

Biofísica i immunoteràpia contra el càncer

Interacció entre cèl·lules T i hidrogels sintetitzats amb heparina

Maximilià Soler i Subirana

1497326

Treball supervisat per Jordi Faraudo, científic titular a l'ICMAB

Treball de Fi de Grau en Física
Convocatòria pel Juliol de 2023



Universitat Autònoma de Barcelona

Facultat de Ciències

Departament de Física

Barcelona, Espanya

15 de Setembre de 2022

Biofísica i immunoteràpia contra el càncer

Interacció entre cèl·lules T i hidrogels sintetitzats amb heparina

Maximilià Soler i Subirana

Resum

En el present treball es durà a terme una investigació de la interacció electroestàtica entre dues molècules; la heparina, altament electronegativa; i la proteïna CCL21, recentment considerada una de les més electropositives al cos humà. Els resultats d'aquest estudi permetran de valorar si aquesta interacció pot ajudar al desenvolupament de cèl·lules T al laboratori per a la lluita contra el càncer, mitjançant la tècnica coneguda com a immunoteràpia.

A més, es consideraran les altres forces intermoleculars que hi puguin jugar un paper clau, i.e. interaccions de Van der Waals, ponts d'hidrogen i la mateixa energia emmagatzemada al medi. Tot això ens durà a concloure, finalment, si és una interacció crucial per a la immunoteràpia tot fent ús dels programes de visualització i edició de molècules disponibles per a l'abast de tothom.

Sumari

| | | |
|----------|---|-----------|
| 1 | Introducció i objectius | 3 |
| 1.1 | Motius | 3 |
| 1.2 | Procediments | 3 |
| 1.3 | Electroestàtica | 3 |
| 1.4 | Rotacions i torsions | 4 |
| 1.5 | Hipòtesis i objectiu final | 4 |
| 2 | Mètode | 6 |
| 2.1 | Equació de Poisson-Boltzmann | 6 |
| 2.2 | PB per a medis salins | 8 |
| 2.3 | Acoblament molecular | 9 |
| 3 | Resultats | 10 |
| 3.1 | Selecció del model proteic | 10 |
| 3.2 | Càlcul electrostàtic superficial de la proteïna | 10 |
| 3.3 | Heparina | 12 |
| 3.4 | Interacció molecular | 13 |
| 4 | Conclusions i comentari final | 16 |
| A | Apèndix | 17 |
| A.1 | Bibliografia | 17 |
| A.2 | Annexos | 18 |
| A.2.1 | Estudi i selecció del millor model virtual | 18 |
| A.2.2 | Obtenció dels arxius per l'APBS | 21 |
| A.2.3 | Preparació dels Arxius per a l'AutoDock | 22 |
| A.3 | Declaracions | 26 |
| A.3.1 | Declaració d'autoria del Treball de Fi de Grau | 26 |
| A.3.2 | Declaració d'Extensió del Treball de Grau | 27 |

1 Introducció i objectius

1.1 Motius

En el present treball s'estudiarà la interacció electroestàtica, entre dues molècules, la proteïna CCL21, pertanyent a la membrana cel·lular, i un hidrogel compost principalment d'heparina, per a la millora d'una de les noves teràpies contra el càncer, la immunoteràpia.

Aquesta consisteix, a gran escala, en millorar l'efectivitat de les cèl·lules T del mateix pacient que pateix un càncer i les defenses del qual no poden lluitar adequadament contra el tumor. Per a poder millorar la condició d'aquestes cèl·lules, és necessari extreure-les i tractar-les al laboratori. Això implica haver de trobar un entorn apte per a les cèl·lules T, que ha de ser similar al de l'interior del cos humà.

Per al correcte desenvolupament d'aquestes, es necessitarà descobrir i estudiar les principals interaccions entre les seves membranes, que és per on poden unir-se i interactuar amb el medi, i els diferents teixits del cos humà. Aquí és on indaga aquest treball: l'estudi de la interacció electroestàtica d'una proteïna de la membrana cel·lular altament electropositiva amb un entorn similar al d'un gangli limfàtic, altament electronegatiu.

Finalment, es determinarà si en aquesta unió entre CCL21 i heparina hi intervenen altres interaccions considerables, com ara enllaços covalents o ponts d'hidrogen, excloent la ja mencionada electroestàtica.

1.2 Procediments

Per tal d'obtenir un conjunt de cèl·lules T, es fa ús del mètode d'afèresi, que separa les cèl·lules seleccionades de la resta de components de la sang. Per al cultiu de microorganismes, se solen utilitzar recipients considerats bidimensionals, com per exemple la placa de Petri, que consisteix en un recipient circular pla. No obstant, el que es vol és cultivar aquestes cèl·lules de forma tridimensional, ja que comporta una eficàcia superior a l'hora de reproduir-les. Aquestes necessiten un entorn similar al d'un gangli, la matriu extracel·lular, per a poder multiplicar-se. Un cop feta la selecció, es comprova que les cèl·lules responguin adequadament a la interacció a les proteïnes del càncer de la manera prevista. Finalment, aquestes cèl·lules són reincorporades al torrent sanguini pacient per tal que puguin combatre d'una manera efectiva el tumor maligne.

En tractar-se de cèl·lules del propi pacient, la taxa de rebuig és mínima, de manera que s'aconsegueix un major grau d'eficàcia al mateix temps que s'evita l'alta agressivitat d'altres tipus de tractament [1].

1.3 Electroestàtica

A l'interior de les cèl·lules, distribuïdes pel medi intracel·lular, hi ha una gran quantitat de macromolècules. Aquestes, formades per diferents parts unides per diversos tipus d'enllaços, poden perdre algunes de les zones dèbilment connectades al cos principal. La causa d'aquest fenomen és l'agitació tèrmica: al món microscòpic, la temperatura fa que tots els components estiguin en moviment constant. Aquests components, generalment àtoms o petites molècules, quan se separen de l'estructura principal solen perdre electrons, que hi romanen, de manera que la transformen en una molècula carregada negativament, anomenada macro-ió. És per això que diem que aquesta molècula queda carregada. Els àtoms que han marxat s'anomenen contra-ions, ja que la suma de totes les seves càrregues cancel·la la del macro-ió.

Cal, a més, mencionar els aminoàcids: estructures moleculars compostes d'un grup carboxil ($-COOH$) i un grup amino ($-NH_2$) que composen les proteïnes. Tot i que acostumen a ser neutres en medis com ara l'aigua pura, alguns presenten certa càrrega elèctrica. És important recalcar que, una gran part de les macromolècules en medis intracel·lulars mencionades anteriorment, solen ser proteïnes. Aquestes s'encarreguen de moltes funcions dins el cos: són el principal component de l'estructura física de les cèl·lules, el principal element utilitzat en els processos cel·lulars i transport de materials, i interaccions entre diferents cèl·lules entre molts altres.

Un exemple interessant aplicable al nostre camp seria el de l'ADN, ja que els nucleòtids que el composen acostumen a ionitzar-se en interactuar amb aquest entorn salí que és el medi intracel·lular. Això és possible perquè no presenten enllaços covalents: l'energia que comparteixen dos àtoms junts és inferior a la que tenen per separat, i, normalment l'energia sempre tendeix a romandre a l'estat més baix i estable possible. Per

tant doncs, els és més fàcil separar-se ja que requereixen poca energia per a realitzar aquest allunyament, que el medi els la proporciona.

Aquests ions ara confronten un dilema: si s'allunyen generen entropia, però per a dispersar-se necessiten energia ja que provenen d'una molècula on hi han perdut electrons. El resultat final és, doncs, que adquireixen la màxima entropia que els és possible amb la mínima energia, com hem vist a l'exemple anterior. Als medis cel·lulars objecte d'aquest estudi, té lloc un fenomen conegut com a *capa difusa de càrrega*, que consisteix en què els àtoms formen una espècie de núvol que rodeja el macro-ió. Aquesta capa, usualment es forma a una distància de pocs nanòmetres i cancel·la la càrrega més enllà d'aquesta distància, generant una càrrega total neutra vista des de fora [2]. D'aquesta manera, un macro-ió que es trobi a prop d'aquesta zona, no interaccionarà amb un altre macro-ió fins travessar aquesta capa que contraresta l'energia del macro-ió. Aquests àtoms del núvol, juntament amb les càrregues a la superfície del macro-ió, s'anomenen *dobla capa elèctrica*.

Per altre banda, cal aclarir que en el present treball es farà ús de l'expressió $k_B T$ per a fer referència a l'energia tèrmica per molècula. A més, quan apareix un sistema l'energia del qual supera amb ordres de magnitud aquest factor depenent de la temperatura, es diu que és ionitzable: és el cas mencionat anteriorment de l'ADN.

Una observació interessant a realitzar és que, amb les condicions esmentades anteriorment, quan dos macro-ions es troben molt a prop un de l'altre, les interaccions entre cada component dins el núvol i la geometria de la superfície de cada macro-ió tenen un paper crucial. Aquestes interaccions, de les quals dependrà de quina manera les macromolècules encaixaran entre elles, s'anomenen *estructures estereoespecífiques*. Per a abordar aquest problema i determinar el potencial a la superfície d'aquests macro-ions, s'utilitza l'equació de Poisson-Boltzmann, que s'explicarà detalladament a la secció de Mètodes.

1.4 Rotacions i torsions

Cada molècula presenta certa geometria que pot modificar en interactuar amb altres components. Els enllaços rodables (rotatable bonds) que exhibeix una molècula són tots aquells que no es troben en un anell (per exemple, el benzè) i consisteixen en un enllaç senzill, sense comptar els grups amino i carboxil, ni els hidrògens a cada residu. Pel que fa a les torsions, fan referència a com giren aquests enllaços lliures en contacte amb altres components per a assolir una geometria idònia que maximitzi l'energia d'unió entre aquestes.

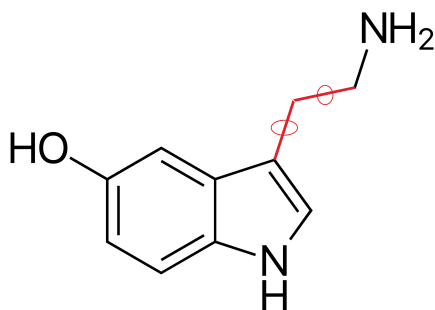


Figura 1: Enllaços rotables de la serotonina. Imatge d'ús públic sense copyright [18].

