços però això no afecta al treball que s'ha realitzat [19].

En aquesta mecànica l'energia potencial del sistema ve donada per una funció analítica anomenada camp de forces.

3.2.1. Camp de forces

Els camps de forces són expressions matemàtiques que estableixen paràmetres per calcular l'energia potencial d'un sistema d'àtoms o partícules en MM i DM. Els paràmetres de les funcions energia es poden obtenir del treball experimental o a través de càlculs de MQ [20].

El mètode tracta les molècules com una col·lecció de partícules que s'ajunten per forces harmòniques simples. Els diferents tipus de forces són descrites en funcions individuals, el sumatori de les quals constitueix l'energia potencial molecular total o l'energia estàtica de la molècula [21]. L'equació 1 descriu aquesta energia total a partir de la suma dels diferents paràmetres de les molècules.

$$E_{total} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross} \quad (1)$$

Les funcions potencials que apareixen en l'equació 1 es descriuen a continuació:

- E_{str} → Energia de *Stretching*, energia associada a la distància entre els nuclis de dos àtoms.
- E_{bend} → Energia de *Bending*, energia associada a la modificació de l'angle entre tres àtoms enllaçats.
- E_{tors} → Energia de Torsió, representa l'energia associada a la rotació per un conjunt de 4 àtoms enllaçats consecutivament.

- E_{vdw} → Energia de van Der Waals, energia associada a la repulsió o atracció entre àtoms que no estan directament enllaçats, sense tenir en compte la contribució electrostàtica.
- E_{el} → Energia Electrostàtica, energia associada a la repulsió o atracció entre àtoms que no estan directament enllaçats degut a la seva càrrega parcial.
- E_{cross} → Termes de Creuament, termes addicionals per descriure en alguns casos l'acoblament entre les diferents coordenades del sistema, com el que es dona entre algunes flexions i torsions.

Totes les funcions potencials anomenades anteriorment contenen constants de forces o d'interacció. Per això, totes les constants s'han d'ajustar per a predir una geometria determinada a partir d'uns components coneguts. Llavors, aquestes constants s'utilitzen per calcular les estructures de nous components. La precisió d'aquestes constants, per tant, és crítica per als càlculs de la MM. Per aquest motiu, hi ha una gran varietat de camps de forces que estan dissenyats per diferents propòsits [22]. Els dos grups principals són els camps de forces clàssics i els polaritzables. Els camps de forces clàssics ignoren el moviment dels electrons i únicament descriuen l'evolució de la posició dels nuclis al llarg del temps. Entre aquests camps de forces s'inclouen AMBER, CHARMM, CVFF, etc. Pel que fa als camps de força polaritzables, inclouen la polarització, és a dir, el canvi de la distribució de la càrrega d'una molècula en resposta a l'ambient [23]. En aquest grup s'inclouen camps com X-Pol, DRF90, PIPF, etc.

En aquest treball s'han utilitzat camps de forces clàssics per a realitzar la MM i la DM. Són menys costosos i disponibles comercialment i s'utilitzen extensivament per la simulació de biomolècules. En concret, s'ha utilitzat AMBER, que és un grup de camps de forces i un conjunt de programes de simulació molecular [24].

Els camps de força que s'han utilitzat han sigut: ff14SB i GLYCAM_06j-1. El ff14SB és una versió del FF99SB i és el recomanat per la simulació de proteïnes i àcids nucleics en AMBER. Per altre banda, el GLYCAM_06j-1 és utilitzat quan intervenen sucres en la simulació. En aquest treball com a sistemes d'estudi s'han utilitzat proteïnes glicosilades i sense glicosilar. Per aquest motiu, quan s'ha realitzat la simulació de la proteïna 3qum sense glicosilar només s'ha fet servir el camp de força ff14SB, mentre que per les glicosilades s'han utilitzat ambdós camps.

3.3. Dinàmica molecular

La DM clàssica permet estudiar la dinàmica de processos complexos que es donen en sistemes biològics com la estabilitat de proteïnes, canvis conformacionals, plegament proteic, reconeixement molecular o transport iònic. És una eina que permet la simulació d'interaccions entre àtoms o molècules i calcular l'evolució real del sistema. Utilitza les lleis de Newton, tractant els enllaços i els angles com motlles, per obtenir una aproximació de la posició de les estructures moleculars i la seva evolució al llarg del temps [25].

Els objectius de la DM són: (1)Obtenir visions mitjanes d'un sistema; (2) Estudiar canvis en un sistema induït per pertorbacions externes; (3) Obtenir la termodinàmica d'un sistema i les seves interaccions; i (4) Ajudar a refinar les estructures a partir de les estructures dels rajos X o NMR.

La DM és determinista ja que la posició de les partícules en un temps determinat s'obté a partir de les condicions del sistema a temps t . Les equacions de moviment s'integren, calculant la posició, velocitat i força sobre les partícules en intervals de temps petits on s'assumeix que l'acceleració i la força són constants. Les partícules es van movent i en les noves posicions es recalculen les velocitats i les forces successivament. D'aquesta forma es va generant una trajectòria que descriu com evoluciona la posició i la velocitat de les partícules al llarg del temps [18].

La DM depèn de la teoria ergòdica que assumeix que les mitjanes dels conjunts estadístics obtinguts experimentalment són equivalents a les mitjanes del temps obtingudes per DM. Això permet fer córrer simulacions llargues i poder fer conclusions sobre les propietats d'un sistema basat en la seva simulació.

3.4. Simulació dels sistemes d'estudi

Per la simulació i dinàmica de les proteïnes d'estudi, s'ha utilitzar AMBER, com ja s'ha comentat anteriorment en l'apartat 3.2.1 *Camp de forces*. Per accedir a aquest software, s'ha utilitzat el servidor BETA de l'institut de Química Computacional i Catàlisi de la UdG.

L'estructura de la proteïna d'estudi, que ja s'ha descrit en l'apartat 1.5, s'ha obtingut de la pàgina del *PDB*. Per la preparació, simulació i anàlisi de resultats, s'ha seguit el tutorial B1: *Simulating a small fragment of DNA* de la pàgina web: ambermd.org/tutorials/ [26]. Encara que en aquest tutorial es tracti un fragment de DNA, s'han adaptat les condicions per fer la simulació d'una proteïna. A més, per simular àcids nucleics i proteïnes amb AMBER, es fa servir el mateix camp de força, el ff14SB.

3.4.1. Preparació dels sistemes d'estudi

Abans de començar la simulació, s'ha hagut de preparar la proteïna 3QUM per adaptar-la a la finalitat de l'estudi. La 3QUM del *pdb* conté dues unitats de la PSA, la P i la Q i dos anticossos monoclonals en la seva estructura resolta per difracció de raigs X. Per l'anàlisi computacional d'aquest treball s'ha escollit la cadena P.

En totes les variants de la 3QUM que s'estudien en aquest treball, s'han hagut de fer alguns canvis perquè el programa tLeap del software AMBER reconegués correctament la proteïna i les interaccions. Els canvis han sigut: canviar CYX per CYS, en referència als ponts di-sulfur, i ASN per NLN, ja que es dona una N-glicosilació en aquest aminoàcid (Asn 61).

En el cas de la 3QUM sense glicosilar, s'han eliminat les cadenes de sucre de l'arxiu PDB mitjançant l'editor de text Vi del software *Linux*, ja que l'arxiu original contenia una N-glicosilació triantenària en la Asn 61.

Pel que fa a les 3QUM glicosilades, amb i sense fucosa, s'ha eliminat com en el cas anterior la glicosilació que portava l'arxiu del PDB. L'objectiu d'aquest treball és l'estudi de la 3QUM amb una glicosilació de dues ramificacions, és a dir, biantenària i l'efecte de la fucosa en la proteïna. Un cop eliminada la cadena de sucres triantenària, s'ha procedit a construir la nova cadena de sucres amb el programa tLeap. S'ha utilitzat com a referència el manual d'AMBER: Amber 2015 Reference Manual per les instruccions i la nomenclatura de les molècules de sucres que s'han afegit. També s'han afegit els enllaços i els angles diedres perquè la unió i la posició entre les molècules sigui la correcta. Amb el software VMD s'ha visualitzat la posició de les molècules de la cadena de sucres per comprovar que fos la correcta.

Després de realitzar els canvis mencionats, s'ha creat una topologia amb una càrrega neutra afegint ions a l'estructura. En el cas de la proteïna 3QUM, s'han afegit ions Na^+ per neutralitzar la càrrega. Llavors s'ha procedit a solvatar la proteïna afegint una caixa octaèdrica d'aigua segons el model *TIP3PBOX* de 8 àngstroms [26]. Posteriorment, s'han creat els arxius:

- *Prmtop* → L'arxiu que conté la topologia molecular i defineix la connectivitat i els paràmetres del model actual. Conté informació estàtica ja que no canvia al llarg de la simulació.
- *Inpcrd* → Conté les coordenades moleculars i la informació. Varia al llarg de la simulació.

3.4.2. Minimització del sistema d'estudi

Abans de començar la dinàmica molecular és recomanable realitzar una minimització del sistema per evitar que hi hagi inestabilitat de les molècules en la simulació computacional. També és interessant deixar que la caixa de solvent, en aquest cas aigua, es relaxi en un estadi d'equilibri abans de començar la DM. Això permetrà que l'aigua s'estabilitzi al voltant del solut per arribar a un equilibri de la densitat [26].

Per aquest motiu, s'han realitzat dos passos en la minimització del sistema. En el primer pas s'ha mantingut la proteïna fixa amb restriccions i només s'han minimitzat les posicions de l'aigua i dels ions. Un cop s'han minimitzat, s'ha procedit al segon pas on s'han minimitzat tot el sistema.