1.5 Hipòtesis i objectiu final

L'objectiu final d'aquesta investigació és estudiar i valorar les interaccions entre dues importants macromolècules (CCL21 i heparina) per a crear un entorn quasi-perfecte per a les cèl·lules T mencionades amb anterioritat i la seva reproducció. Per a aconseguir-ho, aquestes interaccions suposadament electroestàtiques han de garantir un enllaç suficientment fort que permeti adherir les cèl·lules T a un hidrogel i així simular un entorn que s'assembli molt a l'interior d'un gangli limfàtic. Examinant-ne les interaccions implicades, es pot determinar si l'hidrogel és l'adequat per a la simulació, o bé si s'hauria de considerar una proteïna diferent a la que és objecte d'estudi.

En el cas d'aquest estudi, es fa servir un hidrogel creat principalment amb heparina i un químic anomenat polietilenglicol per a simular l'entorn del gangli limfàtic. A través de les investigacions i els resultats experimentals obtinguts al laboratori, s'ha descobert que una proteïna específica que es troba a la membrana cel·lular, anomenada CCL21, carregada electropositivament, promet una de les proteïnes més electropositives detectades a la membrana cel·lular. Aquesta troballa suggereix una forta interacció electrostàtica entre aquesta proteïna i l'heparina, una de les molècules més electronegatives conegudes fins avui dia, que serà un dels principals punts focals d'aquest estudi.

En el marc de la hipòtesi proposada s'investigarà si la interacció entre la proteïna CCL21 i l'heparina és basa únicament en interaccions electroestàtiques, o bé si n'hi ha d'altres prou significatives que desmenteixin aquesta troballa. A més, serà crucial identificar els llocs d'unió utilitzats per cada component i determinar-ne la disposició geomètrica resultant. En definitiva, caldrà validar si aquesta interacció és prou eficaç per a contribuir significativament als estudis d'immunoteràpia.

L'objectiu principal d'aquest estudi serà explorar les principals interaccions entre aquests dos elements, inclosos els càlculs electroestàtics que els regeixen. S'utilitzaran mètodes tant analítics com numèrics per a investigar i resoldre aquestes interaccions.

En resum, els objectius, punt per punt, seran:

- Determinar la millor estructura digital de la molècula CCL21.
- Estudiar-la.
- Trobar-ne el potencial electroestàtic a la superfície.
- Realitzar els mateixos passos per a una estructura fiable de l'heparina.
- Comprovar la interacció i geometria assolida (el punt i orientació més probables d'unió) d'aquestes dues molècules.
- Verificar si es tracta d'una interacció purament electroestàtica o hi intervenen d'altres.
- Determinar finalment si es prou bona per a l'estudi en immunoteràpia.

2 Mètode

2.1 Equació de Poisson-Boltzmann

A partir de la definició dels macro-ions esmentada anteriorment, s'estudia el cas simple d'un macro-ió carregat negativament envoltat de càrregues positives a pocs nanòmetres de la superfície. En aquest cas, es pot aproximar un camp elèctric lineal i uniforme E amb direcció $-\hat{x}$, a causa de la curta distància. Si es considera una distribució uniforme, es pot definir $-\sigma_q$ com a distribució superficial de càrrega del macro-ió (Cm^{-2}), mentre ρ_q (Cm^{-3}) defineix la distribució volumètrica del núvol de càrregues positives que l'envolta. És a dir, aplicant la llei de Gauss, es té que el camp elèctric a la superfície és:

$$\mathcal{E}_{sup} = \frac{-\sigma_q}{\varepsilon} \quad (1)$$

On ε representa la permitivitat relativa del medi aquós (també anomenada constant dielèctrica de l'aigua) que se sol establir a 78.4 a temperatura ambient. Pel que fa a les càrregues del núvol, altre cop per la llei de Gauss es pot deduir que:

$$\frac{d\mathcal{E}}{dx} = \frac{\rho_q}{\varepsilon} \quad (2)$$

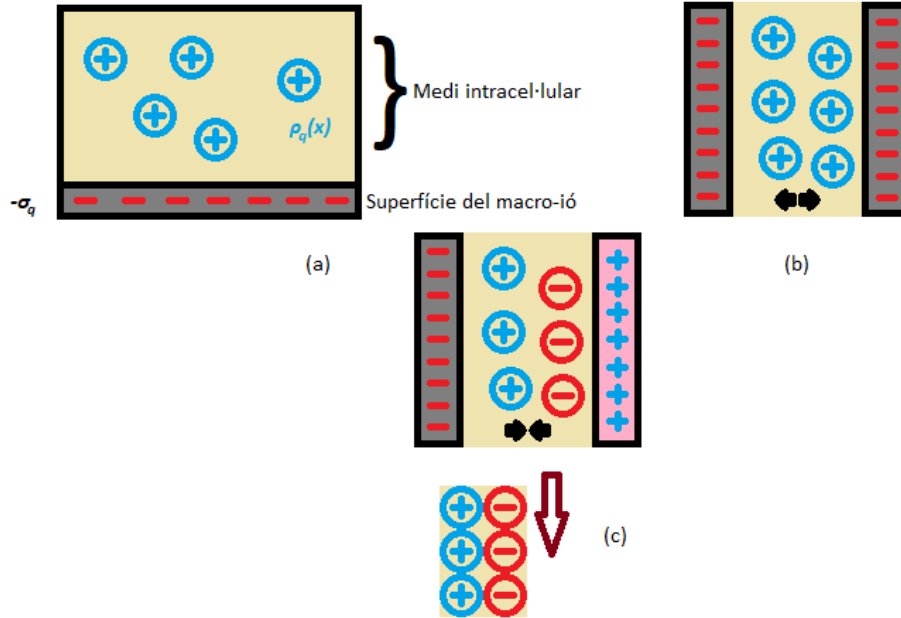


Figura 2: (a) Esquema simplificat del macro-ió rodejat pel núvol carregat positivament. Podem observar les distribucions de càrrega esmentades anteriorment, recalcant que només hi ha pocs nm de separació entre superfície i càrregues lliures. (b) Dos macro-ions de símbol igual interactuant: hi ha repulsió entre ells, però els contraions es mantenen pressionats per ambdues bandes. (c) Quan dos macro-ions de símbol diferent s'acosten, els contraions s'ajunten i surten dispersats, guanyant entropia.

Un cop determinades aquestes condicions, es pot començar a plantejar com determinar els potencials de cada banda. És interessant començar pel cas més simple: el núvol està carregat de ions monovalents positius. Sorgeix, però, un problema: no es pot determinar el camp elèctric sense saber la distribució de càrregues i; per tant, s'han de determinar aquestes. A la vegada, hom pot veure que en estar carregades aquestes partícules, elles mateixes afecten la distribució de les altres. És per això que hem de determinar dos aspectes a la vegada: la distribució de càrregues que afecta el potencial, i el potencial que afecta la distribució de càrregues.

Tenint en compte que el potencial pot ser expressat en funció de la distribució de càrregues, es pot definir una distribució mitjana de càrregues $\langle \rho_q \rangle$, que representa la mitjana d'aquestes partícules pertanyents al núvol, i és possible en tenir innumerables contraions repartits al voltant del macro-ió en estudi, ja que es tracta d'un element biològic de gran mida. És a dir, cada ió notará l'efecte de molts altres, i les fluctuacions en el potencial seran més aviat mínimes. Aquesta aproximació s'anomena aproximació del camp mitjà, i el potencial $V(x)$ que la regeix té com a nom camp mitjà (o mean field en anglès). Per simplicitat se segueix escrivint ρ_q .

Tanmateix, s'ha de determinar la manera de trobar el potencial elèctric a partir de la densitat de càrrega i, per contra, com trobar la densitat de càrrega a partir del potencial elèctric. Començant per la concentració de contraions, i.e. $c_+(x)$ (cal recordar que seguim en el cas simplificat monovalent), suposant un medi aquós no salí, la concentració lluny de la superfície tendeix a zero. Per altra banda, el potencial electroestàtic de qualsevol contraió és $eV(x)$, ja que s'ha considerat un camp constant a x . Suposant que cada ió es mou de manera independent en un potencial fix $V(X)$ (com hem mencionat abans), es pot veure que es tracta d'una distribució de Boltzmann, que podem escriure tal que $c_+(x) = c_0 e^{-eV(x)/k_B T}$, amb $c_0 = c_+(0)$ tal que $V(0) = 0$ (és a dir, la concentració a la superfície), i suposant que $q = +e$, ja que es tracta del cas monovalent sense sals al medi.

En aquest senzill cas, podem definir ρ_q com la concentració $c_+(x)$ per la càrrega $q = +e$, i tenint en compte que $\mathcal{E} = -dV/dx$ per al camp establert anteriorment i; tenint en compte 2, s'arriba a l'equació de Poisson $\frac{d^2 V}{dx^2} = -\rho_q/\epsilon$. Cal remarcar, a més, que $-\sigma_q$ és representada pels aminoàcids carregats negativament presents a la superfície de la macromolècula, i ρ_q correspon a la distribució dels contraions més les sals presents en aquest medi, com solen ser-ho Na^+ i Cl^- . A més, la càrrega de tots aquests components dependrà del pH del medi.

Aquí, apareix el problema de vincular aquestes dues equacions. Per una banda; tenim les càrregues fixes a la superfície, que vénen determinades per la relació de Poisson; per l'altra, però, tenim les lliures que són determinades a partir de la distribució de Boltzmann. És ràpid de veure que, tant una com l'altra, s'afecten mútuament: el canvi de potencial modifica la concentració d'ions, i el canvi de l'última en modifica el primer. En conseqüència, s'han de determinar a la vegada, fet que acostuma a ser complex ja que no sol existir una solució analítica en ser no-lineal, llevat d'alguns casos molt simples amb geometries molt senzilles. Hom podria pensar de fer una aproximació lineal, però aquesta només seria vàlida per a un potencial inferior a $k_B T$, que com bé s'ha mencionat prèviament, no succeeix en l'àmbit d'estudi. En unir les dues relacions, s'obté:

$$\frac{d^2 V}{dx^2} = -\frac{ec_0 e^{-eV(x)/k_B T}}{\epsilon} \quad (3)$$

Aquesta equació pot millorar-se per a entendre'n millor el comportament. Un canvi de terme interessant és el del potencial adimensional reescalat, que permet comparar la relació directa, l'ordre de magnitud, entre l'energia tèrmica de les molècules a temperatura ambient, $k_B T$ amb el potencial de la superfície en el medi. Dit altrament, representa el potencial electroestàtic que ve alliberat a causa de l'agitació tèrmica a certa temperatura, com l'ambient, o en aquest cas, la corporal. Es pot reescriure com:

$$\bar{V} = \frac{eV(x)}{k_B T} \quad (4)$$

Això modifica l'equació anterior tal que:

$$\frac{d^2 \bar{V}}{dx^2} = -\frac{e^2 c_0 e^{-\bar{V}}}{\epsilon k_B T} \quad (5)$$

On hem reescrit $c_+(x)$ com $c_0 e^{-\bar{V}}$.

Aquest raonament fa que Niels Bjerrum (1879-1958) trobi una altra magnitud d'ampli interès d'estudi: la longitud de Bjerrum. Aquesta indica a partir de quina distància la interacció electroestàtica entre dues partícules de càrrega e és equivalent a l'energia tèrmica kT , o en altres paraules, quant es poden acostar dues molècules de càrrega de mateix signe sense que es repel·leixin, considerant que es disposa d'una energia de valor $K_B T$. En aigua a temperatura ambient amb ions monovalents, es dona a 0.71nm. La fórmula que estableix és:

$$l_B = \frac{e^2}{4\pi\epsilon k_B T} \quad (6)$$

Que correspon a equiparar la fórmula de l'energia elèctrica, $E = kq_1q_2/r$, amb l'energia tèrmica, $E = K_B T$, amb ϵ la constant dielèctrica de l'aigua i tenint dues càrregues idèntiques d'igual signe. Aïllant r , es troba el valor mencionat. Introduïda a l'equació anterior, acaba proporcionant la coneguda equació de Poisson-Boltzmann sense sals en el medi i amb ions monovalents:

$$\frac{d^2 \bar{V}}{dx^2} = -4\pi l_B c_0 e^{-\bar{V}} \quad (7)$$

Utilitzada per a la superfície que interessa analitzar [3]. Aquesta equació, com que acostuma a ser no-lineal i; per tant, irresoluble manualment, es calcula de forma numèrica amb l'aplicació APBS (Automatic Poisson-Boltzmann Solver) al VMD, programa que és utilitzat en aquest treball per a estudiar les estructures en detall. El procés seguit és descrit amb més detall a l'annex.

Com a tota equació diferencial, cal afegir condicions inicials i de contorn per a trobar la solució desitjada. A partir de la llei de Gauss mencionada a 1, aplicant el potencial adimensional trobem que:

$$\frac{d\bar{V}}{dx}_{sup} = 4\pi l_B \frac{\sigma_q}{e} \quad (8)$$

Amb aquesta condició, i establint que $\bar{V}(0) = 0$ i que:

$$\frac{d\bar{V}}{dx}_{\infty} = 0 \quad (9)$$

Es poden determinar les solucions per a aquest cas simplificat. Per a una solució analítica, necessitaríem una funció que en segona derivada fos equivalent a la seva exponencial. El llibre de P. Nelson [2], pàgina 236, en cita les més conegudes. Aquest concepte serà important per a entendre el següent apartat.

2.2 PB per a medis salins

El nombre d'aminoàcids àcids i bàsics determina la càrrega d'una proteïna. Ja siguin a la pròpia proteïna o al medi, aquests interactuen entre ells deixant carregada positiva o negativament la molècula estudiada. Això condueix a pensar en el comportament les càrregues mòbils de la superfície mencionada anteriorment i els ions presents a la dissolució: hi ha càrregues positives i negatives, sobretot en medis aquosos com és el cos humà. Conseqüentment, és necessari reescriure l'equació de Poisson-Boltzmann per a diferents tipus de càrregues, en medis salins.

En els casos realistes, on normalment no tenim solució analítica, s'han de considerar tots els factors: el macro-ió serà segurament heterogeni, carregat de forma positiva i negativa a diferents zones de la molècula; els ions presentaran una situació similar, i, a més, seran presents en un medi salí, que provoca que la concentració a l'infinit c_{∞} de les sals mai sigui exactament zero. A causa d'això, serà millor canviar la condició inicial del potencial: se n'ha d'establir un tal que $V(x)$ tendeixi a zero a grans distàncies. Això duu a: $c_+(x) = c_{\infty} e^{-eV(x)/k_B T}$ i $c_-(x) = c_{\infty} e^{-(-e)V(x)/k_B T}$, suposant que ens trobem en un medi salí on tenim components de càrregues e i $-e$.

Unint aquestes dues concentracions a l'equació de Poisson-Boltzmann ja esmentada, s'obté:

$$\frac{d^2 \bar{V}}{dx^2} = -\frac{1}{2} 8\pi l_B c_{\infty} \left(e^{-\bar{V}} - e^{\bar{V}} \right) \quad (10)$$

On escrivim $-1/2 \cdot 8$ per a la següent simplificació: Peter Debye (1884-1966), de manera similar a Bjerrum, estipula una magnitud interessant en aquesta expressió: la longitud de Debye, que en aquests casos es representa com $\lambda_D = (8\pi l_B c_{\infty})^{-1/2}$. Aquesta magnitud ens indica fins a on persisteixen els efectes electroestàtics d'una solució. En altres paraules, a distàncies més grans que la longitud de Debye, els camps elèctrics comencen a patir un efecte d'apantallament que decau de forma exponencial. Per a l'aigua a temperatura ambient, és aproximadament 0.7nm, suposant 100mM per a cada ió.

Aquesta longitud pot ser implementada dins l'equació de Poisson-Boltzmann com:

$$\frac{d^2\bar{V}}{dx^2} = -\frac{1}{2}\lambda_D^{-2} \left(e^{-\bar{V}} - e^{\bar{V}} \right) \quad (11)$$

Finalment cal remarcar doncs, que s'utilitza el programari APBS [12] per a la resolució numèrica ja que l'analítica en aquest cas és impossible en ser geomètricament molt complexa.

No cal fer esment de l'ampli rang d'usos que proveeix aquesta equació en tots els diferents camps de la física: electroestàtica, sistemes biològics, sistemes electroquímics, sistemes biofísics, ciència de materials, entre molts altres. Això és degut a la gran quantitat de situacions realistes on intervé la presència de diferents càrregues en un medi determinat, com és el medi aquós salí del cos humà en aquest cas.

En altres paraules: en aquest treball es farà ús de l'equació de Poisson-Boltzmann per a trobar el potencial a la superfície de la proteïna CCL21 i a l'heparina, mitjançant el programa APBS, que proporciona la solució numèrica de l'equació amb unes condicions prèviament estipulades, com a qualsevol equació diferencial, mencionades a l'apèndix. Per a preparar les coordenades d'aquesta proteïna, a més, s'utilitzarà del programari VMD, present a l'apèndix.

2.3 Acoblament molecular

A l'àmbit d'interaccions moleculars hi és present un dels factors més importants, l'acoblament molecular. Aquest consisteix en determinar les posicions més favorables, i.e. que tinguin l'energia d'unió més elevada segons els diferents factors que hi contribueixen, a partir de diferents mètodes i aproximacions analítiques i numèriques basats en diversos models de simulacions moleculars. El conjunt d'aquestes funcions s'anomena, en anglès, *scoring functions*.

En aquest estudi es farà ús del programa AutoDock [15], que té la següent fórmula per al càlcul de l'energia d'unió:

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + (V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf}) \quad (12)$$

On V representa cada avaluació dels dos participants en l'acoblament: L representa el *ligand*, la molècula que modificarà la seva forma per a acoblar-se a la proteïna o receptor, representada per P . En altres paraules, $L-L$ i $P-P$ són les energies de cada element per separat (energies intramoleculars) i $P-L$ les energies entre els dos components. A més, s'avaluen aquestes energies tant en estat separat (*unbound*) com ajuntat (*bound*). El terme ΔS_{conf} representa l'entropia perduda en ajuntar-se lligand i receptor.

Cada terme V es calcula de la següent forma:

$$V = W_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + W_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{elect} \sum_{i,j} \frac{q_i q_j}{e(r_{ij}) r_{ij}} + W_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{(-r_{ij}^2/2\sigma^2)} \quad (13)$$

On W són les constant de pes (*weighting constant*) optimitzades a partir de constants d'unió determinades experimentalment. Cada terme representa un contribuent a l'energia d'unió diferent: el primer de tots representa el causat a partir de les forces de Van der Waals, fent ús d'un potencial basat en interaccions de dispersió i repulsió. El següent és el que apareix en presència de ponts d'hidrogen. El tercer és el conegut potencial de Coulomb per a interaccions electrostàtiques. El final és el terme representatiu del potencial del solvent on es troben els components per acoblar. Les explicacions més detallades són al document adjunt de l'AutoDock [17].

Com s'ha mencionat abans a la introducció, el programa fa ús del nombre de torsions al lligand, per a trobar la configuració més adient en contacte amb la proteïna o receptor. Per defecte, el nombre d'avaluacions està configurat a 10. És a dir, el programa troba les deu configuracions amb les energies d'unió més elevades segons les torsions assignades al lligand.

3 Resultats

3.1 Selecció del model proteic

La proteïna desitjada per a l'estudi és la CCL21. Al *Protein Databank*, lloc on s'emmagatzemen les estructures moleculars de forma digital, s'hi troben dos models: 2L4N i 5EKI. Aquestes estructures de proteïnes han estat obtingudes de forma experimental, detallades als annexos. En comparar les estructures, es determina que l'estructura més adient per a aquest estudi és la 5EKI, més concretament la cadena E (ja que el model al PDB presenta sis cadenes similars de la mateixa proteïna). Aquesta selecció és a causa de la important presència d'uns aminoàcids a la terminal C (a partir del 70) de l'estructura 5EKI per a la interacció amb altres proteïnes com la CCR7, que no són presents a la 2L4N. El protocol seguit a la comparació s'explica en detall als annexos.

Cal remarcar però, que cap de les dues estructures es considera completa. La cadena E de la 5EKI manca els 5 primers aminoàcids, però no s'han considerat crucials per a aquest estudi. L'aparença d'aquesta molècula és la següent:

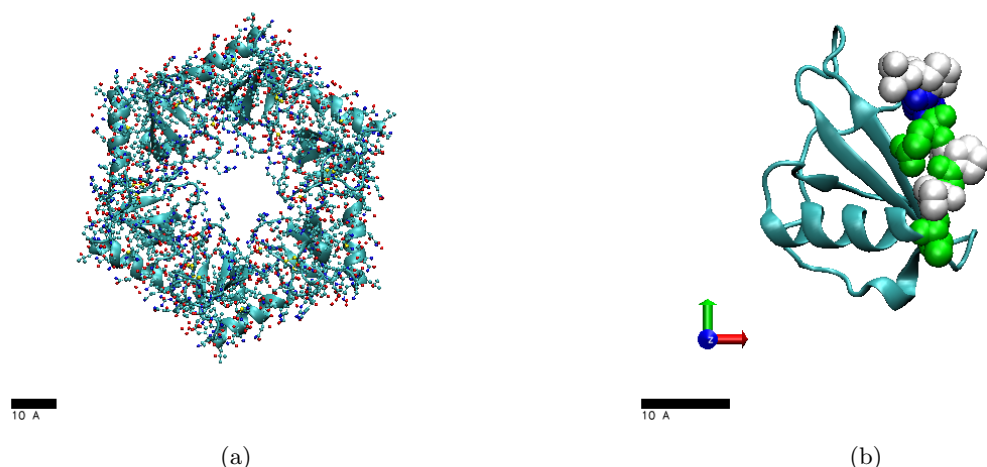


Figura 3: (a) Estructura de codi 5EKI, que presenta sis cadenes similars a causa del mètode d'obtenció de cristal·lografia de rajos X. Presenta 3639 àtoms, sense hidrògens. (b) Cadena E de l'estructura acabada de mencionar, que es considera el més bo (veure Annex) amb la terminal C d'aquest model, lloc crucial per a la interacció amb la proteïna CCR7. Presenta 583 àtoms, sense hidrògens. Els gràfics utilitzats per a l'estructura base són NewCartoon. A la primera fotografia utilitzem CPK per a veure els residus més clarament, i a la segona indiquem la terminal amb VDW colorats segons el tipus de residu.