3.4.3. Dinàmica molecular del sistema d'estudi

Un cop s'ha minimitzat el sistema de treball, s'ha procedit a realitzar la dinàmica del mateix. El primer pas de la dinàmica és augmentar la temperatura del sistema ja que s'ha estat treballant a 0 K. Es realitza amb una restricció en el moviment de la proteïna perquè es doni l'escalfament sense grans fluctuacions. La finalitat és aconseguir una dinàmica a temperatura i pressió constant.

D'aquesta manera s'ha fet una primera dinàmica de 20 ps amb unes restriccions febles per començar a escalfar el sistema perquè s'equilibri. Llavors s'ha seguit realitzant la dinàmica sense restriccions durant 100 ps més, per seguir estabilitzant la proteïna i arribar a la temperatura de 300 K i pressió constant.

Després d'aquesta etapa de relaxació i equilibri del sistema s'ha procedit a realitzar la dinàmica durant un període de temps més elevat. La durada de cada pas de la dinàmica ha sigut de 2 fs i s'ha utilitzat l'algoritme SHAKE. Aquest algoritme aplica una constricció als enllaços on intervien hidrògens, els manté fixes i elimina el seu moviment. D'aquesta manera s'eliminen els enllaços amb major freqüència d'oscil·lació del sistema i es pot augmentar el temps de cada pas sense induir inestabilitat en el sistema que realitza la dinàmica molecular [26].

El servidor amb el que s'ha realitzat el treball admet com a màxim cues de fins a 8 dies. Un cop finalitzava aquest temps es tornava a utilitzar l'arxiu *.rst* (restart) i *.mdcrd* (conté la trajectòria de la dinàmica molecular) per continuar la dinàmica en el punt on havia finalitzat l'anterior. D'aquesta forma s'han realitzat dinàmiques de les diferents proteïnes d'estudi durant diferents temps, ja que depèn de la mida del sistema que la DM vagi més o menys ràpid. Si la DM ha corregut durant més temps implica que la proteïna té més probabilitat d'assolir una conformació determinada i augmentar la seva estabilitat. En aquest treball els temps de simulació per les 3 proteïnes està comprés entre 36 i 58 ns.

Per analitzar l'estabilitat dels sistemes d'estudi s'han analitzat els ponts d'hidrogen entre la proteïna, mentre que amb l'anàlisi del rmsd s'ha estudiat la convergència de les dinàmiques. S'entén com convergència l'evolució del sistema cap a la situació d'equilibri, és a dir, assolir la conformació més estable.

3.4.4. Rmsd

El rmsd dona informació del grau de similitud entre dues estructures i es calcula segons l'equació 2. D és la distància entre cada parell d'àtoms equivalents en dos estructures superposades, mentre que n és el nombre total d'àtoms.

$$rmsd = \sqrt{\frac{\sum_i d_i^2}{n}} \quad (2)$$

Si el valor del rmsd és 0 vol dir que les dues estructures són idèntiques mentre que si el valor augmenta implica que les estructures es van diferenciant. Els valors de rmsd són uns indicadors fiables de la variabilitat entre proteïnes similars o conformacions alternatives de la mateixa proteïna [27]. També s'utilitza per analitzar la convergència de les simulacions, és a dir, si s'ha arribat a una situació final d'equilibri i s'ha assolit la conformació més estable, on el valor de rmsd roman constant.

En aquest treball l'estructura que s'ha agafat de referència per el càlcul del rmsd ha sigut la conformació de la proteïna en la segona minimització, on s'ha minimitzat tota l'estructura. A partir d'aquest punt, s'han comparat les estructures i s'ha analitzat com ha anat variant la conformació a mesura que es duia a terme la dinàmica.

3.4.5. Anàlisi dels ponts d'hidrogen i les distàncies

L'anàlisi dels ponts d'hidrogen que s'estableixen entre les molècules de sucre i la proteïna permet donar informació de l'estabilitat ja que aporten estabilitat a l'estructura secundària del sistema. Això implica que les molècules de sucre s'apropen a les molècules de la proteïna per adoptar una conformació diferent que pot augmentar l'estabilitat. Els ponts d'hidrogen que duren bastants *frames* impliquen un enllaç estable i que té importància en l'estabilitat del sistema. Els que es donen durant pocs *frames* no es tenen en compte ja que són interaccions puntuals amb poca influència. Els *frames* són arxius que es generen en intervals determinats que representen una imatge de la situació de la proteïna en aquell moment.

Els criteris per considerar que es tracta d'un pont d'hidrogen es basen en la distància entre acceptor i donador i l'angle establert en l'enllaç, com es pot veure en la figura 4. La distància (R) ha de ser menor de 3 Å i l'angle (θ) més gran de 135°.