Es calcula, a més, la càrrega total de la cadena E, que proporciona un resultat de 7e. Per a la interpretació dels arxius pdb, es poden consultar els annexos.

3.2 Càlcul electrostàtic superficial de la proteïna

Mitjançant l'APBS es resolen les equacions de Poisson-Boltzmann mencionades a la secció anterior. És interessant recalcar les condicions prèvies establertes. Establint unes concentracions de 100mM NaCl. Col·locant una temperatura de 298.15K (T ambient) i comparant-la amb 309.65 (T mitjana del cos humà), una permeabilitat del medi de 78.54 i un pH de 7.0, es troben les següents figures:

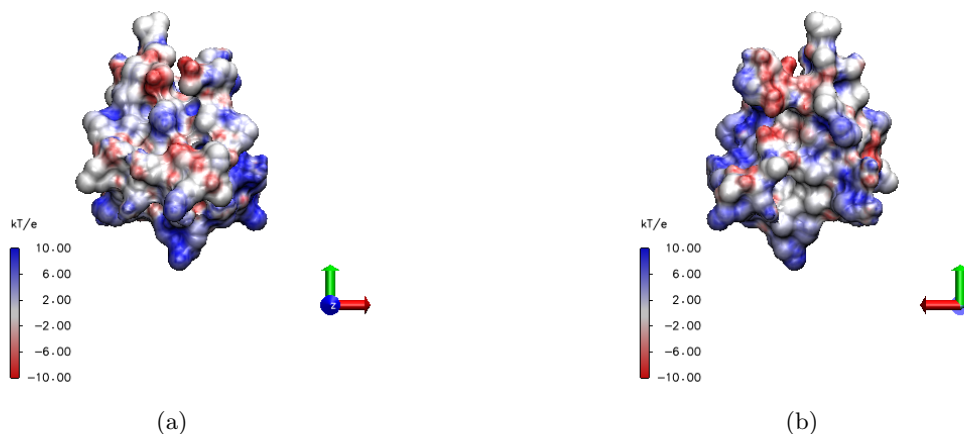


Figura 4: (a) Representació del potencial electroestàtic a la superfície de la proteïna CCL21 (5EKI) a una temperatura corporal de 309.65K. (b) Mateixa figura, rotada 180° a l'eix y. Els càlculs han estat realitzats amb APBS (consultar Annex) i les representacions amb VMD (Surf + Volume).

Cal recalcar que s'ha obtingut aquesta imatge fent servir un rang de potencials de -10 a +10 unitats de kT/e que representa el potencial adimensional. Tenint en compte l'equació 4 es pot veure que $V(x)$ s'expressa en unitats de kT/e multiplicats per aquest potencial adimensional. A més, a temperatura corporal, kT/e té un valor aproximat de 26.68mV, pel que es té una rang de 266mV a -266mV segons els valors trobats amb més freqüència a l'arxiu dx. Com que les figures obtingudes amb la temperatura ambient són molt similars a les del cos, s'han mostrat a l'annex.

Es pot veure que la zona blava, sobretot a la imatge a 180° , correspon aproximadament a la terminal que interessa estudiar, vista abans. Es pot deduir que es tracta d'una estructura molt inhomogènia, que de fet en part pot semblar útil a l'hora d'adherir-se a l'heparina, de càrrega negativa, ja que les zones positives que veiem al gràfic són relativament lluny de les negatives, que afavoreix la unió sense repel·lir-se.

3.3 Heparina

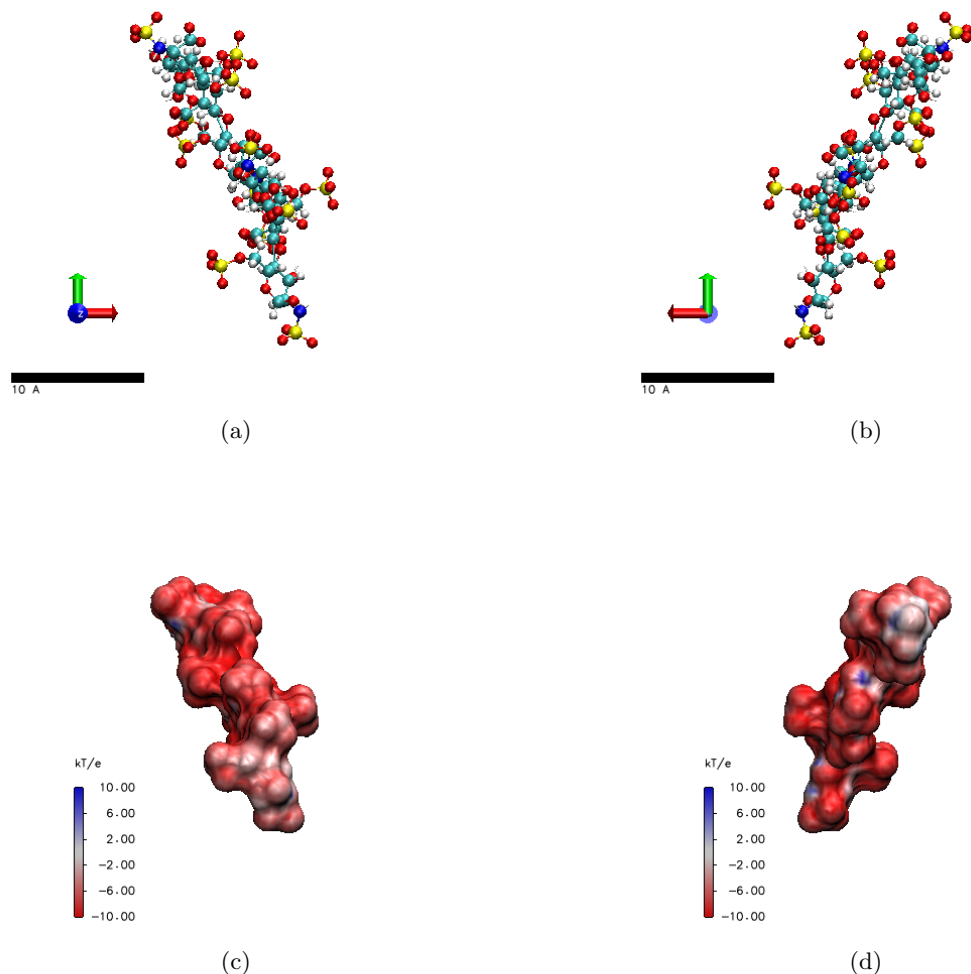


Figura 5: (a) Estructura general de l'heparina [14], amb 203 àtoms, hidrògens inclosos. (b) Estructura girada a 180° sobre l'eix y. (Tant (a) com (b) estan representades en format CPK, amb el programa VMD) (c) Potencial a la superfície de l'heparina, a temperatura corporal (veure resta de condicions a l'apèndix). (d) Potencial superficial de l'heparina, a 180° en l'eix y. El color blau representa la tendència cap a càrregues positives i, el vermell, negatives. (Tant (c) com (d) estan representades en format Surf-Volume, al VMD).

A les dues primeres imatges s'observa l'estructura general de l'heparina amb escala [14]. Al segon parell, es pot observar el comportament del potencial a la superfície mitjançant l'APBS. Cal fixar-se, com destaca, la distribució gairebé homogènia de càrregues a la superfície, on gairebé no hi és present cap càrrega positiva ni neutra, llevat d'alguns punts molt discrets. Comparant a amb b, podem veure que la zona electropositiva més destacada és la cantonada inferior dreta, on hi tenim un intens color blau. A més, correspon justament amb una part dels residus 70-77 de la cadena E, que segons els documents respecte al model 5EKI (mencionats a l'apèndix), són primordials en les interaccions d'aquesta proteïna. Per altra banda, l'heparina, tot i tenir càrrega negativa a tota la superfície, sembla ser més intensa a la part superior. Això condueix a pensar que, en una primera aproximació, les dues estructures s'unirien per aquella zona.

3.4 Interacció molecular

Seguidament, després de realitzar l'estudi dels potencials electroestàtics a la superfície per a cadascuna de les dues molècules, es fa ús del programa AutoDock, que es pot consultar als annexos, per a trobar la millor posició d'acoblament del *ligand*, la nostra heparina; amb el receptor, que és l'estructura 5EKI. Es consideren quatre casos: interacció de l'heparina amb la zona de potencial més electropositiu de la proteïna CCL21 (imatge a Annexos) i amb la cua d'aquesta (residus 70-77), la qual ha estat esmentada anteriorment com a zona de gran interès. Dins aquests dos aspectes a considerar, podem definir les torsions de l'heparina (consultar Introducció) com a rotables o no rotables, que sumen un total de quatre casos.

En haver realitzat els càlculs i preparacions necessàries, explicades amb detall a l'annex, s'obté que el cas amb l'energia d'unió (o binding energy) més elevada és el cinquè sobre deu posicions que ha calculat el programa, de la interacció, amb heparina flexible, amb la cua de la proteïna. Cal recordar, com està mencionat a mètodes, que el programa busca les posicions del lligand respecte la proteïna que roman rígida, que proporcionen l'energia d'unió més forta. La taula obtinguda amb el programa és la següent:

| Rank | Sub-Rank | Run | Binding Energy | Cluster RMSD | Reference RMSD |
|------|----------|-----|----------------|--------------|----------------|
| 1 | 1 | 5 | -27.72 | 0.00 | 120.86 |
| 2 | 1 | 7 | -26.54 | 0.00 | 124.74 |
| 3 | 1 | 9 | -22.08 | 0.00 | 123.46 |
| 4 | 1 | 1 | -20.19 | 0.00 | 125.23 |
| 5 | 1 | 8 | -15.95 | 0.00 | 124.49 |
| 6 | 1 | 3 | -15.00 | 0.00 | 124.32 |
| 7 | 1 | 10 | -14.53 | 0.00 | 123.45 |
| 8 | 1 | 6 | -14.20 | 0.00 | 126.05 |
| 9 | 1 | 2 | -13.76 | 0.00 | 124.21 |
| 10 | 1 | 4 | -11.24 | 0.00 | 123.47 |

Figura 6: Taula de les deu millors posicions trobades pel programa AutoDock al cas heparina flexible - cua CCL21.