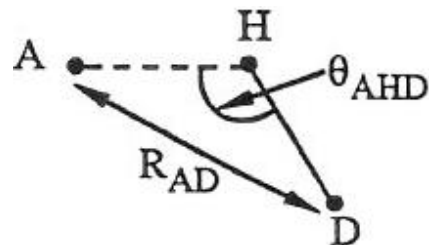


Figura 4. Pont d'hidrogen entre un donador (D) i un acceptor (A). [29]

La distància del pont d'hidrogen també és important per tindre informació de com ha variat la posició relativa de les dues molècules a mesura que es realitza la dinàmica.

4. RESULTATS

Els temps de simulació de les proteïnes d'aquest estudi, així com els frames que s'han generat al llarg de la dinàmica, es recullen en la *taula 1*. Com es pot observar, el temps que han estat les diferents glicoformes, G1, G2 i G3, realitzant la dinàmica molecular han sigut diferents. Això s'ha degut a que s'han anat generant els arxius progressivament i s'han anat enviant al servidor per realitzar la dinàmica. Per aquest motiu, no s'han pogut enviar tots a la vegada i els temps de simulació han variat bastant.

En la dinàmica que s'ha realitzat, els *frames* s'han generat cada 200 fs en la G1 fins els 8350 ps i després cada 1000 fs. En la G2 i la G3 al llarg de tota la dinàmica s'han generat cada 1000 fs. Aquest canvi va ser degut a la limitació d'espai de disc del servidor BETA. Si els *frames* tenen un pas molt curt, llavors es genera un volum d'arxius major però ocupa més espai. Degut a aquesta limitació, s'ha sacrificat precisió per poder continuar el treball. Per aquest motiu, la *figura 5* està en funció dels *frames* i no dels temps, com la resta de gràfics.

Taula 1. Temps de simulació computacional de les diferents proteïnes en nanosegons (ns) així com els frames generats.

Sistema d'estudi	Temps de simulació [ns]	Frames
G1: 3QUM sense glicosilar	57,640	65742
G2: 3QUM glicosilada	45,241	45642
G3: 3QUM glicosilada sense fucosa	36,677	36748

4.1. Rmsd i convergència de les trajectòries

Els valors de rmsd que s'han obtingut de les tres glicoformes s'han representat en la *figura 5* i s'han calculat a partir de l'*equació 2*.

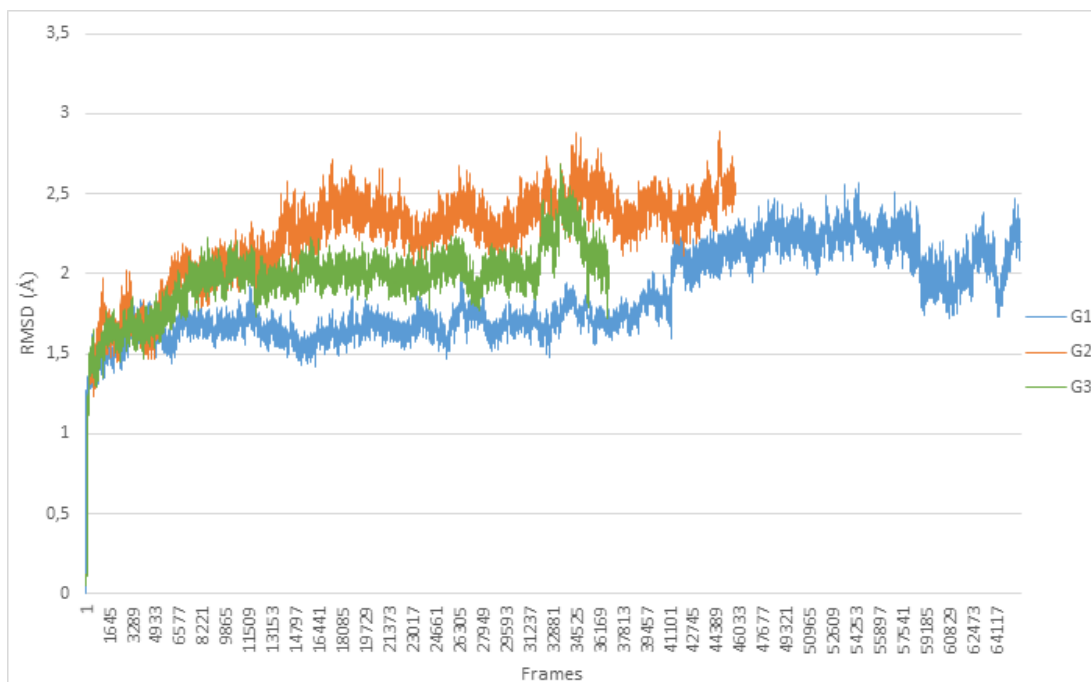


Figura 5. Representació dels rmsd de les diferents glicoformes de la proteïna d'estudi. S'han representat en colors diferents: blau (G1), taronja (G2) i verd (G3).