La resta de taules es poden consultar als annexos. L'energia d'aquest cas és de -27.72kcal/mol, tal i com proporciona el programa. En esbrinar els diferents components d'aquesta energia, s'observa que l'energia per interacció electroestàtica segueix sent la dominant, i no només al cinquè cas, sinó a tots. L'arxiu dlj amb les dades corresponents es pot trobar a l'Annex. Per tant, es continua l'estudi del cinquè cas. Consultant l'arxiu dlj trobem que l'energia electroestàtica representa uns -33.00kcal/mol. Es té, a més, que el conjunt d'energies per Van der Waals, ponts d'hidrogen i del solvent sumen en total -2.77kcal/mol, molt petit comparat amb l'energia electroestàtica d'aquest cas. L'energia total interna és, a més, -2.42kcal/mol, que es correspon amb l'energia del sistema no-lligada (unbound) i per tant s'anul·len mútuament. Finalment, l'energia lliure per torsió, que és de signe contrari, representa +8.05kcal/mol del total. La suma d'aquests termes ens proporciona l'energia lliure d'unió esmentada amunt.

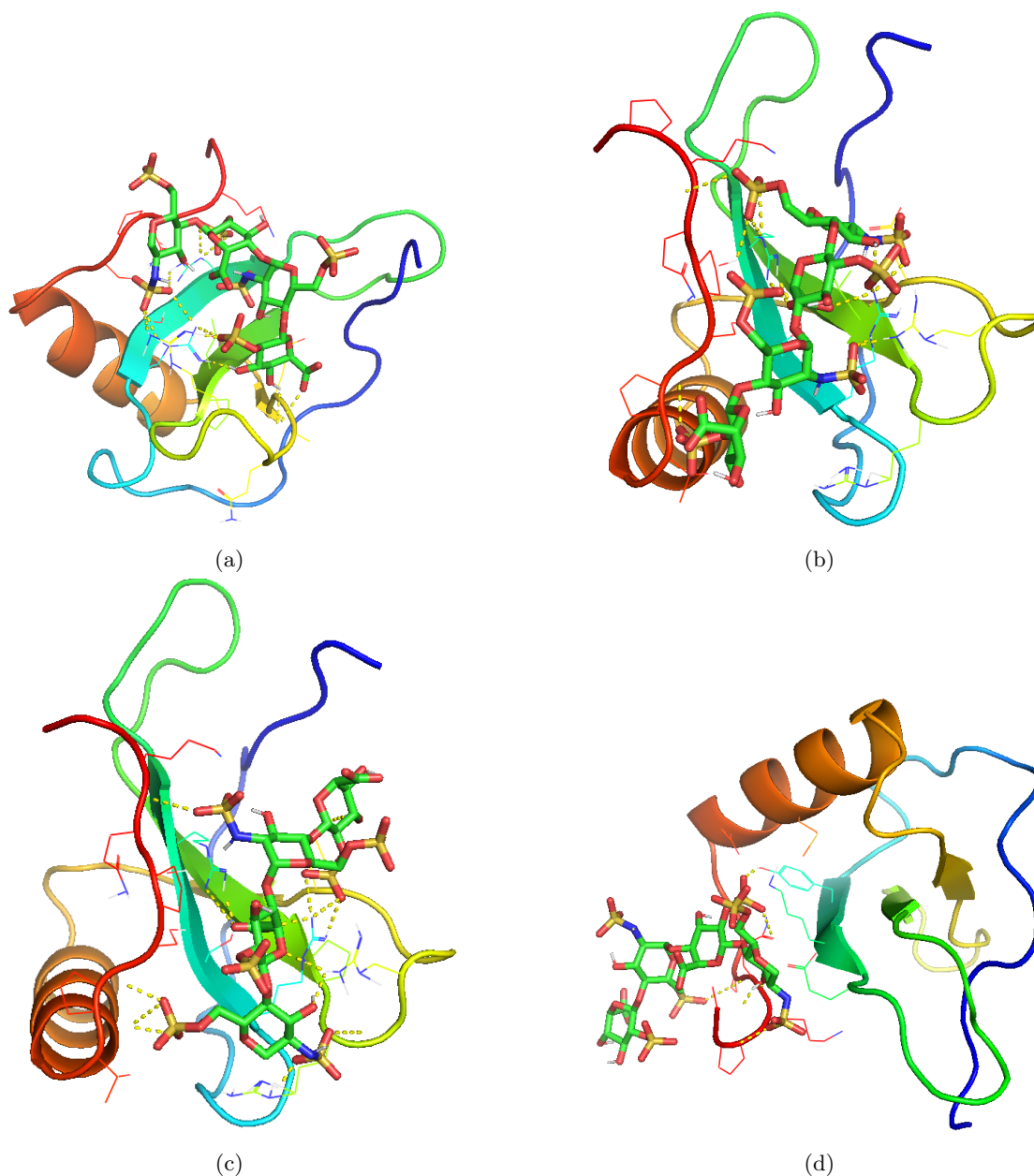


Figura 7: (a) Configuració trobada per a l'heparina, flexible, a la zona de càrrega superficial positiva més elevada. (b) Configuració trobada per a l'heparina, flexible, a la cua de la proteïna CCL21 (residus 70-77). (c) Configuració trobada per a l'heparina, rígida, a la zona de càrrega superficial positiva més elevada. (d) Configuració trobada per a l'heparina, rígida, a la cua de la proteïna CCL21 (residus 70-77). Per a aquestes imatges ha estat escollit el millor angle per a visualitzar el màxim nombre d'enllaços presents, senyalitzats de color groc discontinu.

Gràcies a l'opció de seleccionar la zona d'interès, s'han pogut determinar aquestes imatges. El motiu pel qual s'han considerat tant l'heparina flexible com rígida, és a causa del gran nombre de torsions presents al lligand, un total de 27. Com bé esmenta el manual de l'AutoDock, a partir de 10 torsions cap amunt els resultats deixen de ser tan fiables. Per tant s'han considerat els casos rotables i no-rotables.

L'altre motiu pel qual s'han estudiat dues zones diferents d'interacció és a causa de l'interès a ambdues. La zona de la cua (70-77) és, segons estudis previs i el mateix document publicat al PDB per l'estructura 5EKI [10], de gran importància a l'hora d'interactuar amb altres compostos, en especial la macromolècula

CCR7, una altra considerada com a alternativa d'estudi a la CCL21, amb menys electropositivitat que aquesta última. Per altra banda, s'ha estudiat també la zona corresponent, segons l'APBS, a la superfície més carregada positivament de l'estructura 5EKI per la proteïna CCL21.

Els resultats per a la interacció electroestàtica més elevada ja han estat mencionats abans, que de fet corresponen a la imatge (b). Pel que fa a (a), el resultat més elevat ha estat de -24.14kcal/mol a la tercera configuració o *run*, amb una energia electroestàtica de -29.52kcal/mol, altres intermoleculars (Van der Waals, ponts d'hidrogen, solvent) de -2.67kcal/mol, i una energia lliure de torsió de +8.05kcal/mol. Podem observar que en aquest cas també es dona una situació molt similar a (b), on predomina l'energia electroestàtica.

Pel que fa als rígids, tenim primer (c), on l'energia lliure d'unió és, aproximadament, -18.88kcal/mol a tots els casos (ja que no hi ha torsions). L'energia electroestàtica trobada ha estat de -25.47kcal/mol, i altres intermoleculars de -1.47kcal/mol. L'energia lliure de torsió és de +8.05kcal/mol. Tot i ser l'heparina rígida, es repeteixen els resultats anteriors, tot i que aquest cop l'energia lliure no supera el llindar dels -20.00kcal/mol, que és la interacció trobada de forma natural més forta que es coneix. En cas de superar-lo, com es veu en els casos flexibles, es considera una interacció rellevant a l'àmbit biofísic.

Finalment, pel cas (d), es té una energia lliure d'unió de -3.26kcal/mol, molt inferior al que s'esperava. Probablement sigui a causa d'haver establert l'heparina rígida en una àmplia zona com és la de la terminal C de la proteïna CCL21. L'energia electroestàtica és de -6.26kcal/mol, i les altres intermoleculars, -5.06kcal/mol. En aquest cas no s'obté el que s'esperava, però tot i així es poden considerar els resultats obtinguts com a prou bons, ja que l'heparina rígida a la zona més electropositiva de la superfície de la proteïna ha donat resultats força elevats.

4 Conclusions i comentari final

En aquest estudi, s'ha fet un esforç per a identificar i corroborar la interacció electroestàtica entre dues molècules que mostren potencial per a més investigacions al camp de la immunoteràpia.

La interacció electroestàtica trobada i estudiada entre l'heparina, medicament utilitzat com a anticoagulant; i la proteïna CCL21, present a les membranes cel·lulars com ara de les cèl·lules T; ha demostrat ser molt rellevant en el sentit biofísic, com s'ha mencionat abans.

Tant la proteïna com el lligand (heparina) han demostrat presentar uns potencials de superfície molt elevats i de signe contrari que en facilita la interacció. Al començament, la inhomogeneïtat de càrrega superficial de la proteïna semblava poder crear cert obstacle a l'hora d'acoblar-la amb el lligand. No obstant, després d'haver seleccionat l'àrea d'interès (terminal C) mencionada per diversos estudis per la seva elevada interacció amb altres molècules, s'ha pogut observar que la interacció electroestàtica és primordial, tornant gairebé irrelevantes les interaccions per Van der Waals, ponts d'hidrogen o la mateixa energia del solvent.

Conseqüentment, es pot considerar que la interacció electroestàtica entre aquests dos elements compleix la hipòtesi estipulada prèviament, i.e. és la interacció principal de les dues molècules de manera rellevant. Per tant, pot convertir-se en element d'estudi per a les teràpies contra el càncer.

Aquest fet acabat de mencionar es pot comprovar a l'apartat anterior, on, s'ha pogut demostrar que en tots els casos, la interacció electroestàtica és la major força intermolecular. S'ha de descartar, però, la interacció de l'heparina rígida amb la terminal C de la proteïna, que com era d'esperar, ha donat uns resultats allunyats dels esperats. Tot i aquest petit incís, els altres tres casos han demostrat aquest fet, ja sigui en rígida o flexible, de la importància d'aquesta interacció envers les altres interaccions intermoleculars.