Es pot observar en la *figura 5* que les glicoformes han donat diferents resultats. La G1 té un canvi important al voltant dels 40.000 frames i una oscil·lació al final. La G2 encara està oscil·lant al final de la trajectòria i la G3 presenta una variació important aproximadament als 32.000 frames, poc abans del final de la simulació.

4.2. Anàlisi dels ponts d'hidrogen

S'han analitzat els ponts d'hidrogen que estableix la cadena de sucres amb la proteïna, per les glicoformes G2 i G3. Els resultats es recullen en la *taula 2*. Únicament s'han considerat els ponts d'hidrogen que s'han establert a partir del 10 % dels frames totals generats per cada glicoforma. S'ha establert aquest llindar per considerar els que tenen major importància i tenen un paper en l'estabilitat de la proteïna.

Taula 2. Ponts d'hidrogen que s'estableixen entre la proteïna i la cadena de sucres i representen com a mínim un 10% dels frames totals generats en la simulació computacional. La nomenclatura dels sucres és la que utilitza el programa AMBER. OSA: àcid siàlic en posició terminal; 2MA: manosa amb enllaç 1→2. La numeració dels sucres es mostra en la *figura 3*.

	Acceptor	Donador	Frames [% respecte total]	Distància mitjana
G2: 3QUM glicosilada	ASP_107@OD1	OSA_260@H4O	29,0	2,71
	SER_96@O	OSA_260@H8O	20,9	2,79
	Acceptor	Donador	Frames [% respecte total]	Distància mitjana
G3: 3QUM glicosilada sense fucosa	2MA_261@O6	ASN_100@H	18,3	2,89
	2MA_261@O3	ARG_101@HH21	13,4	2,87
	2MA_261@H61	ASN_100@H	11,3	2,84

4.2.1. Evolució de les distàncies

S'ha analitzat l'evolució de les distàncies dels ponts d'hidrogen més representatius que són els que apareixen en la *taula 2*. D'aquesta forma s'obté una visió de com s'han anat apropant els àtoms per generar l'enllaç i en quin moment de la simulació s'ha donat aquest apropament.

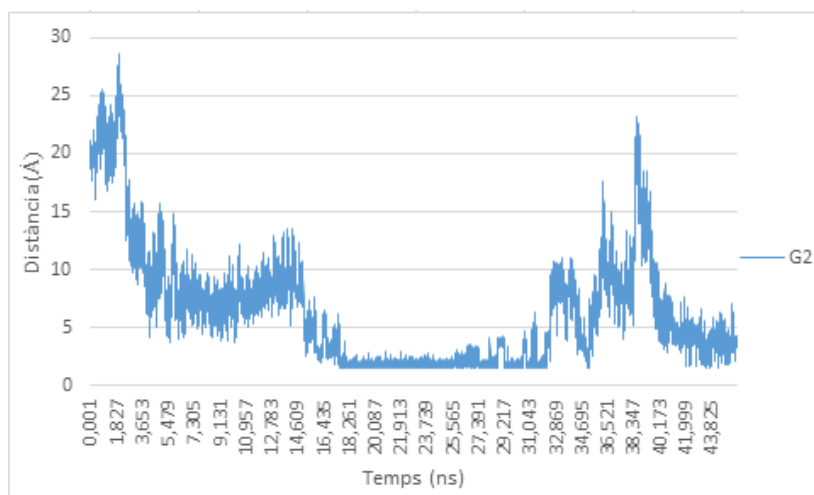


Figura 5. Evolució del pont d'hidrogen entre els àtoms ASP_107@OD1 i OSA_260@H4O de la G2.

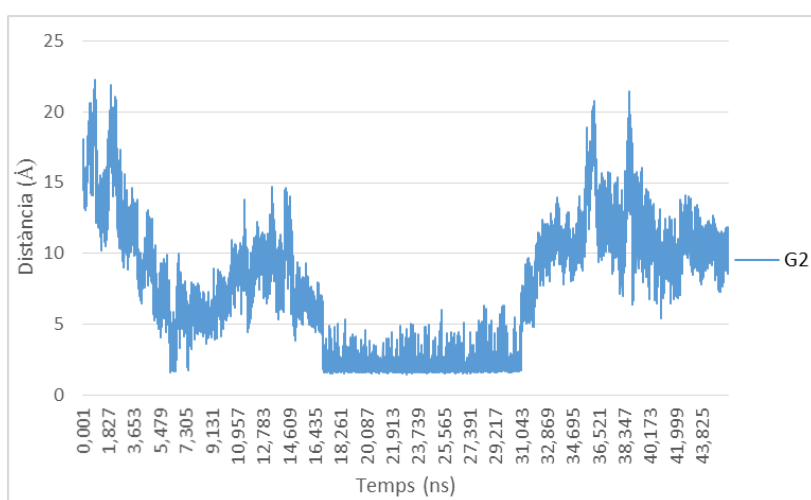


Figura 6. Evolució del pont d'hidrogen entre els àtoms SER_96@O i OSA_260@H8O de la G2.