Per altra banda, però, cal profunditzar l'estudi de la interacció d'aquestes dues molècules. Si bé aquest treball ha estat una primera aproximació, és clar que resulta necessari una investigació detallada de tots els factors, e.g. un programa que permeti considerar l'heparina sencera amb totes les seves torsions, juntament amb una proteïna CCL21 no rígida, per a uns resultats més fiables. L'ús d'aquest tipus de programes sol estar restringit a ordinadors molt potents, pel que s'ha hagut de fer ús de programes més simplificats, ja que l'objectiu d'aquest treball és una primera aproximació de la interacció electroestàtica entre les dues molècules mencionades anteriorment, i no pas la familiarització i desenvolupament de programes d'estudi d'interaccions entre macromolècules.

Per finalitzar, personalment, he pogut eixamplar els meus coneixements en biofísica de manera exhaustiva. La familiarització dels programes ha estat veloç, i he estat capaç d'entendre tot el procés que té lloc darrere les interaccions moleculars i la seva conseqüent transcripció al món digital. Els obstacles han estat mínims, ja que tots els programes proporcionen guies i tutorials d'accés públic i gratuït per a facilitar la investigació científica. Gràcies a tots els coneixements adquirits durant aquesta experiència, considero que em poden ajudar àmpliament al món de la investigació física i, no solament a mi, sinó a tothom que ho desitgi i faci ús d'aquests programes tan fortament útils.

A Apèndix

A.1 Bibliografia

- [1] <https://www.youtube.com/watch?v=eckgmqiBBRo> “Avancem en la immunoteràpia contra el càncer amb materials 3D amb Ignacio Crespo” per ICMAB-CSIC
- [2] Nelson, P., “Biological Physics: Energy, Information, Life,” *Entropic forces at work*, pp. 230-231, First Edition (2003).
- [3] Nelson, P., “Biological Physics: Energy, Information, Life,” *Entropic forces at work*, pp. 232-235, First Edition (2003).
- [4] <https://www.ks.uiuc.edu/Research/vmd/> Pàgina principal del programari VMD.
- [5] <https://www.rcsb.org/structure/2l4n> Primer model considerat, 2L4N.
- [6] https://files.rcsb.org/pub/pdb/validation_reports/14/2l4n/2l4n_full_validation.pdf Informe sobre l’obtenció del model mencionat i els procediments aplicats.
- [7] <https://pubs.acs.org/doi/10.1021/bi201601k> Document informatiu sobre el model 2L4N de la proteïna CCL21.
- [8] <https://www.ks.uiuc.edu/Training/Tutorials/vmd/tutorial-html/> Tutorial utilitzat en aquest treball.
- [9] <https://www.rcsb.org/structure/5EKI> Segon model utilitzat.
- [10] <https://pubs.acs.org/doi/10.1021/acs.biochem.6b00304> Document informatiu del model 5EKI.
- [11] https://files.rcsb.org/pub/pdb/validation_reports/ek/5eki/5eki_full_validation.pdf Informe d’obtenció del model 5EKI. pp. 4, 9-10.
- [12] <https://server.poissonboltzmann.org/pdb2pqr> Convertidor en línia d’arxius PDB/PSF a PQR i APBS.
- [13] <https://github.com/MaxTFG/TFG> Arxius treballats per a consultar.
- [14] <https://amarolab.ucsd.edu/covid19.php> Arxius generats per a l’estudi de l’heparina.
- [15] <https://autodock.scripps.edu/download-autodock4/> Descàrrega de l’AutoDock.
- [16] <https://ccsb.scripps.edu/mgltools/downloads/> Descàrrega de les altres eines que fem servir amb l’AutoDock.
- [17] <https://autodock.scripps.edu/wp-content/uploads/sites/56/2022/04/AutoDock4.2.6UserGuide.pdf> Guia de l’usuari per l’AutoDock.
- [18] <https://en.wikipedia.org/wiki/Serotonin> Imatge sense drets d’autor de la serotonina.

A.2 Annexos

A.2.1 Estudi i selecció del millor model virtual

En fer ús del programa VMD, s'hauran d'utilitzar models virtuals de proteïnes principalment del *Protein Databank*, on hi ha emmagatzemades les que interessa estudiar en aquest treball. El primer model seleccionat per a estudiar és l'anomenat 2L4N, de la proteïna CCL21 com s'ha esmentat abans [4] [5]. Per a saber fer anar el programa, la pàgina conté alguns tutorials essencials. Els utilitzats en aquest treball han estat els mencionats a la bibliografia [8].

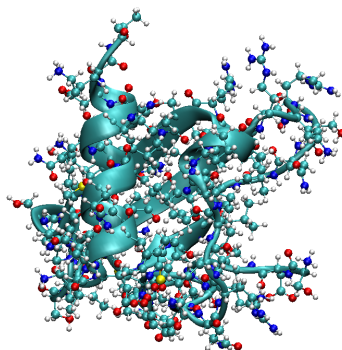


Figura 8: Estructura general del model 2L4N

La determinació de l'estructura d'aquesta proteïna s'ha realitzat mitjançant el mètode NMR (Nuclear Magnetic Resonance), que consisteix en col·locar la proteïna en un camp magnètic potent i constant mentre els seus àtoms són afectats per un altre camp magnètic feble i oscil·lant. En llegir la freqüència obtinguda d'aquest procés, se'n poden determinar els elements i molècules (aminoàcids) d'aquesta proteïna de manera força eficient [6].

En analitzar aquesta proteïna, descobrim que el model virtual només inclou els 70 primers aminoàcids, mentre menciona que en falten uns 40 a la terminal C.

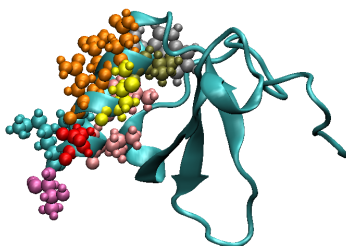


Figura 9: Terminal C del model 24LN (59 a final)

És interessant revisar el document adjunt ja que ens menciona els llocs d'unió (*binding sites*) de la proteïna CCL21 amb la CCR7, que és la proteïna localitzada a la capa més exterior de les cèl·lules T que facilita la interacció electrostàtica i és d'estudi primordial en aquest treball, ja que d'elles depèn el possible cultiu en 3D al laboratori [7]. En són els següents:

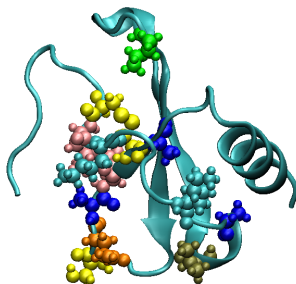


Figura 10: Residus 9-11, 16, 19, 21, 36, 47-50, 52 i 53 del model 24LN

En descobrir la manca d'aminoàcids a la terminal C, es decideix cercar si existeix un model virtual amb més aminoàcids determinats a la terminal C ja que són crucials a l'hora de la interacció amb l'heparina, perquè s'ha demostrat que el motiu pel qual la proteïna CCL21 atrau tan eficientment les cèl·lules T és gràcies a aquesta terminal que altres similars menys eficients no contenen. Es troba un model força interessant, de la mateixa proteïna CCL21, anomenat 5EKI, que s'ha determinat mitjançant cristal·lografia de rajos X [9].

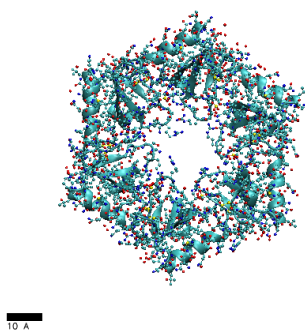


Figura 11: Estructura general del model 5EKI

S'observa que el model conté 6 formes de la mateixa proteïna (a causa de la cristal·lografia), pel que es decideix buscar quina és la més adient. Finalment es troba que la cadena E no conté cap col·lisió entre àtoms [10] [11] (el model no té àtoms massa a prop entre ells que puguin causar problemes a l'hora de calcular interaccions) i a més és el que menys residus mancants proporciona. S'escull finalment, doncs, utilitzar aquesta cadena per a l'estudi.

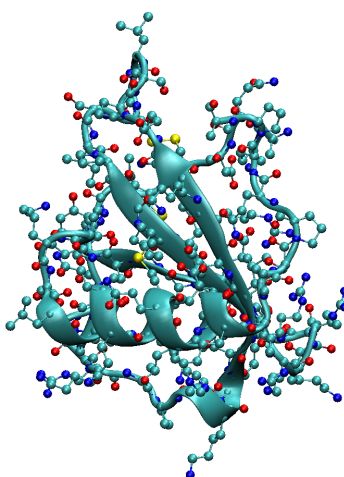


Figura 12: Cadena E del model 5EKI

Fent un cop d'ull al document adjunt d'aquest model es veu que menciona que els aminoàcids de la terminal C a partir del 70 fins el 79, són de gran importància en la interacció d'aquesta proteïna, els quals el model 2L4N no conté. Finalment, se selecciona el model 5EKI, sent el que més interessa utilitzar per a l'estudi, tot i no contenir els 5 primers aminoàcids de la terminal N, que de moment, es consideren no-crucials per als càlculs que ens interessin realitzar.

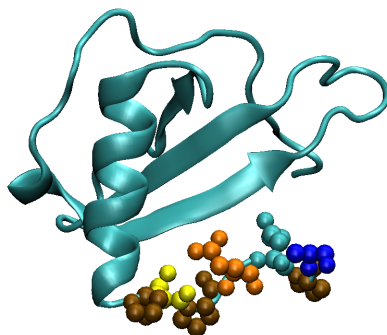


Figura 13: Terminal C (70-79) de la cadena E (5EKI)

A.2.2 Obtenció dels arxius per l'APBS

Per al correcte estudi i càlculs amb el model escollit, s'ha de perfeccionar la nostra proteïna, és a dir, afegir-hi els elements que puguin faltar-hi. En general solen faltar només àtoms d'hidrogen, però s'escull fer-ho amb el VMD mitjançant uns arxius de topologia que l'ajuda a interpretar cada aminoàcid de la proteïna i determinar quins elements falten. Es poden utilitzar els convertidors en línia mencionats a la bibliografia [12], que afegeixen pel seu compte els hidrògens i altres àtoms que falten a la proteïna.

A més, un cop obtinguda aquesta nova estructura, s'utilitzarà el programa *PDB2PQR* que permet proporcionar una càrrega i un radi a cada element de la proteïna amb el pH desitjat. Amb tot això es pot procedir a utilitzar el programa APBS, que proporcionarà una imatge tridimensional del potencial a la superfície d'aquesta proteïna.

A part, convindria calcular la càrrega total de la proteïna, que es pot fer sumant la segona columna de l'arxiu PQR amb Excel o similar.