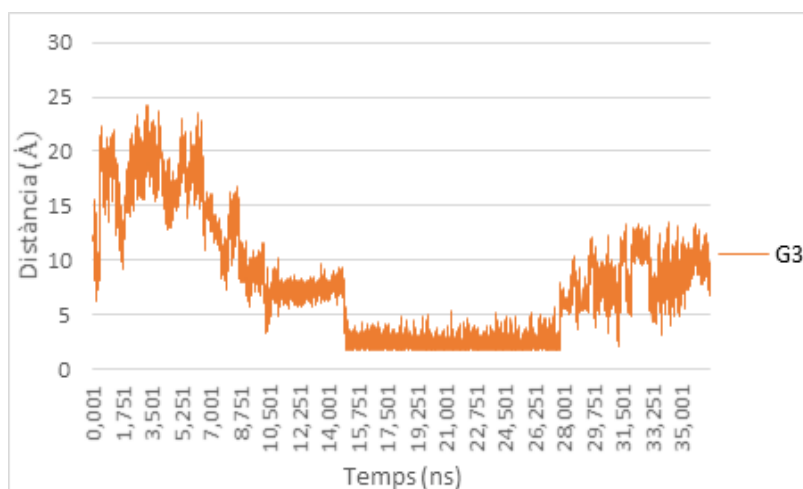


Figura 7. Evolució del pont d'hidrogen entre els àtoms 2MA_261@O3 i ARG_101@HH21 de la G3.

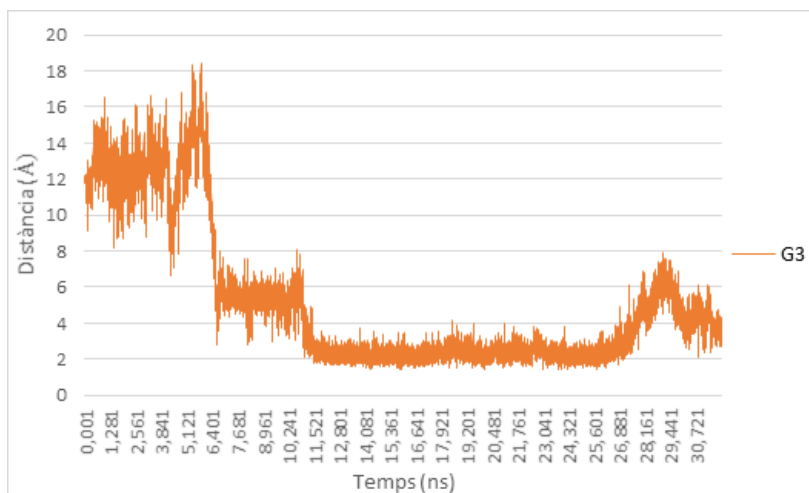


Figura 8. Evolució del pont d'hidrogen entre els àtoms 2MA_261@H61 i ASN_100@H de la G3.

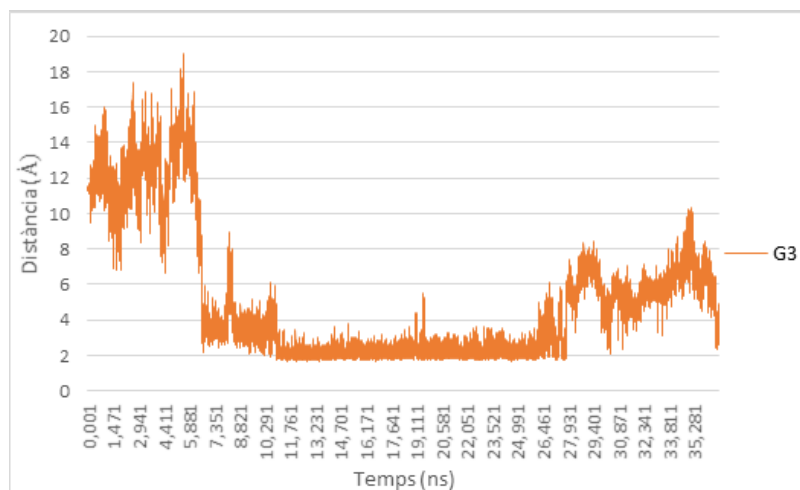


Figura 9. Evolució del pont d'hidrogen entre els àtoms 2MA_261@O6 i ASN_100@H de la G3.

Es pot observar en les dos glicoformes que els àtoms es van apropant fins arribar a una distància menor a 3 Å entre els 17 i 32 ns en la G2 i entre 11 i 28 ns en la G3. Llavors, la distància entre els àtoms augmenta de valor i pateix algunes variacions, sobretot en la G2 on la oscil·lació en la distància és superior que en la G3.

5. DISCUSSIÓ

Com s'ha comentat prèviament, un dels factors que s'ha de tenir en compte alhora de comparar els resultats obtinguts de les diferents glicoformes, és que els temps de simulació han sigut diferents. Això implica que el resultat podria ser diferent si es comparaessin totes les glicoformes amb el mateix temps de simulació. Tot i així, com no s'ha disposat de més temps s'han comparat els resultats amb els temps de simulació disponibles.