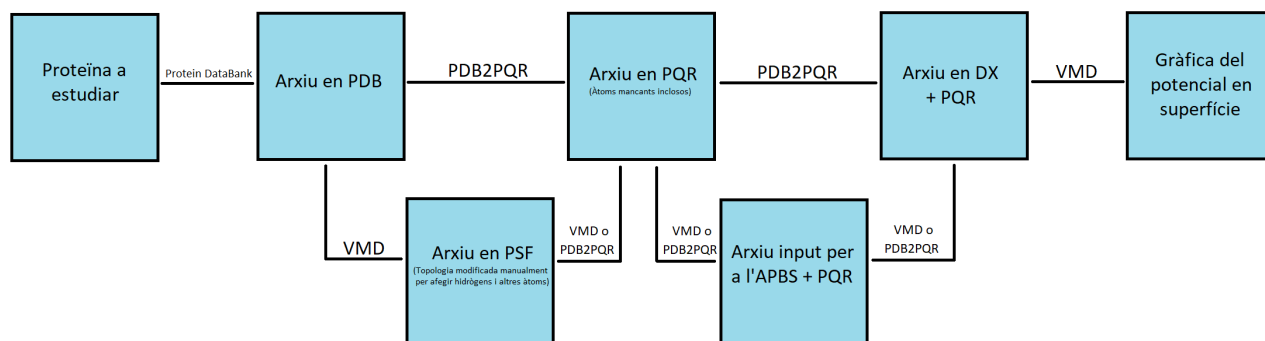


Figura 14: Procés seguit a la primera part.

La disposició dels arxius PQR és la següent: la primera columna especifica el tipus d'element (ATOM / HETATM), el segon l'ordre, el tercer la posició dins l'aminoàcid, i el quart l'aminoàcid al qual pertany. La cinquena columna menciona el chainID, a quina cadena pertany l'àtom. Les tres següents especifiquen les posicions X Y Z, i la següent és la càrrega, que en sumar tota la columna s'obté la càrrega total de la proteïna. Finalment l'última columna és el radi de cada àtom. Per a consultar els arxius utilitzats, es pot veure el GitHub a [13].

Els resultats per a la temperatura ambient del potencial a la superfície són:

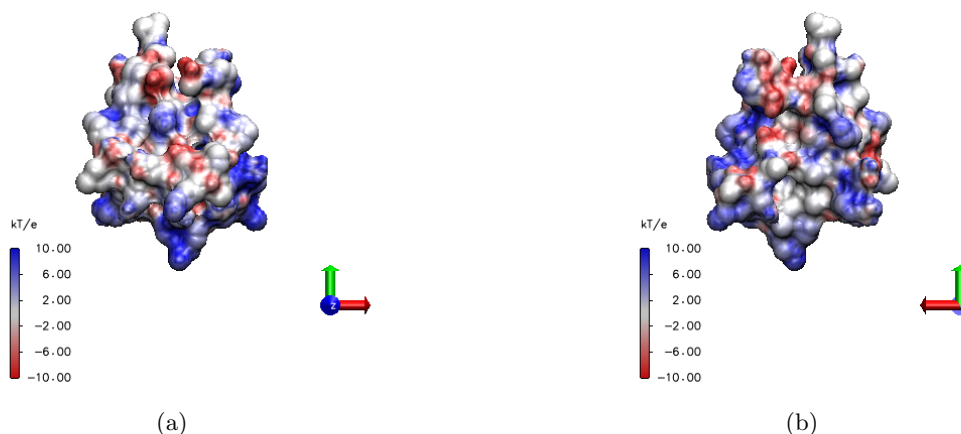


Figura 15: (a) Representació del potencial electrostàtic a la superfície de la proteïna CCL21 (5EKI) a una temperatura corporal de 298.65K. (b) Mateixa figura, rotada 180° a l'eix y. Els càlculs han estat realitzats amb APBS i les representacions amb VMD (Surf + Volume).

Pel que fa a l'heparina, es repeteix el mateix procés esmentat abans amb els arxius trobats a [14] dins la carpeta "*spikepbsamarolab.tar.gz*", amb l'arxiu "*complex.pqr*" trobat a "*pqrs*" dins *rbs_hep* i "*complex.in*" trobat a "*ins*". Tot això al convertidor en línia proporciona les imatges mencionades en l'apartat de l'heparina.

A.2.3 Preparació dels Arxius per a l'AutoDock

Descarregant els programes dels enllaços [15] i [16], s'hi troben les diferents eines que es faran servir durant l'estudi de l'acoblament molecular. Primer de tot, però, és essencial crear una carpeta per a destinar-hi tots els arxius que es generin durant els càlculs. En aquesta carpeta, a més, és interessant copiar-hi els arxius *AD4.1_bound.dat* i *AD4_parameters.dat* (que es troben dins la carpeta de MGL Tools i AutoDockTools) que serveixen per a que el programa pugui llegir i interpretar tots els àtoms i altres paràmetres relacionats amb la lectura d'estructures moleculars digitalitzades.

A més, en aquesta carpeta, també convé col·locar-hi els executables *autogrid4.exe* i *autodock4.exe*, que serviran per a més endavant.

Seguidament, és convenient executar OpenBabelGUI, (o PyMOL si es prefereix) per a transformar l'extensió de l'arxiu de provinença, en el nostre cas, l'heparina està en format pqr, i es vol en un que el programa AutoDockTools pugui llegir com a *ligand*, per exemple, com s'ha fet aquí, de pqr a pdb. Cal remarcar que s'ha escurçat l'heparina, ja que presentava més de 32 torsions, i el programa només en pot fer els càlculs per a aquesta xifra com a màxim. S'han seleccionat els residus de l'1 al 4, que presenten 27 torsions en total. S'han escollit basats en el gràfic del potencial a la superfície de la imatge c, on sembla ser més electronegativa.

Un cop obtingut el format desitjat i retallat de l'heparina, s'executa el programa AutoDockTools. Aquí, per simplificar la feina, es pot establir el directori a la carpeta creada fent servir ús de l'opció File;Set. Totes les instruccions d'interès es poden descarregar al manual que es troba al mateix enllaç que el programa AutoDock [15]. Un cop fet, es pot obrir el nostre receptor (5EKI) i començar a treballar amb ell. Per a evitar problemes, convé eliminar aigües, buscar i reparar àtoms que puguin faltar, i afegir els hidrògens polars que no presenti, per a juntar després els no-polars (merge non-polar). Tot aquests passos es troben dins la secció de Edit, al programa mateix. Finalment, cal computar les càrregues de Kollman, que en aquest cas és 8.688e, i repartir la càrrega total dels residus que pugui haver quedat mal repartida.

Ara, a la secció Ligand, es busca l'heparina que s'ha preparat. Repetint els mateixos passos que abans, tenint en compte que estigui ben seleccionada, aquest cop se seleccionen tots els hidrògens (All hydrogens)

i en comptes de computar Kollman, es computa Gasteiger. Després es retorna a Ligand i se selecciona (a Choose) el que hem preparat. Si tot surt bé, permetrà guardar-lo com a arxiu pdbqt fent clic a output.

Com que ja s'ha escurçat l'heparina i es tenen les dues molècules preparades, en el nostre cas no cal parar atenció a la secció de residus flexibles. Anant directament a Grid, se seleccionen una macromolècula receptora (5EKI) i un ligand (heparina). Ara s'obre la GridBox, que és un cub per a escollir la zona del receptor que interessa estudiar. En aquest treball interessen els residus del 70 al 76, pel que se seleccionen amb el ratolí i s'ajusta la capsa (gridbox) de forma que inclogui els 7 residus. En tancar, es marca l'opció File ¿ Save del Gridbox.

Ara es pot guardar la configuració del grid amb l'opció output. Un cop realitzats tots els passos anteriors, es procedeix a clicar Run. Allà se selecciona Autogrid, i es fa ús del directori per a seleccionar *autogrid4.exe* i el mapa del grid que s'acaba de guardar. Si tot va bé, generarà un arxiu per al docking (o acoblament) que interessa calcular.

Finalment doncs, a Docking, s'escull l'opció de Rigid Filename (volem el receptor rígid), i s'escull un ligand. A Search Parameters s'escull Genetic Algorithm. En tractar-se d'un estudi més detallat, s'ha marcat l'opció de "large" al nombre d'avaluacions. Ara se salva de la mateixa manera que el grid, un arxiu dock (amb l'opció Lamarckian). Després, pel mateix procediment que abans, a Run s'executa l'AutoDock. Això pot durar un bon temps, depenent de la precisió que s'hagi escollit i el nombre de torsions del ligand.

Quan finalment ha acabat el càlcul, l'ordinador genera un fitxer dlq, que conté els resultats de l'estudi. Allà es pot observar els diferents tipus d'interaccions que s'han hipotetitzat, i a la columna Analyze, es pot observar cada "Run", posició, ponts d'hidrogen, i els altres paràmetres i condicions desitjades.

Per al següent càlcul, corresponent a la zona de potencial electropositiu, se selecciona la zona d'interès amb el *grid* i es repeteix el procés anterior. Per als casos rígids, al menú del lligand se selecciona "escollir torsions" i es clicca l'opció fer totes les torcions no-rotables. Així ja tenim tots els casos llestos.

Un cop fets els arxius dlq, as'entra al menú Analyze (on s'han de reestablir lligand, receptor i grid) i s'escull el run amb l'energia lliure d'unió més elevada, després es clicca el botó "and" i s'activen els ponts d'hidrogen. Tot seguit es clicca l'opció de "save complex" per al receptor i "save current" per al lligand. Es guarden les imatges en el format més favorable (en el nostre cas pdbqt per a interpretar-ho amb PyMol) i s'obren totes juntes per a cada cas. A PyMol se selecciona mostrar els enllaços entre heparina i receptor (a ligand sites), i es poden eliminar aigües i hidrogens no polars en cas que n'hi hagi.

Amb això s'obtenen totes les imatges necessàries per cada cas. És interessant, a més, considerar les taules trobades per a cada interacció:

| Rank | Sub-Rank | Run | Binding Energy | Cluster RMSD | Reference RMSD |
|------|----------|-----|----------------|--------------|----------------|
| 1 | 1 | 3 | -24.14 | 0.00 | 127.61 |
| 1 | 2 | 4 | -22.27 | 1.61 | 127.70 |
| 2 | 1 | 2 | -23.48 | 0.00 | 127.50 |
| 3 | 1 | 5 | -22.93 | 0.00 | 127.11 |
| 4 | 1 | 7 | -22.82 | 0.00 | 128.53 |
| 5 | 1 | 10 | -18.86 | 0.00 | 128.92 |
| 6 | 1 | 9 | -18.34 | 0.00 | 130.22 |
| 7 | 1 | 8 | -18.29 | 0.00 | 129.39 |
| 8 | 1 | 1 | -17.82 | 0.00 | 126.86 |
| 9 | 1 | 6 | -16.40 | 0.00 | 127.95 |

Figura 16: Taula dels runs per a la interacció heparina (flexible) amb la zona de potencial electropositiu més elevat de la CCL21.

| Rank | Sub-Rank | Run | Binding Energy | Cluster RMSD | Reference RMSD |
|------|----------|-----|----------------|--------------|----------------|
| 1 | 1 | 6 | -18.88 | 0.00 | 124.50 |
| 1 | 2 | 9 | -18.88 | 0.01 | 124.50 |
| 1 | 3 | 5 | -18.88 | 0.04 | 124.52 |
| 1 | 4 | 2 | -18.88 | 0.04 | 124.52 |
| 1 | 5 | 3 | -18.88 | 0.05 | 124.52 |
| 1 | 6 | 10 | -18.88 | 0.02 | 124.50 |
| 1 | 7 | 7 | -18.88 | 0.05 | 124.53 |
| 1 | 8 | 1 | -18.88 | 0.02 | 124.51 |
| 1 | 9 | 8 | -18.87 | 0.01 | 124.50 |
| 1 | 10 | 4 | -18.87 | 0.03 | 124.52 |

Figura 17: Taula dels runs per a la interacció heparina (rígida) amb la zona de potencial electropositiu més elevat de la CCL21.

| Rank | Sub-Rank | Run | Binding Energy | Cluster RMSD | Reference RMSD |
|------|----------|-----|----------------|--------------|----------------|
| 1 | 1 | 1 | -3.26 | 0.00 | 123.43 |
| 1 | 2 | 3 | -3.26 | 0.01 | 123.43 |
| 1 | 3 | 8 | -3.26 | 0.01 | 123.43 |
| 1 | 4 | 5 | -3.26 | 0.01 | 123.43 |
| 1 | 5 | 6 | -3.26 | 0.01 | 123.43 |
| 1 | 6 | 10 | -3.26 | 0.01 | 123.43 |
| 1 | 7 | 2 | -3.26 | 0.01 | 123.44 |
| 1 | 8 | 4 | -3.26 | 0.01 | 123.43 |
| 1 | 9 | 7 | -3.26 | 0.02 | 123.43 |
| 1 | 10 | 9 | -3.26 | 0.01 | 123.44 |

Figura 18: Taula dels runs per a la interacció heparina (rígida) amb la terminal C de la CCL21.

Per altra banda, la zona d'interès del potencial de superfície ha estat la següent:

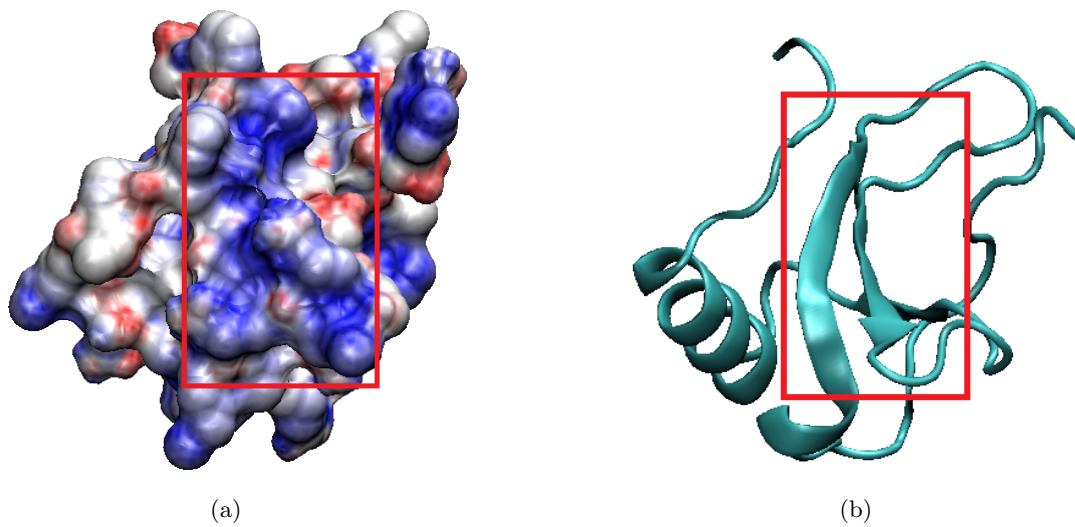


Figura 19: Zona d'interès per a col·locar el grid segons el potencial a la superfície més electropositiu. (a) Mapa de l'APBS. (b) Zona corresponent, aproximadament, en Cartoon.

A.3 Declaracions

A.3.1 Declaració d'autoria del Treball de Fi de Grau

Jo, Maximilià Soler Subirana, amb Document Nacional d'Identitat 21770721W, i estudiant del Grau en Física de la Universitat Autònoma de Barcelona, en relació amb la memòria del treball de final de Grau presentada per a la seva defensa i avaluació durant la convocatòria de Juliol del curs 2022-2023, declaro que

- El document presentat es original i ha estat realitzat per la meva persona.
- El treball s'ha dut a terme principalment amb l'objectiu d'avaluar l'assignatura de treball de grau en física en la UAB, i no s'ha presentat prèviament per a ser qualificat en l'avaluació de cap altra assignatura d'aquesta universitat ni de cap altra.
- En el cas de continguts de treballs publicats per terceres persones, l'autoria està clarament atribuïda, citant les fonts degudament.
- En el casos en els que el meu treball s'hagi realitzat en col·laboració amb altres investigadors i/o estudiants, es declara amb exactitud quines contribucions es deriven del treball de tercers i quines es deriven de la meva contribució.
- A l'excepció del punts esmentats anteriorment, el treball presentat és de la meva autoria.

Signat:

A.3.2 Declaració d'Extensió del Treball de Grau

Jo, Maximilià Soler Subirana, amb Document Nacional d'Identitat 21770721W, i estudiant del Grau en Física de la Universitat Autònoma de Barcelona, en relació amb la memòria del treball de final de Grau presentada per a la seva defensa i avaluació durant la convocatòria de Juliol del curs 2022-2023, declaro que:

- El nombre total de paraules (segons el comptatge proposat) incloses en les seccions des de la introducció fins a les conclusions és de X paraules.
- El nombre total de figures es de X .
- En total el document, comptabilitza: paraules + x 200 paraules/figura =
Que compleix amb la normativa en ser inferior a 10000.

Signat:

Llistat de figures

| | | |
|----|---|----|
| 1 | Enllaços rotals de la serotonina. Imatge d'ús públic sense copyright [18]. | 4 |
| 2 | (a) Esquema simplificat del macro-ió rodejat pel núvol carregat positivament. Podem observar les distribucions de càrrega esmentades anteriorment, recalcant que només hi ha pocs nm de separació entre superfície i càrregues lliures. (b) Dos macro-ions de símbol igual interactuant: hi ha repulsió entre ells, però els contraions es mantenen pressionats per ambdues bandes. (c) Quan dos macro-ions de símbol diferent s'acosten, els contraions s'ajunten i surten dispersats, guanyant entropia. | 6 |
| 3 | (a) Estructura de codi 5EKI, que presenta sis cadenes similars a causa del mètode d'obtenció de cristal·lografia de rajos X. Presenta 3639 àtoms, sense hidrògens. (b) Cadena E de l'estructura acabada de mencionar, que es considera el més bo (veure Annex) amb la terminal C d'aquest model, lloc crucial per a la interacció amb la proteïna CCR7. Presenta 583 àtoms, sense hidrògens. Els gràfics utilitzats per a l'estructura base són NewCartoon. A la primera fotografia utilitzem CPK per a veure els residus més clarament, i a la segona indiquem la terminal amb VDW colorats segons el tipus de residu. | 10 |
| 4 | (a) Representació del potencial electroestàtic a la superfície de la proteïna CCL21 (5EKI) a una temperatura corporal de 309.65K. (b) Mateixa figura, rotada 180° a l'eix y. Els càlculs han estat realitzats amb APBS (consultar Annex) i les representacions amb VMD (Surf + Volume). | 11 |
| 5 | (a) Estructura general de l'heparina [14], amb 203 àtoms, hidrògens inclosos. (b) Estructura girada a 180° sobre l'eix y. (Tant (a) com (b) estan representades en format CPK, amb el programa VMD) (c) Potencial a la superfície de l'heparina, a temperatura corporal (veure resta de condicions a l'apèndix). (d) Potencial superficial de l'heparina, a 180° en l'eix y. El color blau representa la tendència cap a càrregues positives i, el vermell, negatives. (Tant (c) com (d) estan representades en format Surf-Volume, al VMD). | 12 |
| 6 | Taula de les deu millors posicions trobades pel programa AutoDock al cas heparina flexible - cua CCL21. | 13 |
| 7 | (a) Configuració trobada per a l'heparina, flexible, a la zona de càrrega superficial positiva més elevada. (b) Configuració trobada per a l'heparina, flexible, a la cua de la proteïna CCL21 (residus 70-77). (c) Configuració trobada per a l'heparina, rígida, a la zona de càrrega superficial positiva més elevada. (d) Configuració trobada per a l'heparina, rígida, a la cua de la proteïna CCL21 (residus 70-77). Per a aquestes imatges ha estat escollit el millor angle per a visualitzar el màxim nombre d'enllaços presents, senyalitzats de color groc discontinu. | 14 |
| 8 | Estructura general del model 2L4N | 18 |
| 9 | Terminal C del model 24LN (59 a final) | 19 |
| 10 | Residus 9-11, 16, 19, 21, 36, 47-50, 52 i 53 del model 2L4N | 19 |
| 11 | Estructura general del model 5EKI | 20 |
| 12 | Cadena E del model 5EKI | 20 |
| 13 | Terminal C (70-79) de la cadena E (5EKI) | 21 |
| 14 | Procés seguit a la primera part. | 21 |
| 15 | (a) Representació del potencial electroestàtic a la superfície de la proteïna CCL21 (5EKI) a una temperatura corporal de 298.65K. (b) Mateixa figura, rotada 180° a l'eix y. Els càlculs han estat realitzats amb APBS i les representacions amb VMD (Surf + Volume). | 22 |
| 16 | Taula dels runs per a la interacció heparina (flexible) amb la zona de potencial electropositiu més elevat de la CCL21. | 23 |
| 17 | Taula dels runs per a la interacció heparina (rígida) amb la zona de potencial electropositiu més elevat de la CCL21. | 24 |
| 18 | Taula dels runs per a la interacció heparina (rígida) amb la terminal C de la CCL21. | 24 |
| 19 | Zona d'interès per a col·locar el grid segons el potencial a la superfície més electropositiu. (a) Mapa de l'APBS. (b) Zona corresponent, aproximadament, en Cartoon. | 25 |