Si es comparen els rmsd es pot observar que les estructures han variat bastant a partir de la conformació que tenien després de la segona minimització. Però en cap de les tres glicoformes es podria concloure que s'ha arribat a l'equilibri. Les estructures s'han anat estabilitzant, però la variació al final de la simulació és alta ja que encara es donen alguns pics i oscil·lacions importants. Això es deu a que les estructures inicials s'han obtingut modificant l'estructura de la proteïna 3QUM triantenària. La conformació inicial s'ha imposat de manera aproximada, i els resultats indiquen que és necessari més temps de simulació per arribar a la conformació que tindrà la proteïna en condicions reals.

Pel que fa als ponts d'hidrogen de la *taula 2*, es poden observar diferències entre la G2 i la G3. S'han generat enllaços durant més *frames* respecte el total i amb unes distàncies mitjanes menors en la G2. Això implica que aquests ponts d'hidrogen són més estables, i per tant, més importants per l'estabilitat de la proteïna. A més, els àtoms que han establert els enllaços són diferents en les dos glicoformes. Per tant, donat que la única diferència que hi ha entre les dues proteïnes és la presència o no de fucosa, es pot determinar que la presència de fucosa té un paper clau en el plegament proteic i influeix en l'estabilitat del sistema.

Pel que fa a la variació de la distància entre els àtoms quan estableixen els ponts d'hidrogen, s'observen uns resultats bastant similars en les dos glicoformes. Segueixen el mateix patró en les distàncies dels enllaços, és a dir, els àtoms s'apropen un temps determinat i llavors s'allunyen amb alguns pics en la distància d'enllaç. Aquest allunyament dels àtoms implica un canvi conformacional que ha sigut més acusat en la G2 ja que es veu una variació important en la distància dels enllaços cap als 38 ns de simulació. És possible que al llarg d'aquest canvi conformacional s'hagin donat nous ponts d'hidrogen, però com el temps de la dinàmica ha sigut limitat no s'han tingut en compte.

Per tant, no s'observa un efecte important de la presència o no de fucosa en la distància dels ponts d'hidrogen. Els sistemes d'estudi haurien d'estar més temps realitzant la dinàmica per obtenir uns resultats més concrets, ja que s'observa una tendència de les gràfiques de tornar a reduir el valor de la distància entre els àtoms al final del temps simulat.

Finalment, cal destacar que les simulacions computacionals que s'han realitzat no han generat residus contaminants ni han provocat patiment animal com podria passar en un laboratori. A més, és un treball relacionat amb projectes del Departament de Biologia per desenvolupar marcadors més eficients alhora de diagnosticar la presència de càncer de pròstata.

6. CONCLUSION

The main objective of building the model, making a computer simulation and modelling the PSA glycoforms G1, G2 and G3 has been achieved, with a simulation time between 36 ns and 58 ns.

Rmsd analysis has been carried out to study the convergence of the simulation and shows that more time is needed in order to achieve the final and more stable conformation. This is due to the fact that the starting structures for the glycosylated forms have been modified from the available X-ray protein structure.

Hydrogen bonds have been studied and some differences between G2 and G3 have been found. The G2's hydrogen bonds last longer and have a shorter distance which means that are more relevant for protein stabilization.

Having analyzed and discussed the results, we conclude that the fucose plays an important role in protein folding and affects system stability.

7. BIBLIOGRAFIA

- [1] L. M. Clemente, "Calidad del semen y del esperma." 2010. [Online]. Available: <http://www.hola.com/salud/enciclopedia-salud/2010031245149/hombre/generales/calidad-del-semen-y-del-esperma/>. [Accessed: 15-Mar-2016].
- [2] G. Tabarés, C. M. Radcliffe, S. Barrabés, M. Ramírez, N. Aleixandre, W. Hoesel, R. A. Dwek, P. M. Rudd, R. Peracaula, and R. de Llorens, "Different glycan structures in prostate-specific antigen from prostate cancer sera in relation to seminal plasma PSA." *Glycobiology*, vol. 16, no. 2, pp. 132–145, 2006.
- [3] NCBI, "KLK3 kallikrein related peptidase 3 [Homo sapiens (human)]." 2016. [Online]. Available: <http://www.ncbi.nlm.nih.gov/gene/354>. [Accessed: 19-Apr-2016].
- [4] C. Borgoño, I. P. Michael, and E. P. Diamandis, "Human tissue kallikreins: physiologic roles and applications in cancer." *Mol. Cancer Res.*, vol. 2, no. 5, pp. 257–280, 2004.
- [5] P. Leidinger, A. Keller, L. Milchram, C. Harz, M. Hart, A. Werth, H. P. Lenhof, A. Weinhäusel, B. Keck, B. Wullich, N. Ludwig, and E. Meese, "Combination of Autoantibody Signature with PSA Level Enables a Highly Accurate BloodBased Differentiation of Prostate Cancer Patients from Patients with Benign Prostatic Hyperplasia." *PLoS One*, 2015.
- [6] X. Filella, D. Truan, J. Alcover, R. Molina, P. Luque, F. Coca, and A. M. Ballesta, "Antígeno prostático específico y sus fracciones en el diagnóstico del cáncer de próstata." *Med. Clin. (Barc.)*, 2003.
- [7] M. A. Khan, A. W. Partin, H. G. Rittenhouse, S. D. Mikolajczyk, L. J. Sokoll, D. W. Chan, and R. W. Veltri, "Evaluation of Prostate Specific Antigen for Early Detection of Prostate Cancer in Men With a Total Prostate Specific Antigen Range of 4.0 to 10.0 Ng/ML." *J. Urol.*, vol. 170, no. 3, pp. 723–726, 2003.
- [8] R. Peracaula, S. Barrabés, A. Sarrats, P. M. Rudd, and R. de Llorens, "Altered glycosylation in tumours focused to cancer diagnosis." *Dis. Markers*, vol. 25, pp. 207–218, 2008.
- [9] G. Durand and N. Seta, "Protein glycosylation and diseases: Blood and urinary oligosaccharides as markers for diagnosis and therapeutic monitoring." *Clin. Chem.*, vol. 46, no. 6, pp. 795–805, 2000.
- [10] S. Gilgunn, P. J. Conroy, R. Saldova, P. M. Rudd, and R. J. O'Kennedy, "Aberrant PSA glycosylation-a sweet predictor of prostate cancer." *Nat. Rev. Urol.*, vol. 10, no. 2, pp. 99–107, 2013.
- [11] UniProt, "Glycosylation." 2015. [Online]. Available: <http://www.uniprot.org/help/carbohydr>. [Accessed: 02-Mar-2016].
- [12] S. Wildt and T. U. Gerngross, "The humanization of N-glycosylation pathways in yeast." *Nat. Rev. Microbiol.*, vol. 3, no. 2, pp. 119–28, 2005.
- [13] A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, and M. E. Etzler, "Essentials of Glycobiology." 2nd editio. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2009.
- [14] E. A. Stura, B. H. Muller, M. Bossus, S. Michel, C. Jolivet-Reynaud and F. Ducancel, "Crystal structure of human prostate specific antigen (PSA) in Fab sandwich with a high affinity and a PCA selective antibody." 2011. [Online]. Available: <http://www.rcsb.org/pdb/explore.do?structureId=3QUM>. [Accessed: 02-Mar-2016].
- [15] U. S. D. of E. Office of Science, "Scientific Discovery through Advanced Computing." p. 28, 2000.
- [16] Michael P. Allen, "Introduction to Molecular Dynamics Simulation." 2004.
- [17] W. J. Hehre, "A Guide to Molecular Mechanics and Quantum Chemical Calculations." 2003.
- [18] P. Salvador, "Tema 7. Computer Simulations." UdG, Girona, 2015.

- [19] C. D. Sherril, "Introduction to Molecular Mechanics." School of Chemistry and Biochemistry Georgia Institute of Technology, Georgia, 2014.
- [20] J. W. Ponder and D. A. Case, "Force fields for protein simulations." *Advances in Protein Chemistry*. 2003.
- [21] K. Vanommeslaeghe, O. Guvench and A. D. MacKerell Jr, "Molecular Mechanics." 2014.
- [22] T. W. Shattuck, "Molecular Mechanics Tutorial." Colby College, Maine, 2008.
- [23] C. M. Baker, "Polarizable force fields for molecular dynamics simulations of biomolecules." *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 5, no. 2, pp. 241–254, 2015.
- [24] Amber, "Ambermd." 2016. [Online]. Available: www.ambermd.org. [Accessed: 12-Apr-2016].
- [25] Actoforlibraries.org, "What is Molecular Dynamics." 2016. [Online]. Available: <http://www.actoforlibraries.org/what-is-molecular-dynamics/>. [Accessed: 13-Apr-2016].
- [26] R. Walker, "TUTORIAL B1 Simulating a DNA polyA-polyT Decamer." 2015. [Online]. Available: <http://ambermd.org/tutorials/basic/tutorial1/>. [Accessed: 10-Apr-2016].
- [27] O. Carugo and S. Pongor, "A normalized root-mean-square distance for comparing protein three-dimensional structures." *Proteina Sci.*, 2001.
- [28] J. E. Nettleship, "Structural Biology of Glycoproteins." InTech, 2012.
- [29] S. Corporation, "Hydrogen bond." 2013. [Online]. Available: http://lammps.sandia.gov/doc/pair_hbond_dreiding.htmlhttp://lammps.sandia.gov/doc/pair_hbond_dreiding.html. [Accessed: 01-Apr-2016].